

Available online at www.sciencedirect.com



JOURNAL OF CHROMATOGRAPHY B

Journal of Chromatography B, 847 (2007) 245-261

www.elsevier.com/locate/chromb

# The retention properties of nucleobases in alkyl $C_8$ -/ $C_{18}$ - and IAM-chromatographic systems in relation to log $P_{ow}$

Haibin Luo, Chuanqi Zheng, Yuen-Kit Cheng\*

Department of Chemistry, Hong Kong Baptist University, Kowloon Tong, Hong Kong Received 19 July 2006; accepted 9 October 2006 Available online 27 October 2006

#### Abstract

In order to evaluate the differences in the partition properties of 35 structurally congeneric nucleobases of biological interests in octanol–water biphasic, alkyl  $C_8/C_{18}$ , and IAM systems, a comparative chromatographic study was performed. Comparing with the reversed-phase  $C_8/C_{18}$  retention data, most of the purines possessed weaker IAM retention except for those with specific H-bond and/or electrostatic interactions. Quantitative correlations between the experimental log  $P_{ow}$  literature values and the IAM,  $C_8$ , and  $C_{18}$  log *k* were evaluated ( $R^2 = 0.943$ , 0.794, and 0.767, respectively). Although IAM retention correlated significantly better (larger  $R^2$  value) with the log  $P_{ow}$  values statistically, the latter was revealed apparently behaving more like (slope approaching unity) alkyl  $C_8/C_{18}$  retention and hence also has the same shortcoming in under-representing analytes capable of forming short-term H-bond/electrostatic interactions with polar head-groups of phospholipids. A chemically meaningful structure-retention model ( $q^2 = 0.824$  and  $R^2 = 0.968$ ) was derived, in which the hydrophobic interaction is identified as the underlying factor for the retention of purines in IAM system modulated non-trivially by H-bond/electrostatic interactions.

Keywords: Nucleobases; Retention; Partition; IAM; C8/C18; Octanol-water log P

#### 1. Introduction

Native and derivatized nucleic-acid bases (nucleobases) are readily found as drug motifs. They have rather complicated and subtle protonation/deprotonation states unlike most of the small organic drug molecules which usually have straightforward mono-basic/mono-acidic properties. Corollarily, the transport properties of nucleobases are not readily deducible from the available physico-chemical properties based on other small organic drug molecules. Recent research efforts mainly focused on the facilitative diffusion of nucleobases, but there exists documented evidence of unmediated passive transport of nucleobases and nucleosides which should be particularly important for the apolar or weakly polar synthetic analogs yet without identified transporters. Even in the facilitative case, the solubility or lipophilicity of a nucleobase in biomembrane is a pre-requisite prior to the passage aided by a transporter.

1570-0232/\$ - see front matter © 2006 Elsevier B.V. All rights reserved. doi:10.1016/j.jchromb.2006.10.009

In general, drug transport via the gastrointestinal system is an intricate active/passive transcellular or paracellular process. Nevertheless, the knowledge of unmediated transports of small drug molecules passively through biomembrane is still indispensable per se, not only because of the voluminous partition data already being accumulated hitherto, but also of their indispensable role in constructing the full picture of numerous drug absorption processes, especially for the transcellular cases. Therefore, the affinity of a small drug candidate to a lipophilic environment is clearly instrumental in its ultimate pharmacological and therapeutic values. Both the mobility of molecules within the lipid (diffusivity) and solubility in the membrane together determine the total unmediated passive transport as a rule. For drug molecules of similar sizes, the former factor level off to a very good approximation leading to the remaining dominant role of solubility (often expressed in terms of partition coefficients and recently as chromatographic retention) in unmediated passive drug transport.

Traditionally, octanol-water partition coefficient (often expressed as log P, or log D for pH-dependent cases) was adopted as a convenient lipophilicity parameter in investigating and predicting the passive transport properties of oral drug

<sup>\*</sup> Corresponding author. Tel.: +852 3411 7066; fax: +852 3411 7348. *E-mail address:* ykcheng@hkbu.edu.hk (Y.-K. Cheng).

compounds [1–5]. Even the ADME (absorption, distribution, metabolism, and excretion) of many drug series have been shown qualitatively correlated well with  $\log P/\log D$  [4]. However,  $\log$ P/log D has unavoidable shortcoming in fully rationalizing ionizable drug compounds in biomembrane-water systems [2,3], due to the inadequacy of the octanol-water biphasic system in emulating the intrinsically more complex interactions between biomembranes and ionizable molecules [2]. Accordingly, several phospholipid-modified biomembrane-like systems based on the major eukaryotic phospholipid phosphatidylcholine (PC) have been developed for the remedy. Immobilized-artificialmembrane (IAM) stationary phase [3,6–8], which is prepared by phospholipids covalently bonded to a propylamino-silica support, has been popular for its high-throughput screening potential [3], and is useful as a physico-chemical model of drug-membrane partitioning similarly as liposome membranes [6–11]. The logarithmic capacity factor using IAM (log  $k_{\text{IAM}}$ ) has been correlated well with log  $P_{ow}$  statistically and some other bioactivities [8–11]. As a result,  $\log k_{\text{IAM}}$  has been recently treated as a convenient reference parameter for predicting the lipophilic properties of small drug molecules. However, the conceptual basis for the similarity (or difference) has not been investigated which will definitely impede the general use of this relatively low-cost convenient approach.

Most IAM retention data reported so far were obtained using stationary phases containing the single-chain IAM.PC.DD, or the double-chain IAM.PC.DD2 and IAM.PC.MG (Fig. 1) [3]. Residual amino groups are end-capped with methylglycolate (MG) in IAM.PC.MG, but with  $C_{10}$  and  $C_3$  alkyl chains in IAM.PC.DD2. Comparing with the IAM.PC.DD2 phase, singlechain IAM.PC.DD lacks the glycerol backbone motif found in the natural phospholipids (Fig. 1). Barbato and co-workers observed that IAM.PC.MG retention data are better predictors of the interactions between drugs and biomembranes than IAM.PC.DD and suggested that a biologically representative phospholipid stationary phase in IAM was crucial to obtain reliable lipophilicity parameters [2]. Some other studies demonstrated that IAM.PC.DD2 was more hydrophobic than the single-chain IAM.PC.DD counterpart [8] and a comparison of IAM data for 68 compounds also indicated better correlation for the two double-chain columns [8b]. A case in point, the comparison of the intermolecular interactions between 6thioguanine and IAM.PC.DD2 or IAM.PC.DD phases were proposed here and schematized as in Fig. 2. The single-chain phase lacking the glycerol moiety, would not be able to give the corresponding mechanistic retention as in IAM.PC.DD2 which encodes the more correct interactions between ionizable (usually through protonation/deprotonation) analytes and the polar head-groups of phospholipids besides the extra hydrophobic chains.

