



## 论文

# 人类白细胞抗原 HLA-B\*5701 和 B\*5801 肽亲和特异性

张亚兰<sup>①</sup>, 梅虎<sup>①②\*</sup>, 王青<sup>②</sup>, 谢江安<sup>②</sup>, 吕娟<sup>②</sup>, 潘显超<sup>②</sup>, 谭文<sup>②</sup><sup>①</sup> 重庆大学生物工程学院, 生物流变科学与技术教育部重点实验室, 重庆 400044;<sup>②</sup> 重庆大学生物工程学院, 重庆 400044

\* 联系人, E-mail: meihu@cqu.edu.cn

收稿日期: 2012-07-02; 接受日期: 2012-07-11

国家自然科学基金面上项目(批准号: 61073135)和重庆市自然科学基金重点项目(批准号: CSTC, 2009BA5068)资助

**摘要** 全基因组关联分析研究发现, abacavir 的严重毒副作用具 HLA-B\*5701 限制性, 而仅在 4 个氨基酸位点存在差异的 B\*5801 携带者却无明显反应, 推测是由于 B\*5801 无法结合致敏相关特异性多肽所致. 因此, 比较分析 B\*5701 和 B\*5801 的多肽亲和特性及 B\*5701 特异性肽库是探索 abacavir 限制性毒副作用机理的首要内容. 本文基于 IEDB 数据库, 采用 VHSE 结构表征方法, 建立了 HLA-B\*5701 和 B\*5801 多肽亲和活性的支持向量机预测模型, 其最优线性模型的  $R^2$ , 10-fold 交互验证  $Q^2$  和外部预测  $R_{PRE}^2$  分别为 0.7530, 0.7037, 0.6153(B\*5701)和 0.6074, 0.5966, 0.5762(B\*5801). 研究表明, 位于结合凹槽 B 口袋附近的 45(MET-THR)和 46(ALA-GLU)位突变残基对 B\*5701 和 B\*5801 结合肽 P2 位的氨基酸选择特异性影响较小, 二者在 P2 位优先选择的氨基酸种类基本一致, 仅在优先选择顺序上有所差异; 位于 C 口袋和 E 口袋间的 97 位(VAL-ARG)突变残基明显改变了 P7 位氨基酸的选择偏好. 对于 HLA-B\*5701, 其 P7 位优先选择大体积正电性氨基酸(ARG, HIS, LYS), 而 B\*5801 则优先选择非极性的疏水性氨基酸, 且正电性氨基酸对结合极为不利.

**关键词**人类白细胞抗原  
B\*5701  
B\*5801  
SVM P-I 理论  
VHSE

近年来, 随着高通量和高精度基因测序技术的发展, 全基因组关联分析(genome wide association study, GWAS)被越来越多地应用到药物严重不良反应(serious adverse drug reactions, SADR)研究中. 研究发现, 具有遗传异质性的药物毒副作用(Type B)<sup>[1,2]</sup>与人类白细胞抗原(human leucocytes antigen, HLA)基因密切相关. 最早引起广泛关注的是抗艾滋病药物阿巴卡韦(abacavir)的 SADR, 其临床症状表现为: 发热、皮疹、胃肠道症状、头昏或心神不宁等不良反应,

严重可危及生命<sup>[3]</sup>. 2002 年, Hetherington 等人<sup>[4]</sup>和 Mallal 等人<sup>[5]</sup>研究发现, abacavir 的遗传异质性毒副作用与 HLA-B\*5701 等位基因密切相关, 而这种关联在白种人群中尤为显著<sup>[6]</sup>. Chessman 等人<sup>[7]</sup>研究发现, abacavir 可刺激 CD8<sup>+</sup>T 细胞释放细胞因子而引发超敏反应, 而该反应具 HLA-B\*5701 限制性, 提示 abacavir 并非直接与 T 细胞发生作用, 而是依赖传统抗原呈递过程来激发 T 细胞免疫反应.

针对药物的遗传异质性毒副作用, Martin 等

人<sup>[8]</sup>、Gerber 和 Pichler<sup>[9]</sup>提出了与以往“半抗原”观点不同的“P-I”理论。“P-I”理论认为, 除以半抗原方式引发超敏反应外, 药物也可能通过非共价键与 HLA-peptide(pHLA)形成 pHLA-drug 复合物, 共同激发 T 细胞免疫反应, 而随后的相关实验也陆续证实了这种非共价作用的存在<sup>[10-13]</sup>。

根据“P-I”理论, 药物以非共价键与 pHLA 复合物相互作用是引发 SADR 的关键一步, 这种非共价结合改变了 pHLA 与 TCR 接触端分子表面的特征, 从而干扰了正常 T 细胞的免疫应答, 导致 SADR 的发生。HLA-B\*5701 和 B\*5801 同属于 HLA-B17 家族<sup>[14]</sup>, 且仅在 4 个氨基酸位点上存在差异, 因此二者的多肽亲和和特征具较高的相似性。然而, 为何 abacavir 与 pHLA-B\*5701 作用可以激发特异性 CD8<sup>+</sup>T 细胞的免疫应答, 而与 pHLA-B\*5801 则反应较弱。目前多数观点倾向认为, 是由于 B\*5801 无法结合致敏相关特异性多肽所致<sup>[11]</sup>。2010 年, Mandvi 等人<sup>[15]</sup>从接受 abacavir 治疗的 HLA-B\*5701 阳性患者的抗原递呈细胞中提取出具有免疫原性的天然多肽, 研究表明该天然多肽可选择性激发 abacavir 特异性 T 细胞的免疫应答。因此, 比较和分析 HLA-B\*5701 和 B\*5801 的多肽亲和特性是探索 abacavir 的 B\*5701 限制性 SADR 机理的首要研究内容。

