Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 17 (2007) 1082–1090
Bioorganic & Medicinal Chemistry Letters

Quantitative structure–activity relationship (QSAR) of tacrine derivatives against acetylcholinesterase (AChE) activity using variable selections

Mankil Jung,* Jungae Tak, Yongnam Lee and Youngae Jung

Department of Chemistry, Yonsei University, Seoul 120-749, Republic of Korea

Received 19 June 2006; revised 16 October 2006; accepted 8 November 2006
Available online 10 November 2006

Abstract—A diverse approach to the quantitative structure–activity relationship (QSAR) of tacrine derivatives against acetylcholinesterase (AChE) activity was studied using variable selections of stepwise multiple linear regression (MLR), genetic algorithm (GA)-MLR, and simulated annealing (SA)-MLR. AChE activity (log RA) of tacrine derivatives was expressed with acceptable explanation (95.5–95.9%) and good predictive power (94.5–95.2%), respectively, in the models. The best equation was obtained from simulated annealing (SA) MLR with greater explanatory capability and better prediction, with a smaller standard error than other methods. The resulting models with the given descriptors illustrate the significant roles of hydrophobic and electrostatic interaction on increasing AChE activity, but hydrophilic and topological feature of molecules were shown to decrease AChE activity.

Alzheimer's disease (AD) is one of the most widespread and prevalent diseases in Western societies;[1] it is a progressive, chronic, and neurodegenerative disorder accompanied by a loss of memory and cognition. Although many factors are involved in AD, its pathogenesis and cause are unclear. The cholinergic hypothesis[2] represents one of the most useful approaches involved in the design of new agents for the treatment of AD. This concept has led to the development and commercialization of several drugs in the United States and Europe for the symptomatic treatment of AD (Fig. 1). Acetylcolinesterase (AChE) inhibitors such as tacrine (THA, Cognex®),[3–5] donepezil (Aricept®),[6–9] rivastigmine (Exelon®),[10] and galanthamine (Reminyl®)[11,12] increase the brain acetylcoline (ACh) levels by preventing the degradation of the released neurotransmitter, thereby enhancing neurotransmission at cholinergic synapses.[2] Tacrine, the first drug approved in the United States for the treatment of AD, was a potent and reversible AChE inhibitor but reported a number of side effects, low selectivity, and hepatotoxicity,[13]

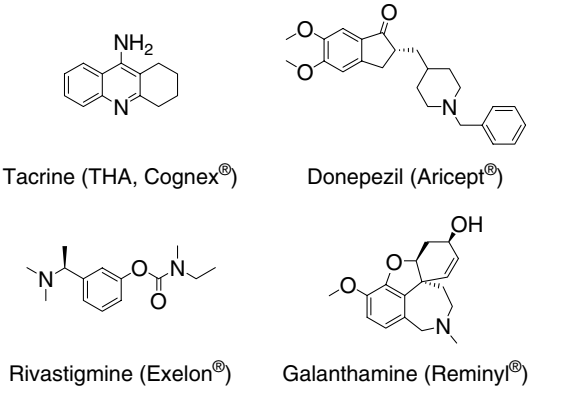


Figure 1. Commercial drugs for AChE inhibitors.

and current research has focused on developing a new drug for the treatment of AD with improved AChE activity and reduced side effects.

The search for tacrine analogues or related new candidates is still of interest to medicinal chemists involved in AD research.[14,15] In this letter, we report on a diverse variable selection approach for a quantitative structure–activity relationship (QSAR) study of tacrine derivatives and suggest new analogues for the treatment of AD.

Keywords: Tacrine; Multiple linear regression (MLR); RV criterion; Quantitative structure–activity relationship (QSAR); Acetylcholinesterase (AChE).
* Corresponding author. Tel.: +82 2 2123 2648; fax: +82 2 364 7050; e-mail: mjung@yonsei.ac.kr

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.11.022

The AChE activity of 80 compounds (a training set of 68 compounds and a test set of 12 compounds from the whole available compounds with known activity), 11*H*-indeno-[1,2-*b*]-quinolin-10-ylamine derivatives,[16] thio-pyranoquinolines,[17] pyranoquinolines and benzonaphthyridines,[18–20] tacrine-E2020 hybrids,[21] bis-tacrine congeners,[22] and tacrine–hurprine heterodimers[23] were taken from published results and used for the present QSAR study. The log value of the relative activity (RA) of these compounds was used and was defined as[24]

$$\log \mathrm{RA} = -\log[(\text{analogues } \mathrm{IC}_{50})/(\text{tacrine } \mathrm{IC}_{50})].$$

Therefore, the compounds more active than tacrine have a positive term of log RA value, while the others less active than tacrine have a negative term of log RA value. The observed AChE activity (log RA) and the calculated activity of tacrine derivatives are represented in Table 1. Geometry optimization of the molecules was calculated by the semiempirical quantum-mechanical method with a PM3 model.[25] The equilibrium geometries were carried into TITAN Pro software.[26]

Regarding the descriptors' calculation and generation, the numerical descriptors are momentous features of the structures of the molecules and can be divided as constitutional, electrostatic, geometrical, topological, and physicochemical descriptors. A total of 133 descriptors was calculated for each compound in the data set by different kinds of software: 110 descriptors were calculated from PreADME on the Web,[27] and 23 descriptors were obtained using the BioMedCAChe 6.10.[28] Table 2 summarizes the descriptors used in this study. The calculated values of the descriptors used in this work are shown in Table 3. Since 133 molecular descriptors are available for QSAR analysis, and only a subset of the whole set is statistically significant in terms of correlation with biological activities, the obtainment of an optimal QSAR model through variable selection needs to be the focus. We selected only the variables that contain the information necessary for the modeling. After this, we selected a total of 32 descriptors with good correlation coefficients ($R \geqslant \pm 0.80$), the variables containing the information necessary for the modeling. We performed the residual analysis and verified the normal distribution of $N(0,1)$ against each model.

