



Article

How Many Messenger RNAs Can Be Translated by the START Mechanism?

Laurence Despons and Franck Martin *

Institut de Biologie Moléculaire et Cellulaire, “Architecture et Réactivité de l’ARN” CNRS UPR9002, Université de Strasbourg, 2, allée Konrad Roentgen, F-67084 Strasbourg, France; despons@unistra.fr

* Correspondence: f.martin@ibmc-cnrs.unistra.fr

Received: 29 September 2020; Accepted: 5 November 2020; Published: 8 November 2020



Abstract: Translation initiation is a key step in the protein synthesis stage of the gene expression pathway of all living cells. In this important process, ribosomes have to accurately find the AUG start codon in order to ensure the integrity of the proteome. “Structure Assisted RNA Translation”, or “START”, has been proposed to use stable secondary structures located in the coding sequence to augment start site selection by steric hindrance of the progression of pre-initiation complex on messenger RNA. This implies that such structures have to be located downstream and at an optimal distance from the AUG start codon (i.e., downstream nucleotide +16). In order to assess the importance of the START mechanism in the overall mRNA translation process, we developed a bioinformatic tool to screen coding sequences for such stable structures in a 50 nucleotide-long window spanning the nucleotides from +16 to +65. We screened eight bacterial genomes and six eukaryotic genomes. We found stable structures in 0.6–2.5% of eukaryotic coding sequences. Among these, approximately half of them were structures predicted to form G-quadruplex structures. In humans, we selected 747 structures. In bacteria, the coding sequences from Gram-positive bacteria contained 2.6–4.2% stable structures, whereas the structures were less abundant in Gram-negative bacteria (0.2–2.7%). In contrast to eukaryotes, putative G-quadruplex structures are very rare in the coding sequence of bacteria. Altogether, our study reveals that the START mechanism seems to be an ancient strategy to facilitate the start codon recognition that is used in different kingdoms of life.

Keywords: ribosome; translation initiation; START; secondary structures; mRNA

1. Introduction

In all living organisms, translation is a universal mechanism that enables the biosynthesis of proteins according to the nucleic acid-encoded genetic information. The platform that allows the sequential assembly of amino acids is the ribosome, a macromolecular machine found in all three kingdoms of life. Translation consists of four steps: translation initiation, elongation, termination and ribosome recycling. Among these, the rate-limiting step is, most of the time, translation initiation and is therefore considered as the master regulator step, although there are examples where the elongation rate can dominate. In this initiation phase, the ribosome has to accurately localize the start codon, most frequently encoded by an AUG triplet. Importantly, the cell integrity depends directly on the efficiency and the fidelity of this critical step. Therefore, living organisms have evolved sophisticated molecular mechanisms that ensure proper translation initiation in order to avoid aberrant protein synthesis. These mechanisms have been extensively studied in the last three decades in prokaryotes [1,2], in eukaryotes [3–10] and in archaea [11]. Interestingly, rapid cellular adaptation is achieved essentially by post-transcriptional events. Among these, localization of the start codon is a critical step that is not only essential for cell maintenance, it is also essential to modify the proteome. For instance, in prokaryotes, the proteome is changed in order to adapt

quickly to new environmental conditions. In eukaryotes, the translational landscape can be drastically modified to allow fast and efficient stress response to various stimuli (infection, nutrient deprivation, developmental stages, etc.). Localization of the AUG start codon can be a challenging task due to various obstacles. For instance, mRNAs have the propensity to fold and form stable secondary structures such as stem-loops [12], tertiary structures [13,14] or more sophisticated structures termed G-quadruplexes [15]. Translation initiation efficiency is greatly influenced by the presence of secondary structures in the 5'UTR. In addition, it is well established that structures directly at or in close proximity to a start codon are inhibitory because they prevent efficient ribosome loading [16]. However, in the case of internal ribosome entry sites (IRES), structures located close to the start codons are usually crucial for efficient translation [17].

In prokaryotes, the ribosome binding site (RBS) that encompasses the "Shine-Dalgarno" (SD) sequence must be accessible and not involved in structures to allow efficient translation initiation. During initiation, the 3' end of the 16S ribosomal RNA anneals with the SD sequence on the mRNA in order to ensure efficient assembly of the pre-initiation complex. Indeed, hindering of the RBS in a structure is an efficient way to down-regulate translation initiation. In fact, these structures are considered as genuine switches that can turn on translation by unwinding them, thereby liberating the RBS for efficient translation initiation. Many examples of such structures in the RBS leading to SD sequence hindering have been described [18–20]. The largest acidic ribosomal protein S1 is required for efficient unfolding of RNA duplexes hindering the SD sequence [21,22]. The folding/unfolding of these structures is also controlled by various mechanisms including *cis*-acting elements such as riboswitches that can sense several biophysical parameters like temperature, pH or the presence of specific ligands or metabolites or regulatory RNAs or proteins [23]. Indeed, translation initiation requires the anchoring of the ribosomal RNA on the SD sequence. Nowadays this point is very well documented ([1] and references therein). Nevertheless, a recent genome-wide analysis revealed that a large fraction (15–100%) of prokaryotic transcripts is translated by another mechanism that is SD-independent. These mRNAs have no 5'UTR (leaderless mRNAs) or do not contain any SD-like sequences [24]. These SD-independent translation events occur in bacterial and organellar genomes. Moreover, in the entire Gram-negative bacterial phylum Bacteroidetes, the ribosome does not use SD interactions to initiate translation [25]. Apart from the fact that the 5'UTR of these mRNAs is usually unfolded, the molecular basis of such an SD-independent translation initiation mechanism remains currently elusive [24].

In eukaryotes, the 5'UTRs are generally much longer and consequently more likely to harbor stable secondary structures. The median length of 5'UTRs in eukaryotes ranges from 53 nucleotides in yeast to 218 nucleotides in humans [26–30]. First, an mRNA activation step triggers the assembly of the eIF4F complex on the 5' cap. This complex contains eIF4E, the cap binding protein, the RNA helicase eIF4A and its auxiliary factor eIF4B and the platform protein eIF4G. Then, the assembly of the translation machinery occurs at the 5' cap, and a so-called scanning step enables the 5'-3' sliding of the 43S particle (which contains the 40S ribosomal subunit, the multi subunit factor eIF3, the scanning factors eIF1, 1A and 5 and the so-called ternary complex (comprising eIF2 and the initiator tRNA^{Met})) in order to localize the start codon. This scanning step has been extensively described by detailed mechanistic and functional studies [3,31] but has been effectively observed only very recently [32,33]. The presence of these stable structures in both prokaryotes and eukaryotes requires a specific mechanism in order to enable proper access to the AUG start codon. Specific RNA helicases that can unwind such structures are found in prokaryotes [34,35] and in eukaryotes [36].

Stable secondary structures can also be found downstream of the start codon in the coding sequence. These sequences, when they are located at the proper distance from the start codon, can influence translation initiation efficiency both in prokaryotes [37] and in eukaryotes [38]. High-throughput methods to find secondary structures in early coding sequences have been conducted [39–42]. These studies converged on the finding that the region around the start codon of a great majority of mRNAs is, most of the time, poorly structured.

