

## 苏云金杆菌杀虫晶体蛋白活性预测的支持向量机模型

# A Prediction Model for the Activity of Insecticidal Crystal Proteins from *Bacillus thuringiensis* Based on Support Vector Machine

林 毅\* 蔡福营 张光亚

LIN Yi\* , CAI Fu-Ying and ZHANG Guang-Ya

华侨大学生物工程与技术系, 工业生物技术福建省高校重点实验室, 泉州 362021

Department of Bioengineering & Biotechnology, Huaqiao University, Key Laboratory of Industrial Biotechnology of Fujian Province Universities, Quanzhou 362021, China

**摘 要** 藉均匀设计(UD)方法, 构建了苏云金杆菌(*Bt*)杀虫晶体蛋白氨基酸组成特征与其杀虫活性之间关系的支持向量机(SVM)模型。当惩罚系数为 0.01、epsilon 值为 0.2、gamma 值为 0.05、域值为 0.5 时, 该模型对 *Bt* 杀虫晶体蛋白杀虫活性的预测平均准确率达 73%。

**关键词** 苏云金杆菌, 均匀设计, 支持向量机, 杀虫晶体蛋白, 杀虫活性预测

中图分类号 Q811.4 文献标识码 A 文章编号 1000-3061(2007)01-0127-06

**Abstract** A quantitative structure-property relationship (QSPR) model in terms of amino acid composition and the activity of *Bacillus thuringiensis* insecticidal crystal proteins was established. Support vector machine (SVM) is a novel general machine-learning tool based on the structural risk minimization principle that exhibits good generalization when fault samples are few; it is especially suitable for classification, forecasting, and estimation in cases where small amounts of samples are involved such as fault diagnosis; however, some parameters of SVM are selected based on the experience of the operator, which has led to decreased efficiency of SVM in practical application. The uniform design (UD) method was applied to optimize the running parameters of SVM. It was found that the average accuracy rate approached 73% when the penalty factor was 0.01, the epsilon 0.2, the gamma 0.05, and the range 0.5. The results indicated that UD might be used an effective method to optimize the parameters of SVM and SVM and could be used as an alternative powerful modeling tool for QSPR studies of the activity of *Bacillus thuringiensis* (*Bt*) insecticidal crystal proteins. Therefore, a novel method for predicting the insecticidal activity of *Bt* insecticidal crystal proteins was proposed by the authors of this study.

**Key words** *Bacillus thuringiensis*, uniform design, support vector machine, insecticidal crystal proteins, activity prediction

苏云金杆菌(*Bacillus thuringiensis*, *Bt*)是一类非常重要的昆虫病原体, 在芽孢形成过程中伴随晶体

Received: July 20, 2006; Accepted: September 7, 2006.

This work was supported by the grants from the National Natural Sciences Foundation of China (No. 40601046), the Program for New Century Excellent Talents in Fujian Province Universities, and the Natural Science Foundation of Fujian Province (No. B0510011).

\* Corresponding author. Tel: +86-595-22692031, Email: lyhxm@hqu.edu.cn

国家自然科学基金(No.40601046)福建省高等学校新世纪优秀人才支持计划资助基金和福建省自然科学基金(No.B0510011)资助。

© 中国科学院微生物研究所期刊联合编辑部 <http://journals.im.ac.cn>

蛋白的产生。不少晶体蛋白被证实具有特异杀虫活性,通常称其为杀虫晶体蛋白(Insecticidal crystal proteins, ICPs)。目前已发现 Bt 杀虫晶体蛋白对许多重要的农作物害虫,包括鳞翅目、鞘翅目、双翅目、膜翅目、螨类和线虫,具有特异性的毒杀作用,而对人畜无害<sup>[1]</sup>。

支持向量机(support vector machine, SVM)<sup>[2]</sup>是近年来国际上新兴的一种机器学习方法,由于出色的学习性能,该技术已成为当前的研究热点。且已被成功地应用于生物信息学中基因微阵列表达模式<sup>[3]</sup>、蛋白质家族<sup>[4]</sup>、转录起始点<sup>[5]</sup>、蛋白质亚细胞定位<sup>[6-8]</sup>、蛋白质折叠<sup>[9]</sup>等方面。

