Supporting Information for

Predicting Skin Permeation Rate from Nuclear Magnetic Resonance Spectra

Nan An, John-Hanson Machado, Yuechuan Tang, Jakub Kostal and Adelina Voutchkova-Kostal

Department of Chemistry, The George Washington University, Washington, DC 20052

Corresponding Author:

Adelina Voutchkova-Kostal. Tel.: 202-994-6121, Fax: 202-994-5873,

Email: avoutchkova@email.gwu.edu

Table of Contents

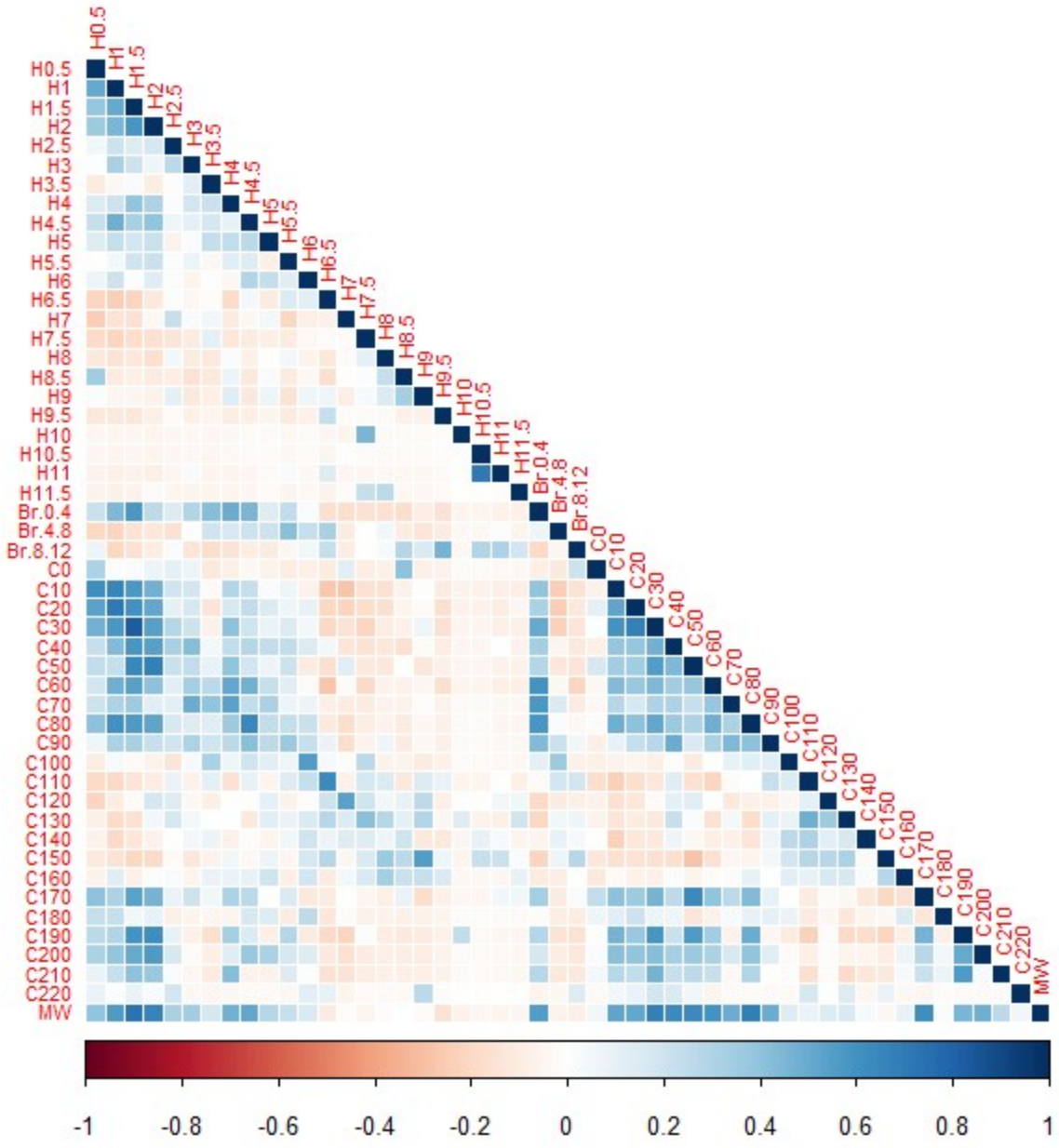

Figure S1. Correlation of the initial descriptors. Proton resonance bins are 0.5 ppm in width, for example, H1.5 stands for the integration of protons between 1.5-2 ppm. Carbon resonance bins are 10 ppm in width, for example, C10 stands for number of peak between 10-20 ppm. Br.0.4, Br.4.8 and Br.8.12 stands for integration of broad peaks within 0-4 ppm, 4-8 ppm and 8-12+ ppm respectively.

Table S1. Dataset with designation of training/validation set, experimental log K_p and that predicted by the PLS model and QikProp.

No.	Validation Set	Name	CAS	Exp log P[1]	Exp log K_p[1] (cm/s)	QikProp log K_p[2] (cm/s)	QSDAR log K_p (cm/s)
1		1,1,1-Trichloroethane	71-55-6	2.49	-5.9	-5.08	-5.73
2		1,2,3-Trihydroxybenzene	87-66-1	0.97	-6.37	-6.97	-5.75
3	*	1,2,4-Trihydroxybenzene	533-73-3	0.55	-7.46	-7.14	-7.57
4		1,3,5-Trihydroxybenzene	108-73-6	0.16	-6.07	-7.31	-5.76
5		12-Deoxyphorbol 13-isobutyrate	25090-74-8	2.30	-5.99	-7.09	-5.72
6		2,3-Butanediol	513-85-9	-0.92	-7.62	-6.33	-8.79
7		2,4,6-Trichlorophenol	88-06-2	3.69	-4.7	-5.40	-4.75
