EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

Original article:

QSAR STUDY OF ANTI-PRION ACTIVITY OF 2-AMINOTHIAZOLES

Prasit Mandi[1,2], Chanin Nantasenamat[1,2]*, Kakanand Srungboonmee[1], Chartchalerm Isarankura-Na-Ayudhya[2], Virapong Prachayasittikul[2]

[1] Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[2] Department of Clinical Microbiology and Applied Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
* Corresponding author: E-mail: mtcnt@mahidol.ac.th, Tel: +662 441 4371, Fax: +662 441 4380

ABSTRACT

2-aminothiazoles is a class of compounds capable of treating life-threatening prion diseases. QSAR studies on a set of forty-seven 2-aminothiazole derivatives possessing anti-prion activity were performed using multivariate analysis, which comprised of multiple linear regression (MLR), artificial neural network (ANN) and support vector machine (SVM). The results indicated that MLR afforded reasonable performance with a correlation coefficient (*r*) and root mean squared error (RMSE) of 0.9073 and 0.2977, respectively, as obtained from leave-one-out cross-validation (LOO-CV). More sophisticated learning methods such as SVM provided models with the highest accuracy with *r* and RMSE of 0.9471 and 0.2264, respectively, while ANN gave reasonable performance with *r* and RMSE of 0.9023 and 0.3043, respectively, as obtained LOO-CV calculations. Descriptor analysis from the regression coefficients of the MLR model suggested that compounds should be asymmetrical molecule with low propensity to form hydrogen bonds and high frequency of N content at topological distance 02 in order to provide good activities. Insights from QSAR studies is anticipated to be useful in the design of novel derivatives based on the 2-aminothiazole scaffold as potent therapeutic agents against prion diseases.

Keywords: QSAR, 2-aminothiazole, anti-prion, multiple linear regression, artificial neural network, support vector machine

INTRODUCTION

Prion diseases or transmissible spongiform encephalopathy are progressive fatal brain disorders affecting human and animal. This class of disease includes Creutzfeldt-Jakob Disease (CJD), Gerstmann-Sträussler-Scheinker syndrome (GSS), Familial Fatal Insomnia (FFI) and Kuru in human; Scrapie in sheep and goat; and Bovine Spongiform Encephalopathy (BSE) in cattle (Miller, 2009; Prusiner, 1998). Moreover, prion diseases are transmissible and can be infected among individuals of the same or different species. Treatment for this class of diseases is not yet available (Weissmann et al., 2002).

Although the precise underlying molecular mechanisms of this disease remains unclear, all prion diseases are associated with the conversion of normal cellular prion protein (PrP^{C}) into the pathogenic isoform (PrP^{Sc}) (Race et al., 1987). The accumulation of PrP^{Sc}, which is rich in β-sheets and is protease-resistant, leads to the precipitation of fibrillar structures and eventually culminating in brain damage, which are the hallmark of cases with clinical prion diseases (Butler et al., 1988).

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

A diverse array of compounds, both small and large molecules, has been extensively evaluated for their anti-prion properties. Small molecules that have been tested include acridine analogues (Dollinger et al., 2006; Korth et al., 2001; May et al., 2003, 2006), statins (Kempster et al., 2007), 2,4-diphenylthiazole and 2,4-diphenyloxazole amides (Heal et al., 2007), pyrazolones (Kimata et al., 2007), indole-3-glyoxamides (Thompson et al., 2009), pyridyl hydrazones (Kawasaki et al., 2007) and 2-aminothiazole analogues (Gallardo-Godoy et al., 2011). In addition, larger molecules such as polyanionic chemotype (i.e. suramin, pentosan polysulfate) or polycationic chemotype (i.e. dendritic polyamines and cationic polysaccharides) (Yudovin-Farber et al., 2005) have also been reported to exhibit anti-prion activity in cells. Among all molecules, 2-aminothiazole is considered a potent class of compound that can reduce satisfactory amount of PrP^{Sc} in neuroblastoma cells (ScN2a-cl3) and can be orally absorbed when formulated appropriately in liquid rodent diet to achieve steady-state brain concentration (Gallardo-Godoy et al., 2011). This finding shows that this class of compound is promising therapeutic agents for treating prion diseases.

The concept of quantitative structure-activity relationship (QSAR) has been widely used in drug discovery and design (Isarankura-Na-Ayudhya et al., 2008; Nantasenamat et al., 2006, 2008b, 2009, 2010, 2012; Piacham et al., 2006, 2009; Prachayasittikul et al., 2007, 2010, 2011; Suksrichavalit et al., 2008, 2009; Thippakorn et al., 2009; Worachartcheewan et al., 2009, 2011a, b). Computational formulation of predictive QSAR models is an important step in the drug design process as it helps avoiding unnecessary experimentation by seeking out relationships that exist between a compound's structure and its respective activities as well as elucidating the physicochemical characteristics that constitute the biological activity of interest. Therefore, the aim of this study is to investigate the molecular properties of 2-aminothiazole derivatives that are responsible for the anti-prion activity of aminothiazole derivatives via the construction of predictive QSAR models for predicting the EC_{50} values using traditional learning methods and machine learning techniques.

MATERIAL AND METHODS

Data set

A data set of forty-seven 2-aminothiazole derivatives and their anti-prion activity were obtained from the work of Gallardo-Godoy et al. (2011). The chemical structures of the compounds are shown in Figure 1. The anti-prion activity was represented by the EC_{50} value, indicating 50 % maximal effective concentration of PRP^{Sc} in prion-infected neuroblastoma cells (ScN2a-cl3). EC_{50} values were subjected to data transformation by taking the negative logarithm to the base of 10 in order to obtain a more uniformly distributed data. This was performed according to the following equation:

$$pEC_{50} = -\log(EC_{50}) \qquad (1)$$

where pEC_{50} is the negative logarithm of EC_{50} values.

Geometry optimization and descriptors calculation

Chemical structures of the compounds were drawn using ChemAxon Marvin (ChemAxon Ltd., 2011) and the molecular geometries were optimized at the density functional theory (DFT) level using Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP) in combination with the 6-31G(d) basis set.