对杀虫晶体蛋白的杀虫活性鉴定一般采用传统的生物学测定法。本研究藉均匀设计方法,构建了 ICPs 氨基酸组成与其特异杀虫活性之间关系的支持向量机模型,并得到了很好的预测结果,可望用于杀

虫晶体蛋白活性的快速预测。

# 1 材料与方法

## 1.1 数据来源

本研究所用的 108 条杀虫晶体蛋白序列及其杀虫活性来自 NCBI 蛋白质数据库(<http://www.ncbi.nlm.nih.gov>)和 CFS(<http://www.glf.cfs.nrcan.gc.ca/science/research/netintro99-e.html>)。其中 44 条 ICPs 杀鳞翅目(Lepidoptera, L)害虫,24 条 ICPs 杀鞘翅目(Coleoptera, C)害虫,20 条 ICPs 杀双翅目(Diptera, D)害虫,20 条 ICPs 毒杀鳞、双、鞘以外的害虫(Other, O)。所用的蛋白序列的 ID 号及其对应的杀虫活性见表 1。大体上把杀虫晶体蛋白根据他们的杀虫活性分为以上这四类,并分别令 L = 1, C = 2, D = 3, O = 4 作为支持向量机的输出数据。

表 1 蛋白序列的 ID 号及其对应的杀虫活性  
Table 1 ID of amino acid sequence and their insecticidal activity

Insecticidal activity	ID							
Lepidoptera	AAA22331	AAA22340	AAA22344	AAA22353	AAA22410	AAB82749	CAA29898	AAC32850
	CAA30396	CAA37933	AAA22346	AAA22348	CAA44633	AAA82114	AAA22341	AAA98959
	AAB00376	AAA22335	AAA22342	CAA40536	AAA22333	CAA85764	CAA41122	CAA00882
	BAA77213	CAA30312	CAA38099	CAA01235	AAB97923	AAB49768	AAS68357	AAS68355
	AAS68356	AAS68354	AAA22614	AAA22613	POA370	P05519	P05068	Q45733
	P19415	P09662	AAG33380					
AAA21119	Coleoptera	CAA34983	AAA22336	AAA22334	CAA42469	AAA22351	AAA21117	AAA21118
	AAF76375	AAB32872	AAY28601	CAA68482	AAW82872	Q06117	Q03749	BAC07226
	AAA22334	POA379	Q45744	P17969	Q45708	Q45707	AAA01896	AAA01454
Diptera	AAA22335	CAA68485	AAA22614	AAA22352	CAA60504	AAC97162	CAA63860	CAA68875
	AAB93476	I32932	CAA26943	CAA67328	CAA78519	AAB93477	CAA30312	AAA22614
	AAA22613	POA370	P05519	P09662				
Other	EAL12386	EAL11303	BAA19712	AAF76290	1BLI	8A3HA	5CEVF	AAA22355
	AAA22356	AAA22357	AAA22358	AAA67693	AAA67694	I34547	EAL12386	EAL11303
	BAA19712	AAF76290	1BLI	8A3HA	5CEVF			

研究表明 Bt 杀虫晶体蛋白的杀虫活性与氨基酸组成有明显的关联性<sup>[10]</sup>。鉴于 Bt 杀虫晶体蛋白的分子量差别很大,所以以各杀虫晶体蛋白的 20 种氨基酸组成和氨基酸残基数为输入数据。这样 ICPs 序列可表示为如下特征向量:

$$x_j^\rho = [x_{j,1}^\rho, x_{j,2}^\rho, \dots, x_{j,20}^\rho, x_{j,21}^\rho]^T \quad (1)$$

式中  $\rho$  表示蛋白质的类别,  $j$  表示每类蛋白质的样本个数,  $x_{j,i}^\rho$  ( $i = 1, 2, \dots, 20$ ) 表示具有  $\rho$  类杀虫活性的蛋白质第  $j$  个蛋白质序列中  $i$  种基本氨基酸出现的频率数,  $x_{j,21}^\rho$  表示具有  $\rho$  类杀虫活性的蛋白质第  $j$  个蛋白质序列的氨基酸残基数。特征向量中元素的顺序按照 20 种氨基酸的字母顺序排列。所有杀虫晶体蛋白的氨基酸组成分析由 Bioedit 软件完成。

## 1.2 支持向量机

支持向量机是 Vapnik 等<sup>[2]</sup>根据统计学习理论提出的一种新的机器学习方法,其最大特点是根据结构风险最小化原则,尽量提高学习的泛化能力,即由有限的训练集样本得到的小误差仍能够保证对独立的测试集小的误差。应用支持向量机进行分类研究的基本思想可简述为:首先将输入空间的样本通过某种非线性函数关系映射到一个特征空间中(维数可能较高),在此特征空间中构造最优分类超平面,使两类样本在此特征空间中可分。映射函数仅与低维输入向量和特征空间的点积有关,此映射函数点积可用一核函数来替代,从而可避免“维数灾难”,因而可解决高维特征问题。支持向量机总是寻

找全局最优解,且可防止过学习。其判别函数为

$$f(x) = \text{sgn} \left[ \sum_{i=1}^N a_i y_i k(x, x_i) + b \right] \quad (2)$$

$k(x_i, x_j)$  称为核函数,核函数的选取应使其为特征空间的一个点积,即存在函数  $\Phi$ , 使  $\Phi(x_i) \cdot \Phi(x_j) = k(x_i, x_j)$ 。业已证明,核函数  $k(x_i, x_j)$  只要满足 Mercer 条件即可满足上述要求<sup>[13]</sup>。常用的核函数有:

多项式核函数

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (3)$$

径向基核函数

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

Sigmoid 核函数

$$k(x_i, x_j) = \tanh [c + x_i \cdot x_j] \quad (5)$$

支持向量机方法最初是针对二类别的分类而提出的,如何将其有效地推广到多类别分类? J. Weston<sup>[11]</sup>提出了一种新的多类 SVM 算法(qp-mc-sv)。它在构造决策函数时,同时考虑所有的类,将原始优化问题推广为:

$$\varphi(w, \xi) = \frac{1}{2} \sum_{m=1}^k (w_m g w) + C \sum_{l=1}^l \sum_{m \neq y_l} \xi_l^m \quad (6)$$

$$S(w_i g x_i) + b y_i \geq (w_m g x_i) + b_m + 2 - \xi_i^m \quad (7)$$

$$\xi_i^m \geq 0, i = 1 \dots l, m \in \{1 \dots k\} y_i \quad (8)$$

相应地决策函数变为:

$$f(x) = \arg \max_k [w_i g x + b_i], i = 1 \dots k \quad (9)$$

当  $k$  取 2 时,和两类分类问题等价。

本研究用 Thorsten Joachims 编写的 SVM 程序,该程序是由 C 语言编写的,对于学术用途者可免费从以下网址下载: <http://www.cs.ucl.ac.uk/staff/M.Sewell/winsvm/winSVM.exe>。文中所用的核函数为径向基核函数。采用均匀设计法优化支持向量机运行参数。

### 1.3 均匀设计法

均匀设计由我国数学家方开泰教授所创造<sup>[12]</sup>,它是将数论和多元统计相结合的一种新颖的试验方法,其核心思想是用确定性方法寻找空间中均匀分布的点集来代替 Monte Carlo 中的随机数。它通过提高试验点“均匀分散”的程度,使试验点具有更好的代表性及能用较少的试验获得较多的信息。为了方便参数的筛选,我们特定义以下三个特征指标<sup>[13]</sup>:

$$Q = \sum_{i=1}^p \rho(i) \gamma N \quad (10)$$

$$Q_i = \rho(i) \gamma \text{obs}(i) \quad (11)$$

$$Mcc\Delta = \frac{1}{\rho} \sqrt{\frac{\sum_{i=1}^p \left( \frac{\rho(i)}{\rho(i) + \alpha(i)} \right) g \sum_{i=1}^p \left( \frac{\rho(i)}{\rho(i) + \alpha(i)} \right)}{\sum_{i=1}^p \left( \frac{\rho(i)}{\rho(i) + \alpha(i)} \right)}} \quad (12)$$

式中  $Q$  为总分类精度,  $Q_i$  为每类样本的分类精度,  $Mcc\Delta$  为相关系数,  $N$  为样本总数,  $\rho$  为样本类别数,  $\text{obs}(i)$  为类别  $i$  的样本数,  $\rho(i)$  为  $i$  类样本的正确分类数,  $\alpha(i)$  为  $i$  类样本中被错分为其他类别的样本数,  $\alpha(i)$  为其他类别的样本被错分为  $i$  类样本的数目。

## 2 结果与分析

### 2.1 基于均匀设计的支持向量机的优化

当采用径向基核函数时,有 3 个运行参数需要优化,分别为惩罚系数  $C$  值、epsilon 和 gamma,进行 3 因素 10 水平的均匀设计优化,所得的均匀设计表如表 2 所示。为了能量化的比较拟合值与真实值之间的差别,我们引进一个域值  $\delta$ ,当拟合值  $X$  (由计算机模型得出的分类值)与真实值  $x$  (人为赋予杀虫晶体蛋白分类值)满足  $|X - x| \leq \delta$  时,定义为拟合值是可取的。我们分别令  $\delta = 0.001, 0.005, 0.01, 0.05, 0.1, 0.5$ 。按照均匀设计得出的实验方案进行拟合实验,得出图 1。

表 2 均匀设计表

Table 2 Uniform design U10(10<sup>3</sup>)

Levels	These factors		
	C	epsilon	gamma
N1	1	0.4	0.001
N2	100	0.1	0.03
N3	1000000	0.5	0.1
N4	10000	0.3	0.3
N5	100000	0.8	0.01
N6	1000	1	0.04
N7	0.001	0.6	0.02
N8	10	0.7	0.5
N9	0.01	0.2	0.05
N10	0.1	0.9	0.2

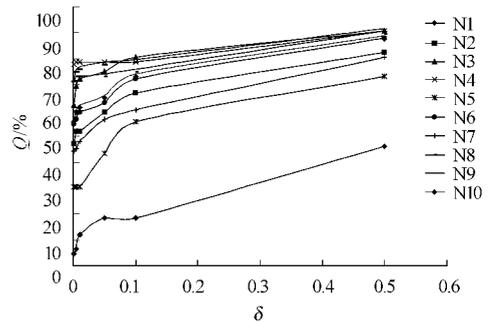


图 1 预测精度随域值变化的关系曲线

Fig.1 The relation curve of expected prediction accuracy with range

$Q$  stand for accuracy and  $\delta$  stand for range.

方案拟合的总精度较高,明显优于其他 6 组参数。为了测试模型的稳定性我们引进了相关系数  $Mcc\Delta$

$$Mcc(i) = \frac{\rho(i)n(i) - u(i)d(i)}{\sqrt{[\rho(i)+u(i)][\rho(i)+d(i)]n(i)+u(i)]n(i)+d(i)}} \quad (13)$$

是不可用的,因为  $n(i)$  (非  $i$  类样本的正确分类数),  $u(i)$  和  $d(i)$  可能为 0 值,使得公式没有数学意义。因为马修斯相关系数主要与精度和灵敏度相关,所以我们用精度和灵敏度来代替马修斯相关系数,从而得出一个新的相关系数  $Mcc\Delta$ 。

$$Pre(i) = \frac{\rho(i)}{\rho(i)+u(i)} \quad (14)$$

$$Rec(i) = \frac{\rho(i)}{\rho(i)+d(i)} \quad (15)$$

$$Mcc\Delta(i) = \frac{1}{\rho} \sqrt{\sum_{i=1}^{\rho} Pre(i)g \sum_{i=1}^{\rho} Rec(i)} \quad (16)$$

