

Supporting Information

Integrating *In Silico* and *In Vitro* Approaches to Predict Drug Accessibility to the Central Nervous System

Yan-Yan Zhang,^{1,*} Houfu Liu,¹ Scott G. Summerfield,² Christopher N. Luscombe,³ and Jasminder Sahi^{1,†}

¹ Drug Metabolism and Pharmacokinetics, Platform Technology and Science China, GlaxoSmithKline R&D, Shanghai, China

² GlaxoSmithKline R&D, David Jack Centre for R&D, Park Road, Ware, Hertfordshire, SG12 0DP, UK

³ Computational and Structural Chemistry, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK

***Corresponding Author**

E-mail: yanyan.6.zhang@gsk.com; Phone: +86-21-61590546

Drug Metabolism and Pharmacokinetics, Platform Technology and Science China,
GlaxoSmithKline R&D, Shanghai, China.

†Present Addresses

Sanofi R&D, 19F Tower III, Kerry Center, 1228 Middle Yan'an Road, Jing An District,
Shanghai 200040, China.

Table S1. Information on 54 CNS drugs, including generic names, calculated descriptors, and rat brain penetration values (total brain-to-blood ratio, K_p ; unbound brain-to-blood ratio, $K_{p,uu}$).

Drug Name	MW	Rotatable Bonds	Rings	Aromatic Rings	Most Acidic	Most Basic	cLogP	HBD	N+O	TPSA	Ionized ^b	Brain Penetration	
					pKa	pKa						Total	Unbound
AMANTADINE ^a	151	0	3	0		10.71	2	1	1	26.02	Base	6.70	3.00
AMITRIPTYLINE ^a	277	3	3	2		9.76	4.85	0	1	3.24	Base	10.07	1.33
ARIPIPRAZOLE	448	7	4	2	13.51	7.46	5.31	1	5	44.81	Base	9.00	1.64
ATOMOXETINE ^a	255	6	2	2		9.8	3.94	1	2	21.26	Base	19.30	2.13
BUPROPION ^a	240	3	1	1		8.22	3.21	1	2	29.1	Base	16.17	7.07
CARBAMAZEPINE ^a	236	1	3	2			2.38	1	3	46.33	Neutral	1.39	0.76
CHLORPROMAZINE ^a	319	4	3	2		9.2	5.5	0	2	6.48	Base	27.30	4.27
CITALOPRAM ^a	324	5	3	2		9.78	3.13	0	3	36.26	Base	6.83	1.30
CLOMIPRAMINE	315	4	3	2		9.2	5.92	0	2	6.48	Base	22.09	2.77
CLOZAPINE ^a	327	1	4	2		7.35	3.56	1	4	30.87	Neutral	22.96	4.26
DIAZEPAM ^a	285	1	3	2		2.92	2.96	0	3	32.67	Neutral	2.86	0.96

DONEPEZIL ^a	380	6	4	2	8.62	4.6	0	4	38.77	Base	3.45	2.36	
DOXEPIN ^a	279	3	3	2	9.76	4.09	0	2	12.47	Base	20.65	3.11	
ETHOSUXIMIDE ^a	141	1	1	0	10.73	0.4	1	3	46.17	Neutral	1.06	1.50	
FLUOXETINE ^a	309	6	2	2	9.8	4.57	1	2	21.26	Base	36.58	5.23	
FLUPHENAZINE ^a	438	6	4	2	8.21	4.32	1	4	29.95	Base	24.60	1.58	
GALANTAMINE	287	1	4	1	8.91	1.02	1	4	41.93	Base	1.50	1.16	
HALOPERIDOL ^a	376	6	3	2	13.96	8.05	3.85	1	40.54	Base	16.73	2.09	
IMIPRAMINE	280	4	3	2	9.2	5.04	0	2	6.48	Base	22.98	3.42	
ISOCARBOXAZID ^a	231	4	2	2	12.02	3.12	0.99	2	5	67.16	Neutral	0.43	0.18
LAMOTRIGINE ^a	256	1	2	2	5.87	2.53	2	5	90.71	Neutral	5.06	4.06	
LOXAPINE ^a	328	1	4	2	7.18	3.98	0	4	28.07	Neutral	14.23	2.95	
MAPROTILINE ^a	277	4	4	2	10.54	4.52	1	1	12.03	Base	14.04	0.98	
MEMANTINE	179	0	3	0	10.7	3.03	1	1	26.02	Base	19.58	6.06	
MEPROBAMATE ^a	218	6	0	0	0.92	2	6	104.64	Neutral	0.87	0.77		
MESORIDAZINE ^a	387	4	4	2	8.19	4.64	0	3	23.55	Base	1.00	0.10	
METOCLOPRAMIDE ^a	300	7	1	1	9.04	2.23	2	5	67.59	Base	0.79	0.40	

