

## HLA-A3 supermotif defined by quantitative structure–activity relationship analysis

Pingping Guan<sup>1</sup>, Irimi A.Doytchinova and Darren R.Flower

Edward Jenner Institute for Vaccine Research, Compton, Newbury, Berkshire RG20 7NN, UK

<sup>1</sup>To whom correspondence should be addressed.  
E-mail: pingping.guan@jenner.ac.uk

**Activation of a cytotoxic T cell requires specific binding of antigenic peptides to major histocompatibility complex (MHC) molecules. This paper reports a study of peptides binding to members of the HLA-A3 superfamily using a recently developed 2D-QSAR method, called the additive method. Four alleles with high phenotype frequency were included in the study: A\*0301, A\*1101, A\*3101 and A\*6801. The influence of each of the 20 amino acids at each position of the peptide on binding was studied. A refined A3 supertype motif was defined in the study.**

**Keywords:** A3-supertype/additive method/PLS

### Introduction

Cytotoxic T cell responses play a crucial role in mitigating viral infection and tumour formation (Madden, 1995; Ogasawara, 2001; Ryu *et al.*, 2001; Wei *et al.*, 2001). Antigenic peptides are produced by the proteolytic processing of endogenously synthesized proteins. These peptides bind to the complex formed by the major histocompatibility complex (MHC) molecule and  $\beta_2$ -microglobulin in the endoplasmic reticulum; the complex is then transported to the cell surface (Yewdell and Bennink, 1992; Reichstetter *et al.*, 1999; Wild *et al.*, 1999). Cytotoxic T lymphocytes recognize antigenic peptides presented by MHC molecules on the surface of antigen presenting cells initiating a concomitant cellular immune response (Bjorkman, 1990; Mannie, 2001).

A series of experiments have been carried out to study peptides binding to class I MHC molecules. Results from pool sequencing of naturally processed epitopes isolated from MHC molecules, peptide binding experiments and, most importantly, the X-ray crystallographic study of these molecules showed that MHC class I molecules bind short peptides, most of which are nonamers (Bjorkman *et al.*, 1987; Bjorkman, 1990; Falk *et al.*, 1991; Madden *et al.*, 1991; Ruppert *et al.*, 1994). Specific anchor residues at position 2 and at the C-terminus are of primary importance for binding (Bouvier and Wiley, 1994).

Epitope identification is the initial step in epitope based vaccine design and often begins with an initial *in silico* motif search. However, the process is complicated by MHC polymorphism. Several hundred HLA alleles have been found and the phenotype frequencies of the alleles vary, which makes it difficult to identify a motif that is effective in the whole population. Sette and Sidney grouped the class I alleles into superfamilies on the basis of the overlap between binding motifs (supermotifs) (Sette and Sidney, 1998). Four major

superfamilies have been identified: HLA-A2 (Altfeld *et al.*, 2001), HLA-A3 (Kawashima *et al.*, 1999), HLA-B7 (Coyle and Gutierrez-Ramos, 2001) and HLA-B44 (Sette and Sidney, 1998). If an epitope is identified that is reactive to a whole superfamily, then the rate of vaccine design process can be greatly accelerated.

In this work, we applied a recently developed 2D-QSAR technique, the additive method (Doytchinova *et al.*, 2002), to study the binding affinities of peptides to one of the four MHC class I superfamilies: the HLA-A3 superfamily. The additive method is based on the Free–Wilson concept (Kubinyi and Kehrhn, 1976), whereby each substituent makes an additive and independent contribution to the biological activity. Additional terms were added to the basic QSAR model to account for interactions between adjacent (1–2 interactions) and every second (1–3 interactions) side-chain. For a nonamer peptide the model is represented by

$$pIC_{50} = \text{constant} + \sum_{i=1}^9 P_i + \sum_{i=1}^8 P_i P_{i+1} + \sum_{i=1}^7 P_i P_{i+2} \quad (1)$$

where  $pIC_{50}$  is the binding affinity expressed in  $p$ -units (negative decimal logarithm of  $IC_{50}$  values), the constant

accounts for the peptide backbone contribution,  $\sum_{i=1}^9 P_i$  is the

sum of amino acid contributions at each position,  $\sum_{i=1}^8 P_i P_{i+1}$

is the sum of adjacent peptide side-chain interactions and

$\sum_{i=1}^7 P_i P_{i+2}$  is the sum of every second side-chain interactions.

The method was used to define a supermotif for peptides binding to the HLA-A2 supertype (Doytchinova and Flower, 2002). In this work, the method was used to study peptides bound to the HLA-A3 superfamily. The four most widespread A3-alleles (Sidney *et al.*, 1996a) were chosen: A\*0301, A\*1101, A\*3101 and A\*6801.

