Chapter 8

Gene Expression: The Flow of Genetic Information from DNA to RNA to Protein

Sections to study

8.1 The genetic code

8.2 Transcription: From DNA to RNA

8.3 Translation: From mRNA to protein

8.4 Differences in gene expression between prokaryotes and eukaryotes

8.5 Comprehensive example: Computerized analysis of gene expression in *C. elegans*

8.6 The effect of mutations on gene expression and gene function

8.1 The genetic code

A gene's nucleotide sequence is colinear with the amino acid sequence of the encoded peptide.

- **1960s – Charles Yanofsky, *E.coli* tryptophan synthetase subunit TrpA.**
- **Isolated a large number of *trpA* mutants.**
- **Build a fine genetic map of *trpA* mutations.**
- **Purified and determined amino acid sequence of TrpA mutants.**

Charles Yanofsky
(1925-)

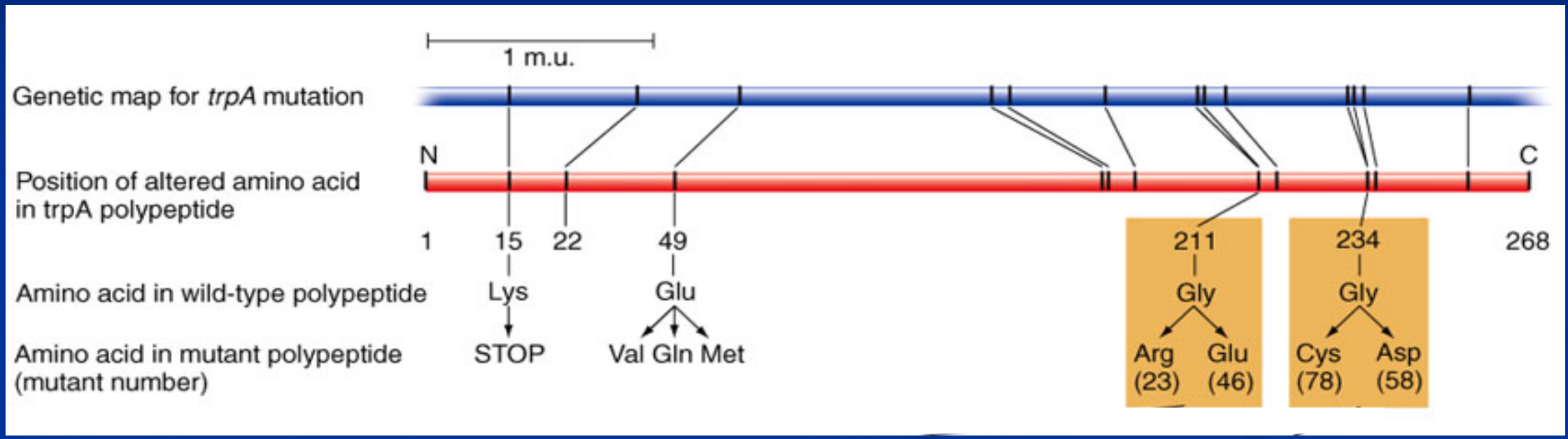

1 m.u.
Genetic map for trpA mutation
N
C
Position of altered amino acid in trpA polypeptide
1
15
22
49
211
234
268
Amino acid in wild-type polypeptide
Lys
Glu
Gly
Gly
Amino acid in mutant polypeptide (mutant number)
STOP
Val Gln Met
Arg (23)
Glu (46)
Cys (78)
Asp (58)

Fig. 8.4

Evidence that a codon is composed of more than one nucleotide

4 Nucleotides:

A
G
T
C

20 Amino acids:

Ala	Leu
Arg	Lys
Asn	Met
Asp	Phe
Cys	Pro
Gln	Ser
Glu	Thr
Gly	Trp
His	Tyr
Ile	Val

1 nt/a.a. ⟶ A, G, T, C (4 combinations)

2 nt/a.a. ⟶ AA, AG, AT, AC
GA, GG, GT, GC
TA, TG, TT, TC
CA, CG, CT, CC (4×4=16 combinations)

3 nt/a.a. ⟶ AAA, AAG, AAT, AAC
AGA, AGG, AGT, AGC
ATA, ATG, ATT, ATC
ACA, ACG, ACT, ACC
GAA, GAG, GAT, GAC
GGA, GGG, GGT, GGC
GTA, GTG, GTT, GTC
GCA, GCG, GCT, GCC
…….. (4×4×4=64 combinations)

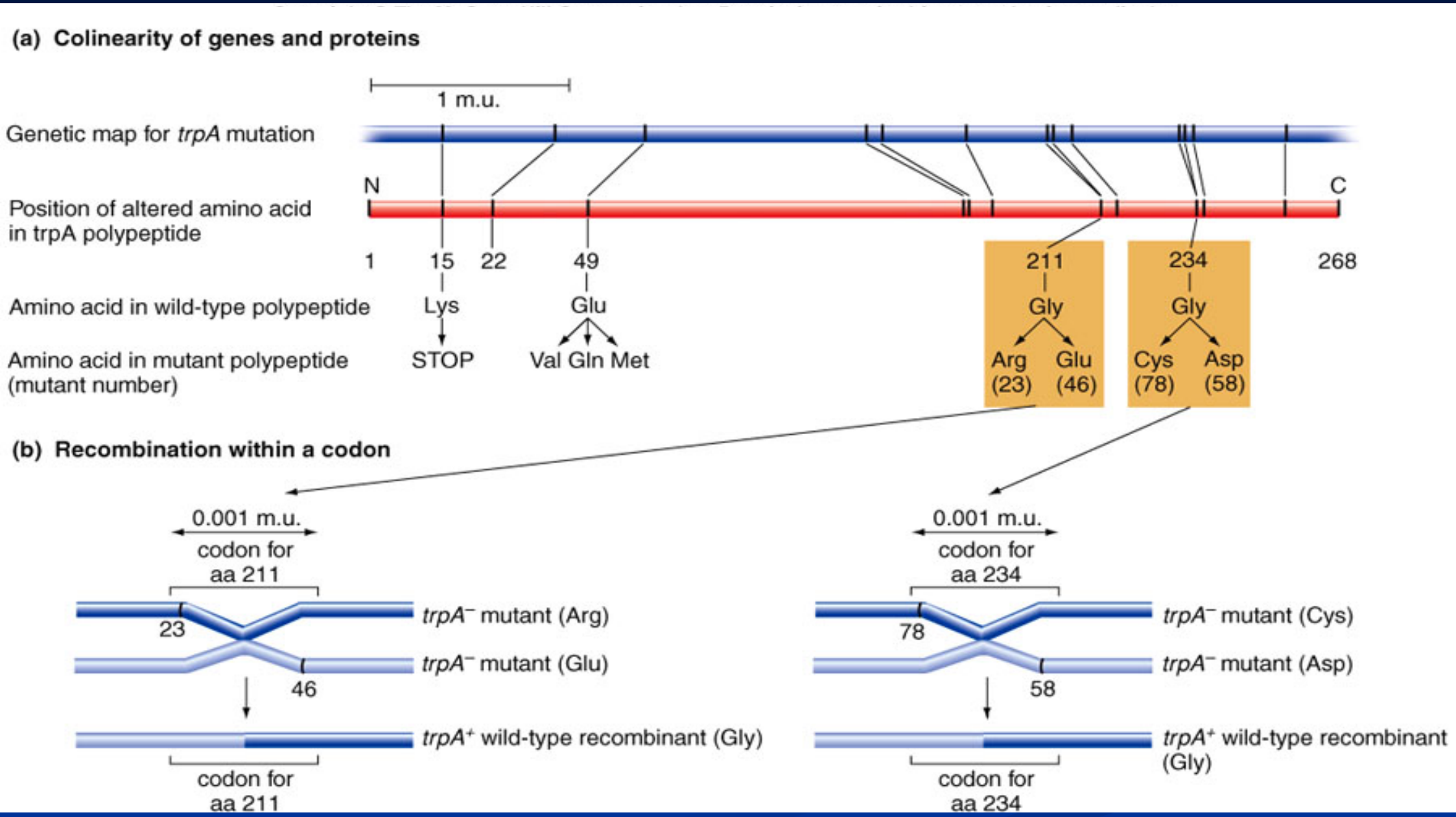

(a) Colinearity of genes and proteins
1 m.u.
Genetic map for trpA mutation
N
C
Position of altered amino acid in trpA polypeptide
1
15
22
49
211
234
268
Amino acid in wild-type polypeptide
Lys
Glu
Gly
Gly
Amino acid in mutant polypeptide (mutant number)
STOP
Val Gln Met
Arg (23)
Glu (46)
Cys (78)
Asp (58)
(b) Recombination within a codon
0.001 m.u.
codon for aa 211
0.001 m.u.
codon for aa 234
23
trpA⁻ mutant (Arg)
78
trpA⁻ mutant (Cys)
trpA⁻ mutant (Glu)
trpA⁻ mutant (Asp)
46
58
trpA⁺ wild-type recombinant (Gly)
trpA⁺ wild-type recombinant (Gly)
codon for aa 211
codon for aa 234

Fig. 8.4

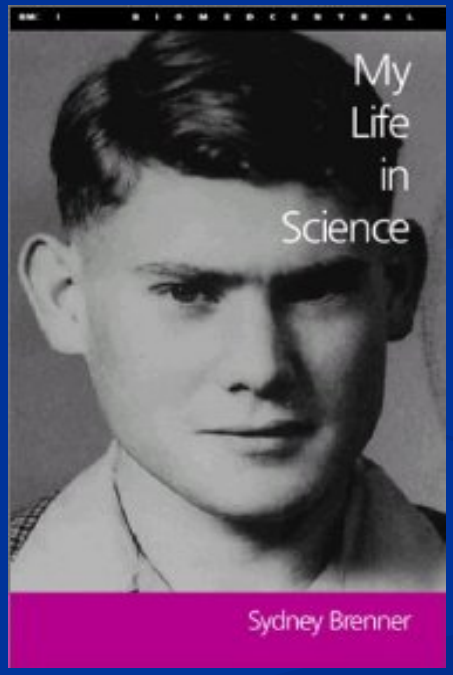
Interpretation of the results

- A codon is composed of more than one nucleotide.
  - Different point mutations may affect the same amino acid.
- Each nucleotide is part of only a single codon.
  - Each point mutation altered only one amino acid.

Evidence for a triplet code

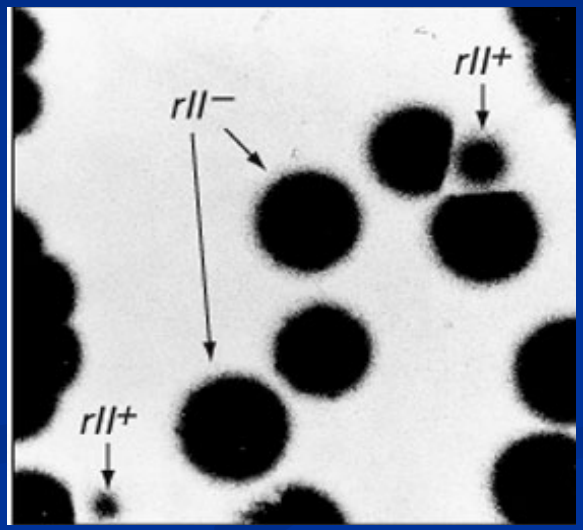
- 1955 – Francis Crick and Sydney Brenner
- Generate mutations of bacteriophage T4 *rIIB* gene with proflavin.

Francis Crick
(1916-2004)



Sydney Brenner
(1927-)

T4 strain	*E. coli* strain B	*E. coli* strain K(λ)
rII$^-$	Large, distinct	No plaques
rII$^+$	Small, fuzzy	Small, fuzzy



Intragenic suppression of *rIIB* mutations

- **Intragenic suppression:**

  The restoration of gene function by one mutation canceling another in the same gene.

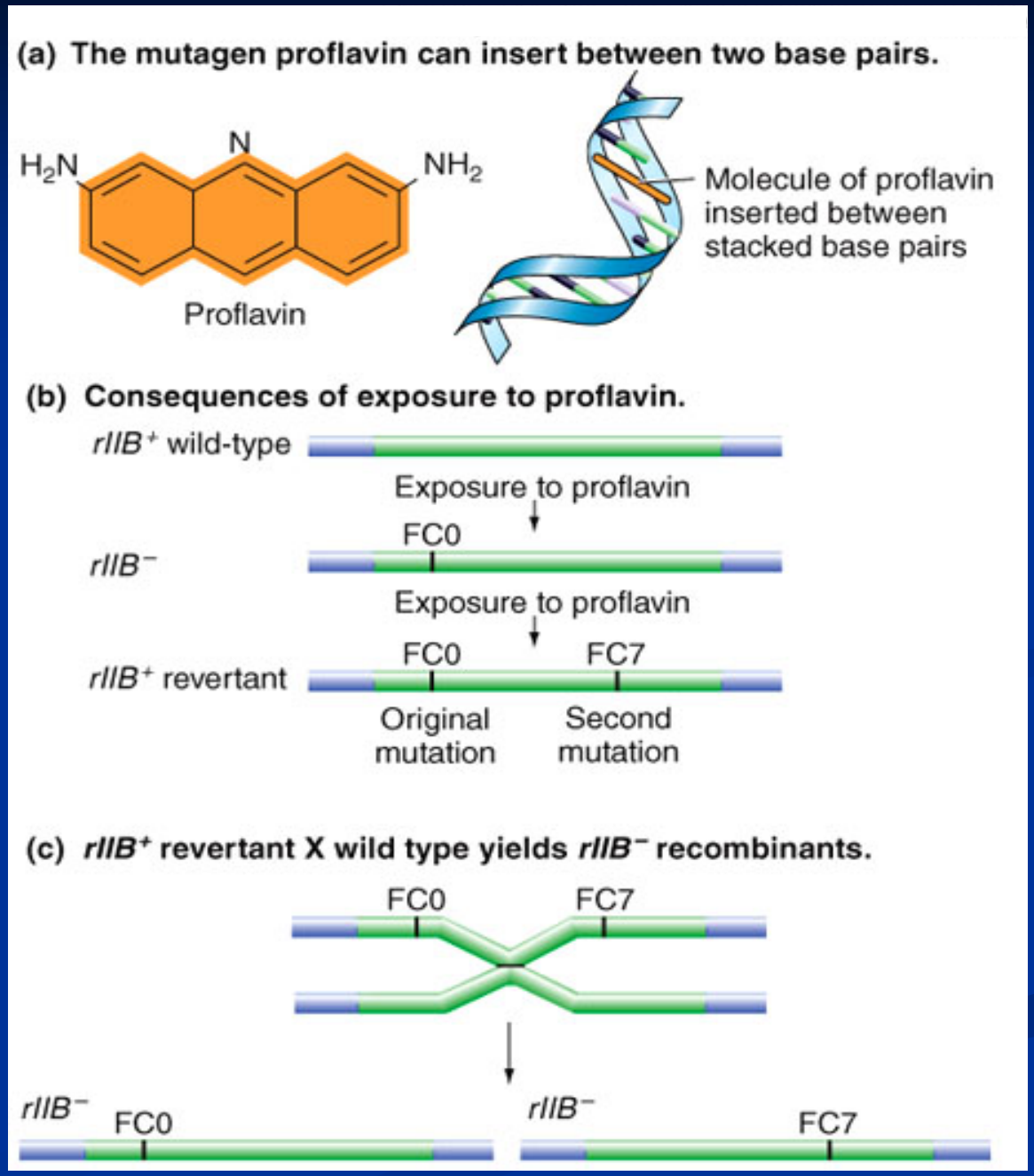


Fig. 8.5

(d) Different sets of mutations generate either a mutant or a normal phenotype.

Proflavin-induced mutations (+) insertion (–) deletion	Phenotype
– or +	Mutant
– – or + +	Mutant
– – – – or – – – – – or + + + + or + + + + +	Mutant
– +	Wild type
– – – or – – – – – – or + + + or + + + + + +	Wild type

Fig. 8.5

Interpretation:

- Each gene has a single starting point which establishes a **reading frame**.
  - **Frameshift mutations**: Insertions or deletions of base pairs that alter the grouping of nucleotides into codons.
- A codon is composed of three nucleotides.
- Most amino acids are specified by more than one codon.

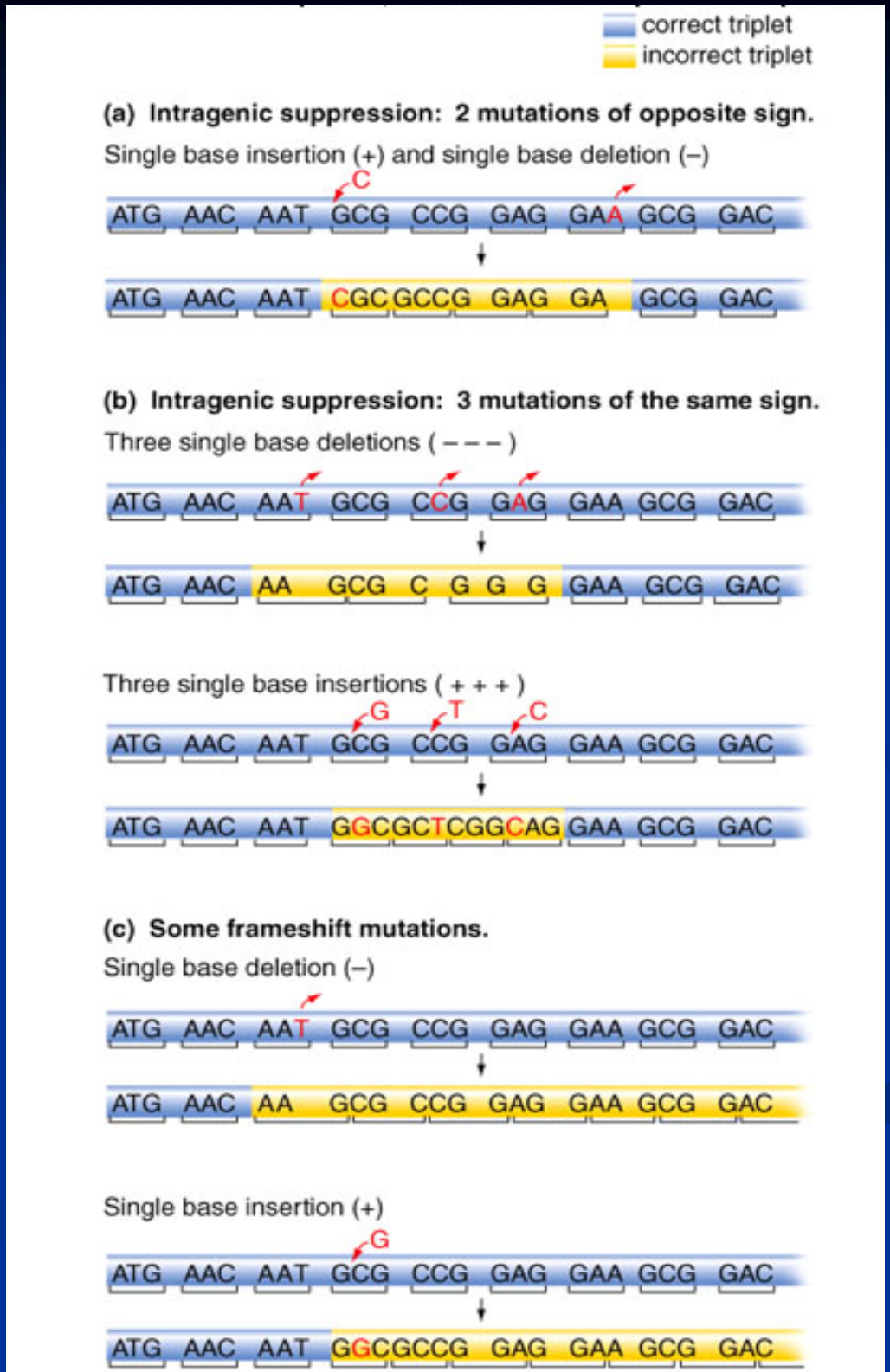
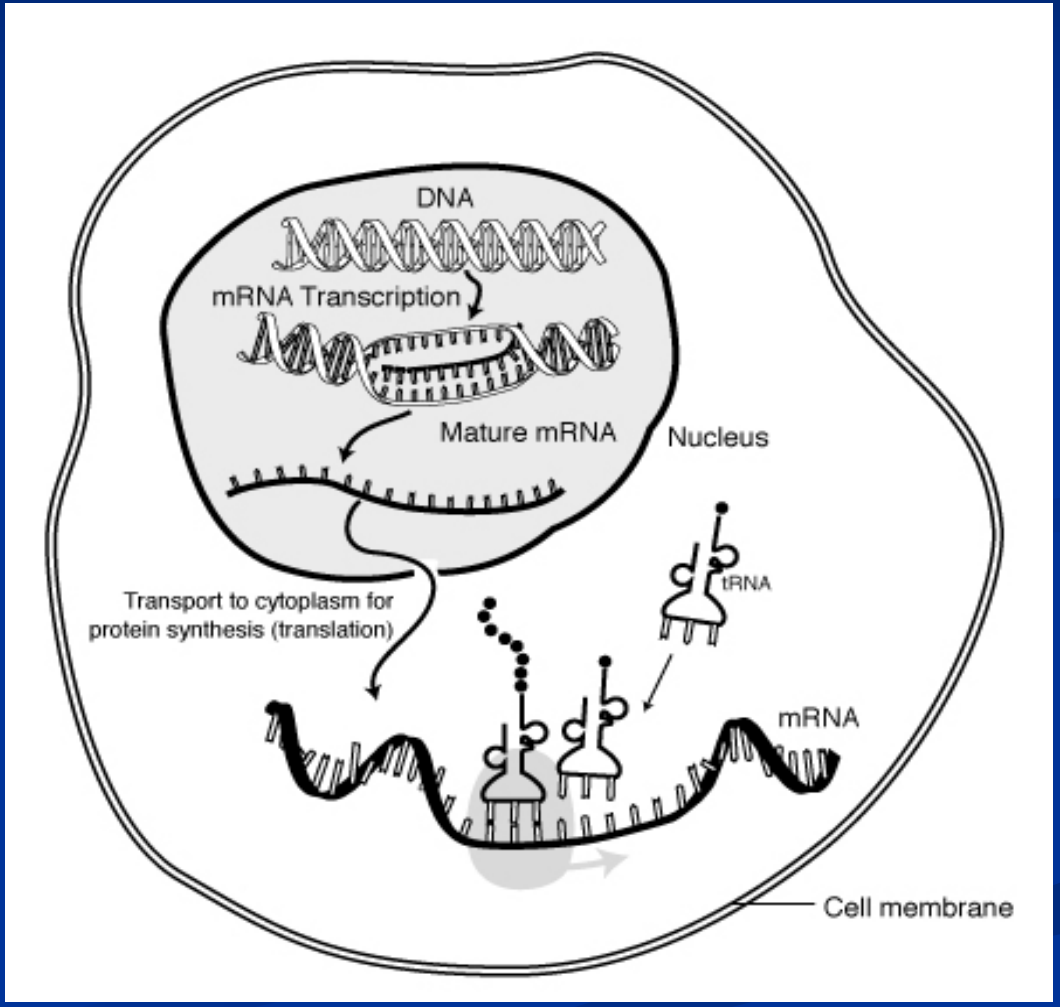


