Final Report

Real-time Gene Expression Profiling of Single Live Cells of *Shewanella oneidensis*

PI: Xiaoliang Sunney Xie

Summary

The overall objective of this proposal is to make real-time observations of gene expression in live *Shewanella oneidensis* cells with high sensitivity and high throughput. Gene expression, a central process to all life, is stochastic because most genes often exist in one or two copies per cell. Although the central dogma of molecular biology has been proven beyond doubt, due to insufficient sensitivity, stochastic protein production has not been visualized in real time in an individual cell at the single-molecule level. We report the first direct observation of single protein molecules as they are generated, one at a time in a single live *E. coli* cell, yielding quantitative information about gene expression [*Science* 2006; 311: 1600-1603].

We demonstrated a general strategy for live-cell single-molecule measurements: detection by localization. It is difficult to detect single fluorescence protein molecules inside cytoplasm - their fluorescence is spread by fast diffusion to the entire cell and overwhelmed by the strong autofluorescence. We achieved single-molecule sensitivity by immobilizing the fluorescence protein on the cell membrane, where the diffusion is much slowed. We learned that under the repressed condition protein molecules are produced in bursts, with each burst originating from a stochastically-transcribed single messenger RNA molecule, and that protein copy numbers in the bursts follow a geometric distribution.

We also simultaneously published a paper reporting a different method using β -galactosidase as a reporter [*Nature* **440**, 358 (2006)]. Many important proteins are expressed at low levels, inaccessible by previous proteomic techniques. Both papers allowed quantification of protein expression with unprecedented sensitivity and received overwhelming acclaim from the scientific community. The Nature paper has been identified as one of the most-cited papers in the past year [<http://esi-topics.com/>].

We have also an analytical framework describing the steady-state distribution of protein concentration in live cells, considering that protein production occurs in random bursts with an exponentially distributed number of molecules. This model allows for the extraction of kinetic parameters of gene expression from steady-state distributions of protein concentration in a cell population, which are available from single cell data obtained by fluorescence microscopy. [*Phys. Rev. Lett.* **97**, 168302 (2006)].

A major objective in the Genome to Life (GtL) program is to monitor and understand the gene expression profile of a complete bacterial genome. We developed genetic and imaging methods for sensitive protein expression profiling in individual *S. oneidensis* cell. We have made good progress in constructing YFP-library with several hundred chromosomal fusion proteins and studied protein expression profiling in living *Shewanella oneidensis* cells. Fluorescence microscopy revealed the average abundance of specific proteins, as well as their noise in gene

expression level across a population. We also explored ways to adapt our fluorescence measurement for other growth conditions, such as anaerobic growth.

Final Report

Tremendous progress has been made to track gene expression at the mRNA level by DNA arrays and at the protein level by mass spectrometry. Although current DNA microarray and mass spectrometry technologies have started to address compelling biological problems at a genome-wide scale, they suffer from a few disadvantages: (1) They cannot continuously monitor temporal evolution of expression; multiple samples have to be taken in order to evaluate the response to a stimulus or an environmental change. (2) They cannot follow fast gene expression processes on the time scale of minutes. The low time resolutions prevent studies of transient gene expression processes, for example, those involved in cell division. (3) They are not sensitive to low copy number gene products, which often play a prominent role in cellular sensing, signaling and gene regulation. (4) They can only provide averaged results of large populations of cells rather than behaviors of individual cells. Because of the lack of synchronization among single cells, transient and stochastic gene expression events are often masked in the population measurements.

For all these reasons, techniques that continuously monitor gene expression dynamics in live cells are in high demand. In this project, we take a systems biology approach and embark on high-throughput and real-time detection of low and high copy number of gene products in individual live bacterial cells.

Developing assays for real time observation of protein expression in living bacteria

Gene expression is a central process to all lives. At the individual cell level, gene expression can often be viewed as a single molecule process, because many genes exist in one or a few copies per cell.

Taking the *lac* operon of *E. coli* for example, transcription by RNA polymerase is initiated upon a stochastic dissociation event of the repressor from the operator region of DNA, generating one mRNA molecule. A burst of fusion protein molecules will be synthesized by multiple ribosome molecules bound to the mRNA, yielding fluctuating protein production in time (Fig. 1A). In order to observe these processes in a cell, methods of protein detection with extremely high sensitivity is required.

We have developed two methodologies for probing gene expression in living cells with single protein molecule sensitivity.

