Multivariate Discrimination between Modes of Toxic Action of Phenols

Aynur O. Aptula[a], Tatiana I. Netzeva[b], Iva V. Valkova[c], Mark T. D. Cronin[b], Terry W. Schultz[d], Ralph Kühne[a] and Gerrit Schüürmann[a]*

[a] Department of Chemical Ecotoxicology, UFZ Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany.
[b] School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, England.
[c] Department of Chemistry, Faculty of Pharmacy, Medical University-Sofia, 2 Dunav St., 1000 Sofia, Bulgaria.
[d] Department of Comparative Medicine, College of Veterinary Medicine, The University of Tennessee, 2407 River Drive, Knoxville, Tennessee, 37996-4500 USA.

Abstract

A set of 221 phenols, for which toxicity data to the ciliate *Tetrahymena pyriformis* were available, was subjected to stepwise linear discriminant analysis (LDA) in order to classify their toxic mechanisms of action. The compounds were *a priori* grouped into the following four mechanisms according to structural rules: polar narcotics, weak acid respiratory uncouplers, pro-electrophiles and soft electrophiles. Hydrophobicity with and without correction for ionisation (log K_{ow}, log $D_{ow}{}^{u}$), acidity constant (pK_a), frontier orbital energies (E_{LUMO}, E_{HOMO}) and hydrogen-bond donor and acceptor counts were used as molecular descriptors. LDA models employing 3 – 6 variables achieved 86 – 89% overall correct classification of the four mechanisms, with more varied performance for respiratory uncouplers and pro-electrophiles. For the latter, a separate model was developed that discriminated compounds undergoing metabolic activation from compounds with different mechanisms very accurately. Model validation was performed by evaluating the simulated external prediction through LDA models built from complementary subsets.

1 Introduction

Phenols are widely used both in industry and as consumer products. Widely used derivatives include 2,4-dichlorophenol and 2,4,5-trichlorophenol, which are precursors of the herbicides 2,4-dichlorophenoxy acetic acid and 2,4,5-trichlorophenoxy acetic acid, respectively, and chlorophenols, which are themselves used as bactericides, fungicides and herbicides [1, 2]. Cresols form an important group of disinfectants, and some naturally occurring phenols such as thymol (2-isopropyl-5-methylphenol) and carvacrol (5-isopropyl-2-methylphenol) are also known for their antiseptic action. Environmentally important nitrophenols include 3-trifluoromethyl-4-nitrophenol (TFM), which has been used as larval lampricide to control the sea lamprey (*Petromyzon marinus*) in the Great Lakes for more than 30 years [3].

The toxicity of phenols involves a number of different mechanisms and modes of action (MOA) [4, 5]. The ability to act as oxidative uncouplers is associated with pK_a values (negative logarithms of ionisation constant) in the range 3.8 to 8.5 [6]. QSAR investigations have suggested that this uncoupling of oxygen consumption from ATP synthesis is not restricted to mitochondria-containing species, but also observed in prokaryotic organisms such as bacteria [7, 8]. The energy of the lowest unoccupied molecular orbital, E_{LUMO}, has been demonstrated to discriminate various MOAs [5]. As such, E_{LUMO} quantifies the electron affinity of the molecule (cf. [9]), and may reflect both the tendency of phenols to attack electron-rich sites of endogenous macromolecules directly, and their ability to undergo metabolic activation following 1-electron reduction [10].

Existing methods for classifying compounds according to MOAs can be grouped into two types of approaches – a

* Corresponding author: e-mail: gs@uoe.ufz.de

Key words: phenols, linear discriminant analysis, mechanism of action, external validation

Abbreviations: CDF, canonical discrimination function; E_{HOMO}, energy of highest occupied molecular orbital; E_{LUMO}, energy of lowest unoccupied molecular orbital; LDA, linear discriminant analysis; log $D_{ow}{}^{u}$, decadic logarithm of the octanol/water partition coefficient corrected for ionisation; log K_{ow}, decadic logarithm of the octanol/water partition coefficient; MOA, mode of action; N_{Hdon}, number of hydrogen bond donor centres; N_{Hacc}, number of hydrogen bond acceptor centres; pK_a, negative decadic logarithm of the acidity constant; TFM, 3-trifluoromethyl-4-nitrophenol.

  0931-8771/02/0105-0012 $ 17.50+.50/0

qualitative approach, based on simple structural characteristics (such as the presence of a certain substituent), or a quantitative approach based on statistical analyses of physico-chemical properties [11]. The first approach is simple and relatively successful for phenols with only few substituents. However, it is restricted to the type of the substituents in the training set, and its application is limited when substituents associated with different MOAs are present in a molecule. Classification based on physico-chemical properties also has some disadvantages. These include the availability and use of the descriptors, the difficulty of mechanistic interpretation with some types of descriptors, and the fact that the property profile of the initial compound may differ significantly from the metabolically activated toxicant.

The aim of this study was to derive a descriptor-based classification of 221 phenols with respect to four pre-assigned mechanisms of action, using compounds with existing toxicity data with the ciliate *Tetrahymena pyriformis*, and various physico-chemical properties known to be associated with phenol toxicity.

2 Materials and Methods

2.1 Assignment of Toxic Mechanisms of Action

The mechanisms of toxic action of the phenols were assigned *a priori* following simple structural rules developed earlier for the growth inhibition assay with *Tetrahymena pyriformis* [5]. Phenols with more than one nitro group or more than three halogen groups are classified as respiratory uncouplers. Such compounds, in the case of eukaryotic cells, impair the pH and electrochemical gradient across the inner mitochondrial membrane [12]. Note that according to this structural classification scheme, picric acid (2,4,6-trinitrophenol) is also assigned an oxidative uncoupling mechanism, although its high acidity (pK_a below 0.5) is likely to make the anionic form highly prevalent in both the low-pH intermembrane area (cytosol) and the high-pH matrix of the mitochondria, thus preventing this compound from pumping protons into the matrix efficiently [6, 12].

Phenols with either a hydroxy group or an amino group in the 2- or 4-position may act as pro-electrophiles, i.e. they may be metabolised to more toxic forms [13]. Soft electrophiles have one nitro group, but generally no more than one halogen group. Their toxicity can be attributed to the alkylation of essential protein thiol or amino groups, or to oxidative stress produced by free radical formation [14, 15]. The group of pro-redox cyclers (not included in the analysis due to their small number) contains 2- or 4-amino or hydroquinones with no unsubstituted aromatic carbon atom [5], with distinct E_{LUMO} and the SOMO energy (singly occupied molecular orbital energy of the radical anion) properties [10]. By default, compounds not classified according to these structural rules were defined as polar narcotics. Narcosis is a reversible state of arrested activity of protoplasmic structures resulting from exposure to the appropriate xenobiotic [16]. The polar narcotic effect of the phenols can be distinguished from non-polar narcosis by various electrophysiological and biochemical variables [17].

A total of 221 phenols were taken from the TETRATOX database [18]. The phenols were classified into four MOA groups as listed in Table 1: 153 polar narcotics, 18 respiratory uncouplers, 27 pro-electrophiles, and 23 soft electrophiles. Note that with different organisms, a different classification scheme may be indicated (e.g. algae are sensitive towards inhibitors of photosynthesis), and that the endpoint under investigation may mask more specific types of responses. Moreover, there is certainly some overlap between the different MOAs, which is ignored in the present modeling approach for the sake of simplicity.

2.2 Calculation of Molecular Descriptors

A judiciously selected set of molecular parameters, that offered a clear mechanistic interpretation, was used for the representation of the relevant physico-chemical properties of the compounds.

The logarithm of the octanol/water partition coefficient, $\log K_{ow}$, was calculated using the MedChem (ver 3.53) software [19]. In addition, hydrophobicity, corrected for ionisation (considering only the undissociated compound fraction $f_u = 1/[1 + 10^{pH\text{-}pKa}]$) was calculated as $\log D_{ow}{}^{u} = f_u * \log K_{ow}$ (cf. [10]), using pH = 7.35 according to the conditions of the ciliate test. The negative logarithms of the ionisation constant (pK_a) were obtained from the MicroQSAR package [20].

