Supporting Information

for

Development of R-group Fingerprints Based on the Local Landscape from an Attachment Point of a Molecular Structure

Shunsuke Tamura,[†] Tomoyuki Miyao,[†,‡] and Kimito Funatsu[*,†,‡,§]

[†]Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan

[‡]Data Science Center, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan

[§]Department of Chemical System Engineering, School of Engineering, The University of Tokyo, 7-3-1 Hongo. Bunkyo-ku, Tokyo 113-8656, Japan

*To whom correspondence should be addressed:

Tel: +81-3-5841-7751, Fax: +81-3-5841-7771, E-mail: funatsu@chemsys.t.u-tokyo.ac.jp

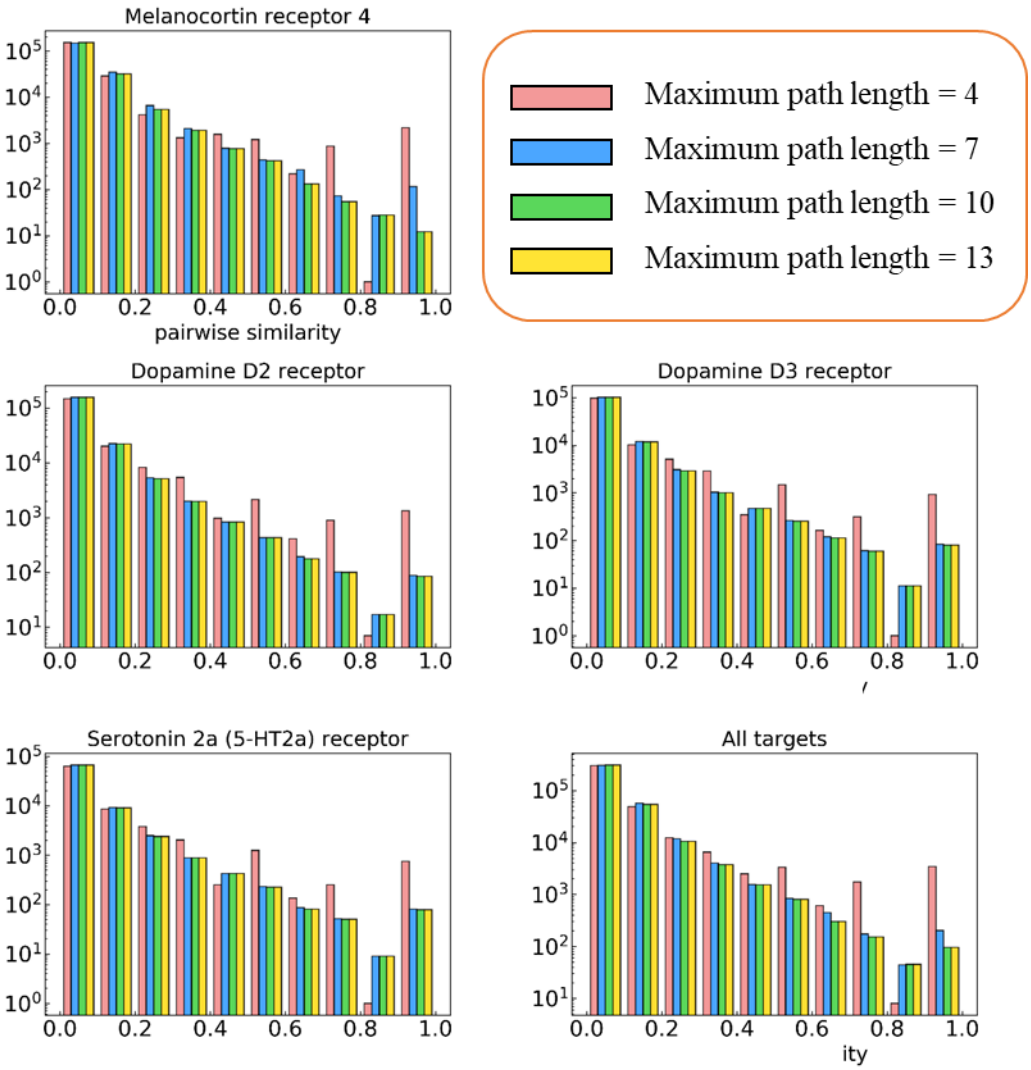

Figure S1. Distributions of the pairwise Tanimoto similarity values with different maximum path lengths. Reported are the histograms of the pairwise Tanimoto similarity values for compounds in a target dataset using R-path fingerprints with different maximum path lengths. The maximum path lengths varied from four to 13. The minimum sub-path length for the R-path fingerprints was set to one.

