BCMB 3100 – Chapters 36-38
Transcription & RNA Processing

• Definition of gene
• RNA Polymerase
• Gene coding vs template strand
• Promoter
• Transcription in E. coli
• Transcription factors
• mRNA processing

Biological information flow

```
DNA ➔ Replication ➔ Transcription ➔ RNA ➔ Translation ➔ Protein
```

- a DNA sequence that is transcribed (includes genes that do not encode proteins)

- “______________________” encode proteins or RNA essential for normal activities of the cell (e.g. enzymes in basic metabolic pathways, tRNAs and rRNAs)

**Four Classes of RNA in living organisms (review)**

*Messenger RNA (mRNA)* - linear “copies” of DNA that encode genetic information. Encode primary structure of protein. ~1-3% of total RNA, relatively unstable (discovered by Jacob & Monod).

*Non-coding RNA*

*Ribosomal RNA (rRNA)* - ~80% of total RNA, part of ribosomes (translation machinery)

*Transfer RNA (tRNA)* - ~15% of total RNA, 73-95 nucleotides long, carry activated amino acids to ribosomes during translation

*(Small RNA)* - may have catalytic activity and/or associate with proteins to enhance activity, some involved with RNA processing (includes snRNA and microRNA, the latter involved in mRNA degradation, translation inhibition and chromatin remodeling), (long non-coding RNAs (long ncRNAs)): functions being determined.
RNA Synthesis \textit{(E. coli)}

Transcription

\[ \text{DNA} \rightarrow \text{RNA} \]

RNA Polymerase (450 kd) \textit{(1960, Hurwitz; Weiss)}

\(\alpha_2\beta\beta'\omega\sigma\) (holoenzyme)

- Recognizes promoter &
- Initiates synthesis
- Forms phosphodiester bond
- Binds rNTPs

\(\alpha_2\beta\beta'\omega\) (core enzyme)

\begin{table}[h]
\centering
\caption{Additional classes of RNA}
\label{table:37.2}
\begin{tabular}{lll}
\hline
RNA & Size (nucleotides) & Function \\
\hline
Small nuclear RNA (snRNA) & Less than 300 & Components of RNA splicing machinery \\
Small nucleolar RNA (snoRNA) & Less than 300 & rRNA biogenesis and modification \\
MicroRNA (miRNA) & 20–25 & Regulates use of mRNA \\
Small interfering RNA (siRNA) & 20–25 & Antiviral defense mRNA degradation \\
Piwi-interacting RNA (piRNA) & 29–30 & Gene regulation \\
Long noncoding RNA (lncRNA) & Greater than 200 & Gene regulation \\
\hline
\end{tabular}
\end{table}
See Fig 36.1  Structure of *T. aquaticus* RNA pol core enzyme

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Gene</th>
<th>Number</th>
<th>Mass (kd)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td><em>rpoA</em></td>
<td>2</td>
<td>37</td>
<td>Required for assembly of core enzyme; interacts with regulatory factors</td>
</tr>
<tr>
<td>β</td>
<td><em>rpoB</em></td>
<td>1</td>
<td>151</td>
<td>Takes part in all stages of catalysis</td>
</tr>
<tr>
<td>β’</td>
<td><em>rpoC</em></td>
<td>1</td>
<td>155</td>
<td>Binds to DNA; takes part in catalysis</td>
</tr>
<tr>
<td>ω</td>
<td><em>rpoZ</em></td>
<td>1</td>
<td>10</td>
<td>Required to restore denatured polymerase to its native form</td>
</tr>
<tr>
<td>σ70</td>
<td><em>rpoD</em></td>
<td>1</td>
<td>70</td>
<td>Takes part in promoter recognition</td>
</tr>
</tbody>
</table>

Table 36.1  Subunits of *E. coli* RNA polymerase

*Rockefeller A Short Course, Second Edition
© 2013 W.H. Freeman and Company*
RNA Synthesis

1) Initiation
2) Elongation
3) Termination

RNA polymerase is responsible for these function in RNA synthesis

*E. coli* RNA polymerase synthesizes all major types of RNA:
- mRNA
- tRNA
- rRNA
- small RNA

