

Data Management Plan

We will publish our results in peer-reviewed journals, and will choose open-access journals whenever possible to allow broad accessibility. Importantly, proteomic datasets are similar to genomic datasets in being dense and having great value for continued “harvesting” for additional future studies. For example, the identification and quantification of hundreds to thousands of proteins in each proteome dataset is useful for experimental design of new quantitative assays, through the identification of mass spectrometry compatible tryptic peptides (also known as Peptide Atlases). Moreover, complex proteome datasets are often useful to other researchers studying gene and proteins found in the datasets, but not mentioned in the manuscripts due to lack of space. In addition, proteomics and transcriptomic datasets are of great value in systems biology modeling studies that aim to understand cellular regulation and expression in response to environmental stimuli. Although such modeling efforts are still quite novel, in particular for environmental microbes, we are committed to making our datasets available to the oceanographic and biological/biochemical communities in the future.

We plan a multi-pronged approach to data management, in order to reach multiple audiences that will have an interest in the datasets. First, we will work with BCO-DMO to incorporate our experimental results. BCO-DMO has begun to incorporate genomic and proteomic results (e.g. metaproteomic results from the CoFeMUG expedition have recently been added to BCO-DMO) to store our compiled results. Second, we will continue to publish our proteomic datasets with each Journal’s supplemental materials that include protein and tryptic peptide identifications and relative abundance values for the global proteomes, when those journals allow supplemental materials. Third, we are creating a full archive of both proteomic results (protein and peptide identifications, relative abundances of proteins within global proteomes) as well as mass spectral libraries (“RAW files”) on our WHOI proteome laboratory website (www.whoi.edu/sites/proteomicslab). This process has already begun, with results of recently published studies available on this site. Our local website will be particularly valuable in making available the many larger mass spectra datasets for those who wish to re-search the spectra for additional protein identifications, that will benefit from the continually improving genome models and annotation (particularly for eukaryotic genomes as gene models improve with additional cDNA data input that confirm/correct the locations of difficult to predict introns).

Finally, the Department of Energy’s Joint Genome Institute’s Integrated Microbial Genomics (IMG) website has just begun to incorporate proteomic and peptide datasets into their bioinformatics capability (<http://img.jgi.doe.gov/cgi-bin/w/main.cgi>). In preparation for this proposal, we have contacted the scientists at JGI-IMG and offered our already published datasets for inclusion into their website. They have responded quite positively (see attached letter). IMG is one of the most popular microbial/phytoplankton bioinformatics interfaces at this time hosting almost 7000 bacterial, archaeal, eukaryotic, plasmid, and viral genomes. It is considered by many in microbial ecology to be the “go-to” site for examining microbial genomes. Konstantinos Liolios at IMG has encouraged us to submit the proteomes associated with multiple experimental conditions (as planned in this proposal), since “it adds significant value to the datasets”. The inclusion of our datasets in their system should provide excellent availability and visibility of the data, as well as and synergy to the existing genomic data.

This combination of using an Ocean Sciences database (BCO-DMO), our proteomics laboratory website (WHOI), published supplemental materials, and the Joint Genome Institute – Integrated Microbial Genomics portal, will provide access of our ocean acidification phytoplankton datasets to a broad community of oceanographers, mass spectrometrists, interested readers, and microbial ecologists/systems biologists, respectively.