NISR Research GRANT

2007 Young Investigator Research Grant

Comprehensive analysis of the effect of *E. coli* proteins on the translation reaction and its application to protein production

Tomoaki MATSUURA

Department of Bioinformatic Engineering, Osaka University

Research aims

The protein translation reaction, one of the most important regulators of cell behavior, involves the interactions of a large number of components, and can thus be seen as an intermolecular interaction network. It has been demonstrated experimentally that 36 proteins and the ribosomes are sufficient to carry out protein translation.[1] These minimal protein components include the ribosomal proteins, initiation, elongation, and release factors, aminoacyl-tRNA synthetases, and enzymes involved in energy regeneration. While the genome of *Escherichia coli* contains more than 4000 genes,[2] constituting a very large interaction network,[3] the number of protein components constituting the protein synthesis system corresponds to only 2.1% of the genes encoded in the genome. Thus, only small subsets of the protein components are required for protein synthesis. On the other hand, a number of previous studies, including protein–protein interaction (PPI) network analysis in *E. coli*,[3] indicated that protein components constituting the minimal protein synthesis system interact with a large number of other proteins. To gain a deeper understanding of the protein translation system, it is important to identify not only the proteins that interact physically but also those that interact functionally, *i.e.*, those that affect the activity of the translation machinery. While there have been many studies to characterize the properties of such individual proteins in detail, there have been no previous attempts to search for such proteins in the entire genome. Hence, it is not clear what fraction of gene products of *E. coli* are linked functionally with the minimal protein synthesis system.

Methods

In the present study, we performed an experimental search for open reading frame (ORF) products of the *E. coli* genome that affect the activity of the translation system. The PURE system is an *E. coli*-based protein synthesis system composed of the minimal protein components[1]. As this is a reconstituted system comprised of defined components, it differs from other conventional cell extract-based *in vitro* translation systems, which contain a number of unidentified components. We measured the effects of 4194 individual ORF products on the green fluorescent protein (GFP) synthesis reaction using the PURE system.

Results

By measuring the effects of individual ORF products on the GFP synthesis reaction using the PURE system, we demonstrated that at least 12% of the 4194 ORF products of *E. coli* can affect the activity of the system (Fig. 1).[4] We designated these as functional modifiers of the protein synthesis reaction composed of minimal protein components.

We found 10 ORF products that completely suppressed the GFP synthesis reaction, in addition to many that decreased the efficiency to a lesser extent. These genes included those encoding the transcriptional repressor LacI (GFP DNA construct used contained the lacO region), toxins (ChpA, ChpB, RelE, and YoeB) known to exhibit ribosome-dependent nuclease activity, and the nuclease Rnt. Thus, our data were consistent with previous observations. On the other hand, functionally uncharacterized proteins were also found to be lethal for the reaction. To verify that the beneficial components are effective when added to the reaction as purified proteins, six proteins were overexpressed and purified from *E. coli*. While the optimum concentration for each protein differed, addition of six purified

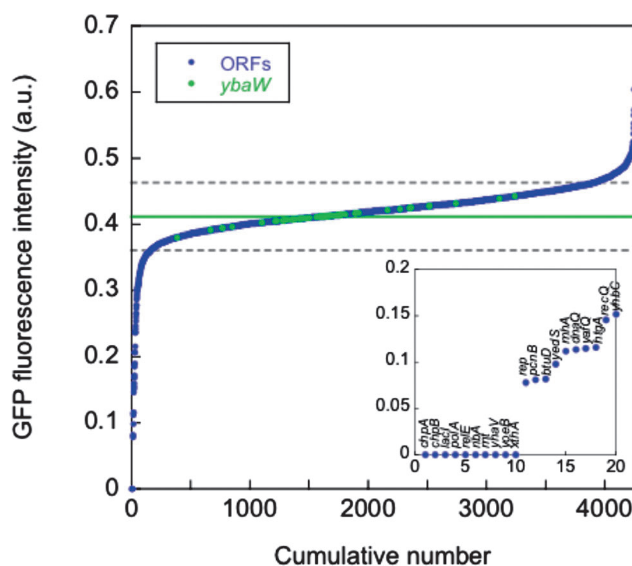

Fig. 1. The effect of *E. coli* ORF products on the GFP synthesis reaction Vertical axis shows the GFP fluorescence intensity obtained after 3 h in the presence of each ORF products during the GFP synthesis reaction. Numbers on the horizontal axis indicate the ranking of each ORF gene assigned according to its fluorescence intensity among the 4194 ORFs and 62 independent measurements of the *ybaW* gene. Number 1 is the ORF that gave the lowest value.

proteins resulted in an increase in the fluorescence intensity. Furthermore, when all six proteins (HrpA, Orn, PhnH, SlyD, Tig, and TrxC) were added at the optimum concentrations, the fluorescence intensity increased by 2.44-fold.

The PPI network shows the physical interactions between the proteins, and such networks from various organisms have been investigated; the results of these studies have indicated that proteins are highly connected with each other. As the protein translation system is embedded in such a large interaction network, we were also interested in the topological relationships between the minimal components and those that are functionally linked with them in the PPI network of *E. coli*, which may provide insight into the topological structure and evolution of the protein synthesis system. We thus mapped each of the components involved in the protein synthesis reaction on the PPI network of *E. coli*.[3] Network analyses indicated that functional modifiers seem to be spread across the PPI network rather than clustering close to the minimal protein components.[4] A possible interpretation of this observation in relation to the evolutionary process of the protein synthesis system is that the system may have evolved by first establishing a system with minimal components, and then adding components further outside rather than integrating them into the network composed of the minimal components.

Conclusion

The protein translation reaction, one of the most important regulators of cell behavior, can be operated by 91 protein components (55 ribosomal proteins+36 factors), representing only 2.1% of the entire complement of ORFs in *E. coli*. Through comprehensive analysis, we showed experimentally that at least 12% of the *E. coli* genes affected the activity of the minimal protein synthesis system. The comprehensive data obtained in this study will be useful in future studies on individual proteins as well as functional genomics and systems biology. While we are still at the starting point of investigating the precise roles of the modifier proteins on the reaction, combining the experimental data with the PPI network may provide insight into how these proteins interact with the translation machinery. Moreover, our data may be useful not only for technologies utilizing *in vitro* translation systems, but also for designing *in vitro* translation systems with significantly improved performance than those available at present and also and for protein production *in vivo*, such as by coexpressing beneficial components together with a protein of interest.

References

1) Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
2) Riley, M. *et al. Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res.* **34**, 1–9 (2006).
3) Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli. Nature* **433**, 531–537 (2005).
4) Kazuta, Y. *et al.* Comprehensive analysis of the effects of *Escherichia coli* ORFs on protein translation reaction. *Mol Cell Proteomics* **7**, 1530–1540 (2008).