Translation efficiency is maintained at elevated temperature in *Escherichia coli*

Received for publication, October 5, 2017, and in revised form, November 22, 2017. Published, Papers in Press, November 28, 2017, DOI 10.1074/jbc.RA117.000284

Gareth J. Morgan§, David H. Burkhardt§, Jeffery W. Kelly‡, and Evan T. Powers‡†

From the ‡Departments of Chemistry and Molecular Medicine and §Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, and §California Institute of Quantitative Biosciences and Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, California 94158

Edited by Ronald C. Wek

Cellular protein levels are dictated by the balance between gene transcription, mRNA translation, and protein degradation, among other factors. Translation requires the interplay of several RNA hybridization processes, which are expected to be temperature-sensitive. We used ribosome profiling to monitor translation in *Escherichia coli* at 30 °C and to investigate how this changes after 10–20 min of heat shock at 42 °C. Translation efficiencies are robustly maintained after thermal heat shock and after mimicking the heat-shock response transcriptional program at 30 °C by overexpressing the heat shock σ factor encoded by the *rpoH* gene. We compared translation efficiency, the ratio of ribosome footprint reads to mRNA reads for each gene, to parameters derived from gene sequences. Genes with stable mRNA structures, non-optimal codon use, and those whose gene product is cotranslationally translocated into the inner membrane are generally less highly translated than other genes. Comparison with other published datasets suggests a role for translational elongation in coupling mRNA structures to translation initiation. Genome-wide calculations of the temperature dependence of mRNA structure predict that relatively few mRNAs show a melting transition between 30 and 42 °C, consistent with the observed lack of changes in translation efficiency. We developed a linear model with six parameters that can predict 38% of the variation in translation efficiency between genes, which may be useful in interpreting transcriptome data.

The regulation of the rate of protein synthesis is not completely understood. Cells continuously alter protein levels and stoichiometry to maintain a correctly-balanced proteome. In rapidly dividing cells such as *Escherichia coli*, the major process by which protein levels are reduced is dilution as the cell grows and divides (1). Therefore, steady-state protein levels are largely determined by differential synthesis. The regulation of transcription is the best-studied layer of expression control, but differential translation also plays an important role in the expression of many genes (2–8). Over a population of cells, mRNA levels generally correlate with protein levels, but there are large differences between individual proteins (3, 5, 9). A clear example in prokaryotes is the differential translation of individual genes from polycistronic transcripts (3, 7).

Control of translation has been intensely studied for many years, but there is little consensus on the relative roles of different factors and no way to predict how well-translated a particular sequence will be (8). Translation rate is generally determined by a combination of initiation and elongation rates, both of which are governed by hybridization of rRNA, mRNA, and tRNA sequences with each other. Therefore, a gene's mRNA sequence can potentially encode the rate of its translation as well as its amino acid sequence. Translational initiation is slower than elongation, so initiation is rate-limiting for the translation of most mRNAs (10).

Ribosome profiling by deep sequencing measures the distribution of ribosomes on mRNA sequences within a cell and hence how often a particular mRNA is translated (3, 6, 11). Ribosome profiling data are a population-averaged snapshot of which mRNA molecules are being translated. The ratio of ribosome-protected footprint read counts to total mRNA read counts for each gene is a measure of the relative rate of translation, defined as the translation efficiency (TE),² for that gene. Surprisingly, differences in TE measured by ribosome profiling correlate weakly, if at all, with sequence-specific factors that affect translation of individual genes, and the determinants of differential TE across transcriptomes remain unclear (3, 8, 12). Two recent papers came to strikingly different conclusions based on similar data: Burkhardt et al. (13) suggest that ORF-wide mRNA structure is the primary determinant of TE differences in the *E. coli* genome, whereas Del Campo et al. (14) suggest that mRNA structure has little effect on ribosome density and highlight the role of structure in a gene's 5′-untranslated region.

The role of RNA/RNA hybridization in controlling translation suggests that changes in temperature may differentially alter the translation of different mRNAs. A rapid rise in temperature induces a well-characterized transcriptional program

---

²The abbreviations used are: TE, translation efficiency; CPM, counts per kilobase per million; IMP, inner membrane protein; nt, nucleotide; TCE, trichloroethanol; Ni-NTA, nickel-nitritirotiacetic acid; SRP, signal recognition particle; RDM, rich defined medium; MM, minimal medium; CAI, codon adaptation index; tAI, tRNA adaptation index; RBS, ribosome-binding site.
E. coli translation at 30 and 42 °C

Table 1

| Strain/plasmid               | Condition                     | Replicates | Mapped mRNA reads | Mapped footprint reads |
|------------------------------|-------------------------------|------------|-------------------|-----------------------|
| K12 MG1655                  | 30 °C                         | 3          | 2286101.2         | 8745226.3             |
|                              |                               |            | 3389070.0         | 12268435.9            |
| K12 MG1655                  | 42 °C, 10 min                 | 3          | 2126384.0         | 12824504.5            |
|                              |                               |            | 1166227.8         | 10406361.1            |
|                              |                               |            | 2150863.9         | 8729665.6             |
| K12 MG1655                  | 42 °C, 20 min                 | 3          | 3563684.7         | 10737917.1            |
|                              |                               | (1 excluded)| 1289734.9        | 2728949.3             |
|                              |                               |            | 3373485.2         | 7860057.0             |
| K12 MG1655/empty pBAD       | 30 °C, 0.2% arabinose, 20-min induction | 4 | 369812.8          | 4473425.6             |
|                              |                               |            | 753555.8          | 8317688.9             |
|                              |                               |            | 1286659.0         | 7554082.3             |
|                              |                               |            | 4606106.5         | 5530391.1             |
| K12 MG1655/pBAD-rpoH wild-type | 30 °C, 0.2% arabinose, 20-min induction | 3 | 1136587.9          | 3418731.7             |
|                              |                               |            | 2096730.6          | 6476605.5             |
|                              |                               |            | 997466.2           | 3833833.4             |
| K12 MG1655/pBAD-rpoH I54N   | 30 °C, 0.2% arabinose, 20-min induction | 2 | 4974062.5          | 5657884.6             |
|                              |                               |            | 5940206.5          | 5945953.5             |

called the heat-shock response (15, 16), whereby expression of chaperones and proteases is increased to mitigate heat-induced unfolding and aggregation of the proteome. Production of new, misfolding-prone proteins is a major source of protein aggregation and toxicity during heat shock (17, 18). We reasoned that differential translational control may be important in rapidly increasing the translation of chaperones and proteases and by reducing the concentrations of heat-labile proteins, making heat shock a potentially useful tool with which to investigate TE. We therefore asked whether there are gene-specific TE changes at different temperatures.

Here, we use ribosome profiling to quantify the relationship between mRNA abundance and ribosome occupancy at 30 °C and under heat-shock conditions (42 °C) in the well-studied model organism, E. coli K12 MG1655. We find that TE for all measured genes is very similar between 30 °C and after 10 or 20 min of heat shock at 42 °C despite widespread changes in transcription and translation levels. mRNA structure and codon use both play significant roles in determining TE. RNA stability predictions suggest that few mRNAs undergo structural transitions in the temperature range studied. Unrelated to our original hypothesis, we did observe one striking and unexpected pattern: a distinctly lower TE for inner membrane proteins compared to cytosolic proteins, which we hypothesize is linked to cotranslational export from the cytosol.

