Computer Analyses of Translation Initiation Sites in Complete Genomes

Rintaro Saito [1] rsaito@mag.keio.ac.jp
Yuko Osada [2] t94092yo@sfc.keio.ac.jp
Chiyo Yasui [2] t95938cy@sfc.keio.ac.jp
Masaru Tomita [2] mt@sfc.keio.ac.jp

[1] Graduate School of Media and Governance
[2] Department of Environmental Information
Keio University
5322 Endo, Fujisawa 252, Japan

1 Introduction

Although the Shine-Dalgarno (SD) sequence is widely accepted as the signal sequence for ribosome-mRNA binding in procaryotic translation initiation, this sequence is not well conserved among genes within or among species. This is especially striking in *Mycoplasma genitalium*, where no obvious conserved sequence can be observed upstream of start codons. The exact significance of the different SD sequences in the translation initiation process is still not well understood.

We have conducted computer analyses of the complete genome sequence in several bacteria to investigate how the sequence patterns correlate to different conditions. Sequences from the GenBank database were used for the analyses. Our results suggest the possible existence of alternative translation initiation mechanisms which do not utilize Shine-Dalgarno sequences.

2 Summary and Conclusions

- We have calculated entropy values of translation initiation sites in genes classified according to the start codon used(AUG, GUG or UUG). The results show that entropy values near the start codon tend to be much lower in genes which use GUG or UUG rather than AUG(Figure 1). This suggests that these genes require less conservation in the Shine-Dalgarno sequence, possibly because of an alternative mechanism for translation initiation. The same tendencies exist in *H.influenzae* and *Synechocystis* PCC6803.

- We scanned the sequences with respect to potential of hybridization with the 3'-terminus of 16SrRNA, an process which is thought to be essential in translation initiation. Here, the variation in free energy was calculated along the mRNA sequence in a window function to infer the likely hybridization site.

  All genes annotated in the complete genome sequences of *E.coli*, *M.genitalium*, *H.influenzae* and *Synechocystis* PCC6803 were used for this analysis. In *E.coli* and *H.influenzae*, a drop in free energy is observed around position -14, implying conservation of the SD sequence. However, no significant drop can be seen in *M. genitalium* or *Synechocystis* PCC6803 (Figure 3).

- To investigate the possibility of an alternative sequence pattern to account for the lack of SD sequences in *M.genitalium*, triplet frequencies around the start codon were analyzed. Normalized frequency values were used to account for local triplet bias.

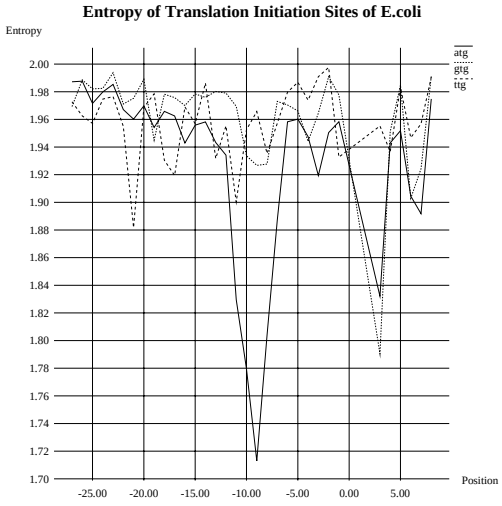

Figure 1

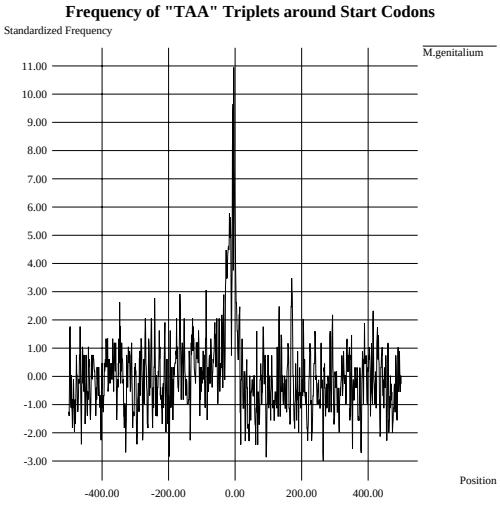

Figure 2

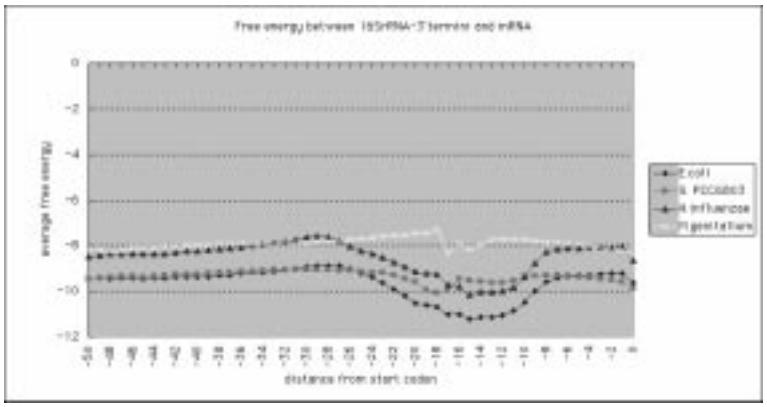

Figure 3

An outstanding peak in the frequency of the triplet "TAA" was observed between positions -27 to -4 (Figure 2). The fact that the complementary triplet "TTA" does not exist in the 3'-terminus of 16S rRNA of *M. genitalium* leads us to suspect the existence of an alternative mechanism for procaryotic translation initiation.

Acknowledgments

This work is supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, "Genome Science", from The Ministry of Education, Science, Sports and Culture in Japan.

References

[1] Loechel, S., Inamine, J. and Hu, P., "A novel translation initiation region from *Mycoplasma genitalium* that functions in *Escherichia coli*," *Nuc. Acid. Res.*, 19:6905–6911, 1991.

[2] Stormo, G., Schneider, T., and Gold, L., "Characterization of translation initiation sites in *E.coli*" *Nuc. Acid. Res.*, 10:2971–2997, 1982.