The stepwise multiple linear regression (MLR) procedure was used for model selection and is a common method used in QSAR studies. The MLR method performed by the software package SPSS 12.0[29] was used for selection of the descriptors. The main goal of the present work was the development of MLR to predict AChE activity. As the first step, correlation analysis was performed for the selected 32 descriptors calculated for each compound. From a correlation matrix, five descriptors were shown, and these were chosen for the QSAR equation. The QSAR equation was obtained using stepwise multiple regression techniques following the multilinear form

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

In this equation, Y is relative activity, X_1, X_2, and X_n are the descriptors, the intercept (α) and the unstandardized coefficients of the descriptors ($\beta_1, \beta_2, \ldots, \beta_n$) were determined using the least squares method, and n is the number of descriptors. Error term is ε.

Computational experiments were performed with simulated annealing (SA) and genetic algorithms (GA) for the optimization problems resulting from the RV criterion[30] that can be found in the literature, although the methodology discussed can be extended to any numerical criterion that may be considered of interest. The RV indicator was proposed by Robert and Escoufier[30] to measure similarity between two configurations of points in a linear space. The RV-coefficient, for a data matrix (with p variables/columns) X, and for the regression of these columns on a k-variable subset, is given by;

$$\mathrm{RV} = \frac{\mathrm{tr}(XX^t \cdot (P_v X)(P_v X)^t)}{\sqrt{\mathrm{tr}((XX^t)^2) \cdot \mathrm{tr}(((P_v X)(P_v X)^t)^2)}}$$

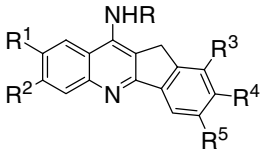
where P_v is the matrix of orthogonal projections on the subspace defined by the k-variable subset. This definition is equivalent to the expression used in the code, which only requires the correlation matrix of the data under consideration.

The SA algorithm begins with a Monte Carlo method for examining the equations of state and frozen states of n-body systems.[31] Simulated annealing has been used in various combinatorial optimization problems and has been particularly successful in circuit design problems.[32] SA is a stochastic technique derived from statistical thermodynamics for finding near-global optimum solutions to complex optimization problems.

The SA algorithm proceeds stepwise through a search space defined by all possible solutions to the optimization problem. After each iteration (e.g., after a variable has been removed), the value of the cost function for the new step is compared to that of the previous step.[33] If the new solution is better than the old one, the removal of the variable is confirmed. If the new solution is worse than the old one, there is still a probability, p, for the removal of the variable to be accepted. This offers the algorithm the possibility of jumping out of a local optimum.[34,35] Otherwise, the removal of the variable will be discarded, and the previous step will be the starting point for the next attempt to eliminate a variable.

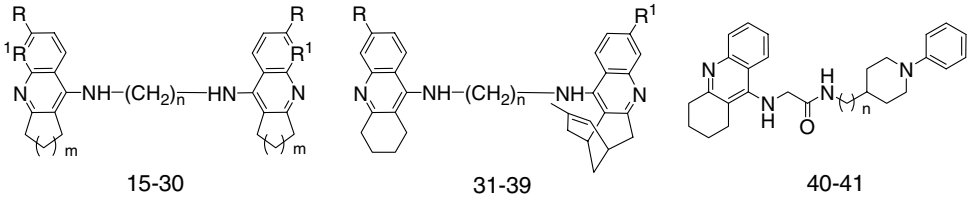
The GA is derived from Darwin's theory of natural selection and evolution, and is expected to solve these defects. A genetic algorithm is controlled by biological evolution rules.[36] These are stochastic optimization methods that have been stimulated by evolutionary principles. The explicit view of a genetic algorithm is that it explores many probable solutions concurrently. Lately, GA has been united with MLR for variable selection, and successful applications have been found in QSAR studies.[37,38] In addition, the descriptors or the ratios of the numbers of variables and compounds are statistically well correlated.[39] GA has the possibility to provide some undesired combinations of variables from the statistical point of view. Also, model validation

Table 1. Structures, observed (log RA) and calculated AChE activity of tacrine derivatives



1-11 12-14

Mol. ID	Structural features						AChE activity of tacrine derivatives						
	R	R^1	R^2	R^3	R^4	R^5	Obs.[a]	Pred.[c] (model 1)	Pred.[c] (model 2)	Pred.[c] (model 3)	Resid. (model 1)	Resid. (model 2)	Resid. (model 3)
1	H	NH_2	H	H	H	H	−1.373	−1.101	−0.995	−0.789	−0.272	−0.378	−0.584
2	H	H	Cl	H	H	H	−1.415	−1.120	−1.483	−1.243	−0.295	0.068	−0.172
3	H	H	H	H	OCH_3	H	−1.415	−1.408	−1.459	−1.197	−0.007	0.044	−0.218
4	H	H	H	H	H	OCH_3	−1.236	−1.407	−1.459	−1.197	0.171	0.223	−0.039
5	H	H	H	CH_3	H	H	−1.193	−1.268	−1.490	−1.107	0.075	0.297	−0.086
6	H	H	H	H	H	CH_3	−1.265	−1.271	−1.443	−1.107	0.006	0.178	−0.158
7	H	H	H	H	Cl	H	−1.334	−1.155	−1.483	−1.243	−0.179	0.149	−0.091
8	CH_3	H	H	H	H	H	−0.716	−1.338	−1.371	−0.996	0.622	0.655	0.280
9[b]	H	H	H	H	H	H	−0.435	−1.365	−1.375	−1.160	0.930	0.940	0.725
10[b]	H	H	F	H	H	H	−0.681	−1.181	−1.156	−0.956	0.500	0.475	0.275
11[b]	H	H	H	OCH_3	H	H	−0.806	−1.393	−1.510	−1.197	0.587	0.704	0.391
12	CHOH	S	—	—	—	—	−0.852	−1.184	−1.084	−1.319	0.332	0.232	0.467
13	CHOH	SO	—	—	—	—	−0.745	−0.653	−0.580	−0.880	−0.092	−0.165	0.135
14[b]	CO	S	—	—	—	—	−0.745	−1.313	−1.194	−1.475	0.568	0.449	0.730

