RESEARCH

Molecular bases for strong phenotypic effects of single synonymous codon substitutions in the *E. coli ccdB* toxin gene

Priyanka Bajaj[1,2], Munmun Bhasin[1] and Raghavan Varadarajan[1*]

Abstract

Background Single synonymous codon mutations typically have only minor or no effects on gene function. Here, we estimate the effects on cell growth of ~ 200 single synonymous codon mutations in an operonic context by mutating almost all positions of *ccdB*, the 101-residue long cytotoxin of the *ccdAB* Toxin-Antitoxin (TA) operon to most degenerate codons. Phenotypes were assayed by transforming the mutant library into CcdB sensitive and resistant *E. coli* strains, isolating plasmid pools, and subjecting them to deep sequencing. Since autoregulation is a hallmark of TA operons, phenotypes obtained for *ccdB* synonymous mutants after transformation in a RelE toxin reporter strain followed by deep sequencing provided information on the amount of CcdAB complex formed.

Results Synonymous mutations in the N-terminal region involved in translation initiation showed the strongest non-neutral phenotypic effects. We observe an interplay of numerous factors, namely, location of the codon, codon usage, t-RNA abundance, formation of anti-Shine Dalgarno sequences, predicted transcript secondary structure, and evolutionary conservation in determining phenotypic effects of *ccdB* synonymous mutations. Incorporation of an N-terminal, hyperactive synonymous mutation, in the background of the single synonymous codon mutant library sufficiently increased translation initiation, such that mutational effects on either folding or termination of translation became more apparent. Introduction of putative pause sites not only affects the translational rate, but might also alter the folding kinetics of the protein in vivo.

Conclusion In summary, the study provides novel insights into diverse mechanisms by which synonymous mutations modulate gene function. This information is useful in optimizing heterologous gene expression in *E. coli* and understanding the molecular bases for alteration in gene expression that arise due to synonymous mutations.

Keywords Operon, Evolution, Fitness, Elongation, Misfolding, Ribosome codon usage

*Correspondence:
Raghavan Varadarajan
varadar@iisc.ac.in
[1] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India
[2] Present address: Department of Bioengineering and Therapeutic Sciences, University of CA – San Francisco, San Francisco, CA 94158, USA

Introduction

Synonymous codon substitutions once thought to be genomic background noise, have now been widely acknowledged to have the capacity to alter protein expression, conformation and function [1]. However, it is still generally thought that single synonymous mutations that preserve the identity of the amino acid and do not alter the resulting protein sequence should have minimal or no effects on cellular function or organismal fitness. Nevertheless, in most sequenced genomes, synonymous codons are used with different frequencies

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 2 of 19

[2]. Codon usage bias is different for different organisms, varies across different genes and also specific loci between genes [3, 4]. Rare codons are often found at the N-terminal regions of ORFs in prokaryotes and eukaryotes [5]. One hypothesis that is consistent with this observation is that rare codons act as a 'ramp' to reduce translational velocity at the beginning of the gene, also known as the 'ramp hypothesis' [6]. Several studies suggest that poorly adapted codons at the N-terminus slow ribosome progression to increase the translational efficiency of the gene [6, 7]. Many studies also report that decreased mRNA structure at the N-terminus increases gene expression [8], however there have been conflicting studies as well [9].

Protein synthesis by ribosomes takes place at non-uniform rates on mRNA [10]. Synonymous mutations are known to alter gene expression by changing the translation rate via varied mechanisms such as codon usage bias [3], t-RNA abundance [11], and generation of an anti-Shine Dalgarno (aSD) sequence within the gene leading to ribosomal stalling [10]. Other ways include change in the mRNA structure due to alteration in base pairing [12], or altering the mRNA steady-state levels due to either change in mRNA synthesis or degradation levels [13]. However, typically multiple synonymous mutations are required for observable phenotypic effects.

In principle, translation is regulated at three different stages, initiation, elongation and termination. Translation initiation is a crucial step in protein biogenesis [14]. Finding the open reading frame (ORF) and ribosome loading on the mRNA takes place at the initiation step, which is the rate-limiting step and largely controls the frequency of translation of a certain mRNA [15, 16]. The translation efficiency of an mRNA, i.e., the amount of protein produced per unit mRNA is primarily determined by the accessibility of the ribosome binding site, nature of the start codon, occurrence of A/U rich codons disfavouring mRNA secondary structure in the beginning of the coding gene, position of the Shine-Dalgarno (SD) sequence relative to the start codon and its complementarity to 16S rRNA [15, 17, 18]. A growing body of evidence suggests that not only initiation, but also elongation plays a predominant role in controlling translation efficiency of the corresponding protein. During elongation, amino acids are added to the nascent chain one at a time. The elongation rate is non-uniform with periods of rapid movement separated by pauses [16, 19–21]. Translation may also be associated with a step-by-step folding process in which partial domain folding events may be required to ensure correct folding of the entire protein [22]. It is thought that some synonymous mutations can affect protein folding by affecting or targeting cotranslational folding processes [23, 24] that are altered by transient ribosome pausing [25]. Alterations in translation termination may be affected by the RNA structure at the end of the coding region. Translation termination can contribute to the production of altered protein isoforms by extending the C-terminal end due to translational read-through of a stop codon [26]. It is thus evident that translation efficiency of a gene is governed by the initiation, elongation and termination phases [27], but determining the relative contribution of each phase to protein abundance continues to be a challenging task.

In the context of operons, since there are two or more genes that are being translated, this adds an additional layer of complexity, as mutations can selectively affect translation efficiency of some genes of the operon relative to others [28, 29]. We used the *ccdAB* toxin-antitoxin (TA) operon as a model system to infer the effects of synonymous mutations on the expression and associated phenotypes of a toxin gene, that lead to altered fitness of the organism [30, 31]. The *ccd* operon from F-plasmid contains two genes, *ccdA* and *ccdB*, which encode for the homodimeric, labile antitoxin CcdA and the homodimeric stable toxin CcdB, respectively. The two proteins form a stable complex which in turn binds to the cognate *ccd* promoter and represses transcription. However, under conditions of cellular stress or plasmid loss, the labile CcdA antitoxin is degraded and CcdB binds to its cellular target, DNA Gyrase, compromising DNA replication and ultimately leading to cell death (Fig. S1). Since, autoregulation is a hallmark of many TA operons, the efficiency of complex formation determines whether the operon is being repressed or derepressed, which in turn dictates its in vivo transcriptional levels [30]. CcdB mutants can affect binding to CcdA, thereby altering CcdA-CcdB operonic regulation [32]. Either altered toxin:antitoxin ratio or structural changes exhibited by CcdB can modulate CcdB expression in cells [31, 33]. In the present work, we measure the fitness effects of single synonymous codon mutations spread throughout the entire *ccdB* gene. We attempt to address important issues such as 1) the relative codon-specific contribution to protein abundance for the initiation, elongation, and termination phases, 2) identification of the location of synonymous mutations that exhibit the largest phenotypic and codon-specific effects on protein synthesis, 3) understanding the molecular bases behind the observed phenotypic effects, 4) phenotypic effects of increasing translation initiation through addition of an N-terminal synonymous mutation to the existing single synonymous codon mutant library.

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 3 of 19

Results

Phenotypes of single synonymous codon mutant library of *ccdB* in its operonic context

A single synonymous codon mutant library of the globular cytotoxin gene, *ccdB* was made in the *ccd* operon. The *ccd* operon was cloned in pUC57 vector, a high copy number vector in order to get an amplified phenotypic response to distinguish mutant phenotype from WT [32]. Each position of *ccdB* was mutated to all possible degenerate codons via inverse PCR methodology [34]. All mutants were placed in identical regulatory contexts. The pooled synonymous mutant library was transformed in the CcdB resistant strain, Top10Gyr. The DNA recovered from this library was further transformed in the CcdB sensitive strain and RelE reporter strain, the latter strain is resistant to the action of CcdB and harbours a RelE reporter gene downstream of the *ccd* promoter containing the consensus Shine Dalgarno (SD) sequence [32]. Following transformation and plating, DNA was recovered from pooled transformants and subsequently deep sequenced. The fractional representation of each mutant in each condition was estimated and a good correlation between the two biological replicates of the resistant strain ($r=0.99$), sensitive strain ($r=0.97$), and RelE reporter stain ($r=0.99$) was observed when a threshold of a minimum of a 20 reads was taken in both replicates of the resistant strain as described previously [35] (Fig. 1A). Of 257 possible synonymous mutants, information for ~200 CcdB mutants were available in the resistant strain. Each synonymous mutant was assigned two variant scores, namely, Enrichment Score^{CcdB} (ES^{CcdB}) and Enrichment Score^{RelE} (ES^{RelE}), based on their phenotypic activity, i.e., cell growth versus cell death, which in turn is based on CcdB toxicity in the sensitive strain and RelE toxicity in the RelE reporter strain, respectively (Figs. S2 and S3) (see methods) [32]. ES^{CcdB} scores reflect free toxin protein levels in the cell. Higher levels of free toxin will result in decreased cell growth. Based on K-means clustering algorithm, a machine learning algorithm

used for partitioning a dataset into distinct, non-overlapping groups or clusters because of certain similarities, we classified synonymous mutations as hyperactive if ES$^{CcdB} < 0.7$, i.e., with a killing efficiency significantly higher than the WT and inactive if ES$^{CcdB} > 1.8$, i.e., with a killing efficiency significantly lower than the WT (Fig. S4). Throughout the manuscript we assume that *ccdA* translational efficiency is unaffected by synonymous mutations in *ccdB* and that [CcdA]$_{TOT}$ is proportional to the amount of *ccdAB* mRNA. This assumption can be justified by other studies which showed that synonymous mutations in a gene do not affect the expression levels of the upstream reporter gene in an operon [36] and that the secondary structures of mRNA in adjacent ORFs are independent of each other [29]. Therefore, any change in ES^{CcdB} directly reflects a change in translational efficiency of the *ccdB* gene as we define the translational efficiency as the amount of functional CcdB produced per mRNA per unit time. Phenotypes for 15 synonymous *ccdB* mutants inferred from deep sequencing were validated in a low-throughput manner by spotting culture dilutions of each mutant in both the resistant and the sensitive strains. A good correlation of $r=0.95$ was observed between the ES^{CcdB} scores and normalised CFU of mutants relative to WT, thereby validating the deep sequencing results (Fig. 1B). For instance, synonymous mutants with distinct phenotypes like T7_ACT, R13_AGG and R15_AGG having ES$^{CcdB} \sim 100$ fold lower than the WT exhibit growth only in the undiluted fraction, while mutations like P35_CCA, V33_GTC, R10_AGG and R10_CGA that have ES^{CcdB} at least twofold higher than the WT show visible growth even in the highest diluted fraction, i.e., 10,000 fold diluted.

To measure how efficiently the CcdAB complex is formed, the single synonymous codon mutant library was also transformed in the RelE reporter strain from which the mutants were screened based on RelE toxicity, indicated by their ES^{RelE} scores [32] (see Methods). The ES^{RelE} score for WT is 1 and due to the low dynamic range of

(See figure on next page.)