In the first approach, we use a fast maturing fluorescent protein called Venus as a gene expression reporter. The key for achieving single molecule sensitivity was to immobilize the fluorescent protein reporter on the cell membrane, by constructing a chimeric fluorescent protein reporter *tsr-venus*, which contains a membrane localization sequence. It is normally difficult to detect single protein molecules inside cytoplasm - their fluorescence is spread by fast diffusion to

the entire cell during the image acquisition time, and overwhelmed by the strong cellular autofluorescence. However molecules on cell membrane diffuse much slower, and therefore can be detected individually with our sensitive microscope. Using this approach, we recorded [movies](#) of growing *E. coli* cells to study the real-time expression from the *lac* operon in repressed state in real-time (Fig. 1B).

Figure 1. (a) Experimental design for live-cell observations of gene expression. Tsr-Venus is expressed under the control of *lac* repressor, which binds tightly to the *lac* operator on DNA. Transcription of one mRNA by an RNA polymerase results from an infrequent and transient dissociation event of repressor from DNA. Multiple copies of protein molecules are translated from the mRNA by ribosomes. Upon being assembled into *E. coli*'s inner membrane, Tsr-Venus protein molecules can be detected individually by a fluorescence microscope. (b) Real-time monitoring of the expression of *tsr-venus* under the control of repressed *lac* promoter. A sequence of fluorescent images (yellow) overlaid with simultaneous DIC images (gray) of *E. coli* cells expressing single-molecules of Tsr-Venus on agarose gel pad of M9 medium is shown.

Another approach is based on the enzymatic reporter β -galactosidase (Fig. 2). A single copy of β -galactosidase generates many copies of fluorescence molecules when a cell trapped in a microfluidic device was treated with a fluorogenic substrate, resulting enzymatic amplification in signal and single molecule sensitivity. This technique has been applied to probe gene expression from *E. coli* as well as individual budding yeast and mouse embryonic stem cells, demonstrating its generality.

Figure 2. (a) Schematic diagram of the two-layer microfluidic chamber used for the enzymatic assay. Cells are trapped inside a volume of 100 μ l chamber. (b) Enzymatic reaction: hydrolysis of the synthetic substrate FDG by the β -gal yields a fluorescent product, fluorescein. (c) Real-time detection of β -gal production in a living cell. Discrete jumps in β -gal number are due to burst-like production of proteins. (d) Histogram of copy number of β -gal molecules per burst. The black line is the single exponential fit to the histogram

We monitor the expression of our reporters Tsr-YFP and the native β -galactosidase, respectively, from the repressed lac promoter on the chromosomal DNA. With both reporter systems, we observe that reporter protein production occurs in bursts, with the number of molecules per burst following an exponential distribution. These observations were predicted only theoretical previously, accounted for by the competition between mRNA degradation by nuclease and translation by ribosome. Our single-molecule experiment demonstrates the ability to provide quantitative real-time information on gene expression in a live cell. We have also developed theory to explain the new data.

These works on *E. coli* were major breakthroughs that stimulated a great deal of interest in the scientific community. News articles about this work can be found in this website <http://bernstein.harvard.edu/pages/NewsReports.html>.

Single-cell protein expression profiling in *Shewanella oneidensis*

We developed tools for protein expression profiling in living *Shewanella oneidensis* cells with high sensitivity. This required the construction of a chromosomal fluorescent protein fusion library followed by analysis with fluorescence microscopy. Our reporter library offers three advantages. First, chromosome integration ensures that all native promoter elements, known or

unknown, are intact. Second, a reporter fusion allows exact one-to-one quantitation with the target protein and measurements of post-transcriptional regulation. Third, the specific choice of a yellow fluorescent protein (YFP) as a reporter allows visualization of localization and even more importantly, the potential sensitivity for single-molecule imaging in individual cells.

Reporter library construction

Construction of the reporter library presented a challenge due to the inefficient genetic methods in *S. oneidensis*. However, we developed a cloning strategy to construct a library of about 250 chromosomal YFP fusions at the C-terminus of the target protein. We also inserted the YFP gene into the transcript of several stable RNAs. The complete list of constructed reporter strains appears in Table 1.

To facilitate rapid chromosome integrations, the plasmid pVS148.1 was constructed from the conjugable suicide vector pMK2017 by replacing the tetracycline resistance gene with a gentamycin resistance gene and fusing the YFP variant Venus behind the Gateway destination cassette. Homologous sequences of 500 to 1000 bp of the target locus were PCR amplified and inserted into the entry vector pMK2010 with BP clonase before transfer to pVS148.1 with LR clonase. The resulting suicide vector was transformed into the *dap* auxotroph WM3064 containing the pRK2013 conjugation helper plasmid, followed by conjugation into *S. oneidensis* MR-1 and selection on gentamycin LB plates to generate the reporter mutant (Figure 3). Resulting clones were screened by locus specific PCR and positive mutants were subsequently sequenced.

