Part IV => DNA and RNA

§4.5  RNA Transcription
§4.5a Transcriptional Machinery
§4.5b RNA Processing
Section 4.5a:
Transcriptional Machinery
Synopsis 4.5a

- Transcription is the process of copying or “transcribing” the deoxyribonucleotide sequence within the nontemplate/coding/sense strand of DNA into a corresponding single-stranded ribonucleotide sequence.

- Such transcription results in the production of three major forms of RNA:
  1. Ribosomal RNA (rRNA)—a constituent of ribosomes
  2. Transfer RNA (tRNA)—a component of translational machinery
  3. Messenger RNA (mRNA)—a messenger for protein synthesis

- Together with rRNA and tRNA, numerous other forms of such “noncoding RNA” or ncRNA species are transcribed—such ncRNAs account for close to 80% of total RNA transcription—the other 20% of course being the coding mRNA!

- In eukaryotes, RNA transcription is mediated by three RNA polymerases (RNAPs) designated RNAPI, RNAPII and RNAPIII.

- In addition to RNAPs, a plethora of DNA-binding proteins called “transcription factors” as well as their interactands (called co-regulators/co-activators/co-respressors) are required to initiate transcription in eukaryotes.
## Functional Diversity of RNAs

### (1) RNA translation and protein synthesis

<table>
<thead>
<tr>
<th>RNA Type</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
<td>Coding for protein synthesis</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
<td>Component of ribosomal translational machinery</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
<td>Delivery of amino acids to ribosomes during mRNA translation</td>
</tr>
</tbody>
</table>

### (2) RNA transcription and DNA replication

<table>
<thead>
<tr>
<th>RNA Type</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
<td>Post-transcriptional RNA splicing (a component of spliceosome)</td>
</tr>
<tr>
<td>snoRNA</td>
<td>small nucleolar RNA</td>
<td>Modification of RNA nucleotides (e.g., methylation)</td>
</tr>
<tr>
<td>telRNA/TeRC</td>
<td>telomerase RNA</td>
<td>Synthesis of telomeres during DNA replication (see §4.4)</td>
</tr>
</tbody>
</table>

### (3) Gene regulation

<table>
<thead>
<tr>
<th>RNA Type</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
<td>RNA silencing/ degradation</td>
</tr>
<tr>
<td>piRNA</td>
<td>piwi-interacting RNA</td>
<td>Silencing of transposons (dynamic DNA segments)</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro RNA</td>
<td>RNA silencing/ degradation</td>
</tr>
<tr>
<td>IncRNA</td>
<td>long noncoding RNA</td>
<td>Gene regulation via interaction with transcription factors</td>
</tr>
</tbody>
</table>
RNA transcription is largely mediated by three RNA polymerases (RNAPs) designated RNAPI, RNAPII and RNAPIII.

Like DNA polymerases, RNAPs also require Mg$^{2+}$ divalent ion as a cofactor—the Mg$^{2+}$ ion neutralizes negative charge on incoming NTPs by virtue of its ability to coordinate with phosphoryl groups (cf protein kinases—see §2.5).

RNAPs recognize different and variable promoter sequences within DNA.

RNAPs are specialized for specific tasks:
- RNAPI is primary involved in the synthesis of rRNA precursors
- RNAPII is largely dedicated to mRNA synthesis
- RNAPIII displays a level of functional versatility in that it mediates the synthesis of precursors of tRNA as well as 5S rRNA, snRNA, miRNA, and lncRNA.
- Using the noncoding/antisense DNA strand as a template together with ribonucleoside triphosphates (NTPs), RNAPs undertake the gradual synthesis or polymerization of an RNA strand in a reaction driven by the hydrolysis of pyrophosphate (PPi).

- The sequence of RNA strand is guided by virtue of the ability of incoming NTPs—e.g., ATP, CTP, GTP, and UTP—to engage in Watson-Crick base pairing with the corresponding bases in the template DNA strand.

- Although devoid of 3’→5’ exonuclease activity, RNAPs nevertheless harbor a varying degree of “proofreading” abilities—thus, they not only stall upon encountering the failure of an incoming NTP to properly base pair with the corresponding template DNA base at the active site but also frequently backtrack to add a level of quality control to their work.

- Unlike DNA replication, the failure to properly proof-read the mRNA transcript is not problematic due to its transient nature, coupled with the rather high copy number.
RNA Chain Growth

- Like DNA replication, RNA transcription also proceeds in the 5' → 3' direction.