MIANSERIN	264	0	4	2		6.92	3.76	0	2	6.48	Neutral	18.02	3.15
MIDAZOLAM ^a	326	1	4	3		7.42	3.42	0	3	30.18	Base	2.02	1.08
MIRTAZAPINE ^a	265	0	4	2		6.67	2.81	0	3	19.37	Neutral	6.80	3.83
MORPHINE ^a	285	0	5	1	10.26	9.12	0.57	2	4	52.93	Base	0.40	0.27
NORTRIPTYLINE	263	3	3	2		10.47	4.32	1	1	12.03	Base	12.48	1.65
OLANZAPINE ^a	312	1	4	2		7.24	2.86	1	4	30.87	Neutral	8.50	2.14
PEMOLINE ^a	176	1	2	1			0.46	1	4	64.68	Neutral	0.55	0.55
PERPHENAZINE ^a	404	6	4	2		8.21	4.01	1	4	29.95	Base	24.91	7.12
PHENELZINE ^a	136	3	1	1		5.55	1.03	2	2	38.05	Neutral	2.34	1.48
PHENYTOIN ^a	252	2	3	2	6.46		2.08	2	4	58.2	Acid	1.23	0.66
PRIMIDONE ^a	218	2	2	1	11.5		0.88	2	4	58.2	Neutral	0.46	0.38
QUETIAPINE ^a	384	6	4	2		7.06	2.99	1	5	48.3	Neutral	4.03	1.38
RISPERIDONE ^a	410	4	5	3		8.76	2.71	0	6	64.16	Base	0.30	0.21
RITANSERIN	478	5	5	4		8	5.28	0	4	37.61	Base	1.99	0.33
RIZATRIPTAN ^a	269	5	3	3		9.56	1.18	1	5	49.74	Base	0.09	0.04
SELEGILINE ^a	187	4	1	1		8.67	2.84	0	1	3.24	Base	2.65	1.08

SERTRALINE ^a	306	2	3	2		9.85	5.35	1	1	12.03	Base	38.59	2.67
SUMATRIPTAN ^a	295	6	2	2	11.24	9.54	0.74	2	5	65.2	Base	0.03	0.04
TACRINE ^a	198	0	3	2		8.95	3.27	1	2	38.91	Base	5.94	1.61
TEMAZEPAM ^a	301	1	3	2	10.68		2.34	1	4	52.9	Neutral	2.27	0.80
THIORIDAZINE ^a	371	4	4	2		8.93	6.2	0	2	6.48	Base	11.30	2.83
THIOTHIXENE ^a	444	5	4	2		8.56	3.23	0	5	43.86	Base	1.51	0.13
TIAGABINE ^a	376	6	3	2	4.14	9.26	2.77	0	3	40.54	Zwitterion	1.44	0.62
TRAZODONE ^a	372	5	4	3		7.09	3.85	0	6	45.78	Neutral	3.35	1.49
TRIFLUOPERAZINE ^a	408	4	4	2		8.39	4.89	0	3	9.72	Base	19.10	3.34
VENLAFAKINE ^a	277	5	2	1		8.91	3.27	1	3	32.7	Base	5.20	1.91
ZALEPLON ^a	305	4	3	3		0.3	1.44	0	6	74.29	Neutral	0.83	0.59

^a The rat K_p values were determined by steady-state design.

^b The ionized state was determined as neutral when the compound has the most acid pKa>7.4 or the most basic pKa<7.4.

Table S2. The importance of physicochemical descriptors to the rat K_{p,uu} classification QSAR model.

