informa
healthcare

HQSAR study of β-ketoacyl-acyl carrier protein synthase III (FabH) inhibitors

ALI ASHEK[1,2], AMOR A. SAN JUAN[1,2], & SEUNG J. CHO[1]

[1]*Korea Institute of Science and Technology, Biochemicals Research Center, PO Box 131, Cheongryang, Seoul 130-650, South Korea, and* [2]*Department of Biomolecular Science, University of Science & Technology, Daejon, South Korea*

Abstract
The enzyme FabH catalyzes the initial step of fatty acid biosynthesis via a type II fatty acid synthase. The pivotal role of this essential enzyme combined with its unique structural features and ubiquitous occurrence in bacteria has made it an attractive new target for the development of antibacterial and antiparasitic compounds. Predictive hologram quantitative structure activity relationship (HQSAR) model was developed for a series of benzoylamino benzoic acid derivatives acting as FabH inhibitor. The best HQSAR model was generated using atoms and bond types as fragment distinction and 4–7 as fragment size showing cross-validated q^2 value of 0.678 and conventional r^2 value of 0.920. The predictive ability of the model was validated by an external test set of 6 compounds giving satisfactory predictive r^2 value of 0.82. The contribution maps obtained from this model were used to explain the individual atomic contributions to the overall activity. It was confirmed from the contribution map that both ring A and ring C play a vital role for activity. Moreover hydroxyl substitution in the ortho position of ring A is favorable for better inhibitory activity. Therefore the information derived from the contribution map can be used to design potent FabH inhibitors.

Keywords: *HQSAR, FabH, benzoylaminobenzoic acid*

Introduction

Fatty acid biosynthesis, the first stage in membrane lipid biogenesis, is carried out by two different fatty acid synthase (FAS) systems in bacteria, plants and animals [1]. In the type I system of animals, including humans, FAS is a single multifunctional polypeptide that catalyzes all the reactions in the elongation pathway [1,2]. On the other hand, in the type II systems of bacteria [3], plants [4], and protozoa [5], fatty acid synthesis is catalyzed by a series of small, soluble proteins that are each encoded by a discrete gene existing as separate proteins. The ubiquitous type II fatty acid synthase (FAS) in bacteria is not only essential to cell survival but also exhibits structural and organizational differences from that in higher organisms, such as humans. Thus, the bacterial fatty acid synthesis pathway offers several unique targets for selective inhibition by chemotherapeutic agents.

Among the FAS II system enzymes FabH, a β-keto-acyl-ACP synthase is found both in gram positive and negative bacteria. FabH performs a critical role in the regulation of the entire pathway by catalyzing the condensation of malonyl-ACP with acetyl-CoA. In addition FabH is structurally distinct and its active site residues are also conserved in different bacterial species [6]. So the successful development of novel and broad-spectrum FabH inhibitors would add a new antibacterial to the shrinking arsenal of antibiotics available to combat against the infections by resistant organism.

QSAR techniques have proven to be extremely valuable in pharmaceutical research to create predictive models as a valuable tool to facilitate the discovery of enzyme inhibitors [7,8]. HQSAR [9] has several potential advantages over existing methods for QSAR. It avoids not only the need for molecular

RIGHTS LINK

alignment and conformation specification inherent in CoMFA [10] and CoMSIA [11] but also the selection and calculation or measurement of the physicochemical descriptors required by classical QSAR. By removing the necessity for molecular alignment, models by HQSAR can be obtained more rapidly than other techniques. It makes HQSAR readily applicable on large datasets, such as combinatorial libraries or database subsets that are not amenable to analysis by existing QSAR methods. In our previous study [12] we developed a 3D QSAR model for a set of benzoylaminobenzoic acid derived FabH inhibitors and superimposed the contour map into the active site of FabH to find out important amino acid residues responsible for ligand binding. Here we have conducted the HQSAR study using the same training and test set to explore individual atomic contribution to molecular bioactivity with visual display of active centers in a compound, i.e., a display of the fragments that most likely contribute to the compound's activity. The results of this study can provide some clues about which new compounds to synthesize, or what possible chemical modifications should bring in existing (tested) compounds. Moreover this can also be employed as a predictive model to virtually screen a library of potential candidate compounds.

