

# The Frequency of Internal Shine–Dalgarno-like Motifs in Prokaryotes

Gaurav D Diwan<sup>1,2,\*</sup> and Deepa Agashe<sup>1,\*</sup>

<sup>1</sup>National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

<sup>2</sup>SASTRA University, Thanjavur, India

\*Corresponding author: gauravd@ncbs.res.in; dagashe@ncbs.res.in.

Accepted: April 27, 2016

## Abstract

In prokaryotes, translation initiation typically depends on complementary binding between a G-rich Shine–Dalgarno (SD) motif in the 5′ untranslated region of mRNAs, and the 3′ tail of the 16S ribosomal RNA (the anti-SD sequence). In some cases, internal SD-like motifs in the coding region generate “programmed” ribosomal pauses that are beneficial for protein folding or accurate targeting. On the other hand, such pauses can also reduce protein production, generating purifying selection against internal SD-like motifs. This selection should be stronger in GC-rich genomes that are more likely to harbor the G-rich SD motif. However, the nature and consequences of selection acting on internal SD-like motifs within genomes and across species remains unclear. We analyzed the frequency of SD-like hexamers in the coding regions of 284 prokaryotes (277 with known anti-SD sequences and 7 without a typical SD mechanism). After accounting for GC content, we found that internal SD-like hexamers are avoided in 230 species, including three without a typical SD mechanism. The degree of avoidance was higher in GC-rich genomes, mesophiles, and N-terminal regions of genes. In contrast, 54 species either showed no signature of avoidance or were enriched in internal SD-like motifs. C-terminal gene regions were relatively enriched in SD-like hexamers, particularly for genes in operons or those followed closely by downstream genes. Together, our results suggest that the frequency of internal SD-like hexamers is governed by multiple factors including GC content and genome organization, and further empirical work is necessary to understand the evolution and functional roles of these motifs.

**Key words:** Shine–Dalgarno sequence, anti-SD affinity, hexamer frequency, GC content, translational pausing.

## Introduction

The process of translation may be divided into three broad steps: initiation, elongation, and termination. In prokaryotes, initiation typically begins when the Shine–Dalgarno (SD) sequence in an mRNA molecule is recognized by a complementary anti-Shine–Dalgarno (anti-SD) sequence on the 3′ tail of the 16S ribosomal RNA. The full 70S ribosome complex then assembles and proceeds to translate the mRNA (reviewed by Simonetti et al. 2009). The classical view of translation elongation posits that the rate of ribosome translocation on mRNA molecules is determined primarily by the codon usage and tRNA content of bacterial cells (Varenne et al. 1984; Sørensen et al. 1989). However, recent studies have highlighted the significant contribution of other factors such as mRNA secondary structure (Del Campo et al. 2015; Gorochowski et al. 2015), the charge on amino acids in the ribosome exit tunnel (Sabi and Tuller 2015), and internal SD-like sequences in the transcript (Li et al. 2012). In particular,

sequences in the coding region that are complementary to the anti-SD sequence (i.e., internal SD-like sequences) can act as pause sites for the ribosome (Wen et al. 2008; Li et al. 2012; Fluman et al. 2014), and may limit the amount of free ribosomes available for translation initiation (Li et al. 2012). Thus, if SD-like sequences in the coding region decrease the overall translation elongation rate in prokaryotes, the occurrence of such sequences should be deleterious. Li et al. (2012) further suggested that purifying selection against internal SD-like sequences may also act against codons that compose such motifs, driving codon use across bacteria and archaea.

Although internal SD-like motifs are suggested to be deleterious (Li et al. 2012), recent work has uncovered instances where ribosomal pausing due to internal SD-like sequences is beneficial. For example, in *Escherichia coli*, ribosomal pauses at the beginning of genes encoding membrane-bound proteins (codons 16–60) aid co-translational folding of the synthesized

polypeptide chain and are thus important for accurate targeting of these proteins (Fluman et al. 2014). Internal SD-like motifs can also facilitate programmed ribosomal frameshifts, which can lead to the production of alternate proteins from the same transcript. For example, in *E. coli*, production of the polypeptide chain release factor 2 (Weiss et al. 1988) and the *tau* subunit of DNA polymerase III (Larsen et al. 1994; Chen et al. 2014) depend on programmed frameshifts aided by internal SD-like sequences in the respective genes. In cases where long stretches of rare codons in the mRNA cause ribosomal drop-off, internal SD-like sequences may stabilize the interaction between the mRNA and the ribosome and prevent ribosomes from falling off. Indeed, SD-like sites are found more frequently before larger clusters of rare codons in *E. coli*, especially in highly expressed genes (Ponnala 2010). Finally, SD-like sequences at domain boundaries improved the solubility of a synthase and GFP protein chimera in *E. coli* (Vasquez et al. 2015). Thus, depending on their location in the genome, SD-like motifs in coding regions of prokaryotes could potentially evolve under positive rather than purifying selection.

Given this evidence for context-dependent benefits of internal SD-like motifs, it is difficult to make general predictions about the evolution of such motifs across prokaryotic genomes. The issue becomes particularly clouded by recent work suggesting weak or no selection against internal SD-like motifs. For instance, a recent study shows that ribosomal pausing due to internal SD-like sequences is rare in *E. coli* (Mohammad et al. 2016), in stark contrast to the earlier report by Li et al. (2012). In addition, some prokaryotes either have unusual anti-SD sequences (Lim et al. 2012) or do not use the SD mechanism of translation initiation (Accetto and Avguštin 2011; Kramer et al. 2014) and thus may not face selection to avoid internal SD-like sequences. Finally, since SD sequences are typically G-rich (Shine and Dalgarno 1975a, b), they are also more likely to occur in GC-rich genomes. Thus, it is important to separate the impact of GC content from selection on internal SD-like sequences. Ultimately, internal SD-like motifs in different genes, gene regions or organisms may face opposing and varying selection pressures including direct selection acting on translational pauses, as well as indirect selection via GC and/or mRNA structure and stability.

To address these issues, we analyzed the frequency and genomic location of internal SD-like motifs in 284 prokaryotes. We included 249 eubacteria and 28 archaeobacteria whose anti-SD sequences in the 3' 16S rRNA tails were previously characterized (Nakagawa et al. 2010), and seven prokaryotes that have unusual anti-SD sequences (Lim et al. 2012) and/or do not use the SD mechanism to initiate translation (Kramer et al. 2014). The latter seven species serve as negative controls since we expect negligible selection acting on internal SD-like sequences in these species. To facilitate comparison across species, we focused on internal hexanucleotide sequences (hexamers) in coding regions. Following an earlier analysis of

internal SD-like motifs (Li et al. 2012), we calculated the anti-SD affinity of all possible 4,096 ( $4^6$ ) hexamers as the free energy of binding between the hexamer and the predicted anti-SD sequence of each species. For each species, we estimated the expected frequency of each hexamer given the organism's genome GC (using randomized genomes). We then tested whether the observed frequencies of high-affinity (SD-like) hexamers were lower than expected, as predicted by purifying selection against such motifs. Finally, we analyzed the location of internal SD-like hexamers within genomes. We show that internal SD-like sequences are not universally avoided in prokaryotic genomes, and highlight various factors that may influence their evolution.