Results

Translation efficiency varies across the E. coli genome

We sequenced total RNA and ribosome footprints from several sets of E. coli K12 MG1655 cultures growing exponentially in rich defined media (Table 1). We investigated the effect of heat shock by sequencing libraries from bacteria growing at 30 °C and after 10 and 20 min of growth at 42 °C. Any changes in translation that we observed could be directly caused by temperature-dependent RNA hybridization or by downstream effects of genes expressed at high temperature. To differentiate between these possibilities, we also compared bacteria expressing either wild-type or I54N σH protein from a pBAD plasmid, which mimics the heat-shock transcriptional program, to bacteria containing an empty pBAD vector at 30 °C (19). σH, encoded by the rpoH gene, is the RNA polymerase σ factor responsible for the transcription of the canonical heat-shock proteins such as the chaperones DnaK and GroEL (15). The activity of σH protein is inhibited by several factors, including DnaK, but this repression is alleviated by the I54N mutation (20, 21).

For each sample, mRNA and ribosome footprint cDNA libraries were prepared and sequenced. Reads were aligned to the MG1655 genome using Bowtie (22). Table 1 shows the total number of reads mapped to non-rRNA genes for each library. The number of reads per gene was calculated and normalized with EdgeR (23) to give mRNA and footprint counts per kilobase per million (CPKM) for each gene. Footprint reads were adjusted to remove the influence of the elevated ribosome density seen at the beginning of genes (3). We focus on protein-coding genes without unusual translational events such as frameshifting. For each gene, and for each replicate, we calculated TE, the ratio of footprint CPMK to mRNA CPMK. Homologous genes such as tufA and tufB, which encode the elongation factor EF-Tu, were considered as a single gene, and their counts were summed. CPMK values for both mRNA and footprints were reproducible across biological replicates (Fig. S1), although the footprint reads from one replicate of the 20-min heat-shock condition differ substantially from those of the other replicates, suggesting a problem with library construction, and the data were excluded from our analysis. Raw and normalized gene read counts for each library are shown in Table S1.

Under all conditions, global transcription and translation patterns are similar, with CPMK values varying over 1000-fold between genes (Fig. 1, A and B). Footprint levels correlate well with mRNA levels (Fig. 1, A and B, $R^2 = 0.80$ for 30 °C log-transformed data), indicating that transcript level is a primary determinant of the overall translation level. Translation effi...
Translation efficiency varies across the genome and within individual operons. Most genes have TE values within a 20-fold range (0.2–4), although extreme examples such as the yobF-cspC operon vary by more than 100-fold (Fig. 1, C and D). Translation levels for measurable proteins correlate with those observed previously (3, 6) and with two separate proteomic abundance measurements (Fig. S1) (24, 25).

**Translation efficiency is maintained at elevated temperature**

Because translation initiation is controlled by the interplay between different RNA hybridization events, we expected that temperature might differentially affect the TE of different genes. However, the measured TEs of genes do not significantly change between 30 and 42 °C (Fig. 2 A), despite changes in absolute translation levels between conditions. Variability between samples limits our ability to detect small changes in TE; however, the 1182 genes for which we obtained replicate data in both conditions (black points in Fig. 2 A) show the smallest differences in TE between those conditions. We looked for altered ratios of footprint and mRNA counts using EdgeR (23) and anota (26); neither method identified genes with significantly altered TE (data not shown). TE values measured in all conditions are very similar (Fig. 2 B), indicating that neither heat nor translation of the rpoH regulon affects translation.

Both heat and overexpression of σH lead to up-regulation of genes associated with the heat-shock response. Gene expression changes were similar to those observed previously (data not shown) (16). Fig. 3 shows plots of footprint versus mRNA counts for a selection of canonical heat-shock genes under different conditions. The strong induction of genes such as the chaperonin subunit groL indicates that the cells are responding to heat stress by transcribing genes from the σH regulon. Because the data in each case have a strong tendency to fall on a line, it is clear that translation efficiency (the slopes of the lines) is maintained at different temperatures and widely varying expression levels. Individual genes within polycistronic operons often have differing ribosome densities, indicative of differential translation (3, 6, 7). The patterns of translation within individual operons, which are very similar in these data to those described previously (3), are also maintained at different temperatures, as exemplified by the rpsP (27) and rpsM (28) operons in Fig. 4.

**Activation of σH is controlled post-translationally**

Thermal regulation of translation has been described for several *E. coli* genes (29–31). We therefore asked whether these observations are reflected in the ribosome-profiling data. The heat shock σ factor, σH, is regulated at several levels as follows: translation of its mRNA is limited by secondary structure, and the protein is rapidly bound by DnaK and the signal recognition particle (SRP) and delivered to the membrane to be degraded by the FtsH protease (21, 29, 32, 33). As the downstream targets of σH are produced, they compete with RNA polymerase apoenzyme for binding to σH, which turns off the transcriptional response. We can use transcription of the σH regulon genes as a measure of the protein’s activity and thereby distinguish between translational and post-translational regulation. Fig. 5 shows σH activity as measured by transcription of the groL and dnaK genes, which encode two major heat shock chaperones, as a function of its translation. At 30 °C, the activity of wild-type σH is only weakly dependent on its translation level, either from the chromosome or a plasmid. Translation of rpoH mRNA is similar at 30 °C and after 10 min at 42 °C (mRNA CPKMs of 393 ± 35 at 30 °C and 460 ± 26 at 42 °C after 10 min; footprint

---

**Figure 1. Translation varies across the genome and within operons.** Ribosome footprint levels are strongly correlated with mRNA levels at both 30 °C (A) and after 10 min of heat shock at 42 °C (B). Each point represents a gene. The blue lines show linear regression fits, and the dashed red lines are 1:1 ratios. The ratio of footprint to mRNA reads varies by over 100-fold across the genome, as illustrated by the yobF-cspC operon, highlighted in A and B. Histograms of translation efficiencies at 30 °C (C) and after 10 min heat shock at 42 °C (D) show similar distributions.
**E. coli translation at 30 and 42 °C**

**Figure 2. TE is maintained during heat shock.** A. plots of TE per gene at 30 °C versus TE after 10 min at 42 °C. Black points represent genes for which at least two replicate TE measurements were made at both temperatures. Red points represent genes for which only one measurement was made in one or both conditions. yobF and cspC are highlighted. B. Scatterplot matrix showing pairwise TE correlations between conditions. Panels on the diagonal are density histograms of TE distributions for each dataset. Panels below the diagonal show TE distributions. Panels above the diagonal show the $R^2$ values for each individual correlation.

CPKMs of 456 ± 86 at 30 °C and 518 ± 7.4 at 42 °C after 10 min. However, transcription of both groL and dnaK increases to a much greater extent after 10 min at 42 °C (open circles in Fig. 5) than is seen following the overexpression of wild-type $\sigma^H$ at 30 °C (solid circles in Fig. 5). After 20 min at 42 °C, translation of rpoH remains at a similar level, but the $\sigma^H$ protein activity decreases (open triangles in Fig. 5). Finally, direct overexpression of I54N $\sigma^H$ results in a large transcriptional activity (solid triangles in Fig. 5). This pattern is consistent with post-translational repression of $\sigma^H$ by DnaK and SRP being the primary means of control at ambient or mild heat-shock conditions.

E. coli has two other known RNA thermometers. Similarly to rpoH, the translation of the small heat-shock protein IbpA is known to be controlled by RNA secondary structure that occludes its ribosome-binding site at low temperature (30). We do not observe enough reads for ibpA to be able to reliably assess its translation, but its low expression in itself suggests that its translation is not activated. The cold-shock protein CspA’s mRNA contains a motif that activates its translation at low temperature (31, 34). However, cspA is strongly translated (TE of 4.6 ± 1.1 at 30 °C and 4.5 ± 0.77 after 10 min at 42 °C), in agreement with a previous observation that it is highly expressed at 37 °C (34).