8	*	2,4-Dichlorophenol	120-83-2	3.06	-4.7	-5.39	-4.92
9	*	2-Amino-4-nitrophenol	99-57-0	1.26	-6.67	-8.10	-6.17
10		2-Butanone	78-93-3	0.29	-5.78	-5.80	-5.65
11	*	2-Chlorophenol	95-57-8	2.15	-4.96	-5.22	-5.19
12		2-Ethoxyethanol	110-80-5	-0.32	-7.06	-5.70	-7.67
13		2-Heptanol	543-49-7	2.31	-5.04	-5.37	-4.52
14	*	2-Hydroxypropyl nicotinate	61379-44-0	1.16	-7.55	-6.30	-8.24
15		2-Nitrophenol	88-75-5	1.79	-4.08	-6.71	-4.15
16		1,4-Diamino-2-nitrobenzene	5307-14-2	0.53	-6.81	-8.07	-7.45
17	*	2-Phenylethanol	60-12-8	1.36	-5.23	-5.15	-5.36
18		3,4-xylenol	95-65-8	2.30	-4.92	-5.16	-4.96
19	*	3-Phenylpropanol	122-97-4	1.88	-4.84	-5.06	-4.77
20		4-Amino-2-nitrophenol	119-34-6	0.96	-6.04	-7.91	-5.90
21	*	4-Bromophenol	106-41-2	2.59	-4.92	-5.37	-4.89
22	*	2-Chloro-5-hydroxytoluene	59-50-7	3.10	-4.71	-5.52	-4.06
23		1-Chloro-2,4-diaminobenzene	5131-60-2	0.85	-6.31	-6.55	-6.62
24		4-Ethylphenol	123-07-9	2.58	-4.93	-5.35	-4.74
25		4-Hydroxybenzyl alcohol	623-05-2	0.25	-6.26	-6.23	-6.98
26		4-Hydroxy-methylphenylacetate	155380-11-3	1.60	-5.26	-6.12	-5.06
27		4-Hydroxyphenylacetamide	17194-82-0	0.07	-6.89	-7.22	-6.84
28		4-Hydroxyphenylacetic acid	156-38-7	0.75	-6.06	-7.12	-5.24
29		4-Phenylbutanol	3360-41-6	2.35	-4.74	-4.96	-4.23
30	*	5-Fluorouracil	51-21-8	-0.81	-6.82	-7.74	-7.17
31		5-Phenylpentanol	10521-91-2	3.04	-4.53	-4.86	-4.30
32		6-Phenylhexanol	2430-16-2	3.53	-4.56	-4.77	-4.44
33		8-Methoxypsoralen	298-81-7	2.00	-5.12	-5.38	-4.64
34		Ethylic acid	64-19-7	-0.17	-6.53	-7.28	-6.70
