

# Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF Archive): Unique Resources for Biological Research

Masanari KITAGAWA,<sup>1,4</sup> Takeshi ARA,<sup>2</sup> Mohammad ARIFUZZAMAN,<sup>1</sup> Tomoko IOKA-NAKAMICHI,<sup>3</sup>  
Eiji INAMOTO,<sup>3</sup> Hiromi TOYONAGA,<sup>3</sup> and Hirotada MORI<sup>1,2,3,\*</sup>

*Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan<sup>1</sup>, Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0035, Japan<sup>2</sup>, CREST, JST (Japan Science and Technology)<sup>3</sup>*

<sup>4</sup>Present address: Dragon Genomics Center, Takara Bio Inc. 7870-15, Sakura-cho, Yokkaichi, Mie 512-1211, Japan

(Received 4 June 2004; revised 28 July 2005; published online 9 January 2006)

## Abstract

Based on the genomic sequence data of *Escherichia coli* K-12 strain, we have constructed a complete set of cloned individual genes encoding Histidine-tagged proteins with or without GFP fused for functional genomic analysis. Each clone encodes a protein of predicted ORF attached by Histidines and seven spacer amino acids at the N-terminal end, and five spacer amino acids and GFP at the C-terminal end. *Sfi*I restriction sites are generated at both the N- and C-terminal boundaries of ORF upon cloning, which enables easy transfer of ORF to other vector systems by cutting with *Sfi*I. Expression of cloned ORF is under the control of an IPTG-inducible promoter, which is strictly repressed by *lacI<sup>q</sup>* repressor gene product. The set of cloned ORFs described here should provide unique resources for systematic functional genomic approaches including (i) construction of DNA microarray, (ii) production and purification of proteins, (iii) analysis of protein localization by monitoring GFP fluorescence and (iv) analysis of protein–protein interaction.

**Key words:** *Escherichia coli*; bacterial functional genomics; ORF clone; resource

## 1. Introduction

More than 200 microbial complete genome sequences are now available from publicly accessible databases, such as DDBJ at the National Institute of Genetics (<http://www.ddbj.nig.ac.jp>), GenBank at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) and EMBL-EBI at European Bioinformatics Institute (<http://www.ebi.ac.uk>).<sup>1</sup> Among them, two prokaryotes, the gram-positive bacterium *Bacillus subtilis* and the gram-negative bacterium *Escherichia coli*, as well as the unicellular eukaryote *Saccharomyces cerevisiae* have been used extensively as model organisms for basic biological research.<sup>2–4</sup> Complete genome sequence

revealed that although these microorganisms are among the most thoroughly studied genetic systems, <50% of their respective genes had been characterized experimentally. Genome sequence data are valuable resources not only for information complementary to traditional biological approaches but for development of new approaches known as ‘functional genomics’, a research area of computational analysis of complete genomes, followed by experimental testing of emerging hypotheses.<sup>5,6</sup>

Genome sequence information throws new light on the nature and interrelationship among bacterial genes, including ancestries, gene families, modules and motifs in comparative terms.<sup>7–9</sup> Comparison of the chromosomal location of genes within and between species might also reveal aspects of genome evolution and functional coupling between genes.<sup>10–12</sup> Detailed information on the structure and function of genome and of individual

Communicated by Michio Oishi

\* To whom correspondence should be addressed. Tel. +81-743-72-5660, Fax. +81-743-72-5669, E-mail: hmori@gtc.naist.jp

gene products in particular are most abundant and comprehensive for *E. coli*, and approximately half of the gene products have been characterized by using a variety of genetical, biochemical and molecular biological techniques. On the basis of extensive data resulting from these analyses, genes of known function were classified into a number of distinct categories.<sup>13</sup> The importance of *E. coli* genome as a leading model system seems to have gained rapid recognition for future studies of functional genomics and systems biology.

To further clarify the nature of genes with unknown function, however, a variety of resources, such as individual cloned genes, disruption or deletion mutants for each of the predicted genes would be particularly useful. Complete genome sequence data permitted to construct such resources in *Mycoplasma genitalium*, *Bacillus subtilis* and *Saccharomyces cerevisiae*.<sup>14–17</sup> Construction of genome-wide clones in *E. coli* has a long history including plasmid clones,<sup>18</sup> cosmid library,<sup>19–21</sup> and phage lambda clones.<sup>22,23</sup>

Furthermore, increasing number of proteins, called ‘moonlighting proteins’, having two or more distinct functions are accumulating recently.<sup>24</sup> It would therefore be a good opportunity to reevaluate reported function of known genes and to explore new function by using large-scale resources such as cloned genes reported here as well as deletion and other mutants.

Here, we describe a complete clone set of genes in *Escherichia coli* that should allow us to perform systematic functional analyses not only of genes of unknown function but of those of known function to survey their function. Each of the predicted ORF except start and stop codons was amplified by PCR followed by cloning into multicopy plasmid vector. Each product is Histidine tagged at its N-terminal end and GFP fused at its C-terminal. Expression of cloned genes is directed by the P<sub>T5-lac</sub> promoter that can be activated by IPTG but normally repressed by *lacI<sup>q</sup>* placed in *cis*.

## 2. Materials and Methods

### 2.1. Strains and growth conditions

*E. coli* K-12, strain AG1 [*recA1 endA1 gyrA96 thi-1 hsdR17(r<sub>K</sub><sup>-</sup>m<sub>K</sub><sup>+</sup>) supE44 relA1*] that exhibits high transformation efficiency, purchased from Stratagene, Inc. was used for all plasmid construction and other experiments.

Cells were usually grown in Luria–Bertani (LB) medium [1% Bacto Tryptone (BD Diagnostic Systems), 0.5% yeast extract (BD Diagnostic Systems), 0.5% NaCl] containing 50 µg/ml of ampicillin (Meiji Seika Kaisha, Ltd), 30 µg/ml of kanamycin (Wako Pure Chemical Industries, Ltd) or chloramphenicol (Nacalai Tesque, Inc.) as required. Isopropyl-β-D-thiogalactoside (IPTG) was used at 0.1 mM to induce expression of cloned gene.

