On the Parametrization of the Toxicity of Organic Chemicals to *Tetrahymena pyriformis*. The Problem of Establishing a Uniform Activity

Suresh Babu Mekapati and Corwin Hansch*

Department of Chemistry, Pomona College, Claremont, California 91711

Received December 22, 2001

In this report we illustrate the importance of making an effort to ensure that all of the dependent variables in a QSAR are associated with one mechanism of action. If this cannot be established, it will not be possible to compare the QSAR with others acting on different biological systems. That is, such information cannot be used to develop a science of chemical−biological interactions. A study of the action of a large set of phenols acting on *Tetrahymena pyriformis* made by Cronin and Schultz is analyzed to illustrate the problem.

INTRODUCTION

The field of quantitative structure−activity relationships (QSAR) continues to develop at very rapid pace with new approaches of all kinds constantly being published. Naturally those working in this area are eager to include as many congeners in one equation as possible. Rather little thought is given to forming equations that can be compared to other QSAR. We believe that this is most essential if a science of chemical−biological interactions is to be constructed. It is a mistake to lump together molecules that do not operate by the same mechanism. That is the dependent variable is not uniform. To understand this problem one needs rather large sets of homologues that have been tested in the same laboratory to ensure reliable data. In a recent publication on the toxicity of a wide variety of 268 organic chemicals to the protozoa *Tetrahymena pyriformis* Cronin and Schultz derived the following QSAR.

$$\log 1/C = 0.60\ \log P - 0.37\ \mathrm{LUMO} - 0.981$$

$$N = 253,\ r^2 = 0.751,\ s = 0.391 \qquad (1)$$

Obviously this is not a very satisfactory result since r^2 is very low. They made a study of a wide variety of parameters used in correlation analysis and emphasized the importance of using "clearly transparent" parameters for correlation analysis. That is, parameters that can be clearly understood in mechanistic terms that would allow comparison with other QSAR. Certainly, if we are ever to develop a science to rationalize the interactions of chemicals with living organisms or parts there of (DNA, enzymes, cells, etc.) this is the only way to do it. So far none of these new approaches allows one to do comparative QSAR. Certainly logP and quantum mechanical parameters such as HOMO and LUMO have been very well justified in countless studies. Our current database of 17 200 QSAR contains 8600 for biological reactions and 8600 from mechanistic organic chemistry for comparison. There are an increasing number based on molecular orbital parameters (MO). Of the biological QSAR 66 are based on LUMO or HOMO. However, our program keeps track of parameters tried for developing each QSAR. Thus we find 134 instances where such efforts were made. The difference in these numbers, 68, represents instances where the Hammett type parameters were better. Actually for practical purposes the two approaches are similar. MO parameters have the advantage in that they can be used for more complex molecules where Hammett parameters are lacking. Hammett parameters have the advantage that in our system they can be automatically loaded for regression analysis. It would have been completely impossible to have developed our system using quantum chemical parameters because of the time needed for their calculations.

* Corresponding author phone: (909)621-8445; fax: (909)607-7726; e-mail: atessier@pomona.edu.

RESULTS

So much for the right side of a QSAR equation. A much more difficult problem is that of ensuring that the left side, dependent variable, is uniform in composition. Here it is almost impossible to be sure that all "congeners" are acting by a uniform mechanism. There are many ways that the components of a cell can be effected by a set of organic compounds. About the only assurance of a uniform mechanism is a high r^2 *with* a q^2 close in value. What constitutes high enough is not clear, but it depends heavily on the quality of the experimental work. For the present study we have a set a level of r^2 and q^2 of 0.90.

With this background we now consider 133 phenols (Table 1) from the work of Cronin and Schultz.[1] Starting with 133 data points we reduced the set to 106 data points by jackknifing on LUMO + MlogP until we reached an r^2 of 0.908 with $q^2 = 0.902$. We assume that such quality of fit would ensure a reasonable uniformity of action. The result is QSAR 2 and 3.

$$\log 1/C = -0.61(\pm 0.07)\mathrm{LUMO} + 0.62(\pm 0.04)\mathrm{MlogP} - 0.96(\pm 0.13)$$

$$n = 106,\ r^2 = 0.908,\ s = 0.232,\ q^2 = 0.902 \qquad (2)$$

Doing the same with σ, QSAR 3 is obtained.

$$\log 1/C = 0.75(\pm 0.09)\sigma + 0.62(\pm 0.04)\mathrm{ClogP} - 1.08(\pm 0.13)$$

$$n = 109,\ r^2 = 0.909,\ s = 0.229,\ q^2 = 0.902 \qquad (3)$$

10.1021/ci010250e CCC: $22.00
Published on Web 05/24/2002

Table 1: Growth Inhibition Concentration (IGC_{50}) of Phenols against *Tetrahymena Pyriformis*