Here we performed genome-wide analyses of the beginning of the coding sequences in order to identify putative secondary structures that might influence translation initiation efficiency. Indeed, the presence of stable structures, when located downstream at the optimal distance from the AUG start codon, can greatly facilitate translation initiation by a novel mechanism that we named the Structure Assisted RNA Translation (START) mechanism [38]. More precisely, the START mechanism relies on the presence of stable secondary structures located in the coding region. These structures are located at the appropriate distance from the start codon (+16 to +19) that enables the assembly of the pre-initiation complex on the AUG start codon without unfolding of the structures. In fact, these structures at this position slow down, by steric hindrance, the progression of the scanning pre-initiation and thereby augment the initiation efficiency. Thus, their precise position downstream +16 to +19 is crucial for efficient localization of the start codon.

In order to evaluate the importance of this START mechanism in the cell, we underwent a genome-wide screening of secondary structures located in the coding sequence downstream of the AUG. We screened six eukaryotic genomes and eight prokaryotic genomes.

2. Results and Discussion

During protein biosynthesis, the ribosome slides 5' toward 3' on the mRNA and orchestrates the sequential addition of amino acids by recruiting the cognate aminoacyl-tRNAs accordingly to the codon sequence. Various biochemical and structural studies have shown that the fully assembled ribosome covers a region encompassing approximately 25 to 30 nucleotides of the messenger RNA [43,44] (Figure 1A). In the initiation step, the ribosome is positioned with the AUG codon in its P-site. Traditionally, the initiating ribosome is located on the mRNA by the so-called "toe-printing" experiment. The principle of this method is to map the position of the ribosome on the mRNA in the initiation phase, thanks to a primer extension assay that leads to a reverse transcriptase arrest, called a "toe-print", induced by the presence of the ribosome. Indeed, the position of the "toe-print" corresponds to the 3' edge of the ribosome on the mRNA. By convention, the A of the AUG start codon is numbered +1. During translation initiation, the ribosome is positioned on the AUG start codon in its P-site. Toe-printing experiments performed by many researchers over the last three decades have shown that when the ribosome is positioned on the AUG, the typical "toe-print" is usually detected at position +15 to +17 (Figure 1A). Recently, we have described a novel translation mechanism that we named "Structure Assisted RNA Translation", or START [38]. The mRNAs that are translated by this mechanism harbor a stable secondary structure that is located downstream nucleotide +19. By steric hindrance, these structures assist the ribosome to accurately localize the AUG start codon and therefore improve the translation efficiency of these mRNAs [38] (and references therein). In order to evaluate how many mRNAs are translated by the START mechanism, we performed genome-wide screenings for the presence of putative secondary structures located in this critical area on six eukaryotic organisms and eight prokaryotic organisms. The screenings were performed on the coding sequences, and we targeted our search in a window of 50 nucleotides. We focused our efforts in order to find structures located at the optimal distance required for the START mechanism, precisely from +16 to +65 (Figure 1B). The strategy was to detect putative secondary structures by calculating the minimal free energy (MFE) of the folded structure. In order to select significant structures, we proceeded as follows. In the first step, we calculated the MFE of a putative structure (MFE_{struc}). Then in the second step, the corresponding selected sequence was randomized by mixing the nucleotides present in this 50 nt window. This randomization was repeated 100 times for each sequence, and we calculated the MFE of the resulting randomized sequence (MFE_{rand}). Only the structures with an $MFE_{\text{struc}}/MFE_{\text{rand}}$ ratio equal or higher than 2 were retained. An additional criterion for selecting the structures was used. We retained the top 25% of the structures, i.e., the structures that had an MFE_{struc} equal or lower than the value of the quantile 0.25. This value was calculated from the distribution of MFE_{struc} values of all structures of a given species. Among the selected structures, we also predicted the presence

of putative G-quadruplex structures. These G-quadruplex structures are formed by the stacking of so-called G-quartets composed by 8, 12, 16, 20 G or even more G residues (Figure 1C).

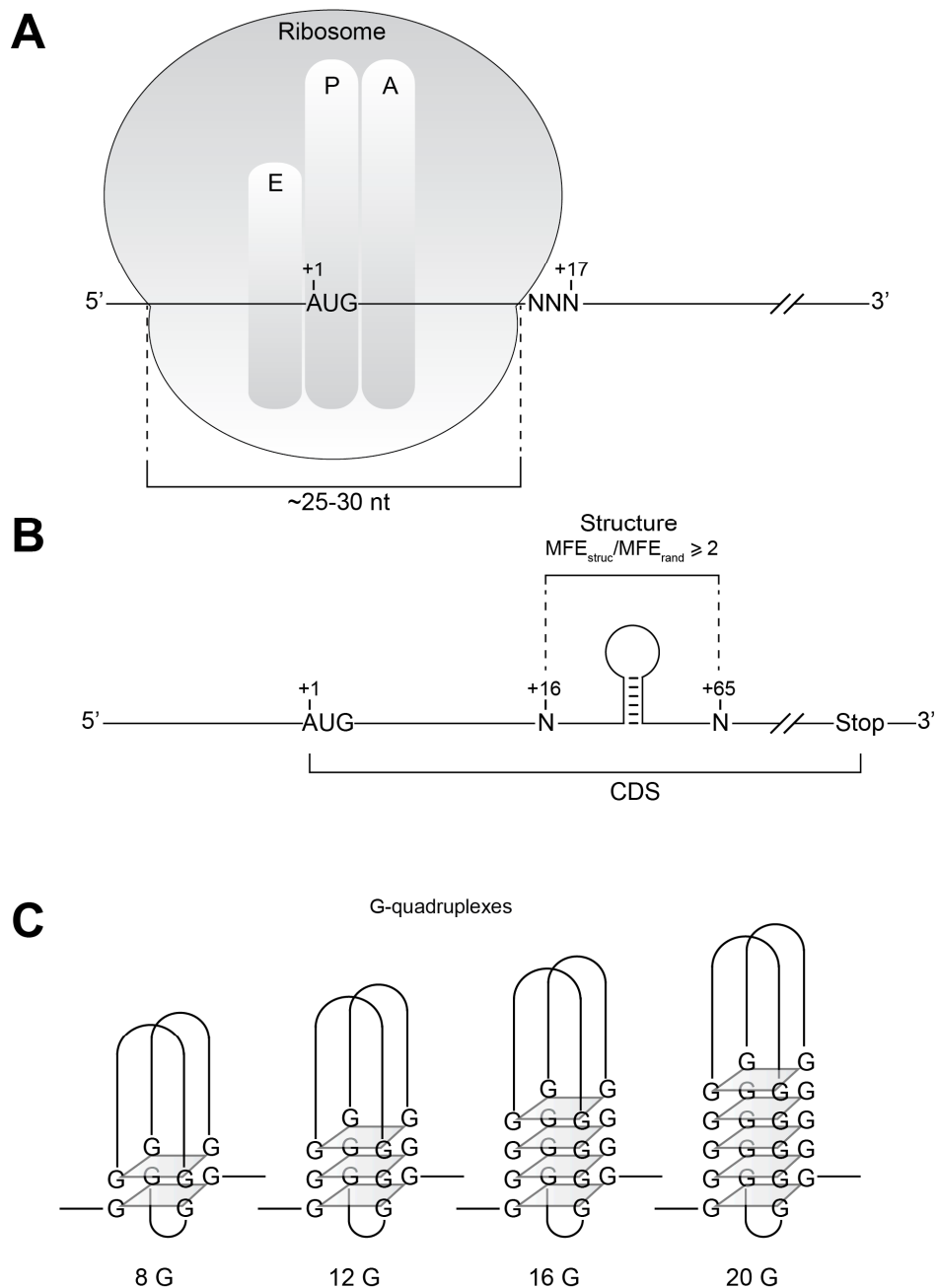


Figure 1. (A) During translation initiation, the assembled ribosome covers 15 to 17 nucleotides downstream from A (number +1 by convention) of the AUG start codon. In the Structure Assisted RNA Translation (START) mechanism, the AUG start codon is positioned in the P-site of the ribosome by stable secondary structures located immediately downstream of the +17 nucleotide. (B) In order to identify such putative new secondary structures, we screened the window +16 to +65 for putative secondary structures in the coding sequences (CDS). We focused on secondary structures with a minimal free energy (MFE_{struct}) that is at least two times higher than the mfe of the same sequence that was randomized 100 times (MFE_{rand}). (C) In addition, we also screened the same +16 to +65 window in CDS for G-quadruplex structures that contain 8, 12, 16 or 20 G-quartets.