### 2.2. PCR primers

A pair of PCR primers was designed to amplify each of the predicted ORFs of *E. coli* which starts from the second codon to the last amino acid codon omitting the initiation and termination codons. Each primer contains 3 or 2 additional bases at its 5′ end followed by 20 or 21 bases of ORF-specific sequence for the N- or C-terminus, respectively. Additional nucleotides were included to facilitate directional cloning and to generate *Sfi*I restriction sites. Thus, all N-terminal primers have the sequence 5′-GCC(20N)-3′, where 20N is specific for each ORF beginning at the second codon. Similarly, all C-terminal primers have the sequence 5′-(21N)CC-3′, where 21N is specific to each ORF ending with the last amino acid codon. The resulting PCR fragment has the sequence of 5′-GCC-2nd through amino acid codon -GG-3′. All primers were phosphorylated before use in PCR reaction.

In the following special cases, primer sequences were modified to circumvent fortuitous generation of *Sfi*I restriction sites. (i) If the sequence of ORF happens to be ATG NNN NNG GCC N...N NNN GGC CNN NNN TAA, the PCR-amplified fragment would be GCC NNN NNG GCC N...N NNN GGC CNN NNN GG and the resultant plasmid after cloning would have the sequence ‘...GGCCCTGAG(GGCC NNN NNG GCC) N...N NNN GGC CNN NNN (GGCCTATGCGGCC)...’, generating additional *Sfi*I sites (underlined) besides these at the peripheries (in parentheses). To avoid this complication, the G residue of the third NNG codon was replaced by an appropriate residue without changing the amino acid encoded, and the third residue of the last amino acid codon was changed from GGC to GGT.

If the third residue of the last amino acid codon is G, and the ORF sequence is ‘ATG NN...NN NNG TAA’, the PCR-amplified fragment would have the sequence ‘GCC ATG NN...NN NNG GG’, and this fragment has an *Sfi*I site even if the last G residue was unexpectedly removed by exonuclease activity contaminated with polymerase or restriction enzyme. To avoid this, the last amino acid codon was replaced by a synonymous codon which lacks G at the third residue. However, when the last codon was Met(ATG) or Trp(TGG), no modification was made because of the lack of done synonymous codon.

### 2.3. PCR amplification

PCR amplification reaction was performed in 96-formatted well plates. To minimize PCR errors during amplification, 1 U of KOD DNA polymerase (TOYOBO CO., Ltd.) was used in 25 µl reaction mixture containing 20–30 ng Kohara phage clone DNA or *E. coli* genomic DNA, 1.0 µM of each primer and 200 µM dNTPs. A recombinant form of *Thermococcus kodakaraensis* KOD DNA Polymerase having 3′→5′ exonuclease-dependent proofreading activity and exhibits very high

fidelity was used.<sup>25,26</sup> Reactions were performed according to manufacturer's instruction and the amplified fragments were blunt-ended by using KOD polymerase. PCR reactions were run by 25 cycles at 95°C for 15 s, 64°C for 15 s, 72°C for 4 min followed by a final incubation at 72°C for 5 min. All reactions were performed by an automatic sequencing reaction robot PRIZM877 (Applied BioSystems) to avoid errors arising from manual operation.

About 99% of ORFs (4217) could be successfully amplified by KOD DNA polymerase; many of the unsuccessful ORFs were >2 kb, and most of them (~90%) could be amplified by DynaZyme (Finzymes Oy). The remaining ORFs (~1% of total) could be prepared by using LATaq DNA polymerase (TAKARA BIO, Inc.).

#### 2.4. PCR fragment purification

All PCR products were purified by agarose gel electrophoresis (0.7–1.5% gel) in 1× TAE buffer depending on the length of amplified fragments. Gels were stained with SyberGOLD (Molecular Probes, Inc.) and visualized by 480 nm light to minimize DNA damage. Images were then photographed and checked for their sizes. PCR products of expected size were cut out and eluted by using MagExtractor (TOYOBO Co., Ltd) according to the manufacturer's instruction.

#### 2.5. Cloning of individual ORFs into pCA24N vector

DNA fragments purified by agarose gel electrophoresis were individually ligated with vector pCA24N DNA that had been digested with *Stu*I and dephosphorylated. Ligation was performed for 1 h at 16°C by using ligation kit (TAKARA BIO, Inc.). The ligated DNA was ethanol precipitated, washed with 70% ethanol, vacuum dried, and resuspended in 10  $\mu$ l of H<sub>2</sub>O. DNA solution (1  $\mu$ l) was used for electro-transformation, selecting for chloramphenicol resistance on LB plate. Competent cells (50  $\mu$ l) prepared according to Laboratory Manuals<sup>27</sup> were mixed with 1  $\mu$ l of DNA solution in an ice-cold 0.2-cm cuvette (Bio-Rad Laboratories, Inc.), treated at 2.5 kV with 25 mF and 200-ohms, followed by addition of 1 ml of SOC medium (2% Bacto Tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM glucose). Cells were then transferred to each of 96 formatted deep wells, and allowed to recover for 1 h at 37°C with shaking at 250 r.p.m. before plating on selective media.

#### 2.6. Confirmation of structure of ORF clones

Plasmid clones were purified by 96 well format Multi Screen Plasmid DNA purification kit (Millipore Corporation) and their structure was confirmed by digestion with *Bgl*II or *Sfi*I restriction enzyme and examining by agarose gel electrophoresis. Clones having expected gel pattern were further analyzed by direct sequencing of

PCR-amplified products made by using pCA-F primer 5'-GGCGTATCACGAGGCCCTTTCGTCTTACC-3' and pCA-R3 primer 5'-TTGCATCACCTTACCCTCTCCACTGACAG-3'. The primers used for sequencing were F-CA primer 5'-CATTAAAGAGGAGAAATTA-ACTATGAGAGG-3' from His-tagged side, and R-CA primer 5'-CATCTAATTCAACAAGAATTGGGACAACTC-3' from GFP side. The sequencing reaction was performed using ABI PRISM model 377, 3700 or 3730, according to the standard Big-dye protocol (Applied BioSystems).

