

• 研究论文 •

## 基于定量序效模型的计算机辅助虚拟疫苗库设计

周 鹏<sup>a,b</sup> 李志良<sup>\*,a,b</sup> 田菲菲<sup>a</sup> 张梦军<sup>c</sup>

(<sup>a</sup>重庆大学化学化工学院 重庆 400044)

(<sup>b</sup>化学生物传感与计量学国家重点实验室 长沙 410082)

(<sup>c</sup>第三军医大学医学检验系 重庆 400040)

**摘要** 给出了基于定量序效模型(QSAM)的计算机辅助虚拟疫苗库设计方案,并以此为基础成功组建了一个合理规模的 HLA-A\*0201 限制性 CTL 表位库,其实现流程如下: (1)从天然氨基酸 516 种理化性质经主成分分析(PCA)得到一种新的氨基酸描述子:氨基酸综合性质得分(SP-score); (2)基于 SP-score 结合遗传-偏最小二乘(GA-PLS)技术建立 QSAM 模型; (3)利用 QSAM 模型作为评价工具采用遗传算法(GA)优化 CTL 表位种群; (4)统计优秀种群中 20 种氨基酸分别在抗原肽序列不同位置出现频率  $f$ ; (5)保留  $f > F$  ( $F$  为平均出现概率,对于任意氨基酸为 1/20)的氨基酸作为该位置的有利残基类型参与虚拟组合库的构建。

**关键词** 定量序效模型; 虚拟疫苗库; HLA-A\*0201 分子; CTL 表位; 遗传算法

## QSAM-based Computer-aided Virtual Vaccine Library Design

ZHOU, Peng<sup>a,b</sup> LI, Zhi-Liang<sup>\*,a,b</sup> TIAN, Fei-Fei<sup>a</sup> ZHANG, Meng-Jun<sup>c</sup>

(<sup>a</sup> College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400044)

(<sup>b</sup> State Key Laboratory of Chemo/Biosensing and Chemometrics, Changsha 410082)

(<sup>c</sup> Department of Medical Analysis, Third Military Medical University, Chongqing 400040)

**Abstract** Upon QSAM-based virtual vaccine library project, a reasonable human leukocyte antigen (HLA)-A\*0201-restricted cytotoxic T lymphocyte (CTL) epitope library was designed. The process is as follows: (a) synthetical property score (SP-score) of amino acids is derived from 516 physicochemical properties by principal component analysis (PCA); (b) based on the SP-score, genetic algorithm-partial least square regression (GA-PLS) is used to construct the quantitative sequence-activity model (QSAM); (c) employing QSAM as evaluation tool, CTL epitopes are optimized by GA; (d) frequency ( $f$ ) of each amino acid in different position of excellent antigen peptides is calculated; (e) the amino acid in condition of  $f > F$  ( $F$  is the random probability, of 1/20 for the natural amino acid) is reserved to be favorable residue type at this position and a structural element of the combinatorial library.

**Keywords** quantitative sequence-activity model; virtual vaccine library; HLA-A\*0201; CTL epitope; genetic algorithm

组合化学(combichem)是将化学结构单元按某种组合方式连接起来,合成包含大量有机分子的化学库,并从中筛选具有某种药理活性物质的一套策略. 通常组合

化学能够在短时期内产生众多有机分子,因此该法现已成为新药研制与开发环节的重要工具. 另一方面,伴随着分子免疫学发展,人们逐渐认识到入侵到体内的外源

\* E-mail: ggootc@163.com; zlli2662@163.com

Received March 10, 2006; revised May 19, 2006; accepted July 12, 2006.

化学生物传感与计量学国家重点实验室基金(No. 05-12-1)、重庆大学创新基金(No. 04-10-10)和第三军医大学青年教师自主创新基金(No. 05-5-28)资助项目.

