IS 2-2

Toward Modeling a Cell Using *Escherichia coli*

Hirotada Mori

Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101,
Institute of Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0017

E. coli is undoubtedly one of the best studied organisms and enormous amount of information that has been collected over the past half century (e.g., genome sequence, mutant phenotypes, metabolic and regulatory networks, etc.), we now have detailed knowledge about gene regulation, protein activity, several hundred enzyme reactions, metabolic pathways, macromolecular machines, and regulatory interactions for this model organism. However, understanding how all these processes interact to form a *living* cell will require further characterization, quantification, data integration, and mathematical modeling, that is Systems Biology or Synthetic Biology [1]. No organism can rival *E. coli* with respect to the amount of available basic information and experimental tractability for the technologies needed for this undertaking. A focused, systematic effort to understand the *E. coli* cell will accelerate the development of new post-genomic technologies, including both experimental and computational tools. It will also lead to new technologies applicable to other organisms, from microbes to plants, animals, and humans.

E. coli is not only the best studied free-living model organism but is also an extensively used microbe for industrial applications, especially for production of small molecules of interest and is an excellent representative of Gram-negative commensal bacteria.

E. coli may represent a perfect model organism for novel biology directing whole-cell modeling and simulation.

1. The *E. coli* genome sequence.

The complete *E. coli* genome sequence was determined through independent efforts by US and Japanese groups using two different strains of *E. coli* K-12, MG1655 and W3110, respectively [2-6]. These strains diverged from the same ancestral strain about 50 years ago resulting in slight but significant differences including the large inversion between ribosomal *rrnD* and *rrnE* genes in W3110 [7]. Complete genome sequence analysis revealed precise differences between the two strains. Comparison of the genome sequence so far revealed relatively infrequent, 1 per 10^5 base substitution mutations in the protein coding region [8]. Recently, the Japanese group confirmed differences between the genomes of MG1655 and W3110 by PCR-based direct sequencing and found that only 9 bases in 8 different ORFs represent actual sequence differences besides large scale insertions or deletions, such as IS elements [9]. In summary, the *E.*

coli genome thus seems tolerate large rearrangements such as insertion, deletion and recombination than micro-scale nucleotide changes.

1.1 ORF prediction and annotation

The recently revised *E. coli* genome sequence was estimated to contain about 4,300 genes of which about half of those have not been characterized experimentally [10, 11]. In total about 3,600 genes can be assigned or predicted a function with reasonable assurance based on biochemical experiments and computational analysis in *E.coli* and other microorganisms. Of the remaining about 700 genes, 650 show sequence similarity to genes of unknown function in other bacteria, whereas 50 show no obvious similarity to any known genes. The assignment of function to these unknown genes is one of the major targets of functional genomics in *E. coli*. In addition to their fundamental importance for understanding *E.coli* biology, these functional assignments are significant for three other reasons: complete functional assignment will result in discovery of new physiological and biochemical pathways, will facilitate functional assignment in other bacteria and will lead to identification of new targets for antibiotic design other for biotechnological development.

Besides, elucidation of individual gene functions which is basic to understanding of a cell, systematic analysis of relationships between genes or gene products is also a significant target that is just starting to be explored systematically.

2. Post genome sequencing project

Nearly half of the total ORFs in *E.coli* is not experimentally confirmed their physiological function. The latest estimation reveals that about 700 to 800 of total ORFs have no attributable function. Therefore, a high priority will be placed on the development of novel, high-throughput technologies to identify their function. A new comprehensive "molecular tool kit" would be required to define unknown gene functions and to assign potentially new roles to known genes. A large number of valuable plasmid constructs, *E. coli* strains, assay tools, and other biological materials useful for analyzing gene function have been constructed, developed, collected, and tested. These will form key resources for getting basic and comprehensive knowledge on *E. coli* biology and for exploiting these data with *E. coli* for other organisms. Furthermore, methods for genome-wide analysis have rapidly evolved to make best use of the rich DNA sequence information ("-ome" is a Greek suffix for "whole").

2. 1 Experimental resources

DNA sequence analysis has identified about 4,300 protein coding genes in *E. coli*. The power of genetically tractable model organisms resides in the fact that they can facilitate the global and systematic analysis of physiological gene function in vivo. Precise genetic manipulation is particularly important for functional genomics. Genome sequence data has permitted the design of oligo DNA primers for precise amplification of entire ORFs and generation of a complete set of histidine-tagged ORF clones (with or

without fusion to GFP gene) [10, 12].

