Codon optimality is the primary contributor to the exceptional mutational sensitivity of CcdA antitoxin in its operonic context

Soumyanetra Chandra[1#], Kritika Gupta[1#], Shruti Khare[1], Pehu Kohli[1], Aparna Asok[1], Sonali Vishwa Mohan[2], Harsha Gowda[2], Raghavan Varadarajan [1*]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India

[2]Institute of Bioinformatics, Bangalore-560100, India

[#]Authors have contributed equally

[*]Corresponding Author

Raghavan Varadarajan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India, Email : varadar@iisc.ac.in, Phone: +91-80-22932612, Fax: +91-80-23600535

Abstract

Deep mutational scanning studies suggest that single synonymous mutations are typically silent and that most exposed, non active-site residues are tolerant to mutations. Here we show that the *ccdA* antitoxin component of the *E.coli ccdAB* toxin-antitoxin operonic system is unusually sensitive to mutations when studied in the operonic context. A large fraction (~80%) of single codon mutations, including many synonymous mutations in the *ccdA* gene show inactive phenotypes that are correlated with the *E.coli* codon usage frequency but retain native-like binding affinity towards cognate toxin, CcdB. Therefore, the observed phenotypic effects are largely not due to alterations in protein structure or stability, consistent with the fact that a large region of CcdA is intrinsically disordered. In select cases, proteomics studies reveal altered ratios of CcdA:CcdB protein levels *in vivo*, suggesting that these mutations likely alter relative translation efficiencies of the two genes in the operon. We extend these results by predicting and validating single synonymous mutations that lead to loss of function phenotypes in the *relBE* operon upon introduction of rarer codons. Thus, in their native context, genes are likely to be more sensitive to both synonymous and non-synonymous point mutations than inferred from previous saturation mutagenesis studies.

Keywords: codon preference, saturation mutagenesis, translational efficiency, ribosome pausing, protein and RNA levels, protein degradation

Introduction

The effects of mutations on protein function have been studied extensively in mono-cistronic proteins or in isolated, heterologous expression systems. Such studies commonly rely on activity or phenotypic assays that help establish gene function or identify residue specific contribution to protein function. Studies of bacterial genomes reveal a high occurrence of functionally related proteins proximally spaced on a DNA stretch, under a shared promoter [1]–[5]. Such genetic systems, referred to as operons, often exhibit characteristic signal dependent, tightly regulated gene expression from a common promoter. The different gene products are then translated from the same mRNA molecule and these genes are usually separated by less than 20 base pairs [6]. The translation of genes clustered in operons is often dependent on the space between genes [7]. Consequently, DNA and mRNA sequence features are expected to play an important role in gene expression and regulation in operons. Moreover, transcription and translation are known to be spatiotemporally coupled in prokaryotes [8]–[10] which further complicates the study of mutational effects on gene functions in operonic systems.

Mutations that lead to changes in the protein's amino acid sequence are often expected to alter the protein structure, stability and/or activity. However, the effect of mutations can be manifold and can affect genetic information, encoded not solely in the amino acid sequence but also in the DNA and RNA sequences. Unlike missense mutations that alter the amino acid code, synonymous mutations alter only the DNA (and mRNA) sequence and not the amino acids, and arise due to the degeneracy of the genetic code [11]. Synonymous mutations were classically believed to be silent in their effects on protein function. While this is often true, there have been several pioneering studies that show prominent effects of synonymous mutations on protein function [12]–[14]. Synonymous mutations are also reported to be associated with over 50

human diseases [15]. Some of the mechanisms known to be involved in the altered phenotypes, observed upon introduction of multiple synonymous mutations, are due to the effects on protein expression and folding [16], [17], transcription [18], as well as mRNA stability due to changes in secondary structure or altered mRNA decay rates [19], [20]. Such effects are often found to be brought about by the bias due to unequal codon usage [21]–[23] and tRNA abundance [24] as well as effects on translation initiation [25], elongation or termination [17]. Such surprising effects seen in the case of synonymous mutations indicate that a similar impact on gene functions can also be produced by modifications in DNA sequences due to non-synonymous mutations. Such mutational consequences of altered DNA sequence on transcription and translation might have significant phenotypic effects, including when studied in an operonic context. In tightly regulated operonic systems, downstream functions are often acutely dependent on the relative levels of gene products of the operonic genes.

Type II Toxin-Antitoxin (TA) systems, where the antitoxin and the toxin proteins are expressed as part of a single operon, serve as an attractive model to study co-regulation of gene expression as well as transcriptional and translational coupling [26]. These TA systems usually code for an upstream labile antitoxin protein responsible for negating the toxicity of the downstream toxin protein, the latter can cause growth arrest or cell death [27], [28]. Maintaining a perfect balance between the antitoxin and toxin proteins is therefore essential for proper functioning of TA modules and small changes in gene expression caused by altered DNA sequences, both in coding or non-coding regions can potentially be detrimental to this balance. Although the distance between the stop codon of the first gene and start codon of the second gene can vary, bioinformatics studies have shown that in the case of type II TA systems it is typically between 0-5 nucleotides [29]. In some cases, the two ORFs even overlap with each other [27]. Post-

translationally, the two proteins are subjected to differential degradation. The ratio of antitoxin to toxin is expected to diminish during stress conditions. In the case of drug tolerant persisters, some studies have suggested that the accumulation of the toxin leads to metabolic dormancy by slowing vital cell processes such as replication, transcription and translation, though this is a controversial area where many ambiguities remain [30]–[32].

Although conditional co-operativity (co-operative interaction of TA complex with the operator DNA) is known to be a dominant mechanism in regulating the transcription of most TA systems [33], mechanisms that can lead to differential levels of the toxin and the antitoxin protein from a common mRNA are not well understood. For conditional co-operativity to manifest, the default assumption is that the antitoxin is present at significantly higher levels than the toxin, a notion which has not been rigorously tested. Given the higher rate of degradation for the antitoxin due to its disordered nature, the antitoxin needs to be produced at a higher rate than the toxin, for the cells to be viable. While protein levels can in principle be regulated at both the transcriptional as well as the translational level, RNA-seq data from *E.coli* indicate that mRNA levels for the toxin and the antitoxin are not significantly different, as is expected for genes encoded by a single polycistronic mRNA [27]. Ribo-seq data available for TA systems indicate that the protein synthesis rates (calculated from ribosome densities) for the antitoxin are at least two fold higher than the toxin, for six TA pairs for which sufficient data is available [27]. These observations together indicate that there are indeed nucleotide sequence dependent features, which allow regulation at the level of translation, leading to differential synthesis rates for the two genes in the operon. We therefore investigated the molecular bases for such regulation by carrying out saturation mutagenesis studies of the antitoxin *ccdA,* from the CcdAB TA system in *E.coli*.

The *E. coli* F-plasmid borne *ccd* operon encodes a labile 72 residue CcdA antitoxin, that prevents killing of the bacterial cells by binding to the 101-residue toxin, CcdB. Both genes are co-expressed in low amounts in F-plasmid bearing *E. coli* cells, and their expression is autoregulated at the level of transcription [34]. If the cell loses F-plasmid, the labile CcdA is degraded by the ATP dependent Lon protease, releasing CcdB from the complex to act on its DNA gyrase target. This eventually leads to cell death [35]. Therefore, mutations which disrupt CcdA antitoxin function, can lead to cell death under normal conditions. Using saturation mutagenesis, we find that the labile antitoxin CcdA apparently shows considerably higher mutational sensitivity than most proteins studied to date [36]. Several synonymous point mutations in CcdA lead to loss of function in a manner correlated to the *E. coli* codon preference. Our data suggests that reduction in optimality of the codons upon mutation results in decreased levels of CcdA protein, likely by altering the translation efficiency. The reduced CcdA: CcdB protein ratio *in vivo* in such mutants, results in heightened toxic effects on bacterial cells. Moreover, the increased CcdB: CcdA ratio results in upregulation of *ccd* mRNA expression, further amplifying the effect of mutation, resulting in cell death. Improved translation initiation of the downstream toxin CcdB, upon mutations in the *ccdA* gene, especially in the ribosome binding site (RBS) of CcdB, also appears to contribute to increased toxicity in cells. The *ccd* and likely most other TA operons, are highly sensitive genetic circuits that can be used to probe effects of mutations on gene function, *in vivo*. We observe that a large number of single codon variants of the *ccdA* gene (both synonymous as well as non-synonymous mutations) produce inactive phenotypes. In contrast, most studies till date suggest that multiple synonymous mutations are required to observe a discernible phenotype. Subtle differences in fitness that are often undetectable in laboratory experiments can become fixed in the population over long

timescales [37], [38] and laboratory selection conditions can differ greatly from those experienced during natural selection. Hence, to properly understand the contribution of single mutations, it is useful to have experimental systems, such as the present one, where small changes in protein activity have substantial phenotypic effects that are largely independent of external conditions.

Materials and Methods

Preparation of a single-site saturation mutagenesis library of CcdA

A 986 bp region comprising of the complete *ccdAB* operon along with its upstream and downstream regulatory regions (based on F-plasmid sequence deposited in GenBank) was synthesized at GenScript (USA) and cloned in the EcoRV site of the pUC57 plasmid vector. A few mutations were introduced in this cloned *ccdAB* construct relative to F-plasmid sequence, in order to introduce restriction enzyme sites to facilitate cloning of *ccdA* or *ccdB* mutant libraries (Supplementary Figure S1). A single-site saturation mutagenesis (SSM) library of *ccdA* in this operonic construct was then constructed. Mutagenic primers for all 71 positions (2 - 72) of CcdA were designed such that the degenerate codon (NNK) was at the 5' end of each 21 bp forward primer. A non- overlapping adjacent 21 bp reverse primer, along with the forward primer, was used to amplify the entire pUC57-*ccdAB* plasmid by inverse PCR methodology using Phusion DNA polymerase [39]. The primers were obtained in 96-well format from the PAN Oligo facility at Stanford University. A master-mix was made for performing the PCR reactions for all positions in a 96 well format. Following quantification, an equal amount of PCR product (~200ng) at each position was added to get a pool of PCR products of all required fragments.

Gel-band purification of the pooled PCR product at the required size (~3.6 Kb) was done using a Fermentas GeneJET Gel Extraction Kit. After purification, pooled PCR product was phosphorylated, followed by ligation. The ligated product was transformed into high efficiency (10^9 CFU/µg of pUC57 plasmid DNA) electro-competent *E. coli* Top10Gyr cells. 5µg of the total ligated PCR product was transformed into 2mL of competent cells (4 aliquots of 500µl each) and the cells were plated on LB agar plates containing 100µg/mL ampicillin, for selection of transformants. Plates were incubated for 18-20 hrs at 37˚C. Around 2000-3000 colonies obtained on each plate were washed off from a total of 8 LB agar plates (150mm diameter) into a total volume of 100 mL of LB media. Pooled plasmid was purified using a Qiagen plasmid maxiprepkit.

Construction of Top10 Gyr strain

The *E. coli* Top10 Gyr strain was constructed using homologous recombination-based methodology [40]. Briefly, a GyraseA fragment of DNA Gyrase enzyme having the mutation R462C was amplified from the parent strain *E. coli* CSH501. To enable insertion of this fragment at a particular locus in the *E. coli* Top10 genome, the fragment had homology at the 5` end with the Gyrase sequence upstream of the mutation and at the 3` end with *kan* gene from pKD4 plasmid. The Kan cassette was added for selection of recombinants. The *kan* gene containing FRT sites, was amplified such that it has homology at the 3' end with Gyrase sequence, downstream of the R462C mutation. This Gyrase-kan cassette was transformed in *E. coli* Top10 strain by electroporation and transformants growing on Kanamycin plates were selected. The mutation was confirmed by DNA sequencing. The *kan* gene was later removed by transforming the strain with plasmid pCP20 expressing the FRT-FLP recombinase, as described previously [40].

Sample preparation for deep sequencing and data processing

The purified plasmid library was simultaneously transformed into high efficiency electro-competent cells of Top10 (sensitive) and Top10 Gyr (resistant) strains, having equal efficiencies of 10^9 CFU/µg of pUC57 plasmid DNA. Approximately equal number of transformants (~10^8 in number) were scraped into 10 mL LB media, and directly used for plasmid purification using a Thermo Scientific Plasmid purification kit. These transformations and subsequent plasmid isolation procedures were performed in three biological replicates. The CcdA gene was PCR amplified from each purified plasmid sample, with primers (Sigma-Aldrich) containing condition specific, six-base long Multiplex Identifier (MID) tags. The resulting 270 bp long PCR products containing the full *ccdA* gene, were pooled, gel-band purified, and sequenced using Illumina Sequencing, on the NovaSeq6000 Platform, at Macrogen, South Korea. The sequencing was performed in a paired-end manner. The overall quality of the sequencing data was assessed using the FASTQC software. Further analysis was performed using our in-house Perl scripts. Reads were separated into 'bins' based on their MID tags. The downstream primer sequence was used to identify forward and reverse reads in each bin. A Phred score cutoff of 20 and a minimum read length cutoff of 75 were used to filter low quality reads. Reads were converted into FASTA format and aligned with the *ccdA* gene sequence using the WATER program of the EMBOSS package [41], [42] Default values were used for all the parameters except the Gap Opening Penalty, which was increased to 20. Forward and reverse reads of the same read pair were merged together if the overlapping regions were identical. Only those reads which cover the entire *ccdA* gene length were considered for further analysis. Reads with insertions, deletions and multiple mutations were omitted, and only mapped single mutants were used for further data analyses.

Assignment of mutant phenotypes based on the deep sequencing data

Read numbers for mutants at all 71 positions (2-72) in CcdA were analyzed. Mutants with <20 reads in the resistant strain were not considered for analysis. The total number of reads in different conditions was calculated. Read numbers for each mutant at a given condition were normalized to the total number of reads in that condition. Mutational fitness was defined as the ratio of the normalized reads in the sensitive strain to the normalized reads in the resistant strain, for a given mutant. Absolute fitness score for WT was 1.6, 1.6 and 1.4 in replicates 1, 2 and 3 respectively. We further normalized the mutants' fitness scores with respect to the WT fitness score to obtain the Relative fitness score.

Absolute fitness score of a mutant i, ABS_{seq}^{i} = (# of read of mutant i in Top10 / total # of reads in Top10)/ (# of read of mutant i in Top10 Gyr / total # of reads in Top10Gyr) Equation 1

Relative fitness score of a mutant i, $RFS_{seq}^{i} = ABS_{seq}^{i} / ABS_{seq}^{wt}$ Equation 2

$RFS_{seq}$is the fitness of a mutant relative to the WT construct. The RFS_{seq} scores were calculated for all mutants in the library that had >20 reads in the resistant strain, separately for each replicate. RFS_{seq} values used in the work are the average of the three replicates.

For each residue in the CcdA protein, reads for all 32 possible mutant codons (NNK codons) were analyzed in the sensitive vs. resistant strain, and phenotypes were assigned in terms of the RFS_{seq} scores. Positions showing codon specific mutation effects were analyzed separately.

***In vivo* activity of individual single-site CcdA mutants**

Selected single-site synonymous mutants of CcdA (in pUC57-*ccdAB* construct) were designed and synthesized from GenScript and sequence confirmed (see Table 1). 40 ng of each plasmid

was individually transformed into the sensitive and the resistant strains and plated on LB-amp agar plates in ten-fold serial dilutions and grown at 37˚C for 16 hrs. The number of CFUs for the transformants in Top10 strain obtained in different dilution plates was counted to confirm their *in vivo* activity. For easy readout, serially diluted transformation mix samples were also spotted using 3µL volumes per spot to visualize differences among mutant phenotypes.

Monitoring the relative levels of CcdA and CcdB using a quantitative proteomics approach

1. ***Selection of proteolytic peptides from CcdA and CcdB for MRM experiments:***

In order to identify proteolytic peptides, *E. coli* lysates were resolved on SDS-PAGE, and gel bands corresponding to CcdA and CcdB were excised and subjected to proteolytic digestion using trypsin. Peptide digests were analyzed on a Thermo Orbitrap Fusion™ Tribrid™ Mass Spectrometer interfaced with Easy-nLC 1000 nanoflow liquid chromatography system (Thermo Scientific, Bremen, Germany). Peptides were loaded onto a trap column (75 µm x 2 cm, Magic C18AQ, 5 µm, 100 Å, Michrom Biosciences Inc., Auburn, CA) using 0.1% formic acid at a flow rate of 3 µl/min. The peptides were then resolved on an analytical column (75 µm x 20 cm, Magic C18AQ, 3 µm, 100 Å, Michrom Biosciences Inc, Auburn, CA) at a flow rate of 350 nl/min using a linear gradient of 5-35% B (0.1% formic acid in 95% acetonitrile) over 18 min and a run time of 30 min. The MS and MS/MS scans were acquired at a mass resolution of 120,000 and 30,000 at 200 m/z, respectively. Full MS scans were acquired in the m/z range of 350-1550. Precursor ions with single charge or unassigned charge were rejected. Dynamic exclusion of fragmented precursor ions was set to 30 sec. Precursor ions were isolated using a quadrupole mass filter with isolation width of 2 m/z. Higher energy collision dissociation was used as the fragmentation method with 32% normalized collision energy. Proteome discoverer

(version 1.4) software suit (Thermo Fisher Scientific, Bremen Germany) was used to analyze the data. Three peptides were selected for CcdA and two for CcdB. Peptides ITVTVDSDSYQLLK and LFVDVQSDIIDTPGR was used for quantitation of protein CcdA and CcdB, respectively while others were used for identification purposes only.

