Supplementary Information for

Evolutionary conservation of codon optimality reveals hidden signatures of co-translational folding

Sebastian Pechmann & Judith Frydman
Department of Biology and BioX, Stanford University, Stanford, CA-94305

Correspondence to: jfrydman@stanford.edu

Figure 1 Definition and computation of the codon usage. **(a)** Schematic illustration of the definition of the codon usage. The codon usage in this work is derived from the occurrence of each codon in an ORF, weighted by the corresponding transcript abundance. For each codon, the weighted codon counts are summed over all ORFs. The relative scale of the resulting 61 codon usage values is linearly rescaled so that the maximum value is 1. (**b)** The codon usage approximates how often each codon is presented for translation. Experimentally measured abundances of ribosome-associated mRNAs are only available for *S. cerevisiae*. The codon usage computed with abundances of ribosome-associated mRNAs correlates very highly with the codon usage computed with steady-state mRNA expression levels. Because only mRNA expression levels are available for all the other yeasts in our evolutionary analysis, we used mRNA expression levels to compute the codon usage in this study. **(c)** An even more detailed description of how often each codon is translated is given by incorporating both the abundance of only the mRNAs that are associated to ribosomes, as well as the density or occupancy of how many ribosomes are associated to these mRNAs. Ribosome occupancies experimentally measured by ribosome profiling (Science 324, 218-223, 2009) are incorporated into the computation of the codon usage as an additional multiplicative factor. The resulting codon usage vector is almost perfectly correlated to the simplified counterpart that only considers total mRNA expression levels. Thus, the simplified description of only using mRNA expression levels that allows the evolutionary analyses across ten yeasts in in full compliance with experimental ribosome profiling data, underlining the robustness of this approximation. **(d)** As in (c), but with ribosome densities measured by polysome profiling (PNAS, 100, 3889-3894, 2003). This independently measured experimental dataset of ribosome densities further supports the approach and approximation of only using mRNA expression levels as a proxy in the computation of the codon usage, thus allowing an evolutionary analyses.

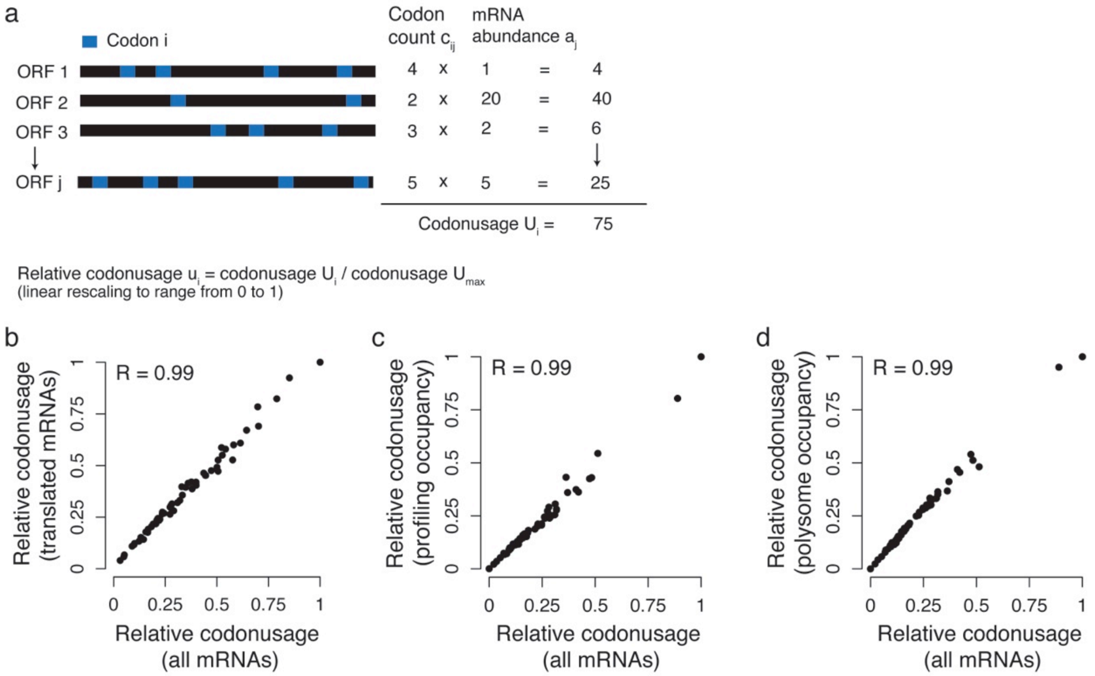

Figure 2 Derivation and validation of the normalized translational efficiency scale. (**a**) Comparison of the classical translational efficiency and codon usage scales for each codon. (**b**) The normalized translational efficiency (nTE) scale is derived from diving the cTE (supply) by the codon usage (demand), linearly rescaled to a maximum value of 1. In nTE, codons are optimal if supply exceeds demand, and nonoptimal otherwise. (**c**) The nTE scale (red line) exhibits a characteristic shallow shape of a plateau-like middle region with an inflexion point and a tail of low translational efficiency that are highly non-random. 10000 random normalized scales are computed by dividing each cTE value by a random codon usage value. Shown are the mean random normalized scale (black line), and 1 and 2 standard deviations (grey shadings). The nTE scale has a distinct tail of low efficiency codons that deviates significantly from the random profile (p < 0.001 for each of the 15 lowest efficiency codons). The middle region is significantly higher than expected (p < 0.01 for codon 15 – 22; p < 0.05 for codon 22 – 44). This can only be explained by the fact that tRNA supply and demand are more closely matched for most codons than expected by chance. The cTE (green line) and codon usage (blue line) scales are shown for reference. Of not, each scale in this plot is sorted in ascending order, illustrating their distributions, not the comparison of specific codons.

