# Synonymous codon substitutions perturb co-translational protein folding and significantly impair cell fitness

Ian M. Walsh[a], Micayla A. Bowman[a] & Patricia L. Clark[a,b,1]

[a]Department of Chemistry & Biochemistry and
[b]Department of Chemical & Biomolecular Engineering,
University of Notre Dame, Notre Dame, IN 46556 USA

[1]*Correspondence:*
pclark1@nd.edu
01-574-631-8353

## Abstract

In the cell, proteins are synthesized from N- to C-terminus and begin to fold during translation. Co-translational folding mechanisms are therefore linked to elongation rate, which varies as a function of synonymous codon usage. However, synonymous codon substitutions can affect many distinct cellular processes, which has complicated attempts to deconvolve the extent to which synonymous codon usage promotes or frustrates proper protein folding *in vivo*. Although previous studies have shown that some synonymous changes can lead to different final structures, other substitutions will likely be more subtle, perturbing only the protein folding pathway without altering the final structure. Here we show that synonymous codon substitutions encoding a single essential enzyme lead to dramatically slower cell growth. These mutations do not prevent active enzyme formation but instead alter the folding mechanism, leading to enhanced degradation. These results support a model where synonymous codon substitutions can impair cell fitness by altering co-translational protein folding mechanisms. Synonymous codons changes can therefore have a significant impact on fitness even when the native structure is preserved.

## Significance

Many proteins are incapable of refolding *in vitro* yet fold efficiently to their native state in the cell. This suggests that more information than the amino acid sequence is required for proper folding of these proteins. Here we show that synonymous mRNA mutations can alter the protein folding mechanism *in vivo*, leading to changes in cellular fitness. This work highlights the important role of synonymous codon selection in supporting efficient protein production *in vivo*.

## Keywords

elongation rate, translation, ribosome, co-translational folding, folding pathway, folding intermediate, protease degradation, ClpXP, in vivo, protein design

## Introduction

Synonymous codon substitutions alter the mRNA coding sequence but preserve the encoded amino acid sequence. For this reason, these substitutions were historically considered to be phenotypically silent and often disregarded in studies of human genetic variation [1,2]. In recent years, however, it has become clear that synonymous substitutions can significantly alter protein function *in vivo*, through a wide variety of mechanisms that can change protein level [3-5], translational accuracy [6,7], secretion efficiency [8,9], the final folded structure [1,10-12] and post-translational modifications [13]. The full range of synonymous codon effects on protein production is, however, still emerging and much remains to be learned regarding the precise mechanisms that regulate these effects.

One mechanism that has long been proposed but has scant evidence to support its significance *in vivo* is the extent to which synonymous codon substitutions can perturb co-translational folding mechanisms. In general, rare synonymous codons tend to be translated more slowly than their common counterparts [14-17]. Moreover, rare synonymous codons tend to appear in clusters, creating broader patterns of codon usage [18], many of which are conserved through evolution [19-21]. The folding rates of many protein secondary and tertiary structural elements are similar to their rate of synthesis [22,23], lending conceptual support to the hypothesis that even subtle changes in elongation rate could alter folding mechanisms [24]. In theory, reducing the rate of translation elongation by synonymous common-to-rare codon substitutions could provide the N-terminal portion of a nascent protein with more time to adopt a stable tertiary structure before C-terminal portions are synthesized and emerge from the ribosome exit tunnel [25-27]. Depending on the specific native structure of the encoded protein, such extra time could be either advantageous or detrimental to efficient folding [28]. However, cells contain an extensive network of molecular chaperones to facilitate the folding of challenging protein structures, including several that associate with nascent polypeptide chains during translation [29-32]. Thus, it remains unclear whether a synonymous codon-derived alteration to elongation rate and co-translational folding mechanism could be sufficiently perturbative to rise above the buffering provided by the cellular chaperone network.

Here we show that position-specific synonymous codon changes in the coding sequence of an enzyme essential for *E. coli* growth can have a dramatic impact on growth rate. We tested a variety of mechanistic origins for this growth defect, including changes to the folded protein structure, expression level, enzymatic activity, mRNA abundance and/or activation of a cell stress response. Our results are consistent with synonymous substitutions altering the co-translational folding mechanism, leading to the formation of intermediates that are more susceptible to post-translational degradation. These results demonstrate that changes to synonymous codon usage can significantly affect protein folding *in vivo* and therefore have broad implications for both protein design and the interpretation of disease-associated synonymous mutations.

## Results

### *Synonymous codon substitutions impair* E. coli *growth rate*

To develop a system to test connections between synonymous codon usage, co-translational folding and cell fitness, we used chloramphenicol acetyltransferase (CAT), a water-soluble, homotrimeric *E. coli* enzyme with a complex tertiary structure [33] (**Fig. 1**). An early study showed that synonymous codon substitutions near the middle of the coding sequence (**Fig. 2a**) led to lower specific activity for CAT synthesized by *in vitro* translation [11]. CAT is

essential for *E. coli* growth in chloramphenicol (cam) [34], enabling us to use growth rate in the presence of cam as a convenient fitness assay. Furthermore, because CAT is not part of an operon or regulatory network, we hypothesized that it would be unlikely for feedback regulation of other genes to mask the effects of CAT synonymous codon changes on enzyme function [35]. Crucially, although CAT cannot be refolded to its native structure after dilution from chemical denaturants, the native structure is resistant to unfolding up to 80ºC (**Fig. S1**), highlighting the crucial role of *in vivo* folding intermediates on attainment of the native CAT structure.

We transformed *E. coli* with a plasmid encoding the previously described synonymous CAT coding sequence variant [11] under a titratable promoter but detected no discernable difference in growth versus *E. coli* producing CAT from the wild type (WT) coding sequence (**Fig. 2b, S2a**). Compared to WT-CAT, this synonymous construct contains a larger number of common codons (**Fig. 2a**), which leads to increased protein accumulation *in vitro* due to an overall faster translation elongation rate [11,16,25]. Consistent with these *in vitro* results, we detected more CAT in *E. coli* transformed with the coding sequence enriched in common codons (**Fig. 2c**). We hypothesized that this higher intracellular CAT concentration could mask a defect in specific activity. To test this, we used a Monte Carlo simulation method [18,36] (see Supplemental Methods) to design and select an alternative synonymous CAT coding sequence, Shuf1, which we predicted would be more likely to produce a WT-like amount of CAT. In the Shuf1 coding sequence, the local synonymous codon usage patterns are very different from WT but the global codon usage frequencies are very similar (**Fig. 2a, S3**). To avoid known effects of 5' synonymous codon substitutions on translation initiation [5,37-40], the first 46 codons of Shuf1 are identical to the WT coding sequence. *E. coli* produced CAT from the Shuf1 coding sequence at levels indistinguishable from WT-CAT (**Figure 2c**) but exhibited a significantly slower growth rate (**Fig. S2a**).

We hypothesized that we could further exacerbate the Shuf1-CAT growth defect by adapting a strategy developed by Hilvert and coworkers to couple subtle changes in enzyme function to *E. coli* growth rate [41]. This strategy involves encoding an ssrA tag at the C-terminus of the protein of interest, selectively enhancing its degradation by the protease ClpXP and leading to correspondingly lower intracellular protein concentrations. The smaller number of remaining enzymes face increased selective pressure to be highly functional. Addition of the ssrA tag did not affect CAT structure, stability or specific activity (**Fig. S2b-d**) but did lead to a dramatic growth defect for *E. coli* expressing Shuf1-CAT, versus the ssrA-tagged WT CAT coding sequence, in the presence of cam (**Fig. 2b**). This defect also led to a lower minimum inhibitory concentration for *E. coli* expressing Shuf1- versus WT-CAT (**Fig. S2e**).

*Shuf1 CAT mRNA, protein not toxic*

A major challenge of all *in vivo* experiments is discerning the precise origin of an observed effect. For example, a recent study showed that synonymous codon substitutions can lead to toxicity at the mRNA level even in the absence of protein production [42]. To test whether production of Shuf1-CAT mRNA and/or protein is inherently toxic, we compared growth rates of *E. coli* expressing WT or Shuf1-CAT in the absence of cam. These growth rates were indistinguishable (**Fig. 2d**), indicating that the Shuf1 defect is specifically related to impaired CAT enzyme function. Moreover, in the presence of cam the growth defect was partially suppressed at higher inducer concentrations (**Fig. 2b**), contrary to the larger growth defect expected if the Shuf1-CAT mRNA and/or protein were inherently toxic.

To test whether Shuf1-CAT induces a cell stress response, we used mass spectrometry to compare the abundances of 1277 proteins in *E. coli* expressing ssrA-tagged CAT from either the WT or Shuf1 coding sequence. There was no significant difference detected in the level of most proteins, including known stress-associated molecular chaperones and proteases (**Fig. 3a**). Taken together, these results support a model where the Shuf1-CAT growth defect is due to a direct loss of CAT protein activity, rather than an indirect effect on other cell functions.

*Native CAT proteins are indistinguishable*

Synonymous codon substitutions can lead to a wide range of effects on the encoded protein, including changes to translational fidelity (decoding accuracy) [6] and the native structure [1,10,12,17]. As a next test of the mechanism by which Shuf1 codon changes alter cell growth rate, we compared the solubility of CAT produced from the WT and Shuf1 coding sequences. In both cases, CAT was detected only in the soluble fraction of the cell lysate (**Fig. S4a**), indicating the Shuf1 growth defect is not due to CAT aggregation. Likewise, the purified CAT proteins produced from each mRNA sequence had indistinguishable secondary and tertiary structure (**Fig. S4b**), indistinguishable resistance to chemical and thermal denaturation (**Fig. 3b, S4c**) and indistinguishable specific activity (**Fig. 3c**). We also used mass spectrometry to compare the molecular weights of CAT translated from these coding sequences. These masses were indistinguishable to within one mass unit and matched the expected molecular weight of 25,953 Da. Taken together, these results demonstrate that CAT production from the Shuf1 coding sequence does not prohibit formation of the native, active CAT structure.

*Shuf1 coding sequence impairs native CAT protein production*

We noticed that addition of the ssrA tag led to a larger reduction in intracellular accumulation of CAT produced from the Shuf1 versus WT coding sequence (**Fig. 2c, S4a**). To determine whether this decrease in Shuf1-CAT was due to a defect arising from Shuf1 transcription and/or mRNA half-life, we quantified the levels of WT and Shuf1 mRNA. These levels were indistinguishable (**Fig. S5a**), supporting a model where the Shuf1 synonymous codon changes affect intracellular CAT concentration at the translational level, likely due to a greater susceptibility to protein degradation.

To explore this further, we subjected native, purified ssrA-tagged CAT produced *in vivo* from the WT or Shuf1 coding sequences to an *in vitro* ClpXP degradation assay [43,44]. Consistent with our other analyses of the native CAT structures and stabilities (above), the native proteins were equally resistant to degradation by ClpXP (**Fig. 4a**). This result suggests CAT produced from the Shuf1 mRNA sequence is more susceptible to degradation by ClpXP only prior to acquiring its native structure. Because the ssrA tag is located at the CAT C-terminus, this degradation is presumably post-translational, occurring after release of the nascent chain from the ribosome. To further test this model, we next tested (*i*) whether the Shuf1 growth defect was dependent on ClpXP activity *in vivo* and (*ii*) whether we could develop a physical mechanism for the impact of synonymous codon substitutions on Shuf1-CAT folding.