2. Mass spectrometry protein sample preparation:

To quantify the levels of CcdA and CcdB proteins in the *E.coli* lysate, 10 ml cultures of *E. coli* Top10Gyr cells transformed individually with different CcdA mutants were grown till saturation (OD_{600} of 1.0) under shaking conditions at 37˚C, 180rpm. Cells were pelleted, washed twice with PBS, and lysed by sonicating in lysis buffer (2% SDS in 100mM triethyl ammonium bicarbonate (TEABC) buffer, pH 8.5). The lysates were clarified by centrifugation at 12,000 g for 30 min at 4°C. Proteins were precipitated with ice-cold acetone to remove SDS and re-constituted in 100mM TEABC. Bicinchoninic acid (BCA) assay kit (Thermo-Scientific) was used to estimate protein concentration. An equal amount of protein from each sample was reduced with 10mM DTT at 60°C for half an hour. The samples were cooled to room temperature and proteins were alkylated by adding iodoacetamide at a final concentration of 20mM and incubating at room temperature in the dark for 15 min. Proteins were subjected to proteolytic digestion overnight by adding sequencing grade trypsin (Promega) at an enzyme: substrate ratio of 1:20. For LC-MS-MRM experiments, samples were spiked with stable isotope labeled peptides (SpikeTides^{TM} TQL peptides were purchased from JPT Peptide Technologies, Germany) before adding trypsin. After overnight incubation, the reaction was quenched with 10% formic acid and centrifuged at 12,000 g to remove any precipitate. Sample clean-up was performed using reverse phase spin columns (Harvard apparatus) according to the manufacturer's protocol.

3. SRM/MRM analysis:

LC-MS/MRM analysis was performed with an electrosprayionization-triple quadrupole mass spectrometer (QTRAP 6500; SCIEX) coupled with an HPLC system (Agilent 1260 systems; Agilent Technologies, Santa Clara, CA). The peptides were resolved on a reverse phase C18 column (Agilent Poroshell 120 EC-C18 2.7µm, 4.6x50mm) at a flow rate of 350 µl/min and a run time of 10 min. The LC method used was as follows: 10% - 30% solvent B from 0-3 min, 30%-50% solvent B from 3-5 min, 50%-95% solvent B from 5-7 min and 95%-5% solvent B from 7-10 min. The data was acquired in SRM/MRM mode using a dwell time of 25 msec. Each target peptide and corresponding SIL peptide were measured. Standard curves were generated by spiking different amounts of SIL peptides (0.5, 0.1, 0.25. 0.5, 1, 5, 10, 20 pmol) into a constant amount of *E. coli* digest background.

4. LC-MS/MRM data analysis:

MRM data was analysed using Skyline 3.5 [43] and composite peak AUCs were calculated after Savitsky-Golay smoothening. Each transition was inspected and XICs for each transition was generated. Response curves were generated by plotting the ratio of heavy: light peptide over different concentrations of spiked heavy peptide. Linear regression analysis was carried out to determine the linearity of the response curve. Also, to validate the analytical method, regression analysis of variance (ANOVA) of the linear regression data measurements (at a 95% confidence level)was performed. In addition, to illustrate the reliability of LC-MS/MRM analysis, the %RSD was calculated, which was within 20% in all cases. Peaks corresponding to endogenous target peptides were identified based on the retention time of the corresponding SIL peptide as it

co-elutes with the endogenous peptide. Welch's test was used to determine the statistical significance of abundance differences between the wild type and mutants.

Estimation of *ccd* specific mRNA levels using qRT-PCR

For RNA isolation and quantification, 3 ml of culture was grown for each of the mutants of *ccdA* transformed in *E. coli* Top10 Gyr strain. Cells were grown to saturation under shaking conditions at 37˚C, 180rpm, pelleted and total RNA was extracted by the RNAsnap method [44]. Chromosomal and plasmidic DNA was removed by treatment with 2 units of DNase1 (New England Biolabs) for 2 hours at 37˚C, followed by Sodium Acetate-ethanol precipitation of RNA. RNA was quantified by nanodrop spectrophotometric estimation, and quality was assessed by agarose gel electrophoresis, prior to downstream processing. 2µg of total RNA was taken in a sterile, RNase-free microcentrifuge tube and 0.5µg of random hexamers (Sigma-Aldrich) was added to serve as primer for cDNA synthesis along with DEPC (Sigma-Aldrich) treated double autoclaved MilliQ water. The mixture was heated to 60°C for 10 minutes to melt secondary structure within the template. The tube was cooled immediately on ice to prevent secondary structure from reforming. To this mix, 200 units of SuperScript III Reverse Transcriptase (Invitrogen) were added, along with dNTPs, RT Reaction Buffer and 25 units of Ribonuclease Inhibitor following the manufacture's protocol. The reaction mix was incubated for 60 minutes at 37°C. cDNA was directly used for quantitative PCR (Q-PCR) analysis. Q-PCR was set up with *ccdA* as well as *ccdB* gene specific primers (250nM) using the cDNA template and Bio-Rad iQ SYBR Green Supermix (1x) in a Bio-Rad iQ5 machine. A reaction with no reverse transcriptase was used as a negative control. 16S rRNA was used as the internal positive control for total RNA quantification.

The sequences for the forward and reverse *ccdA* specific primers were

5’GAAGCAGCGTATTACAGTGACAGTTG3’ (binding 2-28 bp region of *ccdA*) and 5’CGTCAGCAAAAGAGCCGTTCA3’ (binding 181-202 bp region of *ccdA*) respectively.

The sequences for the forward and reverse *ccdB* specific primers were

5’GAGAGAGCCGTTATCGTCTGTTTGTG3’ (binding 28-54 bp region of *ccdB*) and

5’GATAACGGAGACCGGCACAC3’ (binding 208-228 bp region of *ccdB*) respectively.

Both the gene specific amplicons for the q-PCR reaction were 200bp.

The sequences for the forward and reverse 16S rRNA specific primers were

5’ CCTAACACATGCAAGTCGAACGG 3’ and

5’ CTCCATCAGGCAGTTTCCCAGAC 3’ respectively.

The fold change in RNA levels for different mutants was calculated by following the formulae,

Fold change = $2^{-\Delta\Delta Ct}$ Equation 3

where, $\Delta\Delta Ct = \Delta Ct_{mut} - \Delta Ct_{WT}$

and $\Delta Ct = Ct_{exp} - Ct_{control}$, (“exp” refers to gene of interest i.e. *ccdA* or *ccdB* and “control” refers to 16S rRNA). Ct stands for threshold cycle of amplification in PCR reaction and is automatically provided by the Bio-Rad iQ5 machine.

Computational prediction of mRNA secondary structure

The initial 150 bp of the *ccdAB* transcript were submitted to mFOLD [45] and RNAfold [46] for prediction of secondary structure. All energy parameters were set at default values. Detailed output was obtained in the form of structure plots with reliability information, single strand frequency plots, and energy dot plots. Local mRNA secondary structural elements encompassing the region of synonymous inactive mutations were further analyzed.

Calculation of parameters describing various DNA sequence features

The codon usage information was taken from the *E. coli* codon usage frequency table at the GenScript website (https://www.genscript.com/tools/codon-frequency-table). The codon usage frequency fraction (CUF) was used to calculate the relative codon usage for mutants, using the following formula:

relative codon usage = (CUF of mutant codon/ CUF of most common codon of mutant amino acid) / (CUF of WT codon/ CUF of most common codon of WT amino acid). Equation 4

The tRNA abundance raw data [47] depicts the abundance of tRNA molecules in an *E. coli* cell that decode one or more codons. We assigned the tRNA abundance values to the respective codons that they decode, and calculated the relative tRNA abundance values associated with each codon using a similar normalization as described above for calculation of relative codon usage.

Relative tRNA abundance = tRNA abundance of mutated codon/ tRNA abundance of WT codon Equation 5

The ribosome stalling data [48] describes the observed averaged pausing of ribosome on each codon at the A-site for all genes in *E. coli*. The ribosomal pausing was estimated by introducing each codon individually into a 6X-His leader sequence of the histidine operonic system and quantified by downstream HisD-LacZ fusion protein (β-lactamase) activity. For the current work, the relative ribosome stalling for each mutation was calculated by dividing the mutant codon pause value by the WT codon pause value.

The relative GC content for mutations was calculated by dividing the fractional GC content of

mutant codon by that of WT codon. The number of base changes is the number of positions in a 3 bp codon where the mutant codon varies from the WT codon.

To investigate if there is any significant difference in the distribution of these sequence parameters across the inactive and active classes of mutants, we chose to use the non-parametric Mann-Whitney test since our data is ordinal and does not follow a normal distribution.

Calculation of anti-Shine Dalgarno interaction energies for synonymous mutants in *ccd* mRNA

The difference in the interaction energy between single synonymous mutants in *ccd*A and the WT sequence, with the consensus anti-Shine Dalgarno (aSD) sequence (5' CCUCCUAU 3'), was calculated for a window of eight nucleotides using the RNAsubopt program from the RNAVienna package 2.4.3 [46]. The calculations were performed for ten such windows encompassing the mutant codon. The average energy difference for each mutant across these ten windows was compared for all the available synonymous active and inactive mutants. Such aSD like sequences are known ribosomal pause sites [25].

The binding energies for different mutants at the ribosome binding site of CcdB in the *ccdA* gene (residue positions 70 and 71) to the anti-Shine Dalgarno (aSD) sequence (5' CCUCCUAU 3') was also calculated similarly, to look for possible *ccdA* mutants (at residue positions 70 and 71) that can alter the translation initiation of CcdB.

Yeast surface display of CcdA library and single mutants to probe CcdB binding

The CcdA libraries, individually recovered both from the resistant and the sensitive strains, were amplified from the pUC57-*ccdAB* plasmid vector and recombined into the yeast surface

display vector, pETcon (Addgene plasmid # 41522) between NdeI and XhoI sites, using yeast homologous recombination in *S. cerevisiae* EBY100 strain. Positive clones, sufficient in number to cover the library diversity, were obtained on selective SDCA plates. The transformant pools obtained were grown in liquid SDCAA media for 48 hrs and stored in aliquots of 10^9 cells per ml in SDCA media containing 20% glycerol at -70˚C. The same methodology was used for cloning WT and individual synonymous mutants of CcdA into pETcon. The CcdA coding region was cloned as a C-terminal fusion to the yeast Aga2p protein with a C-terminal c-myc tag under control of the GAL promoter. The library was displayed on the yeast surface as described [49]. Cells were induced and grown in SGCA media containing 2% galactose at 30˚C for 16 hrs. The surface expression of CcdA was monitored by incubation of 10^6 cells with anti-c-myc antibody raised in chicken (1:400 dilution), that binds to the Myc tag at the C-terminus of CcdA. Anti-chicken IgG conjugated AlexaFluor-488 (1:300) was used as the secondary antibody. CcdB binding to the surface expressed CcdA was assessed by incubating cells with a fixed concentration of 2nM biotinylated CcdB (or varying concentrations between 0.1pM and 200nM for titration experiments) and binding of streptavidin conjugated AlexaFluor-633 secondary antibody (1:2000) to biotinylated CcdB was monitored. All the primary and secondary antibodies were obtained from Invitrogen. Labelled cells were analyzed on a BD FACSARIA III. The titration curves were fit to a single ligand binding model to obtain the dissociation constants as described previously [50], [51]. Double plots for mean fluorescence intensities for both expression and binding for the CcdA library as well as for single synonymous mutants were analyzed to monitor differences with respect to the WT.

Monitoring the effect of single synonymous mutations in the RelBE operon

Design of mutants: The WT sequence for the RelBE operon from *E.coli* K12MG1655 strain was retrieved from the NCBI database. The entire transcription unit of 910 bp [52] including another gene *relF*, which is expressed as a part of the same operon, was cloned in the EcoRV site of the pUC57 vector at GenScript. Five different constructs having synonymous mutations in the relB gene were synthesized with a stop codon in the *relE* toxin gene to ensure normal cell growth even after inactivation of the antitoxin, for ease of cloning. Three base synonymous substitutions are only possible with serine codons. Three base synonymous substitutions with varying codon preferences were made in three of the serine codons, S3, S16 and S28, all at the N-terminal half of the *relB* coding sequence. One representative two base substitution at position R7 was also tested. A single base substitution at the D77 residue was made to increase the strength of the SD sequence for *relE* and was used as a positive control mutant. To obtain an easy phenotypic readout, the introduced stop codon in the *relE* gene was reverted, using the inverse PCR strategy described earlier [39].

Phenotype testing: Due to absence of a strain that is resistant to toxin action, we employed a toxin deletion strain (*E. coli* BW25113 Δ*relE*) for the propagation of mutants. By virtue of this deletion, the WT chromosomal operon is de-repressed, leading to an excess of antitoxin in the cell, relative to the WT strain (*E. coli* BW25113 *relBE*). The inverse PCR products for individual mutants were phosphorylated and ligated, as described previously. Ligation mixes were individually transformed into the *E. coli* BW25113 ΔrelE strain, to recover the plasmid. Single clones were isolated, and sequence confirmed. Equal amounts of the mutant plasmids (~30 ng) were transformed in *E. coli* BW25113 WT strain (having chromosomal *relBE* genes intact) to test their phenotype, transformation mixes were plated on LB-amp agar plates in ten-fold serial dilutions and grown at 37˚C for 16 hrs. The number of CFUs for the transformants in

E. coli BW25113 WT strain obtained in different dilution plates was counted. Serially diluted transformation mix samples were also spotted using 2μL volumes per spot to visualize differences among mutant phenotypes.

Results

Unexpectedly high mutational sensitivity in the CcdA antitoxin gene, in its native operonic context

The *ccd* operon comprises of the *ccdA* and *ccdB* genes encoding a labile antitoxin and a DNA-gyrase poisoning toxin respectively, under the control of a single auto-regulated promoter [53], [54]. The 72 residue, labile, disordered CcdA antitoxin protein binds the CcdB toxin tightly, preventing CcdB toxin interaction with its cellular target, bacterial DNA-gyrase [55]. Mutations in CcdA that impair its antitoxic function are thus expected to promote cell death (Figure 1A). To understand the consequences of mutations in CcdA on bacterial fitness, the *ccd* operon consisting of the promoter, *ccdA* and *ccdB* genes (Figure 1B) was cloned into the pUC57 plasmid. Restriction digestion sites introduced in the construct for ease in cloning resulted in a few changes with respect to the WT sequence of the *ccdAB* operon, at certain positions in the non-coding region upstream of the CcdA gene (Supplementary Figure S1). A site-saturation mutagenesis (SSM) library of CcdA was created in this operonic construct using NNK primers (Supplementary Figure S2). This library comprising of both synonymous and non-synonymous mutants of CcdA was subjected to deep sequencing following growth under selection and control conditions, in sensitive and resistant (having a mutated DNA-gyrase, defective in CcdB binding) strains of *E. coli* respectively. The relative fitness score of each CcdA mutant was computed

from the available sequencing reads in selective (sensitive strain) versus control (resistant strain) growth conditions relative to WT (Supplementary Figure S2 and Methods). The RFS_{seq} score is thus defined as the relative ratio of fraction of mutant reads in sensitive to resistant strains, normalized to WT reads. The deep sequencing was performed in three biological replicates, showing a high correlation (Pearson Correlation Coefficient = 0.98 approximately, Supplementary Figure S3) amongst themselves. The relative fitness score (RFS_{seq}) was calculated and assigned only to those mutants having a minimum of twenty reads in the resistant strain, for each replicate. The fitness scores averaged over three replicates were used to classify mutants based on their activity. The WT CcdA has an RFS_{seq} value of 1. Of 2272 possible NNK single codon CcdA mutations (71 positions * 32 codons), fitness values could be assigned to 1558 mutants. A drastic reduction in cellular growth in the sensitive strain indicated a substantial prevalence of mutants exhibiting an inactive phenotype in the library (Figure 1C and Supplementary Figure S4). CcdA mutants with lower relative fitness are expected to have decreased antitoxic activity, thus failing to rescue CcdB mediated cell death. The low incidence of active mutations demonstrates the extreme sensitivity of the *ccdAB* operonic construct to mutations in *ccdA*. We classified mutants having RFS_{seq} values $\leq$ 0.1 as inactive. Surprisingly, ~80% of all CcdA mutants display an inactive phenotype with RFS_{seq} values $\leq$ 0.1 (Figure 1C-D). A phenotypic assay was conducted for a subset of CcdA mutants by individual transformation in resistant and sensitive strains, and dilution plating of the single mutants. This showed good agreement with the deep sequencing derived mutational fitness scores (Supplementary Figure S4, Table 1). Phenotypic assays of individual single mutants with RFS_{seq} values less than 0.1, confirmed that they show significant growth defects in the sensitive strain (Supplementary Figure S4).

