

## LECTURE PRESENTATIONS

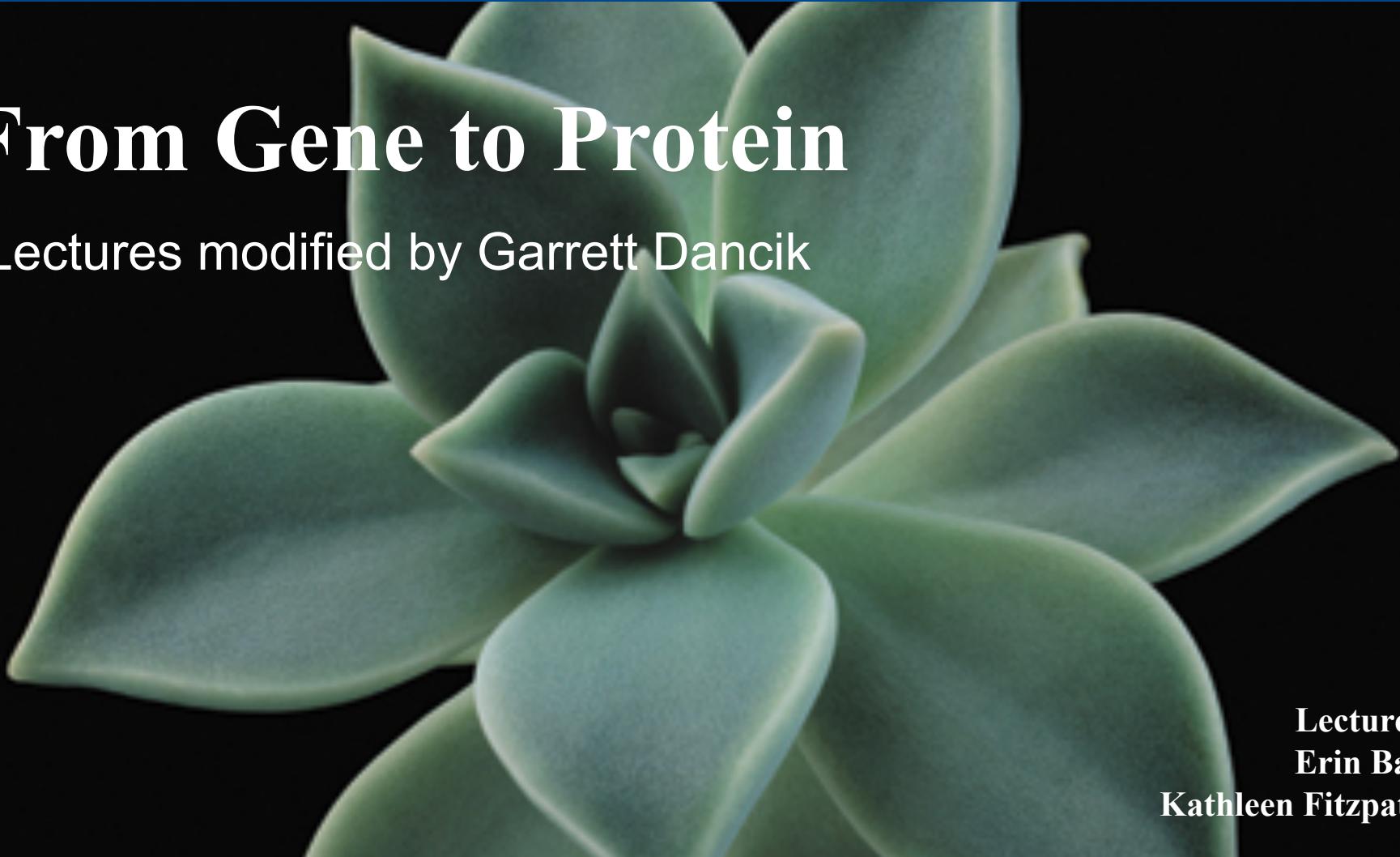
For CAMPBELL BIOLOGY, NINTH EDITION

Jane B. Reece, Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, Robert B. Jackson

# Chapter 17

## From Gene to Protein

Lectures modified by Garrett Dancik

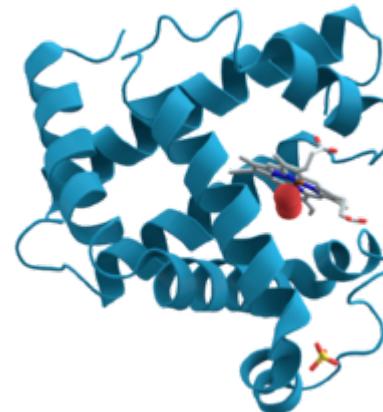
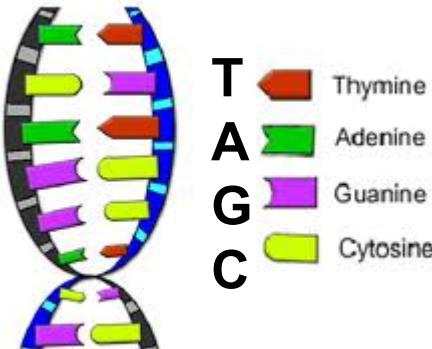


Lectures by  
Erin Barley

Kathleen Fitzpatrick

# Overview of gene expression

Central Dogma of  
Molecular Biology:



4-character alphabet → 20-character alphabet

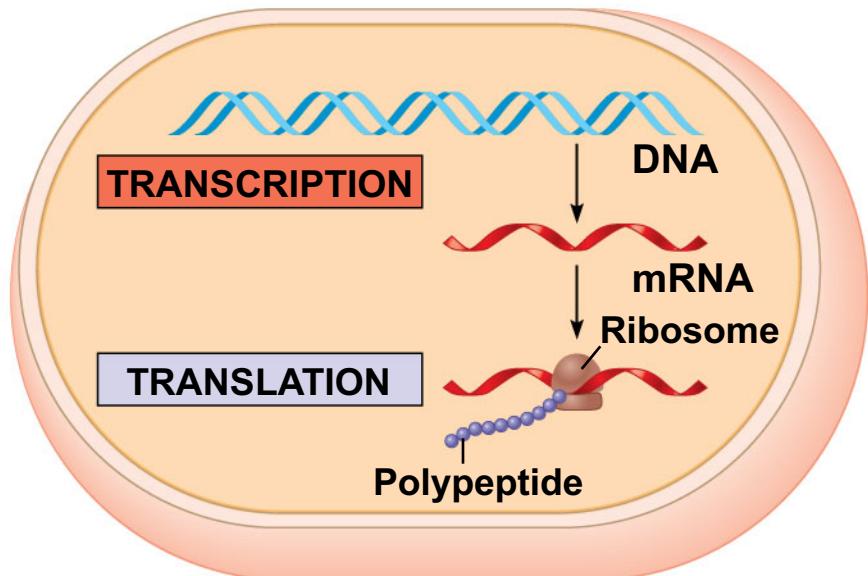
- Questions:
  - A gene (DNA) is physically located on what molecule?
  - Where in a cell is DNA stored?

# Basic Principles of Transcription and Translation

- RNA is the bridge between genes and the proteins for which they code
- **Transcription** is the synthesis of RNA under the direction of DNA
- Transcription produces **messenger RNA (mRNA)**
- **Translation** is the synthesis of a polypeptide, using information in the mRNA
  - What are the building blocks of polypeptides?
- Where does translation occurs?

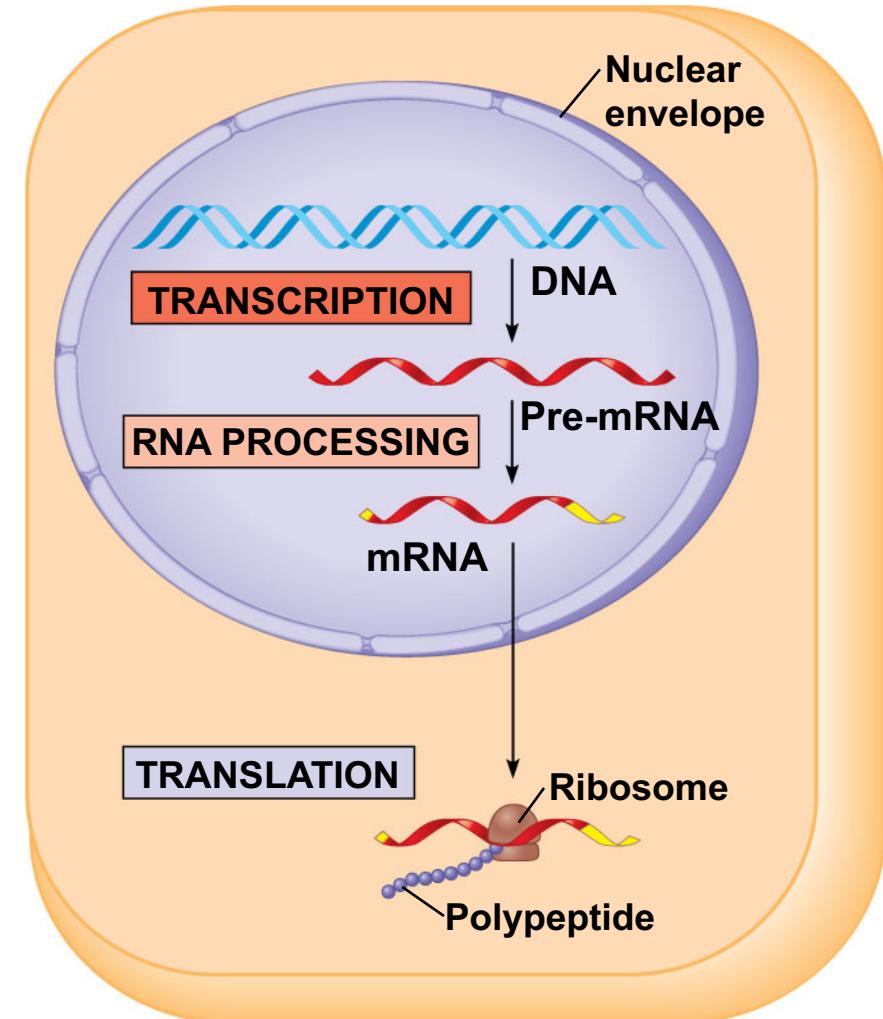
- A **primary transcript** is the initial RNA transcript from any gene prior to processing
- The *central dogma* is the concept that cells are governed by a cellular chain of command:  
DNA → RNA → protein

