Characterizing the binding motifs of 11 common human HLA-DP and HLA-DQ molecules using \textit{NNAlign} \\
Massimo Andreatta$^1$ and Morten Nielsen$^1$ \\

$^1$Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark \\

EXTENDED ABSTRACT \\

\textbf{Introduction:} Unlike MHC class I, which samples peptides from cytosolic proteins, MHC class II molecules present short peptide sequences derived from extracellular proteins. Human MHC class II molecules are heterodimers consisting of an alpha and a beta chain encoded on chromosome 6 in one of three HLA loci: DR, DP and DQ. Compared with DR molecules, the specificities of DP and DQ molecules have only been studied to a limited extent, and their binding motifs are poorly characterized and understood. The scarcity of binding data for DP and DQ molecules is mainly due to the relative difficulty, compared to HLA-DR, of obtaining experimental binding data for these molecules, but also to the common assumption that DR molecules are more important in mediating immune responses. However, a growing number of reports associate certain DP and DQ alleles to several diseases, such as diabetes type I and celiac disease, as well as to cancer. The recent publication of larger data sets of peptide binding data to DP and DQ molecules opens the possibility to use data-driven bioinformatics methods to accurately define the binding motifs of these molecules. \\

\textbf{Methods:} \textit{NNAlign} is a neural network-based method specifically designed to identify short linear motifs contained in large peptide data sets. The method has been shown to perform significantly better than any other publicly available method for MHC class II binding prediction, including HLA-DP and DQ molecules. Here, we applied \textit{NNAlign} on a large data set of 17,092 measured peptide-MHC class II affinities for 5 HLA-DP and 6 HLA-DQ molecules among the most common in the human population to characterize their specificities and binding motifs. \\

\textbf{Results:} For what concerns HLA-DP, there appears to be a common pattern in all the 5 variants under consideration, with primary anchor positions at P1 and P6 with preference for hydrophobic and aromatic residues. Some variants show an additional hydrophobic anchor at P9 and other minor differences, but in general there appears to be a consistent overlap in the binding specificities of all 5 molecules. The same cannot be said for HLA-DQ, where most of the molecules have very different anchor positions, anchor spacing and amino acid preferences. Thus, there does not seem to be a supertypical mode of binding for DQ, and each variant appears to be characterized by a distinct binding specificity. Importantly, the sequence logo representation provides a quantitative measure of the relevance of each position in the binding core,
and the relative importance of each amino acid, in determining the specificities of a given molecule, a differentiation that was not obtained in previous reports. This study first and foremost demonstrates the power of the NNAalign method to, in a fully automated manner, identify and characterize the receptor-binding motif from a set of peptide binding data. Secondly, it underlines the importance of generating such peptide data sets in order to carry out receptor-binding motif characterizations, gain insights into the peptide-binding repertoire of MHC molecules and reveal details about which amino acids and amino acid positions are critical for binding and, potentially, for peptide immunity.

Figure 1: Sequence logos for 5 HLA-DP molecules. Hydrophobic amino acids are shown as black, acidic amino acids as red, basic amino acids as blue, neutral and polar amino acids as green and pink. All variants appear to share two main hydrophobic/aromatic anchors at P1 and P6, with an additional P9 anchor for some variants.

References:
