Rapid optimization of gene dosage in *E. coli* using DIAL strains

Joshua T Kittleson, Sherine Cheung and J Christopher Anderson

**Abstract**

**Background:** Engineers frequently vary design parameters to optimize the behaviour of a system. However, synthetic biologists lack the tools to rapidly explore a critical design parameter, gene expression level, and have no means of systematically varying the dosage of an entire genetic circuit. As a step toward overcoming this shortfall, we have developed a technology that enables the same plasmid to be maintained at different copy numbers in a set of closely related cells. This provides a rapid method for exploring gene or cassette dosage effects.

**Results:** We engineered two sets of strains to constitutively provide a trans-acting replication factor, either Pi of the R6K plasmid or RepA of the ColE2 plasmid, at different doses. Each DIAL (different allele) strain supports the replication of a corresponding plasmid at a constant level between 1 and 250 copies per cell. The plasmids exhibit cell-to-cell variability comparable to other popular replicons, but with improved stability. Since the origins are orthogonal, both replication factors can be incorporated into the same cell. We demonstrate the utility of these strains by rapidly assessing the optimal expression level of a model biosynthetic pathway for violacein.

**Conclusions:** The DIAL strains can rapidly optimize single gene expression levels, help balance expression of functionally coupled genetic elements, improve investigation of gene and circuit dosage effects, and enable faster development of metabolic pathways.

**Background**

Optimizing desired outcomes by varying a design parameter is a staple of almost every engineering field, from mechanical engineers tweaking blade angles on a wind turbine to civil engineers altering the timing of traffic lights. Similarly, genetic engineers alter gene expression levels to optimize some desirable phenotype. Strong overproduction of single proteins can impose a metabolic burden on *E. coli*, and often a lower expression level leads to improved phenotype [1]. In multi-subunit proteins and genetic circuits, expression of particular proteins often needs to be balanced for proper function (e.g. [2,3], and [4]). Extensive work has established methods for achieving expression of a gene or operon at a particular level, including control of transcription using standard promoter sets [5], modulation of RNA processing [6], and control of translation through ribosome binding site (RBS) manipulation [7]. However, using these tools to investigate the desired expression level of a single gene or operon requires cloning for each level to be tested. Using inducible promoter systems to probe multiple expression levels can rapidly determine an approximate desired expression level, but does not provide a genetically encoded solution, which can be useful for downstream applications. For optimizing multi-operon constructs, fewer tools exist. Generating large numbers of repeats in the genome is labor intensive [8], while strategies that increase plasmid copy number upon induction provide a narrow range of copy numbers [9], cause runaway replication [10,11], or are incompatible with constructs using the common P*BAD* promoter [12]. To address these shortcomings and allow researchers to explore the effect of copy number on genetic devices, we have exploited an underutilized control point: plasmid copy number.

Genetic engineers working in *E. coli* are blessed with a wide range of plasmid systems and plasmid copy numbers to choose from, ranging from single copy BACs to ~500 copy pUC plasmids [13]. However, to take advantage of copy number differences, each gene or device of interest has to be cloned into different plasmids,
necessarily changing the local genetic context along with the copy number. A notable exception to this rule is the gamma origin of the R6K plasmid, which requires the trans-acting Pi protein to initiate replication [14]. Different alleles of pir integrated into the E. coli genome are known to support R6K plasmids at different copy numbers (e.g. pir+ and pir116 [15]), meaning that the genetic context on the plasmid itself remains unaltered. Similarly, the orthogonal replicon of ColE2-P9 also uses a trans-acting factor, RepA, to support the origin of replication [16].

In this work, we generate two sets of strains bearing different alleles (DIAL) of pir or repA to support the same plasmid at a wide range of copy numbers. We then characterize the copy number, cell-to-cell variability, and stability of plasmids in both sets of DIAL strains. We illustrate their utility on a model system by examining expression of the violacein biosynthesis pathway [17] at different copy numbers. The results demonstrate that artificially re-regulating replication factor expression from the genome can produce stable plasmid copy numbers, that phenotype varies with copy number, and that DIAL strains can accelerate development of genetic devices.