Physicochemical Descriptors		Number of Questions ^b	Raw Selection	Number of Chances ^d	Percent Selection	Mean Score Improvement ^f	Mean Score When Used ^g	Mean Depth When Used ^h
Name ^a	Details of GSK Definition		Frequency ^c		Frequency ^e		When Used ^g	When Used ^h
Count of negatively ionisable								
gskneg	groups (e.g. carboxylic acid, tetrazole)	1	22	93	23.66	0.02572	0.03843	2.636
Num_SpiroAtoms		1	16	147	10.88	0.008041	0.03443	4.688
IsChiral		1	29	279	10.39	0.01697	0.07664	5.207
Num_Rings		4	47	498	9.438	0.01134	0.04142	4.234
Count of positively ionisable								
gskpos	groups (e.g. basic amines, amidines)	2	24	272	8.824	0.01463	0.05732	5.708
Num_StereoAtoms		4	51	640	7.969	0.01425	0.05129	4.882
Num_StereoBonds		2	26	336	7.738	0.01699	0.05623	4.846

	Count of HBD without							
gskhbd	including any strong acid substructures	3	37	524	7.061	0.01237	0.05929	5
Num_H_Donors		4	44	645	6.822	0.01206	0.04578	5.114
Num_H_Donors_Lipinski		5	50	764	6.545	0.01157	0.0457	5.18
HBD_Count		3	32	498	6.426	0.01133	0.03819	5.312
Num_H_Acceptors		8	61	1048	5.821	0.01105	0.04744	5.557
Num_BridgeHeadAtoms		1	5	96	5.208	0.009887	0.0449	4.8
gskacidclass	Categories (0-7) according to acidic pKa values	7	48	965	4.974	0.01467	0.06251	4.646
HBA_Count		6	49	1000	4.9	0.01115	0.06838	5.939
Num_AromaticRings		4	26	585	4.444	0.01209	0.06432	5.269
Num_H_Acceptors_Lipinski		9	47	1072	4.384	0.01114	0.04115	4.66
gskhba	Count of HBA excluding basic amines and anilines etc..	6	37	862	4.292	0.01036	0.05674	4.703
gskbaseclass	Categories (0-7) according to	7	74	1752	4.224	0.01343	0.0612	4.986

basic pKa values

Num_Rotatable_Bonds	10	52	1348	3.858	0.01022	0.05086	5.423
Num_Aromatic_Bonds	13	62	1880	3.298	0.01125	0.05741	5.161
Num_Bonds	32	88	3311	2.658	0.01101	0.05064	4.83
Num_Atoms	29	62	2596	2.388	0.01247	0.05398	5.323
Num_Bridge_Bonds	3	6	254	2.362	0.008872	0.02861	4.333
Molecular_Surface_Area	49	125	7635	1.637	0.01539	0.06141	5.376
Molecular_Mass	49	118	7246	1.628	0.01228	0.05774	5.11
Molecular_Fractional_PolarSurfaceArea	49	116	7218	1.607	0.01084	0.05381	5.474
Molecular_SAVol	49	105	6829	1.538	0.01064	0.05783	5.562
Molecular_Fractional_PolarSASA	50	104	6793	1.531	0.01224	0.07016	5.337
Molecular_SASA	49	112	7399	1.514	0.01117	0.05472	5.464
Molecular_Polar_SurfaceArea	49	104	7761	1.34	0.01228	0.0609	4.942
Molecular_Weight	49	109	8233	1.324	0.0112	0.05148	5.092
LogD	49	134	10186	1.316	0.0121	0.05633	5.515
Molecular_Solubility	49	117	8979	1.303	0.01293	0.06375	5.709

Molecular_PolarSASA	50	116	9298	1.248	0.009913	0.0402	5.043
ALogP	49	119	9912	1.201	0.01174	0.06091	5.731

^a The descriptor names if without GSK definitions refer to the properties in Pipeline Pilot (BIOVIA, Inc. San Diego, CA).

^b The number of distinct questions for data splitting associated with the descriptor.

^c The number of times the descriptor was selected for a split.

^d The number of times the descriptor was considered for a split and a split was possible (meaning that the number of samples in each child node would have been at least Minimum Samples Per Node as specified in the learner).

^e The percent of the time the Descriptor was selected for a split when a split was possible.

^f The mean over all possible splits of the improvement in node purity according to the scoring function specified by Split Method (Gini) in the learner.

