

VIRTUAL SEARCHING OF DUMMY TEMPLATE FOR SINENSETIN BASED ON 2D MOLECULAR SIMILARITY USING CHEMDB TOOL

Iqmal Tahir^{1,2,*}, Mohd Noor Ahmad², AKM Shafiqul Islam², and Dachyar Arbain²

¹Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara, Yogyakarta, Indonesia, 55281

²Centre of Excellence for Advanced Sensor Technology, University Malaysia Perlis, 01000 Kangar Perlis, Malaysia

Received July 24, 2012; Accepted November 28, 2012

ABSTRACT

Sinensetin-like molecules have been searched virtually to find a dummy template for Molecular Imprinted Polymer (MIP) synthesis. The aim of this study is to find a novel adsorbent for Solid Phase Extraction (SPE) of sinensetin based on tailored method using dummy template. Virtual searching was done using ChemDB online chemical database containing 5 million commercially available chemicals. Sinensetin as target was converted into SMILES code then it was run with threshold of molecular similarity score of 0.5. There were 90 chemical structures given as the output mainly flavone analogue, three flavonols and several non-flavonoid compounds. The suggested results are three dummy templates that derived from luteolin (flavon) or quercetin (flavonol) with slight alteration by methoxylation process.

Keywords: sinensetin; molecular imprinted polymer; chemical database; molecular similarity; dummy template

ABSTRAK

Molekul-molekul menyerupai sinensetin telah ditelusuri secara maya untuk menemukan templat dumi yang digunakan pada sintesis polimer tercetak molekul (MIP) sinensetin. Tujuan penelitian adalah untuk dapat menemukan adsorben baru untuk ekstraksi fasa padat (SPE) sinensetin yang dibuat berdasarkan pencetakan molekul menggunakan templat dumi. Pencarian maya dilakukan menggunakan database ChemDB secara online yang memiliki 5 juta data senyawa komersial. Struktur sinensetin diubah menjadi kode SMILES kemudian dijalankan pada ambang nilai keserupaan molekul 0,5. Terdapat 90 struktur kimia sebagai output terutamanya golongan flavon, tiga senyawa flavonol dan beberapa senyawa bukan flavonoid. Hasil kajian memperoleh tiga templat dumi yang diturunkan dari luteolin (flavon) atau quercetin (flavonol) melalui proses metoksilasi.

Kata Kunci: sinensetin; polimer tercetak molekul; database kimia; keserupaan molekul; templat dumi

INTRODUCTION

Sinensetin or (Fig. 1) is a bioactive commonly used for treatments of several health problems especially for kidney stone, diabetes mellitus and gout. It can be isolated from herbs like *Orthosiphon stamineus* Benth using solvent extraction and chromatographic technique [1]. Akowuah et al. has separated sinensetin in the microscale using high performance thin layer chromatography (HPTLC) for quantitative determination [2]. For macroscale separation, especially to produce sinensetin as herb product, solvent extraction method is used [3]. If necessary, it needs special techniques such as sonication [1] or microwave-assisted extraction [4]. The alternative of solvent extraction is the Molecular Imprint Solid Phase Extraction (MISPE) process that uses a specific sorbent material related to the target

compound. Molecular imprinting provides increased selectivity and sensitivity over conventional SPE's packing materials [5]. Due to the specific nature of the interaction between the rebinding template and the chemical functionalities within the pores or cavities of the cross linked polymer matrix, molecular imprint polymer

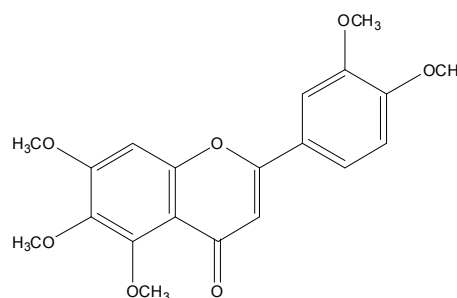


Fig 1. Molecular structure of sinensetin

* Corresponding author. Tel/Fax : +62-274-545188
Email address : iqmal@ugm.ac.id

can discriminate between closely related compounds. This is very useful to extract specific flavonoid compounds available in the plant with many related compounds. Xie et al. [6] and Song et al. [7] have reported isolation of some flavonoids using MISPE method, but still there is no special sorbent for sinensetin. Thus developing MISPE for sinensetin is still important to find a good and selective adsorbent in the column of SPE for large scale application.

