Freie Universität Berlin

MSc Thesis

Prediction of polyadenylation motives with the help of 3' UTR secondary structure

Moritz Kiekeben
Matrikelnummer: 4179384
moritz.k@fu-berlin.de

Betreuer: Prof. Dr. Peter N. Robinson
Eingereicht bei:

Berlin, June 23, 2016

Abstract

Abstract

Declaration of authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references.

I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future theses submitted. Further rights of reproduction and usage, however, are not granted here.

This paper was not previously presented to another examination board and has not been published.

June 23, 2016

Moritz Kiekeben

Contents

1 Introduction

1.1 Motivation

The polyadenylation signal(PAS) is part of the 3' untranslated region(UTR) of almost every messenger RNA. It is the binding site for a protein complex that causes the formation of the polyadenylation tail(poly(A)-tail). This tail of adenosines prevents the degradation of the mRNA in the cytoplasm effectively affecting the half-life of a mRNA and thereby raising the chance to be translated into a protein by the ribosome. As such, variants that have an impact on the efficiency of the PAS are potentially pathogenic due to their interference with the gene expression. Chen et al. wrote a review about diseases caused by PAS-affecting mutations. [REFERENCE]
Prediction of polyadenylation sites has been attempted several times with moderate success in the last twenty years. [REFERENCES/EXAMPLES] Seeing that we cannot predict PASs with high accuracy prompts the question if we have enough knowledge about the structure of polyadenylation sites. Analyzing the secondary structure of the 3' UTR might give us more insight and allows the use of more characteristics for the prediction.
Accurately identifying functional PASs allows for better detection of disease causing mutations.

1.2 Background

1.2.1 History

Polyadenylation (poly(A)) is the process of adding adenosines to the 3' end of precursor mRNA (pre-mRNA), i.e. messenger RNA that has been recently transcribed and has yet to undergo any post-transcription steps like splicing or polyadenylation.
The enzyme polyadenylate polymerase which catalyzes the reaction was first discovered in 1960. [EA60] But the protein's function was revealed about a decade later. [EVN71] At that time it was still a mystery why most mRNAs had polyadenylated tails.
In the early 1980s the role of the poly(A)-tail was largely elucidated: nuclear export, translation and stability of the mature mRNA. (reviewed in [Bra81]). It took another decade to discover the two primary proteins that mediate the 3' end cleavage and polyadenylation, namely cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CStF). The number of participating proteins has since increased and recent publications mention several dozen cis-acting and auxiliary proteins or polypeptides working together in a large complex. [EUA13]
2008 it became clear that most genes have more than one polyadenylation site, resulting in transcripts with different lengths but identical coding sequence. [SJH+08] The new finding was termed alternative polyadenylation

(APA). Since the 3' untranslated Region (3' UTR) has many binding sites for mRNA-repressive micro RNAs (miRNA), APA is taking part in the gene expression. [SNS$^+$08] mRNA is a biomolecule generated by the enzyme RNA polymerase during the DNA transcription. The primary role of mRNA is to transport the genetic information stored by the DNA to the ribosome in the cytoplasm. The ribosome then translates the mRNA into a protein.

1.2.2   messenger ribonucleic acid (mRNA)

1.2.3   Polyadenylation

References

[Bra81]   G. Brawerman. The Role of the poly(A) sequence in mammalian messenger RNA. *CRC Crit. Rev. Biochem.*, 10(1):1–38, 1981.

[EA60]   M. EDMONDS and R. ABRAMS. Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *J. Biol. Chem.*, 235:1142–1149, Apr 1960.

[EUA13]   R. Elkon, A. P. Ugalde, and R. Agami. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, 14(7):496–506, Jul 2013.

[EVN71]   M. Edmonds, M. H. Vaughan, and H. Nakazato. Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc. Natl. Acad. Sci. U.S.A.*, 68(6):1336–1340, Jun 1971.

[SJH+08] Y. Shen, G. Ji, B. J. Haas, X. Wu, J. Zheng, G. J. Reese, and Q. Q. Li. Genome level analysis of rice mRNA 3'-end processing signals and alternative polyadenylation. *Nucleic Acids Res.*, 36(9):3150–3161, May 2008.

[SNS+08] R. Sandberg, J. R. Neilson, A. Sarma, P. A. Sharp, and C. B. Burge. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, 320(5883):1643–1647, Jun 2008.