Effect of Insertion Sequences on Genom Stability in *Escherichia coli* K-12

Abstract of Ph.D thesis

Umenhoffer Kinga

Supervisor: Dr. Pósfai György

Ph.D School of Biology

HAS BRC, Institute of Biochemistry

University of Szeged, Faculty of Science and Informatics

2010

Szeged

Introduction

Escherichia coli K-12 is one of the best understood and most thoroughly analyzed organisms. *Escherichia coli* K-12 and O157:H7 were amongst the first bacteria chosen for whole genome sequencing. By far the highest number of individual genome sequences is available for this species, providing us an extraordinarily useful tool for genomic comparisons. Moreover, *Escherichia coli* is an ideal model for studies of processes involved in bacterial genome evolution.

Genomes of *Enterobatericeae* in general, and of *Escherichia coli* strains in particular are now recognized to be mosaics in which a backbone of conserved genes in conserved order is interspersed with strain-specific horizontally transmitted “islands”. Our group`s earlier goal was to construct an improved *Escherichia coli* by using scarless genomic reduction of the sequenced K-12 strain MG1655 to physically realize a bacterium whose genome consists of the backbone elements common to most *E. coli.*

E. coli is littered with genes that mediate horizontal gene transfer, including IS (insertion sequence) elements, transposases, defective phages, integrases, and site specific recombinases. Amongst them IS elements are the small, genetically compact units, which generally encode no functions other than those involved in their mobility. These are “selfish” DNA elements, which can translocate, duplicate, and be maintained in the genome even if detrimental to the host, behaving in some sense like an infectious agent. In a typical bacterial cell, IS elements generate a significant share of genetic variation. IS elements, when present in more than one copy, also provide repeats that lead to inversions, duplications, and deletions mediated by homologous recombination. Genome alterations due to IS translocation occur surprisingly frequently and many commonly used laboratory strains have unrecognized genome alterations because of this. IS hopping from chromosome to plasmids is also common, leading to difficulties in cloning, expression, and library construction. IS elements can also pass from strain to strain during laboratory manipulations. Having a unique IS free *Escherichia coli* (MDS42) strain in our hand, we aimed to investigate the role of IS elements on genome evolution and their impact during synthetic biological-biotechnical applications.

During our study we used an extremely mutable “IS-trap” plasmid. A

recombinant ectopic gene, constructed for vaccine development, and composed of a synthetic gene of the structural capsid protein VP60 of rabbit haemorrhagic disease (RHD) virus fused to the gene of the B subunit of cholera toxin (CTX) was very unstable in conventional *E. coli* hosts. In contrast, in the IS-free MDS42, reduced-genome strain the plasmid remained much more stable. We investigated the reasons of the plasmid instability in detail, and extended the analysis to large shotgun sequencing clone libraries.

Aims

We wanted to investigate the following questions in our study:

- What is the general frequency of IS transposition in *E. coli*?
- Does stress increase the IS transposition, similar to pontmutational rate?
- What are the molecular details of the instability in our ”IS-trap” plasmid?
- How can the IS inactivated plasmid become dominant in the population in very short time? What is the mutation/selection dynamics of the cells harbouring the unstable plasmid?
- Is the observed case of IS-mediated instability of *ctxvp60* unique, or is it a far more widespread, general phenomenon? Can we find analogous processes by analysing raw data from various shotgun sequencing projects?

Methods

Transformation of bacterial cells

Bacterial growth measurements and statistics

Growth retardation effect measurements of transformants bearing toxic and control inducible plasmids in liquid medium

