# Understanding Prediction Systems for HLA-Binding Peptides and T-Cell Epitope Identification

Liwen You[1,2,3], Ping Zhang[3], Mikael Bodén[4], and Vladimir Brusic[3,5]

[1] School of Information Science, Computer and Electrical Engineering,
Halmstad University, Halmstad, Sweden
[2] Department of Theoretical Physics, Lund University, Lund, Sweden
[3] School of Land, Crop, and Food Sciences, and
[4] School of Information Technology and Electrical Engineering,
University of Queensland, Brisbane QLD, Australia
[5] Cancer Vaccine Center, Dana-Farber Cancer Institute, Boston MA, USA
liwen@thep.lu.se, p.zhang2@uq.edu.au, m.boden@uq.edu.au,
vladimir_brusic@dfci.harvard.edu

**Abstract.** Peptide binding to HLA molecules is a critical step in induction and regulation of T-cell mediated immune responses. Because of combinatorial complexity of immune responses, systematic studies require combination of computational methods and experimentation. Most of available computational predictions are based on discriminating binders from non-binders based on use of suitable prediction thresholds. We compared four state-of-the-art binding affinity prediction models and found that nonlinear models show better performance than linear models. A comprehensive analysis of HLA binders (A*0101, A*0201, A*0301, A*1101, A*2402, B*0702, B*0801 and B*1501) showed that non-linear predictors predict peptide binding affinity with high accuracy. The analysis of known T-cell epitopes of survivin and known HIV T-cell epitopes showed lack of correlation between binding affinity and immunogenicity of HLA-presented peptides. T-cell epitopes, therefore, can not be directly determined from binding affinities by simple selection of the highest affinity binders.

## 1 Introduction

Major histocompatibility complex (MHC) molecules present peptides, derived from antigens and host proteins, on the cell surface. The recognition of presented peptides by TCD8[+] cells is necessary for recognition of infected or pathologically mutated cells and induction of cellular immune responses and subsequent elimination of tumors and infected cells. Human MHC is known as the human leukocyte antigen (HLA). Antigen processing and presentation involves primarily three steps: proteasomal cleavage, translocation of cleaved fragments by transporter associated with antigen processing, and HLA-peptide binding. The HLA/peptide binding is by far the most discriminative step: natural prevalence of HLA-binding peptides is in the range of 0.1-5% for any given protein

of which some 20% remain functionally relevant [1,2]. Therefore, computational prediction and modeling of HLA/peptide binding can greatly facilitate peptide screening, with tremendous savings in time and experimental effort.

The HLA peptide binding prediction can be approached as simple classification problem of discriminating binders from non-binders. However, peptide binding is necessary but it does not guarantee an immune response. A binding affinity metric like inhibition concentration ($IC_{50}$) of a standard probe quantifies HLA/peptide binding. Given a large number of binding data recently made available [3] we have extended the approach by studying peptide binding as a regression problem.

Many different attempts have been made to predict MHC peptide binding. There are primarily three approaches: by structure modelling, data-driven using peptide sequences and their binding affinities, or by the combination thereof. Sequence-based approaches can be further categorized into motif/profile based methods [4,5,6,7,8,9] and machine learning methods which use Artificial Neural Networks (ANN) [10,11], Hidden Markov Models (HMM) [12], or Support Vector Machines (SVM) [13,14,15,16,17,18,19]. An example of combined method is the adaptive double threading [20]. The prediction methods have been compared for accuracy of classification (binders vs. non-binders) [3,21,22]. However different data sets are used to build models and evaluation data vary between different studies making it intrinsically difficult to compare predictor performance.

Recently, a comprehensive experimental relative binding affinity analysis of a complete overlapping peptide library derived from the tumor-associated antigen survivin [23] was reported for eight different types of HLA class I molecules. Also, a large data set of peptide binding affinities became available at the Immune Epitope Database and Analysis Resource (IEDB, www.immuneepitope.org) [3]. Combining these two data sets, we analysed the factors that affect accuracy of prediction models and explored the correlation of peptide binding and immunogenicity. The results of this study provide an improved understanding of the prediction systems design issues and their use for identification of HLA-binding peptides and T-cell epitopes.

