

# 用支持向量机预测 HIV-1 整合酶抑制剂活性

刘 昕, 谭建军, 陈慰祖, 刘 斌, 李 杉, 王存新  
(北京工业大学 生命科学与生物工程学院, 北京 100124)

**摘 要:** HIV 整合酶可催化病毒复制周期中的整合过程, 即将 HIV 反转录产物 cDNA 整合入宿主基因组, 它是病毒复制过程中不可缺少的酶, 也是抗 HIV 药物设计的重要靶点. 构建嘧啶酮类 (pyrimidones) HIV-1 整合酶抑制剂定量构效关系模型, 有助于进一步了解影响抑制剂活性的结构因素. 本文应用 CoMFA 软件计算了 68 个化合物的拓扑、分子极化、亲水性等结构参数, 用所选的结构参数作为支持向量机 (support vector machine, SVM) 的输入, 建立起非线性的支持向量机回归模型. 研究表明: 支持向量机算法与分子结构参数的有机集成, 可为 HIV 整合酶抑制剂的结构与活性数据建立起预测模型, 为抗 HIV 药物设计提供生物学信息.

**关键词:** 支持向量机; HIV-1 整合酶抑制剂; 定量构效关系; 嘧啶酮

中图分类号: R 914.2

文献标志码: A

文章编号: 0254-0037(2013)04-0634-07

## Support Vector Machine Applied to Predicting the Activity of HIV-1 IN Inhibitors

LIU Xin, TAN Jian-jun, CHEN Wei-zu, LIU Bin, LI Shan, WANG Cun-xin  
(College of Life Science and Bio-engineering, Beijing University of Technology, Beijing 100124, China)

**Abstract:** HIV integrase (IN) catalyzes the integration process in the viral life cycle. IN helps the viral reverse transcription product cDNA integrating into the host chromosome. As an indispensable enzyme, IN is also an important target for designing and developing the novel anti-HIV drugs. Constructing the QSAR model of the HIV-1 IN pyrimidones inhibitors can help for understanding of the structural factors. In this paper, the CoMFA software was used to calculate topological descriptors, polarizable descriptors and hydrophilic descriptors and other structural parameters for 68 compounds. The structural parameters were selected as inputs of support vector machine (SVM) to establish the non-linear regression model. Results show that the SVM algorithm combined with QSAR can establish the forecasting model for inhibitors, and provide biological information for the design of anti-HIV drugs.

**Key words:** support vector machine (SVM); HIV-1 integrase inhibitors; quantitative structure activity relationship (QSAR); pyrimidones

HIV 利用整合酶 (integrase, IN) 通过 3'-加工 (3'-processing)、链转移 (strand transfer, ST) 和切口修复 (gap repairing, GR) 3 个过程将自己的遗传物质整合到受感染的细胞中, 它在病毒复制过程中起到重要作用. 由于人类细胞中没有 HIV 整合酶类

似物, 因此, 设计以整合酶为靶点的抗 HIV 药物对人体正常细胞毒性较小, 同时整合酶采用单一的活性位点去适应 2 种不同的 DNA 底物<sup>[1]</sup>. 这在一定程度上限制了 HIV 产生对整合酶抑制剂的耐药株, 所以整合酶是研发抗 HIV 药物一个非常有意义的靶点.

收稿日期: 2011-12-07.

基金项目: 国家自然科学基金资助项目 (21173014); 科技部国际合作项目 (2010DFA31710).

作者简介: 刘 昕 (1974—), 女, 讲师, 主要从事分子设计和蛋白质组学方面的研究, E-mail: dbliuxin@bjut.edu.cn.

通信作者: 王存新 (1943—), 男, 教授, 主要从事生物物理学方面的研究, E-mail: cxwang@bjut.edu.cn.