One approach for systematic functional analysis is to make use of gene deletion (replacement) obtained by homologous recombination. Targeted gene replacement, once thought to be difficult in *E. coli,* can now be dealt with by using the techniques developed by Wanner and his colleague [13]. The systematic attempts to construct these resources for functional genomics is now rapidly raising *E.coli* to one of the leading organisms in the field of functional genomics.

2.1.1. Plasmid library of predicted ORFs (ASKA library; A Complete Set of *E. coli* K-12 ORF Archive)

ORF clones will provide a basic genetic tool for studying gene function, since they provide template for PCR amplification and for preparing purified gene products etc. To clone all the genes of *E. coli*, a plasmid vector with the following properties was constructed: 1) high copy number plasmid, 2) IPTG inducible expression of cloned ORF and repression of expression by $lacI^q$, 3) a Histidine tag coding sequence attached to the N-terminal of ORF, 4) in-frame fusion with GFP coding sequence at the C terminal end, 5) generation of *Sfi*I restriction sites at both boundaries of the cloned ORF, 6) possibility of GFP fragment removal by *Not*I [12]. The whole set of PCR amplified ORF fragments were cloned into the *Stu*I site of this vector. As far as we know, these clones represent the only comprehensive collection of *E.coli* ORFs that is currently publicly available. The structure of these clones is shown in Fig. 1.

2.1.2 Deletion mutant library (KEIO collection; Knock-out library of *E. coli* in frame deletion of ORFs)

E. coli has long been thought as one of difficult organisms for linear DNA transformation because of the presence of intracellular exonucleases. In contrast, genes can be directly disrupted in *Saccharomyces cerevisiae* by transformation with PCR fragments encoding a selectable marker having only 35 nt of flanking DNA homologous to the chromosome [14]. On the other hand, it has long been known that many bacteriophages encode their own homologous recombination systems [15]. It was recently shown that the λ Red (γ, β, *exo*) function promotes a greatly enhanced rate of recombination over that exhibited by the *recBC sbcB* or *recD* mutants when using linear DNA. Wanner and his colleagues developed a valuable

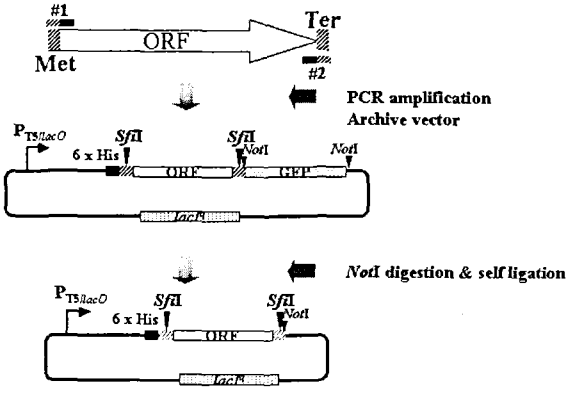

Fig. 1 Construction of ASKA library

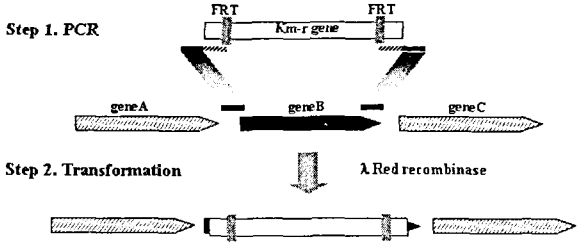

Fig. 2 Construction of KEIO collection

protocol based on the λ Red system that provides an efficient way to isolate replacement mutants using PCR fragments encoding an antibiotic resistance gene and having only 40 – 50 nt of flanking regions [13]. A comprehensive and clean deletion mutant library for all *E. coli* ORFs using the λ Red system is now finished and also ready for distribution as an open resource for academic consortium [16]. The design of this collection is illustrated in Fig. 2.

2.1.3 Large deletion mutants.

Two recent publications describe deletions of large chromosomal segments from the *E. coli* genome. Kim and his colleagues made two large libraries of independent transposon mutants using modified Tn5 transposons with two different selection markers and precisely mapped their chromosomal location [17]. The method allows to integrate the mapped transposon insertion carrying *loxP* site from each of the mutant libraries into the same chromosome followed by excision of the flanking genomic segments by site-specific Cre mediated *loxP* recombination. Alternately, Blattner and his colleagues used a deletion "by design approach" and have already generated lineages in which more than 10% of the genome has been eliminated without loss of viability [18]. Another large-scale deletion construction has been reported by Hashimoto et al[19]. Basically the approach consists of markerless gene replacement, and has already generated a genome lacking more than 25% of the original genome DNA. Intermediates have been saved and descendents with larger deletions are being generated.