*ClpXP preferentially degrades Shuf1-CAT folding intermediates*

A prediction of the model described above is that the Shuf1 codon-dependent growth defect is exacerbated by post-translational degradation of ssrA-tagged CAT by cellular proteases, specifically ClpXP. ClpXP is the major *E. coli* protease responsible for degrading ssrA-tagged polypeptides under log-phase growth [43,45]. In general, less stably-folded proteins are more susceptible to degradation by ClpXP than more stable substrates [46-48], presumably

because less energy is required to unfold unstable protein structures and expose the chains to the ClpXP protease active sites [49]. To test whether ClpXP degradation is the key mechanism impairing growth when *E. coli* expresses CAT from the Shuf1 coding sequence, we induced expression of WT and Shuf1 CAT in *E. coli* W3110, which lacks ClpX [46,50], and compared growth in this background to the parent strain, in the presence of cam. ClpX deletion enhanced growth only of cells expressing ssrA-tagged CAT from the Shuf1 coding sequence (**Fig. 4b**). Likewise, omission of the ssrA tag enhanced growth only for *E. coli* expressing ClpX; there was no effect on cells lacking ClpX (**Fig. 4c**). These results confirm that the major effect of the Shuf1 synonymous codon substitutions is enhanced degradation of ssrA-tagged CAT by ClpXP.

*Shuf1 synonymous codon substitutions alter the CAT folding mechanism*

As the next step towards a physical mechanism for the impact of synonymous codon substitutions on Shuf1-CAT folding, we noticed that after five hours of high-level induction in the presence of cam the growth rate of cells expressing the Shuf1 mRNA sequence increased to match the growth rate of cells expressing the WT sequence (**Fig. 2b, right panel**; note similar slopes at induction times >5 hr). This increase in growth rate for *E. coli* transformed with the Shuf1 plasmid was not due to the appearance of suppressor mutations, as cells taken from the endpoints of these growth assays had a reproducible growth lag when subsequently diluted and grown under identical conditions (**Fig. S5b**). In all of the cell growth assays described above, *E. coli* were induced to express CAT and at the same time challenged with cam. We hypothesized that this simultaneous challenge to both produce CAT and acetylate the antibiotic might amplify the importance of rapidly producing a sufficient pool of native CAT, highlighting the defect created by the increased susceptibility of CAT folding intermediates to degradation. Further, we hypothesized that even if only a small fraction of CAT translated from the Shuf1 coding sequence attains its native fold, the protease resistance of the native CAT structure (**Fig. 4a**) will eventually lead to the accumulation of a pool of native CAT sufficient to support a WT-like growth rate, regardless of the precise folding mechanism.

A direct prediction of the hypothesis above is that providing cells with more time to accumulate native CAT prior to cam addition should reduce or eliminate the Shuf1 growth defect. To test this prediction, we modified our growth assay to induce CAT production in the overnight culture, then diluted cells into fresh growth medium in the presence of cam and inducer. Overnight induction was sufficient to suppress the Shuf1-CAT growth defect, but only when a high concentration of inducer was used (**Fig. 5**). Based on these results, we hypothesized that (1) high levels of induction enable more copies of Shuf1 CAT to fold to its native, active structure and (2) providing more time for CAT protein folding prior to cam addition can suppress the Shuf1 growth defect. In support of these hypotheses, we found that just 30 min of induction prior to cam addition was sufficient to suppress the Shuf1 growth defect, but only at high inducer conditions (**Fig. S5c**).

*mRNA secondary structural stability does not explain Shuf1 growth defect*

The results above suggest the Shuf1 synonymous codon substitutions impair CAT co-translational folding by altering the rate of translation elongation. *In vitro*, synonymous codons have been shown to alter elongation rate either by altering the rate of decoding [51] or by altering downstream mRNA stability, which can impede ribosome translocation [52]. *In vivo*, there is some evidence that stable mRNA stem-loop structures can alter the elongation rate of the ribosome [53-55], although other studies have detected no difference [37,56,57]. Although the overall predicted mRNA stability of the WT and Shuf1 genes are similar, a predicted stable 3' stem-loop structure in Shuf1 is not present in the WT coding sequence (**Fig. S6a**). To test

whether this structure is responsible for the Shuf1 growth defect, we created chimeric mRNA sequences with only the 5', middle or 3' portion of the wild type sequence substituted with the Shuf1 sequence (**Fig. S7a**) but observed no growth defect for the chimera bearing the 3' portion of Shuf1 had no impact on growth rate (**Fig. S7b**). Moreover, growth rates for these chimeras correlated more closely with the difference in relative codon usage frequencies than measures of mRNA stability (**Fig. S6b**). Taken together, these results indicate that translation elongation rate differences arising from changes in codon usage frequencies is a more likely origin of the Shuf1 growth defect than changes in mRNA secondary structure.

## Discussion

Most of our current understanding of protein folding mechanisms is derived from studies of small proteins that refold reversibly when diluted from chemical denaturants. However, only a small number of proteins can refold robustly *in vitro*, even though many more can be maintained in a stable state once extracted from the cell [24,58,59]. This suggests that (i) the conformations adopted early during the folding process are crucial to successful folding and (ii) the cellular environment supports the formation of early folding intermediates that are distinct from the conformations populated upon dilution from denaturant. Indeed, there is substantial evidence that molecular chaperones are crucial to the successful folding of many complex proteins *in vivo* [29-32]. Although it has been hypothesized that synonymous codon changes could alter elongation rate and modify folding mechanisms *in vivo*, it has thus far been challenging to find evidence to support this hypothesis from experiments performed *in vivo*, possibly due to buffering provided by molecular chaperones.

Our results demonstrate that, during synthesis, the folding of nascent CAT polypeptide chains is sensitive to synonymous codon-induced changes to translation elongation rate. Although in all cases the nascent chains produced using different synonymous codon patterns remain capable of achieving the native, protease-resistant CAT trimer structure, translation of the wild type mRNA sequence led to the formation of CAT folding intermediates that are less susceptible to degradation by cellular proteases once released from the ribosome (**Fig. 6**). Crucially, these results demonstrate that although the CAT native structure is indistinguishable regardless of synonymous codon usage, the folding mechanism differs significantly, leading to increased degradation and an adverse effect on cell fitness.

These results are consistent with a small but growing number of studies indicating that synonymous codon substitutions can perturb protein folding mechanisms [1,10,12,60] and highlight strategies for uncovering such perturbations even when they do not alter the final protein structure. In contrast, recent *in vitro* single molecule force-unfolding experiments have shown that some small, ribosome-bound natively-folded domains fold via similar mechanisms on and off the ribosome [61,62]. However, as these studies noted, forced unfolding measured by molecular tweezers cannot capture the transient folding of a nascent chain during its synthesis, and hence what is measured in these experiments is the effect of close proximity of the ribosome surface, rather than co-translational folding. The very robust folding behavior of these well-characterized, reversible folding models may indeed lead to indistinguishable folding behavior during translation, a model supported by recent force-feedback folding measurements [63]. Of note, the model proteins selected for these studies are smaller than >75% of proteins in the *E. coli* proteome [24], whereas all known examples of synonymous codon-derived alterations to co-translational folding are much larger (e.g., [1,9,10,64]). We are not aware of an *in vitro* folding mechanism for a protein >175 aa long that is preserved during co-translational folding.

Our CAT results demonstrate that synonymous changes to mRNA coding sequences can significantly perturb folding of the wild type protein sequence even in the presence of its repertoire of relevant molecular chaperones. This result suggests that mRNA sequences have likely evolved alongside molecular chaperones to most efficiently support folding of the broad repertoire of protein structures produced *in vivo*. Although our understanding of co-translational folding mechanisms is still in its infancy, these results imply that it should ultimately be possible to rationally design mRNA coding sequences in order to enhance co-translational folding and to identify disease-associated synonymous codon substitutions most likely to adversely affect protein folding, particularly for large or otherwise complex proteins.

## Methods

### *Cell growth assays*

A single colony of *E. coli* KA12 [66] or W3110 [50] transformed with a pKT-CAT plasmid from a freshly streaked LB-amp plate was used to inoculate 20 mL of LB plus 100 µg/mL ampicillin (LB-amp) and grown overnight with shaking at 37°C. Unless otherwise specified, all cultures contained 100 µg/mL ampicillin and no tetracycline. Overnight cultures were used to innoculate fresh LB-amp to an optical density at 600 nm ($OD_{600}$) of 0.05, to which was added 35 µg/mL chloramphenicol (unless otherwise specified) and the indicated concentration of tetracycline inducer (0-1600 ng/mL), transferred to one well of a 12-well plate and incubated at 37°C with continuous shaking in a Synergy H1 microplate reader (BioTek). Growth was measured as the increase in $OD_{600}$. The linear portion of the growth curve was fit to a straight line and the slope was taken as the growth rate.

## Acknowledgements

We thank Matt Champion for performing the mass spectrometry experiments, Don Hilvert for the kind gift of the pKT and pKTS plasmids and Peter Chein, Don Hilvert, Jeff Nivala and Mark Akeson for sharing *E. coli* strains with us. We are grateful to Anabel Rodriguez, Gabriel Wright and Scott Emrich for helpful discussions. This project was supported by grants GM120733 and GM105816 from the National Institutes of Health.

## References


1. Kimchi-Sarfaty, C, et al. (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 315:525-8.
2. Komar, AA (2016) The Yin and Yang of codon usage. *Hum Mol Genet* 25:R77-R85.
3. Subramaniam, AR, et al. (2013) A serine sensor for multicellularity in a bacterium. *Elife* 2:e01501.
4. Radhakrishnan, A, et al. (2016) The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell* 167:122-132 e9.
5. Bhattacharyya, S, et al. (2018) Accessibility of the Shine-Dalgarno Sequence Dictates N-Terminal Codon Bias in E. coli. *Mol Cell* 70:894-905 e5.
6. Drummond, DA and Wilke, CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341-52.
7. Daidone, V, et al. (2011) An apparently silent nucleotide substitution (c.7056C>T) in the von Willebrand factor gene is responsible for type 1 von Willebrand disease. *Haematologica* 96:881-7.
8. Pechmann, S, Chartron, JW, and Frydman, J (2014) Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol* 21:1100-5.
9. Zhang, G, Hubalewska, M, and Ignatova, Z (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 16:274-80.
10. Sander, IM, Chaney, JL, and Clark, PL (2014) Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc* 136:858-61.
11. Komar, AA, Lesnik, T, and Reiss, C (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett* 462:387-91.
12. Buhr, F, et al. (2016) Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol Cell* 61:341-51.
13. Zhang, F, Saha, S, Shabalina, SA, and Kashina, A (2010) Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. *Science* 329:1534-7.
14. Sorensen, MA and Pedersen, S (1991) Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J Mol Biol* 222:265-80.
15. Tuller, T, et al. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141:344-54.
16. Shah, P and Gilchrist, MA (2011) Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc Natl Acad Sci U S A* 108:10231-6.
17. Boel, G, et al. (2016) Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature* 529:358-363.
18. Clarke, TFt and Clark, PL (2008) Rare codons cluster. *PLoS One* 3:e3412.
19. Chaney, JL, et al. (2017) Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput Biol* 13:e1005531.
20. Jacobs, WM and Shakhnovich, EI (2017) Evidence of evolutionary selection for cotranslational folding. *Proc Natl Acad Sci U S A* 114:11434-11439.
21. Pechmann, S and Frydman, J (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* 20:237-43.
22. Dill, KA and MacCallum, JL (2012) The protein-folding problem, 50 years on. *Science* 338:1042-6.

