JOURNAL OF PHYSICAL ORGANIC CHEMISTRY
J. Phys. Org. Chem. 2003; **16:** 811–817
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1003/poc.643

Nitrobenzene toxicity: QSAR correlations and mechanistic interpretations†

Alan R. Katritzky,[1]* Polina Oliferenko,[1] Alexander Oliferenko,[1] Andre Lomaka[1,2] and Mati Karelson[2]

[1]Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, PO Box 11720, Gainesville, Florida 32611-7200, USA
[2]Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu, 51014, Estonia

Received 8 January 2003; revised 10 March 2003; accepted 12 March 2003

ABSTRACT: The overall five-parameter QSAR correlation [$R^2 = 0.723$, $R^2_{cv} = 0.676$, $s = 0.42$ in terms of $\log(IGC_{50}{}^{-1})$] based on CODESSA-PRO methodology for the aquatic toxicity of 97 substituted nitrobenzenes to the ciliate *Tetrahymena pyriformis* supports previous conclusions that hydrophobicity and electrophilic reactivity control nitrobenzene toxicity. Correcting for the ionization of acidic species (picric and nitrobenzoic acids) improves the results: $R^2 = 0.813$, $R^2_{cv} = 0.787$, $s = 0.346$. Consideration of the total set of 97 compounds suggests two mechanisms of toxic action. A subset containing 43 compounds favorably disposed to reversible reduction of nitro group with respect to the single occupied molecular orbital energy, E^{SOMO} correlated well with just four theoretically derived descriptors: $R^2 = 0.915$, $R^2_{cv} = 0.890$, $s = 0.276$. Another set of 49 substances predisposed to aromatic nucleophilic substitution modeled well ($R^2 = 0.915$, $R^2_{cv} = 0.888$, $s = 0.232$) with five descriptors. Copyright © 2003 John Wiley & Sons, Ltd.

KEYWORDS: aromatic nitro compounds; modes of toxic action; QSAR/QSPR theoretical descriptors; CODESSA

INTRODUCTION

Nitroaromatics are hazardous chemicals that display several manifestations of toxicity, including skin sensitization,[1] immunotoxicity,[2] germ cell degeneration,[3] inhibition of liver enzymes[4] and also a conjectured carcinogenicity.[5] Nitrobenzene toxicity to the aquatic ciliate *Tetrahymena pyriformis* has been extensively studied by several groups of workers[1,6–8] with the use of 2D and 3D QSAR methodologies. Cronin *et al.*[1] showed that there are multiple modes of nitrobenzene toxic action: several factors are operative, with hydrophobicity and electrophilic reactivity being the most important. Hydrophobicity is considered to control transport from the medium to the site of action, whereas the electrophilicity is an intrinsic reactivity pattern.

Electrophilic reactivity of nitrobenzenes can be considered from two standpoints: (i) as due to nitro group reduction and (ii) as the tendency to act as an electrophile in S_NAr reactions.[1] In an attempt to quantify the reactivity in S_NAr reactions, Mekenyan *et al.*[9] discriminated between skin allergenic and non-allergenic species with the help of two quantum chemical descriptors, the energy of the lowest unoccupied molecular orbital (E^{LUMO}) and the difference in E^{LUMO} from the parent nitro compound to the corresponding anionic Meisenheimer complex (ΔE^{LUMO}).

The reduction of a nitro group can occur by at least two mechanisms: the single-step reduction with the nitroreductase and the so-called redox cycling, during which multiple back-oxidation of the reduced nitro compound occur.[10] To this end, Schmitt *et al.*[11] proposed an appropriate quantum chemical descriptor, the energy of the singly occupied molecular orbital (E^{SOMO}). According to their conclusion, compounds of high redox cycling ability fall into a well-defined window of E^{SOMO} variation, from 0.55 to −0.3 eV. We will show below that for nitrobenzenes, it is more appropriate to consider an upper limit of E^{SOMO} variation rather than a window.

The application of other theoretically derived descriptors to the modeling of nitrobenzene toxicity was exemplified recently by Agrawal and Khadikar,[12] who built multiple regression models based solely on topological descriptors. This limitation presumably emphasized bulk molecular properties related to the capability of a molecule to approach and associate with the binding sites while neglecting the specific reactivity of the compounds. On the other hand, failure to take account of molecular topology caused Cronin and co-workers[1,6] to postulate a separate toxicity mechanism for *para*-substituted nitrobenzenes, which they clarified as statistical outliers.

**Correspondence to:* A. R. Katritzky, Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, PO Box 11720, Gainesville, Florida 32611-7200, USA.
E-mail: katritzky@chem.ufl.edu
†Dedicated to Professor Marek Krygowski on the occassion of his 65th birthday.

For a set of 39 monosubstituted nitrobenzenes the best five-parameter equation produced by Agrawal and Khadikar[12] had the following statistical characteristics: $R^2 = 0.801$; $R^2_{cv} = 0.771$, $s = 0.252$, $F = 27$. However, this model includes (a) the Szeged index and the PI index with mutual intercorrelation of 0.989 and (b) a pair of indicator variables Ip2 and Ip3 intercorrelated to the extent of 0.622.

