Electronic Supplementary Material (ESI) for MedChemComm.

Supplementary materials

The Emerging Chemical Patterns Applied in Predicting Human Toll-Like Receptor 8 Agonists

Shuheng Huang [1, 2], Hu Mei [1, 2, *], Duo Zhang [2], Yubin Ren [2], MuliadiYeremia Kevin [2], Xianchao Pan [2, 3, *]

[1] Key Laboratory of Biorheological Science and Technology (Ministry of Education), Chongqing University, Chongqing 400044, China

[2] College of Bioengineering, Chongqing University, Chongqing 400044, China

[3] Department of Medicinal Chemistry, College of Pharmacy, Southwest Medical University, Luzhou, Sichuan 646000, China

Corresponding Author:

*Hu Mei, E-mail address: meihu@cqu.edu.cn

*Xianchao Pan; E-mail address: panxc@swmu.edu.cn

Table of contents

Table S1.

The information of 286 samples

AID [a]	Serial Number [b]	SID [c]	Smile Structure	Activity	EC50 (μM)
766330	70001	174498091	o3c1c(c2c(nc1N)cccc2)cc3C	Active	
	70002	174498092	o3c1c(c2c(nc1N)cccc2)cc3CCC	Active	
	70003	174498093	o3c1c(c2c(nc1N)cccc2)cc3CCCCCC	Active	
	70004	174503346	o3c1c(c2c(nc1N)cccc2)cc3CCCCC	Active	
	70005	174513636	o1c2c(cc1CCCC)ccnc2N	Active	46.2
	70006	174523914	o1c2c(cc1CCC)ccnc2N	Active	24.4
	70007	174523915	o3c1c(c2c(nc1N)cccc2)cc3	Inactive	
766334	70008	174487933	o3c1c(c2c(nc1N)cccc2)cc3CCCO	Active	
	70009	174487934	o3c1c(c2c(nc1N)cccc2)cc3CCOCC	Active	
	70010	174487935	o3c1c(c(nc2ccccc21)N)cc3CCCC	Active	
	70011	174492975	o3c1c(c2c(nc1N)cccc2)cc3C4CC4	Active	
	70012	174492976	o3c1c(c2c(nc1N)cccc2)cc3CC4CCCCC4	Active	
	70013	174492977	o3c1c(c2c(nc1N)cccc2)cc3CC[NH3+]	Active	
	70014	174498061	o3c2nc(c1c(cccc1)c2cc3CCCC)N	Inactive	
	70015	174498088	o3c1c(c2c(nc1N)cccc2)cc3Cc4ccccc4	Active	
	70016	174498089	o3c1c(c2c(nc1N)cccc2)cc3CO	Active	
	70017	174498090	o3c1c(c2c(nc1N)cccc2)cc3CCCCO	Active	
	70018	174498094	o3c1c(c2c(nc1N)cccc2)cc3C(C)(C)C	Active	
	70019	174503345	o3c1c(c2c(nc1N)cccc2)cc3CCO	Active	
	70020	174508536	o3c1c(c2c(nc1N)cccc2)cc3-c4ccccc4	Active	
	70021	174513657	o3c1c(c2c(nc1N)cccc2)cc3C4CCCC4	Active	
	70022	174513658	o3c1c(c2c(nc1N)cccc2)cc3CC4CCCC4	Active	
	70023	174513659	o3c1c(c2c(nc1N)cccc2)cc3CC(O)C	Active	
	70024	174518672	o3c1c(c2c(nc1N)cccc2)cc3CCC(C)C	Active	
	70025	174518695	o3c1c(c2c(nc1N)cccc2)cc3C(O)(C)C	Active	
	70026	174523911	o3c1c(c2c(nc1N)cccc2)cc3C(O)CC(C)C	Active	
	70027	174523912	o3c1c(c2c(nc1N)cccc2)cc3C[NH3+]	Active	
	70028	174523917	o3c1c(c2c(nc1N)cccc2)cc3CC(C)C	Active	
474846	70029	103746829	O=C2N(c1nc(nc(c1N2)N)OCCCC)Cc3cc(ccc3)CC(=O)OC	Inactive	
1138188	70030	242634216	O4C(C2C(C(c1cc(c(c(c1)OC)OC)OC)C2C)c5c(nc3c(cccc3)c54)O)(C)C	Inactive	
706910	70031	136937159	n21c(c(nc1cncc2)-c3ccccc3)[NH2+]C(C)(C)C	Inactive	
	70032	163312889	n21c(c(nc1cncc2)-c3ccccc3)[NH2+]Cc4ccccc4	Inactive	
	70033	163314443	O(c1ccc(cc1)Nc2n3c(nc2CCCC)c(ncc3)N)C	Inactive	
	70034	163314444	[O-]c1c(cccc1)-c3nc2n(ccnc2)c3[NH2+]C4CCCCC4	Inactive	
	70035	163317909	n21c(c(nc1cccc2)-c3ccncc3)[NH2+]C4CCCCC4	Inactive	

