

## 基于机器学习方法的激素敏感脂肪酶抑制剂活性预测

吕 巍<sup>1</sup>      薛 英<sup>1,2,\*</sup>

<sup>1</sup> 四川大学化学学院, 教育部绿色化学与技术重点实验室, 成都 610064;

<sup>2</sup> 四川大学生物治疗国家重点实验室, 成都 610041)

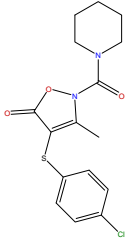
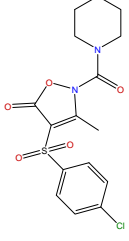
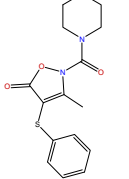
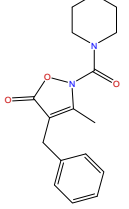
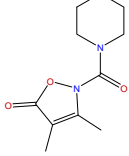
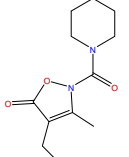
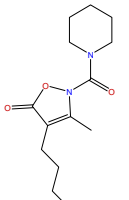
## Activity Prediction of Hormone-Sensitive Lipase Inhibitors Based on Machine Learning Methods

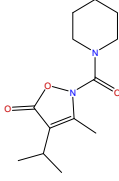
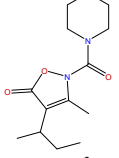
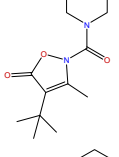
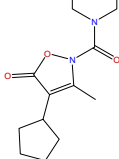
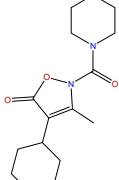
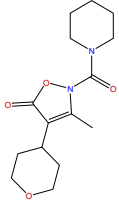
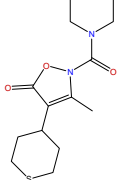
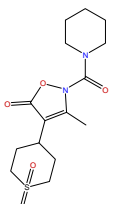
LÜ Wei<sup>1</sup>      XUE Ying<sup>1,2,\*</sup>

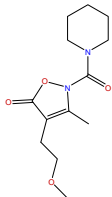
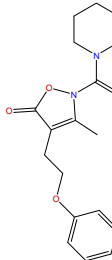
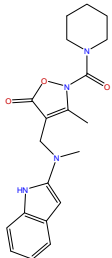
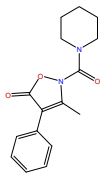
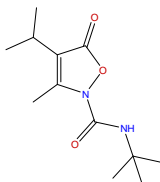
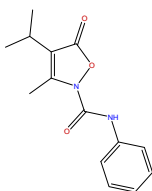
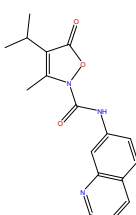
(<sup>1</sup>Key Laboratory of Green Chemistry and Technology, Ministry of Education, College of Chemistry, Sichuan University, Chengdu 610064, P. R. China; <sup>2</sup>State Key Laboratory of Biotherapy, Sichuan University, Chengdu 610041, P. R. China)

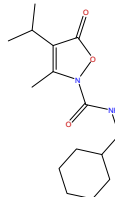
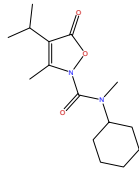
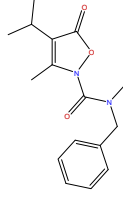
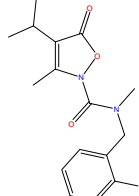
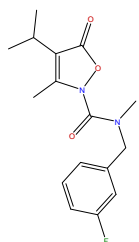
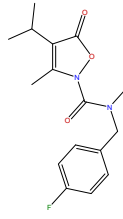
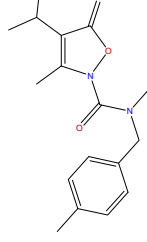
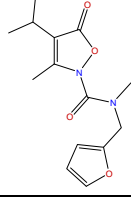
\*Corresponding author. Email: yxue@scu.edu.cn; Tel: +86-28-85418330.

