



## Prediction of class I T-cell epitopes: evidence of presence of immunological hot spots inside antigens

K.N. Srinivasan<sup>1</sup>, G.L. Zhang<sup>2</sup>, A.M. Khan<sup>2</sup>, J.T. August<sup>1,3</sup> and V. Brusic<sup>2,4,\*</sup>

<sup>1</sup>Division of Biomedical Sciences, Johns Hopkins Medicine in Singapore, 41-Science Park Road, Lobby C, The Gemini, Singapore 117610, <sup>2</sup>Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, <sup>3</sup>Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA and <sup>4</sup>Department of Microbiology, National University of Singapore, 5 Science Drive 2, Singapore 117597

Received on January 15, 2004; accepted on March 1, 2004

### ABSTRACT

**Motivation:** Processing and presentation of major histocompatibility complex class I antigens to cytotoxic T-lymphocytes is crucial for immune surveillance against intracellular bacteria, parasites, viruses and tumors. Identification of antigenic regions on pathogen proteins will play a pivotal role in designer vaccine immunotherapy. We have developed a system that not only identifies high binding T-cell antigenic epitopes, but also class I T-cell antigenic clusters termed immunological hot spots.

**Methods:** MULTIPRED, a computational system for promiscuous prediction of HLA class I binders, uses artificial neural networks (ANN) and hidden Markov models (HMM) as predictive engines. The models were rigorously trained, tested and validated using experimentally identified HLA class I T-cell epitopes from human melanoma related proteins and human papillomavirus proteins E6 and E7. We have developed a scoring scheme for identification of immunological hot spots for HLA class I molecules, which is the sum of the highest four predictions within a window of 30 amino acids.

**Results:** Our predictions against experimental data from four melanoma-related proteins showed that MULTIPRED ANN and HMM models could predict T-cell epitopes with high accuracy. The analysis of proteins E6 and E7 showed that ANN models appear to be more accurate for prediction of HLA-A3 hot spots and HMM models for HLA-A2 predictions. For illustration of its utility we applied MULTIPRED for prediction of promiscuous T-cell epitopes in all four SARS coronavirus structural proteins. MULTIPRED predicted HLA-A2 and HLA-A3 hot spots in each of these proteins.

**Contact:** vladimir@i2r.a-star.edu.sg

\*To whom correspondence should be addressed.

### 1 INTRODUCTION

Molecules of adaptive immune responses diversified very rapidly in early vertebrates. Major histocompatibility complex (MHC) molecules play a vital role in the regulation of immune responses (Hudson and Ploegh, 2002; Watts and Amigorena, 2001). Foreign and host proteins are degraded by specialized intracellular mechanisms to short antigenic peptides. The primary function of MHC molecules is to bind and present antigenic peptides on the cell surface for recognition by antigen-specific T-cell receptors (TCRs) of lymphocytes. These processing and presentation mechanisms are essential processes for cellular immune recognition of antigens. MHC class I peptides are primarily generated by the proteasome complex, and are translocated from the cytosol into the lumen of the endoplasmic reticulum (ER) by a transporter associated with antigen processing. In the ER, peptides are loaded onto the MHC class I molecules and are exported to the cell surface for presentation to TCRs. Short peptides presented to TCRs, termed T-cell epitopes, are critical elements for understanding the basis of immunity (Parker *et al.*, 1994; Van Kaer, 2002; Britschgi *et al.*, 2003). Precise identification of T-cell epitopes is a prerequisite for accurate epitope mapping and for design of vaccines and immunotherapies. Peptides that bind to more than one MHC allelic variant ('promiscuous peptides') are important because they are relevant to higher proportions of the human populations and are targets for vaccine and immunotherapy development.