In eukaryotic coding sequences, we found a significant number of structures (Supplementary Tables S1 and S2). For example, in humans 747 structures have been retained (Figure 2A). Among these, an important part is composed by G-quadruplexes, although the ratio is variable from one to another organism. Indeed, almost half of the detected structures are G-quadruplexes in humans and mice, whereas G-quadruplexes are rare in *Caenorhabditis elegans*. Remarkably, G-quadruplexes are almost absent in *Saccharomyces cerevisiae*. However, the proportion of mRNAs containing structures are higher in yeast reaching ~2.5% of the total coding sequences. This proportion is almost as high as in yeast in *C. elegans* (~2%) but is significantly lower in the human and mouse (less than 1%) (Figure 2B).

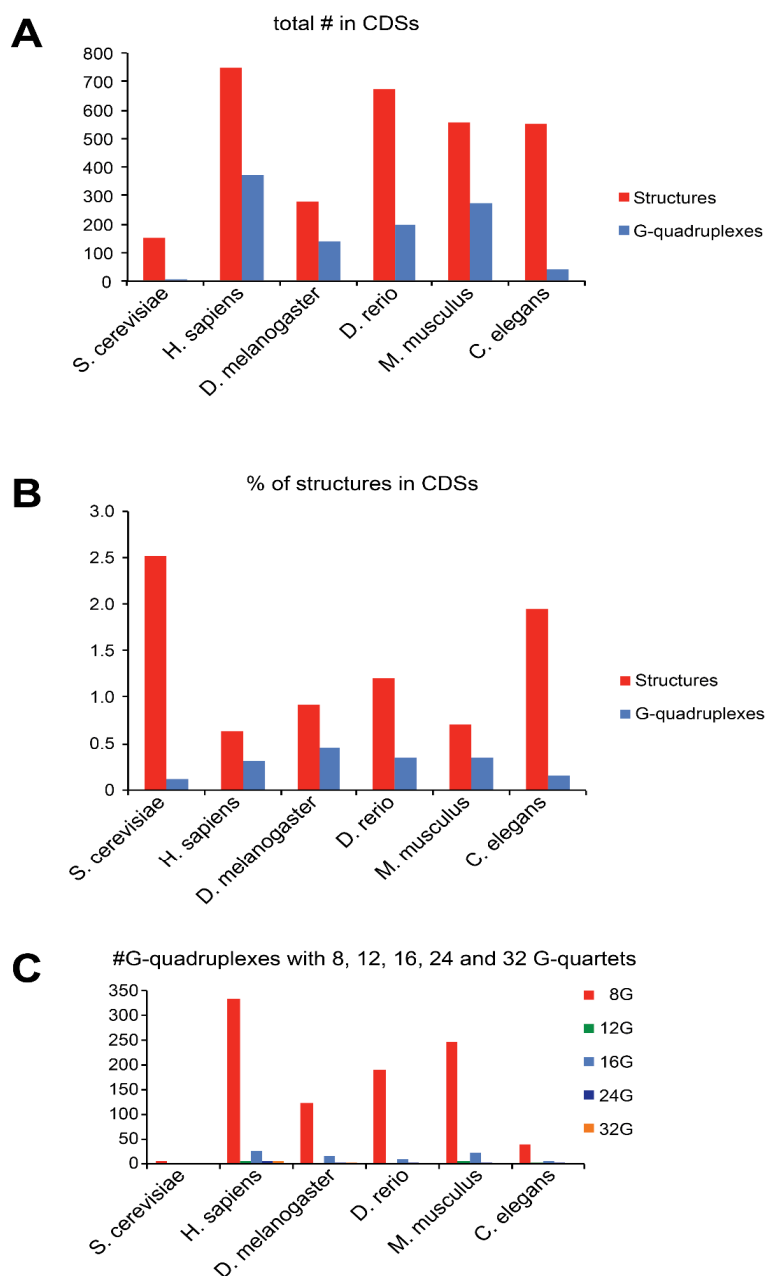


Figure 2. (A) The histogram represents the total number of putative secondary structures (red) and G-quadruplexes (blue) found in the +16 to +65 window of CDS that have a $MFE_{\text{struc}}/MFE_{\text{rand}}$ ratio of at least 2 in the CDSs from six eukaryotes. (B) The histogram represents the percentage of CDSs that contained secondary putative structures (red) and G-quadruplexes (blue). (C) The histogram represents the number of G-quadruplexes with 8, 12, 16, 24 and 32 G-residues in each organism.

We performed the same analysis on prokaryotic species. We screened the coding sequences from four Gram-positive bacteria and four Gram-negative bacteria (Supplementary Tables S1 and S3). In this case, the diversity of the result of our screening is much higher (Figure 3A). We found bacteria that contain many structures like *Bacillus anthracis* with 167 structures and bacteria that contain very few structures, such as *Pseudomonas aeruginosa* with only 10 structures. Concerning the type of structures, G-quadruplexes are rare in all the bacteria tested. However, 168 and 161 G-quadruplexes have been found in *Escherichia coli* and *P. aeruginosa*, respectively, mainly in the coding sequences [45], indicating that G-quadruplexes are also present in bacteria but very rarely in the +16 to +65 window. Altogether, our data seem to indicate that the appearance of such structures in this area of the coding sequence is rather a eukaryotic feature. In agreement with the observation that G-quartets are pretty common in the coding sequences of eukaryotes, specific RNA helicases dedicated to G-quadruplexes such as DHX36 and DHX9 have been found in eukaryotic genomes [46]. In addition, in humans G-quadruplexes are not always present because they are only transiently folded under specific conditions [47]. This suggests that the START mechanism would be activated only in specific conditions that would induce the formation of G-quadruplexes in the +16 to +65 window, thereby adding another layer of regulation in the translation process.

Interestingly, the proportion of structures selected is in general much lower in Gram-negative bacteria compared to Gram-positive bacteria, suggesting that START is more common in Gram-positive bacteria (Figure 3B). Translation initiation in Gram-positive bacteria is far less understood than in Gram-negative bacteria. Indeed, the presence, the nature and the use of the SD sequence are features that remain rather poorly studied in Gram-positive bacteria. Since the importance of the SD sequence is lower, the START mechanism might be another alternative used for translation initiation, which might explain that more structures are found by our screen in Gram-positive bacteria.

In prokaryotes, the transcription process has been shown to be coupled to translation. The time needed to fold a hairpin helix is in the millisecond range [48,49]. Translation initiation frequency by the ribosome is rather in the second range [50]. Therefore, secondary structures in the coding region have enough time to fold before translation initiation occurs, which indicates that the START mechanism is indeed possible even in prokaryotes when transcription and translation are coupled.

