Supplementary Material

Ligand-based Discovery of New Potential acetylcholinesterase inhibitors for Alzheimer's disease treatment

Y. Cañizares-Carmenate,[a†] N-H. Nam,[b†] R. Díaz-Amador,[c] N.T. Thuan,[b] P.T.P. Dung,[b] F. Torrens,[d] H. Pham-The,[b] F. Perez-Gimenez[e] and J.A. Castillo-Garit[e,f,*]

Table of Contents

Section S1. Preparation of the dataset

In this study, we used a freely available dataset downloaded from the ChEMBL database [1]. The dataset was curated from ChEMBL database, following this steps:

(1) Since May 20th 2020 we started collecting the data from ChEMBL database. In the Bioassay search, we used keyword "Acetylcholinesterase inhibitor". The original output includes 28,304 cases with bioactivity expressed as Ki, IC_{50}, %Inh… (very heterogeneous). We then only extracted compounds with IC_{50} measurements.
(2) We the discarded compounds with inconclusive data (>, <, ≥, ≥) and those with blank data. This step results in 12,806 cases.
(3) Regarding the experimental conditions, we only selected compounds tested on single protein in vitro assay with organism mentioned as homo sapiens (human protein). Then only 3117 compounds satisfied our queries.
(4) By comparing molecular weight and manual structure similarity inspection, we found 539 duplicated compounds.
(5) Salts and metal complexes were detected and discarded. After this step, only 2142 compounds were kept.
(6) Compounds detected by ChEMBL as "outside typical range" và "Potential transcription error" according to Data Validity module, we only obtained 1975 compounds which is the final number of data curated.

The dataset is finally composed by 1975 compounds of great structural diversity and with reported values of IC_{50} enzyme inhibition against AChE.

Section S2. Molecular descriptors

Table S1. Descriptors used in the Piecewise model

Descriptor	Definition	Descriptor Type
nArCONHR	number of secondary amides (aromatic)	Functional group counts--Basic descriptors
nS	number of Sulfur atoms	Constitutional indices--Basic descriptors
nR12	number of 12-membered rings	Ring descriptors--Basic descriptors
P_VSA_s_1	P_VSA-like on I-state, bin 1	P_VSA-like descriptors--Intrinsic State
nCs	number of total secondary C(sp3)	Functional group counts--Basic descriptors
nR10	number of 10-membered rings	Ring descriptors--Basic descriptors
nR=Ct	number of aliphatic tertiary C(sp2)	Functional group counts--Basic descriptors
nHM	number of heavy atoms	Constitutional indices--Basic descriptors
nHM	number of heavy atoms	Constitutional indices--Basic descriptors
nCrs	number of ring secondary C(sp3)	Functional group counts--Basic descriptors

Table S2. Descriptors used in the SVM model (RBF kernel)

Descriptor	Definition	Descriptor Type
nCs	number of total secondary C(sp3)	Functional group counts--Basic descriptors
GGI7	topological charge index of order 7	2D autocorrelations--Topological charge autocorrelations
SpMax2_Bhs	largest eigenvalue n. 2 of Burden matrix weighted by I-state	Burden eigenvalues--Largest eigenvalues
SpMin1_Bhs	smallest eigenvalue n. 1 of Burden matrix weighted by I-state	Burden eigenvalues--Smallest eigenvalues
nR12	number of 12-membered rings	Ring descriptors--Basic descriptors
SIC3	Structural Information Content index (neighborhood symmetry of 3-order)	Information indices--Indices of neighborhood symmetry
totalcharge	total charge	Constitutional indices--Basic descriptors
nP	number of Phosphorous atoms	Constitutional indices--Basic descriptors
nArCOOR	number of esters (aromatic)	Functional group counts--Basic descriptors
P_VSA_s_3	P_VSA-like on I-state, bin 3	P_VSA-like descriptors--Intrinsic State
DBI	Dragon branching index	Topological indices--Vertex degree-based indices
GATS3i	Geary autocorrelation of lag 3 weighted by ionization potential	2D autocorrelations--Geary autocorrelations

Section S3. SVM models (kernel selection)

Table S3. Results using SVM linear kernel.

		Training	
		Observed	
		1	-1
Predicted	1	826	0
	-1	661	0
Se: 1 Sp: 0 Pr: 0.5554 Acc: 0.5555			

		Test	
		Observed	
		1	-1
Predicted	1	273	0
	-1	215	0
Se: 1 Sp: 0 Pr: 0.5594 Acc: 0.5594			

Table S4. Results using SVM quadratic kernel.

		Training	
		Observed	
		1	-1
Predicted	1	826	0
	-1	661	0
Se: 1 Sp: 0 Pr: 0.5554 Acc: 0.5555			

		Test	
		Observed	
		1	-1
Predicted	1	273	0
	-1	215	0
Se: 1 Sp: 0 Pr: 0.5594 Acc: 0.5594			

Table S5. Results using SVM cubic kernel.

		Training	
		Observed	
		1	-1
Predicted	1	826	0
	-1	659	2
Se: 1 Sp: 0.003 Pr: 0.5562 Acc: 0.5598			

		Test	
		Observed	
		1	-1
Predicted	1	272	1
	-1	215	0
Se: 0.9903 Sp: 0 Pr: 0.5585 Acc: 0.5574			

References

[1] D. Mendez, A. Gaulton, A.P. Bento, J. Chambers, M. De Veij, E. Félix, María P. Magariños, Juan F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, Chris J. Radoux, A. Segura-Cabrera, A. Hersey, and Andrew R. Leach, *ChEMBL: towards direct deposition of bioassay data*, Nucleic Acids Research 47 (2019), pp. D930-D940.