Protein and mRNA Quantitation

DOI: 10.1002/anie.201005969

Genome-wide Correlation between mRNA and Protein in a Single Cell**

*Edward S. Yeung**

gene expression · mRNA · proteins · single cells · single-molecule studies

Since the completion of the Human Genome Project a decade ago, much progress has been made towards determining and understanding the transcriptome and the proteome. After all, DNA is no more than a template which is systematically converted into mRNA which controls protein synthesis and life itself. A major difficulty when studying transcriptome and sproteome is that gene expression in each cell is unique. Even within a given tissue, age, cell cycle, environment, signal transduction, and other stochastic events all contribute to this diversity. On the other hand, the sensitivity of standard analytical methods typically requires a large number of cells for the comprehensive quantification of mRNAs or proteins. Since there is no such thing as an "average" cell, our understanding of the biological system is incomplete. Advances in transcriptomics and proteomics at the level of single cells are therefore badly needed.

The determination of nucleic acids and proteins inside single cells dates back half a century. The total nucleic acid content was measured by using absorption photometry,[1] and various hemoglobins, the major proteins in red blood cells, were separated and detected from a single cell by using fiber electrophoresis[2] and later by capillary electrophoresis.[3] However, detecting just the major components of either type is far from the comprehensive "omics" that defines the cell. Enzymes are highly specific and efficient biochemical catalysts, so many proteins present at only a few copies per cell can still be critical to biological function. A corollary is that only small amounts of each type of mRNA are needed for translation. The challenge is to identify and quantify each type of mRNA and each protein within a cell regardless of whether they are the major components or exist in only trace amounts.

An intermediate solution towards comprehensive proteomics in single cells is that of protein profiling.[4] The information content can be extremely high, as has been demonstrated by conventional two-dimensional (2D) gel electrophoresis. To cope with the small amounts present in individual cells, protein separation is achieved by capillary electrophoresis and sensitive detection is possible by laser-induced fluorescence. There are still difficulties in implementing the same two separation modes (isoelectric focusing and SDS polyacrylamide sizing) in the capillary format, particularly since the proteins must be fluorescently labeled for detection, so identification and correlation with traditional 2D gel data bases remain elusive. The separation efficiencies are also below those of slab gel electrophoresis and the information content suffers as a result.

It is clear that single-molecule detection and identification of mRNA and proteins is required to characterize the single-cell transcriptome and proteome. Advances in single-molecule spectroscopy over the past two decades have led to simple and rugged instrumentation based on fluorescence sensing, like flow cytometry and fluorescence microscopy. Stability issues aside, mRNA can be quantified with high specificity by using the appropriate hybridization (complementary) probe even without amplification by the polymerase chain reaction (PCR). One such approach is fluorescence in situ hybridization (FISH).[5] Fluorescence provides a linear response over a large dynamic range as long as the degree of photobleaching is negligible.

The analogous scheme for detecting single proteins in single cells does not work well. Even though fluorescent antibodies are effective probes for proteins, their binding affinities are often not large enough for single-molecule sensing. Furthermore, comprehensive libraries of antibodies are not available. Fortunately, gene splicing has facilitated the co-expression of the protein of interest with a signal-producing protein that can be detected at the single-molecule level such as β-galactosidase,[6] luciferase,[7] and green fluorescent protein (GFP).[8] Of these, the first two can be highly sensitive since large numbers of signaling product molecules can be accumulated over time, but the fact that further reaction is required means that quantitation is indirect, is dependent on the conditions, and requires adding foreign substrates. In the original GFP system it takes two hours for the translated protein to fold to the efficient fluorescent form; in addition, this system photobleaches easily. However, with the availability of a complete chromosomally affinity-tagged library of *Escherichia coli*[9] single-cell proteomics became

[*] Prof. E. S. Yeung
Ames Laboratory-USDOE, Iowa State University
118 TASF, Ames, IA 50011 (USA)
Fax: (+1) 515-294-0105
E-mail: yeung@ameslab.gov

[**] The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under contract DE-AC02-07CH11358. This work was supported by the Director of Science, Office of Basic Energy Sciences, Division of Chemical Sciences.

possible, since yellow fluorescent protein (YFP) translational fusion can be implemented.[10]

Armed with these technological innovations, Taniguchi, Li, et al. performed an elegant comprehensive determination of the proteome in single cells by using a YFP fusion library for the model organism *E. coli*.[11] The intact cells were imaged by fluorescence microscopy (Figure 1). Clearly a large

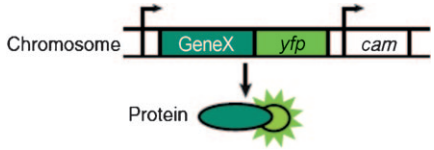

Figure 1. Quantitative imaging of a YFP fusion library.[11] Each library strain has a YFP translationally fused to the C terminus of a protein in its native chromosomal position. Cam: chloramphenicol acetyltransferase gene.

number of cells have to be imaged in this type of comprehensive study. The solution is to take advantage of the volume size and the high degree of multiplicity afforded in microfluidics. Cells were screened 96 at a time and automatic image analysis was developed to generate the huge amount of data. About 4000 cells for each strain can be imaged in 25 s. For selected proteins that are cytoplasmic membrane- or DNA-binding, localization of the YFP fluorescence in the expected locations was evident. Intensities were first corrected for autofluorescence by using a nonfluorescent strain and then compared to the average intensities per isolated membrane-bound protein found in similar cell images. The individual cell volumes were recorded at the same time to convert the calculated protein numbers to the concentrations inside the cell. The approach was validated by using mass spectrometry and by Western blotting.

