

Supplementary Material

Table S1

Observed pK_a values for the carboxylic acids.

N	CAS no.	Chemical Name	Exp. pKa
meta- and para-substituted benzoic acids			
1	000051-44-5	3,4-dichlorobenzoic acid	3.64
2	000051-36-5	3,5-dichlorobenzoic acid	3.54
3	000057-66-9	4-[(dipropylamino)sulfonyl]benzoic acid/probenecid	3.40
4	000062-23-7	4-nitrobenzoic acid	3.44
5	000065-85-0	benzoic acid	4.19
6	000074-11-3	4-chlorobenzoic acid	3.98
7	000093-09-4	2-naphthalenecarboxylic acid	4.17
8	000093-07-2	3,4-dimethoxybenzoic acid	4.36
9	000098-73-7	4-(1,1-dimethylethyl)benzoic acid	4.40
10	000099-96-7	4-hydroxybenzoic acid	4.54
11	000099-94-5	4-methylbenzoic acid	4.37
12	000099-50-3	3,4-dihydroxybenzoic acid	4.48
13	000099-34-3	3,5-dinitrobenzoic acid	2.82
14	000099-10-5	3,5-dihydroxybenzoic acid	4.04
15	000099-05-8	3-aminobenzoic acid	4.74
16	000099-04-7	3-methylbenzoic acid	4.27
17	000100-09-4	4-(methoxy)benzoic acid	4.47
18	000121-92-6	3-nitrobenzoic acid	3.46
19	000121-34-6	4-hydroxy-3-(methoxy)benzoic acid	4.51
20	000149-91-7	3,4,5-trihydroxybenzoic acid	4.21
21	000150-13-0	4-aminobenzoic acid	4.85
22	000455-38-9	3-fluorobenzoic acid	3.86
23	000456-22-4	4-fluorobenzoic acid	4.14

24	000528-45-0	3,4-dinitrobenzoic acid	2.82
25	000530-57-4	4-hydroxy-3,5-bis(methoxy)benzoic acid	4.34
26	000535-80-8	3-chlorobenzoic acid	2.81
27	000536-66-3	4-(1-methylethyl)benzoic acid	4.35
28	000585-76-2	3-bromobenzoic acid	3.81
29	000586-89-0	4-acetylbenzoic acid	3.70
30	000586-76-5	4-bromobenzoic acid	4.00
31	000586-38-9	3-(methoxy)benzoic acid	4.09
32	000619-86-3	4-(ethoxy)benzoic acid	4.45
33	000619-66-9	4-formylbenzoic acid	3.77
34	000619-65-8	4-cyanobenzoic acid	3.55
35	000619-64-7	4-ethylbenzoic acid	4.35
36	000619-21-6	3-formylbenzoic acid	3.84
37	000619-05-6	3,4-diaminobenzoic acid	3.49
38	001132-21-4	3,5-dimethoxybenzoic acid	3.97
39	001877-72-1	3-cyanobenzoic acid	3.60
40	002215-77-2	4-(phenyloxy)benzoic acid	4.52
41	003739-38-6	3-(phenyloxy)benzoic acid	3.92
42	004052-30-6	4-(methylsulfonyl)benzoic acid	3.64
43	005438-19-7	4-(propyloxy)benzoic acid	4.46
44	007496-53-9	4-(glycylamino)benzoic acid	4.20

ortho-substituted benzoic acids

45	000050-85-1	2-hydroxy-4-methylbenzoic acid	3.40
46	000050-84-0	2,4-dichlorobenzoic acid	2.68
47	000050-79-3	2,5-dichlorobenzoic acid	2.47
48	000050-78-2	2-(acetyloxy)benzoic acid	3.49
49	000050-31-7	2,3,6-trichlorobenzoic acid	1.50
50	000050-30-6	2,6-dichlorobenzoic acid	1.59
51	000059-07-4	4-amino-2-(ethyloxy)benzoic acid	5.09
52	000061-68-7	2-[(2,3-dimethylphenyl)amino]benzoic acid	4.20
53	000065-49-6	4-amino-2-hydroxybenzoic acid	3.66
54	000069-72-7	2-hydroxybenzoic acid	2.97