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

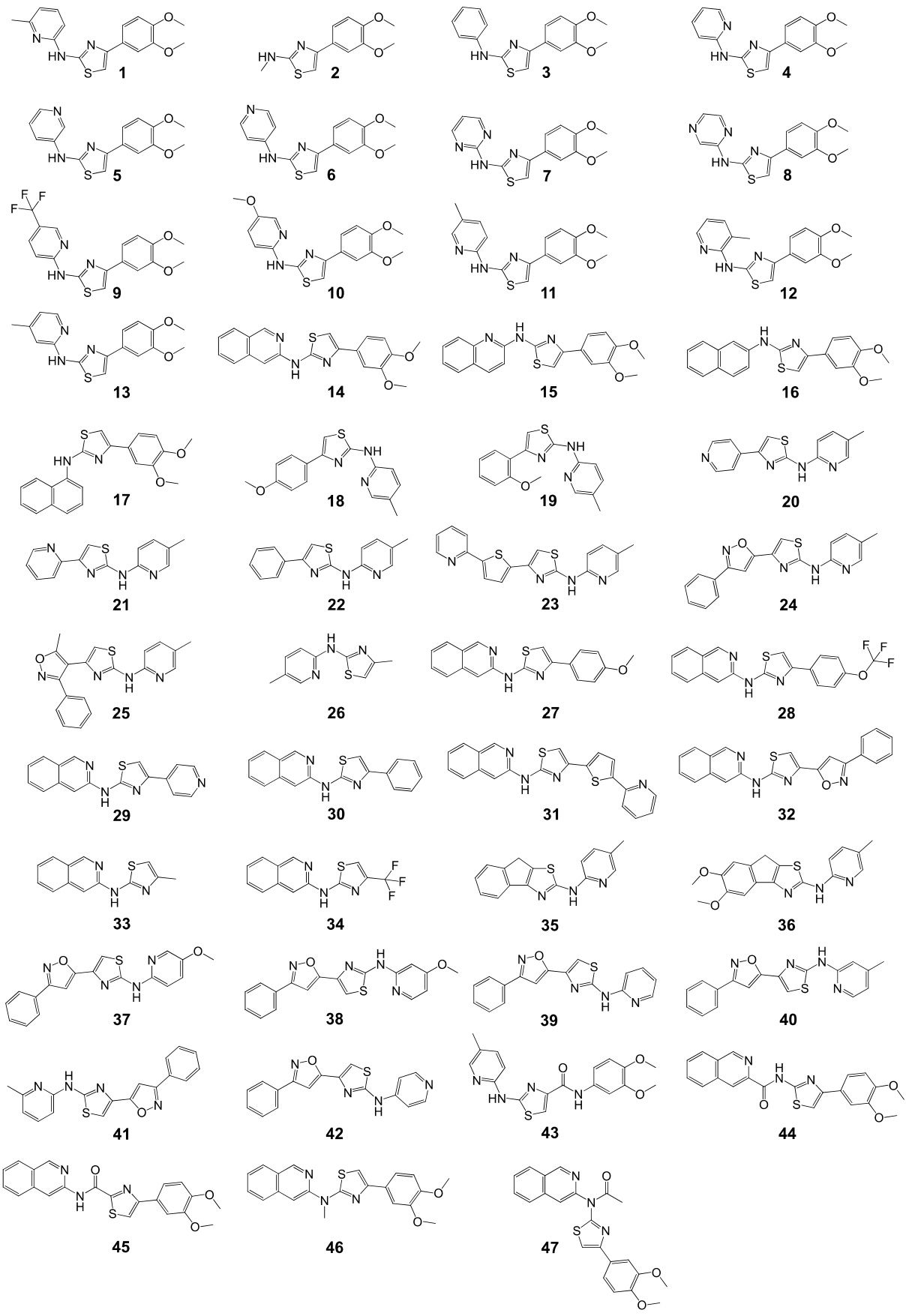

Figure 1: Chemical structure of 2-aminothiazole derivatives

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

Such quantum chemical calculation was calculated using Gaussian 09 (Frisch et al., 2009). Subsequently, a set of quantum chemical descriptors was obtained from the low energy conformer of the structures as generated by Gaussian 09. The quantum chemical descriptors used in this study included the total energy of the molecule (E_{total}), highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), dipole moment (μ) of the molecule, electron affinity (EA), ionization potential (IP), energy difference of HOMO and LUMO ($HOMO\text{-}LUMO_{gap}$), Mulliken electronegativity (χ), Hardness (η), Softness (S), Electrophilicity (ω), Electrophilic index (ω_i), most negative atom in the molecule (Q_{neg}), most positive atom in the molecule (Q_{pos}) and the mean absolute atomic charge (Q_m) (Karelson et al., 1996; Parr et al., 1978, 1999; Parr and Pearson, 1983; Thanikaivelan et al., 2000).

The low energy conformers were then submitted for further generation of an additional set of 3,224 molecular descriptors using DRAGON version 5.5 (Talete srl, 2007). This descriptor set spanned 22 categories and comprises of 48 constitutional descriptors, 119 topological descriptors, 47 walk and path counts, 33 connectivity indices, 47 information indices, 96 2D autocorrelation, 107 edge adjacency indices, 64 Burden eigenvalues, 21 topological charge indices, 44 eigenvalue-based indices, 41 randic molecular profiles, 74 geometrical descriptors, 150 RDF descriptors, 160 3D-MoRSE descriptors, 99 WHIM descriptors, 197 GETAWAY descriptors, 154 functional group counts, 120 atom-centered fragments, 14 charge descriptors, 29 molecular properties, 780 2D binary fingerprints and 780 2D frequency fingerprints.

Feature selection

Descriptors having constant value and pairs of variables with correlation coefficient greater than 0.9 were removed using the Unsupervised Forward Selection algorithm (UFS, version 1.8) (Whitley et al., 2000) as described previously (Nantasenamat et al., 2005). The descriptors were further removed using stepwise multiple linear regression using SPSS Statistics 18.0 (IBM Corporation, 2011) in order to obtain important descriptors that will subsequently be used in correlating with anti-prion activities of 2-aminothiazole derivatives.