性物质是以表位(epitope)形式为免疫系统所识别,而在该过程中表达于细胞表面的主要组织相容性复合体(MHC)是否与特定表位发生结合是引发后续免疫级联反应的关键,故寻找能够与 MHC 分子发生有效识别的短肽片段是开发新型疫苗的重要途径.然而由于 MHC 的遗传复杂性和个体多样性,使得该过程变得异常困难<sup>[1]</sup>,为此人们尝试通过组合化学手段合成包含大量肽段的组合肽库,并进一步通过免疫检测筛选具有潜在亲和活性的抗原决定簇序列.然而,由于组合化学的模块交叉连接形式导致被合成的目标分子随反应次数呈几何增长,尽管高通量筛选(HTS)技术的出现为快速测定大量样品提供了有效解决途径,但对于一个包含几千万甚至上亿个分子的组合库而言,对其合成及检测仍是一个耗时耗力的过程.鉴于此,近年来发展起来的计算机辅助虚拟库设计(CAVD)方法<sup>[2]</sup>成为组合化学研究的一个热点,其追求通过理论计算使得在最大限度减少合成目标基础上提高组合库的命中概率.

定量构效关系(QSAR)一直是国内外一个相当活跃的研究领域,通常使用 QSAR 方法来研究生物序列与活性问题亦被称为定量序效模型(QSAM)<sup>[3]</sup>.尽管对生物大分子的 QSAM 研究仍有很大困难,但其间亦不乏成功例子,如 Kidera<sup>[4]</sup>的因子得分法, Hellberg<sup>[5]</sup>定义的  $z$  标度, Collantes<sup>[6]</sup>发展的全向表面积(ISA)/侧链电荷指数(ECI)以及本文作者<sup>[7]</sup>提出的肽/蛋白质结合模式虚拟筛选(GVSC)等方法,已在 QSAM 研究中取得了一定成效.在前人研究基础上,本文收集了 20 种天然氨基酸的 516 种理化性质,通过主成分分析(PCA)技术得到了一种新的氨基酸描述子:氨基酸综合性质得分(SP-score).进而使用 SP-score 成功地建立了 266 个细胞毒性 T 淋巴细胞(CTL)表位对 HLA-A\*0201 分子亲和性的 QSAM 模型,并以该模型作为评价基础,利用遗传算法(GA)设计了一个在统计学意义上具有较大命中概率的虚拟组合疫苗库,该组合库有效地将 CTL 表位合成空间从  $5.12 \times 10^{11}$  降低至  $2.82 \times 10^6$ .

## 1 原理与方法

### 1.1 氨基酸综合性质得分

为了综合考虑处于溶液状态氨基酸残基的各类性质,我们收集了 516 种氨基酸理化参数.它们主要反映氨基酸残基以下几方面信息:(1)电性特征:如净电荷、分子极性; (2)立体特征:如标准 van der Waals 体积、分子尺度、侧链体积; (3)疏水特征:如溶剂化自由能、分配系数; (4)其他:如  $\alpha$  碳原子化学位移、组成参数、频率等.由于不同性质之间可能存在较大信息重叠,且

直接使用 516 个参数来表征序列中单个氨基酸残基将会造成实际应用过于复杂,因此这里结合经典多维数据处理技术——主成分分析(PCA),进行变量压缩提取.首先按列对原始变量矩阵  $X_{20 \times 516}$  进行标准化处理,在此基础上运用 PCA 计算得到得分矩阵.经观察发现,该矩阵前 19 个主成分已经包含原始变量所有信息,故仅取这 19 列主成分得分作为替代矩阵,它们分别解释原始变量矩阵 33.47%, 14.76%, 11.65%, 7.33%, 5.65%, 4.77%, 3.66%, 2.68%, 2.30%, 2.29%, 1.97%, 1.79%, 1.57%, 1.36%, 1.18%, 1.15%, 1.05%, 0.75% 和 0.64% 的方差,累积解释方差为 100.00%.这里称 20 种氨基酸的 19 个显著主成分得分为氨基酸综合性质得分(synthetical property score of amino acids, SP-score).主成分分析使用统计学软件 SPSS 13.0 实现.

### 1.2 HLA-A\*0201 限制性 CTL 表位实验数据

266 个人类 HLA-A\*0201 限制性 CTL 表位皆为 9 肽,并具有自由的 N 和 C 末端.该组数据来自 JenPep 数据库<sup>[8]</sup>,其中包括 110 个高亲和性的抗原肽( $IC_{50} \leq 50$  nmol/L,  $pIC_{50} \geq 7.301$ ), 118 个中等亲和性的抗原肽( $50$  nmol/L  $< IC_{50} \leq 500$  nmol/L,  $7.301 > pIC_{50} \geq 6.301$ )和 38 个低亲和性的抗原肽( $IC_{50} > 500$  nmol/L,  $pIC_{50} < 6.301$ ).  $IC_{50}$  值为采用不同剂量的待测肽与 0.5 nmol/L 的放射性标记的 HBVc18227 (FLPSDYEPSV) CTL 表位肽(对照肽)和 HLA-A\*0201 的复合物在室温下共孵育 2 h,测定待测肽序列将对照肽/HLA-A\*0201 复合物中 50% 的对照肽置换下来的浓度.

### 1.3 定量序效模型

Golbraikh 等<sup>[9]</sup>最近研究表明,交叉检验  $Q^2$  值大小与模型预测能力并没有明显直接关系,对模型预测能力的评价只能通过外部样本集即测试集来进行.故这里按一定规则将 266 个抗原肽样本集随机划分为一个包含 200 个样本的训练集和一个包含 66 个样本的测试集.对于一组肽类似物,每个位置上的氨基酸残基的理化性质可由 19 个 SP-score 所表征.当使用 SP-score 对本文研究的抗原九肽序列进行表征时,共产生  $9 \times 19 = 171$  个 SP-score 描述子,我们用  $V_1 \sim V_{171}$  来分别表示,其中  $V_1 \sim V_{19}$  依次为位置 1 上的 19 个 SP-score 描述子,  $V_{20} \sim V_{38}$  依次为位置 2 上的 19 个 SP-score 描述子,以此类推.由于抗原肽的不同残基以及用于描述同一残基的不同变量对结合能力的贡献是不相同的,因此在建立 QSAM 模型之前需进行变量筛选.本文采用遗传-偏最小二乘(GA-PLS)技术实现变量选择及 QSAM 建模工作,所用软件为 PLS\_Toolbox Version 3.0 基于 Matlab 7.0 环境实现,相关参数:初始群体大小: 100; 最大遗传代数: 200;

变异概率: 0.5%; 交叉互换: 2 点; 交互检验: 留一法; 数据预处理: 标准化; 收敛标准: 80% 个体达到一致; 其他均为默认设置. 筛选结果: 主成分数 5; 适应度 (RMSECV) 0.392.

#### 1.4 虚拟疫苗库设计

利用 QSAM 模型作为指导工具通过 GA 来构造高性价比的定向虚拟疫苗库, 称为 QSAM-based Virtual Vaccine Library. 其具体实现过程如下:

(1) 随机生成规模为  $N$  的初始遗传种群, 其中每一个个体的染色体串代表一个表位(抗原肽段), 而个体的等位基因则为 20 种天然氨基酸代码(即算法采用氨基酸字符进行编码).

(2) 执行选择、交叉和变异算子, 进行遗传迭代操作:

(a) 选择算子(select operator): 通过解码器将个体的氨基酸序列转化成对应的 SP-score 描述子, 利用 QSAM 模型计算出每一个个体的适应度(pIC<sub>50</sub> 值), 以此为基础采用转轮盘法按个体适应度值选择优良种群. 选择概率  $p_i$  通过下式计算:

$$p_i = (\text{pIC}_{50i} - \text{pIC}_{50\min}) / \sum_{j=1}^N (\text{pIC}_{50j} - \text{pIC}_{50\min})$$

其中 pIC<sub>50min</sub> 为种群的最低适应度值;