本文基于 IEDB(Immune Epitope Databass)数据库<sup>[16,17]</sup>中 HLA-B\*5701 和 B\*5801 的多肽亲和数据, 采用支持向量机(support vector machine, SVM)结合多元逐步回归(multiple stepwise regression, MSR)变量筛选方法<sup>[18]</sup>建立了 HLA-B\*5701 和 B\*5801 多肽亲和活性的定量构效关系(quantitative structural-activity relationships, QSAR)模型, 取得了较优的建模结果, 根据所建模型对 B\*5701 和 B\*5801 多肽结合特征进行了细致的比较研究, 结果对于探寻 abacavir 的 B\*5701 限制性 SADR 机理以及相关表位疫苗设计均具有理论意义。

## 1 原理与方法

### 1.1 数据来源与处理

截止 2011 年 10 月, 从 IEDB 数据库共收集到 HLA-B\*5701 相关多肽序列 1895 条, HLA-B\*5801 相关多肽序列 3069 条。去除活性缺失及重复序列后,

共得到 1155 条 HLA-B\*5701 相关 9 肽和 1867 条 HLA-B\*5801 相关 9 肽。为降低无活性样本引起的建模误差, 活性样本: 无活性样本数设定为 1:1, 训练集: 测试集样本数也设定为 1:1。对于 HLA-B\*5701, 具亲和活性样本共计 198 条, 无活性样本 957 条( $EC_{50} > 20000$ ), 即随机选择 99 条活性样本, 99 条无活性样本组成训练集, 余下 99 条活性样本连同随机选择的 99 条无活性样本组成测试集。对于 HLA-B\*5801, 活性样本共计 448 条, 无活性样本 1419 条, 即随机抽取 224 条活性样本和 224 条无活性样本组成训练集, 剩下 224 条活性样本和随机抽取的 224 条无活性样本作为测试集。样本活性标度为半数有效浓度的负对数( $-\lg EC_{50}$ )。

### 1.2 结构表征

氨基酸结构描述符 VHSE(principal component score vector of hydrophobic, steric, and electronic properties)来源于 Mei 等人<sup>[19]</sup>的研究。其从历年文献和数据库中的 200 余种物化性质和参数中, 经分类和筛选得到 50 种天然氨基酸的性质和参数, 其中疏水性性质 18 个, 立体性质 17 个, 电性性质 15 个。针对 3 个性质和参数矩阵, 经主成分特征提取产生 2, 2 和 4 个主成分得分, 分别累计解释原始数据矩阵 74.33%, 78.68% 和 79.97% 的方差。当用 8 个主成分得分来替代原始数据矩阵, 并作为氨基酸结构描述子时, 原始数据的信息损失相对较少。为方便起见, 他们将该矢量称之为 VHSE。其中 VHSE<sub>1</sub> 和 VHSE<sub>2</sub> 代表氨基酸的疏水性特征; VHSE<sub>3</sub> 和 VHSE<sub>4</sub> 代表氨基酸的立体特征; VHSE<sub>5</sub>~VHSE<sub>8</sub> 则代表氨基酸的电性特征(表 1)。

对于本文 HLA-B\*5701 和 B\*5801 结合 9 肽, 当采用 VHSE 进行结构表征时, 每个样本可产生  $72(9 \times 8 = 72)$  个 VHSE 描述子变量, 依次用  $V_{11}, V_{12}, V_{13}, \dots, V_{ij}, \dots, V_{98}$  表示, 其中下标  $i$  表示氨基酸残基位点,  $j$  表示 VHSE 描述子序号, 如  $V_{13}$  表示肽的第 1 个位点(P1)的 VHSE<sub>3</sub> 描述子。

### 1.3 MSR-SVM 建模

采用 SVM 结合 MSR 的方法, 即 MSR-SVM 进行变量筛选和建模, 即在多元逐步回归变量筛选结果的基础上采用 SVM 进行建模。MSR 是一种“有进有出”的算法, 当变量数相对较少时, 它是寻找最优子

空间的一种快速有效的方法. 同时, SVM 可有效避免多元线性回归中的小样本建模和过拟合等问题.

SVM<sup>[18,20,21]</sup>起初是用于解决线性可分情况下两类样本的分类问题, 其核心思想是找到一个最优分类超平面, 使两类样本的分类间隔(margin)最大化. 对于非线性问题, 首先经一个非线性映射  $\Phi$ , 将样本映射到一个高维特征空间, 然后用线性方法来解决. 支持向量机不仅可以解决分类问题, 同时也可解决回归问题, 即寻求一个最优超平面, 使得在给定精度  $\varepsilon(\varepsilon \geq 0)$  条件下可以无误差地拟合  $y$ , 即所有样本点到最优超平面的距离都不大于  $\varepsilon$ . 高维映射经核函数:  $K(x, xi) = \Phi(x) \cdot \Phi(xi)$  实现, 其是支持向量机的核心技术, 可有效解决维数灾难及计算复杂性等问题.

根据对已有 HLA 与 peptide 亲和特性研究经验<sup>[22-24]</sup>, 采用线性核和 RBF 核为代表对 SVM 回归建模进行研究. 模型质量评价采用  $R^2$  (determinant coefficients), 10 折交叉验证  $Q^2$  (10-fold crossvalidation  $R^2$ ), RMS (root of mean squared error) 和  $R_{PRE}^2$  (determinant coefficients of test set) 进行综合评价. 其中,  $R_{PRE}^2$  定义为外部测试集过原点的决定系数.

## 2 结果与分析

### 2.1 MSR-SVM 建模

对于 HLA-B\*5701 和 HLA-B\*5801, 其 MSR-SVM 分别筛选出 4 和 3 组最优变量子集, 建模结果见表 2 和 3. 可见, 采用 RBF 核的 SVM 结果略优于线性 SVM 建模结果, 但总体上无显著性差异.

