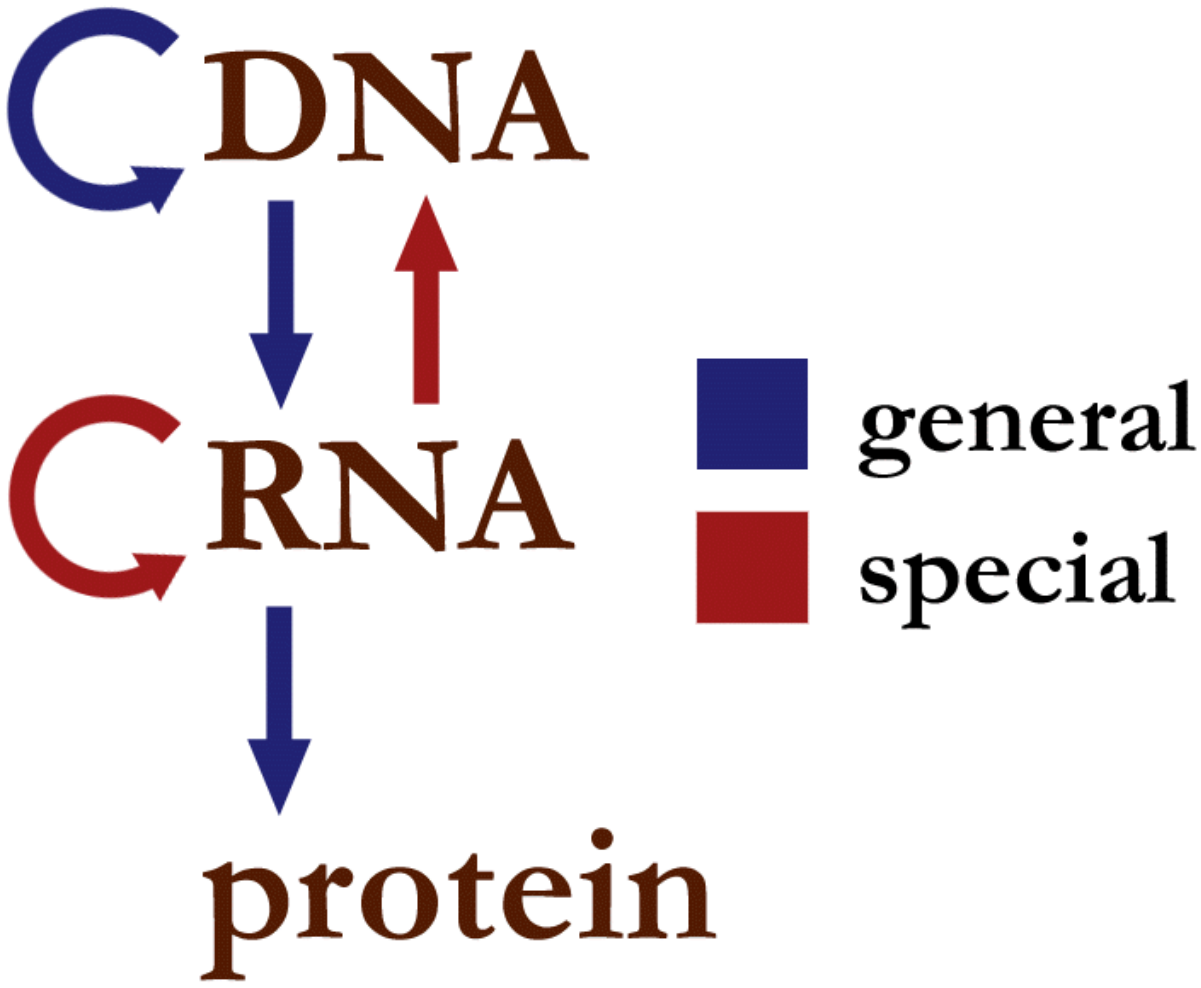