由以上三个公式得

$$Mcc\Delta = \frac{1}{\rho} \sqrt{\sum_{i=1}^{\rho} \left( \frac{\rho(i)}{\rho(i)+u(i)} \right) g \sum_{i=1}^{\rho} \left( \frac{\rho(i)}{\rho(i)+d(i)} \right)} \quad (17)$$

上式中  $Pre(i)$  为第  $i$  类的精度,  $Rec(i)$  为第  $i$  类的灵敏度。  $Mcc\Delta$  相关系数为模型的总相关系数,它虽然没有马修斯相关系数局部灵敏,但是它更能从整体水平上判断一个模型的可靠程度。我们对 N3、N4、N8、N9 实验方案进行总相关系数评估,如图 2 所示。

这个参数,在低数据量的情况下,马修斯相关系数 (Matthews correlation coefficient)<sup>[14]</sup>:

的相关系数在各个域值范围内都在 90% 以上,这说明这几种试验方案都是可靠的,但这只能证明模型总体趋于稳定,都是可用的,并不能判断出这四组试验方案的参数那一个最优。为了找出最优参数组合,使总精度随域值变化的更加灵敏,我们分别测出了这 4 组方案的分精度  $Q_i$ ,如图 3 所示。可以看出,方案 N9 惩罚系数为 0.01,  $\epsilon$  值为 0.4,  $\gamma$  值为 0.05 拟合效果最好。所以后续训练及测试均采用上述参数。

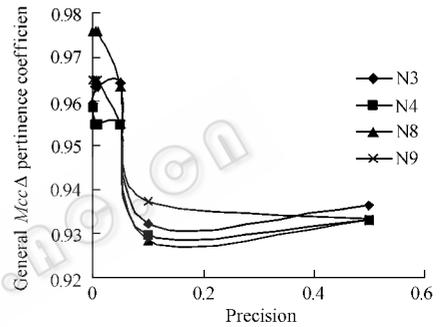


图 2 总修订马修斯系数与预测精度之间的关系曲线  
Fig.2 The relation curve of the general  $Mcc\Delta$  pertinence coefficient with precision

从图 2 可以看出,试验方案 N13、N4、N8 和 N9

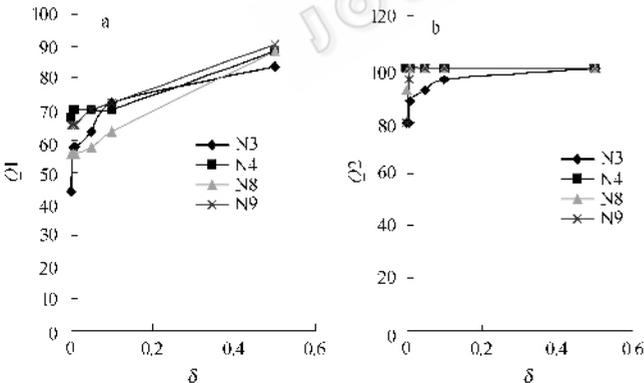


图 3 预测分精度与域值之间的关系曲线

Fig.3 The relation curve of the part expected prediction accuracy with range  
 $Q_1, Q_2, Q_3, Q_4$  all stand for part accuracy and  $\delta$  stand for range.