^g The mean over only the splits for which the descriptor was used as the splitting criterion.

^h The mean depth of the tree at which the descriptor was used as the splitting criterion. (The top node of each tree has a depth of 0.)

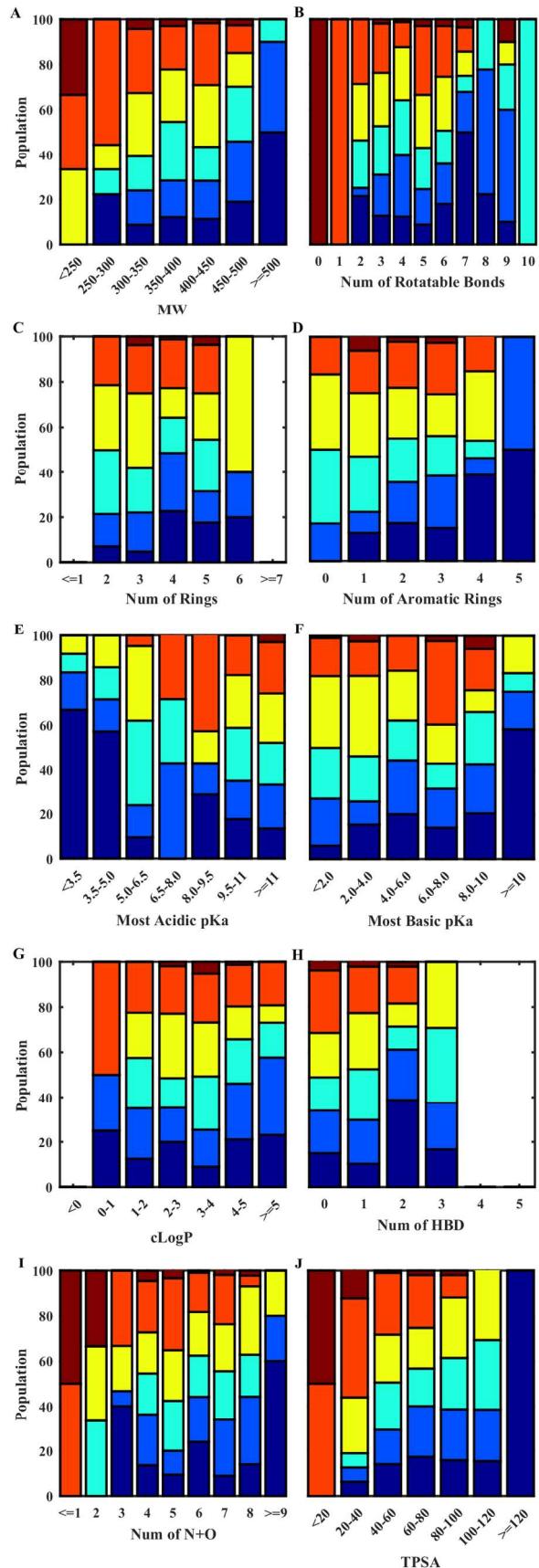


Figure S1. Distribution of physicochemical descriptors for the dataset of 385 compounds with rat $K_{p,uu}$ data determined by brain/blood (plasma) AUC ratio method. (A) MW, (B) number of rotatable bonds, (C) number of rings, (D) number of aromatic rings, (E) most acidic pKa, (F) most basic pKa. (G) cLogP, (H) number of HBD, (I) Lipinski hydrogen bonding acceptors as N+O atoms, and (J) TPSA. The coloring represents the range of $K_{p,uu}$ values. Dark blue, $K_{p,uu} < 0.1$; Blue, $0.1 \leq K_{p,uu} < 0.3$; Cyan, $0.3 \leq K_{p,uu} < 0.5$; Yellow, $0.5 \leq K_{p,uu} < 1.0$; Red, $1.0 \leq K_{p,uu} < 3.0$; Dark red, $K_{p,uu} \geq 3.0$. Since only a small fraction of compounds (<5%) in this rat $K_{p,uu}$ subset had MW<300, number of rotatable bonds≤1, number of aromatic rings≥4, most acidic pKa at 6.5–9.5, most basic pKa≥10, cLogP<1, N+O≥9, and TPSA≥120, the frequency distribution of compounds located in these physicochemical regions should be interpreted with caution.

