Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 853 (2007) 114–122
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Complex retention behavior of pyrimidines on biomembrane-mimic immobilized-artificial-membrane phase

Hai-Bin Luo [1], Chuanqi Zheng, Yuen-Kit Cheng *

Department of Chemistry, Hong Kong Baptist University, Kowloon Tong, Hong Kong

Received 23 November 2006; accepted 27 February 2007
Available online 21 March 2007

Abstract

The influence of the chemical substitutions on the interfacial interactions of pyrimidines with the phospholipid-mimic immobilized-artificial-membrane (IAM) chromatographic stationary phase was evaluated. Monocyclic pyrimidine nucleic acid bases (nucleobases) were revealed behaving differently from their bicyclic purine counterparts substantially. The computed electrostatic potential surfaces for both the IAM phase and the interacting nucleobases are intuitive in deconvoluting the retention patterns of pyrimidines molecularly. A structure-retention model has also been derived using quantitative 3D-QSAR methodology pertinent to the IAM-retention of pyrimidines for the potential use in molecular design. IAM phase is found particularly suitable in assessing the retention of pyrimidines with bulky or elongated exocyclic substituents in the biological context than the alkyl-based chromatographic counterparts.

Keywords: Pyrimidine; IAM; Retention; CoMSIA; ESP; Modeling

1. Introduction

Besides the ubiquitous role of cytosine and thymine/uracil as an important part of the genetic basis, various other natural or synthetic monocyclic pyrimidine bases are commonly used for biomedical and pharmaceutical purposes [1]. Quite a number of studies have been carried out towards the understanding of the bioactivities and retention properties of pyrimidines, usually fewer than 10 compounds unfortunately, using both experimental and theoretical approaches. Among these, the alkyl-type reversed-phase liquid chromatography (RPLC) studies for the separation of pyrimidines and their derivatives have been well documented [2-6], but the results might not be directly relevant in the biological milieu. Instead, commercially available immobilized-artificial-membrane (IAM) columns (Fig. 1a), which are currently prepared by linking phospholipids or their analogues at monolayer density to silica supporting particles, presumably provide biologically more relevant retention data than those from the largely hydrophobic (except for the residual uncapped polar groups on the solid support) alkyl-type stationary phases [7]. Not uncommon, as a result, IAMs have been developed as an important *in vitro* physico-chemical screening device to predict unfacilitated passive transport of oral drugs through cellular membranes and proved to be useful as a promising tool in modeling the membrane/solute interactions similarly as liposomal membranes [7–12].

Quantitative structure-retention relationships (QSRRs) for over thirty purine structural congeners on Waters XTerra^{TM} MS C_{18} [13,14], Alltech Hypersil MOS-2 C_8 [14], and IAM [14] columns using comparative molecular field analysis (CoMFA) [15], a 3D-QSAR technique, have been elucidated. For pyrimidine bases, however, in view of their smaller skeletons than those of purines, their properties are in principle relatively more influenced by the functional substitutions except for the simplest cases. Thus, the equilibrium retention properties of pyrimidine bases in chromatographic systems, especially of IAM, are not likely deducible from the available 2D/3D physico-chemical properties of purine molecules.

One of the main objectives herein is to investigate the physico-chemical factors influencing the retention patterns and interfacial interactions of pyrimidines with phospholipids in IAM. The other is to derive structure-retention models which

* Corresponding author. Tel.: +852 34117066; fax: +852 34117348.
E-mail address: ykcheng@hkbu.edu.hk (Y.-K. Cheng).
[1] Present address: School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou 510080, People's Republic of China.

1570-0232/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2007.02.059

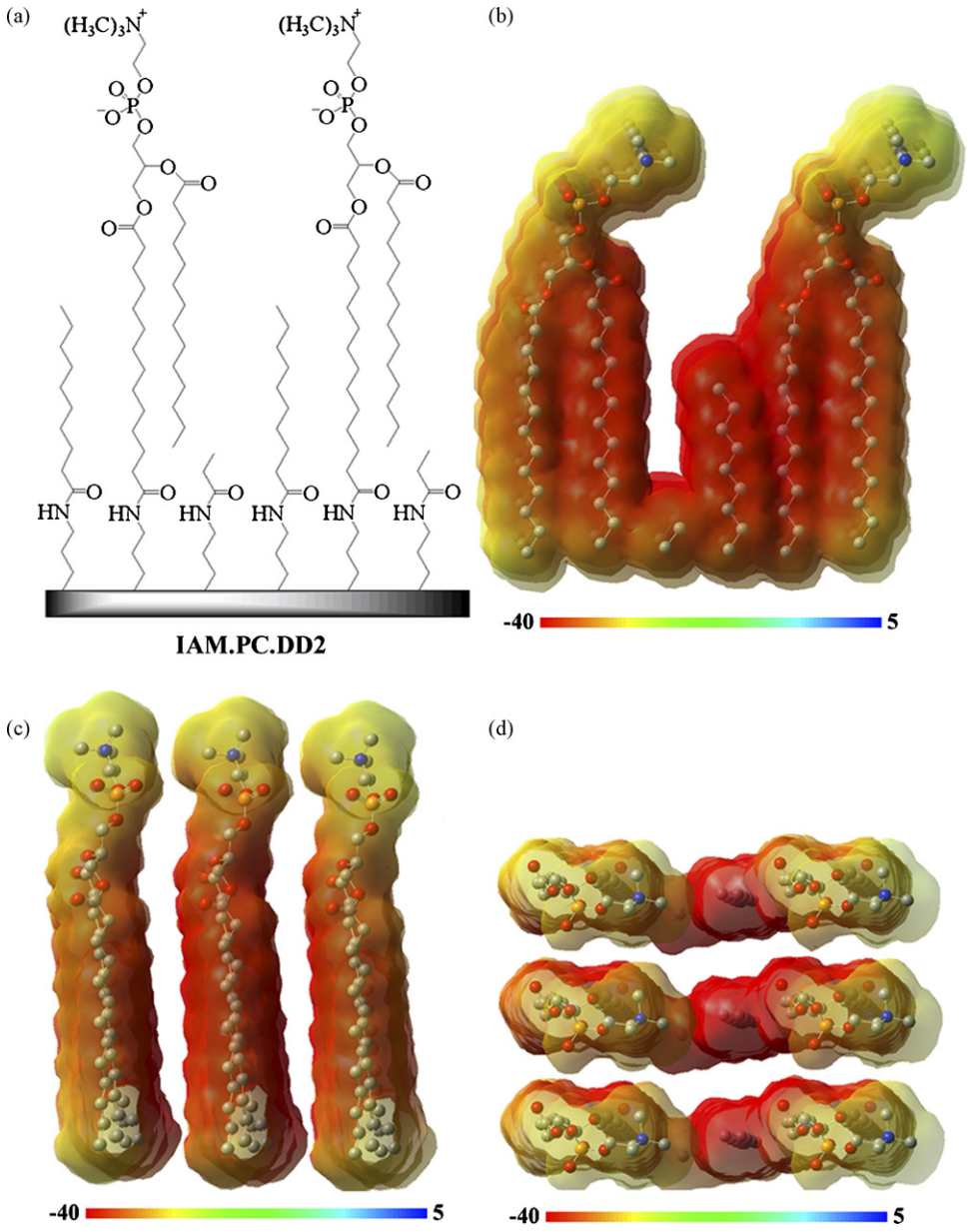

Fig. 1. Schematic structure and the calculated electrostatic potential (ESP) surfaces of the essential components in the IAM.PC.DD2 chromatographic stationary phase. Six PC molecules (2×3) were used to represent the IAM.PC.DD2 retention phase. In (b)–(d), the ESP surfaces were displayed separately along the three orthogonal Cartesian directions. The color-coded ESP is expressed as the potential energy felt by a probe of positive unit charge at a particular point in space. The ESP was calculated at the PM3 level using Gaussian 03 [22] and being mapped on an isosurface of the total SCF electronic density at a value of 0.0004 au. The color bars range from red (electronegative) through green to blue (electropositive) with the mapped ESP values from −40 to 5 hartrees linearly. In real situation, the ESP of the bottom part of the IAM will be modified by the silica support which is not important in this study and will not change our conclusion there obtained. Please refer to the web-version for colors.

are capable of elucidating the quantitative relationships relating the 3D structural descriptors of the pyrimidines to their retention data on various columns using both the CoMFA and comparative molecular similarity indices analysis (CoMSIA) [16,17] approaches, especially on the IAM.PC.DD2 stationary phase. Comparison with the purine counterparts is important in obtaining a complete picture of the biointerfacial properties covering the most representative nucleobases and their analogues. In the present study, the Alltech C_8, Waters C_{18}, and Regis IAM.PC.DD2 columns have been employed for the comparison. The obtained chemometric models were also cross-validated and evaluated with respect to the size and conformational effects of large substituents. Quantum-mechanically derived electrostatic potential surfaces of the IAM stationary phase and selected analytes were performed in dissecting the convoluted cases.

2. Materials and methods

2.1. Materials

All pyrimidines were purchased from Sigma–Aldrich of the highest purities available and used without further purification