We also performed similar approach for large scale deletion construction by Wanner's method using our oligo DNA primer set used for construction of KEIO collection.

By examining representatives of these collections for growth defects under specific environmental conditions, the effects of losing many genes can be traced simultaneously. Data generated in this type of experiment will be a great asset for cell modeling. Such approaches are likely to be especially advantageous when single gene mutations display no discernable phenotypic changes.

2.1.4 Others approaches for functional geneomics

Two different types of technology were recently developed to study protein expression and protein-protein interaction. One is based on novel tandem affinity purification (TAP) of tag fusion protein. This is a generic procedure to purify target proteins expressed at their natural level under native conditions [20]. To investigate heteromeric protein complexes of unknown composition, standard systems for protein overexpression may lead to the assembly of overexpressed proteins as non-physiological complexes. To overcome this problem, a TAP tag fusion cassette was developed which encodes calmodulin-binding peptide (CBP), a TEV protease recognition site, and proteinA of *Staphylococcus aureus* (ProtA).

The other method depends on the genome-wide, registered collection of *E. coli* bioluminescent reporter gene fusions. Each of the random fusions of *E. coli* chromosomal DNA fragment to the *Photorhabdus luminescens luxCDBE* reporter gene was precisely mapped by sequencing [21]. To identify and quantify changes in expression level, the authors tested this type of fusion and analyzed alterations in expression level of heat shock, SOS response and oxidative stress genes.

2.2 Transcriptome analysis: To exploit the rapid progress in genome research, many novel techniques have been developed including DNA microarray or DNA chip technology that are extremely useful for analysis of global gene expression [22]. Generally, a DNA microarray is defined as an orderly arrangement of tens to hundreds of thousands of unique DNA molecule of known sequence, usually on a glass slide. Unique DNA molecule are either individually synthesized on a rigid silicon plate (generally referred to as DNA chips and developed by Affymetrix Co.) or prepared from pre-synthesized DNA (synthetic oligo-nucleotides or PCR products) that are spotted and immobilized on a slide glass. The use of DNA microarrays to study *E. coli* gene regulation was first illustrated by Blattner and colleagues, and has rapidly expanded and applied to study various aspects of transcriptional regulation [23]. To elucidate whole transcriptional regulatory network, several systematic approaches using DNA microarray have been performed [24, 25]. These analyses were done not only under different growth conditions but also using either overexpression or deletion of a target regulatory gene. As described above, large experimental resources including complete sets of clones and deletion mutants of all of the predicted genes of *E. coli* are being established and those resources will contribute to accelerate such systematic transcriptome analyses. Accumulation of the results from such large-scale analyses using DNA microarrays will also assist more traditional biological research, and construction of large databases will have a wide beneficial impact. Public availability of such data is not limited to supplemental data derived from publications on the ftp site of journal publishers but is also available from integrated databases such as the KEGG database as a systematic collection of microarray data [26].

The accumulation of publicly available DNA microarray analyses data promotes rapid computational analysis of gene expression profiles. In general, bacterial genes form operon in which multiple ORFs are transcribed from the same promoter to form a single mRNA transcript. Prediction of operons using DNA microarray experiments is one approach used to reconstruct gene regulatory networks at the whole genome scale [27]. Reconstruction of global regulatory networks using genome-scale gene expression data sets has also been reported recently [28, 29].

2.3 Proteome

2.3.1 One of the basic technologies for global analysis of cellular proteins has been developed by O'Farrell as two-dimensional polyacrylamide gel electrophoresis [30]. *E. coli* has a long history of protein cataloging by 2D gel [31]. The major limitations of 2D electrophoresis are related to the difficulty in assigning gene identities to observed spots and the lack of spots for proteins that do not separate well on 2D gels. In addition, the method displays a general bias against membrane proteins and proteins of low abundance. However, matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS) has greatly contributed to accelerate protein identification [32], and a new gel based separation system (RFHR) significantly expanded the separation range of proteins [33]. The assignment of 2D gel data will be coordinated between two major ongoing initiatives, the Cybercell project (Ellison, personal communication) and the efforts of a Japanese group (Wada , personal communication). in addition, several alternative methods based on the analysis of peptides from whole cell lysates by LC/MS were developed and successfully used [34, 35].

