

# 基于偏最小二乘法的抗类风湿性免疫活性肽定量构效建模研究

程媛, 曹慧, 徐斐, 于劲松, 袁敏, 吕翠翠

(上海理工大学医疗器械与食品学院, 上海 200093)

**摘要:** 本研究采用偏最小二乘法 (PLS) 建立了抗类风湿性免疫活性肽的 QSAR 模型。首先从常见氨基酸各原子间静电、立体和疏水等与生物活性直接相关的非键作用方式, 筛选出适合于免疫活性肽定量-构效建模的氨基酸描述符, 再采用此描述符对 47 个肽数据集进行定量构效关系建模。结果表明, Z-scales 氨基酸描述符实现了对氨基酸残基的多位置、多变量的定量结构描述, 最适合用于描述免疫活性肽的物化性质。根据 Z-scales 描述符, 经主成分分析和偏最小二乘法建立的模型有良好的可靠性和预测能力, 模型的复相关系数  $R^2=0.986$ , 均方根误差  $RMSE=0.253$ , 留一交叉验证相关系数  $Q^2_{LOO}=0.893$ , 外部验证相关系数  $Q^2_{ext}=0.971$ 。通过对免疫活性肽构效关系的研究为类风湿性关节炎耐受原的筛选、设计开辟了广阔空间, 同时也为功能食品的开发和创新提供了新手段。

**关键词:** 偏最小二乘法; 免疫; 肽; 定量构效关系

文章编号: 1673-9078(2014)10-176-181

DOI: 10.13982/j.mfst.1673-9078.2014.10.030

## Quantitative Structure-activity Relationship Model of Anti-rheumatoid Immune-active Peptides Based on Partial Least Square Method

CHENG Yuan, CAO Hui, XU Fei, YU Jin-song, YUAN Min, LV Cui-cui

(School of Medical Instruments and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** In this study, partial least squares (PLS) method was used to build a quantitative structure-activity relationship (QSAR) model for anti-rheumatoid immune activity of peptides. First, based on the bioactivity-related non-bonding effects (such as electrostatic, steric, and hydrophobic interactions) of each atom, the amino acids descriptor best suited for QSAR model of immune activity of peptides was selected to construct a QSAR model for a 47-peptide data set. The results showed that Z-scales amino descriptor achieved multiple-position and multivariate quantitative structure description for amino acid residues and was the most suitable for describing physicochemical properties of anti-rheumatoid immune-active peptides. Using the Z-scales descriptor, the model built on principle component analysis and PLS showed good reliability and predictive ability, with multiple correlation coefficient of 0.986, root mean square error (RMSE) of 0.253, leave-one-out cross-validation correlation coefficient of 0.893, and external validation coefficient of 0.971. Based on the structure-activity relationship of immune-active peptides, this study provides new opportunities for screening and design of new tolerogens for rheumatoid arthritis, thus providing a new approach for the development innovative functional food products.

**Key words:** partial least squares; immunity; peptides; quantitative structure-activity relationship

类风湿性关节炎 (rheumatoid arthritis, RA) 是临床常见的多发病, 是一种以关节滑膜炎为特征的慢性全身性自身免疫性疾病, 发病率约占我国人口的

收稿日期: 2014-04-21

基金项目: 国家自然科学基金 (31201421); 国家“863”计划项目 (2013AA102207); 上海市研究生创新基金项目 (JWCXSL1302)

作者简介: 程媛 (1988-), 女, 硕士研究生, 研究方向为功能性配料及添加剂

通讯作者: 曹慧 (1976-), 女, 博士, 副教授, 研究方向为功能性配料及添加剂

0.35~0.4%, 其中致残率可达 15%, 然而目前对 RA 尚没有一个较好的治疗方法。最近在研究人类 RA 的过程中, 一些学者报道了口服 II 型胶原产生免疫耐受可使 RA 的发病过程受到抑制, 其机制为 T 细胞对 II 型胶原分子中特异性肽段的识别作用。而有效的免疫耐受的激发依赖于: 抗原递呈细胞 (Antigen Processing Cells, APCs) 加工处理蛋白抗原成 8~10 个氨基酸组成的短肽, 并与其表面的 HLA-DR4 分子形成 HLA-DR4/肽复合物, 再经分泌途径呈现于 APCs 细胞表面, 供受体特异性 T 细胞表面受体 (TCR) 所识别,