Inactive synonymous mutants and codon specific effects in *ccdAB* operonic construct

The majority of synonymous mutants (synonymous with respect to WT sequence, with the same amino acid sequence encoded by an alternate codon in DNA and mRNA) show severe growth defects (Figure 2A and Table 2). The growth defects are inferred in terms of the number of colonies formed (and thus deep sequencing reads) by each mutant on plates following transformation in sensitive TOP1O *E.coli* cells and overnight growth at 37°C (see Methods). These synonymous inactive mutants are dispersed throughout the length of the CcdA protein, with mutations at only a few specific positions at the N- and C-terminal regions showing RFS_{seq} scores greater than 1. We could confidently assign RFS_{seq} scores to only 62 synonymous mutants having adequate representation in the library in terms of the numbers of reads. Since we used NNK (K= G/T) primers during mutagenesis and library preparation steps, we only have 50% of all 64 codons at each position represented in the library. These 62 variants bear single codon mutations at 36 positions out of the 72 residues in CcdA. 49 of these 62 available synonymous mutants (~79%), have RFS_{seq} scores less than 0.1 and can therefore be classified as highly inactive. The central stretch of CcdA (residues 20-55) was especially sensitive to synonymous mutations, with all documented synonymous mutations in this region being highly inactive. Synonymous mutations at the N terminal and C- terminal regions of CcdA resulted in both active and inactive variants. We also observed that most synonymous active mutants were hyperactive in comparison to WT with RFS_{seq}> 1.5. Such a wide distribution of activity in single synonymous mutants is unprecedented.

The mutational effects observed in case of synonymous mutations led us to investigate codon specific effects amongst non-synonymous CcdA mutants. We found several non-synonymous mutations at multiple positions having distinct phenotypes for different codons that code for the

same mutant amino acid (Figure 2B-C). Since NNK codons were used for mutagenesis and preparation of the SSM library, we could only investigate phenotypes of mutated codons having G/T at the 3rd nucleotide position. Therefore, the phenotypic differences between codons could be investigated for a subset of mutated amino acid residues that are encoded by more than one codon, namely Alanine, Glycine, Leucine, Arginine, Proline, Threonine, Serine and Valine. We also found that phenotypic differences amongst codons in non-synonymous mutants were heightened in the N- and C- terminal positions of CcdA, while all mutations in the CcdA central stretch were inactive irrespective of the identity of the mutated codons as well as amino acids (Figure 2B-C). These results indicate the observed effects are not due to introduction of a mutated amino acid residue, but due to changes in the DNA sequences. The CcdA library under study comprises of a very small number of active mutants, which are both synonymous and non-synonymous and most are substantially more active in comparison to the WT construct (Figure 2). These active mutants were populated at the N and C-terminal residues, where mutations lead to the most diverse range of activities (Figure 2).

Growth defects caused by *ccdA* mutations are not due to CcdB binding defects

The growth defects caused by *ccdA* mutations are only observed in the Top 10 strain sensitive to CcdB toxicity and not in the Top10 Gyr strain, where the mutated Gyrase is resistant to CcdB mediated growth arrest. This clearly indicates that the growth defect displayed by *ccdA* mutations is caused by decreased inhibition of CcdB toxin in the sensitive strain. CcdA is known to rescue growth arrest by binding to CcdB with very high affinity, thus preventing CcdB from binding and poisoning DNA gyrase [56], [57]. We therefore characterized the ability of CcdA mutants (displaying a loss of function phenotype) to bind CcdB, using yeast surface display (YSD), where CcdA-CcdB binding can be decoupled from toxic effect of CcdB on bacterial

growth. CcdA was fused to *S. cerevisiae* Aga2p surface protein, displayed on the surface and its binding to purified CcdB toxin protein was measured using FACS (Figure 3). Although CcdA shows high mutational sensitivity throughout its length based on the phenotypic study in the operonic construct in *E. coli* (Figure 3A), the yeast surface display (YSD) study revealed only minor differences in CcdB binding between the two libraries procured from sensitive and resistant *E. coli* strains respectively (Figure 3B-C). The YSD CcdA libraries were constructed from plasmid DNA isolated from the sensitive and resistant *E. coli* strains by cloning in YSD, pETcon vector (Figure 3B). Moreover, studies of individual mutants using YSD reveal that all CcdA synonymous mutants and WT protein bound equivalently to CcdB (Figure 3D), irrespective of the corresponding distinct mutant phenotypes in the native operonic context in *E. coli*. Synonymous mutations have previously been suggested to affect translation rate and co-translational folding, thus altering protein activity [12], [13], [58]. Such folding defects are unlikely in the case of CcdA, since its C-terminal domain, involved in CcdB binding is intrinsically disordered and remains natively unfolded in unbound conditions. Further, we find no discernable difference in the distribution of phenotypes or the RFS_{seq} values between mutations at CcdB interacting and CcdB non-interacting residues in the C-terminal CcdB binding domain of CcdA, classified based on the available structure of CcdA C-terminal domain bound to dimeric CcdB [57] (Supplementary Figure S5A). These observations indicate that mutational effects on binding affinity to CcdB contribute insignificantly to the mutant phenotypes observed for *ccdA* in its native operonic context.

We also studied the distribution of mutational phenotypes (for non-synonymous mutations) at different classes of functionally important residues in the N-terminal domain of CcdA, identified based on the available NMR structure of the DNA bound dimeric CcdA N terminal domain [59].

We found that mutations in the DNA binding residues in the CcdA N-terminal domain have no distinguishable difference in terms of the growth phenotype, when compared to other non-interacting exposed residues (Supplementary Figure S5B). Surprisingly, mutations at buried positions in the N terminal domain show significantly reduced activity and more severe growth defects, especially in the case of mutations to non-aliphatic residues, when compared to any other classes of CcdA residues (Supplementary Figure S5C). This suggests that mutations that can disrupt the N-terminal buried core of CcdA can induce increased protein degradation or proteolysis, ultimately producing the observed inactive phenotypes.

E. coli **genome codon usage correlates well with growth phenotypes of single codon mutants in CcdA**

The frequency of different degenerate codons (for any amino acid) in the genome of an organism varies significantly. Codons with higher frequency in the genome are referred to as optimal codons. We investigated the relationship between the phenotypes of mutants and relative codon usage (in the *E.coli* genome) of the introduced codon. We find a weak positive correlation between the relative fitness of CcdA mutants and corresponding relative codon usage values. The distribution of relative codon usage values amongst active and inactive CcdA mutants was significantly different (Figure 4A). It has been suggested that having a higher percentage of optimal codons in a gene leads to higher amounts of expressed protein, across various kingdoms of life [21], [22], [60], [61]. Decreased levels of CcdA protein *in vivo* will result in the presence of unbound CcdB molecules that will poison bacterial DNA Gyrase, resulting in cell death. While there is a large fraction of inactive mutants in the *ccdA* single mutant library, we observe that ~70% of *ccdA* mutants displaying an active phenotype harbor optimal codons with similar or higher codon usage frequency relative to the WT codon. The observed high mutational

sensitivity in the central region of the *ccdA* gene might result from the high prevalence of optimal codons in this region (Supplementary Figure S6A). Thus, most mutations at these positions lead to the introduction of rarer codons, possibly leading to lower CcdA protein expression and subsequent cell death. RFS_{seq} values of mutants were weakly correlated with the relative tRNA abundance for the mutated codon, relative to WT. The distribution of relative tRNA abundance values amongst active and inactive mutants was similar (Figure 4B). There was no correlation of the RFS_{seq} values with other sequence features such as relative GC content and number(s) of base changes upon mutation. (Figure 4C-D).

We also studied the distribution of these DNA sequence features among the active and inactive variants for the smaller subsets of synonymous mutations, as well as for the non-synonymous substitutions which showed codon specific effects. Inactive synonymous mutations show significantly lower relative codon usage and tRNA abundance values than active synonymous mutants (Figure 5A-B). There are also statistically significant differences in the distributions of the relative codon usage values, but not tRNA abundance between active and inactive classes of non-synonymous mutations that show codon specific effects (Figure 5C-D).

Strength of CcdB ribosome binding site correlates with CcdA mutant phenotype

A primary sequence element that is known to be a strong determinant of the efficiency of translational coupling is the Shine Dalgarno sequence. Early studies by Das and Yanofsky in the Trp operon showed that translation initiation at any start site located near a functional stop codon may be influenced by the strength, sequence and location of the SD region and its spacing from the start codon [62]. The Ribosome Binding Site (RBS) for the *ccdB* gene lies within the 72 amino acid CcdA coding region (*ccdA* gene), primarily between the 70th and 71st amino acids.

Therefore, mutations in these residues may affect the relative CcdA and CcdB levels by changing the strength of the RBS for CcdB.

An analysis of all available codon substitutions at the residues 70 and 71 of CcdA from the deep sequencing data indicated a strong correlation between the predicted strength of the RBS for CcdB expression and the depletion ratio of these mutants (Figure 6). The molecular basis of one such mutation, D71R was probed. Since two of the arginine codon mutants had an active phenotype, it is unlikely that the inactive phenotype observed for the AGG arginine codon is due to effects on protein stability or protein-protein interaction. It is known that the consensus sequence for the RBS, having the maximal strength is 5' AGGAGG 3' (based on the corresponding complementary sequence of 16S rRNA, 5'CCUCCU 3'). Therefore, it is likely that a change from the suboptimal RBS for CcdB present in the *ccdAB* operon (5'AGGGAC 3'), to a consensus RBS (as is predicted in case of D71R_AGG mutation), increases the CcdB: CcdA ratio in the cell, thereby leading to cell death.

Mutations in CcdA lead to altered levels of CcdA proteins and therefore modified CcdA:CcdB ratio in cell

Codons for inactive CcdA mutants have significantly lower relative codon usage in comparison to active mutants, suggesting that the observed mutational phenotypes might result from alterations in translational rate of the CcdA or CcdB gene products. To further probe protein levels in mutants, we carried out a proteomics study. The extremely low levels of CcdA and CcdB proteins *in vivo* cannot be accurately characterized using classical methods like SDS-PAGE or Western Blotting. The ratio of CcdA: CcdB in WT was found to be slightly larger than 1 in the proteomics study (Figure 7A). Interestingly the inactive CcdA mutants show a lower CcdA:CcdB ratio, while hyperactive mutants show a higher CcdA:CcdB protein ratio in

comparison to WT (Figure 7A). The introduction of suboptimal codons thus appears to reduce the translation of CcdA, decreasing the CcdA:CcdB ratio of proteins in the cell, causing more severe growth arrest. The inactive D71R_AGG mutant (CcdB RBS site mutation in *ccdA*) also exhibits a lower CcdA:CcdB ratio, likely due to increased translation of CcdB protein.

Increased levels of operonic mRNA in inactive mutants confirm decrease in CcdA:CcdB ratio

CcdA inactive and hyperactive mutants have relatively reduced and elevated levels of CcdA proteins relative to WT respectively. The *ccdAB* expression at the transcription level is autoregulated by the CcdA-CcdB protein complex in a CcdA:CcdB ratio dependent way [34], [53]. When CcdA:CcdB protein level $\geq$ l, elongated multimeric CcdA-CcdB complexes with alternating dimeric $CcdA_2$ and $CcdB_2$ are formed which bind to the *ccdAB* operator region with high affinity, repressing gene expression. On the other hand, when CcdA:CcdB < 1 (protein level), hetero-hexameric complexes of one CcdA dimer bound to two CcdB dimers ($CcdB_2$-$CcdA_2$-$CcdB_2$) are formed that have low binding affinity to the operator region and leads to a derepressed state of operon, allowing transcription to proceed [63].

To study the possible effects of CcdA sequence changes resulting from single codon mutations on gene expression, we quantified *ccdAB* operon specific RNA using RT-qPCR using both *ccdA* and *ccdB* gene specific primers. Relative amounts of *ccdA* and *ccdB* encoding transcripts were near identical in WT and were not significantly changed in case of most mutations (ratio of *ccdA* and *ccdB* specific mRNA levels ≈ 1) (Figure 7B), as expected for transcription from an operon. However, significant increases in the mRNA levels for both the *ccdA* and *ccdB* amplicons were observed in case of all inactive mutants (Figure 7B). This is consistent with the lowered CcdA:CcdB protein levels in the inactive mutants observed in the previous section. A lowered

CcdA:CcdB protein ratio will result in de-repression of the operon, leading to upregulation in mRNA production (Figure 7C). Similarly, the WT like mRNA level observed for the hyperactive mutant, R57A can be explained by an increased CcdA protein level, caused by mutation from a rare to a more optimal codon that maintains the repressed state of the operon. Analysis of whole genome RNA-seq and ribosome profiling data available for the *E. coli* genome indeed indicate that while the mRNA levels for most type II TA systems were similar for the toxin and the antitoxin consistent with our observations in *ccdAB*, the antitoxin protein levels can be upto two fold higher than the toxin under normal growth conditions [27]. This indicates that differential translation of the *ccdA* and *ccdB* genes and/or differential proteolysis of the two proteins play a vital role in maintaining the toxin-antitoxin protein levels *in vivo* (Figure 7D). Altered DNA sequences affecting the relative levels of the toxin-antitoxin proteins can thus be detrimental to functions of TA systems.

Investigation of possible effects of ribosomal pausing on CcdA mutant phenotype

An analysis of bacterial genomes has suggested that Shine Dalgarno (SD) like sequences within the coding region of genes could be potential ribosomal pause sites and are therefore under negative selection [25], [64]. To examine if mutations in CcdA generate such potential pause sites, the difference in interaction energies with the *E. coli* ribosomal anti-SD sequence between single synonymous mutants and the WT sequence was calculated for a window of 10 nucleotides using the RNAsubopt program in the Vienna RNA package [46]. The data indicate that the inactive and active synonymous mutations show no distinguishable difference in anti-SD interaction (Supplementary Figure S7A). The extent of ribosome pausing at each codon and the associated effects on *in vivo* translation speed have been estimated previously [48]. In the present study, we did not observe any significant differences in distribution of these relative ribosome

stalling values for different codons between synonymous inactive and active mutants of CcdA (Supplementary Figure S7B).

Effect of single codon synonymous mutations in another TA module, RelBE

To test if other TA modules are also sensitive to synonymous mutations, the RelBE system was used as a test case. Unlike the plasmidic *ccdAB* system, *relBE* is a chromosomal toxin-antitoxin system, expressed as a part of the *relBEF* operon [65]. The RelE toxin is a site-specific, ribosome dependent mRNA endonuclease. Expression of free RelE leads to bacterial growth inhibition. Mild overexpression of RelE is also reported to increase persister cell frequency [66]. While unrelated in sequence, the RelB protein from *E. coli* is similar to the CcdA antitoxin in both fold and thermodynamic properties, and inhibits the action of RelE by forming a tight complex. The RelBE complex acts as a transcriptional repressor and autoregulates its own expression [65]. Homologs of this system are found in many bacterial species including pathogens like *Haemophilus influenzae.* An analysis of ribosome profiling data for the *E. coli* K12 MG1655 strain indicates that the translation efficiency for the *relB* gene is 3.6-fold higher than the *relE* gene [27], although both the codon and tRNA adaptation indices are similar for both genes, similar to what we observed in the CcdAB system.

We studied five single codon synonymous mutations, distributed across the length of the relB gene in a native *relBE* operonic context, including both three and two base substitutions. These mutations were to both optimal and rarer codons. A control mutation at residue 77 of *relB* that strengthens the SD sequence of RelE was also studied. This is expected to lead to an increase in RelE levels and inhibition of cell growth (Figure 8). Using a similar screen of cell death versus growth as in the case of CcdAB system, we monitored

the growth of single synonymous mutants in the WT (*E. coli* BW25113) as well as in the toxin deleted (*E. coli* BW25113 Δ*relE*) strain. Most *relB* variants with rarer mutated codons had a decreased growth in the WT strain, indicating that the substitution with rarer synonymous codon can indeed result in a loss of function phenotype (Figure 8B-C). We found that two mutations of *relB* with higher codon usage with respect to WT codon display improved growth phenotype. This observed correlation of phenotype with codon usage agrees with observations in case of the *ccdAB* phenotypic study of CcdA mutants investigated using deep mutational scanning. The *relB* mutant phenotypes indicate the existence of a relatively general phenomenon wherein a change in DNA sequence upon synonymous mutations can have observable phenotypic effects for toxin-antitoxin operon systems, where small perturbations in the system can compromise cellular fitness.