## 2   Materials and Methods

Four different prediction methods were used in this study. Support Vector Regression (SVR) and epitope information were used to build a regression model to predict HLA-peptide binding affinities using the first of the data sets. The three models used at IEDB, namely ANN and two matrix methods were used as comparison predictors. We defined prediction performance criteria to enable fair comparisons between models. Specifically, we used match curve area and correlation coefficient to compare model performance.

### 2.1   Datasets

Survivin, a member of the apoptosis inhibitor protein family, is one of a limited number of shared tumor-associated antigens that is over-expressed in the

majority of human cancers. There is an intense interest in using survivin as a target for therapeutic CTL response. Bachinsky et al. [23] used a high-throughput technique to identify peptides derived from survivin that bind eight HLA class I alleles: HLA-A*0101, HLA-A*0201, HLA-A*0301, HLA-A*2402, HLA-A*1101, HLA-B*0702, HLA-B*0801, HLA-B*1501. A library of 134 overlapping nonamers spanning the full length of the survivin protein (UniProt O15392 with 142 amino acids) was experimentally screened for peptides capable of binding each allele. Binding to each allele was reported as a percentage relative to a positive control peptide for that allele as values from 0 to >100%. An arbitrary cutoff of 30% of the control was used as a positive cutoff for experimental binders. Therewith, they identified nineteen HLA-A*0201, zero HLA-A*0101, seven HLA-A*0301, twelve HLA-A*1101, twenty-four HLA-A*2402, six HLA-B*0702, six HLA-B*0801 and eight HLA-B*1501 binding peptides.

Friedrichs et al. [24] collected a set of survivin-derived peptides, which can induce HLA restricted CTL responses. Two peptides reported as survivin-derived nonamer T-cell epitopes are HLA-A*0201 restricted $_{96}$LTLGEFLKL$_{104}$ and A* 2402 restricted $_{20}$STFKNWPFL$_{28}$.

Another set of proteins that has been comprehensively studied in T-cell responses are HIV proteins. We analyzed all HIV protein T-cell epitopes available in the HIV molecular immunology database (www.hiv.lanl.gov/content /immunology). In addition, we analysed mutations of a small set of HLA-restricted CD8$^+$ T-cell epitopes.

Peters et al. [3] have made public a set of 48,828 quantitative peptide-binding affinity measurements relating to 48 different mammalian MHC class I alleles. They used this data to establish a set of predictions with one neural network method (IEDB ANN) and two matrix-based prediction methods (IEDB SMM and IEDB ARB) and compared them with other available online predictors. In this study, we only used eight nonamer datasets of the eight HLA alleles of interest in this study. The data set (which we denote as the IEDB data set) was downloaded from (mhcbindingpredictions.immuneepitope.org).The datasets used in this study were: A*0101 (1157 peptides), A*0201 (3089), A*0301 (2094), A*1101 (1985), A*2402 (197), B*0702 (1262), B*0801 (708), and B*1501 (978).

## 2.2   SVM Regression Model and Peptide Coding Using Extra Epitope Information

The SVM is firmly based on statistical learning theory. It can be used to solve both classification and regression problems by optimizing given generalization bounds. Its regression form (SVR) is based on a loss function that ignores errors within a certain distance of the true value (we use the $\varepsilon$-insensitive loss function). In SVMs data is implicitly projected into a high-dimensional feature space using a kernel function. We employed the Gaussian kernel, $K(x, z) = \exp(-||x - z||^2/\sigma^2)$, where $x$ and $z$ are two samples and $\sigma$ is a kernel parameter. The Gaussian kernel requires peptides to be represented as numerical vectors. A sparse orthogonal coding was used to represent peptides, with each amino acid encoded by 20 bits (19 bits set to zero and 1 bit set to one). Hence, a nonamer is

represented in a 180-dimensional space. The coding vector was extended by nine positions encoding the shape of the binding motif for each studied allele. The final coded peptide vector had 189 elements and we refer to it as the extended sparse coding.

### 2.3   IEDB Prediction Models

IEDB ANN, SMM and ARB prediction models are used by IEDB website and all three methods predict the quantitative binding affinity. The ANN is a nonlinear model and the other two generate scoring matrices. They have been used as benchmarking predictions [3] and we compared the SVR model performance with them.

