

文章编号:1000-582x(2001)04-0154-05

# HMG-CoA 还原酶阻滞剂的多维多元定量构效关系\*

王远强<sup>1</sup>, 叶楠<sup>1</sup>, 李志良<sup>1</sup>, 白石<sup>Y2</sup>, 宫下<sup>Y2</sup>, 佐佐木<sup>S2</sup>

(1. 重庆大学 化学化工学院, 重庆 400044;

2. 日本国立技术科学大学化学与材料工程学院, 日本丰桥 441)

**摘要:**采用比较分子场(CoMFA)分析法和偏最小二乘回归(PLS)法对十多种 HMG-CoA 还原酶阻滞剂的构效关系进行了多维定量建模研究。在描述变量数远远超过样本数的情况下,经典的 Hansch 方法等均不适应,而笔者的方法可给出良好结果。

**关键词:**多维多元定量构效关系;比较分子场分析;分子建模

中国分类号:R914.2

文献标识码:A

血清中胆固醇的增加将引起动脉硬化和心肌梗塞等心血管疾病,因此人们期望血清中胆固醇减少。业已发现大部分(约 80%)胆固醇是由肝脏依下列路线(图 1)生化合成的。其过程相当复杂,历经约十多步。其中重要的一步是由 3-羟基-3-甲基戊二酰辅酶 A (HMG-CoA)生成 Mevalonate 酯,该步反应受到 HMG-CoA 还原酶催化,因此研究和开发 HMG-CoA 阻滞剂以能有效地阻止胆固醇的合成而引人注目。

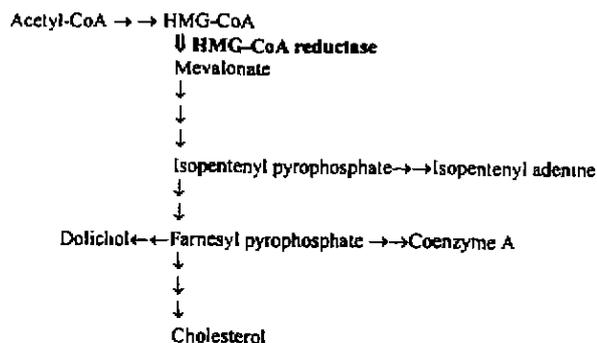


图 1 胆固醇是由肝脏依上述生化路线合成  
(其过程复杂,历经十多步)

目前世界上市售药物有近十种,但各有其优缺点,并非理想药物。研究和创制 HMG-CoA 酶阻滞剂类新药仍是一项重要课题。新近研究了一类新的 HMG-CoA 酶阻滞剂 I, I-双(对氟苯基)-C, E-二羟基-8-取代-F, H-壬二烯酸(8-取代 BFDNA)的阻滞活性。笔者则以比较分子场(CoMFA)及偏最小二乘回归

(PLS)对其进行多维定量构效关系(3D-QSAR)研究,取得良好结果。

## 1 原理与算法

所谓比较分子场(CoMFA)方法是由美国 D. E. Cramer III 教授<sup>[1,2]</sup>提出的一种新颖重要的多维定量构效关系(3D-QSAR)研究方法和计算机辅助药物设计工具<sup>[2]</sup>。该法构思新颖,算法合理。CoMFA 认为(1)产生相互作用通常是非共价作用;(2)非共价相互作用导致生物活性及其变化。在 Cramer 原法<sup>[1]</sup>中只考虑了电子与立体效应,笔者则进一步考察了疏水效应的影响<sup>[3]</sup>,使该法得以扩展则功能更强,效果更好。其主要步骤如下:(1)分子周围的三维空间以格子状(正或长方体)分割,各格子点上以探针原子(probe atom)计算相互作用能。本文采用 +1 价电荷 SP<sup>3</sup> 杂化碳原子(SP<sup>3</sup>-C<sup>+</sup>)作探针。(2)以这些相互作用能作为 3D-QSAR 结构描述子变量并以 PLS 建模。(3)将所得 PLS 模型变换成为多元线性回归形式(MLR-like)的模型,由各变量的系数描绘出等高线,从而使化合物的三维结构性质得以阐明。

