A Rapid Method for Disrupting Genes in the Escherichia coli Genome

Chinatsu Kato, Ryusuke Ohmiya, and Takeshi Mizuno*
Laboratory of Molecular Microbiology, School of Agriculture, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan

Received April 28, 1998

The entire genomic sequence of Escherichia coli has recently been completed. To gain insight into the function of the vast array of yet uncharacterized open reading frames (ORFs), a variety of new genetic tools will be required. Here we examined a genetic system, using an integration plasmid vector (named pINT007), for rapid construction of disruption mutants of any ORF in E. coli. It was found that the vector allows us to rapidly construct a disruption mutant for any gene on the chromosome as a co-integrate, furthermore, resolution of the resulting co-integrate can be surely accomplished by using a pair of the bla (ampicillin resistant) genes on the vector as a positive-selection marker.

Key words: E. shercichia coli; gene disruption; genome analysis; plasmid vector

The entire genomic sequences have recently been completed for a number of bacterial species including Escherichia coli, Bacillus subtilis, Haemophilus influenzae, and Methanococcus jannaschii (archaeal species). Among them, E. coli is obviously the organism of choice to be learned more deeply through systematic post-sequencing programs, because we know so much more about the biology of E. coli than about any other living organism. The paper by Blattner et al., showed that there are about 4300 genes (or open reading frames, ORFs) in the entire genome of E. coli. Importantly, however, about 40 percent of these sequences are not obviously related to genes of known function. To gain insight into the functions of these uncharacterized genes, the post-sequencing approaches should include a systematic isolation of disruption (or deletion) mutants for the vast array of these uncharacterized ORFs. To do so, a variety of new genetic tools will be required. Here we examined a genetic system for rapidly constructing disruption mutants of any gene in wild-type E. coli.

Gene replacement on the E. coli chromosome can be done by a variety of techniques. One of the ways is called linear DNA-transformation. An appropriate linear DNA, containing a mutated or deleted gene flank ed by homologous regions of the chromosome, is transferred into recombination-proficient strains, such as recBC, sbcBC, or recD. Double cross over recombination between the E. coli chromosome and both ends of the linear DNA fragment results in gene replacement in these particular genetic backgrounds at a high frequency. A disadvantage of this method is that it is restricted to specific nuclease-deficient recombination-proficient genetic backgrounds. Another general method involves co-integration of a plasmid containing the mutated gene of interest into the E. coli chromosome by single cross-over recombination, followed by resolution (or excision) of the co-integrate (see below). The key to this procedure is the use of appropriate vectors that cannot replicate under the conditions used for selection of the co-integrate. The best-used vectors include ColE1-derived plasmids that do not replicate in polA mutants, a temperature-sensitive pSC101 replicon, and a phage-based vector. Although such a co-integration scheme has been successfully and widely used, a problem is that resolution of the co-integrate occurs at a relatively low frequency and may not always give the replacement desired. However, if there is a positive-selection device for monitoring resolution of the excised vector, this problem would be overcome. Recently, Link et al., reported such a integration vector that carries a positive-selection marker, the sacB gene encoding levansucrase. Previously, we also described briefly another temperature-sensitive integration vector (named pINT-vector), carrying a sophisticated positive-selection maker for the resolution step. In light of the completion of the E. coli genomic sequence, here we wanted to evaluate our vector more extensively, to see whether or not it can be used for rapidly constructing disruption mutants for any gene of interest.

Our integration vector (named pINT007) has the following characteristics (Fig. 1). It carries a pSC101-derived temperature-sensitive replicon (OriS) and the neo gene (Km, kanamycin resistance). It also carries a pair of the bla gene (Ap, ampicillin resistance), nonetheless, neither of which are functional; one of the bla genes (bla-ΔN) is truncated at the 5'-proximal region of its coding-sequence and the other has a frame shift mutation in the 3'-proximal PstI site (bla-ΔC). It should be emphasized that the OriS-replicon is located between these inactivated bla genes, which are tandemly aligned on pINT007. Provided that the events shown in Fig. 1A take place in E. coli, use of pINT007 should enable us to construct a desired disruption mutation of any gene (say gene A) on the chromosome. First, a DNA fragment encompassing the gene A coding-sequence, which is truncated at both the 5'- and 3'-ends, is cloned into the multi-cloning sites (MCS) (step 1 in Fig. 1A). The resulting pINT007-derivative is then transferred into a wild-type E. coli strain at 30°C. The transformant is allowed briefly to grow in a liquid medium (e.g., Luria-broth) containing kanamycin at 30°C. The cells are then plated

* To whom correspondence should be addressed: Takeshi Mizuno; Tel: +81-52-789-4089; Fax: +81-52-789-4091
Fig. 1. Schematic Representation of the Strategy Used for the Construction of Disruption Mutants on the E. coli Chromosome. (A) The structure of the integration plasmid-vector (pINT007) is presented. This plasmid carries the temperature-sensitive replication origin, the neo gene, the inactivated bla genes, and multi-cloning sites (MCS). (B) A set of lacZ coding-sequences with varied lengths were inserted into the multi-cloning sites, as indicated. Other details are given in the text.

on kanamycin plates at the nonpermissive temperature (42°C) to select for chromosomal cointegrates as Km− cells (step 2). The presumed cointegrates are picked up, and then allowed to grow in a liquid medium containing kanamycin at 30°C. To select certain cells in which the cointegrated plasmid is excised via homologous recombination between the bla−ΔN and bla−ΔC sequences, the culture is plated on ampicillin plates at 30°C. Only the cells that have excised the vector sequences containing the OriTS-replicon should gain the functional bla gene and be able to grow under these conditions (step 3). In the selected cells, the gene replacement event should have occurred, and they should carry a disruption mutation in the gene-A coding-sequence on the chromosome.