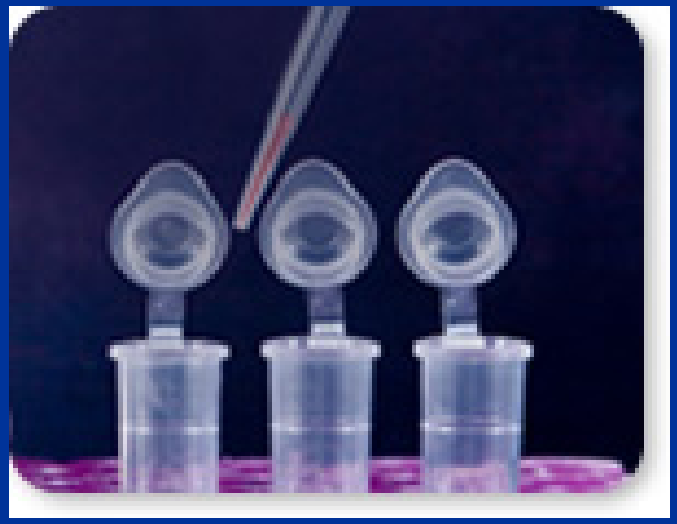
Fig. 8.6

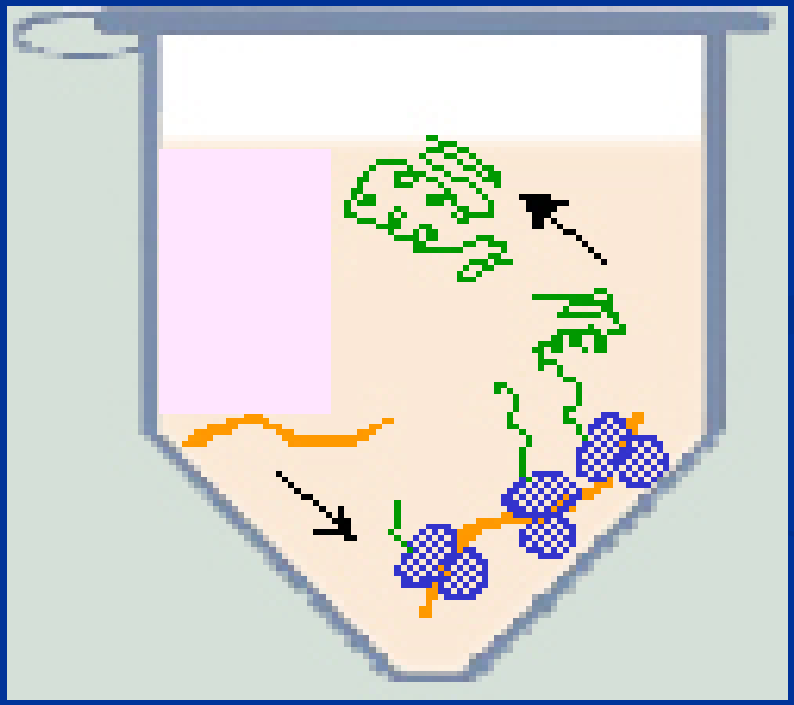
Cracking the code: Which codons represent which amino acids?

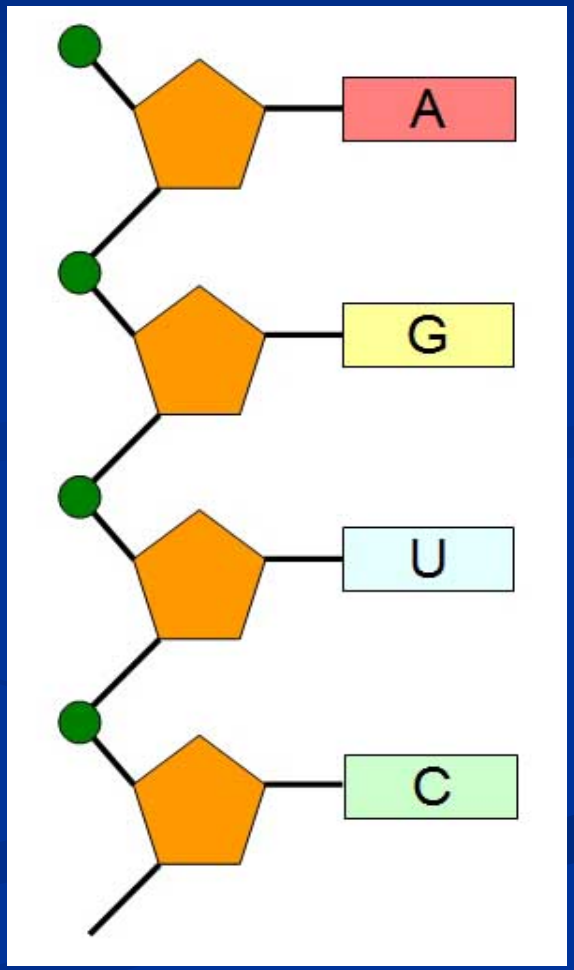
- In 1950s, the discovery of **messenger RNAs**, molecules for transporting genetic information.
  - Protein synthesis takes place in cytoplasm deduced from radioactive tagging of amino acids.



- **The development of two techniques**
  - ***In vitro* translation systems**
  - **Synthesis of artificial mRNAs**

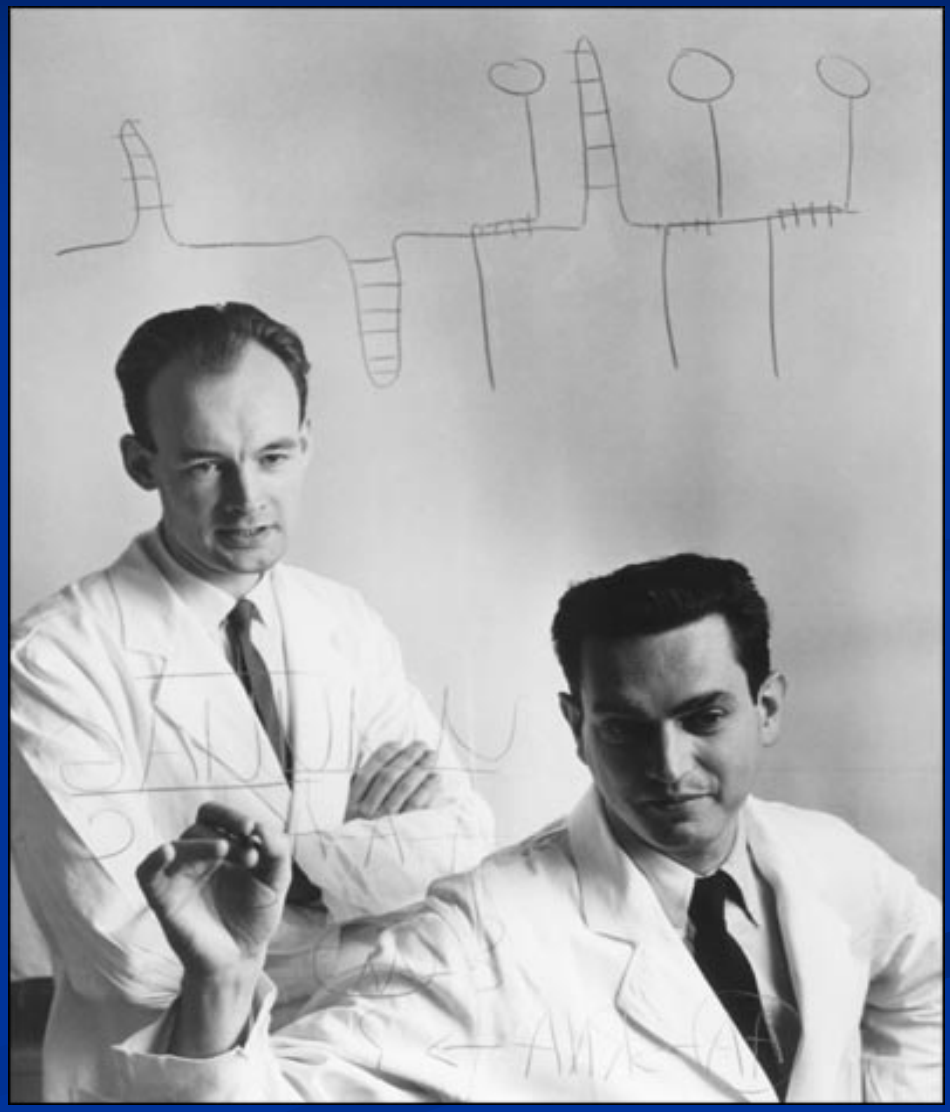


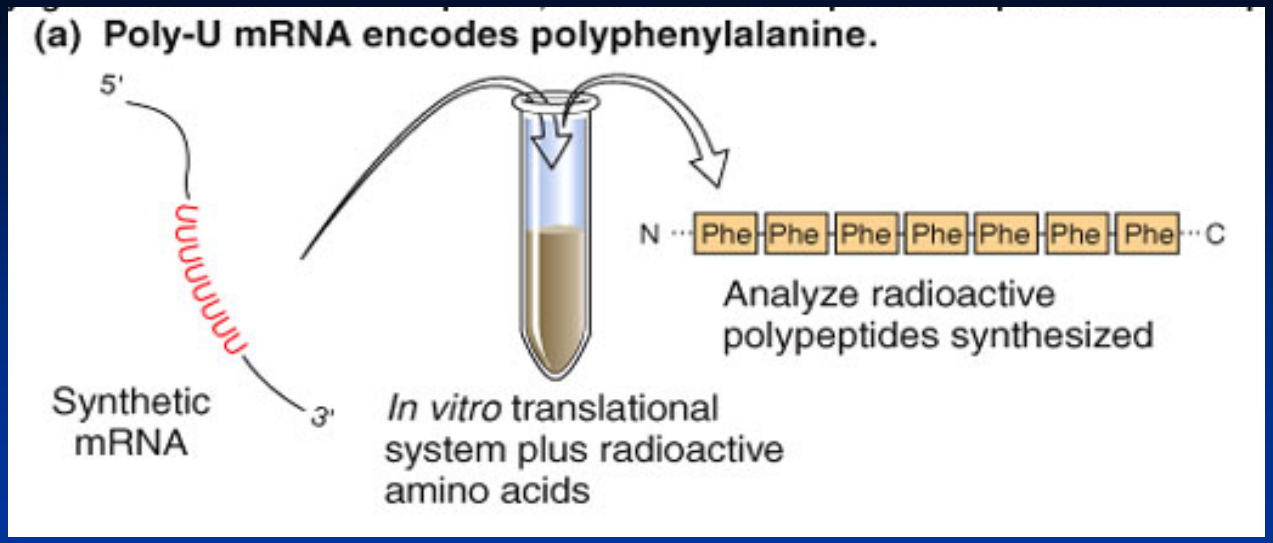




Nirenberg and Matthaei's experiment

- 1961 – Marshall Nirenberg and Heinrich Matthaei
- *In vitro* translation of synthetic poly-U mRNA





(b) Analyzing the coding possibilities.

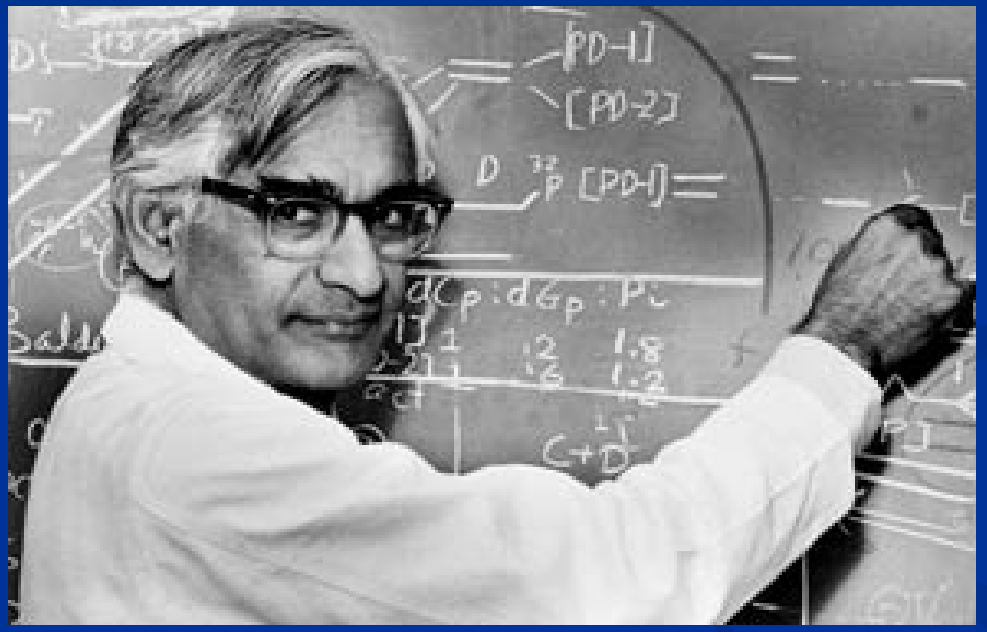
Synthetic mRNA	Polypeptides synthesized
	Polypeptides with one amino acid
poly-U UUUU …	Phe-Phe-Phe …
poly-C CCCC …	Pro-Pro-Pro …
poly-A AAAA …	Lys-Lys-Lys …
poly-G GGGG …	Gly-Gly-Gly …

UUU – Phe
CCC -- Pro
AAA -- Lys
GGG -- Gly

Fig. 8.7a, b

Khorana's experiment

- **Har Gobind Khorana**
- **Synthesis and translation of mRNAs with repeating nucleotides.**



Har Gobind Khorana
(1922-2011)

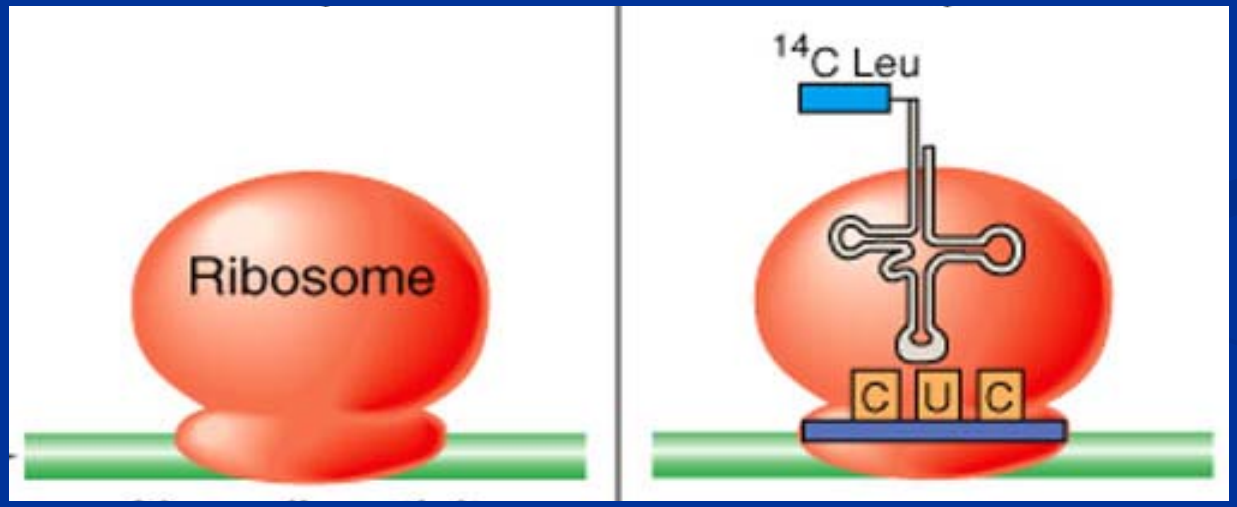
Repeating dinucleotides	Polypeptides with alternating amino acids
poly-UC UCUC …	Ser-Leu-Ser-Leu …
poly-AG AGAG …	Arg-Glu-Arg-Glu …
poly-UG UGUG …	Cys-Val-Cys-Val …
poly-AC ACAC …	Thr-His-Thr-His …
Repeating trinucleotides	**Three polypeptides each with one amino acid**
poly-UUC UUCUUCUUC …	Phe-Phe.... and Ser-Ser.... and Leu-Leu....
poly-AAG AAGAAGAAG …	Lys-Lys.... and Arg-Arg.... and Glu-Glu....
poly-UUG UUGUUGUUG …	Leu-Leu.... and Cys-Cys.... and Val-Val....
poly-UAC UACUACUAC …	Tyr-Tyr.... and Thr-Thr.... and Leu-Leu....
Repeating tetranucleotides	**Polypeptides with repeating units of four amino acids**
poly-UAUC UAUCUAUC …	Tyr-Leu-Ser-Ile-Tyr-Leu-Ser-Ile...
poly-UUAC UUACUUAC …	Leu-Leu-Thr-Tyr-Leu-Leu-Thr-Tyr...
poly-GUAA GUAAGUAA …	none
poly-GAUA GAUAGAUA …	none

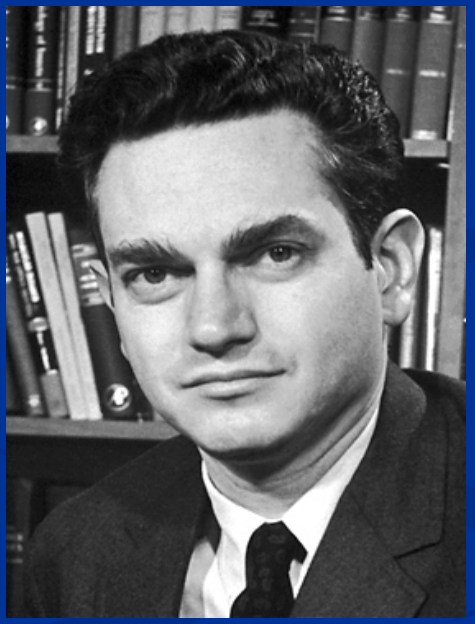
UCU, CUC – Ser, Leu
AGA, GAG – Arg, Glu
UGU, GUG – Cys, Val
ACA, CAC – Thr, His

Fig. 8.7b

Nirenberg and Leder's experiment

- 1965 - Nirenberg and Philip Leder
- Mix synthetic 3 nucleotide mRNAs with tRNAs charged with radioactive amino acid.