### 1.1 三维空间相互作用能

(1) 静电相互作用能 一般用下式表示:

$$E_e = Z_i Z_j e^2 / 4\pi r_{ij} \quad (1)$$

此处 e 为点电荷量,即 e = 1.602 177 33x10<sup>-19</sup>C(库仑),

\* 收稿日期:2001-03-26

基金项目:国家新药基金(96-901-5-4)、教育部春晖计划启动基金(1999-38)、霍英东青年教师基金(1998-03-28)、日本政府文部省与科学振兴会(YC0506-326)资助项目。

作者简介:王远强(1979-),重庆綦江人,重庆大学本科生。

( $\epsilon$  为介电常数, 其真空值

$\epsilon_0 = 8.854187817 \times 10^{-12}$  F/m, 1986调整值),  $r_{ij}$  为两原子之间距离,  $Z_i, Z_j$  为原子相应电荷(例如用分子轨道计算, 则取密立根电荷)。以此可计算

$$E_c = Z_i Z_j e^2 / 4\pi\epsilon_0 r_{ij} = 2.3070956 \times 10^{-28} Z_i Z_j / r_{ij} \quad [\text{J}] \quad (2b)$$

$$E_c = 2.3070956 \times 10^{-28} N A Z_i Z_j / r_{ij} = 6.0221376 \times 10^{23} \times 2.3070956 \times 10^{-28} Z_i Z_j / r_{ij} = 1.3893548 \times 10^{-4} Z_i Z_j / r_{ij} \quad [\text{J/mol}] \quad (2c)$$

$$E_c = 3.3184170 \times 10^{-5} Z_i Z_j / r_{ij} \quad \text{Cal/mol} \quad (2c)$$

若  $r_{ij}$  以  $\text{\AA}$  为单位, 则

$$E_c = 331.84170 Z_i Z_j / r_{ij} \quad [\text{KCal/mol}] \quad (2d)$$

若探针原子(团)带电荷为 +1, 则静电相互作用能可依下式算出:(a) 介电常数一定的场合( $\epsilon = \epsilon_0$ ),

$$E_c = 331.84170 \sum_j^{\text{All atoms}} Z_j / r_{ij} \quad (2e)$$

(b) 介电常数随距离变化( $\epsilon = r_{ij}\epsilon_0$ ),

$$E_c = 331.84170 \sum_j^{\text{All atoms}} Z_j / r_{ij}^2 \quad (2f)$$

(2) 立体相互作用能 立体相互作用能可由下式

$$E_i = k(1.0/r_{ij}^{12} - 2.0/r_{ij}^6) = k(1.0r_{ij}^{-12} - 2.0r_{ij}^{-6}) \quad (3)$$

描述, 此处  $k$  为比例常数,  $r_{ij}$  为两原子  $i, j$  间距离 [ $\text{\AA}$ ]。式(3)中右边第1项为斥力, 第2项为引力。以 Toripos5.2 力场计算探针原子  $i$  所受的作用, 则可以下式替换上式表述:

$$E_i = \sum_j^{\text{All atoms}} (k_i k_j)^{1/2} \{ [(r_i + r_j)/r_{ij}]^{12} - 2.0[(r_i + r_j)/r_{ij}]^6 \} \quad [\text{Kcal/mol}] \quad (3a)$$

此处  $k_i, k_j$  为常数,  $r_i, r_j$  为原子  $i, j$  的 van der Waals 半径 [ $\text{\AA}$ ]。本文 CoMFA 的探针原子若用  $\text{SP}^3\text{C}$ , 则具体为:

$$E_i = \sum_j^{\text{All atoms}} (0.107k_j)^{1/2} \{ [(1.7 + r_j)/r_{ij}]^{12} - 2.0[(1.7 + r_j)/r_{ij}]^6 \} \quad [\text{kcal/mol}] \quad (3b)$$