近年来,以 HIV-1 整合酶为靶标的抗 HIV-1 药物研究进入到一个崭新的时期,人们不断寻找新的具有抑制活性的化合物,很多抑制剂被筛选出来,一些抑制剂已经进入临床研究. 迄今为止所报道的 HIV-1 整合酶抑制剂,主要有二酮酸(Diketo Acid, DKA)及二酮酸衍生物类整合酶抑制剂、多羟基化芳香族化合物类整合酶抑制剂、苯甲酸衍生物类整合酶抑制剂、核苷酸类整合酶抑制剂和肽类整合酶抑制剂<sup>[2]</sup>. 其中,只有二酮酸类化合物显示有效的细胞内抗病毒活性,实验表明该类化合物通过抑制整合酶 2 个催化功能中的链转移而获得抗病毒活性的<sup>[3-4]</sup>. 由于二酮酸结构对整合酶的抑制机理明确,因而成为具有成药前景的药效团. 美国 Merck 公司对数十万个类似二酮酸结构的小分子化合物进行随机筛选,得到了第 1 批具有整合酶链转移抑制活性的化合物 L-371988 和 L-708906<sup>[5]</sup>. 2007 年 10 月,美国 Merck 公司的整合酶抑制剂 Raltegravir (MK-0518) 获得美国 FDA 批准上市,成为唯一的以整合酶为靶点的抗 HIV 药物<sup>[6]</sup>.

本文利用计算机辅助药物设计的基本原理,针对二酮酸类化合物嘧啶酮的结构特点,通过模拟药物与受体生物大分子的相互作用,用支持向量机(support vector machine, SVM) 回归方法分析结构与活性的内在联系,为今后预测化合物活性、药物设计、分子结构改造、耐药研究等提供一定的帮助.

## 1 原理与方法

### 1.1 支持向量机

Vapnik 等<sup>[7]</sup>从 20 世纪 60 年代末开始研究小样本情况下机器学习的基本理论,到 90 年代中期在统计学习理论的基础上提出了支持向量机算法. 它能较好地解决小样本、非线性和高维模式识别问题,在坚实的理论基础上建立简单明了的数学模型,使得支持向量机的应用研究受到广泛的关注,在人脸识别、纹理图像分析、故障诊断、信用风险评估等方面得到较好的应用,目前在计算机辅助药物设计中被广泛应用<sup>[8-11]</sup>.

SVM 应用于分类问题上,在线性可分的情况下, SVM 就是建立一个最优分类超平面,使得可分的 2 类数据到该平面的最小距离最大化. 对于非线性问题, SVM 解决思路是通过一个非线性映射把原始数据映射到一个高维的特征空间,使得新数据集在该特征空间上是线性可分的,由此建立的最优分

离超平面在原始空间内就是一个超曲面<sup>[12-13]</sup>.

支持向量机不仅可解决分类问题,同时也可用来解决回归问题,解决的思路与分类问题相似<sup>[14]</sup>. 但与分类问题不同的是,回归所求的超平面是使所有样本点到超平面的距离为最小. 对于线性回归问题<sup>[15]</sup>, 给定样本集  $(x_i, y_i)$ , 其中  $i = 1, 2, \dots, n$ ;  $x_i \in \mathbb{R}^d$  是第  $i$  个学习样本的输入值,且为一个  $d$  维列向量  $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T$ ;  $y_i \in \mathbb{R}$  是对应的目标值,寻求一个最优超平面,使得在给定精度  $\varepsilon$  ( $\varepsilon \geq 0$ ) 条件下无误差地拟合  $y$ , 即所有样本点到最优超平面的距离都不大于  $\varepsilon$ .

此时的寻优问题变为最小化模型复杂度,其等价于  $\min_{w, b} \frac{1}{2} \|w\|^2$  转化成相应的二次规划问题,即

$$\min_{w, b} \frac{1}{2} \|w\|^2$$

$$\text{s. t. } y_i((w \cdot x_i) + b) \geq 1 \quad i = 1, 2, \dots, l$$

考虑到允许误差的情况,可引入松弛变量  $\xi$  和惩罚系数  $C$  相应凸二次规划问题为

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

$$\text{s. t. } ((w \cdot x_i) + b) - y_i \leq \varepsilon + \xi_i \quad i = 1, 2, \dots, l$$

$$y_i - ((w \cdot x_i) + b) \leq \varepsilon + \xi_i^* \quad i = 1, 2, \dots, l$$

$$\xi_i, \xi_i^* \geq 0$$

将上面的优化问题转化成对偶问题后可解得最优回归函数为

$$f(x) = (w \cdot x) + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (x \cdot x_i) + b$$

式中:  $0 \leq \alpha_i, \alpha_i^* \leq C$ . 与分类问题一样,这里也只有小部分样本支持向量的  $\alpha_i, \alpha_i^*$  不为 0. 对于非线性回归,可通过支持向量机的核心技术即引入核函数将样本映射到高维特征空间求解. 目前常用于支持向量机的核函数主要有以下 4 种类型<sup>[16]</sup>:

线性核(linear kernel):  $K(x, x_i) = x \cdot x_i$ ;

多项式核(polynomial kernel):  $K(x, x_i) = (\alpha_1 x \cdot x_i + \alpha_2)^p$ ;

径向基核(RBF kernel):  $K(x, x_i) = \exp(-\gamma \|x - x_i\|^2)$ ;

Sigmoid 核(sigmoid kernel):  $K(x, x_i) = \tanh(\alpha_1 x \cdot x_i + \alpha_2)$ .

定量构效关系(quantitative structure activity relationship, QSAR)借助分子的结构参数,用数学或者统计学方法定量研究分子结构与生物活性之间关系. 到目前为止,主成分分析、偏最小二乘法、启发

式方法、人工神经网络等方法在 QSAR 研究中已十分广泛,但也存在一定的局限性<sup>[17-18]</sup>。随着药学研究的发展,基于量子化学计算的分子模拟和计算机辅助设计给 QSAR 研究带来了引人注目的成就。本实验采用台湾大学林智仁提供的网络共享算法 libSVM 结合 QSAR 研究中的分子力场研究对 68 个嘧啶酮类(pyrimidones) HIV-1 整合酶抑制剂的生物活性进行了定量构效关系建模。

## 1.2 样本数据来源和结构参数计算

样本数据选用一组相似化合物嘧啶酮类(Pyrimidones) HIV-1 整合酶抑制剂,包括 68 个化合物分子的结构及其相应的生物活性值,即  $IC_{50}$  ( $\mu\text{mol/L}$ )。为了保证这些生物数据的一致性,选用

的化合物全部由 Merck 研究所合成<sup>[19-20]</sup>,为了利用算法得到较好的平均分布情况,抑制剂分子的活性用  $pK_i$  ( $-\lg IC_{50}$ ) 表示。

为了建立分子模型,首先将化合物(如表 1 所示)分为训练集和测试集。训练集分子用来建立预测模型,测试集分子用来判断模型的预测能力。训练集最终包含 52 个分子,测试集包括 16 个分子(表 1 中带\* 的分子)。分组过程中,先根据化合物的结构将结构相似的分子分成几大类,同样可按照化合物的活性值将所有分子进行递增排列。然后,结合 2 种不同分类方式选择训练集和测试集分子,选择时应尽可能包含所有结构类型,同时保证 2 个集合的分子活性值均能覆盖整个活性值范围。

表 1 用于构效关系研究的生物分子的结构及其抑制活性

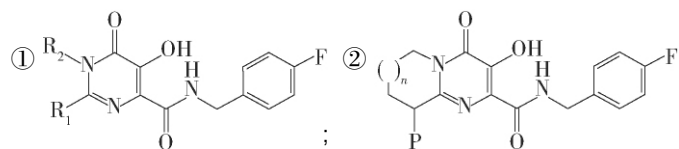
Table 1 Structure of biomolecular and inhibitory activity

公共骨架 1 <sup>①</sup>					公共骨架 2 <sup>②</sup>				
No.	R <sub>1</sub>	R <sub>2</sub>	IC <sub>50</sub> / ( $\mu\text{mol}\cdot\text{L}^{-1}$ )	pK <sub>i</sub>	No.	P	n	IC <sub>50</sub> / ( $\mu\text{mol}\cdot\text{L}^{-1}$ )	pK <sub>i</sub>
001		CH <sub>3</sub>	0.004	2.40	038		0	0.007	2.15
005*		CH <sub>3</sub>	0.041	1.39	039		2	0.550	0.26
006		CH <sub>3</sub>	0.032	1.49	040		1	0.063	1.20
007*		CH <sub>3</sub>	0.070	1.15	041		1	0.004	2.40
008		CH <sub>3</sub>	0.012	1.92	042		1	0.005	2.30
010		CH <sub>3</sub>	0.011	1.96	043		1	0.004	2.40
012		CH <sub>3</sub>	0.008	2.10	044		2	0.062	1.21
014*		CH <sub>3</sub>	0.060	1.22	045	H	0	0.008	2.10
015		H	0.030	1.52	046	H	1	0.005	2.30
016*		CH <sub>3</sub>	0.050	1.30	047	H	2	0.019	1.72