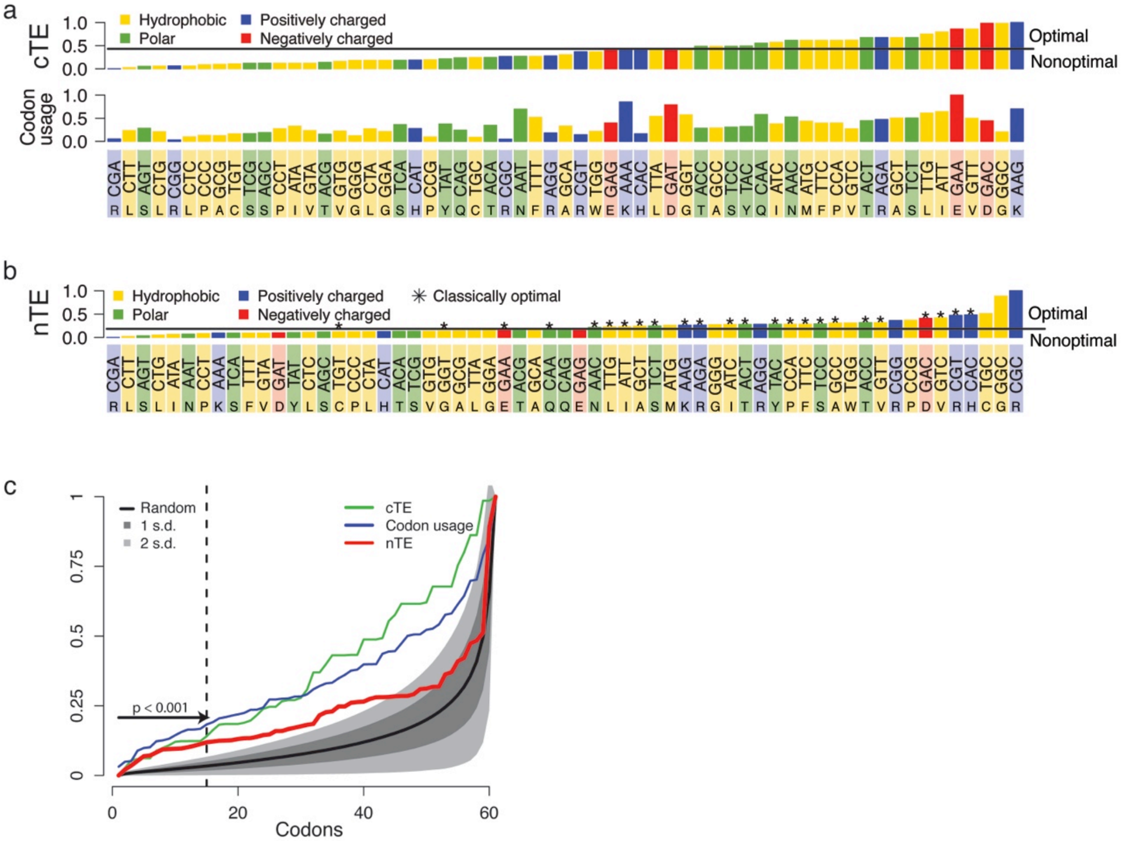

Figure 3 Evolutionary conservation of the short "dip" of low translational efficiency at the beginning of coding sequences. **(a)** Average normalized translational efficiency (red) and classical translational efficiency (black) profiles for *S. cerevisiae* (Scer), *S. paradoxus* (Spar) and *S. bayanus* (Sbay). The solid lines indicate the expected mean and the dashed lines +/- 2 standard deviations. (**b)** The very short region of low normalized translational efficiency is evolutionarily conserved across 10 yeasts including *C. glabrata* Cgla), *D. hansenii* (Dhan), *K. lactis* (Klac), *S. kluyveri* (Sklu), *S. mikatae* (Smik), *S. pombe* (Spom), and *Y. lipolytica* (Ylip) **(c)** The evolutionarily conserved region of low classical translational efficiency, if not normalized, is much longer (*Cell* **141**, 344-54, 2010).