Discussion

Most high-throughput mutagenesis studies coupled to deep sequencing, have shown little or no effect of synonymous mutations on protein function or phenotype, with a few exceptions [67], [68]. One limitation in such studies is the use of artificial stress conditions, such as antibiotic stress, that complicates direct evolutionary interpretation [36]. These studies also commonly employ heterologous proteins over-expressed in *E. coli* or other model organisms, limiting the ability to understand the intrinsic *in vivo* determinants of such phenotypes, and the effect of synonymous mutations in a native context on fitness. In the current work, we have attempted to look at mutational effects in the bacterial *ccdA* antitoxin gene on phenotype, in its native TA operonic context, using saturation mutagenesis coupled to deep sequencing. These studies reveal

that most single codon substitutions, including synonymous mutations along the length of *ccdA* lead to a loss of function phenotype. Many non-synonymous mutations also show codon specific phenotypes, meaning that significantly different phenotypes are observed upon introduction of different codons that encode the same mutated amino acid residue. These studies demonstrate that there can be strong selection for specific synonymous codons. This complicates interpretation and use of dN/dS ratios for studying selection pressure especially in operons. The overall correlation between the mutant phenotype and relative codon usage of mutant with respect to WT suggests that single codon mutations in CcdA likely influence the translation efficiency of the CcdA, thus altering the relative levels of CcdA and CcdB proteins in the cell. A high frequency of optimal codons in the central region of *ccdA* gene suggests that the observed high mutational sensitivity at this region is because any codon introduced upon mutation is expected to be suboptimal relative to the WT codon. At the *ccdA* gene extremities there is larger variation in codon usages values. Hence mutated codons at these regions may be more or less optimal relative to WT. Altered DNA sequence can directly affect ribosome assembly, ribosome stalling, rate of translation, or promote altered mRNA structure, which in turn can have multifaceted consequences on protein production. The quantitative proteomics study reveals that for loss of function mutants, the *in vivo* CcdA: CcdB protein level ratio is lower than that of WT, resulting in cell death. In contrast, mutants showing higher fitness than WT, have a higher CcdA: CcdB protein ratio. The prevalence of high phenotypic variations among mutants exclusively at the N-terminal (first 20 residues) and C-terminal residues (56-72) can also be a result of mutational effects on translation initiation of *ccdA* and *ccdB* gene. Mutations near the *ccdA* gene terminus, at residue positions 70 and 71 (that also constitute the ribosome binding site of CcdB) appear to affect translational efficiency of the downstream CcdB gene. Mutants at

these positions which are predicted to show improved ribosome binding exhibit a higher *in vivo* CcdB:CcdA protein ratio and higher toxicity in *E. coli*. Mutations that decrease the CcdA:CcdB ratio result in de-repression of the operon, this is confirmed by the observed increased *ccd* mRNA expression levels in these mutants. The elevated operonic mRNA level is expected to amplify the mutational effect on differential translation efficiency of CcdA and CcdB.

Earlier studies which observe a strong influence of codon rarity and codon pair bias on protein level [23], [69] typically used over-expressed proteins. In natural contexts, many bacterial genes are organized in operons to ensure co-regulation of genes involved in similar pathways and functions. In the special case of toxin-antitoxin operons, fine tuning of the relative amounts of antitoxin and toxin is essential and directly linked to bacterial fitness. Given the abundance and diversity of these TA systems in many bacterial species, including the pathogen *Mycobacterium tuberculosis*, it is reasonable to assume that the cell must have evolved multiple mechanisms to ensure their stringent spatial and temporal regulation. We find that these relative protein levels can be perturbed through changes in DNA sequences. Although effects of synonymous mutations are often attributed to co-translational folding defects and altered activity of protein, we find that most of the inactive *ccdA* mutants are not CcdB binding defective, based on yeast surface display studies. We also observed that single mutations in DNA interacting residues of the CcdA N-terminal domain did not affect growth phenotype more drastically than mutations in other residues in CcdA, indicating no significant loss of DNA binding in such mutants. In contrast however, the heightened inactivity was observed in case of mutations at buried positions in the structured N-terminal domain. This suggests probable disruption of the hydrophobic core in the N-terminal domain, leading to increased proteolysis and decreased CcdA protein level. Besides the amino acid substitution related effects observed in the N-terminal domain, we also

observe large phenotypic variation in the *ccdA* activity amongst a subset of missense mutants coded by different codons in both the N and C terminal regions of CcdA that cannot be explained by the codon usage bias or tRNA abundance. Interestingly, these effects occur in spans of about 40 bases which is on the order of the footprint of a single ribosome on mRNA in prokaryotes [70]. Thus the high diversity in phenotype for synonymous codons exclusively at the two termini of *ccdA* is likely to be caused altered translation initiation rates for CcdA and CcdB in the operon with point mutations at the N- and C-terminus positions of the *ccdA* gene respectively.

Previous studies on prokaryotic translation also suggest mRNA structure and accessibility near the translation start site play important roles in translation initiation, a rate limiting step of translation [71]–[73]. Though we find no significant correlation of the phenotype with change in GC content or number of base changes upon mutations in these regions, we suspect that local mRNA structural changes upon mutations in the 5' end of the ccdA gene affect the translation initiation of CcdA while those in the in the 3' end of the ccdA gene affect the translation initiation of CcdB, leading to the large codon-specific phenotypic variation observed at both termini of the *ccdA* gene. While the RNA structures of the *ccdAB* transcript predicted by the MFold [45] and RNAfold [46] web servers show that several single codon mutations can affect the predicted mRNA structural architecture of *ccdA* mRNA and its stability, no consistent difference is observed amongst the structures for mutants displaying neutral, inactive and hyperactive phenotypes. Since mRNA stability, transcription and translation are coupled in prokaryotes, to understand mutational effects of mRNA structure on protein levels and activity, one must determine the *in vivo* mRNA structure of the mutants which is beyond the scope of the present study.

The *ccd* operon used in the present experiments differs slightly in sequence from the one on F-plasmid (Supplementary Figure S1), resulting in a slightly reduced CcdA: CcdB protein ratio (Figure 7A). This might be due to the presence of a suboptimal *ccdA* RBS in the non-coding region (Supplementary Figure S1) in our WT CcdAB construct. The high copy number pUC57 plasmid would further enhance sensitivity to mutations in the operonic context. Overall, the present *ccdAB* system allows reporter independent monitoring of *in vivo* phenotypes, that are highly sensitive to small perturbations in the regulation of the system, which otherwise would remain undetected. The observed phenotypic effects in CcdA mutants are caused by alteration in the relative translation efficiencies of *ccdA* and *ccdB* genes in the operon. Mutant phenotypes were also found to be correlated with the relative codon usage in case of selected single synonymous variants of the *relB* gene in its RelBE TA operon. The current study therefore showcases the utility of TA operons for providing general insights into the plethora of mechanisms by which point mutations can affect phenotype. While multiple previous studies have shown effects of altered codon usage upon protein level and translation efficiency or rate, all of them exclusively use constructs with a large number of codon mutations significantly altering the codon bias of the whole gene [74]–[76]. This is the first report that describes such dramatic phenotypic effects of variants bearing single codon mutations distributed across the length of the gene.

Figures

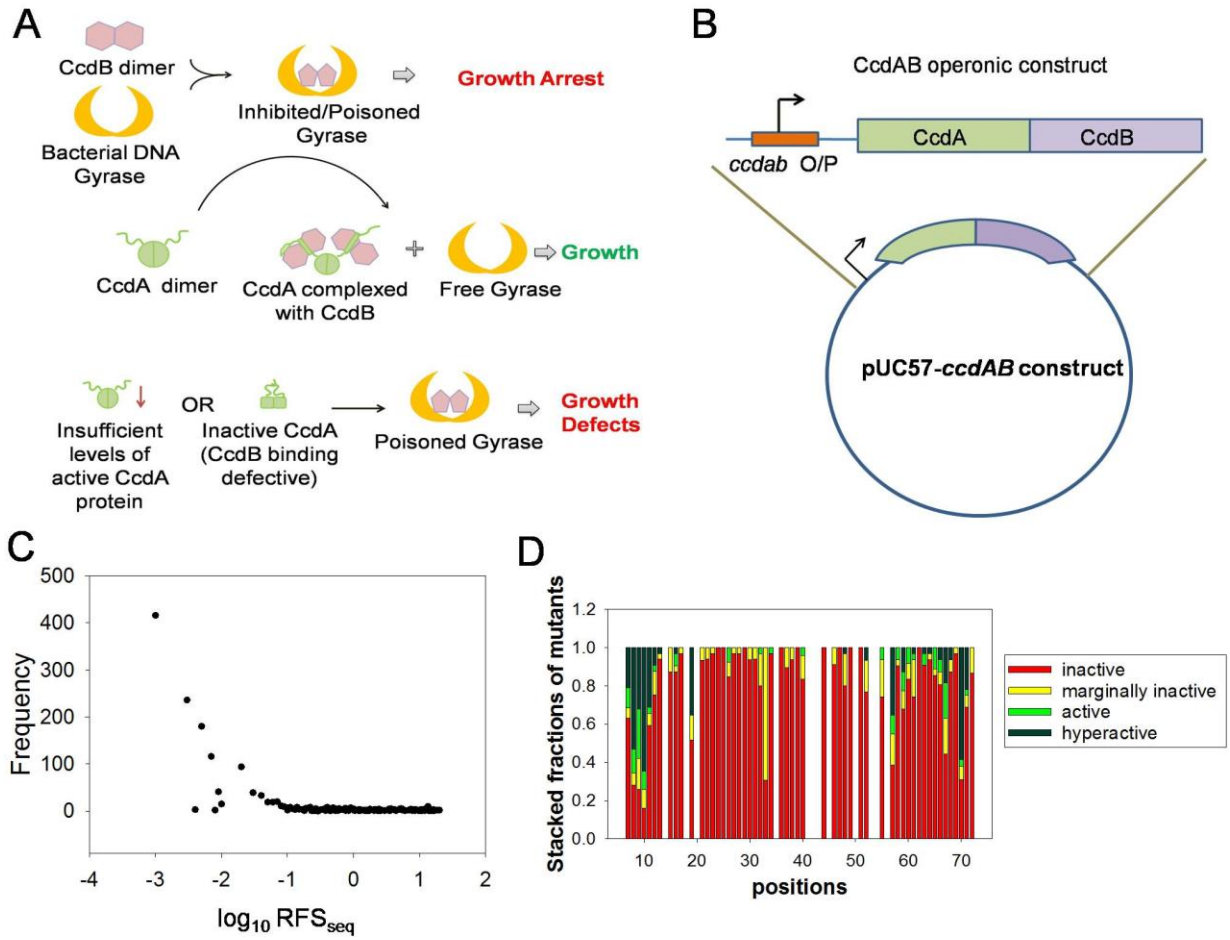


Figure 1. High Mutational Sensitivity of *ccdA* antitoxin gene in its native operonic context in *E.coli* (A) A schematic overview of the CcdAB TA module and the molecular mechanisms by which CcdB toxin triggers growth arrest in cells (top), how antitoxin CcdA rescues growth arrest by binding and sequestering CcdB (middle) and how mutations in CcdA can affect the antitoxic functions in the CcdAB TA system and promote CcdB mediated cell toxicity (bottom). (B) Schematic representation of *ccdAB* operon cloned in pUC57 vector. (C) Distribution of the Relative Fitness Score (RFS_{seq}) values for the *ccdA* mutants in the library, inferred through deep sequencing. A low RFS_{seq} value is indicative of low antitoxic activity of the corresponding *ccdA* mutant in an operonic context, relative to WT (RFS_{seq} =1). (D) Phenotypic landscape for

different residues in CcdA inferred from deep sequencing data. For each residue position in CcdA, the fraction of mutants showing inactive ($RFS_{seq} \leq 0.1$), marginally active ($0.1 < RFS_{seq} < 0.7$), active ($0.7 \leq RFS_{seq} \leq 1.5$) and hyperactive phenotypes ($RFS_{seq} > 1.5$) is plotted. Only positions which have data for more than 10 mutants available have been plotted here. A large number of mutants across the length of *ccdA* display a loss of a function phenotype.

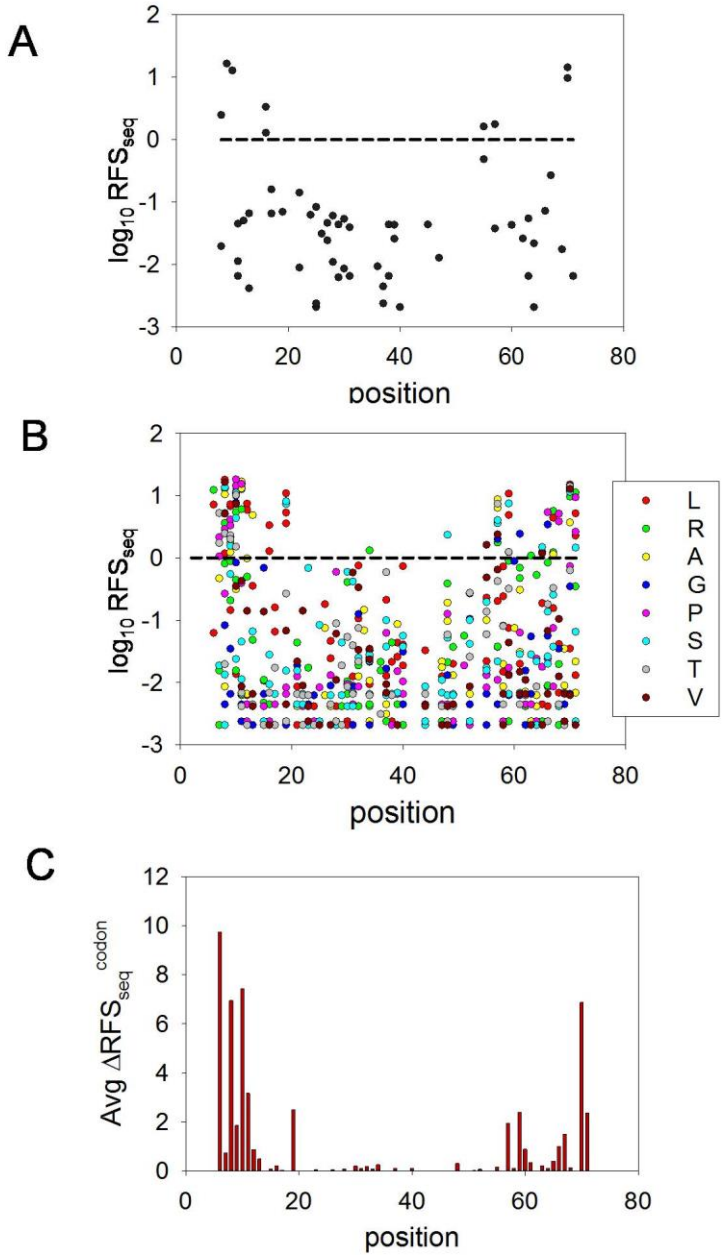


Figure 2. Inactive phenotype of synonymous mutants and codon specificity effects for non-synonymous mutants are non-uniformly distributed across the length of ccdA(A) Distribution of $\log_{10}RFS_{seq}$ scores for all 62 available single synonymous mutants as a function of residue position. The WT RFS_{seq} score of 1 ($\log_{10}RFS_{seq} = 0$) is depicted as a black dashed line. Loss of function mutations with low RFS_{seq} values are distributed across the length of the *ccdA* gene. A few mutants show WT like or higher RFS_{seq} values and are largely clustered at the

terminal regions of the gene. Phenotype of each synonymous mutant of CcdA is described in Table 2. (B) Distribution of $\log_{10}RFS_{seq}$ scores for all non-synonymous mutants that show differences in activity among synonymous codons encoding the same mutant amino acid. Codons encoding different amino acids are depicted in different colors. (C) Averaged pairwise absolute differences in the RFS_{seq} scores amongst codons as a function of residue position in CcdA. The differences between RFS_{seq} values of each pair of synonymous codons encoding the same mutant amino acids are calculated. The average of the absolute differences of all mutant codon pairs for a residue position is estimated as follows: Average $\Delta RFS_{seq}^{codon} = (\sum |RFS_{seq}^{a} - RFS_{seq}^{b}|) / {}^{n}C_{2}$, where a and b are different codons encoding a missense mutation and n is the total number of such codons at a position that have data for fitness scores. Regions close to the termini show the greatest variation in mutational effects, whereas the central region is largely populated by loss of function mutations.

