Markerless *Escherichia coli* rrn Deletion Strains for Genetic Determination of Ribosomal Binding Sites

Selwyn Quan*, Ole Skovgaard†, Robert E. McLaughlin‡, Ed T. Buurman‡, and Catherine L. Squires§

Department of Biology, Stanford University, Stanford, CA*; Department of Science, Systems and Models, Roskilde University, Roskilde, Denmark†; Department of Biosciences, Infection Innovative Medicines Unit, AstraZeneca R&D Boston, Waltham, MA‡; Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA§;

¹Present address: Vaccine Research and Development, Pfizer, Pearl River, NY

Genbank accession numbers CP011320, CP011321, CP011322, CP011323, and CP011324
Running Title: Markerless rnr Strains of *E. coli*

Keywords: rrl, rrs, duplication, genomic instability, ribosome

Corresponding author: Ed T. Buurman, Pfizer Vaccine Research and Development,
401 North Middletown Road, Pearl River, NY 10964; Ph: 781 354 1481 E: Ed.Buurman@pfizer.com
ABSTRACT

Single copy rrn strains facilitate genetic ribosomal studies in Escherichia coli. Consecutive markerless deletion of rrm operons resulted in slower growth upon inactivation of the fourth copy, which was reversed by supplying tRNA genes encoded in rrm operons in trans. Removal of the sixth, penultimate rrm copy led to a reduced growth rate due to limited rrm gene dosage. Whole genome sequencing of variants of single copy rrm strains revealed duplications of large stretches of genomic DNA. The combination of selective pressure, resulting from the decreased growth rate, and the six identical remaining scar sequences, facilitating homologous recombination events, presumably leads to elevated genomic instability.
INTRODUCTION

The majority of antibiotic classes currently in clinical use act by inhibiting ribosome function. Their utility is threatened by the emergence of microbial resistance, mainly due to compound efflux (Thaker et al. 2010), covalent modification of the ribosome (Goossens 2009), or alteration of the antibiotic (Zhanel et al. 2012). The binding interactions of the antibiotics with the ribosome are predominantly established with its RNA components (Poehlsgaard and Douthwaite 2005), which in E. coli are encoded by seven virtually identical rrn operons: rrnA through rrnH (rrnF was renamed as rrnG) (Fig. S1A). This seven-fold redundancy limits the introduction of resistance caused by binding site mutations, mainly due to the recessive nature of resistance mutations that are suppressed by the presence of six wild type alleles that lead to rapid cell death (Orelle et al. 2013). In addition, the multiplicity of wild type alleles will tend to revert the mutant allele due to rapid gene conversion. This extremely low resistance frequency is from a clinical perspective a very attractive feature for an antibiotic and as a consequence the ribosome has remained an attractive antibacterial target despite decades of macrolide and aminoglycoside usage. From a drug discovery perspective, however, this redundancy makes definition of a structural foundation guiding iterative chemistry programs challenging.

One recent example is negamycin, a natural compound that mediates its antibacterial activity by inhibition of translation (Mizuno et al. 1970) and for which as many as ten different ribosomal binding sites have been described (Schroeder et al. 2007, Olivier et al. 2014, Polikanov et al. 2014a). The location of a single binding site through which the compound mediated its activity, required for structure-based design of improved analogs, remained elusive for forty years. Negamycin-resistant mutants of wild-type E. coli have been isolated, but the mutations affected transport
of the compound into the cell rather than binding to the ribosome (McKinney et al. 2015). In order to obtain genetic evidence for negamycin’s functional binding site, helix 34 of the small subunit, the redundancy of wild type *E. coli* ultimately needed to be circumvented by the use of markerless single *rrn* strains (Olivier et al. 2014, Polikanov et al. 2014a). Although this tool has found use in various seminal scientific advances of ribosomal biology (Bollenbach et al. 2009, Orelle et al. 2013, Polikanov et al. 2014b, Orelle et al. 2015), its construction was mentioned briefly (Bollenbach et al. 2009). Here we describe the construction in detail, showing that sequential deletion of four *rrn* genes leads to a reduced growth rate due to limiting expression of tRNAs, before upon deletion of the sixth copy the cellular ribosomal content becomes growth-limiting. Furthermore, although colony morphology suggested stability of the strains (Bollenbach et al. 2009), whole genome analyses revealed elevated genomic instability due homologous recombination between the scars remaining upon *rrn* deletion.

MATERIALS AND METHODS
SI contains a list of strains (Table S1), a detailed *rrn* deletion strategy (Fig S1, S2 and S3) and a map of pKD4-16 (Fig S5). Additional whole genome sequencing data is provided in Fig S4. *E. coli* strains SQ37, SQ88, SQ2203, SQ110 and SQ171 used for genomic sequencing were obtained from the Coli Genetic Strain Center at Yale University. Complete nucleotide sequences of these strains have been deposited in Genbank under accession numbers CP011320, CP011321, CP011322, CP011323, CP011324, respectively.
RESULTS AND DISCUSSION

A set of single *rrn* deletion strains was constructed using *E. coli* MG1655 (Blattner *et al.* 2012). Seven strains, each with a single deletion spanning one of the *rrn* operons, were made using the recombineering method (Fig. S1, Table S1) (Datsenko and Wanner 2000). Upon resolving the kanamycin-resistance cassette from one operon, leaving a 85 bp insertion (“scar”), a new kanamycin-resistance cassette was introduced into another *rrn* operon which was subsequently resolved, and so on (Fig. S2, Fig. S3A). Deletions of the *rrn* operons were confirmed by both PCR (data not shown) and Southern blots (Fig. S3B). However, all *rrn* operons encode the RNA components for the small and large ribosomal unit (*rrs* and *rrl* genes, respectively) that are interspersed with a number of different tRNA genes which would have prevented removal of more than five operons (Fig. S1B). Therefore, plasmid ptRNA67 (Zaporojets *et al.* 2003) was introduced to provide tRNA genes of Ala-1B, Ile-1, Trp, Asp-1, Thr-1 and Glu-2 *in trans* (Asai *et al.* 1999).