35		Aldosterone	52-39-1	1.08	-8.17	-7.70	-7.79
36		Amylobarbital	57-43-2	2.07	-6.14	-7.01	-5.83
37		Aniline	62-53-3	0.90	-5.05	-5.35	-4.72
38		Anisole	100-66-3	2.11	-4.8	-4.30	-4.17
39		Aspirin	50-78-2	1.19	-5.69	-6.92	-6.14
40	*	Atropine	51-55-8	1.83	-7.8	-7.29	-7.98
41		Barbital	57-44-3	0.65	-7.47	-7.25	-7.24
42		Phenylmethanal	100-52-7	1.48	-4.47	-5.66	-3.83
43		Benzene	71-43-2	2.13	-4.41	-4.15	-4.97
44		Benzoic acid	65-85-0	1.87	-5.15	-6.22	-4.64
45		Benzyl alcohol	100-51-6	1.10	-5.41	-5.17	-6.02
46		Benzyl nicotinate	94-44-0	2.40	-4.87	-4.77	-5.12
47		2-Hydroxynaphthalene	135-19-3	2.70	-5.04	-4.96	-4.77
48		Bromodichloromethane	75-27-4	2.00	-3.82	-5.08	-2.74
49	*	Methenyl tribromide	75-25-2	2.40	-3.75	-5.08	-3.27
50	*	Butobarbital	125-40-6	1.73	-7.23	-7.03	-7.44
51		Butyl nicotinate	6/3/6938	2.27	-4.86	-5.50	-4.12
52		Butyric acid	107-92-6	0.79	-6.27	-6.85	-6.87
53		Caffeine	58-08-2	-0.07	-6.46	-6.65	-6.59
54		Catechol	154-23-4	0.88	-6.07	-8.16	-5.37
55		Chlorodibromomethane	124-48-1	2.16	-3.78	-5.08	-2.36
56		Methane trichloride	67-66-3	1.97	-3.92	-5.08	-2.72
57		Chloroxylenol	88-04-0	3.27	-4.72	-5.67	-4.35
58		Chlorpheniramine	132-22-9	3.38	-6.19	-6.19	-5.73
59		Codeine	76-57-3	1.19	-7.58	-7.55	-7.28
60		Cortexolone	152-58-9	3.08	-7.68	-7.42	-7.96
61		Cortexone	64-85-7	2.88	-6.9	-7.07	-6.98
62		Corticosterone	50-22-6	1.94	-7.26	-7.48	-7.43
63		Cortisone	53-06-5	1.47	-8.56	-7.92	-9.72
64		Dexamethasone	50-02-2	1.83	-7.53	-7.42	-7.72