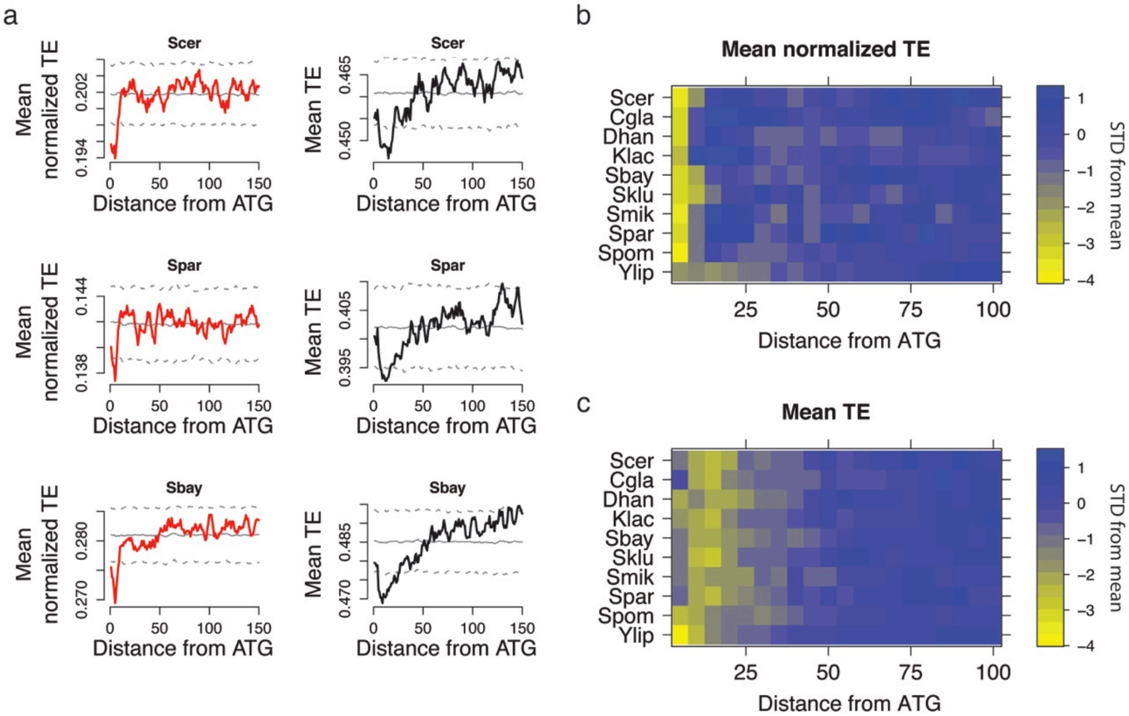

Figure 4 Curation of sequence alignments of codon optimality. (**a**) Distribution of the number of orthologs per alignment for a larger set of 13 yeasts. Most alignments contain 9 or 10 orthologs. (**b**) Distribution of the number of orthologs for 375 alignments with matching *S. cerevisiae* PDB structures in the 10 yeasts with available gene expression data. Choosing a more stringent requirement of at least 7 orthologs per alignment only discards a very low number of alignments. (**c**) Normalized microarray gene expression levels of 10 yeasts allow to curate two groups of high and low expression alignments.

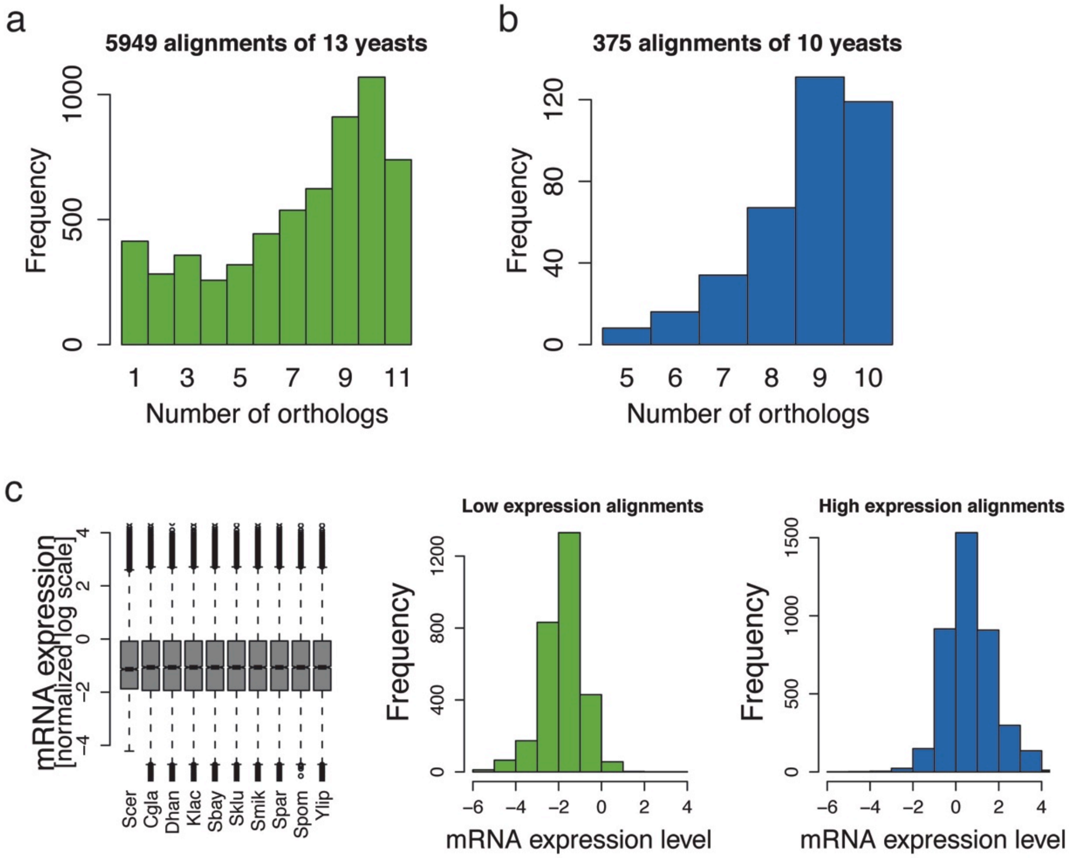

Figure 5 Site-specific evolutionary conservation of codon optimality is independent of amino acid biases. **(a–c)** In this control analysis, we employed a randomization of the simplified sequence alignments of codon optimality that observes the distribution of optimal and nonoptimal codons in the genetic code. Herein, in each alignment randomization step for each codon in the coding sequence alignment a synonymous codon for a given amino acid is randomly chosen, and its species-specific optimality assign to that position. (**a**) The clear majority of analyzed ORFs is under selective pressure for the site-specific evolutionary conservation of both optimal and nonoptimal codons. Highly expressed genes naturally contain a high fraction of optimal codons, thus the randomization just based on the genetic code on average reduces this fraction. As a result, an overall lower fraction of optimal codons (compared to the natural sequences) leads to a lower significance threshold for significantly conserved optimal codons. Accordingly, we find slightly fewer highly expressed ORFs under selective pressured for conserving nonoptimal codons than for conserving optimal codons. The opposite observation can be made for lowly expressed genes that generally have a low fraction of optimal codons. (**b**) The randomization solely based on the genetic code removes any expression bias at codon level in the sequences, thus the significance thresholds for highly and lowly expressed genes distribute similarly. (**c**) The site-specific conservation scores are discrete, thus not always separate at exactly the 5% quantiles. In these cases, we systematically tested at a more stringent significance level. The distributions of the significance levels for the different alignments in our analysis are shown. (**d–e**) As above, but the overall fraction of optimal codons of the true sequence is maintained during the randomization. In this extended randomization, in each randomization step for each codon in the coding sequence alignment a synonymous codon for a given amino acid is randomly chosen, and its species-specific optimality assign to that position. After a full sequence is randomized, random sites are changed to optimal or nonoptimal to restore the overall fraction of optimal codons of the true sequence. (**d**) Under this extended randomization almost all analyzed ORFs are under selective pressure for site-specific evolutionary conservation of both optimal and nonoptimal codons, independent of the levels of expression. (**e**) The distribution of significance thresholds reflects the higher faction of optimal codons in highly expressed genes. (**f**) Tests for site-specific conservation were generally performed at a stringent significance level. (**g**) Distributions of the number of significantly conserved non-optimal (red) and optimal (blue) codons as fraction of the total sequence length. (**h**) The number of alignments that contain more significantly conserved sites than expected by chance does not depend on the 5' coding region. When omitting the first 50 codons from this analysis, the same number of ORFs appears under selective pressure.