### 2.2 支持向量机的测试

对支持向量机而言,由于训练样本集的大小有限,训练后对训练集外的输入响应如何,直接决定了支持向量机的性能。对预测结果的评价基于两种检验方法,一种是 Jackknife 检验,另一种为  $k$ -fold cross-validation 检验,这两种检验是较为客观和严格

的方法。在 Jackknife 检验方法中,每一蛋白质依次从数据库中取出作为测试蛋白,而剩余的蛋白质作为训练集。在  $k$ -fold cross-validation ( $k$ -CV) 检验方法中,随机将数据库分为  $k$  个子集合,依次取出一个子集作为测试集,而其余的  $k-1$  个子集作为训练集,此过程循环  $k$  次。

在此研究中我们采用 11 倍交叉验证( 11-fold cross-validation )检测方法,把上述 108 组数据大约随机平均分成 11 个子集合。任取一个子集合作为测试样本,其余 10 组作为训练样本,对支持向量机模型进行测试,共循环了 11 次。在测试中,可以根据实际需要调整域值,在本测试中定义域值为 0.5,即预测结果在域值为 0.5 这个范围内测试结果是可取的。训练和测试结果如表 3 所示。

在表 3 中我们给出了 4 类杀虫晶体蛋白的训练和测试精确度,并计算了总体水平上的精确度。从

表 3 训练与测试的结果

Table 3 The results of training and testing

Circulation	Lepidoptera		Coleoptera		Diptera		Other		Accuracy rate	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	85.7	0	100	100	88.9	50	72.2	100	87	62.5
2	95.3	100	100	100	94.1	0	66.7	100	91.8	70
3	90	75	100	100	81.2	75	84.2	100	86.7	80
4	82.5	50	100	0	83.3	50	88.2	33.3	77.6	40
5	82.5	50	100	100	78.9	0	76.8	66.7	84.7	60
6	90	75	100	100	83.3	66.7	77.8	50	88.8	70
7	83.3	50	100	100	100	100	78.9	100	86.7	60
8	92.5	50	100	100	73.7	100	84.2	100	88.8	80
9	85	100	100	75	94.7	100	94.7	100	91.8	90
10	90	100	100	75	84.2	100	68.4	100	85.7	90
11	90	100	100	100	84.2	100	73.7	100	89.5	100
Average	87.9	68.2	100	86.3	86	67.4	78.7	86.4	87.2	73

### 3 讨论

苏云金杆菌杀虫剂由于其具有较强的致病力、毒性持久、对人畜无害等优点,又能进行大规模的工业化生产,因此在害虫生物防治中得到较为广泛的应用。目前已建立了多种目标昆虫的生物测定模式<sup>[15]</sup>。在生物学测试中要考虑很多因素如害虫本身的生殖、生存潜能等生物学特性,还与温度、湿度和光照等主要环境因子等有关。生物学测试过程繁琐、耗时耗力耗财。更有甚者,对于一些害虫仍未建立起一套行之有效的方法。

随着计算机技术的飞速发展,利用计算机建模来预测苏云金杆菌杀虫晶体蛋白的活性不失为一种行之有效的方法。近年来,很多机器学习方法如遗传算法、人工神经网络、支持向量机等被引进到生物学领域,用这些模型来预测酶的最适 pH 值<sup>[16]</sup>、最适温度<sup>[17]</sup>,预测蛋白质的高级结构、优化发酵反应体系等已经取得了可喜的效果。但是,应用支持向量机来预测苏云金杆菌杀虫晶体蛋白活性的研究至今还没有文献报道。

最后需要指出的是,尽管本文采用了均匀设计

训练和测试结果可以看出,拟合的效果总体好于测试的效果,随着数据源的变化其拟合和测试的效果变化也很大。在这 11 次循环测试中,预测最差的准确率为 40%,最好的为 100%,一般都高于 70%,并且总的平均预测精度达到 73%。在这 11 次循环测试中,四类杀虫晶体蛋白的平均预测精度分别为 68%(鳞翅目)、86.3%(鞘翅目)、67.4%(双翅目)和 86.4%(其它),鳞翅目和双翅目预测精度接近,鞘翅目和“其他”类型的预测精度接近,这可能和序列的同源性有关,序列同源性越高其预测精度越接近。