Next, we examined the possibility that the differences we observed in the different organisms were actually reflecting the GC content of the coding sequences from each genome. Thus, we wondered whether a high GC content in the coding sequences would be more prone to structures and favor the START mechanism. In the eukaryotes tested, we observed that the GC content was very similar, around 50% in all the genomes analyzed here, the lowest GC content being for yeast with 41.66% (Figure 4A). This is an interesting point because yeast is the eukaryotic organism that contains the highest proportion (2.5%) of structures in the beginning of its coding sequence, indicating that there is no correlation between the GC content and the number of selected secondary structures. By calculating the MFE for each structure, we can also evaluate the stability of these structures. We investigated the stability of the top 25% of the structures found in each organism (Figure 4B). The more stable structures were found in humans and mice, while the less stable were found in yeast and *C. elegans*. When looking at the stability of the selected structures, there was a perfect correlation with the GC content. This is rather expected since mice and humans also contain the highest numbers of G-quadruplexes. This suggests that G-quadruplexes are generally more stable than other types of secondary structures.

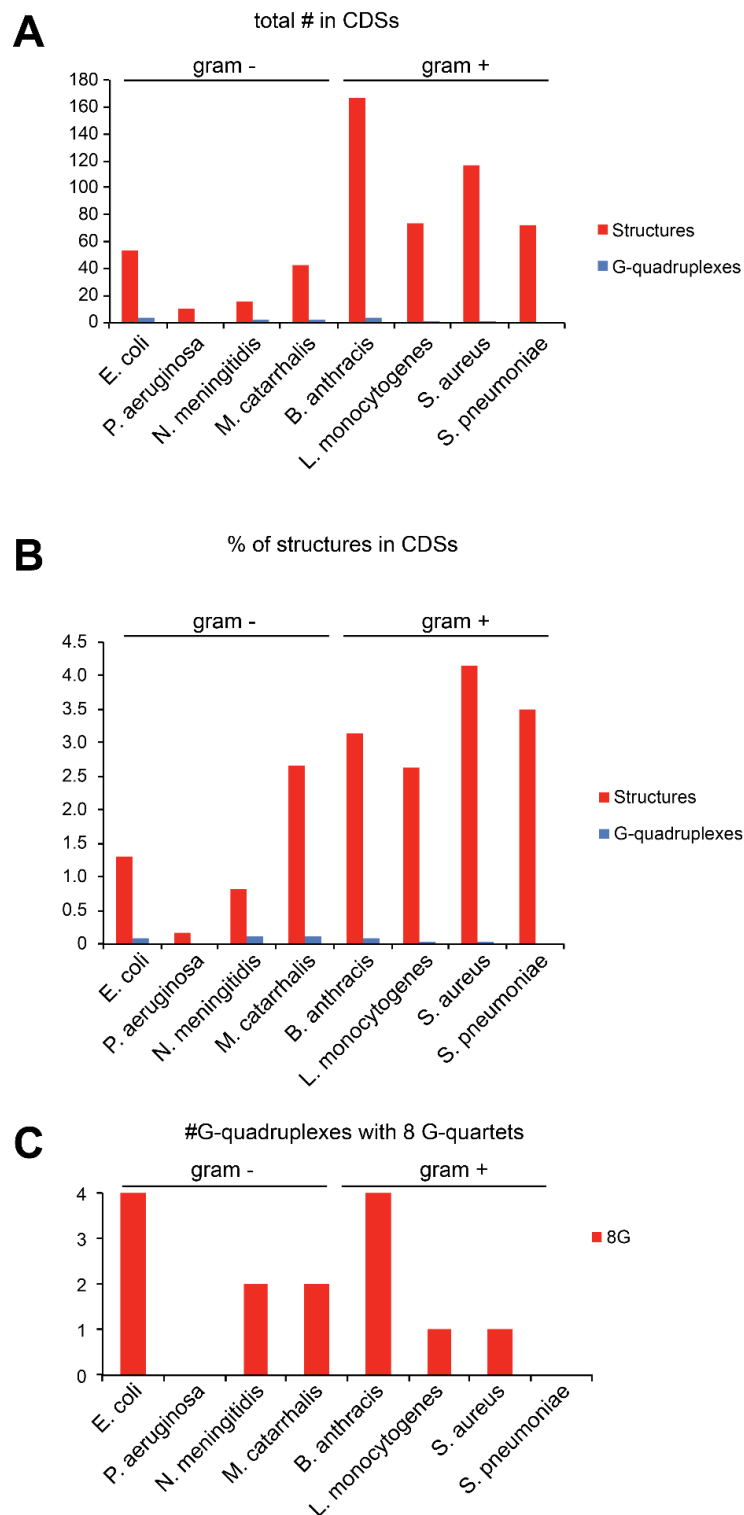


Figure 3. (A) The histogram represents the total number of putative secondary structures (red) and G-quadruplexes (blue) found in the +16 to +65 window of CDS that have a $MFE_{\text{struct}}/MFE_{\text{rand}}$ ratio of at least 2 in the CDSs of eight bacteria. Four Gram-negative bacteria are shown on the left and four Gram-positive are shown on the right. (B) The histogram represents the percentage of CDSs that contained secondary putative structures (red) and G-quadruplexes (blue). (C) The histogram represents the number of G-quadruplexes with 8 G residues in each organism.