The proteins in 1009 of the 1018 strains studied were found to fit a gamma distribution, as was suggested previously in a limited study.[6] The measured protein levels can thus be used to understand noise and dynamics in gene expression.[12] By using pairs of selected genes, it was found that while intrinsic noise (inherent to random statistics) dominates at low numbers of protein copies, extrinsic noise (cell-to-cell variability) is a common feature of all highly expressed genes.[11] Another interesting feature is that the *E. coli* system shows greater extrinsic noise compared to yeast[12] even for proteins with similar copy numbers. However, the reason behind this is unclear.

The numbers of mRNA copies inside each cell, down to the single-molecule level, were also determined in a similar manner (Figure 2). In order to allow simultaneous imaging of the mRNA and the corresponding translation product, a different color label was utilized to hybridize with the same *yfp* gene by using FISH. The approach was validated in control experiments by using real-time PCR for quantitation of the mRNA. Unlike the translation product that can be imaged in real time, the cell must be fixed for FISH imaging. Fortunately, the YFP fluorescence levels were found to correlate well in the cells before and after fixation.

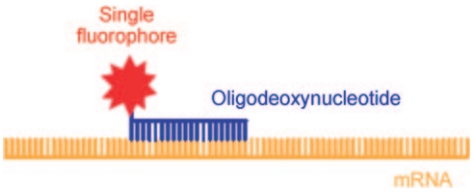

Figure 2. The mRNA of a tagged gene can be detected by FISH against the *yfp* mRNA sequence by using a DNA oligomer probe that is labeled with a single Atto594 fluorophore.[11]

The simultaneous determination of the mRNA and the corresponding protein of every gene in a cell (Figure 3) sheds unprecedented light on the mechanism of expression. Previous studies have shown that proteins are produced in bursts

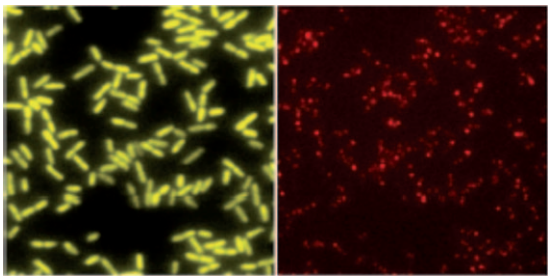

Figure 3. Images of proteins in a YFP fusion strain (yellow, left) and the corresponding Atto-594-labeled mRNA (red, right) in single cells.[11]

that are stochastic in time and in amounts, consistent with having low numbers of mRNA copies.[10] The question remains whether the copy numbers of mRNA and those of the translation product are correlated. Surprisingly, for the 129 strains studied that have highly expressed genes, no correlation was found between the mRNA copy number and the protein copy number at a given point in time, being the moment of fixation. This can be explained by the different lifetimes of the two biomolecules within a cell. Proteins are accumulated products, while mRNA are the instantaneous messengers that are degraded in minutes. Naturally, one would still expect that the mRNA levels integrated over time would correlate with the protein levels.

While this study represents a major step towards comprehensive single-cell transcriptomics and proteomics, further developments are desirable. To pin down the correlation between mRNA numbers and protein numbers, one would like to follow them through an entire cell cycle. While imaging proteins can be achieved in real time, the use of FISH at present allows only a single point measurement. Also, to truly reflect function, one needs to follow the enzymatic reactions themselves rather than the protein levels, that is, measure the metabolites. Those are of course even more challenging to monitor because of the lack of available tags and because of the diversity of these compounds.

Received: September 23, 2010

[1] J. E. Edström, *Nature* **1953**, *172*, 809.
[2] G. T. Matioli, H. B. Niewisch, *Science* **1965**, *150*, 1824.
[3] S. J. Lillard, E. S. Yeung, R. M. A. Lautamo, D. T. Mao, *J. Chromatogr. A* **1995**, *718*, 397.
[4] D. Cohen, J. A. Dickerson, C. D. Whitmore, E. H. Turner, M. M. Palcie, O. Hindsgaul, N. J. Dovichi, *Annu. Rev. Anal. Chem.* **2008**, *1*, 165.
[5] J. M. Levsky, S. M. Shenoy, R. C. Pezo, R. H. Single, *Science* **2002**, *297*, 836.
[6] L. Cai, N. Friedman, X. S. Xie, *Nature* **2006**, *440*, 358.
[7] S. H. Kang, S. Lee, E. S. Yeung, *Angew. Chem.* **2010**, *122*, 2657; *Angew. Chem. Int. Ed.* **2010**, *49*, 2603.
[8] R. Y. Tsien, *Annu. Rev. Biomed.* **1998**, *67*, 509.
[9] G. Butland, et al., *Nature* **2005**, *433*, 531.
[10] J. Yu, J. Xiao, X. Ren, K. Lao, X. S. Xie, *Science* **2006**, *311*, 1600.
[11] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X. S. Xie, *Science* **2010**, *329*, 533.
[12] J. R. S. Newman, et al., *Nature* **2005**, *441*, 840.