Owing to the great pharmacological potential of chemically modified nucleobases, quantitative structure-retention relationship has been developed to interpret the relationship relating molecular structures to chromatographic retention of purine bases, but mainly on the alkyl phases [11–13]. Our previous chromatographic study resulted in a meaningful structure-retention correlation on a  $C_{18}$  column [13].



Fig. 1. The schematic structures of IAM.DD.MG, IAM.PC.DD, IAM.PC.DD2, C<sub>8</sub>, and C<sub>18</sub> stationary phases.

In the present work, a comparative study of the IAM.PC.DD2/ C<sub>8</sub>/C<sub>18</sub>-HPLC retention data of a set of 35 structurally congeneric purines (see Appendix A for the chemical structures) was performed (the aqueous component of the mobile phase buffered at the physiological pH). The experimental log  $P_{ow}$  values of purines retrieved from the literature were regressed with the log  $k_{IAM}/\log k_{C18}/\log k_{C18}$  values. 3D chemometric approaches were also employed in order to interpret and explain the retention



Fig. 2. The schematic drawings illustrating the essential difference of the possible (favorable) intermolecular interactions between 6-thioguanine and IAM.PC.DD2 (a) or IAM.PC.DD (b) stationary phase temporally. The compound will interact with the stationary phase in other configurations (including the hydrophobic aliphatic chains down below the polar head-groups because of the dynamic nature of the equilibrium partitioning). The CPK renderings on the LHS are not to scale (C = cyan; H = gray; N = blue; O = red; and S = yellow). Refer to the online version for the colors.

mechanism with respect to the molecular structures. We aim at dissecting the differences in the retention mechanism of nucleobases between alkyl and double-chain IAM.PC.DD2 phases molecularly in correlation with their partition properties.

#### 2. Experimental

#### 2.1. Instrumental and chemicals

A Waters 2695 HPLC system with a 996 photodiodearray detector set to 254 nm for detection was used. Commercially available Alltech Hypersil MOS-2 C<sub>8</sub> (150 mm × 4.6 mm,  $5 \,\mu$ m), Waters Xterra<sup>TM</sup> MS C<sub>18</sub> (150 × 4.6 mm, 5  $\mu$ m), and Regis IAM.PC.DD2 (150 × 4.6 mm, 12  $\mu$ m, Morton Grove, IL, USA) columns, were used for all the HPLC experiments. All analytes were dissolved at ~0.2 mg/ml in an HPLC-grade methanol solution using 1,3-dimethyl-5-fluorouracil (Sigma) as the internal standard. The column temperature and flow rate were kept at 25 °C and 1.0 ml/min, respectively. An ORION 720A pH meter was used for the pH determinations. Mobile phases used for the IAM-HPLC experiments were 10:90, 15:85, and 20:80 (% by v/v) methanol/sodium phosphate buffer (35 mM, pH ~7.3), whereas those for the C<sub>8</sub>-HPLC experiments were 20:80, 30:70, and 40:60 (% by v/v), and for the C<sub>18</sub>-HPLC experiments was 20:80 (% by v/v). 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 times for the sample peak and column void volume/time, respectively. The latter was estimated by citric

acid [3,14] measured at 210 nm. Extra-column (dead) volume (0.247 ml) was determined in the absence of column in the Waters 2695 HPLC system.

The retention time was determined one-at-a-time for each analyte isocratically. All the retention times were averages of triplicate or quadruplicate measurements of the same sample solution resulting in relative standard deviation (RSD) less than 3.0% for each analyte. The RSD of the internal standard in each case was smaller than 1%.

The following analytes were purchased from Sigma–Aldrich: 6-benzyloxypurine (1), 6-benzylaminopurine (2), kinetin (3), 2,6-dichloropurine (5), 2-amino-6-methylthiopurine (7), 6ethoxylpurine (8), 6-amine-9-ethylpurine (9), 2-amino-6bromopurine (10), 2-amino-6-chloropurine (11), 6-bromopurine (12), ethyl adenine-9-acetate (13), 6-chloropurine (14), caffeine (15), 6-thioguanine (16), 6-cyanopurine (17), 8chlorotheophylline (18), 6-methylpurine (19), adenine (20),  $N^2$ ,9-diacetylguanine (21), β-hydroxyethyltheophylline (22), 2,6-diaminopurine (23), dyphylline (24), 2-aminopurine (25), theobromine or 3,7-dimethylxanthine (26), purine (27), 3methylxanthine (28), guanine (29), 8-methylxanthine (30), hypoxanthine (31), 6-thioxanthine (32), 2-thiopurine (34), and uric acid (35).

3-Isobutyl-1-methylxanthine (4), 6-dimethylaminopurine (6), and xanthine (33) were obtained from Acros Organics. All chemicals (Appendix A) are of the highest purity available commercially and used without further purification. HPLC-grade methanol was obtained from Lab-Scan Ltd.

### 3. Results and discussion

All the chromatographic experiments were performed under analytical and isocratic conditions, thus complication associated with nonlinear chromatographic retention (critical in preparative separations otherwise) is not a concern here. Typical maximum chromatographic peak width of the studied nucleobases from all the columns used is about 0.15 min (FWHM), which amounts to a maximum precision error in log  $k_{C8}/\log k_{C18}/\log k_{IAM}$  about  $3 \times 10^{-2}$  by simple propagation error estimation. The retention data using the IAM phase are summarized in Table 1 (see Appendix B for the full data set).