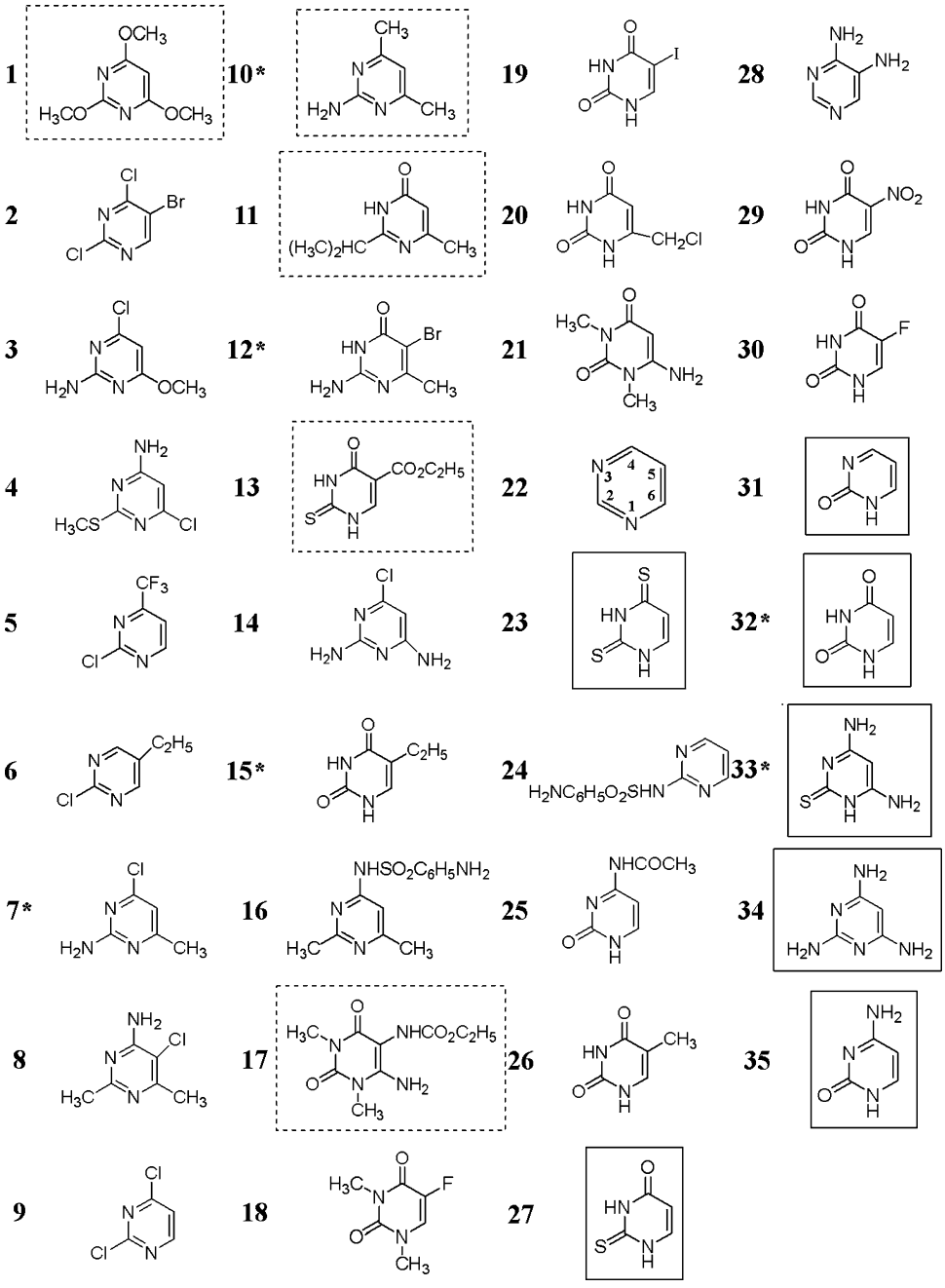

Fig. 2. Chemical structures of the 35 pyrimidines. The analytes comprising the test set are marked with asterisks and the remaining 29 analytes are taken as the training set. The conventional atom numbering of the core ring for all the structures is given in analyte **22**. Refer to Section 2 for the chemical names. See text for the specific discussion of the boxed analytes.

(the bracketed bold-face number corresponds to the chemical structures in Fig. 2): 2,4,6-trimethoxypyrimidine (**1**), 2,4-dichloro-5-bromopyrimidine (**2**), 2-amino-4-chloro-6-methoxypyrimidine (**3**), 4-amino-6-chloro-2-methyl-thiopyrimidine (**4**), 2-chloro-4-(trifluoromethyl)pyrimidine (**5**), 2-chloro-5-ethylpyrimidine (**6**), 2-amino-4-chloro-6-methylpyrimidine (**7**), 4-amino-5-chloro-2,6-dimethylpyrimidine (**8**), 2,4-dichloropyrimidine (**9**), 2-amino-4,6-dimethylpyrimidine (**10**), 2-isopropyl-6-methyl-4-pyrimidinol (**11**), 2-amino-5-bromo-6-methyl-4-pyrimidinol (**12**), 5-carbethoxy-2-thiouracil (**13**), 2,6-diamino-4-chloropyrimidine (**14**), 5-ethyluracil (**15**), sulfisomidine (**16**), ethyl(6-amino-1,2,3,4-tetrahydro-1,3-dimethyl-2,4-dioxo-5-pyrimidinyl)-carbamate (**17**), 1,3-dimethyl-5-fluorouracil (**18**), 5-iodouracil (**19**), 6-(chloromethyl)uracil (**20**), 6-amino-1,3-dimethyluracil (**21**), pyrimidine (**22**), dithiouracil (**23**), sulfadiazine (**24**), N^4-acetylcytosine (**25**), thymine (**26**), 2-thiouracil (**27**), 4,5-diaminopyrimidine (**28**), 5-nitrouracil (**29**), 5-fluorouracil (**30**), 2-thiopyrimidine (**31**), uracil (**32**), 4,6-diamino-2-thiopyrimidine (**33**), 2,4,6-triaminopyrimidine (**34**), and cytosine (**35**). NaH_2PO_4 and Na_2HPO_4 used for preparing the pH buffers were obtained from BDH. HPLC-grade methanol was obtained from Lab-Scan.

Chromatograms were obtained using a Waters 2695 HPLC system equipped with a 996 photodiode-array detector set. The Alltech Hypersil MOS-2 C_8 (150 mm × 4.6 mm, 5 μm), Waters XTerra^{TM} MS C_{18} (150 mm × 4.6 mm, 5 μm), and Regis IAM.PC.DD2 (150 mm × 4.6 mm, 12 μm) columns, were used for the HPLC experiments. Retention detection was performed at 254 nm. Each pyrimidine base together with the internal standard 1,3-dimethyl-5-fluorouracil (**18**) was dissolved at the concentration of ~0.2 mg/ml in an HPLC-grade methanol solution and chromatographed individually and isocratically. Mobile phases consisted of 20:80 (v/v%) methanol/sodium phosphate buffer (35 mM, pH 7.3). The column temperature and flow rate were kept at 25 °C and 1.0 ml/min, respectively. The retention data expressed as capacity factor (k) were calculated as $t_R/t_0 - 1$, where t_R and t_0 are the retention time of sample peak and column void volume time, respectively. The latter was estimated by citric acid measured at 210 nm [7,18]. All the retention measurements performed within 3 weeks were triplicates resulting in relative standard deviations less than 2.1% and 1.0% for each sample solution and the internal standard, respectively.

2.2. Molecular modeling and statistical analysis

All the modeling and 3D-QSAR calculations were performed by the program suite Sybyl (Linux version 6.9.2) [19]. The electrostatic potential surfaces (Section 3.1.4) were obtained by Gaussian 03 and the computational details can be found in the captions of Figs. 1 and 4.

Based on the experimental pK_a values of 12 representative pyrimidines (Appendix A), exocyclic amino groups tend to enhance the pK_a values relative to the bare pyrimidine **22**, whereas halogen substituents seem to offset this enhancement. Thymine and uracil are exceptions and thus the 14 uracil analogues in this study (Fig. 2) should give pK_a higher than 8 and mainly adopt their unionized forms at pH 7.3, whereas for the remaining 19 analytes (except **28** and **34)** should possess pK_a lower than 6. Therefore, the 35 pyrimidines in this study were reasonably modeled in their neutral forms at pH 7.3 (experimental condition). The conspicuous exception is 2,4,6-triaminopyrimidine (**34**) with a pK_a value of 6.81 which is within ±0.5 of pH 7.3. Therefore, it might still be significantly protonated at the experimental condition.