对于 HLA-B\*5701 多肽亲和模型(表 2), 兼顾模型拟合和预测结果、模型复杂度及可解释性, 选择 ID 5 为最优变量子集, 其所含变量数为 14, 基于该变量子集的线性 SVM 模型为最优模型. 图 1 为外部测试集多肽亲和活性的观察值与预测值通过原点的线性回归图, 可见模型具有较高的预测准确性, 模型决定系数  $R_{PRE}^2$  为 0.6153, 斜率为 0.9818.

对于 HLA-B\*5801, ID 4 为最优变量子集, 其包含 15 个变量, 线性核 SVM 模型为最优模型(表 3). 图 2 为外部测试集多肽亲和活性的观察值与预测值通过原点的线性回归图, 可见该模型同样具有较高的预测准确性, 模型决定系数  $R_{PRE}^2$  为 0.5762, 斜率为 0.9759.

为进一步验证模型质量, 随机划分训练集和预

表 1 20 种氨基酸的 VHSE 描述子

AA	VHSE1	VHSE2	VHSE3	VHSE4	VHSE5	VHSE6	VHSE7	VHSE8
Ala A	0.15	-1.11	-1.35	-0.92	0.02	-0.91	0.36	-0.48
Arg R	-1.47	1.45	1.24	1.27	1.55	1.47	1.30	0.83
Asn N	-0.99	0.00	-0.37	0.69	-0.55	0.85	0.73	-0.80
Asp D	-1.15	0.67	-0.41	-0.01	-2.68	1.31	0.03	0.56
Cys C	0.18	-1.67	-0.46	-0.21	0.00	1.20	-1.61	-0.19
Gln Q	-0.96	0.12	0.18	0.16	0.09	0.42	-0.20	-0.41
Glu E	-1.18	0.40	0.10	0.36	-2.16	-0.17	0.91	0.02
Gly G	-0.20	-1.53	-2.63	2.28	-0.53	-1.18	2.01	-1.34
His H	-0.43	-0.25	0.37	0.19	0.51	1.28	0.93	0.65
Ile I	1.27	-0.14	0.30	-1.80	0.30	-1.61	-0.16	-0.13
Leu L	1.36	0.07	0.26	-0.80	0.22	-1.37	0.08	-0.62
Lys K	-1.17	0.70	0.70	0.80	1.64	0.67	1.63	0.13
Met M	1.01	-0.53	0.43	0.00	0.23	0.10	-0.86	-0.68
Phe F	1.52	0.61	0.96	-0.16	0.25	0.28	-1.33	-0.20
Pro P	0.22	-0.17	-0.50	0.05	-0.01	-1.34	-0.19	3.56
Ser S	-0.67	-0.86	-1.07	-0.41	-0.32	0.27	-0.64	0.11
Thr T	-0.34	-0.51	-0.55	-1.06	0.01	-0.01	-0.79	0.39
Trp W	1.50	2.06	1.79	0.75	0.75	-0.13	-1.06	-0.85
Tyr Y	0.61	1.60	1.17	0.73	0.53	0.25	-0.96	-0.52
Val V	0.76	-0.92	0.17	-1.91	0.22	-1.40	-0.24	-0.03

测集样本, 采用相同的变量子集和线性核参数进行 10 次重复建模, 均取得了较优的建模和预测结果. 所得 10 个模型的  $R^2$ ,  $Q^2$ ,  $R_{PRE}^2$  的平均值分别为: 0.6958, 0.6661, 0.6557(B\*5701) 和 0.6027, 0.5926, 0.5796(B\*5801).

## 2.2 HLA-B\*5701 和 HLA-B\*5801 多肽结合特性

HLA-B\*5701 和 B\*5801 同属于 HLA-B17 家族.

对于由  $\alpha 1$  和  $\alpha 2$  结构域组成的肽槽, 两者氨基酸序列基本一致, 仅在 4 个氨基酸位点上存在差异(图 3). 除了 103 号位氨基酸位于结合口袋之外, 45, 46 和 97 号位氨基酸分别位于结合凹槽的 B 口袋和 E 口袋<sup>[25]</sup>, 分别与亲和 9 肽的 P2 位氨基酸(一级锚定位点)和 P7 位氨基酸(二级锚定位点)发生相互作用. 因此, 上述 3 个位点差异主要影响 B\*5701 和 B\*5801 结合 9 肽的 P2 和 P7 位氨基酸残基选择偏好.

表 2 HLA-B\*5701 多肽亲和活性 SVM 建模结果

ID	变量数	Kernal	C	$\epsilon$	$\gamma$	$R^2$	RMS	$Q^2(10-fold)$	$R_{PRE}^2$
1	72	Linear	1.9691	0.0317	-	0.7083	0.7013	0.6055	0.5386
		RBF	16.4872	0.0811	1.3591	0.9957	0.6542	0.6544	0.6031
2	32	Linear	42.1016	0.3012	-	0.7882	0.5918	0.7273	0.6149
		RBF	12.0623	0.1716	1.1994	0.9316	0.5730	0.7449	0.6244
3	25	Linear	10.6449	0.0672	-	0.7450	0.6042	0.7121	0.5746
		RBF	39.5508	0.0026	0.5000	0.8320	0.5845	0.7282	0.5765
4	18	Linear	10.000	0.3679	-	0.7712	0.6078	0.7068	0.6130
		RBF	5.3526	0.3012	1.3591	0.8140	0.5840	0.7270	0.6465
5	14	Linear	27.1828	0.3679	-	0.7530	0.6037	0.7037	0.6153
		RBF	10.0000	0.3679	1.3591	0.7964	0.6009	0.7061	0.6761