Computational methods have been used for the prediction of T-cell epitopes and are now a standard methodology (Schirle *et al.*, 2001; Yu *et al.*, 2002). *In silico*, T-cell epitope mapping using computational models is emerging as a new approach for the study of peptide vaccines (De Groot *et al.*, 2001). A number of predictive methods for MHC

classes I and II binding peptides are available, including those based on binding motifs (Rammensee *et al.*, 1995), quantitative matrices (Parker *et al.*, 1994), artificial neural networks (ANNs) (Honeyman *et al.*, 1998), hidden Markov models (HMMs) (Mamitsuka, 1998), multivariate statistical approaches (Guan *et al.*, 2003), support vector machines (Zhao *et al.*, 2003) and decision trees (Savoie *et al.*, 1999). Computational strategies for promiscuous class II binding peptides using multiple quantitative matrices (Sturniolo *et al.*, 1999) have been used for vaccine development in cancer (Kobayashi *et al.*, 2001) and infectious disease (Panigada *et al.*, 2002).

In our prediction of promiscuous class I T-cell epitopes, we made predictions of T-cell epitope hot spots in nucleocapsid protein of the severe acute respiratory syndrome coronavirus (SARS-CoV). MULTIPRED, a computational system developed for human leukocyte antigen (HLA) classes I-A2 and I-A3 binding, predicts individual 9-mer T-cell epitopes and also promiscuous class I regions as immunological hot spots, based on HMM (Brusic *et al.*, 2002) and ANN models (Zhang *et al.*, 2003).

Severe acute respiratory syndrome, an outbreak of atypical pneumonia was first reported in Guangdong Province, China in November 2002 and spread to other parts of the world (Rota *et al.*, 2003; Booth *et al.*, 2003). Genome analysis of SARS-CoV revealed the virus to be of completely new pathogenic strain and distantly related to other CoV members (Ruan *et al.*, 2003; Holmes and Enjuanes, 2003; Holmes, 2003). The four major structural proteins of SARS-CoV are: surface spike (S), nucleocapsid (N), envelope (E) and membrane (M) (Marra, 2003; Holmes, 2003). The packaging of the genome to form the viral nucleocapsid is by the N protein, which is incorporated into virions by intracellular budding through a membrane containing three proteins: the S glycoprotein, the M glycoprotein and the small E protein (Kuo and Masters, 2002). It has been demonstrated that antibodies to SARS N proteins are predominant among the early responses to infection (Shi *et al.*, 2003; Liu *et al.*, 2004).

## 2 METHODS

### 2.1 Algorithm

We used ANN and HMM as the prediction engines. The ANN learning algorithm in MULTIPRED is the error back-propagation with sigmoid activation function. The ANN is a three-layer network with structure 267-4-1. The inputs to the ANN are binary strings representing the virtual peptides; the outputs are the binding scores ranged from 1 to 9. In the training dataset, scores 8 and 9 denote high binding affinity; 6 and 7 moderate binding affinity; 4 and 5 low binding affinity. Scores less than 4 denote non-binding. The maximum number of the ANN training cycles is set to 300. The training was repeated four times, and four sets of weights were obtained. The value of learning momentum was 0.5 and of learning rate

was 0.001. Algorithm details of neural network can be found in Brusic *et al.* (1998). The HMM algorithm, training and testing were described earlier (Brusic *et al.*, 2002).

### 2.2 Training and testing of MULTIPRED models

Peptide data containing both binding and non-binding 9-mer peptides were extracted from literature sources, MHCPEP (Brusic *et al.*, 1994), and a set of HLA non-binding peptides (Brusic, unpublished data) for HLA-A2 and HLA-A3 alleles. The dataset had a total of 2962 (604 binders and 2358 non-binders) 9-mer peptides representing 15 different HLA-A2 alleles and 2216 (680 binders and 1536 non-binders) 9-mer peptides for eight different HLA-A3 alleles. The available dataset was divided into training and testing datasets. The training set for a given allele contained virtual peptides that are known to bind other alleles and the test set included all peptides with known binding affinity for the allele to be tested, as described earlier (Brusic *et al.*, 2002). The performances of MULTIPRED ANN and HMM in predicting promiscuous binders to different HLA alleles were tested by a number of trained ANN and HMM models, one model for the prediction of peptide binding to each selected HLA allele. Models for the prediction of alleles with small number of peptides in the dataset could not be tested reliably and were excluded.

