A non-conformational QSAR study for plant-derived larvicides against Zika *Aedes aegypti* L. vector

Laura M. Saavedra [a,*], Gustavo P. Romanelli [b, c], Pablo R. Duchowicz [a,*]

[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA). CONICET. UNLP. Diag. 113 y 64. C.C. 16. Sucursal 4. 1900 La Plata. Argentina.

[b] Centro de Investigación y Desarrollo en Ciencias Aplicadas "Dr. J.J. Ronco" (CINDECA). Departamento de Química. Facultad de Ciencias Exactas. CONICET. UNLP. Calle 47 No. 257. B1900AJK La Plata. Argentina.

[c] Cátedra de Química Orgánica. Centro de Investigación en Sanidad Vegetal (CISaV). Facultad de Ciencias Agrarias y Forestales. Universidad Nacional de La Plata. Calles 60 y 119 s/n. B1904AAN La Plata. Argentina.

[*]*Corresponding authors:* Tel.: +54 221 425 7430. +54 221 425 7291; fax: +54 221 425 4642. E-mail addresses: laurasaavedra@gmail.com (L.M.S.); pabloducho@gmail.com (P.R.D.)

(Total 46 pages including this cover page)

Supplementary Material

Table 1S. The experimental LC_{50} larvicidal activity of plant-derived compounds analysed in this QSAR study. Molecules from 1 to 145 were collected by Geris et al. 2012.

ID	Smiles Code	LC_{50} (µg. mL^{-1}) experimental	Reference
1^	OC(=O)CCCCCCC/C=C\CCCCCCCC	8.8	Rahuman et al. (2008)
2	OC(=O)CCCCCCC/C=C\C/C=C\CCCCC	18.2	Rahuman et al. (2008)
3	C(CCCCCCCC[C@H](O)[C@@H]1O[C@H](CC1)[C@H]1O[C@H]([C@@H](CCCCCC[C@H](CC2=CC(=O)O[C@@H]2C)O)O)CC1)CCCCC	141.1	Ye et al. (1996)
4*	C(CCCCCCCC[C@H](O)[C@@H]1O[C@H](CC1)[C@H]1O[C@H]([C@@H](CCCCCC[C@H](CC2=CC(=O)O[C@@H]2C)O)O)CC1)CCC	30.4	Ye et al. (1996)
5^	C(CCCCCCCC[C@H](O)[C@@H]1O[C@H](CC1)[C@H]1O[C@H]([C@@H](CCCC[C@H](CC2=CC(=O)O[C@@H]2C)O)O)CC1)CCCCC	75.7	Ye et al. (1996)
6^	C(CCCCCCCC[C@H](O)[C@@H]1O[C@H](CC1)[C@H]1O[C@H]([C@@H](CC[C@@H](CCCCCCCC2=CC(=O)O[C@@H]2C)O)O)CC1)CCC	8.3	Ye et al. (1996)
7	C(CCCCCCCC[C@H](O)[C@@H]1O[C@H](CC1)[C@@H](CCCC[C@@H](CCCCC[C@H](O)CC1=C[C@H](OC1=O)C)O)O)CCC	9.5	Ho et al. (2003)
8	c1ccc(s1)c1ccc(s1)c1cccs1	0.02	Wat et al. (1981)
9	C1(=SS=C(C=C1)C#C)C#C/C=C\C=C	0.08	Wat et al. (1981)
10*	c1(ccco1)c1ccc(o1)C#CC=C	3.9	Perich et al. (1995)
11	c1(ccc(o1)C)c1ccc(o1)C#CC=C	3.9	Perich et al. (1995)
12*	c1(ccc(o1)C)c1ccc(o1)c1ccco1	3.9	Perich et al. (1995)
13	C1(=C[C@H]2[C@@H](C(=C1OC)CC=C(C)C)C(=O)c1c(O2)ccc(c1O)CC=C(C)C)O	19.4	Ee et al. (2004)
14	c1c(c2c(cc1)C(=O)C(=CC2=O)C)O	5.4	Sreelatha et al. (2010)
15*	c1(c(c2c(cc1)C(=O)C(=CC2=O)C)O)O	13.6	Sreelatha et al. (2010)
16	c1(c2c(ccc1)[C@@H](O)[C@@H](C)CC2=O)O	1.3	Sreelatha et al. (2010)
17	Oc1c2c(c(c3c4c(c(O)cc3)C(=O)C=C(C4=O)C)cc1)C(=O)C(=CC2=O)C	40.7	Sreelatha et al. (2010)
18	Oc1c(c2ccc3c(C(=O)C=C(C3=O)C)c2O)cc(c2c1C(=	31.2	Sreelatha et

	O)CCC2=O)C		al. (2010)
19	O=C1c2c(C(=O)c3c1cccc3)ccc(c2)C	3.3	Cheng et al. (2008)
20^	c1cc2c(cc1)C(=O)C(=C(C2=O)CC=C(C)C)O	26.3	Rodrigues et al. (2005)
21	c1cc2c(cc1)C(=O)C(=C(C2=O)CC=C(C)C)OC(=O)C	9.6	Ribeiro et al. (2009)
22^	c1cc2c(cc1)C(=O)C(=C(C2=O)CCC(C)C)O	4.7	Ribeiro et al. (2009)
23	c1c(c2c(cc1)C(=O)C=CC2=O)O	3.6	Ribeiro et al. (2009)
24^	c1c(c2c(cc1)C(=O)C=CC2=O)OC(=O)C	4.6	Ribeiro et al. (2009)
25*	c1c(c2c(cc1)C(=O)C=CC2=O)OC	7.9	Ribeiro et al. (2009)
26*	c1c(c2c(cc1)C(=O)C(=CC2=O)Br)O	1.4	Ribeiro et al. (2009)
27	c1c(c2c(cc1)C(=O)C(=CC2=O)Br)OC(=O)C	1.2	Ribeiro et al. (2009)
28	c1c(c2c(cc1)C(=O)C(=CC2=O)Br)OC	9.7	Ribeiro et al. (2009)
29	c1c(c2c(cc1)C(=O)C=C(C2=O)Br)O	0.9	Ribeiro et al. (2009)
30	c1c(c2c(cc1)C(=O)C=C(C2=O)Br)OC(=O)C	7.3	Ribeiro et al. (2009)
31*	c1c(c2c(cc1)C(=O)C=C(C2=O)Br)OC	5.8	Ribeiro et al. (2009)
32	c1ccc2c(C(=O)C=C(C2=O)C)c1	15	Ribeiro et al. (2009)
33	C1=CC(=O)C=CC1=O	90	Sousa et al. (2010)
34*	C1=C(C(=O)C=CC1=O)C	61	Sousa et al. (2010)
35	C1=C(C(=O)C(=CC1=O)C)C	42	Sousa et al. (2010)
36	C1=C(C(=O)C=C(C1=O)C)C	57	Sousa et al. (2010)
37	C1=C(C(=O)C=CC1=O)C(C)C	33	Sousa et al. (2010)
38^	C1=C(C(=O)C=C(C1=O)C)C(C)C	48	Sousa et al. (2010)
39^	c1(cc2c(c(c1)O)C(=O)c1c(C2=O)cc(cc1O)O)C	1.9	Yang et al. (2003)
40^	c1c2c(C(=O)c3c(C2=O)c(cc(c3O)O)O)ccc1	19.6	Yang et al. (2003)
41	c1c2c(C(=O)c3c(C2=O)cc(C(=O)O)cc3)ccc1	16.3	Ee et al. (2009)

42*	c1c2c(C(=O)c3c(C2=O)cc(cc3)CO)ccc1	15.4	Ee et al. (2009)
43^	c1c2c(C(=O)c3c(C2=O)ccc(c3O)C)ccc1	1.8	Ee et al. (2009)
44	c1c2c(C(=O)c3c(C2=O)cc(c(c3OC)C=O)O)ccc1	7.4	Ee et al. (2009)
45*	c1c2c(C(=O)c3c(C2=O)cc(c(c3O)C=O)O)ccc1	15	Ee et al. (2009)
46	c1c2c(C(=O)c3c(C2=O)ccc(c3O)C=O)ccc1	15	Ee et al. (2009)
47	c1(c2c(C(=O)c3c(C2=O)ccc(c3O)C)ccc1OC)O	15	Ee et al. (2009)
48	c12c(cc(c(c1OC)[C@H]1OC[C@@H]3[C@@]1(CO[C@H]3c1c(cc3c(c1OC)OCO3)OC)OC(=O)C)OC)OCO2	2.1	Perumalsamy et al. (2010)
49	O1[C@H]([C@H]2[C@H]([C@H](OC2)c2cc3OCOc3cc2)C1)c1cc2OCOc2cc1	10.5	Perumalsamy et al. (2010)
50^	O1c2c(OC1)cc(cc2OC)CC=C	10.8	Lichtehnstein et al. (1974)
51	O1c2c(OC1)c(c(cc2OC)CC=C)OC	11.5	Lichtehnstein et al. (1974)
52*	O1c2c(OC1)cc(c(c2OC)OC)CC=C	11	Lichtehnstein et al. (1974)
53*	O(c1cc(CC=C)ccc1OC)C	102.5	Scotti et al. (2014)
54	c1c(/C=C/C=O)cccc1	24.4	Santos et al. (2010)
55^	c1(cc(CC=C)ccc1OC)O	73.2	Scotti et al. (2014)
56^	O1c2c(OC1)ccc(c2)CC=C	49	Santos et al. (2010)
57	O(c1cc(ccc1O)C=O)C	514.2	Scotti et al. (2014)
58	c1c(c(ccc1)O)C=O	136	Santos et al. (2010)
59	c1cc(ccc1)O	194	Santos et al. (2010)
60*	c1c(c(ccc1)O)O	243	Santos et al. (2010)
61	c1(cccc(O)c1)O	577	Santos et al. (2010)
62	c1(ccccc1OC)O	177	Santos et al. (2010)
63*	O1c2c(OC1)ccc(c2)C=O	200	Siddiqui et al. (2008)
64^	O=c1oc2c(cc1)ccc(OC)c2	22.5	Oranday et al. (2008)
65^	O=C1O[C@@H](CC=C1)/C=C/c1ccccc1	15	Jewers et al.

			(1972)
66	O1c2c(C(=O)CC1(C)C)cc(cc2)C=C	9	Albuquerque et al. (2004)
67*	C(=O)(c1c(ccc(c1)C=C)O)C=C(C)C	50	Albuquerque et al. (2004)
68	c1ccc(cc1)/C=C/c1cc(cc(c1)OC)OC	2	Ioset et al. (2001)
69	c1ccc(cc1)/C=C/c1c(c(cc(c1CC=C(C)C)O)O)CC=C(C)C	4.0	Ioset et al. (2001)
70^	c1ccc(cc1)/C=C/c1cc(c(c(O)c1)CC=C(C)C)O	4.0	Ioset et al. (2001)
71^	O1[C@@H](CC(=O)c2c1cc(O)cc2O)c1ccc(O)cc1	3.7	Ho et al. (2003)
72	O([C@@H]1[C@@H](O)[C@H](O)[C@H](O[C@H]1Oc1cc2O[C@@H](CC(=O)c2c(O)c1)c1ccc(O)cc1)CO)[C@@H]1O[C@H]([C@H](O)[C@@H](O)[C@H]1O)C	0.1	Rajkumar et al. (2008)
73*	O([C@@H]1[C@@H](O)[C@H](O)[C@H](O[C@H]1Oc1cc2O[C@@H](CC(=O)c2c(O)c1)c1ccc(OC)cc1)CO)[C@@H]1O[C@H]([C@H](O)[C@@H](O)[C@H]1O)C	0.1	Rajkumar et al. (2008)
74*	O([C@@H]1[C@@H](O)[C@H](O)[C@H](O[C@H]1Oc1cc2oc(cc(=O)c2c(O)c1)c1ccc(O)cc1)CO)[C@@H]1O[C@H]([C@H](O)[C@@H](O)[C@H]1O)C	0.1	Rajkumar et al. (2008)
75	o1c2c(c(OC)c3OCOc3c2)c(=O)c(OC)c1c1cc2OCOc2cc1	0.5	Midiwo et al. (2005)
76	O1[C@H]2[C@H](c3c(OC2)cc(OC)c(OC)c3)C(=O)c2c1c1c(OC(C=C1)(C)C)cc2	5	Ollis et al. (1967)
77*	O1[C@H]2[C@](O)(c3c(OC2)cc(OC)c(OC)c3)C(=O)c2c1c1c(OC(C=C1)(C)C)cc2	5	Ollis et al. (1967)
78	O1[C@@H]2COc3c(cc(OC)c(OC)c3)[C@@H]2C(=O)[C@@H]2C=CC3=C(C=CC(O3)(C)C)[C@@H]12	24.6	Vasconcelos et al (2009)
79*	O=C1C[C@H](CC=C1C)C(=C)C	43.8	Simas et al. (2004)
80	O1[C@]2(CC[C@@H](C1(C)C)CC2)C	60	Araujo et al. (2003)
81*	Oc1c(=O)cc(C(C)C)ccc1	2.9	Jang et al. (2005)
82	OC/C=C(/CCC=C(C)C)\C	81.6	Simas et al. (2004)
83^	O[C@@](CCC=C(C)C)(C)C=C	96.6	Perumalsamy et al. (2010)
84	C(CC(=C)C=C)C=C(C)C	66.4	Perumalsamy et al. (2010)
85	O[C@H]1[C@@H](CC[C@H](C1)C)C(C)C	365.8	Pandey et al. (2013)
86*	C1[C@]2(CC[C@@](C1)(O2)C)C(C)C	751	Santos et al. (2010)

