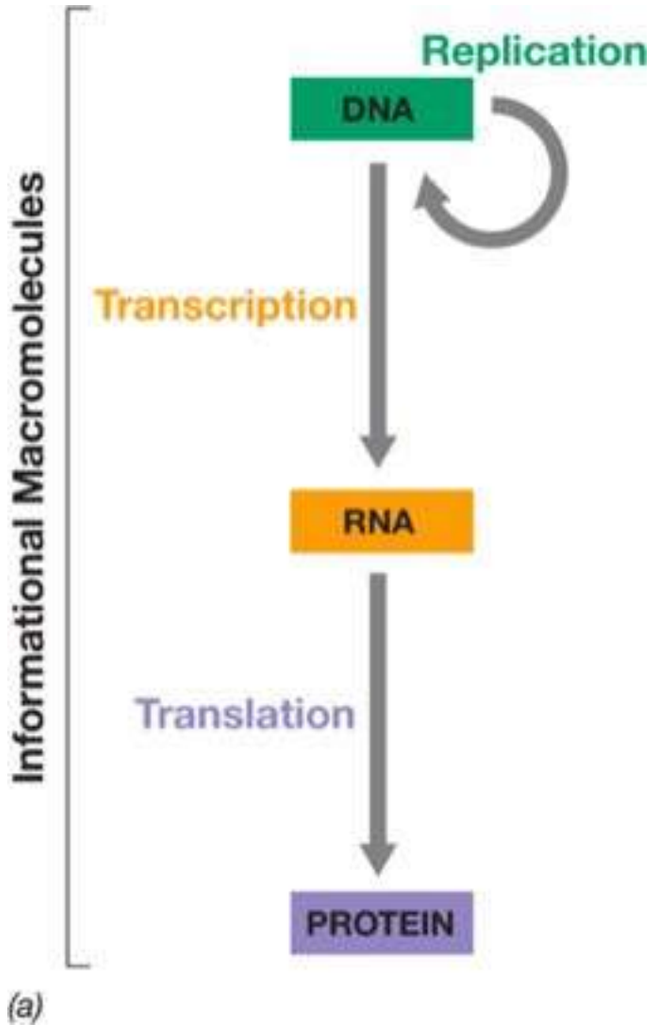
L05a

Recap

Biological Information Flow is highly coupled in Bacteria and Archaea



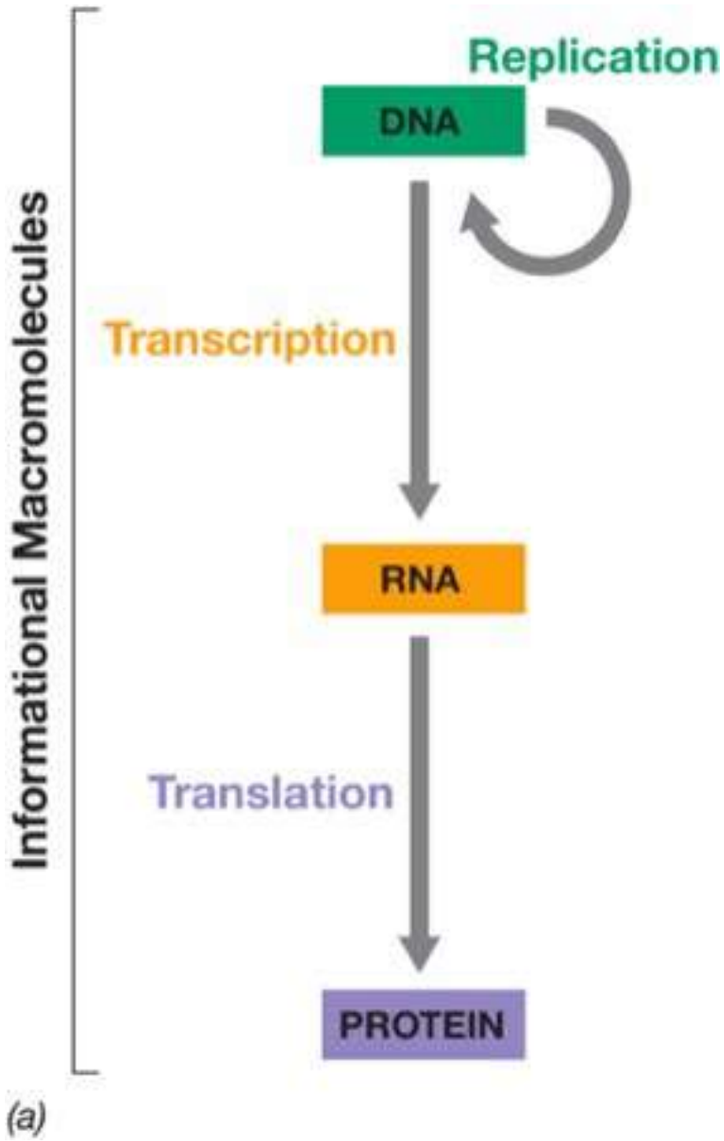
Madigan et al. 2020

The molecular processes of genetic information flow can be divided:

1. **Replication:** DNA double helix is duplicated. Replication is catalyzed by enzyme **DNA polymerase**
2. **Transcription:** Transfer of genetic information from DNA to RNA. Transcription is catalyzed by enzyme **RNA polymerase**
3. **Translation:** Formation of a polypeptide using genetic information transferred to mRNA by DNA. Process occurs on ribosome

- ***A ribosome initiates translation of an mRNA before RNA polymerase has finished synthesizing it***
- ***Rapidly growing cells to produce proteins at a maximal rate + rapid adaptation to changes in growth conditions by quickly expressing new protein sets required***

High diversity of RNAs in the cell



Madigan et al. 2020

……and only one DNA

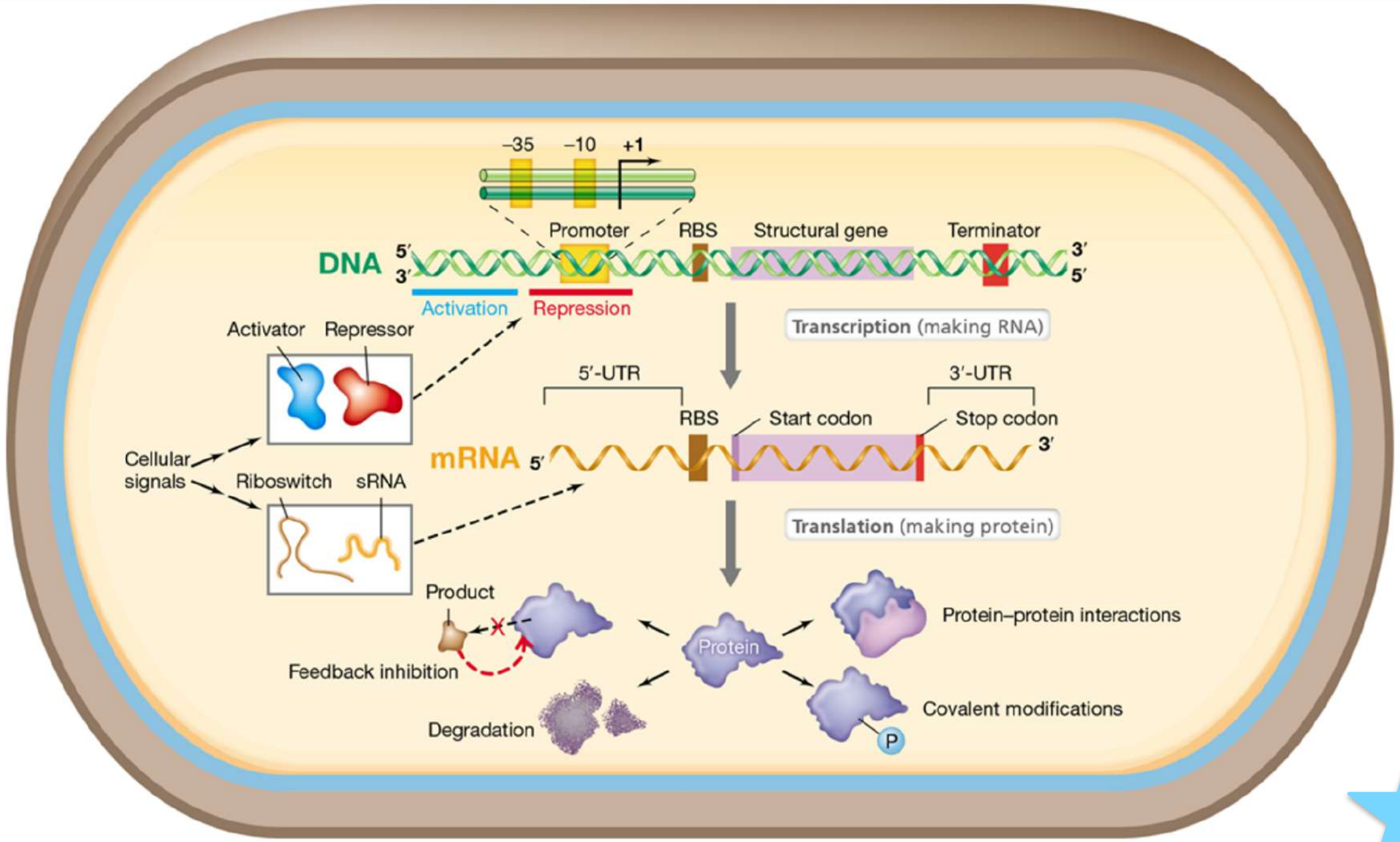
- **Messenger RNAs** (mRNAs) are single-stranded molecules that carry the genetic information from DNA to the ribosome

- **Transfer RNAs** (tRNAs) help convert the genetic information in the nucleotide sequences of RNA into a defined sequence of amino acids in proteins

- **Ribosomal RNAs** (rRNAs) are important catalytic and structural components of the ribosome

……and more small RNAs

Replication, Transcription, Translation and Regulation are example of molecule-molecule interaction



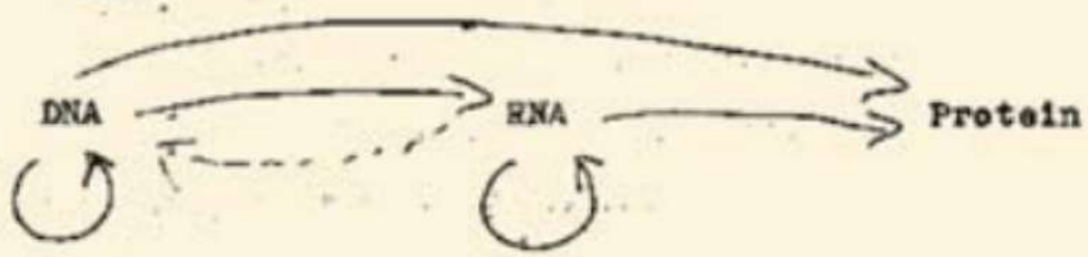
Madigan et al. 2020

property	*E. coli*	budding yeast	mammalian (HeLa line)
cell volume	0.3–3 μm^3	30–100 μm^3	1000–10,000 μm^3
proteins per μm^3 cell volume	2–4 × 10^6		
mRNA per cell	10^3–10^4	10^4–10^5	10^5–10^6
proteins per cell	~10^6	~10^8	~10^{10}
mean diameter of protein	4–5 nm		
genome size	4.6 Mbp	12 Mbp	3.2 Gbp
number protein coding genes	4300	6600	21,000
regulator binding site length	10–20 bp	5–10 bp	
promoter length	~100 bp	~1000 bp	~10^4–10^5 bp
gene length	~1000 bp	~1000 bp	~10^4–10^6 bp (with introns)
concentration of one protein per cell	~1 nM	~10 pM	~0.1–1 pM
diffusion time of protein across cell ($D \approx 10\ \mu m^2/s$)	~0.01 s	~0.2 s	~1–10 s
diffusion time of small molecule across cell ($D \approx 100\ \mu m^2/s$)	~0.001 s	~0.03 s	~0.1–1 s
time to transcribe a gene	<1 min (80 nts/s)	~1 min	~30 min (incl. mRNA processing)
time to translate a protein	<1 min (20 aa/s)	~1 min	~30 min (incl. mRNA export)
typical mRNA lifetime	3 min	30 min	10 h
typical protein lifetime	1 h	0.3–3 h	10–100 h
minimal doubling time	20 min	1 h	20 h
ribosomes/cell	~10^4	~10^5	~10^6
transitions between protein states (active/inactive)	1–100 µs		
time scale for equilibrium binding of small molecule to protein (diffusion limited)	1–1000 ms (1 µM–1 nM affinity)		
time scale of transcription factor binding to DNA site	~1 s		
mutation rate	10^{-8}–10^{-10}/bp/replication		

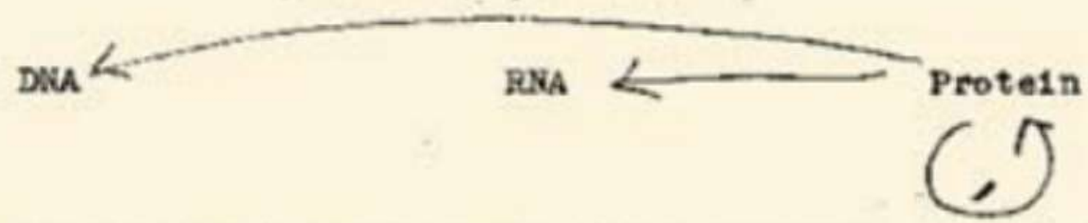
Ideas on Protein Synthesis (Oct. 1956)

The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it. That is, we may be able to have



but never

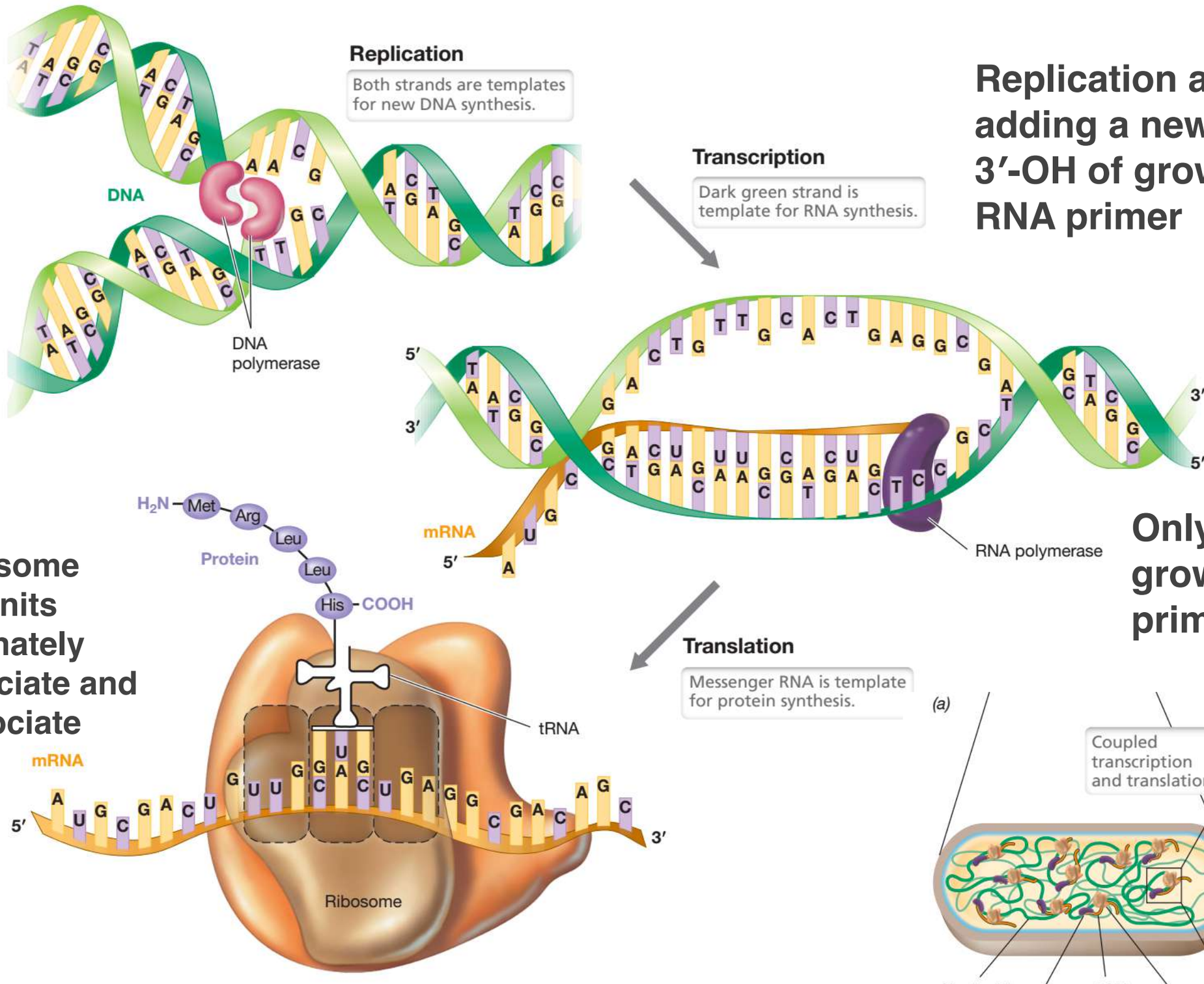


where the arrows show the transfer of information.

Notes of Francis Crick on the central dogma

Early draft for article published as: Crick, F.H.C. (1958): On Protein Synthesis. Symp. Soc. Exp. Biol. XII, 139-163. The 1958 paper did not include this visual depiction which later appeared in a 1970 Nature paper.