(3) 疏水相互作用 分子的三维亲水性表示为分子亲脂势  $E = \text{MLP}$  (molecular lipophilicity potential):

$$E_p = \text{MLP} = \sum_{j=1}^{\text{All atoms}} f_j / (1 + d_j) \quad (4)$$

式中  $f_i$  为原子疏水性常数,  $d_j$  为原子  $j$  与探针原子  $C(\text{sp}^3)$  的距离 [ $\text{\AA}$ ]。Ghose 等<sup>[5]</sup> 对有机物中原子类型进行了分类得 110 种。原子的辛醇/水分配系数  $P$  的对数  $\log P$  可由原子疏水性常数  $f_i$  线性组合而成。可由最小二乘(PLS)法算出各  $f_i$  值。由已求出的值及上式(4)并考虑到  $d_{j=0}$  场合可算出分子的  $\log P$  值。作为分子的三

维疏水性表达方法还可参见 HINT 和 GRID 方法, 此处从略。

## 1.2 结构最优化与分子场

(1) 结构最优化: 先求出阻滞剂的最优化结构, 以寻找具有生理活性的活性构型(active conformation)。进行构造最优化时, 首先对化合物中主键  $C_8 - C_9$  有关的二面角  $\varphi_1, \varphi_2$  及  $\varphi_3$  以一定刻度进行回转, 求得生成热最小的构型, 然后对此构型作结构最优化处理导出最优(适)结构。此处生成热的计算和结构最优化有 MOPAC AM1 完成。(2) 重合过程: 各阻滞剂分子的重合过程按如下步骤完成。此处分子的重合采用 GUMMOS 的 ROT/MOVE 命令,  $1^\circ$  各分子以  $C_8$  处进行重构, 以  $C_8$  为原点;  $2^\circ$  取代基的轴方向以  $C_8$  定位则  $C_8 -$  (轴(键)与轴重合;  $3^\circ$  在  $C_9$  原子的  $x - y$  平面上的位置, 使分子以  $C_8 -$  (轴(键)即轴回转, 其中  $C_9$  原子位于(轴的下方向。分子的重合位置及坐标从略。(3) 场的建立或作成: CoMFA 解析的原子探针为带电荷 +1 的  $\text{SP}^3$  杂化  $C$ 。在本研究中考察了立体、静电和疏水相互作用这 3 种场, 对单一或组合的三类场均进行解析。其中静电的相互作用能以介电率( $\epsilon$ )一定算出, 其 Milliken 电荷值由 MOPAC AM1 算出<sup>[6]</sup>。格子点上立体与静电相互作用能超过  $\pm 30\text{kcal/mol}$  场合则改为  $\pm 30\text{kcal/mol}$ 。

## 1.3 偏小二乘回归(PLS)<sup>[10,11]</sup>

偏小二乘回归(PLS) 参见文献<sup>[8,9]</sup> (1) 数据标准化处理: 采用自定标化(autoscaling) 处理, 使(a) (的描述变量均值为 0; (b) 三类型场的方差一定; (c) 活性的均值为零, 方差为 1。具体表达式为:

$$X_{ik} = x_{ik} - x_{ik} \quad (i = 1 \sim n, k = 1 \sim d) \quad (5a)$$

$$\sum_{i,k} (x_{ik}^a - x_{ik}^c)^2 = \sum_{i,k} (x_{ik}^a - x_{ik}^c)^2 =$$

$$\sum_{i,k} (x_{ik}^b - x_{ik}^d)^2 \quad (i = 1 \sim n, k = 1 \sim g) \quad (5b)$$

$$y_i = (y_i - \bar{y}) / \sigma_y \quad (i = 1 \sim n) \quad (5c)$$

式中  $n$  为样本数,  $d$  为全部描述子变量数,  $g$  为格子点数,  $\sigma_y$  为  $y$  的标准差(自由度  $d_f = n - 1$ )。(2) 偏小二乘建模(PLS modelling): 由于描述子变量数远远超过了样数( $m > n$ ), 故采用经典 Hansch 法(MLR) 等均不适宜。故采用新颖的化学计量学方法之一即偏最小二乘(PLS) 进行此复杂建模:

$$X = \sum_{a=1}^{a=4} t_a p_a + e \quad (6)$$

$$y = \sum_{a=1}^{a=4} b_a t_a p_a + f \quad (7)$$

此处  $t_a, p_a$  为特征(latent) 变量,  $b_a$  为灵敏参数。

## 1.4 软件与程序

本研究使用 CoMFA 程序为自编 CoMFA - HUTUT。编制程序的软件环境分别如下:其中(1)生成程序为 OS = VAX/VMS,语言为 VAX/FORTRAN;(2)PLS 解析程序为 OS = VAX/VMS,语言为 VAX/FORTRAN;作图程序为 MS - DOS,语言为 BASICA 或 VAX/FORTRAN。

## 2 数据获取与处理

### 2.1 酶阻滞剂

新 HMG - CoA 还原酶阻滞剂及其阻滞活性以

表 1 HMG - CoA 还原酶阻滞剂活性

编号	取代基 Substituent X	VO	VN	VOP	VNP	- Epol	IC - 50	活性 Activity	pIC50
1	3a 1 - Me - Tet; 3e0.023	102.47	0.00	93.73	8.74	1.70	0.037/43	7.43	
2	40 Tet	88.58	0.00	79.81	8.74	1.31	0.33	6.48	
3	59 iPr	90.70	0.00	82.23	8.43	1.38	0.20/23	6.70	
4	Ph	96.95	11.99	86.42	22.80		3.0	5.48	
5	2,4 - diCl - Ph	103.49	34.48	91.41	42.65	140.0	3.85		
6	2 - Me - Ph	104.29	18.58	91.18	31.05		9.6	5.02	
7	3 - Me - Ph	99.91	22.82	88.32	34.51		25.0	4.60	
8	6 CN	65.44	0.00	64.54	1.00	0.36	8.0	5.10	
9	58 Me	63.45	0.00	62.52	1.00	0.39	7.1/7.9	5.15	
10	42 1 - Et - Tet	98.48	16.87	89.66	25.64		1.1	5.95	
11	44 1 - iPr - Tet	103.85	25.31	95.00	34.15		14.0/13.	4.85	
12	3b 2 - Me - Tet	87.73	14.58	78.87	23.41		30.0	4.52	
13	43 2 - Et - Tet	90.34	25.79	81.50	34.58		200.0	3.70	
14	H	49.77	0.00			0.05		4.59	
15	Et	80.00	0.00					6.39	
16	Bt	98.15	6.11					6.30	
17	CHF2	71.08	0.00			0.58		5.11	
18	CF3	74.01	0.00			0.72		5.82	
19	53 1 > ph, ph:0.19						0.19		
20	1 > 4F, H; > 250						> 250		
21	1 > H, 4F; 210						210		
22	54 Di[2F4MPh]						0.58		
23	Di[4F3MPh]						0.16		
24	46 2Pi - Tet						> 300		
25	47 1 - MEM						3.7		
26	48 2 - MEM						240		
27	49 2 - Bt						> 300		
28	50 2[2M4FPh]						0.029		
29	51 Ph, 2 - 4FPh						0.044		
30	52 4FPh, Ph						0.051		

### 2.2 CoMFA 描述变量

将上述最优化结构作为分子场,以  $SP^3$  杂化 C(+I) 原子作探针计算其立体、静电、疏水相互作用能 ( $E_s, E_e, E_p = MLP$ ), 在三维点上均是描述变量元。场的范围和格点间隔选取内见表 2。此模型称为 Model1, 则建模的三维分子场总格点数为  $N_g = 567$ , 每点均含电子、立体、疏水三类参数, 则总描述(子)变量数目高达  $m = 3 \times 567 = 1601$  个。而所考察的对象仅 13 种化合物(即样本数为  $n = 13$ )。可见其描述子数远远超过了样本数( $m = 1601 \gg n = 13$ )。