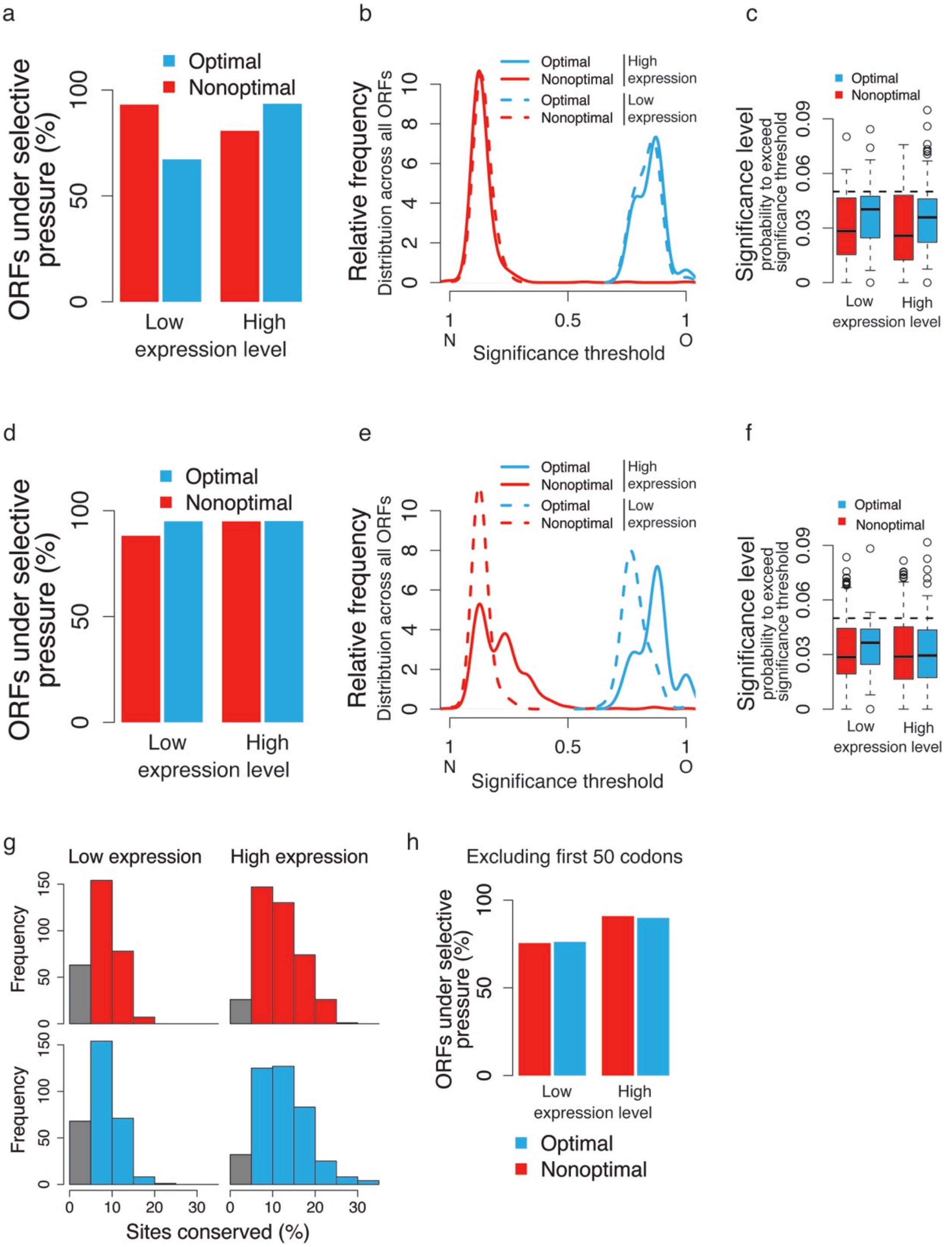
a
Optimal
Nonoptimal
ORFs under selective pressure (%)
0
50
100
Low
High
expression level
b
Optimal
Nonoptimal
High expression
Optimal
Nonoptimal
Low expression
Relative frequency
Distribtuion across all ORFs
0
2
4
6
8
10
1
N
0.5
1
O
Significance threshold
c
Optimal
Nonoptimal
Significance level
probability to exceed significance threshold
0
0.03
0.06
0.09
Low
High
expression level
d
Optimal
Nonoptimal
ORFs under selective pressure (%)
0
50
100
Low
High
expression level
e
Optimal
Nonoptimal
High expression
Optimal
Nonoptimal
Low expression
Relative frequency
Distribtuion across all ORFs
0
2
4
6
8
10
1
N
0.5
1
O
Significance threshold
f
Optimal
Nonoptimal
Significance level
probability to exceed significance threshold
0
0.03
0.06
0.09
Low
High
expression level
g
Low expression
High expression
Frequency
0
50
100
150
Frequency
0
50
100
150
0
10
20
30
0
10
20
30
Sites conserved (%)
h
Excluding first 50 codons
ORFs under selective pressure (%)
0
50
100
Low
High
expression level
Optimal
Nonoptimal