	70036	163319732	n21c(c(nc1cccc2)-c3ccccc3)[NH2+]C(C)(C)C	Inactive	
	70037	163319733	n21c(c(nc1cccc2)-c3ccccc3)[NH2+]Cc4ccccc4	Inactive	
	70038	163321381	n21c(c(nc1c(ncc2)N)-c3ccccc3)[NH2+]C4CCCCC4	Inactive	
	70039	163326567	n21c(c(nc1cncc2)-c3ccccc3)[NH2+]C4CCCCC4	Inactive	
	70040	163328162	n31c(nc(c1[NH2+]Cc2ccccc2)CCCC)c(ncc3)N	Inactive	
	70041	163328163	n21c(c(nc1cncc2)-c3ccncc3)[NH2+]Cc4ccccc4	Inactive	
	70042	163328164	[O-]c1c(cccc1)-c3nc2n(ccnc2)c3[NH2+]Cc4ccccc4	Inactive	
	70043	163331739	n21c(c(nc1cncc2)-c3ccncc3)[NH2+]C(C)(C)C	Inactive	
	70044	163331740	[O-]c1c(cccc1)-c3nc2n(cccc2)c3[NH2+]Cc4ccccc4	Inactive	
	70045	163335187	n21c(c(nc1c(ncc2)N)-c3ccc(cc3)-c4ccccc4)[NH2+]C5CCCCC5	Inactive	
	70046	163335188	n21c(c(nc1cccc2)-c3ccncc3)[NH2+]C(C)(C)C	Inactive	
	70047	163335189	n21c(c(nc1cncc2)-c3ccncc3)[NH2+]C4CCCCC4	Inactive	
	70048	163335190	n21c(c(nc1cccc2)-c3ccncc3)[NH2+]Cc4ccccc4	Inactive	
	70049	163335191	Oc1c(cccc1)-c3nc2n(ccnc2)c3[NH2+]C(C)(C)C	Inactive	
	70050	163338585	Oc1c(cccc1)-c3nc2n(cccc2)c3[NH2+]C(C)(C)C	Inactive	
706908	70051	103602263	s1c2c(nc1CCC)c(nc3ccccc32)N	Active	1.32
	70052	163314442	o1c6c(c(c1[NH2+]CCCCCC[NH2+]c3oc2c(c(cnc2C)CO)c3Nc4nccccc4)Nc5nccc5)c(cnc6C)CO	Active	3.37
	70053	163317906	o1c3c(c(c1[NH3+])Nc2nccccc2)c(cnc3C)CO	Inactive	
	70054	163317907	o1c4c(c(c1[NH2+]C2CCCCC2)Nc3nccccc3)c(cnc4C)CO	Active	6.93
	70055	163317908	P(=O)(OCC)(OCC)C[NH2+]c2oc1c(c(cnc1C)CO)c2Nc3nccccc3	Inactive	
	70056	163317910	o1c3c(c(c1[NH2+]C(C)(C)C)Nc2nccccc2)c(cnc3C)CO	Inactive	
	70057	163321377	o1c3c(c(c1NC(C)(C)C)Nc2nccnc2)c(cnc3C)CO	Inactive	
	70058	163321378	o1c4c(c(c1[NH2+]C2CCCCC2)Nc3nccnc3)c(cnc4C)CO	Active	10.58
	70059	163321379	o1c3c(c(c1[NH2+]CCC(=O)OC(C)(C)C)Nc2nccccc2)c(cnc3C)CO	Active	7.64
	70060	163321380	o1c4c(c(c1Nc2ccc(cc2)OC)Nc3nccccc3)c(cnc4C)CO	Inactive	
	70061	163324730	o1c3c(c(c1[NH2+]CCCCC)Nc2cc(ccc2)[N+](=O)[O-])c(cnc3C)CO	Active	0.85
	70062	163324731	o1c4c(c(c1[NH2+]Cc2ccccc2)Nc3nccccc3)c(cnc4C)CO	Inactive	
	70063	163324732	o1c4c(c(c1[NH2+]CCN2CCOCC2)Nc3nccccc3)c(cnc4C)CO	Inactive	
	70064	163328159	o1c4c(c(c1[NH2+]Cc2ccccc2)Nc3nccnc3)c(cnc4C)CO	Active	9.79
	70065	163328160	O(c1c(c(cnc1C)CO)-c3nc2n(cccc2)c3[NH2+]C4CCCCC4)Cc5ccccc5	Inactive	
	70066	163328161	[Si](C[NH2+]c2oc1c(c(cnc1C)CO)c2Nc3nccccc3)(C)(C)C	Inactive	
	70067	163331734	o1c3c(c(c1[NH2+]CCCCC)Nc2ccccc2)c(cnc3C)CO	Active	2.25
	70068	163331735	o1c3c(c(c1[NH2+]CCCCC)Nc2nccccc2)c(cnc3C)CO	Inactive	
	70069	163331736	o1c3c(c(c1[NH2+]C(C)C)Nc2nccccc2)c(cnc3C)CO	Inactive	
	70070	163331737	o1c3c(c(c1[NH2+]C(CCC)C)Nc2nccccc2)c(cnc3C)CO	Active	9.01
	70071	163331738	Clc1c(c(ccc1)C)Nc3oc2c(c(c[nH+]c2C)CO)c3Nc4nccccc4	Inactive	
	70072	163335186	o1c4c(c(c1[NH2+]Cc2ccccc2)Nc3ccccc3)c(cnc4C)CO	Active	1.68
	70073	163338580	Fc1cc(ccc1)Nc2c3c(oc2[NH2+]CCCCC)c([nH+]cc3CO)C	Active	0.37
	70074	163338581	o1c3c(c(c1[NH2+]CCCC)Nc2nccccc2)c(cnc3C)CO	Inactive	