### Methods

#### Peptides

Information on peptide sequences and binding affinities was obtained from the JenPep database. JenPep (Blythe *et al.*, 2002) is a database containing data about the epitopes that bind to MHC alleles. The data were collected from previous publications (Kast *et al.*, 1994; Van de Burg *et al.*, 1995; Threlkeld *et al.*, 1997; Kawashima *et al.*, 1998; Wang *et al.*, 1998; Chang *et al.*, 1999; Scognamiglio *et al.*, 1999;) and are publicly available at the URL <http://www.jenner.ac.uk/JenPep>. Only nonamers were included in this study. The HLA-A\*0301 allele set included 72, the A\*1101 set included 62, the A\*6801 set included 38 and the A\*3101 set included 31 peptides. Among the selected peptides, some bound to more than one allele. On the other hand, some amino acids appeared at a position only once. The binding

Table I. Models generated by the additive method

Model	A*1101		A*0301		A*3101		A*6801	
	Amino acids	Amino acids and interactions	Amino acids	Amino acids and interactions	Amino acids	Amino acids and interactions	Amino acids	Amino acids and interactions
No. of peptides	62	62	72	70	30	31	38	37
$q^2_{\text{LOO}}^a$	0.458	0.428	0.436	0.305	0.482	0.453	0.531	0.370
$\text{NC}^b$	2	3	6	4	3	6	4	4
$\text{SEP}^c$	0.572	0.593	0.680	0.699	0.710	0.727	0.594	0.664
$q^2_{\text{CV5}}^d$	0.451	0.405	0.428	0.300	0.478	0.448	0.493	0.343
$\text{SEP}/\text{affinity range (\%)}^e$	15.7	16.2	19.5	20.1	21.3	21.8	13.2	14.8
$r^2$	0.829	0.977	0.959	0.972	0.892	0.990	0.959	0.974
$\text{SEE}^e$	0.321	0.119	0.181	0.136	0.325	0.098	0.175	0.136
$F$ ratio	143.005	821.098	246.895	557.374	71.356	399.955	194.845	297.481

<sup>a</sup> $q^2$  obtained after leave-one-out cross-validation.

<sup>b</sup>Optimal number of components.

<sup>c</sup>Standard error of prediction.

<sup>d</sup> $q^2$  obtained by cross-validation in five groups.

<sup>e</sup>Standard error of estimate.

affinity value ( $IC_{50}$ ) was used to quantify the interaction between the peptide and the MHC molecule. The  $IC_{50}$  values were measured by a competition assay based on the inhibition of the binding of the radiolabeled standard peptide to detergent-solubilized MHC molecules (Sidney *et al.*, 1996b).

#### Matrix construction

Equation 1 was written for each peptide in the study. The computer-generated matrix consisted of 6181 columns. The number of the rows was equal to the number of peptides in the study and the data for one peptide were contained in one row. The first column represented the dependent variable ( $pIC_{50}$ ). The next 180 represented the single amino acid contributions, 3200 represented the contributions of adjacent amino acid interactions (1–2 interactions) and 2800 columns were used for the calculation of the side chain interactions of amino acids at every second position (1–3 interactions). The existence of each amino acid and each interaction were recorded in the matrix. If present, the matrix recorded 1 in the corresponding column, otherwise the element would be 0. Any column containing only 0 was deleted during matrix construction.

#### Statistics

The partial least-squares (PLS) method implemented in SYBYL6.7 was used in the calculation (Sybyl, 1999).

PLS is an effective technique for relating the target property to its molecular structure. In mathematical terms, PLS relates a matrix of dependent variables  $Y$ , in this case the binding affinity data, to a matrix of descriptors  $X$  (Espinosa-Mansilla *et al.*, 2001). It is especially good when the number of properties is equal to or greater than the number of compounds. PLS is a variation of the principal component analysis (PCA)-based regression method. It finds families of variables that are correlated with the activity of the compounds and generates a set of latent variables, each of which is correlated to one of the family. Increasing the number of latent variables indicates a better explanation of the biological activity. PLS surpassed other statistical methods such as PCA in its ability to find such latent variables or factors, that both capture the greatest amount of variance and also are the best at correlating the  $X$  and  $Y$  blocks. In other words, PLS tries to find the maximum covariance between the variables that gave the model better predictive power.

The experimental  $IC_{50}$  values ( $pIC_{50}$ ) were used as the dependent variable in the study. Both the column filtering and the scaling

were turned off. The optimal number of components was found by running cross-validation using SAMPLS (Bush *et al.*, 1993).

#### Cross-validation

Cross-validation (CV) estimates the predictivity of the model (Wold *et al.*, 1995). In this technique the data were randomly divided into groups and the activities of the compounds in one group were predicted using the model generated by the rest of the data. The leave-one-out CV (LOO-CV) incorporated in SYBYL6.7 was used, in which each compound in the model was omitted once only. The following parameters were generated by the calculation and were used to assess the predictive ability of the models: the cross-validated coefficient  $q^2$  and the standard error of prediction  $SEP$ :

