Published in final edited form as:

Integr Biol (Camb). 2013 November ; 5(11): 1393–1406. doi:10.1039/c3ib40120k.

Changes in Translational Efficiency is a Dominant Regulatory Mechanism in the Environmental Response of Bacteria†

Ronald C. Taylor², Bobbie-Jo M. Webb Robertson², Lye Meng Markille³, Margrethe H. Serres⁴, Bryan E. Linggi¹, Joshua T. Aldrich¹, Eric A. Hill³, Margaret F. Romine³, Mary S. Lipton³, and Steven Wiley¹

¹Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352 USA

²Computational Biosciences Division, Pacific Northwest National Laboratory, Richland, WA 99352 USA

³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352 USA

⁴Bay Paul Center for Molecular Evolution, Marine Biological Laboratory, Woods Hole, MA 02543 USA

Abstract

To understand how cell physiological state affects mRNA translation, we used *Shewanella oneidensis* MR-1 grown under steady state conditions at either 20% or 8.5% O₂. Using a combination of quantitative proteomics and RNA-Seq, we generated high-confidence data on >1000 mRNA and protein pairs. By using a steady state model, we found that differences in protein-mRNA ratios were primarily due to differences in the translational efficiency of specific genes. When oxygen levels were lowered, 28% of the proteins showed at least a 2-fold change in expression. Transcription levels were significantly altered for 26% of the protein changes; translational efficiency was significantly altered for 46% and a combination of both was responsible for the remaining 28%. Changes in translational efficiency were significantly correlated with the codon usage pattern of the genes and measurable tRNA pools changed in response to altered O₂ levels. Our results suggest that changes in the translational efficiency of proteins, in part due to altered tRNA pools, is a major determinant of regulated alterations in protein expression levels in bacteria.

Introduction

How cells regulate the expression of proteins in response to environmental changes is one of the most fundamental questions in biology. At the most basic level, changes in the synthesis of specific transcripts can give rise to alterations in protein levels, but the multitude of steps between transcription and translation provide many different regulatory opportunities¹. Traditionally, regulatory mechanisms have been investigated at the level of individual proteins, but advances in high-throughput sequencing technologies and mass spectrometry

†Electronic Supplementary Information (ESI) available: Supplementary methods, Supplementary data, Supplementary figures S1–S4 and Supplementary tables S1–S5. See DOI: 10.1039/b000000x/

© The Royal Society of Chemistry [year]

⁵To whom correspondence should be addressed; Tel: 01 509 371 6551; steven.wiley@pnnl.gov.

Author contributions

HSW conceived the project; RCT, BWR, MSL and HSW designed research; MLM, JTA and EAH performed the experiments; RCT, BWR, BL, JTA, MHS, MFR and HSW analyzed the data; RCT, BWR, GS and HSW wrote the paper.

has enabled new global approaches for understanding protein expression². These approaches provide the potential for understanding larger, systems-level regulatory mechanisms that can provide insight into fundamental biological design principles³.

Several recent studies have revisited questions regarding the relationship between gene expression and protein levels. In general those studies have suggested that the correlation between the absolute levels of transcripts and protein levels is rather poor, although the reason cited for the lack of strong correlation is somewhat varied^{3–6}. Differences in translational efficiency between transcripts appear to play a prominent role^{7, 8}, although variations in protein turnover have been proposed to be important as well⁹. Some studies have shown that changing cell growth rates or cellular stress states can change the relationship between mRNA and protein abundance in cells^{10, 11}, but stress responses and changing growth rates are known to directly affect the translational machinery^{12, 13}. It has been suggested that in the case of yeast, translational efficiency of mRNAs can change in response to environmental alterations¹⁴, although later investigations have concluded that translational efficiency of individual transcripts are constant across conditions¹⁵. Thus, it is currently unclear whether changes in translational efficiency or protein turnover at the level of the individual transcript are a general cellular response to an altered environment.

An underlying technical challenge in understanding global regulatory mechanisms is the need to combine multiple types of frequently noisy high-throughput data, which can lead to the introduction of systematic errors⁶. Indeed, an analysis of mRNA-protein correlations has shown that much of the poor correlation is due to noise in the experimental measurements¹⁶. Some studies have addressed this issue by focusing on simplified experimental systems or a subset of transcripts/protein pairs so that errors can be reduced by the careful use of calibration standards^{3, 11}. Although these studies have shown the potential of current proteomics and transcriptional profiling technologies to understand regulation of translation², the application of these technologies to understand the regulation of protein expression in non-model systems has been limited.

To extend the usefulness of global proteomics and transcriptomics profiling for understanding the regulation of protein expression in response to environmental perturbations, we used cells grown under steady state conditions in chemostats¹⁷. Because the growth rate of cells in chemostats can be controlled independently of alterations in their environment, we can explore the response of cells independent of their proliferation rate. Furthermore, by using a relatively high exchange rate for the chemostats, we can create conditions that greatly reduce the effect of differential protein turnover rates on protein expression levels. We combined this reduction of experimental complexity with global quantification of both mRNA and protein expression using multiple biological and technical replicates. Quantification of mRNA levels was done using RNA-Seq¹⁸ whereas protein abundance was quantified using the Accurate Mass-and-Tag (AMT) approach¹⁹. Both measurement technologies are capable of producing quantitative measurements over an extremely broad dynamic range. By averaging multiple technical and biological replicates, we reduced the intrinsic noise of these measurements to greatly improve the reliability of our data sets.

Using this combined approach, we were able to confidently estimate the translational efficiency of over 1000 genes under different environmental conditions. We found that significant changes in protein translational efficiency were observed more frequently than changes in gene expression when environmental conditions were shifted. These changes in translational efficiency were correlated with changes in the codon usage frequency of the genes. In addition, the pattern of tRNA expression also appeared to change with environmental conditions. Our results indicate that changes in translational efficiency is a

major mechanism responsible for the modulation of protein abundance in bacteria and could be mediated by shifts in the abundance or modification of cellular tRNA pools. Because specific metabolic pathways and functions appear to be correlated with altered translational efficiency, it could represent a mechanism for optimizing bacterial function in response to altered environment conditions.

Results

Steady-state relationship between protein and mRNA levels

The conversion of an mRNA transcript into a protein is a complex dynamic process that involves multiple transcript-specific steps. If a cell is at steady state, however, the rate at which any given protein is synthesized is equal to its rate of degradation, or loss. Thus,

$$[Prot]_x K_{tx} = [mRNA]_x T_{effx} \quad [\text{Eq. 1}]$$

where $[Prot]_x$ is the steady state concentration of protein x , $[mRNA]_x$ is the concentration of its corresponding message, K_{tx} is the protein-specific turnover rate constant and T_{effx} is the translational efficiency of mRNA x . Rearrangement yields

$$\frac{[Prot]_x}{[mRNA]_x} = \frac{T_{eff}}{K_{tx}} \quad [\text{Eq. 2}]$$

Thus, the steady state ratio of an individual protein to its corresponding mRNA is simply the ratio of the translational efficiency of its message to the turnover rate constant of the protein. Note that both T_{eff} and K_t are “lumped” constants that include a variety of different processes. For example, T_{eff} includes ribosomal binding and codon-dependent translation rates²⁰. Similarly, K_t would include the rate of protein degradation and export from the cell (if relevant). Because both T_{eff} and K_t are generally gene product-specific, the relationship between the levels of proteins and their cognate messages should be different between various genes, as has indeed been observed^{3–6}. Importantly, if the protein:mRNA ratio of regulated genes remains constant at steady state following an environmental change, this implies that regulation of those genes is at the transcriptional rather than post-transcriptional level.

Equation 2 is only valid under steady state conditions, which is difficult to achieve at the level of individual cells, especially under batch culture conditions. However, chemostats can be used to achieve steady state growth of a population of cells by adjusting conditions so that the rate of cell proliferation in the chemostat is equivalent to the rate at which cells are removed from the system. Under these conditions, the rate at which total cell proteins are being produced is equivalent to the rate at which they are being lost due to both degradation and removal. In the case of bacteria, culture medium is pumped through the system at a relatively rapid rate, resulting in the turnover of the culture on the order of a few hours. Under these circumstances, the dilution rate of the chemostat, D , is much faster than typical turnover rates of bacterial proteins^{21, 22} and thus $K_t D$. Importantly, because the value of K_t for proteins from chemostat samples will essentially be the same, differences in the protein-mRNA ratios should be almost entirely due to differences in the translational efficiency of the different mRNAs.