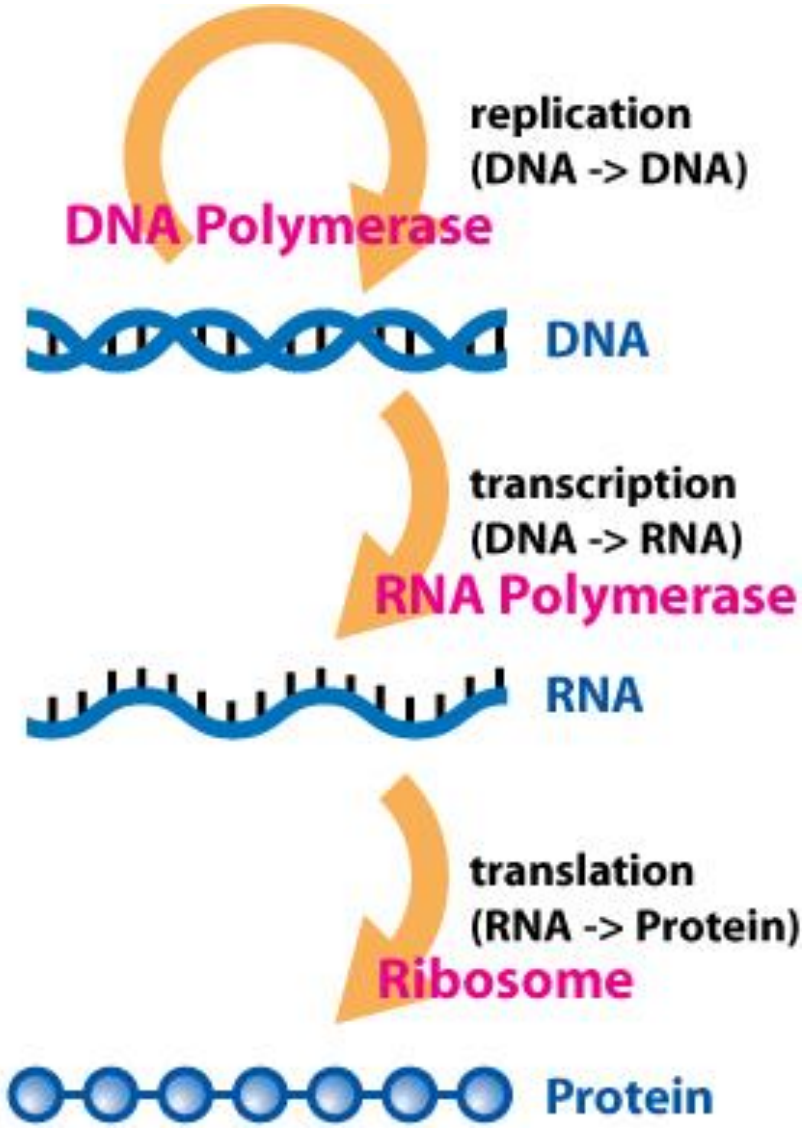
Lecture 5 - Biochemistry