### 2.4   Performance Evaluation Methods

To compare two classifiers discriminating binders vs. non-binders, area under ROC (Receiver operating characteristic curve; the AUC value) compares overall performance of classifiers and does not require a decision threshold to be determined. For regression, the correlation coefficient between predicted and true binding affinities were used. To assess potential epitopes/binders along protein sequences, we used "match curve" plots of the number of true binders in the top N ranked predicted affinities (y-axis) vs. N (x-axis). If most of true epitopes can be found within a short list of top ranked predicted binders, the prediction system is very useful for screening epitopes along protein sequences.

## 3   Study Design

This study has three parts designed to understand prediction systems for HLA-binding peptides, the relationship between binding affinities and known T-cell epitopes for survivin and selected HIV proteins, and the relationship of natural epitopes and their mutants.

### 3.1   SVR HLA-Binding Predictor

The Gaussian kernel and the extended sparse coding SVR were combined with the IEDB datasets to build a regression model for binding affinity prediction for each HLA allele. Since vast majority of $IC_{50}$ values from the IEDB database for the eight HLA alleles are within 1 to 50,000 nM, we transformed the binding affinities to the range of 0 and 1 by using $1 - \log(\text{binding affinity})/\log(50000)$ as described in [25].

The model building was done using single-level five-fold cross-validation for each allele. We used the test data in each run to tune regularization and kernel parameters. We got five different regression models from the cross-validation process. We repeated the five-fold cross-validation five times. Regression performance was reported as the mean value of correlation coefficients from the five runs. In addition, we chose the models from a single cross-validation run which

gave the best cross-validation performance. These five regression models were used as a committee for the corresponding single allele to get predicted binding affinity values for all yet unseen peptides.

## 3.2   Prediction of HLA-Binding Peptides in Survivin Protein

The comprehensive experimental relative binding affinity analysis of the complete survivin peptide set presented an opportunity to evaluate how different predictive models perform against an independent data set representing a complete protein. We applied our regression model on the survivin dataset and retrieved prediction results from the IEDB prediction servers to compare their performances. Comparison was based on the correlation coefficient calculated from the predicted binding affinities and the experimentally measured relative binding affinities.

## 3.3   Prediction of T-Cell Epitopes on HIV Proteins and Survivin Protein

We used the SVR model to predict binding affinities for known T-cell epitopes on HIV proteins and survivin within the eight alleles. When a known epitope is longer than a nonamer, we took the highest binding affinity of all possible nonamers within the epitope as its binding affinity. We also did one site mutagenesis on eight T-cell epitopes of HIV proteins and survivin in order to compare their affinities to corresponding ascendant epitopes and find mutation patterns.

# 4   Results

## 4.1   Cross-Validation Performance on the Eight HLA-Alleles

Table 1 shows the correlation coefficient (with standard deviation) of SVR models on the eight HLA-alleles using cross-validation. Most of correlation coefficient performances are satisfactory except for the B*0801 allele. Figure 1 shows the binding affinity distributions of data sets for HLA-A*0201, A*2402 and B*0801 and the horizontal line in each subplot denotes the binding affinity value, $\log_{10}(500)$. From the analysis of experimental data from [3], $\log_{10}(500)$ was taken as a threshold for binder and non-binder, which means that for a peptide with $\log_{10}(IC_{50})$ value less than $\log_{10}(500)$ it should be treated as a binder. Although the threshold, $\log_{10}(500)$, is arbitrary, it enables objective separation of binders from non-binders. From Figure 1, it is clear that there are very few samples (only 21) with binding affinities less than the threshold versus the total around 700 samples for B*0801 allele. For A*2402 alleles, there are only 197 samples in total, but the predictor performance is superior to that of B*0801. A possible explanation is that the dataset is more balanced. For the A*0201 allele the dataset is slightly unbalanced, however, there are still about 1000 samples with $IC_{50}$ values below the binding threshold. This observation holds for other alleles. More samples with stronger binding affinities seem to imply better prediction performance.