To produce MIP for sinensetin in SPE application, sinensetin should be treated as template, but the

availability of high purity sinensetin is limited and very expensive. Therefore, it is not advantageous to use pure sinensetin as template to produce MIP. Thus, an alternative molecule to be used as a dummy template is suggested. Dummy template has some similarities with target, including the physical and chemical properties [8]. Several papers has been reported the use of dummy template to synthesize MIP which showed good performance [9-11]. Many others give low grade MIP with weak binding to the target [12-13]. Searching of the dummy template can be determined by "trial and error" approach using the analogue molecule. A systematic searching of dummy template is possible by adopting molecular similarity analysis that is a good approach popularly in drug design methodology to find the analogue molecule [14-15]. Actually, there are many successful stories to find a new drug based on molecular similarity combine with quantitative structure-activity relationships [16-18]. In practice, the MIP learns the similarity of several properties such as mass, structure, steric, lipophilic and electronic properties from the dummy template.

There are many tools available for molecular similarity analysis such as BRUTUS [19], ShaEP [20], SMSD [21] and many others. Some review papers discussed the application of molecular similarity in the scope of drug design areas [22-25]. Another tool for molecular similarity analysis is ChemDB that uses an online chemical database [26] and can be accessed freely for researchers. In this paper we report a study of the use of ChemDB tool for analysis of sinensetin similarity to find a suitable dummy template for MIP applications.

METHODS

Molecular similarity analysis using ChemDB can be accessed online freely on the website: <http://cdb.ics.uci.edu/cgi-bin/ChemicalSearchWeb.psp>.

The procedure are simplified using input files based on SMILES code (Simplified Molecular Input Files Entry String) of the lead molecules i.e. sinensetin. SMILES code of sinensetin (Fig. 1) is COc1cc2oc(c3ccc(OC)c(OC)c3)cc(=O)c2c(OC)c1OC. Similarity option is

determined with similarity score threshold value of 1–0.5. Four parameters are used to filter off molecular similarity covering of molecular weight, number of rotatable bonds, number of hydrogen donor atoms and calculated log P. The two former parameters represent of steric descriptors, the third and the fourth are assumed to describe electronic and lipophilic properties, respectively. Descriptor log P refers to the logarithm value of the partition coefficient (P), which is defined as the ratio of concentration of neutral species in octanol divided the concentration of neutral species in water [14].

Running the input file will allow ChemDB to search for the target from the database from the server. Searching will explore from the database that contains almost 5 million commercially molecules from the electronic catalogs of over 150 chemical vendors. The output will display as a molecular sketch and the values of four parameters according to the order of the similarity score.

RESULT AND DISCUSSION

In MIP synthesis, the role of template molecules is very important for the formation of the imprinted pores in the polymer. Most of the publications on MIP have reported that the imprinted pores will give different characteristics between MIP and the NIP (control polymer) that is produced without template. Therefore the availability of a template should be firstly addressed. In some cases, the template molecule is not easily available due to limited stock of chemicals, very expensive price and unstable chemicals, or is very toxic for technical application. In this situation, template should be replaced with other alternative molecule which is well known as dummy template. In the case of MIP for sinensetin, the above mentioned problems are encountered. Therefore here we report the process to select a suitable dummy template for sinensetin based on molecular similarity using ChemDB tool.

Molecular Similarity Analysis

Molecular similarity analysis for sinensetin on ChemDB is the first step that should be done. After running the input file using SMILES code of sinensetin, the results are obtained as shown on the list containing of 90 molecular models that are possible to be selected as dummy template. In Fig. 2, we can see the value of similarity score of the models resulted by ChemDB. Score close to the value of 1, represent that the model is closely similar to sinensetin. The less value of similarity score indicates the less molecular association. When the threshold of similarity score is determined at 0.5, then 90 models of molecules have