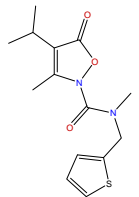
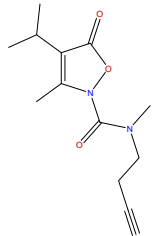
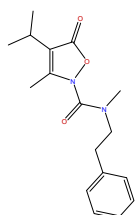
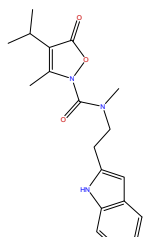
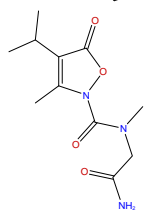
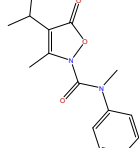
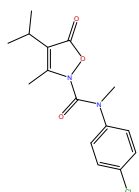
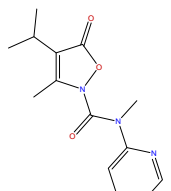
**Table S1** All Molecules Used in this Article

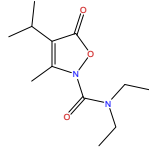
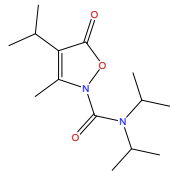
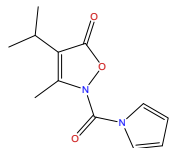
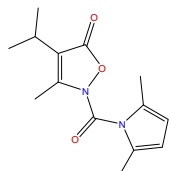
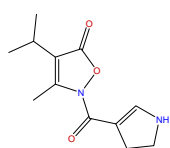
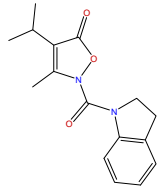
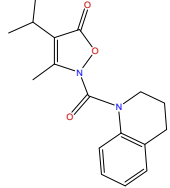
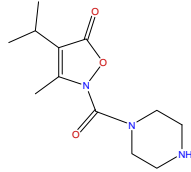
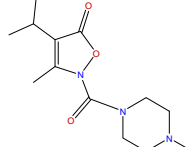
ID	Molecule Structure	ID in Ref.	IC50 (nM)	Reference
1		1	6	1
2		2	35	1
3		3	6	1
4		4	5	1
5		5	13	1
6		6	18	1
7		7	7	1

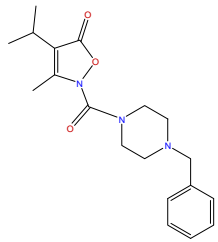
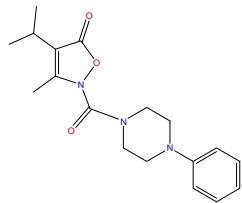
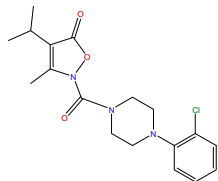
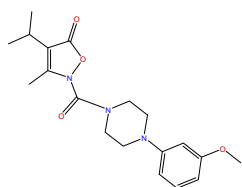
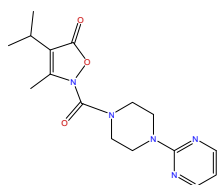
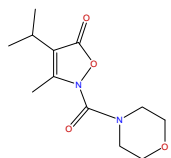
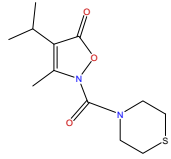
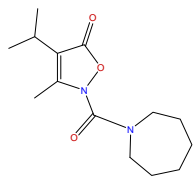
8		8	2	1
9		9	5	1
10		10	14	1
11		11	6	1
12		12	11	1
13		13	71	1
14		14	11	1
15		15	55	1

16		16	10	1
17		17	4	1
18		18	17	1
19		19	14	1
20		20	>1000	1
21		21	>1000	1
22		22	NA	1

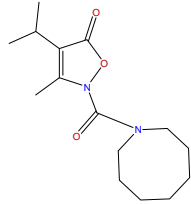
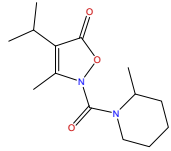
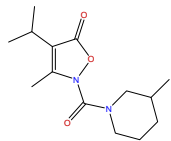
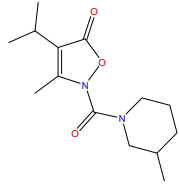
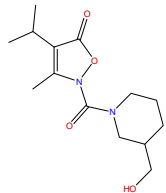
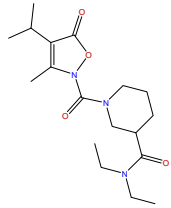
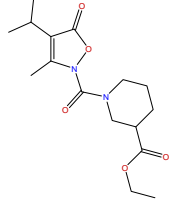
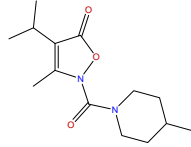
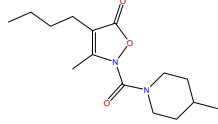
23		23	72	1
24		24	9	1
25		25	5	1
26		26	4	1
27		27	5	1
28		28	111	1
29		29	408	1
30		30	4	1