Results and Discussion

ColE2 as a Trans-Activated Origin

A previous report [18] identified a minimal 32 bp region of ColE2 sufficient to support replication of a plasmid when repA is provided in trans. We first recapitulated this behaviour by transforming a plasmid bearing the ColE2 minimal origin (pBjk2164-jtk2619) into cells constitutively expressing the trans-acting repA gene from a plasmid (MC1061 + pBca9145-jtk2627). Although we observed transformants, they exhibited small colony morphology equivalent, while an R6K plasmid in the DIAL strains spans the range of ~1-60 copies per genome equivalent. This covers nearly the entire range of copy numbers, that phenotype varies with copy number, and that non-essential factors contributing to robustness would be captured. This yielded colonies with morphologies indistinguishable from untransformed cells (data not shown).

Orthogonality of R6K and ColE2

We next examined if ColE2 and R6K were orthogonal origins of replication. Plasmids bearing either an R6K or ColE2 origin of replication were transformed or cotransformed into cells expressing pir, repA, or both. We observed colonies only when a plasmid was transformed into cells possessing the cognate replication factor, as expected.

Construction of DIAL Strains

To generate cells expressing diverse levels of the trans-acting factors Pir and RepA, we integrated expression cassettes with randomized ribosome binding sites (RBSs) into the genome, thereby creating strain sets JTK160 and JTK164 for pir and repA, respectively. We subsequently visualized copy number variation by transforming the libraries with reporter plasmids constitutively expressing sfGFP [20] (pBjk2741-jtk2828 or pBjk2807-jtk2828). After a preliminary fluorescence analysis of 380 clones of each type (data not shown), 24 were selected for further investigation. After a second round of fluorescence measurement, 10 variants of each type that spanned the range of observed fluorescence levels were chosen for full characterization. The sequences of the selected RBSs are reported in table 1.

Characterization of DIAL Strains: Copy Number, Cell-to-Cell Variability, and Stability

We characterized three important properties of plasmids in the DIAL strains: copy number, cell-to-cell variation, and stability. To estimate the copy number supported by each strain, we employed qPCR to examine plasmid content both at mid log and at stationary phase (Figure 1). Based on this analysis, a ColE2 plasmid in the DIAL strains spans the range of ~1-60 copies per genome equivalent, while an R6K plasmid in the DIAL strains spans the range of ~5-250 copies per genome equivalent.

Table 1 RBSs of repA and pir expression variants

| Strain | Variant | Sequence |
|--------|---------|----------|
| JTK160 (repA) | A | TCTAGAATACCTGATG |
|          | B | GTGAGAAAGGGAAGTG |
|          | C | CCGAAGAAAGGTAGATG |
|          | D | GCCGAAAGGTGAAATG |
|          | E | GTGAAAGGTGCTCTGT |
|          | F | CTCAGAAAGAAATAG |
|          | G | TGTGAGAAAGGTGATG |
|          | H | TTAGAACAAAGAGATG |
|          | I | GGGAGAAAGATGATG |
|          | J | GGGAGAAAACAAAATG |
| JTK164 (pir) | A | GCTGAAAGGTTGATG |
|          | B | TAAGAAATAGGTTG |
|          | C | GGCGAAGGTGCTAG |
|          | D | TGCGAAAGGCTGATG |
|          | E | GTGCGAAGGCTGATG |
|          | F | GTGGAATGGATG |
|          | H | AGCGAAGGATG |
|          | I | ACTGAAAGGCTGATG |
|          | J | ATCGGAAACATAGGATG |

Start codons are underlined.
reported plasmid copy numbers, from single copy to nearly pUC levels. We observed that the pUC plasmid exhibited 4-5 fold increased copy numbers at stationary phase, while the p15a, R6K, and CoLE2 plasmids showed ~2-fold or lower changes.

