Poster H-8

Prediction of Peptide Binding to MHC Class I Using Sliding Motifs and Regression Trees

Authors:
Ana Paula Sales (*Computational Biology and Bioinformatics Program, Duke University*)
Georgia D. Tomaras (*Department of Surgery, Duke University Medical Center*)
Thomas B. Kepler (*Department of Biostatistics and Bioinformatics, Duke University*)

Short Abstract: Accurate prediction of peptide binding to Major Histocompatibility Complex class I molecules (MHCI) can accelerate vaccine development. We present a novel approach that integrates experimental and computational methods to elucidate the factors that affect MHCI-peptide binding and to identify potential epitopes for vaccines against pathogens causing emerging infectious diseases.

Long Abstract:

Background

The binding between peptides and Major Histocompatibility Complex class I molecules (MHCI) plays an important role in the immune system. Recognition of MHCI-bound peptides by the T cell receptor leads to the activation of T cells and the initiation of the cellular immune response, which culminates with the clearance of mutated or infected cells. Because MHCI binding is a prerequisite for T cell activation, understanding how MHCI interacts with peptides and being able to predict MHCI-peptide binding is of both theoretical and practical value. Theoretically, it is important for understanding the basis of immunity and pathogenesis. Practically, it promotes advances in biomedicine by accelerating the development of vaccines and drugs against infectious agents, cancer and autoimmune diseases.

In order to learn precisely which properties of peptides determine how they interact with MHCI, we use an integrated approach combining experimental and computational methods. An innovative high throughput experimental platform is used to generate binding affinity measurements between MHCI molecules and peptides. This data is then analyzed computationally and used in the construction of a regression tree algorithm to predict what factors contribute, positively or negatively, to MHCI interaction with peptides. Ultimately we will use this method to identify potential epitopes for the development of vaccines against pathogens responsible for emerging/re-emerging infectious diseases.

Experimental Methods

We have selected target proteins from Variola major, Ebola virus, and the multi-drug-resistant Mycobacterium tuberculosis. For each one of the target proteins, 9-mer peptides overlapping by 8 amino acids have been synthesized and screened for binding against 8 MHCI alleles that together cover a large fraction of the US population.

The platform used to screen these peptides consists of 96-well microtiter plates, in which

each well is coated with properly folded MHCI-peptide complex composed of the MHCI heavy chain of a single allele, the MHCI light chain (beta-2-microglobulin) and a placeholder peptide. The affinity measurements are then obtained in a series of three steps: First, the beta-2-microglobulin and the placeholder peptide are removed, leaving only the heavy chain attached to the bottom of the well in an unfolded configuration. Second, a candidate peptide, extra beta-2-microglobulin, and a fluorescently labeled anti-MHCI antibody are added to each well and incubated overnight. Third, any unbound molecule is removed and the fluorescence signal of each well is recorded. The antibody is designed to bind only to a properly folded MHCI heavy chain, so that the fluorescence measure from each well is proportional to the number of MHCI molecules bound to peptides and thereby indicative of the affinity between the specific MHCI allele and the candidate peptide.

Binding motifs redefined

Peptides eluted from MHCI molecules are generally (but not always) 8 to 10 amino acids long and display two or more highly conserved amino acids at the so-called “anchor positions.” In this context the amino acids in a peptide are labeled 1 through 9, starting with the first amino acid at the amino terminus end of the peptide, and binding motifs are defined by the amino acid types required at each anchor position. This motif definition implies that two 9-mer peptides overlapping by 8 amino acids would not be able to be bind strongly to the same MHCI allele since the anchor residues in one of the peptides would be off by one position. However, data produced in our laboratory, as well as described in the literature, contradicts this motif definition.

Here we define a binding motif by two or more residue types and the distance between these residues. For example, “Alanine-7-Hydrophobic” is a binding motif in which the essential residues are of the type “alanine” and “hydrophobic” and they must be separated by 7 amino acids. Note that with this definition of binding motif, a residue type could be an amino acid itself or a physical/chemical property. Using this definition a motif is allowed to slide along the length of the peptide. As a result, perfect alignment between the peptide and the MHCI groove is not a prerequisite for binding. Moreover, peptides are allowed to hang off the edge of the groove and protrude in the middle of it, and thus it is now possible for two 9-mer peptides overlapping by 8 amino acids to both be binders of the same MHCI allele.

Regression trees

We use the binding motifs described above in combination with a regression tree algorithm to model the MHCI-peptide interaction. A tree is created by a series of recursive binary splits, chosen to maximize homogeneity within a node and heterogeneity between nodes. For each MHCI allele, a large number of trees are constructed composing a “random forest.” The estimated binding affinity of a peptide is obtained by averaging the prediction of the individual trees in the forest. This approach deals naturally with complicated interactions between factors and so allows flexible and interpretable models for peptide binding to MHCI. It is a useful method for characterizing the binding patterns and clarifying how the main factors identified in the analysis combine to influence MHCI-peptide interaction.

Results

We have obtained binding affinities for over 1500 peptides so far, generating a dataset comprised of peptides with a broad range of binding affinities. This has provided us the unique opportunity to use both binders and non-binders in the construction of a MHCI-peptide binding model. Due to the flexible nature of the binding motifs proposed here, this method can be further expanded to be used with MHC class II. We will illustrate the efficacy of this method using both the dataset described here as well as other publicly available MHC-peptide binding data.