Figure 17.3



(a) Bacterial cell

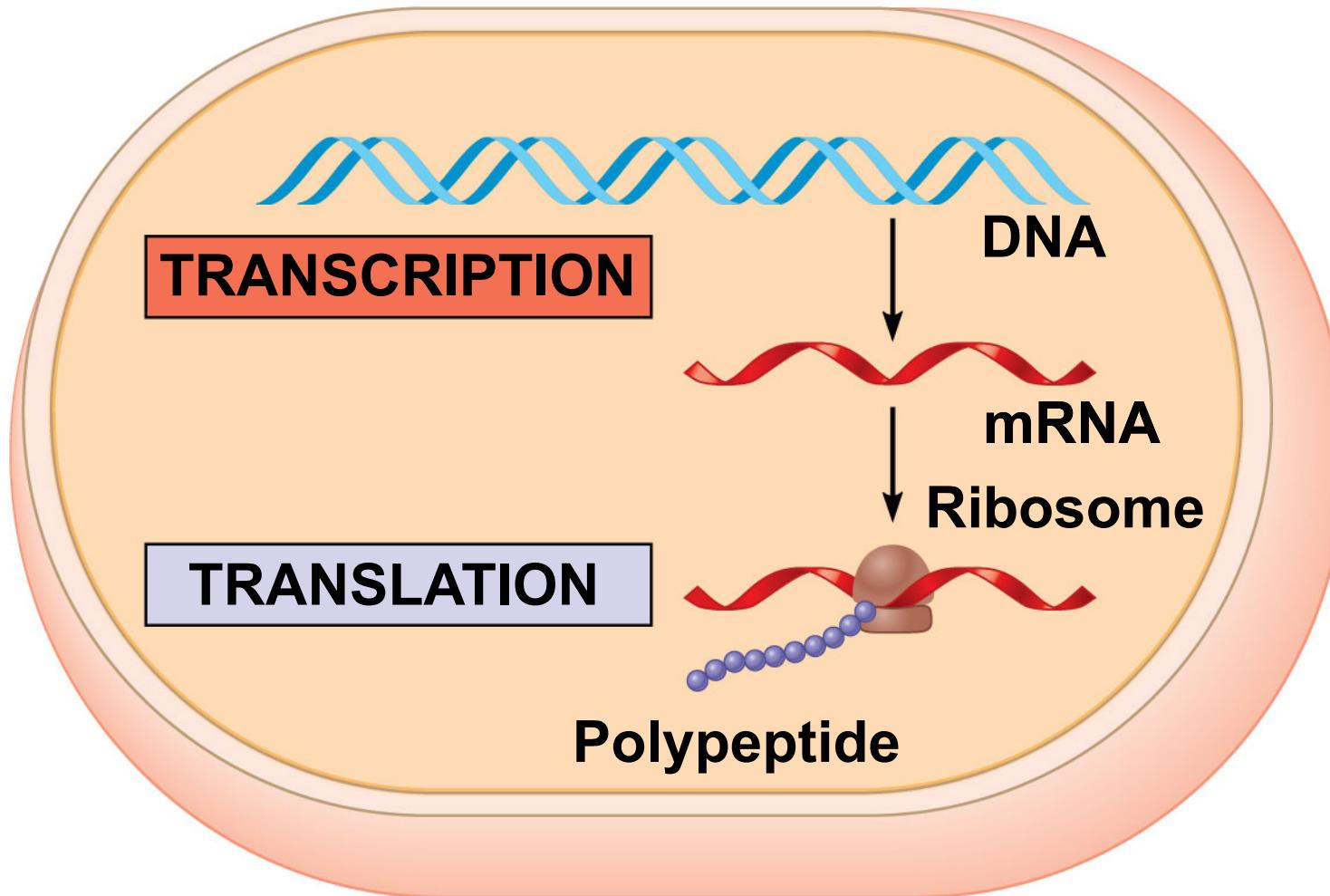
© 2011 Pearson Education, Inc.



(b) Eukaryotic cell

Figure 17.3a-1

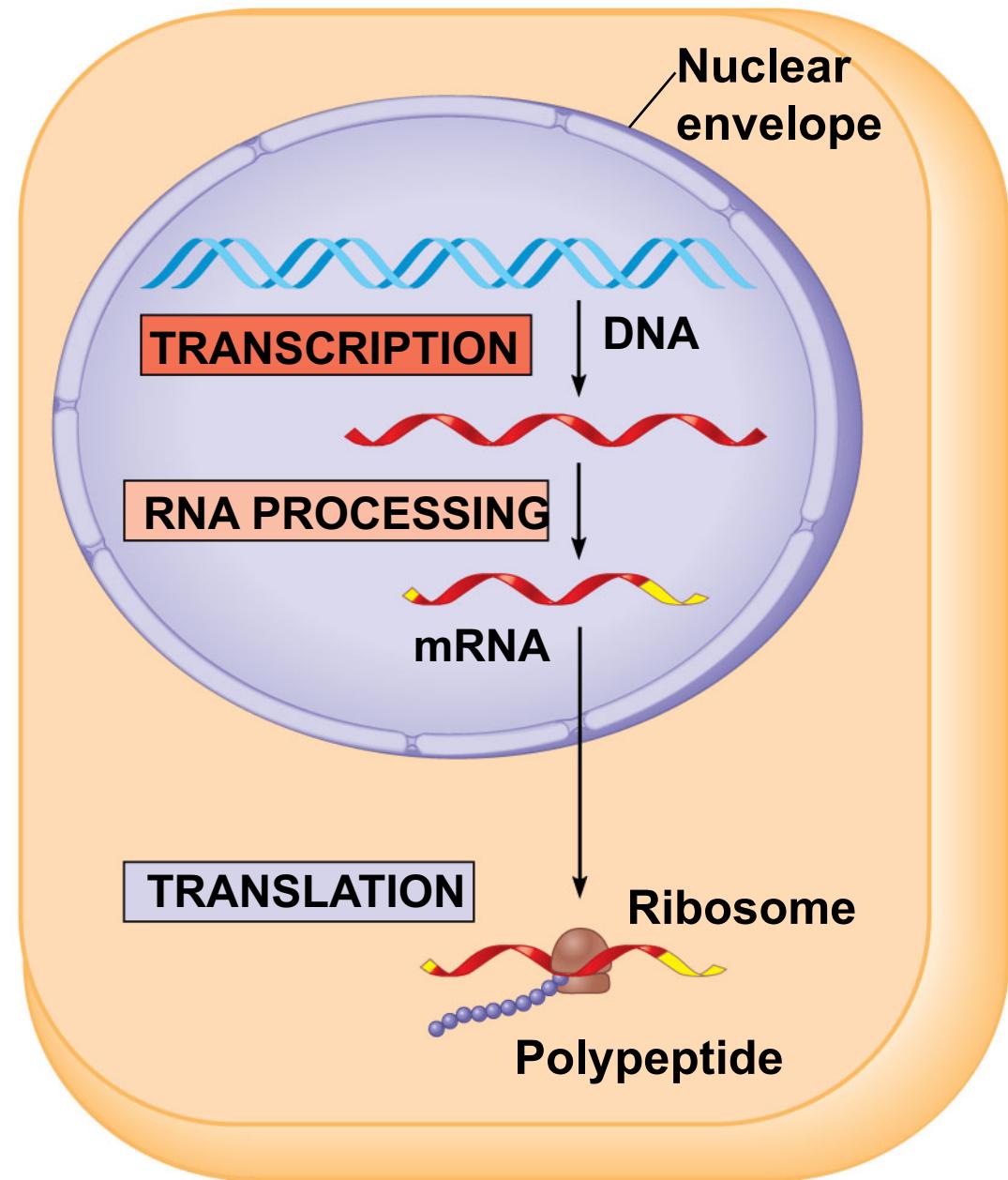
# Transcription and Translation in Prokaryotes



- In prokaryotes, translation of mRNA can begin before transcription has finished

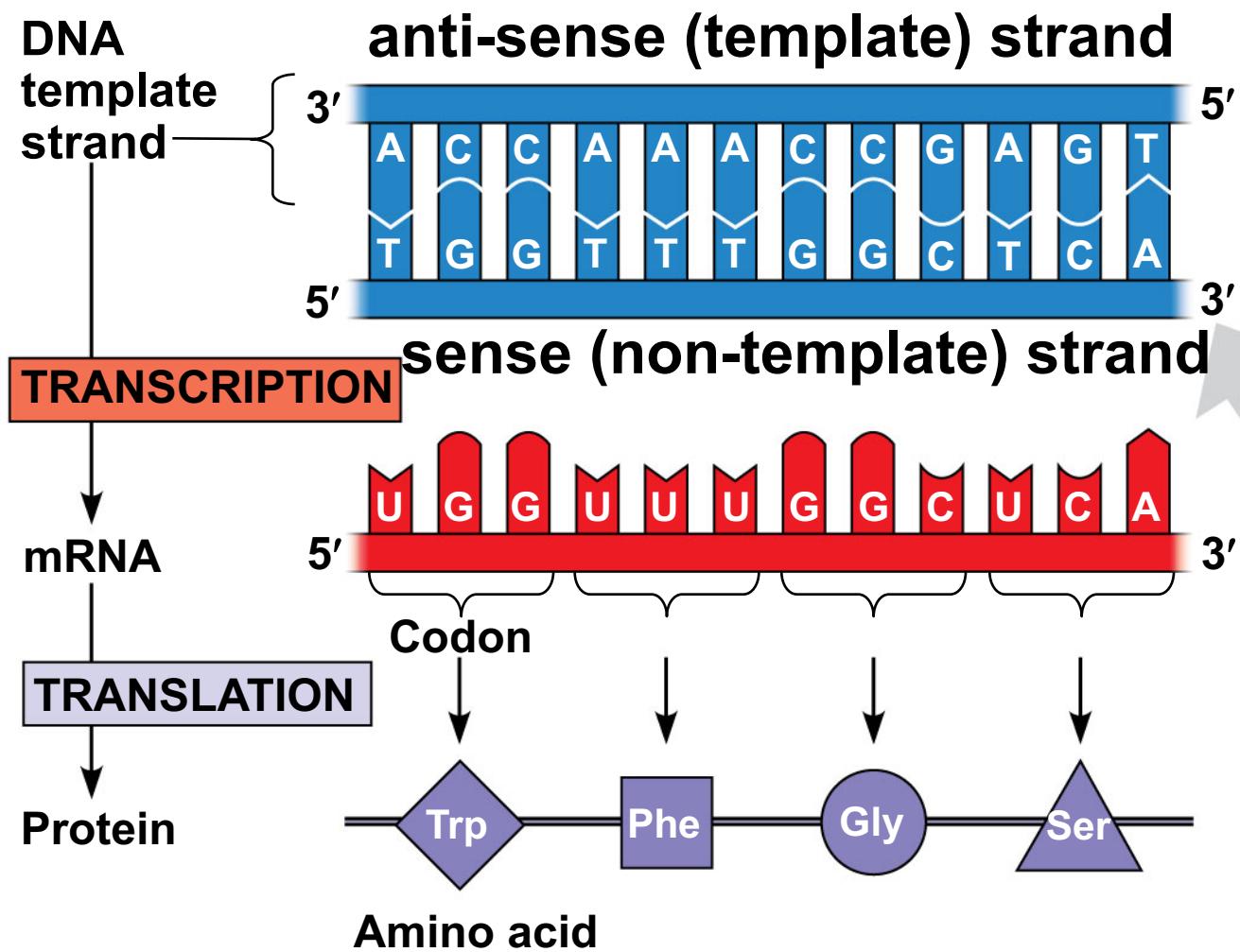
# Transcription and Translation in Eukaryotes

- In a eukaryotic cell, the nuclear envelope separates transcription from translation
- Eukaryotic RNA transcripts are modified through RNA processing to yield finished mRNA



# The Genetic Code

- The genetic code refers to the code (rules) that governs how a DNA sequence is converted to a protein (amino acid) sequence
- There are 20 amino acids, but only four nucleotide bases in DNA (and RNA)
  - What is the minimum number of nucleotides that could code for an amino acid?