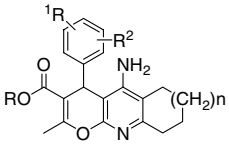


15-30 31-39 40-41

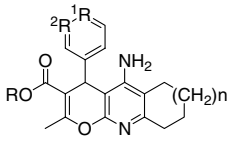
Mol. ID	Structural features				AChE activity of tacrine derivatives						
	R	R^1	*m*	*n*	Obs.[a]	Pred.[c] (model 1)	Pred.[c] (model 2)	Pred.[c] (model 3)	Resid. (model 1)	Resid. (model 2)	Resid. (model 3)
15	H	CH	2	6	2.376	1.900	1.517	1.870	0.476	0.859	0.506
16[b]	H	CH	2	7	2.409	1.861	1.584	1.673	0.548	0.825	0.736
17	H	CH	2	8	2.318	2.223	1.962	2.111	0.095	0.356	0.207
18	F	CH	2	6	2.568	2.284	2.052	2.288	0.284	0.516	0.280
19	F	CH	2	7	2.744	2.424	2.267	2.407	0.320	0.477	0.337
20	F	CH	2	8	2.677	2.579	2.508	2.531	0.098	0.169	0.146
21	Cl	CH	2	6	2.744	2.346	1.779	1.945	0.398	0.965	0.799
22	Cl	CH	2	8	3.045	2.639	2.292	2.221	0.406	0.753	0.824
23	H	N	2	6	1.841	1.714	1.380	1.559	0.127	0.461	0.282
24	H	N	2	8	2.244	2.086	1.813	1.792	0.158	0.431	0.452
25	H	CH	3	6	2.195	1.939	2.170	2.333	0.186	−0.045	−0.208
26	H	CH	3	7	2.091	2.092	2.414	2.462	−0.001	−0.323	−0.371
27	H	CH	3	8	2.318	2.238	2.681	2.597	0.080	−0.363	−0.279
28[b]	H	CH	1	6	0.462	1.661	0.731	1.283	−1.199	−0.269	−0.821
29[b]	H	CH	1	7	0.647	1.810	0.899	1.382	−1.163	−0.252	−0.735
30[b]	H	CH	1	8	1.180	1.958	1.093	1.488	−0.778	0.087	−0.308
31	H	H	*n* = 8		2.398	2.852	2.553	3.034	−0.454	−0.155	−0.636
32	H	Cl	*n* = 6		2.613	2.803	2.244	2.816	−0.190	0.369	−0.203
33	H	Cl	*n* = 7		2.780	2.894	2.482	2.961	−0.114	0.298	−0.181
34	H	Cl	*n* = 8		2.793	3.024	2.740	3.112	−0.231	0.053	−0.319
35	H	Cl	*n* = –$(CH_2)_3$–NMe–$(CH_2)_3$–		2.807	2.638	2.575	3.025	0.169	0.232	−0.218
36	Cl	Cl	*n* = 6		2.659	2.993	2.411	2.888	−0.334	0.248	−0.229
37	Cl	Cl	*n* = 7		2.793	3.123	2.672	3.043	−0.330	0.121	−0.250
38	Cl	Cl	*n* = 8		2.710	3.281	2.940	3.203	−0.571	−0.230	−0.493

Table 1 (*continued*)

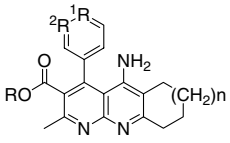
Mol. ID	Structural features				AChE activity of tacrine derivatives						
	R	R^1	*m*	*n*	Obs.[a]	Pred.[c] (model 1)	Pred.[c] (model 2)	Pred.[c] (model 3)	Resid. (model 1)	Resid. (model 2)	Resid. (model 3)
39	Cl	Cl	*n* = –$(CH_2)_3$–Nme–$(CH_2)_3$–		2.849	2.888	2.779	3.121	−0.039	0.070	−0.272
40			*n* = 2		1.570	1.122	1.486	1.297	0.448	0.084	0.273
41			*n* = 1		1.340	0.966	1.355	1.254	0.374	−0.015	0.086