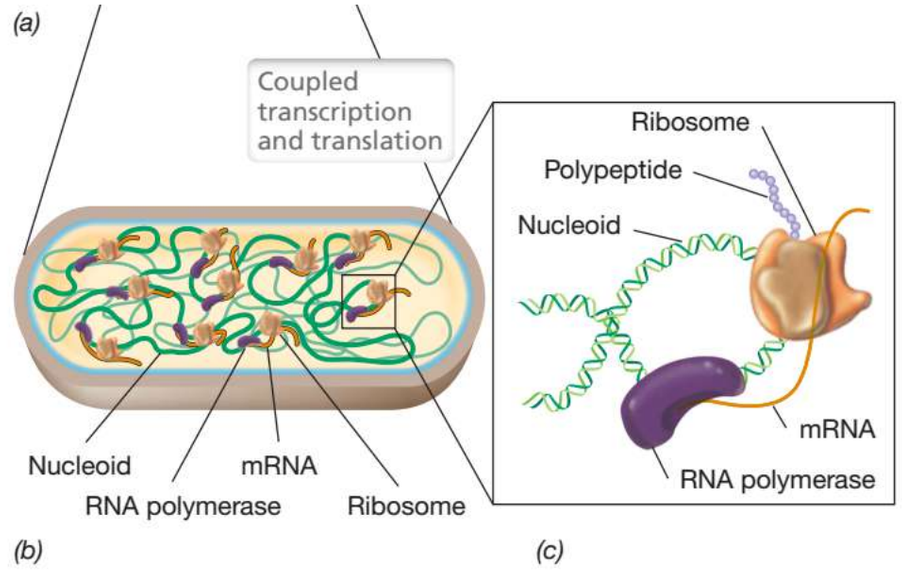
Replication, Transcription, Translation



Replication always 5′—> 3′, adding a new nucleotide to 3′-OH of growing chain, RNA primer

Only one chain grows: 5′—> 3′, no priming

Ribosome subunits alternately associate and dissociate



Ribosome-binding site at 5′ end of mRNA, RBS is complementary at 3′ end of the 16S rRNA part

Madigan et al. 2020

Illustration by David S. Goodsell, The Scripps Research Institute.

doi: 10.2210/rcsb_pdb/goodsell-gallery-001

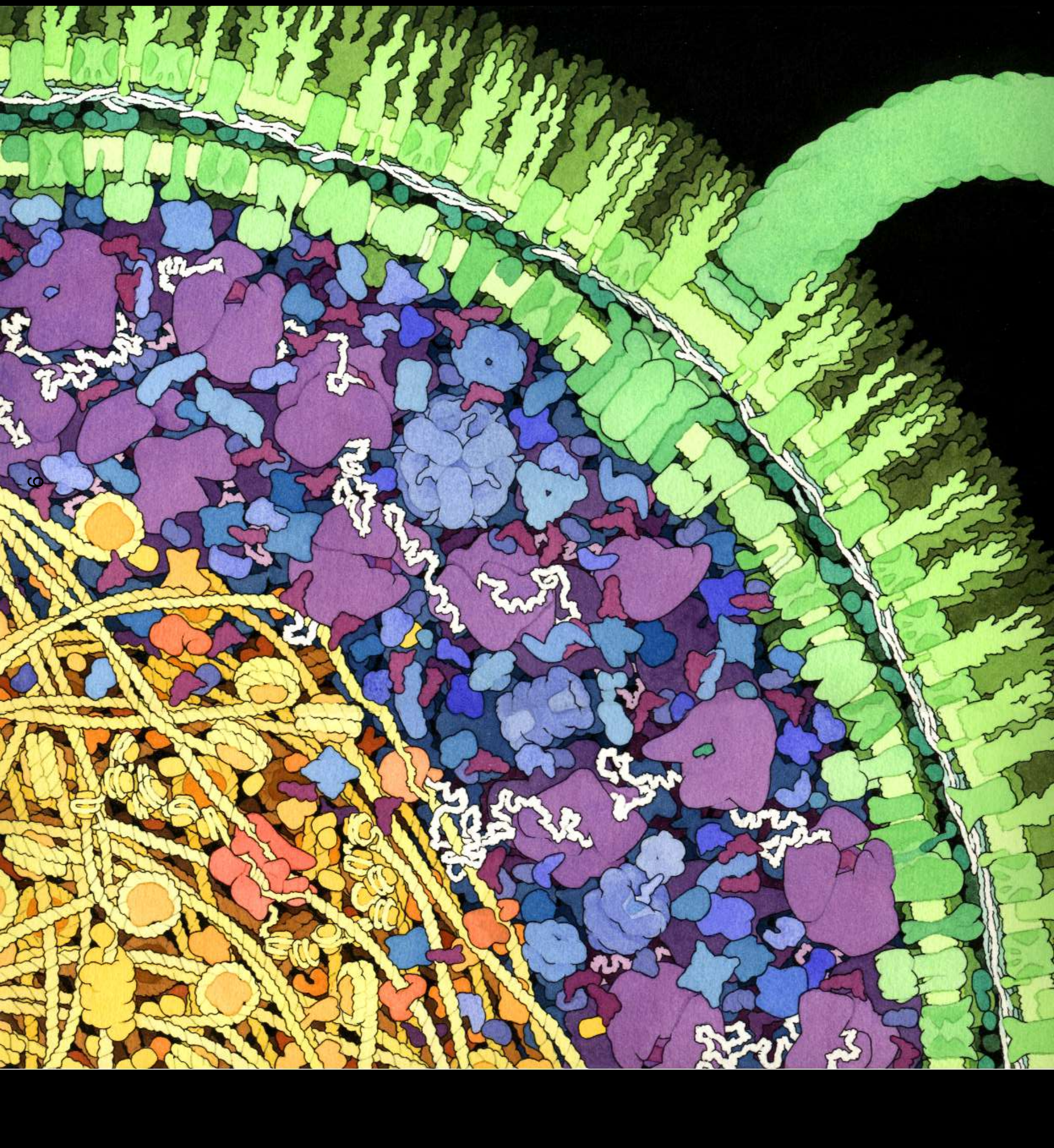


Figure Legend

Acknowledgement: Illustration by David S. Goodsell, The Scripps Research Institute. doi: 10.2210/rcsb_pdb/goodsell-gallery-001

This illustration shows a cross-section of a small portion of an Escherichia coli cell.

The cell wall, with two concentric membranes studded with transmembrane proteins, is shown in green.

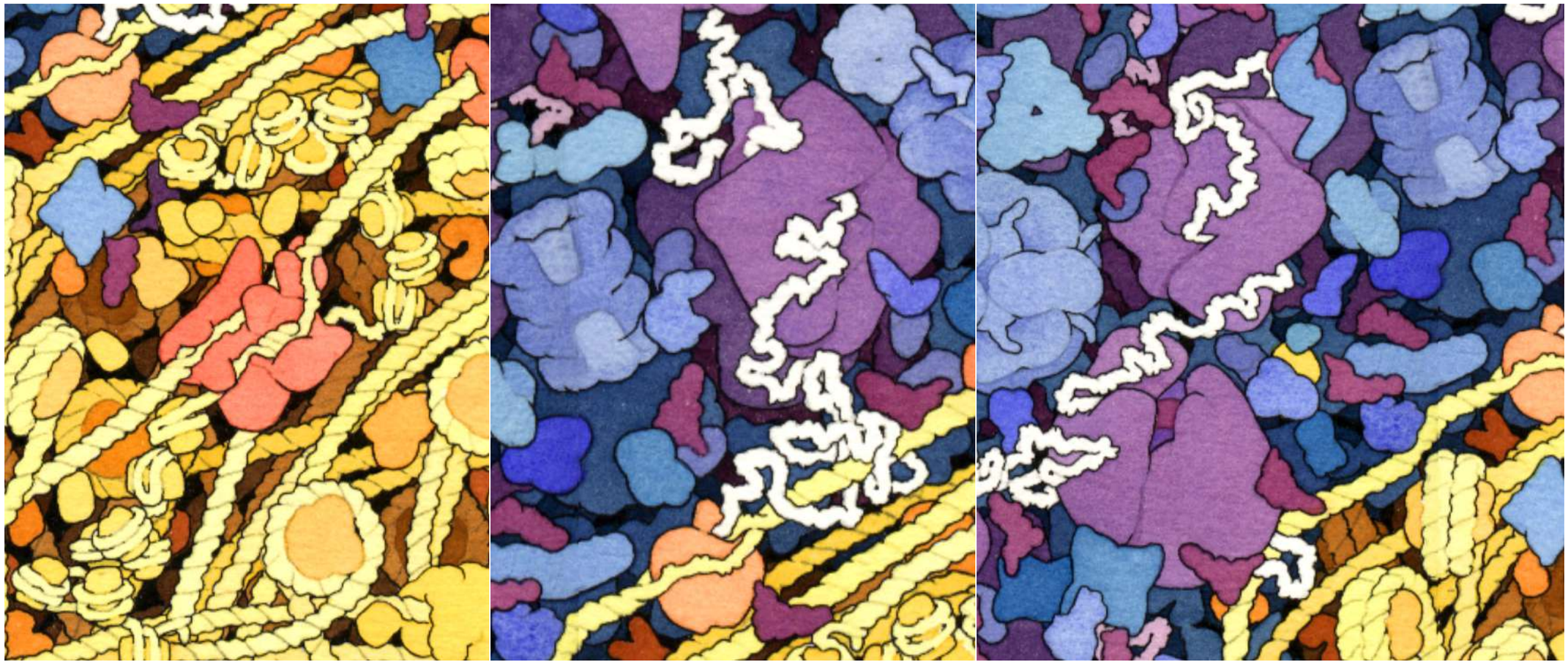
A large flagellar motor crosses the entire wall, turning the flagellum that extends upwards from the surface.

The cytoplasmic area is colored blue and purple. The large purple molecules are ribosomes and the small, L-shaped maroon molecules are tRNA, and the white strands are mRNA.
Enzymes are shown in blue.

The nucleoid region is shown in yellow and orange, with the long DNA circle shown in yellow, wrapped around HU protein (bacterial nucleosomes). In the center of the nucleoid region shown here, you might find a replication fork, with DNA polymerase (in red-orange) replicating new DNA.

Illustration by David S. Goodsell, The Scripps Research Institute.
doi: 10.2210/rcsb_pdb/goodsell-gallery-001

Escherichia coli, 1999



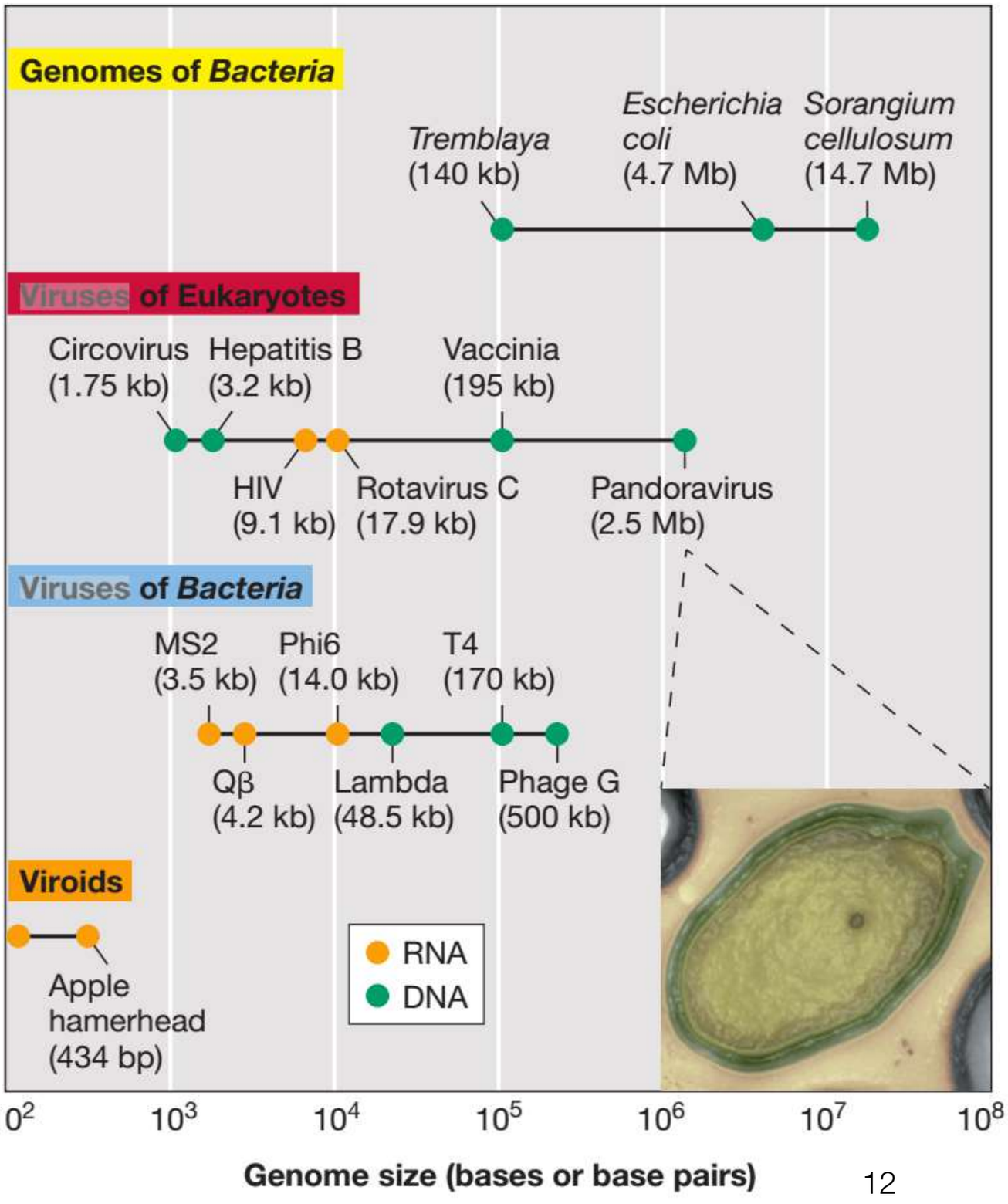
Long DNA circle shown in yellow, wrapped around HU protein (bacterial nucleosomes). In the center of the nucleoid region shown here, you might find a replication fork, with DNA polymerase (in red-orange) replicating new DNA.

You might find RNA polymerase (in pale orange) synthesizing mRNA (in white).

The large purple molecules are ribosomes and the small, L-shaped maroon molecules are tRNA, and the white strands are mRNA.

Illustration by David S. Goodsell, The Scripps Research Institute
doi: 10.2210/rcsb_pdb/goodsell-gallery-001

Genome size range



1 kb = 10^3 bp (base pairs)
1 Mb = 10^6 bp
1 Gb = 10^9 bp

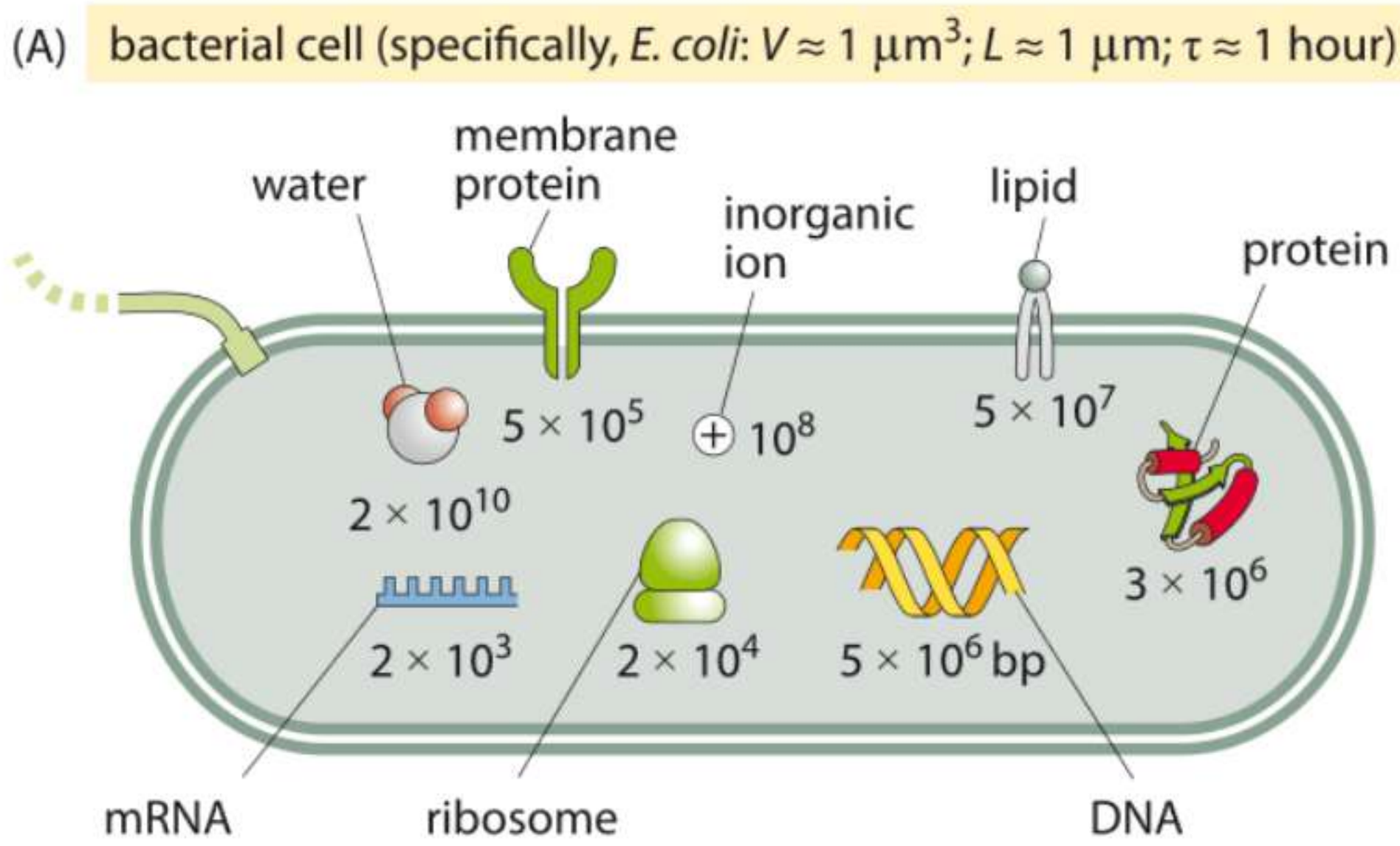
1 kb = 10^{-6} pg
1 Mb = 10^{-3} pg
1 Gb = 1pg

Doležel et al., 2003
Base pair # = mass in pg x 0.978 10^9

1 kb = 0.33 µm
1 Mb = 0.33 mm
1 Gb = 0.33 m

Dickerson et al., 1982
1 bp = 0.33 nm

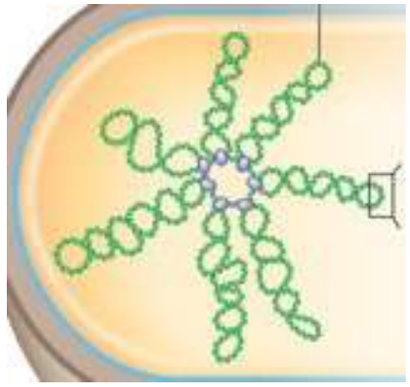
A number game!