进一步结合成 TCR-HLA/DR4-抗原肽三分子复合体, T 细胞被活化、增值并分泌各种免疫因子而启动免疫应答, 其中与 HLA-DR4 分子结合的短肽即为 T 细胞表位, 它在 T 细胞活化过程中起着关键的作用。

目前对抗原 T 细胞表位鉴定及构效关系的研究方法主要有两种。传统的方法是采用先合成大量的交叠肽, 再通过多肽结合试验或 T 细胞因子抑制效应检测来选取表位肽。这种方法预测 T 细胞表位的效率较低, 工作量大。通过建立抗类风湿免疫活性肽定量构效模型, 大大缩小了目标活性肽的选取范围, 为进一步实验测量活性奠定基础, 提高预测效率, 并为活性肽与 MHC-II 类分子 (即 HLA-DR4 分子) 相互作用的机理提供依据, 具有很高的参考价值。

偏最小二乘回归法 (Partial Least-Squares Regression, PLS) 是采用信息综合与筛选技术的新型多元回归建模方法, 它能克服变量多重相关性在系统建模中的不良作用。在利用多个自变量  $X = (X_1, X_2, \dots, X_p)$  进行回归建模时, PLS 将自变量系统中的信息重新组合, 有效地提取对因变量 Y 解释性最强, 又最能概括自变量系统 X 中信息的综合变量 (即主成分)  $t_0, t_1, \dots, t_i (i \leq p)$ , 剔除多重相关信息和无解释意义信息的干扰, 从而克服了自变量多重相关性在系统建模中的不良作用, 得到一个更为可靠的分析结果<sup>[1]</sup>。因而, 本研究基于肽库和文献中查阅到的免疫活性肽, 采用偏最小二乘算法<sup>[2]</sup>建立了 HLA-DR4\*0401 限制性 T 细胞表位的 QSAR 模型, 并运用该模型对 II 型胶原分子中具有免疫活性的肽段进行了预测。

## 1 材料与方法

### 1.1 PLS 建模方法

本研究采用 PLS 对抗类风湿免疫活性肽进行定量构效建模。具体方程如下:

$$pIC_{50} = a_0 + a_1x_1 + a_2x_2 + \dots + a_ix_i$$

其中抗类风湿免疫活性肽的活性大小为  $pIC_{50}$  值,  $pIC_{50} = -\lg(IC_{50})$ <sup>[3]</sup>。  $pIC_{50}$  值是短肽主链的贡献加上每个位置上氨基酸残基的贡献之和<sup>[1]</sup>,  $a_0$  为主链的贡献;  $a_i$  为抗原肽的每个位置上氨基酸残基的贡献。偏最小二乘回归用  $simca-p$ <sup>[4]</sup>完成。

### 1.2 氨基酸描述符表征

从常见氨基酸分子的三维空间结构出发, 基于氨基酸各原子间静电、立体和疏水等与生物活性直接相关的非键作用方式, 选择 Z-scales、C 标度描述符、

ISA-ECI 描述符和 MS-WHIM 描述符对免疫活性多肽进行表征。

### 1.3 建模数据采集

从 AntiJen 数据库和相关文献<sup>[5]</sup>中收集到 47 个抗类风湿免疫活性肽, 这些肽来自于 HLA-DRB1\*0401 限制性表位, 每个肽均含有 13 个氨基酸并有  $IC_{50}$  (酶的活性降低至 50% 时抑制剂的浓度)。为有效的建立构效模型, 将 IC 值转化为对数形式, 并将其按降序排列, 每隔三个肽选取一个肽做预测集, 余下部分做训练集, 这样训练集中的肽可以代表抗类风湿免疫活性肽广泛的活性和结构。

