The molecular toolbox for chromosomal heterologous multiprotein expression in *Escherichia coli*

Katrin Richter and Johannes Gescher

Institut für Angewandte Biowissenschaften, Angewandte Biologie, Karlsruher Institut für Technologie, Fritz-Haber-Weg 2, D-76131 Karlsruhe, Germany

**Abstract**

Heterologous multiprotein expression is the tool to answer a number of questions in basic science as well as to convert strains into producers and/or consumers of certain compounds in applied sciences. Multiprotein expression can be driven by plasmids with the disadvantages that the gene dosage might, in some cases, lead to toxic effects and that the continuous addition of antibiotics is undesirable. Stable genomic expression of proteins can forgo these problems and is a helpful and promising tool in synthetic biology. In the present paper, we provide an extract of methods from the toolbox for chromosome-based heterologous expression in *Escherichia coli*.

**Introduction**

Synthetic biology is an emerging field in applied sciences. In most cases, it aims to programme cells for the production of compounds and/or the consumption of certain substrates. This programming is dependent on the genetic manipulation of the host organism and often the heterologous expression of multiple proteins within one cell. In the present review, we provide a survey of the molecular toolbox we have in our hands for the controlled and simultaneous heterologous expression of multiple proteins from the chromosome of the model organism *Escherichia coli*. The present paper is inspired by our attempts to programme *E. coli* to become an efficient dissimilatory metal-reducing organism.

**Genomic deletions and insertions**

**Recombineering**

The construction of mutants that express heterologous genes from the chromosome requires efficient allelic exchange procedures. Most of the methods used rely on phage-based homologous recombination systems, and the process itself is often called ‘recombineering’ (recombination-mediated genetic engineering). Here, a single-stranded annealing protein and an exonuclease catalyse the recombination between the host genome and a linear DNA fragment. For this purpose, either the *Red* operon of lambda phage or the *RecE* and *RecT* genes from Rac phage [1,2] are expressed during the recombination event. The phage recombination genes can be expressed from easily curable (usually temperature-sensitive) plasmids [3,4] or from defective prophages controlled by a temperature-dependent repressor [5]. For site-directed recombination, a selectable marker is flanked by two homology arms that are defining the site of integration. The efficiency of recombineering depends on the length of the homology arms [5–7], but, in most studies, regions of 35–60 nt were suggested to be sufficient. Nevertheless, in a number of cases, longer homology regions seem to be necessary or at least useful to accelerate mutant formation. In contrast with the length of the homologue regions, Zhang et al. [6] were able to show that the distance between the left and right homology arm only has a minor influence on the recombination efficiency since constructs with distances between 0 and 3100 nucleotides were integrated into the genome at similar rates, although the integration rate of constructs with 0-nt distance was slightly decreased. Datsenko and Wanner [3] published in 2000 a recombineering method that is still widely used. This strategy is based on the transformation of linear DNA for homologous recombination to the chromosome [3]. Like most bacteria, *E. coli* is not transformable with linear DNA owing to the activity of its exonuclease V. Nevertheless, this can be overcome by the expression of bacteriophage lambda recombination genes. This so called Red system includes three genes that encode the 5′→3′ exonuclease, the recombinase and an inhibitor of the major *E. coli* exonuclease and the host recombination complex, RecBCD [1]. Depending on the linear DNA fragment, not only can this method be used for gene deletion, but also genes can be integrated into the chromosome [8].

In most of the methods, host DNA is deleted via subsequent insertion of a selectable marker and in some cases a gene of interest. Often, the marker sequence needs to be excised after a successful mutation, in order to reuse the technique again or to avoid an accumulation of genes leading to resistance against antibiotics in the strain. There are two ways of removing the marker sequence from the genome. One is to replace it using counter-selection methods.
achieving clean and ‘scarless’ mutations [7,9,10]; the other is to excise the unwanted marker sequences by site-specific recombination mechanisms (Figure 1). Mostly, the Flp-FRT system of Saccharomyces cerevisiae [11] and the Cre-lox system of phage P1 [12] are used for this purpose. These recombinases are highly specific and can be used in a wide variety of organisms [13]. They promote site-specific recombination within short target sequences, the FRT (Flp recognition target) and the loxP (P1 locus of crossing over) site respectively. Flp- and Cre-mediated excision of resistance genes leaves behind unwanted sequences, so-called ‘scars’, with a length of 50–70 nt [14]. For further Flp/Cre-mediated deletion or integration events, this scar can reduce the recombination efficiency, probably due to homology of the FRT/loxP sequence to the scar in the chromosome [15,16]. Hence, multiple engineered genomic integrations and deletions demand other techniques to remove the resistance genes from the chromosome.