- Unlike DNA replication, RNA transcription involves copying only one strand of DNA—and the RNAPs harbor an intrinsic ability to initiate transcription without an RNA primer.

- Ribonucleoside triphosphates (NTPs) are added to the growing 3'-end via nucleophilic attack of the 3'-OH group of the ribose sugar on the α-phosphoryl group of incoming NTP with concomitant release of PPI.

- Like DNA replication, RNA transcription also requires the formation of a “transcription bubble” in a manner akin to “replication bubble”—recall §4.4.
During transcription, DNA double helix transiently opens up forming what is called the "transcription bubble" so as to allow the synthesis of an RNA strand—such a bubble apparently travels along with the transcriptional machinery as it moves forward.

- DNA strand that serves as a template for the synthesis of an RNA strand is called the "antisense" or the "noncoding" strand—the non-template DNA strand is accordingly referred to as the "sense" or "coding" strand.

- Save for U→T replacement, the nucleotide sequence of the newly transcribed RNA strand is identical to the non-template/coding/sense DNA strand.

- What may serve as a non-coding strand during the transcription of one gene can also serve as a coding strand during the transcription of another gene—ie both strands of DNA may carry coding information!

- Only about 20% of DNA coding sequences (or genes) encode mRNA directed for protein synthesis—what do the remaining 80% genes encode?!
DNA Supercoiling During Transcription

Formation of a transcription (or replication) bubble results in overwinding (positive supercoiling) of DNA ahead of the bubble while an equivalent unwinding (negative supercoiling) occurs behind it.

In a manner akin to their role in DNA replication:

1. **DNA helicases** facilitate the unwinding of DNA double helix to generate a transcription bubble.
2. **Topoisomerases** help relieve the resulting strain that builds up around the transcription bubble.
Cis-Acting Elements: Promoters, Enhancers & Silencers

- Before RNAP can undertake the transcription of a particular gene, it must dock within the vicinity of its transcriptional start site (TSS)—such region on the coding strand flanking the TSS has come to be known as the “promoter”

- Gene promoters typically span -1000bp upstream (toward the 5’-end of coding strand from TSS) and +100bp downstream (toward the 3’-end of coding strand from TSS)—TSS is a core component of cis-acting promoters!

- The cis-acting promoters not only serve as a docking site for RNAP but also for many trans-acting transcription factors required for the stabilization of RNAP and subsequent transcriptional initiation

- In addition to promoters, eukaryotic genes are also characterized by two additional cis-acting elements called “enhancers” and “silencers”—that can be located up to 1000s of bps upstream or downstream of TSS and may also overlap with each other!

- The enhancers recruit transcriptional activators and thus positively affect gene transcription

- The silencers recruit transcriptional repressors and thus negatively gene transcription

- The ultimate expression of a gene depends upon a subtle interaction between various proteins bound at the promoter, enhancer and silencer
- RNAPs have specific requirements for the recognition of their target promoters—in particular, RNAPII promoters display a rather high degree of complexity and diversity.

- A vast majority of RNAPII promoters (the promoters of protein-encoding genes) usually harbor one or more of the following core cis-acting promoter sequences/elements (wrt the coding strand!) in order to recruit various trans-acting transcription factors (TFs) required for transcription initiation:

  1. **CAAT box**—located around -80bp upstream of TSS, it serves as a docking site for basal transcription factors such as the core binding factors (CBFs) or nuclear factors (NFs).  
  2. **BRE (TFIIB recognition element)**—located around -40bp upstream of TSS, it acts as the recognition site for transcription factor IIB (TFIIB).  
  3. **TATA box**—located around -30bp upstream of TSS, it serves as the binding site for TFIID, or strictly a component of TFIID called the TATA-binding protein (TBP)—one of the very few TFs that binds to DNA in the minor groove so as to facilitate the bending of DNA.  
  4. **INR (initiator) element**—located around the TSS, this pyrimidine(Y)-rich sequence usually serves as an alternative to TATA box in many gene promoters and helps to recruit TFIID.  
  5. **MTE (motif ten element)**—located around +20bp downstream of TSS, it usually acts as a substitute for TATA box in its ability to recruit TFIID in conjunction with the Inr element.  
  6. **DPE (downstream promoter element)**—located around +30bp downstream of TSS, it also serves as a substitute for TATA box in that it is recognized by TFIID in conjunction with the Inr element.
Basal Transcription Factors (BTFs): A Few Examples

