Journal of Theoretical Biology 402 (2016) 54–61

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

In silico analysis of 5′-UTRs highlights the prevalence of Shine–Dalgarno and leaderless-dependent mechanisms of translation initiation in bacteria and archaea, respectively

Ambuj Srivastava [a], Prerana Gogoi [a], Bhagyashree Deka [b], Shrayanti Goswami [c], Shankar Prasad Kanaujia [a,*]

[a] Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India
[b] Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur 784028, Assam, India
[c] Department of Biotechnology, National Institute of Technology Durgapur, Durgapur 713205, West Bengal, India

HIGHLIGHTS

- Prokaryotic genes are closely packed and generally have an intergenic length of < 40 bp.
- Bacteria mostly contain SD-dependent genes while archaea contains leaderless genes.
- SD-led and leaderless genes show opposite correlation with the genome size.
- Like SD-led genes, leaderless genes also utilize ATG as the most dominant start codon.

ARTICLE INFO

Article history:
Received 1 October 2015
Received in revised form
29 April 2016
Accepted 2 May 2016
Available online 4 May 2016

Keywords:
Prokaryotes
Ribosome
Protein translation
Untranslated regions
Start codon

ABSTRACT

In prokaryotes, a heterogeneous set of protein translation initiation mechanisms such as Shine–Dalgarno (SD) sequence-dependent, SD sequence-independent or ribosomal protein S1 mediated and leaderless transcript-dependent exists. To estimate the distribution of coding sequences employing a particular translation initiation mechanism, a total of 107 prokaryotic genomes were analysed using *in silico* approaches. Analysis of 5′-untranslated regions (UTRs) of genes reveals the existence of three types of mRNAs described as transcripts with and without SD motif and leaderless transcripts. Our results indicate that although all the three types of translation initiation mechanisms are widespread among prokaryotes, the number of SD-dependent genes in bacteria is higher than that of archaea. In contrast, archaea contain a significantly higher number of leaderless genes than SD-led genes. The correlation analysis between genome size and SD-led & leaderless genes suggests that the SD-led genes are decreasing (increasing) with genome size in bacteria (archaea). However, the leaderless genes are increasing (decreasing) in bacteria (archaea) with genome size. Moreover, an analysis of the start-codon biasness confirms that among ATG, GTG and TTG codons, ATG is indeed the most preferred codon at the translation initiation site in most of the coding sequences. In leaderless genes, however, the codons GTG and TTG are also observed at the translation initiation site in some species contradicting earlier studies which suggested the usage of only ATG codon. Henceforth, the conventional mechanism of translation initiation cannot be generalized as an exclusive way of initiating the process of protein biosynthesis in prokaryotes.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Protein biosynthesis is comprised of mainly three steps: initiation, elongation and termination. Translation initiation is one of the rate-limiting steps during regulation of differential protein biosynthesis (Brenneis and Soppa, 2009). Although the overall protein translation process is evolutionarily conserved

Abbreviations: bp, base pair; CDS, coding sequence; DNA, Deoxyribonucleic Acid; HGT, horizontal gene transfer; IF, initiation factor; mRNA, messenger Ribonucleic Acid; nt, nucleotide; ORF, open reading frame; RNA, Ribonucleic Acid; SD, Shine Dalgarno; TIS, translation initiation site; tRNA, transfer RNA; TSS, transcription start site; UTR, untranslated region

* Corresponding author.
E-mail addresses: spkanaujia@iitg.ernet.in, spkanaujia@gmail.com (S.P. Kanaujia).

http://dx.doi.org/10.1016/j.jtbi.2016.05.005
0022-5193/© 2016 Elsevier Ltd. All rights reserved.

among the three domains of life i.e. bacteria, archaea and eukarya, the mechanism of translation initiation and its underlying components exhibit a significant divergence. In bacteria, usually the Shine–Dalgarno (SD) sequence (typically GGAGG) at the 5′-UTR of an mRNA is involved in initiating the process of protein translation (Shine and Dalgarno, 1974). The SD sequence is generally located at around 10 bp upstream to the start codon of a coding sequence (CDS) and pairs up with its complementary sequence (CCUCC, known as anti-SD sequence) present at the 3′-end of the 16S rRNA of the small ribosomal (30S) subunit (Shine and Dalgarno, 1975). In bacteria, the interaction between the SD and the anti-SD sequences facilitates the formation of a pre-initiation complex comprising of 30S subunit and three initiation factors (IF1, IF2 and IF3) around the start codon of the mRNA (Jacob et al., 1987). In eukaryotes, the protein translation, as proposed, is initiated by a scanning mechanism wherein the ribosome scans along the mRNA to locate the start codon, most preferably, AUG (Kozak, 1983). In summary, firstly, the pre-initiation (43S) complex consisting of the small ribosomal (40S) subunit, the initiator tRNA and GTP-bound eIF2 is assembled. Subsequently, the 43S complex attaches to the 5′-cap of the mRNA already bound with other initiation factors such as eIF4F complex (Sonenberg and Hinnebusch, 2009). The resulting initiation (48S) complex scans along the 5′-UTR of the mRNA until it encounters an AUG codon which is surrounded by a particular sequence named as the Kozak sequence (gccgccRccAUGG; where R denotes a purine nucleotide, small and capital letters represent semi and fully conserved nucleotides, respectively) (Kozak, 1999). However, in archaea, the exact mechanism of protein translation initiation has not yet been clearly understood. Archaea, being a prokaryote, possess mRNAs which are polycistronic, lack poly-A tail and 5′-cap and thus, similar to that of the bacteria (Keeling and Doolittle, 1995; Tolstrup et al., 2000). On the contrary, most of the archaeal mRNAs either lack SD motif or are completely devoid of 5′-UTR, known as leaderless mRNAs. It has also been reported that homologues of five eukaryotic translation initiation factors (eIF1, eIF1A, eIF2, eIF5B and eIF6) are present in archaea (Dennis, 1997; Kyrpides and Woese, 1998; Benelli and Londei, 2011).