To determine the utility of this approach, we used *Shewanella oneidensis* MR-1. This microbe has been studied extensively using chemostat cultures^{17, 23}, has been used for predictive metabolic modeling²⁴ as well as being investigated by both transcriptome and

proteome-based studies^{17, 25}. As a test perturbation, we grew *Shewanella* under both normal aerobic as well as low oxygen conditions (20% and 8.5% air saturation). Cell proliferation rates were maintained constant using a dilution rate of 0.1 h^{-1} for both conditions. At steady state, triplicate samples were taken 10–24 h apart and subjected to both transcriptional profiling using RNA-Seq and proteomics analysis (Fig. 1). To ensure the robustness of the RNA preparation, different methods for processing the RNA-Seq samples were evaluated (e.g. with or without rRNA depletion). As detailed in Supplemental Methods and Data, few if any significant differences were observed in the results using alternate sample RNA preparation approaches (Fig. S4). However, we did find lower variation in the data between samples using rRNA depletion, most likely because of the higher number of mRNA-specific reads ($\sim 10^9$). Significantly, the variance between the ratios of technical replicates was similar to that of biological replicates (0.03 versus 0.05) whereas the variance between the ratios across conditions was relatively high (17.8), indicating that within the accuracy of our measurements, the cells are at steady state (also see Figs S3E and S3F). Using deep sequencing of rRNA-depleted samples, we could detect expression of all 4436 annotated protein-encoding genes reported for *Shewanella* under at least one condition, with a dynamic range of approximately 6 orders of magnitude (Table S1). We observed 795 genes that displayed at least a 2-fold change in expression in response to altered oxygen concentrations across the three biological replicates ($p < 0.05$, 2-tailed T-test), with about half associated with detectable proteins ($N=373$; see below). Lowering the O_2 concentration from 20% to 8.5% resulted in a halving of cell density (from 0.52 OD_{600} to 0.26 OD_{600}). Using Gene Ontology annotations, we observed a significantly decreased expression of genes involved in catabolic processes and increased in those associated with nucleic acid salvage and biosynthetic pathways (Table S2). There was also a decrease in the expression of genes involved in oxidoreductase activities using heme donors (GO:0016675) and an increase in oxidoreductase activities using aldehyde donors (GO:0016903). Thus, changing the O_2 levels resulted in a significant change in cell metabolism and gene expression even though cell growth rates remained constant.

Proteomics measurements showed a high degree of consistency between samples at the peptide level, with the coefficient of variation between ratios of biological replicates essentially identical to those between technical replicates (37 versus 32; see Figs S3A–D). Reproducibility of measurements improved considerably when we integrated the peptide measurements to generate absolute protein estimates, with the coefficient of variation between the ratios of all the biological replicates averaging 2.2. We observed a total of 28,326 unique peptides that could be mapped to at least 2822 different proteins, 2197 of which were identified by more than a single peptide. Although we observed peptides corresponding to only 64% of the genome, they accounted for 87% of all detected mRNA transcripts by mass.

In conjunction with the proteomics and transcriptomics measurements, we also performed a compositional profile of the samples including a determination of cell number, dry mass, total protein and RNA as well as the distribution of the RNA species between tRNA, rRNA and other noncoding species. We used this information to convert the proteomics and transcriptomics measurements into absolute abundance measurements, with a likely average error range of approximately 2-fold (see Supplemental Methods and Data for details and assumptions). Converting the RNA-Seq measurements to absolute number of mRNA molecules per cell yielded a range of approximately 3,200 to 4,100, with the higher values being associated with cell growth under low O_2 conditions (see Supplemental Methods and Data for details). Data from peptide measurements were used to estimate protein abundance using a variation of the intensity-based absolute quantification (iBAQ) approach³. In brief, the sums of the average ion intensity of protein-specific peptides were normalized against protein size and total protein content of the cell, yielding an estimated protein dynamic range

of 6 orders of magnitude. Converting the estimated protein abundance to absolute number of molecules yielded a value of $2.6\text{--}2.7 \times 10^6$ proteins per cell, similar to the value of 2.35×10^6 proteins per cell estimated for *E. coli*²⁶. Calculations of protein copies per mRNA yielded mean values of 570 and 650 for cells grown under anoxic and aerobic conditions, respectively, similar to the value of approximately 540 for *E. coli*²⁷.

Data on the estimated absolute levels of each mRNA and protein species was filtered in a progressive fashion, as outlined in Fig. 2, and used to evaluate both the relationship between mRNA and protein levels under specific conditions and between conditions. Applying stringent data filters generally excluded low abundance proteins, mostly because the probability of repeatedly observing a given peptide dropped together with abundance. Because the levels of specifically regulated proteins can also vary considerably between conditions, very stringent data filters also tended to remove regulated proteins. To maintain as many protein observations as possible, we implemented a progressive filter, removing lower quality data only as necessary for the particular analysis (see below and Supplemental Methods and Data for details).

To estimate any potential biases in our dataset, we compared the genes present in our filtered proteomics and transcriptomics datasets using the top-level functional subsystem annotations from the SEED database (<http://www.theseed.org>), which were used to assign genes to one of 12 functional categories (see Table S3 for classification). The distribution of the number of genes and the levels of expressed mRNA and expressed protein from the unfiltered datasets were then calculated. As shown in Fig. 3, the fraction of genes assigned to each category ranged from 4% to 19%. At the level of mRNA expression, there was a significant underrepresentation of genes encoding carbohydrate and amino acid metabolism, cofactor and lipid biosynthesis and respiration. Genes encoding motility and chemotaxis functions were overrepresented at the mRNA level, but underrepresented at the protein level. Because many of them were extracellular and membrane proteins (e.g. pili and flagella), their relatively lower protein levels could be due to loss during sample collection or resistance to the enzymatic digestion needed for the proteome analysis²⁸. Genes involved in protein translation (e.g. ribosomal proteins), were overrepresented at the mRNA level and more so at the protein level. Except for cell signalling proteins, which were mostly absent from the proteome, the relative distribution of mRNA and protein expression was similar in the other categories (within a factor of 2).

The progressive filter that we used on our data sets did not cause a significant change in the relative distribution of either mRNA or proteins between functional categories, with the exception of the genes assigned to the cell signalling classification. These genes showed some of the lowest levels of expression at the mRNA level and were essentially removed by the abundance filtering. Thus, with perhaps the exception of the cell signalling genes, the relationships between mRNA and protein levels that we see in the filtered datasets are similar to those observed in the unfiltered ones.

In our unfiltered data set, the relationship between steady state mRNA and protein abundance was highly variable (Fig. 4A), ranging from ~650,000 protein molecules per mRNA molecule to less than 1. Some of this high dynamic range is likely to be artifactual due to underestimation of specific mRNA and protein species. When an abundance filter was used to remove low-quality data points (see below), the protein to mRNA ratio was reduced to between ~30,000 - 30. The overall Pearson's correlation coefficient (cc) between mRNA and protein abundance in the first aerobic sample (cc = 0.56, $P < 0.001$) is similar to what has been reported in other studies^{3, 7, 29}. Spearman's rank-order correlation coefficient was very similar ($r_s = 0.55$, $P < 0.001$). As shown by the dotted circles in Fig. 4A, filtering the data to remove low-confidence data points (see below) improved the correlation between

mRNA and protein only slightly ($cc = 0.59$, $P < 0.001$). The correlation between mRNA and protein abundance for the other 5 samples was very similar, with the Pearson's correlation coefficient ranging between 0.53–0.57.

To determine whether there is a general relationship between mRNA and protein abundance, as modelled by Eq. 2, the observed abundances of both mRNA and protein for each gene in the unfiltered data sets were ranked by mRNA abundance and divided into 25 bins of approximately 100 genes each. The average of the mRNA and protein abundance for each bin was then calculated. As shown in Fig. 4B, there is a strongly linear correlation between average mRNA and protein levels above an average transcript abundance of 0.05 copies per cell (Pearson's cc : ~ 0.98 , P -value < 0.001). Below this level, protein levels appear uncorrelated with mRNA levels, likely because most of the low-abundance proteins are represented by only one or two peptides, and because of errors in measuring low-abundance mRNAs (Table S1). From this data, we established our abundance threshold for high quality data to be > 2 peptides and > 0.1 mRNA molecules per cell. The slope of the mRNA versus protein plot, which should be proportional to average T_{eff} values, was 0.87 for cells grown in normal O_2 and 0.76 for cells grown under low O_2 , indicating that our alterations in growth conditions have only a small effect on the average translational efficiency of proteins.