$pIC_{50}$  表示, 参见图 2, 并列于表 1。其中 X 为取代基, 可分别为四唑 - 5 - 基(Tet), 1 - 甲基四唑 - 5 - 基, 异丙基, 苯基, 2,4 - 二氯苯基, 2 - 甲基苯基, 3 - 甲基苯基, 氰基, 1 - 乙基四唑 - 5 - 基, 1 - 丙基四唑 - 5 - 基, 2 - 甲基四唑 - 5 - 基, 2 - 乙基四唑 - 5 - 基等 13 种。以 C8 为原点, 以 C8 - (轴(键)为轴回转, 寻找能量最低构型, 从而导出最优结构。此过程以 MOPAC AM1 完成。

### 2.3 处理方法

由样本数远小于变量数( $n \ll m$ ), 故一般方法包括经典 Hansch 方法或多元线性回归(MLR) 均不适应<sup>[4]</sup>。现代化学计量学方法如主成份回归(PCR), 偏最小二乘回归(PLS) 可发挥重要作用<sup>[4]</sup>。笔者采用 PLS 对此复杂体系进行建模与预测, 它亦是笔者所用 CoMFA 方法的重要组成部分。

## 3 结果与讨论

### 3.1 CoMFA及PLS处理结果

用CoMFA及PLS对表1所示13/18种HMG-CoA还原酶抑制剂取代BFDNA化合物活性进行多维QSAR建模。其校正相关系数 $R_c$ 随模型复杂性或组分数目 $A$ 的增加而增加。说明,采用的组分数越多,模型对体系拟合效果越好。这意味着在获取有关有用信息的同时,可能拟合了有害的噪声情形,即所谓“过拟”现象。当采用“离去”(leave-one-out)的交互校验(cross-validation)方法<sup>[9]</sup>进行预测时,发现其预测相关系数 $R_p$ 则以组分数 $A=3$ 效果最佳。(均见表3,作用从略)。以交互校验预测误差PRESS对组分数 $A$ 作图亦知 $A=3$ 时最小。故宜采用 $A=3$ 的Model1模型。

表2 CoMFA的场范围及点间隔

场范围(Å)			格点间隔(Å)		总格点数( $N_t$ )
x	y	z	dx	dz	
[-10, +6]	[-8, +8]	[-6, +6]	2.0		567

表3 CoMFA模型的预测与校正相关系数

组分数 $A$	1	2	3	4	5	6
校正相关系数 $R_c^2$	0.751	0.835	0.915	0.970	0.986	0.995
预测相关系数 $R_p^2$	-0.241	0.335	0.455	0.429	0.435	0.344

$(pF-Ph)_2CH-CZ=CH-(CH=CH)-CHOH-(CH_2)_n-(CHOH)_n-CH_2-COOH$

其中X为取代基,可分别为四唑-5-基,1-甲基四唑-5-基,异丙基,苯基,2,4-二氯苯基,2-甲基苯基,3-甲基苯基,氨基,1-乙基四唑-5-基,1-丙基四唑-5-基,2-甲基四唑-5-基,2-乙基四唑-5-基等

表4 CoMFA模型的单因素影响

组分数	1	2	3	4	5	6
立体场	$R_c$ 0.747	0.876	0.986	0.987	0.994	0.999
	$R_p$ 0.343	0.417	0.539	0.579	0.667	0.675
静电场	$R_c$ 0.688	0.819	0.892	0.938	0.960	0.979
	$R_p$ -0.320	-0.554	-0.442	-0.573	-0.561	-0.613
疏水场	$R_c$ 0.366	0.691	0.866	0.893	0.930	0.967
	$R_p$ -0.389	-0.042	0.449	0.644	0.219	-0.673