Although the translation initiation mechanism of SD-led genes is well established, the same is not very well understood in the cases of leaderless and non-SD-led genes. The first experimental data asserting the existence of an alternate mechanism for translation initiation in leaderless genes was obtained from *in vitro* studies in *Sulfolobus solfataricus* where authors demonstrated that the translation of 5′-UTR-deleted mRNA can still be initiated (Condo et al., 1999). It is proposed that a possible mechanism of translation initiation of leaderless genes could be the utilization of an undissociated 70S ribosome which along with an initiator tRNA directly forms an initiation complex at the 5′-end of the transcript (Moll et al., 2004; Udagawa et al., 2004; Andreev et al., 2006). It is suggested that the start codon itself might play a crucial role in commencing the process of translation initiation. Furthermore, it was hypothesized that signals downstream to the start codon called “downstream boxes” may bind with the 16S rRNA and help in the translation initiation (Shean and Gottesman, 1992); however, this was experimentally disproven later on (Moll et al., 2002b). Although bacteria possess a significant number of mRNAs with non-SD-led 5′-UTR, the mechanism of translation initiation of these mRNAs is not well explored (Chang et al., 2006). It is suggested that the translation initiation of non-SD-led mRNAs in bacteria is promoted by a ribosomal protein S1 which interacts with the pyrimidine-rich sequences upstream to the start codon (Boni et al., 1991; Moll et al., 2002a; Laursen et al., 2005). Nonetheless, several bacteria and almost all archaea do not encode the ribosomal protein S1 homologue suggesting the existence of another alternate mechanism of protein translation initiation of these mRNAs.

In this study, we performed a comprehensive analysis of the upstream region of prokaryotic CDS to estimate the distribution of different kinds of mRNAs. In total, 107 prokaryotic (archaea: 14 and bacteria: 93) genomes were examined and the distribution of different types of mRNAs were obtained. Moreover, we derived a correlation between the genome size and the leaderless & SD-led genes which show some important aspects in bacterial and archaeal genomes.

2. Materials and methods

2.1. Data collection

All the archaeal and bacterial genomic sequences and their coding sequences were downloaded from the GenBank database available at the National Centre for Biotechnology Information (NCBI). The names and accession numbers of the species are provided in the Supplementary data (Tables S1 and S2). The gene positions in a given genome were obtained from genome database available at NCBI and subsequently the genes were located and their upstream sequences from the stop codon of the previous gene and to the start of the current gene were extracted using home-built shell scripts. It is to be noted that all mRNA sequences were represented with T in place of U.

2.2. Data analysis

To analyse the 5′-UTR regions of genes, the intergenic regions of each genome considered in this study were extracted using the home-built shell scripts. Subsequently, the distribution of intergenic length was plotted. To calculate the length distribution of genes, these were divided in five groups on the basis of their intergenic length: (I) less than 40 bp, (II) between 40 and 100 bp, (III) between 100 and 500 bp, (IV) between 500 and 1000 bp and (V) above 1000 bp. The genes, having intergenic length of less than 40 bp, were considered to be internal genes of a polycistronic mRNA; otherwise monocistronic. The cut-off of 40 bp was chosen based on the previous studies suggesting that most of the operons are found to possess less than 40 bp intergenic region at the 5′-UTR (Tolstrup et al., 2000; de Hoon et al., 2004; Torarinsson et al., 2005). In this study, only SD-led and leaderless genes have been analysed. The SD content at the 5′-UTR of each gene was searched between -4 and -12 positions from the translation initiation site (TIS). The consensus sequence ‘AGGAGGT’ was used as a search pattern for predicting SD-containing genes. The genes containing at least 4 nt out of 5 nt of the consensus sequence at their 5′-UTR were considered as the SD-led genes. This cut-off was chosen based on the observation that ribosome is able to bind mRNAs having at least four complementary bases to 16S rRNA at their 5′-UTR. It is to be noted that only those genes having an intergenic length of more than 10 nt were considered for searching the SD consensus sequence for each genome. To estimate the number of leaderless genes, the promoters were searched upstream to start codon at or near its intended place (position) to transcription start site (TSS) assuming that coincidence of TIS and TSS would refer to a leaderless gene. It is to be noted that only monocistronic genes were searched for being leaderless. In bacteria, -10 and -35 promoter elements were searched between the positions -7 to -13 and -25 to -40, respectively, using the position weight matrix of sigma A promoter from DBTBS database (Promoter database of *Bacillus subtilis*) (Sierro et al., 2008). The position weight matrix was used to search the motifs using the sliding window approach. In archaea, TATA box, BRE^d element, BRE^u and INR box

Download English Version:

https://daneshyari.com/en/article/4495798

Download Persian Version:

https://daneshyari.com/article/4495798

Daneshyari.com