- The genetic code is a triplet code where a 3-nucleotide DNA word codes for a 3-nucleotide mRNA word (a **codon**) which codes for an amino acid

- During transcription, one of the two DNA strands, called the **template strand**, provides a template for ordering the sequence of complementary nucleotides in an RNA transcript
- The template strand is always the same strand for a given gene
- During translation, the mRNA base triplets, called **codons**, are read in the 5' to 3' direction
- Each codon specifies the amino acid (one of 20) to be placed at the corresponding position along a polypeptide

- Codons along an mRNA molecule are read by translation machinery in the 5' to 3' direction
- Each codon specifies the amino acid (one of 20) to be placed at the corresponding position along a polypeptide

# *Cracking the Code*

- All 64 codons were deciphered by the mid-1960s
- Of the 64 triplets, 61 code for amino acids; 3 triplets are “stop” signals to end translation
- The genetic code is redundant (more than one codon may specify a particular amino acid) but not ambiguous; no codon specifies more than one amino acid
- Codons must be read in the correct **reading frame** (correct groupings) in order for the specified polypeptide to be produced

# Properties of the Genetic Code

- 64 triplets
  - 61 code for amino acids
  - 3 triplets are “stop” codons
- The genetic code is redundant but not ambiguous

Second mRNA base				Third mRNA base (3' end of codon)			
U	C	A	G	U	C	A	G
U	UUU Phe UUC UUA Leu UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	UCA G	UCA G	UCA G
C	CUU CUC CUA Leu CUG	CCU CCC CCA Pro CCG	CAU His CAC CAA Gln CAG	CGU CGC CGA Arg CGG	UCA G	UCA G	UCA G
A	AUU AUC Ile AUA AUG Met or start	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	UCA G	UCA G	UCA G
G	GUU GUC Val GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU GGC GGA Gly GGG	UCA G	UCA G	UCA G

# *The Universality of the Genetic Code*

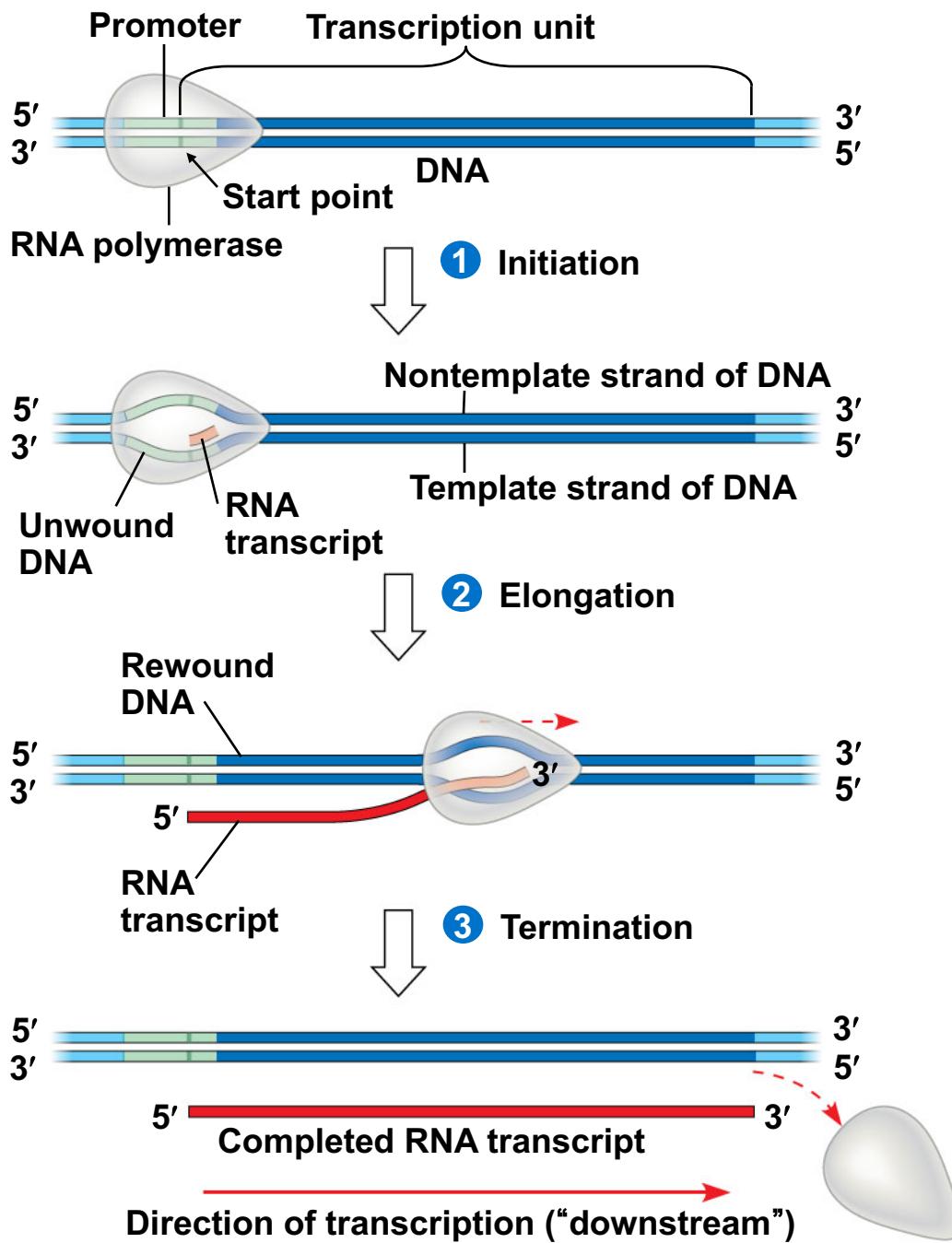
- The genetic code is nearly universal, shared by the simplest bacteria to the most complex animals
- Genes can be transcribed and translated after being transplanted from one species to another
- Examples:  
<http://blogs.scientificamerican.com/observations/2011/09/12/jellyfish-genes-make-glow-in-the-dark-cats/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4006694/>



**Pig expressing a jellyfish gene**

# Transcription is the DNA-directed synthesis of RNA: *a closer look*

- Transcription consists of three steps
  - Initiation
    - The DNA sequence where RNA polymerase attaches is called the **promoter**
  - Elongation
    - **RNA polymerase** pries the DNA strands apart and hooks together the RNA nucleotides
    - The RNA is complementary to the DNA template strand, but uracil replaces thymine
  - Termination
    - in bacteria, the sequence signaling the end of transcription is called the **terminator**

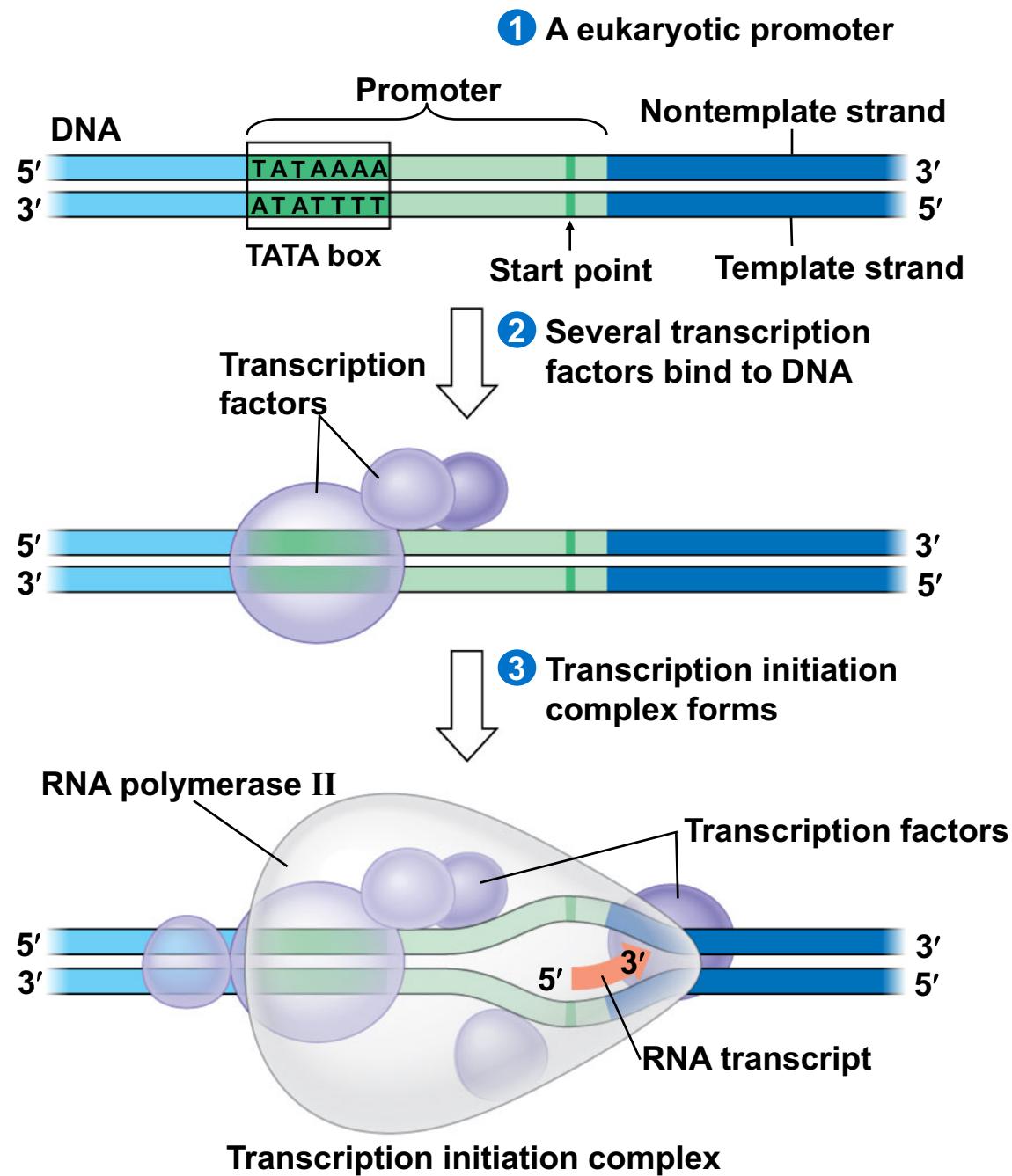


# *RNA Polymerase Binding and Initiation of Transcription*

- Promoters signal the transcriptional **start point** and usually extend several dozen nucleotide pairs upstream of the start point
- **Transcription factors** mediate the binding of RNA polymerase and the initiation of transcription
- The completed assembly of transcription factors and RNA polymerase II bound to a promoter is called a **transcription initiation complex**
- A promoter called a **TATA box** is crucial in forming the initiation complex in eukaryotes