The initial structures of the pyrimidines were optimized (Tripos force field) with a non-bond cutoff of 8 Å using the Powell conjugate-gradient algorithm. The convergence criterion was set to 0.001 kcal/mol. The parent pyrimidine base (**22**) served as an alignment template for superposition. The steric and electrostatic interaction energies were calculated respectively, with the Lennard–Jones and Coulomb potential functions adopting the empirical Gasteiger–Hückel (GH) partial-atomic charges. An sp^3-hybridized carbon atom with a formal charge of +1 was employed as the interaction probe. The steric and electrostatic interaction energies were truncated at 25 kcal/mol. The dimensions of the CoMFA lattice were set to 20 Å × 20 Å × 20 Å with the grid points 2.0 Å apart.

In CoMSIA, similarity is expressed parametrically in terms of the steric (S), electrostatic (E), hydrophobic (H), and H-bond donor/acceptor (D/A) fields, which are more readily envisaged as the various spatial contributions of intermolecular interactions than the traditional CoMFA (which partitioned merely into the steric and electrostatic interactions) [16,17]. CoMSIA similarity indices were calculated using the same lattice box as in the CoMFA calculations. A Gaussian-type functional form for the parametric fields with the default attenuation factor of 0.3 was used in the CoMSIA approach.

The robustness of the partial-least-squares (PLS) analysis embedded in CoMFA/CoMSIA was addressed by internal cross-validation using the leave-one-out (LOO) procedure [20,21]. The PLS regression together with the LOO cross-validation was employed to determine the optimal number of principal components (PCs) and to formulate the CoMFA/CoMSIA models using the standard procedures.

3. Results and discussion

All retention data are expressed in terms of log k in the following discussion unless stated otherwise. The numbering of the analytes (in bold face) refers to the decreasing retention time order on the C_8 column (Appendix B).

3.1. Comparison of the retention data from various columns

The HPLC retention data (Appendix B) of the 35 pyrimidines were plotted in Fig. 3a in a decreasing retention-time order with respect to the Alltech C_8 chromatographic experiments. All the analytes were retained longer on the more hydrophobic C_{18} than on the C_8 column with an almost linear correlation (Fig. 3b), which corroborated the same putative retention mechanism for these pyrimidines on the two alkyl phases.

Owing to the larger particle size of the IAM.PC.DD2 column (12 μm) than those of the alkyl C_8/C_{18} counterparts (5 μm) and the reduced hydrophobicity, most of the pyrimidines gave significantly shorter IAM retention times under identical chromatographic conditions than on the alkyl phases. In spite of this, quantitative correlations ($R^2 = 0.833$ and 0.826, respectively) are found between IAM and C_8/C_{18} retention (Fig. 3c and d) with most data points lying above the ideal line resulting in an average absolute retention reductions in log k of 0.419 and 0.559 (i.e., $|C_8 - IAM|$ and $|C_{18} - IAM|$), respectively. Only about one-third of the analytes (**14**, **19**, **23**, **24**, **27**, **28**, **30–32**, **35**) gave fairly similar C_8 and IAM retention ($C_8 - IAM \leq 0.2$), whereas 4,6-diamino-2-thiopyrimidine (**33**) and 2,4,6-triaminopyrimidine (**34**) were retained exceptionally longer on the IAM column (Fig. 3c) with their retention increments ($IAM - C_8$) at 0.390 and 0.493, respectively. These subtle aspects of the IAM-retention of pyrimidines are appeared to be governed by factors including the electrostatic/H-bond effects, the steric/hydrophobic effects, and the mixed effects as dissected in more details below separately.

3.1.1. Electrostatic/H-bond effects

The eight analytes possessing only exocyclic amino or carbonyl or thiocarbonyl groups (**23**, **27**, **28**, **31–35**, boxed in solid lines in Fig. 2), all gave retention reductions ($C_8 - IAM$) less

