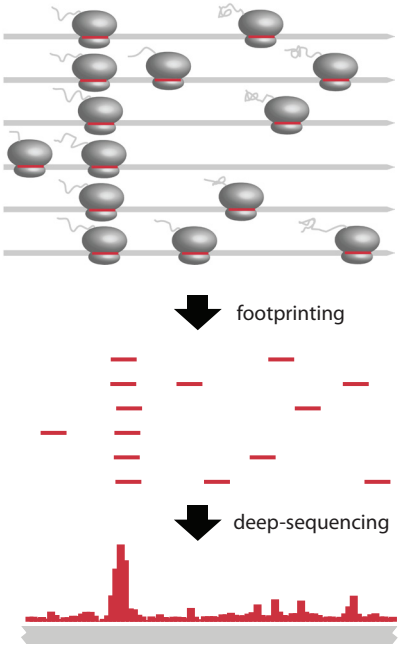
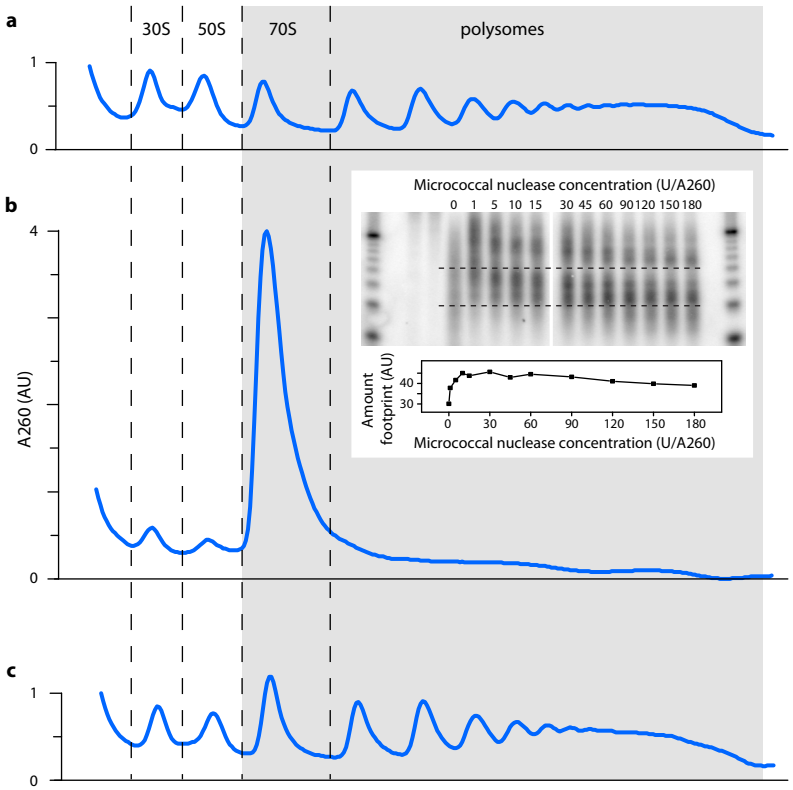
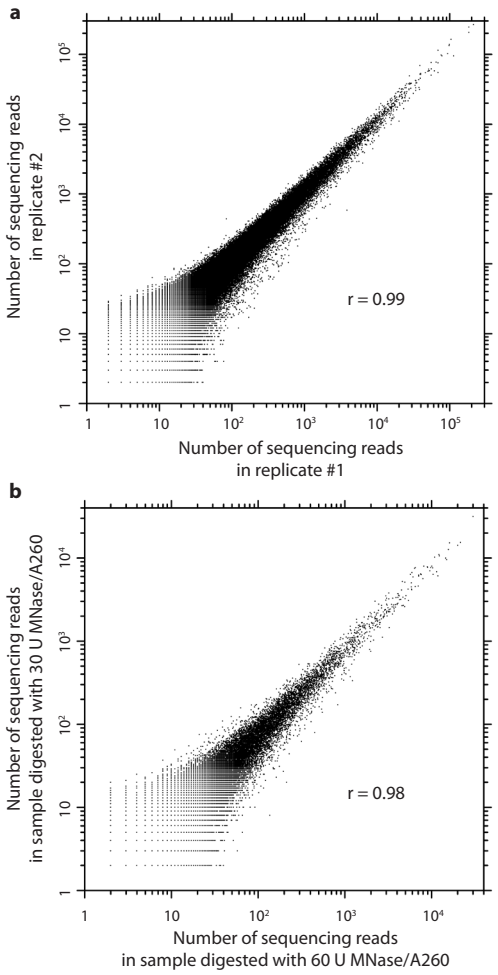
doi:10.1038/nature10965