~ 1.6 mm long is the DNA (base pair and genome size) —-> need to be twisted, tight packaged and supercoiled

DNA Replication

Inserting supercoils —> topoisomerase to prevent break and stabilization



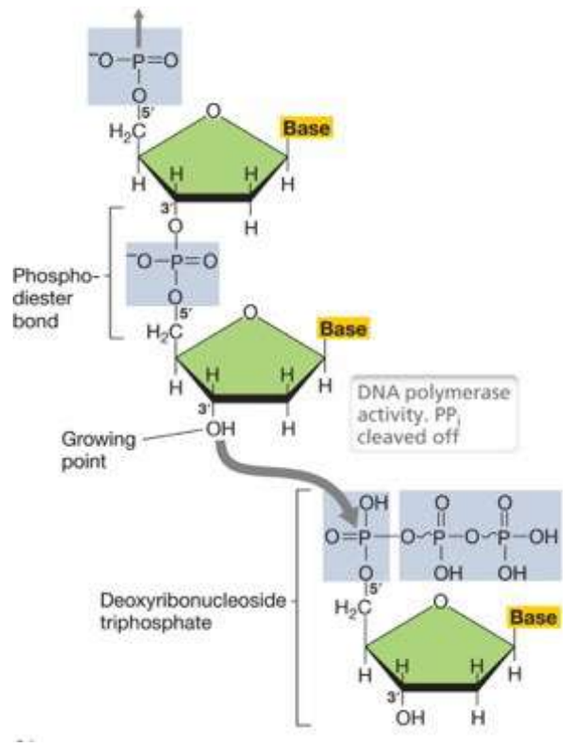
Madigan et al. 2020

Removing supercoils —> topoisomerase to access information for replication and transcription

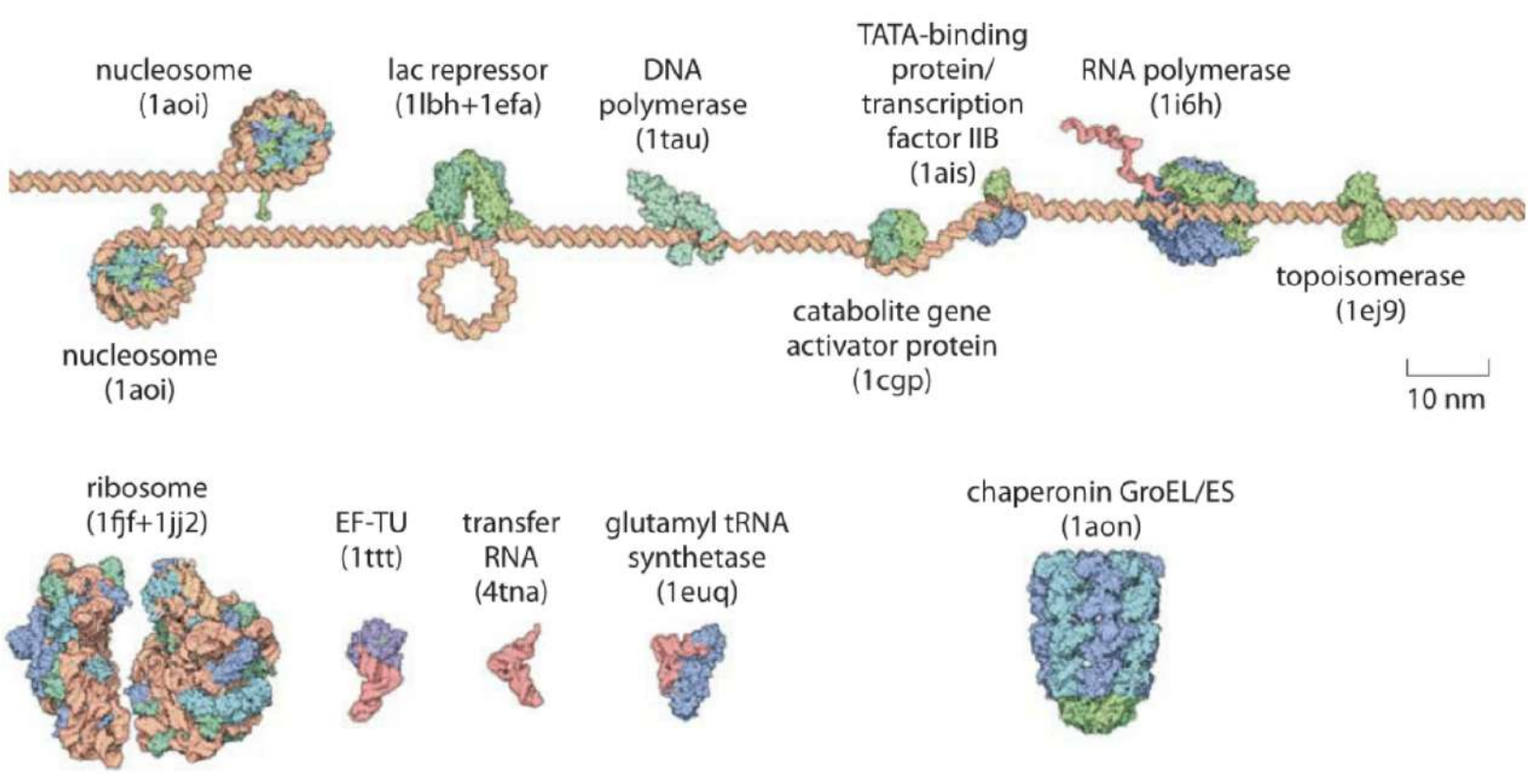
- Inserting supercoils into DNA **requires energy from ATP,** whereas releasing supercoils does not

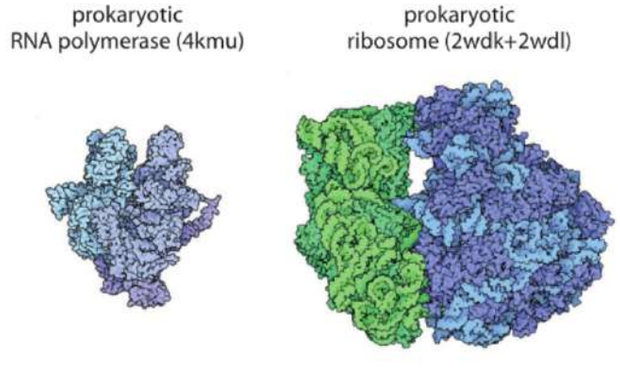
Duplicating DNA —> DNA polymerase III (Replication always 5′ —> 3′)

Strand stabilization —> gyrase and proteins
Cutting and sawing the strands
Repairing mistake and breaks



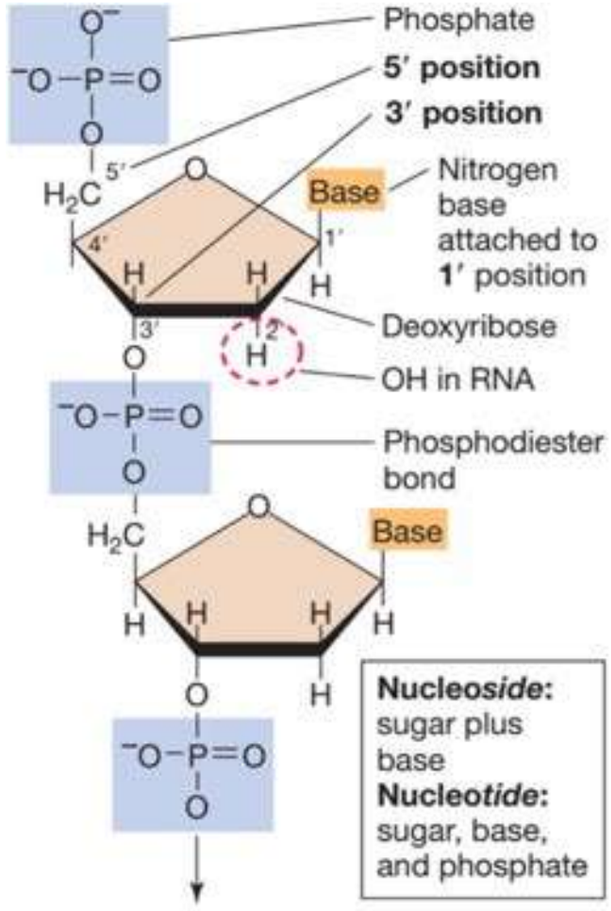
Structures of the machines of the central dogma

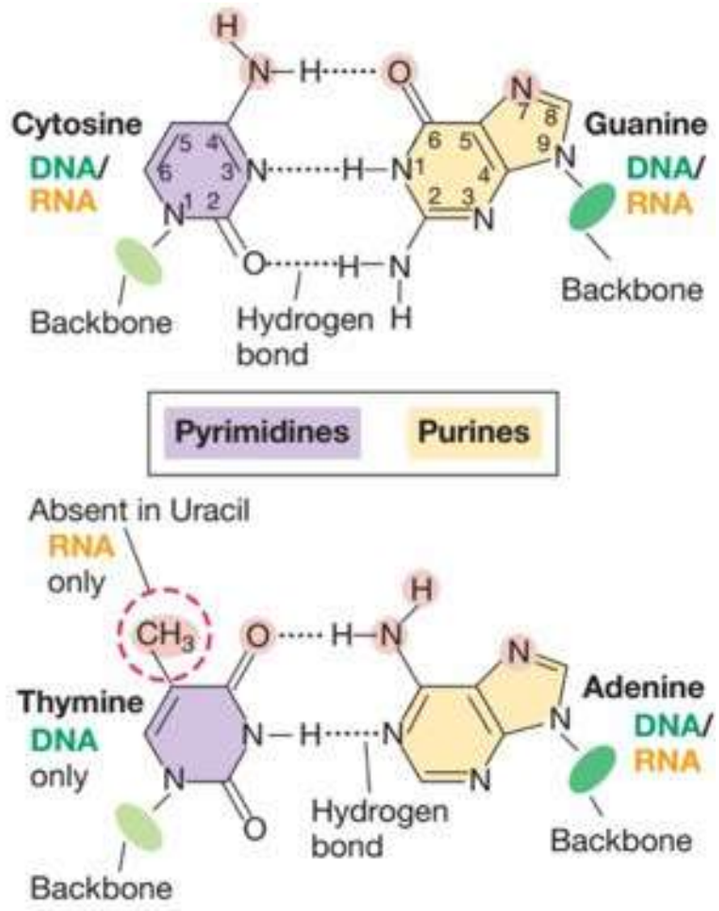




- All shown drawn to scale relative to the DNA substrate
- PDB database notations
- Replication, transcription and translation

Background: DNA structure, I

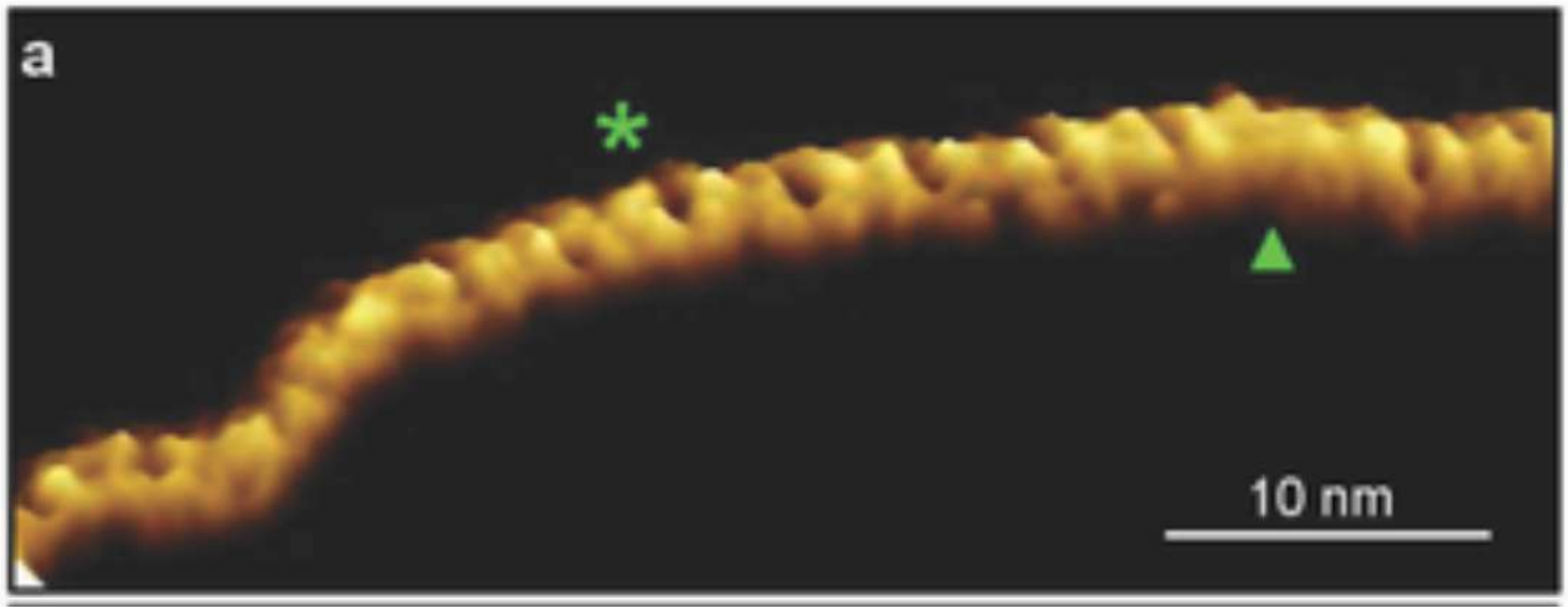




Madigan et al. 2020

- **The sequence of nucleotides** in a DNA or RNA molecule **is its primary structure and encodes the genetic information**
- In cells, DNA is double-stranded kept in place by **H-bonds between bases**
- **Specific base pairing: A - T & G - C**
- **2 strands are complementary in base sequence** —> faithful replication of the molecule
- **2 strands are in an antiparallel fashion** (5′ —> 3′ and 3′ —> 5′)

Background:DNA structure, II

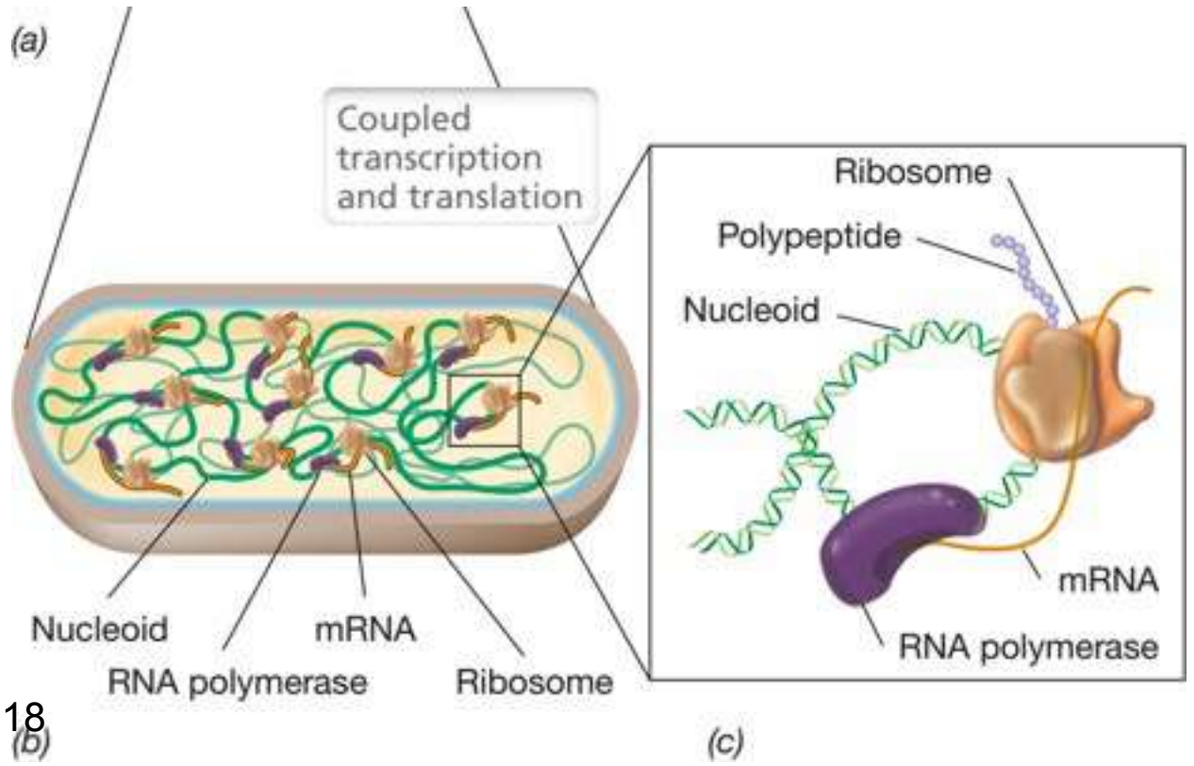


Pyne et al., 2014

- Complementary and antiparallel strands are wrapped around each other: **double helix**
- The helix naturally forms two distinct grooves, **the major groove** (green asterisks) **and the minor groove** (green triangle)
- **Most proteins** that **interact** specifically with DNA bind in **the major groove**
- **Double helix is a regular structure, some atoms of each base are always exposed in the major groove (and some in the minor groove) —> important in interactions with proteins**