(b) 交叉算子(crossover operator): 研究证实, 交叉点个数直接影响着遗传算法的收敛性和寻优能力. 交叉点过多将可能破坏连续的优秀基因块, 而过少则难以保证种群遗传多样性和个体间的信息交流能力. 本文采用 2 点交叉进行计算, 即随机选择父母个体上两个切断点作为氨基酸序列的互换位置;

(c) 变异算子(mutation operator): 由于本文遗传算法个体采用氨基酸字符编码, 故最简单的变异操作就是按一定概率将某一位置的残基以其他氨基酸类型替代. 这里设置变异概率为 2%, 即一次循环迭代中 100 个残基位置平均发生 2 次变异.

(3) 收敛条件: 种群的平均适应度值连续 10 代未有显著改进或达到最大循环次数则视为收敛.

(4) 分别计算由遗传算法优化得到的优秀种群中每一个肽段位置 20 种氨基酸出现的频率, 对于出现频率大于随机平均概率(1/20)的氨基酸则被视为该位置的有利残基类型, 反之则为不利残基类型. 保留每个位置的有利残基作为虚拟疫苗库的结构单元.

## 2 结果及分析

### 2.1 定量序效模型分析

采用化学计量学软件包 Simca-p 10.0 对 GA-PLS 筛

选得到的最优变量子集进行深入的数据挖掘. 所得 PLS 模型采用 5 个显著主成分累积解释了  $Y$  变量 76.6% 的方差, 交叉检验解释  $Y$  变量的方差为 65.8%, 可见模型具有较高的稳定性. 图 1 为训练集 200 个样本在前两个主成分上的得分分布散点图, 可以看到抗原肽亲和和活性大小从该图左下角到右上角呈递增分布, 表现出较强的规律性. 另外, 绝大多数抗原肽都落在该图 95% 置信度的 Hotelling  $T^2$  椭圆置信圈内, 仅有 153 和 222 号样本点超出该范围. 经分析发现, 153 号抗原肽的第二位疏水性锚定残基被生理环境下亲水性的谷酰胺(glutamine)所取代, 由此推测这种变化将引起该序列的活性下降(pIC<sub>50</sub> 观测值为 5.501); 而对于 222 号抗原肽而言其活性远低于其他序列(pIC<sub>50</sub> 观测值仅为 4.301), 推测造成这种情况可能是由于实验值偏低或者样本性质特殊的缘故. 基于上述分析可以认为, 153 和 222 号抗原肽表现异常是由于自身结构和活性的特殊性所致, 同时这两个样本在图 1 上的分布趋势也表明, 它们对 HLA-A\*0201 分子具有较低的亲和性. 图 2 是 266 个抗原肽亲和和活性的计算值与实验观测值相关情况, 从该图可以直观地看出本文模型具有良好的拟合能力和稳定性能, 其对测试集 66 个样本预测结果的复相关系数  $R^2_{\text{pred}}$  以及均方根误差 RMSEP 分别为 0.713 和 0.447, 且无明显异常点.

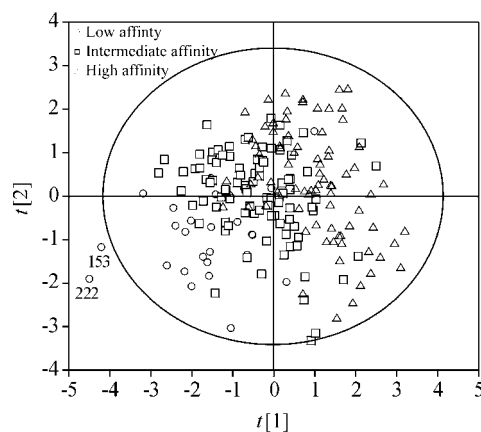


图 1 训练集 200 个抗原肽在前两个主成分( $t[1]/t[2]$ )的得分散点图

Figure 1 Scoring distribution scatter of the first and second principal components ( $t[1]/t[2]$ ) in PLS model for 200 samples in training set

