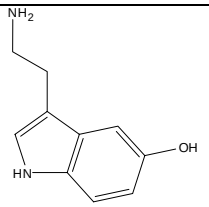
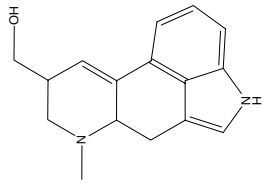
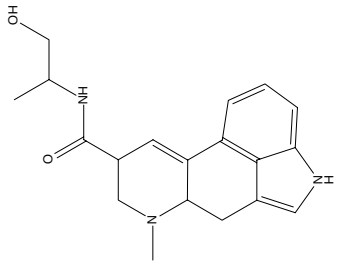
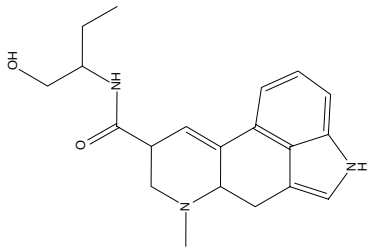
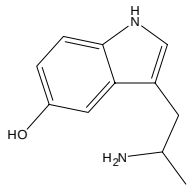
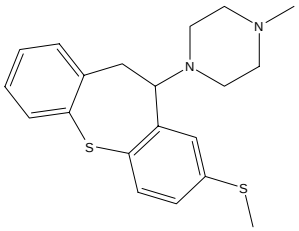
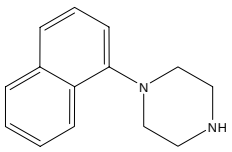
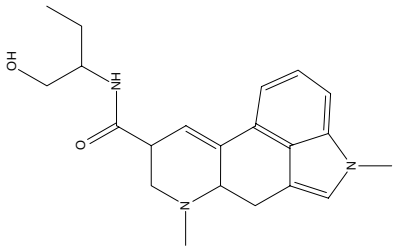
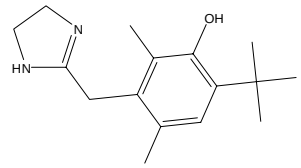
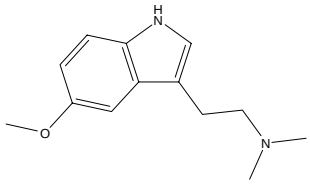
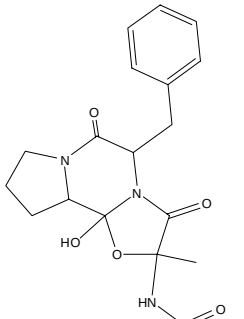
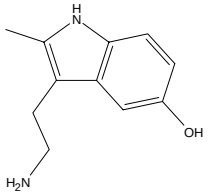
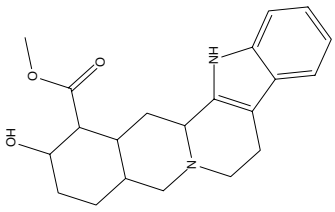
Combinatorial QSAR Modeling of Specificity and Subtype Selectivity of Ligands that Bind to Serotonin Receptors 5HT1E and 5HT1F

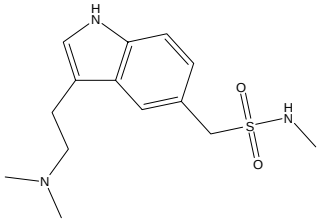
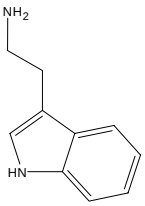
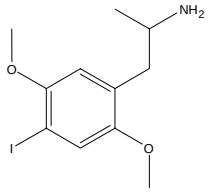
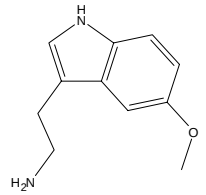
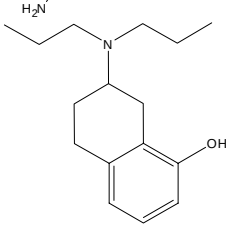
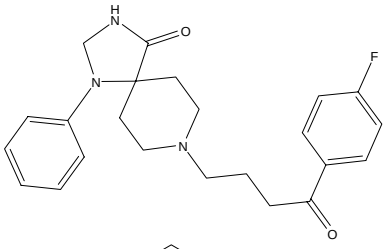
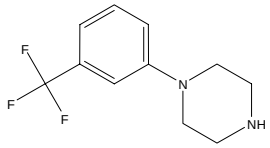
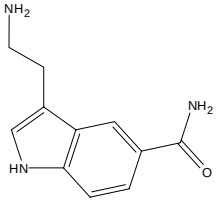
Xiang S. Wang[ǂ], Hao Tang[ǂ†], Alexander Golbraikh[ǂ], and Alexander Tropsha[ǂ*]

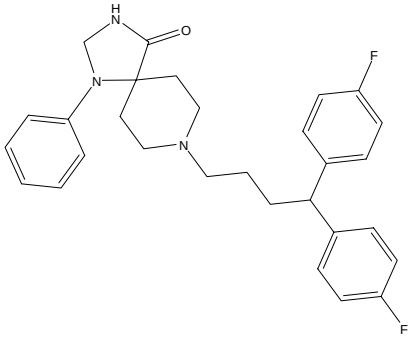
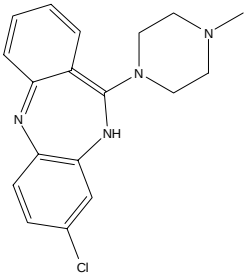
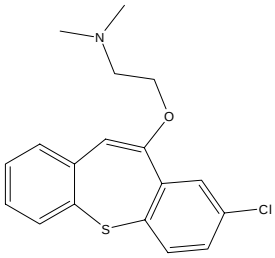
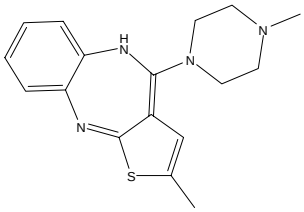
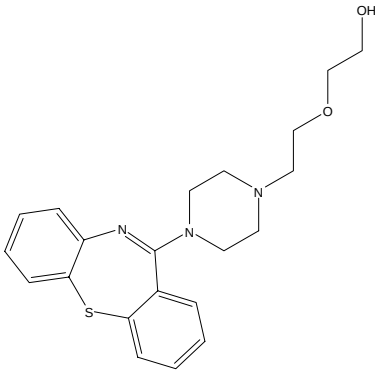
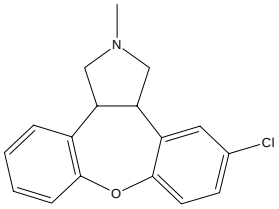
Supporting Information