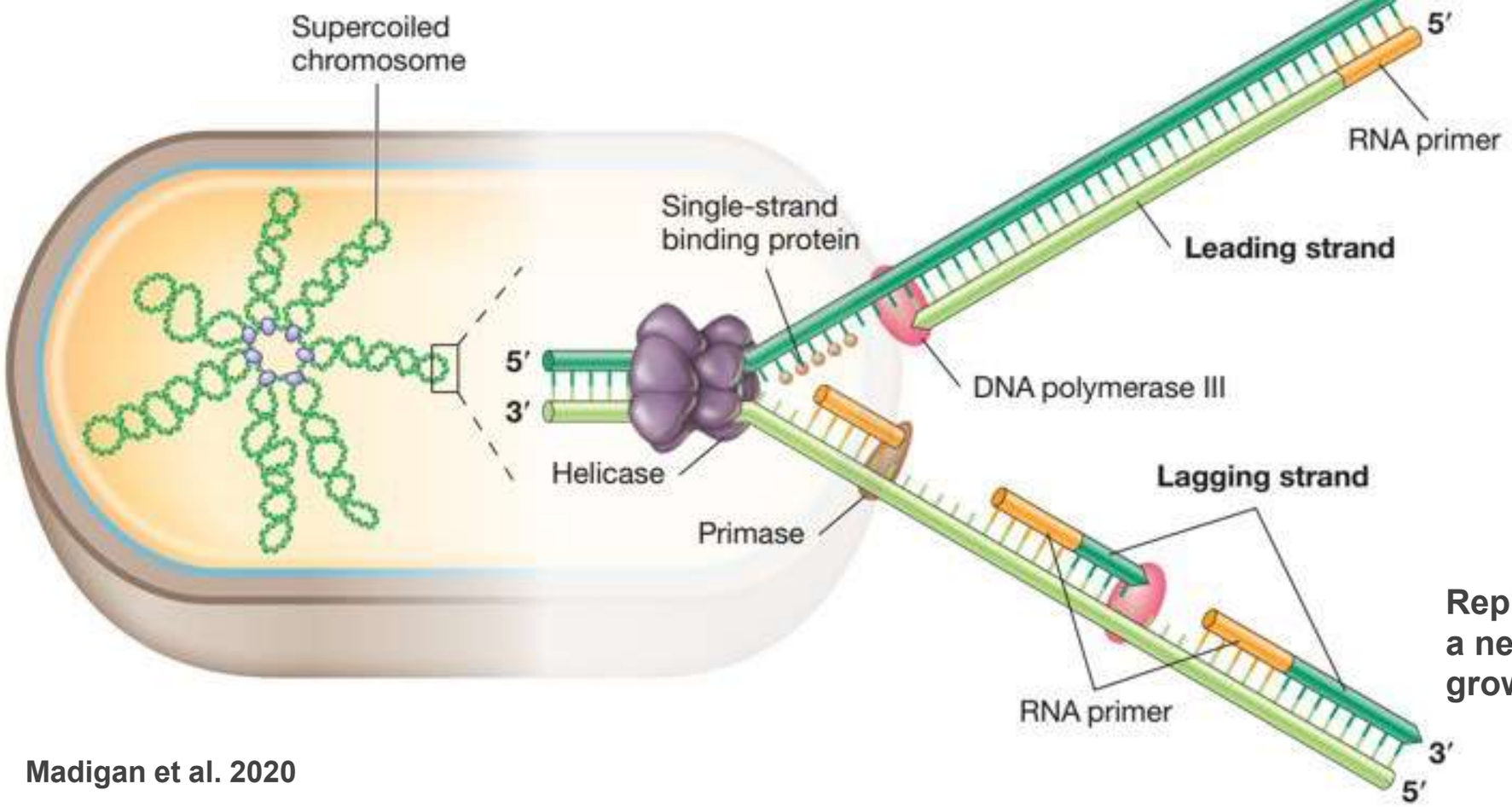
Genes and Operons in *E.coli*

- **Genes encoding enzymes that function in steps of the same biochemical pathway** are sometimes **clustered groups** is called an **operon**
- An operon is **transcribed** to form **a single mRNA** that encodes several different proteins and **is regulated as a unit**
- Genome analysis has revealed 4288 possible **protein-encoding genes** that account for **88%** of the ***E. coli* genome**
- Approximately **1%** of genome encodes **tRNAs and rRNAs**, and the remaining genes consist of regulatory sequences that may or may not be transcribed (but are not translated) or have other functions

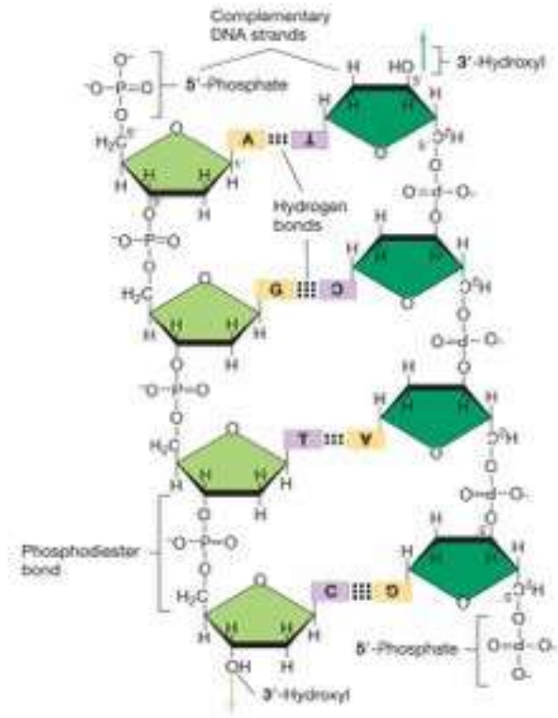


Madigan et al. 2020

DNA replication, I



Madigan et al. 2020

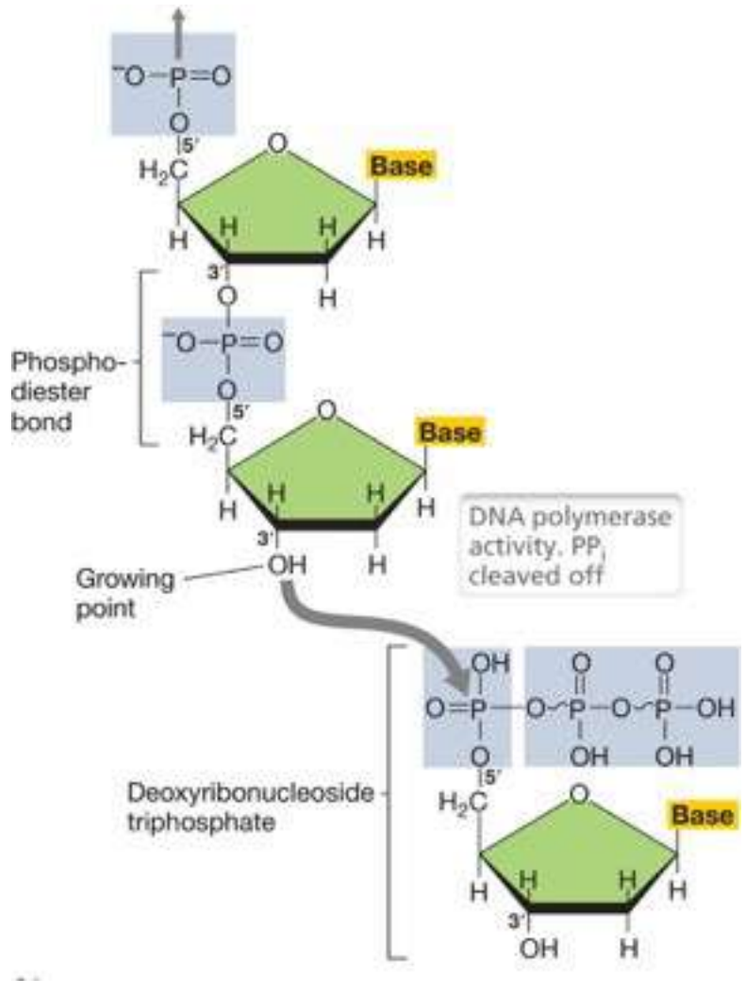


Replication always 5′ to 3′, adding a new nucleotide to the 3′-OH of the growing chain

- DNA synthesis begins at a single site on chromosome, **origin of replication (oriC), where DnaA binds and opens up the double helix**
- **Bi-directional replication, two forks, opposite directions**
- **Stabilization** of strands by helicase (DnaB), and its helper loader protein (DnaC)
  Two helicases are loaded, one onto each strand, facing in opposite directions
- **Two primase and two DNA polymerase enzymes** are loaded onto the DNA behind the helicases and initiation of DNA replication begins
- As replication proceeds, replication fork appears to move along DNA

DNA replication, II

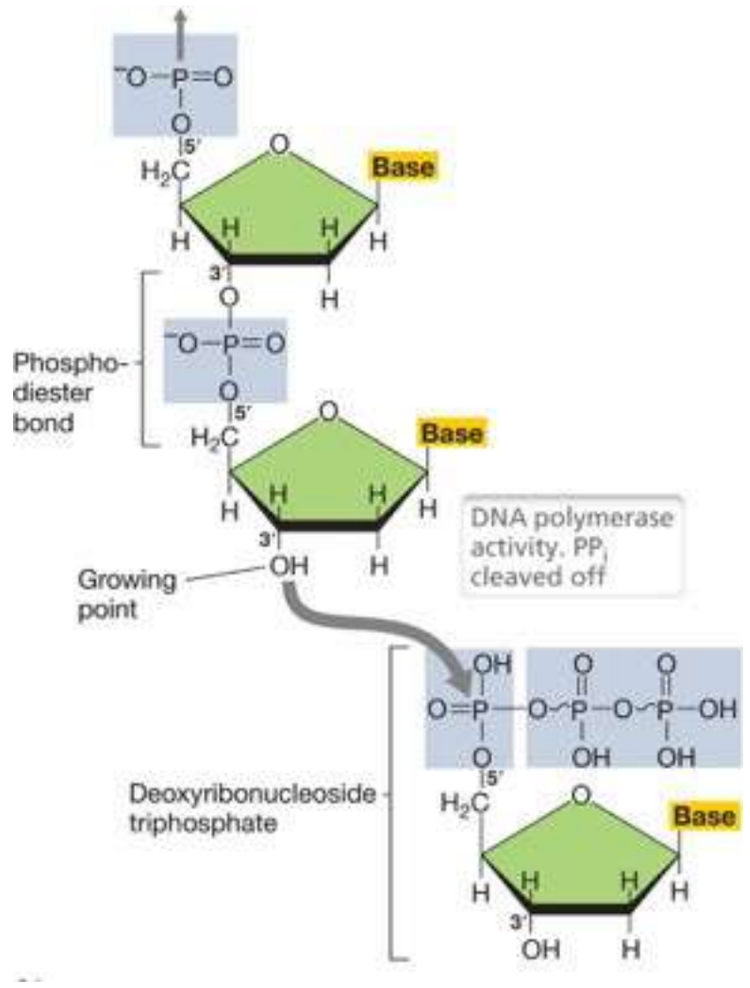
- DNA replication **semiconservative**
- DNA replication **always** proceeds from the **5′ end to the 3′** end, the 5′-phosphate of the incoming nucleotide being attached to the 3′-hydroxyl of the previously added nucleotide
- Enzymes that catalyze the polymerization of deoxynucleotides: **DNA polymerases**, 5 in *E.coli*
- **DNA Pol III** is the primary enzyme for replicating DNA (DNA Pol I plays a lesser role)



Madigan et al. 2020

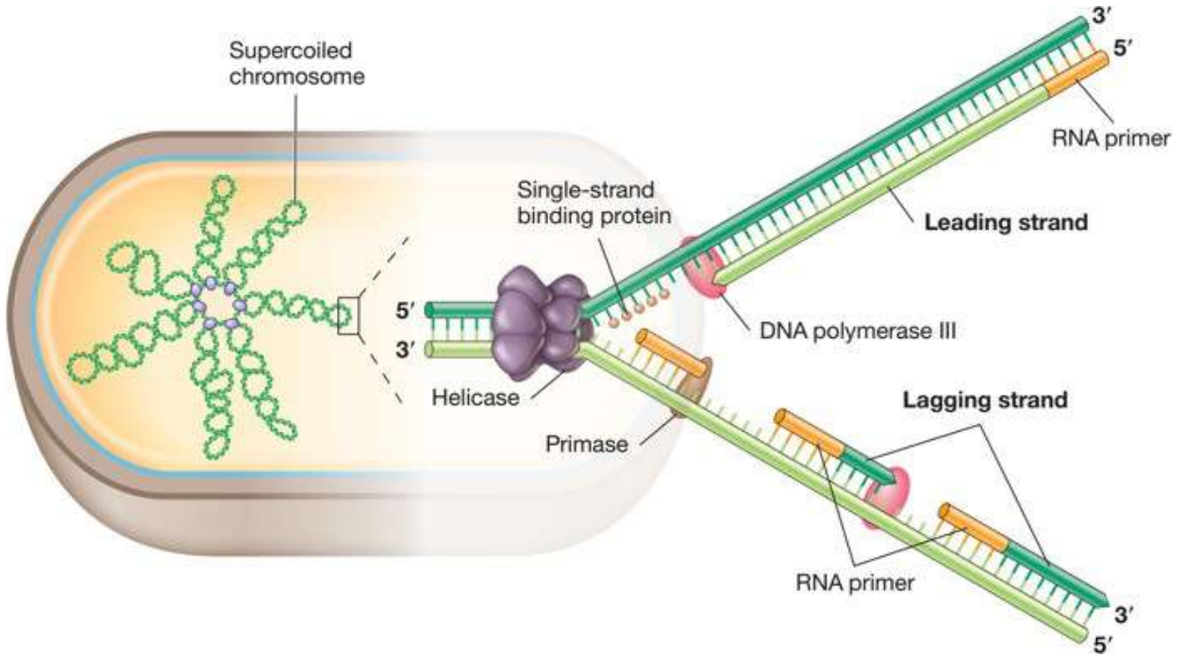
DNA replication, II

- Other DNA polymerases function to repair damaged DNA
- **All DNA polymerases synthesize DNA in the 5′—> 3′ direction**
- DNA polymerases need a preexisting 3′-OH group —> primer is a short stretch of RNA
- When the DNA helix is first opened, **primase makes this RNA primer, synthesizing a short stretch** (11–12 nucleotides) of RNA complementary in base pairing to the template strand DNA
- At the growing end of this RNA primer is a 3′-OH group to which DNA polymerase adds the first deoxyribonucleotide
- RNA primer is removed and replaced with DNA



Madigan et al. 2020

DNA replication, III



Madigan et al. 2018

Only one RNA primer
Continuous synthesis by DNA Pol III
Rate: 1000 nucleotides per second

Multiple RNA primers
Discontinuous synthesis by DNA Pol III

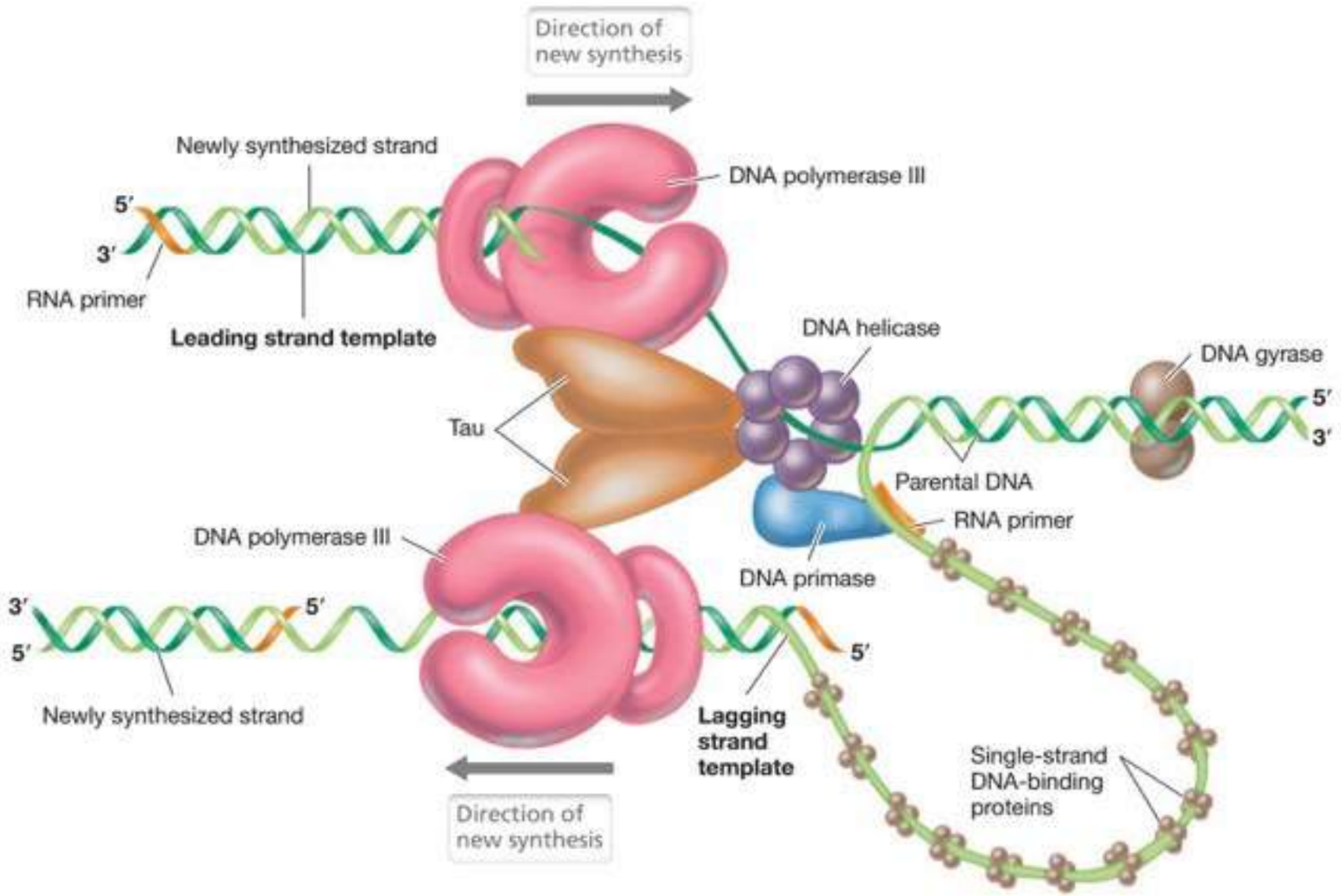
- Lagging strand forms from several short DNA fragments that are combined later to yield a continuous strand of DNA
- After synthesizing the RNA primer, primase is replaced by DNA Pol III
- DNA Pol III is held on the DNA by a "sliding clamp," which encircles and slides along the single template strands of DNA
- Replication fork contains two polymerase core enzymes and two sliding clamps, one set for each strand
- After assembly on the lagging strand, the elongation activity of DNA Pol III adds deoxyribonucleotides sequentially until it reaches previously synthesized DNA at this point, activity of DNA Pol III stops
- To complete DNA synthesis, **DNA Pol I catalyzes two different reactions: besides synthesizing DNA, Pol I has a 5′ —> 3′ exonuclease activity that removes the RNA primer**
- When the primer has been excised and replaced with DNA, Pol I is released, **the very last phosphodiester bond in replicating DNA is made by DNA ligase**
- **DNA ligase seals nicks in DNA**s that have an adjacent 5′-PO_4^{2-} and 3′-OH (DNA Pol I and Pol III are unable to do), and along with DNA Pol I, it also **participates in DNA repair**

