**Determination of a predictive cleavage motif for eluted major histocompatibility complex class II ligands**

CD4$^+$ T cells have a major role in regulating immune responses. They are activated by recognition of peptides mostly generated from exogenous antigens through the major histocompatibility complex (MHC) class II pathway. Identification of epitopes is important and computational prediction of epitopes is used widely to save time and resources. Although there are algorithms to predict binding affinity of peptides to MHC II molecules, no accurate methods exist to predict which ligands are generated as a result of natural antigen processing. We utilized a dataset of around 14,000 naturally processed ligands identified by mass spectrometry of peptides eluted from MHC class II expressing cells to investigate the existence of sequence signatures potentially related to the cleavage mechanisms that liberate the presented peptides from their source antigens. This analysis revealed preferred amino acids surrounding both N- and C-terminuses of ligands, indicating sequence-specific cleavage preferences. We used these cleavage motifs to develop a method for predicting naturally processed MHC II ligands, and validated that it had predictive power to identify ligands from independent studies. We further confirmed that prediction of ligands based on cleavage motifs could be combined with predictions of MHC binding, and that the combined prediction had superior performance. However, when attempting to predict CD4$^+$ T cell epitopes, either alone or in combination with MHC binding predictions, predictions based on the cleavage motifs did not show predictive power. Given that peptides identified as epitopes based on CD4$^+$ T cell reactivity typically do not have well-defined termini, it is possible that motifs are present but outside of the mapped epitope. Our attempts to take that into account computationally did not show any sign of an increased presence of cleavage motifs around well-characterized CD4$^+$ T cell epitopes. While it is possible that our attempts to translate the cleavage motifs in MHC II ligand elution data into T cell epitope predictions were suboptimal, other possible explanations are that the cleavage signal is too diluted to be detected, or that elution data are enriched for ligands generated through an antigen processing and presentation pathway that is less frequently utilized for T cell epitopes.