#### 2.7. GFP fluorescence

Colonies were formed on LB plates containing 30  $\mu$ g/ml chloramphenicol with (1 mM) or without IPTG. GFP fluorescence of individual colonies was visualized by 480 nm light and photographed by IMAGE FREEZER AE-6905 (ATTO CORPORATION).

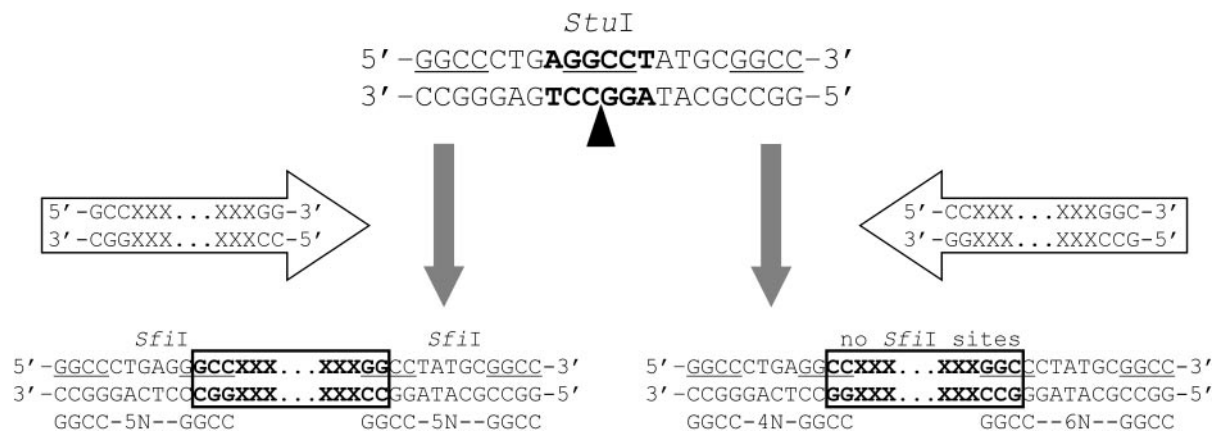
#### 2.8. Protein purification

Cells producing Histidine-tagged protein from the clone were grown at 37°C in 5 ml of LB medium supplemented with 30  $\mu$ g/ml chloramphenicol to OD<sub>600</sub> 0.3. Samples were taken 2 h after the addition of 1 mM IPTG. The cells were collected by centrifugation (10 000 r.p.m. × 3 min at 4°C) and resuspended in 400  $\mu$ l of cold buffer I [50 mM Sodium phosphate (pH 7.0), 200 mM NaCl, Proteinase inhibitor (Hoffmann-La Roche Ltd)]. Thereafter, all manipulations were done at 4°C. Crude cell extracts were obtained by sonication (5 × 5 s, level 3, Astrason ultrasonic processor) and centrifugation (16 000 r.p.m. × 15 min). Crude cell extracts were loaded onto a 30  $\mu$ l Nickel (Ni<sup>2+</sup>)-column [prepared according to manufacturer's instructions (QIAGEN, Inc) and equilibrated with buffer I]. Affinity chromatography of the extracts was performed at 4°C. Loaded columns were washed three times with 1 ml of buffer 2 [Buffer I contains 20 mM imidazole and 0.05% *n*-octyl- $\beta$ -glucoside (Pierce Biotechnology, Inc)] and proteins were eluted with buffer 3 [50 mM Tris-HCl (pH 6.8), 2% SDS, 0.1% bromophenol blue, 10% glycerol, 100 mM DTT, 6 M Urea and 250 mM imidazole]. Eluted proteins were analyzed by 7.5–15% gradient SDS-PAGE (BIOCRAFT Co., Ltd.) followed by Coomassie Brilliant Blue staining.

### 3. Results and Discussions

#### 3.1. Construction of cloning vector pCA24N

To clone all the genes of *Escherichia coli*, we constructed a plasmid vector with the following properties: (i) high copy number plasmid, (ii) IPTG-inducible expression of cloned ORF repression of expression by *lac*I<sup>r</sup>, (iii) Histidine tag attached to the N-terminal end of ORF, (iv) in-frame fusion with GFP at the C-terminal



**Figure 1.** Generation of *Sfi*I sites by cloning a target ORF only in correct direction. The top line is the cloning site of pCA24N. To clone a target ORF, blunt-ended PCR-amplified protein-coding region with three (GCC) and two (GG) additional nucleotides at the N- and C-terminal, respectively, was inserted into the *Stu*I site (filled triangle). Ligation of blunt-ended 5'-GCCXXX...XXXGG-3' fragment (X indicates nucleotides of target gene) generates *Sfi*I sites at both boundaries of the fragment when the ligation was done in correct direction (left) but not in opposite direction (right). Open arrows show PCR-amplified fragments with arrowhead indicating direction of the ORF.

end, (v) generation of *Sfi*I restriction sites at both boundaries of cloned ORF (Fig. 1) (see below). The construction of pCA24N is shown schematically in Fig. 2. Expression of a target ORF inserted at the *Stu*I site is directed by the IPTG-inducible promoter,  $P_{T5-lac}$ . Each ORF is fused in-frame with Histidine tag 7 spacer amino acids at the N-terminal end (Fig. 3).

The gene encoding a mutated form of GFP<sup>28</sup> was placed immediately downstream of ORF so as to produce fusion protein. This is mainly for detecting localization of protein product in the host cell. GFP was also supposed to serve as an indicator for correct cloning of amplified ORF DNA. It was designed, however, to permit removal of the ORF after removing GFP by digestion with *Not*I enzyme followed by self-ligation. The final structure of the cloned ORF after removing GFP is shown in Fig. 4. Clones that have *Not*I restriction site(s) within ORF can be prepared separately by using partial digestion with *Not*I enzyme.

