PNAS
www.pnas.org

Supplementary Information for

Codon usage influences fitness through RNA toxicity

Pragya Mittal, James Brindle, Julie Stephen, Joshua B. Plotkin, Grzegorz Kudla

Corresponding author: Grzegorz Kudla
Email: gkudla@gmail.com

This PDF file includes:

Supplementary text
Figs. S1 to S7
Table S1
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Dataset S1

www.pnas.org/cgi/doi/10.1073/pnas.1810022115

Supplementary text

Materials and Methods

Genes, plasmids and bacterial strains

We used a collection of 347 individually cloned full-length synonymous variants of the GFP gene. 154 of these variants came from our previous study [1], while the others were ordered as gBlocks from Integrated DNA Technologies (IDT), generated from existing variants by DNA shuffling [2] or made by site-directed mutagenesis, as described below. The coding sequences of all variants are provided as a fasta file. The GFP library was cloned into the Gateway entry plasmid pGK3, and then into Gateway destination plasmids pGK8, a T7 promoter-based expression plasmid, and pGK16, an expression plasmid with a bacterial *trp*/*lac* promoter [1]. Expression from both plasmids is IPTG-inducible; pGK8 produces untagged GFP, whereas pGK16 encodes a 28-codon 5'-terminal tag with weak mRNA secondary structure, known to facilitate expression [1]. pGK8 was used for GFP expression in the strain BL21-Gold(DE3) [*E. coli* B F^- *omp*T *hsd*S(r_B^- m_B^- *dcm*+ Tet^r *gal* λ(DE3) *end*A Hte], in C41/C43 strains (Lucigen) [3], and in evolved suppressor strains (see below). pGK16 was used for expression in the DH5α strain [F^- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-]

Growth assays

For growth rate analysis, three independent colonies of *E. coli* cells carrying each construct of GFP were grown overnight at 37°C in a 96-well plate with 2 ml wells, with constant shaking (320 rpm), until the cultures reached saturation. Following that, the culture was diluted 1:100 in 200 µl of LB containing ampicillin (100 µg/ml) in a 96-well

plate (Cat No. 655180, Cellstar). The plate was covered with its lid and placed in an automated plate reader (Tecan Infinite M200 Pro/ Tecan Sunrise). After an hour of incubating the plate at the appropriate temperature, typically 37°C, with constant orbital shaking (amplitude-1.5 mm, frequency-335 rpm), the cultures were induced with 1 mM IPTG. Subsequent to induction, plates were incubated in the plate reader with constant shaking as before. To avoid condensation while adding IPTG we retained the lid of the plate in the plate reader chamber which was maintained at 37°C and we avoided prolonged manipulation of plate outside the plate reader. To avoid excessive evaporation of cultures and test for potential contamination, we placed media without bacterial cultures in the external rows and columns of each plate, and only used internal wells for experiments. Optical density (OD) was measured at 595nm and GFP fluorescence was measured with excitation at 485 nm and emission at 515 nm, at fixed time intervals over a period of 6-8 hrs (or 24 hrs in the low-temperature growth assays). LB-only wells were used to normalize the background OD and fluorescence. Bacterial growth rate and fluorescence represent means from three independent experiments, with three replicate measurements in each experiment.

We calculated the growth rates of IPTG-induced cultures as the slope of $\log_2(OD)$ against time, normalized to the slope of non-induced cultures. Thus:

$$\text{growth rate} = \lambda_{\text{induced}} / \lambda_{\text{noninduced}},$$

where $\lambda = (\log_2(OD_t) - \log_2(OD_0)) / t$; OD_0 and OD_t represent the optical densities at the beginning and end of a time interval, and t represents the duration of the time interval.

We defined the time interval as the interval between 1 h and 2.5 h after induction, with a further restriction that the OD of cells is between 0.1 and 1.0.

This formula gives a negative growth rate when the OD of the induced culture decreases over time, seen for example for GFP_170 in Figure 1B. We explain the slight reduction of OD by the lysis of a fraction of cells, mediated by expression of the toxic constructs. Indeed, we could observe cell lysis of GFP_170-expressing cells under the microscope. Typically, the growth rate formula is applied to exponentially growing cells, and only gives positive values for such cells. In our experiments, although non-induced cells grew exponentially (Figure 1A), the growth rate of induced cells changed over time (Figure 1B), due to the combination of reduced growth following the expression of toxic constructs, partial cell lysis, and emergence of suppressors. Thus, the formula is only meant to approximate the behaviour of cells and provide a combined estimate of toxicity. Graphs were plotted using GraphPad Prism 7 Software.

Cell viability assay

The viability of bacterial cultures was estimated by spot assay or more quantitatively by measuring the colony forming units (cfu). For spot assays, BL21 strains carrying a subset of GFP constructs were grown in 2 ml LB with ampicillin, overnight in 14 ml falcon tubes with snap cap, at 37°C. Following growth to saturation, cells were diluted 1:100 in LB containing ampicillin (100 µg/ml) and allowed to grow until OD reached ~0.5. Cultures in exponential phase were then diluted in LB (a factor of 10 between each step) and spotted on to LB plates containing Ampicillin (100 µg/ml) and 1 mM IPTG. Plate containing no IPTG were used as control to show equal number of cells were spotted. A volume of 10 µl was used for spotting.

For quantitative measurements, exponential phase cultures were induced with 1 mM IPTG. Following induction, 1 ml of culture was aliquoted at every 1 hr interval and appropriately diluted (depending on OD) and 100 µl of appropriate dilution was spread on LB-Agar plates containing ampicillin (100 µg/ml). For each culture two different dilutions were spread at each time point in duplicate. Plates were then incubated at 37°C until colonies appeared on them. Viability was assessed by counting the colony forming units (cfu/ml) from the plates.

Microscopic analyses of viability

Microscopy slides were prepared as previously published [4]. Briefly, two plain microscopy slides were cleaned with absolute ethanol. One of the two plastic covers of a gene frame (ABgene; 10 mm x 10 mm) was removed and the adhesive side pressed onto the centre of a glass slide. 1.5% Low Melting Point (LMP) agarose was dissolved in MQ water. 60 µl of the warm agarose solution was pipetted into the centre of each gene frame. The second glass slide was placed on top of the gene frame, avoiding the formation of any air bubbles. The sandwiched slides were allowed to set at 4°C for one hour. Then the upper glass slide was removed by sliding off gently from the agarose bed. BL21 cultures were grown in LB with ampicillin until OD reached ~0.2-0.3, following which they were induced with 1 mM IPTG. Un-induced cultures served as control. Subsequent to induction, aliquots were taken at appropriate time points. 1 µl propidium iodide (Life Technologies, 1mg/ml solution), that stains dead cells preferentially, was

added to the aliquots. The tubes were then incubated at room temperature in dark for 5 minutes. 4 µl of culture was mounted onto the agarose bed and evenly spread on the agarose bed by turning the slide up and down. A clean glass coverslip was adhered to the upper adhesive side of the gene frame avoiding any air bubbles.