Transcription in *E. coli*

1) RNA polymerase searches for initiation sites (~2000 in 4,000,000 bp)
2) Unwinds DNA to produce single-stranded template
3) Selects correct ribonucleotide and catalyzes the formation of phosphodiester bonds (totally processive)
4) Detects termination signals
5) Interacts with activators & repressor proteins
Requirements for Transcription

- DNA template
- RNA polymerase
- Transcription factors
- NTPs (ATP, CTP, GTP, UTP)
- Mg++

**Note:** RNA Polymerase does NOT require a primer. RNA chains can be initiated de novo.

Mechanism of elongation is the same as for DNA Polymerase: nucleophilic attack by 3’-OH on α-phosphate of NTP

- RNA polymerase reaction
**E. coli** transcription rate: 30-85 nucleotides/sec

Error rate = 10⁻⁶

RNA Polymerase has no exonuclease activity

---

**Orientation of a gene**
mRNA versus Template and Coding strands of DNA

5’-GGCGGAACGCGCGAGUUAUCCCACAGGCCAGGUUCCGCUUGCGCGCUU-3’
3’-CCGCGGTTGCCGCTTAAAGGCTGCTCGGATCGAGGCAACGGCCA-5’
5’-GGCGCGACGCGCGAATTATCCCAAGGCGGAGTTTGGCTGCGGCGAC-3’

mRNA
Template (antisense) strand of DNA
Coding (sense) strand of DNA

Figure 36.2
Biochemistry: A Short Course, Second Edition
© 2013 W.A. Freeman and Company

Transcription is initiated at promoter sites on the DNA template

Promoter sequences for housekeeping genes from 10 bacteriophage and bacterial genes (coding strand)

See Fig. 36.4

Strong promoters correspond to consensus sequence (once in 2 sec)
Weak promoters have substitutions (~ once in 10 minutes)
The subunit of RNA polymerase is responsible for specific initiation of transcription.

σ recognizes promoter sequences.

σ70 recognizes promoters of house keeping genes.

In eukaryotes, transcription factors are required for formation of transcription complex.
**E. coli σ subunits**

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Gene</th>
<th>Genes transcribed</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ³⁰⁰</td>
<td>rpoD</td>
<td>Many</td>
<td>TTGACA</td>
</tr>
<tr>
<td>σ⁴⁴</td>
<td>rpoN</td>
<td>Nitrogen metabolism</td>
<td>None</td>
</tr>
<tr>
<td>σ¹⁸</td>
<td>rpoS</td>
<td>Stationary phase</td>
<td>?</td>
</tr>
<tr>
<td>σ²²</td>
<td>flaI</td>
<td>Flagellar synthesis and chemotaxis</td>
<td>TAA</td>
</tr>
<tr>
<td>σ¹²</td>
<td>rpoH</td>
<td>Heat shock</td>
<td>CTTGAA</td>
</tr>
<tr>
<td>σ³⁵⁵²</td>
<td>gene 55</td>
<td>Bacteriophage T4</td>
<td>None</td>
</tr>
</tbody>
</table>

⁷N represents any nucleotide.
Closed promoter complex shifts to open complex, 18bp DNA is unwound, RNA chain is initiated with pppA (or pppG) at 5’ terminus.

~10 ribonucleotides are added in 5’→3’ direction. σ falls off to core enzyme. NusA binds core enzyme → elongation mode.

Transcription pauses at poly G/C’s & is stopped at termination signals: RNA hairpin followed by several U’s; or Rho protein.