**Table 1.** SVR cross-validation correlation coefficient (r) performance on HLA-alleles

| Allele | A*0101 | A*0201 | A*0301 | A*1101 | A*2402 | B*0702 | B*0801 | B*1501 |
|---|---|---|---|---|---|---|---|---|
| Size | 1157 | 3089 | 2094 | 1985 | 197 | 1262 | 708 | 978 |
| Mean(r) | 0.781 | 0.847 | 0.766 | 0.823 | 0.669 | 0.812 | 0.287 | 0.726 |
| Std | 0.005 | 0.002 | 0.004 | 0.002 | 0.003 | 0.004 | 0.073 | 0.008 |
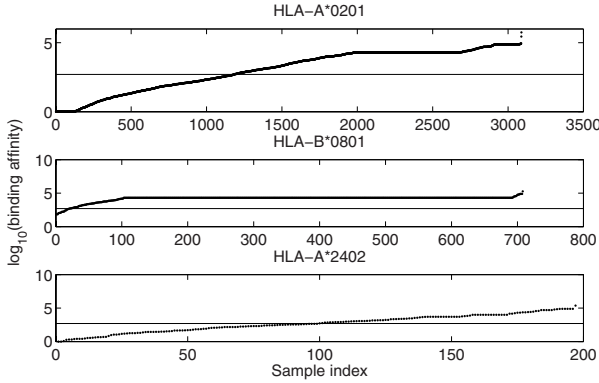


**Fig. 1.** Experimental binding affinity plots for A*0201, A*2402 and B*0801 data sets. The horizontal lines denote the binding affinity values, $\log_{10}(500)$.
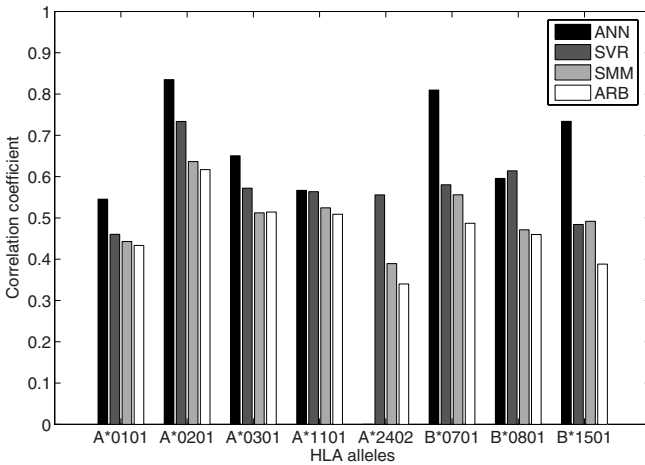


**Fig. 2.** Correlation coefficient performance comparison between IEDB ANN, SMM, ARB and SVR models on the whole survivin dataset

### 4.2  Prediction of Binding Affinity along Survivin Protein Sequence

For each of the eight HLA alleles we calculated the correlation coefficient between experimental binding values and predicted binding affinities. $\log_{10}(IC_{50})$ values of all nonamers in the survivin sequence were predicted using IEDB ANN, SMM and ARB prediction tools (tools.immuneepitope.org/analyze/html/ mhc_binding.html). Figure 2 illustrates performance differences between models (HLA-A*2402 ANN method is not available from the IEDB web server). IEDB ANN is generally superior, followed by SVR. SMM and ARB show similar performance inferior to the other two.

For screening potential epitopes/binders in long protein sequences, it is useful to look at the match curve to assess how many peptides are typically required to test to identify all true epitopes/binders. Figure 3 shows the match curves for the eight alleles using the IEDB ANN and our SVR models using experimental data for survivin peptides. For A*0101 there are no binders according to the experimental settings in [23] and for A*2402 results for IEDB ANN are unavailable. The classification performance of ANN and SVR is very similar. For A*0201, A*0301 and B*0801, SVR is slightly better than the IEDB ANN. Most of experimentally determined binders are within a few numbers of the predicted top peptides. The worst performance by this measure is for the A*2402 allele, where 19 of 23 binders are within predicted top 65 peptides.
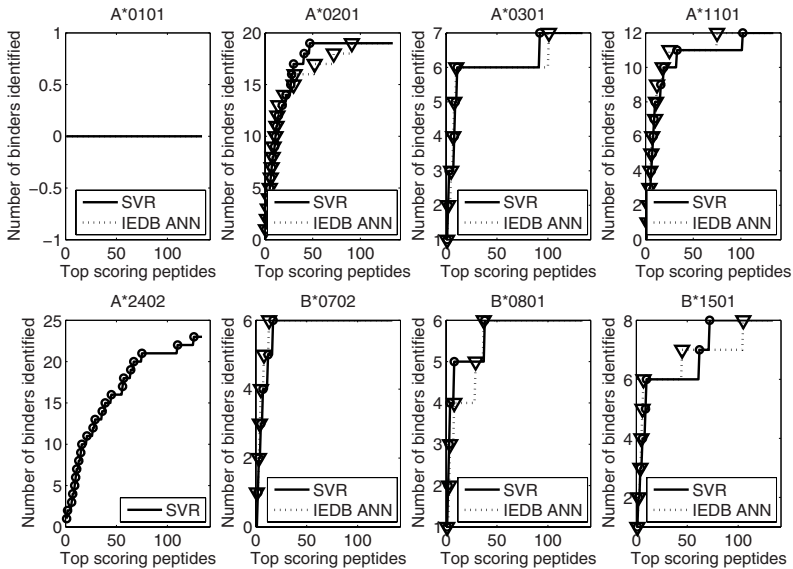