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Subunit</th>
<th>Full Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFIID</td>
<td>TBP</td>
<td>TATA binding protein</td>
<td>Binds to TATA box</td>
</tr>
<tr>
<td>TFIID</td>
<td>TAF1</td>
<td>TBP-associated factor 1</td>
<td>Facilitates binding of TBP to TATA box</td>
</tr>
<tr>
<td>TFIIB</td>
<td></td>
<td>Transcription factor IIB</td>
<td>Binds to BRE; tethers TFIID to TFIIF</td>
</tr>
<tr>
<td>TFIIA</td>
<td></td>
<td>Transcription factor IIA</td>
<td>Stabilizes the binding of TFIID to TATA box</td>
</tr>
<tr>
<td>TFIIE</td>
<td></td>
<td>Transcription factor IIE</td>
<td>Tethers TFIIF to TFIIH</td>
</tr>
<tr>
<td>TFIIF</td>
<td></td>
<td>Transcription factor IIF</td>
<td>Tethers TFIIB to TFIIE</td>
</tr>
<tr>
<td>TFIIH</td>
<td></td>
<td>Transcription factor IIH</td>
<td>Harbors helicase activity; facilitates unwinding of DNA</td>
</tr>
</tbody>
</table>

- Expression of all protein-encoding genes relies upon the involvement of RNAPII coupled with a set of core/general proteins termed “basal transcription factors (BTFs)” or “general transcription factors (GTFs)” — such a macromolecular complex is called the basal/general “transcriptional machinery”

- Without the BTFs, RNAPII would not even be able to bind to the target gene promoters!

- BTFs (shown above and many more) are required for transcriptional initiation or a basal/low level of transcriptional activity of all mRNAs — but they alone are not sufficient to account for the expression of specific genes in large quantities in a highly temporal and spatial manner

- If gene expression only relied upon BTFs and RNAPII to mediate transcription of all genes, all cells will harbor a very similar set of proteins — and life would be so boring, for all cells would be alike!

- On the contrary, only a small fraction of genes are expressed in a given cell type — such a phenomenon is referred to as “differential gene expression” — thanks largely to gene-specific transcription factors
Gene-Specific Transcription Factors: A Few Examples

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Name</th>
<th>DNA-Binding Motif</th>
<th>Cis-Acting Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP</td>
<td>Catabolite activator protein</td>
<td>Helix-turn-helix (HTH)</td>
<td>AATGTGATCTAGATCACATTT</td>
</tr>
<tr>
<td>HIF1</td>
<td>Hypoxia-inducible factor 1</td>
<td>Helix-loop-helix (HLH)</td>
<td>(A/G)CGTG</td>
</tr>
<tr>
<td>ERα</td>
<td>Estrogen receptor α</td>
<td>Zinc finger (ZF)</td>
<td>AGGTCANNTGACCT</td>
</tr>
<tr>
<td>Jun</td>
<td>Jun oncoprotein</td>
<td>Leucine zipper (LZ)</td>
<td>TGAC(G)TCA</td>
</tr>
</tbody>
</table>

- Differential gene expression is essential to life not only for cellular homeostasis but more so for the development of differentiated cells and tissues—e.g. pancreatic β-cells vs erythrocytes

- For example, pancreatic β-cells synthesize insulin in copious amounts but no detectable hemoglobin, whereas the reverse scenario prevails in erythrocytes

- How do cells accomplish such a remarkable feat? In addition to BTFs, specific cell types also express a distinct and unique set of cell-specific or gene-specific transcription factors

- Such gene-specific transcription factors run into thousands and they can activate (activators) or repress (repressors) the expression of specific genes so to provide a tight regulatory control depending on the needs of the cell in sync with its micro-environment

- Like their basal counterparts, gene-specific transcription factors also exert their action by virtue of their ability to bind to cis-elements within the promoters of their target genes in a sequence-specific manner and/or via specific interaction with other transcription factors

- What are HTH, HLH, ZF and LZ DNA-binding motifs?! These are super-secondary structural motifs within the DNA-binding (DB) domains of respective transcription factors—recall §2.3 and see §4.7—directly involved in DNA-binding and recognition
Preinitiation Complex—Overview

- The initial assembly of BTFs along with RNAPII at a gene promoter is referred to as the “preinitiation complex (PIC)”

- Within the PIC, TBP (a component of TFIID) reigns supreme in its ability to lead/escort the “pack” of other BTFs to gene promoters—how does it do that?

