SUPPORTING INFORMATION

Validation of Acetylcholinesterase Inhibition Machine Learning Models for Multiple Species

Patricia A. Vignaux*, Thomas R. Lane, Fabio Urbina, Jacob Gerlach, Ana C. Puhl, Scott H. Snyder, and Sean Ekins*

Collaborations Pharmaceuticals, Inc., 840 Main Campus Drive, Lab 3510, Raleigh, NC 27606, USA.

*To whom correspondence should be addressed. E-mail patricia@collaborationspharma.com and sean@collaborationspharma.com Phone: 215-687-1320

Table of Contents

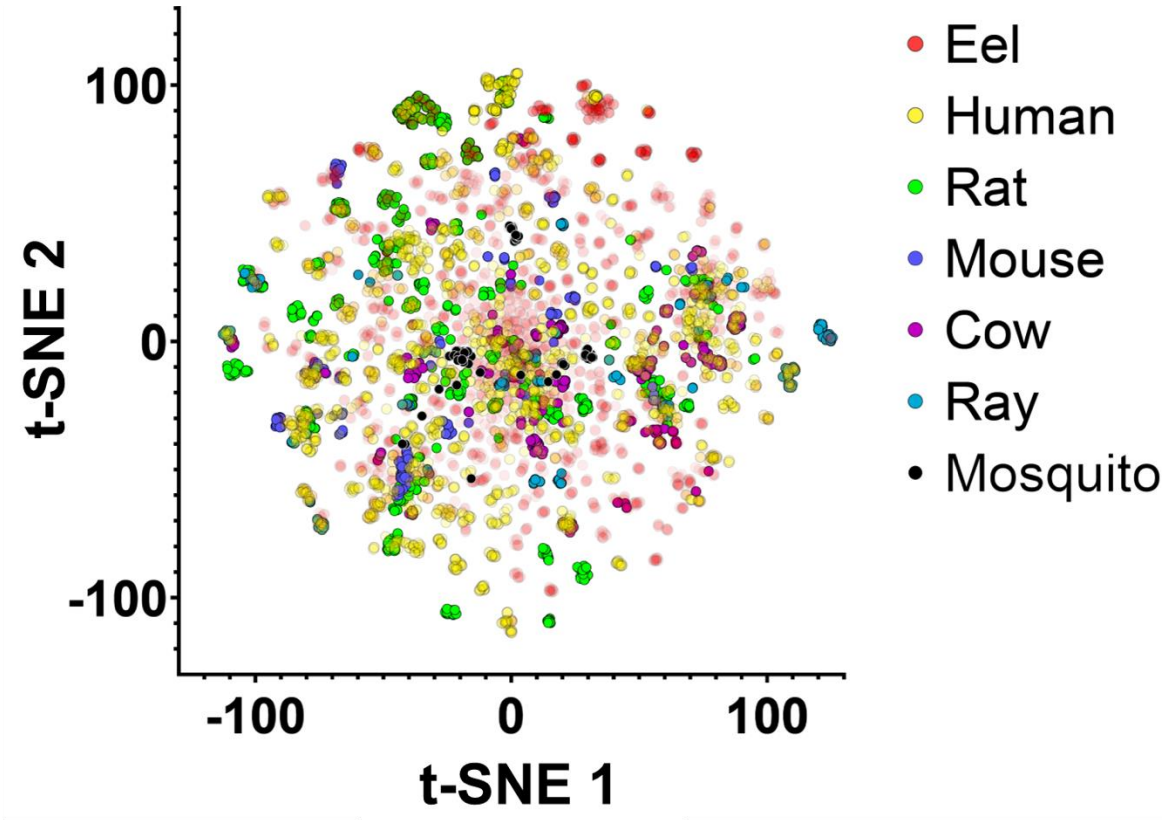

Figure S1. t-SNE Comparison of Training Sets for AChE Classification Models for Seven Different Species. The chemical space represented by condensing the ECFP6 fingerprints of the 10748 molecules used to build the classification models shows extensive overlap between the either the human or eel datasets and the other five species.

