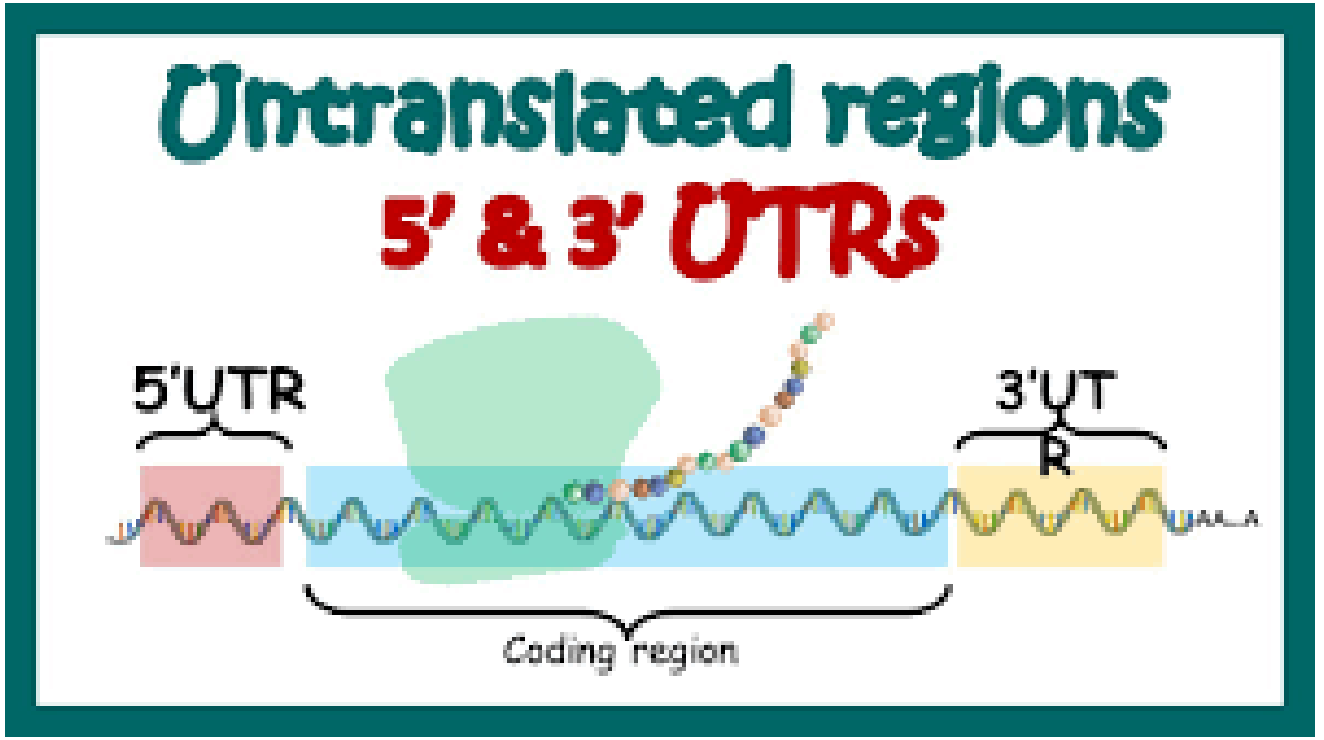
An untranslated region (or UTR)

Fereshteh Noroozi

- An untranslated region (or UTR) including two sections on a strand of mRNA: 5' UTR (or leader sequence) The 5′ UTR, or leader sequence, begins at the 5′ terminal end and ends one nucleotide before the AUG start site, 3' UTR (or trailer sequence: follows the termination codon at the 3′-end of the mRNA) in each side of coding region of mature mRNA.
- This two sections are usually not translated into protein.



History

- In late 1970s after the first mRNA molecule was fully sequenced.
- In 1978, the 5' UTR of the human gamma-globin(A group of proteins found in blood plasma, which contain high levels of antibodies) mRNA was fully sequenced.
- In 1980, a study was conducted on the 3' UTR of the duplicated human alpha-globin genes.

Evolution

- These often long sequences were once thought to be useless or junk mRNA that has simply accumulated over evolutionary time. However, it is now known that the untranslated region of mRNA is involved in many regulatory aspects of gene expression in eukaryotic organisms. The importance of these non-coding regions is supported by evolutionary reasoning, as natural selection would have otherwise eliminated this unusable RNA.