2.3.2 Protein localization: Even though the bacterial cell does not have the complex intracellular compartments that are hallmarks of eukaryotic cells, comprehensive information about the location of proteins within the bacterial cell is important for understanding their functions and interactions. Large-scale analyses of protein localization in *S. cerevisiae* have been reported [36, 37]. We also performed the comprehensive analysis of protein localization using clones of individual ORFs fused with GFP protein under non-induced growth conditions[38]. Localized GFP fluorescence was successfully observed for about 4,000 out of 4,300 genes tested, and the patterns of localization were classified roughly into 4 distinct categories.

2.3.3 Protein-protein interaction: Most cellular processes are carried out by multiprotein complexes. The identification and analysis of these protein complexes provides further insight into physiological function and molecular mechanisms of the functional units. Following identification and cataloging of all the proteins expressed in a cell, global analysis of protein-protein interaction becomes critical for understanding cellular processes. In *S. cerevisiae*, genome-scale analysis of protein complexes has been performed using the yeast two-hybrid system [39-41], protein chips [42] or affinity tagged system [43]. In *E. coli*, two comprehensive analyses using affinity-tagged proteins, chromosomally tagged with TAP (J. Greenblatt, personal communication) and plasmid clones containing an histidine tag [44], to identify protein-protein interaction, as was done in *S. cerevisiae*, are now underway. These will allow identification by mass spectrometry of proteins that co-purify with the tagged baits, that are thus candidate interacting proteins. Burtland et al has recently reported interaction network from nearly 1000 *E. coli* ORFs using TAP tagged BAIT proteins [45]. On the other hand, we also performed high-throughput analysis using His tagged plasmid clones (ASKA library) as Bait proteins. Out of 4,300 total ORFs, about 2,700 ORFs were successfully analyzed from the plasmid clones and candidates that can interact with His-tagged bait proteins were identified [44]. The total number of observed interactions in this set of 2,700 proteins amounts to about 11,000 potential interactions. In protein-protein interaction analysis by pull-down assay using such as TAP and His tagged bait proteins, problems for flase positive and negative cannot be avoided and we have tried to pick out reliable interaction from the total interaction that we identified using physiological function of ORFs interacted. The algorithm itself is very simple and 78 novel putative protein complexes could be extracted (Fig 3). Also 60 novel protein interaction was identified as functional component (Fig. 3).

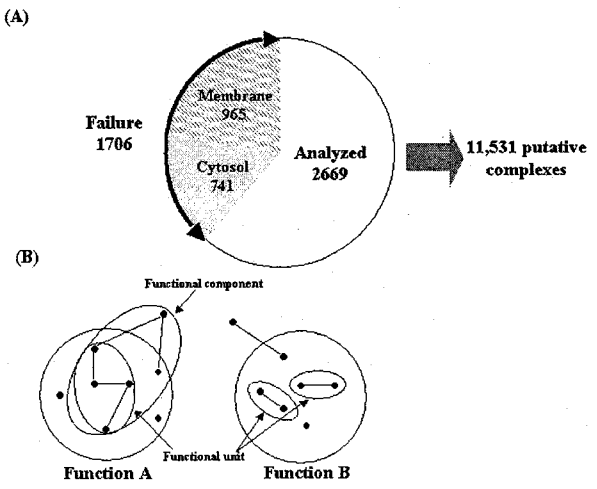

Fig. 3 Protein-protein interaction

2.4 Metabolome

A substantial portion of the *E. coli* genome encodes enzymes that interconvert metabolites, synthesize

cofactors, and regulate small molecule metabolism. Metabolites can in turn control gene expression and are allosteric regulators of enzymes. The metabolome can be described as the total complement of metabolites in a cell [46]. Metabolome analyses can be performed using several approaches, such as metabolite profiling, flux analysis using isotopic tracer, and pathway reconstruction etc., and allows to gain insight into metabolic and physiological responses within a cell. Global metabolite profiling will provide deeper insight not only into metabolism but also into cellular physiology and functional genomics [47]. This approach has been used for functional genomics studies in plants [48] and yeast [49]. Recently, Soga and colleagues developed a powerful analytical method using capillary electrophoresis-electrospray ionization mass spectrometry (CE-ESI-MS) that dramatically increases the number of metabolites that can be measured simultaneously [50-52].