The energies of the highest occupied and lowest unoccupied molecular orbital, E_{HOMO} and E_{LUMO} respectively, were calculated using the AM1 Hamiltonian [21] as implemented in MOPAC 6.0.

To count hydrogen bond donor and acceptor centres, N_{Hdon} and N_{Hacc} respectively, the following identification rules, implemented in TSAR (ver 3.3, Oxford Molecular Limited, Oxford, England), were used: donor centres: (O–H), (N–H), (S–H); acceptor centres: (–O–), (C=O), (–C#N), (–N=), tertiary nitrogen excluding N–C (sp^2), thio groups (–S–H), C=S.

2.3 Development of the Classification Model

Linear discriminant analysis (LDA) was used to classify the compounds according to the four *a priori* assigned mechanisms of toxic action on the basis of their molecular descriptors. Recent applications of this technique demonstrate its suitability for the modelling of categorical data [22–26]. For the purpose of LDA modelling, a value of 1 was assigned to the polar narcotics, 2 to weak acid respiratory uncouplers, 3 to pro-electrophiles, and 4 to the soft electrophiles (cf. Table 1). LDA was performed using STATISTICA software (STATISTICA '99 Edition package, version 5.5 A (Statsoft 1999). In this package, prior probabilities were computed from group size (0.692 for group 1, 0.081 for group 2, 0.122 for group 3, and 0.104 for group 4), and the quality of the

Table 1. Compounds with ciliate toxicity, mode-of-action assignment and calculated molecular descriptors.

N°	Name	Toxicity – log 1/IGC_{50} [mmol/L]	MOA class	*group**	log K_{ow}	pK_a	E_{LUMO}	E_{HOMO}	N_{Hdon}
Polar narcotics									
1	1,3,5-trihydroxybenzene	−1.26	1	1	0.16	8.45	0.25	−9.16	3
2	2-(*tert*)-butyl-4-methylphenol	1.30	1	1	3.80	11.39	0.46	−8.76	1
3	2,3,5-trichlorophenol	2.37	1	1	3.58	6.75	−0.56	−9.49	1
4	2,3,5-trimethylphenol	0.36	1	1	2.92	10.48	0.38	−8.81	1
5	2,3,6-trimethylphenol	0.28	1	1	2.92	10.63	0.36	−8.84	1
6	2,3-dichlorophenol	1.28	1	2	2.84	7.44	−0.25	−9.39	1
7	2,3-dimethylphenol	0.12	1	1	2.42	10.50	0.38	−8.93	1
8	2,4,5-trichlorophenol	2.10	1	2	3.58	7.37	−0.51	−9.32	1
9	2,4,6-tribromophenol	2.03	1	2	3.92	6.80	−0.62	−9.50	1
10	2,4,6-tribromoresorcinol	1.06	1	1	4.37	5.72	−0.66	−9.31	2
11	2,4,6-trichlorophenol	1.41	1	2	3.37	6.35	−0.50	−9.39	1
12	2,4,6-trimethylphenol	0.28	1	2	2.97	10.88	0.42	−8.70	1
13	2,4,6-tris (dimethylaminomethyl) phenol	−0.52	1	1	0.92	8.96	0.53	−8.65	1
14	2,4-dibromophenol	1.40	1	1	3.31	7.87	−0.30	−9.33	1
15	2,4-dichlorophenol	1.04	1	1	2.96	7.97	−0.19	−9.23	1
16	2,4-difluorophenol	0.60	1	2	1.95	8.58	−0.32	−9.29	1
17	2,4-dimethylphenol	0.07	1	2	2.47	10.58	0.44	−8.77	1
18	2,5-dichlorophenol	1.13	1	1	2.96	7.58	−0.29	−9.31	1
19	2,5-dimethylphenol	0.08	1	2	2.47	10.22	0.38	−8.85	1
20	2,6-di-(*tert*)-butyl-4-methylphenol	1.80	1	1	5.63	12.55	0.50	−8.62	1
21	2,6-dichloro-4-fluorophenol	0.80	1	1	2.80	6.75	−0.57	−9.38	1
22	2,6-dichlorophenol	0.74	1	2	2.63	6.78	−0.26	−9.37	1
23	2,6-difluorophenol	0.47	1	2	1.75	7.51	−0.32	−9.46	1
24	2,6-dimethoxyphenol	−0.60	1	2	1.10	9.92	0.39	−8.61	1
25	2-allylphenol	0.33	1	1	2.55	10.28	0.36	−9.04	1
26	2-bromo-4-methylphenol	0.60	1	1	2.85	8.67	0.02	−9.05	1
27	2-bromophenol	0.33	1	2	2.36	9.34	−0.01	−9.24	1
28	2-chloro-4,5-dimethylphenol	0.69	1	1	3.10	8.85	0.09	−8.89	1
29	2-chloro-5-methylphenol	0.39	1	1	2.65	8.54	0.06	−9.07	1
30	2-chlorophenol	0.18	1	2	2.16	8.55	0.07	−9.19	1
31	2-cyanophenol	0.03	1	1	1.60	6.98	−0.43	−9.58	1
32	2-ethoxyphenol	−0.36	1	1	1.85	10.11	0.42	−8.73	1
33	2-ethylphenol	0.16	1	1	2.50	10.20	0.40	−8.99	1
34	2-fluorophenol	0.19	1	1	1.72	8.73	0.02	−9.28	1
35	2-hydroxy-4,5-dimethylacetophenone	0.71	1	2	2.86	10.37	0.35	−8.76	1
36	2-hydroxy-4-methoxyacetophenone	0.55	1	1	1.98	9.67	−0.28	−9.33	1
37	2-hydroxy-4-methoxybenzophenone	1.42	1	1	3.58	9.67	−0.51	−9.03	1
38	2-hydroxy-5-methylacetophenone	0.31	1	1	2.41	10.23	−0.32	−9.19	1
39	2-hydroxyacetophenone	0.08	1	1	1.92	9.90	−0.33	−9.32	1
40	2-hydroxybenzylalcohol	−0.95	1	2	0.44	9.92	0.20	−9.24	2
41	2-hydroxyethylsalicylate	−0.08	1	2	1.56	9.92	−0.54	−9.42	2
42	2-isopropylphenol	0.80	1	2	2.90	10.40	0.41	−8.99	1
43	2-methoxy-4-propenylphenol	0.75	1	1	2.58	9.88	−0.05	−8.50	1
44	2-methoxyphenol	−0.51	1	1	1.32	9.99	0.40	−8.79	1
45	2-phenylphenol	1.09	1	1	3.09	9.55	−0.04	−8.74	1
46	2-(*tert*)-butylphenol	1.30	1	2	3.30	11.10	0.42	−8.97	1
47	3,4,5-trimethylphenol	0.93	1	1	2.87	10.25	0.41	−8.76	1
48	3,4-dichlorophenol	1.75	1	1	3.17	8.63	−0.20	−9.26	1
49	3,4-dimethylphenol	0.12	1	2	2.42	10.32	0.43	−8.80	1
50	3,5-dibromosalicylaldehyde	1.64	1	2	3.42	6.20	−0.93	−9.66	1
51	3,5-dichlorophenol	1.57	1	1	3.29	8.18	−0.28	−9.54	1
52	3,5-dichlorosalicylaldehyde	1.55	1	2	3.07	6.20	−0.90	−9.58	1
53	3,5-diiododsalicylaldehyde	2.34	1	2	3.87	6.37	−0.90	−9.72	1
54	3,5-dimethoxyphenol	−0.09	1	1	1.60	9.35	0.42	−8.94	1
55	3,5-dimethylphenol	0.11	1	1	2.47	10.15	0.39	−8.98	1
56	3,5-di-(*tert*)-butylphenol	1.64	1	2	5.13	10.32	0.47	−8.93	1
57	3-acetamidophenol	−0.16	1	1	0.49	9.92	0.22	−8.73	2
58	3-bromophenol	1.15	1	1	2.64	9.03	−0.05	−9.34	1
59	3-chloro-4-fluorophenol	1.13	1	2	2.72	8.96	−0.26	−9.25	1

Table 1. (cont.)