### 1.4 数据的正交信号修正

在进行 PLS 回归时, 经常遇到这样的情况: 第一个主成分 (或特征变量) 可以解释很高比例过程变量 X 的变化, 但不能很好解释质量变量 Y 的变化。只有包含更多的主成分, 模型性能才会慢慢得到改善。从模型的解释和鲁棒性考虑, 包含过多主成分的模型是不可取的。为提高模型的预测能力, 采用应用化学计量学领域中的正交信号修正法 (Orthogonal signal correction, OSC) 对原始 X 变量进行处理, 将与 Y 变量正交的信息滤除, 再进行 PLS 建模。

### 1.5 模型评估

采用标准回归系数 ( $R^2$ ) 和均方根误差 (RMSE)<sup>[6]</sup> 评价回归方程的拟合能力, 模型验证采用留一交叉验证法 (leave-one-out crossvalidation, LOO-CV), 外部验证系数  $Q^2_{ext}$ <sup>[7]</sup> 用于评价模型的预测能力。各系数的计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^{training} (y_{obs} - y_{est})^2}{\sum_{i=1}^{training} (y_{obs} - y_{trave})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{training} (y_{obs} - y_{est})^2}{n}}$$

$$Q^2_{LOO} = 1 - \frac{\sum_{i=1}^{training} (y_{obs} - y_{pre})^2}{\sum_{i=1}^{training} (y_{obs} - y_{ave})^2}$$

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{test} (y_{obs} - y_{est})^2}{\sum_{i=1}^{test} (y_{obs} - y_{trave})^2}$$

其中:  $y_{obs}$  表示肽的活性测量值;  $y_{est}$  表示肽活性估计值;  
 $y_{train}$  表示训练集的活性平均值  $y_{pre}$  表示留一交叉验证法得到的活  
 性预测值;  $y_{ave}$  表示整个数据集中肽活性的平均值。

## 2 结果与讨论

### 2.1 不同氨基酸描述符对抗类风湿性免疫活

### 性肽的定量构效关系建模及评价

采用四种氨基酸描述符对 47 个抗类风湿性免疫活性肽的性质进行表征, 所获得的矩阵经过正交信号修正后, 再用 PLS 法进行定量-构效建模, 模型选取 1~35 号肽为训练集, 36~47 号肽为预测集, 肽的观测值与预测值如表 1 所示, 不同氨基酸描述符的偏最小二乘模型参数如表 2 所示。

表 1 抗类风湿性免疫活性肽序列及其活性的观测与计算值

Table 1 Observed and calculated values of the sequences of anti-rheumatoid immune active peptides and their activities