Mutation rate measurements

Mutational spectrum analysis

Recombinant DNA constructions

RNA isolation from colonies and RT-PCR

Cell imaging using confocal LSM

Bioinformatic analysis of *E. coli* IS sequences in raw shotgun sequencing data

Results

In an earlier study, it was shown that a chimeric gene (*ctxvp60*), constructed for vaccine development, and composed of a synthetic gene of the structural capsid protein VP60 of rabbit haemorrhagic disease virus fused to the gene of the B subunit of cholera toxin (CTX) was very unstable in conventional *E. coli* hosts. Upon transformation by plasmid pCTXVP60, heterogeneous colonies of transformants were obtained on agar plates (small, slow growing, and normal growth). A vast majority, 92 %, of the large colonies carried plasmids with IS insertions, while 8 % displayed deletions or no alteration detectable by PCR. In the case of IS insertions, we detected IS1, IS3, and IS5 translocations, all of them occurring in the 5' end of *ctxvp60*. In contrast, the recombinant plasmid was stable in IS-less MDS42. Transformants in MDS42 showed homogenic colonies, plasmid DNA recovered yielded unaltered restriction digestion pattern and nucleotide sequence. To understand the causes of the host-dependent stability of the plasmid, we've planned to investigate the molecular details of this behaviour.

To determine whether cloning of pCTXVP60 increases the IS-mediated mutation rate, bystander mutation tests were performed, detecting mutations in a neutral, chromosomal gene (*cycA*). The appearance of *cycA* mutant (d-cycloserine resistant) colonies was counted in MG1655 cells carrying pCTX (control, stable plasmid), and in cells carrying the pCTXVP60 plasmid. The overall mutation rate increased 2-fold in MG1655 carrying pCTXVP60, compared to MG1655 carrying pCTX. However, the IS transposition rate was elevated 4-fold. Four different IS elements contributed to this elevated transposition rate. Furthermore, 32 % of the chromosomal *cycA* IS mutants also carried an IS in pCTXVP60. This significantly high proportion of double IS mutants can not be caused by independent single transposition events. Our hypothesis is that stress-induced IS transposition, combined with fast selection of IS-inserted mutants could provide an explanation.

To test whether protein overexpression *per se* increases the frequency of IS transposition in growing cells, a culture of MG1655 carrying pPROEX HT-CAT (expressing a well-tolerated protein, chloramphenicol-acetyltramsferase /CAT/) was divided into two portions, and the frequency and types of mutations were studied in 20-

tube fluctuation tests with or without induction of CAT expression. We found that the number of transpositions into *cycA* was 136% higher in expression-induced cells. Point mutation frequencies remained virtually unchanged. We conclude that overexpression of even a well-tolerated protein leads to elevated IS transposition.

For better understanding of the toxic phenomenon of *ctxvp60,* we cloned it as an inducible construct in plasmid pSG1144. Upon induction, a strong growth retardation effect was observed, possibly caused by the chimeric gene. The gene *ctxvp60* contains a large number of rare Arg codons. We assumed that translation of CTXVP60 exhausted the rare $tRNA_{Arg}$ pool of the cells resulting in the toxic effect. We tested this hypothesis by synthesizing two different versions of *ctxvp60,* that were different only in codon usage. Comparing the induced expression of both constructs (optimized and dezoptimized codon usage for *E. coli*) with the original *ctxvp60* showed that the extreme codon usage had no negative effect on the growth of the cell. Additionally, a frameshift version of *ctxvp60* was constructed, but this modification did not eliminate the toxic effect either. All together, this indicated that neither the *ctxvp60* gene/CTXVP60 protein *per se,* nor the supposed rare tRNA depletion phenomenon were the cause of the toxic effect of pCTXVP60.

Bioinformatic analysis of the fusion gene revealed an artificially formed ORF out of the original reading frame (ORF238). Translation of ORF238 results in an extremely Leu-rich (102 Leu residues) protein, which is predicted to have four putative transmembrane domains. To determine whether this small artificial ORF is responsible for the abnormal phenotype, we cloned it as an inducible construct. Upon induction, ORF238 produced the same growth retardation as seen with the original pCTXVP60 plasmid. We concluded that ORF238 was the culprit of the growth defect. Confocal laser scanning microscopy imaging of the cells expressing either CTXVP60 or ORF238 supported our findings. Both construct showed abnormally long cells with multiple nucleoids.

Based on sequencing 12 IS mutant pCTXVP60, we defined that insertions occurred in the 5’ third of the fusion gene, in the *ctx* part or near the 5’ end of *vp60*. Since the 5’ end of ORF238 is located in the joint between *ctx* and *vp60,* we concluded that IS insertions in this region might block the transcription, relieving the growth retardation.