23. Plaxco, KW, Simons, KT, and Baker, D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277:985-94.
24. Braselmann, E, Chaney, JL, and Clark, PL (2013) Folding the proteome. *Trends Biochem Sci* 38:337-44.
25. Chaney, JL and Clark, PL (2015) Roles for Synonymous Codon Usage in Protein Biogenesis. *Annu Rev Biophys* 44:143-66.
26. Hanson, G and Coller, J (2018) Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* 19:20-30.
27. Brule, CE and Grayhack, EJ (2017) Synonymous Codons: Choose Wisely for Expression. *Trends Genet* 33:283-297.
28. Jacobson, GN and Clark, PL (2016) Quality over quantity: optimizing co-translational protein folding with non-'optimal' synonymous codons. *Curr Opin Struct Biol* 38:102-10.
29. Balchin, D, Hayer-Hartl, M, and Hartl, FU (2016) In vivo aspects of protein folding and quality control. *Science* 353:aac4354.
30. Kramer, G, Shiber, A, and Bukau, B (2018) Mechanisms of Cotranslational Maturation of Newly Synthesized Proteins. *Annu Rev Biochem*.
31. Pechmann, S, Willmund, F, and Frydman, J (2013) The ribosome as a hub for protein quality control. *Mol Cell* 49:411-21.
32. Preissler, S and Deuerling, E (2012) Ribosome-associated chaperones as key players in proteostasis. *Trends Biochem Sci* 37:274-83.
33. Leslie, AG (1990) Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 A resolution. *J Mol Biol* 213:167-86.
34. Shaw, WV (1967) The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant Escherichia coli. *J Biol Chem* 242:687-93.
35. Eames, M and Kortemme, T (2012) Cost-benefit tradeoffs in engineered lac operons. *Science* 336:911-5.
36. Rodriguez, A, Wright, G, Emrich, S, and Clark, PL (2018) %MinMax: A versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding. *Protein Sci* 27:356-362.
37. Mustoe, AM, et al. (2018) Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* 173:181-195 e18.
38. Goodman, DB, Church, GM, and Kosuri, S (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342:475-9.
39. Kudla, G, Murray, AW, Tollervey, D, and Plotkin, JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324:255-8.
40. Welch, M, et al. (2009) Design parameters to control synthetic gene expression in Escherichia coli. *PLoS One* 4:e7002.
41. Neuenschwander, M, Butz, M, Heintz, C, Kast, P, and Hilvert, D (2007) A simple selection strategy for evolving highly efficient enzymes. *Nat Biotechnol* 25:1145-7.
42. Mittal, P, Brindle, J, Stephen, J, Plotkin, JB, and Kudla, G (2018) Codon usage influences fitness through RNA toxicity. *Proc Natl Acad Sci U S A* 115:8639-8644.
43. Gottesman, S, Roche, E, Zhou, Y, and Sauer, RT (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* 12:1338-47.
44. Nivala, J, Marks, DB, and Akeson, M (2013) Unfoldase-mediated protein translocation through an alpha-hemolysin nanopore. *Nat Biotechnol* 31:247-50.
45. Keiler, KC, Waller, PR, and Sauer, RT (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271:990-3.

46. Flynn, JM, Neher, SB, Kim, YI, Sauer, RT, and Baker, TA (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* 11:671-83.
47. Lee, C, Schwartz, MP, Prakash, S, Iwakura, M, and Matouschek, A (2001) ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol Cell* 7:627-37.
48. Sauer, RT, et al. (2004) Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* 119:9-18.
49. Ortega, J, Singh, SK, Ishikawa, T, Maurizi, MR, and Steven, AC (2000) Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol Cell* 6:1515-21.
50. Gottesman, S, Clark, WP, de Crecy-Lagard, V, and Maurizi, MR (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of Escherichia coli. Sequence and in vivo activities. *J Biol Chem* 268:22618-26.
51. Haase, N, Holtkamp, W, Lipowsky, R, Rodnina, M, and Rudorf, S (2018) Decomposition of time-dependent fluorescence signals reveals codon-specific kinetics of protein synthesis. *Nucleic Acids Res* 46:12186-12187.
52. Choi, J, et al. (2018) How Messenger RNA and Nascent Chain Sequences Regulate Translation Elongation. *Annu Rev Biochem* 87:421-449.
53. Chen, C, et al. (2013) Dynamics of translation by single ribosomes through mRNA secondary structures. *Nat Struct Mol Biol* 20:582-8.
54. Nackley, AG, et al. (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 314:1930-3.
55. Burkhardt, DH, et al. (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife* 6:811.
56. Nakahigashi, K, et al. (2014) Effect of codon adaptation on codon-level and gene-level translation efficiency in vivo. *BMC Genomics* 15:1115.
57. Beaudoin, JD, et al. (2018) Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat Struct Mol Biol* 25:677-686.
58. Xia, K, et al. (2007) Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proc Natl Acad Sci U S A* 104:17329-34.
59. Park, C, Zhou, S, Gilmore, J, and Marqusee, S (2007) Energetics-based protein profiling on a proteomic scale: identification of proteins resistant to proteolysis. *J Mol Biol* 368:1426-37.
60. Yu, CH, et al. (2015) Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Mol Cell* 59:744-54.
61. Guinn, EJ, Tian, P, Shin, M, Best, RB, and Marqusee, S (2018) A small single-domain protein folds through the same pathway on and off the ribosome. *Proc Natl Acad Sci U S A* 115:12206-12211.
62. Kaiser, CM, Goldman, DH, Chodera, JD, Tinoco, I, Jr., and Bustamante, C (2011) The ribosome modulates nascent protein folding. *Science* 334:1723-7.
63. Tian, P, et al. (2018) Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc Natl Acad Sci U S A* 115:E11284-E11293.
64. Kim, SJ, et al. (2015) Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science* 348:444-8.
65. Li, C, et al. (2011) FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol* 11:92.
66. Kast, P, Asif-Ullah, M, Jiang, N, and Hilvert, D (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: the importance of electrostatic catalysis. *Proc Natl Acad Sci U S A* 93:5043-8.

67. Gill, SC and von Hippel, PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182:319-26.
68. Shaw, WV (1975) Chloramphenicol Acetyltransferase from Chloramphenicol-Resistant Bacteria, ed. Hash, JH (Academic Press, New York), pp 737-755.
69. Frase, H, Toth, M, Champion, MM, Antunes, NT, and Vakulenko, SB (2011) Importance of position 170 in the inhibition of GES-type beta-lactamases by clavulanic acid. *Antimicrob Agents Chemother* 55:1556-62.
70. Frase, H, et al. (2011) Identification of products of inhibition of GES-2 beta-lactamase by tazobactam by x-ray crystallography and spectrometry. *J Biol Chem* 286:14396-409.
71. Bosserman, RE, et al. (2017) WhiB6 regulation of ESX-1 gene expression is controlled by a negative feedback loop in Mycobacterium marinum. *Proc Natl Acad Sci U S A* 114:E10772-E10781.
72. Cox, J, et al. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13:2513-26.
73. Zuker, M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406-15.
74. Martin, A, Baker, TA, and Sauer, RT (2005) Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* 437:1115-20.

## Figure Legends

**Figure 1**. Chloramphenicol acetyltransferase (CAT) has a complex tertiary and quarternary structure. (**A**) The native homotrimeric structure (PDBID: 3CLA) [33]. (**B**) Schematic representation of the complex topology of the CAT monomer structure. Secondary structure elements are shown in rainbow order. Polka dots indicate the H β-strand in the central β-sheet contributed from an adjacent monomer. (**C**) Close up of the trimer interface. Highlighted in color are β-strands in the central β-sheets contributed from two different polypeptide chains.

**Figure 2**. CAT encoded by the synonymous Shuf1 sequence impaired growth in the presence of chloramphenicol (cam). (**A**) Relative codon usage in WT (black), Komar [11] (green), and Shuf1 (gray) CAT coding sequences. Positive values correspond to clusters of common codons and negative values represent clusters of rare codons, calculated over a sliding window of 17 codons [36]. (**B**) Growth curves of *E. coli* expressing ssrA-tagged CAT variants challenged with cam under low (200 ng/mL) or high (1,600 ng/mL) concentrations of inducer. (**C**) Relative abundance of untagged (solid bars) or ssrA-tagged (hatched bars) CAT accumulated in cells determined by quantitative western blotting of cell lysates. (**D**) Growth curves in the absence of cam. In all figures, data points represent the mean ± SD of at least three independent experiments; ** $p < 0.01$, *** $p < 0.001$, Welch's t-test.

**Figure 3**. Native states of CAT variants are indistinguishable. (**A**) Relative abundance of *E. coli* proteins upon expression of WT or Shuf1 CAT. Twelve *E. coli* molecular chaperones and AAA+ ATPases are shown in red; 1264 other *E. coli* proteins are shown in black. No significant upregulation of chaperones or ATPases was observed for *E. coli* expressing Shuf1. (**B**) Thermal denaturation of CAT monitored by far-UV CD spectroscopy at 205 nm. (**C**) Acetyltransferase activity of purified, native CAT, normalized to WT. Data points represent the mean ± SD; n=3.

**Figure 4**. Shuf1 CAT *in vivo* folding intermediates, but not the native structure, are more susceptible to ClpXP degradation than WT CAT. (**A**) *In vitro* ClpXP degradation of native, purified, ssrA-tagged CAT trimers. The CAT band is marked with an arrowhead. (**B-C**) Selective effects of ssrA-tagging and ClpX deletion on the Shuf1 growth defect. (**B**) In the ClpX deletion strain (W3110 *ΔclpX*), a large increase in growth rate relative to the parent strain is observed only for ssrA-tagged Shuf1. Other constructs grow slightly slower in the absence of ClpX. U, uninduced cell culture. (**C**) Impact on growth rate of removing the ssrA tag. Omitting the ssrA tag has no effect on growth in the ClpX knockout (hatched bars). In the presence of ClpX (solid bars), there is a much larger increase in growth upon ssrA tag deletion for Shuf1 than WT, indicating Shuf1 is more susceptible to ClpXP degradation than WT. Data points represent the mean ± SD; n=3 biological replicates.

**Figure 5**. Growth curves for *E. coli* induced with (**B**) high (1,600 ng/mL) or (**C**) low (200 ng/mL) concentrations of inducer prior to challenge with chloramphenicol. Data points represent the mean ± SD; n=3 biological replicates.

**Figure 6**. Proposed model for Shuf1 growth defect. Synonymous changes in the Shuf1 coding sequence alter the local rate of translation, affecting the conformation of CAT after release from the ribosome. Post-translational Shuf1 folding is therefore slower, leaving folding intermediates susceptible to degradation by ClpXP. Some molecules can evade degradation and eventually fold to the native state.