65		Diclofenacb	15307-86-5	4.51	-5.73	-5.51	-5.91
66		Diethyl ether	60-29-7	0.89	-5.37	-5.57	-5.14
67		Diethylcarbamazine	90-89-1	0.37	-7.01	-7.60	-6.96
68		Digitoxin	71-63-6	1.85	-8.37	-8.83	-8.44
69		Ephedrine	299-42-3	1.13	-5.72	-6.94	-6.04
70		Estradiol	50-28-2	4.01	-6.14	-6.23	-5.14
71		1,3,5(10)-Estratriene-3,16a,17b-triol	50-27-1	2.45	-7.96	-6.81	-8.40
72		Estrone	53-16-7	3.13	-6	-6.41	-5.25
73	*	Ethanol	64-17-5	-0.31	-6.53	-6.03	-6.60
74		Ethyl benzene	100-41-4	3.15	-3.36	-4.29	-3.13
75	*	Diethyl ether	60-29-7	0.89	-6.02	-5.57	-6.37
76		Ethyl nicotinate	614-18-6	1.32	-5.28	-5.69	-5.04
77		Etorphineb	14521-96-1	2.79	-5.88	-7.62	-5.92
78	*	Fentanyl	437-38-7	4.05	-5.56	-5.62	-5.45
79		Fluocinonide	356-12-7	3.19	-6.33	-7.35	-5.95
80	*	Hydrogencarboxylic acid	64-18-6	-0.46	-6.6	-7.97	-6.47
81		1-Hexanecarboxylic acid	111-14-8	2.42	-5.03	-6.57	-5.16
82		Hexanoic acid	142-62-1	1.92	-5.18	-6.66	-5.56
83		Hexyl nicotinate	23597-82-2	3.51	-4.83	-5.31	-4.32
84		11,17,21-Trihydroxypregn-4-ene-3,20-dione	8063-42-1	3.26	-6.3	-7.79	-5.48
85		Hydrocortisone hemisuccinate	83784-20-7	1.89	-6.76	-9.38	-7.09
86		Hydrocortisone hexanoate	3593-96-2	4.48	-5.3	-7.35	-5.13
87		Hydrocortisone octanoate	6678-14-4	5.49	-4.77	-7.16	-5.02
88		Hydrocortisone propinatea	6677-98-1	2.80	-6.17	-7.48	-6.27
89		Pregn-4-ene-3,20-dione, 11,17,21-trihydroxy-, (11β)-	50-23-7	1.61	-7.85	-7.67	-8.51
90		Hydromorphone	466-99-9	1.60	-8.03	-8.64	-8.49
91		Hydroquinone	123-31-9	0.59	-6.51	-6.26	-6.59
92		Hydroxypregnenolone	12041-98-4	3.71	-6.78	-3.56	-6.80
93		Hydroxyprogesterone	68-96-2	3.17	-6.78	-6.90	-7.20
94		Ibuprofen	15687-27-1	3.50	-5.3	-6.02	-6.42
95		Indomethacinb	53-86-1	4.27	-5.85	-6.17	-5.91
96		Isoquinoline	119-65-3	2.08	-5.29	-4.50	-5.62
97	*	Lidocaine	137-58-6	2.44	-6.17	-6.71	-6.21
98		1-Hydroxy-3-methylbenzene	108-39-4	1.96	-5.29	-5.40	-5.12
99		Meperidine	57-42-1	2.72	-5.7	-6.90	-6.06
100		Methanol	67-56-1	-0.77	-6.7	-6.28	-7.02
101	*	Methyl 4-hydroxybenzoate	99-76-3	1.96	-5.44	-6.30	-4.10
102		Methyl nicotinate	93-60-7	0.83	-5.77	-5.85	-5.61
103		3-Hydroxy-1-nitrobenzene	554-84-7	2.00	-5.73	-7.10	-6.29
104	*	Morphine	57-27-2	0.89	-8.3	-8.71	-9.10
105		(+)-2-(6-Methoxy-2-naphthyl)-propionic acid	22204-53-1	3.18	-5.74	-5.83	-5.88
106		1-Butanol	71-36-3	0.88	-6.08	-5.84	-6.12
107	*	1-Decanol	112-30-1	4.57	-4.47	-5.26	-4.69
108	*	1-Heptanol	111-70-6	2.62	-4.91	-5.55	-5.32
109		1-Hexanol	111-27-3	2.03	-5.27	-5.65	-5.40
110		Nicotine	54-11-5	1.17	-5.47	-7.31	-5.60
111		Nitroglycerine	55-63-0	1.62	-5.44	-10.51	-5.27
112		1-Octanol	111-87-5	3.00	-4.69	-5.45	-4.78
113	*	1-Nonanol	143-08-8	3.77	-4.68	-5.36	-5.03
114		1-Pentanol	71-41-0	1.51	-5.7	-5.74	-5.64
115		1-Propanol	71-23-8	0.25	-6.32	-5.93	-6.62
116		1-Hydroxy-2-methylbenzene	95-48-7	1.95	-5.28	-5.18	-5.31
117		1-Heptanecarboxylic acid	124-07-2	3.05	-4.93	-6.47	-4.76
118		1,2-Benzenediamine	95-54-5	0.15	-6.85	-6.32	-7.37
119		Ouabain	630-60-4	-2.00	-9.66	-9.36	-9.40
120		4-Chlorophenol	106-48-9	2.39	-4.92	-5.37	-4.52
121	*	1-Hydroxy-4-methylbenzene	106-44-5	1.94	-5.23	-5.40	-5.03
122	*	1-Butanecarboxylic acid	109-52-4	1.39	-5.96	-6.76	-6.75
123		Phenobarbital	50-06-6	1.47	-6.66	-6.85	-7.45
124		Phenol	108-95-2	1.46	-5.52	-5.20	-5.51
125		Piroxicam	36322-90-4	3.06	-7.37	-6.44	-8.09
126		4-Hydroxynitrobenzene	100-02-7	1.91	-5.73	-7.10	-6.59
127		1,4-Benzenediamine	106-50-3	-0.30	-7.13	-6.55	-7.53
128	*	Pregnenolone	145-13-1	4.22	-6.38	-6.00	-5.79
129		Progesterone	57-83-0	3.87	-5.5	-6.38	-5.32
130	*	Propionic acid	79-09-4	0.33	-6.02	-6.95	-6.56
131		Resorcinol	108-46-3	0.80	-7.03	-6.26	-7.43
132		Salicylic acid	69-72-7	2.26	-5.53	-6.88	-5.72
133		Scopolamine	51-34-3	0.98	-7.87	-7.41	-8.34
134		Styrene	100-42-5	2.95	-3.75	-4.10	-3.51