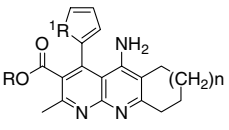
42-55



56-59



60-76



77-80

Mol. ID	Structural features				AChE activity of tacrine derivatives						
	R	R^1	R^2	*n*	Obs.[a]	Pred.[c] (model 1)	Pred.[c] (model 2)	Pred[c] (model 3)	Resid. (model 1)	Resid. (model 2)	Resid. (model 3)
42	Et	H	H	1	−1.079	−1.672	−1.713	−1.600	0.593	0.634	0.521
43	Et	4-CH_3	H	1	−1.146	−1.587	−1.776	−1.572	0.441	0.630	0.426
44	Et	4-Cl	H	1	−1.172	−1.458	−1.693	−1.631	0.286	0.521	0.459
45	Et	4-CN	H	1	−1.426	−1.836	−1.860	−1.725	0.410	0.434	0.299
46	Et	3-NO_2	H	1	−1.163	−1.727	−1.240	−1.277	0.564	0.077	0.114
47	Et	H	H	0	−1.477	−1.895	−1.996	−1.837	0.418	0.519	0.360
48	Et	H	2-OCH_3	0	−2.125	−1.889	−2.121	−1.883	−0.236	−0.004	−0.242
49	Et	3-OCH_3	4-OCH_3	0	−2.176	−2.094	−2.350	−2.065	−0.082	0.174	−0.111
50[b]	Et	4-OCH_3	H	1	−0.825	−1.732	−1.739	-1.635	0.907	0.914	0.810
51	Et	H	2-OCH_3	1	−1.631	−1.737	−1.821	−1.635	0.106	0.190	0.004
52	Et	3-OCH_3	4-OCH_3	1	−1.407	−1.860	−2.037	−1.812	0.453	0.630	0.405
53[b]	Et	4-OCH_3	3-OCH_3	2	−2.263	−1.760	−1.790	−1.622	−0.503	−0.473	−0.641
54	Et	H	3-OCH3	2	−1.859	−1.694	−1.491	−1.444	−0.165	−0.368	−0.415
55	Et	H	2-OCH3	2	−2.067	−1.635	−1.580	−1.444	−0.432	−0.487	−0.623
56	Et	N	CH	0	−1.673	−1.897	−1.767	−1.739	0.224	0.094	0.066
57	Et	N	CH	1	−1.331	−1.686	−1.485	−1.502	0.355	0.154	0.171
58	Et	N	CH	2	−1.788	−1.633	−1.253	−1.316	−0.155	−0.535	−0.472
59	Et	CH	N	1	−1.706	−1.712	−1.485	−1.502	0.006	−0.221	−0.204
60	Et	CH	CH	0	−1.962	−1.063	−1.590	−1.122	−0.899	−0.372	−0.840
61	Et	CH	CH	1	−0.801	−0.928	−1.315	−0.888	0.127	0.514	0.087
62[b]	Et	CH	CH	2	−0.768	−0.851	−1.090	−0.706	0.083	0.322	−0.062
63[b]	Et	$CHOCH_3$	H	0	−2.023	−1.118	−1.647	−1.172	−0.905	−0.376	−0.851
64	Et	$CHOCH_3$	H	1	−1.839	−0.984	−1.349	−0.928	−0.855	−0.490	−0.911
65	Et	$CHOCH_3$	H	2	−1.016	−0.848	−1.109	−0.740	−0.168	0.093	−0.276
66	Et	H	$CHOCH_3$	1	−0.824	−0.903	−1.349	−0.928	0.079	0.525	0.104
67	Et	$CHOCH_3$	$CHOCH_3$	1	−0.859	−0.958	−1.653	−1.108	0.099	0.794	0.249
68	Et	CHF	CH	1	−0.596	−0.728	−1.062	−0.685	0.132	0.466	0.089
69	Et	CHF	CH	2	−1.477	−0.656	−0.841	−0.502	−0.821	−0.636	−0.975
70	Et	CHCl	CH	1	−0.859	−0.660	−1.302	−0.923	−0.199	0.443	0.064
71	Et	CH	CHNO2	1	−0.713	−0.924	−0.852	−0.570	0.211	0.139	−0.143
72	Et	$CHOCH_3$	$CHOCH_3$	2	−1.551	−0.927	−1.413	−0.921	−0.624	−0.138	−0.630
73	*i*-Pr	CH	CH	1	−0.824	−1.215	−1.156	−0.907	0.391	0.332	0.083
74	Et	N	CH	0	−1.768	−0.939	−1.332	−1.177	−0.829	−0.436	−0.591
75	Et	N	CH	1	−1.000	−0.783	−1.056	−0.943	−0.217	0.056	−0.057
76	Et	N	CH	2	−1.561	−0.698	−0.861	−0.607	−0.863	−0.700	−0.954
77	Et	R^1 = S		0	−2.298	−1.962	−2.343	−2.343	−0.336	0.045	0.045
78	Et	R^1 = S		1	−2.248	−1.806	−2.060	−2.110	−0.442	−0.188	−0.138
79	Et	R^1 = S		2	−2.012	−1.656	−1.843	−1.926	−0.356	−0.169	−0.086
80	Et	$NCOCH_e$		1	−1.084	−1.790	−1.726	−1.536	0.706	0.642	0.452

[a] Relative activity (log RA).
[b] Test set.
[c] Model 1 (MLR), model 2 (SA-MLR), model 3 (GA-MLR).

Table 2. Definition of the used descriptors from each model

Descriptors	Definition
log *P*	Octanol–water partition coefficient, calculated by Ghose's atom additive method
SE	Steric energy (kcal/mol), the optimization to find a low-energy structure
SIK1	Shape index of order 1 (κ_1), quantifying the number of cycles in the chemical sample
SIKA3	Shape index of order 3 (κ_3), quantifying the degree of branching toward the center of the chemical sample
VCI1	First-order (bond) valence molecular connectivity index ($^1\chi^V$) for the chemical sample
VDWSURFA	Two-dimensional van der Waals surface area
NORB	Number of rigid bonds
WNSA1	Surface-weighted charged partial negative surface area, first type
SFE	Water solvation free energy, calculated by Ghose's atom additive method

is necessary to choose the best variable combination from GA.[40]

In this work, we performed SA and GA on a 3.2-GHz Pentium IV, and the R statistical software package, called Subselect,[41] is free software. The quality of each model was proven by parameters such as explained variance (R^2), cross-validated R^2 (Q^2), the standard error (SE), the variance ratio of calculated and observed activities (F), the predicted residual sum of squares (PRESS), and the standard error of prediction (SDEP).

MLR analysis with stepwise selection was employed to model the structure–activity relationships into the software package SPSS. From the result of the stepwise addition of terms, we set up the QSAR model 1 and showed the 95% confidence intervals of the regression coefficients. In addition, explained variance (R^2) and cross-validated R^2(Q^2) were well estimated over 94%. AChE activity (log RA) of tacrine derivatives was expressed with acceptable statistical significance in model 1:

Model 1

$$\begin{aligned}\log \mathrm{RA} = &-5.094(\pm 0.456) + 0.207(\pm 0.047)\log P \\ &- 0.017(\pm 0.008)\mathrm{SE} + 0.679(\pm 0.096)\mathrm{VCI1} \\ &- 0.258(\pm 0.040)\mathrm{SIK1} - 0.112(\pm 0.013)\mathrm{SFE}\end{aligned}$$

$n = 68$, $R = 0.978$, $R^2 = 0.957$, $Q^2 = 0.946$, SE = 0.404, $F = 278.5$ (df 5, 62), PRESS = 10.101, SDEP = 0.224.