	70075	163338582	o1c3c(c(c1NC(CC(C)(C)C)(C)C)Nc2[nH+]cccc2)c(c[nH+]c3C)CO	Inactive	
	70076	163338583	o1c3c(c(c1[NH2+]CC(=O)OCC)Nc2ncccc2)c(cnc3C)CO	Inactive	
	70077	163338584	o1c4c(c(c1[NH2+]C(c2ccccc2)C)Nc3ncccc3)c(cnc4C)CO	Active	4.27
484312	70078	103758846	OC(Cn1c2c(nc1CNCC)c(nc3ccccc32)N)(C)C	Inactive	
	70079	103758847	OC(Cn1c2c(nc1C[NH+](CC)CC)c([nH+]c3ccccc32)N)(C)C	Inactive	
	70080	103758910	OC(Cn1c2c(nc1CN(C(=O)C)CC)c(nc3ccccc32)N)(C)C	Inactive	
	70081	103758911	OC(Cn1c2c(nc1CN(C(=O)CCCCCCCCCCCCCCCC)CC)c(nc3ccccc32)N)(C)C	Inactive	
	70082	103758912	OC(Cn1c2c(nc1CN(C(=[NH+]CC)NCCCN(C)C)CC)c(nc3ccccc32)N)(C)C	Inactive	
	70083	103758913	OC(Cn1c2c(nc1C[NH+](CCOCCOCC[NH3+])CC)c(nc3ccccc32)N)(C)C	Inactive	
	70084	103758978	OC(Cn1c2c(nc1CN(CCC#N)CC)c([nH+]c3ccccc32)N)(C)C	Inactive	
	70085	103758979	OC(Cn1c2c(nc1C[NH+](CCC[NH3+])CC)c(nc3ccccc32)N)(C)C	Inactive	
	70086	103758981	OC(Cn1c2c(nc1CNCCCCCCCCCCC)c(nc3ccccc32)N)(C)C	Inactive	
	70087	103758982	n1(c2c(nc1C)c(nc3ccccc32)N)Cc4ccccc4	Inactive	
	70088	103758983	n1(c2c(nc1CC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70089	103759059	n1(c2c(nc1CCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70090	103759060	n1(c2c(nc1CCCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70091	103759061	n1(c2c(nc1CCCCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70092	103759062	n1(c2c(nc1CCCCCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70093	103759063	n1(c2c(nc1CCCCCCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70094	103759134	n1(c2c(nc1CCCCCCCCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70095	103759135	n1(c2c(nc1CCCCCCCCCCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70096	103759136	n1(c2c(nc1CCCCCCCCCCCCCCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70097	103759137	n1(c2c(nc1CCC=C)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70098	103759209	n1(c2c(nc1CCC#C)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70099	103759210	Clc1[nH+]c4c(c2n(c(nc21)CCCCC)Cc3ccccc3)cccc4	Inactive	
	70100	103759211	O=C2Nc1c(cccc1)-c3n(c(nc32)CCCCC)Cc4ccccc4	Inactive	
	70101	103759212	n1(c2c(nc1CCCCC)c([nH+]c3ccccc32)-c4ccccc4)Cc5ccccc5	Inactive	
	70102	103759213	ONc1[nH+]c4c(c2n(c(nc21)CCCC)Cc3ccccc3)cccc4	Inactive	
	70103	103759285	n1(c2c(nc1CCCC)c([nH+]c3ccccc32)N[NH3+])Cc4ccccc4	Inactive	
	70104	103759286	n1(c2c(nc1CCCC)c[nH+]c3ccccc32)Cc4ccccc4	Inactive	
	70105	103759287	n1(c2c(nc1CNCC)c([nH+]c3ccccc32)N)Cc4ccccc4	Inactive	
	70106	103759289	Clc1nc4c(c2n(nnc21)Cc3ccccc3)cccc4	Inactive	
	70107	103759290	n2(nnc3c(nc1c(cccc1)c32)N)Cc4ccccc4	Inactive	
	70108	103759372	O=C3N(c1c(cnc2ccccc21)N3C(=O)NCCC)Cc4ccccc4	Inactive	
389403	70109	103647001	O(C(=O)N2Cc1c(cccc1)N=C(N)C2)CC	Inactive	100
	70110	103647002	O=C(N2Cc1c(cccc1)N=C(N)C2)N(CC)CC	Inactive	100
719614	70111	103396795	n1(c2c(nc1)c(nc3ccccc32)N)CC(C)C	Inactive	270
	70112	103463768	OC(Cn1c2c(nc1)c(nc3ccccc32)N)(C)C	Inactive	270
	70113	103759288	OC(Cn1c2c(nc1CCCC)c(nc3ccccc32)N)(C)C	Active	0.1

