Transcription is the first stage in gene expression, and the principal step at which it is controlled.

- **Regulatory proteins** determine whether a particular gene is available to be transcribed by RNA polymerase.

- Transcription involves synthesis of an RNA chain representing one strand of a DNA duplex.
DNA vs. RNA

1. DNA is double-stranded; RNA is single-stranded.
2. DNA has deoxyribose sugars; RNA has ribose sugars.
3. DNA contains thymine; RNA contains uracil.

 RNA Replication vs. DNA Replication

RNA replication:
- Requires no priming
- Has many more initiation sites
- Is slower (50–100 b/sec vs. 1000 b/sec)
- Has lower fidelity
- Is more processive
A **transcription unit** is a region from the **promoter** to the **terminator**.

### RNA Replication: Initiation

- RNA synthesis begins at the **promoter**.
- **RNA polymerase** and accessory proteins assemble on DNA at this site.
- The first base of the synthesized RNA is **+1**.
RNA Replication: Promoters

Prokaryotic promoter consensus sequences:
-35 TTGACA ........-10 TATAAT ......+1

Eukaryotic promoter consensus sequences (pol II):
Enhancer ... / ... CCAAT ... / ... -25 TATAAT ...... +1

<table>
<thead>
<tr>
<th>Prokaryotes</th>
<th>Eukaryotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximal element(s)</td>
<td>Distal elements</td>
</tr>
<tr>
<td>Promoter</td>
<td>Proximal elements</td>
</tr>
<tr>
<td>Structural gene(s)</td>
<td>Structural gene</td>
</tr>
</tbody>
</table>

Transcription bubble

- RNA synthesis takes place within a "transcription bubble," in which DNA is transiently separated into its single strands, and the template strand is used to direct synthesis of the RNA strand.

- The RNA chain is synthesized from the 5' end toward the 3' end.

- The 3'-OH group of the last nucleotide added to the chain reacts with an incoming nucleoside 5' triphosphate. The incoming nucleotide loses its terminal two phosphate groups (γ and β); its α group is used in the phosphodiester bond linking it to the chain.
RNA Replication: Elongation

Core enzyme = $\alpha, \alpha_2, \beta, \beta'$

Holoenzyme = $\alpha, \alpha_2, \beta, \beta', \sigma$

Rho factor = $\rho$
The complete enzyme or holoenzyme in *E. coli* has a molecular weight of ~465 kD.

The β and β′ subunits together make up the catalytic center. The β subunit can be crosslinked to the template DNA, the product RNA, and the substrate ribonucleotides.

The α subunit is required for assembly of the core enzyme. When phage T4 infects *E. coli*, the α subunit is modified by ADP-ribosylation of an arginine. The α subunit plays a role in promoter recognition. The α subunit also plays a role in the interaction of RNA polymerase with some regulatory factors.

The σ subunit is concerned specifically with promoter recognition.

The holoenzyme (α2ββ′σ) can be separated into two components, the core enzyme (α2ββ′) and the sigma factor (the σ polypeptide).

Only the holoenzyme can initiate transcription.

Sigma factor ensures that bacterial RNA polymerase binds in a stable manner to DNA only at promoters. The sigma "factor" is usually released when the RNA chain reaches 8-9 bases, leaving the core enzyme to undertake elongation.

Core enzyme has the ability to synthesize RNA on a DNA template, but cannot initiate transcription at the proper sites.
Molecular Diagnostics  Fundamentals, Methods and Clinical Applications
Second Edition

- Sigma factor introduces a major change in the affinity of RNA polymerase for DNA. The holoenzyme has a drastically reduced ability to recognize loose binding sites—that is, to bind to any general sequence of DNA. The association constant for the reaction is reduced by a factor of ~10⁴, and the half-life of the complex is <1 second. So sigma factor destabilizes the general binding ability very considerably.

- Sigma factor also confers the ability to recognize specific binding sites. The holoenzyme binds to promoters very tightly, with an association constant increased from that of core enzyme by (on average) 1000 times and with a half-life of several hours.

