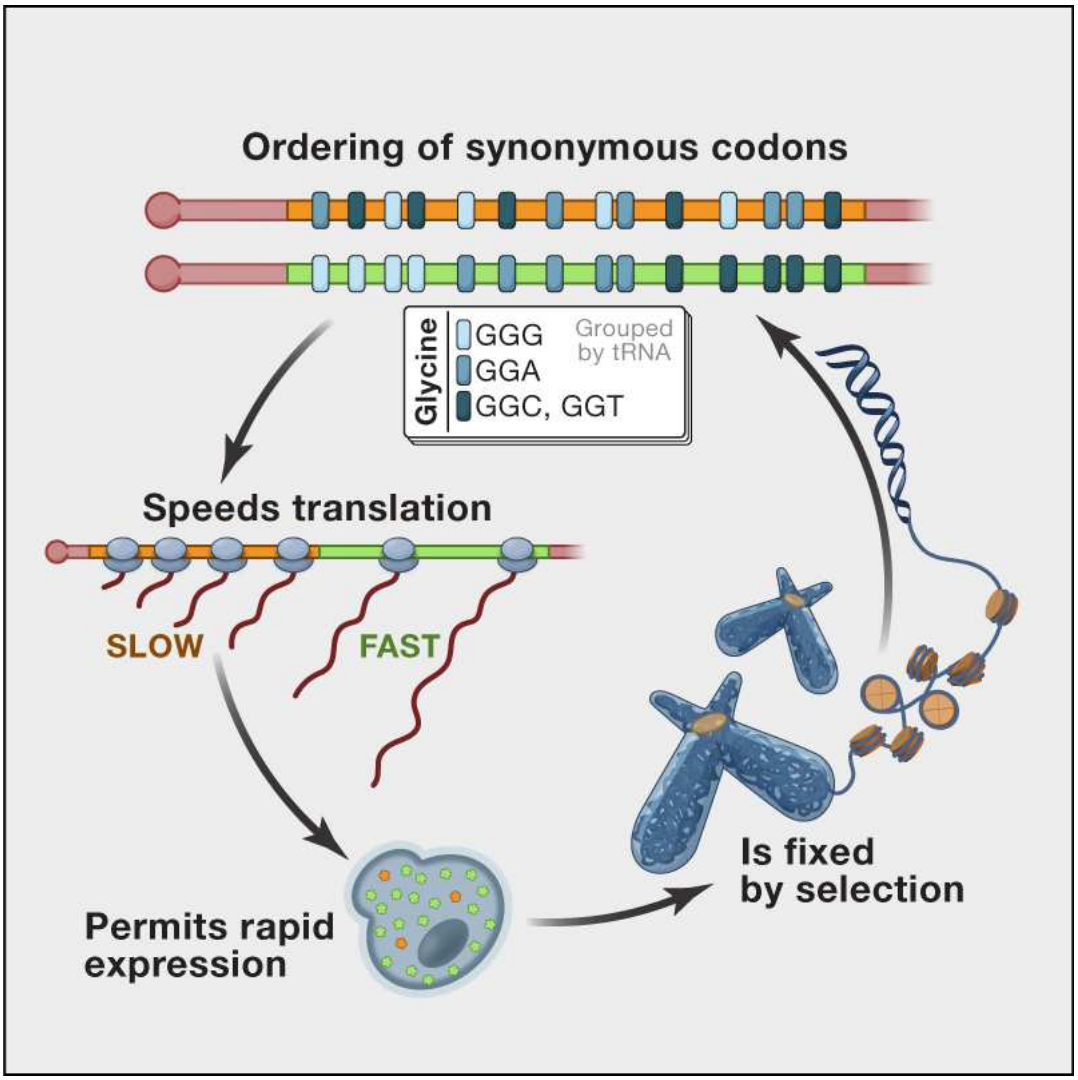
In This Issue



mRNA Sets Ribosomal Speed Limits

PAGE 344 and PAGE 355

Cells control the efficiency of translation with various protein regulatory factors. Using computational approaches, both Tuller et al. and Cannarozzi et al. now find that translational efficiency is also hardwired into the transcript sequence. Tuller et al. show that messages often start with codons that correspond to rare transfer RNAs (tRNAs). This forms an early elongation “ramp” that fosters slow translation and may maximize the efficiency of protein production. Cannarozzi et al. find that once a codon for a particular amino acid is used, the subsequent codons for that amino acid are likely to be the same, despite the availability of multiple codons for each amino acid. This correlation is highly conserved across organisms and is shown to increase translation speed, suggesting that tRNAs are recycled and channeled by the ribosome.

Cell

A Role for Codon Order in Translation Dynamics

Gina Cannarozzi,[2,3,4] Nicol N. Schraudolph,[2,4] Mahamadou Faty,[1,4] Peter von Rohr,[2] Markus T. Friberg,[2] Alexander C. Roth,[2,3] Pedro Gonnet,[2] Gaston Gonnet,[2,3,*] and Yves Barral[1,*]
[1]Institute of Biochemistry, ETH Zurich, 8093 Zurich, Switzerland
[2]Institute of Computational Science, ETH Zurich, 8092 Zurich, Switzerland
[3]Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Genopode, 1015 Lausanne, Switzerland
[4]These authors contributed equally to this work
*Correspondence: gonnet@inf.ethz.ch (G.G.), yves.barral@bc.biol.ethz.ch (Y.B.)
DOI 10.1016/j.cell.2010.02.036

SUMMARY

The genetic code is degenerate. Each amino acid is encoded by up to six synonymous codons; the choice between these codons influences gene expression. Here, we show that in coding sequences, once a particular codon has been used, subsequent occurrences of the same amino acid do not use codons randomly, but favor codons that use the same tRNA. The effect is pronounced in rapidly induced genes, involves both frequent and rare codons and diminishes only slowly as a function of the distance between subsequent synonymous codons. Furthermore, we found that in *S. cerevisiae* codon correlation accelerates translation relative to the translation of synonymous yet anticorrelated sequences. The data suggest that tRNA diffusion away from the ribosome is slower than translation, and that some tRNA channeling takes place at the ribosome. They also establish that the dynamics of translation leave a significant signature at the level of the genome.

INTRODUCTION

Translation of coding sequences into proteins by the ribosome underlies the expression of genomes into cellular and organismal functions. This process is mediated by tRNAs, which provide the code that associates each sense nucleotide triplet (codon) with a given amino acid. Each tRNA is charged on one end with a specific amino acid by its respective aminoacyl tRNA synthetase. At its other end, the tRNA exposes a 3-nucleotide sequence (anticodon) that recognizes specific codons of messenger RNAs at the acceptor site of the ribosome. In doing so, the tRNAs ensure that coding sequences are reproducibly translated into the same polypeptides. Thus, each of the 61 sense codons requires that at least one specific tRNA decodes it always into the same amino acid. Because there are more sense codons than amino acids, groups of codons are synonymous, i.e., they code for the same amino acid. Frequent amino acids can be encoded by up to six alternative codons. Ideally, these synonymous codons should be recognized and translated each by their own tRNA, presenting the corresponding anticodon sequence.

However, numerous tRNAs compete with each other at the acceptor site of ribosomes, until the correct tRNA is stably selected. Two observations suggest that this competition antagonizes translation efficiency.

First, evolution favored the emergence of multivalent tRNAs that can recognize more than one synonymous codon. This allows reducing the number of tRNAs needed, and hence, tRNA complexity. Consequently, most organisms translate the 61 sense codons with less than 61 tRNAs. Multivalent tRNAs use non-Watson-Crick base pairing to recognize several synonymous codons, called isoaccepting codons.

Second, the different tRNA species are differentially expressed: some tRNAs are more abundant than their synonymous cognates. As a consequence, synonymous codons are not equivalent and are not used randomly (Ikemura, 1985; Sharp et al., 1993): Codons decoded by frequent tRNAs are more frequent in coding sequences than their synonyms (Ikemura, 1985; Dong et al., 1996; Duret, 2000). This bias is strongest in highly expressed genes, indicating that codon composition has an impact on translation efficiency. This notion has been experimentally verified: Replacement of rare codons with frequent synonymous codons strongly improves the efficiency with which a sequence is translated in a given organism (Gustafsson et al., 2004).

Despite these mechanisms to simplify and optimize it, translation remains a fairly slow process in eukaryotes (2 amino acids per sec. on average) compared to bacteria (15 amino acids per sec.). Furthermore, depending on the physiological conditions and the transcript, the rate of amino acid incorporation varies substantially in eukaryotes, ranging from 1 to 10 amino acids per second (Spirin, 1999). Here, we provide evidence that translation speed can be modulated through tRNA recycling at the ribosome.

RESULTS

We investigated the distribution of pairs of synonymous codons in coding sequences (analyzed in all reading frames) of the yeast

Table 1. Codon Co-occurrence

(A) Co-occurrence Counts

tRNA	Ser1	Ser2	Ser3	Ser4
Ser1	**45392**	20797	9564	25702
Ser2	21119	**11766**	5101	13534
Ser3	9581	5150	**2607**	6296
Ser4	25381	13980	6463	**21029**

tRNA		Ser1		Ser2	Ser3	Ser4	
	Codon	TCC	TCT	TCA	TCG	AGC	AGT
Ser1	TCC	**6443**	**10525**	7831	3748	3814	5713
	TCT	**10412**	**18012**	12966	5816	6575	9600
Ser2	TCA	7707	13412	**11766**	5101	5272	8262
Ser3	TCG	3647	5934	5150	**2607**	2573	3723
Ser4	AGC	3906	6200	5543	2724	**3737**	**5030**
	AGT	5897	9378	8437	3739	**4982**	**7280**

(B) Standard Deviations from Expected

tRNA	Ser1	Ser2	Ser3	Ser4
Ser1	**16.62**	−5.31	−3.35	−12.98
Ser2	−2.53	**8.09**	1.12	−4.79
Ser3	−2.77	1.88	**6.34**	−2.09
Ser4	−15.81	−1.86	−0.68	**21.16**

tRNA		Ser1		Ser2	Ser3	Ser4	
	Codon	TCC	TCT	TCA	TCG	AGC	AGT
Ser1	TCC	**6.55**	**6.16**	−2.86	0.60	−6.23	−6.19
	TCT	**5.30**	**12.02**	−4.36	−4.68	−5.35	−7.16
Ser2	TCA	−3.82	−0.15	**8.09**	1.12	−5.78	−1.33
Ser3	TCG	−0.71	−2.92	1.88	**6.34**	−0.85	−1.98
Ser4	AGC	−5.14	−10.55	−2.93	1.53	**13.44**	**9.34**
	AGT	−3.90	−9.78	0.05	−2.15	**8.91**	**10.33**

(C) Percent Deviation from Expected

tRNA	Ser1	Ser2	Ser3	Ser4
Ser1	**7.35**	−3.46	−3.30	−7.34
Ser2	−1.65	**7.56**	1.56	−3.91
Ser3	−2.74	2.63	**13.15**	−2.56
Ser4	−8.91	−1.51	−0.84	**15.06**

tRNA		Ser1		Ser2	Ser3	Ser4	
	Codon	TCC	TCT	TCA	TCG	AGC	AGT
Ser1	TCC	**8.39**	**6.05**	−3.13	0.97	−9.52	−7.77
	TCT	**5.22**	**9.03**	−3.65	−5.87	−6.30	−6.90
Ser2	TCA	−4.19	−0.13	**7.56**	1.56	−7.57	−1.43
Ser3	TCG	−1.17	−3.68	2.63	**13.15**	−1.66	−3.17
Ser4	AGC	−7.82	−12.36	−3.81	2.95	**24.38**	**13.92**
	AGT	−4.89	−9.41	0.06	−3.42	**13.32**	**12.68**

(D)

Grouped by:	Parsimony Rule		Extended Wobble		Individual tRNA (No grouping)		Global Standard deviations
Amino Acid	Isoaccepting	Nonisoaccepting	Isoaccepting	Nonisoaccepting	Self	Other	
Alanine	6/2/0	0/2/6	6/2/0	0/2/6	2/0/0	0/0/2	21.73
Arginine	7/3/2	5/9/10	7/4/3	5/8/9	4/0/0	0/7/5	17.50
Glycine	4/0/2	6/0/4	6/0/2	4/0/4	3/0/0	2/0/4	18.18
Isoleucine	3/2/0	0/0/4	3/2/0	0/0/4	2/0/0	0/0/2	19.50

Table 1. *Continued*

(D)

Grouped by:	Parsimony Rule		Extended Wobble		Individual tRNA (No grouping)		Global Standard deviations
Leucine	7/3/0	2/12/12	7/4/1	2/11/11	4/0/0	0/6/6	20.26
Proline	4/2/2	1/3/4	4/2/2	1/3/4	2/0/0	0/0/2	9.12
Serine	10/0/0	0/14/12	10/2/0	0/12/12	4/0/0	0/7/5	30.81
Threonine	6/0/0	2/0/8	8/0/0	0/0/8	3/0/0	2/0/4	16.54
Valine	4/2/0	2/0/8	6/2/0	0/0/8	3/0/0	2/0/4	18.27
Total:	51/14/6	18/40/68	57/18/8	12/36/66	33/6/0	6/20/40	

Codon reuse measured over all pairs comprised of one codon and the next one that codes for the same amino acid. See also Table S1. (A) Co-occurrence counts, (B) standard deviations, (C) percent deviation from expected tRNA (left) and codon (right) reuse for Serine, coded by 6 codons and translated by 4 tRNA molecules. Positive deviations from expected indicate selection for tRNA reuse (bold, above 3; underlined, between 1 and 3; standard deviations, SD). (D) Codon pairs grouped into those with isoacceptors (sharing a tRNA) and those without, by parsimony or extended wobble rules, or by individual tRNA (no isoacceptors other than self). Within each group, pairs were classified as favored ($\geq +3$ SD), neutral (between -3 and $+3$ SD), or disfavored (≤ -3 SD); counts are tabulated as favored/neutral/disfavored for each amino acid.

genome (*Saccharomyces cerevisiae*). This genome is well annotated, providing precise information about which sequences are coding. For this genome, the analysis was conducted for the nine amino acids (alanine, arginine, glycine, isoleucine, proline, leucine, serine, threonine, and valine) that have at least two isoaccepting codons. First, we computed a correlation matrix for each amino acid: All instances of codons for a given amino acid were extracted, and the identity of the next synonymous codon in the same reading frame was determined. Based on this, the frequency of each possible codon pair was established and compared to the frequency expected under the assumption of random distribution, given the observed codon frequencies (product of individual frequencies). The results were then expressed as number of standard deviations from the expected (Z transform) or percent deviation from the expected, respectively. Note that we are not looking at pairs of consecutive codons but pairs of consecutive synonymous codons, which may be separated by any number of codons for other amino acids.

These correlation matrices show that successive synonymous codons are not chosen independently from one another (see correlation matrix for serine in Table 1; all other amino acids in Table S1 and Supplemental Information, available online). Indeed, in all matrices, identical codons followed each other more frequently than expected and more than any other off-diagonal element. For serine, the diagonal ranges between +6.3 and +13.4 standard deviations (SD) away from expected values (up to 24% more often than expected). More remarkably, favored pairs were not limited to the diagonal, indicating that bias was not simply directed toward reuse of the same codon. Most favored pairs (positive deviations by more than 3 SD) associated codons translated by the same tRNA using the parsimony of the wobbling rule of Percudani (Percudani et al., 1997). The Percudani rule states that tRNAs wobble with a synonymous codon only if there is no better tRNA for that codon. This is proposed to improve translation fidelity and to be favored in eukaryotes. In contrast, the extended wobbling rule of Crick (Crick, 1966) states that all tRNAs wobble and read all compatible codons. This hypothesis is well documented for bacteria.

In the example of serine, all four pairs of nonidentical codons decoded by the same tRNA in the parsimony rule (codons TCC and TCT, read by tRNA Ser1, and codons AGC and AGT, read by tRNA Ser4) are correlated at least 5.3 and up to 9.3 SD more frequently than expected (Table 1). No other pair of codons is otherwise favored (above 3 SD), except the pairs of identical codons. Conversely, pairs of codons read by nonisoaccepting tRNAs are nearly all underrepresented. Taking all relevant amino acids in consideration, positive deviations are much more frequent for isoaccepting codon pairs than for nonisoaccepting (Figure 1A). Thus, consecutive encodings of the same amino acid favored the usage of codons translated by the same tRNA.

Interestingly, among the three additional pairs of serine codons that are slightly overrepresented (between 1 and 3 SD), we find the TCA, TCG and TCG, TCA pairs. These pairs involve the two remaining codons, which are not predicted to be read by the same tRNAs according to the parsimony rule (each of these codons has its own tRNA), but would be in the extended wobble rule. Thus, among the 13 pairs of serine codons (out of 36 possible) that are overrepresented, we find all 12 pairs that associate isoaccepting codons according to the wobble theories. Out of the 24 pairs that associate codons that cannot be read by the same tRNA, only one was slightly overrepresented, and 17 were underrepresented by at least one and up to 10.5 standard deviations. Among the isoaccepting codon pairs, those defined by the parsimony rule were the most strongly overrepresented.