序号	肽	测量值	预测值			
			Z-scales	MS-WHIM	C 标度	ISA-ECI
1	AGFKAEQGPKGEP	7.68	7.68	7.67	7.44	7.61
2	AGFKGEQGPKGAP	7.62	7.50	7.54	7.25	7.64
3	AGFKGEQGPKAEP	7.40	7.40	7.42	7.23	7.41
4	AAFKGEGQGPKGEP	7.28	7.25	7.48	7.34	7.332
5	AGFKGEQAPKGEP	7.19	7.22	7.1	6.86	7.35
6	AGFKGEAGPKGEP	6.94	6.94	6.98	6.72	6.8
7	AGFKGEQGPKGEP	6.74	6.63	6.89	6.54	6.77
8	AKFYRPTAFSG	6.70	6.80	6.82	6.82	6.65
9	PGVKGHRGYPGLD	6.43	6.56	6.41	6.7	6.4
10	PGIAGFKGEQGP	6.40	6.33	6.01	6.38	6.42
11	GPFGKARGPPGPT	6.34	6.28	6.16	6.74	6.22
12	SGFQGLPGPPGPP	6.34	6.32	6.17	6.59	6.25
13	AGYKGEQGPKGEP	6.27	6.40	6.15	6.1	6.26
14	QGFQGNPGEPEGEP	6.22	6.31	6.19	6.58	6.33
15	TKLTYAAAVQAAL	6.15	6.06	5.95	6.32	6.11
16	EGMRFAKGYISGY	6.00	5.89	5.71	6.27	5.94
17	PGVKGESGSPGEN	5.72	5.85	5.81	5.58	5.67
18	AKFYRDPTAFSG	5.70	5.77	5.94	5.83	5.68
19	AGFAGEQGPKGEP	5.69	5.69	5.54	5.87	5.69
20	QGARGQPGVMGFP	5.62	5.49	5.4	5.56	5.53
21	KAMLQAMAILTGG	5.54	5.53	5.39	5.44	5.6
22	PGIAGAPGFPGAR	5.50	5.73	5.39	5.63	5.43
23	VEIRAEGNSRFTY	5.30	5.21	5.28	5.05	5.45
24	SSFSQARDSQAQS	5.26	5.20	5.2	5.2	5.3
25	AGVQGAPGPAGEE	5.06	4.81	5.13	4.94	5.01
26	GPVSGPMGPRGPP	5.04	4.82	5.13	5.09	5.18
27	PGVMGFPGPKGAN	5.03	5.14	5.06	5.01	4.98
28	EGMRFDKGYISGY	5.00	4.86	4.88	4.67	5.1
29	FGLQLALTEGMRF	5.00	5.02	5.11	4.88	4.9
30	QYMRADQAAGGLR	4.82	4.92	4.78	4.94	4.7
31	AGFKGEQGPKGEP	4.74	4.86	4.86	4.89	4.74
32	PGLRGLPGKDGET	4.47	4.39	4.42	4.46	4.72

转下页

接上页

33	AGFKGEQGPKEP	4.35	4.51	4.45	4.57	4.5
34	DGIAGPKGPPGER	4.32	4.28	4.35	4.08	4.11
35	AGFKGEQGPKEP	4.10	4.29	4.2	4.36	4.15
36	AGFKGAQGPKEP	5.90	5.79	5.71	5.97	5.88
37	FGLQLELTEGMRF	4.30	4.51	4.72	4.19	4.26
38	AGFKGDQGPKEP	5.65	5.42	5.46	5.61	5.7
39	AGFKGEQGAKEP	6.11	6.01	5.9	6.03	6.09
40	KAMLQDMAILTGG	5.22	5.11	5.07	5.01	5.44
41	AGFKGEQGPAGEP	7.11	7.23	7.07	7.26	6.96
42	QYMRAAQAAGGLR	6.00	6.11	5.99	6.07	5.62
43	AGFKGEQGPKEA	6.37	6.31	6.5	6.37	6.47
44	SSFSQDRDSQAQS	5.16	4.97	5.1	4.96	5.33
45	TKLTYDEAVQAAL	5.00	5.00	5.13	5.2	5.09
46	AGFKGEQGPKEP	6.85	6.73	7	6.64	6.88
47	AGFKGEQGPKEP	6.74	6.63	6.89	6.55	6.77

表 2 不同氨基酸描述符对抗类风湿性免疫活性肽定量构效建模的影响

Table 2 Effect of different amino acid descriptors on building QSAR models for anti-rheumatoid immune active peptides

No.	Descriptors	Model	A <sup>a</sup>	R <sup>2</sup> cum <sup>b</sup>	Q <sup>2</sup> Loo <sup>c</sup>	RMSE <sup>d</sup>	Q <sup>2</sup> ext <sup>e</sup>
1	z-scale	PLS	2	0.986	0.893	0.253	0.971
2	MS-WHIM	PLS	3	0.987	0.881	0.277	0.949
3	ISA-ECI	PLS	4	0.99	0.913	0.425	0.965
4	C 标度	PLS	3	0.954	0.783	0.328	0.967

注：a, A表示主成分数；b, R<sup>2</sup>cum表示相关系数；c, Q<sup>2</sup>Loo表示交叉验证相关系数；d, RMSE表示均方根误差；e, Q<sup>2</sup>ext表示外部验证系数。