135		Sucrose	57-50-1	-3.07	-8.84	-8.99	-8.98
136		Sufentanyl	56030-54-7	3.95	-5.4	-5.68	-5.59
137		Testosterone	58-22-0	3.32	-6.27	-6.49	-6.12
138		Tetrachloroethylene	127-18-4	3.40	-4.86	-5.02	-4.74
139		Thymol	89-83-8	3.30	-4.74	-5.34	-4.95
140	*	Toluene	108-88-3	2.73	-3.58	-4.34	-3.22
141	*	Trichloroethylene	79-01-6	2.42	-4.03	-4.98	-3.02
142		Urea	57-13-6	-2.11	-7.39	-8.76	-7.63
143	*	Water	7732-18-5	-1.38	-6.53	-7.40	-6.29

1,2 See reference 1 and 2.

*: indicates the compound is in first validation set. (EXT1)

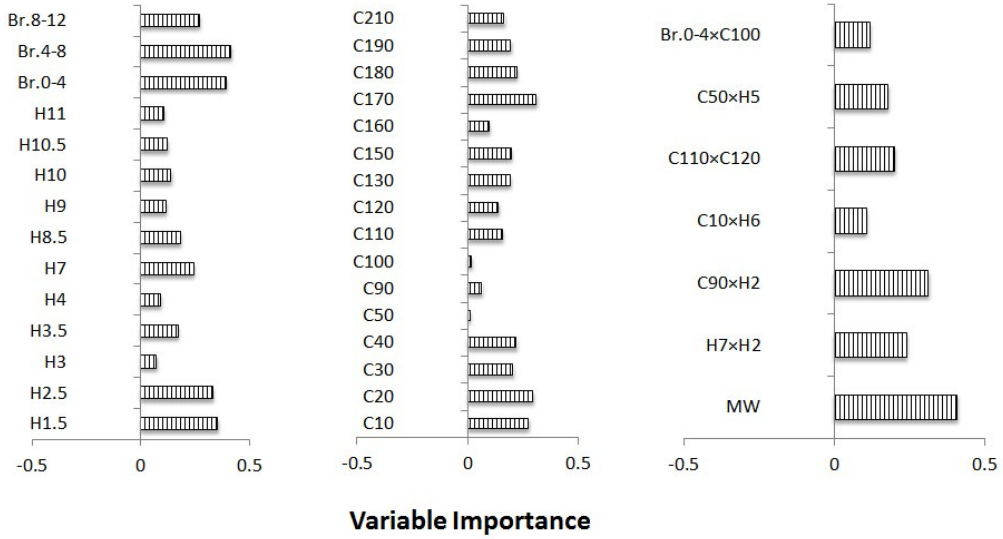

Figure S2. Absolute standardized coefficients for descriptors in the QSDAR model. Proton resonance bins are 0.5 ppm in width, for example, H1.5 stands for the integration of protons between 1.5-2 ppm. Carbon resonance bins are 10 ppm in width, so C10 for example represents number of resonances between 10-20 ppm. Br.0-4, Br.4-8 and Br.8-12 represents integration of broad peaks within 0-4 ppm, 4-8 ppm and 8-12 ppm ranges respectively.

Table S2. Full details of the data sources and assessment of 20 selected compounds in additional external validation set with high fidelity log K_p values

Name	CAS	Klimisch Binary Score[a,3]	Klimisch Cat[a,3]	Critical Criteria Failed (ToxRTool Reference Number)[4]	Ref	Donor Compartment pH	Receiver Compartment pH	Body Part	Exp log K_p (cm/s)	QikProp log K_p (cm/s)
bromoacetic acid	79-08-3	17	1	None	Xu et al 2002	7	7.4	Breast	-6.41	-7.02
dibromoacetic acid	631-64-1	17	1	None	Xu et al 2002	7	7.4	Breast	-6.14	-6.86
bromochloroacetic acid	5589-96-8	17	1	None	Xu et al 2002	7	7.4	Breast	-6.35	-6.88
trichloroacetic acid	76-03-9	17	1	None	Xu et al 2002	7	7.4	Breast	-6.28	-6.85
chloroacetic acid	79-11-8	17	1	None	Xu et al 2002	7	7.4	Breast	-6.51	-7.13
dichloroacetic acid	79-43-6	17	1	None	Xu et al 2002	7	7.4	Breast	-6.28	-6.99
bromochloroacetonitrile	83463-62-1	17	1	None	Xu et al 2002	7	7.4	Breast	-4.35	-6.26
1,1-dichloropropanone	513-88-2	17	1	None	Xu et al 2002	7	7.4	Breast	-4.92	-5.96
chloroacetonitrile	107-14-2	17	1	None	Xu et al 2002	7	7.4	Breast	-4.56	-6.34
dibromoacetonitrile	3252-43-5	17	1	None	Xu et al 2002	7	7.4	Breast	-4.33	-6.26
dichloroacetonitrile	3018-12-0	17	1	None	Xu et al 2002	7	7.4	Breast	-4.38	-6.36
chloral hydrate	302-17-0	17	1	None	Xu et al 2002	7	7.4	Breast	-5.97	-6.51
famotidine	76824-35-6	16	1	10	Degim et al 2003	Not Reported	7.4	Abdomen	-8.35	-10.45
nizatidine	76963-41-2	16	1	10	Degim et al 2003	Not Reported	7.4	Abdomen	-7.98	-8.37
ranitidine	71130-06-8	16	1	10	Degim et al 2003	Not Reported	7.4	Abdomen	-7.61	-8.01
nimesulide	51803-78-2	16	1	10	Degim et al 2003	Not Reported	7.4	Abdomen	-6.55	-7.23
etodolac	41340-25-4	16	1	10	Degim et al 2003	Not Reported	7.4	Abdomen	-5.68	-5.91
DL-histidine	71-00-1	18	1	None	Sznitowska et al 1993	7.3	7.3	Thigh	-7.81	-9.83
L-Lysine	56-87-1	18	1	None	Sznitowska et al 1993	7.3	7.3	Thigh	-7.33	-12.42
L-Aspartic acid	56-84-8	18	1	None	Sznitowska et al 1993	7.3	7.3	Thigh	-7.59	-11.30

[a]Ignoring de facto category 3 called for by the ToxRTool for one critical criteria

Table S3. Comparison of log K_p, predicted log K_p and NMR chemical shifts and pK_a associated with phenols in the training and validation set.

