Peptide binding specificities of HLA-B*5701 and B*5801

ZHANG YaLan[1], MEI Hu[1,2*], WANG Qing[2], XIE JiangAn[2], LV Juan[2],
PAN XianChao[2] & TAN Wen[2]

[1]*Key Laboratory of Biorheological Science and Technology (Ministry of Education), Chongqing University, Chongqing 400044, China;*
[2]*College of Bioengineering, Chongqing University, Chongqing 400044, China*

Recently, genome wide association studies showed that there is a strong association between abacavir-induced serious, idiosyncratic, adverse drug reactions (ADRs) and human leukocyte antigen-B*5701 (HLA-B*5701). Studies also found that abacavir-induced ADRs were seldom observed in patients carrying the HLA-B*5801 subtype. HLA-B*5801 of the same serotype (B17) as B*5701 differs by only 4 amino acids from B*5701. It is believed that because of these sequence differences, HLA-B*5801 cannot bind the specific peptides which are required for HLA-B*5701 to stimulate the T cell immune response. Thus, the difference in peptide binding profiles between HLA-B*5701 and B*5801 is an important clue for exploring the mechanisms of abacavir-induced ADRs. VHSE (principal component score vector of hydrophobic, steric, and electronic properties), a set of amino acid structural descriptors, was employed to establish QSAR models of peptide-binding affinities of HLA-B*5701 and B*5801. Optimal linear SVM (support vector machine) models with high predictive capabilities were obtained for both B*5701 and B*5801. The R^2 (coefficient of determination), Q^2 (cross-validated R^2), and R_{PRE}^2 (R^2 of test set) of two optimal models were 0.7530, 0.7037, 0.6153 (B*5701) and 0.6074, 0.5966, 0.5762 (B*5801), respectively. For B*5701 and B*5801, the mutations in positions 45 (MET-THR) and 46 (ALA-GLU) have little influence on the selection specificity of the P2 position of the bound peptide. However, the mutation in position 97 (VAL-ARG) greatly influences the selection specificity of the P7 position. HLA-B*5701 prefers the bulky and positively charged amino acids at the P7 position. In contrast, HLA-B*5801 prefers the non-polar hydrophobic amino acids at the P7 position while positively charged amino acids are unfavored.

human leukocyte antigen, B*5701, B*5801, SVM, P-I concept, VHSE

The rapid development of high-throughput and high-precision whole-genome sequencing technologies has led to the frequent use of genome-wide association studies (GWAS) to detect connections between serious, idiosyncratic, adverse drug reactions (ADRs) and human leucocytes antigens (HLAs) [1,2]. The association between abacavir hypersensitivity and HLA-B*5701 has been studied intensively. The clinical symptoms of abacavir hypersensitivity are fever, rash, nausea, vomiting, gastrointestinal symptoms, and lethargy or malaise, and can be life-threatening [3]. In 2002, Hetherington, Mallal *et al*. [4,5] proposed that abacavir hypersensitivity was strongly associated with HLA-B*5701, particularly among Caucasians [6]. Moreover, Chessman [7] observed that abacavir could activate abacavir-specific CD8+ T cells to produce IFNγ and TNFα and induce HLA-B*5701-restricted hypersensitivity. Thus abacavir hypersensitivity reactions may depend on the conventional MHC-I antigen presentation pathway.

In order to explain idiosyncratic adverse drug reactions, in 2006, Gerber [8,9] explored the P-I concept, or more

*Corresponding author (email: meihu@cqu.edu.cn)

generally, the pharmacologic interaction of drugs with immune receptors. The idea is that drugs can bind directly to the immune receptor molecules, namely the T cell receptor (TCR) and pHLA complex, by non-covalent interaction to elicit T-cell immune responses. The "P-I" concept was then proved by a series of experiments [10–13].

