



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Leader Sequence

SK Samal, University of Maryland, College Park, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by PS Lovett, volume 2, pp 1078–1079, © 2001, Elsevier Inc.

Glossary

A cap It is the structure at the 5' end of eukaryotic messenger RNA (mRNA) and is introduced after transcription by linking the terminal phosphate of 5' guanosine triphosphate (GTP) to the terminal base of the mRNA. The added G and sometimes other bases are methylated.

Coding sequence The sequence of a gene that represents a protein sequence.

Genome Total genetic information carried by a cell or organism.

Negative sense The sense of DNA or RNA that is complementary in sequence to the positive sense, or the coding sense.

Operon A unit of bacterial gene expression and regulation, including structural genes and control elements in DNA, recognized to regulate gene product(s).

Preinitiation complex A promoter-based complex of an RNA polymerase and initiation protein competent to initiate transcription.

Repressor A protein that inhibits expression of a gene. It may act to prevent transcription by binding to an operator site in DNA or to prevent translation by binding to RNA.

Splicing The precise ligation of blocks of noncontiguous coding sequences (exons) in cellular or viral pre-mRNAs with excision of the intervening noncoding sequences (introns).

Transcription attenuation A mechanism that depends upon the ability of external circumstances to influence ribosome movement in the leader region.

Translation attenuation A mechanism that controls the ability of RNA polymerase to read through an attenuator, which is an intrinsic terminator located at the beginning of a transcription unit.

The messenger RNA (mRNA) region that precedes the coding sequence for a gene is called the 'leader sequence'. This region is also known as the 'five prime untranslated region (5' UTR)' (Figure 1). It begins at the +1 position (where transcription starts) and ends one nucleotide before the start codon of the coding region. Leader sequences have the propensity for forming secondary structures (stem-loops) by base pairing of complementary sequences. In prokaryotes, the leader sequences are usually short. In eukaryotes and viruses, the leader sequences may vary from few nucleotides to several hundred nucleotides. It usually contains the site for ribosome binding. At times, the leader sequence may be translated into a short-leader peptide, which can affect the transcription of the rest of the mRNA.

Leader sequences can regulate downstream expression at the levels of transcription or translation in bacteria and can modulate downstream translation in eukaryotes. Leader sequences in viruses can play an important role in the regulation of gene expression, replication, and pathogenicity. Mutations in the leader sequences of cellular mRNAs can have implications for disease and tumorigenesis.

Transcription in Bacteria

Transcription attenuation comprises one level of regulation for many amino acid biosynthetic operons in enteric bacteria. Nucleotide sequences within the leader cause the formation of a domain of secondary structure which acts as a transcription termination signal for bacterial RNA polymerase. Transcription initiated in the upstream promoter terminates within the leader so as to prevent RNA polymerase from

entering the structural genes of an operon. Transcription termination is relieved when the intracellular concentration of the end-product amino acid of the operon-specified enzymes falls below some minimal level. The level of the end-product amino acid is sensed by the translation of a short-leader-encoded open reading frame (ORF) immediately upstream of the transcription termination signal; the ORF contains one or more codons for the operon end-product amino acid. Low intracellular levels of the end-product amino acid prevent high-level charging of the cognate transfer RNA (tRNA), resulting in ribosomal pausing at leader codons for the end-product amino acid. The paused ribosome interferes with the secondary structure of the transcription terminator causing the formation of a second configuration in the mRNA, the attenuator, which allows transcription to enter the downstream operon-coding sequence.

Transcription antitermination involves the formation of a transcription termination structure in leader mRNA, which is either inhibited or facilitated by the interaction of a protein (or a tRNA molecule) with leader mRNA sequences. For example, the trp RNA-binding attenuation protein (TRAP) plus tryptophan binds to the leader sequence of the *Bacillus subtilis* trp operon causing the formation of a transcription terminator. In the absence of tryptophan, TRAP fails to bind to the leader sequence and an antiterminator structure forms allowing transcription to enter the operon. Other operons that follow this general pattern of regulation include the *bgl* operon of *Escherichia coli*, the *pur*, *pyr*, *hut*, *lic*, and *glp* operons and the *sac* regulon of *B. subtilis*, the *ami* operon of *Pseudomonas*, and the *nas* regulon of *Klebsiella*. Aminoacyl-tRNA synthetases in Gram-positive bacteria are also regulated by antitermination. The uncharged tRNA interacts with the leader sequences to

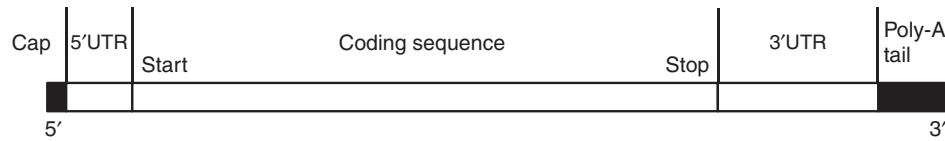


Figure 1 The structure of a typical human protein-coding mRNA, including the untranslated regions (UTRs).

promote the formation of an antiterminator structure allowing transcription to enter the tRNA synthetase-coding sequence.

Translation in Bacteria

Translation attenuation regulates several antibiotic-inducible, antibiotic-resistance genes (e.g., *cat*, *erm*). A domain of secondary structure in leader mRNA sequesters the ribosome-binding site for the downstream resistance determinant, preventing translation initiation. Antibiotic-induced ribosome stalling in a short ORF within the leader causes destabilization of the secondary structure, which frees the ribosome-binding site allowing translation of a coding sequence whose protein product can neutralize the antibiotic.