For each relevant amino acid (coded by more than two codons and read by more than two tRNAs; i.e., alanine, arginine, glycine, isoleucine, leucine, proline, serine, threonine, and valine), we counted the number of isoaccepting and nonisoaccepting codon pairs that were overrepresented by more than 3 SD, those that were underrepresented by more than 3 SD, or those that were neutral (between -3 and 3 SD from expected, Table 1D). This analysis was made using either the parsimony or the extended wobble rules to assign tRNAs and codons. For all amino acids, the overrepresentation of isoaccepting pairs was strong under the parsimony rule, and further increased in the extended rule. Clearly, codon correlation was strongly linked to

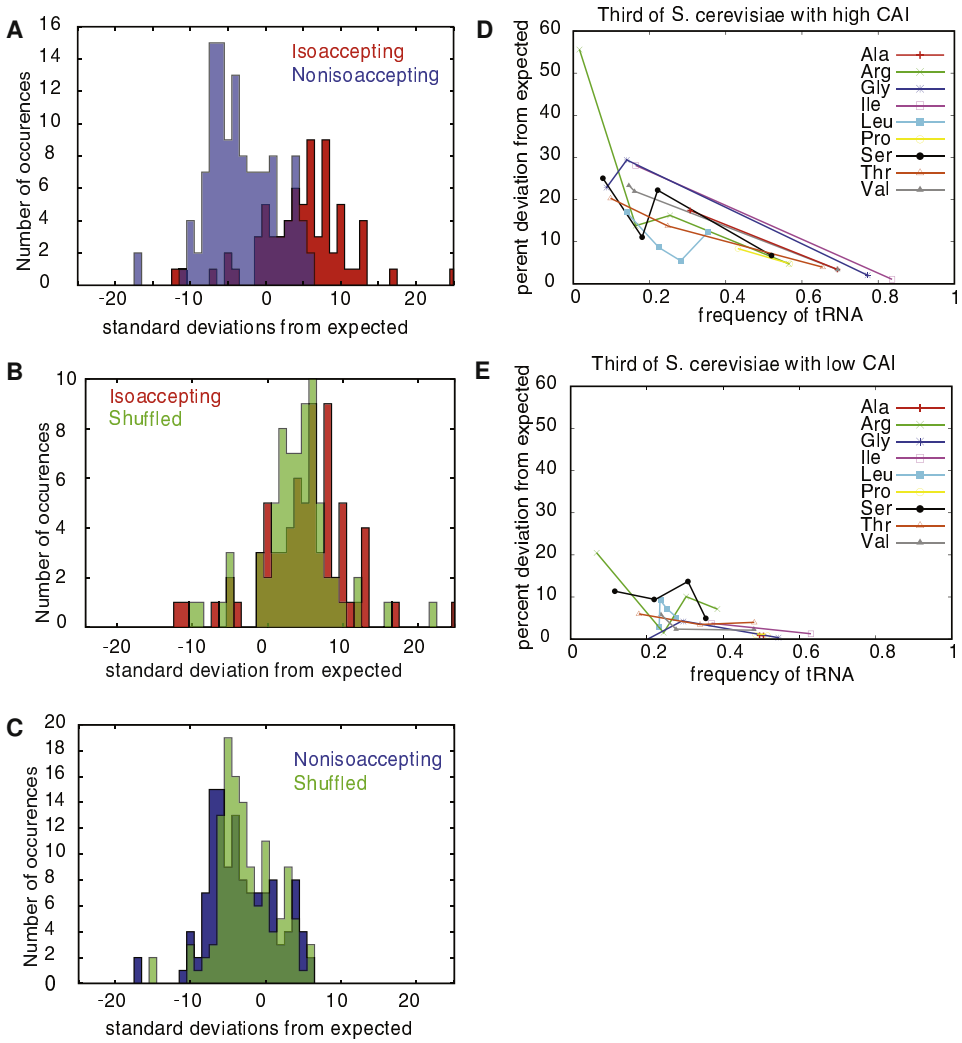


Figure 1. Controls for Mechanisms Other Than tRNA Recycling

(A) Observed z-scores (standard deviations from expected) for isoaccepting codons pairs (red) and nonisoaccepting codon pairs (blue), according to the parsimony of wobbling rule.

(B) Control for codon bias in isoaccepting codons. The codons within each gene were shuffled while maintaining the amino acid sequence. The mean of the distribution of the deviations from expected of the naturally occurring isoaccepting pairs is significantly (p-value 0.045) more positive (red) than that of the shuffled isoaccepting pairs (green).

(C) Control for codon bias in nonisoaccepting codons. The naturally occurring nonisoaccepting pairs (blue) are significantly more negative than those for the shuffled genes (green). The means of the two distributions are different with a p-value < 0.06.

(D) Control for tRNA abundance: Correlation between tRNA frequency and autocorrelation in highly expressed genes. Data based on the third of the *S. cerevisiae* genome with the highest CAI values. Correlation coefficient = −0.77, p-value = 0.00001).

(E) Control for tRNA abundance: As above, but in the least expressed genes. Data obtained from the third of the genome with the lowest CAI. Correlation coefficient = −0.5, p-value = 0.06.

the capacity of considered codons to be read by the same tRNA. Thus, we conclude that subsequent synonymous codons are correlated according to their reading tRNAs.

Several phenomena can result in codon correlation: (1) different genes are enriched in different codons, perhaps due to local variations in GC content, or (2) there is a selection pressure for codon ordering in open reading frames. In the first case, the correlation observed at the genomic level would be due to the accumulation of given codons in specific genes and should remain if codon distribution is shuffled in each gene individually. In the second case, such codon shuffling would erase correlation. In both cases, correlation should disappear when codon distribution is shuffled throughout the entire genome. Thus, we studied how autocorrelation changed with (Figure 1B and 1C, green) and without (Figure 1B, red and 1C, blue) shuffling the codons within each gene. For the shuffled genes, autocorrelation decreased for isoacceptor pairs (Figure 1B) and increased for nonisoaccepting pairs (Figure 1C). The hypothesis that the two

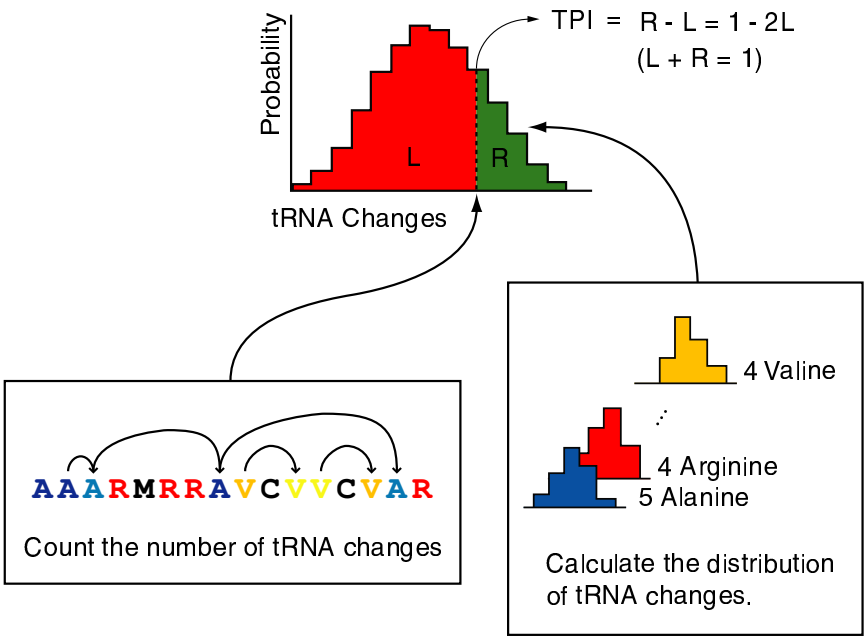


Figure 2. Computation of the tRNA Pairing Index for a Protein Sequence

Different amino acids are shown in different colors, with one shade of color per tRNA. tRNA pairing is quantified as follows: First, the number of changes of isoacceptor tRNA is summed. This total number of observed changes is then compared to the distribution of all possible numbers of changes, computed (by convolution) from those for each amino acid. The TPI is 1-2p, where p is the percentile (i.e., the value of the cumulative density function) of the global distribution at the given number of changes.

distributions have different means was confirmed at a p-value of 0.05 by Monte Carlo simulations. Thus, autocorrelation was not simply due to codon bias at the gene level, but to codon ordering within genes. In contrast, no significant nucleotide triplet ordering was found for noncoding DNA (data not shown).

If the correlation effect was simply due to accumulation of frequent codons in genes with biased codon composition, this effect should also be highest for frequent codons and not observed for rare codons. However, this was not the case. For example, the serine tRNA Ser3 (Table 1) is the least frequent and yet among the most correlated tRNAs (reuse 13% more frequent than expected). To study the effect of tRNA abundance on correlation, tRNA autocorrelation was plotted for each tRNA as a function of its usage frequency. Furthermore, instead of carrying out this analysis on the full genome, we compared tRNA autocorrelations in the third of the *S. cerevisiae* genes with the highest Codon Adaptation Index, CAI (Figure 1D) and in the third with the lowest CAI (Figure 1E). The CAI measures the bias toward usage of frequent codons. The genes with a high CAI are the most highly expressed. Remarkably, for the high CAI genes, tRNA autocorrelation is negatively correlated with tRNA usage (correlation coefficient of −0.77, p-value = 0.00001 by Monte Carlo simulations). In other words, in highly expressed genes, reuse of isoaccepting codons is strongest for rare tRNAs. For the third of the genome with the lowest CAI, a negative correlation is also observed although it is much weaker (correlation coefficient = −0.58, p value = 0.066). Autocorrelation was overall weaker in these genes. Thus, for infrequent tRNAs the pressure toward correlation is stronger, particularly in highly expressed genes.

These observations establish that evolutionary pressure selects for reusage of isoacceptor codons at successive intervals. This effect is not restricted to frequent codons and frequent tRNAs and is absent in noncoding DNA. Thus, these data suggest that reusage, i.e., recycling, of the same tRNA at successive encodings of the same amino acid may speed up translation, or favor fidelity.

tRNA Pairing Index and Its Correlation with Expression

Next, we measured isoacceptor codon autocorrelation at the gene level, using a tRNA pairing index (TPI). For an example of TPI, consider an amino acid X that occurs seven times in a protein and is translated by tRNAs A and B. We extract the corresponding codons from the gene sequence and represent them as a string of seven symbols, e.g., AABABBB, depending on the tRNA that decodes them. Highly autocorrelated cases are AAABBBB and BBBBAAA. The most anticorrelated case is BABABAB. The number of tRNA changes along the string quantifies autocorrelation (e.g., three changes in AABABBB). This number can be summed for all relevant amino acids (more than two codons and two tRNAs), giving a total number of changes in a given sequence. This observed number of changes is then compared to the average number of changes in random codon sequences coding for the same protein in which the random codons are drawn from the global codon distribution of the genome. Efficient recursions of these individual distributions for each amino acid have been presented (Friberg et al., 2006). The distributions of the individual amino acids are then convolved to a global background distribution. The TPI index is then defined as 1-2p, where p is the percentile (i.e., the value of the cumulative density function) of the global distribution at the given number of tRNA changes, as shown in Figure 2. By definition, the TPI ranges from −1 for the maximal number of tRNA changes (perfectly anticorrelated) to +1 for the minimal number of tRNA changes (perfectly autocorrelated). As expected from the correlation data, the distribution of the TPIs of all yeast genes was biased toward positive values (average TPI was 0.124).

To investigate how the TPI behaved in genes that are under variable pressure for rapid expression, TPI values were analyzed in the genes upregulated at least ten times in response to: 1-cell cycle progression (Cho et al., 1998; Spellman et al., 1998), 2-diauxic shift (DeRisi et al., 1997), 3-DNA damage (Gasch et al., 2001), 4-changes in zinc levels (Lyons et al., 2000), 5-phosphate deprivation (Ogawa et al., 2000), 6-ER stress

Table 2. tRNA Correlation in Individual Genes

(A)

Experiment	#	TPI	P Value	CAI	P Value
Cell cycle	32	**0.475**	**0.0020**	0.219	0.084
Diauxic shift	30	**0.428**	**0.0072**	0.220	0.078
DNA damage	68	**0.445**	**0.00008**	0.233	0.0029
Zinc levels	33	**0.461**	**0.0020**	0.256	0.0034
Phosphate dep.	24	**0.248**	0.19	0.197	0.29
ER stress	72	**0.267**	0.37	0.181	0.42
Sporulation	160	−0.0311	0.0019	0.152	0.00002
Glucose to glycerol	207	0.152	0.28	0.166	0.0087
Mating pheromone treatment	275	0.0052	0.0020	0.148	<0.00001
Arsenic time all	53	**0.405**	**0.0012**	0.185	0.46

(B)

Experiment		#	TPI	P Value	CAI	P Value
arsenic	fast	19	0.579	0.0013	0.192	0.36
arsenic	slow	17	0.211	0.30	0.168	0.34
Pheromone arrest	fast	72	0.395	0.00027	0.258	0.000010
elutriation	fast	73	0.524	<0.00001	0.232	0.0025
Cdc15 arrest	fast	74	0.345	0.0026	0.236	0.0016
Pheromone arrest	slow	74	−0.037	0.022	0.160	0.029
elutriation	slow	75	0.0071	0.069	0.146	0.00022
Cdc15 arrest	slow	75	0.0027	0.063	0.161	0.033

(A) Average TPIs (range: −1 to1) and codon adaptation indices (CAIs) for groups of genes that have been shown to be upregulated at least tenfold under the given conditions. Average CAI of the yeast database is 0 and the average TPI is 0.124. Groups with an average TPI higher than twice the genomic average are in bold in 3rd column. Groups with high and highly significant TPI ($p<0.01$) are in bold in 4th column. (B) TPI of rapidly and slowly responding genes upon arsenic poisoning or during cell cycle progression.

(Travers et al., 2000), 7-sporulation (Chu et al., 1998), 8-change from glucose to glycerol metabolism (Roberts and Hudson, 2006), 9-mating pheromone treatment (Roberts et al., 2000) and 10-arsenic treatment (Haugen et al., 2004). Their average TPI and CAI values were computed and compared to those of 100,000 random groups of the same size to assess their significance (Table 2A). Seven out of these ten categories (bold in 3rd column) showed an average TPI much higher than the genome average, and in five of these cases this TPI was highly significant ($p < 0.01$, bold in fourth column). Four of these five conditions correspond to acute responses (DNA damage, arsenic intoxication, zinc deprivation and diauxic shift). The fifth one, cell cycle progression, is the main fitness parameter for yeast. The genes induced in sporulation showed a negative TPI average that was highly significant. The genes induced by pheromone showed a neutral TPI (0.0052). Thus, codon correlation was highly increased in genes contributing to rapid growth or to acute stress responses.

Strikingly, in four out of the five categories with high TPI, the TPI values were clearly more significant than the corresponding CAI values (DNA damage, cell cycle, arsenic response and diauxic shift). These categories correspond to genes that are very dynamically regulated, i.e., that are rapidly turned on upon induction. Furthermore, the significance of the TPI correlated well with the rapidity with which they need to be regulated. The TPI was highest and most significant for genes induced by DNA damage, slightly lower for cell cycle genes and lower for genes involved in the diauxic shift. DNA damage must be repaired rapidly within the time of one cell cycle, while the timescale of the cell cycle is shorter than that of the diauxic shift. Most likely, cells are under high pressure to very rapidly fight drugs as potent as arsenic. The fact that in these genes the TPI values were clearly more significant than the CAI values suggests that codon order, and not just bias, was the primary cause for nonrandom codon usage.

To test whether high TPI contributes to induction speed, we investigated whether the genes whose transcription is induced fastest and slowest in response to the same stimulus show different TPIs (Table 2B). The gene categories found above to show a high TPI and for which the kinetics of induction are available (cell-cycle and arsenic response) were sorted further into rapidly and slowly induced groups, and their average TPI was compared. The genes most rapidly induced in response to arsenic poisoning showed the highest and most significant average TPI (0.579), compared to the genes that reacted more slowly (0.211; Table 2B). Thus, the selection for high TPI was strongest for the genes under highest pressure for rapid induction. The CAIs for the fast and slow groups were not significantly different from the average, indicating that codon order rather than bias correlated best with a rapid response to arsenic.

Time course data for cell cycle genes (Spellman et al., 1998) were obtained from cells synchronized at two different cell cycle stages: G1 (pheromone and centrifugal elutriation) and mitosis (*cdc15-2* mutant cells). For each experiment, the data was sorted by induction speed (see Methods) and the fastest 10% and the slowest 10% were compared. In all cases, the average TPI for the fast groups was higher than genomic average, while the TPI for the slowest groups was low. Thus, in rapidly induced genes a strong pressure selects for codons decoded by the same tRNA at consecutive encodings of the same amino acid. Therefore, the reuse of isoaccepting codons may support rapid translation.

