Dissertation Defense - Patrick Ford Combs, PhD Bioinformatics and Computational Biology
November 16, 2021 2:00 - 4:00 PM

All are invited to attend the defense. For more information please contact Graduate Coordinator at dstgerma@gmu.edu

Candidate: Patrick Ford Combs

Program: PhD, Bioinformatics and Computational Biology

Date: Tuesday, November 16, 2021

Time: 2:00 PM

Zoom Link: https://gmu.zoom.us/j/94993547493

Title: ProTeSSA: A New Method for Secondary Structure Assignment Based on Topology

Committee Chair: Dr. Iosif Vaisman

Committee Members: Dr. Dmitri Klimov, Dr. Patrick Gillevet

ABSTRACT:

Secondary structures (SSs), such as helices and strands, are local, thermodynamically-stable regions of a protein that are critical to its global structure and function. The process of evaluating a protein's three-dimensional structure and determining each residue's secondary structure classification is known as secondary structure assignment (SSA). SSA is difficult because SSs frequently deviate from their theoretical ideals. So, while there are currently over 20 SSA algorithms, each uses different sets of cutoffs and parameters leading to disagreements in assignments. A better SSA method is needed because SS is used to build protein classification databases, score homology model templates, and train SS prediction algorithms. Here, a new method, ProTeSSA (PROtein Tessellation-based SSA), is introduced. Based on the Delaunay tessellation of a protein's alpha-carbon coordinates in combination with a novel application of persistent homology, ProTeSSA machine learning models were trained using structure author(s)', DSSP, and STRIDE SSAs. These models achieved greater than 85% accuracy and low helix as strand and strand as helix misclassifications. Furthermore, the topological feature space was clustered producing a K-means classifier that performs accurate SSA without any preconceived definitions of SS or influence from other methods. These models provide a new, accurate means to perform SSA without cutoffs or parameters. And, the cluster model opens the door to new ideas about the definitions of SS. ProTeSSA is currently available for use at omics.gmu.edu/protessa.