		log1/C								
s. no.	substituents	obsd	calcd eq 2	Δ	calcd eq 3	Δ	LUMO	MlogP	ClogP	σ
1	2,6-di-OCH_3	−0.60	−0.49	−0.11	−0.79	0.19	0.388	1.15	1.11	−0.54
2	3-NH_2[a]	−0.52	−1.17	0.65	−1.05	0.53	0.541	0.21	0.25	−0.16
3	2-COOH[a,b]	−0.51	0.73	−1.24	0.62	−1.13	−0.457	2.26	2.19	0.45
4	2-OCH_3	−0.51	−0.38	−0.13	−0.46	−0.05	0.392	1.32	1.32	−0.27
5	4-CH_2CN	−0.38	−0.44	0.06	−0.39	0.01	0.064	0.90	0.90	0.18
6	2-OCH_2CH_3	−0.36	−0.07	−0.29	−0.11	−0.25	0.421	1.85	1.85	−0.24
7	3-OCH_3	−0.33	−0.25	−0.08	−0.01	−0.32	0.434	1.58	1.57	0.12
8	2-CH_3	−0.30	0.01	−0.31	−0.01	−0.29	0.397	1.95	1.92	−0.17
9	3-OC_2H_5-4−OCH_3	−0.30	−0.18	−0.13	−0.10	−0.20	0.318	1.58	1.79	−0.17
10	H	−0.21	−0.29	0.08	−0.16	−0.05	0.397	1.47	1.48	0
11	4-CH_3	−0.18	−0.02	−0.16	0.02	−0.20	0.428	1.94	1.97	−0.17
12	3-$NHCOCH_3$	−0.16	−0.64	0.48	−0.62	0.46	0.205	0.73	0.49	0.21
13	2,4,6-tri-NO_2[a,b]	−0.16	1.15	−1.31	1.69	−1.85	−2.534	0.89	1.64	2.34
14	4-OCH3	−0.14	−0.32	0.18	−0.30	0.16	0.303	1.34	1.57	−0.27
15	3,5-di-OCH_3	−0.09	−0.20	0.11	0.10	−0.19	0.415	1.64	1.61	0.24
16	4-NH_2[a,b]	−0.08	−1.21	1.13	−1.42	1.34	0.439	0.04	0.25	−0.66
17	3-CH_3	−0.06	0.03	−0.09	0.10	−0.16	0.377	1.96	1.97	−0.07
18	3-CN	−0.06	0.41	−0.47	0.33	−0.39	−0.514	1.70	1.60	0.56
19	4-OC_2H_5	0.01	−0.04	0.05	0.05	−0.04	0.337	1.81	2.10	−0.24
20	4-F	0.02	0.10	−0.08	0.16	−0.14	0.059	1.77	1.92	0.06
21	2-CN	0.03	0.35	−0.32	0.41	−0.38	−0.509	1.61	1.60	0.66
22	2,4-di-CH_3	0.07	0.22	−0.15	0.17	−0.10	0.403	2.30	2.42	−0.34
23	2,5-di-CH_3	0.08	0.28	−0.20	0.25	−0.17	0.345	2.33	2.42	−0.24
24	3,5-di-CH_3	0.11	0.27	−0.16	0.35	−0.24	0.383	2.35	2.47	−0.14
25	3,4-di-CH_3	0.12	0.16	−0.04	0.25	−0.13	0.436	2.23	2.42	−0.24
26	2,3-di-CH_3	0.12	0.31	−0.19	0.22	−0.10	0.386	2.42	2.37	−0.24
27	2,4-di-NH_2[a,b]	0.13	−1.68	1.81	−2.44	2.57	0.544	−0.61	−0.61	−1.32
28	2-C_2H_5	0.16	0.34	−0.18	0.33	−0.17	0.386	2.47	2.45	−0.15
29	2-Cl	0.18	0.36	−0.18	0.43	−0.25	0.029	2.15	2.16	0.23
30	2-F	0.19	0.09	0.10	0.03	0.16	0.013	1.71	1.72	0.06
31	4-C_2H_5	0.21	0.31	−0.10	0.36	−0.15	0.435	2.47	2.50	−0.15
32	3-C_2H_5	0.23	0.29	−0.05	0.42	−0.19	0.402	2.40	2.50	−0.07
33	3,4-di-NO_2[a,b]	0.27	1.29	−1.02	1.16	−0.89	−1.850	1.79	1.82	1.49
34	2,3,6-tri-CH_3	0.28	0.62	−0.34	0.37	−0.09	0.382	2.92	2.82	−0.41
35	2,4,6-tri-CH_3	0.28	0.62	−0.34	0.33	−0.05	0.433	2.97	2.87	−0.51