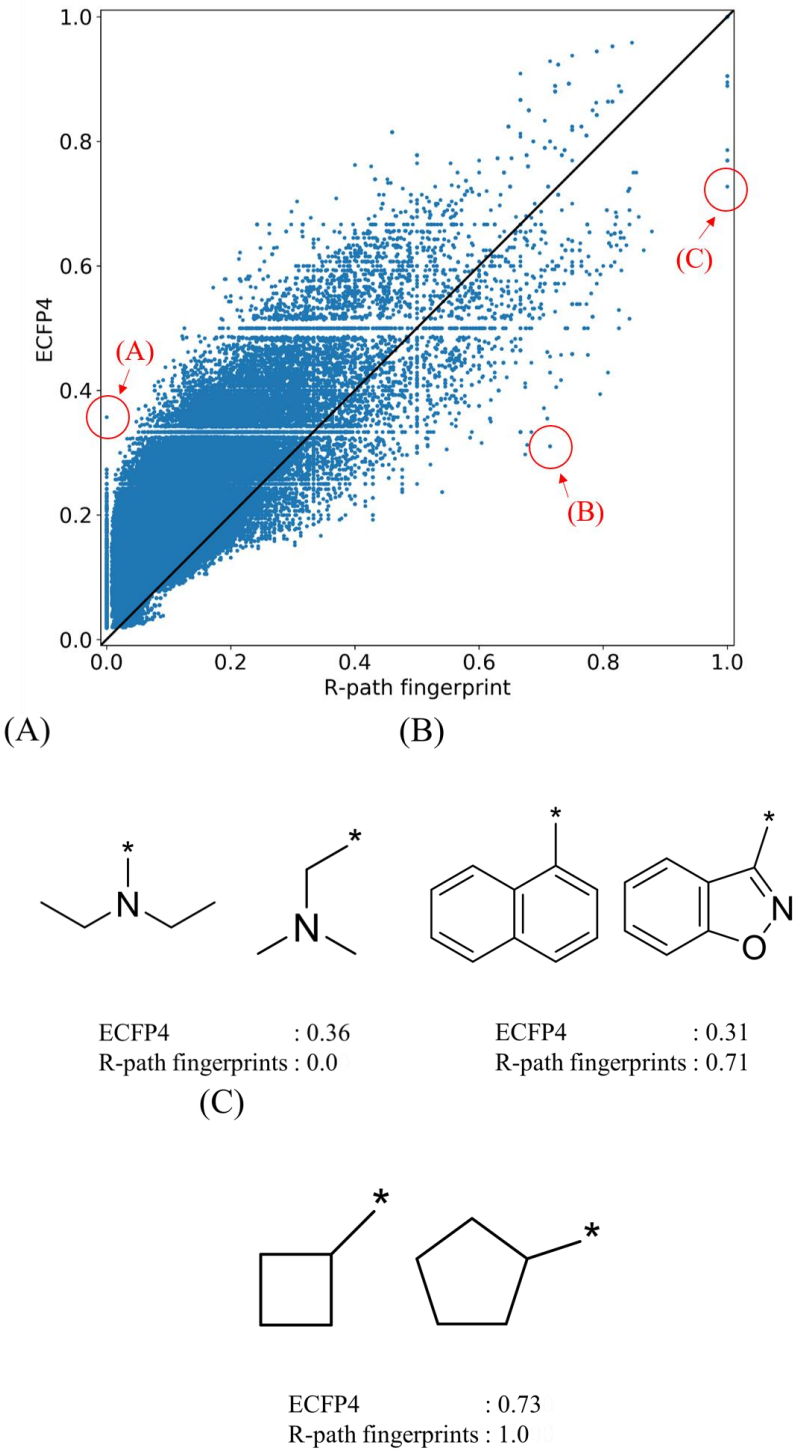

Figure S2. Comparison of the pairwise Tanimoto similarity between R-path fingerprints and ECFP4. Reported is a scatter plot of the pairwise Tanimoto similarities between R-path fingerprints and ECFP4. Representative pairs, A, B and C, with similarity values are shown below.

