The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria

Gene-Wei Li1, Eugene Oh1 & Jonathan S. Weissman1

SUMMARY

Protein synthesis by ribosomes takes place on a linear substrate but at non-uniform speeds. Transient pausing of ribosomes can affect a variety of co-translational processes, including protein targeting and folding. These pauses are influenced by the sequence of the messenger RNA.

Using an orthogonal ribosome7,8 possessing an altered anti-SD sequence, we show that pausing is due to hybridization between the mRNA and 16S ribosomal RNA of the translating ribosome. In protein-coding sequences, internal SD sequences are disfavoured,which leads to biased usage, avoiding codons and codon pairs that resemble canonical SD sites. Our results indicate that internal SD-like sequences are a major determinant of translation rates and a global driving force for the coding of bacterial genomes.

INTRODUCTION

Our current understanding of sequence-dependent translation rates in vivo derives largely from pioneering work begun in the 1980s. These studies, which measured protein synthesis times using pulse labelling, established that different mRNAs could be translated with different elongation rates. In particular, messages decoded by less abundant tRNAs were translated slowly, although this effect was exaggerated by the overexpression of mRNA,which can lead to the depletion of available tRNAs. Even with fixed tRNA usage, different synonymously coded mRNAs were translated at different rates13. This result, together with the observation of biased occurrence of adjacent codon pairs14, suggested that tRNA abundance is not the only determinant of elongation rates.

Further investigations addressingwhat determines the rate of translation in vivo, however, have been hampered by the limited temporal and positional resolution of existing techniques.

RESULTS

To provide a high-resolution view of local translation rates, we used the recently developed ribosome profiling strategy3–5 to map ribosome occupancy along each mRNA. We focused on two distantly related bacterial species, the Gram-negative bacterium Escherichia coli and the Gram-positive bacterium Bacillus subtilis. Several observations argued that ribosome transit time is proportional to the occupancy at each position. With our genome-wide view of local translation rates, we confirmed established examples of peptide-mediated stalling at transcripts secM15 and tnaC16 in E. coli and mifM17 in B. subtilis (Fig. 1a and Supplementary Fig. 6). Strikingly, in addition to these known stalling sites, the observed ribosome occupancy was highly variable across coding regions, as illustrated for secA in Fig. 1a.