36	2-Br	0.33	0.53	−0.20	0.56	−0.23	−0.049	2.35	2.36	0.23
37	2-allyl	0.33	0.41	−0.08	0.37	−0.04	0.348	2.55	2.50	−0.14
38	2,3,5-tri-CH_3	0.36	0.64	−0.28	0.47	−0.11	0.360	2.92	2.87	−0.31
39	2-NH_2-4-t−C_4H_9[a]	0.37	0.24	0.13	−0.20	0.57	0.522	2.44	2.44	−0.86
40	3-OCH_3-2−CHO	0.38	0.17	0.21	0.51	−0.13	−0.454	1.37	1.90	0.54
41	3-F	0.38	0.22	0.16	0.36	0.02	0.025	1.93	1.92	0.34
42	2-Cl-5-CH_3	0.39	0.65	−0.26	0.69	−0.30	0.058	2.65	2.65	0.16
43	5-NH_2-2−OCH_3[a,b]	0.45	−1.09	1.54	−1.31	1.76	0.357	0.15	0.15	−0.43
44	2,3-di-NO_2[a,b]	0.46	1.26	−0.80	1.16	−0.70	−1.799	1.79	1.82	1.49
45	2,6-di-F	0.47	0.51	−0.04	0.10	0.37	−0.321	2.05	1.76	0.12
46	4-*i*-C_3H_7	0.47	0.57	−0.10	0.61	−0.14	0.446	2.90	2.90	−0.15
47	2-NH_2-4−NO_2[b]	0.48	0.59	−0.11	−0.26	0.74	−0.976	1.53	1.18	0.12
48	3-NO_2[a]	0.51	0.99	−0.48	0.60	−0.09	−1.159	2.0	1.85	0.71
49	4-CN	0.52	0.29	0.23	0.41	0.12	−0.412	1.60	1.60	0.66
50	2,6-di-NO_2[a,b]	0.54	1.09	−0.55	1.21	−0.67	−1.952	1.37	1.82	1.56
51	4-Cl	0.55	0.47	0.08	0.64	−0.07	0.094	2.39	2.49	0.23
52	4-CH_3-2−NO_2[a]	0.57	1.11	−0.54	0.84	−0.27	−0.967	2.37	2.35	0.61
53	2-Br-4-CH_3	0.60	0.82	−0.22	0.74	−0.14	−0.010	2.85	2.85	0.06
54	2,4-di-F	0.60	0.45	0.15	0.23	0.37	−0.318	1.95	1.96	0.12
55	3-*i*-C_3H_7	0.61	0.60	0.01	0.69	−0.08	0.398	2.90	2.90	−0.04
56	2,6-di-Cl-4-NO_2[a,b]	0.63	1.75	−1.12	1.56	−0.93	−1.441	2.94	2.76	1.24
57	4-C_3H_7	0.64	0.64	0	0.71	−0.07	0.432	3.00	3.03	−0.13
58	4-NO	0.65	0.33	0.32	0.44	0.21	−0.799	1.29	1.36	0.91
59	2-NO_2	0.67	0.77	−0.10	0.65	0.02	−1.014	1.79	1.85	0.78
60	4-Br	0.68	0.64	0.04	0.73	−0.05	0.020	2.59	2.64	0.23
61	2-Cl-4,5-di-CH_3	0.69	0.94	−0.25	0.84	−0.15	0.053	3.10	3.10	−0.01
62	4-Cl-2-CH_3	0.70	0.69	0.01	0.79	−0.09	0.134	2.78	2.93	0.06
63	4-OC_4H_9	0.70	0.64	0.06	0.65	0.05	0.330	2.90	3.16	−0.32
64	3-*t*-C_4H_9	0.73	0.84	−0.11	0.90	−0.17	0.416	3.30	3.30	−0.10
65	4-CH_3-3−NO_2	0.74	1.13	−0.39	0.74	0.01	−1.111	2.27	2.27	0.54
66	2,6-di-Cl	0.74	0.91	−0.17	0.91	−0.17	−0.258	2.75	2.64	0.46
67	2-CH_2Cl-4−NO_2[a]	0.75	1.28	−0.53	1.06	−0.31	−1.195	2.42	2.37	0.90
68	3-Cl-5-OCH_3	0.76	0.56	0.20	0.85	−0.09	0.060	2.50	2.51	0.49
69	2-NH_2-4−Cl[a,b]	0.78	0.06	0.72	−0.34	1.12	0.172	1.81	1.71	−0.43
70	2-*i*-C_3H_7	0.80	0.58	0.22	0.49	0.31	0.408	2.88	2.70	−0.15