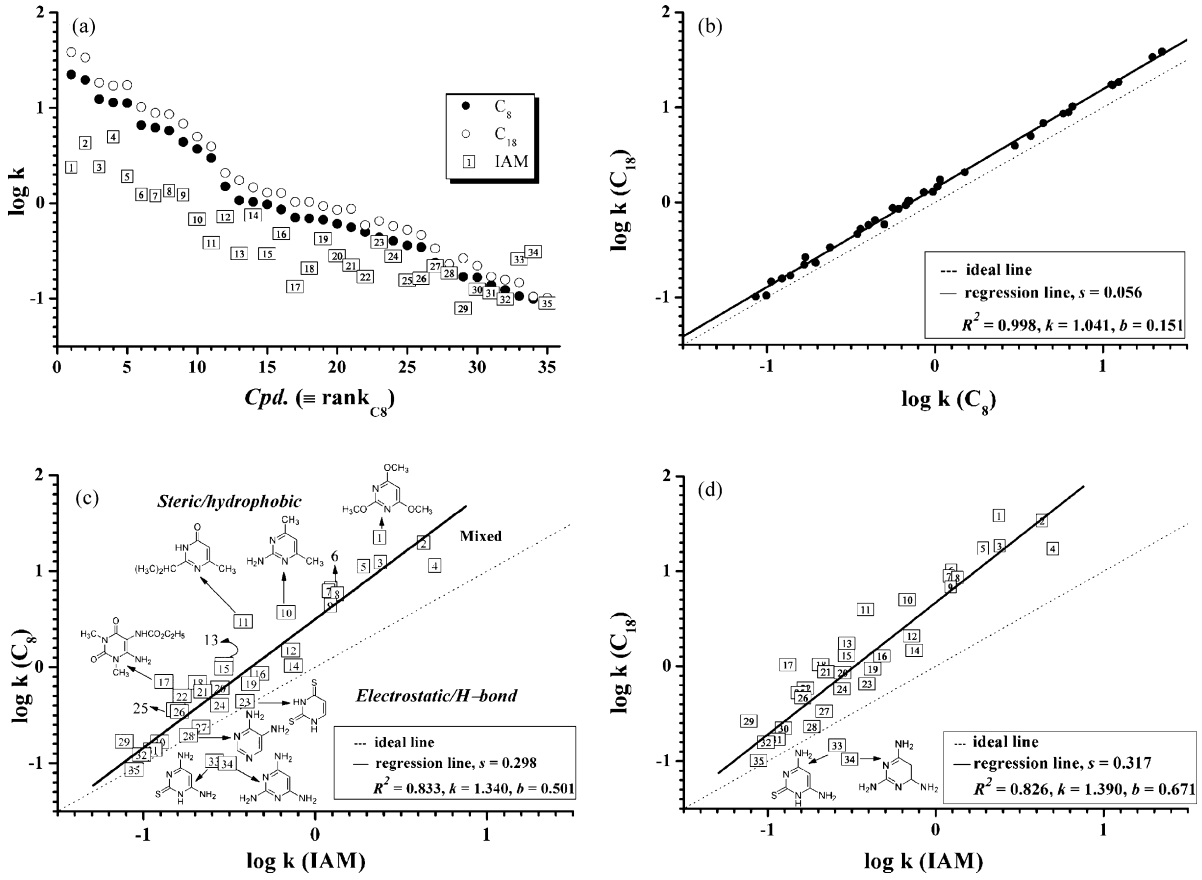

Fig. 3. Correlations of retention data (log k) on various columns. In (b)–(d), the solid line is the actual linear regression curve and the dotted straight line indicated a would-be ideal correlation case (slope = 1 and intercepts = 0) for reference. R, Pearson correlation coefficient; k, slope of linear regression; b, Y-intercept of linear regression; s, standard deviation of residuals. In (c), the retention behaviors of the analytes above the regression line mostly attributed to the steric/hydrophobic interactions; those below the ideal line were mainly predominated by the electrostatic/H-bond interactions; and others by the mixed effects.

than 0.2, putatively interacting strongly with the oxygen atoms in the glycerol backbone or phosphate ester of phospholipids via directional H-bonds (by exocyclic amino group) and may even possess electronegative attraction to the ammonium group of phospholipids (by carbonyl/thiocarbonyl). H-bond or electrostatic interactions compensate the retention reduction effects of the larger particle size and the lower hydrophobicity of the IAM phase. Analyte **22**, a bare pyrimidine, eluted after **23**, **27**, **28**, **33**, and **34** on the C_8 column (Appendix B), but it gave the shortest retention time among them on the IAM phase instead, which corroborates the strong interactions between phospholipids and these latter pyrimidines (**23**, **27**, **28**, **33**, and **34**).

3.1.2. Steric/hydrophobic effects

Alkyl or bulky functional substituents might attenuate or even disrupt the H-bond and electrostatic interactions with IAM. Indeed, for the five analytes **1**, **10**, **11**, **13**, and **17** (Fig. 2, boxed in dashed lines) with their retention reductions ($C_8 - IAM$) larger than 0.400, the steric/hydrophobic effects dominate the intermolecular interactions between these analytes and phospholipids although **10**, **11**, **13**, and **17** have the potential to interact with phospholipids via electrostatic and/or H-bond interactions. Hence, much stronger retention on the C_8 column was observed than on the IAM counterpart (Fig. 3c).

3.1.3. Mixed effects

For the remaining 20 analytes unaccounted for yet (except **16** and **24**), mixed effects induced by halogen substituents in addition to the above factors govern the retention behaviors of these analytes. Most of them lie in the regions between the dotted ideal and the actual regression lines. Although **16** and **24** possess relatively large substitutents, the rigidity and planarity of the benzene rings apparently render them to have favorable interactions with the IAM phase which is an interesting observation worth pursuing further by some other experimental or computational means.

3.1.4. Electrostatic potential (ESP) surfaces

The calculated electrostatic potential (ESP) surfaces (Figs. 1 and 4) for the IAM.PC.DD2 phase and selected analytes using Gaussian 03 [22] are valuable in interpreting and understanding the interfacial interactions between the analytes and phospholipids in IAM discussed above, especially in some convoluted cases. In IAM.PC.DD2, the highly electronegative glycerol backbone and phosphate ester oxygen atoms (red regions in Fig. 1b) have a strong attractive potential to the electropositive (blue) regions of the nucleobases, provided that the size, shape, and orientation are compatible. For illustration, consider analytes **33**, **34**, and **35** (Fig. 4a–c), the amino substituents