In the MLR model, n is the compound's number, F is the F-ratio, and df is degree of freedom. The regression coefficient in the model is significant at the 95% level. Model 1 could explain and predict 95.7% and 94.6% of the descriptors, respectively. The residuals of model 1 have the average value, $-1.8E{-}16$. The calculated values according to model 1 are represented in Table 1. AChE activity relates to the blood–brain barrier, so AChE activity should be considered in terms of the lipophilicity and the electronic contribution in each molecule. The positive coefficient of molecular hydrophobicity (log *P*) is considered as brain uptake of drugs and the descriptor (VCI1)[42] indicates the degree of branching, connectivity of atoms, and the unsaturation in the molecule, and in a number of cases, it has been found to be significantly correlated with the hydrophobic property of the molecules. Whereas the negative coefficient of the solvation free energy (SFE) expresses hydrophilicity, and the steric energy (SE) and the Kier shape index first-order (SIK1)[43,44] influence active site of AChE. We schematized that the model is able to explain good correlation with the scatterplot (Fig. 2).

A SA algorithm focuses on a substitute for the full set around a used RV criterion. The whole process of variable selection is repeated a fixed number of times for monotonically decreasing annealing temperatures.[45] After every change in the annealing temperature, the algorithm starts selecting variables from the complete set of variables. An initial k-variable ranging five variables of a whole set of p variables is randomly selected for this algorithm because five variables showed good explained variance (R^2) and cross-validated R^2 (Q^2) values than other things. In this work, we used 2000 iterations and 20 solutions. These selected variables used MLR analysis with a stepwise selection, and we estimated selected variables according to significance at the 95% level repeatedly. The best selected variables were employed to model SA-MLR. The model was established, leave-one-out (LOO) was calculated, and the activity of AChE was represented by the resulting equation. The best equation derived was the following (model 2):

Model 2

$$\begin{aligned}\log \mathrm{RA} = &-1.993(\pm 0.561) + 0.764(\pm 0.218)\mathrm{SIKA3} \\ &+ 0.104(\pm 0.262)\mathrm{NORB} - 0.081(\pm 0.013)\mathrm{SFE} \\ &- 0.027(\pm 0.003)\mathrm{VDWSURFA} \\ &+ 0.027(\pm 0.004)\mathrm{WNSA1}\end{aligned}$$

$n = 68$, $R = 0.979$, $R^2 = 0.959$, $Q^2 = 0.952$, SE =0.395, $F = 290.8$ (df 5, 62), PRESS = 9.691, SDEP = 0.233.

Model 2 shows excellent equation statistics and cross-validation parameters. From the result of SA-MLR the regression coefficients in model 2 are significant at the 95% level. Model 2 could explain and predict 95.9% and 95.2%, respectively, of the variance of the AChE activity. The estimated values using model 2 are shown in Table 1; also, the SE has a very small value and the residuals of model 2 have the almost absolute zero, −8.12E−16. The positive coefficient of the Kier shape κα index third-order (SIKA3)[43,44] stresses the importance of branching at the extreme part of the molecule for AChE activity. The number of rigid bonds (NORB) indicates flexible bonds of tacrine on the end position and the increment of NORB increases AChE

Table 3. Values of the descriptors in the data set in this work

Mol. ID	log *P*	SE	SIK1	SIKA3	VCI1	VDWSURFA	NORB	WNSA1	SFE
1	2.687	−21.759	12.719	1.427	6.297	221.967	35.000	28.190	−18.510
2	3.988	−22.644	12.719	1.506	6.575	225.480	33.000	32.296	−14.110
3	3.218	−19.841	13.648	1.610	6.621	240.135	36.000	29.924	−15.250
4	3.218	−19.905	13.648	1.610	6.621	240.135	36.000	29.924	−15.250
5	3.938	−21.946	12.719	1.376	6.514	225.131	36.000	26.084	−13.360
6	3.938	−22.007	12.719	1.437	6.508	225.131	36.000	26.084	−13.360
7	3.988	−20.616	12.719	1.506	6.575	225.480	33.000	32.296	−14.110
8	3.827	−6.205	12.719	1.437	6.559	228.801	36.000	27.317	−15.060
9	3.470	−21.47	11.796	1.273	6.098	207.899	33.000	26.176	−13.830
10	3.610	−23.366	12.719	1.420	6.197	212.308	33.000	26.657	−16.450
11	3.218	−20.495	13.648	1.544	6.627	240.135	36.000	29.924	−15.250
12	1.295	−7.704	11.111	1.537	6.354	205.608	30.000	24.951	−18.420
13	0.158	−6.047	12.055	1.668	6.854	214.992	31.000	26.724	−24.660
14	0.956	−8.821	11.111	1.456	6.207	200.371	28.000	25.035	−18.620
15	7.493	17.126	26.234	5.056	14.170	460.302	72.000	92.761	−25.720
16	6.198	7.854	27.184	5.364	14.390	473.823	72.000	95.543	−27.220
17	8.286	16.743	28.135	5.796	15.170	494.766	76.000	108.316	−25.400
18	7.772	11.582	28.135	5.460	14.370	469.120	72.000	94.213	−30.960
19	8.168	12.594	29.089	5.824	14.870	486.352	74.000	101.906	−30.800
20	8.565	12.739	30.044	6.210	15.370	503.584	76.000	109.898	−30.640
21	8.529	16.524	28.135	5.725	15.126	495.465	72.000	117.017	−26.280
22	9.322	17.781	30.044	6.490	16.126	529.929	76.000	134.379	−25.960
23	5.802	7.212	26.234	5.006	13.890	456.591	70.000	88.096	−27.380
24	6.594	3.931	28.135	5.743	14.890	491.055	74.000	103.291	−27.060
25	8.286	33.472	28.135	5.796	15.170	494.766	78.000	108.316	−25.400
26	8.682	33.700	29.089	6.178	15.670	511.998	80.000	116.544	−25.240
27	9.078	34.363	30.044	6.580	16.170	529.230	82.000	125.072	−25.080
28	6.700	12.38	24.343	4.361	13.170	425.838	66.000	70.994	−26.040
29	7.097	12.992	25.288	4.697	13.670	443.070	68.000	77.722	−25.880
30	7.493	13.697	26.234	5.056	14.170	460.302	70.000	84.750	−25.720
31	9.128	25.990	31.592	6.144	17.275	553.570	87.000	134.758	−26.070
32	8.854	23.047	30.643	5.771	16.752	536.687	83.000	130.607	−26.670
33	9.250	27.014	31.592	6.117	17.252	553.919	85.000	139.621	−26.510
34	9.646	28.689	32.542	6.480	17.752	571.151	87.000	148.934	−26.350
35	7.641	31.924	32.542	6.332	17.332	590.323	88.000	152.408	−29.650
36	9.372	24.672	31.592	6.090	17.230	554.268	83.000	144.490	−26.950
37	9.768	26.375	32.542	6.451	17.730	571.500	85.000	153.955	−26.790
38	10.164	26.345	33.495	6.811	18.230	588.732	87.000	163.720	−26.630
39	8.159	30.002	33.495	6.659	17.810	607.904	88.000	167.480	−29.930
40	4.989	22.556	24.684	5.322	12.488	441.391	66.000	72.846	−30.860
41	4.665	21.270	23.728	4.931	11.988	424.159	64.000	70.002	−30.700
42	3.255	−18.952	20.280	3.119	9.532	348.861	50.000	46.600	−10.590
43	3.722	−19.539	21.240	3.336	9.943	366.093	53.000	45.224	−10.120
44	3.773	−18.871	21.240	3.430	10.010	366.442	50.000	55.298	−10.870
45	3.291	−18.715	22.203	3.380	9.916	372.591	51.000	51.814	−11.200
46	3.209	−19.806	23.168	3.622	10.031	376.999	52.000	61.056	−13.680
47	2.859	−15.041	19.322	2.817	9.032	331.629	47.000	38.527	−10.750
48	2.606	−17.107	21.240	3.118	9.561	363.865	50.000	41.802	−12.170
49	2.353	−7.099	23.168	3.522	10.084	396.100	53.000	38.292	−13.590
50	3.002	−17.467	22.203	3.531	10.055	381.097	53.000	50.413	−12.010
51	3.002	−16.897	22.203	3.423	10.061	381.097	53.000	50.413	−12.010
52	2.749	−11.852	24.135	3.837	10.584	413.332	56.000	47.151	−13.430
53	3.146	−8.675	25.104	4.176	11.084	430.565	59.000	52.863	−13.270
54	3.399	−10.568	23.168	3.870	10.555	398.329	56.000	56.159	−11.850
55	3.398	−13.825	23.168	3.753	10.561	398.329	56.000	56.159	−11.850
56	1.547	−13.872	19.322	2.796	8.882	329.773	46.000	39.106	−14.240
57	1.943	−17.087	20.280	3.097	9.382	347.005	49.000	47.188	−14.080
58	2.339	−11.024	21.240	3.424	9.882	364.237	52.000	52.701	−13.920
59	1.943	−15.586	20.280	3.097	9.382	347.005	49.000	47.188	−14.080
60	3.995	−5.610	19.322	2.730	8.955	329.233	46.000	35.783	−17.980
61	4.391	−4.348	20.280	3.028	9.455	346.465	49.000	43.686	−17.820
62	4.787	0.322	21.240	3.352	9.955	363.697	52.000	49.023	−17.660
63	3.742	−4.281	21.240	3.128	9.478	361.469	49.000	38.820	−19.400
64	4.138	−3.069	22.203	3.438	9.978	378.701	52.000	47.261	−19.240
65	4.535	−1.940	23.168	3.772	10.478	395.933	55.000	52.856	−19.080