At high growth rates, as much as 70% of *E. coli*’s resources are devoted to the translation machinery and protein synthesis (Russell and Cook 1995). The ribosome forms the core of the translation machinery and the growth rate is proportional to both cellular ribosome content and specific peptide elongation rate (Dennis *et al.* 2004). Feedback regulation is one mechanism of rRNA regulation in *E. coli*. *E. coli* strains with a limited number of *rrn* operons maintain ribosome content by increased transcription of the remaining operons; conversely, the presence of additional *rrn* copies does not change ribosome content (Gyorfy *et al.* 2015).

Sequential deletion of three *rrn* operons did not significantly alter the doubling time of strains when grown in Luria Broth at 37°C, but the additional deletion of a fourth operon resulted in an increase from 33 to 43 min (30%) (*rrnGBA vs *rmGADE*).
p=0.0003, rrmGBA vs rrmGBAD p=0.0018 (Fig. 1)). Deletion of the fifth operon, rrmH, required the addition of tRNA genes on a plasmid since this deletion would remove all copies of the tRNA-Ile and tRNA-Ala genes encoded within the rrm operon spacers (Fig. S1B). Despite the removal of the fifth operon, the growth rate of the ΔrrnGADEH ptRNA67 strain was significantly higher than that of the ΔrrnGADE strain, with doubling time of 38 and 43 min respectively (p=0.0089), suggesting that the lower growth rate of the ΔrrnGADE strain was caused largely by limiting amounts of tRNA (Fig. 1). This observation was confirmed by failure to increase the growth rate of the ΔrrnGADE strain upon introduction of rrm encoding plasmid pK4-16 (p=0.87). The largest decrease in growth rate, increasing the doubling time from 38 to 60 min, was observed upon deletion of the sixth rrm copy, leaving a single chromosomal rrm operon (Fig. 1). Introduction of pK4-16 restored the growth rate close to the level observed in the ΔrrnGADEH ptRNA strain (p=0.00002).

For the single rrm strains to serve as genetic tools and select for mutations in rrm, whole-genome-sequences of strains SQ37, SQ88, SQ2203, SQ110 and SQ171 were determined. Mapping of the sequence reads confirmed correct deletion of the appropriate rrm operons, and the density of sequence reads in SQ37, containing a single deletion, and SQ88, containing five deletions, was evenly distributed across the genome (Fig. S4). However, in single rrm operon strains SQ2203, SQ110 and SQ171 genomic regions were found in which the density of sequence reads was doubled, reflecting DNA duplications (Fig. S4). The duplications always seemed to have occurred between two scar sites. In an extension of this work, whole-genome-sequences of multiple colonies of SQ110 and SQ2203 as well as negamycin-resistant mutants derived therefrom (Olivier et al. 2014, McKinney et al. 2015) were determined, and in some of these the duplication pattern was altered (Fig. 2). This
observation suggests that the combination of the severe growth defect observed
upon removal of the penultimate *rrn* operon (Fig. 1) and the presence of six identical
scar sequences spread through the genome promote recombination events.

ACKNOWLEDGMENTS

This work was supported by NIH grant R01-GM24751 to C.L.S.
REFERENCES


McKinney, D. C., N. Bezdenejnih-Snyder, K. Farrington, J. Guo, R. E. McLaughlin *et al.*, 2015 Illicit transport via the dipeptide transporter Dpp is irrelevant to efficacy of negamycin in mouse thigh models of *Escherichia coli* infection. ACS Infect. Dis 1: 222-230.


Fig. 1 Growth rates of *E. coli* *rrn* deletion strains grown in Luria Broth medium at 37° with and without the tRNA plasmid, ptRNA67, and rRNA plasmid, pK4-16 (n≥3).

- **Strain background**
  - WT
  - SQ37(1)
  - SQ40(2)
  - SQ49(3)
  - SQ53(4)
  - SQ78(4)
  - SQ88(5)
  - SQ110(6)
  - SQ171(7)

- **Doubling time (min)**
  - WT: 20
  - SQ37: 30
  - SQ40: 40
  - SQ49: 50
  - SQ53: 60
  - SQ78: 70
  - SQ88: 80
  - SQ110: 90
  - SQ171: 100

- Symbols:
  - ○ no plasmid
  - ■ + ptRNA67
  - ○ + ptRNA67 + pK4-16
  - ▲ + pK4-16
Fig. 2 Relative sequence read coverage of multiple isolates of *E. coli* strains SQ2203 and SQ110 and negamycin-resistant mutants derived thereof (Olivier et al 2014, McKinney et al 2015) reveal unstable regions of genomic duplication. Regions which showed ~2x increased relative coverage, reflecting duplication, are indicated with a black bar. (A) SQ2203 yielded two variants (blue, yellow) in additional to the predicted, single fold coverage genome (grey). (B) All wild type SQ110 isolates that were sequenced contained the same duplication (brown) but different arrangements were found among negamycin-resistant mutants. ‘Spikes’ in the plots represent regions of homology of mainly tRNA and rrn genes/operons in resident plasmids, with genomic sequences.