The percentage of binders represents ~25% of the training dataset, while non-binders represent the remaining 75%. To optimize the disproportionate numbers of binders and non-binders in the training dataset, new training datasets were constructed using a novel approach in which one or more copies of binders (up to 10 copies) were used in the training datasets (Zhang *et al.*, manuscript in preparation). We trained ANN models to each of the HLA-A2 and HLA-A3 alleles using 10 sets of data to find the composition of training data that result in best predictive performance of the training system. The predictive performance was assessed by the sensitivity (SE), specificity (SP) and receiver operating characteristic (ROC) analysis as described previously (Brusic *et al.* 2002). MULTIPRED can perform a 10-fold internal crossvalidation and calculate  $A_{\text{ROC}}$  values (measure of overall prediction accuracy) of high, moderate and low binders. The accuracy of prediction is poor for values of  $A_{\text{ROC}} < 70\%$ , good for values of  $A_{\text{ROC}} > 80\%$ , and excellent for values of  $A_{\text{ROC}} > 90\%$  (described in Brusic *et al.*, 2002). Peptides that are predicted to bind to multiple HLA alleles are considered promiscuous T-cell epitope candidates.

## 3 RESULTS

The performances of the different MULTIPRED ANN models containing 1–10 copies of binders (HLA-A2 and HLA-A3) were compared. The results showed that the binder/non-binder composition of a dataset influences predictive performance of the model. ANN models trained on the raw dataset containing a single copy of both binders and non-binders produced inferior prediction results. ANN models trained using datasets

**Table 1.** Experimental HLA-A3 restricted peptides and their respective IC<sub>50</sub> values of human melanoma related proteins (gp100, and tyrosinase-Tyr, Reynolds *et al.* (1998)) were compared with MULTIPRED ANN and HMM models for A3 predictions

HLA-A3 restricted peptides	IC <sub>50</sub> (nM)	ANN		HMM	
		Score	Rank* (Predicted)	Score	Rank* (Predicted)
Gp100 <sub>460</sub> GTATLRLVK	3.1	44.08	6 (653)	48.54	15 (653)
Gp100 <sub>609</sub> VVLASLIYR	17.9	44.47	4 (653)	48.62	14 (653)
Tyr <sub>395</sub> SIFEQWLQR	45.5	54.00	1 (521)	53.61	2 (521)
Tyr <sub>14</sub> QTSAGHFPR	16.1	38.37	8 (521)	49.62	4 (521)

When top 10% of the predicted peptides were considered as potential T-cell epitopes, MULTIPRED could predict all the experimental HLA-A3 restricted peptides. \*Numbers in the parentheses indicate the total number of 9mer peptides predicted for that protein.

containing four and six copies of binders provided higher  $A_{roc}$  values. Because the performance of ANN models with four and six copies of binders were comparable, the simpler ANN model with four copies was chosen for the prediction of HLA-A2 promiscuous peptides. For HLA-A3, the ANN model with six copies was found to be more accurate in predicting low (L), moderate (M) and high (H) binding peptides.