87*	OC([C@H]1CCC(=CC1)C)(C)C	111.8	Perumalsamy et al. (2010)
88	O[C@@]1(C(C)C)CCC(=CC1)C	64.8	Perumalsamy et al. (2010)
89^	C1C(=C(C)C)CC=C(C1)C	15.3	Perumalsamy et al. (2010)
90	C(C1=CCC(=CC1)C)(C)C	17.1	Perumalsamy et al. (2010)
91^	[C@H]1(C(C)C)CC=C(C=C1)C	32.1	Perumalsamy et al. (2010)
92	O=C1[C@H]2C([C@H](C2)C(=C1)C)(C)C	93.2	Perumalsamy et al. (2010)
93^	[C@@]12(CC([C@H](C1)C=C2)(C)C)C	69.3	Perumalsamy et al. (2010)
94	O[C@@H]1[C@]2(C([C@@H](C1)CC2)(C)C)C	94.9	Perumalsamy et al. (2010)
95	O=C1[C@]2(C([C@H](CC2)C1)(C)C)C	657	Santos et al. (2010)
96*	OC1C2(C(C(C1)CC2)(C)C)C	598	Santos et al. (2010)
97	C1=C[C@H]2C[C@@H]1[C@@H]([C@@H]2CCO)CCO	785	Santos et al. (2010)
98	C1=C[C@H]2C[C@@H]1C[C@@H]2O	759	Santos et al. (2010)
99	C1([C@@H]2[C@H]1CC=C(C2)C)(C)C	19.2	Perumalsamy et al. (2010)
100	O=C1CC(C)(C)C=CC=C1C	130.4	Perumalsamy et al. (2010)
101	C(c1ccc(cc1)C)(C)C	51	Santos et al. (2010)
102*	Oc1c(C(C)C)ccc(c1)C	81	Santos et al. (2010)
103	Oc1cc(C(C)C)ccc1C	69	Santos et al. (2010)
104	O[C@@](CC/C=C(/CCC=C(C)C)\C)(C)C=C	17	Simas et al. (2004)
105*	OC/C=C(/CC/C=C(/CCC=C(C)C)\C)\C	13	Simas et al. (2004)
106	O=C(C[C@@H](c1ccc(cc1)C)C)C=C(C)C	30	Neves et al. (1999)
107	O1[C@@H]2C[C@@]3(C(=C[C@@H]2C(=C)C1=O)C(=C)CCC3)C	10	Cantrell et al. (2005)
108	C1([C@H]2[C@H](C1)C(=C)CC/C=C(/CC2)\C)(C)C	88.3	Cantrell et al. (2005)
109	C=C1CC[C@@H]2C(C)([C@H](OC=O)CC[C@@H]2[C@@H]1CC/C(=C\C(=O)O)/C)C	0.8	Geris et al. (2008)
110*	C=C1CC[C@@H]2C(C)([C@@H](CC[C@@H]2[C@@H]1CC/C(=C\C(=O)O)/C)O)C	87.3	Geris et al. (2008)

111	C1CCC([C@H]2[C@H]1[C@@H]1[C@@H]3[C@H]([C@H]2O)OC(=O)[C@@H]3c2c(C1)occ2)(C)C	50.1	Omena et al. (2006)
112	C1CCC([C@H]2[C@H]1[C@@H]1[C@H]([C@@H]([C@H]2O)O)[C@H](c2c(C1)occ2)C(=O)O)(C)C	14.7	Omena et al. (2006)
113*	C1CCC([C@H]2[C@H]1[C@@H]1[C@H]([C@@H]([C@H]2O)O)[C@H](c2c(C1)occ2)C(=O)OC)(C)C	21.8	Omena et al. (2006)
114*	C1CCC([C@H]2[C@H]1[C@@H]1[C@@H](Cc3c(C1)occ3)C[C@H]2OC=O)(C)C	186.2	Pimenta et al. (2006)
115^	O([C@H]1[C@@]2([C@@H]([C@@]3([C@@H]([C@H]1O)C(C(=O)C=C3)(C)C)C)CC[C@@]1(C2=CC[C@H]1C1=C[C@H](OC1=O)OC)C)C)C(=O)C=C(C)C	2.1	Siddiqui et al. (2008)
116	O([C@@H]1[C@@H]2[C@]([C@@H]3[C@](C4=CC[C@H]([C@@]4(CC3)C)c3ccoc3)([C@@H]1O)C)(C)C=CC(=O)C2(C)C)C=O	21	Siddiqui et al. (2008)
117	O1[C@@H]2[C@@H]3[C@]([C@@H]4[C@](C5=CC(=O)O[C@@H](c6ccoc6)[C@@]5(CC4)C)([C@@H]2O)C)(C)[C@@H](CC(=O)[C@@]13C)OC(=O)C(=C(C)C)C	83	Siddiqui et al. (2008)
118	O1[C@H]([C@@]2(CC[C@@H]3[C@]([C@]42[C@@H](C1=O)O4)(C)[C@@H](C(=O)[C@]1([C@@]3(C)/C=C/C(=O)OC)O[C@@]1(C)C)O)C)c1ccoc1	69	Freitas et al. (2009)
119	O([C@H]1[C@]2([C@]34[C@](C3)([C@@H](CC4)[C@H]3C[C@H](O[C@H]3O)[C@H]3OC3(C)C)CC[C@@H]2[C@@]2([C@@H](C1)C([C@H](O)CC2)(C)C)C)C)C(=O)C	4.8	Freitas et al. (2009)
120	O1c2cc(cc(c2OC1)c1c2c(cc(/C=C/C=C/C(=O)N3CCCCC3)c1)OCO2)CC/C=C/C(=O)N1CCCCC1	45	Siddiqui et al. (2008)
121*	O1c2cc(cc(c2OC1)c1c2c(ccc1/C=C/C=C/C(=O)N1CCCCC1)OCO2)CC/C=C/C(=O)N1CCCCC1	40	Siddiqui et al. (2004)
122	O=C(NC(C)(C)C)CCCCCCCCCCC	25	Siddiqui et al. (2004)
123^	O=C(NC(C)(C)C)CCCCCCCCCCCCCCC	27	Siddiqui et al. (2004)
124	O=C(N1CCCCC1)/C=C/CCc1cc2OCOc2cc1	17	Siddiqui et al. (2004)
125	O=C(N1CCCC1)/C=C/C=C\C=C\CCCCCc1cc2OCOc2cc1	20	Siddiqui et al. (2004)
126^	O=C(N1CCCC1)/C=C/C=C\C=C\CCCc1cc2OCOc2cc1	25	Siddiqui et al. (2004)
127	O=C(/C=C/C=C\CCCCC/C=C/c1cc2OCOc2cc1)NCCC(C)C	30	Siddiqui et al. (2005)
128	O=C(/C=C/C=C\CCC)NCCC(C)C	23	Siddiqui et al. (2005)
129	O=C(N1CCCC1)/C=C/CCCCCCCCCCCCC	64	Siddiqui et al. (2005)
130	O=C(/C=C/C=C\CCCCCCCCCCCCC)NCC(C)C	25	Siddiqui et al. (2005)

131^	O=C(/C=C/C=C\CCCCCCCCCCCCCCC)NCC(C)C	25	Siddiqui et al. (2005)
132^	O=C(NCC(C)C)/C=C\C=C\CCC/C=C/c1cc2OCOc2cc1	25	Siddiqui et al. (2005)
133	c1ccc(c(c1)C(=O)OC[C@H](CCCC)CC)C(=O)OC[C@H](CCCC)CC	79	Katade et al. (2006)
134*	c1(ccc(cc1OC)CCC(=O)C[C@@H](CCC)O)O	4.3	Rahuman et al. (2008)
135^	c1(ccc(cc1OC)/C=C/C(=C/C(=O)CCCCC)/O)O	9.8	Rahuman et al. (2008)
136	c1(ccc(cc1OC)CCC(=O)C[C@@H](CCCCC)O)O	18.2	Rahuman et al. (2008)
137	CCCCCCCCCCCCCCC	96.7	Rahuman et al. (2008)
138	C12=C(C(=O)C(=O)C(=C(C)C)C[C@H]2[C@H](CC1)O)C	5.8	Madhu et al. (2010)
139*	C12=C(C)CC(=O)C(=C(C)C)C[C@H]2[C@H](CC1)O	13.6	Madhu et al. (2010)
140	CCCSSCCC	6	Balandrin et al (1988)
141	O1[C@@]23C(=C)[C@@]4(C=CC(=O)OC4(C)C)C[C@@H](OC(=O)C)[C@]3([C@@]34[C@@]1(C(=O)O[C@](C)(C4=C)[C@@H]2OC(=O)C)[C@@H](OC3=O)C)C	7.3	Santos et al. (2003)
142*	O1[C@@H]([C@@]2(O)[C@@]3([C@]4([C@H](C[C@@](C3=C)(C2=O)C)C(=C)[C@@]2(CC4)C(OC(=O)C=C2)(C)C)C)C1=O)C	10	Santos et al. (2003)
143	O1C(=O)[C@]2(O)[C@@]3(C(=C)[C@@]1([C@H](OC(=O)C)C1=C([C@]4(C(OC(=O)C=C4)(C)C)CC[C@]31C)C)C)C(=O)O[C@H]2C	11	Santos et al. (2003)
144	o1c(cc(=O)c(O)c1)CO	204.5	Siddhardha et al. (2010)
145	O[C@H]1[C@@H](CC[C@H](C1)C)C(=C)C	297	Santos et al. (2011)
146^	C1(=CC[C@H]2[C@@H](C1)C2(C)C)C	150	Santos et al. (2010)
147^	C1(=CC[C@H](C(=C)C)CC1)C	18.1	Cheng et al (2009)
148*	C(CC(=C)C=C)C=C(C)C	35.8	Cheng et al. (2013)
149^	[C@]12([C@@H](C1)C(=C)CC2)C(C)C	27.3	Cheng et al. (2013)
150*	C1(=CC=C(CC1)C)C(C)C	14.7	Cheng et al (2009)
151^	O1O[C@@]2(CC[C@@]1(C(C)C)C=C2)C	9.6	Torres et al. (2008)
152	[C@@H]12C(C(=C)[C@@H](C1)CC2)(C)C	67	Perumalsamy et al. (2009)

153^	O[C@H]1C[C@H](CC=C1C)C(=C)C	32.9	Govindarajan et al. (2012)
154	OCC[C@H](CCC=C(C)C)C	49.9	Ali et al. (2013)
155^	O(C/C=C(/CCC=C(C)C)\C)C=O	58.5	Ali et al. (2013)
156^	O=C1C[C@H](CCC1=C(C)C)C	48.7	Waliwitiya et al. (2009)
157^	O[C@@]1(C(C)C)CCC(=CC1)C	64.8	Perumalsamy et al. (2009)
158^	c1(cc(cc(c1)O)O)C	64.1	Perumalsamy et al. (2009)
159*	c1(cc(c(c(c1)O)O)O)C	67.1	Perumalsamy et al. (2009)
160^	[C@H]1(C(C)C)CC/C(=C\CCC(=C)/C=C\1)/C	63.6	Govindarajan et al. (2010)
161^	[C@]1([C@H](C=CCC1)C(=C)C)(C)C=C	43.4	Kiran et al. (2006)
162*	C1C/C(=C\C=C/CC/C(=C\1)/C)/C	28.3	Kiran et al. (2006)
163	O[C@]1([C@@H](CCC=C(C)C)C)CCC(=CC1)C	33.2	Rajkumar et al. (2010)
164^	O[C@](C=C)(CC/C=C(\C)/CCC=C(C)C)C	13.4	Ali et al. (2013)
165*	[C@]123[C@H]([C@]4([C@H](CC1)C(CCC4)(C)C)C)CC[C@H](C2)C(=C)C3	57	Cheng et al. (2009)
166	O(c1ccc(cc1)/C=C/C)C	42	Cheng et al. (2004)
167	O(Cc1ccccc1)C(=O)c1ccccc1	6.8	Jantan et al. (2005)
168	O(Cc1ccccc1)C(=O)c1c(O)cccc1	6.8	Jantan et al. (2005)
169*	O=C/C=C/c1ccccc1	24	Cheng et al. (2004)
170*	O(C/C=C/c1ccccc1)C(=O)C	33	Cheng et al. (2004)
171*	O(C(=O)/C=C/c1ccccc1)C	26	Cheng et al. (2004)
172	O(C(=O)/C=C/c1ccccc1)CC	33	Cheng et al. (2004)
173	O=C(/C(=C/c1ccc(cc1)OC)/CC)O	18.9	Kim et al. (2008)
174	O=C(O)/C=C/c1ccc(cc1)OC	61	Kim et al. (2008)
175^	c1(ccc(cc1)OC)CC=C	46.4	Govindarajan et al (2010)
176^	CCCC(=O)OCCCCCC	74.9	Tabaca et al. (2012)

177*	CC(=O)CCCCCCCCC	14.4	Tabaca et al (2012)
178	O=C1[C@]2(C([C@H](CC2)C1)(C)C)C	657	Santos et al. (2010)
179^	O1[C@]2([C@@H]1C[C@H](CC2)C(=C)C)C	517	Santos et al. (2011)
180*	O[C@H]1[C@@H](CC[C@H](C1)C)C(C)C	404	Santos et al. (2011)
181*	O=C1[C@@H](CC[C@H](C1)C)C(C)C	508	Santos et al. (2011)
182	OC([C@H]1CCC(=CC1)C)(C)C	331.7	Padey et al. (2013)
183^	O1[C@@]2([C@@H]1CCC(=C)[C@H]1[C@@H](C(C1)(C)C)CC2)C	125	Silva et al (2008)
184*	CCCCCCOC(=O)C(C)C	106.3	Tabaca et al. (2012)
185^	CCCCCCOC(=O)[C@@H](C)CC	107.7	Tabaca et al. (2012)
186	C(CCCCCOC(=O)C)CC	148.9	Tabaca et al. (2012)
187	O=Cc1ccccc1	50	Cheng et al. (2009)
188*	O=Cc1ccc(cc1)O	50	Cheng et al. (2009)
189	c1(ccccc1)CCC=O	50	Cheng et al. (2009)
190^	c1(ccccc1)/C=C\C(=O)O	50	Cheng et al. (2009)
191^	c1(ccccc1)/C=C/CO	50	Cheng et al. (2009)
192^	O([C@H]1[C@@]2(C([C@H](C1)CC2)(C)C)C)C(=O)C	50	Cheng et al. (2009)
193	OC(=O)CCCCCCCCCCCCCCC	57.2	Rahuman et al. (2010)
194^	O(c1cc(CC=C)ccc1OC)C	43.0	Kihampa et al. (2010)
195	c1cc2c(cc1)C(=O)C(=C(C2=O)N1CCOCC1)Cl	13.9	Kim et al (2013)
196*	c12c(occ1)C[C@@H]1[C@@H]3[C@@H]([C@H]([C@H]4[C@@]1(CCCC4(C)C)C)O)OC(=O)[C@@H]23	50.1	Omena et al. (2006)
197*	c12c(occ1)C[C@@H]1[C@@H]([C@H]2C(=O)O)[C@H]([C@H]([C@H]2[C@@]1(CCCC2(C)C)C)O)O	14.7	Omena et al. (2006)
198^	c12c(occ1)C[C@@H]1[C@@H]([C@H]2C(=O)OC)[C@H]([C@H]([C@H]2[C@@]1(CCCC2(C)C)C)O)O	21.8	Omena et al. (2006)
199	O[C@@H]1CC2=CC[C@H]3[C@H]4[C@@]([C@	46	Kihampa et