Slides were imaged using using 100X Lens on a Zeiss Axio-Observer Z1 inverted microscope (Carl Zeiss UK, Cambridge, UK), with a ASI MS-2000 XY stage (Applied Scientific Instrumentation, Eugene, OR). Samples were illuminated using brightfield or a Lumencor Spectra X LED light source (Lumencor Inc, Beaverton, OR) complete with Chroma #89000ET single excitation and emission filters (Chroma Technology Corp., Rockingham, VT) and acquired on an Evolve EMCCD camera (Photometics, Photometrics, Tucson, AZ). GFP and RFP channels were used to image GFP and Propidium iodide (GFP-excitation: 470/22 nm, dichroic: 495 nm emission: 520/28 nm, RFP- excitation: 542/33 nm, dichroic: 562 nm emission: 593/40 nm). Image was captured using Micromanager (https://open-imaging.com/). For each microscopy slide, at least 10 independent fields were imaged in multi-channel acquisition mode, whilst remaining as unbiased as possible in order to obtain a true representation of the cell number and morphology of cells in the culture. Acquired images were analysed using ImageJ software.

Generation of additional mutated constructs

To prevent the ribosomes from translating, we mutated the Shine-Dalgarno (SD) sequence in seven GFP constructs. All mutations were performed in the pGK3 plasmid[1], using a site directed mutagenesis protocol [5], employing AccuPrime™ Pfx DNA polymerase (Thermo Fisher Scientific). The RBS site aaGAAGGA was changed to

tgTTCTCT. The oligos used were SD_mut_Forward and SD_mut_Reverse primers (see List of oligos). The mutations were confirmed by sequencing and the constructs were then sub-cloned into pGK8 using Gateway cloning. The constructs were then transformed into BL21 cells and growth rates and fluorescence were analysed as described above. Constructs expressing: 1) 132 nt fragments of GFP_012 and GFP_170 with and without start and stop codons ("Frag" and "Frag_(s+s)"), 2) GFP_012 and GFP_0170 with stop codons at 136th and 157th codon ("Stop1" and "Stop2"), and 3) GFP_012 and GFP_170 with transcription terminator sequence inserted at 492 nt position ("TT"), were generated as follows: 132 nt of GFP_170 and its corresponding region on GFP_012 were PCR-amplified using oligos containing BamHI and EcoRI sites (see List of oligos), for cloning into pGK3 plasmid. Start and stop codons were also added to the respective oligos in case of Frag_(s+s) constructs. To introduce TAA stop codons at 136th and 157th codon positions, site directed mutagenesis was carried out using specific oligos on pGK3-GFP_012 and pGK3-GFP_170 plasmids. In the same way we introduced stop codons in all three reading frames at the 157th and 215th codon positions of plasmid pGK3-GFP_170. To introduce Transcription Terminator (TT), 5'end phosphorylated oligos containing 57 nt sequence of TT, were self-annealed and cloned into the HpaI site of GFP_012 and GFP_170, on pGK3 plasmid. To fuse FLAG tags with the toxic fragment of GFP_170 (514-642bp) in all three reading frames, we amplified the GFP fragment from pGK3- GFP_170 with a forward primer containing a BamHI site and three individual reverse primers containing the FLAG tag in three different reading frames along with an EcoRI site. All the above constructs were cloned into pGK3, confirmed by sequencing and subcloned into pGK8 by Gateway cloning. All pGK8 constructs were then transformed into the BL21 strain and growth and fluorescence were analysed as above.

DNA shuffling was performed as previously reported [2], with minor modifications. Briefly, an incomplete DNase I digestion of equimolar concentrations of the two variants

was carried out in the presence of 5 mM $MnCl_2$. Mn^{2+} ions in the reaction ensure DNaseI digests both strands of DNA at approximately the same sites [6]. To achieve controlled digestion DNaseI treatment was performed for only 2 minutes at 15°C before inactivating the enzyme at 90°C for 5 minutes. Digested products were assembled by primerless assembly to obtain larger fragments of expected size. Assembly PCR was performed using Q5 high fidelity DNA polymerase (NEB) and PCR conditions were as follows: Annealing temperature: 45°C, extension time: 30 secs for 40 cycles. The above step was followed by re-amplification with oligos pENTR_seq_U6 and pENTR_seq_L3. We obtained 36 GFP constructs from this experiment that were made of randomly shuffled fragments of GFP_012 and GFP_170. The shuffled variants encoding the GFP protein sequence were cloned into the pGK16 vector using Gateway cloning and transformed into DH5α for analysis of growth phenotype.

To make synonymous mutations in the region spanning nts 534-642 in GFP_170, we designed degenerate oligos in five windows of 20-25 base pairs. In each window all wobble positions were mutated synonymously, allowing all possible changes at a given position. Site directed mutagenesis was performed using oligo sets A, B, C, D and E (see list of oligos) and AccuPrime Pfx DNA Polymerase (Thermo Fisher Scientific). All mutagenesis were carried out on the pGK3-GFP_170 plasmid. Mutations were confirmed by sequencing and the constructs were then sub-cloned into pGK8. The number of mutations per construct that we generated ranged from 2-9 and we obtained 98 constructs from five sets of PCRs. Single mutations were also generated in the region spanning 540-620 bp. Each wobble position was mutated synonymously, allowing all possible changes. We generated 36 constructs such that each construct had only one synonymous mutation

per construct at a given codon in the region. Codons which were exactly the same between GFP_012 and GFP_170 were not mutated.

Bovine mitochondrial 2-oxoglutarate carrier protein (OGCP) constructs: wild type OGCP (OGCP_WT), OGCP with Shine-Dalgarno sequence changed from GAAGGA to TTCTCT and with no start codon (OGCP_noRBS), OGCP with *E. coli*-optimized codons (OGCP_CO), were purchased as gBlocks from IDT.

mKate2 constructs: A mKate2 gene fusion with the toxic fragment of GFP_170 was also ordered as a gBlock from IDT. The fragment contained BamHI and EcoRI sites for cloning into the pGK3 plasmid. The mKate2 gene by itself was amplified from the mKate-GFP_170 fusion construct using primers containing BamHI and EcoRI sites for cloning into pGK3. All constructs were confirmed by sequencing and subcloned into pGK8 by Gateway cloning.

Isolation and validation of genetic suppressors

BL21 cells carrying several GFP variants (both toxic and non-toxic) were plated on LB agar supplemented with Ampicillin (100 μg/ml) and 1 mM IPTG. We obtained two kinds of colonies on the plates: highly fluorescent small colonies and large white colonies. We picked primarily the green colonies and a few white colonies for further analyses. All the colonies that were picked were plated on LB Agar+Amp plates. All green and some white colonies grew on Amp plates while some of the white colonies couldn't grow any further on Amp plates. 37 colonies that grew on Amp plates (30 green and 7 white) were selected for further study. The growth rate and GFP fluorescence levels were measured for all colonies as described above.