- Binding of TBP to TATA box just upstream of TSS induces local DNA bending that ultimately results in its partial unwinding and melting

- The resulting conformational change within DNA subsequently facilitates the recruitment of other BTFs that are required to “feed” the template strand of DNA into the active site of RNAPII and help it recognize the TSS

- In particular, TFIIH harbors DNA helicase activity that results in the formation of a “transcription bubble”

- How does TBP bend/melt DNA and facilitate the assembly of PIC?!
While its saddle-like shape may suggest that it straddles the DNA double-helix at the TATA box, the concave face of TBP rather binds to minor grooves along the helical axis of DNA—in lieu of binding perpendicular to it in a manner akin to the binding of most transcription factors within the major groove.

In this manner, TBP acts like a pair of forceps/pliers that compress and bend the DNA inwards so as to partially unwind/melt it.
DNA bending at the promoter region upon the binding of TBP aids subsequent binding of TFIIB to BRE site upstream of TATA box—while the engagement of TFIIA via protein-protein interactions helps stabilize the “territorial” control of TBP over TATA box.
RNAPII with its various subunits or components—such as a sliding clamp, wall, fork, jaw, lobe and protrusion—in complex with partially unwound DNA and other BTFS during the formation of PIC at a gene promoter.
RNAPII in complex with partially-unwound DNA during the act of transcription of a mRNA precursor

- RNAPII is a large multi-subunit complex (~500kD) harboring both catalytic and regulatory functions—in particular, the clamp subunit of RNAPII accounts for its processivity by virtue of its ability to act as a sliding clamp and thereby preventing the dissociation of RNAPII from its DNA template.

- Using one of the DNA strands as a template and NTPs, RNAPII synthesizes a complementary strand of RNA in the 5’→3’ sense in the presence of a plethora of so-called “transcription elongation factors.”

- This newly transcribed RNA strand—the primary transcript or pre-mRNA—will ultimately serve as a precursor of fully mature mRNA.
Exercise 4.5a

- What are the functions of the three eukaryotic RNA polymerases?

- What are the advantages of having multiple types of promoters and enhancers?

- Summarize the roles of basal transcription factors in the formation of transcription preinitiation complex (PIC) with RNAPII

- Why are gene-specific transcription factors important?
Section 4.5b: RNA Processing

Lasso/Lariat
a looped rope for restraining an object or an animal
- Newly transcribed RNA transcripts are called the “primary transcripts”—a primary transcript that ultimately becomes processed into mRNA is referred to as “precursor mRNA” or simply pre-mRNA

- In addition to pre-mRNA, primary transcripts also include precursors for other RNAs such as pre-rRNA and pre-tRNA

- In eukaryotes, pre-mRNAs are comprised of coding regions (exons) interspersed with non-coding regions (introns)—thus, prior to the genetic information stored in exons can be decoded by the ribosomal translational machinery, introns must be spliced out in a process called “RNA splicing”

- The presence of introns in pre-mRNAs is important for generating a diverse array of protein products from a single gene—thus, exons can be spliced out of a pre-mRNA and stitched together via various combinations to generate not one but many mature mRNAs in a process referred to as “alternative splicing”—diversity is indeed the spice of life!

- Major mechanisms involved in RNA processing include covalent modification, endonucleolytic cleavage, and splicing

- Prior to being translated into a sequence of amino acids, the mRNA sequence is read as a contiguous set of triplets of nucleotides referred to as CODONS
RNA Processing: Overview

- Primary transcripts, or pre-mRNAs, harbor three types of segments or regions:
  (a) **Exons**—expressed (coding) regions throughout the transcript
  (b) **Introns**—intervening (noncoding) regions between exons
  (c) **UTRs**—untranslated regions that solely occur at the 5’-end (5’-UTR) and 3’end (3’-UTR)

- Prior to pre-mRNAs can be decoded, they undergo what is referred to as “RNA processing” that can be broadly divided into two major categories:
  (1) **mRNA modification**—involves the addition of a 5’-cap and poly(A) tail
  (2) **mRNA splicing**—involves the excision of introns and the fusion of exons
mRNA Modification: 5’-cap

