Laying the Foundation for an Ocean Protein Portal

Authors*: Danie Kinkade, Adam Shepherd, Nick Symmonds, David Gaylord, Dawn Moran, Matthew Mcilvin, Noelle Held, Mak Saito

*Woods Hole Oceanographic Institution

The measurement of the biological inventory of proteins within an organism, known as proteomics, has emerged as an important new biological methodology within the past decade. Because proteins are involved in nutrient transport and catalysis of reactions, they can be used as biomarkers of environmental stresses and key biochemical processes. When this new technology is applied to environmental communities of microbes, typically called metaproteomics for its inclusion of a biological community, it has shown potential to significantly improve the understanding of ocean ecology and biogeochemistry by allowing a broad diagnosis of ecosystems across large geographical distances. In this manner the measurement of proteins has great potential as a tool for ocean scientists interested in detecting changes in ocean ecosystems.

Yet being a relatively new data type based in mass spectra, proteomic datasets have specific informatics complexities. Moreover, these datasets are not easily accessed by the broader oceanographic communities. The goal of the Ocean Protein Portal project is to develop a prototype that will enable expert and novice users alike, including oceanographers, molecular biologists, and biochemists to interrogate large ocean datasets to discover proteins of environmental importance and their distributions throughout the oceans. This paper will present an overview of project objectives including progress made to date on portal components such as text- and sequence-based search capability of protein datasets, and geospatial visualization and table output of the query protein’s occurrence in the oceans. In addition we will present results of a community workshop (to be held in May 2017) for ocean protein data generators and consumers designed to advance standards and best practices of protein data analyses, and provide feedback for refinement of the portal workflow.