Materials and methods

Data sets

Forty-three molecules selected for the present study were taken from the published work by Zhe Nie et al. [6]. The structures of the compounds and their biological data are given in Table I–III. The 2D QSAR models were generated using a training set of 37 molecules and predictive power of the resulting models was evaluated using a test set of six molecules (Table I–III marked with ⋆). The test compounds were selected manually such that the structural diversity and wide range of activity in the data set were included. The biological activity used in the present study was expressed as

$$pIC_{50} = -\log IC_{50} \qquad (1)$$

where IC_{50} is the concentration (μM) of the inhibitor producing 50% inhibition of *Entercoccus faecalis* FabH. pIC_{50} values were used as dependent variables in the HQSAR analysis.

HQSAR analysis

The input data set for HQSAR analysis consists of the 2D chemical structures [13] and their associated biological data. HQSAR analysis involves three main steps: the generation of substructural fragments for each of the molecules in the training set; the encoding of these fragments in holograms; and correlation of the latter with the available biological data. For this purpose we used the novel molecular hologram representation devised by Tripos associates as generated by the HQSAR package [14].

The input molecule is broken into all possible structural fragments (including branched, cyclic, and overlapping fragments) containing user defined minimum (M) and maximum (N) number of atoms. Each unique fragment in the dataset is assigned a specific large integer by means of cyclic redundancy check (CRC) algorithm. Each of these integers corresponds to a bin in an integer array of fixed length L (L is generally in the range 50–500). Bin occupancies are incremented according to the fragments generated. Thus, all generated fragments are hashed [15] into array bins in the range 1 to L. This array is called molecular hologram, [16] and bin occupancies are the descriptor variables. The use of hashing greatly reduces the size of molecular hologram but leads to a phenomenon called 'fragment collision'. During fragment generation, identical fragments are always hashed to same bin, and the corresponding occupancy for that bin is incremented. However, as the hologram length is generally smaller than the total number of unique fragments, different unique fragments can hash to the same bin causing 'collision' between fragments. In order to reduce the probability of identical or similar fragment collisions occurring, values of L are selected to be prime numbers (default values of which are 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401). Computation of the molecular holograms for a dataset of structures yields a data matrix of dimension R × L, where R is the number of compounds in the training set and L is the length of the molecular hologram. Standard PLS analysis is then applied to identify a set of orthogonal explanatory variables (components) that are linear combinations of the original L variables. Leave one out [17] and cross-validation is applied to determine the number of components that yield an optimally predictive model. Once an optimal model is identified, PLS yields a mathematical equation that relates the molecular hologram bin values to the corresponding biological activity of each compound in the data set.
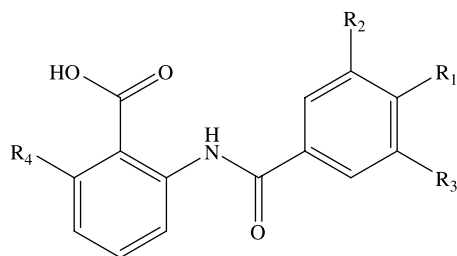
$$\text{Activity}_i = c_o + \Sigma_L c_{il} x_{il} \qquad (2)$$

Where x_{il} is the occupancy value of the molecular hologram of compound i at position or bin l, c_{il} is the coefficient for that bin derived from the PLS analysis, L is the length of the hologram, Activity_i is the biological activity, and c_o is a constant.

HQSAR contribution maps

The results of the HQSAR analysis is graphically displayed as a color-coded structure diagram in which the color of each atom reflects the contribution of that

Table I. Diethyl sulfonamide and [3-phenoxybenzoylamino] benzoic acid derivatives as inhibitors of FabH.