To validate that the mutation that affects the survival of cells on IPTG is located on the chromosome and not on the plasmid itself, we cured the strains of the plasmid. For curing, the colonies were streaked on LB Agar plates in absence of Ampicillin repeatedly for at least 3-4 rounds. Colonies obtained after growing without antibiotic selection were further replica plated on LB Agar and LB Agar+Amp. Colonies that grew on only LB Agar but not on LB Agar+Amp were cured of the plasmid. These cured strains were re-transformed with the same GFP variants from which they were isolated. After re-transformation these cured strains were plated on LB Agar+Amp+IPTG plates. We obtained only bright green colonies from the re-transformed cured strain that was originally bright green. However, the cured strain from white colonies, on retransformation with the same GFP plasmid, produced a mix of green and white colonies on IPTG plates.

To further validate that the mutation was not located on the plasmid we isolated plasmids from the 37 colonies and transformed them into fresh competent BL21 strain and assayed the growth and fluorescence. The phenotype was the same as in the parental strains, showing that the isolated plasmids did not carry any mutations that affected the phenotype.

To identify the genomic mutations that conferred the suppressor phenotype we selected 22 suppressors (green=18, white=4) for genome sequencing. We also sequenced the genomic DNA from two independent BL21 parental colonies to serve as reference and control during the analysis of genome sequences.

Analysis of genome sequence and variant calling

Chromosomal DNA was isolated using the Wizard® Genomic DNA Purification Kit (Promega, U.S.A.) according to the manufacturer's instructions. The concentration of genomic DNA was estimated by Qubit dsDNA BR Assay Kit (ThermoFischer Scientific). Quantitation and quality control of genomic DNA was performed on a Bioanalyzer (Edinburgh Genomics UK). Genomic DNA samples were supplied in required concentrations for Nextera XT Library preparation, followed by 250-bp paired-end HiSeq Illumina sequencing (Edinburgh Genomics, UK).

The reads were mapped onto the reference genome sequence of BL21-Gold(DE3) (GenBank Accession ID CP001665.1) with default settings using bwa [7]. PCR duplicates were marked using Picard tools (http://broadinstitute.github.io/picard/). Genomic variants (SNPs, indels and insertions) were called using GATK [8]. We used GATK haplotype caller with ploidy=1, stand_call_conf=30 and stand_emit_conf=10. Variants were filtered with parameter settings: DP<9.0 and QUAL<10.0. Bedtools [9] was used to detect unique variation in our suppressor strains in comparison to the control strain and the reference genome. Finally, the identified variations were confirmed by targeted PCR amplification followed by Sanger sequencing. As the lac promoter is duplicated in the BL21-Gold(DE3) strain, wild type lac promoter (P_{lacWT}) and the *lac*UV5 promoter (P_{lacUV5}) driving the expression of T7 polymerase, we obtained dual peaks in targeted sequencing of P_{lac}UV5 promoter region. To resolve this, we carried out a detailed analysis of this region by extracting all read pairs where one read of the pair was mapped on to an unduplicated region and the read pairs were unambiguously assigned to the specific loci on the genome. The genome sequencing results can be accessed on https://www.ncbi.nlm.nih.gov/sra/srp149903.

Statistical analyses

We annotated the GFP sequences with a range of sequence-derived parameters and experimental measurements. The codon adaptation index (CAI) was calculated as in [1] using codon optimality scores from [10]. GC3 content (GC content at the third positions of codons) was calculated by dividing the number of G- and C-ending codons by the total number of codons. Folding energy within the window (-4 to +38) relative to the translation start codon [1] was calculated using hybrid-ss-min from the UNAFold package [11]. Translation initiation rate was calculated in a window from -40 to +60 relative to the start codon using the RBS Calculator [12]. Growth rate was calculated as described in the "growth assays" section above. OD was measured 3 hours after IPTG induction, after subtraction of LB-only background. OD and fluorescence measurements from a previous study [1] were used after converting their units to units measured in the present study with a linear least squares model. Protein level measurements by Coomassie staining and RNA measurements by Northern blotting were from a previous study [1]. Protein abundance per cell was calculated by dividing protein fluorescence by OD.

To map the toxicity-determining region of GFP_170 based on the DNA shuffling experiment, we used Student's t-test for each synonymous position *i* to compare the growth rates of variants in which position *i* was derived from GFP_170 and from GFP_012. We applied a Bonferroni correction for 239 tests, resulting in a p-value cutoff of 0.0002 (0.05/239). In this analysis, positions 532-640 from GFP_170 were associated with significantly slower growth of shuffled variants. We conservatively defined a slightly larger fragment (nts 512-645 of GFP_170) as the putative toxicity-

determining region. We subsequently narrowed down this region based on the results of mutagenesis experiments.

Regression analyses were performed in the R software package. Correlations reported in the text are quantified by the Spearman rank correlation coefficient and its associated p-value. We performed multiple regression analyses in order to quantify the relative importance of the various predictor variables in determining growth rates and optical densities. The output of these analyses, shown below, highlights the predominant influence of toxic mRNA fragments on growth:

Multiple regression. Dependent variable: growth rate, BL21 cells

Coefficients	Estimate	Std. Error	tvalue	Pr(>|t|)	Significance
GFP_155_nt490-720	-0.30111	0.05081	-5.927	1.48E-07	***
GFP_170_nt514-645	-0.41221	0.04715	-8.743	2.05E-12	***
CAI	0.24554	0.1678	1.463	0.148444	
GC3	-0.03029	0.09525	-0.318	0.751524	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1077 on 62 degrees of freedom

Multiple R-squared: 0.6612, Adjusted R-squared: 0.6394

F-statistic: 30.25 on 4 and 62 DF, p-value: 5.752e-14

Multiple regression. Dependent variable: growth rate, DH5a cells

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	Significance

GFP_155_nt490-720	-0.19791	0.05044	-3.924	0.000183	***
GFP_170_nt514-645	-0.16322	0.03054	-5.344	8.37E-07	***
CAI	0.17294	0.17654	0.98	0.330226	
GC3	-0.09352	0.08833	-1.059	0.292866	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1071 on 80 degrees of freedom

Multiple R-squared: 0.3628, Adjusted R-squared: 0.331

F-statistic: 11.39 on 4 and 80 DF, p-value: 2.296e-07

Multiple regression. Dependent variable: OD, BL21 cells

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	Significance
GFP_155_nt490-720	-0.20718	0.02904	-7.135	2.29E-11	***
GFP_170_nt514-645	-0.40684	0.03874	-10.501	< 2e-16	***
CAI	0.36495	0.08492	4.298	2.82E-05	***
GC3	-0.04033	0.06678	-0.604	0.547	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09221 on 180 degrees of freedom