Table 1. (Continued)

s. no.	substituents	log1/C obsd	log1/C calcd eq 2	log1/C Δ	log1/C calcd eq 3	log1/C Δ	LUMO	MlogP	ClogP	σ
71	2,6-di-Cl-4-F	0.80	1.13	−0.33	1.06	−0.26	−0.568	2.80	2.82	0.52
72	4-Cl-3-CH_3	0.80	0.91	−0.11	0.89	−0.09	0.093	3.10	2.98	0.16
73	4-I	0.85	0.84	0.02	0.85	0.00	0.023	2.91	2.90	0.18
74	3-Cl	0.87	0.58	0.29	0.74	0.13	0.019	2.50	2.50	0.37
75	4-NH_2-2−NO_2[a,b]	0.88	0.32	0.56	−0.49	1.37	−1.119	0.96	0.81	0.12
76	6-NH_2-2,4-di−CH_3[a,b]	0.89	−0.23	1.12	−0.85	1.74	0.442	1.62	1.57	−1.00
77	4-*t*-C_4H_9	0.91	0.81	0.10	0.82	0.09	0.463	3.31	3.30	−0.20
78	3,4,5-tri-CH_3	0.93	0.56	0.37	0.47	0.46	0.430	2.87	2.87	−0.31
79	2-NH_2[a,b]	0.94	−0.87	1.81	−1.19	2.13	0.474	0.62	0.62	−0.66
80	2,5-di-NO_2[a]	0.95	1.51	−0.56	1.16	−0.21	−2.261	1.75	1.82	1.49
81	2-sec-C_4H_9	0.98	0.68	0.30	0.84	0.14	0.445	3.08	3.23	−0.12
82	2,4-di-Cl	1.04	1.09	−0.05	1.11	−0.07	−0.245	3.06	2.97	0.46
83	4-Cl,3-C_2H_5	1.08	1.14	−0.06	1.22	−0.14	0.140	3.51	3.51	0.16
84	2,4-di-NO_2	1.08	1.24	−0.16	1.21	−0.13	−1.888	1.67	1.82	1.56
85	3-I	1.12	0.91	0.22	0.98	0.14	−0.070	2.93	2.90	0.35
86	5-F-2-NO_2	1.13	1.02	0.11	1.06	0.07	−1.297	1.91	2.10	1.12
87	3-Cl-4-F	1.13	0.91	0.22	0.94	0.19	−0.292	2.72	2.73	0.43
88	2,5-di-Cl	1.13	1.14	−0.01	1.21	−0.08	−0.325	3.06	2.97	0.60
89	3-Br	1.14	0.72	0.42	0.85	0.29	−0.074	2.63	2.64	0.39
90	6-*t*-C_4H_9-2,4-di−CH_3	1.16	1.44	−0.28	1.10	0.06	0.455	4.30	4.15	−0.54
91	2,3,5,6-tetra-F[b]	1.17	0.94	0.23	0.78	0.39	−0.994	2.07	2.03	0.80
92	2-NH_2-4-Cl-5−NO_2[a]	1.17	0.75	0.42	0.29	0.88	−0.961	1.80	1.87	0.28
93	4-Br-2,6-di-CH_3	1.17	1.25	−0.08	1.03	0.14	0.085	3.63	3.53	−0.11
94	4-Cl-3,5-di-CH_3	1.20	1.11	0.09	1.15	0.05	0.147	3.48	3.48	0.09
95	2,6-di-NO_2-4−CH_3	1.23	1.63	−0.40	1.40	−0.17	−1.894	2.29	2.32	1.39
96	4-*t*-C_5H_{11}	1.23	1.13	0.10	1.17	0.06	0.469	3.83	3.83	−0.17
97	4-Br-3,5-di-CH_3	1.27	1.23	0.04	1.25	0.03	0.109	3.63	3.63	0.09
98	4-BR-6-Cl	1.28	1.43	−0.15	1.20	0.08	−0.233	3.61	3.12	0.46
99	2,3-di-Cl	1.28	0.97	0.31	1.14	0.14	−0.262	2.84	2.85	0.60
100	4-cy-C_5H_9	1.29	0.97	0.32	1.01	0.28	0.436	3.54	3.54	−0.14
101	2-*t*-C_4H_9-4−CH_3	1.30	1.11	0.19	0.94	0.36	0.477	3.80	3.70	−0.37
102	2-*t*-C_4H_9[a,b]	1.30	0.83	0.47	0.76	0.54	0.436	3.31	3.20	−0.20
103	4-NH_2-2−CH_3[a,b]	1.31	−0.76	2.07	−1.27	2.58	0.435	0.75	0.70	−0.83
104	2,6-di-Br-4-NO_2[a]	1.35	2.15	−0.80	1.81	−0.46	−1.453	3.57	3.16	1.24
105	2,4-di-Br	1.40	1.26	0.14	1.33	0.07	−0.348	3.22	3.32	0.46
106	2,4,6-tri-Cl	1.41	1.64	−0.23	1.54	−0.13	−0.502	3.69	3.39	0.69
107	4-NO_2[a,b]	1.42	0.88	0.54	0.65	0.77	−1.065	1.91	1.85	0.78
108	3,5-di-Cl	1.57	1.47	0.10	1.52	0.05	−0.285	3.62	3.30	0.74
109	2-Cl-4-NO_2	1.59	1.26	0.33	1.13	0.46	−1.264	2.33	2.34	1.01
110	3,5-di-*t*-C_4H_9	1.64	1.94	−0.30	1.96	−0.32	0.471	5.13	5.13	−0.20
111	4-Cl-6-NO_2	1.64	1.69	−0.05	1.33	0.31	−1.346	2.93	2.67	1.01
112	penta-F[b]	1.64	1.84	−0.20	0.91	0.73	−1.296	3.23	2.17	0.86
113	4-OC_6H_{13}	1.64	1.46	0.18	1.30	0.34	0.338	4.22	4.22	−0.32
114	2,6-di-I-4-NO_2	1.71	2.10	−0.39	1.97	−0.26	−1.421	3.52	3.54	1.14
115	4,6-di-NO_2-2−CH_3	1.72	1.48	0.24	1.37	0.35	−1.825	2.13	2.27	1.39
116	3-CH_3-4−NO_2[a,b]	1.73	1.19	0.54	0.86	0.87	−1.006	2.47	2.27	0.71
117	2,4-di-Cl-6-NO_2	1.75	1.92	−0.17	1.77	−0.02	−1.579	3.07	3.09	1.24
118	3,4-di-Cl	1.75	1.26	0.49	1.35	0.41	−0.236	3.33	3.18	0.60
119	4-Br-2,6-di-Cl	1.78	1.55	0.24	1.64	0.15	−0.515	3.52	3.54	0.69
120	2,6-di-*t*-C_4H_9-4−CH_3	1.80	2.23	−0.43	1.87	−0.07	0.514	5.63	5.43	−0.57
121	4-Cl-2-*i*-C_3H_7-5−CH_3	1.85	1.41	0.44	1.54	0.31	0.113	3.92	4.21	0.01
122	2,4,6-tri-Br	2.03	1.86	0.17	1.88	0.15	−0.621	3.92	3.94	0.69
123	4-OC_7H_{15}	2.03	1.79	0.24	1.63	0.40	0.338	4.75	4.75	−0.32
124	penta-Cl[a,b]	2.05	2.83	−0.78	2.92	−0.87	−0.978	5.12	4.71	1.43
125	4-Cl-2-NO_2[a,b]	2.05	1.33	0.72	1.33	0.72	−1.230	2.47	2.67	1.01
126	2,4,5-tri-Cl	2.10	1.70	0.41	1.77	0.32	−0.556	3.72	3.60	0.83
127	4-*t*-C_8H_{17}	2.10	1.96	0.14	2.16	−0.06	0.474	5.16	5.42	−0.17
128	2,3,5,6-tetra-Cl	2.22	1.96	0.27	2.30	−0.08	−0.817	3.88	4.00	1.20
129	2,3,5-tri-Cl[a]	2.37	1.62	0.75	1.88	0.49	−0.578	3.58	3.60	0.97
130	4-C_9H_{19}	2.47	2.64	−0.17	2.68	−0.21	0.429	6.21	6.21	−0.13
131	2-CH_3-3,4,5,6-tetra−Br	2.57	2.67	−0.10	2.95	−0.38	−0.882	4.97	5.20	1.07
132	penta-Br[b]	2.66	2.79	−0.13	3.42	−0.76	−1.193	4.85	5.48	1.47
133	2,3,4,5-tetra-Cl	2.71	2.12	0.59	2.43	0.28	−0.752	4.21	4.21	1.20

