When researchers first sequenced the human genome, they were astonished by how few traditional genes encoding proteins were scattered along those 3 billion DNA bases. Instead of the expected 100,000 or more genes, the initial analyses found about 35,000 and that number has since been whittled down to about 21,000. In between were megabases of “junk,” or so it seemed.

“This year”, 30 research papers, including six in Nature and additional papers published by Science, sound the death knell for the idea that our DNA is mostly littered with useless bases. A decadelong project, the Encyclopedia of DNA Elements (ENCODE), has found that 80% of the human genome serves some purpose, biochemically speaking. “I don’t think anyone would have anticipated even close to the amount of sequence that ENCODE has uncovered that looks like it has functional importance,” says John A. Stamatoyannopoulos, an ENCODE researcher at the University of Washington, Seattle.

Beyond defining proteins, the DNA bases highlighted by ENCODE specify landing spots for proteins that influence gene activity, strands of RNA with myriad roles, or simply places where chemical modifications serve to silence stretches of our chromosomes. These results are going “to change the way a lot of [genomics] concepts are written about and presented in textbooks,” Stamatoyannopoulos predicts.
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 → Functional
• _______ - 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.
### Table 37.2 Additional classes of RNA

<table>
<thead>
<tr>
<th>RNA</th>
<th>Size (nucleotides)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small nuclear RNA (snRNA)</td>
<td>Less than 300</td>
<td>Components of RNA splicing machinery</td>
</tr>
<tr>
<td>Small nucleolar RNA (snoRNA)</td>
<td>Less than 300</td>
<td>rRNA biogenesis and modification</td>
</tr>
<tr>
<td>MicroRNA (miRNA)</td>
<td>20–25</td>
<td>Regulates use of mRNA</td>
</tr>
<tr>
<td>Small interfering RNA (siRNA)</td>
<td>20–25</td>
<td>Antiviral defense mRNA degradation</td>
</tr>
<tr>
<td>Piwi-interacting RNA (piRNA)</td>
<td>29–30</td>
<td>Gene regulation</td>
</tr>
<tr>
<td>Long noncoding RNA (IncRNA)</td>
<td>Greater than 200</td>
<td>Gene regulation</td>
</tr>
</tbody>
</table>

RNA Synthesis  (*E.coli*)

**Transcription**

DNA $\rightarrow$ RNA

RNA Polymerase  (450 kd)  (1960, Hurwitz; Weiss)

*recognizes promoter & initiates synthesis*

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

forms phosphodiester bond

binds rNTPs

(see Table 36.1 & Fig. 36.1)

$\alpha_2\beta\beta'\omega$  (core enzyme)
Table 36.1  Subunits of *E. coli* RNA polymerase

<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>rpoA</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>rpoB</td>
<td>1</td>
<td>151</td>
<td>Takes part in all stages of catalysis</td>
</tr>
<tr>
<td>β’</td>
<td>rpoC</td>
<td>1</td>
<td>155</td>
<td>Binds to DNA; takes part in catalysis</td>
</tr>
<tr>
<td>ω</td>
<td>rpoZ</td>
<td>1</td>
<td>10</td>
<td>Required to restore denatured polymerase to its native form</td>
</tr>
<tr>
<td>σ70</td>
<td>rpoD</td>
<td>1</td>
<td>70</td>
<td>Takes part in promoter recognition</td>
</tr>
</tbody>
</table>

See Fig 36.1  Structure of *T. aquaticus* RNA pol core enzyme

Scaffold for assembly, interact with regulatory proteins

α subunits

β subunit

ω subunit

β’ subunit

catalytic subunit

DNA binding

Table 36.1
Rockefeller University Press, 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 \textit{de novo}.