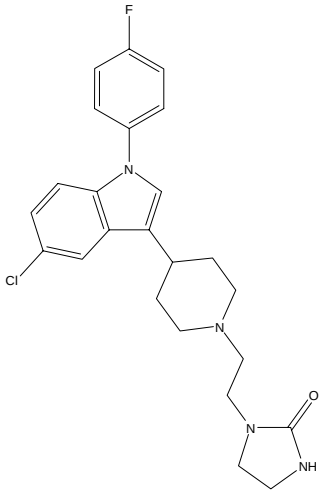
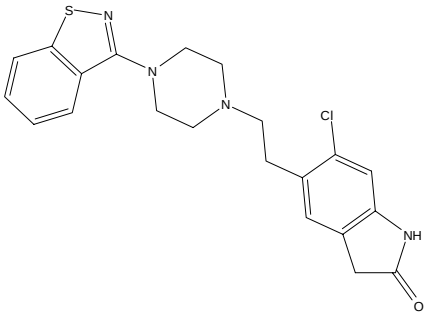
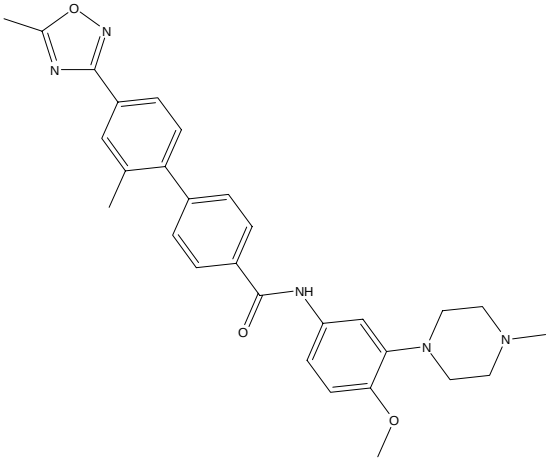
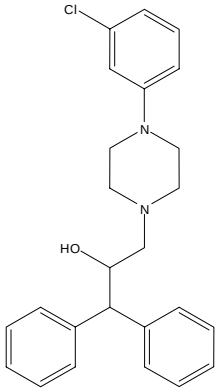
Table 1. The 5HT1E Dataset from PDSP Ki Database Used to Build the *k*NN QSAR Models.

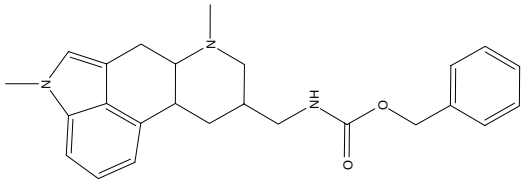
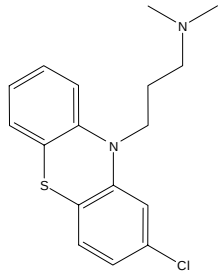
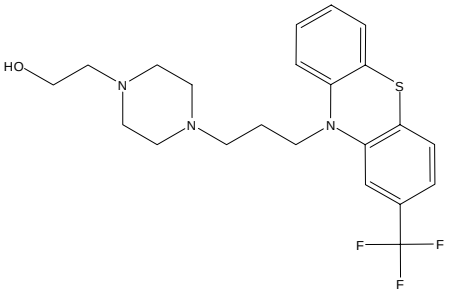
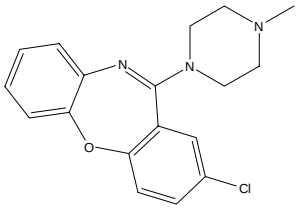
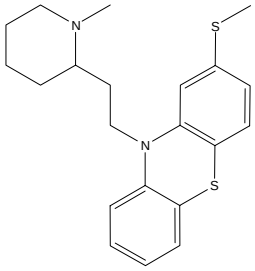
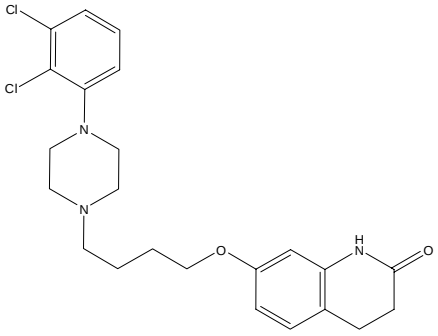
Num.	Structure	PDSP ID	Name	Ki(nM)	Reference
1		6637	5-Hydroxy Tryptamine	7.53	Zgombick JM, et al., 1992
2		6638	Lysergol	42.83	Zgombick JM, et al., 1992
3		6639	Ergonovine	87.55	Zgombick JM, et al., 1992
4		6640	Methylergonovine	89.06	Zgombick JM, et al., 1992
5		6641	5-HT,a-Me	120.61	Zgombick JM, et al., 1992
6		6642	Methiothepin	179.53	Zgombick JM, et al., 1992
7		6643	1-NP	207.96	Zgombick JM, et al., 1992

8		6644	Methysergide	247.29	Zgombick JM, et al., 1992
9		6645	Oxymetazoline	417.93	Zgombick JM, et al., 1992
10		6646	DMT,5-Me	526.4	Zgombick JM, et al., 1992
11		6647	Ergotamine	466.6	Zgombick JM, et al., 1992
12		6648	5-HT,2-Me	814.91	Zgombick JM, et al., 1992
13		6649	Yohimbine	1264.46	Zgombick JM, et al., 1992

