Transcriptome Analysis of *Bacillus subtilis* by DNA Microarray Technique

Choong-Min Kang[1,2], Ken-ichi Yoshida[2], Masayuki Matsunaga[2], Kazuo Kobayashi[3], Minoru Ueda[4], Naotake Ogasawara[3], and Yasutaro Fujita[2]

[1]Department of Food Science, Youngdong University, [2]Department of Biotechnology, Fukuyama University, [3]Graduate School of Biological Science, Nara Institute of Science and Technology (NAIST), [4]Biomedical Group, Takara Shuzo Co.

Abstract

The complete genome sequence of a Gram-positive bacterium *Bacillus subtilis* has recently been reported and it is now clear that more than 50% of its ORFs have no known function (1). To study the global gene expression in *B. subtilis* at single gene resolution, we have tested the glass DNA microarrays in a step-wise fashion. As a preliminary experiment, we have created arrays of PCR products for 14 ORF whose transcription patterns have been well established through transcriptional mapping analysis. We measured changes in mRNA transcript levels between early exponential and stationary phase by hybridizing fluorescently labeled cDNA (with Cy3-UTP and Cy5-UTP) onto the array. We then compared the microarray data to confirm that the transcription patterns of these genes are well consistent with the known Northern analysis data. Since the preliminary test has been successful, we scaled up the experiments to ~94% of the 4,100 annotated ORFs for the complete genome sequence of *B. subtilis*. Using this whole genomic microarray, we searched genes that are catabolite-repressive and those that are under the control of σ^{Y}, one of the functionally unknown ECF sigma factors. From these results, we here report that we have established DNA microarray techniques that are applicable for the whole genome of *B. subtilis*.

Introduction

For biologists, the biggest and probably the most important question is 'what is life'? As one of the critical steps to address this question, we humans have launched the projects for sequencing the genomes of various organisms from bacteria to human. Until now, complete genome sequences of more than 17 organisms, including a Gram-positive bacterium *Bacillus subtilis* have been produced. The next and more challenging task is how to characterize the biological functions of the newly found genes from the genome sequencing projects. From the international genome sequencing project, we now know that *B. subtilis* has 4,100 open reading frames (1). However, we also realized that the function, expression and regulation of more than 50% of those genes remained unknown. Then, how can we characterize or identify the unknown genes?

In *B. subtilis*, we have been searching for a wide variety of sequence motifs which encodes structural domains, such as DNA-binding or ATP-binding domain to get the clues for the functions of the genes. This approach has been successful and more than 100 genes have been identified in this way. People have also mutagenized each unknown gene using the same strain and protocol, and these approach has also given us important ideas for their functions. Another approach to study the function of a gene is to determine its pattern of expression. In fact, many *Bacillus* laboratories have already conducted Northern analysis and now for more than 700kb regions of *B. subtilis* genome, transcriptional mapping has been finished.

Even though these approaches have provided us important information on the function and expression of unknown genes, they are dealing with only single gene or genes in the same operon at a time. To study the expression pattern of many genes more systematically, DNA microarray technique recently has been developed, with which we can analyse the expression of thousands of genes in a single hybridization experiment (2).

Method: cDNA Microarray technology

DNA microarray technique usually involves 4 major steps; 1) preparation of target cDNA, 2) preparation of probe DNA on the microarray, 3) hybridization, and 4) scanning and computer analysis (3). First, for the preparation of target cDNA, RNA has to be extracted from both the control and experimental cells. For example, with or without stress or exponential cells and stationary cell, and so on. Total RNA from each sample is fluorescently labeled with either Cy3-UTP or Cy5-UTP to produce tagged cDNA in a single round of reverse transcription reaction. Cy3 and Cy5 have such distinctive excitement and emission wavelengths that you can have two separate signals from the same array. In addition, to directly compare relative message abundance between the two samples, the unrelated RNA sample such as human RNA is added as an internal control.

Two individually labeled cDNA samples are then pooled and hybridized to the probe DNA on the microarray. After the overnight hybridization, the array is washed to remove unhybridized materials and subsequently optically scanned. Then resulting data are analyzed in the computer to produce gene-pie or bar graphs.