DNA replication, IV

Replication proteins aggregate to form a large replication complex: **replisome** (facilitates the sequential activities)

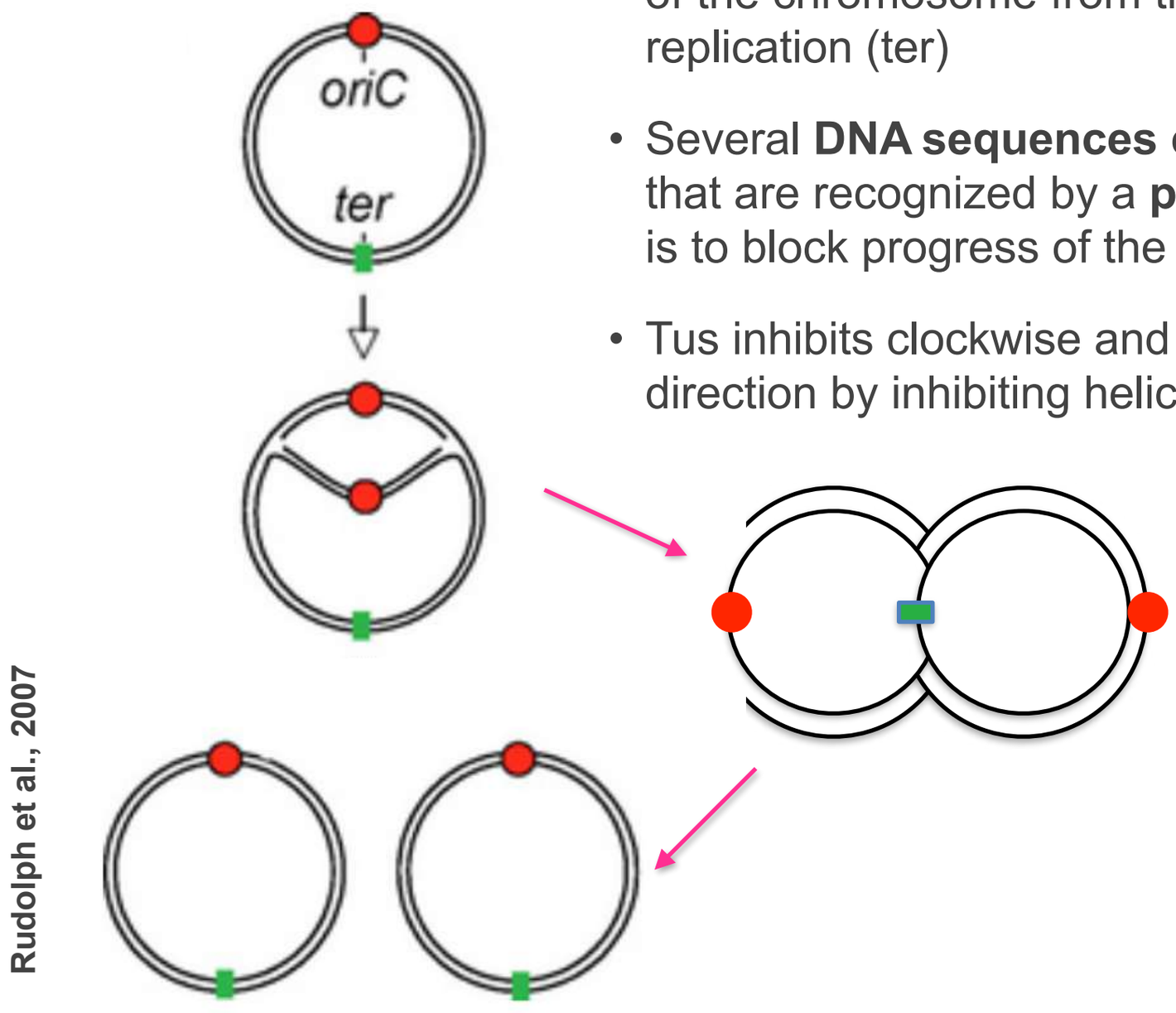
The lagging strand of DNA actually loops out to allow the replisome to move smoothly along both strands, the complex literally pulling the DNA template through it as replication proceeds

In addition to the replisome, **helicase and primase** form their own subcomplex within the replisome called the **primosome**



Madigan et al. 2018

DNA replication, V



Rudolph et al., 2007

- Replication forks collide at a site located on the opposite side of the chromosome from the origin called **terminus** of replication (ter)
- Several **DNA sequences called Ter sites (2 groups of 3)** that are recognized by a **protein** called **Tus**, whose function is to block progress of the replication forks
- Tus inhibits clockwise and counterclockwise replication direction by inhibiting helicase
- Topoisomerase canalizes separation of catenated daughter DNA circles

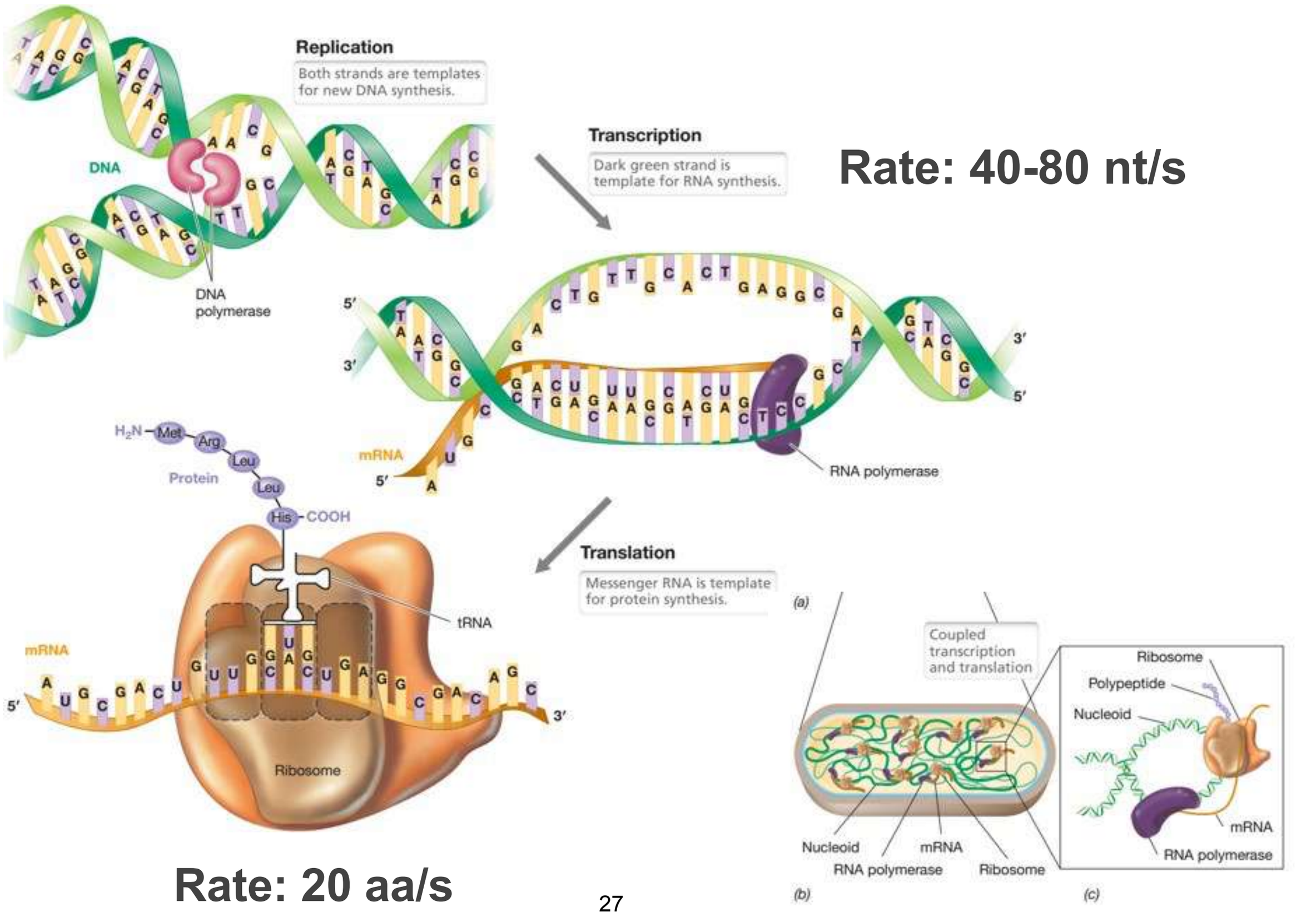
Proofreading

- **Mutation rates in cells are extremely low, between 10^{-8} and 10^{-11} errors per base pair inserted**
- **DNA Pol III inserts bases according to the base-pairing rules**
- **Proofreading**: both **DNA Pol I and Pol III possess a 3′—> 5′ exonuclease activity** that can remove such mismatched nucleotides
- The polymerase detects the **error** because incorrect base pairing causes a **slight distortion** in the **topology** of the double helix
- After the removal of a mismatched nucleotide, the polymerase then gets a second chance to insert the correct nucleotide
- **Exonuclease proofreading occurs in Bacteria and Archaea (Euk)**

organism	errors per base or codon	BNID and measurement methods
		transcription
E. coli	10^{-4}	111146, transition mutations based on sequencing at very high (10^6) coverage (2013)
E. coli	10^{-5}	105212, *In vitro* selection for rifampicin resistance and increased leakiness of an early, strongly polar nonsense mutation of lacZ (1983, 1986)
E. coli	10^{-4}	103453, activity in strains carrying lacZ mutations (1981)
S. cerevisiae	2×10^{-6}	110019, RNA pol II, determined *in vitro* (2008)
S. cerevisiae	2×10^{-4}	105213, RNA pol III, determined based on selectivity (2007)
C. elegans	4×10^{-6}	111144, determined using bar coded sequencing (2013)
		translation
E. coli	3×10^{-4}	105069, Lys-tRNA, reporter system for frequency of each type of misreading error (2007)
E. coli	$1–4\times10^{-3}$	105215, identify cases that do not contain the amino acid cysteine responsible for the missense substitution (1983)
E. coli	$10^{-4}–10^{-3}$	103454, identify cases that do not contain the amino acid cysteine responsible for the missense substitution (1977, 1983)
B. subtilis	4×10^{-3}	105466, GFP with nonsense mutation, also find 2.4% for frame-shift (!) (2010)
S. cerevisiae	$0.5–2\times10^{-5}$	105216, measurement of rescue rate of inactivating mutations of type III chloramphenicol acetyl transferase (1998)

- ***E. coli* mutation rate on the order of 10^{-10} mutations/bp/replication**

Transcription & Translation



Rate: 40-80 nt/s

Rate: 20 aa/s

Madigan et al. 2018

RNA synthesis: Transcription, I

- **Transcription—RNA synthesis off of a DNA template**
- 3 typeRNA: messenger (mRNA), transfer (tRNA), ribosomal (rRNA) and several other minor classes of RNA exist but most of these function in regulation
- Single strand RNA is both a genetic and a functional molecule
- **mRNA encodes genetic information from the genome and carries it to the ribosome**
- **rRNAs play both a structural and a functional role in ribosomes**
- **tRNAs function as the carriers of aa the ribosome for protein synthesis**

RNA synthesis: Transcription, II

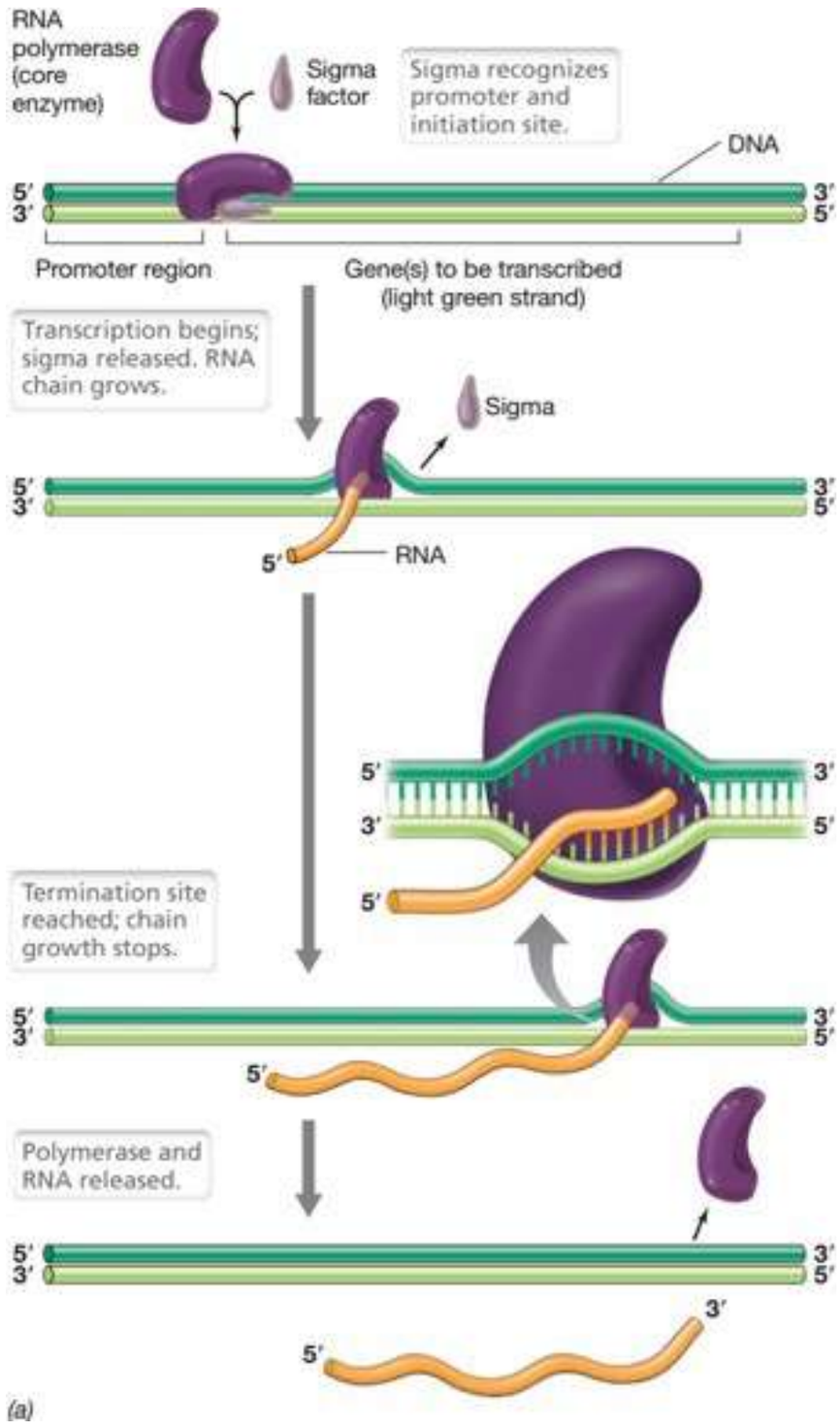
- RNA contains ribose and uracil (instead thymine; U=A).
- RNA **primary** structure (sequence of nucleotides) allows them to fold and exploit **complementary base pairing**
- RNA **secondary** structure refers to this **folding**, and RNA **functional role** plays in cell may depend critically on its secondary structure:

1. **mRNAs** are unfolded, exist for only **a few minutes** before enzymes called **ribonucleases** degrade them—rapid turnover adapt to changing environmental conditions and halt the translation of mRNAs whose products are no longer needed

2. **rRNAs and tRNAs** (referred to as stable RNAs) are **long-lived** because **their secondary structures** prevent ribonuclease attack

RNA synthesis: Transcription, III

- Transcription is catalyzed by the enzyme **RNA polymerase**
- RNA polymerase forms **phosphodiester bonds but between the ribonucleotides**
- **Polymerization** is driven by energy released from the **hydrolysis of 2 energy-rich phosphate bonds** of the incoming ribonucleoside triphosphates
- During elongation of an RNA chain, ribonucleoside triphosphates are added to the 3′-OH of the ribose of the preceding nucleotide
- **Only one chain growth is 5′—> 3′**
- **No priming**
- RNA polymerase must first **recognize** initiation sites on the DNA: **promoters**



Madigan et al. 2018

RNA synthesis: Transcription, IV

- RNA polymerase, 5 subunits: **β, β′, α (2 copies), ω, σ**
- Subunits form an enzyme complex called the **RNA polymerase holoenzyme**
- **Sigma, σ**, is not as tightly bound as the other subunits and easily **dissociates** to yield the RNA polymerase core enzyme, α2ββ′ω
- The core enzyme alone synthesizes RNA

TABLE 4.3 Sigma factors in *Escherichia coli*

Name[a]	Upstream recognition sequence[b]	Function
σ^{70} RpoD	TTGACA	For most genes, major sigma factor for normal growth
σ^{54} RpoN	TTGGCACA	Nitrogen assimilation
σ^{38} RpoS	CCGGCG	Stationary phase, plus oxidative and osmotic stress
σ^{32} RpoH	TNTCNCCTTGAA	Heat shock response
σ^{28} FliA	TAAA	For genes involved in flagella synthesis
σ^{24} RpoE	GAACTT	Response to misfolded proteins in periplasm
σ^{19} FecI	AAGGAAAAT	For certain genes in iron transport

Gene regulation one the presence or absence of σ

Madigan et al. 2018

RNA synthesis: Transcription, IV

- **σ recognizes the appropriate site on the DNA for transcription** to begin (σ dissociates from holoenzyme once a short sequence of RNA has been formed)
- Several **σ, most used σ^{70}**
- Several **promoters w. 2 highly conserved regions**
- Upstream the transcription start site:
  - **A. 10 bases upstream, the -10 region, or Pribnow box; consensus sequence of TATAAT**
  - **B. 35 bases upstream consensus sequence is TTGACA, -35 region**