Fig. 1 Phenotypic effects of single synonymous codon mutants of CcdB in an operonic context. **A** Correlation between biological replicates. Following deep sequencing, the fraction of each mutant normalised to the WT value is estimated for all mutants in resistant (extreme left panel), sensitive (middle panel) and RelE reporter strains (extreme right panel). The (0,0) point represent the WT. **B** In vivo activity of CcdB mutants in both resistant (Top10Gyr) and sensitive (Top10) strains validate deep sequencing results in a low-throughput manner (left panel). Mutants are tenfold serially diluted and spotted on LB Agar plates. Corresponding ES^{CcdB} scores of individual mutants obtained from deep sequencing are mentioned in grey in the bottom panel. The mutants are ordered according to the residue number from right to left. A correlation of $r=0.96$ ($P < .001$) is observed between the ES^{CcdB} scores and normalised CFU count of mutants relative to WT in the sensitive (Top10) strain obtained in the spotting assay. Error bars for duplicate experiments are also shown (right panel). **C** Effects of synonymous mutations on ES^{CcdB} and ES^{RelE} scores as a function of location within the gene. Top panel shows position averaged ES^{CcdB} scores for synonymous mutants as a function of residue number. The bottom panel shows position averaged ES^{RelE} scores for synonymous mutants as a function of residue number. The N-terminal region shows the largest variation. Black dashed line represents the WT score. The entire data is divided into N-terminal (1–13), middle (14–86) and C-terminal (87–101) region. The secondary structure of CcdB (from PDB id 3VUB) is shown at the bottom
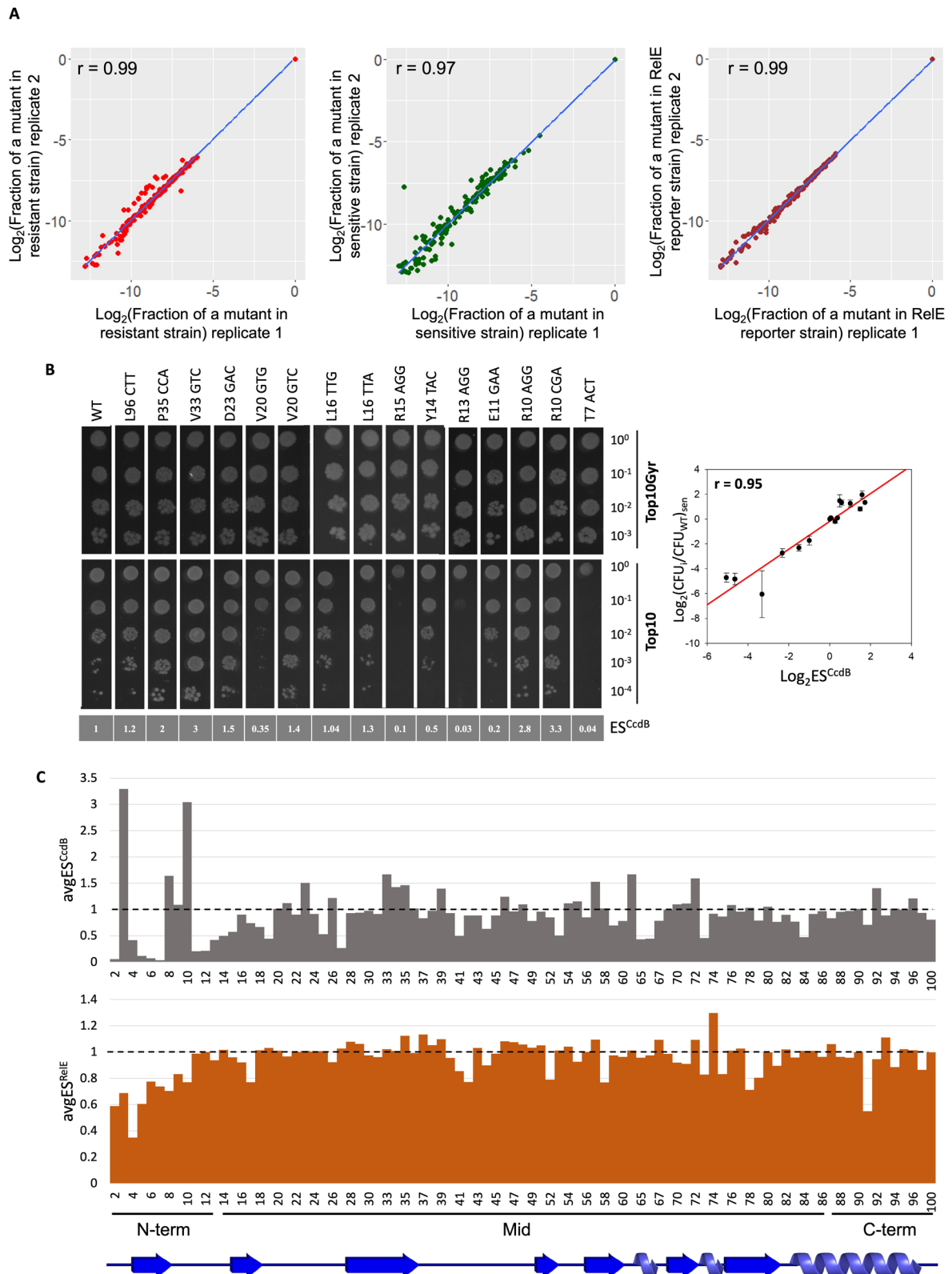
Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 4 of 19

Fig. 1 (See legend on previous page.)

ES^{RelE} scores, we classify mutants with $ES^{RelE} > 1$ as having a repressing phenotype and $ES^{RelE} < 1$ as having a derepressing phenotype [32]. The ES^{RelE} scores combined with the ES^{CcdB} scores provide insights into the molecular mechanisms responsible for the observed phenotypes for the synonymous mutations in the sensitive strain (Table 1). Based on these two screens, all synonymous mutations are classified into four mutational categories, namely, 1) Hyperactive and Derepressing denoted as 'H+D'; 2) Hyperactive and Repressing denoted as 'H+R'; 3) Inactive and Derepressing denoted as 'I+D'; and 4) Inactive and Repressing denoted as 'I+R', throughout the text.

Synonymous mutations in the N-terminal region of CcdB display large and diverse phenotypic effects

While most synonymous mutants showed a near neutral phenotype, i.e., similar to the WT, with ES^{CcdB} and ES^{RelE} scores close to 1, a significant number show altered phenotypes (Fig. 1C). Ribosome profiling studies show that ribosomes cover 20–30 bases at a stretch [37]. We examined codon specific contributions to protein abundance from three different parts of the gene, i.e., the N-terminal (residues 1–13), middle (residues 14–86) and the C-terminal (residues 87–100) regions of the *ccdB* gene. Values of ES^{CcdB} and ES^{RelE} averaged over synonymous mutations for each position were plotted to understand the overall trend. The most diverse ES^{CcdB} phenotypes relative to WT are displayed by the mutants of N-terminal region residues. The middle region shows the second highest variation in phenotypic effects. C-terminal amino acids show the least diverse phenotypic effects both

in the context of CcdB toxicity as well as RelE toxicity (Fig. 1C).

Estimating the importance of codon usage for N-terminal, middle and C-terminal region

We next examined correlations of various codon usage parameters, such as ΔGC content, RCU, CAI, RtrnaA, tAI and RCU_{dV} with ES^{CcdB}, ES^{RelE}, ES^{CcdB}_{dV} and ES^{RelE}_{dV} (see Methods section for parameter descriptions) for the entire dataset (Fig. S5A). As expected, different codon usage parameters (RCU, CAI, RtrnaA and tAI) are well correlated with each other. A positive correlation between ES^{CcdB} and codon usage parameters show that in general, enhancement in translational efficiency of CcdB is associated with decreased codon optimality (Fig. S5A). To further understand how codon bias (RCU) impacts protein abundance (ES^{CcdB}) at different locations in the gene, we plotted a moving average of the two parameters over a sliding window of 5 mutants (Fig. 2A). The data showed a clear trend, i.e., ES^{CcdB} increased with increased RCU for N-terminal residues, indicating mutations to more frequently used codons decreased CcdB levels in the cell. However, for middle and C-terminal residues, this was not the case. Effects of codon usage are largest at the N-terminus, compared to the rest of the gene (Fig. 2A).

63% (15/24) of the N-terminal region mutants belong to the 'H+D' class. Amongst these, synonymous mutations to rarer codons typically display extremely low ES^{CcdB} and ES^{RelE} scores. 21% (5/24) of the N-terminal region mutants belong to the 'I+D' class. Several of these synonymous mutations are at the R10 residue.

Table 1 Possible molecular mechanisms for the observed phenotypes

Phenotypic category	Notation	Parameter values	Inference
Inactive and Derepressing	I+D	$ES^{CcdB} > 1.8$ and $ES^{RelE} < 1$	Misfolded protein
Inactive and Repressing	I+R	$ES^{CcdB} > 1.8$ and $ES^{RelE} > 1$	Reduced translational efficiency
Hyperactive and Derepressing	H+D	$ES^{CcdB} < 0.7$ and $ES^{RelE} < 1$	Well folded, increased translational efficiency
Hyperactive and Repressing	H+R	$ES^{CcdB} < 0.7$ and $ES^{RelE} > 1$	Faster folding kinetics in vivo or slightly increased translational efficiency

(See figure on next page.)

Fig. 2 Synonymous codon usage plays an important role in regulating gene expression. **A** Altered CcdB levels (ES^{CcdB}) as a function of change in codon usage (RCU) relative to WT upon synonymous substitution. Mutants in N-terminal, middle and C-terminal regions are shown in blue, green and pink, respectively. **B**-**C** Left panel shows the corresponding values of ES^{CcdB} (x-axis), ES^{RelE} (y-axis) and RCU (color bar) for each mutant of the (**B**) N-terminal region and (**C**) C-terminal region. WT score is represented by black lines. Right panel shows the correlation between experimentally determined relative fitness values with various sequence-based parameters for the (**B**) N-terminal region and (**C**) C-terminal region. **B**-**C** The right panel shows the Pearson's correlation coefficient values and their corresponding *P* values. Gradation of white to red colour reflects positive correlation whereas the gradation of white to blue color reflects negative correlation. Non-significant correlations, i.e., correlations with *P* > 0.05 are shown with a cross 'X' mark. Parameter definitions are in the Materials and Methods section
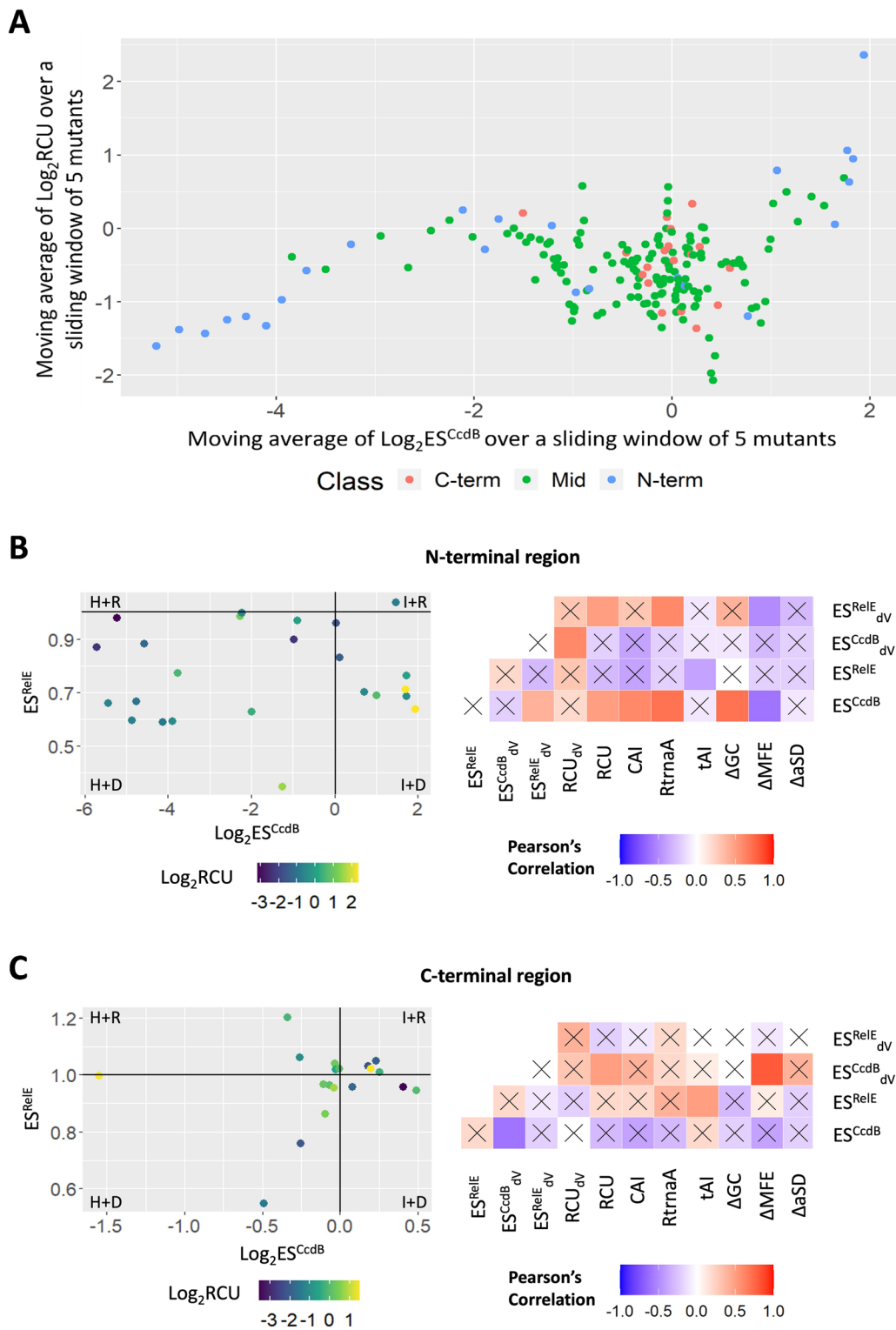
Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 6 of 19