Multivariate analysis

The independent variables (e.g. quantum chemical and molecular descriptors) and the dependent variable (e.g. pEC_{50}) (Table 1) were subjected to multivariate analysis using traditional learning methods, namely multiple linear regression (MLR) and machine learning techniques, particularly support vector machine (SVM) and artificial neural network (ANN). The learning approaches were selected to establish a relationship between the structures of 2-aminothiazole and their anti-prion activity.

The MLR models were calculated to obtain the following equation:

$$Y = B_0 + \sum B_n X_n \qquad (2)$$

where Y is the pEC_{50} of the 2-aminothiazole derivatives, B_0 is the intercept or the base value of pEC_{50} and B_n are the regression coefficients of descriptors X_n. MLR was calculated using the Waikato Environment for Knowledge Analysis (Weka), version 3.4.5 (Witten et al., 2011).

SVM is a machine learning method developed by Vapnik and coworkers that is based on the statistical learning theory (Cortes and Vapnik, 1995; Vapnik, 1998), which is a powerful technique for classification (Du et al., 2008; Liu et al., 2004) and regression (Dong et al., 2005; Li et al., 2007; Nantasenamat et al., 2008a; Niu et al., 2007; Zhao et al., 2006). An in-depth treatment of the concepts and theory of SVM has previously been reviewed (Cristianini and Shawe-Taylor, 2000; Sánchez, 2003). Briefly, SVM was origin-

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

Table 1: Molecular descriptors and pEC50 value of 2-amonothaiazole derivatives

Compound	ISH	MATS8e	G2p	F02[N-N]	nHDon	EC_{50} (μM)	pEC_{50}
1	0.897	0.151	0.171	2	1	3.010	-0.479
2	0.874	0.142	0.186	1	1	32.000	-1.505
3	0.870	0.241	0.174	1	1	8.200	-0.914
4	0.878	0.142	0.175	2	1	1.220	-0.086
5	0.870	0.142	0.202	1	1	28.000	-1.447
6	0.850	0.198	0.175	1	1	6.380	-0.805
7	0.918	0.032	0.191	4	1	3.940	-0.596
8	0.842	0.032	0.177	2	1	15.600	-1.193
9	0.912	0.212	0.165	2	1	2.530	-0.403
10	0.873	-0.06	0.194	2	1	1.000	0
11	0.858	0.158	0.158	2	1	0.790	0.102
12	0.851	0.158	0.158	2	1	7.290	-0.863
13	0.844	0.151	0.171	2	1	1.000	0
14	0.807	0.124	0.156	2	1	0.110	0.959
15	0.834	0.124	0.168	2	1	0.430	0.367
16	0.847	0.187	0.161	1	1	0.390	0.409
17	0.863	0.187	0.161	1	1	32.000	-1.505
18	0.827	0.383	0.177	2	1	3.030	-0.481
19	0.905	0.383	0.191	2	1	29.000	-1.462
20	0.856	-0.02	0.186	2	1	1.570	-0.196
21	0.857	-0.138	0.186	2	1	0.860	0.066
22	0.833	0.022	0.184	2	1	7.880	-0.897
23	0.804	0.077	0.174	2	1	1.230	-0.089
24	0.909	-0.144	0.174	2	1	0.940	0.027
25	0.914	0.212	0.170	2	1	26.000	-1.415
26	0.962	0.365	0.177	2	1	32.000	-1.505
27	0.754	0.292	0.159	2	1	0.340	0.469
28	0.860	0.151	0.173	2	1	32.000	-1.505
29	0.842	-0.081	0.164	2	1	0.310	0.509
30	0.939	-0.06	0.171	2	1	3.080	-0.489
31	0.907	0.036	0.170	2	1	2.010	-0.303
32	0.821	-0.128	0.187	2	1	0.800	0.097
33	0.942	0.027	0.193	2	1	32.000	-1.505
34	0.954	-0.429	0.193	2	1	8.660	-0.938
35	0.954	0.027	0.165	2	1	2.440	-0.387
36	0.873	0.182	0.170	2	1	5.590	-0.747
37	0.866	-0.289	0.159	2	1	0.230	0.638
38	0.856	-0.168	0.173	2	1	0.250	0.602
39	0.885	-0.189	0.163	2	1	0.600	0.222
40	0.859	-0.166	0.222	2	1	1.270	-0.104
41	0.884	-0.166	0.160	2	1	0.820	0.086
42	0.877	-0.328	0.163	1	1	4.420	-0.645
43	0.873	-0.203	0.161	2	2	2.440	-0.387
44	0.840	-0.033	0.176	1	1	8.180	-0.913
45	0.890	0.218	0.186	1	1	10.000	-1
46	0.827	0.064	0.164	2	0	0.140	0.854
47	0.853	-0.054	0.152	2	0	0.081	1.092

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

ally developed as a linear classifier and was therefore limited to classification problems. However, with the introduction of the ε-insensitive loss function the technique is also capable of performing non-linear regression problems

$$L_\varepsilon(y, f(x,w)) = \begin{cases} |y - f(x,w)| - \varepsilon \\ 0 \end{cases} \quad \begin{matrix} \text{for } |y - f(x,w)| \geq \varepsilon \\ \text{otherwise} \end{matrix} \tag{3}$$

where ε denotes the tube size and it represents the approximation accuracy of the training data samples. Support vector regression seeks to find an *f(x)* function for which there is at most ε deviation from the experimental value of y_i for all training samples while trying to be as flat as possible. The loss function gives little consideration to error as long as it is less than ε and their is minimal deviation from it. As SVM is essentially a linear classifier, it must project the input variables onto a higher dimensional feature space by means of kernel transformation as described by the following equation:

$$K(x,y) = \langle \phi(x) \cdot \phi(y) \rangle \tag{4}$$

where K represents a kernel function and ϕ represents a mapping function from the input space onto the feature space. A series of hyperplane is then added to the newly generated higher dimensional feature space where the maximal-margin hyperplane that maximizes the distance between support vector hyperplanes is identified and used in reaching a solution. Popular kernel functions are comprised of linear, polynomial and radial basis function (RBF). A commonly used kernel function is RBF, which is used in this study is described below:

$$K(x,y) = e^{-(\gamma \langle x-y, x-y \rangle)^2} \tag{5}$$

In order to obtain good predictive performance for the SVM models, an empirical search of the SVM parameters is needed as there are no universal sets of parameters that perform well for all types of problems. The two parameters of the RBF kernel included the complexity parameter (C) and the gamma (γ) parameter, which were optimized in order to obtain the ideal configuration for the SVM model. Parameter optimization was performed using a two-level grid search that is comprised of an initial coarse grid search where the values of C and γ were adjusted using an exponential increase in the value. Subsequently, a local grid search of the optimal regions discovered in the coarse grid search was selected for further refinement of the model using a much smaller increment of steps.