[a] Data points not included in deriving eq 2. [b] Data points not included in deriving eq 3.

We see that the results with σ and LUMO are in good agreement. In the jackknifing procedure the program scans all data points and then indicates which point on dropping yields the largest improvement in r^2. This process was repeated until q^2 of essentially 0.9 was achieved. Table 1 lists the starting group of 133 phenols MlogP (values used by Cronin and Schultz), ClogP, σ, and LUMO values. Tables 2 and 3 record the outliers for each run.

The above study suggests that all the phenols are not operating by the same mechanism. Of course some of the

Table 2: Growth Inhibition Concentration (IGC_{50}) of Phenols Containing Amino Group against *Tetrahymena pyriformis*

		log1/C				
s. no.	substituents	obsd	calcd eq 6	Δ	σ	B1
2	3-NH_2	−0.52	−0.56	0.04	−0.16	2.00
16	4-NH_2[a]	−0.08	−0.87	0.79	−0.66	2.00
27	2,4-di-NH_2	0.13	0.20	−0.07	−1.32	2.35
39	2-NH_2-4-t−C_4H_9	0.37	0.48	−0.11	−0.86	2.35
43	5-NH_2-2−OCH_3	0.45	0.75	−0.30	−0.43	2.35
47	2-NH_2-4−NO_2[a]	0.48	1.09	−0.61	0.12	2.35
69	2-NH_2-4−Cl	0.78	0.75	0.03	−0.43	2.35
75	4-NH_2-2−NO_2[a]	0.88	2.57	−1.69	0.12	2.70
76	6-NH_{2-2}-4-di−CH_3[a]	0.89	2.59	−1.70	−1.00	2.87
79	2-NH_2	0.94	0.61	0.33	−0.66	2.35
92	2-NH_2-4-Cl-5−NO_2	1.17	1.19	−0.02	0.28	2.35
103	4-NH_2-2−CH_3	1.31	1.22	0.09	−0.83	2.52

[a] Data points not included in deriving equation.

misfits could be due to experimental error. The standard of $r^2 = 0.9$ and $q^2 = 0.9$ is not often achieved in QSAR studies. This suggests excellent experimental work. The results show that we must be careful about trying to lump together chemicals that might contain common function groups without first checking for uniformity of action.

The jackknifing process can be misleading in that it forces the data to fit the model from which one starts. We have assumed from many examples that σ and Clog P are a reasonable model for the phenols, moreover both LUMO and σ yield the same result, and $q^2 \cong r^2$, so that there is evidence from two points of view that we are talking above a uniform mechanism of action.