表 3 20 种天然氨基酸的 Z-scales 描述符

Table 3 Z-scales structure descriptor of 20 kinds of natural amino acids

氨基酸	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>
Ala A	0.07	-1.73	0.09
Val V	-2.69	-2.53	-1.29
Leu L	-4.19	-1.03	-0.98
Ile I	-4.44	-1.68	-1.03
Pro P	-1.22	0.88	2.23
Phe F	-4.92	1.3	0.45
Trp W	-4.75	3.65	0.85
Met M	-2.49	-0.27	-0.41
Lys K	2.84	1.41	-3.14
Arg R	2.88	2.52	-3.44
His H	2.41	1.74	1.11
Gly G	2.23	-5.36	0.3
Ser S	1.96	-1.63	0.57
Thr T	0.92	-2.09	-1.4
Cys C	0.71	-0.97	4.13
Tyr Y	-1.39	2.32	0.01

Asn N	3.22	1.45	0.84
Gln Q	2.18	0.53	-1.14
Asp D	3.64	1.13	2.36
Glu E	3.08	0.39	-0.07

由表 2 可见, Z-scales 描述符最适合用于对抗类风湿性关节炎免疫活性肽的结构进行描述。Z-scales<sup>[8]</sup>主要由与氨基酸的亲水性、立体形状和电性有关的 3 个参数 Z<sub>1</sub>、Z<sub>2</sub>、Z<sub>3</sub>组成。其中, Z<sub>1</sub> 主要与氨基酸的亲水性有关, Z<sub>2</sub> 与氨基酸的立体形状有关, Z<sub>3</sub> 与氨基酸的电性参数有关(见表 3)。通过这三个参数对氨基酸残基进行表征,可实现氨基酸残基多位置、多变量的定量结构描述。本研究结果与文献资料报道<sup>[9-10]</sup>,在免疫活性肽与 HLA-DR4 分子结合的过程中,分子的疏水性、电性和立体结构是影响其结合力的主要因素相一致。

在 MS-WHIM 氨基酸结构描述符中,第 1 个指数与静电势有关,体现氨基酸残基电荷的正负性和侧链芳香族/脂肪族的结构信息;第 2 个指数受负静电势的影响;第 3 个指数多为负值,主要与它们的电性和线性侧链有关。采用 MS-WHIM 结构描述符建立的多肽

QSAR 模型的预测能力较强,且可用来描述非天然氨基酸,解决含非天然氨基酸多肽的定量描述问题,但采用 MS-WHIM 结构描述符建立的 QSAR 模型,物理意义不明确<sup>[11]</sup>。ISA-ECI 描述符用 ISA 近似地表示侧链基团的疏水性,主要与氨基酸分子的体积有关;用 ECI 表示氨基酸侧链基团的极性,即氨基酸侧链每个原子电荷数  $q$  的绝对值之和,ISA-ECI 描述符的特点是能显示侧链和受体之间的偶极作用。但 ISA-ECI 描述符来源简单,描述符个数少,对于复杂的模型解释能力较弱<sup>[12]</sup>。C 标度描述符是经过对 167 个氨基酸结构参数进行主成分分析后,从中提取出的与氨基酸残基一维结构性质密切相关的 3 个结构描述参数<sup>[13]</sup>。但 C 标度描述符建立的抗类风湿性免疫活性肽 QSAR 模型的内部拟合和外部预测的能力均较差。

