Progress Report (09/10/02)

"Interdisciplinary Study of *Shewanella oneidensis* MR-1's Metabolism & Metal Reduction"

This Progress Report describes the status of the study supported by Department of Energy (DOE)'s *Genomes to Life* (formerly *Microbial Cell Project*) DE-FG03-01ER63218 to PI Dr. E. Kolker at BIATECH (formerly at ISB) during the period from September 15, 2001 until August 31, 2002 (first year).

Since our project became part of the *Shewanella Federation*, we focused our work mostly on analysis of different types of data produced by global high-throughput technologies to characterize gene and protein expression as well as getting a better understanding of the cellular metabolism. After an understandable lag, the *Shewanella oneidensis* data is currently being produced by the labs participating in the *Shewanella Federation*, but as agreed upon by other members of the *Federation*, we took advantage of relevant data available during first year. Our research has been primarily concentrated on development of general data analysis approaches based on available gene & protein expression data for two similar to *Shewanella* gram-negative, gamma-subgroup microorganisms *Escherichia coli* & *Haemophilus influenzae*. The developed approaches will be transferred to analysis of the *Shewanella* data during the second year of this project. These studies are being also complemented with metabolic capability and transformation experiments with the *Shewanella oneidensis*. Specifically, first year activities include development of:

1. new labeling technique for quantitative proteomics, so called methyl esterification labeling approach, complementary to currently available methods;
2. new algorithm for *de novo* protein sequencing;
3. new statistical model for spectral analysis of arbitrary shape data;
4. one of the **first** analyses of the transcriptome of the entire microorganism;
5. new approach to predict operon structures and transcripts within untranslated regions;
6. the **first** control protein experimental mixtures with known physico-chemical characteristics for high-throughput proteomics experiments;
7. the **first** statistical models for peptide and protein identifications for high-throughput proteomics analysis;
8. one of the **first** assessments of the entire microorganism's proteome and reconstruction of its metabolism;
9. *Shewanella* metabolic capability experiments with minimal media on aerobically & anaerobically grown cells and transformation experiments.

The **status** of this study after first year was also discussed during the recent visit of the PI to DOE on July 22nd, 2002.

Several collaborations have been established within the *Shewanella Federation* with PNNL labs of James Fredrickson (Yuri Gorby & Margaret Romine) and Richard Smith (Gordon Anderson & Mary Lipton), ORNL lab of Z. Zhou (Alex Beliaev), USC lab of Kenneth Nealson (Sasha Tsapin), UCSD lab of Bernhard Palsson (Jason Papin) & Genomatica (Chris Schilling).

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

This project, supported by DOE, resulted in 6 published papers and 2 papers are being submitted (3 more are currently in preparation). These recent papers (with people supported by this project in **bold**) include:

- Goodlett DR, Keller A, Watts JD, Newitt R, Yi E, Purvine S, Eng JK, vonHaller P, Aebersold R, & **Kolker E** (2001) Differential Isotopic Labeling of Peptides for Quantitation and *de novo* Sequence Derivation. *Rapid Communications in Mass Spectrometry*, 15: 1214-21.
- **Kolker E**, **Tjaden BC**, Hubley R, Trifonov EN, & Siegel AF (2002) Spectral Analysis of Distributions: Finding Periodic Components in Eukaryotic Enzyme Length Data. *OMICS A Journal of Integrative Biology*, 6(1): 123-30.
- **Tjaden BC**, Haynor D, **Stolyar S**, Rosenow C, & **Kolker E** (2002) Identifying Operons and Untranslated Regions of Transcripts Using *Escherichia coli* RNA Expression Analysis. *Bioinformatics*, 18(S1): 337-44.
- **Tjaden BC**, Saxoni R, **Stolyar S**, Haynor D, **Kolker E**, & Rosenow C (2002) Transcriptome Analysis of *Escherichia coli* using High-Density Oligonucleotide Probe Arrays. *Nucleic Acid Res.*, 30(7): 3732-8.
- Keller A, Purvine S, **Nesvizhskii AI**, **Stolyar S**, Goodlett DR, & **Kolker E** (2002) Experimental protein mixture for validating tandem mass spectrometry analysis. *OMICS A Journal of Integrative Biology*, 6(2): 207-12.
- Keller A, **Nesvizhskii AI**, **Kolker E**, & Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry*, Sept. issue, in press.
- **Kolker E**, **Stolyar S**, Goodlett DR, Purvine S, Yi E, **Nesvizhskii AI**, Xie T, Keller A, Eng JK, **Picone AF**, Siegel AF, Hood L, Reilly TJ, Makarova KS, Palsson BO, & Smith AL (2002) Metabolic Reconstruction of *Haemophilus influenzae* Rd KW20 using Protein Expression Profiling, subm.
- **Nesvizhskii AI**, Keller A, Aebersold R, & **Kolker E** (2002) Estimating the accuracy of protein identifications made by MS/MS and database search, subm.