**Figures**

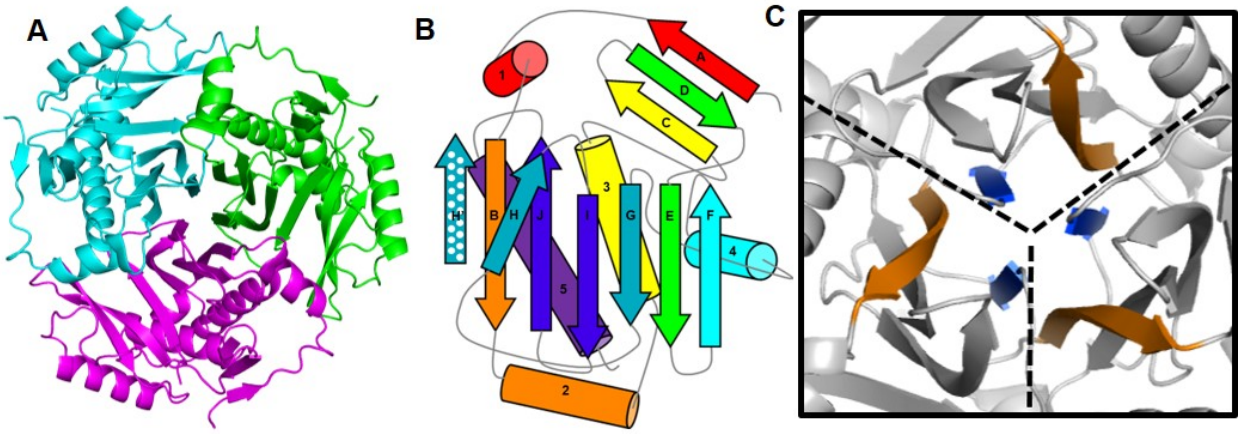


**Figure 1**

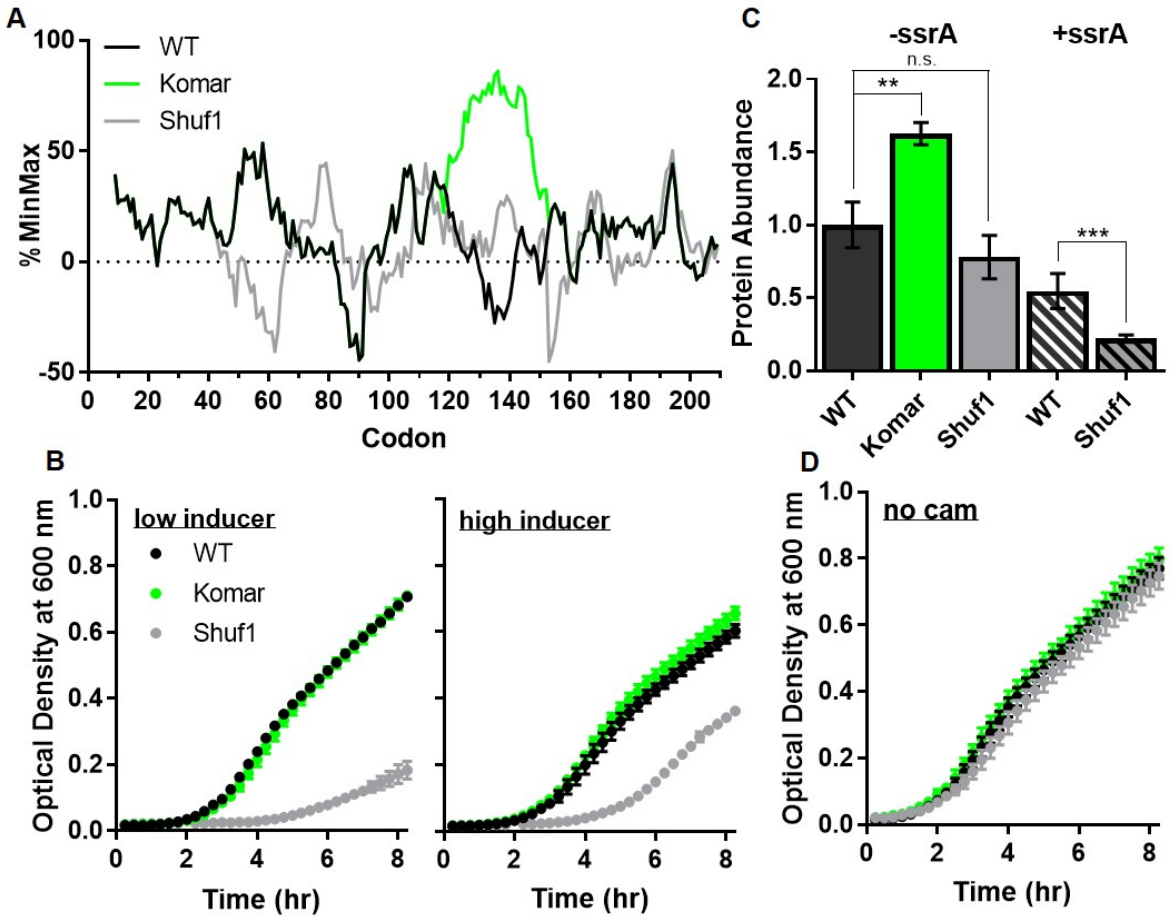
A
WT
Komar
Shuf1
% MinMax
Codon
C
-ssrA
+ssrA
n.s.
**
***
Protein Abundance
WT
Komar
Shuf1
WT
Shuf1
B
low inducer
high inducer
WT
Komar
Shuf1
Optical Density at 600 nm
Time (hr)
D
no cam
Optical Density at 600 nm
Time (hr)

**Figure 2**

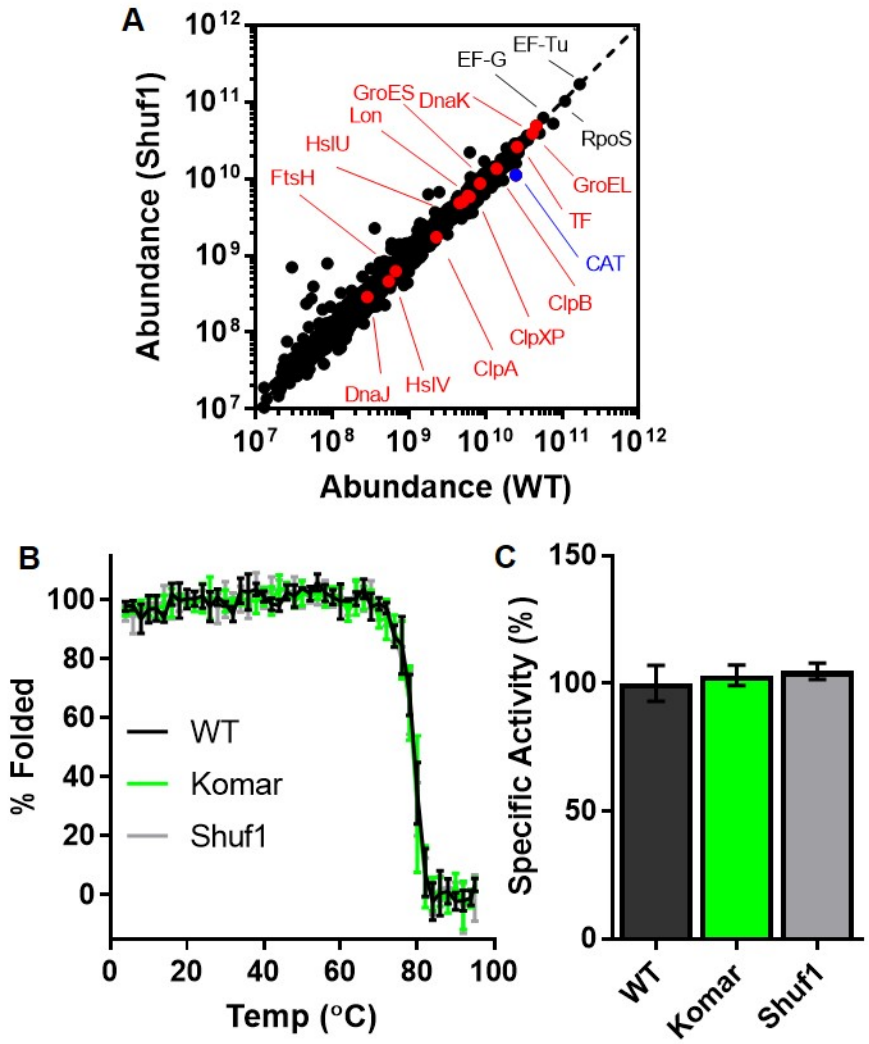
A
Abundance (Shuf1)
Abundance (WT)
EF-G
EF-Tu
GroES
DnaK
Lon
HslU
FtsH
RpoS
GroEL
TF
CAT
ClpB
ClpXP
ClpA
HslV
DnaJ
B
% Folded
Temp (°C)
WT
Komar
Shuf1
C
Specific Activity (%)
WT
Komar
Shuf1

**Figure 3**

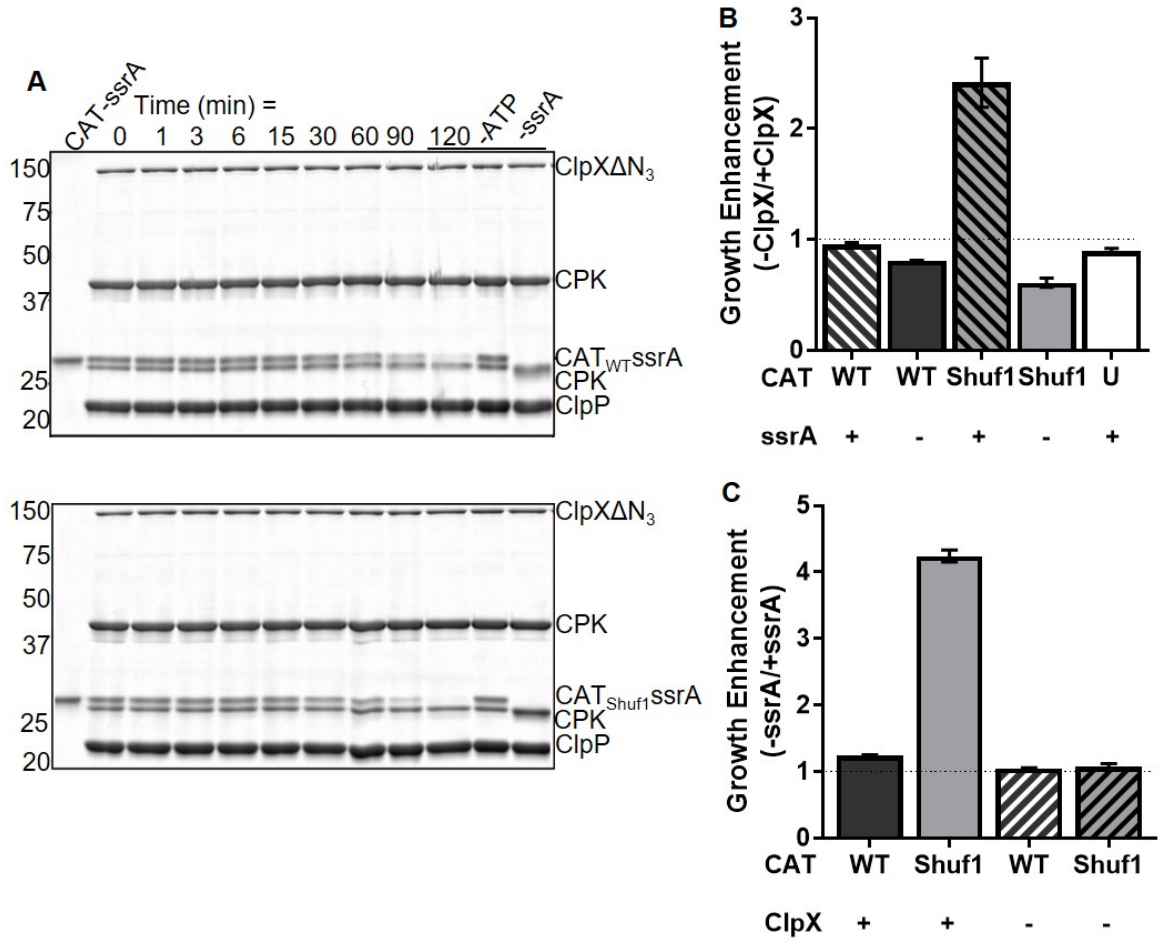
A
CAT-ssrA
Time (min) =
0 1 3 6 15 30 60 90 120 -ATP -ssrA
150
75
50
37
25
20
ClpXΔN3
CPK
CATWTssrA
CPK
ClpP
ClpXΔN3
CPK
CATShuf1ssrA
CPK
ClpP
B
Growth Enhancement
(-ClpX/+ClpX)
CAT WT WT Shuf1 Shuf1 U
ssrA + - + - +
C
Growth Enhancement
(-ssrA/+ssrA)
CAT WT Shuf1 WT Shuf1
ClpX + + - -