### 3.1 Validation of MULTIPRED prediction using experimental binders

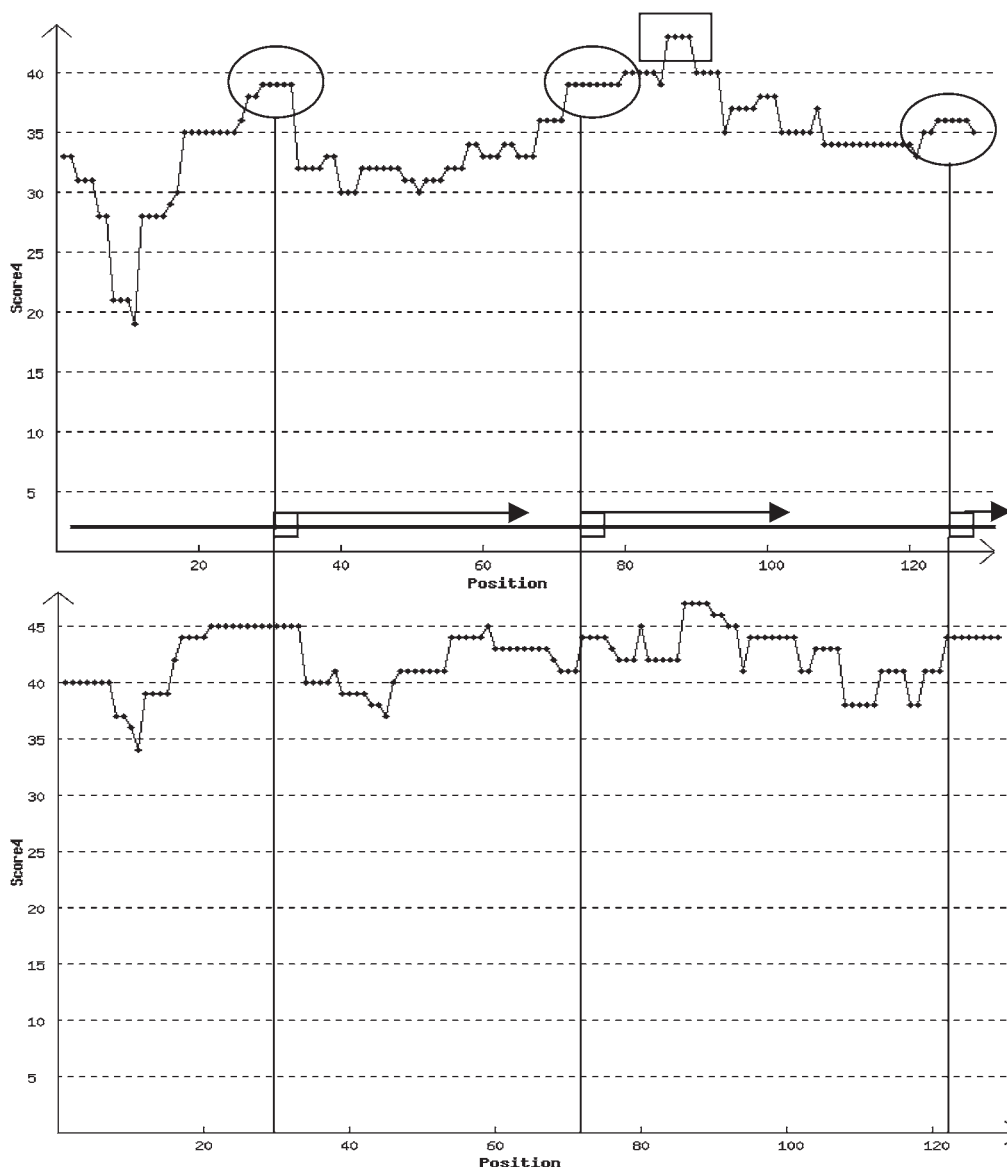
The prediction performance of MULTIPRED for HLA-A2 and HLA-A3 binding was assessed using experimentally known binders. HLA-A2 and HLA-A3 restricted peptides from four melanoma associated proteins, gp100, tyrosinase, tyrosinase-related protein 2 and melanocortin-1 receptor, (Reynolds *et al.*, 1998) were used for validation of MULTIPRED. All duplicate 9-mer peptides in the training dataset were removed and the models were re-trained for prediction of promiscuous peptides to HLA-A2 and HLA-A3. When top 10% of the predicted peptides were considered as potential T-cell epitopes, MULTIPRED could predict most of the HLA-A2 and all the HLA-A3 restricted peptides of the four proteins tested, suggesting that the performance and accuracy of MULTIPRED is reliable. Of 28 known binders tested for HLA-A2, both MULTIPRED ANN and HMM predicted 27 peptides within top 10% of the scores. Within top 5%, ANN predicted 22 peptides and HMM 24 peptides. Hence the prediction accuracy was 96% for top 10% prediction (both methods) and 78.5% and 85.7% for ANN and HMM top 5% prediction, respectively. The prediction accuracy of HLA-A3 peptides tested was 100% for both top 10% and 5% of the predicted peptides (Table 1). To assess the accuracy of individual 9-mer predictions, we compared predictions of HPV E7 HLA-A2 binding peptides with experimental binding measured by Kast *et al.* (1994). HPV E7 is 98 amino acids long

and contains six 9-mer HLA-A2 binders (7–15, 11–19, 12–20, 82–90, 84–92 and 85–93). Of these, four peptides were within top 5% and five were within top 10% of predictions. Top 5% predictions contained one false positive and top 10% predictions contained five false positives.