Codon Correlation Enhances Translation Efficiency in *S. cerevisiae*

On the basis of these observations, we tested whether codon autocorrelation impacts translation speed. To this end, we designed a technique to compare the relative rates of translation of two sequences encoding the same peptide, in vivo. This method does not provide absolute translation rates, which can be modulated by many more parameters.

Our strategy relies on the fact that the distance between objects moving behind each other on a linear path is directly related to their velocity. For example, cars following each other on a highway get closer to each other when they slow down and are more dispersed when they speed up. Therefore, the local density of cars along the highway is inversely proportional to the local velocity of traffic, provided that the entry flux remains constant. Because ribosomes start translating at the beginning

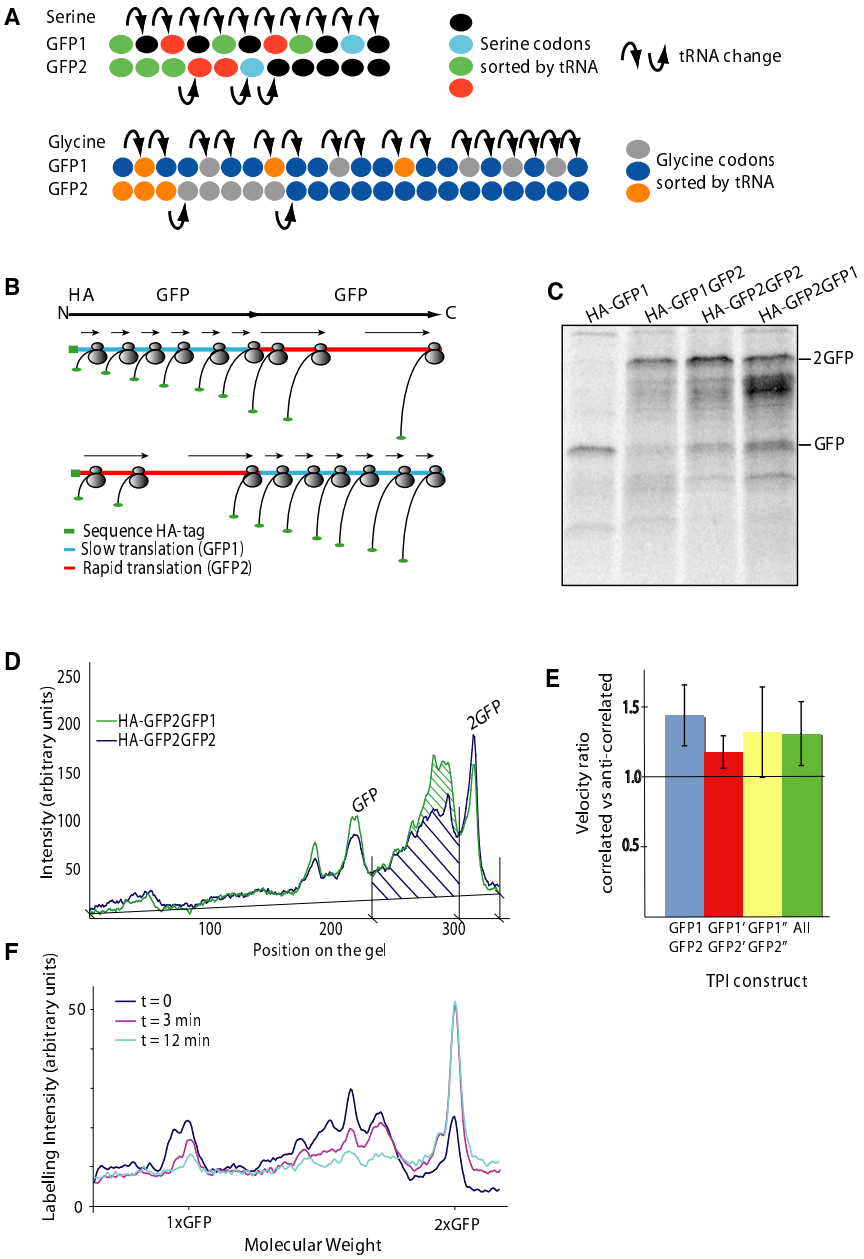


Figure 3. Codon Order Influences the Speed of Translation in Yeast Cells

(A) Two different green fluorescent proteins were synthesized with the same codons but ordered either to minimize (GFP1) or maximize (GFP2) isoacceptor tRNA reuse for each amino acid. For the construct sequences, see Table S2.

(B) Translation of double constructs containing two GFP proteins (HA-GFP1GFP2, HA-GFP2GFP1, HA-GFP2GFP2). Nascent chain length varies in proportion to ribosome position, while nascent chain density varies inversely proportional to local translation speed.

(C) Distribution of nascent chains and final products after a 3 min. labeling pulse for the indicated fusion genes (autoradiogram of PAGE gel). A construct expressing only 1 GFP fused to HA (HA-GFP1) shows the size of the first GFP domain (first lane).

(D) Signal intensity as a function of gel position for the indicated fusion genes. The faster translating GFP2 (purple) leads to more product while GFP1 (hatched green) accumulates more intermediary products 3 min after addition of the label. The first half of each construct is GFP2 and serves as a control.

(E) Data for three pairs consisting of one GFP2 (correlated codons) and one GFP1 (anticorrelated codons). Within pairs, sequences differed only in codon order. For all three constructs, the correlated were translated faster than the anticorrelated constructs. For each pair, data are based on at least 3 repeats (maximum 5). Error bars denote one standard deviation. See Table S3 for details on the constructs and tRNA usage.

(F) Pulse chase experiments. Cells were first labeled with ^{35}S methionine and cysteine. At t = 0 radioactivity was chased within excess of cold amino acids. Samples were taken at t = 0 (black), t = 3 min (purple) and t = 12 min (light blue). Nascent chains were separated by electrophoresis. Signal intensity along the gel is shown as in (D). Intermediary bands were extended into higher molecular weight products during the chase.

and stop at the end of the open reading frame, this rule applies to them too: Ribosome density along the transcript inversely reflects the local rate of translation (Figure 3B). Since each ribosome carries a nascent chain, the length of which is directly related to the position of the ribosome on the transcript, the abundance of nascent chains at given lengths directly reflects translation speed at the corresponding positions on the transcript. We thus investigated the effect of codon correlation on the density distribution of nascent chains. Local translation rates along a sequence might be influenced by codon sequence, the amino acids to be incorporated, the nascent chain folding (Kowarik et al., 2002), and by cotranslational binding of interaction partners. To focus on codon effects, we compared the translation rates of sequences coding for the exact same peptide, green fluorescent protein (GFP), a protein that does not interact with cellular factors, and changed only codon distribution.

Two DNA fragments (GFP1 and GFP2) were constructed with identical nucleotide and codon composition and coding both for the same GFP protein, but differing in the order of codon distribution (see Figure 3A, Table S2 and Table S3, and Supplemental Information). The overall codon distribution was chosen to match that of average *S. cerevisiae* genes. In GFP1, nonisoaccepting codons alternated regularly (anticorrelated codons), forcing a change of the tRNA at each occurrence of the same amino acid, for all relevant amino acids. Thus, the TPI value of GFP1 was −1. In contrast, codon distribution in GFP2 minimized

the number of events in which the ribosome had to change isoacceptor tRNAs (correlated codons, TPI of +1).

To compare the speed of translation through these sequences, the two fragments were fused into a single DNA string encoding a GFP-GFP fusion protein. The following combinations were constructed: GFP1-GFP2, GFP2-GFP2, GFP2-GFP1. A sequence encoding three repeats of the HA-epitope (3XHA) was added in frame to the 5′ end of each fusion to allow immunoprecipitation of nascent chains using an anti-HA antibody. Each open reading frame was put under the control of the galactose-inducible promoter *GAL1-10*, on a plasmid. Expression of the fusion gene in yeast was induced with galactose and the products were labeled with ^{35}S-labeled methionine. Upon cell lysis, the nascent chains and the final products were immunoprecipitated using anti-HA antibodies, and separated according to size by electrophoresis. Autoradiograms of the gels were used to quantify and compare the distribution of the nascent chains along the different transcripts.

These experiments established that the pattern of nascent chain distribution was highly reproducible for a given sequence, but distinct from one construct to the other (Figure 3C). Thus nucleotide order indeed affected translation speed. One band was observed for all constructs, at the size of a single GFP. In cold chase experiments, the band disappeared, while the full-length product continued to accumulate (Figure 3F, t = 0 black, t = 3 min. pink, t = 12 min. light blue). Thus, this band corresponds to a transient pause upon synthesis of the first GFP, and not to abortive transcription or translation. Strikingly, this band was always present, also for alternative GFP constructs (see below). Thus, it was dictated by peptide and not nucleotide sequence. Potentially, GFP induces the ribosome to pause while it folds, as is already documented for other peptides (Kowarik et al., 2002). Otherwise, the patterns of nascent chain distribution were clearly different when comparing GFP1 and GFP2 sequences. For example, when the two constructs started with the same GFP-coding sequence (such as GFP2 in Figures 3C and 3D) but diverged for the second copy of GFP, differences were not observed in the lower part of the gel corresponding to the translation of GFP2, but were obvious in the upper part (GFP1 versus GFP2). In all combinations, nascent chain density, and hence ribosome density, was highest for the region of the transcript corresponding to GFP1 and lowest for GFP2. Since the only difference between GFP1 and GFP2 is that codons were anticorrelated in GFP1 and correlated in GFP2, these data indicate that the correlated sequence was translated significantly faster than the anticorrelated one.

To assess the generality of this observation, two additional GFP pairs were constructed (GFP1′/GFP2′ and GFP1″/GFP2″; Table S2). While codon composition was different between pairs, it was not within pairs, where only the order of the codons varied. All sequences coded for the same peptide, i.e., GFP. GFP1′ and GFP1″ had a TPI of −1, while GFP2′ and GFP2″ had a TPI of +1. Furthermore, in the GFP1″/GFP2″ pair, the number of rare codons was increased by 30% (Table S3). As above, the effect of codon correlation on relative translation rates was determined by pulse-labeling. Again, the codon-correlated sequences were always expressed more efficiently than their anticorrelated variant (Figure 3E). Because in each pair both sequences were identical in terms of codon usage, the effects observed should be only due to codon order. To ascertain this conclusion, we considered whether unfavorable codon placement could affect the results. This could be the case, for example, if the frequency of adjacent, nonsynonymous, rare codons was increased in the slow constructs, or if unfavorable mRNA secondary structures, such as G-tetraplexes, were introduced. However, careful inspection of the sequences did not reveal accumulation of such features in the slow sequences. Thus, the variations in translation speed observed were primarily explained by codon correlation.

Precise quantification of thirteen distinct experiments carried out with the three GFP pairs showed that anticorrelated sequences were covered on average with 29% more nascent chains than correlated sequences. Furthermore, this result was not significantly affected by the increase in rare codon frequency in the GFP1″/GFP2″ pair. Thus, autocorrelation of rare codons also promoted translation speed. Since nascent chain density inversely relates to ribosome speed, we conclude that on average fully correlated sequences are expressed 29% faster than their fully anticorrelated counterparts. Thus, autocorrelation of isoaccepting codons substantially speeds up translation in vivo. This is consistent with codon correlation being strongest in genes under pressure for rapid expression.

tRNA Correlation as a Function of Distance between Encoded Amino Acids

Next, we investigated whether autocorrelation correlates with codon proximity. For each amino acid, codon correlation was determined as a function of the number of intervening, nonsynonymous codons between the paired codons. For example, for leucine at distance 20, the frequency of each isoacceptor codon pair was determined for all successive leucines that are 20 amino acids apart. These frequencies were examined for each amino acid individually or summed over all amino acids. To simplify the statistical analysis, frequencies were combined to form seven bins covering subsequent distance ranges such that the number of counts in each bin was approximately equal. The percentage deviation from expected value was then plotted as a function of intervening distance (Figure 4A, red lines/+ marks).

For the yeast genome, this study revealed that the bias toward consecutive use of isoaccepting codons slowly decays with distance. This decay is not observed when the codon distribution is shuffled within genes (Figure 4A, green dashes/x marks) or within the genome (Figure 4A, short blue dashes/✳ marks). The difference between the natural (red/+) and within-gene shuffled (green/x) sequences reflects the impact of codon order on correlation, while the difference between the within-genome shuffled (blue/✳) and within-gene shuffled sequences reflects the impact of codon bias. Shuffling the codons within the genome should give an autocorrelation of 0. Deviations from 0 give a visual estimate of the variance. When all amino acids are considered together, the effect of codon order is small but present for *S. cerevisiae*. At the amino acid level (Figure S1), codon order had a variable impact, highest for the amino acids with six codons (leucine, arginine, and serine). In all cases where autocorrelation was significant, this significance decayed slowly with distance (alanine and threonine) or not at all (leucine, serine, arginine, and proline).

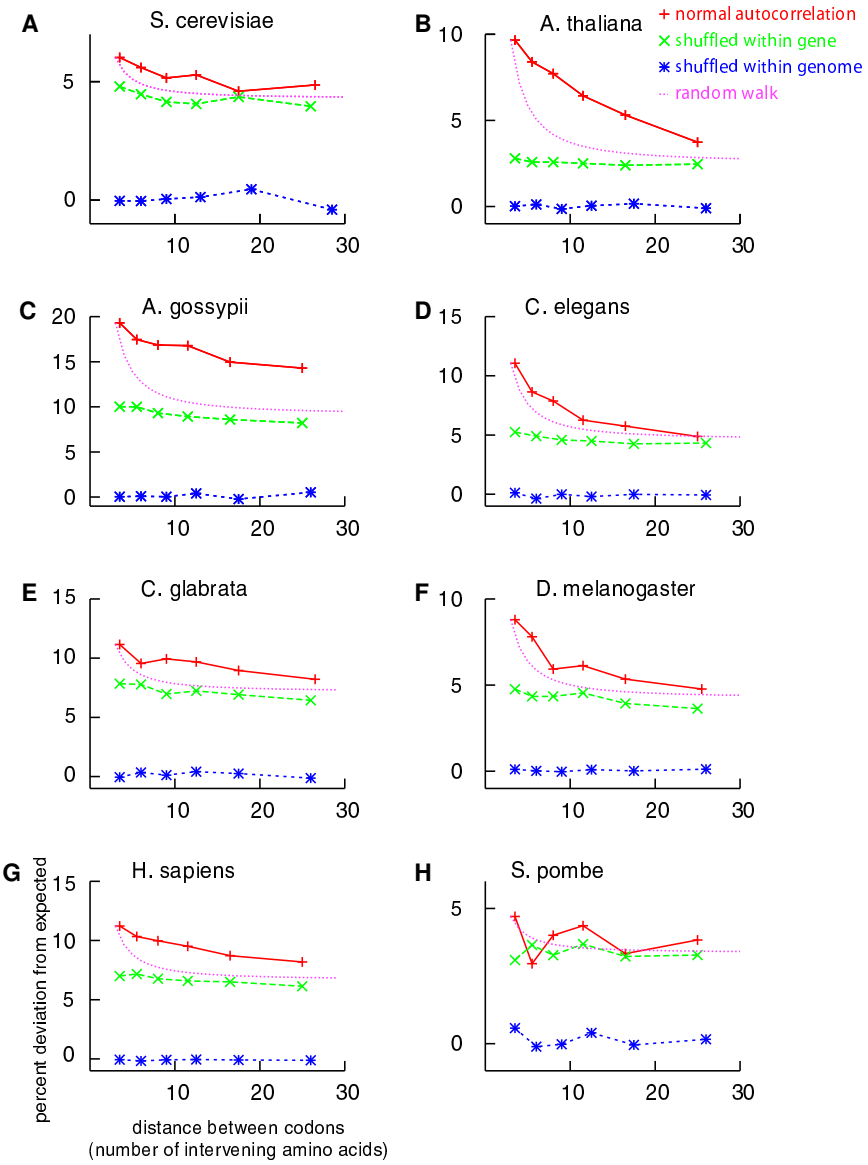


Figure 4. Percent Deviation from Expected Autocorrelation as a Function of the Distance between the Paired Codons

Autocorrelation is shown in red (with + marks) for the sum of all amino acids in the indicated genomes. Also shown are autocorrelation values after shuffling the codons within the genes (green, x marks), and throughout the genome (blue, ✱ marks). The pink dotted line shows the probability of returning to the origin (random walk in three dimensions model). See also Figure S1.