N°	Name	Toxicity – log 1/IGC_{50} [mmol/L]	MOA class	*group**	log K_{ow}	pK_a	E_{LUMO}	E_{HOMO}	N_{Hdon}
60	3-chloro-5-methoxyphenol	0.76	1	2	2.50	8.85	0.03	−9.22	1
61	3-chlorophenol	0.87	1	1	2.49	9.10	0.04	−9.30	1
62	3-cyanophenol	−0.06	1	1	1.60	8.61	−0.50	−9.59	1
63	3-ethoxy-4-hydroxybenzaldehyde	0.02	1	1	1.80	7.80	−0.44	−9.09	1
64	3-ethoxy-4-methoxyphenol	−0.30	1	2	1.69	9.94	0.32	−8.36	1
65	3-ethylphenol	0.23	1	2	2.50	10.07	0.40	−9.04	1
66	3-fluorophenol	0.38	1	2	1.92	9.29	0.03	−9.37	1
67	3-hydroxy-4-methoxybenzylalcohol	−0.99	1	2	0.29	9.67	0.33	−8.68	2
68	3-hydroxyacetophenone	−0.38	1	2	1.46	9.19	−0.46	−9.40	1
69	3-hydroxybenzaldehyde	0.09	1	2	1.44	9.00	−0.53	−9.43	1
70	3-hydroxybenzoic acid	−0.81	1	2	1.56	4.08	−0.57	−9.52	2
71	3-hydroxybenzylalcohol	−1.04	1	2	0.44	9.83	0.12	−9.26	2
72	3-iodophenol	1.12	1	2	2.90	8.88	−0.04	−9.34	1
73	3-isopropylphenol	0.61	1	1	2.90	10.10	0.42	−9.02	1
74	3-methoxyphenol	−0.33	1	2	1.57	9.65	0.36	−8.88	1
75	3-phenylphenol	1.35	1	1	3.23	9.63	−0.15	−8.95	1
76	3-(*tert*)-butylphenol	0.73	1	1	3.30	10.10	0.43	−9.01	1
77	4-(*tert*)-octylphenol	2.10	1	1	5.16	9.92	0.47	−8.85	1
78	4-(*tert*)-butylphenol	0.91	1	2	3.30	10.30	0.47	−8.91	1
79	4,6-dichlororesorcinol	0.97	1	2	2.08	7.28	−0.25	−9.03	2
80	4-allyl-2-methoxyphenol	0.42	1	2	2.40	10.23	0.25	−8.89	1
81	4-benzyloxyphenol	1.04	1	2	3.34	10.60	0.24	−8.61	1
82	4-bromo-2,6-dichlorophenol	1.78	1	2	3.52	6.40	−0.52	−9.44	1
83	4-bromo-2,6-dimethylphenol	1.17	1	1	3.63	10.00	0.10	−8.99	1
84	4-bromo-3,5-dimethylphenol	1.27	1	2	3.63	9.70	0.11	−9.06	1
85	4-bromo-6-chloro-2-cresol	1.28	1	1	3.61	8.20	−0.22	−9.23	1
86	4-bromophenol	0.68	1	2	2.64	9.34	0.02	−9.20	1
87	4-butoxyphenol	0.70	1	1	3.16	10.60	0.34	−8.61	1
88	4-chloro-2-isopropyl-5-methylphenol	1.85	1	1	4.41	10.03	0.14	−8.93	1
89	4-chloro-2-methylphenol	0.70	1	2	2.98	9.67	0.12	−9.01	1
90	4-chloro-3,5-dimethylphenol	1.20	1	2	3.48	9.65	0.14	−8.99	1
91	4-chloro-3-ethylphenol	1.08	1	2	3.51	9.54	0.14	−9.04	1
92	4-chloro-3-methylphenol	0.80	1	1	2.98	9.55	0.13	−9.04	1
93	4-chlorophenol	0.55	1	2	2.49	9.43	0.10	−9.12	1
94	4-chlororesorcinol	0.13	1	2	1.58	8.11	0.01	−8.98	2
95	4-cyanophenol	0.52	1	1	1.60	7.95	−0.40	−9.57	1
96	4-ethoxyphenol	0.01	1	2	2.10	10.50	0.33	−8.61	1
97	4-ethylphenol	0.21	1	1	2.50	10.00	0.43	−8.92	1
98	4-fluorophenol	0.02	1	2	1.92	9.89	0.07	−9.09	1
99	4-heptyloxyphenol	2.03	1	1	4.75	10.70	0.35	−8.59	1
100	4-hexyloxyphenol	1.64	1	1	4.22	10.70	0.35	−8.60	1
101	4-hexylresorcinol	1.80	1	2	3.45	9.63	0.30	−8.80	2
102	4-hydroxy-2-methylacetophenone	0.19	1	2	1.95	10.05	−0.35	−9.33	1
103	4-hydroxy-3-methoxyacetophenone	−0.12	1	2	1.27	9.85	−0.39	−9.00	1
104	4-hydroxy-3-methoxybenzonitrile	−0.03	1	2	1.42	7.96	−0.44	−9.10	1
105	4-hydroxy-3-methoxybenzylalcohol	−0.70	1	2	0.29	9.90	0.17	−9.03	2
106	4-hydroxy-3-methoxybenzylamine	−0.97	1	1	0.28	9.27	0.19	−8.87	2
107	4-hydroxy-3-methoxyphenethylalcohol	−0.18	1	2	0.47	9.92	0.28	−8.78	2
108	4-hydroxyacetophenone	−0.30	1	1	1.46	8.05	−0.38	−9.43	1
109	4-hydroxybenzaldehyde	0.27	1	2	1.44	7.62	−0.44	−9.49	1
110	4-hydroxybenzamide	−0.78	1	1	0.33	9.23	−0.25	−9.44	2
111	4-hydroxybenzoic acid	−1.02	1	1	1.56	4.58	−0.49	−9.60	2
112	4-hydroxybenzophenone	1.02	1	2	3.07	8.89	−0.49	−9.40	1
113	4-hydroxybenzylcyanide	−0.38	1	1	0.90	9.52	0.06	−9.30	1
114	4-hydroxyphenethylalcohol	−0.83	1	1	0.52	9.92	0.29	−9.06	2
115	4-hydroxyphenylacetic acid	−1.50	1	2	0.75	4.49	−0.21	−9.48	2
116	4-hydroxypropiophenone	0.05	1	1	1.98	8.85	−0.36	−9.41	1
117	4-iodophenol	0.85	1	2	2.90	9.20	0.02	−9.24	1
118	4-isopropylphenol	0.47	1	1	2.90	10.30	0.44	−8.92	1
119	4-methoxyphenol	−0.14	1	2	1.57	10.20	0.31	−8.65	1

Table 1. (cont.)