表 3 HLA-B\*5801 多肽亲和活性 SVM 建模结果

ID	变量数	Kernal	C	$\epsilon$	$\gamma$	$R^2$	RMS	$Q^2(10-fold)$	$R_{PRE}^2$
1	72	Linear	10.0000	0.3679	-	0.6213	0.8510	0.5674	0.5709
		RBF	4.7237	0.1180	1.3591	0.7759	0.8197	0.5977	0.5929
2	35	Linear	25.5359	0.2497	-	0.6420	0.8076	0.6090	0.5088
		RBF	23.9888	0.1716	1.3591	0.7917	0.7733	0.6404	0.5453
3	28	Linear	65.2082	0.2346	-	0.6167	0.8019	0.6164	0.4968
		RBF	21.1700	0.3012	1.3591	0.7293	0.7789	0.6362	0.5537
4	15	Linear	27.1828	0.3679	-	0.6074	0.8220	0.5966	0.5762
		RBF	69.4138	0.1945	1.3591	0.7137	0.7585	0.6564	0.6018

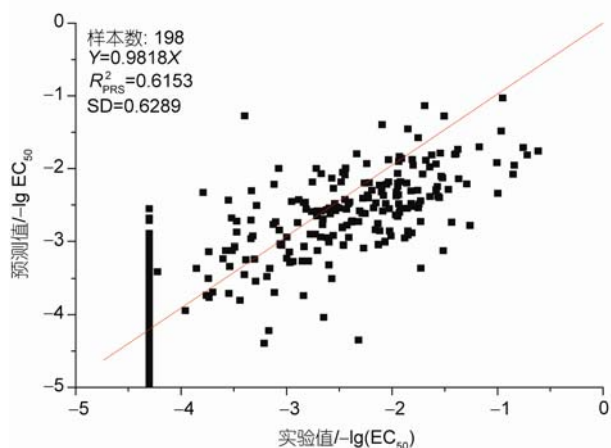


图 1 HLA-B\*5701 多肽亲和活性的观察值与预测值

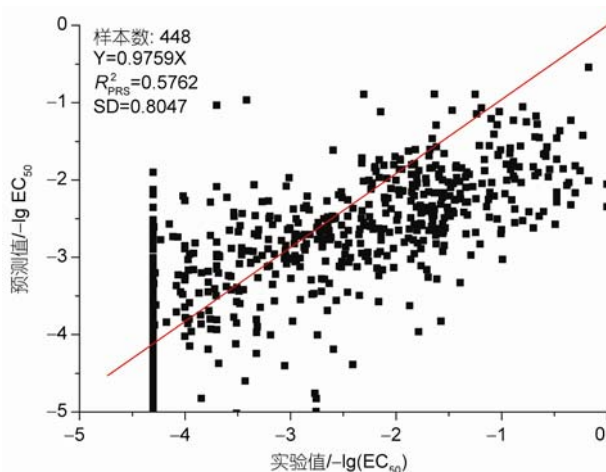


图 2 HLA-B\*5801 多肽亲和活性的观察值与预测值

对于 B\*5701, 其 B 口袋由 TYR7, ASN66 和 GLU63 等大体积氨基酸组成袋口, 由 VAL24, GLY26, MET45 和 ALA46 等组成偏疏水性质的底部 (图 4A); 而 B\*5801 的 45 和 46 号位氨基酸为亲水性的 THR 和负电性的 GLU. 由于 46 号位氨基酸位于 B

口袋底部较深位置, 因此对 B 口袋特性及 P2 位氨基酸选择偏好影响相对较小. 总体上看, B\*5801 的 B 口袋更具负电性和亲水特征.

97 号氨基酸残基位于肽结合凹槽的 E 口袋和 C 口袋之间, 主要影响 P7 号位氨基酸的选择特异性.