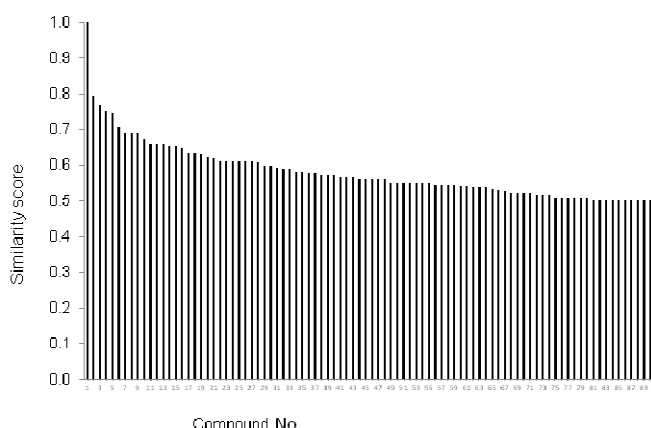


Fig 2. The value of similarity score resulted by ChemDB associated to sinensetin molecule

Table 1. Classification of compound type of the model similar to sinensetin

No	Flavonoid type	Quantity
1	Flavone	79
2	Flavonol	3
3	Other flavonoid (chalcone, flavanone, flavan-3-ol, isoflavone, anthocyanidin and flavanonol)	0
4	Non flavonoid	9

score to suit the criteria.

Type of the models resulted by molecular similarity analysis are varied and the classification of the compounds are shown in Table 1. They are flavone analogues, the other are three models of flavonol type and several models of non-flavonoid compounds. Surprisingly, the result showed that there are no flavonoid compound emerged from the type of chalcone, flavanone, flavan-3-ols, isoflavone, anthocyanidin and flavanonol. Table 1 shows only two types of flavonoid i.e. flavon itself (like the type of sinensetin) and flavonol. Therefore there are only two molecules of this type which are possible to be selected as dummy template for sinensetin. Other possible dummy templates using non-flavonoid type compounds are rejected because there are no strong relationships of molecular structures, especially from the molecular volume descriptor although in this molecular similarity analysis is still included in the result. The molecular volume descriptor did not include in the analysis.

In our analysis, we used four descriptors that are selected to represent sinensetin molecularly, namely molecular weight, number of hydrogen donor atoms, number of rotatable bonds and calculated log P. Table 2 shows the range of recapitulation of the descriptors for 90 model compounds resulted from ChemDB associated to sinensetin structure.

In QSAR analysis, molecular weight (MW) descriptor usually is classified as steric descriptor

because it represents size and atomic variety available in the structure [14]. The data showed that a range value of MW is large from 253.72 to 608.59. MW of sinensetin itself is 372.4. Most models have MW less than MW of sinensetin as shown on Fig. 3(b). The models are mostly flavon analogues which contain hydroxy and methoxy groups. The more hydroxy groups attached instead of methoxy groups decrease MW. Option that some models have hydroxy group also can be used in the strategy of selection dummy template in the next part.

MW descriptor can give a bias prediction especially for molecules that contains heavy atoms like halide chains (fluoride, chloride, bromide or iodide). If molecular size of dummy template is larger than the real target, it will cause the MIP with less selectivity [27]. In this analysis, there also several models that contains halide atom, i.e model no 46, 47 and 48. Halide atoms have atomic size larger than hydrogen, the atom that has been substituted; therefore the molecular size of dummy template is larger than sinensetin. Thus the cavity on MIP will have different size of pores.

MW data also could cause bias data especially to model of dummy template contain cyclic site bound to the flavon skeleton. The effect of cyclic site is a larger pore size, although in sinensetin there is no cyclic site. For sinensetin, all of the side chains are methoxy, therefore a recommended model using cyclic chain will give different sizes of pores in the MIP. Consequently, using dummy template with cyclic chain will tend to produce MIP with lower selectivity to sinensetin.