	70114	163315609	OC(Cn2c1c3c(nc(c1nc2)N)cc(cc3)C(=O)OC)(C)C	Inactive	270
	70115	163315610	OC(Cn1c2c(nc1CCCC)c(nc3cc(ccc32)C(=O)OC)N)(C)C	Active	41.3
770386	70116	174508709	O(c1nc2c(c(n1)N)NC(=O)CN2Cc3cc(ccc3)CN4CCCC4)CCCC	Active	9
1068404	70117	103481674	O(Cc3n(c1c(c(nc2ccccc21)N)n3)CC(O)(C)C)CC	Active	4.5
	70118	194146855	O(C(=O)c3cc2nc(c1nc(n(c1c2cc3)CCOCc4ccccc4)CCCC)N)C	Active	31
	70119	194146856	OC(Cn1c2c([nH+]c1CC)c(nc3cc(ccc32)C(=O)OC)N)(C)C	Inactive	
	70120	194146857	OC(Cn1c3c(nc1Cc2ccccc2)c(nc4cc(ccc43)C(=O)OC)N)(C)C	Inactive	
	70121	194160693	OC(Cn1c2c(nc1CCCCC)c(nc3cc(ccc32)C(=O)OC)N)(C)C	Active	7.2
	70122	194167603	O(C(=O)c3cc2nc(c1nc(n(c1c2cc3)CCOC)CCCC)N)C	Inactive	
	70123	194167604	OC(Cn1c2c(nc1C)c(nc3cc(ccc32)C(=O)OC)N)(C)C	Inactive	
	70124	194167605	O(C(=O)c3cc2nc(c1nc(n(c1c2cc3)CC(O)C)CCCC)N)C	Active	1.5
	70125	194181308	OC(Cn1c2c(nc1CCC)c(nc3cc(ccc32)C(=O)OC)N)(C)C	Inactive	
	70126	194181309	O(C(=O)c3cc2nc(c1nc(n(c1c2cc3)CCO)CCCC)N)C	Active	4.4
1187004	90001	312355083	c1cccc2c1NC(=[NH2+])C2CCCC	Inactive	
	90002	312355084	c1cccc2c1NC(=N)C2(CCCC)O	Inactive	
	90003	194147973	c1cccc2n(c(nc21)N)CCCC	Active	7.3
	90004	312355085	c1cccc2n(c(nc21)N)CCCCC	Active	3.23
	90005	194157023	c1cccc2n(c(nc21)N)CCCCCC	Active	3.96
	90006	312355086	c1cccc2n(c(nc21)N)Cc3ccccc3	Inactive	
	90007	312368827	c1cccc2n(c(nc21)N)Cc3cccc(c3)C[NH3+]	Inactive	
	90008	312355087	c1cccc2n(c(nc21)N)C(=O)OCCCC	Inactive	
	90009	312355088	c1cccc2n(c(nc21)NC(=O)C)CCCCC	Inactive	
	90010	312355089	c1cccc2cc3c(cc21)nc(n3CCCCC)N	Inactive	
	90011	312355090	c1cc3c(c2n(c(nc21)N)CCCCC)cccc3	Inactive	
	90012	312355091	c3ccc1c(ccc2n(c(nc21)N)CCCCC)c3	Active	3.16
	90013	312355092	n1cccc2n(c(nc21)N)CCCCC	Inactive	
	90014	312355093	c1nccc2n(c(nc21)N)CCCCC	Inactive	
	90015	312355094	c1cncc2n(c(nc21)N)CCCCC	Inactive	
	90016	312355095	c1ccnc2n(c(nc21)N)CCCCC	Inactive	
	90017	312355096	c1(cccc2n(c(nc21)N)CCCCC)C	Active	1.13
	90018	312355097	c1c(ccc2n(c(nc21)N)CCCCC)C	Active	4.57
	90019	312355098	c1cc(cc2n(c(nc21)N)CCCCC)C	Active	7.21
	90020	312355099	c1ccc(c2n(c(nc21)N)CCCCC)C	Active	6.61
	90021	312355100	c1(cccc2n(c(nc21)N)CCCCC)OC	Active	3.74
	90022	312355101	c1(cccc2n(c(nc21)N)CCCCC)F	Inactive	
	90023	312355102	c1(cccc2n(c(nc21)N)CCCCC)Cl	Inactive	
	90024	312355103	c1(cccc2n(c(nc21)N)CCCCC)C(F)(F)F	Inactive	
	90025	312355104	c1(cccc2n(c(nc21)N)CCCCC)Br	Inactive	
	90026	312355105	c1(cccc2n(c(nc21)N)CCCCC)CC	Active	1.65

	90027	312355106	c1(cccc2n(c(nc21)N)CCCCC)N(C)C	Active	7.12
	90028	312355107	c1(cccc2n(c(nc21)N)CCCCC)-c3ccccc3	Inactive	
	90029	312355108	c1(cccc2n(c(nc21)N)CCCCC)Cc3ccccc3	Inactive	
	90030	312355109	c1(cccc2n(c(nc21)N)CCCCC)OCc3ccccc3	Inactive	
	90031	312355110	c1(cccc2n(c(nc21)N)CCCCC)[O-]	Active	5.01
	90032	312355111	c1(cccc2n(c(nc21)N)CCCCC)[N+](=O)[O-]	Inactive	
	90033	312355112	c1(cccc2n(c(nc21)N)CCCCC)N	Active	6.6
	90034		c1cccc2c3c(c(nc21)N)oc(n3)CCCC	Active	
	90035		c1cccc2n3c(c(nc21)N)nc(c3)CCCC	Active	3.05
	90036		c1cccc2c3c(c(nc21)N)oc(c3)CCCC	Active	
	90037		c1cccc2c3c(c(nc21)N)cc(s3)CCCC	Active	
	90038		c1cccc2c3c(c(nc21)N)cc([nH]3)CCCC	Active	
	90039		c1cccc2c3c(c(nc21)N)NN(CCCC)N3	Active	
	90040		c1cccc2c3c(c(nc21)N)nc(s3)NCCC	Active	3.94
	90041		c1cccc2c3c(c(nc21)N)nc(o3)CCCC	Active	0.18
	90042		c1cccc2c3c(c(nc21)N)nc(s3)CCCC	Active	
	90043		c1cccc2c3c(c(nc21)N)NN(CCCC)C3	Active	0.056
	90044		c1cccc2c3c(c(nc21)N)nc([nH]3)NCCC	Active	
	90045		c1cccc2c3c(c(nc21)N)sc(c3)CCCC	Active	
	90046		c1cccc2c3c(c(nc21)N)nc([nH]3)CCCC	Active	
	90047		c1cccc2c3c(c(nc21)N)CC(=C3)CCCC	Active	
	90048		c1cccc2c3c(c(nc21)N)nc([nH]3)COCC	Active	
	90049		c1cccc2c1N=C(C3=NC(CCCC)CN32)N	Active	7.99
	90050		c1cccc2n3c(c(nc21)N)nc(n3)CCCC	Inactive	
	90051		c1cccc2c3c(c(nc21)N)nc(s3)NC(=O)CC	Inactive	
	90052		c1cccc2n3c(c(nc21)N)nnc3CCCC	Inactive	
	90053		c1cccc2c3c(c(nc21)N)nc(o3)NCCC	Inactive	
	90054		c1cccc2c3n(c(nc21)N)nc(n3)CCCC	Inactive	
	90055		c1cccc2c3c(c(nc21)N)sc(n3)CCCC	Inactive	
	90056		c1cccc2c3n(c(nc21)N)cc(n3)CCCC	Inactive	
	90057		c1cccc2c3c(c(nc21)N)C=C(CCCC)C3	Active	
	90058		c1cccc2c3c(c(nc21)N)[nH]c(c3)CCCC	Inactive	
	90059		c1cccc2c(c(c([nH+]c21)N)CCCCC)Cc3cccc(c3)C[NH3+]	Active	0.15
	90060		c1cccc2c(c(c(nc21)N)CCCCC)Cc3ccc(cc3)CN	Active	0.12
	90061		c1cccc2c(c(c(nc21)N)CCCCC)CCCC[NH3+]	Active	0.19
	90062		c1cccc2c(c(c(nc21)N)CCCCC)CCCCC[NH3+]	Active	0.25
	90063		c1ccc(c2cc(c(nc21)N)CCCCC)Cc3cccc(c3)C#N	Inactive	
	90064		c1ccc(c2cc(c(nc21)N)CCCCC)Cc3cccc(c3)C[NH3+]	Active	0.049
	90065		c1ccc(c2cc(c(nc21)N)CCCCC)Cc3ccc(cc3)C#N	Inactive	

