A Large Number of Hypothetical Proteins are Differentially Expressed during Stress in *Desulfovibrio vulgaris*

Elliott C. Drury, Alyssa M. Redding, Aindrila Mukhopadyay, Katherine H. Huang, Terry C. Hazen, Adam P. Arkin, Judy D. Wall, Dwayne A. Elias*

ESPP Environmental Stress Pathway Project

DOE GENOMICS:GTL ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT

VIMSS Virtual Institute for Microbial Stress and Survival | Miami University | University of Missouri | University of Washington | Diversa | Sandia National Laboratories | Oak Ridge National Laboratory | Berkeley | http://vimss.lbl.gov/

Abstract

Hypothetical and conserved hypothetical proteins consistently make up >30% of sequenced bacterial genomes. It is likely that many of these proteins serve significant functions ranging from regulation to presently unknown steps in carbon or electron flux.

Expression profiles for the expected 1167 hypothetical and conserved hypothetical proteins in *D. vulgaris* were obtained from VIMSS/ESPP transcriptomic and MS-based iTRAC proteomic datasets from control culture data over 10 environmental stresses. The genes were divided into two groups; those in polycistronic operons and those that are monocistronic.

We are presently able to confirm the expression at the mRNA level for 837 genes and for 247 genes at both the mRNA and protein level, while there was no evidence for either mRNA or protein detectable for 83 genes. We are testing six mutants isolated from a random transposon library. As the library sequencing continues, we will test the interrupted hypothetical and conserved hypothetical proteins either to assign a function or to confirm the putative assignments.

Expression of Hypothetical Genes

The sequenced genome of *Dv. vulgaris* suggests that there are 280 conserved hypothetical proteins and 887 hypothetical proteins. In order to attempt to assign putative functions to these proteins, we first probed all available microarray and proteomics data to eliminate the genes that had never shown expression on either level.

Table 1: Hypothetical and conserved hypothetical proteins with evidence of expression

Protein Annotation	Operon?	Genes Annotated	Micro-array	Diversa Proteomics	Diversa and iTRAC
Hypothetical	Operonic	438	409	409	70 (16%)
Conserved	Operonic	176	166	166	64 (36%)
Hypothetical	Monocistronic	449	405	405	65 (14%)
Conserved	Monocistronic	104	102	102	48 (46%)

Once those hypothetical and conserved hypothetical genes for which expression was not detected were eliminated, those showing expression were grouped according to their microarray expression profiles.

Table 2: Definitions for Gene Grouping based on Microarray Expression Profiles

Category	Definition
No Expression	No record of binding of the RNA to the oligonucleotide of the microarray in any control or stress experiments.
Normal Expression	Differential expression was not detected and was not in the top 1/8 of genes expressed.
High Expression	Differential expression was not detected but was in the top 1/8 of genes expressed.
Differential Expression to One Stress	mRNA level showed a $\log_2$ R value >= 1.5 change relative to the experimental control in greater than 20% of the time points in only one stress condition.
Differential Expression to Multiple Stresses	mRNA level showed a $\log^2$ R value >= 1.5 change relative to the experimental control in greater than 20% of the time points in more than one stress condition.

Expression Profiling of Monocistronic Hypothetical Genes

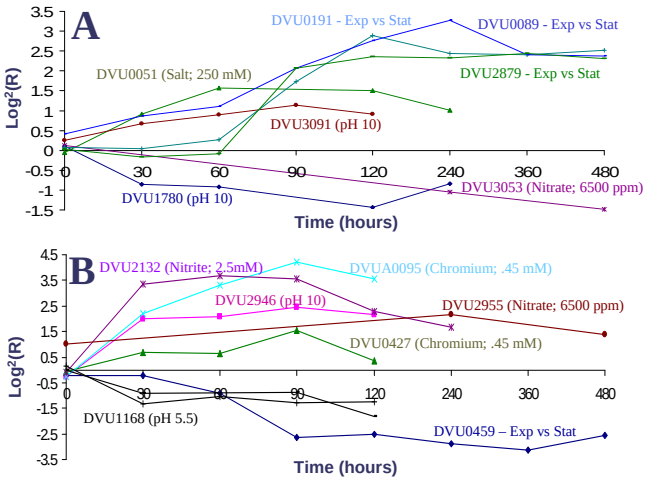

Figure 1: Microarray expression profiling of several monocistronic **(A)** conserved hypothetical and **(B)** hypothetical proteins from various stresses with differential expression in a single stress. This profiling yields clues to the function of the protein and allows for a physiologically based annotation of the gene. Exp vs Stat= Exponential growth vs stationary phase.