### 2.2 虚拟组合疫苗库实现

设置遗传算法种群大小  $N$  和最大迭代次数  $C$  分别为 500 和 200, 通过自编应用程序进行 QSAM-based Virtual Vaccine Library 设计, 最终统计遗传优化得到的种群中, 20 种天然氨基酸在抗原肽 9 个残基位置分别出现的频率分布情况参见图 3, 从该图可以看到不同氨基

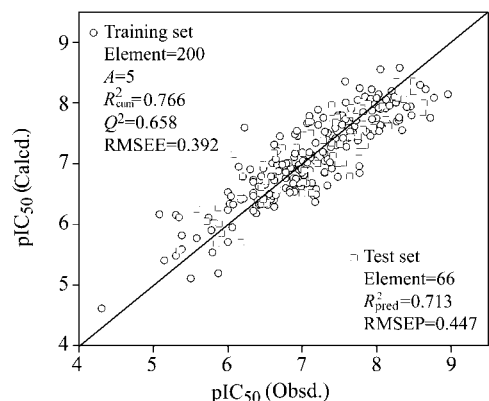


图2 266个抗原肽的亲合活性计算值与观测相关情况  
Figure 2 Plot of the calculated versus the observed binding affinities of 266 antigen peptides

酸类型在不同位置分布趋势具有较为明显的不均匀性。Parker 等<sup>[10]</sup>提出的 IBS (independent binding of side chain)假说指出, 抗原决定簇与蛋白质分子复合体总的结合能量可视为不同位置残基对自由能独立贡献的简单加和, 鉴于此我们认为可以将表位序列上不同位置残基分别考虑, 故定义遗传算法优化得到的种群中在抗原肽某一位置出现频率大于平均概率(1/20)的氨基酸为该位置对亲和活性有利的残基类型并加以保留(图3中处于横线以上的氨基酸), 由此确定最终设计得到的 HLA-A\*0201 限制性 CTL 表位虚拟组合库所包含的结构单元, 参见表1。可以看到这里参与组合库构建的每个残基位置的氨基酸数目已经由全部20种减少到3~8种, 从组合数学的角度来看, 这种改变将引起组合库整