Figure 6 The link between conserved codon optimality and protein secondary structures is independent of amino acid biases. (**a,b**) To test for independence of amino acid biases, each position was assigned the probability of the given amino acid being encoded by an optimal codon based on the distribution of optimal and nonoptimal codons in the genetic code. For example, *Ala* is encoded by 2 optimal and 2 nonoptimal codons in *S. cerevisiae*, thus each *Ala* contributes 0.5 sites of optimal codons. We first tested for a general inherent association between codon optimality and secondary structure due to amino acid biases. For example, in the case of α-helices, we tested for association between the numbers of optimal sites and helices. For both highly and lowly expressed genes, we find no significant associations between codon optimality of randomized sequence and helices or coil regions. There is a very weak but significant enrichment of optimal codons in sheets. Sites of significantly conserved codon optimality also coincide with higher conservation of the encoded amino acids. To verify that highly conserved amino acids do not bias a link between codon optimality and secondary structure, we tested for association only for sites that are significantly conserved in our evolutionary analysis. No statistically significant associations could be found for helices and coil regions. A weak enrichment of optimal codons is sheets could be detected. For comparison, conserved codon optimality in the true observed sequences associates significantly with protein secondary structures in both highly and lowly expressed genes (see Figures 4 and 5 in the main text). Importantly, the significant associations between conserved codon optimality and sheets deviate dramatically from the trends found for randomized sequences, and the enrichment of conserved optimal codons in helices and coil regions, structural elements that can fold co-translationally, cannot be found in randomized sequences. Importantly, all amino acids that are not encoded by equal numbers of optimal and nonoptimal codons in nTE have equal secondary structure propensities for helix and sheet formation (Chou & Fasman, *Biochemistry* **13**, 222–245, 1974). (**c**) Associations between conserved codon optimality and sequence hydrophobicity. Significantly conserved optimal and nonoptimal codons as determined with the cTE scale are compared to the corresponding calculations with the nTE scale for highly expressed and lowly expressed genes, as well as the sequences of the available PDB structures. (**d**) Associations between protein secondary structure from experimental PDB structures and conserved optimal and nonoptimal codons that appear in clusters for the normalized translational efficiency scale. Associations are even stronger and more significant than for predicted secondary structures and significantly conserved sites that do not necessarily appear in clusters. (**e**) The fraction of alignments with available PDB structures that appear under selective pressure agree with the corresponding distributions for predicted secondary structures. (**f–h**) Evolutionary conservation of RNA secondary structure. (**f**) We repeated our complete analysis for the evolutionary conservation of predicted RNA secondary structure in 10 yeasts. Homogeneous distributions of significance thresholds for the site-specific conservation of unpaired and paired nucleotides. (**g**) Fraction of alignments that appear under selective pressure for site-specific conservation of RNA secondary structure. Both unpaired and paired nucleotides seem to be only in about half of the alignments under general site-specific selection. Most of this stems from a clear preference for unpaired nucleotides at the start site of messages, which facilitates translation initiation. (**h**) In comparison, codon optimality is much more strongly conserved in a site-specific manner than RNA secondary structure.

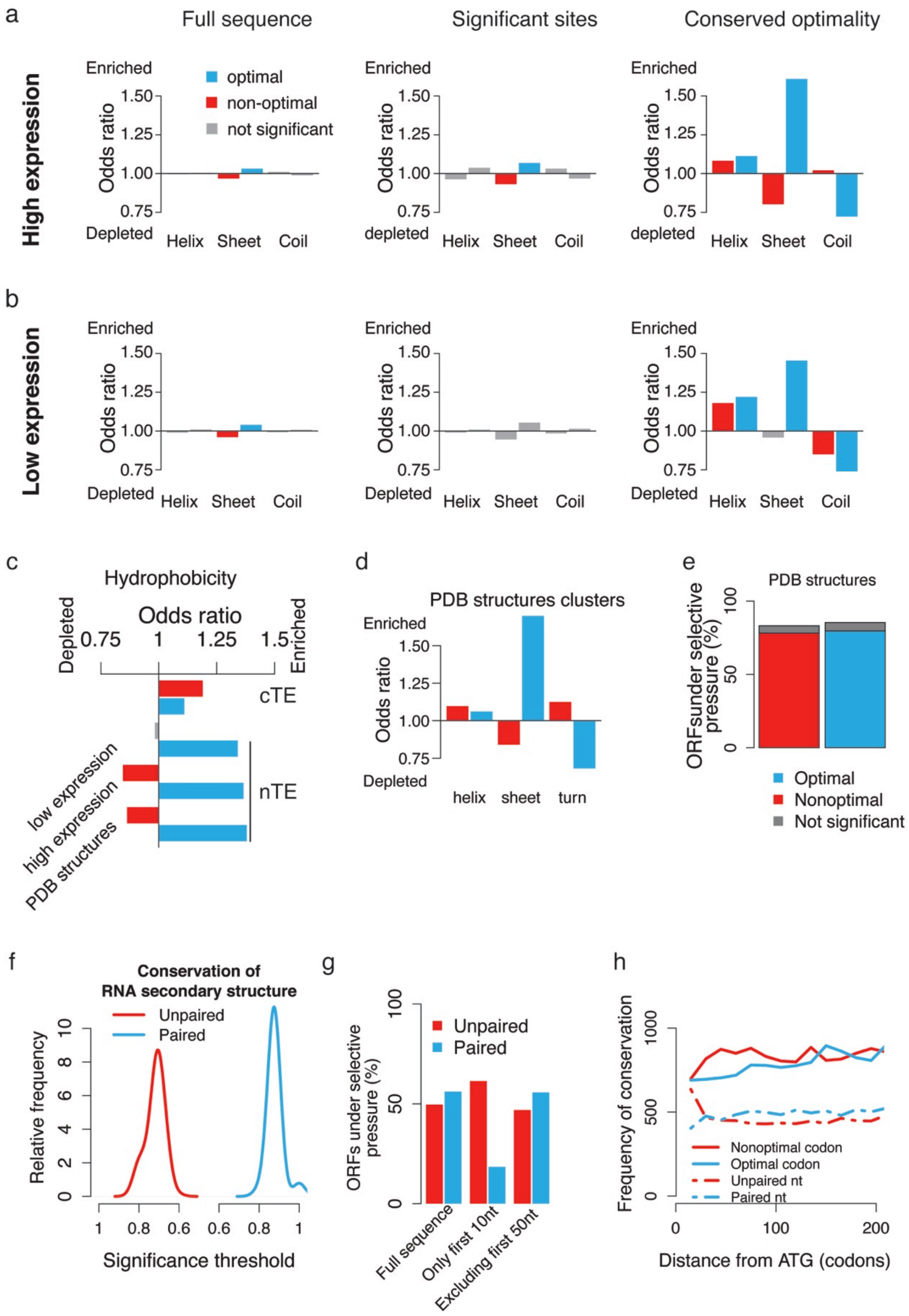
a
Full sequence
Significant sites
Conserved optimality
High expression
Enriched
Depleted
depleted
Odds ratio
1.50
1.25
1.00
0.75
optimal
non-optimal
not significant
Helix
Sheet
Coil
b
Low expression
c
Hydrophobicity
Odds ratio
Depleted
Enriched
0.75
1
1.25
1.5
cTE
nTE
low expression
high expression
PDB structures
d
PDB structures clusters
helix
sheet
turn
e
PDB structures
ORFsunder selective pressure (%)
0
50
100
Optimal
Nonoptimal
Not significant
f
Conservation of RNA secondary structure
Unpaired
Paired
Relative frequency
0
2
4
6
8
10
1
0.8
0.6
Significance threshold
g
ORFs under selective pressure (%)
Full sequence
Only first 10nt
Excluding first 50nt
h
Frequency of conservation
0
500
1000
Nonoptimal codon
Optimal codon
Unpaired nt
Paired nt
0
100
200
Distance from ATG (codons)