Fig. 2 (See legend on previous page.)

Thus, in this case, it appears that a synonymous mutation from a rarer to a more frequent codon might result in misfolding of the protein in vivo (Fig. 2B). In the

N-terminal region, a positive correlation of ES^{CcdB} with RCU ($r = 0.54$, $P < 0.01$), CAI ($r = 0.64$, $P < 0.001$) and RtrnaA ($r = 0.71$, $P < 0.001$), as well as ES^{CcdB}_{dV} with RCU

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 7 of 19

$_{dV}$ ($r = 0.61$, $P < 0.001$), implies that introduction of rarer codon mutants at the N-terminus increases in vivo translational efficiency of the protein (Fig. 2B). In the present case, a positive correlation of ΔaSD with tAI ($r = 0.46$, $P < 0.05$) and a negative correlation with RCU$_{dV}$ ($r = $ -0.5, $P < 0.01$) (Fig. S5B), indicate that the introduction of rarer codons enhances the likelihood of formation of an aSD sequence, that in turn leads to stalling of the ribosome at the N-terminus. The data also might indicate that translational efficiency increases when the ribosome encounters a codon which has a significantly different codon usage than the WT. It will be interesting to see if this is a general feature across all genes.

In the middle region, ES^{CcdB} and ES^{RelE} show no correlation with codon usage parameters, suggesting that alterations in synonymous codon bias in the middle region does not contribute to altering the translation efficiency of CcdB (Fig. S5C).

In the C-terminal region, most mutants lie either in the 'H + D' or 'I + R' categories, implying that the phenotypes are primarily determined by translational efficiency (Fig. 2C). ES^{CcdB} is negatively correlated to codon usage (RCU, $r = $ -0.34, $P = 0.16$; CAI, $r = $ -0.33, $P = 0.06$), indicating that in contrast to the N-terminus, optimal codons at the C-terminus enhance translation efficiency of the gene in vivo, thereby increasing its protein abundance (Fig. 2C, Fig. S5D).

mRNA secondary structure dictates translation initiation

We next explored how predicted mRNA stability is correlated with observed phenotypes. For the prediction of secondary structure, a stretch beginning 18 bases prior to the start codon of the transcript was used. This included the putative Shine-Dalgarno (SD) sequence and the complete 306 base pair *ccdB* transcript for both the wild type and all single-site synonymous mutants. However, given that the entire ccd operon is 526 bp long, it is not possible to accurately predict its secondary structure. Also, since the same *ccdA* sequence is employed for all *ccdB* mutants, it is unlikely that mutations near the start codon of *ccdA* mRNA would exert distinct and differential effects on the mRNA structures of different *ccdB* mutants. Additional support for this assertion comes from previous work which suggests that structures of individual ORFs in an operon are relatively insulated from each other [29]. In the operonic mRNA, the transcribed regions of *ccdA* and *ccdB* that are most likely to interact are the 3' end of the *ccdA* that contains the *ccdB* RBS with the 5' region of *ccdB*. This region has been taken into consideration when predicting the structure of the *ccdB* mutant transcript.

In the N-terminal region, a negative correlation of ΔGC with ΔMFE ($r = $ -0.61, $P < 0.001$) indicates that as expected, higher GC content is associated with increased

mRNA stability (Fig. S5B). ES^{CcdB} is positively correlated with ΔGC ($r = 0.7$, $P < 0.001$) and negatively correlated with ΔMFE ($r = $ -0.59, $P < 0.001$) (Fig. 2B), suggesting that stable mRNA might reduce translation initiation, for example by occluding the RBS, whereas absence of structure might enhance RBS binding to the anti-SD sequence of the ribosome [3, 38] or alternatively, the start codon may be more efficiently recognised by the initiator t-RNA [8]. Another important feature is the accessibility of the SD sequence for interaction with the anti-SD sequence on the ribosome and how this might be modulated by synonymous mutations. We therefore predicted the secondary structure of a 59 base stretch starting at a $(U)_4$ stretch, 16 bases upstream of the start of the SD sequence and terminating at an $(A)_3$AGA sequence (residue R10) for the various mutants in the N-terminal region (Fig. S6). We observed an overall trend wherein occlusion of the RBS was associated with a higher value of ES^{CcdB}.

Evolutionary pressure drives codon selection

We also estimated evolutionary conservation for the WT *ccdB* gene both at the nucleotide and residue level. We found 62% (63/101) of the WT *ccdB* gene codons were the most conserved codons, 3% (3/101) were the least conserved codons, while the remaining 35% (35/101) have an intermediate conservation level (Fig. 3A). We further evaluated the evolutionary conservation levels of the synonymous mutant codons and compared it with their fitness effects. Largely, synonymous mutations to the most evolutionarily conserved synonymous codon display a hyperactive phenotype (Fig. 3B), indicating that synonymous codons are under selection pressure and that these mutant codons were not selected during the course of evolution in the context of the *E.coli* CcdB gene, because they will cause an increase in toxin levels leading to cell death.

We further investigated the impact of evolutionary conservation at the residue level, where each residue is assigned a conservation score. Interestingly, synonymous mutations at conserved residues in the N-terminal region are enriched in either the 'H + D' or 'I + D' categories, implying that as mentioned above, these mutations either increase *ccdB* translational efficiency and CcdB protein levels, or result in a folding defect in the protein, respectively (Fig. 3C).

In order to compare the codon usage between *ccdA* and *ccdB* in the wild-type context, the ratio of the CAIs, i.e. CAI(*ccdA*):CAI(*ccdB*) was calculated for *E. coli* and a few other bacterial species which contribute the bulk of *ccdAB* sequence (Fig. 3D-G). The CAI ratio was found to be higher than 1, consistent with a higher expressed protein ratio (CcdA/CcdB), a result in concordance with
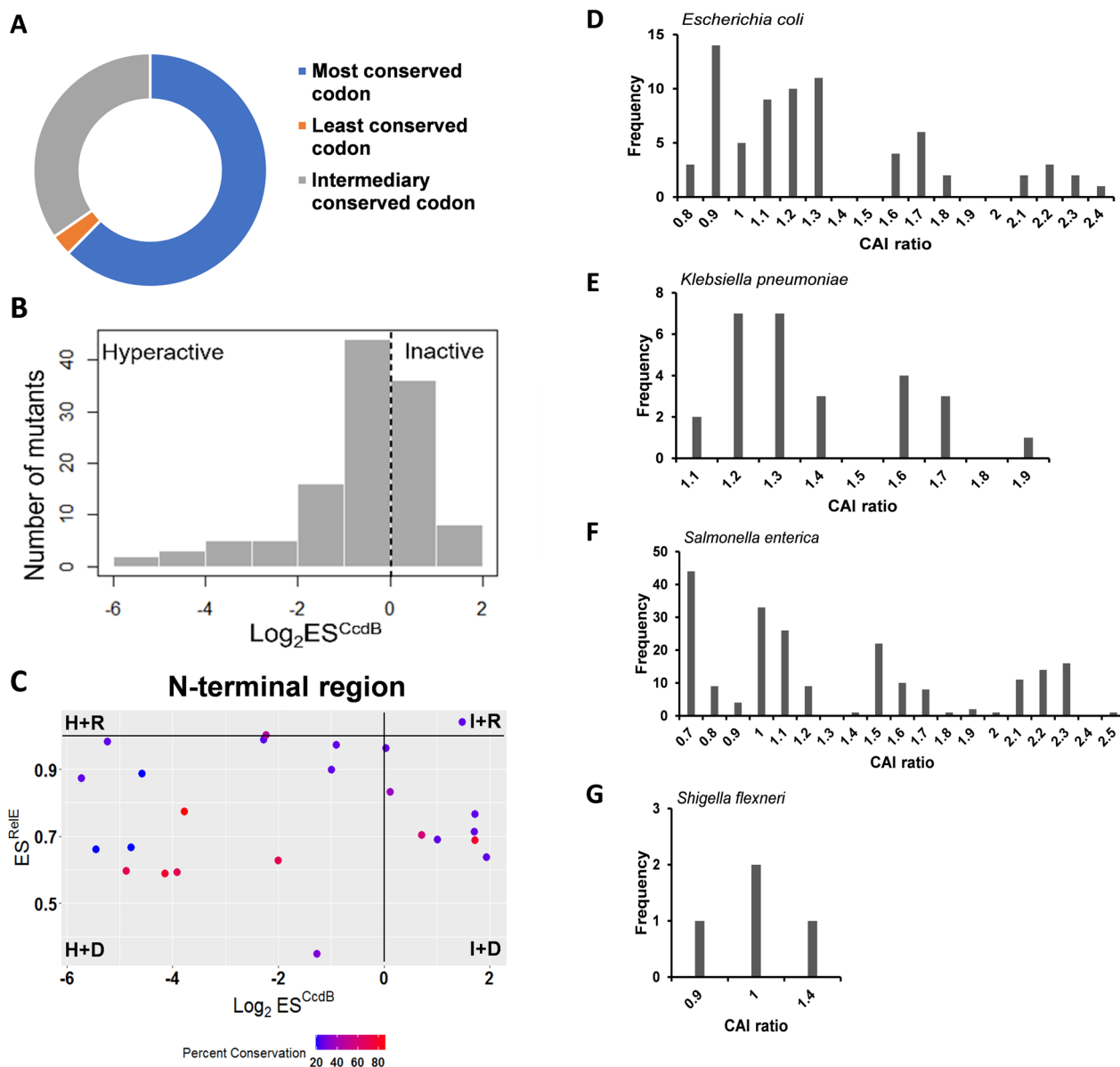
Fig. 3 Codon selection is under evolutionary pressure. **A** Percent distribution of WT codons in context of evolutionary conservation at the nucleotide level. **B** Some CcdB positions do not have the WT codon as the most conserved codon. Instead one of the synonymous counterparts is the most conserved codon. Histogram represents the frequency distribution of ES^{CcdB} scores for synonmous mutants with the most conserved codon for the corresponding WT amino acid. **C** Corresponding values of ES^{CcdB} (x-axis), ES^{RelE} (y-axis) and evolutionary conservation at the residue level (color bar) for each mutant of the N-terminal region is plotted. Percent conservation is proportional to the conservation level of the residue. Distribution of Codon Adaptation Index (CAI) ratios in wild-type (WT) *ccdA* and WT *ccdB* genes across different bacterial species: **D** *Escherichia coli*, **E** *Klebsiella pneumoniae*, **F** *Salmonella enterica*, and **G** *Shigella flexneri*. The CAI ratio provides insights into the codon usage of these genes within each respective species

experimental proteomics data for *E. coli* reported in a previous study [35].

