

Transporter Associated with Antigen Processing Preselection of Peptides Binding to the MHC: A Bioinformatic Evaluation

Irini Doytchinova, Shelley Hemsley, and Darren R. Flower¹

TAP is responsible for the transit of peptides from the cytosol to the lumen of the endoplasmic reticulum. In an immunological context, this event is followed by the binding of peptides to MHC molecules before export to the cell surface and recognition by T cells. Because TAP transport precedes MHC binding, TAP preferences may make a significant contribution to epitope selection. To assess the impact of this preselection, we have developed a scoring function for TAP affinity prediction using the additive method, have used it to analyze and extend the TAP binding motif, and have evaluated how well this model acts as a preselection step in predicting MHC binding peptides. To distinguish between MHC alleles that are exclusively dependent on TAP and those exhibiting only a partial dependence on TAP, two sets of MHC binding peptides were examined: HLA-A*0201 was selected as a representative of partially TAP-dependent HLA alleles, and HLA-A*0301 represented fully TAP-dependent HLA alleles. TAP preselection has a greater impact on TAP-dependent alleles than on TAP-independent alleles. The reduction in the number of nonbinders varied from 10% (TAP-independent) to 33% (TAP-dependent), suggesting that TAP preselection is an important component in the successful *in silico* prediction of T cell epitopes. *The Journal of Immunology*, 2004, 173: 6813–6819.

Peptides that bind to class I MHC molecules are presented on the cell surface for inspection by CD8 T cells. Such peptides are derived from a variety of sources, including proteins of the host proteome and pathogenic microorganisms such as viruses and bacteria. Peptides, cleaved by proteasomes, are selectively transported by TAP from the cytosol into the lumen of the endoplasmic reticulum (ER)² where they are bound by MHC molecules before export to the cell surface and potential recognition by T cells (1). TAP is an ATP-dependent peptide transport protein that belongs to the ATP-binding cassette family of transporters. This family transports a wide range of molecules, from small sugars to large polypeptides, across membranes (2). There are two TAP proteins (TAP-1 and TAP-2), which form a transmembrane heterodimer. Both proteins encode one hydrophobic transmembrane domain and one ATP binding domain (3). Extant experimental studies have shown that TAP prefers peptides of eight or more amino acids with hydrophobic or basic residues at the C terminus (4, 5). TAP-mediated Ag presentation is important not only for cytosolic Ags, but also for most epitopes within membrane/secretory proteins (6). The TAP-dependent pathway is the principal processing route for peptides binding HLA-A1, HLA-A3, HLA-A11, HLA-A24, HLA-B15, and HLA-B27 (7–9). Some peptides are able to access the ER via other, TAP-independent mechanisms. Examples of alleles exhibiting only partial dependence on TAP include HLA-A2, HLA-A23, HLA-B7, and HLA-B8 (10–13).

The largest and most systematic screen for determining the sequence specificity of TAP was performed by Uebel et al. (14). A

peptide-binding motif for TAP has been defined by van Endert et al. (15). Hydrophobic aromatic amino acids are preferred at the C terminus, position 3 (p3) and position 7 (p7). Hydrophobic or positively charged residues are preferred at position 2 (p2), whereas aromatic or acidic residues are preferred at position 1 (p1). Proline at positions p1 and p2 has a strong negative effect on TAP binding. TAP preferences were confirmed by combinatorial peptide libraries (16), artificial neural networks studies (17), and matrix-based scoring functions (18, 19). The results of such studies are consistent with TAP contributing significantly to epitope selection.

Recently, we proposed a method for predicting the binding affinity of peptide-protein interactions, which we called the additive method (20). The additive method assumes that each amino acid makes an additive and constant contribution to the biological activity regardless of amino acid variation in the rest of the peptide, with possible interactions between amino acids being accounted for by cross-terms. The method was applied initially to predict the affinity of peptides binding to HLA-A*0201 molecule (20) and was then extended to 10 other human MHC class I (21, 22), three murine class I (our unpublished observations), three human class II (23), and six murine class II molecules (our unpublished observations). Most of these predictive models can be accessed via the MHCpred server (URL: <http://jenner.ac.uk/MHCpred>) (24). The additive method is universal, being equally applicable to any peptide-protein interaction. We have subsequently used this method in the cyclical optimization of high affinity peptides binding to HLA-A*0201, generating superbinders and anchorless epitopes (25). In the present study we used the additive method to develop a TAP binding prediction model, analyzing and extending the TAP binding motif. Importantly, we also evaluated how well this model acts as a preselection step in predicting MHC binding peptides. To distinguish between fully and partially TAP-dependent alleles, two datasets were examined. Peptides binding to HLA-A*0201 were selected as representatives of HLA alleles exhibiting partial TAP dependence, and peptides binding to HLA-A*0301 represented fully TAP-dependent HLA alleles.

Edward Jenner Institute for Vaccine Research, Compton, Berkshire, United Kingdom
Received for publication April 23, 2004. Accepted for publication May 30, 2004.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ Address correspondence and reprint requests to Dr. Darren R. Flower, Edward Jenner Institute for Vaccine Research, Compton, Berkshire, U.K. RG20 7NN. E-mail address: darren.flower@jenner.ac.uk

² Abbreviations used in this paper: ER, endoplasmic reticulum; PLS, partial least squares; ROC, receiver operating characteristic.

