



THYMINE IN DIFFERENT FRAMES OF UNTRANSLATED REGIONS OF NUCLEIC ACIDS

RAJASEKARAN E.* AND JACOB A.

Dept. of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Karunya Nagar, Coimbatore- 641 114, TN, India.

*Corresponding Author: Email- ersekaran@gmail.com

Received: May 07, 2012; Accepted: February 07, 2013

Abstract- Earlier study on thymine distribution in coding frames of mRNAs reveals that frame 1 prefers to have a defined number of thymine. The same in untranslated regions (UTRs) are further looked in to check this distribution in different frames. The results reveal that there is no such preference over the thymine distribution in different frames. Also confirms the earlier report of defined number of thymine in different frames of coding regions.

Keywords- thymine distribution, untranslated regions, coding regions

Citation: Rajasekaran E. and Jacob A. (2013) Thymine in Different Frames of Untranslated Regions of Nucleic Acids. International Journal of Bioinformatics Research, ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 5, Issue 1, pp.-282-284.

Copyright: Copyright©2013 Rajasekaran E. and Jacob A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Introduction

The essential literature on untranslated regions is now available in digital form globally. A review on functional characterization of eukaryotic mRNA untranslated regions is available [3].

During protein synthesis the initiation of translation mainly dependent on regulatory elements present in untranslated regions (UTRs). The regulatory proteins bind to these regions and initiate the synthesis. The synthesis can either be activated or suppressed by mutation in UTRs. The up or down regulation of synthesis can cause disease [10]. This well ordered UTR regions influence the rate of translation. A comparison of single gene in different species on UTR and coding regions are reported [5]. There are functional RNAs which are again expressed by the regulatory elements [2]. Palanivelu and coworkers have developed tool for prediction of patterns present in UTRs [6].

The role of thymine in different frames mRNA sequences are reported by us [1]. It reports that the thymine in protein coding mRNAs is distributed with preference. Particularly the frame1 prefers to have definite amount of thymine. This is because of the thymine in frame 1 refers to codon XTX (where X stands for A, T, G or C) which codes for large hydrophobic residues which again responsible for adding carbon adequately [4,9]. Further reduction in thymine is observed [7] which are responsible for hydration of proteins during evolution [8]. Our question is, whether the thymine distribution is same in frames of UTRs as in the case of coding mRNAs? This has been addressed here.

Methodology

The [Fig-1] gives the details on different regions of mRNAs. The 5'UTRs are just before the start codon and the 3'UTRs are immedi-

ately after stop codon. The regions between start and stop codons are coding regions.

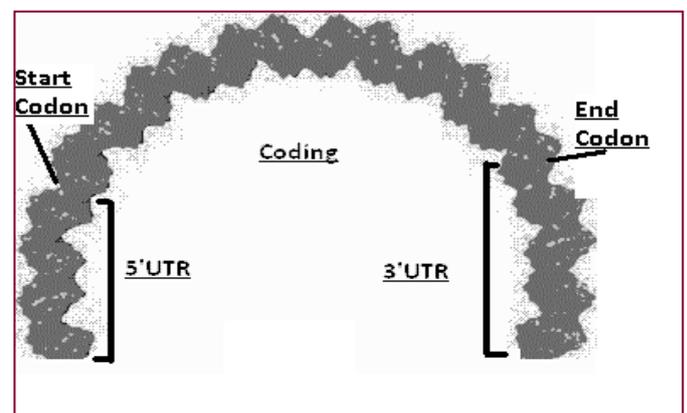


Fig. 1- Representation of coding sequence (CDS) and untranslated regions (UTRs)

The mRNA sequences of human are taken in genbank format from NCBI and extracted the 3'UTR, 5'UTRs and coding sequences using WIBR UTR extractor (<http://jura.wi.mit.edu/bioc/tools/utrs/>). The thymine content in all 6 frames of different sequence is computed using XTX tool available online. It tabulates the number of thymine in all six frames plus the total number of nucleotides. From this table the fraction of thymine in each frame is computed. This calculation is carried out for all sequences of different regions and species. In fact 100 sequences are taken from each species and computed in one go by using modified version of XTX tool called DNAFRAME. The mean and standard deviation of the thymine fraction in different sequences are calculated using in house software.

The mean value is taken as thymine fraction in that particular frame. In a particular frame, the variation in thymine content in different sequences is calculated using standard deviation value. This is carried out for all frames, regions and species. The values are tabulated as shown in [Table-1] and plotted graph for comparison. The results on species comparison are not given here as it is not the focus of this research work.