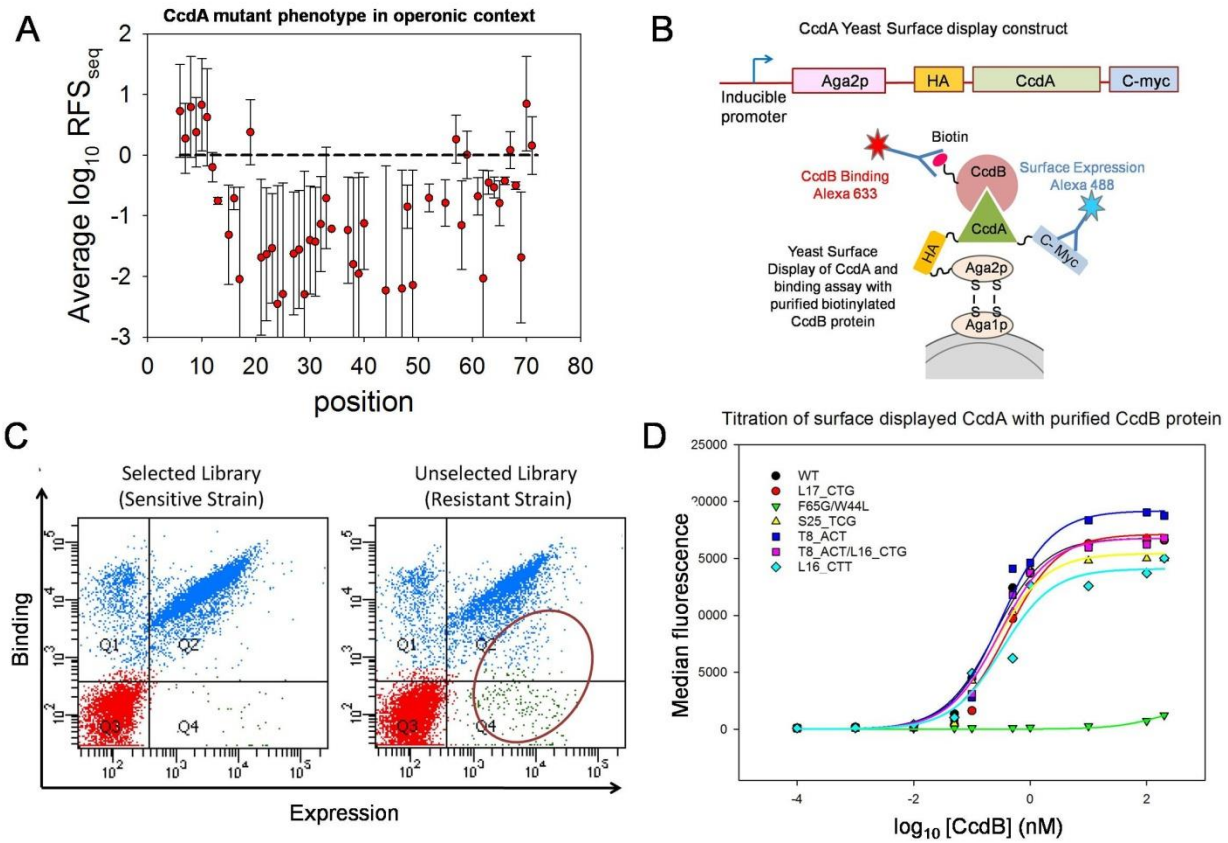


Figure 3. Observed growth defects in CcdA mutants are typically not caused by loss of CcdB binding activity. (A) Average $\log_{10}RFS_{seq}$ values (red circles) as a function of residue position. The error bars represent the associated standard deviation and the dashed line indicates the WT RFS_{seq} value of 1. (B) Schematic representation showing (top) a monocistronic fusion construct of CcdA gene for (bottom) yeast surface display and subsequent probing of surface expression and CcdB binding using flow cytometry (*see Methods*). (C) The CcdA mutant libraries derived from the resistant and sensitive strains of *E. coli* were cloned in the YSD vector, surface displayed and probed for surface expression and CcdB binding using FACS. The resistant strain derived library shows only marginally lower CcdB binding than the sensitive strain derived library, indicating that the number of CcdA mutants having significantly lower CcdB binding activity in comparison to WT CcdA is very low (population marked with red

circle). (D) Binding of individual CcdA mutants to WT CcdB probed by yeast surface display. CcdB binding affinity was assessed by titrating the cells displaying an individual CcdA mutant with various concentrations of biotinylated CcdB, followed by binding to streptavidin conjugated AlexaFluor-633. Labelled cells were analyzed on a BD FACSARIA III. The binding affinity was calculated by fitting median fluorescence intensity to a one site binding model. All single-site synonymous mutants of CcdA(showing significant variation in phenotypes in the operonic study, see Table 1) showed comparable binding affinity to WT (see Table 3). The double mutant F65G/W44L was taken as a negative control that shows impaired CcdB binding.

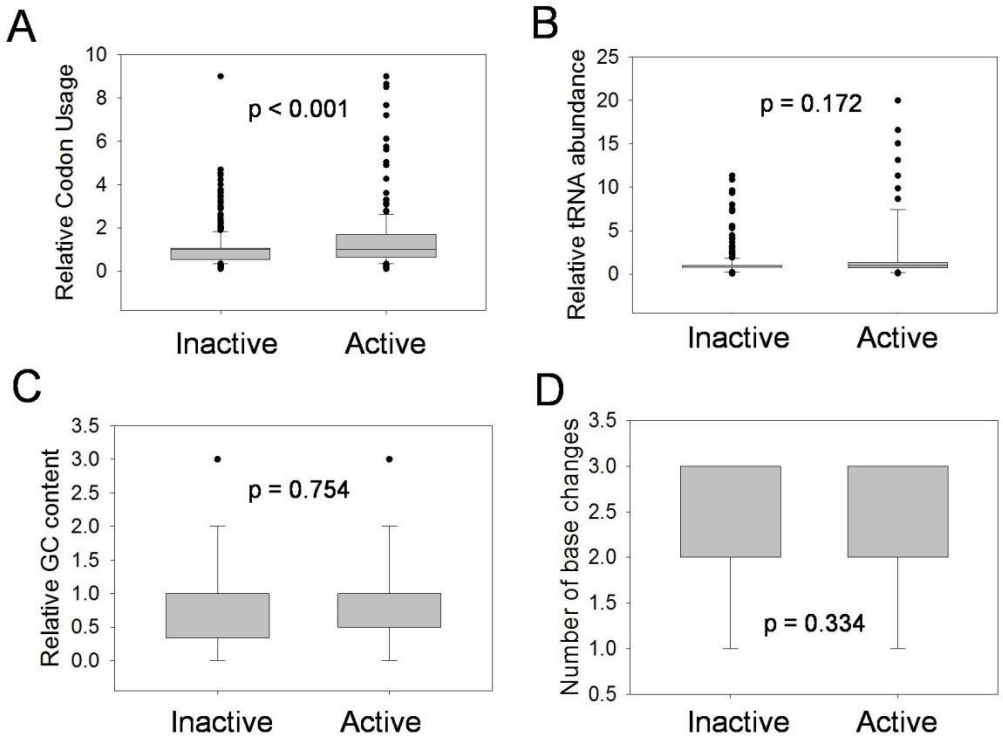


Figure 4. Relation between DNA sequence features and growth phenotypes of CcdA mutants. All available mutants in the CcdA library (N=1528) were divided into active (N=314) and inactive (N=1214) classes based on an RFS_{seq} score cut-off of 0.1. Mutant distribution was plotted with respect to various sequence based features, namely relative codon usage (A), relative tRNA abundance (B), relative GC content (C) and number of base changes upon mutation (D) calculated for mutant codon with respect to WT codon (*See Methods*). The distribution of sequence feature scores for inactive and active mutants have been plotted as box plots, where the central straight line depicts the median, the box the interquartile range (IQR), edges of the box the first and third quartile (Q1 and Q3), the whiskers minimum (Q1-1.5IQR) and maximum (Q3+1.5IQR) of the distribution and the black circles the outliers. Statistical significance was assessed by a Mann-Whitney Rank Sum Test. A p-value >0.05 indicates that there is no statistically significant difference between the distributions of the indicated parameter for active and inactive mutants, at a 95% confidence interval.

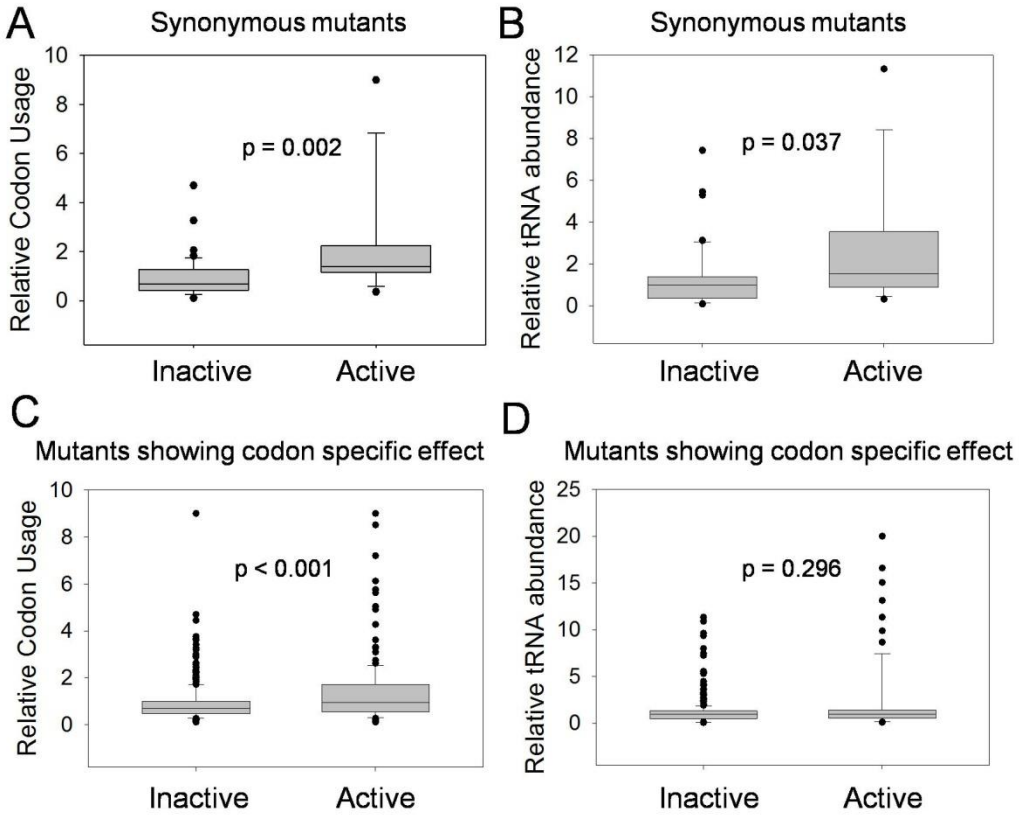


Figure 5. Phenotype of CcdA synonymous mutants and non-synonymous mutants that show codon specificity correlates well with codon preference. Single synonymous mutants (with respect to WT) in CcdA (N=62) were classified as active (N=14) and inactive (N=48) based on an RFS_{seq} score cut-off of 0.1. Similarly, we also examined fitness effects across CcdA missense mutants encoded by different codons (for mutated amino acids encoded by more than one codon). Such non-synonymous mutants showing codon specific effects (N= 550) were divided into active (N= 156) and inactive (N= 394) classes. Mutant distribution was analyzed with respect to relative codon usage (A and C) and relative tRNA abundance (B and D) of mutant relative to the WT codon (*see Methods*). (A-B) For active CcdA synonymous mutants, both relative codon usage and tRNA abundance are significantly higher than for corresponding inactive mutants (C-D) Active, non-synonymous CcdA variants that show codon specific effects,

exhibit a higher relative codon usage value than inactive ones. Statistical significance was assessed by a Mann-Whitney Rank Sum Test.

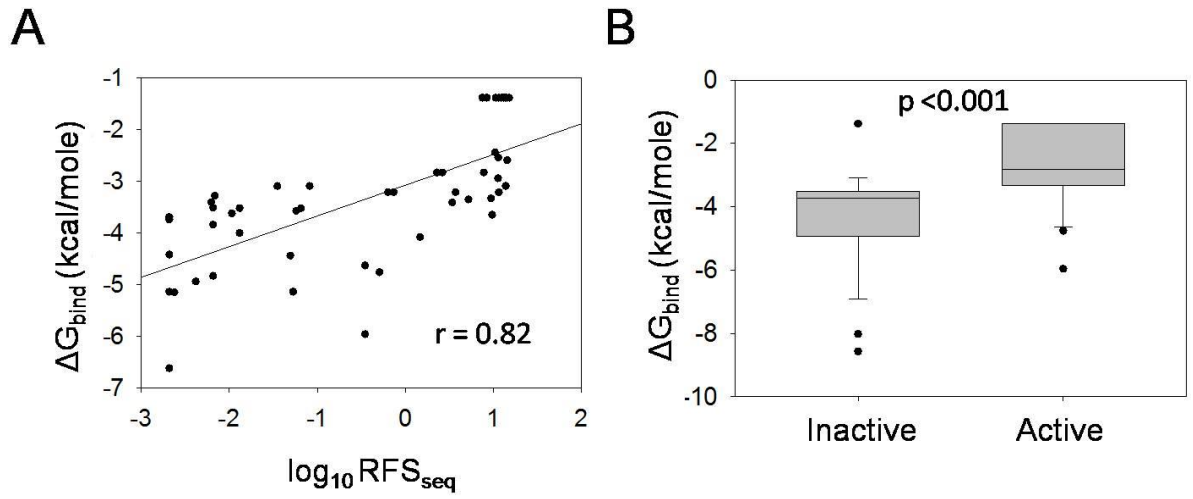


Figure 6. Effects of mutations in the CcdB RBS site in *ccdA* gene on phenotype in thenative operonic context. The free energy of binding to the anti-Shine Dalgarno sequence on the ribosome, denoted as ΔG_{bind} (CcdB RBS) was calculated for each of the possible single codon mutations at the CcdA residue positions 70 and 71 that constitute the ribosome binding site (RBS) of the downstream CcdB gene. Calculations were done using the RNAsubopt program in the Vienna RNA package. (A) The ΔG_{bind} (CcdB RBS) values show a strong positive correlation with the corresponding $\log_{10}RFS_{seq}$ values. This indicates that mutations at residues 70 or 71 of CcdA that improve ribosome binding and thus CcdB translation initiation, can lead to severe growth defects, resulting in inactive phenotypes. (B) Distribution of ΔG_{bind} (CcdB RBS) values for inactive and active mutants at residue positions 70 and 71 of CcdA are significantly different. Statistical significance was assessed by a Mann-Whitney Rank Sum Test.

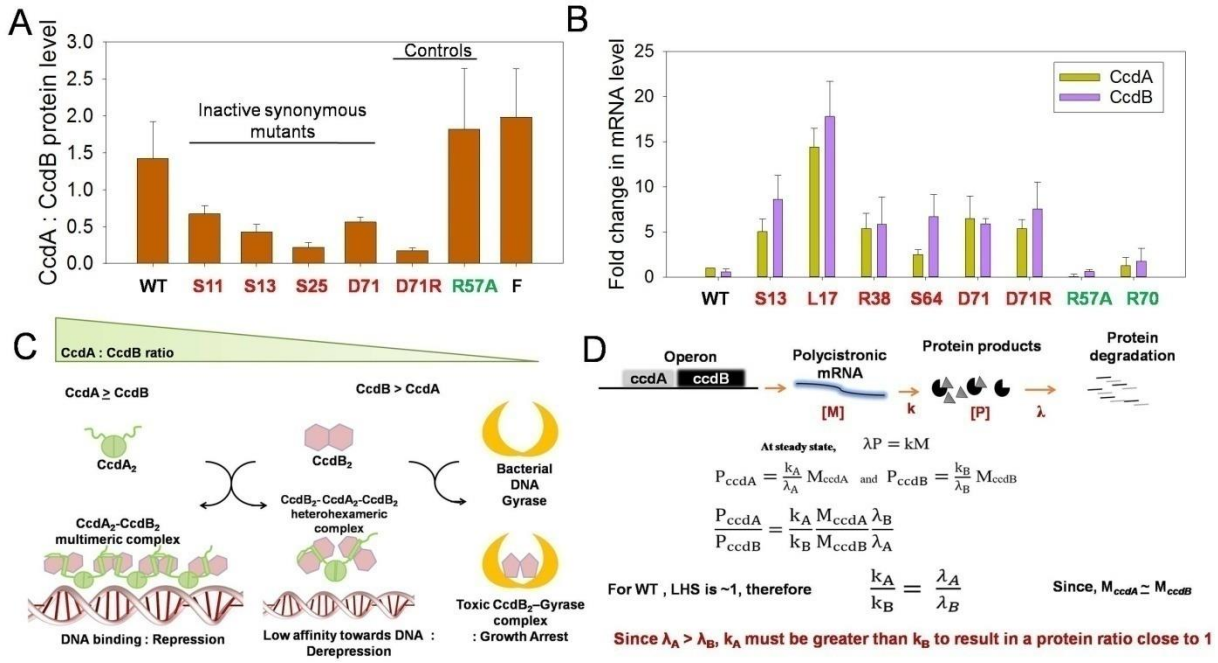


Figure 7. Decreased CcdA: CcdB protein ratio leads to loss of function phenotype for inactive synonymous CcdA mutants and also increases *ccdAB* mRNA levels by feedback autoregulation of the operon. (A) Relative levels of CcdA and CcdB peptides in *E. coli* (Top 10 Gyr strain) lysates were determined for WT as well as selected single synonymous mutants of CcdA using a quantitative proteomics approach. All synonymous inactive mutants studied have a decreased CcdA: CcdB protein ratio. The inactive non-synonymous mutant D71R which is predicted to improve CcdB translation due to a stronger CcdB RBS, also displays a decreased CcdA: CcdB protein ratio relative to WT. An active R57A mutant shows a higher CcdA: CcdB protein ratio than WT (See Table1 for phenotypic scores of mutants). F indicates the value for F-plasmid. Mutants showing active and inactive phenotypes in the operonic study are marked in red and green respectively. (B) Inactive synonymous mutants show higher levels of *ccdAB* mRNA, while active mutants have lower or WT comparable levels of *ccdAB* mRNA in Top10 Gyr *E. coli* cells. The mRNA levels of *ccdAB* specific transcripts were determined using q-RTPCR. *ccdA* and *ccdB* regions were both independently quantitated. Mean Ct represents the

threshold cycle for amplification obtained from duplicate samples from two different experiments. Fold change in *ccdA* and *ccdB* transcript levels are with respect to the WT Ct values (see Methods). Error bars indicate estimated standard deviation of the measurements. (C) Schematic model for CcdA: CcdB ratio dependent autoregulation of *ccdAB* operon transcription. (D) Differential translational efficiency of *ccdA* and *ccdB* genes in the operon. The steady state levels of CcdA (P_{ccdA}) and CcdB (P_{ccdB}) proteins are assumed to scale linearly as a function of their mRNA concentration (M_{ccdA} and M_{ccdB}). [M] and [P] represents mRNA and protein concentrations respectively. Translation efficiency (k_A for *ccdA* and k_B for *ccdB*) is the rate of protein production per mRNA. λ_A and λ_B are the protein degradation rate constants of CcdA and CcdB.