Comp No[a]	Chemical name	log K_p	C shift (ppm)	H shift (ppm)	pK_a	Predicted log K_p
T70	Estradiol	-6.14	154.9	6.00	10.33	-5.14(+)
T126	4-Hydroxynitrobenzene	-5.73	160.4	12.15	7.04	-6.59(-)
T28	4-Hydroxyphenylacetic acid	-6.06	155.5	6.00	4.00, 9.51	-5.24(+)
V101	Methyl 4-hydroxybenzoate	-5.44	163.7	9.95	8.50	-4.10(+)
V104	Morphine	-8.3	140.2	5.92	10.26	-9.10(-)
V22	2-Chloro-5-hydroxytoluene	-4.71	155.0	6.00	9.08	-4.06(+)
V9	2-Amino-4-nitrophenol	-6.67	148.3	8.20	7.82	-6.17(+)

[a] T: training set; V: validation set

Table S4. Comparison of log K_p, predicted log K_p and NMR chemical shifts and pK_a associated with carboxylic acids in the training and validation set.

Comp NO.[a]	Chemical Names	log K_p	C shift (ppm)	H shift (ppm)	pK_a	Predicted log K_p
T94	Ibuprofen	-5.30	179.3	10.75	4.85	-6.42(-)
V122	Pentanoic Acid	-5.96	176.3	11.00	5.01	-6.75(-)
V130	Propanoic Acid	-6.02	178.5	11.00	4.75	-6.56(-)
T28	4-Hydroxyphenylacetic Acid	-6.06	155.5	6.00	4.00, 9.51	-5.24(+)
T52	Butanoic Acid	-6.27	175.9	11.00	4.91	-6.87(-)