Table 1 Tests of association between conserved optimal and conserved nonoptimal codons, and protein sequence and structural features. The Cochran–Mantel–Haenszel as implemented in the statistics package R was used to estimate the strength of the enrichment or depletion (odds ratio) of conserved codon optimality, and its statistical significance.

	high expression (n=404)		low expression (n=302)	
Feature	odds ratio	p-value	odds ratio	p-value
	conserved optimal codons			
helix	1.113	$3.51 * 10^{-11}$	1.18	$3.79 * 10^{-14}$
sheet	1.609	$1.12 * 10^{-116}$	1.454	$6.61 * 10^{-43}$
coil	0.723	$1.47 * 10^{-101}$	0.743	$1.98 * 10^{-47}$
hydropohobicity	1.366	$2.82 * 10^{-26}$	1.341	$1.98 * 10^{-19}$
disorder	0.796	$2.54 * 10^{-26}$	0.796	$6.59 * 10^{-16}$
	conserved nonoptimal codons			
helix	1.083	$9.15 * 10^{-7}$	1.22	$4.81 * 10^{-21}$
sheet	0.802	$1.73 * 10^{-19}$	0.958	0.21
coil	1.021	0.164	0.85	$8.17 * 10^{-16}$
hydropohobicity	0.845	$3.43 * 10^{-9}$	0.983	0.648
disorder	0.781	$2.97 * 10^{-29}$	0.713	$2.27 * 10^{-33}$

Table 2 Tests of association between conserved optimal and conserved nonoptimal codons in clusters, and protein sequence and structural features. As in Table 1, but only significantly conserved optimal codons that appear within 4 residues of other conserved optimal codons, and nonoptimal codons in clusters respectively, are considered.

	high expression (n=404)		low expression (n=302)	
Feature	odds ratio	p-value	odds ratio	p-value
	conserved optimal codons in clusters			
helix	1.1	$8.64 * 10^{-9}$	1.21	$2.89 * 10^{-13}$
sheet	1.845	$1.75 * 10^{-188}$	1.452	$1.01 * 10^{-24}$
coil	0.68	$4.87 * 10^{-136}$	0.728	$3.54 * 10^{-39}$
hydropohobicity	1.57	$1.84 * 10^{-76}$	1.52	$7.24 * 10^{-309}$
disorder	0.782	$1.04 * 10^{-29}$	0.873	$4.22 * 10^{-5}$
	conserved nonoptimal codons in clusters			
helix	1.107	$2.69 * 10^{-9}$	1.31	$2.1 * 10^{-27}$
sheet	0.738	$7.57 * 10^{-32}$	0.98	0.65
coil	1.038	0.02	0.792	$4.47 * 10^{-23}$
hydropohobicity	0.833	$1.36 * 10^{-9}$	1.084	0.047
disorder	0.767	$3.42 * 10^{-30}$	0.64	$1.05 * 10^{-40}$

Table 3 Tests of association between conserved optimal and conserved nonoptimal codons and protein sequence and structural features for all *S. cerevisiae* proteins with available PDB structures. Solvent accessible surface area (ASA) is computed with the DSSP program.