Marshall W. Nirenberg
(1927-2010)

Fig. 8.8

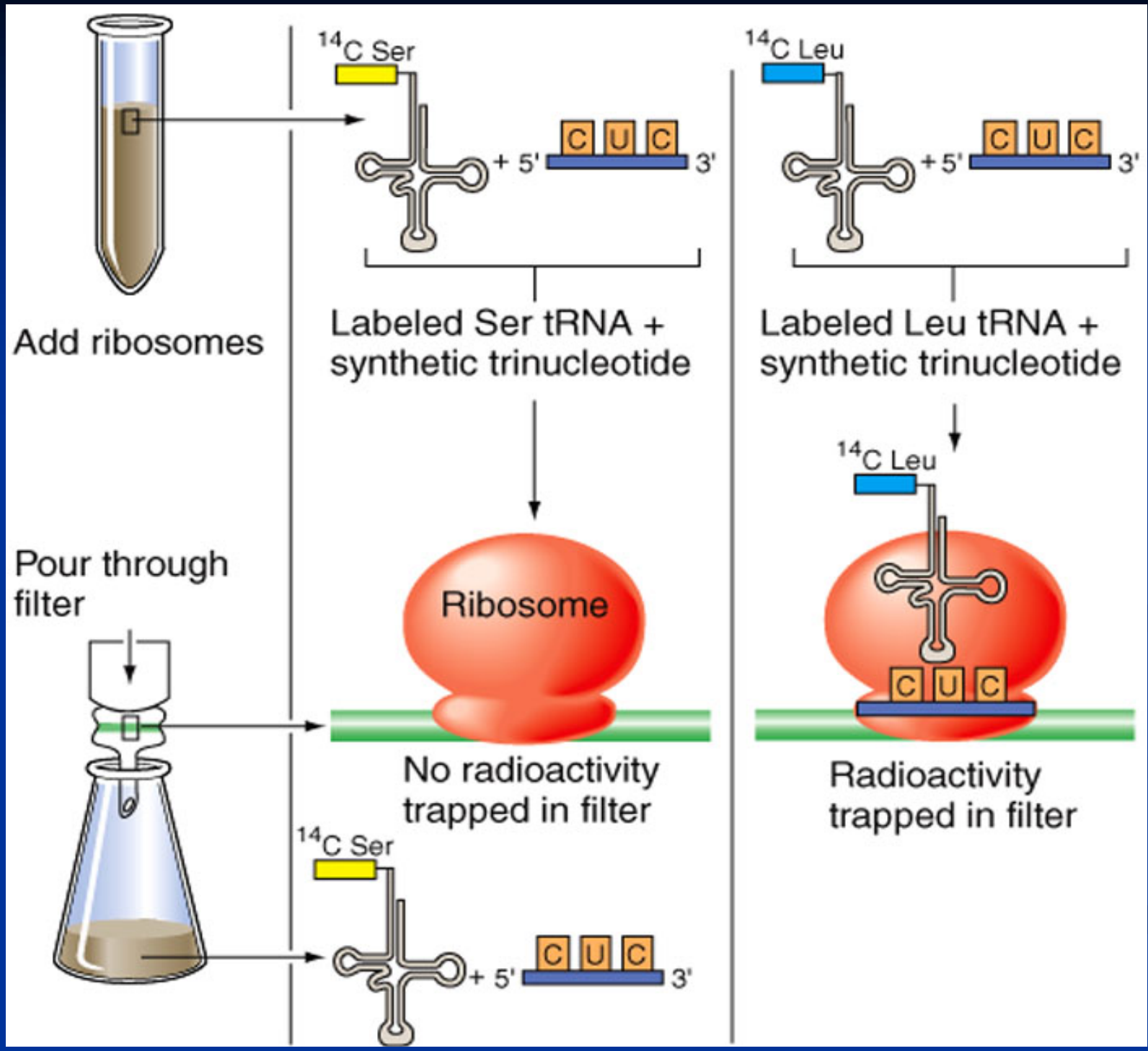

^{14}C Ser
C U C
+ 5'
3'
^{14}C Leu
C U C
+ 5'
3'
Add ribosomes
Labeled Ser tRNA + synthetic trinucleotide
Labeled Leu tRNA + synthetic trinucleotide
^{14}C Leu
C U C
Pour through filter
Ribosome
No radioactivity trapped in filter
Radioactivity trapped in filter
^{14}C Ser
C U C
+ 5'
3'

Fig. 8.8

The Genetic Code: 61 triplet codons represent 20 amino acids; 3 triplet codons signify stop

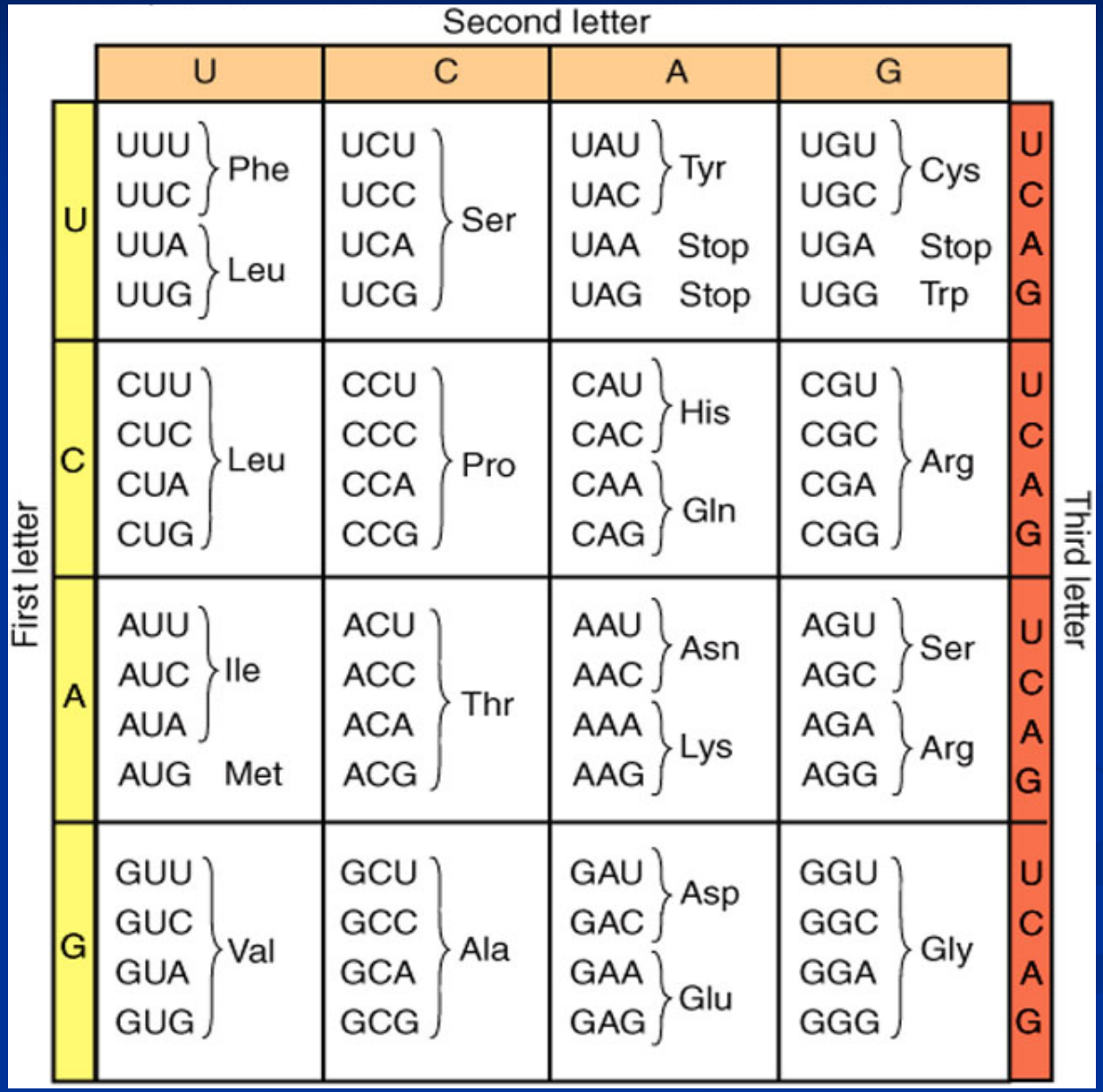


Fig. 8.3

Polarities: 5’-to-3’ in mRNA corresponds to N-to-C in polypeptide

5’-AAA UUU-3’ —*In vitro* translation→ N-Lys-Phe-C (Actual result)

5’-AAA UUU-3’ —*In vitro* translation→ N-Phe-Lys-C (Not found)

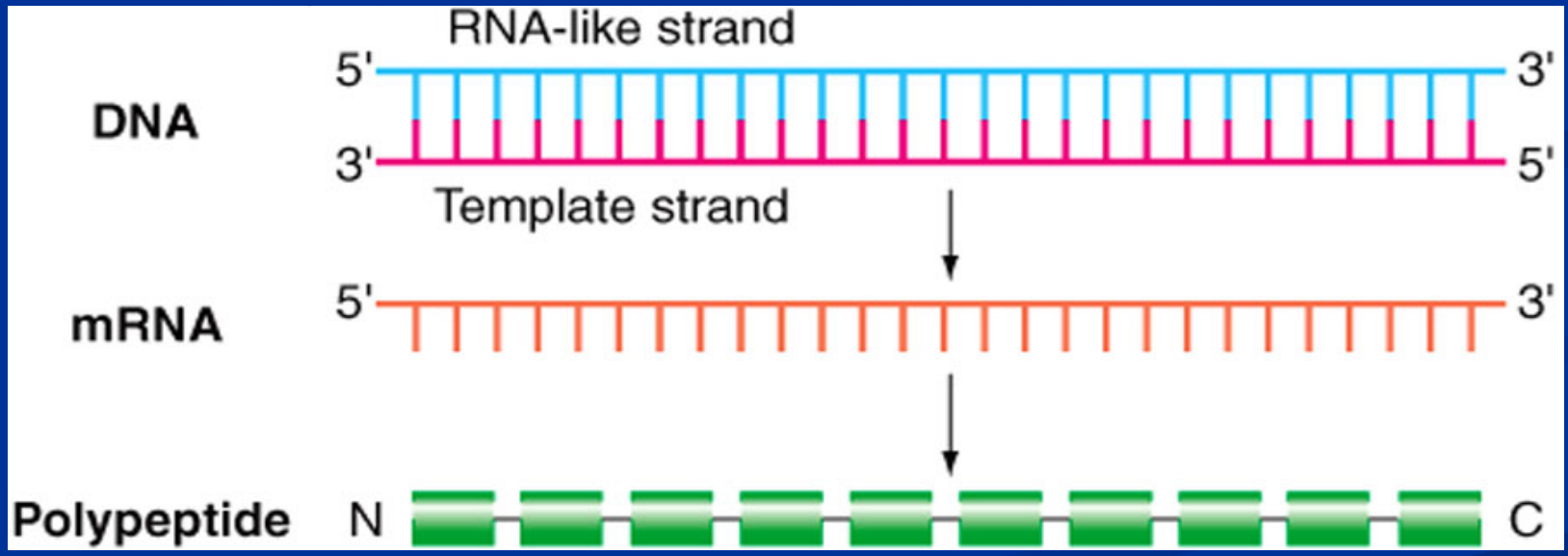


Fig. 8.9

Nonsense codons cause termination of a polypeptide chain

Repeating tetranucleotides		Polypeptides with repeating units of four amino acids
poly-UAUC	UAUCUAUC ...	Tyr-Leu-Ser-Ile-Tyr-Leu-Ser-Ile...
poly-UUAC	UUACUUAC ...	Leu-Leu-Thr-Tyr-Leu-Leu-Thr-Tyr...
poly-GUAA	GUAAGUAA ...	none
poly-GAUA	GAUAGAUA ...	none

The three stop codons:

UAA (*ocher*)

UAG (*amber*)

UGA (*opal*)

Fig. 8.7 b

The Nobel Prize in Physiology or Medicine 1968

"for their interpretation of the genetic code and its function in protein synthesis"

Robert W. Holley

1/3 of the prize

USA

Cornell University
Ithaca, NY, USA

Har Gobind Khorana

1/3 of the prize

USA

University of Wisconsin
Madison, WI, USA

Marshall W. Nirenberg

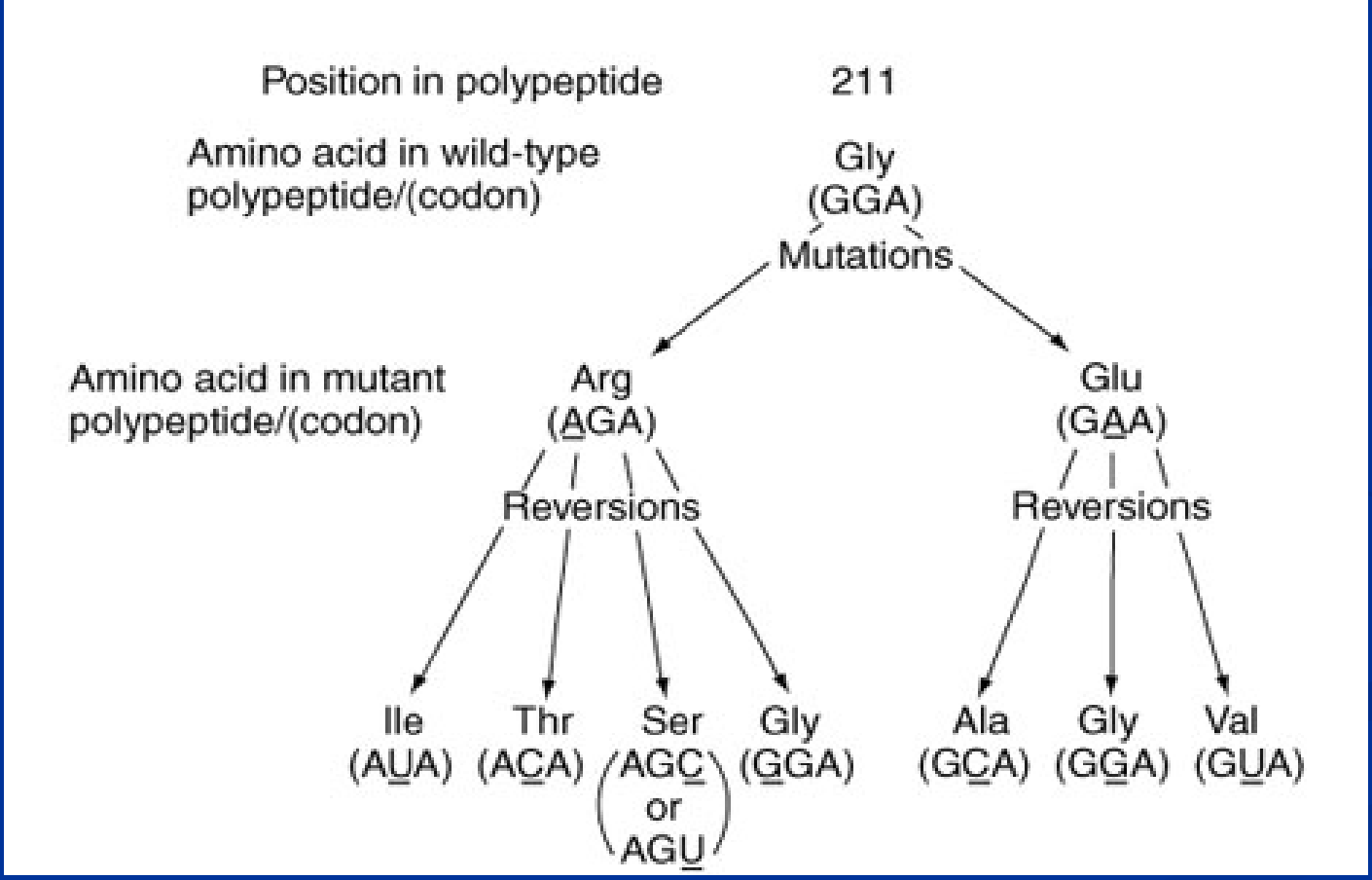
1/3 of the prize

USA

National Institutes of Health
Bethesda, MD, USA

Do living cells construct polypeptides according to same rules as *in vitro* experiments?

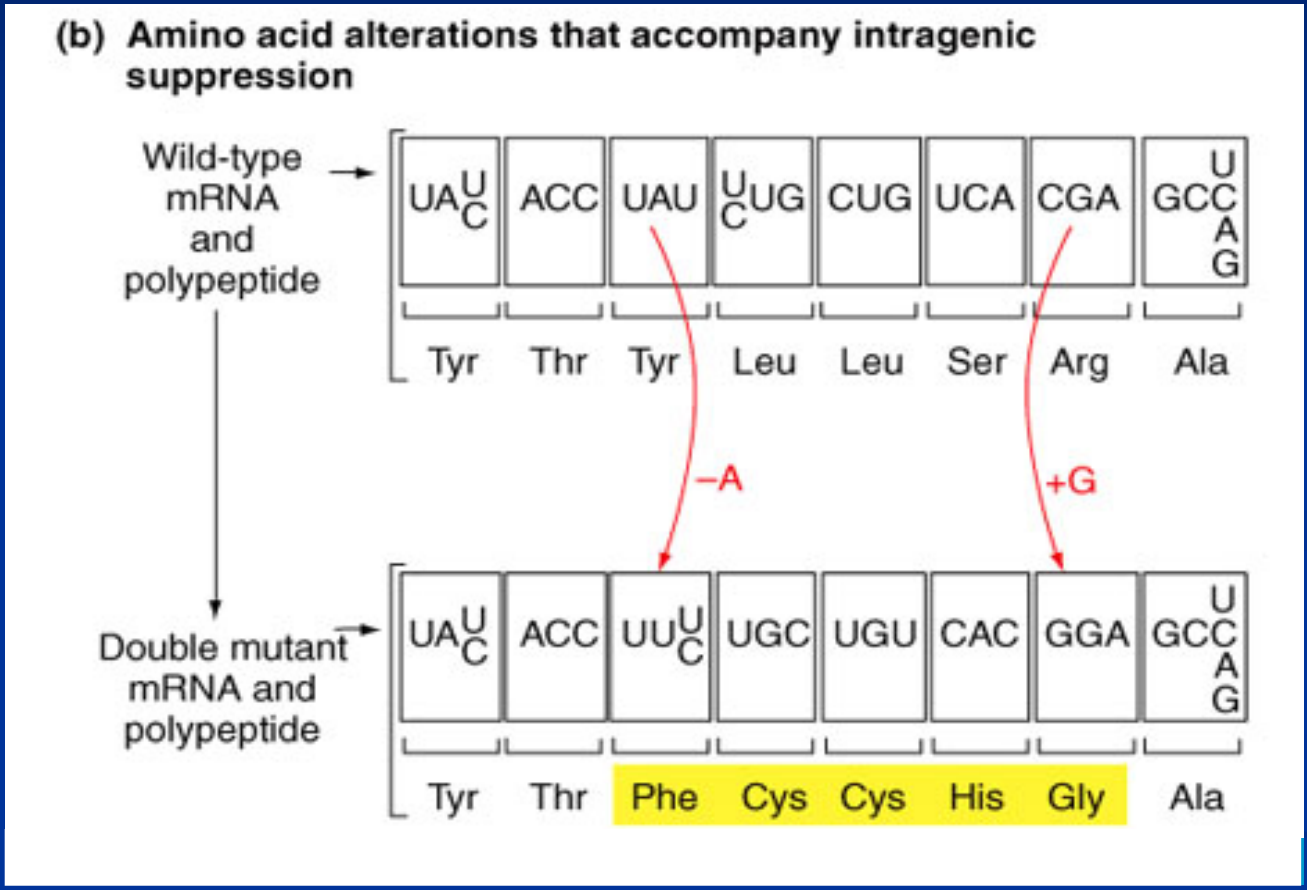
- **Charles Yanofsky**
- ***trp*$^{-}$ mutants of *E.coli* tryptophan synthetase subunit**
- **Single-base substitutions can explain the amino acid substitutions of *trp*$^{-}$ mutations and *trp*$^{+}$ revertants.**



Charles Yanofsky
(1925-)

Fig. 8.10 a

- **Proflavin treatment generates *trp*$^{-}$ mutants. Further treatment generates some *trp*$^{+}$ revertants.**



It makes sense only if codons do not overlap and are read from a fixed starting point with no pauses separating the adjacent triplets.