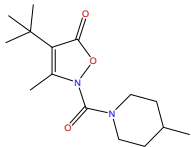
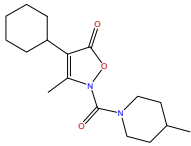
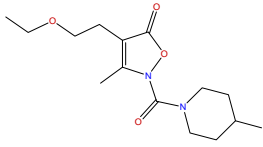
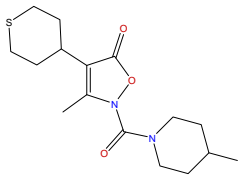
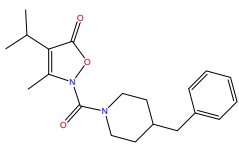
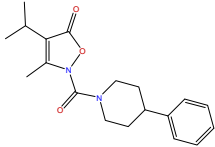
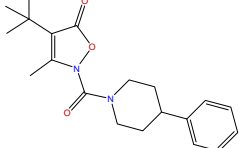
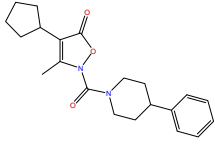
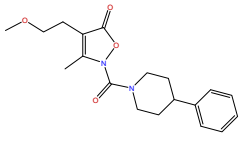
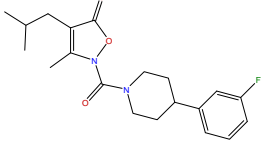
31		31	2	1
32		32	>1000	1
33		33	180	1
34		34	NA	1
35		35	>1000	1
36		36	7	1
37		37	17	1
38		38	240	1

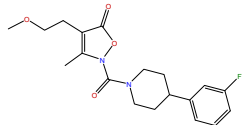
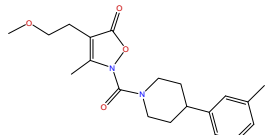
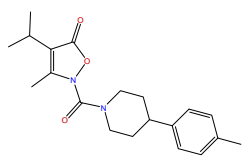
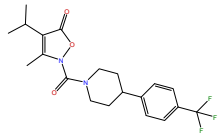
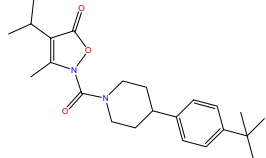
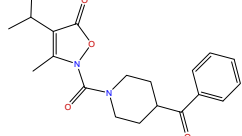
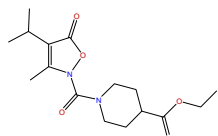
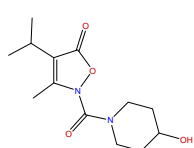
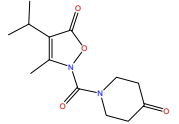
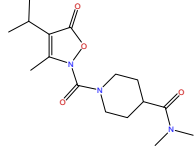
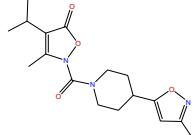
39		39	>1000	1
40		40	NA	1
41		41	15	1
42		42	NA	1
43		43	72	1
44		44	>1000	1
45		45	60	1
46		46	>1000	1
47		47	314	1

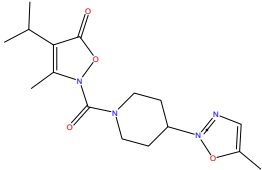
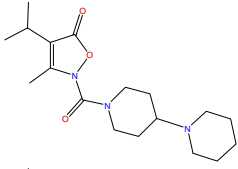
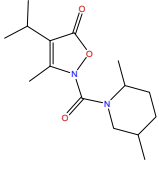
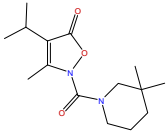
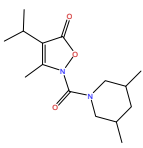
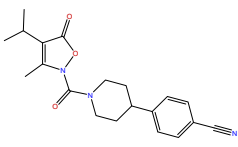
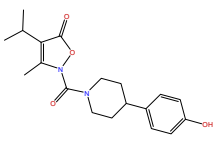
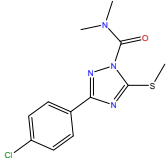
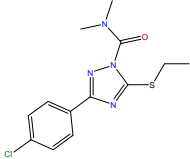
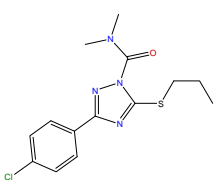
48		48	65	1
49		49	6	1
50		50	30	1
51		51	1	1
52		52	67	1
53		53	52	1
54		54	8	1
55		55	62	1