TABLE 4.3 Sigma factors in *Escherichia coli*

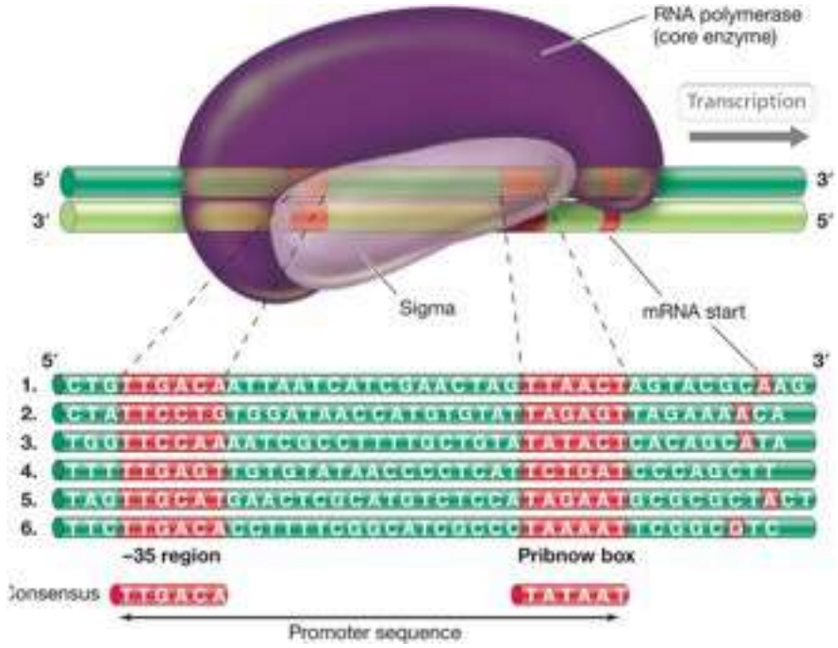
Name[a]	Upstream recognition sequence[b]	Function
σ^{70} RpoD	TTGACA	For most genes, major sigma factor for normal growth
σ^{54} RpoN	TTGGCACA	Nitrogen assimilation
σ^{38} RpoS	CCGGCG	Stationary phase, plus oxidative and osmotic stress
σ^{32} RpoH	TNTCNCCTTGAA	Heat shock response
σ^{28} FliA	TAAA	For genes involved in flagella synthesis
σ^{24} RpoE	GAACTT	Response to misfolded proteins in periplasm
σ^{19} FecI	AAGGAAAAT	For certain genes in iron transport

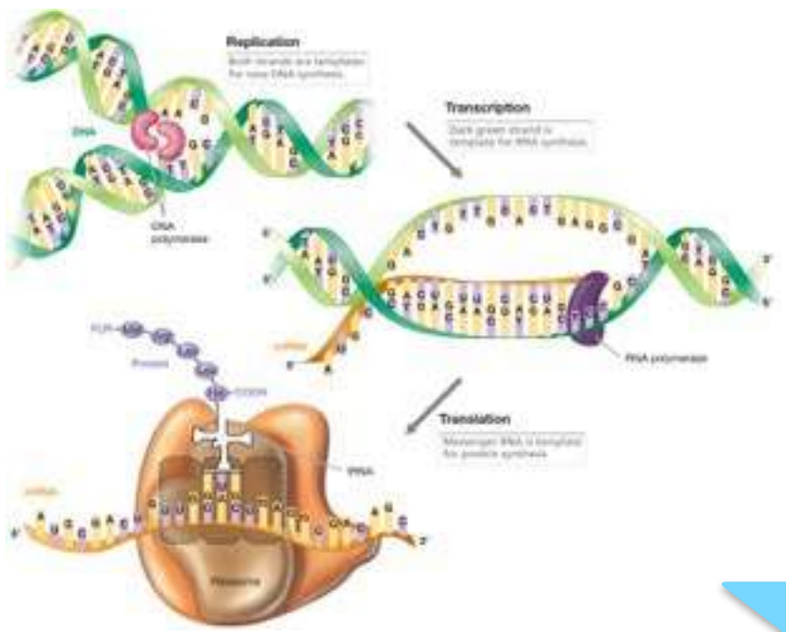
Gene regulation one the presence or absence of σ

Madigan et al. 2018

RNA synthesis: Transcription, V

- Transcription begins at a unique base just downstream from -35 and the Pribnow box
- **Sigma recognizes the promoter sequences on the 5′—>3′ (dark green) strand of DNA**
- **RNA polymerase** core enzyme will actually transcribe the light green strand (that runs 3′—>5′) b/c **core enzyme synthesizes 5′—>3′ direction**





Madigan et al. 2018

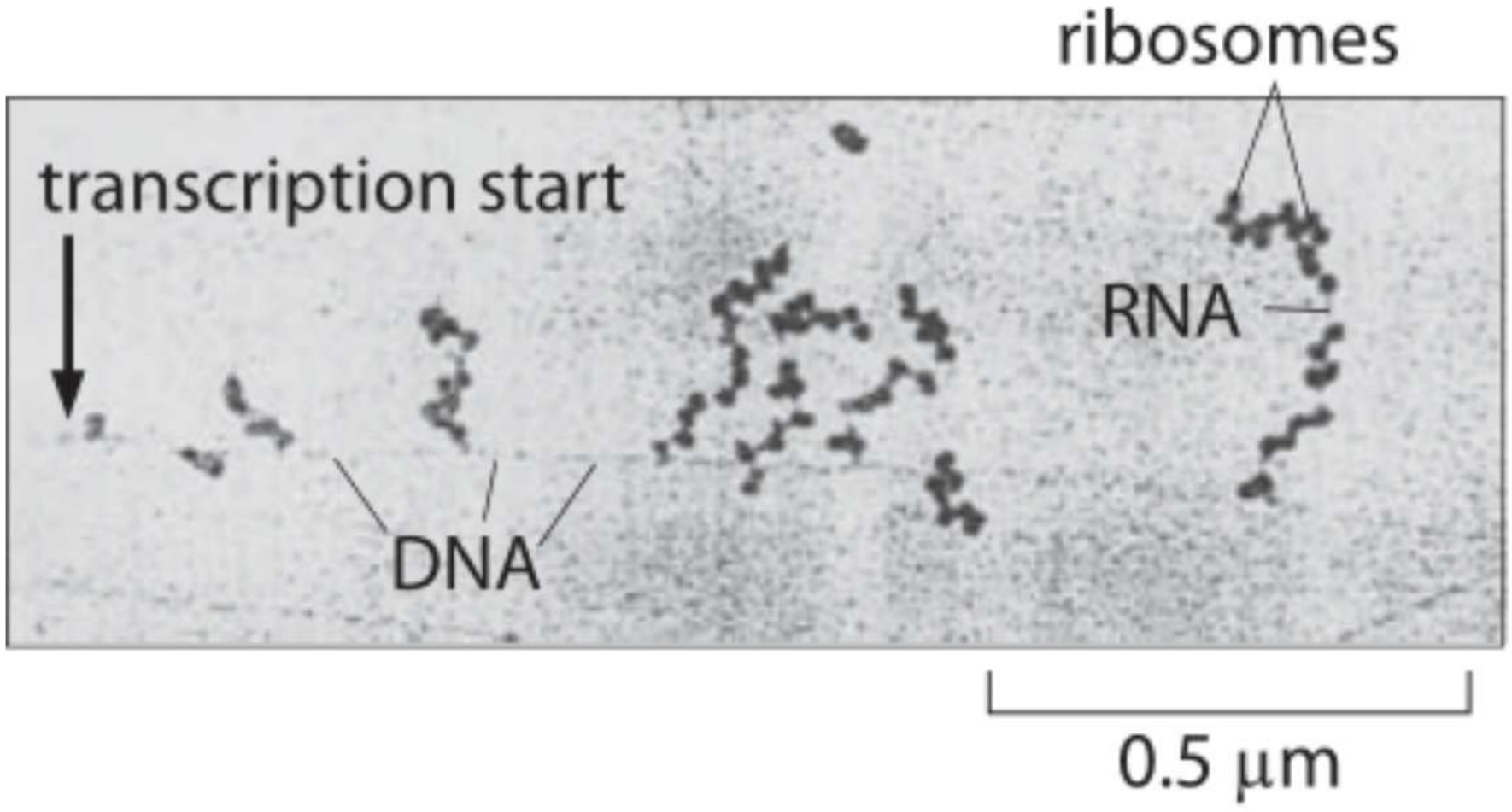
RNA synthesis: Transcription, VI

- RNA polymerase **stop** at transcription terminator sequence a **GC-rich sequence** containing an **inverted repeat with a central non repeating segment**
- When such a DNA sequence is transcribed, the **RNA forms a stem–loop structure by intra-strand base pairing**
- Stem–loops followed by a run of adenines in the DNA template (which yield a run of uridines in the mRNA) are strong transcription terminators because a stretch of U:A base pairs are formed that hold RNA and DNA together
- U:A base pairs is very weak 2 H- bonds not 3 H-bonds as in T:A pairs
- RNA polymerase pauses at the stem–loop, and the **DNA and RNA dissociate**

RNA synthesis: Transcription, VII

- **Terminator protein Rho catalyzes stop transcription**
- Rho binds tightly to RNA and moves down the chain toward the RNA polymerase–DNA complex
- Once **RNA polymerase has paused at a Rho-dependent termination site** (a specific sequence on the DNA template), Rho causes both the **RNA and RNA polymerase to be released from the DNA**, thus terminating transcription

Transcription



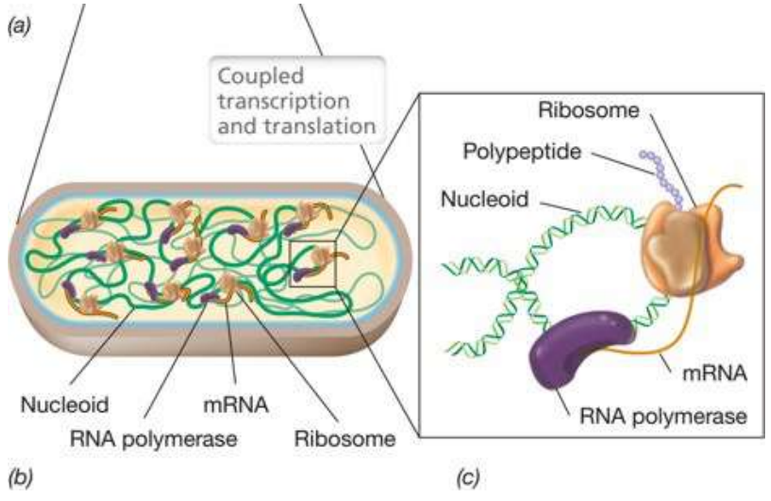
Electron microscopy image of simultaneous transcription and translation.

The image shows bacterial DNA and its associated mRNA transcripts, each of which is occupied by ribosomes. (Adapted from O. L. Miller et al., Science 169:392, 1970.)

Transcriptional units

- **Genetic information is organized into transcriptional units**, segments of DNA that are transcribed into a single RNA molecule bounded by their initiation and termination sites
- Single gene or two or more genes (cotranscribed genes), **operon**
- Most genes encode proteins
- **Genes encode nontranslated RNAs as rRNA (16S, 23S, 5S**; S refers to Svedberg units, a measure of particle size), and their genes are **cotranscribed** to form a single transcriptional unit that also includes a **tRNA**
- Transcriptional unit is subsequently "**processed**" by proteins that cut them to form the individual rRNAs or tRNAs
- Assembling genes for the same biochemical pathway or genes needed under the same conditions into an **operon —> coordinated expression**
- Operon into a single mRNA called a **polycistronic mRNA** contain multiple open reading frames,
- When polycistronic mRNA is translated, **several polypeptides are synthesized sequentially by the same ribosome**

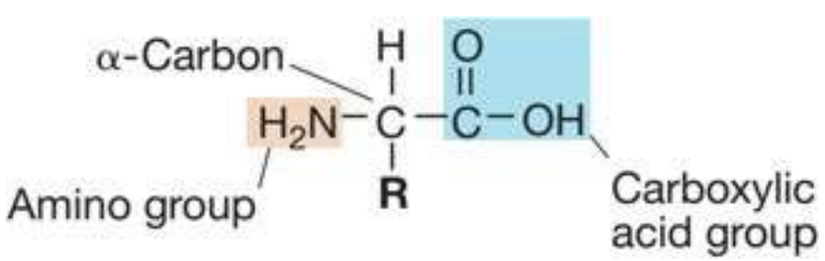
Translation

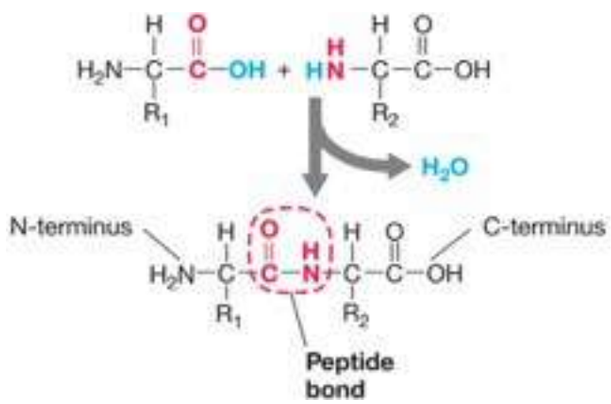


Madigan et al. 2018

- Once transcription has occurred, the mRNAs are translated into protein
- Translation requires many proteins and RNAs (mRNA) and cellular structure, the ribosome

- Proteins are polymers of amino acids, organic compounds that contain both an **amino group** (−NH2) and a **carboxylic acid group** (−COOH) attached to the α-carbon
- Bonds between the carboxyl carbon of one amino acid and the amino nitrogen of a second (formed through the **elimination of water**) are known as **peptide bonds: strong bonds**

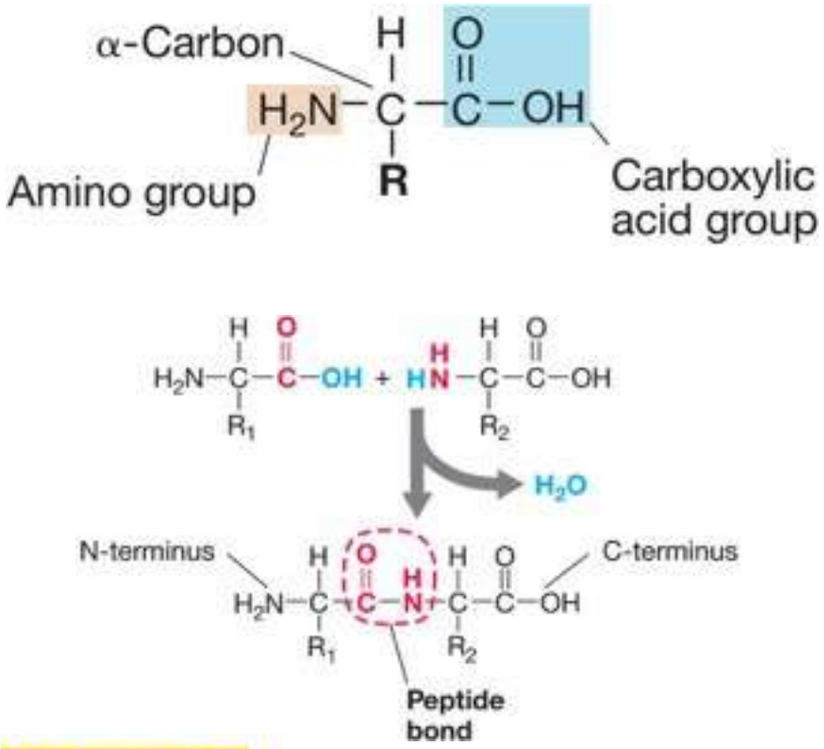




Madigan et al. 2018

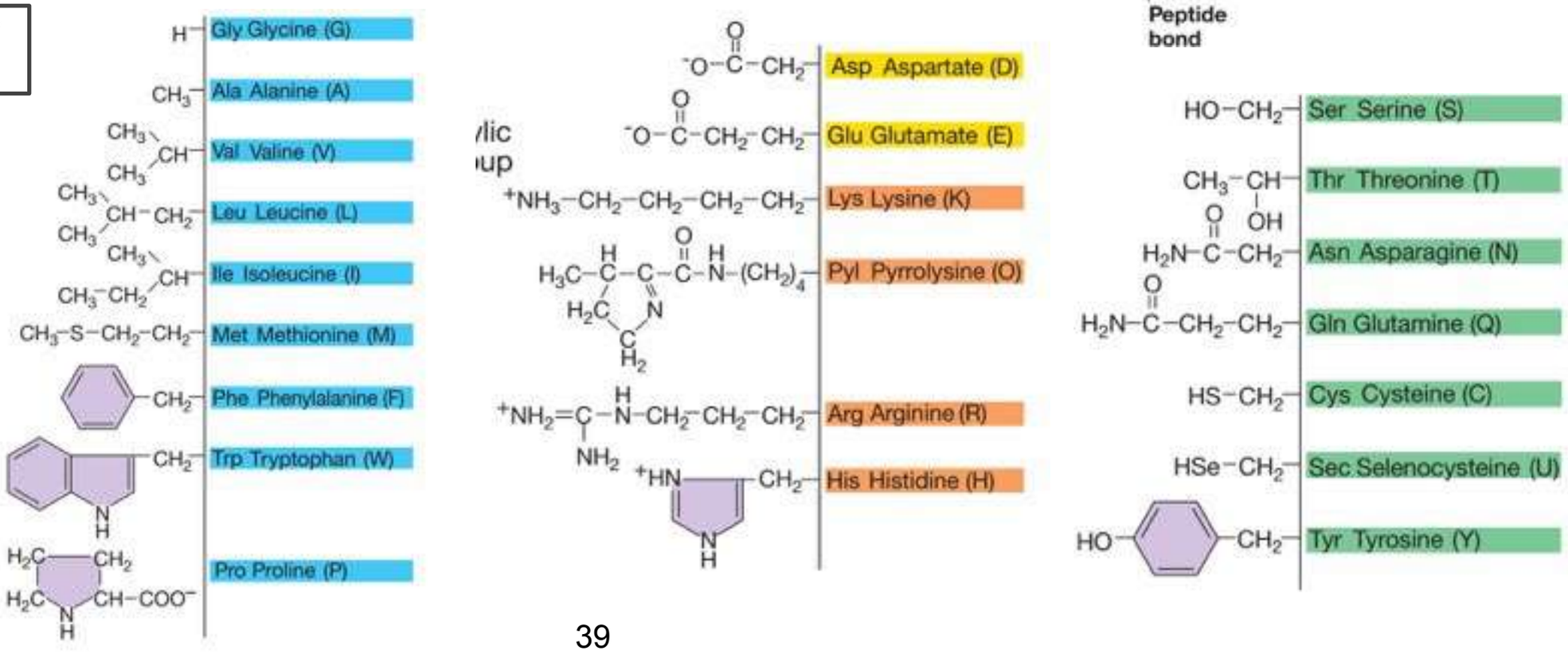
Background: Amino Acid

- L- amino acids form proteins
- Difference in size, charge, behavior according to pH, light absorbing properties
- ***E. coli:* ~2000 different proteins, # dependent nutrients and growth strategy**