	H](CC4)[C@H](C)/C=C/[C@H](C(C)C)CC)(CC[C@@H]3[C@]2(CC1)C)C		al. (2010)
200	c1cc2c(cc1)C(=O)C(=C(C2=O)N1CCCC1)Cl	5.7	Kim et al (2013)
201	C(#CC#CCC/C=C\C=C\C(=O)NCC(C)C)C	100	Clifford et al. (2002)
202	c1(c2c(c(cc1)O)cccc2)O	3.4	Kim et al (2013)
203	C1(=C(C(=O)c2c(C1=O)cccc2)O)CC=C(C)C	3.7	Ribeiro et al. (2009)
204*	C1(=C(C(=O)c2c(C1=O)cccc2)OC(=O)C)CC=C(C)C	9.6	Ribeiro et al. (2009)
205	C1(=O)c2c(C(=O)C=C1)cccc2O	3.6	Ribeiro et al. (2009)
206*	C1(=O)c2c(C(=O)C(=C1)C)ccc(c2O)c1cc(c2c(C(=O)CCC2=O)c1C)O	40.7	Ribeiro et al. (2009)
207*	C1(=O)c2c(C(=O)C=C1[C@H](CC=C(C)C)C)c(ccc2O)O	3.9	Chen et al. (2003)
208	c12c(C(=O)C(=O)C=C2)cccc1	1.8	Kim et al (2013)
209	c12c(cccc1)oc(=O)cc2	76.4	Wang et al. (2012)
210	c12c(cc(cc1)O)oc(=O)cc2	132.7	Wang et al. (2012)
211*	c12c(cc(cc1)OC)oc(=O)cc2	44.4	Wang et al. (2012)
212	c12c(cc(cc1)OCC)oc(=O)cc2	40.5	Wang et al. (2012)
213*	c12c(cc(c(c1)OC)OC)oc(=O)cc2	60.5	Wang et al. (2012)
214^	c12c(cc3c(c1)cco3)oc(=O)cc2	34.2	Wang et al. (2012)
215	c12c(c(c(cc1)OC)CC=C(C)C)oc(=O)cc2	13.1	Wang et al. (2012)
216^	c12c(c(c3c(c1)cco3)OC)oc(=O)cc2	45.1	Wang et al. (2012)
217	c12c(oc(=O)cc2)cc2c(c1OC)cco2	73.5	Wang et al. (2012)
218^	c12c(oc(=O)cc1)c(c1c(cco1)c2)OCC=C(C)C	2.9	Wang et al. (2012)
219^	c12c(oc(=O)cc1)c(c1c(cco1)c2OC)OC	6.8	Wang et al. (2012)
220*	c12c(oc(=O)cc1)cc1c(cco1)c2OCC=C(C)C	13	Wang et al. (2012)
221	c12c(oc(=O)cc1)c1c(occ1)cc2	26.4	Wang et al. (2012)
222	c12c(oc(=O)c(c2C)Br)c(c(c(c1)Br)O)Br	2.2	Deshmukh et al. (2008)

223*	c12c(oc(=O)c(c2C)Br)cc(O)cc1	78	Deshmukh et al. (2008)
224	c12c(oc(=O)c(c2C)Br)cc(OC(=O)C)cc1	89.6	Deshmukh et al. (2008)
225	Brc1ccc(c2ccc([C@H]3C[C@H](c4c(C3)cccc4)c3c(O)c4c(oc3=O)cccc4)cc2)cc1	8.2	Jung and Moon (2011)
226	FC(F)(F)c1ccc(COc2ccc([C@H]3C[C@H](c4c(C3)cccc4)c3c(O)c4c(oc3=O)cccc4)cc2)cc1	9.3	Jung and Moon (2011)
227	C(CCCC/C=C/C=C/C(=O)NCCC(C)C)CCC	28	Gulzar et al. (2013)
228	C(CCCC/C=C/C=C/C(=O)NCCC(C)C)CCCCCCC	22	Gulzar et al. (2013)
229^	C(CCCC/C=C/C=C/C(=O)N1CCCC1)CCCCCCCC	31	Gulzar et al. (2013)
230^	C(CCC/C=C/C=C/C(=O)NCCC(C)C)C	20	Gulzar et al. (2013)
231^	C(CCC/C=C/C=C/C(=O)NCC(C)C)CCCCCCCCC	25	Gulzar et al. (2013)
232*	o1c2c(c(=O)c(OC)c1c1ccccc1)ccc1occc21	16.1	Perumalsamy et al. (2015)
233	O1C(C=Cc2c1ccc1c2oc(c(OC)c1=O)c1ccccc1)(C)C	20.6	Perumalsamy et al. (2015)
234*	OC(=O)CCCCCCCCCCCCCCC	43	Perumalsamy et al. (2015)
235^	o1c2c(c(OC)c(C(=O)CC(=O)c3ccccc3)cc2)cc1	25.8	Perumalsamy et al. (2015)
236	o1c2c(cc1)cc1c(=O)c3c4c(cccc4)OCc3oc1c2	37.6	Perumalsamy et al. (2015)
237^	OC(=O)CCCCCCC/C=C/CCCCCCCC	32.2	Perumalsamy et al. (2015)
238*	OC(=O)CCCCCCCCCCCCCCCCCCC	60.5	Perumalsamy et al. (2015)
239	OC(=O)CCCCCCCCCCCCCCCCCCCCC	86.8	Perumalsamy et al. (2015)
240^	OC(=O)CCCCCCC/C=C\C/C=C\C/C=C\CC	22.6	Perumalsamy et al. (2015)
241^	c1cc2c(cc1)C(=O)C(=C(C2=O)CC=C(C)C)O	8.8	Kim et al. (2013)
242	C1([C@@H]2C[C@H]1C(=CC2)C)(C)C	79.1	Cheng et al. (2009)
243*	C1([C@@H]2C[C@H]1C(=C)CC2)(C)C	27.7	Govindarajan et al (2010)
244	O(C)c1c(/C=C/C)cc(OC)c(OC)c1	25.9	Perumalsamy et al. (2015)
245	C1=C[C@@H]2[C@H]([C@H]([C@H]1C2)CO)CO	1062	Santos et al. (2010)
246^	O([C@H]1[C@@]2([C@@H]([C@@]3([C@@H]([C@H]1O)C(C(=O)C=C3)(C)C)C)CC[C@@]1(C2=C	0.7	Siddiqui et al. (2008)

	C[C@H]1C1=CC(=O)O[C@@H]1O)C)C)C(=O)C		
247	c12c(c(c(c(c1)Br)O)Br)oc(=O)c(c2C)Br	2.2	Deshumukh et al. (2008)
248*	c12c(cc(cc1)O)oc(=O)c(c2C)Br	34.4	Deshumukh et al. (2008)
249	c12c(c(c(c(c1)Br)OC(=O)C)Br)oc(=O)c(c2C)Br	78	Deshumukh et al. (2008)
250*	c12c(cc(cc1)OC(=O)C)oc(=O)c(c2C)Br	89.6	Deshumukh et al. (2008)
251	c12c(cccc1)OC(=O)[C@H]([C@@H]2O)c1cc2c(cc1)[C@@H](C[C@H](C2)c1ccc(cc1)c1ccc(cc1)Br)C	8.2	Jung and Moon (2011)
252	c12c(cccc1)SC(=O)[C@H]([C@@H]2O)c1cc2c(cc1)[C@@H](C[C@H](C2)c1ccc(cc1)c1ccc(cc1)Br)C	41.4	Jung and Moon (2011)
253*	c12c(cccc1)OC(=O)[C@H]([C@@H]2O)c1cc2c(cc1)[C@@H](C[C@H](C2)c1ccc(cc1)OCc1ccc(cc1)C(F)(F)F)C	9.3	Jung and Moon (2011)
254*	c12c(c(c(c(c1OC)OC)OC)OC)c(=O)cc(o2)c1ccc(cc1)OC	7.2	Rao et al. (1990)
255*	c12c(c(cc(c1)O)O)c(=O)c(co2)c1ccc(cc1)O	10.8	Rao et al. (1990)
256	c12c(ccc(c1)O)c(=O)c(co2)c1ccc(cc1)O	7.6	Rao et al. (1990)
257^	c12c(ccc(c1)O)c(=O)c(co2)c1ccc(cc1)OC	10.7	Rao et al. (1990)
258*	c12c(c(cc(c1)OC(=O)C)OC(=O)C)c(=O)c(co2)c1ccc(cc1)OC(=O)C	4	Rao et al. (1990)
259	c12c(ccc(c1)OC(=O)C)c(=O)c(co2)c1ccc(cc1)OC(=O)C	2	Rao et al. (1990)
260^	c12c(ccc(c1)OC(=O)C)c(=O)c(co2)c1ccc(cc1)OC	2.5	Rao et al. (1990)
261*	[C@@]12([C@H](C(C(=O)C=C1)(C)C)[C@]([C@@H]([C@@]1([C@H]2CC[C@@]2(C1=CC[C@@H]2C1=C[C@H](OC1=O)O)C)C)OC(=O)C)(C)O)C	0.6	Champagne et al. (1992)
262	[C@@]12([C@H](C(C(=O)C=C1)(C)C)[C@]([C@@H]([C@@]1([C@H]2CC[C@@]2(C1=CC[C@@H]2C1=CC(=O)O[C@@H]1O)C)C)OC(=O)C)(C)O)C	0.7	Champagne et al. (1992)
263^	c1(ccc(c(c1)OC)OC)/C=C\C	59	Santos et al. (2011)

* Validation set compound

^ Test set compound

Table 2S. List of mathematical equations used in the present study.

$$RMS = \sqrt{\frac{\sum_{i=1}^{N}\left(p_i^{exp} - p_i^{pred}\right)^2}{N}}$$

$$VIF = \frac{1}{1 - R_{ij}^2}$$

$$h_i = x_i(\mathbf{X}^T\mathbf{X})^{-1}x_i^T$$

$$h^* = 3(d+1)/N_{train}$$

$$b_j^s = \frac{S_j\, b_j}{S_{p_{exp}}}$$

$$k = \frac{\sum_{i=1}^{N_{test}}\left(p_i^{exp} p_i^{pred}\right)^2}{\sum_{i=1}^{N_{test}}\left(p_i^{pred}\right)^2}$$

$$k' = \frac{\sum_{i=1}^{N_{test}}\left(p_i^{exp} p_i^{pred}\right)^2}{\sum_{i=1}^{N_{test}}\left(p_i^{\exp}\right)^2}$$

$$p_0^{exp} = kp^{pred}$$

$$p_0^{pred} = k' p^{exp}$$

$$R_0^2 = 1 - \frac{\sum_{i=1}^{N_{test}}\left(p_i^{exp} - p_{0i}^{exp}\right)^2}{\sum_{i=1}^{N_{test}}\left(p_i^{exp} - p_{av}^{exp}\right)^2}$$

$$R_0^{'2} = 1 - \frac{\sum_{i=1}^{N_{test}} \left(p_i^{pred} - p_{0i}^{pred} \right)^2}{\sum_{i=1}^{N_{test}} \left(p_i^{pred} - p_{av}^{pred} \right)^2}$$

$$R_m^2 = R_{test}^2 \left(1 - \left| \sqrt{R_{test}^2 - R_0^2} \right| \right)$$

$$A\% = \frac{TP + TN}{TOTAL}$$

$$SE = \frac{TP}{P}$$

$$SP = \frac{TN}{N}$$

S: standard deviation; *N*: number of molecules; p_i^{exp} : experimental activity for compound *i*; p_i^{pred} : predicted activity for compound *i*; *d*: number of descriptors; *RMS*: root mean square error; *VIF* : variance inflation factor; R_{ij}^2 : squared correlation coefficient between a given descriptor and the remaining ones of the model; h_i : leverage for compound *i*; x_i: descriptor vector for *i*; $\mathbf{X}$: model matrix for the training set (train); N_{train} : number of molecules in the training; N_{test} =number of molecules in the test set; h^* : warning leverage; b_j : regression coefficient for the *j*th descriptor; b_j^s : standardized b_j ; S_j : standard deviation for the *j*th descriptor; $S_{p_{exp}}$: standard deviation for the experimental activity; p_{0i}^{exp} and p_{0i}^{pred} : calculated activities for *i* in the test set in regressions through the origin of p^{exp} against p^{pred} and p^{pred} against p^{exp}, respectively; R_0^2 and $R_0^{'2}$: squared correlation coefficients for regressions through the origin in the test set of p^{exp} against p^{pred} and p^{pred} against p^{exp}, respectively; p_{av}^{pred} : average value for p^{pred} in the test set; p_{av}^{exp} : average value for p^{exp} in the test set; R_m^2 :

modified squared correlation coefficient; R^2_{test} : squared correlation coefficient between observed and predicted values for the test set; *A%*: accuracy percent; *TP* : True Positive, toxic compounds correctly predicted; *TN* : True negative, non-toxic compounds correctly predicted; *TOTAL*: all compounds correctly predicted; *SE*: Sensitivity; *P*: all compounds predicted as toxic; *SP*: Specificity; *N*: all compounds predicted as non-toxic.

Table 3S. A brief description of different molecular descriptors calculated by means of PaDEL, Mold^2, EPI Suite and QuBiLs softwares.