的方法对支持向量机的结构进行了优化,但在各因素水平的选择上仍带有一定的随意性,如果经过精心的选择,支持向量机的检测效果还会有所改善。同时,如果在支持向量机训练之前对样本集进行优化和选择,如采用主成分分析法,可望进一步提高拟合和测试的效果。而且本文只考虑了 20 种氨基酸的频率分布和氨基酸的残基数,排除了其它影响因素,这是一种最简单的情形。同时样本中噪声的影响也不可忽视。我们在此只注重模型的建立,对于模型的进一步优化和机理性的探讨,将是以后工作的重点。

### REFERENCES (参考文献)

- [ 1 ] Guan X(关雄). Research on *Bacillus thuringiensis* strain 8010. Beijing : Science Press(科学出版社),1997.
- [ 2 ] Vapnik V. Statistical Learning Theory. New York : Wiley Press , 1998.
- [ 3 ] Brown M , Grundy W , Lin D , et al . Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* , 2000 , **97** ( 1 ) : 262 - 267 .
- [ 4 ] Jaakkola T , Diekhans M , Haussler D . Using the Fisher kernel

- International Conference on Intelligent systems for Molecular Biology. Menlo Park, CA : AAAI Press , 1999.
- [ 5 ] Zien A , Ratsch G , Mika S , *et al* . Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics* , 2000 , **16** ( 9 ) : 799 - 807 .
- [ 6 ] Hua SJ , Sun ZR . Support vector machine approach for protein subcellular localization prediction. *Bioinformatics* , 2001 , **17** ( 8 ) : 721 - 728 .
- [ 7 ] Cai YD , Liu XJ , Xu XB , *et al* . Support vector machines for prediction of protein subcellular location. *Mol Cell Biol Res Commun* , 2000 , **4** ( 4 ) : 230 - 233 .
- [ 8 ] Cai YD , Liu XJ , Xu XB , *et al* . Support vector machines for prediction of protein subcellular location by incorporating quasi-sequence-order effect. *J Cell Biochem* , 2002 , **84** ( 2 ) : 343 - 348 .
- [ 9 ] Ding CHQ , Dubchak I . Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* , 2001 , **17** ( 4 ) : 349 - 358 .
- [ 10 ] Beron CM , Curatti L , Salerno GL . New strategy for identification of novel cry-type genes from *Bacillus thuringiensis* strains. *Appl Environ Microbiol* , 2005 , **71** ( 2 ) : 761 - 765 .
- [ 11 ] Weston J , Watkins C . Multiclass support vector machines. *Department of Computer Science Egham , Surrey TW20 OEX , England* , TR CSD TR9804 , 1998 .
- [ 12 ] Fang KT . The uniform design : application of number-theoretic methods in experimental design. *Acta Math Appl Sin* , 1980 , **3** : 363 - 372 .
- [ 13 ] Zhang SW( 张绍武 ) , Pan Q( 潘泉 ) , Chen RS( 陈润生 ) , *et al* . Classification of protein homo-oligomers using support vector machine. *Prog Biochem Biophys*( 生物化学与生物物理进展 ) , 2003 , **30** ( 6 ) : 879 - 883 .
- [ 14 ] Matthews BW . Comparison of predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta* , 1975 , **405** ( 2 ) : 442 - 451 .
- [ 15 ] Gao MY( 高梅影 ) , Li RS( 李荣森 ) , Dai SY( 戴顺英 ) , *et al* . New coleopterancidai strains of *Bacillus thuringiensis* and production of coleopterancide. *Acta Microbiologica Sinica*( 微生物学报 ) , 1999 , **39** ( 6 ) : 515 - 520 .
- [ 16 ] Zhang GY( 张光亚 ) , Fang BS( 方柏山 ) . A model for amino acid composition and optimum pH in GP11 xylanase based on neural networks. *Chinese Journal of Biotechnology* ( 生物工程学报 ) , 2005 , **21** ( 4 ) : 658 - 661 .
- [ 17 ] Zhang GY( 张光亚 ) , Fang BS( 方柏山 ) . A model for amino acid composition and optimum temperature in F/10 xylanase. *Journal of Huaqiao University*( 华侨大学学报 ) 2005 , **26** ( 2 ) : 191 - 194 .