According to the "P-I concept", non-covalent interactions among a drug, TCR, and pHLA, which alter the interface properties between pHLA and TCR are a key step to elicit ADRs. HLA-B*5701 and B*5801 of the same HLA-B17 serotype have very similar peptide repertoires and differ in their amino acid sequences at only 4 positions [14]. However, research indicated that abacavir-specific $CD8^+$ T cells were only stimulated in the presence of HLA-B*5701. It is generally accepted that specific peptides involved in abacavir-induced ADRs cannot be selected by HLA-B*5801 for T cell recognition [11]. In 2010, Bharadwaj [15] successfully eluted an antigenic ligand from HLA-B*5701-positive patients treated with abacavir. This ligand was then proven to selectively stimulate abacavir specific T cells. Thus, the difference in peptide binding profiles between HLA-B*5701 and B*5801 is very important for exploring the mechanisms of abacavir-induced ADRs.

Herein, we describe the use of a support vector machine (SVM) and multiple step-wise regression (MSR) [16] to establish quantitative structural-activity relationship (QSAR) models for predicting the binding affinities of peptides bound to HLA-B*5701 and B*5801. The QSAR models can provide quantitative characterizations of the differences in peptide-binding profiles of HLA-B*5701 and B*5801. These results should prove important for exploring the mechanisms of HLA-B*5701-restricted ADRs as well as for designing epitope-based vaccines.

1   Principles and methods

1.1   Data processing

In October 2011, 1895 9-mer peptide ligands of HLA-B*5701 and 3069 9-mer peptide ligands of HLA-B*5801 were extracted from the Immune Epitope Database (IEDB) [17,18]. After removing duplicate samples and samples with missing values, 1155 and 1867 ligands were finally obtained for HLA-B*5701 and HLA-B*5801, respectively. The ligands with $EC_{50}>20000$ were considered to be non-binders. In order to decrease the modeling error brought about by non-binders, the number of binders and non-binders were kept the same in both training and test sets. A total of 99 binders and 99 non-binders of HLA-B*5701 were randomly selected from 198 binders and 957 non-binders to construct a training set. The remaining 99 binders together with 99 non-binders randomly selected were used to construct a test set. A total of 224 binders and 224 non-binders of HLA-B*5801 were randomly selected

from 448 binders and 1419 non-binders to build a training set. The remaining 224 binders and 224 non-binders randomly selected were used to build a test set. The binding affinity of peptide ligand was defined as the negative logarithm of 50% effective concentration ($-lgEC_{50}$).

1.2   Structural description of peptide ligand

VHSE, a set of amino acid descriptors, was derived from the research of Mei *et al*. [19]. Fifty physico-chemical properties of 20 coded amino acids, consisting of 18 hydrophobic properties, 17 steric properties, and 15 electronic properties, were used for principal component analysis (PCA). For the matrices of hydrophobic, steric, and electronic properties, the first 2, 2, and 4 principal components accounted for 74.33%, 78.68%, and 77.97% variance of the original data matrices, respectively. That is to say, the hydrophobic, steric, and electronic properties of 20 coded amino acids can be characterized by 8 principal component scores with minimal loss of information. These 8 score vectors are called the VHSE (principal components score vectors of hydrophobic, steric, and electronic properties) descriptors [19]. For the 20 amino acids, VHSE1 and VHSE2 are related to hydrophobic properties, VHSE3 and VHSE4 to steric properties, and VHSE5 to VHSE8 to electronic properties (Table 1).

For the 9-mer peptide ligand described in this paper, each sample can be characterized by 72 (9×8) VHSE descriptors denoted by V_{ij}, in which i indicates the position of the amino acid and j is the index of the VHSE descriptor. For example, V_{13} represents the VHSE3 descriptor of position 1.

1.3   MSR-SVM modeling

MSR combined by SVM was used to screen variables and construct QSAR models. MSR is an effective method to find an optimal subset in the original variable space, especially when the number of variables is not too large. SVM can avoid an over-fitting problem which often occurs in MSR modeling. Thus, MSR-SVM is a good choice for use in QSAR modeling for small datasets.