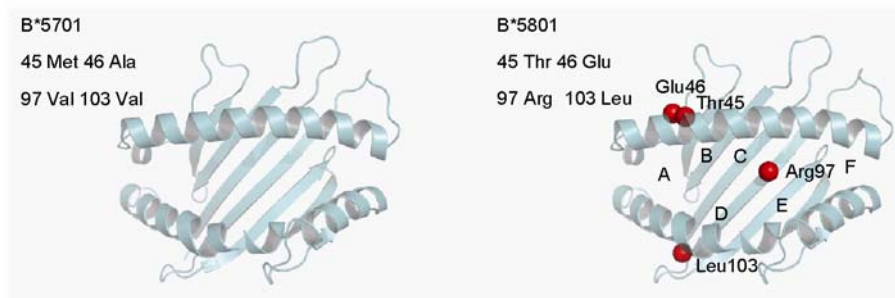


图 3 HLA-B\*5701 和 B\*5801 三维结构模型

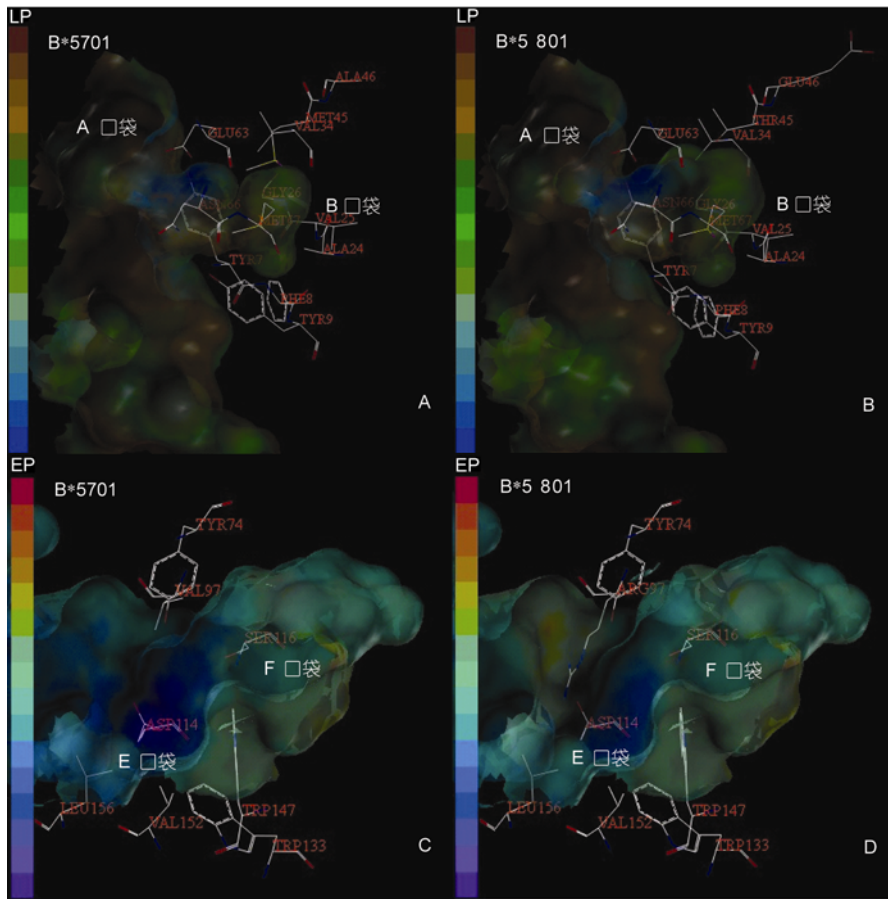


图 4 HLA-B\*5701 和 B\*5801 的 B 口袋和 E 口袋

A: B\*5701 结合凹槽中 B 口袋的疏水性质; B: B\*5801 结合凹槽中 B 口袋的疏水性质; C: B\*5701 结合凹槽中 E 口袋的电性性质; D: B\*5801 结合凹槽中 E 口袋的电性性质

对于 B\*5701 和 B\*5801, 其 E 口袋袋口主要由 VAL152, TRP147, LEU156 和 TRP133 等氨基酸组成, 口袋底部则由负电性的 ASP114 和 SER116 等组成(图 4C). 对于 B\*5701, 其 97 号为小体积疏水性的 VAL, 空间位阻相对较小; 而对于 B\*5801, 其 97 号氨基酸为空间位阻较大的正电性的 ARG. 总体看来, B\*5701 的 E 口袋更具负电性, 空间位阻较小, 更易结合大体积的氨基酸残基; 而 B\*5801 的 E 口袋负电性相对较小, 且当 P7 位为大体积的氨基酸残基时, 存在空间位阻效应.

(1) HLA-B\*5701 多肽亲和特征. 对于 B\*5701, 其最优 SVM 线性模型各变量的贡献权重如图 5 所示. 可见, 14 个 VHSE 变量共涉及 P1, P2, P3, 4, P5, P7 和 P9 共 7 个位点. 从变量权重的绝对值大小看, 对结合影响最大的为 P2 和 P9 位点, 其次为 P1, P3 和 P7 位点, 贡献最小的为 P4 和 P5 位点. 而 P6 和 P8 位点对 B\*5701-Peptide 的亲合模型无贡献.

根据最优模型对各位点模拟突变的预测结果, 分别进行标准化处理, 得到 B\*5701 结合肽各位点的氨基酸选择偏好(表 4). 由于进行了标准化处理, 所以数值大小可表征各位点的氨基酸选择偏好, 即数值越大表示对结合越有利, 数值越小表示对结合越不利.

对于 P2 位, 其包含 VHSE3, VHSE4, VHSE5 和

VHSE6 共 4 个变量, 涉及氨基酸的立体和电性性质. 根据表 4, 同时结合 P2 位变量载荷大小, 综合分析表明, P2 位为 ALA, GLY, SER, THR 和 CYS, 即小体积或极性的氨基酸对亲和活性有利; 而 TRP, TYR, GLU 和 ARG 等大体积或带电荷的氨基酸对结合不利(表 4).

P9 位包含 3 个 VHSE 变量, 分别表征疏水、立体和电性性质. 根据表 4 并结合 3 个变量的载荷大小, 综合分析表明, TRP, PHE, TYR, LEU 和 MET 等大体积、疏水性的氨基酸对活性有利; 反之, PRO 和 ASP 等小体积或带电荷的氨基酸对亲和活性不利.

P7 位拥有 2 个 VSHE 变量, 分别涉及立体和电性特征. 分析表明, P7 位优先选择 ARG, HIS, LYS 等大体积的正电性氨基酸; 而小体积或疏水性的氨基酸 GLY, ALA 和 PRO 对结合不利.

P3 位包含 2 个 VHSE 变量, 均为电性特征. 分析表明, P3 位偏向于结合 CYC, PHE, TRP 和 TYR 等疏水、大体积的氨基酸; 而 GLY, GLU 和 ASP 等亲水或带电荷的氨基酸则不利于结合. P1, P4 和 P5 位各包含一个 VHSE 变量, 涉及电性和立体特征.

(2) HLA-B\*5801 多肽亲和特征. 与 B\*5701 模型一样, HLA-B\*5801 最优模型共涉及 P1, P2, P3, P4, P5, P7 和 P9 共 7 个位点(图 6). P2 和 P9 位对亲合活性影响最大, 其次为 P7, P3, P1 和 P4 位点, 贡献最小