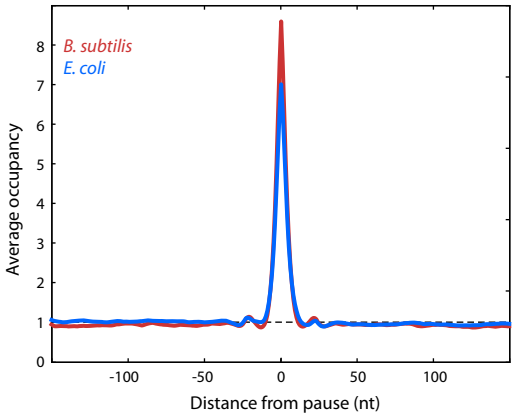
Supplementary Figure 1. Schematic of ribosome profiling experiment for quantification of ribosome occupancy along mRNA. The protocol for bacterial ribosome profiling with flash freezing was described in detail by Oh *et al*.[5]. Polysome-containing cell lysate was treated with micrococcal nuclease to generate ribosome-protected mRNA fragments. The mRNA fragments were converted into a sequenceable DNA library. Regions of mRNA that have higher ribosome occupancy give rise to more sequenced fragments.



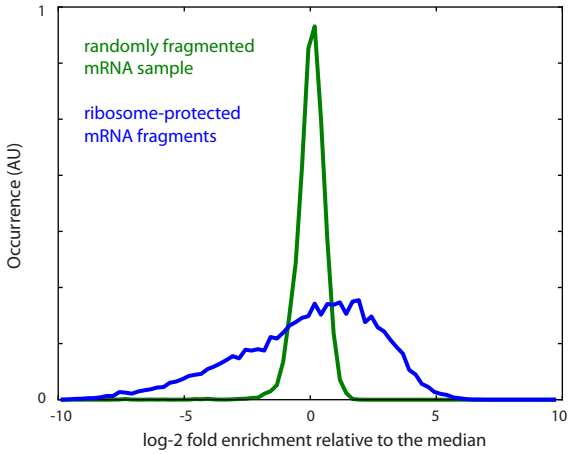
Supplementary Figure 2. Polysome profiles at different stages of ribosome footprinting for *E. coli.* a, Polysome profile of flash frozen and pulverized cell lysate. 77% of total RNA was in assembled ribosomes (shaded area), of which 87% was in the polysome fraction. **b,** Polysome profile after treatment with micrococcal nuclease at 25°C for one hour. The amount of RNA in the assembled ribosome (shaded area) was the same as that of the undigested lysate, indicating that assembled ribosomes stayed intact during footprinting. Consistent with this observation, nuclease protection assay (inset) showed a constant level of ribosome-protected fragment over a range of micrococcal nuclease concentrations. The amount of footprint (between dashed lines) as a function of nuclease concentration was plotted. In the ribosome profiling experiments we used 60 U of nuclease per 1 A_{260} unit of RNA. Nuclease protection assay was previously described by Ingolia et al.[3], using the mirVana miRNA detection kit. [α^{32}P] UTP labeled anti sense probe (*gapA* gene of *E. coli*) was generated using the MAXIscript kit with T7 RNA polymerase. **c,** Polysome profile after incubation at 25°C for one hour without micrococcal nuclease. The fraction of RNA in the assembled ribosome (shaded area) was again the same as that of the undigested lysate. The ratio of 70S particles to polysome fractions increased during the 25°C incubation, which is likely due to breakage of mRNA in between ribosomes.