The basic principle of SVM [16,20,21] is to find an optimal hyperplane to separate two classes of data with a large margin between them. In a nonlinear case, the data are mapped into a high-dimensional feature space where a linear decision boundary is computed. SVM can be applied not only to classification problems but also to regression. SVM can fit a target function with a given accuracy ε ($\varepsilon \geqslant 0$), when the distances between data points and the optimal hyperplane are less than ε. The kernel function, i.e., $K(x, xi)=\Phi(x)\cdot\Phi(xi)$, is the core technology of SVM, and its use helps to avoid the problems of dimension disaster and decrease computational complexity.

Based on experience in predicting peptide binding affini-

ties [22–24], the radial basis function (RBF) and linear kernel function were used for SVM modeling. R^2 (coefficient of determination), Q^2 (10-fold cross-validated R^2), RMS (root of mean squared error), and R_{PRE}^2 (R^2 of a test set) were used to evaluate the resulting QSAR models. Herein, R_{PRE}^2 was defined as the R^2 of a regression line through the origin for the test set.

2   Results

2.1   MSR-SVM modeling

After screening by MSR, 4 and 5 optimal variable subsets were obtained for HLA-B*5701 and HLA-B*5801, respectively. SVM models established by using each variable subset are shown in Table 2 and Table 3. The modeling performances of the RBF kernel were better than that of the linear kernel, but there were no significant differences between them.

After considering the robustness, predictive powers, complexities and interpretabilities of the SVM models, model 5 with 14 VHSE variables (Table 2) and model 4 with 15 VHSE variables (Table 3) were selected as the optimal linear models of HLA-B*5701 and HLA-B*5801, respectively.

For HLA-B*5701 and HLA-B*5801, the binding affinities of the test sets were predicted by model 5 and model 4, respectively. From Figures 1 and 2, it can be seen that model 5 and model 4 are predictive for the external datasets. The R_{PRE}^2 and slopes of the two regression lines were 0.6153, 0.9818 (B*5701) and 0.5762, 0.9759 (B*5801), respectively.

Table 1   VHSE scales for 20 coded amino acids

AA		$VHSE_1$	$VHSE_2$	$VHSE_3$	$VHSE_4$	$VHSE_5$	$VHSE_6$	$VHSE_7$	$VHSE_8$
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

Table 2   The performance of the SVM models on the peptide binding affinities of HLA-B*5701

ID	No. of variables	Kernal	C	ε	γ	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

Table 3   The performance of the SVM models on the peptide binding affinities of HLA-B*5801

ID	No. of variables	Kernal	C	ε	γ	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
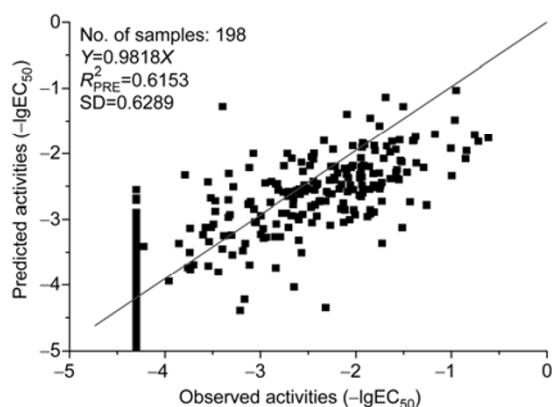
Figure 1   Observed vs. predicted peptide binding affinities of HLA-B*5701.
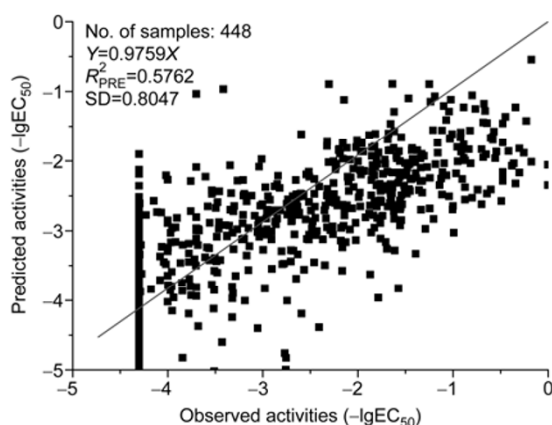
Figure 2   Observed vs. predicted peptide binding affinities of HLA-B*5801.