Effect of environmental change on transcription versus translational efficiency

It has been previously reported that changes in protein abundance in bacteria do not always correlate with changes in gene expression^{11,30}. Our ability to reliably estimate values of both T_{eff} and transcript levels for a large number of proteins across two conditions allowed us to determine the extent to which protein expression is regulated at the level of transcription versus translational efficiency. Our RNA-Seq measurements showed an average percent coefficient of variation (%CV) of between 20–25% for transcripts corresponding to detected proteins across all biological replicates. However, there was more noise in the protein measurements, with an average %CV of 48% across all the biological replicates. Variability between proteomics measurements was more pronounced for low abundance proteins, likely due to their lower number of associated peptide measurements.

To generate a set of high-quality proteomics data to compare across conditions, we eliminated proteins that had a %CV $> 33\%$ across all three biological replicates. For log-normal distributions, this would provide a minimum 95% confidence limit for < 2 -fold variation in expression. This resulted in a reduced set of 1,258 high-quality protein-mRNA pairs (Fig. 2; Table S1). Despite the bias towards higher abundance proteins in the filtering step, the resulting set of proteins spanned nearly 5 orders of magnitude of abundance. Although this set of high-confidence protein-mRNA pairs showed very little variation across biological replicates (Figs 5A, 5B), there were significant differences in their expression under different experimental conditions (Figs 5C, 5D), with greater than a 2-fold change in at least 296 different mRNA species and 357 proteins (Table S1). The percent of O_2 -regulated mRNA observed for protein-mRNA pairs ($\sim 23\%$) was higher than all mRNA species as a whole ($\sim 14\%$), indicating that genes that produce detectable proteins were preferentially regulated.

When we correlated changes in mRNA expression with changes in protein expression, we obtained a Pearson's correlation coefficient of only 0.49 (Fig. 5E). Interestingly, this is worse than the correlation between absolute mRNA and protein abundance ($cc = 0.56$, Fig. 4A). The lack of strong correlation between changes in mRNA and protein levels in our steady state chemostat cultures indicates that T_{eff} is changing as well. Indeed, when we compared the distribution of changes in mRNA, protein and T_{eff} between our two steady state cultures, we observed similar magnitude of changes (Fig. 5F). This suggests that

alterations in post-transcriptional processes can significantly alter the level of proteins generated from cellular mRNAs.

To facilitate the exploration of transcriptional versus post-transcriptional regulation of the >1200 protein-mRNA pairs in our data set, we visualized the data using an approach where changes in gene transcription and T_{eff} are simultaneously displayed as ratios of the values obtained under two conditions (Fig S5). When there are no changes in mRNA or protein expression, all data points will cluster in the centre of the graph within the limits dictated by data noise (Fig. 6, top left panel). In agreement with this notion, data from our biological replicates formed an oblong cluster in the centre of the plot (Fig. 6, top right panel), with the lateral stretch reflecting the greater variability in the proteomics versus the transcriptomics measurements. Almost all of the data was clustered within a 2-fold range from the centre, indicated by the blue box in the figure. In contrast, data comparing relative transcript levels and T_{eff} from cells grown under normal versus low O₂ conditions were shifted from the centre of the plot (Fig. 6, bottom left panel), indicating changes in both mRNA levels and T_{eff} for a large fraction of the proteins. To provide a visual indication of the reliability of the different data points, we encoded the number of peptides used to estimate protein abundance in terms of the size of the data point and encoded the %CV of the data in terms of colour. Thus, larger and darker data points should be more reliable than smaller and lighter ones. Significantly, many of the shifted data points are from highly confident data.

These “double-bubble” plots are particularly effective for spotting behavioural trends of different groups of genes and proteins. For example, when integrated mRNA and protein data specific for ribosomal proteins were plotted in the double-bubble format, they formed a diagonal pattern (Fig. S5, bottom right panel), suggesting that ribosomal protein levels remained relatively constant despite changes in their underlying mRNA levels. A close examination of the data showed that the average level of ribosomal proteins was approximately 10% higher at normal versus low O₂ conditions (9,200 versus 8,400 copies per cell) whereas the corresponding average ribosomal protein mRNA level was 40% lower (7.2 versus 12.2 copies per cell). A similar pattern was seen for translational initiation and termination factors (Table S1). This data is consistent with previous studies on post-transcriptional mechanisms that regulate the levels of ribosomal proteins and their associated partners³¹.

Out of the 1,258 filtered gene-protein pairs used in this analysis, 376 (or about 30%) showed a significant (e.g. > 2-fold), change in T_{eff} as compared to 297 genes (24%) that displayed a significant change in transcription. Of the genes that showed significant changes in mRNA levels, only 43% of the time was this accompanied by a significant change in protein expression due to compensatory changes in T_{eff} . For example, the average 2.9-fold increase in the expression of 111 genes was accompanied by no significant change in the corresponding proteins because of a concurrent decrease in translational efficiency (Table S1).

Of the proteins that showed at least a 2-fold change in expression (356 out of 1258 proteins, or 28%), transcriptional regulation alone was responsible for 92 (26%) and changes in T_{eff} were responsible for 163 (46%). In particular, the absolute fold change for T_{eff} across these 356 proteins was significantly larger than for the corresponding transcription level changes based on a Wilcoxon signed rank test (P-value < 1.5×10^{-5}). A combination of both mRNA levels and T_{eff} was responsible for the remainder 101 (28%). Thus, changes in T_{eff} constitute a major mechanism that dictates protein expression levels in response to environmental changes in *Shewanella*.

There were 376 proteins that displayed a significant change in T_{eff} as determined by a two-tailed t-test (P-value <0.05) and by a >2-fold change. These included enzymes, transporters, regulators, and structural proteins. Their functions were analysed using BioCyc metabolic pathways and SEED cellular role categories. We detected an increased T_{eff} for 269 proteins for cells grown at 20% versus 8.5% O₂. Approximately 17% of these proteins were directly involved in transcription or translation, including enzymes involved in RNA methylation; MetK for conversion of L-methionine to S-adenosyl-L-methionine and the enzymes TrmJ and RlmB for methylation of tRNA and 23S rRNA, respectively. The group also included respiratory proteins, such as cytochromes, cytochrome c reductase, and CymA, a key component of MR1's respiratory chain^{32,33}. Some of the proteins were involved in degradation of amino acids (alanine, glutamate and isoleucine), known carbon sources for MR1. Others were associated with motility and sensing of environmental signals (e.g. flagella proteins and methyl-accepting chemotaxis proteins). Proteins responding to several types of stresses (cold, heat, phage, oxidative, iron depletion) also had increased T_{eff} . In addition, the group included two NAD-dependent malate dehydrogenases converting malate to either pyruvate (SfcA) or oxaloacetate (Mdh), and their T_{eff} increased 4 and 3 fold, respectively.

The remaining 107 proteins had higher T_{eff} under low O₂ conditions. Some were related to respiration and energy metabolism; nitrate reductase, formate dehydrogenase, NADH-ubiquinone reductase, and ATP synthase, likely responding to the reduced oxygen level. Proteins for amino acid biosynthesis, rather than degradation, had increased T_{eff} under low O₂ conditions, including those involved in the biosynthesis of histidine, isoleucine, leucine, proline, threonine as well as uptake of arginine. Proteins involved in sensing and responding to the external environment also had increased T_{eff} (e.g., TonB receptors and methyl-accepting chemotaxis receptors). One of these receptors is a member of the ArgR regulon, and might have a role in arginine acquisition.

A pathway enrichment analysis of the proteins with altered T_{eff} , showed that the higher T_{eff} under aerobic conditions (20% vs 8.5% O₂) affected lipoate biosynthesis (P-value 0.002), isoleucine degradation (P-value 0.007), S-adenosyl-L-methionine cycle (P-value 0.026), and gluconeogenesis (P-value 0.026). A higher T_{eff} under low O₂ conditions appeared to affect leucine biosynthesis (P-value 4.14×10^{-5}), histidine biosynthesis (P-value 0.007), proline biosynthesis (P-value 0.009), and pyrimidine deoxyribonucleoside degradation pathway (P-value 0.028), the latter possibly reflecting an increase in nucleotide degradation under low oxygen conditions.