**Translation from the open reading frame of ssrA increases during heat shock**

E. coli has several mechanisms to rescue ribosomes that have stalled on an mRNA molecule, the best-understood of which is the tmRNA/ssrA system (35, 36). The tmRNA molecule, encoded by the ssrA gene, binds to ribosomes with a stalled nascent peptide, which may be caused by an mRNA lacking a stop codon. The tmRNA molecule releases the ribosome from the mRNA and encodes for the translation of a short peptide tag that directs the resulting peptide for degradation by the ClpXP protease. ssrA mutants show a growth defect at 43 °C compared with wild-type strains (37). Ribosome footprint counts from the ORF portion of the ssrA gene increase following heat shock (Fig. 6A). To further investigate whether heat shock leads to increased tmRNA/ssrA activity, we used a reporter strain where the native ssrA ORF sequence is replaced with a hexahistidine peptide tag (37, 38). More His-tagged material was detected by immunoblotting of whole-cell extracts after 20 min at 42 °C compared with 30 °C (Fig. 6B). The proteins tagged by this strain are not degraded and can be purified Ni-NTA affinity chromatography. Levels of purified, His$_6$-tagged peptides, corrected for total cellular protein, are higher in heat-shocked cells than in cells grown at 30 °C (Fig. 6C). These data indicate that ribosome stalling and rescue increase during heat shock.

**Translation efficiency differences between genes are partly determined by ORF-wide mRNA structure**

Because translation efficiency does not change significantly between conditions, but varies widely between genes (Fig. 2), it must be regulated by the cell. There are several known genespecific factors that can influence translation in E. coli. Translation initiation is thought to be rate-limiting in most cases, and this is controlled by a combination of ribosome binding to mRNA, mRNA secondary structure, and codon use. We therefore examined the relationship between these metrics and TE, expecting that factors known to influence translation rate would correlate with differences in log(TE). We initially describe the results for the 30 °C dataset.

Recent work suggests that E. coli mRNA molecules are organized into ORF-wide structures, and that the extent of these structures determines TE (13). We calculated mRNA folding free energies using the Vienna RNA software package (39). We found a negative correlation between TE and the predicted stability of an ORF’s mRNA sequence, corrected for gene length ($\Delta G$ per nucleotide, Fig. 7A, $R^2 = 0.16$, Spearman’s $\rho = 0.43$). This correlation is weaker than that between TE and mRNA stability calculated using constraints from in vivo RNA structure measurements, $\rho = 0.52$ (13). Exclusion of genes for which only one replicate TE measurement was obtained at 30 °C resulted in an improved correlation between TE and $\Delta G$ per
E. coli translation at 30 and 42 °C

nucleotide ($R^2 = 0.20, \rho = 0.47$). The predicted mRNA stability of the translation initiation region (30 nt upstream and downstream of the gene’s translation start site) also correlates with TE (Fig. 7B, $R^2 = 0.11$), and these two mRNA stability parameters correlate weakly ($R^2 = 0.03$).

Other factors that influence translation initiation are less well-correlated with TE. Similarly to previous studies, we see no correlation between calculated ribosome-binding site strength and TE ($R^2 = 0.0$). The effect of a gene’s start codon is smaller than we expected, given the large effects observed by mutation of AUG start codons (40). Fig. 7C shows the distribution of translation efficiencies for genes as a function of their start codon. Although genes with non-AUG start codons are, on average, less well-translated than those with AUG start codons, the effect on TE is small (mean TEs of 1.31 for AUG versus 0.988 for non-AUG, $p < 10^{-10}$, Welch’s t test), and there is no significant difference in TE between UUG and GUG codons. We observe correlations between TE and the following two measures of codon use: the tRNA adaptation index (41, 42), tAI ($R^2 = 0.12$), and the codon adaptation index, CAI ($R^2 = 0.14$). CAI and tAI are strongly correlated ($R^2 = 0.70$) across the genome. We use tAI in preference to CAI for further analysis because tAI can be calculated from the populations of tRNA genes in a genome, whereas CAI is an empirical measure based on the codons used in highly-expressed genes (42).

Inner membrane proteins are significantly less well-translated than other classes of proteins

All E. coli proteins are translated in the cytosol but many are co- or post-translationally exported to the periplasm, inner or outer membranes, or secreted from the cell (43). Distributions of TE as a function of protein location, taken from the consensus locations defined in Diaz-Mejia et al. (44), are shown in Fig. 8A. Inner membrane proteins (IMPs) have significantly lower TE than proteins destined for any other cellular compartment (mean TEs of 0.801 for IMPs and 1.40 for others, $p < 10^{-15}$, Welch’s t test). Notably, outer membrane proteins and periplasmic proteins, which are also exported from the cytosol, have very similar patterns of translational efficiency to cytosolic proteins. The TE of a protein could be influenced by its mode of translocation. Most IMPs are cotranslationally translocated into the membrane via the SecYEG or YidC translocons, in a process dependent on the SRP (45). Bacteria also use post-translational pathways for export, which depend on the SecB chaperone. SRP clients have been identified using selective ribosomal profiling and are predominantly IMPs (46). The TE of SRP clients is significantly lower than that of non-clients, supporting the hypothesis that some feature of SRP-dependent synthesis affects translational efficiency (Fig. 8B). Cotranslational translocation of IMPs could potentially cause aggregation of ribosome-nascent chain complexes upon lysis of the bacteria, thereby depleting ribosomes that are translating IMPs and reducing apparent TE. The lysis buffer contains detergent (see “Experimental procedures”) in part to avoid this. We analyzed ribosomal profiling data on the Gram-negative bacteria Caulobacter crescentus (47) and found a similar reduction in TE for IMPs to that seen in E. coli (Fig. 8C), suggesting that this feature is conserved in at least these closely-related bacteria.

Temperature-dependent mRNA structural transitions are uncommon between 30 and 42 °C

The correlation between ORF structure and TE raises the question of whether general thermal unfolding of mRNA structures might be expected to increase TE. The lack of observed differences in TE at different temperatures suggests that this is not a major effect between 30 and 42 °C. To further investigate this, we calculated the temperature dependence of mRNA stability for the ORFs (Fig. 9, A–C) and initiation regions (Fig. 9, D–F) of all protein-coding genes using the RNA heat program from the Vienna package. The resulting melting curves (also known as thermograms) show peaks at temperatures where unfolding events occur. Example melting curves are shown in Fig. 9, A and D. The midpoint temperatures of these events (11,399 events in all genes and 5104 in genes with temperature-
dependent TE data for ORFs and 9233 events in all genes and 4024 in genes with temperature-dependent TE data for initiation regions are shown as histograms in Fig. 9, B and D. Most gene ORFs have similar predicted melting curves, with a large transition around 90 °C. Only 724 of 5104 ORF transitions in genes with temperature-dependent TE data occur between 30 and 42 °C, of which 262 have larger estimated enthalpies than background thermal fluctuations ($RT$, 0.6 kcal/mol at 30 °C, where $R$ is the ideal gas constant). This threshold excludes any structures that would transiently unfold at 30 °C and allows us to exclude small transitions in the melting curves such as that observed for $dnaJ$ (Fig. 9A). The genes whose ORFs have transitions with enthalpies estimated to be above $RT$ in the 30 – 42 °C range are highlighted in Fig. 9C. Similarly, 79 genes with temperature-dependent TE data have transitions with enthalpy greater than $RT$ in their initiation regions (Fig. 9F). Predicted transitions occur in genes with a wide range of TE (red points in Fig. 9, C and F).

These predictions are in broad agreement with the maintenance of TE for the majority of genes at these temperatures. This relative lack of thermal transitions suggests that mRNA structure may be selected to be insensitive to temperature changes in this range so that changes in TE with varying temperature are avoided. However, the RNA structure calculations are not well-calibrated for such long molecules, and these results may not be representative of the mRNA structural ensembles in living cells.