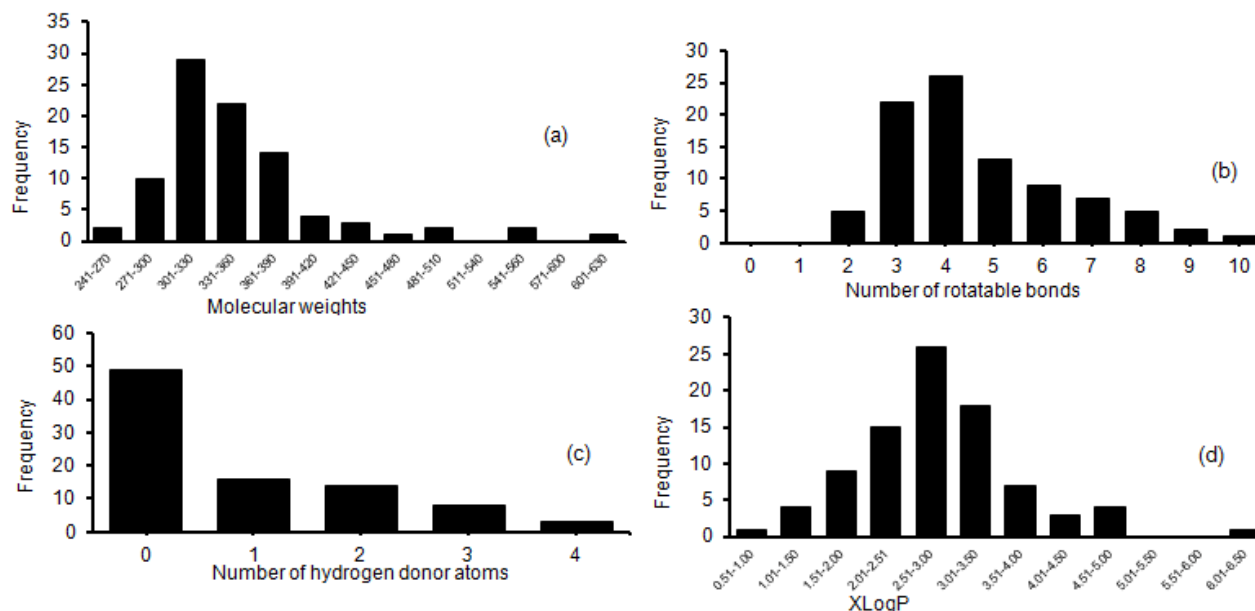
A descriptor of number of rotatable bonds is also representing steric structure. They can show a quantity of non cyclic single bond available in the structure because it was defined as any single bond, not in a ring, bound to a nonterminal heavy (i.e., non-hydrogen) atom [28]. As shown in sinensetin structure, it contains minimum of one rotatable bond in the skeleton of flavonoid and the other will be distributed on the side chains. Actually, all of the models especially from flavon and flavonol type compounds fulfill the criteria.

Fig. 3(b) showed that the number of rotatable bond of the most models has 3 or 4 bonds, and it is less than the value of sinensetin which has 6 rotatable bonds. Actually as shown in sinensetin structure, it contains one rotatable bond in the skeleton of flavonoid and the rest are contributed in the side chain (Fig. 1). It can be explained that most models have less number of side chain that are distributed in the skeleton.

Next is a descriptor of number of hydrogen donor atom. This descriptor represents of functional sites available in the molecule, especially which contains sites with possibilities to give great differences of electronegativity between two poles [14]. Examples of

Table 2. Recapitulation of the descriptors resulted from ChemDB associated to sinensetin structure

Criteria	Molecular weight	Rotatable bonds	H-Bond donors	XLogP
Minimum value	253.72	2	0	0.66
Maximum value	608.59	10	4	6.03
Average	347.20	4.57	0.90	2.81
Sinensetin	372.4	6	0	2.54

**Fig 3.** Histogram of descriptors used for molecular similarity analysis

these sites are hydroxyl, amine, carboxylic acid, etc. In this case, this value of the descriptor is contributed mainly by hydroxyl sites. From sinensetin structure, the value of this descriptor is minimum or zero. It is indicated that the functional sites in the model should be methoxy that are similar in sinensetin. In reality there are many flavon and flavonol compounds that are existed to have a combination of methoxy and hydroxyl sites. Fig. 3(c) shows a distribution of model that is related to the number of hydrogen donor atom. If the number of hydrogen donor atom is zero, it means the model does not contain hydroxyl sites. As seen in Fig. 3(b) that more than half of the models follow this category. They are like sinensetin itself that does not bind a chain side containing hydrogen donor atom.