ID	R$_1$	R$_2$	R$_3$	R$_4$	PIC$_{50}$
			Structure		
1	F	H	(methylsulfonyl-diethylamino)	H	5.08
2*	Br	H	(methylsulfonyl-diethylamino)	H	5.80
3	Ph	H	(methylsulfonyl-diethylamino)	H	5.80
4*	Br	Me	(methylsulfonyl-diethylamino)	H	3.80
5	OMe	H	(methylsulfonyl-diethylamino)	H	4.94
6	(N-methylpiperidinyl)	H	(methylsulfonyl-diethylamino)	H	5.66
7	(methoxyphenyl)	H	(methylsulfonyl-diethylamino)	H	5.21
8	(cyclopentylmethoxy)	H	(methylsulfonyl-diethylamino)	H	5.68
9	H	H	(methoxyphenyl)	H	5.57
10	F	H	(methoxyphenyl)	H	5.42

Table I – *continued*

ID	R₁	R₂	R₃	R₄	PIC₅₀
			Structure		
	R_1	R_2	R_3	R_4	PIC_{50}
11	Br	H		H	4.96
12		H		H	4.6
13		H		H	4.42
14		H		H	5.49
15		H		H	5.92
16		H		H	6.54
17		H		H	6.57
18		H		H	4.66
19		H		H	6.96
20		H		H	7.25
21	Br	H		OH	7.21

Table II. Aromatic substitutions on the para position of [3-phenoxy-benzoylamino] benzoic acid derivatives as inhibitors of FabH.
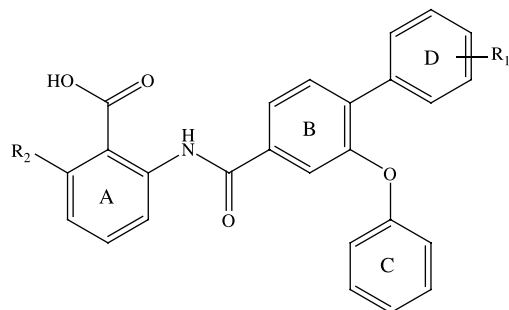
ID	R$_1$	R$_2$	pIC$_{50}$
		Structure	
22	CF$_3$	H	7.02
23	Me	H	6.80
24	CO$_2$H	H	5.68
25	OH	H	6.39
26	OEt	H	6.66
27	SO$_2$Me	H	7.55
28	OCF$_3$	H	6.33
29*	iPr	H	6.10
30	3-Me−4-F	H	6.62
31	2,4-di-F	H	6.80
32	3,4-di-F	H	6.48
33	3-Me−4-Cl	H	6.60
34*	3-Cl−4-F	H	6.24
35	H	OH	8.40

atom to the molecules overall activity. The colors at the red end of the spectrum (red, red–orange, and orange) reflect poor (or negative) contributions, while colors at the green end (yellow, green–blue, and green) reflect favorable (positive) contributions. Atoms with intermediate contributions are colored white. By default, HQSAR specifies the maximal common structure (MCS) based on the similar backbone present in all of the compounds from the training set and denoted by cyan color. It should be noted that the contribution to activity of the atoms involved in MCS were ignored since the template is common to all structures and thus, does not provide a distinguishing feature among the compounds in the dataset.

Predictive r *squared* (r^2 *pred*)

To validate the derived HQSAR models, biological activities of an external test set were predicted using models derived from the training set. The predictive ability of the models is expressed by predictive r^2 value, which is analogous to cross-validated r^2 (q^2) and is calculated by using the formula

$$r^2_{pred} = \frac{SD - PRESS}{SD} \quad (3)$$

Table III. Ring A and C modified [3-Phenoxybenzoylamino] benzoic acid derivatives as inhibitors of FabH.
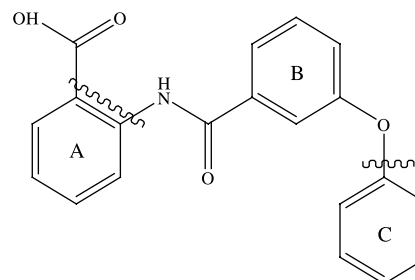
ID	A	C	pIC$_{50}$
		Structure	
36		Ph	5.00
37		Ph	4.37
38		Ph	5.22
39		Ph	6.39
40	Ph	4-Pyr	5.00
41	Ph	3-CO$_2$H-Ph	5.43
42	Ph	4-CO$_2$H-Ph	5.36
43	Ph	4-F-Ph	5.30

where SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules.

Result and discussion

HQSAR investigation was performed using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H) and donor and acceptor (DA). Several combinations of these parameters were considered using the default fragment size (4–7), as follows: A/B, A/B/C, A/B/C/H, and A/C/DA. HQSAR analysis was performed over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins and optimum number of components (LV) were selected

Table IV.   HQSAR analyses for various fragment distinctions using default fragment size (4−7); ($LV_{max} = 8$).