Fig. 8.10 b

The genetic code is almost, but not quite, universal

- Almost all living organisms use the same genetic code.
  - Translational system from one organism can use mRNA from another organism to generate protein.
  - Comparisons of DNA and protein sequence reveal perfect correspondence between codons and amino acids in almost all organisms.

A few exceptions to the genetic code

Codon	Most organisms	Ciliates	Yeast mitochondria	Human mitochondria
UAA, UAG	Stop	Gln		
UGA	Stop	Cys		Trp
CUA	Leu		Thr	
AGG, AGA	Arg			Stop
AUA	Ile			Met

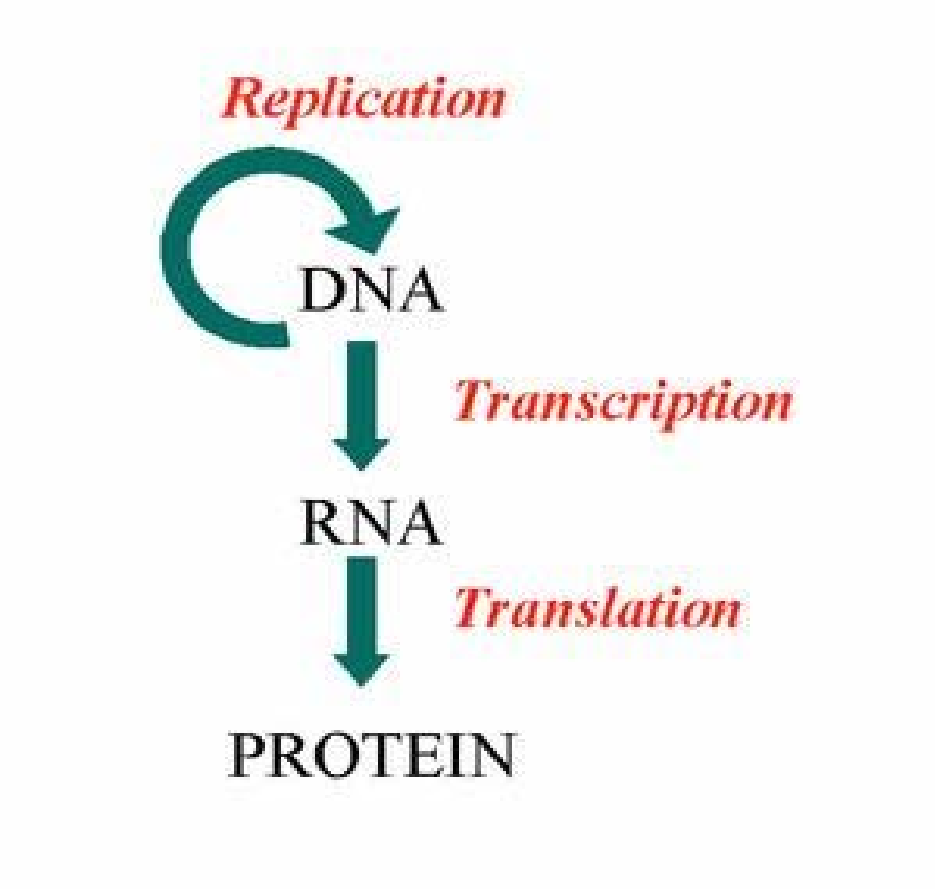
CUG specifies Ser in *Candida albicans*.

Summary of the genetic code

- The code consists of triplet codons, each of which specifies an amino acid.
- The code includes three stop codons, UAA, UAG, and UGA
- The position of the initiation codon (usua. AUG) establishes a reading frame.
- 5’- 3’ direction of mRNA corresponds with N-terminus to C-terminus of polypeptide.
- The code is nonoverlapping, degenerate and universal.

The Central Dogma

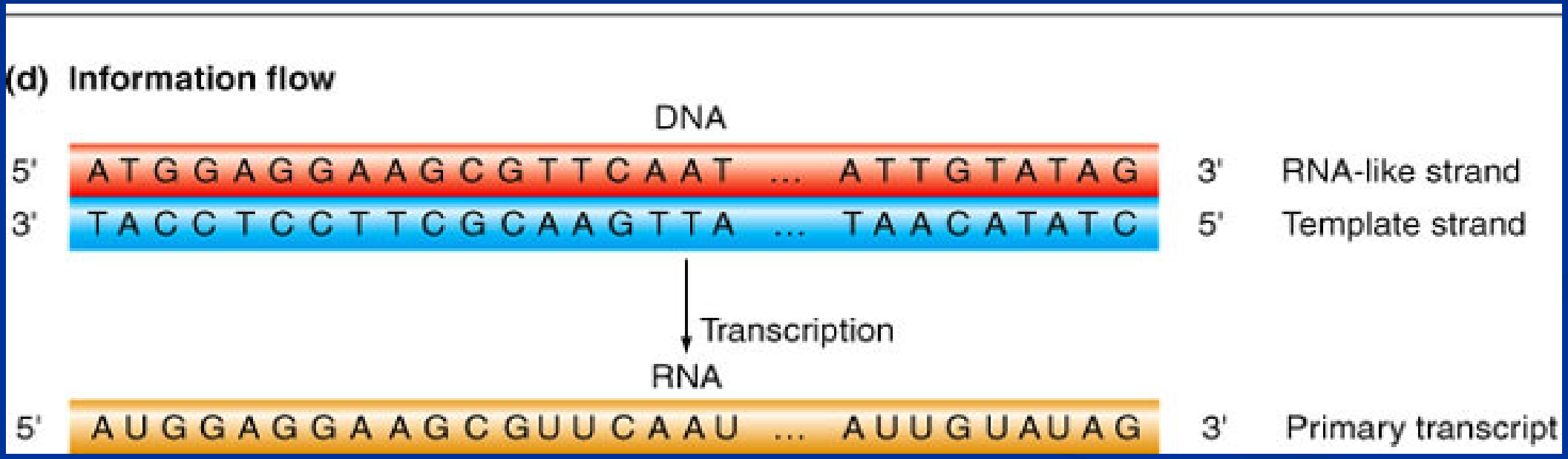
- **Proposed by Francis Crick in 1957.**
- ***Within each cell, genetic information flows from DNA to RNA to protein.***



Francis Crick
(1916-2004)

8.2 Transcription: From DNA to RNA

Transcription **is the conversion of DNA-encoded information to its RNA-encoded equivalent.**



- ***Sense strand***: **or RNA-like strand. The DNA strand in a gene which has the same sequence as the mRNA.**
- ***Antisense strand***: **or template strand.**

Fig. 8.11 d

- ***RNA polymerase*** **catalyzes transcription.**
- ***Promoters*** **signal RNA polymerase where to begin transcription.**
- **RNA polymerase adds nucleotides to the growing RNA polymer in 5’ to 3’ direction.**
- ***Terminator sequences*** **tell RNA polymerase where to stop transcription.**

Promoters of 10 different bacterial genes

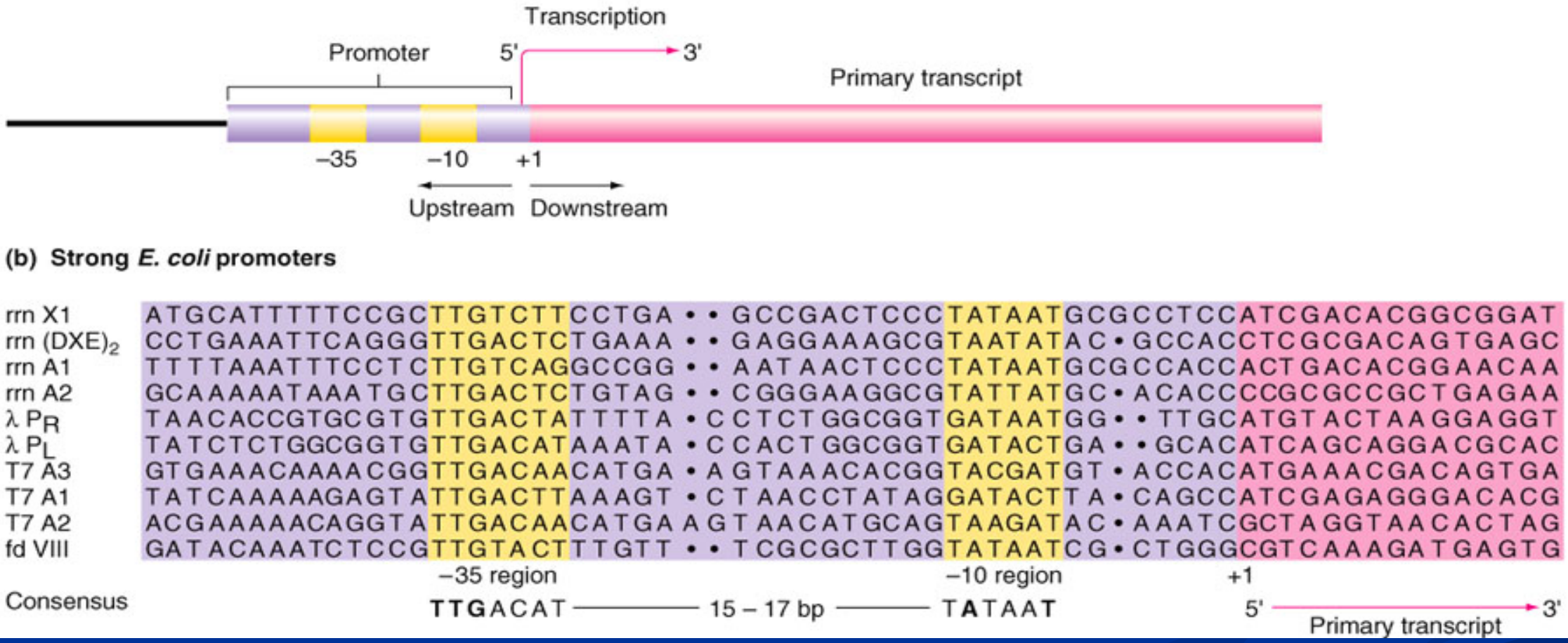


Fig. 8.12

Initiation of transcription

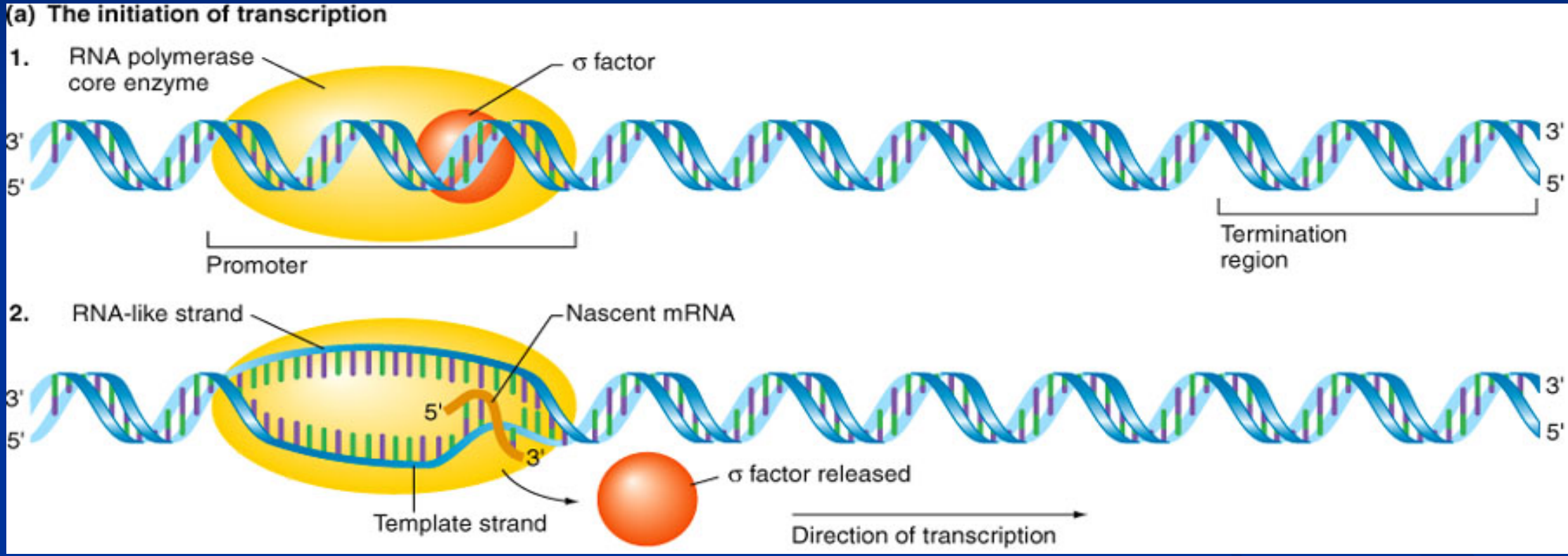


Fig. 8.11 a

Elongation

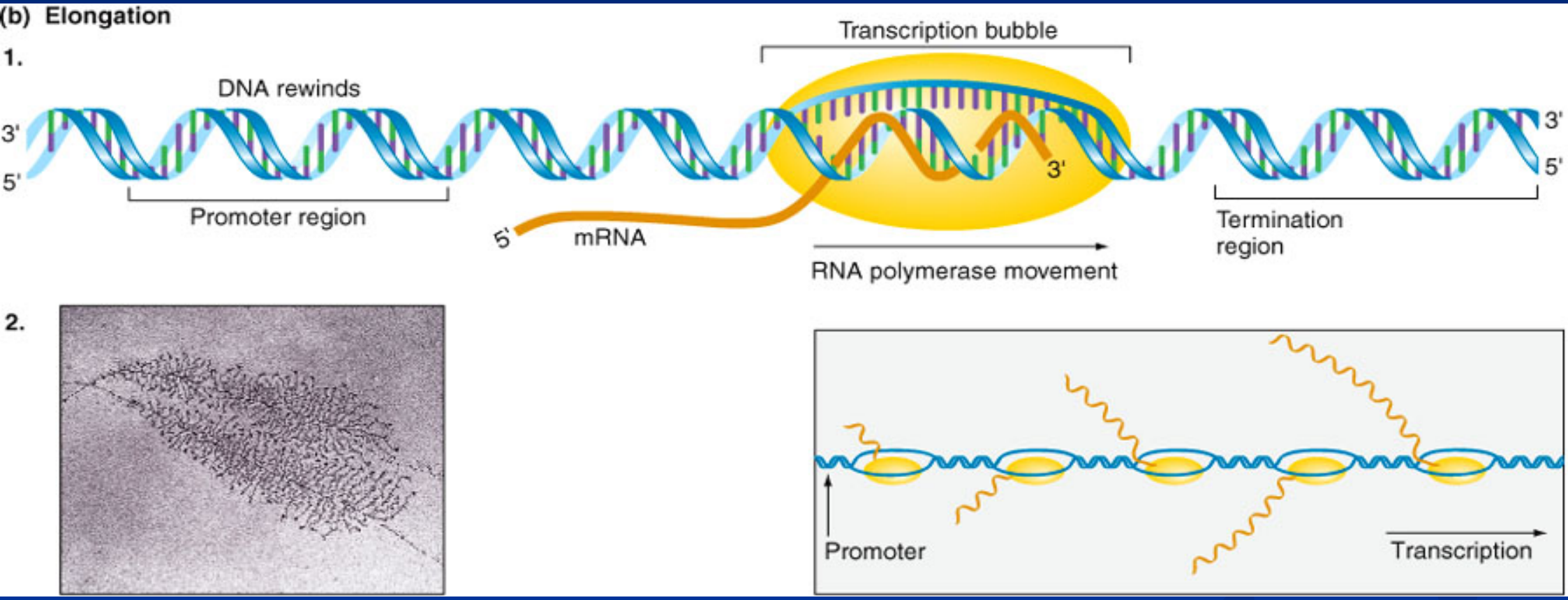


Fig. 8.11 b

Termination

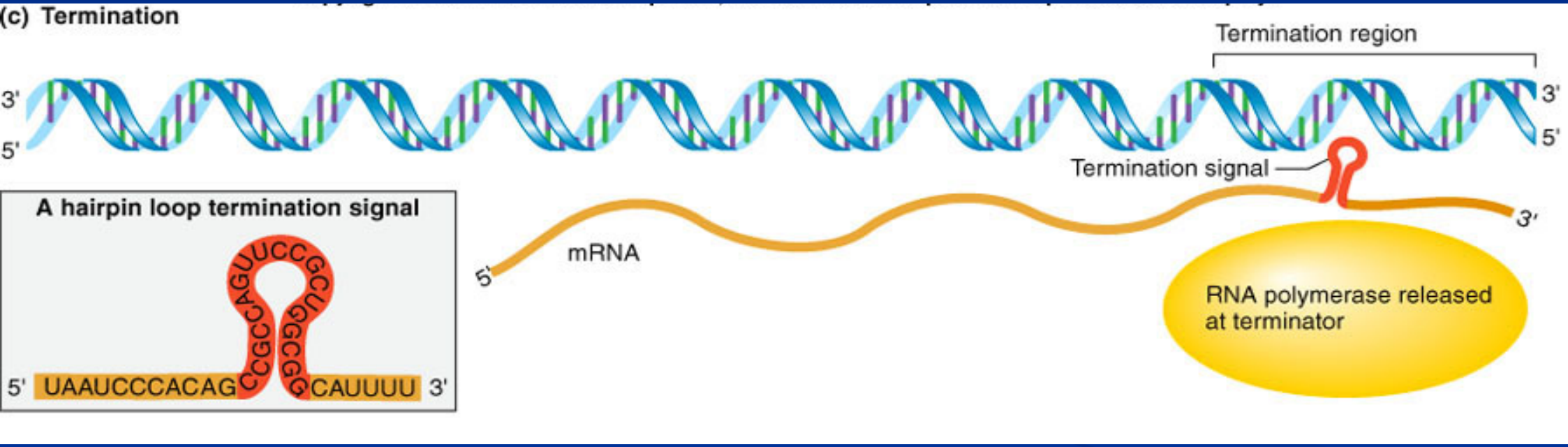


Fig. 8.11 c

In eukaryotes, RNA is processed after transcription

- A 5’ methylated cap and a 3’ poly-A tail are added.
- Structure of the methylated cap.