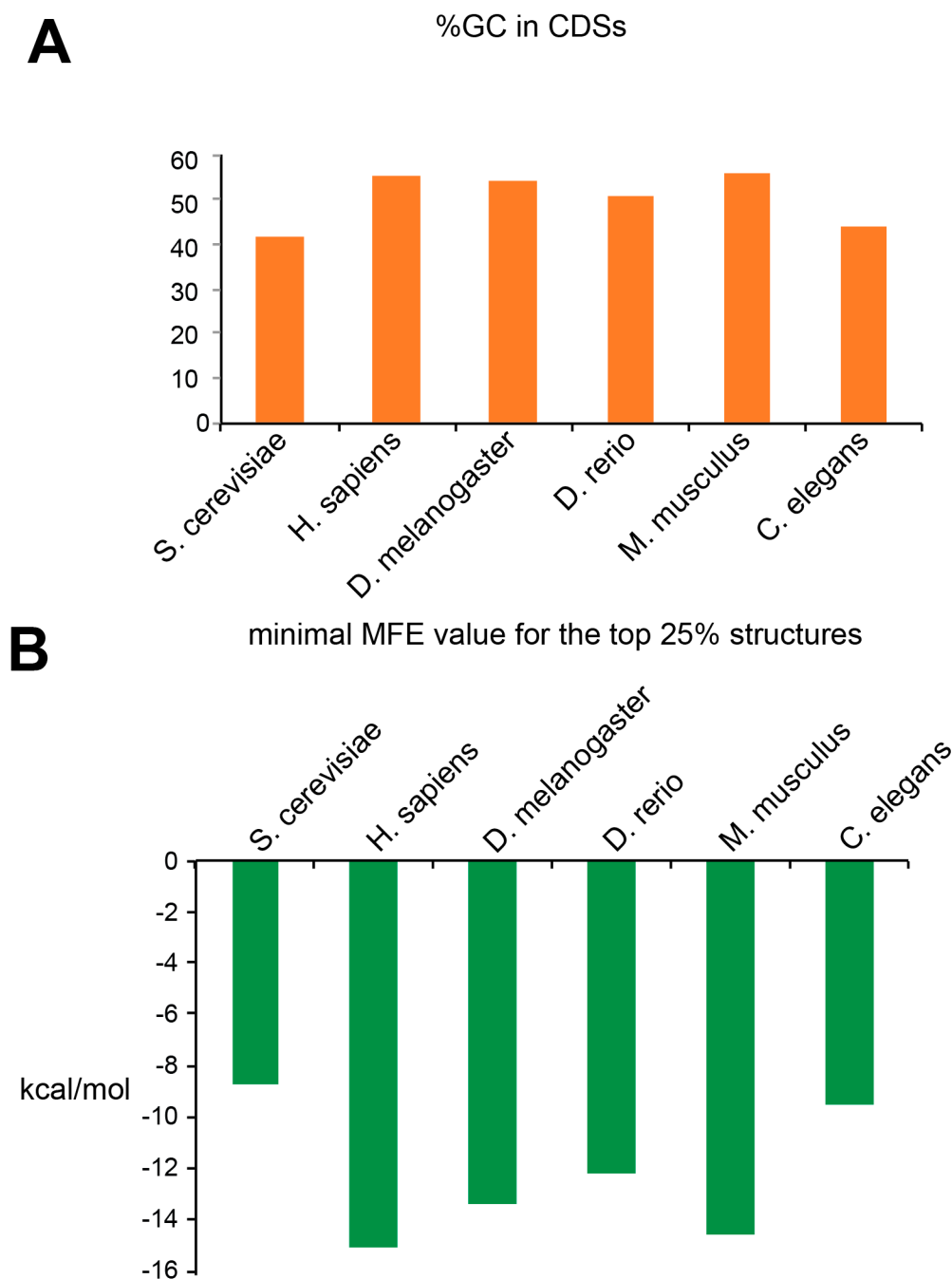


Figure 4. (A) The histogram represents the percentage of GC in the analyzed CDSs in eukaryotes. (B) The histogram represents the lowest minimal free energy of the top 25% of the structures found in the analyzed CDSs from eukaryotes.

In prokaryotes, the situation is significantly different. The GC contents of Gram-negative bacteria were generally higher than those in Gram-positive bacteria, the extreme example being *P. aeruginosa* (65.41%) (Figure 5A). Like in eukaryotes, there was a perfect correlation between the GC content and the stability of the top 25% of the structures (Figure 5B). Nevertheless, the Gram-negative bacteria do contain fewer structures than Gram-positive bacteria, as previously mentioned. This is particularly striking for *P. aeruginosa*, the bacteria with the highest GC content and the least structures selected (only 10 structures selected). Another interesting point is the presence of a lot of structures in Gram-positive bacteria, however with a medium stability (Figure 5B). In conclusion, the GC content is a good predictor for the presence of structures in eukaryotic coding sequences but not for prokaryotic

coding sequences. This is probably due to the high occurrence of G-quadruplexes, a eukaryotic feature as previously mentioned.

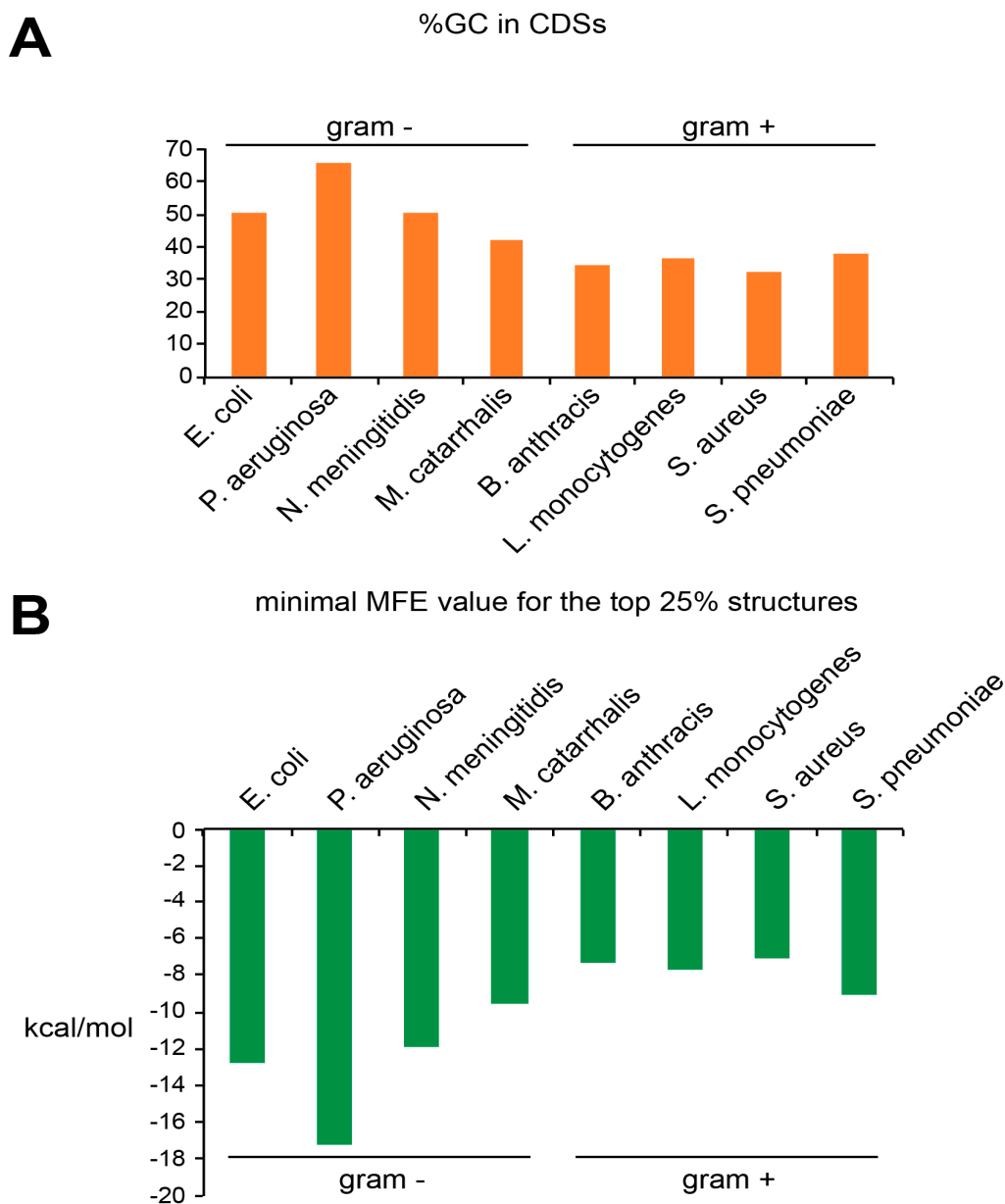


Figure 5. (A) The histogram represents the percentage of GC in the analyzed CDSs in Gram-negative and Gram-positive bacteria. (B) The histogram represents the lowest minimal free energy of the top 25% of the structures found the analyzed CDSs from Gram-negative and Gram-positive bacteria.

Next, we analyzed the types of mRNA that contained the selected structures by examining their GO terms (Supplementary Table S4). In eukaryotes, no specific biological process was enriched among the selected mRNAs. In contrast, in bacteria, the translation process was the first hit for all the bacteria tested here. Therefore, according to the structures that were selected, the START mechanism seems to be particularly important for translation regulation in bacteria. When looking at the functions of the selected mRNAs, the first hits for all the organisms tested were ATP binding or metal ion binding. Concerning the localization, the selected mRNA code, in general, for proteins were almost equally distributed in all the compartments of the cell.

Finally, we had a deeper look at the most stable structures that we found in human mRNAs (Figures 6 and 7, Supplementary Table S5). The stability of these structures ranged from -41.8 to -18.5 kcal/mol for classical structures (Figure 6). Structures containing G-quadruplexes reached the highest stability of -72 kcal/mol (Figure 7).