[a] T: training set; V: validation set

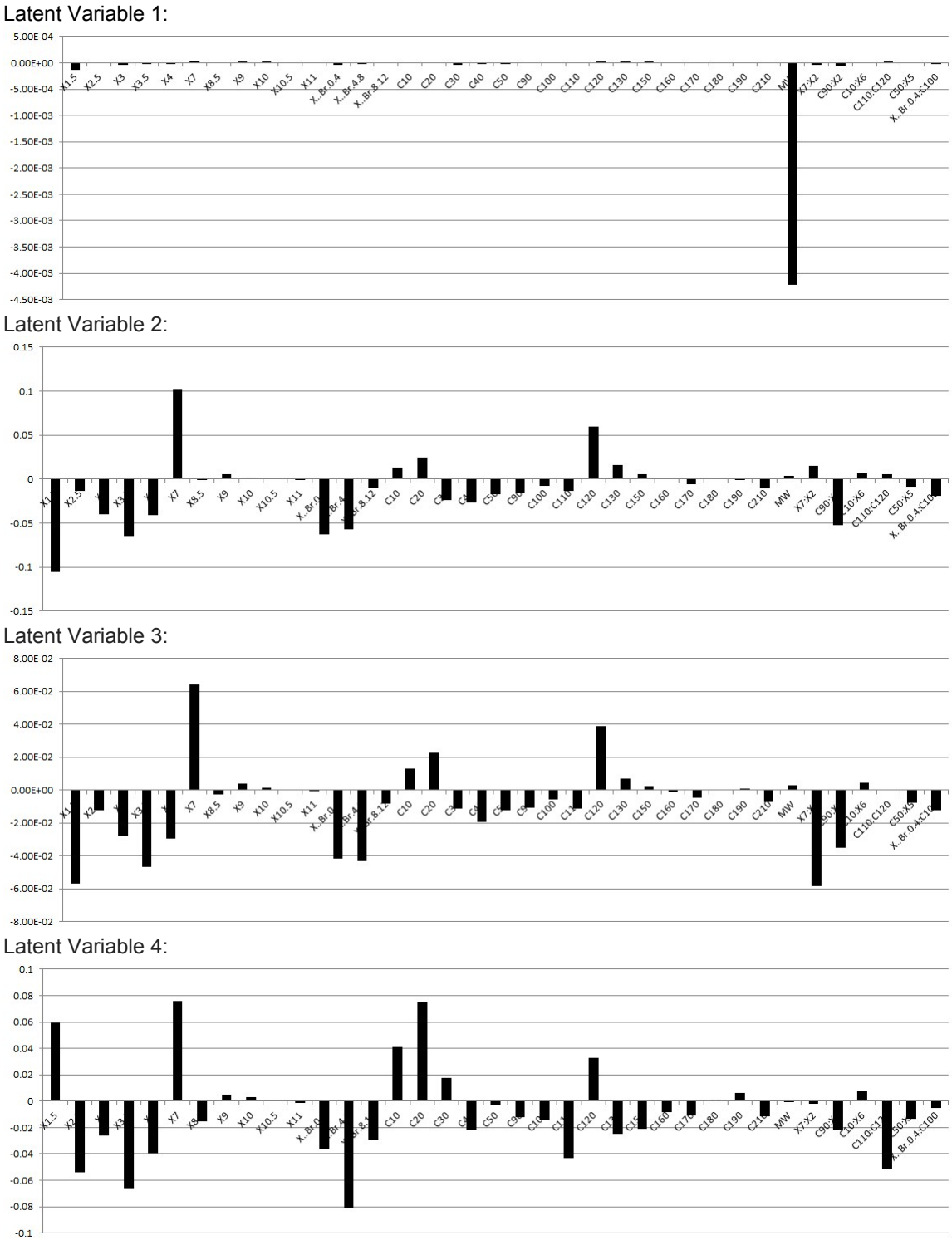

Figure S3. Descriptor contribution of the first four latent variables

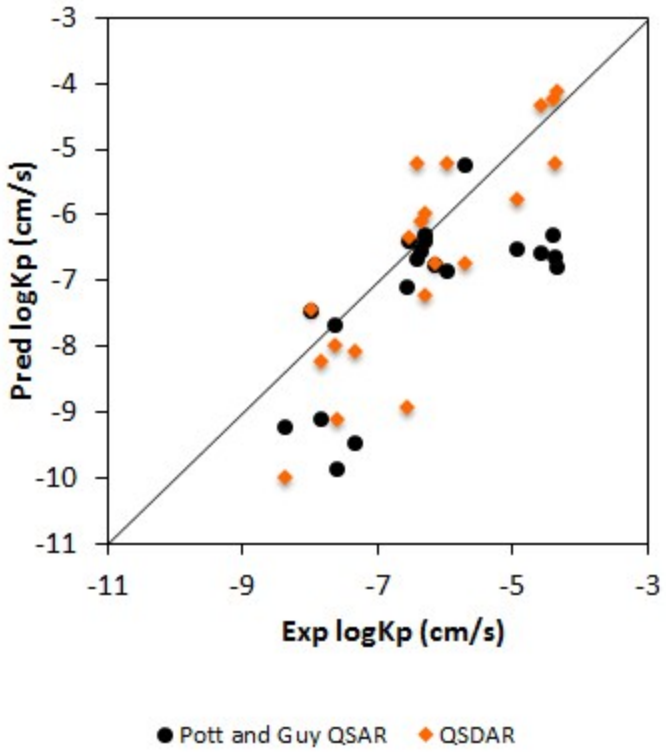

Figure S4. Classic QSAR for skin permeation applied for benchmarking.

Potts and Guy classic QSAR[5] with the equation:

$$\log K_p = 0.71\log P - 0.0061MW - 6.3$$

ESI references

(1) Katritzky, A. R.; Dobchev, D. A.; Fara, D. C.; Hur, E.; Tamm, K.; Kurunczi, L.; Karelson, M.; Varnek, A.; Solov'ev, V. P. *J. Med. Chem.* **2006**, *49*, 3305.

(2) Jorgensen, W. L. In *QikProp version 3.0;*; v. 3.0 ed.; Schrodinger, LLC: New York, NY., 2003.

(3) Klimisch, H. J.; Andreae, M.; Tillmann, U. *Regulatory Toxicology and Pharmacology* **1997**, *25*, 1.

(4) Schneider, K.; Schwarz, M.; Burkholder, I.; Kopp-Schneider, A.; Edler, L.; Kinsner-Ovaskainen, A.; Hartung, T.; Hoffmann, S. *Toxicology Letters* **2009**, *189*, 138.

(5) Potts, R. O.; Guy, R. H. *Pharm. Res.* **1992**, *9*, 663.