	90066		c1ccc(c2cc(c(nc21)N)CCCCC)Cc3ccc(cc3)C[NH3+]	Active	0.038
	90067		c1ccc(c2cc(c(nc21)N)CCCCC)Cc3ccccc3C[NH3+]	Active	1
	90068		c1ccc(c2cc(c(nc21)N)CCCCC)Cc3ccccc3	Inactive	
	90069		c1ccc(c2cc(c([nH+]c21)N)CCCCC)Cc3cccc(c3)C(=O)N	Inactive	
	90070		c1ccc(c2cc(c(nc21)N)CCCCC)-c3cccc(c3)C[NH3+]	Inactive	
	90071		c1ccc(c2cc(c(nc21)N)CCCCC)-c3ccc(cc3)C[NH3+]	Active	0.699
	90072		c1ccc(c2cc(c(nc21)N)CCCCC)CCC[NH3+]	Active	0.091
	90073		c1ccc(c2cc(c(nc21)N)CCCCC)CCCC[NH3+]	Active	0.027
	90074		c1ccc(c2cc(c(nc21)N)CCCCC)CCCCC[NH3+]	Active	0.009
	90075		c1ccc(c2cc(c(nc21)N)CCCCC)CCCCCC[NH3+]	Active	0.056
	90076		c1ccc(c2cc(c(nc21)N)CCCCC)CCCC(=O)N	Active	2.181
	90077		c1ccc(c2cc(c(nc21)N)CCCCC)CCCCNC(=N)N	Active	2.862
	90078		c1cc(cc2cc(c(nc21)N)CCCCC)CCCC[NH3+]	Active	0.727
	90079		c1cc(cc2cc(c(nc21)N)CCCCC)CCCCC[NH3+]	Active	0.519
	90080		c1cc(cc2cc(c(nc21)N)CCCCC)CCCCCC[NH3+]	Active	1.016
	90081		c1c(ccc2cc(c(nc21)N)CCCCC)CCCC[NH3+]	Active	0.06
	90082		c1c(ccc2cc(c(nc21)N)CCCCC)CCCCC[NH3+]	Active	0.05
	90083		c1c(ccc2cc(c(nc21)N)CCCCC)CCCCCC[NH3+]	Active	0.085
	90084		c1(cccc2cc(c(nc21)N)CCCCC)CCCC[NH3+]	Inactive	
	90085		c1c(cc(c2cc(c(nc21)N)CCCCC)CCCCC[NH3+])CCCCC[NH3+]	Active	0.621
	80001		c1nc(n(c1)CCCCC)N	Active	28.4
	80002		c1nc(n(c1-c2ccccc2)CCCCC)N	Inactive	
	80003		c1(nc(n(c1-c2ccccc2)CCCCC)N)-c3ccccc3	Inactive	
	80004		c1(nc(n(c1)CCCCC)N)-c2ccccc2	Active	2.48
	80005		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)C	Active	2.5
	80006		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)CC	Active	2.5
	80007		c1(nc(n(c1)CCCCC)N)-c2cc(ccc2)C	Active	2
	80008		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)C	Active	2.5
	80009		c1(nc(n(c1)CCCCC)N)-c2c(c(ccc2)C)C	Active	1.5
	80010		c1(nc(n(c1)CCCCC)N)-c2c(cc(cc2)C)C	Active	1.5
	80011		c1(nc(n(c1)CCCCC)N)-c2c(ccc(c2)C)C	Active	1.5
	80012		c1(nc(n(c1)CCCCC)N)-c2c(cccc2C)C	Active	2.5
	80013		c1(nc(n(c1)CCCCC)N)Cc2ccccc2	Active	2.7
	80014		c1(nc(n(c1)CCCCC)N)CCc2ccccc2	Active	1.8
	80015		c1(nc(n(c1)CCCCC)N)-c2c3c(ccc2)cccc3	Active	1.5
	80016		c1(nc(n(c1)CCCCC)N)-c2cc3c(cc2)cccc3	Inactive	
	80017		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)-c3ccccc3	Inactive	
	80018		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)OCc3ccccc3	Inactive	
	80019		c1(nc(n(c1)CCCCC)N)-c2cnccc2	Active	21.31