**Minimal linear model can predict trends in translation efficiency**

Although no one factor can reliably predict translation efficiency, it is possible that similar combinations of sequence- or gene-specific factors control TE of subsets of genes (5). We used parameters calculated from each gene’s sequence and their consensus locations to fit the 30 °C log-transformed TE data to a linear model. The parameters used were as follows: the
**Figure 6.** Translation from the coding region of the ssrA gene increases during heat shock. A, both mRNA and footprint levels increase after 10 min at 42 °C and remain elevated after 20 min at 42 °C. Colors represent replicate cultures. B, *E. coli* carrying an ssrA-his<sub>6</sub> reporter construct or an equivalent wild-type ssrA construct were grown at 30 °C and then either maintained at 30 °C or heat-shocked at 42 °C for 20 min. Cell extracts were run on SDS-polyacrylamide gels. Representative total protein (left) and anti-H<sub>6</sub> immunoblot (right) visualization is shown (n = 4). C, cell extracts of *E. coli* grown as for B were purified on Ni-NTA columns to isolate proteins tagged with a hexahistidine peptide by the tmRNA system. Eluted peptides were separated on an SDS-polyacrylamide gel and visualized by silver staining.

**Figure 7.** ORF mRNA structure, translation start region mRNA structure, start codon identity, and tRNA availability correlate with TE. A, correlation between ORF mRNA secondary structure stability (normalized for gene length) and TE. B, correlation between mRNA initiation region (±30 nt from initiation site) secondary structure stability and TE. C, smoothed density histogram of translation efficiencies for the three most highly-represented start codons. Curves are scaled so that their integrals are equal. Numbers of genes included in the analysis are indicated. D, correlation between gene tRNA adaptation index (42) and TE.
gene’s start codon; GC content; protein location; coding sequence length; predicted RBS-binding strength; mRNA folding free energy of the entire ORF, the initiation region (30 nt upstream and downstream of the start site, as for Fig. 9D), the 5’ UTR (120 nt immediately upstream of the ORF), the 5’ end of the ORF (120 nt downstream) and the combination of those regions; folding free energy divided by length of the whole ORF (free energy per nucleotide, as for Fig. 9A); and the calculated CAI and tAI for the full protein and its N-terminal 40 residues. Parameters and errors for the model are in Table S2. The
model's predictions correlate with the input data with an adjusted R² of 0.41, much better than the best individual parameter, tAI. However, not all of these parameters have a significant impact on the performance of the model. A simpler model that includes six parameters (GC content, tAI, the predicted ORF mRNA folding free energy per nucleotide, the predicted mRNA folding free energy of the initiation region, the presence of an ATG start codon, and whether a protein is predicted to localize to the inner membrane) correlated with its input data with an adjusted R² of 0.38. This indicates that more than one-third of the variation in TE between genes can be predicted by these parameters. To estimate how well ribosome footprint levels are predicted by the model, we compared the correlations between measured and predicted footprint CPKMs with measured mRNA CPKMs. Calculated footprint CPKMs correlate with measured mRNA CPKMs with an R² of 0.88, an improvement on the direct correlation between measured mRNA and footprint levels (R² = 0.81). Parameters for the model are in Table 2. Fig. 10 shows predicted TE values and the residuals for footprint levels predicted by a linear fit to the mRNA levels are shown in red. Curves are scaled so that their integrals are equal.

**Table 2**
Parameters for reduced linear model

| Parameter | Estimate | Standard error | Mean square error | F-statistic | p value | log_{10} p value |
|-----------|----------|----------------|------------------|-------------|--------|-----------------|
| Offset    | -0.582   | 0.157          | 63.0             | 230         | 1.37 \times 10^{-50} | -49.9 |
| GC        | 5.49     | 0.486          | 248              | 907         | 8.00 \times 10^{-10} | -179 |
| tAI       | 6.64     | 0.388          | 41.9             | 152.9       | 1.70 \times 10^{-34} | -33.8 |
| locIM     | -0.291   | 0.0234         | 8.42             | 30.7        | 3.19 \times 10^{-8}  | -7.50 |
| startATG  | 0.152    | 0.0274         | 215              | 784         | 7.36 \times 10^{-15} | -157.1 |
| mfePerNuc | 10.2     | 0.445          | 99.5             | 363         | 1.50 \times 10^{-77} | -76.8 |
| 5’ mfe    | 0.00451  | 0.00250        |                  |             |        |                  |

**Figure 10**. Linear model with six gene-specific parameters can explain 38% of the variation in translation efficiency. Translation data from three replicates at 30 °C were used to create a linear model to predict TE from gene sequence-specific data. A, predicted versus measured TE values. Each point represents a single measurement for a gene. The red line shows a linear regression fit of the data. B, fold differences between the predicted and measured footprint CPKM values are plotted as smoothed density histograms. The residuals for footprint levels predicted by a fit to the model are shown in blue, and the residuals for footprint levels predicted by a linear fit to the mRNA levels are shown in red. Curves are scaled so that their integrals are equal.

**Translation efficiency varies between growth conditions**

To investigate whether the gene-specific factors identified here can explain TE differences in other conditions, we analyzed published ribosome profiling data on *E. coli*. We calculated per-gene TEs from eight datasets (3, 13, 14, 48, 49) and measured the correlations between different conditions (Fig. 11). Details of each dataset are shown in Table 3. Three experiments on *E. coli* growing in rich defined media (RDM, see under “Experimental procedures”) measured TEs that correlated strongly with each other and with those measured in this work (R² values 0.58–0.77). However, weaker correlations were observed between data from *E. coli* grown in RDM and in either Luria-Bertani (LB) media or M9 minimal media (MM), either under optimal growth conditions or following heat or osmotic shock. The different *E. coli* K12 strains used in the experiments (MG1655 for the RDM experiments and MC4100 for others) may affect TE. However, the weak correlations between the LB- and MM-derived datasets suggest that other factors are responsible. Some of these differences in TE may also be attributable to differences in library preparation, sequencing depth, or data processing, although the correlations between RDM datasets from three different groups suggest that the ribosome profiling protocol itself is robust at the resolution of individual genes.

The minimal linear model derived for the 30 °C data was able to describe the data from *E. coli* grown in RDM with a similar degree of accuracy to the training data (Fig. 12, A, upper panels, and B). However, fitting data from *E. coli* grown in LB media at either 37 or 47 °C was less successful (Fig. 12, A, lower panels, and B). The LB- and MM-derived TE measurements correlate less well with tAI than the RDM TE measurements, but the correlations between all datasets and ORF mRNA stability are similar (Fig. 12C). We fit each dataset individually to a 21-parameter linear model as described for the 30 °C dataset above. The models could describe up to half of the TE variation in the


**Table 3**

Other datasets used in this study

The following abbreviations are used: RDM, rich defined media; LB, Luria-Bertani media; MM, M9 minimal media.

| Identifier (first author) | Refs. | Accession no. | Media | Temperature °C |
|--------------------------|-------|---------------|-------|---------------|
| Burkhardt et al.         | 13    | GSE77617      | RDM   | 37            |
| Li et al.                | 3     | GSE53767      | RDM   | 37            |
| Mohammad et al.          | 49    | GSE72899      | RDM   | 37            |
| Del Campo et al.         | 14    | GSE63817      | LB    | 37            |
| Bartholomäus et al.      | 48    | GSE68762      | LB    | 37            |
| Bartholomäus et al.      | 48    | GSE68762      | MM    | 37            |
| Bartholomäus et al.      | 48    | GSE68762      | MM+OS | 37            |