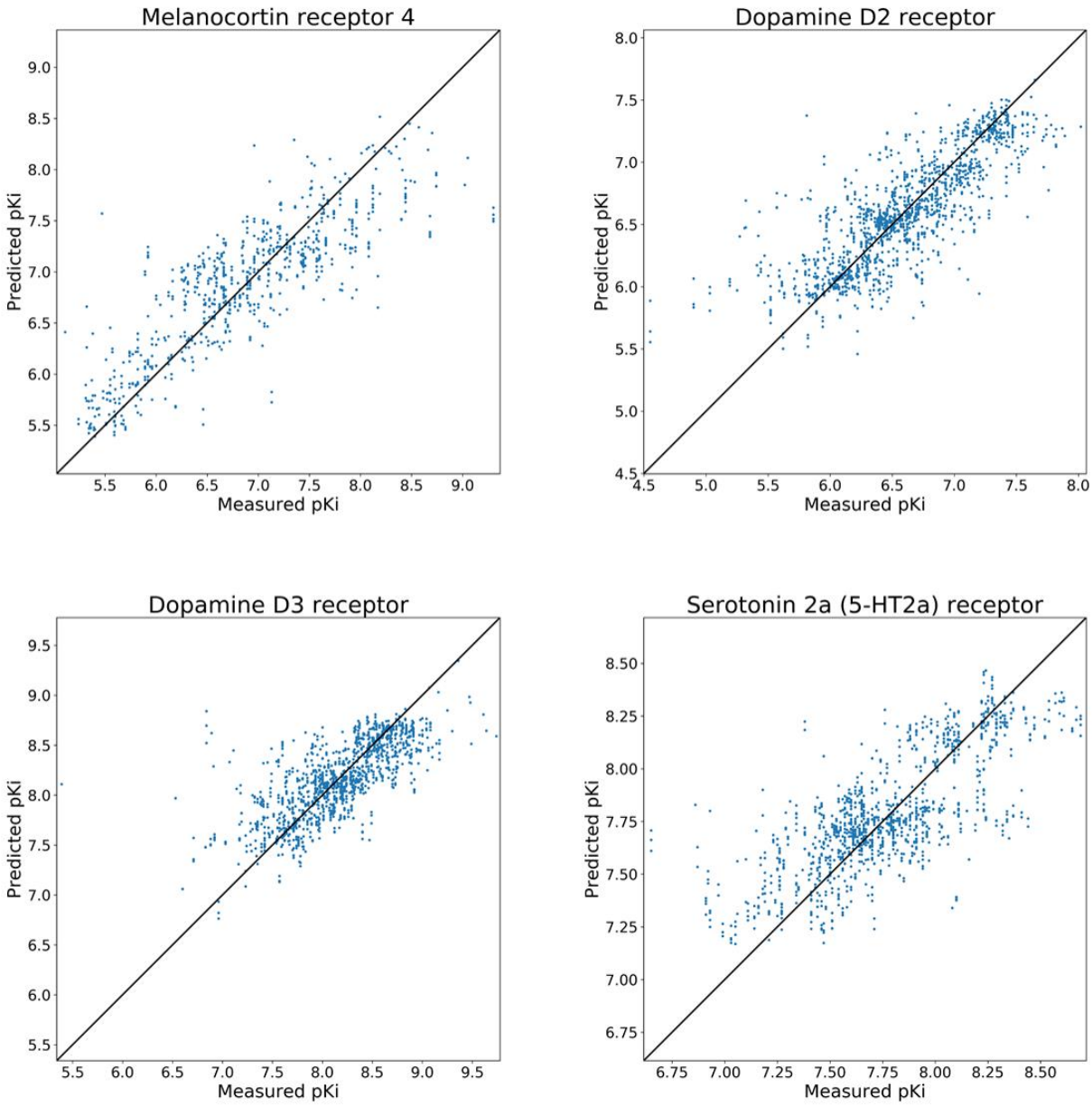

Figure S3. Measured versus predicted plots for each target with ECFP4.

