Bull. Korean Chem. Soc. **2012**, Vol. 33, No. 4 1123
http://dx.doi.org/10.5012/bkcs.2012.33.4.1123

Articles

Prediction Models of P-Glycoprotein Substrates Using Simple 2D and 3D Descriptors by a Recursive Partitioning Approach

Jong Young Joung,[†,‡] Hyoungjoon Kim,[‡] Hwan Mook Kim,[§] Soon Kil Ahn,[#] Ky-Youb Nam,[¶,*] and Kyoung Tai No[‡,a,*]

[†]*Bioinformatics and Molecular Design Research Center, Seoul 120-749, Korea*
[‡]*Department of Biotechnology and Translational Research Center for Protein Function Control, Yonsei University, Seoul 120-749, Korea.* [*]*E-mail: ktno@yonsei.ac.kr*
[§]*College of Pharmacy, Gachon University of Medicine and Science, Yeonsu, Inchon 406-799, Korea*
[#]*Division of Life Sciences, University of Incheon, Incheon 406-772, Korea*
[¶]*YOUAI Co., Ltd. Suwon 443-766, Korea.* [*]*E-mail: kyn@youai.co.kr*
Received October 31, 2011, Accepted December 21, 2011

P-gp (P-glycoprotein) is a member of the ATP binding cassette (ABC) family of transporters. It transports many kinds of anticancer drugs out of the cell. It plays a major role as a cause of multidrug resistance (MDR). MDR function may be a cause of the failure of chemotherapy in cancer and influence pharmacokinetic properties of many drugs. Hence classification of candidate drugs as substrates or nonsubstrate of the P-gp is important in drug development. Therefore to identify whether a compound is a P-gp substrate or not, *in silico* method is promising. Recursive Partitioning (RP) method was explored for prediction of P-gp substrate. A set of 261 compounds, including 146 substrates and 115 nonsubstrates of P-gp, was used to training and validation. Using molecular descriptors that we can interpret their own meaning, we have established two models for prediction of P-gp substrates. In the first model, we chose only 6 descriptors which have simple physical meaning. In the training set, the overall predictability of our model is 78.95%. In case of test set, overall predictability is 69.23%. Second model with 2D and 3D descriptors shows a little better predictability (overall predictability of training set is 79.29%, test set is 79.37%), the second model with 2D and 3D descriptors shows better discriminating power than first model with only 2D descriptors. This approach will be used to reduce the number of compounds required to be run in the P-gp efflux assay.

Key Words : ADME prediction, P-Glycoprotein, Recursive partitioning, 2D Descriptors, 3D Descriptors

Introduction

Absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties are very important in the drug discovery.[1,2] Unfavorable ADMET of new drug candidates cause over 40% of drug failures in drug development.[3] Bioavailability of drugs as well as possibility of drug-drug interaction is strongly influenced by interaction of drugs with ABC-transporters, because the ABC-multidrug transporters are constitutively expressed in many organs, such as hepatocytes, the intestine and the kidney.[4] Inhibition of the thoroughly studied transporters, ABCB1, ABCC1 and ABCG2, has been advocated as a mechanism for the restoration of drug sensitivity.[5] And it is known that cholestatic forms drug-induced liver damage result from a drug- and metabolite-mediated inhibition of hepatobiliary transporter systems, such asABCB1, ABCB4, ABCG2, ABCG5 and ABCG8.[6] Therefore, interaction with ABC transporters has the critical roles in determination of the clinical usefulness, side effects and toxicity risks of drugs. P-glycoprotein (P-gp) is a member of the ATP-binding cassette (ABC) family of transporters and known as ABCB1 in ABCB subfamily.[7] P-gp is the product of the multi drug resistance (MDR) gene and an ATP dependent efflux transporter that affects the absorption, distribution and excretion of clinically important drugs.[8] Over-expression of this protein, which may result in MDR is a major cause of the failure of cancer chemotherapy, and decrease efficacy of antibiotics.[9,10] The prediction of P-gp substrates, which facilitates early identification and elimination of drug candidates of low efficacy or high potential of MDR.[11-17] However, the accurate prediction of P-gp remains a challenge due to the complexity of the understanding physiological mechanisms and the lack of high quality data. For increasing the predictability, many complex statistical machine learning methods were carried out using supervised methods such as artificial neural network (ANN), Bayesian network and support vector machine (SVM). However, many molecular descriptors in P-gp models did not provide a better interpretable understanding