**Figure 11.** TE distributions vary between experimental conditions. Per-gene TE values were calculated from published data \(3, 13, 14, 48, 49\) and compared with TE values determined in this work. The scatterplot matrix shows pairwise TE correlations between conditions. **Panels on the diagonal** are density histograms of TE distributions for each dataset. **Panels below the diagonal** show TE correlations. **Panels above the diagonal** show the Pearson’s \(R^2\) values for each individual correlation.
Discussion

We have used ribosome profiling to measure transcription and translation in E. coli at 30 °C and after 10 and 20 min of heat shock at 42 °C. Translation rates are strongly dependent on mRNA levels, but the ratio of footprint to mRNA reads, TE, varies across the genome (Fig. 1). Because translation initiation is thought to be governed by RNA hybridization, we expected to observe widespread changes in TE at different temperatures. However, TE values were maintained at elevated temperatures (Fig. 2), despite large changes in mRNA levels between conditions (Figs. 3 and 4). A gene’s TE is correlated with both the predicted structure of the mRNA for a gene’s whole-coding region and the predicted structure of its mRNA in the translation initiation region, which for most genes are predicted to be maintained over the temperature range investigated here (Figs. 7 and 9). Inner membrane proteins tend to be translated at a lower rate than other proteins (Fig. 8). A simple linear model can predict one-third of the variation between translation of each mRNA (Fig. 10). This information is needed to incorporate heat shock into computational models of protein homeostasis such as FoldEco (50) and may be of use in predicting protein levels in other organisms.

This work uses a mild heat-shock protocol, which is insufficient to activate E. coli’s known RNA thermometers, although it does result in increased transcription of σH regulon genes (Figs. 3 and 5) and increased translation of the ssrA ORF (Fig. 6). The TE of those genes for which we obtained replicate mRNA and ribosome footprint data appears to be very robust to the temperature changes used here. It is possible that transient changes in translation occur before the 10-min time point measured here, because the transcriptional effects of the heat-shock response peak around 5 min after a temperature shift (51). It is likely that higher temperatures could cause greater changes in translation. Observations by two groups show that E. coli in LB media subjected to a temperature jump from 37 to 47 °C (48) or from 30 to 45 °C (52) do show translational changes in 129 and 115 genes, respectively, including members of the σH regulon. Genes whose translation increased upon heat shock had weak mRNA secondary structure content in their 5′ regions, suggesting that melting of these structures at high temperature may contribute to increased translation (48). TE of ibpA increased at 47 °C but TE of rpoH did not (48), supporting our observation that activation of σH is driven by transcriptional and post-translational mechanisms. However, only 12 genes are identified as differentially regulated by both groups. We measured replicate data at 30 and 42 °C for 30 out of 129 and 57 out of 115 genes, and we found that these genes did not exhibit an altered TE in our experiments. The lack of TE changes reported here may be partly due to low expression of genes with differential translational regulation under the conditions used here. E. coli strains typically exhibit growth defects above 45 °C (52), which may related to the differences in translation observed at higher temperatures than were used here.

Translation of inner membrane proteins

The observation that inner membrane proteins and SRP client proteins (Fig. 8) are translated less efficiently than other proteins may be due to the effects of cotranslational translocation into the inner membrane via the SRP pathway. However, it is possible that depletion of insoluble ribosome–nascent chain complexes during library preparation may be affecting the data. Translocation is faster than translation (43) and therefore unlikely to be rate-limiting. Moreover, SRP binding is not associated with translational pausing in E. coli (46). However, the geometry and steric constraints of the inner membrane may place a limit on the number of ribosomes that can simultaneously translate a particular gene. Polysomes in solution adopt a helical conformation that minimizes the space between ribosomes while maximizing the separation between emerging nascent chains (53). On a planar membrane surface, however, the density of ribosomes on an mRNA molecule may be limited by the packing of translocons and the requirement that the nascent chains are oriented in the same direction. This may be due to limiting numbers of translocons (54) or another factor. Nascent membrane proteins may also require physical separation to allow their helices to pack correctly and avoid inter-protein interactions.
contacts that might lead to aggregation. *E. coli* mRNA molecules that encode membrane proteins have been shown to segregate to the membrane independently of translation (55), and the as-yet undetermined factors responsible for this could also be involved in reducing ribosome density at the membrane.

**Predicting translation from sequence**

The coding and non-coding regions of mRNA can affect translation by several known mechanisms. mRNA structure and complementarity to the ribosome anti-Shine-Dalgarno sequence can affect both translation initiation and elongation (56). The genetic code is degenerate: most amino acids are encoded by several codons, but organisms use them selectively (41, 57). Codon use can alter the translational elongation rate as ribosomes wait for more or less abundant aminoacyl tRNAs (41, 57). Codon use can alter the translational elongation rate as ribosomes wait for more or less abundant aminoacyl tRNAs (41, 57–59), but different mRNA sequences also have different structural propensities which influence TE (58, 60–62). The 5′ end of open reading frames seems to be particularly important in determining how a gene is translated (58, 63, 64). The relative contribution of these effects is the subject of much research and debate (8, 65). Natural genomes have evolved under constraints maintained against temperature changes that might be expected to differentially alter RNA/RNA interactions. Prediction of the temperature dependence of mRNA structure suggests that there are relatively few structural transitions in the 30–42 °C range (Fig. 9), consistent with the hypothesis that *E. coli* has evolved to minimize temperature-dependent translational changes. Prediction of structures formed by long RNA molecules is challenging. The simple approach taken here has less predictive ability than the measurement of RNA structures in *vivo* or in *vitro*. Recent technical advances in such measurements (13, 69) should lead to better structure prediction algorithms, and restraints from these experiments can already improve predictions (13). However, the ability of the model presented here to predict TE from sequence-derived parameters could be useful for interpreting existing or new transcriptome data, whether from *E. coli*, other bacteria, or metagenomics studies. Because the necessary parameters are easily calculated or predicted from sequences, this approach suggests a way to refine estimates of protein levels from only a genome sequence. It remains to be seen how widespread these patterns of translational control are, both within *E. coli* strains and in other organisms.

**Experimental procedures**

The procedure for ribosome footprinting and cDNA library preparation was modified slightly from that published in Oh et al. (6). Detailed protocols are given in Refs. 70, 71.

**Growth conditions**

*E. coli* K12 MG1655 cells were grown in 200-ml cultures of EZ MOPS rich defined media (RDM, Teknova) at 30 °C, 200 rpm. A total of 18 cultures were grown in the batches shown in Table 1, starting from fresh overnight cultures of the same glycerol stocks of bacteria. For the temperature-shift experiments, cultures of untransformed cells were grown to *A*<sub>600</sub> of 1 in media without antibiotics at 30 °C, then diluted 3-fold into fresh media at either 30 or 42 °C, grown for 10 or 20 min, and then harvested at an *A*<sub>600</sub> of between 0.4 and 0.5. We used shaking water baths to stabilize the media temperature. For σ<sup>H</sup> expression, cultures were transformed with a pBAD vector containing either wild-type or I54N σ<sup>H</sup> or with an empty pBAD vector. Cultures were grown in media containing 0.1 mg/ml ampicillin to an *A*<sub>600</sub> of 0.2, induced with 0.2% arabinose, grown for a further 20 min, and then harvested at an *A*<sub>600</sub> of between 0.4 and 0.5. We used a high arabinose concentration to avoid the stochastic “all-or-nothing” activation of the arabinose promoter observed at the single cell level due to varying levels of the AraE transporter protein across populations (72).