We have found over the years that QSAR for phenols may be correlated by σ, σ^-, or σ^+ depending on the test system and the type of phenol.[2−4] With σ^+ we have come to associate radical reactions.[5] With σ^- we expect ionization of the phenols to be significant. Hence we have checked the role of those parameters with our present data to obtain QSAR 4 and 5 via jackknifing.

$$\log 1/C = 0.49(\pm 0.06)\sigma^- + 0.61(\pm 0.04)\mathrm{ClogP} - 1.08(\pm 0.31)$$

$$n = 100,\ r^2 = 0.909,\ s = 0.236,\ q^2 = 0.902 \qquad (4)$$

$$\log 1/C = 0.53(\pm 0.08)\sigma^+ + 0.62(\pm 0.04)\mathrm{ClogP} - 0.98(\pm 0.13)$$

$$n = 101,\ r^2 = 0.909,\ s = 0.235,\ q^2 = 0.902 \qquad (5)$$

The difference between these three parameters is small as might be anticipated from their high degree of collinearity σ Vs σ^+ $r^2 = 0.919$. It would seem that through resonance as modeled by σ^- or σ^+ is not of great importance, however, a better choice of substituents would be necessary to firmly validate this point. Scanning the points dropped to achieve QSAR 2 and 3 we note that those containing an NH_2 or NO_2 function were often poorly fit. Employing the 12 examples containing an amino group and removing four by jackknifing (4-NH_2; 2-NH_2, 4-NO_2; 4-NH_2, 2-NO_2; 6-NH_2, 2,4-di-Me) we obtained the following QSAR.

$$\log 1/C = 0.62(\pm 0.46)\sigma + 4.22(\pm 1.55)\mathrm{B1} - 8.91(\pm 3.52)$$

$$n = 8,\ r^2 = 0.909,\ s = 0.214,\ q^2 = 0.833 \qquad (6)$$

QSAR 6 is not as sharp a correlation as we would like. Note q^2. It is interesting that now we find that σ (eq 6) and the Verloop sterimol parameter for ortho substituents yield a better result than logP and σ. B1 indicates that there is a positive steric effect of ortho substituents. This parameter is not often used by the QSAR community, but we have

Table 3: Growth Inhibition Concentration (IGC_{50}) of Phenols Containing Nitro Group against *Tetrahymena Pyriformis*

		log1/C								
s. no.	substituents	obsd	calcd eq 9	Δ	calcd eq 10	Δ	LUMO	MlogP	ClogP	σ^+
13	2,4,6-tri-NO_2	−0.16	0.07	−0.23	0.15	2.34	−2.534	0.89	1.64	2.37
33	3,4-di-NO_2[a]	0.27	0.87	−0.56	−0.42	1.49	−1.850	1.79	1.82	1.50
44	2,3-di-NO_2	0.46	0.89	−0.43	−0.23	1.49	−1.799	1.79	1.82	1.50
45	2,6-di-F[a,b]	0.47	1.77	−1.30	−1.66	0.12	−0.321	2.05	1.76	0.14
48	3-NO_2[a,b]	0.51	1.32	−0.81	−1.03	0.71	−1.159	2.00	1.85	0.71
50	2,6-di-NO_2	0.54	0.60	−0.06	−0.08	1.56	−1.952	1.37	1.82	1.58
52	4-CH_3,2−NO_2[a,b]	0.57	1.61	−1.04	−1.32	0.61	−0.967	2.37	2.35	0.48
56	2,6-di-Cl,4-NO_2[a,b]	0.63	1.65	−1.02	−0.79	1.24	−1.441	2.94	2.76	1.01
59	2-NO_2[a,b]	0.67	1.29	−0.62	−0.80	0.78	−1.014	1.79	1.85	0.79
75	4-NH_2,2−NO_2[b]	0.88	0.82	0.06	−0.78	0.12	−1.119	0.96	0.81	−0.51
80	2,5-di-NO_2	0.95	0.64	0.31	0.26	1.49	−2.261	1.75	1.82	1.50
84	2,4-di-NO_2[b]	1.08	0.79	0.29	0.47	1.56	−1.888	1.67	1.82	1.58
86	5-F,2-NO_2	1.13	1.21	−0.08	−0.09	1.12	−1.297	1.91	2.10	1.13
92	2-NH_2-4-Cl-5−NO_2[b]	1.17	1.32	−0.15	−0.84	0.28	−0.961	1.80	1.87	−0.48
95	2,6-di-NO_2-4−CH_3	1.23	1.10	0.14	0.19	1.39	−1.894	2.29	2.32	1.27
104	2,6-di-Br-4-NO_2[a]	1.35	1.96	−0.61	−0.27	1.24	−1.453	3.57	3.16	1.09
107	4-NO_2	1.42	1.33	0.10	−0.05	0.78	−1.065	1.91	1.85	0.79
109	2-Cl-4-NO_2	1.59	1.44	0.16	0.13	1.01	−1.264	2.33	2.34	0.90
111	4-Cl-6-NO_2	1.64	1.70	−0.06	0.02	1.01	−1.346	2.93	2.67	0.90
114	2,6-di-I-4-NO_2	1.71	1.95	−0.24	−0.20	1.14	−1.421	3.52	3.54	1.07
115	4,6-di-NO_2-2−CH_3[a,b]	1.72	1.05	0.67	0.70	1.39	−1.825	2.13	2.27	1.27
116	3-CH_3-4−NO_2	1.73	1.64	0.09	−0.02	0.71	−1.006	2.47	2.27	0.72
117	2,4-di-Cl-6-NO_2	1.75	1.65	0.10	0.17	1.24	−1.579	3.07	3.09	1.01
125	4-Cl-2-NO_2[a]	2.05	1.52	0.53	0.43	1.01	−1.230	2.47	2.67	0.90

[a] Data points not included in deriving eq 9. [b] Data points not included in deriving eq 10.