Fragment distinction	q^2	SE_{cv}	r^2	SEE	LV	Length
A/B	.678	.573	.920	.285	5	151
A/B/C	.621	.621	.932	.264	5	307
A/B/C/H	.493	.719	.793	.460	5	97
A/C/DA	.622	.611	.846	.390	4	71

based on the PLS analyses that gave the least cross-validated standard error SE_{cv}. The results of HQSAR analyses for the 37 training set compounds using several fragment distinction combinations are summarized in Table IV. The best model was generated using atoms and bond types as fragment distinction showing cross-validated $r^2(q^2)$ value of 0.678 and noncross-validated r^2 value of 0.92. We have tried to find a better model using other fragment lengths. Thus here we have repeated the analysis with different fragment sizes using the best fragment distinction obtained from previous step to check its influence on key statistical parameters. The statistical results for the different fragment sizes evaluated (2−5, 3−6, 4−7, 5−8, 6−9, and 7−10) are summarized in Table V. Still the best statistical result was obtained with the fragment size 4−7. Like other QSAR techniques HQSAR can also be used to predict the activity of structurally related inhibitors from its fingerprint. The predictive ability of the best HQSAR model derived using the 37 training set molecules was validated by predicting pIC_{50} values for external test set of 6 compounds (marked asterisk in Table I–III), giving satisfactory predictive r^2 value of 0.82. The observed and predictive activities of both training set and test set were shown in Table VI. All compound of the test were fairly predicted with residual values less than one log unit. Figure 1 shows the graph of observed versus predicted activities of both training set and test set.

Besides predicting the activities of untested molecules, the QSAR model plays an important role to provide hints about the relation of different molecular fragments to biological activity. In HQSAR, the model is graphically represented in the form of contribution maps where the color of each atom reflects the

Table V.   HQSAR analysis for the influence of various fragment sizes using the best fragment distinction (A/B).

Fragment size	q^2	SEcv	r^2	SEE	LV	Length
2−5	.674	.558	.932	.263	5	401
3−6	.677	.579	.915	.295	5	83
4−7	.679	.573	.938	.269	6	53
5−8	.588	.648	.916	.293	5	257
6−9	.621	.622	.922	.282	5	401
7−10	.645	.603	.923	.280	5	401

Table VI.   Experimental activities (pIC_{50}) and predicted activities (PA) with residuals (Δ) by HQSAR.

Compound	pIC_{50}	HQSAR PA	Δ
1	5.08	5.22	−0.14
2★	5.80	5.61	0.19
3	5.80	5.85	−0.05
4★	3.80	3.52	0.28
5	4.94	4.83	0.11
6	5.66	5.87	−0.21
7	5.21	4.99	0.22
8	5.68	5.57	0.11
9	5.57	5.51	0.06
10	5.42	5.43	0.01
11★	4.96	5.16	−0.20
12	4.60	4.47	0.13
13	4.42	4.32	0.10
14	5.49	5.86	−0.37
15	5.92	5.94	−0.02
16	6.54	6.23	0.31
17	6.57	6.85	−0.28
18	4.66	4.68	−0.02
19	6.96	6.33	0.63
20★	7.25	6.78	0.47
21	7.21	7.11	0.10
22	7.02	6.87	0.15
23	6.80	6.61	0.19
24	5.68	6.28	−0.60
25	6.39	6.97	−0.58
26	6.66	6.75	−0.09
27	7.55	7.54	0.01
28	6.33	6.20	0.13
29★	6.10	6.93	−0.83
30	6.62	6.63	−0.01
31	6.80	6.87	−0.07
32	6.48	6.60	−0.12
33	6.60	6.74	−0.14
34★	6.24	6.86	−0.62
35	8.40	7.73	0.67
36	5.00	5.08	−0.08
37	4.37	4.58	−.021
38	5.22	5.54	−0.32
39	6.39	6.45	−0.06
40	5.00	4.95	0.05
41	5.43	5.22	0.21
42	5.36	4.98	0.38
43	5.30	5.43	0.13

contribution of that atom to the molecule's overall activity. From a discussion point of view, the most important fragments of the compound **35** (the most potent inhibitor of the data set) are shown in Figure 2.