### 3.2. Cloning into pCA24N

The number of ORFs used as initial target cloning was 4276 excluding ORFs coded by IS elements. Most of them (4267 out of 4276 ORFs) were successfully cloned into pCA24N in-frame and in correct orientation: 3 out of 9 unsuccessful ORFs, *yhcQ*, *yibP* and *aidB*, could be cloned only in opposite direction probably because even small amount of leak expression from  $P_{T5-lac}$  promoter or upstream region is toxic for the host cell. Remaining six ORFs, *eaeH* (7104 bp), *yheB* (2694 bp), *yhiH* (2685 bp), *yagF* (1968 bp), *ydbD* (2313 bp) and *btuB* (1845 bp), were failed to be amplified by PCR reaction partly due to their length to be amplified. *btuB* mRNA was reported to form complex structure to control binding to ribosome and its translation.<sup>29</sup> This might interfere to be amplified by PCR reaction.

The purified PCR-amplified fragments were then ligated into *Stu*I site of pCA24N vector. After overnight incubation at 37°C, eight colonies were picked out for each ORF and suspended in 1 ml fresh LB medium in 96 formatted deep well, and were also streaked on 1 mM IPTG containing LB plates at 37°C to check in-frame cloning by observing growth and GFP fluorescence upon overproduction of GFP fusion protein encoded by each ORF. After overnight incubation at 37°C, plasmid DNA was extracted and purified by using Multi Screen DNA purification kit, digested with *Sfi*I or *Bgl*II restriction enzymes, and analyzed by agarose gel electrophoresis. Four independent clones, that have the expected structures, were chosen and stored as a DNA mixture to minimize loss of intrinsic genetic information during PCR amplification. Boundaries of cloning sites were confirmed by sequencing.

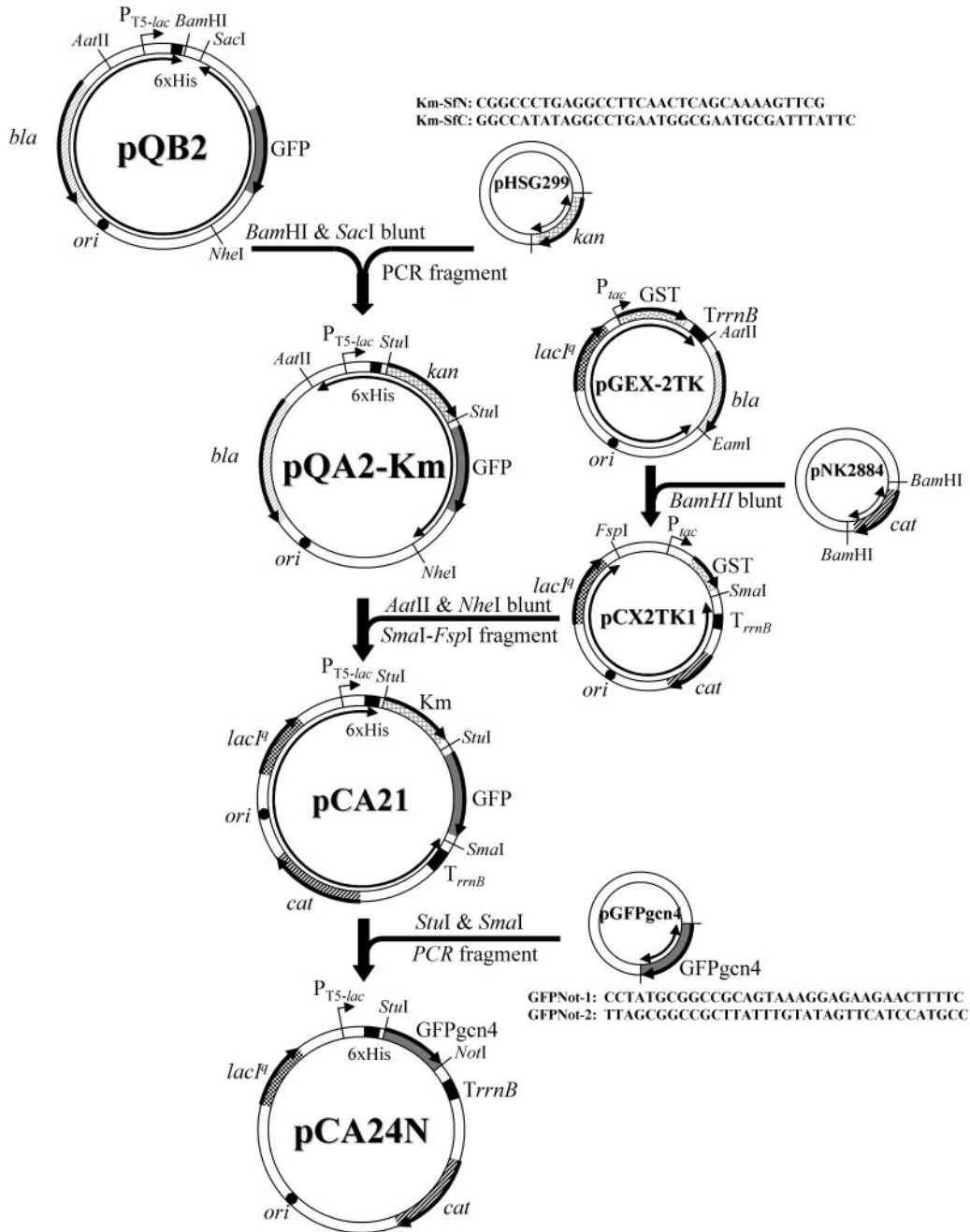
### 3.3. Growth inhibition by IPTG induction

To investigate the influence of over-supply of target gene products, we induced expression of ORF by growing patches of cells on LB agar medium supplemented with or without 1 mM IPTG at 37°C. Out of 4269 ORFs tested, 3301 ORFs showed some growth inhibitory effects by IPTG induction (Fig. 5). Out of these ORFs, 2149 ORFs showed severe growth defects: 1158 ORFs were predicted to have membrane-spanned domain and as membrane proteins. Of 1158 plausible membrane proteins, 1032 ORFs showed severe growth defects when overproduced.

