Transcription in Eukaryotes

Biology – I

Hayder A Giha
Transcription

• Transcription is a DNA-directed synthesis of RNA, which is the first step in gene expression.

• Gene expression, is transformation of the specific nucleotide sequence (genetic information) into a product that influence the cell function.

• The gene is transcribed into RNA and the RNA either translated into protein (mRNA) or it contribute in the process of transcription (rRNA, tRNA) or regulation of transcription (microRNA).

• For transcription of mRNA, the gene sequence in the DNA and its start site need to be identified first.
Types of RNA

• Four types of RNA are known: i. ribosomal (rRNA), transfer (tRNA), messenger (mRNA) and small (sRNA), each have its distinct structure and function

1. Ribosomal RNA (rRNA): accounts for 80% of total RNA in the cell, and complexes with proteins to form the ribosomes. Eukaryotes have different types of rRNA; 18S, 28S, 5S, & 5.8S. The ribosomes are the site for protein synthesis.

2. Transfer RNA (tRNA): Are small RNAs, have a clover-like structure with 4 arms. tRNAs are adaptor molecules involved in protein synthesis. The tRNA recognize the triplet nucleotides (codons) in the mRNA in one hand and hold and deposit the appropriate amino acids in the new polypeptide chain on the other hand.
3. **Messenger RNA (mRNA):** Accounts for about 5% of total RNA, carries the genetic information in the DNA from the nucleus to the cytosol for translation into proteins. It is the most heterogeneous RNA in terms of size and coding information.

4. **Small RNA:** Single stranded RNA molecules, about 21 – 23 nucleotides in length. **MicroRNAs (miRNAs),** are recently discovered subtype. Involved in regulation of gene expression and have the ability to bind mRNA and down regulate its expression.
Gene structure

- The gene is the **minimal linear sequence** of nucleotides in the DNA molecule that **encode** for protein or **structural RNA**.
- The gene sequence is written from the **5'** (5 prime) to **3'**.
- The eukaryotic gene composed of **exons (coding)**, **introns (non-coding)** and **non-coding consensus (regulatory)** sequences.
- The number of **intron** and **exons** differ in **number, size, location and sequences** between the **genes**.
- Any nucleotide before the transcription star site (**5' end**) is located **upstream** (have -ve number e.g. -1) and others located after the star site towards **3' end** are **downstream** (+ve numbers e.g. +5).
Gene structure

- Consensus sequences: are conserved and considered as recognition markers, they are usually bound to transcription factors and other regulatory proteins.
  
  a. Promoter: is a DNA sequence that determine the start site for RNA synthesis, mostly 15-30 bp upstream to the start site, typically has a TATA sequence (known as TATA box).
  
  b. Other sequences that may be needed for promoter function are CAAT box & GC box.

- Basal transcriptional proteins (factors) binds TATA box and link the box to RNA Polymerase II.
Gene structure

- Splice acceptor and Splice donor sequences: are one type of **consensus sequence** found at the 3' ends and 5' ends of the introns.

- **Introns** begin with **GU** nucleotides (splice donor) and end with **AG** nucleotides (splice acceptor) preceded by **pyrimidine-rich** tract. This sequence is essential for **splicing** (removal) of the introns from the primary transcript.

![Intron Diagram]

- 5' splice junction (Donor site)
- Branch point
- 3' splice junction (Acceptor site)
RNA synthesis

- RNA synthesis takes place in the **nucleus** and catalyzed by RNA polymerase using **DNA** template.
- RNA differ from **DNA**: a. Its **single** stranded  b. Contain Uracil (U) instead of thymine (T) c. Much shorter.
- Proteins coding genes are transcribed into **mRNA** that is **translated** in the **cytosol**.
- Regulatory sequences in the mRNA (in the 5' & 3' ends, called **un-translated region** – **UTR**) are important for the **stability** and **translation** efficiency of the mRNA, but are **not translated** into protein.

![Diagram of the structure of a typical human protein coding mRNA including the untranslated regions (UTRs)]
RNA synthesis

• RNA polymerase: There are at least 3 distinct RNA polymerases in eukaryotes:
  a. RNA polymerase I, catalyze synthesis of rRNA
  b. RNA polymerase II, catalyze synthesis of mRNA and microRNA
  c. RNA polymerase III, catalyze synthesis of tRNA

• Transcription and accessory factors: several proteins, transcription factors and accessory proteins (collectively known as pre-initiation complex) are need for initiation of transcription by proper placing of RNA polymerase II to the DNA (just upstream the start site of the gene).
Transcription Promoters

- DNA sequences at start site (recognize TATA BOX)

- RNA polymerase and accessory proteins (transcription factors) bind to the promoter to initiate production of an mRNA transcription.
c. Regulatory regions:

- **Genes** producing **mRNA**, have 2 components:
  a. **Coding region** which is **transcribed** into **mRNA**, and ultimately **translated** into **protein**
  b. **Regulatory region** which is consist to 2 classes of sequences.
    i. Basal **promoters**: a class ensure basal expression
    ii. **DNA regulatory elements**: regulate expression.
1. Basal promoters
   - This sequences have 2 components.
   i. Proximal components: generally the TATA box, which direct RNA polymerase II to the correct site for binding to the promotor region. The protein-DNA interaction at TATA box ensure fidelity of initiation
   ii. Distal components: Includes CAAT, GC boxes & others.
      - The CAAT & GC boxes bind specific proteins, this interaction determines the frequency of transcription initiation.
      - Mutation of the distal components reduce the frequency of transcription 10 to 20 folds.
2. **Enhancers and repressor response elements**

- Are **DNA sequences** that regulate gene expression by enhancing (enhancer element) or repressing (repressing element) expression. The result is increased or decreased rate of initiation, respectively.
- The above effects can be responses to **hormones**, **chemicals**, etc.
- The **response elements** could be found **upstream** or **downstream** the initiation site.
- Can regulate **more than one gene**, including genes that are found **hundreds or thousands bp away** but in the **same chromosome**.
The elements are orientation-independent i.e. can have effects on genes found 5′ to 3′ or 3′ to 5′ in relation to the DNA element.

The same element can be enhancer for one gene and repressor for another gene.

The regulatory effects of the DNA elements are limited by some proteins.
d. **Basal transcription complex:**

- Basal transcription needs in addition to the RNA polymerase certain transcription factors called A, B, C, D, E, F & H, some are composed of several subunits.
- Generally are describe as **TFII** (transcription factor, class II gene) e.g. TFIIA.
- **TFIID** consists of TATA binding protein (TBP) plus 8 to 10 TBP-associated factors.
**Synthesis of single stranded RNA from DNA**

- The RNA polymerase is DNA-dependent RNA polymerase, since it uses the information in the DNA to produce complementary sequence.
- It uses only one strand of the gene as a template (described as **template strand**). The other DNA strand, known as **coding strand**, have a similar sequence to the produced RNA, except for T replaced by U in the RNA.
- The RNA polymerase reads the DNA from 3′ to 5′ direction.
- The product is a single strand complementary to the template strand synthesized from 5′ to 3′ direction.
RNA processing

• The immediate product of transcription in the nucleus is called primary transcript or heterogeneous nuclear RNA (hnRNA), which is larger than the mRNA seen in the cytoplasm.
• The hnRNA contain exons (coding sequence) and introns (non-coding) segments.
• The introns are removed and exons are joined together at specific sites called donor and acceptor sequences by a mechanism known as RNA processing.
RNA processing

A. **Addition of a 5′ cap:** Immediately after initiation of RNA synthesis, the 5′ end of the RNA is capped by methyl guanosine residue which protects it from degradation by 5′ exonuclease, the cap also help the binding of the mRNA with the ribosomes during translation.
B. Addition of poly(A) tail: The hnRNA contain a conserved AAUAAA consensus sequence known as polyadenylation signal several hundreds of nucleotide before the 3′ end.

- The AAUAAA signal is recognized by specific endonuclease which cuts the RNA downstream, after this signal by 20 nucleotides, the cut part is removed.

- To the new 3′ end, up to 250 adenine nucleotides are added by poly(A) polymerase enzyme, forming the poly(A) tail.
RNA processing

c. Splicing or intron removal: Within the gene there are splice sites that indicates beginning (GU) and ending (AG) of each intron in the hnRNA.

- Introns are removed and exons are spliced (joined) together to form the mature mRNA.

- This process is done by, 5 small nuclear RNAs (snRNA): U1, U2, U5 and U4/6, and more than 50 proteins, collectively known as small nuclear ribonucleo proteins snRNPs (pronounced “snurps”)

- The snurps together with the primary transcript (pre-mRNA) form special structure called spliceosome, that form mature mRNA.

- The mature mRNA pass through nuclear membranes pores to the cytoplasm for translation.
hnRNA → capping and poly-A addition (nucleus) → splicing by spliceosome (snRNA and snRNPs in the nucleus) → mRNA

- hnRNA: 5' cap (Me Gppp), 5' UTR, 5' splice site, intron, 3' splice site, 3' UTR
- mRNA: 5' cap (Me Gppp), 5' UTR, 5' splice site, 3' splice site, 3' UTR, poly-A tail

Splicing involves the removal of introns and the joining of exons. The poly-A tail is added during mRNA processing, facilitating its transport to the cytoplasm for translation.
The End