We next examined sfGFP expression in samples of each of the cells by flow cytometry (Figure 2) to determine cell-to-cell variability. The CoLE2 and R6K plasmids generally exhibit similar distributions to p15a and pUC origins. Strains JTK160I and JTK164E, which have mean expression levels within 25% of p15a levels, have coefficients of variance that fall within 25% of p15a levels. Similarly, JTK160J and JTK164I, which have mean expression levels within 25% of pUC levels, have coefficients of variance that fall within 25% of pUC levels. It is unclear if the relatively high coefficients of variance for JTK160A and JTK164A are a result of noise at low fluorescence (as evidenced by the very high MCI061 coefficients of variance) or due to true variance in plasmid copy number or GFP expression.

Finally, we monitored the stability of plasmids in mid-level expression variants of both $pir$ and $repA$ after 100 generations without selection (Figure 3). Both $pir$ and $repA$ variants exhibited high stability, losing the plasmid in only 5.2% or 0.5% of cells, respectively. These numbers fall between the stability of the control p15a and pUC plasmids, which lost the plasmid in 23.5% or .25% of cells, respectively.

**Optimization of Violacein Expression**

As a simple demonstration of the utility of the DIAL strains, we optimized the expression level of the violacein biosynthesis operon, VioABCDE. While moderate production levels of the deeply purple metabolite are tolerated in *E. coli*, high levels can be toxic or cause instability [17]. We cloned the operon behind both a weak and a strong constitutive promoter to illustrate the flexibility afforded by controlling copy number. Figure 4 shows the colonies that result from transforming both weakly and strongly expressed operons (jtk3070 and jtk3080, respectively) into all of the DIAL strains. Production of violacein, as indicated by purple coloration, clearly increases as copy number increases, up to some threshold level. Beyond that threshold, colonies begin to grow at reduced rates or not at all. The large colonies in the high copy strains (JTK160G, JTK160I, and JTK164H) with a strong promoter were sequenced and confirmed to be escape mutants in which a fragment of the strong constitutive promoter has recombined out, demonstrating the risk of overstressing the cells.

**Conclusions**

We have developed and characterized two sets of strains that support the R6K and CoLE2 origin of replication at a wide range of different copy numbers enabling rapid exploration of gene and circuit dosage. To accomplish this, we placed the *trans*-acting replication factors from...
each replicon under artificial transcriptional regulation in the genome, leaving only the origins of replication on the plasmids themselves. Although negative feedback relying on elements 5' of the trans-acting factor open reading frame has been implicated as one factor helping to maintain stable copy numbers in both ColE2 [21] and R6K [14], we found that engineered cells stably maintained plasmid copy numbers despite removal of all 5' regulatory elements. This is consistent with the existence of additional feedback mechanisms, as has been suggested for both R6K [22] and ColE2 [23].

To generate copy number diversity in the DIAL strains, we created a library of RBS variants of the trans-acting replication factor in the genome. Although other strategies, such as the use of an inducible promoter or a library of promoters, could also have achieved diverse levels of trans-acting factor expression, varying the RBS enables compatibility with genetic circuits employing any promoter and maintains a consistent noise profile across strains due to stochastic transcription effects [24]. We employed a novel mechanism of generating RBS libraries in the genome: lambda red based integration of Splicing by Overlap Extension (SOEing) [25] PCR products. Multiplex automated genome engineering [26] has also been employed for creating libraries of modified genomic RBSs. While that process is a powerful method for modifying genes already in the genome, this work required simultaneous modification and introduction into the genome of an exogenous gene. In such cases, PCR based integration is an excellent option for library construction, particularly where a relatively small number of variants (<10,000) is sufficient to isolate the desired functionality.

The DIAL strains are the first tool capable of systematically varying genetic circuit dosage without altering the local genetic context. Previous studies examining the impact of circuit dosage in prokaryotes have been largely theoretical (e.g. [27,28]), and in eukaryotes focus only on low (~1-2) copy numbers (e.g. [29]). Because the theoretical predictions suggest that circuit dosage
has a significant impact on the function of some genetic circuits, it is important to empirically verify the robustness or fragility of different circuit architectures. Using the DIAL strains, network behaviour and expression noise can be rapidly assessed at a wide variety of different circuit dosages.