续表 1

公共骨架 1 <sup>①</sup>					公共骨架 2 <sup>②</sup>				
No.	R <sub>1</sub>	R <sub>2</sub>	IC <sub>50</sub> / ( $\mu\text{mol}\cdot\text{L}^{-1}$ )	pK <sub>i</sub>	No.	P	n	IC <sub>50</sub> / ( $\mu\text{mol}\cdot\text{L}^{-1}$ )	pK <sub>i</sub>
017		CH <sub>3</sub>	0.033	1.48	048		1	0.012	1.92
018		CH <sub>3</sub>	0.015	1.82	049		2	0.016	1.80
019		CH <sub>3</sub>	0.020	1.70	050*		1	0.007	2.15
023		CH <sub>3</sub>	0.050	1.30	052		1	0.005	2.30
024*		CH <sub>3</sub>	0.021	1.68	054*		1	0.014	1.85
025*		CH <sub>3</sub>	0.062	1.21	055		2	0.006	2.22
026*		CH <sub>3</sub>	0.100	1.00	057*		1	0.009	2.05
028		CH <sub>3</sub>	0.010	2.00	061*		0	0.009	2.05
029		CH <sub>3</sub>	0.190	0.72	062		1	0.006	2.22
030*		CH <sub>3</sub>	0.100	1.00	063*		2	0.008	2.10
035		CH <sub>3</sub>	0.020	1.70	064		1	0.003	2.52
036*		CH <sub>3</sub>	0.020	1.70	066*		1	0.009	2.05
037		CH <sub>3</sub>	0.030	1.52	067*		2	0.017	1.77
					068		1	0.007	2.15

注: 带\* 的分子为测试集分子, 其他为部分训练集分子.



所有分子结构均在 SGI 工作站上, 由 LINUX 系 统运行 SYBYL7.3 软件包的 CoMFA 进行计算. 分

子能量优化处理中,将为所有原子赋予电荷,优化过程中化合物被赋予 Gasteiger-Hückel 电荷,用 Tripos 力场和最陡下降法进行 5 000 步能量优化,能量变化收敛标准为 0.041 8 kJ/mol,再用共轭梯度法进行 1 000 步能量优化,收敛标准为 0.021 kJ/mol.

### 1.3 描述符的筛选

在定量构效关系研究中,十分重要的一个环节是对分子结构特征进行描述,描述符的选择与确定是关键.本文利用 SYBYL7.3 软件包的 CoMFA 模块计算 68 个嘧啶酮类化合物的拓扑、静电、分子极化、亲水性等结构参数.在文献 [21] 的基础上选用 6 个描述符对化合物的分子结构进行表征.

分子总表面积 (total area),主要描述分子的表面情况及活性,通常情况下表面积越大分子活性越强;正电荷分布表面积 (total charge weighted partial positive surface area),是一个与正电荷相关的参数,代表受体与化合物之间的结合能力,值越大,分子的结合作用越强烈;亲水区 (hydrophilic region) 和极化度 (polarizability) 标定分子聚集和分解的能力及分子极化的能力,能反映氢键受体和供体区域分布;局部最小作用能 (local interaction energy minimum)、相互作用能矩 (integy moment),是反映分子偶极矩及亲水性的指标.

由于不同描述符的量纲不同,数值相差较大,为了提高建模的准确率,本文对数据进行归一化处理,即以每个变量与样本平均值的差再除以样本方差.

## 2 结果与讨论

### 2.1 SVM 模型的建立

根据 CoMFA 的计算,首先将分子场的数值作为自变量,将活性值作为因变量,然后选择留一法 (leave-one-out) 进行交叉验证分析,将主成分值设为 5,并设置 Column Filtering 的值为 2.0 kcal/mol (1 cal = 4.186 8 J).得到交叉验证系数  $q^2$  为 0.672,非交叉验证相关系数  $r^2$  为 0.912,预测值的标准误差 SEE 为 0.246,显著性检验  $F$  为 137.652.

一般来说, $r^2$  值越大,SEE 值越小,表示相关性越好,模型的预测能力越强.而  $F$  主要用于判断样本间是否有显著差别,其值越大,说明样本间差别越显著.所得模型的交叉验证系数  $q^2 > 0.5$ ,非交叉验证系数  $r^2 > 0.9$ ,且估计标准误差较小,表明该模型具有较高的可靠性和预测能力.