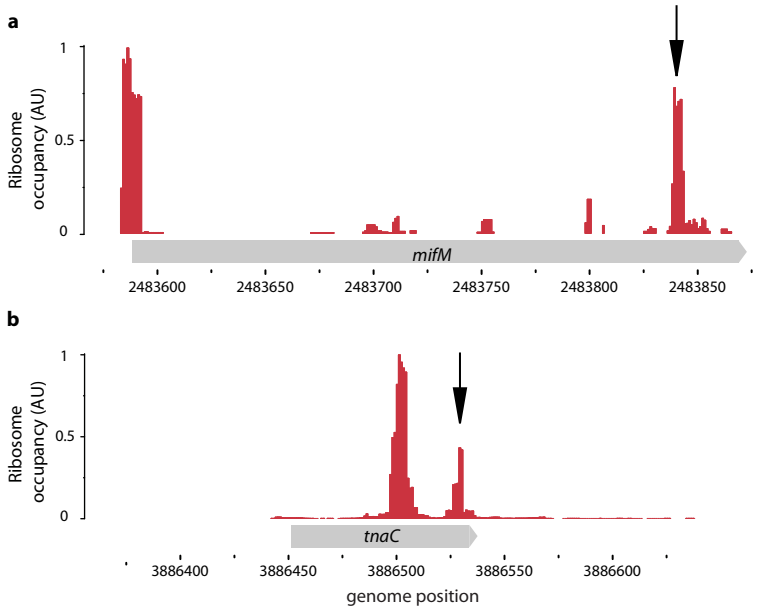
Supplementary Figure 3. Reproducibility of bacterial ribosome profiling. a, Reproducibility among biological replicates. Each dot corresponds to the number of sequencing reads mapped to a particular position on mRNA in ribosome profiling experiments from two separate cultures of *E. coli*. The Pearson correlation coefficient is 0.99. **b,** Effect of micrococcal nuclease (MNase) concentration. Lysate of *B. subtilis* was treated with either 60 U or 30 U of MNase per 1 A_{260} unit of RNA, and the number of ribosome protected fragments at each position on mRNA were compared. The Pearson correlation coefficient of 0.98, confirming that nuclease digestion introduces negligible bias at the working concentration of MNase. In addition, the correlation between Shine-Dalgarno-like sequences and pausing was unaffected by the 2-fold change in the amount of MNase.



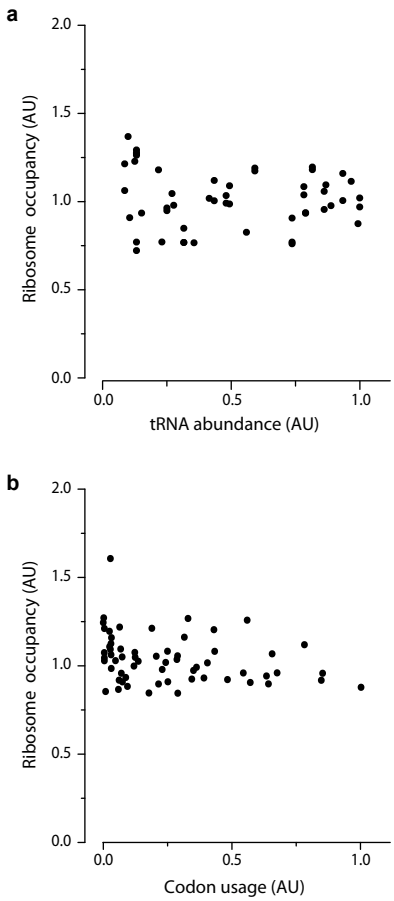
Supplementary Figure 4. Average ribosome density before and after translational pausing sites. Pauses with ribosome occupancy at 5-fold greater than the mean were aligned at position 0. The ribosome occupancy surrounding a pause was normalized by the mean occupancy of the message, and averaged over all pausing sites. There is no loss of ribosome density immediately before and after pauses, indicating that translation within coding sequences is a continuous process with negligible internal initiation and early termination at the pausing sites. Furthermore, this observation also argues that there is negligible ribosome movement after cells were flash frozen, which would lead to depletion of ribosome density after pausing sites.