Multiple R-squared: 0.5559, Adjusted R-squared: 0.546

F-statistic: 56.32 on 4 and 180 DF, p-value: < 2.2e-16

Western blotting

For Western blotting, BL21strains carrying non-toxic and toxic GFP constructs +/- RBS were grown in 2 ml LB with ampicillin, overnight in snap cap tubes, at 37°C. Following growth to saturation, cells were diluted 1:100 in LB containing ampicillin (100 µg/ml) and allowed to grow until OD reached ~0.5 and then induced with 1mM IPTG. Un-induced samples were collected before adding IPTG as control. After 1.5 h of induction, 1-2 ml of cultures were pelleted. Pellets were re-suspended in standard RIPA buffer and briefly sonicated in presence of Protease inhibitor (Roche) to lyse the cells. The lysate was further spun at 14000rpm for 5 mins to get rid of debris and the total protein was estimated by BCA assay (Pierce BCA protein estimation kit). 10µg of protein was resolved on 10% Bis-Tris gel. Prestained PageRuler protein ladder (ThermoFisher Scientific) was used as standard. Following electrophoresis the gel was transferred onto Nitrocellulose membrane using iblot2 gel transfer device (Invitrogen). The following antibodies were used for detection: Polyclonal Anti-GFP antibody (ab290, abcam), 1:5000, and goat anti-rabbit IgG-HRP conjugate (Santa Cruz Biotech, SC2030), 1:10000.

In the case of Flag fusion constructs, cells were grown and processed as described above. 13 µg of protein was resolved on 4-12% Bis-Tris gel. Prestained Benchmark protein ladder (Invitrogen) was used as standard. The following antibodies were used for detection, Flag M2 Monoclonal antibody (F3165, Sigma), 1:2000 and goat anti-mouse IgG-HRP conjugate (Santa Cruz Biotech, SC2031), 1:10000. The membranes were developed by soaking in Chemiluminescent substrate (Protein simple) and blots were imaged on Imagequant LAS4000 (GE Healthcare).

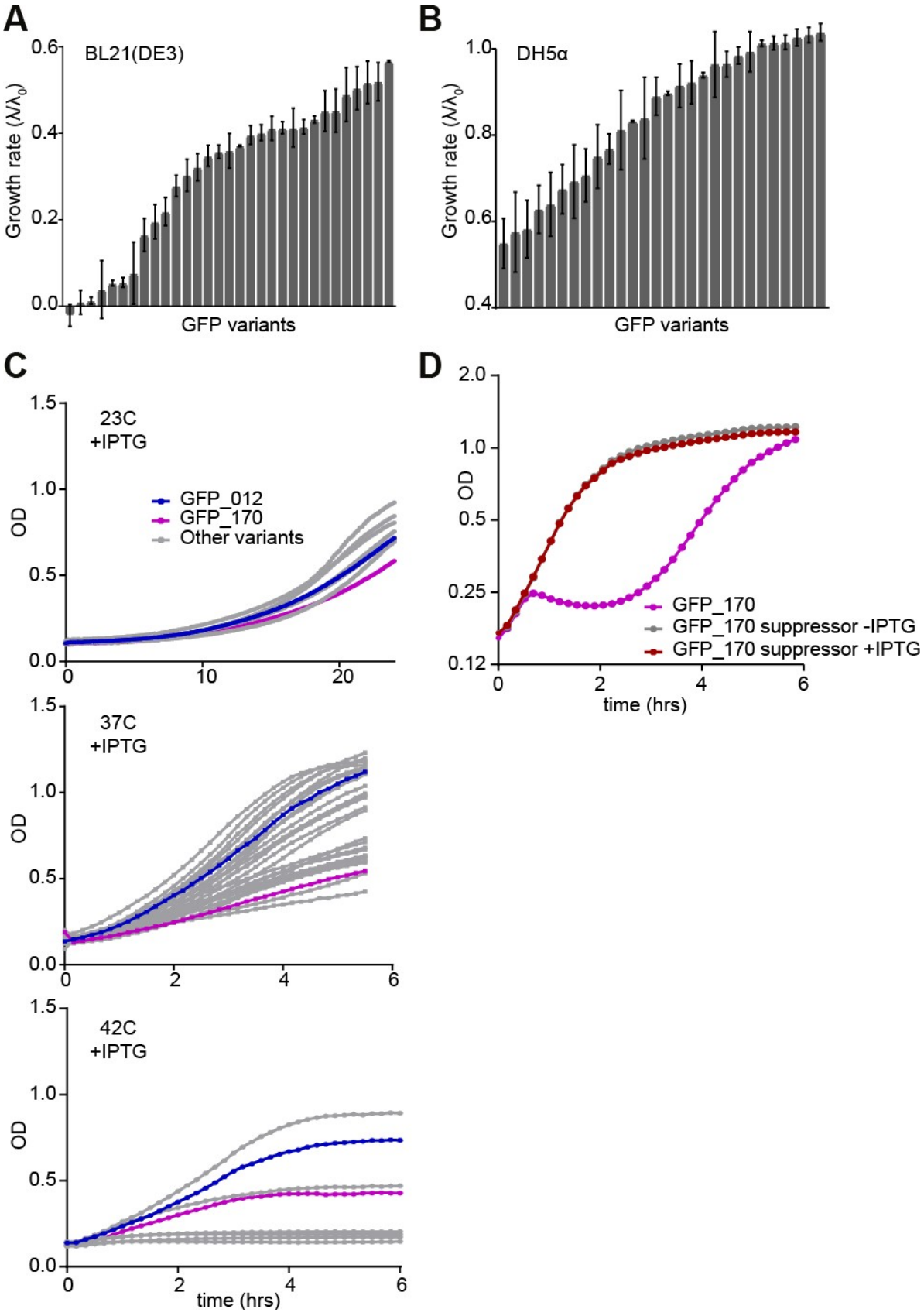
A
BL21(DE3)
Growth rate (λ/λ0)
GFP variants
B
DH5α
Growth rate (λ/λ0)
GFP variants
C
23C
+IPTG
GFP_012
GFP_170
Other variants
OD
37C
+IPTG
OD
42C
+IPTG
OD
time (hrs)
D
OD
GFP_170
GFP_170 suppressor -IPTG
GFP_170 suppressor +IPTG
time (hrs)

Fig. S1. Growth phenotypes of GFP variants. (A-B) Growth rates in BL21 cells (A) and DH5α (B) in the presence of IPTG, sorted from minimum to maximum growth rate in each strain. **(C)** Growth curves of DH5α cells at different temperatures (23°C, 37°C and 42°C) in presence of IPTG. At 23°C there are minor variations in growth of cells expressing GFP variants, at 37°C there are large variations, and at 42°C, some of the GFP variants fail to grow altogether. GFP_012 (non-toxic, blue), GFP_170 (toxic, magenta), other variants (grey). The growth curves represent averages of at least 6 replicates. (D) Growth curve of BL21 cells expressing GFP_170 (magenta); suppressor isolated after back-diluting cells expressing GFP_170 in presence (red) and absence (grey) of IPTG. The suppressor strain has similar growth phenotypes both in presence and absence of IPTG.