由于径向基 RBF 核函数可工作在高维空间并且善于处理非线性函数依赖,本文选择径向基作为

核函数.对  $pK_i$  建立 SVM 模型时,本实验通过结合线性搜索法完成参数寻优,优化后的参数值分别为 best  $C = 0.80$ , best  $\gamma = 0.52$ ,所得模型的交叉验证相关系数平方  $q^2$  (squared correlation coefficient) 和均方根误差 (mean squared error, MSE) 分别是:训练集  $q^2 = 0.911$ , MSE = 0.040; 测试集  $q^2 = 0.833$ , MSE = 0.036.可见,使用径向基核函数 SVM 模型对数据的预报能力较强.

### 2.2 活性预测

本实验获得训练集活性预测值与实验值  $R^2 = 0.937$ ,测试集的活性预测  $R^2 = 0.804$ ,都达到较高的相关性,如图 1、2 所示.

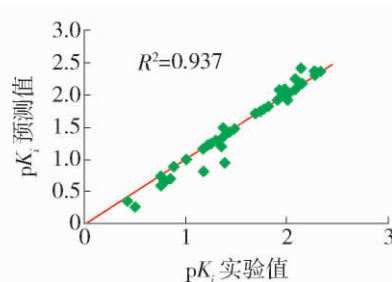


图 1 训练集分子活性预测值与实验值相关性  
Fig. 1 Correlation of experimental and predicted  $pK_i$  in the training set

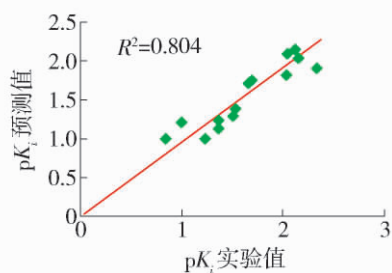


图 2 测试集分子活性预测值与实验值相关性  
Fig. 2 Correlation of experimental and predicted  $pK_i$  in the test set

整合酶活性中心是由 DDE 基序整合 2 个二价金属离子组成,是整合酶发挥催化功能的区域.研究表明<sup>[22]</sup>,二酮酸类化合物结构中的 1,3-二羰基能与整合酶催化核心区 DDE 基序上的二价金属结合,从而阻止了催化活性位点与宿主 DNA 结合,选择性抑制了链转移的过程(图 3).

二酮酸类化合物中羧基的存在一般使化合物在体内的生物利用度不高,羧基如果被其他含氮杂环结构取代,则许多化合物具有较好的生物活性<sup>[23]</sup>,尽管这些化合物不属于典型的二酮酸类结构,但都具有明显的二酮酸电子等排体结构特征<sup>[24]</sup>.本实

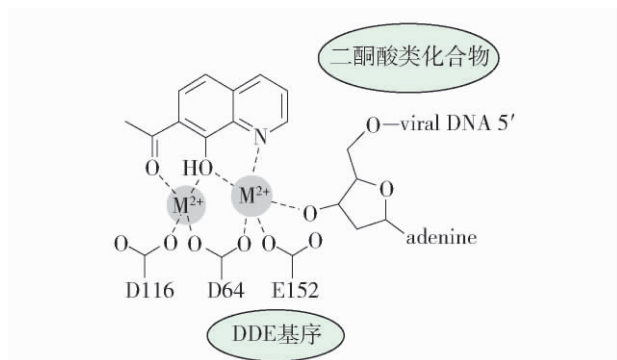


图3 二酮酸类化合物与位于整合酶活性位点的金属离子结合

Fig. 3 Diketo acids chelating the metal ions inside the integrase active site

验中,以 N-甲基嘧啶(图4)为母核结构的化合物 001 对链转移的  $IC_{50}$  为  $0.001 \mu\text{mol/L}$ ,二羟基嘧啶结构化合物 015 对链转移的  $IC_{50}$  为  $0.03 \mu\text{mol/L}$ .