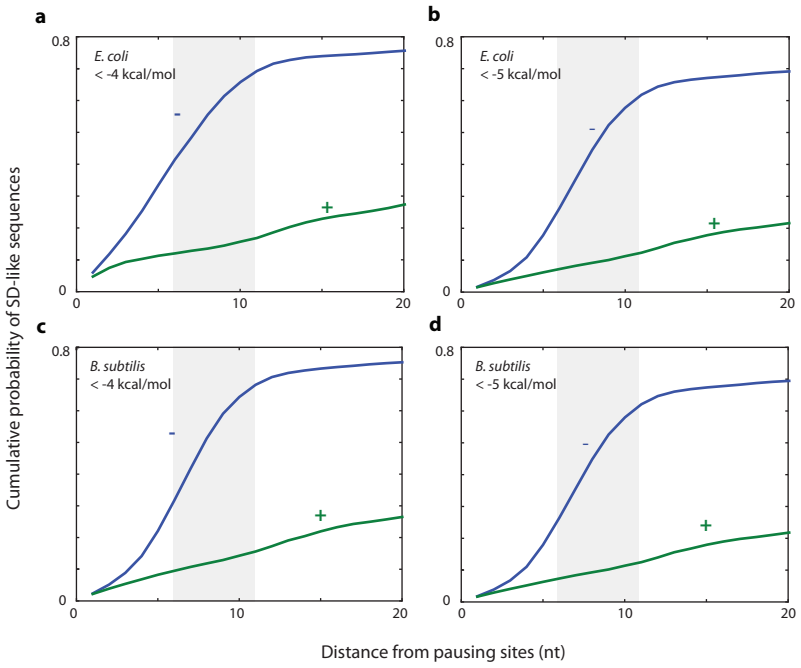
Supplementary Figure 5. Variation of ribosome occupancy and randomly fragmented mRNA sample for *E. coli*. Ribosome footprints (blue) and randomly fragmented mRNA (green) were converted to sequenceable DNA libraries using the same protocol. The frequency of sequencing reads at each codon on each message was normalized to the median frequency of the message. Histograms of $\log_2$ enrichment relative to the median were plotted for codons in the genes that have at least 10 sequencing reads per codon on average. Ribosome occupancy exhibited greater variations than that introduced during the conversion of RNA fragments into a sequenceable DNA library.



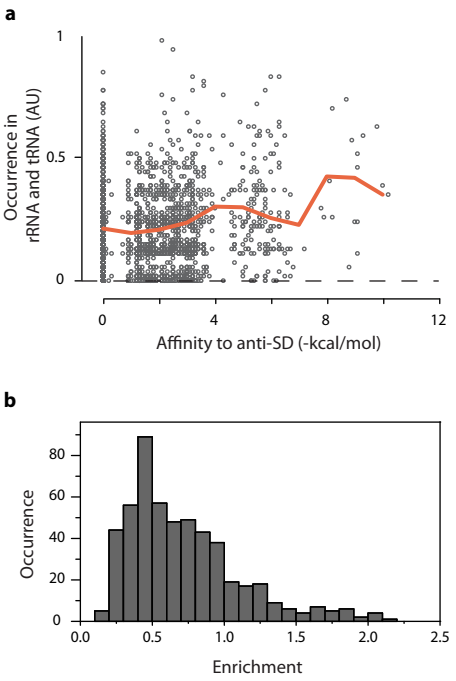
Supplementary Figure 6. Ribosome occupancy profile of genes with translational stalling sites. a, The *mifM* gene in *B. subtilis*. **b,** The *tnaC* gene in *E. coli*. The arrows point at the position of known stalling sites. In *tnaC* we observed a second ribosome queuing ~30 nt upstream the known stalling site. The presence of this second ribosome immediately before the stalling site would be difficult to detect using conventional assays based on primer extension. It is plausible that a trailing ribosome is queuing behind the ribosome that is stalled downstream.