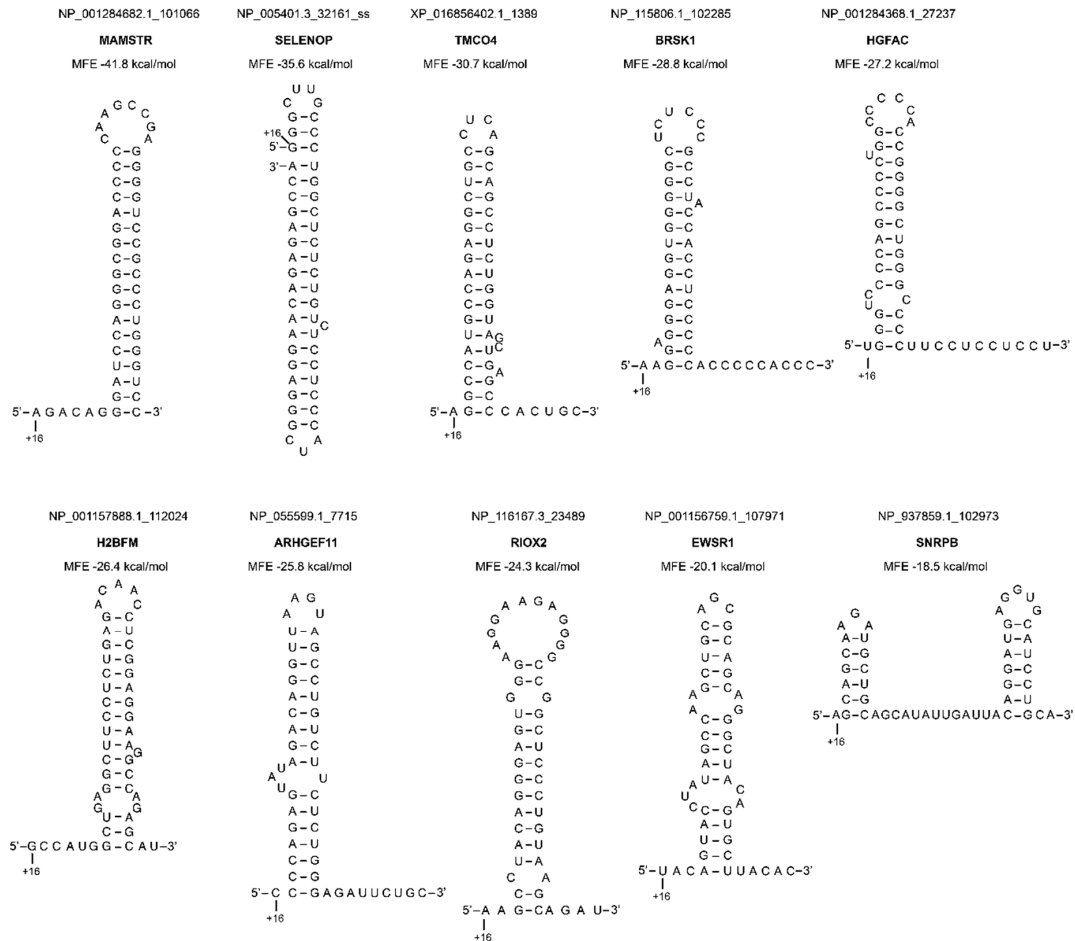


Figure 6. Representation of the ten best secondary structures selected in the human genome. The position of nucleotide +16 is indicated. The names of the genes are shown in bold.

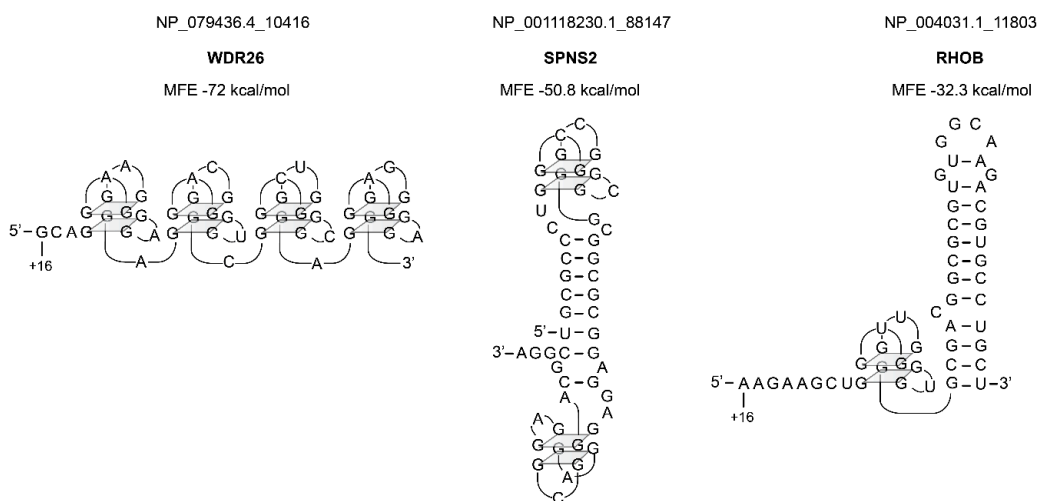


Figure 7. Representation of three structures containing G-quadruplexes selected in the human genome. The position of nucleotide +16 is indicated. The names of the genes are shown in bold.

3. Materials and Methods

Data Processing

Data processing was performed using seven homemade Python scripts available as supplementary data. They must be executed in the following order: (1) `search_stem_loop.py`, (2) `parse_uniprotdata.py`, (3) `analyze_stem_loop.py`, (4) `G_quartets.py`, `GO_terms.py`, `get_rnafold_structures.py`. The Python module named `toolbox.py` is imported by some of these scripts (supplementary file: `Python_scripts.zip`). Coding DNA Sequences (CDSs) from genomic sequences of six eukaryotic and eight prokaryotic species were downloaded from the NCBI site <ftp://ftp.ncbi.nlm.nih.gov/genomes/>.

All CDSs from each species were first carefully checked in order to remove the ones that were non-nuclear or abnormal, like those annotated as pseudogenes. The RNAfold program from the ViennaRNA package 2 [51] was used to predict the minimum free energy (MFE) of secondary structures of RNA sequences located between the +16 and +65 positions, the A from the AUG initiator codon being numbered +1 in each CDS. Detection of G-Quadruplex structures was also incorporated into the RNAfold structure prediction algorithm. The number of guanines involved in G-Quadruplex structures was determined by counting the number of “+” characters in the bracket notation representing secondary structure printed by RNAfold. The MFE_{struc} calculated by RNAfold for a given structure was compared to the mean MFE_{rand} value obtained when the same sequence was randomized 100 times. The ratio $MFE_{\text{struc}}/MFE_{\text{rand}}$ allows to assess the reliability of the secondary structure prediction. The structures with a ratio $MFE_{\text{struc}}/MFE_{\text{rand}}$ equal or higher than 2 were selected. Importantly, the MFE value can be biased by the GC content of the RNA sequence. Therefore, the GC content was also calculated for each 50 nt RNA sequence submitted to RNAfold. PostScript files with plots of secondary structure graphs were produced by RNAfold for a few selected 50 nt CDS sequences. The functions of some CDSs are provided in the description lines of FASTA files downloaded from the NCBI site. Additional functional information can be obtained using the Gene Ontology (GO) resource. For each analyzed species, the relationship between a CDS and its associated GO terms was done parsing two text files, reviewed (Swiss-Prot database) and unreviewed (TrEMBL database), downloaded from the UniProt site <https://www.uniprot.org/>. We are aware that this selection approach is not fully exhaustive and might lead to loss of some putative interesting structures. A major drawback comes from the fact that randomization of sequences with a lot of repeated sequences is not expected to notably change the mfe. Indeed, these types of structures are most likely selected by our approach. Nevertheless, we decided to favor this approach that is more stringent than others in order to obtain a list of strong candidates rather than expanding our list by using alternative methods that might include false-positive structures. In other words, we choose to potentially omit false-positive structures rather than include in our selection false-negative structures that would appear with less stringent approaches.