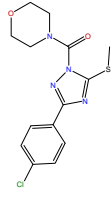
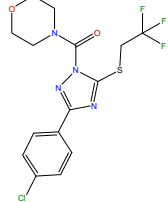
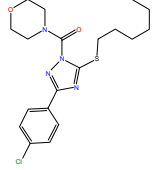
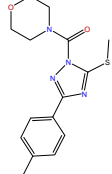
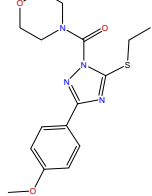
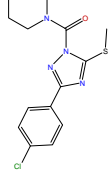
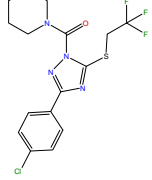
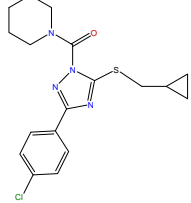
56		56	900	1
57		57	93	1
58		58	20	1
59		59	3	1
60		60	180	1
61		61	>1000	1
62		62	28	1
63		63	6	1
64		64	1	1

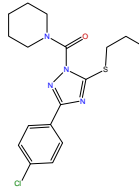
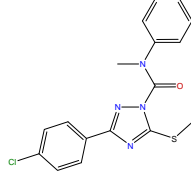
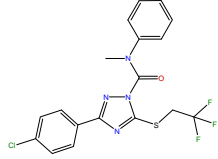
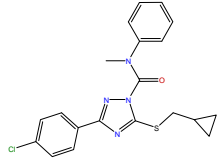
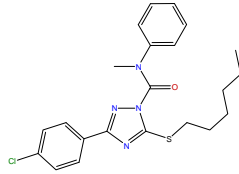
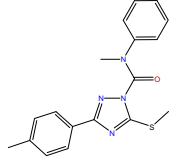
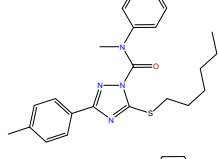
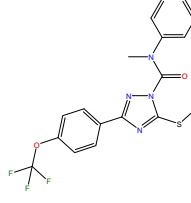
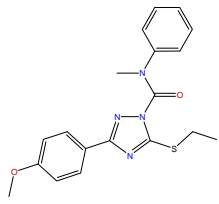
65		65	3	1
66		66	4	1
67		67	3	1
68		68	6	1
69		69	102	1
70		70	12	1
71		71	10	1
72		72	24	1
73		73	11	1
74		74	12	1

75		75	20	1
76		76	23	1
77		77	37	1
78		78	160	1
79		79	480	1
80		80	250	1
81		81	6	1
82		82	166	1
83		83	200	1
84		84	>1000	1
85		85	28	1

86		86	>1000	1
87		87	NA	1
88		88	460	1
89		89	7	1
90		90	326	1
91		91	>1000	1
92		92	>1000	1
93		8	20	2
94		2a	210	2
95		2b	60	2

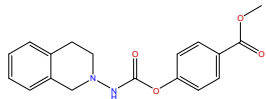
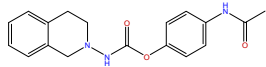
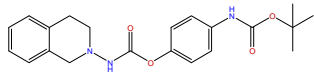
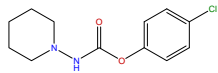
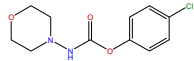
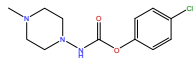
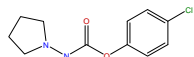
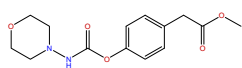
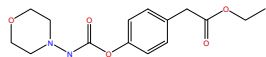
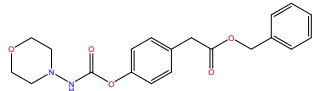
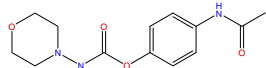
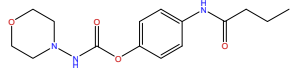
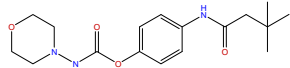
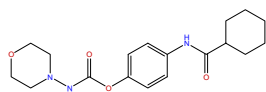
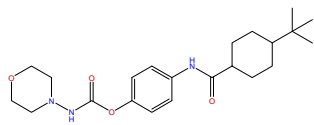
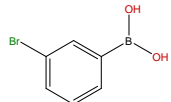
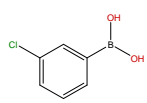
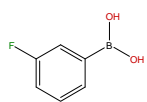
96		2c	30	2
97		2d	600	2
98		2e	50000	2
99		2f	100	2
100		2g	900	2
101		2h	100	2
102		2i	1500	2
103		2j	2000	2
104		2k	50000	2