Software	Class	Symbol	Description
PaDEL	Autocorrelation	*AATS5v*	Average Broto-Moreau autocorrelation - lag 5 / weighted by van der Waals volumes
		AATS5i	Average Broto-Moreau autocorrelation - lag 5 / weighted by first ionization potential
	Klekota-Roth fingerprint	*KR1592*	Presence of chemical substructures [!#1]c1[cH][cH][cH]c([!#1])c1[!#1]
		KR3584	Presence of chemical substructure Cc1cccc(c1)c2cccc(C)c2
		KRC4736	Count of chemical substructure Oc1ccc(O)cc1
	Substructure fingerprint	*Sub282*	Presence of chemical substructure [OX2;$([r5]1@C(!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C1),$([r6]1@C(!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C@C1)]
	Ring count	*nRing*	Number of rings
	Pubchem fingerprint	*PC34*	Hierarchic element counts of 2S
		PC199	Presence of SMARTS Patterns of any ring size 6
		PC777	Complex SMARTS patterns of count CC1CCC(O)CC1
	2D-Atom pairs fingerpring	*AP653*	Presence of O-Br at topological distance 9
	Barysz matrix	*VE2_Dze*	Average coefficient sum of the last eigenvector from Barysz matrix / weighted by Sanderson electronegativities
		VE1_Dze	Coefficient sum of the last eigenvector from Barysz matrix / weighted by Sanderson electronegativities
	MACCS Keys	*M16*	QAA@1
Mold^2	Non-conformational	*D178*	All-Path Wiener - sum of the edges in the shortest paths between all pairs of non-hydrogen atoms
		D589	Highest eigenvalue from Burden matrix weighted by polarizabilities order-2
		D590	Highest eigenvalue from Burden matrix weighted by polarizabilities order-3

Table 4S. The $\log LC_{50}$ experimental and predicted values obtained according to Eq. 1 and the leverage values for each compound. The warning leverage is $h^* = 0.1805$.

ID	LC_{50} (µg. mL-1) exp.	LC_{50} (µg. mL-1) pred.	Leverage (h_i)
1^	0.94	1.70	0.01
2	1.26	1.69	0.01
3	2.15	1.44	0.01
4*	1.48	1.44	0.01
5^	1.88	1.44	0.01
6^	0.92	1.44	0.01
7	0.98	1.52	0.01
8	-1.72	-1.37	0.33
9	-1.10	-0.95	0.33
10*	0.59	1.12	0.01
11	0.59	1.07	0.01
12*	0.59	0.95	0.01
13	1.29	0.83	0.01
14	0.73	0.55	0.14
15*	1.13	1.17	0.01
16	0.10	0.93	0.14
17	1.61	0.98	0.09
18	1.49	1.12	0.09
19	0.52	0.88	0.01
20^	1.42	1.18	0.01
21	0.98	1.17	0.01
22^	0.68	1.23	0.01
23	0.55	0.63	0.14
24^	0.66	0.60	0.14
25*	0.90	0.62	0.14
26*	0.14	0.28	0.14
27	0.07	0.27	0.14
28	0.99	0.28	0.14
29	-0.06	0.30	0.14
30	0.86	0.29	0.14
31*	0.76	0.29	0.14
32	1.17	1.31	0.01
33	1.95	1.78	0.01
34*	1.79	1.74	0.01
35	1.62	1.72	0.01
36	1.76	1.67	0.01
37	1.52	1.69	0.01
38^	1.68	1.61	0.01
39^	0.28	0.79	0.01
40^	1.29	0.82	0.01
41	1.21	0.84	0.01
42*	1.19	0.88	0.01

43^	0.26	0.84	0.01
44	0.87	0.78	0.01
45*	1.18	0.78	0.01
46	1.18	0.82	0.01
47	1.18	0.77	0.01
48	0.32	0.56	0.01
49	1.02	0.68	0.01
50^	1.03	1.62	0.01
51	1.06	1.61	0.01
52*	1.04	1.60	0.01
53*	2.01	1.66	0.01
54	1.39	1.52	0.01
55^	1.86	1.66	0.01
56^	1.69	1.63	0.01
57	2.71	1.75	0.01
58	2.13	1.95	0.01
59	2.29	2.23	0.01
60*	2.39	2.17	0.01
61	2.76	2.12	0.01
62	2.25	2.13	0.01
63*	2.30	1.68	0.01
64^	1.35	1.38	0.01
65^	1.18	1.20	0.01
66	0.95	1.40	0.01
67*	1.70	1.23	0.01
68	0.30	0.82	0.01
69	0.60	0.80	0.01
70^	0.60	0.80	0.01
71^	0.57	0.77	0.01
72	-0.91	-0.93	0.74
73*	-0.99	-0.94	0.74
74*	-1.07	-0.96	0.74
75	-0.33	0.65	0.01
76	0.70	1.28	0.09
77*	0.70	1.27	0.09
78	1.39	1.38	0.09
79*	1.64	1.51	0.01
80	1.78	2.19	0.01
81*	0.46	1.58	0.01
82	1.91	1.45	0.01
83^	1.98	1.52	0.01
84	1.82	1.42	0.01
85	2.56	2.19	0.01
86*	2.88	2.17	0.01
87*	2.05	1.80	0.01
88	1.81	1.85	0.01
89^	1.19	1.56	0.01
90	1.23	1.73	0.01
91^	1.51	1.96	0.01

92	1.97	1.91	0.01
93^	1.84	2.14	0.01
94	1.98	2.58	0.005
95	2.82	2.28	0.01
96*	2.78	2.58	0.005
97	2.89	2.08	0.01
98	2.88	2.45	0.01
99	1.28	1.63	0.01
100	2.12	1.71	0.01
101	1.71	1.81	0.01
102*	1.91	1.80	0.01
103	1.84	1.77	0.01
104	1.23	1.41	0.01
105*	1.11	1.34	0.01
106	1.48	1.18	0.01
107	1.00	1.18	0.01
108	1.95	1.38	0.01
109	-0.10	1.21	0.01
110*	1.94	1.21	0.01
111	1.70	1.26	0.01
112	1.17	1.27	0.01
113*	1.34	1.27	0.01
114*	2.27	1.28	0.01
115^	0.33	-0.24	0.20
116	1.32	0.93	0.01
117	1.92	1.60	0.09
118	1.84	1.70	0.55
119	0.68	0.82	0.55
120	1.65	1.40	0.09
121*	1.60	1.42	0.09
122	1.40	1.95	0.01
123^	1.43	1.92	0.01
124	1.23	1.30	0.01
125	1.30	1.10	0.01
126^	1.40	1.10	0.01
127	1.48	1.17	0.01
128	1.36	1.76	0.01
129	1.81	1.91	0.01
130	1.40	1.74	0.01
131^	1.40	1.74	0.01
132^	1.40	1.17	0.01
133	1.90	1.67	0.01
134*	0.63	1.53	0.01
135^	0.99	1.08	0.01
136	1.26	1.53	0.01
137	1.99	2.05	0.01
138	0.76	1.23	0.01
139*	1.13	1.23	0.01
140	0.78	0.27	0.33

141	0.86	1.59	0.09
142*	1.00	0.93	0.01
143	1.04	1.58	0.09
144	2.31	1.82	0.01
145	2.47	1.89	0.01
146^	2.18	1.63	0.01
147^	1.26	1.56	0.01
148*	1.55	1.42	0.01
149^	1.44	1.94	0.01
150*	1.17	1.79	0.01
151^	0.98	2.02	0.01
152	1.83	2.25	0.01
153^	1.52	1.54	0.01
154	1.70	1.93	0.01
155^	1.77	1.44	0.01
156^	1.69	1.96	0.01
157^	1.81	1.85	0.01
158^	1.81	2.08	0.01
159*	1.83	1.95	0.01
160^	1.80	1.37	0.01
161^	1.64	1.71	0.01
162*	1.45	1.39	0.01
163	1.52	1.35	0.01
164^	1.13	1.41	0.01
165*	1.76	1.35	0.01
166	1.62	1.56	0.01
167	0.83	0.84	0.01
168	0.83	0.83	0.01
169*	1.38	1.52	0.01
170*	1.52	1.58	0.01
171*	1.41	1.40	0.01
172	1.52	1.40	0.01
173	1.28	1.21	0.01
174	1.79	1.37	0.01
175^	1.67	1.69	0.01
176^	1.87	2.09	0.01
177*	1.16	2.02	0.01
178	2.82	2.28	0.01
179^	2.71	2.33	0.55
180*	2.61	2.19	0.01
181*	2.71	2.19	0.01
182	2.52	1.80	0.01
183^	2.10	2.17	0.55
184*	2.03	2.08	0.01
185^	2.03	2.08	0.01
186	2.17	2.02	0.01
187	1.70	1.95	0.01
188*	1.70	1.82	0.01
189	1.70	1.91	0.01

190^	1.70	1.42	0.01
191^	1.70	1.63	0.01
192^	1.70	1.80	0.01
193	1.76	1.92	0.01
194^	1.63	1.66	0.01
195	1.14	1.16	0.01
196*	1.70	1.20	0.01
197*	1.17	1.20	0.01
198^	1.34	1.20	0.01
199	1.66	1.21	0.01
200	0.75	1.16	0.01
201	2.00	1.29	0.01
202	0.53	1.44	0.01
203	0.57	1.18	0.01
204*	0.98	1.17	0.01
205	0.56	0.63	0.14
206*	1.61	1.12	0.09
207*	0.59	1.16	0.01
208	0.25	1.25	0.01
209	1.88	1.46	0.01
210	2.12	1.40	0.01
211*	1.65	1.38	0.01
212	1.61	1.38	0.01
213*	1.78	1.33	0.01
214^	1.53	1.18	0.01
215	1.12	1.29	0.01
216^	1.65	1.18	0.01
217	1.87	1.18	0.01
218^	0.46	1.18	0.01
219^	0.83	1.18	0.01
220*	1.11	1.18	0.01
221	1.42	1.23	0.01
222	0.35	0.74	0.01
223*	1.89	0.96	0.01
224	1.95	2.02	0.50
225	0.92	1.18	0.09
226	0.97	1.33	0.09
227	1.45	1.74	0.01
228	1.34	1.74	0.01
229^	1.49	1.68	0.01
230^	1.30	1.74	0.01
231^	1.40	1.74	0.01
232*	1.21	0.74	0.01
233	1.31	1.38	0.09
234*	1.63	1.92	0.01
235^	1.41	0.76	0.01
236	1.58	1.30	0.09
237^	1.51	1.70	0.01
238*	1.78	1.90	0.01

239	1.94	1.89	0.01
240^	1.35	1.53	0.01
241^	0.95	1.18	0.01
242	1.90	1.93	0.01
243*	1.44	1.93	0.01
244	1.41	1.52	0.01
245	3.03	2.16	0.01
246^	-0.17	-0.23	0.20
247	0.35	0.74	0.01
248*	1.54	0.96	0.01
249	1.89	1.82	0.50
250*	1.95	2.02	0.50
251	0.92	1.24	0.09
252	1.62	1.15	0.09
253*	0.97	1.35	0.09
254*	0.86	0.71	0.01
255*	1.03	0.75	0.01
256	0.88	0.77	0.01
257^	1.03	0.76	0.01
258*	0.60	0.73	0.01
259	0.30	0.75	0.01
260^	0.40	0.76	0.01
261*	-0.21	-0.25	0.20
262	-0.13	-0.24	0.20
263^	1.77	1.52	0.01

*BMS validation set compound

^BSM test set compound

Table 5S. Values of each descriptor included in the established QSAR model.

ID	*M*16	*PC*34	*PC*199	*KR*1592	*AP*653	*SubC*282	*D*589
1^	0	0	0	0	0	0	3.20
2	0	0	0	0	0	0	3.21
3	0	0	0	0	0	0	3.41
4*	0	0	0	0	0	0	3.41
5^	0	0	0	0	0	0	3.41
6^	0	0	0	0	0	0	3.41
7	0	0	0	0	0	0	3.35
8	0	1	0	0	0	0	4.26
9	0	1	0	0	0	0	3.92
10*	0	0	0	0	0	0	3.67
11	0	0	0	0	0	0	3.71
12*	0	0	0	0	0	0	3.81
13	0	0	0	0	0	0	3.91
14	0	0	0	1	0	0	3.57
15*	0	0	0	0	0	0	3.63
16	0	0	0	1	0	0	3.25
17	0	0	1	0	0	0	4.31
18	0	0	1	0	0	0	4.20
19	0	0	0	0	0	0	3.87
20^	0	0	0	0	0	0	3.62
21	0	0	0	0	0	0	3.64
22^	0	0	0	0	0	0	3.58
23	0	0	0	1	0	0	3.50
24^	0	0	0	1	0	0	3.52
25*	0	0	0	1	0	0	3.51
26*	0	0	0	1	0	0	3.79
27	0	0	0	1	0	0	3.80
28	0	0	0	1	0	0	3.79
29	0	0	0	1	0	0	3.77
30	0	0	0	1	0	0	3.78
31*	0	0	0	1	0	0	3.78
32	0	0	0	0	0	0	3.52
33	0	0	0	0	0	0	3.14
34*	0	0	0	0	0	0	3.17
35	0	0	0	0	0	0	3.18
36	0	0	0	0	0	0	3.22
37	0	0	0	0	0	0	3.20
38^	0	0	0	0	0	0	3.27
39^	0	0	0	0	0	0	3.95
40^	0	0	0	0	0	0	3.92
41	0	0	0	0	0	0	3.91
42*	0	0	0	0	0	0	3.87
43^	0	0	0	0	0	0	3.90
44	0	0	0	0	0	0	3.96