Figure 17.8

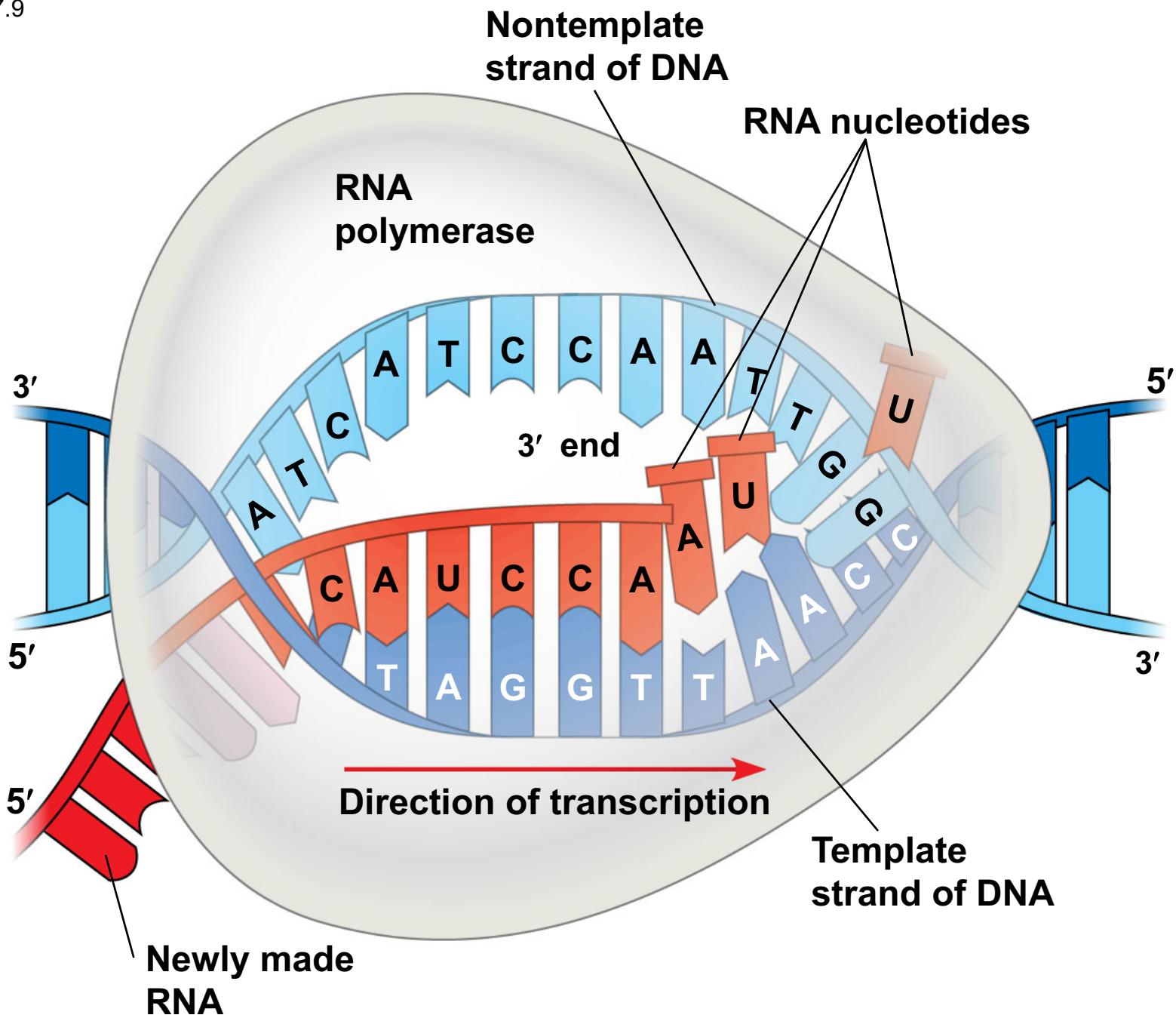
- Transcription in eukaryotes



# *Elongation of the RNA Strand*

- As RNA polymerase moves along the DNA, it untwists the double helix, 10 to 20 bases at a time
- Transcription progresses at a rate of 40 nucleotides per second in eukaryotes
- A gene can be transcribed simultaneously by several RNA polymerases
- Nucleotides are added to the 3' end of the growing RNA molecule

Figure 17.9



# *Termination of Transcription*

- The mechanisms of termination are different in bacteria and eukaryotes
- In bacteria, the polymerase stops transcription at the end of the terminator and the mRNA can be translated without further modification
- In eukaryotes, RNA polymerase II transcribes the polyadenylation signal sequence (AAUAAA); the RNA transcript is released 10–35 nucleotides past this polyadenylation sequence

# Eukaryotic cells modify RNA after transcription

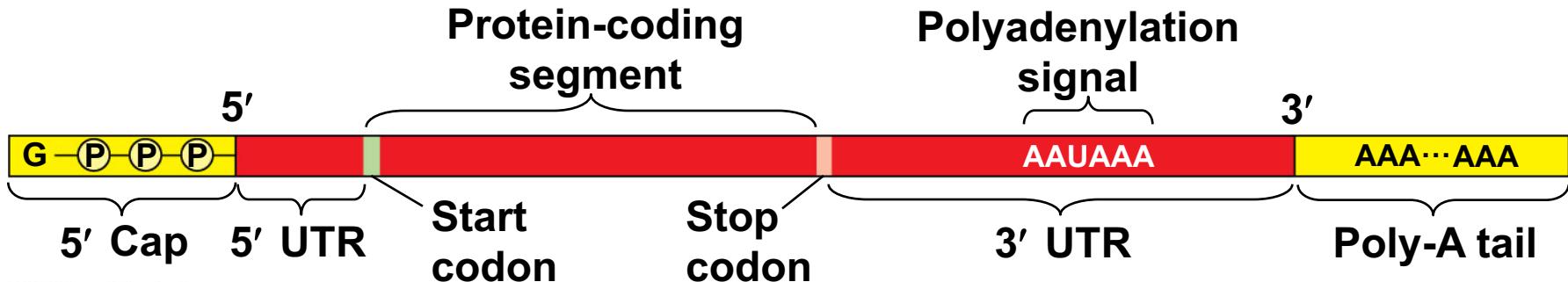
- Enzymes in the eukaryotic nucleus modify pre-mRNA (**RNA processing**) before the genetic messages are dispatched to the cytoplasm
- During RNA processing, both ends of the primary transcript are usually altered
- Also, usually some interior parts of the molecule are cut out, and the other parts spliced together

# Alteration of mRNA Ends

- Each end of a pre-mRNA molecule is modified in a particular way
  - The 5' end receives a modified nucleotide **5' cap**
  - The 3' end gets a **poly-A tail**
- These modifications share several functions
  - They seem to facilitate the export of mRNA
  - They protect mRNA from hydrolytic enzymes
  - They help ribosomes attach to the 5' end

# mRNA in Eukaryotes

- Note that pre-mRNA
  - Does *not* contain the 5' Cap or Poly-A tail
  - Does contain introns which are spliced out

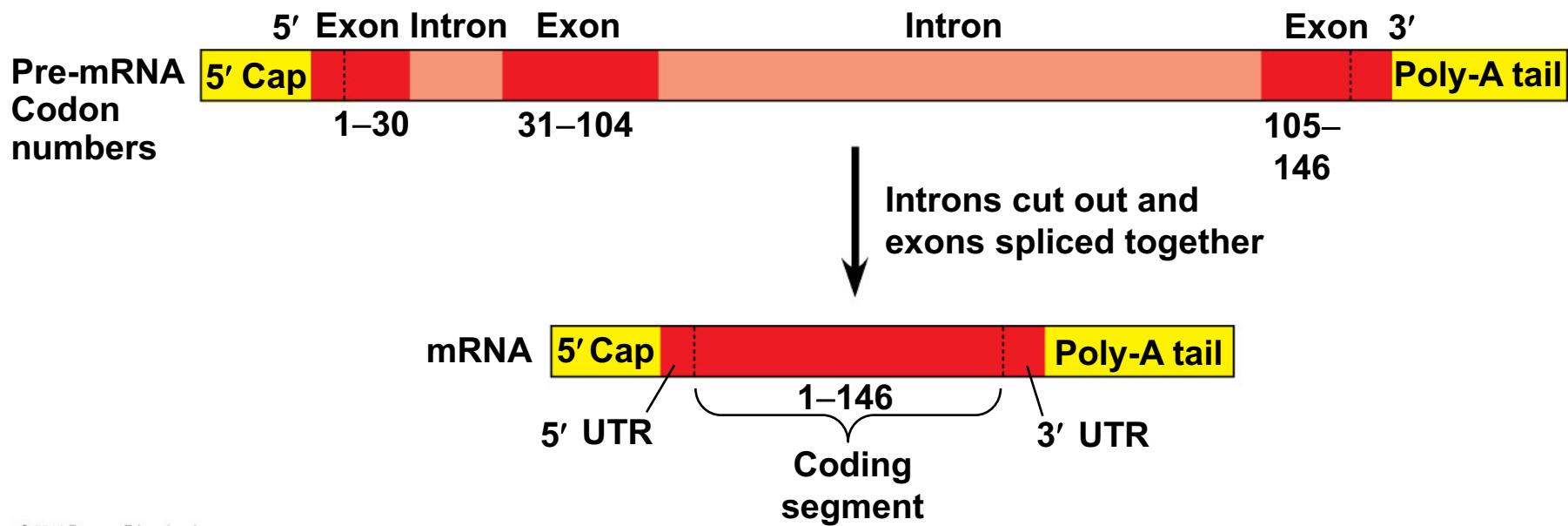


© 2011 Pearson Education, Inc.

# Split Genes and RNA Splicing

- Most eukaryotic genes and their RNA transcripts have long noncoding stretches of nucleotides that lie between coding regions
- These noncoding regions are called intervening sequences, or **introns**
- The other regions are called **exons** because they are eventually expressed, usually translated into amino acid sequences
- **RNA splicing** removes introns and joins exons, creating an mRNA molecule with a continuous coding sequence