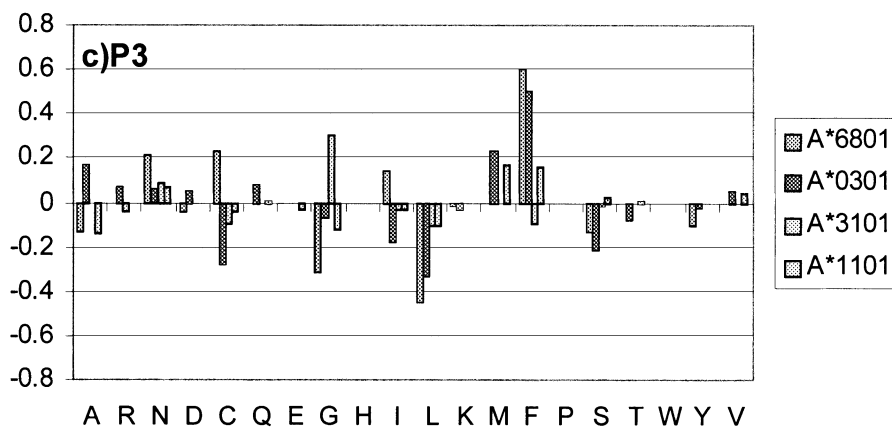
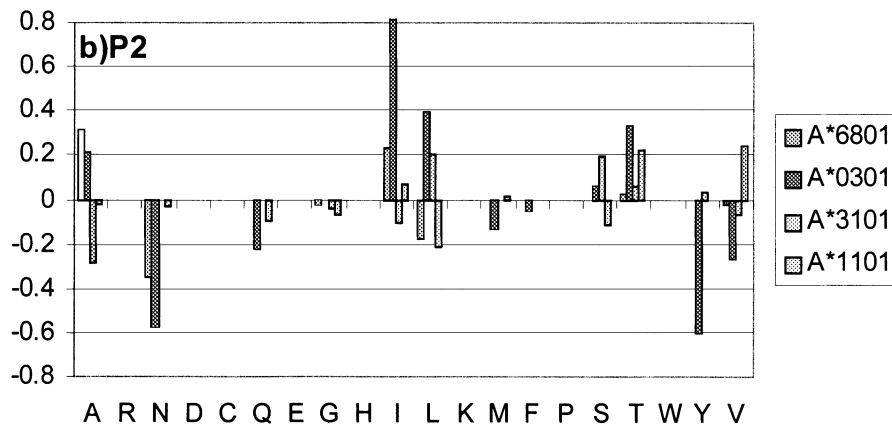
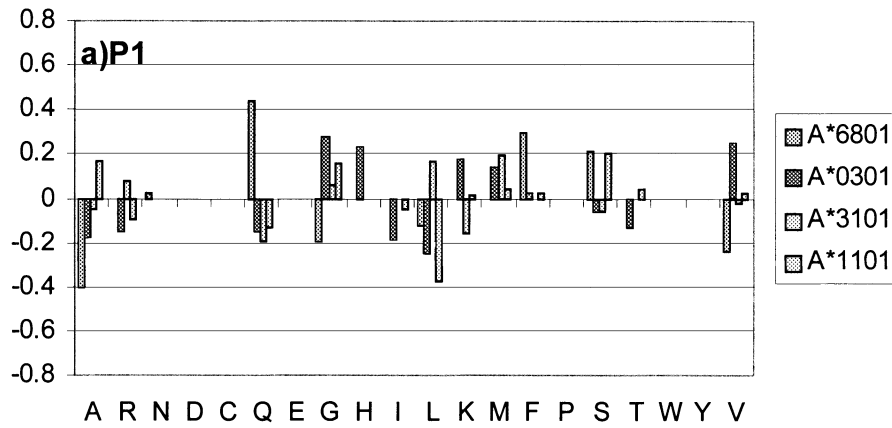
$$q^2 = 1 - \frac{\sum_{i=1}^n (pIC_{50 \text{ exp}} - pIC_{50 \text{ pred}})^2}{\sum_{i=1}^n (pIC_{50 \text{ exp}} - pIC_{50 \text{ mean}})^2}$$

$$SEP = \frac{\sqrt{\sum_{i=1}^n (pIC_{50 \text{ exp}} - pIC_{50 \text{ pred}})^2}}{p - 1}$$

where  $p$  represents the number of the peptides omitted,  $pIC_{50 \text{ pred}}$  and  $pIC_{50 \text{ exp}}$  are the values predicted by LOO-CV for the binding affinity and from the binding experiments, respectively. A more robust cross-validation test was also performed, in which the data were divided into five groups, a number of parallel models were developed from the reduced data with one of the groups omitted and the affinities of the excluded peptides were predicted. The mean of the  $q^2$  values from 20 runs is given as  $q^2_{\text{CV5}}$ . As the affinity range for the separate alleles was slightly different, the ratio of the  $SEP$  to affinity range was used as an additional assessment of the statistical validity of the models. Ratios close to 10% are indicative of good QSAR models. The non-cross-validated models were assessed by the explained variance  $r^2$ , standard error of estimate ( $SEE$ ) and  $F$  ratio.