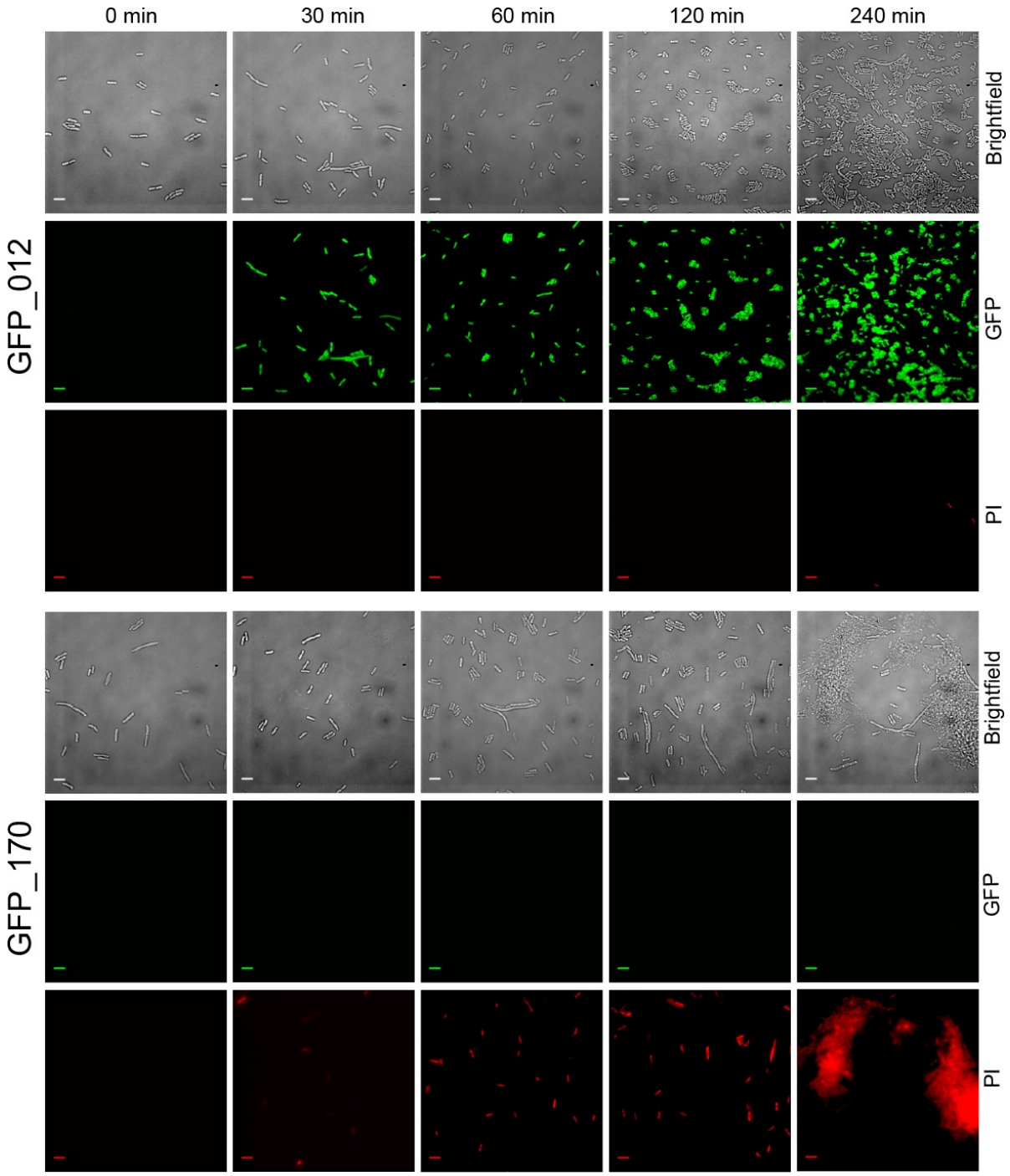

Fig. S2. Microscopic analysis of cell viability. Cell viability was estimated for BL21 cells expressing GFP_012 (non-toxic variant) and GFP_170 (toxic variant). Brightfield images give an estimate of cell morphology and densities. GFP and RFP channels were

used to determine the number of cells expressing GFP and the number of dead cells stained by Propidium Iodide (PI) respectively. At 0 min (just before IPTG induction) GFP_012 and GFP_170 cultures have similar cell densities and morphology. For cells expressing GFP_012, we see a steady increase in cell number after induction and GFP expression appears after 30 mins of induction. There is no significant cell death (PI stained cells) at any given time point. For cells expressing GFP_170 cell densities do not increase rapidly and most cells lose their morphology. We see a rapid increase in number of dead cells and the severity of the phenotype can be estimated at 240 min time point when PI staining shows only dead cells or debris from the dead cells. GFP expression is not seen for GFP_170 due to a strong mRNA secondary structure at its 5' end, impeding its translation. The scale bar is 5 μm.