Formation of an RNA hairpin:

Dyad symmetry

5’~ACCTGCTCAGGACCTCTTGAGCAACCT~3’
3’~TGAGACGCTGACTGATGATA~5’

5’~AUCUCAGGAACCUUCUGAGCACAUC~3’

RNA hairpin

See Fig. 36.6

See Fig. 36.11
• Rho-dependent termination of transcription (E. coli)
• RNA pol is stalled at pause site
• Rho binds to new RNA, destabilizes RNA-DNA hybrid
Prokaryote RNA Polymerase synthesizes
mRNA, tRNA and rRNA

Eukaryotes have three RNA Polymerases
RNA Pol I: 18S, 5.8S, 28S rRNA
RNA Pol II: mRNA
RNA Pol III: tRNA, 5S rRNA, small RNA
**Housekeeping genes:**
* encode proteins required for basic metabolism
* have strong or weak promoters depending upon level of protein required

**Regulated (differentially expressed) genes:**
* often regulated at level of transcription

**Activators:** regulatory proteins that bind DNA & increase rate of transcription of weak promoters.
* Often interact with RNA polymerase → increase RNA polymerase binding or increase rate of transcription bubble formation (opening) or increase rate of primer formation

activators can be allosterically regulated
**Repressors**: regulatory proteins that bind DNA and repress transcription

Repressors have many mechanisms for repressing transcription.

These include:
* preventing RNA polymerase from binding promoter
* inhibition of initiation reactions (e.g. transition bubble formation, primer synthesis, promoter clearance)

Repressors are allosterically regulated**

**inducer**: ligands that bind to, and inactivate, repressors

**corepressor**: ligands that bind to, and activate, repressors

---

**Ligand activation of an Activator**

- Strategies for regulating transcription initiation by regulatory proteins

(a) An activator with bound ligand stimulates transcription.
Ligand inactivation of an Activator

(b) An activator stimulates transcription. In the presence of ligand, the activator is inhibited.

A repressor is inactivated by binding of an inducer

(c) A repressor prevents transcription. Binding of ligand (inducer) to the repressor inactivates the repressor and allows transcription.
A corepressor allows a repressor to repress transcription

(d) In the absence of ligand, the repressor does not bind to DNA. Repression occurs only when ligand (corepressor) is present.

Organization of the genes for proteins required to metabolize lactose: the Lac Operon

- Three coding regions of lac operon are cotranscribed from the $P_{lac}$ promoter
- $lacI$ is under the control of the $P$ promoter

Lac repressor

$P_{lac}$

$P$

$\beta$-galactosidase

Lactose permease

Thiogalactoside transacetylase

Operators: repressor binding sites

Recall lactose = Gal$\beta$-1,4-Glc

The Lac operon encodes a poycistronic mRNA

See Fig. 36.17
Figure 36.16

**Lactose removed**

**Lactose added**

**$\beta$-Galactosidase ($\mu$g)**

**Total bacterial protein ($\mu$g)**

Fig. 36.16

---

Figure 36.18

**Repressor–inducer complex does not bind DNA**

Fig. 36.18
Binding of *lac repressor* to the *lac operon*

- Tetrameric *lac* repressor interacts simultaneously with two sites near the *lac* promoter
- DNA loop forms
- RNA pol can still bind to the promoter
- Binding of the *lac* repressor to the *lac* operon is inhibited by the inducer *allolactose*

---

- Formation of allolactose from lactose

- **NOTE:** *β*-galactosidase catalyzes both hydrolysis of lactose and formation of some allolactose
<table>
<thead>
<tr>
<th>Carbon Source</th>
<th>Relative transcription from lac Operon</th>
<th>Reason?</th>
</tr>
</thead>
<tbody>
<tr>
<td>bacteria</td>
<td>Escape synthesis (very low level)</td>
<td></td>
</tr>
<tr>
<td>bacteria</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>bacteria</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 36.19**

*Binding site for catabolite activator protein (CAP)*

CAP-cAMP (same as CRP-cAMP) stimulates transcription at the lac operon.
• Transcription from the Lac promoter occurs in the presence of the inducer allolactose

• However, transcription is further activated by binding of cAMP to the activator CRP-cAMP which converts relatively weak promoter to a strong promoter

CRP = cAMP Regulatory Protein

cAMP is produced in when glucose is limiting

(b) RNA polymerase holoenzyme binds to the promoter and also contacts the bound activator, which increases the rate of transcription initiation.
cAMP production

