三维全息原子场作用矢量用于嘌呤衍生物定量结构保留相关的研究

仝建波[①] 周 鹏[②③] 张生万[①*] 周 原[②③] 梅 虎[②③] 曾 晖[②③]
李美萍[①] 李志良[②③*]

(① 山西大学化学化工学院, 太原 030006; ② 重庆大学化学化工学院, 重庆 400044; ③ 生物医学工程教育部与重庆市重点实验室, 重庆 400044. * 联系人, E-mail: zswan@sxu.edu.cn, zlli2662@163.com)

摘要 采用本实验室新近提出的三维全息原子场作用矢量(3D-HoVAIF)对 33 个嘌呤衍生物进行定量结构保留相关(QSRR)的研究, 运用逐步回归(SMR)筛选变量后, 用偏最小二乘(PLS)回归建模与交互校验复相关系数, 标准偏差分别为 R^2_{cum} =0.966 与 Q^2_{cum} =0.860 及 SD=0.112, 均优于文献值, 色谱保留预测模型具有良好的稳定性和预测能力. 研究结果表明: 3D-HoVAIF 能较好地表征该类分子结构信息, 有物化意义明确及结果易于解释等特点, 值得进一步推广应用.

关键词 三维全息原子场作用矢量 定量结构保留相关 色谱参数预测 嘌呤衍生物

分子结构与色谱保留之间的定量关系即定量结构保留相关(QSRR)[1,2]的构建, 需要引入相关化合物的描述符(子). 长期以来, 人们在探寻定量结构保留相关、预测色谱保留指数、选择色谱分离条件及探索保留机理[3]等方面做了许多有意义的工作. 例如利用分子几何结构、拓扑性质及连接特征和各种物化参数等进行结构描述, 然后建立QSRR模型来预测化合物色谱保留值[4~10]. 但上述方法是基于 2D结构描述符提出的, 由于 2D描述符通常难以再现分子真实空间立体结构, 并没有考虑到化合物和固定相及流动相的相互作用, 因而对有机或生物大分子建立更有效的QSRR模型存在一定困难. 在这种背景下 3D描述方法得以迅速发展, 并成为QSRR分子结构表征的主流趋势, Gramatica等[11]利用Todeschini发展的加权全局不变分子指数(WHIM)[12~14]对多氯联苯色谱保留指数进行计算; 最近Luo等[15]利用Cramer等[16]提出的比较分子场法(CoMFA)研究嘌呤衍生物的色谱保留行为. 遗憾的是WHIM指数是通过不同物理量对原子空间坐标进行加权变换产生对旋转和平移不变量而得到的, 计算过程相对复杂, 因而难以得到广泛应用, 而CoMFA弊端是在进行一组分子研究时首先要对样本进行空间结构叠合, 构象重叠, 另外空间网格划分、变量数目控制及势场探针选取也是不容忽视的问题. 鉴此, 本文考虑到 3D分析方法具有意义明确、计算准确、可解释性强等特点, 同时结合 2D表征技术简便而迅速的优势, 运用本实验室最近提出的 3D分子结构信息表征技术——三维全息原子场作用矢量(3D-HoVAIF)建立嘌呤衍生物色谱保留行为与其结构参数之间的定量关系, 得到了满意的结果.

1 原理与方法

在分子电性距离矢量[17~19]基础上, 本实验室最近提出一种全新 3D分子结构描述符: 三维全息原子场作用矢量(3D-HoVAIF). 该法从分子立体结构两个不变量——原子相对距离和原子自身性质出发, 基于 3 种常见的与生物活性直接相关的原子间非键作用方式(静电、立体和疏水作用), 在不借助任何实验参数及母需样本构象重叠的前提下得到了用于表征分子结构特征的多维矢量描述符. 有机分子中常见原子有氢、碳、氮、磷、氧、硫、氟、氯、溴、碘, 它们分别属于元素周期表的 5 个主族(ⅠA, ⅣA, ⅤA, ⅥA, ⅦA), 按其所处元素周期表划分为 5 大类, 为更精确描述原子结构特征, 进一步把不同主族中原子按其杂化状态细分为 10 类, 这样一个分子内部不同类别原子之间的相互作用情况就有 55 种. 3D-HoVAIF采用静电、立体和疏水 3 种势能来表达不同的作用形式, 这样对于一个有机化合物分子将有 3×55=165 对原子作用项来表征分子结构信息.

(1) 静电作用场. 静电作用可用经典的 Coulomb 定理(式(1))来表述.

$$E_{mn}(E)=\sum_{i\in m,j\in n}\frac{e^2}{4\pi\varepsilon_0}\cdot\frac{Z_i\cdot Z_j}{r_{ij}}\ (1\leq m\leq 10,\ m\leq n\leq 10),\quad(1)$$

式(1)中 r_{ij} 是原子间 Euclid 距离，单位为 nm; e 为单位电荷电量 1.6021892×10^{-19}C; ε_0 为真空中的介电常数 8.85418782×10^{-12}C^2/(J·m); Z 为原子净电荷数，以电子为单位; m 和 n 为原子所属种类. 通过该式计算分子中所有原子间的静电作用势，并按所属类型将其分别计入 55 个静电作用项中.