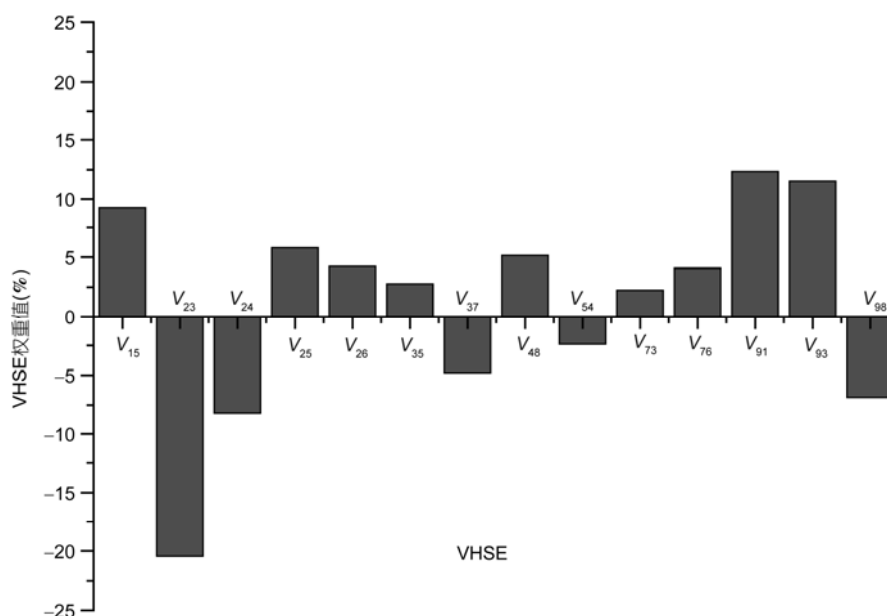


图 5 HLA-B\*5701 模型的 VHSE 变量权重图

的为 P5 位. 而 P6 和 P8 位对 HLA-B\*5801-Peptide 的亲模型无贡献. 上述结果与 B\*5701 结合肽各位点贡献基本一致.

根据最优模型对各位点模拟突变的预测结果, 分别进行标准化处理, 得到 B\*5801 结合肽各位点的

氨基酸选择特性(表 5)

对于 P2 位, 其包含 VHSE3, VHSE5, VHSE7 和 VHSE8 这 4 个变量, 涉及氨基酸的立体和电性性质. 根据各变量载荷并结合表 5 分析表明, GLY, ALA, SER, CYS 和 THR, 即小体积不带电荷的氨基酸对结

表 4 HLA-B\*5701 多肽结合特性

	P1	P2	P3	P4	P5	P7	P9
Ala A	0.0220	1.7232	-0.2980	-0.4760	0.9249	-1.3160	-0.3307
Arg R	1.5494	-0.9150	-0.3108	0.8301	-1.2714	1.6815	-0.5240
Asn N	-0.5498	0.0843	-0.9079	-0.7970	-0.6859	0.4621	-0.1673
Asp D	-2.6776	-0.3920	-1.4196	0.5565	0.0135	0.7873	-1.0499
Cys C	0.0037	0.9124	1.3781	-0.1890	0.2052	0.6832	-0.0351
Gln Q	0.0886	-0.1320	0.2204	-0.4100	-0.1579	0.3969	-0.1027
Glu E	-2.1555	-1.2830	-1.8985	0.0160	-0.3610	-0.0750	-0.4899
Gly G	-0.5316	1.4043	-1.9901	-1.3390	-2.2824	-2.1160	-0.6203
His H	0.5053	0.0477	-0.5345	0.6456	-0.1949	1.1300	-0.3905
Ile I	0.3007	0.2016	0.2944	-0.1280	1.8026	-1.0660	0.7957
Leu L	0.2185	-0.1810	0.0442	-0.6220	0.8042	-0.8990	1.1088
Lys K	1.6373	-0.2560	-0.5466	0.1279	-0.8045	0.8287	-0.2505
Met M	0.2286	-0.3570	0.8561	-0.6760	0.0001	0.2728	1.0666
Phe F	0.2499	-0.8300	1.2632	-0.1970	0.1591	0.6654	1.2803
Pro P	-0.0127	0.1988	0.1570	3.5562	-0.0477	-1.2360	-2.2402
Ser S	-0.3241	1.3127	0.3774	0.1072	0.4085	-0.2990	-0.8984
Thr T	-0.0557	1.0692	0.6445	0.3937	1.0570	-0.2640	-0.6617
Trp W	0.7492	-2.0560	1.2512	-0.8520	-0.7483	0.7489	2.0720
Tyr Y	0.5310	-1.3540	1.1003	-0.5200	-0.7337	0.7432	1.1683
Val V	0.2225	0.8013	0.3191	-0.0270	1.9129	-1.1300	0.2696

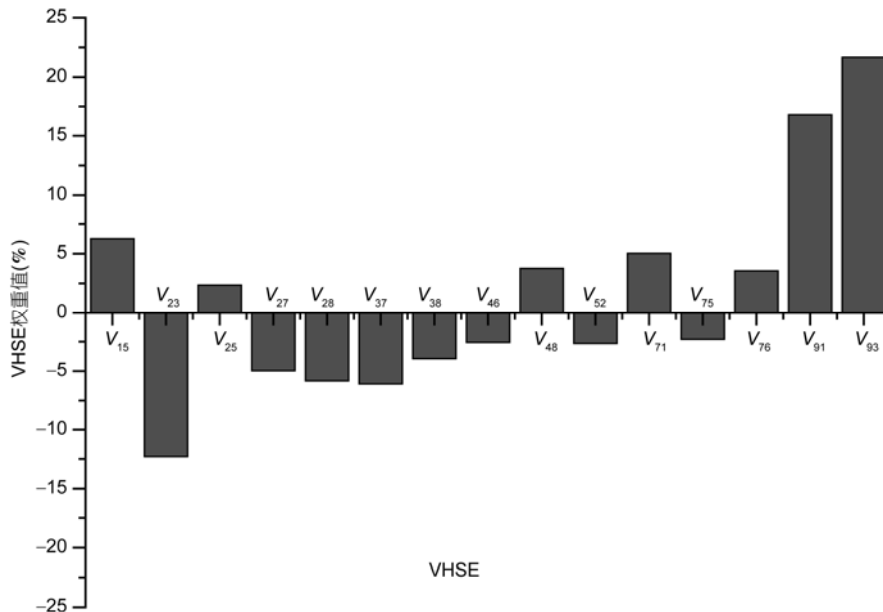


图 6 HLA-B\*5801 模型的 VHSE 变量权重图

合有利, 而 ARG 和 TRP 等大体积或带电荷的氨基酸对结合不利. 与 B\*5701 相比, 虽然两者在 P2 位优先选择的氨基酸种类基本一致, 但在优先选择顺序上却有所差异.