- In the absence of glucose, enzyme III (EIII) transfers a phosphate group to adenylate cyclase leading to CRP-cAMP increases
- CRP-cAMP activates transcription of other genes
- See Fig. 21.21 & 21.22 & 21.23

In the presence of glucose the phosphate is transferred to glucose

Phosphoenolpyruvate-dependent sugar phosphotransferase system

Conformational changes in CRP caused by cAMP binding

- α-Helices of each monomer of the cAMP-CRP dimer fit into major groove of DNA
Structure of CRP-cAMP and DNA complex

- Both subunits have a cAMP bound at the allosteric site
- Each subunit has an α-helix in DNA major groove

Common eukaryotic promoter elements

Fig. 37.3

Eukarotes have different mechanism for recognition of cis-acting elements than bacteria

The eukaryote TATA, CAAT and GC boxes & other cis-acting elements are recognized by proteins other than RNA polymerase.

Cis-acting element:
DNA sequences that regulate expression of gene located on same DNA molecule

Inr: initiator element

DPE: downstream core promoter element

RNA polymerase II promoter

Enhancer
-24 to -32

Inr
-3 to +5

TATA box

Promoter

or

Enhancer

Inr

DPE

Promoter

+28 to +32

RNA polymerase I promoter

-150 to -200

28S

UPE

rnr

18S

28S

Promoter

RNA polymerase III promoter

Type I: 5S rRNA

Type II: tRNA

A block

C block

A block/B block

Figure 37.3

© 2013 W.H. Freeman and Company
Transcription initiation in eukaryotes: basal transcription apparatus

TFII: transcription factor for RNA Polymerase II

TBA: TATA-box-binding protein

CTD: carboxyl-terminal domain of RNA Pol II

TFIIH: opens double helix and phosphorylates CTD, allowing RNA Pol II to leave promoter; marks change from initiation to elongation phase

Trans acting elements: proteins that recognize cis-acting elements and regulate RNA synthesis; commonly called transcription factors

Enhancer: cis-acting element that can be present on either DNA strand and can exert stimulatory action over distances of thousands of base pairs

Most eukaryotic transcription factors (TFs) interact with multiple proteins to form large complexes that interact with transcriptional machinery to activate or repress transcription.

Mediator: complex of 25-30 subunits that interact with transcription machinery before initiation begins. Acts as a bridge between enhances and promoter-bound RNA Pol II
37.3 Gene Expression Is Regulated by Hormones

Steroid hormones are powerful regulatory molecules that control gene expression.

Estradiol controls the genes in the development of female secondary sex characteristics.

Estradiol exerts its effects by forming a complex with a specific receptor protein called the estrogen receptor.

The estrogen receptor is part of a larger class of regulatory proteins called nuclear hormone receptors, all of which are activated by binding of small molecules or ligands.

Structure of two nuclear hormone-receptor domains Fig. 37.10
Ligand binding (estradiol) to nuclear hormone-receptor (estradiol receptor) Fig. 37.11

Binding of ligand to nuclear hormone-receptor induces conformation change that generates site for binding of coactivator. Fig. 37.12
Ligands that activate a nuclear hormone receptor are called agonists, while ligands that inhibit the receptor are called antagonists.

Some cancers dependent on action of estradiol-receptor complex. Growth of these cancers can be slowed by administering receptor antagonists, such as tamoxifen and raloxifene.

Such antagonist to the estrogen receptor are called selective estrogen receptors modulators (SERMS).

Coactivators can stimulate transcription by loosening the interaction between histones and DNA, making the DNA more accessible to the transcription machinery.

A common means of weakening the interaction of histones with DNA is by acetylation of histones on specific lysine residues.

Histones can also be modified by other means, such as methylation and phosphorylation.
ATP-citrate lyase located in nucleus generates acetyl CoA that is used by histone acetyltransferases (HATS) to modify histones.

HATs are components of coactivators or are recruited by coactivators.

Acetylation reduces affinity of histones for DNA and generates a docking site for other components of transcription machinery. These components bind to the acetylated histones at acetyllysine binding domains also called bromodomains.