Codon Correlation in Other Eukaryotic Genomes

Next, we asked whether codon ordering is also present in other eukaryotes. Autocorrelation and its decay as a function of distance were computed for all amino acids for the genomes of *Arabidopsis thaliana* (Figure 4B), the filamentous fungus *Ashbya gossypii* (Figure 4C), *Caenorhabditis elegans* (Figure 4D), *Candida glabrata* (Figure 4E), *Drosophila melanogaster* (Figure 4F), *Homo sapiens* (Figure 4G), and *S. pombe* (Figure 4H). This analysis indicates that codon order has a strong impact on codon correlation in all organisms, except *S. pombe* and that the effect always decays with distance. In contrast, the effect due to codon bias (correlation upon codon shuffling at the gene level, green) showed no distance-dependent decay, as expected. Codon order had the strongest impact on codon correlation (difference between the red and green curves) in *A. gossypii* and *A. thaliana*. It had a strong impact in all animal genomes (*C. elegans, D. melanogaster,* and *H. sapiens*). In *A. thaliana, C. elegans*, and *D. melanogaster*, the effect of codon order decayed rapidly, while it decayed slowly in all other species. Thus, although in some organisms codon bias is the strongest cause of codon autocorrelation, codons are ordered beyond this effect and in a distance-dependent manner in virtually all genomes investigated.

DISCUSSION

tRNA Recycling Promotes Efficient Translation

Together, our results establish that sequences supporting tRNA reusage are expressed more efficiently than sequences that impose tRNA changes. Five main arguments lead to this conclusion. First, sequences varying in codon order, but not in codon composition or in the encoded amino acid sequence, are translated faster when the codons are correlated, at least in yeast. Second, all genomes investigated are biased toward autocorrelated sequences. Third, autocorrelation is strongest in highly expressed genes in yeast, and particularly in genes that are under pressure for rapid induction. Fourth, pressure for codon correlation is strongest for rare codons, especially in highly expressed genes, arguing that codon correlation strongly helps translation. Fifth, codon order decays with the distance separating two synonymous codons, suggesting that it reflects a memory-effect taking place during translation. Based on all these observations, we suggest that codon correlation allows the actual reuse of tRNAs by the ribosome (see below). In our experiments, the average gain in terms of speed was an impressive 30%. An average augmentation of 30% in translation speed is remarkable for a process that has been optimized by selection since the early days of evolution. Furthermore, a 30% gain in response speed is likely to have an important and decisive impact in the context of a competitive environment and on the time scale of many generations.

Our observation has three main corollaries. First, it provides an interesting and quantitative approach to evaluate the relative

contributions of the parsimony and extended wobble mechanisms for codon-tRNA assignments in vivo. Crick hypothesized that while the first two positions of the codon triplet strictly observe the base pairing rules, the third position is allowed to 'wobble'–the 5′ end of the anticodon can form hydrogen bonds with several bases at the 3′ end of the codon (Crick, 1966). While this wobbling is common in prokaryotes, our data suggest that in eukaryotes the parsimony rule is more relevant.

Observing that eukaryotes have more synonymous tRNAs than prokaryotes, Percudani introduced the "parsimony of wobbling" rule for eukaryotes. According to this hypothesis, codons only wobble when there is no perfect tRNA. The fact that synonym tRNAs have been retained in evolution is a strong argument for restricted codon reading and might reflect the need for higher specificity of decoding. A small amount of correlation is observed between codons read by the same codon in the extended wobble rule, arguing that some cross-reading takes place. However, autocorrelation is much stronger between codons read by the same tRNA under the parsimony rules. Thus, our data indicate that the parsimony rule is favored in vivo.

As a second corollary, our findings underline the selection pressure being exerted on codons. Indeed, our data indicate that beyond the selection pressure exerted on the nature of the amino acid being encoded, the codon choice is also under selection. This selection does not only reflect tRNA availability, but also the advantage there is in reusing the same tRNA (see below). Possibly, when the nature of the amino acid is not absolutely crucial, selective pressure might be more on codon choice. Statistical analysis of coding sequences suggested that codon distribution contributes to the modulation of translation speed along a given transcript (Thanaraj and Argos, 1996; Zhang et al., 2009), perhaps to adapt production speed to folding kinetics of the product (Kimchi-Sarfaty et al., 2007). It will be interesting to determine whether breaks in codon correlation also contribute to this process.

Third, our studies indicate that codon correlation is a predictor of genes under strong pressure for rapid and efficient translation. In unicellular organisms, such genes most likely promote stress-response and cell proliferation. It would be interesting to determine which classes of genes are under pressure for rapid expression in multicellular organisms. Interestingly, two classes of genes do not favor increased correlation: the genes involved in mating of haploids and in meiosis, in diploids. Mating induced genes show a neutral TPI, indicating that selection promotes neither codon autocorrelation nor anticorrelation. Therefore, there is apparently no selection for responding rapidly to partners. In contrast, codons were anticorrelated in meiotic genes. Interestingly, meiosis is induced by nitrogen starvation, i.e., when translation is limited by amino acid availability rather than tRNA complexity. Interestingly, theoretical modeling has established that codon usage helps to regulate translation during starvation. Indeed, frequently used tRNAs are exhausted most rapidly upon amino acid depletion (Elf et al., 2003; Dittmar et al., 2005). As a consequence, starvation response genes, the expression of which needs to be optimal under amino acid depletion, are enriched in rare codons (Elf et al., 2003). Thus, if codon autocorrelation leads to tRNA reusage, this process might speed up elongation when amino acids are abundant (see below), and slow it down under low amino acids levels. Therefore, the negative value of the TPI in meiosis genes is in excellent agreement with the hypothesis that codon autocorrelation promotes tRNA recycling. In summary, a systematic analysis of TPI of individual genes will be highly informative about the conditions of their expression.

Why Is Codon Autocorrelation Beneficial?

Figure 5 presents three scenarios for tRNAs behavior upon leaving the ribosome. They might diffuse away quickly, relative to translation speed, and be rapidly mixed with isoacceptors (Model A, Figure 5A). If so, no selection pressure would fix mutations that promote successive codons to be similar or dissimilar and codons should not be correlated beyond the correlation caused by codon bias.

In contrast, if tRNA diffusion is slow relative to both reloading and translation (Model B), a recently used tRNA would be more likely than any of its isoacceptors to still be in the vicinity of the ribosome at the next occurrence of the same amino acid (Figure 5B). In this case, it becomes advantageous to profit from its presence and reuse it. Hence, successive occurrences of isoaccepting codons would be likely to translate faster than when the ribosome must wait for arrival of another, appropriate tRNA. The advantage would be strongest for amino acids that are read by several competing tRNAs (keeping the same tRNA reduces complexity) and for codons that are close in the gene sequence (when the ribosome arrives at the second codon, the tRNA is then more likely to still be around). Thus, autocorrelation would be predicted to decay steeply at increasing intervals. If tRNA diffusion is modeled as a random walk in three dimensions, the probability that the tRNA comes back to the ribosome should decay with each time step (adjacent amino acids being translated) t as $b = O(t^{-3/2})$.

Finally, the tRNAs might remain physically associated with the ribosome (Model C, Figure 5C). Codon autocorrelation would enhance translation speed in this model, as in model B, but autocorrelation would now be predicted to decay much more slowly.

All our data establish that codon order is not neutral and disprove Model A. Thus, tRNA diffusion away from the ribosome is slower than translation and acylation. This last point confirms that tRNA acylation is not limiting translation (Zouridis and Hatzimanikatis, 2008). However, why should tRNA diffusion be slower than translation? The answer might come from comparing models B and C. Model B predicts that the decay of autocorrelation with the distance separating subsequent synonymous codons is sharp, while Model C predicts it to be slower. The random diffusion model, a random walk in three dimensions starting from the first point of the autocorrelation in the real sequences (red/+) to the average of the within-gene shuffled sequences (green dashes/x), is shown for each genome (Figure 4, purple dotted line). For all genomes, autocorrelation decays more slowly than the diffusion model predicts. Only in the nematode and fly does the decay approach the diffusion model. Thus, at least in *A. gossypii*, *C. glabrata*, *H. sapiens*, *A. thaliana*, and *S. cerevisiae,* Model C best explains our data. Thus, our data suggest that tRNAs are recycled through binding of out-going tRNAs to the ribosome. This association might be particularly strong for Leu, Ser, Arg and Pro tRNAs. Furthermore, given the

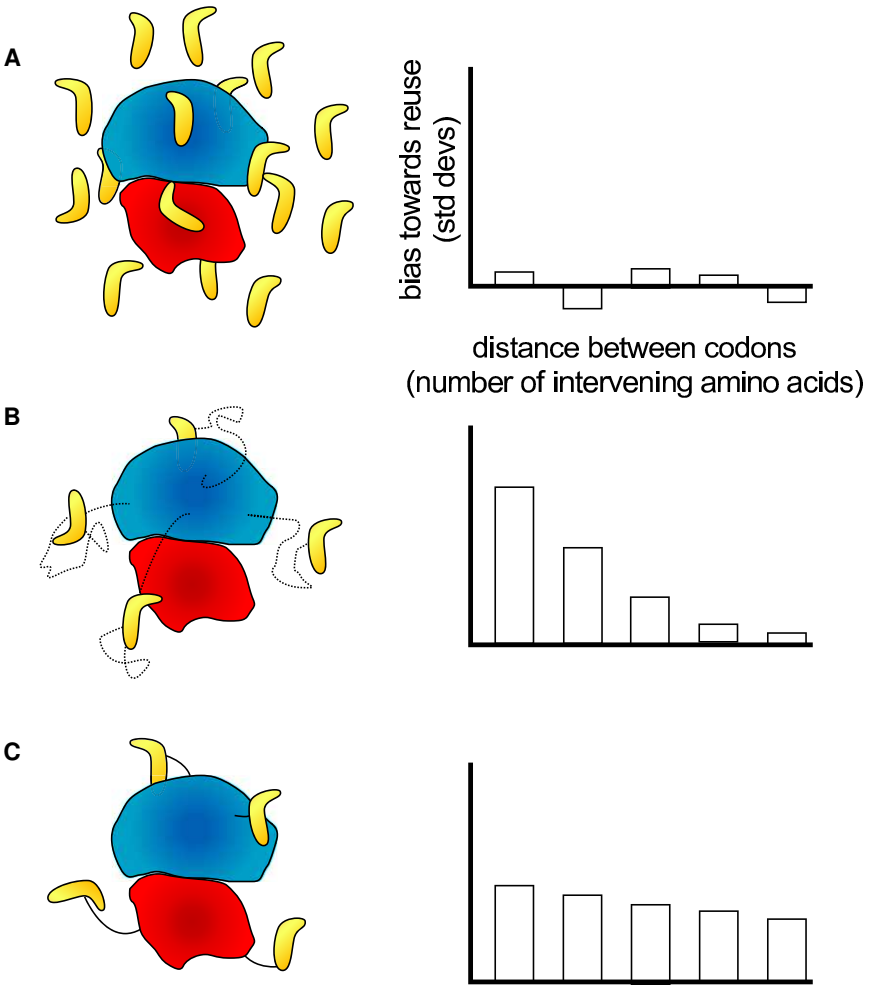


Figure 5. Three Hypotheses for Codon Correlation

The ribosome and tRNA size and shape are adapted from an *E. coli* crystal structure (Schuwirth et al., 2005).

(A) Hypothesis A: the tRNAs diffuse rapidly and can hardly be reused. We expect no significant codon autocorrelation.

(B) Hypothesis B: the tRNAs diffuse more slowly than the translation rate but move freely around the ribosome. If modeled by a random walk in three dimensions, an exponential decay of the autocorrelation is expected with time.

(C) Hypothesis C: the tRNAs remain associated with the ribosome. A slower decay of the autocorrelation than for Model B is expected.

phylogenetic relationships of fungi, animals and plants, tRNA recycling may have emerged early in eukaryotic evolution. We suggest that tRNA recycling contributes to the optimization of translation speed through local reduction of tRNA complexity around the ribosome.

Previous studies suggested that upon recharging, the tRNAs might remain bound with the tRNA-synthetase complex, which might itself remain associated with the ribosome (Irvin and Hardesty, 1972; Petrushenko et al., 2002). The tRNAs may remain bound to the elongation factor (Gaucher et al., 2000). Similarly, data from the Deutscher lab indicated that tRNAs might be channeled to the ribosome (Negrutskii and Deutscher, 1991; Stapulionis and Deutscher, 1995). Our observations are compatible with such ideas and argue in favor of at least some level of tRNA channeling taking place at the ribosome.

EXPERIMENTAL PROCEDURES

Yeast and Bacterial Strains and Methods

S^{35} labeling was carried out in YYB384 (MATa his3Δ200 ura3-52 trp1-Δ63 leu2 lys2-801 ade2-101), an S288c derivative. DNA amplification was carried out in *E. coli* XL1-Blue (supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac). All yeast and bacterial media and methods are standard (CurrentProtocolsMB).

Production of the Synthetic Genes

Oligonucleotides of 110 to 125 nucleotides were designed, such that one served as a template for amplification by the two others by PCR. The products (around 300 nucleotides) contained 30 nucleotides overlapping sequences at their extremities with each other and the cloning vector to allow recombination in vivo. Recombination of the fragments generated the GFP coding sequence and inserted it behind the *GAL1-10* promoter in a 2μ plasmid. Positive clones were screened by visualization of GFP fluorescence and confirmed by Western blotting and sequencing.

Radioactive Labeling

Cells grown overnight at 30°C in SGal-Leu liquid medium, were inoculated in 100 ml of SGal-Leu-Met-Cys liquid medium at OD_{600} = 0.2. After 3 hr at 30°C (OD_{600} = 0.5) with constant shaking, 8.0 OD_{600} cells were collected by centrifugation (3000 rpm, 2 min). The supernatant was removed and the cells resuspended in 300 μl of SGal-Leu and equilibrated at 22°C. At t_0, 0.3 mCi ^{35}SMet/Cys-Promix (Amersham-Pharmacia) were added. After 3 min incubation at 22°C, 50 μl of stopping buffer (Cycloheximide, 10mg/ml, NaF 1M) were added to stop freeze. The cells were briefly vortexed, rapidly spun down (13,000 rpm, 5 s), the supernatant was removed, and the pellet was shock frozen in liquid nitrogen.

Immunoprecipitation and Detection of Nascent Chains

100 μl of ice-cold lysis buffer (IP buffer with 1% SDS, Cycloheximide 0.1 mg/ml, NaF 10 mM) were added to the frozen pellet. An equal volume of acid-washed glass beads (Sigma) was used to break the cells by vortexing 4 min at 4°C (8 × 30 s with 30 s intervals on ice). 900 μl of ice-cold IP buffer (Tris Cl [pH 8.0] 50 mM, KCl 100 mM, SDS 0.1%, Triton X-100 1.0%, DOC 0.3%, EDTA 5.0 mM, Yeast protease inhibitor cocktail (Sigma, 1%) and (PMSF 0.1 mM) were added, the cell lysate was mixed, and cleared by centrifugation for 15 min at 14,000 rpm (4°C). The cleared lysate was pre-incubated with 50 μl of protein A-Sepharose (50% slurry in PBS) for 1 hr at 4°C to eliminate proteins binding nonspecifically to the beads. The lysate was then recovered and incubated with 2 μg of anti-HA antibodies (Santa Cruz) overnight at 4°C on a rotating wheel. Antibody-nascent chains complexes were recovered the next day by adding 75 μl of protein A-Sepharose beads (50% slurry in PBS) and incubating for 1 hr at 4°C on a rotating wheel. The beads were recovered by centrifugation (30 s at 1400 rpm). The supernatant was kept for control. The Sepharose beads were subsequently washed 4 times with 1 ml of ice-cold IP buffer. The beads were transferred into a new tube, and the immune complexes were eluted with 20 μl of 1.5× Laemmli buffer and incubation at 90°C for 5 min. For

separation of the nascent chains, an SDS-PAGE gradient gel 10 to 20% was prerun for 15 min at 15 mA at 4°C, the slots were rinsed with running buffer, and loaded with the radio-labeled samples (15 μl). These were separated at 15 mA at 4°C until the migration front exited the gel. The gel was then dried on a Whatmann paper and exposed on a Phospholmager plate for 72 hr. For pulse chase analysis of protein synthesis, the same procedure was applied, with the following adaptations. After 3 min of labeling, a first sample of 300 μl was collected, while 100 μl of nonradioactive Met/Cys saturated solution were added to the remainder of the cells for the chase. Additional samples (325μl) were collected at t = 6, 9 and 15 min. All samples were subsequently treated as described above.

Databases

For the computational studies, NCBI release 36 of the *S. cerevisiae* genome, NCBI Release 36 of the human genome, WormBase Release 170 of the *C. elegans* genome, NCBI release 84 of *Ashbya gossypii*, Release 81 of *S. pombe*, Release 80 of *Candida glabrata*, the November 2005 release of *Arabidopsis thaliana* from NCBI, and FlyBase Release 4.3 of *Drosophila melanogaster* were obtained from EMBL (Kulikova et al., 2004). All genome processing and computational scripts were written using the Darwin software package (Gonnet et al., 2000).