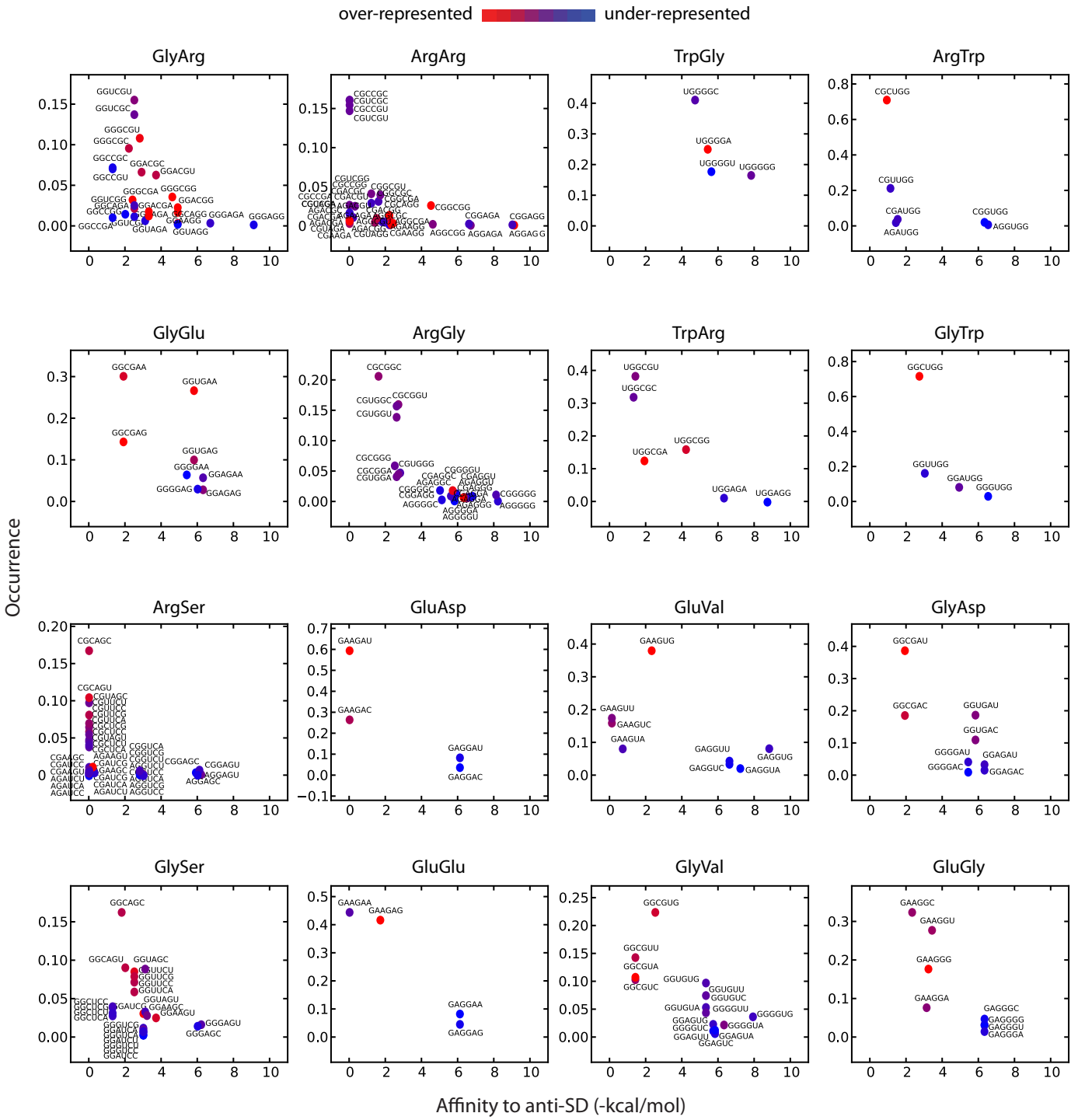
Supplementary Figure 7. Average ribosome occupancy of codons. a, Ribosome occupancy relative to the corresponding tRNA abundance in *B. subtilis*. Similar to Fig. 1c and d, average ribosome occupancy of each codon relative to their respective tRNA abundance is plotted. Codons with undetermined tRNA abundance were not included. The codon-specific ribosome occupancy was uncorrelated with the tRNA abundance. **b,** Ribosome occupancy relative to codon usage in *E. coli*. The codon usage was calculated from a group of 321 highly expression genes that have at least 500 sequencing reads per codon on average in the dataset. The average ribosome occupancy was calculated from 2,255 genes with at least 10 sequencing reads per codon.