Bromodomains are also present in chromatin-remodeling machines, ATP-powered complexes that make DNA in chromatin more accessible.

---

**Structure of histone acetyltransferase.**

Fig. 37.13

Histone acetylation activates transcription by:

1. reducing affinity of histones for DNA
2. Recruiting components of transcription machinery
3. Initiating remodeling of chromatin structure
Acetyllysine residues in histones interact with acetyllysine-binding domain called **bromodomain** (present in many TFs).

**TAF**: TATA-box-binding protein associated factors (has pair of bromodomains);

**Chromatin-remodeling engines**: use hydrolysis of ATP to shift position of nucleosomes; can open up regions of chromatin for transcription
The acetylation of histones is not an irreversible reaction. Genes may need to be expressed at certain times and then be repressed.

The acetylation of histones is not an irreversible reaction. Genes may need to be expressed at certain times and then be repressed.

Histone deacetylases catalyze the removal of acetyl groups from histones, resulting in the inhibition of transcription.

All covalent modifications of histone are reversible.

---

**Table 37.3 Selected histone modifications**

<table>
<thead>
<tr>
<th>Histone modified*</th>
<th>Amino acid modification</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>H4 K8</td>
<td>Acetylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H3 K14</td>
<td>Acetylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H3 K27</td>
<td>Methylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H3 R17</td>
<td>Methylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H2B S14</td>
<td>Phosphorylation</td>
<td>Apoptosis initiation</td>
</tr>
</tbody>
</table>

*The letters K, R, and S are the single-letter abbreviations for lysine (K), arginine (R) and serine (S). The numbers with these letters refer to the location in the primary structure.

---

37.4 Histone Acetylation Results in Chromatin Remodeling

Histone Deacetylases Contribute to Transcriptional Repression

---

Table 37.3

Biotechnology A Short Course, Second Edition
Many primary transcripts must be further processed to be active. Such transcripts include: tRNA, rRNA and mRNA in eukaryotes

Types of transcript processing
1. removal of nucleotides
2. addition of nucleotides
3. covalent modification of nucleotides

Thus, in some cases the mature transcript includes different bases or modifications NOT encoded by the corresponding gene!!

Fig. 36.13

Primary bacterial transcript prior to processing

16S rRNA  tRNA  23S rRNA  5S rRNA

Fig. 36.13
38.2 Transfer RNA Is Extensively Processed

RNA polymerase III catalyzes the synthesis of precursors to tRNA.

RNase P removes nucleotides from the 5' end of the precursor, while the CCA-adding enzyme adds nucleotides to the 3' end.

Bases and riboses are also modified.

Many eukaryotic tRNA precursors contain an intron that is removed by an endonuclease, and the resulting products are joined by a ligase.
Examples of base modifications in RNA. Fig. 36.14

(A) 6-Dimethyladenylate

(B) Uridylate

Ribothymidylyate

Pseudouridylylate
mRNA Processing

Prokaryote mRNA is NOT further processed. 1° transcript is directly translated

Eukaryote mRNA IS processed: cleavage, covalent modification, addition of nucleotides & splicing

mRNA processing steps:

5’ capping; 3’ polyadenylation; splicing

Covalent modification of the ends of the transcript increases RNA stability.

Formation of 5’ cap on eukaryotic mRNA
5’ capping (7-methylguanylate-5'ppp-5’-mRNA)

See Fig. 38.3
• Guanylate base is methylated at N-7, Step (3)
• 2-Hydroxyl groups of last two riboses may also be methylated
3' polyadenylylation of eukaryotic mRNA:

* cleavage 10-20 nucleotides downstream from polyadenylation site (AAUAAA) (CPSF binds consensus & endonuclease binds CPSF)
* polyadenylation ($\leq$ 250 As) by poly A polymerase $\rightarrow$ poly A tail

(a) Polyadenylation begins when RNA polymerase II transcription complex synthesizes a polyadenylation signal at the 3' end of an mRNA precursor.

(b) CPSF binds to the consensus sequence and forms a complex containing an RNA endonuclease. The endonuclease catalyzes cleavage of the transcript downstream of the polyadenylation sequence, forming a new 3' end. Poly A polymerase can then bind to the end of the mRNA precursor.