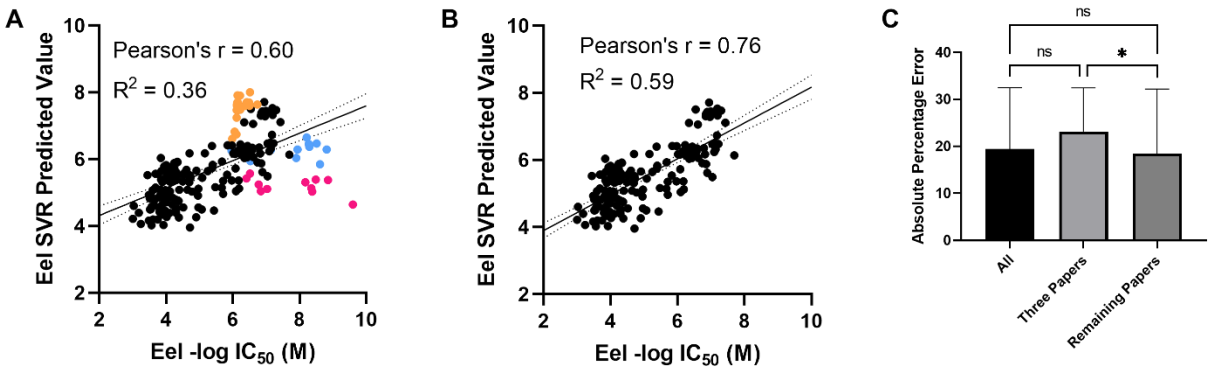

Figure S2. Conjugate synthesis papers in the eel test set compared to the potency of their parent compounds. A) Three Papers. Three publications in the eel test set that that show notable divergence between measured IC50 values and those predicted by the model. Orange molecules are brominated phenylacetic acid/tacrine hybrids from Cheng et al. 2018 [64], blue compounds are 3-aminocumarin-N-benzylpyridinium conjugates from Khunnawutmanotham et al. 2018[65], and magenta compounds are ligustrazine derivatives from Wu et al. 2018 [66]. *Ligustrazine was shown to be not active in human AChE, but is not present in eel model. B) Remaining Set. Removal of these Three Papers from the eel test set increases the correlation between the measured and predicted -log(M) IC50 values. Pearson Correlation and Linear regression performed in GraphPad Prism 9.4.1, Slope = 0.53, Y-intercept = 2.83. C) Comparison of the Absolute Percentage Error of the compounds in the complete eel test set, the compounds from the selected Three Papers, and the compounds in the Remaining Set when predicted with the eel SVR model. P value = 0.0170.