### 3.1. Nontrivial substituent effect of IAM.PC.DD2 versus $C_8/C_{18}$ phases

Most of the purines revealed faster elution in the IAM.PC.DD2 column than those by the alkyl  $C_{18}$  and  $C_8$  counterparts under identical chromatographic conditions (20% methanol in sodium phosphate buffer at pH 7.3) albeit with moderate univariate linear correlations ( $R^2 = 0.850$  and 0.865, respectively) as shown in Fig. 3a and d. Assuming the same retention mechanism for the alkyl columns, which is reasonable, the robustness of the  $C_8/C_{18}$  retention data is reflected by the nearly exact correspondence of the retention order in the two plots (see Appendix B for detailed retention data). The polar head-groups in the IAM phase pointing towards the mobile phase partially shield the hydrophobic alkyl chains from the

analytes. Thus, weaken the average overall apolar interaction between the analytes and the stationary phase considerably, resulted in faster elution than that by the alkyl ones for analytes retained more by hydrophobicity (provided that the Donnan effect caused by the residual silanol groups were suitably masked by the sodium phosphate added). The exceptions were those compounds (Fig. 3f: 16, 21, 23, 25, 31-34) with retention falling nearly on or below the ideal (dotted) lines (slope = 1and intercepts = 0) with potential H-bond and/or electrostatically interacting substituents, such as exocyclic -NH2/-NHR and =S/=O groups at positions 2 and 6, respectively. These groups are capable of interacting with the polar PC head-groups spatially but most likely temporally because of the dynamic nature of partition equilibrium and the flexibility of the PC chains. The analytes are not static and still likely to interact in various configurations with the stationary phase including situation deep below the polar head-groups which interact with the hydrophobic aliphatic chains. 6-Thioguanine (16) as suggested in Fig. 2a can adopt a specific orientation to interact temporally with the IAM.PC.DD2 via both H-bond at position 1, 2, or 9 (with the phospholipid carbonyl glycerol backbone oxygen or the phosphate ester oxygen atoms) and electrostatic interactions of S at position 6 (with the ammonium group of IAM.PC.DD2). The GH (Gasteiger-Hückel) partial atomic charge of the sulphur atom is -0.322 |e| (Appendix C). As a result, 6-thioguanine gave a stronger retention on the IAM column (2.941 min) than on the  $C_8$  (2.391 min) and  $C_{18}$  (2.465 min) counterparts. Simply by changing the  $S_{(6)}$  (the subscript for atomic numbering is bracketed to avoid confusion) atom to O<sub>(6)</sub> resulted in analyte 29. The IAM retention dropped considerably (Fig. 3c and f) due to the compromise of the favorable specific interactions caused by the shortening of the  $C_{(6)} = O_{(6)}$  bond and the concomitant increase in the negative partial charge (-0.408 |e|) on O<sub>(6)</sub> possibly induce repulsion with the negatively-charged phosphate moiety. This biointerfacial effect on purines will not be readily reflected in the log  $P_{ow}$  values or  $C_8/C_{18}$  retention data. Clearly, H-bond and electrostatic interactions demonstrated an important influence on the retention behaviors of these purines on IAM in a complicated manner superimposing on the baseline hydrophobic retention as found in  $C_8/C_{18}$ .

For xanthine (**33**), almost the same experimental retention times were observed on the IAM and  $C_8$  columns at identical chromatographic condition (Fig. 3f). Based on a structural point of view, the steric occupancy at the ring position  $N_{(1)}$  between the two keto groups of xanthine might attenuate or even disrupt the specific H-bond and electrostatic interactions with IAM. Indeed, hydrophobic  $N_{(1)}$ -substituted methyl derivatives of xanthine (Fig. 3e: **4**, **15**, **18**, **22**, **24**) indicated much longer retention on the  $C_8$  column than on the IAM.PC.DD2 counterpart with data points considerably lying above the ideal line.

Interestingly, analyte **1** (6-benzyloxypurine) had a rather long retention time of 127.533 min on the C<sub>8</sub> column, but merely 21.644 min on the IAM counterpart which was only slightly longer than that of analyte **2** (6-benzylaminepurine, 17.310 min). Once again, this implies biomembranes possess non-trivial retention mechanism, at least for nucleobases, different from that of the alkyl-based stationary phases.

The IAM_HPLC r	retention data of	the nuri	es and their	r available lo	ng P	literature values
THE PAINT-TH LC I	ciciliuon uata or	une purn	ies and then		JE I OW	incrature values

Compound 1		2	3	6	7	8	9	$\log k_{\text{IAM}}$		$\log P_{\rm ow}$
								10% MeOH	20% MeOH	
1		—н		-O-CH <sub>2</sub>		—H	—Н	1.366	0.995	
2		—Н		-NH-CH <sub>2</sub>		—H	—Н	1.263	0.887	1.57
3	CU	-H		-NH-CH <sub>2</sub>		H	H	0.795	0.439	1.00
4	$-CH_3$	=0	<i>i</i> -butyl	=0		-H	-H	0.625	0.313	1.29
5		-Cl		-Cl		-H	-H	0.451	0.122	
6		H		$-N(CH_3)_2$		—н	-H	0.340	0.011	
7		$-NH_2$		-SCH <sub>3</sub>		-H	-H	0.276	0.040	
8		—н		$-OCH_2CH_3$		-H	-H	0.175	-0.056	
9		-H		$-NH_2$		-н	$-C_2H_5$	0.126	-0.236	
10		-NH <sub>2</sub>		-Br		-H	-H	0.095	-0.139	
11		$-NH_2$		-CI		-H	-H	0.035	-0.235	
12		-H		-BI		-H		0.022	-0.242	
13		H		-NH <sub>2</sub>		-H	$-CH_2CO_2C_2H_5$	0.015	-0.280	
14	CU	-H	CU	-0	CU	-H	-н	-0.041	-0.319	0.07
15	-сп3	=0	-Сп3		-сп3	<u>–</u> п		-0.034	-0.320	-0.07
10	-п					-п	-п	-0.150	-0.520	-0.07
1/	CU	—п —0	CU	-CN			-п 11	-0.135	-0.445	0.95
10	-СП3	_U	-СП3	0 CU		-01	-n 11	-0.180	-0.082	-0.85
19				-CH3 NH.		U	-п ц	-0.200	-0.491	0.00
20	U			-N112 		U		-0.209	-0.510	-0.09
21			CH.	=0	C.H.OH	_H	-00013	-0.209	-0.490	
22	CH3	-O -NHa	-0113	-NH2	-02114011	_H	_н	-0.207	-0.557	
23	-CH	=0	-CH	=1(112)		_H	-11	-0.346	-0.654	
25	CH3	-O -NH2	-0113	H	-c2113011c112011	_H	_Н	-0.330 -0.423	-0.634	
25	—н	=0	$-CH_2$	=0	-CH2	_н	11	-0.433	-0.654	-0.78
20		—н	CIIS	_H	city	—н	—Н	-0.453	-0.671	-0.37
28	—н	=0	$-CH_2$	=0		—н	_H	-0.481	-0.695	-0.72
29	—н	-NH2	eny	=0		—Н	_H	-0.525	-0.692	-0.96
30	—н	=0	—Н	=0		-CH <sub>2</sub>	-H	-0.527	-0.740	0.70
31	—н	-H		=0		-н	-H	-0.702	-0.848	-1.11
32	—н	=0	-H	=5		—н	-H	-0.704	-0.918	
33	-н	=0	-H	=0		—Н	-H	-0.782	-0.935	-0.73
34	—н	=S	-H	_H		—Н	-H	-0.852	-0.998	0.70
		~						1.400	1 (75	0.17