Materials and Methods

Peptide training and test sets

A set of 163 polyaniline nonameric peptides was used as a training set. Originally, using the peptide AAASAAAAY as the parent peptide, a set was prepared that included all natural amino acids except cysteine substituted at each position. This set was used to develop an artificial neural network-based prediction model (17). Because the binding affinities for the peptides were presented in the form of a histogram and used IC_{50} values relative to the reference peptide RRYNASTEL, we initially extracted these from the graph and then calculated raw, nonnormalized IC_{50} values, which we present as $-\log IC_{50}$ values (pIC_{50}). This set was used to develop an additive model for TAP binding.

Two sets of nonameric peptides were used to test the predictive ability of the additive model. Set A consisted of 47 analogues of the peptide ALAKAAAAY (data from Fig. 1 of Ref. 17). Originally, affinities were presented as IC_{50} values relative to the parent peptide ALAKAAAAY. Set B included 38 nonamers (data from Tables II and IV of Ref. 17), with affinities presented as IC_{50} values relative to the reference peptide RRYNASTEL.

To assess the TAP contribution to T cell epitope selection, the additive scoring function derived in this study was applied to a set of 317 A*0201 binders, 239 A*0201 nonbinders, 76 A*0301 binders, and 237 A*0301 nonbinders. Binders were extracted from our in-house database AntiJen, a development of JenPep (26, 27). Nonbinders were gifts from Dr. V. Brusic. Receiver operating characteristic (ROC) curves (28) were used to measure the prediction rate of binders vs nonbinders. Two variables, sensitivity (true binders/total binders) and $1 - \text{specificity}$ (false binders/total nonbinders), were calculated at different cutoffs. The area under the curve (A_{ROC}) is a single quantitative measure of the predictive ability and varies from 0.5 for random prediction to 1.0 for perfect prediction.

Additive method

The additive method (20) was applied to the training set of 163 polyaniline nonameric peptides. A matrix of 172 columns and 163 rows was generated. The number of columns corresponded to 171 independent variables (nine positions \times 19 amino acids) and one dependent variable (pIC_{50}). The number of rows equaled the number of peptides in the training set. Each peptide was represented by a binary string of 171 ones and naughts. A term is equal to 1 when a particular amino acid at a particular position is present and is equal to 0 when it is absent. The matrix was solved by partial least squares (PLS) linear regression, as implemented in SYBYL 6.7 (29). PLS is a so-called projection method (30) that can handle matrixes with more variables than observations and with noisy and highly collinear data. In this situation, conventional statistical methods, such as multiple regression, produce over-fitted models, i.e., models that fit the training data well, but are unreliable in prediction. PLS forms new variables, named principal components, as linear combinations of the initial variables and then uses them as predictors of the dependent variable, TAP binding affinity in our

case. The additive model was assessed using the correlation coefficient r^2 , and its predictive power was validated by the two external test sets and presented as r_{pred} , the correlation coefficient between the predicted and experimentally derived relative IC_{50} values.

Results

TAP additive model

The TAP additive model is shown in Table I. Two principal components explain 99.9% of the variance in the set. Positive and negative coefficients indicate amino acids that make positive and negative contributions to TAP binding affinity, respectively. The most positive coefficient belongs to Phe at p9, followed by Phe, Tyr, and Trp at p3. The most negative value corresponds to Ser at p9, followed by Pro at p2, and Asp and Gly at p9.

By using the additive method, a TAP binding motif was defined. Amino acids that increase TAP binding affinity >5 -fold (0.699 log unit) were identified as preferred; amino acids that decrease affinity >10 -fold (1 log unit) were identified as deleterious (Table I). No amino acid is strongly preferred at p1, but Glu, Asp, and Pro are deleterious. Trp has the highest positive contribution at p2, and Pro and Asp have the most negative contributions. Two groups of amino acids make significant contributions at p3; the first group includes Phe, Tyr, and Trp (each making an equal contribution of 1.125 log units), and the second group comprises Ile, Met, and Val (coefficients of 0.824). Only Asp and Gly are detrimental at p3. Low value positive and negative contributions are characteristic of p4. Trp is the preferred amino acid at p5, whereas Ile is preferred at p6; Pro and Tyr at p7; and Ser, Thr, and Gly at p8. There are no strongly disfavored amino acids at any of these four positions. P9 is very sensitive to changes; Phe and Tyr are favored, and there are many disfavored amino acids, including Ser, Asp, Gly, Asn, Thr, Glu, His, and Ala.

External validation

Two sets were used to test the predictive ability of the additive model. Both of them were taken from Ref. 17, but because the reference peptides used in the calculation of relative IC_{50} values were different, they are given as two separate test sets: A and B. Test set A included 47 peptides, and test set B comprised 38 peptides. The binding affinities were calculated by the additive model and are presented as the logarithm of the relative IC_{50} value ($IC_{50test}/IC_{50reference}$). The correlation between the predicted and