PDB structures (n=357)	conserved optimal codons		conserved nonoptimal codons	
Feature	odds ratio	p-value	odds ratio	p-value
helix	1.092	$2.16 * 10^{-4}$	1.058	0.016
sheet	1.499	$2.08 * 10^{-53}$	0.895	$9.961 * 10^{-5}$
turn	0.702	$4.94 * 10^{-33}$	1.131	$2.84 * 10^{-4}$
hydrophobicity	1.381	$7.35 * 10^{-26}$	0.863	$2.62 * 10^{-5}$
buried (ASA < 50 Å^2)	1.377	$2.80 * 10^{-45}$	0.960	0.056
	Significantly conserved codons in clusters			
helix	1.061	0.038	1.097	$7.25 * 10^{-4}$
sheet	1.698	$2.15 * 10^{-69}$	0.843	$5.18 * 10^{-7}$
turn	0.682	$1.97 * 10^{-27}$	1.125	$1.67 * 10^{-4}$
hydrophobicity	1.586	$4.35 * 10^{-40}$	0.824	$5.16 * 10^{-6}$
buried (ASA < 50 Å^2)	1.395	$5.30 * 10^{-35}$	0.977	0.36

Table 4 Tests of association between conserved optimal and conserved nonoptimal codons and protein sequence and structural features for all *S. cerevisiae* proteins with available PDB structures and the classical definition of codon optimality.

PDB structures (n=357)	conserved optimal codons		conserved nonoptimal codons	
feature	odds ratio	p-value	odds ratio	p-value
helix	1.074	0.017	0.997	0.914
sheet	1.36	$8.09 * 10^{-20}$	0.989	0.728
turn	0.797	$4.09 * 10^{-10}$	0.955	0.134
hydrophobicity	1.112	0.0087	1.188	$6.20 * 10^{-7}$
buried (ASA < 50 Å^{2})	1.22	$7.00 * 10^{-14}$	1.212	$2.62 * 10^{-15}$
	Significantly conserved codons in clusters			
helix	1.012	0.742	0.969	0.376
sheet	1.572	$4.59 * 10^{-27}$	0.930	0.088
turn	0.789	$3.69 * 10^{-7}$	0.940	0.134
hydrophobicity	1.187	$7.79 * 10^{-4}$	1.548	$1.43 * 10^{-26}$
buried (ASA < 50 Å^{2})	1.299	$2.20 * 10^{-13}$	1.247	$1.30 * 10^{-11}$

Table 5 The link between conserved codon optimality and protein secondary structures is independent of amino acid biases. To test for independence of amino acid biases, we tested for association between optimal and nonoptimal codons and secondary structure elements for randomized sequences. For each amino acid, a random synonymous codon was chosen, thus inherent amino acid biases would persist this randomization.

	high expression (n=404)		low expression (n=302)	
Feature	odds ratio	p-value	odds ratio	p-value
	optimal vs. nonoptimal codons fully randomized sequence			
helix	0.996	0.71	0.990	0.40
sheet	**1.032**	**0.03**	**1.040**	**0.03**
coil	0.989	0.24	0.993	0.51
	optimal vs. nonoptimal codons positions in the fully randomized sequences that show significant conservation of codon optimality			
helix	1.037	0.09	1.009	0.78
sheet	**1.069**	**0.03**	1.054	0.24
coil	0.968	0.12	1.015	0.60

Table 6 Positional preference of codon optimality in a-helices in *S. cerevisiae*. All α-helices longer than 6 residues (n = 2829) were extracted from all available *S. cerevisiae* PDB structures and aligned at their beginning. For each position, the fraction of nonoptimal codons was tested with Fisher's exact test for independence from all other positions, and the p-values were corrected for multiple-testing with the Benjamini & Hochberg method. A distinct pattern of preferred optimal and nonoptimal codons emerges between positions -1 and 4, thus stretching just across the first helix turn.

Position	**odds ratio**	**p-value**
-3	0.949	0.22
-2	0.937	0.16
-1	**1.147**	**0.02**
1	**0.916**	**0.05**
2	**1.281**	**5.54 * 10^{-9}**
3	**1.111**	**0.02**
4	**0.796**	**5.81 * 10^{-8}**
5	0.943	0.19
6	1.002	0.97

Table 7 Codon optimality in nTE. Optimal (O) and nonoptimal (N) codons in the ten closely related yeast species analyzed in this study as defined in nTE.

Codon	Cgla	Dhan	Klac	Sbay	Sklu	Smik	Spar	Spom	Ylip	Scer
TTT	N	N	N	N	N	N	N	N	N	N
TTC	O	O	N	O	O	O	O	O	O	O
TTA	N	N	N	N	N	N	N	N	O	N
TTG	O	N	O	O	O	N	N	N	N	O
TCT	O	O	O	O	O	O	O	O	O	O
TCC	O	O	O	O	O	O	O	O	O	O
TCA	N	N	N	N	N	N	N	N	N	N
TCG	N	N	N	N	N	N	N	O	N	N
TAT	N	N	N	N	N	N	N	N	O	N
TAC	O	O	O	O	O	O	O	O	N	O
TGT	N	N	N	N	N	N	N	N	O	N
TGC	O	O	O	O	O	O	O	O	O	O
TGG	O	O	O	O	O	O	O	O	O	O
CTT	N	N	N	N	N	N	N	N	O	N
CTC	N	O	N	N	N	N	N	O	N	N
CTA	N	N	N	N	N	N	N	N	N	N
CTG	N	O	N	N	N	N	N	N	N	N
CCT	N	N	N	N	N	N	N	O	O	N
CCC	N	N	N	N	N	N	N	O	N	N
CCA	O	O	O	O	O	O	O	N	N	O
CCG	O	O	O	O	O	O	O	O	N	O
CAT	N	N	N	N	N	N	N	N	N	N
CAC	O	O	O	O	O	O	O	O	O	O
CAA	N	N	N	N	O	N	N	N	N	N
CAG	N	O	O	O	N	N	N	O	N	N
CGT	O	O	O	O	O	O	O	O	N	O
CGC	O	O	O	O	O	O	O	O	N	O
CGA	N	N	N	N	N	N	N	N	O	N
CGG	O	O	O	O	O	O	O	O	O	O
ATT	O	N	O	O	O	O	O	O	O	O
ATC	O	O	O	O	O	O	O	O	O	O
ATA	N	N	N	N	N	N	N	N	O	N
ATG	O	O	O	O	O	O	O	O	O	O
ACT	O	O	O	O	O	O	O	O	O	O
ACC	O	O	O	O	O	O	O	O	N	O
ACA	N	N	N	N	N	N	N	N	N	N
ACG	O	N	O	N	N	N	N	O	N	N
AAT	N	N	N	N	N	N	N	N	O	N
AAC	O	O	O	N	N	O	O	O	N	O
AAA	N	N	N	N	N	N	N	N	N	N
AAG	O	O	O	O	O	O	O	O	O	O
AGT	N	N	N	N	N	N	N	N	N	N
AGC	O	O	O	O	O	N	O	O	O	N
AGA	O	O	O	O	O	O	O	N	N	O
AGG	O	O	O	O	O	O	O	O	O	O