4. Conclusions

Altogether, our analysis suggests that the START mechanism is putatively an important mechanism that is widely used in the three kingdoms of life to regulate translation of numerous mRNAs. The fact that RNA can modulate their folding in order to change their secondary structures on microsecond time scale [48,49] also suggests that the START mechanism can be used transiently for mRNA subclasses in specific conditions. The so-called “strucluomes” of total RNA have been investigated both in vitro [52] and in vivo with DMS [53,54] and SHAPE [55]. Our study shed light on a novel feature indicating that a careful examination of putative secondary structures in the +16 to +65 window in the previously published genome-wide structural data might bring novel insights in the putative use of the START mechanism. In eukaryotes, it will be of particular interest to investigate the consensus sequences flanking the start codons of mRNAs that are potentially translated by the START mechanism. Since most of the eukaryotic start codons are embedded in the so-called Kozak sequence [56,57], the START mechanism possibly requires alternative consensus sequences around the start codon. In the future, the next step will be to confirm the implication of the selected structures in the START mechanism for

efficient translation initiation. This can be done for instance by transplanting these selected structures in the +16 to +65 areas of reporter mRNAs in order to evaluate their impact on translation initiation efficiency. Finally, our study indicates that the START mechanism can be used in various prokaryotic and eukaryotic organisms. This suggests that the localization of the start codon by steric hindrance with downstream RNA secondary structures might be seen as an ancient mechanism that was maintained during evolution most probably because of its efficiency and simplicity.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/21/8373/s1>.

Author Contributions: F.M. and L.D. designed the experimental strategy. L.D. conceived the bio-informatic tools and performed the database screenings. F.M. analyzed the produced data, made the figures and wrote the article. All authors have read and agreed to the published version of the manuscript.

Funding: F.M. is funded by ‘Agence Nationale pour la Recherche’ (ANR-17-CE12-0025-01, ANR-17-CE11-0024, ANR-20-COVI-0078), by ‘Fondation pour la Recherche Médicale’ (project CoronaIRES), by University of Strasbourg and by the ‘Centre National de la Recherche Scientifique’.

Acknowledgments: We gratefully thank Gilbert ERIANI for helpful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rodnina, M.V. Translation in Prokaryotes. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a032664. [[CrossRef](#)] [[PubMed](#)]
2. Tollerson, R.; Ibba, M. Translational regulation of environmental adaptation in bacteria. *J. Biol. Chem.* **2020**, *295*, 10434–10445. [[CrossRef](#)] [[PubMed](#)]
3. Hinnebusch, A.G.; Ivanov, I.P.; Sonenberg, N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **2016**, *352*, 1413–1416. [[CrossRef](#)] [[PubMed](#)]
4. Hinnebusch, A.G. Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem. Sci.* **2017**, *42*, 589–611. [[CrossRef](#)] [[PubMed](#)]
5. Aylett, C.H.S.; Ban, N. Eukaryotic aspects of translation initiation brought into focus. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160186. [[CrossRef](#)]
6. Akirtava, C.; McManus, J. Control of translation by eukaryotic mRNA transcript leaders—Insights from high-throughput assays and computational modeling. *Wiley Interdiscip. Rev. RNA* **2020**, e1623. [[CrossRef](#)] [[PubMed](#)]
7. Shen, L.; Pelletier, J. General and Target-Specific DExD/H RNA Helicases in Eukaryotic Translation Initiation. *Int. J. Mol. Sci.* **2020**, *21*, 4402. [[CrossRef](#)]
8. Pelletier, J.; Sonenberg, N. The Organizing Principles of Eukaryotic Ribosome Recruitment. *Annu. Rev. Biochem.* **2019**, *88*, 307–335. [[CrossRef](#)]
9. Merrick, W.C.; Pavitt, G.D. Protein Synthesis Initiation in Eukaryotic Cells. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a033092. [[CrossRef](#)]
10. Shirokikh, N.E.; Preiss, T. Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions. *Wiley Interdiscip. Rev. RNA* **2018**, *9*, e1473. [[CrossRef](#)]
11. Schmitt, E.; Coureux, P.-D.; Monestier, A.; Dubiez, E.; Mechulam, Y. Start Codon Recognition in Eukaryotic and Archaeal Translation Initiation: A Common Structural Core. *Int. J. Mol. Sci.* **2019**, *20*, 939. [[CrossRef](#)] [[PubMed](#)]
12. Brion, P.; Westhof, E. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113–137. [[CrossRef](#)]
13. Batey, R.T.; Rambo, R.P.; Doudna, J.A. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed.* **1999**, *38*, 2326–2343. [[CrossRef](#)]
14. Pleij, C.W.A.; Rietveld, K.; Bosch, L. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **1985**, *13*, 1717–1731. [[CrossRef](#)] [[PubMed](#)]
15. Kwok, C.K.; Marsico, G.; Balasubramanian, S. Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a032284. [[CrossRef](#)]