Metabolites control gene expression of enzymes as a feed back regulation as mentioned above and on the other hand, cellular matabolic activities are controlled by enzyme concentration in a cell. And this is clear the importance of analysis of enzymes both for their activities and amounts in a cell. To perform this, we purified enzymes and produced antibodies against those using ASKA library. Purified enzymes were supplied to measure enzymatic parameters and antibodies to measure protein amounts in a cell by Western blotting analysis (Fig. 4).

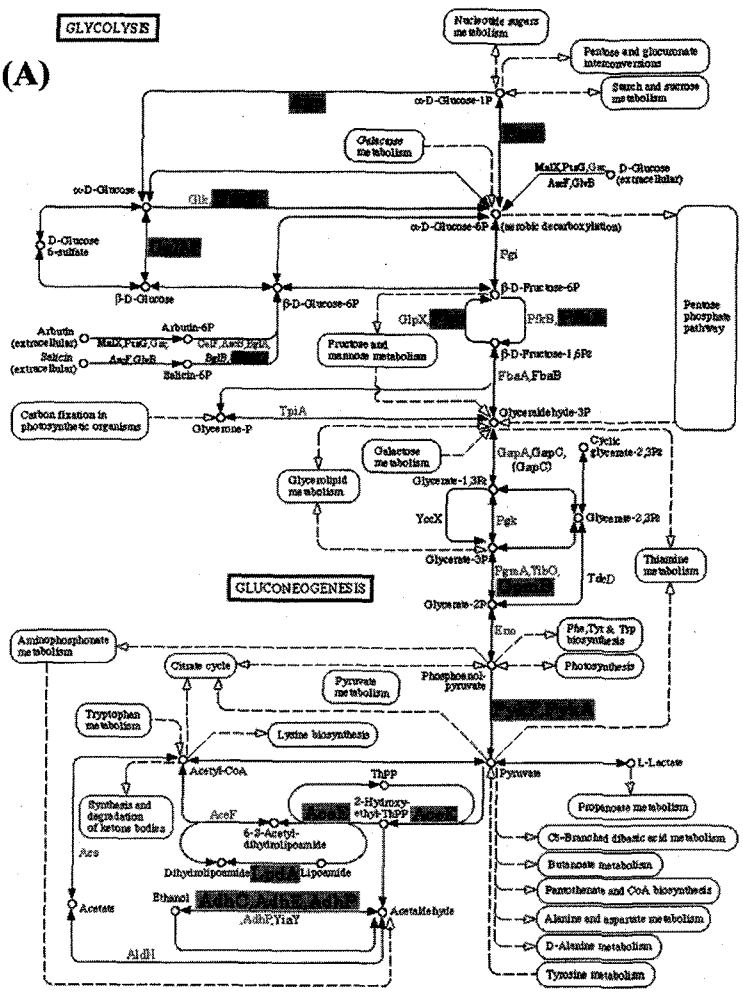

Fig. 4 Quantification of enzymes
(A) Antibody production in Glycolysis

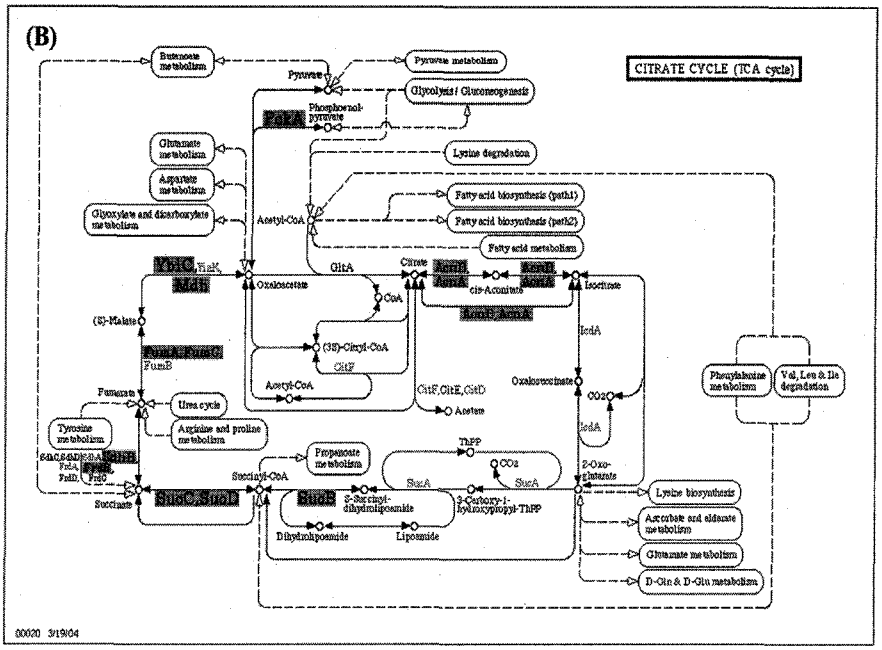

Fig. 4 Quantification of enzymes
(B) Antibody production in TCA cycle

(C)

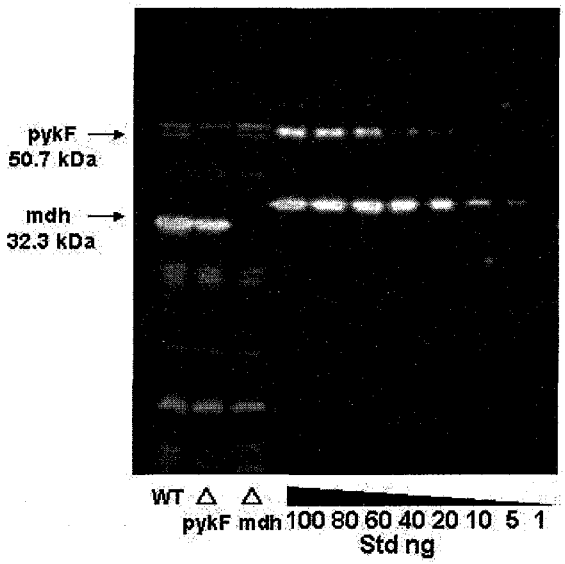

Fig. 4 Quantification of enzymes
(c) Western analysis

3. E.coli genome as a model for systems biology.

Biology itself is now clearly standing at a turning point between the past descriptive science and the emerging modern quantitative systems biology. Understanding the causal relationships between genotype and phenotype will require a very significant expansion of the traditional toolbox used by molecular

biologists. It must include concepts and techniques from many other scientific disciplines such as physics, mathematics, numerical analysis, stochastic processes, and control theory. Many novel tools must be developed to understand how dynamic, robust but adapting, and developing systems can emerge from the information buried in the genome [53]. The genome sequencing efforts and the subsequent bioinformatics analyses have not only defined the molecular parts for a number of living organisms but opened up the possibilities to reconstruct the metabolic pathways. The stoichiometric coefficients for each enzyme in the *E. coli* metabolic map were assembled to construct a genome-specific stoichiometric matrix, and the matrix was used to define the system's characteristics and the capabilities of this organism's metabolism [54]. In that report, the authors showed the result of comparisons between the *in silico* predictions and experimental observations using deletions of genes in the central metabolic pathways. And now, these approaches expand to the genome-scale reconstruction not only of metabolic network [55, 56] but also of heterogeneous types of network including transcription and translation [28, 57, 58]. A shift in biology from a component-based perspective to a systems view of the cell is occurring based on genome sequence accumulation and high-throughput post-genomic data generation. Modeling cellular functions according to a systems biology is not new but this approach is now expanding to reach the genome-scale and the total number of genes and biochemical elements integrated into a single model has now reached ~2000 [59]. In parallel, several software environments for quantitative simulation of cellular processes including metabolic pathways, based on the numerical integration of rate equations, have been developed [60-63].

4. International consortium for large scale *E. coli* modeling

The International *E. coli* Alliance (IECA, http://www.EcoliCommunity.org) was launched in November 2002 in UK to tackle the fundamental biological problem in developing the first comprehensive computational model of a living cell. IECA's mission is to consolidate global efforts to understand a living bacterial cell and construct a mathematical model. An *E.coli* cell model will have immediate practical benefits in biology and bioengineering and should significantly contribute to advancing the field of computational systems biology. The generation of a computerized *E. coli* cell will also add powerful new tools to our existing arsenal for functional discovery, including virtual experimentation and mathematical simulation. Ultimately, these biological and computational tools could be useful in both drug discovery and in the design of bioenhanced nanomachines. Furthermore, development of a virtual system for experimentation on the *E. coli* cell will be extremely useful for understanding more complex cells and contribute to the development and validation of *in silico* models of human cells and whole multicellular organisms. Biology is now evolving to become a "big science" and the tiny *E. coli* is well positioned to become one of the giant players in the new biology era, based on the determining role it played in the field of molecular genetics.

The next IECA international conference will be planed to be held in Korea in 2006.