N°	Name	Toxicity – log 1/IGC_{50} [mmol/L]	MOA class	*group**	log K_{ow}	pK_a	E_{LUMO}	E_{HOMO}	N_{Hdon}
120	4-phenylphenol	1.39	1	2	3.20	9.55	−0.10	−8.68	1
121	4-propylphenol	0.64	1	1	3.03	10.30	0.43	−8.91	1
122	4-(*sec*)-butylphenol	0.98	1	1	3.43	10.30	0.45	−8.91	1
123	4-(*tert*)-pentylphenol	1.23	1	1	3.83	10.30	0.46	−8.90	1
124	5-bromo-2-hydroxybenzylalcohol	0.34	1	2	1.60	9.34	−0.13	−9.29	2
125	5-bromovanillin	0.62	1	2	1.92	6.06	−0.70	−9.33	1
126	5-fluoro-2-hydroxyacetophenone	0.04	1	2	2.17	9.79	−0.64	−9.39	1
127	5-methylresorcinol	−0.39	1	2	1.31	9.46	0.26	−8.95	2
128	5-pentylresorcinol	1.31	1	2	3.42	9.49	0.29	−8.93	2
129	6-(*tert*)-butyl-2,4-dimethylphenol	1.16	1	2	4.30	12.50	0.44	−8.69	1
130	α,α,α-trifluoro-4-cresol	0.62	1	1	2.88	8.68	−0.34	−9.80	1
131	ethyl-3-hydroxybenzoate	0.48	1	2	2.51	9.09	−0.45	−9.44	1
132	ethyl-4-hydroxy-3-methoxyphenylacetate	−0.23	1	1	1.53	9.92	0.04	−8.86	1
133	ethyl-4-hydroxybenzoate	0.57	1	2	2.51	8.92	−0.38	−9.50	1
134	Isovanillin	−0.14	1	1	1.28	8.89	−0.49	−9.11	1
135	3-cresol	−0.06	1	2	1.97	10.10	0.39	−9.02	1
136	Methyl-3-hydroxybenzoate	−0.05	1	1	1.99	9.21	−0.48	−9.45	1
137	Methyl-4-hydroxybenzoate	0.08	1	1	1.99	9.05	−0.40	−9.54	1
138	Methyl-4-methoxysalicylate	0.62	1	2	2.49	9.21	−0.31	−9.35	1
139	Nonylphenol	2.47	1	2	6.20	10.40	0.43	−8.92	1
140	2-cresol	−0.30	1	1	1.97	10.26	0.40	−8.96	1
141	2-vanillin	0.38	1	2	1.65	7.91	−0.45	−9.12	1
142	4-cresol	−0.18	1	1	1.97	10.26	0.43	−8.88	1
143	4-cyclopentylphenol	1.29	1	1	3.54	9.92	0.43	−8.90	1
144	Phenol	−0.21	1	2	1.48	9.99	0.40	−9.11	1
145	Resorscinol	−0.65	1	1	0.80	9.44	0.28	−8.98	2
146	Salicylaldehyde	0.42	1	1	1.81	8.34	−0.43	−9.50	1
147	Salicylaldoxime	0.25	1	1	1.10	9.92	−0.19	−8.99	2
148	Salicylamide	−0.24	1	2	1.28	8.36	−0.19	−9.51	2
149	Salicylhydrazide	0.18	1	1	0.85	9.92	−0.36	−9.57	3
150	Salicylhydroxamic acid	0.38	1	1	0.88	8.78	−0.46	−9.66	3
151	Salicylic acid	−0.51	1	2	2.19	2.98	−0.46	−9.51	2
152	Syringaldehyde	0.17	1	2	0.99	7.62	−0.50	−8.94	1
153	Vanillin	−0.03	1	1	1.28	7.40	−0.48	−9.14	1
Weak acid respiratory uncouplers									
154	2,3,4,5-tetrachlorophenol	2.71	2	2	4.06	6.22	−0.72	−9.45	1
155	2,3,5,6-tetrachlorophenol	2.22	2	2	3.85	5.24	−0.82	−9.63	1
156	2,3,5,6-tetrafluorophenol	1.17	2	1	2.07	6.00	−1.00	−9.88	1
157	2,3-dinitrophenol	0.46	2	1	1.98	5.15	−1.94	−10.66	1
158	2,4,6-trinitrophenol	−0.16	2	1	1.59	0.15	−2.53	−11.42	1
159	2,4-dichloro-6-nitrophenol	1.75	2	1	3.07	4.75	−1.58	−9.88	1
160	2,4-dinitrophenol	1.08	2	1	1.79	4.08	−1.89	−10.76	1
161	2,5-dinitrophenol	0.95	2	2	1.79	5.22	−2.27	−10.62	1
162	2,6-dichloro-4-nitrophenol	0.63	2	1	2.74	4.12	−1.44	−10.18	1
163	2,6-diiodo-4-nitrophenol	1.71	2	1	3.72	5.37	−1.43	−10.24	1
164	2,6-dinitro-4-cresol	1.23	2	1	2.29	3.99	−1.90	−10.35	1
165	2,6-dinitrophenol	0.54	2	2	1.79	3.71	−1.95	−10.66	1
166	3,4,5,6-tetrabromo-2-cresol	2.57	2	2	4.97	6.42	−0.88	−9.49	1
167	3,4-dinitrophenol	0.27	2	2	1.98	5.42	−1.87	−10.74	1
168	4,6-dinitro-2-cresol	1.72	2	2	2.29	4.32	−1.82	−10.50	1
169	Pentabromophenol	2.66	2	1	4.85	4.57	−1.19	−9.69	1
170	Pentachlorophenol	2.05	2	2	4.32	4.70	−0.98	−9.58	1
171	Pentafluorophenol	1.64	2	2	2.21	5.86	−1.30	−9.94	1
Precursors to soft electrophiles									
172	1,2,3-trihydroxybenzene	0.85	3	2	0.21	9.03	0.03	−9.16	2
173	1,2,4-trihydroxybenzene	0.44	3	1	0.21	9.54	0.16	−8.64	3
174	2,3-dimethylhydroquinone	1.41	3	1	1.24	9.98	0.22	−8.58	2
175	2,4-diaminophenol	0.13	3	1	−0.61	4.48	0.54	−8.00	3
176	2-amino-4-(*tert*)-butylphenol	0.37	3	1	2.44	5.10	0.51	−8.31	2
177	2-aminophenol	0.94	3	2	0.62	9.28	0.47	−8.35	2

Table 1. (cont.)

N°	Name	Toxicity – log 1/IGC_{50} [mmol/L]	MOA class	*group**	log K_{ow}	pK_a	E_{LUMO}	E_{HOMO}	N_{Hdon}
178	3,5-di-(*tert*)-butylcatechol	2.11	3	1	4.53	10.23	0.33	−8.63	2
179	3-aminophenol	−0.52	3	1	0.25	9.83	0.55	−8.56	2
180	3-methylcatechol	0.28	3	2	1.38	9.96	0.28	−8.84	2
181	4-acetamidophenol	−0.82	3	2	0.49	10.12	0.28	−8.46	2
182	4-amino-2,3-dimethylphenol	1.44	3	2	1.15	5.28	0.32	−8.84	2
183	4-amino-2-cresol	1.31	3	1	0.75	5.65	0.44	−8.21	2
184	4-aminophenol	−0.08	3	2	0.25	8.50	0.43	−8.28	2
185	4-chlorocatechol	1.06	3	1	1.98	9.01	0.01	−8.97	2
186	4-methylcatechol	0.37	3	2	1.37	9.96	0.34	−8.72	2
187	5-amino-2-methoxyphenol	0.45	3	2	0.15	5.01	0.48	−8.22	2
188	5-chloro-2-hydroxyaniline	0.78	3	1	1.71	3.77	0.16	−8.55	2
189	6-amino-2,4-dimethylphenol	0.89	3	1	1.62	5.18	0.41	−8.57	2
190	Bromohydroquinone	1.68	3	1	1.78	8.29	−0.20	−8.93	2
191	Catechol	0.75	3	2	0.88	9.36	0.30	−8.88	2
192	Chlorohydroquinone	1.26	3	2	1.40	8.34	−0.12	−8.90	2
193	Hydroquinone	0.47	3	1	0.59	9.91	0.23	−8.72	2
194	Methoxyhydroquinone	2.20	3	2	0.47	9.61	0.25	−8.55	2
195	Methylhydroquinone	1.86	3	1	0.98	9.99	0.24	−8.63	2
196	Phenylhydroquinone	2.01	3	2	2.43	9.74	−0.22	−8.65	2
197	Tetrachlorocatechol	1.70	3	2	4.29	5.83	−0.79	−9.36	2
198	Trimethylhydroquinone	1.34	3	2	1.69	10.41	0.24	−8.49	2
Soft electrophiles									
199	2,6-dibromo-4-nitrophenol	1.36	4	2	3.14	4.03	−1.45	−10.22	1
200	2-amino-4-chloro-5-nitrophenol	1.17	4	1	1.80	0.14	−0.99	−9.16	2
201	2-amino-4-nitrophenol	0.48	4	2	1.18	2.71	−0.98	−9.08	2
202	2-chloro-4-nitrophenol	1.59	4	2	2.33	5.64	−1.26	−10.03	1
203	2-chloromethyl-4-nitrophenol	0.75	4	2	2.42	7.15	−1.19	−10.14	1
204	2-nitrophenol	0.67	4	2	1.85	7.22	−1.19	−9.91	1
205	2-nitroresorcinol	0.66	4	1	1.56	6.21	−1.32	−9.59	2
206	3-fluoro-4-nitrophenol	0.94	4	2	1.79	6.40	−1.28	−10.24	1
207	3-hydroxy-4-nitrobenzaldehyde	0.27	4	2	1.47	6.00	−1.66	−10.21	1
208	3-methyl-4-nitrophenol	1.73	4	1	2.27	7.29	−1.00	−9.96	1
209	3-nitrophenol	0.51	4	1	1.85	8.36	−1.15	−9.94	1
210	4-amino-2-nitrophenol	0.88	4	1	0.81	3.34	−1.12	−8.96	2
211	4-chloro-2-nitrophenol	2.05	4	2	2.66	6.48	−1.38	−9.84	1
212	4-chloro-6-nitro-3-cresol	1.64	4	1	3.16	6.40	−1.36	−9.79	1
213	4-hydroxy-3-nitrobenzaldehyde	0.61	4	1	1.47	4.50	−1.46	−10.24	1
214	4-methyl-2-nitrophenol	0.57	4	2	2.35	7.11	−1.14	−9.65	1
215	4-methyl-3-nitrophenol	0.74	4	1	2.27	8.58	−1.09	−9.65	1
216	4-nitro-3-(trifluoromethyl)-phenol	1.65	4	2	2.77	6.13	−1.59	−10.50	1
217	4-nitrocatechol	1.17	4	2	1.66	6.70	−1.16	−9.76	2
218	4-nitrophenol	1.42	4	1	1.85	7.65	−1.07	−10.07	1
219	4-nitrosophenol	0.65	4	2	1.36	6.48	−0.80	−9.57	1
220	5-fluoro-2-nitrophenol	1.13	4	1	2.09	6.04	−1.45	−10.21	1
221	5-hydroxy-2-nitrobenzaldehyde	0.33	4	1	1.75	6.35	−1.51	−10.28	1