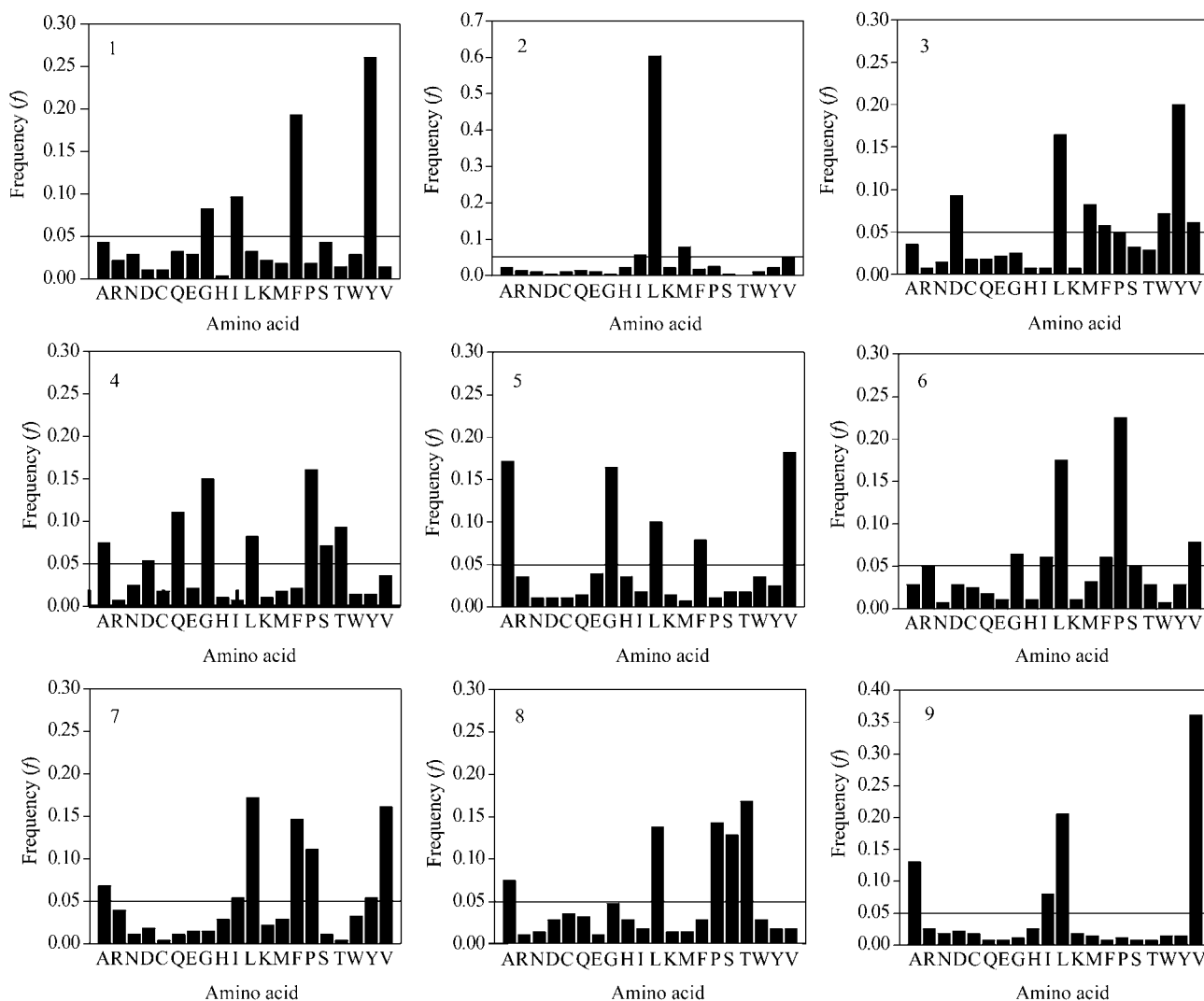


图3 20种天然氨基酸在遗传算法优化得到的抗原肽种群中9个位置的出现频率分布情况  
Figure 3 Frequency distributions of 20 natural amino acids at the nine positions at antigen peptide population optimized by genetic algorithm

体规模呈几何级数的降低. 简单计算表明, 本文设计的 CTL 表位库将 9 肽全合成空间由  $5.12 \times 10^{11}$  ( $20^9$ ) 降低至  $2.82 \times 10^6$  ( $4 \times 3 \times 7 \times \dots \times 4$ ), 这使得 CTL 表位的组合合成及疫苗活性筛选成为可能. 下面进一步对抗原肽不同位置所包含的氨基酸类型进行分析(抗原肽 1~9 号位置以下统一称为 P1~P9):

(1) HLA-A\*0201 限制性 CTL 表位的 P2 和 P9 位残基为抗原肽的锚定残基(anchor residue), 这两处位置氨基酸类型直接决定抗原肽与 HLA-A\*0201 分子的亲和能力. Falk 等<sup>[11]</sup>指出, HLA-A\*0201 分子在其肽结合沟槽(peptide-binding cleft)对应抗原肽序列 P2, P9 处形成疏水口袋, 因此使得抗原肽相应位置通常出现强疏水性的氨基酸, 特别是 P2 的亮氨酸(L)和 P9 的缬氨酸(V)/亮氨酸(L)为经典的锚定残基类型. 从图 3 统计分布情况可以清楚地看到本文计算结果非常符合实际情况, 在 P2 位置亮氨酸(L)出现的频率达到了 60.36%, 远远高于其他残基类型. 另外, 蛋氨酸(M)和异亮氨酸(I)这两种锚定残基允许出现的氨基酸类型也略高于平均值; 而在 P9 位置高于平均概率的 4 种氨基酸分别为缬氨酸(V)、丙氨酸(A)、亮氨酸(L)和异亮氨酸(I), 显然这些氨基酸侧链都为疏水性的烃基, 符合锚定残基要求.