<sup>a</sup> See Appendix D for more information.

#### 3.2. Solvent effect

Table 1

1

-<u>N</u>

Mobile phases used for the IAM-HPLC experiments are consisting of 20:80, 15:85, or 10:90 (% by v/v) methanol/sodium phosphate buffer (35 mM, pH 7.3). The decrease in the methanol content generally increases the overall polarity of the mobile phase. As a result, all the analytes were retained significantly longer by the IAM column in 10% methanol than those in 20% methanol (Table 1) though these two conditions possess a near linear retention relationship (log  $k_{20\%}$  MeOH = 0.910 × log  $k_{10\%}$  MeOH -0.269,  $R^2$  = 0.989, and s = 0.056) in support of essentially the same retention mechanism still in operation. Therefore, the retention data in 10% methanol were used in the subsequent chemometric calculation for enhanced precision.

#### 3.3. Correlation between log $P_{ow}$ and log $k_{IAM}$

For the 13 purines with available experimental log  $P_{ow}$  values (Table 1), quantitative correlations ( $R^2 = 0.943$  and 0.916) with the log  $k_{IAM}$  determined in IAM-HPLC have been observed (Table 2, Fig. 4a and b). As a result, log  $k_{IAM}$  readily determined consistently by IAM-HPLC can still act as a robust indicator of the classical "lipophilic" properties determined by log  $P_{ow}$  for purines associated with biomembrane-like systems. The log  $P_{ow}$  values of the other purines based on this regression model, may not be easily obtained experimentally, can be readily predicted (e.g., Appendix E).

Of particular interests here, the regression line distinguishes two groups at the physiological pH 7.3 according to their



Fig. 3. Correlations of various retention data (log  $k_{C8}/\log k_{C18}/\log k_{IAM}$ ). The solid lines are the linear regression curves and the dotted straight lines correspond to the ideal correlation case (slope = 1 and intercepts = 0) for reference.  $R^2$  = Pearson correlation coefficient; k = slope of regression; and b = y-intercept of regression.

Table 2 Correlations between log  $P_{ow}$  and experimental log k for the selected purines given in Table 1

	C <sub>8</sub>			IAM				
	20% MeOH	30% MeOH	40% MeOH	10% MeOH	15% MeOH	20% MeOH		
$\overline{R^2}$	0.794	0.805	0.814	0.916	0.924	0.943		
k	1.060	1.242	1.573	1.412	1.608	1.558		
b	-0.227	0.167	0.751	0.005	0.193	0.426		
s	0.466	0.453	0.443	0.297	0.282	0.244		
F	42	45	48	120	134	183		

R, k, b, s, and F are the regression coefficient, slope of regression line, intercept of regression line, standard deviation of residuals, and F-test values, respectively.



Fig. 4. Correlations of the classical "lipophilic" parameter log  $P_{ow}$  to various log k's. The solid lines are the linear regression curves and the dotted straight lines correspond to the ideal correlation case (slope = 1 and intercepts = 0) for reference.  $R^2$  = Pearson correlation coefficient; k = slope of regression; and b = y-intercept of regression.

protonation/deprotonation states and chemical properties (Fig. 4a and b). Except for xanthine (**33**) and uric acid (**35**), the other purine analytes should mainly present in their neutral forms in 20% methanol/sodium phosphate buffer at pH 7.3 (Fig. 4e, RHS) [15]. Analytes **26**, **29**, and **31** fall below the

ideal dotted line; whereas the partially ionized xanthine (**33**) lies above. This appearance of the two groups is even more pronounced in the  $C_8/C_{18}$  phases (Fig. 4c and d). A closer inspection reveals that **26**, **29**, and **31** similarly contain an aromatic keto-amino motif ( $-HN_{(1)}-C_{(6)}O-$ ). With no basic

ring nitrogen but two acidic amino protons, analyte **28** tends to follow the ideal trend. Any specific interactions of **2**, **4**, and **15** with the IAM phase would be shielded or disrupted very effectively by the hydrophobic group at the N<sub>(1)</sub> position or in the presence of more than one methyl/alkyl groups. Their predicted log  $P_{ow}$  values were thus underestimated according to the ideal linear model (Fig. 4a and b). A Cl group at position 8 in analyte **18** seems to offset this effect. Bare or simple-substituted purines (**20**, **27**) are also slightly underestimated according to the ideal model (Fig. 4a and b). This subtle behavior warrants further investigation by simulation study for a molecular explanation.

Comparing with the retention data on the C<sub>8</sub> and C<sub>18</sub> columns (Appendix B) at identical chromatographic conditions, relatively weaker statistical correlations ( $R^2 = 0.794$  and 0.767, respectively) between log *k* and log  $P_{ow}$  were observed (Fig. 4c and d) presumably due to the absence of polar headgroups. Interestingly, however, log  $P_{ow}$  versus log  $k_{C8/C18}$  (20% methanol) has slope approaching unity, explaining why log  $k_{C8/C18}$  have been regarded performing well in estimating log  $P_{ow}$  statistically hitherto. Here we see that the reason is more a chance cancellation (reflected by the low  $R^2$  values) of data above and below the regression line in a congeneric series of analytes (Fig. 4c and d). Higher percentage of methanol in the mobile phase gave slight improvement in the regression  $R^2$  values but deflecting the slope from unity (Table 2). This observation was also found in other studies [16–18].