Warne *et al.*[13] used a more extended set of descriptors including electronic and thermodynamic features, and Fukui-type atomic properties (superdelocalizabilities, electronic densities of FMO). Principal component analysis together with multiple regression analysis suggested general narcosis, polar narcosis and uncoupling toxicity as the three main modes of substituted halobenzene toxicity against *Vibrio fisheri* for a restricted set of 19 halonitrobenzenes ($R^2 = 0.86$). The effect of different chemical narcotics on *Tetrahymena pyriformis* was investigated by Bearden and Schultz.[14] Various aromatics display distinct types of narcosis: the toxicity of aromatics with strong electron-releasing amino and hydroxy groups was explained by polar narcosis mechanism.[15,16]

Estrada and Uriarte[17] applied their original Topological Sub-Structural Molecular Design (TOPS-MODE) approach based on topological descriptors to a data set of 43 substituted nitrobenzenes. The equation obtained ($R^2 = 0.901$, $R^2_{cv} = 0.900$, $s = 0.22$) utilizes four spectral moment variables composed in turn of 17 sub-structural and two graph-theoretical fragments. Although mechanistic interpretation of the correlation is complex, it can be used for the prediction of molecular toxicity through the summation of structural group toxicity contributions.

In our CODESSA-PRO approach we can avoid significant descriptor intercorrelations; our usual cutoff for this value is 0.5. The descriptor pool of our new Windows software, CODESSA-PRO, possesses hundreds of constitutional, topological, geometric, electrostatic and quantum chemical descriptors. We have successfully applied our methodology to the modeling of diverse physical properties and of chemical reactivity.[18,19] Aspects of the toxicity and genotoxicity of aromatic species were recently investigated with the CODESSA approach.[20,21]

The aim of this work was to establish reliable QSAR models of nitrobenzene toxicity and to throw light on the mechanisms of action of the title compounds. The study outlined below consists of three parts. First, we constructed a set of correlations on all 97 nitro compounds, which produced squared correlation coefficients varying from 0.66 for three-parameter to 0.815 for five-parameter models.

We next treated the toxicity as a multi-dimensional activity, as the multiple modes of nitrobenzene toxic action have been pointed out many times.[9,11] We partitioned the whole database of 97 nitrobenzenes into two overlapping clusters (total 62 compounds) based on mechanistic considerations: (i) for 43 compounds causing the appearance of the oxidative stress in a living cell (due to the redox cycling while nitro group reduction) and (ii) for 49 species that are predisposed to nucleophilic attack. We also modeled the 35 compounds belonging to neither cluster.

DATA AND EXPERIMENTAL

The training sets for the present investigation were created from two data sets containing nitrobenzenes, described by Agrawal and Khadikar,[12] and Cronin and Schultz.[8] Molecules were modeled using the MM+ method of Hyperchem. Final optimizations were performed with the MOPAC computer program[22] using the AM1 semiempirical method.[23] Constitutional, topological, geometrical, charge-related, semiempirical and molecular-, atomic-, bond-type descriptors were calculated with the CODESSA-PRO software package.[24] The AM1 semiempirical method of Hyperchem was also used for the calculation of the energy of the singly occupied molecular orbital of radical anions generated by one-electron reduction E^{SOMO} of the restricted Hartree–Fock open-shell method.

RESULTS

The data set of 97 nitrobenzenes includes recent data on toxicity[8] as summarized in Table 1. Descriptor definitions are referenced in the Discussion section. Our overall five-descriptor correlation [Eqn (1)] had $R^2 = 0.724$, $R^2_{cv} = 0.676$, $F = 48$, $s = 0.42$.

$$\begin{aligned}\log 1/IGC_{50} = & -28.428(\pm 2.856) \\ & + 0.109(\pm 0.011) E^{\min}_{ne}(\mathrm{C-C}) \\ & + 0.395(\pm 0.068)^2 \chi \\ & + 0.015(\pm 0.003) WPSA^{(1)} \\ & - 10.707(\pm 3.108) FPSA^{(3)} \\ & + 4.598(\pm 1.164) HASA^{(1)}_{TMSA} \end{aligned} \quad (1)$$

The degree of deprotonation in aqueous solutions depends on pK_a. Hence, in the second treatment, nitrobenzoic and picric acids with highly negative pK_a were substituted in the data set by their anions. This resulted in a substantially improved five-parametered equation [Eqn (2–1)]: $R^2 = 0.815$, $R^2_{cv} = 0.789$, $s = 0.348$, and also the disappearance of picric and nitrobenzoic acids from the outliers. However, five outliers remain in the model: 6-bromo-1,3-dinitrobenzene, 4-chloro-2-nitrophenol, 2,6-dichloro-4-nitrophenol, 3,4-dinitrophenol, 4-methyl-3-nitrophenol (outliers, nevertheless, are included in all our calculations of R^2, etc.). The plot of predicted versus experimental values of $\log 1/IGC_{50}$ is exemplified in Fig. 1. The best three-parameter equation [Eqn (2–2)] had $R^2 = 0.754$, $R^2_{cv} = 0.736$, $s = 0.348$.