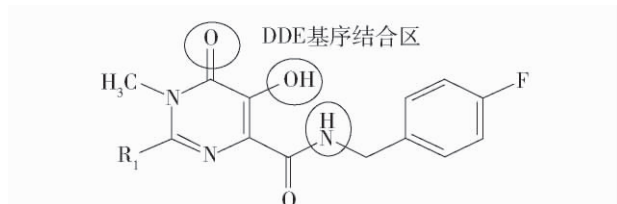


图4 N-甲基嘧啶母核

Fig. 4 N-methylpyrimidone

在嘧啶酮类化合物中,模板 N-甲基嘧啶母核中二酮酸类似结构通过结合二价金属离子与整合酶核心区 DDE 基序结合<sup>[25]</sup>,而对位氟苯酰胺结构参与了临近疏水口袋的相互作用.杂环上的 2 位,即  $R_1$  位置可有各种取代.

利用支持向量机对嘧啶酮类化合物活性预测所得预测值与实验值有较高的一致性,可见该模型通过训练学习,能较好地反映资料本身所蕴含的变化规律,从而比较可靠地预测化合物活性值.

在定量构效关系研究中,将 QSAR 方法获得的描述符应用于支持向量机中,这些实验还都是尝试性和初步的,对结构特征量的分析和选择有待进一步探索.对一些参数的选取(如核函数参量等),虽然目前有一些求取的方法,仍然存在一定的人为性,但是支持向量机在结构功能预测中已经显示了一定的应用前景.

### 3 结论

1) 利用 CoMFA 计算出构效关系中重要的参数,结合支持向量机的优势,通过 2 种方法相互配

合,对 HIV-1 整合酶抑制剂的结构和活性数据样本建立预测模型,获得训练集  $R^2 = 0.937$ ,测试集  $R^2 = 0.804$ ,有较好的预测能力.

2) 定量构效关系研究的关键步骤是选择合适的计量学方法建立结构参数与活性数据之间的定量关系模型.化合物结构与生物活性之间存在线性关系,也可能存在非线性关系,大多数是非线性关系.因此,在解决小样本、非线性问题上,有优势的 SVM 建模具有广阔的发展前景,本实验进行了有益的尝试.

3) 本研究对支持向量机在分析化合物生物活性与其结构、物理化学性质之间的定量变化规律进行了探索,希望今后在阐明构效机理方面做更多的工作.

### 参考文献:

- [1] POMMIER Y, JOHNSON A A, MARCHAND C. Integrase inhibitors to treat HIV/AIDS [J]. *Nat Rev Drug Discov*, 2005, 4(3): 236-248.
- [2] 姜晓华,龙亚秋. 结构多样的 HIV-1 整合酶抑制剂: 过去、现在和未来 [J]. *有机化学*, 2004(24): 1380-1388.  
JIANG Xiao-hua, LONG Ya-qiu. Structurally diverse HIV-1 integrase inhibitors: past, present and perspective [J]. *Chin J Org Chem*, 2004(24): 1380-1388. (in Chinese)
- [3] DAYAM R, SANCHEZ T, CLEMENT O, et al.  $\beta$ -Diketo acid pharmacophore hypothesis: 1. discovery of a novel class of HIV-1 integrase inhibitors [J]. *J Med Chem*, 2005, 48(1): 111-120.
- [4] HAZUDA D J, PETER F, MARC W, et al. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells [J]. *Science*, 2000, 287(5453): 646-650.
- [5] EGBERTSON M S. HIV integrase inhibitors: from diketoacids to heterocyclic templates: a history of HIV integrase medicinal chemistry at merck west point and merck rome (IRBM) [J]. *Current Topics in Medicinal Chemistry*, 2007, 7(13): 1251-1272.
- [6] STEIGBIGEL R, KUMAR P, ERON J, et al. Results of BENCHMRK-2, a phase III study evaluating the efficacy and safety of MK-0518, a novel HIV-1 integrase inhibitor, in patients with triple-class resistant virus [C]// 14th Conference on Retroviruses and Opportunistic Infections, Los Angeles, February 25-28, 2007.
- [7] VAPNIK V. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995.
- [8] HUA S, SUN Z. A novel method of protein secondary