#### Results

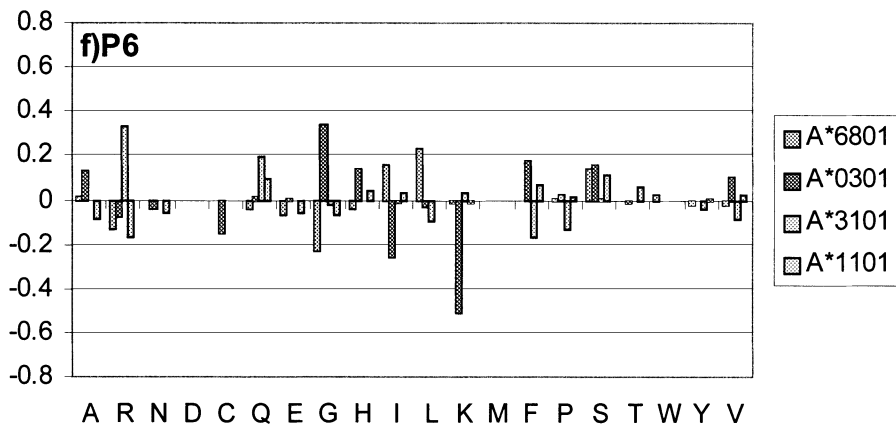
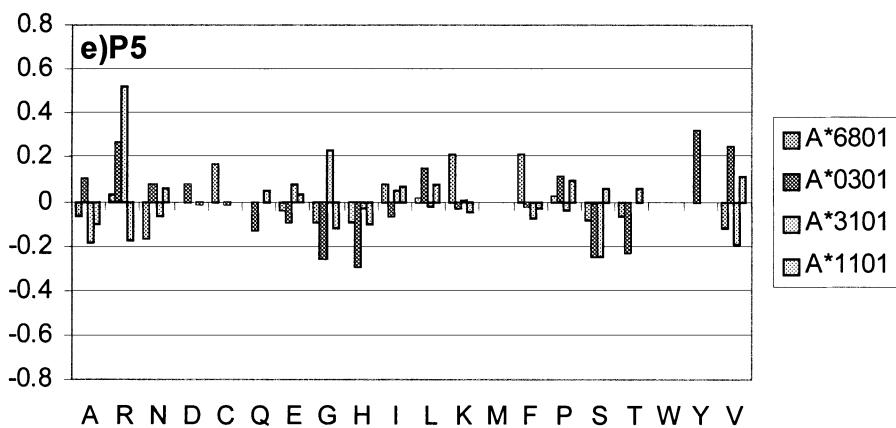
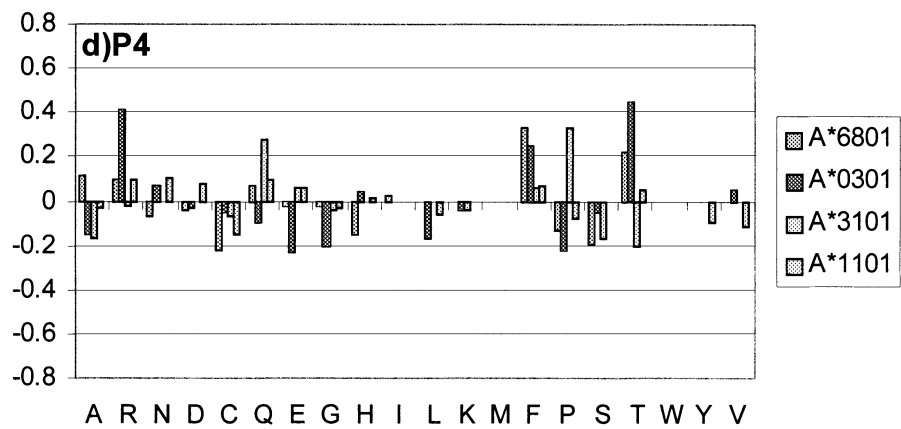
The additive models were generated for four alleles: A\*1101, A\*0301, A\*3101 and A\*6801. The statistical parameters are



given in Table I. Two types of models were created: one based solely on the amino acid contributions (amino acids models) and another based on both amino acid contributions and amino acid interactions (amino acids and interactions models). The number of peptides included in the two types differs for

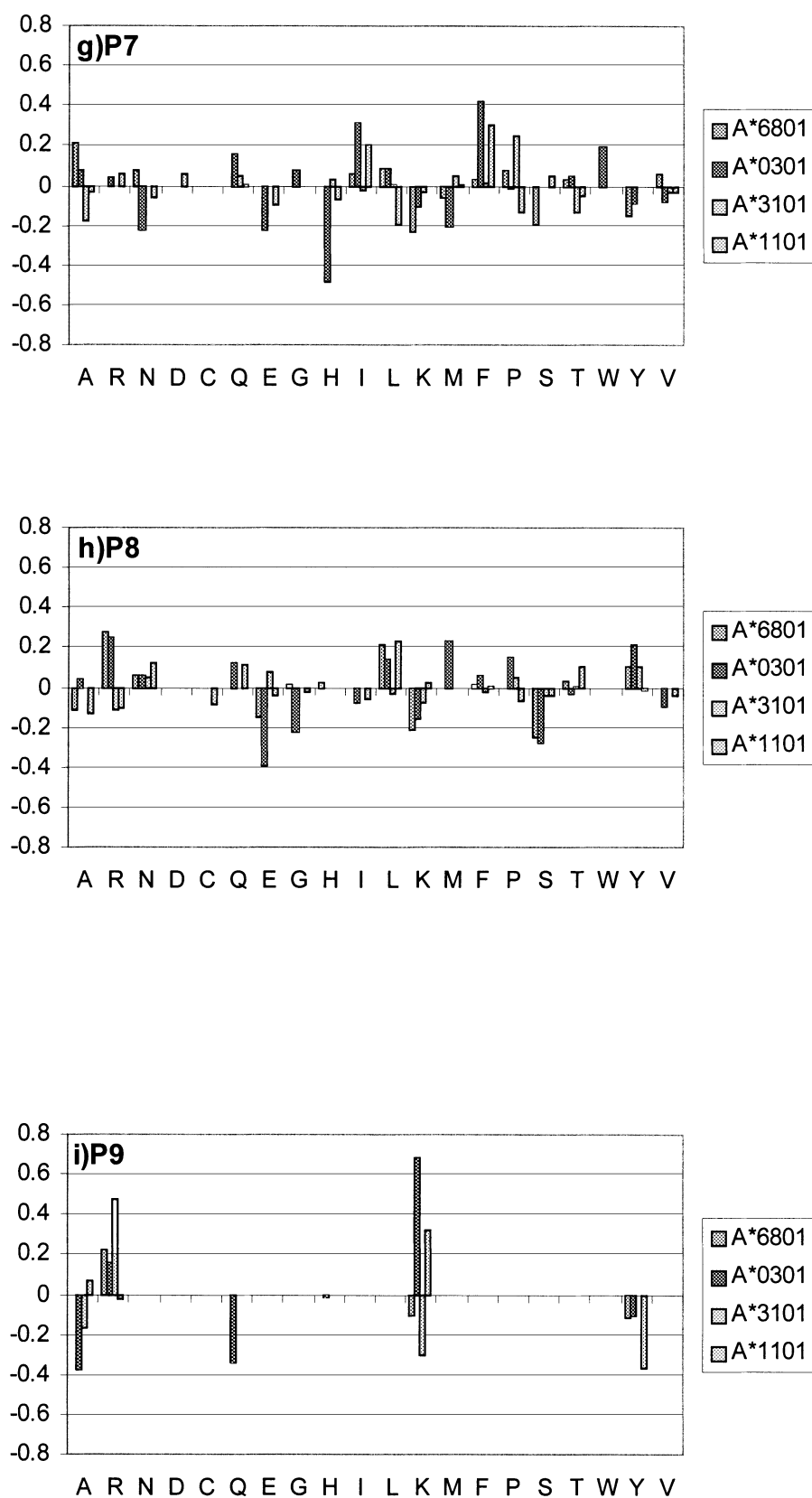
A\*0301 and A\*6801 alleles, because one or two peptides poorly predicted by LOO-CV were excluded.