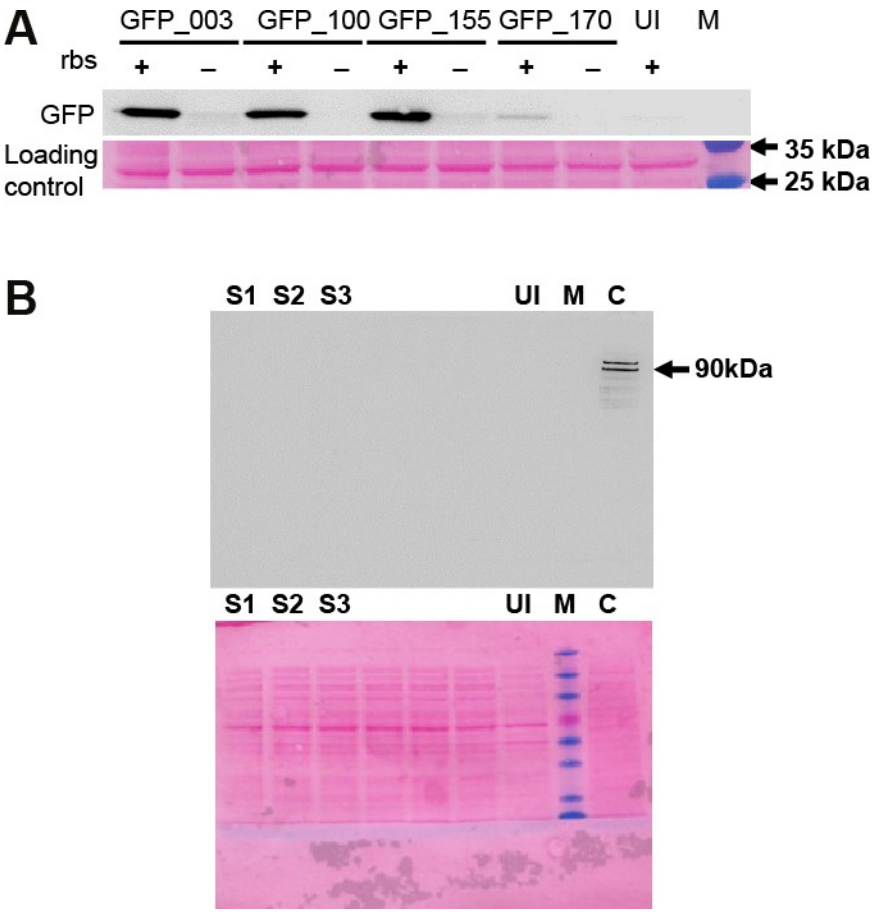

Fig. S3. Measurement of GFP expression by Western blotting (A) Expression of four toxic variants of GFP in the presence and absence of RBS. UI, uninduced control; M, marker. GFP expression was analysed by probing with anti-GFP polyclonal antibody (abcam 290). Ponceau stained blot shows equal loading. **(B)** GFP_170 toxic fragment (nt 514-645) expression fused to FLAG tag in all three reading frames (S1, S2, and S3) was analysed by probing with monoclonal Anti-FLAG (F3165 sigma). UI, uninduced control; M, marker; C, control sample expressing two Flag-tagged proteins of size 116 and 90 kDa. No FLAG expression was detected from S1, S2 or S3 constructs.

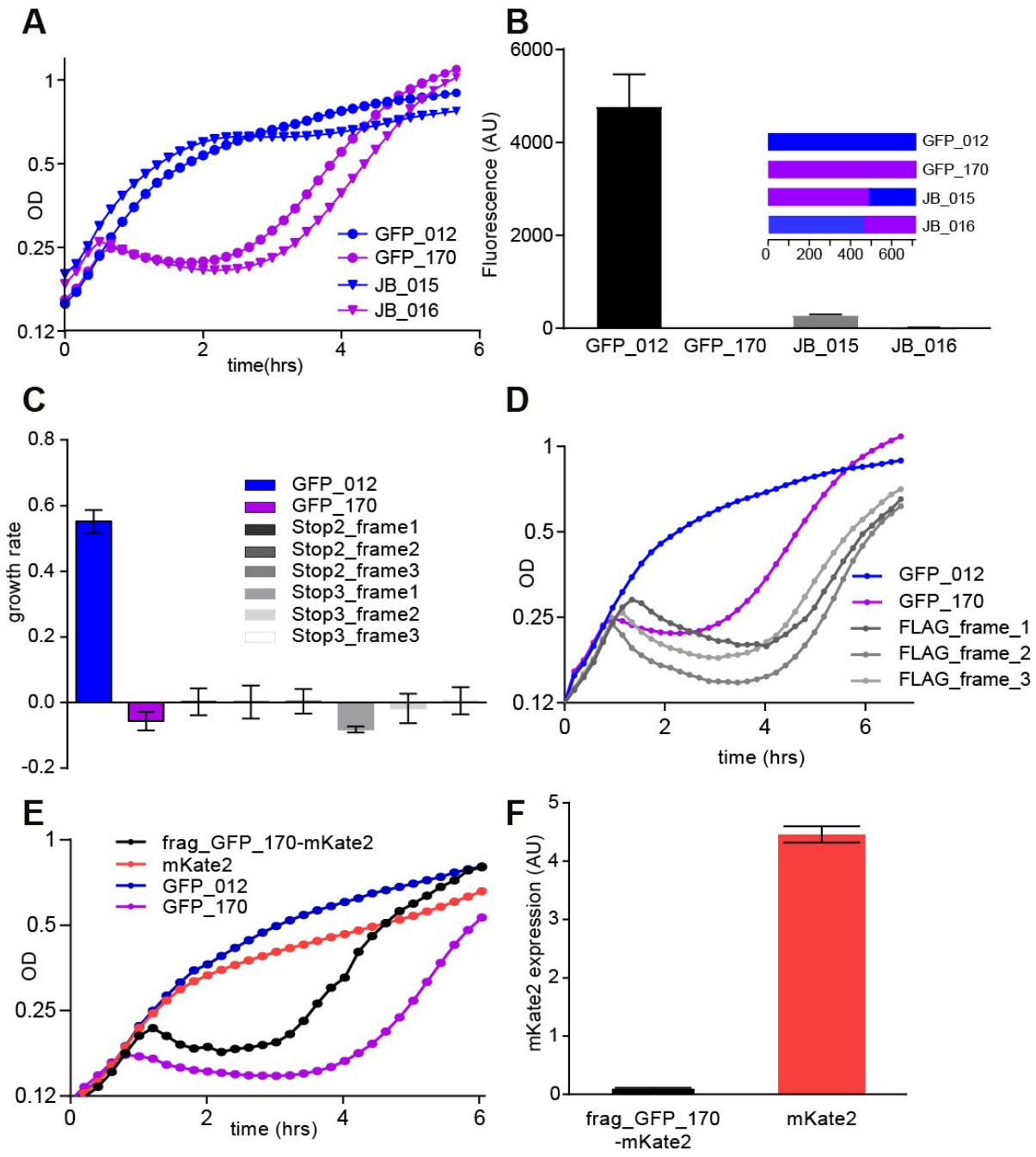

Fig. S4. The toxic element resides near the 3'end of GFP_170 and toxicity is independent of translation. (A) Growth curve for BL21 cells expressing constructs GFP_012, GFP_170 and their shuffled variants JB_015 and JB_016. JB_015 consists of GFP_170 (nts 1-497) and GFP_012 (498-720); JB_016 consists of GFP_012 (1-449) and GFP_170 (450-720). **(B)** Fluorescence of the shuffled constructs. JB_015 is non-toxic

and shows a low level of fluorescence; JB_016 and GFP_170 are toxic and almost non-fluorescent. **(C)** Growth rate of cells expressing GFP_170 constructs with internal stop codons before and after the toxic fragment (nt 514-645) in all three reading frames. TAA stop codons were inserted at nucleotide positions 469 (stop2_frame1), 470 (stop2_frame2) and 471 (stop2_frame3) upstream of the toxic fragment and 643 (stop3_frame1), 644 (stop3_frame2) and 645 (stop3_frame3) downstream of toxic fragment. (D) Growth curves of constructs having toxic fragment from GFP_170 fused to FLAG tag at the 3' end in all three reading frames. All three constructs retain toxicity. (E) Growth curves of mKate2 and toxic GFP_170 fragment fused to mKate2 at the 5' end. Fusion construct retains toxicity (F) Expression of mKate2. No fluorescence is detected when mKate2 is fused with the toxic fragment from GFP_170.