**Figure 4**

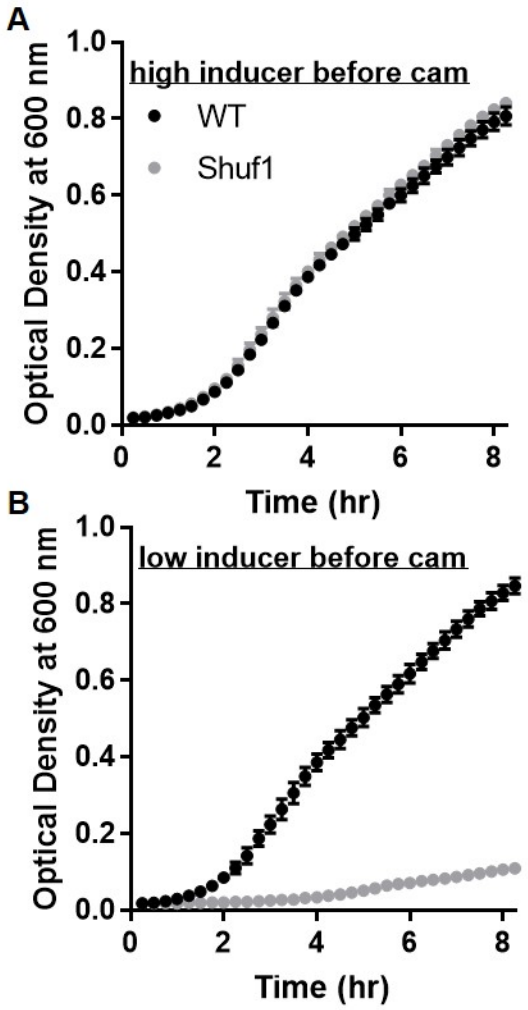
A
high inducer before cam
WT
Shuf1
Optical Density at 600 nm
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
Time (hr)
B
low inducer before cam
Optical Density at 600 nm
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
Time (hr)

**Figure 5**

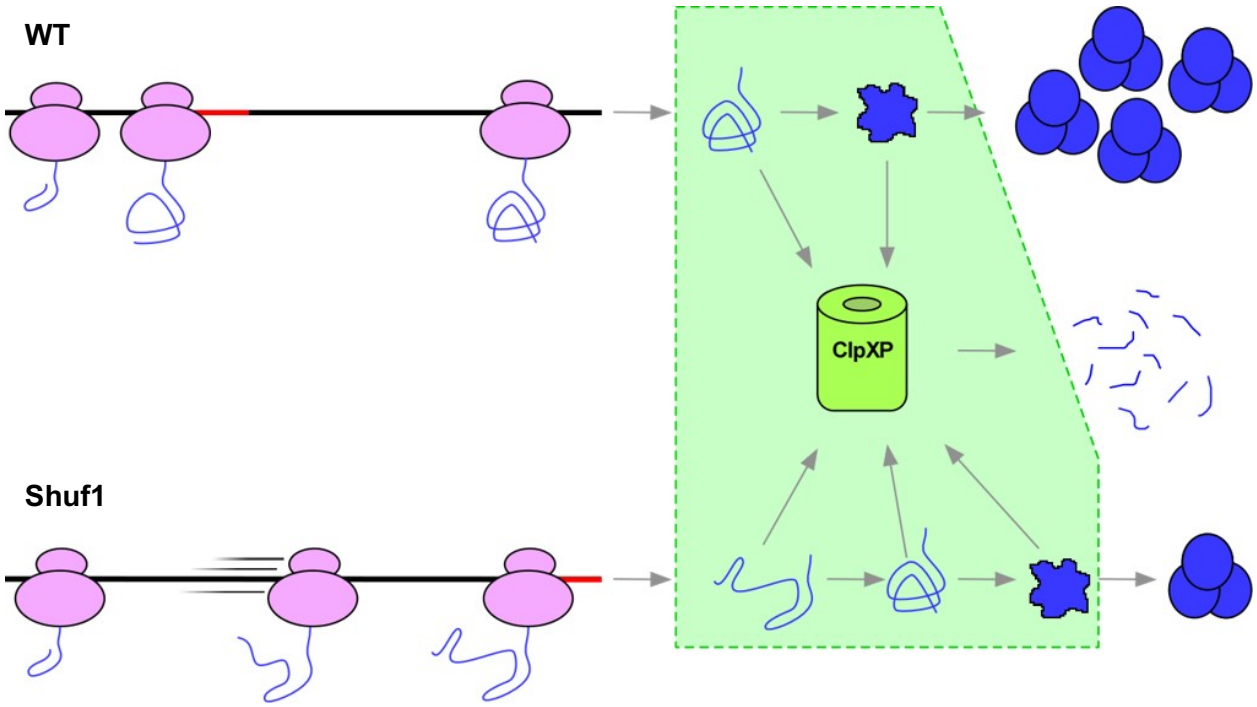


**Figure 6**

**Supplemental Information for:**

# Synonymous codon substitutions impair protein folding and cell fitness

Ian M. Walsh[a], Micayla A. Bowman[a] & Patricia L. Clark[a,b,1]

[a]Department of Chemistry & Biochemistry and
[b]Department of Chemical & Biomolecular Engineering,
University of Notre Dame, Notre Dame, IN 46556 USA

[1]*To whom correspondence should be addressed:*
pclark1@nd.edu
01-574-631-8353

**Contents**

Supplemental Methods

Supplemental Figures 1-6

## Supplemental Methods

*Synonymous CAT coding sequence design*

The Shuf1 synonymous mutant sequence was designed using a Monte Carlo algorithm. The first 46 codons (excluding the His tag) were left unchanged to avoid altering translation initiation [5, 37-40]. For remaining amino acid positions, a synonymous codon was chosen at random. After each random sequence was designed, its codon usage frequencies were analyzed using the %MinMax algorithm [18, 36], which compares the codon usage frequency of a coding sequence to hypothetical sequences that encode the identical amino acid sequence using either the most common (+100%) or most rare (-100%) codons possible. This simple algorithm has been shown to accurately predict the relative effects of synonymous codon changes on elongation rate [10, 19]. The %MinMax codon usage profile of each synonymous sequence was compared to the WT profile. Profiles were filtered to select those with the same number of 17-codon windows of rarer-than-average codon usage, a sequence-wide average codon usage very similar to the wild type mRNA sequence (%MinMax within +/- 3% of WT), and likewise very similar extreme codon usage values (maximum and minimum %MinMax values within +/- 5% of WT). Out of one million sequences generated, three fulfilled these criteria. The coding sequence with the lowest Pearson correlation coefficient compared to WT was chosen as Shuf1.

*Molecular biology*

All CAT constructs were synthesized by GeneArt (Invitrogen). The pKT and pKTS plasmids were a generous gift from Donald Hilvert (ETH, Zurich). CAT constructs were cloned into pKT and pKTS using FastCloning [65]. All primers were ordered from IDT as standard desalting primers and resuspended to 1 mg/mL in $ddH_2O$. pKT and pKTS were amplified using primers pKT F (ACTAGTGCGGCCGCTTGATAAA) and pKT R (ATGTATATCTCCTTCTTAA AGTTAAACAAAATTATTTCTAGAGGGAAA) and pKTS F (GCGGCGAACGATGAAAACTATG C) and pKT R, respectively. The CAT F primer (AGAAGGAGATATACATATGCATCACC ATCACCATCACCATAACTATACAAAATTTG) was used to amplify all CAT constructs. Unique primers were required for the reverse direction. WT R pKT (AAGCGGCCGCACTAGTTTATTTTAATTTA CTGTTACACAACTCTTGTAG) was used to clone WT and the common variant of CAT [11] into pKT, WT R pKTS (TTTCA TCGTTCGCCGCTTTTAATTTACTGTTACACAACTCTTGTAGCCGATTAATAAAGC) for pKTS, Shuf1 R pKT (AAGCGGCCGCACTAGTTTACTTCAATTTAGAATTACATAACT CCTGTAAT) was used to clone Shuf1 CAT into pKT, and Shuf1 R pKTS (TTTCATCGTTCGCCGCCTTCAATTTAGAATTACATAACTCCT GTAATCTGTTAATGAAGC) for pKTS.

To make the chimeric CAT constructs, the 5', center or 3' region of the Shuf1 coding sequence were amplified and cloned into the WT CAT sequence. All hybrid CAT constructs were cloned into pKTS. Primers were ordered from IDT as standard desalting primers. To make the 5' chimera, the WT plasmid was amplified using the primers WTsN F (GCTATCCTGCCCC TATTCATCCGATATTGATCAATTTATGGTGAATTATTTATCGGT) and WTsN R (ATATCATCA CGGGGTAAAACTTATACGCTGAATCATCCAATGACTTTTTTAACGTC) and codons 58-111 of Shuf1 using the primers Shuf1 N F (TACCCCGTGATGATATACTTAATTGCCCA) and Shuf1 N R (ATAGGGGCAGGATAGCGCAC). To make the middle chimera, the WT plasmid was amplified using the primers WTsM F (TGACTATTTTGCGCCGATTATAACAATGGCAAAATAT CAGCAAGAAGGG) and WTsM R (GATCGATGTCCGAGCTGTATGGGCAACTCAGTGCTGAA AA) and codons 112-172 of Shuf1 using the primers Shuf1 M F (AGCTCGGACATCGATCAAT

TTATGG) and Shuf1 M R (CGGCGCAAAATAGTCAGTAAAGTTAGCTA). To make the 3' chimera, the WT plasmid was amplified using the primers WTsC F (TAATTCTAAATTGAAGG CGGCGAACGATGAAAACTATGC) and WTsC R (TGGCCATGGTTATGATGGGTGCAAAATA ATCGGTAAAATTAGCAA) and codons 173-220 of Shuf1 using the primers Shuf1 C F (ATCAT AACCATGGCCAAATACCAACAG) and Shuf1 C R (CTTCAATTTAGAATTACATAACTCCTGT AATCTGTTAATGAAGC).

PCR reactions were prepared using 5 µL of supplied 10x *Pfu*Ultra II buffer (Agilent), 2.5 mM dNTPs (AMRESCO), 2 ng/µL of forward and reverse primer, 0-3 mM $MgCl_2$, 0.2 ng/µL DNA, and 1 µL *Pfu*Ultra II Fusion HS DNA polymerase (Agilent). The total reaction volume was 50 µL. The PCR protocol started with 2 min at 95°C followed by 18 cycles of: 95°C for 30 sec, $T_m$-5°C for 1 min, 72°C for 2.5 min, then a final step of 72°C for 10 min.

PCR products were treated with 1 µL *DpnI* for 1.5 hours at 37°C. The insert and vector were then mixed together in 1:1, 1:2, 1:4, 2:1, and 4:1 ratios and 10 µL were incubated with 100 µL chemically competent *E. coli* DH5α for 30 min on ice. Cells were heat shocked at 42°C for 45 sec, incubated on ice for 2 min, and recovered at 37°C with 500 µL of SOC media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) for 1 hr. The entire transformation reaction was spread on an LB-agar plate with 100 µg/mL ampicillin and incubated overnight at 37°C. Single colonies were grown in liquid culture and DNA was extracted (Wizard Plus minipreps, Promega) using the manufacturer's directions. The entire gene along with its promoter and transcription terminator was sequenced.