Table I. Additive QSAR model for peptides binding to TAP molecule^a

	p1	p2	p3	p4	p5	p6	p7	p8	p9
Ala	0.400	-0.030	-0.240	0.094	-0.173	-0.140	-0.097	-0.178	-0.808
Arg	0.487	0.558	0.347	0.492	0.290	0.527	0.190	-0.016	0.198
Asn	0.444	-0.533	-0.367	0.140	-0.351	0.448	-0.010	0.109	-1.406
Asp	-1.240	-1.074	-1.145	0.191	-0.062	0.022	-0.634	0.109	-1.809
Gln	-0.683	0.558	-0.237	0.316	0.151	0.381	-0.236	-0.174	-0.471
Glu	-1.349	0.035	-0.668	-0.082	-0.351	0.226	-0.053	-0.174	-1.114
Gly	-0.276	-0.791	-1.103	-0.268	-0.430	0.381	-0.685	0.711	-1.687
His	-0.522	-0.706	0.280	0.492	-0.038	0.184	0.093	-0.192	-0.950
Ile	-0.085	0.336	0.824	0.094	0.415	0.749	0.491	-0.091	-0.251
Leu	-0.351	0.637	0.347	0.094	0.591	0.022	0.424	0.146	-0.304
Lys	0.186	0.222	0.125	0.395	-0.086	0.381	-0.394	-0.259	-0.068
Met	-0.027	0.491	0.824	0.316	0.415	-0.189	0.491	0.146	-0.439
Phe	-0.648	0.190	1.125	0.492	0.348	-0.029	0.491	-0.091	1.174
Pro	-1.094	-1.945	0.083	0.395	0.591	0.050	0.792	0.234	-0.633
Ser	-0.073	0.433	0.347	-0.272	-0.086	-0.189	-0.685	0.887	-2.352
Thr	-0.243	0.222	0.011	0.094	0.017	-0.172	-0.146	0.711	-1.292
Trp	-0.419	0.859	1.125	0.492	0.892	-0.136	0.269	0.586	0.308
Tyr	-0.546	0.433	1.125	0.395	0.494	-0.293	0.792	0.074	0.762
Val	-0.012	0.558	0.824	0.249	-0.011	-0.117	0.366	0.146	-0.384
ASC	9.085	10.611	11.147	5.363	5.792	4.636	7.339	5.034	16.410

^a The constant is 6.223. Statistics of the model: $r^2 = 0.999$; principal components = 2. ASC, absolute sum of contributions.

the experimental $\log IC_{50\text{relative}}$ (r_{pred}) was used to assess model predictability. The correlation graphs for both sets are presented in Fig. 1. Set A has an r_{pred} of 0.717 (Fig. 1, *top panel*), and set B has

an r_{pred} of 0.832 (Fig. 1, *bottom panel*). The high predictive ability of the TAP additive model confirmed the applicability of the additive method for TAP binding affinity prediction.

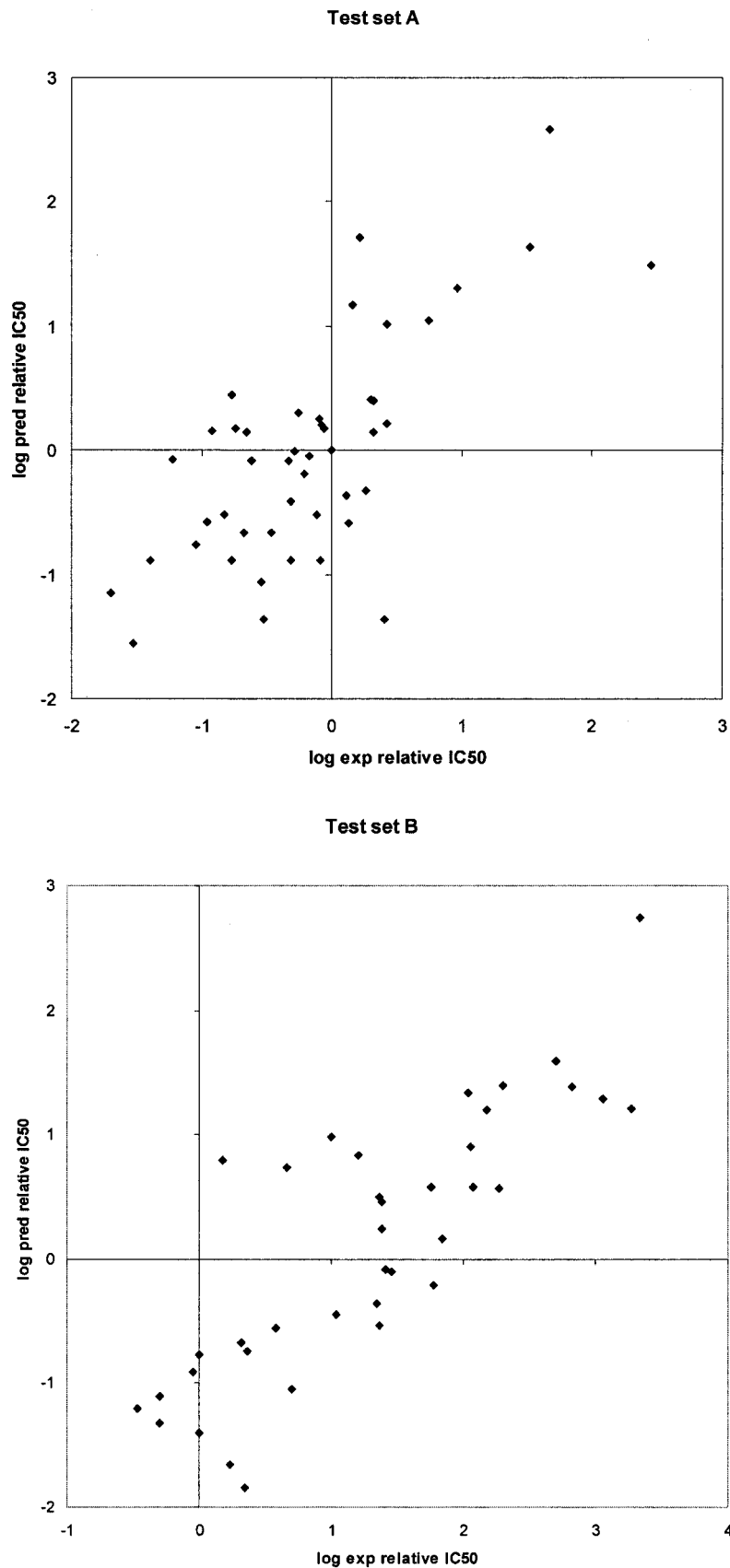


FIGURE 1. External validation of the TAP additive model. Test set A has $r_{pred} = 0.717$, and test set B has $r_{pred} = 0.832$.