(*continued on next page*)

Table 3 (*continued*)

Mol. ID	log *P*	SE	SIK1	SIKA3	VCI1	VDWSURFA	NORB	WNSA1	SFE
66	4.138	−7.845	22.203	3.438	9.978	378.701	52.000	47.261	−19.240
67	3.886	−6.469	24.135	3.743	10.507	410.937	55.000	43.800	−20.660
68	4.531	−7.713	21.240	3.222	9.555	350.874	49.000	44.120	−20.440
69	4.909	−7.395	21.240	3.336	9.933	364.046	49.000	52.210	−18.100
70	4.927	−2.809	22.203	3.539	10.055	368.106	52.000	49.490	−20.280
71	4.345	−8.648	23.168	3.528	9.954	374.603	51.000	57.863	−20.910
72	4.282	0.752	25.104	4.078	11.007	428.169	58.000	49.363	−20.500
73	1.493	−21.717	19.322	2.748	8.921	327.377	45.000	36.370	−21.470
74	1.890	−21.650	20.280	3.047	9.421	344.609	48.000	44.282	−21.310
75	3.476	−7.651	21.240	3.330	9.805	361.841	51.000	49.645	−21.150
76	4.138	−3.605	22.203	3.438	9.978	363.697	52.000	45.858	−17.090
77	1.941	−6.012	18.367	2.695	9.396	322.301	44.000	37.186	−8.810
78	2.337	−5.975	19.322	3.003	9.896	339.533	47.000	45.083	−8.650
79	2.734	−5.542	20.280	3.314	10.396	356.765	50.000	50.476	−8.490
80	1.815	−11.790	22.203	3.298	9.986	382.863	53.000	50.391	−14.960

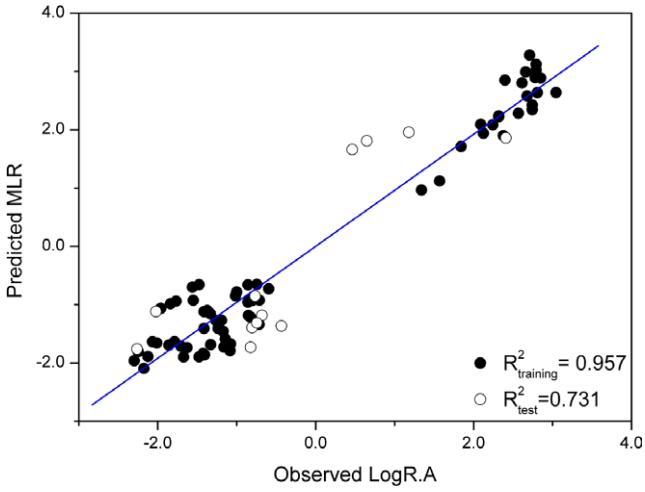


Figure 2. Scatterplots of observed versus calculated RA for AChE activity using model 1 in Orgin 7.0.