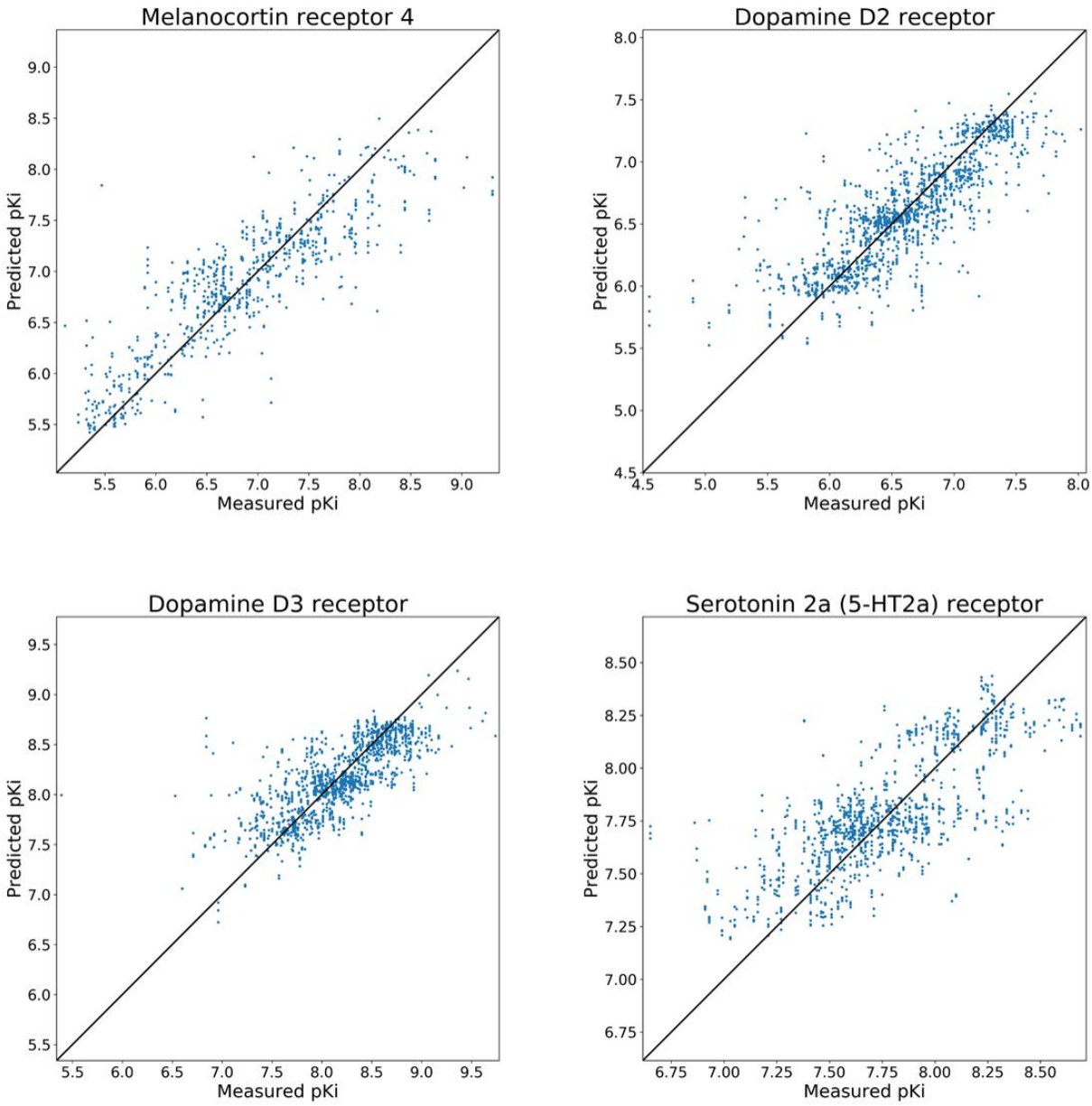

Figure S4. Measured versus predicted plots for each target with R-path fingerprints.

Table S1. Core Structures for Melanocortin receptor 4 (A), Dopamine D2 receptor (B), Dopamine D3 receptor (C) and Serotonin 2a (5-HT2a) receptor (D).

A

MMS Number	SMILES (core)
0	[R1]c1ccccc1N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3)Cl)NC(=O)[C@H]4Cc5ccccc5CN4
1	[R1]Cc1ccccc1N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3)Cl)NC(=O)[C@H]4Cc5ccccc5CN4
2	[R1]N1CCN(CC1)C(=O)[C@@H](Cc2ccc(cc2)Cl)NC(=O)[C@H]3Cc4ccccc4CN3
3	[R1]NCC1(CCCCC1)N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3Cl)Cl)NC(=O)CCN
4	[R1]C(=O)NCC1(CCCCC1)N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3Cl)Cl)NC(=O)CCN
5	[R1]CC1(CCCCC1)N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3Cl)Cl)NC(=O)CCN