Prof. Sebastian Maerkl

Central Dogma



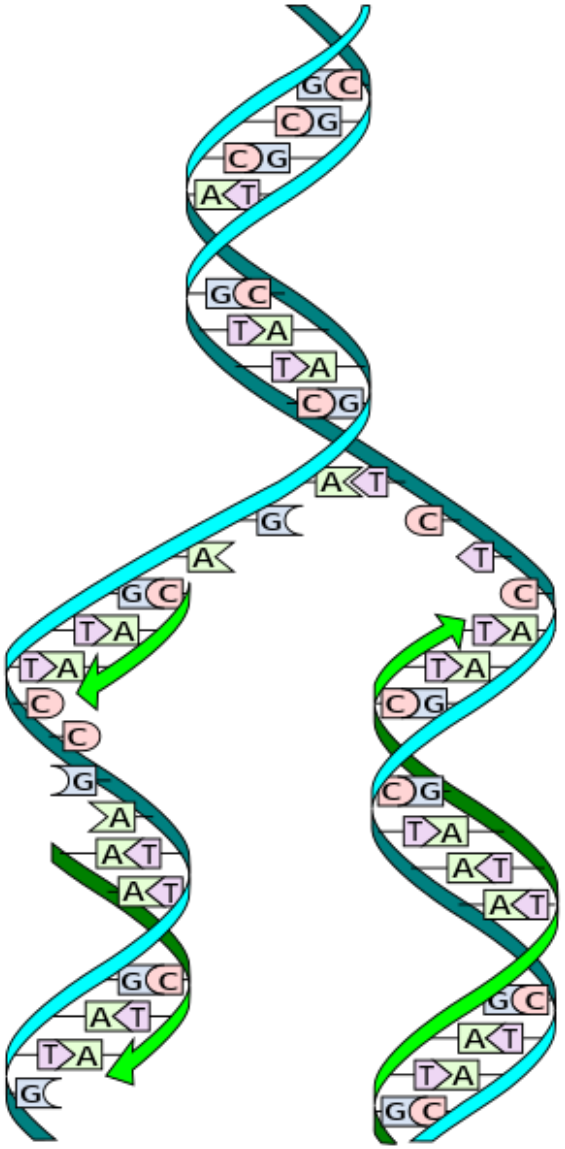


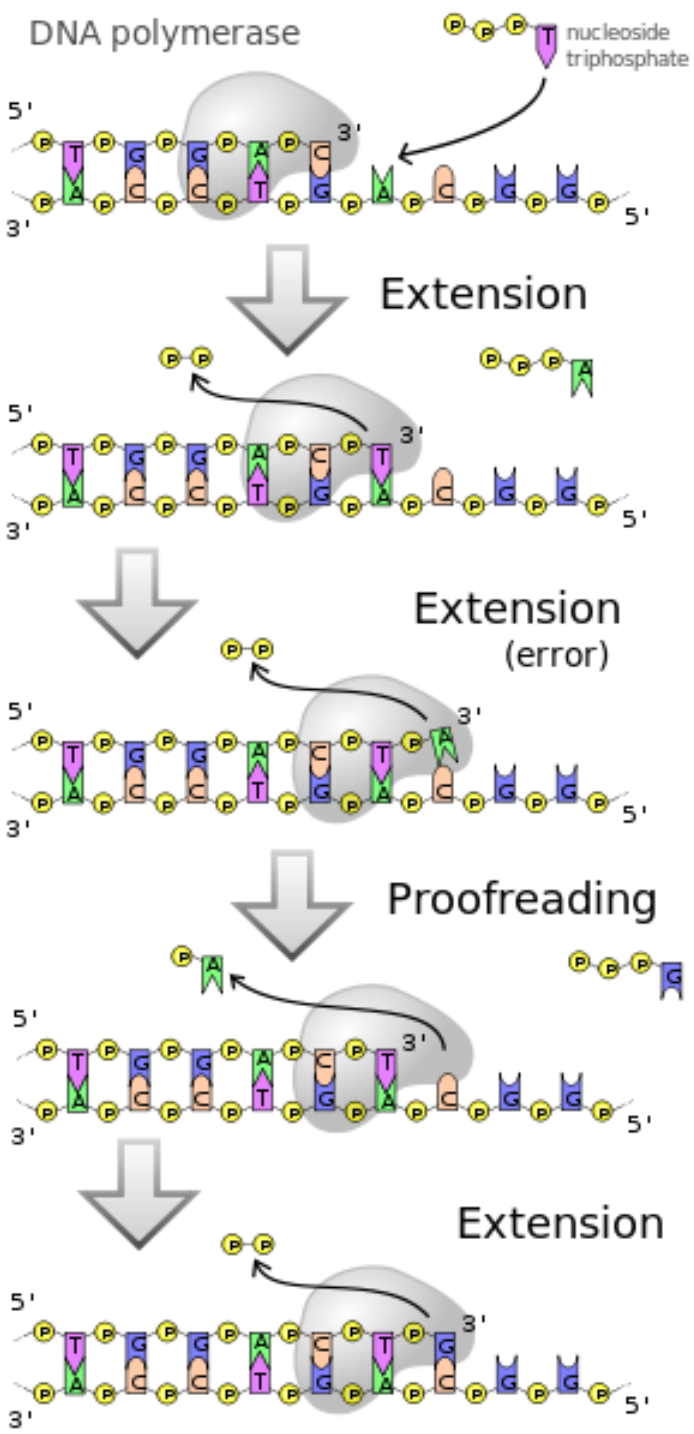
Reverse transcription: reverse transcriptase