912 QSAR where B1 works better than the classical Taft parameter Es.

Considering the LUMO and MlogP model we obtain QSAR 7 and 8

$$\log 1/C = -0.14(\pm 0.27)\mathrm{LUMO} + 0.13(\pm 0.64)\mathrm{B1} + 2.85(\pm 1.54)$$

$$n = 8,\ r^2 = 0.891,\ s = 0.186,\ q^2 = 0.755 \qquad (7)$$

$$\mathrm{LOG}1/C = 0.31(\pm 0.32)\mathrm{MlogP} + 2.508(\pm 1.455)\mathrm{B1} - 5.391(\pm 3.275)$$

$$n = 8,\ r^2 = 0.882,\ s = 0.186,\ q^2 = 0.652 \qquad (8)$$

Clearly B1 is a parameter of major importance as we saw in QSAR 7 and 8. Observe the confidence intervals for MlogP in QSAR 8 and for B1 in QSAR 7.

Considering now the 24 nitrophenols in the Cronin-Schultz model we obtain QSAR 9.

$$\log 1/C = 0.51(\pm 0.29)\mathrm{LUMO} + 0.50(\pm 0.19)\mathrm{MlogP} + 0.91(\pm 0.69)$$

$$n = 15,\ r^2 = 0.860,\ s = 0.222,\ q^2 = 0.773 \qquad (9)$$

It was necessary to omit *nine* data points to obtain a modest correlation. The sign of the LUMO term is positive while that of QSAR 1 and 2 is negative. This clearly points to a different mechanism of action.

For comparison with σ model we have formulated QSAR 10.

$$\log 1/C = -1.01(\pm 0.36)\sigma^+ + 0.36(\pm 0.26)\mathrm{ClogP} + 1.67(\pm 0.92)$$

$$n = 15,\ r^2 = 0.897,\ s = 0.222,\ q^2 = 0.830 \qquad (10)$$

The quality of eqs 9 and 10 is essentially the same. The σ^+ term suggests radical toxicity.

DISCUSSION

Two things are clear from the above analysis. LUMO and σ can be used interchangeably as we have noted before.[4] Second, it is a mistake to lump together chemicals for correlation analysis before one has some evidence for their compatability. With some experience in mechanistic organic chemistry and biochemical interactions these problems can be mitigated in the design of the data set. For example the NO_2 function easily undergoes radical reduction in a variety of systems. Hence if one were looking for a strong electron attracting substituent the CN group would be a better choice.

Now doing a comparative study of $X-C_6H_4-NH_2$ on *Tetrahymena pyriformis* using results from Schultz's[1] laboratory, we obtain QSAR 11.

$$\log 1/C = 0.63(\pm 0.07)\mathrm{Clog\ P} - 0.38(\pm 0.25)\mathrm{B1}_2 - 0.46(\pm 0.33)$$

$$n = 54,\ r^2 = 0.859,\ s = 0.291,\ q^2 = 0.845$$
outliers: 4-F, 4-Cl, 4-NO_2, 4-NH_2 (11)

Of the outliers, one does not know how to classify the 4-nitroaniline and 4-aminoaniline is extremely easily oxidized. We have no explanation for the two halogens. All attempts to force an electronic term into QSAR 11 failed. The steric term B1 is weak but definitely significant. Equation 11 is not at all like QSAR 3. This is solid support for our omitting the anilines from QSAR 2 and 3.

Turning now to a study by Cronin and Schultz[1] on the action of $X-C_6H_4NO_2$ on *Tetrahymena pyriformis* we can formulate QSAR 12 by omitting compounds having two or more nitro groups.

$$\log 1/C = 0.60(\pm 0.18)\mathrm{Clog\ P} + 0.43(\pm 0.25)\sigma^- - 1.05(\pm 0.48)$$

$$n = 27,\ r^2 = 0.901,\ s = 0.168,\ q^2 = 0.863 \qquad (12)$$

A hallmark of QSAR of nitro compounds is that σ^- is so often present. The following are a few examples out of many that illustrate the point.