14		6650	Sumatriptan	2366.35	Zgombick JM, et al., 1992
15		6651	Tryptamine	2,559.00	Zgombick JM, et al., 1992
16		6652	DOI	2,970.00	Zgombick JM, et al., 1992
17		6653	5-Methoxytryptamine	3,151.00	Zgombick JM, et al., 1992
18		6654	DPAT	3,333.00	Zgombick JM, et al., 1992
19		6656	Spiperone	5,051.00	Zgombick JM, et al., 1992
20		6657	TFMPP	6301.28	Zgombick JM, et al., 1992
21		6658	5-CT	6552.26	Zgombick JM, et al., 1992

22		11343	Fluspirilene	2,230.00	Schotte A, et al., 1996
23		11344	Clozapine	698	Schotte A, et al., 1996
24		11345	Zotepine	700	Schotte A, et al., 1996
25		11346	Olanzapine	2209	Schotte A, et al., 1996
26		11347	Quetiapine	1826	Schotte A, et al., 1996
27		11348	ORG-5222	24.2	Schotte A, et al., 1996

28		11349	Sertindole	430	Schotte A, et al., 1996
29		11350	Ziprasidone	819.5	Schotte A, et al., 1996
30		11921	GR 127935	3,981.07	Price GW, et al., 1997
31		11923	BRL 15572	6,309.57	Price GW, et al., 1997

32		15504	Metergoline	949.13	Boess FG, et al., 1994
33		24812	Chlorpromazine	344	PDSP certified data
34		24911	Fluphenazine	540	PDSP certified data
35		24978	Loxapine	1,399.00	PDSP certified data
36		25181	Thioridazine	194	PDSP certified data
37		25385	Aripiprazole	8,000.00	PDSP certified data

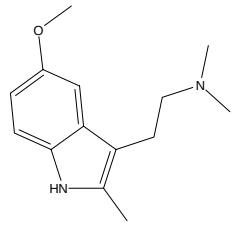
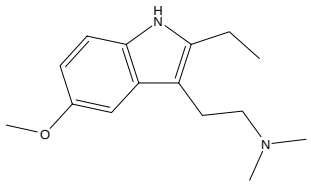
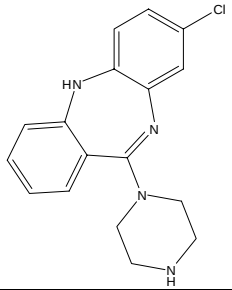
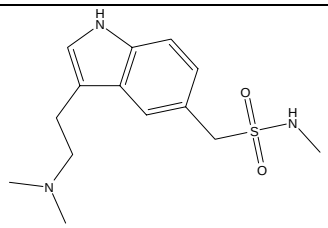
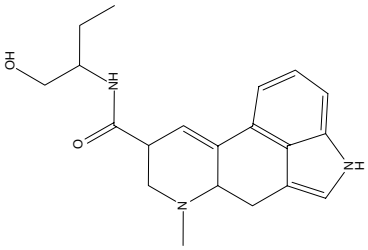
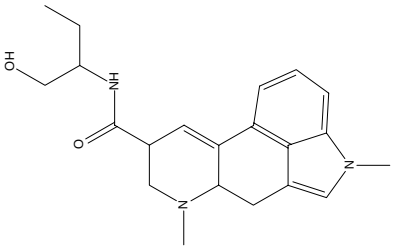
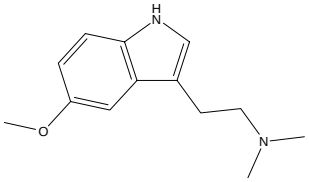
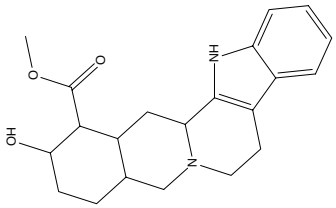
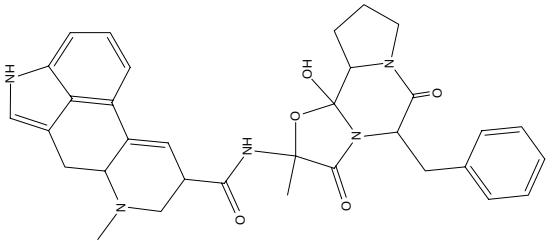
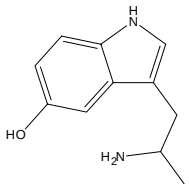
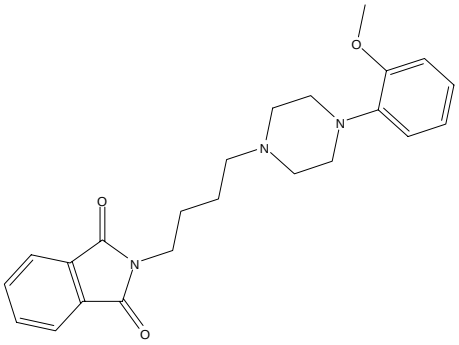
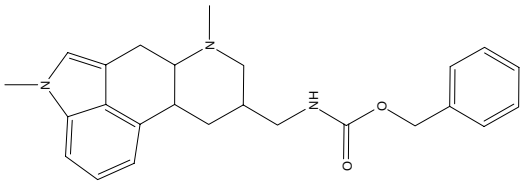
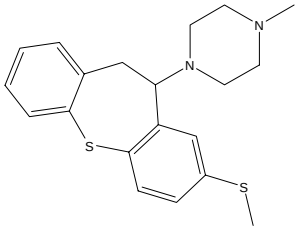
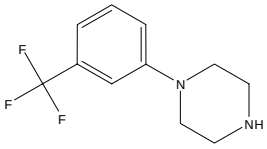
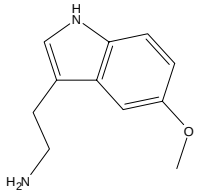
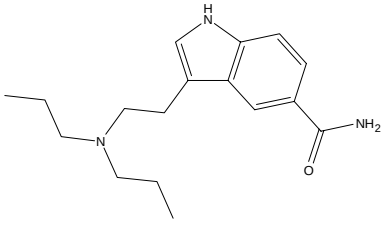
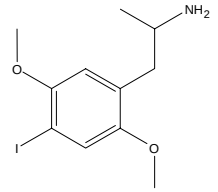
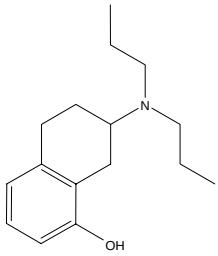
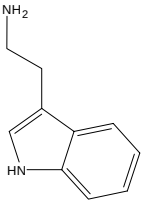
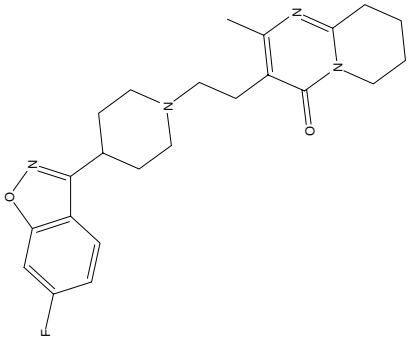
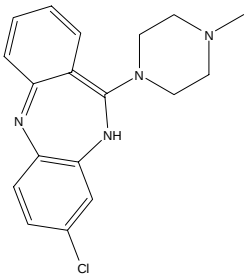
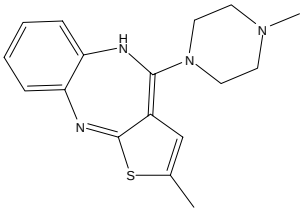
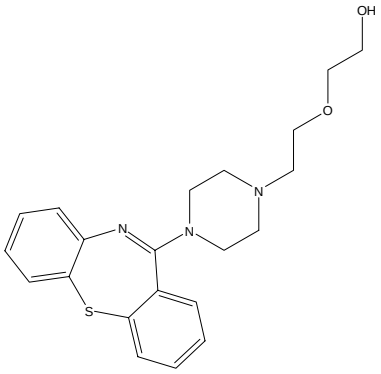
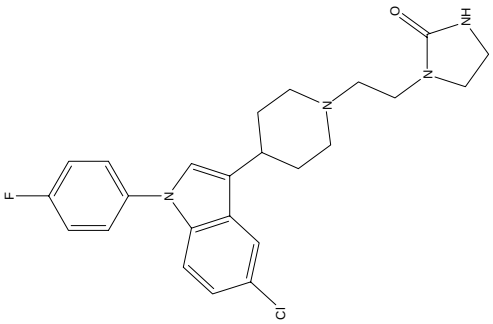
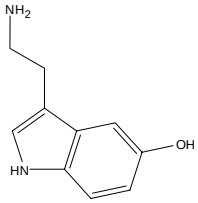
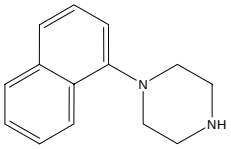
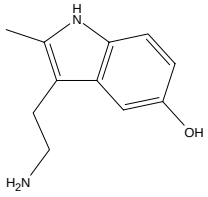
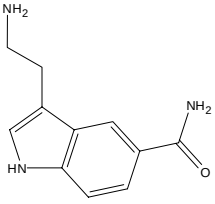
38		27973	DMT,2-Me-5-methoxy	1,800.00	Glennon RA, et.al., 2000
39		27974	EMDT	520	Glennon RA, et.al., 2000
40		41613	N-Desmethylclozapine	393.4	PDSP certified data

