On Base-Pairing Potential Between 16S rRNA and 5' UTR in Archaebacterial Genomes

Yuko Osada[1,2] osada@sfc.keio.ac.jp
Rintaro Saito[1,2] rsaito@sfc.keio.ac.jp
Masaru Tomita[1,3] mt@sfc.keio.ac.jp

[1] Laboratory for Bioinformatics
[2] School of Media and Governance
[3] Department of Environmental Informatics

Keywords: free-energy, 16sRNA, translation initiation

1 Introduction

The Shine-Dalgarno (SD) sequence [4] of *E. coli* is known to be a signal to initiate translation. The widely accepted model is that the 3' end of 16S rRNA base-pairs with the SD sequence in the first step of ribosome binding to mRNA. However, archaebacteria have been supposed to have systems of translation different from those of eubacteria and eucaryotes. Further, some eubacteria, such as *M. genitalium*, do not seem to have SD-like sequences. Rather than directly displaying consensus sequences using software such as Sequence Logos [3], we have developed a method to systematically analyze base-pairing potentials between the 3' end of 16S rRNA and the 5' UTRs of all genes. The base-pairing potentials between the two sequences are calculated as free-energy values, and those values at each position in the 5' UTR are averaged over all of the genes, in order to visualize the base-pairing trend of the organism as a whole.

2 Method

The complete genome sequences of the following 12 procaryotes and *S. cerevisiae* chromosome I and chromosome XII were obtained from the National Center for Biotechnology Information (NCBI) (ftp://ncbi.nlm.nih.gov/): *H. influenzae Rd*, *M. genitalium*, *M. jannaschii*, *H. pylori*, *A. fulgidus*, *B. burgdorferi*, *Synechocystis PCC6803*, *M. pneumoniae*, *E. coli K-12*, *M. thermoautotrophicum*, *B. subtilis*, *A. aeolicus*. The 3' end sequence of each organism is systematically aligned with the 5' UTR of all genes in the organism, with a window size of 20 bps from positions 0 to -50, originating at the start codons. In the alignment, small gaps (bulges) and mismatches (interior loops) are permitted. For each window, the standard dynamic programming method is applied to find the best alignment with the lowest free-energy value. The free-energy value is calculated on the bases of the parameter of Turner [5]. Free-energy values are computed for each base position (from 0 to -50) for each gene. We then take the average free-energy values at each base position over all genes. These averaged values are plotted to visualize the organism's overall tendency of base-pairing between 16S rRNA and the 5' UTR.

3 Results and Discussion

The average free-energy values drop sharply about 15 bps upstream from the start codon in *E. coli* , which is consistent with the model, in which the 3' end of 16S rRNA base-pairs with the SD sequence locates around that position. *B. subtilis* shows a deeper drop and its location is shifted a few bases to the left, more distant from the start codon . *H. influenzae*, *H. pylori*, and *A. aeolicus* show patterns similar to that of *E. coli*, indicating that the organisms have basically the same translation initiation mechanism (Figure 1(a)) [1].

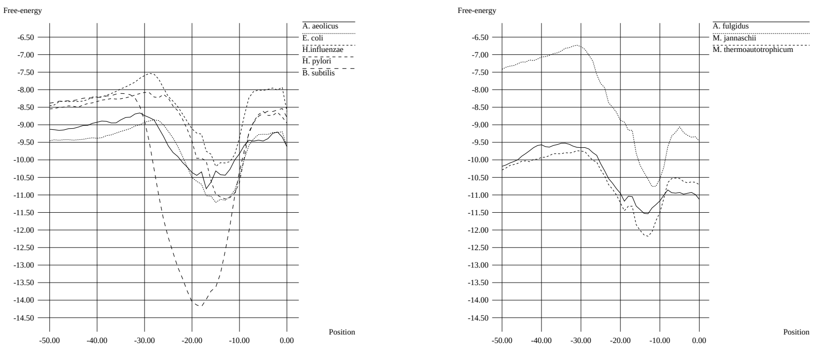

Figure 1: (a) Eubacteria: Free-energy value sharply goes down around SD, then goes up again. (b) Archaebacteria: Free-energy value goes down, but it goes up again slightly .

Each of the 3 archaebacteria, *A. fulgidus*, *M. jannaschii*, and *M. thermoautotrophicum*, shows a free-energy drop similar to that shown in eubacteria (Figure 1(b)), suggesting that archaebacteria employ the eubacterial translation initiation mechanism with base-pairing between 16S rRNA and the 5' UTR. The shapes of the drop for archaebacteria are significantly different from those for eubacteria. In the case of eubacteria, the shape of the drop is symmetric, and the flanking regions of the drop are equally high. The shape of archaebacteria, however, is asymmetric, and the right flanking region is not as high as the left.

We consider that this asymmetry is due to the initiation mechanism of archaebacteria. Saito and Tomita [2] proposed a hybrid model of archaebacteria's translation initiation, in which ribosome scans mRNAs from the 5' to 3' direction (as is the case in eucaryotes) then binds to the SD-like sequence (as is the case in eubacteria). Our results are consistent with this model. SD-like sequences presented upstream of the real SD adsequences could mislead the ribosome and results in a "wrong" translation, thereafter negatively selected over the course of evolution. On the other hand, SD-like sequences presented downstream of the real SD sequences would seldom mislead ribosome, thus they are not the subjects of negative selection. This unbalanced, negative selection of SD-like sequences might have resulted in the asymmetric shape of the free energy drop.

References

[1] Osada, Y., Saito, R., and Tomita, M., Analysis of base-pairing potentials between 16S rRNA and the 5' UTR for traslation initiation in various prokaryotes, *Bioinformatics*, 15:578–581, 1999.

[2] Saito, R., Tomita, M., Computer analyses of archaebacterial genomes suggest that th ey employ both eucaryotic and eubacterial mechanisms in translation initiation, *Gene*, 238:79–83, 1999.

[3] Schneider, TD., Reading of DNA sequence logos: Prediction of major groove binding by information theory, *Meth. Enzym.*, 274:445–455, 1996.

[4] Shine, J. and Dalgarno, L., The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA:Complementarity to nonsense triplets and ribosome binding sites. *Proc. Nat. Acad. Sci. USA*, 71:1342–1346, 1974.

[5] Turner DH., Sugimoto N., Jaeger JA., Longfellow CE., Freier SM. and Kierzek R., Improved parameters for prediction of RNA structure, *Cold Spring Harb. Symp. Quant. Biol.*, 52:123–133, 1987.