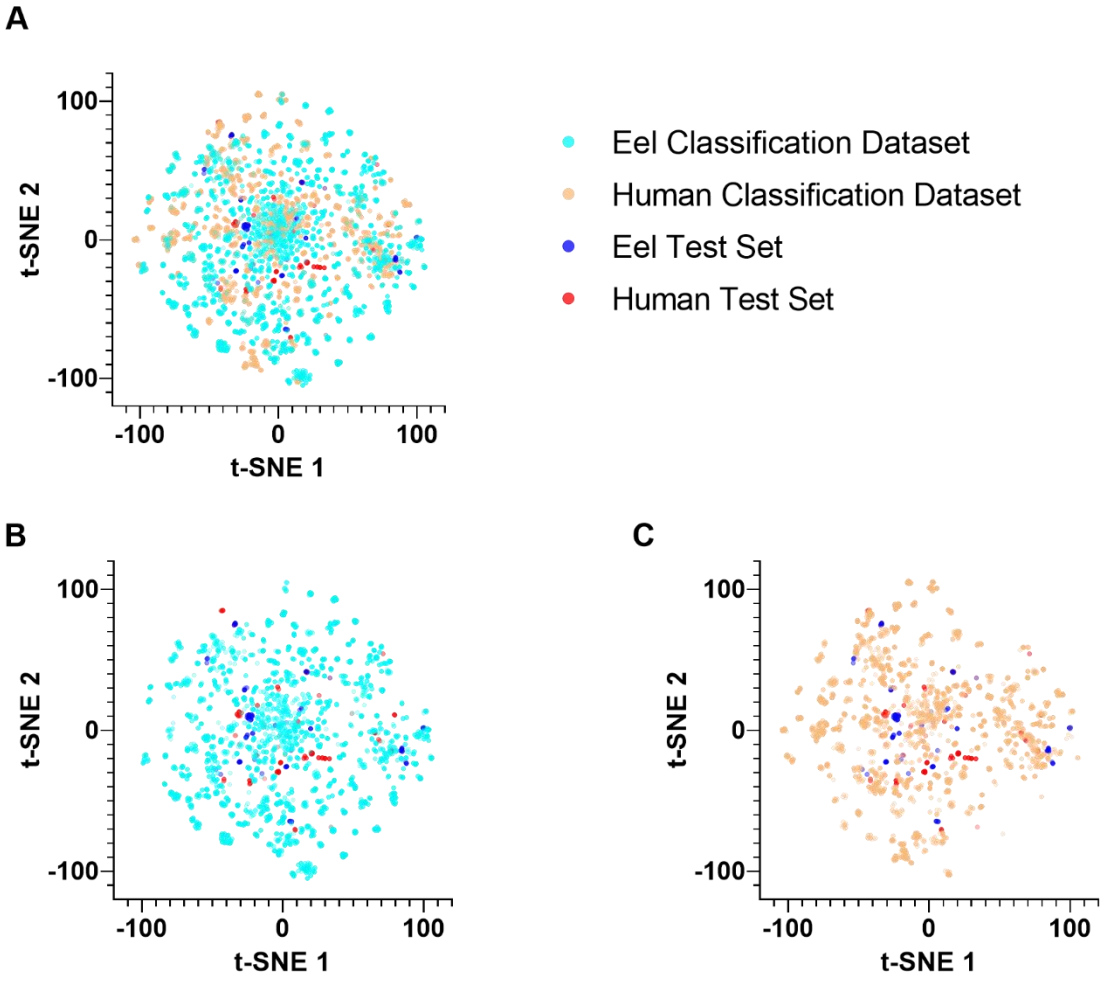

Figure S3. t-SNE plot showing comparative chemical spaces of human and eel continuous datasets. A) Human Test Set and Eel Test Set within the chemical spaces of both the Human and Eel Continuous datasets. B) Both Test Sets compared to Human dataset. C) Both test sets compared to Eel dataset.

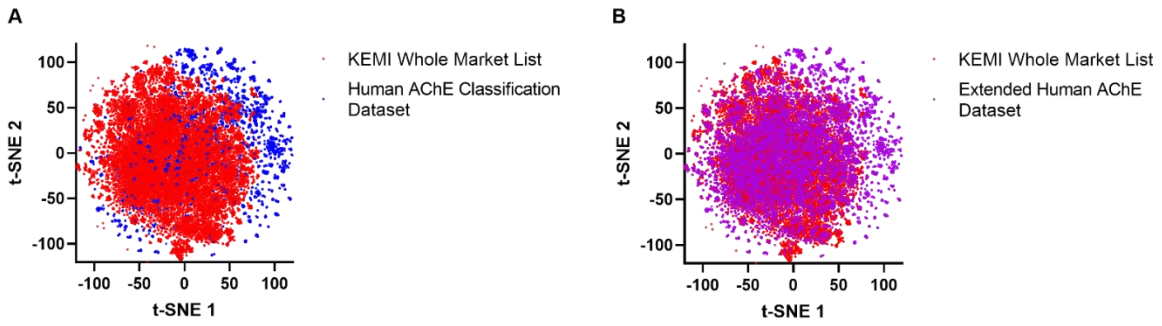

Figure S4. t-SNE plot showing comparative chemical spaces of the KEMI library with A) The dataset of the human AChE classification model and B) the dataset of the Extended Human Classification model supplemented with the Tox21 AChE Screen data.