	80020		c1(nc(n(c1)CCCCC)N)-c2c(noc2C)C	Active	18.8
	80021		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)O	Active	5.1
	80022		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)CO	Active	8.05
	80023		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)N	Active	8
	80024		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)C(=O)N	Active	20.24
	80025		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)C(=O)OC	Active	4.04
	80026		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)C(=O)OC	Active	11.1
	80027		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)OC	Active	1.6
	80028		c1(nc(n(c1)CCCCC)N)-c2cc(ccc2)OC	Active	2.2
	80029		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)OC	Active	2.94
	80030		c1(nc(n(c1)CCCCC)N)-c2c(cc(cc2)OC)C	Active	1.5
	80031		c1(nc(n(c1)CCCCC)N)-c2c(c(ccc2)C)OC	Active	1.5
	80032		c1(nc(n(c1)CCCCC)N)-c2c(c(ccc2)OC)OC	Active	2.5
	80033		c1(nc(n(c1)CCCCC)N)-c2cc(c(c(c2)OC)OC)OC	Active	3.08
	80034		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)Cl	Active	1.36
	80035		c1(nc(n(c1)CCCCC)N)-c2cc(ccc2)Cl	Active	1.93
	80036		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)Cl	Active	6.83
	80037		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)F	Active	2.35
	80038		c1(nc(n(c1)CCCCC)N)-c2cc(ccc2)F	Active	2.57
	80039		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)F	Active	1.96
	80040		c1(nc(n(c1)CCCCC)N)-c2c(cc(cc2)F)C	Active	2.01
	80041		c1(nc(n(c1)CCCCC)N)-c2c(cccc2)C(F)(F)F	Active	1.64
	80042		c1(nc(n(c1)CCCCC)N)-c2cc(ccc2)C(F)(F)F	Active	1.98
	80043		c1(nc(n(c1)CCCCC)N)-c2ccc(cc2)C(F)(F)F	Active	2.48
	80044		n1c(c(c(nc1N)NCCC)I)C	Inactive	
	80045		n1c(c(c(nc1N)NCCCC)I)C	Active	1.64
	80046		n1c(c(c(nc1N)NCCCCC)I)C	Active	3.7
	80047		n1c(c(c(nc1N)NCCCCCC)I)C	Inactive	
	80048		n1c(c(c(nc1N)N(C)CCCC)I)C	Inactive	
	80049		n1c(c(c(nc1N)N(Cc2ccccc2)CCCC)I)C	Inactive	
	80050		n1c(c(c(nc1N)OCCCC)I)C	Inactive	
	80051		n1c(c(c(nc1N)SCCCC)I)C	Inactive	
	80052		n1c(cc(nc1N)C#CCCC)C	Inactive	
	80053		n1c(cc(nc1N)CCCCC)C	Inactive	
	80054		n1c(c(c(nc1N)CCCCC)I)C	Inactive	
	80055		n1c(c(c(nc1N)NCCCC)Cl)C	Active	20
	80056		n1c(c(c(nc1N)NCCCC)Br)C	Active	8.5
	80057		n1c(c(c(nc1N)NCCCC)F)C	Active	34
	80058		n1c(cc(nc1N)NCCCC)C	Active	22

	80059		n1ccc(nc1N)NCCCC	Active	73.2
	80060		n1cc(c(nc1N)NCCCC)I	Inactive	
	80061		n1c(cc(nc1N)NCCCC)N	Inactive	
	80062		n1c(cc(nc1N)NCCCC)OC	Inactive	
	80063		n1c(cc(nc1N)NCCCC)Cl	Inactive	
	80064		n1c(c(c(nc1N)NCCCC)I)N	Inactive	
	80065		n1c(c(c(nc1N)NCCCC)I)OC	Inactive	
	80066		n1c(c(c(nc1N)NCCCC)I)Cl	Inactive	
	80067		n1c(c(c(nc1N)NCCCC)Cc2ccccc2)C	Active	1.2
	80068		n1c(c(c(nc1N)NCCCC)CCCN)C	Active	2.31
	80069		n1c(c(c(nc1N)NCCCC)CCCCN)C	Active	0.3
	80070		n1c(c(c(nc1N)NCCCC)CCCCCN)C	Active	0.3
	80071		n1c(cc(nc1N)NCCCC)-c2ccccc2	Active	6.7
	80072		n1c(c(c(nc1N)NCCCC)I)-c2ccccc2	Inactive	
	80073		n1c(c(c(nc1N)NCCCC)CCCN)-c2ccccc2	Inactive	
	80074		n1c(c(c(nc1N)NCCCC)CCCCN)-c2ccccc2	Active	2.7
	80075		n1c(c(c(nc1N)NCCCC)CCCCCN)-c2ccccc2	Active	2.35

[a] Pubmed BioAssay record ID (https://pubchem.ncbi.nlm.nih.gov/#)

[b] Assigned molecular serial number in this work

[c] PubChem Substance ID

Table S2.

The performances of step-wise linear discriminant analysis on 132 training samples

	Entered	Removed	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	nN		0.422	0.178	0.171	0.911
2	fHpbVSA		0.519	0.270	0.258	0.862
3	BV12 -DRY		0.563	0.318	0.302	0.837
4	BV12 -OH2		0.609	0.371	0.351	0.806
5	ID6 -DRY		0.622	0.386	0.362	0.800
6	mlogpUB		0.631	0.398	0.370	0.795
7	tAromVSA		0.642	0.412	0.378	0.789
8	HB7 -O		0.656	0.430	0.393	0.780
9	f-veVSA		0.669	0.447	0.407	0.771
10	BV11 -DRY		0.678	0.459	0.414	0.766
11	D13-DRY		0.687	0.471	0.423	0.760
12	R-OH2		0.699	0.488	0.437	0.751
13	fHplVSA		0.706	0.499	0.444	0.746
14	Mv		0.715	0.511	0.453	0.741
15		nN	0.725	0.525	0.464	0.733
16	ID5 -DRY		0.724	0.524	0.467	0.731
17	W4		0.734	0.539	0.479	0.722
18		BV12 -OH2	0.733	0.537	0.482	0.720
19	ID7 -DRY		0.745	0.555	0.497	0.710

Table S3.