Table 1- The mean thymine fraction in different frames of different regions is shown

Frame Number	5' UTRs	Coding	3' UTRs
1	0.055	0.084	0.088
2	0.062	0.066	0.089
3	0.061	0.054	0.091
4	0.056	0.1	0.09
5	0.061	0.081	0.09
6	0.063	0.056	0.088

Results and Discussion

Our earlier work on this topic concluded that frame 1 of coding mRNA sequences prefers to maintain definite amount of thymine in order to translate adequate number of large hydrophobic residues to maintain required carbon content in the corresponding protein. This result is confirmed again here. Beyond that what happens in thymine distribution in pre(5'UTRs) and post(3'UTRs) coding sequences? To compare this in UTRs, hundred UTRs (both 3' and 5') human mRNAs are taken. These extracted 3'UTR, 5'UTR and coding sequences are grouped into different file and thymine contents were analysed.

The thymine fraction (thymine in frame1 divided by total number of bases) in all these hundred sequences are computed and grouped based on thymine content as shown in [Table-2]. From this a plot [Fig-2] of thymine fraction versus number of sequence is obtained. The fraction (0.09 here) at frequent number of sequence is taken as the probable thymine content in the frame. This fraction is computed as distribution mean. This calculation was repeated for each frame of CDS and UTRs. The fraction obtained is tabulated as shown in [Table-1].

Table 2- Number of sequences with different fractions of thymine in one frame is shown as example

Thymine Fraction	No. of Sequence
0.01	0
0.02	1
0.03	1
0.04	1
0.05	3
0.06	6
0.07	8
0.08	16
0.09	28
0.1	17
0.11	10
0.12	7
0.13	1
0.14	1

A graph showing this variation of thymine in different frames in different regions is obtained as shown in [Fig-3]. The blue line (◆) indicates the thymine fraction in different frames of 3' UTR. Similarly the 5'UTR and coding regions are shown in brown (■) and green (▲). The variation in thymine fraction in different frames is observed in coding regions and not in UTRs. Mild variation may be discussed in 5'UTR compared to 3'UTR. Overall there is no prefer-

ence of thymine fraction in different frames of UTRs. As discussed earlier [1], frame 1 and 4 are having higher thymine fraction while frame 3 and 6 least fraction. The fraction of thymine in frame 1 is ~0.09. A preliminary result on comparison of this fraction in different species gives the same values. A complete result on this value will be published elsewhere.

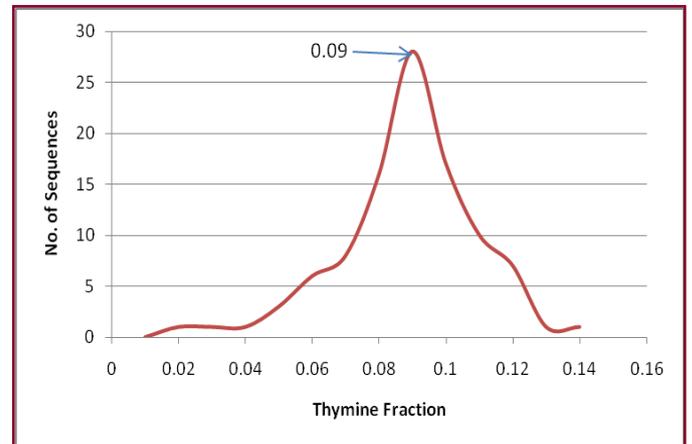


Fig. 2- Graph showing number sequences at different fraction of thymine.

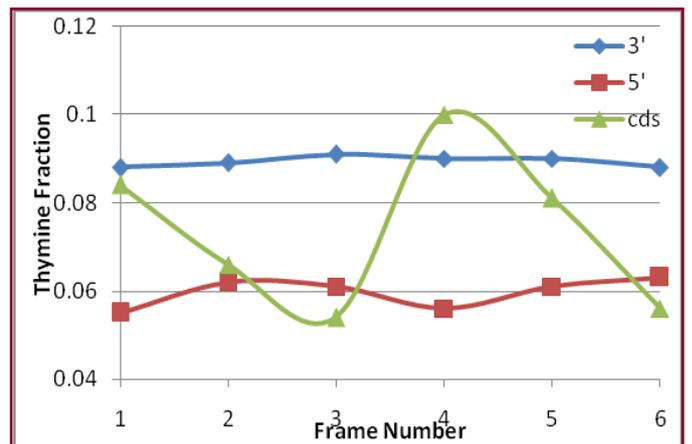


Fig. 3- Thymine fraction in different frames of human 3'UTRs, 5'UTRs and coding regions.