16. Gualerzi, C.O.; Pon, C.L. Initiation of mRNA translation in bacteria: Structural and dynamic aspects. *Cell. Mol. Life Sci.* **2015**, *72*, 4341–4367. [[CrossRef](#)]
17. Mailliot, J.; Martin, F. Viral internal ribosomal entry sites: Four classes for one goal. *Wiley Interdiscip. Rev. RNA* **2017**, *9*, e1458. [[CrossRef](#)]
18. Mearls, E.B.; Jackter, J.; Colquhoun, J.M.; Farmer, V.J.; Matthews, A.J.; Murphy, L.S.; Fenton, C.; Camp, A. Transcription and translation of the sigG gene is tuned for proper execution of the switch from early to late gene expression in the developing *Bacillus subtilis* spore. *PLoS Genet.* **2018**, *14*, e1007350. [[CrossRef](#)]
19. Masachis, S.; Darfeuille, F. Type I Toxin-Antitoxin Systems: Regulating Toxin Expression via Shine-Dalgarno Sequence Sequestration and Small RNA Binding. In *Regulating with RNA in Bacteria and Archaea*; American Society for Microbiology: Washington, DC, USA, 2018; Volume 6, pp. 171–190.
20. Masachis, S.; Tourasse, N.J.; Lays, C.; Faucher, M.; Chabas, S.; Iost, I.; Darfeuille, F. A genetic selection reveals functional metastable structures embedded in a toxin-encoding mRNA. *eLife* **2019**, *8*. [[CrossRef](#)]
21. Duval, M.; Korepanov, A.; Fuchsbauer, O.; Fechter, P.; Haller, A.; Fabbretti, A.; Choulier, L.; Micura, R.; Klaholz, B.P.; Romby, P.; et al. *Escherichia coli* ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biol.* **2013**, *11*, e1001731. [[CrossRef](#)]
22. Byrgazov, K.; Grishkovskaya, I.; Arenz, S.; Coudevylle, N.; Temmel, H.; Wilson, D.N.; Djinić-Carugo, K.; Moll, I. Structural basis for the interaction of protein S1 with the *Escherichia coli* ribosome. *Nucleic Acids Res.* **2014**, *43*, 661–673. [[CrossRef](#)] [[PubMed](#)]
23. Chiaruttini, C.; Guillier, M. On the role of mRNA secondary structure in bacterial translation. *Wiley Interdiscip. Rev. RNA* **2020**, *11*, e1579. [[CrossRef](#)] [[PubMed](#)]
24. Scharff, L.B.; Childs, L.; Walther, D.; Bock, R. Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. *PLoS Genet.* **2011**, *7*, e1002155. [[CrossRef](#)]
25. Accetto, T.; Avguštin, G. Inability of *Prevotella bryantii* to Form a Functional Shine-Dalgarno Interaction Reflects Unique Evolution of Ribosome Binding Sites in Bacteroidetes. *PLoS ONE* **2011**, *6*, e22914. [[CrossRef](#)]
26. Leppek, K.; Das, R.; Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 158–174. [[CrossRef](#)]
27. Pesole, G.; Mignone, F.; Gissi, C.; Grillo, G.; Licciulli, F.; Liuni, S. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* **2001**, *276*, 73–81. [[CrossRef](#)]
28. Lynch, M.; Scofield, D.G.; Hong, X. The Evolution of Transcription-Initiation Sites. *Mol. Biol. Evol.* **2005**, *22*, 1137–1146. [[CrossRef](#)]
29. Mignone, F.; Gissi, C.; Liuni, S.; Pesole, G. Untranslated regions of mRNAs. *Genome Biol.* **2002**, *3*, reviews0004-1. [[CrossRef](#)]
30. Jan, C.H.; Friedman, R.C.; Ruby, J.G.; Bartel, D.P. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nat. Cell Biol.* **2011**, *469*, 97–101. [[CrossRef](#)] [[PubMed](#)]
31. Hinnebusch, A.G. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.* **2014**, *83*, 779–812. [[CrossRef](#)]
32. Archer, S.K.; Shirokikh, N.E.; Beilharz, T.H.; Preiss, T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nat. Cell Biol.* **2016**, *535*, 570–574. [[CrossRef](#)] [[PubMed](#)]
33. Shirokikh, N.E.; Dutikova, Y.S.; Staroverova, M.A.; Hannan, R.D.; Preiss, T. Migration of small ribosomal subunits on the 5' untranslated regions of capped messenger RNA. *Int. J. Mol. Sci.* **2019**, *20*, 4464. [[CrossRef](#)]
34. Qu, X.; Wen, J.-D.; Lancaster, L.; Noller, H.F.; Bustamante, C.D.; Tinoco, I. The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nat. Cell Biol.* **2011**, *475*, 118–121. [[CrossRef](#)] [[PubMed](#)]
35. Takyar, S.; Hickerson, R.P.; Noller, H.F. mRNA Helicase Activity of the Ribosome. *Cell* **2005**, *120*, 49–58. [[CrossRef](#)] [[PubMed](#)]
36. Parsyan, A.; Svitkin, Y.V.; Shahbazian, D.; Gkogkas, C.G.; Lasko, P.; Merrick, W.C.; Sonenberg, N. mRNA helicases: The tacticians of translational control. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 235–245. [[CrossRef](#)]
37. Jagodnik, J.; Chiaruttini, C.; Guillier, M. Stem-Loop Structures within mRNA Coding Sequences Activate Translation Initiation and Mediate Control by Small Regulatory RNAs. *Mol. Cell* **2017**, *68*, 158.e3–170.e3. [[CrossRef](#)]
38. Eriani, G.; Martin, F. START: STructure-Assisted RNA Translation. *RNA Biol.* **2018**, *15*, 1250–1253. [[CrossRef](#)]

39. Bhattacharyya, S.; Jacobs, W.M.; Adkar, B.V.; Yan, J.; Zhang, W.; Shakhnovich, E.I. Accessibility of the Shine-Dalgarno Sequence Dictates N-Terminal Codon Bias in *E. coli*. *Mol. Cell* **2018**, *70*, 894.e5–905.e5. [CrossRef]
40. Boël, G.; Letso, R.; Neely, H.; Price, W.N.; Wong, K.-H.; Su, M.; Luff, J.D.; Valecha, M.; Everett, J.K.; Acton, T.B.; et al. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nat. Cell Biol.* **2016**, *529*, 358–363. [CrossRef]
41. Goodman, D.B.; Church, G.M.; Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **2013**, *342*, 475–479. [CrossRef]
42. Kudla, G.; Murray, A.W.; Tollervey, D.; Plotkin, J.B. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science* **2009**, *324*, 255–258. [CrossRef]
43. Wolin, S.L.; Walter, P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* **1988**, *7*, 3559–3569. [CrossRef]
44. Steitz, J.A. Polypeptide Chain Initiation: Nucleotide Sequences of the Three Ribosomal Binding Sites in Bacteriophage R17 RNA. *Nat. Cell Biol.* **1969**, *224*, 957–964. [CrossRef]
45. Shao, X.; Zhang, W.; Umar, M.I.; Wong, H.Y.; Seng, Z.; Xie, Y.; Zhang, Y.; Yang, L.; Kwok, C.K.; Deng, X. RNA G-Quadruplex Structures Mediate Gene Regulation in Bacteria. *mBio* **2020**, *11*. [CrossRef]
46. Murat, P.; Marsico, G.; Herdy, B.; Ghanbarian, A.T.; Portella, G.; Balasubramanian, S. RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* **2018**, *19*, 1–24. [CrossRef]
47. Yang, S.Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A.G.; Wong, J.M.Y.; Monchaud, D. Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* **2018**, *9*, 1–11. [CrossRef]
48. Crothers, D.; Cole, P.; Hilbers, C.; Shulman, R. The molecular mechanism of thermal unfolding of *Escherichia coli* formylmethionine transfer RNA. *J. Mol. Biol.* **1974**, *87*, 63–88. [CrossRef]
49. Porschke, D. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophys. Chem.* **1974**, *1*, 381–386. [CrossRef]
50. Kennell, D.; Riezman, H. Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *J. Mol. Biol.* **1977**, *114*, 1–21. [CrossRef]
51. Lorenz, R.; Bernhart, S.H.F.; Zu Siederdisen, C.H.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [CrossRef] [PubMed]
52. Del Campo, C.; Bartholomäus, A.; Fedyunin, I.; Ignatova, Z. Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet.* **2015**, *11*, e1005613. [CrossRef]
53. Burkhardt, D.H.; Rouskin, S.; Zhang, Y.; Li, G.-W.; Weissman, J.S.; A Gross, C. Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *eLife* **2017**, *6*, e22037. [CrossRef] [PubMed]
54. Zhang, Y.; Burkhardt, D.H.; Rouskin, S.; Li, G.-W.; Weissman, J.S.; Gross, C.A. A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. *Mol. Cell* **2018**, *70*, 274.e7–286.e7. [CrossRef]
55. Mustoe, A.M.; Busan, S.; Rice, G.M.; Hajdin, C.E.; Peterson, B.K.; Ruda, V.M.; Kubica, N.; Nutiu, R.; Baryza, J.L.; Weeks, K.M. Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **2018**, *173*, 181.e18–195.e18. [CrossRef]
56. Kozak, M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **1987**, *196*, 947–950. [CrossRef]
57. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger rNAS. *Nucleic Acids Res.* **1987**, *15*, 8125–8148. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).