P9 位包含 2 个 VHSE 变量, 分别表征疏水 (VHSE1) 和立体性质 (VHSE3). 与 B\*5701 相比, P9 位的 VHSE1 和 VHSE3 对活性的贡献明显增大, 而表征电性特征的 VHSE8 则对活性没有贡献. 结合表 5 分析表明, P9 位优先选择 TRP, PHE 和 TYR 等大体积疏水性质的氨基酸, 而 GLY, SER 和 ALA 等小体积或亲水性质的氨基酸对结合不利.

P7 位包含 3 个 VHSE 变量, 分别表征疏水和电性性质. 根据模型对 P7 位氨基酸突变 9 肽的亲亲和活性预测结果分析发现, PHE, TRP 和 MET 等疏水性质的氨基酸对结合有利; 同时电性性质对其也有一定影响, 如带负电荷的 ASP 有利于结合, 但带正电荷氨基酸 (如 ARG 和 LYS) 则对结合不利. 可见, 受 97(VAL-ARG) 位突变残基的影响 (图 4D), HLA-B\*5801 和 B\*5701 在 P7 位氨基酸选择偏好上有明显差异, 即 B\*5801 优先结合疏水性的氨基酸, 且正电性性质的氨基酸对结合不利; 而 B\*5701 优先结合大体积的正电性氨基酸.

P3 和 P4 位均包含 2 个表征电性性质的 VHSE 变

量. 综合分析表明, P3 位为 TRP, CYC, MET 和 PHE 等对结合有利, 而 PRO, ARG 和 LYS 等氨基酸对结合不利. P4 位则优先选择结合 PRO, 而 ASN 对结合不利.

P1 和 P5 位各包含 1 个 VHSE 变量, 分别表征电性和疏水性质. 结果表明, P1 位为 LYS 和 ARG, 即正电性的氨基酸对结合有利; P5 位为 CYC 和 GLY 等氨基酸对结合有利.

与 B\*5701 相比, 两者的氨基酸选择特性在 P7 位有明显差异, P1, P2, P3, P4, P5 和 P9 位无显著差异, 仅在优先选择的顺序上有所不同. 因此, 与 abacavir 的 HLA-B\*5701 限制性 SADR 相关的特异性多肽, 极可能与 B\*5701 结合肽 P7 位氨基酸残基的选择特性密切相关.

### 3 结论

全基因组关联分析结果显示, abacavir 的遗传异质性毒性与 HLA-B\*5701 基因密切相关. 根据最新提出的 P-I 理论, abacavir 与 pHLA-B\*5701 的非共价作用是引发 SADR 的关键一步, 并具 HLA-B\*5701 限制性, 且同属于 B17 家族的 B\*5801 等基因型无明显的 SADR 发生, 推测是由于 B\*5801 无法结合致敏相关

表 5 HLA-B\*5801 多肽结合特性

	P1	P2	P3	P4	P5	P7	P9
Ala A	0.0220	1.4903	0.1028	0.2300	1.1116	-0.5600	-1.0046
Arg R	1.5494	-1.6150	-1.4959	-0.3250	-1.4511	-1.3910	0.3333
Asn N	-0.5497	0.2908	0.0890	-1.1610	0.0021	-0.2750	-0.7384
Asp D	-2.6776	-0.4980	-0.4315	-0.4320	-0.6704	0.9392	-0.8366
Cys C	0.0036	1.0765	1.2337	-0.9250	1.6727	1.2042	-0.2799
Gln Q	0.0886	0.0415	0.4395	-0.5820	-0.1218	-0.9080	-0.2851
Glu E	-2.1555	-0.9990	-0.6320	0.1203	-0.3974	-0.5610	-0.4437
Gly G	-0.5315	2.4617	-0.3819	-0.2510	1.5299	-0.9400	-2.1704
His H	0.5055	-0.7820	-1.1074	-0.3440	0.2485	0.2547	0.0974
Ile I	0.3007	-0.1370	0.2041	0.9530	0.1351	0.1213	0.8001
Leu L	0.2185	-0.0260	0.4008	0.4182	-0.0688	0.4776	0.8123
Lys K	1.6373	-0.8770	-1.2050	-0.3370	-0.7029	-1.7150	0.0354
Met M	0.2288	0.1265	1.0827	-0.5760	0.5278	1.2249	0.7885
Phe F	0.2499	-0.4510	1.0462	-0.3350	-0.6070	1.9984	1.4359
Pro P	-0.0128	-0.6600	-2.4771	3.5702	0.1653	-0.8120	-0.2997
Ser S	-0.3241	1.2149	0.3555	-0.0960	0.8551	-0.4510	-1.1425
Thr T	-0.0557	0.6862	0.2473	0.3036	0.5066	-0.4080	-0.5884
Trp W	0.7491	-1.0690	1.3104	-0.5650	-2.0614	1.3980	2.0852
Tyr Y	0.5311	-0.6050	1.0372	-0.5600	-1.5979	0.7082	1.2017
Val V	0.2225	0.3306	0.1818	0.8927	0.9241	-0.3060	0.1994



的特异性多肽所致. 因此, 比较分析 HLA-B\*5701 和 B\*5801 的多肽亲和特性是探索 abacavir 的 B\*5701 限制性 SADR 机理的重要研究内容.