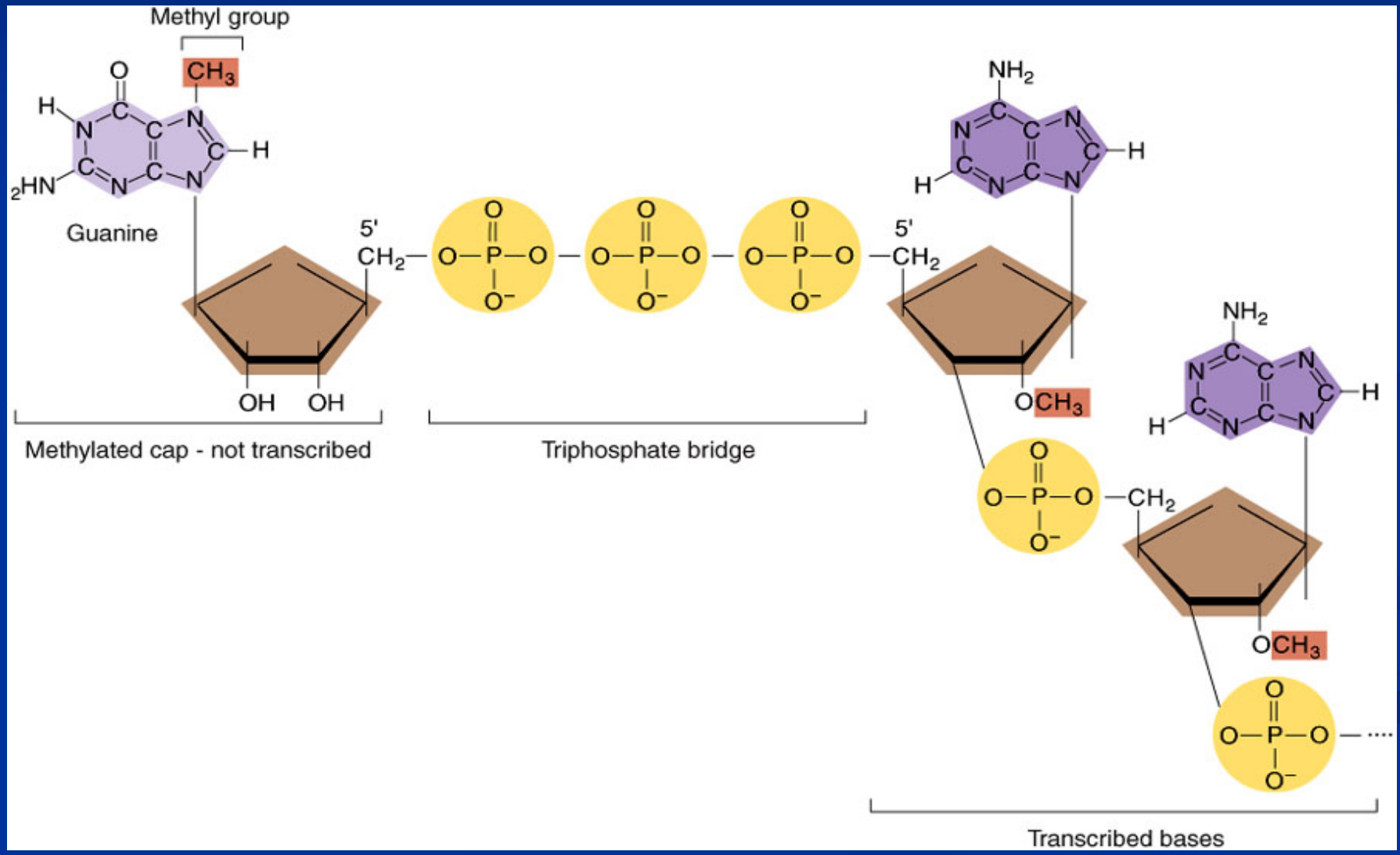


Fig. 8.13

Poly-A tail is added to 3’ end of mRNA

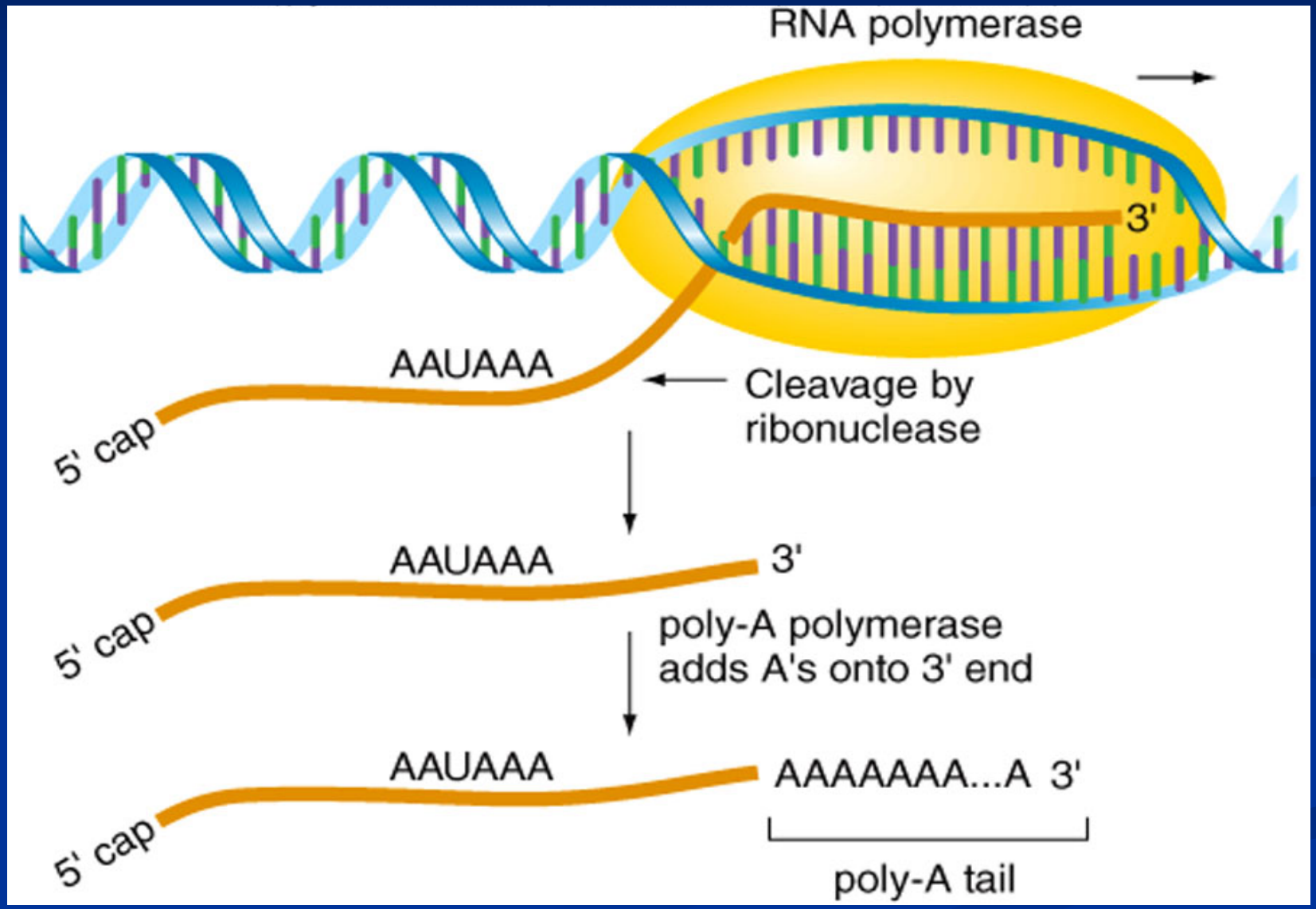
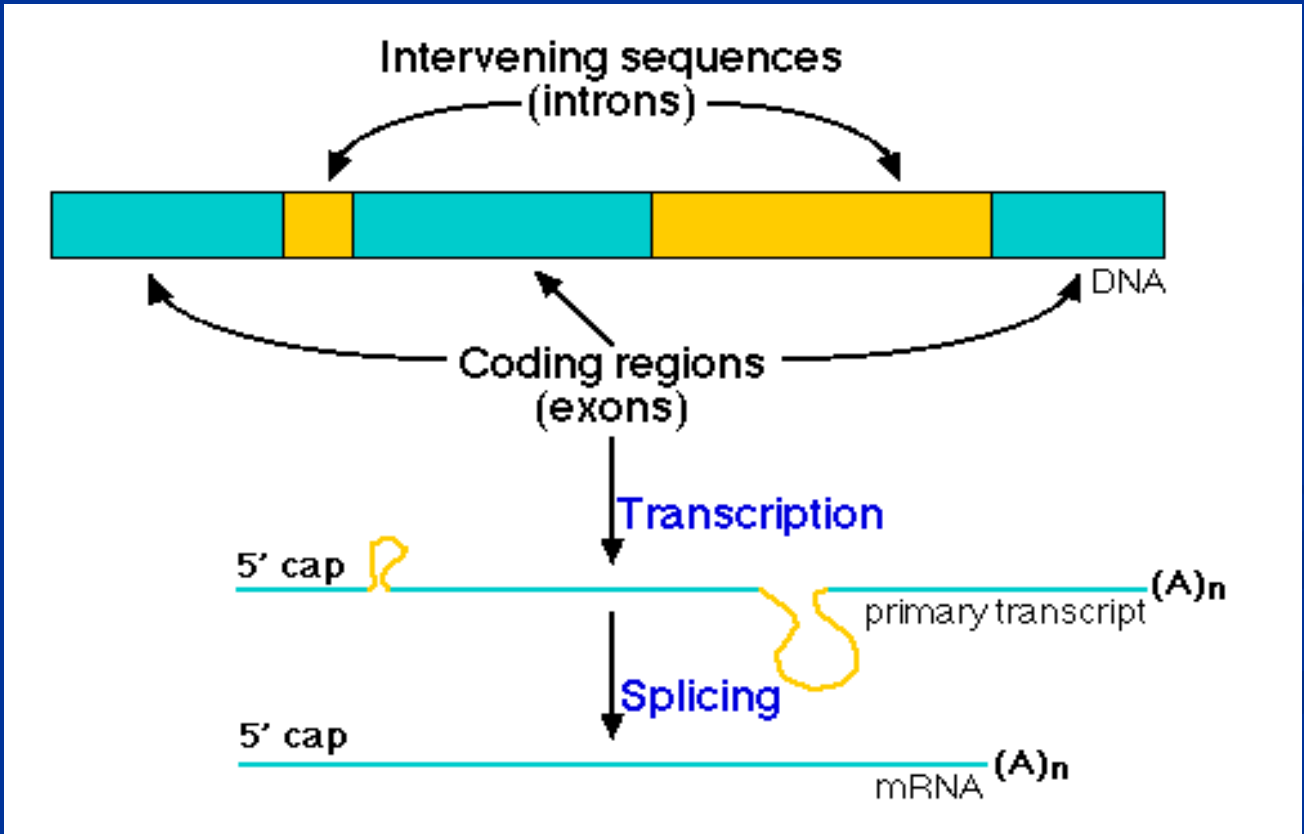


Fig. 8.14

RNA splicing removes introns

- ***Exons*** **– sequences found in a gene's DNA and mature mRNA (expressed regions)**
- ***Introns*** **– sequences found in DNA but not in mRNA (intervening regions)**



Human dystrophin gene underlying Duchenne muscular dystrophy (DMD)

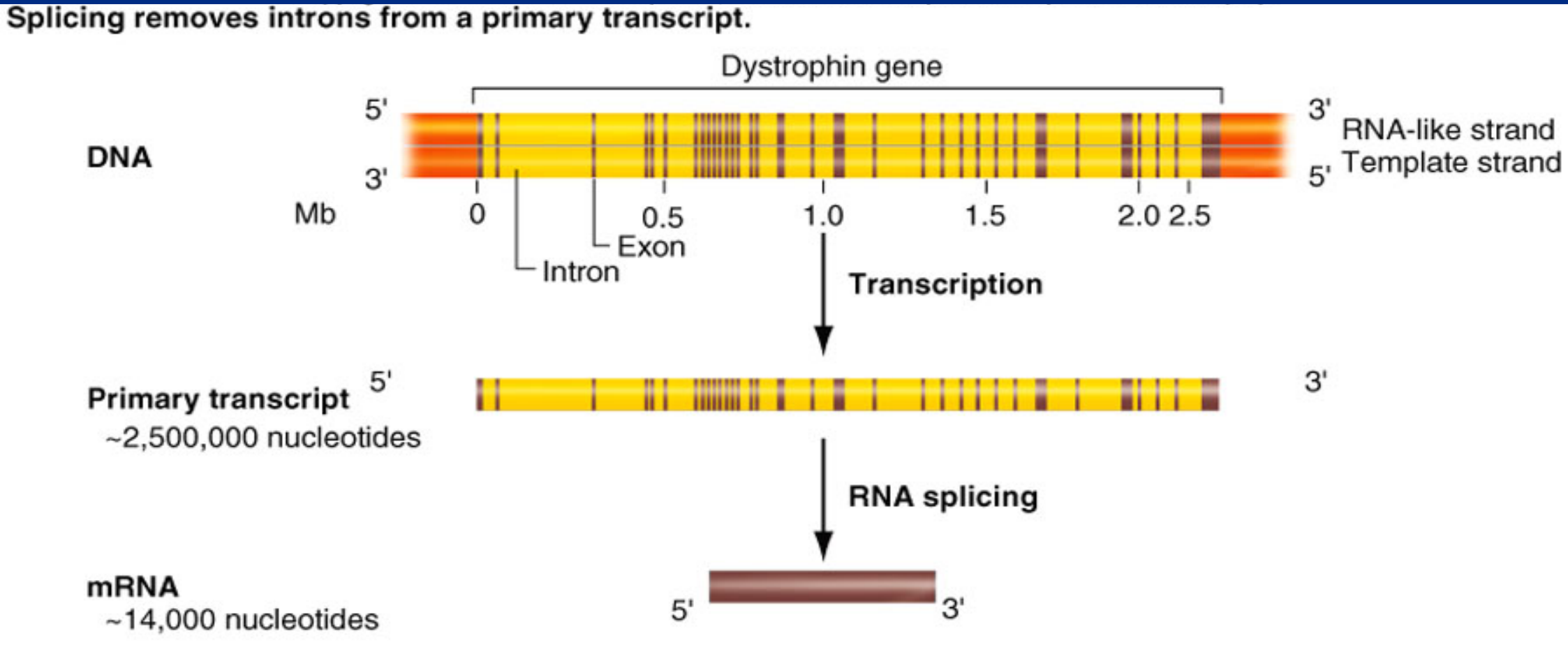


Fig. 8.15

How RNA processing splices out introns and joins adjacent exons

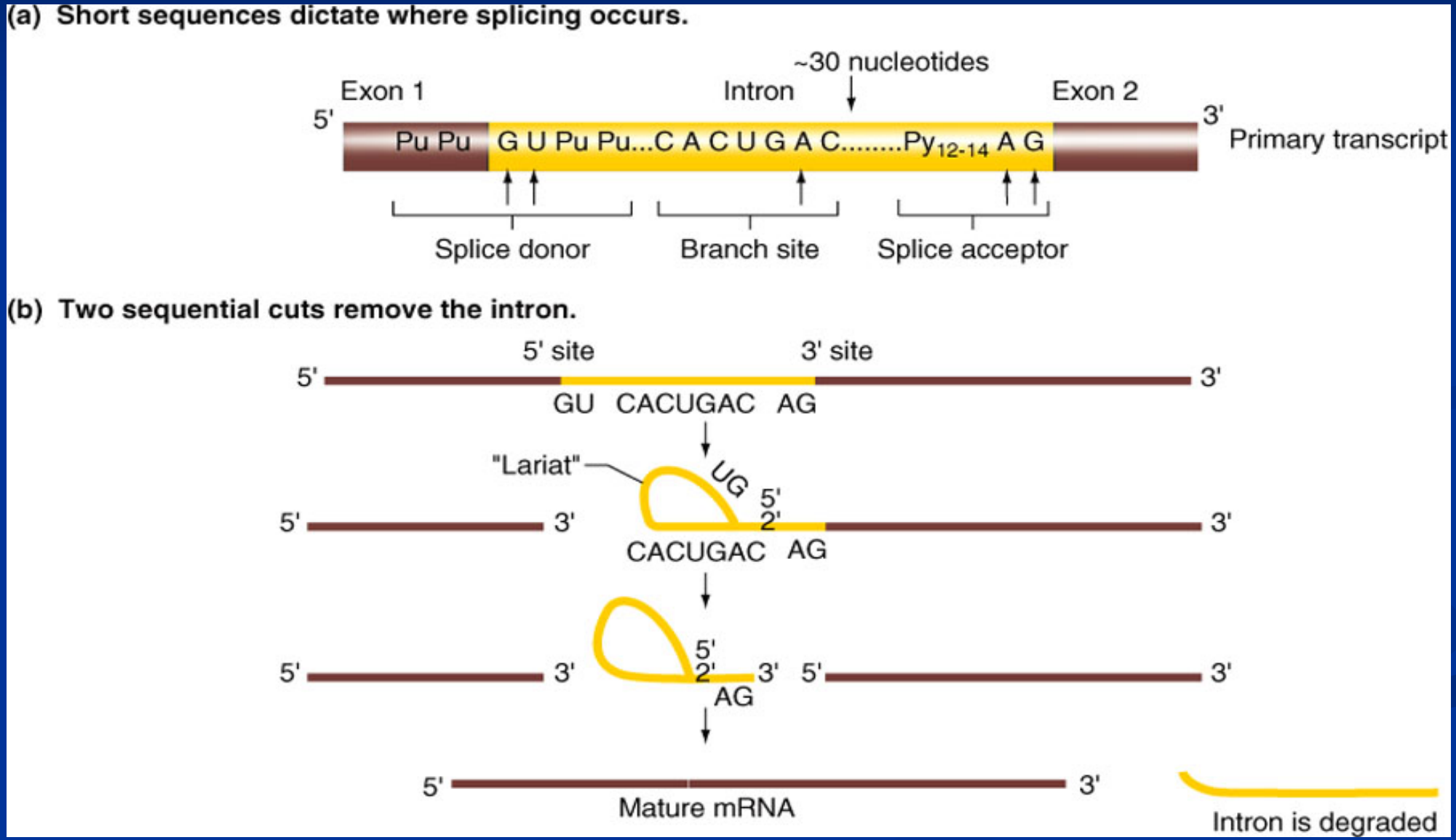


Fig. 8.16

Splicing is catalyzed by spliceosomes.

- **Ribozymes** – RNA molecules that act as enzymes
- Ensures that all splicing reactions take place in concert

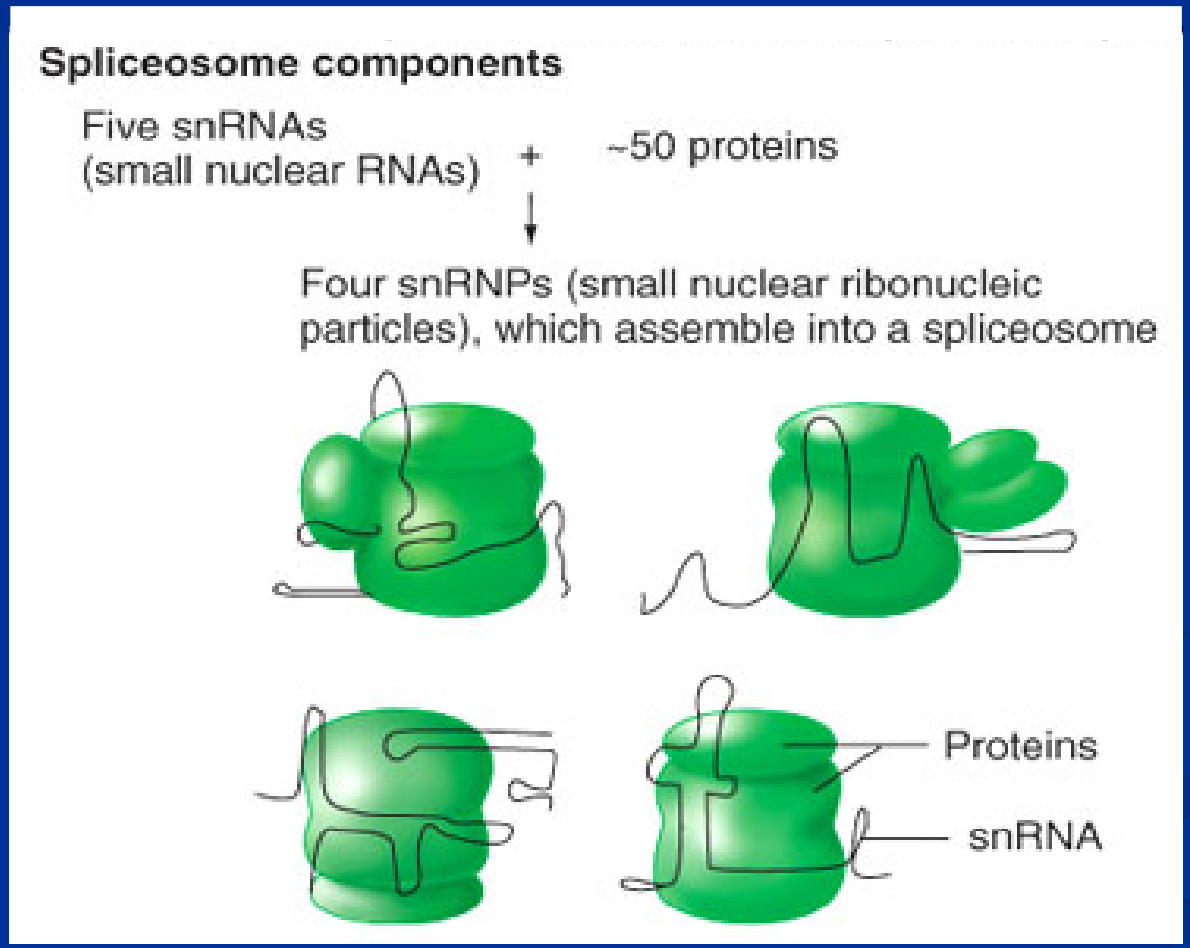


Fig. 8.17

Alternative splicing:

Production of different mature RNAs from the same primary transcript by joining different combinations of exons.