45*	0	0	0	0	0	0	3.95
46	0	0	0	0	0	0	3.92
47	0	0	0	0	0	0	3.96
48	0	0	0	0	0	0	4.13
49	0	0	0	0	0	0	4.04
50^	0	0	0	0	0	0	3.27
51	0	0	0	0	0	0	3.28
52*	0	0	0	0	0	0	3.28
53*	0	0	0	0	0	0	3.23
54	0	0	0	0	0	0	3.34
55^	0	0	0	0	0	0	3.23
56^	0	0	0	0	0	0	3.26
57	0	0	0	0	0	0	3.15
58	0	0	0	0	0	0	3.00
59	0	0	0	0	0	0	2.77
60*	0	0	0	0	0	0	2.81
61	0	0	0	0	0	0	2.86
62	0	0	0	0	0	0	2.85
63*	0	0	0	0	0	0	3.21
64^	0	0	0	0	0	0	3.46
65^	0	0	0	0	0	0	3.61
66	0	0	0	0	0	0	3.44
67*	0	0	0	0	0	0	3.59
68	0	0	0	0	0	0	3.92
69	0	0	0	0	0	0	3.94
70^	0	0	0	0	0	0	3.94
71^	0	0	0	0	0	0	3.96
72	0	0	1	0	0	2	3.96
73*	0	0	1	0	0	2	3.96
74*	0	0	1	0	0	2	3.98
75	0	0	0	0	0	0	4.06
76	0	0	1	0	0	0	4.07
77*	0	0	1	0	0	0	4.08
78	0	0	1	0	0	0	3.98
79*	0	0	0	0	0	0	3.35
80	0	0	0	0	0	0	2.80
81*	0	0	0	0	0	0	3.30
82	0	0	0	0	0	0	3.41
83^	0	0	0	0	0	0	3.34
84	0	0	0	0	0	0	3.43
85	0	0	0	0	0	0	2.80
86*	0	0	0	0	0	0	2.81
87*	0	0	0	0	0	0	3.12
88	0	0	0	0	0	0	3.08
89^	0	0	0	0	0	0	3.32
90	0	0	0	0	0	0	3.18
91^	0	0	0	0	0	0	2.99

92	0	0	0	0	0	0	3.02
93^	0	0	0	0	0	0	2.84
94	0	0	0	0	0	0	2.48
95	0	0	0	0	0	0	2.72
96*	0	0	0	0	0	0	2.48
97	0	0	0	0	0	0	2.89
98	0	0	0	0	0	0	2.58
99	0	0	0	0	0	0	3.25
100	0	0	0	0	0	0	3.19
101	0	0	0	0	0	0	3.11
102*	0	0	0	0	0	0	3.12
103	0	0	0	0	0	0	3.14
104	0	0	0	0	0	0	3.44
105*	0	0	0	0	0	0	3.49
106	0	0	0	0	0	0	3.63
107	0	0	0	0	0	0	3.62
108	0	0	0	0	0	0	3.46
109	0	0	0	0	0	0	3.60
110*	0	0	0	0	0	0	3.60
111	0	0	0	0	0	0	3.56
112	0	0	0	0	0	0	3.55
113*	0	0	0	0	0	0	3.55
114*	0	0	0	0	0	0	3.54
115^	0	0	0	0	0	1	3.83
116	0	0	0	0	0	0	3.83
117	0	0	1	0	0	0	3.81
118	1	0	0	0	0	0	3.84
119	1	0	0	0	0	1	3.60
120	0	0	1	0	0	0	3.97
121*	0	0	1	0	0	0	3.95
122	0	0	0	0	0	0	2.99
123^	0	0	0	0	0	0	3.02
124	0	0	0	0	0	0	3.53
125	0	0	0	0	0	0	3.69
126^	0	0	0	0	0	0	3.69
127	0	0	0	0	0	0	3.63
128	0	0	0	0	0	0	3.15
129	0	0	0	0	0	0	3.03
130	0	0	0	0	0	0	3.17
131^	0	0	0	0	0	0	3.17
132^	0	0	0	0	0	0	3.63
133	0	0	0	0	0	0	3.23
134*	0	0	0	0	0	0	3.34
135^	0	0	0	0	0	0	3.71
136	0	0	0	0	0	0	3.34
137	0	0	0	0	0	0	2.91
138	0	0	0	0	0	0	3.59

139*	0	0	0	0	0	0	3.58
140	0	1	0	0	0	0	2.92
141	0	0	1	0	0	0	3.81
142*	0	0	0	0	0	0	3.83
143	0	0	1	0	0	0	3.82
144	0	0	0	0	0	0	3.10
145	0	0	0	0	0	0	3.04
146^	0	0	0	0	0	0	3.25
147^	0	0	0	0	0	0	3.31
148*	0	0	0	0	0	0	3.43
149^	0	0	0	0	0	0	3.00
150*	0	0	0	0	0	0	3.12
151^	0	0	0	0	0	0	2.93
152	0	0	0	0	0	0	2.74
153^	0	0	0	0	0	0	3.33
154	0	0	0	0	0	0	3.01
155^	0	0	0	0	0	0	3.41
156^	0	0	0	0	0	0	2.99
157^	0	0	0	0	0	0	3.08
158^	0	0	0	0	0	0	2.89
159*	0	0	0	0	0	0	3.00
160^	0	0	0	0	0	0	3.47
161^	0	0	0	0	0	0	3.19
162*	0	0	0	0	0	0	3.45
163	0	0	0	0	0	0	3.48
164^	0	0	0	0	0	0	3.44
165*	0	0	0	0	0	0	3.49
166	0	0	0	0	0	0	3.31
167	0	0	0	0	0	0	3.91
168	0	0	0	0	0	0	3.91
169*	0	0	0	0	0	0	3.34
170*	0	0	0	0	0	0	3.30
171*	0	0	0	0	0	0	3.44
172	0	0	0	0	0	0	3.45
173	0	0	0	0	0	0	3.60
174	0	0	0	0	0	0	3.47
175^	0	0	0	0	0	0	3.21
176^	0	0	0	0	0	0	2.88
177*	0	0	0	0	0	0	2.94
178	0	0	0	0	0	0	2.72
179^	1	0	0	0	0	0	3.32
180*	0	0	0	0	0	0	2.80
181*	0	0	0	0	0	0	2.80
182	0	0	0	0	0	0	3.12
183^	1	0	0	0	0	0	3.45
184*	0	0	0	0	0	0	2.88
185^	0	0	0	0	0	0	2.89

186	0	0	0	0	0	0	2.94
187	0	0	0	0	0	0	2.99
188*	0	0	0	0	0	0	3.10
189	0	0	0	0	0	0	3.03
190^	0	0	0	0	0	0	3.43
191^	0	0	0	0	0	0	3.26
192^	0	0	0	0	0	0	3.12
193	0	0	0	0	0	0	3.02
194^	0	0	0	0	0	0	3.23
195	0	0	0	0	0	0	3.64
196*	0	0	0	0	0	0	3.61
197*	0	0	0	0	0	0	3.61
198^	0	0	0	0	0	0	3.61
199	0	0	0	0	0	0	3.60
200	0	0	0	0	0	0	3.64
201	0	0	0	0	0	0	3.54
202	0	0	0	0	0	0	3.41
203	0	0	0	0	0	0	3.62
204*	0	0	0	0	0	0	3.64
205	0	0	0	1	0	0	3.50
206*	0	0	1	0	0	0	4.20
207*	0	0	0	0	0	0	3.64
208	0	0	0	0	0	0	3.57
209	0	0	0	0	0	0	3.40
210	0	0	0	0	0	0	3.45
211*	0	0	0	0	0	0	3.46
212	0	0	0	0	0	0	3.46
213*	0	0	0	0	0	0	3.50
214^	0	0	0	0	0	0	3.62
215	0	0	0	0	0	0	3.54
216^	0	0	0	0	0	0	3.62
217	0	0	0	0	0	0	3.62
218^	0	0	0	0	0	0	3.62
219^	0	0	0	0	0	0	3.62
220*	0	0	0	0	0	0	3.62
221	0	0	0	0	0	0	3.58
222	0	0	0	0	0	0	3.98
223*	0	0	0	0	0	0	3.81
224	0	0	0	0	1	0	3.82
225	0	0	1	0	0	0	4.15
226	0	0	1	0	0	0	4.03
227	0	0	0	0	0	0	3.17
228	0	0	0	0	0	0	3.17
229^	0	0	0	0	0	0	3.21
230^	0	0	0	0	0	0	3.16
231^	0	0	0	0	0	0	3.17
232*	0	0	0	0	0	0	3.99

233	0	0	1	0	0	0	3.98
234*	0	0	0	0	0	0	3.02
235^	0	0	0	0	0	0	3.97
236	0	0	1	0	0	0	4.05
237^	0	0	0	0	0	0	3.20
238*	0	0	0	0	0	0	3.03
239	0	0	0	0	0	0	3.04
240^	0	0	0	0	0	0	3.34
241^	0	0	0	0	0	0	3.62
242	0	0	0	0	0	0	3.01
243*	0	0	0	0	0	0	3.01
244	0	0	0	0	0	0	3.35
245	0	0	0	0	0	0	2.82
246^	0	0	0	0	0	1	3.82
247	0	0	0	0	0	0	3.98
248*	0	0	0	0	0	0	3.81
249	0	0	0	0	1	0	3.99
250*	0	0	0	0	1	0	3.82
251	0	0	1	0	0	0	4.10
252	0	0	1	0	0	0	4.17
253*	0	0	1	0	0	0	4.01
254*	0	0	0	0	0	0	4.01
255*	0	0	0	0	0	0	3.98
256	0	0	0	0	0	0	3.96
257^	0	0	0	0	0	0	3.97
258*	0	0	0	0	0	0	3.99
259	0	0	0	0	0	0	3.97
260^	0	0	0	0	0	0	3.97
261*	0	0	0	0	0	1	3.84
262	0	0	0	0	0	1	3.83
263^	0	0	0	0	0	0	3.34

*BMS validation set compound

^BSM test set compound

Table 6S. The $\log_{10} LC_{50}$ predicted values by Eq. 1 of the unknown set ($n = 237$) and the leverage value for each compound. The limit leverage is $h^* = 0.1805$.

ID	Smiles Code	LC_{50} (µg. mL^{-1}) pred.	Leverage (h_i)	Reference
1	CCCCCCCCCC[C@@H](O)C[C@H]1OCC[C@@H]1[C@H]1O[C@H](CC1)[C@@H](CCCCCCCCCC[C@H](O)CC1=CC(=O)O[C@@H]1C)O	1.41	0.01	Cepleanu et al. (1993)
2	CCCCCCCCCC[C@@H](O)C[C@H]1OCC[C@@H]1[C@H]1O[C@H](CC1)[C@@H](CC[C@H](CCCCCCC[C@H](O)CC1=CC(=O)OC1)O)O	1.41	0.01	Cepleanu et al. (1993)
3	CCC[C@H](CC[C@@H](O)CC[C@H](O)[C@H]1O[C@@H](CC1)[C@H]1O[C@H](CC1)[C@H](CCCCCCCCCCC[C@H](CC1=CC(=O)OC1)O)O)O	1.44	0.01	Cepleanu et al. (1993)
4	c1ccc(cc1)C#CC#CC#C	1.26	0.01	Arnason et al. (1983)
5	c1(ccc(s1)C#CC=C)C#C/C=C/C	1.26	0.01	Arnason et al. (1983)
6	c1cc(ccc1)C#CC#C/C=C\C=O	1.18	0.01	Arnason et al. (1983)
7	C#CC#CC#C/C=C/C=C/CCCOC(=O)C=C(C)C	1.20	0.01	Arnason et al. (1983)
8	c1ccc(o1)/C=C/C#CC#CC#C	1.19	0.01	Arnason et al. (1983)
9	C1=C([C@H]2[C@@H](C=C1)C(=O)c1c(O2)c2c(c(c1O)C(C=C)(C)C)OC(C=C2)(C)C)O	1.61	0.09	Ee et al. (2009)
10	c1(c2c(ccc1)[C@@H](O)C[C@@H](C2=O)C)O	0.90	0.14	Sreelatha et al. (2010)
11	Oc1c(C2=C(C(=O)c3c(C2=O)c(O)ccc3)C)ccc2c1C(=O)C=C(C2=O)C	0.26	0.24	Sreelatha et al. (2010)
12	c1c(cc2c(C(=O)C=CC2=O)c1)CC[C@]1([C@@H](CCC(=O)[C@@H]1C)C)C	1.19	0.01	Ioset et al. (2001)
13	c1c(cc2c(C(=O)C=CC2=O)c1)CC[C@]1([C@@H](CC[C@H]2[C@]1(C)O2)C)C	1.88	0.55	Ioset et al. (2001)
14	c1c(cc2c(C(=O)C=CC2=O)c1)CC[C@]1([C@@H](CC[C@H]([C@@]1(C)O)O)C)C	1.19	0.01	Ioset et al. (2001)
15	c1c(cc2c(C(=O)C=CC2=O)c1)CCC(=C(C)C)[C@@H](CCCO)C	1.08	0.01	Ioset et al. (2001)
16	O1[C@]2([C@@H](C([C@@H]1CC2)(C)C)CCc1cc2c(cc1)C(=O)C=CC2=O)C	1.14	0.01	Ioset et al. (2001)
17	O=C1C([C@@H](CCc2cc3c(cc2)C(=O)C=CC3=O)C(=C)CC1)(C)C	1.07	0.01	Ioset et al. (2001)