(2) 立体作用场. 立体作用采用 Lennard-Jones 方程来描述这种作用方式(式(2)).

$$E_{mn}(S)=\sum_{i\in m,j\in n}\varepsilon_{ij}\cdot D\cdot\left[\left(\frac{R_{ij}^*}{r_{ij}}\right)^{12}-2\cdot\left(\frac{R_{ij}^*}{r_{ij}}\right)^{6}\right]$$
$$(1\leq m\leq 10,\ m\leq n\leq 10),\quad(2)$$

式(2)中 ε_{ij} 为原子对势能阱深[20,21]; D 为经验推导的原子间作用能校正常数，取 0.01[22]; $R_{ij}^*=(C_h\cdot R_{ii}^*+C_h\cdot R_{jj}^*)/2$，为校正后的原子对van der Waals半径，校正因子 C_h 当 sp^3 杂化时取 1.00, sp^2 杂化取 0.95, sp杂化取 0.90[22].

(3) 疏水作用场. 疏水作用使用Kellogg等提出的hint方法[23~27]来定义计算两原子间疏水作用表达式(式(3)).

$$E_{mn}(H)=\sum_{i\in m,j\in n}S_i\cdot a_i\cdot S_j\cdot a_j\cdot e^{-r_{ij}}\cdot T_{ij}$$
$$(1\leq m\leq 10,\ m\leq n\leq 10),\quad(3)$$

式(3)中 S 为原子溶剂可及面积，是以水分子(半径为 0.14 nm)为探针在原子表面滚动其球心形成的表面面积[23]; a 为原子疏水性常数[24]; T 是作用形式的二值判别函数，以表明不同类型原子疏水作用的熵效应变化方向[25~29].

2 结果与讨论

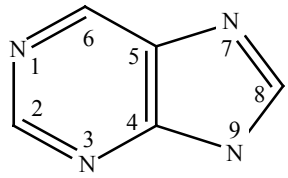
2.1 嘌呤衍生物结构与色谱保留数据

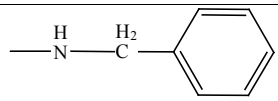
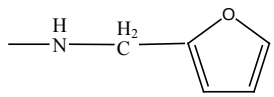
本文研究的 33 个嘌呤衍生物的高效液相色谱的保留数据——容量因子对数值取自文献[15](其分子结构式与保留数据见表 1). 采用高效液相色谱仪(Waters 2695)带二极管阵列光度检测器(Waters 996; 测定波长 254 nm)，测定时取进样量 10 mL 将嘌呤衍生物及内标 1,3-二甲基-5-氟尿嘧啶溶解在乙腈、甲醇、水(体积比为 1∶1∶3)的混合溶剂中，以 Waters XterraTM MS C18 硅胶柱(柱温 25℃)，流动相为 20%(体积分数)甲醇水溶液，流速 1 mL/min，以硝酸钠(测定波长 210 nm)的出峰时间作为柱子的死时间，将容量因子的对数值作为保留参数.

2.2 模型的建立

使用 Chemoffice 8.0 自动构建 33 个嘌呤衍生物的初始分子立体结构，用 Chem3D 自带的 MOPAC 半经验量子化学软件在 AM1 水平上最终优化得到分子结构(截断值 0.0001 kJ/mol)，并采用 Mülliken 布居分析法以单点形式计算出原子的净电荷数量，将分子中每个原子的空间位置及电荷分别以笛卡儿坐标和净电子数目的形式输入实验室自编的 C 语言应用程序 3D-HoVAIF.EXE 加以处理，得到分子 3D-HoVAIF 描述符. 3D-HoVAIF 最多含 165 个描述符，由于所研究的 33 个嘌呤衍生物不含 v_{13}, v_{16}, v_{28}, v_{32}, v_{33}, v_{34}, v_{46}, v_{47}, v_{48}, v_{49}, v_{50}, v_{52}, v_{68}, v_{71}, v_{83}, v_{87}, v_{88}, v_{89}, v_{101}, v_{102}, v_{103}, v_{104}, v_{105}, v_{107}, v_{123}, v_{126}, v_{138}, v_{142}, v_{143}, v_{144}, v_{156}, v_{157}, v_{158}, v_{159}, v_{160}, v_{162}，因此最终得到 129 个非零矢量. 使用这 129 个非零矢量进行逐步线性回归(SMR)，并通过设定适当的 F 临界值，使每个变量按其重要性的大小顺序逐步引入方程并采取有进有出的原则，选出 31 个显著性矢量. 逐步回归由统计学软件 SPSS 13.0 完成，其回归过程各方程标准化回归系数见本文网络版附表 1; 同时将 SMR 每一步得到的原始变量矩阵经偏最小二乘回归(PLS)建模，并同时使用 LOO-CV 法计算模型的 Q^2_{cum} 大小，以其达到最大值时变量数目来确定最终模型. 这里 PLS 由化学计量学软件 Simca-p 10.0 实现. 需要说明的是，对于 SMR 每一步生成原始变量矩阵 X，其 PLS 隐变量(Latent Variable)数目的确定按照 Simca-P 10.0 默认标准得到，即通过 NIPALS 算法进行迭代并依据解释原始变量方差大小依次计算各个主成分，同时检测每次得到的主成分对模型交互检验复相关系数(Q^2_{cum})的贡献，当某一个被引入的主成分对模型的 Q^2_{cum} 相关性小于 0.097 时即认为该主成分已不显著，并停止计算更多的主成分. 每一步 PLS 建模所得到的主成分数、复相关系数(R^2_{cum})以及交互校验的复相关系数(Q^2_{cum})随变量引入的变化情况见本文网络版附表 1. 可以看到当取 5, 11, 17, 20, 24 个变量时模型 Q^2_{cum} 值出现极值，其中取 17 个变量时达到最大值，由上述变量组合分别建立 PLS 模型，建模过程中相关统计参数见表 2.