Artificial neural network (ANN) implementing the back-propagation of error algorithm is an interconnected feed-forward network of neuronal nodes that essentially mimicks the inner workings of the brain. The principles of ANN have been described previously (Nantasenamat et al., 2005, 2007a, b). Briefly, a typical ANN architecture is a network comprising of three interconnected layers: input layer, hidden layer and output layer (Zupan and Gasteiger, 1999). Information from the molecular descriptors is first sent to the input layer where they are subsequently relayed onto nodes of the hidden layer for further processing and finally sent to the output layer. The interconnections of nodes of the various layers are assigned a randomized weight value. Therefore, to achieve reasonable stabilization of the resulting values the calculations were performed for 10 times and their average values were used. Similar to SVM, the parameters in ANN were also optimized using an empirical trial-and-error search. The ANN parameters that were investigated are comprised of the number of nodes in the hidden layer, the learning epoch size, the learning rate and the momentum. ANN calculations implementing the back-propagation of error algorithm were performed using Weka, version 3.4.5.

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

Data sampling

Leave-one-out cross-validation (LOO-CV) was used in separating the data set into a training set and testing set. LOO-CV is a practical and reliable approach suited for small data sets as it allows the best economical usage of the available data. Briefly, the concepts of LOO-CV involve the leaving out of one data sample as the testing set while employing the remaining N-1 samples as the training set. In this manner, each of the samples of the data set had a chance to be used as the testing set.

Outlier identification

Compounds having a standardized residual value exceeding ± 2 were identified as an outlier and were subsequently removed from the data set. The standardized residual values were calculated according to the following equation:

$$x_{IJ}^{stn} = \frac{x_{IJ} - \bar{x}}{\sum_{I=1}^{N}(x_{IJ} - \bar{x}_{IJ})^2 / N} \qquad (6)$$

where x_{IJ}^{stn} is the standardized residual, x_{IJ} is the value of interest, $\bar{x}$ and $\bar{x}_J$ are the mean value and N is the sample size.

Model evaluation

To assess the performance of the developed QSAR models, correlation coefficient (r) was employed as it describes the degree of correlation between the predicted and experimental values. Furthermore, root mean squared error (RMSE) was used as a measure of the predictive error of the model. RMSE was calculated according to the following equation:

$$RMSE = \sqrt{\frac{\sum_{I=1}^{N}(x_{PRED} - x_{EXP})^2}{N}} \qquad (7)$$

where x_{PRED} represents the predicted pEC_{50}, x_{EXP} represents the actual pEC_{50}, and N is the number of compounds in the data set.

RESULTS AND DISCUSSION

The aim of this study was to develop QSAR models for elucidating the inherent structure-activity relationship of 2-aminothiazoles by correlating the calculated physicochemical descriptors with their respective anti-prion activities. Therefore, in order to formulate accurate predictive QSAR models a wide array of molecular descriptors was obtained from quantum chemical and Dragon calculations. The descriptors were obtained from geometrically optimized structures calculated at the B3LYP/6-31G(d) level of theory. The molecular descriptor and pEC_{50} values for each compound are shown in Table 1.

Feature selection

The multi-collinear and redundant descriptors obtained from the Dragon software package were reduced from 3,224 to 59 descriptors using UFS. An additional set of 15 quantum chemical descriptors derived from Gaussian 09 were combined with a set of 59 Dragon descriptors and the descriptors were then subjected to stepwise multiple linear regression in order to select relevant descriptors. The resulting subset of molecular descriptors is comprised of the standardized information content on the leverage equality (ISH), Moran autocorrelation of lag 8 weighted by Sanderson electronegativity (MATS8E), 2nd component symmetry directional WHIM index/ weighted by polarizability (G2p), frequency of N - N at topological distance 02 (F02[N-N]) and the number of hydrogen bond donor (nHDon).

Prediction of anti-prion activity using MLR

The five selected molecular descriptors were then used in the construction of QSAR models using MLR. The initial MLR model provided moderate predictive performance yielding correlation coefficients of 0.7877 and 0.7227, respectively, for the training set (r_{tr}) and testing set (r_{cv}). The root mean squared error was 0.4470 and 0.5049, respectively, for the training set ($RMSE_{tr}$) and

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

testing set ($RMSE_{cv}$). In efforts to develop robust QSAR models that are free from outlying compounds, a standardized residual cutoff criteria of ± 2 was employed. This led to the identification of compound number 8, 16 and 17 as outliers. This was reiteratively performed until no outliers could be identified from the data set. The models and their predictive performance after outlier removal are shown in Table 2.

The best MLR model was observed to be model 5 which provided high correlation coefficients (r_{tr} = 0.9356 and r_{cv} = 0.9073) and small error ($RMSE_{tr}$ = 0.2487 and $RMSE_{cv}$ = 0.2977) after removing a total of 9 outliers. The equation for QSAR model 5 is shown below:

$$pEC_{50} = -7.756(ISH) - 1.1942(MATS8e) - 20.2741(G2p) + 0.7055(F02[N\text{-}N]) - 0.6117(nHDon) + 9.3543 \quad (8)$$

The 2-aminothiazole derivatives that were identified as outliers are comprised of compounds 7, 8, 12, 16, 17, 22, 25, 28 and 40. Investigations of the molecular structures were carried out in efforts to gain an in-depth understanding of reasons behind the outlying compounds. The structures of compounds 7 and 8 were flagged as outliers because they have more than one nitrogen atom on the substituent ring at the amine arm. This suggested that compounds with high anti-prion activity should not have more than one nitrogen atom on substituent sites. Compound 12 is different from other structurally related compounds because it is the only compound with a methyl group at the ortho position of the substituent ring. Compounds 16 and 17 had naphthalene as the substituent group and such polyaromatic ring did not bear any nitrogen atoms therefore making it hydrophobic and therefore differentiating them from other compounds.