Autocorrelation of Codon and tRNA Usage

The autocorrelation results in Table 1 and Table S2 were computed as follows. For each sequence and for each of the nine amino acids with isoaccepting tRNAs (Ala, Arg, Gly, Iso, Leu, Pro, Ser, Thr, and Val) the number of consecutive pairs of codons were counted. The expected number of consecutive pairs was computed as the products of the frequencies of the individual codons of each pair in the database. A Z-transform, subtracting the expected counts from the observed and dividing by the standard deviation (estimated assuming a binomial distribution) was performed and the results expressed as standard deviations from the expected value. The same results were expressed as percentages by subtracting the expected counts from the observed counts and dividing by the expected counts.

Correlation of TPI with Gene Expression

To determine how TPI data correlates with expression data, we obtained groups of genes that are upregulated when subjected to various conditions from "Expression Connection," (Ball, 2001) http://db.yeastgenome.org/cgi-bin/expression/expressionconnection.pl. For each group in Table 2, the TPI and the average TPI were computed. The average values were compared to average values from groups of equally many randomly picked genes (10^5 repetitions), and p values were computed. The CAI values were computed using the method of Sharp and Li (Sharp and Li, 1987).

Expression data was sorted into fast and slow groups (Table 2B) in the following manner: the expression connection returns 55 genes that are upregulated by $\geq$ 10-fold when exposed to $NaAsO_2$ (Haugen et al., 2004). Expression levels were available for 0.5, 2 and 4 hr. The intensity ratio 2 hr/0.5 hr was used to split the genes into fastest third (the third with the largest indices), middle third, and the slowest third (the third with the smallest indices).

Similarly, the expression data of (Spellman et al., 1998) was analyzed. Time course data were reported after synchronizing cells at three different parts of the cell cycle (pheromone arrest in G1, sampled every 7 min. for 140 min., centrifugal elutriation in G1, sampled every 30 min. for 6.5 hr, and late in mitosis, sampled every 10 min. for 300 min.). The ratio of expression from each time pair t/(t-1) was computed for the 800 genes implicated in the cell cycle and the highest 3 ratios averaged. The data in Table 2B is for the 10% of the genes with the highest and lowest of these averages.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and three tables and can be found with this article online at doi:10.1016/j.cell.2010.02.036.

ACKNOWLEDGMENTS

We acknowledge the Swiss National Science Foundation (Project no. 107782, Origin and Function of Codon Bias) and the ETH Zurich for funding support. We thank the Computational Biochemistry Research Group, Steven A. Benner, Ari Helenius, three anonymous reviewers and Lara Szewczak for helpful discussions.

Received: November 8, 2007
Revised: June 5, 2009
Accepted: February 18, 2010
Published: April 15, 2010

REFERENCES

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W. (1998). A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle. Mol. Cell *2*, 65–73.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P., and Herskowitz, I. (1998). The Transcriptional Program of Sporulation in Budding Yeast. Science *282*, 699–705.

Crick, F.H.S. (1966). Codon-anticodon pairing: The Wobble Hypothesis. J. Mol. Biol. *19*, 548–555.

DeRisi, J., Iyer, V., and Brown, P. (1997). Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. Science *278*, 680–686.

Dittmar, K.A., Sorensen, M.A., Elf, J., Ehrenberg, M., and Pan, T. (2005). Selective Charging of tRNA Isoacceptors Induced by Amino-acid Starvation. EMBO Rep. *6*, 151–157.

Dong, H., Nilsson, L., and Kurland, C.G. (1996). Co-variation of tRNA Abundance and Codon Usage in Escherichia coli at Different Growth Rates. J. Mol. Biol. *260*, 649–663.

Duret, L. (2000). TRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. *16*, 287–289.

Elf, J., Nilsson, D., Tenson, T., and Ehrenberg, M. (2003). Selective Charging of tRNA Isoacceptors explains Patterns of Codon Usage. Science *300*, 1718–1722.

Friberg, M.T., Gonnet, P., Barral, Y., Schraudolph, N.N., and Gonnet, G.H. (2006). Measures of Codon Bias in Yeast, the tRNA Pairing Index and Possible DNA Repair Mechanisms. In Algorithms in Bioinformatics: 6th Intl. Workshop (WABI), (Beucher, P. & Moret, B. M. E., eds), vol. 4175, of Lecture Notes in Bioinformatics pp. 1–11, Springer Verlag, Berlin, Zurich, Switzerland.

Gasch, A., Huang, M., Metzner, S., Botstein, D., Elledge, S., and Brown, P. (2001). Genomic Expression Responses to DNA-Damaging Agents and the Regulatory Role of the Yeast ATR Homolog Mec1p. Mol. Biol. Cell *12*, 2987–3003.

Gaucher, E.A., Miyamoto, M.M., and Benner, S.A. (2000). Function-structure analysis of proteins using covarion-based evolutionary approaches: Elongation factors. Proc. Natl. Acad. Sci. USA *98*, 548–552.

Gonnet, G.H., Hallett, M.T., Korostensky, C., and Bernadin, L. (2000). Darwin v. 2.0: an interpreted computer language for the biosciences. Bioinformatics *16*, 101–103.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon Bias and Heterologous protein expression. Trends Biotechnol. *22*, 346–353.

Haugen, A., Kelley, R., Collins, J., Tucker, C., Deng, C., Afshari, C., Brown, J., Ideker, T., and Van Houten, B. (2004). Integrating Phenotypic and Expression Profiles to Map Arsenic-Response Networks. Genome Biol. *5*, R95.

Ikemura, T. (1985). Codon Usage and tRNA Content in Unicellular and Multicellular Organisms. Mol. Biol. Evol. *2*, 13–34.

Irvin, J.D., and Hardesty, B. (1972). Binding of Aminoacyl Transfer Ribonucleic Acid Synthetases to Ribosomes from Rabbit Reticulocytes. Biochemistry *11*, 1915–1920.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I., Sauna, Z., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity. Science *315*, 525–528.

Kowarik, M.T., Kung, S., Martoglio, B., and Helenius, A. (2002). Protein Folding during Cotranslational Translocation in the Endoplasmic Reticulum. Mol. Cell *10*, 769–778.

Kulikova, T., Aldebert, P., Althorpe, N., Baker, W., Bates, K., Browne, P., van den Broek, A., Cochrane, G., Duggan, K., Eberhardt, R., et al. (2004). The EMBL Nucleotide Sequence Database. Nucleic Acids Res. *32*, D27–D30.

Lyons, T., Gasch, A., Gaither, L., Botstein, D., Brown, P., and Eide, D. (2000). Genome-wide Characterization of the Zap1p Zinc-Responsive regulon in yeast. Proc. Natl. Acad. Sci. USA *97*, 7957–7962.

Negrutskii, B.S., and Deutscher, M.P. (1991). Channeling of Aminoacyl-tRNA for Protein Synthesis in vivo. Proc. Natl. Acad. Sci. USA *88*, 4991–4995.

Ogawa, N., DeRisi, J., and Brown, P. (2000). New Components of a System for Phosphate Accumulation and Polyphosphate Metabolism in Saccharomyces cerevisiae Revealed by Genomic Expression Analysis. Mol. Biol. Cell *11*, 4309–4321.

Percudani, R., Pavesi, A., and Ottonello, S. (1997). Transfer RNA Gene Redundancy and Translational Selection in Saccharomyces cerevisiae. J. Mol. Biol. *268*, 322–330.

Petrushenko, Z.M., Budkevich, T.V., Shalak, V.F., Negrutskii, B.S., and El'skaya, A.V. (2002). Novel Complexes of Mammalian Translation Elongation Factor eEF1A·GDP with Uncharged tRNA and Aminoacyl-tRNA Synthetase. Eur. J. Biochem. *269*, 4811–4818.

Roberts, G., and Hudson, A. (2006). Transcriptome Profiling of Saccharomyces cerevisiae during a Transition from Fermentative to Glycerol-based Respiratory Growth Reveals Extensive Metabolic and Structural Remodeling. Mol. Genet. Genomics *276*, 170–186.

Roberts, C., Nelson, B., Marton, B., Stoughton, R., Meyer, M., Bennett, H., He, Y., Dai, H., Walker, W., Hughes, T., et al. (2000). Signaling and Circuitry of Multiple MAPK Pathways Revealed by a Matrix of Global Gene Expression Profiles. Science *287*, 873–880.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H.D. (2005). Structures of the Bacterial Ribosome at 3.5 Å Resolution. Science *310*, 827–834.

Sharp, P.M., Stencio, M., Peden, J.F., and Lloyd, A.T. (1993). Codon usage: mutational bias, translational selection, or both? Biochem. Soc. Trans. *21*, 835–841.

Sharp, P.M., and Li, W.-H. (1987). The Codon Adaptation Index–a Measure of Directional Synonymous Codon Usage Bias, and its Potential Applications. Nucleic Acids Res. *15*, 1281–1295.

Spellman, P., Sherlock, G., Zhang, M., Iyer, V., Anders, K., Eisen, M., Brown, P., Botstein, D., and Futcher, B. (1998). Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. Mol. Biol. Cell *9*, 3273–3297.

Spirin, A.S. (1999). Ribosomes (New York: Kluwer Academic/Plenum), pp. 243.

Stapulionis, R., and Deutscher, M.P. (1995). A Channeled tRNA Cycle during Mammalian Protein Synthesis. Proc. Natl. Acad. Sci. USA *92*, 7158–7161.

Thanaraj, T., and Argos, P. (1996). Ribosome-mediated translational pause and protein domain organization. Protein Sci. *5*, 1594–1612.

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S., and Walter, P. (2000). Functional and Genomic Analyses Reveal an Essential Coordination between the Unfolded Protein Response and ER-Associated Degradation. Cell *101*, 249–258.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. *16*, 274–280.

Zouridis, H., and Hatzimanikatis, V. (2008). Effect of Codon Distributions and tRNA Competition on Protein Translation. Biophys. J. *95*, 1018–1033.

Cell

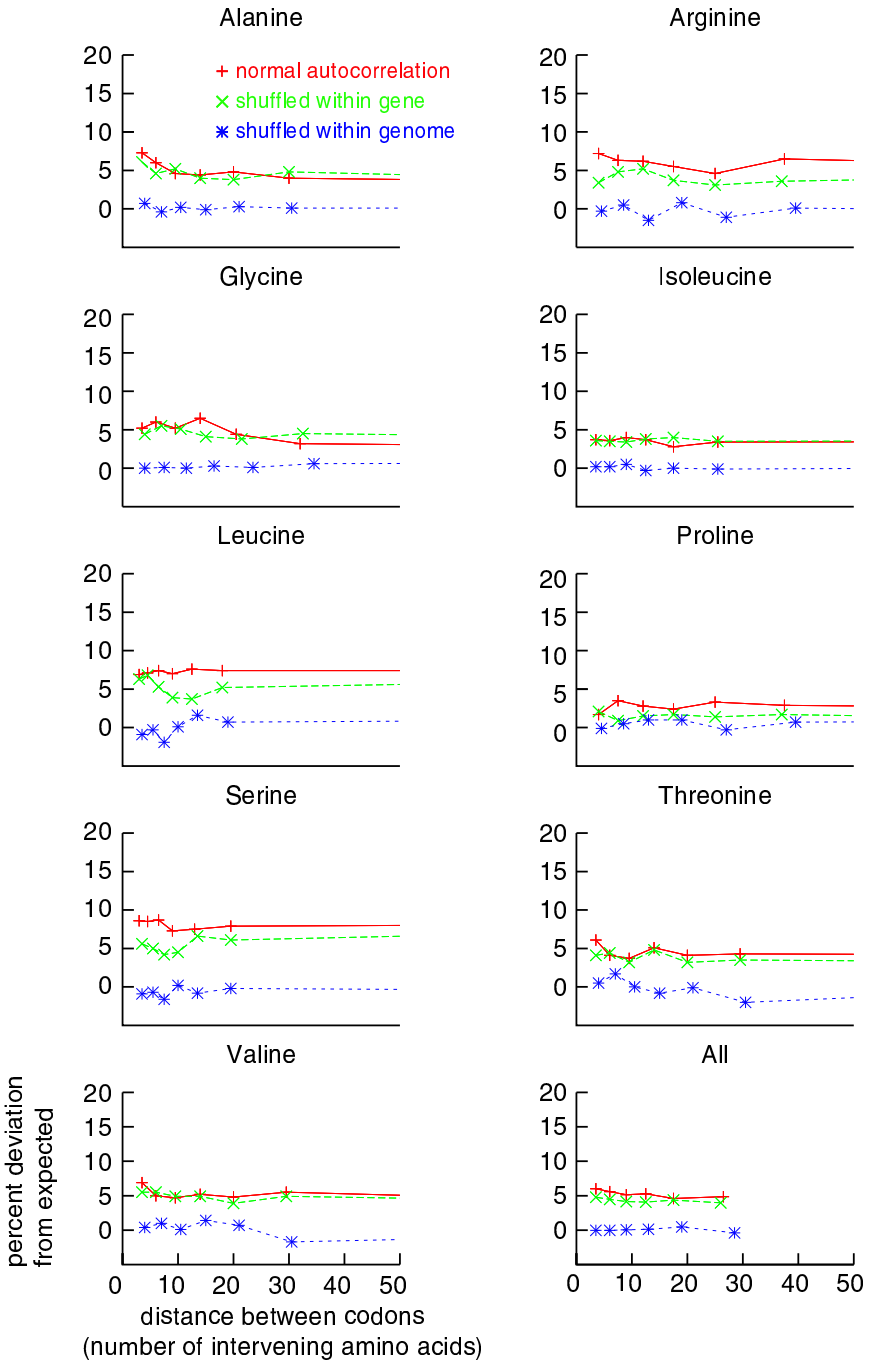


Figure S1. Autocorrelation for Each Amino Acid as a Function of Distance between Codons, Related to Figure 4

Decay of the percent deviation from expected autocorrelation as a function of the distance between the paired codons for each of the amino acids in *S. cerevisiae* individually and summed (red lines). Also shown are the decay of the autocorrelation for the codons shuffled within genes (green dashes) and shuffled within the genome (short blue dashes). Each amino acid was binned differently such as to keep approximately the same number of counts per bin.

Table S1. Co-occurence Counts, Related to Table 1

Co-occurrence counts, standard deviations (Z-score) and percent deviation from expected codon (top) and tRNA (bottom) pairing frequencies for individual amino acids in *S. cerevisiae.* Positive correlations above 3 standard deviations are bold-faced; those between 1 and 3 standard deviations are underlined. Isoaccepting codons are in boldfaced boxes.