表 1 嘌呤衍生物结构及高效液相色谱保留数据



编号	1	2	3	6	7	8	9	logk_{OBS}	logk_{EST}	logk_{Err}
1		−H				−H	−H	1.683	1.818	−0.135
2	$-CH_3$	=O	*i*-butyl	=O		−H	−H	1.521	1.559	−0.038
3		−H				−H	−H	1.196	1.118	0.078
4		−Cl		−Cl		−H	−H	0.787	0.701	0.086
5		−H		−OEt		−H	−H	0.771	0.830	−0.059
6		−H		$-N(CH_3)_2$		−H	−H	0.770	0.459	0.311
7	$-CH_3$	=O	$-CH_3$	=O		−Cl	−H	0.701	0.695	0.006
8	$-CH_3$	=O	$-CH_3$	=O	$-CH_3$	−H		0.624	0.554	0.070
9		−H		$-NH_2$		−H	$-CH_2CO_2C_2H_5$	0.531	0.494	0.037
10		$-NH_2$		$-SCH_3$		−H	−H	0.529	0.527	0.002
11	$-CH_3$	=O	$-CH_3$	=O	$-C_2H_4OH$	−H		0.420	0.317	0.103
12		−H		$-NH_2$		−H	$-C_2H_5$	0.403	0.479	−0.076
13		−H		−Br		−H	−H	0.318	0.186	0.132
14		−H		−Cl		−H	−H	0.225	0.178	0.047
15		$-NH_2$		−Br		−H	−H	0.220	0.255	−0.035
16		−H		−CN		−H	−H	0.174	0.242	−0.068
17		$-NH_2$		−Cl		−H	−H	0.131	0.245	−0.114
18	−H	$-NHCOCH_3$		=O		−H	$-COCH_3$	0.052	0.057	−0.005
19	−H	=O	$-CH_3$	=O	$-CH_3$	−H		0.026	0.083	−0.057
20		−H		$-CH_3$		−H	−H	0.005	0.071	−0.066
21	−H	=O	−H	=S		−H	−H	−0.099	−0.459	0.360
22	−H	=O	−H	=O		$-CH_3$	−H	−0.168	−0.111	−0.057
23		−H		$-NH_2$		−H	−H	−0.244	−0.288	0.044
24	−H	=O	$-CH_3$	=O		−H	−H	−0.276	−0.156	−0.120
25		$-NH_2$		$-NH_2$		−H	−H	−0.300	−0.230	−0.070
26	−H	$-NH_2$		=S		−H	−H	−0.331	−0.271	−0.060
27		−H		−H		−H	−H	−0.340	−0.331	−0.009
28		$-NH_2$		−H		−H	−H	−0.389	−0.269	−0.120
29	−H	=O	−H	=O	−H	=O	−H	−0.428	−0.499	0.071
30	−H	=O	−H	=O		−H	−H	−0.537	−0.508	−0.029
31	−H	$-NH_2$		=O		−H	−H	−0.610	−0.525	−0.085
32	−H	=S	−H	−H		−H	−H	−0.618	−0.520	−0.098
33	−H	−H		=O		−H	−H	−0.626	−0.557	−0.069

本文利用新型三维全息原子场作用矢量对嘌呤衍生物色谱保留行为及其化学结构进行定量结构保留关系研究，取得模型的相关统计量(复相关系数 R^2_{cum}、交互检验复相关系数 Q^2_{cum})与文献[15]结果见表 2. 由此可看出，使用 3D-HoVAIF 描述符对嘌呤衍生物建立 QSRR 模型包含(17 个原始变量)4 个主成分，解释了*Y*变量 98.29%方差，交互检验所解释*Y*变量方差为 92.74%，可见模型有很高自身拟合能力以及对外部样本的预测能力. 其中模型复相关系数 R^2_{cum} 与 Luo等[15]使用CoMFA[16]拟合的结果相当，而交互检验 Q^2_{cum} 明显高于文献[15]使用同类线性方法的建模结果. 值得注意的是文献[15]在计算过程中使用了