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

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

Error rate = $10^{-6}$

RNA Polymerase has no exonuclease activity

**BCMB 3100 - Chapter 36-38**

Transcription & RNA Processing

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

mRNA versus Template and Coding strands of DNA

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

Fig. 36.2
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

GTGCGTCTTGAATTTTA CTCCTGGCCTGATTACCTG TACGTACTAATGC CACGACGACCAAGGA AP2
GGCGGTCTTGAATTAAAT CCACTGGCCCTGATACCTGA GTCACATCAACGACGACG AP1
TGAGCTCTTGAATTTAAAT CATCGAATTATCTTACGTAG GACCAGGTTCTACGTAA rpo

CCCAGGCTTTACATTAT CGTCGGGCTTCGTATGTGT GTGAATTGTAGACGCG bac

CCCAGGCTTTACATTAT CGTCGGGCTTCGTATGTGT GTGAATTGTAGACGCG lacUV5

ATCTAAGGAAGCTCTTTAT ATCGAATTACCTCTACGTACGATTATCTCACGGATT TAT

TTCTCTTTGATCAGCGCGGT ATTAAATCCCTATTAGCGCCACTGACACGGAA rpoA0

TAAATGCTTTGACCTCTGAG CGGGAAGGCCATTTATGC ACCACCCGCGGCTGATAG rpoA2

TCCATGCTACACACTTACATTTATCGCATTATCGCTATCTGTGTTA ATTTACATACATAGCC gpdP1

TTATCAGGTCACTATTTCTGCATATTGTATTATGCCTATTAGCATGATGATG +1

Consensus sequence: TTGACA ** TATAAT **

−35 region

−10 region

Pribnow box/ TATA box

Strong promoters correspond to consensus sequence (once in 2 sec)

Weak promoters have substitutions (~ once in 10 minutes)

_________ 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 α subunits

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Gene</th>
<th>Genes transcribed</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ^{70}</td>
<td>rpoD</td>
<td>Many</td>
<td>TTGACA</td>
</tr>
<tr>
<td>σ^{54}</td>
<td>rpoN</td>
<td>Nitrogen metabolism</td>
<td>None</td>
</tr>
<tr>
<td>σ^{38}</td>
<td>rpoS</td>
<td>Stationary phase</td>
<td>?</td>
</tr>
<tr>
<td>σ^{28}</td>
<td>flaI</td>
<td>Flagellar synthesis and chemotaxis</td>
<td>TAAA</td>
</tr>
<tr>
<td>σ^{22}</td>
<td>rpoH</td>
<td>Heat shock</td>
<td>CTTGAA</td>
</tr>
<tr>
<td>σ^{55}</td>
<td>gene 55</td>
<td>Bacteriophage T4</td>
<td>None</td>
</tr>
</tbody>
</table>

<sup>6</sup>N represents any nucleotide.
TRANSCRIPTION PROBLEM  Complete by Breakout Session on October 15.

(a) The piece of prokaryotic DNA shown below has a promoter for a housekeeping gene that is recognized by $\delta 70$. Circle and name the two consensus regions of the promoter of the prokaryotic gene. (Hint: The underlined A in the coding strand is the transcription start site. Be sure to circle the correct strand.)

b) Transcribe the RNA that would be transcribed by *E.coli* RNA polymerase from the dsDNA shown below. Be sure to indicate the 5’ and the 3’ end of the transcript.

Coding strand
5’-GCTGGGTTGACACCTACTCAAACGGCTTATAATGGATTTGCAGCTACAT…3’
3’-CGACCCAACTGTGGATGAGTTTGCCGAATATTACTAAGTCGATGTA…5’

Template strand

---

**BCMB 3100 - Chapter 36-38**  
Transcription & RNA Processing

- Definition of gene
- RNA Polymerase
- Gene coding vs template strand
- Promoter
  - Transcription in *E. coli*
- Transcription factors
- mRNA processing
Initiation of transcription in *E. coli*

**Holoenzyme binds DNA**

- **RNA polymerase holoenzyme** binds nonspecifically to DNA.

**Searches for promoter by sliding along DNA (670 bp/sec)**

- The holoenzyme conducts a one-dimensional search for a promoter.

**Based on **\( \sigma \)** finds correct promoter**

- When a promoter is found, the holoenzyme and the promoter form a closed complex.