- structure prediction with high segment overlap measure: support vector machine approach [J]. *J Mol Biol*, 2001, 308(2): 397-407.
- [9] BURBIDGE R, TROTTER M, BUXTON B, et al. Drug design by machine learning: support vector machines for pharmaceutical data analysis [J]. *Computers and Chemistry*, 2001, 26(1): 5-14.
- [10] HONG S J, WEISS S M. Advances in predictive models for data mining [J]. *Pattern Recognition Letters*, 2001, 22: 55-61.
- [11] GAOA J B, GUNNB S R, HARRISB C J. SVM regression through variational methods and its sequential implementation [J]. *Neurocomputing*, 2003, 55(1): 151-167.
- [12] 金宏. 统计模拟肽及蛋白质的性质和活性 [D]. 浙江大学化学系, 2010.  
JIN Hong. Statistical modeling of the property and activity of peptides and proteins [D]. Hangzhou: Department of Chemistry, Zhejiang University, 2010. (in Chinese)
- [13] CORTES C, VAPNIK V. Support vector networks [J]. *Machine Learning*, 1995, 20(3): 273-293.
- [14] 梅虎, 梁桂兆, 周原, 等. 支持向量机用于定量构效关系建模的研究 [J]. *科学通报*, 2005, 50(16): 1703-1708.  
MEI Hu, LIANG Gui-zhao, ZHOU Yuan, et al. Modeling research on quantitative structure and activity relationship using support vector machine [J]. *Chinese Science Bulletin*, 2005, 50(16): 1703-1708. (in Chinese)
- [15] FLACH P A. On the state of the art in machine learning: a personal review [J]. *Artificial Intelligence*, 2001, 131(1): 199-222.
- [16] BELOUSOV A I, VERZAKOV S A, FRESE J V. A flexible classification approach with optimal generalization performance: support vector machines [J]. *Chemom Intell Lab Syst*, 2002, 64(1): 15-25.
- [17] 纪彩虹. 定量构效关系的原理、方法及其研究进展 [J]. *甘肃联合大学学报: 自然科学版*, 2011, 25(2): 58-62.  
JI Cai-hong. The principle, methods, realization process and research status of QSAR/QSPR [J]. *Journal of Gansu Lianhe University: Natural Science Edition*, 2011, 25(2): 58-62. (in Chinese)
- [18] 冯丽萍, 郭正元, 梁菁, 等. 几种 QSAR 建模方法的研究进展与应用 [J]. *农业环境科学学报*, 2007, 26(增刊): 651-655.  
FENG Li-ping, GUO Zheng-yuan, LIANG Jing, et al. Research progress and application in the several QSAR modeling method [J]. *Journal of Agro-Environment Science*, 2007, 26(S): 651-655. (in Chinese)
- [19] CRISTINA G, EMANUELA N, ESTER M, et al. Discovery and synthesis of HIV integrase inhibitors: development of potent and orally bioavailable N-Methyl pyrimidones [J]. *J Med Chem*, 2007(50): 4953-4975.
- [20] ESTER M, OLAF K, CRISTINA G, et al. Design and synthesis of bicyclic pyrimidinones as potent and orally bioavailable HIV-1 integrase inhibitors [J]. *J Med Chem*, 2008(51): 861-874.
- [21] JING S, HAI F C, HAI R X, et al. Comparative study of factor Xa inhibitors using molecular docking/SVM/HQSAR/3D-QSAR methods [J]. *QSAR Comb Sci*, 25, 2006(1): 25-45.
- [22] MARCHAND C, JOHNSON A A, SEMENOVA E, et al. Mechanisms and inhibition of HIV integration [J]. *Drug Discovery Today: Disease Mechanisms*, 2006(3): 253-260.
- [23] 郭涤亮, 刘冠男, 周宇, 等. HIV 整合酶抑制剂的研究进展 [J]. *有机化学*, 2010, 30(4): 477-485.  
GUO Di-liang, LIU Guan-nan, ZHOU Yu, et al. Research progress in HIV integrase inhibitors [J]. *Chinese Journal of Organic Chemistry*, 2010, 30(4): 477-485. (in Chinese)
- [24] SATO M, MOTOMURA T, ARAMAKI H, et al. Novel HIV-1 integrase inhibitors derived from quinolone antibiotics [J]. *J Med Chem*, 2006(49): 15061508.
- [25] PACE P, SPIESER S A H, SUMMA V. 4-Hydroxy-5-pyrrolinone-3-carboxamide HIV-1 Integrase Inhibitors [J]. *Bioorganic & Medicinal Chemistry Letters*, 2008, 18(14): 3865-3869.

(责任编辑 吕小红)