55	000083-40-9	2-hydroxy-3-methylbenzoic acid	2.95
56	000088-65-3	2-bromobenzoic acid	2.88
57	000089-86-1	2,4-dihydroxybenzoic acid	3.11
58	000089-56-5	2-hydroxy-5-methylbenzoic acid	3.15
59	000089-55-4	5-bromo-2-hydroxybenzoic acid	2.66
60	000089-52-1	2-(acetylamino)benzoic acid	3.40
61	000091-52-1	2,4-dimethoxybenzoic acid	4.36
62	000091-40-7	2-(phenylamino)benzoic acid	3.99
63	000092-70-6	3-hydroxy-2-naphthalenecarboxylic acid	2.79
64	000096-97-9	2-hydroxy-5-nitrobenzoic acid	2.12
65	000099-60-5	2-chloro-4-nitrobenzoic acid	2.14
66	000118-92-3	2-aminobenzoic acid	4.95
67	000118-91-2	2-chlorobenzoic acid	2.89
68	000119-90-1	2-methylbenzoic acid	3.98
69	000129-66-8	2,4,6-trinitrobenzoic acid	0.65
70	000133-90-4	3-amino-2,5-dichlorobenzoic acid	3.40
71	000303-38-8	2,3-dihydroxybenzoic acid	2.91
72	000303-07-1	2,6-dihydroxybenzoic acid	1.05
73	000321-14-2	5-chloro-2-hydroxybenzoic acid	2.65
74	000445-29-4	2-fluorobenzoic acid	3.27
75	000490-79-9	2,5-dihydroxybenzoic acid	2.95
76	000552-16-9	2-nitrobenzoic acid	2.17
77	000577-56-0	2-acetylbenzoic acid	4.13
78	000579-75-9	2-(methoxy)benzoic acid	3.90
79	000609-99-4	2-hydroxy-3,5-dinitrobenzoic acid	0.70
80	000610-30-0	2,4-dinitrobenzoic acid	1.42
81	000632-46-2	2,6-dimethylbenzoic acid	3.35
82	000652-32-4	2,3,5,6-tetrafluoro-4-methylbenzoic acid	2.00
83	000947-84-2	2-biphenylcarboxylic acid	3.46
84	001466-76-8	2,6-bis(methoxy)benzoic acid	3.44
85	001521-38-6	2,3-dimethoxybenzoic acid	3.98
86	001918-00-9	3,6-dichloro-2-(methoxy)benzoic acid	1.97

87	002243-42-7	2-(phenyloxy)benzoic acid	3.53
88	002438-04-2	2-(1-methylethyl)benzoic acid	3.63
89	002516-96-3	2-chloro-5-nitrobenzoic acid	2.17
90	003970-35-2	2-chloro-3-nitrobenzoic acid	2.02
91	004727-29-1	2-[(phenylamino)carbonyl]benzoic acid	2.50
92	005344-49-0	2-chloro-6-nitrobenzoic acid	1.34
93	021327-86-6	2-chloro-6-methylbenzoic acid	2.75
94	025784-02-5	2-[(2-amino-2-oxoethyl)amino]benzoic acid	4.20

aliphatic carboxylic acids

95	000050-21-5	2-hydroxypropanoic acid	3.86
96	000053-86-1	[1-[(4-chlorophenyl)carbonyl]-2-methyl-5-(methyloxy)-1H-indol-3-yl]acetic acid/indomethacin	4.50
97	000061-78-9	N-[(4-aminophenyl)carbonyl]glycine	3.80
98	000061-33-6	3,3-dimethyl-7-oxo-6-[(phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid/benzylpenicillin	2.74
99	000061-32-5	6-([[2,6-bis(methyloxy)phenyl]carbonyl]amino)-3,3-dimethyl-7-oxo-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid/methicillin	2.77
100	000064-19-7	acetic acid	4.76
101	000068-11-1	mercaptoacetic acid	3.55
102	000075-99-0	2,2-dichloropropanoic acid	1.79
103	000075-98-9	2,2-dimethylpropanoic acid	5.03
104	000076-93-7	hydroxy(diphenyl)acetic acid	3.05
105	000076-05-1	trifluoroacetic acid	0.52
106	000076-03-9	trichloroacetic acid	0.51
107	000077-06-5	gibberellic acid	4.00
108	000079-43-6	dichloroacetic acid	1.26
109	000079-31-2	2-methylpropanoic acid	4.84
110	000079-14-1	hydroxyacetic acid	3.83
111	000079-11-8	chloroacetic acid	2.87
112	000079-09-4	propanoic acid	4.88
113	000079-08-3	bromoacetic acid	2.89
114	000081-25-4	3,7,12-trihydroxycholan-24-oic acid	4.98
115	000085-34-7	(2,3,6-trichlorophenyl)acetic acid	3.70
116	000086-87-3	1-naphthalenylacetic acid	4.23
117	000087-51-4	1H-indol-3-ylacetic acid	4.75