We have developed a scoring scheme to identify class I regions termed ‘immunological hot spots’ within antigens that are based on high scoring individual 9-mers within a window of 30 amino acids. Immunological hot spots are thus defined as antigenic regions of up to 30 amino acids that are predicted to bind multiple HLA alleles. For validation of hot spot predictions, a test dataset for peptides to HLA-class I alleles were taken from a set of 240 9-mer peptides of human papillomavirus type 16 E6 and E7 proteins reported by Kast *et al.* (1994). All duplicate 9-mers peptides pertaining to E6 and E7 proteins were removed from the training dataset. The class I epitopes were predicted by the use of both MULTIPRED models. The results were sorted to the average score of top four 9-mers within the region (across all the alleles studied for promiscuous prediction) and regions of immunological hot spots were identified. By use of this strategy, our models were successful in identifying class I hot spots for E6 and E7 proteins. MULTIPRED ANN and HMM HLA-A3 output against experimental data from human papillomavirus protein E6 (Kast *et al.*, 1994) is shown in Figure 1.

HPV protein E7 and E6 HLA-A2 hot spots were predicted for validation of MULTIPRED ANN and HMM models. The known HLA-A2 hot spots for protein E7 (E7:7–20 and E7:82–94) that were previously demonstrated (Kast *et al.*, 1994), were predicted by MULTIPRED ANN and HMM models, with a false negative prediction for E7:7–20 by HLA-A2 ANN prediction at threshold >60. Similarly, the known HLA-A2 hot spot for protein E6 (E6 7–34) was predicted with MULTIPRED HMM and ANN models. ANN and HMM predictions produced similar results, with a false positive at position 85–105. The results of validation of hot spot predictions by MULTIPRED models suggests that the overall performances of were reliable and a single MULTIPRED model could make reasonably accurate predictions of peptides binding to multiple alleles of HLA molecules, and also for variants of HLA supertypes that lack experimental binding data. These predictions will be further improved by increased number of training datasets and additional rigorous testing strategies.

The testing results on E6 and E7 proteins indicate that with available datasets ANN models appear more accurate for prediction of HLA-A3 hot spots, while HMM models appear more accurate for HLA-A2 predictions. Therefore we selected HMM as a method of choice for prediction of HLA-A2 T-cell hot spots and ANN for prediction of HLA-A3 hot spots. Testing results provided a basis for determination of prediction thresholds: A2 scores were calculated as sum of individual predictions for eight HLA-A2 variants and A3 scores as sum of individual predictions for six HLA-A3 variants. The suitable thresholds for both ANN A2 and HMM A2 hot spot



**Fig. 1.** Predictions of the start positions (horizontal arrows) of HLA-A3 hot spots from human papillomavirus Type 16 protein E6: 33–67, 75–101 and 125–151 (Kast *et al.*, 1994). ANN prediction is shown at the top and HMM at the bottom. All three hot spots have been captured with ANN A3 predictions with the threshold >35 with a false positive shown inside the square. ANN predictions appear to be more accurate than HMM predictions for HLA-A3 hot-spots.

prediction thresholds were set to >60, while ANN A3 and HMM A3 thresholds were set to >35.

### 3.2 Prediction of SARS-CoV proteins immunological hot spots

To illustrate the utility of MULTIPRED for prediction of immunological hot spots, we analyzed four structural proteins from the SARS-CoV. The SARS-CoV protein sequences were retrieved from NCBI GenBank database (AY283798). The four proteins were submitted to MULTIPRED for prediction

of class I (HLA-A2, A3) T-cell epitope hot spots. The hot spots were derived from consensus predictions of both (ANN and HMM) models in MULTIPRED. The results of the analysis showed that SARS-CoV E and M proteins were predicted to possess one hot spot each to HLA-A2 (E 1–52 and M 4–83) and HLA-A3 (E 1–76 and M 74–220) supertypes. Further, we have identified two HLA-A2 (N 202–253 and N 296–347) and four HLA-A3 (N 79–115, N 229–280, N 291–326 and N 341–392) immunological hot spots in SARS N protein. As for SARS S protein, the system predicted three hot spots to

