Electronic Supplementary Material (ESI) for Integrative Biology

Steady State Analysis of Integrated Proteomics and Transcriptomics Data Shows Changes in Translational Efficiency a Dominant Regulatory Mechanism in the Environmental Response of Bacteria

Ronald C. Taylor[2], Bobbie-Jo M. Webb Robertson, Lye Meng Markillie[3], Margrethe H. Serres[4], Bryan E. Linggi[1], Joshua T. Aldrich[1], Eric A. Hill[3], Margaret F. Romine[3], Mary S. Lipton[3] and H. Steven Wiley[1]

[1]Environmental Molecular Sciences Laboratory, [2]Computational Biosciences, [3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, [4]Bay Paul Center for Molecular Evolution, Marine Biological Laboratory, Woods Hole, MA 02543

Supplementary Information

Contents:

SUPPLEMENTAL METHODS

Transcriptomics. Measurement of gene expression was performed in RNA-Seq experiments conducted across the two conditions of normal oxygen (20% O_2, normal aerobic) and low oxygen (8.5% O_2, anoxic). Three technical replicates were done for sample1 (aerobic) and three for sample4 (anoxic), for a total of six RNA-Seq runs on a Life Technologies SOLiD 4 sequencer, following rRNA depletion (see Figure 1A for sample designations). Runs were performed on two additional anoxic samples (sample5 and sample6) with separate library preparations as technical duplicates and RNA depletion (four runs total). Two additional aerobic samples (sample2 and sample3) were each run four times, twice after rRNA depletion, and twice using total RNA (eight runs total). These 12 additional runs were performed on a SOLiD 5500xl sequencer.

A second set of samples for transcriptomics analysis was collected and processed to ensure consistency and repeatability of the results from the first set of samples. In this case, duplicate technical replicate libraries were prepared from cells grown under both aerobic condition and anoxic conditions. rRNA was not removed from these samples, which were analyzed on a SOLiD 5500xl sequencer.

The 50-base short read sequences produced by the SOLiD 4 sequencer were mapped in color space using the Life Technologies BioScope software version 1.3 using the reference genome and called genes from *Shewanella oneidensis* strain MR-1 (RefSeq Assembly ID GCF_000146165.2). The 75-base reads produced by the SOLiD 5500xl sequencer were similarly mapped using the Life Technologies LifeScope software version 2.5, using the same reference genome.

The sequence of *Shewanella oneidensis* strain MR-1 (as two FASTA file entries, for one chromosome and one plasmid), as well as the locations of the *Shewanella* called genes in GTF file format were used for read alignment and mapping in BioScope and LifeScope using the default parameters, with starting base positions of 0 and 15, each run allowing a maximum of two mismatches in the "seed" region of 25 bases. Reads passing the initial seeding phase were extended with each base match receives a score of +1, while mismatches received the default value of -2. Extension proceeds until a maximum score is reached. The read is assigned to the chromosome location giving the highest score.

A filtering file was used that included the rRNA sequences for *S. oneidensis* in addition to the adapter and bar code sequences before mapping the remaining short reads to the reference genome. Each SOLiD RNA-Seq run produced (1) a gene rollup file, with the base counts summed to a single value across the entire gene length, and with a RPKM value (Bullard et al, 2010) also given for each gene; (2) a BAM file containing the sequence of every mapped read and its mapped location; (3) two pairs of *.wig files (one pair for the two strands on the chromosome, one pair for the two strands of the plasmid) giving the mapped counts at each base position; and (4) a statistics summary on the set of reads as to what was filtered, what passed the filtering and was successfully mapped and what did not map after filtering. The fraction of total RNA that mapped to rRNA, mRNA and non-protein coding regions was used in conjunction with the total amount of RNA extracted from each cell to determine the estimated amount of each type of RNA species on a per-cell basis. The sum of the reads associated with each protein-encoding gene was divided by the sum of the reads associated with all protein-encoding genes to yield the fraction of total mRNA associated with each expressed gene. These were then

converted to copies of mRNA per cell using the total mass of mRNA per cell and the molecular weight of each mRNA species (see Table S1 for individual calculations).

Proteomics. To measure protein abundance, mass spectrometry (MS) experiments were run across the two conditions of low oxygen (8.5% O_2, anoxic) and normal oxygen (20% O_2, aerobic). Six biological samples were used: three samples each for the two conditions, with four technical replicates for each sample (G1, G2, G3, G4), for a total of 24 measurements on the peptide sets.

Each MS experiment was performed using 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. Full MS spectra were recorded at a resolution of 100k (for ions at *m/z* 400) over the range of *m/z* 400-2000 with an automated gain control (AGC) set at 1×10^6 ions. MS/MS of the ten most abundant parent ions were done in the ion trap at a normalized collision energy setting of 40% in the data-dependent mode with an AGC target value of 3×10^4 ions. Precursor ion activation was performed with an isolation width of 2 Da, a minimal intensity of 500 counts, and an activation time of 30 ms.

The Accurate Mass and Time (AMT) tag approach (Hixson et al, 2006; Zimmer et al, 2006) was applied to produce quantitative peptide abundance data. This method matches LC-MS features to a previously generated database using the metrics monoisotopic mass and normalized elution time (NET). Peptide sequences were identified using the SEQUEST v.27 (rev. 12) search engine. The feature database was populated using identifications having an XCorr >= 2.0, 2.5, or 3.5 for 1+, 2+, or >=3+ if seen once, XCorr >= 1.8, 2.2, or 3.2 for 1+, 2+, or >=3+ if seen >= 2 times, no cleavage rules and minimum length of six. Features from the 1-D analysis were matched to this database and filtered using a uniqueness probability of 0.51 to ensure specificity of the match.

