Analyzing the γδ T cell receptor sequences of patients with celiac disease

Celiac disease is linked to increased intestinal proportions of T cells expressing a γδ T cell receptor (γδ TCR), which constitute only a small fraction of the overall T-cell population in healthy individuals. To elucidate the relationship these γδ TCRs have with celiac disease, we analyze sequences sampled from the blood and intestines of patients with celiac disease, patients who manage their celiac disease with a gluten-free diet, and disease-free control subjects. An array of sequence- and structure-based approaches is used to investigate what distinguishes the γδ TCRs of healthy patients from those with celiac disease. First, we characterize the diversity of the subjects' γδ TCR repertoires using the Morisita-Horn similarity index, which quantifies the overlap in the sets of γδ TCR sequences present in a pair of samples. Second, we investigate the chemical properties of the sequences in a non-position-specific manner. For example, we compare the proportions of charged, hydrophobic, aromatic, etc. residues comprising the sequences of each group. Next, we utilize a variety of methods to search for motifs, or repeated patterns, in the sequences; for instance, we implement the Smith-Waterman algorithm to locally align all possible pairs of sequences to cluster similar sequences. Finally, we analyze theoretical structures of these γδ TCRs produced by the ModWeb homology modeling server.