Table 1. Experimental and predicted toxicities to *Tetrahymena pyriformis* and descriptor values of nitrobenzenes[a]

ID	Name	E_{SOMO}	Log1/IGC_{50}					
			Exp.	Eqn (2-1)	Eqn (3-1)	Eqn (4)	Eqn (5)	Eqn (6)
1	Nitrobenzene	1.13	0.14	0.31	−0.04	0.16	**0.11**	0.25
2	2-Nitrophenol	0.97	0.67	0.43	−0.13	0.66	**0.78**	0.39
3	2-Nitrotoluene	1.12	0.05	0.18	−0.13	0.68	**0.15**	0.11
4	2-Chloronitrobenzene[b]	0.86	0.68	0.74	0.22	**0.80**	0.64	0.69
5	2-Bromonitrobenzene[b]	0.79	0.75	0.80	0.54	**0.84**	0.74	**0.76**
6	2-Nitrobenzyl alcohol	0.84	−0.16	−0.07	−0.79	0.01	0.02	−0.13
7	2-Nitrobiphenyl	0.63	1.3	1.38	0.15	1.97	**1.69**	1.40
8	3-Nitroaniline	1.21	0.03	0.32	−0.98	0.04	**0.35**	0.27
9	3-Nitrophenol	0.96	0.51	0.62	−0.71	−0.12	**0.58**	0.59
10	3-Nitrotoluene	1.12	0.05	0.40	0.13	0.66	**0.13**	0.35
11	3-Chloronitrobenzene[b]	0.81	0.73	0.70	0.48	**0.75**	0.59	0.67
12	3-Nitrobenzonitrile[b]	0.61	0.45	0.64	0.65	**0.56**	−0.15	0.61
13	3-Nitrobenzaldehide[b]	0.71	0.14	0.31	0.48	**0.31**	0.33	0.26
14	1,3-Dinitrobenzene[b,c]	0.24	0.89	1.17	**1.07**	**0.97**	**−0.47**	1.17
15	3-Nitroanisole[b]	0.98	0.67	0.95	0.24	**0.57**	**0.99**	0.93
16	4-Nitrotoluene	1.05	0.17	0.28	0.26	0.62	0.10	0.22
17	4-Ethylnitrobenzene	1.03	0.8	0.25	0.44	0.92	0.25	0.19
18	4-Nitroanizole[b]	1.08	0.54	1.05	0.35	**0.42**	1.26	1.03
19	4-Chloronitrobenzene[b]	0.77	0.43	0.75	0.78	**0.72**	0.66	0.71
20	4-Bromonitrobenzene[b]	0.66	0.38	0.77	0.93	**0.84**	0.78	0.74
21	4-Nitrobenzyl alcohol	0.92	0.1	0.12	−0.33	0.01	**0.21**	0.07
22	4-Nitrobenzamide[b]	0.33	0.18	−0.12	−0.27	**−0.16**	−0.18	−0.18
23	4-Nitrobenzaldehyde[b,c]	0.18	0.2	0.33	**0.60**	**0.43**	0.40	**0.28**
24	2-Nitrobenzoic acid[c]	0.46	−1.64	−1.16	**−1.57**	0.03	0.09	−1.27
25	3-Nitrobenzoic acid	0.66	−1.09	−1.27	−0.59	−0.16	**−0.75**	−1.38
26	4-Nitrobenzoic acid[c]	0.11	−0.86	−1.17	**−0.97**	0.05	−0.05	**−1.29**
27	2-Nitrobenzamide[b,c]	0.41	−0.72	−0.21	**−0.61**	**−0.31**	0.02	−0.27
28	3-Nitrobenzyl alcohol	1.1	−0.22	0.07	−0.74	0.12	**0.20**	0.02
29	3-Nitrobenzamide[b]	0.76	−0.19	−0.16	−0.23	**−0.37**	0.23	−0.22
30	2,4,6-Trinitrophenol	3.56	−0.16	0.00	2.04	2.15	**−0.27**	0.02
31	4-Nitrophenylacetonitrile	0.65	0.13	0.35	0.43	0.70	**−0.19**	0.30
32	2-Nitrobenzaldehyde[b,c]	0.37	0.17	0.07	**0.32**	**0.43**	0.36	0.01
33	4-Fluoronitrobenzene	0.87	0.25	0.33	0.54	**0.06**	0.20	0.28
34	3,4-Dinitrophenol[b,c]	−0.1	0.27	1.25	**0.57**	**0.42**	0.23	**1.28**
35	3-Nitroacetophenone[b]	0.76	0.32	0.54	0.51	**0.37**	0.28	0.51
36	5-Hydroxy-2-nitrobenzaldehyde[b,c]	0.52	0.33	0.36	**−0.01**	**0.14**	0.86	0.34
37	2,3-Dinitrophenol[b,c]	−0.2	0.46	1.06	**0.57**	**0.86**	0.24	1.08
38	2-Amino-4-nitrophenol	1.14	0.48	0.52	−1.46	−0.56	**0.64**	0.50
39	3,5-Dinitrobenzyl alcohol[c]	0.24	0.53	0.70	**0.53**	1.21	−0.50	0.71
40	2,6-Dinitrophenol[b,c]	0.12	0.54	0.86	**0.73**	**0.87**	1.04	0.88
41	4-Nitrobenzonitrile[b,c]	0.25	0.57	0.63	**0.76**	**0.60**	−0.07	0.59
42	4-Methyl-2-nitrophenol	0.96	0.57	0.83	−0.13	0.80	**0.84**	0.82
43	2,6-Dichloro-4-nitrophenol[c]	0.53	0.63	1.51	**0.98**	1.20	2.12	1.52
44	2-Chloro-6-nitrotoluene	0.79	0.68	0.55	0.54	1.43	**0.64**	0.50
45	Ethyl-4-nitrobenzoate[b,c]	0.14	0.71	0.72	**0.94**	1.86	0.55	0.71
46	4-Methyl-3-nitrophenol	0.93	0.74	0.83	−0.66	0.43	**0.67**	0.82
47	2-Chloromethyl-4-nitrophenol	0.74	0.75	0.94	−0.03	0.86	**0.65**	0.93
48	4-Chloro-2-nitroyoluene	0.8	0.82	0.61	0.61	1.47	**0.70**	0.57
49	4-Nitrophenetole[b]	1.08	0.83	1.05	0.53	**0.80**	1.44	1.03
50	2,4,6-Trimethylnitrobenzene	1.03	0.86	0.89	0.23	1.45	**0.76**	**0.86**
51	6-Methyl-1,3-dinitrobenzene[c]	0.22	0.87	1.06	**1.23**	1.58	0.03	1.06
52	4-Amino-2-nitrophenol	1.2	0.88	0.85	−1.72	−0.45	**1.10**	**0.85**
53	2,5-Dinitrophenol[b,c]	−0.5	0.95	0.98	**0.95**	**1.14**	0.10	1.00
54	2,4-Dichloronitrobenzene[b]	0.52	0.99	1.13	1.09	**1.07**	**1.49**	1.12
55	3-Bromonitrobenzene	0.77	1.03	0.73	0.63	0.94	**0.73**	0.70
56	2,3-Dichloronitrobenzene	0.59	1.07	1.13	0.84	**1.15**	**1.09**	1.10
57	2-Nitrobenzonitrile[b,c]	0.48	1.08	0.51	**0.68**	**0.67**	0.95	**0.47**
58	2,4-Dinitrophenol[b,c]	0.17	1.08	0.92	**0.96**	**0.75**	0.69	0.93
59	3,4-Dinitrobenzyl alcohol	1.08	1.09	1.05	0.53	1.47	**1.40**	1.03
60	5-Fluoro-2-nitrophenol[b]	0.71	1.13	0.61	0.33	**0.68**	0.78	0.61

Continues

Table 1. Continued

ID	Name	E_{SOMO}	Log1/IGC_{50}					
			Exp.	Eqn (2-1)	Eqn (3-1)	Eqn (4)	Eqn (5)	Eqn (6)
63	3-Methyl-4-bromonitrobenzene[b]	0.67	1.16	0.71	0.85	**1.27**	0.66	0.67
64	3,4-Dichloronitrobenzene[b,c]	0.5	1.16	1.12	**1.27**	**1.06**	1.04	1.10
65	2-Amino-4-chloro-5-nitrophenol	0.99	1.17	0.78	−0.73	0.30	**1.48**	0.77
66	4-Nitrobenzyl chloride[b,c]	−0.22	1.18	0.89	**1.01**	**1.15**	0.00	0.87
67	2,6-Dinitro-4-cresol[c]	0.12	1.23	1.30	**1.24**	2.03	1.05	1.34
68	1,2-Dinitrobenzene[b,c]	−0.05	1.25	1.19	**0.85**	**1.00**	−0.34	1.20
69	1,4-Dinitrobenzene[b,c]	−0.38	1.3	1.37	**1.22**	**1.16**	−0.41	1.38
70	2,6-Dibromo-4-nitrophenol[c]	0.49	1.35	1.39	**1.34**	1.33	2.52	1.41
71	2,5-Dibromonitrobenzene[b,c]	0.45	1.37	1.12	**1.33**	**1.42**	1.53	1.11
72	4-Nitrophenol	1.09	1.42	0.75	−0.56	−0.17	**0.96**	0.73
73	4-Butoxynitrobenzene[b]	1.07	1.42	1.35	1.04	**1.53**	1.70	1.35
74	2,4,6-Trichloronitrobenzene[b,c]	0.28	1.43	1.71	**1.42**	**1.23**	1.61	1.73
75	2,3,4-Trichloronitrobenzene[b,c]	0.29	1.51	1.42	**1.47**	**1.42**	1.89	1.41
76	5-Methyl-1,2-dinitrobenzene[b,c]	−0.06	1.52	1.33	**1.41**	**1.50**	−0.03	1.35
77	2,4,5-Trichloronitrobenzene[b,c]	0.26	1.53	1.60	**1.75**	**1.47**	2.17	1.60
78	3-Nitrobiphenyl	0.77	1.57	1.25	0.26	2.31	**1.24**	1.26
79	2-Chloro-4-nitrophenol	0.8	1.59	1.16	0.24	0.76	**1.50**	1.16
80	4-Chloro-6-nitromcresol	0.76	1.64	1.48	−0.05	1.08	**1.81**	1.49
81	2,6-Diiodo-4-nitrophenol[c]	0.48	1.71	1.84	**1.73**	1.41	2.89	**1.87**
82	4,6-Dinitro-2-cresol[c]	0.31	1.72	1.35	**1.03**	1.66	0.78	1.39
83	3-Methyl-4-nitrophenol	1.08	1.73	0.90	−0.59	0.15	**1.06**	0.88
84	2,4-Chloro-6-nitrophenol[c]	0.38	1.75	1.37	**1.27**	1.80	2.11	1.37
85	2,3,4,5-Tetrachloronitrobenzene[b,c]	0.05	1.78	2.06	**2.03**	**1.64**	2.42	2.08
86	2,3,5,6-Tetrachloronitrobenzene[b,c]	0.1	1.82	2.14	**1.76**	**1.65**	2.30	2.17
87	4-Nitrodiphenylamine	0.6	1.89	1.67	0.51	2.54	**1.79**	1.70
88	4-Chloro-2-nitrophenol	0.65	2.05	0.73	0.34	1.15	**1.41**	**0.72**
89	6-Iodo-1,3-dinitrobenzene[b,c]	−0.2	2.12	1.68	**2.03**	**1.79**	0.16	1.71
90	2,4,6-Trichloro-1,3-dinitrobenzene[b,c]	−0.23	2.19	2.45	**2.42**	**2.40**	1.33	**2.52**
91	1,2-Dinitro-4,5-dichlorobenzene[b,c]	−0.55	2.21	2.22	**2.27**	**2.29**	0.66	2.27
92	6-Bromo-1,3-dinitrobenzene[b,c]	−0.17	2.31	1.57	**1.95**	**1.98**	0.12	1.60
93	2,4,5-Trichloro-1,3-dinitrobenzene[b,c]	−0.42	2.59	2.40	**2.55**	**2.62**	1.29	2.46
94	4,6-Dichloro-1,2-dinitrobenzene[b,c]	−0.51	2.42	2.21	**2.31**	**2.20**	0.74	2.26
95	2,3,5,6-Tetrachloro-1,4-dinitrobenzene[b,c]	−1.09	2.74	2.80	**2.80**	**3.01**	1.46	2.88
96	1,3-Dimethyl-2-nitrobenzene	1.06	0.3	0.51	0.09	1.13	0.45	0.46
97	2,3-Dimethylnitrobenzene	1.09	0.56	0.49	0.13	1.11	0.30	0.45