* Two complementary subsets, *group 1* and *group 2*, are defined as described in the section *Materials and Methods.*

discriminant functions was characterised by Wilks' λ, Fisher's *F* test value and the p level of statistical significance. Note that Wilks' λ is a test parameter according to the U statistics. Its value ranges between 0 and 1, with values close to 0 indicating significant differences between the group means, and values close to 1 indicating the group means are only marginally, or not at all different. The LDA results are reported as percentage of correctly classified cases.

Validation

In order to assess the predictivity of the LDA models, the following strategy was selected to avoid the pitfalls occasionally observed with the conventional leave-one-out method [27]. Within each pre-assigned mechanism, the compounds were re-numbered according to toxicity, and then subdivided into two equalised complementary subsets by taking alternate chemical. Through this approach, both

subsets cover essentially the same toxicity range as well as the same number of compounds from each MOA class, and can be used for subset-specific LDA model validation as well as external prediction (*group1* compound MOAs from *group2* model and vice versa). Previous experience shows that this kind of simulated external validation, through the use of complementary subsets, is significantly more realistic than the leave-one-out procedure in assessing the true prediction power of regression models [27].

3 Results and Discussion

Hydrophobicity (log K_{ow} or log $D_{ow}{}^{u}$), compound acidity (pK_a) and electron affinity (E_{LUMO}) are known to be used to discriminate between polar narcosis, oxidative uncoupling and (direct or metabolically induced) electrophilic mechanisms of toxic action [5–8]. Further, pro-electrophiles require metabolic activation, which in case of oxidative pathways could be modelled by E_{HOMO}. This parameter quantifies the ionisation potential and thus also characterises the ability of the molecule to donate electrons to reaction partners [9]. Since polar narcotics differ from nonpolar narcotics by their greater ability for hydrogen-bond interactions [28], corresponding measures of the hydrogen-bond donating and hydrogen-bond accepting capability in terms of N_{Hdon} and N_{Hacc} appear to be meaningful for discrimination of mechanisms. In addition, hydrogen bonding may also play a role in fixating the toxicant in the course of bioreactive interactions with endogenous macromolecules.

Table 2. Discrimination between modes of action of the 221 phenols based on individual molecular descriptors.

Descriptors	Wilks' λ	*F* value	p level	% Correct classification
E_{LUMO}	0.312	159.6	0	80.5
E_{HOMO}	0.391	112.5	0	79.2
pK_a	0.558	57.3	0	71.5
N_{Hdon}	0.647	39.5	0	69.2
N_{Hacc}	0.734	26.3	1.54E-14	69.2
$\log D_{ow}{}^{u}$	0.756	23.3	4.07E-13	68.8
$\log K_{ow}$	0.871	10.7	1.36E-6	69.7

Table 3. Squared intercorrelation coefficients between the molecular descriptors for the total set of 221 phenols.

	pK_a	$\log K_{ow}$	$\log D_{ow}{}^{u}$	E_{LUMO}	E_{HOMO}	N_{Hdon}	N_{Hacc}
pK_a	1						
$\log K_{ow}$	0.01	1					
$\log D_{ow}{}^{u}$	0.48	0.50	1				
E_{LUMO}	0.55	0	0.15	1			
E_{HOMO}	0.35	0	0.07	0.81	1		
N_{Hdon}	(−)0.01	(−)0.28	(−)0.21	0.04	0.09	1	
N_{Hacc}	(−)0.19	(−)0.14	(−)0.31	(−)0.42	(−)0.31	0.01	1

The results of LDA utilising one parameter are summarised in Table 2. Of the seven descriptors tested, E_{LUMO} is the best single discriminator, yielding 81% correct classification of the 221 phenols with good statistics (Wilks' $\lambda = 0.312$, $F_{3,217} = 159.6$). It follows that the present LDA confirms, on a statistical basis, previous findings about associations between E_{LUMO} and certain mechanisms [5].

Interestingly, E_{HOMO} provides an almost equivalent discrimination (79% correct classification), which can be explained by its high intercorrelation with E_{LUMO} for the phenol set under investigation ($r^2 = 0.81$, see Table 3). All other descriptors provide similar classification rates of approximately 70%, with $\log K_{ow}$ showing the greatest value for Wilks' λ (0.871), thus providing the poorest model. As can be seen from Table 3, intercorrelations between the molecular parameters are well below 50% except for E_{LUMO} vs. E_{HOMO} (s.a.), pK_a vs. E_{LUMO} ($r^2 = 0.55$) and pK_a vs. $\log D_{ow}{}^{u}$ ($r^2 = 0.48$).

Stepwise linear discriminant analysis revealed that in multivariable models, the use of $\log K_{ow}$ is, in most cases, significantly superior to $\log D_{ow}{}^{u}$. This finding is somewhat surprising and contrasts with recent QSAR investigations, which demonstrate that for the phenols, the use of $\log K_{ow}$ may in fact mask toxicity contributions from specific modes of action through a substantial overestimation of the hydrophobicity-driven toxicity component [10].

The best LDA models employing three to six variables are listed in Table 4. The three-variable model using $\log K_{ow}$, pK_a and E_{LUMO} correctly classified 86.4% of the compounds. However, the modelling of individual MOA classes showed significant differences in correct predictions. The largest group of compounds, the polar narcotics (153 compounds) achieved the best degree of classification (96.1%), whilst only 37% of the pro-electrophiles were correctly classified. Both the 18 respiratory uncouplers and the 23 soft electrophiles are identified reasonably well (83%). The three canonical discriminant functions (CDFs), which form the basis of the linear discriminant analysis were:

$$CDF1 = -0.197 \log K_{ow} + 0.036\ pK_a + 2.619\ E_{LUMO} + 0.776$$

$$CDF2 = 0.617 \log K_{ow} + 0.493\ pK_a - 1.045\ E_{LUMO} - 5.745$$

$$CDF3 = 0.670 \log K_{ow} - 0.532\ pK_a + 1.468\ E_{LUMO} + 3.262$$

For CDF1, CDF2 and CDF3, Wilks' λ is 0.241, 0.796 and 0.970, respectively, which indicates that the third function contributes only a very small amount to the overall classification. The respective χ^2 test values are 308.2, 49.3 and 6.6, showing that CDF3 is significant at the 99% level ($p = 0.01$).