It is seen from the contribution map that the ring A has a favorable contribution to the activity. This is supported from the fact that when benzene ring is replaced by thiophene ring (**36**, **37**) it loses potency. But the ortho carbon of ring A was found red or orange−red in compound **12**, **13** and **28** (Figure 3) whereas in the rest of the compounds it had positive or neutral contribution to the activity. Interestingly when the hydrogen of this ortho carbon was substituted by hydroxyl group (in compound **35** and **39**) it showed
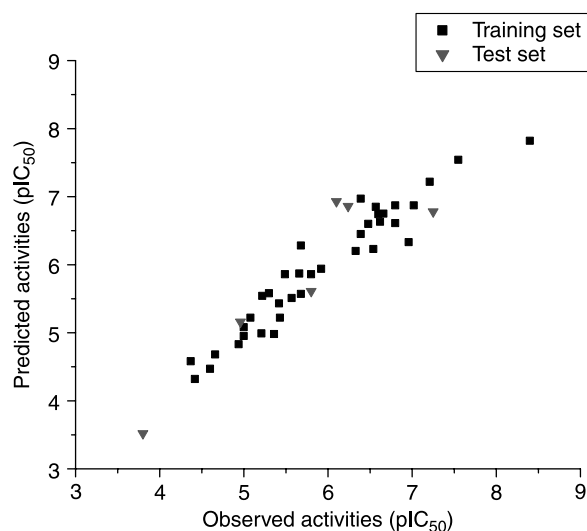
Figure 1. Correlation between the observed and the predicted activities (pIC$_{50}$).

a positive contribution (Figure 2). So our HQSAR contribution map recommends the placement of a hydroxyl group in ortho carbon of ring A for better activity.

From the contribution map of compound **35** we can tell that the green colored ring C is also important for activity. This is in agreement with the fact that the inhibitory activities of the compounds **10** and **20** with phenoxy are higher than those of their diethyl sulfonamide derivatives, compound **1** and **3**. Along

with these regions, the presence of yellow color in ortho and meta carbon of ring D in Figure 2 indicates its positive contribution to activity. The result is consistent with the fact that when the para position of the ring B is substituted by phenyl it shows greater activity. This can be seen with compound **20** in which the incorporation of phenyl group in the para position of ring B results in an increase of activity as compared to compound **9**. Although HQSAR results are subject to chance correlation, the information obtained from these contribution maps can be used for further development of FabH inhibitors.

Conclusion

Here we have applied the principle of HQSAR analysis to study the 2D QSAR of benzoylamino benzoic acid derivatives as FabH inhibitors. The model developed after optimizing the atom and bond type parameters possesses both good internal and external consistency, which indicates that this model is statistically robust with good correlative and predictive power. Moreover the contribution maps obtained from this model was used to explain the individual contribution of the atoms to the overall activity of the compound. It identified rings A and C as positive contributors for activity and evaluated the presence of *o-hydroxyl* as critical for activity. Thus this study can contribute to the design of structurally related FabH inhibitors.
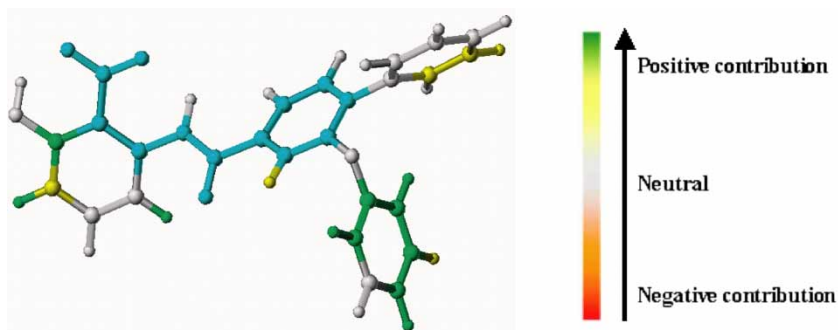