Another factor that determines protein expression levels is the strength of the SD sequence. To characterize the binding between the putative SD sequence in the respective *ccdA* and *ccdB* genes with the 16S rRNA 3' ends

across different organisms, the interaction energies with the ribosome anti-Shine Dalgarno (anti-SD) sequence (5'CACCUCCU 3') were calculated. The results suggest a stronger interaction of the anti-SD with the SD sequence of the ccdA gene in *E. coli*, *K. pneumoniae* and *S. enterica*, whereas in the case of *S. flexneri* the anti-SD interacts

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 9 of 19

Table 2 Interaction energy values calculated using RNAsubopt between anti-SD sequence and SD sequence upstream of *ccdA* and *ccdB* gene across different bacterial species

	ccdA		*ccdB*	
	Putative SD sequence	MFE (kcal/mol)	Putative SD sequence	MFE (kcal/mol)
Escherichia coli	AAAGAGGT	-6.7	AGGGACTG	-3.7
Klebsiella pneumoniae	TCCGGAGT	-6.1	AGGGACTG	-3.7
Salmonella enterica	AAAGAGGT	-6.7	AGGAACTG	-3.7
Shigella flexneri	AGGTGTAA	-5.8	CGGAGCCT	-5.8
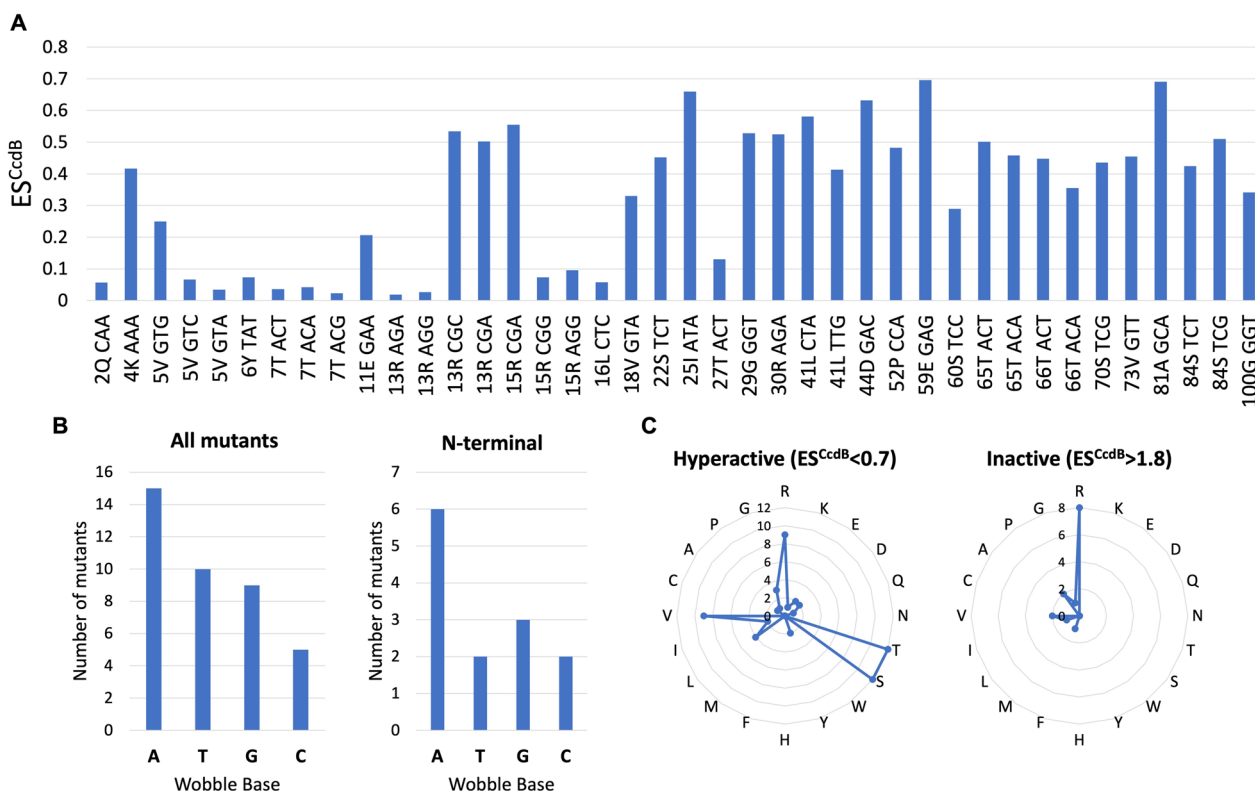
Fig. 4 Synonymous mutations enhancing translational efficiency of the *ccdB* gene. **A** ES^{CcdB} values of synonymous mutants belonging to 'H + D' class ($ES^{CcdB} < 0.7$ and $ES^{RelE} < 1$). **B** Enrichment of A/T at the wobble base for all synonymous mutants belonging to the 'H + D' class (left panel) and for the subset lying in the N-terminal region (right panel). **C** Synonymous codon mutational pattern for mutations with substantially altered ES^{CcdB} as a function of amino acid. Amino acid identity is shown at the circumference of the radar plot. Each circle represents the number of mutants (shown on the vertical axis) obtained for each amino acid, showing a hyperactive (left panel) or inactive (right panel) phenotype

equally well with the putative SD's of *ccdA* and *ccdB* (Table 2). Stronger binding of the ribosome to the SD is expected to lead to more efficient translation initiation.

Characteristics of synonymous mutations that enhance translational efficiency

The 'H + D' class ($ES^{CcdB} < 0.7$, $ES^{RelE} < 1$) of mutations in this study are associated with an increase in the $[CcdB]_{TOT}/[CcdA]_{TOT}$ ratio. These synonymous mutants likely result in enhanced translational efficiency of *ccdB*.

Synonymous mutations in the N-terminal region, especially the 5–13 residue stretch showed the lowest ES^{CcdB} scores, implying these synonymous mutations result in the maximum fold increase in translational efficiency. Altering the single synonymous codon in the N-terminal region (R13_AGA) resulted in an ~100-fold decrease in ES^{CcdB} (Fig. 4A).

Rare codons in *E.coli* generally end with A/T, and rare codons ending with A/T are known to correlate with increased expression compared to synonymous

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 10 of 19

mutations ending with G/C [11]. This observation is consistent with our results. This association also forms a link to the mRNA transcript secondary structure. If secondary structure is the dominant factor, we would expect a disproportionate enrichment of A over T due to G-U wobble base pairing. GU base pairing is known to be negatively associated with translation efficiency [39]. Indeed, nucleotide triplets with A in the wobble positions are enriched in the 'H + D' mutant class (Fig. 4B). Disproportionate enrichment of A over T is more prominent at the N-terminus (Fig. 4B). Rare codons in the N-terminal region with increased A/T content likely increase translation initiation, thereby increasing the efficiency with which the gene is translated.

Arginine displays the maximum phenotypic and codon specific effects

We analysed synonymous mutants giving either a hyperactive phenotype ($ES^{CcdB} < 0.7$) or an inactive phenotype ($ES^{CcdB} > 1.8$). Synonymous mutations of R, T, S and V are enriched in the former, while R alone is enriched in the latter class (Fig. 4C). R and S are encoded by six codons whereas V and T are encoded by four codons.

We also measured the variation of mutational phenotypes amongst the synonymous mutants of the same residue, by analysing the coefficient of variation of ES^{CcdB} scores for each CcdB residue. The coefficient of variation is highest for mutations in arginine, serine, and leucine, suggesting that amino acids with the maximum number of degenerate codons have the largest codon specific effects (Fig. S7A). Of the three, arginine showed the maximum phenotypic effects, depicted by diverse ES^{CcdB} scores (Fig. S7B), perhaps because it encodes part of an anti-SD sequence (AGG) which in turn will modulate the translational rate.

Single synonymous codon substitutions combined with an N-terminus hyperactive synonymous mutation display enhanced mutational sensitivity

We further incorporated an N-terminal hyperactive synonymous mutation (K4_AAG → K4_AAA) in the background of the existing single synonymous codon mutant library, therefore, generating an exhaustive double-site synonymous mutant library in the presence of a Parent Hyperactive Mutation (PHM). The ES^{CcdB} score of the single K4_AAA codon mutation is 0.42 relative to the K4_AAG WT sequence. The enhanced toxicity implies that a larger amount of CcdB is produced with K4_AAA relative to K4_AAG. This double-site synonymous mutant library was transformed in the three strains as described previously. A good correlation was observed between the two biological replicates in each of the resistant ($r=0.99$), sensitive ($r=0.72$) and RelE reporter strain ($r=0.99$) (Fig. 5A). The number of reads for most mutants in the sensitive strain was very low because of the enhanced toxicity displayed by the K4_AAA mutation due to which the correlation between the two biological replicates decreased. Of the ~257 possible mutants, we obtained information for ~160 mutants. Data analysis and assignment of mutant score in the form of ES'^{CcdB} and ES'^{RelE} for each mutant was done as described previously [32] (Figs. S2 and S3). Here, ES'^{CcdB} and ES'^{RelE} score for WT K4_AAA is 1. The ''' superscript indicates the scores are associated with the K4_AAA mutant library. There is an issue with the ES'^{RelE} scores because ES^{RelE} of the K4_AAA mutation in the WT *ccdB* gene background is 0.35 and is the minimum of the ES^{RelE} values (Fig. S3). Hence, in the double mutant library it is hard to get ES^{RelE} scores lower than this, a plausible reason why ES'^{RelE} scores for most mutants are generally ~1 or >1 (Fig. S3). Therefore, we do not interpret ES'^{RelE} values for the double mutant library. 15 mutants were individually constructed, and ES'^{CcdB} scores obtained for these mutants were validated by screening on plates.