A

RelB mutants	WT codon	Mutant codon	Relative codon usage
S3	AGC	TCT	0.68
R7	CGT	AGA	0.19
S16	TCT	AGC	1.47
S28	TCT	AGC	1.47
D77	GAT	GAG	0.51

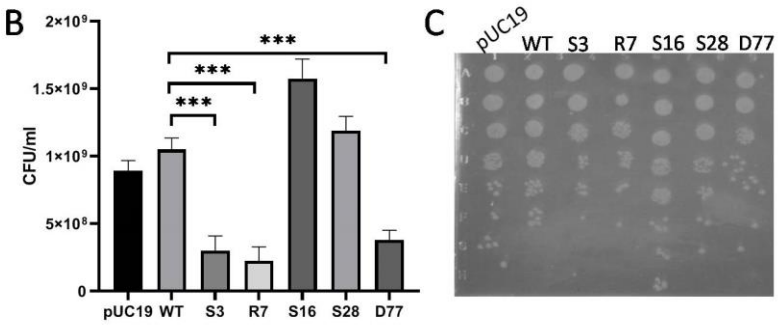


Figure 8. Phenotypic effects of select single synonymous substitutions in *relB* antitoxin gene of the RelBE operon. (A) List of various synonymous mutations made in the *relB* gene in the native operon context along with their relative codon usage upon mutation with respect to the WT codons. A single base substitution at the D77 residue of RelB which increases the strength of the SD sequence for downstream toxin RelE was used as a control mutant. (B-C) Growth of single synonymous mutants in *E. coli* BW25113, sensitive strain. Following transformation, ten-fold dilutions were plated (B) as well as spotted (C) on LB-Agar plates. The number of colonies formed in the dilution plate with ~50-100 colonies were counted for each mutant and CFU/ml was calculated and reported. (B) The D77 synonymous mutant, expected to increase the levels of the downstream RelE toxin, displays an inactive phenotype. RelB synonymous mutants with rarer codons display decreased growth, while optimal codon bearing mutants exhibit higher growth relative to WT. The error bars indicate standard error. '*' indicate significant differences from the WT (*** p value <0.001, paired t-test).

Tables

Table 1. Validation of mutant phenotypes inferred from deep sequencing

Growth phenotype on plates (Supplementary Figure S4) for individual CcdA mutants were compared with values of RFS_{seq} obtained from deep sequencing.

Mutant Name	Mutant Codon	Phenotype[a]	Description	RFS_{seq}
S11	TCT	Inactive *	Synonymous	0.011
L17	CTG	Inactive	Synonymous	0.066
V22	GTG	Inactive	Synonymous	0.009
S25	TCG	Inactive	Synonymous	0.002
T31	ACG	Inactive	Synonymous	0.007
R38	AGG	Inactive	Synonymous	0.007
S64	AGT	Inactive	Synonymous	0.002
D71	GAT	Inactive	Synonymous	0.007
D71R	AGG	Inactive	Non-synonymous	0.000
V28	GTG	Inactive	Synonymous	0.011
S13	AGT	Inactive	Synonymous	0.066
V55	GTG	Moderately inactive	Synonymous	0.486
S29	AGT	Inactive	Synonymous	0.044
V9	GTG	Hyperactive	Synonymous	16.505
R70	CGG	Hyperactive	Synonymous	9.673
R57	AGG	Hyperactive	Synonymous	1.765
V55	GTT	Hyperactive	Synonymous	1.624
L16	CTG	Active	Synonymous	1.286
V55S	TCG	Inactive	Non-synonymous	0.017
S11G	GGT	Inactive	Non-synonymous	0.002
G63L	CTG	Inactive	Non-synonymous	0.002
N69L	CTG	Inactive	Non-synonymous	0.000
A66P	CCT	Hyperactive	Non-synonymous	5.416
R57A	GCG	Hyperactive	Non-synonymous	8.718
M52I	ATT	Hyperactive	Non-synonymous	2.062
T30S	TCG	Moderately inactive *	Non-synonymous	0.598
F58L	CTG	Active	Non-synonymous	0.730
T8	ACT	Hyperactive	Synonymous	2.482

[a]mutant phenotype is assigned based on the deep sequencing derived RFS_{seq} values. These results correlate well with growth of individual mutants on plate, except in case of only two mutants, namely S11_TCT and T30S_TCG where growth on plates (see Supplementary Figure S4B) and deep sequencing results do not agree. These cases have been marked with an *.

Table 2. Mutational sensitivities of all available synonymous mutant codons at each residue position of CcdA library in operonic context.

The codons (synonymous with respect to WT sequence of CcdA) have been colored based on the deep sequencing phenotypic activity (RFS_{seq} scores) in operonic context. Dark green depicts hyperactivity ($RFS_{seq}>1.5$), while light green denotes WT like activity ($0.7 \leq RFS_{seq} \leq 1.5$). Yellow and red indicate marginally inactive mutants ($0.1<RFS_{seq}< 0.7$) and highly inactive mutants ($RFS_{seq} \leq 0.1$) respectively.

Residue position	Synonymous mutant codons		
8	ACG	ACT	
9	GTG		
10	GAT		
11	AGT	TCG	TCT
12	GAT		
13	AGT	TCG	TCT
16	CTG	CTT	
17	CTG	CTT	TTG
19	GCT		
22	GTG	GTT	
24	ATT		
25	AGT	TCG	TCT
26	GGG		
27	CTT	TTG	
28	GTG	GTT	
29	AGT	TCG	TCT
30	ACG	ACT	
31	ACG	ACT	
36	GCG		
37	AGG	CGG	
38	AGG	CGG	
39	CTT	TTG	
40	AGG	CGG	
45	AAG		
47	GAG		
55	GTG	GTT	
57	AGG	CGT	
60	GAG		
62	AAT		
63	GGG	GGT	
64	AGT	TCG	
66	GCG		
67	GAT		
69	AAT		
70	CGG	CGT	
71	GAT		

Table 3. Binding affinities of CcdA mutants and WT calculated using titration of yeast surface displayed CcdA molecules with purified CcdB.

The dissociation constants (K_d) were obtained by fitting the fluorescence intensity titration traces to a single ligand binding model. Single codon synonymous mutants L17_CTG and S25_TCG mutants display an inactive phenotype, while T8_ACT and L16_CTT have an active phenotype based on the deep sequencing studies.

CcdA	**K_d (nM)**
WT	0.24±.09
L17_CTG	0.4±0.1
S25_TCG	0.24±0.07
T8_ACT	0.31±0.12
L16_CTT	0.32±0.06
T8_ACT/L16_CTG	0.29±0.09
F65G/W44L	>200

Author Contributions: K.G. prepared the SSM library. S.C. carried out the experiments except for the quantitative proteomics (S.V.M. and H.G.) and the experiments with RelBE system (P.K). S.K. processed the deep sequencing data. A.A helped in acquiring FACS data. R.V, S.C. and K.G. designed the experiments, analyzed the overall data and wrote the manuscript with critical inputs and review from all other authors.

Conflict of interest statement: The authors declare no conflict of interest.

Data and Material availability: The data relevant to the figures in the paper have been made available within the article and in the Supplementary Information section. The deep sequencing data are submitted in NCBI SRA database and to be released upon acceptance of manuscript (BioProject ID: PRJNA795874) All materials generated in this study are available from the lead contact (varadar@iisc.ac.in) without restriction.

Funding: This work was funded in part by a grant to RV from the Department of Biotechnology (grant number BT/COE/34/SP15219/2015, DT.20/11/2015), Government of India, from Science and Engineering Research Board (SERB), Government of India (grant number EMR/2017/0040S4 and SR/S2/JCB-10/2007 (JC Bose Fellowship)). We also acknowledge funding for infrastructural support from the following programs of the Government of India: DST FIST, UGC Centre for Advanced study, Ministry of Human Resource Development (MHRD), and the DBT IISc Partnership Program. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

References

[1] F. Jacob and J. Monod, "On the Regulation of Gene Activity," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 26, pp. 193–211, 1961.

[2] C. Yanofsky, "Attenuation in the control of expression of bacterial operons," *Nature*, vol. 289, no. 5800, pp. 751–758, 1981.

[3] G. Balázsi, A.-L. Barabási, and Z. N. Oltvai, "Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 22, pp. 7841–7846, May 2005.

[4] M. N. Price, A. P. Arkin, and E. J. Alm, "The Life-Cycle of Operons," *PLOS Genet.*, vol. 2, no. 6, p. e96, Jun. 2006.

[5] H. Zhang, Y. Yin, V. Olman, and Y. Xu, "Genomic arrangement of regulons in bacterial genomes," *PLoS One*, vol. 7, no. 1, pp. e29496–e29496, 2012.

[6] G. Moreno-Hagelsieb and J. Collado-Vides, "A powerful non-homology method for the prediction of operons in prokaryotes.," *Bioinformatics*, vol. 18 Suppl 1, pp. S329-36, 2002.

[7] Y. Chemla *et al.*, "A possible universal role for mRNA secondary structure in bacterial translation revealed using a synthetic operon.," *Nat. Commun.*, vol. 11, no. 1, p. 4827, Sep. 2020.

[8] R. Byrne, J. G. Levin, H. A. Bladen, and M. W. Nirenberg, "The in vitro formation of a DNA-ribosome complex," *Procee.dings Natl. Acad. Sci. United States Am.*, vol. 52, no. 1, pp. 140–148, Jul. 1964.

[9] O. L. J. Miller, B. A. Hamkalo, and C. A. J. Thomas, “Visualization of bacterial genes in action.,” *Science*, vol. 169, no. 3943, pp. 392–395, Jul. 1970.

[10] G. S. Stent, “The operon: on its third anniversary. Modulation of transfer RNA species can provide a workable model of an operator-less operon.,” *Science*, vol. 144, no. 3620, pp. 816–820, May 1964.

[11] F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, “General Nature of the Genetic Code for Proteins,” *Nature*, vol. 192, no. 4809, pp. 1227–1232, 1961.

[12] A. A. Komar, T. Lesnik, and C. Reiss, “Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation.,” *FEBS Lett.*, vol. 462, no. 3, pp. 387–391, Dec. 1999.

[13] I. M. Sander, J. L. Chaney, and P. L. Clark, “Expanding Anfinsen’s principle: contributions of synonymous codon selection to rational protein design.,” *J. Am. Chem. Soc.*, vol. 136, no. 3, pp. 858–861, Jan. 2014.

[14] C. Kimchi-Sarfaty *et al.*, “A ‘silent’ polymorphism in the MDR1 gene changes substrate specificity.,” *Science*, vol. 315, no. 5811, pp. 525–528, Jan. 2007.

[15] Z. E. Sauna and C. Kimchi-Sarfaty, “Understanding the contribution of synonymous mutations to human disease.,” *Nat. Rev. Genet.*, vol. 12, no. 10, pp. 683–691, Aug. 2011.

[16] F. Buhr *et al.*, “Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations.,” *Mol. Cell*, vol. 61, no. 3, pp. 341–351, Feb. 2016.

[17] M. V Rodnina, “The ribosome in action: Tuning of translational efficiency and protein folding.,” *Protein Sci.*, vol. 25, no. 8, pp. 1390–1406, Aug. 2016.

[18] F. Zhao *et al.*, “Genome-wide role of codon usage on transcription and identification of potential regulators.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 6, Feb. 2021.

[19] V. Presnyak *et al.*, “Codon optimality is a major determinant of mRNA stability.,” *Cell*, vol. 160, no. 6, pp. 1111–1124, Mar. 2015.

[20] G. Hanson and J. Coller, “Codon optimality, bias and usage in translation and mRNA decay.,” *Nat. Rev. Mol. Cell Biol.*, vol. 19, no. 1, pp. 20–30, Jan. 2018.

[21] S. Karlin, J. Mrázek, and A. M. Campbell, “Codon usages in different gene classes of the Escherichia coli genome.,” *Mol. Microbiol.*, vol. 29, no. 6, pp. 1341–1355, Sep. 1998.

[22] P. M. Sharp, T. M. Tuohy, and K. R. Mosurski, “Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes.,” *Nucleic Acids Res.*, vol. 14, no. 13, pp. 5125–5143, Jul. 1986.

[23] T. E. F. Quax, N. J. Claassens, D. Söll, and J. van der Oost, “Codon Bias as a Means to Fine-Tune Gene Expression.,” *Mol. Cell*, vol. 59, no. 2, pp. 149–161, Jul. 2015.

[24] T. E. Gorochowski, Z. Ignatova, R. A. L. Bovenberg, and J. A. Roubos, “Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate.,” *Nucleic Acids Res.*, vol. 43, no. 6, pp. 3022–3032, Mar. 2015.

[25] G.-W. Li, E. Oh, and J. S. Weissman, “The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria.,” *Nature*, vol. 484, no. 7395, pp. 538–541, Mar. 2012.

[26] S. Masachis, N. J. Tourasse, S. Chabas, O. Bouchez, and F. Darfeuille, “FASTBAC-Seq: Functional Analysis of Toxin–Antitoxin Systems in Bacteria by Deep Sequencing,” in

Methods in Enzymology, vol. 612, pp. 67–100, 2018.

[27] H. S. Deter, R. V Jensen, W. H. Mather, and N. C. Butzin, "Mechanisms for Differential Protein Production in Toxin-Antitoxin Systems.," *Toxins (Basel).*, vol. 9, no. 7, Jul. 2017.

[28] N. Goeders and L. Van Melderen, "Toxin-antitoxin systems as multilevel interaction systems.," *Toxins (Basel).*, vol. 6, no. 1, pp. 304–324, Jan. 2014.

[29] D. P. Pandey and K. Gerdes, "Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes.," *Nucleic Acids Res.*, vol. 33, no. 3, pp. 966–976, 2005.

[30] D. W. Holden, J. Errington, and G. Storz, "Type II Toxin-Antitoxin Systems and Persister Cells," *MBio*, vol. 9, no. 5, pp. e01574-18, 2018.

[31] F. Goormaghtigh *et al.*, "Reassessing the Role of Type II Toxin-Antitoxin Systems in Formation of Escherichia coli Type II Persister Cells.," *MBio*, vol. 9, no. 3, Jun. 2018.

[32] A. Vandervelde, R. Loris, J. Danckaert, and L. Gelens, "Computational Methods to Model Persistence.," *Methods Mol. Biol.*, vol. 1333, pp. 207–240, 2016.

[33] I. Cataudella, A. Trusina, K. Sneppen, K. Gerdes, and N. Mitarai, "Conditional cooperativity in toxin-antitoxin regulation prevents random toxin activation and promotes fast translational recovery.," *Nucleic Acids Res.*, vol. 40, no. 14, pp. 6424–6434, Aug. 2012.

[34] H. Afif, N. Allali, M. Couturier, and L. Van Melderen, "The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system.," *Mol. Microbiol.*, vol. 41, no. 1, pp. 73–82, Jul. 2001.

[35] L. Van Melderen, M. H. Thi, P. Lecchi, S. Gottesman, M. Couturier, and M. R. Maurizi, "ATP-dependent degradation of CcdA by Lon protease. Effects of secondary structure and heterologous subunit interactions.," *J. Biol. Chem.*, vol. 271, no. 44, pp. 27730–27738, Nov. 1996.

[36] K. Gupta and R. Varadarajan, "Insights into protein structure, stability and function from saturation mutagenesis.," *Curr. Opin. Struct. Biol.*, vol. 50, pp. 117–125, Jun. 2018.

[37] L. A. Miosge *et al.*, "Comparison of predicted and actual consequences of missense mutations.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 37, pp. E5189-98, Sep. 2015.