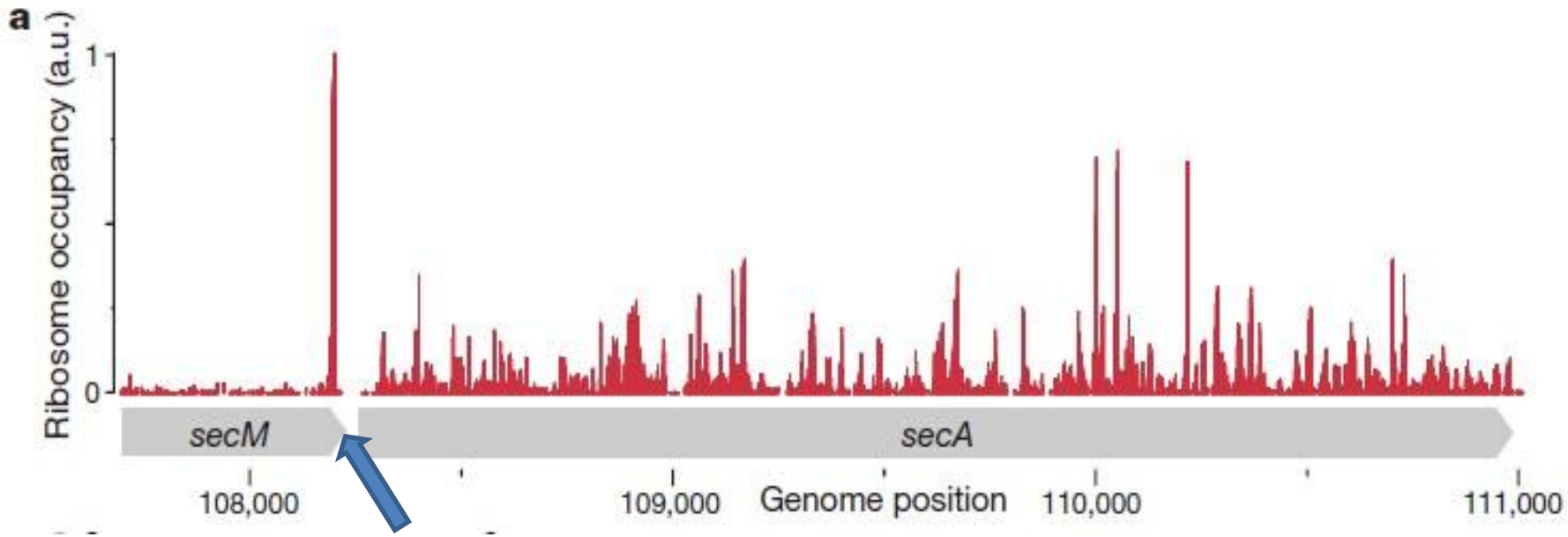


Figure 1 | Analysis of translational pausing using ribosome profiling in bacteria. a, Validation of the ribosome stalling site in the *secM* mRNA.

We first sought to determine whether the identity of the codon being decoded could account for the differences in local translation rates, by examining the average ribosome occupancy for each of the 61 codons in the ribosomal A-site.

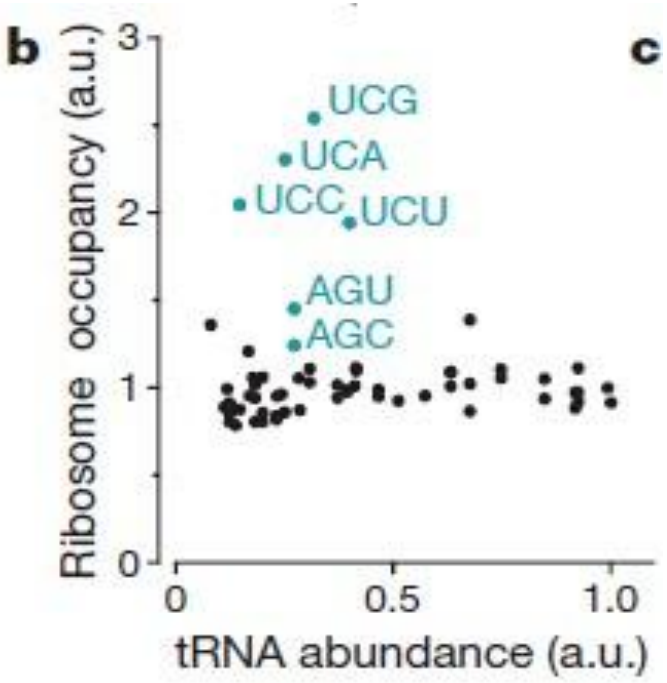
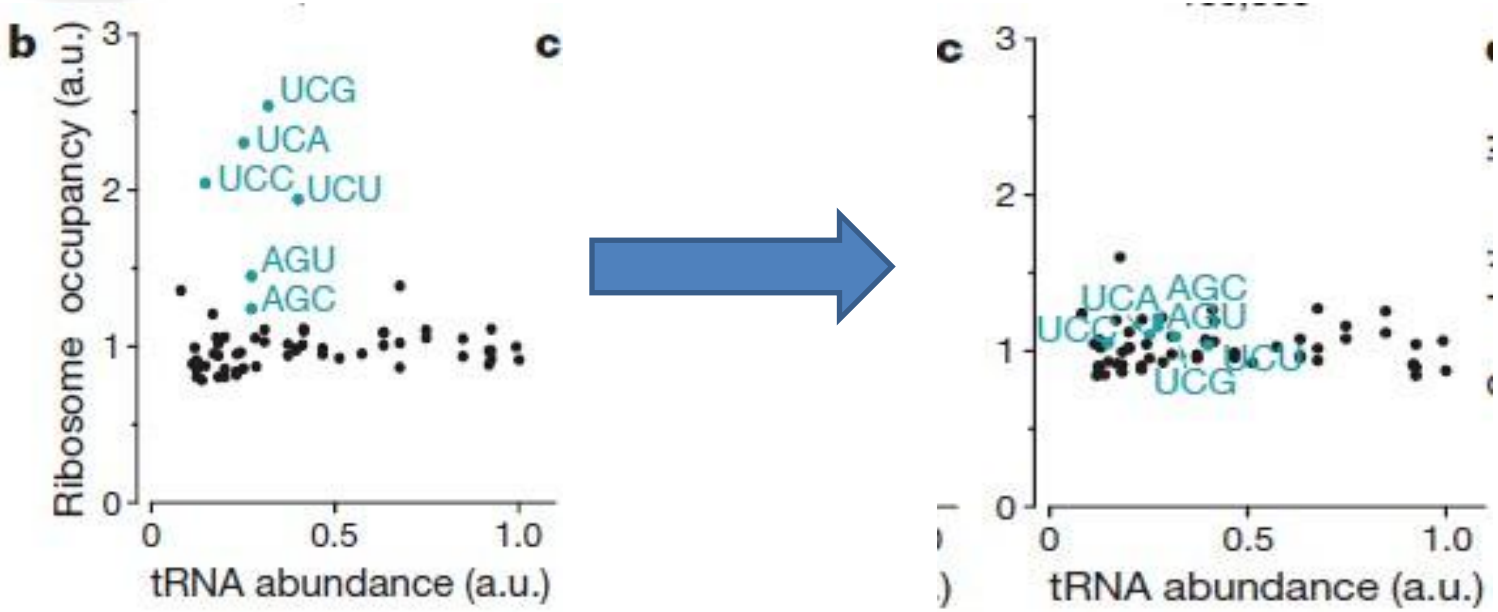