*CAT protein purification*

To purify CAT for *in vitro* analyses, 1 L cultures were inoculated as above with 1600 ng/mL tetracycline. Cultures were grown at 37°C with continuous shaking for 6.5 hours. Cells were pelleted at 4,000x*g* for 15 min and resuspended in nickel affinity binding buffer (500 mM NaCl, 20 mM imidazole, 20 mM sodium phosphate pH 8). Cells were lysed by repeated cycles of freezing at -80°C for 30 min followed by thawing at room temperature. Lysozyme (final concentration 1 mg/mL) was added after the first freezing cycle. Thawed lysates were incubated for 30 min at 4°C with frequent inversions to facilitate lysis. In total, cells were frozen and thawed four times. After lysis, cells were homogenized with 25 mM $MgCl_2$ and 80 µL DNase I (Invitrogen). The lysates were again incubated at 4°C for 30 min with frequent inversions. Alternatively, lysates were homogenized by sonication for 12 min using 30 sec on 30 sec off. After lysis, the lysate was pelleted at 14,000x*g* for 15 min and CAT was purified from the supernatant using nickel affinity chromatography (His Spin Trap (small volumes), HisTrap HP (large volumes); GE Healthcare). After addition of the cell supernatant, columns were washed with binding buffer and eluted with elution buffer (500 mM NaCl, 500 mM imidazole, 20 mM sodium phosphate pH 8). CAT protein concentration was determined by absorbance at 280 nm, using a calculated CAT extinction coefficient of 30,350 (for His-CAT) or 31,630 $M^{-1}$ $cm^{-1}$ (for His-CAT-ssrA) [67].

*Far-UV circular dichroism spectroscopy*

Far-UV CD spectra were collected using a J-815 CD Spectropolarimeter (JASCO). CAT protein samples were diluted to a concentration of 4 µM in 20 mM phosphate buffer pH 8. Three spectra were collected from 190-250 nm with a 1 second integration time and a 2 nm bandwidth, averaged and from this an averaged spectrum of the blank buffer solution (collected under identical conditions) was subtracted. The effect of increasing temperature on the far-UV CD spectrum was recorded every 2°C from 4-95°C. Samples were equilibrated for 2 min at each

temperature prior to data collection. To normalize the spectra, the average signal prior to and after the unfolding transition were set to 100% and 0% folded, respectively.

*Tryptophan fluorescence emission spectroscopy*

Purified CAT protein was diluted to 2 μM and incubated overnight in 100 mM Tris buffer pH 7.5 with 0-4 M guanidinium HCl at 4°C. Tryptophan emission spectra were collected using a PTI QM-6 fluorimeter (Birmingham, NJ). An excitation wavelength of 280 nm was used (2 nm slit widths) and emission was measured from 300-380 nm (4 nm slit widths; 3 sec integration time). To normalize the spectra, the average fluorescence emission intensity prior to and after the unfolding transition were set to 100% and 0% folded, respectively. The resulting curves were then fit to a sigmoidal function.

*Acetyltransferase activity assay*

CAT enzymatic activity was measured as described previously [68]. Briefly, CAT was diluted to a concentration of 6.25 μg/mL in 1 mL of reaction mix (94 mM Tris pH 7.8, 83 μM DTNB, 190 μM acetyl-CoA and 155 μM chloramphenicol). Free CoA-SH produced by the CAT reaction subsequently reacts with DTNB to make TNB, which absorbs at 412 nm. Absorbance at 412 nm was monitored for 5 min in a DU 530 UV-Vis Spectrophotometer (Beckman Coulter). To determine the specific activity from these curves, the initial velocity of the reaction was calculated by fitting data points from 0.25-1 min to a straight line. The initial velocity was divided by the extinction coefficient for TNB (0.0136 $\mu M^{-1}$ $cm^{-1}$) to convert to units of enzyme activity and the amount of CAT to calculate activity in units per microgram. One unit of activity converts 1 nmol of chloramphenicol and acetyl-CoA to chloramphenicol-3-acetate and CoA per min. The average activity of WT His-CAT was set to 100%; results for other CAT proteins were normalized to this value.

*Mass spectrometry*

The molecular weight of CAT was measured using HPLC-ESI-MS as described previously [69, 70], with minor modifications. Briefly, purified CAT was diluted to 0.5 mg/mL in 25 mM ammonium bicarbonate pH 8.0 and 5 μg was injected per sample. A 20 min gradient from 10-80% A-B (A: $H_2O$, 0.1% formic acid; B: acetonitrile, 0.1% formic acid) with the first 3 min diverted to waste was performed. Positive mode electron spray ionization was acquired on a micro TOF-QII (Bruker, Billerica, MA) mass spectrometer with 2 sec spectral averaging. The MaxEnt software package (Bruker) was used for deconvolution. The deconvoluted mass measured for each CAT variant was 25,953 Da, identical to the predicted mass.

*E. coli* expressing WT or Shuf1 His-CAT-ssrA were grown to an $OD_{600}$ of 0.3 at 37°C in LB with amp and induced with 1000 ng/mL tc for 3 hr. Cells (from 10 mL cultures) were pelleted at 4,000x*g* for 15 min and stored at -80°C. Cell pellets were lysed using a bead mill (BioSpec OK, USA) 3 x 30 sec in 1% SDS 100 mM ammonium bicarbonate (ABC) and clarified by centrifugation at 12,000x*g* for 10 min. Whole cell protein lysates were quantified by BCA assay (Thermo Scientific) according to the manufacturer's instructions.

Whole cell lysates were prepared for proteomics as previously described [71], with minor modifications. All reagents were from Sigma unless otherwise described. First, 50 μg of each lysate was suspended in 5% SDS 100mM ABC and reduced and alkylated with dithiothreitol and iodoacetamide. Samples were then digested with trypsin using Strap columns (Protifi, NY)

according to manufacturer's instructions, with the following alterations. Mass-spectrometry grade trypsin (Promega, WI) was added at 1:40 (m/m) and digested overnight at 37°C. Peptides were eluted from the digestion column and desalted using a 10 mg HLB SPE column (Waters, MA) following the manufacturer's recommended conditions. Desalted peptide digests were dried and stored at -20°C until LC/MS/MS analysis.

Each digested sample (1 µg) was analyzed in triplicate using LC-MS/MS with a Waters nanoUHPLC Acquity (Billerica, MA) and a Q-Exactive Oribitrap (Thermo, San Jose, CA). RAW files from the Nano UHPLC-MS/MS acquisition were subjected to tandem spectral match and quantification using label free quantification [72]. The current KA12 *E. coli* proteome (UP000000625) concatenated with the WT/Shuf1 His-CAT-ssrA sequences and common contaminants was used to assign peptides. Data were normalized using median fold change.

*Calculation of local mRNA secondary structure*

A sliding window analysis was used to predict local stable structure in overlapping 30 nt-long segments of the CAT coding sequence, using the Quikfold algorithm to calculate the free energy of mRNA folding [73]. All predicted stable structures were plotted.

*Correlation between growth and mRNA properties*

Growth was calculated as the %WT $OD_{600}$ after 4 hours of growth in cam. The change in %MinMax was calculated by summing the absolute value of the differences between a given construct and WT. This number was divided by the average %MinMax of the gene as a proxy for protein production. The strongest local mRNA secondary structure element per gene was calculated as above. These data were fit to a line using Prism.

*Western blotting*

Cultures were inoculated as described above and induced for 6.5 hours. Aliquots were pelleted at 14,000x*g* for 15 min and the cell pellets were resuspended to an $OD_{600}$ of 15 in 100 mM Tris pH 7.5. To this was added one half-volume of 3x SDS gel loading buffer and β-mercaptoethanol (BME) to 5 mM. Samples were boiled for 40 min, separated using denaturing polyacrylamide gel electrophoresis, transferred to a PVDF membrane and probed using a mouse anti-His antibody (Promega) and goat anti-mouse secondary antibody conjugated to alkaline phosphatase. In some experiments, the 7x-His tag was replaced with an HA tag (YPYDVPDYA) and the primary antibody used was rabbit α-HA (Sigma). Intensities of CAT protein bands were quantified using ImageJ.

*mRNA quantification*

Cultures were inoculated as described above and grown for 2.5 hr until the cells achieved an $OD_{600}$ of about 0.3. RNA was purified using the RNeasy kit (Qiagen), according to manufacturer protocol. RNA concentrations were determined by absorbance at 260 nm. Purified RNA was stored at -80°C.

RNA samples were treated with DNase to remove genomic DNA or plasmid contaminants. To 10 µL of RNA (about 0.2 µg) 10 µL of DNase buffer (10 mM Tris pH 7.5, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$) was added with 2 units of RNase-free DNase (Ambion/Invitrogen).

Samples were incubated at 37°C for 30 min, then 1 µL of 100 mM EDTA was added and samples were incubated at 65°C for 10 min to denature the DNase. Reverse transcription (RT) was performed using the iScript RT kit (BioRad). RNA was diluted to about 100 ng/µL into RT reaction mix (20 µL total, 4 µL 5x buffer, 2 µL random hexamer primers, 1 µL reverse transcriptase). Control reactions lacking RT were used to ensure the PCR amplification was specific for mRNA. To perform the RT, the following protocol was used on a MyCycler (BioRad): 5 min at 25°C, 30 min at 42°C, 5 min at 85°C. The resulting cDNA was diluted fivefold for quantitative PCR (qPCR) in a 96 well plate using a CFX96 Touch Real-Time PCR Detection System (BioRad). The reaction volume was 5 µL and included 1 µL of cDNA, 60 nM primers (IDT) and 2.5 µL SsoAdvanced SYBR Green Supermix (BioRad). To amplify all CAT constructs the primers CAT qPCR F (CCATCACCATAACTATACAAAATTTGATGTAAAAAATT) and CAT qPCR R (AAACTTATACGCTGAATCATCCAATGA) were used. To amplify *E. coli* GAPDH, the primers GAPDH qPCR F (CGGTACCGTTGAAGTGAAAGAC) and GAPDH qPCR R (ACCAGTTGCTTCAGCGAC) were used. $\Delta C_T$ values were calculated by taking the difference in $C_T$ values for each CAT construct with its corresponding GAPDH $C_T$. $\Delta\Delta C_T$ values were calculated by taking the difference in $\Delta C_T$ from WT CAT. This was converted to fold change in RNA level by raising 2 to the power of $-\Delta\Delta C_T$.

*Growth rate as function of induction time*

Cultures were inoculated as described above in 20 mL of LB with amp and grown in shake flasks at 37°C. Cultures were grown to an $OD_{600}$ of ~0.35 at 37°C and induced with 1600 ng/mL tc. Cells were induced from 0-60 min. Then a 1 mL aliquot was pelleted at 21,000x*g* for 5 min and resuspended in LB with amp and cam. These were diluted twofold into a 24 well plate. Growth was monitored with double orbital shaking for 5 hours. Growth from 1-3 hours after chloramphenicol addition was fit to a line and the slope was used as the growth rate. Growth rates were normalized to the average growth rate of both constructs after 60 min induction.

*ClpX knockout effect on growth*

The ClpX knockout was in the *E. coli* W3110 background. Knockout and parent strain were cultured as described above. These strains were induced with 50 ng/mL tc. The effect of ClpX or ssrA tag on growth was calculated by dividing the endpoint $OD_{600}$. No effect of deletion would render a value of 1.

*Contact maps*

Contact maps were generated using VMD software support and the crystal structure of $CAT_{III}$ (PDBID: 3CLA) [33]. VMD was developed with by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

*ClpXP degradation assay*

ClpX$\Delta N_3$ or ClpP were overexpressed in *E. coli* and purified using $Ni^{2+}$ affinity and size exclusion chromatography, as described previously [74]. ClpXP degradation were performed in PD buffer (25 mM HEPES pH 7.6, 100 mM KCl, 10 mM $MgCl_2$, 10% v/v glycerol, 2 mM BME at 30°C. CATssrA purified as above (2 µM) was incubated with 300 nM $ClpX_6$ (as 600 nM ClpX$\Delta N_3$), 900 nM $ClpP_{14}$, 5 mM ATP (Sigma), 0.2 mg/ml creatine phosphokinase (CPK, Sigma), and 1.6 mM creatine phosphate (Sigma). At indicated timepoints, an aliquot of the degradation reaction mix was removed and combined with SDS-PAGE loading dye. The degradation reaction was monitored by SDS-PAGE.