[a] Predicted values of toxicity were obtained by external validation and regression analysis; those calculated by the latter procedure are given in bold.
[b] Nitrobenzenes included in the sub-set of compounds predisposed for nucleophilic attack.
[c] Nitrobenzenes included in the sub-set of redox cyclers.

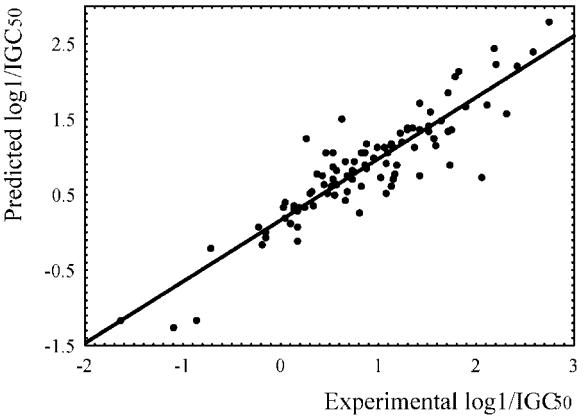

Figure 1. Plot of the predicted log1/IGC_{50} vs experimental log1/IGC_{50} for the whole data set of 97 nitrobenzenes of Eqn (2-1)

$$\log 1/IGC_{50} = 13.154(\pm 4.779) + 0.060(\pm 0.007)E_{en}^{min}(C-C) + 0.605(\pm 0.059)^{2}\chi - 0.165(\pm 0.070)E^{SOMO} - 2.731(\pm 0.379)\overline{V}_{O} + 4.654(\pm 0.798)FNSA_{PNSA}^{(2)} \quad (2\text{-}1)$$

$$\log 1/IGC_{50} = 605.191(\pm 88.849) - 1.150(\pm 0.167)E_{en}^{max}(N-O) + 0.562(\pm 0.079)^{1}\chi^{\nu} + 0.313(\pm 0.080)E_{C}^{min}(C-C) \quad (2\text{-}2)$$

As already mentioned, the reduction of the nitro group is expected to be one manifestation of the toxicity of substituted nitrobenzenes, as the energy of the singly

occupied molecular orbital E^{SOMO} of the corresponding radical anion generated by the one-electron reduction is a characteristic of the radical stability. Nitroaromatic compounds are a group of substances that may cause oxidative stress in living cells because of redox cycling. It has been shown[11] that for a number of aromatic redox cyclers, including two nitrofurans, four polycyclic aromatic quinones, *p*-nitrobenzoic acid and tetramethylbenzoquinone, an E^{SOMO} of -0.30 to 0.55 could tentatively indicate redox cycling ability. However, we have now found by means of E^{SOMO} variation that better results [Eqn (3-1)] for nitrobenzenes can be obtained if a lower limit of -1.09 for the E_{SOMO} value is applied: for this subset of 43 compounds $R^2 = 0.933$ $R^2_{cv} = 0.912$, $s = 0.249$, with 4,6-dinitro-2-cresol as a sole outlier. The two-parameter equation [Eqn (3-3)] also provides the reasonable statistics: $R^2 = 0.828$, $R^2_{cv} = 0.802$, $s = 0.383$.