The amino acids models are more predictive than the amino acids and interactions models, as is evident from the  $q^2$  values. The difference ranges from about 3% for A\*1101 and A\*3101



to 13% for A\*0301 and 16% for A\*6801. This is because some amino acid interactions occur only once in the model. Such interactions create a column in the matrix with only one value and many zeros. They appear as missing terms in the

equation used to predict the binding affinity of a peptide carrying such interactions. The prediction error is proportional to the number of missing terms. The number of missing terms in amino acids models is lower and because of this their



**Fig. 1.** The amino acid contributions at each position of the peptide. (a) Position 1, (b) position 2, (c) position 3, (d) position 4, (e) position 5, (f) position 6, (g) position 7, (h) position 8 and (i) position 9.

predictivity is higher. Increasing the number of peptides in the set under study will reduce the number of unique amino acids and amino acid interactions. The  $q^2_{CV5}$  values are close to the

corresponding  $q^2_{LOO}$  and this indicates the statistical stability of the models. The ratio  $SEP/affinity$  range is from 13 to 22%, further substantiating the statistical validity of the models.

**Table II.** Binding pockets of the A3 superfamily

	Pocket A										
	5	7	59	63	66	99	159	163	167	171	
A*1101	M	Y	Y	E	N	Y	Y	R	W	Y	
A*0301	-	-	-	-	-	-	-	T	-	-	
A*3101	-	-	-	-	-	-	-	T	-	-	
A*6801	-	-	-	N	-	-	-	T	-	-	
	Pocket B										
	7	9	24	34	45	63	66	67	70	99	
A*1101	Y	Y	A	V	M	E	N	V	Q	Y	
A*0301	-	F	-	-	-	-	-	-	-	-	
A*3101	-	T	-	-	-	-	-	-	H	-	
A*6801	-	Y	-	-	-	N	-	-	-	-	
	Pocket C										
	9	22	70	73	74	97	99	114	116		
A*1101	Y	F	Q	T	D	I	Y	R	D		
A*0301	F	-	-	-	-	-	-	-	-		
A*3101	T	-	H	I	-	M	-	Q	-		
A*6801	Y	-	-	-	-	M	-	-	-		
	Pocket D										
	99	114	155	156	159	160					
A*1101	Y	R	Q	Q	Y	L					
A*0301	-	-	-	L	-	-					
A*3101	-	Q	-	L	-	-					
A*6801	-	-	-	W	-	-					
	Pocket E										
	97	114	116	147	152	155	156				
A*1101	I	R	D	W	A	Q	Q				
A*0301	-	-	-	-	E	-	L				
A*3101	M	Q	-	-	V	-	L				
A*6801	M	-	-	-	V	-	W				
	Pocket F										
	73	77	80	81	84	95	116	118	123	124	143
A*1101	T	D	T	L	Y	I	D	Y	Y	I	T
A*0301	-	-	-	-	-	-	-	-	-	-	-
A*3101	I	-	-	-	-	-	-	-	-	-	-
A*6801	-	-	-	-	-	-	-	-	-	-	-

In contrast,  $r^2$  was found to be slightly lower for the single amino acid models than for the amino acids and interaction models. The decrease in  $r^2$  shows that the amino acid side-chain interactions are important for the variance explanation in the binding of the peptides to class I MHC molecules and should be considered in the modelling of the binding process.

Overall, the amino acid and interaction models were considered to have better explanatory power and were used to draw bar-charts for the amino acid contributions at each position, which are shown in Figure 1. Amino acids with contributions  $>0.2$  were considered as preferred for a particular allele at the specific position and those with contributions  $<-0.2$  were considered as deleterious. Residues identified as preferred for two or more A3-alleles without being deleterious for any other allele were considered as preferred for the A3 supermotif.

**Table III.** A3 superfamily binding motif defined by the additive method

Preferred	SM	IT	F	FRQ	-	S	FI	RLY	R
Position	1	2	3	4	5	6	7	8	9
Deleterious	ALQ	N	L	S	GHS	-	-	KSE	Y

Residues identified as deleterious for two or more alleles were considered as deleterious in the common motif.