Figure 17.11

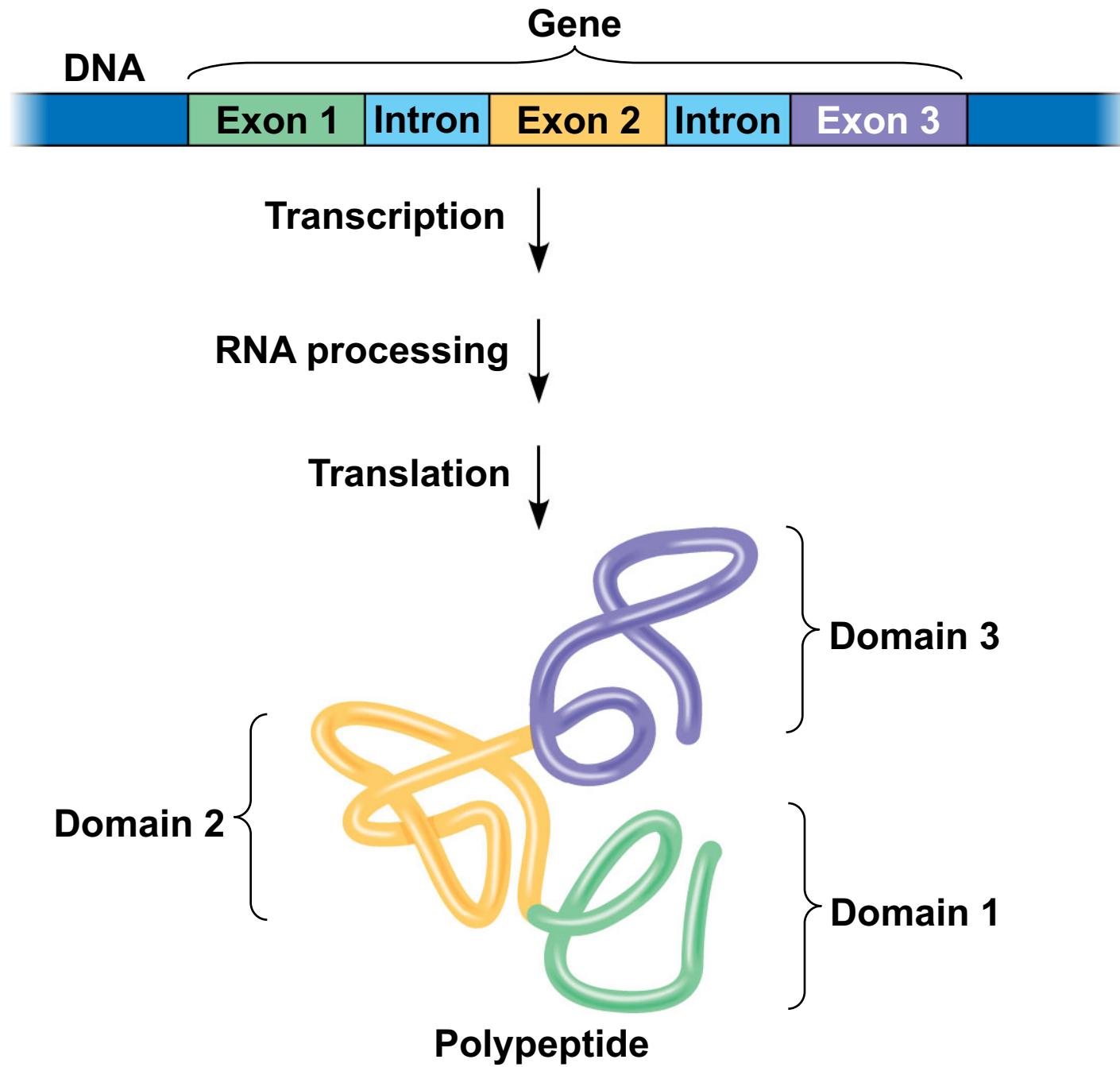


# *The Functional and Evolutionary Importance of Introns*

- Some introns contain sequences that may regulate gene expression
- Some genes can encode more than one kind of polypeptide, depending on which segments are treated as exons during splicing
- This is called **alternative RNA splicing**
- Consequently, the number of different proteins an organism can produce is much greater than its number of genes

- Proteins often have a modular architecture consisting of discrete regions called **domains**
- In many cases, different exons code for the different domains in a protein
- Exon shuffling may result in the evolution of new proteins

Figure 17.13



# Concept 17.5: Mutations of one or a few nucleotides can affect protein structure and function

- **Mutations** are changes in the genetic material of a cell or virus
- **Point mutations** are chemical changes in just one base pair of a gene
- The change of a single nucleotide in a DNA template strand can lead to the production of an abnormal protein

### Wild-type hemoglobin

#### Wild-type hemoglobin DNA



### Sickle-cell hemoglobin

#### Mutant hemoglobin DNA



#### mRNA



#### mRNA



#### Normal hemoglobin



© 2011 Pearson Education, Inc.

#### Sickle-cell hemoglobin



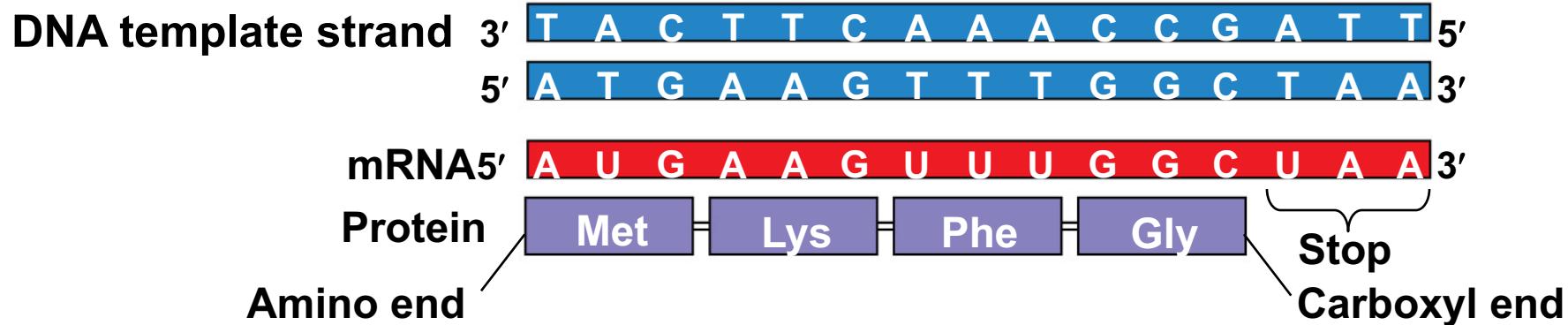
# Types of Small-Scale Mutations

- Point mutations within a gene can be divided into two general categories
  - Nucleotide-pair substitutions
  - One or more nucleotide-pair insertions or deletions

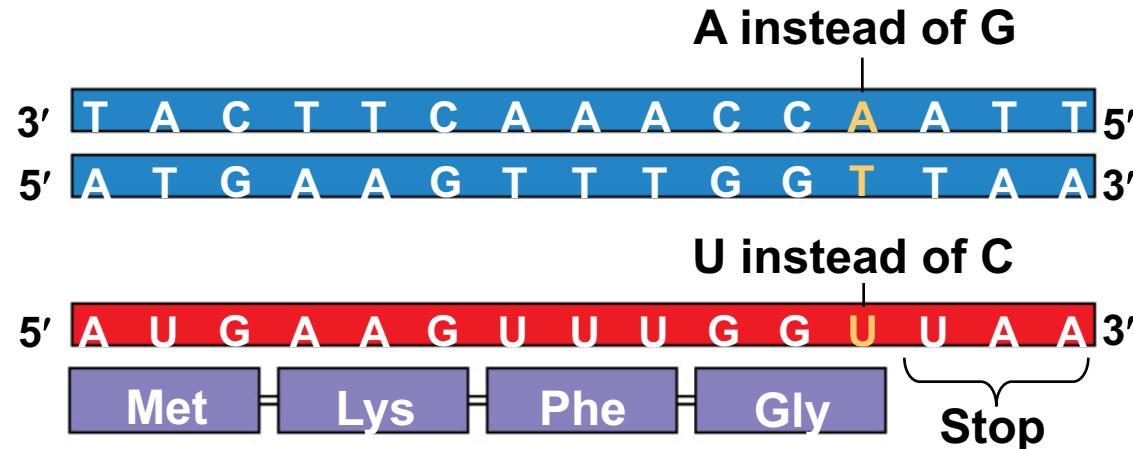
# *Substitutions*

- A **nucleotide-pair substitution** replaces one nucleotide and its partner with another pair of nucleotides
- **Silent mutations** have no effect on the amino acid produced by a codon because of redundancy in the genetic code
- **Missense mutations** still code for an amino acid, but not the correct amino acid
- **Nonsense mutations** change an amino acid codon into a stop codon, nearly always leading to a nonfunctional protein

## Wild type



### (a) Nucleotide-pair substitution: silent



# Wild type

**DNA template strand** 3' T A C T T C A A A C C G A T T 5'  
5' A T G A A G T T T G G C T A A 3'

The diagram illustrates the correspondence between an mRNA sequence and its encoded protein. The mRNA sequence is shown as a red horizontal bar with the 5' end on the left and the 3' end on the right. The sequence is: mRNA5' A U G A A G U U U G G G C U A A 3'. Below the mRNA, the protein structure is shown as a series of four purple boxes, each representing an amino acid: Met, Lys, Phe, and Gly. An arrow points from the 'Amino end' label to the first box (Met). Another arrow points from the 'Carboxyl end' label to the last box (Gly). A bracket above the last two boxes (Gly and Stop) indicates that they are part of the same codon.

## (a) Nucleotide-pair substitution: missense

**T instead of C**

The diagram shows two horizontal rows representing the sugar-phosphate backbones of a DNA double helix, oriented in opposite directions (3' to 5' and 5' to 3'). The top row is labeled "3'" at its left end and "5'" at its right end. The bottom row is labeled "5'" at its left end and "3'" at its right end. The nucleotide bases are shown as colored squares: Adenine (blue), Thymine (orange), Cytosine (green), Guanine (purple), and Tymidine (yellow). In the middle of the sequence, there is a difference between the two strands: the top strand has a Thymine base (orange) where the bottom strand has a Cytosine base (green). This visualizes a mutation where a Cytosine base has been replaced by a Thymine base.

AUGAAAGUUUAGCUAAG

**A instead of G**

5' A U G A A G U U U A G C U A A 3'

Met = Lys = Phe = Ser Stop

## Wild type

DNA template strand 3' T A C T T C A A A C C G A T T 5'

5' A T G A A G T T T G G C T A A 3'

mRNA 5' A U G A A G U U U G G C U A A 3'

Protein Met = Lys = Phe = Gly Stop

Amino end Carboxyl end

### (a) Nucleotide-pair substitution: nonsense

**A instead of T**

**T instead of C**

3' T A C A T C A A A C C G A T T 5'

5' A T G T A G T T T G G C T A A 3'

**U instead of A**

5' A U G U A G U U U G G C U A A 3'

Met Stop

# *Insertions and Deletions*

- **Insertions and deletions** are additions or losses of nucleotide pairs in a gene
- These mutations have a disastrous effect on the resulting protein more often than substitutions do
- Insertion or deletion of nucleotides may alter the reading frame, producing a **frameshift mutation**

## Wild type

DNA template strand 3' T A C T T C A A A C C G A T T 5'

5' A T G A A G T T T G G C T A A 3'

mRNA 5' A U G A A G U U U G G C U A A 3'

Protein Met = Lys = Phe = Gly Stop Carboxyl end

Amino end

**(b) Nucleotide-pair insertion or deletion: frameshift causing immediate nonsense**

The diagram illustrates two RNA sequences. The top sequence, labeled "Extra A", shows a frameshift mutation where the reading frame is shifted by one nucleotide. The mutated sequence is: 3' T A C A T T C A A A A C G G A T T 5'. The bottom sequence, labeled "Extra U", shows a frameshift mutation where the reading frame is shifted by one nucleotide. The mutated sequence is: 5' A U G U A A G U U U U G G G C U A A 3'. In both cases, the mutated sequence is highlighted in blue or red, and the original sequence is shown below it.

## 1 nucleotide-pair insertion

# Wild type

DNA template strand 3' T A C T T C A A A C C G A T T 5'  
5' A T G A A G T T T G G C T A A 3'

mRNA 5' A U G A A G U U U U G G C U A A 3'

**Protein** Met Lys Phe Gly

## Amino end

**Stop**

## Carboxyl end

**(b) Nucleotide-pair insertion or deletion: frameshift causing extensive missense**

The diagram shows two horizontal blue lines representing the DNA strands. The top strand is labeled '3' at its left end and ends with '5'. The bottom strand is labeled '5' at its right end and ends with '3'. Above the top strand, the sequence reads T A C T T C A A C C G A T T. A yellow arrow points from the word 'missing' to the fourth Adenine (A) nucleotide in the sequence, which is highlighted with a blue square.