Peptide data were normalized by conditions within each set of technical replicates by linear regression normalization. Briefly, the ion intensity of each peptide was $\log_2$ transformed followed by linear regression using all combinations of the samples. The sample pairs showing the greatest correlations were used as the baseline for each condition. The coefficients from the baseline sample were then used to normalize within each condition. The normalized peptide intensity values from each technical replicate were then averaged and converted into absolute protein abundance using a modification of the "intensity-based absolute quantification" (iBAQ) method (Schwanhausser et al, 2011). First, the sum of the ion intensity values of all of the peptides from each protein was divided by the sum of the ion intensities in each sample to yield the fraction of the total proteins comprised by each individual protein. This fractional value was then converted into absolute amount of protein per cell and then finally copies of protein per cell by using the calculated protein per cell and molecular weight of each protein (see Table S1 for individual calculations).

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. Datasets were placed in a small number of flexible HBase tables using locus tag, peptide ID, parent protein and counts from RNA-Seq corresponding to each gene (rnaSeqCount), taxonomy, and genome tables that can contain different fields in their records. The datasets are annotated with background information on the genes, genome and taxonomy also stored in the central warehouse. The datasets, as well as annotation information, are finally combined in a gene-centric warehouse table that can then be easily accessed by Java programs to export in a

variety of formats from the warehouse into software tools for statistical analysis and visualization. Ties between the peptide counts in the peptide file and the corresponding RNA-Seq gene counts are made through the peptide rollups to their parent protein, where the proteinID is set to match the locus tag that is the primary key for the records in the rnaSeqCount table that stores the RNA-Seq count data on a per-gene basis. All calculations on protein rollups and peptide predictions were performed within the data warehouse system and stored back into linked tables.

Simulated data sets

A random set of 1200 mRNA abundance and T_{eff} value pairs was extracted from the original data set of cells grown under normal aerobic conditions to serve as the baseline set. The expression of mRNA was randomly changed using a Gaussian distribution with a standard deviation of 60(Saucier, 2000). The values of T_{eff} were changed in a similar way, using a standard deviation of 30. When generating modified data sets, random Gaussian noise was introduced into replicate data using a %CV of 10% for the mRNA data and 25% for the protein data. See Table S5 and Fig. S5 for example simulated data set.

Supplemental Data

Assumptions and Data Assessment

1. Calculations of net protein and RNA per cell in chemostat-grown Shewanella

To convert data from RNA-Seq and proteomics measurements, it is necessary to have accurate and consistent measurements on cell numbers per sample as well as their protein and RNA content. Unfortunately, most protein assays have inherent limitations because of interference by non-proteinaceous materials in the samples and because the reactivity of calibration standards do not necessarily match that of the sample. Similarly, methods to estimate RNA recovery have inherent limitations because they cannot correct for the effects of variable cell lysis (Johnson et al, 2005). As an alternate approach, we elected to base our protein estimates on the dry mass of the samples because there have been extensive studies relating the composition and RNA:protein ratios of bacterial cells to growth rates (Bremer & Denis, 1996; Cox, 2003). Based on these studies, protein content should be ~49% of the dry mass. At a growth rate of 0.1 h^{-1}, the RNA should be 10% of the protein mass. In a series of careful calibration studies comparing *Shewanella* dry cell mass, cell number and protein and RNA recovery, we found that 5.5 x 10^8 cells/mL corresponded to an OD_{600} = 0.2363 and a dry mass of 0.171g/L. The equation to convert optical density (Absorbance @ 600 nm) to dry weight (g/L) was Y=0.805319x - 0.019207, where Y=dry weight g/L, and x=optical density at 600 nm. The R-squared value of this trend line was 0.9966 (for 16 points). The calculated mass per cell was 309 fg/cell with an estimated protein content of 152 fg/cell and RNA content of 15 fg/cell. Measured protein content by BCA was 260 fg/cell, 71% greater than the estimate based on dry weight measurement. Measured RNA content after extraction and purification (by A_{260} using a NanoDrop) was 3 fg/cell, or an estimated recovery of 20%. The RNA recovery is within the range reported for other studies (Johnson et al, 2005). Because the intent of this study was to investigate relative quantitative changes between conditions and because the growth rate of the cell between conditions did not change, we used these estimated recoveries to normalize the data derived from the different samples.

Sample	Total reads	Mapped to Filter	Mapped to Genome	Mapped to genes	Intergenic	Not mapped	mapQV (<10)	% Mapped
MR1 HiO2a	128,461,457	92,462,735	17,577,533	10,771,744	2,577,150	18,421,189	4,228,639	85.7%
MR1 HiO2b	133,894,010	93,298,906	22,619,373	13,664,222	3,249,762	17,975,731	5,705,389	86.6%
MR1 LiO2a	114,106,443	55,968,033	41,786,380	12,375,891	3,398,383	16,352,030	26,012,106	85.7%
MR1 LiO2b	113,263,713	64,336,083	32,560,254	11,560,157	2,741,032	16,367,376	18,259,065	85.5%

Table 1. Mapping of sequencing reads in duplicate samples from aerobic (MR1_HiO2) and anoxic (MR1_LiO2) chemostat cultures.