HLA-A2 (S 848–880, S 937–994 and S 1181–1231) and eight hot spots to HLA-A3 (S 270–411, S 1026–1147, S 743–828, S 134–163, S 76–115, S 9–53, S 418–465 and S 886–950). These results indicate the presence of immunological hot spots of both HLA-A2 and HLA-A3 molecules in all SARS-CoV structural proteins. Similar patterns have been observed in 10 dengue virus proteins (data not shown). We propose that T-cell epitopes tend to cluster in certain regions of protein antigens in a HLA supertype-dependent manner. These regions therefore represent immunological hot spots containing multiple T-cell epitopes.

#### 4 DISCUSSION AND CONCLUSION

Our strategy of peptide-based vaccines is to identify promiscuous T-cell epitopes that are representative of large proportion of the human population. The majority of publicly available methods has not been properly assessed for predictive accuracy and do not predict promiscuous peptides for a broad range of HLA alleles. In this context, we have developed a computational system MULTIPRED that identifies promiscuous peptides across HLA-A2 and HLA-A3 alleles, and also a scoring scheme for prediction of immunological hot spots of class I molecules, using MULTIPRED ANN and HMM models. The system was trained and rigorously tested using experimentally known peptides, human melanoma-related proteins and human papillomavirus type 16 proteins E6 and E7. It was found that ANN model could predict HLA-A3 with more accuracy than the HMM model, while HMM appeared to be more accurate for HLA-A2 predictions.

Severe acute respiratory syndrome was a great threat both to public health and economy affecting more than 30 countries around the globe and was of great concern due to formidable morbidity and mortality. Although SARS looked a devastating pandemic and the outbreak was deemed to be under control, the World Health Organization (2003, [http://www.who.int/csr/sars/country/table2003\\_09\\_23/en/](http://www.who.int/csr/sars/country/table2003_09_23/en/)) has urged health authorities not to be contented. Hence there is a need to design a more efficient vaccine to combat the deadly SARS. Current therapeutic strategies to SARS involve the use of convalescent plasma (Burnouf and Radosevich, 2003), glucocorticoids (Li *et al.*, 2003), interferons (Cinatl *et al.*, 2003), but still remains empirical. Peptide-based vaccines offer several potential advantages over the conventional whole proteins in terms of high specificity in eliciting immune responses, ease of manufacturing and quality control and proven successful against specific allergy (Alexander *et al.*, 2002), malaria (Lopez *et al.*, 2001) and certain types of tumors (Tanaka *et al.*, 2003). In this study, MULTIPRED ANN and HMM models identified immunological hot spots in four structural proteins of SARS-CoV. The results show that there are several overlapping hot spots of multiple 30 amino acid regions. Our system could thus predict not only high binding individual 9-mer peptides but also regions of immunological

hot spots in an antigen, which could have potential therapeutic significance as peptide vaccines. This bioinformatics approach to vaccine design increases the efficiency of T-cell epitope screening and will be further enhanced by additional experimental data and enrichment of training datasets.

#### ACKNOWLEDGEMENTS

This research was supported by the National Institutes of Health Grants, NIAID R37 AI 41908 (JTA), and U19 AI 56541 (KNS, GZ, JTA, VB); Biomedical Research Council Grant, Singapore 03/1/55/20/282 (KNS, GZ, JTA, VB); and the National University of Singapore Graduate Scholarship (AMK). The authors are also grateful to the Agency for Science, Technology and Research, Singapore.