The optimized model 5 and model 4 were further validated by 10 repeated random samplings. By using the same variable subsets and parameters, 10 linear SVM models were obtained for each model. For HLA-B*5701 and HLA-B*5801, the average R^2, Q^2, R_{PRE}^2 of the 10 models were 0.6958, 0.6661, 0.6557 (B*5701) and 0.6027, 0.5926, 0.5796 (B*5801), respectively.

2.2   Peptide binding specificities of HLA-B*5701 and HLA-B*5801

HLA-B*5701 and B*5801 belong to the HLA-B17 serotype, and differ at 4 amino acid positions in the peptide-binding groove composed by the α1 and α2 domains (Figure 3). Residues 45 and 46 are located in the B pocket of the peptide-binding groove and can interact with the dominant anchor residue in position 2 (P2) of the 9-mer peptide. Residue 97 located between pocket C and E can interact with the secondary anchor residue in position 7 (P7) [25]. Therefore, mutations at these 3 positions play important roles in the binding specificities of the P2 and P7 positions. Residue 103 is not in the peptide binding groove, so it has no significant effect on peptide binding specificities.

The entrance of the B pocket of B*5701 is mainly composed of the bulky residues TYR7, ASN66 and GLU63, and the bottom primarily consists of the hydrophobic residues VAL24, GLY26, MET45 and ALA46 (Figure 4A). Compared to B*5701, position 45 of B*5801 is substituted by a hydrophilic THR and position 46 by a negatively charged GLU (Figure 4B). Thus, pocket B of B*5801 is more hydrophilic than that of B*5701. However, the negatively charged GLU46 located in the far end of the bottom pocket has little influence on the selection specificity of P2 of a bound peptide.

Residue 97 is located between the E and C pockets of the binding groove. For both B*5701 and B*5801, pocket E is mainly composed of VAL152, TRP147, LEU156, TRP133, ASP114 and SER116 (Figure 4C). It is important to note that the negatively charged ASP114 and SER116 located in the bottom of the E pocket have proved to be closely related to the abacavir-induced ADRs. Position 97 of B*5701 is occupied by a small hydrophobic VAL, whereas position 97 of B*5801 is occupied by a bulky and positively charged ARG. Thus, the E pocket of B*5701 is more negatively charged and is less sterically hindered than that of B*5801.