[a]Member of Translational Research Center for Protein Function Control, Korea

of P-gp substrates. Previously, we developed the machine learning method of genetic algorithm-grid search-SVM (GA-GS-SVM) for developing model of hERG toxicity prediction.[18] The selected descriptors could not easily interpretable for understating of binding between hERG ion channel and inhibitors.

In this study, Recursive partitioning (RP) method used in this work, we published that RP was widely known to be fast compared to other methods and provides easily interpretable results.[19] It is a simple statistical method but, we can understand the physical meaning of descriptors and provide interpretable information of P-gp to a medicinal chemist. For developing model of P-gp substrates, the well-known 2-dimensional (2D) descriptors were used for classification model of P-gp substrate. For increasing the accuracy, we additionally developed 3-dimensional (3D) descriptors for developing the classification model of P-gp substrates.

Methods

Data Sets. A total of 261 compounds which were classified as P-gp substrates or nonsubstrates used in this work were cautiously assembled from literature.[10,15,16,20] After removing 12 compounds of overlap data, we constructed SD file format of 261 compounds using PreADMET S/W.[21] The data set consists of 146 substrates and 115 nonsubstrates. In the 2D RP model, data set was randomly divided into two parts in Table 1. One is a training set with 209 compounds, containing 120 substrates and 89 nonsubstrates, and the other is a test set with 52 compounds, containing 26 substrates and 26 nonsubstrates. In case of 3D RP model, data set was randomly divided into two parts, a training set contained 198 compounds and a test set contained 63 compounds in Table 2.

Molecular Descriptors. We used 2D and 3D descriptors. First, we selected six descriptors which can explain their physical meaning. These descriptors are No_H_donors, No_H_acceptors, Topological_PSA, SKlogP_value, 2D_VDW_Surface, Molecular weight. We calculated value of descriptors using PreADMET. There is high coefficient of correlation (> 0.9) between molecular weight and surface area. And there is also high coefficient of correlation (> 0.85) between sum of HBD, HBA and polar surface area. Commonly, descriptors show high correlations are discarded because they may not have significant statistical meaning. But, we thought they have physically important meaning to describe feature of substrates or non-substrates compounds.

In order to increase the predictability, 3D descriptors, polar surface area and logP, were employed to build the RP model. The additional 3D descriptors are polar surface area based on Connolly surface and logP based on Solvation Free Energy Density model. To calculate values of the 3D descriptors reliably, all the substrate and non-substrate were optimized by MMFF force field.

The polar surface area (C_PSA) was calculated based on the method to generate three dimensional molecular surfaces, Connolly surface.[22] This method, which was implemented in in-house descriptor program, calculates exposed atomic surface area of each atoms in a given molecule. The total PSA is the sum of the exposed atomic surface area of polar atoms, which are nitrogen, oxygen, phosphate and sulfur atoms. The surface areas of our in-house program were calculated about 10 test molecules to be compared with surface areas from the available commercial software, Discovery Studio®.[23]

The LogP descriptor was implemented based on the Solvation Free Energy Density (SFED) Model, which calculated various solvation energy values including solvation energy in the pure water, octanol solvation energy, and partition coefficient.[24] In order to calculate logP values using SFED model, the grid points, of which distribution depends on the conformation of a molecule, are generated surround a molecule. Therefore the logP descriptor based on the SFED model reflects the three dimensional information of a given

Table 1. Performance parameters - accuracy, sensitivity, specificity, kappa, Matthews correlation coefficient- for RP model (2D) corresponding to P-gp training and test sets