表 2　不同方法应用于 33 个嘌呤衍生物建模结果比较

编号	方法	模型	优化方法	电荷类型	m[a]	l[b]	R^2_{cum}[c]	Q^2_{cum}[d]	SD[e]	F[f]
1	CoMFA	PLS	B3LYP/6-311G	NPA[g]	-	7	0.984	0.738	0.085	223
2	CoMFA	PLS	B3LYP/6-311G	MPA[h]	-	7	0.983	0.722	0.089	204
3	CoMFA	PLS	B3LYP/6-311G	MMFF94	-	9	0.993	0.751	0.059	362
4	CoMFA	PLS	B3LYP/6-311G	GH[i]	-	4	0.947	0.752	0.148	125
5	CoMFA	PLS	B3LYP/6-311G	GM[j]	-	4	0.938	0.693	0.160	106
6	CoMFA	PLS	Tripos field	MMFF94	-	4	0.946	0.685	0.149	122
7	CoMFA	PLS	Tripos field	GH[i]	-	4	0.919	0.621	0.183	79
8	CoMFA	PLS	Tripos field	GM[j]	-	5	0.937	0.613	0.164	80
9	3D-HoVAIF	MLR	AM1	MPA[h]	8	-	0.841	0.689	0.277	16
10	3D-HoVAIF	PLS	AM1	MPA[h]	5	2	0.888	0.746	0.192	246
11	3D-HoVAIF	PLS	AM1	MPA[h]	11	3	0.959	0.846	0.121	721
12	3D-HoVAIF	PLS	AM1	MPA[h]	17	4	0.966	0.860	0.112	859
13	3D-HoVAIF	PLS	AM1	MPA[h]	20	4	0.963	0.849	0.115	815
14	3D-HoVAIF	PLS	AM1	MPA[h]	24	5	0.976	0.858	0.093	1250

a) 逐步回归过程筛选变量数; b) 偏最小二乘回归选取主成分数; c) 复相关系数; d) 交互检验复相关系数; e) 标准偏差; f) F-检验值; g) 自然布局分析; h) Mülliken 布居分析; i) Gasteiger-Hückel; j) Gasteiger-Marsili