To determine the net distribution of RNA species, duplicate samples of cells grown under either aerobic or anoxic conditions were analyzed by RNA-Seq using the rRNA, linker and barcode sequences in the filter files. The short reads were then mapped to all genomic features of *S. oneidensis* MR-1. Over 85% of the total reads were mapped to either sequences in the filter file or genome.

The distribution of the filtered and mapped counts to the different species was calculated and by using the estimated amount of total RNA per cell, the distribution could be converted into femtograms of the different species per cell.

During the preparation of the RNA-Seq libraries for this run, the recovery of the small tRNA species was not rigorously controlled and thus they comprised only ~2% of the total RNA species. This is quite a bit less than the estimates of tRNA pools of rapidly growing *E. coli* (up to 20% of the total) or the estimated rRNA:tRNA ratio of ~6:1 (Neidhardt & Umbarger, 1996). This magnitude of tRNA would give rise to an ~12% overestimation of the mRNA content of all samples. Because the main focus of this study was a comparative analysis between different samples and because the tRNA was essentially absent from all of the analyzed samples, we ignored this correction.

Based on our estimated recovery of RNA and distribution of the different RNA species, we calculate an abundance of mRNA between 2,600-4,300 molecules per cell, with an average of $3.2x10^3$ copies under aerobic conditions and $4.1x10^3$ copies under anoxic conditions (a 30% increase), compared to a 35% reduction of rRNA under anoxic conditions. The amount of rRNA under aerobic conditions translates to about 5,700 ribosomes and under anoxic conditions to 3,800 ribosomes. If the rRNA is all assembled into functional ribosomes, then there is an ~1:1 ratio of ribosomes to the number of mRNA molecules available. This contrasts with rapidly growing *E. coli* where there are about 7- to 14-fold greater number of ribosomes than mRNA molecules (Bremer & Denis, 1996).

Sample	% rRNA	% mRNA	% ncRNA	% Intergenic	fg RNA /cell	fg rRNA /cell	Fg ncRNA /cell	fg mRNA /cell
MR1_HiO2a	87.4	6.1	4.1	2.4	17.0	14.9	0.7	1.0
MR1_HiO2b	84.7	8.7	3.7	2.9	17.0	14.4	0.6	1.5
MR1_LiO2a	78.0	10.2	7.1	4.7	12.0	9.4	0.8	1.2
MR1_LiO2b	81.8	9.6	5.1	3.5	12.0	9.8	0.6	1.2

Table 2. Distribution of mapped reads to different species of RNA in samples from aerobic (MR1_HiO2) and anoxic (MR1_LiO2) chemostat cultures.

Estimates of the number of specific protein species per cell in *Shewanella* require knowledge of the fraction of total protein that is detectible by MS. Although we can detect the expression of essentially all *Shewanella* genes at some level, it is not clear that all of these transcripts give rise to functional proteins. Global analysis of yeast protein expression suggests that all transcripts give rise to some level of protein (Ghaemmaghami et al, 2003), but comparable studies in bacteria are not available. Based on the estimated amount of protein per cell and average protein molecular weight, the number of protein molecules is ~$2.6\text{-}2.7x10^6$ per cell, but the distribution of these numbers to the individual protein species will depend on which ones are actually expressed. There is a clear bias in the abundance of transcripts that are associated with expressed proteins (Fig. S1), with an average abundance of 0.88 mRNA copies/cell when a protein was detected and 0.24 mRNA copies/cell when a protein was absent. Still, the proteins from some abundant mRNAs are likely absent because they are secreted from the cells (e.g., CsgB,

nucleation component of extracellular curlin monomers) or because they are transmembrane proteins and poorly digested by trypsin (e.g., the preprotein translocase subunit SecG). This is consistent with previous studies (Vuckovic et al, 2013) and the observed underrepresentation of membrane transport and motility proteins as compared to their transcripts (see Fig. 2A).

As discussed in the main text, there is a strong correlation between the average transcript expression level and protein expression (Fig. 2D). Thus, it seems reasonable to assume that the