The presence of a nonpolar benzene ring at the thiazole arm of compound 22 differentiated it from the rest of the compounds. Compound 25 was the only compound to have a methyl group emanating from the oxazole substituent. Similarly, compound 28 was the only compound to have the highly electronegative trifluoromethoxy group. Although the molecular structure of compound 40 did not significantly differentiate it from other compounds; interesting it was observed that the compound had the highest G2p value of 0.222 as compared to the mean value of 0.17417 and a standard deviation of 0.0137.

The relative importance of the molecular descriptors can be discerned from the regression coefficient values from the QSAR question of model 5. The molecular descriptors listed in order of decreasing importance is as follows: G2p > ISH > MATS8e > F02[N-N] > nHDon. G2p is a WHIM index that describes the second component symmetry directional weight by atomic polarizabilities. Taken together it was observed that potent molecules tend to have lower symmetry of atomic polarizabilities at second component to have high potency. Particularly, ISH is a descriptor that provides information on the molecular symmetry in terms of shape and the results suggested that potent compounds had asymmetrical structures. MATS8e is a Mo-

Table 2: Summary of predictive performance of MLR model

Model	r_{tr}	r_{cv}	RMS_{tr}	RMS_{cv}	Outliers
1	0.7877	0.7277	0.447	0.5049	8, 16, 17
2	0.8366	0.7729	0.389	0.4548	28
3	0.8621	0.7957	0.3529	0.4269	7, 12, 40
4	0.9091	0.8772	0.2979	0.3449	22, 25
5	0.9356	0.9073	0.2487	0.2977	No Outliers

r_{Tr}: correlation coefficient of training set
r_{CV}: correlation coefficient of leave-one-out cross-validation testing set
RMS_{tr}: root mean squared error of training set
RMS_{cv}: root mean squared error of testing set

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

ran autocorrelation that essentially describes the autocorrelation of the atomic electronegativities. Lowering the correlation of electronegativities between atoms at different topological distance of lag 8 provided compounds with higher potency. F02[N-N] is a 2D frequency fingerprint describing the counting distance between N and N atoms at topological distance 02. Therefore, high activity compounds should have high frequency of nitrogen atoms at topological distance 02. nHdon is a descriptor describing the number of hydrogen bond donor present in the molecule. As such, compounds with higher number of hydrogen donor are unlikely to exhibit good anti-prion activity. Looking at this holistically, it can be summarized that the physicochemical properties necessary for anti-prion activity should be molecules that are asymmetrical with low propensity to form hydrogen bonds and possessing high frequency of nitrogen content at topological distance 02.

In order to discern the physicochemical properties governing the bioactivity of 2-aminothiazoles, the compounds were categorized as active, inactive and intermediate compounds using the following as cut-off criteria: <1 µM, >10 µM and 1-10 µM, respectively. It was observed from Table 3 that active compounds tend to have lower values for G2p, ISH, MATS8e and nHDon while having higher values of F02[N-N]. Of particular note is compound 47, which had the best anti-prion activity with an EC_{50} value of 0.081 µM while also having the least G2p value of 0.152. This observation confirmed the importance of G2p that is low symmetry of atomic polarizabilities at the second component.

Prediction of anti-prion activity using SVM and ANN

Additionally, QSAR model was constructed with SVM employing the same data set as MLR model 5. In order to obtain an optimal set of SVM parameters for pEC_{50} prediction, the C and γ of parameters were optimized using a two-level grid search. Initially, a coarse grid search (Figure 2a) was performed by exponentially adjusting the values of C and γ such that the exponent *n* value in 2^n was varied from -19 to 19 using an incremental step of 2, which correspondingly sampled the regions of C and γ from 2^{-19} to 2^{19} using an incremental step of 2^2. This is then followed by a local grid search (Figure 2b) of the optimal regions identified in the coarse grid search. The optimal values for C and γ were $2^{4.75}$ (26.9087) and $2^{-3.25}$ (0.1051) respectively. The SVM model demonstrated good predictive performance as observed from correlation coefficients of r_{tr} = 0.9578 and r_{cv} = 0.9471 as well as root mean squared errors of $RMSE_{tr}$ = 0.2022 and $RMSE_{cv}$ = 0.2264.

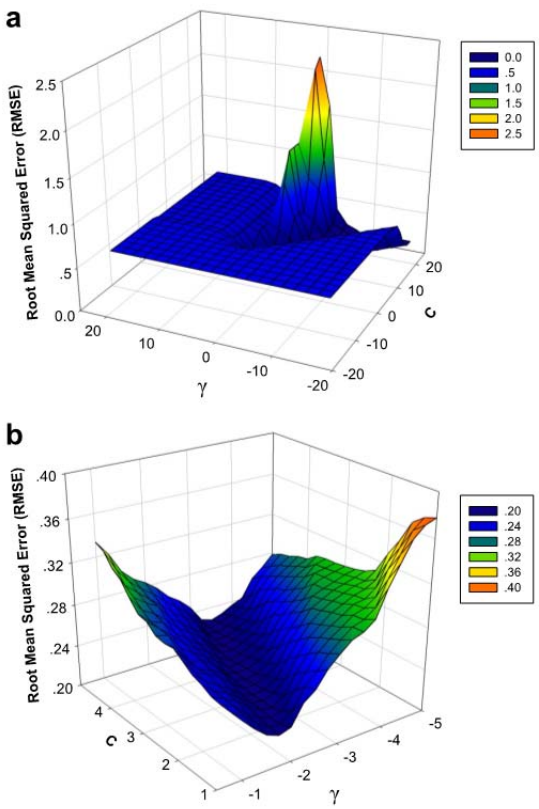

Figure 2: Optimization of SVM parameters comprising of an initial coarse grid search (a) and local grid search (b)

EXCLI Journal 2012;11:453-467 – ISSN 1611-2156
Received: July 19, 2012, accepted: July 27, 2012, published: August 15, 2012

Table 3: Summary of the predictive performance of MLR, SVM and ANN models for predicting the anti-prion activity of 2-aminothiazole derivatives after outlier removal