Table 2. The 5HT1F Dataset from PDSP Ki Database Used to Build the *k*NN QSAR Models.

Num.	Structure	PDSP ID	Name	Ki(nM)	Reference
1		6750	Sumatriptan	22.95	Adham N, et al., 1992 [6]
2		6751	Methylergonovine	30.95	Adham N, et al., 1992 [6]
3		6752	Methysergide	33.94	Adham N, et al., 1992 [6]
4		6753	DMT,5-Me	37.08	Adham N, et al., 1992 [6]
5		6755	Yohimbine	91.60	Adham N, et al., 1992 [6]
6		6756	Ergotamine	170.41	Adham N, et al., 1992 [6]

7		6757	5-HT,a-Me	182.99	Adham N, et al., 1992 [6]
8		6758	NAN-190	203.58	Adham N, et al., 1992 [6]
9		6759	Metergoline	339.92	Adham N, et al., 1992 [6]
10		6761	Methiothepin	648.83	Adham N, et al., 1992 [6]
11		6763	TFMPP	1001.00	Adham N, et al., 1992 [6]
12		6764	5-Methoxytryptamine	1166.00	Adham N, et al., 1992 [6]
13		6765	5-CT,DP	1613.00	Adham N, et al., 1992 [6]

14		6766	DOI	1739.00	Adham N, et al., 1992 [6]
15		6767	8-OH-DPAT	1772.00	Adham N, et al., 1992 [6]
16		6768	Tryptamine	2409.00	Adham N, et al., 1992 [6]
17		11351	Risperidone	1240.00	Schotte A, et al., 1996 [12]
18		11354	Clozapine	130.00	Schotte A, et al., 1996 [12]
19		11355	Olanzapine	310.00	Schotte A, et al., 1996 [12]

20		11356	Quetiapine	2240.00	Schotte A, et al., 1996 [12]
21		11357	Sertindole	360.00	Schotte A, et al., 1996 [12]
22		15510	5-Hydroxy Tryptamine	10.00	Boess FG, et al., 1994 [7]
23		15515	1-NP	53.85	Boess FG, et al., 1994 [7]
24		15521	5-HT,2-Me	414.93	Boess FG, et al., 1994 [7]
25		15523	5-CT	720.71	Boess FG, et al., 1994 [7]