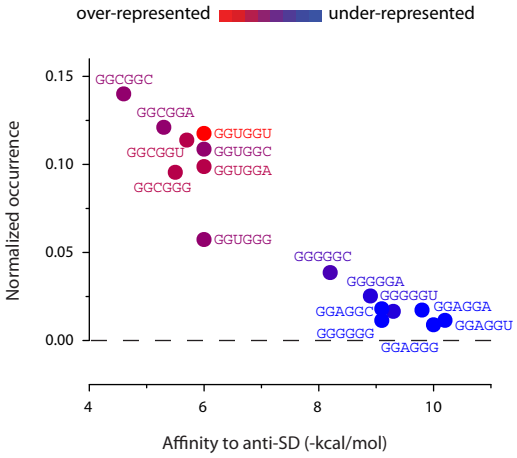
Supplementary Figure 8. Fraction of pauses associated with SD-like sequences. Cumulative probability of having SD-like sequences either upstream (-) or downstream (+) from pausing sites was plotted against the distance from the pausing sites in *E. coli* (**a** and **b**) and in *B. subtilis* (**c** and **d**). The cumulative probability is the probability of having at least one SD-like sequence within a certain distance from the pausing site. SD-like sequences were defined as hexamer sequences with affinity to aSD < -4 kcal/mol (**a** and **c**) or < -5 kcal/mol (**b** and **d**). Pausing sites with ribosome occupancy greater than 10-fold of the mean (~ 2 pauses/gene) were included in this analysis. ~70% of the pauses were associated with SD-like sequences upstream (shaded).



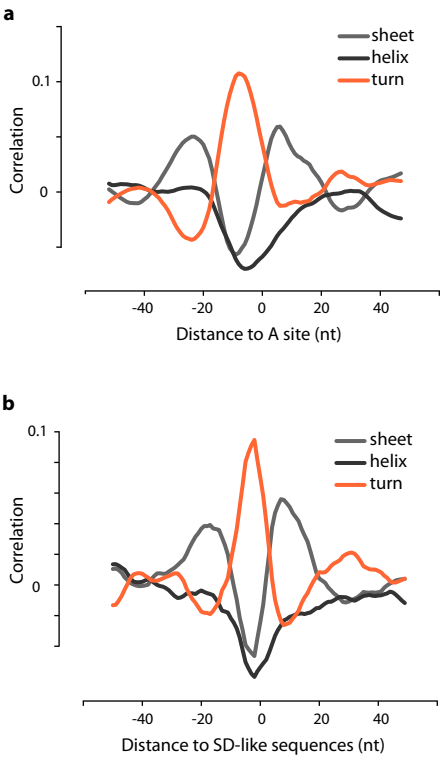
Supplementary Figure 9. Occurrence of SD-like sequences. a, The occurrence of hexamer sequences in rRNA and tRNA relative to the affinity to anti-SD in *E. coli*. The orange line shows the average occurrence within a bin size of 0.5 kcal/mol. Unlike hexamers in protein coding sequences, strong SD-like hexamers were not avoided. **b,** Histogram of enrichment of internal SD-like sequences in the mRNA of 533 bacterial species in the AMPHORA[35] database. The enrichment level of each species was calculated based on its GC content. SD-like hexamers (with predicted hybridization energy < -7 kcal/mol) were avoided in the majority of bacterial species. The avoidance of SD-like sequences is one of many forces, including GC bias and mutational bias, that determine the genome composition.