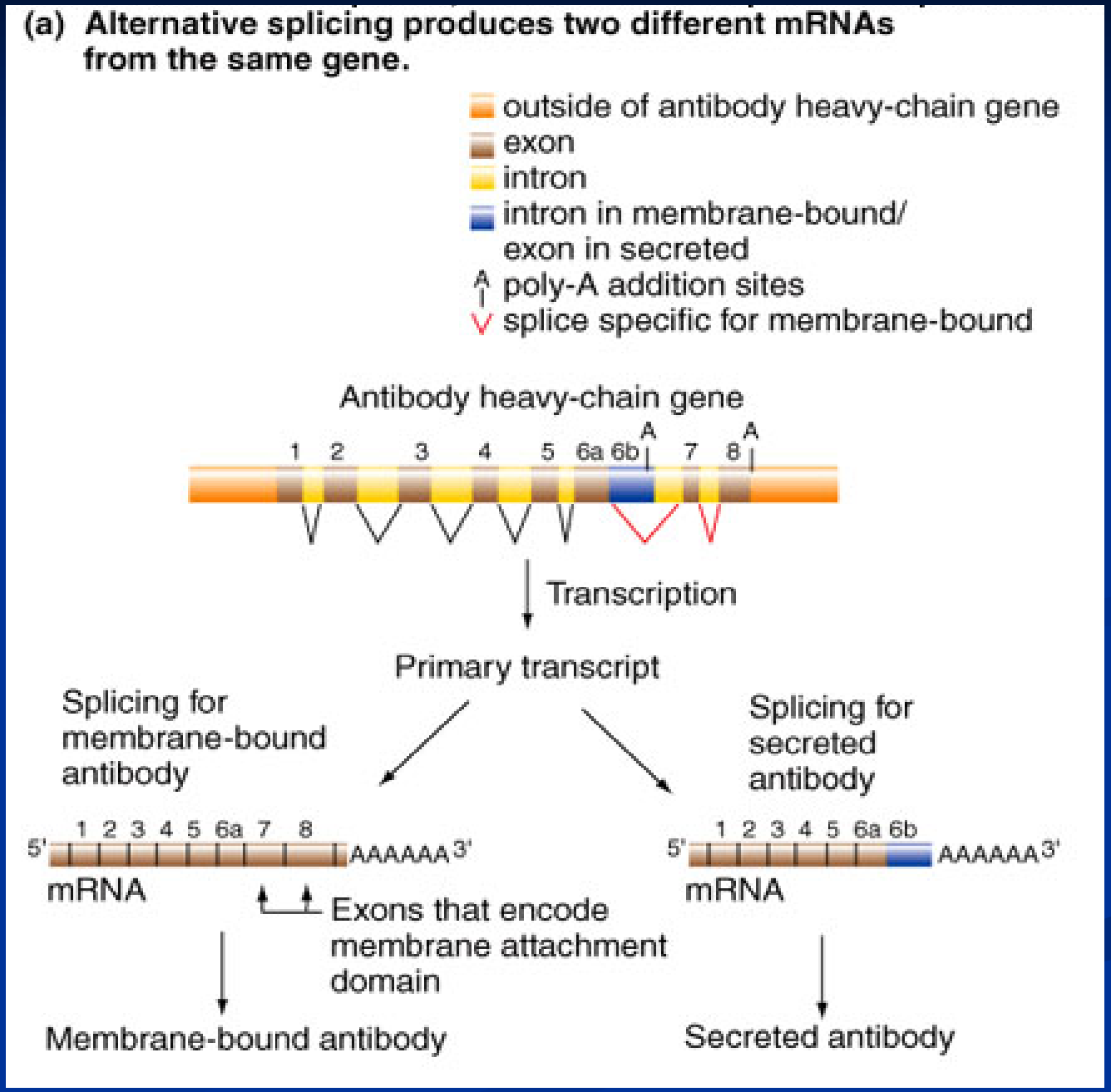


Fig. 8.18

8.3 Translation: From mRNA to protein

8.3 Translation: From mRNA to protein

Translation **is the process in which the codons carried by mRNA direct the synthesis of polypeptides from amino acids according to the genetic code.**

- ***Transfer RNAs*** **(tRNAs) mediate translation of mRNA codons to amino acids.**
  - **Short, single-stranded, 74-95 nucleotides.**
  - **tRNAs carry *anticodon* on one end.**
    - **Three nucleotides complementary to an mRNA codon**
  - **Base pairing between an mRNA codon and a tRNA anticodon directs amino acid incorporation into a growing polypeptide.**
  - **Charged tRNA is covalently coupled to its amino acid.**

Each tRNA has a primary, secondary, and tertiary structure

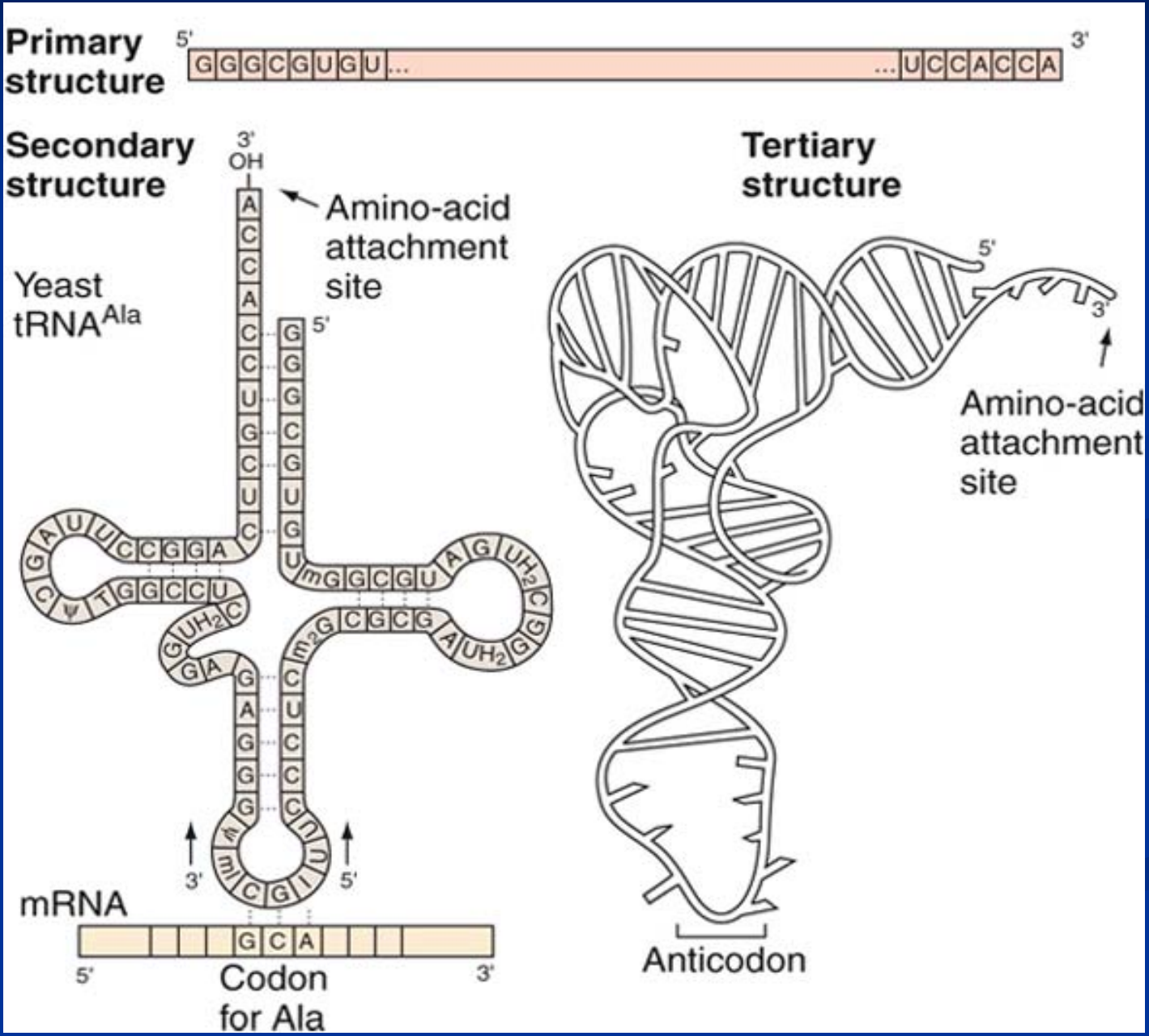


Fig. 8.19 b

Many tRNAs contain modified bases

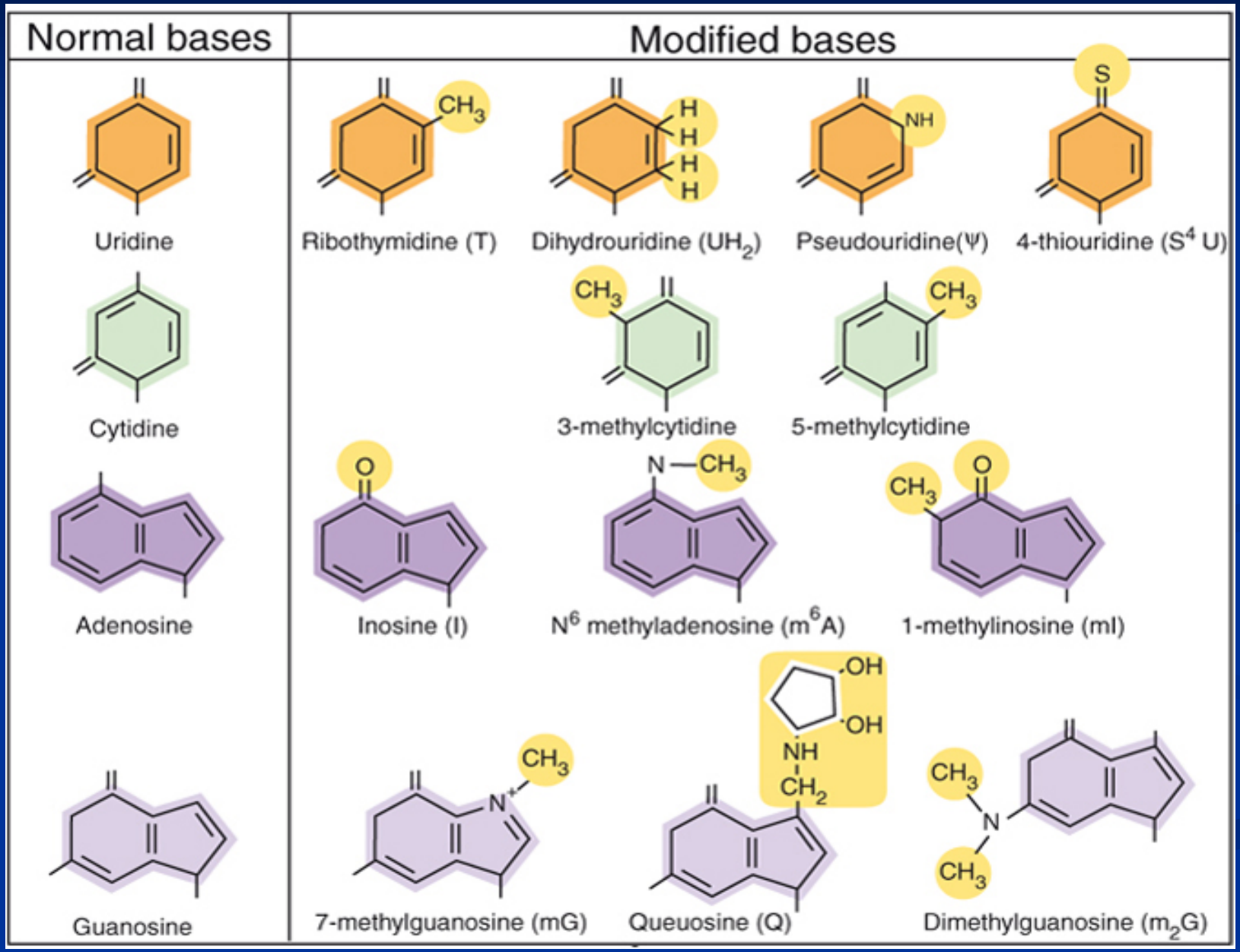


Fig. 8.19 a

Aminoacyl-tRNA synthetase catalyzes attachment of tRNAs to corresponding amino acid

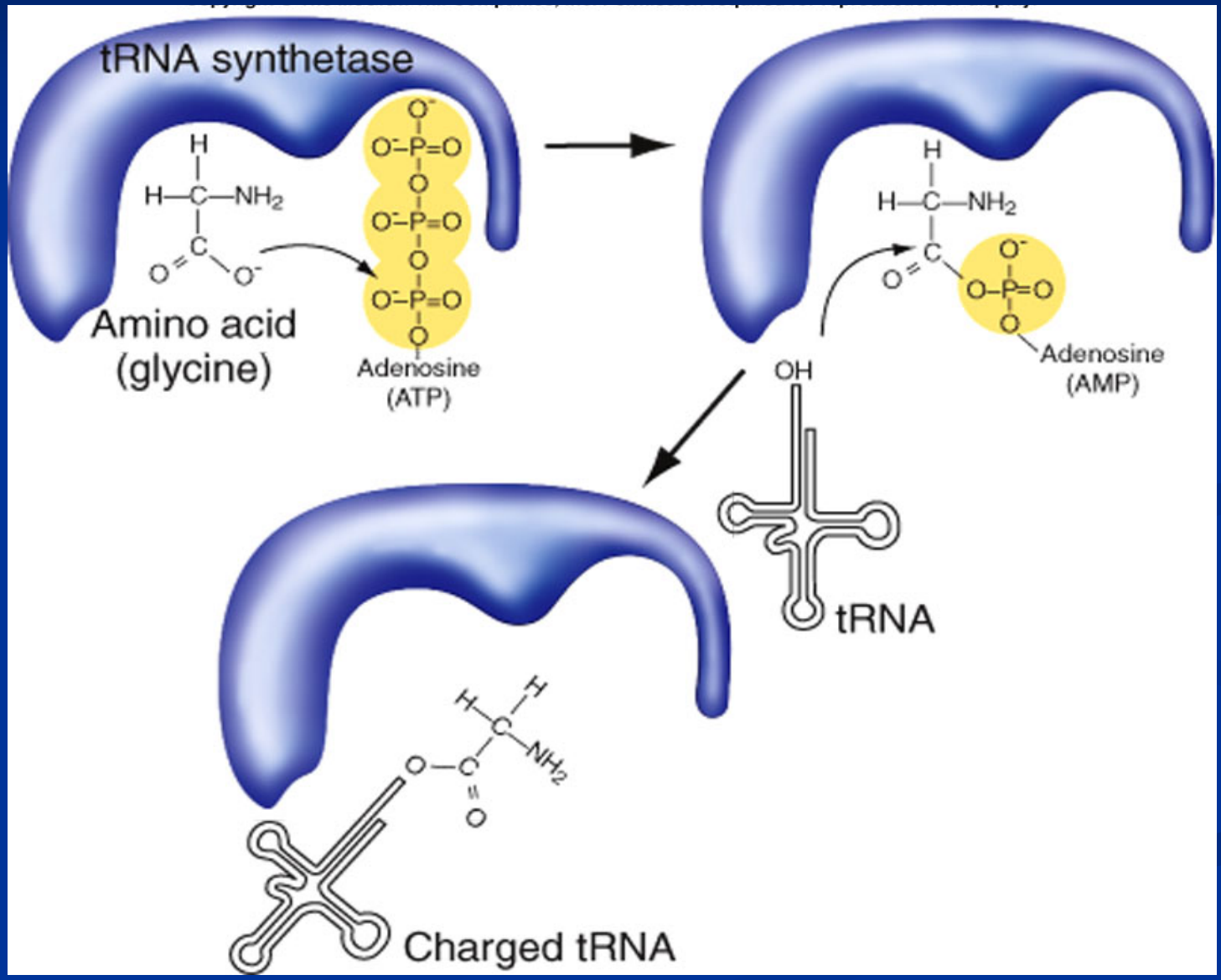


Fig. 8.20

Base pairing between mRNA codon and tRNA anticodon determines where incorporation of amino acid occurs