Of great practical use, the DIAL strains offer a rapid, facile mechanism for determining desired expression levels, making it a tool with broad applicability in genetic engineering. The trivial operation of transforming the same plasmid into different strains is sufficient to provide information on the maximum tolerated expression level for a given protein, pathway, or circuit, and screening of viable colonies reveals the optimal expression level for a desired phenotype. We demonstrated this simple capability by optimizing expression of the violacein biosynthesis pathway, which in excess produces moderate toxicity in *E. coli*. Regardless of whether that starting point was a weakly or a strongly expressed operon, deeply purple yet healthy cells were isolated when matched with the appropriate strain. Knowing the optimal gene dosage can be leveraged to change the context of a gene or operon without altering the phenotype. Since the copy number is known, any change in protein dosage resulting from changing the context of the system (such as by integration into the genome) can be compensated for by using existing tools such as the RBS calculator [7] or a set of standard promoters [5].

Importantly, the ColE2 and R6K origins are orthogonal and can co-exist in the same cell, and the two sets of DIAL strains were designed to enable ready combination of both trans factors into a single strain by P1 transduction. Having a single set of cells with both orthogonal origins allows both the relative and absolute levels of two genes or sets of genes to be optimized by, for example, co-transformation into a pool of competent cells. Although R6K has already seen widespread use because of its ability to split into trans and cis elements, having a variety of copy number variants available for both R6K and ColE2 provides an even more powerful toolset for expression level optimization and balancing, circuit dosage investigation, and novel selection schemes.

**Methods**

**Media**

Strains were propagated in LB broth and LB agar plates, with addition of 100 μg/ml ampicillin sodium salt, 50 μg/ml spectinomycin dihydrochloride pentahydrate, 25 μg/ml kanamycin sulphate, and/or 10 μg/ml trimethoprim if appropriate.

**Plasmids**

Plasmids were constructed using BglBrick standard assembly [30]. Full sequences of plasmids are available in Additional file 1, Table S1. The replicon of ColE2-P9 [16] is referred to as ColE2 for convenience. The gamma origin of R6K is similarly referred to simply as the R6K origin.

**Genomic RBS Library Construction**

Template plasmids pBjk2648r-jtk2951 and pBjk2741-jtk3041, illustrated schematically in Figure 5, were first constructed. Splicing by overlap extension SOEing PCR [25] with degenerate oligos (table 2) was used to
Figure 5 Schematic of templates constructed for genomic integration. Plasmids pBjk2648r-jtk2951 and pBjk2741-jtk3041 both contain: 1) a ~500 bp homology arm (HA) matching the 5' end of the genomic insertion site 2) a constitutive promoter (PCON) 3) the ORF of the trans-acting replication factor (Trans) 4) a terminator 5) a kanamycin resistance cassette (KnR) flanked by FRT sites and 6) a ~500 bp HA matching the 3' end of the genomic insertion site. Full sequences can be found in Additional file 1, Table S1.

Plasmid Copy Number Estimation by qPCR
Plasmid copy number per genome equivalent was estimated using the relative quantitation method described previously [33]. Briefly, cells were subcultured 1:100 into fresh media and grown until mid-log or stationary phase before total DNA isolation using QIAamp DNA Mini kits (Qiagen) according to the manufacturer’s instructions. DNA samples and 10-fold serial dilutions of a purified calibrator dna for a given sample in arbitrary units, and then calculating their ratio.

Flow Cytometry
Cells grown to stationary phase were subcultured 1:100 in fresh media, grown until mid-log, resuspended in PBS, and then examined on a Coulter Epics XI-MCl instrument with a 488 nm excitation wavelength and 525 nm emission bandpass filter.

Stability Analysis
Single colonies were picked and grown to stationary phase under selection. Cells were then subcultured 1:106 and grown back to stationary phase without selection, which corresponds to 20 generations of growth. The dilution and regrowth was repeated serially for 100 generations, at which point dilutions of cells were plated on non-selective media, and colonies were examined for sGFP fluorescence as an indicator of plasmid presence.

Additional material
Additional file 1: Lists the full sequences of plasmids and strain modifications used in this study.

