DOE/ER/20319-1

Final Report

DOE Grant #: DE-FG02-98ER20319
Title: High Throughput Technologies for Functional Analysis of Archaeal Genomes
Principal Investigator: Najib M.A. El-Sayed, Ph.D., The Institute for Genomic Research

I- Specific Aims

The proposal focused on the development of high-density microarrays for two methanogenic archaea, *Methanoccus jannaschii* and *Methanobacterium thermoautotrophicum.*

The specific aims of this project were as follows:

1) To design primers to each predicted open reading frame (ORF) in M. jannaschii and *M. thermoautotrophicum* to allow the amplification of a unique target sequence that will represent the corresponding coding region on a complete genome chip.
2) To amplify each target sequence from *M. jannaschii* and *M. thermoautotrophicum* and verify that these PCR products are the expected DNA fragment.
3) To establish a relational database that will track the production of target DNAs and the nucleotide sequence used to represent each ORF .

II- Results

Oligonucleotide primers for coding regions in both genomes were designed using Primer 3.0 (Whitehead Institute), an algorithm that searches for good candidate regions for priming by minimizing stem-loop structures and self-annealing. After optimizing the PCR process (Amount of DNA template, oligo concentration, source of Taq polymerase, and tubes used), nearly all 1869 ORFs of *M. thermoautotrophicum* and 1737 ORFs from *M. jannaschii* were amplified (success rates: 99.2% and 98.6%, respectively). Genomic DNAs that served as templates for reactions were at a final concentration of 5-50 pg/μl. A portion of each reaction (5 μl) was analyzed by agarose gel electrophoresis and visualized by Ethidium bromide. PCR products were evaluated with respect to yield and expected size, and scored using a PCR scoring tool developed at TIGR to allow data to be easily viewed, entered into the microarray database and tracked. About 3% of the PCR products were audited by direct sequencing to further confirm that the amplified DNA is the correct sequence. All the amplicons for each genome have been printed on glass slides.

Obtaining the maximal benefit from microarray analysis requires the development of databases capable of effectively capturing the data as well as tools to make that data accessible to the laboratory scientist. A relational database has been created specifically to deal with expression analyses in microbial organisms. The microarray database allows the tracking of the all materials necessary for construction of the array, the samples to be studied by hybridization, and the resulting data that are generated. Tables exist to allow for the storage of information regarding primers, PCR products, arrayed slides, samples and probes used to interrogate the arrays, etc... The relational nature of the underlying database allows links to be made associating expression levels directly to the underlying annotated genes and their functional categories.

III- Validation of microarrays – steps beyond the original goals

As both genome chips are now completed, we have initiated expression studies with collaborators: Dr. John Reeve (Department of Microbiology, University of Ohio) and Dr Douglas S. Clark from the Department of Chemical Engineering, University of California, Berkeley.

To examine the feasibility of using microarrays for expression analysis in archaea, we are carrying out a pilot study in collaboration with Dr. Reeve to examine hydrogen-dependent regulation of methane gene transcription in *M. thermoautotrophicum* by essentially reproducing a detailed Northern analysis reported by Reeve *et al.* (J. Bact. 179: 5975-5986). Toward that purpose, we have constructed, in addition to the whole genome arrays, an additional subarray containing a subset of 80 ORFs encoding proteins involved in hydrogen metabolism and methanogenesis in *M. thermoautotrophicum*. The *M. thermoautotrophicum* arrays were hybridized with two populations of RNA derived from cells growing under different conditions (Fig. 1) and obtained from John Reeve's laboratory.

DISCLAIMER

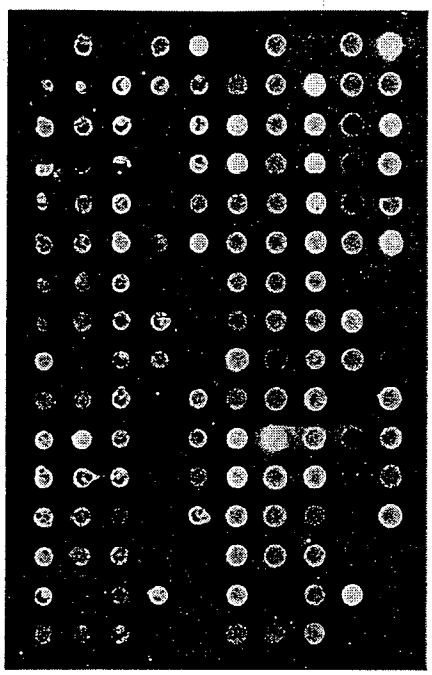

Figure 1. Photograph of the hybridization of cy3 (red) and Cy5 (green) labeled first strand cDNA probes to a test microarray. The red probe is derived from *M. thermoautotrophicum* total RNA under conditions of high H2 availability and the green probe from total RNA under conditions of low H2 availability. Eighty *M.t.* methanogenesis ORFS were PCR-amplified, purified and printed onto a SuperAmine™ (TeleChem) substrate in 3x SSC (columns 1-5) or 60% DMSO (columns 6-10). The printed microarrays were processed to remove unbound material, chemically couple the DNA to the SuperAmine surface and block unreacted amine groups to reduce non-specific binding of RNA (The DNA was cross-linked using a UV Stratalinker (Stratagene, La Jolla, CA) at 120 mjoules/cm^2. Following a rinse in 0.1% SDS for 2 x 5' at R.T. to remove unbound DNA, the unreacted amine groups on the slide were removed using a succinic anhydride/1-methyl-2-pyrrolidinone/boric acid solution, then rinsed with dH_2O at R.T. The DNA was denatured by immersing the slides in dH_2O at 95°C for 2 min). The printed microarray was reacted with the labelled probes in 5X SSC and 0.1% SDS and was performed at 55°C. Expression was assayed by scanning the arrays using a laser confocal scanner.

Another ongoing collaboration with the research group of D.S. Clark at U.C. Berkeley focuses on a microarray analysis of *M. jannaschii* cultivated at extreme temperatures and pressures, a research interest of the Clark laboratory for over 10 years. A high temperature-pressure bioreactor in the Clark laboratory is capable of continuous operation at temperatures and pressures up to 200°C and 550 atm, respectively. This unique capability is enabling TIGR and the Clark lab to analyze the physiology of extremophiles in relation to temperature and pressure, probe the environmental limits of microbial life, and explore the production of novel metabolites and natural products *in extremis.*