Table II. Correlation coefficients between different quantitative matrices

Matrix	Peters et al. (18)	Bhasin and Raghava (19)
Bhasin and Raghava (19)	-0.683	
This study	-0.838	0.594

Comparison with other quantitative matrices

Peters et al. (18) generated a consensus scoring matrix for TAP binding affinity prediction. The matrix elements represent $\log(\text{IC}_{50})$ values for TAP binding that add up to the $\log(\text{IC}_{50})$ value of the peptide. A low matrix entry at a given position corresponds to an amino acid well suited for TAP binding. Bhasin and Raghava (19) developed a quantitative matrix using a support vector machine-based method to model the TAP binding affinity of peptides. These two quantitative matrices were compared with the matrix generated by the additive method in the present study. The correlation coefficients are given in Table II. Good correlation was found between Peters' and additive matrices ($r = -0.838$). The negative value of the correlation coefficient comes from the TAP affinity presentation; in Peters' model, affinity is presented as $\log(\text{IC}_{50})$, but in the additive model affinity it is given as $-\log(\text{IC}_{50})$. Correlation coefficients for positions 1, 2, 3, 7, and 9 were higher than 0.9. Moderate correlations were found between the matrix of Bhasin and Raghava and that of Peters and between Bhasin and Raghava's and additive matrices ($r = -0.683$ and $r =$

0.594, respectively). Again, the highest correlation coefficients were found for positions 1, 2, 3, and 9. The prediction ability of TAP quantitative matrices was similar for the anchor positions and differs only for the nonanchors.

TAP preselection of MHC binders

Many studies have suggested that the peptide selectivity exhibited by TAP binding may contribute to epitope selection (18, 31–33). To assess quantitatively the impact of this preselection for alleles that are strongly or weakly TAP dependent, two ROC analyses were performed. HLA-A*0201 was considered a representative of alleles that receive peptides via both TAP-dependent and TAP-independent mechanisms. This set included 317 A*0201 binders and 239 MHC nonbinders (Fig. 2A). For fully TAP-dependent alleles, HLA-A*0301 was selected. This set included 76 A*0301 binders and 237 MHC nonbinders (Fig. 2B). TAP and HLA (23) additive scoring functions were applied to predict binders and nonbinders. In both cases, HLA scoring functions gave better predictions than TAP scoring functions. However, TAP scores were better for fully TAP-dependent A*0301 than for partially TAP-dependent A*0201 ($A_{ROC} = 0.874$ vs $A_{ROC} = 0.721$). As is evident from the number of false and true negatives at different TAP cutoffs (Table III), a lower TAP cutoff ($-\log\text{IC}_{50}$, <3.00) is recommended for A*0201 peptides preselection than for A*0301 ($-\log\text{IC}_{50}$, <5.00). Increasing the TAP cutoff drastically increased the number of false negatives for A*0201, but did not affect the number of false negatives for A*0301. Thus, a TAP cutoff of 3.00 eliminated only 24 of the nonbinders (10%) for