- UTRs are known to play crucial roles in the post-transcriptional regulation of gene expression, including modulation of the transport of mRNAs out of the nucleus and of translation efficiency , subcellular localization and stability. Regulation by UTRs is mediated in several ways. Nucleotide patterns or motifs located in 5’ UTRs and 3’ UTRs can interact with specific RNA-binding proteins.

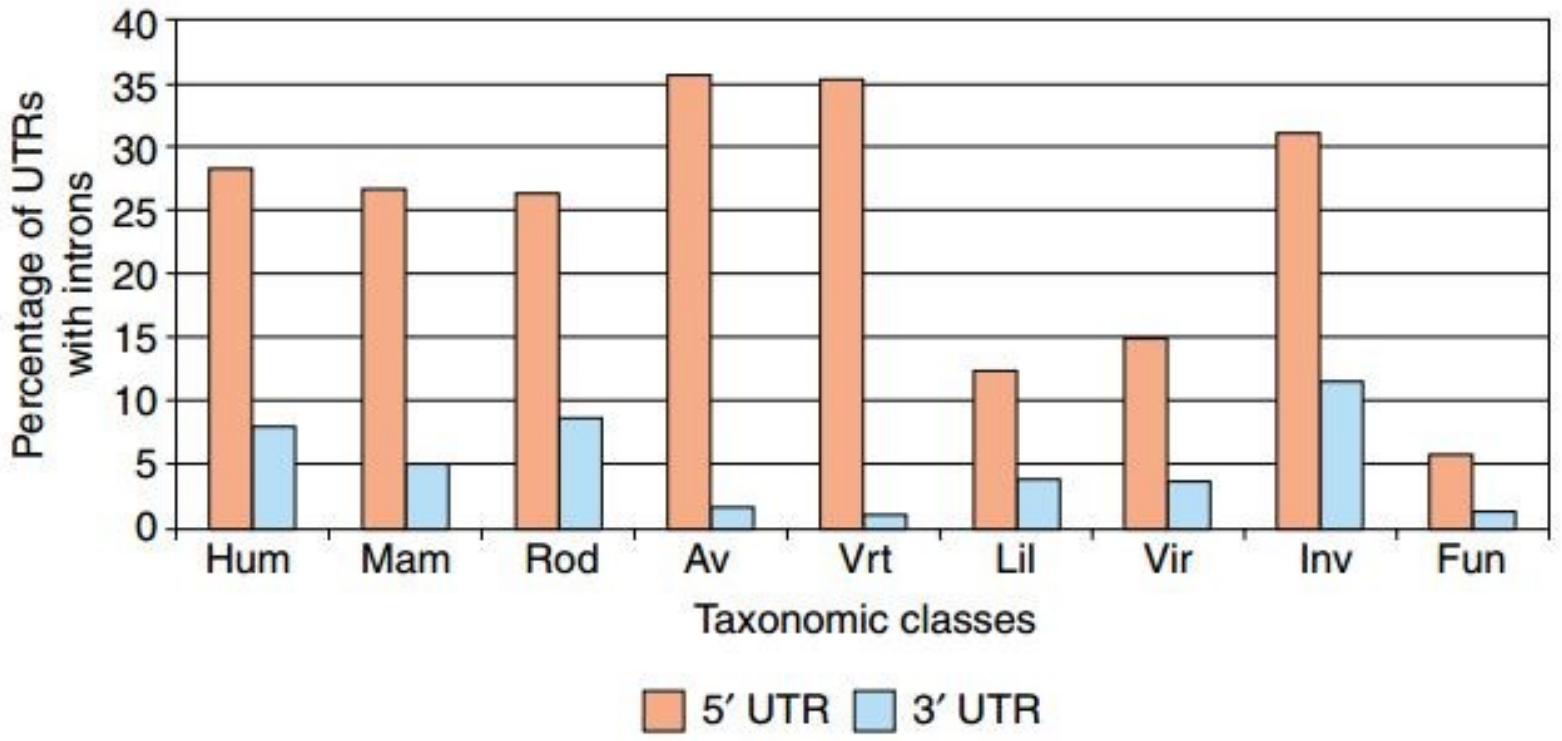
Structural features of untranslated regions

The average length of 5' UTRs is roughly constant over diverse taxonomic classes and ranges between 100 and 200 nucleotides, whereas the average length of 3' UTRs is much more variable, ranging from about 200 nucleotides in plants and fungi to 800 nucleotides in humans and other vertebrates. It is striking that the length of both 5' and 3' UTRs varies a lot within a species, ranging from a dozen nucleotides to a few thousand.