In summary, contrary to the alkyl-silica phases, IAM.PC. DD2 possesses not only two hydrophobic chains but also a polar head-group. On the other hand, the simple hydroxyl head-group of octanol only captures partially the complexity of the polar head-groups of phosphatidylcholine in IAM.

#### 3.4. Correlation between ClogP and log $k_{IAM}$

The ClogP values of the 35 analytes generated based on the fragmental methods implemented in ChemDraw Ultra (Version 7.0.1) were moderately correlated to IAM log  $k_{20\% \text{ MeOH}}$  ( $R^2 = 0.709$ , slope = 0.421, Fig. 5), which indicated that parametric method such as ClogP, though convenient, is inadequate in predicting the experimental IAM retention.



Fig. 5. Correlation of log  $k_{IAM}$  to ClogP. The solid lines are the linear regression curves and the dotted straight lines correspond to the ideal correlation case (slope = 1 and intercepts = 0) for reference.  $R^2$  = Pearson correlation coefficient; k = slope of regression; and b = y-intercept of regression.

#### 3.5. 3D Chemometric analyses

For a better understanding in terms of spatial features, the IAM retention data were subjected to 3D quantitative structureretention relationships using both comparative molecular field analysis (CoMFA) [19] and comparative molecular similarity indices analysis (CoMSIA) [20]. Chemometric models based on CoMSIA which combined various molecular fields (see Appendix F for the detailed methodology and results) gave reasonable regressed relationships high in cross-validated regression coefficients ( $q^2 \ge 0.768$ ) and conventional regression coefficients ( $R^2 \ge 0.955$ ) in the (Fig. 6a) where the hydrophobic interaction is identified as the underlying factor for the retention of purines.

The corresponding contour plots of the CoMSIA models allowed the correlation of experimental retention data with changes in the steric/electrostatic (Fig. 6b) and hydrophobic (Fig. 6c) molecular-field contributions and aid in the optimal molecular design of nucleobases in transporting through biomembranes. Bulky groups at ring position 6 (analytes 1–3, 6, 8, 19, Appendix A) and methyl groups at position 1 (4, 15,



Fig. 6. The regression results of the chemometric models. In (a), the fitted vs. the experimental log  $k_{IAM}$  for the 35 analytes in 10% methanol are plotted. The dotted straight lines correspond to the ideal correlation case (slope = 1 and intercepts = 0) for reference. For clarity, the data sets were shifted upward 0.5, 1.0, and 1.5 units for the CoMSIA models with the H (hydrophobic), S + E + H (combined steric, electrostatic, and hydrophobic), and all-field (combined steric, electrostatic, hydrophobic, and H-bond donor/acceptor) fields, respectively. In (b), blue (+) and red (-) contours refer to regions where electropositive substitutions are favorable and unfavorable (w.r.t. longer retention), respectively, whereas green (S) and yellow (NS) contours refer to the sterically favored and disfavored areas, respectively. In (c), orange (HO) and purple (HI) contours indicate regions where hydrophobic and hydrophic groups are favorable, respectively. Refer to the online version for colors.

**18**) clearly enhanced the overall retention according to the CoM-SIA models. The extra methyl group of analyte **15** at position 1 brought on a dramatic increase in the retention relative to **26**. However, bulky groups at position 7 decreased the retention (e.g., compare **15**, **22**, and **24**). The hydrophobic methyl group of compound **30** at position 8 close to the purple region (Fig. 6c) resulted in a shorter elution time than that of **28** at position 3.

#### 4. Conclusions

IAM, especially those of the double-chain types, has been revealed in this study as a robust mimic of the biological environment in emulating the solute/membrane interactions. Log  $k_{\text{IAM}}$  furnishes itself as a useful reference lipophilic parameter which is unmatched by the alkyl-silica counterparts and at the same time much readily to be carried out than the more costly liposome/micelle electrokinetic chromatography [21]. This is especially so for the evaluation of compounds capable of forming subtle specific hydrogen-bond or electrostatic interaction at the physiological pH (=7.3 as in the present experimental conditions) such as the nucleobases studied here of which partition data are difficult to be obtained in single experiments consistently. The IAM baseline-retention of the purines is determined by hydrophobic interaction with the hydrophobic chains clearly revealed by the CoMSIA analysis, modulated by specific interactions with polar head-groups which are not readily reflected by the traditional "shake-flask" log  $P_{ow}$  values. For example, xanthine analogs with methyl or alkyl substituents at the  $N_{(1)}$ position will attenuate the IAM-retention consistently more than expected from a classical log  $P_{ow}$  prediction. The current results could serve as a reference for the biointerfacial properties of synthetic nucleobases for pharmaceutical purposes.

#### Acknowledgements

We cordially acknowledge the support from HKBU Faculty Research Grant (FRG00-01/II-72) and RGC CERG Grant (HKBU 2035/02P) making this work possible. The careful reading and the suggestions from an anonymous reviewer are highly appreciated.

Supporting information available: chemical structures of the 35 purine bases, retention data on the various columns, GH partial-atomic charges and geometries of selected purines, literature log  $P_{\rm ow}$  values, plots of the predicted log  $P_{\rm ow}$ , chemometric procedures and analyses, and selected chromatograms on the IAM.PC.DD2 column.