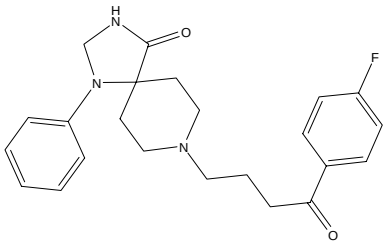
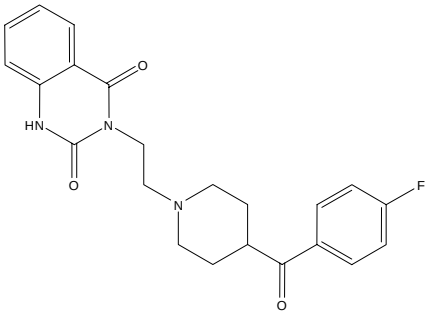
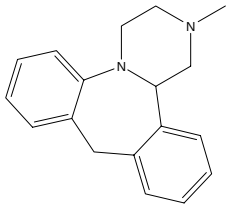
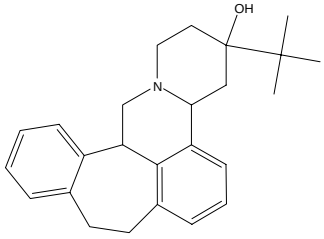
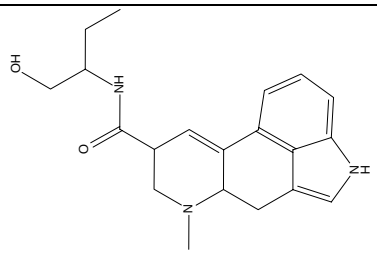
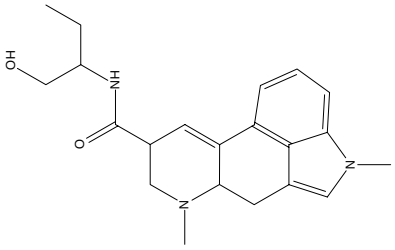
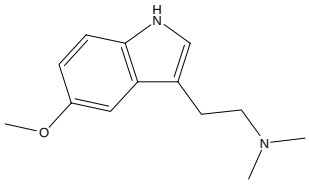
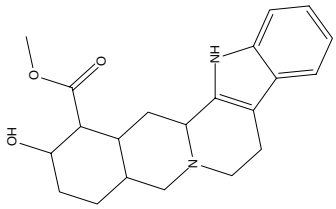
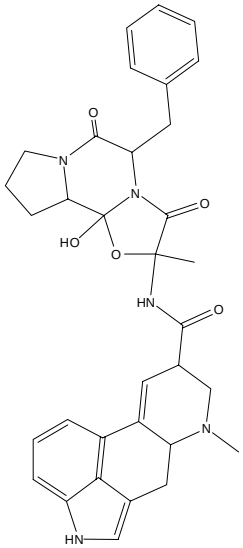
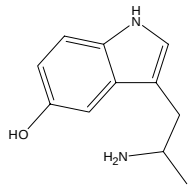
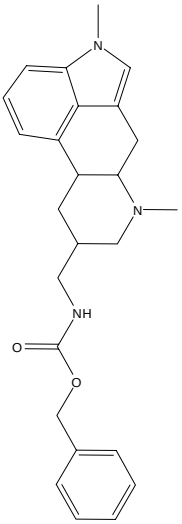
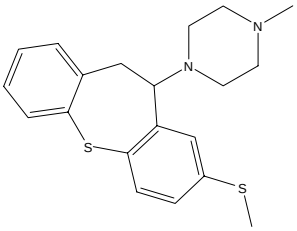
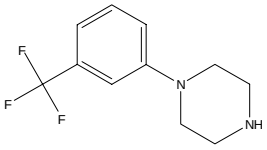
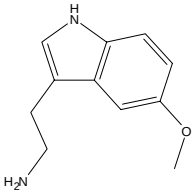
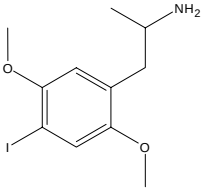
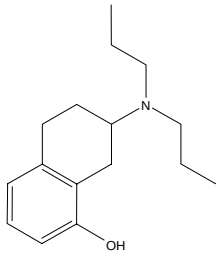
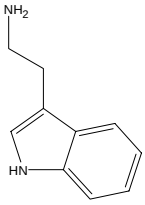
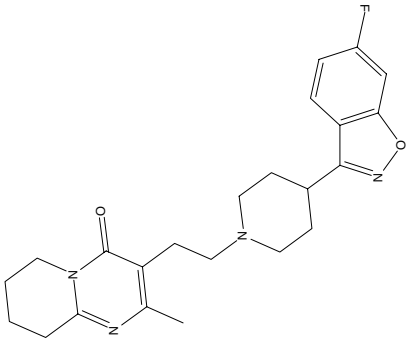
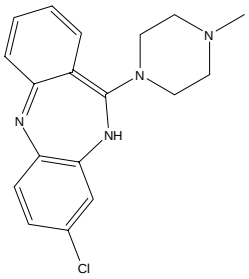
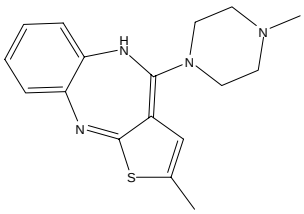
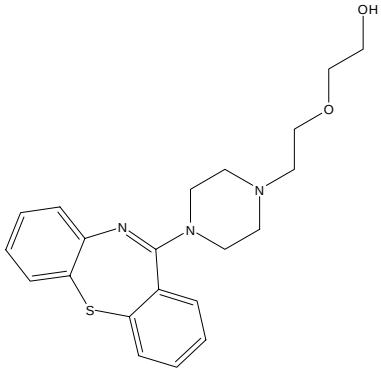
26		15911	Spiperone	3.98	Boess FG, et al., 1994 [7]
27		15912	Ketanserin	1.25	Boess FG, et al., 1994 [7]
28		15913	Mianserin	12.58	Boess FG, et al., 1994 [7]
29		15916	Butaclamol, (-)	3890.45	Boess FG, et al., 1994 [7]

Table 3. The 5HT1E/5HT1F selectivity Dataset from PDSP Ki Database Used to Build the *k*NN Models.