[38] J. I. Boucher, D. N. A. Bolon, and D. S. Tawfik, "Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature.," *Protein Sci.*, vol. 25, no. 7, pp. 1219–1226, Jul. 2016.

[39] P. C. Jain and R. Varadarajan, "A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library," *Anal. Biochem.*, vol. 449, no. 1, pp. 90–98, 2014.

[40] K. A. Datsenko and B. L. Wanner, "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 12, pp. 6640–6645, Jun. 2000.

[41] P. Rice, I. Longden, and A. Bleasby, "EMBOSS: the European Molecular Biology Open Software Suite.," *Trends Genet.*, vol. 16, no. 6, pp. 276–277, Jun. 2000.

[42] T. Carver and A. Bleasby, "The design of Jemboss: a graphical user interface to EMBOSS.," *Bioinformatics*, vol. 19, no. 14, pp. 1837–1843, Sep. 2003.

[43] B. MacLean *et al.*, “Skyline: an open source document editor for creating and analyzing targeted proteomics experiments,” *Bioinformatics*, vol. 26, no. 7, pp. 966–968, Apr. 2010.

[44] M. B. Stead *et al.*, “RNAsnap^{TM}: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria.,” *Nucleic Acids Res.*, vol. 40, no. 20, p. e156, Nov. 2012.

[45] M. Zuker, “Mfold web server for nucleic acid folding and hybridization prediction,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3406–3415, Jul. 2003.

[46] A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neuböck, and I. L. Hofacker, “The Vienna RNA websuite.,” *Nucleic Acids Res.*, vol. 36, no. Web Server issue, pp. W70-4, Jul. 2008.

[47] H. Dong, L. Nilsson, and C. G. Kurland, “Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates.,” *J. Mol. Biol.*, vol. 260, no. 5, pp. 649–663, Aug. 1996.

[48] F. F. V Chevance, S. Le Guyon, and K. T. Hughes, “The effects of codon context on in vivo translation speed.,” *PLoS Genet.*, vol. 10, no. 6, p. e1004392, Jun. 2014.

[49] G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, and K. D. Wittrup, “Isolating and engineering human antibodies using yeast surface display,” *Nat. Protoc.*, vol. 1, no. 2, pp. 755–768, 2006.

[50] U. Rathore *et al.*, “Glycosylation of the core of the HIV-1 envelope subunit protein gp120 is not required for native trimer formation or viral infectivity,” *J Biol Chem*, vol. 292, no. 24, pp. 10197–10219, 2017.

[51] S. Chandra, G. Chattopadhyay, and R. Varadarajan, “Rapid Identification of Secondary

Structure and Binding Site Residues in an Intrinsically Disordered Protein Segment," *Front. Genet.*, vol. 12, p. 2173, 2021.

[52] F. W. Bech, S. T. Jørgensen, B. Diderichsen, and O. H. Karlström, "Sequence of the relB transcription unit from Escherichia coli and identification of the relB gene.," *EMBO J.*, vol. 4, no. 4, pp. 1059–1066, Apr. 1985.

[53] J. E. Tam and B. C. Kline, "Control of the ccd operon in plasmid F.," *J. Bacteriol.*, vol. 171, no. 5, pp. 2353–2360, May 1989.

[54] T. Miki, K. Yoshioka, and T. Horiuchi, "Control of cell division by sex factor F in Escherichia coli: I. The 42.84–43.6 F segment couples cell division of the host bacteria with replication of plasmid DNA," *J. Mol. Biol.*, vol. 174, no. 4, pp. 605–625, 1984.

[55] P. Bernard *et al.*, "The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase.," *J. Mol. Biol.*, vol. 234, no. 3, pp. 534–541, Dec. 1993.

[56] S. Maki, S. Takiguchi, T. Horiuchi, K. Sekimizu, and T. Miki, "Partner switching mechanisms in inactivation and rejuvenation of Escherichia coli DNA gyrase by F plasmid proteins LetD (CcdB) and LetA (CcdA).," *J. Mol. Biol.*, vol. 256, no. 3, pp. 473–482, Mar. 1996.

[57] N. De Jonge *et al.*, "Rejuvenation of CcdB-Poisoned Gyrase by an Intrinsically Disordered Protein Domain," *Mol. Cell*, vol. 35, no. 2, pp. 154–163, 2009.

[58] I. A. Krasheninnikov, A. A. Komar, and I. A. Adzhubeĭ, "[Role of the rare codon clusters in defining the boundaries of polypeptide chain regions with identical secondary structures in the process of co-translational folding of proteins].," *Dokl. Akad. Nauk SSSR*,

vol. 303, no. 4, pp. 995–999, 1988.

[59] T. Madl *et al.*, “Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA.,” *J. Mol. Biol.*, vol. 364, no. 2, pp. 170–185, Nov. 2006.

[60] D. C. Shields and P. M. Sharp, “Synonymous codon usage in Bacillus subtilis reflects both translational selection and mutational biases.,” *Nucleic Acids Res.*, vol. 15, no. 19, pp. 8023–8040, Oct. 1987.

[61] P. M. Sharp and K. M. Devine, “Codon usage and gene expression level in Dictyostelium discoideum: highly expressed genes do ‘prefer’ optimal codons.,” *Nucleic Acids Res.*, vol. 17, no. 13, pp. 5029–5039, Jul. 1989.

[62] A. Das and C. Yanofsky, “Restoration of a translational stop-start overlap reinstates translational coupling in a mutant trpB’-trpA gene pair of the Escherichia coli tryptophan operon.,” *Nucleic Acids Res.*, vol. 17, no. 22, pp. 9333–9340, Nov. 1989.

[63] A. Vandervelde *et al.*, “Molecular mechanism governing ratio-dependent transcription regulation in the ccdAB operon.,” *Nucleic Acids Res.*, vol. 45, no. 6, pp. 2937–2950, Apr. 2017.

[64] F. Mohammad, C. J. Woolstenhulme, R. Green, and A. R. Buskirk, “Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling.,” *Cell Rep.*, vol. 14, no. 4, pp. 686–694, Feb. 2016.

[65] I. Cherny, M. Overgaard, J. Borch, Y. Bram, K. Gerdes, and E. Gazit, “Structural and thermodynamic characterization of the Escherichia coli RelBE toxin-antitoxin system: indication for a functional role of differential stability.,” *Biochemistry*, vol. 46, no. 43, pp.

12152–12163, Oct. 2007.

[66] R. A. Fasani and M. A. Savageau, “Molecular mechanisms of multiple toxin-antitoxin systems are coordinated to govern the persister phenotype.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 27, pp. E2528-37, Jul. 2013.

[67] G. Bhagavatula, M. S. Rich, D. L. Young, M. Marin, and S. Fields, “A Massively Parallel Fluorescence Assay to Characterize the Effects of Synonymous Mutations on TP53 Expression.,” *Mol. Cancer Res.*, vol. 15, no. 10, pp. 1301–1307, Oct. 2017.

[68] E. D. Kelsic, H. Chung, N. Cohen, J. Park, H. H. Wang, and R. Kishony, “RNA Structural Determinants of Optimal Codons Revealed by MAGE-Seq.,” *Cell Syst.*, vol. 3, no. 6, pp. 563-571.e6, Dec. 2016.

[69] G. Boël *et al.*, “Codon influence on protein expression in E. coli correlates with mRNA levels.,” *Nature*, vol. 529, no. 7586, pp. 358–363, Jan. 2016.

[70] F. Mohammad, R. Green, and A. R. Buskirk, “A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution.,” *Elife*, vol. 8, Feb. 2019.

[71] S. M. Studer and S. Joseph, “Unfolding of mRNA secondary structure by the bacterial translation initiation complex.,” *Mol. Cell*, vol. 22, no. 1, pp. 105–115, Apr. 2006.

[72] C. O. Gualerzi and C. L. Pon, “Initiation of mRNA translation in bacteria: structural and dynamic aspects.,” *Cell. Mol. Life Sci.*, vol. 72, no. 22, pp. 4341–4367, Nov. 2015.

[73] A. M. Mustoe, M. Corley, A. Laederach, and K. M. Weeks, “Messenger RNA Structure Regulates Translation Initiation: A Mechanism Exploited from Bacteria to Humans.,” *Biochemistry*, vol. 57, no. 26, pp. 3537–3539, Jul. 2018.

[74] M. A. Sørensen, C. G. Kurland, and S. Pedersen, “Codon usage determines translation rate in Escherichia coli,” *J. Mol. Biol.*, vol. 207, no. 2, pp. 365–377, 1989.

[75] M. Robinson *et al.*, “Codon usage can affect efficiency of translation of genes in Escherichia coli,” *Nucleic Acids Res.*, vol. 12, no. 17, pp. 6663–6671, Sep. 1984.

[76] C.-H. Yu *et al.*, “Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding.,” *Mol. Cell*, vol. 59, no. 5, pp. 744–754, Sep. 2015.

Supplementary Information

GGAGATGGCCAAAACCCCAAGTTACGGATCTTCCTCTCCCTCCGCACAGCGTTACATCCCGTCAGCACAGCATGT

AGTGCCTCATACAGTTGCCCATGGCACTATATGTTGTGTTGTATCTCTGGACTGTGATGCGCCGCGCAGGGGCGG

Eco RI

AAAACAGCGATATGATGATTTTCTCAGCGTTGTACACTTCCGGAAAGTCGTTTATTCAAATAAAGTCGAATTCATAC

GAAACGGGAATGCGGTAATTACGCTTTGTTTTTATAAGTCAGATTTTAATTTTTATTGGTTAACATAACGAAAGGTA

PsiI

AAATACATAAGGCTTATAAAAAGCCAGATAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATAC

Bgl 2

TGATATGTATACCCGAAGTATGTCAAAAAGATCTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCGACA

GCTATCAGTTGCTCAAGGCGTATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCG

TCTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAAGGGATGGCTGAGGTCGCCCGGTTTATTGAAATGAACG

Dra I

GCTCTTTTGCTGACGAGAACAGGGACTGGTGAAATGCAGTTTAAAGTTTACACCTATAAAAGAGAGAGCCGTTATC

GTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCAC

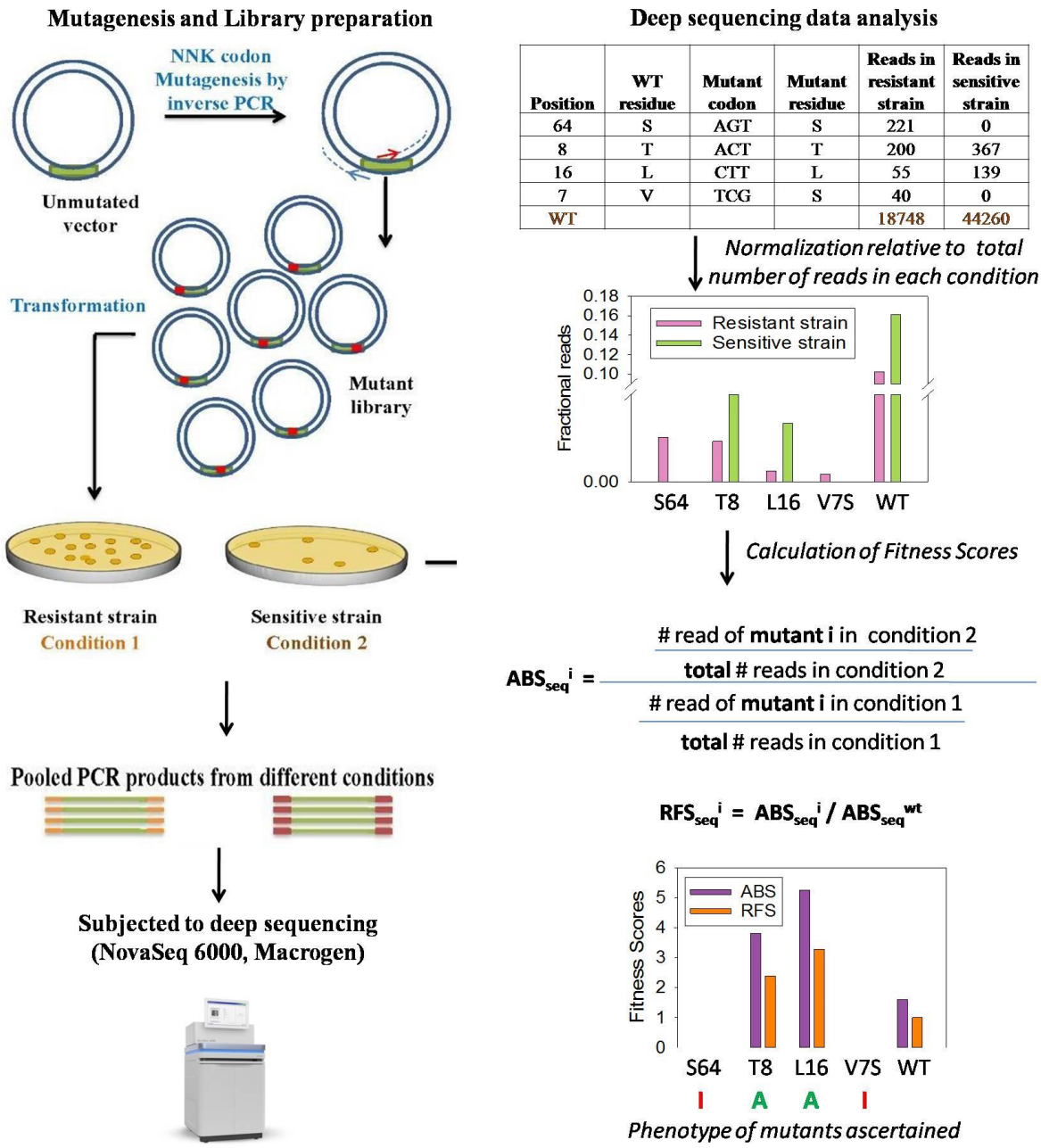
GTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGA

CCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACA

Bam HI

TCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTGAGGATCCGTTATACAC

Figure S1. Sequence of *ccdAB* operon used in the study. Sequence of *ccd* operon synthesized and cloned in pUC57 vector at GenScript (USA). Restriction sites have been underlined and marked. The *ccdA* coding region is in green and *ccdB* is in purple. Mutations in the construct with respect to the native F-plasmid borne *ccd* sequence were introduced to facilitate cloning of the *ccdA* and *ccdB* libraries and are shown in red. Following the *ccdB* gene, the first few residues of the *resD* gene have been included with a stop codon at the second residue position and a mutation to introduce a BamH1 site.



Position	WT residue	Mutant codon	Mutant residue	Reads in resistant strain	Reads in sensitive strain
64	S	AGT	S	221	0
8	T	ACT	T	200	367
16	L	CTT	L	55	139
7	V	TCG	S	40	0
WT				18748	44260

Figure S2. Saturation mutagenesis coupled to deep sequencing to infer mutational sensitivities. Single-site saturation mutagenesis libraries are generated using inverse PCR based methodology that involves use of non-overlapping primers with the mutant NNK codon at the 5' end of the forward primer. Mutant phenotypes are assayed by transforming into resistant

(condition 1) and sensitive strains (condition 2). Pooled plasmids from different conditions are used as template to amplify the gene of interest with primers having condition specific tags. Pooled PCR products are subject to deep sequencing (NovaSeq6000 platform, Macrogen) in three biological replicates and the data is analyzed to obtain the relative fitness scores (RFS_{seq}) of the mutants relative to WT (*See Methods*). The number of reads for each mutant in a condition is first normalized with respect to the total number of reads in that condition. The ratio of relative fraction of reads in the sensitive (condition 2) to that in resistant strain (condition1) gives us the absolute fitness scores (ABS_{seq}). To compare mutants relative to WT, we use relative fitness scores (RFS_{seq}). The RFS_{seq} score is the ABS_{seq} of mutant normalized with respect to WT. WT RFS_{seq} =1. Mutants with lower RFS_{seq} show reduced growth in the sensitive strain, relative to WT and therefore are inactive (I), while mutants with RFS_{seq} scores similar to or higher than WT are active (A).