### Appendix A. Chemical structures of the 35 purine bases used in this study



Appendix A	( <i>Continued</i> )
------------	----------------------



#### ۸ div A (Continued)

Appendix A (Continued)		Appendix A (Continued)			
Compound	Structures	Compound	Structures		
Appendix A (Continued)CompoundStructures11 $\begin{array}{c} & & & \\$	20	$NH_2 \\ N \\ N \\ N \\ N \\ N \\ H$			
12	Br N N H Br N H Br N H Br N H Br N H Br N H Br N H Br N H Br N H H Br N H H Br N H H Br N H H Br N H H H H H H H H H H	21	HN HN COCH <sub>3</sub> N <sub>2</sub> ,9-diacet		
13	$N$ $N$ $N$ $N$ $N$ $N$ $N$ $CH_2CO_2Et$ ethyl adenine-9-acetate	22	H <sub>3</sub> C <sub>N</sub> β-hydroxye		
14	CI N N N N N N N N N N	23	H <sub>2</sub> N H <sub>2</sub> N 2,6-diamino		
15	$H_3C_N$ $H_3C_N$ $H_3C_N$ $CH_3$ $CH_3$ $CH_3$ $C$	24	$H_{3C} \cap C_{2} h$		
16	$HN H_2N N H_2N N H_2N N H_2N H_2N H_2N H$	25	$H_2N$ $N$ $N$ $N$ $N$ $N$ $N$ $N$ $N$ $N$		
17	CN $N$ $N$ $N$ $N$ $N$ $N$ $N$ $N$ $N$	26	HN HN HN HN HOP		
18	$H_{3}C_{N} \xrightarrow{N} K_{N} CI$	27	N N N N N N H purine		
19	$\begin{array}{c} \text{CH}_3\\ \text{N} \xrightarrow{\text{CH}_3}\\ \text{N} \xrightarrow{\text{N}}\\ \text{N} \xrightarrow{\text{N}}\\ \text{H} \end{array}$ 6-methylpurine	28	HN O HN O $CH_3$ $H_3$		

COCH<sub>3</sub> tylguanine Ç₂H₄OH Ń ĊH₃ ethyltheophylline 1<sub>2</sub> K N N H opurine H₃OHCH₂OH lline N N H rine



ne



inthine



## Appendix B. Retention data $(\log k)$ on the various columns<sup>a</sup>

Column	IAM	IAM	IAM	C <sub>8</sub>	C <sub>8</sub>	C <sub>8</sub>	C <sub>18</sub>
Compound	10%	15%	20%	20%	30%	40%	20%
	MeOH	MeOH	MeOH	MeOH	MeOH	MeOH	MeOH
1	1.366	1.281	0.995	1.824	1.329	0.714	2.068
2	1.263	1.074	0.887	1.519	1.032	0.449	1.739
3	0.795	0.600	0.439	0.998	0.564	0.059	1.206
4	0.625	0.463	0.313	1.348	0.874	0.344	1.559
5	0.451	0.267	0.122	0.446	0.129	-0.244	0.655
6	0.340	0.165	0.011	0.586	0.182	-0.242	0.778
7	0.276	0.164	0.040	0.336	-0.009	-0.395	0.531
8	0.175	0.070	-0.056	0.610	0.264	-0.128	0.789
9	0.126	-0.071	-0.236	0.250	-0.123	-0.507	0.418
10	0.095	-0.025	-0.139	0.007	-0.283	-0.621	0.182
11	0.035	-0.114	-0.235	-0.064	-0.347	-0.642	0.139
12	0.022	-0.120	-0.242	0.090	-0.199	-0.522	0.229
13	0.015	-0.127	-0.280	0.375	-0.019	-0.427	0.516
14	-0.041	-0.206	-0.319	0.020	-0.254	-0.557	0.192
15	-0.054	-0.179	-0.320	0.400	-0.003	-0.392	0.643
16	-0.150	-0.231	-0.320	-0.571	-0.759	-0.976	-0.378
17	-0.155	-0.318	-0.443	-0.044	-0.310	-0.636	0.112
18	-0.186	-0.414	-0.682	0.072	-0.237	-0.573	0.368
19	-0.200	-0.353	-0.491	-0.146	-0.413	-0.677	0.008
20	-0.269	-0.366	-0.510	-0.240	-0.546	-0.833	-0.001
21	-0.269	-0.378	-0.490	-0.409	-0.645	-0.892	-0.275
22	-0.287	-0.392	-0.531	0.190	-0.205	-0.587	0.397
23	-0.340	-0.426	-0.557	-0.599	-0.799	-1.017	-0.433
24	-0.386	-0.521	-0.654	0.089	-0.326	-0.674	0.353
25	-0.423	-0.511	-0.642	-0.530	-0.739	-0.951	-0.377
26	-0.433	-0.539	-0.654	-0.168	-0.502	-0.782	0.029
27	-0.453	-0.544	-0.671	-0.439	-0.651	-0.860	-0.320
28	-0.481	-0.569	-0.695	-0.381	-0.654	-0.900	-0.172
29	-0.525	-0.579	-0.692	-0.465	-0.950	-1.112	-0.194
30	-0.527	-0.612	-0.740	-0.528	-0.783	-1.007	-0.320
31	-0.702	-0.771	-0.848	-0.839	-0.981	-1.115	-0.624
32	-0.704	-0.730	-0.918	-0.870	-1.028	-1.179	-0.762
33	-0.782	-0.784	-0.935	-0.907	-1.049	-1.190	-0.658
34	-0.852	-0.826	-0.998	-0.990	-1.070	-1.199	-0.860
35	-1.483	-1.292	-1.675	-1.270	-1.303	-1.420	-1.073

<sup>a</sup>The RSDs of the retention data for the internal standard were all smaller than 1%.







Appendix D. The literature experimental  $\log P_{ow}$  values

Compound	Name	$\log P_{\rm ow}$	pH	Method	Temperature
2	6-Benzylaminopurine	1.57 <sup>a</sup>	-	Shake-flask	25 °C
4	3-Isobutyl-1-methylxanthine	1.29 <sup>b</sup>	7.4	Shake-flask	25 °C
15	Caffeine	$-0.07^{\circ}$	_	_	_
16	6-Thioguanine	$-0.07^{\circ}$	8.2	_	_
18	8-Chlorotheophylline	$-0.85^{d}$	7.4	_	_
20	Adenine	$-0.09^{\circ}$	_	_	_
26	Theobromine	$-0.78^{\circ}$	_	_	_
27	Purine	$-0.37^{\circ}$	_	_	_
28	3-Methylxanthine	$-0.72^{b}$	7.4	Shake-flask	25 °C
29	Guanine	-0.96 <sup>c</sup>	7.4	_	_
31	Hypoxanthine	-1.11 <sup>e</sup>	7.4	_	_
33	Xanthine	$-0.73^{\circ}$	_	_	_
35	Uric acid	$-2.17^{f}$	7.0	Shake-flask	25 °C

<sup>a</sup> W.E. Shafer. Physiol. Plantarum. 78 (1990) 43.

<sup>b</sup> T. Hasegawa, K. Miyamoto, S. Wakusawa, K. Takagi, R. Apichartpichean, T. Kuzuya, M. Nadai, T. Horiuchi, Int. J. Pharmaceut. 58 (1990) 129.