- **Nascent pre-mRNAs** (as short as 30nt long)—while still being transcribed—undergo a process referred to as “mRNA capping”

- In mRNA capping, the 5’-end of pre-mRNAs is modified with 7-methylguanosine (m7G) via the formation of a 5’-5’ triphosphate bridge (TPB)

- Additionally, the 2’-OH group of the ribose sugars within the first two nucleotides @ the 5’-end being capped may also be methylated

- mRNA capping (or the 5’-cap)—that usually occurs co-transcriptionally—serves the following functions:
  1. Facilitates mRNA splicing by spliceosome
  2. Regulates nuclear export of mRNA
  3. Promotes mRNA translation by ribosomes
  4. Prevents degradation by 5’-exonucleases
mRNA Modification: Poly(A) tail

- In eukaryotes, termination of newly transcribed RNA transcripts is poorly defined—synthesis of RNA transcripts on DNA template is usually terminated by the cleavage of 3’-end by an endonuclease between two conserved sequences: AAUAAA and G/U-rich

- Because of such imprecise cleavage, the 3’-ends of primary RNA transcripts would be highly heterogeneous—i.e., they are without a well-defined sequence or a marker for the recognition of their 3’-ends

- To overcome such 3’-end heterogeneity, primary RNA transcripts are subjected to polyadenylation—a post-transcriptional process wherein as many as 300 adenine nucleotides are added to the 3’-ends of primary RNA transcripts courtesy of poly(A) polymerase (using ATP as a donor)

- The repeating stretch of adenine nucleotides at the 3’-ends of mature mRNAs has come to be known as the “poly(A) tail”—in the cytosol, poly(A) tail exists in complex with poly(A)-binding protein or PAB for short (see §4.6)

- Poly(A) tail serves the following functions:
  1. Regulates nuclear export of mRNA
  2. Promotes mRNA translation by ribosomes
  3. Prevents degradation by 3’-exonucleases
mRNA splicing involves removing (or excising out) introns and splicing (or fusing/joining) together exons.

mRNA splicing is a co-transcriptional process and usually takes place following the 5’-capping of the primary transcript during the elongation stage of transcription but prior to polyadenylation of the 3’-end post-transcriptionally.
In silico nucleotide sequence analysis of exon-intron junctions of eukaryotic pre-mRNAs suggests that they have a rather high degree of homology:

1. The 5′-end of the intron is marked by an invariant GU sequence
2. The 3′-end of the intron is marked by an invariant AG sequence

The 5′-GU and 3′-AG thus define the exon-intron boundaries, or rather they serve as markers or splice sites for the splicing reaction—wherein the introns are excised out and exons are spliced together to generate a mature mRNA ready to be exported out of the nucleus for the next phase of its life (the ribosomal translational machinery).
mRNA Splicing: Splicing Reaction

RNA splicing reaction is mediated by the spliceosome—a large macromolecular machine comprised of small nuclear RNAs (snRNAs) and proteins located within the nucleus—and occurs as follows:

1. The 2’-OH group of an intronic adenosine (usually located close to the 3’-end of an intron) launches a nucleophilic attack on the 5’-phosphate moiety at the 5’-end of intron—the resulting 2’-5’-phosphodiester bond releases the 3’-end of the preceding 5’-exon coupled with the generation of a lariat/lasso within the intron preceding the 3’-exon.

2. The newly liberated 3’-OH group of the 5’-exon nucelophilically attacks the 5’-phosphate at the 5’-end of the 3’-exon—the resulting 3’-5’ phosphodiester bond splices together the two exons coupled with the elimination of intron lariat.

3. The intron lariat is subsequently degraded.
mRNA Splicing: Alternative Splicing

- Spliceosome does not merely stick together all exons within a pre-mRNA but rather it is selective in how the various exons are spliced together so as to generate not one but many mRNAs from a single pre-mRNA—such a clever act performed by the spliceosome is called “alternative splicing”

- In alternative splicing, introns of one mRNA may serve as exons in another!

- Alternative splicing offers a powerful evolutionary advantage in its ability to diversify gene products (eg proteins) so as to meet specific and specialized demands of various cells and tissues
Exercise 4.5b

- Summarize co-transcriptional and post-transcriptional modifications of eukaryotic mRNA

- What is the advantage of starting RNA processing before transcription is complete?

- Discuss the advantages, in terms of protein structure and evolution that result from alternative mRNA splicing