Num.	Structure	PDSP ID	Name	5-HT1E Ki(nM)	5-HT1F Ki(nM)	Reference
1		6751	Methylergonovine	89.06	30.95	Adham N, et al., 1992 [6]
2		6752	Methysergide	247.29	33.94	Adham N, et al., 1992 [6]
3		6753	DMT,5-Me	526.40	37.08	Adham N, et al., 1992 [6]
4		6755	Yohimbine	1264.46	91.60	Adham N, et al., 1992 [6]
5		6756	Ergotamine	466.60	170.41	Adham N, et al., 1992 [6]

6		6757	5-HT,a-Me	120.61	182.99	Adham N, et al., 1992 [6]
7		6759	Metergoline	949.13	339.92	Adham N, et al., 1992 [6]
8		6761	Methiothepin	179.53	648.83	Adham N, et al., 1992 [6]
9		6763	TFMPP	6301.28	1001.00	Adham N, et al., 1992 [6]
10		6764	5-Methoxytryptamine	3151.00	1166.00	Adham N, et al., 1992 [6]
11		6766	DOI	2970.00	1739.00	Adham N, et al., 1992 [6]

12		6767	8-OH-DPAT	826.00	1772.00	Adham N, et al., 1992 [6]
13		6768	Tryptamine	2559.00	2409.00	Adham N, et al., 1992 [6]
14		11351	Risperidone	1320.00	1240.00	Schotte A, et al., 1996 [12]
15		11354	Clozapine	698.00	130.00	Schotte A, et al., 1996 [12]
16		11355	Olanzapine	2209.00	310.00	Schotte A, et al., 1996 [12]
17		11356	Quetiapine	1826.00	2240.00	Schotte A, et al., 1996 [12]

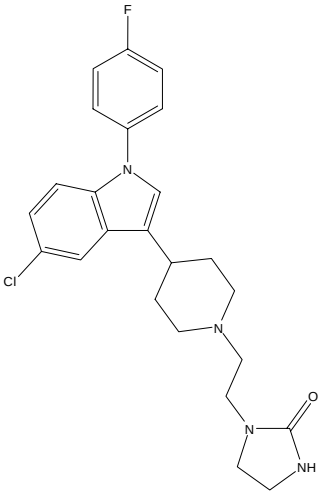
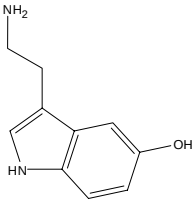
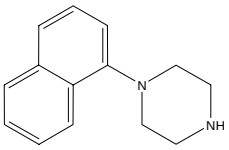
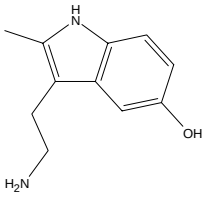
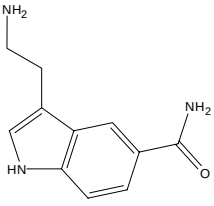
18		11357	Sertindole	430.00	360.00	Schotte A, et al., 1996 [12]
19		15510	5-Hydroxy Tryptamine	7.53	10.00	Boess FG, et al., 1994 [7]
20		15515	1-NP	207.96	53.85	Boess FG, et al., 1994 [7]
21		15521	5-HT,2-Me	814.91	414.93	Boess FG, et al., 1994 [7]
22		15523	5-CT	6552.26	720.71	Boess FG, et al., 1994 [7]

Table 4. The consensus prediction of 27 5HT1E agonist (AID726).

CID	Pred. Ki	St. Dev.	Model Coverage	Coverage Ratio	Exp. EC50
6603611	5.86	0.44	117	100.00	7.95
3245407	6.17	0.49	117	100.00	6.71
601380	5.66	0.37	117	100.00	6.17
197774	5.61	0.46	117	100.00	6.32
68551	6.85	0.40	117	100.00	7.42
6088	5.84	0.45	117	100.00	6.98
1832	5.89	0.34	117	100.00	7.04
2761023	6.15	0.44	115	98.29	7.80
152215	6.21	0.43	115	98.29	8.17
51703	6.26	0.40	115	98.29	7.62
2998624	6.24	0.45	112	95.73	6.73
2998813	6.16	0.47	109	93.16	5.88
916985	6.39	0.39	107	91.45	5.97
66259	6.25	0.29	106	90.60	7.01
5282386	5.92	0.37	101	86.32	7.06
6603257	6.00	0.29	97	82.91	6.31
660499	5.82	0.40	91	77.78	6.33
6602844	5.34	0.20	88	75.21	6.09
5281073	7.01	0.14	84	71.79	7.43
6364821	5.74	0.40	77	65.81	5.93
3239520	5.76	0.35	67	57.26	5.46
659395	5.93	0.42	66	56.41	5.76
2965107	5.67	0.47	48	41.03	6.12
733831	5.76	0.39	42	35.90	5.12
3235665	5.64	0.27	41	35.04	5.64
3240509	5.35	0.22	32	27.35	5.74
2523521	5.46	0.26	32	27.35	5.74

Table 5. The consensus prediction of 24 5HT1E antagonist (AID749).

CID	Pred. Ki	St. Dev.	Model Coverage	Coverage Ratio	Exp. EC50
91430	6.45	0.66	116	99.15	5.04
659735	6.09	0.32	112	95.73	5.08
659940	5.98	0.31	109	93.16	5.74
1881562	6.02	0.32	101	86.32	5.16
130606	5.62	0.21	91	77.78	5.43
2082098	5.58	0.22	85	72.65	5.54
2112599	6.24	0.41	82	70.09	5.46
650038	6.29	0.55	79	67.52	5.41
648878	5.98	0.31	74	63.25	6.00
1263400	5.58	0.38	68	58.12	5.58
868608	5.73	0.25	67	57.26	5.63
890649	5.93	0.43	57	48.72	5.98
653297	6.04	0.61	57	48.72	6.20
5740802	5.99	0.30	53	45.30	5.28
666746	6.28	0.35	40	34.19	5.44
3242053	5.53	0.31	33	28.21	5.55
3236803	5.34	0.21	27	23.08	5.67
2899176	6.42	0.65	22	18.80	5.31
3245258	5.34	0.21	21	17.95	5.71
597363	5.70	0.37	21	17.95	5.68
3237865	5.33	0.19	15	12.82	5.65
3244140	5.36	0.19	13	11.11	5.54
660939	5.53	0.37	13	11.11	5.73
2860584	5.78	0.45	7	5.98	5.58

Table 6. The Performance of *k*NN QSAR Classification Models Using Four Different Thresholds of Selectivity Index.