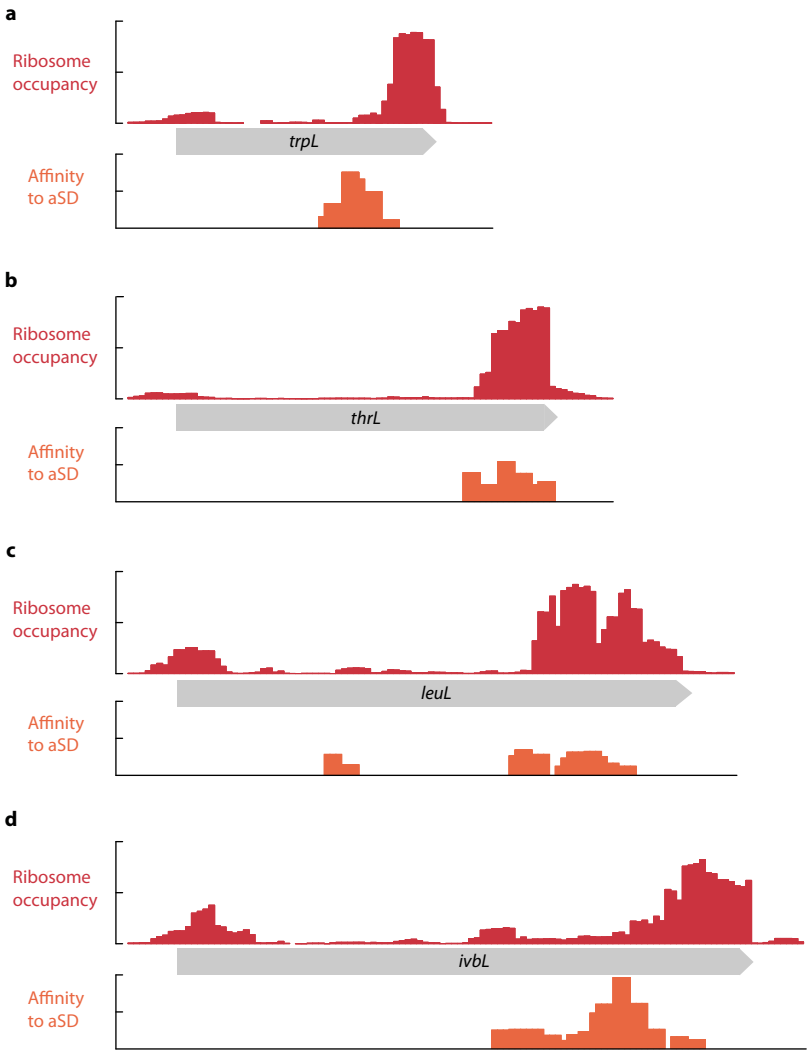
Supplementary Figure 10. Disenrichment of codon pairs that resemble SD sequences in *E. coli*. We calculated the normalized occurrence of codon pairs (y-axes) and the enrichment relative to single codon usage (color coded) for 16 pairs of amino acids that can be coded with SD sites (< -6 kcal/mol). The occurrence was normalized within each group of codon pairs encoding the same pairs of amino acids. Similar to Gly-Gly pairs, strong SD-like codon pairs appear less often than what is expected from the single codon usage.



Supplementary Figure 11. Occurrence of GGNGGN sequences that are not aligned to Gly-Gly pairs in *E. coli* protein coding sequences. The occurrence of GGNGGN that does not encode two glycine codons were plotted against the affinity to the anti-SD site. The colour coding represents the enrichment in occurrence after correcting for the usage of single trinucleotide sequence. The fact that the same trend exists regardless of reading frame information supports the notion that the preference of codon pairs stems from properties of the sequence, rather than properties of the codon or the tRNA.



Supplementary Figure 12. Correspondence of protein structure and ribosome pausing. a, Correlation between protein secondary structures and ribosome occupancy profiles. Translational pauses were over-represented in places where the newly synthesized polypeptides correspond to turns in a protein. **b,** Correlation between protein secondary structures and SD-like sequences. SD-like sequences are also over-represented in regions encoding protein turns. The pause sites, including at most Gly-Gly residues, are over-represented in protein turns and unstructured regions. Therefore pausing could potentially facilitate independent folding of adjacent structural motifs. An important caveat is that pausing at SD sites occurs when the amino acid residues translated from SD sites, such as Gly-Gly, are still within the ribosome exit tunnel. Whether the structured region would be outside the exit tunnel therefore needs to be determined on a case-by-case level.



Supplementary Figure 13. Ribosome pausing near the end of leader peptide sequences of amino acid biosynthesis operons. Ribosome density is low at the beginning of leader sequences and high near the end. Slow translation near the stop codon may provide additional protection for the structural mRNA elements to promote transcription termination.