Reduction of $X-C_6H_4NO_2$ by $CH_3{}^{\bullet}CHOH$ in N_2O saturated solution[2]

$$\log k = 0.85(\pm 0.15)\sigma^- + 8.26(\pm 0.11)$$

$$n = 13,\ r^2 = 0.932,\ s = 0.125,\ q^2 = 0.911 \qquad (13)$$

Reduction of $X-C_6H_4NO_2$ by milk xanthine oxidase under anaerobic conditions[3]

$$\log k = 0.98(\pm 0.16)\sigma^- - 0.35(\pm 0.23)\mathrm{B5}_2 + 2.13(\pm 0.27)$$

$$n = 26,\ r^2 = 0.884,\ s = 0.201,\ q^2 = 0.865$$
outliers: 4-SO_3^-, 2-CHO, 4-SO_2NH_2 (14)

I_{50} (lethal concentration) of $X-C_6H_4NO_2$ to fathead minnows[6]

$$\log 1/C = 0.83(\pm 0.14)\sigma^- + 0.21(\pm 0.20)\mathrm{B5}_4 + 0.38(\pm 0.10)$$

$$n = 30,\ r^2 = 0.984,\ s = 0.226,\ q^2 = 0.850$$
outliers: 2-Me, 5-NO_2; 2-NH_2, 4-Me, 5-NO; 2-OH, 3-Me, 5-NO_2 (15)

Equations 11 and 12 support our removal of amino and nitro compounds in the formulation of QSAR for phenols. One needs to give serious thought to the selection of substituents in the design of a congeneric set of molecules. They must be selected so that the major factors, electronic, steric, and hydrophobic, have a reasonable range or else one cannot assess their importance or lack of importance. For example, in selecting a strong electron withdrawing substituent one would want to avoid NO_2 because of its easy reduction to a toxic radical. The choice of CN would be much better. In fact, the aromatic OH can present problems because of its easy oxidation to a radical form.[2] One needs considerable experience with mechanistic organic chemistry as well as how biological systems respond to chemicals to obtain a map of even a small region in chemical−biological space.

Finally, we are interested in understanding the relationship between Hammett parameters and quantum chemical parameters such as E_{LUMO} or HF/6-316 (eq 16). Equations 2 and 3 are valuable in this respect. Another good example that comes to our attention is the comparison of σ and HF/6-316. The latter parameter was calculated for a set of benzoic acids from which QSAR 16 was derived.[7]

$$\sigma = -1.83(\pm 0.16)\mathrm{HF/6{-}316} + 11.33(\pm 0.98)$$

$$n = 32,\ r^2 = 0.946,\ s = 0.114,\ q^2 = 0.937$$

outliers: 3-COOMe, 4-COOMe (16)

Equations 2, 3, and 16 provide confidence that when it is possible to use them, Hammett parameters are just as satisfactory as MO parameters and far less time-consuming to use. The problem with outliers plagues all aspects of QSAR our program enables us to isolate QSAR that have one or more datapoints omitted. For example searching our set of 8600 QSAR from physical organic chemistry for instances that have one or more datapoints omitted finds 3406; 1220 have two or more outliers. These results are based on far more accurate data than biological QSAR. Hence, one cannot give up when finding outliers.

SUMMARY

The work of Schultz carefully testing a large set of miscellaneous phenols has provided an excellent set of data to illustrate the importance of trying to establish the uniformity of the dependent variable in the formulation of QSAR. Unfortunately such opportunities are rare. However, there is no perfect solution to the problem of understanding chemical biological interactions. It must be remembered that jackknifing yields a QSAR that conforms to the initial model selected. In the present case a two-parameter equation is based on electronic and hydrophobic parameters. Thanks to the MO calculations we can use two somewhat different perspectives to arrive at essentially the same conclusion.

ACKNOWLEDGMENT

This research was supported by NIEHS Grant RO1 ES 07595.

REFERENCES AND NOTES

(1) Cronin, M. T.; Schultz, T. W. Development of Quantitative Structure−Activity Relationships for the Toxicity of Aromatic Compounds to Tetrahymena pyriformis: Comparative Assessment of the Methodologies. *Chem. Res. Toxicol*. **2001**, *14*, 1284−1295.

(2) Selassie, C. D.; Shusterman, A. J.; Kapur, S.; Verma, R. P.; Zhang, L.; Hansch, C. On the Toxicology of Phenols to fast growing cells. A QSAR model for a radical-based toxicity. *J. Chem. Soc., Perkin Trans. 2* **1999**, 2729−2733.

(3) Garg, R.; Kapur, S.; Hansch, C. Radical Toxicity of Phenols; A Reference point for obtaining perspective in the formulation of QSAR. *Med. Res. Rev.* **2001**, *21*, 7382.

(4) Garg, R., Kapur, A.; Hansch, C. Comparative QSAR: On the toxicology of the phenolic OH moiety. *Crit. Rev. Toxicol*. **2000**, *31*, 223−245.

(5) Hansch, C.; Gao, H. Comparative QSAR: Radical reactions of benzene derivatives in Chemistry and Biology. *Chem. Rev.* **1997**, *97*, 2995−3059.

(6) Hansch, C.; Gao, H.; Hoekman, D. A *Generalised Approach to Comparative QSAR in Comparative QSAR*; Devillers, J., Ed.; Taylor and Francis: 1998.

(7) Kurita, Y.; Takayama, C. An Analytical Method to Determine Reference Orbitals for Localized Frontier Orbitals (LFOs) and its Application to the Correlation Analysis Between LFO Energies and Acidities of Conjugate Cations of Heterocyclic Bases and Monosubstituted Benzoic Acids. *J. Am. Chem. Soc.* **1997**, *101*, 5593−5595.

CI010250E