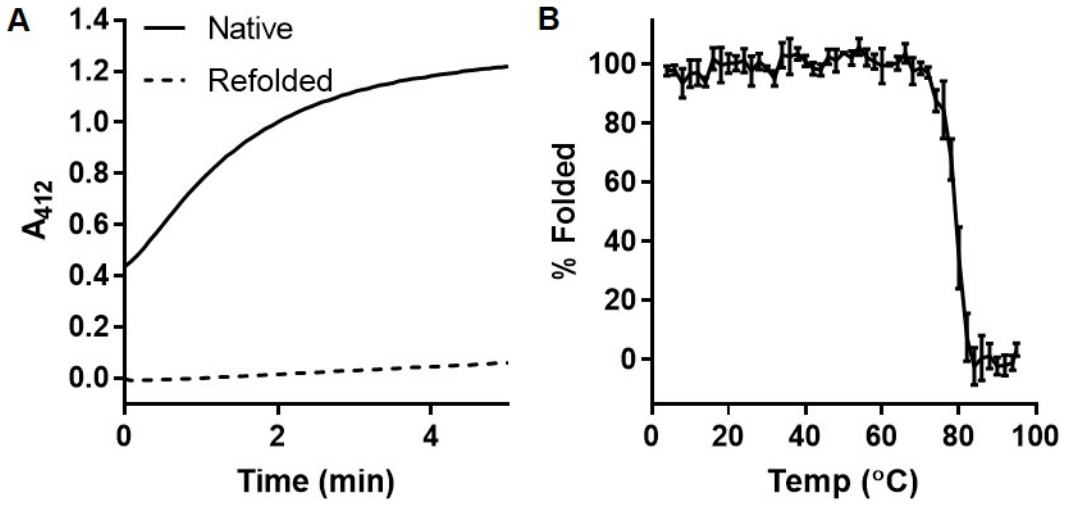


**Figure S1**

The native CAT structure is very thermostable but cannot be refolded after chemical denaturation. (**A**) CAT activity, expressed as the change in absorbance of TNB over time in the presence of native or refolded CAT (see Methods). Refolded CAT does not have significant acetyltransferase activity. (**B**) Thermal denaturation of native, purified CAT, monitored as the change in the far-UV CD signal at 205 nm; n=3, data are represented as mean ± SD.

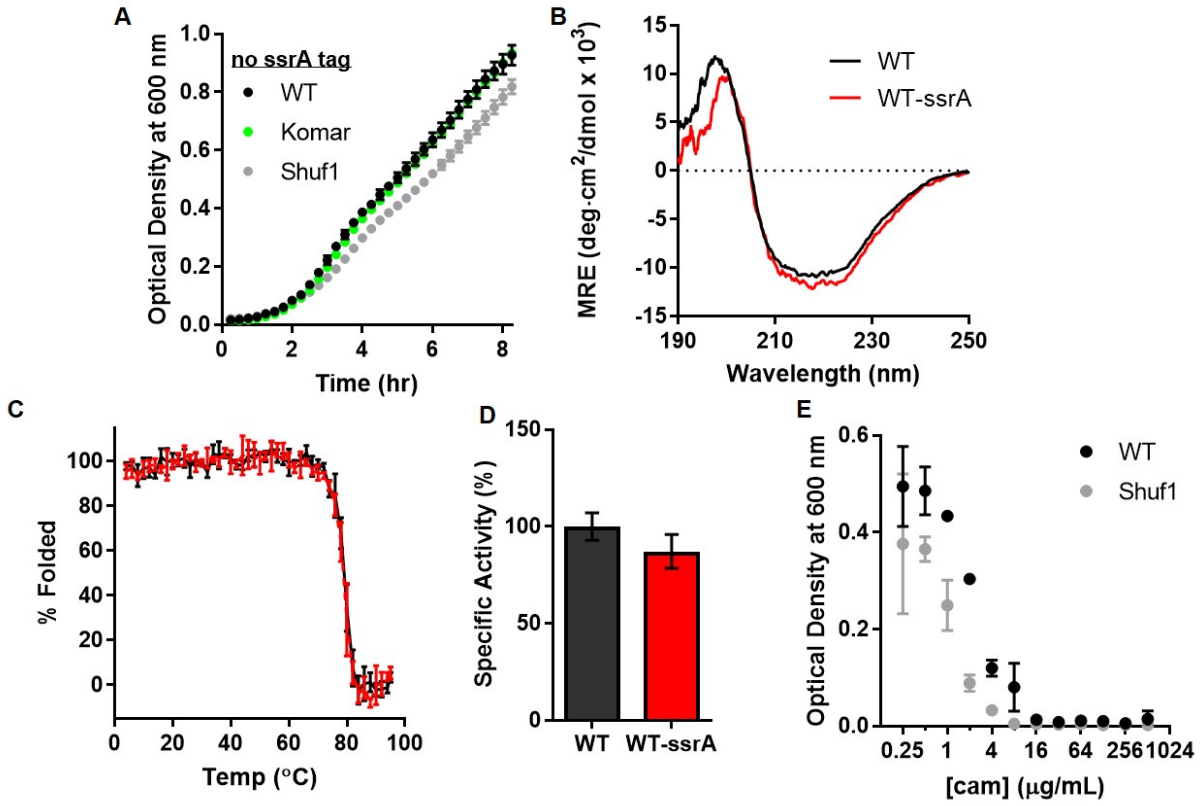


**Figure S2**

The C-terminal ssrA tag does not affect properties of native, purified CAT, but does affect chloramphenicol resistance of *E. coli* expressing CAT from the Shuf1 coding sequence. (**A**) Growth curves for cell expressing CAT codon variants without an ssrA tag. (**B**) Far-UV CD spectra of CAT and CAT-ssrA. (**C**) Thermal denaturation of CAT and CAT-ssrA monitored by changes in the far-UV CD signal at 205 nm. (**D**) Acetyltransferase activity of CAT and CAT-ssrA are indistinguishable; p=0.12, Welch's t-test; n=3, data are represented as mean ± SD. (**E**) *E. coli* were induced to express CAT from the WT or Shuf1 coding sequence for 5 hr and then tested for minimum inhibitory concentration of chloramphenicol (cam).

```
  1 Met-His-His-His-His-His-His-His-Asn-Tyr-Thr-Lys-Phe-Asp-Val-Lys-Asn-
 WT ATG CAT CAC CAT CAC CAT CAC CAT AAC TAT ACA AAA TTT GAT GTA AAA AAT
 Kr ATG CAT CAC CAT CAC CAT CAC CAT AAC TAT ACA AAA TTT GAT GTA AAA AAT
 Sh ATG CAT CAC CAT CAC CAT CAC CAT AAC TAT ACA AAA TTT GAT GTA AAA AAT

 18 Trp-Val-Arg-Arg-Glu-His-Phe-Glu-Phe-Tyr-Arg-His-Arg-Leu-Pro-Cys-Gly-
 WT TGG GTT CGC CGT GAG CAT TTT GAG TTT TAT CGG CAT CGT TTA CCA TGT GGT
 Kr TGG GTT CGC CGT GAG CAT TTT GAG TTT TAT CGG CAT CGT TTA CCA TGT GGT
 Sh TGG GTT CGC CGT GAG CAT TTT GAG TTT TAT CGG CAT CGT TTA CCA TGT GGT

 35 Phe-Ser-Leu-Thr-Ser-Lys-Ile-Asp-Ile-Thr-Thr-Leu-Lys-Lys-Ser-Leu-Asp-
 WT TTT AGC TTA ACA AGC AAA ATT GAT ATC ACG ACG TTA AAA AAG TCA TTG GAT
 Kr TTT AGC TTA ACA AGC AAA ATT GAT ATC ACG ACG TTA AAA AAG TCA TTG GAT
 Sh TTT AGC TTA ACA AGC AAA ATT GAT ATC ACG ACG TTA AAA AAG TCA TTG GAT

 52 Asp-Ser-Ala-Tyr-Lys-Phe-Tyr-Pro-Val-Met-Ile-Tyr-Leu-Ile-Ala-Gln-Ala-
 WT GAT TCA GCG TAT AAG TTT TAT CCG GTA ATG ATC TAT CTG ATT GCT CAG GCC
 Kr GAT TCA GCG TAT AAG TTT TAT CCG GTA ATG ATC TAT CTG ATT GCT CAG GCC
 Sh GAT TCA GCG TAT AAG TTT TAC CCC GTG ATG ATA TAC TTA ATT GCC CAA GCA

 69 Val-Asn-Gln-Phe-Asp-Glu-Leu-Arg-Met-Ala-Ile-Lys-Asp-Asp-Glu-Leu-Ile-
 WT GTG AAT CAA TTT GAT GAG TTG AGA ATG GCG ATA AAA GAT GAT GAA TTG ATC
 Kr GTG AAT CAA TTT GAT GAG TTG AGA ATG GCG ATA AAA GAT GAT GAA TTG ATC
 Sh GTT AAC CAA TTT GAC GAG CTC CGA ATG GCG ATT AAA GAT GAT GAA CTC ATT

 86 Val-Trp-Asp-Ser-Val-Asp-Pro-Gln-Phe-Thr-Val-Phe-His-Gln-Glu-Thr-Glu-
 WT GTA TGG GAT TCA GTC GAC CCA CAA TTC ACC GTA TTC CAT CAA GAA ACA GAG
 Kr GTA TGG GAT TCA GTC GAC CCA CAA TTC ACC GTA TTC CAT CAA GAA ACA GAG
 Sh GTG TGG GAC TCC GTC GAT CCG CAA TTT ACT GTA TTC CAT CAG GAG ACT GAA

103 Thr-Phe-Ser-Ala-Leu-Ser-Cys-Pro-Tyr-Ser-Ser-Asp-Ile-Asp-Gln-Phe-Met-
 WT ACA TTT TCA GCA CTG AGT TGC CCA TAC TCA TCC GAT ATT GAT CAA TTT ATG
 Kr ACA TTT TCA GCA CTG AGT TGC CCA TAC TCA TCC GAT ATT GAT CAA TTT ATG
 Sh ACT TTT AGT GCG CTA TCC TGC CCC TAT AGC TCG GAC ATC GAT CAA TTT ATG

120 Val-Asn-Tyr-Leu-Ser-Val-Met-Glu-Arg-Tyr-Lys-Ser-Asp-Thr-Lys-Leu-Phe-
 WT GTG AAT TAT TTA TCG GTA ATG GAA CGT TAT AAA AGT GAT ACC AAG TTA TTT
 Kr GTG AAT TAT TTA TCG GTA ATG GAA CGT TAT AAA TCT GAT ACA AAA CTG TTC
 Sh GTG AAT TAC CTG TCG GTC ATG GAA AGG TAT AAG TCC GAC ACA AAA TTA TTT

137 Pro-Gln-Gly-Val-Thr-Pro-Glu-Asn-His-Leu-Asn-Ile-Ser-Ala-Leu-Pro-Trp-
 WT CCT CAA GGG GTA ACA CCA GAA AAT CAT TTA AAT ATT TCA GCA TTA CCT TGG
 Kr CCG CAG GGC GTG ACA CCG GAA AAC CAT CTG AAC ATT TCT GCG CTG CCT TGG
 Sh CCC CAG GGC GTG ACG CCA GAA AAC CAT CTA AAC ATA TCC GCG CTG CCA TGG

154 Val-Asn-Phe-Asp-Ser-Phe-Asn-Leu-Asn-Val-Ala-Asn-Phe-Thr-Asp-Tyr-Phe-
 WT GTT AAT TTT GAT AGC TTT AAT TTA AAT GTT GCT AAT TTT ACC GAT TAT TTT
 Kr GTT AAT TTT GAT AGC TTT AAT TTA AAT GTT GCT AAT TTT ACC GAT TAT TTT
 Sh GTT AAC TTT GAC TCG TTC AAC TTA AAT GTA GCT AAC TTT ACT GAC TAT TTT

171 Ala-Pro-Ile-Ile-Thr-Met-Ala-Lys-Tyr-Gln-Gln-Glu-Gly-Asp-Arg-Leu-Leu-
 WT GCA CCC ATT ATA ACA ATG GCA AAA TAT CAG CAA GAA GGG GAT AGA CTG TTA
 Kr GCA CCC ATT ATA ACA ATG GCA AAA TAT CAG CAA GAA GGG GAT AGA CTG TTA
 Sh GCG CCG ATC ATA ACC ATG GCC AAA TAC CAA CAG GAA GGG GAT AGG CTC TTG

188 Leu-Pro-Leu-Ser-Val-Gln-Val-His-His-Ala-Val-Cys-Asp-Gly-Phe-His-Val-
 WT TTG CCG CTC TCA GTA CAG GTT CAT CAT GCA GTT TGT GAT GGC TTC CAT GTT
```

```
 Kr TTG CCG CTC TCA GTA CAG GTT CAT CAT GCA GTT TGT GAT GGC TTC CAT GTT
 Sh TTA CCC TTG AGC GTC CAG GTG CAT CAC GCC GTG TGT GAC GGC TTT CAT GTT

205 Ala-Arg-Phe-Ile-Asn-Arg-Leu-Gln-Glu-Leu-Cys-Asn-Ser-Lys-Leu-Lys-Ala-
 WT GCA CGC TTT ATT AAT CGG CTA CAA GAG TTG TGT AAC AGT AAA TTA AAA GCG
 Kr GCA CGC TTT ATT AAT CGG CTA CAA GAG TTG TGT AAC AGT AAA TTA AAA GCG
 Sh GCA CGC TTC ATT AAC AGA TTA CAG GAG TTA TGT AAT TCT AAA TTG AAG GCG

222 Ala-Asn-Asp-Glu-Asn-Tyr-Aal-Leu-Ala-Ala-***
 WT GCG AAC GAT GAA AAC TAT GCG CTG GCG GCG TAA
 Kr GCG AAC GAT GAA AAC TAT GCG CTG GCG GCG TAA
 Sh GCG AAC GAT GAA AAC TAT GCG CTG GCG GCG TAA
```