Selectivity Index ε_j	**Activities**	**Size**	**Acceptable model #**	**Best model** CCR_{train}	**Best model** CCR_{test}
0.14 - 7	1	17	0	0.80	0.50
< 0.14, > 7	0	5			
0.2 - 5	1	15	134	0.84	0.93
< 0.2, > 5	0	7			
0.33 - 3	1	13	281	0.88	1.00
< 0.33, > 3	0	9			
1	1	2	0	0.50	0.50
< 1, > 1	0	20			

Table 7a. 2 x 2 Confusion Matrix of the Training Set in the Best *k*NN QSAR Classification Model Using MolConnZ Descriptors.

	Observed		
Predicted	**1**	**0**	**Total**
1	7	3	10
0	0	5	5
Total	7	8	15

Table 7b. 2 x 2 Confusion Matrix of the Validation Test Set in the Best *k*NN QSAR Classification Model Using MolConnZ Descriptors.

	Observed		
Predicted	**1**	**0**	**Total**
1	6	0	6
0	0	1	1
Total	6	1	7

Table 8. Ten Best *k*NN QSAR Classification Models of 5HT1E/5HT1F Selectivity with the Highest CCR Values for All Test Sets Using MOE Descriptors.

Model No.	Nearest Neighbors No.	CCR_{train}	Confusion Matrix						Statistics for the Models				
			N(1)	N(0)	TP	TN	FP	FN	SE	SP	EN(1)	EN(0)	CCR_{test}
1	3	0.86	4	2	4	2	0	0	1.00	1.00	2.00	2.00	1.00
2	4	0.85	6	3	5	3	0	1	0.83	1.00	2.00	1.71	0.92
3	3	0.85	5	3	4	3	0	1	0.80	1.00	2.00	1.67	0.90
4	3	0.85	5	3	4	3	0	1	0.80	1.00	2.00	1.67	0.90
5	1	0.83	5	3	4	3	0	1	0.80	1.00	2.00	1.67	0.90
6	3	0.86	4	3	3	3	0	1	0.75	1.00	2.00	1.60	0.88
7	3	0.85	3	4	3	3	1	0	1.00	0.75	1.60	2.00	0.88
8	1	0.94	7	1	5	1	0	2	0.71	1.00	2.00	1.56	0.86
9	1	0.92	7	3	5	3	0	2	0.71	1.00	2.00	1.56	0.86
10	1	0.94	6	1	4	1	0	2	0.67	1.00	2.00	1.50	0.83

N(1) = number of non-selective ligands, N(0) = number of selective ligands, TP = true positive (non-selective ligands predicted as non-selective ligands), FP = false positives (selective ligands predicted as non-selective ligands), FN = false negatives (non-selective ligands predicted as selective ligands), TN = true negative (selective ligands predicted as selective ligands), SE = sensitivity = TP/N(1), SP = specificity = TN/N(0), EN - the normalized enrichment, EN(1) = (2TP * N(0))/(TP * N(0) + FP * N(1)), EN(0) = (2TN * N(1))/(TN * N(1) + FN * N(0)), and CCR = correct classification rate.

Table 9. Ten Best *k*NN QSAR Classification Models of 5HT1E/5HT1F Selectivity with the Highest CCR Values for All Test Sets Using Frequent Subgraph Descriptors.

Model No.	Nearest Neighbors No.	CCR_{train}	Confusion Matrix						Statistics for the Models				
			N(1)	N(0)	TP	TN	FP	FN	SE	SP	EN(1)	EN(0)	CCR_{test}
1	1	0.73	5	2	4	2	0	1	0.80	1.00	2.00	1.67	0.90
2	1	0.73	5	1	4	1	0	1	0.80	1.00	2.00	1.67	0.90
3	1	0.76	4	1	3	1	0	1	0.75	1.00	2.00	1.60	0.88
4	1	0.72	4	2	3	2	0	1	0.75	1.00	2.00	1.60	0.88
5	2	0.71	4	1	3	1	0	1	0.75	1.00	2.00	1.60	0.88
6	1	0.72	3	2	2	2	0	1	0.67	1.00	2.00	1.50	0.83
7	1	0.73	5	2	3	2	0	2	0.60	1.00	2.00	1.43	0.80
8	2	0.84	4	4	3	3	1	1	0.75	0.75	1.50	1.50	0.75
9	1	0.82	4	1	2	1	0	2	0.50	1.00	2.00	1.33	0.75
10	2	0.81	4	2	2	2	0	2	0.50	1.00	2.00	1.33	0.75

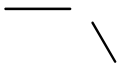
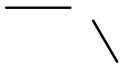
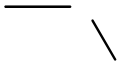
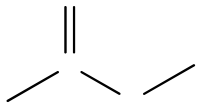
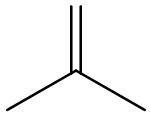
N(1) = number of non-selective ligands, N(0) = number of selective ligands, TP = true positive (non-selective ligands predicted as non-selective ligands), FP = false positives (selective ligands predicted as non-selective ligands), FN = false negatives (non-selective ligands predicted as selective ligands), TN = true negative (selective ligands predicted as selective ligands), SE = sensitivity = TP/N(1), SP = specificity = TN/N(0), EN - the normalized enrichment, EN(1) = (2TP * N(0))/(TP * N(0) + FP * N(1)), EN(0) = (2TN * N(1))/(TN * N(1) + FN * N(0)), and CCR = correct classification rate.

Table 10. Ten Best *k*NN QSAR Classification Models of 5HT1E/5HT1F Selectivity with the Highest CCR Values for All Test Sets Using Molecular Hologram Fingerprints.