Some of the changes in T_{eff} that we observe between cells grown at normal versus low O₂ conditions are likely to be gene specific. For example, it is known that ribosomal proteins can bind and regulate the translation of their own transcripts^{31,34}. We also observed several operons that appeared to display specific post-transcriptional regulation. An example is the *hypAEDCB* operon that encodes the NiFe hydrogenase maturation complex³⁵. When cells were shifted to low oxygen conditions, the mRNA of this operon went up by approximately 2 fold. However, the corresponding protein levels went up 15–40 fold (Table S1). Proteins from flanking operons or those that encode the NiFe hydrogenase itself showed no disparity between transcript and protein levels, suggesting that translation of the polycistronic message encoding HypAEDCB is specifically regulated.

Although many of the changes that we observe in T_{eff} between cell grown under normal versus low O₂ conditions are likely to be specific, it also seemed possible that global mechanisms could be important as well. For example, a major determinant of T_{eff} is thought to be the codon usage bias (CUB) of the genes, because of the correlation between specific tRNA abundances and their frequency in abundant proteins^{36–38}. It has recently been shown

that the pattern of codon usage of highly expressed genes is altered when cells are grown under different environmental conditions, suggesting that the tRNA pools might be dynamically regulated³⁹. To investigate this possibility, we quantified the expression profiles of the tRNA pools of cells grown under either normal or low O₂ by using a lower RNA size range selection prior to constructing RNA-Seq libraries. The reads corresponding to each tRNA gene were then quantified.

As shown in Table 1, cells grown under normal aerobic conditions displayed a different pattern of measured tRNA expression than cells grown under low O₂ conditions, with 18 out of 34 (~53%) being significantly different at a P-value of 0.05. Some tRNA transcripts showed decreased expression under low O₂ conditions, such as those encoding tRNA-Asp (6-fold decrease) whereas other showed an increase in measured expression, such as tRNA-Cys. However, there were no obvious changes in the pattern of amino acids incorporated into proteins under low O₂ conditions that correlated with changes in the tRNA pools, with the exception of cysteine, which increased by 41% (Table S4). We also did not observe any changes in the expression of specific amino acid tRNA synthetases that corresponded with changes in the tRNA pools. In general, changes in the measured levels of any particular tRNA isotype was due to coordinate changes in the measured levels of all of its constituent gene copies (Table S4).

Although the quantitative relationship between the levels of tRNA measured by RNA-Seq and levels of functional tRNA isoforms is not clear, these data are consistent with the idea that different environmental conditions can change the composition of the available tRNA pools which could then change T_{eff} because of the codon usage pattern of specific genes. If this were the case, then we would expect that changes in T_{eff} be correlated with a specific codon usage pattern.

To test this hypothesis, we first created a high-confidence list of genes and their associated T_{eff} by eliminating proteins represented by either few peptides (<3) or low-abundance mRNAs (<0.1 copy/cell). After eliminating proteins that were expressed under a single conditions and those that were too short to allow for accurate CUB estimation⁴⁰, we were left with a set of 854 proteins, representing ~86% of the protein mass and ~64% of the mRNA transcripts of the cells (see Fig. 2). We then looked at the relationship between changes in T_{eff} and CUB using a linear model in which the dependent variable was the log₂ change in the T_{eff} values of each gene under different conditions ($\log_2(\delta T_{eff})$) and the independent variables being the codon frequencies of the genes normalized to codons per thousand. By using linear regression analysis of all 61 codons, we found that the frequency of one or more codons was indeed predictive of $\log_2(\delta T_{eff})$, with a P-value of less than 6.8×10^{-14} (F-statistic = 4.52). The model had an overall Root Mean Square Error (RMSE) of 0.84. There was a significant relationship (P-value < 1×10^{-50}) between the true observed and fitted $\log_2(\delta T_{eff})$ values of the model (cc = 0.51; Fig. 7A). As a point of comparison, the relationship between changes in mRNA levels and protein levels (Fig. 5E) showed similar correlation values (cc = 0.49).

The predictive nature of the regression model was not due to a single codon or set of related codons as the correlation coefficient between the frequency profiles of all pairs of codons ranged from -0.39 to 0.49. In addition, the correlations between any single codon and $\log_2(\delta TE)$ were small (-0.23 < cc < 0.19). Stepwise regression identified a 9-codon model that was as predictive as the full 61-codon model (Table 2). The RMSE of this model was 0.87 versus 0.84 for the full model. Although the correlation was not quite as high, there is a clear relationship between these 9 codons and the observed changes in translation efficiency within this experiment (Fig. 7B). Together, these data support the idea that the CUB of

genes is a significant determinant of the extent to which T_{eff} changes in response to environmental perturbations.

Discussion

There are numerous mechanisms that regulate the response of microbes to a change in their environment. These include changes in the community composition as well as alterations in the complement of proteins that dictate the behavior of individual cells. Most investigations on how individual cells are regulated have focused on transcriptional control mechanisms, but there are other levels that could be important, such as translational control and protein turnover. The lack of suitable technologies to reliably screen for alternate regulatory mechanisms on a global level has made it difficult to determine how common alternate mechanisms are and the role they might play in regulating microbial responses. To address this challenge, we developed an integrated transcriptomics-proteomics analysis approach to define the relative roles of gene transcription and translational efficiency in dictating steady state levels of proteins in a bacterial system.

As a test case, we used *Shewanella* grown at steady state in chemostats under normal aerobic or low oxygen conditions and performed quantitative global proteomics and RNA-Seq measurements. By using chemostats, we were also able to greatly reduce the confounding effect of protein turnover rates on mRNA-protein relationships. Other investigators have attempted to infer the turnover rate of bacterial proteins from protein:mRNA ratios obtained from cells grown in chemostat cultures⁹. However, careful measurements of protein turnover in continuous culture have shown that only 2–7% of the proteins turn over with a half life of < 1h, with the remaining being extremely long lived⁴¹. Thus, the turnover rate of the chemostat can be considered the de facto turnover rate of most cellular proteins, providing us with a simple approach for identifying factors that influence translational efficiency of native *Shewanella* proteins.

We observed a total of 28,326 unique peptides that mapped to at least 2822 proteins, of which 2197 were characterized by more than a single peptide. Although these corresponded to only 64% of the postulated 4436 protein-encoding genes in *Shewanella*, they corresponded to 87% of the mRNA transcripts by mass. The use of multiple biological and technical replicates allowed us to reduce the affect of experimental noise on our estimates of T_{eff} values. Even when we filtered our data sets to eliminate proteins found in only one condition or characterized by low abundance mRNA (<0.1 copies/cell), low abundance protein (<3 peptides) or showing a %CV of >33%, we were still left with a group of 1051 protein-mRNA pairs that represented >90% of all of the protein mass and ~70% of the mRNA mass of the cells. The size, quality and diversity of this protein-mRNA data set allowed us to critically assess important relationships between gene expression and protein expression levels.

The results of our study are consistent with a growing body of literature that suggests that post-transcriptional mechanisms are extremely important in dictating the levels of proteins that arise from a given amount of mRNA^{2, 3, 10, 20, 27, 30,42}. We found that the correlation between mRNA and protein expression in *Shewanella* was relatively good at the level of individual genes (cc = 0.56), and that 31% of the steady state relationship between genes and proteins could be explained by transcript abundance alone. A variety of different mechanisms have been described that could contribute to the remainder of the difference between transcript and protein abundance, such as ribosome binding^{43, 44} and codon usage³⁸. Binning the data into 25 equal groups ranked by absolute levels of gene expression averaged out much of the noise in the data and yielded a strong linear correlation (cc ~0.98; Fig. 4B), supporting the proposed linear relationship between protein concentrations and T_{eff}

(see eq. 1). Such a relationship implies that a change in mRNA levels should give rise to a directly proportional increase in protein levels. However, there was actually a very poor correlation between changes in mRNA and protein expression in response to lowering oxygen levels (cc ~0.49), suggesting that T_{eff} is not a constant and that alterations in it could be responsible for significant changes in protein levels.

It has been suggested that translational efficiency is better described as a saturable rather than a linear process²⁰. If this is the case, then T_{eff} should decrease as mRNA levels increase and visa versa. Indeed, we observed a negative correlation between changes in mRNA levels and T_{eff} (Fig. S6; slope = -0.41, cc = -0.37, $p < 0.001$), suggesting that this effect could be significant contributor to the poor correlation between changes in mRNA levels and protein abundance. For example, of the 297 mRNA transcripts that showed a significant change in their level of expression, only 128 (43%) were accompanied by a significant change in the protein they encoded. However, saturation kinetics cannot account for changes in T_{eff} in the absence of a corresponding change in mRNA abundance. Importantly, of the proteins that showed at least a 2-fold change in expression (356 out of 1258 proteins), only 54% showed a significant change in the level of the corresponding transcript whereas 74% showed a significant change in T_{eff} . Thus, it appears that a change in translational efficiency independent of mRNA levels is a dominant mechanism for altering protein levels in bacteria.