5' **A U G A A G U U** **G** G C U A A ... 3'

**Met    Lys    Leu    Ala** ...

**U missing**

## 1 nucleotide-pair deletion

## Wild type

DNA template strand 3' T A C T T C A A A C C G A T T 5'  
5' A T G A A G T T T G G C T A A 3'

mRNA 5' A U G A A G U U U G G G C U A A 3'

**Protein** → Met = Lys = Phe = Gly → **Stop Carboxyl end**

**no end**

**(b) Nucleotide-pair insertion or deletion: no frameshift, but one amino acid missing**

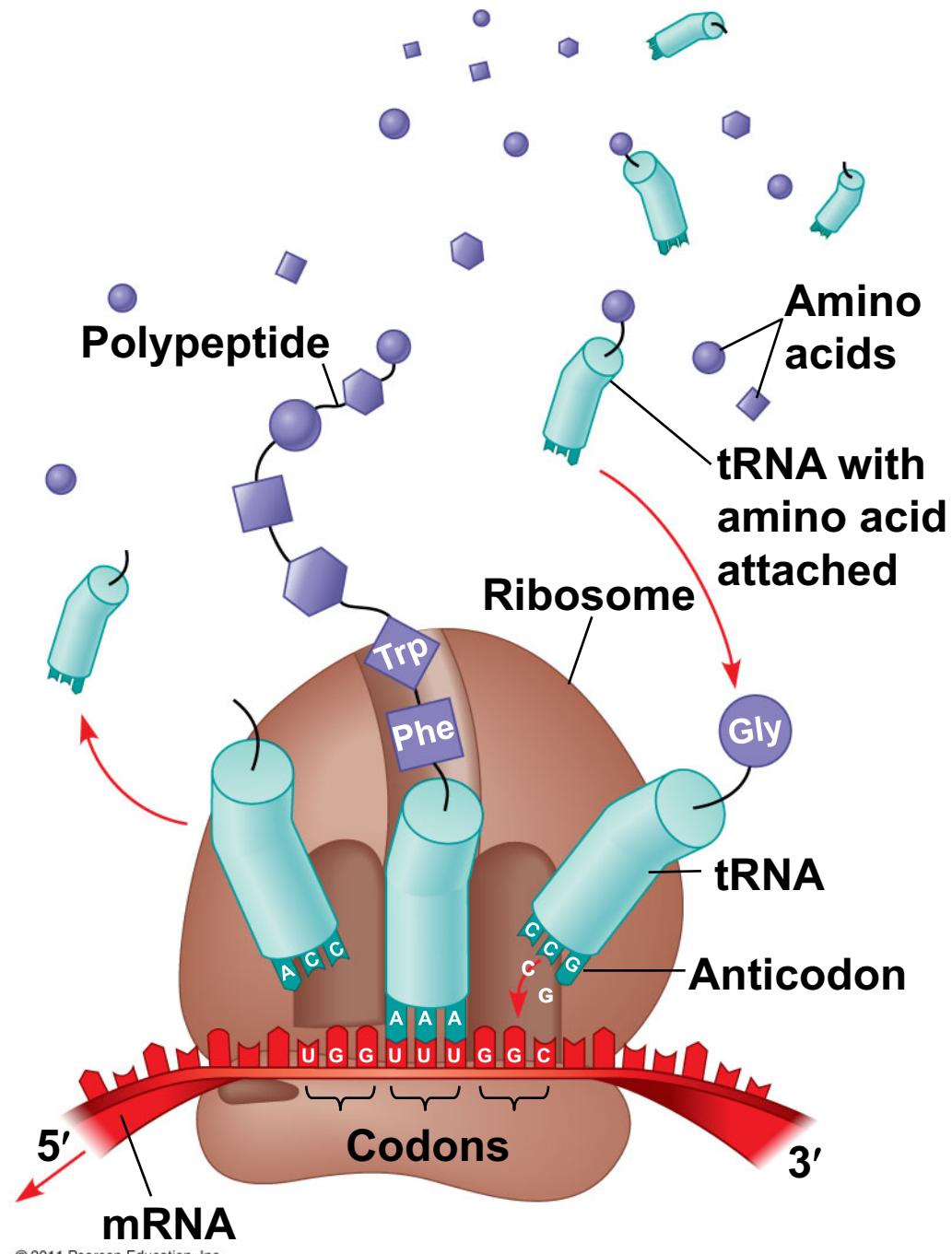
The diagram illustrates a mismatch between the template strand (top) and the target strand (bottom). The template strand has a sequence of T T C, with the third base highlighted in blue. The target strand has a sequence of T A C. A vertical orange line connects the third position of both strands, indicating the mismatch.

The diagram illustrates a mutation in an mRNA sequence. The original sequence is: 5' A U G U U U G G C U A A A 3'. The first three nucleotides (AUG) code for Methionine (Met). The next four nucleotides (UUU) code for Phenylalanine (Phe). The following two nucleotides (GG) code for Glycine (Gly). The last three nucleotides (CUA) are a stop signal, indicated by a bracket labeled "Stop". A red box highlights the mutated nucleotide at position 4, which is a Guanine (G), instead of the normal Uracil (U). This change shifts the reading frame, skipping the codon for Phe and starting the next codon at the third position (UAA), which is a stop signal.

### 3 nucleotide-pair deletion

# Translation

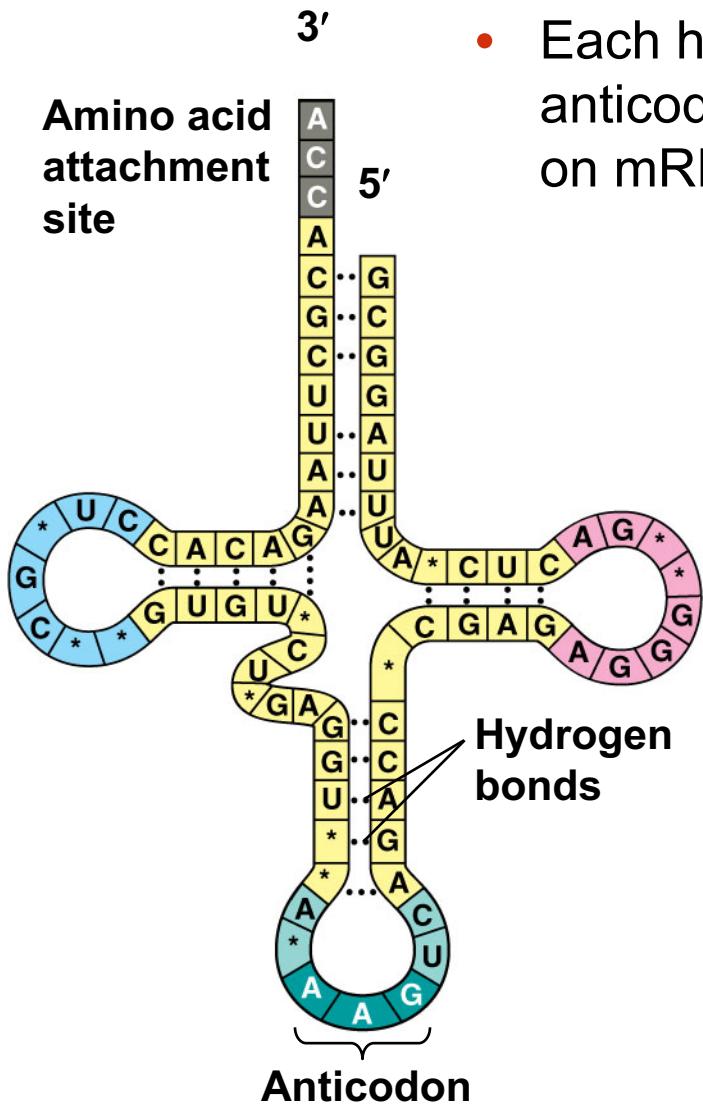
- A cell translates a mRNA message into protein with the help of **transfer RNA (tRNA)**
- tRNA transfer amino acids to the growing polypeptide in a ribosome
- Translation is a complex process in terms of its biochemistry and mechanics



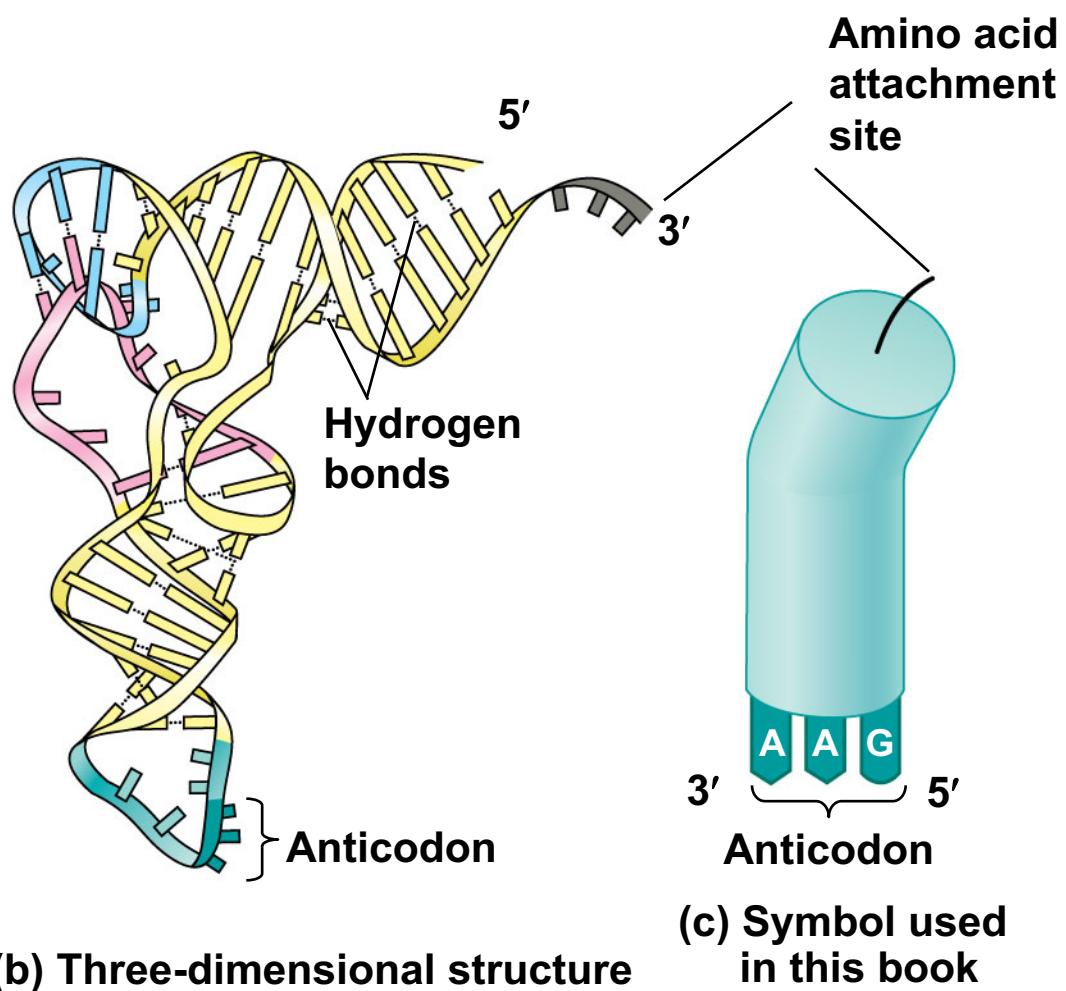
# tRNA structure

Figure 17.15

- Each carries a specific amino acid on one end
- Each has an **anticodon** on the other end; the anticodon base-pairs with a complementary codon on mRNA