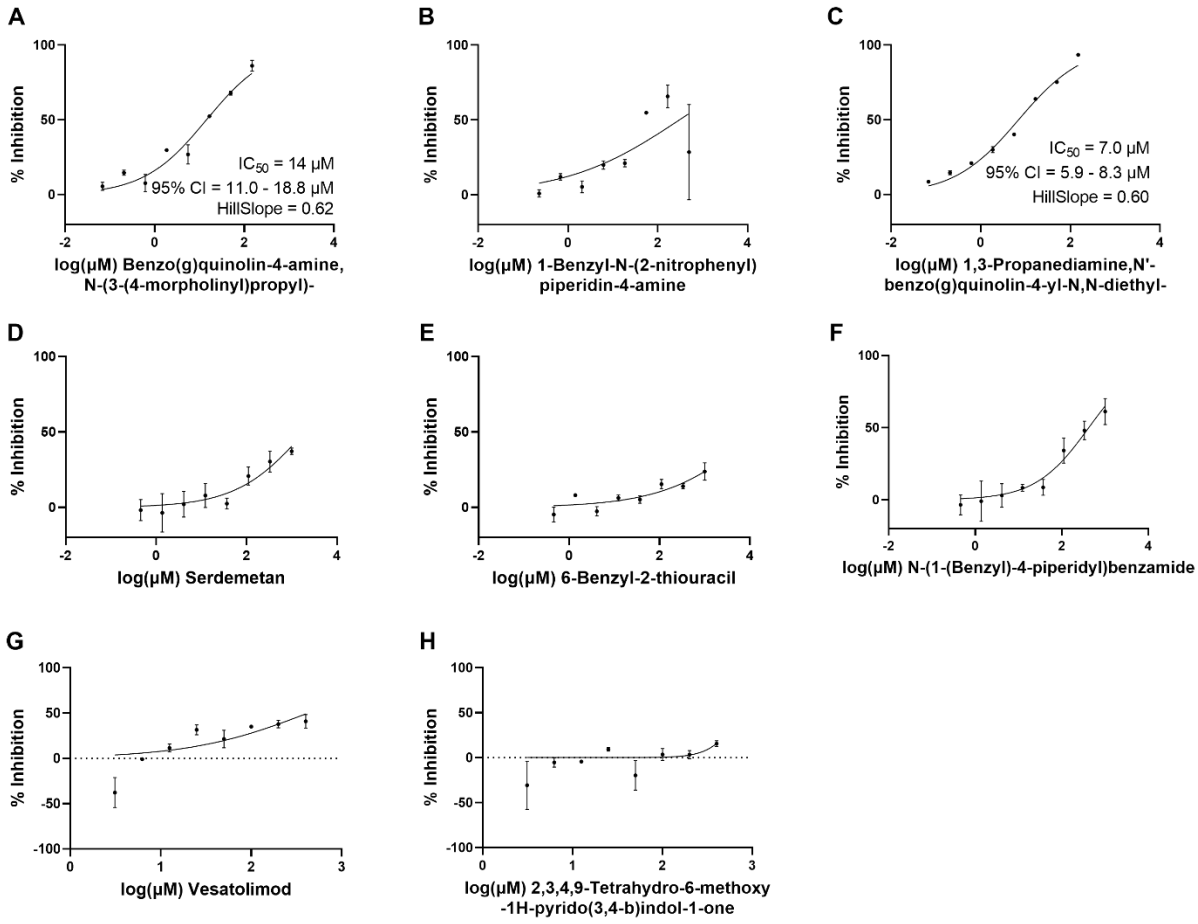

Figure S5. Dose-response of eight compounds from the Compounds of Emerging Concern List on human AChE. Compounds diluted in 1:3 (A-F) or 1:2 (G-H) dilution series in technical triplicate. Curves represent log(inhibitor) vs. response with variable slope (4 parameters, top and bottom constrained to 0 and 100%) generated in GraphPad Prism 9.4.1. Error bars represent standard deviation and are only visible when they exceed the size of the point.

Figure S6. MegaAChE website for prediction of AChE in multiple species (megaache.collaborationspharma.com).

Table S1. Internal 5-Fold cross validation statistics of Extended Human Classification model and 100 μM activity threshold. Human AChE training set: 3560 active/10,382 total compounds.

Model	AUC	F1 Score	precision	recall	accuracy	specificity	Cohen's Kappa	MCC
DL*	0.95	0.83	0.91	0.76	0.89	0.96	0.75	0.76
ADA	0.97	0.89	0.91	0.88	0.93	0.95	0.84	0.84
BNB	0.94	0.82	0.84	0.81	0.88	0.92	0.74	0.74
kNN	0.98	0.94	0.94	0.95	0.96	0.97	0.91	0.91
LREG	0.98	0.91	0.91	0.91	0.94	0.95	0.86	0.86
RF	0.99	0.93	0.96	0.90	0.95	0.98	0.89	0.89
SVC	0.99	0.94	0.95	0.94	0.96	0.98	0.92	0.92
XGB	0.99	0.93	0.94	0.93	0.95	0.97	0.90	0.90

DL = Deep learning; ADA = Adaboosted decision trees; BNB = Bernoulli Naïve Bayes; kNN = K-nearest neighbors; LREG = LogisticRegression; RF = Random forest; SVC = Support vector classification; XGB = XGBoost

Table S2. Predicted hits from CompTox libraries against the Extended Human Classification model, sorted by AD score.