One of the most interesting observations in our study was the non-concordance between changes in the expression level of rRNA, ribosomal proteins and their encoding mRNAs. Because ribosome capacity is typically limiting for growth, increased microbial growth rates are usually associated with an increase in rRNA synthesis and the number of ribosomes per cell¹². Thus, the synthesis of rRNAs and ribosomal proteins are thought to be tightly coordinated. When *Shewanella* is shifted from normal to low O₂ in a chemostat, however, cell density decreases while the growth rate remains constant. Consistent with a constant growth rate, the number of ribosomal proteins per cell remained constant. However, we observed an approximately 40% drop in the level of rRNA and a paradoxical 70% increase in the levels of ribosomal protein mRNAs. The levels of rRNA present at normal O₂ levels is sufficient to assemble approximately 5700 ribosomes per cell, while we estimate that there are between 5000–9600 copies of ribosomal proteins per cell (see Supplemental Methods and Data). The levels of rRNA under low oxygen conditions, however, are sufficient for only 3,800 ribosomes, much lower than the estimated levels of ribosomal proteins. This suggests that ribosomal proteins are in excess and that rRNAs could be limiting for ribosomal biogenesis under low O₂ conditions. Despite the total number of ribosomal proteins remaining constant between conditions, following the shift to low O₂ conditions, they fall from constituting approximately 10% to 8% of total detectable proteins concurrently with a 22% drop in the average T_{eff} of all transcripts. The increase in mRNA encoding ribosomal proteins under low O₂ conditions could be a mechanism to compensate for a context-dependent fall in their translational efficiency.

The mechanisms that give rise to alterations in T_{eff} are likely to be complex and to involve both global and message-specific regulatory factors. It has previously been suggested that a shift in cell environment could change the pattern of codon usage through alterations in either the expression or modification of specific tRNAs^{39, 45}. Consistent with this idea, we observed a significant change in the pattern of detectable tRNAs in *Shewanella* shifted from 20% to 8.5% O₂. The observed changes in tRNA levels are unlikely to be an artifact of sample preparation because they were very reproducible across multiple, independently generated RNA-Seq libraries (Table S4). However, the underlying mechanisms that gave rise to the observed change in measurable levels are not known. Many tRNAs are subject to extensive post-transcriptional modifications, which could potentially interfere with the

ligation and amplification technologies used in their detection. Thus it is not possible to conclude that a 6-fold drop in the measured levels of specific tRNAs, such as the product of the tRNA-Asp gene, accurately reflect a change in abundance or instead reflect a change in chemical modification.

The mechanistic basis underlying the relationship we observe between codon usage pattern and changes in T_{eff} are also not clear. The codons that are most strongly associated with changes in T_{eff} are neither rare nor obviously associated with synonymous codon usage (Table 2), although most do appear associated with changing tRNA pools (Table 1). Changes in tRNA modification could underlie changes in apparent tRNA abundances as well as the association of codon usage pattern to changes in T_{eff} ⁴⁵. For example, the expression of tRNA (uridine-5-oxyacetic acid methyl ester) 34 synthase (SO_2435) was increased 2–3 fold at both the mRNA and protein level when cells were shifted to low O₂ (Table S1). This enzyme is responsible for modification of uridine in position 34 (the wobble position) of tRNAs, and is critical for efficient decoding of G-ending Pro, Ala, Ser and Val codons⁴⁶ and to allow reading of codons for those amino acids that end in C (e.g. CCC (Pro) and GCC (Ala))^{47, 48}. Interestingly, recognition of 4 of the top 9 codons associated with changes in T_{eff} values is altered by uridine-5-oxyacetic acid modifications (Table 2).

Altered competition between different transcripts for a limited pool of ribosomes due to changes in transcript abundance could also affect the T_{eff} values of genes⁴⁹. At the relatively low ratio of ribosomes to transcripts that we observe under our experimental conditions, however, T_{eff} should be more strongly affected by tRNA abundance⁵⁰. The CUB of the cohort of expressed genes is thought to reflect evolutionary adaptation for optimal growth under a particular environmental condition¹. Organisms such as *Shewanella* display a high degree of metabolic flexibility, indicating that they have evolved to thrive under a variety of environments⁵¹. Optimizing the translation of mRNA into protein under multiple conditions would present a particular challenge and would probably necessitate the evolution of multiple levels of control. Regulation of tRNA pools and their enzymatic modification could be part of a multi-tiered mechanism by which protein expression is optimized in response to environmental changes. Any alterations in the complex translational machinery would register as a change in translational efficiency, necessarily modifying the relationship between mRNA levels and protein abundance. Consistent with this idea, it has been shown that bacteria growth rates can change global physiological parameters that affect both the efficiency of transcription factors⁵² and the translational machinery¹⁰. However, our work demonstrates that a change in growth rate per se is not required to change global physiological control mechanisms. Rather, any significant alteration in the physiological state of the cell will likely suffice.

Conclusions

Our studies also suggest that although many microbial proteins are regulated at the level of transcription, posttranscriptional mechanisms appear to have a stronger affect on net protein abundance. This highlights the needs to include proteomics measurements in the analysis of microbial systems. The integrated and quantitative approach we have developed for identifying the level at which proteins are regulated provides a foundation for more detailed mechanistic studies to understand how cells control protein expression and for explorations of why T_{eff} changes under different cell environments.

Materials and Methods¹

Cell culture

A 15-liter New Brunswick Bioflow 3000 reactor (New Brunswick Scientific, Edison, NJ) operated at a 7-liter working volume at 30°C was used to grow chemostat cultures of *S. oneidensis* MR-1 in modified M1 media supplemented with 18mM lactate. The gas flow rate and agitation were kept at 3.5 liters/min and 350 rpm, respectively, and the dissolved oxygen tension (DOT) was maintained at 20% or 8.5% by automatically changing the ratio of N₂ and air in the gas mix. The pH was maintained at 7.0 by addition of 2 M HCl. Culture growth was constantly monitored by measuring and recording optical density using a custom-made system²⁴. All oxygen concentrations in the manuscript are expressed as a percentage of air saturation.

Reactors were inoculated with 1mL/L of overnight culture grown in TSB without dextrose and maintained in batch mode until late-logarithmic stage. Continuous cultures were started by pumping medium of the same composition at desired dilution rate (D). Samples for further analysis were taken after at least five volume changes at steady-state conditions. Steady-state was inferred from stability of the following parameters: OD₆₀₀ (no more than 3% variation between measurements), acid addition rate, O₂ concentration in incoming gas, and O₂ and CO₂ concentrations in the off-gas. Samples for ash-free dry organic weight (AFDW) were taken and analyzed as described²⁴. The sampling procedure took no more than 4 minutes.

Sample preparation for transcriptome analysis

The RNA was extracted using Invitrogen TRIzol® Reagent followed by genomic DNA removal and clean up using Qiagen RNase-Free DNase Set kit and Qiagen Mini RNeasy™ kit. The Agilent 2100 Bioanalyzer was used to assess the integrity of the RNA samples. All the RNA samples in this experiment have an RNA Integrity Number between 9 and 10. Invitrogen's RiboMinus™ Eukaryote kit for RNA-seq and bacterial probe from Invitrogen's RiboMinus™ Transcriptome Isolation Kit (Bacteria) were used to remove the 16s and 23 s rRNA from the RNA samples⁵³. The SOLiD® Total RNA-Seq Kit was used to construct template cDNA for RNA-Seq and this was performed accordingly to the protocol recommended by Applied Biosystems. Briefly, the mRNA was fragmented using chemical hydrolysis followed by ligation with strand-specific adapters and reverse transcribed to generate cDNA. Fragments of cDNA between 150 to 250 bp were subsequently isolated by electrophoresis in 6% Urea-TBE acrylamide gel. For analysis of tRNA, size selection was between 70 to 250 bp. The gel-isolated cDNA went through 15 cycles of amplification to produce enough templates for the SOLiD™ EZ Bead™ system to generate template bead library for ligation base sequencing on the SOLiD™ instruments.