References

1. Church, G., *From systems biology to synthetic biology.* Nat Mol Systems Biol, 2005. 10: p. 1038.

2. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12.* Science, 1997. 277(5331): p. 1453-74.
3. Aiba, H., et al., *A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 28.0-40.1 min region on the linkage map.* DNA Res, 1996. 3(6): p. 363-77.
4. Itoh, T., et al., *A 460-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 40.1-50.0 min region on the linkage map.* DNA Res, 1996. 3(6): p. 379-92.
5. Oshima, T., et al., *A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.* DNA Res, 1996. 3(3): p. 137-55.
6. Yamamoto, Y., et al., *Construction of a contiguous 874-kb sequence of the Escherichia coli -K12 genome corresponding to 50.0-68.8 min on the linkage map and analysis of its sequence features.* DNA Res, 1997. 4(2): p. 91-113.
7. Hill, C.W. and B.W. Harnish, *Inversions between ribosomal RNA genes of Escherichia coli.* Proc Natl Acad Sci U S A, 1981. 78(11): p. 7069-72.
8. Itoh, T., et al., *A low rate of nucleotide changes in Escherichia coli K-12 estimated from a comparison of the genome sequences between two different substrains.* FEBS Lett, 1999. 450(1-2): p. 72-6.
9. Horiuchi, T., in preparation.
10. Mori, H., et al., *Functional genomics of Escherichia coli in Japan.* Res Microbiol, 2000. 151(2): p. 121-8.
11. Riley, M., Annotation meeting at Wood's Hole. 2005.
12. Kitagawa, M.M., H., in preparation.
13. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. 97(12): p. 6640-5.
14. Baudin, A., et al., *A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae.* Nucleic Acids Res, 1993. 21(14): p. 3329-30.
15. Smith, G.R., *Homologous recombination in procaryotes.* Microbiol Rev, 1988. 52(1): p. 1-28.
16. Baba, T., in preparation.
17. Yu, B.J., et al., *Minimization of the Escherichia coli genome using a Tn5-targeted Cre/loxP excision system.* Nat Biotechnol, 2002. 20(10): p. 1018-23.
18. Kolisnychenko, V., et al., *Engineering a reduced Escherichia coli genome.* Genome Res, 2002. 12(4): p. 640-7.
19. Hashimoto, M., et al., *Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome.* Mol Microbiol, 2005. 55(1): p. 137-49.
20. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. 17(10): p. 1030-2.
21. Van Dyk, T.K., et al., *A genomic approach to gene fusion technology.* Proc Natl Acad Sci U S A, 2001. 98(5): p. 2555-60.
22. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays.* Nat Biotechnol, 1996. 14(13): p. 1675-80.
23. Tao, H., et al., *Functional genomics: expression analysis of Escherichia coli growing on minimal and rich media.* J Bacteriol, 1999. 181(20): p. 6425-40.

24. Oshima, T., et al., *Transcriptome analysis of all two-component regulatory system mutants of Escherichia coli K-12.* Mol Microbiol, 2002. 46(1): p. 281-91.
25. Masuda, N. and G.M. Church, *Regulatory network of acid resistance genes in Escherichia coli.* Mol Microbiol, 2003. 48(3): p. 699-712.
26. Kanehisa, M., et al., *The KEGG databases at GenomeNet.* Nucleic Acids Res, 2002. 30(1): p. 42-6.
27. Sabatti, C., et al., *Co-expression pattern from DNA microarray experiments as a tool for operon prediction.* Nucleic Acids Res, 2002. 30(13): p. 2886-93.
28. Gutierrez-Rios, R.M., et al., *Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles.* Genome Res, 2003. 13(11): p. 2435-43.
29. Herrgard, M.J., M.W. Covert, and B.O. Palsson, *Reconciling gene expression data with known genome-scale regulatory network structures.* Genome Res, 2003. 13(11): p. 2423-34.
30. O'Farrell, P.H., *High resolution two-dimensional electrophoresis of proteins.* J Biol Chem, 1975. 250(10): p. 4007-21.
31. Vanbogelen, R.A., *in Escherichia coli and Salmonella: Cellular and Molecular Biology. ed. Neidhart. F. C.* 1996.
32. Figeys, D., et al., *Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes.* Electrophoresis, 1998. 19(10): p. 1811-8.
33. Wada, A., et al., *A 5 kDa protein (SCS23) from the 30 S subunit of the spinach chloroplast ribosome.* FEBS Lett, 1993. 319(1-2): p. 115-8.
34. Corbin, R.W., et al., *Toward a protein profile of Escherichia coli: comparison to its transcription profile.* Proc Natl Acad Sci U S A, 2003. 100(16): p. 9232-7.
35. Gygi, S.P., B. Rist, and R. Aebersold, *Measuring gene expression by quantitative proteome analysis.* Curr Opin Biotechnol, 2000. 11(4): p. 396-401.
36. Ross-Macdonald, P., et al., *Large-scale analysis of the yeast genome by transposon tagging and gene disruption.* Nature, 1999. 402(6760): p. 413-8.
37. Kumar, A., et al., *Subcellular localization of the yeast proteome.* Genes Dev, 2002. 16(6): p. 707-19.
38. Niki、H., *in preparation.* 2005.
39. Fromont-Racine, M., J.C. Rain, and P. Legrain, *Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens.* Nat Genet, 1997. 16(3): p. 277-82.
40. Ito, T., et al., *A comprehensive two-hybrid analysis to explore the yeast protein interactome.* Proc Natl Acad Sci U S A, 2001. 98(8): p. 4569-74.
41. Uetz, P., et al., *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.* Nature, 2000. 403(6770): p. 623-7.
42. Zhu, H., et al., *Global analysis of protein activities using proteome chips.* Science, 2001. 293(5537): p. 2101-5.
43. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. 415(6868): p. 141-7.
44. Arifuzzaman, M., in preparation.
45. Butland, G., et al., *Interaction network containing conserved and essential protein complexes in*