Data set	Accuracy, %	Sensitivity, %	Specificity, %	Kappa[a]	Matthews correlation coefficient (MCC)[b]
Training	78.95 (165/209)	75.83 (91/120)	83.15 (74/89)	0.58	0.58
Test	69.23 (36/52)	65.38 (17/26)	73.08 (19/26)	0.38	0.39

[a]$Kappa = \frac{Accuracy - E}{1 - E}$, $E = \frac{(TP+FN)(TN+FP)+(FP+TP)(FN+TP)}{(TP+FP+FN+TN)}$. [b]$MCC = \frac{(TP \times TN)-(FP \times FN)}{\sqrt{(TP+FN)(TN+FP)(FP+TP)(FN+TP)}}$

Table 2. Performance parameters - accuracy, sensitivity, specificity, kappa, Matthews correlation coefficient- for RP model (3D) corresponding to P-gp training and test sets

Data set	Accuracy, %	Sensitivity, %	Specificity, %	Kappa	Matthews correlation coefficient (MCC)
Training	79.29 (157/198)	83.78 (93/114)	73.56 (64/82)	0.58	0.58
Test	79.37 (50/63)	79.31 (23/29)	79.41 (27/34)	0.59	0.59

molecule. In the previous work, the logP values were used to predict experimental logP values of peptide structures and involved in structure optimization of peptides.[25,26] This work confirmed that the logP model of SFED could describe relative stability of different conformers of an amino acid in the aqueous systems. From this result, availability of the logP model in SFED to explain the three dimensional state of molecules were validated and the rationale of usefulness of logP as the 3D descriptor was confirmed.

Computational Procedure. The computational procedure in this work is outlined as the following: The RP classification system for this study was trained by the *Gini Impurity* scoring function. The value of 1/100 of samples was considered as the minimum number of samples in any node. The maximum tree depth was 5 and the default values were accepted for the maximum number of generic splits and the number of knots per variable. The predictability of RP system during the training process was evaluated by means of 5-fold cross-validation. The RP training was performed using Pipeline Pilot.[27]

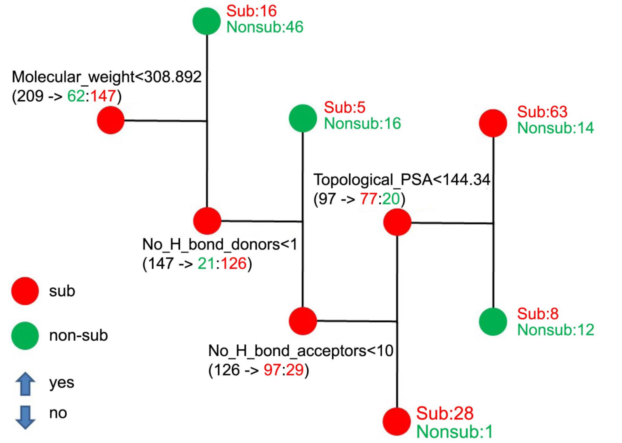

Figure 1. Decision tree and 2D descriptors built with 209 compounds of P-gp substrates and non-substrates in the training set. Two classes were used: non-substrates (Nonsub; green) and substrates (Sub; red). 2D descriptors from PreADMET were used. The classified number of substrates and non-substrates for training set is shown. The arrows show the direction of the branches.

Resultsand Discussions

In order to measure the prediction accuracy for the substrates and nonsubstrates of P-gp with this model, we introduced true positives (TP), true negatives (TN), falsepositives (FP), false negatives (FN), sensitivity SE = TP/(TP+FN), specificity SP = TN/(TN+FP). Sensitivity is the prediction accuracy for positive examples and specificity is the prediction accuracy for negative examples.[28,29] In this study, positives are substrates and negatives are nonsubstrates.