XlogP descriptor represents calculated lipophilicity parameters which is very useful to measure similarity for application in MIP synthesis. Based on the definition of log P, molecules with lower log P are more polar and vice versa [14]. From the result (Table 2), values of calculated log P for 90 models are varied from 0.66 until 6.03 with the average value of 2.81. Log P that is equal to 2.81 indicates model have semipolar property. If we see the distribution of log P for all models, it is clear that almost all of the models have XlogP in the range of 2.01-2.50 and it is less than sinensetin which has XlogP of 2.54. It can be easily understood as many models

contains hydroxyl site to replace methoxy, consequently their lipophilicity increase (XlogP will decrease).

Calculated log P descriptor is very useful in MIP synthesis especially for the selection of the porogenic solvent. Based on the XlogP value for sinensetin and most of sinensetin-like model which has value of 2.5, it is classified for semi polar properties and then a semi polar solvent can be chosen to be used in the synthesis [29]. Therefore in the synthesis of MIP using this compound as dummy template, semi polar solvent like tetrahydrofuran, chloroform or others can be used.

Selection of Dummy Template

The models for sinensetin from ChemDB list can be selected as dummy template but some of them are still expensive and available in a limited stock, so they are not suitable although they are available commercially from the vendors. As an alternative, we can modify a common flavonoid compound. From the result, there are two types of flavonoid that are possible as a dummy template i.e. flavone and flavonol. In the market, a common compound of flavone is luteolin and of flavonol is quercetin. Their prices are relatively cheap and available in enough stock. The other flavons and flavonols are very expensive and need special