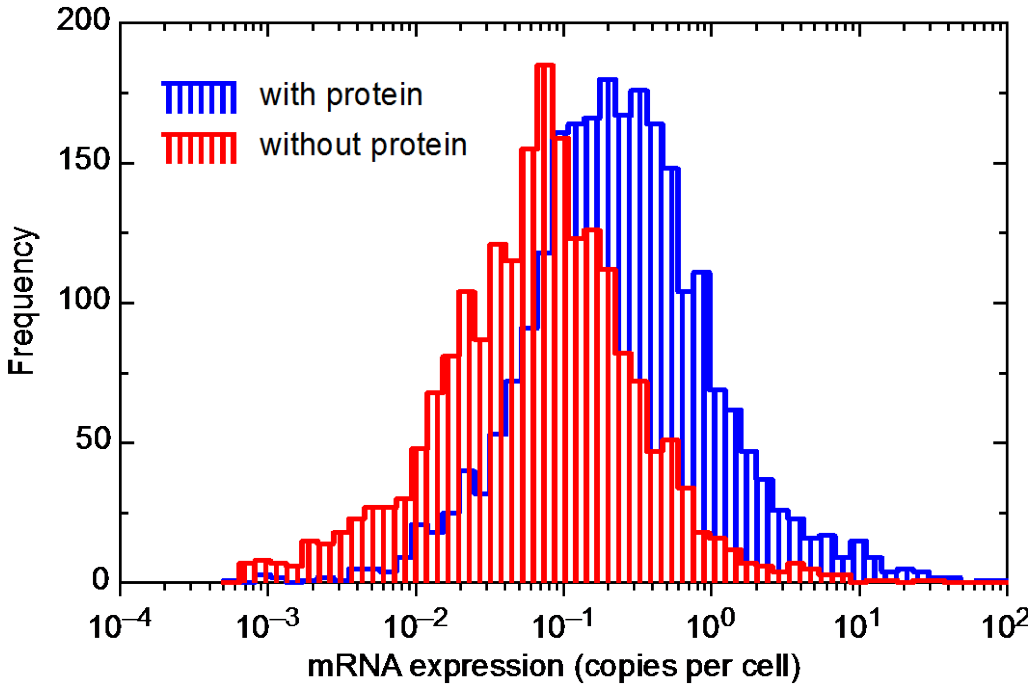

Fig. S1: frequency of mRNA abundance for transcripts associated with detected proteins (blue) or non-detected proteins (red). Transcripts from cells grown under aerobic conditions were sorted into the two groups and then placed in 50 bins based on log expression levels.

amount of protein that is not detectable is proportional to the average expression level of their transcripts. In terms of copies per cell, mRNAs corresponding to proteins with detectable expression comprise 81% of all transcripts numerically (and 87% of total mRNA mass because of their longer average length). If the proportionally between protein and mRNA expression holds, then we are missing between 13-19% of the proteins on a copies per cell basis. This would suggest that our estimates for protein copies per cell are only overestimated by 15-23% because of non-detectable proteins.

Although virtually all genes are expressed to some level, there are clear differences in their expression pattern when cells are changed from aerobic to anoxic conditions. As would be expected, some proteins are observed only under one condition. Out of the 2600 genes in which corresponding mRNA transcripts and peptides were observed under at least one condition, only 2123 (82%) were observed under both conditions. However, these corresponded to >99% of the total protein mass under either condition (99.5% of the proteins under aerobic and 99.2% of the proteins under anoxic conditions) and ~95% of the transcripts (93% under aerobic and 97.6% under anoxic conditions). This suggests that the protein/gene pairs that are only observed under a

single condition are generally expressed at low levels and the "disappearance" of the proteins could be a result of sensitivity limitations of MS-based proteomics measurements. Thus, there is an inherent difficulty in measuring changes in translational efficiency between conditions for low abundance proteins. In part to correct for this bias, we limited our comparisons not only to proteins found in both conditions, but also to those that were relatively abundant (>3 peptides).

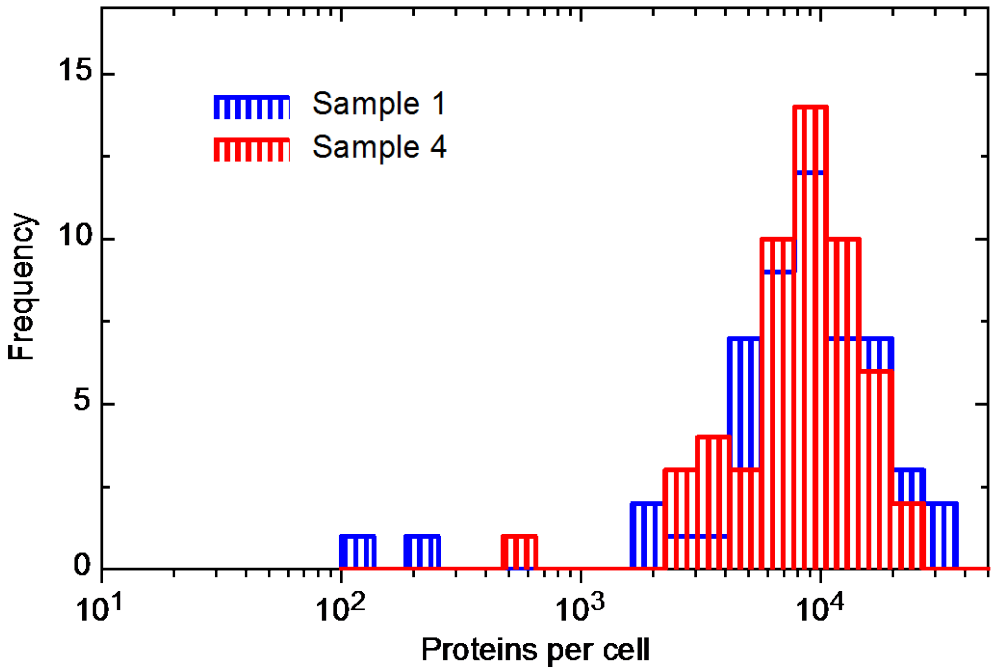

Fig. S2: count frequency of ribosomal proteins from cells grown under aerobic (blue) or anoxic (red) conditions. The calculated expression levels of all 53 detectable ribosomal proteins (both large and small subunit) were sorted into 20 bins by log expression levels. Median expression was 8400 for cells grown under aerobic conditions and 9200 for those grown under anoxic conditions.

As a check on the magnitude of our estimates of absolute protein copies per cell, we compared the average number of ribosomal proteins per cell to the estimated copies of 16S rRNA, assuming that there should be an average stoichiometry of 1:1 (Kaczanowska & Ryden-Aulin, 2007). We estimate that there are sufficient ribosomal proteins to make between 8400-9,200 ribosomes (Fig. S2). A missing protein correction would reduce this to 6800-8000 ribosomes, which is within 2-fold of our estimated range of 3800-5700 copies of 16S rRNA. However, it is not clear that the stoichiometry of 16S rRNA and ribosomal proteins are indeed 1:1 and as discussed in the main text, the non-coordinate nature of changes in ribosomal proteins, their mRNAs and rRNA abundance suggests that there might be some differential regulation of the various ribosomal components.