Transcriptomics

Three technical replicates were used for cell samples grown under normal aerobic conditions (20% O₂) or low O₂ conditions (8.5% O₂), for a total of six RNA-Seq runs on a Life Technologies SOLiD 4 sequencer, following rRNA depletion. Also, replicate runs were performed on two additional biological samples from each of the two conditions (eight additional runs). Two of the normal aerobic O₂ samples (samples 2 and 3; Fig. 1) were also run in duplicate before rRNA depletion (four additional runs). These 12 additional runs were performed on a SOLiD 5500xl sequencer. Mapping to the *Shewanella oneidensis* strain MR-1 reference genome (RefSeq Assembly ID GCF_000146165.2) was performed using

¹Further information on methods can be found in Supplemental Materials and Data.

the default settings in the Life Technologies software provided for the SOLiD sequencers (BioScope ver. 1.3 for the SOLiD 4, LifeScope ver. 2.5 for the SOLiD 5500xl).

Sample Preparation for Proteomics

To measure protein abundance, four mass spectrometry (MS) analyses were run on the six biological samples obtained from anoxic and aerobic conditions with four technical replicates for each sample (total of 24 runs). The *Shewanella oneidensis* MR-1 cell pellets were reconstituted in 1.0 mL of 100 mM NH_4HCO_3 , pH 8.4 buffer and transferred to a 1.5-mL siliconized microcentrifuge tubes. Half of each sample was then transferred to a PULSE-ND tube and the cells were then lysed with a PBI Barocycler NEP 3229 using FT500 PULSE Tubes for ten cycles, holding at 35000 psi for 20 seconds each time. The cell lysates were mixed to a homogenized state and the volume was determined using a pipette. The sample concentration was determined with a BCA Protein Assay using a microplate reader. The sample was then brought to 8M urea and 10mM dithiothreitol. The sample was incubated at 60°C for 30 minutes and was then diluted 10-fold with 100 mM NH_4HCO_3 to reduce the salt concentration and brought to 1 mM CaCl_2 . The sample was digested for 3 hours with trypsin at 37°C at a concentration of 1 unit trypsin/50 units protein. After trypsin incubation, the sample was snap frozen and placed in a -80°C freezer overnight. Sample was then cleaned up using a 1 mL/50 mg C18 SPE column using the following steps: Conditioned column with 3 mL of MeOH then rinsed the column with 2 mL acidified water (0.1% TFA). The sample solution was slowly passed through the column then the column containing the sample was washed with 4 mL of 95:5 H₂O:ACN, 0.1% TFA. The column was allowed to go to dryness and the “needles” below the column were blotted dry. Collection tubes were placed under the column and while tubes were closed off from vacuum, 1 mL of 80:20 ACN:H₂O, 0.1% TFA was added. The elution buffer flowed slowly through the column until dry. The sample was removed from SPE vacuum chamber and concentrated in a speed-vac to a volume of 50µL. A sufficient amount of 25 mM NH_4HCO_3 was added to dissolve any particulate matter on the tube sides, and a BCA protein assay was performed on the sample to determine the final concentration. The sample was then quick frozen in liquid nitrogen and stored at -80°C until analysis.

Mass Spectrometry Sample Analysis

Each MS experiment was done on a Thermo Fisher Scientific LTQ-Velos Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) coupled with an electrospray ionization (ESI) interface using etched tips made in-house. Peptide samples were analyzed by LC-MS using a previously described high pressure LC system⁵⁴ with a 60 cm long, 75 µm inner diameter fused silica capillary (Polymicro Technologies, Phoenix, AZ) packed in-house with 3-µm Jupiter C18 stationary phase (Phenomenex, Torrance, CA). Mobile phase A consisted of 0.1% formic acid in water and mobile phase B, of 80% acetonitrile/20% water/0.1% formic acid. The LC column was coupled via electrospray ionization to the mass spectrometer⁵⁵. The Accurate Mass and Time (AMT) tag approach^{19,56,57} was applied to produce quantitative peptide abundance data. Rollups of the peptide abundance data into protein measurements were performed using a modification of the ‘intensity-based absolute quantification’ (iBAQ) method³. Using the iBAQ technique, we summed the average intensity counts (across technical replicates) of all peptides matching to a specific protein and then corrected for the average number of expected peptides to provide a final value as a proxy for the protein level (see Supplemental Methods and Data for details).

Data Integration and Analysis

Transcriptomics and proteomics text files were parsed and uploaded into tables in a data warehouse constructed using Hadoop and HBase on a PNNL Linux cluster. Combined data

sets, annotated with background information on the genes and genome also stored in the central warehouse, were then exported in a variety of formats from the warehouse into software tools for statistical analysis and visualization (see Supplemental Methods and Data for details).

Gene Ontology Enrichment analysis was performed using Bioconductor package “GOSStats” in R⁵⁸. Gene ontology annotations were obtained from the Comprehensive Microbial Resource (JCVI) available at <http://www.geneontology.org/GO.downloads.annotations.shtml>. All analyses in Table S2 have a p-value < 0.005.

Codon usage pattern

To determine whether there was a relationship between codon usage and changes in T_{eff} , we used the set of 854 mRNA:protein pairs that passed both quality and abundance filters and were detectable under both normal aerobic and low O₂ conditions (see Fig. 2). The codon usage of each gene was then obtained from the Codon Usage Database (<http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=211586>). We used a standard linear regression model to relate the change in translation efficiency for protein j between the normal O₂ and low O₂ conditions, TE_{jH} and TE_{jL} , respectively, to the codon frequency of each protein. The regression model defines the dependent variable as the log2 change in TE_{jH} to TE_{jL} and the independent variables as the codon frequencies (normalized to codons per thousand), where $C_j^{(k)}$ is the frequency of codon k for the j -th protein:

$$\log_2(\delta(TE_j)) = \log_2\left(\frac{TE_{jH}}{TE_{jL}}\right) = \beta_0 + \beta_1 C_j^{(1)} + \beta_2 C_j^{(2)} + \dots + \beta_{61} C_j^{(61)} \quad [\text{Eq. 3}]$$

A standard linear regression was performed using the ‘LinearModel.fit’ function within the Statistics Toolbox of MatLab R2012b. An evaluation of the specific codons with the largest contribution to the model was identified using stepwise regression (LinearModel.stepwise). Correlation between predicted and observed change in T_{eff} using a 7-fold cross-validation approach repeated 1000 times yielded an average correlation of 0.37 with a standard deviation of 0.01.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was performed in the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory (PNNL) and is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the PNNL Foundational Scientific Focus Area. This work was also funded in part by NIH grant 2 P41 GM103493-11.

References

1. Gingold H, Pilpel Y. Mol Syst Biol. 2011; 7:481. [PubMed: 21487400]
2. Vogel C, Marcotte EM. Nature reviews Genetics. 2012; 13:227–232.
3. Schwanhauser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Nature. 2011; 473:337–342. [PubMed: 21593866]
4. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C. Molecular bioSystems. 2009; 5:1512–1526. [PubMed: 20023718]

5. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M. *Mol Syst Biol*. 2011; 7:548. [PubMed: 22068331]
6. Ning K, Fermin D, Nesvizhskii AI. *Journal of proteome research*. 2012; 11:2261–2271. [PubMed: 22329341]
7. Picard F, Milhem H, Loubiere P, Laurent B, Coccagn-Bousquet M, Girbal L. *BMC genomics*. 2012; 13:528. [PubMed: 23036066]
8. Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, Nichols RJ, Typas A, Gross CA, Kramer G, Weissman JS, Bukau B. *Cell*. 2011; 147:1295–1308. [PubMed: 22153074]
9. Dressaire C, Gitton C, Loubiere P, Monnet V, Queinnec I, Coccagn-Bousquet M. *PLoS Comput Biol*. 2009; 5:e1000606. [PubMed: 20019804]
10. Klumpp S, Zhang Z, Hwa T. *Cell*. 2009; 139:1366–1375. [PubMed: 20064380]
11. Maier T, Schmidt A, Güell M, Kühner S, Gavin AC, Aebersold R, Serrano L. *Molecular Systems Biology*. 2011; 7:511. [PubMed: 21772259]
12. Bremer, H.; Denis, PP. *Escherichia coli and Salmonella: cellular and molecular biology*. 2. Neidhardt, FC.; Curtiss, R., III; Ingraham, JL.; Lin, ECC.; Low, KB.; Magasanik, B.; Reznikoff, WS.; Riley, M.; Schaechter, M.; Umberger, HE., editors. ASM Press; Washington, D.C: 1996. p. 1553-1569.
13. Christensen SK, Mikkelsen M, Pedersen K, Gerdes K. *Proc Natl Acad Sci U S A*. 2001; 98:14328–14333. [PubMed: 11717402]
14. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. *Molecular & cellular proteomics: MCP*. 2002; 1:323–333. [PubMed: 12096114]
15. Tuller T, Kupiec M, Rupp E. *PLoS Comput Biol*. 2007; 3:e248. [PubMed: 18159940]
16. Nie L, Wu G, Zhang W. *Biochemical and biophysical research communications*. 2006; 339:603–610. [PubMed: 16310166]
17. Elias DA, Tollaksen SL, Kennedy DW, Mottaz HM, Giometti CS, McLean JS, Hill EA, Pinchuk GE, Lipton MS, Fredrickson JK, Gorby YA. *Arch Microbiol*. 2008; 189:313–324. [PubMed: 18030449]
18. Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH. *J Bacteriol*. 2009; 191:3203–3211. [PubMed: 19304856]
19. Pasa-Tolic L, Masselon C, Barry RC, Shen Y, Smith RD. *BioTechniques*. 2004; 37:621–624. 626–633, 636. passim. [PubMed: 15517975]
20. Brockmann R, Beyer A, Heinisch JJ, Wilhelm T. *PLoS computational biology*. 2007; 3:e57. [PubMed: 17381238]
21. Mandelstam J. *The Biochemical journal*. 1958; 69:110–119. [PubMed: 13535591]
22. Pine MJ. *Annual review of microbiology*. 1972; 26:103–126.
23. Tang YJ, Hwang JS, Wemmer DE, Keasling JD. *Applied and environmental microbiology*. 2007; 73:718–729. [PubMed: 17098921]
24. Pinchuk GE, Hill EA, Geydebrekht OV, De Ingeniis J, Zhang X, Osterman A, Scott JH, Reed SB, Romine MF, Konopka AE, Beliaev AS, Fredrickson JK, Reed JL. *PLoS computational biology*. 2010; 6:e1000822. [PubMed: 20589080]
25. Elias DA, Monroe ME, Marshall MJ, Romine MF, Beliaev AS, Fredrickson JK, Anderson GA, Smith RD, Lipton MS. *Proteomics*. 2005; 5:3120–3130. [PubMed: 16038018]
26. Neidhardt, FC.; Umberger, HE. *Escherichia coli and Salmonella: cellular and molecular biology*. 2. Neidhardt, FC.; Curtiss, R., III; Ingraham, JL.; Lin, ECC.; Low, KB.; Magasanik, B.; Reznikoff, WS.; Riley, M.; Schaechter, M.; Umberger, HE., editors. ASM Press; Washington, D.C: 1996. p. 13-16.
27. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. *Nature biotechnology*. 2007; 25:117–124.
28. Peng M, Taouatas N, Cappadona S, van Breukelen B, Mohammed S, Scholten A, Heck AJR. *Nature Methods*. 2012; 9:524–525. [PubMed: 22669647]
29. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. *Nature*. 2003; 425:737–741. [PubMed: 14562106]
30. Jayapal KP, Philp RJ, Kok YJ, Yap MG, Sherman DH, Griffin TJ, Hu WS. *PloS one*. 2008; 3:e2097. [PubMed: 18461186]

31. Kaczanowska M, Ryden-Aulin M. Microbiology and Molecular Biology Reviews. 2007; 71:477–494. [PubMed: 17804668]
32. Gescher JS, Cordova CD, Spormann AM. Mol Microbiol. 2008; 68:706–719. [PubMed: 18394146]
33. Marritt SJ, McMillan DG, Shi L, Fredrickson JK, Zachara JM, Richardson DJ, Jeuken LJ, Butt JN. Biochemical Society transactions. 2012; 40:1217–1221. [PubMed: 23176457]
34. Mathy N, Pellegrini O, Serganov A, Patel DJ, Ehresmann C, Portier C. Mol Microbiol. 2004; 52:661–675. [PubMed: 15101974]
35. Bock A, King PW, Blokesch M, Posewitz MC. Advances in microbial physiology. 2006; 51:1–71. [PubMed: 17091562]
36. Karlin S, Mrázek J. Journal of Bacteriology. 2000; 182:5238–5250. [PubMed: 10960111]
37. Reis, Md. Nucleic acids research. 2004; 32:5036–5044. [PubMed: 15448185]
38. Tuller T, Waldman YY, Kupiec M, Ruppin E. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:3645–3650. [PubMed: 20133581]
39. Gingold H, Dahan O, Pilpel Y. Nucleic acids research. 2012; 40:10053–10063. [PubMed: 22941644]
40. Supek F, Vlahovick K. BMC bioinformatics. 2005; 6:182. [PubMed: 16029499]
41. Nath K, Koch AL. The Journal of biological chemistry. 1970; 245:2889–2900. [PubMed: 4912536]
42. Picard F, Dressaire C, Girbal L, Cocaïgn-Bousquet M. Comptes rendus biologiques. 2009; 332:958–973. [PubMed: 19909919]
43. Salis HM, Mirsky EA, Voigt CA. Nature biotechnology. 2009; 27:946–950.
44. Kudla G, Murray AW, Tollervey D, Plotkin JB. Science. 2009; 324:255–258. [PubMed: 19359587]
45. Novoa EM, Ribas de Pouplana L. Trends Genet. 2012; 28:574–581. [PubMed: 22921354]
46. Nasvall SJ, Chen P, Bjork GR. RNA. 2007; 13:2151–2164. [PubMed: 17942742]
47. Nasvall SJ, Chen P, Bjork GR. RNA. 2004; 10:1662–1673. [PubMed: 15383682]
48. Gabriel K, Schneider J, McClain WH. Science. 1996; 271:195–197. [PubMed: 8539617]
49. Baumgartner BL, Bennett MR, Ferry M, Johnson TL, Tsimring LS, Hasty J. Proc Natl Acad Sci U S A. 2011; 108:21087–21092. [PubMed: 22160690]
50. Heinrich R, Rapoport TA. Journal of theoretical biology. 1980; 86:279–313. [PubMed: 7442295]
51. Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner TS, Nealson KH, Osterman AL, Pinchuk G, Reed JL, Rodionov DA, Rodrigues JL, Saffarini DA, Serres MH, Spormann AM, Zhulin IB, Tiedje JM. Nature reviews Microbiology. 2008; 6:592–603.
52. Berthoumieux S, de Jong H, Baptist G, Pinel C, Ranquet C, Ropers D, Geiselmann J. Mol Syst Biol. 2013; 9:634. [PubMed: 23340840]
53. Chen Z, Duan X. Methods Mol Biol. 2011; 733:93–103. [PubMed: 21431765]
54. Adkins JN, Mottaz HM, Norbeck AD, Gustin JK, Rue J, Clauss TR, Purvine SO, Rodland KD, Heffron F, Smith RD. Molecular & cellular proteomics: MCP. 2006; 5:1450–1461. [PubMed: 16684765]
55. Livesay EA, Tang K, Taylor BK, Buschbach MA, Hopkins DF, LaMarche BL, Zhao R, Shen Y, Orton DJ, Moore RJ, Kelly RT, Udseth HR, Smith RD. Analytical chemistry. 2008; 80:294–302. [PubMed: 18044960]
56. Hixson KK, Adkins JN, Baker SE, Moore RJ, Chromy BA, Smith RD, McCutchen-Maloney SL, Lipton MS. Journal of proteome research. 2006; 5:3008–3017. [PubMed: 17081052]
57. Zimmer JS, Monroe ME, Qian WJ, Smith RD. Mass spectrometry reviews. 2006; 25:450–482. [PubMed: 16429408]
58. Falcon S, Gentleman R. Bioinformatics. 2007; 23:257–258. [PubMed: 17098774]

Insight, innovation, integration

A large number of studies have shown that changes in mRNA levels are not always accompanied by a corresponding change in protein levels. Here, we show that in bacteria, this is mostly due to widespread changes in the translational efficiency of specific mRNAs, partially because of environmentally-modulated functional tRNA pools. We achieved this understanding by using a steady-state mathematical model in combination with a steady-state experimental cell system. The integration of RNA-Seq data with quantitative global proteomics measurements as well as the use of numerous biological and technical replicates yielded a dataset of very high quality and precision, thus allowing us to determine the relationship between more than 1000 mRNA-protein pairs.