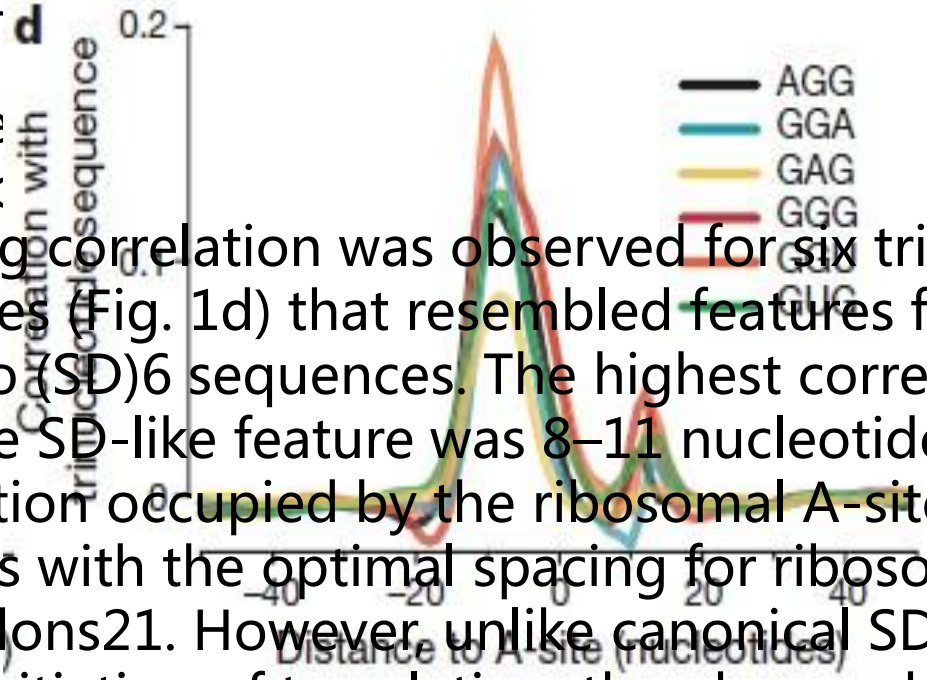
Figure1b

Surprisingly, there was little correlation between the average occupancy of a codon and existing measurements of the abundance of corresponding tRNAs (Fig. 1b, c and Supplementary Fig. 7). Most notably, the six serine codons had the highest ribosome occupancy for E. coli cultured in Luria broth (Fig. 1b).

Indeed, serine-associated pauses were greatly decreased in glucose-supplemented MOPS medium (Fig. 1c). The increase in serine codon occupancy when glucose becomes limiting confirmed our ability to capture translation rates at each codon. However, the identity of the A-site codon, which had less than a twofold effect on ribosome occupancy (Fig. 1c), could not account for the large variability in ribosome density along messages.