(a) Two-dimensional structure

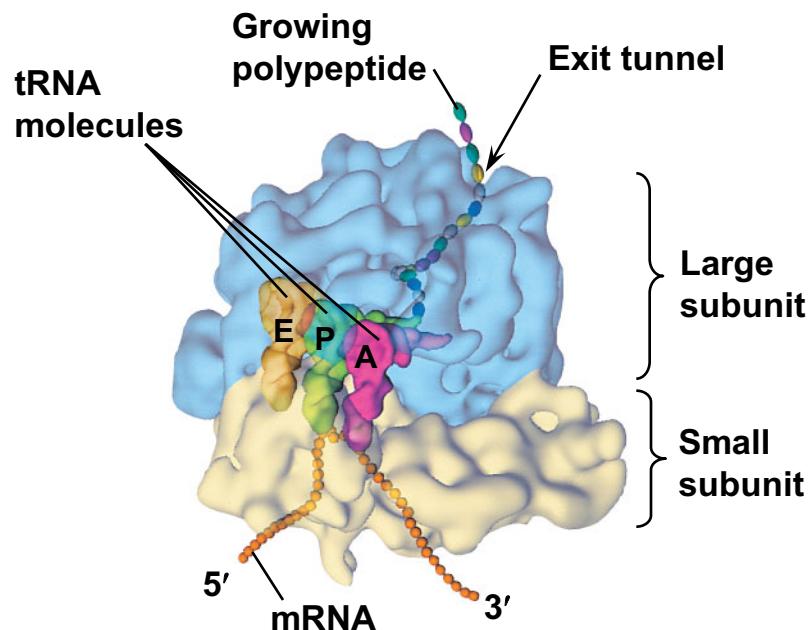


(c) Symbol used  
in this book

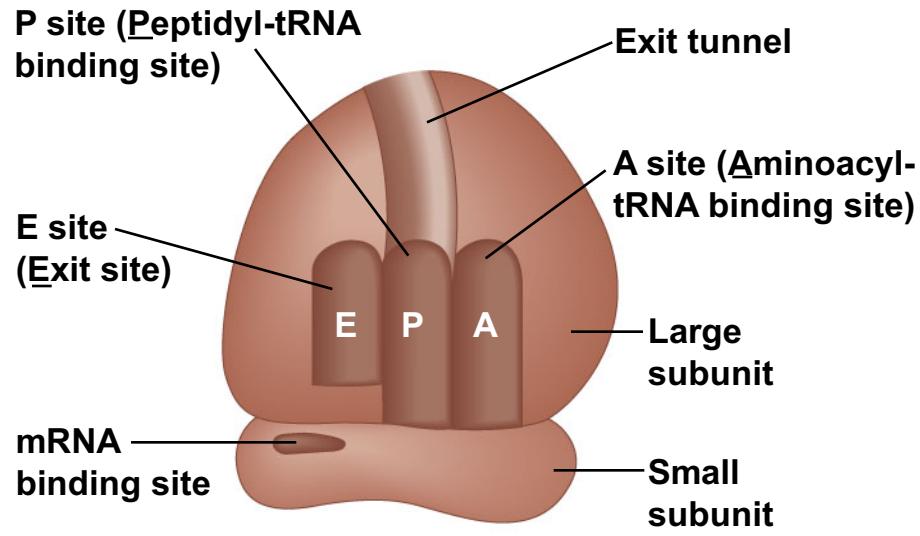
# *Ribosomes*

- Ribosomes facilitate specific coupling of tRNA anticodons with mRNA codons in protein synthesis
- The two ribosomal subunits (large and small) are made of proteins and **ribosomal RNA (rRNA)**
- Bacterial and eukaryotic ribosomes are somewhat similar but have significant differences: some antibiotic drugs specifically target bacterial ribosomes without harming eukaryotic ribosomes

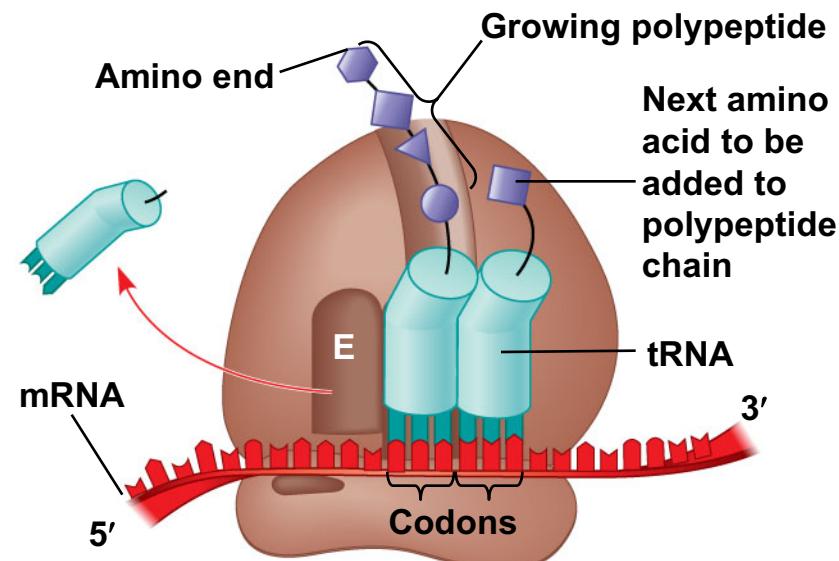
Figure 17.17



(a) Computer model of functioning ribosome



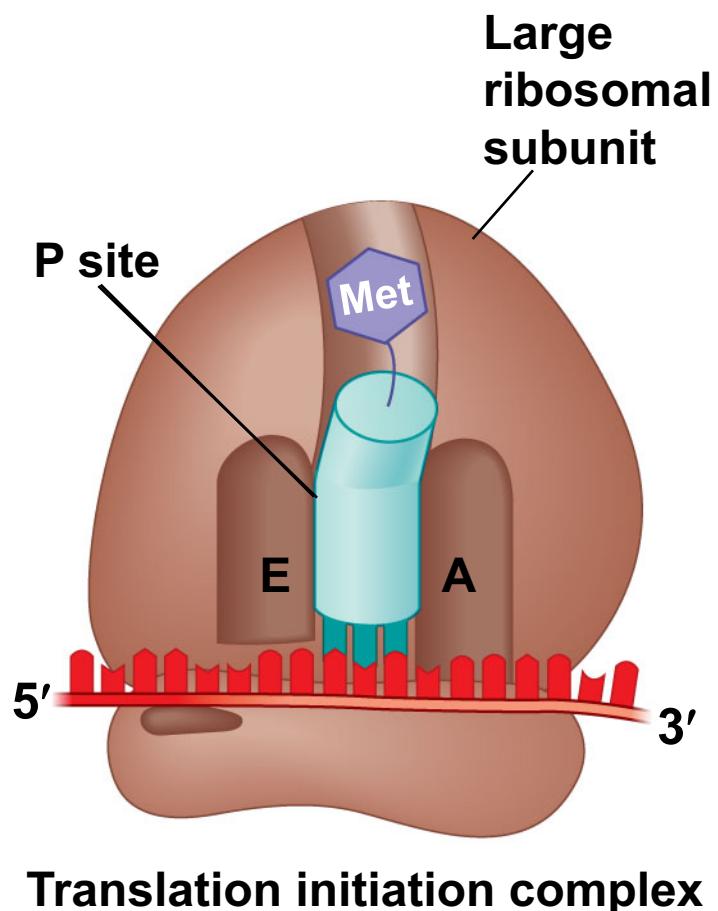
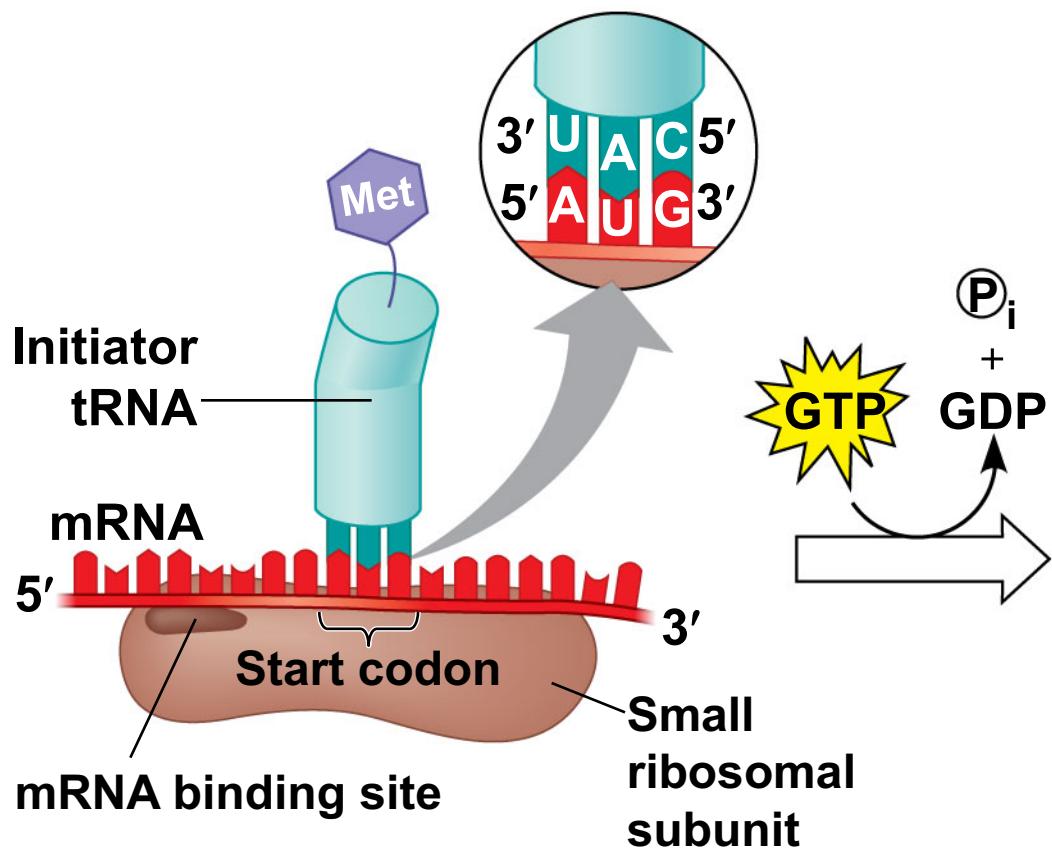
(b) Schematic model showing binding sites