6	[R1]C1(CCCCC1)N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3Cl)Cl)NC(=O)CCN
7	[R1]C(=O)N[C@H](Cc1ccc(cc1)Cl)C(=O)N2CCN(CC2)C3(CCCCC3)CNC(=O)Cc4ccccc4
8	[R1][C@H](Cc1ccc(cc1)Cl)C(=O)N2CCN(CC2)C3(CCCCC3)CNC(=O)Cc4ccccc4
9	[R1]NCc1cccnc1N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3Cl)Cl)NC(=O)CCN
10	[R1]c1cccnc1N2CCN(CC2)C(=O)[C@@H](Cc3ccc(cc3Cl)Cl)NC(=O)CCN
11	[R1]C(=O)N1CCN(CC1)c2ccc(cc2[C@H](CC(C)C)N)C(F)(F)F
12	[R1]CC(=O)N1CCN(CC1)c2ccc(cc2[C@H](CC(C)C)N)C(F)(F)F
13	[R1]c1nc2ccc(cc2n1CCCN3CCCCC3)C(=O)N(CCC(C)C)CCC(C)C
14	[R1]CC1(CCCCC1)N2CCN(CC2)C(=O)[C@H]3CN(C[C@@H]3c4ccc(cc4)Cl)C(C)C
15	[R1]C1(CCCCC1)N2CCN(CC2)C(=O)[C@H]3CN(C[C@@H]3c4ccc(cc4)Cl)C(C)C
16	[R1]NCC1(CCCCC1)N2CCN(CC2)C(=O)[C@H]3CN(C[C@@H]3c4ccc(cc4)Cl)C(C)C
17	[R1]CNCC1(CCCCC1)N2CCN(CC2)C(=O)[C@H]3CN(C[C@@H]3c4ccc(cc4)Cl)C(C)C
18	[R1]c1ccccc1N2CCN(CC2)C(=O)[C@@H]3CN(C[C@H]3c4ccc(cc4)Cl)C(C)C