In the first RP model using 2D descriptors, finally four kinds of descriptors, which are Molecular weight, No_H_donors, No_H_acceptors, and Topological_PSA, are used to make a RP model. Table 1 summarizes the performance parameters – accuracy, sensitivity, specificity, kappa, and MCC – involving the training andtest sets. Matthews Correlation Coefficient (MCC) values for the training and test sets were 0.58 and 0.38 where they were above 0, indicating improved prediction compared to random prediction.[30,31] Kappa values for the training and test sets were 0.58 and 0.38, respectively. According to the guideline for interpreting kappa values, kappa of 0.41-0.60 is considered to be of moderate agreement,[32] and thus, we concluded that our RP model is a predictive one. Using training set, Accuracy, sensitivity and specificity of our model is 78.95%, 75.83% and 83.15% as respectively (Table 1). This model has 9 *nodes*, 5 *leaf nodes* and 4 *depths*. The model is illustrated in Figure 1. We construct five-fold cross validation model to improve the performance of our model using training set. Using test set, we validated whether our model is robust or not. If the results of training set are similar with the results of validation set, we can predict that our model will be robust and have good performance with novel compounds. We obtained slightly lower results of test set than the results of training set. In the 2D RP model, accuracy of test set is 69.23%, sensitivity is 65.38%, and specificity is 73.08%. For improving the predictability, we developed codes of two 3D descriptors that were logP and PSA of molecule. In the second RP model using 2D and 3D descriptors, the number of final descriptors which are used for making a RP model is four. Four descriptors are Molecular weight, No_H_donors, SFED_logP, and C_PSA. Table 2 was tabulated the performance parameters. The second model with 2D and 3D descriptors shows better accurate rates not only training but also test sets than the model with 2D descriptors. Accuracy of test set is 79.37%, sensitivity is 79.31%, and specificity is 79.41%.

Many statistic methods have been developed for prediction and physicochemical properties characterization of P-gp substrates.[28] Various research groups used these statistic methods and complicated descriptors for classification P-gp

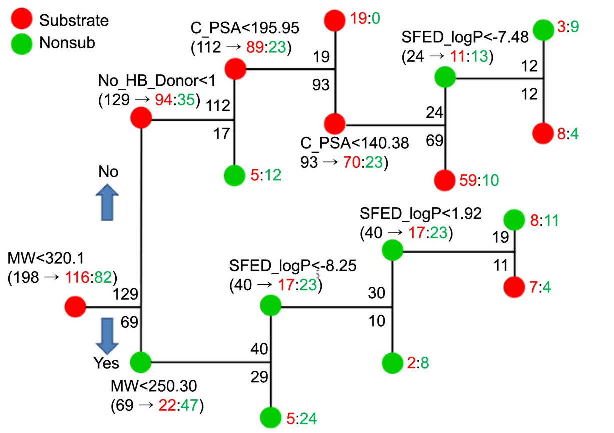

Figure 2. Decision tree and 2D and 3D descriptors built with 198 compounds of P-gp substrates and non-substrates in the training set. Two classes were used: non-substrates (Nonsub; green) and substrates (Sub; red). 2D descriptors from PreADMET were used, 3D descriptors were developed in this study. The classified number of substrates and non-substrates for training set is shown. The arrows show the direction of the branches.

substrates. In this study, our purpose is to suggest the P-gp classification model that can be easily interpreted by medicinal chemists, and easily applicable. Because the P-gp substrates and inhibitors have structural diversity, it is difficult to identify the common structural elements. Therefore intuitive and interpretable descriptors in this suggested RP model can help medicinal chemist to develop and investigate P-gp inhibitors. Additionally, we tried to perform the SVM training with same descriptors for increasing the accuracy. The results show a little increase of accuracy in Table S3. Even the SVM model did not easily be interpreted in the most cases, the model statistically demonstrated the increasing accuracy compare with the RP model.

The First model adopted four 2D descriptors which are the MW, the number of H-bond donors, the number of H-bond acceptors and Topological_PSA. In the result of the first model, about large-sized molecules (MW > 308.892), at least one H-bond donors, many H-bond acceptors (> 10), and high Topological_PSA (< 144.34) are important elements for the P-gp substrate recognition. The second model was described by two 2D and two 3D descriptors which were the MW, the number of H-bond donors, SFED_logP, 3D_PSA. In the results, about large-sized molecules (MW > 320.1), at least one H-bond donors, low logP (< 1.92), and highly polar C_PSA (> 195.95) are important elements for the P-gp substrate recognition in Figure 2. We tabulated all prediction results and category of function in Table S1 and used descriptors in Table S2. Based on analysis of the molecular size, the MW shows that small molecules cannot be P-gp substrates.