118	000087-08-1	3,3-dimethyl-7-oxo-6-[(phenyloxy)acetyl]amino-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid/phenoxymethylpenicillin	2.79
119	000088-09-5	2-ethylbutanoic acid	4.71
120	000090-64-2	hydroxy(phenyl)acetic acid	3.41
121	000093-76-5	[(2,4,5-trichlorophenyl)oxy]acetic acid	2.83
122	000093-72-1	2-[(2,4,5-trichlorophenyl)oxy]propanoic acid	2.84
123	000094-82-6	4-[(2,4-dichlorophenyl)oxy]butanoic acid	4.95
124	000094-81-5	4-[(4-chloro-2-methylphenyl)oxy]butanoic acid	6.20
125	000094-75-7	[(2,4-dichlorophenyl)oxy]acetic acid	2.73
126	000094-74-6	[(4-chloro-2-methylphenyl)oxy]acetic acid	3.13
127	000097-61-0	2-methylpentanoic acid	4.79
128	000098-89-5	cyclohexanecarboxylic acid	4.90
129	000099-66-1	2-propylpentanoic acid	4.60
130	000102-32-9	(3,4-dihydroxyphenyl)acetic acid	4.25
131	000103-82-2	phenylacetic acid	4.31
132	000104-03-0	(4-nitrophenyl)acetic acid	3.85
133	000104-01-8	[4-(methyloxy)phenyl]acetic acid	4.36
134	000107-94-8	3-chloropropanoic acid	3.99
135	000107-92-6	butanoic acid	4.82
136	000116-53-0	2-methylbutanoic acid	4.81
137	000117-34-0	diphenylacetic acid	3.94
138	000120-36-5	2-[(2,4-dichlorophenyl)oxy]propanoic acid	3.10
139	000122-88-3	[(4-chlorophenyl)oxy]acetic acid	3.10
140	000122-59-8	(phenyloxy)acetic acid	3.17
141	000123-76-2	4-oxopentanoic acid	4.64
142	000141-82-2	propanedioic acid	2.85
143	000144-49-0	fluoroacetic acid	2.59
144	000300-85-6	3-hydroxybutanoic acid	4.41
145	000305-03-3	4-[4-bis(2-chloroethyl)amino]phenyl]butanoic acid/chlorambucil	5.75
146	000306-08-1	[4-hydroxy-3-(methyloxy)phenyl]acetic acid	4.41
147	000327-97-9	3-[(2E)-3-(3,4-dihydroxyphenyl)-2-propenoyl]oxy]-1,4,5-trihydroxycyclohexanecarboxylic acid/chlorogenic acid	2.66
148	000331-25-9	(3-fluorophenyl)acetic acid	4.13
149	000348-10-7	[(2-fluorophenyl)oxy]acetic acid	3.08