**Fig. 3.** Match curves for experimentally identified binders on survivin protein

### 4.3  Predictions of T-Cell Epitopes

Figure 4 displays the predicted binding affinities, $\log_{10}(IC_{50})$, using SVR for known HIV T-cell epitopes. The upper horizontal line indicates $\log_{10}(5000)$ and

followed by $\log_{10}(500)$ and $\log_{10}(50)$. In [10], a peptide with $\log_{10}(IC_{50})$ less than $\log_{10}(5)$ is a very good binder; good binder with affinity between $\log_{10}(5)$ and $\log_{10}(50)$; intermediate binder between $\log_{10}(50)$ and $\log_{10}(500)$ and low affinity binder between $\log_{10}(500)$ and $\log_{10}(5000)$. In B*1501, all epitopes are low affinity binders. In B*0801 predictions are not informative, reflecting poor training set. For other alleles, T-cell epitopes show a broad range of binding affinities, most of which are between $\log_{10}(50)$ and $\log_{10}(5000)$ .
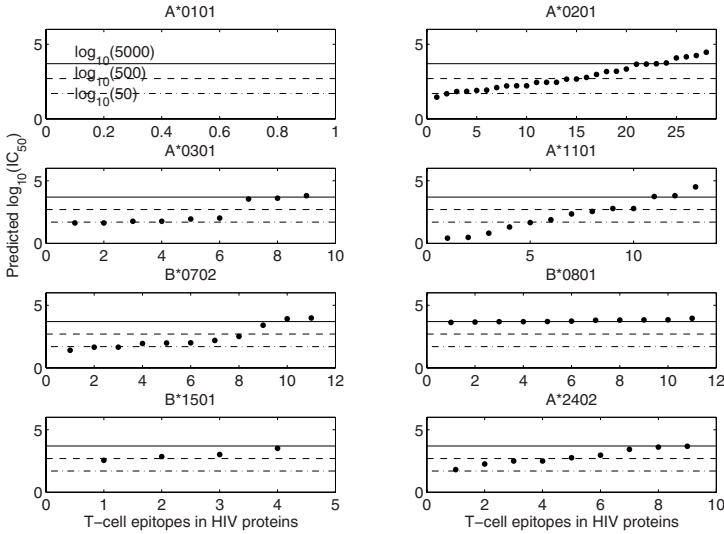


**Fig. 4.** SVR predicted binding affinities for known HIV T-cell epitopes

We compared the predicted binding affinities with experimental data for known survivin T-cell epitopes, shown in Table 2. Both known T-cell epitopes have moderate binding affinity. The 96-104 epitope is the third highest binder to A*0201 while 20-28 is the second highest binder to A*2402 of all survivin peptides [23]. Predicted binding affinities are approximately within 2-fold concentration of their experimental affinities indicating excellent correlation.

