Supplementary Information

Compound Pathway Model to Capture SAR Progression: Comparison of Activity Cliff-Dependent and -Independent Pathways

Dagmar Stumpfe, Dilyana Dimova, Kathrin Heikamp, and Jürgen Bajorath*

Supplementary Table S1

Table S1. Compound type statistics.

No.	ChEMBL Target ID	Target	#Cpds	#hpCPs				#C	#D						Potency (pK_i)		
				all	B	in D	acPaths		all	singl.	w/o B and C	with only B	with B and C	with only C	Min	Max	Range
1	4617	Phenylethanolamine N-methyltransfe-rase	148	29	3	5	21	8	14	1	2	10	1	0	3.05	8.54	5.49
2	249	Neurokinin 1 receptor	211	43	1	3	39	18	21	13	0	7	0	1	3.77	11.96	8.19
3	299	Protein kinase C alpha	168	31	5	7	19	18	16	2	1	9	2	2	4.47	9.03	4.56
4	2243	Anandamide amidohydrolase	101	26	3	0	23	9	10	0	3	1	0	6	4.92	10.03	5.11
5	248	Leukocyte elastase	192	10	2	1	7	22	19	0	5	0	7	7	2.77	10.66	7.89
6	3795	Melanocortin receptor 1	134	11	3	5	3	11	13	5	3	5	0	0	4.63	9.77	5.14
7	211	Muscarinic acetylcholine receptor M2	199	18	2	4	12	18	19	7	7	5	0	0	4.96	10.38	5.42
8	3837	Cathepsin L	201	26	5	8	13	21	20	4	3	11	0	2	3.12	10.60	7.48
9	268	Cathepsin K	272	38	6	16	16	32	27	1	8	18	0	0	3.87	11.52	7.66
10	1997	Equilibrative nucleoside transporter 1	118	11	2	4	5	4	11	3	4	4	0	0	5.31	10.74	5.43
11	1855	Gonadotropin-releasing hormone receptor	267	56	15	12	29	16	26	1	3	12	5	5	3.39	9.70	6.31
12	4822	Beta-secretase 1	116	17	3	2	12	7	11	5	4	2	0	0	4.08	9.52	5.44

13	5071	G protein-coupled receptor 44	376	44	7	14	23	31	37	3	3	18	5	8	4.48	9.10	4.62
14	3759	Histamine H4 receptor	334	39	8	13	18	31	33	1	6	13	7	6	3.42	10.40	6.97
15	4561	Neuropeptide Y receptor type 5	234	43	5	7	31	21	23	1	1	11	2	8	5.22	9.70	4.48
16	213	Beta-1 adrenergic receptor	175	21	6	4	11	17	17	4	2	7	1	3	3.61	10.00	6.39
17	344	Melanin-concentrating hormone receptor 1	870	104	21	31	52	90	87	2	7	32	28	18	3.57	10.47	6.90
18	1889	Vasopressin V1a receptor	318	34	5	8	21	34	31	11	0	13	7	0	3.27	10.30	7.03
19	2954	Cathepsin S	371	71	7	18	46	32	37	1	14	14	2	6	3.53	9.82	6.29
20	2014	Nociceptin receptor	599	111	15	5	91	58	59	16	0	14	12	17	5.00	10.70	5.70
21	4308	Bradykinin B1 receptor	415	100	15	14	71	28	41	3	10	19	3	6	3.62	11.00	7.38
22	219	Dopamine D4 receptor	436	12	4	3	5	73	43	1	13	8	0	21	4.05	9.60	5.56
23	244	Coagulation factor X	1198	301	26	48	227	108	119	1	35	25	26	32	3.59	11.40	7.81
24	245	Muscarinic acetylcholine receptor M3	343	46	9	6	31	30	34	5	6	10	0	13	4.82	11.00	6.18
25	234	Dopamine D3 receptor	881	94	15	27	52	125	88	7	18	26	12	25	4.17	10.01	5.84
26	214	Serotonin 1a (5-HT1a) receptor	938	45	9	25	11	109	93	2	24	16	2	49	2.54	10.85	8.31
27	1800	Corticotropin releasing factor receptor 1	477	71	11	8	52	38	47	6	3	14	9	15	5.11	9.70	4.59
28	228	Serotonin transporter	1099	35	15	9	11	123	109	5	17	21	3	63	2.04	10.52	8.48

29	204	Thrombin	808	75	21	20	34	82	80	4	19	23	13	21	2.84	12.19	9.35
30	259	Melanocortin receptor 4	1273	138	28	29	81	80	127	5	5	51	45	21	3.65	9.40	5.74
31	231	Histamine H1 receptor	187	19	1	0	18	30	18	5	3	1	0	9	4.10	9.72	5.62
32	3371	Serotonin 6 (5-HT6) receptor	888	18	7	5	6	136	88	9	21	3	1	54	1.38	9.85	8.48
33	3798	Calcitonin gene-related peptide type 1 receptor	246	30	5	1	24	29	24	5	1	2	0	16	5.00	11.00	6.00
34	284	Dipeptidyl peptidase IV	276	18	7	5	6	33	27	1	5	8	0	13	3.22	10.52	7.31
35	1836	Prostanoid EP4 receptor	194	26	5	7	14	18	19	0	7	3	0	9	4.89	10.70	5.81
36	264	Histamine H3 receptor	1515	121	27	15	79	176	151	9	31	9	8	94	4.49	10.51	6.02
37	1945	Melatonin receptor 1A	215	9	2	6	1	28	21	1	9	0	0	11	4.98	11.19	6.21
38	1946	Melatonin receptor 1B	262	19	7	4	8	20	26	1	8	1	0	16	5.10	12.26	7.16
39	1914	Butyrylcholinesterase	159	36	9	1	26	14	15	0	2	0	0	13	1.92	9.57	7.65

For all 39 data sets, the ChEMBL Target ID and the number of compounds (#Cpds) are provided. In addition, the distribution of highly potent activity cliff partners (#hpCPs), *C* and *D* compounds are reported. Only hpCPs are considered that yield qualifying pathways with positive potency progression. These hpCPs are subdivided into *B* compounds, *D* compounds, and compounds in activity cliff-dependent pathways (acPaths). Furthermore, *D* compounds are partitioned into singletons, undetected *D*

compounds (*D* w/o *B* and *C*), *D* compounds detected only by activity cliff-dependent pathways (#D from only B), *D* compounds detected only by activity cliff-independent pathways (#D from only C), and *D* compounds in merging pathways (#D from B and C). For each data set, the lowest compound potency (Min), highest potency (Max), and the potency range are listed as pK_i values.