Integrating Transcriptomics and Proteomics Data for Identifying Mechanisms Regulating the Response of Microbes to Environmental Change

H. Steven Wiley[1*] (steven.wiley@pnnl.gov)**,** Ronald C. Taylor [2], Meng L, Markillie[3], Mary S. Lipton[3], Gretta Serres[4] and Margie Romine[3]

[1]Environmental Molecular Sciences Laboratory, [2]Computational Biosciences, [3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, [4]Marine Biological Laboratory, Woods Hole, MA 02543

Project Goals: The general goal was to identify global regulatory strategies used by microbes when responding to environmental changes. By integrating simultaneous proteomics and transcriptomics measurements, we sought to determine which cellular processes were regulated at the transcriptional versus post-transcriptional level. Using *Shewanella* as an initial test case, we developed approaches for accurately quantifying the level of protein and mRNA expression, integrating these data sets and visualizing their relationships on a global level. This approach was then evaluated for its ability to identify novel regulatory mechanisms and to obtain insights into regulatory strategies used by microbes in responding to environmental changes.

There are numerous mechanisms that regulate the response of microbial communities to changes in their environment. These include changes in the community composition as well as alterations in the complement of proteins that dictate the behavior of individual cells. Most investigations on how individual cells are regulated have focused on transcriptional control mechanisms, but there are other levels that could be important, such as translational control and protein turnover. The lack of suitable technologies to screen for alternate regulatory mechanisms on a global level has made it difficult to determine how common alternate mechanisms are and the role they might play in regulating microbial responses. To address this need, we developed a general and global approach for discriminating transcriptional from post-transcriptional regulatory mechanisms. The basic approach is to simultaneously quantify and compare the levels of proteins and mRNA transcripts on a global level. Proteins whose levels change in concordance with changes in their mRNA transcripts are considered transcriptionally regulated. Those that show discordance between changes in mRNA and protein levels are potentially regulated at a post-transcriptional level.

As a test case, we used *Shewanella* grown at steady state in chemostats under either high or low oxygen conditions. Samples were collected on three different days and subjected to quantitative proteomics and transcriptomics analysis. For proteomics, quadruplicate technical replicates were run for all three sets of biological replicates and peptides identified using the AMT approach (accurate mass and time). Transcriptional profiling was performed using SOLiD sequencing with triplicate technical replicates. The proteomics data were normalized and converted to relative protein levels by combining average ion intensity measurements and peptide counts followed by corrections for protein length. The mRNA levels of each measured protein was normalized and corrected for gene length. The average protein:mRNA ratio for each gene was then calculated for each condition. To visualize changes in the protein:mRNA ratios for these extremely

large data sets, we created a “double ratio” plot where the mRNA ratio for each gene under the two treatment conditions was plotted against the ratios of their protein:mRNA ratios. These data were plotted together with information on peptide counts and ion intensity measurements using the exploratory data analysis program Aabel. This allowed us to readily identify proteins that displayed evidence of posttranscriptional regulation.

The absolute amounts of protein and mRNA expressed in *Shewanella* showed a general correlation at the level of individual genes (C.C. between 0.35-0.37). Binning the data into 25 equal groups ranked by absolute levels of gene expression showed much stronger correlation (C.C. ~0.98), suggesting that in general, the most abundant transcripts give rise to the most abundant proteins. Comparing changes of mRNA and protein expression in response to high and low oxygen conditions yielded similar results (C.C. ~0.43), thus supporting the hypothesis that the dominant mode of regulating protein levels is through transcriptional control. Nevertheless, there were several hundred proteins that showed evidence of posttranscriptional regulation. For example, the relative expression level of ribosomal proteins was essentially invariant despite changes in transcript abundance for several of them. Other proteins involved in regulating mRNA translation displayed a similar pattern. A clear case of post-transcriptional regulation was observed in the operon (*hypAEDCB*) which encodes the NiFe hydrogenase maturation complex. During high oxygen conditions, the mRNA of this operon goes down by about 2-3 fold. However, the corresponding protein levels go down 15-40 fold. Proteins from flanking operons or those that encode the NiFe hydrogenase showed no disparity between transcript and protein levels, suggesting that translation of the polycistronic message encoding HypAEDCB is specifically regulated. Preliminary pathway enrichment analysis of proteins displaying evidence for posttranscriptional regulation shows enrichment of leucine biosynthesis (p-value=1.8e-4), formate to TMAO electron transfer/anaerobic respiration (p-value=0.0023/0.013) and nitrogen metabolism (p-value=0.0085).

Our studies suggest that although most microbial proteins are regulated at the level of transcription, a significant fraction is regulated at the posttranscriptional level, particularly those involved in translation itself. This highlights the needs to include proteomics measurements in the analysis of microbial systems. The integrated approach we have developed for identifying the level at which proteins are regulated should also prove useful in the analysis of more complex microbial systems.

This research was performed in the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory (PNNL) and is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the PNNL Foundational Scientific Focus Area.