## Materials and Methods

### Genomes Analyzed

All analyses were carried out in R (R Development Core Team 2015) using the 'seqinr' package (Charif et al. 2005). We downloaded ffn files containing coding nucleotide sequences for the primary chromosome of 284 prokaryotes from the NCBI ftp site. The set included 277 species whose anti-SD analyzed was previously described (Nakagawa et al. 2010); six species with unusual anti-SD sequences (Lim et al. 2012); and one species that does not use the anti-SD mechanism of translation initiation (Kramer et al. 2014). The 3' 16S tail sequences for the first set of 277 organisms were obtained from the study by Nakagawa et al. (2010). However, functionally not all bases in the 3' tail might be involved in the anti-SD::SD interaction. Thus, we used the highly conserved 8-nt region from the above sequences as the predicted anti-SD sequence of the organism. The list of organisms analyzed and summary of results is given in [supplementary table S1, Supplementary Material](#) online.

### Calculating anti-SD Affinity and Hexamer Frequency

We used the minimum free energy of hybridization of hexamers to the anti-SD sequence as a proxy for the anti-SD affinity of each hexamer. We calculated the minimum free energy of hybridization of all possible 4,096 hexamers to the respective anti-SD sequence of each prokaryote using the RNAsubopt program in the Vienna RNA package (Lorenz et al. 2011). The binding energy was predicted using default program parameters (temperature = 37°C, energy range = 1), allowing contribution from dangling ends (Li et al. 2012). The program calculated several anti-SD affinities for every hexamer, but we considered the lowest binding energy in every case as a representation of the maximum binding ability of the hexamer. For anti-SD affinity calculations at the optimal growth temperature of organisms, we changed the default temperature value in the program to the respective growth temperature of the organism. We obtained the optimal growth

temperatures of 211 organisms in our dataset from a previous report (Lobry and Necsulea 2006).

Next, we used the command 'count' from the `seqinr` package in R to calculate the gene-wise frequency of each hexamer in each prokaryote as: count of the hexamer in gene/total number of possible hexamers in the gene. We then calculated the observed hexamer frequency (HF) (henceforth "HF<sub>obs</sub>") as the mean frequency of a hexamer across all genes in a genome. We could then test whether HF<sub>obs</sub> for each hexamer is correlated with its anti-SD affinity. However, this would give us 4,096 correlations that span a wide range of anti-SD affinities, with no empirical basis to distinguish between hexamers with similar affinity values. Thus, we decided to group hexamers into classes such that hexamers within a class would have similar anti-SD affinities. We could then combine the frequencies of all hexamers within a class and analyze their frequencies across genomes. As is generally observed for RNA–RNA interactions, the strength of the SD::anti-SD interaction presumably depends on the number as well as the identity of paired bases. We chose to classify hexamers by the number of paired bases. We calculated the maximum possible binding affinity given a pairing of three, four, or five bases of a hexamer and the anti-SD sequence for each organism (called aff<sub>3</sub>, aff<sub>4</sub>, and aff<sub>5</sub>). Using `RNAsubopt` as above, we determined these maximum affinity values for each organism by calculating the binding affinity of all possible 3-, 4-, and 5-mers to the anti-SD sequence of each organism. We then partitioned the hexamer affinity values calculated earlier into four bins, each representing the range of possible affinity values if the hexamer and anti-SD binding involved  $\leq 3$ , 3–4, 4–5, or  $\geq 5$  bases (bins: 0 to aff<sub>3</sub>, aff<sub>3</sub> to aff<sub>4</sub>, aff<sub>4</sub> to aff<sub>5</sub>, and aff<sub>5</sub> to highest affinity; [supplementary table S2, Supplementary Material online](#)). Since the highest affinity bin (aff<sub>5</sub> to highest) represents maximum binding between the SD and anti-SD sequence, we denoted hexamers in this bin as "SD-like hexamers" and focused on these hexamers in our subsequent analysis.

#### Accounting for Genome GC Content and Identifying Genomes That Avoid SD-like Hexamers

We generated two sets of 250 randomized genomes to determine the expected HF given the genome GC content of an organism. We chose the number of randomized genomes such that the mean HF did not change with increasing number of randomizations (see [Supplementary Methods and supplementary fig. S1, Supplementary Material online](#)). For each organism, we generated 250 randomized genomes, such that synonymous codons within each gene were shuffled without replacement. Thus, we maintained the amino acid sequence, the gene, and genome GC% of original genomes, as well as the dinucleotide frequencies (Spearman's  $\rho \geq 0.96$ , [supplementary fig. S2, Supplementary Material online](#)), but scrambled any local associations between specific

codons. A similar protocol has been used previously to determine selection acting on various features of protein coding sequences (Katz and Burge 2003; Goroehowski et al. 2015). Next, we generated 250 more randomized genomes per organism by scrambling the sequence of every gene (except the start and stop codon), sampling each base with a probability equal to its proportion in the wild type gene. Thus, we maintained the GC% of individual genes as well as the genome, but scrambled the amino acid sequence.

We calculated mean HF values for each hexamer in each randomized genome as described earlier. We then calculated HF<sub>exp (shuff or GC)</sub> as the average HF across all randomized genomes per organism. Finally, we calculated the corrected HF for each hexamer as:  $\Delta HF_{shuff (or GC)} = HF_{obs} - HF_{exp (shuff or GC)}$ . Within each hexamer affinity bin (0–aff<sub>3</sub>, aff<sub>3</sub>–aff<sub>4</sub>, aff<sub>4</sub>–aff<sub>5</sub>, and aff<sub>5</sub>–highest), we thus obtained multiple  $\Delta HF_{shuff}$  values for each organism (one per hexamer). For each affinity bin, we tested whether the values of HF<sub>obs</sub> were significantly different from HF<sub>exp</sub> using a nonparametric, paired Wilcoxon signed rank test (with *P*-value correction for multiple comparisons using the Benjamini–Hochberg method). To help visualize these patterns within each affinity bin, we show the median  $\Delta HF_{shuff}$  value for all hexamers in each bin ([fig. 2 and supplementary fig. S6, Supplementary Material online](#)). For a given genome,  $\Delta HF_{shuff} < 0$  indicates that the observed frequency of hexamers within the affinity bin is typically lower than expected. Hence, genomes with median  $\Delta HF_{shuff} < 0$  and *P* values  $< 0.05$  in the aff<sub>5</sub>–highest bin were marked as those that avoid SD-like hexamers. Similarly, genomes with *P* values  $\geq 0.05$  were marked as genomes that do not avoid SD-like hexamers. Genomes with *P* values  $< 0.05$  and median  $\Delta HF_{shuff} > 0$  in the aff<sub>5</sub>–highest bin were marked as genomes that are enriched in SD-like hexamers.

#### Calculating Codon Use and Codon Occurrence in High-Affinity Hexamers

If avoidance of SD-like motifs drives codon use (Li et al. 2012), we would expect that codons that frequently occur in SD-like hexamers (hexamers in the bin aff<sub>5</sub>–highest) should be relatively rare in coding regions. To test this, we first calculated the relative use of each synonymous codon encoding each amino acid. For each genome, we determined the Relative Synonymous Codon Usage (RSCU) value of each codon as the number of times the codon occurs in the entire coding region divided by the total number of occurrences of the amino acid it encodes. Next, using the codon composition table of all SD-like hexamers, we calculated the probability that a given codon occurs in SD-like hexamers. Thus, a codon with an occurrence value of 1 would always occur in SD-like hexamers (e.g., AGG has an occurrence value of 0.91), whereas a codon with a value of 0 would never occur in SD-

like hexamers (e.g., AAA). For each organism, we calculated average RSCU for all codons with similar occurrence values.

### Determining the Relative Position and Potential Function of Internal SD-like Hexamers

We determined the relative position of SD-like hexamers (bin:  $\text{aff}_5$ –highest) by dividing all genes in an organism into 20 equal parts, counting the total number of SD-like hexamers in each part across genes, and dividing it by the total number of SD-like hexamers in the entire coding region. This normalized value is the fraction of SD-like hexamers that occur in a particular region across all genes in that organism. We then calculated the mean fraction of SD-like hexamers across all organisms in a given category (shown in fig. 5) and tested whether this fraction is significantly different from the expected value of 5% for each gene region.

For each organism, we determined the Minimum Folding Energy (MFE) of the first 51 nts of genes either with or without an N-terminal SD-like hexamer in the first 51 bases. We ensured that the length of genes in the second group was comparable (within  $\text{mean} \pm 2 \text{ SD}$ ) to that of the first group. We used the RNAfold program (Lorenz et al. 2011) to calculate MFE at a temperature of 37°C for each gene, and obtained the average value for genes in each group of each species. Finally, we compared the values of mean MFE for genes with and without an N-terminal SD-like hexamer across all species. More negative MFE values for the first group would indicate that N-terminal SD-like hexamers are generally likely to increase mRNA structure in the N-terminal regions of genes.

Next, we determined whether a focal gene containing a C-terminal SD-like hexamer (within the final 5% of the gene) was often followed by another gene with an overlapping reading frame. To do this, we calculated the length of the intergenic region between genes with a C-terminal SD-like hexamer and the succeeding gene. If the length of the intergenic region was  $\leq 0$ , we noted the gene as an overlapping gene. We also determined the distance between the centre of the SD-like hexamer in the C-terminal region and the start codon of the next gene. Next, we generated a frequency distribution of distances for each organism and calculated the mean frequency for each of the following categories: (i) organisms that avoid internal SD-like hexamers; (ii) organisms that do not avoid internal SD-like hexamers; (iii) organisms that are enriched in internal SD-like hexamers; and (iv) organisms that do not use the SD mechanism of translation initiation.

Finally, to determine whether a gene is part of an operon, we identified focal genes with a C-terminal SD-like hexamer using their PID from NCBI PTT files. We retrieved the operon (or “transcription unit”) information for 253 organisms from our list that were also described in the database created by Moreno-Hagelsieb and Collado-Vides (2002). We determined whether

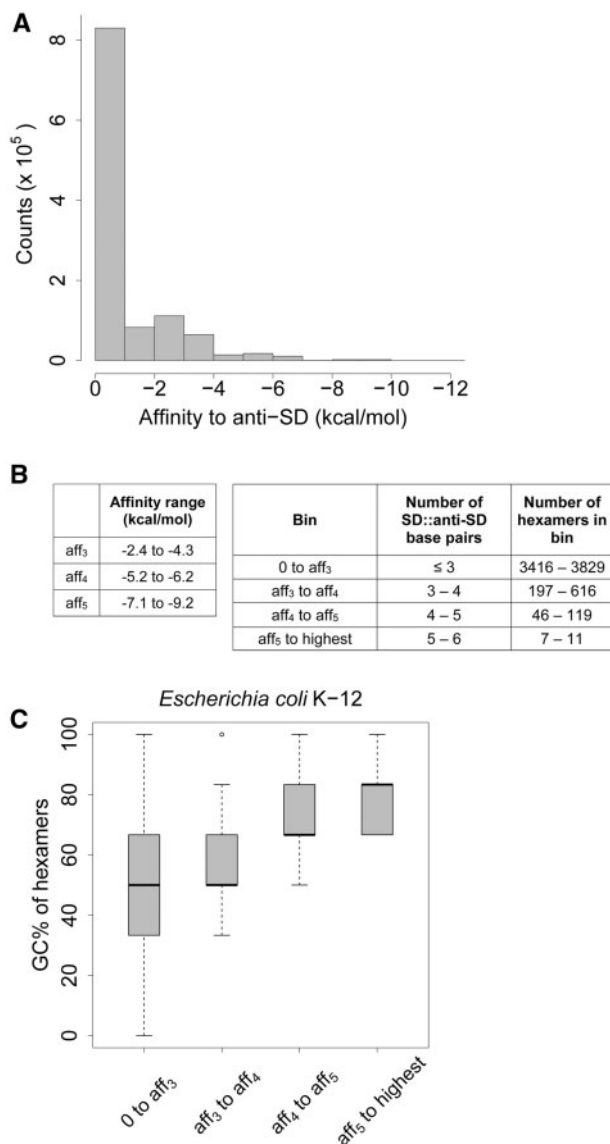
the focal genes were part of operons by checking if they belonged to a transcription unit with two or more genes. As a comparison, we sampled an equal number of randomly selected genes that did not have a C-terminal SD-like hexamer, and determined whether each of them was a part of an operon.

## Results

### The Frequency of Internal High Affinity Hexamers Varies Across Organisms

In a previous analysis of 277 prokaryotes, Nakagawa et al. (2010) described 18 unique 16S rRNA tail sequences with a highly conserved core region (5'—ACCUCCU U/A—3'). We used these eight nucleotides as the predicted anti-SD sequence of each organism (though the region is highly conserved, there are minor variations in the sequence across species). For each species, we calculated the binding affinity of all possible 4,096 hexamers to the predicted anti-SD sequence of the species (henceforth “affinity”; more negative values indicate higher affinity). We found that most hexamers had no affinity to the anti-SD sequence (fig. 1A). Instead, in 273 bacteria the hexamer GGAGGU had the highest anti-SD affinity, with values ranging from  $-10.1$  to  $-10.7$  kcal/mol depending on the specific anti-SD sequence (supplementary table S2, Supplementary Material online). Given the highly skewed distribution of binding affinities, we binned hexamers into four affinity classes denoting an increasing number of paired bases between the SD and the anti-SD sequence: (i) 0– $\text{aff}_3$ , (ii)  $\text{aff}_3$ – $\text{aff}_4$ , (iii)  $\text{aff}_4$ – $\text{aff}_5$ , and (iv)  $\text{aff}_5$ –highest. The values  $\text{aff}_3$ ,  $\text{aff}_4$ , and  $\text{aff}_5$  represent the highest affinity of 3-, 4-, and 5-mer nucleotide sequences binding to the anti-SD sequence of the organism (fig. 1B, supplementary table S2, Supplementary Material online). The binning allowed us to test whether there was stronger selection to avoid hexamers in the highest-affinity bin ( $\text{aff}_5$ –highest; henceforth “SD-like hexamers”), since these hexamers most closely resembled SD sequences.

When we directly examined the relationship between raw HF and anti-SD affinity values, we found that the relationship was nonlinear and varied across organisms (supplementary fig. S3, Supplementary Material online). Typically, high-affinity hexamers were rare in AT-rich genomes (supplementary fig. S3A, Supplementary Material online) compared with GC-rich genomes (supplementary fig. S3C, Supplementary Material online), suggesting that the strength of selection to avoid high affinity hexamers may vary across organisms as a function of genome GC. The SD sequence is usually G-rich, and accordingly we observed that the GC content of hexamers increased across affinity bins (fig. 1C, supplementary fig. S4, Supplementary Material online). Thus, if genome GC imposes a sufficiently strong constraint, GC-rich genomes may frequently contain internal high-affinity hexamers. Consistent with this expectation, we found that HF is positively correlated



**FIG. 1.**—(A) Histogram of anti-SD affinities of 4,096 hexamers across 277 organisms. (B) Range of anti-SD affinity values used to define affinity bins, and the range of number of hexamers present in each bin across the 277 species. (C) GC content of hexamers across affinity bins for *E. coli* K-12 substr. W3110.

with genomic GC content across affinity bins, except for the bin containing the lowest-affinity hexamers (supplementary fig. S5, Supplementary Material online). Although GC-rich genomes appear to be enriched in SD-like hexamers relative to low-GC genomes, we need to explicitly control for the expected variation in HF given genome GC content.

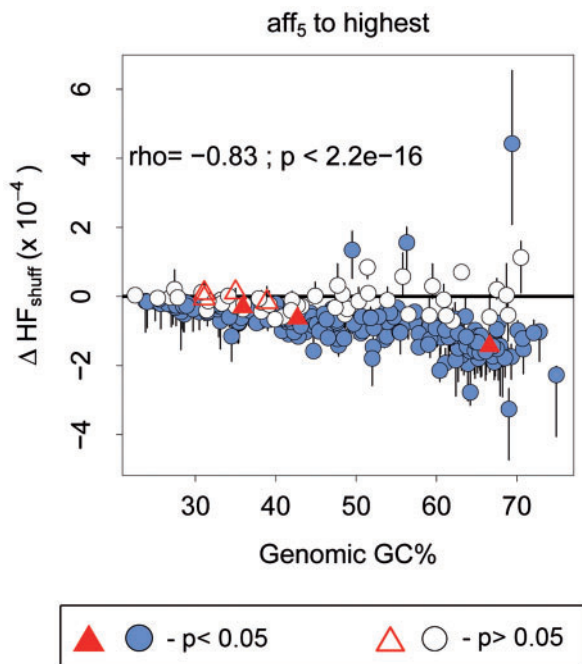
### SD-like Hexamers Are Not Universally Avoided in the Coding Regions of Prokaryotes

To calculate expected HF given the genome GC content, we generated 250 randomized genomes per organism where

synonymous codons within each gene were shuffled but the amino acid sequence was maintained. The codon shuffling allowed us to determine the expected frequency of hexamers when genome (and gene) GC content was retained but the codon arrangement was altered such that new hexamers could be formed. We determined whether observed HF values were significantly different from the expected mean HF of the 250 shuffled genomes for all hexamers in every bin. For each organism, we then calculated the corrected frequency of each hexamer ( $\Delta HF_{shuff}$ ) as  $HF_{obs} - HF_{exp}$ .

To analyze broad patterns within an affinity bin, we calculated the median corrected HF value for all hexamers in each affinity bin. A positive value of median  $\Delta HF_{shuff}$  indicates that most hexamers in that bin are used more often than expected given the organism's genome GC and codon order. We found that median  $\Delta HF_{shuff}$  decreased with genome GC% in all affinity bins, so that GC-rich genomes had more negative  $\Delta HF_{shuff}$  values (fig. 2, supplementary fig. S6, Supplementary Material online). Thus, the degree of avoidance of SD-like hexamers was stronger in GC-rich genomes (Spearman's  $\rho = -0.83$ ;  $P < 2.2e-16$  for colored circles below the black line in fig. 2). Furthermore, the negative correlation between median  $\Delta HF_{shuff}$  and GC content was the strongest for hexamers in the highest affinity bin, indicating that GC-rich genomes are especially depleted in SD-like hexamers (fig. 2; compare with supplementary fig. S6, Supplementary Material online). However, this pattern may be generally true for all G-rich hexamers (since high-affinity hexamers have  $\geq 4$  guanine bases on average). To test whether the correlation is specific to only SD-like hexamers, we also analyzed the correlation between median corrected HF ( $\Delta HF_{shuff}$ ) of low affinity G-rich hexamers and genome GC%. We found that this correlation was significantly weaker ( $\rho = -0.51$ ; supplementary fig. S7, Supplementary Material online) than that observed for high-affinity hexamers ( $\rho = -0.83$ ; Fisher's  $z$  test comparing correlation coefficients:  $P < 0.01$ ). These results are consistent with the hypothesis that selection specifically decreases the frequency of SD-like hexamers rather than G-rich hexamers in general.

Of the 277 genomes that use the SD-mechanism of translation initiation, most (227) had a significantly lower frequency of SD-like hexamers than expected (colored circles below black line in fig. 2), consistent with purifying selection against SD-like hexamers. The overall pattern of avoidance of SD-like hexamers was observed even when we focused on the single highest-affinity hexamer for each species: its frequency was typically lower than expected given the genome GC content (paired Wilcoxon rank sum test,  $P < 2.2e-16$ ; supplementary fig. S8, Supplementary Material online). However, in 47 species from diverse clades (supplementary fig. S9, Supplementary Material online), the frequency of SD-like hexamers was no different than expected given their genome GC content (white circles in fig. 2). Thus, internal SD-like hexamers are not avoided in all prokaryotes. It is also noteworthy that



**FIG. 2.**—Median and interquartile range of corrected HF ( $\Delta HF_{shuff} = HF_{obs} - HF_{exp}$ ) for hexamers as a function of genome GC%, in the highest affinity hexamer bin. The horizontal solid black line indicates no difference in observed and expected HF. Red triangles indicate organisms that do not use the SD mechanism. Colored circles and triangles indicate genomes for which  $HF_{obs}$  is significantly different from  $HF_{exp}$ . Spearman's rho and *P*-value is given for the relationship between  $\Delta HF_{shuff}$  and genome GC%, for colored circles below 0.

three genomes were significantly enriched in SD-like hexamers, suggesting that these motifs may potentially evolve under positive selection in some genomes (colored circles above the black line in fig. 2). We observed similar patterns of avoidance of internal SD-like motifs when we used a different randomization protocol to calculate expected HF given genome GC (supplementary fig. S10A–D, Supplementary Material online; see Methods). In this protocol, we scrambled genomes such that all translational meaning (except the start and stop codon) was lost, but gene GC was maintained. Using this randomization, we identified many more genomes (128) that do not show evidence for purifying selection against internal SD-like hexamers, and eight genomes that were enriched in SD-like hexamers (supplementary fig. S10D, Supplementary Material online).

Lastly, we analyzed HF in the genomes of seven prokaryotes (six described by Lim et al. 2012 and one described by Kramer et al. 2014) that have unusual anti-SD sequences and/or that do not use the SD mechanism of translation initiation. We expected that these genomes should face weak or negligible selection to avoid internal SD-like hexamers. Since their annotated anti-SD sequences do not have high affinity for any hexamer, we used the anti-SD affinity values from *E. coli* K-12

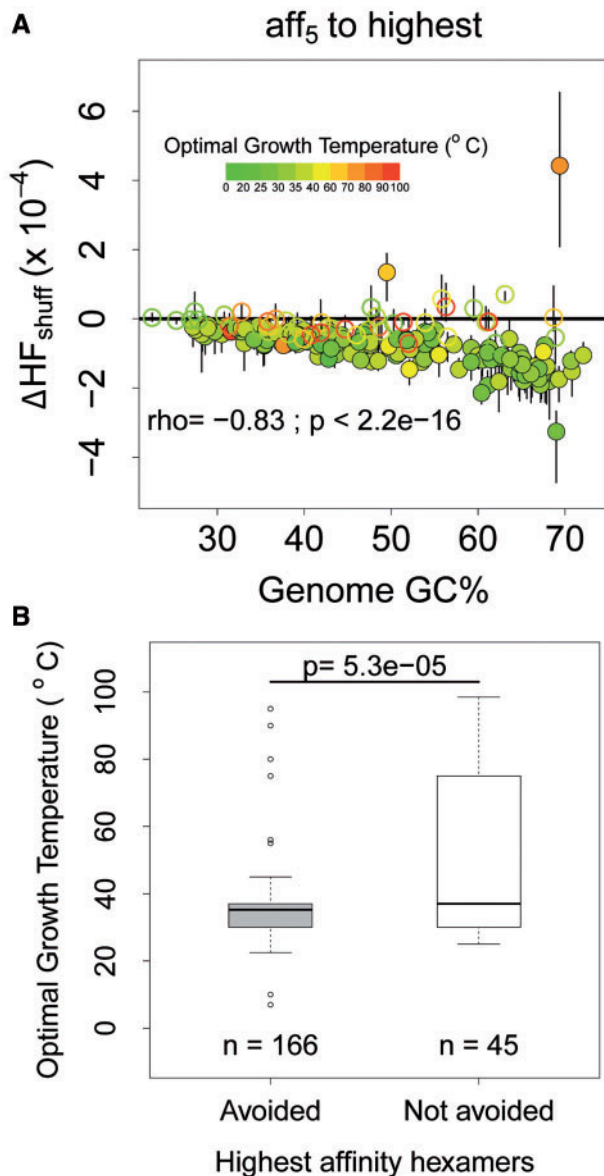
(the anti-SD sequence of the archaeon from Kramer et al. 2014 is not well annotated). We found that three of these genomes showed significantly lower HF than expected (filled red triangles in fig. 2; supplementary fig. S10D, Supplementary Material online). Thus, even with weak or no selection against internal SD-like motifs, some genomes are significantly depleted in such motifs. As observed for species that use the SD initiation mechanism, the magnitude of the deviation from expected HF was higher for the GC-rich genome (*Haloflexax volcanii*), although the sample size is too low in this case to derive firm conclusions. Together, these results suggest that the observed depletion of internal SD-like hexamers may be independent of their impact on ribosomal pausing, and that other selective pressures may also play a role in determining the frequency of such motifs.

### Avoidance of Internal High Affinity Hexamers Is Correlated with Optimal Growth Temperature

For the above analysis, we calculated SD::anti-SD affinity at 37°C, whereas the growth temperature of different organisms varies substantially. Temperature should affect the stability of the SD::anti-SD interaction, and hence growth temperature may affect the strength of selection to avoid SD-like hexamers. To test this, we re-calculated affinity values of hexamers for 211 organisms in our dataset for which optimal growth temperatures were reported previously (Lobry and Necsulea 2006). We observed that the set of hexamers with the highest affinity to the anti-SD sequence varied with optimal growth temperature. As above, we calculated the median  $\Delta HF_{shuff}$  for each species and identified organisms that avoid the new set of high affinity hexamers (filled circles below black line in fig. 3A). We found the same pattern that we had observed earlier: a negative correlation between HF and genome GC content, independent of optimal growth temperature (Spearman's rho = -0.83; Generalised linear model:  $\Delta HF_{shuff} \sim GC * t_{opt}$ ;  $p_{GC} = 2.28e-06$ ;  $p_{t_{opt}} = 0.09$ ;  $p_{GC * t_{opt}} = 0.09$ ). Thus, even when we calculated the strength of SD::anti-SD binding at different temperatures, GC-rich genomes still showed stronger avoidance of internal SD-like hexamers. Overall, the optimal growth temperature of organisms that avoid high affinity hexamers was significantly lower than that of organisms that do not avoid internal SD-like motifs ( $p = 5.3e-05$ , Wilcoxon rank sum test; fig. 3B). These results indicate that high growth temperatures may reduce the strength of selection to avoid high-affinity hexamers.

### Codon Use and the Frequency of Internal SD-like Sequences

Li et al. (2012) proposed that selection against internal SD-like hexamers may also drive selection against codons that compose such motifs (and therefore show high affinity to the 16S rRNA tail). However, the generality of this hypothesis remains unclear. Based on our results described above, we predicted



**FIG. 3.**—(A) Median and interquartile range of corrected HF ( $\Delta HF_{shuff} = HF_{obs} - HF_{exp}$ ) for hexamers as a function of genome GC%, for the highest affinity bin. The horizontal solid black line indicates no difference in observed and randomized HF. Points are colored according to the optimal growth temperature of the organism. Filled circles indicate that  $HF_{obs}$  is significantly different from  $HF_{exp}$ . Spearman's rho and  $P$ -values are given for the relationship between  $\Delta HF$  and genome GC% (for filled circles below 0). (B) Boxplots showing optimal growth temperatures for genomes that avoid SD-like hexamers versus genomes that do not (based on SD::anti-SD affinity calculations at the respective growth temperatures). The distributions are significantly different from each other (Wilcoxon rank sum test).

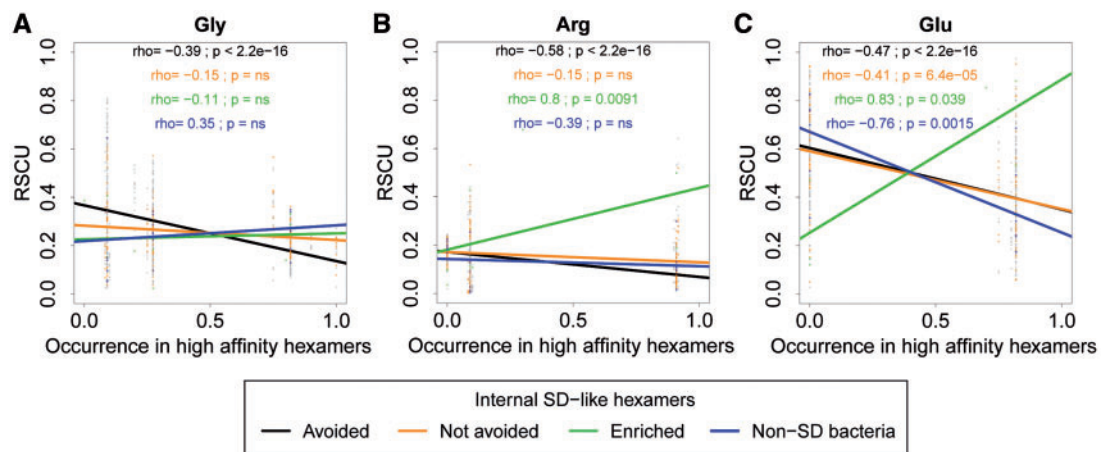
that genomes showing patterns consistent with selection against internal SD-like hexamers should also avoid codons that frequently occur in SD-like motifs. In contrast, occurrence in SD-like motifs should not affect codon use in genomes that do not avoid internal SD-like hexamers, are enriched in the

motifs, or that do not use the SD mechanism. We tested this by determining the relationship between RSCU and the probability that the codon occurs in SD-like hexamers (i.e., hexamers in the highest affinity bin). Most codons do not occur in SD-like hexamers (supplementary fig. S11, Supplementary Material online) and hence we only focus on codons for Glycine, Arginine, and Glutamic acid.

As predicted, we found that in the 227 species that avoid internal SD-like hexamers (see previous section), codons that occur frequently in SD-like hexamers are typically avoided in coding sequences (i.e., have lower RSCU values; black points and lines in fig. 4). We observed this relationship for all three amino acids examined. In the 47 species that do not avoid SD-like hexamers, the negative relationship between RSCU and occurrence in high-affinity hexamers was significant only for Glutamic acid (orange lines in fig. 4;  $P < 0.01$ —Fisher's  $z$  test for correlation coefficients). In stark contrast, in the three species that are enriched in internal SD-like hexamers, the correlation between RSCU and occurrence in SD-like hexamers was positive for Arginine and Glutamic acid (green lines in fig. 4). Finally, in the seven species that do not use the SD mechanism, we found no relationship between RSCU and codon occurrence in SD-like hexamers for Glycine and Arginine (blue points and lines in fig. 4), and a significantly negative relationship for Glutamic acid. Overall, genomes that show patterns consistent with selection against SD-like hexamers also show an associated and consistent decrease in the relative use of codons that compose these hexamers, as suggested by Li et al. (2012). However, it is important to note that we cannot gauge causality from these correlations, or infer the impact of selection on internal SD-like motifs on codons that do not typically compose SD-like hexamers.

### SD-like Hexamers Are Enriched in Specific Gene Regions

Next, we determined the relative location of SD-like hexamers in genes from each species. To normalize for gene length, we divided each gene into 20 parts of equal length and determined the distribution of SD-like hexamers in each part. We observed a stark depletion of SD-like hexamers in the first 5% of the gene length (N-terminus) for all genomes except those that do not use the SD mechanism of translation initiation (fig. 5A). This consistent pattern suggested that N-terminal SD-like hexamers may be especially deleterious. A potential mechanism underlying this pattern may be that early ribosome pauses during translation have a greater probability of causing ribosomal traffic jams (Tuller et al. 2010). However, we could not directly test this hypothesis given the paucity of ribosomal profiling data for most species in our dataset. Alternatively, the presence of N-terminal SD-like hexamers may be associated with strong RNA secondary structure that reduces protein expression levels (Kudla et al. 2009; Allert et al. 2010; Goodman et al. 2013). Indeed, we found that genes containing an N-terminal SD-like hexamer have lower MFE, that is,



**Fig. 4.**—RSCU of (A) Glycine, (B) Arginine, and (C) Glutamic acid codons as a function of the relative occurrence of the codon in SD-like hexamers (aff<sub>5</sub> to highest). Black points and regression lines indicate genomes that avoid internal SD-like hexamers, orange points, and regression lines indicate genomes that do not avoid internal SD-like hexamers, green points, and regression lines indicate genomes that are enriched in internal SD-like hexamers (from fig. 2), and blue points and regression lines indicate genomes that do not use the SD mechanism of translation initiation. Spearman's rho and significance values for each category are indicated in each panel in the same color. Slopes of the black and orange lines in the first two panels were significantly different from each other (Fisher's z test for correlation coefficients,  $P < 0.01$  in each case).

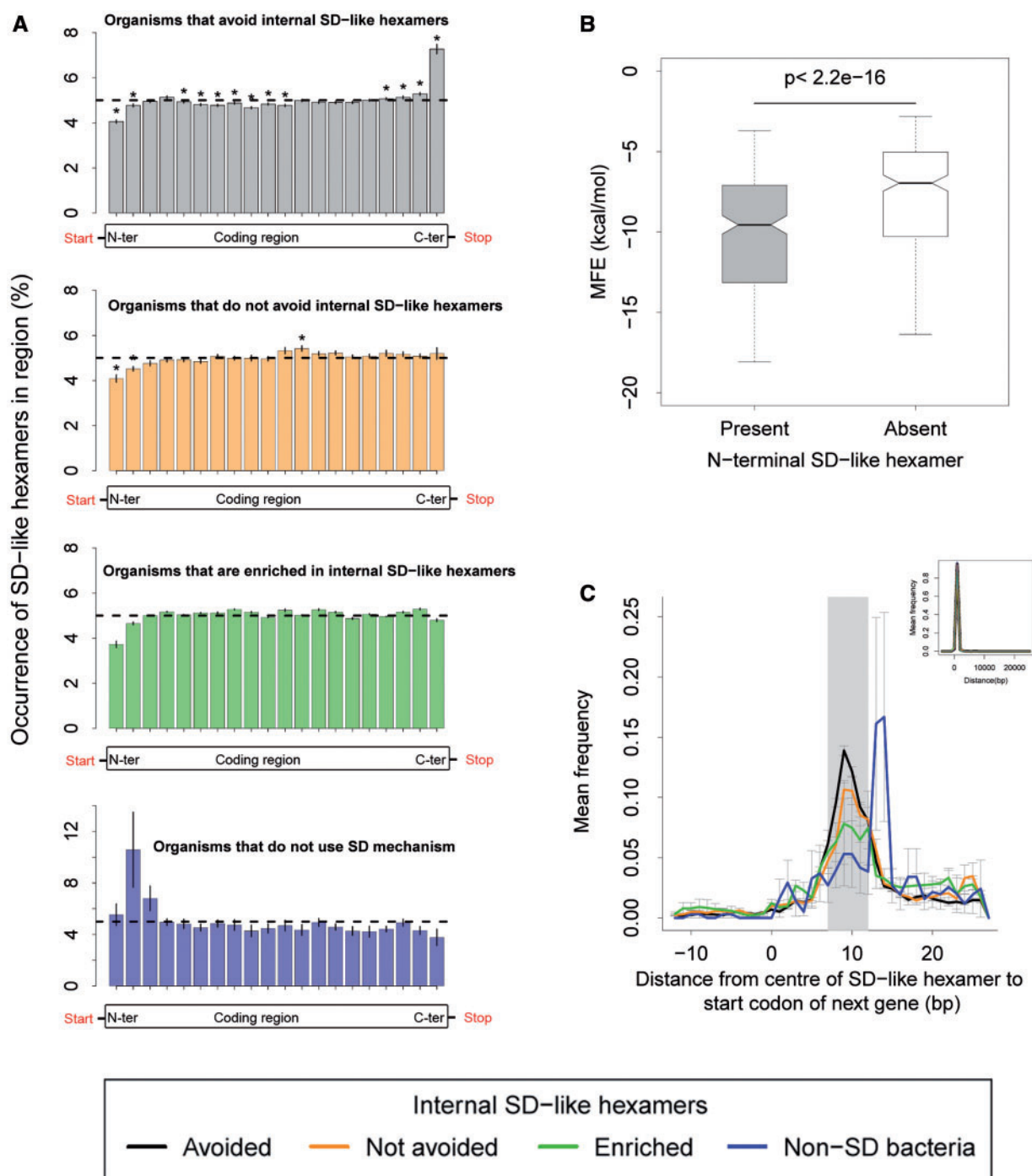
stronger secondary structure, compared to genes without an N-terminal hexamer (fig. 5B;  $P < 2.2e-16$ , Wilcoxon rank sum test). Thus, selection for reduced N-terminal secondary structure may lead to the apparent depletion of internal SD-like hexamers at the N-terminal end of coding regions.

In contrast to the N-terminal depletion of SD-like hexamers, we observed a relative enrichment of SD-like hexamers at the C-terminal ends of genes in genomes that avoid internal SD-like hexamers (fig. 5A). Thus, in these genomes, SD-like hexamers seem to be concentrated towards the C-terminal ends of genes. Note that genomes that do not avoid or those enriched in SD-like hexamers generally have more SD-like hexamers (supplementary fig. S12, Supplementary Material online;  $P = 0.001$  and  $P = 0.006$ , respectively, Wilcoxon rank sum test). Hence, despite a lack of relative enrichment at the C-terminal end, these genomes may still have many C-terminal SD-like hexamers. A previous study with *H. volcanii* suggested that C-terminal SD-like hexamers may serve as translation initiation sites for downstream genes (Kramer et al. 2014). Indeed, we observed that the start codons of downstream genes were 7–12 bp away from the middle of the C-terminal SD-like hexamer across all organisms, except in species that do not use the SD mechanism for translation initiation (fig. 5C). The 7–12 bp range matches the distance between the ribosome binding site and the start codon in most *E. coli* genes (Osterman et al. 2013), suggesting that C-terminal internal SD-like hexamers are probably used as translation initiation sites for the downstream gene. This was further supported by our finding that a higher proportion of genes with a C-terminal SD-like hexamer are succeeded by overlapping genes (i.e., genes whose start codon occurred before the stop codon of the previous gene), as opposed to genes

without a C-terminal SD-like hexamer (supplementary fig. S13A, Supplementary Material online;  $p < 2.2e-16$ , Wilcoxon rank sum test). Genes with a C-terminal SD-like hexamer are also more likely to be part of an operon (supplementary fig. S13B, Supplementary Material online;  $P < 2.2e-16$ ; Wilcoxon rank sum test). Together, these data suggest that C-terminal ends of genes may be enriched in SD-like hexamers as a result of the compact organization of prokaryotic genomes.

Lastly, we determined the relative location of SD-like hexamers in genes belonging to different COG categories (supplementary fig. S14, Supplementary Material online). We recovered the pattern of C-terminal enrichment of internal SD-like hexamers in all categories except A (RNA processing and modification) and B (chromatin structure and dynamics). We also observed non-terminal enrichment of SD-like hexamers in genes from four COG categories: A, B, Z (cytoskeleton) and W (extracellular structure). Interestingly, genes in these COG categories generally had fewer high-affinity hexamers compared to the mean frequency across categories (significantly lower for Z and W,  $P < 0.05$ , one-sample Wilcoxon rank sum test corrected for multiple comparisons using the Benjamini–Hochberg method, supplementary fig. S15, Supplementary Material online), suggesting strong avoidance of SD-like hexamers in genes relevant for these functions. However, it is not clear why internal SD-like motifs may be particularly deleterious in genes with these specific functions, and why these genes show atypical patterns of internal enrichment in SD-like motifs. We also note that the number of organisms with genes in these categories was low (25–130), and it is possible that improved annotation will alter these results.





**Fig. 5.**—Functional and positional analysis of the occurrence of internal SD-like hexamers (aff<sub>5</sub>—highest affinity). (A) The distribution of SD-like hexamer occurrence (mean  $\pm$  SEM) across the relative length of all genes in 284 organisms, for genomes that avoid, do not avoid, are enriched in internal SD-like hexamers and organisms that do not use the SD mechanism of translation initiation. Dashed lines indicate the random expectation for occurrence in each region (gene length/20 regions = 5%). In each panel, asterisks indicate gene regions where the occurrence was significantly different from this expectation (one-sample Wilcoxon signed rank test with Benjamini–Hochberg correction for multiple comparisons,  $P < 0.05$ ) (B) N-terminal MFE of genes with an N-terminal SD-like hexamer (grey bars) versus those without (white bars) for 277 organisms that use the SD mechanism of translation initiation. The two distributions were significantly different from each other (paired Wilcoxon test,  $P < 2.2e-16$ ). (C) The mean frequency distribution ( $\pm$ SEM) of the distance from the centre of a C-terminal SD-like hexamer to the start codon of the next gene (inset plot shows the entire distribution). The grey rectangle indicates the most common distances between the ribosome binding site and the start codon of genes in *E. coli* (7–12 bp).

## Discussion

Ever since its discovery in 1975, the SD sequence has dominated our thinking of how prokaryotes regulate translation initiation—the first step in the important process of protein production. While the SD::anti-SD interaction is thought to be critical for translation initiation across prokaryotes, it is also clear that misplaced SD-like sequences in an mRNA can be deleterious. In fact, a previous analysis of a few bacterial genomes suggested that selection against such internal SD-like motifs is a major evolutionary force acting on bacterial genomes (Li et al. 2012). However, our analysis of 284 prokaryotic genomes suggests that this may be an overly simplified view of the evolution of SD-like motifs. We show that internal SD-like hexamers are not universally avoided in prokaryotes; conversely, selection against internal SD-like motifs is not the only factor that determines the frequency of SD-like hexamers. Although our results are largely consistent with selection against internal SD-like hexamers, we identified 50 species (~17% of analyzed) in which SD-like hexamers are either not avoided, or are significantly enriched relative to the null expectation based on the GC content of the organism. Some species that do not use the SD mechanism of translation initiation are also significantly depleted in SD-like hexamers, indicating that factors other than selection against internal SD-like motifs are also important drivers of their frequency. For instance, we found that N-terminal SD-like hexamers are more likely to confer stronger mRNA secondary structure, which is a major determinant of protein levels (Goodman et al. 2013). Hence, selection favoring weaker 5' mRNA structure could indirectly lead to selection against SD-like motifs, or *vice versa*. Finally, our results suggest that selection favoring compact genome organization may also favor SD-like motifs at the C terminal ends of genes, resulting in highly variable selection on these motifs across genes.

This study also highlights the importance of accounting for genome GC content while analyzing selection on prokaryotic genomes. For instance, we observed that mean HF was positively correlated with genome GC% (supplementary fig. S5, Supplementary Material online), which would imply GC-rich organisms are enriched in internal SD-like hexamers. When we accounted for genome GC we observed the opposite relationship, so that the corrected HF was negatively correlated with genome GC (fig. 2). However, genome GC% was not correlated with proxies of translational selection (e.g., growth rate described in Vieira-Silva and Rocha 2010 or the strength of codon usage bias as calculated by Sharp 2005; data not shown). Hence, the apparent depletion of internal SD-like motifs in GC-rich genomes may arise simply because they are more likely to have G-rich SD-like motifs. Thus, we suggest that several gene- and genome-level features of prokaryotes are associated with SD-like motifs, and may have important impacts on their evolution.

Recent experiments support these conclusions, demonstrating that the presence of SD-like hexamers does not

always impact protein production or fitness as expected. For instance, *in vitro* translation assays of a synthetic gene showed that neither the rate of ribosome translocation in the middle of an a gene nor the protein production rate were affected by a large change in the anti-SD affinity (from 0 to  $-6$  kcal/mol) of an internal hexamer (Borg and Ehrenberg 2015). In another study, we generated a panel of *Methylobacterium extorquens* strains carrying internal point mutations in a key enzyme-coding gene. The mutations were predicted to substantially alter affinity to the anti-SD sequence. However, anti-SD affinity was not correlated with strain fitness, suggesting that selection against internal SD-like motifs is neither consistently strong nor predictable (Agashe et al. 2016). Lastly, Mohammad et al. (2016) showed that ribosomes do not necessarily pause at internal SD-like motifs in *E. coli*; hence, such motifs may have relatively weak impacts on global translation rates. This report thus questions the hypothesis of translational pausing due to internal SD-like sequences. However, it is possible that internal SD-like hexamers occur only in specific locations and excessive occurrence of such hexamers is deleterious. Indeed, the significant depletion in internal SD-like hexamers in most species that we analyzed suggests significant selection against such motifs. Additional ribosome profiling studies on different bacterial species may shed light on the exact role of SD-like hexamers in various contexts.

Specifically, the 50 species in our analysis that are not depleted in SD-like sequences are good candidates for ribosome profiling studies. These species do not belong to specific clades, and hence the lack of internal SD-like hexamers cannot be attributed to clade-specific characteristics or evolutionary history. A number of mechanisms may be responsible for weakened selection against internal SD motifs in these 50 species. Selection against SD-like hexamers may decrease if the SD::anti-SD interaction itself is weaker; for example, at high temperature. Indeed, we found that regardless of their GC content, organisms that do not avoid internal SD-like hexamers have higher optimal growth temperatures. Second, selection against internal SD-like motifs may diminish in organisms that do not rely strongly on the SD mechanism of translation initiation. For instance, Nakagawa et al. (2010) suggested that species that possess relatively few genes with a 5' SD motif (correcting for genome GC content) may instead rely on alternative mechanisms such as RPS1-mediated translation initiation. However, we found that lower dependence on SD-based initiation is not associated with stronger avoidance of internal SD-like hexamers (supplementary fig. S16, Supplementary Material online). Furthermore, four of seven species in our analysis that do not use the SD mechanism also avoid internal SD-like hexamers (filled triangles below black line in fig. 2). Together, these results suggests that the prevalent mechanism of translation initiation in different organisms is not correlated with selection to avoid internal SD-like sequences. Finally, selection against internal SD-like motifs may be weaker in slow-growing bacteria if protein

production rate is not limiting and translational selection is weak. In support of this hypothesis, we observed that organisms that do not avoid internal SD-like hexamers have lower growth rates (from Vieira-Silva and Rocha 2010) compared to organisms that do avoid internal SD-like hexamers (supplementary fig. S17, Supplementary Material online). However, it is unclear whether fast growth rate is a result of faster protein production associated with avoidance of internal SD-like hexamers, or *vice versa*. Thus, although it is possible that the 50 species that do not avoid internal SD-like hexamers evolve under relatively weak translational selection owing to their slow growth rates, we caution against assigning causality to the relationship between growth rate and avoidance of internal SD-like hexamers.

An alternative hypothesis is that these 50 genomes face strong positive selection for other genomic features that opposes purifying selection against internal SD-like hexamers. For instance, strong selection favoring specific codons may reduce the strength of selection against internal SD-like hexamers. We observed that species that avoid internal SD-like sequences also avoid codons that form SD-like hexamers, but the relationship is weak or absent in species that do not avoid internal SD-like hexamers. Although previous authors have interpreted this correlation to suggest that selection against SD-like hexamers drives codon use (Li et al. 2012), it is plausible that strong selection favoring specific codons could instead increase the number of SD-like hexamers in the latter set of species. Quantifying selection on genomic features associated with SD-like hexamers may thus shed light on the variability in the apparent strength of selection against SD-like motifs across prokaryotes.

In summary, our results call for a more nuanced view of selection acting on SD-like as well as other motifs in prokaryotes. Building on previous single-species experimental and bioinformatic work, our analysis of 284 prokaryotes undermines the universality of selection against SD-like sequences in coding regions of prokaryotes. Instead, we present various contexts in which SD-like hexamers may evolve effectively neutrally or under positive selection. However, to evaluate these roles, we need empirical information on the functional and evolutionary significance of internal SD-like motifs. Critical experiments include the introduction of SD-like hexamers at various locations within genes and testing their impact on ribosomal pausing, protein folding and protein targeting. Together, these analyses can lead to a deeper understanding of prokaryotic genome and protein evolution as well as the process of translation.

## Supplementary Material

Supplementary methods, table S1, and figures S1–S17 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

We thank Aswin Seshasayee, Szabolcs Semsey, and members of the Agashe lab for discussion; Aswin Seshasayee for allowing us to use his lab server space; and Aalap Mogre and Supriya Khedkar for help with R scripts and databases. This study was supported by the National Centre for Biological Sciences (NCBS), a fellowship from the University Grants Commission (GDD), and an INSPIRE Faculty fellowship (DA) from the Department of Science and Technology (Grant number IFA-13 LSBM-64).

## Literature Cited

- Accetto T, Avguštin G. 2011. Inability of *Prevotella bryantii* to form a functional Shine–Dalgarno interaction reflects unique evolution of ribosome binding sites in bacteroidetes. *PLoS One* 6: e22914.
- Agashe D, et al. 2016. Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. *Mol Biol Evol.* msw035. doi: 10.1093/molbev/msw035.
- Allert M, Cox JC, Hellinga HW. 2010. Multifactorial determinants of protein expression in prokaryotic open reading frames. *J Mol Biol.* 402:905–918.
- Borg A, Ehrenberg M. 2015. Determinants of the rate of mRNA translocation in bacterial protein synthesis. *J Mol Biol.* 427:1835–1847.
- Charif D, Thioulouse J, Lobry JR, Perrière G. 2005. Online synonymous codon usage analyses with the *ade4* and *seqinR* packages. *Bioinformatics* 21:545–547.
- Chen J, et al. 2014. Dynamic pathways of –1 translational frameshifting. *Nature* 512:328–332.
- Del Campo C, Bartholomäus A, Fedyunin I, Ignatova Z. 2015. Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function. *PLoS Genet.* 11:e1005613.
- Fluman N, Navon S, Bibi E, Pilpel Y. 2014. mRNA-programmed translation pauses in the targeting of *E. coli* membrane proteins. *Elife* 3:e03440.
- Goodman DB, Church GM, Kosuri S. 2013. Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342:475–479.
- Gorochowski TE, Ignatova Z, Bovenberg RAL, Roubos JA. 2015. Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* 43:3022–3032.
- Katz L, Burge CB. 2003. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.* 13:2042–2051.
- Kramer P, Gäbel K, Pfeiffer F, Soppa J. 2014. *Haloflex volcanii*, a prokaryotic species that does not use the Shine Dalgarno mechanism for translation initiation at 5'-UTRs. *PLoS One* 9:e94979.
- Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324:255–258.
- Larsen B, Wills NM, Gesteland RF, Atkins JF. 1994. rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. *J Bacteriol.* 176:6842–6851.
- Li G-W, Oh E, Weissman JS. 2012. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484:538–541.
- Lim K, Furuta Y, Kobayashi I. 2012. Large variations in bacterial ribosomal RNA genes. *Mol Biol Evol.* 29:2937–2948.
- Lobry JR, Necşulea A. 2006. Synonymous codon usage and its potential link with optimal growth temperature in prokaryotes. *Gene* 385:128–136.
- Lorenz R, et al. 2011. ViennaRNA package 2.0. *Algorithms Mol Biol.* 6:26.

- Mohammad F, Woolstenhulme CJ, Green R, Buskirk AR. 2016. Clarifying the translational pausing landscape in bacteria by ribosome profiling. *Cell Rep.* 14:686–694.
- Moreno-Hagelsieb G, Collado-Vides J. 2002. A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics* 18(Suppl 1):S329–S336.
- Nakagawa S, Niimura Y, Miura K.-I, Gojobori T. 2010. Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci.* 107:6382–6387.
- Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA. 2013. Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.* 41:474–486.
- Ponnala L. 2010. A plausible role for the presence of internal Shine–Dalgarno sites. *Bioinform Biol Insights* 4:55–60.
- R Development Core Team. 2015. R: a language and environment for statistical computing. doi: 10.1038/sj.hdy.6800737.
- Sabi R, Tuller T. 2015. A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics* 16:S5.
- Sharp PM. 2005. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33:1141–1153.
- Shine J, Dalgarno L. 1975a. Determinant of cistron specificity in bacterial ribosomes. *Nature* 254:34–38.
- Shine J, Dalgarno L. 1975b. Terminal-sequence analysis of bacterial ribosomal RNA. Correlation between the 3'-terminal-polypyrimidine sequence of 16-S RNA and translational specificity of the ribosome. *Eur J Biochem.* 57:221–230.
- Simonetti A, et al. 2009. A structural view of translation initiation in bacteria. *Cell Mol Life Sci.* 66:423–436.
- Sørensen MA, Kurland CG, Pedersen S. 1989. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol.* 207:365–377.
- Tuller T, et al. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141:344–354.
- Varenne S, Buc J, Lloubes R, Lazdunski C. 1984. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol.* 180:549–576.
- Vasquez KA, Hatridge TA, Curtis NC, Contreras LM. 2015. Slowing translation between protein domains by increasing affinity between mRNAs and the ribosomal anti-Shine–Dalgarno sequence improves solubility. *ACS Synth Biol.* 5: 133–145.
- Vieira-Silva S, Rocha EPC. 2010. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet.* 6:e1000808.
- Weiss RB, Dunn DM, Dahlberg AE, Atkins JF, Gesteland RF. 1988. Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *Embo J.* 7:1503–1507.
- Wen J-D, et al. 2008. Following translation by single ribosomes one codon at a time. *Nature* 452:598–603.

Associate editor: Mar Alba