Able to do donate or accept H+

- Ionizable: acidic
- Ionizable: basic
- Nonionizable polar
- Nonpolar (hydrophobic)



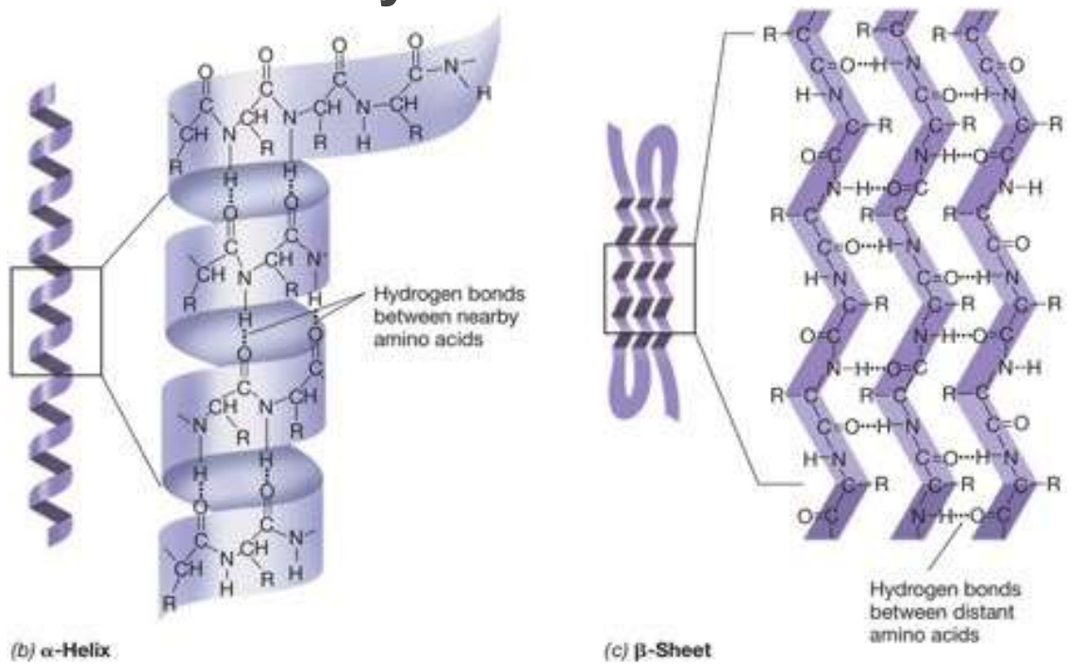
Madigan et al. 2018

Background: Protein, I

Primary structure

F-R-A-N-C-E-S-C-A-M-A-L-F-A-T-T-I

Secondary structure



Madigan et al. 2018

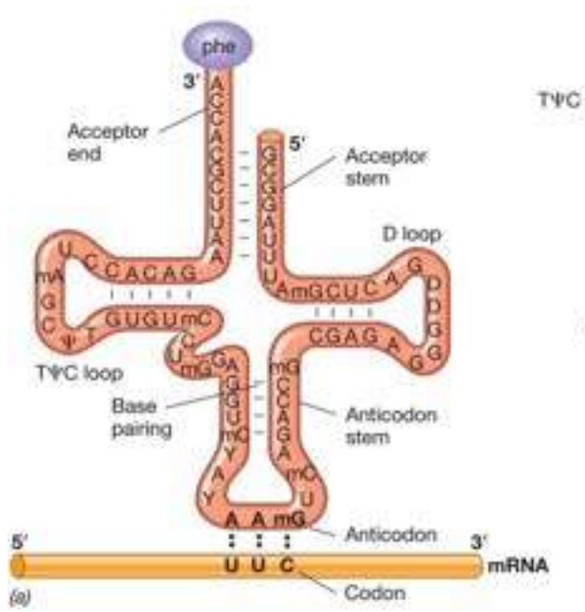
- Catalytic (enzyme) and structural proteins
- **Linear sequence of amino acids** in a polypeptide is its **primary structure**
- Primary structure determines —> **folding pattern of polypeptide —> biological activity**
- **Single aa change** —>change in activity
- Polypeptide folds to form a more stable structure by **H-bonding between the oxygen and nitrogen atoms** of two peptide bonds —> the **secondary structure**
- Secondary structure: **1. a-helix** (polypeptide wound around a cylinder) **2. b-sheet** (a repeated “back and forth” type of folding)

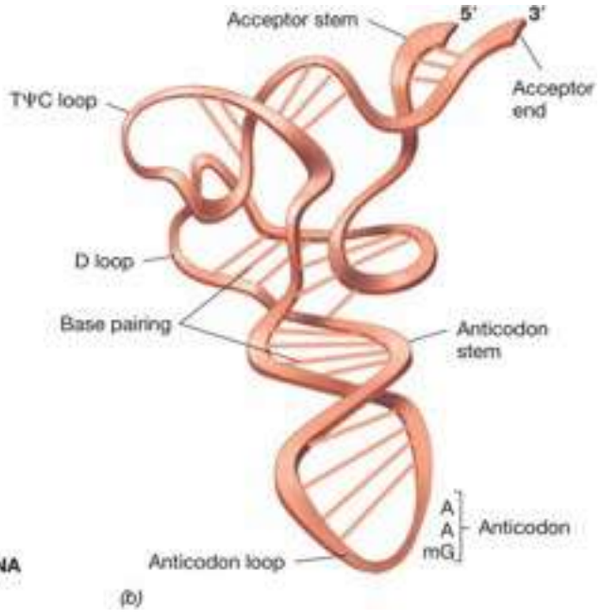
Background: Protein, II

- **Tertiary structure** depends largely on **hydrophobic interactions** (with lesser contributions hydrogen bonds, ionic bonds, and disulfide bonds) **3D shape of the polypeptide**
- Many proteins consist of two or more polypeptides (subunits): **quaternary structure**
- **Stabilization by disulfide bonds**: cysteines located in two different polypeptides are joined, the disulfide bond links the two subunits in **quaternary structure**
- Exposure to **heat, pH, certain chemicals** —> folding—> **denaturation**
- Denaturation: **loss of structure (II, III, IV) —> loss of biological properties**
- Peptide bonds are usually not broken, the denatured polypeptide retains its primary structure
- Depending on the severity of the denaturing conditions, the polypeptide may folds back or not
- If refolding is not correct, the protein is effectively “dead” and is **degraded by proteases** to release its amino acids for new protein synthesis

tRNA, I

- **Transfer RNA**s function to **carry amino acids** to translation machinery
- Each tRNA contains a specific **three-nucleotide sequence called anticodon**, group of **three-bases that recognizes a codon on mRNA**
- **Three-base code for an amino acid**
- **Correct** aa (called **cognate amino acid**) is linked to a specific tRNA by an enzyme called an **aminoacyltRNA synthetase**

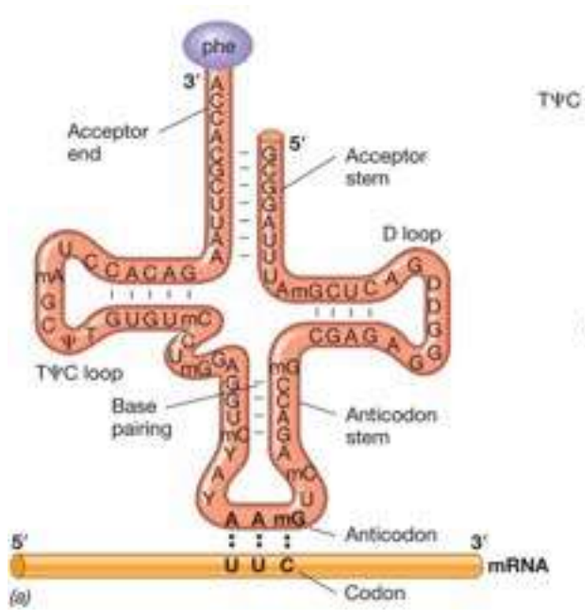


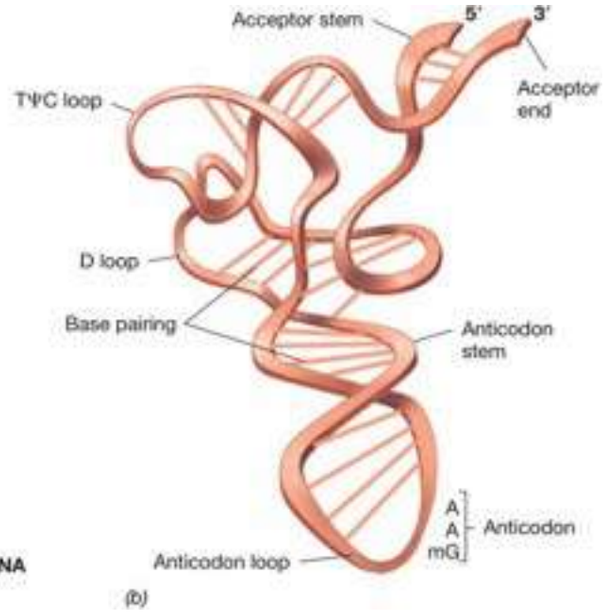


Madigan et al. 2018

tRNA, II

- **For each amino acid, a separate aminoacyltRNA synthetase exists** that specifically binds both cognate amino acid and tRNA containing corresponding anticodon
- **~ 60 different tRNAs in prokaryotic cells**
- **Single-stranded** molecule contains extensive **secondary structure**
- **Modified bases after transcription**: pseudouridine ψ, inosine, dihydrouridine D, ribothymidine, methyl guanosine, dimethyl guanosine, methyl inosine





Madigan et al. 2018

tRNA, III

- The mature and active tRNA also contains extensive **double-stranded regions** formed by internal **base pairing** when single-stranded molecule folds back on itself
- A tRNA is often depicted in shape of a cloverleaf (trifoglio)
- Some regions of tRNA secondary structure are named after modified bases found there (TψC and D loops) or after their functions (anticodon loop and acceptor stem)
- **Bases are in close proximity allowing base pairing in loop**s

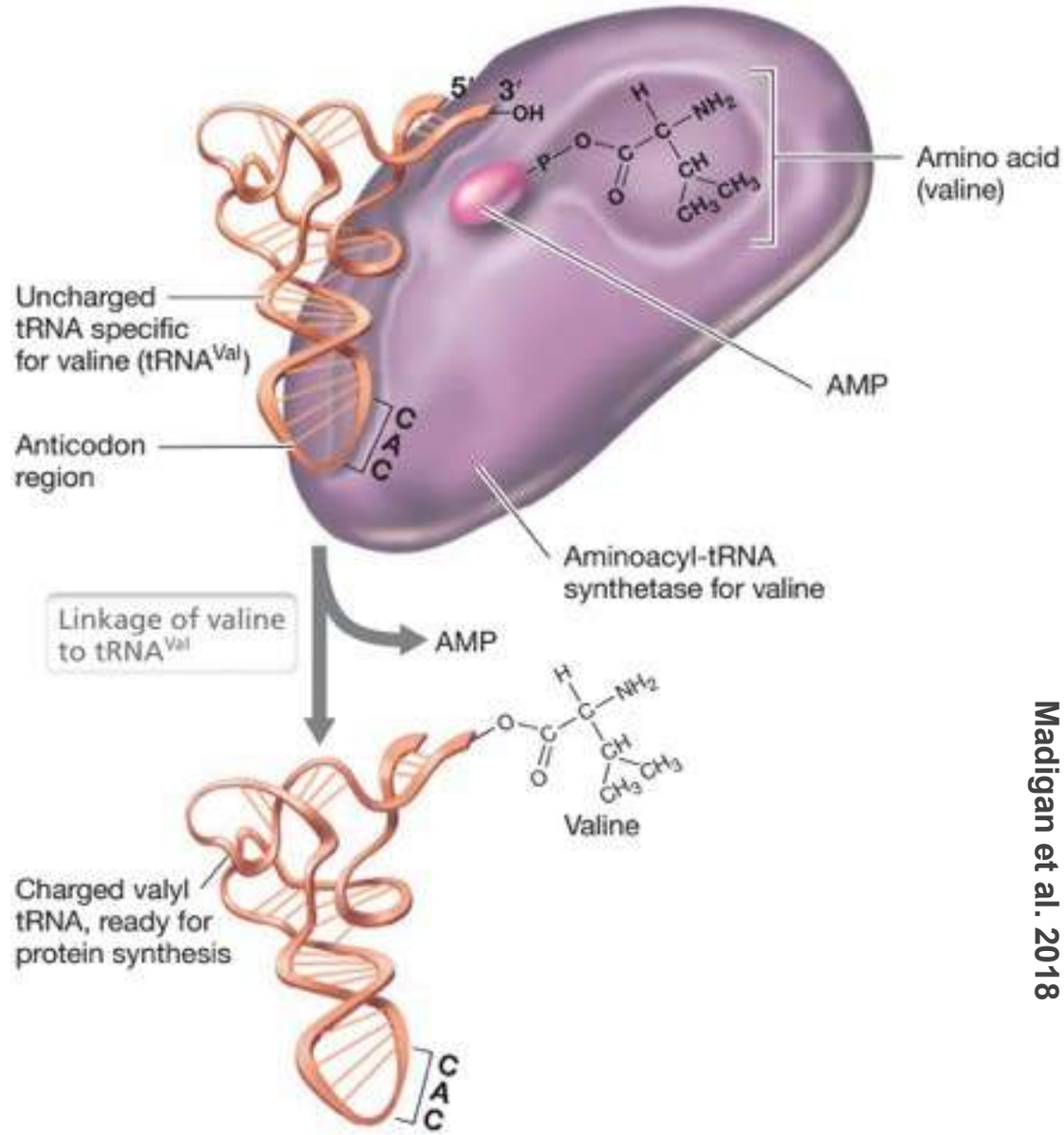
tRNA, IV

- **At the 3′ end** (the acceptor end) of all tRNAs are three unpaired nucleotides: **cytosine-cytosine-adenine (CCA)**
- CCA nucleotide is added sequentially by a protein called CCA-adding enzyme, using CTP and ATP as substrates
- **Cognate amino acid is covalently attached to terminal adenosine of CCA end of its corresponding tRNA**
- From this location, amino acid is incorporated into growing polypeptide chain on ribosome

tRNA, V

The specific reaction between amino acid and tRNA catalyzed by the aminoacyl-tRNA synthetase:

1. **Activation of aa by reaction with ATP**: Amino acid + ATP <—> aminoacyl—AMP + P—P
2. Aminoacyl-AMP intermediate remains bound to tRNA synthetase until collision with appropriate tRNA molecule
3. **Activated aa is attached to CCA stem** of its tRNA to form a **charged tRNA**: Aminoacyl—AMP + tRNA <—> Aaminoacyl—tRNA + AMP
4. Pyrophosphate (PPi) formed in first reaction is split into two molecules of inorganic phosphate



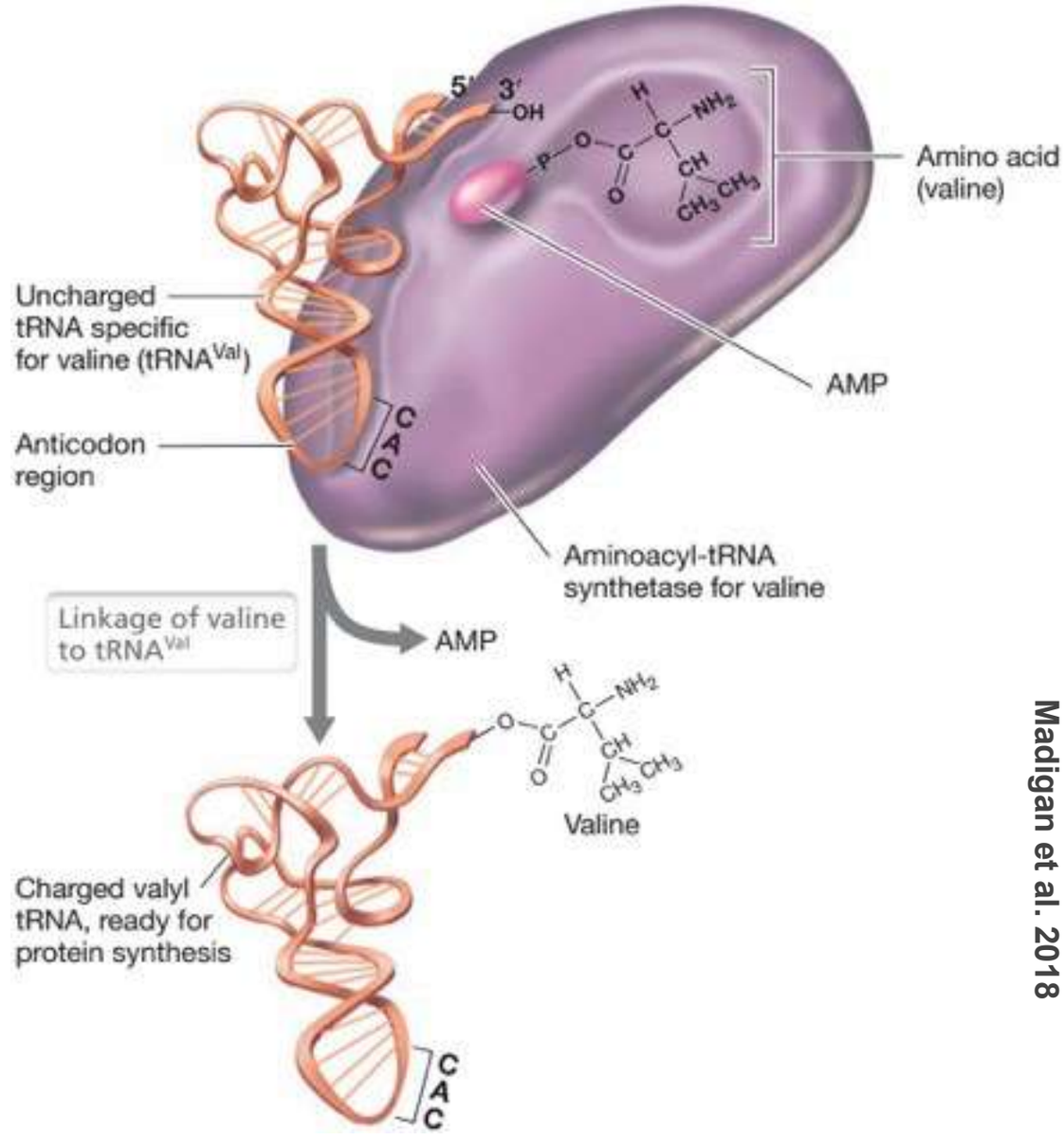
Madigan et al. 2018

tRNA, VI

The specific reaction between amino acid and tRNA catalyzed by the aminoacyl-tRNA synthetase:

- Total of 2 energy-rich phosphate bonds are expended to charge a tRNA with its cognate amino acid

- **After activation and charging, aminoacyl-tRNA leaves** synthetase



Madigan et al. 2018

Background: The Genetic Code, I

- **The heart of genetic information transfer is the correspondence between the nucleic acid template and the amino acid sequence of a polypeptide**
- *This correspondence is rooted in the genetic code*

TABLE 4.4 The genetic code as expressed by triplet base sequences of mRNA

Codon	Amino acid	Codon	Amino acid	Codon	Amino acid	Codon	Amino acid
UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine
UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine
UUA	Leucine	UCA	Serine	UAA	None (stop signal)	UGA	None (stop signal)
UUG	Leucine	UCG	Serine	UAG	None (stop signal)	UGG	Tryptophan
CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine
CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine
CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine
CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine
AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine
AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine
AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine
AUG (start)[a]	Methionine	ACG	Threonine	AAG	Lysine	AGG	Arginine
GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine
GUC	Valine	GCC	Alanine	GAC	Aspartic acid	GGC	Glycine
GUA	Valine	GCA	Alanine	GAA	Glutamic acid	GGA	Glycine
GUG	Valine	GCG	Alanine	GAG	Glutamic acid	GGG	Glycine

[a]AUG encodes *N*-formylmethionine at the beginning of polypeptide chains of *Bacteria*.