Table 3 shows natural epitopes and their *in silico* mutant versions with the highest predicted binding affinities. The affinity change varies from 2 to 10 fold. For A*0201, the mutant happens at the second position from T to M or L; for A*0101, it is at the ninth position from E to Y; for A*1101, at the seventh position from C to F; and for A*2402, at the second position from T to Y. This example illustrates

**Table 2.** Predicted binding affinities ($IC_{50}$) of known survivin T-cell epitopes

| Allele Name | Start Position | End Position | Peptide | $IC_{50}$ (SVR) | Experimental $IC_{50}$ (from [23]) |
|---|---|---|---|---|---|
| A*0201 | 96 | 104 | LTLGEFLKL | 893 | 430 nM |
| A*2402 | 20 | 28 | STFKNWPFL | 1290 | 740 nM |

a possible vaccine engineering applications with modified peptides. In a successful case study of immunotherapy with modified survivin 96-104 peptide LMLGE-FLKL in a liver metasthasis of pancreatic cancer the result was a complete remission of metasthasis [26]. Table 4 shows *in silico* mutations of natural HIV epitopes and their mutants with the highest binding affinities.

**Table 3.** Mutant epitopes with the highest binding affinities vs. natural epitopes

| Allele Name | Natural Epitope | $IC_{50}$ (SVR) | Mutant Epitopes | $IC_{50}$ (SVR) |
|---|---|---|---|---|
| A*0101 | QFEELTLGE | 33144 | QFEELTLGY | 2443 |
| A*0201 | STFKNWPFL | 2077 | SMFKNWPFL | 891 |
| A*0201 | LTLGEFLKL | 893 | LLLGEFLKL | 231 |
| A*1101 | LAQCFFCFK | 21 | LAQCFFFFK | 3 |
| A*2402 | STFKNWPFL | 1290 | SYFKNWPFL | 163 |

**Table 4.** Differences between natural HIV epitopes and their mutants resulting in improved binding affinity

| Allele | $P_5$ | $P_4$ | $P_3$ | $P_2$ | $P_1$ | $P_{1'}$ | $P_{2'}$ | $P_{3'}$ | $P_{4'}$ |
|---|---|---|---|---|---|---|---|---|---|
| A*0201 | Q → Y | W/R → M<br>G → L/P<br>L → Q<br>Y/F → L<br>P/G → V<br>I → Y | R → M<br>T → K<br>Q → R | | | A → I | G → F<br>K → F | | E → F<br>G → F<br>P → L<br>Y → L<br>A → Y |
| A*0301 | | Y → V<br>D/Y → L<br>C → P | D → M | | | Q → R | | | C → F |
| A*1101 | | Y → V | Q → R<br>T → K | | Q → F | A → R<br>Q → V | | W → A<br>M → A | R → K<br>A/T → F<br>A → Y |
| A*2402 | S → Y | | R → Y | | | Q → V | | G → A | A/P → F<br>P → L |
| B*0702 | | V → P<br>T → L<br>E/F → M | R → Y<br>T → K | | | | | G → A | G → L<br>I → K<br>K/W → F |
| B*0801 | D → R | V/G → P<br>G/P → L | Y → R | | | Q → R | | | G/Y/V → L<br>C → F |
| B*1501 | | P → V | Y → R<br>D → K | | | Q → R | | | |

## 5  Conclusions

IEDB ANN model is the best among the four models, followed by the SVR model. The predictions by SMM and ARB models are inferior to them. The

two non-linear methods produced more accurate predictions than two linear methods. We also found that, provided that training sets are representative, the more training samples result in a better prediction performance.

IEDB ANN and SVR models performed similarly in scanning potential binders when tested on survivin sequence. Both true epitopes were within top 2.5% of predicted binders. Therefore, *in silico* models can save significant experimental time and costs in screening potential targets. By analyzing predicted binding affinities of known HIV T-cell epitopes, we found that the range of binding affinities varies for different alleles; and the range of binding affinities include high, moderate, and low affinity. Most of the survivin epitopes are intermediate affinity binders compared to their one-site mutated descendants, some of which are high-affinity binders. The change of $IC_{50}$ varies from 2 to 10 fold for mutated epitopes and most of them were at epitope anchors or auxiliary anchors. These phenomena indicate that high binding affinity binding and immunogenicity are not necessary correlated.

## 6    Discussion

Binding affinity alone is not sufficient to describe the interaction between HLA allele molecules and peptides. Other factors like dissociation rate or stability of each complex are also the determinants of the interaction. Classification of interaction into binders and non-binders only is not informative of immunogenic properties of peptides. Known survivin-derived T-cell epitopes are low affinity binders to their respective HLA molecules.