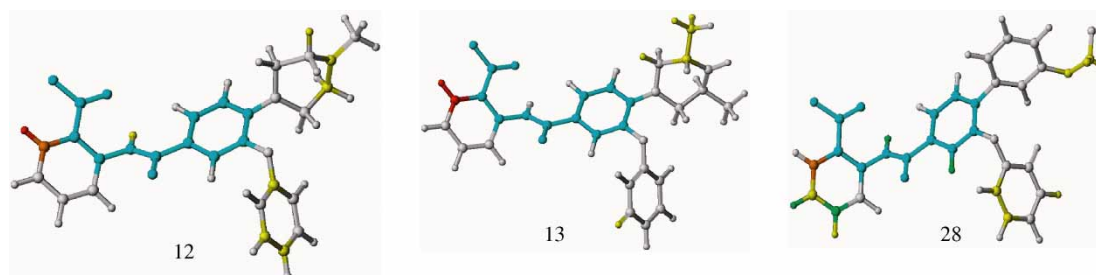Figure 2. Contribution map of compound **35** (see colour online).

Figure 3. Negative contribution of ring A ortho carbon of compound **12**, **13** and **28** (see colour online).

References

[1] Jayakumar A, Tai MH, Huang WY, Al Feel W, Hsu M, Abu-Elheiga L, Chirala SS, Wakil SJ. Human fatty acid synthase: Properties and molecular cloning. Proc Natl Acad Sci USA 1995;92:8695–8699.

[2] Chirala SS, Huang WY, Jayakumar A, Sakai K, Wakil SJ. Animal fatty acid synthase: Functional mapping and cloning and expression of the domain I constituent activities. Proc Natl Acad Sci USA 1997;94:5588–5593.

[3] Cronan JE, Rock CO. Biosynthesis of Membrane Lipids. In: Neidhardt FC. et al. editors. *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology. 2nd ed. Washington DC: American Society for Microbiology; 1996. p 612–636.

[4] Clough RC, Matthis AL, Barnum SR, Jaworski JG. Purification and characterization of 3-ketoacyl–acyl carrier protein synthase-III from spinach: A condensing enzyme utilizing acetyl-coenzyme-A to initiate fatty acid synthesis. J Biol Chem 1992;267:20992–20998.

[5] Waller RF, Ralph SA, Reed MB, Su V, Douglas JD, Minnikin DE, Cowman AF, Besra GS, McFadden GI. A type II pathway for fatty acid biosynthesis presents drug targets in *Plasmodium falciparum*. Antimicrob Agents Chemother 2003;47:297–301.

[6] Nie Z, Perretta C, Lu J, Su Y, Margosiak S, Gajiwala KS, Cortez J, Nikulin V, Yager KM, Appelt K, Chu S. Structure-based design, synthesis, and study of potent inhibitors of β-ketoacyl–acyl carrier protein synthase III as potential antimicrobial agents. J Med Chem 2005;48:1596–1609.

[7] Van de Waterbeemd H. The history of drug research: From Hansch to the present. Quant Struct Act Relat 1992;11:200–204.

[8] Van de Waterbeemd H. Recent progress in QSAR technology. Drug Des Discov 1993;9:277–285.

[9] Heritage TW, Lowis DR. Molecular hologram QSAR. In rational drug design: Novel methodology and practical applications. ACS Symposium Series 1999;719:212–215.

[10] Crammer III RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 1988;110:5959–5967.

[11] Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 1994;37:4130–4146.

[12] Ashek A, Cho SJ. A combined approach of docking and 3D QSAR study of β-ketoacyl–acyl carrier protein synthase III (FabH) inhibitors. Bioorg Med Chem 2006;14:1474–1482.

[13] Ash S, Cline MA, Homer RW, Hurst T, Smith GB. "SYBYL line notation (SLN): A versatile language for chemical structure representation. J Chem Inf Comp Sci 1997;37:71–79.

[14] HQSAR Software. Tripos Associates: 1699 South Hanley Road, Suite 303, St. Louis, MO 63144.

[15] Knuth DE. Sorting and searching: The art of computer programming. 2nd ed. Reading, MA: Adison-Wesley; 1973.

[16] Flower DR. On the properties of bit string-based measures of chemical similarity. J Chem Inf Comput Sci 1998;38:379–386.

[17] Cramer III RD, Bunce JD, Patterson DE, Frank IE. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant Struct Act Relat 1988;7:18–25.