18	c1c(c2c(c(c1)C)C(=O)C(=O)C(=C2)C)C(C)C	1.09	0.01	Tiew et al. (2003)
19	C1(=C(C(=O)C(=C(C1=O)O)CCCCCCCCCCC)O)CC1=C(C(=O)C(=C(C1=O)O)CCCCCCCCCCCC)O	0.50	0.01	Kiprono et al. (2004)
20	c12c(cc(c([C@H]3OC[C@@H]4[C@@]3(CO[C@H]4c3c(cc4c(c3OC)OCO4)OC)OC(=O)C)c2)OC)OCO1	0.61	0.01	Perumalsamy et al. (2009)
21	c12c(ccc([C@H]3OC[C@H]4[C@@H](OC[C@H]34)c3ccc(c(OC)c3)OC)c2OC)OCO1	0.68	0.01	Cabral et al. (2009)
22	c12c(ccc([C@H]3OC[C@H]4[C@@H](OC[C@H]34)c3cc(c(c(OC)c3)OC)OC)c2OC)OCO1	0.63	0.01	Cabral et al. (2009)
23	c1c([C@@H]2O[C@@H](c3cc(c(c(OC)c3)OC)OC)CC2)cc(OC)c(c1OC)OC	0.64	0.01	Cabral et al. (2009)
24	O(c1cc(CC=C)cc(c1OC)OC)C	1.64	0.01	Marston et al. (2000)
25	O(c1cc(/C=C\C)cc(c1OC)OC)C	1.50	0.01	Marston et al. (2000)
26	O(c1cc(CCC(=O)OC)c(OC)cc1OC)C	1.62	0.01	Ioset et al. (2001)
27	O1C([C@H]1COc1c2c(occ2)cc2oc(=O)ccc12)(C)C	1.97	0.55	Marston et al. (2000)
28	o1c2c(c3c(c(OC)c2OC)C=CC(=O)C3)cc1	1.20	0.01	De Oliveira et al. (2005)
29	o1c2c(c3c(c(OC)c2OC)C2=C(C(=O)C3)[C@H]3C(=O)Oc4c([C@@H]23)c(c(c2c4occ2)OC)OC)cc1	1.10	0.09	De Oliveira et al. (2005)
30	O(c1c2c(c(c3c1ccc(=O)o3)CC=C(C)C)occ2)C	1.18	0.01	De Oliveira et al. (2005)
31	C1(=O)c2c(cc(c(c2)OC)OC)OCC1	1.51	0.01	Yenesew et al. (2006)
32	c1c(cc(c(c1)CC(=O)/C=C/c1ccccc1O)O)OC	0.72	0.01	Midiwo et al. (2005)
33	C1=Cc2c(c(c3c(c2O)C(=O)C[C@H](c2ccc(cc2)OC)O3)CC=C(C)C)OC1(C)C	1.39	0.09	Yenesew et al. (2006)
34	C1=Cc2c(c(c3c(c2O)C(=O)C[C@H](c2ccc(cc2)O)O3)CC=C(C)C)OC1(C)C	1.40	0.09	Yenesew et al. (2006)
35	o1c2c(c(=O)c(O)c1c1cc(O)c(O)cc1)c(O)cc(O)c2	0.70	0.01	Midiwo et al. (2005)
36	O1[C@H]2[C@H](c3c(OC2)cc2OCOc2c3)C(=O)c2c1c1c(OC(C=C1)(C)C)cc2	1.26	0.09	Ollis et al. (1967)
37	O1[C@@H]2[C@@](O)(c3c(OC2)cc2OCOc2c3)C(=O)c2c1c1c(OC(C=C1)(C)C)cc2	1.26	0.09	Ollis et al. (1967)
38	o1c2COc3c(cc(OC)c(OC)c3)c2c(=O)c2c1c1c(OC(C=C1)(C)C)cc2	1.25	0.09	Vasconcelos et al. (2009)
39	O=CC[C@H](CCC=C(C)C)C	1.81	0.01	Simas et al. (2004)

40	O1[C@@H]2[C@@H]3[C@@H](CCC3=C)C(=C)CC[C@H]2C(=C)C1=O	1.22	0.01	Neves et al. (1999)
41	O1[C@@H]2[C@@H]3[C@H]([C@](OC(=O)C)(CC[C@H]2C(=C)C1=O)C)CCC3=C	1.25	0.01	Neves et al. (1999)
42	O1[C@@H]2[C@@H]3[C@@H](CCC3=C)C(=C)CC[C@H]2[C@H](C1=O)C	1.43	0.01	Neves et al. (1999)
43	[C@@H]1([C@@H](CC[C@]2([C@]31O[C@]1([C@@]4([C@]3([C@@H]([C@@H]([C@]4(C(C)C)O)OC)[C@@]2(C)C1)O)C)O)O)C)O	0.29	0.20	Barreiros et al. (2007)
44	O[C@H]1CC[C@@]2(C)[C@@H]3[C@@]1(C)C[C@@H]3C[C@@]1(C)[C@@H]2CC=C2[C@@H]3CC(C)(C)CC[C@@]3(C(=O)O)CC[C@]12C	1.72	0.09	Njoku et al. (1997)
45	O1[C@@]23[C@]4([C@@H]([C@@]5([C@@H](C[C@H]4OC(=O)C)C(C(=O)C=C5)(C)C)C)CC[C@]2([C@@H](OC(=O)[C@@H]13)c1ccoc1)C)C	2.37	0.68	Gurulingappa et al. (2009)
46	O1[C@@]23[C@]4([C@@H]([C@@]5([C@@H](C[C@H]4OC(=O)C)C(C(=O)C=C5)(C)C)C)CC[C@]2([C@@H](C(=O)[C@@H]13)c1ccoc1)C)C	1.71	0.55	Gurulingappa et al. (2009)
47	O1[C@@H]2[C@@]3([C@@H]4[C@@]([C@H](OC(=O)C)C[C@H]3C(C(=O)[C@H]12)(C)C)(C1=CC(=O)O[C@@](O)([C@@]1(CC4)C)c1ccoc1)C)C	1.20	0.62	Gurulingappa et al. (2009)
48	O([C@H]1[C@@]2([C@@H]([C@@]3([C@@H](C1)C(C(=O)C=C3)(C)C)C)CC[C@@]1(C2=CC(=O)[C@H]1c1ccoc1)C)C)C(=O)C	0.93	0.01	Gurulingappa et al. (2009)
49	O[C@H]1[C@]2([C@H]([C@@]3([C@H](C1)C(C(=O)C=C3)(C)C)C)CC[C@]1(C2=CC(=O)[C@@H]1c1ccoc1)C)C	0.93	0.01	Gurulingappa et al. (2009)
50	O([C@H]1[C@@]2([C@@H]([C@@]3([C@@H](C1)C(C(=O)C=C3)(C)C)C)CC[C@@]1(C2=CC(=O)O[C@]1(O)c1ccoc1)C)C)C(=O)C	0.40	0.22	Gurulingappa et al. (2009)
51	O=C1C([C@H]2[C@]([C@@H]3[C@]([C@@H](C2)O)(C)C2=CC(=O)[C@H](c4ccoc4)[C@@]2(CC3)C)(C)[C@@H]2[C@@H]1O2)(C)C	1.74	0.55	Gurulingappa et al. (2009)
52	O1[C@@]23[C@]([C@@H](OC(=O)[C@@H]12)c1ccoc1)(CCC=C3C)C	1.83	0.55	Kiprop et al. (2005)
53	O1[C@@]23[C@]4([C@@H]([C@]56[C@@H](CC4=O)C(O[C@H]5CC(=O)OC6)(C)C)CC[C@]2([C@@H](OC(=O)[C@@H]13)c1ccoc1)C)C	2.41	0.68	Kiprop et al. (2005)
54	O1[C@@]23[C@]4([C@@H]([C@]56C(=C(O)C4=O)C(O[C@H]5CC(=O)OC6)(C)C)CC[	2.32	0.68	Kiprop et al. (2005)

	C@]2([C@H](OC(=O)[C@@H]13)c1ccoc1)C)C			
55	O1[C@@]23[C@]4([C@H](CC[C@]2([C@H](OC(=O)[C@@H]13)c1ccoc1)C)[C@@]([C@]1(O)[C@@]4(O)OC(C1=O)(C)C)(C)/C=C/CC(=O)OC)C	0.54	0.55	Kiprop et al. (2007)
56	O1[C@]2([C@]3([C@@H]([C@](O)([C@@]4(C(=O)[C@]3(OC)O)C(OC(=O)C=C4)(C)C)C)C[C@@H](OC)[C@@]2(C)C(=O)c2ccoc2)C)[C@H]1C(=O)OC	1.65	0.55	Kiprop et al. (2007)
57	O1[C@H]([C@@]2(CCC=C(C2=CC1=O)C)C)c1ccoc1	1.04	0.01	Kiprop et al. (2007)
58	O1[C@]2([C@H]3[C@]4(O)C=CO[C@@H]4O[C@@H]([C@]12[C@]1([C@H](O)[C@H]2OC[C@@]4(C(=O)OC)[C@@H]2[C@@]2([C@H](OC(=O)/C(=C/C)/C)C[C@H]4OC(=O)C)CO[C@@](O)(C(=O)OC)[C@H]12)C)C3)C	0.59	0.55	Shivakumar et al. (2011)
59	O1[C@@](CCC=C(C)C)(C=Cc2c1c(cc1c2[nH]c2c1cccc2)C)C	0.86	0.01	Ramsewak et al. (1999)
60	Oc1c(c2[nH]c3c(c2cc1)cc(c(OC)c3)C)C/C=C(/CCC=C(C)C)\C	0.85	0.01	Ramsewak et al. (1999)
61	O1[C@@](CCC=C(C)C)(C=Cc2c1c(cc1c2[nH]c2c1ccc(c2)O)C)C	0.82	0.01	Ramsewak et al. (1999)
62	O1c2c3c4c(cc2OC1)ccnc4C(=O)c1c3cccc1	1.46	0.09	Feitosa et al. (2009)
63	CO[C@H]1CC[C@@H](N(C)[C@@H]1C)/C=C/C=C/C=C/CCCC	1.51	0.01	Bandara et al. (2000)
64	O=c1n(c(=O)n(c2ncn(c12)C)C)C	1.97	0.01	Laranja et al. (2003)
65	O1c2cc(/C=C/CCCCCCC/C=C/C(=O)NCC(C)C)ccc2OC1	1.32	0.01	Siddiqui et al. (2003)
66	C(CCCCC/C=C/CCCCC)CC/C=C/C(=O)NCCCC(C)C	1.49	0.01	Siddiqui et al. (2003)
67	C(CCCCC/C=C/CCCCC)/C=C\C=C\C(=O)NCCCC(C)C	1.49	0.01	Siddiqui et al. (2004)
68	O1c2cc(/C=C/CCCCC/C=C/C=C/C(=O)NCCC(C)C)ccc2OC1	1.17	0.01	Siddiqui et al. (2004)
69	O=C(NCCC(C)C)/C=C/C=C/CCC	1.76	0.01	Siddiqui et al. (2004)
70	O=C(/C=C/C=C/CCCCC)N1CCCC1	1.69	0.01	Siddiqui et al. (2004)
71	O=C(/C=C/C=C/CCCCCCC)NCC(C)C	1.74	0.01	Siddiqui et al. (2004)
72	O=C(/C=C/C=C/CCCCCCC)N1CCCC1	1.68	0.01	Siddiqui et al. (2004)
73	O=C(NCC(C)C)/C=C/C=C/CCCCC	1.74	0.01	Siddiqui et al. (2004)

74	O=C(N1CCCC1)CCCCCCCCCCCCCCC	1.92	0.01	Siddiqui et al. (2004)
75	O=C(/C=C/CCCCCCC)N1CCCC1	1.93	0.01	Siddiqui et al. (2004)
76	C(CCCCC/C=C/CCCCC)/C=C/C=C/C(=O)NCC(C)C	1.49	0.01	Siddiqui et al. (2004)
77	O1c2cc(CCC/C=C/C=C/C(=O)NCCC(C)C)ccc2OC1	1.17	0.01	Siddiqui et al. (2004)
78	O1c2cc(/C=C/C=C/C=C/CCC(=O)N3CCCCC3)ccc2OC1	1.16	0.01	Sidiqui et al. (2002)
79	O1c2cc(/C=C/C=C/C(=O)N3CCCCC3)ccc2OC1	1.18	0.01	Sidiqui et al. (2002)
80	[C@@H]1(C=C(C(=O)O1)CCCCCCCCCC#C)C	1.72	0.01	Ratnayake et al. (2001)
81	[C@@H]1(C=C(C(=O)O1)CCCCCCCCCC=C)C	1.71	0.01	Ratnayake et al. (2001)
82	c12c(c(ccc2C)C(C)C)C=C(C(=O)C1=O)C	1.09	0.01	Balandrin et al. (1988)
83	c1(ccc(cc1OC)/C=C/C=C)OC	1.36	0.01	Bandara et al. (2005)
84	C1(=C(C(=O)C(=C(C1=O)C)C)OC)OC	1.64	0.01	Likhovidov et al. (2010)
85	c1c(c2c(c(c1C(=O)N[C@@H](Cc1ccccc1)C(=O)O)O)C(=O)O[C@@H](C2)C)Cl	0.82	0.01	Ondeyka et al. (2003)
86	O1[C@@]2([C@H](OC(=O)C)[C@@](O[C@H]([C@@H]2O)/C=C/C=C/C=C/c2oc(=O)cc(OC)c2C)([C@H]1CC)C)C	1.08	0.01	Likhovidov et al. (2010)
87	O1[C@@]2([C@@H](O)CCC(=O)C2=C(O)c2c1cc(c(c1c(O)c3C(=C4[C@](Oc3cc1C)(C(=O)OC)[C@H](OC(=O)c1ccccc1)CCC4=O)O)c2O)C)CO	1.19	0.09	Ondeyka et al. (2003)
88	O1[C@@H](Cc2c(c(O)c3c(c2)C(=O)C(=C(C2=C(OC)C(=O)c4c(c(O)c5c(C[C@H](OC5=O)C)c4)C2=O)C3=O)OC)C1=O)C	0.92	0.09	Ondeyka et al. (2003)
89	O1[C@@H](Cc2c(c(O)c3c(c2)cc(OC)c(C2=C(OC)C(=O)c4c(c(O)c5c(C[C@H](OC5=O)C)c4)C2=O)c3O)C1=O)C	0.97	0.09	J.G. Ondeyka et al (Chapter 10) (2003)
90	O[C@@H]1[C@@]([C@H]2[C@@]([C@@]3([C@@H](CC2)Cc2c3n3c4c2cc2c(c4C(=O)[C@@H]3C(=C)C)[C@@H](O)[C@H]3C2=CC(OC3(C)C)(C)C)C)(CC1)C)(C)/C=C/C=C(\C)/C(=O)O	1.44	0.09	Ondeyka et al. (1997)
91	O1[C@H](CC/C/1=C\C(=C\c1ccc([N](=O)O)cc1)\C)c1oc(OC)c(c(=O)c1C)C	0.36	0.55	Nair et al. (1995)
92	O1[C@H](C/C(=C/C(=C/C(=C/C(=C/c2ccc([N](=O)O)cc2)/C)/C)/C)/C1)c1oc(OC)c(c(=O)c1C)C	0.36	0.55	Nair et al. (1995)
93	o1c(/C(=C/C(=C/c2ccc([N](=O)O)cc2)/C)/C)	1.55	0.55	Nair et al. (1995)