Alanine

Codons						
	Co-occurence Counts				sum row	frequency row
	Ala1		Ala2			
	GCC	GCT	GCA	GCG		
GCC	7796	11929	9214	3672	32611	0.22262043731
GCT	12030	21170	14944	5591	53735	0.36682435984
GCA	9149	15041	14133	5269	43592	0.29758272065
GCG	3559	5589	5290	2111	16549	0.1129724822
sum column	32534	53729	43581	16643	146487	1
frequency col.	0.22209479339	0.36678340057	0.29750762866	0.11361417737	1	
	Standard Deviations From Expected					
	Ala1		Ala2			
	GCC	GCT	GCA	GCG		
GCC	**6.66797803008**	-0.3069783869	-5.1272912253	-0.5503313403		
GCT	0.91439110867	**11.1856959338**	-8.7361911909	-6.7206456764		
GCA	-5.6006797378	-7.9417437248	**10.706506618**	**4.57288663952**		
GCG	-1.9453158429	-6.3045231681	**5.31395263626**	**5.35718848659**		
	Percent Deviation From Expected					
	Ala1		Ala2			
	GCC	GCT	GCA	GCG		
GCC	7.63892123493	-0.2689826062	-5.0300990313	-0.8926125775		
GCT	0.80219676019	7.41227921702	-6.5215507016	-8.4201957773		
GCA	-5.5007296431	-5.9280289566	8.9756473928	6.38707666851		
GCG	-3.1682335307	-7.9226777255	7.44490091926	12.2752288793		
tRNAs						
	Co-occurrence Counts					
	Ala1	Ala2	sum row	freq row		
Ala1	52925	33421	86346	0.58944479715		
Ala2	33338	26803	60141	0.41055520285		
sum column	86263	60224	146487			
freq col.	0.58887819397	0.41112180603	1			
	Standard Deviations From Expected					
	Ala1	Ala2				
Ala1	**11.4033980102**	-12.669000838				
Ala2	-12.679097636	**14.493100861**				
	Percent Deviation From Expected					
	Ala1	Ala2				
Ala1	4.08620403106	-5.8529526157				
Ala2	-5.8666695477	8.40323650354				

Arginine

Codons								
	Co-occurence Counts						sum row	Frequency row
	Arg1			Arg2	Arg3	Arg4		
	CGA	CGC	CGT	CGG	AGA	AGG		
CGA	747	482	992	398	3393	1921	7933	0.0678138517
CGC	576	574	993	332	2822	1560	6857	0.05861585543
CGT	991	1029	2895	613	7539	3253	16320	0.13950864236
CGG	425	390	548	300	1942	1106	4711	0.04027115283
AGA	3488	2877	7697	1962	27939	11784	55747	0.47654339984
AGG	1852	1484	3215	1078	11938	5847	25414	0.21724709784
sum column	8079	6836	16340	4683	55573	25471	116982	1
frequency col.	0.06906190696	0.05843634063	0.13967960883	0.04003179976	0.47505599152	0.21773435229	1	
	Standard Deviations From Expected							
	Arg1			Arg2	Arg3	Arg4		
	CGA	CGC	CGT	CGG	AGA	AGG		
CGA	**8.5275088027**	0.8574285236	-3.503745984	**4.51933810251**	-6.2196690332	**4.69577842104**		
CGC	**4.71709529789**	**8.672415863**	1.14262314514	**3.47474602709**	-7.7382036459	1.74503684166		
CGT	-4.0733354761	2.44895061215	**13.0174055813**	-1.5818413942	-2.5141816885	-5.1181176212		
CGG	**5.53227085233**	**6.92151562431**	-4.3014560034	**8.11925462499**	-6.3174352285	2.51684345835		
AGA	-5.9325071289	-6.7640552396	-1.0523628233	-5.763337445	**10.1725943435**	-3.3943932442		
AGG	2.32956293613	-0.0287572338	-5.7068520115	1.90923078007	-1.2981139891	**4.31776315121**		
	Percent Deviation From Expected							
	Arg1			Arg2	Arg3	Arg4		
	CGA	CGC	CGT	CGG	AGA	AGG		
CGA	36.3466843953	3.97443569414	-10.475643551	25.325804768	-9.9670240668	11.2148951643		
CGC	21.6325377245	43.2500330884	3.67692049276	20.9480354458	-13.368055982	4.48729715162		
CGT	-12.074482318	7.8977052714	26.9975677695	-6.1714069036	-2.7591404681	-8.4545096846		
CGG	30.6282954235	41.6669326759	-16.721202779	59.0754039607	-13.225659539	7.82390812078		
AGA	-9.4024592073	-11.684821933	-1.1522073881	-12.083095924	5.49808023993	-2.9167561372		
AGG	5.51868995955	-0.0741471956	-9.4319657656	5.95967548694	-1.1187952632	5.66547616943		
tRNAs								
	Co-occurrence Counts							
	Arg1	Arg2	Arg3	Arg4	sum row	freqency row		
Arg1	9279	1343	13754	6734	31110	0.26593834949		
Arg2	1363	300	1942	1106	4711	0.04027115283		
Arg3	14062	1962	27939	11784	55747	0.47654339984		
Arg4	6551	1078	11938	5847	25414	0.21724709784		
sum column	31255	4683	55573	25471	116982	1		
freq col.	0.26717785642	0.04003179976	0.47505599152	0.21773435229	1			
	Standard Deviations From Expected							
	Arg1	Arg2	Arg3	Arg4				
Arg1	**11.005866561**	2.7807954642	-9.0203981683	-0.4971656762				
Arg2	2.95652239229	**8.11925462499**	-6.3174352285	2.51684345835				
Arg3	-7.3008804163	-5.763337445	**10.1725943435**	-3.3943932442				
Arg4	-2.9891741602	1.90923078007	-1.2981139891	**4.31776315121**				
	Percent Deviation From Expected							
	Arg1	Arg2	Arg3	Arg4				
Arg1	11.6350837289	7.83776687915	-6.9354655811	-0.5863207376				
Arg2	8.28848814884	59.0754039607	-13.225659539	7.82390812078				
Arg3	-5.5884491886	-12.083095924	5.49808023993	-2.9167561372				
Arg4	-3.5207069159	5.95967548694	-1.1187952632	5.66547616943				

Glycine

Codons						
	Co-occurence Counts				sum row	frequency row
	Gly1		Gly2	Gly3		
	GCC	GCT	GCA	GCG		
GCC	5702	10712	6290	3446	26150	0.19713086021
GCT	10850	31191	11748	6649	60438	0.45560974874
GCA	6224	11667	7758	4076	29725	0.22408087265
GCG	3416	6707	3984	2233	16340	0.12317851839
sum column	26192	60277	29780	16404	132653	1
frequency col.	0.19744747574	0.45439605588	0.22449548823	0.12366098015	1	
	Standard Deviations From Expected					
	Gly1		Gly2	Gly3		
	GCC	GCT	GCA	GCG		
GCC	**7.64795542041**	-11.253310888	**5.59966773231**	**3.77907950871**		
GCT	-10.395602628	**25.2638224063**	-16.491363591	-9.8215834338		
GCA	**4.738203636**	-16.704752586	**13.6276685558**	**6.69387117999**		
GCG	**3.38127875338**	-8.5740648768	**5.28681672733**	**4.76105849422**		
	Percent Deviation From Expected					
	GCC	GCT	GCA	GCG		
	Gly1		Gly2	Gly3		
GCC	10.4342875848	-9.8502933763	7.14485834295	6.56409363383		
GCT	-9.0781910006	13.5755010126	-13.414287254	-11.036151025		
GCA	6.0464500232	-13.622072129	16.2573166495	10.8867430435		
GCG	5.88007101685	-9.6679843559	8.60746083272	10.5106126159		
tRNAs						
	Co-occurrence Counts					
	Gly1	Gly2	Gly3	sum row	freq row	
Gly1	58455	18038	10095	86588	0.65274060896	
Gly2	17891	7758	4076	29725	0.22408087265	
Gly3	10123	3984	2233	16340	0.12317851839	
sum column	86469	29780	16404	132653		
freq col.	0.65184353162	0.22449548823	0.12366098015			
	Standard Deviations From Expected					
	Gly1	Gly2	Gly3			
Gly1	**11.1796865036**	-10.874115634	-6.1741478461			
Gly2	-11.545052917	**13.6276685558**	**6.69387117999**			
Gly3	-5.3359642353	**5.28681672733**	**4.76105849422**			
	Percent Deviation From Expected					
	Gly1	Gly2	Gly3			
Gly1	3.56680916538	-7.205324611	-5.7207909538			
Gly2	-7.6643539616	16.2573166495	10.8867430435			
Gly3	-4.9583813037	8.60746083272	10.5106126159			

Isoleucine

Codons						
	Co-occurence Counts				sum row	frequency row
		Ile1	Ile2			
	ATC	ATT	ATA			
ATC	12793	20879	11662		45334	0.25731199946
ATT	20855	38900	21947		81702	0.4637337314
ATA	11617	21743	15787		49147	0.27895426914
sum column	45265	81522	49396	0	176183	1
frequency col.	0.25692036122	0.46271206643	0.28036757235	0	1	
	Standard Deviations From Expected					
		Ile1	Ile2			
	ATC	ATT	ATA			
ATC	**10.9859820915**	-0.7178931413	-9.6520740791			
ATT	-0.9994810185	**6.35752127746**	-6.7975193632			
ATA	-9.3274757749	-7.0908316146	**17.8150304367**			
	Percent Deviation From Expected					
		Ile1	Ile2			
	ATC	ATT	ATA			
ATC	9.83729672477	-0.4652273085	-8.2468008665			
ATT	-0.6474582053	2.89779976509	-4.1891496637			
ATA	-7.9977491898	-4.3881706229	14.5710292857			
tRNAs						
	Co-occurrence Counts					
	Ile1	Ile2	sum row	freq row		
Ile1	93427	33609	127036	0.72104573086		
Ile2	33360	15787	49147	0.27895426914		
sum column	126787	49396	176183	1		
freq col.	0.71963242765	0.28036757235	1			
	Standard Deviations From Expected					
	Ile1	Ile2				
Ile1	**9.57355227566**	-11.910479304				
Ile2	-11.941760306	**17.8150304367**				
	Percent Deviation From Expected					
	Ile1	Ile2				
Ile1	2.19622833141	-5.6371609332				
Ile2	-5.6768482778	14.5710292857				

Leucine

Codons								
	Co-occurence Counts						sum row	Frequency row
	Leu1	Leu2	Leu3		Leu4			
	TTA	TTG	CTC	CTT	CTA	CTG		
TTA	20956	19998	3870	9443	10388	7587	72242	0.2786899159
TTG	19733	22352	3873	9056	9985	7913	72912	0.28127459301
CTC	3942	3786	1074	2011	2139	1807	14759	0.05693619319
CTT	9357	8679	2040	4577	4746	3785	33184	0.12801481367
CTA	10435	9992	2064	4631	5515	4318	36955	0.14256230229
CTG	7796	7945	1860	3544	4226	3797	29168	0.11252218193
sum column	72219	72752	14781	33262	36999	29207	259220	1
frequency col.	0.27860118818	0.28065735669	0.05702106319	0.12831571638	0.14273204228	0.11267263328	1	
	Standard Deviations From Expected							
	Leu1	Leu2	Leu3		Leu4			
	TTA	TTG	CTC	CTT	CTA	CTG		
TTA	**6.0865637385**	-2.0280190257	-3.9157543678	1.83214877609	0.7713422628	-6.224585592		
TTG	-4.2416418367	**13.7573315689**	-4.448424618	-3.15653016	-4.2210931164	-3.3881320839		
CTC	-2.6704316473	-5.5795851262	**8.02499841374**	2.70276493849	0.70919563451	**3.54419105439**		
CTT	1.18509507482	-6.6943829306	**3.41052794451**	**4.92882977706**	0.14048845772	0.75894816603		
CTA	1.40088376307	-3.8051715474	-0.9452278208	-1.6255212484	**3.34339821463**	2.40883211264		
CTG	-3.7222063539	-2.7091230917	**4.84142374793**	-3.2718299951	0.9810804239	**8.96311172121**		
	Percent Deviation From Expected							
	Leu1	Leu2	Leu3		Leu4			
	TTA	TTG	CTC	CTT	CTA	CTG		
TTA	4.1203608817	-1.3674247104	-6.0523559812	1.86860899462	0.74435024821	-6.7901350168		
TTG	-2.8570829821	9.2297518339	-6.8434974637	-3.2039690679	-4.0538443877	-3.678395107		
CTC	-4.1313254653	-8.5997789006	27.6180304813	6.1879617992	1.5388807422	8.66327133444		
CTT	1.21035088107	-6.8110275335	7.81175656387	7.49105484483	0.20226054221	1.23220690315		
CTA	1.35292401036	-3.6608548899	-2.0507363123	-2.3388753055	4.55645023613	3.70292044381		
CTG	-4.0638656858	-2.946585397	11.8332592772	-5.3093257531	1.50825487463	15.535514174		
tRNAs								
	Co-occurrence Counts							
	Leu1	Leu2	Leu3	Leu4	sum row	freqency row		
Leu1	20956	19998	13313	17975	72242	0.2786899159		
Leu2	19733	22352	12929	17898	72912	0.28127459301		
Leu3	13299	12465	9702	12477	47943	0.18495100687		
Leu4	18231	17937	12099	17856	66123	0.25508448422		
sum column	72219	72752	48043	66206	259220	1		
freq col.	0.27860118818	0.28065735669	0.18533677957	0.25540467557	1			
	Standard Deviations From Expected							
	Leu1	Leu2	Leu3	Leu4				
Leu1	**6.0865637385**	-2.0280190257	-0.6753406227	-3.6356410208				
Leu2	-4.2416418367	**13.7573315689**	-5.1625340149	-5.507472742				
Leu3	-0.5150952408	-8.7700645289	**8.81318575853**	2.14915972861				
Leu4	-1.4596589946	-4.7303366695	-1.4439446628	**7.70297150135**				
	Percent Deviation From Expected							
	Leu1	Leu2	Leu3	Leu4				
Leu1	4.1203608817	-1.3674247104	-0.5683700306	-2.5795133161				
Leu2	-2.8570829821	9.2297518339	-4.3237132391	-3.8882136723				
Leu3	-0.4340560437	-7.3616852193	9.18788449561	1.89576294704				
Leu4	-1.0365156983	-3.3457782514	-1.2731421583	5.73110828738				

Isoleucine

Codons						
	Co-occurence Counts				sum row	frequency row
		Ile1	Ile2			
	ATC	ATT	ATA			
ATC	12793	20879	11662		45334	0.25731199946
ATT	20855	38900	21947		81702	0.4637337314
ATA	11617	21743	15787		49147	0.27895426914
sum column	45265	81522	49396	0	176183	1
frequency col.	0.25692036122	0.46271206643	0.28036757235	0	1	
	Standard Deviations From Expected					
		Ile1	Ile2			
	ATC	ATT	ATA			
ATC	**10.9859820915**	-0.7178931413	-9.6520740791			
ATT	-0.9994810185	**6.35752127746**	-6.7975193632			
ATA	-9.3274757749	-7.0908316146	**17.8150304367**			
	Percent Deviation From Expected					
		Ile1	Ile2			
	ATC	ATT	ATA			
ATC	9.83729672477	-0.4652273085	-8.2468008665			
ATT	-0.6474582053	2.89779976509	-4.1891496637			
ATA	-7.9977491898	-4.3881706229	14.5710292857			
tRNAs						
	Co-occurrence Counts					
	Ile1	Ile2	sum row	freq row		
Ile1	93427	33609	127036	0.72104573086		
Ile2	33360	15787	49147	0.27895426914		
sum column	126787	49396	176183	1		
freq col.	0.71963242765	0.28036757235	1			
	Standard Deviations From Expected					
	Ile1	Ile2				
Ile1	**9.57355227566**	-11.910479304				
Ile2	-11.941760306	**17.8150304367**				
	Percent Deviation From Expected					
	Ile1	Ile2				
Ile1	2.19622833141	-5.6371609332				
Ile2	-5.6768482778	14.5710292857				

Proline

Codons						
	Co-occurence Counts				sum row	frequency row
	Pro1		Pro2			
	CCC	CCT	CCA	CCG		
CCC	3242	5737	6803	2472	18254	0.15945143256
CCT	5729	11400	14040	4413	35582	0.310814116
CCA	6947	14100	20272	5311	46630	0.4073200559
CCG	2379	4343	5324	1968	14014	0.12241439553
sum column	18297	35580	46439	14164	114480	1
frequency col.	0.15982704403	0.3107966457	0.40565164221	0.12372466806	1	
	Standard Deviations From Expected					
	Pro1		Pro2			
	CCC	CCT	CCA	CCG		
CCC	**6.08610077209**	0.86772499192	-7.2308879235	**4.53814388467**		
CCT	0.57177598287	**3.41396131974**	-3.5071577721	0.16336487169		
CCA	-6.0587470311	-3.4882101212	**10.7948227911**	-6.1915570673		
CCG	2.97011414893	-0.1931787225	-4.908754865	**5.66562667152**		
	Percent Deviation From Expected					
	Pro1		Pro2			
	CCC	CCT	CCA	CCG		
CCC	11.1231891927	1.12312466893	-8.1267274603	9.45462638713		
CCT	0.73913085431	3.08564034167	-2.7289701481	0.24143491316		
CCA	-6.7858988496	-2.7079455467	7.17116299727	-7.9434725097		
CCG	6.21407262543	-0.2870894463	-6.3467840512	13.502828313		
tRNAs						
	Co-occurrence Counts					
	Pro1	Pro2	sum row	freq row		
Pro1	26108	27728	53836	0.47026554857		
Pro2	27769	32875	60644	0.52973445143		
sum column	53877	60603	114480	1		
freq col.	0.47062368973	0.52937631027	1			
	Standard Deviations From Expected					
	Pro1	Pro2				
Pro1	**5.49268026919**	-5.2733164374				
Pro2	-5.2707842144	**5.076034968**				
	Percent Deviation From Expected					
	Pro1	Pro2				
Pro1	3.04502647325	-2.7070754138				
Pro2	-2.7031865513	2.40317446037				