Imaging of protein expression levels and noise

We imaged the single-cell protein expression profile for a subset of the reporter strains. After aerobic growth, cell samples were pelleted by centrifugation, resuspended in 1-2 microliters, and spread between two glass coverslips. Samples were imaged with a 100x NA 1.35 oil-immersion objective using a liquid nitrogen-cooled CCD camera with 514 nm fluorescence excitation. Phase contrast and fluorescence images were taken at each location. A custom image analysis program identified cell boundaries from the phase contrast image and calculated the corresponding fluorescence intensity per cell, normalized by cell size.

The average expression level per cell, as well as the variation in expression level across a population, for a sample data set appears in Table 2. We observed that most genes are expressed at low levels, although some examples, such as the RNA polymerase alpha subunit, were expressed at high levels, as expected. We also observed the expression of hypothetical proteins with unknown functions. One interesting example was, SO0521, which has an unknown function but is expressed at levels higher than even RNA polymerase alpha subunit.

Different genes exhibit different levels of expression noise, as measured by the variation across the population, normalized by the mean expression. The noisiest genes in the measured set were the RNA polymerase alpha subunit, a sigma factor, and a hypothetical protein. The least noisy genes included an isomerase, peptidase, and pyrophosphorylase. Our imaging method for

measuring expression noise with sensitive microscopy is applicable to other biological systems as well.

Development of *S. oneidensis* imaging techniques

Imaging of YFP in *S. oneidensis* presented several unique challenges. Because the metabolism of *S. oneidensis* is of particular interest, the imaging techniques must be adapted for different media and growth conditions. For sensitive imaging, though, a key requirement is growth in media with low autofluorescence while maintaining desirable properties of cell metabolism. We found that modified M1 medium supplemented with 30 mM lactate as a carbon source and 0.05 % yeast extract provided healthy growth at 30 degrees Celsius with very low autofluorescence. The carbon source could be altered without affecting autofluorescent properties.

Another aspect of *S. oneidensis* imaging is the applicability of YFP expression profiling under anaerobic growth conditions. Fluorescent proteins require oxygen to form mature, fluorescent chromophores. Under anaerobic growth, we found that cells exhibited virtually no fluorescence, even if YFP was translated. However, we developed a technique for measuring YFP expression level under this condition. Cells are translationally arrested with chloramphenicol, removed from anaerobic growth, and incubated aerobically with shaking at room temperature for one hour. After this treatment, we could again quantitate YFP expression levels for fluorescence even for cells grown under anaerobic conditions. The use of a translation inhibitor minimized changes in protein expression once cells were removed from anaerobic growth.

Conclusions

Recent developments on fluorescent proteins and microscopy techniques have allowed the probing of single molecules in a living bacterial cell with high specificity, millisecond time resolution, and nanometer spatial precision. The DOE/GtL project allowed us to develop the key methods of single-molecule detection and to demonstrate a few examples to illustrate how these experiments are contributing to the quantitative understanding of the fundamental processes in a living cell.

We developed genetic and imaging methods for sensitive protein expression profiling in individual *S. oneidensis* cell. We constructed a YFP reporter library with several hundred chromosomal fusion proteins. Fluorescence microscopy revealed the average abundance of specific proteins, as well as their noise in gene expression level across a population. We also explored ways to adapt our fluorescence measurement for other growth conditions, such as anaerobic growth.

Figure 3. Chromosome-integrated YFP-fusion reporter mutants. A suicide vector generated using the Gateway method integrates into the chromosome of *S. oneidensis*. The destination vector generates YFP fusions to the C-terminus of the desired protein under control of the native promoter. An origin of transfer facilitates conjugation into the target strain, while the R6K origin prevents replication of the plasmid in the host in the absence of the *pir* gene.

Table 1: List of constructed YFP chromosomal fusions.