CPSF = cleavage & polyadenylation specificity factor
(c) The endonuclease dissociates and the new 3' end of the RNA is polyadenylated by the activity of poly A polymerase.
Triose phosphate isomerase gene (nine exons and eight introns)

Many eukaryote genes have exons & introns

>3400 bp

[Diagram of Triose phosphate isomerase gene showing nine exons and eight introns.]

Consensus sequences at splice sites in vertebrates

GU_______A_________AG

Exon

Intron

Exon

Splice site

GU

AG

Exon

Intron

Exon

5' splice site consensus sequence

Branch site consensus sequence

3' splice site consensus sequence

See Fig. 38.6
Splicing takes place on ____________: complexes of 45 proteins & 5 RNAs called small nuclear RNA (snRNA): U1, U2, U4, U5, U6

snRNA associates with proteins → small nuclear ribonucleoproteins (snRNPs).
Intron removal in mRNA precursors

(a) The spliceosome positions the adenylate residue at the branch site near the 5′ splice site. The 2′-hydroxyl group of the adenylate attacks the 5′ splice site.

(b) The 2′-hydroxyl group is attached to the 5′ end of the intron, and the newly created 3′-hydroxyl group of the exon attacks the 3′ splice site.
(c) As a result, the ends of the exons are joined, and the intron, a lariat-shaped molecule, is released.

Formation of a spliceosome

(a) As soon as the 5' splice site exits the transcription complex, a U1 snRNP binds to it.
(b) Next, a U2 snRNP binds to the branch site within the intron.

(c) When the 3' splice site emerges from the transcription complex, a U5 snRNP binds, and the complete spliceosome assembles around a U4/U6 snRNP.
Alternative splicing is a powerful mechanism for expanding protein diversity.

In alternative splicing, a pre-mRNA can be spliced in different patterns, generating proteins with different functions.

**38.3 Messenger RNA Is Modified and Spliced**

**Clinical Insight**

Most Human Pre-mRNAs Can Be Spliced in Alternative Ways to Yield Different Proteins

Alternative splicing can generate a membrane-bound versus soluble antibody. **Fig. 38.10**
Transcription and splicing are coordinated by the carboxyl-terminal domain (CTD) of RNA polymerase II.

Functions of the CTD include:

1. Recruiting enzymes to synthesize the 5’ cap.
2. Recruiting components of the splicing complex.
3. Recruiting an endonuclease that cleaves the pre-mRNA to expose the site for poly A addition.

The CTD: coupling transcription to premRNA processing. Fig. 38.11
38.4 RNA Can Function As a Catalyst

Some RNAs, called ribozymes, function as catalysts.

Introns from certain organisms can self-splice—that is, excise themselves.

Self-splicing or Group 1 introns were initially identified in rRNA from Tetrahymena.

Group 1 introns require guanosine as a cofactor.
mRNA Processing

Prokaryote mRNA is NOT further processed. 1º transcript is directly translated

Eukaryote mRNA IS processed: cleavage, covalent modification, addition of nucleotides & splicing

mRNA processing steps
1. covalent modification of the ends of the transcript increases RNA stability.
   a) 5'end modification: capping (7-methylguanylate-5'ppp-5'-mRNA)
   b) 3' end modification :
      * cleavage 10-20 nucleotides downstream from polyadenylation site (AAUAAA) (CPSF binds consensus & endonuclease binds CPSF)
      * polyadenylation (≤ 250 As) by poly A polymerase → poly A tail
2. **splicing**: removal of some internal pieces (introns) of the 1° transcript and rejoining of the remaining pieces (exons).

Junctions between introns and exons = **splice sites**. Splice sites have consensus 5', 3' and branch sequences required for splicing.

Splicing takes place on **spliceosomes**: complexes of 45 proteins & 5 RNAs called small nuclear RNA (snRNA): U1, U2, U4, U5, U6

snRNA associates with proteins → small nuclear ribonucleoproteins (snRNPs).