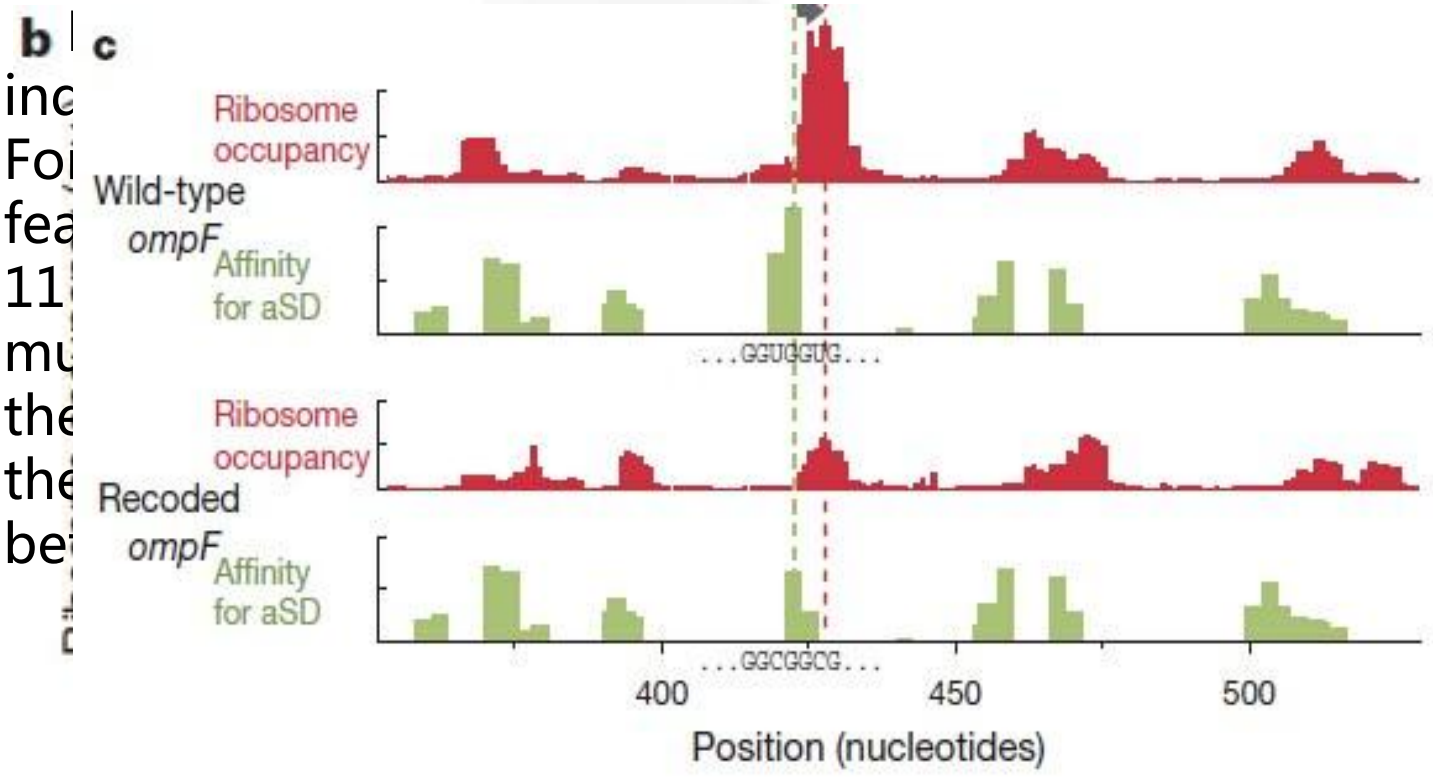


What, then, are the sequence features that cause slow translation? we calculated the crosscorrelation function between les and the presence the mRNA independ



Strong correlation was observed for six trinucleotide sequences (Fig. 1d) that resembled features found in Shine–Dalgarno (SD)6 sequences. The highest correlation occurred when the SD-like feature was 8–11 nucleotides upstream of the position occupied by the ribosomal A-site. This spacing coincides with the optimal spacing for ribosome binding at start codons21. However, unlike canonical SD sites, which enable initiation of translation, the observed pauses were associated with SD-like features within the body of coding regions.

As predicted by a model in which the interaction between mRNA and the aSD site of the 16S rRNA drives pausing, the predicted hybridization free energy of a hexanucleotide to the aSD sequence was a strong indicator of its average downstream ribosome occupancy (Fig. 2b).