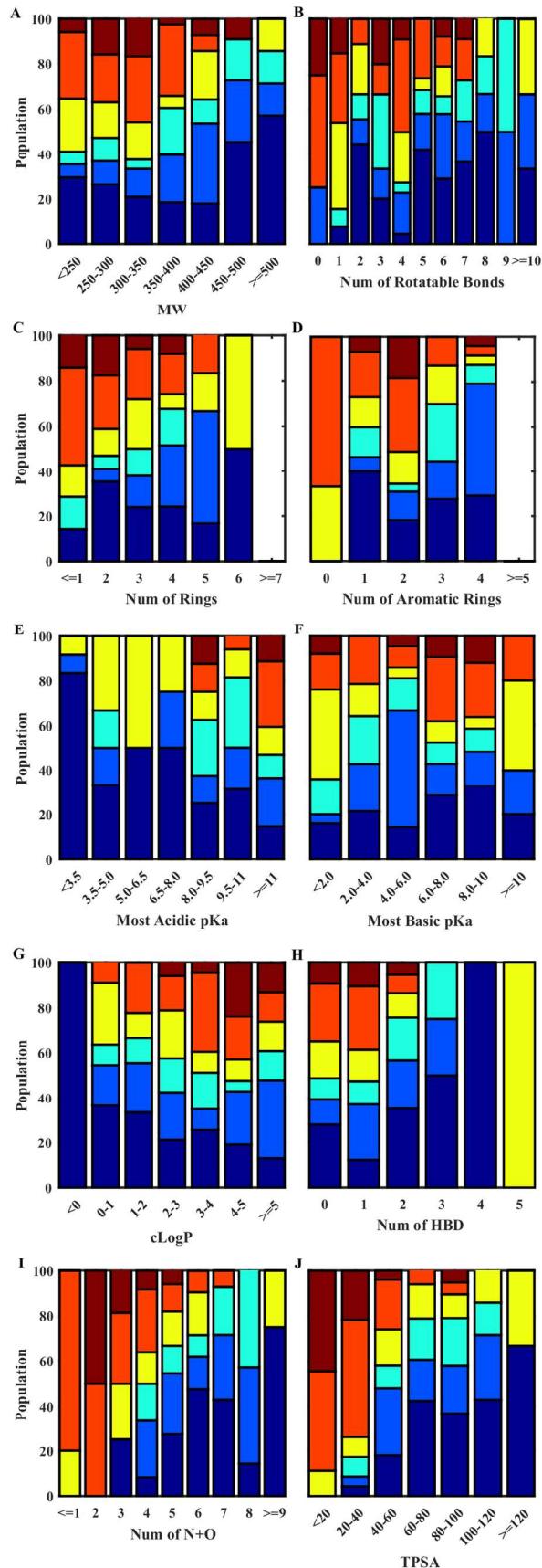


Figure S2. Distribution of physicochemical descriptors for the dataset of 144 compounds with rat $K_{p,uu}$ data determined by steady-state design. (A) MW, (B) number of rotatable bonds, (C) number of rings, (D) number of aromatic rings, (E) most acidic pKa, (F) most basic pKa. (G) cLogP, (H) number of HBD, (I) Lipinski hydrogen bonding acceptors as N+O atoms, and (J) TPSA. The coloring represents the range of $K_{p,uu}$ values. Dark blue, $K_{p,uu} < 0.1$; Blue, $0.1 \leq K_{p,uu} < 0.3$; Cyan, $0.3 \leq K_{p,uu} < 0.5$; Yellow, $0.5 \leq K_{p,uu} < 1.0$; Red, $1.0 \leq K_{p,uu} < 3.0$; Dark red, $K_{p,uu} \geq 3.0$. Since only a small fraction of compounds (<5%) in this rat $K_{p,uu}$ subset had number of rotatable bonds ≥ 9 , number of rings ≥ 6 , most acidic pKa at 5.0–8.0, most basic pKa ≥ 10 , cLogP < 0 , number of HBD ≥ 3 , N+O ≥ 9 , and TPSA ≥ 120 , the frequency distribution of compounds located in these physicochemical regions should be interpreted with caution.