GTT	O	O	O	O	O	O	O	O	O	O
GTC	O	O	O	O	O	O	O	O	O	O
GTA	N	N	N	N	N	N	N	N	O	N
GTG	N	N	N	N	N	N	N	N	N	N
GCT	O	O	O	O	O	O	O	O	O	O
GCC	O	O	O	O	O	O	O	O	N	O
GCA	O	N	N	N	N	N	N	N	N	N
GCG	N	O	N	N	N	N	N	O	N	N
GAT	N	N	N	N	N	N	N	N	N	N
GAC	O	O	O	O	O	O	O	O	O	O
GAA	N	N	N	O	N	N	N	N	N	N
GAG	N	N	O	O	N	N	N	O	N	O
GGT	N	N	N	N	N	N	N	N	O	N
GGC	O	O	O	O	O	O	O	O	O	O
GGA	N	N	N	N	N	N	N	N	N	N
GGG	O	O	O	O	O	O	O	O	O	O

Table 8 Codon optimality in cTE. Optimal (O) and nonoptimal (N) codons in the ten closely related yeast species analyzed in this study as defined in cTE.

codon	Cgla	Dhan	Klac	Sbay	Sklu	Smik	Spar	Spom	Ylip	Scer
TTT	N	N	N	N	N	N	N	N	N	N
TTC	O	O	O	O	O	O	O	O	O	O
TTA	N	O	N	N	N	N	N	N	N	N
TTG	O	O	O	O	O	O	O	O	N	O
TCT	O	O	O	O	O	O	O	O	O	O
TCC	O	O	O	O	O	O	O	O	O	O
TCA	N	N	N	N	N	N	N	N	N	N
TCG	N	N	N	N	N	N	N	N	N	N
TAT	N	N	N	N	N	N	N	N	N	N
TAC	O	O	O	O	O	O	O	O	O	O
TGT	O	O	O	O	O	O	O	N	N	O
TGC	N	N	N	N	N	N	N	O	O	N
TGG	N	N	N	N	N	N	N	N	N	N
CTT	N	N	N	N	N	N	N	O	O	N
CTC	N	N	N	N	N	N	N	O	O	N
CTA	O	N	N	N	N	N	N	N	N	N
CTG	N	N	N	N	N	N	N	N	N	N
CCT	N	N	N	N	N	N	N	O	N	N
CCC	N	N	N	N	N	N	N	O	O	N
CCA	O	O	O	O	O	O	O	N	N	O
CCG	N	N	N	N	N	N	N	N	N	N
CAT	N	N	N	N	N	N	N	N	N	N
CAC	O	O	O	O	O	O	O	O	O	O
CAA	O	O	O	O	O	O	O	O	N	O
CAG	N	N	N	N	N	N	N	N	O	N
CGT	O	O	O	O	O	O	O	O	N	O
CGC	N	N	N	N	N	N	N	O	N	N
CGA	N	N	N	N	N	N	N	N	O	N
CGG	N	N	N	N	N	N	N	N	N	N
ATT	O	O	O	O	O	O	O	O	N	O
ATC	O	O	O	O	O	O	O	O	O	O
ATA	N	N	N	N	N	N	N	N	N	N
ATG	N	N	N	N	N	N	N	N	N	N
ACT	O	O	O	O	O	O	O	O	N	O
ACC	O	O	O	O	O	O	O	O	O	O
ACA	N	N	N	N	N	N	N	N	N	N
ACG	N	N	N	N	N	N	N	N	N	N
AAT	N	N	N	N	N	N	N	N	N	N
AAC	O	O	O	O	O	O	O	O	O	O
AAA	N	N	N	N	N	N	N	N	N	N
AAG	O	O	O	O	O	O	O	O	O	O
AGT	N	N	N	N	N	N	N	N	N	N
AGC	N	N	N	N	N	N	N	O	N	N
AGA	O	O	O	O	O	O	O	N	N	O

AGG	N	N	N	N	N	N	N	N	N	N
GTT	O	O	O	O	O	O	O	O	O	O
GTC	O	O	O	O	O	O	O	O	O	O
GTA	N	N	N	N	N	N	N	N	N	N
GTG	N	N	N	N	N	N	N	N	N	N
GCT	O	O	O	O	O	O	O	O	O	O
GCC	O	O	O	O	O	O	O	O	O	O
GCA	N	N	N	N	N	N	N	N	N	N
GCG	N	N	N	N	N	N	N	N	N	N
GAT	N	N	N	N	N	N	N	N	N	N
GAC	O	O	O	O	O	O	O	O	O	O
GAA	O	O	O	O	O	O	O	N	N	O
GAG	N	N	N	N	N	N	N	O	O	N
GGT	O	O	O	O	O	O	O	O	O	O
GGC	N	N	N	N	N	N	N	N	O	N
GGA	N	N	N	N	N	N	N	N	N	N
GGG	N	N	N	N	N	N	N	N	N	N