150	000372-09-8	cyanoacetic acid	2.45
151	000404-98-8	[(3-fluorophenyl)oxy]acetic acid	3.13
152	000405-79-8	[(4-fluorophenyl)oxy]acetic acid	3.13
153	000405-50-5	(4-fluorophenyl)acetic acid	4.24
154	000462-60-2	N-(aminocarbonyl)glycine	3.89
155	000467-69-6	9-hydroxy-9H-fluorene-9-carboxylic acid/flurenol	1.09
156	000473-81-4	2,3-dihydroxypropanoic acid	3.55
157	000501-52-0	3-phenylpropanoic acid	4.66
158	000503-74-2	3-methylbutanoic acid	4.77
159	000503-66-2	3-hydroxypropanoic acid	4.51
160	000515-30-0	2-hydroxy-2-phenylpropanoic acid	3.53
161	000516-05-2	methylpropanedioic acid	3.12
162	000539-35-5	6-(4-oxo-1,3-thiazolidin-2-yl)hexanoic acid/mycobacidin	5.10
163	000581-96-4	2-naphthalenylacetic acid	4.25
164	000588-32-9	[(3-chlorophenyl)oxy]acetic acid	3.07
165	000588-22-7	[(3,4-dichlorophenyl)oxy]acetic acid	2.92
166	000594-61-6	2-hydroxy-2-methylpropanoic acid	3.61
167	000595-46-0	dimethylpropanedioic acid	3.15
168	000595-37-9	2,2-dimethylbutanoic acid	5.03
169	000598-78-7	2-chloropropanoic acid	2.80
170	000598-72-1	2-bromopropanoic acid	2.97
171	000601-75-2	ethylpropanedioic acid	2.96
172	000614-61-9	[(2-chlorophenyl)oxy]acetic acid	3.05
173	000616-62-6	propylpropanedioic acid	2.99
174	000617-31-2	2-hydroxypentanoic acid	2.89
175	000622-47-9	(4-methylphenyl)acetic acid	4.37
176	000646-07-1	4-methylpentanoic acid	4.84
177	000689-13-4	N-formyl-N-hydroxyglycine	3.50
178	000940-64-7	[(4-methylphenyl)oxy]acetic acid	3.21
179	001643-15-8	[(3-methylphenyl)oxy]acetic acid	3.20
180	001759-53-1	cyclopropanecarboxylic acid	4.83
181	001798-99-8	[(3-bromophenyl)oxy]acetic acid	3.09

182	001798-11-4	[(4-nitrophenyl)oxy]acetic acid	2.89
183	001821-12-1	4-phenylbutanoic acid	4.76
184	001877-75-4	{[4-(methoxy)phenyl]oxy}acetic acid	2.31
185	001877-73-2	(3-nitrophenyl)acetic acid	3.97
186	001878-91-7	[(4-bromophenyl)oxy]acetic acid	3.13
187	001878-88-2	[(3-nitrophenyl)oxy]acetic acid	2.95
188	001878-87-1	[(2-nitrophenyl)oxy]acetic acid	2.90
189	001878-85-9	{[2-(methoxy)phenyl]oxy}acetic acid	3.23
190	001878-82-6	[(4-cyanophenyl)oxy]acetic acid	2.93
191	001878-68-8	(4-bromophenyl)acetic acid	4.19
192	001878-66-6	(4-chlorophenyl)acetic acid	4.19
193	001878-65-5	(3-chlorophenyl)acetic acid	4.14
194	001878-49-5	[(2-methylphenyl)oxy]acetic acid	3.23
195	001879-58-9	[(3-cyanophenyl)oxy]acetic acid	3.03
196	001879-56-7	[(2-bromophenyl)oxy]acetic acid	3.13
197	002088-24-6	{[3-(methoxy)phenyl]oxy}acetic acid	3.14
198	002270-20-4	5-phenylpentanoic acid	4.88
199	002976-75-2	(1-naphthalenyloxy)acetic acid	3.20
200	003813-05-6	(4-chloro-2-oxo-1,3-benzothiazol-3(2H)-yl)acetic acid	3.04
201	005292-21-7	cyclohexylacetic acid	4.80
202	006324-11-4	[(2-hydroxyphenyl)oxy]acetic acid	3.02
203	010502-44-0	hydroxy[4-(methoxy)phenyl]acetic acid	3.42
204	014387-10-1	(4-ethylphenyl)acetic acid	4.37
205	015307-86-5	{2-[(2,6-dichlorophenyl)amino]phenyl}acetic acid/ diclofenac	4.15
206	015687-27-1	2-[4-(2-methylpropyl)phenyl]propanoic acid/ibuprofen	4.45
207	016484-77-8	2-[(4-chloro-2-methylphenyl)oxy]propanoic acid	3.68
208	016563-41-0	3-(1-naphthalenyloxy)propanoic acid	4.00
209	018046-21-4	[4-(4-chlorophenyl)-2-phenyl-1,3-thiazol-5-yl]acetic acid/fentiazac	3.60
210	020225-24-5	2-ethylpentanoic acid	4.71
211	022071-15-4	2-[3-(phenylcarbonyl)phenyl]propanoic acid/ketoprofen	4.45
212	022131-79-9	[3-chloro-4-(2-propen-1-yloxy)phenyl]acetic acid/alcofenac	4.29
213	022204-53-1	2-[6-(methoxy)-2-naphthalenyl]propanoic acid/naprosyn	4.15