B

MMS Number	SMILES (core)
0	[R1]C(=O)NCCCCN1CCN(CC1)c2ccccc2OC
1	[R1]CCCCN1CCN(CC1)c2cc(nc(n2)C(C)(C)C)C(F)(F)F
2	[R1]C(=O)NN1c2ccccc2C(=Nc3c1ccc(c3)Cl)N4CCN(CC4)C
3	[R1]N1c2ccccc2C(=Nc3c1ccc(c3)Cl)N4CCN(CC4)C
4	[R1]NN1c2ccccc2C(=Nc3c1ccc(c3)Cl)N4CCN(CC4)C
5	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
6	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
7	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
8	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
9	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4

10	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
11	[R1][C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
12	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
13	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
14	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
15	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
16	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
17	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
18	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
19	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
20	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
21	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
22	[R1][C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
23	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
24	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4

C

MMS Number	SMILES (core)
0	[R1]C(=O)NCCCCN1CCN(CC1)c2ccccc2OC
1	[R1]CCCCN1CCN(CC1)c2cc(nc(n2)C(C)(C)C)C(F)(F)F
2	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
3	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
4	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
5	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
6	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
7	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
8	[R1][C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
9	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
10	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
11	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
12	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
13	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
14	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
15	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
16	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
17	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
18	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4

19	[R1][C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
20	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
21	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4

D

MMS Number	SMILES (core)
0	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
1	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
2	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
3	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(oc3)cccc4
4	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
5	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
6	[R1][C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
7	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3nccc4c3CCO4
8	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
9	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
10	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
11	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCO4
12	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
13	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
14	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4
15	[R1]N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
16	[R1]C(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
17	[R1][C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
18	[R1]CC(=O)N[C@H]1CC[C@@H](CC1)CCN2CCN(CC2)c3c4c(ccc3)OCO4
19	[R1][C@H]1CC[C@@H](CC1)CCN2CCC(CC2)c3c4c(ccc3)OCC4

Table S2. Profiles of MMS for Melanocortin receptor 4 (A), Dopamine D2 receptor (B), Dopamine D3 receptor (C) and Serotonin 2a (5-HT2a) receptor (D).

A

MMS Number	Number of CPDs	Potency (pK_i)		MW (sub)		Number of Heavy Atoms (sub)	
		Max	Min	Max	Min	Max	Min
0	45	8.19	6.31	175.25	30.05	13	2
1	34	7.95	6.31	185.24	56.09	13	4
2	31	8.40	5.11	182.33	106.15	13	8
3	44	8.44	6.27	189.11	57.11	13	4
4	29	8.44	6.27	161.1	67.07	11	4
5	41	8.44	6.27	199.02	72.13	13	5
6	37	8.44	6.31	213.05	30.05	13	2
7	32	7.96	6.00	146.21	29.06	11	2
8	31	7.96	6.00	175.21	58.06	13	4
9	33	8.17	5.55	164.18	43.04	12	3
10	32	8.17	5.55	181.19	72.09	13	5
11	66	8.19	5.24	197.70	106.15	12	7
12	41	7.13	5.24	178.06	92.12	11	6
13	32	9.05	6.96	175.25	74.10	13	5
14	49	9.30	5.57	190.14	88.13	13	6
15	49	9.30	5.57	188.17	30.05	13	2
16	42	9.30	5.92	175.13	73.11	12	5
17	38	9.30	5.92	161.10	67.07	11	5
18	30	7.04	5.32	157.23	26.02	11	2

B

MMS Number	Number of CPDs	Potency (pK_i)		MW (sub)		Number of Heavy Atoms	
		Max	Min	Max	Min	Max	Min
0	40	8.02	5.81	243.02	29.06	10	2
1	48	7.76	6.03	179.08	95.08	12	7
2	59	7.11	4.55	203	43.09	13	3
3	42	7.2	4.55	246.03	86.11	11	6
4	47	6.74	4.55	231.01	71.1	12	5
5	78	7.82	5.95	175.25	29.06	13	2
6	73	7.82	5.95	189.23	29.02	12	2
7	58	7.82	6.69	189.02	44.03	11	3