Most of its known T-cell epitopes of tumor antigen survivin are low affinity binders, which might offer an explanation for lack of response to antigens in cancer patients, self-tolerance, and subdominance [23]. It is unclear which epitopes within a given tumor-associated antigen should be selected to circumvent tolerance and hence serve as the best target in anti-tumor vaccination. One challenge for vaccine design is to enhance the immunogenicity of weak antigens and prevent silencing of active T-cell clones. One possible strategy is to optimize tumor-associated antigen epitope analogs for priming. The *in silico* mutation analysis demonstrated that the optimization should target mainly anchor or auxiliary anchor positions.

## References

1. Brusic, V., Zeleznikow, J.: Computational binding assays of antigenic peptides. Letters in Peptide Sci. 6, 313–324 (1999)
2. Yewdell, J.W.: Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. Immunity 25, 533–543 (2006)

3. Peters, B., Bui, H.H., Frankild, S., Nielson, M., Lundegaard, C., Kostem, E., Basch, D., Lamberth, K., Harndahl, M., Fleri, W., Wilson, S.S., Sidney, J., Lund, O., Buus, S., Sette, A.: A community resource benchmarking predictions of peptide binding to MHC-I molecules. PLoS Comput. Biol. 2(6), 574–584 (2006)

4. Rammensee, H.G., Bachmann, J., Emmerich, N.P., Bachor, O.A., Stevanovic, S.: SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50(3-4), 213–219 (1999)

5. Parker, K.C., Bednarek, M.A., Coligan, J.E.: Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152(1), 163–175 (1994)

6. Udaka, K., Wiesmuller, K.H., Kienle, S., Jung, G., Tamamura, H., Yamagishi, H., Okumura, K., Walden, P., Suto, T., Kawasaki, T.: An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries. Immunogenetics 51(10), 816–828 (2000)

7. Guan, P., Doytchinova, I.A., Zygouri, C., Flower, D.R.: MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. Applied Bioinformatics 2(1), 63–66 (2003)

8. Peters, B., Sette, A.: Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6(132) (2005)

9. Bui, H.H., Sidney, J., Peters, B., Sathiamurthy, M., Sinichi, A., Purton, K.A., Mothe, B.R., Chisari, F.V., Watkins, D.I., Sette, A.: Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. Immunogenetics 57(5), 304–314 (2005)

10. Buus, S., Lauemoller, S.L., Worning, P., Kesmir, C., Frimurer, T., Corbet, S., Fomsgaard, A., Hilden, J., Holm, A., Brunak, S.: Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. Tissue Antigens 62(5), 378–384 (2003)

11. Brusic, V., Bucci, K., Schonbach, C., Petrovsky, N., Zeleznikow, J., Kazura, J.W.: Efficient discovery of immune response targets by cyclical refinement of QSAR models of peptide binding. Journal of Molecular Graphics and Modelling 19(5), 405–411 (2001)

12. Mamitsuka, H.: Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. Proteins 33(4), 460–474 (1998)

13. Dönnes, P., Elofsson, A.: Prediction of MHC class I binding peptides, using SVMHC. BMC Bioinformatics 3(25) (2002)

14. Zhao, Y., Pinilla, C., Valmori, D., Martin, R., Simon, R.: Application of support vector machines for T-cell epitopes prediction. Bioinformatics 19, 1978–1984 (2003)

15. Riedesel, H., Kolbeck, B., Schmetzer, O., Knapp, E.W.: Peptide binding at class I major histocompatibility complex scored with linear functions and support vector machines. Genome Informatics 15(1), 198–212 (2004)

16. Yang, Z.R., Johnson, F.C.: Prediction of T-cell epitopes using biosupport vector machines. J. Chem. Inf. Model 45(5), 1424–1428 (2005)

17. Bozic, I., Zhang, G.L., Brusic, V.: Predictive vaccinology: optimisation of predictions using support vector machine classifiers. In: Gallagher, M., Hogan, J.P., Maire, F. (eds.) IDEAL 2005. LNCS, vol. 3578, pp. 375–381. Springer, Heidelberg (2005)

18. Zhang, G.L., Bozic, I., Kwoh, C.K., August, J.T., Brusic, V.: Prediction of supertype-specific HLA class I binding peptides using support vector machines. Journal of Immunological Methods 320(1-2) (2007)