2.2.1   Peptide binding specificity of HLA-B*5701

The weight coefficients of the 14 VHSE variables included in model 5 are shown in Figure 5. These 14 VHSE variables
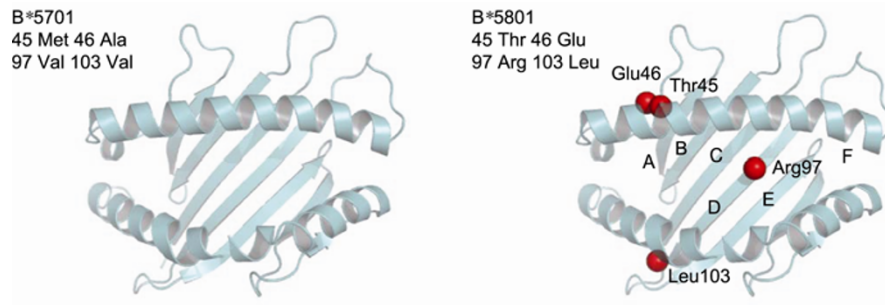
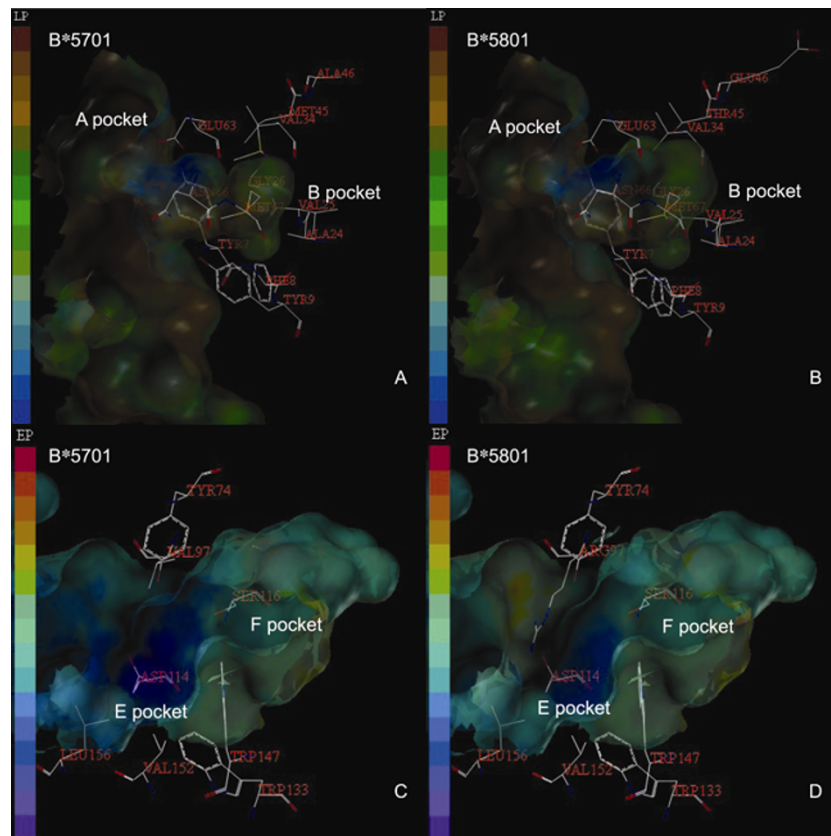**Figure 3**   Models of HLA-B*5701 and B*5801.

Figure 4   The B and E pockets of HLA-B*5701 and B*5801. A, Lipophilic potential properties of the B pocket of B*5701. B, Lipophilic potential properties of the B pocket of B*5801. C, Electrostatic potential properties of the E pocket of B*5701. D, Electrostatic potential properties of the E pocket of B*5801.

characterize the hydrophobic, steric, and electronic properties of P1, P2, P3, P4, P5, P7 and P9 positions. From the absolute values of the weight coefficients, it can be seen that P2 and P9 significantly contribute to the binding affinities of the bound peptides, followed by P1, P3 and P7, then P4 and P5. P6 and P8 do not contribute to the binding affinities.

Templated by an "AAAAAAAAA" sequence, a virtual 9-mer peptide library was constructed by a single point mutation at each position using each of the other 19 amino acids. For each position, a total of 19 mutated peptides together with the template were predicted by model 5. Then,

the predicted binding affinities of the 20 peptides were subjected to auto-scaled treatment. Thus, the scales at 9 amino acid positions (Table 4) can characterize the binding profile of B*5701.

In model 5, 4 VHSE variables of the P2 position represent steric (V_{23}, V_{24}) and electronic properties (V_{25}, V_{26}). According to the peptide binding profile shown in Table 4, small or polar amino acids like ALA, GLY, SER, THR and CYS are favored at this position, whereas TRP, TYR, GLU and ARG are disfavored.