The performance comparison between SVM and ECP models

model	No. of descriptors	Training set(%)			Validation set(%)		
		Acc	Sen	Spe	Acc	Sen	Spe
SVM	190	77.5	81.8	73.5	74.7	69.4	79.1
	17	77.3	77.0	77.4	72.2	61.1	81.4
	10	75.1	88.5	63.5	74.7	75.0	74.4
	6	77.5	81.8	73.5	74.7	69.4	79.1
ECP	6[a]	83.3	93.4	74.6	81.0	80.6	81.4

a: the best ECP model with 6 descriptors.

Table S4.

The results of 30 repeated ECP modeling

model	number of descriptors	number of ECPs	Training set(%)			Validation set(%)		
			Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
1	6	28	84.1	83.6	84.5	79.7	75	83.7
2		30	83.3	80.7	85.3	78.5	70	87.2
3		34	86.4	88.5	84.5	75.9	72.2	79.1
4		23	79.5	88.4	69.8	73.4	71.4	74.5
5		27	86.4	79.6	91	77.2	79.1	75
6		26	91	90.2	91.4	70.9	63	81.8
7		26	81.8	85.7	78.3	73.4	76.5	71.1
8		27	88.6	88.3	88.9	74.7	88.6	63.6
9		26	81.1	88.9	75.6	78.5	79.1	77.8
10		22	79.5	84.2	76	70.9	70	71.8
11		33	90.9	90.2	91.4	75.9	72.2	79.1
12		32	87.1	86.4	87.7	72.2	76.3	68.3
13		25	83.3	83.1	83.6	73.4	71.4	75.7
14		21	87.9	84.9	89.9	70.9	63	81.8
15		32	87.1	88.1	86.3	83.5	87.9	80.4
16		32	89.4	91.9	87.1	74.7	88.6	63.6
17		24	85.6	83.1	87.7	78.5	78.9	78
18		31	84.8	86.9	83.1	74.7	77.8	72.1
19		15	85.6	82.7	87.5	73.4	76.5	71.1
20		27	87.9	81.1	92.4	75.9	72.2	79.1
21		31	88.6	88.3	88.9	74.7	81.1	69.1
22		29	81.1	87.1	75.7	70.9	88.6	56.8
23		24	85.6	86.4	84.9	75.9	78.9	73.2
24		29	84.8	86.6	83.1	86.1	83.3	87.8
25		22	83.3	87.5	79.4	83.5	87.9	80.4
26		22	83.3	87.5	79.4	83.5	87.9	80.4
27		25	83.3	93.5	74.3	77.2	77.1	77.3
28		21	79.5	85.5	74.3	77.2	77.1	77.3
29		26	86.4	92.7	81.8	73.4	71.4	75.7
30		27	86.4	91.5	82.2	77.2	76.3	78
Average			85.1	86.8	83.5	76.2	77.3	75.7
Std Dev			3.2	3.5	6	4	7.1	6.9

Table S5.

The 13 ECPs discovered from the non-agonists of TLR8 (ECPG2)

Patterns	Class	Patterns Information	Growth Rate	Target-Class Support(%)	Rest Support(%)	Contribution
P15	-1	BV12-OH2>21.3 BV12-OH2≤24.8, ID6-DRY≤81.2	6	19.7	3.3	16.9
P16	-1	BV12-OH2>23.6 BV12-OH2≤48 BV12-DRY≤32 ID6-DRY>81.2 fHpbVSA_v1.0≤0.4	9.5	15.5	1.7	14
P17	-1	BV12-OH2>24.8 BV12-OH2≤45.7	infinity	23.9	0	23.9
P18	-1	BV12-OH2≤17.8	8.6	14.1	1.6	12.6
P19	-1	BV12-OH2≤23.6 BV12-DRY≤28.7 ID6-DRY>81.2 fHpbVSA_v1.0≤0.4	infinity	18.3	0	18.3
P20	-1	BV12-OH2≤34.1 BV12-DRY≤40.8 mlogpUB_v1.0>7.5	infinity	47.9	0	47.9
P21	-1	BV12-OH2≤46.9 BV12-DRY≤28.7	2.7	62	23	45.2
P22	-1	BV12-OH2≤48 BV12-DRY≤28.7 ID6-DRY>61.3 fHpbVSA_v1.0≤0.4	3.4	39.4	11.5	30.5
P23	-1	BV12-OH2≤48 BV12-DRY≤28.7 nN_v1.0>3.5, fHpbVSA_v1.0≤0.4	9.5	31.0	3.3	28.0
P24	-1	BV12-OH2≤48 nN_v1.0>3.5, fHpbVSA_v1.0≤0.4	2.3	45.1	19.7	31.4
P25	-1	BV12-DRY≤40.1 fHpbVSA_v1.0>0.4	7.2	35.2	4.9	30.9
P26	-1	nN_v1.0>3.5, fHpbVSA_v1.0>0.4	infinity	23.9	0	23.9
P27	-1	mlogpUB_v1.0>7.5	2.6	69.0	26.2	50.0

Table S6.