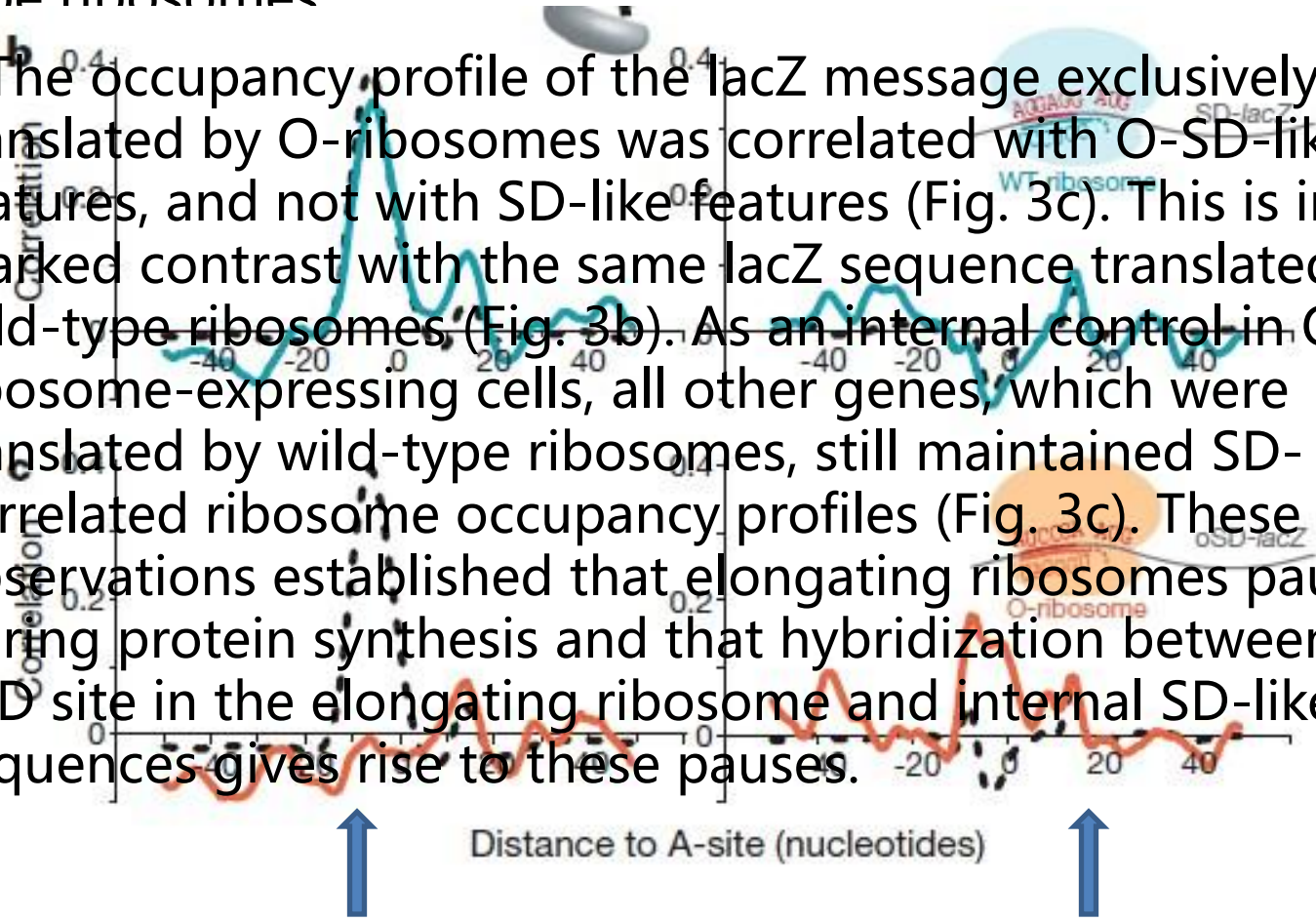
We next sought to evaluate directly whether the excess footprint density seen at internal SD-like sequences was due to pausing of elongating ribosomes rather than attempted internal initiation,driven by SD–aSD interactions .To distinguish between these possibilities, we used a previously described orthogonal ribosome (Oribosome) system in which a mutant form of the 16S rRNA with an altered aSD site is expressed together with wild-type 16S rRNA8. O-ribosomes containing the mutant 16S RNA will only translate a target mRNA that has the corresponding orthogonal SD (O-SD) sequence before its start codon. Conversely, a message whose translation is driven by the O-SD sequence will only be translated by O-ribosomes, and not by wild-type ribosomes. This system thus allows one to determine the source of regions of excess ribosome footprints, because elongating O-ribosomes would pause at internal O-SD sequences, whereas attempted internal initiation would still occur at SD sequences as a result of the cellular pool of wild-type ribosomes.

We compared the ribosome occupancy profiles of a lacZ message that was translated by either O-ribosomes or wild-type ribosomes.