## 2.2 采用 Z-scales 描述符建立的抗类风湿性免疫活性肽模型的评价

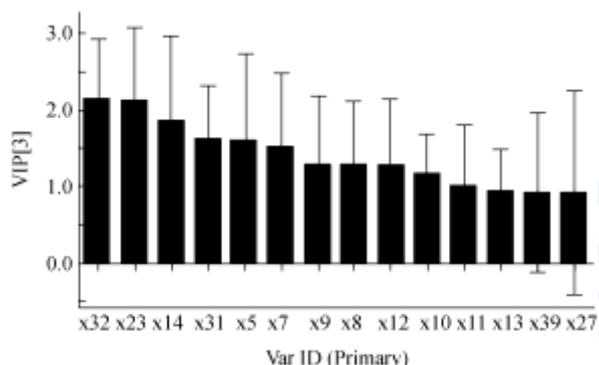


图 1 PLS 模型投影重要性指标图

Fig.1 Plot of variable importance for projection in PLS model

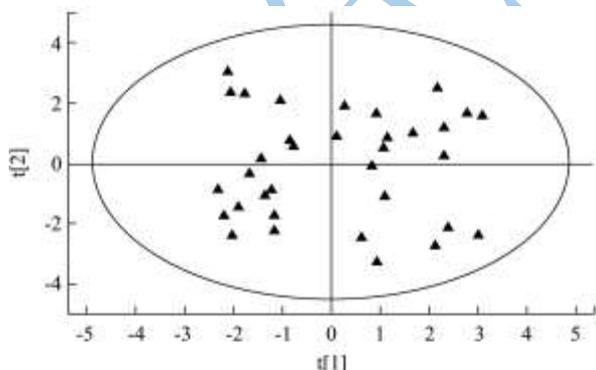


图 2 PLS 得分散点图

Fig.2 Plot of PLS scores

图 1 为抗类风湿性免疫活性肽构效模型的投影重要性指标图,从图中可以看出第十一位氨基酸的立体特性(x32)和亲水性(x31),第八位氨基酸的立体特征(x23),第五位氨基酸的立体特性(x14),第三位氨基酸的亲水性(x7),第二位氨基酸的立体特性(x5)

与肽活性相关性较大。

图 2 是抗类风湿性免疫活性肽构效模型前 2 个主成分的得分分布散点图,训练集中所有样本均在 Hotelling  $T^2$  椭圆图的 95% 置信区间内,说明样本集中不存在异常样本<sup>[14]</sup>。

观察抗类风湿性免疫活性肽模型空间中心规格化距离图(图 3)可以发现除 X20、23、24、26、30 样本重构质量较差外,其他样本到模型 X 空间中心规格化距离都小于 5% 置信度下的临界值 1.345,说明该模型可以解释大部分样本。