Serine

Codons								
	Co-occurence Counts						sum row	Frequency row
	Ser1		Ser2	Ser3	Ser4			
	TCC	TCT	TCA	TCG	AGC	AGT		
TCC	6443	10525	7831	3748	3814	5713	38074	0.15638580148
TCT	10412	18012	12966	5816	6575	9600	63381	0.2603322079
TCA	7707	13412	11766	5101	5272	8262	51520	0.2116141328
TCG	3647	5934	5150	2607	2573	3723	23634	0.09707469749
AGC	3906	6200	5543	2724	3737	5030	27140	0.1114753021
AGT	5897	9378	8437	3739	4982	7280	39713	0.16311785823
sum column	38012	63461	51693	23735	26953	39608	243462	1
frequency col.	0.15613114162	0.26066080127	0.21232471597	0.09748954662	0.11070721509	0.16268657943	1	
	Standard Deviations From Expected							
	Ser1		Ser2	Ser3	Ser4			
	TCC	TCT	TCA	TCG	AGC	AGT		
TCC	**6.545482282**	**6.15560678154**	-2.8623757234	0.59847695961	-6.2317006466	-6.1925195433		
TCT	**5.29844069902**	**12.0153413797**	-4.3577360712	-4.6774841224	-5.3511065925	-7.1574244578		
TCA	-3.8197378578	-0.1530893951	**8.09125704713**	1.11695392574	-5.7834792176	-1.32959824		
TCG	-0.7133551204	-2.9224380027	1.88172660294	**6.34106795382**	-0.8541274453	-1.9821625344		
AGC	-5.1358746812	-10.549644903	-2.9262842934	1.52731027118	**13.4448397648**	**9.33570092995**		
AGT	-3.903533637	-9.779600981	0.05484862346	-2.1482597194	**8.91082975204**	**10.3300589996**		
	Percent Deviation From Expected							
	Ser1		Ser2	Ser3	Ser4			
	TCC	TCT	TCA	TCG	AGC	AGT		
TCC	8.38522675408	6.05175820937	-3.1302527475	0.97480565133	-9.5150694919	-7.7674978778		
TCT	5.21690850631	9.02525856108	-3.6511848169	-5.8745078229	-6.2954360189	-6.8976982618		
TCA	-4.1879859735	-0.1284099169	7.56040725026	1.55970213944	-7.5677294706	-1.4270830458		
TCG	-1.1654027454	-3.6759831854	2.62884611787	13.147704944	-1.6608094801	-3.171305373		
AGC	-7.8208157669	-12.359242983	-3.8089903003	2.95304811059	24.3762127917	13.9216886281		
AGT	-4.8937853025	-9.4055053721	0.05868743368	-3.424999588	13.3170086511	12.68003042		
tRNAs								
	Co-occurrence Counts							
	Ser1	Ser2	Ser3	Ser4	sum row	freqency row		
Ser1	45392	20797	9564	25702	101455	0.41671800938		
Ser2	21119	11766	5101	13534	51520	0.2116141328		
Ser3	9581	5150	2607	6296	23634	0.09707469749		
Ser4	25381	13980	6463	21029	66853	0.27459316033		
sum column	101473	51693	23735	66561	243462	1		
freq col.	0.41679194289	0.21232471597	0.09748954662	0.27339379451	1			
	Standard Deviations From Expected							
	Ser1	Ser2	Ser3	Ser4				
Ser1	**16.6182153894**	-5.3123724612	-3.3548633742	-12.981790928				
Ser2	-2.5307774459	**8.09125704713**	1.11695392574	-4.7852639495				
Ser3	-2.7716330665	1.88172660294	**6.34106795382**	-2.0853760448				
Ser4	-15.805680771	-1.8556690037	-0.6839108072	**21.1645420745**				
	Percent Deviation From Expected							
	Ser1	Ser2	Ser3	Ser4				
Ser1	7.34616863062	-3.4556895962	-3.3040996497	-7.3373296971				
Ser2	-1.6491356771	7.56040725026	1.55970213944	-3.9136569117				
Ser3	-2.735514453	2.62884611787	13.147704944	-2.5596499622				
Ser4	-8.9104590632	-1.511455622	-0.835733369	15.0559459669				

Threonine

Codons						
	Co-occurence Counts				sum row	frequency row
	Thr1		Thr2	Thr3		
	ACC	ACT	ACA	ACG		
ACC	7445	11732	9474	4313	32964	0.2103422752
ACT	11603	19112	15961	7018	53694	0.3426197708
ACA	9479	15779	15653	7204	48115	0.30702034253
ACG	4322	7065	7138	3418	21943	0.14001761148
sum column	32849	53688	48226	21953	156716	1
frequency col.	0.20960846372	0.34258148498	0.30772863013	0.14008142117	1	
	Standard Deviations From Expected					
	Thr1		Thr2	Thr3		
	ACC	ACT	ACA	ACG		
ACC	**6.58869519699**	**4.28987750347**	-6.8782876943	-4.5506849469		
ACT	**3.40759629949**	**5.63049387729**	-4.624056893	-5.9505086044		
ACA	-6.2415893629	-5.7992597154	**7.31179071399**	**5.77719228852**		
ACG	-4.1522430633	-5.3461068912	**4.79588011212**	**6.2699727873**		
	Percent Deviation From Expected					
	Thr1		Thr2	Thr3		
	ACC	ACT	ACA	ACG		
ACC	7.74967817887	3.88868791376	-6.6045817434	-6.5973896987		
ACT	3.09455274398	3.90022564051	-3.4023779327	-6.6945382905		
ACA	-6.0118247008	-4.2728567796	5.71806161246	6.8839942174		
ACG	-6.032008433	-6.0163569239	5.70916368886	11.1976261782		
tRNAs						
	Co-occurrence Counts					
	Thr1	Thr2	Thr3	sum row	freq row	
Thr1	49892	25435	11331	86658	0.55296204599	
Thr2	25258	15653	7204	48115	0.30702034253	
Gly3	11387	7138	3418	21943	0.14001761148	
sum column	86537	48226	21953	156716		
freq col.	0.5521899487	0.30772863013	0.14008142117			
	Standard Deviations From Expected					
	Thr1	Thr2	Thr3			
Thr1	**11.1908820084**	-8.2828132688	-7.636925744			
Thr2	-8.8232999959	**7.31179071399**	**5.77719228852**			
Thr3	-6.9012551317	**4.79588011212**	**6.2699727873**			
	Percent Deviation From Expected					
	Thr1	Thr2	Thr3			
Thr1	4.26384940275	-4.6204702775	-6.6575837545			
Thr2	-4.9329600561	5.71806161246	6.8839942174			
Thr3	-6.0222981563	5.70916368886	11.1976261782			

Valine

Codons						
	Co-occurence Counts				sum row	frequency row
	Val1		Val2	Val3		
	GTC	GTT	GTA	GTG		
GTC	6551	11732	5968	5615	29866	0.20074474378
GTT	11721	23622	11885	10611	57839	0.38876566113
GTA	5942	11766	7689	6631	32028	0.21527665753
GTG	5562	10535	6609	6337	29043	0.19521293757
sum column	29776	57655	32151	29194	148776	1
frequency col.	0.2001398075	0.38752890251	0.21610340377	0.19622788622	1	
	Standard Deviations From Expected					
	Val1		Val2	Val3		
	GTC	GTT	GTA	GTG		
GTC	**7.57315519675**	1.52993315872	-6.1869479985	-3.2725313843		
GTT	1.40449570429	**8.75308197112**	-5.7402059848	-7.213806279		
GTA	-5.9765391535	-6.0545371536	**9.44945264605**	**4.46242388793**		
GTG	-3.3539176828	-7.0589937356	**4.29113576336**	**8.61726644504**		
	Percent Deviation From Expected					
	Val1		Val2	Val3		
	GTC	GTT	GTA	GTG		
GTC	9.59659486383	1.36566996331	-7.5322806211	-4.1897498168		
GTT	1.25358586086	5.3881524714	-4.9139507848	-6.5079218898		
GTA	-7.3022164848	-5.2029274922	11.0908868242	5.50875112454		
GTG	-4.3123184669	-6.3971727506	5.30104222235	11.1940392244		
tRNAs						
	Co-occurrence Counts					
	Val1	Val2	Val3	sum row	freq row	
Val1	53626	17853	16226	87705	0.5895104049	
Val2	17708	7689	6631	32028	0.21527665753	
Val3	16097	6609	6337	29043	0.19521293757	
sum column	87431	32151	29194	148776		
freq col.	0.58766871001	0.21610340377	0.19622788622			
	Standard Deviations From Expected					
	Val1	Val2	Val3			
Val1	**11.3575050822**	-8.5561558475	-7.9775775118			
Val2	-8.6869680982	**9.44945264605**	**4.46242388793**			
Val3	-7.8966150242	**4.29113576336**	**8.61726644504**			
	Percent Deviation From Expected					
	Val1	Val2	Val3			
Val1	4.04434567755	-5.8055651613	-5.7185196079			
Val2	-5.9178733243	11.0908868242	5.50875112454			
Val3	-5.6871428853	5.30104222235	11.1940392244			

Table S2. Sequences of the Double Constructs of GFP Used in the Experiment, Related to Figure 3A

The eleven serine codons present in the double construct are highlighted in yellow. GFP1, GFP1' and GFP1'' are constructed to maximize the number of changes of isoaccepting tRNA used in translation while GFP2, GFP2' and GFP2'' are constructed to minimize this number.

GFP1	ATG	AGC	AAA	GGT	GAA	GAG	CTA	TTC	ACC	GGG	GTC	GTA	CCA	ATT	TTG	GTC	GAA	TTA	GAC	GGC	GAC	GTG	AAC
GFP2	ATG	AGC	AAA	GGG	GAA	GAA	CTA	TTC	ACA	GGG	GTA	GTA	CCA	ATA	CTA	GTA	GAA	CTA	GAC	GGG	GAC	GTA	AAC
GFP1'	ATG	TCT	AAG	GGT	GAA	GAG	TTA	TTC	ACT	GGA	GTT	GTA	CCA	ATT	TTG	GTT	GAA	CTT	GAC	GGT	GAC	GTG	AAC
GFP2'	ATG	TCA	AAA	GGT	GAA	GAA	CTT	TTC	ACA	GGT	GTT	GTT	CCT	ATA	CTT	GTT	GAA	TTA	GAC	GGT	GAC	GTT	AAC
GFP1''	ATG	TCT	AAG	GGT	GAA	GAG	TTA	TTT	ACT	GGA	GTT	GTA	CCA	ATT	TTG	GTT	GAA	CTT	GAT	GGT	GAT	GTG	AAT
GFP2''	ATG	TCA	AAA	GGT	GAA	GAA	CTT	TTT	ACA	GGT	GTT	GTT	CCT	ATA	CTT	GTT	GAA	TTA	GAT	GGT	GAT	GTT	AAT

...	GGT	CAC	AAG	TTC	TCC	GTC	TCA	GGA	GAA	GGC	GAG	GGT	GAC	GCC	ACG	TAC	GGG	AAA	CTC	ACC	TTG	AAG	TTC
...	GGA	CAC	AAA	TTC	AGT	GTG	AGT	GGA	GAA	GGA	GAA	GGA	GAC	GCA	ACA	TAC	GGA	AAA	CTG	ACA	CTG	AAA	TTC
...	GGG	CAC	AAA	TTC	TCA	GTT	TCT	GGT	GAG	GGA	GAA	GGT	GAC	GCT	ACA	TAC	GGG	AAG	TTA	ACT	TTG	AAA	TTC
...	GGT	CAC	AAA	TTC	TCA	GTT	TCG	GGT	GAA	GGT	GAA	GGT	GAC	GCT	ACA	TAC	GGT	AAA	TTA	ACA	TTA	AAA	TTC
...	GGG	CAT	AAA	TTT	TCA	GTT	TCT	GGT	GAG	GGA	GAA	GGT	GAT	GCT	ACA	TAT	GGG	AAG	TTA	ACT	TTG	AAA	TTT
...	GGT	CAT	AAA	TTT	TCA	GTT	TCG	GGT	GAA	GGT	GAA	GGT	GAT	GCT	ACA	TAT	GGT	AAA	TTA	ACA	TTA	AAA	TTT

...	ATA	TGC	ACA	ACT	GGC	AAA	TTA	CCC	GTT	CCA	TGG	CCC	ACG	CTA	GTA	ACT	ACA	TTC	TCC	TAC	GGT	GTT	CAA
...	ATA	TGC	ACA	ACA	GGC	AAA	CTC	CCA	GTG	CCA	TGG	CCA	ACG	CTT	GTG	ACG	ACC	TTC	TCA	TAC	GGC	GTC	CAA
...	ATA	TGC	ACG	ACT	GGT	AAG	CTA	CCT	GTA	CCA	TGG	CCA	ACA	TTA	GTT	ACT	ACA	TTC	TCG	TAC	GGA	GTG	CAA
...	ATA	TGC	ACA	ACA	GGT	AAA	TTA	CCT	GTT	CCT	TGG	CCT	ACG	TTA	GTT	ACG	ACT	TTC	AGT	TAC	GGT	GTT	CAA
...	ATA	TGT	ACG	ACT	GGT	AAG	CTA	CCT	GTA	CCA	TGG	CCA	ACA	TTA	GTT	ACT	ACA	TTT	TCG	TAT	GGA	GTG	CAA
...	ATA	TGT	ACA	ACA	GGT	AAA	TTA	CCT	GTT	CCT	TGG	CCT	ACG	TTA	GTT	ACG	ACT	TTT	AGT	TAT	GGT	GTT	CAA

...	TGT	TTC	AGT	AGA	TAC	CCA	GAC	CAC	ATG	AAA	CAG	CAC	GAC	TTT	TTT	AAG	TCT	GCA	ATG	CCT	GAA	GGA	TAC	GTG
...	TGT	TTC	TCA	AGA	TAC	CCG	GAC	CAC	ATG	AAA	CAA	CAC	GAC	TTT	TTT	AAA	TCG	GCA	ATG	CCC	GAA	GGC	TAC	GTC
...	TGC	TTC	TCT	CGT	TAC	CCT	GAC	CAC	ATG	AAG	CAG	CAC	GAC	TTC	TTC	AAA	AGT	GCA	ATG	CCA	GAA	GGT	TAC	GTT
...	TGC	TTC	AGT	CGT	TAC	CCA	GAC	CAC	ATG	AAA	CAA	CAC	GAC	TTC	TTC	AAA	TCT	GCT	ATG	CCA	GAA	GGT	TAC	GTT
...	TGT	TTT	TCT	CGT	TAT	CCT	GAT	CAT	ATG	AAG	CAG	CAT	GAT	TTT	TTT	AAA	AGT	GCA	ATG	CCA	GAA	GGT	TAT	GTT
...	TGT	TTT	AGT	CGT	TAT	CCA	GAT	CAT	ATG	AAA	CAA	CAT	GAT	TTT	TTT	AAA	TCT	GCT	ATG	CCA	GAA	GGT	TAT	GTT

...	CAA	GAA	AGG	ACC	ATT	TTT	TTT	AAA	GAC	GAT	GGC	AAC	TAC	AAG	ACA	AGA	GCC	GAG	GTT	AAA	TTT	GAA	GGT	GAT
...	CAA	GAA	AGA	ACC	ATA	TTT	TTT	AAA	GAC	GAT	GGC	AAC	TAC	AAA	ACC	AGA	GCG	GAA	GTC	AAA	TTT	GAA	GGT	GAT
...	CAA	GAG	CGG	ACT	ATT	TTC	TTC	AAG	GAC	GAC	GGG	AAC	TAC	AAA	ACG	AGA	GCT	GAA	GTA	AAG	TTC	GAG	GGT	GAC
...	CAA	GAA	CGT	ACT	ATA	TTC	TTC	AAA	GAC	GAC	GGA	AAC	TAC	AAG	ACT	CGG	GCT	GAA	GTG	AAG	TTC	GAA	GGA	GAC
...	CAA	GAG	CGG	ACT	ATT	TTT	TTT	AAG	GAT	GAT	GGG	AAT	TAT	AAA	ACG	AGA	GCT	GAA	GTA	AAG	TTT	GAG	GGT	GAT
...	CAA	GAA	CGT	ACT	ATA	TTT	TTT	AAA	GAT	GAT	GGA	AAT	TAT	AAG	ACT	CGG	GCT	GAA	GTG	AAA	TTT	GAA	GGA	GAT

...	ACT	TTA	GTT	AAC	CGA	ATA	GAA	TTG	AAG	GGG	ATT	GAT	TTT	AAA	GAG	GAT	GGT	AAC	ATA	CTT	GGT	CAC	AAA	TTA
...	ACT	CTT	GTT	AAC	AGG	ATC	GAA	TTA	AAA	GGT	ATC	GAT	TTT	AAA	GAA	GAT	GGT	AAC	ATC	TTA	GGT	CAC	AAG	TTA
...	ACT	TTG	GTT	AAC	CGT	ATA	GAA	CTA	AAG	GGA	ATT	GAC	TTC	AAA	GAG	GAC	GGT	AAC	ATT	TTA	GGG	CAC	AAG	TTG
...	ACT	TTA	GTG	AAC	AGA	ATA	GAG	CTA	AAG	GGA	ATA	GAC	TTC	AAG	GAG	GAC	GGA	AAC	ATT	CTA	GGA	CAC	AAG	CTA
...	ACT	TTG	GTT	AAT	CGT	ATA	GAA	CTA	AAG	GGA	ATT	GAT	TTT	AAA	GAG	GAT	GGT	AAT	ATT	TTA	GGG	CAT	AAG	TTG
...	ACT	TTA	GTG	AAT	AGA	ATA	GAG	CTA	AAG	GGA	ATA	GAT	TTT	AAG	GAG	GAT	GGA	AAT	ATT	CTA	GGA	CAT	AAG	CTA