To determine the reasonableness of our estimates of protein:mRNA ratios and calculated translational efficiency (T_{eff}) values, we compared them to previous measurements in *E. coli*. Because the ratio of mRNA to ribosomes in our *Shewanella* cultures is relatively low, T_{eff} should be mostly dictated by protein elongation rates (Heinrich & Rapoport, 1980). The value of T_{eff} is given by Protein/mRNA * K_{dx}. (see Eq. 2 in the main text). Because of dilution rate of the

chemostats is much faster than the turnover rate of most bacterial proteins, D~ K_{dx}. The D of the bioreactors is 1.57×10^{-3} min^{-1}. Taking the total number of protein molecules per cell as 2.65×10^{6} and the mRNA molecules as 3.2×10^{3} copies under normal aerobic conditions, the global protein/mRNA ratio would be ~830, translating to a mean T_{eff} of ~1.3 min^{-1}. The average protein length is 330 AAs, thus the average rate of synthesis would be about 429 AAs per minute, or ~7 AAs per second. This compares to an average rate in *E. coli* estimated to be between 12-21 AAs per second (Young & Bremer, 1976). Our values thus seem consistent, especially considering that the experiments in *E. coli* were performed at 37°C whereas the *Shewanella* experiments were conducted at 30°C.

2. Reproducibility of technical and biological replicates

To ensure the reliability of the datasets we used for our analysis, we used multiple technical and biological replicates. The technical replicates for the proteomics measurements employed multiple mass spectrometers and separation columns to eliminate any inherent instrumentation biases. The biological replicates were used to check for steady state conditions in our chemostat cultures since, by definition, the levels of gene and protein expression should not change over time.

Reproducibility of proteomics measurements was evaluated by the ion current measurement of peptides observed over multiple runs. There was close agreement between the ion current measurements of peptides from samples run on different instruments, with only a handful of measurements (~0.1%) falling outside a relatively narrow cluster (Fig. S3A). A similar result was obtained from sequential biological samples run on the same instrument and column (Fig. S3B). Comparing the average ion current from the peptides seen in all of the aerobic biological samples (~5600 peptides) across different runs smoothed out the data, resulting in a tighter correlation (Fig. S3C). A similar result was seen from averaging the ion current of each peptide across all 4 runs and comparing different biological samples (Fig. S3D). These results suggest that any biological variability in the chemostat cultures is less than the intrinsic noise of the instrument measurements. Comparing RNA-Seq data from technical and biological replicates yielded almost perfect correlations, except at the very low end of the scale (Figs. S3E-F), demonstrating the extremely high reproducibility of the RNA-Seq measurements and the lack of any significant variability of the steady state chemostat samples taken from specific conditions.

3. Effect of removing rRNA on mRNA abundance measurements

The ribosome depletion step that we use in our study has been reported to bias some of the estimates of mRNA abundance by removing some of the message. To check for this, we took two biological replicates of the samples grown under normal aerobic conditions and looked for the effect of ribosome depletion. The rRNA was depleted according to the protocol of Chen and Duan (Chen & Duan, 2011), in which probes specific for bacterial 16S and 23S rRNA are hybridized to the extracted total RNA and then removed by magnetic beads. Typically, this removed 85% of the rRNA species, lowering the value from 86% to 13% of the total RNA.

We plotted the RPKM values of each gene from depleted and non-depleted replicates against each other and against biological replicates. As shown in Fig S4, even though there were 4-fold greater total gene-specific counts in the depleted versus non-depleted samples (~10.5M versus ~2.6M), there was no obvious change in the variation between samples due to the depletion step itself. On a more quantitative level, of the 4241 protein-encoding genes detected in this analysis, only 29 (0.7%) showed a 2-fold or greater reduction in measured gene expression levels

following ribosomal depletion. Most of these genes were expressed at extremely low levels, with only 3 associated with detectable proteins.