(See figure on next page.)
Fig. 5 Phenotypic effects of single synonymous codon mutants in the background of an N-terminal hyperactive synonymous mutation, inferred through deep sequencing. **A** Correlations between the biological replicates. The fraction of each mutant normalised to the 'WT' value is estimated for all mutants in resistant (extreme left panel), sensitive (middle panel) and RelE reporter strains (extreme right panel). The (0,0) point represent the WT **B** In vivo activities of a few individual mutants to validate deep sequencing inferred phenotypes in a low-throughput manner, by transforming individual constructs in resistant (Top10Gyr) and sensitive (Top10) strains (left panel). These mutants are made in the background of WT K4_AAA. Mutants are tenfold serially diluted and spotted on LB Agar plates. ES'^{CcdB} scores for individual mutants obtained from deep sequencing are mentioned in the bottom panel. Y6_TAA and WTF' are used as controls. Y6_TAA is a stop codon mutation at the sixth position of CcdB, thereby making it non-functional. WTF' is the WT *ccdAB* operon sequence as present in the F plasmid. These two should grow equally well in both resistant and sensitive strains. A correlation of r = 0.95 (*P* < .001) is observed between the ES'^{CcdB} scores and normalised CFU count of mutants relative to WT in the sensitive (Top10) strain obtained in the spotting assay. Error bars for the duplicate experiments are also shown (right panel). **C** Effects of double-site synonymous mutations as a function of location in the gene. ES'^{CcdB} scores averaged over all mutants for each position. Black dashed line represents the WT K4_AAA score. As done previously, the entire data is divided into N-terminal (1–13), middle (14–86) and C-terminal (87–101) region
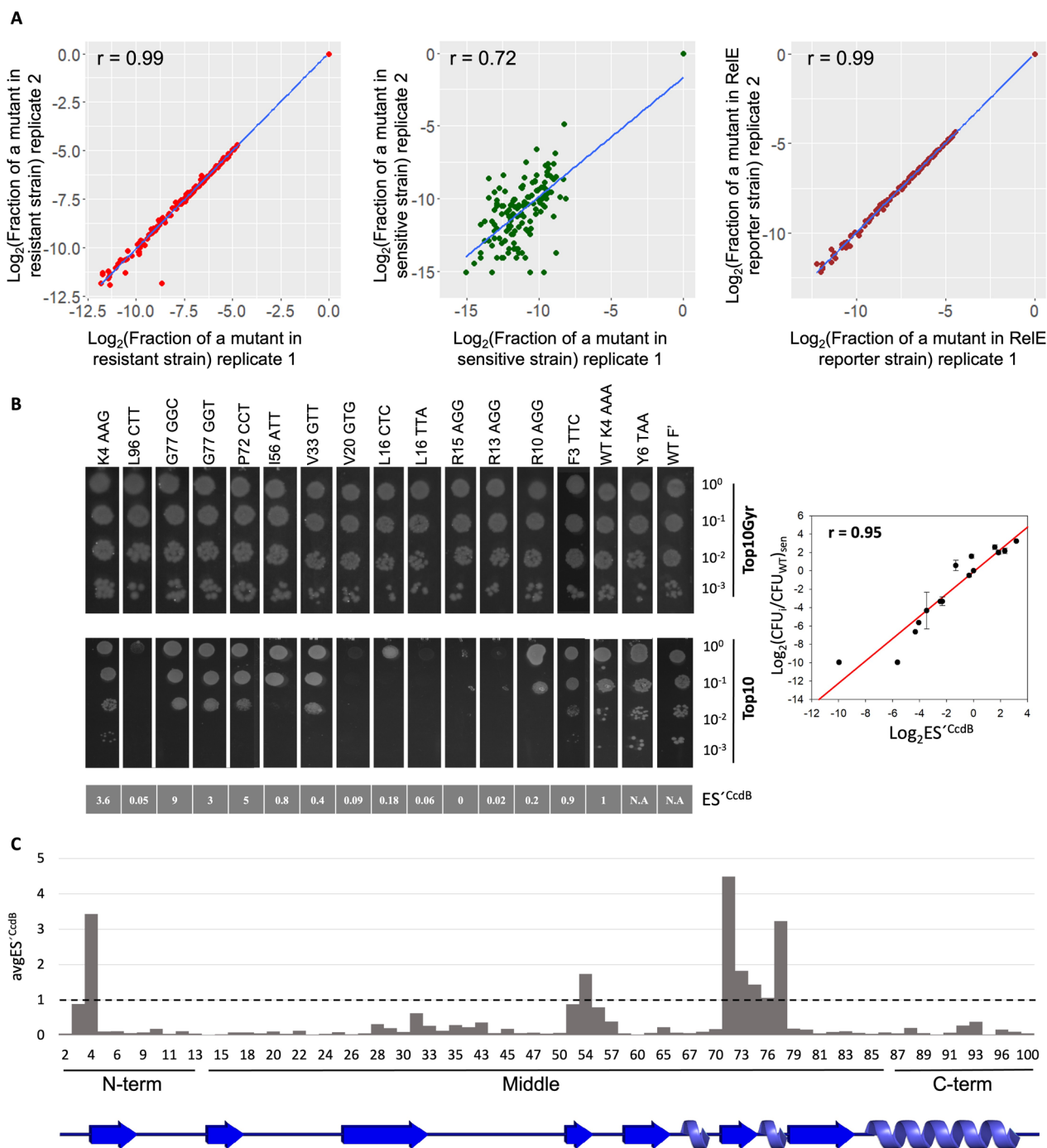
Fig. 5 (See legend on previous page.)

A good correlation of $r = 0.95$ was observed between ES'^{CcdB} obtained from deep sequencing data and normalised colony count of mutants relative to WT obtained from individually spotting dilutions of mutants transformed in the sensitive strain on plate (Fig. 5B).

We plotted $avgES'^{CcdB}$ scores for each position, that are obtained after taking an average of all the mutant scores for each position (Fig. 5C). ~82% (127/155) of the mutants displayed ES'^{CcdB} scores less than 0.2, i.e., ~ eightfold higher toxicity than the WT. We observed that synonymous mutations in specific stretches such as residues 52–57 and 72–77 showed higher values of ES'^{CcdB} than synonymous mutations in other locations. Here, in contrast to what was seen for the

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 12 of 19

single synonymous codon mutant library, we observe that mutants lying in the middle region of the gene show the most diverse phenotypes (Fig. 5C).

Reduced translational rate affects folding kinetics

A possible way by which such synonymous changes can lead either to different final structures or more likely enhanced yield, is by perturbing the protein folding pathway [40, 41]. This could occur for example, by a change in translation kinetics, which varies as a function of translation pause sites [26]. To probe if the synonymous mutations in CcdB could generate such potential ribosomal pause sites, the difference in interaction energies of the ribosome with the anti-SD (aSD) sequence associated with the synonymous mutations relative to WT, were calculated for a window of 10 nucleotides using the RNAsubopt program in the Vienna RNA package [42]. A correlation study was conducted separately for the N-terminal (Fig. S5A) and the remaining (middle and C-terminal) region (Fig. S5B). Consistent with the previous observation of the K4_AAG synonymous mutant library, reduced mRNA structure in the N-terminal region increases CcdB translational efficiency, depicted by positive correlation of ES$^{'CcdB}$ with ΔGC ($r=0.3$, $P=0.14$) and negative correlation with ΔMFE ($r=-0.4$, $P=0.07$). A slower progression of the ribosome at the beginning of the gene is shown by the negative correlation of ES$^{'CcdB}$ with ΔaSD ($r=-0.4$, $P=0.11$). In the middle and the C-terminal region, in contrast, a weak positive correlation of ES$^{'CcdB}$ with ΔaSD ($r=0.1$, $P=0.35$) suggests that translational pausing is accompanied by increased.

Gyrase binding activity, implying increased yield of properly folded protein likely by a process involving cotranslational folding. Negative correlation of ES$^{'CcdB}$ ($r=-0.2$, $P<0.05$) with ΔGC and positive correlation with ΔMFE ($r=0.2$, $P<0.05$) suggest that unstable mRNA results in diminished CcdB activity, possibly because loss of mRNA structure might lead to enhanced *ccdAB* mRNA degradation (Fig. 6A, S8).

Synonymous mutations in residue stretches 52–57 and 72–77 show a more inactive phenotype compared to the rest of the mutants (Fig. 5C), but do not result in formation of predicted translational pause sites (52–57: avgΔaSD=0.96, avgRtrnaA=1.23; 72–77: avgΔaSD=0.55, avgRtrnaA=1.4) (Table S2). From the positive ΔaSD values, it appears that the WT codons introduced translational pause sites which were abrogated upon synonymous mutation. These residues are part of a β-strand (52-57)and a short helical turn followed by a loop (72–77) respectively (Fig. 6B). The data suggest that the absence of predicted translational pause sites in these regions may cause the protein to misfold inside the cell.

We observed that the generation of aSD sequences and rarer codons due to mutations of residues lying in the α-helical regions (65–67, 73–75 and 84–99) [43] increased CcdB activity (avgES$^{'CcdB}$=0.25, avgRCU=0.86, avgΔaSD=-0.33) (Table S3). It is possible that mutations which reduce translation rate in fast folding structural elements like an α-helix, promote either a more stable protein conformation or a higher yield of properly folded protein. On analysing ES$^{'CcdB}$ and Hydropathy Index values for each CcdB residue as a function of position, we observed stretches 52–57 and 72–77 show positive HI values (Fig. 6C), suggesting that hydrophobic regions seem to have special translation kinetic requirements that ensure proper folding of the protein. These analyses suggest lack of translational pausing (high ΔaSD) due to synonymous mutations of hydrophobic residues (HI > 0) decreases the amount of properly folded CcdB protein (high ES$^{'CcdB}$). Therefore, occurrence of non-optimal codons and internal SD-like sequences in specific regions of the sequence as well as enhanced mRNA stability due to synonymous substitutions reduce the translation rate of the gene, that in turn enhances the yield of folded protein. These analysis are in agreement with a whole genome analysis of *E.coli* that shows an over-representation of non-optimal codons in alpha-helical signal peptides [44], Another study demonstrated that changing rare codons to optimal codons in signal peptides resulted in decreased protein expression [45].