Using this cDNA microarray technique, we asked whether we can adopt it to analyze the expression pattern of *B. subtilis* genome. To address this, we have evaluate the DNA microarray technique in a step-wise fashion. First, we selected 4 different groups of genes from the *gntZ-ywaA* genomic region. In fact, this is the region that has been sequenced, mutagenized, and transcription-mapped in Dr. Fujita's lab. Among 155 genes in this region, we have chosen 14 genes that showed different transcription patterns between exponential and stationary phase. That is, in Northern analysis, *yxaI* and *yxcE* showed transcription at exponential phase but not much at stationary phase. In contrast, five genes including *asnH*, *yxbA*, *hutM*, *yxjC*, and *yxjF* were

transcribed mainly at stationary phase. *iolF*, *pepT*, *galE* and *msmX* genes were transcribed at both exponential and stationary phases, and the other three genes (*yxaG*, *yxdM*, and *yxeK*) did not show any transcription at either exponential or stationary phase.

With these Northern data in hand, we examined whether microarray technique can provide similar expression pattern for the chosen 14 genes. To do that, we first purified RNA from cells harvested at exponential and stationary phase. 100 μg of each RNA was reverse-transcribed to make labeled cDNA with Cy3 or Cy5 using a mixture of specific reverse primers for the 14 genes. Then the labeled target cDNAs were hybridized to an array that has been spotted with PCRed DNA of the 14 genes. After washing, the plate was scanned for Cy3 and Cy5 by GMS 418 Scanner and the collected data has been analysed by the ImaGene software (3).

Results

Microarray data for the selected 14 genes are well consistent with the known Northern data

In the presentation, I will present microarray analysis data for the 14 genes that shows very similar results with the known Northern analysis data. We used *rrnA* as a negative control by not adding its specific primer in the labeling reaction. In addition, we used human TRF gene as a positive control by adding TRF RNA and its primer to the labeling reaction.

Similar patterns of gene expression for the 14 genes are observed even when more primers (448 and 4,100 primers) are used.

Since the preliminary experiments with the small number of genes (14 genes) has been successful, we wanted to increase the number of primers in larger scale to see if the signals for the 14 genes were specific. For this, we used 448

primers, which included the selected 14 genes as well as other 434 genes. From this analysis, we could observe the similar pattern of expressions of the 14 genes, which strongly suggested that the previous signals from the experiment with 14 primers were not nonspecific. We also conducted similar experiments with the whole 4,100 primers and we confirmed that DNA microarray technique can provide very specific expression patterns for the chosen 14 genes.

Using DNA microarray technique, we were able to find candidates for σ^Y*-dependent genes and catabolite-repressive genes*

To test whether we can apply this newly developed DAN microarray technique for other systems in *B. subtilis*, we chose two model systems to search for 1) genes which are under the control of a ECF sigma factor σ^Y and 2) genes that are catabolite-repressive.

To search for σ^Y regulon, we compared the transcriptome level between *B. subtilis* cells with overexpressed σ^Y and the ones without overexpressed σ^Y. For the overexpression of σ^Y in vivo, we subcloned *sigY* structural gene under the control of IPTG-inducible P_{spac} promoter in pDG148 plasmid (4). After introducing this recombinant plasmid into *B. subtilis* cell, two kinds of total RNAs, one from cells grown with IPTG and the other without IPTG, were purified. These RNAs were then used for cDNA labeling and hybridization as described earlier. Here, I will show you that we have found some possible σ^Y-dependent genes, which includes the genes in *sigY* operon itself. Since we can have indirect clues for the physiological functions of σ^Y by studying the putative σ^Y-dependent genes, we are going to conduct genetic studied with these newly found genes, such as construction of deletion mutants and searching for phenotypes.

We have also done the similar experiment to look for catabolite repressive genes using RNA samples from cells grown in the presence of glucose and the ones without glucose. I will present the results of this experiment.

Future perspective

Since we now know that we can use DNA microarray technique for transcriptional analysis in *the* genome scale of *B. subtilis*, we now want to establish a transcriptional regulation network of *B. subtilis* through the technology. The main regulatory components of transcription in bacteria are sigma factors, response regulators of the two-component systems and the regulatory proteins that have helix-turn-helix domain. From the *B. subtilis* genome sequencing project, we have found 19 sigma factors, 34 response regulators, and 200 possible HTH regulatory proteins. Therefore, to understand the transcriptional regulation of *B. subtilis* more comprehensively, we want to establish the transcriptional regulation network by studying each of the above regulatory component through the DNA microarray technology developed in our lab. To do so, we are going to either overexpress or disrupt genes of those putative regulatory components.

References

1. Kunst. F. *et al.*, Nature 1997. **390**:249-256
2. Review of Microarray technology. Nature Genetics **21**:1-56
3. Stanley D. Rose. 1998. Journal of the Association for Laboratory Automation **3**:53-56
4. Stragier, P., C. Bonamy, and C. Karmazyn-Campelli. 1988. Processing of a sporulation sigma factor in Bacillus subtilis: How morphological structure could control gene expression. Cell **52**:697-704