$$\begin{aligned}\log 1/IGC_{50} = &-1.899(\pm 0.197)\\ &+ 0.007(\pm 0.001)^{HA}PSA^{(2)}\\ &- 0.447(\pm 0.056)E^{LUMO}\\ &+ 0.005(\pm 0.001)WPSA^{(2)}_{PPSA}\\ &+ 19.431(\pm 8.225)R^{max}_{C}\\ &- 158.175(\pm 21.029)^{HD}FCPSA^{(2)} \qquad (3\text{-}1)\end{aligned}$$

$$\begin{aligned}\log 1/IGC_{50} = &-4.416(\pm 0.758) - 0.760(\pm 0.069)^2\chi\\ &- 0.521(\pm 0.050)E^{LUMO}\\ &- 10.800(\pm 1.679)^{HD}FPSA^{(2)}\\ &+ 0.385(\pm 0.099)E_C \qquad (3\text{-}2)\end{aligned}$$

$$\begin{aligned}\log 1/IGC_{50} = &-25.158(\pm 2.216)\\ &+ 0.088(\pm 0.009)E^{min}_{en}(C-C)\\ &+ 0.052(\pm 0.007)\alpha \qquad (3\text{-}3)\end{aligned}$$

In addition to free radical reduction, some nitrobenzenes can undergo S_NAr nucleophilic attack by low molecular proteins containing soft nucleophiles such as the amino group of lysine or the sulfhydryl group of cysteine.[9] To investigate the correlation between the predisposition of nitrobenzenes towards nucleophilic substitution and their toxicity, nucleophilic nitrobenzenes were transformed from the whole set to a subset comprising 49 molecules which provided the correlation [Eqn (4)]: $R^2 = 0.915$, $R^2_{cv} = 0.888$, $F = 93$, $s = 0.232$, with no outliers. The substances which could undergo a nucleophilic attack were chosen according to the classical criteria outlined in[25] (i) the presence of a 'good' leaving group (NO_2, F, Cl) and (ii) the presence of electron-withdrawing groups (NO_2, CN, COH, COOH, $CONR_2$).

$$\begin{aligned}\log 1/IGC_{50} = &-13.633(\pm 4.498) + 0.096(\pm 0.010)N^{occ}\\ &+ 8.887(\pm 2.003)n^{max}_{A}\\ &- 163.417(\pm 18.417)^{HD}FCPSA^{(2)}\\ &- 0.543(\pm 0.147)\Delta E^{LUMO}_{HOMO}\\ &+ 1.971(\pm 0.377)RNCG \qquad (4)\end{aligned}$$

Finally, a correlation equation [Eqn (5)] ($R^2 = 0.819$, $R^2_{cv} = 0.757$, $s = 0.316$) was derived for 35 nitrobenzenes included in neither of the subsets and was used in cross-validation testing. This correlation has two outliers: 3-methyl-4-nitrophenol and 4-chloro-2-nitrophenol.

$$\begin{aligned}\log 1/IGC_{50} = &\ 12.174(\pm 4.305)\\ &+ 0.013(\pm 0.002)PPSA^{(1)}\\ &- 0.083(\pm 0.017)E^{min}_{ee}(O)\\ &+ 40.008(\pm 6.216)N^{rings}/N_A\\ &+ 2.293(\pm 0.902)\bar{P}_C \qquad (5)\end{aligned}$$

A cross-validation test was performed for 10 randomly chosen compounds on the basis of the model built for the remaining 87 compounds. The model obtained [Eqn (6)] is characterized with slightly better squared correlation coefficient, $R^2 = 0.852$, and includes the same descriptors as the equations discussed above.

$$\begin{aligned}\log 1/IGC_{50} = &\ 13.841 + 0.063E^{min}_{en}(C-C)\\ &+ 0.642^2\chi - 0.169E^{SOMO}\\ &- 2.886\overline{V}_O + 4.625FNSA^{(2)} \qquad (6)\end{aligned}$$

Overall toxicity values of nitrobenzene predicted by Eqns (2-1), (3-1), (4), (5) and (6) are listed in Table 1.

DISCUSSION

Analysis of the descriptors used in Eqns (1)–(5) shows close agreement of the results obtained with the known modes of nitrobenzene toxicity. In general terms, these correlation equations contain two types of descriptors: those describing molecular bulk properties (including transport properties such as membrane permeability) and those representing chemical reactivity of the substance under study. Molecular polarizability and branching are directly related to hydrophobicity,[26] so the semiempirically derived molecular dipole polarizability, α, and topological indices such as those of Kier and Hall ($^m\chi^\nu$) and Randic ($^m\chi$) make significant positive contributions to the target toxicity. Increasing the molecular

surface area, and thus revealing the sites of hydrogen bond donation, should also increase hydrophilicity.