(c) Schematic model with mRNA and tRNA

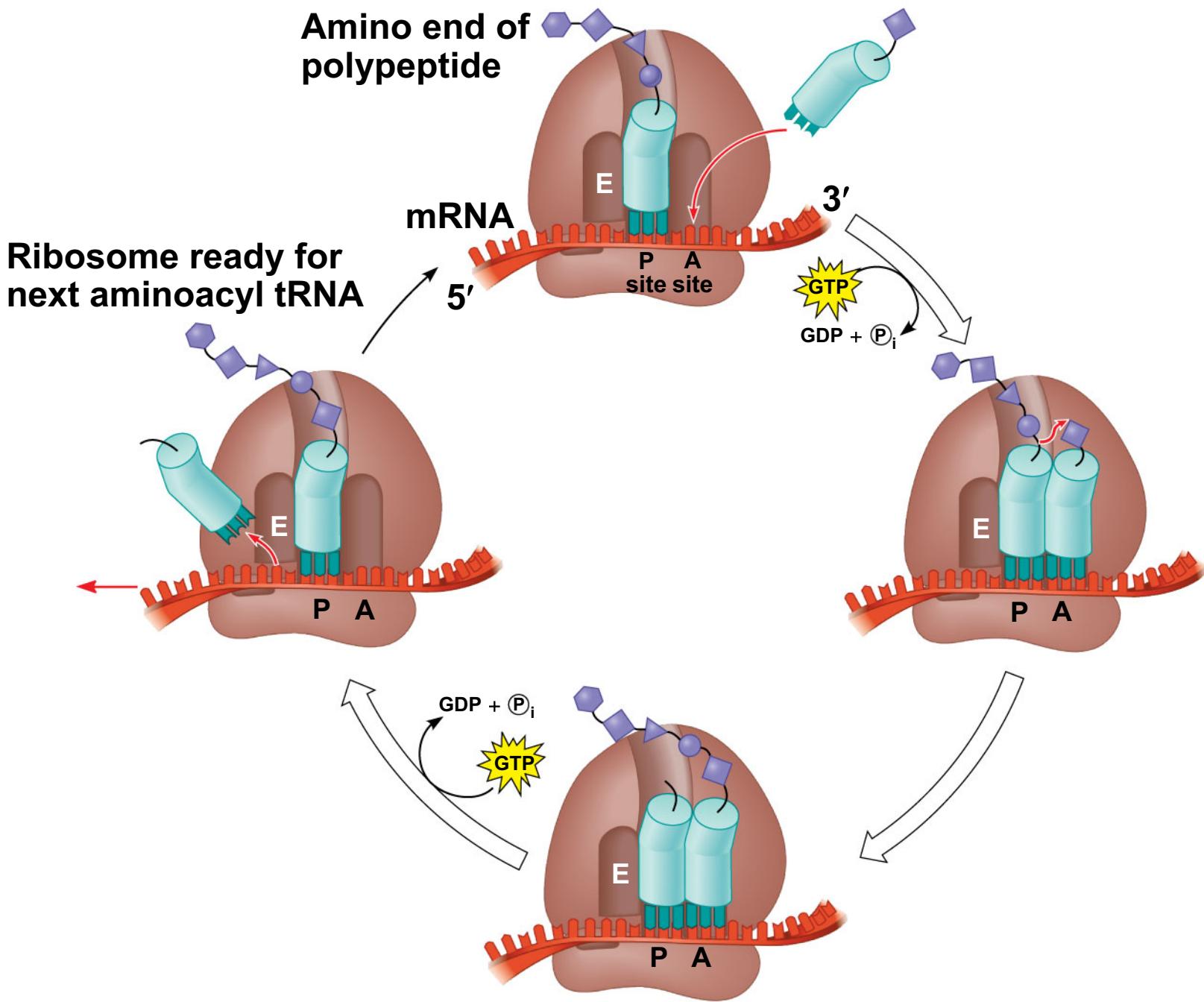
- A ribosome has three binding sites for tRNA
  - The **P site** holds the tRNA that carries the growing polypeptide chain
  - The **A site** holds the tRNA that carries the next amino acid to be added to the chain
  - The **E site** is the exit site, where discharged tRNAs leave the ribosome
- As with transcription, there are three steps to translation
  - Initiation
  - Elongation
  - Termination

# Initiation of Translation



# *Elongation of the Polypeptide Chain*

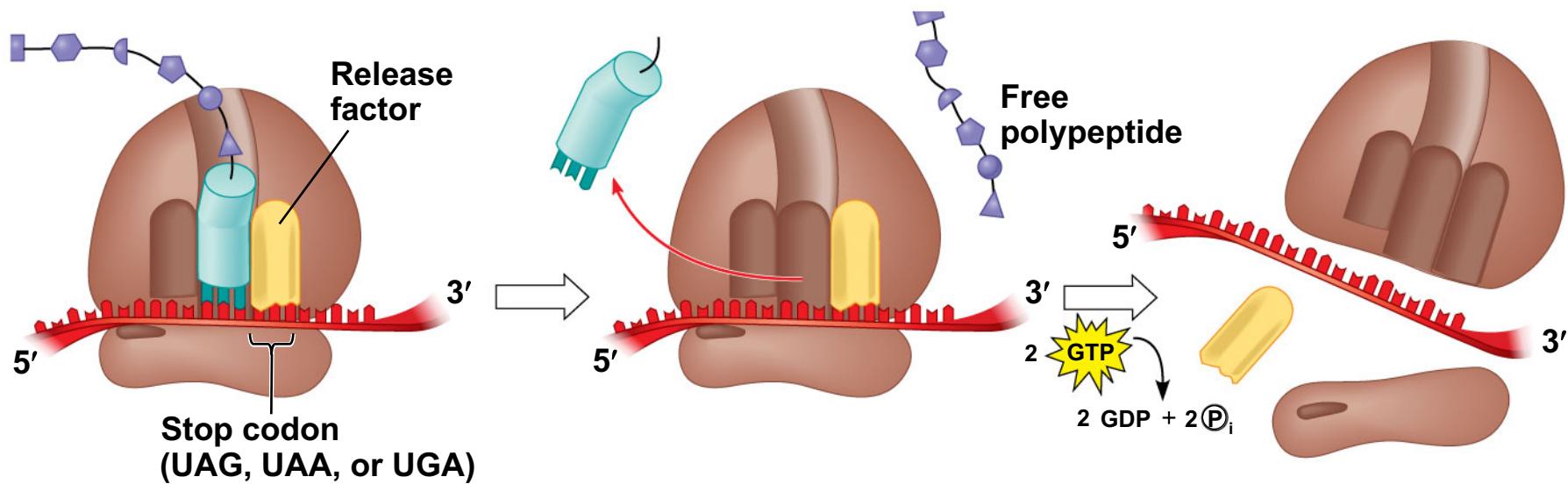
- During the elongation stage, amino acids are added one by one to the preceding amino acid at the C-terminus of the growing chain
- Each addition involves proteins called elongation factors and occurs in three steps: codon recognition, peptide bond formation, and translocation
- Translation proceeds along the mRNA in a 5' to 3' direction



# *Termination of Translation*

- Termination occurs when a stop codon in the mRNA reaches the A site of the ribosome
- The A site accepts a protein called a release factor
- The release factor causes the addition of a water molecule instead of an amino acid
- This reaction releases the polypeptide, and the translation assembly then comes apart

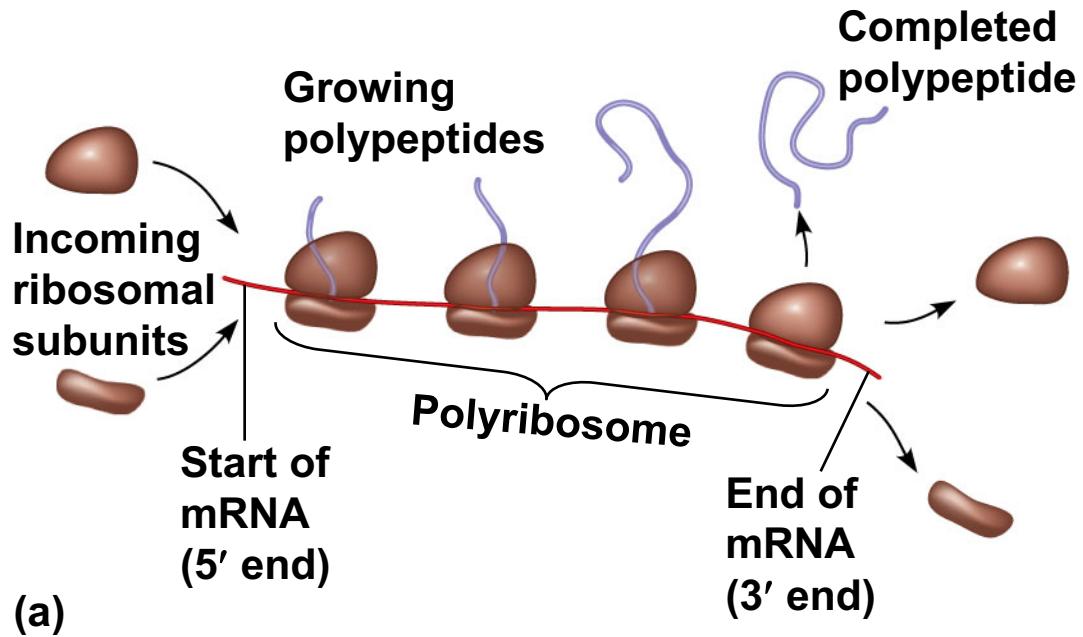
# Termination of Translation



© 2011 Pearson Education, Inc.

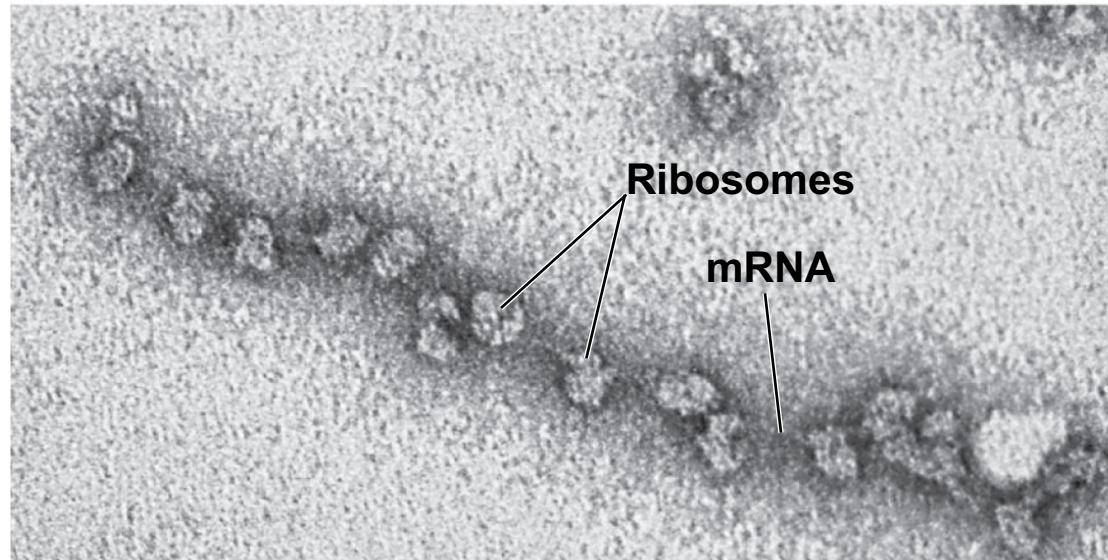
# Polyribosomes

- A number of ribosomes can translate a single mRNA simultaneously, forming a **polyribosome** (or **polysome**)
- Polyribosomes enable a cell to make many copies of a polypeptide very quickly



(a)

(b)



# Completing and Targeting the Functional Protein

- Often translation is not sufficient to make a functional protein
- Polypeptide chains are modified after translation or targeted to specific sites in the cell
  - Proteins targeted to specific sites in the cell contain a **signal sequence**
  - What organelle functions as a "highway" within a cell?

# What is a gene?

- a region of DNA that can be expressed to produce a final functional product, either
  - a polypeptide or
  - an RNA molecule