The occupancy profile of the lacZ message exclusively translated by O-ribosomes was correlated with O-SD-like features, and not with SD-like features (Fig. 3c). This is in marked contrast with the same lacZ sequence translated by wild-type ribosomes (Fig. 3b). As an internal control in O-ribosome-expressing cells, all other genes, which were translated by wild-type ribosomes, still maintained SD-correlated ribosome occupancy profiles (Fig. 3c). These observations established that elongating ribosomes pause during protein synthesis and that hybridization between the aSD site in the elongating ribosome and internal SD-like sequences gives rise to these pauses.

Wild-type ribosome

Orthogonal ribosome

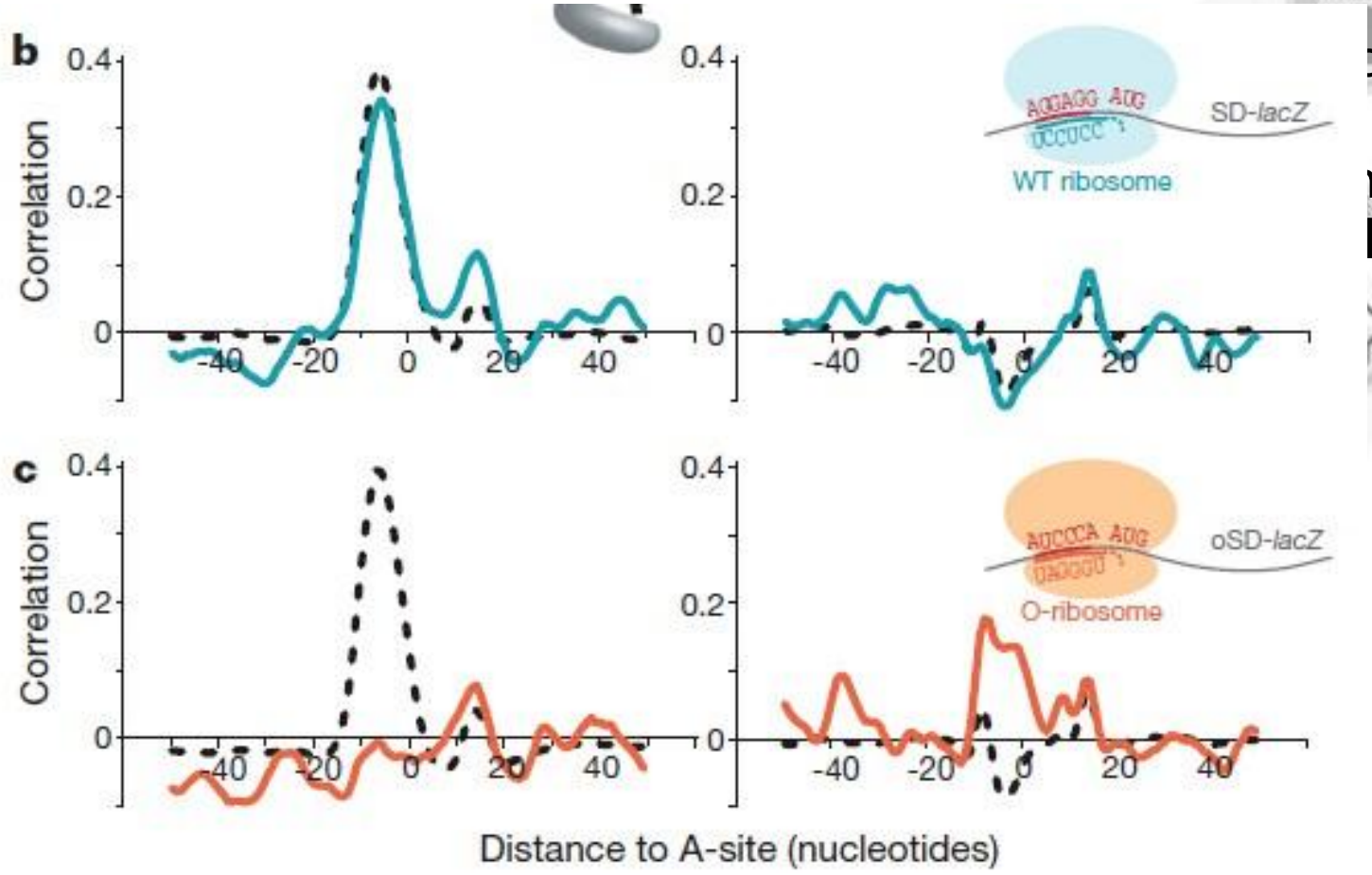


SD-like sequences

O-SD-like sequences

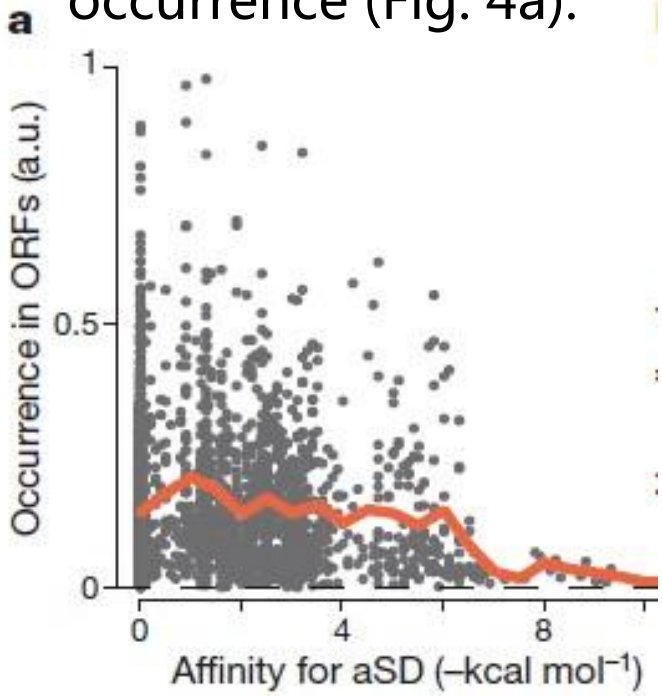
MATHOD

METHODS



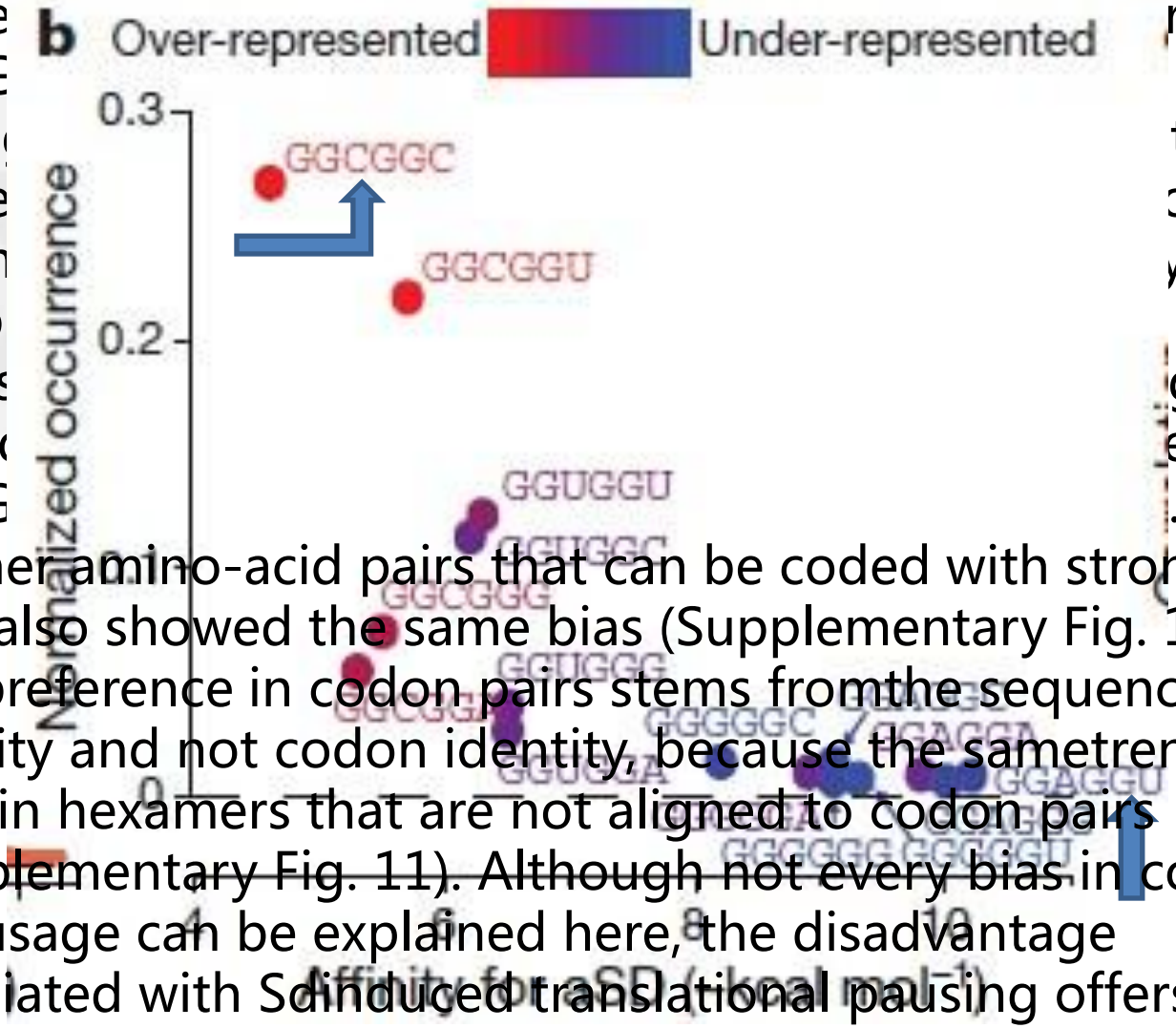
The expectation value is obtained by dividing the summation by. $N - |i|$.

