

LECTURE 23

TRANSCRIPTION IN EUKARYOTES

A. Carried out by three different RNA polymerases, each of which contains more than 10 subunits (i.e., these are very complex enzymes). All three require several assisting proteins, collectively called *transcription factors*.

RNAP I → ribosomal RNAs (5.8S, 18S, 28S) and SnoRNAs found in the nucleolus

RNAP II → all nuclear genes encoding proteins (mRNAs) found in the nucleus

RNAP III → all tRNAs, the 5S rRNA, and all snRNAs found in the nucleus

B. Initiation of RNA synthesis:

1. For RNAP II (protein-coding genes), initiation requires several transcription factors that assist binding to promoter sites. Promoter sites recognized by RNAP II (and associated protein factors) are several conserved elements that are located upstream from the transcription start point (the +1 base). The consensus sequences of the conserved elements are...
 - a) -30 = TATAAAA [TATA homology or Goldberg-Hogness box]
 - b) -80 = GGCCAATCT [CAAT box]
 - c) GGGCGG [GC box] & ATTTGCAT [octamer box] often present but occur in different positions and in different copy numbers
2. The TATA homology is found in all eukaryotic promoters known to date. The remaining “consensus” sites are found but not necessarily in the same promoter. All “consensus” sites affect binding efficiency of RNA polymerase/transcription factors.
3. RNAP I and RNAP III utilize some of the same transcription factors as RNAP II but the promoters are quite different. RNAP III utilizes internal promoters (i.e., within the transcriptional units).

C. Elongation of RNA via RNAP II (protein-coding genes):

1. Elongation of the RNA chain is similar to that in prokaryotes except that a 7-methyl guanosine (7-MG) cap is added to the 5' end when the growing RNA chain is fairly short (20-30 bases in length).
 - (a) The 7-MG cap is “attached” by an unusual 5'-5' triphosphate linkage and serves to protect the growing RNA from degradation by nucleases. This “capping” is part of RNA processing in eukaryotes.

D. Termination of RNA synthesis:

1. Transcription by RNAP II (for protein-coding genes) is not really terminated, in the sense that transcription continues for 1,000 - 2,000 bases *after* or *downstream* from the site that ultimately will become the 3' end of the mature transcript.
 - (a) The “functional” transcript actually results from endonucleolytic cleavage of the primary transcript.
 - (b) Cleavage occurs 10-30 bases downstream from the conserved sequence AAUAAA.
 - (c) After cleavage, an enzyme [*poly(A) polymerase*] adds about 200 adenine (A) bases to the 3' ends. This is called *polyadenylation* or the addition of poly-A tails.
 - (i) The function of poly-A tails is to increase stability of the transcript and to assist in transport of the mRNA from the nucleus to the cytoplasm. This is another part of RNA processing in eukaryotes.
2. Termination of transcription via RNAPI and RNAP III is via response to discrete termination signals.

E. Transcription *enhancers* and *silencers*:

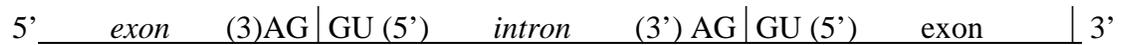
1. These are sequences that exist upstream or downstream of the transcribe sequence, and that enhance or depress rates of transcription.
2. Most enhancers and silencers are not well characterized, can exist close to or distant (>1,000 bp) from the gene(s) they affect.

INTERVENING SEQUENCES IN EUKARYOTES

A. General facts:

1. Most eukaryotic genes are “split” (have intervening sequences), including protein-coding genes and tRNA & rRNA genes. Exceptions include histones and a few others.
2. There are a few introns in prokaryotes. Most are found in viruses and an archebacteria.
3. Introns were “discovered” by the presence of R loops, single-stranded regions observed when DNA-mRNA hybrids (heteroduplexes) were visualized by electron microscopy. Observation of R-loops fit well with earlier observations that pre-mRNAs (hnRNAs) were far larger than predicted based on length of mRNA in cytoplasm and/or amino acid sequence.
4. In general, the amount of intron sequence per “gene far exceeds the amount of exon sequence. Some examples are given in the text; others are listed below.

- (a) Human insulin – 3 exons, 2 introns (intron \approx 1/2 of gene)
 - (b) Chicken ovalbumin – 8 exons, 7 introns (intron \approx 3/4 of gene)
 - (c) Dihydrofolate reductase – 56 exons, 55 introns (intron \approx 95% of gene)
5. Some introns can be quite large. One intron in the *Ubx* gene of *Drosophila*, for example, is roughly 70,000 bp in length.
 6. The only features shared in common by all introns in protein-coding genes are splice sites.



B. Splicing: there are three distinct types of intron excision, one for tRNA, one rRNA, and one for mRNA.

1. For tRNA and rRNA:

- (a) Splicing of tRNA is a two-stage model, involving a *splicing endonuclease* that cuts at the ends of introns, and a *splicing ligase* that joins the two halves of the tRNA together.
- (b) Splicing of rRNA occurs via self-splicing or “autocatalytic” cleavage.
 - (i) this autocatalytic cleavage involves no external energy source and no enzyme(s), but rather is simply a series of phosphoester bond transfers.

2. For mRNA (transcripts of protein-coding genes):

- (a) Splicing is carried out by complex RNA/protein structures called *spliceosomes*.
 - (i) Involves five small snRNA molecules (U1, U2, U4, U5, & U6) that range in size from 100 to 215 nucleotides
 - (ii) snRNA U3 is found only in the nucleolus, and hence is assumed to be involved in ribosome formation
 - (iii) snRNAs do not exist as free RNA molecules but rather are complexed with several proteins into *snRNPs* (small nuclear ribonucleoproteins)
- (b) The initial step is cleavage at the 5' intron splice site (\downarrow GU-intron), followed by intramolecular phosphodiester linkage between the 5' carbon of the G residue at the cleavage site and the 2' carbon of a *conserved* A residue near the 3' end of the intron, forming a *lariat* structure.
- (c) The next step is cleavage of the 3' splice site, followed by joining of the two exons via a 5' to 3' phosphodiester linkage.
- (d) The spliced, processed mRNA is then transported to the cytoplasm for translation.

C. Why introns?

1. Some pertinent factoids:

- (a) The intron arrangement in eukaryotes is quite old. Comparisons of homologous genes across great taxonomic breadth show both similar intron number and location within genes.
 - (i) Examples include α - and β -globin genes in mammals, birds, and amphibians. Introns of these genes are highly conserved in number and placement (but not in sequence), indicating that the structural arrangement is over 500 million years old.
- (b) *Sequence* differences between introns (even within the same species) are large, indicating that there is little constraint on intron sequence *per se*.
- (c) There also are numerous examples in introns of insertions or transposable elements; suggesting that possible functions (?) of introns likely are not related to specific sequences of DNA bases.

2. The most interesting speculation about possible function of introns is *exon shuffling*.

- (a) Most, proteins have several *domains*. These domains include substrate recognition, cofactor recognition, catalytic regions, allosteric functions, et cetera. Examples previously discussed include the DNA polymerases (e.g., DNAP-I, the Kornberg enzyme).
- (b) Many proteins (e.g., enzymes) share one or more domains, and via “exon shuffling” new “genes” coding for similar but different proteins could evolve through recombination of different domains (exons). The evolutionary advantage would be to make “new” genes into single transcriptional units.