Method	r_{tr}	r_{cv}	RMS_{tr}	RMS_{cv}
MLR	0.9356	0.9073	0.2487	0.2977
SVM	0.9578	0.9471	0.2022	0.2264
ANN	0.9399	0.9023	0.2544	0.3043

r_{tr}: correlation coefficient of training set
r_{cv}: correlation coefficient of leave-one-ut cross-validation testing set
RMS_{tr}: root mean squared error of training set
RMS_{cv}: root mean squared error of testing set

An optimal set of ANN parameters was achieved by means of an empirical trial-and-error search for the number of hidden node, the number of learning epoch, learning rate and momentum as shown in Figure 3. The results indicated that the optimal number of hidden nodes and learning epoch to use is 2 and 100, respectively, while the optimal configuration for the learning rate and momentum is 0.1 and 0.0, respectively. The predicted outputs were derived from the average of 10 runs of network training. It was observed that the ANN models could provide good predictive models for predicting the anti-prion activity as observed from the correlation coefficients of r_{tr} = 0.9399 and r_{cv} = 0.9023 as well as root mean squared errors of $RMSE_{tr}$ = 0.2544 and $RMSE_{cv}$ = 0.3043.

Of the tested learning methods, SVM was demonstrated to be the best learning approach for predicting the anti-prion activity while ANN and MLR gave similar level of performance. Summary of the predictive performance of the QSAR models for predicting anti-prion activity of 2-aminothiazole is shown in Table 3 while scatter plots of the predicted versus experimental values for the training and cross-validation testing sets of QSAR models are shown in Figure 4.

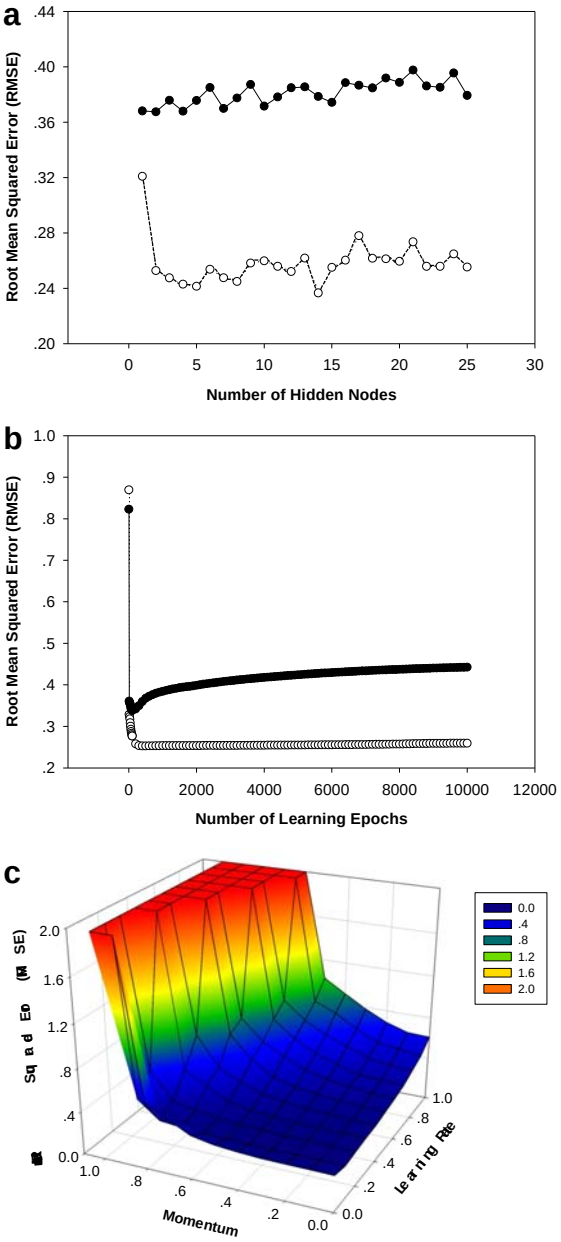

Figure 3: Optimization of ANN parameters comprising of the number of hidden nodes (a), the number of learning epochs (b) and the learning rate and momentum

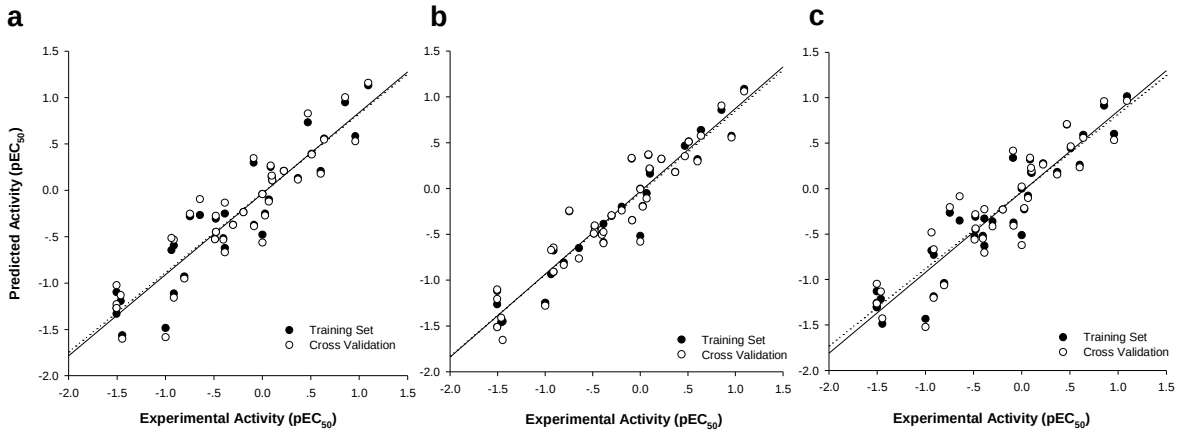

Figure 4: Plot of the experimental versus predicted values of pEC_{50} as obtained from MLR (a), SVM (b), and ANN (c) calculations

CONCLUSION

In this study, traditional and machine learning approaches were employed in the development of QSAR models of 2-aminothiazole derivatives in order to establish a structure-activity relationship between the molecular descriptors of the compounds with their respective anti-prion activity. The MLR model indicated that high potency compound should be an asymmetrical molecule having low propensity to form hydrogen bonds and possess high frequency of nitrogen content at topological distance 02. Furthermore, the predicted anti-prion activities of 2-aminothiazoles using SVM were found to be in good agreement with their experimental values. Overall, the results suggested the robustness of the developed QSAR models in predicting the anti-prion activity as well as providing key insights on the physiochemical properties that is inherently found in the investigated 2-aminothiazole derivatives. Such knowledge serves as a general guideline for future structural modifications of 2-aminothiazoles as therapeutic agents against prion with potentially higher potency and less toxicity.