214	029679-58-1	2-[3-(phenyloxy)phenyl]propanoic acid/fenoprofen	4.50
215	032857-63-9	[4-(1,1-dimethylethyl)phenyl]acetic acid	4.42
216	036330-85-5	4-(4-biphenyl)-4-oxobutanoic acid/fenbufen	4.51
217	038194-50-2	((1E)-5-fluoro-2-methyl-1-[[4-(methylsulfinyl)phenyl]methylidene]-1H-inden-3-yl)acetic acid/sulindac	4.70
218	040828-46-4	2-[4-(2-thienylcarbonyl)phenyl]propanoic acid/suprofen	3.91
219	040843-25-2	2-([4-[(2,4-dichlorophenyl)oxy]phenyl]oxy)propanoic acid	3.43
220	053808-88-1	[3-(4-chlorophenyl)-1-phenyl-1H-pyrazol-4-yl]acetic acid/ionazolac	4.30
221	055335-06-3	[(3,5,6-trichloro-2-pyridinyl)oxy]acetic acid	2.68
222	055863-26-8	(11-oxo-6,11-dihydrodibenzo[b,e]thiepin-2-yl)acetic acid/tiopinac	3.71
223	058667-63-3	N-(3-chloro-4-fluorophenyl)-N-(phenylcarbonyl)alanine	3.72
224	069335-91-7	2-[4-[[5-(trifluoromethyl)-2-pyridinyl]oxy]phenyl]oxy]propanoic acid	3.12
225	069806-34-4	2-[4-[[3-chloro-5-(trifluoromethyl)-2-pyridinyl]oxy]phenyl]oxy]propanoic acid	2.90
226	074103-06-3	5-(phenylcarbonyl)-2,3-dihydro-1H-pyrrolo[1,2-a]pyrrole-1-carboxylic acid/ketorolac	3.49
227	089894-13-3	[(4-chloro-3-nitrophenyl)oxy]acetic acid	2.96
228	104273-73-6	4-(cyclopropylcarbonyl)-3,5-dioxocyclohexanecarboxylic acid	5.32

Table S2

Summary of the results obtained from the three methods at level C and E. The values in bold are the lowest RMSE and the values in italics are the next lowest RMSE.