#### REFERENCES

- Alexander,C., Kay,A.B. and Larche,M. (2002) Peptide-based vaccines in the treatment of specific allergy. *Curr. Drug Targets Inflamm. Allergy*, **1**, 353–361.
- Booth,C.M., Matukas,L.M., Tomlinson,G.A., Rachlis,A.R., Rose,D.B., Dwosh,H.A., Walmsley,S.L., Mazzulli,T., Avedano,M., Derkach,P. *et al.* (2003) Clinical features and short-term outcomes of 144 patients with SARS in the greater Toronto area. *JAMA*, **289**, 2801–2809.
- Britschgi,M., von Greyerz,S., Burkhart,C. and Pichler,W.J. (2003) Molecular aspects of drug recognition by specific T cells. *Curr. Drug Targets*, **4**, 1–11.
- Brusic,V., Petrovsky,N., Zhang,G. and Bajic,V.B. (2002) Prediction of promiscuous peptides that bind HLA class I molecules. *Immunol. Cell Biol.*, **80**, 280–285.
- Brusic,V., Rudy,G. and Harrison,L.C. (1994) MHCPEP: a database of MHC-binding peptides. *Nucleic Acids Res.*, **22**, 3663–3665.
- Brusic,V., Rudy,G., Honeyman,G., Hammer,J. and Harrison,L. (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics*, **14**, 121–130.
- Burnouf,T. and Radosevich,M. (2003) Treatment of severe acute respiratory syndrome with convalescent plasma. *Hong Kong Med. J.*, **9**, 309.
- Cinatl,J., Morgenstern,B., Bauer,G., Chandra,P., Rabenau,H. and Doerr,H.W. (2003) Treatment of SARS with human interferons. *Lancet*, **362**, 293–294.
- De Groot,A.S., Bosma,A., Chinai,N., Frost,J., Jesdale,B.M., Gonzalez,M.A., Martin,W. and Saint,A. (2001) From genome to vaccine: *in silico* predictions, *ex vivo* verification. *Vaccine*, **19**, 4385–4395.
- Guan,P., Doytchinova,I.A., Zygouri,C. and Flower,D.R. (2003) MHCpred: a server for quantitative prediction of peptide–MHC binding. *Nucleic Acids Res.*, **31**, 3621–3624.
- Holmes,K.V. and Enjuanes,L. (2003) Virology. The SARS coronavirus: a postgenomic era. *Science*, **300**, 1377–1378.
- Holmes,K.V. (2003) SARS coronavirus: a new challenge for prevention and therapy. *J. Clin. Invest.*, **111**, 1605–1609.