<sup>c</sup> C. Hansch, A. Leo, D. Hoekman, Exploring QSAR: hydrophobic, electronic, and steric constants, ACS, Washington, 1995.

<sup>d</sup> D.R. Sanvordeker, S. Pophristov, A. Christensen, Drug. Devel. Ind. Pharm. 3 (1977) 149.

<sup>e</sup> N. Kolassa, K. Pfleger, W. Rummel, Eur. J. Pharmacol. 9 (1970) 265.

<sup>f</sup> A. Nahum, C. Horvath, J. Chromatogr. 192 (1980) 315.

Appendix E. Predicted log  $P_{ow}$  based on the regression model in Fig. 4a except for the five analytes with large substitution groups at position 7 or 9 (9, 13, 21, 22, and 24) which do not rank among the 13 analytes used for the regression



Appendix F. Chemometric procedures and analyses

A further parametric CoMFA [1] technique, comparative molecular similarity indices analysis (CoMSIA) [2] based on Gaussian-type potential functions, has been shown insensitive to small changes in the structural alignment of compounds or the orientation of the grid. In CoMSIA, the steric (S), electrostatic (E), hydrophobic (H), and H-bond donor/acceptor (D/A) molecular fields [2-4] are used. The initial models of all purines prepared with the Tripos Sybyl 6.9.2 [5] were optimized using the standard Tripos force field and their relevant partial-atomic charges were adopted from the empirical Gasteiger-Hückel (GH) set. The parent purine base (27) served as an alignment template for the superposition. Following the standard CoMFA procedure [5], each purine compound was mapped onto a 3D lattice  $(20 \text{ Å} \times 20 \text{ Å} \times 20 \text{ Å})$  with grid points 2.0 Å apart. An sp<sup>3</sup>-hybridized carbon atom with a charge of +1 was employed as the interaction probe. The cut-off interaction energies were set to 30 kcal/mol. CoMSIA similarity indices descriptors were calculated using the same lattice box as in the CoMFA calculations. The attenuation factor  $\alpha$  was set to 0.3. The internal partial-least-squares (PLS) analysis followed by the leave-oneout (LOO) cross-validation [6-8] was used to determine the optimal number of principal components (PCs) and to formulate the 3D-QSRR models using the standard implementation in the Sybyl package.

The CoMFA/CoMSIA PLS results based on the GH-charge formalisms were summarized in Table A1. More rigorous Y-randomization test [6–8] normally expected to generate low  $q^2$  and  $R^2$  values for valid models has been performed to ensure the robustness of our CoMFA/CoMSIA models. In the present work, none of the Y-randomization evaluations (Table A2) gave comparable  $q^2$  values to those of the leave-one-out results (Table A1). In addition, the leave-four-out cross-validation results (not shown) are quite similar to the leave-one-out ones.

Table A1	
Summary of the CoMFA/CoMSIA-QSRR results <sup>a,b</sup>	

-					
Model	$q^2$	$R^2$	PCs	F	S
CoMFA	0.93641	0.936	4	110	0.153
CoMSIA					
S + E	0.674	0.886	4	59	0.204
Н	0.768	0.955	5	124	0.130
A+D	0.263	0.435	1	25	0.436
S + E + H	0.824	0.968	4	227	0.108
all-field	0.832	0.974	4	286	0.097

<sup>a</sup> S = steric, E = electrostatic, H = hydrophobic, A = H-bond acceptor, D = Hbond donor, all-field = (S + E + A + D + H). q, R, PC, F, and s are the crossvalidated regression coefficient, conventional Pearson regression coefficient, optimal number of principal components, statistical F value, and standard error of estimate, respectively.

<sup>b</sup> The maximum number of PCs in the cross-validated runs was set to six for consistent comparison.

Table A2The Y-randomization results<sup>a</sup>

	Model	Model								
	CoMFA		CoMSIA	$\sqrt{S+E+H}$	CoMSIA	/all-field	eld			
Run	$\overline{q^2}$	PCs		PCs	$-Q^{2}$	PCs				
1	-0.087	5	-0.015	2	0.169	6				
2	0.080	1	0.078	1	0.152	5				
3	-0.062	1	0.005	2	-0.044	1				
4	-0.103	3	-0.258	1	-0.230	5				
5	-0.160	1	-0.244	1	-0.254	1				
6	0.333	3	0.360	4	0.312	2				
7	0.101	1	0.121	1	0.231	3				
8	0.220	2	0.254	2	0.201	2				
9	-0.110	2	0.003	2	0.121	1				
10	-0.232	2	-0.155	3	0.005	2				
11	0.123	1	0.129	2	0.210	3				
12	-0.111	2	-0.067	1	-0.264	2				
13	0.187	1	0.210	2	0.145	1				
14	0.432	3	0.455	3	0.470	4				
15	0.231	1	0.3	3	0.346	3				
16	-0.180	2	-0.145	2	-0.123	1				
17	-0.154	1	-0.055	2	-0.056	1				
18	0.328	3	0.289	2	0.356	2				
19	0.164	1	0.210	2	0.299	5				
20	0.011	1	0.200	1	0.042	1				
21	0.124	2	0.213	1	0.127	2				
22	-0.456	2	-0.426	4	-0.356	2				
23	-0.312	1	-0.215	3	-0.378	2				
24	-0.175	4	-0.287	6	-0.211	6				
25	0.399	4	0.421	4	0.411	3				
26	0.456	4	0.471	5	0.478	6				
27	0.126	1	0.101	1	0.099	1				
28	0.211	2	0.231	1	0.231	2				
29	0.244	2	0.333	3	0.294	1				
30	-0.222	1	-0.312	1	-0.106	1				
LOO <sup>b</sup>	0.747	4	0.824	4	0.832	4				

Abbreviations: see Table A1.

<sup>a</sup> In principle, lower  $q^2$  values than 0.5 as expected for the valid models are yielded in the *Y*-randomization tests to ensure the robustness of our resulting models. Leave-one-out cross-validation results from Table A1 for comparison.