Compound	CAS #	AD	Library
Benzo(g)quinolin-4-amine, N-(3-(4-morpholinyl)propyl)-	66667-67-2	0.96	CEC
Bucricaine	316-15-4	0.95	CEC
Despropionyl para-fluorofentanyl	122861-41-0	0.95	CEC/SUSDAT
N'-(1,2,3,4-tetrahydroacridin-9-yl)octane-1,8-diamine	249290-19-5	0.94	SUSDAT
N-(3-(Morpholino)propyl)naphthalen-2-amine	93762-06-2	0.93	CEC/CompTox/KEMI/SUSDAT
4-Morpholinepropanamine, N-1-naphthalenyl-	5235-82-5	0.93	CEC/SUSDAT
1-(1-(Phenylmethyl)-4-piperidinyl)-1H-benzimidazole	84176-75-0	0.92	CEC/KEMI/SUSDAT
NN'-dimethyl-99'-bisacridinium nitrate	2315-97-1	0.92	CEC
Quindecamine	19056-26-9	0.92	CEC/KEMI/SUSDAT
1-Benzyl-N-(2-nitrophenyl)piperidin-4-amine	7255-89-2	0.90	CEC/KEMI/SUSDAT
1-Benzylpiperidine-4-carbaldehyde	22065-85-6	0.89	CEC/KEMI/SUSDAT
(2-Naphthalenyl)perfluorobutanamide	582-00-3	0.88	PFAS
2,2,3,3,3-pentafluoro-N-naphthalen-2-ylpropanamide	21970-68-3	0.88	PFAS

Mindoperone	52157-83-2	0.88	CEC
AL 3264	118420-47-6	0.88	CEC
Mepixanox	17854-59-0	0.88	CEC/KEMI/SUSDAT
Benzamide, N-(2,4-dinitrophenyl)-4-nitro-	37632-91-0	0.88	CEC/CompTox/KEMI/SUSDAT
2-Methylacridine-9-carbaldehyde	70401-29-5	0.88	CEC/SUSDAT
Buterizine	68741-18-4	0.87	CEC
Picumast	39577-19-0	0.87	CEC
9-Acridinamine, N-(4-(4-ethyl-1-piperazinyl)phenyl)-	143069-08-3	0.87	CEC
Lauralkonium chloride	19486-61-4	0.87	CEC
1,3-Propanediamine, N'-benzo(g)quinolin-4-yl-N,N-diethyl-	56297-68-8	0.87	CEC
Riboflavin	83-88-5	0.87	CEC/CompTox/KEMI/SUSDAT
Serdemetan	881202-45-5	0.86	SUSDAT
3-(2-{[(2-Methoxyphenyl)methyl]amino}ethyl)quinazoline-2,4(1H,3H)-dione	1028307-48-3	0.86	CEC/SUSDAT
7-(ß-D-galactopyranosyloxy)-2-oxo-N-[3-(2367-tetrahydro-13-dimethyl-26-dioxo-1H-purin-8-yl)propyl]-2H-1-benzopyran-3-carboxamide	79181-90-1	0.86	CEC
2-[(1-benzyl-4-piperidyl)amino]butyronitrile	84196-14-5	0.86	CEC
12-Oxospirostan-3-yl acetate	915-35-5	0.86	CEC/SUSDAT

Bis(1,3)benzodioxolo(5,6-a:4',5'-g)quinolizinium, 6,7-dihydro-13-methyl-	30243-28-8	0.86	SUSDAT
Cinepaxadil	69118-25-8	0.86	CEC
E 5880	128420-61-1	0.86	CEC
6-chloro-9-[[3-[(2-chloroethyl)amino]propyl]amino]-2-methoxyacridine dihydrochloride	17070-45-0	0.86	CEC/SUSDAT
Diaminoanthra[912-cde]benzo[rst]pentaphene-510-dione	58019-27-5	0.86	CEC
2-({[2-(2,5-Dimethoxyphenyl)ethyl]amino}methyl)phenol	919797-17-4	0.86	CompTox/SUSDAT
N-(2-(6-Methoxy-1-methyl-9H-carbazol-2-yl)ethyl)-N-methylacetamide	93841-60-2	0.86	CEC/KEMI/SUSDAT
2,3,4,9-Tetrahydro-6-methoxy-1H-pyrido(3,4-b)indol-1-one	17952-87-3	0.85	CEC/KEMI/SUSDAT
moxicoumone	17692-56-7	0.85	CEC
Ilmofosine	83519-04-4	0.85	CEC
1-Benzyl-4-((4-fluorophenyl)amino)piperidine-4-carbonitrile	61085-37-8	0.85	CEC/SUSDAT
1-Benzyl-4-((4-chlorophenyl)amino)piperidine-4-carbonitrile	972-20-3	0.85	CEC/SUSDAT
EINECS 288-151-0	85665-61-8	0.85	CEC/CompTox/SUSDAT
N,N'-Bis(2-(((5-((dimethylamino)methyl)furan-2-yl)methyl)sulfanyl)ethyl)-2-nitroethene-1,1-diamine	72126-78-4	0.85	SUSDAT
Atevirdine	136816-75-6	0.85	CEC
Oxypertine	153-87-7	0.85	CEC/CompTox/KEMI/SUSDAT