Three VHSE variables of the P9 position characterize hydrophobic (V_{91}), steric (V_{93}) and electronic (V_{98}) proper-
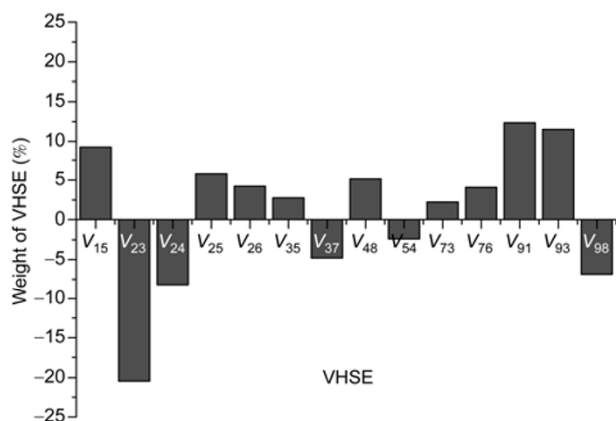
Figure 5 The weight coefficients of the VHSE variables of the HLA-B*5701 model.
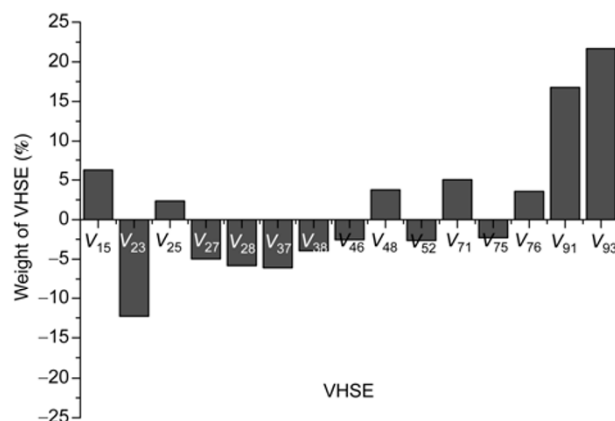

Figure 6 The weight coefficients of the VHSE variables of the HLA-B*5801 model.

ties. Table 4 indicates that bulky or hydrophobic amino acids like TRP, PHE, TYR, LEU and MET are favored at this position, while PRO and ASP are disfavored.

The profile of P7 is characterized by 2 VHSE variables representing steric (V_{71}) and electronic (V_{76}) properties respectively. According to Table 4, the bulky and positively charged ARG, HIS and LYS are desirable at P7, whereas small and hydrophobic GLY, ALA and PRO are undesirable.

Two VHSE variables of the P3 position represent electronic properties (V_{35}, V_{37}). The hydrophobic or bulky CYC, PHE, TRP and TYR are favored at this position, whereas GLY, GLU and ASP are disfavored. For positions P1 (V_{15}), P4 (V_{48}) and P5 (V_{54}), each has only 1 VHSE variable. Therefore, these positions have less influence on peptide binding.

*2.2.2   Peptide binding specificity of HLA-B*5801*

For HLA-B*5801, 15 VHSE variables included in model 4 are related to P1, P2, P3, P4, P5, P7 and P9 positions. From the weight coefficients of the VHSE variables (Figure 6), it can be concluded that P2 and P9 have the most influence on the binding affinity, followed by P7, P3, P1 and P4, then P5. P6 and P8 have no contributions to the binding affinity according to model 4.

Based on the standardized predicted values of the mutated peptides, the peptide binding profiles of B*5801 are shown in Table 5, and were used for the following analysis. The 4 VHSE variables of the P2 position represent steric (V_{23}, V_{25}) and electronic properties (V_{27}, V_{28}) respectively. From Table 5, it can be deduced that the small or neutral amino acids GLY, ALA, SER, CYS and THR are preferred, while the charged or bulky ARG and TRP are not preferred at this position. This profile is similar to that of B*5701, which only differs in the order of preference.

Two VHSE variables of P9 describe the hydrophobic (V_{91}) and steric (V_{93}) properties. According to the weight coefficients, the contributions of these 2 variables are larger

than that of B*5701. However, compared with B*5701, the electronic property (V_{98}) has no contribution to the peptide binding affinity. According to Table 5, the bulky or hydrophobic TRP, PHE and TYR are favored at P9, whereas GLY, SER and ALA are disfavored.

