Final Summary Report for Award #ER63972

Title of Project: Identification and Characterization of Prokaryotic Regulatory Networks

Institution: Washington University

Principal Investigator: Gary D. Stormo, 314-747-5534

Project Results: We have completed our characterization of both the transcriptional regulatory network and post-transcriptional regulatory motifs in *Shewanella*.

For the transcriptional regulatory network we used both comparisons to the well studied *E. coli* regulatory interactions, and comparative analysis on the sequenced genomes of several *Shewanella* species. We identified over 200 putative DNA binding motifs that we predict control the expression of over 1000 transcription units. Nearly all of our predicted motifs are conserved in additional, newly sequenced *Shewanella* genomes that were not included in our motif-discovery process, lending support to their validity. Additionally, there are other types of supporting evidence, such as co-expression patterns and GO-term enrichment, for many of the motifs, again supporting the reliability of the predicted motifs. And by comparing the predicted regulons with expression data obtained from growth on various electron acceptors, we have predicted several regulons, and their associated motifs (and in a few cases the transcription factor) that we think are involved in responding to those conditions. This work has been published in:

[Liu J, Xu X, Stormo GD.](#) (2008) The cis-regulatory map of *Shewanella* genomes. *Nucleic Acids Res.* 36:5376-90.

For the analysis post-transcriptional regulatory motifs we first developed an improved algorithm for detection of conserved motifs that are composed of both secondary structure and sequence constraints. That method has been published in:

[Xu X, Ji Y, Stormo GD.](#) (2007) RNA Sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment. *Bioinformatics* 23:1883-91.

We then used five sequenced *Shewanella* genomes to identify significantly conserved secondary structure motifs in the 5' UTR regions of the predicted transcription units and trained a support vector machine to distinguish likely real motifs from false positives. From over 1000 transcription units we identified 166 putative regulatory motifs, many of which correspond to known regulatory elements in other species, such as attenuators, riboswitches and other known post-transcriptional regulatory motifs. Many others are conserved in the newly sequenced *Shewanella* genomes, suggesting they are valid motifs. This work has now been published in:

[Xu X, Ji Y, Stormo GD.](#) (2009) Discovering cis-regulatory RNAs in *Shewanella* genomes by Support Vector Machines. *PLoS Comput Biol.* 5:e1000338.