**Figure S3**

CAT coding sequences used in this study. Codon substitutions are highlighted. WT, wild type CAT; Kr, synonymous variant described previously [11]; Sh, Shuf1 synonymous variant.

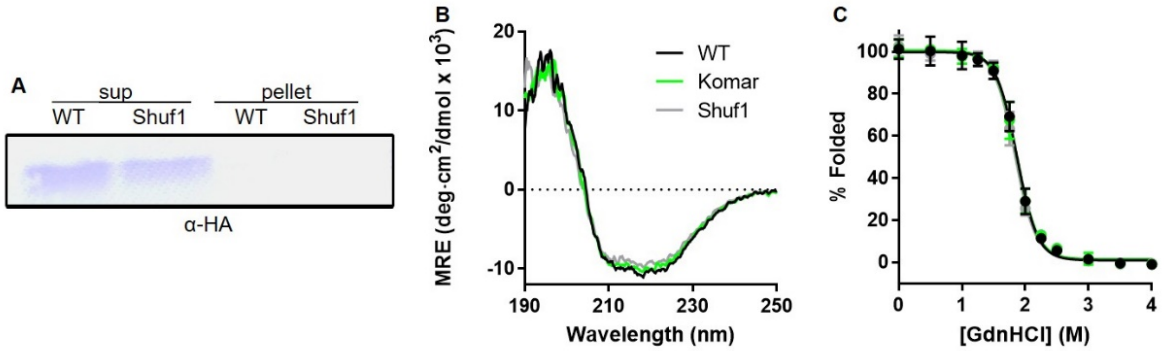


**Figure S4**

The native CAT protein structures expressed from the WT, Komar [11] and Shuf1 coding sequences are indistinguishable. (**A**) Partitioning of CAT-ssrA into soluble (sup) and insoluble (pellet) cell lysate fractions. (**B**) Far-UV CD spectra. (**C**) Guanidinium hydrochloride denaturation of CAT monitored as change in ratio of tryptophan fluorescence emission intensity at 330 versus 349 nm; n=3, data are represented as mean ± SD.

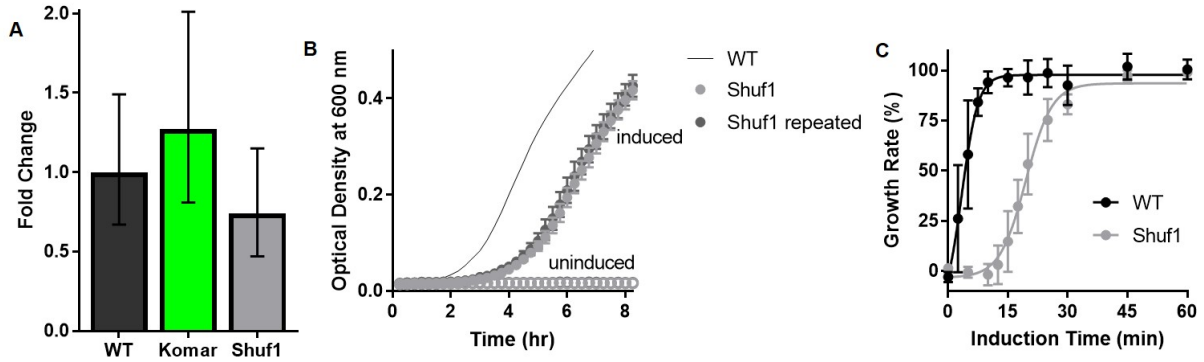


**Figure S5**

(**A**) Fold change in CAT mRNA levels relative to WT, as determined by RT-qPCR. (**B**) Growth curves of cells taken from the endpoint of a growth curve expressing Shuf1-CAT in cam compared to initial growth curves, with or without induction of CAT. (**C**) Growth rate of cells expressing CAT from the WT or Shuf1 coding sequences with varying induction time prior to addition of cam. More time is required to produce a sufficient amount of native CAT to support growth in the presence of cam when CAT expressed from the Shuf1 versus WT coding sequence; n=3 biological replicates, data are represented as mean ± SD.

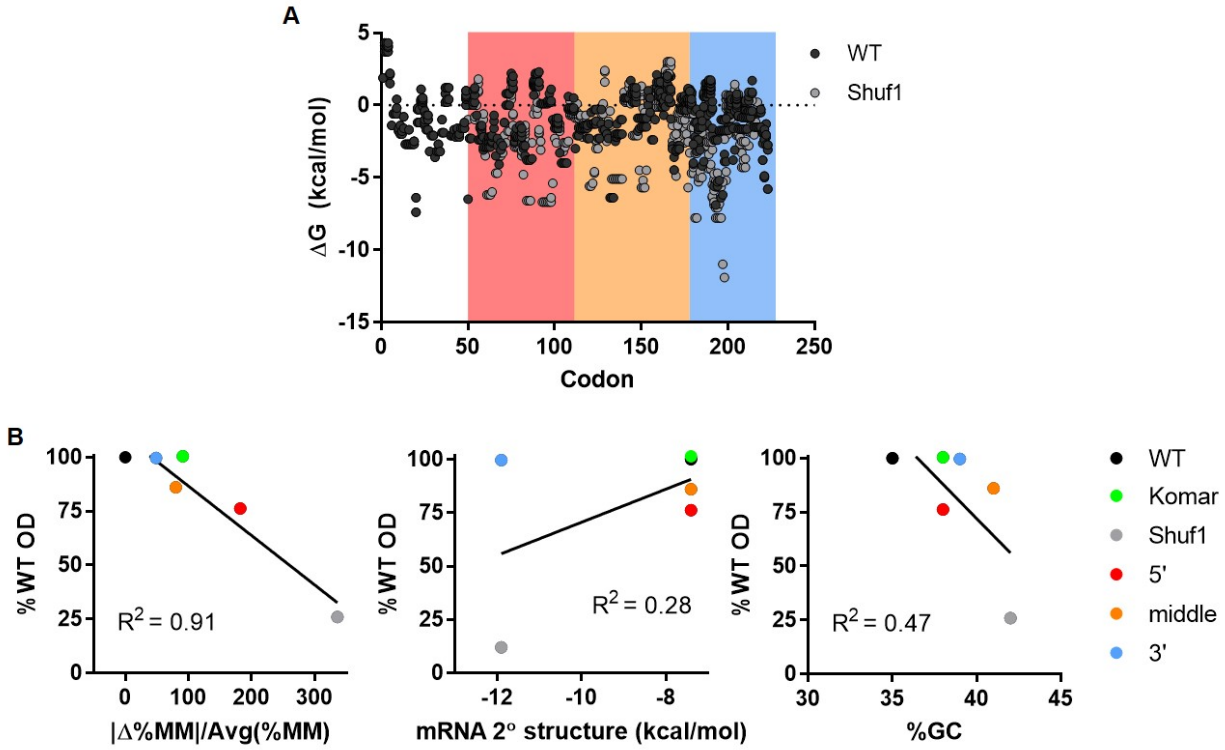


**Figure S6**

Growth defect correlates with codon usage and not mRNA structure or GC content. A) mFold predictions of local mRNA secondary structure over 30 nt windows. Background color indicate chimera construct boundaries. B) Correlation between change in %MinMax divided by the average %MinMax, strongest local mRNA secondary structure element, and %GC with growth. Growth is plotted as the %WT $OD_{600}$.

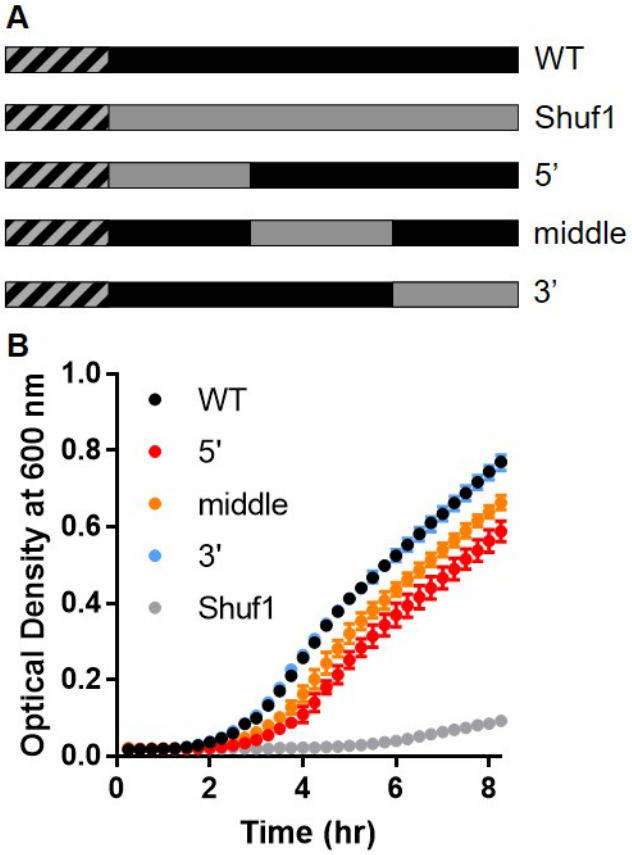


**Figure S7**

Growth of *E. coli* expressing CAT chimeras. (**A**) Schematic of chimeric mRNA sequences; black, WT; gray, Shuf1. The hatched portion represents the most 5' 50 codons, which were identical in each sequence. (**B**) Growth curves for the 5', middle, and 3' chimeras compared to WT and Shuf1. Data points represent the mean ± SD; n=3 biological replicates.