Escherichia coli. Nature, 2005. 433(7025): p. 531-7.

46. Tweeddale, H., L. Notley-McRobb, and T. Ferenci, *Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis.* J Bacteriol, 1998. 180(19): p. 5109-16.
47. Fiehn, O., *Metabolomics--the link between genotypes and phenotypes.* Plant Mol Biol, 2002. 48(1-2): p. 155-71.
48. Fiehn, O., et al., *Metabolite profiling for plant functional genomics.* Nat Biotechnol, 2000. 18(11): p. 1157-61.
49. Raamsdonk, L.M., et al., *A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations.* Nat Biotechnol, 2001. 19(1): p. 45-50.
50. Soga, T., et al., *Pressure-assisted capillary electrophoresis electrospray ionization mass spectrometry for analysis of multivalent anions.* Anal Chem, 2002. 74(24): p. 6224-9.
51. Soga, T., et al., *Simultaneous determination of anionic intermediates for Bacillus subtilis metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry.* Anal Chem, 2002. 74(10): p. 2233-9.
52. Soga, T., et al., *Quantitative metabolome analysis using capillary electrophoresis mass spectrometry.* J Proteome Res, 2003. 2(5): p. 488-94.
53. Ehrenberg, M., et al., *Systems biology is taking off.* Genome Res, 2003. 13(11): p. 2377-80.
54. Edwards, J.S. and B.O. Palsson, *The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities.* Proc Natl Acad Sci U S A, 2000. 97(10): p. 5528-33.
55. Reed, J.L., et al., *An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).* Genome Biol, 2003. 4(9): p. R54.
56. Forster, J., et al., *Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network.* Genome Res, 2003. 13(2): p. 244-53.
57. Shen-Orr, S.S., et al., *Network motifs in the transcriptional regulation network of Escherichia coli.* Nat Genet, 2002. 31(1): p. 64-8.
58. Allen, T.E., et al., *Genome-scale analysis of the uses of the Escherichia coli genome: model-driven analysis of heterogeneous data sets.* J Bacteriol, 2003. 185(21): p. 6392-9.
59. Reed, J.L. and B.O. Palsson, *Thirteen years of building constraint-based in silico models of Escherichia coli.* J Bacteriol, 2003. 185(9): p. 2692-9.
60. Hucka, M., et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.* Bioinformatics, 2003. 19(4): p. 524-31.
61. Tomita, M., et al., *E-CELL: software environment for whole-cell simulation.* Bioinformatics, 1999. 15(1): p. 72-84.
62. Mendes, P. and D.B. Kell, *MEG (Model Extender for Gepasi): a program for the modelling of complex, heterogeneous, cellular systems.* Bioinformatics, 2001. 17(3): p. 288-9.
63. Goryanin, I., T.C. Hodgman, and E. Selkov, *Mathematical simulation and analysis of cellular metabolism and regulation.* Bioinformatics, 1999. 15(9): p. 749-58.