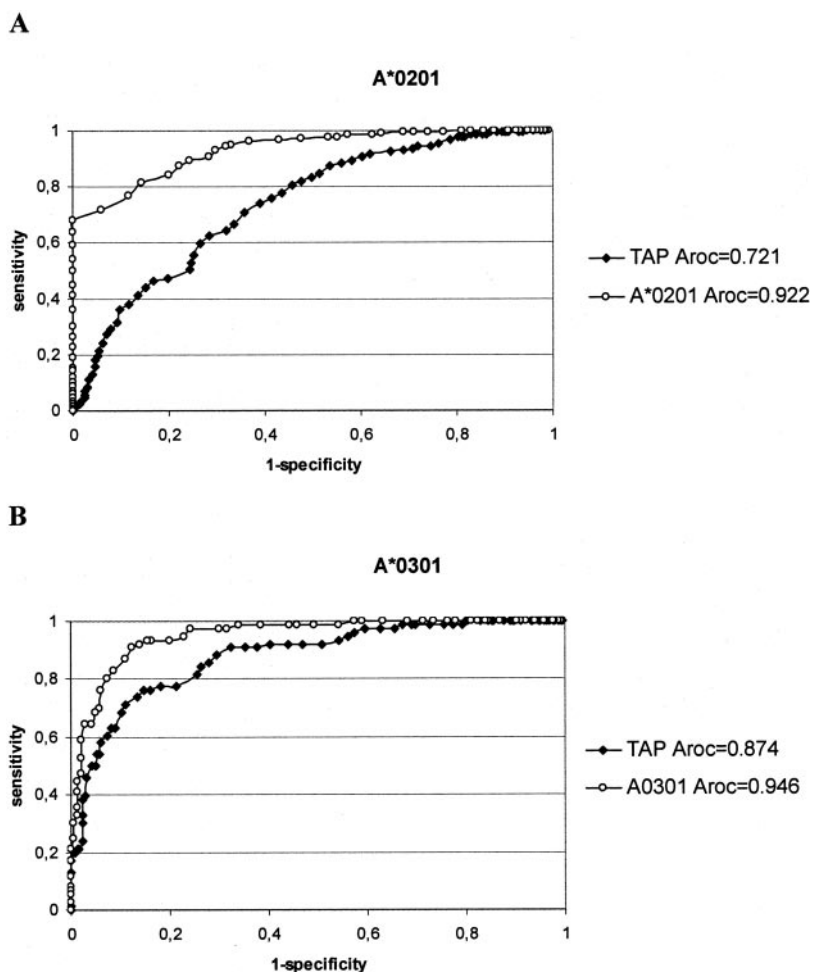


FIGURE 2. ROC curves of TAP, HLA-A*0201 (A) and HLA-A*0301 (B), additive scoring models. The sensitivity of the prediction (true binders/total binders) is plotted vs $1 - \text{specificity}$ (false binders/total nonbinders). Predictions of binders/nonbinders are made at different cutoffs using TAP additive models (●) and HLA additive models (○).

Table III. TAP preselection at different $-\log IC_{50}$ cutoffs

$-\log IC_{50}$ Cutoff	A*0201		A*0301	
	False negatives	True negatives	False negatives	True negatives
3.00	2	24	0	13
4.00	8	47	0	38
5.00	34	100	1	80

A*0201, whereas at a TAP cutoff of 5.00, 80 A*0301 nonbinders (33%) were eliminated. Unsurprisingly, TAP preselection was more efficient for fully TAP-dependent alleles than for partially TAP-dependent alleles.

Positions 1, 2, 3, and 9 were the most significant for binding to TAP (18, 31, 32). To test their contributions to MHC binding prediction, models considering the amino acid contributions at these positions were developed and applied to A*0201 binder/non-binder test set (Fig. 3). The model, which considered only positions 1, 2, 3, and 9 (TAP_1239 model), gave predictions close to the TAP model considering all positions (TAP_all model). A model based on positions 2 and 9 performed better (TAP_29 model; $A_{ROC} = 0.703$). The ability of the last model to preselect A*0201 binders was impressive (*upper right* of the ROC curve), predicting 66 (28%) true and only two false nonbinders at lower cutoffs. At the same time, it gave a high number of false binders (*lower left* of the ROC curve), which reduced its overall accuracy. Models considering positions 1 and 9 (TAP_19 model; $A_{ROC} = 0.563$), positions 3 and 9 (TAP_39 model; $A_{ROC} = 0.598$), and position 9 only (TAP_9 model; $A_{ROC} = 0.479$) exhibited significantly inferior performance (data not shown).

Discussion

The additive method was applied to a set of 163 nonameric peptides that bound the human TAP heterodimer. The binding affinity of a peptide is the sum of the quantitative contributions of each amino acid at each of the nine positions. The predictive ability of the derived model was assessed using two test sets, comprising a total of 85 peptides. A good correlation between predicted and experimental IC_{50} values was found for both sets. Different TAP additive scoring functions were applied to a set of 317 A*0201 binders and 239 nonbinders and to a set of 76 A*0301 binders and 237 nonbinders. TAP preselection had a higher impact on fully TAP-dependent alleles than on partially TAP-dependent ones. The

reduction in the number of nonbinders varied from 10% (TAP-independent) to 33% (TAP-dependent). Comparing the preselection ability of different TAP additive scoring models, the model based on all positions (TAP_all model) performed best.

Because TAP transport precedes HLA binding, a conflict will only arise between positions that are deleterious for TAP binding but preferred for HLA binding, not between TAP-preferred and HLA deleterious positions. The absolute sum of contributions (Table I) indicates that positions 1, 2, 3, and 9 exhibit the greatest variation in amino acid preference. It is widely assumed that positions 2 and 9 are the primary anchors, and positions 1 and 3 are secondary anchors for MHC binding. Supermotifs for certain major MHC class I supertypes are shown in Table IV, together with those amino acids that are preferred for TAP at the same positions. Most HLA alleles prefer peptides with hydrophobic or aromatic amino acids at their C termini; only the A3 binding motif has positively charged amino acids (Arg or Lys) in this study. Phe, Tyr, and Trp are the preferred amino acids at the C terminus of TAP binding peptides, whereas Arg makes a small positive contribution, and Lys makes a negligible contribution (Table I). Ile, Leu, and Val exhibit moderate negative values (<0.4 log units). Ser, Asp, Gly, Asn, Thr, and Glu are all detrimental for TAP binding, and this provides a possible explanation of why few human class I MHC ligands have these amino acids at their C termini.

There is a great variety of preferred amino acids at anchor position 2 in HLA motifs (Table I). A2 and A3 supertypes prefer hydrophobic amino acids, A24 prefers aromatic, B7 prefers Pro, B27 prefers positively charged amino acids, and B44 prefers negatively charged ones. All these amino acids make positive contributions to TAP binding (Table I), except for Pro and Asp. At position 2, Pro is a preferred anchor for B7, whereas Asp is preferred for B44. These are the only points of conflict between TAP and HLA binding preferences. The deleterious effects of Pro and Asp suggest that ligands with Pro and Asp at position 2 are unlikely to be transported into the lumen of the ER via a TAP-dependent mechanism. Fortunately, these ligands often bear Phe and Tyr at their C termini, which are strongly preferred by TAP, indicating a potential compensating effect for Pro and Asp.