	5′ UTR				3′ UTR			
	Number of sequences	Average length	Maximum length	Minimum length	Number of sequences	Average length	Maximum length	Minimum length
Humans	1,203	210.2	2,803	18	1,247	1,027.7	8,555	21
Other mammals	142	141.3	936	20	148	441.1	3,324	37
Rodents	638	186.3	1,786	16	457	607.3	3,354	19
Aves	59	126.4	620	17	56	651.9	3,990	21
Other vertebrates	105	164.0	1,154	15	111	446.5	2,858	31
Invertebrates	5,464	221.9	4,498	14	3,736	444.5	9,142	15
Liliopsidae	144	129.8	715	17	127	273.3	1,605	22
Other Viridiplantae	1,471	103.0	1,355	12	1,699	207.7	1,911	13
Fungi	388	134.0	1,088	16	326	237.1	1,142	25

Structural features of untranslated regions

- The genomic region corresponding to the UTRs of an mRNA may contain introns, more frequently in the 5’ than in the 3’ UTR.



Structural features of untranslated regions

The base composition of 5' and 3' UTR sequences also differs; the G+C content of 5' UTR sequences is greater than that of 3' UTR sequences. This difference is more marked in mRNAs from warm-blooded vertebrates, whose G+C content is about 60% for 5' UTRs and 45% for 3' UTRs.

In particular, it has emerged that genes localized in large GC-rich regions of a chromosome (heavy isochores: Isochores are large DNA segments (≫300 kb on average)) have shorter 5'UTRs and 3'UTRs than genes located in GC-poor isochores.

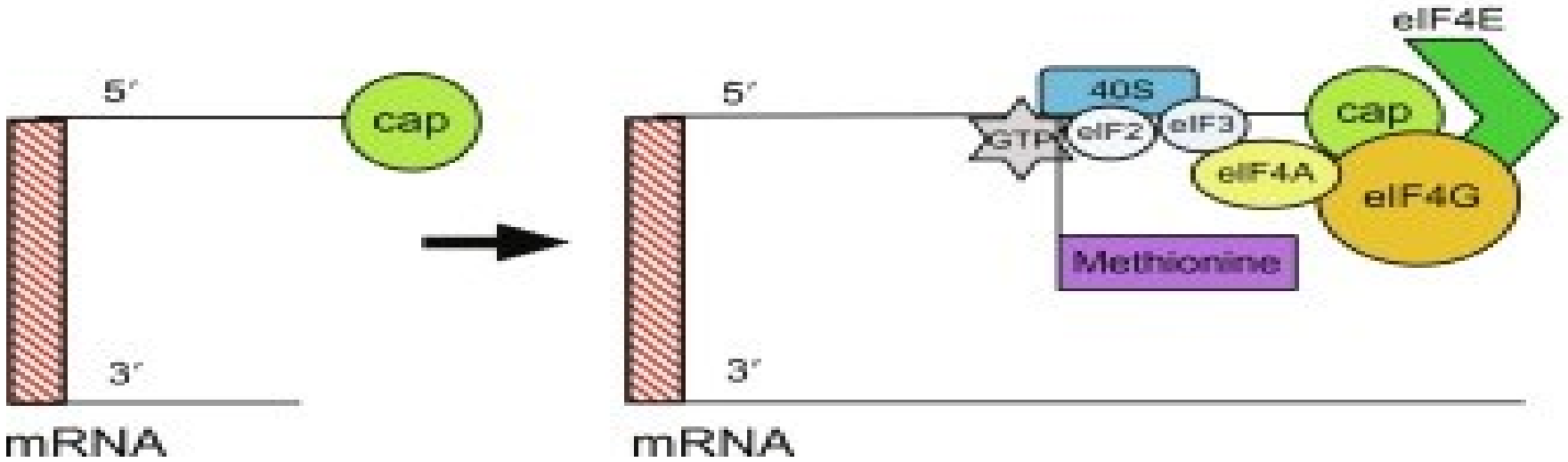
Control of translation efficiency

Translation of mRNAs can vary in efficiency, so that the amount of protein produced is modulated. Structural features of the 5’ UTR have a major role in the control of mRNA translation.

Control of translation efficiency

Under normal conditions, following the transport of an mRNA from the nucleus to the cytoplasm, the eIF4F protein complex assembles at the cap. This complex consists of three subunits: eIF4E, the cap-binding protein; eIF4A, which has RNA helicase activity; and eIF4G, which interacts with various other proteins, including polyadenylate-binding protein. The ATPdependent helicase activity of eIF4A, stimulated by the RNAbinding protein eIF4B, unwinds any secondary structure in the mRNA, thus creating a 'landing platform' for the small (40S) ribosomal subunit .