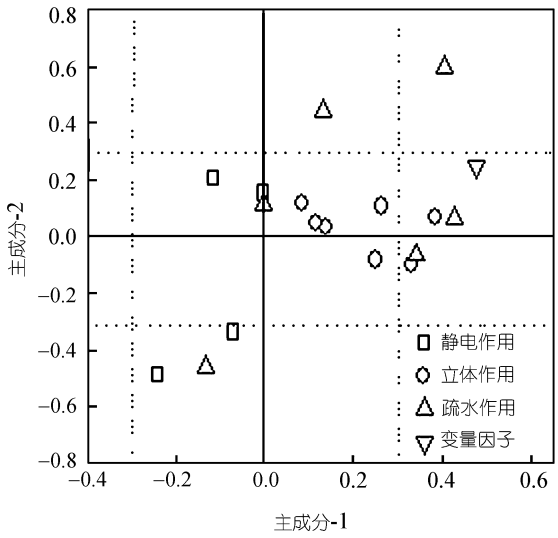


图 1　原始变量对前两个主成分的载荷贡献

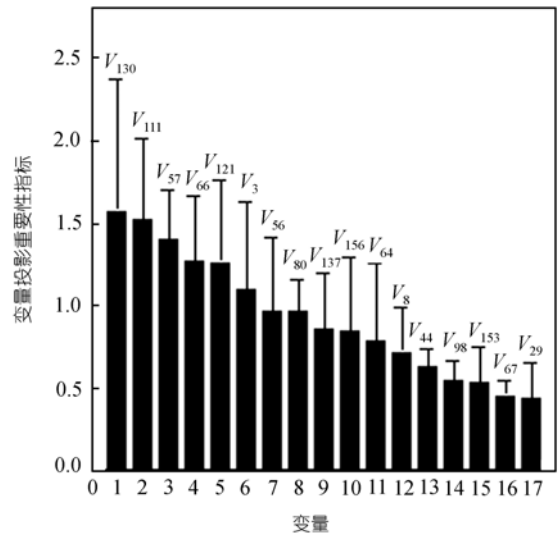


图 2　嘌呤衍生物的变量投影重要性指标图

相当复杂的高等级量化方法, 而此处只用了相对简便的中等级量化方法. 图 1 是 17 个变量对前两个主成分载荷的贡献情况, 可看到静电、立体和疏水作用对前两个主成分都有不同程度的贡献, 其中对 PC1 载荷贡献较大(大于经验值 0.3)为疏水(v_{111}, v_{121}, v_{130})和立体项(v_{57}, v_{66}), 它们都与 Y(容量因子对数值)呈较大的正相关, 对 PC2 载荷贡献较大的为疏水(v_{130}, v_{137}, v_{155})和静电(v_3, v_{44})作用. 另外原始变量对模型显著性贡献大小被绘制成变量投影重要性(VIP)指标(见图 2), 其中贡献最大的前 8 项(v_{130}, v_{111}, v_{57}, v_{66}, v_{121}, v_3, v_{56}, v_{80})有疏水、立体和静电作用, 可知嘌呤衍生物容量因子对数值与其分子疏水性、立体性和电性密切相关, 尤以疏水和立体性与容量因子对数值更为密切. 如以该 8 项变量与嘌呤衍生物容量因子对数值建立多元线性回归模型, 各项变量的相关统计参数见表 2. 图 3 为 33 个样本在 PLS 前两个主成分得分空间分布情况, 其中绝大多数样本都落在 95%置信度 Hotelling T^2 椭圆置信圈内, 而超出此范围的 1 和 2 号样本分别在 6 和 3 位上有一个大基团(6-苯甲胺基, 3-异丁基), 故结构略显特殊. 由此表明 3D-HoVAIF 能够恰当表现化合物的结构特征, 并在统计模型中作出正确反映. 以原始变量(X 矩阵)第一个主成分得分 t[1]对容量因子对数值(Y 矩阵)第一个主成分得分 u[1]作图(图 4), 可知原始变量 t[1]与 u[1]成线性相关, 几乎所有样本都均匀分布于过原点 45°直线周围, 无明显异常点. 图 5 是使用 3D-HoVAIF 建立的 QSRR 模型的计算值

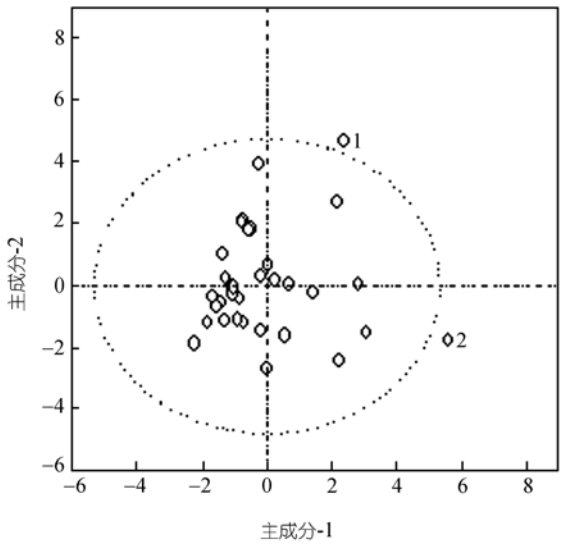


图 3 33 个嘌呤衍生物在前两个主成分的得分分布散点图

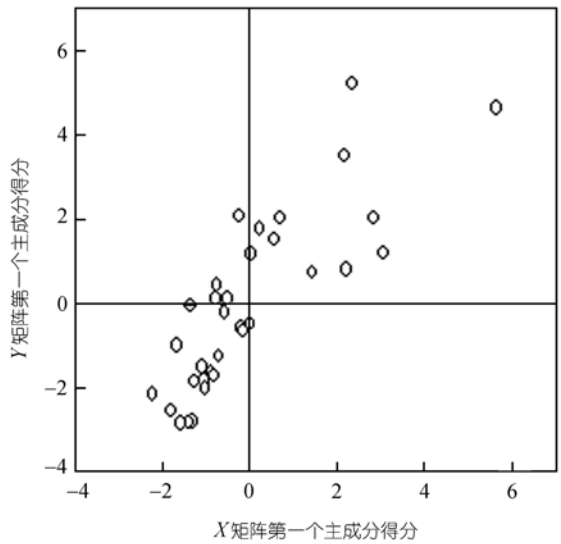


图 4 X 矩阵第一个主成分得分 t[1]对 Y 矩阵第一个主成分得分 u[1]作图

与实验观测值的相关情况(R^2_{cum} =0.966; Q^2_{cum} = 0.860), 显然所建模型具有良好的稳定性能和泛化能力.