RT-PCR experiments from IS mutant and original pCTXVP60 plasmids confirmed our conception.

Growth characteristics of strains transformed with toxic or mutant plasmids were monitored in liquid medium. The results correlated well with the orserved growth characteristics of transformants on agar plates. The original pCTXVP60 i) causes slow growth ii) cells pick up an IS insertion in *ctxvp60*, due IS transposition from the chromosome iii) cells harboring an IS-inactivated ORF238 resume normal growth and quickly become dominant in the culture. By using strains that possess various numbers of IS-es, we observed that the time needed for evolution of a non-toxic plasmid variant in the culture inversely correlated with the number of IS-es in the host genome. Additionally, even a single copy of an IS present in the host genome had a marked effect on plasmid stability.

In silico analysis of sequence libraries supported our theory that the case of *ctxvp60* is not unique, but far more widespread mechanism of self-defence against not tolerated, ectopic recombinant genes. Data from 295 shotgun genome sequencing projects were analyzed with stringent criteria searching for IS contamination from cloning host genome. We've found a total of 109 IS-containing reads with 8 different kinds of IS elements. This elevated rate could be explained by processes analogous to those, which have been seen when cloning *ctxvp60*.

Recombined products that are toxic for *E. coli* may accumulate neutralizing, IS-mediated mutant clones, which can become dominant by their faster growth and may cause undesired genotypic and phenotypic changes. We've shown that this phenomenon is widespread and leads to inactivation of biotechnological/synthetic-biological products and can diminish productivity. Removing all IS elements from the genome of a bacterial cell results in a significant increase in genomic stability and phenotypic uniformity, yielding an improved cellular chassis with reduced evolutionary potential.

Results in brief

- We identified specific mutations in an unstable plasmid (pCTXVP60), which caused severe growth retardation in *E. coli*. These mutations were capable of eliminating the toxic effect caused by pCTXVP60.
- We demonstrated that virtually all mutations resulted from IS insertions. We identified their types, positions and insertional frequency.
- We identified a hydrophobic protein as the toxic byproduct of the *ctxvp60* clone, and showed that IS elements spontaneously landing in the clone relieve the cell from the stress.
- We showed that the IS insertions act by blocking transcription of *orf238.*
- We showed that IS-inserted clones that stopped expressing ORF238 resumed normal growth and rapidly became dominant in the culture, due to their rapid selection.
- We demonstrated that stress significantly induces the mobility of IS element.
- Bioinformatic analysis of *E. coli* IS sequences in raw shotgun sequencing data demonstrated, that IS-mediated clone instability is a more wide-spread phenomenon, than anticipated.
- We demonstrated that removal of all IS elements from the genome of a bacterial cell results in a significant increase in plasmid (genomic) stability. We conclude, that delaying the genetic adaptation in clean-genome MDS42 - serving as host in SB applications - is beneficial in both laboratory and industrial settings.

List of publications

Umenhoffer K, Feher T, Baliko G, Ayaydin F, Posfai J, Blattner FR, Posfai G
Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications.
MICROB CELL FACT 9: 10.1186/1475-2859-9-38- p. (2010)
IF: 3.432

Feher T, Karcagi I, Gyorfy Z, Umenhoffer K, Csorgo B, Posfai G
Scarless engineering of the Escherichia coli genome
In: Gerdes S, Osterman AL

Microbial Gene Essentiality: Protocols and Bioinformatics: Methods in Molecular Biology series. Vol. 416..
TOTOWA, USA: HUMANA PRESS INC., 2008. pp. 251-259.

Posfai G, Plunkett G, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR
Emergent properties of reduced-genome Escherichia coli
SCIENCE 312: 1044-1046 (2006)
IF: 30.927

Feher T, Cseh B, Umenhoffer K, Karcagi I, Posfai G
Characterization of cycA mutants of Escherichia coli - An assay for measuring in vivo mutation rates
MUTAT RES-FUND MOL M 595: 184-190 (2006)
IF: 3.340