The patterns of chemical reactivity of the nitrobenzenes are expressed in the present QSAR correlation in terms of molecular, bond and atomic characteristics. The HOMO–LUMO energy gap, $\Delta E_{\mathrm{HOMO}}^{\mathrm{LUMO}}$, and the total molecular electrostatic interaction, E_{C}, account for general stability of a molecule. Quantum chemical descriptors such as the number of occupied electronic levels, N^{occ}, and the maximum atomic orbital electronic population, $n_{\mathrm{A}}^{\mathrm{max}}$, reflect the distribution of electron density within a molecule and its concentration at a particular atom, respectively. Descriptor $n_{\mathrm{A}}^{\mathrm{max}}$ effectively divides the data set into those containing halogen and others. Halogen substituents can donate electron density through the π-system of an aromatic ring to the nitro group, which increases the reduction potential of the latter. Such molecular features as the partial positive charged surface molecular area [$PPSA^{(1)}$] and the fractional negative charged surface molecular area [$FNSA^{(2)}$] increase toxicity. The highest values for these descriptors are found for halonitrobenzenes. These descriptors are thus indicators of the susceptibility to aromatic nucleophilic substitution. Similar considerations should apply to the weighted and fractional positively charged surface molecular areas, $WPSA^{(1)}$ and $FPSA^{(3)}$, respectively.

Bond characteristics such as the electron–nuclear attraction for the C—C, $E_{\mathrm{en}}^{\mathrm{min}}$(C—C) and N—O bonds, $E_{\mathrm{en}}^{\mathrm{max}}$(N−O), the electron–electron repulsion, $E_{\mathrm{ee}}^{\mathrm{min}}$(C−C), and the Coulombic interaction of the C—C bond, $E_{\mathrm{C}}^{\mathrm{min}}$(C−C), reflect the strength of these bonds, and thus their resistance towards attack. The electron–nuclear attraction for the N—O bond correlates the toxicities of the full 97 nitrobenzene data set and the sub data set of 43 nitrobenzenes assumed to be redox cyclers with the highest values of R^2 (0.601 and 0.706, respectively), thus revealing radical reduction of the nitro group as a main toxic manifestation. In contrast to the decrease in toxicity caused by increasing values of the C—C bond descriptors, an increase in the N—O bond stability increases toxicity. Combined with the correlation of toxicity with the stability of the nitrobenzene anion radicals (in terms of E^{SOMO}), this illuminates the mechanism of the nitroreductase reduction of the nitro group. Stepwise enzymatic reduction of a strong N—O bond through the unstable anion radical intermediate could facilitate the formation of the corresponding hydroxylamine. By contrast, a more chemically active substrate having weaker C—C bonds and low value of E^{SOMO}, could leave the enzyme active site already after the first reduction step as a nitroxyl radical. Whereas hydroxylamine is a moderately active metabolic intermediate, nitroxyl radicals aggressively attack lipids and vigorously react with free oxygen and metal ions (Cu^{2+}), causing oxidative stress[10] and DNA damage.[5]

Atomic characteristics, including the electron–electron repulsion for an oxygen atom, $E_{\mathrm{ee}}^{\mathrm{min}}$(O), and the one-electron reactivity index for a carbon atom, $R_{\mathrm{C}}^{\mathrm{max}}$, measure intrinsic reactivity in radical reactions. Both $E_{\mathrm{ee}}^{\mathrm{min}}$(O) and E^{LUMO} reflect the direct dependence between toxicity and the electron attraction of an atom in a molecule, whereas $R_{\mathrm{C}}^{\mathrm{max}}$ suggests the involvement of nitrobenzenes into the radical coupling. An atomic property such as the relative negative charge, *RNCG*, models polar interactions between species.[27] Finally, the average valency for oxygen atom, $\overline{V}_{\mathrm{O}}$, together with such molecular features as the hydrogen acceptor partial surface area, ${}^{\mathrm{HA}}PSA^{(2)}$, and the hydrogen acceptor surface area divided by the total molecular surface area, $HASA_{TMSA}^{(1)}$, indirectly relate to the number of oxygen atoms in a compound: nitrobenzenes bearing additional nitro groups are more toxic. ${}^{\mathrm{HA}}PSA^{(2)}$ and $HASA^{(1)}$ can also be interpreted in relation to the polar narcosis type of action; aromatic molecules with hydrogen acceptor groups are readily incorporated into lipid membrane bilayers leading to biological disfunction. By contrast, hydrogen donor abilities of nitroaromatics do not increase narcosis manifestation since the fractional charge weighted partial surface area of hydrogen donors, ${}^{\mathrm{HD}}FCPSA^{(2)}$, has a negative contribution to the total toxicity. More likely this descriptor can be related to the hydrophilicity of a molecule.

The average bond order for a carbon atom, $\overline{P}_{\mathrm{C}}$, together with the relative number of rings, $N^{\mathrm{rings}}/N_{\mathrm{A}}$, represent saturation and bulk properties and can be indirectly associated with the molecular hydrophobicity.

To reveal mechanisms of toxicity by construction of theoretically derived particular correlations, individual QSAR analyses were performed for nucleophilic nitrobenzenes and for potential redox cyclers. The subset of 49 nucleophilic nitrobenzenes selected according to classical criteria provided a good correlation of toxicity with the above-discussed CODESSA descriptors. This is also the case for another subset containing 43 nitrobenzenes, which react predominantly as promoters of the oxidative stress. There is considerable overlapping between these two subsets, namely by 30 compounds. This suggests that there are no completely separate modes of action since the toxicants in a living cell undergo different paths of biotransformation so that they can overlap in the models discussed.

Correlation of the toxicity values of the 35 compounds that do not belong either to the Redox Cycling Model or the Nucleophilic Substitution Model gives an equation that shows no fundamental difference to the others, except for a lower value of R^2. The types of descriptors used are virtually the same as in the models discussed above.