**Harvesting and lysis**

Cells were harvested by rapid vacuum filtration through a 90-mm diameter 0.2-μm pore filter, scraped off the filter with a Scoopula, then immediately plunged into a 50-ml tube full of liquid nitrogen. Harvesting time from decanting to freezing was between 90 and 120 s. The frozen cells were scraped off the Scoopula into the bottom of the tube. Nitrogen was allowed to evaporate at −80 °C, and cell pellets were stored frozen at −80 °C until lysis, typically overnight. Frozen cell pellets were...
lysed at liquid nitrogen temperatures using a bead beater with steel tubes, silicone caps, and a single 5-mm steel ball (BioSpec). Frozen cell pellets were decanted into pre-cooled tubes containing 600 μl of frozen lysis buffer (100 mM NH₄Cl, 10 mM MgCl₂, 5 mM CaCl₂, 20 mM Tris-Cl, pH 8, 0.1% Nonidet P-40, 0.4% Triton X-100, 50 μg/ml chloramphenicol, 100 units/ml DNase I). All components were RNase-free. Chromatophenol in the lysis buffer stalled the ribosomes on mRNA. A pre-cooled ball was added, the tubes were capped and shaken at full speed on a Mini-Beadbeater-1 for 6 cycles of 10 s, and were re-cooled in liquid nitrogen for 45–60 s between cycles. This treatment gave satisfactory lysis of the cells without apparent thawing; increasing the number of cycles ran the risk of splitting the silicone caps.

Ribosome footprinting and polysome profiling

Cell lysate was thawed in the steel tubes at 30 °C, incubated on ice for 10 min, then transferred into 1.7-ml Eppendorf LoBind DNA tubes, and centrifuged at 14,000 rpm in a micro-centrifuge for 10 min at 4 °C. The supernatant was removed, and a 30-μl aliquot was snap-frozen and stored at −80 °C; this sample was used for total cellular RNA. Two 180-μl aliquots of the remaining lysate (~100 μg of total RNA each) were used for polysome profiling and ribosome footprinting. One aliquot was incubated with 2.3 μl of 33.3 mg/ml (43,000 units) of streptococcal nuclelease S7 (Roche Applied Science) for 1 h at 25 °C to produce ribosome-protected footprints; the other was incubated without the nuclease as a control to confirm that cells were translating and that the nuclease treatment had digested the mRNA to leave monosomes. The nuclease digestion was quenched with EDTA, and the lysates were cooled on ice. The lysates were immediately loaded onto a 10–40% sucrose gradient (in buffer containing 100 mM NH₄Cl, 10 mM MgCl₂, 5 mM CaCl₂, 20 mM Tris-Cl, pH 8, 100 mM chloramphenicol, and 2 mM dithiothreitol) in Beckman centrifuge tubes and centrifuged at 35,000 rpm for 2.5 h in a Beckman SW-41 rotor. Polysome profiles were measured by pushing the gradients out of the tube with 60% sucrose solution and monitoring RNA absorbance at 260 nm. The fractions corresponding to the center of the monosome peak were collected for the nuclease-digested samples, pooled, and frozen at −80 °C.

RNA extraction

RNA was extracted from the total RNA and monosome fractions by hot acid phenol chloroform (Thermo Fisher Scientific AM9720) (73). RNA was precipitated with isopropyl alcohol after adding sodium acetate and GlycoBlue (Thermo Fisher Scientific) as a coprecipitant. Total RNA was enriched for mRNA by purification with MegaClear (Thermo Fisher Scientific AM1908) and RiboMinus (Thermo Fisher Scientific K155004) kits to remove small RNAs and rRNA, respectively, following the manufacturer’s instructions. The remaining RNA was fragmented by incubation at 95 °C, pH 9.3, in sodium carbonate buffer for 40 min. Monosome fractions and fragmented mRNA were loaded onto a 15% polyacrylamide TBE-urea gel (Thermo Fisher Scientific) and run at 200 V for 1 h. Gels were stained with SYBR Gold (Thermo Fisher Scientific), and the band corresponding to footprint-sized oligonucleotides was excised. RNA was extracted from the gel by crushing the gel slice and shaking at 70 °C for 10 min. Gel fragments were removed by filtration, and the RNA was precipitated as above.

cDNA sequencing library preparation

The extracted RNA footprints and fragments were dephosphorylated by incubating with T4 polynucleotide kinase (New England Biolabs) without ATP. RNA was ligated to a 17-nucleotide adaptor (CTGTTAGCCACCATCAAT) (IDT) using truncated T4 RNA ligase. The ligated products were purified using a Zymo RNA cleanup column, which removed unligated adaptor and concentrated the RNA. Ligated RNA was reverse-transcribed by Superscript III (Thermo Fisher Scientific) using a primer complementary to the adaptor sequence, which also contained the sequences necessary for PCR amplification separated by a peptide spacer (IDT; sequence: 5′phos-GATCG-TCGGACTGTAGAAGCTCTGAGTGTGCGCCGTATCATT-isp18-CACTCA-isp18-CAAGCAGAAGACGCCATACGAATTGATGGTGCTACAG, where isp18 is a hexa-ethylene glycol spacer). Full-length cDNA was gel-puriﬁed as before from a 10% polyacrylamide TBE-urea gel. RNA was hydrolyzed in 0.1 M NaOH, pH 13, 95 °C for 40 min, leaving single-stranded cDNA. cDNA was circularized using CircLigase (Epiceri) to produce single-stranded and circular DNA molecules which included the two complementary sequences for PCR amplification needed to make Illumina sequencing libraries. Footprint libraries tend to be contaminated with specific rRNA sequences, which were removed by hybridization with 5′-biotinylated oligonucleotides (IDT; sequences: TCTCCGGGGGTAGACGTGTTCGTGCAGGGAGTACGACCTGCCTGTCG). Libraries were amplified by 8–12 cycles of PCR with Phusion polymerase (New England Biolabs). Primers encoded an Illumina TruSeq barcode sequence at the 5′ end of the insert to allow for multiplexing. Amplified libraries were gel-purified on 8% polyacrylamide-amine TBE gels; the sample that had experienced the largest number of cycles without showing large overamplification products was excised from the gel and extracted overnight at 20 °C.

Sequencing and alignment

Libraries were quantified by Agilent BioAnalyzer and pooled to give a final sequencing library containing 12 barcoded samples. Libraries were sequenced on an Illumina HiSeq 2500 at the sequencing core facility at The Scripps Research Institute. Each barcoded sample typically gave 10–20 million single-end, 100-bp reads. Base calling and demultiplexing were done with Illumina’s Casava software. Adapter sequences were removed from reads using the fastx_clipper application (developed by A. Gordon and G. J. Hannon) to leave footprint or fragment sequences. These were aligned to the E. coli K12 MG1655 genome (NC_000913.2) using Bowtie (22). Most samples resulted in between one and three million mapped reads; the unmapped reads were mostly due to contaminating rRNA,
E. coli translation at 30 and 42 °C

which was not removed from our libraries as successfully as was reported previously. The number of reads mapped to a particular nucleotide was counted using an in-house Python script (modified from that used in Oh et al. (6)) that averaged each read over the central nucleotides in the sequence. Because footprint read lengths are non-uniform in bacteria, the exact position of the ribosome peptidyltransferase site on each read cannot be precisely determined. The resulting .wig files were processed with in-house R scripts using a variety of analysis packages as described below.

