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.

I. Introduction

The activation of CD4+ helper T cells is essential for the development of adaptive immunity against pathogens [1]. A critical step in CD4+ T cell activation is the recognition of epitopes presented by MHC class II molecules [2]. MHC class II molecules are heterodimers expressed on the surface of professional antigen presenting cells that bind peptide fragments derived from protein antigens. X-ray crystallographic studies demonstrated that the MHC class II epitope binding site consists of a groove and several pockets provided by a β-sheet and two α-helices. Unlike class I, the class II binding groove is open at both ends. As a result, peptides binding to class II molecules tend to be of variable length, but typically between 13 and 25 residues.

A hallmark of the MHC class II binding peptide groove is that there are four major pockets. These pockets accommodate side chains of residues 1, 4, 6, and 9 of a 9-mer core region of the binding peptide. This core region interaction largely determines binding affinity and specificity [3]. In addition, peptide residues immediately flanking the core region have been indicated to make contact with the MHC molecule outside of the binding groove, and to contribute to MHC-peptide interaction. MHC class II molecules are highly polymorphic, and this polymorphism largely corresponds with differences along the peptide binding groove. However, the binding motifs derived for MHC class II molecules are highly degenerate, and many promiscuous peptides have been identified that can bind multiple MHC class II molecules. Promiscuous peptides are a prime target for vaccine and immunotherapy and computational tools have been developed to facilitate systematic scanning for promiscuous peptides.

Computational prediction of MHC class II epitopes is of important theoretical and practical value, as experimental dentification is costly and time consuming. Many computational methods exist to predict peptide MHC binding.

Experimentally determined affinities data have formed the basis of many peptide-MHC binding prediction methods and has enabled classification of binding and nonbinding peptides. Highly sophisticated computer science algorithms such as artificial neural networks [4], HMMs [5], support vector machines (SVMs) [6] and methods derived from computational chemistry, such as QSAR analysis [7] and structure-based approaches [8] have been previous deployed to solve this problem.

Dimitrov et al. [9] and Wang et al. [10] have published detailed assessment of MHC Class II peptide binding predictors. In the above mentioned assessment, it was clear that each predictor for a specific genetic loci (including its alleles) worked independently i.e., no knowledge sharing occured between predictors although they might be predicting alleles of the same genetic loci. 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 the same gene. We tie pairs of MHC Class II alleles together and jointly learn the prediction/classifier parameters. This approach allows for transfer learning via high dimensional bayesian transfer hierarchies [11]. Our results are then compared to the state of the art support vector machine methods [6], GLMNET [12] and we show that our joint classifier either matches or outperforms the SVM and GLMNET methods consistently.

II. Method

Section text here.

A. Dataset

SubSection text here.

B. Features

SubSection text here.

C. Optimization Setup

SubSection text here.

D. Algorithm

SubSection text here.

III. Experimental Results

Section text here.

A. Dataset

Section text here.

IV. Conclusion

Section text here.

V. Future Work

Section text here.

VI. Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank...

References

[1] M. K. Jenkins, A. Khoruts, E. Ingulli, D. L. Mueller, S. J. McSorley, R. L. Reinhardt, A. Itano, and K. A. Pape, "In vivo activation of antigen-specific cd4 t cells," *Annual Review of Immunology*, vol. 19, no. 1, pp. 23–45, 2001.

[2] M. Rudolph, R. Stanfield, and I. Wilson, "How tcrs bind mhcs, peptides, and coreceptors," *Annual Review of Immunology*, vol. 24, no. 1, pp. 419–466, 2006.

[3] E. Y. Jones, L. Fugger, J. L. Strominger, and C. Siebold, "Mhc class ii proteins and disease: a structural perspective," *Nat Rev Immunol*, vol. 6, no. 4, pp. 271–282, Apr. 2006. [Online]. Available: http://dx.doi.org/10.1038/nri1805

[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.

[11] G. Elidan, B. Packer, G. Heitz, and D. Koller, "Convex point estimation using undirected bayesian transfer hierarchies," *CoRR*, vol. abs/1206.3252, 2008.

[12] J. H. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2 2010.