	cc(OC)c(c1=O)C			
94	O[C@@H]1C[C@@H](C[C@@H](/C=C/[C@H]([C@@H](OC(=O)/C=C/C=C/C=C/C=C/C=C/C[C@@H]([C@@H]([C@H](C[C@H](C[C@@H](C[C@@H](C1)O)O)O)O)C)O)C(C)C)C)O)O	1.21	0.01	Nair et al. (1989)
95	O[C@@H]1C[C@@H](C[C@@H](/C=C/[C@H]([C@@H](OC(=O)/C=C/C=C/C=C/C=C/C=C/C[C@@H]([C@@H]([C@H](C[C@H](C[C@@H](C[C@@H](C1)O)O)O)O)C)O)[C@@H](C)CC)C)O)O	1.21	0.01	Nair et al. (1989)
96	O=C1[C@H](CCCN=C(N)N)NC(=O)[C@@H]([C@@H](C(=O)O)NC(=O)[C@H](CCCN=C(N)N)NC(=O)[C@H](C)NC(=O)C(=C)NC(=O)CC[C@H](C(=O)O)NC(=O)[C@@H]([C@@H](/C=C/C(=C/[C@@H](C)[C@H](Cc2ccccc2)OC)/C)N1)C)C	1.06	0.01	Kiviranta et al. (1992)
97	C1(=O)C=C[C@H]2CCCN12	2.22	0.01	Jizba et al. (1992)
98	C1(=O)C=C[C@@]2(CCCN12)O	2.15	0.01	Jizba et al. (1992)
99	c1(c(c2c(c(c1C)O)[C@@]1(C(=CC(=O)[C@@H](C1=O)OC(=O)C)O2)C)OC(=O)C)O	0.97	0.01	Bomfim et al. (2009)
100	o1c2c(c(=O)c(c3ccc(OC)cc3)c1)ccc(O)c2	0.76	0.01	Pluempanupat et al. 2012
101	O([C@@H]1[C@]2(C([C@@H](C1)CC2)(C)C)C)C(=O)C	1.80	0.01	Waliwitiya et al. (2009)
102	O(CC[C@H](CCC=C(C)C)C)C=O	1.88	0.01	Ali et al. 2013
103	OC(=O)/C=C(/CCC=C(C)C)\C	1.39	0.01	Ali et al. 2013
104	O[C@@]1([C@@H]2[C@@H]([C@H](CC1)C(C)C)C=C(CC2)C)C	1.66	0.01	Cheng et al. (2013)
105	O[C@]1([C@H]2C([C@H]3[C@@](C2)([C@@H](CC3)C)CC1)(C)C)C	1.83	0.01	Cheng et al. (2013)
106	OC([C@H]1C[C@H]([C@](CC1)(C)C=C)C(=C)C)(C)C	1.68	0.01	Cheng et al. (2013)
107	OC([C@H]1C[C@@H]2[C@@](CC1)(CCCC2=C)C)(C)C	1.66	0.01	Cheng et al. (2013)
108	c1c(cc(cc1)C)CC	1.99	0.01	Miles et al. (2000)
109	c1(c(cc(c(c1)C)CCC(C)C)C)C	1.74	0.01	Miles et al. (2000)
110	c1(c(CC/C=C/C(=O)C)cccc1)C(=O)C	1.30	0.01	Miles et al. (2000)
111	c12c(cccc1)C(=O)O/C/2=C/CCC	1.48	0.01	Miles et al. (2000)
112	O1[C@H]2[C@@H](c3c1cc(OC)cc3)COc1c2ccc(O)c1	0.75	0.01	Pluempanupat et al. (2012)

113	c1(c(/C=C/CCC)cccc1)C(=O)O	1.59	0.01	Miles et al. (2000)
114	c1(c(/C=C/C)cc(c(c1)OC)OC)OC	1.52	0.01	Momin and Nair (2002)
115	c1(ccc(cc1)C)[C@@H](C)CC(=O)C=C	1.36	0.01	Lee et al. (2001)
116	c1(ccc(cc1)C)/C(=C/C(=O)C=C)/C	1.18	0.01	Nair et al. (1998)
117	c1(cc(c(cc1)OC)O)CC=C	1.66	0.01	Momin et al. (2000)
118	c1(ccc(cc1)OC)/C=C\C	1.56	0.01	Kelm et al. (1997)
119	c1(ccc(c(c1)OC)OC)/C=C\C	1.52	0.01	Kelm et al. (1997)
120	c1(cc(cc(c1CC=C(C)C)OC)O)/C=C/c1ccccc1	0.80	0.01	Ioset et al. (2001)
121	c1(c(c(cc(c1CC=C(C)C)OC)O)CC=C(C)C)/C=C/c1ccccc1	0.80	0.01	Ioset et al. (2001)
122	c1(/C=C/c2ccccc2)cc(c(c(O)c1)CC=C(C)C)O	0.80	0.01	Ioset et al. (2001)
123	c1(/C=C/c2ccccc2)cc(c2c(OC([C@@H](C2)O)(C)C)c1)O	0.79	0.01	Ioset et al. (2001)
124	c1(/C=C\c2ccccc2)cc(OC)cc(OC)c1	0.82	0.01	Ioset et al. (2001)
125	c1(/C=C/c2ccccc2)cc(O)c(c(c1CC=C(C)C)OC)CC=C(C)C	0.79	0.01	Ioset et al. (2001)
126	c1(/C=C/c2ccccc2)c(c(O)cc2OC([C@H](Cc12)O)(C)C)CC=C(C)C	0.79	0.01	Ioset et al. (2001)
127	C1(=O)O[C@@H](C=C1CCCCCCCCCCC#C)C	1.72	0.01	Ratnayake et al. (2001)
128	C1(=O)O[C@@H](C=C1CCCCCCCCCCC=C)C	1.71	0.01	Ratnayake et al. (2001)
129	C1(=O)O[C@H](CCCC)[C@H]2CCCC=C12	2.04	0.01	Momin and Nair (2001)
130	C(CC=C(C)C)/C(=C/C=O)/C	1.41	0.01	Kelm et al. (1997)
131	O=C/C=C(\CCC=C(C)C)/C	1.41	0.01	Kelm et al. (1997)
132	O=Cc1ccc(C(C)C)cc1	1.67	0.01	Zahran and Abdelgaleil (2010)
133	[C@@]12([C@@H](C[C@@H](CC1)C(=C)C)C(=C)CCC2)C	1.44	0.01	Momin et al. (2000)
134	C1C[C@@H]2[C@H](OC(=O)C2=C)/C=C(\CC/C=C/1\C)/C	1.30	0.01	Lee et al. (1971)
135	C1C[C@@H]2[C@H](OC(=O)C2=C)[C@@H]2[C@](CC/C=C/1\C)(C)O2	2.08	0.55	Lee et al. (1971)
136	C12=CC[C@H]3[C@H](CC[C@@]4([C@@H]3CC[C@H]4[C@@H](C)CC[C@@H](C(C)C)CC)C)[C@]1(CC[C@H](C2)O[C@@H]1[C@H]([C@H]([C@H]([C@H](O1)O)O)O)	0.75	0.22	Amin et al. (2012)

	O)C			
137	C#CC#CCC/C=C\C=C\C(=O)NCC(C)C	1.35	0.01	Clifford et al. (2002)
138	C#CC#CCC/C=C\C=C\C(=O)NC[C@H](C)CC	1.35	0.01	Clifford et al. (2002)
139	C(#CC#CCC/C=C\C=C\C(=O)NC[C@H](C)CC)C	1.29	0.01	Clifford et al. (2002)
140	CO[C@@H]1CC[C@@H](N(C)[C@H]1C)/C=C/C=C/C=C/CCCC	1.51	0.01	Bandara et al. (2000)
141	N1([C@@H](CCCC1)C)C(=O)CC1CCCCC1	1.81	0.01	Pridgeon et al. (2007)
142	N1([C@@H](CCCC1)C)C(=O)CCCCCCCCC	1.96	0.01	Pridgeon et al. (2007)
143	N1([C@@H](CCCC1)C)C(=O)CCCCCCCCCCCC	1.94	0.01	Pridgeon et al. (2007)
144	N1([C@@H](CCCC1)C)C(=O)CCCCCC	1.99	0.01	Pridgeon et al. (2007)
145	N1([C@@H](CCCC1)C)C(=O)CCC1CCCCC1	1.75	0.01	Pridgeon et al. (2007)
146	N1(CCCC[C@@H]1C)C(=O)[C@H]1CC[C@H](CC1)C	1.83	0.01	Pridgeon et al. (2007)
147	N1(CCC[C@@H](C1)C)C(=O)CC1(CCCCC1)C	1.75	0.01	Pridgeon et al. (2007)
148	N1(CCC[C@@H](C1)C)C(=O)CCC1CCCCC1	1.75	0.01	Pridgeon et al. (2007)
149	N1(CCC[C@@H](C1)C)C(=O)CCCCCC	1.94	0.01	Pridgeon et al. (2007)
150	N1(CCC[C@@H](C1)C)C(=O)CC1CCCCC1	1.80	0.01	Pridgeon et al. (2007)
151	N1(CCC(CC1)C)C(=O)CCCCCCCCC	1.91	0.01	Pridgeon et al. (2007)
152	N1(CCC(CC1)C)C(=O)CCCC1CCCCC1	1.71	0.01	Pridgeon et al. (2007)
153	N1(CCC(CC1)C)C(=O)C1CCCCC1	1.84	0.01	Pridgeon et al. (2007)
154	N1(CCC(CC1)C)C(=O)CCCCCCCCCCC	1.91	0.01	Pridgeon et al. (2007)
155	N1(CCC(CC1)C)C(=O)CCC1CCCCC1	1.75	0.01	Pridgeon et al. (2007)
156	N1(CCCC[C@@H]1CC)C(=O)C1CCCCC1	1.84	0.01	Pridgeon et al. (2007)
157	N1(CCCC[C@@H]1CC)C(=O)CCC1CCCCC1	1.73	0.01	Pridgeon et al. (2007)
158	N1(CCCC[C@@H]1CC)C(=O)CC	2.04	0.01	Pridgeon et al. (2007)
159	N1(CCCC[C@@H]1CC)C(=O)CCC1CCCC1	1.72	0.01	Pridgeon et al. (2007)

160	N1(CCCC[C@@H]1CC)C(=O)CCCCCCCC	1.94	0.01	Pridgeon et al. (2007)
161	N1(CCC[C@@H](C1)Cc1ccccc1)C(=O)CCCCCCC	1.58	0.01	Pridgeon et al. (2007)
162	N1(CCC(CC1)Cc1ccccc1)C(=O)CC1CCCCC1	1.51	0.01	Pridgeon et al. (2007)
163	N1(CCC[C@@H](C1)Cc1ccccc1)C(=O)CCCCCCCC=C	1.58	0.01	Pridgeon et al. (2007)
164	N1(CCC(CC1)Cc1ccccc1)C(=O)CCC1CCCCC1	1.55	0.01	Pridgeon et al. (2007)
165	N1(CCCC[C@@H]1CC)C(=O)CCCCCCCCC=C	1.71	0.01	Pridgeon et al. (2007)
166	N1(CCCC[C@@H]1C)C(=O)CCCCCCCCC=C	1.71	0.01	Pridgeon et al. (2007)
167	N1(CCCC[C@@H]1Cc1ccccc1)C(=O)CCCCCCCC=C	1.57	0.01	Pridgeon et al. (2007)
168	N1(CCC[C@@H](C1)CC)C(=O)CCCCCCCCC=C	1.71	0.01	Pridgeon et al. (2007)
169	N1(CCC[C@@H](C1)C)C(=O)CCCCCCCCC=C	1.71	0.01	Pridgeon et al. (2007) Pridgeon et al. (2007)
170	N1(CCC[C@@H](C1)Cc1ccccc1)C(=O)CCCCCCCC=C	1.58	0.01	Pridgeon et al. (2007)
171	N1(CCC(CC1)CC)C(=O)CCCCCCCCC=C	1.71	0.01	Pridgeon et al. (2007)
172	N1(CCC(CC1)C)C(=O)CCCCCCCCC=C	1.71	0.01	Pridgeon et al. (2007)
173	N1(CCC(Cc2ccccc2)CC1)C(=O)CCCCCCCCC=C	1.59	0.01	Pridgeon et al. (2007)
174	C1=CC(=O)c2c(C1=O)cc(cc2)CC[C@H]1C(CCC(=O)[C@@H]1C)(C)C	1.21	0.01	Ioset et al. (2000)
175	C1=CC(=O)c2c(C1=O)cc(cc2)CCC(=C(C)C)[C@H](C)CCCO	1.08	0.01	Ioset et al. (2000)
176	C1=CC(=O)c2c(C1=O)ccc(c2)CC[C@@H]1[C@]2(C)CC[C@H](C1(C)C)O2	1.14	0.01	Ioset et al. (2000)
177	C1=CC(=O)c2c(C1=O)ccc(c2)CC[C@H]1C(=C)CCC(=O)C1(C)C	1.07	0.01	Ioset et al. (2000)
178	C1=CC(=O)c2c(C1=O)ccc(c2)CC[C@]1([C@H](CC[C@@H]2[C@]1(C)O2)C)C	1.88	0.55	Ioset et al. (2000)
179	C1=CC(=O)c2c(C1=O)ccc(c2)CC[C@]1([C@H](CC[C@H](O)[C@@]1(C)O)C)C	1.19	0.01	Ioset et al. (2000)
180	C1=C(C(=O)c2c(C1=O)c(ccc2)O)C	0.55	0.14	Maniafu et al. (2009)
181	C1(=O)c2c(C(=O)C(=C1OC)C)cccc2O	0.49	0.14	Sreelatha et al. (2010)
182	c12c(oc(=O)c(c2C)Br)c(c(c(c1)Br)OC(=O)C)Br	0.74	0.01	Deshmukh et al. (2008)
183	O1[C@@H]([C@H](c2c1ccc(c2)/C=C/C)C)c	0.78	0.01	Chauret et al.