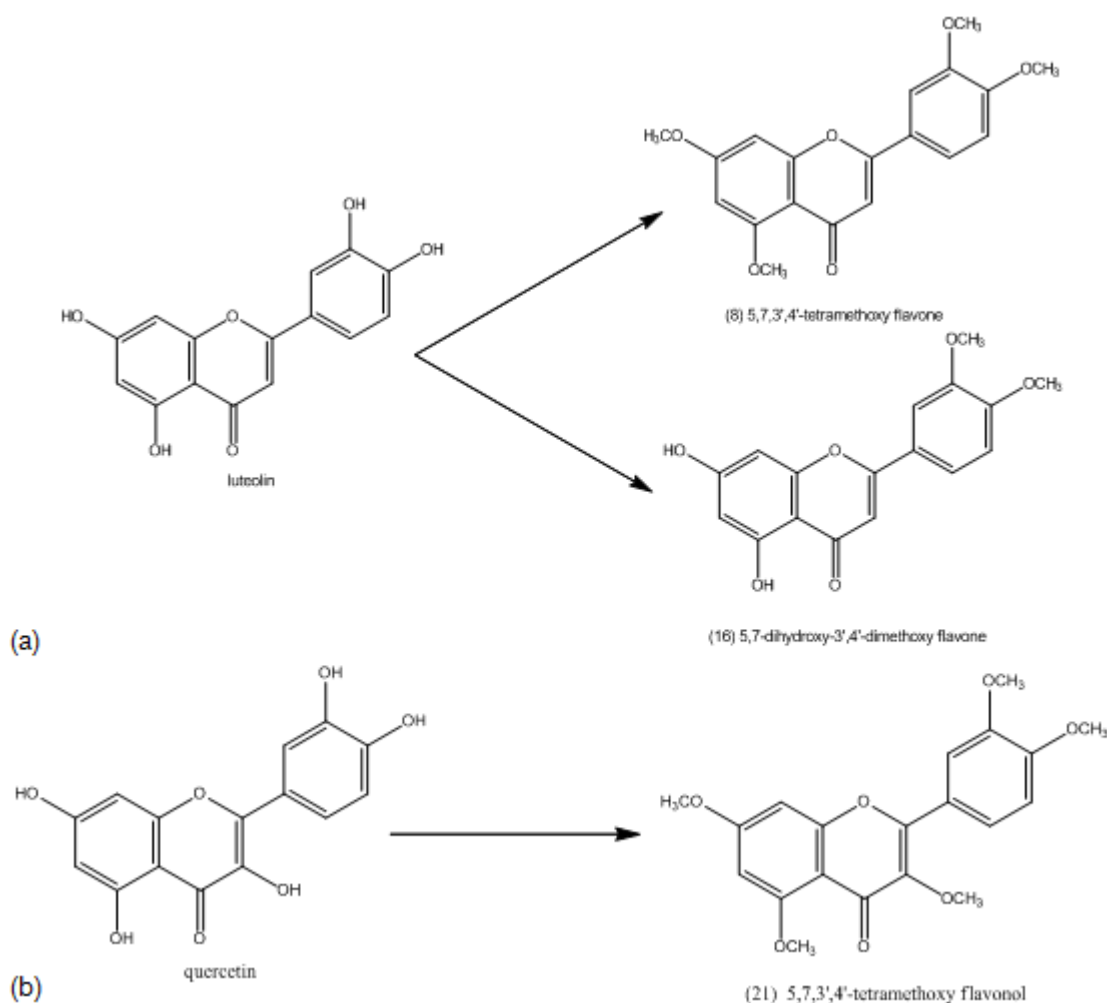


Fig 4. Scheme of methoxylation process of (a) luteolin and (b) quercetin

request to give enough stock as the dummy template for sinensetin. Molecularly, both of them contain hydroxyl sites in the skeleton. As the dummy template for sinensetin, it will be modified into methoxy to comply the characteristic of the result of molecular similarity analysis using ChemDB. Hypothetical scheme reactions for both modifications are given on Fig. 4.

There are four hydroxyl sites in luteolin structure, therefore methoxylation can give many products that are full or part modification of hydroxyl sites. Starting from a flavon type molecules of luteolin (5,7,3',4'-tetrahydroxy flavone), there are two dummy templates matched with ChemDB result (Fig. 4(a)). The first result is compound No 8 that all of the side chain is methoxy (similarity score is equal 0.689). The second is a compound No 16 that is 5,7-dihydroxy-3',4'-dimethoxy flavone (similarity score is equal 0.646).

If a flavonol type molecule is used as the precursor, we can select quercetin. Modification of quercetin that contains five hydroxyl sites will give several compounds because partial or full methoxylation to the hydroxyl

sites. If all of the hydroxyl sites can be changed to methoxy, the modification will give a product (Fig. 4(b)) that is close in its structure to the model No 21 (similarity score is equal 0.617).

Modification of luteolin or quercetin by methoxylation process will give several products. The process can be done using Claisen condensation of luteolin using a precursor which has ability to donor methyl site like dimethyl sulfate or methyl iodide under reflux condition [30-31]. The product hypothetically can be used as dummy template for synthesis of MIP. The cavity on MIP will exist after removing dummy template during washing process and then it can be applied for sorption of sinensetin.

CONCLUSION

To provide a dummy template for sinensetin molecule in the synthesis of molecular imprinted polymer to produce special sorbent applied in MISPE for sinensetin, modified compounds of luteolin (flavon)

or quercetin (flavonol) can be used by methoxylation process. These modified compounds are proposed based on molecular similarity analysis using ChemDB tool. More results of preparation of MIP using these dummy templates for sinensetin will be reported in our next publications.