#### Primary anchor positions

Positions 2 and 9 are generally accepted as the primary anchor positions for peptides bound to the A3 superfamily (Garrett *et al.*, 1989; Matsamura *et al.*, 1992; Falk *et al.*, 1994; Gavioli *et al.*, 1994). The side chain of the amino acid at position 2 falls into pocket B and the C-terminus is buried in pocket F of the MHC molecule (Madden *et al.*, 1991; Saper *et al.*, 1991; Vasmatzis *et al.*, 1996). The results from our study show a great variety of preferred amino acids for position 2 (Figure 1b). A\*6801 prefers Ala and Ile, A\*1101 prefers Thr and Val, A\*0301 prefers Ala, Ile, Leu and Thr and A\*3101 prefers Leu and Ser. This variety is explained by the polymorphism of the residues forming the bottom of pocket B, which surround position 9 of the MHC molecule (Table II). Phe9 in A\*0301 is substituted to Tyr9 in A\*6801 and A\*1101 and to Thr9 in A\*3101 (Schönbach *et al.*, 2000). The hydroxyl group of Tyr9 points towards the inside of the pocket and prevents larger amino acids from reaching the bottom of the pocket (Sudo *et al.*, 1995). Because of this, larger residues such as Leu are deleterious for A\*6801 and A\*1101 but are preferred for A\*0301. The change from Glu63 to Asn63 in A\*6801 and A\*1101 also changes the conformation of the pocket and stops large amino acids from binding (Vasmatzis *et al.*, 1996). Additionally, a previous study of pocket B revealed Val67 was reoriented in A\*6801 and affected amino acid selection (Guo *et al.*, 1993).

Compared with position 2, the amino acid preference at position 9 was uniform (Figure 1i). The side chain of the amino acid at position 9 of the peptide extends into pocket F, which is conserved among the HLA-A3 alleles. Positively charged amino acids such as Arg and Lys were preferred at this position. In members of the A3 family, two negatively charged Asp residues, Asp77 and Asp116, are present in pocket F (Table II). The residues provide a negatively charged environment to accommodate the positively charged amino acids. Falk and Röttschke postulated that Asp77 and Asp116, together with Asp residues at positions 74 and 77, were absolutely required in all the class I alleles that had an acidic C-terminus binding pattern (Falk and Röttschke, 1993). This ability could be abolished by the absence of any of the Asp residues at these positions. Tyr was deleterious in this position, mainly because of the bulky aromatic ring that could not fit into the narrow pocket.

The size of the side chains also affected peptide preference. A\*6801 and A\*3101 preferred Arg, A\*1101 favoured the smaller residue Lys, while A\*0301 accepted both. The difference in size seemed to be important in determining the binding affinity of the peptide. Zhang and co-workers found that the expression of A\*1101 was abolished when Lys at position 9 was changed to Arg, also a positively charged amino acid but significantly larger (Zhang *et al.*, 1993).

### Secondary anchor positions

The presence of the primary anchor residues alone does not induce stable peptide binding and several other positions are also crucial to successful binding (Zhang *et al.*, 1993). The amino acid at position 3 was considered important as it produced stable binding together with amino acids at position 2 and 9 in previous experiments, whereas amino acids at position 2 and 9 alone did not (DiBrino *et al.*, 1993). In the present study, position 3 preferred the hydrophobic residue Phe (Figure 1c). The side chain of the residue at position 3 extends into pocket D and contacts the aromatic side chains of the two conserved Tyr residues at positions 99 and 159. Previous peptide binding experiments by Sidney *et al.* gave a similar result (Sidney *et al.*, 1996b).

Another possible secondary anchor position is position 7 (Rammensee *et al.*, 1995). Hydrophobic residues were preferred at this position. Phe and Ile were strongly preferred by A\*0301 and A\*1101. Peptide binding studies showed either position 3 or 7, together with residues at positions 2 and 9, induced stable binding of the peptide (Sidney *et al.*, 1996b).

### Other positions

The study of the crystal structure of Aw68 (Silver *et al.*, 1992) suggested that positions 1, 4 and 8 pointed away from the peptide–MHC complex and towards the T cell receptor. In our study, Ser and Met were preferred at position 1, Phe, Arg and Tyr were favoured at position 4, Arg, Tyr and Leu were slightly favoured at position 8, while Ser, Lys and Glu were deleterious. The variance of amino acids accepted at these positions showed that they may not make significant contributions to the binding of the peptide to MHC molecules, yet these positions could be important in antigen recognition by T cells.

In the structure of Aw68, residues at positions 5 and 6 lie across the top of the binding groove and have contact with the T cell receptor. In the present study, no amino acid was preferred at these two positions. For deleterious residues, Ser was disfavoured at position 5. As with the discussion above, these positions were not particularly important in the binding of the peptide and they might participate in reactions with T cells.

### Discussion

This paper defined an epitope-binding motif for the A3 superfamily using the 2D-QSAR additive method. The superfamily classification is based on the peptide binding specificities of the alleles. Class I MHC alleles A\*1101, A\*0301, A\*3101 and A\*6801 bound to similar peptides. Sequence analysis showed that only 11 of the residues inside the binding pockets are polymorphic (Table I). A good, if incomplete, consensus was found in the preferences at the primary anchor positions 2 and 9. Thr and short hydrophobic residues such as Ala and Ile were favoured at P2 and nearly all the peptides bound to A3 alleles had positively charged residues Arg or Lys at the C-terminus.

The amino acids involved in peptide binding are similar in HLA-A2 and the A3 family. Pocket B interacts with the side chain of the residue at position 2, which was one of the anchor positions in nearly all the MHC class I alleles. Most of the amino acids in pocket B are conserved in the A2 and A3 families; both families accept hydrophobic residues. The amino acid at sequence position 9 of the MHC protein is important in peptide binding in the two families. Alleles with small to

medium sized residues (Phe9 or Thr9) were able to accept residues with long side chains such as Leu, such as A\*3101, A\*0301 and A\*0201. On the other hand, only small residues such as Ala and Val, could bind to A\*6801, A\*1101 and A\*0206, all of which had the larger residue Tyr9.