Discussion

In its operonic context, single synonymous codon mutations in *ccdB* toxin display a wide variety of fitness effects. The consistency of our findings with other reports [46–49] validates our approach to measure the effects of synonymous substitutions. In most prior studies, multiple individual synonymous mutants need to be introduced to see an observable phenotypic effect [11, 20], complicating interpretation of the data. In contrast, in the present *ccd* system, we observe strong phenotypic effects with single synonymous substitutions. Here, we employ two different phenotypic readouts. In the CcdB sensitive Top10 strain, we measure ES^{CcdB} which is a measure of the amount of free CcdB toxin (uncomplexed to CcdA), a higher value of ES^{CcdB} implies a lower amount of free CcdB toxin. In the CcdB resistant strain, Top10GyrA, we measure ES^{RelE}, which is a measure of the amount of the CcdA:CcdB complex. A higher value of ES^{RelE} implies a higher amount of complex. Mutational effects are amplified because of the transcriptional autoregulation of the operon and use of a high copy number pUC plasmid. It is challenging to directly measure the extremely low levels of CcdA and CcdB proteins that are present upon
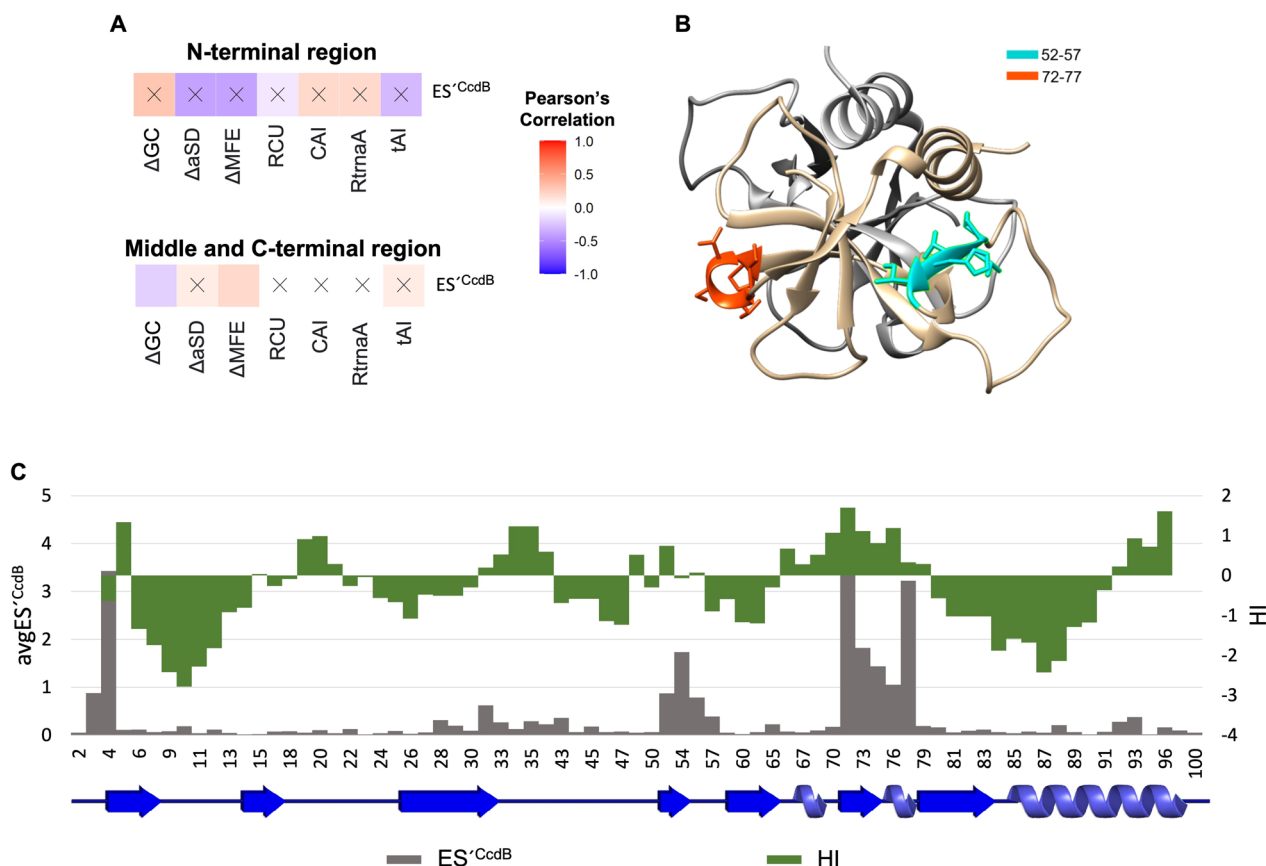
Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 13 of 19

Fig. 6 Molecular bases for observed phenotypes. **A** Correlations between experimentally derived mutational scores and sequence-based parameters for the double-site synonymous mutants. Top panel shows these correlations for the N-terminal region. Bottom panel shows these correlations for the middle and C-terminal region. White to red color gradation reflects positive correlation whereas white to blue color gradation reflects negative correlation. Non-significant correlations, i.e., correlations with $P > 0.05$ are shown with a cross 'X' mark. **B** Elucidating structural properties associated with distinct phenotypic effects. One monomer of CcdB is shown in light grey while residue stretches 52–57 (cyan) and 72–77 (orange) are mapped on the other monomer (tan) of the 3VUB crystal structure. The stretches are part of a β-strand and loop regions. **C** Comparison of the Hydropathy Index (green) and average ES′^{CcdB} score (grey) as a function of CcdB position. Hydrophobic residues shown with positive HI values tend to have a higher avgES′^{CcdB} score than the majority of the mutants

expression in the operon in vivo, using classical methods like SDS-PAGE or Western Blotting. We have previously used mass spectrometric methods to measure relative levels of CcdA and CcdB in a study where we explored the phenotypic effects of synonymous mutations in *ccdA* [35]. In that study, the results from a similar genetic screen were consistent with loss of function phenotypes being associated with a decreased CcdA:CcdB ratio in vivo measured through quantitative proteomics ([35] Fig. 5). However, this approach is quite laborious and can only be applied to a limited number of mutants. In previous studies where we examined phenotypic expression of a *ccdB* mutant library under control of the P$_{BAD}$ promoter we showed that for non-synonymous mutants, mutant activity phenotype was correlated with the amount of soluble protein, which in turn was correlated with the decrease in stability associated with the mutation ([50]

Fig. 3, [32] Figs. 2, and 5). Given these two prior validations, we are confident of the reliability of our inferences in the present study, which used two different phenotypic readouts to ascertain the molecular basis of the observed phenotypes associated with *ccdB* synonymous mutations in the CcdB sensitive strain Top 10.

A thorough understanding of effects of codon bias is central to fields as diverse as biotechnology and molecular evolution. Our results agree with other studies indicating that the initiation phase is the major contributing factor to translational efficiency. Several mutations in the N-terminal region result either in significantly decreased (residues 2, 4, 5, 6) or increased (residues 3, 8,9, 10) values of ES^{CcdB} (Fig. S2) corresponding to increased or decreased levels of CcdB protein. Modeling studies (Fig. S6) indicate that synonymous mutations in the N-terminal stretch likely affect the accessibility of the

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 14 of 19

ccdB Shine-Dalgarno sequence to ribosomes, however this needs to be confirmed by ribosome profiling experiments. Of note, an earlier (54-57)study found that the codon selection at the second position of the LacZ gene is determined by factors governing gene regulation at the initiation step of translation [51]. Stenström and Isaksson updated this study by measuring the effect of synonymous mutations from position 2 to 5 and reported that the mRNA base sequence in the early coding region of LacZ is the major determinant for the apparent efficiency of translation initiation and/or early elongation [52].

Analysis of [3] evolutionary conservation also provides insights into factors controlling gene expression and translation [53–55], especially at the N-terminus, likely because translation initiation plays a major role in modulating the rate and efficiency with which the gene will be translated. Although mutations in the N-terminal region show the largest diversity of phenotypes, several synonymous mutations in other regions, result in significant changes in fitness. This suggests that changes in elongation rate can also influence protein yield, although there was no clear pattern of association between relative codon preference and ES^{CcdB} values in the middle and C-terminal regions (Table S4).

The single synonymous codon mutant library when generated in the presence of an N-terminus hyperactive synonymous mutation enhanced mutational sensitivity. It is likely that the N-terminus hyperactive synonymous mutation sufficiently increased translational initiation rate, such that mutational effects in the elongation and termination phase can be more easily observed in this background. We did not observe large phenotypic effects for synonymous mutations in the C-terminal region of CcdB in this library also (low values of ES'^{CcdB}, Fig. S2) suggesting that synonymous mutations do not significantly affect termination. In future studies it would be interesting to examine effects of synonymous mutations at the C-terminal residue, 101 to see if these have a larger effect than synonymous mutations at other C-terminal proximal residues.

The choice of codons affects translation velocity, which in turn might affect the final conformation or amount of properly folded protein [56, 57]. In silico analyses revealed that introduction of potential pause sites in the middle of the gene through synonymous mutations, resulted in increased CcdB activity, likely by a process involving co-translational (sequential) folding. Our data suggest that synonymous mutations at different secondary structural elements likely alter translational rate which in turn alters the folding kinetics of the protein. From the results obtained through comparative studies of the mutational scores with Hydropathy Index parameters of the WT gene, we speculate that synonymous

mutations at hydrophobic residues in the 52–57 and 72–77 stretches (eg: 73V_GTT, $ES'^{CcdB} = 1.82$, RCU = 1.4, RtrnaA = 4.05, HI = 1.11) disrupt the highly synchronised protein folding process, thereby exposing hydrophobic patches that lead to misfolding, aggregation and lower levels of protein synthesis.

Changes in protein structure and function due to change in translation kinetics and cotranslational folding pathway have been observed for other proteins as well, such as chloramphenicol acetyltransferase [20, 40], suf1 [58], and *Echinococcus granulosus* fatty acid binding protein1 (EgFABP1) [41]. Another study reported that a silent mutation of the Ile codon AUC to a rare AUU in the coding sequence of the human MDR1 protein changes translation velocity and affects cotranslational folding. This results in a protein with altered conformation and affinity to its substrates [23]. Other studies show slowly translating codon clusters frequently occur at domain boundaries [59, 60], suggesting that translational pausing at rare codons may provide a time delay for optimal sequential folding at defined locations of the nascent polypeptide emerging from the ribosome. A systematic study on protein folding shows that cotranslational folding takes place under quasi-equilibrium conditions, provided translation is slower than folding [61].

We have earlier reported the effects of single synonymous codon mutations on the *ccdA* gene in its native operonic context [35]. In that study, synonymous mutations to rarer codons decreased translational efficiency of CcdA eventually leading to more cell death than the WT [35]. On the contrary, in the present study we observed synonymous mutations to rarer codons increase CcdB translational efficiency, prominently for the N-terminal region. In both cases variable effects were observed especially at the N-terminus. Due to the smaller length of *ccdA* gene (72 amino acids), it was not divided into three segments to study the effects of synonymous mutations on initiation, elongation and termination phases separately. The study, only examined effects of synonymous mutations in CcdA in an operonic context on cell survival. It was not possible to measure the effects of synonymous mutations on CcdA folding as CcdA is an intrinsically disordered protein. However, in the present study, in addition to CcdB toxicity assay, use of the RelE reporter assay helps clarify molecular mechanisms responsible for phenotypic effects seen for synonymous mutations in CcdB. In addition to highlighting molecular correlates of phenotypic changes associated with synonymous mutations, this study also outlines a novel approach to probe changes in co-translational folding and assembly associated with such single synonymous codon mutations. Such studies help to understand the molecular bases of alteration in gene expression and protein activity

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 15 of 19

arising due to synonymous mutations. It will be interesting to see if such phenotypic effects are observed in other systems. The challenge is to design sensitive readouts wherein small changes in protein activity, result in observable phenotypes. Further focus on development of strategies that can provide direct evidence of transient or permanent perturbations in protein structure arising due to synonymous mutations is also needed.

Conclusion

Synonymous mutations which do not change amino-acid identity, typically have only minor or no effects on gene function. Using sensitive genetic screens in the context of the *ccdAB* bacterial toxin-antitoxin operon, we demonstrate that many single synonymous codon mutations of the *ccdB* toxin gene display significant phenotypic effects in an operonic context. The largest effects were seen for synonymous mutations in the N-terminal region involved in translation initiation. Synonymous mutations that affected either folding or translation termination were also identified. Lack of translational pausing due to synonymous mutations in hydrophobic residue stretches, was found to decrease the amount of properly folded CcdB protein. Exploring the molecular determinants of these synonymous mutant phenotypes provides interesting insights into protein activity, folding, evolution as well as regulation of gene expression in bacteria.

Materials and methods

Plasmids and host strains

WT *ccdAB* operon cloned in pUC57 vector (pUCccd) is used as the starting plasmid for making the library of mutants. Two *E. coli* host strains were used, Top10Gyr, containing the GyrA462 mutation resistant to the activity of the CcdB toxin [32], and Top10 which is sensitive to the action of CcdB. These were used for phenotypic screening of CcdB synonymous mutants [32]. A third strain is a RelE reporter strain, namely Top10Gyr harbouring the pBT vector containing a RelE reporter gene downstream of the *ccd* promoter with a strong Ribosome Binding Site (RBS) formed by introducing a consensus SD sequence. The RelE reporter strain is sensitive to the level of the RelE toxin expressed from the *ccd* operon [32].