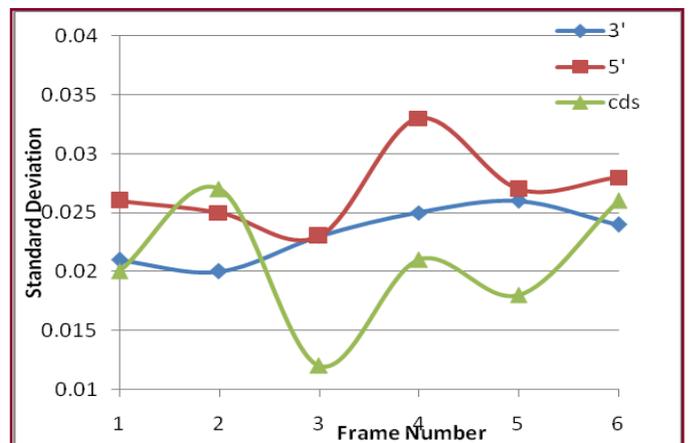


Fig. 4- Thymine fraction variation in different frames of 3'UTRs, 5'UTRs and coding regions as calculated from standard deviation.

One can calculate the standard deviation of the distribution curve. [Fig-2] shows the distribution of data set. If the distribution curve is

narrow, then the standard deviation will be less. In biological term the particular frame prefers to have single specific fractional value. Other side the broad curve indicates the variable fraction of thymine in the frame. So the standard deviation value for each frames of different regions are computed and tabulated as shown in [Table-1]. Again the graph obtained for this number is shown in [Fig-4].

Overall the coding regions have less standard deviation compared to UTRs with exception to frame 2 & 6. This exception is due to the thymine in frame 2 and 6 does not alter the translation of amino acid in the protein. As the standard deviation is less in frame 1, 3, 4 and 5 of coding sequence, it try to maintain defined number of thymine for protein translation. Frame 3 has the least standard deviation. This may be important for protein translation. That is frame 3 should have less defined number of thymine. Any adjustment in thymine content in frame 2 and 6 can be tolerated.

Generally the standard deviation is high in 5'UTR, stating that least preference of thymine distribution in different frames. That is varying number of thymine observed in different frames. To be specific, it can be any number in any frame. Compared to 5'UTR, the 3'UTR has less standard deviation. From the mean and standard deviation calculation of distribution, it is observed that high variation of thymine fraction in different frames is observed in coding regions while it is minimal in 5'UTR and no variation in 3'UTR.

Conclusion

Thymine distribution analysis in different frames of coding and UTRs is conclude that high variation of thymine fraction in different frames is observed in coding regions while it is not so in 5'UTR and no variation in 3'UTR. Thymine content is not specific in 3' and 5'UTRs. Thymine in coding frames are ultimately important in producing proteins with adequate hydrophobicity.

References

- [1] Anandagopu P., Suhanya S., Jayaraj and Rajasekaran E. (2008) *Bioinformation*, 2, 304-307.
- [2] Vaz C., Ahmad H.M., Sharma P., Gupta R., Kumar L., Kulshreshtha R. and Bhattacharya A. (2010) *BMC Genomics*, 11, 288.
- [3] Pesole G., Grillo G., Larizza A. and Liuni S. (2000) *Briefings in Bioinformatics*, 3, 236-249.
- [4] Jayaraj V., Suhanya R., Vijayasathy M., Anandagopu P. and Rajasekaran E. (2009) *Bioinformation*, 3, 409-412.
- [5] Mukesh M., Mishra B.P., Kataria R.S., Sobti R.C., Ahlawat S.P. (2006) *J. DNA Seq. Map.* 17(2), 94-8.
- [6] Marikannu R. and Palanivelu P. (2005) *Ind. J. Biotech.*, 4, 21-38.
- [7] Rajasekaran E. and Anandagopu P. (2010), *J. Adv. Biotech*, 9, 9-10.
- [8] Rajasekaran E., Rajadurai M., Vinobha C.S. and Senthil R. (2008) *Comput. Intelli. in Bioinfo.*, 1, 109-114.
- [9] Rajasekaran E., Vinobha CS., Vijayasathy M., Senthil R. and Shankarganesh P. (2009) *International Association of Computer Science and Information Technology- Spring Conference, IEEE*, 5, 452- 453.
- [10] Chatterjee S. and Pal J.K. (2009) *Biol. Cell* 101, 251-262.