The five residues that directly interacted with the peptide in the F pocket are identical in both the A3 family and HLA-B27 (Leu81, Asp116, Tyr123, Thr143 and Trp147). Arg and Lys bound to pocket F and interacted with negatively charged residues Asp116 or Asp77 in both the A3 family and HLA-B27. B27 had been shown to accept hydrophobic residues such as Leu, Ala and Tyr because of their interaction with Leu81, Tyr123, Thr143 and Trp147 in the binding pocket (Jardetzky *et al.*, 1991). In the present study, the specificity at position 9 was restricted to Arg and Lys only; both Ala and Tyr had deleterious effects on peptide binding. This suggests a possible difference in the conformation of the binding pocket in spite of sequence similarity. Also, this may be the result of a change in conformation after the binding of other amino acids in the peptide.

A peptide binding motif for the HLA-A3 superfamily has been defined previously by Sidney *et al.* (Sidney *et al.*, 1996b) and Rammensee *et al.* (Rammensee *et al.*, 1995). Some useful similarities can be found on comparing the present motif with those defined by the above two groups. The amino acid preferences for the primary anchor residues are similar. All the motifs show preference for Arg and Lys at position 9 and have a preference for various hydrophobic residues at position 2, such as Ile and Thr. The preferences for secondary anchor residue positions 3 and 7 in the three motifs are hydrophobic amino acids such as Phe. The motif defined in the present paper, while in good agreement with previous motifs, is more extensive, covering all nine positions that contact the MHC molecule.

To conclude, in order to bind to members of the HLA-A3 superfamily, the peptide has to satisfy the requirements shown in Table III. The A3 superfamily good binder has to have a small to medium sized residue at position 2, such as Ile or Thr, and a positively charged residue Arg at position 9. Phe at either position 3 or 7 is also required for stable binding. Ser is well accepted at positions 1 and 6. Although positions 4 and 8 are more solvent-exposed than MHC-bound, they show also some well-defined preferences. Position 4 requires Phe, Arg and Gln and position 8 requires Arg, Leu and Tyr. In the present study, the additive method was shown to be an effective method for analysing peptide–MHC interactions. To make the results publicly accessible, all the models derived in the present study have been incorporated into a program for MHC-binding prediction and are freely available via the Internet at <http://www.jenner.ac.uk/MHCPred> (Flower *et al.*, 2002). We shall extend the application of our method to other MHC alleles in the future. Moreover, the additive method allows us to identify epitopes with high binding affinity that can be used to develop vaccines with high, non-ethnically biased population coverage.

### References

- Altfield, M.A. *et al.* (2001) *J. Virol.*, **75**, 1301–1311.
- Bjorkman, P.J. (1990) *Annu. Rev. Biochem.*, **59**, 253–288.
- Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D.C. (1987) *Nature*, **329**, 506–512.
- Blythe, M., Doytchinova, I.A. and Flower, D.R. (2002) *Bioinformatics*, **18**, 434–439.
- Bouvier, M. and Wiley, D.C. (1994) *Science*, **265**, 398–402.