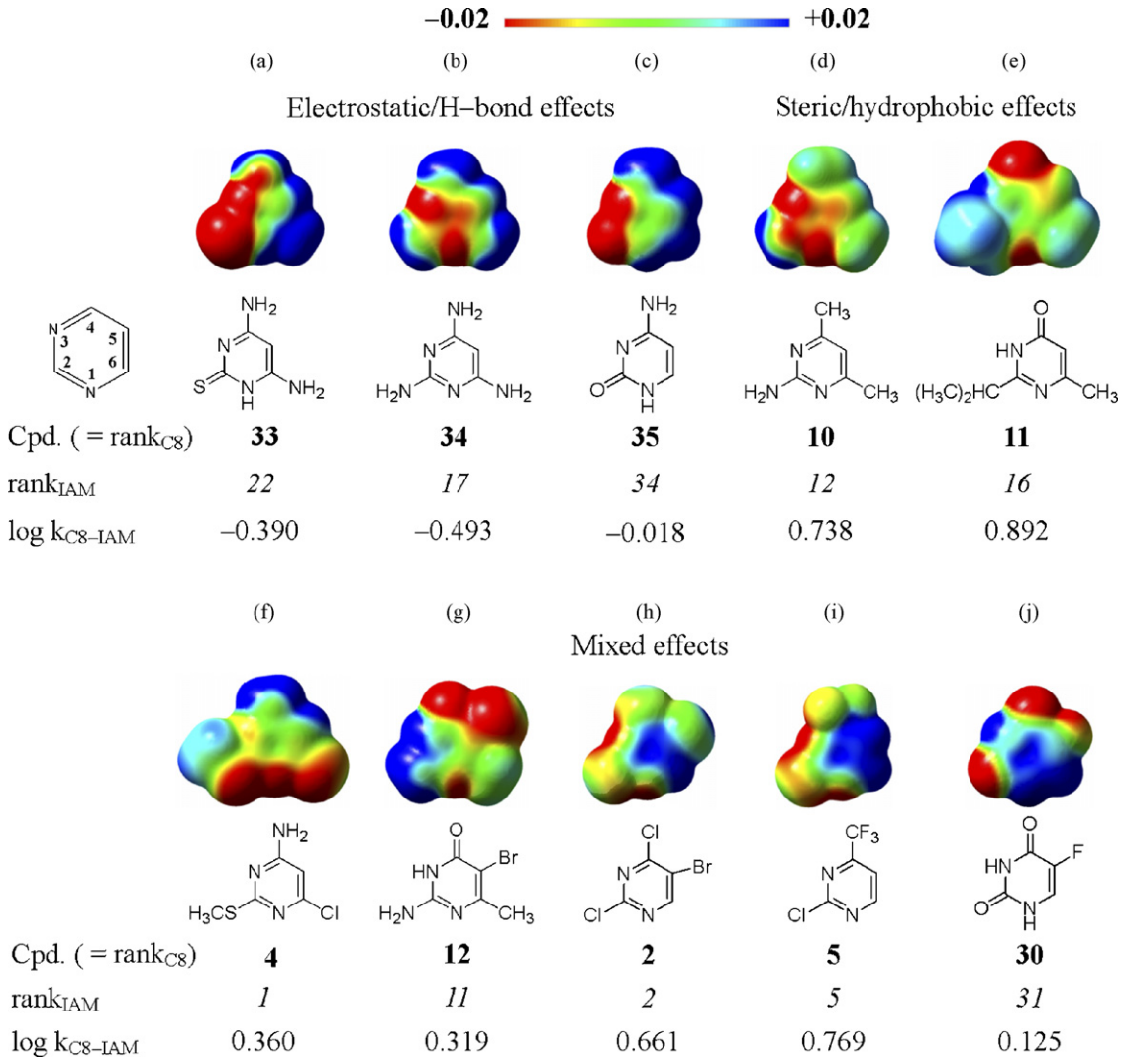

Fig. 4. Calculated electrostatic potential (ESP) surfaces for ten pyrimidines at the DFT B3LYP/6–31G(*d*) level using Gaussian 03 [22]. The color-coded ESP is expressed as the potential energy felt by a probe of positive unit charge at a particular point in space. The ESP was mapped on an isosurface of the total SCF electronic density at a value of 0.0004 au. The color bar on the top ranges from red (electronegative) through green to blue (electropositive) with the mapped ESP values from −0.02 to +0.02 hartrees linearly. Please refer to the web-version for colors.

(electropositive blue regions) are strong H-bond donors. The electronegative potential at the $S_{(2)}$ or $O_{(2)}$ (the subscript for atom numbering in nucleobase is bracketed to avoid confusion in the main text) position of **33** or **35** may orient the analyte to interact with the outermost electropositive ammonium groups of phospholipids temporarily (since partition experiment is based on dynamic equilibrium). In contrast, the H-bond and electrostatic interactions between these three analytes and the C_8 phase were negligible. Thus, they gave longer retention on the IAM column than on the C_8 counterpart.

The methyl and isopropyl groups in analytes **10** (Fig. 4d) and **11** (Fig. 4e) are mainly non-polar (pale blue and green regions), which decrease electrostatic interactions and increase steric repulsion with phospholipids. That is why the retention data of both clearly lie above the regression line (Fig. 3c).

For analytes **4** and **12** (Fig. 4f and g, respectively), the coexistence of the comparably significant H-bond (blue), steric/hydrophobic (green), and electronegative (red) regions are observed. None of the interaction fields serves as the dominant driving force and hence their retention patterns are attributed to the mixed effects. Intuitively, the F, Cl, and Br functional substitutions may serve as electron-withdrawing groups. However, in terms of the ESP surfaces, the Cl, Br, and CF_3 substituents in analytes **2** (Fig. 4h) and **5** (Fig. 4i) are surrounded by the green/yellow-mixed regions which imply more or less hydrophobic properties like methyl group. Interestingly, the $F_{(5)}$ and $Br_{(5)}$ groups in analytes **30** (Fig. 4j) and **12** (Fig. 4g) are in the red regions indicative of the electron-withdrawing properties possibly due to the neighboring carbonyl $C_{(6)}$=O groups. To confirm our conjecture, the same procedures were employed to generate the ESP surfaces of 5-bromopyrimidine, 5-fluoropyrimidine, and 5-bromouracil (see Appendix C). The $F_{(5)}$ and $Br_{(5)}$ groups of 5-fluoropyrimidine and 5-bromopyrimidine are found to be located in the green areas, whereas the $Br_{(5)}$ group of 5-bromouracil is in the vicinity of the red/green-mixed regions similar to **30** as expected. Therefore, one should be cau-

Table 1
Summary of the partial-least-squares analyses on the training set[a]

Column	Model	R^2	q^2	PC	s	Relative contributions to the models (%)				
						S	E	H	A	D
C_8	CoMFA	0.269	0.075	1	0.632	34.9	65.1	–	–	–
	CoMSIA									
	S + E	0.502	0.102	2	0.532	24.8	75.2	–	–	–
	H	0.815	0.468	4	0.337	–	–	100.0	–	–
	S + E + H	0.920	0.598	4	0.222	23.5	28.8	47.7	–	–
IAM	CoMFA	0.339	0.183	1	0.442	31.0	69.0	–	–	–
	CoMSIA									
	S + E	0.296	0.180	1	0.437	9.9	90.1	–	–	–
	H	0.893	0.554	5	0.185	–	–	100.0	–	–
	S + E + H	0.872	0.661	3	0.194	18.8	29.7	51.5	–	–
	All-field	0.974	0.593	6	0.093	13.1	19.0	35.0	12.4	20.6

S: steric, E: electrostatic, H: hydrophobic, A: H-bond acceptor, D: H-bond donor, and all-field: (S + E + A + D + H). *R*, *q*, PC, and *s* are the conventional Pearson regression coefficient, cross-validated regression coefficient, optimal number of principal components, and standard error of estimate, respectively.
[a] The maximum number of PCs included in our cross-validated runs was set to six for consistency.

tious in evaluating the effects of halogen substituents on the retention patterns of pyrimidines in IAM phase.