Position 1 is the next most sensitive position for TAP binding after position 9. It is thought to be a secondary anchor for MHC binding, and the side chain occupies pocket A (34). However, the TAP_19 model ($A_{ROC} = 0.563$) suggests that this position is not overly important for TAP transport. Additionally, the highest negatively contributing amino acids for TAP affinity,

FIGURE 3. ROC curves of different TAP models for HLA-A*0201. The sensitivity of the prediction (true binders/total binders) is plotted vs $1 - \text{specificity}$ (false binders/total nonbinders). Predictions of binders/nonbinders are made at different cutoffs using the TAP_all model (●; all positions included), the TAP_1239 model (X; only positions 1, 2, 3, and 9 included), and the TAP_29 model (○; only positions 2 and 9 included).

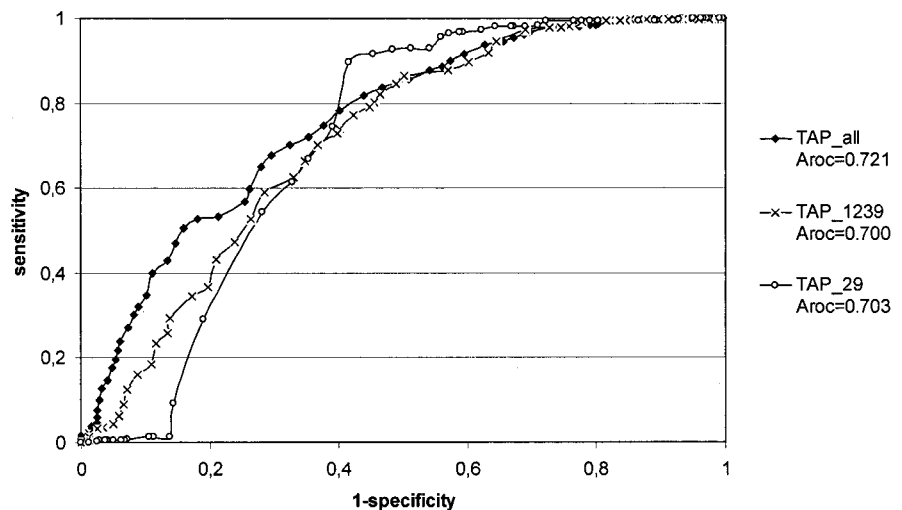


Table IV. HLA supermotifs (TAP motif included for comparison)

Supertypes	Position 2	C-Terminal
A2	Hydrophobic	Hydrophobic
A3	Hydrophobic	Positively charged
A24	Aromatic	Hydrophobic/aromatic
B7	Pro	Aromatic/hydrophobic
B27	Positively charged	Aromatic/hydrophobic
B44	Negatively charged	Aromatic/hydrophobic
TAP	Hydrophobic/positively charged	Aromatic/positively charged

Glu and Asp, are common at position 1 in many HLA ligands (35).

Phe, Tyr, and Trp at position 3 have the highest positive contribution to TAP binding after Phe⁹, whereas Asp and Gly contribute negatively. The side chain at position 3 occupies pocket D in the MHC binding groove, and it is thought to be an important secondary anchor (36). A wide range of amino acids, including Asp and Gly, are available at this position in different MHC ligands, which point to the moderate importance of this position for TAP transport.

Weak amino acid contributions to TAP binding were seen at positions 4, 5, 6, 7, and 8 (Table I). Similar results have been reported by others (18, 31, 32). The primary interaction of TCR is with residues 5–8 of a class I MHC binding nonapeptide (36). Thus, Ag recognition by a TCR is in the region of the peptide where TAP exerts minimal selection. Moreover, TAP transport is only one part of the complexity inherent in the emerging picture of class I presentation (6, 37).

MHC class I presentation has long been assumed to be, broadly speaking, a linear process beginning with the ubiquitin-mediated transport of proteins to the proteasome, a multimeric protease responsible for digestion of most cytosolic protein (38); this step is followed by binding to TAP. However, this simple picture is becoming more complicated by the day. Peptides are unfolded and cleaved by cytosolic proteases other than the proteasome. Tripeptidyl peptidase II is the only properly characterized such proteolytic enzyme (39), but may be one of many more. Peptides generated by the proteasome or tripeptidyl peptidase II may be degraded by cytosolic proteases such as leucine aminopeptidase (40) and thimet oligoendopeptidase (41). Peptides entering the ER pool are trimmed by ER-associated aminopeptidase (42) and other proteases, such as furin (43), leukocyte-derived arginine aminopeptidase-1 (44), or puromycin-resistant aminopeptidase (45), before binding to MHCs. The process can be decomposed into a set of peptide cleavage and peptide binding steps, each of which will require modeling using bioinformatic techniques, such as our additive method, which can predict the peptide specificity of each stage (where peptides are cleaved) and associated on- and off-rates (how rapidly peptide cleavage occurs). Because the TAP additive model considers the last nine positions before the C-terminal, it is applicable to any peptide extended at the N terminus. N termini extensions do not affect the TAP binding affinity prediction, which is a particular advantage of quantitative methods. However, because of the inherent complexity of immunological presentation, one cannot account for its dynamic behavior solely using bioinformatic methods to predict individual steps. Rather, we will need to supplement them with well-understood mathematical models, similar to those used in metabolic control theory. These methods, which can account for substrate fluxes within complex multicomponent metabolic pathways, allow ready incorporation of quantitative aspects of bioinformatic models and may help us better un-

derstand why particular peptide Ags become immunodominant (46). In particular, it is clear that peptide transport into the ER proceeds via both TAP-mediated and TAP-independent pathways. Because there are no clear sequence signals differentiating peptides transported via these several mechanisms, and the molecular components of TAP-independent pathways are not yet understood, it will require both experimentation to identify and characterize other transporters as well as in silico informatics techniques to produce robust methods able to reliably predict the complexities of Ag presentation.