Fig. 1.

Schematic of the data sampling approach. Cells were grown to steady state at 20% O₂ in a chemostat and samples were taken at approximately 10h intervals at a cell density of 1.5×10^9 /ml ($OD_{600}=0.52$). Triplicate RNA-Seq and quadruplicate MS-based proteomics measurements were made from each sample. After equilibration of the chemostat to 8.5% O₂ for at least 48 h, samples were taken at 16–24h intervals at a cell density of 6.3×10^8 /ml ($OD_{600}=0.26$). The processing of all samples was the same. See Supplemental Materials and Data for details. The graph is representational in that the time used to reach steady state was far greater than the sampling intervals.

Fig. 2.

Workflow for integrating transcriptomics and proteomics data. Data from the RNA-Seq and proteomics analysis of each sample indicated in A was expressed in copies per cell and combined at the level of individual samples and pairs of samples. The first aerobic O₂ sample was paired with the first low O₂ sample, the second with the second and so forth. Because the set of expressed genes under both conditions are not completely congruent, the combined sets contain fewer members than the parent sets. See text for details on the how the data was combined and filtered.

Fig. 3. Properties of the combined proteomics and transcriptomics data from each sample. (A) Distribution of the number of genes, the mRNA abundance and protein abundance in each sample. Genes that had functional assignments in the SEED database were used in this comparison (3276 out of 4429) and the different subsystems were assigned general categories as indicated in Table S3. Data shown is for the first aerobic sample. Similar results were obtained for the other samples. (B) The effect of the progressive filtering on the category distribution of measured mRNA copies (top set of bars) and protein copies (bottom set of bars). The color code and sample used is the same as 2A. See Supplemental Materials and legend for Fig. 2 for description of the filters used.

Fig. 4.

Protein abundance as a function of mRNA expression levels. (A) The estimated number of mRNA copies for each detectible protein in the first aerobic sample is plotted against the estimated copy number of the protein. Regression analysis was against the log of the protein and mRNA data. Open circles are all (2393) protein-mRNA pairs. Blue dots comprise the quality-filtered subset of the data (1630 pairs) (B) Protein-mRNA pairs (unfiltered) were ranked by mRNA copy number and binned into 25 equal groups. The average of the mRNA and protein copies of each bin are shown together with the standard error of the mean. Linear regression yields a $cc = 0.93\text{--}0.94$ ($P\text{-value} < 0.001$) using all data points. Omitting the lowest two values of each group gives a $cc = 0.98\text{--}0.99$ ($P\text{-value} < 0.001$).

Fig. 5.

Effect of changing O_2 levels on both mRNA and protein expression levels. (A) Proteomics data from the three biological replicates of cells grown under aerobic conditions was filtered to remove proteins that displayed a %CV of >33%. Shown is a comparison of the data from the first two samples. (B) The levels of mRNA corresponding to the proteins shown in top panel of 4A are shown for comparison. (C) Proteins from samples grown under anaerobic conditions were first filtered to remove those that displayed a %CV>33% and then corresponding proteins from both samples were compared. (D) The levels of mRNA corresponding to the proteins shown in 4C are shown for comparison. (E) The log2 values of the fold change in mRNA levels due to low O_2 conditions versus the log2 values of the corresponding proteins. The data corresponds to those shown in 4C and 4D. The line is from linear regression analysis. (F) The data from 4E was assigned to 50 different bins based on the log2 change in protein levels (dotted grey), mRNA levels (red) or calculated T_{eff} values (blue). The Y-values are the count frequencies (out of 1258) of the data in each bin.

Fig. 6.

Double-bubble plots of the change in both mRNA expression and T_{eff} in cells grown under aerobic and low O_2 conditions. Top left panel: data of the mRNA and calculated T_{eff} values from cells grown under normal aerobic conditions plotted against the same data with introduced Gaussian noise to simulate biological replicates. Top right panel: data of the mRNA and calculated T_{eff} values of samples from two biological replicates of cells grown under normal aerobic conditions plotted against each other. Bottom left: quality-filtered data of the mRNA and calculated T_{eff} values of samples from cells grown under normal aerobic conditions plotted against data from cells grown under 8.5% O_2 conditions (N=1258). Bottom right panel: data shown in the bottom left panel was filtered to include only the ribosomal proteins (N=48). As indicated in the legend, the size of each data point is proportional to the number of peptides used in the abundance calculation and the color corresponds to the %CV of the data calculated from the variance of each protein from all biological replicates.

Fig. 7.

Observed versus predicted $\log_2(\delta T_{eff})$ based on either a full or reduced codon usage model. (A) Linear regression analysis of the usage of all 61 codons in each gene and their observed $\log_2(\delta T_{eff})$ was used to build a linear model that could predict the relationship. Shown is the model prediction of each gene plotted against the observed value. (B) Stepwise regression identified a 9 codon model that was as predictive as the full 61 codon model (see Table 2). Shown is the predicted $\log_2(\delta T_{eff})$ for each gene plotted against the observed value using the reduced model.

Table 1

Changes in tRNA pools as measured by RNA-Seq

Amino acid	Codon	Gene Copy	Level ^a (Aerobic)	Level ^a (low O ₂)	Fold change	p-value ^b
Ala	GCA	3	38.9	35.7	-1.09	0.612
Ala	GCC	2	122	240	1.95	0.009
Arg	CGT	6	128	274	2.13	0.008
Arg	CGG	1	21.2	31.6	1.49	0.094
Arg	AGA	1	2.67	3.89	1.46	0.050
Arg	AGG	1	390	468	1.20	0.443
Asn	AAC	5	104	85.0	-1.23	0.224
Asp	GAC	4	380	59.7	-6.36	0.002
Cys	TGC	2	10.1	31.2	3.06	0.002
Gln	CAA	3	105	57.9	-1.82	0.079
Glu	GAA	6	214	237	1.11	0.352
Gly	GGC	6	39.6	104	2.61	0.001
Gly	GGA	1	206	82.1	-2.50	0.009
His	CAC	2	36.3	47.4	1.30	0.005
Ile	ATC	3	158	105	-1.50	0.072
Leu	TTA	3	14.8	44.3	2.98	0.016
Leu	TTG	2	1070	1440	1.36	0.019
Leu	CTA	2	25.5	18.6	-1.37	0.057
Leu	CTG	2	294	43.9	-6.71	0.029
Leu	CTC	1	45.0	33.1	-1.36	0.353
Lys	AAA	8	495	803	1.62	0.016
Met	ATG	8	1590	265	-5.99	0.016
Phe	TTC	3	1.85	4.83	2.61	0.016
Pro	CCA	3	31.1	33.2	1.07	0.580
Pro	CCC	1	5.53	0.480	-11.6	0.011
Ser	TCA	3	21.0	4.36	-4.82	0.001
Ser	AGC	2	84.2	78.6	-1.07	0.574
Ser	TCC	1	190	25.9	-7.34	0.005

Amino acid	Codon	Gene Copy	Level ^a (Aerobic)	Level ^a (low O ₂)	Fold change	p-value ^b
Thr	ACA	3	13.9	8.09	-1.72	0.050
Thr	ACC	2	9.45	16.7	1.76	0.055
Trp	TGG	1	4.41	4.67	1.06	0.537
Tyr	TAC	4	136	128	-1.06	0.459
Val	GTA	5	90.1	71.1	-1.27	0.171
Val	GTC	2	10.3	0.641	-16.0	0.028

^a Average number of reads mapping to the indicated tRNA genes from three replicate measurements x 10⁻³.

^b p-values for each codon based on a 2-sample t-test

Table 2

Codons associated with changes in translational efficiency

Codon ^a	P-Value ^b	RMSE ^c
GCG	2.45×10^{-9}	0.919
GCC	5.42×10^{-6}	0.903
CGC	1.49×10^{-4}	0.894
UAC	3.53×10^{-4}	0.889
CCC	2.25×10^{-3}	0.883
GGU	7.55×10^{-3}	0.880
ACC	1.07×10^{-2}	0.878
CUU	1.72×10^{-2}	0.876
UCG	4.60×10^{-2}	0.874
9 Codon Model	1.30×10^{-25}	0.874
61 Codon Model	6.75×10^{-24}	0.844

^a Codons used in the reduced model in order of their selection.^b Compared against the full model.^c Associated Root Means Square Error when compared against the full model.