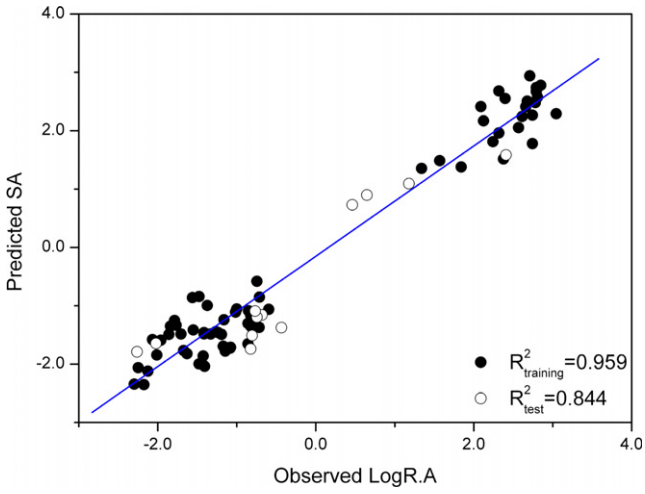


Figure 3. Scatterplots of observed versus calculated RA for AChE activity using model 2 in Orgin 7.0.

activity. In addition, the surface-weighted charged partial surface area (WNSA1)[46,47] makes an important contribution for binding to enzyme active site. However, the van der Waals term (VDWSURFA) reflects the potential related to the difference between the inter-atomic distance and the sum of the van der Waals radii. We show that model 2 is able to explain the scatterplot (Fig. 3).

The GA-based variable selection method used far fewer molecular descriptors, which is important for using the derived model in the QSAR study. These models also have higher predictive accuracy, especially cross-validated predictive accuracy. A typical QSAR table consists of compound rows and descriptor columns. The optimization goal is to find a combination of descriptors (variable selection to generate a set of descriptors). With a known set of variables, a GA finds a k-variable subset that is best with a given RV criterion. We found that each cardinality k was ranging five variables and that an initial population of 200 k-variable subsets was randomly selected by a whole set of p variables because five variables showed the best explained variance (R^2) and cross-validated R^2 (Q^2) values than other variables. The generation of 10,000 was used as the stopping RV criterion. To investigate how well GA performed, we also compared the descriptors to the significant probabilities with all equations obtained from GA. If descriptors did not correlate with the significant probabilities at 95%, they were excluded from the variable set. The GA-MLR was processed in the same way as the SA method. A LOO was performed to calculate the activity of the molecules by the resulting equation. The performance of the GA is measured by comparison to the root-mean-square of prediction (RMSEP) of the model proposed by the GA. Model 3 was derived from GA variable selection using the RV criterion:

Model 3

$$\begin{aligned}\log \mathrm{RA} = &-3.588(\pm 0.359) + 0.129(\pm 0.054)\log P \\ &+ 0.111(\pm 0.028)\mathrm{NORB} - 0.096(\pm 0.012)\mathrm{SFE} \\ &- 0.017(\pm 0.003)\mathrm{VDWSURFA} \\ &+ 0.020(\pm 0.005)\mathrm{WNSA1}\end{aligned}$$

$n = 68$, $R = 0.977$, $R^2 = 0.955$, $Q^2 = 0.945$, SE = 0.415, RMSEP = 0.440, SDEP = 0.223, $F = 263.3$ (df 5, 62), PRESS = 10.655.

Model 3 with five variables could explain 95.5% of the variance and predict 94.5% of the variance. The residuals of Model 3 showed the almost absolute zero, −1.23E−15. The presence of two negative descriptors indicates that the large molecule surface area (VDW-SURFA) and high solvation on water (SFE) decrease AChE activity. The positive coefficient of the molecular hydrophobicity ($\log P$), the number of rigid bonds (NORB), and the partial negative surface area (WNSA1) are important descriptors in favor for the AChE activity. However, Kier shape index third-order (SIKA3) is better descriptor than the hydrophobicity descriptor ($\log P$) because model 2 was derived from excellent QSAR equation by using SIKA3. The calculated values according to model 3 are presented in Table 1. Through this model we designed a scatterplot (Fig. 4).

In this QSAR study, diverse approach for the variable selection of stepwise MLR, SA-MLR, and GA-MLR showed good predictivity. The analysis of the multiple linear regression models gives an insight into the characteristic features of the tacrine derivatives. The binding to enzyme active site and the hydrophobicity increase AChE activity. The results of the QSAR study suggest that the hydrophobicity value and rigidity of compounds make increased inhibitory potency of tacrine derivatives against respective AChE activity with preferential order as SA-MLR > GA-MLR > MLR. However, hydrophilic and topological feature of molecules were shown to decrease AChE activity. The data presented in Table 1 show that the observed and the predicted activities are very close to each other evidenced by low values of residual activity. Finally, we found that SA-MLR is the best variable selection method compared to other methods because this method has shown greater explanatory capability and better prediction, with a smaller standard error than other methods.

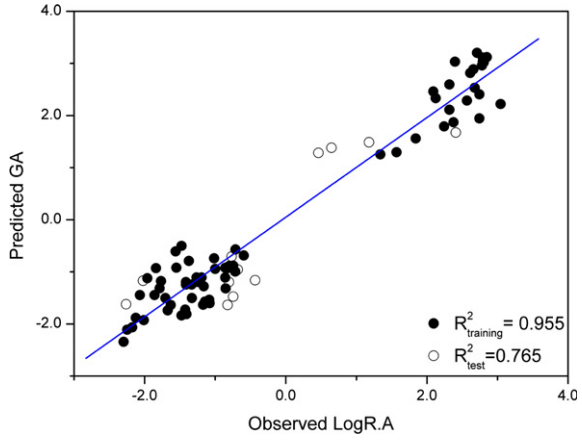


Figure 4. Scatterplots of observed versus calculated RA for AChE activity using model 3 in Orgin 7.0.

Acknowledgment

This work was supported by the Korea Research Foundation (Grant KRF-2003-015-C00380).