The smallest drug substrate of P-gp is "Isosafrole (MW is 162.19 amu)" in the training and test set. Isosafrole is a precursor for 3,4-(Methylenedioxy)phenyl-2-propanone (MDP2P) which is converted into the psychoactive drug MDMA ('ecstasy'). The 20 molecules of P-gp substrate include 10 central nervous system (CNS) drugs like Isosafrole, L_dopa, Phenobarbital, Phenytoin, Cimetidine, Nortriptyline, Protriptyline, Morphine, Promazine and Ondansetron, which have MW under 300.00, but 51 non-substrate, which have MW under 300.00, only include 13 CNS drugs. The P-gp plays a major role in drug efflux at the blood-brain barrier, and may be an underlying factor in the variable responses of patients to CNS drugs.[33] The P-gp is involved in multiple drug resistance in tumour therapy because it can efflux structurally and functionally diverse compounds.[34] The 28 compounds among the substrate data are anticancer drugs in Figure 3, but only 5-compounds among the 115 nonsubstrate dataset are anticancer compounds. Antibiotics are the most clinically important substrates of P-gp efflux systems, the 37 compounds among substrate data are antibiotics compounds in Figure 3. Resistance to antibiotics occurs typically as a result of drug inactivation/modification, target alteration and reduced accumulation owing to decreased permeability and/or increased efflux.[35]

In the work of Xue *et al.*,[29] they built the P-gp substrate prediction model using SVM with 2D descriptors like a number of H-bond donors, sum of charge weighted solvent accessible surface areas of positive charged atoms and many topological descriptors. In the work of Pajeva and Wiese,[36] a general pharmacophore pattern was proposed for the verapamil of P-gp substrate that involved two hydrophobic features, three H-bond acceptors and one H-bond donor. In our study, molecular features used in RP models show consistency with this general pharmacophore pattern. Number of H-bond donors is very important descriptor, both RP models suggest that a compound with the zero number of H-bond donors is discriminated as a nonsubstrate. In the work of Wang *et al.*,[15] unsupervised machine learning approach was explored with some molecular descriptors include *SHbint3* descriptor which definition is E-state descriptors of potential internal H-bond strength that described the H-bond in a molecule in spatial distance. The role of the number of H-bonds suggested as an essential structural requirement for P-gp recognition. Although we cannot determine how many hydrogen bonds are essential for the identification of substrate of P-gp, it seems that many hydrogen bonds are concerned in specificity of substrates of P-gp.

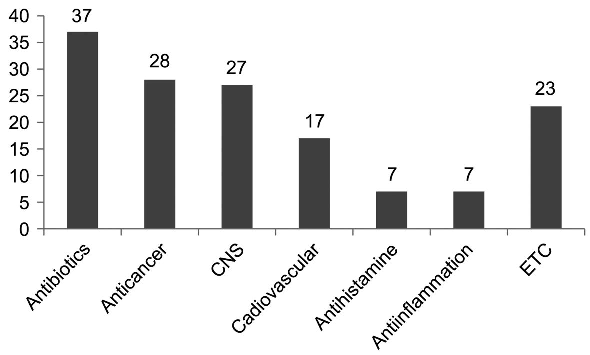

Figure 3. Distribution of the 146-substrate data set, in terms of category of functions.

P-gp, ABC-multidrug transporters, are distributed and expressed in many organs and found especially in the blood-brain barrier, the bile-canalicular membrane of hepatocytes, apical lumen of the intestine and the kidney.[37,38] The ability of drugs to interact with P-gp, therefore, is strongly related with bioavailability of drugs, as well as drug-drug interactions. In metabolism phases in the liver, P-gp exports metabolites into the bile. In the aspect of metabolism process, P-gp trap metabolites and assist exportation of these metabolites, which become hydrophilic after interacting with CYP enzymes. Consequently, the PSA would be important feature to distinguish whether the compounds are substrate or not. In Figure 2, the RP model using 3D descriptors shows that PSA discriminate substrates with high polarity (C_PSA > 195.95) from dataset.

