Available online at www.sciencedirect.com
ScienceDirect
Chemometrics and Intelligent Laboratory Systems 90 (2008) 97–107
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Prediction of aqueous toxicity for heterogeneous phenol derivatives by QSAR

Pablo R. Duchowicz, Andrew G. Mercader *, Francisco M. Fernández, Eduardo A. Castro

Research Institute of Theoretical and Applied Physical Chemistry (INIFTA), Theoretical Chemistry Division, Faculty of Exact Sciences, Chemistry Department, La Plata National University, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina

Received 25 June 2007; accepted 12 August 2007
Available online 22 August 2007

Abstract

We provide QSAR models for the growth inhibition of the ciliated protozoan *Tetrahymena pyriformis* by 250 mechanistically diverse phenolic compounds. The simultaneous linear regression analysis on 1338 topological, geometrical, and electronic molecular descriptors over 200 molecules leads to a seven-parameter relationship with $R=0.851$ and leave more out $R_{1-60\%-o}=0.730$, while a model based on flexible descriptors improves to $R=0.880$ and $R_{1-60\%-o}=0.812$. An external test set of 50 related derivatives demonstrates that both models show good predictive power with rms = *0.418* and rms = *0.352*, respectively, comparing fairly well with previously reported Artificial Neural Networks with similar rms. Finally, we employ the best QSAR equation to estimate the unknown aqueous toxicity of 74 structures.

Keywords: QSAR studies; Flexible descriptors; Replacement method; Phenol derivatives; *Tetrahymena pyriformis*; IGC_{50}

1. Introduction

Organic chemicals carrying the structure of phenol have been in production since the 1860s, and include a wide number of applications in various industries such as textile, leather, paper and oil. For example, salicylic acid is used during the production of Aspirin and other pharmaceuticals; chlorophenols are utilized in Agriculture to manufacture a range of pesticides; alkylphenols are involved in the production of surfactants and detergents; bisphenol A is used to synthesize epoxy resins for paint coatings and mouldings, and in polycarbonate plastics, familiar in CDs and domestic electrical appliances. Despite their great importance, the main drawback of employing phenolic compounds is the resulting pollution both in aquatic and terrestrial ecosystems. Therefore, the accurate estimation of the adverse environmental impact poses great interest in the scientific community, together with a convenient way to regulate their production [1,2].

It is known that performing a toxicological experiment for a given substance is not an easy task as it usually results expensive, requires time and, furthermore, an analysis of such dimension should consider multiple environments and all biological interactions with the living organisms of the ecosystems, data that quite often are not available [3]. A generally accepted strategy for overcoming the absence of experimental measurements in complex biological systems is the analysis based on Quantitative Structure–Activity Relationships (QSAR) [4], in the present study this analysis will be done on aqueous toxicity of phenol derivatives. An obvious advantage of this sort of studies is to minimize animal testing. The ultimate role of formulating the QSAR is to suggest mathematical models estimating the toxicities by relying on the assumption that these are determined solely by the molecular structures of the phenolic compounds. The structure is therefore translated into the so-called molecular descriptors, describing some relevant feature of the compounds, with mathematical formulae obtained from Chemical Graph Theory, Information Theory, Quantum Mechanics, etc. [5,6]. There exist more than a thousand available descriptors in the literature, and one has to decide how to select those that characterize the property under consideration in the best possible manner.

* Corresponding author. Fax: +54 221 425 4642.
E-mail address: andrewmercader@yahoo.com (A.G. Mercader).