Model No.	Nearest Neighbors No.	CCR_{train}	Confusion Matrix						Statistics for the Models				
			N(1)	N(0)	TP	TN	FP	FN	SE	SP	EN(1)	EN(0)	CCR_{test}
1	1	0.94	4	1	3	1	0	1	0.75	1.00	2.00	1.60	0.88
2	1	0.93	4	2	3	2	0	1	0.75	1.00	2.00	1.60	0.88
3	1	0.88	4	1	3	1	0	1	0.75	1.00	2.00	1.60	0.88
4	1	0.87	4	2	3	2	0	1	0.75	1.00	2.00	1.60	0.88
5	1	0.86	4	2	3	2	0	1	0.75	1.00	2.00	1.60	0.88
6	2	0.83	4	1	3	1	0	1	0.75	1.00	2.00	1.60	0.88
7	4	0.82	4	1	3	1	0	1	0.75	1.00	2.00	1.60	0.88
8	2	0.80	3	4	3	3	1	0	1.00	0.75	1.60	2.00	0.88
9	1	0.86	6	2	4	2	0	2	0.67	1.00	2.00	1.50	0.83
10	2	0.83	3	5	3	3	2	0	1.00	0.60	1.43	2.00	0.80

N(1) = number of non-selective ligands, N(0) = number of selective ligands, TP = true positive (non-selective ligands predicted as non-selective ligands), FP = false positives (selective ligands predicted as non-selective ligands), FN = false negatives (non-selective ligands predicted as selective ligands), TN = true negative (selective ligands predicted as selective ligands), SE = sensitivity = TP/N(1), SP = specificity = TN/N(0), EN - the normalized enrichment, EN(1) = (2TP * N(0))/(TP * N(0) + FP * N(1)), EN(0) = (2TN * N(1))/(TN * N(1) + FN * N(0)), and CCR = correct classification rate.

Table 11. 25 Most Important MolconnZ 4.05 Descriptors in *k*NN Models of 5HT1E/5HT1F Subtype Selectivity Datasets.

Rank[a]	Des. Index	Freq.[b]	Description	Class	Illustration
1	**diam**	22	graph diameter	shape indices	
2	**rad**	21	graph radius	shape indices	
3	**mulrad**	21	multiplicity of radius	shape indices	
4	**Xvp6**	20	chi valence path 6	chi indices	
5	**nd4**	20	vertex degree counts	vertex counts	
6	**etyp12**	18	count of occurrences where an atom with one connection is bonded to an atom with two connections	edge counts	
7	**SsF**	18	sum of atom level E-State values for all fluorine atoms in the molecule	atom-type Estate	
8	**nXc3**	17	counts of simple clusters	cluster counts	
9	**dXvp3**	17	chi valence difference path 3	chi indices	
10	**n2Pag23**	17	vertex alpha-gamma counts	vertex counts	
11	**K2**	17	kappa simple indices	kappa indices	
12	**dXvp4**	16	chi valence difference path 4	chi indices	
13	**n3Pad33**	16	vertex alpha-delta counts	vertex counts	
14	**n4Pae11**	16	count of occurrences where an atom with one connection is separated by four bonds from an atom with one connections	vertex counts	
15	**SHsOH**	16	sum of atom-type hydrogen E-State values of all hydrogen atoms of the hydroxyl group in the molecule	atom-type Estate	
16	**SsOH**	16	sum of atom level E-State values for all hydroxyl oxygen atoms in the molecule	atom-type EState	
17	**WT**	16	total Wiener number	shape indices	
18	**dXvp6**	15	chi valence difference path 6	chi indices	
19	**nHsOH**	15	counts of all hydrogen atoms of single bond to the hydroxyl group in the molecule	atom-type counts	
20	**nssCH2**	15	counts of all CH2 groups of single bond to others in the molecule	atom-type counts	
21	**dXvp7**	15	chi valence difference path 7	chi indices	
22	**Xvp9**	15	chi valence path 9	chi indices	
23	**k3**	15	kappa simple indices	kappa indices	
24	**Sester**	15	sum of E-State values of ester atoms in the molecule	group-type Estate	
25	**SdO**	14	sum of atom level E-State values for all sp2 oxygen atoms in the molecule	atom-type Estate	

[a] *k*NN rank is based on the frequency of each descriptor occurred.
[b] Frequency is the number of times each descriptor occurred in 60 Models with $CCR_{train} \geq 0.75$ and $CCR_{test} \geq 0.75$.

Table 12. 25 Most Important MOE Descriptors in *k*NN Models of 5HT1E/5HT1F Subtype Selectivity Datasets.

Rank[a]	Des. Index	Freq.[b]	Description
1	**GCUT_SLOGP_0**	78	atomic contribution to logP using the Wildman and Crippen SlogP method
2	**PEOE_VSA-2**	66	sum of v_i where q_i is in the range [-0.15,-0.10)
3	**std_dim1**	66	standard dimension 1
4	**chi1v**	65	atomic valence connectivity index.
5	**vsa_pol**	64	approximation to the sum of VDW surface areas of polar atoms
6	**dipoleX**	64	*x* component of the dipole moment
7	**PEOE_VSA+6**	63	sum of v_i where q_i is greater than 0.3
8	**vsa_don**	63	approximation to the sum of VDW surface areas of hydrogen bond donors
9	**vol**	63	van der Waals volume calculated using a grid approximation
10	**VAdjEq**	62	vertex adjacency information (equality)
11	**logS**	61	log of the aqueous solubility
12	**SMR_VSA5**	60	sum of v_i such that R_i is in (0.44,0.485]
13	**ASA_H**	60	water accessible surface area of all hydrophobic ($\|q_i\|$<0.2) atoms
14	**FCASA+**	60	fractional CASA+ calculated as CASA+ / ASA
15	**dens**	60	mass density
16	**b_1rotN**	60	number of rotatable single bonds
17	**rgyr**	60	radius of gyration
18	**reactive**	59	indicator of the presence of reactive groups
19	**E_str**	59	bond stretch potential energy
20	**E_tor**	59	torsion (proper and improper) potential energy
21	**a_heavy**	59	number of heavy atoms $\#\{Z_i \mid Z_i > 1\}$
22	**b_1rotR**	59	fraction of rotatable single bonds
23	**PEOE_PC+**	59	total positive partial charge
24	**petitjean**	59	value of (diameter - radius) / diameter
25	**PEOE_VSA_PPOS**	58	total positive polar van der Waals surface area

[a] *k*NN rank is based on the frequency of each descriptor occurred.
[b] Frequency is the number of times each descriptor occurred in 181 Models with $CCR_{train} \geq 0.75$ and $CCR_{test} \geq 0.75$.