Level C	r-fold	Set	PLS		SVM		RBFNN	
			q ²	RMSE	q ²	RMSE	q ²	RMSE
Systematic Sampling	4 Fold	All	0.782	0.422	0.897	0.288	<i>0.878</i>	<i>0.316</i>
		Ortho	<i>0.812</i>	<i>0.408</i>	0.797	0.398	0.746	0.472
		Para/Meta	<i>0.917</i>	<i>0.121</i>	0.915	0.117	0.847	0.147
		Aliphatic	<i>0.893</i>	<i>0.288</i>	0.899	0.28	0.874	0.317
	7 Fold	All	0.778	0.424	<i>0.895</i>	<i>0.291</i>	0.927	0.256
		Ortho	0.844	0.361	<i>0.819</i>	<i>0.391</i>	0.774	0.437
		Para/Meta	0.934	0.104	<i>0.915</i>	<i>0.126</i>	0.900	0.129
		Aliphatic	0.920	0.265	0.904	0.27	<i>0.910</i>	<i>0.267</i>
	10 Fold	All	0.779	0.420	0.888	0.295	<i>0.884</i>	<i>0.308</i>
		Ortho	<i>0.845</i>	<i>0.355</i>	0.848	0.352	0.832	0.372
		Para/Meta	<i>0.939</i>	<i>0.101</i>	0.942	0.096	0.942	0.107
		Aliphatic	0.905	0.335	0.907	0.262	<i>0.889</i>	<i>0.298</i>
Random Sampling	4 Fold	All	0.775	0.440	0.891	0.295	<i>0.860</i>	<i>0.338</i>
		Ortho	<i>0.836</i>	<i>0.379</i>	0.836	0.378	0.807	0.407
		Para/Meta	0.927	0.120	<i>0.910</i>	<i>0.129</i>	0.235	0.325
		Aliphatic	0.907	0.267	<i>0.900</i>	<i>0.282</i>	0.883	0.305
	7 Fold	All	0.778	0.426	<i>0.860</i>	<i>0.290</i>	0.912	0.282
		Ortho	0.801	0.382	<i>0.807</i>	<i>0.395</i>	0.753	0.460
		Para/Meta	<i>0.928</i>	<i>0.112</i>	0.925	0.105	0.789	0.177
		Aliphatic	<i>0.903</i>	<i>0.273</i>	0.915	0.251	0.881	0.308
	10 Fold	All	0.778	0.423	0.885	0.300	<i>0.855</i>	<i>0.345</i>
		Ortho	<i>0.827</i>	<i>0.369</i>	0.843	0.352	0.761	0.412
		Para/Meta	0.929	0.097	0.930	0.097	0.878	0.135
		Aliphatic	<i>0.903</i>	<i>0.269</i>	0.911	0.261	0.891	0.295
Level E								
Systematic Sampling	4 Fold	All	0.814	0.391	0.873	0.269	<i>0.849</i>	<i>0.352</i>
		Ortho	<i>0.806</i>	<i>0.417</i>	0.842	0.372	0.690	0.520
		Para/Meta	0.916	0.131	0.916	0.131	0.920	0.126
		Aliphatic	0.875	0.313	0.900	0.275	0.902	0.280
	7 Fold	All	0.821	0.375	<i>0.885</i>	<i>0.301</i>	0.907	0.289
		Ortho	0.836	0.370	<i>0.851</i>	<i>0.354</i>	0.852	0.350
		Para/Meta	<i>0.906</i>	<i>0.133</i>	0.913	0.125	0.869	0.159
		Aliphatic	<i>0.894</i>	<i>0.285</i>	0.909	0.263	0.893	0.293
	10 Fold	All	0.816	0.377	0.881	0.303	<i>0.859</i>	<i>0.339</i>
		Ortho	<i>0.834</i>	<i>0.371</i>	0.861	0.342	0.749	0.423
		Para/Meta	<i>0.917</i>	<i>0.122</i>	0.919	0.119	0.892	0.146

Random Sampling	4 Fold	Aliphatic	<i>0.883</i>	<i>0.299</i>	0.906	0.262	0.762	0.435
		All	0.787	0.407	0.888	0.298	<i>0.864</i>	<i>0.334</i>
		Ortho	<i>0.825</i>	<i>0.390</i>	0.846	0.366	0.712	0.487
	7 Fold	Para/Meta	<i>0.870</i>	<i>0.160</i>	0.908	0.137	0.760	0.220
		Aliphatic	<i>0.877</i>	<i>0.307</i>	0.901	0.280	0.876	0.314
		All	0.814	0.387	0.886	0.289	<i>0.897</i>	<i>0.304</i>
	10 Fold	Ortho	0.879	0.398	<i>0.788</i>	<i>0.405</i>	0.641	0.554
		Para/Meta	0.911	0.129	0.913	0.129	0.848	0.167
		Aliphatic	0.884	0.299	0.905	0.269	<i>0.903</i>	<i>0.279</i>
		All	0.814	0.383	0.886	0.299	<i>0.882</i>	<i>0.311</i>
		Ortho	<i>0.879</i>	<i>0.380</i>	0.832	0.368	0.704	0.483
			Para/Meta	<i>0.908</i>	<i>0.119</i>	0.914	0.116	0.853
		Aliphatic	<i>0.883</i>	<i>0.295</i>	0.904	0.269	0.878	0.312