The prediction results of 46 TLR8 agonists collected from the most recent published papers

Ref	Simile structures	Activity (μM)	Label	Predicted (ECP)	Predicted (Surflex-dock)
50	N/C3=N/C1=CC=CC=C1C=2C=C(OC=23)CCCC	1.6	1	1	1
	N/C2=N/C1=CC=CC=C1/C=C2/CCCCC	0.2	1	1	1
	N/C2=N/C1=CC=CC(CCCCCN)=C1/C=C2/CCCCC	0.009	1	1	1
	N/C2=N/C=1C(=CC=CC=1C)N2CCCCC	1.13	1	1	1
51	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3CC(C)(C)O)C(=O)OC	20.86	1	-1	-1
	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3C[C@@](C)(O)[H])C(=O)OC	9.88	1	1	-1
	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3CCO)C(=O)OC	37	1	1	1
	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3CCN)C(=O)OC	49.8	1	-1	-1
	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3CCCN)C(=O)OC	26.7	1	1	1
	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3CCCCN)C(=O)OC	3.85	1	1	1
	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3CCCCCN)C(=O)OC	2.21	1	1	-1
	N/C2=N/C=1C=C(C=CC=1C3=C2/N=C(/CCCC)N3C/C4=C/C=C(C=C4)CN)C(=O)OC	5.34	1	1	1
52	NC=2/N=C(/NC)C1=CC=CC=C1N=2	6.87	1	1	1
	NC=2/N=C(/NCC)C1=CC=CC=C1N=2	0.77	1	1	1
	NC=2/N=C(/NCCC)C1=CC=CC=C1N=2	0.93	1	1	1
	NC=2/N=C(/NCCCC)C1=CC=CC=C1N=2	0.07	1	1	1
	NC=2/N=C(/NCCCCC)C1=CC=CC=C1N=2	0.10	1	1	1
	NC=2/N=C(/NCCC(C)(C)[H])C1=CC=CC=C1N=2	2.11	1	1	1
	NC=2/N=C(/N[C@@](C)([H])CCC)C1=CC=CC=C1N=2	0.17	1	1	1
	NC=2/N=C(/NCCCOC)C1=CC=CC=C1N=2	0.01	1	1	1
	NC=2/N=C(/NCCCC)C1=CC=CC([F])=C1N=2	0.3	1	1	1
	NC=2/N=C(/NCCCC)C1=CC=C(C=C1N=2)OC	0.3	1	1	1
	NC=2/N=C(/NCCCC)C1=CC=C(C=C1N=2)CO	0.1	1	1	1
	NC=2/N=C(/NCCCC)C1=CC=C(C=C1N=2)ON	2.0	1	1	1
	NC=2/N=C(/NCCCC)C1=CC=C(C=C1N=2)C(=O)N(C)C	0.5	1	1	1
	NC=2/N=C(/NCCCC)C1=CC=C(C=C1N=2)CN(C)C	0.04	1	1	1
	NC=2/N=C(/NCCCC)C1=CC([F])=CC=C1N=2	0.1	1	1	1
	NC=2/N=C(/NCCCC)C1=CC(=CC=C1N=2)OC	1.8	1	-1	1
	NC=2/N=C(/NCCCC)C1=CC(=CC=C1N=2)CO	0.2	1	1	1
	NC=2/N=C(/NCCCC)C1=CC(=CC=C1N=2)C#N	7.1	1	1	1
	NC=2/N=C(/NCCCC)C1=CC(=CC=C1N=2)C(=O)C	1.9	1	1	1
	NC=2/N=C(/NCCCC)C1=CC(=CC=C1N=2)[C@@](C)(O)[H]	5.8	1	-1	1
	NC=2/N=C(/NCCCC)C1=CC(=CC=C1N=2)CC/C3=C/C=C(C=C3)N(C)C	0.5	1	-1	1
	NC=2/N=C(/NCCCC)C1=CC(=CC=C1N=2)CC/C3=C/C=CC=N3	0.3	1	-1	1
	NC=2/N=C(/NCCCC)C1=CC([F])=C([F])C=C1N=2	0.1	1	1	1

	NC=2/N=C(/NCCCC)C1=CC(OC)=C(C=C1N=2)OC	0.1	1	1	1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2C	0.3	1	1	-1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2OC	0.1	1	-1	1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2CC/C3=C/C=CC=C3	0.47	1	-1	-1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2OCCOC	0.2	1	-1	1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2OC/C3=C/C=CC=C3	1.66	1	-1	1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2OC/C3=C/C=CC=N3	18.5	1	-1	1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2OC/C3=C/C=CN=C3	0.53	1	-1	1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2C(=O)NC	0.39	1	1	-1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2C(=O)NC(C)(C)[H]	15.55	1	-1	-1
	NC=1/N=C(/NCCCC)C=2C(N=1)=CC=CC=2C(=O)NC/C3=C/C=CC=C3	1.6	1	-1	-1

50. M. Beesu, H. P. Kokatla and S. A. David, *Methods Mol Biol*, 2017, **1494**, 29-44.
51. P. Larson, T. A. Kucaba, Z. Xiong, M. Olin, T. S. Griffith and D. M. Ferguson, *ACS medicinal chemistry letters*, 2017, **8**, 1148-1152.
52. S. Pieters, D. McGowan, F. Herschke, F. Pauwels, B. Stoops, S. Last, W. Embrechts, A. Scholliers, W. Mostmans, K. Van Dijck, B. Van Schoubroeck, T. Thoné, D. De Pooter, G. Fanning, M. L. Rosauro, M. D. Khamlichi, I. Houpis, E. Arnoult, T. H. M. Jonckers and P. Raboisson, *Bioorganic & medicinal chemistry letters*, 2018, **28**, 711-719.

Table S7.

The prediction performances of the ECP-Surflex model

	Accuracy (%)					
Data set	Training set	Validation set	External test set	Blind test set (%)		
Sample information	132 samples (61 agonists/71 non-agonists)	79 samples (36 agonists/43 non-agonists)	75 samples (52 agonists/23 non-agonists)	4 agonists	8 agonists	34 agonists
Reference	9, 24-27, 29, 32-36	9, 24-27, 29, 32-36	22, 23	50	51	52
ECP	83.3	81.0	80.0	100.0	75.0	64.7
Surflex	69.7	72.2	81.3	100.0	50.0	85.3
ECP-Surflex	86.9	86.2	94.2	100.0	66.7	87.0