Translational repression is well exemplified by certain operons encoding bacterial ribosomal proteins. The translational repressor is a single ribosomal protein encoded by the operon; the nonregulatory function of this protein is to act as a structural component of the ribosome. In several examples, the binding target for the repressor protein in the operon leader sequence mimics the structure or sequence of the ribosomal RNA (rRNA) target for the same protein. Binding of the regulatory protein to leader mRNA is presumably of lower affinity than that for rRNA binding *in vivo*. Leader binding by the repressor interferes with translation of operon mRNA by occluding the ribosome-binding site or by changing the secondary structure of leader.

Translation in Eukaryotes

Translation in eukaryotes is typically initiated by the scanning of a 40S ribosomal preinitiation complex. Scanning begins at the 5'-capped end of the mRNA and halts at the first initiator codon, usually AUG, where translation begins. There are a number of ways in which structural elements in the 5' UTR of mRNAs can influence the translation of the downstream ORF.

1. *Regulation of transcription by small structural elements.* Small structural elements in the 5' UTR can affect the translation of certain mRNAs. For example, a stem-loop structure of around 30 nucleotides formed the iron-response element (IRE), which regulates translation of mRNAs that are responsible for iron storage and metabolism. There are two related cytoplasmic iron-regulatory proteins (IRP1 and IRP2) that bind to this element. Under conditions of iron deprivation, these proteins bind to the IRE and block the binding of translation preinitiation complex, which inhibits translation of downstream ORF. Under conditions of iron plentiful, these proteins undergo posttranslational modifications that cause inactivation and degradation by proteasome. Mutations in the sequence of IRE can lead to decreased stability of the

structure or decreased recognition of IRPs, leading to human disease.

2. *Structured 5' UTRs and translation inhibition.* Around 10% of all mRNAs contain unusually long 5' UTRs. Many of those mRNAs encode proto-oncogenes or genes whose protein products are implicated in cell growth. These long 5' UTRs give the propensity for formation of stable secondary structure that may inhibit translation initiation by blocking the progress of the scanning ribosome. These structured mRNAs are particularly dependent on the activity of the cap-dependent unwinding machinery for their translation. Since many of these structured RNAs are associated with growth control, mutation in the 5' UTR of these mRNAs can cause overexpression of the protein product leading to tumor genesis.

3. *Internal ribosome entry site (IRES).* Certain eukaryotic mRNAs contain an internal ribosome entry site (IRES) prior to the coding sequence. An IRES presumably functions in an analogous manner to a bacterial ribosome-binding site in allowing translation initiation by directly serving as a ribosome-binding target. The presence of an IRES preceding a coding sequence in a eukaryotic mRNA enables an mRNA that is not capped to be translated.

IRESs are responsible for translation of many viral and cellular mRNAs with clinical relevance in humans. Mutations in the region of the IRES have been found to influence the translational efficiency of the mRNA.

Transcription and Translation of Viral mRNAs

The 5' UTRs of viral mRNAs have been shown to play important roles in the regulation of viral transcription and translation. In influenza A virus, 5' UTR contains the signals responsible for RNA replication, transcription, and packaging. The 5' UTR of all picorna virus genomes contains IRES that directs cap-independent internal initiation of protein synthesis. The 5' UTR present in all of the late adenovirus mRNAs are termed 'tripartite leader sequence (TPL)'. TPL is formed by the splicing of three exons. TPL facilitates mRNA transport and accumulation in the cytoplasm and is responsible for the selective translation of the late viral proteins in preference to the cellular proteins. All of the coronavirus mRNAs contain identical leader sequences, which are derived from the 5' end of the genomic RNA. Coronavirus leader RNA regulates and initiates mRNA transcription. The 5' leader sequence of retrovirus RNA is involved in a variety of functions that include RNA elongation, RNA splicing, protein translation, genomic RNA dimerization and packaging, and initiation of reverse transcriptions. The 3' end of the genome of nonsegmented negative-sense RNA viruses is also known as the leader sequence that acts as a promoter for transcription and replication of the viral genome.

See also: Open Reading Frame; Transcription; Translation.

Further Reading

- Henkin TM (2000) Transcription termination in bacteria. *Current Opinion in Microbiology* 3: 149–153.
- Hinnebusch A (1994) Translational control of GCN4: An *in vivo* barometer of initiation-factor activity. *Trends in Biochemical Sciences* 19: 409–414.
- Landick R, Turnbough CL, Yanofsky C, *et al.* (1996) Transcriptional attenuation. In: Neidhardt FC, *et al.*, *Escherichia coli and Salmonella*, 2nd edn., American Society for Microbiology, Washington, DC, 1263–1286.
- Lovett PS Rogers EJ (1996) Ribosome regulation by the nascent peptide. *Microbiological Reviews* 60: 366–385.
- Molhoj M and Dal Degan F (2004) Leader sequences are not signal peptides. *Nature Biotechnology* 22: 1502.
- Nakamoto T (2009) Evolution and the universality of the mechanism of initiation of protein synthesis. *Gene* 432: 1–6.
- Pickering BB and Willis AE (2005) The implications of structured 5' untranslated regions on translation and disease. *Seminars in Cell and Development Biology* 16: 39–47.