Exogenous substrates such as xenobiotics and toxins are caught by P-gp after partitioning into the plasma membrane but before reaching the cytosol and are then exported or flipped at the expense of ATP hydrolysis.[39] Romsicki *et al.*[40] investigated that the lipid-water partitioning coefficient, logP, of P-gp substrates increased as binding affinity of substrates to P-gp in parallel manner. In our RP model using 2D and 3D descriptors, molecules with lower SFED_logP

(< −8.25), which are relatively hydrophilic, are classified as non-substrates. It is consistent with the fact that hydrophilic drugs or toxins are inserted into membrane first and accumulated in the lipid environment. Experimentally the concentration of drug in the membrane is important for the interaction with P-gp.[40] In our 2D and 3D descriptor model, SFED_logP predicted that relatively hydrophilic compounds are substrates among predicted nonsubstrates group.

Conclusion

There are many examples of the significance of drug transporters to the clinical development of drugs that have been described. For instance, drug-drug interactions and modifying molecules for uptake transporters. Identifying molecules that interact with P-gp transporters is important for drug discovery, but it is commonly determined through laborious *in vitro* and *in vivo* studies. Computational classification model can be used to screen large chemical databases of molecules rapidly and propose those likely to bind as substrates for P-gp.

In this study, the RP approach has been used a successful methodology to classify P-gp substrates and nonsubstrates. We used small number of descriptors and interpretable descriptors. The interpretable descriptors mean that we can analyze their physical meaning easily. We developed some 3D descriptors for increasing accuracy. The model developed in this workis easily calculated and suitable for the rapid prediction of P-gp substrate. We expect that our models can accelerate the virtual screening in the early stages of drug discovery while considering whether a compound includes a P-gp substrate specificity or not.

Acknowledgments. This research was supported by Ministry for Health, Welfare & Family Affairs (A085105) Republic of Korea.