Inclusion of N_{Hdon} as next best descriptor results in a significantly better treatment of the pro-electrophiles (70.4%), which might indicate that hydrogen bonding is involved in the metabolic toxification of these compounds. N_{Hacc} is clearly less suited as a further descriptor. Interestingly, the identification of polar narcosis was now poorer

Table 4. Linear discriminant analysis results for the total compound set and for the two complementary subsets.

Variables	Calibration [% correct classification]			External prediction [% correct classification]	
MOA class	all	*group1*	*group2*	*group1*	*group2*
3 ($\log K_{ow}$, pK_a, E_{LUMO})	86.4	89.0	81.3	84.2	85.5
Polar narcotics	96.1	96.1	94.8	96.1	94.8
Weak acid respiratory uncouplers	83.3	88.9	44.4	55.6	77.8
Pro-electrophiles	37.0	46.2	28.6	37.0	37.0
Soft electrophiles	82.6	90.9	83.3	82.6	87.0
4 ($\log K_{ow}$, pK_a, E_{LUMO}, N_{Hdon})	84.2	89.9	79.5	83.3	85.9
Polar narcotics	86.9	96.1	81.8	84.3	94.8
Weak acid respiratory uncouplers	83.3	88.9	44.4	61.1	77.8
Pro-electrophiles	70.4	53.9	85.7	88.9	40.7
Soft electrophiles	82.6	90.9	83.3	87.0	86.9
5 ($\log K_{ow}$, pK_a, E_{LUMO}, E_{HOMO}, N_{Hdon})	89.1	93.6	85.7	88.7	88.7
Polar narcotics	93.5	98.7	93.5	92.8	96.1
Weak acid respiratory uncouplers	77.8	88.9	44.4	66.7	72.2
Pro-electrophiles	77.8	69.2	85.7	92.6	59.3
Soft electrophiles	82.6	90.9	66.7	73.9	86.9
5 ($\log D_{ow}^{u}$, pK_a, E_{LUMO}, E_{HOMO}, N_{Hdon})	87.8	89.0	84.8	86.9	86.4
Polar narcotics	94.8	98.7	93.5	92.2	96.7
Weak acid respiratory uncouplers	55.6	66.7	33.3	55.6	55.6
Pro-electrophiles	81.5	61.5	85.7	92.6	59.3
Soft electrophiles	73.9	72.7	66.7	69.6	73.9
6 ($\log K_{ow}$, pK_a, E_{LUMO}, E_{HOMO}, N_{Hdon}, N_{Hacc})	89.1	92.7	87.5	88.7	88.7
Polar narcotics	94.1	97.4	94.8	94.1	96.1
Weak acid respiratory uncouplers	66.7	77.8	44.4	55.6	61.1
Pro-electrophiles	81.5	76.9	92.9	92.6	66.7
Soft electrophiles	82.6	90.9	66.7	73.9	87.0
2 ($\log K_{ow}$, E_{LUMO})	94.1	95.4	92.0	94.1	93.7
Weak acid respiratory uncouplers	66.7	88.9	44.4	66.7	66.7
All other compounds	96.6	96.0	96.1	96.6	96.1
2 ($\log K_{ow}$, pK_a)	93.7	95.4	91.1	93.2	93.7
Weak acid respiratory uncouplers	66.7	88.9	22.2	44.4	66.7
All other compounds	96.1	96.0	97.1	97.5	96.1
2 (E_{HOMO}, N_{Hdon})	94.6	94.5	91.1	91.0	94.1
Pro-electrophiles	88.9	84.6	92.9	96.3	70.4
All other compounds	95.4	95.8	90.8	90.2	97.4

Statistical rest results of the 3- to 6-variable LDA models in the order as listed above:
Wilks' λ: 0.241, 0.195, 0.170, 0.174, 0.166; *F* value: 46.2, 40.4, 35.2, 34.7, 29.6; p level: < 0.0001 for the first five models. For the three 2-variable models, the statistical parameters are given in the text.
For the definition of the complementary subsets *group1* and *group2*, see Table 1 and the section *Materials and Methods.*

(86.9%), possibly reflecting the above-mentioned fact that polar narcotics possess some capability for hydrogen bonding. Due to the large proportion of polar narcotics (69% of the total compound set), the overall classification of MOA decreased slightly to 84%.

The addition of E_{HOMO} as a fifth molecular parameter increased the correct classification to 89%. As outlined above, E_{HOMO} is a measure of the susceptibility of the compounds to oxidation. Inclusion of this parameter lead to a better classification of pro-electrophiles that might be subject to oxidative biotransformation (77.8%). There is, however, a corresponding decrease in the correct classification of the respiratory uncouplers, and also a similar increase in the classification of the polar narcotics. Replacement of log K_{ow} by log D_{ow}^{u} resulted in a much poorer performance for the group of 18 uncouplers, which is surprising in view of the fact that from mechanistic reasoning, the correction of hydrophobicity for ionisation would make sense. The 5-variable canonical discrimination functions employing log K_{ow} were:

$$CDF1 = 0.126 \log K_{ow} - 0.065\ pK_a - 2.125\ E_{LUMO} - 0.467\ E_{HOMO} - 0.568\ N_{Hdon} - 3.837$$

$$CDF2 = -0.105 \log K_{ow} - 0.134\ pK_a - 2.171\ E_{LUMO} + 2.592\ E_{HOMO} + 1.566\ N_{Hdon} + 22.71$$

$$CDF3 = -0.671 \log K_{ow} + 0.513\ pK_a - 2.638\ E_{LUMO} + 1.891\ E_{HOMO} - 0.788\ N_{Hdon} + 15.07$$

Here, the individual values of Wilks' λ for CDF1, CDF2 and CDF3 were 0.170, 0.581, 0.961, and the χ^2 test 381.3, 117.0, 8.5, respectively. Comparable to the 3-variable model, CDF1 and CDF2 have a high statistical significance ($p <$

0.0001), whilst CDF3 has a significance level of 96% ($p = 0.04$). It follows that according to both Wilks' λ and the χ^2 tests, CDF3 provides only a minor contribution to the MOA discrimination, which is similar for both the 3-variable and the 5-variable functions.

Inclusion of log K_{ow}, pK_a, E_{LUMO}, E_{HOMO}, N_{Hdon} and N_{Hacc} resulted in a model that identified both pro-electrophiles and soft electrophiles reasonably well (81.5% and 82.6% correct classification, respectively). However, it demonstrates a relatively poor performance for the group of oxidative uncouplers (66.7% correctly classified). The latter is again surprising; note that the simple 3-variable model provides a much better classification capability for this mode of action. Indeed, comparative analysis of the five LDA models reveals that the correct classification of oxidative uncouplers decreases with increasing complexity (i.e. the number of variables) of the model.

Since both the pro-electrophiles and the uncouplers showed the greatest variation in correct classification from the five LDA models, separate discriminant analyses were performed to identify these MOAs individually (i.e. assigning a 1 to compounds with the MOA of interest, and a 0 to the combined set of compounds with all other MOAs).