### 3.4. Use of ORF clones

**Construction of DNA microarray:** First, we tried to construct cDNA microarray by using ORF clones as DNA templates for PCR amplification. These clones not

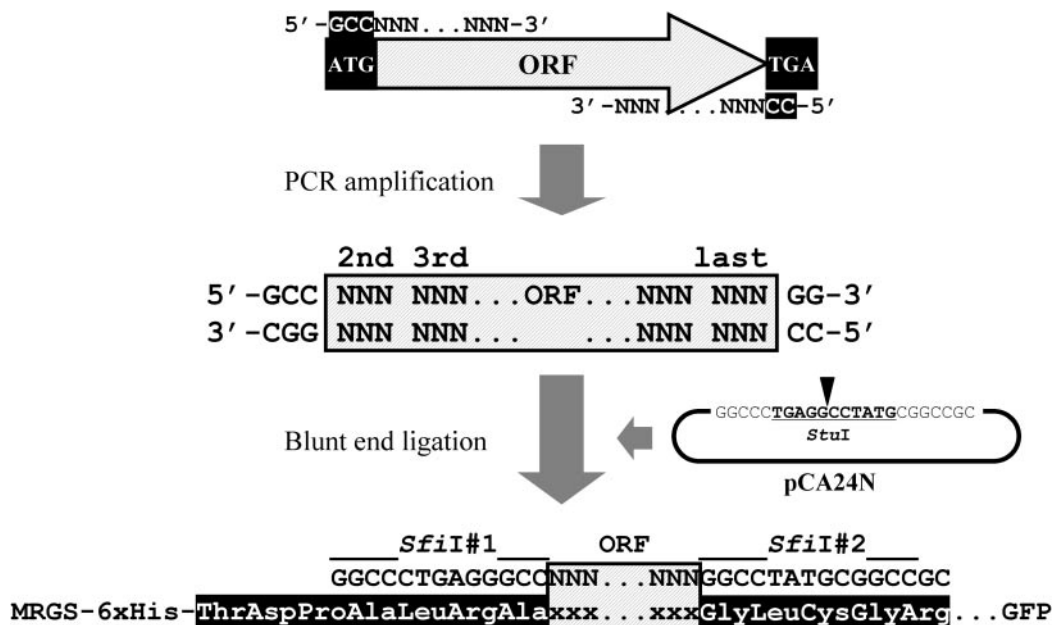




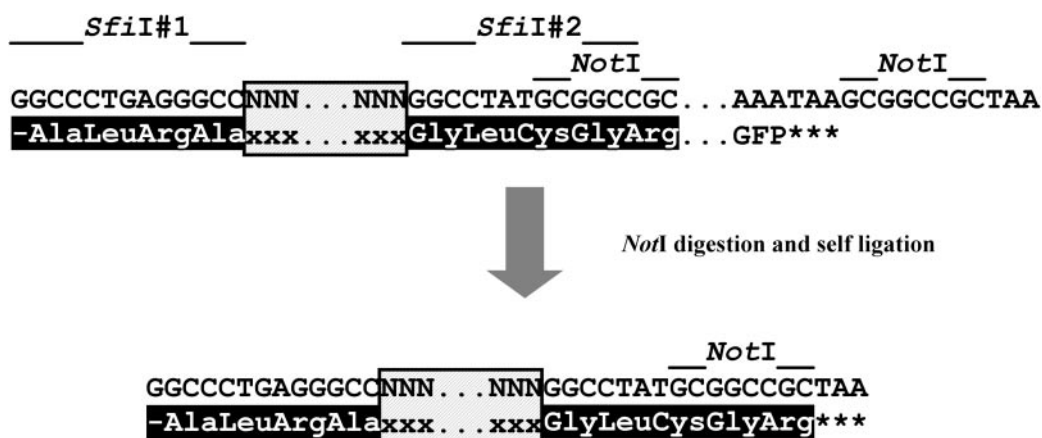
**Figure 2.** Vector construction. The blunt-ended 4.1 kb *Bam*HI and *Sac*I fragment of pQB2,<sup>39</sup> was ligated with the kanamycin resistance fragment (*kan*) that had been amplified from pHSG299,<sup>40</sup> by using the oligo DNA primers: 5'-CGGCCCTGAGGCCTTCAACTCAGCAAAAAGTTTCG-3' and 5'-GGCCATATAGGCCTGAATGGCGAATGCGATTATTC-3' to create pQA2-Km. To prepare a fragment having a *lacI<sup>q</sup>* and terminator, pGEX-2TK,<sup>41</sup> was digested by *Aat*II and *Eam*I and blunt ended. The resulting *lacI<sup>q</sup>* fragment was ligated with blunt-ended *Bam*HI chloramphenicol resistance fragment of pNK2884,<sup>42</sup> to create pCX2TK1. Blunt-ended *Aat*II and *Nhe*I fragment of pQA2-Km and *Sma*I-*Fsp*I blunt-end fragment was ligated to create pCA21. Original GFP is heat labile and have weak or no fluorescence at 37°C or higher temperature. To replace the original GFP by the heat tolerant mutant form, the *Stu*I-*Sma*I fragment of pCA21 was ligated with PCR-amplified GFP fragment from pGFPgcn4,<sup>28</sup> by oligo DNA primers having *Not*I restriction site: 5'-CCTATGCGGCCGCGAGTAAAGGAGAAGAACTTTTC-3' and 5'-TTAGCGGCCGCTTATTTGTATAGTTCATCCATGCC-3'. *Not*I sites of both primers were designed for removal of GFP after cloning. The replication origin and Histidine tag of the resultant vector, named pCA24N, were derived from pGEX-2TK and pQE31 (parental plasmid of pQB2) respectively.