64 possible codons (four UCGA bases taken three at a time = 43)

- An **mRNA triplet of three bases**, called a **codon**, encodes each specific amino acid (codons themselves are encoded by organism's genome)

Background: The Genetic Code, II

- **The heart of genetic information transfer is the correspondence between the nucleic acid template and the amino acid sequence of a polypeptide**
- *This correspondence is rooted in the genetic code*
- An mRNA triplet of three bases, called a **codon**, encodes each specific amino acid (the codons themselves are encoded by the organism's genome)
- 64 possible codons, 22 aa
- **Several aa can be encoded by more than one codon**
- A code such as this that lacks one-to-one correspondence between "word" (= aa) and code (= codon) is called a **degenerate code**
- A **codon** is recognized by specific base pairing with a **complementary sequence on the anticodon**, located on a tRNA (A=U and G=C)

Background: The Genetic Code, III

- 6 different tRNAs in *E.coli* for leucine, one for each codon
- **Some tRNAs can recognize more than one codon**
- 2 lysine codons in *E. coli*, there is only one lysyl tRNA, whose anticodon can base-pair with either AAA or AAG
- **Anticodon forms standard base pairs at only the first 2 positions** of the codon and tolerates **irregular base pairing III position —> wobble**
- If aa is encoded by multiple codons, the codons are typically closely related in base sequence, usually differing at only their third position to allow for wobble
- Not all multiple codons for aa are used at the same frequency, leading to a **codon bias (usage)** that varies from organism to organism
- Codon bias is **correlated with bias in different tRNA concentration**

Translation

- **mRNA is translated beginning with its start codon**
- **START encodes a chemically modified methionine** in Bacteria called N-formylmethionine (AUG at the beginning of a coding region encodes N-formylmethionine, **AUG** within the coding region encodes methionine)
- With a triplet code it is critical for translation to begin at correct nucleotide
- **If it does not, the whole reading frame of the mRNA will be shifted and thus an entirely different (and likely inactive) protein will be made or induction of STOP (polypeptide terminates prematurely)**
- **Reading frame fidelity is governed by interactions between mRNA and rRNA within ribosome**
- In Bacteria, **ribosomal RNA recognizes a specific AUG on mRNA as a start codon with aid of an upstream sequence in mRNA called the ribosome-binding site (RBS) or the Shine–Dalgarno sequence**
- **UAA, UAG, and UGA are stop codons** —> **termination** of translation of a protein- coding sequence on mRNA
- Stop codons are also called nonsense codons, because they interrupt the “sense” of the growing polypeptide when they terminate translation
- **If an mRNA can be translated, it is because it contains an open reading frame (ORF): a START codon followed by a number codons and then a STOP codon in the same reading frame as the start codon**

Protein synthesis

1. **Initiation**
2. **Elongation**
3. **Termination**

mRNA, tRNA, and ribosomes, translation requires a number of initiation, elongation, and termination proteins and the **energy-rich compound guanosine triphosphate (GTP)** to provide the energy for the process

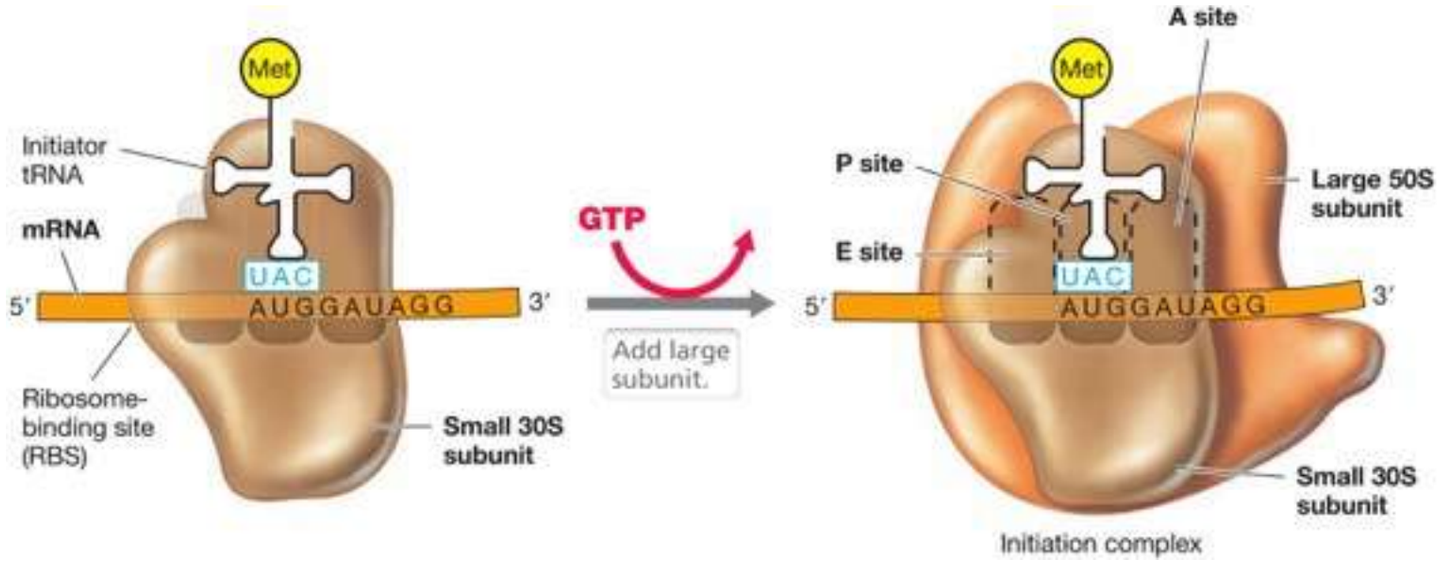
Ribosomes are the sites of protein synthesis, ~ thousands depending on growth rate
Each ribosome (70S) consists of two subunits: 30S and 50S ribosomal subunits

30S subunit =16S rRNA + 21 proteins; 50S subunit = 5S + 23S rRNA + 31 proteins

Ribosome is a **highly dynamic structure** —> subunits alternately associate and dissociate during the translational process and interact with many other proteins (*e.g.* translation factors)

rRNA (ribosomal RNA), a major constituent of the ribosome, **accounting for ~ 2/3 mass**

Initiation

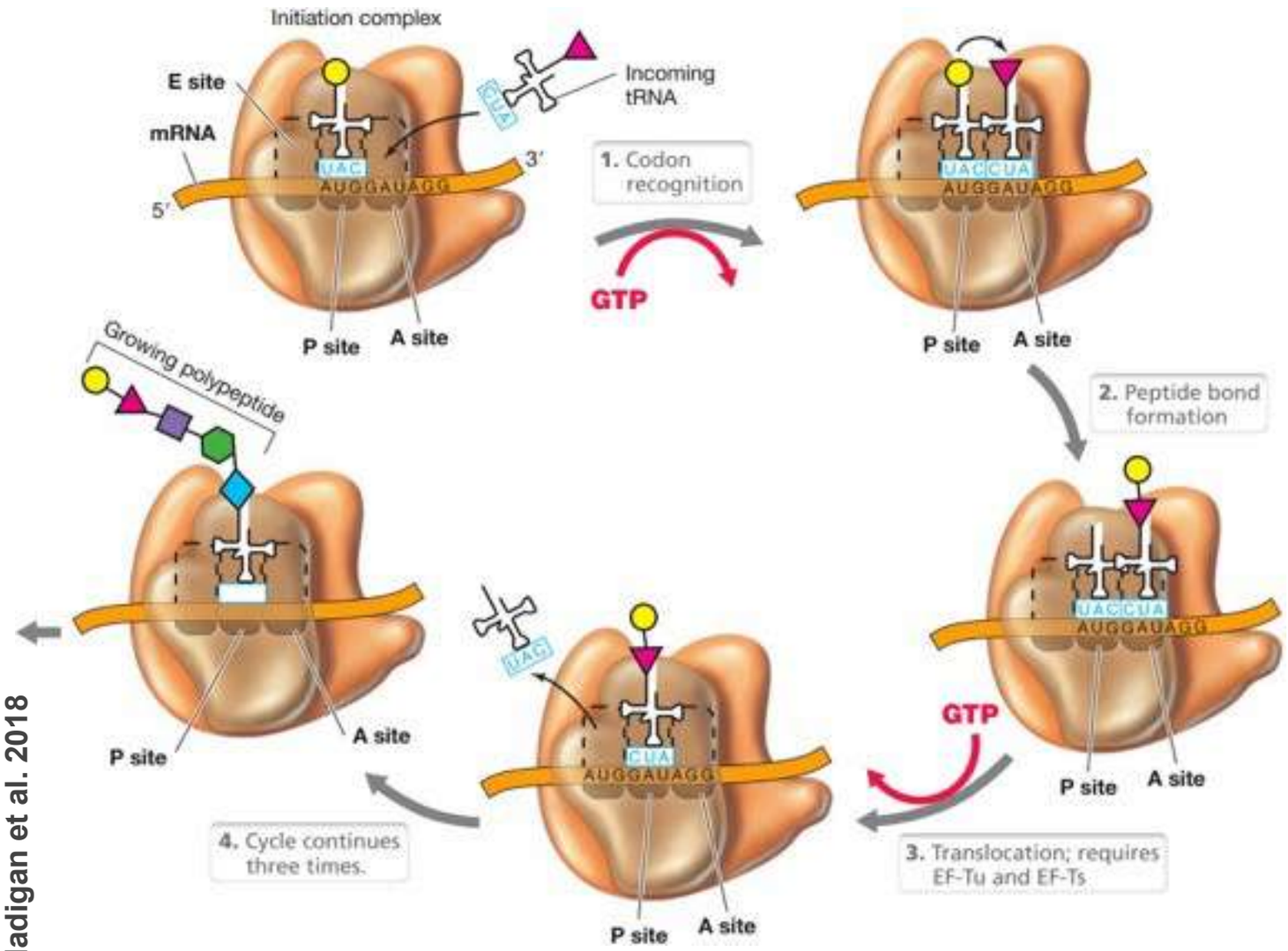


Madigan et al. 2018

- **Initiation complex forms consisting of the 30S subunit, mRNA, formylmethionine tRNA** (initiator tRNA in Bacteria; after polypeptide completion, the formyl group is removed), and **initiation factors** IF1, IF2, and IF3
- **GTP** is also required for this step
- **50S ribosomal subunit is added** to the initiation complex to form the active 70S ribosome
- Just **preceding the start codon on mRNA: sequence** of three to nine nucleotides that compose **ribosome-binding site (RBS)**
- **Ribosome-binding site is toward 5′ end of mRNA and is complementary to base sequences in 3′ end of 16S rRNA part of ribosome**
- Base pairing between these two RNAs holds ribosome–mRNA together in correct reading frame

Elongation, Translocation and Termination, I

mRNA threads through ribosome bound to 30S subunit



Madigan et al. 2018

On 50S subunit:

- **A site** where **incoming charged tRNA first attaches**, and loading of a tRNA into A site is assisted by elongation factor EF-Tu
- **P site** where **growing polypeptide chain** is attached to prior tRNA
- **E site** where free-aa tRNA is released

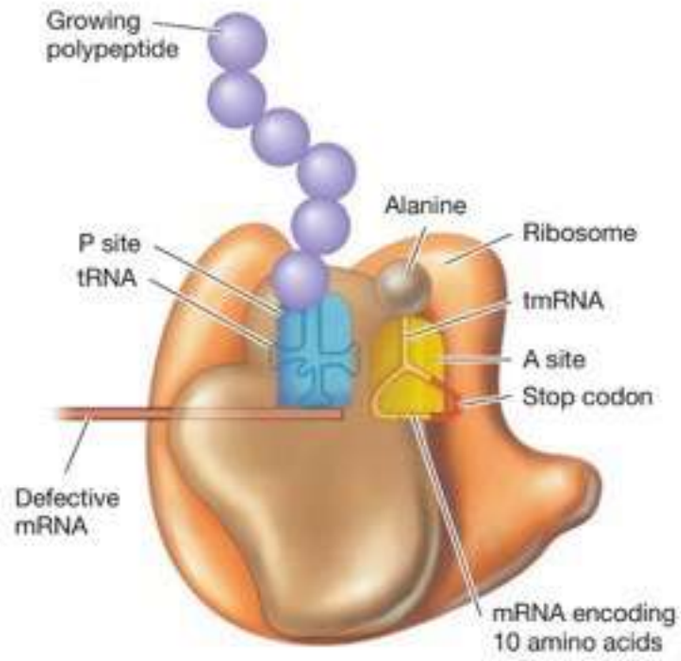
Elongation, Translocation and Termination, II

- **Several ribosomes can translate a single mRNA molecule simultaneously: polysome**
- Ribosomes in polysome that are closest to 5′end (the beginning) of mRNA molecule have short polypeptides attached to them because only a few codons have been read, while ribosomes closest to 3′ end of mRNA have nearly finished polypeptides
- **Translation terminates when ribosome reaches a stop codon because no tRNA binds to a stop codon**
- Proteins called **release factors (RFs) recognize the stop codon and cleave attached polypeptide from final tRNA, releasing finished product**
- Ribosomal subunits dissociate, and 30S and 50S subunits are then free to form new initiation complexes and repeat the process
- rRNA plays major roles in all stages of translation, from initiation to termination
- **Proteins in ribosome is to form a scaffold to position key sequences rRNAs**

Elongation, Translocation and Termination, III

- **16S rRNA facilitates initiation by base pairing with ribosome-binding site on mRNA**, and, along with ribosomal proteins, **holds mRNA in position on either side of the A and P sites**
- **Ribosomal RNA** also plays a role in ribosome subunit association, as well as in **positioning tRNAs on ribosome**
- Although charged tRNAs recognize correct codon by codon–anticodon base pairing, they are also bound to ribosome by interactions between **anticodon stem–loop of tRNA and specific sequences in 16S rRNA**
- Acceptor end of **tRNA base-pairs with sequences of 23S rRNA**
- In addition to roles in mRNA alignment and translocation along transcript, **rRNA also catalyzes actual formation of peptide bonds**
- **Peptidyl transferase reaction occurs on 50S subunit of ribosome and is catalyzed solely by 23S rRNA**
- **23S rRNA also plays a role in translocation and interacts with the elongation factors**
- *Role as the structural backbone of the ribosome, ribosomal RNA plays a major catalytic role in the translation process*

Elongation, Translocation and Termination, IV



Madigan et al. 2018

If a ribosome reaches end of an mRNA molecule and there is **no stop codon**, release factor cannot bind and ribosome cannot be released from mRNA

Small RNA molecule called tmRNA that **frees stalled ribosomes**

tmRNA mimics tRNA (carries alanine) and mRNA (short stretch of RNA that can be translated)

When tmRNA collides with a stalled ribosome, it binds alongside the defective mRNA

Protein synthesis can then proceed, first by adding the alanine on the tmRNA and then by translating the short tmRNA message (signal for specific protease)

TmRNA contains a stop codon that allows release factor to bind and disassemble ribosome —> new protein synthesis

Protein is defective and is subsequently **degraded**

Background: Protein, III

- Some proteins require **subsequent processing** before they become functional
- **Assistance in folding** or, for some enzymes, **incorporation of cofactors** or other nonprotein groups
- Bacteria produce a series of **proteins** called **chaperones** that catalyze a variety of macromolecular folding events:

  Not spontaneously folding or too quick folding

  Refolding partially denatured proteins

  Assembling multiprotein complexes

  Preventing the improper aggregation of proteins

  Untangling RNAs (low temperature)

  Incorporating cofactors into enzyme (Mo in nitrate reductase)

- In *E.coli:* DnaK, DnaJ, GroEL, GroES, DnaK and DnaJ are **ATP-dependent or use ATP energy for activities**

Background: Protein, IV

- Proteins called **translocases transport specific proteins into or through membranes** of Bacteria and Archaea.
- **Sec** translocase system both exports unfolded proteins and inserts integral membrane proteins into cytoplasmic membrane, ATP hydrolysis
- **Tat** (twin-arginine) translocase system transports previously folded proteins through cytoplasmic membrane, PMF
- Most proteins that must be **transported have 15-20 aa** sequence called **signal sequence—at beginning, N-terminus**
- Signal sequences are variable: **a few positively charged amino acids at beginning, a central region of hydrophobic residues, a more polar region at their end**