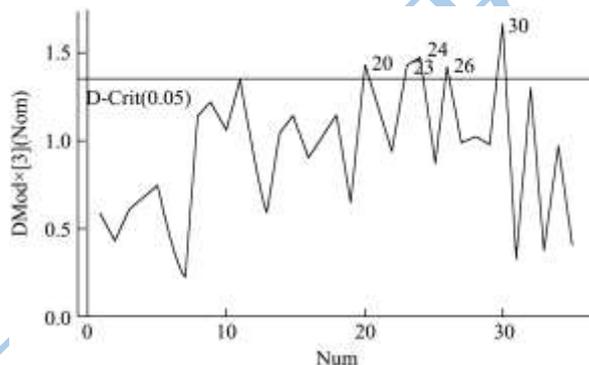


图 3 PLS X 空间中心规格化模型距离

Fig.3 Distance to PLS model in X space

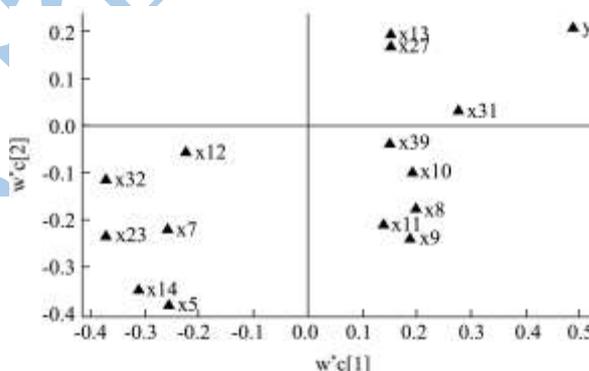


图 4 PLS 载荷图

Fig.4 PLS loading plot

图 4 是抗类风湿性免疫活性肽构效模型的载荷贡献散点图,该图表示各描述子变量与活性之间的相关性。在第一主成分中,第三位氨基酸立体特征 x8 及电性特征 x9,第四位氨基酸亲水性 x10 及立体特征 x11,第五位氨基酸 x13,第九位氨基酸电性特征 x27 和第十一位氨基酸亲水性 x31,第十三位氨基酸电性特征 x39 与肽活性呈正相关,其中 x31 与肽活性的正相关性最强;第二位氨基酸立体特征 x5,第三位氨基酸亲水性 x7,第四位氨基酸电性 x12,第五位氨基酸立体特征 x14,第八位氨基酸立体特征 x23 和第十一位氨基酸立体特征 x32 与活性呈负相关。第二主成分中, x13、x27 和 x31 与肽活性呈正相关,其中 x13 的正相关性最强, x5、x7、x8、x9、x10、x11、x12、x14、

x23、x32 和 x39 与肽活性呈负相关。在与肽活性的正相关变量中，x31 的相关性最强。

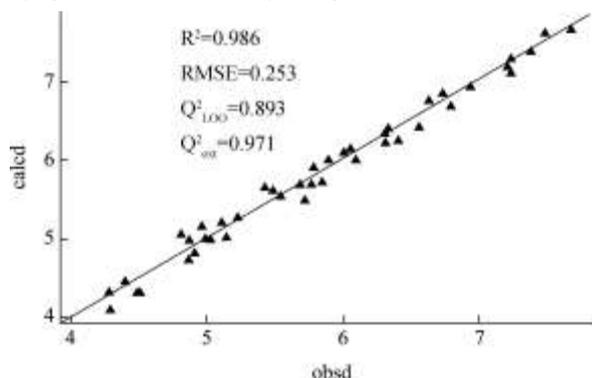


图5 实验值和预测值关系图

Fig.5 Plot of the calculated activities versus the observed activities

图5是抗类风湿性免疫活性肽构效模型的实验值和预测值关系图，样本基本均匀分布在过原点45°直线周围，说明实验室与模型预测值有较好的相关性。依交叉有效指标大小该模型提取了2个主成分，模型的复相关系数  $R^2=0.986$ ，均方根误差  $RMSE=0.253$ ，留一交叉验证相关系数  $Q^2_{LoO}=0.893$ ，外部验证相关系数  $Q^2_{ext}=0.971$ 。Tropsha 曾提出统计参数  $Q^2_{LoO}>0.500$ ， $R^2>0.600$  的模型有效；同时 Tropsha 认为模型的预测能力比拟合能力更为重要，即  $Q^2_{ext}$  的大小更为关键。因此实验结果表明文中建立的抗类风湿性免疫活性肽 QSAR 模型对样本数据的拟合和预测能力都较好，该模型是可靠的。

综合以上分析，采用 Z-scales 描述符建立 HLA-DR4\*0401 限制性表位的 QSAR 预测模型如下：

$$y = -0.03x_1 - 0.14x_2 + 0.04x_3 + 0.09x_4 - 0.21x_5 - 0.2x_7 + 0.04x_8 + 0.02x_9 - 0.08x_{10} - 0.12x_{12} + 0.13x_{13} - 0.24x_{14} - 0.05x_{15} + 0.11x_{16} - 0.01x_{17} - 0.04x_{18} + 0.09x_{19} + 0.02x_{20} - 0.04x_{21} + 0.06x_{22} - 0.24x_{23} - 0.01x_{24} - 0.03x_{25} + 0.05x_{26} + 0.11x_{27} + 0.08x_{28} + 0.06x_{29} - 0.03x_{30} - 0.16x_{31} - 0.24x_{32} - 0.02x_{33} + 0.08x_{34} + 0.02x_{36} + 0.03x_{37} - 0.03x_{38} + 0.04x_{39}$$

### 3 结论

营养与免疫学是近年来兴起的一门学科，主要研究摄入的大分子与机体免疫性疾病之间的相互关系，通过对两者构效关系的研究可以了解食物中组分预防或诱导相关疾病的机制，从而合理指导补充外源物，增强机体的免疫力。本研究采用筛选出的 Z-scales 描述符对抗类风湿性免疫活性肽进行表征，并通过偏最小二乘法对其进行定量构效建模，结果显示该模型的内部拟合能力和外部验证能力均良好。