only function as templates for PCR amplification by one set of common oligo DNA primers for every target ORF, but are also found to serve as an appropriate source for DNA template and provided a stable supply of

DNA fragment. *E. coli* DNA microarray is now commercially available from TAKARA BIO, Inc. The applications using these microarrays have been already published.<sup>30-34</sup>



**Figure 3.** Predicted amino acid sequences at around the N- and C-terminal regions of the cloned ORF. The arrow indicates a target ORF and a pair of primers used for PCR amplification. The bottom nucleotide and amino acid sequences indicate the predicted final structure around the site of cloned ORF. The product proteins should contain 6 Histidine, 7 and 5 amino acids at the N- and C-terminal ends of the target gene, respectively, followed by GFP fragment: 6xHisThrAspProAlaLeuArgAlaXXX...XXXGlyLeuCysGlyArg...GFP, where XXX...XXX indicates second to the last amino acid codon of target gene.



**Figure 4.** Removal of GFP fragment by *Not*I digestion. Each ORF clone has two *Not*I sites, one at the C-terminal spacer region (within *Sfi*I #2) and region directly next to the termination codon of GFP gene. GFP can therefore be removed by *Not*I digestion followed by self-ligation.

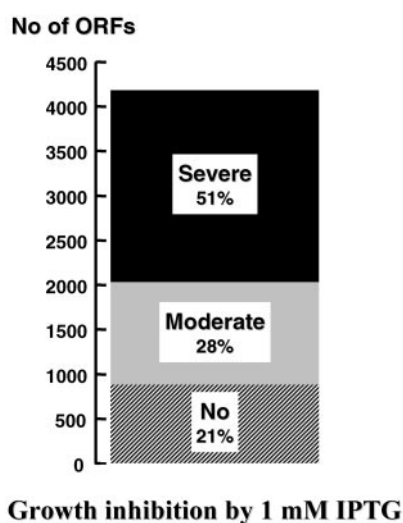
**Production of purified protein:** The present vector was designed as an efficient expression vector, and protein product of each clone can be easily induced by IPTG in quantity and purified by using Ni- or Co-NTA column (Fig. 6). Some proteins, especially most of the membrane proteins, had difficulties to be purified by this method. Few cytosolic proteins showed similar difficulties to be purified. One possibility is that His-tag could not function because of N-terminal 3D-structure. His-tag proteins attached with its C-terminal should be also considered. These purified proteins are also good resource for further analysis, such as protein biochemistry, crystallography and antigens for antibody production. Furthermore, ana-

lysis of co-purified proteins with His-tagged cloned ORF would reveal candidates of interacting proteins. This analysis by using mass-spectrometry for efficient identification of interacting proteins is now underway systematically.

**Protein localization detected by fluorescence from GFP:** Protein localization is undoubtedly one of the important information about protein product to understand its function. When induced by adding IPTG, however, 62% of total ORF clones showed growth inhibition to the host cell. And in some cases, IPTG induction might lead to form inclusion body. To avoid

these artificial effects, cells were grown without IPTG. pCA24N vector has *lacI<sup>f</sup>* for strict suppression of the expression from  $P_{T5-lac}$  promoter without IPTG growth condition; however, very few proportion of the cells showed fluorescence from GFP probably because of leaky expression. After pilot tests by using clones that were known for their cellular localization, this analysis was carried out on the systematic stage (Niki, H., personal communication and kindly provided pictures, Fig. 7).

**Functional analysis of ORF:** Deletion mutation of the target gene might reveal its physiological function by loss of function but clone might also give us many suggestions of its function by over supply of the target gene



**Figure 5.** Effects on growth and GFP fluorescence by IPTG induction. Effects of growth by 1 mM IPTG induction on LB agar plate at 37°C were classified into three categories. The numbers represent the ratio of each of the classified ORFs against the total ORFs of *E. coli*.

product. Applications by using clones have been already published elsewhere.<sup>35-38</sup>

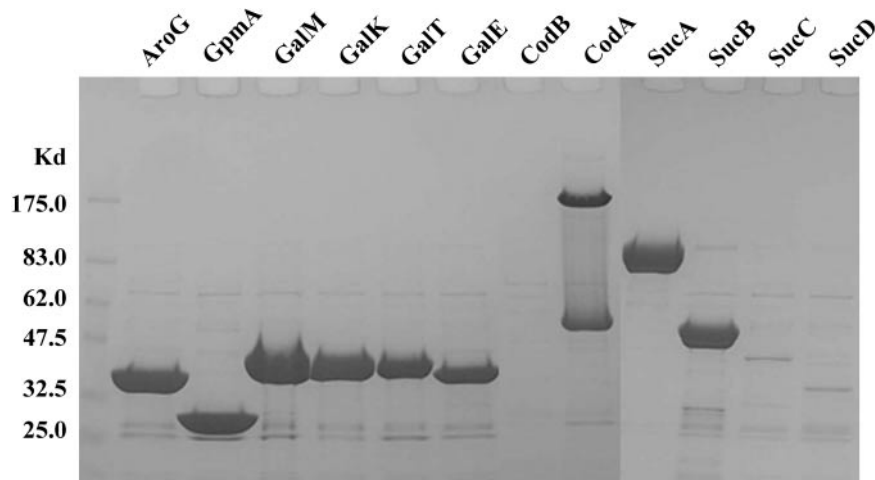
### 3.5. Distribution of clones

Clones are freely available to use for academic purposes with agreement for material transfer and planned to be distributed from multiple sites in Japan and other countries. Initially, these clones are available from Nara Institute of Science and Technology and requests are accepted through our web page (<http://ecoli.aist-nara.ac.jp/Resource/ResourceManage.jsp>).