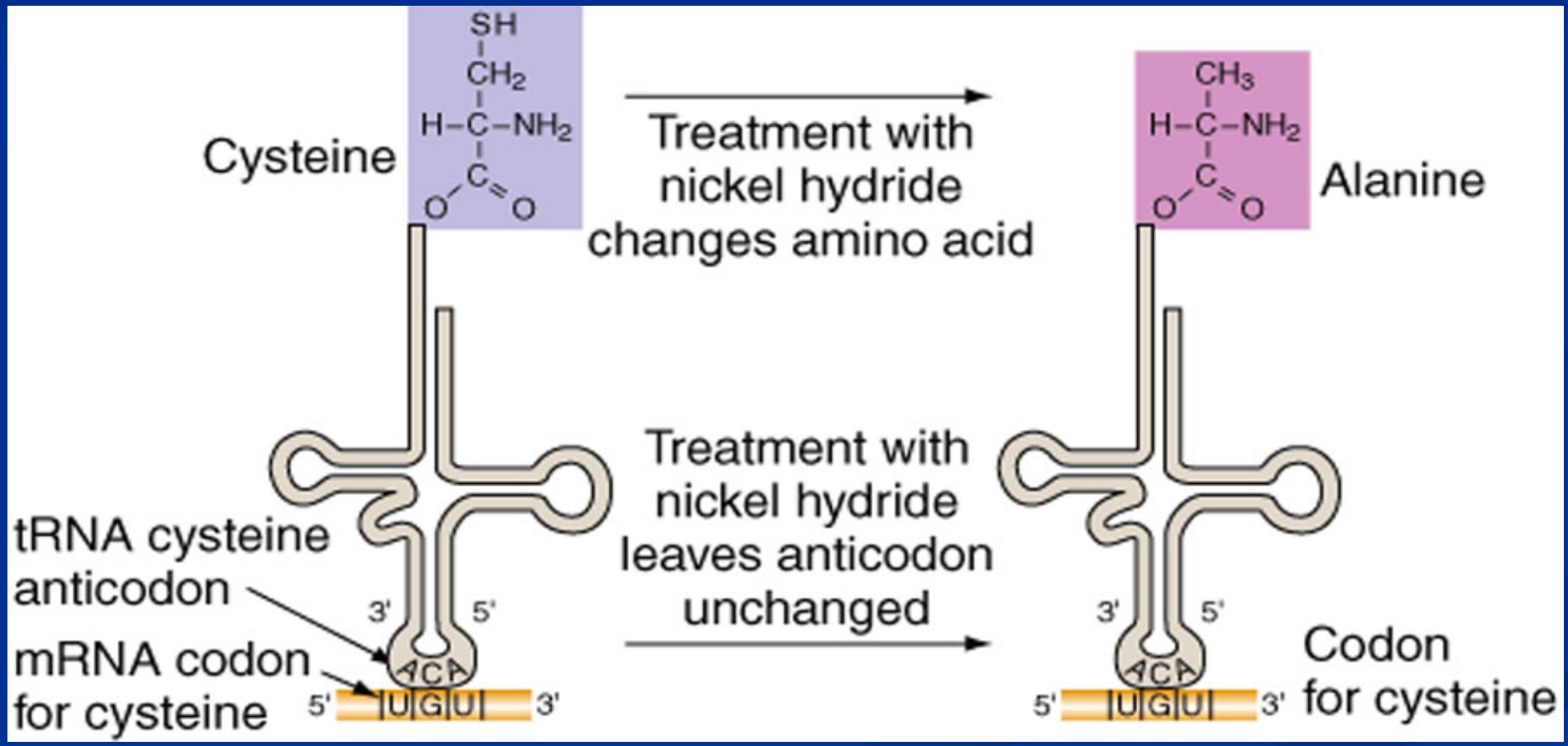


Fig. 8.21

Wobble: Some tRNAs recognize more than one codon for amino acids they carry

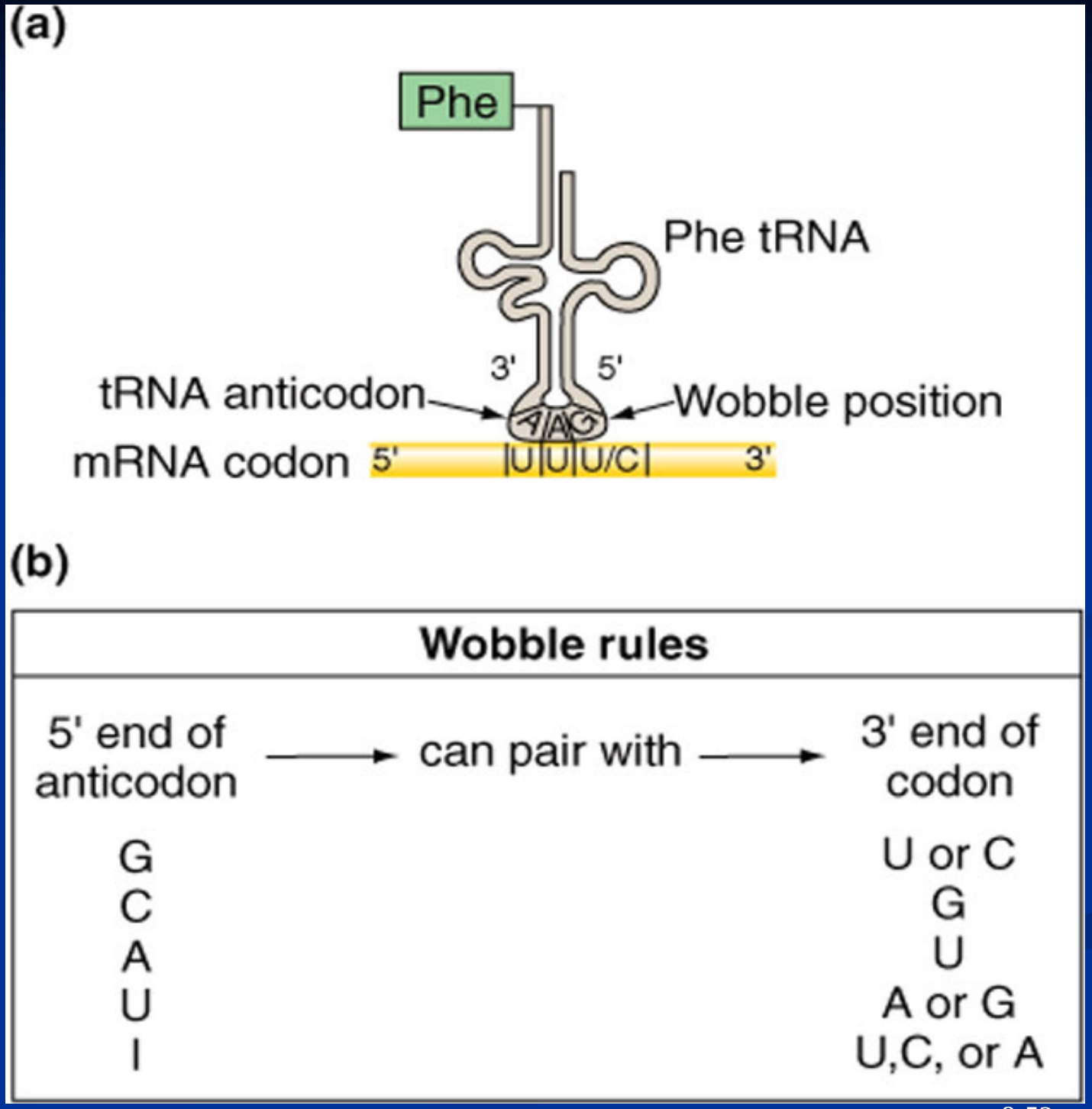


Fig. 8.22

Ribosomes are site of polypeptide synthesis

- Ribosomes are complex structures composed of RNA and protein.

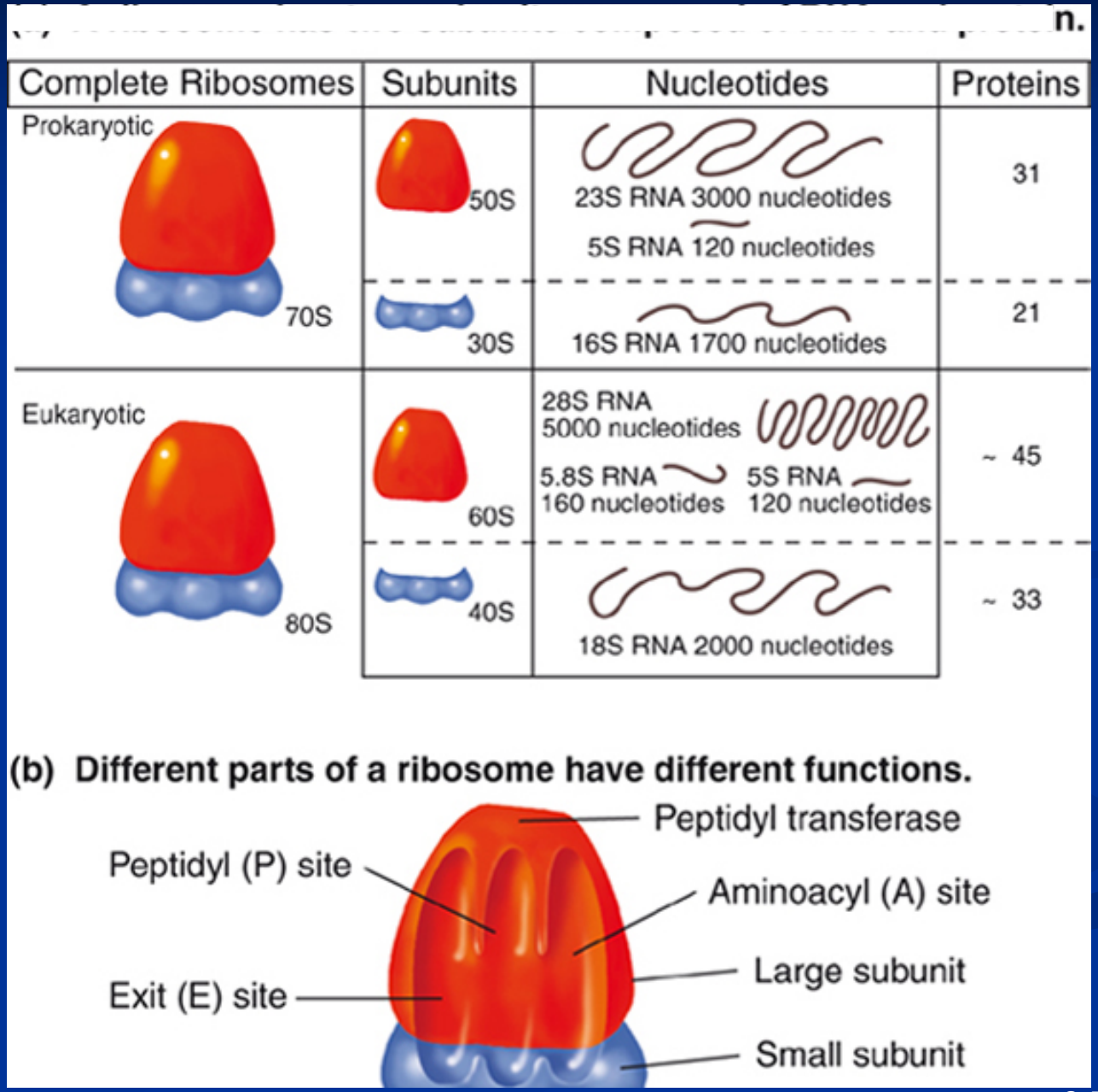


Fig. 8.23

Mechanism of translation

- ***Initiation*** **sets stage for polypeptide synthesis.**
  - **AUG start codon at 5' end of mRNA.**
  - **N-formylmethionine (fMet) on initiation tRNA.**
    - **First amino acid incorporated in bacteria.**
- ***Elongation*** **during which amino acids are added to growing polypeptide.**
  - **Ribosomes move in 5'-3' direction revealing codons.**
  - **Addition of amino acids to C terminus.**
  - **2-15 amino acids per second.**
- ***Termination*** **which halts polypeptide synthesis.**
  - **Nonsense codon recognized at 3' end of reading frame.**
  - **Release factor proteins and halt polypeptide synthesis.**

- **Posttranslational processing can modify a polypeptide's structure.**

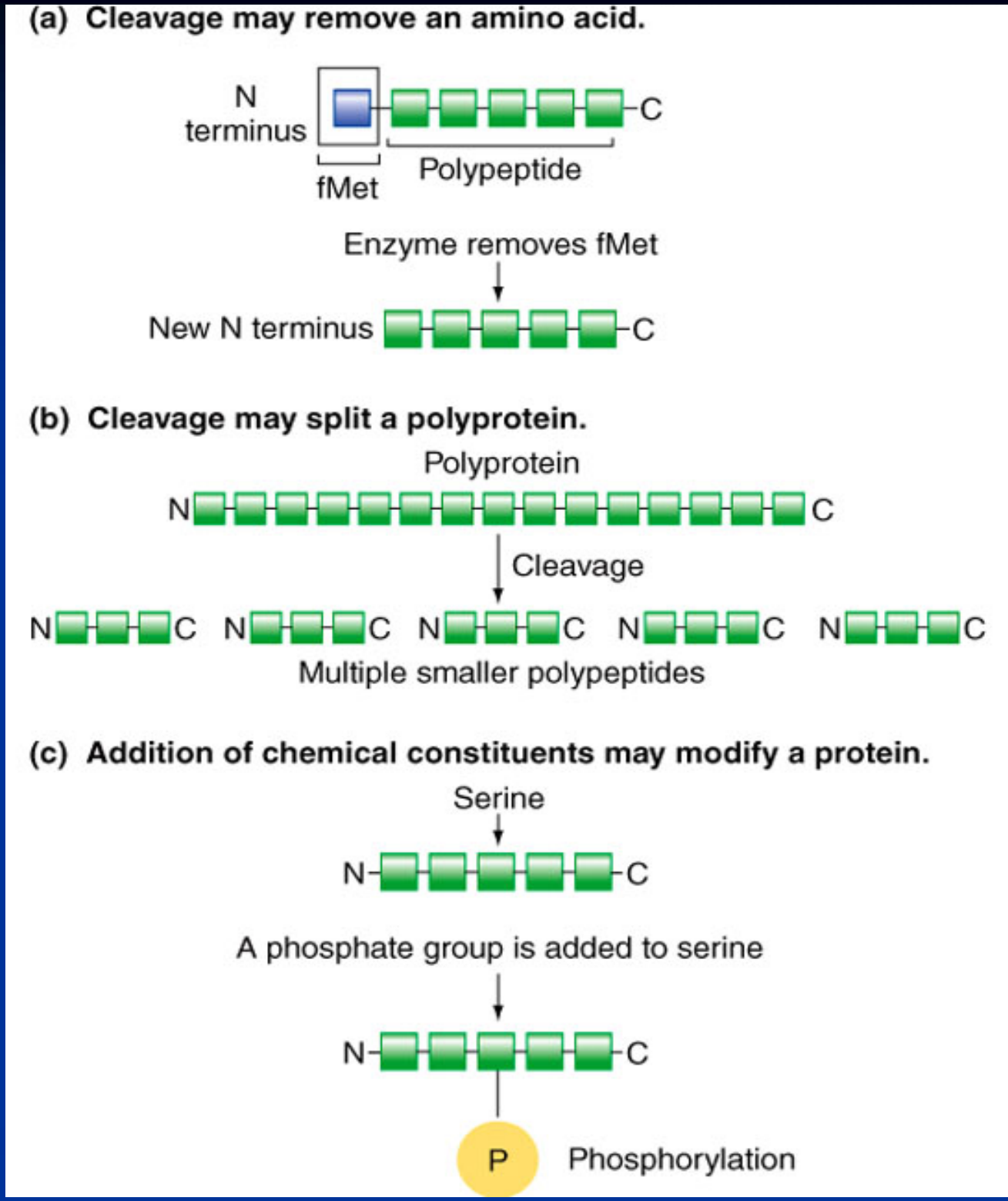


Fig. 8.26

8.4 Differences in gene expression between prokaryotes and eukaryotes

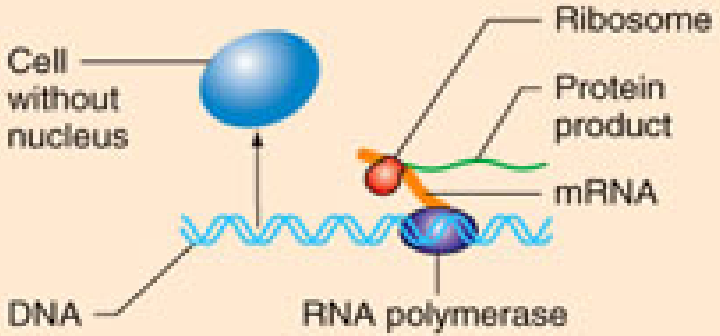
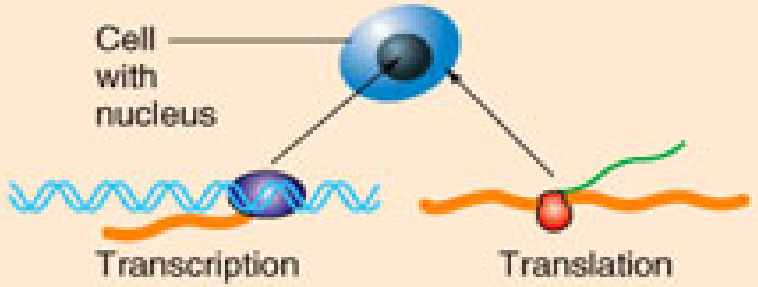
	Prokaryotes	Eukaryotes
Overview	1. No nucleus. Transcription and translation thus take place in the same cellular compartments, and translation is often coupled to transcription.  2. Genes are not divided into exons and introns.  	1. Nucleus separated from the cytoplasm by a nuclear membrane. Transcription takes place in the nucleus, while translation occurs in the cytoplasm. Direct coupling of transcription and translation is thus not possible.  2. The DNA of a gene consists of exons separated by introns; the exons are defined by posttranscriptional splicing, which deletes the introns. 

Table 8.1

	Prokaryotes	Eukaryotes
Transcription	1. One RNA polymerase consisting of five subunits. 2. Primary transcripts are the actual mRNAs; they have a triphosphate start at the 5′ end and no tail at the 3′ end. 	1. Several kinds of RNA polymerase, each containing 10 or more subunits; different polymerases transcribe different genes. 2. Primary transcripts undergo processing to produce mature mRNAs that have a methylated cap at the 5′ end and a poly-A tail at the 3′ end. 
Translation	1. Unique initiator tRNA carries formylmethionine. 2. mRNAs have multiple ribosome binding sites and can thus direct the synthesis of several different polypeptides.  3. Small ribosomal subunit immediately binds to the mRNA's ribosome binding site. 	1. Initiator tRNA carries methionine. 2. mRNAs have only one start site and can thus direct the synthesis of only one kind of polypeptide.  3. Small ribosomal subunit binds first to the methylated cap at the 5′ end of the mature mRNA and then scans the mRNA to find the ribosome binding site. 

8.6 The effect of mutations on gene expression and gene function

Mutations in a gene's coding sequence may alter the gene product.

- ***Silent mutations*** do not alter amino acid specified.
- ***Missense mutations*** replace one amino acid with another.
- ***Nonsense mutations*** change an amino-acid-specifying codon to a stop codon.
- ***Frameshift mutations*** result from the insertion or deletion of nucleotides within the coding sequence.

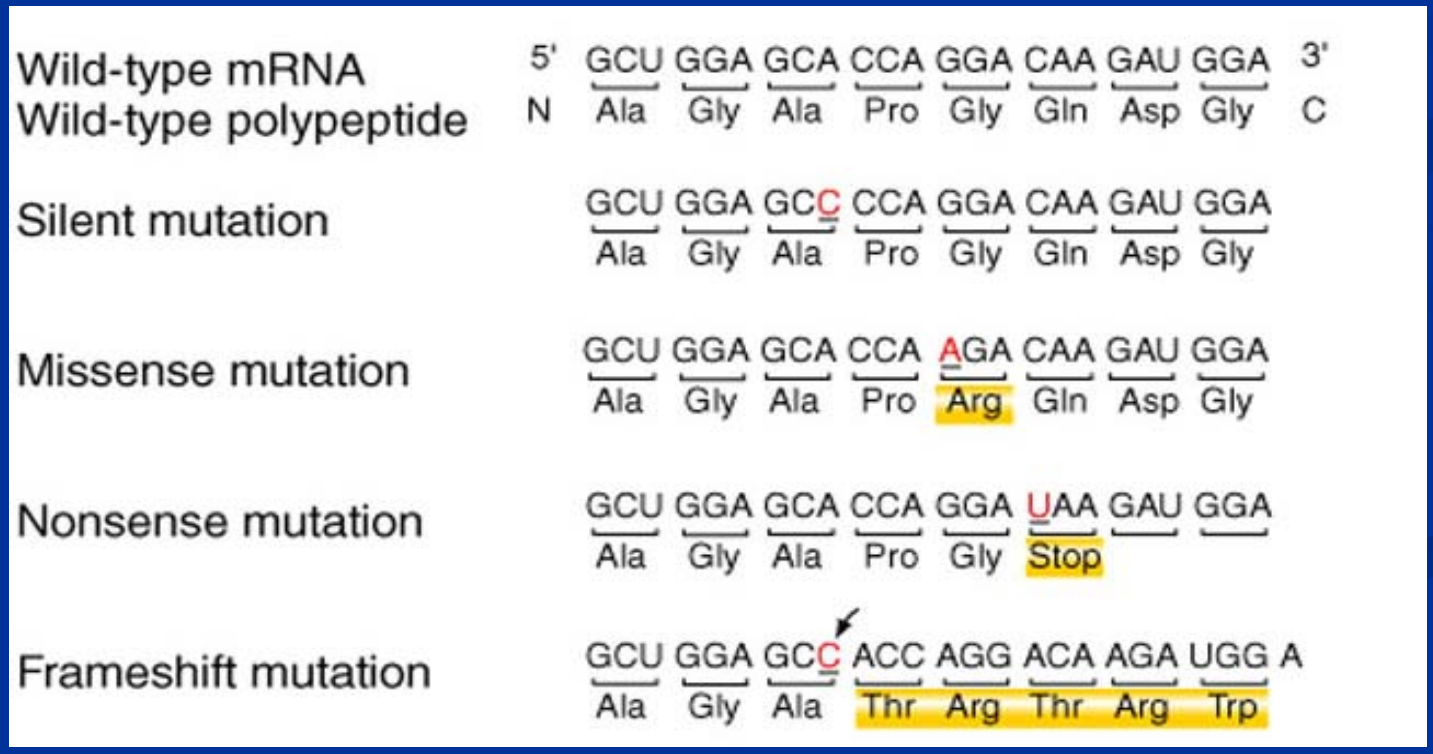


Fig. 8.28 a

- **Mutations outside the coding sequence can also alter gene expression.**
  - **Promoter sequences**
  - **Termination signals**
  - **Splice-acceptor and splice-donor sites**
  - **Ribosome binding sites**

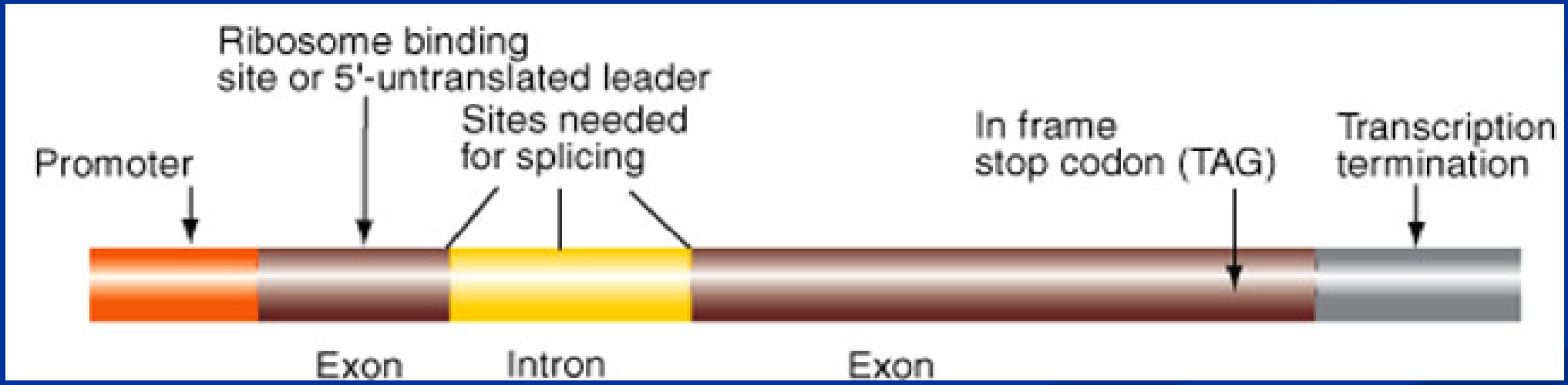
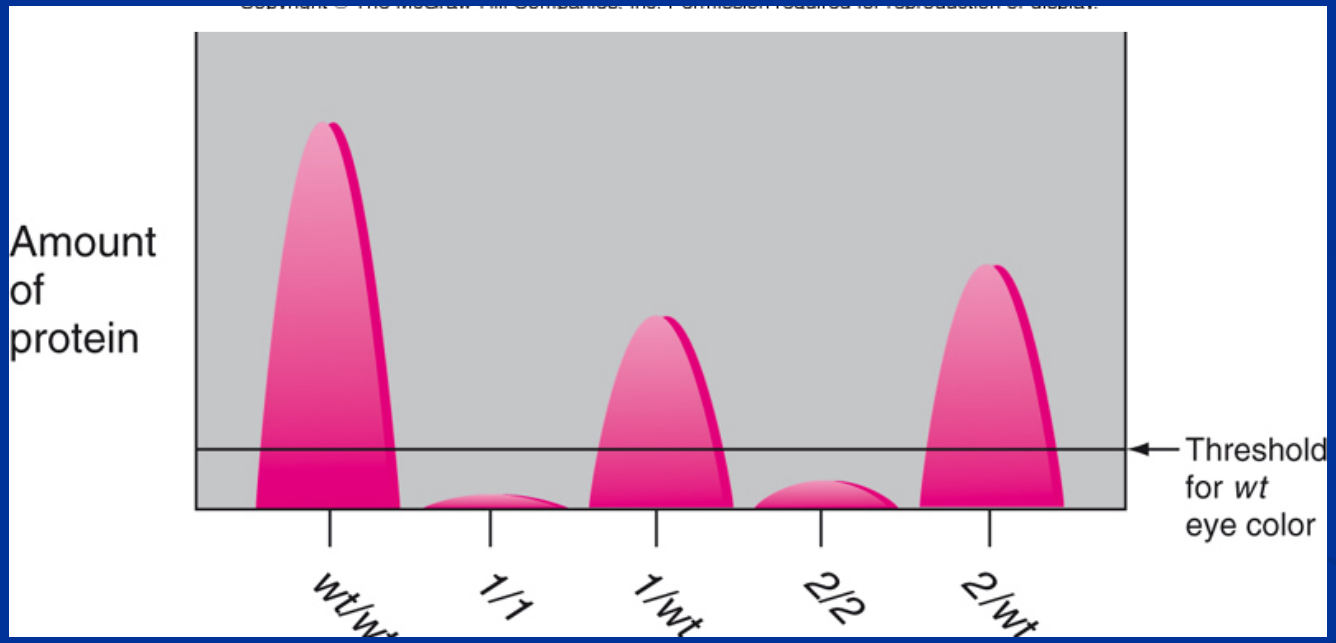


Fig. 8.28 b

Most mutations that affect gene expression reduce gene function

- **Null** or **amorphic mutations** are alleles that completely block the function of a protein.
- **Hypomorphic mutations** produce much less of a protein or a protein with weak but detectable function.