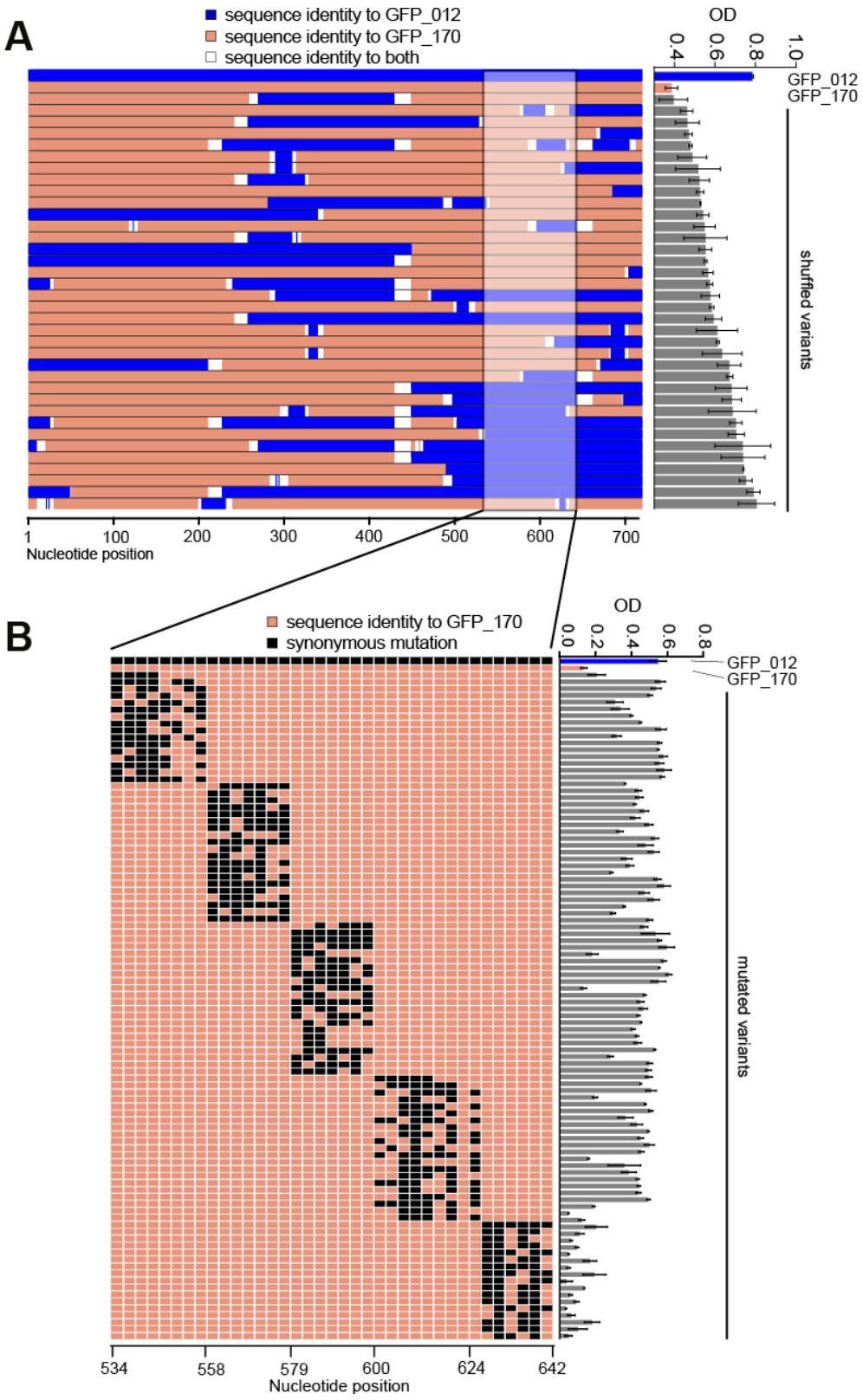
A
sequence identity to GFP_012
sequence identity to GFP_170
sequence identity to both
OD
0.4
0.6
0.8
1.0
GFP_012
GFP_170
shuffled variants
1
100
200
300
400
500
600
700
Nucleotide position
B
sequence identity to GFP_170
synonymous mutation
OD
0.0
0.2
0.4
0.6
0.8
GFP_012
GFP_170
mutated variants
534
558
579
600
624
642
Nucleotide position

Fig. S5. Growth analysis of GFP constructs generated by shuffling and multiple synonymous mutations. (A) 36 constructs were generated by DNA shuffling of GFP_012 (blue) and GFP_170 (orange). All constructs encode full length GFP. Constructs are colour coded according to the sequence identity with GFP_012 and GFP_170. The constructs from top to bottom are arranged in ascending order of their growth (OD 595nm). The highlighted region shows that most constructs having sequence identical to GFP_170 (orange) in 520-620 nt region are toxic. **(B)** An inset of the highlighted area from Panel A summarizes the results of multiple synonymous mutations that were generated in the toxic region. Each row represents a particular mutated variant and each column represents the nucleotide position. Columns highlighted orange and black represent nucleotides identical to GFP_170 and synonymous substitutions respectively. Each construct has 2-9 substitutions. Synonymous mutations in the region 534-624 nt reduce or abolish the toxicity of GFP_170 but any number of synonymous mutations in 627-642 nt region had no effect on toxicity. All data are averages of 9 replicates, +/- SEM.