---

Eukaryotic RNA Polymerases

<table>
<thead>
<tr>
<th>Type</th>
<th>Location</th>
<th>Products</th>
<th>α-amanitin</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Nucleolus</td>
<td>18S, 5.8S, 28S rRNA</td>
<td>Insensitive</td>
</tr>
<tr>
<td>II</td>
<td>Nucleoplasm</td>
<td>mRNA, snRNA</td>
<td>Inhibited</td>
</tr>
<tr>
<td>III</td>
<td>Nucleoplasm</td>
<td>tRNA, 5S rRNA</td>
<td>Inhibited by high concentration</td>
</tr>
<tr>
<td>Mit.</td>
<td>Cytoplasm</td>
<td>Mt RNA</td>
<td>Insensitive</td>
</tr>
</tbody>
</table>
Terminators are distinguished in *E. coli* according to whether RNA polymerase requires any additional factors to terminate in vitro:

- Core enzyme can terminate in vitro at certain sites in the absence of any other factor. These sites are called **intrinsic terminators**.

- Rho-dependent terminators are defined by the need for addition of **rho factor** (**ρ**) in vitro; and mutations show that the factor is involved in termination in vivo.

Two types of feature found in bacterial terminators

- Terminators in bacteria and their phages have been identified as **sequences** that are needed for the termination reaction (in vitro or in vivo).

- The responsibility for termination lies with the **sequences already transcribed** by RNA polymerase. So termination relies on scrutiny of the template or product that the polymerase is currently transcribing.

- Many terminators require a **hairpin** to form in the secondary structure of the RNA being transcribed. This indicates that termination depends on the **RNA product** and is not determined simply by scrutiny of the DNA sequence during transcription.
Intrinsic terminators have two structural features: a hairpin in the secondary structure; and a region that is rich in U residues at the very end of the unit.

Both features are needed for termination. The hairpin usually contains a G+C-rich region near the base of the stem. The typical distance between the hairpin and the U-rich region is 7-9 bases. There are ~1100 sequences in the E. coli genome that fit these criteria, suggesting that about half of the genes have intrinsic terminators.

Both the sequence of the hairpin and the length of the U-run influence the efficiency of termination.

mRNA Processing

DNA

Promoter 5' Exon 1 Intron 1 Exon 2 Intron 2 Exon 3 3'

Transcription

Intron 1

Splicing

mRNA

7 Me G AAAAA

Cap

PolyA tail

Copyright © 2012 F.A. Davis Company
Factors that control gene expression include:

- **Cis elements** are DNA sequences.
- **Trans elements** are proteins.

**Types of Gene Expression**

- **Constitutive**

  ![Constitutive Promoter Diagram]

  - Promoter
  - Coding sequence
  - ATG

- **Inducible**

  ![Inducible Promoter Diagram]

  - Promoter
  - ATG
  - Induction symbols
  - ATG
**Negative control**

- A classic mode of control in bacteria is negative: a **repressor** protein prevents a gene from being expressed.
- The "default state" for such a gene is to be expressed via the recognition of its promoter by RNA polymerase.
- Close to the promoter is another cis-acting site called the **operator**, which is the target for the repressor protein.
- When the repressor binds to the operator, RNA polymerase is prevented from initiating transcription, and **gene expression is therefore turned off**.

**Positive control**

- This is used in bacteria (probably) with about equal frequency to negative control, and it is the most common mode of control in eukaryotes.
- A **transcription factor** is required to assist RNA polymerase in initiating at the promoter.
- The typical default state of a eukaryotic gene is **inactive**: RNA polymerase cannot by itself initiate transcription at the promoter. Several trans-acting factors have target sites in the vicinity of the promoter, and **binding of some or all of these factors enables RNA polymerase to initiate transcription**.
The Lactose Operon

- Jacob and Monod discovered the molecular mechanism of inducible gene expression in E. coli (1960s).
- **Promoter (P):** site of RNA pol binding
- **Repressor protein (R):** trans element
- **Operator (O):** site of R binding
- **Inducer (I):** binds to repressor protein

The lac operon includes cis-acting regulator elements and protein-coding structural genes:

<table>
<thead>
<tr>
<th>DNA</th>
<th>mRNA</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>lac</td>
<td>Repressor</td>
</tr>
<tr>
<td>1040</td>
<td>82</td>
<td>j-lactosidase</td>
</tr>
<tr>
<td>O</td>
<td>I</td>
<td>Permease Transacetylase</td>
</tr>
<tr>
<td>3510</td>
<td>780</td>
<td></td>
</tr>
<tr>
<td>825</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Regulator gene:  
- P: promoter  
- O: operator  
- R: repressor

Structural genes:  
- LacZ: β-galactosidase  
- LacY: Permease  
- LacA: Transacetylase

The Lac Operon

- **Regulator gene (Off):**
  - R
  - (Off)

- **Inducer (On):**
  - R
  - (On)
Modes of Gene Regulation in Prokaryotes

- Induction:
  - Regulator gene
  - Inducer (On)

- Repression:
  - Regulator gene
  - Corepressor (Off)

- Activation:
  - Activator (On)

Types of RNA

- Messenger RNA (mRNA)
- Ribosomal RNA (rRNA)
- Transfer RNA (tRNA)
- Heteronuclear RNA (hnRNA)
  - Heteronuclear RNA is the immediate copy (transcription) of the coding regions of DNA (it is very short lived, it is processed into mRNA)
- Small nuclear RNA (snRNA)
  - They are transcribed by RNA polymerase II or RNA polymerase III and are involved in a variety of important processes such as RNA splicing (removal of introns from hnRNA), regulation of transcription factors or RNA polymerase II, and maintaining the telomeres.
- MicroRNA (miRNA)
  - A microRNA molecule has very few nucleotides (an average of 22) compared with other RNAs.
  - miRNAs are post-transcriptional regulators that bind to complementary sequences on target messenger RNA transcripts (mRNAs), usually resulting in translational repression or target degradation and gene silencing.
- Short interfering RNA (siRNA)
  - silencing RNA, is a class of double-stranded RNA molecules, 20-25 nucleotides in length, that play a variety of roles in biology. The most notable role of siRNA is its involvement in the RNA interference (RNAi) pathway, where it interferes with the expression of a specific gene.
- Double-stranded RNA (dsRNA)
  - Many short RNAs (snoRNA, tncRNA, etc.)
Epigenetics

- Non-sequence-specific, heritable traits
- Transcriptional gene silencing (TGS)
  - Imprinting
  - X-inactivation
  - RNA-induced transcriptional silencing (RITS)
- Post-transcriptional gene silencing (PTGS)
  - RNA-induced silencing complex (RISC)
  - G quartets
- Post-translational protein-protein interactions

Epigenetics: Genomic Imprinting (Silencing)

- Methylation of DNA inhibits transcription of some genes.
- Methylation usually occurs on cytosines or adenines.
  - 5-methyl cytosine
  - N-6 methyl adenine
  - N-4 methyl cytosine
- CpG islands are sites of methylation in human DNA.

CpG Island

...ggagggcgccggccgggacccagagaaaaa
Gccgcagggccggccggccggacccgcacccagacccggg
ccgagggcggg...
Histones affect accessibility of DNA to binding of RNA polymerase and transcription factors.

Epigenetics: The Histone Code

- Distinct combinations of chemical modifications of histones dictate varying responses.
  - Acetylation
  - Phosphorylation
  - Methylation
  - Ubiquitination
  - Ribosylation
  - Glycosylation
Epigenetics: Post-transcriptional Gene Silencing

**siRNA**

Small interfering RNA (siRNA) silences specific genes.

1. **Trigger ds RNA**
2. **Dicer**
3. **siRNA**
4. **Target mRNA**
5. **Direct cleavage**

**miRNA**

1. **Host encoded pri-miRNA**
2. **Drosha**
3. **Dicer**
4. **miRNA**
5. **Target mRNA**
6. **Translation block and cleavage**
Summary

- RNA carries genetic information from DNA to ribosomes for protein synthesis.
- Different types of RNA have different cellular functions.
- RNA is synthesized (transcribed) in a process similar to DNA replication.
- Transcription of RNA is constitutive or inducible. Inducible genes are expressed only when required.
- Epigenetic factors also control expression of genes.