**Calculation of per-gene CPKM values and translational efficiencies**

Reads were mapped to a list of protein-coding gene positions taken from EcoCyc (74). RNA genes, pseudogenes, and phage Ins elements were excluded from the list. Reads mapping to pairs of genes with similar sequences (tufA and tufB; gadA and gadB; yuaE and ydKF; ladA and ldrC; ybiD and yhli; tfqA and tfqQ; rzoD and rzoR; and pinR and pinQ) were aligned randomly to one homolog and the total counts used for determining mRNA and footprint levels, but they were excluded from further analysis. A meta-analysis of reads mapped to well-translated genes (those with at least 128 footprint counts) showed a similar enrichment of ribosome density at the 5' end of genes. For each footprint dataset, we corrected for this bias by dividing the read count at each codon within a gene by the normalized average ribosome density at that position from well-expressed genes (3). Read counts per gene were calculated from either the raw reads (for mRNA counts) or normalized reads (for footprints). We calculated CPKM for each gene in each experiment using the EdgeR package (23), which normalizes the counts per gene by the total number of aligned reads, then corrected for gene length. These CPKM values are the basic measurement of a gene's mRNA level and ribosome density and can be directly compared between datasets. Translation efficiency (footprint CPKM to mRNA CPKM) ratios were calculated pairwise for each gene in each experiment where that gene had at least 64 raw (unnormalized) counts for both mRNA and footprints. The mean TE for each condition was used to assess the influence of gene-specific parameters on translational efficiency.

**ssrA reporter**

*E. coli* strain CH2197, in which ssrA-H6 replaces ssrA in the X90 strain (37, 38), was grown at 30 °C in RDM to an A₆₀₀ of 1, then either maintained at 30 °C or heat-shocked at 42 °C by dilution as above. *E. coli* strain CH2195, in which wild-type ssrA is reintegrated into the ssrA knock-out strain, was used as a control. Cells were harvested by centrifugation and frozen. For immunoblotting, cells were resuspended in SDS-PAGE loading buffer (80 mM Tris-Cl, pH 6.8, 5% (w/v) glycerol, 2% (w/v) SDS, 0.1 mg/ml bromphenol blue), boiled for 5 min, and run on an SDS-polyacrylamide gel containing 0.5% trichloroethanol (TCE). Total protein was quantitated by reaction of TCE with tryptophan residues under UV light (75). Proteins were transferred to a PVDF membrane and blotted using an anti-hexahistidine antibody (Invitrogen catalog no. 372900) and visualized on a LiCor Odyssey scanner. For purification of His-tagged peptides, cells were lysed by resuspension in 8 M urea and purified on Ni-NTA spin columns (Thermo Fisher Scientific). Eluted material was diluted to normalize concentrations by total lysate protein concentration and then visualized by SDS-PAGE and silver staining.

**Gene-specific parameters**

Gene sequence data were taken from EcoCyc (74). We consider only protein-coding genes in this analysis, excluding genes with close homologs as above, selenoproteins, and proteins with frameshifts or stallings sites (fdhF, fdoG, fdnG, prfB, dnaX, secM, and inaC). RNA secondary structures were calculated with the Vienna RNA software package (39). We used the RNAfold program to determine minimal free energy values for every protein-coding ORF and initiation region (defined as the 30 bases upstream and downstream of the first base of the ORF) and the RNAheat program to calculate melting curves at a resolution of 1 °C. Melting transitions were detected by finding peaks in the data. The enthalpies of peaks between 30 and 42 °C were estimated by subtracting the mean heat capacity between 30 and 42 °C from the sum of the heat capacities between 30 and 42 °C. GC content, codon adaptation index, and tRNA adaptation index scores for each gene were calculated with the cac function from the seqinr R package (76) using data from Refs. 41, 57. *E. coli* protein locations were taken from the consensus data in Díaz-Mejía et al. (44). Ribosome-binding site calculations at different temperatures were calculated using the RBS Calculator software (77).

**Other ribosome profiling data**

Publicly-accessible data were downloaded from the GEO database https://www.ncbi.nlm.nih.gov/geo/or from the supplemental data of the relevant paper (3, 13, 14, 19, 48). Translation efficiencies were calculated from per-gene mRNA and footprint read counts, using the authors' normalized count data where available or normalizing with EdgeR as described above. Data for *C. crescentus* in rich (PYE) media was taken from Schrader et al. (47). Homologous genes were identified using the EcoCyc database. Protein localizations were from annotations in the BioCyc *C. crescentus* NA1000 database. The inner membrane class in Fig. 8C is a combination of the single-pass and multipass membrane proteins defined by BioCyc; peripheral membrane proteins were assigned to the cytosol.

**Linear modeling**

Linear models for translation efficiency were calculated using Wolfram Mathematica. Mean TE was fitted to a model incorporating all of the parameters calculated for each gene (Table S2). A second model was calculated using six parameters which contributed most to the fit of the first model (Table 2). We used the model to predict the ribosome footprint CPKMs for each gene given that gene's mRNA CPKM. The correlation between the predicted and measured footprint CPKMs is reported. To visualize the increased information content of the model, we compared the distribution of residuals from the fit of the model to the footprint counts with the residuals calculated from a simple linear regression of the mRNA and footprint CPKMs. The coefficients from the fit to the 30 °C data were used to calculate TE values for all genes, which were then com-

---

J. Biol. Chem. (2018) 293(3) 777–793
pared with measured values from other datasets. Each dataset was also fit with to the 21-parameter model with coefficients allowed to vary freely.

**Author contributions**—G. J. M., J. W. K., and E. T. P. conceptualization; G. J. M. data curation; G. J. M. formal analysis; G. J. M. investigation; G. J. M. visualization; G. J. M. and D. H. B. methodology; G. J. M. writing—original draft; G. J. M., J. W. K., and E. T. P. project administration; G. J. M., D. H. B., J. W. K., and E. T. P. writing—review and editing; D. H. B., J. W. K., and E. T. P. resources; J. W. K. and E. T. P. supervision; J. W. K. and E. T. P. funding acquisition.

**Acknowledgments**—We thank Dr. Jonathan Weismann, Dr. Carol Gross, Dr. Gene-Wei Li, and members of their laboratories for assistance with ribosome-profiling experiments; Dr. Andrew Su and Ryan Thomson for assistance with data analysis; and Dr. Howard Salis for the ribosome-binding site calculations. The ssrA-His6 reporter strain was a gift from Dr. Christopher Hayes of the University of California, Santa Barbara. We thank all of these individuals, Dr. Lila M. Gierasch, and members of the Kelly laboratory for helpful discussions and critical reading of the manuscript.

**References**

1. Nath, K., and Koch, A. L. (1970) Protein degradation in *Escherichia coli*. I. Measurement of rapidly and slowly decaying components. *J. Biol. Chem.* 245, 2889–2900

2. Wikström, P. M., and Björk, G. R. (1988) Noncoordinate translation—level regulation of ribosomal and nonribosomal protein genes in the *Escherichia coli* trmD operon. *J. Bacteriol.* 170, 3025–3031

3. Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635

4. Ohtaka, Y., and Spiegelman, S. (1963) Translational control of protein sequence determinants of protein abundance and noise in *Escherichia coli*. *Cell* 157, 624–635

5. Quax, T. E., Wolf, Y. I., Koehorst, J. J., Wurtzel, O., van der Oost, R., Ran, G. J. M., and Koller, D. (2014) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in *E. coli*. *Cell* 147, 1295–1308

6. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25

7. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140

8. Li, G.-W., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* 25, 117–124

9. Ishihama, Y., Schmidt, T., Rappshler, J., Mann, M., Hartl, F. U., Kern, M. J., and Frishman, D. (2008) Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* 9, 102

10. Larsson, O., Sonenberg, N., and Nadon, R. (2010) Identification of differential translation in genome-wide studies. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21487–21492

11. Wikström, P. M., Lind, L. K., Berg, D. E., and Björk, G. R. (1992) Importance of mRNA folding and start codon accessibility in the expression of genes in a ribosomal protein operon of *Escherichia coli*. *J. Mol. Biol.* 224, 949–966

12. Thomas, M. S., Bedwell, D. M., and Nomura, M. (1987) Regulation of α operon gene expression in *Escherichia coli*. a novel form of translational coupling. *J. Mol. Biol.* 196, 333–345