Analytes might penetrate more into the single-chain IAM.PC.DD1 phase without the glycerol backbone motif found otherwise in natural phospholipids (Appendix D) based only on the spatial consideration. However, lacking the glycerol backbone motif in IAM.PC.DD1 weakens the biological relevant interactions between the oxygen atoms in the glycerol backbone and the analytes, hence the IAM.PC.DD1 phase would not be able to give the more correct mechanistic retention as the IAM.PC.DD2 counterpart in the biological context.

3.2. Chemometric analysis

Herein, some representative CoMFA/CoMSIA models based on the C_8/IAM.PC.DD2 retention data were modelled with the training set consisting of 29 pyrimidines (Fig. 2). Statistical results relating molecular structures to C_8/IAM retention data are summarized in Table 1.

3.2.1. CoMFA

CoMFA PLS analyses resulted in two statistically poor models ($q^2 = 0.075$ and 0.183; $R^2 = 0.269$ and 0.339, respectively) based on the C_8/IAM.PC.DD2 retention data. The steric and electrostatic fields in the CoMFA models were inadequate in interpreting the retention behaviours of the training set in light of the small pyrimidine ring highly affected by bulky ring substitutions. However, for over thirty purines on the C_8/C_{18} columns, the CoMFA models yielded acceptable q^2 values larger than 0.5 [13,14].

3.2.2. CoMSIA

Using the combined steric, electrostatic, and hydrophobic (S + E + H) fields, the derived C_8/IAM CoMSIA models generated significantly improved acceptable q^2 values (0.598 and 0.661, respectively) with the hydrophobic contributions at 47.7% and 51.5%, respectively. Removal either of the steric or the electrostatic field from the models yielded moderate correlations ($q^2 = 0.468$ and 0.554, respectively). In light of negligible H-bond interactions between the analytes and the residual silanol groups on the C_8 column, only the IAM retention data were subjected to the CoMSIA studies for the all-field (S + E + A + D + H) case and a statistically reasonable model ($q^2 = 0.593$; $R^2 = 0.974$) was obtained.

To evaluate the effect of the protonation of **34** at pH 7.3 on the CoMSIA models, its most favorable protonated form (Appendix

Table 2
Predictive evaluation of the derived CoMFA/CoMSIA models on the test set

Column	Model	R^2	q^2	k	b
C_8	CoMFA	0.159	−0.860	0.161	0.071
	CoMSIA				
	H	0.870	0.776	1.215	−0.011
	S + E + H	0.977	0.950	0.903	0.036
IAM	CoMFA	0.329	−0.183	0.340	−0.200
	CoMSIA				
	H	0.600	0.364	1.259	0.064
	S + E + H	0.901	0.753	1.001	0.047
	All-field	0.811	0.601	1.019	−0.010

R, *q*, *k*, and *b* are the conventional Pearson regression coefficient, cross-validated regression coefficient, slope of linear regression, and *Y*-intercept of linear regression, respectively.

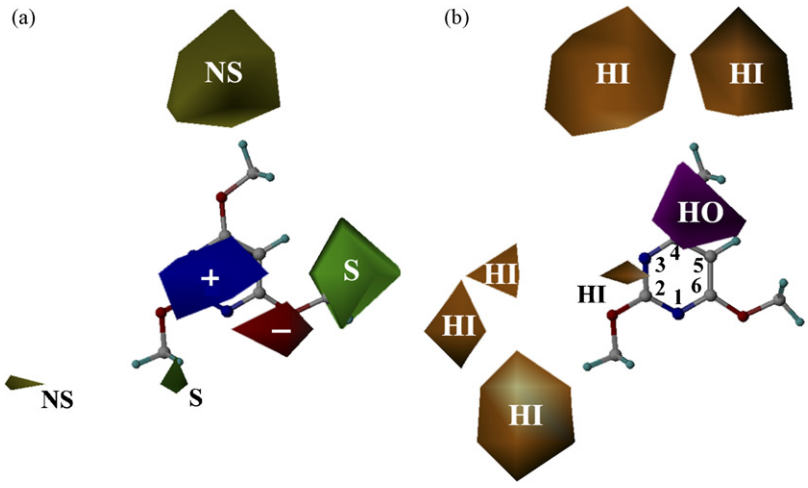

Fig. 5. Contour maps of the CoMSIA models with the combined steric, electrostatic, and hydrophobic fields based on the IAM retention data. Contour plots elucidate the steric and electrostatic features in (a) and hydrophobic features in (b). In (a), green (labeled as "**S**", ≥70% portion) and yellow ("**NS**", ≤8% portion) contours refer to the sterically favored and disfavored areas (w.r.t. longer retention), respectively, whereas blue ("+", ≥70% portion) and red ("−", ≤ 30% portion) contours refer to regions where electropositive substituents are favorable and unfavorable, respectively. In (b), magenta ("**HO**", ≥70% portion) or orange ("**HI**", ≤2.5% portion) contours indicate regions where hydrophobic or hydrophilic groups are favorable. The bracketed percentages are adjustable inequality criteria used in displaying the contours for clarity. Details can be found in [19]. Please refer to the web-version for the alternative color interpretation.

A) along with the other 28 training set nucleobases unchanged were subjected to another IAM CoMSIA analysis. The resulting cross-validated q^2 values for the H, combined S + E + H, and all-field models are 0.592, 0.579, and 0.533, respectively. The same statistical procedure was also repeated for an analysis without **34**, and the relevant q^2 values (0.589, 0.581, and 0.523) are still acceptable. Comparing with the original PLS results (Table 1), the unionized form of **34** seems to increase the performance of IAM CoMSIA models which include non-hydrophobic fields. Therefore, one should be cautious in assessing the influence of the protonated states of pyrimidines with pK_a in the vicinity of the experimental pH.

3.2.3. Cross-validation

Leave-four-out cross-validation was performed to estimate the extent of chance correlation in the IAM CoMSIA models. The means of q^2 values from 30 runs (Appendix E) indicate no significant difference between the LOO and the leave-four-out cases. Moreover, none of the Y-randomization tests [20,21] led to a different conclusion (Appendix E), which further corroborated that the IAM CoMSIA models were not established by chance correlation.

