



MSc Thesis

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

Moritz Kiekeben

Matrikelnummer: 4179384

[moritz.k@fu-berlin.de](mailto:moritz.k@fu-berlin.de)

Betreuer: Prof. Dr. Peter N. Robinson

Eingereicht bei:

Berlin, January 23, 2017

## **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.

January 23, 2017

Moritz Kiekeben

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Background . . . . .	1
1.2.1	Polyadenylation . . . . .	1
1.2.2	Scientific history of polyadenylation . . . . .	1
1.3	Messenger ribonucleic acid (mRNA) . . . . .	2
1.3.1	3' UTR . . . . .	3

# 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  
needed

Prediction of polyadenylation sites has been attempted several times with moderate success in the last twenty years. 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.

reference  
and  
example  
needed

Accurately identifying functional PASs enables us to detect disease-causing mutations easier.

## 1.2 Background

### 1.2.1 Polyadenylation

Polyadenylation (poly(A)) takes place in the cell nucleus and describes 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. Most eukaryotic mRNAs have a polyadenylated tail, except for histone mRNAs. The PAS is a well-conserved six nucleotide long motif usually located ten to thirty bases before the cleavage site. Assuming all four bases are evenly distributed throughout the DNA the same hexamer appears on average every 4096 nucleotides rendering the motif sequence a common occurrence.

reference  
needed,  
use the  
one  
from  
wiki  
"polyadeny-  
lation"

### 1.2.2 Scientific history of polyadenylation

The enzyme polyadenylate polymerase which catalyzes the polyadenylation 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.

reference  
needed

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<sup>+</sup>08] The new finding was termed alternative polyadenylation (APA). Since APA influences the length of the 3' untranslated Region (3' UTR), it is taking part in the gene expression because the longer the 3' UTR the more binding sites for mRNA-repressive micro RNAs (miRNA) a transcript can have. [SNS<sup>+</sup>08]

### 1.3 Messenger ribonucleic acid (mRNA)

RNA is a molecule present in all lifeforms. There are many types of RNA like transfer RNA, ribosomal RNA, microRNA or messenger RNA. Most RNA types play key roles in coding (mRNA), decoding (tRNA), regulation (microRNA) and expression of genes.

mRNA is particularly interesting because it transports genetic information stored by the DNA to the protein synthesis factories of a cell, the ribosome. There, the mRNA specifies the amino acids sequence of a protein, making it an important component of the gene expression.

Every mRNA is composed of three parts: the 5' untranslated region (5' UTR), the coding sequence (CDS) and the 3' untranslated region (3' UTR).

The 5' UTR is described as the region upstream of the start codon. It contains structures to regulate the translation of the transcript, like the 5' guanine cap. Although, it is called UTR, the 5' UTR may contain upstream open reading frames translating to peptides that inhibit the translation of the CDS of the transcript.

The CDS contains the sequence that is translated into an amino acid polymer by the ribosome.

The 3' UTR is defined as the region downstream of the stop codon. (see 1.3.1)

mRNA is generated by the RNA polymerase II protein during the transcription process of a protein-coding gene. While prokaryotic and archaeal mRNA are rarely modified, eukaryotic mRNA is usually subject to many processing steps, including the addition of a 5' cap, splicing and polyadenylation. To distinguish between nascent and processed mRNA, the literature has termed them precursor mRNA and mature mRNA correspondingly.

The 5' cap is a special guanine that is added co-transcriptionally to the 5' end of a mRNA. It is needed for the transport of the mRNA from the nucleus

picture  
of  
mRNA  
struc-  
ture

to the cytoplasm and prevents the degradation by exonucleases. The cap is also important for the initiation of the translation in the ribosome.

Splicing takes place during or after transcription and is responsible for the removal of the introns and the joining of the exons of most eukaryotic pre-mRNAs. Splicing can create different mRNAs and thus different proteins from the same pre-mRNA.

Polyadenylation occurs right after the transcription terminates and describes the process of adding adenosines at the 3' end of the mRNA.

reference  
needed

#### 1.3.1 3' UTR

## 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<sup>+</sup>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<sup>+</sup>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.