In most eukaryotic mRNAs, it is thought that translation initiates at the first AUG codon encountered by the 40S ribosomal subunit as it moves, or scans, 3' along the mRNA from 5'.



Regulation of mRNA stability

The turnover of mRNAs (lifetime between 1 and 3 minutes) is another crucial step in post-transcriptional regulation of gene expression, as changes in mRNA abundance may alter the expression of specific genes by affecting the abundance of the corresponding protein.

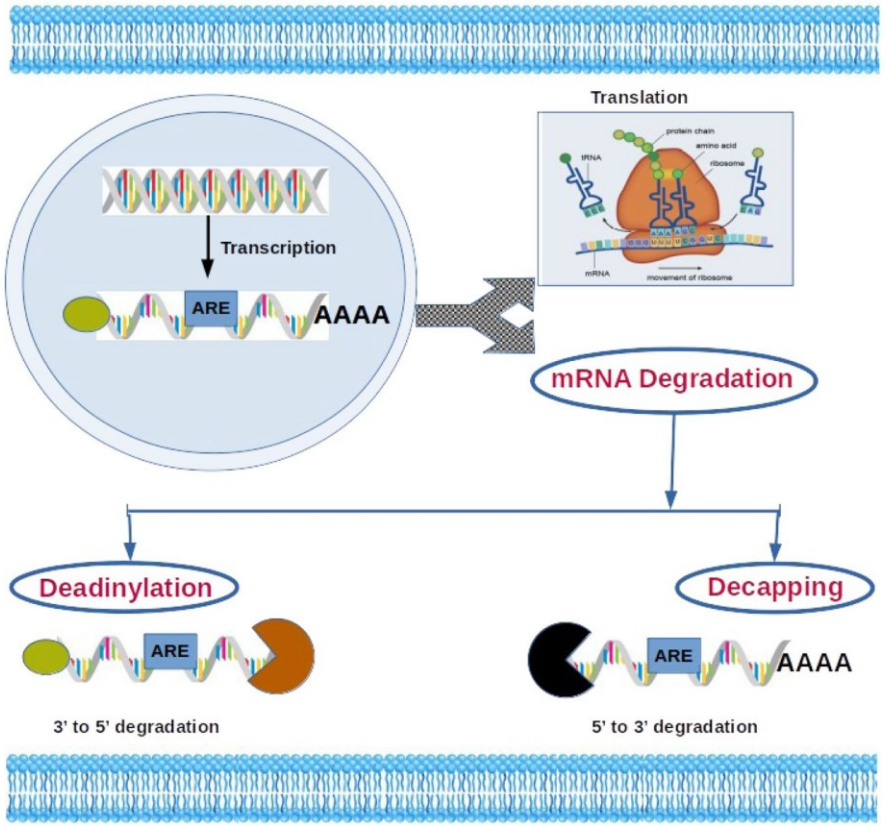
The turnover of an mRNA is mostly regulated by cis-acting elements located in the 3’ UTR, such as the AU-rich elements (AREs), which promote mRNA decay in response to a variety of specific intra- and extra-cellular signals.

Regulation of mRNA stability

Class I AREs by the degradation of all parts of the poly(A) tail at the same rate, generating intermediates with poly(A) tails of 30-60 nucleotides, which are then completely degraded.

Class II AREs mediate asynchronous cytoplasmic deadenylylation, in other words, the poly(A) tail is degraded at differentrates in different transcripts, generating mRNAs without poly(A) tails.

The mRNAs containing class III AREs,they show degradation kinetics similar to those of mRNAs containing class I AREs.



Control of mRNA subcellular localization

The asymmetric localization of some mRNAs leads to an asymmetry of cellular distribution of the encoded proteins; such a situation is clearly more efficient than other possible mechanisms of protein localization, because the same mRNA molecule can serve as a template for multiple rounds of translation.

In all these cases, subcellular localization of mRNA is mediated by cis-acting elements(Cis-regulatory elements (CREs) or Cis-regulatory modules (CRMs) are regions of non-coding DNA which regulate the transcription of neighboring genes.) located in the 3' UTR, but there are also examples of elements in the 5' UTR or even in the coding sequence; these are known as mRNA zip codes and interact with zip-code-binding proteins.

Bioinformatics Perspective

- UTRdb is a database of 5' and 3' untranslated sequences of eukaryotic mRNAs
- Is based on sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs.
- http://utrdb.ba.itb.cnr.it/