**Closed promoter complex shifts to open complex, 18bp DNA is unwound, RNA chain is initiated with pppA (or pppG) at 5’ terminus**

- A conformational change from the closed complex to an open complex produces a transcription bubble at the initiation site. A short stretch of RNA is then synthesized.

- ~10 ribonucleotides are added in 5’→3’ direction. \( \sigma \) 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

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
Fig. 36.9

Fig. 36.12
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

Table 37.1 Eukaryotic RNA polymerases

<table>
<thead>
<tr>
<th>Type</th>
<th>Location</th>
<th>Cellular transcripts</th>
<th>Effects of α-amanitin</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Nucleolus</td>
<td>18S, 5.8S, and 28S rRNA</td>
<td>Insensitive</td>
</tr>
<tr>
<td>II</td>
<td>Nucleoplasm</td>
<td>mRNA precursors and snRNA</td>
<td>Strongly inhibited</td>
</tr>
<tr>
<td>III</td>
<td>Nucleoplasm</td>
<td>tRNA and 5S rRNA</td>
<td>Inhibited by high concentrations</td>
</tr>
</tbody>
</table>

Fig. 37.2 Amanita phalloides
(death cap; destroying angel)
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 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

$lacI$  $lacZ$  $lacY$  $lacA$

Operators: repressor binding sites

Recall lactose = Gal-β-1,4-Glc

The Lac operon encodes a polycistronic mRNA

See Fig. 36.17
Fig. 36.16

Lactose removed

Lactose added

\[
\begin{align*}
\beta\text{-Galactosidase (}\mu\text{g}) & \\
\text{Total bacterial protein (}\mu\text{g})
\end{align*}
\]

Fig. 36. 16

Fig. 36.18

Repressor–inducer complex does not bind DNA

\[
\begin{align*}
\text{lac mRNA} & \\
\beta\text{-Galactosidase} & \\
\text{Permease} & \\
\text{Transacytase}
\end{align*}
\]
Binding of \textit{lac} repressor to the \textit{lac} operon

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

See Fig. 36.18

- Formation of allolactose from lactose
- \textbf{NOTE:} $\beta$-galactosidase catalyzes both hydrolysis of lactose and formation of some allolactose

\textbf{This is a side reaction of the enzyme}
**Fig. 36.19**

**Binding site for catabolite activator protein (CAP)**

**CAP-cAMP**
(same as CRP-cAMP) stimulates transcription at the lac operon

**Carbon Source** | **Relative transcription from lac Operon** | **Reason?**
--- | --- | ---
**bacteria** | 1 | **Escape synthesis (very low level)**
**bacteria** | 50 | **Removal of “catabolic repression” by cAMP-bound activator (CRP-cAMP; cAMP response protein) (also known as catabolite activator protein (CAP))**
• **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** (also known as catabolite activator protein (CAP))

---

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

cAMP is produced in when glucose is limiting.
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

Phosphoenolpyruvate-dependent sugar phosphotransferase system

- Conformational changes in CRP caused by cAMP binding
- $\alpha$-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 $\alpha$-helix in DNA major groove

Common eukaryotic promoter elements

- Eukarotes have different mechanism for recognition of cis-acting elements than bacteria
- The eukaryote TATA & other cis-acting elements are recognized by proteins other than RNA polymerase.
- Promoters may have CAAT and GC boxes

**Enhancer**: DNA sequence with no promoter activity itself, but can greatly enhance activity of promoters, even over distance of 1000s of nucleotides

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

**Inr**: initiator element

**DPE**: downstream core promoter element
**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

**TFIIFH:** 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:** DNA sequence 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
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.

Nuclear hormone receptors bind to specific regions of the DNA called response elements. Thus, the estrogen receptor binds to the estradiol response element (ERE).

Nuclear hormone receptors have two highly conserved domains.

1. The DNA binding domain lies toward the center of the primary structure and is characterized by zinc-finger domains that confer specific DNA binding.