Teoprolol	65184-10-3	0.85	CEC
4-Acridinecarboxamide, 9-amino-N-(3-(dimethylamino)propyl)-	89459-44-9	0.85	CompTox/SUSDAT
2-(1H-Benzimidazol-2-yl)aniline	5805-39-0	0.85	CEC/CompTox/KEMI/SUSDAT
Benzene-1,4-diamine, N-(benzo(g)quinolin-4-yl)-N'-(2-(diethylamino)ethyl)-	127136-27-0	0.85	CEC
Bisdequalinium	16776-40-2	0.85	CEC
N-Stearylbenzylamine	20198-87-2	0.85	CEC/KEMI/SUSDAT
N-Dodecylbenzylamine hydrochloride	2090-62-2	0.85	CEC/SUSDAT
6-Benzyl-2-thiouracil	6336-50-1	0.85	CompTox/SUSDAT
N-(1-(Benzyl)-4-piperidyl)benzamide	971-34-6	0.85	CEC/SUSDAT
Indalpine	63758-79-2	0.84	CEC/KEMI/SUSDAT
N-(2-(6-Methoxy-1-methyl-9H-carbazol-2-yl)ethyl)-N-(phenylmethyl)acetamide	93841-55-5	0.84	CEC/KEMI/SUSDAT
1-Benzyl-4-(m-toluidino)piperidine-4-carbonitrile	972-18-9	0.84	CEC/SUSDAT
1-Benzyl-4-(p-toluidino)piperidine-4-carbonitrile	972-19-0	0.84	CEC/SUSDAT
1-Benzylpyrrolidine	29897-82-3	0.84	CEC/KEMI/SUSDAT
4-Piperidinecarboxaldehyde, 1-benzyl-4-phenyl-	26979-21-5	0.84	CEC/SUSDAT
1H-Imidazole-4-carbaldehyde	3034-50-2	0.84	KEMI
Acridine ethidium heterodimer	68942-32-5	0.84	CEC/SUSDAT

3-benzyl-1-decyl-45-dihydro-1H-imidazolium chloride	94088-59-2	0.84	CEC
Aminoquinol	10023-54-8	0.84	CEC
2-[p-(dimethylamino)styryl]-1-methylpyridinium iodide	2156-29-8	0.83	CEC/SUSDAT
1-Benzylpiperidine	2905-56-8	0.86	CEC/KEMI/SUSDAT
1-{benzyl[2-(2-methoxyphenoxy)ethyl]amino}-3-(9H-carbazol-4-yloxy)propan-2-ol	72955-94-3	0.83	CEC
3-[4-(4-Phenyl-3,4-dihydropyridin-1(2H)-yl)butyl]-1H-indole--hydrogen chloride	73966-59-3	0.83	CEC/SUSDAT
Disodium 6-amino-3-[(4-chloro-5-methyl-2-sulphonatophenyl)azo]-4-hydroxynaphthalene-2-sulphonate	68227-39-4	0.83	CEC
Amopyroquine	550-81-2	0.82	CEC/SUSDAT
milipertine	24360-55-2	0.82	CEC
K201 free base	145903-06-6	0.82	SUSDAT
1-Benzyl-4-(propylamino)piperidine-4-carbonitrile	3560-07-40	0.82	SUSDAT
N~2~-(1-Benzylpiperidin-4-yl)butane-1,2-diamine	84196-18-9	0.82	SUSDAT
ELBANIZINE	110629-41-9	0.82	CEC
oxiramide	13958-40-2	0.82	CEC
1-(2-(4-Methoxyphenyl)ethyl)piperidin-4-amine	85098-70-0	0.82	CEC/KEMI/SUSDAT
2-Methoxy-6-chloro-9-[3-[ethyl(2-chloroethyl)amino]propylamino]acridine dihydrochloride	146-59-8	0.82	CEC/SUSDAT