ACKNOWLEDGEMENTS

C.N. acknowledges support from the Goal-Oriented Research Grant of Mahidol University. P.M. is a post-graduate student grateful for a research assistantship from Mahidol University. Partial support is also acknowledged from Office of the Higher Education Commission and Mahidol University under the National Research Universities Initiative.

REFERENCES

Butler DA, Scott MR, Bockman JM, Borchelt DR, Taraboulos A, Hsiao KK et al. Scrapie-infected murine neuroblastoma cells produce protease-resistant prion proteins. J Virol 1988;62:1558-64.

ChemAxon Ltd. MarvinSketch, version 5.7. Budapest, Hungary, 2011.

Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273-97.

Cristianini N, Shawe-Taylor J. An introduction to support vector machines. Cambridge: Cambridge Univ. Press, 2000.

Dollinger S, Lober S, Klingenstein R, Korth C, Gmeiner P. A chimeric ligand approach leading to potent antiprion active acridine derivatives: design, synthesis, and biological investigations. J Med Chem 2006;49: 6591-5.

Dong N, Lu WC, Chen NY, Zhu YC, Chen KX. Using support vector classification for SAR of fentanyl derivatives. Acta Pharmacol Sin 2005;26:107-12.

Du H, Wang J, Watzl J, Zhang X, Hu Z. Classification structure-activity relationship (CSAR) studies for prediction of genotoxicity of thiophene derivatives. Toxicol Lett 2008;177:10-9.

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR et al. Gaussian 09, Revision A.1. Wallingford, Connecticut, 2009.

Gallardo-Godoy A, Gever J, Fife KL, Silber BM, Prusiner SB, Renslo AR. 2-Aminothiazoles as therapeutic leads for prion diseases. J Med Chem 2011;54:1010-21.

Heal W, Thompson MJ, Mutter R, Cope H, Louth JC, Chen B. Library synthesis and screening: 2,4-diphenylthiazoles and 2,4-diphenyloxazoles as potential novel prion disease therapeutics. J Med Chem 2007;50: 1347-53.

IBM Corporation. SPSS Statistics, version 18. New York: Somers, 2011.

Isarankura-Na-Ayudhya C, Nantasenamat C, Buraparuangsang P, Piacham T, Ye L, Bülow L et al. Computational insights on sulfonamide imprinted polymers. Molecules 2008;13:3077-91.

Karelson M, Lobanov VS, Katritzky AR. Quantum-chemical descriptors in QSAR/ QSPR studies. Chem Rev 1996;96:1027-44.

Kawasaki Y, Kawagoe K, Chen CJ, Teruya K, Sakasegawa Y, Doh-ura K. Orally administered amyloidophilic compound is effective in prolonging the incubation periods of animals cerebrally infected with prion diseases in a prion strain-dependent manner. J Virol 2007;81:12889-98.

Kempster S, Bate C, Williams A. Simvastatin treatment prolongs the survival of scrapie-infected mice. Neuroreport 2007; 18:479-82.

Kimata A, Nakagawa H, Ohyama R, Fukuuchi T, Ohta S, Doh-ura K et al. New series of antiprion compounds: pyrazolone derivatives have the potent activity of inhibiting protease-resistant prion protein accumulation. J Med Chem 2007;50:5053-6.

Korth C, May BC, Cohen FE, Prusiner SB. Acridine and phenothiazine derivatives as pharmacotherapeutics for prion disease. Proc Natl Acad Sci USA 2001;98:9836-41.

Li X, Luan F, Si H, Hu Z, Liu M. Prediction of retention times for a large set of pesticides or toxicants based on support vector machine and the heuristic method. Toxicol Lett 2007;175:136-44.

Liu HX, Zhang RS, Yao XJ, Liu MC, Hu ZD, Fan BT. QSAR and classification models of a novel series of COX-2 selective inhibitors: 1,5-diarylimidazoles based on support vector machines. J Comput Aided Mol Des 2004;18:389-99.

May BC, Fafarman AT, Hong SB, Rogers M, Deady LW, Prusiner SB et al. Potent inhibition of scrapie prion replication in cultured cells by bis-acridines. Proc Natl Acad Sci USA 2003;100:3416-21.

May BC, Witkop J, Sherrill J, Anderson MO, Madrid PB, Zorn JA et al. Structure-activity relationship study of 9-aminoacridine compounds in scrapie-infected neuroblastoma cells. Bioorg Med Chem Lett 2006;16:4913-6.

Miller G. Neurodegeneration. Could they all be prion diseases? Science 2009;326: 1337-9.

Nantasenamat C, Naenna T, Isarankura-Na-Ayudhya C, Prachayasittikul V. Quantitative prediction of imprinting factor of molecularly imprinted polymers by artificial neural network. J Comput Aided Mol Des 2005;19:509-24.

Nantasenamat C, Tantimongcolwat T, Naenna T, Isarankura-Na-Ayudhya C, Prachayasittikul V. Prediction of selectivity index of pentachlorophenol-imprinted polymers. EXCLI J 2006;5:150-63.

Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. Quantitative structure-imprinting factor relationship of molecularly imprinted polymers. Biosens Bioelectron 2007a;22:3309-17.

Nantasenamat C, Isarankura-Na-Ayudhya C, Tansila N, Naenna T, Prachayasittikul V. Prediction of GFP spectral properties using artificial neural network. J Comput Chem 2007b;28:1275-89.

Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine. J Mol Graph Model 2008a;27:188-96.

Nantasenamat C, Piacham T, Tantimongcolwat T, Naenna T, Isarankura-Na-Ayudhya C, Prachayasittikul V. QSAR model of the quorum-quenching N-acyl-homoserine lactone lactonase activity. J Biol Syst 2008b;16:279-93.

Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. A practical overview of quantitative structure-activity relationship. EXCLI J 2009;8:74-88.

Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. Advances in computational methods to predict the biological activity of compounds. Exp Opin Drug Discov 2010;5:633-54.

Nantasenamat C, Li H, Isarankura-Na-Ayudhya C, Prachayasittikul V. Exploring the physicochemical properties of templates from molecular imprinting literature using interactive text mining approach. Chemometr Intell Lab Syst 2012;116:128-36.

Niu B, Lu WC, Yang SS, Cai YD, Li GZ. Support vector machine for SAR/QSAR of phenethyl-amines. Acta Pharmacol Sin 2007;28:1075-86.

Parr RG, Pearson RG. Absolute hardness: companion parameter to absolute electronegativity. J Am Chem Soc 1983;105:7512-16.

Parr RG, Donnelly RA, Levy M, Palke WE. Electronegativity: The density functional viewpoint. J Chem Phys 1978;68:3801.

Parr RG, Szentpály L v, Liu S. Electrophilicity Index. J Am Chem Soc 1999;121: 1922-4.

Piacham T, Isarankura-Na-Ayudhya C, Nantasenamat C, Yainoy S, Ye L, Bülow L et al. Metalloantibiotic Mn(II)-bacitracin complex mimicking manganese superoxide dismutase. Biochem Biophys Res Commun 2006;341:925-30.

Piacham T, Nantasenamat C, Suksrichavalit T, Puttipanyalears C, Pissawong T, Maneewas S et al. Synthesis and theoretical study of molecularly imprinted nanospheres for recognition of tocopherols. Molecules 2009;14:2985-3002.

Prachayasittikul V, Isarankura-Na-Ayudhya C, Tantimongcolwat T, Nantasenamat C, Galla HJ. EDTA-induced membrane fluidization and destabilization: Biophysical studies on artificial lipid membranes. Acta Biochim Biophys Sin 2007;39:901-13.

Prachayasittikul S, Wongsawatkul O, Worachartcheewan A, Nantasenamat C, Ruchirawat S, Prachayasittikul V. Elucidating the structure-activity relationships of the vasorelaxation and antioxidation properties of thionicotinic acid derivatives. Molecules 2010;15:198-214.

Prachayasittikul S, Worachartcheewan A, Nantasenamat C, Chinworrungsee M, Sornsongkhram N, Ruchirawat S et al. Synthesis and structure-activity relationship of 2-thiopyrimidine-4-one analogs as antimicrobial and anticancer agents. Eur J Med Chem 2011;46:738-42.

Prusiner SB. Prions. Proc Natl Acad Sci USA 1998;95:13363-83.

Race RE, Fadness LH, Chesebro B. Characterization of scrapie infection in mouse neuroblastoma cells. J Gen Virol 1987;68:1391-9.

Sánchez VD. Advanced support vector machines and kernel methods. Neurocomputing 2003;55:5-20.

Suksrichavalit T, Prachayasittikul S, Piacham T, Isarankura-Na-Ayudhya C, Nantasenamat C, Prachayasittikul V. Copper complexes of nicotinic-aromatic carboxylic acids as superoxide dismutase mimetics. Molecules 2008;13:3040-56.

Suksrichavalit T, Prachayasittikul S, Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. Copper complexes of pyridine derivatives with superoxide scavenging and antimicrobial activities. Eur J Med Chem 2009;44:3259-65.

Talete srl. DRAGON for Windows (Software for Molecular Descriptor Calculations), version 5.5. Milano, Italy: Talete srl., 2007.

Thanikaivelan P, Subramanian V, Raghava Rao J, Unni Nair B. Application of quantum chemical descriptor in quantitative structure activity and structure property relationship. Chem Phys Lett 2000;323:59-70.

Thippakorn C, Suksrichavalit T, Nantasenamat C, Tantimongcolwat T, Isarankura-Na-Ayudhya C, Naenna T et al. Modeling the LPS neutralization activity of anti-endotoxins. Molecules 2009;14:1869-88.

Thompson MJ, Borsenberger V, Louth JC, Judd KE, Chen B. Design, synthesis, and structure-activity relationship of indole-3-glyoxylamide libraries possessing highly potent activity in a cell line model of prion disease. J Med Chem 2009;52:7503-11.

Vapnik V. Statistical learning theory. New York: Wiley, 1998.

Weissmann C, Enari M, Klohn PC, Rossi D, Flechsig E. Transmission of prions. Proc Natl Acad Sci USA 2002;99 (Suppl 4): 16378-83.

Whitley DC, Ford MG, Livingstone DJ. Unsupervised forward selection: a method for eliminating redundant variables. J Chem Inf Comput Sci 2000;40:1160-8.

Witten IH, Frank E, Hall MA. Data mining: Practical machine learning tools and techniques. 3. ed. Amsterdam: Morgan Kaufmann, 2011.

Worachartcheewan A, Nantasenamat C, Naenna T, Isarankura-Na-Ayudhya C, Prachayasittikul V. Modeling the activity of furin inhibitors using artificial neural network. Eur J Med Chem 2009;44:1664-73.

Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul S, Prachayasittikul V. Predicting the free radical scavenging activity of curcumin derivatives. Chemometr Intell Lab Syst 2011a;109:207-16.

Worachartcheewan A, Prachayasittikul S, Pingaew R, Nantasenamat C, Tantimongcolwat T, Ruchirawat S et al. Antioxidant, cytotoxicity, and QSAR study of 1-adamantylthio derivatives of 3-picoline and phenylpyridines. Med Chem Res 2011b:1-9. DOI 10.1007/s00044-011-9903-y.

Yudovin-Farber I, Azzam T, Metzer E, Taraboulos A, Domb AJ. Cationic polysaccharides as antiprion agents. J Med Chem 2005;48:1414-20.

Zhao CY, Zhang HX, Zhang XY, Liu MC, Hu ZD, Fan BT. Application of support vector machine (SVM) for prediction toxic activity of different data sets. Toxicology 2006;217:105-19.

Zupan J, Gasteiger J. Neural networks in chemistry and drug design. Weinheim: Wiley-VCH, 1999.