3 结语

本文所建模型包含了电性、立体、疏水 3 种经典作用项, 说明 3D-HoVAIF 能够较好地表征嘌呤衍生物的色谱保留行为, 而且所建模型不依赖于实验数据, 与现今流行的三维分子建模方法, 如比较分子场法(CoMFA)相比, 不需要考虑构象的叠加, 物化意义明确, 可解释性强, 并且本实验室已经分别在甾体类

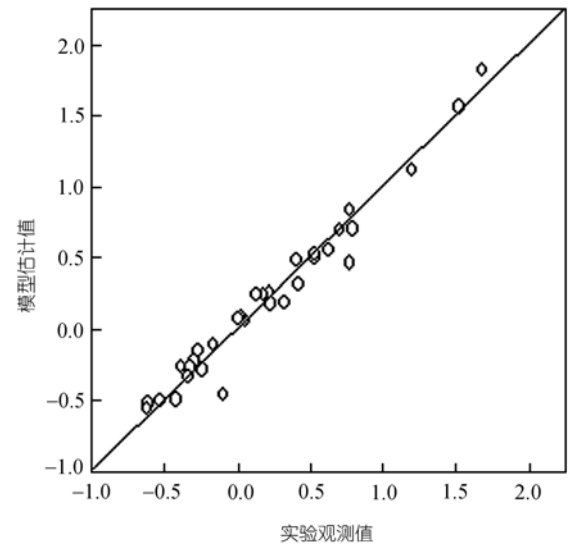


图 5 嘌呤衍生物容量因子实验观测值 log$k$ 与模型估计值相关图

化合物、抗艾滋病药物等的定量构效研究中取得了较好的结果, 有望用于其他含多环结构和多杂原子的复杂分子体系.

致谢 感谢杨力、李志真、蔡绍晢、李晟和、袁晓燕、李盛励、刘树深、李声时、黄莺、梁桂兆、杨善彬等提供的协助. 本工作为山西省工业攻关项目基金(批准号: 2006031204)及重庆市应用基础(批准号: 01-3-6)资助项目.

参 考 文 献

1 Kaliszan R. Quantitative Structure-Chromatographic Retention Relationships. New York: Wiley, 1987
2 Kaliszan R. Structure and Retention in Chromatography. A Chemometric Approach. Amsterdam: Harwood, 1997
3 Mehdi A A, Piotr H, Roman K, et al. Mechanism of separation on cholesterol—silica stationary phase for high-performance liquid chromatography as revealed by analysis of quantitative structure—retention relationships. J Pharm Biomed Anal, 1998, 18: 721—728[DOI]
4 Suzuki T, Timafei S, Iuoras B E, et al. Quantitative structure-enantioselective retention relationships for chromatographic separation of arylalkylcarbinols on pirkle type chiral stationary phases. J Chromatogr A, 2001, 922: 13—23[DOI]
5 Kang J J, Cao C Z, Li, Z. On quantitative structure-retention relationship (QSRR) studies for predicting the gas chromatography retention indices of polycyclic hydrocarbons (PAHs): quasi-length of carbon chain and pseudo-conjugated surface. J Chromatogr A, 1998, 799: 361—367[DOI]
6 Berenice D S J, Renata D D M C A, Rosendo A Y, et al. Prediction of the chromatographic retention of saturated alcohols on stationary phases of different polarity applying the novel semi-em-