REFERENCES

1. Hossain, M.A., Salehuddin, S.M., Kabir, M.J., Rahman, S.M.M., and Rupasinghe, H.P.V., 2009, *Food Chem.*, 113, 1, 185–190.
2. Akowuah, G.A., Zhari, I., Norhayati, I., Sadikun, A., and Khamsah, S.M., 2004, *Food Chem.* 87, 4, 559–566.
3. Pan, Y., Abd-Rashid, B.A., Ismail, Z., Ismail, R., Mak, J.W., Pook, P.C.K., Er, H.M., and Ong, C.E., 2011, *Chem. Biol. Interact.*, 190, 1, 1–8.
4. Wang, J-X., Xiao, X-H., and Li, G-K., 2008, *J. Chromatogr. A*, 1198-1199, 45–53.
5. Farrington, K., Magner, E., and Regan, F., 2006, *Anal. Chim. Acta*, 566, 60–68.
6. Xie, J., Zhu, L., Luo, H., Zhou, L., Li, C., and Xu, X., 2001, *J. Chromatogr. A*, 934, 1-2, 1–11.
7. Song, X., Li, J., Wang, J., and Chen, L., 2009, *Talanta*, 80, 2, 694–702.
8. Sellergren B., 2001, *Molecularly Imprinted Polymers. Man-Made Mimics of Antibodies and their Application in Analytical Chemistry*, Elsevier, Amsterdam.
9. Feás, X., Seijas, J.A., Vázquez-Tato, M.P., Regal, P., Cepeda, A., and Fente, C., 2009, *Anal. Chim. Acta*, 631, 2, 237–244.
10. Wang, J.C., Guo, R.B., Chen, J.P., Zhang, Q., and Liang, X., 2005, *Anal. Chim. Acta*, 540, 2, 307–315.
11. Wang, X-J., Xu, Z-L., Feng, J-L., Bing, N-C., and Yang, Z-G., 2008, *J. Membr. Sci.* 313, 1-2, 97–105.
12. Zhang, H., Song, T., Zhang, W., Hua, W., and Pan, C., 2007, *Bioorg. Med. Chem.*, 15, 18, 6089–6095.
13. Yin, J.F., Meng, Z.H., Du, M.J., Liu, C., Song, M.Y., and Wang, H.L., 2010, *J. Chromatogr. A*, 1217, 33, 5420–5426.
14. Kubinyi, H., 1997, *Drug Discovery Today*, 2, 12, 538–546.
15. Huang, H-J., Yu, H.W., Chen, C-Y., Hsu, C-H., Chen, H-Y., Lee, K-J., Tsai, F-J., and Chen, C-Y-C., 2010, *J. Taiwan Ins. Chem. Eng.*, 41, 6, 623–635.
16. Gallegos, A., Carbó-Dorca, R., Ponec, R., and Waisser, K., 2004, *Int. J. Pharm.*, 269, 1, 51–60.
17. Jeong, J.A., Cho, H., Jung, S.Y., Kang, H.B., Park, J.Y., Kim, J.A., Choo, D.J., and Lee, J.Y., 2010, *Bioorg. Med. Chem. Lett.*, 20, 1, 38–41.
18. Chang, T-T., Sun, M-F., Wong, Y-H., Yang, S-C., Chen, K-C., Chen, H-Y., Tsai, F-J., and Chen, Y-C., 2011, *J. Taiwan Ins. Chem. Eng.*, 42, 4, 580–591.
19. Rönkkö, T., Tervo, A.J., Parkkinen, J., and Poso, A., 2006, *J. Comput.-Aided Mol. Des.*, 20, 4, 227–236.
20. Vainio, M.J., Puranen, J.S., and Johnson, M.S., 2009, *J. Chem. Inf. Model.*, 49, 2, 492–502.
21. Rahman, S.A., Bashton, M., Holliday, G.L., Schrader R., and Thornton J.M., 2009, *J. Cheminf.*, 1, 12.
22. Edgar, S.J., Holliday, J.D., and Willett, P., 2000, *J. Mol. Graphics Model.*, 18, 4-5, 343–357.
23. Saliner, A.G., and Gironés, X., 2005, *J. Mol. Struct. THEOCHEM*, 727, 1-3, 97–106.
24. Good, A.C., 2007, *J. Mol. Graphics Model.*, 26, 3, 656–666.
25. Fukunishi, Y., and Nakamura, H., 2009, *J. Mol. Graphics Model.*, 27, 5, 628–636.
26. Chen, J.H., Linstead, E., Swamidass, J., Wang, D., and Baldi, P., 2007, *Bioinformatics*, 23, 17, 2348–2351.
27. Nantasenamat, C., Li, H., Isarankura-Na-Ayudhya, C., and Prachayasittikul, V., 2012, *Chemom. Intell. Lab. Syst.*, 116, 128–136.
28. Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W., and Kopple, K.D., 2002, *J. Med. Chem.*, 45, 12, 2615–2623.
29. Song, X., Wang, J., and Zhu J., 2009, *Mater. Res.*, 12, 3, 299–304.
30. Ragazzon, P.A., Iley, J., and Missailidis, S., 2009, *Anticancer Res.*, 29, 6, 2285–2293.
31. Li, N.G., Shi, Z.H., Tang, Y.P., Yang, J.P., and Dua, J.A., 2009, *Beilstein J. Org. Chem.*, 5, 60, doi:10.3762/bjoc.5.60.