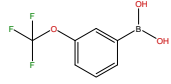
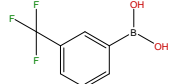
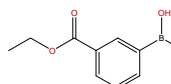
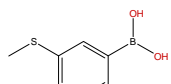
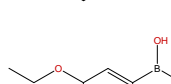
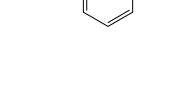
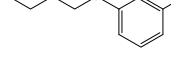
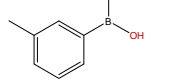
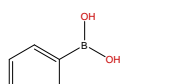
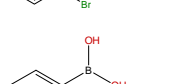
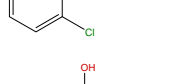
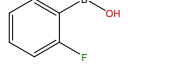
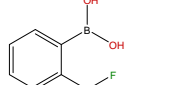
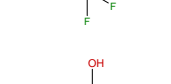
105		3a	2000	2
106		3b	9000	2
107		3c	50000	2
108		3d	4000	2
109		3e	9000	2
110		3f	100	2
111		3g	400	2
112		3h	200	2

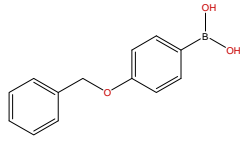
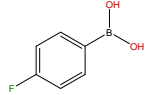
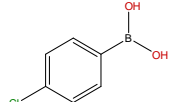
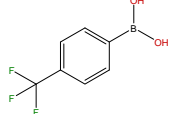
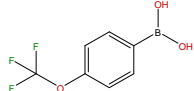
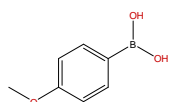
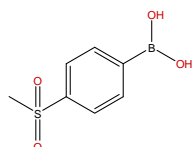
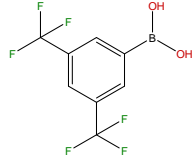
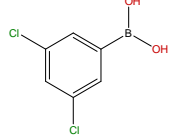
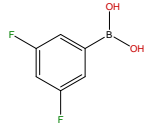
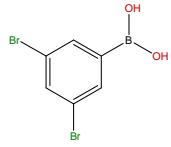
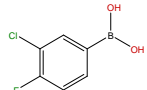
113		3i	50000	2
114		4a	160	2
115		4b	170	2
116		4c	62	2
117		4d	400	2
118		4e	100	2
119		4f	700	2
120		4g	250	2
121		4h	300	2

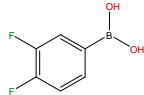
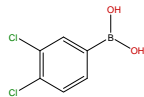
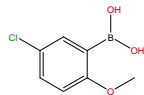
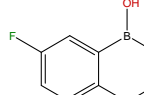
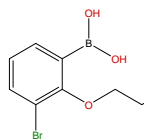
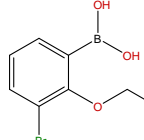
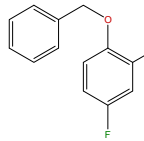
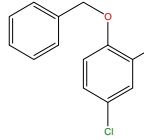
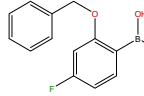
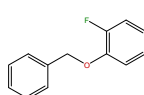
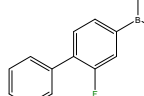
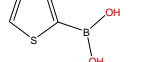
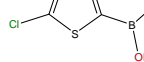
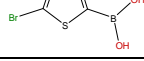
122		5	1100	2
123		6	23	2
124		14	NA	3
125		15	NA	3
126		16	NA	3
127		17	205	3
128		18	6400	3
129		19	360	3
130		20	NA	3
131		21	19	3
132		22	35	3
133		23	32	3
134		24	NA	3
135		25	NA	3
136		26	97	3