8	29	7.47	6.61	121.11	26.02	9	2
9	82	6.7	5.47	189.23	29.02	12	2
10	86	6.79	5.27	175.25	29.06	13	2
11	73	6.7	5.59	189.02	44.03	11	3
12	38	6.39	5.52	121.11	26.02	9	2
13	72	7.28	5.57	162.21	29.06	12	2
14	78	7.28	5.57	163.15	43.04	12	3
15	72	7.28	5.57	174.17	58.06	11	4
16	33	6.98	6.27	121.11	26.02	9	2
17	72	7.16	5.43	175.25	29.06	13	2
18	70	7.16	5.84	189.23	29.02	12	2
19	31	6.86	5.97	121.11	26.02	9	2
20	76	7.52	5.32	163.15	43.04	12	3
21	69	7.52	5.32	162.21	29.06	12	2
22	70	7.52	6.5	174.17	58.06	11	4
23	28	7.42	6.12	121.11	26.02	9	2
24	63	7.16	5.92	189.02	44.03	11	3

C

MMS Number	**Number of CPDs**	**Potency (pK_i)**		**MW (sub)**		**Number of Heavy Atoms**	
		Max	**Min**	**Max**	**Min**	**Max**	**Min**
0	42	9.74	6.89	259.09	29.06	10	2
1	50	9.1	5.39	179.08	95.08	12	7
2	79	9.17	6.84	175.25	29.06	13	2
3	73	9.17	6.84	189.23	29.02	12	2
4	58	9.17	7.89	189.02	44.03	11	3
5	29	9.17	6.84	121.11	26.02	9	2
6	82	8.4	6.6	189.23	29.02	12	2
7	86	8.4	6.6	175.25	29.06	13	2
8	73	8.4	6.6	189.02	44.03	11	3
9	38	8.4	7.17	121.11	26.02	9	2
10	72	8.92	7.3	162.21	29.06	12	2
11	78	8.92	7.3	163.15	43.04	12	3
12	72	8.92	7.3	174.17	58.06	11	4
13	33	8.92	7.7	121.11	26.02	9	2
14	74	8.7	7.14	175.25	29.06	13	2
15	71	8.7	7.18	189.23	29.02	12	2

16	31	8.7	7.64	121.11	26.02	9	2
17	76	9.06	7.56	163.15	43.04	12	3
18	69	9.06	7.45	162.21	29.06	12	2
19	70	9.04	7.56	174.17	58.06	11	4
20	28	9.04	8.33	121.11	26.02	9	2
21	63	8.7	7.19	189.02	44.03	11	3

D

MMS Number	**Number of CPDs**	**Potency (pK_i)**		**MW (sub)**		**Number of Heavy Atoms**	
		Max	**Min**	**Max**	**Min**	**Max**	**Min**
0	79	8.69	7.65	175.25	29.06	13	2
1	73	8.69	7.38	189.23	29.02	12	2
2	58	8.69	7.38	189.02	44.03	11	3
3	29	8.69	7.83	121.11	26.02	9	2
4	82	8.44	7.15	189.23	29.02	12	2
5	86	8.44	7.18	175.25	29.06	13	2
6	73	8.23	7.15	189.02	44.03	11	3
7	38	8.09	7.38	121.11	26.02	9	2
8	72	8.41	6.65	162.21	29.06	12	2
9	78	8.41	6.65	163.15	43.04	12	3
10	72	8.41	6.65	174.17	58.06	11	4
11	33	7.93	7.26	121.11	26.02	9	2
12	74	8.1	6.86	175.25	29.06	13	2
13	71	8.1	6.87	189.23	29.02	12	2
14	31	7.71	6.91	121.11	26.02	9	2
15	76	8.43	7.17	163.15	43.04	12	3
16	69	8.43	6.93	162.21	29.06	12	2
17	70	8.43	7.17	174.17	58.06	11	4
18	28	8.11	7.29	121.11	26.02	9	2
19	63	8.1	6.91	189.02	44.03	11	3