- Honeyman, M.C., Brusica, V., Stone, N.L. and Harrison, L.C. (1998) Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol.*, **16**, 966–969.
- Hudson, A.W. and Ploegh, H.L. (2002) The cell biology of antigen presentation. *Exp. Cell. Res.*, **272**, 1–7.
- Kast, W.M., Brandt, R.M., Sidney, J., Drijfhout, J.W., Kubo, R.T., Grey, H.M., Melief, C.J. and Sette, A. (1994) Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins. *J. Immunol.*, **152**, 3904–3912.
- Kobayashi, H., Lu, J. and Celis, E. (2001) Identification of helper T-cell epitopes that encompass or lie proximal to cytotoxic T-cell epitopes in the gp100 melanoma tumor antigen. *Cancer Res.*, **61**, 7577–7584.
- Kuo, L. and Masters, P.S. (2002) Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J. Virol.*, **76**, 4987–4999.
- Li, X.W., Jiang, R.M. and Guo, J.Z. (2003) Glucocorticoid in the treatment of severe acute respiratory syndrome patients: a preliminary report. *Zhonghua Nei Ke Za Zhi.*, **42**, 378–381.
- Liu, X., Shi, Y., Li, P., Li, L., Yi, Y., Ma, Q. and Cao, C. (2004) Profile of antibodies to the nucleocapsid protein of the severe acute respiratory syndrome (SARS)-associated coronavirus in probable SARS patients. *Clin. Diagn. Lab. Immunol.*, **11**, 227–228.
- Lopez, J.A., Weilenman, C., Audran, R., Roggero, M.A., Bonelo, A., Tiercy, J.M., Spertini, F. and Corradin, G. (2001) A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur. J. Immunol.*, **31**, 1989–1998.
- Mamitsuka, H. (1998) Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. *Proteins*, **33**, 460–474.
- Marra, M.A. (2003) The Genome sequence of the SARS-associated coronavirus. *Science*, **300**, 1399–1404.
- Parker, K.C., Bednarek, M.A. and Coligan, J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, **152**, 163–175.
- Panigada, M., Sturniolo, T., Besozzi, G., Boccieri, M.G., Sinigaglia, F., Grassi, G.G. and Grassi, F. (2002) Identification of a promiscuous T-cell epitope in *Mycobacterium tuberculosis* Mce proteins. *Infect. Immun.*, **70**, 79–85.
- Rammensee, H.G., Friede, T. and Stevanoviic, S. (1995) MHC ligands and peptide motifs: first listing. *Immunogenetics*, **41**, 178–228.
- Reynolds, S.R., Celis, E., Sette, A., Oratz, R., Shapiro, R.L., Johnston, D., Fotino, M. and Bystry, J.C. (1998) HLA-independent heterogeneity of CD8+ T cell responses to MAGE-3, Melan-A/MART-1, gp100, tyrosinase, MC1R, and TRP-2 in vaccine-treated melanoma patients. *J. Immunol.*, **161**, 6970–6976.
- Rota, P.A., Oberste, M.S., Monroe, S.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Janogli, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., et al. (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, **300**, 1394–1399.
- Ruan, Y.J., Wei, C.L., Ee, A.L., Vega, V.B., Thoreau, H., Su, S.T., Chia, J.M., Ng, P., Chiu, K.P., Lim, I., et al. (2003) Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet*, **361**, 1779–1785.
- Savoie, C.J., Kamikawaji, N., Sasazuki, T. and Kuhara, S. (1999) Use of BONSAI decision trees for the identification of potential MHC class I peptide epitope motifs. *Pac. Symp. Biocomput.*, 182–189.
- Schirle, M., Weinschenk, T. and Stevanovic, S. (2001) Combining computer algorithms with experimental approaches permits the rapid and accurate identification of T cell epitopes from defined antigens. *J. Immunol. Methods*, **257**, 1–16.
- Shi, Y., Yi, Y., Li, P., Kuang, T., Li, L., Dong, M., Ma, Q. and Cao, C. (2003) Diagnosis of severe acute respiratory syndrome (SARS) by detection of SARS coronavirus nucleocapsid antibodies in an antigen-capturing enzyme-linked immunosorbent assay. *J. Clin. Microbiol.*, **41**, 5781–5782.
- Sturniolo, T., Bono, E., Ding, J., Radrizzani, L., Tuercio, O., Sahin, U., Braxenthaler, M., Gallazzi, F., Protti, M.P., Sinigaglia, F. and Hammer, J. (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.*, **17**, 555–561.
- Tanaka, S., Harada, M., Mine, T., Noguchi, M., Gohara, R., Azwma, K., Yamada, A., Morinaga, A., Nishikari, M., Tamura, M., et al. (2003) Peptide vaccination for patients with melanoma and other types of cancer based on pre-existing peptide-specific cytotoxic T-lymphocyte precursors in the periphery. *J. Immunother.*, **26**, 357–366.
- Van Kaer, L. (2002) Major histocompatibility complex class I-restricted antigen processing and presentation. *Tissue Antigens*, **60**, 1–9.
- Watts, C. and Amigorena, S. (2001) Phagocytosis and antigen presentation. *Semin. Immunol.*, **13**, 373–379.
- World Health Organization. Cumulative number of reported probable cases of SARS: 1 Nov 2002 to 31 Jul 2003.
- Yu, K., Petrovsky, N., Schonbach, C., Koh, J.Y. and Brusica, V. (2002) Methods for prediction of peptide binding to MHC molecules: a comparative study. *Mol. Med.*, **8**, 137–148.
- Zhang, G., Khan, A.M., Srinivasan, K.N., August, J.T., and Brusica, V. (2003) Neural models for predicting viral vaccine targets. *Proceedings of the Conference on Neuro-Computing and Evolving Intelligence 2003 (NCEI'2003)*, Auckland, New Zealand, KEDRI, pp. 35–36.
- Zhao, Y., Pinilla, C., Valmori, D., Martin, R. and Simon, R. (2003) Application of support vector machines for T-cell epitopes prediction. *Bioinformatics*, **19**, 1978–1984.