3.2.4. Stability of the CoMSIA models to large substituents or multiple conformations

Analytes **13**, **16**, **17**, **24**, and **25** (Fig. 2) were removed from the training set so as to assess the instabilities of the IAM CoMSIA models to bulky or elongated ring substitutions capable of multiple conformations. Similar statistical results were yielded for the S + E + H and all-field models based on IAM retention data (q^2 = 0.661 and 0.602, respectively), whereas significant improvement was observed for the C_8 CoMSIA/S + E + H model (q^2 = 0.707). These results revealed that the CoMSIA/S + E + H model for the C_8 column was more sensitive to the bulky or elongated substitutions than that for the IAM phase, once more argue for using IAM in assessing retention of analytes in biological contexts.

3.2.5. Prediction

By visual inspection, analytes **7**, **10**, **12**, **15**, **32**, and **33** (Fig. 2) comprising a representative subset were selected as a pseudo external test set to minimally evaluate the predictive capability of the CoMFA/CoMSIA models. The recommended criteria for evaluating the predictive power of a resulting model is $R^2 > 0.6$, $q^2 > 0.5$, and $0.85 < k$ (slope of regression line) < 1.15 [20,21]. The test set performed poorly using the CoMFA models as expected (Table 2) compared with the purine case [13] since pyrimidines are considerably more polar molecularly and rotationally flexible. However, the CoMSIA/S + E + H and CoMSIA/all-field models performed reasonably, and the PLS regression coefficients (q^2 and R^2) and slopes of the regression lines satisfied the recommended criteria [20,21].

3.2.6. Contour plots

The 3D contour plots of the CoMSIA model with the combined steric, electrostatic, and hydrophobic fields based on the IAM retention data are shown in Fig. 5. The contour plots aided in the optimal molecular design of nucleobases in transporting through biomembranes. Bulky and non-polar functional groups close to the green (labeled as "**S**") and magenta ("**HO**") areas (ring positions 3, 4, and 5) will enhance retention, that is why an ascending (slower elution) retention order on the C_8/IAM columns is observed for a sub-congeneric series (**32** < **26** < **15**). According to the GH calculations, the $S_{(4)}$ atom (−0.311 |*e*|) of **23** located in the blue ("+") areas is less electronegative than the $O_{(4)}$ atom (−0.412 |*e*|) of **27**, thus **23** eluted after **27** irrespective of the columns used.

4. Conclusion

For pyrimidines and their derivatives as revealed by this study, the chemical nature of exocyclic substituents generally governed the retention behaviors in the IAM chromatographic system. Due to the smaller core, the exocyclic substituents in pyrimidines influence more significantly the biointerfacial interactions with the IAM phase than those in purines. The calculated electronic properties (e.g. ESP surfaces) are important in a more exact and correct understanding towards the interfacial partition properties and retention patterns of the pyrimidines in biomembrane environments. Though successful in the interpretation of purines, the steric and electrostatic molecular fields in the CoMFA models were defective to model the retention behaviors of pyrimidines. The more elaborate CoMSIA models, mainly attributed to the hydrophobic fields, offset the inadequate performance of the CoMFA ones and were revealed as a reasonable tool in modeling the biointerfacial interactions between pyrimidines and IAM.

Acknowledgements

We cordially acknowledge the HKBU Faculty Research Grant (FRG00-01/II-72) and RGC CERG Grant (HKBU 2035/02P) making this work possible.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2007.02.059.

References

[1] (a) D. Bowen, D.H. Johnson, W.M. Southerl, D.E. Hughes, M. Hawkins, Anticancer Res. 19 (1999) 3837;
(b) W. Saenger, Principles of Nucleic Acid Structure, Springer-Verlag, New York, 1984, chapter 7.
[2] P.R. Brown, E. Grushka, Anal. Chem. 52 (1980) 1210.
[3] M. Zakaria, P.R. Brown, J. Chromatogr. 255 (1983) 151.
[4] P.R. Brown, HPLC in Nucleic Acid Research: Methods, Applications, Marcel Dekker, New York, 1984.
[5] P.R. Brown, C.S. Robb, S.E. Geldart, J. Chromatogr. A 965 (2002) 163.
[6] A. Rizzi, H.R.M. Lang, J. Chromatogr. 331 (1985) 33.
[7] A. Taillardat-Bertschinger, P.-A. Carrupt, F. Barbato, B. Testa, J. Med. Chem. 46 (2003) 655.
[8] S.W. Ong, H.L. Liu, X.X. Qiu, G. Bhat, C. Pidgeon, Anal. Chem. 67 (1995) 755.
[9] G.W. Caldwell, J.A. Masucci, M. Evangelisto, R. White, J. Chromatogr. A 800 (1998) 161.
[10] A. Taillardat-Bertschinger, F. Barbato, M.T. Quercia, P.-A. Carrupt, M. Reist, M.I.La. Rotonda, B. Testa, Helv. Chim. Acta 85 (2002) 519.
[11] E. Lazaro, C.R. Rafols, M. Roses, J. Chromatogr. A 1081 (2005) 163.
[12] E.C.Y. Chan, W.L. Tan, P.C. Ho, L.J. Fang, J. Chromatogr. A 1072 (2005) 159.
[13] H. Luo, Y.-K. Cheng, QSAR Comb. Sci. 24 (2005) 968.
[14] H. Luo, C. Zheng, Y.-K. Cheng, J. Chromatogr. B 847 (2007) 245.
[15] R.D. Cramer III, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959.
[16] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, J. Chem. Inf. Comput. Sci. 29 (1989) 163.
[17] G. Klebe, U. Abraham, J. Comput. -Aided Mol. Des. 13 (1999) 1.
[18] H. Luo, Y.-K. Cheng, J. Chromatogr. A 1103 (2006) 356.
[19] Tripos Inc., Tripos Sybyl 6.9.2 for Linux, St. Louis, MO, 2003.
[20] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb. Sci. 22 (2003) 69.
[21] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269.
[22] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, T. Vreven Jr., K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03 (C.02), Gaussian Inc., Wallingford CT, 2004.