Acknowledgements
The authors would like to thank Dr. Tateo Itoh for providing the ColE2-P9 replicon. This work was supported by the National Science Foundation Synthetic Biology Engineering Research Center (SynBERC). JTK was supported by a National Science Foundation Graduate Research Fellowship.

Author details
1Department of Bioengineering, University of California, Berkeley, USA.
2Department of Bioengineering, University of California, San Diego, USA.
3Berkeley National Laboratory, Physical Biosciences Division; QB3: California Institute for Quantitative Biological Research; 327 Stanley Hall, Berkeley, CA 94720, USA.

Authors’ contributions
JTK and JCA designed experiments. JTK and SC carried out experiments. JTK and JCA drafted the manuscript. All authors read and approved the final manuscript.
Competing interests
The authors declare that they have no competing interests.

Received: 30 April 2011 Accepted: 25 July 2011 Published: 25 July 2011

References
1. Jones Prather KL. Low-Copy Number Plasmids as Artificial Chromosomes. In The metabolic pathway engineering handbook: tools and applications. Edited by: Smolke CD. Boca Raton: CRC Press; 2010.
2. Humphreys DP, Carrington B, Bowering LC, Ganesh R, Sehdev M, Smith BJ, King LM, Reeks DG, Lawson A, Poppewell AG. A plasmid system for optimization of Fab’ production in Escherichia coli: importance of balance of heavy chain and light chain synthesis. Protein Expr Purif 2002, 26:309-320.
3. Kim Y, Wang X, Zhang XS, Grigoriu S, Page R, Peti W, Wood TK. Escherichia coli toxin/antitoxin pair MspRI/MspQA regulate toxin CspD. Environ Microbiol 2010, 12:1105-1121.
4. Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of development pathway bifurcation in phase lambda-infected Escherichia coli cells. Genetics 1998, 149:1633-1648.
5. Davis JH, Rubin AJ, Sauer RT. Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res 2011, 39:1311-1341.
6. Smolke CD, Carrier TA, Keasling JD. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. Appl Environ Microbiol 2000, 66:5399-5405.
7. Salis HM, Minsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 2009, 27:946-950.
8. Togna AP, Shuler ML, Wilson DB. Effects of plasmid copy number and runaway plasmid replication on overproduction and excretion of beta-lactamase from Escherichia coli. Biotechnol Prog 1993, 9:31-39.
9. Larsen JE, Gerdes K, Light J, Molin S. Low-copy-number plasmid-cloning vectors amplifiable by derepression of an inserted foreign promoter. Gene 1984, 28:45-54.
10. Wild J, Hradecna Z, Szybalski W. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. Genome Res 2002, 12:1434-1444.
11. Preston A. Choosing a cloning vector. Methods Mol Biol 2003, 235:19-26.
12. Shafferman A, Kotler R, Stalker D, Helinski DR. Plasmid RK6 DNA replication. III. Regulatory properties of the pi initiation protein. J Mol Biol 1982, 161:57-76.
13. Metcalf WW, Jiang W, Warner BL. Use of the rep technique for allele replacement to construct new Escherichia coli hosts for maintenance of RK6 gamma origin plasmids at different copy numbers. Gene 1994, 138:1-7.
14. Horii T, Itoh T. Replication of ColE2 and ColE3 plasmids: the regions sufficient for autonomous replication. Mol Gen Genet 1988, 212:225-231.
15. Akkumal PK, Stephanopoulos G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol 2009, 27:760-765.
16. Chew LC, Tacon WC. Simultaneous regulation of plasmid replication and heterologous gene expression in Escherichia coli. J Biotechnol 1990, 13:47-60.
17. Yasueda H, Horii T, Itoh T. Structural and functional organization of ColE2 and ColE3 replions. Mol Gen Genet 1989, 215:209-216.
18. Mayer M, Schmucker R, Kishan V, Hillen W. Analysis and nucleotide sequence of an origin of DNA replication in Acinetobacter calcoaceticus and its use for Escherichia coli shuttle plasmids. Gene 1990, 87:45-51.