### 参考文献

- [1] 贾玉泉,季常煦,贾利民,等.基于 PLS 的铁路行包运量预测模型研究[J].铁道运输与经济,2006,10(28):80-82  
JIA Yu-quan, JI Chang-xi, JIA Li-min, et al. Research on railway parcel traffic volume forecast model based on PLS [J]. Railway Transport and Economy, 2006, 10(28): 80-82
- [2] Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics [J]. Chemom. Intell. Lab. Syst. 2001, 58: 109-130
- [3] Mukesh C Sharma, Smita Sharma, Kamendra Singh Bhadoriya. QSAR analyses and pharmacophore studies of tetrazole and sulfonamide analogs of imidazo [4,5-b] pyridine using simulated annealing based feature selection [J]. Journal of Saudi Chemical Society, 2012
- [4] O Galtier, O Abbas, Y Le Dréau, et al. Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude [J]. Vibrational Spectroscopy, 2011, 90(3-4): 469-503
- [5] Wen Zhang, Juan Liu, Yan Qing Niu, et al. Improving the prediction of HLA class I-binding peptides using a supertype-based method [J] Journal of Immunological Methods, 2014, 405:109-120
- [6] Wei Long, Jian Xiang, Hongying Wu, et al. QSAR modeling of iNOS inhibitors based on a novel regression method [J]. Chemometrics and Intelligent Laboratory Systems, 2013, 128: 83-88
- [7] Alla P Toropova, Andrey A Toropov, Jovana B Veselinovi, et al. QSAR models for HEPT derivatives as NNRTI inhibitors based on Monte Carlo method [J]. European Journal of Medicinal Chemistry, 2014, 77: 298-305
- [8] Hellberg S, Eriksson L, Jonsson J, et al. Minimum analogue peptide sets (MAPS) for quantitative structure-activity relationships [J]. Int. J. Pept. Protein Res., 1991, 37: 414-424
- [9] Ellen Christina Andersson, Biarke Endel Hansen. Definition of MHC and T cell receptor contacts in the HLA-DR4-restricted immunodominant epitope in type II collagen and characterization of collagen-induced arthritis in HLA-DR4 and human CD4 transgenic mice [J]. Proc. Natl. Acad. Sci. USA, 1998, 95: 7574-7579
- [10] Lars Fugger. Human autoimmunity genes in mice [J]. Current Opinion in Immunology 2000, 12: 698-703
- [11] 王伟,沈生荣,何国庆,等.生物活性多肽定量构效关系研究方法的比较分析[J].中国食品学报,2009,9(2):154-161  
WANG Wei, SHEN Sheng-rong, HE Guo-qing, et al.

- Comparison on the QSAR research methods of bioactivity polypeptides [J]. Journal of Chinese Institute of Food Science and Technology, 2009, 9(2): 154-161
- [12] Wehrens R, Putter H, Buydens L M C. The bootstrap: a tutorial. Chemometr [J]. Intell Lab Syst, 2000(54): 35-52
- [13] 丁俊杰,丁晓琴,赵立峰,等.新型三维氨基酸结构描述符的研究及其在多肽 QSAR 中的应用[J].药学学报, 2005, 40(4):340-346  
DING Jun-jie, DING Xiao-qin, ZHAO Li-feng, et al. New 3D amino acid structure descriptors and its application to the polypeptide QSAR [J]. Acta Phammaceutica Sinica, 2005, 40(4): 340-346
- [14] 张文彤.SPSS 统计分析高级教程[M].北京:高等教育出版社,2004  
ZHANG Wen-tong. SPSS statistical analysis advanced tutorial [M]. Bei Jing: Higher Education Press, 2004
- [15] Gao Li-zhong, Zhou Su-mei. Optimization of extraction technology of schizophyllan produced by solid fermentation [J]. Science and Technology of Food Industry, 2008, 29(2): 214-216