Because translational pausing limits the amount of free ribosomes available for initiating protein synthesis, widespread internal SD-like sequences could decrease bacterial growth rates. Accordingly, we found that strong SD-like sequences are generally avoided in the coding region of E. coli genes: hexamer sequences that strongly bind aSD sites are universally rare, whereas low-affinity hexamers have variable rates of occurrence (Fig. 4a).



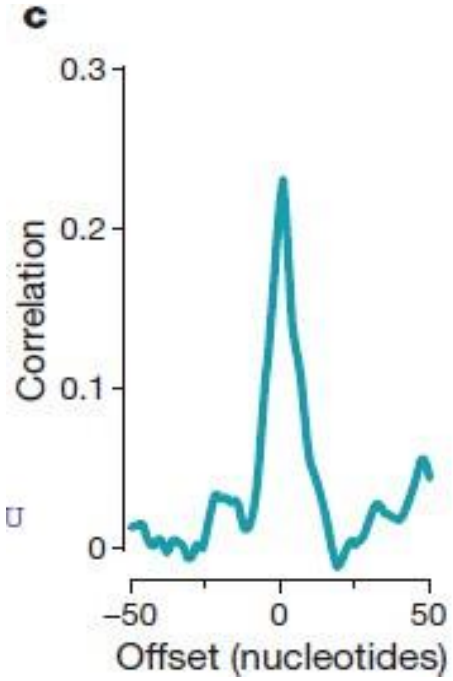
The evolutionary origin of codon selection is often attributed to differences in tRNA abundance because its level is correlated with codon usage18. Instead, we propose that SD-like codons are disfavoured as a result of their interactions with rRNA, and that tRNA expression levels followed codon adaptation.

At the codon-pair level, we can now account for the selection against two consecutive codons that resemble SD seque re coded by GG Gly-Gly codin the aSD seque ongly resem y comp repres ge of indivi e, GGAG .



Other amino-acid pairs that can be coded with strong SD sites also showed the same bias (Supplementary Fig. 10). The preference in codon pairs stems fromthe sequence identity and not codon identity, because the sametrend is seen in hexamers that are not aligned to codon pairs (Supplementary Fig. 11). Although not every bias in codon-pair usage can be explained here, the disadvantage associated with Sdinduced translational pausing offers a clear mechanistic view of why certain codon pairs are avoided.

Despite the selection against internal SD-like sequences, they remain a major driving force for translational pausing. In addition, we found similar pausing patterns between conserved genes in E. coli and B. subtilis (Fig. 4c).



For an mRNA encoding a specific protein, it may not be possible to fully eliminate sequences with affinity for the aSD site without changing the peptide sequence. The optimization for translation rate therefore cannot be achieved only at the level ofmRNA coding: it is also constrained by the requirement to make a functional peptide sequence.

From a more practical perspective, ribosome pausing at internal SD sites presents both a challenge and an opportunity for heterologous protein expression in bacteria. Over expression of eukaryotic proteins with strong internal SD sites would sequester ribosomes and compromise protein yield. Internal SD sequences could be reduced by recoding the gene, which has not been considered in conventional strategies of simple codon optimization or overexpression of rare tRNAs. Conversely, recoding can introduce internal SD sites if pausing is required for co-translational processing. Positioning of internal SD sites therefore adds another dimension to the optimization of heterologous protein expression.

THE END

By YiZhuo Wang