在前期 VHSE 氨基酸结构表征方法研究基础上, 采用 MSR-SVM 建模方法, 分别建立了 HLA-B\*5701 和 B\*5801 的多肽亲和特性的 SVM 模型, 取得了较优的建模结果, 其外部预测能力( $R_{PRE}^2$ )分别达到 0.6153 和 0.5762. 研究结果显示, P2, P9, P1, P3 和 P7 位点是影响多肽亲和活性的主要位点, 而 P4 和 P5 位贡献相对较小, P6 和 P8 位则对亲和活性无贡献. 依据最优线性 SVM 模型, 根据模拟突变的预测结果结合变量权重分析, 详细比较和分析了 B\*5701 和 B\*5801 多肽亲和特性. 研究显示, 45(MET-THR)和 46(ALA-GLU)

号位氨基酸突变, 对 B\*5701 和 B\*5801 的 P2 位氨基酸选择特异性影响不大, 在 P2 位优先选择的氨基酸种类基本一致, 仅优先选择的顺序稍有不同; 对于 97 号位(VAL-ARG)氨基酸突变, 明显改变了 P7 位氨基酸的选择特异性, 如 HLA-B\*5701 优先与大体积的正电性氨基酸结合; 而 B\*5801 则优先与非极性的疏水性氨基酸结合, 且电性氨基酸对结合极为不利. 因此, 与 abacavir 的 HLA-B\*5701 限制性 SADR 相关的特异性多肽, 可能与 B\*5701 结合肽的 P7 位氨基酸残基选择特性密切相关, 即该特异性多肽的 P7 位很有可能为 ARG, HIS 和 LYS. 综上, 本研究可为后续的 HLA-B\*5701 特异性亲和多肽库及 abacavir 的 B\*5701 限制性 SADR 机理研究提供参考依据.

## 参考文献

- 1 Park B K, Pirmohamed M, Kitteringham N R. Role of drug disposition in drug hypersensitivity: a chemical, molecular, and clinical perspective. *Chem Res Toxicol*, 1998, 11: 969–988
- 2 Naisbitt D J, Gordon S, Pirmohamed M, et al. Immunological principles of adverse drug reactions: the initiation and propagation of immune responses elicited by drug treatment. *Drug Saf*, 2000, 23: 483–507
- 3 Rauch A, Nolan D, Martin A, et al. Prospective genetic screening decreases the incidence of abacavir hypersensitivity reactions in the Western Australian HIV cohort study. *Clin Infect Dis*, 2006, 43: 99–102
- 4 Hetherington S, Hughes A R, Mosteller M, et al. Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet*, 2002, 359: 1121–1122
- 5 Mallal S, Nolan D, Witt C, et al. Association between presence of HLA-B\*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet*, 2002, 359: 727–732
- 6 Symonds W, Cutrell A, Edwards M, et al. Risk factor analysis of hypersensitivity reactions to abacavir. *Clin Ther*, 2002, 24: 565–573
- 7 Chessman D, Lethborg T, Kostenko L, et al. Abacavir hypersensitivity in HLA-B57-positive individuals with HIV infection is dependent upon the conventional MHC-I Ag presentation pathway. *Tissue Antigens*, 2007, 69: 373–373
- 8 Martin A M, Nolan D, Gaudieri S, et al. Predisposition to abacavir hypersensitivity conferred by HLA-B\*5701 and a haplotypic Hsp70-Hom variant. *Proc Natl Acad Sci USA*, 2004, 101: 4180–4185
- 9 Gerber B O, Pichler W J. Noncovalent interactions of drugs with immune receptors may mediate drug-induced hypersensitivity reactions. *AAPS J*, 2006, 8: 160–165
- 10 Posadas S, Pichler W. Delayed drug hypersensitivity reactions—new concepts. *Clin Exp Allergy*, 2007, 37: 989–999
- 11 Rozieres A, Vocanson M, Saïd B B, et al. Role of T cells in nonimmediate allergic drug reactions. *Curr Opin Allergy CI*, 2009, 9: 305–310
- 12 Chung W H, Hung S I, Chen Y T. Human leukocyte antigens and drug hypersensitivity. *Curr Opin Allergy CI*, 2007, 7: 317–323
- 13 Adam J, Eriksson K K, Schnyder B, et al. Avidity determines T-cell reactivity in abacavir hypersensitivity. *Eur J Immunol*, 2012, 1–11
- 14 Chessman D, Kostenko L, Lethborg T, et al. Human leukocyte antigen class I-restricted activation of CD8<sup>+</sup>T cells provides the immunogenetic basis of a systemic drug hypersensitivity. *Immunity*, 2008, 28: 822–832
- 15 Bharadwaj M, Illing P, Kostenko L. Personalized medicine for HLA-associated drug-hypersensitivity reactions. *Pers Med*, 2010, 7: 495–516
- 16 Vita R, Zarebski L, Greenbaum J A, et al. The immune epitope database 2.0. *Nucleic Acids Res*, 2010, 38: D854–D862
- 17 Sathiamurthy M, Peters B, Bui H H, et al. An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities. *Immunome Res*, 2005, 1: 1–10
- 18 Trybula W J. Data mining and knowledge discovery. *Annu Rev Inform Sci*, 1997, 32: 197–229
- 19 Mei H, Liao Z H, Zhou Y, et al. A new set of amino acid descriptors and its application in peptide QSARs. *J Pept Sci*, 2005, 80: 775–786

- 20 Sanchez A, David V. Advanced support vector machines and kernel methods. *Neurocomputing*, 2003, 55: 5–20
- 21 Schölkopf B, Smola A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Boston: The MIT Press, 2001
- 22 Tian F, Yang L, Lv F, et al. In silico quantitative prediction of peptides binding affinity to human MHC molecule: an intuitive quantitative structure–activity relationship approach. *Amino Acids*, 2009, 36: 535–554
- 23 Li Z L, Li G R, Shu M, et al. A novel vector of topological and structural information for amino acids and its QSAR applications for peptides and analogues. *Sci China Ser B Chem*, 2008, 51: 946–957
- 24 梅虎, 周原, 廖志华, 等. HLA-A\*0201 限制性 CTL 表位定量结构与活性研究. *化学学报*, 2006, 64: 949–952
- 25 Guan P, Doytchinova I A, Walshe V A, et al. Analysis of peptide-protein binding using amino acid descriptors: prediction and experimental verification for human histocompatibility complex HLA-A\* 0201. *J Med Chem*, 2005, 48: 7418–7425