0169-7439/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2007.08.006

Table 1
Experimental and predicted values for 250 $pIGC_{50}$

no	Compound name	$pIGC_{50}$			
		Exp.	Eq. (4)	Eq. (7)	ANN [10]
1	4-Hydroxyphenylacetic acid	−1.50	−0.42	−0.51	−1.06
2	3-Hydroxybenzyl alcohol	−1.04	−0.41	−0.60	−0.60
3	4-Carboxyphenol	−1.02	−0.40	−0.41	−0.59
4	3-Hydroxy-4-methoxybenzyl alcohol	−0.99	−0.38	−0.46	−0.73
5	4-Hydroxy-3-methoxybenzyl amine	−0.97	−0.26	−0.39	−0.59
6	4-Hydroxyphenethyl alcohol	−0.83	−0.52	−0.62	−0.45
7	3-Carboxyphenol	−0.81	−0.25	−0.32	−0.49
8	4-Hydroxybenzamide	−0.78	−0.33	−0.31	−0.40
9	4-Hydroxy-3-methoxybenzyl alcohol	−0.70	−0.35	−0.47	−0.79
10	2,6-Dimethoxyphenol	−0.60	−0.21	−0.38	−0.50
11	2,4,6-Tris(dimethylaminomethyl) phenol	−0.52	−0.36	−0.47	−0.04
12	Salicylic acid	−0.51	−0.16	−0.27	−0.54
13	2-Methoxyphenol	−0.51	−0.08	−0.13	−0.35
14	5-Methylresorcinol	−0.39	0.00	0.06	−0.25
15	4-Methylcyanophenol	−0.38	0.26	0.25	−0.09
16	3-Hydroxyacetophenone	−0.38	−0.22	−0.43	0.28
17	2-Ethoxyphenol	−0.36	−0.08	−0.20	−0.13
18	4-Acetylphenol	−0.30	−0.31	−0.48	0.30
19	3-Ethoxy-4-methoxyphenol	−0.30	0.02	−0.06	−0.21
20	2-Methylphenol	−0.29	0.20	0.11	−0.26
21	2-Hydroxybenzamide	−0.24	−0.10	−0.19	−0.02
22	Phenol	−0.21	−0.04	−0.04	−0.17
23	4-Methylphenol	−0.18	0.03	0.04	0.00
24	4-Hydroxy-3-methoxyphenethylalcohol	−0.18	−0.28	−0.36	−0.66
25	3-Acetamidophenol	−0.16	0.06	0.07	−0.41
26	3-Hydroxy-4-methoxybenzaldehyde	−0.14	0.05	0.02	0.11
27	4-Hydroxy-3-methoxyacetophenone	−0.12	−0.09	−0.19	0.13
28	3,5-Dimethoxyphenol	−0.09	−0.33	−0.22	−0.27
29	2-Hydroxyethylsalicylate	−0.08	0.47	0.14	0.14
30	3-Methylphenol	−0.06	0.14	0.12	0.01
31	Methyl-3-hydroxybenzoate	−0.05	0.38	0.22	0.24
32	3-Methoxy-4-hydroxybenzaldehyde	−0.03	0.03	−0.01	0.12
33	4-Hydroxy-3-methoxybenzonitrile	−0.03	0.10	0.03	0.16
34	3-Ethoxy-4-hydroxybenzaldehyde	0.01	0.06	−0.02	0.22
35	4-Fluorophenol	0.02	0.40	0.44	0.08
36	2-Cyanophenol	0.03	−0.02	−0.01	0.18
37	5-Fluoro-2-hydroxyacetophenone	0.04	0.34	0.22	0.70
38	2,4-Dimethylphenol	0.07	0.39	0.34	0.18
39	2-Hydroxyacetophenone	0.08	−0.14	−0.32	0.51
40	2,5-Dimethylphenol	0.08	0.40	0.38	0.27
41	Methyl-4-hydroxybenzoate	0.08	0.22	0.08	0.41
42	3,5-Dimethylphenol	0.11	0.26	0.27	0.36
43	4'-Hydroxypropiophenone	0.12	0.10	0.04	0.66
44	2,3-Dimethylphenol	0.12	0.44	0.38	0.16
45	3,4-Dimethylphenol	0.12	0.45	0.39	0.15
46	2-Ethylphenol	0.16	0.41	0.24	0.26
47	Syringaldehyde	0.17	−0.05	−0.05	−0.06
48	Salicylhydrazide	0.18	−0.13	−0.21	−0.12
49	2-Chlorophenol	0.18	0.64	0.64	0.26
50	4-Hydroxy-2-methylacetophenone	0.19	0.13	0.08	0.37
51	4-Ethylphenol	0.20	0.22	0.13	0.43
52	3-Ethylphenol	0.23	0.36	0.22	0.33
53	Salicylaldoxime	0.25	-0.16	0.22	0.31
54	2,3,6-Trimethylphenol	0.28	0.75	0.87	0.65
55	2,4,6-Trimethylphenol	0.28	0.63	0.48	0.66
