Genome scale prediction of two-component signal transducers from the knowledge of regulatory interactions.

Hidemasa Bono
bono@kuicr.kyoto-u.ac.jp
Susumu Goto
goto@kuicr.kyoto-u.ac.jp
Hiroyuki Ogata
ogata@kuicr.kyoto-u.ac.jp
Minoru Kanehisa
kanehisa@kuicr.kyoto-u.ac.jp
Institute of Chemical Research, Kyoto University
Gokanosho, Uji, Kyoto 611, Japan

Abstract

Predicting gene functions from the whole genome sequence is an important problem in a post-genome era. We are developing a function predicting system from the whole genome sequence utilizing the functionally well annotated genome as a reference organism for the knowledge of biologically well known pathways. The databases of gene catalogs and pathways are compiled under the KEGG project. In this paper we show an instance for identifying functions of genes involved in the two-component signal transduction system.

1 Introduction

Analysis of gene functions is the next target of genome projects after the completion of sequencing. An *in silico* approach toward the analysis is being achieved by KEGG (Kyoto Encyclopedia of Genes and Genomes) project[2][3]. The project aims to computerize an enzymatic network of metabolic pathways and other molecular interaction networks making up a cell. Under the project we are developing an automatic system for identifying gene functions[1], called GFIT(Gene Function Identification Tool), utilizing the database of orthologue groups.

Bacteria have devised sophisticated signaling systems for responses to their environment, including two-component signal transduction of proteins that consist of sensory kinase and response regulator. Generally, sensory kinase senses an environmental change, then it phosphorylates a cognate response regulator, which continuously causes transcriptional change to the target genes. Finding candidate components can be done by homology search[4], but the prediction of precise function is not so straightforward because all components have a long conserved region concerning phosphotransfer, which does not tell much about specific pairing.

This paper presents a system for identifying functions of two-component phosphotransfer signal transducers, by making use of GFIT system and binary relation[6] between sensory kinase and response regulator catalogued in KEGG.

2 System and Method

The database of orthologue groups is made from all the protein sequences of the complete genomes taken from the original sources. Orthologue groups are first generated automatically by FASTA program, and then modified by hand to assign same function in the group. Genes with multifunctional domains are separated in the database, so some genes may be clustered in two or more groups.

To predict two-component transducers a database or known two-component signal transducers is required. We are now constructing such a database for the completely sequenced genomes. The data of *E.coli*, taken from Mizuno(1997)[5], are mostly based on experimental observations, so we use them

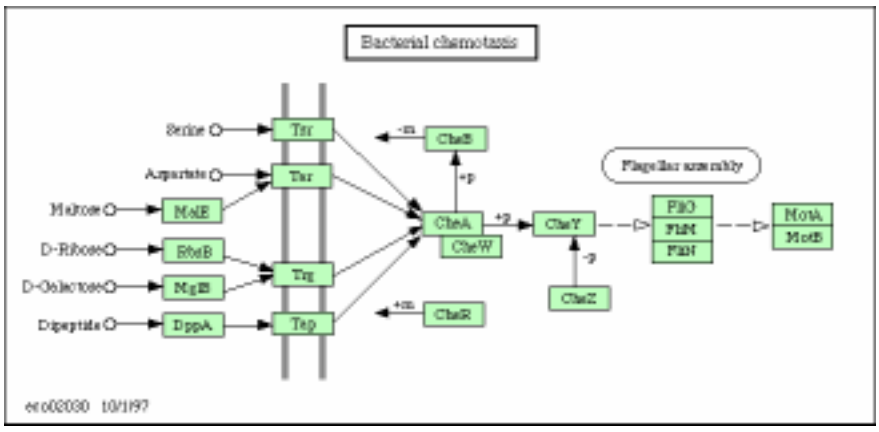

Figure 1: Bacterial chemotaxis regulatory pathway in *E.coli*

as a reference to predict pathways in other organisms. One example of the data is shown in Figure 1. In Figure 1, sensory kinase `CheA` phosphorylates response regulator `CheY`.

In the past we were predicting the function of two-component signal transducers only by homology search or by a conserved motif in the phosphorylation sites, which was like a piece-by-piece prediction. Now, utilizing the feature that a sensory kinase and response regulator form a regulatory pathway or the same operon, we can predict signaling cascades with a higher certainty. Such interactions among all the existing metabolic and regulatory pathways will be included in GFIT system.

Acknowledgement

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Genome Science' from the Ministry of Education, Science, Sports and Culture in Japan. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References

[1] Bono, H., Ogata, H., Goto, S., Fujibuchi, W., Kanehisa, M., "Gene Function Identification System Based on the Metabolic Pathway Database," *Genome Mapping & Sequencing*, page 33, May, 1997.

[2] Goto, S., Bono, H., Ogata, H., Fujibuchi, W., Nishioka, T., Sato, K., Kanehisa, M., "Organizing and computing metabolic pathway data in terms of binary relations ," *Pacific Symposium on Biocomputing '97*, pages 175-186, Jan. 1997.

[3] Kanehisa, M., "A database for post-genome analysis" *Trends in Genetics*, 13, 375-376 (1997)

[4] Mizuno, T., Kaneko, T., Tabata, S., "Compilation of All Genes Encoding Bacterial Two-component Signal Transducers in the Genome of the Cyanobacterium, *Synechocystis* sp. Strain PCC 6803," *DNA Research*, 3, 407-414 (1996)

[5] Mizuno, T., "Compilation of All Genes Encoding Two-component Phosphotransfer Signal Transducers in the Genome of *Escherichia coli*," *DNA Research*, 4, 161-168 (1997)

[6] Ogata, H., Bono, H., Fujibuchi, W., Goto, S., Kanehisa, M., "Analysis of Binary Relations and Hierarchies of Enzymes in the Metabolic Pathways," *Genome Informatics 1996*, pages 128-136, Yebisu, Dec. 1996. Universal Academy Press.