Three VHSE variables of P7 characterize the hydrophobic (V_{71}) and electronic (V_{75}, V_{76}) properties. As indicated in Table 5, the hydrophobic PHE, TRP and MET are preferred at this position. However, the electronic properties (V_{75}, V_{76}) of P7 should also be considered. For instance, a negatively charged ASP is favored at this position, while the positively charged ARG and LYS are disfavored. For B*5801, mutation at position 97 (VAL-ARG) alters the physio-chemical properties of the surface of pocket E, and thus changes the selection specificity at the P7 position. That is, bulky and positively charged residues at P7 are preferred for B*5701, while hydrophobic residues are preferred for B*5801.

P3 (V_{37}, V_{38}) and P4 (V_{46}, V_{48}) both involve 2 VHSE variables for characterization of their electronic properties. From Table 5, it can be deduced that TRP, CYC, MET and PHE are favored at P3, while PRO, ARG and LYS are disfavored. Also, PRO is preferred at P4 position, but ASN should be avoided.

P1 (V_{15}) and P5 (V_{52}) each involve one VHSE variable, representing the electronic and hydrophobic properties, respectively. It can be seen that LYS and ARG are favored at P1, while CYC and GLY are favored at P5 (Table 5).

Overall, there is a notable difference in the binding specificity of P7 between HLA-B*5701 and B*5801, and little differences in P1, P2, P3, P4, P5 and P9. Thus, it can be inferred that the amino acid preference at the P7 position of HLA-B*5701 is closely related to abacavir-induced ADRs.

3   Discussion

Genome-wide association studies indicate that there is a strong association between abacavir-induced ADRs and

Table 4 Peptide binding profile of HLA-B*5701

		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

Table 5 Peptide binding profile of HLA-B*5801

		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

HLA-B*5701. According to the P-I concept, non-covalent binding between abacavir and immune receptors is a key step to elicit ADRs. It is commonly speculated that specific peptides exclusively presented by HLA-B*5701 are involved in abacavir-induced ADRs. Therefore, comparison of peptide binding specificity between HLA-B*5701 and B*5801 is very important for exploring the mechanisms of abacavir-induced ADRs.

Based on a VHSE structural description method, QSAR models of peptide-binding affinities for HLA-B*5701 and B*5801 were established by the MSR-SVM method. Two optimal SVM models with strong predictive power were obtained. The results show that P2, P9, P1, P3 and P7 have significant influence on the binding affinities of both B*5701 and B*5801, followed by P4 and P5. A comparison of the peptide binding profiles of B*5701 and B*5801 indicated that mutations at positions 45 (MET-THR) and 46 (ALA-GLU) have little influence on the selection specificity of the P2 position. However, the mutation at position 97 (VAL-ARG) has significant influence on the selection spec-

ificity of P7. Also, for B*5701, bulky and positively charged residues are preferred at P7, while for B*5801 hydrophobic residues are preferred at this position. Thus, the amino acid preference at P7 of the bound peptides of HLA-B*5701 is inferred to be closely related to abacavir-induced ADRs. In conclusion, these results should prove to be a valuable reference for exploring the mechanisms of abacavir-induced ADRs.

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. Pnas, 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 Cl, 2009, 9: 305–310

12   Chung W H, Hung S I, Chen Y T. Human leukocyte antigens and drug hypersensitivity. Curr Opin Allergy Cl, 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[+] 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   Trybula W J. Data mining and knowledge discovery. Annu Rev Inform Sci, 1997, 32: 197–229

17   Vita R, Zarebski L, Greenbaum J A, *et al*. The immune epitope database 2.0. Nucleic Acids Res, 2010, 38: D854–D862

18   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

19   Mei H, Liao Z H, Zhou Y, *et al*. A new set of amino acid descriptors and its application in peptide QSARs. Biopolymers, 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   Mei H, Zhou Y, Liao Z H, *et al*. Study on quantitative structure-activity relationships of HLA-A* 0201 restrictive CTL epitopes. Acta Chim Sin, 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