2. The ligand binding domain lies toward the carboxyl terminus of the primary structure. Ligand binding causes a structural change that enables the receptor to recruit other proteins to regulate transcription.
Structure of two nuclear hormone-receptor domains Fig. 37.9

Ligand binding (estradiol) to nuclear hormone-receptor (estradiol receptor) Fig. 37.10
Binding of ligand to nuclear hormone-receptor induces conformation change that generates site for binding of coactivator.  Fig. 37.11

Coactivator: proteins that bind to receptor only after is has bound ligand (i.e. steroid binds hormone)

Note: coactivators often referred to as p160 family because size of ~160 kd

Note#2: also examples where nuclear hormone-receptor members repress transcription in absence of corepressor proteins; in that case binding of ligand allows binding of corepressor proteins

Gene Expression Is Regulated by Hormones

Clinical Insight
Steroid-Hormone Receptors Are Targets for Drugs

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.
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.
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>H3 K9</td>
<td>Acetylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H3 K27</td>
<td>Acetylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H4 K16</td>
<td>Acetylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H3 K4</td>
<td>Monomethylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H3 K9</td>
<td>Trimethylation</td>
<td>Transcription repression</td>
</tr>
<tr>
<td>H3 K27</td>
<td>Trimethylation</td>
<td>Transcription repression</td>
</tr>
<tr>
<td>H3 R17</td>
<td>Methylation</td>
<td>Transcription activation</td>
</tr>
<tr>
<td>H2B S14</td>
<td>Phosphorylation</td>
<td>DNA repair</td>
</tr>
</tbody>
</table>

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

Tymoczko et al., Biochemistry: A Short Course, 4e, © 2019 W. H. Freeman and Company
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.

---

BCMB 3100 - Chapter 36-38
Transcription & RNA Processing

- Definition of gene
- RNA Polymerase
- Gene coding vs template strand
- Promoter
- Transcription in E. coli
- Transcription factors
- mRNA processing
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!!
In Eukaryotes 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.
Fig. 38.2 Conversion of yeast (eukaryote) tRNA precursor into mature tRNA

Examples of base modifications in RNA.  Fig. 36. 14

(A) 6-Dimethyladenylate

(B) Uridylate

Ribothymidylate

Pseudouridylate
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 (≤ 250 As) by poly A polymerase → 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 = cleavage & polyadenylation specificity factor

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.
(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

Consensus sequences at splice sites in vertebrates

See Fig. 38.6
Splicing takes place on ___________: complexes of >300 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.
RNA in U1 snRNP has 6 nucleotides that bp to 5' splice site

**U2 and U6 snRNAs form catalytic center of the spliceosome!**

U4 serves as initiator that masks U6 until specific splice sites are aligned

ATP-powered helicases unwind RNA duplex intermediates

---

**Fig. 38.7**

---

**U2 and U6 snRNAs form catalytic center of the spliceosome!**

---

**Fig. 38.8**
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.

Alternative splicing can generate a membrane-bound versus soluble antibody. **Fig. 38.10**
38.3 Messenger RNA Is Modified and Spliced

The Transcription and Processing of mRNA Are Coupled

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 pre-mRNA processing.

Recall: TFIIH of basal transcription apparatus phosphorylates CTD

*signals movement into elongation phase

*phosphorylated CTD binds capping, splicing and polyadenylation factors

*allows growing transcript to be processed

Figure 38.11

Biochemistry, A Short Course, Second Edition
© 2012 W.H. Freeman and Company
Another way to generate diverse proteins:

**RNA Editing**: change in nucleotide sequence of RNA after transcription by processes other than RNA splicing

One example:

*Apolipoprotein B (apo B)*; in liver, 4536 aa (512 kd), apo B-100, transports lipids synthesized by cell as part of lipoprotein

*In small intestine, apo B-48, (240 kd) N-terminal 2152 aa, carries dietary fat. C→U post transcription change, deamination of C by deaminase in small intestine introduces STOP codon (CAA→UAA)*

---

**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).