56	2-Hydroxy-5-methylacetophenone	0.31	0.19	0.09	0.61
57	2-Bromophenol	0.33	0.75	0.82	0.75
58	5-Bromo-2-hydroxybenzyl alcohol	0.34	0.48	0.65	0.27
59	2,3,5-Trimethylphenol	0.36	0.77	0.74	0.75
60	3-Methoxysalicylaldehyde	0.38	0.03	0.04	0.10
61	Salicylhydroxamic acid	0.38	−0.17	−0.25	0.13
62	2-Chloro-5-methylphenol	0.39	0.87	0.85	0.80
63	4-Allyl-2-methoxyphenol	0.42	0.13	0.07	0.32
64	2-Hydroxybenzaldehyde	0.42	0.03	0.09	0.21
65	2,6-Difluorophenol	0.47	0.88	0.64	0.61
66	Ethyl-3-hydroxybenzoate	0.48	0.37	0.14	0.51
67	4-Cyanophenol	0.52	0.00	0.04	0.19
68	4-Propyloxyphenol	0.52	0.11	0.20	0.38
69	4-Chlorophenol	0.55	0.62	0.70	0.52
70	Ethyl-4-hydroxybenzoate	0.57	0.28	0.07	0.70
71	5-Methyl-2-nitrophenol	0.59	0.63	0.77	1.27
72	2-Bromo-4-methylphenol	0.60	0.90	0.90	1.06
73	2,4-Difluorophenol	0.60	1.00	0.97	0.41
74	3-Isopropylphenol	0.61	0.56	0.57	0.61
75	5-Bromovanillin	0.62	0.73	0.54	0.46
76	α,α,α-Trifluoro-4-cresol	0.62	0.21	0.35	0.67
77	Methyl-4-methoxysalicylate	0.62	0.44	0.24	0.34
78	4-Bromophenol	0.68	0.84	0.96	0.70
79	2-Chloro-4,5-dimethylphenol	0.69	1.12	1.08	0.79
80	4-Butoxyphenol	0.70	1.26	1.42	0.78
81	4-Chloro-2-methylphenol	0.70	0.88	0.89	0.90
82	3-tert-Butylphenol	0.73	0.72	0.72	0.93
83	2,6-Dichlorophenol	0.73	0.87	0.78	0.73
84	2-Methoxy-4-propenyIphenol	0.75	1.00	0.82	0.96
85	3-Chloro-5-methoxyphenol	0.76	0.56	0.57	0.46
86	4-Chloro-3-methylphenol	0.80	0.83	0.85	0.64
87	2-Isopropylphenol	0.80	0.55	0.50	0.53
88	2,6-Dichloro-4-fluorophenol	0.80	1.38	1.32	0.87
89	4-Iodophenol	0.85	1.20	1.26	1.10
90	2,2'-Biphenol	0.88	0.66	0.71	0.23
91	4-tert-Butylphenol	0.91	0.60	0.64	0.97
92	3,4,5-Trimethylphenol	0.93	0.66	0.64	0.64
93	2,2',4,4'-Tetrahydroxybenzophenone	0.96	1.39	1.52	0.97
94	4-sec-Butylphenol	0.98	0.72	0.74	1.11
95	3-Hydroxydiphenylamine	1.01	1.30	1.42	0.81
96	4-Hydroxybenzophenone	1.02	0.91	1.01	0.86
97	2,4-Dichlorophenol	1.04	1.23	1.33	1.33
98	2,4,6-Tribromoresorcinol	1.06	1.28	1.15	1.23
99	Benzyl-4-hydroxyphenyl ketone	1.07	0.88	0.96	0.75
100	4-Chloro-3-ethylphenol	1.08	0.91	0.87	1.22
101	2-Phenylphenol	1.09	1.19	1.17	1.19
102	2,5-Dichlorophenol	1.13	1.04	1.19	1.41
103	3-Chloro-4-fluorophenol	1.13	1.23	1.11	0.67
104	3-Bromophenol	1.15	0.91	0.99	0.90
105	6-tert-Butyl-2,4-dimethyIphenol	1.16	1.44	1.62	1.07
106	4-Chloro-3,5-dirnethyIphenol	1.20	0.82	0.94	0.93
107	2-Hydroxybenzophenone	1.23	0.99	0.96	1.15
108	4-tert-Pentylphenol	1.23	1.03	1.04	1.43
109	4-Bromo-3,5-dimethyIphenol	1.27	0.90	1.07	1.31
110	4-Bromo-6-chloro-2-cresol	1.28	1.55	1.72	1.44
111	4-Cyclopentylphenol	1.29	2.06	1.34	1.08
112	2-tert-Butylphenol	1.29	0.72	1.07	0.81
113	2-tert-Butyl-4-methyIphenol	1.30	1.10	1.47	1.13
114	2-Hydroxydiphenylmethane	1.31	1.04	1.15	1.22
115	Butyl-4-hydroxybenzoate	1.33	1.07	1.01	1.35
116	3-PhenyIphenol	1.35	1.21	1.30	1.33