### 3.2 单因素影响

若仅考虑立体、电子、疏水因素的三者之一的单因素影响,仍以表2所示分子场范围扩格子点间隔,分别以CoMFA及PLS处理,其结果见表4。(预测误差PRESS随组分变化从略)。发现校正相关系数均随组分数增大而增加,而预测相关系数( $R_p^2$ )则由于因素不同而改变:(a)立体因素中预测相关关系( $R_p^2$ )随组分数增加而增加;(b)静电因素中 $R_p^2$ 出现负值且呈无规律变化,说明该因素不利于建模;(c)疏水因素中 $R_p^2$ 在 $A=1,2,6$ 出现负值,并非所期望;但 $A=4$ 时,最大( $R_p^2=0.644$ )。

### 3.3 双因素场影响

今考察由立体·电子·疏水三因素中任意两因素

场组合的影响,仍以CoMFA及PLS处理,得到结果见表5。发现校正情况同A、B;预测情况则以立体与疏水场的结果较好且几乎不存在过拟现象( $A=1\sim 6$ ),但 $A=1$ 时为负值,不宜采用。

### 3.4 三因素场影响

进一步考察由立体·电子·疏水三因素中全部因素场组合的影响,仍以CoMFA及PLS处理,得到结果见表6。发现校正情况同A、B、C;预测情况几乎不存在过拟现象( $A=1\sim 6$ ),当 $A=3$ 时,预测能力最强,相关系数最大( $R_p^2=0.456$ )但较两因素最好结果的立体与疏水组合场为劣,另外 $A=1$ 时亦为负值,不宜采用。

表5 CoMFA模型中双因素的影响

CoMFA场中 二因素	组分数(A)					
	1	2	3	4	5	6
立体与静电场	$R_c$ 0.730	0.868	0.954	0.975	0.989	0.995
	$R_p$ 0.061	-0.217	0.365	0.303	0.439	0.373
立体与疏水场	$R_c$ 0.657	0.829	0.934	0.976	0.992	0.995
	$R_p$ -0.067	0.473	0.629	0.641	0.658	0.663
静电与疏水场	$R_c$ 0.624	0.797	0.868	0.958	0.973	0.985
	$R_p$ -0.220	0.232	-0.053	-0.511	-0.504	-0.664

表6 CoMFA模型中三因素的影响

CoMFA场中 三因素	组分数(A)					
	1	2	3	4	5	6
静电/立体/ 疏水场	$R_c$ 0.751	0.835	0.915	0.970	0.986	0.995
	$R_p$ -0.124	0.335	0.456	0.429	0.435	0.344

### 参考文献:

- [1] C III R D, P D E., B J D Comparative molecular field analysis (CoMFA) 1. effect of shape on binding of steroids to carrier proteins, *J Am Chem Soc.*, 1988, 110: 595-9.
- [2] C M, C III R D, O N V Validation of the general purpose tripos 5.2 force field[J]. *J Comput Chem*, 1989, 10: 982.
- [3] F P, S A, C N C. 3D molecular lipophilic potential profiles: a new tool in molecular modeling[J] *J Mol Graphics*, 1988, 6: 182.
- [4] G A K, C G M Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships I. partition coefficients as a measure of hydrophobicity[J] *J Comput Chem*, 1986, 7: 565.
- [5] G P, K B R. An example of 2-block predictive partial least squares regression with simulated data[J]. *Anal Chim Acta*, 1986, 185: 19.
- [6] L W, P J A, W S. Partial least squares method for spectrophotometric analysis of mixtures of humic acid and ligninsulfonate[J]. *Anal. Chem.*, 1983, 55: 643.
- [7] W S, R A, W H, D III W J, The collinearity problem in linear regression. the partial least squares approach to generalized inverse[J]. *SIAM J Sci Statist Comp*, 1984, 5: 735.