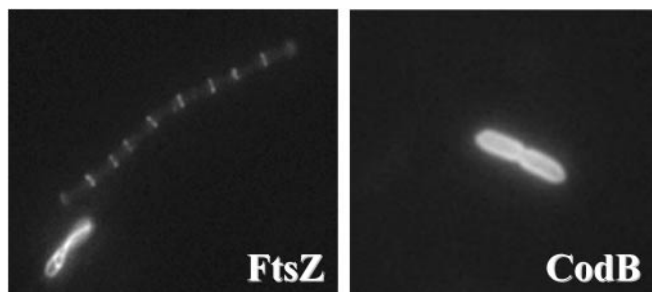
### 3.6. Quality control and related information

To fix the ORF prediction, not only for functional annotation but also for coding region, from a genome sequence is difficult problem and always under evolving stage. To improve the quality of our plasmid clone library, we have been keeping efforts to re-construct target ORF clones based on the latest ORF prediction.

In November 2003, the annotation of *E. coli* genes by international consortium was started according to the correction of genome sequence of *E. coli* by Japanese group organized by Horiuchi (Hayashi, K., Morooka, N., Otsubo, E. et al., manuscript in preparation). Recently, more accurate ORF prediction and annotation was fixed as the first version by the consortium (Riley, M. et al., manuscript in preparation). The total number of ORFs or ORF fragments of *E. coli* K-12 W3110 strain is 4364. Out of 4364 ORFs, 77 were predicted as pseudogenes caused by IS insertion or frameshift mutation. Excluding these pseudogenes and IS or phage-related ORFs, 3986 newly assigned ORFs are required to be established as plasmid clones as ASKA library. According to this information, we found ~900 plasmid clones in our library to be required for minor modification mainly at their start sites, and almost all have been



**Figure 6.** Protein purification by using Ni-NTA column after IPTG induction. The plasmids used to over-express Histidine-tagged proteins were from ASKA library. From 3 ml culture induced by adding 0.1 mM IPTG for 2 h, His-tagged proteins were purified by Ni-NTA column. The whole eluted proteins were analyzed by SDS-PAGE and visualized by Coomassie Brilliant Blue. Some proteins (CodB, SucC and SucD) were hardly purified by this method.



**Figure 7.** GFP fluorescence image of a cell. These images of cells having an *ftsZ* and *codB* gene of our clones were taken by CCD camera without any induction by adding IPTG. The GFP fluorescence is from leaks of transcription from the promoter. The result is consistent with previous observation.<sup>43</sup> Pictures were kindly provided by Dr. Niki, National Institute of Genetics, Mishima, Japan.

already modified and ready for use. At the same time, we tried to read whole sequence of the fragment cloned to check the PCR error and keep one correct clone if available.

As mentioned in Materials and Methods, previous stock of clones are mixture of multiple up to four independent candidates of each target ORF, that had been confirmed their structure by restriction enzyme and sequencing from both side. Some of them might have PCR errors and we will keep efforts to eliminate such kind of errors by sequencing and re-cloning. The latest information about ASKA library is available through our web page (<http://ecoli.aist-nara.ac.jp>).

**Acknowledgements:** This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan, a grant from CREST, JST (Japan Science and Technology) and in part from NEDO (New Energy and Industrial Technology Development Organization) and Inamori Foundation. We thank Hironori Niki (National Institute of Genetics) for providing pictures of protein localization. We also thank Takashi Yura, Katsumi Isono (Kobe University) and Takashi Horiuchi (The Institute for Basic Biology) for manuscript preparation. Funding to pay the Open Access publication charges for this article was provided by NEDO.

## References

- Janssen, P., Audit, B., Cases, I., et al. 2003, Beyond 100 genomes, *Genome Biol.*, **4**, 402.
- Kunst, F., Ogasawara, N., Moszer, I., et al. 1997, The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*, *Nature*, **390**, 249–256.
- Blattner, F. R., Plunkett, G., III, Bloch, C. A., et al. 1997, The complete genome sequence of *Escherichia coli* K-12, *Science*, **277**, 1453–1474.
- Mewes, H. W., Albermann, K., Bahr, M., et al. 1997, Overview of the yeast genome, *Nature*, **387**, 7–65.
- Storms, R. K., Wang, Y., Fortin, N., et al. 1997, Analysis of a 103 kb cluster homology region from the left end of *Saccharomyces cerevisiae* chromosome I, *Genome*, **40**, 151–164.
- Galperin, M. Y and Koonin, E. V. 1999, Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes, *Genetica*, **106**, 159–170.
- Labadan, B. and Riley, M. 1995, Gene products of *Escherichia coli*: sequence comparisons and common ancestries, *Mol. Biol. Evol.*, **12**, 980–987.
- Tatusov, R. L., Koonin, E. V., and Lipman, D. J. 1997, A genomic perspective on protein families, *Science*, **278**, 631–637.
- Riley, M. and Labedan, B. 1997, Protein evolution viewed through *Escherichia coli* protein sequences: introducing the notion of a structural segment of homology, the module, *J. Mol. Biol.*, **268**, 857–868.
- Zipkas, D. and Riley, M. 1975, Proposal concerning mechanism of evolution of the genome of *Escherichia coli*, *Proc. Natl Acad. Sci. USA*, **72**, 1354–1358.
- Overbeek, R., Fonstein, M., D'Souza, M., et al. 1999, The use of gene clusters to infer functional coupling, *Proc. Natl Acad. Sci. USA*, **96**, 2896–2901.
- Uchiyama, I. 2003, MBGD: microbial genome database for comparative analysis, *Nucleic Acids Res.*, **31**, 58–62.
- Riley, M. 1993, Functions of the gene products of *Escherichia coli*, *Microbiol. Rev.*, **57**, 862–952.
- Hutchison, C. A., Peterson, S. N., Gill, S. R., et al. 1999, Global transposon mutagenesis and a minimal Mycoplasma genome, *Science*, **286**, 2165–2169.
- Kobayashi, K., Ehrlich, S. D., Albertini, A., et al. 2003, Essential *Bacillus subtilis* genes, *Proc. Natl Acad. Sci. USA*, **100**, 4678–4683.
- Winzeler, E. A., Shoemaker, D. D., Astromoff, A., et al. 1999, Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis, *Science*, **285**, 901–906.
- Vidan, S. and Snyder, M. 2001, Large-scale mutagenesis: yeast genetics in the genome era, *Curr. Opin. Biotechnol.*, **12**, 28–34.
- Nishimura, A., Akiyama, K., Kohara, Y., et al. 1992, Correlation of a subset of the pLC plasmids to the physical map of *Escherichia coli* K-12, *Microbiol. Rev.*, **56**, 137–151.
- Tabata, S., Higashitani, A., Takanami, M., et al. 1989, Construction of an ordered cosmid collection of the *Escherichia coli* K-12 W3110 chromosome, *J. Bacteriol.*, **171**, 1214–1218.
- Knott, V., Blake, D. J., and Brownlee, G. G. 1989, Completion of the detailed restriction map of the *E. coli* genome by the isolation of overlapping cosmid clones, *Nucleic Acids Res.*, **17**, 5901–5912.
- Birkenbihl, R. P. and Vielmetter, W. 1989, Cosmid-derived map of *E. coli* strain BHB2600 in comparison to the map of strain W3110, *Nucleic Acids Res.*, **17**, 5057–5069.
- Kohara, Y., Akiyama, K., and Isono, K. 1987, The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library, *Cell*, **50**, 495–508.
- Daniels, D. L. and Blattner, F. R. 1987, Mapping using gene encyclopaedias, *Nature*, **325**, 831–832.