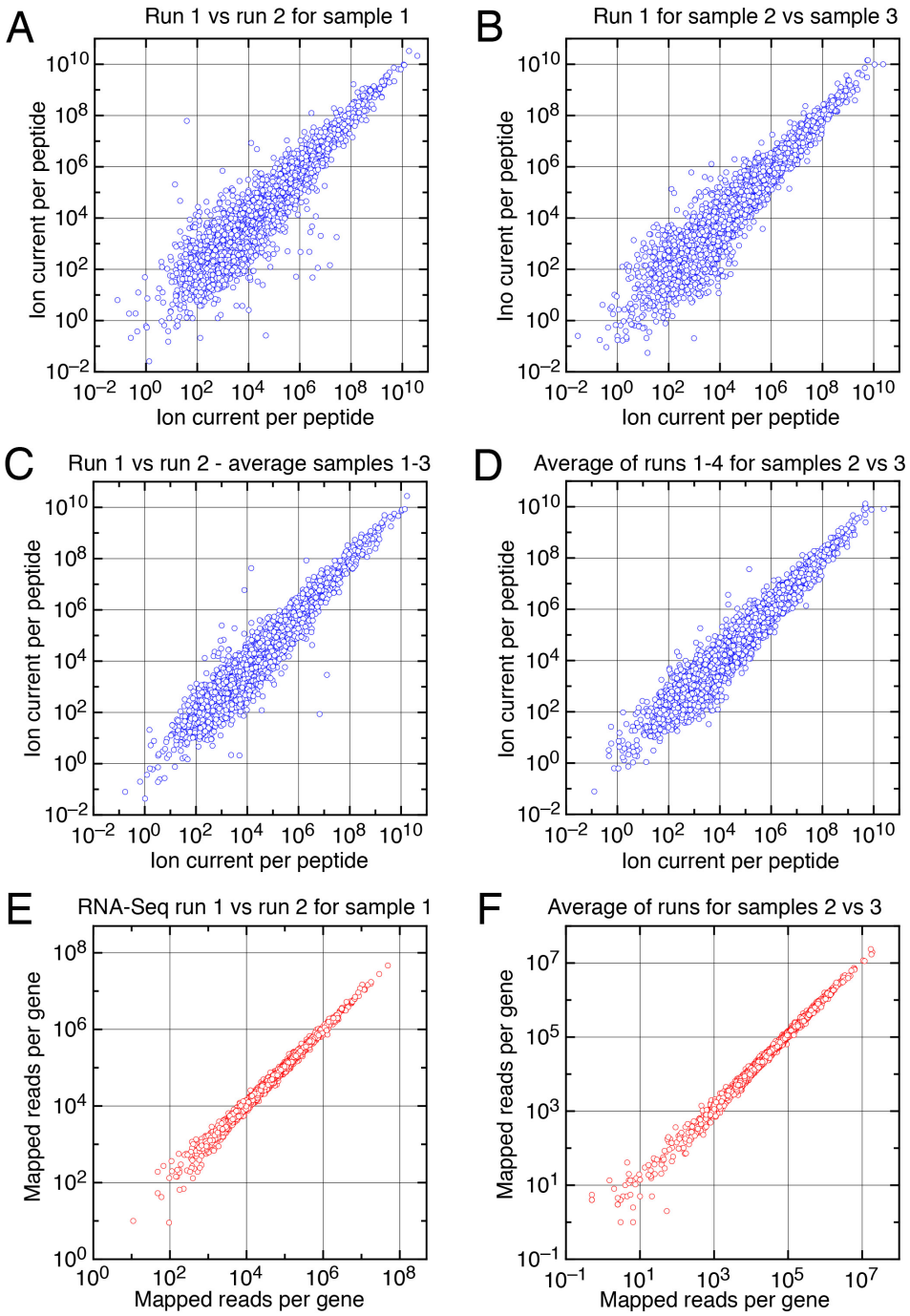

Fig. S3: Reproducibility of biological and technical replicates. (A) Ion intensities of peptides observed in two separate sets of runs are compared. (B) Ion intensities observed from two biological replicates in the same set of runs are compared. (C) The average ion intensity of peptides observed in all biological replicates is compared across two different runs. (D) Average ion intensities observed in all runs for two biological replicates are compared. (E) Technical replicates of RNA-Seq runs for one sample are compared. (F) Biological replicates for RNA-Seq runs are compared.