- [8] D, W S, E U, H S., G J. Multivariate structure - activity relationships between data from a battery of biological tests and an ensemble of structure descriptors; the PLS method[J]. Quant Struct - Act. Relat., 1984, 3: 131.
- [9] M R Analysis of two partial - least - squares algorithms for multivariate calibration[J]. Chemom. Intell Lab Syst, 1987, 2; 187.
- [10] W, G P, E K, O J. Multi - way principal components - and PLS analysis[J], J.Chemom, 1987, 1: 41
- [11] A.PLS regression methods[J]. J Chemom., 1988, 2: 211.
- [12] D M. Thomas E V.Partial least - squares methods for spectral analyses. I. relation to other quantitative calibration methods and the extraction of qualitative information[J]. Anal Chem, 1988, 60: 1 193.
- [13] W G, D III W J. S D R. Principal components analysis and partial least squares regression [J]. Tetrahedr. Comput. Method., 1989, 2: 349.
- [14] D, G M, N B, R, G J A. Quantitative structure - activity relationship (QSAR) analysis using the partial least squares (PLS) method: the binding of polycyclic aromatic hydrocarbons (PAH) to the rat liver 2, 3, 7, 8 - tetrachlorodibenzo - p - dioxin (TCDD) receptor[J]. Quant Struct - Act Relat, 1989, 8: 83.
- [15] M Y, H, C, S S. Multivariate structure - activity relationships analysis of fungicidal and herbicidal thiocarbamates using partial least squares method [J]. Quant. Struct. - Act. Relat., 1992, 7 - 18.
- [16] S J J P. Quntum Chemistry Program Exchange: Program number 455, Indiana University, Bloomington, Indiana; Aikawa Y, Japan Chemistry Program Exchange: P023 (GUMMOS Ver 4.18)
- [17] L Z, L M, S L. Multivariate chemometrics and artificial intelligence for QSAR studies[A], Proc 34 IUPAC, 1993, O7004: 879.
- [18] Multicompoent pharmaceutical and multicompartent pharmacokinetic analysis, Proc 34 IUPAC, 1993, F2607: 350.
- [19] B, E U, J D, W S. Simplified carbon - 13 NMR parameter related to the carcinogenic potency of polycyclic aromatic hydrocarbon[J]. Quant. Struct. - Act. Relat., 1983, 2: 73.
- [20] E L, J J, S, L F, S B, S M., Wold S.; A strategy for ranking environmentally occurring chemical part v. the development of two genotoxicity QSARs for halogenate aliphatic [J]. Acta Chem. Scand., 1990, 44: 50.
- [21] W S. Cross-validatory estimation of the number of components in factor and principal components models[A], Technometrics, 1978, 20(4), 897.

## Multidimensional Multivariate Quantitative Structure - Activity Relationships of HMG-CoA Reductase Inhibitors

WANG Yuan-qiang<sup>1</sup>, YE Nan<sup>1</sup>, LI S. Zhi-liang<sup>1</sup>, MIYASHITA Y<sup>2</sup>, SASAKI S<sup>2</sup>

(1. College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400044, China;

2. College of Molecular Engineering, National University of Technology, Toyohashi 440, Changsha 410082, Japan)

**Abstract:** Multidimensional multivariate quantitative structure - activity relationships (MD/MV QSAR) are modeled for thirteen HMG-CoA reductase inhibitors using partial least squares regression (PLS) and comparative molecular field analysis (CoMFA) modified by introducing molecular lipophilicity potential (MLP) as a new hydrophobic field. Under condition that the number ( $m$ ) of variables is much greater than the number ( $n$ ) of samples,  $m \gg n$ , many techniques such as the classical Hansch multiple linear regression (MLR) method, are not suitable; whereas the developed methods based on the combination of the modified CoMFA and latent PLS modelling suitable and applicable, which provides quite good results for calibration and prediction.

**Key words:** multidimensional multivariate quantitative structure - Activity relationships; HMG-CoA reductase inhibitors; multidimensional multivariate molecular modelling lead optimization partial Least squares(PLS) modified comparative molecular field analysis (CoMFA) molecular lipophilicity potential (MLP) hydrophobic interaction.

(责任编辑 张小强)