13. Morita, M. T., Tanaka, Y., Kodama, T. S., Kyogoku, Y., Yanagi, H., and Yura, T. (1999) Translational induction of heat shock transcription factor α32: evidence for a built-in RNA thermosensor. *Genes Dev.* 13, 655–665

14. Allen, S. P., Polazzi, J. O., Gierse, J. K., and Easton, A. M. (1992) Two novel heat-shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. *J. Bacteriol.* 174, 6938–6947

15. Giuliodori, A. M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Rompy, B., Gualerzi, C. O., and Pol, C. L. (2010) The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol. Cell* 37, 21–33

16. Tomoyasu, T., Gainer, J., Bukau, B., Kanemori, M., Mori, H., Rutman, A. J., Oppenheim, A. B., Yura, T., Yamanaoka, K., and Niki, H. (1995) *Escherichia coli* translation at 30 and 42 °C
coli FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor α32. *EMBO J.* 14, 2551–2560 [CrossRef] Medline

Tomoyasu, T., Ogura, T., Tatsuta, T., and Bukau, B. (1998) Levels of DnaK and DnaJ provide tight control of heat-shock gene expression and protein repair in *Escherichia coli*. *Mol. Microbiol.* 30, 567–581 [CrossRef] Medline

Brandt, A., Spurio, R., Gualerzi, C. O., and Pon, C. L. (1999) Massive presence of the *Escherichia coli* “major cold-shock protein” CspA under non-stress conditions. *EMBO J.* 18, 1653–1659 [CrossRef] Medline

Tu, G. F., Reid, G. E., Zhang, J. G., Moritz, R. L., and Simpson, R. J. (2014) The coding and noncoding architecture of the genome. *Nature* 516, 686–694 [CrossRef] Medline

dos Reis, M., Wernisch, L., and Savva, R. (2003) Unexpected correlations between gene expression and codon usage bias from microarray data. *Nucleic Acids Res.* 31, 59–69 [CrossRef] Medline

Mohammad, F., Woolstenhulme, C. J., Green, R., and Buskirk, A. R. (2016) Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2765–2776 [CrossRef] Medline

Brandt, F., Etchells, S. A., Ortiz, J. O., Elcock, A. H., Hartl, F. U., and Baumeister, W. (2009) The native 3D organization of bacterial polyomes. *Cell* 136, 261–271 [CrossRef] Medline

Matsuyama, S., Fujita, Y., Sagara, K., and Mizushima, S. (1992) Overproduction, purification and characterization of SecD and SecF, integral membrane components of the protein translocation machinery of *Escherichia coli*. *Biochim. Biophys. Acta.* 1122, 77–84 [CrossRef] Medline

Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., and Amster-Choder, O. (2011) Translation-independent localization of mRNA in *E. coli*. *Science* 331, 1081–1084 [CrossRef] Medline

Espah Borujeni, A., Channarassapa, A. S., and Salis, H. M. (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* 42, 2646–2659 [CrossRef] Medline

Sharp, P. M., and Li, W.-H. (1987) The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295 [CrossRef] Medline

Tuller, T., Carmi, A., Vestsigian, K., Navon, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354 [CrossRef] Medline

Sørensen, M. A., Kurland, C. G., and Pedersen, S. (1989) Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* 207, 365–377 [CrossRef] Medline

Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354 [CrossRef] Medline

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blüthgen, N. (2013) Efficient translation initiation dictates codon usage at gene start. * Mol. Syst. Biol.* 9, 675 Medline

Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A., and Roubovs, J. A. (2015) Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* 43, 3022–3032 [CrossRef] Medline

Zhao, W., and Ultsch, G. (2014) The anti-Shine-Dalgarno particle: an essential protein-targeting machine. *Adv. Protein Chem. Struct. Biol.* 86, 151–191 [CrossRef] Medline

Reddy, P., Peterkofsky, A., and McKenney, K. (1985) Translational efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 82, 5656–5660 [CrossRef] Medline

Janssen, B. D., and Hayes, C. S. (2012) The tmRNA ribosome-rescue system. *Adv. Protein Chem. Struct. Biol.* 86, 151–191 [CrossRef] Medline

Roche, E. D., and Sauer, R. T. (2001) Identification of endogenous SrSA-tagged proteins reveals tagging at positions corresponding to stop codons. *J. Biol. Chem.* 276, 28509–28515 [CrossRef] Medline

Hayes, C. S., Bose, B., and Sauer, R. T. (2002) Stop codons preceded by rare arginine codons are efficient determinants of SrSA tagging in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3440–3445 [CrossRef] Medline

Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *algorithms Mol. Biol.* 6, 26 [CrossRef] Medline

Akopian, D., Shen, K., Zhang, X., and Shan, S. (2013) Signal recognition particle: an essential protein-targeting machine. *Annu. Rev. Biochem.* 82, 693–721 [CrossRef] Medline

Díaz-Meja, I. J., Babu, M., and Emili, A. (2009) Computational and experimental approaches to chart the *Escherichia coli* cell-envelope-associated proteome and interactome. *FEBS Microbiol. Rev.* 33, 66–97 [CrossRef] Medline

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blüthgen, N. (2013) Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9, 675 Medline

Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A., and Roubovs, J. A. (2015) Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* 43, 3022–3032 [CrossRef] Medline

Tuller, T., Waldman, Y. Y., Kupiec, M., and Ruppin, E. (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3645–3650 [CrossRef] Medline

Tuller, T., and Zur, H. (2015) Multiple roles of the coding sequence 5′ end in gene expression regulation. *Nucleic Acids Res.* 43, 13–28 [Medline]

Espah Borujeni, A., Cetnar, D., Farasat, I., Smith, A., Lundgren, N., and Salis, H. M. (2017) Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* 45, 5437–5448 [CrossRef] Medline

Andersson, S. G. E., and Kurland, C. G. (1990) Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54, 198–210 [Medline]

Li, G.-W., Oh, E., and Weissman, J. S. (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484, 538–541 [CrossRef] Medline

Avellar-Kucukgoze, I., Bartholomäus, A., Cordero Varela, J. A., Kamil, R. F., Neubauer, P., Budisa, N., and Ignatova, Z. (2016) Discharging tRNAs: a tug of war between translation and detoxification in *Escherichia coli*. *Nucleic Acids Res.* 44, 8324–8334 [CrossRef] Medline

Ruusala, T., Andersson, D., Ehrenberg, M., and Kurland, C. G. (1984) Hyper-accurate ribosomes inhibit growth. *EMBO J.* 3, 2575–2580 [Medline]

Rouskin, S., Zubrardt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705 [CrossRef] Medline

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., and Weissman, J. S. (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550 [CrossRef] Medline

Becker, A. H., Oh, E., Weissman, J. S., Kramer, G., and Bukau, B. (2013) Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nat. Protoc.* 8, 2212–2239 [CrossRef] Medline

Siegel, D. A., and Hu, J. C. (1997) Gene expression from plasmids containing the araBAD promoter at saturating inducer concentrations
represents mixed populations. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8168–8172 CrossRef Medline

73. Ingolia, N. T. (2010) Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol.* **470**, 119–142 CrossRef Medline

74. Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Schröder, I., et al. (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, D605–D612 CrossRef Medline

75. Ladner, C. L., Yang, J., Turner, R. J., and Edwards, R. A. (2004) Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal. Biochem.* **326**, 13–20 CrossRef Medline

76. Charif, D., and Lobry, J. R. (2007) in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (Bastolla, U., Porto, M., Roman, H. E., and Vendruscolo, M., eds) pp. 207–232, Springer Verlag, New York

77. Salis, H. M. (2011) The ribosome-binding site calculator. *Methods Enzymol.* **498**, 19–42 CrossRef Medline

**E. coli translation at 30 and 42 °C**