In the case of separating the class of 27 pro-electrophiles from the combined set of all other compounds, E_{HOMO} and N_{Hdon} lead to a correct classification of 88.9% of these pro-electrophiles, with 95.4% correct classification of the other compounds as having a different MOA, and an overall accuracy of 94.6% (cf. lower part of Table 4). The resultant canonical discriminant function,

$$\text{CDF } (E_{HOMO}, N_{Hdon}) = -0.993\ E_{HOMO;} - 2.048\ N_{Hdon} - 6.513$$

$$\text{Wilks' } \lambda = 0.597,\ F_{2,218} = 73.7,\ \chi^2 = 112.6,\ p < 0.0001$$

is highly statistically significant. The contribution of E_{HOMO} again supports the interpretation that for the phenolic pro-electrophiles under investigation, metabolic toxification is likely to be associated with an oxidative biotransformation route. In order to interpret N_{Hdon}, a further possibility could also be the fact that for the present phenol set, this descriptor simply reflects the structural rule for the *a priori* assignment of pro-electrophiles, since it essentially counts the number of hydroxy and amino groups (by definition [5], precursors to soft electrophiles are compounds with two hydroxy groups, or one hydroxy and one amino group). Comparison of this LDA model with the 5-variable model to classify all four MOAs shows that upon inclusion of more parameters for reactivity, part of the discriminatory power of E_{HOMO} and N_{Hdon} for the pro-electrophiles was lost.

For the group of 18 respiratory uncouplers, a corresponding separate discriminant analysis was only moderately successful with two descriptors. A model based on E_{LUMO} yielded only 61.1% correct classification of the uncouplers (and 96.1% correct classification of the other compounds as exerting a different MOA). With either log K_{ow} and E_{LUMO} or log K_{ow} and pK_a, 66.7% of the uncouplers were correctly identified (and 96.6% and 96.1%, respectively, of the other compounds are correctly classified as exerting a different MOA; cf. lower part of Table 4). No other 2-variable combination provided a better discrimination result. The canonical discrimination functions were as follows:

$$\text{CDF } (\log K_{ow}, E_{LUMO}) = 0.250 \log K_{ow} - 1.823\ E_{LUMO} - 1.006$$

$$\text{Wilks' } \lambda = 0.643,\ F_{2,218} = 60.5,\ \chi^2 = 96.2,\ p < 0.0001$$

$$\text{CDF } (\log K_{ow}, pK_a) = 0.383 \log K_{ow} - 0.523\ pK_a + 3.485$$

$$\text{Wilks' } \lambda = 0.718,\ F_{2,218} = 42.9,\ \chi^2 = 72.4,\ p < 0.0001$$

It is unclear why the inclusion of pK_a does not result in a better degree of classification of the oxidative uncouplers (cf. [6]). Note further that even a 3-variable LDA model based on log K_{ow}, E_{LUMO} and pK_a provided only a 72.2% correct classification of the uncouplers. This is less predictive than the corresponding results of the more complex 4-MOA LDA models based on three to five variables with the inclusion of these three descriptors (see Table 4). When comparing the classification results for uncouplers and pro-electrophiles for the different LDA models in Table 4, one might speculate that in the present descriptor space, phenols of both MOAs have a greater similarity, since the more complex 4-MOA models with higher discrimination power for pro-electrophiles provide a significantly inferior identification of the uncouplers.

Table 4 shows the results of (simulated) external validation of the LDA models using the complementary subsets *group1* and *group2*. It can be noted that preliminary leave-one-out validations based on the total set yielded classification results relatively close to the calibration statistics, which (as noted elsewhere [27]) appears to be too optimistic as measure for the true prediction capability of the models.

Taking the 3-variable model as an example, its calibration with *group2* yields significantly inferior results for the identification of respiratory uncouplers (44.4% correct classification) and pro-electrophiles (28.6%) as compared to *group1* (88.9% and 46.2%, respectively). Surprisingly, external prediction of the *group2* uncouplers from the *group1* model is significantly better than the *group2* calibration (77.8% vs. 44.4%). With respect to the pro-electrophiles, predicted from this 3-variable model, it is striking that the subgroup-specific external predictions are of similar (but low) quality as the corresponding calibrations.

The 5-variable model including log K_{ow} also showed significant differences in the validation statistics of *group1* and *group2*. Interestingly, the external MOA prediction of the *group2* compounds using the *group1* LDA model was better than when using the *group2* calibrated classification model for both respiratory uncouplers (72.2% vs. 44.4%) and soft electrophiles (86.9% vs. 66.7%). In both of these cases, however, the *group1* calibration statistics are reasonably good and in fact much better than the corresponding

group2 statistics (88.9% vs. 44.4% and 90.9% vs. 66.7%, respectively). For the pro-electrophiles, *group1* yielded inferior calibration statistics as compared to *group2* (69.2% vs. 85.7%), but a significantly better external prediction (92.6% vs. 59.3%). It shows that with respect to this MOA, the *group2* model is much more robust than the *group1* model, which is probably caused by corresponding differences in the ranges of chemical structures.

Whilst polar narcotics are correctly identified to 80 – 99% in all 4-MOA models with respect to the total set as well as to the calibration and external validation of *group1* and *group2*, the corresponding range of correct classification of the soft electrophiles is 67 – 91%. With both weak acid respiratory uncouplers and pro-electrophiles, the degree of correct classification depends much more on the specific model and data set. For the overall best LDA model based on the five descriptors $\log K_{ow}$, pK_a, E_{LUMO}, E_{HOMO} and N_{Hdon}, the identification of respiratory uncouplers is inferior to the ones of the other three MOAs when taking into account the complementary subset-specific validation and external prediction. Note that this only moderate success of identifying uncouplers is not changed when omitting picric acid due to its very low pK_a, in which case the statistics are essentially identical. Moreover, external prediction of picric acid classifies this compound as respiratory uncoupler. The latter is also observed for TFM, which contrasts with recent QSAR investigations that suggest a superposition of electrophilic and redox-cycling modes of toxic action for this compound [10].

It should be further noted that in the present LDA models, all seven tri-halogenated phenols are recognized as polar narcotics, in accord with the *a priori* assignment as discussed above. The tri-halogenated phenols may also, under certain circumstances, be considered as oxidative uncouplers. When assigned to this class, however, LDA predicted this MOA correctly for only two (2,4,6-tribromophenol and 2,4,6-tribromoresorcinol) of the seven congeners, and the rate of correct classification for uncoupling phenols was much poorer than before (52% vs. 83% for the 3-parameter LDA model, and 68% vs. 78% for the 5-parameter model). It is our contention, therefore, that the tri-halogenated phenols should be classed as polar narcotics.

As regards the two 2-variable models to separate oxidative uncoupling from the other MOAs (lower part of Table 4), both *group2* calibration and the two external validations indicate that with this specific MOA, the presently selected molecular parameters allow only a moderate degree of correct classification. It suggests that for a better separation of oxidative uncoupling from other specific modes of action, it will be necessary to include a broader range of uncoupling phenols and additional descriptors that encode more specific reactivity patterns of the different MOAs.

By contrast, the separate pro-electrophile model shows good statistics for both calibration and external validation. It follows that for predictive applications, a combined use of this pro-electrophile model together with either the 3-variable 4-MOA model or the 5-variable 4-MOA that includes $\log K_{ow}$ is recommended.

Finally, the median values of E_{LUMO}, $\log K_{ow}$ and pK_a were taken as criteria to identify compounds that represent the four MOA classes particularly well (medians are preferred over arithmetic means due to the occurrence of skewed and bimodal parameter distributions). For the group of 153 polar narcotics, 3-chloro-5-methoxyphenol (#60), 3-chlorophenol (#61), 4-iodophenol (#117) and 4-chlorophenol (#117) have parameter values close to the medians of this MOA (E_{LUMO}: 0.03 eV, $\log K_{ow}$: 2.50, pK_a: 9.54), and the 18 weak respiratory uncouplers (medians: – 1.51 eV, 2.29, 4.95) can correspondingly be represented best by 2,4-dichloro-6-nitrophenol (#159) and 2,6-dichloro-4-nitrophenol (#162). Similarly, typical examples of the 27 pro-electrophiles (medians: 0.28 eV, 1.15, 9.28) are 3-methylcatechol (#180), 4-methylcatechol (#186) and methylhydroquinone (#195), and 4-nitrocatechol (#217) as well as 2-nitrophenol (#204) come closest to the medians of E_{LUMO}, $\log K_{ow}$ and pK_a of the 23 soft electrophiles (– 1.19 eV, 1.85, 6.40).