Generation of a *ccdB* single synonymous codon mutant library in its native operon

Mutagenic primers for all 100 positions (residues 2 to 101) of *ccdB* were designed such that the degenerate codons were at the 5' end of each 21 bp forward primer. The entire pUCccd vector backbone was amplified using non-overlapping adjacent 21 bp primers by inverse PCR methodology [34]. The primers were obtained in 96-well

format from the PAN Oligo facility at Stanford University. A master-mix containing Phusion DNA Polymerase was made for carrying out PCR for all positions in a 96 well format. Following densitometric quantification, an equal amount of PCR product (~ 200 ng) of each position was pooled. Gel-band purification of the pooled PCR product at the required size (~ 3.6 Kb) was done using a Fermentas GeneJET™ Gel Extraction Kit according to the manufacturer's instructions. After purification, pooled PCR product was phosphorylated by T4 PNK, followed by ligation with T4 DNA Ligase. The ligated product was transformed into high efficiency (10^9 CFU/µg of pUC57 plasmid DNA) electro-competent *E. coli* Top10Gyr cells, and subsequently plated on LB agar plates containing 100 µg/mL ampicillin for selection of transformants. Top10Gyr is referred to as the resistant strain in this study as it is resistant to the toxic activity of the CcdB toxin. Plates were incubated for 12–16 h at 37˚C. ~ 100 fold higher number of colonies than the expected library diversity (~ 200 mutants) were obtained. Pooled plasmid was purified using a Qiagen plasmid maxiprep kit.

Growth assay to screen mutants followed by preparation and isolation of barcoded PCR products for multiplexed deep sequencing

The single-synonymous codon master library was purified from Top10Gyr (resistant strain). The library was subsequently transformed and subjected to selection in both Top10 (sensitive strain) and Top10-Gyr harbouring the RelE reporter gene (RelE reporter strain) in two biological replicates. Pooled, purified plasmid samples from each condition were PCR amplified with primers containing a unique six base long Multiplex Identifier (MID) tag. 370-bp long PCR products containing the barcoded *ccdB* gene, were pooled, gel-band purified, and sequenced using Illumina Sequencing, on the Hi-seq 2500 platform at Macrogen, Korea. The WT *ccdAB* operon used in this study has a mutation in the putative SD sequence of *ccdA*, which in turn decreases *ccdA* expression, therefore allowing us to screen for both hyperactive and inactive mutants arising from CcdB mutations that confer increased or decreased toxicity relative to WT, respectively. 15 synonymous mutants were individually made in the same vector, followed by transformation in Top-10Gyr and Top10 for low-throughput validation of deep sequencing inferred phenotypes [32].

Data normalisation

The raw read numbers for the *ccdB* synonymous mutant library in pUC57 vector were normalised to the total number of reads in each condition. This gave an estimate of the fraction of each mutant represented in that condition. Read numbers for all mutants at all 101 positions

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 16 of 19

(1–101) in CcdB were analysed. Mutants having less than 20 reads in the resistant strain were filtered out prior to subsequent analysis. The rationale for using this read cut-off has been provided previously [35]. Two types of variant scores were assigned to each mutant in this study, one is Enrichment Score^{CcdB} (ES^{CcdB}) based on the CcdB toxicity readout while the other is Enrichment Score^{RelE} (ES^{RelE}) based on the RelE toxicity readout [32].

$$F(x_i) = \frac{x_i}{\sum x_i + x_{WT}} \quad (1)$$

$$F(y_i) = \frac{y_i}{\sum y_i + y_{WT}} \quad (2)$$

$$F(z_i) = \frac{z_i}{\sum z_i + z_{WT}} \quad (3)$$

Here, a given mutant is represented by 'i' whereas WT is represented by 'WT'. Number of reads in Top10Gyr resistant strain, Top10 sensitive strain and RelE reporter strain is represented by 'x', 'y' and 'z', respectively. $F(x_i)$, $F(y_i)$ and $F(z_i)$ are the fraction representation of a mutant in resistant, sensitive and RelE reporter strain, respectively.

$$Deepseq\ ratio_{sen_i} = \frac{F(y_i)}{F(x_i)} \quad (4)$$

$$Deepseq\ ratio_{RelE_i} = \frac{F(z_i)}{F(x_i)} \quad (5)$$

$$ES_i^{CcdB} = \frac{Deepseq\ ratio_{sen_i}}{Deepseq\ ratio_{sen_{WT}}} \quad (6)$$

$$ES_i^{RelE} = \frac{Deepseq\ ratio_{RelE_i}}{Deepseq\ ratio_{RelE_{WT}}} \quad (7)$$

For simplicity, these mutational scores are represented as ES^{CcdB} and ES^{RelE} for each mutant throughout the text. ES^{CcdB} and ES^{RelE} scores for the two biological replicates were estimated. An average of the scores from the two replicates for ES^{CcdB} and ES^{RelE} is taken for downstream analysis. The two variant scores are generally indicated in linear scale throughout the text. WT scores for ES^{CcdB} and ES^{RelE} are 1.

Defining codon usage parameters

$$\Delta GCcontent_i = GCcontent_i - GCcontent_{WT} \quad (8)$$

where GC content is the number of Guanine 'G' and Cytosine 'C' bases present in the mutant codon (GC content$_i$) and WT codon (GC content$_{WT}$), respectively.

$$Reltaive\ Codon\ Usage(RCU_i) = \frac{Codon\ Usage\ Frequency_i}{Codon\ Usage\ Frequency_{WT}} \quad (9)$$

Codon usage frequency values used in this study was taken from Genscript website. These values correlate well ($r = 0.98$) with the codon usage frequency values of *E.coli* K12 strain reported in a study [62]. Codon usage frequency is taken from the Bioinformatics tool of Genscript [63] (https://www.genscript.com/tools/codon-frequency-table). These codon usage values are also mentioned in Table S1. The term "rare" and "optimal" codons used in the text are based on the *E.coli* codon usage frequency table taken from the Genscript website (see also Table S1) which indicates how frequently the codon is used for an amino acid in the entire *E.coli* genome.

Codon adaptation index (CAI) is the similarity of codon usage to a reference set of highly expressed genes [64]. We calculated CAI for the synonymous mutants of the *ccdB* gene considering *Escherichia coli* (strain K12) as the selected target organism using the Java codon adaptation tool (JCat) [65].

$$Relative\ tRNA\ abundance(RtrnaA_i) = \frac{Fraction\ of\ tRNA\ out\ of\ total\ tRNA(\%)_i}{Fraction\ of\ tRNA\ out\ of\ total\ tRNA(\%)_{WT}} \quad (10)$$

'Fraction of tRNA out of total tRNA (%)' for each codon of *E.coli* is the fraction of tRNA out of the total tRNA population in *E. coli* [66]. A few degenerate codons are decoded by multiple tRNAs, for simplicity we assume each tRNA binds to its cognate codon with equal probability. Therefore, for such codons we summed the 'Fraction of tRNA out of total tRNA (%)' for its corresponding tRNAs and accordingly calculated the RtrnaA$_i$.

tRNA Adaptation Index (tAI) is defined as the similarity of codon usage to the relative copy numbers of tRNA genes. tAI computes a weight for each codon, based on the number of tRNAs available in the cell that recognize the codon, and the efficiency of the interaction between the different tRNAs and different codons [67]. The score of a coding region is the geometric mean of the weights of all its codons. In order to study the interaction efficiency between a tRNA and a specific codon for the CcdB synonymous mutants, the species-specific tAI (stAI) was calculated using stAI$_{calc}$ [68].

Defining parameters to estimate the extent to which CcdB levels and codon usage vary relative to WT

Let the CcdB synonymous mutant be denoted by 'i' and position 'j'.

If $ES^{CcdB}_i < 1$, then $x_i = 1/ES^{CcdB}_i$.

If $ES^{CcdB}_i > 1$, then $x_i = ES^{CcdB}_i$.

Let x_{minj} be minimum value of x_i for a given position 'j'. For a given mutant 'i' at a particular position 'j'

$$Degree\ of\ Variation\ of\ ES^{CcdB}(ES^{CcdB}_{dv})_{ij} = Log_2(x_{ij}/x_{minj}) \quad (11)$$

where x_{ij} is the value of x for mutant 'i' at position 'j'.

Similarly, if $ES^{RelE}_i < 1$, then $y_i = 1/ES^{RelE}_i$.

If $ES^{RelE}_i > 1$, then $y_i = ES^{RelE}_i$.

Let y_{minj} be minimum value of y_i for a given position 'j'. For a given mutant 'i' at a particular position 'j'

$$Degree\ of\ Variation\ of\ ES^{RelE}(ES^{RelE}_{dv})_{ij} = Log_2(y_{ij}/y_{minj}) \quad (12)$$

where y_{ij} is the value of y for mutant 'i' at position 'j'.

Similarly, if $RCU < 1$, then $z_i = 1/RCU_i$.

If $RCU_i > 1$, then $z_i = RCU_i$.

Let z_{minj} be the minimum value of z_i for a given position 'j'.

For a given mutant 'i' at a particular position 'j'

$$Degree\ of\ Variation\ of\ RCU\ (RCU_{dv})_{ij} = Log_2(z_{ij}/z_{minj}) \quad (13)$$

where z_{ij} is the value of z for mutant 'i' at position 'j'.

Calculation of interaction energies with ribosomal RNA for synonymous mutants in *ccd* mRNA

The interaction energies with the ribosome anti-Shine Dalgarno sequence (5'CACCUCCU 3') were calculated in the *ccd* mRNA to examine whether the synonymous mutations in the *ccdB* gene region led to generation of SD-like sequences. The difference in the interaction energy with the consensus anti-Shine Dalgarno (aSD) sequence between single synonymous mutants in CcdB and the WT sequence, was calculated for a window of ten nucleotides using the RNAsubopt program from the RNA Vienna package 2.4.18 [42].

$$\Delta aSD_i = avg(aSD_i - aSD_{WT}) \quad (14)$$

Since aSD values are negative, a positive value of ΔaSD_i indicates that the mutant shows lower translational pausing than WT.

Computational prediction of mRNA secondary structure

A stretch starting 18 bases upstream of the start codon of the transcript, consisting of the putative SD sequence along with the full length 306 bp *ccdB* transcript for WT and all single synonymous codon mutants was used for prediction of secondary structure by the RNAfold program of RNA Vienna package 2.4.18 [42]. The entire gene was divided into several segments based on a sliding window of 30 bases. Minimum free energy (MFE) values were computed for each segment, and further averaged over the sliding windows for each mutant. The program calculates the minimum free energy (MFE) structure and outputs the MFE structure and its free energy. We assume the mRNA structure of the toxin is unaffected by the mRNA structure of the preceding antitoxin [29].

$$\Delta MFE_i = avgMFE_i - avgMFE_{WT} \quad (15)$$

Nucleotide and protein sequence conservation analysis of *ccdB*

A BlastN search was performed for *WT ccdB* sequence, excluding the *E. coli* K12 strain (taxid:83,333) and other vector sequences with discontiguous megablast program optimized for more dissimilar sequences. The hits were filtered based on ≤95% identity. Query coverage was taken between 90 and 100. Sequences were trimmed from the ends such that they were in-frame with the WT sequence. This resulted in 1554 sequences which were aligned with CcdB WT sequence using Clustal Omega. This alignment was used as an input for finding the degree of conservation for each position in the CcdB nucleotide sequence. The nucleotide preference at each position was further transformed to the preference of codons at the amino acid level. The term 'conserved codons' throughout the text is is used based on the nucleotide conservation analysis done specifically for *ccdB* across different bacterial strains.

For protein sequences, BlastP was performed, excluding the *E. coli* K12 strain (taxid:83,333). The sequences were filtered based on their identity and query coverage using the same cut-offs as for nucleotide sequences. This resulted in 2883 sequences which were further aligned using Clustal Omega. This alignment was further used to find the residue conservation in the protein sequences.

Generation of a double site-synonymous mutant library of CcdB in its native operon

Primers were designed to mutate AAG to AAA at the 4th position in the single synonymous codon mutant library via 3-fragment recombination [69] in pUC57 vector. The K4_AAA mutant is considered as the WT gene for this part of the study, which we named as 'WT K4_AAA'. In this process, other mutations at the first ten positions were lost. Synonymous mutations to all possible degenerate codons were individually generated for these 10 positions via 3-fragment recombination using a Gibson assembly mix. Primers used for amplification of the fragments had 24 bp homology with the gene, and the mutation was present in the middle of the forward primer. These mutants were further pooled with the entire synonymous mutant library in

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 18 of 19

the same ratio as they would be expected to be represented in the library. After synthesising this library and isolating it following transformation in the resistant strain, it was transformed in the sensitive strain and the RelE reporter strain as described previously [32]. Following plating, plasmids from each library were isolated, deep sequenced, and assigned variant scores as ES´CcdB and ES´RelE as described previously [32]. After analysis of the data 14 individual mutants were selected, synthesised and individually transformed in Top10Gyr and Top10 for low-throughput validation.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09817-0.

Additional file 1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Rauscher R, Ignatova Z. Timing during translation matters: synonymous mutations in human pathologies influence protein folding and function. Biochem Soc Trans. 2018;46(4):937–44.
2. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 2011;12(1):32–42.
3. Bhattacharyya S, Jacobs WM, Adkar BV, Yan J, Zhang W, Shakhnovich EI. Accessibility of the Shine-Dalgarno sequence dictates N-terminal codon bias in E. coli. Mol Cell. 2018;70(5):894–905.
4. Muto A, Osawa S. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci U S A. 1987;84(1):166–9.
5. Eyre-walker A, Bulmer M. Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acids Res. 1993;21(19):4599–603.
6. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell. 2010;141(2):344–54.
7. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol. 2013;20(2):237–43.
8. Gu W, Zhou T, Wilke CO. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. PLoS Comput Biol. 2010;6(2):e1000664.
9. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. Science. 2009;324(5924):255–8.
10. Li GW, Oh E, Weissman JS. The anti-shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature. 2012;484(7395):538–41.
11. Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. Science. 2013;342(6157):475–9.
12. Kristofich J, Morgenthaler AB, Kinney WR, Ebmeier CC, Snyder DJ, Old WM, et al. Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. PLoS Genet. 2018;14(8):e1007615.
13. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. High guanine and cytosine content increases mRNA levels in mammalian cells. PLoS Biol. 2006;4(6):e180.
14. Milón P, Rodnina MV. Kinetic control of translation initiation in bacteria. Crit Rev Biochem Mol Biol. 2012;47(4):334–48.
15. Gualerzi CO, Pon CL. Initiation of mRNA translation in bacteria: structural and dynamic aspects. Cell Mol Life Sci. 2015;72(22):4341–67.
16. Tollerson R, Ibba M. Translational regulation of environmental adaptation in bacteria. J Biol Chem. 2020;295(30):10434–45.
17. Cifuentes-Goches JC, Hernández-Ancheyta L, Guarneros G, Oviedo N, Hernández-Sánchez J. Domains two and three of Escherichia coli ribosomal S1 protein confers 30S subunits a high affinity for downstream A/U-rich mRNAs. J Biochem. 2019;166(1):29–40.
18. Saito K, Green R, Buskirk AR. Translational initiation in E. coli occurs at the correct sites genome-wide in the absence of mRNA-rRNA base-pairing. Elife. 2020;9:e55002.
19. Rodnina MV. Translation in prokaryotes. Cold Spring Harb Perspect Biol. 2018;10(9):a032664.
20. Walsh IM, Bowman MA, Santarriaga IFS, Rodriguez A, Clark PL. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. Proc Natl Acad Sci. 2020;117(7):3528–34.
21. Chevance FFV, Le Guyon S, Hughes KT. The effects of codon context on in vivo translation speed. PLoS Genet. 2014;10(6): e1004392.
22. Frydman J. Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem. 2001;70(1):603–47.
23. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV. A"silent" polymorphism in the MDR1 gene changes substrate specificity. Science (1979). 2007;315(5811):525–8.
24. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. J Am Chem Soc. 2014;136(3):858–61.
25. Kramer G, Boehringer D, Ban N, Bukau B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat Struct Mol Biol. 2009;16(6):589–97.
26. Samatova E, Daberger J, Liutkute M, Rodnina MV. Translational control by ribosome pausing in bacteria: how a non-uniform pace of translation affects protein production and folding. Front Microbiol. 2020;11:619430.
27. Fredrick K, Ibba M. How the sequence of a gene can tune its translation. Cell. 2010;141(2):227–9.

Bajaj *et al. BMC Genomics*   (2023) 24:732

Page 19 of 19

28. Quax TEF, Wolf YI, Koehorst JJ, Wurtzel O, van der Oost R, Ran W, et al. Differential translation tunes uneven production of operon-encoded proteins. Cell Rep. 2013;4(5):938–44.

29. Burkhardt DH, Rouskin S, Zhang Y, Li GW, Weissman JS, Gross CA. Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. Elife. 2017;6:6.

30. De Jonge N, Garcia-Pino A, Buts L, Haesaerts S, Charlier D, Zangger K, et al. Rejuvenation of CcdB-Poisoned gyrase by an intrinsically disordered protein domain. Mol Cell. 2009;35(2):154–63.

31. Vandervelde A, Drobnak I, Hadži S, Sterckx YGJ, Welte T, De Greve H, et al. Molecular mechanism governing ratio-dependent transcription regulation in the ccdAB operon. Nucleic Acids Res. 2017;45(6):2937–50.

32. Bajaj P, Manjunath K, Varadarajan R. Structural and functional determinants inferred from deep mutational scans. Protein Sci. 2022;31(7):e4357.

33. Afif H, Allali N, Couturier M, Van Melderen L. The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. Mol Microbiol. 2001;41(1):73–82.

34. Jain PC, Varadarajan R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. Anal Biochem. 2014;449(1):90–8.

35. Chandra S, Gupta K, Khare S, Kohli P, Asok A, Mohan SV, et al. The high mutational sensitivity of ccdA antitoxin is linked to codon optimality. Agashe D, editor. Mol Biol Evol. 2022;39(10):msac187.

36. Lebeuf-Taylor E, McCloskey N, Bailey SF, Hinz A, Kassen R. The distribution of fitness effects among synonymous mutations in a gene under directional selection. Elife. 2019;8:e45952.

37. Ingolia NT. Ribosome footprint profiling of translation throughout the genome. Cell. 2016;165(1):22–33.

38. Lind PA, Andersson DI. Fitness costs of synonymous mutations in the rpsT gene can be compensated by restoring mRNA base pairing. PLoS One. 2013;8(5):e63373.

39. Chan S, Ch'ng JH, Wahlgren M, Thutkawkorapin J. Frequent GU wobble pairings reduce translation efficiency in Plasmodium Falciparum. Sci Rep. 2017;7(1):1–14.

40. Komar AA, Lesnik T, Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. FEBS Lett. 1999;462(3):387–91.

41. Cortazzo P, Cerveñansky C, Marín M, Reiss C, Ehrlich R, Deana A. Silent mutations affect in vivo protein folding in Escherichia coli. Biochem Biophys Res Commun. 2002;293(1):537–41.

42. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008;36(Web Server issue):W70-4.

43. Loris R, Dao-Thi MH, Bahassi EM, Van Melderen L, Poortmans F, Liddington R, et al. Crystal structure of CcdB, a topoisomerase poison from E. coli. J Mol Biol. 1999;285(4):1667–77.

44. Power PM, Jones RA, Beacham IR, Bucholtz C, Jennings MP. Whole genome analysis reveals a high incidence of non-optimal codons in secretory signal sequences of Escherichia coli. Biochem Biophys Res Commun. 2004;322(3):1038–44.

45. Zalucki YM, Beacham IR, Jennings MP. Biased codon usage in signal peptides: a role in protein export. Trends Microbiol. 2009;17(4):146–50.

46. Gingold H, Pilpel Y. Determinants of translation efficiency and accuracy. Mol Syst Biol. 2011;7(1):481.

47. Rodnina MV. The ribosome in action: tuning of translational efficiency and protein folding. Protein Sci. 2016;25(8):1390–406.

48. Choi J, Grosely R, Prabhakar A, Lapointe CP, Wang J, Puglisi JD. How messenger RNA and nascent chain sequences regulate translation elongation. Annu Rev Biochem. 2018;87(1):421–49.

49. Verma M, Choi J, Cottrell KA, Lavagnino Z, Thomas EN, Pavlovic-Djuranovic S, et al. A short translational ramp determines the efficiency of protein synthesis. Nat Commun. 2019;10(1):5774.

50. Tripathi A, Gupta K, Khare S, Jain PC, Patel S, Kumar P, et al. Molecular determinants of mutant phenotypes, inferred from saturation mutagenesis data. Mol Biol Evol. 2016;33(11):2960–75.

51. Looman AC, Bodlaender J, Comstock LJ, Eaton D, Jhurani P, de Boer HA, et al. Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in Escherichia coli. EMBO J. 1987;6(8):2489–92.

52. Stenström CM, Isaksson LA. Influences on translation initiation and early elongation by the messenger RNA region flanking the initiation codon at the 3' side. Gene. 2002;288(1–2):1–8.

53. Agashe D, Sane M, Phalnikar K, Diwan GD, Habibullah A, Martinez-Gomez NC, et al. Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. Mol Biol Evol. 2016;33(6):1542–53.

54. Bailey SF, Alonso Morales LA, Kassen R. Effects of synonymous mutations beyond codon bias: the evidence for adaptive synonymous substitutions from microbial evolution experiments. Genome Biol Evol. 2021;13(9):evab141.

55. Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. Nature. 2022;606(7915):725–31.

56. Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. J Mol Biol. 2008;383(2):281–91.

57. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human Disease. Nat Rev Genet. 2011;12(10):683–91.

58. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol. 2009;16(3):274–80.

59. Komar AA. A pause for thought along the co-translational folding pathway. Trends Biochem Sci. 2009;34(1):16–24.

60. Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, et al. Synonymous codons direct cotranslational folding toward different protein conformations. Mol Cell. 2016;61(3):341–51.

61. O'Brien EP, Ciryam P, Vendruscolo M, Dobson CM. Understanding the influence of codon translation rates on cotranslational protein folding. Acc Chem Res. 2014;47(5):1536–44.

62. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 2000;28(1):292.

63. Codon usage frequency table(chart)-genscript. Available from: https://www.genscript.com/tools/codon-frequency-table. Cited 2021 Sep 12

64. Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987;15(3):1281–95.

65. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 2005;33(suppl2):W526-531.

66. Dong H, Nilsson L, Kurland CG. Co-variation of trna abundance and codon usage inescherichia coliat different growth rates. J Mol Biol. 1996;260(5):649–63.

67. dos Reis M, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. 2004;32(17):5036–44.

68. Sabi R, Volvovitch Daniel R, Tuller T. stAIcalc: tRNA adaptation index calculator based on species-specific weights. Bioinformatics. 2017;33(4):589–91.

69. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343–5.

Publisher's Note