(2) 抗原肽的 P1, P3 和 P7 位残基被 Ruppert 等<sup>[12]</sup>定义为第二锚定残基(secondary anchor residue), 其所起作用仅次于 P2 和 P9 位锚定残基. 根据 Saper 和 Madden 等<sup>[13,14]</sup>报道的 HLA-A\*0201 分子与抗原肽复合物晶体结构可以看到, P1 位置周围(图 4A)分别由 5 个酪氨酸(Y)、一个谷氨酸(E)和一个赖氨酸(K)环绕形成了一个极性口袋, 有利于能够与之形成  $\pi$  键堆积, 并具有氢键受体原子的酪氨酸(Y)残基与之发生结合, 这从图 3 统计结果得到了较好反映. 本文模型同时也将与酪氨酸(Y)结构类似的苯丙氨酸(F)引入虚拟组合库. 另外 Kirksey 等<sup>[15]</sup>的研究发现, 在某些时候(如爱滋病毒逆转录酶的抗原决定簇)抗原肽的 P1 位也常出现异亮氨酸(I)置换的情况, 该类情况亦在本文筛选结果中得到反映. 对于 P3 位置, 图 4D 处形成了一个显著的疏水口袋, Sarobe 等<sup>[16]</sup>研究表明, 此处残基类型倾向于苯丙氨酸(F)和酪氨酸(Y)等带有疏水芳环的氨基酸, 模型给出了较为一致的结果, 同时还包含了色氨酸(W)、亮氨酸(L)等疏水性的氨基酸类型. P7 位周围(图 4E)绝大多数区域为疏水性残基所环绕, 同时该处一侧也被强极性的精氨酸(R)所占据, 从而表现出两性环境, 由此 Saper 等<sup>[13]</sup>认为, 此处出现具有较小疏水侧链的残基类型[如图 3 给出的缬氨酸(V)、丙氨酸(A)和脯氨酸(P)]有利于提高抗原肽的整体亲和性.

(3) 对于抗原肽的 P4, P5, P6 和 P8 位虽所起到的作

用没有锚定残基那么显著, 但是它们也在一定程度上影响着 HLA-A\*0201 分子与 CTL 表位的结合过程. Kubo 等<sup>[17]</sup>曾在其扩展基序图中定义了这些位置的有利和不利残基类型, 经比较发现, 本文筛选结果与该图较为符合, 如 P4 的丝氨酸(S)和苏氨酸(T)、P5 的苯丙氨酸(F)和丙氨酸(A)以及 P8 的脯氨酸(P)等等, 限于篇幅这里不做详细讨论.

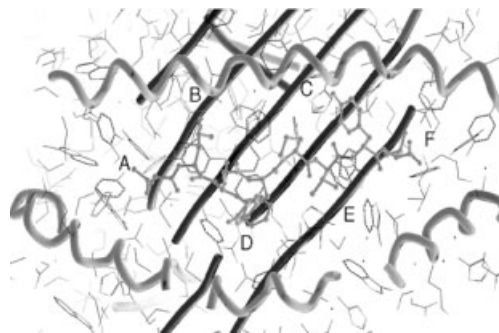


图 4 HLA-A\*0201 分子与抗原肽复合物晶体结构

Figure 4 Crystal structure of HLA-A\*0201/peptide complex

表 1 设计得到的 HLA-A\*0201 限制性 CTL 表位虚拟组合库的结构单元

Table 1 Structural elements of virtual combinational library for HLA-A\*0201-restricted CTL epitope

Position	Number of favorable residue	Favorable residue type
1	4	G, I, F, Y
2	3	I, L, M
3	7	D, L, M, F, W, Y, V
4	8	A, D, Q, G, L, P, S, T
5	5	A, G, L, F, V
6	6	G, I, L, F, P, V
7	7	A, I, L, F, P, Y, V
8	5	A, L, P, S, T
9	4	A, I, L, V