pirical topological index. Anal Chim Acta, 2003, 477: 29—39[DOI]
7 Needham M D, Jurs P C. Quantitative structure-retention relationship studies of polychlorinated dibenzodioxins on gas chromatographic stationary phases of varying polarity. Anal Chim Acta, 1992, 258: 183—198[DOI]
8 邓洪, 黄萍, 胡银玉, 等. 新分子距边矢量用于环境毒物多氯代二苯并呋喃的定量结构保留相关建模. 科学通报, 2005, 50(9): 875—879
9 Liang X, Wang W, Wu W, et al. Quantitative relationship between chromatographic retentions and molecular structures of polychlorinated dibenzo-p-dioxins (PCDDs). Chemosphere, 2000, 41(6): 923—929[DOI]
10 Liang X, Wang W, Wu W, et al. Prediction of the retentions of polybrominated dibenzo-p-dioxins (PBDDs) by using the retentions of polychlorinated dibenzo-p-dioxins (PCDDs). Chemosphere, 2000, 41(6): 917—921[DOI]
11 Gramatica P, Navas N, Todeschini R. 3D-modelling and prediction by WHIM descriptors. Part 9. Chromatographic relative retention time and physico-chemical properties of polychlorinated biphenyls (PCBs). Chemom Intell Lab Syst, 1998, 40: 53—63[DOI]
12 Todeschini R, Lasagni M, Marengo E. New molecular descriptors for 2D and 3D structures. Theory J Chemom, 1994, 8: 263—272
13 Todeschini R, Gramatice P, Provenzani R. Weighted holistic invariant molecular descriptors. Part 2. Theory development and applications on modeling physicochemical properties of polycyclic aromatic hydrocarbons. Chemom Intell Lab Syst, 1995, 27: 221—229[DOI]
14 Bravi G, Gancia E, Mascagni P, et al. MS-WHIM, new 3D theoretical descriptors derived from molecular surface properties: A comparative 3D QSAR study in a series of steroids. J Comput-Aided Mol Des, 1997, 11: 79—92[DOI]
15 Luo H B, Cheng Y K. Quantitative structure-retention relationship of nucleic-acid bases revisited. CoMFA on purine RPLC retention. QSAR Comb Sci, 2005, 24(8): 968—975[DOI]
16 Cramer R D, Patterson D E, Bunce J D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc, 1988, 110: 5959—5967[DOI]
17 Liu S S, Cao C Z, Li Z. Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel Molecular Distance Edge (MDE) vector. J Chem Inf Comput Sci, 1998, 38: 387—394[DOI]
18 Liu S S, Cai S X, Li Z. Molecular Electronegative Distance Vector (MEDV) relating to 15 properties of alkanes. J Chem Inf Comput Sci, 2000, 40: 1337—1348[DOI]
19 Liu S, Yin C, Cai S X, et al. A novel MHDV descriptor for dipeptide QSAR studies. J Chin Chem Soc, 2001, 48: 253—260
20 Levitt M. Protein folding by restrained energy minimization and molecular dynamics. J Mol Biol, 1983, 170: 723—764
21 Levitt M, Perutz M F. Aromatic rings act as hydrogen bond acceptors. J Mol Biol, 1988, 201: 751—754[DOI]
22 Hahn M. Receptor surface models. 1. definition and construction. J Med Chem, 1995, 38(12): 2080—2090[DOI]
23 Kellogg G E, Semus S F, Abraham D J. HINT—a new method of empirical hydrophobic field calculation for CoMFA. J Comput-Aided Mol Des, 1991, 5: 545—552[DOI]
24 Wireko F C, Kellogg G E, Abraham D J. Allosteric modifiers of hemoglobin. 2. Crystallographically determined binding sites and hydrophobic binding/interaction analysis of novel hemoglobin oxygen effectors. J Med Chem, 1991, 34: 758—767[DOI]
25 Kellogg G E, Joshi G S, Abraham D J. New tools for modeling and understanding hydrophobicity and hydrophobic interactions. Med Chem Res, 1992, 1: 444—453
26 Kellogg G E, Abraham D J. KEY, LOCK. LOCKSMITH. Complementary hydrophobicity map predictions of drug structure from a known receptor/receptor structure from known drugs. J Mol Graph, 1992, 10: 212—217[DOI]
27 Nayak V R, Kellogg G E. Cyclodextrin-barbiturate inclusion complexes: a CoMFA/HINT 3D QSAR study. Med Chem Res, 1994, 3: 491—502
28 Hasel W, Hendrikson T F, Still W C. A rapid approximation to the solvent accessible surface areas of atoms. Tetrahed Comp Method, 1988, 1: 103—116[DOI]
29 Pei J F, Wang Q, Zhou J J. Estimating protein-ligand binding free energy: Atomic solvation parameters for partition coefficient and solvation free energy calculation. Proteins: Struct Funct Genet, 2004, 57: 651—664[DOI]

(2005-11-05 收稿, 2006-01-17 接受)