...	GAA	TAT	AAC	TAT	AAT	TCA	CAT	AAT	GTA	TAT	ATT	ATG	GCA	GAT	AAG	CAG	AAA	AAT	GGA	ATC	AAG	GTT	AAT	TTT
...	GAG	TAT	AAC	TAT	AAT	TCC	CAT	AAT	GTT	TAT	ATT	ATG	GCT	GAT	AAG	CAA	AAG	AAT	GGT	ATT	AAG	GTT	AAT	TTT
...	GAA	TAC	AAC	TAC	AAC	TCT	CAC	AAC	GTG	TAC	ATA	ATG	GCA	GAC	AAA	CAG	AAG	AAC	GGT	ATT	AAA	GTT	AAC	TTC
...	GAG	TAC	AAC	TAC	AAC	TCT	CAC	AAC	GTG	TAC	ATT	ATG	GCT	GAC	AAG	CAA	AAG	AAC	GGA	ATT	AAG	GTG	AAC	TTC
...	GAA	TAT	AAT	TAT	AAT	TCT	CAT	AAT	GTG	TAT	ATA	ATG	GCA	GAT	AAA	CAG	AAG	AAT	GGT	ATT	AAA	GTT	AAT	TTT
...	GAG	TAT	AAT	TAT	AAT	TCT	CAT	AAT	GTG	TAT	ATT	ATG	GCT	GAT	AAG	CAA	AAG	AAT	GGA	ATT	AAG	GTG	AAT	TTT

...	AAA	ATT	AGA	CAT	AAT	ATC	GAA	GAT	GGT	AGT	GTG	CAA	CTA	GCT	GAT	CAT	TAT	CAA	CAG	AAT	ACA	CCA	ATT	GGA	GAT
...	AAG	ATT	CGA	CAT	AAT	ATT	GAG	GAT	GGT	TCC	GTT	CAA	TTA	GCT	GAT	CAT	TAT	CAG	CAG	AAT	ACT	CCC	ATT	GGT	GAT
...	AAG	ATA	AGA	CAC	AAC	ATT	GAG	GAC	GGA	TCA	GTA	CAA	CTT	GCT	GAC	CAC	TAC	CAG	CAA	AAC	ACA	CCT	ATA	GGT	GAC
...	AAG	ATT	AGA	CAC	AAC	ATT	GAG	GAC	GGG	TCT	GTA	CAA	CTA	GCT	GAC	CAC	TAC	CAG	CAG	AAC	ACT	CCA	ATT	GGG	GAC
...	AAG	ATA	AGA	CAT	AAT	ATT	GAG	GAT	GGA	TCA	GTA	CAA	CTT	GCT	GAT	CAT	TAT	CAG	CAA	AAT	ACA	CCT	ATA	GGT	GAT
...	AAG	ATT	AGA	CAT	AAT	ATT	GAG	GAT	GGG	TCT	GTA	CAA	CTA	GCT	GAT	CAT	TAT	CAG	CAG	AAT	ACT	CCA	ATT	GGG	GAT

...	GGT	CCT	GTT	TTA	CTT	CCG	GAT	AAT	CAT	TAT	TTG	TCT	ACT	CAA	TCG	GCT	CTG	TCT	AAA	GAT	CCT	AAT	GAG	AAG	CGT
...	GGT	CCT	GTT	TTA	TTA	CCT	GAT	AAT	CAT	TAT	TTG	TCT	ACT	CAG	TCT	GCC	TTG	TCT	AAG	GAT	CCT	AAT	GAG	AAG	CGT
...	GGG	CCA	GTT	TTA	CTA	CCT	GAC	AAC	CAC	TAC	TTG	TCT	ACT	CAA	AGT	GCA	TTA	TCT	AAA	GAC	CCA	AAC	GAA	AAG	AGG
...	GGG	CCA	GTA	TTG	TTG	CCA	GAC	AAC	CAC	TAC	TTG	TCT	ACT	CAG	TCT	GCA	TTG	TCT	AAG	GAC	CCA	AAC	GAG	AAG	AGG
...	GGG	CCA	GTT	TTA	CTA	CCT	GAT	AAT	CAT	TAT	TTG	TCT	ACT	CAA	AGT	GCA	TTA	TCT	AAA	GAT	CCA	AAT	GAA	AAG	AGG
...	GGG	CCA	GTA	TTG	TTG	CCA	GAT	AAT	CAT	TAT	TTG	TCT	ACT	CAG	TCT	GCA	TTG	TCT	AAG	GAT	CCA	AAT	GAG	AAG	AGG

...	GAT	CAT	ATG	GTA	TTA	CTG	GAA	TTT	GTT	ACA	GCG	GCC	GGA	ATC	ACT	CAT	GGT	ATG	GAT	GAA	TTG	TAT	AAA	TAA	TAA
...	GAT	CAT	ATG	GTT	TTG	TTG	GAG	TTT	GTT	ACT	GCC	GCC	GGT	ATT	ACT	CAT	GGT	ATG	GAT	GAG	TTG	TAT	AAG	TAA	TAA
...	GAC	CAC	ATG	GTG	TTG	CTA	GAG	TTC	GTT	ACA	GCT	GCT	GGA	ATT	ACT	CAC	GGT	ATG	GAC	GAA	TTG	TAC	AAG	TGA	TAA
...	GAC	CAC	ATG	GTA	TTG	TTG	GAG	TTC	GTA	ACT	GCA	GCA	GGG	ATT	ACT	CAC	GGG	ATG	GAC	GAG	TTG	TAC	AAG	TGA	TAA
...	GAT	CAT	ATG	GTG	TTG	CTA	GAG	TTT	GTT	ACA	GCT	GCT	GGA	ATT	ACT	CAT	GGT	ATG	GAT	GAA	TTG	TAT	AAG	TGA	TAA
...	GAT	CAT	ATG	GTA	TTG	TTG	GAG	TTT	GTA	ACT	GCA	GCA	GGG	ATT	ACT	CAT	GGG	ATG	GAT	GAG	TTG	TAT	AAG	TGA	TAA

Table S3. tRNA Reusage in All Constructs, Related to Figure 3E

For each amino acid, the number of occurrences of that amino acid, the isoaccepting tRNAs with the codons they read and the exact sequences of occurrences of each tRNA, the number of tRNA changes for each sequence and the increase in the number of rare codons from the GFP to the GFP" sequences are tabulated for all GFP1 and GFP2 constructs.

Number of **Alanines** in GFP	tRNA	Codons	TPI								tRNA changes	change in rare codons GFP" - GFP
8	A	GCT, GCC	GFP1	A	B	A	B	A	A	B	A	6	
	B	GCA, GCG	GFP2	B	B	B	A	A	A	A	A	1	
			GFP1'	A	B	A	B	A	B	A	A	6	
			GFP2'	A	A	A	A	A	B	B	B	1	
			GFP1"	A	B	A	B	A	B	A	A	6	
			GFP2"	A	A	A	A	A	B	B	B	1	0

Number of **Arginines** in GFP	tRNA	Codons								tRNA changes	change in rare codons GFP" - GFP
6	A	CGT, CGC, CGA	GFP1	C	D	C	A	C	A	5	
	B	CGG	GFP2	C	C	C	D	A	A	2	
	C	AGA	GFP1'	A	B	C	A	C	D	5	
	D	AGG	GFP2'	A	A	B	C	C	D	3	
			GFP1"	A	B	C	A	C	D	5	
			GFP2"	A	A	B	C	C	D	3	1

Number of **Glutamines** in GFP	tRNA	Codons										tRNA changes	change in rare codons GFP" - GFP
8	A	CAA	GFP1	A	B	A	B	A	A	B	A	6	
	B	CAG	GFP2	A	A	A	A	A	B	B	B	1	
			GFP1'	A	B	A	B	A	B	A	A	6	
			GFP2'	A	A	A	A	A	B	B	B	1	
			GFP1"	A	B	A	B	A	B	A	A	6	
			GFP2"	A	A	A	A	A	B	B	B	1	0

Number of **Glutamic acids** in GFP	tRNA	Codons																		tRNA changes	change in rare codons GFP" - GFP
16	A	GAA	GFP1	A	B	A	A	B	A	A	B	A	A	B	A	A	B	A	A	10	
	B	GAG	GFP2	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	1	
			GFP1'	A	B	A	B	A	A	B	A	B	A	B	A	B	A	B	A	14	
			GFP2'	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	1	
			GFP1"	A	B	A	B	A	A	B	A	B	A	B	A	B	A	B	A	14	
			GFP2"	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	1	3

Number of **Glycines** in GFP	tRNA	Codons																						tRNA changes	change in rare codons GFP" - GFP
22	A	GAA	GFP1	A	C	A	A	B	A	A	C	A	A	B	A	A	C	A	A	B	A	B	A	B	A	16	
	B	GAG	GFP2	C	C	C	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	B	B	2	
	C		GFP1'	A	B	A	C	A	B	A	C	A	B	A	C	A	B	A	C	A	B	A	C	B	A	21	
			GFP2'	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	C	C	C	C	C	2	
			GFP1"	A	B	A	C	A	B	A	C	A	B	A	C	A	B	A	C	A	B	A	C	B	A	21	
			GFP2"	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	C	C	C	C	C	2	3

Number of **Isoleucines** in GFP	tRNA	Codons	TPI														tRNA changes	change in rare codons GFP" - GFP
12	A	ATT, ATC	GFP1	A	B	A	B	A	B	A	A	A	A	A	A	A	6	
	B	ATA	GFP2	B	B	B	A	A	A	A	A	A	A	A	A	A	1	
			GFP1'	A	B	A	B	A	A	B	A	B	A	B	A		10	
			GFP2'	B	B	B	B	B	A	A	A	A	A	A	A		1	
			GFP1"	A	B	A	B	A	A	B	A	B	A	B	A		10	
			GFP2"	B	B	B	B	B	A	A	A	A	A	A	A		1	2

Number of **Leucines** in GFP	tRNA	Codons																				tRNA changes	change in rare codons GFP" - GFP
19	A	TTA	GFP1	D	B	A	C	B	A	D	A	B	C	A	D	A	C	B	D	A	D	B	18	
	B	TTG	GFP2	D	D	D	D	D	C	C	C	A	A	A	A	A	A	B	B	B	B	B	3	
	C	CTT, CTC	GFP1'	A	B	C	A	B	D	A	B	D	A	B	C	A	D	B	A	B	D	B	18	
	D	CTA, CUG	GFP2'	C	C	A	A	A	A	A	A	D	D	D	D	B	B	B	B	B	B	B	3	
			GFP1"	A	B	C	A	B	D	A	B	D	A	B	C	A	D	B	A	B	D	B	18	
			GFP2"	C	C	A	A	A	A	A	A	D	D	D	D	B	B	B	B	B	B	B	3	-1

Number of **Lysines** in GFP	tRNA	Codons																					tRNA changes	change in rare codons GFP" - GFP
20	A	AAA	GFP1	A	B	A	B	A	A	B	A	B	A	B	A	A	B	A	B	A	A	B	A	16	
	B	AAG	GFP2	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	1	
			GFP1'	B	A	B	A	B	B	A	B	A	B	B	A	B	A	B	A	B	A	B	B	16	
			GFP2'	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	1	
			GFP1"	B	A	B	A	B	B	A	B	A	B	B	A	B	A	B	A	B	A	B	B	16	
			GFP2"	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	1	4

Number of **Prolines** in GFP	tRNA	Codons	TPI											tRNA changes	change in rare codons GFP" - GFP
10	A	CCT, CCC	GFP1	B	A	B	A	B	B	A	B	A	A	7	
	B	CCA, CCG	GFP2	B	B	B	B	B	B	A	A	A	A	1	
			GFP1'	B	A	B	B	A	B	A	B	A	B	8	
			GFP2'	A	A	A	A	B	B	B	B	B	B	1	
			GFP1"	B	A	B	B	A	B	A	B	A	B	8	
			GFP2"	A	A	A	A	B	B	B	B	B	B	1	

Number of **Serines** in GFP	tRNA	Codons													tRNA changes	change in rare codons GFP" - GFP
11	A	TCT, TCC	GFP1	D	A	B	A	D	A	B	D	A	C	A	10	
	B	TCA	GFP2	D	D	D	B	B	C	A	A	A	A	A	3	
	C	TCG	GFP1'	A	B	A	C	A	D	A	B	A	D	A	10	
	D	AGT, AGC	GFP2'	B	B	C	D	D	A	A	A	A	A	A	3	
			GFP1"	A	C	A	B	A	D	A	B	A	D	A	10	
			GFP2"	B	B	C	D	D	A	A	A	A	A	A	3	0

Number of **Threonines** in GFP	tRNA	Codons																tRNA changes	change in rare codons GFP" - GFP
15	A	ACT, ACC	GFP1	A	C	A	B	A	C	A	B	A	B	A	B	A	B	A	14	
	B	ACA	GFP2	B	B	B	B	B	C	C	A	A	A	A	A	A	A	A	2	
	C	ACG	GFP1'	A	B	A	C	A	B	A	B	A	C	A	B	A	B	A	14	
			GFP2'	B	B	B	B	B	C	C	A	A	A	A	A	A	A	A	2	
			GFP1"	A	B	A	C	A	B	A	B	A	C	A	B	A	B	A	14	
			GFP2"	B	B	B	B	B	C	C	A	A	A	A	A	A	A	A	2	0

Number of **Valines** in GFP	tRNA	Codons																		tRNA changes	change in rare codons GFP" - GFP
17	A	GTT, GTC	GFP1	A	B	A	C	A	A	B	A	C	A	A	B	A	C	A	B	A	14	
	B	GTA	GFP2	B	B	B	B	C	C	C	A	A	A	A	A	A	A	A	A	A	2	
	C	GTG	GFP1'	A	B	A	C	A	B	A	C	A	B	A	C	A	B	A	C	A	16	
			GFP2'	A	A	A	A	A	A	A	A	A	C	C	C	C	B	B	B	B	2	
			GFP1"	A	B	A	C	A	B	A	C	A	B	A	C	A	B	A	C	A	16	
			GFP2"	A	A	A	A	A	A	A	A	C	C	C	C	C	B	B	B	B	2	1

Number of **Phenylalanines** in GFP	only one tRNA	Codons														
13	A	TTT	GFP1	B	B	B	B	B	B	B	B	B	B	B	B	B
	B	TTC	GFP2	B	B	B	B	B	B	B	B	B	B	B	B	B
			GFP1'	B	B	B	B	B	B	B	B	B	B	B	B	B
			GFP2'	B	B	B	B	B	B	B	B	B	B	B	B	B
			GFP1"	A	A	A	A	A	A	A	A	A	A	A	A	A
			GFP2"	A	A	A	A	A	A	A	A	A	A	A	A	A

Number of **Tyrosines** in GFP	only one tRNA	Codons												
11	A	TAT	GFP1	B	B	B	B	B	B	B	B	B	B	B
	B	TAC	GFP2	B	B	B	B	B	B	B	B	B	B	B
			GFP1'	B	B	B	B	B	B	B	B	B	B	B
			GFP2'	B	B	B	B	B	B	B	B	B	B	B
			GFP1"	A	A	A	A	A	A	A	A	A	A	A
			GFP2"	A	A	A	A	A	A	A	A	A	A	A

Number of **Cysteines** in GFP	only one tRNA	Codons			
2	A	TGT	GFP1	B	A
	B	TGC	GFP2	B	A
			GFP1'	B	B
			GFP2'	B	B
			GFP1"	A	A
			GFP2"	A	A

Number of **Histidines** in GFP	only one tRNA	Codons										
10	A	CAT	GFP1	B	B	B	B	A	A	A	A	A	A
	B	CAC	GFP2	B	B	B	B	A	A	A	A	A	A
			GFP1'	B	B	B	B	B	B	B	B	B	B
			GFP2'	B	B	B	B	B	B	B	B	B	B
			GFP1"	A	A	A	A	A	A	A	A	A	A
			GFP2"	A	A	A	A	A	A	A	A	A	A

Number of **Asparagines** in GFP	only one tRNA	Codons													
13	A	AAT	GFP1	B	B	B	B	B	A	A	A	A	A	A	A	A
	B	AAC	GFP2	B	B	B	B	B	A	A	A	A	A	A	A	A
			GFP1'	B	B	B	B	B	B	B	B	B	B	B	B	B
			GFP2'	B	B	B	B	B	B	B	B	B	B	B	B	B
			GFP1"	A	A	A	A	A	A	A	A	A	A	A	A	A
			GFP2"	A	A	A	A	A	A	A	A	A	A	A	A	A

Number of **Aspartic acids** in GFP	only one tRNA	Codons																		
18	A	GAT	GFP1	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A
	B	GAC	GFP2	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A
			GFP1'	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
			GFP2'	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
			GFP1"	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
			GFP2"	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A