To evaluate this apparently simple procedure, we did a model experiment using the lacZ gene encoding β-galactosidase. A series of pINT007-derivatives were constructed, in which a set of truncated lacZ coding-sequences with varied lengths in nucleotides were cloned, as schematically shown in Fig. 1B. Polymerase chain reaction (PCR) was used for preparing these DNA segments, of which the 5′-proximal regions are shortened successively, and the 3′-proximal 48-bp coding-sequence is missing. By using these plasmids, the experimental procedures, explained above, were done on MacConkey lactose plates containing appropriate antibiotics. The results for the selection of cointegrates (step 2) are summarized in Table 1, in which is showing the following. (i) When lacZ-DNA segments longer than 1027 bp were used, desired Km−cointegrates (i.e., LacZ+ white colonies) were easily obtained at high frequencies (1.4 × 10−4 to 3.1 × 10−4). (ii) When the pINT007-deriva-
tive carrying the 530-bp lacZ-DNA segment was used, however, such Km−cointegrates appeared on plates at a very low frequency (6 × 10−6), (iii) Furthermore, undesired Km−cointegrates (i.e., LacZ+ red colonies) appeared on plates in all the cases, even when pINT007 itself was used, albeit with very low frequencies. These lacZ+ Km−cointegrates were most likely generated through recombination on the chromosome elsewhere, outside of the lacZ gene (note that when we attempted to detect plasmid DNA in this particular type of cells, no plasmid DNA was detected in most of them).

For each case, several desired LacZ− Km−cointegrates were selected. They were subjected to the next procedure, step-3, to finally resolve the OriTS plasmid DNA from the chromosome. These cells were cultivated
in a liquid medium containing kanamycin at 30°C, and then plated directly on ampicillin plates at 30°C. In each case, Ap’ colonies appeared on the plates at low but statistically constant frequencies (3 × 10^-8 to 5 × 10^-8). Of course, they turned out to be LacZ^- and Km'. Although they are most likely the mutant strains that we originally wanted to establish, we needed to prove this. To do so, the chromosomal DNAs were isolated from each candidate, and then hybridized by Southern’s method with an appropriate DNA probe containing a portion of the lacZ gene. As shown in Fig. 2, each candidate had the hybridization profile, exactly as expected for each lacZ disruption mutant. We succeeded in selecting a set of lacZ disruption mutants using the sophisticated plasmid vector, but through the very simple procedures.

Since there is no reason to assume that this particular integration vector can be used only for the lacZ gene, unless the target gene is essential for the viability of E. coli, pINT007 should allow us to rapidly construct a disruption mutant for any gene on the chromosome as a cointegrate. More importantly, resolution of the resulting cointegrate can be surely accomplished by using a pair of the bla gene-derivatives on the vector as a positive-selection marker.

However, during the course of this study, we noticed several limitations (or disadvantages) of our tool, which we should finally discuss. At step 2, a homologous DNA insert longer than 1000-bp appears to be required to select desired cointegrates at a reasonable frequency. Undesired cointegrates occur at a certain frequency at the chromosomal regions elsewhere. Thus, if a homologous sequence in pINT007 was shorter than 500 bp, the numbers of true cointegrates obtained were smaller than those of the background cointegrates (Table 1). This may be a serious limitation, when considered the fact that, of total 4300 protein-coding sequences in the E. coli genome, about 27% are shorter than 500-bp, and a large portion of them (42%) are annotated as either “unknown” or “hypothetical”. Second, any established disruption mutant should have both the neo and bla genes within a given coding sequence, thereby resulting in the phenotype of Km' and Ap' multidrug resistance. This may also be a disadvantage, when we wanted to further manipulate this particular mutant by introducing another plasmid vector. This problem can partly be overcome by placing the neo gene between the two bla genes and next to the Ori^8 replicon (see Fig. 1). Such a new vector is now under construction. Finally, it should be discussed about the stability of a given disruption mutant. One can assume that if the mutant cells were grown in a medium lacking kanamycin and/or ampicillin, the integrated sequence would easily be removed, and the disrupted gene would revert to the original one. We tested this possibility, and found that this is indeed the case (data not shown). Therefore, another limitation is that resulting disruption mutants should always be maintained in an appropriate medium containing the selective antibiotics. Keeping these in mind, nevertheless, pINT007 is a useful genetic tool to gain quickly and systematically an idea about the functions of the vast array of uncharacterized ORFs found within the E. coli genome.

Acknowledgment
This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

References
9) Hamilton, C. M., Aldea, M., Washburn, K. B., Babitzke, P.,