A major disadvantage of this method is the occurrence of spontaneous point mutations in the sacB gene during positive selection on antibiotic resistance, leading to a high background after negative selection on sucrose [7].

Another counter-selection system is based on the rpsL gene, which encodes for the ribosomal protein S12. This protein is the target of the widely used antibiotic streptomycin, and mutations in the rpsL gene lead to resistance to streptomycin [20]. The resistant phenotype is recessive in a merodiploid genetic background. Hence, a strain is streptomycin-sensitive when both wild-type and mutant allele are expressed [20]. Heermann et al. [4] describe a method for site-directed mutagenesis of E. coli using rpsL counter-selection in combination with lambda red gene-based recombineering [3,6]. Here, a (streptomycin-resistant) E. coli strain is transformed with a linear rpsL-neo cassette containing a wild-type rpsL gene copy that is inserted in the genome and thereby, for instance, replacing a genomic sequence. The wild-type rpsL is dominant over the mutant form and therefore the strain is streptomycin-sensitive and kanamycin-resistant. Then the rpsL-neo cassette can be replaced by a second DNA fragment containing regions homologous with the up- and downstream sequence of the cassette. Positive recombinants will lose the wild-type allele of rpsL and can be selected on streptomycin [4].
In addition to these counter-selection systems with two different selection markers, there is also the possibility to use one selection marker for both positive and negative selection. The first marker of this kind described was tetAR, an operon that encodes the tetracycline-resistance gene tetA and the regulator tetR [20]. The expression of the tetAR genes leads to an alteration in the cell membrane structure, generating cells that are resistant to tetracycline. Additionally, the tetracycline-resistant cells exhibit a hypersensitivity against lipophilic chelating agents such as fusaric or quinalic acids. [21,22]. It is therefore possible to counter-select for clones that have lost tetracycline resistance due to their resistance against fusaric acid (Figure 1).

Another system is the galK-based positive/negative selection via the degradation of galactose and the galactose analogue DOG (2-deoxygalactose). The galactose operon of E. coli consists of four genes for the degradation of galactose. The galK gene encodes the galactokinase, which catalyses the first step of the metabolic pathway. The galactokinase also metabolizes DOG, leading to the toxic derivative 2-deoxygalactose 1-phosphate [23]. The method is dependent on a galK− mutant. In this mutant, a galK cassette can be integrated into the genome using recombineering and positive selection on plates containing galactose as sole carbon source. In a second step, galK can be removed using another piece of linear DNA for galK replacement and plates containing both DOG and glycerol [7] (Figure 1). Since galK is the selection marker in both selection steps, the risk of DOG resistance in the second step is reduced, because the initial positive selection will exclude PCR-generated loss-of-function mutations in the inserted galK gene. This leads to a low background and a high efficiency. A disadvantage of this selection method is that it cannot be used in wild-type E. coli strains, since the galK gene needs to be deleted before the positive selection step.

A modified version of recombineering was published by Posfai et al. [9]. Here, a linear DNA fragment is inserted into the genome using phage recombination techniques. This fragment is composed of a selective marker, flanked by two I-SceI meganuclease sites and two homologous regions (boxes A and C) and replaces a desired segment of the chromosome. However, in contrast with the recombineering methods illustrated already, the integrated fragment carries a third homology domain (box B) fused to box A (depicted in Figure 1). Box B corresponds to the region downstream of box C, and integration of the DNA fragment leads to a duplication of box B in the chromosome. The meganuclease is derived from the yeast Saccharomyces cerevisiae and the 18-nt restriction site of I-SceI is unlikely to occur in other organisms than yeast and some metazoans [24]. After induction of I-SceI expression from a plasmid, double-strand breaks are created and repaired by endogenous double-strand break repair. The short homologous box B regions that can be recombined in a RecA-mediated step support recombination.

All of the methods described have in common that linear DNA is transformed in the E. coli cells, which might limit the efficiency of transformation and DNA integration. The ‘gene gorging’ method [10] bypasses this limitation by linearizing the DNA fragment in vivo. To accomplish this, a donor plasmid containing the DNA fragment flanked by I-SceI recognition sites and a mutagenesis plasmid with the I-SceI meganuclease and the lambda red genes under the control of the inducible arabinose promoter are co-transformed. Mutagenesis is initiated by the induction of the meganuclease and the lambda red genes with arabinose. The meganuclease cuts the donor plasmid and creates linear DNA that can be recombined to the chromosome. Owing to the utilization of a high copy donor plasmid, elevated amounts of linear DNA are produced intracellularly and a higher efficiency of allelic exchange compared with the transformation with linear DNA can be achieved. According to the authors, the efficiency of gene replacement is even high enough to forgo selection [10].