References

1. Kennedy, T. *Drug Disc. Today* **1997**, *2*, 436-444.
2. van de Waterbeemd, H.; Gifford, E. *Nat. Rev. DrugDisc.* **2003**, *2*, 192-204.
3. Chohan, K. K.; Paine, S. W.; Waters, N. J. *Curr. Chem. Biol.* **2008**, *2*, 215-228.
4. Demel, M. A.; Schwaha, R.; Kramer, O.; Ettmayer, P.; Haaksma, E. E.; Ecker, G. F. *Expert Opin. Metab. Toxicol.* **2008**, *4*, 1167-1180.
5. Gottesman, M. M.; Fojo, T.; Bates, S. E. *Nat. Rev. Cancer* **2002**, *2*, 48-58.
6. Pauli-Magnus, C.; Meier, P. J. *Hepatology* **2006**, *44*, 778-787.
7. Higgins, C. F.; Linton, K. J. *Nat. Struct. Mol. Biol.* **2004**, *11*, 918-926.
8. Schinkel, A. H. *Adv. Drug Deliver. Rev.* **1999**, *36*, 179-194.
9. Kim, R. B.; Fromm, M. F.; Wandel, C.; Leake, B.; Wood, A. J.; Roden, D. M.; Wilkinson, G. R. *J. Clin. Invest.* **1998**, *101*, 289-294.
10. Cabrera, M. A.; González, I.; Fernández, C.; Navarro, C.; Bermejo, M. *J. Pharm. Sci.* **2006**, *95*, 589-606.
11. Bain, L. J.; McLachlan, J. B.; LeBlanc, G. A. *Environ. Health Perspect.* **1997**, *105*, 812-818.
12. Litman, T.; Zeuthen, T.; Skovsgaard, T.; Stein, W. D. *Biochim. Biophys. Acta* **1997**, *1361*, 159-168.
13. Seelig, A. *Eur. J. Biochem.* **1998**, *251*, 252-261.
14. Penzotti, J. E.; Lamb, M. L.; Evensen, E.; Grootenhuis, P. D. J. *J. Med. Chem.* **2002**, *45*, 1737-1740.
15. Wang, Y. H.; Li, Y.; Yang, S. L.; Yang, L. *J. Chem. Inf. Model* **2005**, *45*, 750-757.
16. Xue, Y.; Yap, C. W.; Sun, L. Z.; Cao, Z. W.; Wang, J. F.; Chen, Y. Z. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1497-1505.
17. de Cerqueira Lima, P.; Golbraikh, A.; Oloff, S.; Xiao, Y.; Tropsha, A. *J. Chem. Inf. Model.* **2006**, *46*, 1245-1254.
18. Gang, S.; Kim, H.; Oh, W.; Kim, S.; No, K. T.; Nam, K.-Y. *J. Kor. Chem. Soc.* **2009**, *53*, 653-662.
19. Choi, I.; Kim, S. Y.; Kim, H.; Kang, N. S.; Bae, M. A.; Yoo, S. E.; Jung, J.; No, K. T. *Eur. J. Med. Chem.* **2009**, *44*, 2354-2360.
20. Gombar, V. K.; Polli, J. W.; Humphreys, J. E.; Wring, S. A.; Serabjit-Singh, C. S. *J. Pharm. Sci.* **2004**, *93*, 957-968.
21. Bioinformatics & Molecular Design Research Center, Seoul, Korea, PreADMET, version 2.0. 2007; available at http://preadmet.bmdrc.org/
22. Connolly, M. L. *J. Appl. Cryst.* **1983**, *16*, 548-558.
23. Discovery Studio, version 2.1, 2007; Accelrys, Inc., San Diego, Calif.
24. In, Y.; Chai, H. H.; No, K. T. *J. Chem. Inf. Model*. **2005**, *45*, 254-263.
25. Kim, J.; Nam, K.-Y.; Cho, K.-H.; Choi, S.-H.; Noh, J. S.; No, K. T. *Bull. Korean Chem. Soc.* **2003**, *24*(12), 1742-1750,
26. Nam, K.-Y.; Cho, D. H.; Paek, K.; No, K. T. *Chem. Phys. Lett.* **2002**, *364*, 267-272.
27. Accelys Pipeline Pilot, Version 7.0; 2009; available at http://accelrys.com/
28. Huang, J.; Ma, G.; Muhammad, I.; Cheng, Y. *J. Chem. Inf. Model* **2007**, *47*, 1638-1647.
29. Xue, Y.; Li, Z. R.; Yap, C. W.; Sun, L. Z.; Chen, X.; Chen, Y. Z. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1630-1638.
30. Matthews, B. W. *Biochim. Biophys. Acta* **1975**, *405*, 442-451.
31. Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A. F.; Nielsen, H. *Bioinformatics* **2000**, *16*, 412-424.
32. Landis, J. R.; Koch, G. G. *Biometrics* **1977**, *33*, 159-174.
33. Maines, L. W.; Antonetti, D. A.; Wolpert, E. B.; Smith, C. D. *Neuropharmacology* **2005**, *49*, 610-617.
34. Gerhard, F.; Ecker, G. F.; Stockner, T.; Chiba, P. *Drug Discov. Today* **2008**, *13*, 311-317 .
35. Poole, K. *J. Antimicrob. Chemother.* **2005**, *56*, 20-51.
36. Pajeva, I. K.; Wiese, M. *J. Med. Chem.* **2002**, *45*, 5671-5686.
37. van Tellingen, O. *Toxicol. Lett.* **2001**, *120*, 31-41.
38. Thiebaut, F.; Tsuruo, T.; Hamada, H.; Gottesman, M. M.; Pastan, I.; Willingham, M. C. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7735-7738.
39. Gatlik-Landwojtowicz, E.; Aanismaa, P.; Seelig, A. *Biochemistry* **2006**, *45*, 3020-3032.
40. Romsicki, Y.; Sharom, F. J. *Biochemistry* **1999**, *38*, 6887-6896.