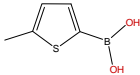
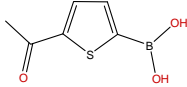
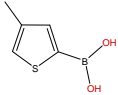
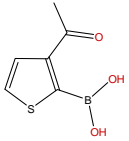
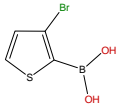
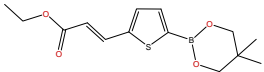
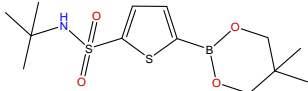
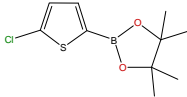
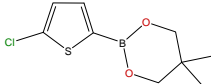
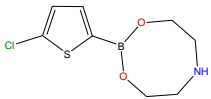
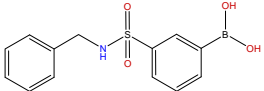
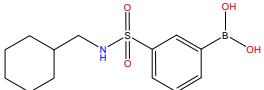
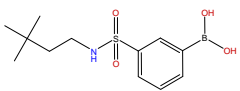
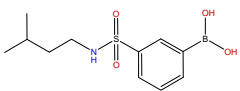


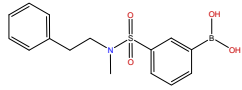
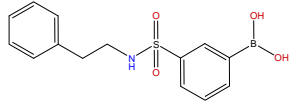
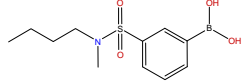
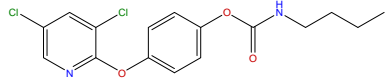
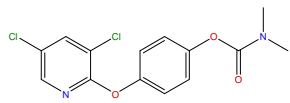
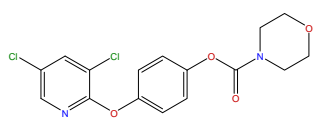
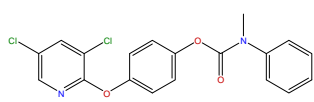
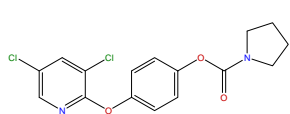
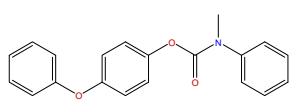
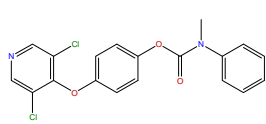
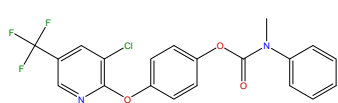
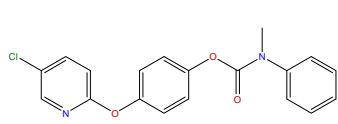
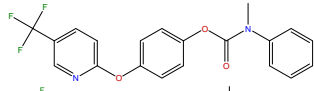
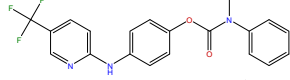
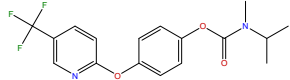
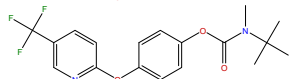
137		27	9952	3
138		28	NA	3
139		29	3	3
140		30	50	3
141		31	184	3
142		32	515	3
143		33	24	3
144		34	716	3
145		35	138	3
146		36	5	3
147		37	NA	3
148		38	3200	3
149		39	47	3
150		40	10	3
151		41	1	3
152		7a	5900	4
153		7b	4400	4
154		7c	5900	4

155		7d	850	4
156		7e	1000	4
157		7f	50000	4
158		7g	4600	4
159		7h	50000	4
160		7i	21000	4
161		7g	50000	4
162		8a	50000	4
163		8b	50000	4
164		8c	50000	4
165		8d	50000	4
166		8e	50000	4
167		8f	3400	4
168		8g	1100	4

169		9a	50000	4
170		9b	50000	4
171		9c	1300	4
172		9d	10000	4
173		9e	50000	4
174		9f	50000	4
175		9g	50000	4
176		10a	50000	4
177		10b	10000	4
178		10c	2200	4
179		10d	50000	4
180		10e	2600	4

181		10f	6000	4
182		10g	50000	4
183		10h	28000	4
184		10i	10000	4
185		10j	50000	4
186		10k	50000	4
187		10l	140	4
188		10m	17	4
189		10n	760	4
190		10o	50000	4
191		10p	50000	4
192		11a	50000	4
193		11b	710	4
194		11c	350	4