Our current quantitative TAP model is our first attempt to address this wider challenge: the modeling of the complete class I presentation pathway. Overall, TAP selection is neither overly precise nor restrictive, allowing a wide variety of peptides to be transported into the ER. Using TAP preselection in binding affinity prediction methods could reduce the number of nonbinders by one-sixth to one-third.

Acknowledgments

We thank Dr. Vladimir Brusich for providing sets of MHC nonbinders.

References

1. Germain, R. N. 1994. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell* 76:287.
2. Monaco, J., S. Cho, and M. Attaya. 1990. Transport protein genes in the murine MHC: possible implications for antigen processing. *Science* 250:1723.
3. Meyer, T. H., P. M. van Endert, S. Uebel, B. Ehring, and R. Tampé. 1994. Functional expression and purification of the ABC transporter complex associated with antigen processing (TAP) in insect cells. *FEBS Lett.* 351:443.
4. Müller, K. M., C. Ebensperger, and R. Tampé. 1994. Nucleotide binding to the hydrophilic C-terminal domain of the transporter associated with antigen processing (TAP). *J. Biol. Chem.* 269:14032.
5. Schumacher, T. N. M., D. V. Kantesaria, M.-T. Heemels, P. G. Ashton-Rickardt, J. C. Shephard, K. Früh, Y. Yang, P. A. Peterson, S. Tonegawa, and H. L. Ploegh. 1994. Peptide length and sequence specificity of the mouse TAP1/TAP2 translocator. *J. Exp. Med.* 179:533.
6. Lautscham, G., A. Rickinson, and N. Blake. 2003. TAP-independent antigen presentation on MHC class I molecules: lessons from Epstein-Barr virus. *Microbios Infect.* 5:291.
7. Brusich, V., P. van Endert, J. Zeleznikow, S. Daniel, J. Hammer, and N. Petrovsky. 1998. A neural network model approach to the study of human TAP transporter. *Silico Biol.* 1:10.
8. de la Salle, H., E. Houssaint, M. A. Peyrat, D. Arnold, J. Salamero, D. Pinczon, S. Stevanovic, H. Bausinger, D. Fricker, E. Gomard, et al. 1997. Human peptide transporter deficiency: importance of HLA-B in the presentation of TAP-independent EBV antigens. *J. Immunol.* 158:4555.
9. Mormung, F., J. J. Neefjes, and G. J. Hämmerling. 1994. Peptide selection by MHC-encoded TAP transporters. *Curr. Opin. Immunol.* 6:32.
10. Henderson, R. A., H. Michel, K. Sakaguchi, J. Shabanowitz, E. Appella, D. F. Hunt, and V. H. Engelhard. 1992. HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. *Science* 255:1264.
11. Guéguen, M., W. Biddison, and E. O. Long. 1994. T cell recognition of an HLA-A2-restricted epitope derived from a cleaved signal sequence. *J. Exp. Med.* 180:1989.
12. Smith, K. D., and C. T. Lutz. 1996. Peptide-dependent expression of HLA-B7 on antigen processing-deficient T2 cells. *J. Immunol.* 156:3755.
13. Khanna, R., S. R. Burrows, D. J. Moss, and S. L. Silins. 1996. Peptide transporter (TAP-1 and TAP-2)-independent endogenous processing of Epstein-Barr virus (EBV) latent membrane protein 2A: implications for cytotoxic T-lymphocyte control of EBV-associated malignancies. *J. Virol.* 70:5357.
14. Uebel, S., W. Kraas, S. Kienle, K.-H. Wiesmüller, G. Jung, and R. Tampé. 1997. Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. *Proc. Natl. Acad. Sci. USA* 94:8976.
15. van Endert, P. M., D. Riganelli, G. Greco, K. Fleischhauer, J. Sidney, A. Sette, and J.-F. Bach. 1995. The peptide-binding motif for the human transporter associated with antigen processing. *J. Exp. Med.* 182:1883.
16. Uebel, S., T. H. Meyer, W. Kraas, S. Kienle, G. Jung, K.-H. Wiesmüller, and R. Tampé. 1995. Requirements for peptide binding to the human transporter associated with antigen processing revealed by peptide scans and complex peptide libraries. *J. Biol. Chem.* 270:18512.
17. Daniel, S., V. Brusich, S. Caillat-Zucman, N. Petrovsky, L. Harrison, D. Riganelli, F. Sinigaglia, F. Gallazzi, J. Hammer, and P. M. van Endert. 1998. Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules. *J. Immunol.* 161:617.
18. Peters, B., S. Bulik, R. Tampé, P. M. van Endert, and H.-G. Holzhütter. 2003. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J. Immunol.* 171:1741.
19. Bhasin, M., and G. P. S. Raghava. 2004. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci.* 13:596.