附表 1　逐步挑选变量过程的标准回归系数及偏最小二乘建模的统计量

模型	b_1	b_2	b_3	b_8	b_{11}	b_{14}	b_{22}	b_{23}	b_{27}	b_{29}	b_{43}	b_{44}	b_{56}	b_{57}	b_{64}	b_{66}	b_{67}	b_{80}	b_{90}	b_{91}	b_{98}	b_{110}	b_{111}	b_{121}	b_{124}	b_{130}	b_{137}	b_{149}	b_{150}	b_{155}	b_{163}	l[a)]	R^2_{cum}	Q^2_{cum}
1																							0.608									1	0.370	0.350
2			–0.555																				0.762									1	0.640	0.429
3			–0.533																				0.870				–0.379					1	0.770	0.519
4			–0.292																				0.778			0.364	–0.416					2	0.856	0.733
5			–0.242																		0.178		0.752			0.178	–0.437					2	0.888	0.746
6			–0.218							0.159											0.188		0.776			0.425	–0.463					2	0.909	0.741
7			–0.082							0.157		–0.182									0.174		0.798			0.491	–0.447					3	0.928	0.787
8			–0.113							0.144		–0.226		0.479							0.136		0.361			0.524	–0.424					3	0.942	0.760
9			0.066							0.141		–0.316		0.661							0.108		0.229			0.616	–0.416				0.178	3	0.942	0.787
10			0.108							0.145		–0.344		0.653				0.120			0.060		0.223			0.604	–0.423				0.177	3	0.948	0.792
11			0.112	0.157						0.151		–0.328		0.704				0.205			0.026		0.215			0.594	–0.432				0.176	3	0.959	0.846
12			0.154	0.180						0.142		–0.340	–0.127	0.777				0.209			0.022		0.260			0.591	–0.427				0.208	3	0.949	0.825
13			0.160	0.181						0.142		–0.352	–0.114	0.803				0.211			0.017		0.230			0.597	–0.147			0.284	0.205	3	0.951	0.793
14			0.166	0.192						0.139		–0.328	–0.132	0.835			–0.053	0.229			0.006		0.214			0.618	–0.136			0.307	0.215	3	0.950	0.791
15			0.195	0.1868						0.148		–0.343	–0.183	0.680		–0.128	–0.061	0.239			0.029		0.511			0.584	–0.132			0.286	0.222	3	0.943	0.800
16			0.164	0.158						0.141		–0.324	–0.122	0.780		–0.471	–0.047	0.224			0.056		0.248	0.449		0.612	–0.142			0.368	0.209	4	0.955	0.833
17			–0.005	0.114						0.146		–0.310	0.004	0.739	–0.138	–0.552	–0.045	0.205			0.072		0.107	0.612		0.576	–0.062			0.332	0.197	4	0.966	0.860
18			–0.010	0.111					–0.055	0.155		–0.326	0.015	0.678	–0.166	–0.529	–0.048	0.212			0.074		0.204	0.546		0.563	–0.135			0.348	0.198	5	0.975	0.816
19			–0.109	0.108	–0.256				–0.057	0.159		–0.289	0.094	0.706	–0.244	–0.771	–0.039	0.194			0.085		0.033	1.121		0.561	–0.126			0.404	0.177	4	0.959	0.822
20			–0.230	0.080	–0.488	0.186			–0.062	0.160		–0.277	0.141	0.505	–0.316	–0.914	–0.024	0.194			0.111		0.020	1.490		0.509	–0.071			0.357	0.162	4	0.963	0.849
21	0.139		–0.222	0.095	–0.521	0.169			–0.055	0.165		–0.264	0.141	0.734	–0.301	–1.029	–0.033	0.174			0.119		0.360	1.725		0.511	–0.132			0.333	0.162	4	0.961	0.846
22	0.380		–0.283	0.164	–0.769	0.262			–0.048	0.174	0.075	–0.230	0.186	1.042	–0.309	–1.365	–0.042	0.132			0.140		–1.069	2.453		0.489	–0.179			0.285	0.148	5	0.973	0.852
23	0.549		–0.305	0.198	–0.885	0.324			–0.041	0.180	0.111	–0.189	0.272	1.182	–0.278	–1.600	–0.046	0.104		–0.057	0.164		–1.530	2.893		0.492	–0.248			0.267	0.148	5	0.976	0.852
24	0.734		–0.559	0.212	–1.103	0.433			–0.043	0.205	0.162	–0.133	0.439	1.027	–0.291	–1.924	–0.007	0.058		–0.109	0.227		–1.780	3.465		0.343	–0.320	0.201		0.208	0.200	5	0.976	0.858
25	1.034		–0.801	0.212	–0.905	0.505			–0.038	0.238	0.214	–0.148	0.268	1.005	–0.317	–1.840	–0.104	0.035		–0.122	0.196		–1.269	2.832	–0.346	0.022	–0.449	0.331		0.217	0.200	4	0.959	0.806
26	1.349		–0.792	0.274	–0.933	0.529		0.105	–0.025	0.271	0.303	–0.141	0.207	1.229	–1.973	–1.973	–0.156	0.005		–0.053	0.199		–1.464	2.901	–0.485	–0.111	–0.505	0.363		0.358	0.204	3	0.929	0.767
27	1.345		–0.840	0.270	–0.994	0.562		0.090	–0.015	0.270	0.301	–0.129	0.247	1.191	–2.034	–2.304	–0.146	0.001		–0.076	0.208	0.030	–1.525	3.050	–0.469	–0.113	–0.379	0.378		0.358	0.206	3	0.927	0.763
28	1.384	0.109	–0.808	0.282	–0.984	0.585		0.097	–0.016	0.271	0.293	–0.126	0.220	1.345	–2.007	–2.007	–0.148	–0.002		–0.064	0.205	0.029	–1.596	3.053	–0.508	–0.111	–0.383	0.351		0.358	0.200	3	0.928	0.772
29	1.400	0.097	–1.601	0.283	–0.976	0.583	–0.026	0.098	–0.016	0.273	0.299	–0.124	0.210	1.370	–2.008	–2.008	–0.156	–0.004		–0.059	0.202	0.027	–1.601	3.035	–0.525	–0.129	–0.398	0.345		0.346	0.199	3	0.920	0.767
30	1.348	0.148	–0.814	0.270	–0.903	0.570	–0.128	0.108	–0.015	0.268	0.275	–0.121	0.197	1.425	–1.922	–1.922	–0.164	0.006	–0.035	–0.063	0.192	0.034	–1.552	2.894	–0.532	–0.094	–0.336	0.234		0.400	0.186	3	0.918	0.756
31	1.325	0.156	–0.806	0.266	–0.898	0.570	–0.132	0.100	–0.017	0.265	0.266	–0.120	0.201	1.426	–1.910	–1.910	–0.160	0.008	–0.037	–0.070	0.192	0.049	–1.556	2.894	–0.520	–0.077	–0.223	0.222	–0.013	0.513	0.185	3	0.914	0.736

a) l: 偏最小二乘回归的最佳主成分