7 Summary and Conclusions

Linear discriminant analysis based on molecular hydrophobicity, acidity, electron affinity, oxidation potential and hydrogen-bond donor capacity allows a satisfactory classification of phenols with respect to four underlying modes of toxic action (MOAs). Whilst 4-MOA models performed reasonably well to identify polar narcotics and soft electrophiles, pro-electrophiles can be significantly better discriminated from other MOAs by using a separate MOA model that differentiates only between pro-electrophiles and compounds exerting other MOAs. With respiratory uncouplers, classification and external prediction is inferior to the other three MOAs, and cannot be improved when discriminating only between uncouplers and other compounds. Moreover, hydrophobicity with correction for ionisation is significantly inferior to conventional $\log K_{ow}$ in contributing to the discrimination between different MOAs. The latter two findings are surprising and need attention in future mechanistic investigations of weak acid respiratory uncouplers. For predictive applications, a combined use of one 4-MOA model and the pro-electrophile model is recommended.

Acknowledgements

This work was supported by the European Union IMAGETOX Research Training Network (HPRN-CT- 1999-00015) and by The University of Tennessee Centre of Excellence in Livestock Disease and Human Health.

References

[1] Ahlborg, U. G., and Thunberg, T., Chlorinated phenols: Occurrence, toxicity, metabolism, and environmental impact. *CRC Crit. Rev. Toxicol. 7*, 1 – 35 (1980).

[2] Koch, R., *Umweltchemikalien. Physikalisch-chemische Daten, Toxizitäten, Grenz- und Richtwerte, Umweltverhalten*. VCI Weinheim 1991, 426 pp.

[3] Hubert, T. D., Jeffry, A. B., Vue, C., and Gingerich, W. H., Uptake, metabolism and elimination of TFM by fish. In: Smith, D. J., Gingerich, W. H., Beconi-Barker, M. G. (Eds.), *Xenobiotics in Fish*. Kluwer Academic / Plenum Publishers, New York 1999, pp. 177 – 187.

[4] Garg, R., Kurup, A., and Hansch, C., Comparative QSAR: On the toxicology of the phenolic OH moiety. *Crit. Rev. Toxicol. 31*, 223 – 245 (2001).

[5] Schultz, T. W., Sinks, G. D., and Cronin, M. T. D., Identification of mechanisms of toxic action of phenols to *Tetrahymena pyriformis* from molecular descriptors. In: Chen, F., Schüürmann, G. (Eds.): *Quantitative Structure-Activity Relationships in Environmental Sciences – VII*. SETAC Press, Pensacola FL 1997, pp. 329 – 342.

[6] Schüürmann, G., Somashekar, R. K., and Kristen, U., Structure-activity relationships for chloro- and nitrophenol toxicity in the pollen tube growth test. *Environ. Toxicol. Chem. 15*, 1702 – 1708 (1996).

[7] Schüürmann, G., Segner, H., and Jung, K., Multivariate mode-of-action analysis of acute toxicity of phenols. *Aquat. Toxicol. 38*, 277 – 296 (1997).

[8] Schultz, T. W., and Cronin, M. T. D., Quantitative structure-activity relationships for weak acid respiratory uncouplers to *Vibrio fischeri*. *Environ. Toxicol. Chem. 16*, 357 – 360 (1997).

[9] Schüürmann, G., Ecotoxic modes of action of chemical substances, in: Schüürmann, G., Markert, B. (Eds.), Ecotoxicology. John Wiley and Spektrum Akademischer Verlag, New York 1998, pp. 665 – 749.

[10] Schmitt, H., Altenburger, R., Jastorff, B., and Schüürmann, G., Quantitative structure-activity analysis of the algae toxicity of nitroaromatic compounds. *Chem. Res. Toxicol. 13*, 441 – 450 (2000).

[11] Livingstone D. J., *Data Analysis for Chemists: Application to QSAR and Chemical Product Design.* Oxford University Press, Oxford 1995.

[12] Mitchell, P., Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature 191*, 144 – 148 (1961).

[13] Lipnick, R. L., Outliers: Their origin and use in the classification of molecular mechanisms of toxicity. *Sci. Total Environ. 109/110*, 131 – 153 (1991).

[14] O'Brien, P. J., Molecular mechanisms of quinone cytotoxicity. *Chem.-Biol. Interactions 80*, 1 – 41 (1991).

[15] Garg, R., Kurup, S., and Hansch, C., Radical toxicity of phenols: a reference point for obtaining perspective in the formulation of QSAR. *Med. Res. Rev. 21*, 73 – 82 (2001).

[16] Kaufman, R. D., Biophysical mechanisms of anesthetic action: historical perspectives and review of current theories. *Anesthesiology 46*, 49 – 62 (1977).

[17] Bradbury, S. P., Henry, T. R., Niemi, G. J., Carlson, R. W., and Snarski, V. M., Use of respiratory-cardiovascular responses of rainbow trout (*Oncorhynchus mykiss*) in identifying acute toxicity syndromes in fish. Part 3: Polar narcotics. *Environ. Toxicol. Chem. 8*, 247 – 261 (1989).

[18] Schultz, T. W., *Tetrahymena* in aquatic toxicology: QSARs and ecological hazard assessment, in: Pauli, W., Berger. S. (Eds.), *Proceedings of the International Workshop on a Protozoan Test Protocol with Tetrahymena in Aquatic Toxicity Testing.* Umweltbundesamt (UBA, Federal Environmental Agency), Research Report No. 106 99 999/03, Berlin (Germany) 1996, pp. 31 – 66.

[19] Leo, A., and Weininger, D. MedChem, Release 3.53. Pomona medicinal Chemistry project, Pomona College, Claremont CA.

[20] Hunter, R. S., Computer calculation of pK_a. In: Turner, J. E., England, M. W., Schultz, T. W., Kwaak, N. J. (Eds.), *QSAR 88. Proceedings of the Third International Workshop on Quantitative Structure-Toxicity Relationships in Environmental Toxicology.* USDOE Office of Scientific and Technical Information, Oak Ridge (Tennessee, USA) 1988, pp. 33 – 41.

[21] Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P., AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc. 107*, 3902 – 3909 (1985).

[22] Iwai, Y., Yoshiba, K., Arai, Y., Schüürmann, G., Loeprecht, B., Fabian, W. M. F., and Suzuki, T., Analysis of the flow patterns of liquid organic compounds between blade electrodes by classification models. *J. Chem. Inf. Comput. Sci. 40*, 988 – 993 (2000).

[23] Mishra, R. K., Garcia-Domenech, R., and Galvez, J., Getting discriminant function of antibacterial activity from physicochemical and topological parameters. *J. Chem. Inf. Comput. Sci. 41*, 387 – 393 (2001).

[24] Suzuki, T., Timofei, S., Kurunczi, L., Dietze, U., and Schüürmann, G., Correlation of aerobic biodegradability of sulfonated azo dyes with the chemical structure. *Chemosphere 45*, 1 – 9 (2001).

[25] Dietze, U., Braunbeck, T., Honnen, W., Köhler, H.-R., Schwaiger, J., Segner, H., Triebskorn, R., and Schüürmann, G., Chemometric discrimination between streams based on chemical, limnological and biological data taken from freshwater fishes and their interrelationships. *J. Aquat. Ecosyst. Stress Recov. 8*, 319 – 336 (2001).

[26] Worth, A. P., and Cronin, M. T. D., The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effect. *J. Mol. Struct. THEOCHEM*, in press (2002).

[27] Boháč, M., Loeprecht, B., Damborsky, J., and Schüürmann, G., Impact of orthogonal signal correction (OSC) on the predictive ability of CoMFA models for the ciliate toxicity of nitrobenzenes. *Quant. Struct.-Act. Relat.*, submitted (2001).

[28] Dearden, J. C., Cronin, M. T. D., Zhao, Y.–H., and Raevsky, O. A., QSAR studies of compounds acting by polar and nonpolar narcosis: An examination of the role of polarisability and hydrogen bonding. *Quant. Struct.-Act. Relat. 14*, 427 – 432 (2000).

received on November 20, 2001; accepted on February 1, 2002