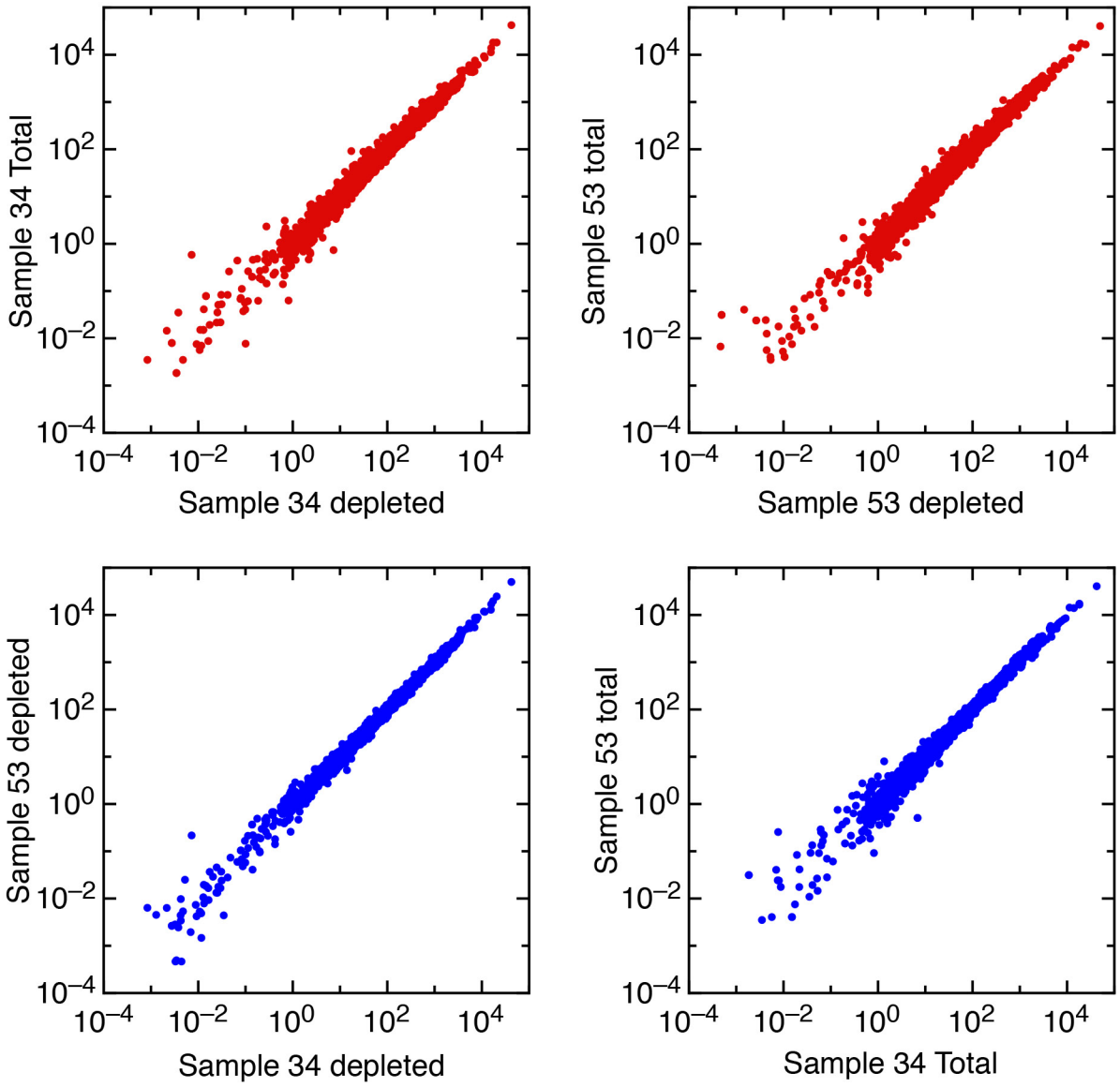

Fig. S4: Effect of ribosome deletion on estimation of gene expression levels in Shewanella. Data from the indicated samples were converted to RPKM values and compared. The most consistent results were obtained from depleted samples, most likely because of their much higher total read depth.

4. Graphic display of changes in transcription versus translational efficiency

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. As shown in Fig. S5 using simulated data, changes in gene expression accompanied by an attendant change in protein expression results in a vertical shift in the data points (Fig. S5, top left panel). If there is a change in T_{eff} without a change in mRNA levels, the

data points will shift horizontally (Fig. S5, top right panel). If there are changes in mRNA levels, but no change in protein levels, then the data points will shift diagonally (Fig. S5, bottom right panel). Finally, if there is a change in both mRNA levels and T_{eff} , the data values will shift out from the center (Fig S5, bottom left panel).

mRNA change
no protein change

Fig. S5 Simulated plots of the effect of the changing transcription and translational efficiency independently. An initial set of 1200 mRNA, protein and T_{eff} values were changed at random as described in Materials and Methods together with the introduction of Gaussian noise. The ratios of the initial and randomly changed data were then plotted. Top left: mRNA levels were altered, but T_{eff} values were kept constant. Top right: T_{eff} values were changed, but mRNA values were kept constant. Bottom left: both mRNA and T_{eff} values were kept the same. Bottom right: mRNA was changed, but the initial protein level was kept the same. The T_{eff} values were then calculated from the protein:mRNA ratios.

4. Effect of saturation kinetics on T_{eff} as mRNA levels change

The linear relationship observed between average mRNA and protein levels (Fig. 4B, main text) suggests that in general, protein levels will increase monotonically in response to changes in mRNA levels. However, it has been proposed that at the level of individual transcripts, protein production rates as a function of mRNA might be better described with saturation kinetics (Brockmann et al, 2007). The steady state balance equation (eg. 1 in main text) will hold regardless of the presence of saturation kinetics because it only describes the relationship between synthetic processes and degradative ones at a given concentration of mRNA (at steady state). If either protein synthesis or degradation displays saturation kinetics, this would be seen as a change in T_{eff} and/or K_d as a function of changing mRNA levels. Thus, a change in translational efficiency could indeed arise in concert with a change in specific mRNA levels due to this effect. Because of the relatively low abundance of ribosomes in our cells (see section 1 above), it seems reasonable to expect that some mRNAs are near saturation with respect to at least some of the steps involved in ribosome binding or elongation. In these cases, an increase in mRNA abundance should be accompanied by an apparent decrease in the associated T_{eff} parameter and visa versa.

To check for a saturation effect, we performed regression analysis using log(mRNA ratio) as the independent variable and log(T_{eff} ratio) as the dependent variable and found a significant negative correlation (Figure S6; slope = -0.41, cc = -0.37, $p < 0.001$). This is consistent with the hypothesis that aspects of translation are saturable and thus some of the changes in T_{eff} that we observe between conditions could be due to this effect. Functionally, this effect would be seen as a change in mRNA abundance associated with a change in protein abundance of a lesser magnitude. This was observed in 73% of the cases where mRNA changes were >2-fold, and thus, the consequence of saturation kinetics could be quite significant. However, saturation kinetics cannot explain how proteins levels can change significantly without a corresponding change in mRNA levels.