**Phage- and transposon-based site-specific integration systems**

Recombineering approaches that allow the researcher to freely choose an integration site for the gene that is supposed to be expressed in E. coli are frequently used, potentially in the majority of cases. Nevertheless, phage- and transposon-based site-specific integration systems have proved to be useful in a number of cases and represent simply another track for genomic insertion that can be followed simultaneously. Integration efficiency of these systems is high and predictable. Moreover, the integration process can be performed in a way that makes it reversible [25,26]. A very interesting suite of so called CRIM (conditional-replication, integration and modular) plasmids was constructed by Haldimann and Wanner [26]. The user can choose between vectors containing five different attP sites. The plasmids are all based on an R6K origin background that necessitates the trans-acting π-protein supplied by pir− or pir-116 host cells. Genes of interest are cloned into these vectors in π-protein-producing hosts. Transformation of constructs into pir− cells that express the specific integrase from a helper plasmid results in stable plasmid integration since the γ-replication origin is not functional in these host cells. All five integration sites can be used since CRIM plasmids with different antibiotic-resistance markers were constructed. Furthermore, Xis/Int helper plasmids that allow excision from the chromosome were constructed as well. Another elegant system is based on T7 site-specific integration at the 3′ end of the glmS gene [27]. The method was initially developed by Nancy Craig’s group [28] and then developed further by Sibley and Raleigh [29]. Basically, a curable plasmid containing the genes necessary for transposition is used that has a multiple cloning site between the Tn7 arms. The gene of interest can be cloned between these arms and the plasmid containing a selectable marker can be transformed into any E. coli strain. Induction of the five tns genes will allow integration of the Tn7 arms and the gene in the middle. Genomic insertion of a selectable marker is not necessary since the method is so efficient that a simple PCR
screen of a limited number of colonies will be sufficient to find cells with the desired insertion [28].

Promoters
For the expression of heterologous genes, it can be of great importance to vary expression levels, (i) for overexpression, (ii) to prevent potentially toxic effects associated with the expression, or (iii) to adjust expression levels to various protein activities of a synthetic pathway. Therefore, in some cases, strain design might depend on not only one promoter, but also a multitude of promoters that guarantee controllable expression of multiple genes. In addition to the established set of lac, ara and tet promoters and their derivatives [30], progress has been made in the isolation of other promoter systems. The propionate promoter is highly adjustable over a wide range of inducer concentrations, provides homogenous expression in individual cells and strong expression (maximal 1500-fold induction) at high propionate concentrations [31–33]. For an alkane-inducible promoter, alkS, the positive regulator of the alkBFGHJKL operon and the alkR promoter of Pseudomonas oleovorans GP01 were assembled to an n-octane-responsive expression cassette which allowed the expression of xylene oxygenase genes in E. coli [34,35]. The nph promoter is responsible for the expression of the glycerol 3-phosphate transport system in E. coli and can be induced with phosphate. The promoter strength and controllability is comparable with that of the tac promoter [36]. Another option would be the use of tightly controlled rhamnose-inducible system. This system uses the rhapBAD promoter which is activated via binding of the RhaS protein together with L-rhamnose. RhaS synthesis is triggered by the activator RhaR which is produced when L-rhamnose becomes available. Similar to the arabinose-inducible araP_BAD promoter, basal expression is relatively low [37–39].

For some applications, it might be necessary to express heterologous genes from constitutive promoters. Different libraries of artificial promoters were described that cover a wide range of promoter activities with small steps of activity increase ranging over up to five orders of magnitude [40–43]. Furthermore, many promoter libraries appear to have a broad host range [40]. The relative strength of promoters can be predicted on the basis of the nucleotide sequence by a mathematical model [PSP (promoter strength predictive) model] [44]. This in turn can be applied for the fine-tuning of gene expression [41].

Conclusions
Original papers describing methods for manipulation of E. coli are always in favour of the method presented and it always seems as though these manipulations are the easiest thing in molecular biology. Nevertheless, we have so far deleted in one E. coli strain 22 genes and inserted in turn 17. Hence, one might allow us a few personal experience-based leads in the end. Simple recombineering leaving scars behind worked well for the deletion of two or three genes, but it was extremely hard to obtain more than this. We have tried all of the other methods mentioned in the text. After one-quarter of the way, we decided to use a mixture of galK/gene gorging-based insertion, selection and counter-selection (using a stronger constitutive promoter in front of galK) as well as a CRIM plasmid and Tn7-based insertion. This, so far, is the best combination in our hands, but the continuous advent of techniques and synthetic genome construction might reveal more and faster ways for genomic engineering in the future.

Funding
We are grateful for financial support from the German Ministry of Education and Research (BMBF, Projekttträger Jülich) under the programme ‘Bioenergie2011’ [grant number 03F0382].

References


