Prediction of MHC Class II binding peptides incorporating bayesian transfer hierarchies

Ravikiran Janardhana
Department of Computer Science
University of North Carolina at Chapel Hill
Email: ravikirn@cs.unc.edu

***Abstract*—T-cells are key players in regulating a specific immune response. Activation of cytotoxic T-cells requires recognition of specific peptides bound to Major Histocompatibility Complex (MHC) class II molecules. MHC-peptide complexes are potential tools for diagnosis and treatment of pathogens and cancer, as well as for the development of peptide vaccines. Only one in 100 to 200 potential binders actually binds to a certain MHC molecule, therefore a good prediction method for MHC class II binding peptides can reduce the number of candidate binders that need to be synthesized and tested for successful design of peptide and protein based vaccines.**

***Index Terms*—MHC class II, Peptide prediction, Machine Learning**

I. Introduction

Vaccines continue to have an enormous and unprecedented positive impact on humanity and its wellbeing. Hundreds of millions of human lives have been saved since the first vaccine was discovered: Edward Jenners smallpox vaccine in 1796 [1]. Yet the need to develop and deploy new vaccines has never been more urgent. Infectious disease causes about 25% of global deaths, particularly in children under five. The leading annual causes of death are 2.9 millions for tuberculosis; 2.5 million for diarrhoeal illnesses, especially rotaviruses; a rapidly escalating 2.3 million for HIV/AIDS; and 1.08 million deaths for malaria. There are no effective vaccines for HIV and Malaria, and the only vaccine available for tuberculosis is of limited utility. Consider also the 35 new, previously unknown, infectious diseases identified in the past 25 years: ebola, SARS, Dengue, West Nile fever, and potentially pandemic H5N1 influenza among them.

Historically, vaccines have been attenuated whole pathogen vaccines such as BCG for TB or Sabins Polio vaccine. Issues of safety have led to the development of other strategies for vaccine development, separately focusing on antigen and epitope vaccines. The epitope is the minimal structure able to evoke an immune response. It is the immunological quantum that lies at the heart of immunity. Epitope-based vaccines have the advantage that many sequences able to induce autoimmunity or adverse reactions can be eliminated. Such vaccines are intrinsically safer: they contain no viable microorganisms and cannot induce microbial disease. However, several significant obstacles must be overcome before epitope-based vaccines can reach the market en masse. One such obstacle is MHC polymorphism.

Major histocompatibility complex (MHC) proteins, also known as human leukocyte antigens (HLA), are glycoproteins which bind within the cell short peptides, also called epitopes, derived from host and/or pathogen proteins, and present them at the cell surface for inspection by T-cells. T cell recognition is a fundamental mechanism of the adaptive immune system by which the host identifies and responds to foreign antigens.

There are two classes of MHC molecules: class I and class II. MHC class I molecules typically present peptides from proteins synthesized within the cell (endogenous processing pathway). MHC class II proteins primarily present peptides derived from endocytosed extracellular proteins (exogenous processing pathway). Both classes of MHC proteins are extremely polymorphic. More than 3500 molecules are listed in IMGT/HLA database [2]. MHC class I proteins are encoded by three loci: HLA-A, HLA-B, and HLA-C. MHC class II proteins also are encoded by three loci: HLA-DR, HLA-DQ, and HLA-DP. The peptide binding site of class I proteins has a closed cleft, formed by a single protein chain (α-chain) [3]. Usually, only short peptides of 8 to 11 amino acids bind in an extended conformation. In contrast, the cleft of class II proteins is open-ended, allowing much longer peptides to bind, although only 9 amino acids actually occupy the site. The class II cleft is formed by two separate protein chains: α and β [3]. In this paper, we will be dealing only with MHC class II proteins.

Many bioinformatics methods exist to predict peptide MHC binding. Experimentally determined affinities data have formed the basis of many peptide-MHC binding prediction methods, able effectively to discriminate binding from non-binding peptides. Such methods include so-called motifs, as well as highly sophisticated computer science algorithmsartificial neural networks [4], HMMs [5], and support vector machines (SVMs) [6] and methods derived from computational chemistry, such as QSAR analysis [7] and structure-based approaches [8].

Dimitrov et al. [9] and Wang et al. [10] have published detailed assessment of MHC Class II peptide binding predictors. In all of these predictions, it is assumed that each of the peptide binding predictors are independent of each other. In this paper, we propose that the prediction of MHC Class II alleles of the same genetic loci (HLA-DR, HLA-DQ and HLA-DP) are dependent on each other since these alleles are nothing but an alternative form of a same gene. We tie pairs of MHC Class

II alleles together and jointly learn the prediction/classifier parameters. These results are then compared to the state of the art support vector machine methods [6] and we show that our joint classifier either matches or outperforms the SVM method consistently.

A. *Subsection Heading Here*

Subsection text here.

1) Subsubsection Heading Here: Subsubsection text here.

II. Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank...

References

[1] D. R. Flower, "Vaccines: Their place in history," *Bioinformatics for Vaccinology*, pp. 1–54, 2008.

[2] J. Robinson, J. A. Halliwell, H. McWilliam, R. Lopez, P. Parham, and S. G. E. Marsh, "The imgt/hla database," *Nucleic Acids Research*, 2012.

[3] C. A. Janeway, P. Travers, M. Walport, and J. D. Capra, "Antigen recognition by b-cell and t-cell receptors," *Immunobiology: The Immune System in Health and Disease*, 2001.

[4] K. Gulukota and C. DeLisi, "Neural network method for predicting peptides that bind major histocompatibility complex molecules," *Methods in Molecular Biology*, vol. 156, September 2000.

[5] H. Noguchi, R. Kato, T. Hanai, Y. Matsubara, H. Honda, V. Brusic, and T. Kobayashi, "Hidden markov model-based prediction of antigenic peptides that interact with mhc class ii molecules," *Journal of Bioscience and Bioengineering*, vol. 94, no. 3, pp. 264 – 270, 2002.

[6] J. Wan, W. Liu, Q. Xu, Y. Ren, D. Flower, and T. Li, "Svrmhc prediction server for mhc-binding peptides," *BMC Bioinformatics*, vol. 7, pp. 1–5, 2006.

[7] I. Doytchinova, V. Walshe, P. Borrow, and D. Flower, "Towards the chemometric dissection of peptide hla-a*0201 binding affinity: comparison of local and global qsar models," *Journal of Computer-Aided Molecular Design*, vol. 19, pp. 203–212, 2005.

[8] M. Davies, C. Hattotuwagama, D. Moss, M. Drew, and D. Flower, "Statistical deconvolution of enthalpic energetic contributions to mhc-peptide binding affinity," *BMC Structural Biology*, vol. 6, pp. 1–13, 2006.

[9] I. Dimitrov, P. Garnev, D. R. Flower, and I. Doytchinova, "Mhc class ii binding predictiona little help from a friend," *Journal of Biomedicine and Biotechnology*, 2010.

[10] P. Wang, J. Sidney, C. Dow, B. Moth, A. Sette, and B. Peters, "A systematic assessment of mhc class ii peptide binding predictions and evaluation of a consensus approach," *PLoS Computational Biology*, 2008.