The results of cross-validation are shown in the last column of Table 1. With the exception of compounds 34 (3,4-dinitrophenol), 88 (4-chloro-2-nitrophenol) and perhaps 57 (2-nitrobenzonitrile), all the compounds are predicted reasonably. In general, the prediction error for the selected compounds does not exceed that of the general QSPR model for the whole of 97 compounds. Based on the cross-validation results, we can conclude

that this general QSPR model [Eqn (6)] can be used as a reliable and efficient predictive tool for the preliminary evaluation of the toxicity of nitrobenzenes.

para-Substituted species turned to show as outliers in our general model as well as in the reduced correlations. As noted by Schüürmann *et al.*,[7] *para*-substituted nitro compounds are special owing to strong conjugation through the benzene ring. Nonetheless, in our models the above compounds are not so strong outliers as in the previous studies,[6,7] possibly because we explicitly accounted for specific features of the molecular structures with the help of topological descriptors.

CONCLUSIONS

Our QSPR model for 97 compounds corrected for acid ionization based on five CODESSA descriptors has $R^2 = 0.815$, $R^2_{cv} = 0.789$, $s = 0.348$.

Two specific toxicity mechanisms, (i) the nitro group reduction followed by the redox cycling and (ii) the nucleophilic S_NAr interaction with soft endogenic nucleophiles, are clearly expressed in terms of charge and quantum chemical descriptors. The SOMO energy variation boundaries applicable to the redox cycling properties of the nitrobenzene series of toxic aromatic compounds are refined with the help of the QSPR approach.

REFERENCES

1. Cronin MTD, Gregory BW, Schultz TW. *Chem. Res. Toxicol.* 1998; **11**: 902–908.
2. Li Q, Minami M, Hanaoka T, Yamamura Y. *Toxicol.* 1999; **137**: 35–45.
3. Shinoda K, Mitsumori K, Yasuhara K, Uneyama C, Onodera H, Takegawa K, Takahashi M, Umemura T. *Arch. Toxicol*. 1998; **72**: 296–302.
4. Sajan M, Reddy G, Kulkarni AP. *Int. J. Toxicol.* 2000; **19**: 285–292.
5. Ohkuma Y, Kawanishi S. *Biochem. Biophys. Res. Commun.* 1999; **257**: 555–560.
6. Dearden JC, Cronin MTD, Schultz TW, Lin DT. *Quant. Struct.–Act. Relat.* 1995; **14**: 427–432.
7. Schüürmann G, Fleming B, Dearden JC, Cronin MTD, Schultz TW. In *QSAR in Environmental Sciences*, Chen F, Schüürmann G (eds). SETAC Press: Pensacola, FL, 1997; 329–342.
8. Cronin MTD, Schultz TW. *Chem. Res. Toxicol.* 2001; **14**: 1284–1295.
9. Mekenyan O, Roberts DW, Karcher W. *Chem. Res. Toxicol.* 1997; **10**: 994–1000.
10. Schmidt RJ, Chung LY. *Arch. Dermatol. Res.* 1992; **284**: 400–408.
11. Schmitt H, Altenburger R, Jastorff B, Schüürmann G. *Chem. Res. Toxicol.* 2000; **13**: 441–450.
12. Agrawal VK, Khadikar PV. *Bioorg. Med. Chem.* 2001; **9**: 3035–3040.
13. Warne MA, Osborn D, Lindon JC, Nicholson JK. *Chemosphere* 1999; **38**: 3357–3382.
14. Bearden AP, Schultz TW. *SAR QSAR Environ. Res.* 1998; **9**: 127–153.
15. Schultz TW, Lin DT, Arnold LM. *Sci. Total Environ.* 1991; **109–110**: 569–580.
16. Schultz TW, Lin DT, Wesley SK. *Qual. Assur.* 1992; **1**: 132–143.
17. Estrada E, Uriarte E. *SAR QSAR Environ. Res.* 2001; **12**: 309–324.
18. Katritzky AR, Maran U, Lobanov VS, Karelson M. *J. Chem. Inf. Comput. Sci.* 2000; **40**: 1–18.
19. Katritzky AR, Petrukhin R, Tatham D, Basak S, Benfenati E, Karelson M, Maran U. *J. Chem. Inf. Comput. Sci.* 2001; **41**: 679–685.
20. Maran U, Karelson M, Katritzky AR. *Quant. Struct.–Act. Relat.* 1999; **18**: 3–10.
21. Katritzky AR, Tatham D, Maran U. *J Chem. Inf. Comput. Sci.* 2001; **41**: 1162–1176.
22. Stewart JJP. *MOPAC Program Package, Quantum Chemistry Program Exchange No. 455*, 1989.
23. Dewar M, Zoebisch EG, Healy EF, Stewart JJP. *J. Am. Chem. Soc.* 1985; **107**: 3902–3909.
24. www.codessa-pro.com.
25. Rauk A. *Orbital Interaction Theory of Organic Chemistry.* Academic Press: New York, 2001.
26. Karelson M, Lobanov VS, Katritzky AR. *Chem. Rev.* 1996; **96**: 1027–1043.
27. Stanton DT, Jurs PC. *Anal Chem.* 1990; **62**: 2323–2329.