SO0082	SO0305	SO0414	SO0480	SO0551	SO4105
SO0085	SO0309	SO0415	SO0481	SO0552	SO4142
SO0094	SO0311	SO0416	SO0482	SO0553	SO4207
SO0140	SO0313	SO0417	SO0484	SO0555	SO4360
SO0142	SO0314	SO0419	SO0486	SO0556	SO4404
SO0145	SO0318	SO0420	SO0487	SO0557	SO4484
SO0146	SO0320	SO0420.2	SO0488	SO0576	SO4485
SO0147	SO0323	SO0421	SO0490	SO0610	SO4570
SO0151	SO0324	SO0422	SO0491	SO0716	SO4572
SO0159	SO0326	SO0423	SO0492	SO0717	SO4591
SO0160	SO0331	SO0425	SO0494	SO0809	SO4606
SO0162	SO0332	SO0426	SO0495	SO0845	SO4746
SO0163	SO0332	SO0428	SO0496	SO0958	
SO0173	SO0335	SO0429	SO0497	SO0r005	
SO0177	SO0336	SO0431	SO0499	SO0r006	
SO0182	SO0337	SO0432	SO0500	SO0r009	
SO0186	SO0338	SO0433	SO0501	SO0t003	
SO0190	SO0342	SO0435	SO0502	SO0t004	
SO0192	SO0345	SO0437	SO0504	SO1415	
SO0198.1	SO0346	SO0438	SO0508	SO1427	
SO0201	SO0347	SO0439	SO0508.1	SO1659	
SO0203	SO0351	SO0440	SO0510	SO1659	
SO0204	SO0353	SO0441	SO0513	SO1777	
SO0205	SO0354	SO0442	SO0515	SO1778	
SO0214	SO0355	SO0444	SO0516	SO1779	
SO0217	SO0360	SO0445	SO0519	SO1780	
SO0219	SO0361	SO0449	SO0520	SO1782	
SO0222	SO0362	SO0450	SO0521	SO2178	
SO0223	SO0364	SO0451	SO0522	SO2201	
SO0226	SO0365	SO0452	SO0523	SO2361	
SO0229	SO0366	SO0453	SO0525	SO2363	
SO0252	SO0367	SO0454	SO0526	SO2727	
SO0256	SO0375	SO0455	SO0528	SO2913	
SO0258	SO0376	SO0456	SO0529	SO2930	
SO0259	SO0377	SO0458	SO0530	SO2931	
SO0264	SO0379	SO0459	SO0536	SO3056	
SO0269	SO0381	SO0462	SO0539	SO3102	
SO0273	SO0390	SO0464	SO0540	SO3141	
SO0277	SO0391	SO0466	SO0541	SO3237	
SO0278	SO0393	SO04666	SO0542	SO3420	
SO0288	SO0396	SO0467	SO0543	SO3432	
SO0290	SO0400	SO0468	SO0545	SO3471	
SO0294	SO0402	SO0471	SO0546	SO3623	
SO0301	SO0406	SO0474	SO0548	SO3667	
SO0301.1	SO0409	SO0476	SO0549	SO3980	
SO0304	SO0410	SO0479	SO0550	SO4048	

Table 2. Single-cell expression profiling of *Shewanella oneidensis*. Units are arbitrary fluorescence units.

ORF	Mean (μ)	Variance (σ^2)	Noise (σ^2/μ^2)	Gene function
SO0146	25	74	0.12	transposase
SO0252	42	819	0.46	ribosome processing
SO0256	200	59702	1.49	RNA polymerase subunit
SO0309	16	49	0.20	unknown
SO0324	54	427	0.15	unknown
SO0346	104	5729	0.53	transcription factor
SO0353	24	60	0.11	unknown
SO0409	55	783	0.26	unknown
SO0419	13	14	0.09	pyrophosphorylase
SO0422	29	703	0.85	biosynthesis
SO0423	57	1151	0.35	transcription factor
SO0425	58	4190	1.26	dehydrogenase
SO0431	83	1072	0.15	phosphatase
SO0433	114	15255	1.17	sigma factor
SO0453	15	8	0.03	isomerase
SO0458	17	84	0.31	unknown
SO0471	222	36359	0.74	dioxygenase
SO0480	18	29	0.09	unknown
SO0486	13	9	0.06	copper transporter
SO0487	21	50	0.12	copper transporter
SO0488	14	45	0.24	copper transporter
SO0490	18	25	0.08	transcription factor
SO0491	34	105	0.09	peptidase
SO0492	39	575	0.38	oxygenase
SO0495	21	44	0.10	transposase
SO0497	27	105	0.14	unknown
SO0501	54	1240	0.42	unknown
SO0508	34	289	0.25	unknown
SO0513	13	18	0.11	unknown
SO0515	12	11	0.08	unknown
SO0521	1207	1028399	0.71	unknown
SO0526	33	518	0.49	acetyltransferase
SO0536	24	209	0.37	redox activity
SO0539	16	36	0.14	transporter
SO4746	45	465	0.23	ATP synthase

Publications resulting from the grant:

Cai, Long; Friedman, Nir; Xie, X. Sunney "Stochastic protein expression in individual cells at the single molecule level," *Nature*, **440**, 358-362 (2006).

Yu, Ji; Xiao, Jie; Ren, Xiaojia; Lao, Kaiqin; Xie, X. Sunney; "Probing Gene Expression in Live Cells, One Protein Molecule at a Time," *Science*, **311**, 1600-1603 (2006).

Friedman, Nir; Cai, Long; Xie, X. Sunney "Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression," *Phys. Rev. Lett.* **97**, 168302 (2006).

Xie, X. Sunney; Choi, Paul J.; Li, Gene-Wei; Lee, Nam Ki; Lia, Giuseppe "Single-Molecule Approach to Molecular Biology in Living Bacterial Cells," *Annu. Rev. Biophys*, **37**, 417-444 (2008).