	1ccc(O)cc1			(1996)
184	o1c(c(c2c1ccc(c2)/C=C/C)C)c1cc(OC)c(O)cc1	0.73	0.01	Chauret et al. (1996)
185	O1[C@@H]([C@H](c2c1ccc(c2)C=O)C)c1ccc(O)cc1	0.79	0.01	Chauret et al. (1996)
186	o1c(c(c2c1ccc(c2)/C=C/C)C)c1ccc(O)cc1	0.78	0.01	Chauret et al. (1996)
187	c1(c(c(c2c(c1)OCO2)OC)[C@H]1OC[C@@]2([C@@H]1CO[C@@H]2c1c(cc2c(c1OC)OCO2)OC)OC(=O)C)OC	0.56	0.01	Park et al. (2005)
188	c1(c(c(c2c(c1)OCO2)OC)[C@H]1OC[C@]2([C@H](OC[C@H]12)c1c(cc2c(OCO2)c1)OC)OC(=O)C)OC	0.61	0.01	Park et al. (2005)
189	O1[C@]23[C@]4([C@H]([C@]56C(=C(O)C4=O)C(O[C@H]5CC(=O)OC6)(C)C)CC[C@@]2([C@H](OC(=O)[C@H]13)c1ccoc1)C)C	2.32	0.68	Liu et al. (2012)
190	O=c1n(CC[C@@]2(c3c(NC2=O)cccc3)O)c(=O)n(c2c1cccc2)C	0.69	0.01	Liu et al. (2012)
191	O=c1n2CCc3c([nH]c4c3cccc4)c2nc2c1cccc2	1.34	0.09	Liu et al. (2012)
192	O=C1N2[C@H](N(c3c1cccc3)C)c1[nH]c3c(c1CC2)cccc3	1.34	0.09	Liu et al. (2012)
193	O(C(=O)CCCCCCC/C=C\C/C=C\CCCCC)CC	1.64	0.01	Perumalsamy et al. (2015)
194	O(C(=O)CCCCCCC/C=C\CCCCCCCC)C	1.65	0.01	Perumalsamy et al. (2015)
195	O(C(=O)CCCCCCC/C=C\CCCCCCCC)CC	1.64	0.01	Perumalsamy et al. (2015)
196	O(C(=O)CCCCCCC/C=C\C/C=C\CCCCC)C	1.65	0.01	Perumalsamy et al. (2015)
197	O(C(=O)CCCCCCC/C=C/C/C=C\CCCCC)CC	1.64	0.01	Perumalsamy et al. (2015)
198	c12c(c(c(cc1CCC(C)C)O)CCC(CCC=C(C)C)C)oc1c(c2=O)c(cc(c1O)OC)O	0.71	0.01	Ee et al. (2009)
199	c12c(c(c(cc1)O)O)oc1c(c2=O)c(c(c2c1C=CC(O2)(C)C)C(C=C)(C)C)O	1.31	0.09	Ee et al. (2009)
200	c12c(cc3c(c1)C(=CC(O3)(C)C)O)oc1c(c2=O)c2c(c(c1)O)OC(C=C2)(C)C	1.26	0.09	Ee et al. (2009)
201	c12c(cc(c(c1CC=C(C)C)OC)O)oc1c(c2=O)c(c(c(c1)OC)CC=C(C)C)O	0.68	0.01	Ee et al. (2009)
202	c12c(c(c(cc1OC)O)CC=C(C)C)oc1c(c2=O)cc(cc1)O	0.78	0.01	Ee et al. (2009)
203	c12c(ccc(c1)CCC1C(=C)CCC(C1(C)C)OC(=O)C=C(C)C)C(=O)C1C(C2=O)O1	1.90	0.55	Ioset et al. (2001)
204	C1OC(C2C1C(OC2)c1cc(c(c(c1)OC)OC)OC)c1cc(c(c(c1)OC)OC)OC	0.64	0.01	Cabral et al. (2009)
205	c12c(cc3c(c1OC[C@H](O)C(C)(C)O)cco3)oc(=O)cc2	1.18	0.01	Marston et al. (1995)

206	c1(ccccc1)/C=C/c1c(c(cc(c1)O)OC)CC=C(C)C	0.80	0.01	Ioset et al. (2001)
207	c1(ccccc1)/C=C/c1cc(c2c(c1)OC([C@H](C2)O)(C)C)O	0.79	0.01	Ioset et al. (2001)
208	c1(ccccc1)/C=C/c1cc(c(c(c1CC=C(C)C)OC)CC=C(C)C)O	0.79	0.01	Ioset et al. (2001)
209	c1(ccccc1)/C=C/c1c(c(cc2c1C[C@@H](C(O2)(C)C)O)O)CC=C(C)C	0.79	0.01	Ioset et al. (2001)
210	c1c(cccc1)/C=C\c1c(c(cc2c1C[C@@H](O2)C(C)(O)C)C)CC=C(C)C	0.79	0.01	Ioset et al. (2001)
211	c12c(ccc3c1C=CC(O3)(C)C)c(=O)c1c(o2)COc2c1cc(c(c2)OC)OC	1.25	0.09	Ollis et al. (1967)
212	c12c(c(c3c(c1)OC(C=C3)(C)C)O)c(=O)c1c(o2)COc2c1cc(c(c2)OC)OC	1.24	0.09	Vasconcelos et al. (2009)
213	C\1(=C/CC/C(=C/[C@@H]2[C@@H]([C@H](C1)OC(=O)C)C(=C)C(=O)O2)/COC(=O)C)/C	1.28	0.01	Neves et al. (1999)
214	C\1(=C/CC/C(=C/[C@@H]2[C@@H]([C@H](C1)OC(=O)C)C(=C)C(=O)O2)/CO)/C	1.29	0.01	Neves et al. (1999)
215	[C@@]12([C@H](C([C@@H](CC1)O)(C)C)CCC(=C)[C@@H]2CC[C@H](CC(=O)O)C)C	1.55	0.01	Geris et al. (2009)
216	[C@@]12([C@H]([C@@](CCC1)(C)C(=O)O)CCC(=C)[C@@H]2CC/C(=C/C(=O)O)/C)C	1.19	0.01	Geris et al. (2009)
217	[C@]12([C@H](C([C@@H](C=C1)OC(=O)C)(C)C)C[C@@H]([C@]1([C@@H]2CC[C@H]2[C@@]31[C@@H](C(=O)O[C@@H]2c1ccoc1)O3)C)OC(=O)C)C	2.41	0.68	Gurulingappa et al. (2009)
218	[C@]12([C@H](C([C@@H](C=C1)O)(C)C)C[C@@H]([C@]1([C@@H]2CC[C@H]2[C@@]31[C@@H](C(=O)O[C@@H]2c1ccoc1)O3)C)OC(=O)C)C	2.41	0.68	Gurulingappa et al. (2009)
219	C1=CCN(CC1)C(=O)CCCCC/C=C/c1ccc2c(c1)OCO2	1.42	0.01	Sidiqui et al. (2002)
220	C(C)C.c12c(c(ccc1)O)C(=O)C(CC2O)C	0.90	0.14	Sidiqui et al. (2002)
221	O[C@@]1(C(=O)[C@@]2([C@]3(C)[C@H](C[C@](C1=O)(C)C2=C)[C@]1(CCC(=O)C(C)(C)[C@@H]1CC3)C)C(=O)OC)C	1.77	0.09	Santos et al. (2003)
222	O(C)C(=O)[C@@]12C(=O)[C@@]([C@](C)(C[C@H]3[C@]1(C)CC[C@H]1C(C(=O)CC[C@]31C)(C)C)C2=C)(C(=O)C)O	1.13	0.01	Santos et al. (2003)
223	O[C@@]1(C(=O)[C@@]2([C@]3(C)[C@H](C[C@](C1=O)(C)C2=C)[C@]1(C=CC(=O)OC(C)([C@@H]1CC3)C)C)C(=O)OC)C	1.06	0.01	Santos et al. (2003)
224	c1(c(c(c(cc1C)O)C=O)O)C(=O)Oc1c(c(c(c(c1)C)C(=O)OC)O)C	0.57	0.01	Vinayaka et al. (2010)

225	C(=O)/C=C(/CCC=C(C)C)\C	1.41	0.01	Ali et al. (2013)
226	C1(C/C=C(/CC/C=C(/C/C=C/1)\C)\C)(C)C	1.39	0.01	Park et al. (2011)
227	c1(cc(c(c(c1)OC)OC)OC)/C=C/C(=O)N1CCC=CC1=O	1.19	0.01	Yang et al. (2002)
228	N1(CCCCC1)C(=O)/C=C/C=C/C=C/c1cc2c(cc1)OCO2	1.11	0.01	Yang et al. (2002)
229	c12c(c(ccc1)O)oc1c(c2=O)c(c(c2c1C=CC(O2)(C)C)C(C=C)(C)C)O	0.65	0.24	Ee et al. (2009)
230	C(C)C.c12c(c(ccc1)O)C(=O)C(CC2O)C	0.90	0.14	Sreelatha et al. (2010)
231	C(C)C.c12c(c(ccc1)O)C(=O)C(C(C2=O)C)c1c(c2c(cc1)C(=O)C(=CC2=O)C)O	0.49	0.24	Sreelatha et al. (2010)
232	C(C)C.c12c(cccc1)C[C@@H](CC2=O)c1ccc(cc1)OCc1ccc(cc1)C(F)(F)F	1.37	0.09	Jung and Moon (2011)
233	C(C)C.c12c(cccc1)C[C@@H](C[C@@H]2O)c1ccc(cc1)OCc1ccc(cc1)C(F)(F)F	1.37	0.09	Jung and Moon (2011)
234	c12c(cccc1)OC(=O)[C@H]([C@@H]2O)[C@@H]1Cc2c(CC1)cccc2	1.37	0.09	Jung and Moon (2011)
235	c12c(cccc1)OC(=O)[C@H]([C@@H]2O)[C@H](Cc1ccccc1)C[C@@H](c1ccc(cc1)c1ccc(cc1)Br)O	1.34	0.09	Jung and Moon (2011)
236	c12c(cc(cc1O)OC)oc(c(c2=O)OC)c1cc(c(cc1)OC)OC	0.69	0.01	Jung and Moon (2011)
237	c12c(cc(cc1OC(=O)C)OC)oc(c(c2=O)OC)c1cc(c(cc1)OC)OC	0.69	0.01	Jung and Moon (2011)

The Replacement Method (RM) procedure

The procedure of the RM technique is as follows: choose d descriptors $\{X_1, X_2, ..., X_d\}$ at random and do a linear regression. Choose one of the descriptors of this set called X_i and replaced it by each of the D descriptors of the set (except itself), keeping the best resulting set.[21] Since one can start replacing any of the d descriptors in the initial model then a regression equation with d variables has d possible paths to achieve the final result; one example of the choice above will develop into path i. Next choose the variable with the greatest relative error in its coefficient (except the one replaced in the previous step) and replace it with all the D descriptors (except itself) keeping again the best set. Replace the entire remaining variable having the greatest relative error in the coefficient and repeat the whole process. The process will be repeated as many times as needed until the set of descriptors remains unchanged. At the end, we have the best model for the path *i*. Proceed in exactly the same way for all possible paths $i = 1, 2, ..., d$ compare the resulting models and keep the best one. Our numerical experiments show that in this way one obtains a model almost as good as the best one with fewer $D!/(D-d)!d!$ linear regressions when this combinatorial number is large.[30]

Data set partitioning

Several standard techniques for designing a rational partition of a data set are currently available in the literature, namely the traditional procedures such as principal compounds analysis (PCA), discriminant analysis (DA), cluster analysis (CA), Kennard-Stone method and D-optimal design, which offer different possibilities and objectives. The CA is proposed for many different application fields and is divided into two methods: hierarchical and partition;

the latter consists of relocation cases by moving them for one cluster to another, starting from an initial cluster, and requires that the cluster numbers shall be defined by the user. For obtaining the global optimality of partition-base clustering many enumeration processes of all possible partitions are needed; the most common algorithm uses a squared error criterion is called k-means algorithm; this method for partitioning the data into k-clusters, where each cluster is determined by its centroid or centre point (the centroid of each cluster is calculated as the mean of all the instances belonging to that cluster). The initialization of the algorithm is based on an initial set of cluster centroids chosen by a heuristic process, where each iteration for each instance is assigned to its nearest cluster centroid depending on the Euclidean distance between two of them, and then the cluster centroids are recalculated.[32-34]

The balanced subsets method (BSM) procedure

The procedure applied to the median lethal concentration (LC_{50}) data set includes the following steps:

A. Prepare a matrix (C) that includes the experimental property and the 18,326 molecular descriptors selected. Now the size of C is $263 \times 26,775$.

B. Remove the linearly dependent variables from the previous matrix and now the new size is $263 \times 10,604$.

C. Standardize C for centring and scaling its matrix elements. This is done for discerning the matrix elements better.

D. Create N_{train}^{0} clusters with the 60 compounds through the k-MCA method; the C matrix is used together with the Euclidean metrics and 90 runs for the numerical optimization algorithm of k-MCA in order to achieve the best solution. This computes N_{train}^{0} cluster centroid location, each centroid of $1 \times 10,604$ size; $N_{train}^{0} = N_{train} - N_{\min \max}$ where N_{train}^{0} is the number of compounds in the training set and $N_{\min \max}$ is the number of compounds that have minimum or maximum values of the experimental property LC_{50} .

E. The training set is (N_{train}) designed by including one compound per cluster, which is the compound that is close to the centroid in each cluster. It also includes the $N_{\min \max}$ compounds.

F. Create (N_{val}) clusters with the remaining $263 - N_{train}$ compounds through the k-MCA method, in the same numerical conditions as described previously. This computes N_{train} cluster centroid locations.

G. The validation set is designed by including one compound per cluster, which is the compound that in nearer to the centroid in each cluster.

H. Finally, the test set (N_{test}) includes the remaining $263 - N_{train} - N_{val}$ compounds.