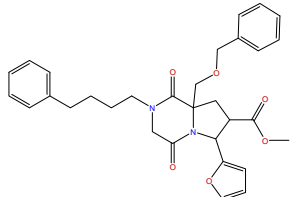
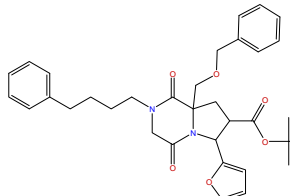
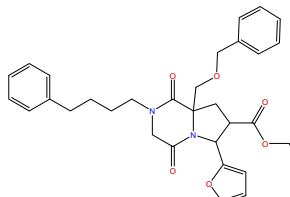
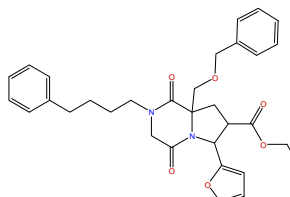
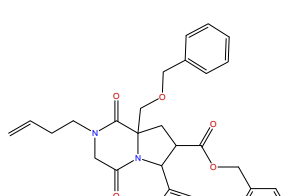
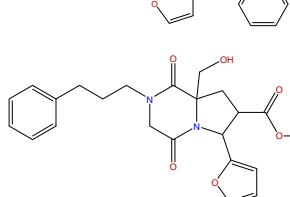
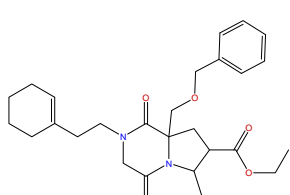
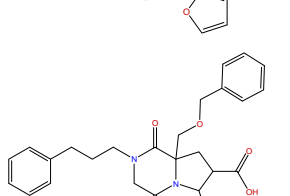
195		11d	50000	4
196		11e	50000	4
197		11f	10000	4
198		11g	50000	4
199		11h	50000	4
200		11i	50000	4
201		11j	50000	4
202		11k	1280	4
203		11l	1090	4
204		11m	1100	4
205		13a	50000	4
206		13b	50000	4
207		13c	50000	4
208		13d	50000	4

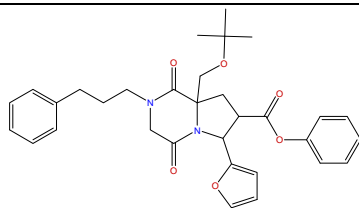
209		13e	50000	4
210		13f	7500	4
211		13g	50000	4
212		9a	350	5
213		9b	50000	5
214		9c	3100	5
215		9d	630	5
216		9e	3900	5
217		10a	5000	5
218		10b	8500	5
219		10c	350	5
220		10d	46	5
221		10e	46	5
222		10f	1230	5
223		11a	5000	5
224		11b	5000	5

225		11c	5000	5
226		11d	1400	5
227		11e	5000	5
228		11f	23	5
229		11g	3200	5
230		11h	5000	5
231		11i	250	5
232		11j	680	5
233		11k	65	5
234		11l	1100	5
235		11m	5000	5
236		12a	39	5
237		12b	250	5
238		12c	320	5

239		12d	26	5
240		12e	71	5
241		12f	1500	5
242		12g	51	5
243		12h	23	5
244		13a	5000	5
245		13b	2400	5
246		13c	120	5
247		13d	1100	5
248		13e	210	5
249		13f	110	5
250		13g	500	5
251		4	120	6



252		5	210	6
253		6	210	6
254		7	760	6
255		8	1300	6
256		9	8500	6
257		10	5500	6
258		11	100000	6
259		12	48000	6



- [1] Lowe, D. B.; Magnuson, S.; Oi, N.; Campbell, A.; Cook, J.; Hong, Z.; Wang, M.; Rodriguez, M.; Achebe, F.; Kluender, H.; Wong, W. C.; Bullock, W. H.; Salhanick, A. I. ; Witman, J. T.; Bowling, M. E.; Keiperb, C.; Clairmont, K. B. *Bioorg. Med. Chem. Lett.*, **2004**, **14**: 3155-3159.
- [2] Ebdrup, S.; Sorensen, L. G.; Olsen, O. H.; Jacobsen, P. *J. Med. Chem.*, **2004**, **47**: 400-410.
- [3] de Jong, J. C.; Sorensen, L. G.; Tornqvist, H.; Jacobsen, P. *Bioorg. Med. Chem. Lett.*, **2004**, **14**: 1741-1744.
- [4] Ebdrup, S.; Jacobsen, P.; Farrington, A. D.; Vedsø, P. *Bioorg. Med. Chem.*, **2005**, **13**: 2305-2312.
- [5] Ebdrup, S.; Hanne, H. F. R.; Christian, F.; Poul, J. *J. Med. Chem.*, **2007**, **50**: 5449-5456.
- [6] Slee, D. H.; Bhat, A. S.; Nguyen, T. N.; Kish, M.; Lundeen, K.; Newman, M. J.; McConnell, S. *J. Med. Chem.*, **2003**, **46**: 1120-1122.