24. Jeffery, C. J. 1999, Moonlighting proteins, *Trends Biochem. Sci.*, **24**, 8–11.
25. Nishioka, M., Mizuguchi, H., Fujiwara, S., et al. 2001, Long and accurate PCR with a mixture of KOD DNA polymerase and its exonuclease deficient mutant enzyme, *J. Biotechnol.*, **88**, 141–149.
26. Takagi, M., Nishioka, M., Kakihara, H., et al. 1997, Characterization of DNA polymerase from *Pyrococcus sp.* strain KOD1 and its application to PCR, *Appl. Environ. Microbiol.*, **63**, 4504–4510.
27. Sambrook, J., Fritsch, E. F., and Maniatis, T. 1998, *Molecular Cloning: A Laboratory Manual*. 2nd Ed.. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
28. Ito, Y., Suzuki, M., and Husimi, Y. 1999, A novel mutant of green fluorescent protein with enhanced sensitivity for microanalysis at 488 nm excitation, *Biochem. Biophys. Res. Commun.*, **264**, 556–560.
29. Nahvi, A., Sudarsan, N., Ebert, M. S., et al. 2002, Genetic control by a metabolite binding mRNA, *Chem. Biol.*, **9**, 1043.
30. Oshima, T., Aiba, H., Masuda, Y., et al. 2002, Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12, *Mol. Microbiol.*, **46**, 281–291.
31. Oshima, T., Wada, C., Kawagoe, Y., et al. 2002, Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*, *Mol. Microbiol.*, **45**, 673–695.
32. Nakahigashi, K., Kubo, N., Narita, S., et al. 2002, HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release factors, and *hemK* knockout induces defects in translational termination, *Proc. Natl Acad. Sci. USA*, **99**, 1473–1478.
33. Kodama, K., Kobayashi, T., Niki, H., et al. 2002, Amplification of Hot DNA segments in *Escherichia coli*, *Mol. Microbiol.*, **45**, 1575–1588.
34. Minagawa, S., Ogasawara, H., Kato, A., et al. 2003, Identification and molecular characterization of the Mg<sup>2+</sup> stimulon of *Escherichia coli*, *J. Bacteriol.*, **185**, 3696–3702.
35. Kawano, M., Oshima, T., Kasai, H., et al. 2002, Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35-amino-acid cell-killing peptide and a cis-encoded small antisense RNA in *Escherichia coli*, *Mol. Microbiol.*, **45**, 333–349.
36. Awano, N., Wada, M., Mori, H., et al. 2005, Identification and functional analysis of *Escherichia coli* cysteine desulfhydrases, *Appl. Environ. Microbiol.*, **71**, 4149–4152.
37. Melnick, J., Lis, E., Park, J. H., et al. 2004, Identification of the two missing bacterial genes involved in thiamine salvage: thiamine pyrophosphokinase and thiamine kinase, *J. Bacteriol.*, **186**, 3660–3662.
38. Proudfoot, M., Kuznetsova, E., Brown, G., et al. 2004, General enzymatic screens identify three new nucleotidases in *Escherichia coli*. Biochemical characterization of SurE, YfbR, and YjjG, *J. Biol. Chem.*, **279**, 54687–54694.
39. Kimata, Y., Iwaki, M., Lim, C. R., et al. 1997, A novel mutation which enhances the fluorescence of green fluorescent protein at high temperatures, *Biochem. Biophys. Res. Commun.*, **232**, 69–73.
40. Takeshita, S., Sato, M., Toba, M., et al. 1987, High-copy-number and low-copy-number plasmid vectors for *lacZ* alpha-complementation and chloramphenicol- or kanamycin-resistance selection, *Gene*, **61**, 63–74.
41. Smith, D. B. and Johnson, K. S. 1988, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase, *Gene*, **67**, 31–40.
42. Kleckner, N., Bender, J., and Gottesman, S. 1991, Uses of transposons with emphasis on Tn10, *Methods Enzymol.*, **204**, 139–180.
43. Ma, X., Ehrhardt, D. W., and Margolin, W. 1996, Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living *Escherichia coli* cells by using green fluorescent protein, *Proc. Natl Acad. Sci. USA*, **93**, 12998–13003.