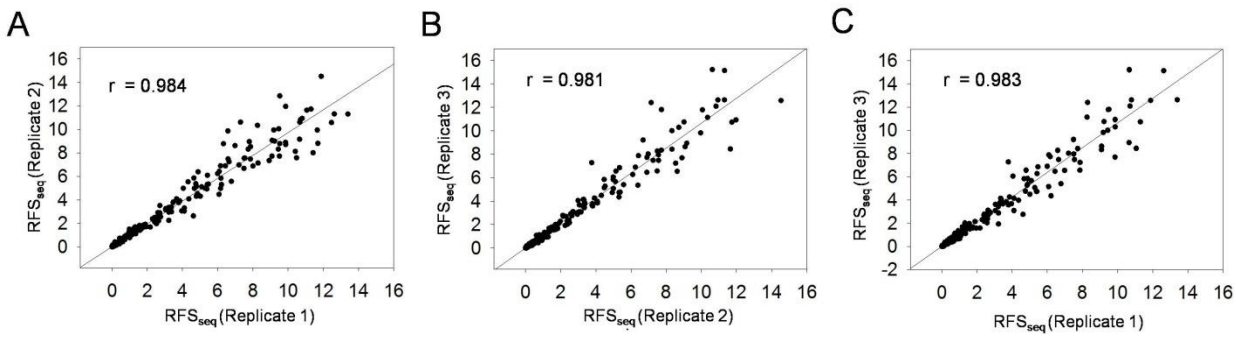


Figure S3. The phenotypic scores (RFS_{seq}) of mutants are highly correlated between biological replicates

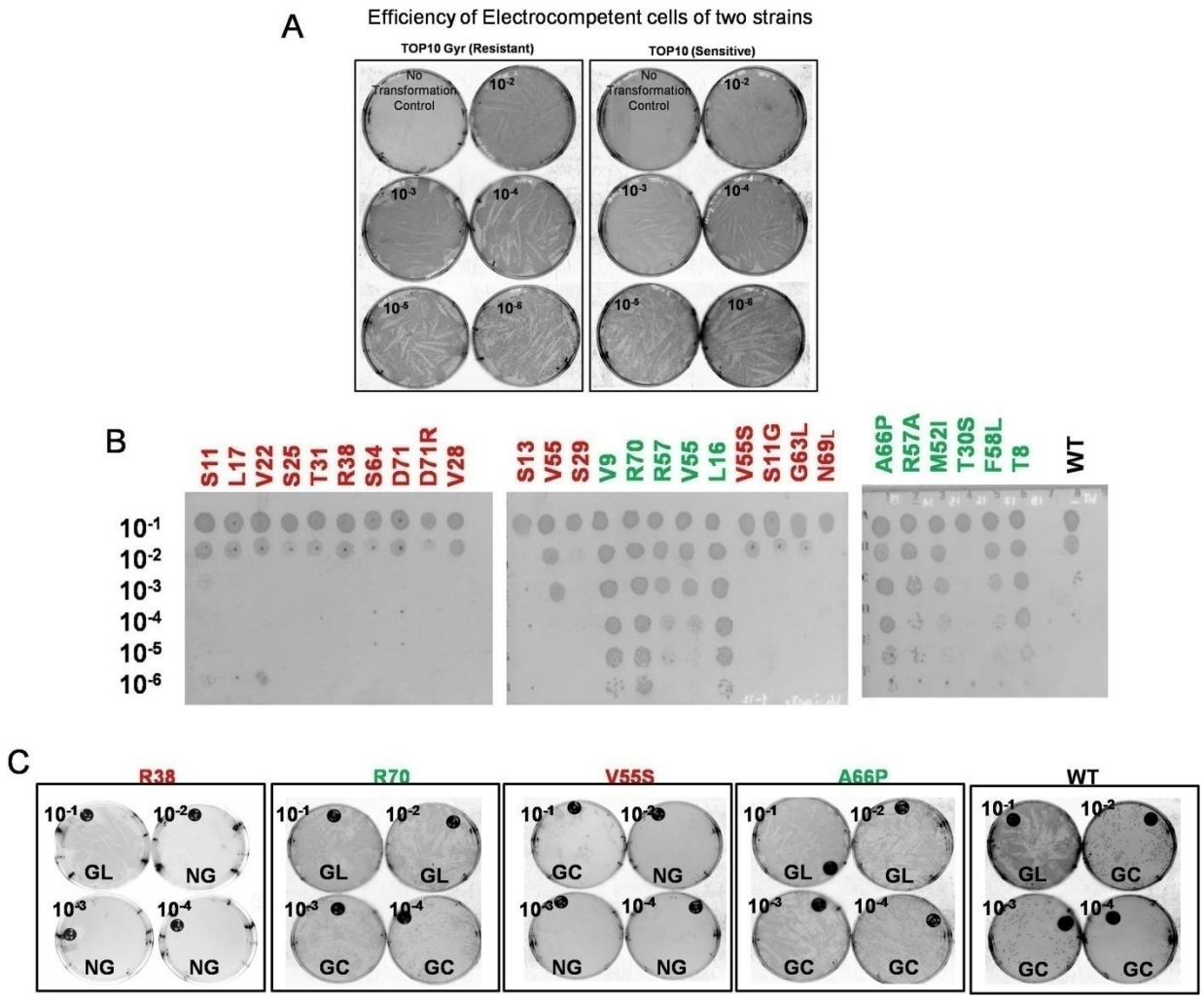


Figure S4. *In-vivo* activities of library and individual mutants of CcdA.(A) Transformation with empty pUC19 control vector shows that the efficiency of the electro-competent cells of the two strains, Top10 Gyr and Top10 is comparable. (B) Growth of single CcdA mutants in pUC57 operonic construct are individually probed, 40 ng of each mutant and WT was transformed in the Top10 sensitive strain and 3µl of 10-fold serial dilutions (10^{-1} to 10^{-6}) of the transformation mix were spotted on LB-Amp agar plates and grown overnight at 37˚C (in two replicates, one is shown here).The mutants labeled in red are inactive based on their deep sequencing phenotype (RFS_{seq} values $\leq$0.1), and mutants labeled in green are active (RFS_{seq} values >0.1). Plate and deep sequencing data are consistent with each other except in case of S11 and T30S. All the

mutants along with the mutated codon and phenotypic information are listed in Table 1. (C) 50µl of serial dilutions 10^{-1} to 10^{-4} of transformation mix of all mutants were plated on individual LB-amp agar plates The growth of four mutants is shown, namely an inactive synonymous, active synonymous, inactive non-synonymous and active non-synonymous mutant, as well as the WT construct, on different dilution plates. GL, GC and NG refer to Growth (Lawn), Growth (Countable colonies) and No Growth respectively to describe the growth patterns of mutants on dilution plates. WT shows observable growth at all dilutions (10^{-1} to 10^{-4}), while inactive mutants fail to grow at the higher dilutions.

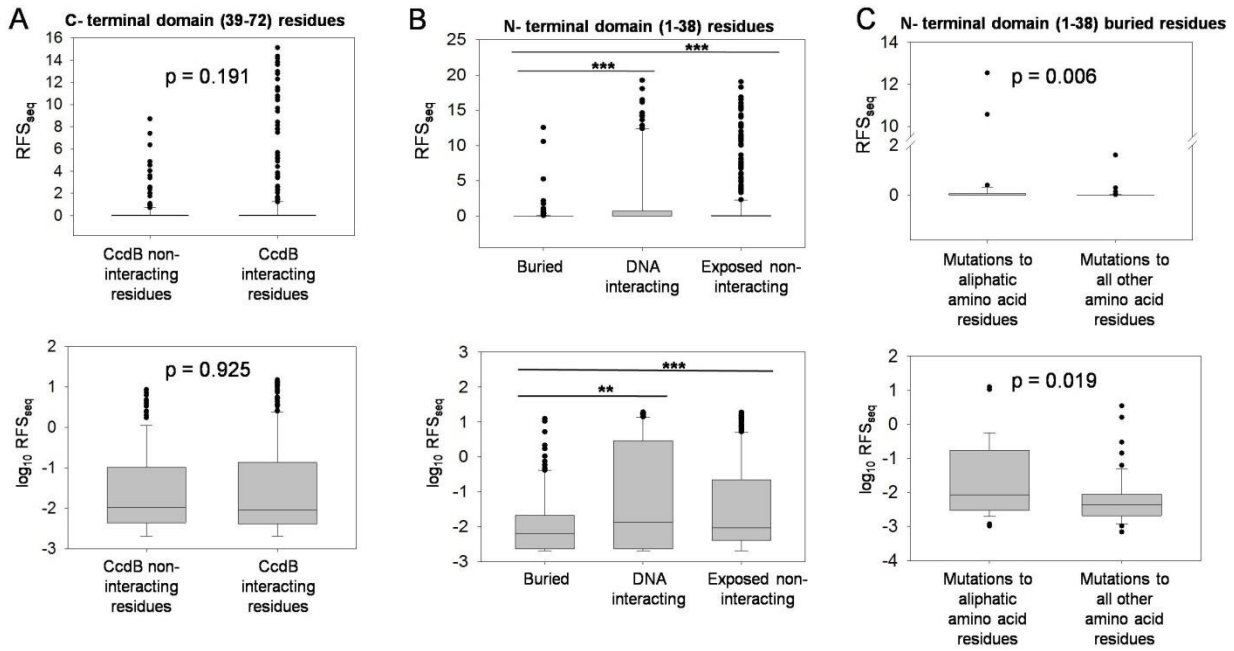


Figure S5. Buried residues in the structured N-terminal domain have higher mutational sensitivity than all other residues. Distribution of RFS_{seq} values for all *ccdA* mutants investigated by deep sequencing was studied to understand the role of different classes of functional residues. (A)Box plots of distribution of RFS_{seq} values in absolute(top) and logarithmic scale (bottom) for CcdB interacting and CcdB non-interacting residues in the disordered C-terminal domain, classified based on the available structure of CcdB bound to the CcdA C-terminal domain (PDB ID : 3G7Z). There was no statistically significant difference between phenotypes of mutations at CcdB interacting and non-interacting sites. (B) Box plots of distribution of RFS_{seq} values in absolute scale (top) and logarithmic scale (bottom) for different N-terminal domain residue classes. The classification of residues was done based on the available structure of the dimeric N-terminal domain of CcdA bound to DNA of the operator/promoter region of the *ccdAB* operon (PDB ID : 2H3C). Mutations at buried residues have considerably lower distribution of RFS_{seq} values (*** means p value < 0.001 and ** means p value < 0.01) than mutations at any other class of residues. (C) Box plots of distribution of

RFS_{seq} values in absolute (top) and logarithmic scale (bottom) for the aliphatic (A, C, L, I, M and V) and all other amino acid substitutions in the N-terminal domain buried residues. The substitutions to all non-aliphatic residues, at the N-terminal domain buried positions show significantly lower RFS_{seq} values than substitutions to aliphatic residues, leading to more drastic inactive phenotype. Statistical significance was assessed by a Mann-Whitney Rank Sum Test. A p-value > 0.05 indicates that there is no statistically significant difference between the distributions of the indicated parameter for active and inactive mutants, at the 95% confidence interval.

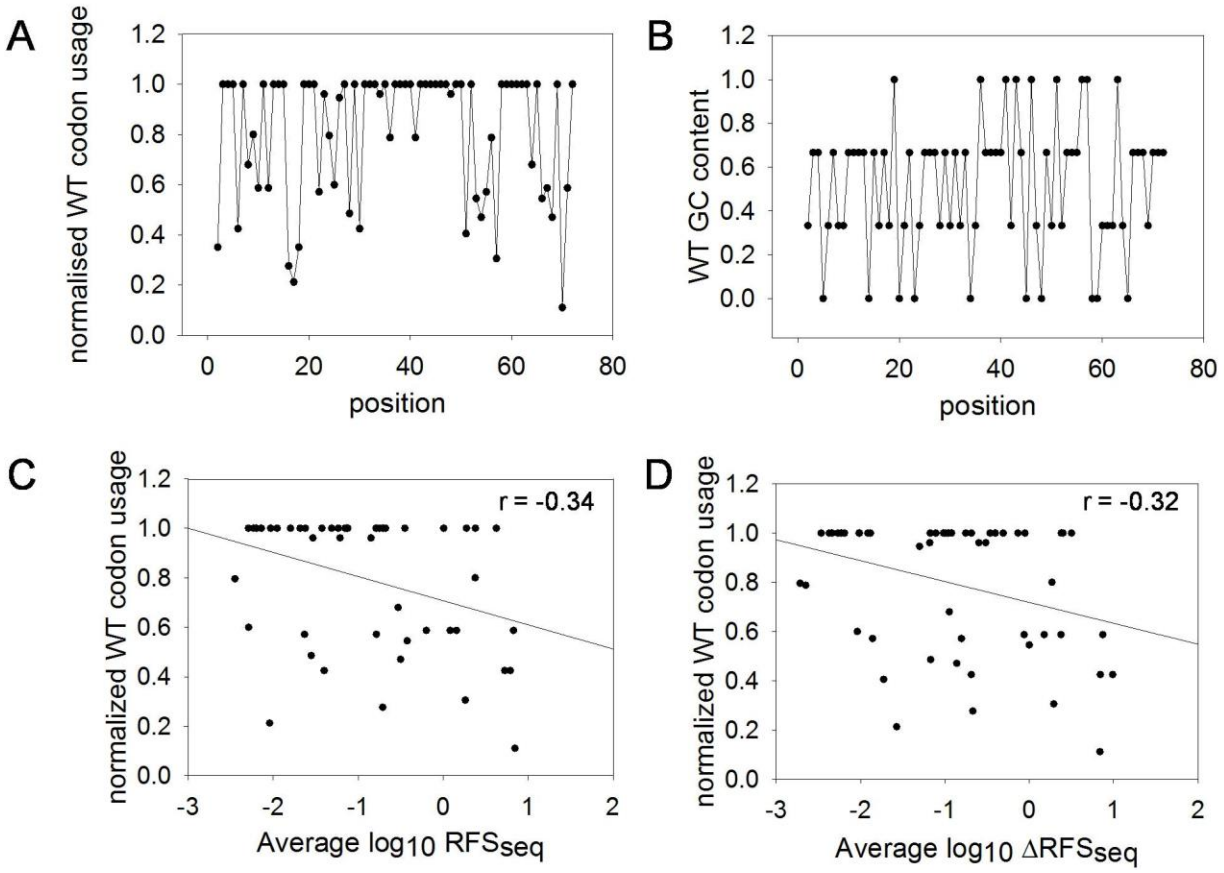


Figure S6. Unequal distribution of inactive phenotypes of mutants across gene length can be partially explained by the non-uniform distribution of rare and preferred codons in the *ccdA* gene. (A) Normalized WT codon usage was calculated by dividing the *E. coli* codon usage of the WT codon by codon usage of the most common codon. A WT codon with normalized codon usage value of 1indicatesthat the WT codon is the most optimal or commonly used in the *E. coli* genome. Interestingly, the central region of CcdA has high numbers of optimal codons, while rarer codons populate the extremities of the *ccdA* gene. (B) GC content of WT codons of the CcdA gene as a function of residue positions. (C) Normalized codon usage values for each residue in the WT *ccdA* gene show modest negative correlation with corresponding position average $\log_{10}RFS_{seq}$. This indicates that residue positions bearing optimal WT codons are more likely to produce inactive phenotypes upon mutations. (D) Normalized codon usage values for

residue positions in WT *ccdA* gene also show a modest negative correlation with average $\log_{10}\Delta RFS_{seq}^{codon}$ values (calculated as described in Figure 2C), indicating that there is larger variation in phenotypes among degenerate codons at residues positions bearing non-optimal WT codons.

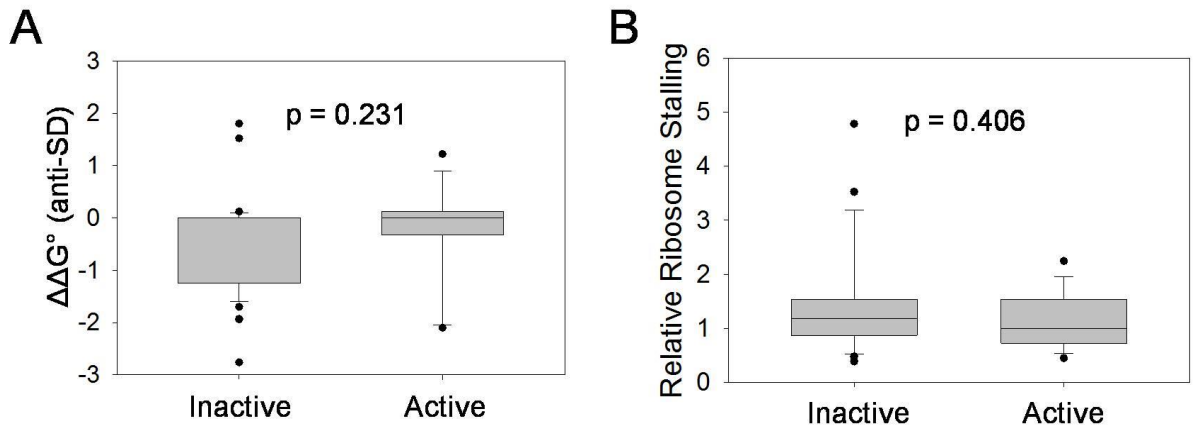


Figure S7. Effects of potential ribosomal pause sites on phenotype of CcdA synonymous mutations. (A) Distribution of differences in anti-SD sequence interaction energies (in kcal) amongst inactive and active synonymous mutants. The change in binding energy to the anti-Shine Dalgarno sequence on the ribosome in a mutant with respect to the WT sequence was calculated for a window of 10 bases using the RNAsubopt program in the Vienna RNA package. (B) Distribution of codon specific relative ribosome stalling[48]amongst inactive and active synonymous mutants. Statistical significance was assessed by a Mann-Whitney Rank Sum Test. A p-value > 0.05 indicates that there is no statistically significant difference between the distributions of the indicated parameter for active and inactive mutants, at the 95% confidence interval.