**Table S2** The Compounds of Training Set, Testing Set, and Independent Validation Set

	ID of the compounds
Training set	1, 2, 3, 5, 6, 7, 8, 9, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23, 27, 30, 32, 33, 34, 36, 37, 39, 40, 41, 42, 43, 44, 45, 46, 49, 51, 52, 53, 56, 57, 59, 61, 62, 64, 65, 67, 68, 71, 73, 75, 76, 77, 78, 81, 83, 84, 89, 91, 95, 96, 98, 101, 103, 104, 105, 106, 107, 108, 109, 110, 112, 113, 114, 115, 119, 120, 121, 122, 123, 125, 126, 127, 128, 133, 138, 139, 140, 141, 143, 145, 146, 147, 148, 150, 151, 152, 154, 155, 156, 158, 159, 161, 165, 167, 168, 169, 171, 175, 176, 177, 178, 179, 184, 187, 188, 189, 191, 192, 194, 196, 197, 199, 200, 201, 202, 204, 207, 208, 209, 210, 214, 215, 216, 217, 221, 222, 223, 225, 227, 228, 229, 234, 235, 236, 237, 239, 240, 241, 243, 244, 245, 246, 247, 248, 251, 252, 253, 254, 255, 256, 257, 260
Testing set	11, 13, 15, 28, 35, 38, 47, 48, 54, 58, 63, 72, 74, 85, 90, 92, 94, 99, 118, 129, 130, 132, 134, 135, 136, 137, 149, 153, 160, 162, 170, 172, 173, 174, 181, 182, 183, 185, 186, 195, 198, 206, 211, 213, 218, 219, 220, 230, 231, 238, 259
Independent validation set	4, 10, 24, 25, 26, 31, 50, 55, 60, 66, 69, 70, 80, 82, 86, 87, 93, 100, 102, 116, 124, 131, 144, 157, 163, 164, 166, 180, 190, 193, 203, 205, 212, 224, 226, 232, 233, 242, 249, 258

**Table S3** Molecular descriptors used in this work

descriptor Class	number of descriptor	descriptors
Simple molecular properties	18	molecular weight, numbers of rings, rotatable bonds, H-bond donors, and H-bond acceptors, element counts
Molecular connectivity and shape	27	molecular connectivity indices, valence molecular connectivity indices, molecular shape Kappa indices, Kappa alpha indices, flexibility index
Electro-topological state	97	Electrotopological state indices, and atom type electrotopological state indices, Wiener index, centric index, Altenburg index, Balaban index, Harary number, Schultz index, PetitJohn R2 index, PetitJohn D2 index, mean distance index, PetitJohn I2 index, information Wiener, Balaban rmsd index, graph distance index
Quantum chemical properties	31	polarizability index, hydrogen bond acceptor basicity (covalent HBAB), hydrogen bond donor acidity (covalent HBDA), molecular dipole moment, absolute hardness, softness, ionization potential, electron affinity, chemical potential, electronegativity index, electrophilicity index, most positive charge on H, C, N, O atoms, most negative charge on H, C, N, O atoms, most positive and negative charge in a molecule, LSum of squares of charges on H, C, N, O and all atoms, mean of positive charges, mean of negative charges, mean absolute charge, relative positive charge, relative negative charge
Geometrical properties	25	length vectors (longest distance, longest third atom, 4th atom), molecular van der Waals volume, solvent accessible surface area, molecular surface area, van der Waals surface area, polar molecular surface area, sum of solvent accessible surface areas of positively charged atoms, sum of solvent accessible surface areas of negatively charged atoms, sum of charge weighted solvent accessible surface areas of positively charged atoms, sum of charge weighted solvent accessible surface areas of negatively charged atoms, sum of van der Waals surface areas of positively charged atoms, sum of van der Waals surface areas of negatively charged atoms, sum of charge weighted van der Waals surface areas of positively charged atoms, sum of charge weighted van der Waals surface areas of negatively charged atoms, molecular rugosity, molecular globularity, hydrophilic region, hydrophobic region, capacity factor, hydrophilic-hydrophobic balance, hydrophilic intery moment, hydrophobic intery moment, amphiphilic moment