- Bush,B.L. and Nachbar,R.B.,Jr. (1993) *J Comput.-Aided Mol. Des.*, **7**, 587–619.
- Chang,M.K., Gruener,N.H., Southwood,S., Sidney,J., Pape,G.R., Chisari,F.V. and Sette,A. (1999) *J. Immunol.*, **162**, 1156–1164.
- Coyle,A.J. and Gutierrez-Ramos,J.C. (2001) *Nature*, **363**, 203–209.
- DiBrino,M., Parker,K.C., Shiloach,J., Knierman,M., Jukszo,J., Turner,R.V., Biddkson,W.E. and Coligan,J.E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 1508–1512.
- Doytchinova,I.A. and Flower,D.R. (2002) *J. Comput. Biol.*, Submitted for publication.
- Doytchinova,I., Blythe,M.J. and Flower,D.R. (2002) *J. Proteome Res.*, **1**, 263–272.
- Espinosa-Mansilla,A., Acedo Valenzuela,M.I., Munoz de la Pena,A., Salinas,F. and Canada,C.F. (2001) *Anal. Chim. Acta*, **427**, 129–136.
- Falk,K. and Röttschke,O. (1993) *Semin. Immunol.*, **5**, 81–94.
- Falk,K., Röttschke,O., Stevanović,S., Jung,G. and Rammensee,H.G. (1991) *Nature*, **351**, 290–296.
- Falk,K., Röttschke,O., Takiguchi,M., Grahovac,B., Gnau,V., Stevanovic,S., Jung,G. and Rammensee,H.G. (1994) *Immunogenetics*, **40**, 238–241.
- Flower,D.R., Doytchinova,I.A., Paine,K., Taylor,P., Blythe,M.J., Lamponi,D., Zygouri,C., Guan,P., McSparron,H. and Kirkbride,H. (2002) In Flower,D.R. (ed.), *Drug Design: Cutting Edge Approaches*. Royal Society of Chemistry, Cambridge, pp. 136–180.
- Garrett,T.P.J., Saper,M.A., Bjorkman,P.J., Strominger,J.L. and Wiley,D.C. (1989) *Nature*, **342**, 692–696.
- Gavioli,R., Kurilla,M.G., De Campos-Lima,P.O., Wallace,L.E., Dolcetti,R., Murray,R.J., Rickinson,A.B. and Masucci,M.G. (1994) *J. Virol.*, **67**, 1572–1578.
- Guo,H.C., Madden,D.R., Silver,M.L., Jardetzky,T.S., Gorga,J.C., Strominger,J.L. and Wiley,D.C. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8053–8057.
- Jardetzky,T.S., Lane,W.S., Robinson,R.A., Madden,D.R. and Wiley,D.C. (1991) *Nature*, **353**, 326–329.
- Kast,W.M., Brandt,R.M.P., Sidney,J., Drijfhout,J., Kubo,R.T., Grey,H.M., Melief,C.J.M. and Sette,A. (1994) *J. Immunol.*, **152**, 3904–3912.
- Kawashima,I., Tsai,V., Southwood,S., Takesako,K., Celis,E. and Sette,A. (1998) *Int. J. Cancer*, **78**, 518–524.
- Kawashima,I., Tsai,V., Southwood,S., Takesako,K., Sette,A. and Celis,E. (1999) *Cancer Res.*, **59**, 431–435.
- Kubinyi,H. and Kehrhahn,O.H. (1976) *J. Med. Chem.*, **19**, 578–586.
- Madden,D.R. (1995) *Annu. Rev. Immunol.*, **13**, 587–622.
- Madden,D.R., Gorga,J.C., Strominger,J.L. and Wiley,D.C. (1991) *Nature*, **353**, 321–325.
- Mannie,M.D. (2001) *Immunol. Res.*, **23**, 1–21.
- Matsamura,M., Fremont,D.H., Peterson,P. and Wilson,I.A. (1992) *Science*, **257**, 927–934.
- Ogasawara,K. (2001) *Jpn. J. Clin. Pathol.*, **49**, 1225–1232.
- Rammensee,H.G., Friede,T. and Stevanovic,S. (1995) *Immunogenetics*, **41**, 178–228.
- Reichstetter,S., Kowk,W.W., Rochik,S., Koelle,D.M., Beaty,J.S. and Netpom,G.T. (1999) *Hum. Immunol.*, **60**, 608–618.
- Ruppert,J., Kubo,R.T., Sidney,J., Grey,H.M. and Sette,A. (1994) *Behring Inst. Mitt.*, No. 94, 48–60.
- Ryu,K.S., Lee,Y.S., Kim,B.K., Park,Y.G., Kim,Y.W., Hur,S.Y., Kim,T.E., Kim,I.K. and Kim,J.W. (2001) *Exp. Mol. Med.*, **33**, 136–144.
- Saper,M.A., Bjorkman,P.J. and Wiley,D.C. (1991) *J. Mol. Biol.*, **219**, 277–319.
- Schönbach, C., Koh,J.L.Y., Sheng,X., Wong,L. and Brusica,V. (2000) *Nucleic Acids Res.*, **28**, 222–224.
- Scognamiglio,P. et al. (1999) *J. Immunol.*, **162**, 6681–6689.
- Sette,A. and Sidney,J. (1998) *Curr. Opin. Immunol.*, **10**, 478–482.
- Sidney,J., Grey,H.M., Kubo,R.T. and Sette,A. (1996a) *Immunol. Today*, **17**, 261–266.
- Sidney,J., Grey,H.M., Southwood,S., Celis,E., Wentworth,P.A., del Guercol,M.F., Kubo,R.T., Chestnut,R.W. and Sette,A. (1996b) *Hum. Immunol.*, **45**, 79–93.
- Silver,M.L., Guo,H.C., Strominger,J.L. and Wiley,D.C. (1992) *Nature*, **360**, 367–369.
- Sudo,T., Kamikawaji,N., Kimura, A., Date,Y., Savoie,C.J., Nakashima,H., Furuichi,E., Kuhara,S. and Sasazuki,T. (1995) *J. Immunol.*, **155**, 4749–4756.
- SYBYL (2000) *SYBYL6.7, Ligand-based Design Manual*. Tripos, St. Louis.
- Threlkeld,S.C. et al. (1997) *J. Immunol.*, **159**, 1648–1657.
- Van de Burg,S.H., Ras,E., Drijfhout,J.W., Benckhuijsen,W.E., Bremers,A.J.A., Melief,C.J.M. and Kast,W.M. (1995) *Hum. Immunol.*, **44**, 189–198.
- Vasmatazis,G., Zhang,C., Cornette,J.L. and DeLisi,C. (1996) *Mol. Immun.*, **33**, 1231–1239.
- Wang,R., Johnston,S.L., Southwood,S., Sette,A. and Rosenberg,S.A. (1998) *J. Immunol.*, **160**, 890–897.
- Wei,W.Z., Ratner,S., Shibuya,T., Yoo,G. and Jani,A. (2001) *J. Immunol. Methods*, **258**, 141–150.
- Wild,M.K., Cambiaggi,A., Brown,M.H., Davies,E.A., Ohno,H., Saito,T. and van der Merwe,P.A. (1999) *J. Exp. Med.*, **190**, 31–41.
- Wold,S. (1995) In van de Waterbeemd,H. (ed.), *Chemometric Methods in Molecular Design*. VCH, Weinheim, pp. 195–218.
- Yewdell,J.W. and Bennink,J.R. (1992) *Adv. Immunol.*, **23**, 447–453.
- Zhang,Q.J., Gavioli,R., Klein,G. and Masucci,M.G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 2217–2221.

Received May 28, 2002; revised October 17, 2002; accepted November 12, 2002