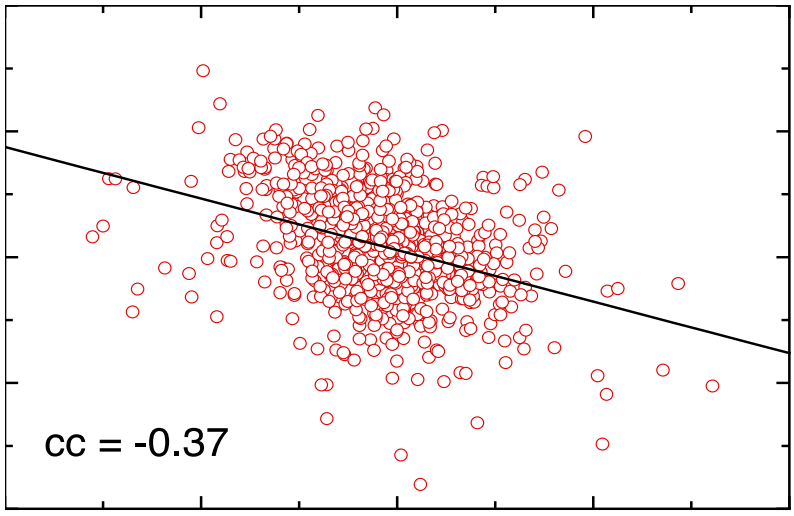

Fig. S6 Correlation between change in mRNA levels and changes in T_{eff}. Filtered data from comparing sample 1 and sample 4 is shown together with the linear regression line. This is the same data set shown in Fig. 5E in the accompanying paper.

References

Bremer H, Denis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella: cellular and molecular biology*, Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (eds), 2nd edn, pp 1553-1569. Washington, D.C.: ASM Press

Brockmann R, Beyer A, Heinisch JJ, Wilhelm T (2007) Posttranscriptional expression regulation: what determines translation rates? *PLoS computational biology* **3:** e57

Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics* **11:** 94

Chen Z, Duan X (2011) Ribosomal RNA depletion for massively parallel bacterial RNA-sequencing applications. *Methods Mol Biol* **733:** 93-103

Cox RA (2003) Correlation of the rate of protein synthesis and the third power of the RNA : protein ratio in Escherichia coli and Mycobacterium tuberculosis. *Microbiology* **149:** 729-737

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425:** 737-741

Heinrich R, Rapoport TA (1980) Mathematical modelling of translation of mRNA in eucaryotes; steady state, time-dependent processes and application to reticulocytes. *Journal of theoretical biology* **86:** 279-313

Hixson KK, Adkins JN, Baker SE, Moore RJ, Chromy BA, Smith RD, McCutchen-Maloney SL, Lipton MS (2006) Biomarker candidate identification in Yersinia pestis using organism-wide semiquantitative proteomics. *Journal of proteome research* **5:** 3008-3017

Johnson DR, Lee PK, Holmes VF, Alvarez-Cohen L (2005) An internal reference technique for accurately quantifying specific mRNAs by real-time PCR with application to the tceA reductive dehalogenase gene. *Appl Environ Microbiol* **71:** 3866-3871

Kaczanowska M, Ryden-Aulin M (2007) Ribosome Biogenesis and the Translation Process in Escherichia coli. *Microbiology and Molecular Biology Reviews* **71:** 477-494

Neidhardt FC, Umbarger HE (1996) Chemical composition of Escherichia coli. In *Escherichia coli and Salmonella: cellular and molecular biology*, Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (eds), 2nd edn, pp 13-16. Washington, D.C.: ASM Press

Saucier R. (2000) Computer Generation of Statistical Distributions. *Army Research Laboratory Technical Reports*. Army Research Laboratory, p. 30.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* **473:** 337-342

Vuckovic D, Dagley LF, Purcell AW, Emili A (2013) Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges. *Proteomics* **13:** 404-423

Young R, Bremer H (1976) Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem J* **160:** 185-194

Zimmer JS, Monroe ME, Qian WJ, Smith RD (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass spectrometry reviews* **25:** 450-482

Supplemental Tables

Table S1: Excel file with two tabs containing the entire original processed data set described in the text. Tab 1 is the main data. Sample labels correspond to the samples described in Figure 1 of the main text. Peptide values (e.g. columns E, R etc) represent the sum of the ion current values of all unique peptides (averaged across all technical duplicates) for that particular protein. "Norm peptide" (e.g. columns F, S etc.) represents the peptide values divided by the expected number of tryptic peptides for each protein. mRNA cts (columns L, Y etc) are the average number of RNA-Seq reads that mapped to that gene across the technical replicates. All intermediate calculations are present and can be seen by using "Column…Unhide" under the Format menu. Tab 2 comprises the quality-filtered data set from samples 1 and 4 used in most of the analyses described in the text.

Table S2: Gene Ontology Enrichment analysis of genes significantly regulated between anerobic and anoxic conditions. Data was pooled from all biological and technical replicates. Significance was calculated in Matlab using a 2-tailed T-test, P=0.05. Tab 1 are genes significantly increased in anoxic relative to aerobic conditions whereas Tab 2 are genes significantly decreased in anoxic relative to aerobic conditions.

Table S3: Functional classification of genes used in the comparative analysis of data filtering approaches.

Table S4: Excel file containing tRNA expression and amino acid usage pattern by Shewanella. Tab I is the tRNA expression data from samples 1 and 4 (see Table 1 in the main text). The data is the number of mapped reads for three independently prepared libraries from each sample. Tab 3 lists the amino acid composition of each protein in Shewanella followed by lists of amino acids weighed by actual protein expression. The Δexpression is the difference in weighed expression between aerobic and anoxic conditions. Below these columns are the sum of the differences at both the absolute and relative levels.

Table S5: Simulated dataset. A set of 1200 mRNA-protein expression pairs was selected randomly from our initial data sets and modified as described in Materials and Methods. This Excel workbook includes the code used to generate simulated differences between the data in the hypothetical experiments described in the Main text.