References and notes

1. Guttman, R.; Altman, R. D.; Nielsen, N. H. *Arch. Fam. Med.* **1999**, *8*, 347.
2. Winkler, J.; Thal, L. J.; Gage, F. H.; Fisher, L. J. *J. Mol. Med.* **1998**, *76*, 555.
3. Summers, W. K.; Majovski, L. V.; Marsh, G. M.; Tachiki, K.; Kling, A. *N. Engl. J. Med.* **1986**, *315*, 1241.
4. Tabarrini, O.; Cecchetti, V.; Temperini, A.; Filipponi, E.; Lamperti, M. G.; Fravolini, A. *Bioorg. Med. Chem.* **2001**, *9*, 2921.
5. Kurz, A. *J. Neural. Transm. Suppl.* **1998**, *54*, 295.
6. Sugimoto, H.; Tsuchiya, Y.; Sugumi, H.; Higurashi, K.; Karibe, I. Y.; Sasaki, A.; Kawakami, Y.; Nakamura, T.; Yamanishi, Y.; Yamatsu, K. *J. Med. Chem.* **1990**, *33*, 1880.
7. Sugimoto, H.; Iimura, Y.; Yamaniohi, Y.; Yamatsu, K. *J. Med. Chem.* **1995**, *38*, 4821.
8. Inoue, A.; Kawai, T.; Wakita, M.; Iimura, Y.; Sugimoto, H.; Kawakami, Y. *J. Med. Chem.* **1996**, *39*, 4460.
9. Sugimoto, H.; Yamanishi, Y.; Iimura, Y.; Yoshiyuki, K. *Curr. Med. Chem.* **2000**, *7*, 303.
10. Jann, M. W. *Pharmacotherapy* **2000**, *20*, 1.
11. Han, S. Y.; Sweeney, J. E.; Bachman, E. S.; Schweiger, E. J.; Forloni, G.; Coyle, J. T.; Davis, B. M.; Joullié, M. M. *Eur. J. Med. Chem.* **1992**, *27*, 673.
12. Sramek, J. J.; Frackiewicz, E. J.; Cutler, N. R. *Drugs* **2000**, *10*, 2393.
13. Relman, A. S. *N. Engl. J. Med.* **1990**, *324*, 349.
14. Brufani, M.; Filocamo, L.; Lappa, S.; Maggi, A. *Drugs Future* **1997**, *22*, 397.
15. Siddiqui, M. F.; Levey, A. I. *Drugs Future* **1999**, *24*, 417.
16. Rampa, A.; Bisi, A.; Belluti, F.; Gobbi, S.; Valenti, P.; Andrisano, V.; Cavrini, A.; Cavalli, A.; Recanatini, M. *Bioorg. Med. Chem.* **2000**, *8*, 497.
17. Tabarrini, O.; Cecchetti, V.; Temperini, A.; Filipponi, E.; Lampertib, M. G.; Fravolinia, A. *Bioorg. Med. Chem.* **2001**, *9*, 2921.
18. Marco, J. L.; Rios, C.; Carreiras, M. C.; Banos, J. E.; Badia, A.; Vivas, N. M. *Bioorg. Med. Chem.* **2001**, *9*, 727.
19. Marco, J. L.; Rios, C.; Garcia, A. G.; Villarroya, M.; Carreiras, M. C.; Martins, C.; Eleuterio, A.; Morreale, A.; Orozco, M.; Luque, F. J. *Bioorg. Med. Chem.* **2004**, *12*, 2199.
20. Leon, R.; Marco-Contelles, J.; Garcia, A. G.; Villarroya, M. *Bioorg. Med. Chem.* **2005**, *13*, 1167.
21. Shao, D.; Zou, C.; Luo, C.; Tang, X.; Li, Y. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4639.
22. Hu, M. K.; Wu, L. J.; Hsiao, G.; Yen, M. H. *J. Med. Chem.* **2002**, *45*, 2277.
23. Camps, P.; Formosa, X.; Munoz-Torrero, D.; Petrignet, J.; Badia, A.; Clos, M. V. *J. Med. Chem.* **2005**, *48*, 1701.
24. Jung, M.; Kim, H. J.; Nam, K. Y.; No, K. T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2994.
25. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
26. TITAN Pro, Wavefuntion, Inc., 1999.
27. PreADME is a web-based application for predicting ADME data. http://www.bmdrc.org/ (accessed 02 2006).
28. BioMedCAChe 6.10, Fujitsu Ltd, 2004.
29. SPSS for Windows, version 12.0, standard version, SPSS, Inc., 2003.
30. Robert, P.; Escoufier, Y. *Appl. Stat.* **1976**, *25*, 257.

31. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1958**, *21*, 1087.
32. Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671.
33. Gualdron, O.; Llobet, E.; Brezmes, J.; Vilanova, X.; Correig, X. *Sens. Actuators, B* **2006**, *114*, 522.
34. Nolle, L.; Armstrong, D. A.; Hopgood, A. A.; Ware, J. A. *Int. J. Knowl. Based Intell. Eng. Syst.* **2002**, *6*, 104.
35. Liao, G. C.; Tsao, T. P. *Electr. Power Syst. Res.* **2004**, *70*, 237.
36. Holland, H. *Adaption in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, 1975.
37. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
38. Leardi, R.; Boggia, R.; Terrile, M. *J. Chemom.* **1992**, *6*, 267.
39. Dunn, W. J., III *Chemom. Intell. Lab. Syst.* **1989**, *6*, 181.
40. Leardi, R. *J. Chemom.* **1994**, *8*, 65.
41. R Project for Statistical Computing. http://www.r-project.org/ (accessed 03 2006).
42. Hall, L. H.; Kier, L. B. *Eur. J. Med. Chem.* **1977**, *4*, 307.
43. Kier, L. B. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109.
44. Kier, L. B. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science Publishers: New York, 1990; p 152.
45. Shen, M.; LeTiran, A.; Xiao, Y.; Golbraikh, A.; Kohn, H.; Tropsha, A. *J. Med. Chem.* **2002**, *45*, 2811.
46. Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323.
47. Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306.