### 3 结论

基于定量序效模型的计算机辅助虚拟疫苗库设计方案, 成败关键在于定量序效模型的预测能力, 很显然, 这受到分子结构表征技术、统计建模方法及样本分布的合理性等诸多因素影响. 为了达到上述目标, 我们收集了 516 种氨基酸理化性质参数, 并通过 PCA 进行变量压缩和降维处理, 得到了一种新的氨基酸描述子: 氨基酸综合性质得分(SP-score). 采用 SP-score 结合 GA-PLS 建模技术对 266 个 HLA-A\*0201 限制性 CTL 表位的亲和性能, 成功建立了 QSAM 模型, 并以此为基础通过 GA 优化得到了一个在统计学意义上具有较高整体活性的虚拟组合疫苗库. 从设计的合理性和过程的可行性来

看, 基于定量序效模型的虚拟组合库设计方案是较为成功的, 并且实践过程中, 使用者可以根据具体情况在此基础上进一步缩小设计得到的组合库的规模. 但是同时应该看到, 由于 QSAM 模型预测精度限制、IBS 假设与实际情况的差异等问题也引起了 QSAM-based Virtual Vaccine Library 的某些偏差(如入选组合库的个别氨基酸类型与实际环境要求不太一致), 其有待进一步完善.

## References

- 1 Hagmann, M. *Science* **2000**, 290, 80.
- 2 Weber, L. *QSAR Comb. Sci.* **2005**, 24, 809.
- 3 Jonsson, J.; Norberg, T.; Carlsson, L.; Gustafsson, C.; Wold, S. *Nucleic Acids Res.* **1993**, 20, 733.
- 4 Kidera, A.; Konishi, Y.; Oka, M.; Ooi, T.; Scheraga, H. *J. Protein Chem.* **1985**, 4, 23.
- 5 Hellberg, S.; Sjostrom, M.; Skagerberg, B.; Wold, S. *J. Med. Chem.* **1987**, 30, 1126.
- 6 Collantes, E. R.; Dunn, W. J. *J. Med. Chem.* **1995**, 38, 2705.
- 7 Zhou, P.; Tian, F. F.; Li, B.; Wu, S. R.; Li, Z. L. *Acta Chim. Sinica* **2006**, 64, 691 (in Chinese).
- (周鹏, 田菲菲, 李波, 吴世容, 李志良, 化学学报, **2006**, 64, 691.)
- 8 Blythe, M. J.; Doytchinova, I. A.; Flower, D. R. *Bioinformatics* **2002**, 18, 434.
- 9 Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Modell.* **2002**, 20, 269.
- 10 Parker, K. C.; Bednarek, M. A.; Coligan, J. E. *J. Immunol.* **1994**, 152, 163.
- 11 Falk, K.; Röttschke, O.; Stefanovic, S.; Jung, G.; Rammensee, H.-G. *Nature* **1991**, 351, 290.
- 12 Ruppert, J.; Sidney, J.; Celis, E.; Kubo, R. T.; Grey, H. M.; Sette, A. *Cell* **1993**, 74, 929.
- 13 Saper, M. A.; Bjorkman, P. J. *J. Mol. Biol.* **1991**, 219, 277.
- 14 Madden, D. R.; Garboczi, D. N. *Cell* **1993**, 75, 693.
- 15 Kirksey, T. J.; Pogue-Caley, R. R.; Frelinger, J. A.; Collins, E. J. *J. Biol. Chem.* **1999**, 274, 37259.
- 16 Sarobe, P.; Pendleton, C. D.; Akatsuka, T. D.; Engelhard, V. H.; Feinstone, S. M.; Berzofsky, J. A. *J. Clin. Invest.* **1998**, 102, 1239.
- 17 Kubo, R. T.; Sette, A.; Grey, H. M.; Appella, E.; Sakaguchi, K.; Zhu, N. Z.; Arnott, D.; Sherman, N.; Shabanowitz, J.; Michel, H. *J. Immunol.* **1994**, 152, 3913.

(A0603103 SONG, J. P.)