moxipraquine	23790-08-1	0.82	CEC
N1N1-bis(2-chloroethyl)-N4-(6-chloro-2-methoxyacridin-9-yl)pentane-14-diamine dihydrochloride	4213-45-0	0.82	CEC
Coumermycin A1	4434-05-3	0.82	0.82
N,N-Bis(2-chloroethyl)-4-[(6-chloro-2-methoxyacridin-9(10H)-ylidene)amino]pentan-1-amine	64046-79-3	0.82	SUSDAT
N-Cyclohexyl-1H-indole-3-methylamine	53924-03-1	0.82	CEC/CompTox/KEMI/SUSDAT
cloxacepride	65569-29-1	0.82	CEC
C.I. Direct Orange 74	6104-56-9	0.82	SUSDAT
Dalbavancin	171500-79-1	0.81	CEC
Barmastine	99156-66-8	0.81	CEC
4-{[4-(Diphenylmethyl)piperazin-1-yl]methyl}-N-ethyl-2-nitroaniline	68740-16-9	0.81	CEC/SUSDAT
Acridine, 9,9'-ethylenediiminobis-	58903-56-3	0.81	CEC
morocromen	35843-07-3	0.81	CEC
2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-one	17952-82-8	0.81	CEC/KEMI
viqualine	72714-74-0	0.81	CEC
4-[3-(3-Ethenylpiperidin-4-yl)propyl]-6-methoxyquinoline--hydrogen chloride (1/1)	79626-11-2	0.81	SUSDAT

3-Benzyl-3-azaspiro(5.5)undecane hydrochloride	1085-84-3	0.81	CEC/SUSDAT
1H-Pyrido(3,4-b)indol-1-one, 2,3,4,9-tetrahydro-6-chloro-	17952-83-9	0.81	CEC/KEMI/SUSDAT
HOE 33342	23491-52-3	0.81	CEC/KEMI/SUSDAT
Carvedilol Phosphate	610309-89-2	0.81	CEC
Carvedilol	72956-09-3	0.81	CompTox/KEMI/SUSDAT
2-[acetyl(4-dodecylphenyl)amino]-4-[[4-methoxy-3-[[3-(4-methoxyphenyl)-13-dioxopropyl]amino]phenyl]amino]-4-oxobutyric acid	17304-65-3	0.81	CEC
2-Deoxycytidine-5-Triphosphate	2056-98-6	0.81	CEC
2-deoxycytidine diphosphate	800-73-7	0.81	CEC
MDL 62873	122173-74-4	0.80	CEC
1'-Benzylspiro(isoquinoline-1(2H),4'-piperidine)-3(4H)-one hydrochloride	15198-52-4	0.80	SUSDAT
5-Chloro-3-ethyl-N-[2-[4-(1-piperidinyl)phenyl]ethyl]-1H-indole-2-carboxamide	868273-06-7	0.80	SUSDAT
N-(3-(3-((4-Chloro-2,5-dimethoxyphenyl)amino)-1,3-dioxopropyl)phenyl)-N2-(2-methylpropionyl)-N2-octadecyl-L-asparagine	31522-23-3	0.80	CEC/KEMI/SUSDAT
lyfarizine	119514-66-8	0.80	CEC
Pyridinium, 1-[2-[[4-[2-(2,6-dichloro-4-nitrophenyl)diazenyl]phenyl]ethylamino]ethyl]-, sulfate (1:1)	64086-81-3	0.80	SUSDAT
4-((2,3-Dihydro-5-benzofuranyl)methyl)piperazine-1-carbaldehyde	85392-07-0	0.80	CEC/KEMI/SUSDAT

ITROCAINIDE	90828-99-2	0.80	CEC
1-Benzyl-4-(butylamino)piperidine-4-carbonitrile	963-08-6	0.80	CEC/KEMI/SUSDAT
1-Benzyl-4-(2-methylanilino)piperidine-4-carbonitrile	972-17-8	0.80	CEC/SUSDAT

CEC = Human Biomonitoring for Europe Chemicals of Emerging Concern Screen list; CompTox = Compilation of 49 lists from the EPA CompTox Chemicals Dashboard; KEMI = KEMI (Swedish Chemicals Agency) Market List; SUSDAT = Merged NORMAN Suspect List