A CoMSIA model with the combined S + E + H contributions (Table A1) indicated relatively high  $q^2$  and  $R^2$  values (0.824 and 0.968), a remarkable improvement over the CoMFA model (0.747 and 0.936, respectively). The CoMSIA model with hydrophobic field alone demonstrated a higher  $q^2$  value (0.768) and a low s value (0.130). In addition, the CoMSIA model with the combined all-five-fields are also statistically significant ( $q^2 = 0.832$  and  $R^2 = 0.974$ ) including the additional H-bond donor/acceptor interactions relative to the CoMSIA/S+E+H model. The linear relationships of the CoMFA/CoMSIA-fitted versus measured retention data of the 35 analytes were plotted in Fig. 6a. The linear correlations indicated the CoMFA/CoMSIA models are adequate to interpret the retention mechanism on the IAM.PC.DD2 phase. The hydrophobic contributions of the S+E+H and all-five-field CoMSIA models are 54.5% and 41.2%, respectively. Based on the statistical results, one can conclude that the hydrophobic factor still play an important baseline role in the IAM-HPLC retention-dependence of the nucleobases.

To consider the deprotonated states of the four analytes **28**, **31**, **33**, and **35** (Fig. 4), their most favorable deprotonated forms with the other nucleobases unchanged were subjected to further CoMFA/CoMSIA analyses. Based on DFT B3LYP/6-31G(d) calculations, the most favorable deprotonated positions of analytes **28**, **31**, and **33** are at the N(9) position except **35** at the N(3). The cross-validated  $q^2$  for the newly derived CoMFA, CoMSIA/H, CoMSIA/S + E + H, and CoMSIA/all-field models are 0.628, 0.782, 0.805, and 0.808, respectively. It has been demonstrated here that only the CoMFA model was moderately sensitive to the deprotonated forms of nucleobases. The same statistical procedures were repeated for each of the four analytes adopting its deprotonated form alone. No significant difference leading a different conclusion was found.

The 3D contour plots of the CoMSIA/S + E + H model allowed the correlation of experimentally determined retention data with changes in the steric/electrostatic (Fig. 6b) and hydrophobic (Fig. 6c) molecular-field contributions. Green area (sterically favorable) indicates a relatively important influence on the retention dependence in Fig. 6b. Bulky groups at the 6 ring position (analytes 1-3, 6, 8, 19, Appendix A) and methyl groups at position 1 (4,15,18) clearly enhanced the overall retention according to the CoMSIA models. The extra methyl group of analyte 15 at position 1 brought on a dramatic increase in the retention relative to 26. However, bulky groups at position 7 decreased the retention (e.g., compare 15, 22, and 24). The hydrophobic methyl group of compound 30 at position 8 close to the purple region (Fig. 6c) resulted in a shorter elution time than that of 28 at position 3.

#### **References for Appendix F**

- D.R. Cramer III, D.E. Paterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959.
- [2] G. Klebe, U. Abraham, T.J. Mietzner, J. Med. Chem. 37 (1994) 4130.
- [3] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, J. Chem. Inf. Comput. Sci. 29 (1989) 163.
- [4] G. Klebe, U. Abraham, J. Comput.-Aided. Mol. Des. 13 (1999) 1.
- [5] Tripos Inc., Tripos Sybyl 6.9.2 for Linux, St. Louis, MO, 2004.
- [6] Tropsha, P. Gramatica, V.K. Gombar, QSAR. Comb. Sci. 22 (2003) 69.
- [7] Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269.
- [8] H. van de Waterbeemd, Chemometric Methods in Molecular Design, VCH, Weinheim, 1995.





#### References

- [1] A. Leo, C. Hansch, D. Elkins, Chem. Rev. 71 (1971) 525.
- [2] F. Barbato, G. di Martino, L. Grumetto, M.I. La Rotonda, Eur. J. Pham. Sci. 22 (2004) 261.
- [3] A. Taillardat-Bertschinger, P.A. Carrupt, F. Barbato, B. Testa, J. Med. Chem. 46 (2003) 655.
- [4] H. van de Waterbeemd, Chemometric Methods in Molecular Design, VCH, Weinheim, 1995.
- [5] T. Suzuki, Y. Kudo, J. Comput. -Aided. Mol. Des. 4 (1990) 155.
- [6] C. Pidgeon, S.W. Ong, H.S. Chol, H.L. Liu, Anal. Chem. 66 (1994) 2701.
- [7] S.W. Ong, H.L. Liu, X.X. Qiu, G. Bhat, C. Pidgeon, Anal. Chem. 67 (1995) 755.
- [8] (a) G.W. Caldwell, J.A. Masucci, M. Evangelisto, R. White, J. Chromatogr. A. 800 (1998) 161;

(b) 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.

[9] P. Barton, A.M. Davis, D.J. McCarthy, P.J.H. Webborn, J. Pharm. Sci. 86 (1997) 1034.

- [10] C.Y. Yang, S.J. Cai, H.L. Liu, C. Pidgeon, Adv. Drug Deliv. Rev. 23 (1996) 229.
- [11] S.P. Assenza, P.R. Brown, J. Chromatogr. 282 (1983) 477.
- [12] I. Baranowska, M. Zydron, Anal. Bioanal. Chem. 373 (2002) 889.
- [13] H.B. Luo, Y.-K. Cheng, QSAR. Comb. Sci. 24 (2005) 968.
- [14] H.B. Luo, Y.-K. Cheng, J. Chromatogr. A. 1103 (2006) 356.
- [15] (a) G. Kortüm, W. Vogel, K. Andrussow, Dissociation constants of organic acids in aqueous solution, Butterworth, London, 1961;
  (b) R. Gargallo, C.A. Sotriffer, K.R. Liedl, B.M. Rode, J. Comput. -Aided Mol. Des. 13 (1999) 611;
  (c) A. Avdeef, Sirius Technical Application notes (STAN), 1, Sirius Analytical Instruments Ltd., Forest Row, UK, 1994, p. 26.
- [16] C. Yamagami, N. Takao, Chem. Express. 6 (1991) 113.
- [17] W.J. Lambert, J. Chromatogr. A. 656 (1993) 469.
- [18] M.J. Ruiz-Angel, S. Carda-Broch, M.C. García-Alvarez-Coque, A. Berthod, J. Chromatogr. A. 1063 (2005) 25.
- [19] D.R. Cramer III, D.E. Paterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959.
- [20] G. Klebe, U. Abraham, T.J. Mietzner, J. Med. Chem. 37 (1994) 4130.
- [21] J.M. Carrozzino, M.G. Khaledi, J. Chromatogr. A. 1079 (2005) 307.