Expression Profiling of Operonic Hypothetical Genes

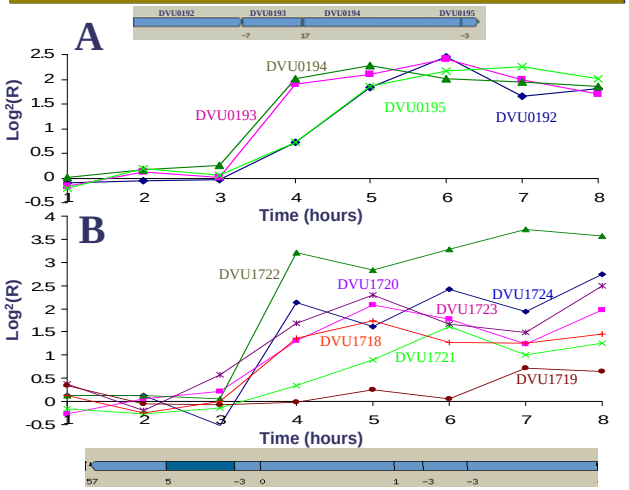

Figure 2: Microarray expression profiling of hypothetical proteins within operons allows for a putative functional assignment by using the profile and gene association. The stress shown here was exponential growth vs stationary phase. **(A)** Up-regulation of a 4 gene operon DVU0192 (adenine specific DNA methyltransferase) and DVU0194 (terminase) are well conserved while DVU0193 and DVU0195 were annotated as hypothetical proteins. **(B)** Seven gene operon containing a phage protein (DVU1724), five hypothetical proteins (DVU1718, DVU1720-1723) and a conserved hypothetical protein (DVU1719). Inserts of operon arrangement were obtained from Microbes Online (http://www.microbesonline.org/).

Use of Transposon Mutants to Assign Functions

Putative functions can be assigned to genes by amino acid sequence similarity in other organisms. However this is challenging for hypothetical and conserved hypothetical genes/proteins. While we have assigned putative functions to some such genes in operons with better-known genes, assessment of the cell physiology after deletion/interruption of these genes is the only way to positively assign function.

A library of ~5000 randomly inserted transposon mutants was constructed (courtesy J. Ringbauer) with the pRL27 plasmid generously donated by Metcalf (1) that has a high frequency transposon with kanamycin resistance. The plasmid was transformed into *D. vulgaris*, and transposition sites were obtained by gDNA sequencing for ~65 isolates. Of these, 3 are in hypothetical genes.

Targeted deletions were also constructed into *D. vulgaris* hypothetical proteins that displayed interesting expression during stress. Deletions were confirmed by Southern blot analysis and antibiotic selection, and are being physiologically assessed.

VIMSS ID	DVU	Annotation	Category	Notes
206615	DVU1175	Hypothetical protein	Differential Expression in One Stress	Monocistronic; down-regulated in high Na
208693	DVU3172	Hypothetical phosphoric diester hydrolase	Differential Expression in Multiple Stresses	Operonic; downstream of *cycA*
408334	DVU1640	Hypothetical protein	Differential Expression in Multiple Stresses	Operonic; 2nd of two genes in operon
209626	DVUA0095	Hypothetical protein; On megaplasmid	Differential Expression in Single Stress	Deletion Mutant; Monocistronic; Up-regulated in chromium
209237-209238	DVU0303-DVU0304	Hypothetical proteins	Differential Expression in Multiple Stresses	Deletion Mutant; two gene operon; dual deletion

Future Plans

1) Complete assessment of the 6 sequenced transposon mutants and targeted deletions.

2) Complete putative functional assignments to hypothetical and conserved hypothetical proteins encoded in operons based on the expression profiles and other annotated genes in the operon.

3) Attempt to assign putative functions to monocistronic hypothetical and conserved hypothetical genes based on the expression profiles.

References

1) Larsen, R. A., M. M. Wilson, A. M. Guss, and W. W. Metcalf. 2002. Genetic analysis of pigment biosynthesis in *Xanthobacter autotrophicus* Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. Arch. Microbiol. 178:193-201.

ACKNOWLEDGEMENT

ESPP is part of the Virtual Institute for Microbial Stress and Survival supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.