Rocket immunoelectrophoresis reveals the amount of xanthine dehydrogenase produced in flies with different genotypes.

Null allele 1 and hypomorphic allele 2 are recessive to wild-type.

Fig. 8.29

Incomplete dominance arises when phenotype varies in proportion to the amount of functional protein

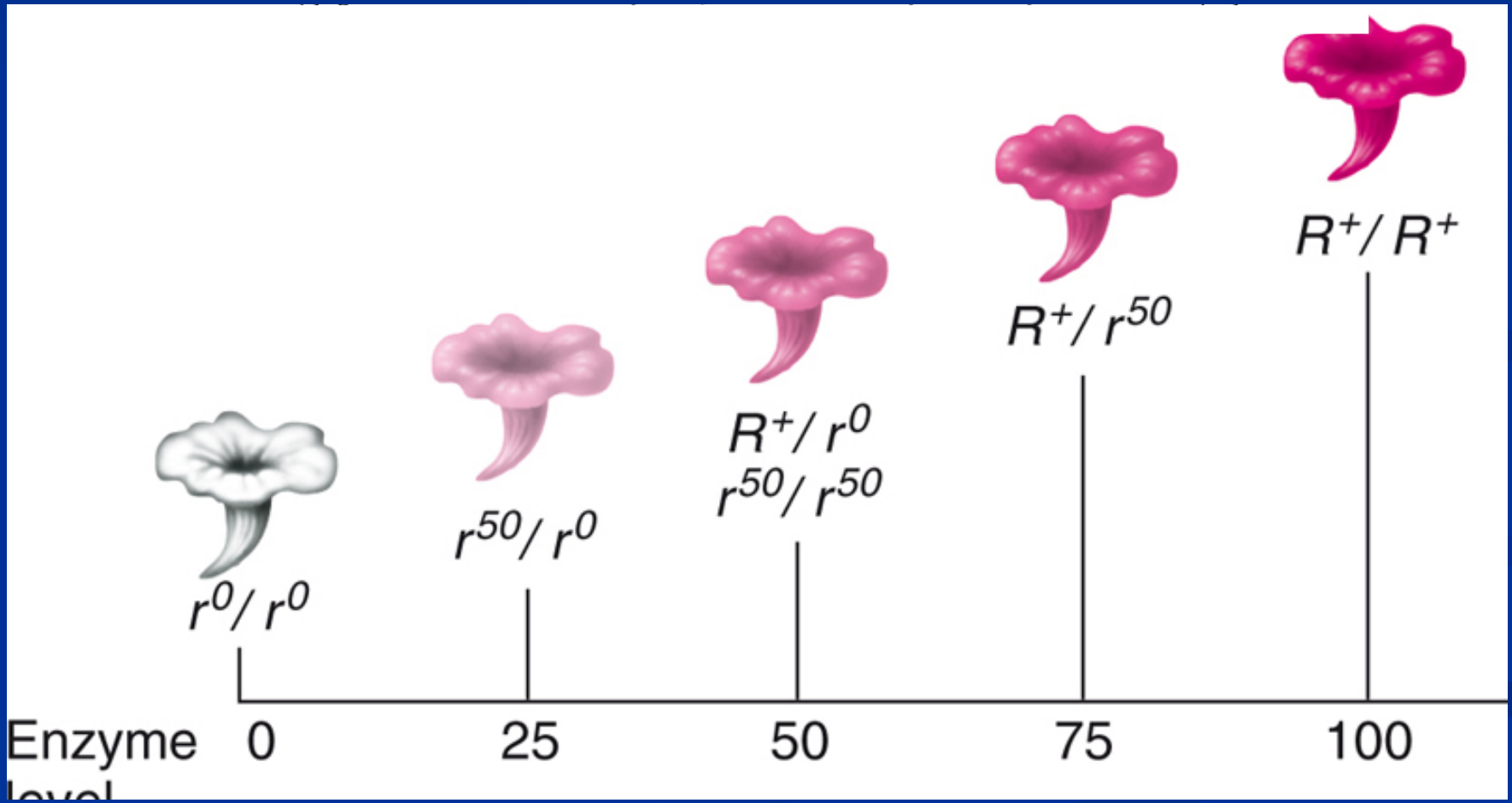
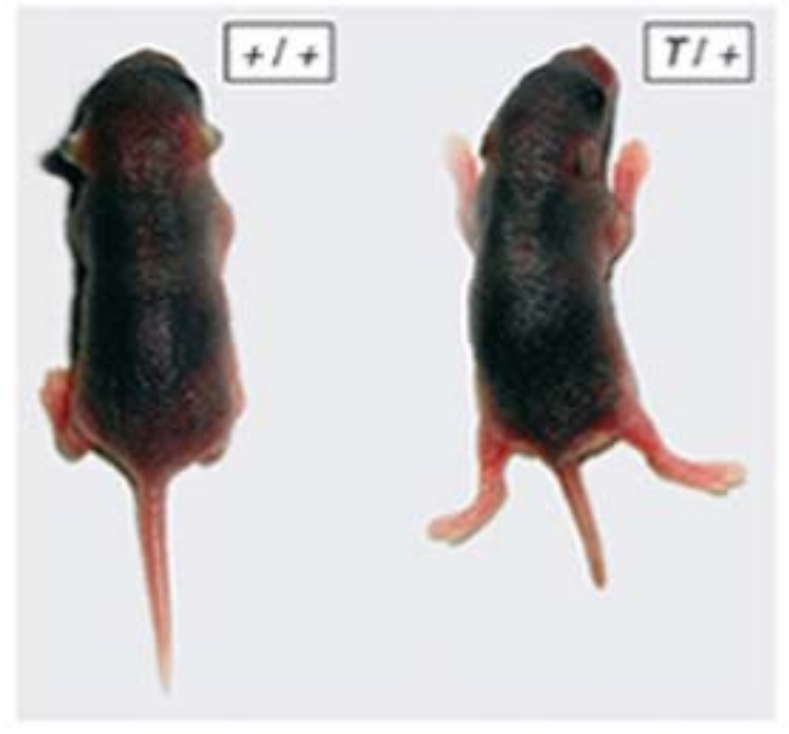


Fig. 8.30

Rarely, loss-of-function mutations are dominant

- **Haploinsufficiency** – one wild-type allele does not provide enough of a gene product.



Heterozygotes for the null mutation of the T locus in mice have short tails.

Fig. 8.31 a

- ***Dominant-negative mutations*: Alleles that block the activity of wild-type alleles of the same gene, causing a loss of function even in heterozygotes.**

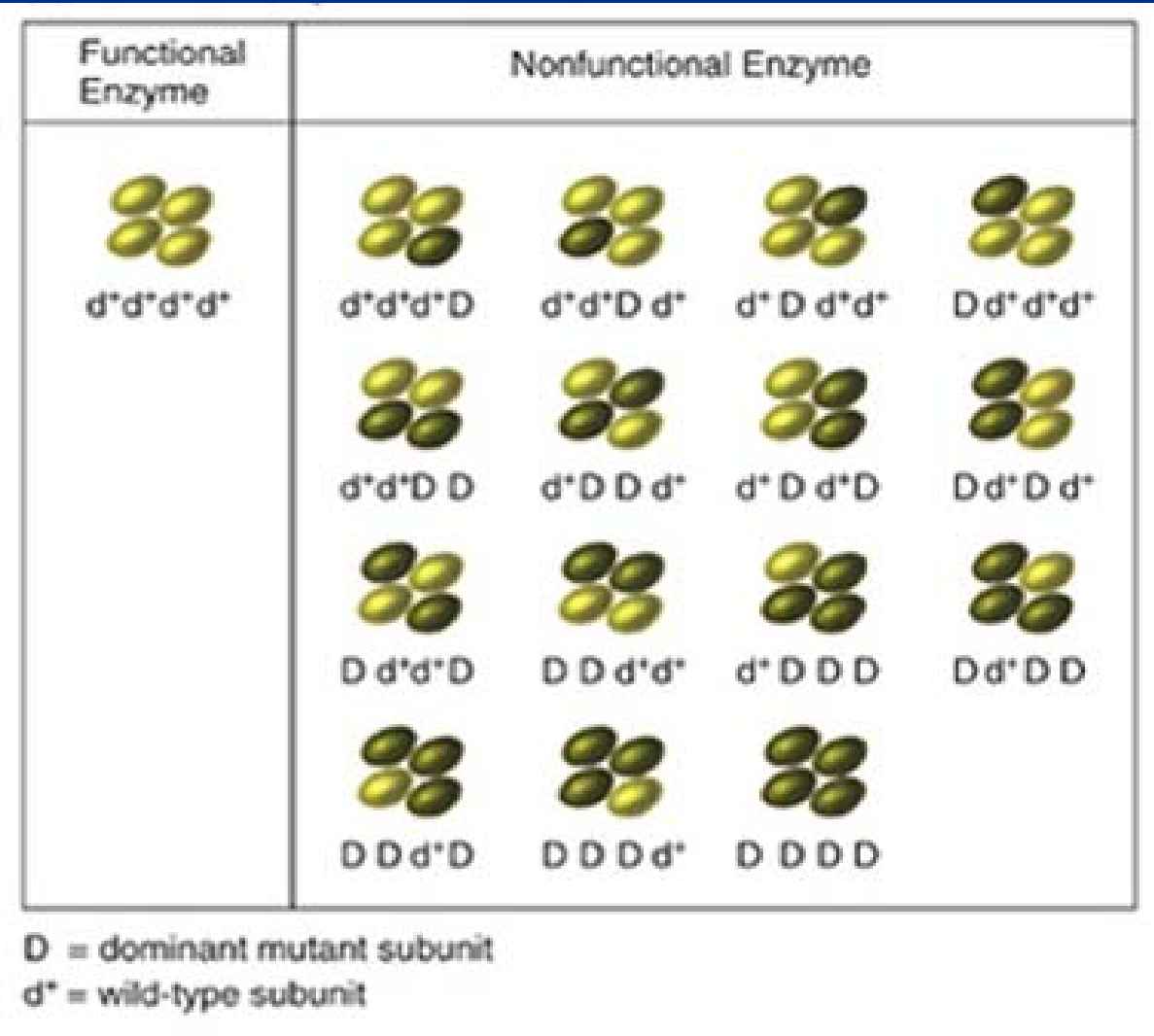
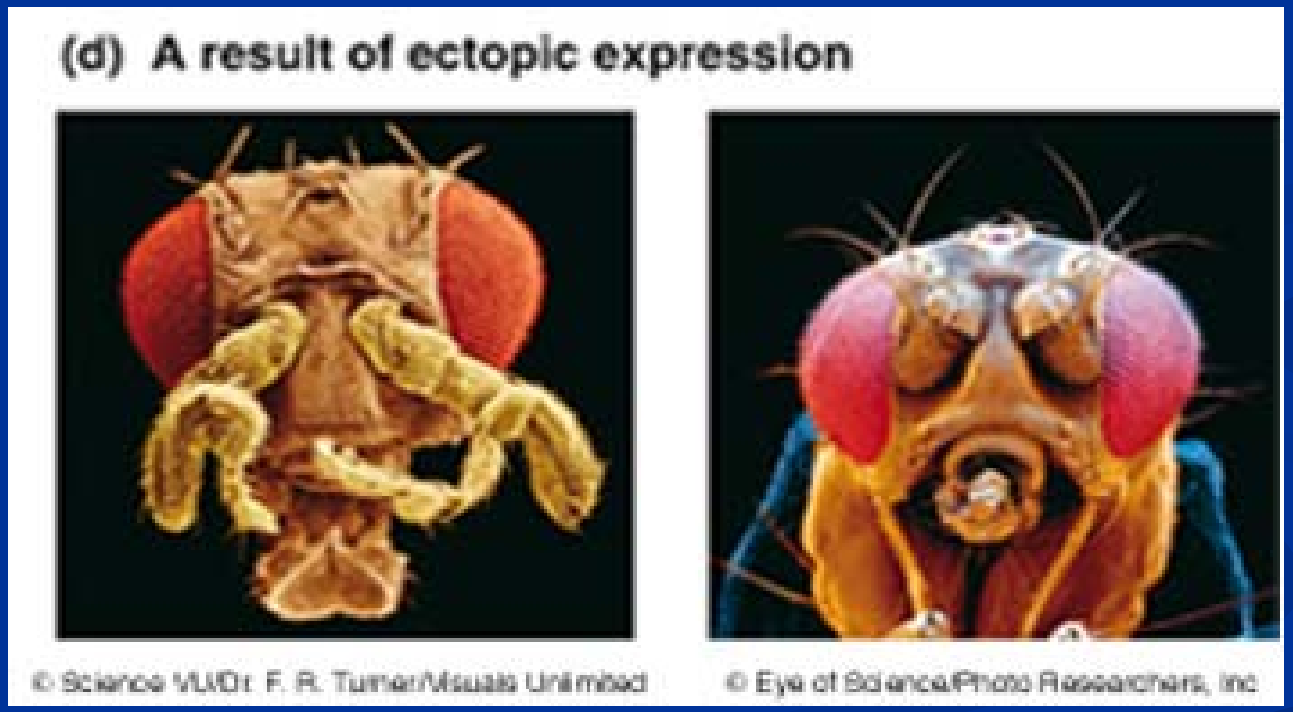


Fig. 8.31 b

Gain-of-function mutations are almost always dominant

- **Hypermorphic mutation**: Rare mutations that enhance a protein function or even confer a new activity on a protein.



***Antennapedia* is a neomorphic mutation**

Fig. 8.31 d

Mutations in genes encoding the molecules that implement gene expression may have global effects

- Usually lethal, i.e. ribosomal proteins.
- Mutations in tRNA genes can suppress mutations in protein-coding genes.
  - Nonsense suppressor tRNAs

Mutations in tRNA genes can suppress mutations in protein-coding genes

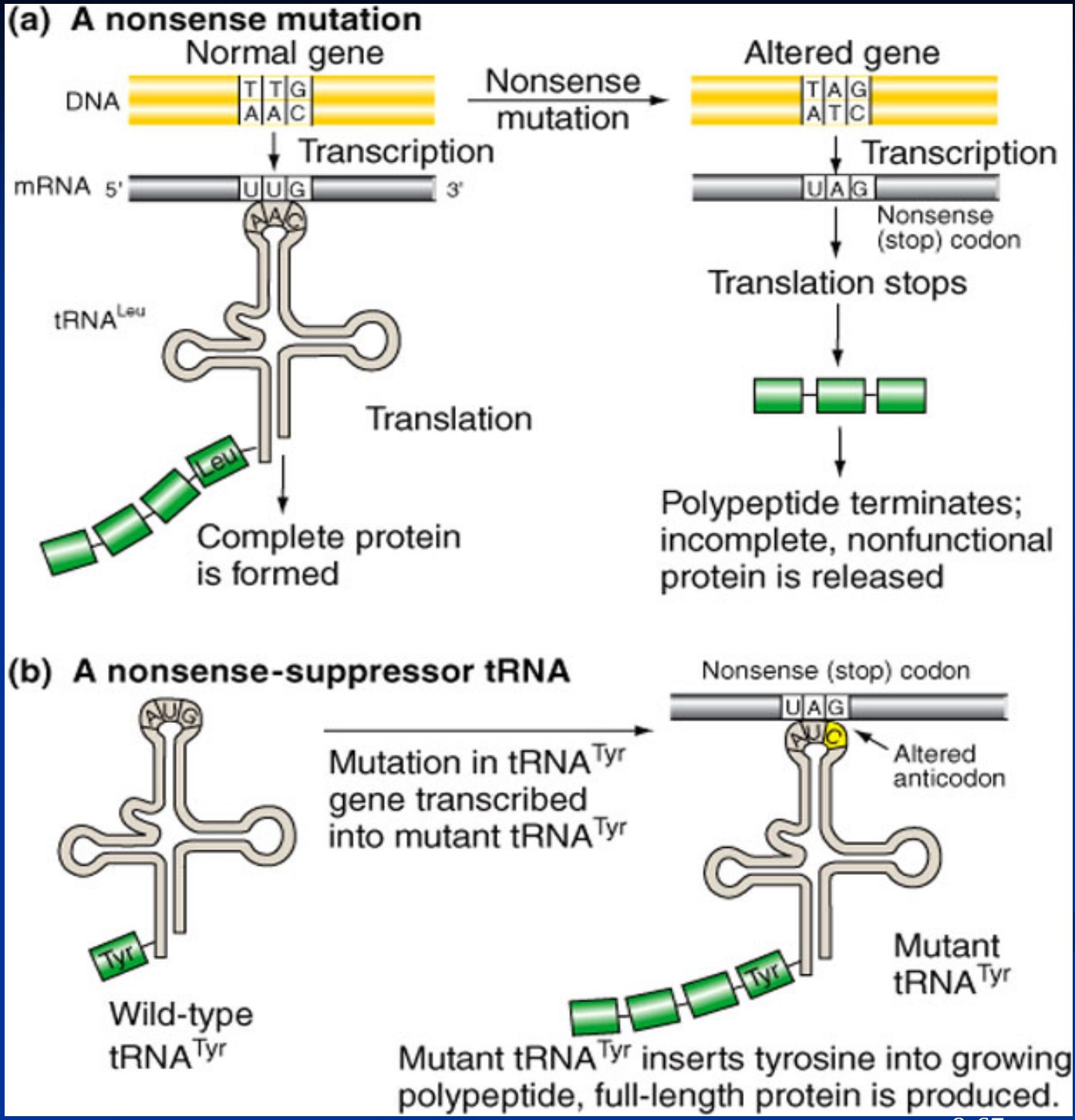


Fig. 8.32