20. Doytchinova, I., M. J. Blythe, and D. R. Flower. 2002. Additive method for the prediction of protein-peptide binding affinity: application to the MHC class I molecule HLA-A*0201. *J. Proteome Res.* 1:263.
21. Guan, P., I. A. Doytchinova, and D. R. Flower. 2003. HLA-A3 supermotif defined by quantitative structure-activity relationship analysis. *Protein Eng.* 16:11.
22. Doytchinova, I., and D. Flower. 2003. The HLA-A2 supermotif: a QSAR definition. *Org. Biomol. Chem.* 1:2648.
23. Doytchinova, I. A., and D. R. Flower. 2003. Towards the in silico identification of class II restricted T-cell epitopes: a partial least squares iterative self-consistent algorithm for affinity prediction. *Bioinformatics* 19:2263.
24. Guan, P., I. A. Doytchinova, C. Zygouri, and D. R. Flower. 2003. MHCpred: a server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Res.* 31:3621.
25. Doytchinova, I. A., V. Walshe, N. Jones, S. Gloster, P. Borrow, and D. R. Flower. 2004. Coupling in silico and in vitro analysis of peptide-MHC binding: a bioinformatics approach enabling prediction of superbinding peptides and anchorless epitopes. *J. Immunol.* 172:7495.
26. Blythe, M. J., I. A. Doytchinova, and D. R. Flower. 2002. JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics* 18:434.
27. McSparron, H., M. J. Blythe, C. Zygouri, I. A. Doytchinova, and D. R. Flower. 2003. JenPep: a novel computational information resource for immunobiology and vaccinology. *J. Chem. Inf. Comput. Sci.* 43:1276.
28. Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30:1145.
29. Tripos. 2004. SYBYL 6.7. Tripos, St. Louis.
30. Eriksson, L., E. Johnansson, N. Kettaneh-Wold, and S. Wold. 2001. Basic concepts and principles of projections. In *Multi- and Megavariate Data Analysis. Principles and Applications*. Umetrics Academy, Umea, pp. 21–41.
31. Gubler, B., S. Daniel, E. A. Armandola, J. Hammer, S. Caillat-Zucman, and P. M. van Endert. 1998. Substrate selection by transporters associated with antigen processing occurs during peptide binding to TAP. *Mol. Immunol.* 35:427.
32. Lankat-Buttgereit, B., and R. Tampé. 1999. The transporter associated with antigen processing TAP: structure and function. *FEBS Lett.* 464:108.
33. Fruci, D., G. Lauvau, L. Saveanu, M. Amicosante, R. H. Butler, A. Polack, F. Ginhoux, F. Lemonnier, H. Firat, and P. M. van Endert. 2003. Quantifying recruitment of cytosolic peptides for HLA class I presentation: impact of TAP transport. *J. Immunol.* 170:2977.
34. Ruppert, J., J. Sidney, E. Celis, R. T. Cubo, H. M. Grey, and A. Sette. 1993. Prominent role of secondary anchor residues in peptide binding to HLA-A*0201 molecules. *Cell* 74:929.
35. Rammensee, H.-G., T. Friede, and S. Stevanović. 1995. MHC ligands and peptide motifs: first listing. *Immunogenetics* 41:178.
36. Garboczi, D. N., P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. 1996. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384:134.
37. Chen, L., and M. Jondal. 2004. Alternative processing for MHC class I presentation by immature and CpG-activated dendritic cells. *Eur. J. Immunol.* 34:952.
38. Kruger, E., U. Kuckelkorn, A. Sijts, and M. Kloetzel. 2003. The components of the proteasome system and their role in MHC class I antigen processing. *Rev. Physiol. Biochem. Pharmacol.* 148:81.
39. Seifert, U., C. Maranon, A. Shmueli, J. F. Desoutter, L. Wesoloski, K. Janek, P. Henklein, S. Diescher, M. Andrieu, H. de la Salle, et al. 2003. An essential role for tripeptidyl peptidase in the generation of an MHC class I epitope. *Nat. Immunol.* 4:375.
40. Saveanu, L., D. Fruci, and P. van Endert. 2002. Beyond the proteasome: trimming, degradation and generation of MHC class I ligands by auxiliary proteases. *Mol. Immunol.* 39:203.
41. Kim, S. I., A. C. Pabon, T. A. Swanson, and M. J. Glucksman. 2003. Regulation of cell surface major histocompatibility complex class I expression by the endopeptidase EP3.4.24.15 (thimet oligopeptidase). *Biochem. J.* 375:111.
42. Serwold, T., F. Gonzalez, J. Kim, R. Jacob, and N. Shastri. 2002. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* 419:480.
43. Lu, J., P. J. Wettstein, Y. Higashimoto, E. Appella, and E. Celis. 2001. TAP-independent presentation of CTL epitopes by Trojan antigens. *J. Immunol.* 166:7063.
44. Tanioka, T., A. Hattori, S. Masuda, Y. Nomura, H. Nakayama, S. Mizutani, and M. Tsujimoto. 2003. Human leukocyte-derived arginine aminopeptidase: The third member of the oxytocinase subfamily of aminopeptidases. *J. Biol. Chem.* 278:32275.
45. Levy, F., L. Burri, S. Morel, A. L. Peitrequin, N. Levy, A. Bachi, U. Hellman, B. J. van den Eynde, and C. Servis. 2002. The final N-terminal trimming of a subaminoterminal proline-containing HLA class I-restricted antigenic peptide in the cytosol is mediated by two peptidases. *J. Immunol.* 169:4161.
46. Flower, D. R. 2003. Towards in silico prediction of immunogenic epitopes. *Trends Immunol.* 24:667.