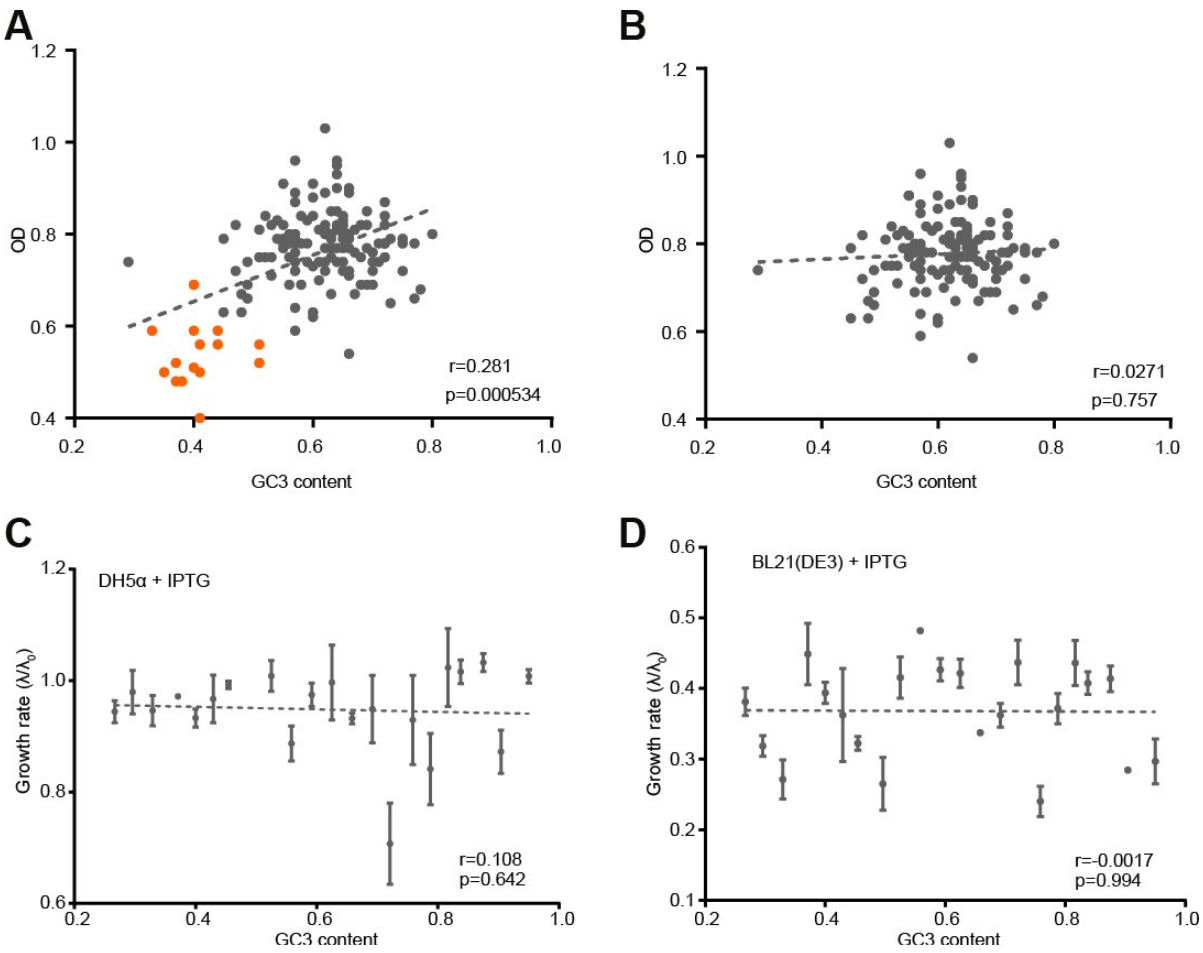

Fig. S6. No correlation between GC3 content and growth rate of GFP variants. (A-B) The correlation between GC3 content and growth (OD 595nm) of GFP variants in BL21 cells is driven by two toxic RNA fragments shared between a number of variants: GFP_155 nt 490-720, and GFP_170 nt 514-645, marked in orange. After removal of these variants (panel B), we no longer see any relationship between GC3 content and growth. **(C-D)** There is no relationship between GC3 content and growth in an independent set of 22 GFP constructs, either in DH5α (C) or BL21 (D) strains. All data are averages of 9 replicates, +/- SEM.

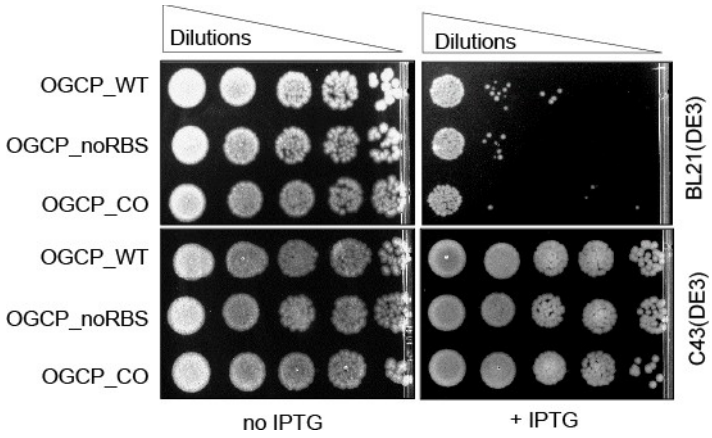

Fig. S7. Spot assay for semi-quantitative estimation of cell viability of BL21 cells expressing OGCP variants. OGCP-WT (wild type OGCP), OGCP_noRBS (OGCP lacking functional RBS) and OGCP_CO (codon-optimized OGCP) variants were cloned in pGK8 plasmid and transformed in BL21 and C43 strains. In the absence of IPTG there are no difference in the viabilities between strains or constructs; in the presence of IPTG, the three constructs are toxic in BL21 cells but not in C43 cells.

Table S1. Analysis of suppressor genotypes. 15/18 green suppressors showed a complete replacement of P_{lacUV5} promoter with P_{lacWT}, 3/18 showed replacement of P_{lacUV5} with $P_{lacWeak}$. 3/4 white suppressors had no changes in the promoter of T7 RNA polymerase, while for 1/4 we could not definitively assign the promoter type.

Strain	Isolated from	Color	Genotype
Sup_01	GFP_003	Green	$P_{lacWeak}$
Sup_02	GFP_003	Green	P_{lacWT}
Sup_03	GFP_003	Green	$P_{lacWeak}$
Sup_04	GFP_003	Green	P_{lacWT}
Sup_05	GFP_003	Green	P_{lacWT}
Sup_06	GFP_003	Green	P_{lacWT}
Sup_07	GFP_003	Green	P_{lacWT}
Sup_08	GFP_003	Green	P_{lacWT}
Sup_10	GFP_003	Green	P_{lacWT}
Sup_12	GFP_003	White	P_{lacUV5}
Sup_14	GFP_069	Green	P_{lacWT}
Sup_15	GFP_069	Green	P_{lacWT}
Sup_17	GFP_069	Green	P_{lacWT}
Sup_18	GFP_069	White	P_{lacUV5}
Sup_21	GFP_183	Green	P_{lacWT}
Sup_22	GFP_183	Green	$P_{lacWeak}$
Sup_24	GFP_183	White	P_{lacUV5}
Sup_26	GFP_155	Green	P_{lacWT}
Sup_30	GFP_155	Green	P_{lacWT}
Sup_34	GFP_100	Green	P_{lacWT}
Sup_35	GFP_100	Green	P_{lacWT}
Sup_37	GFP_170	White	P_{lacUV5}/ $P_{lacWeak}$**

** We observed a mix of two types of reads in sequencing analysis for this strain

References

1. Kudla, G., Murray, A.W., Tollervey, D. & Plotkin, J.B. Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324**, 255-258 (2009).
2. Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* **91**, 10747-10751 (1994).
3. Miroux, B. & Walker, J.E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289-298 (1996).
4. de Jong, I.G., Beilharz, K., Kuipers, O.P. & Veening, J.W. Live Cell Imaging of Bacillus subtilis and Streptococcus pneumoniae using Automated Time-lapse Microscopy. *Journal of visualized experiments : JoVE* (2011).
5. Braman, J., Papworth, C. & Greener, A. Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods Mol Biol* **57**, 31-44 (1996).
6. Lorimer, I.A. & Pastan, I. Random recombination of antibody single chain Fv sequences after fragmentation with DNaseI in the presence of Mn2+. *Nucleic Acids Res* **23**, 3067-3068 (1995).
7. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
8. Van der Auwera, G.A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics* **43**, 11 10 11-33 (2013).
9. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
10. Sharp, P.M. & Li, W.H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281-1295 (1987).
11. Markham, N.R. & Zuker, M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* **453**, 3-31 (2008).
12. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950 (2009).