RNA replication: RNA-dependent RNA Polymerase

DNA Replication

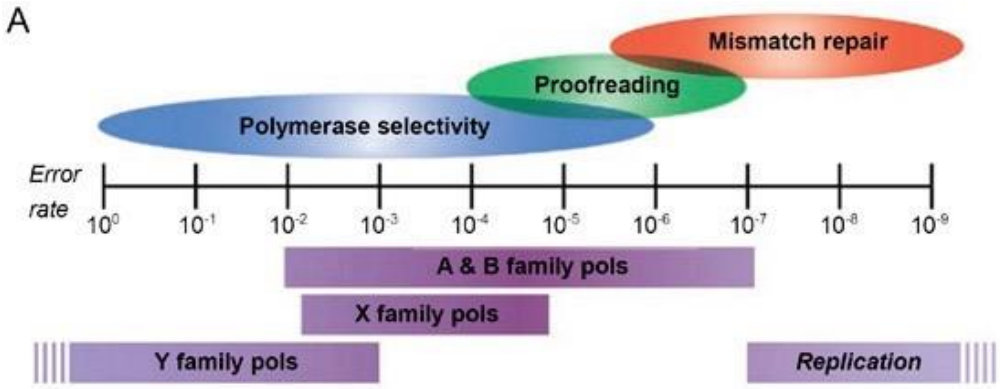
DNA synthesis





5’ -> 3’ elongation
3’ -> 5’ proof-reading (exonuclease)
Strand-displacement

Error rate ~ 10^{-9} (10^{-7})
Processivity > 50 kb (20bp – 70kb)
Synthesis Rate ~750 nucleotides / sec

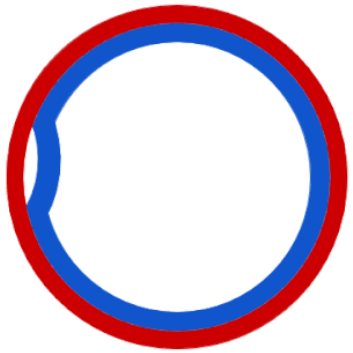


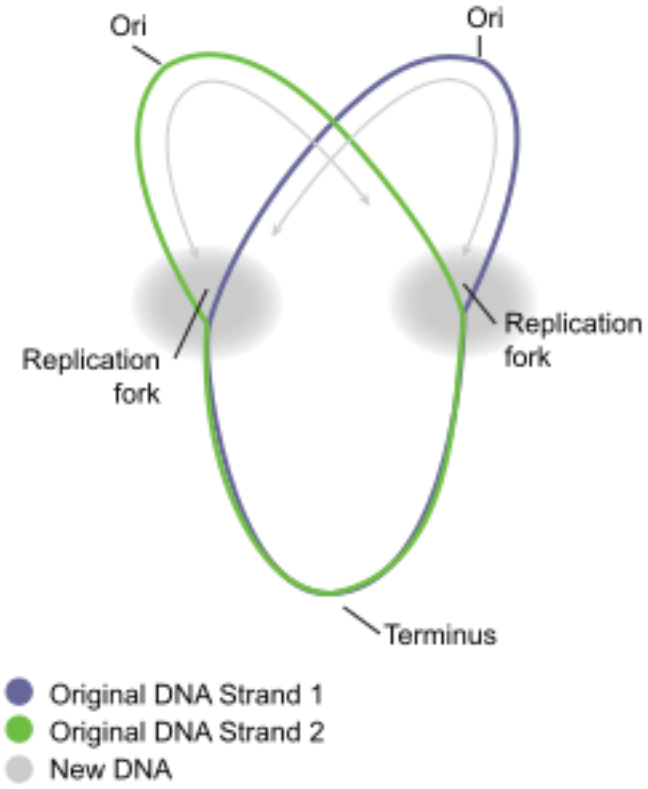
Enzyme	Expt.	Avg. doublings/PCR reaction	Number of clones sequenced	Total bp sequenced	Number of mutations observed	Error rate
Taq	1	20.5 ± 1.2	65	8.8×10^4	54	3.0×10^{-5}
	2	16.7 ± 0.7	37	4.7×10^4	45	5.6×10^{-5}
AccuPrime-*Taq*	1	17.0 ± 1.2	75	1.0×10^5	18	1.0×10^{-5}
	2	16.9 ± 0.6	N.D.	N.D.	N.D.	N.D.
KOD	1	20.8 ± 1.5	70	1.0×10^5	16	7.6×10^{-6}
	2	17.6 ± 0.8	N.D.	N.D.	N.D.	N.D.
Pfu (cloned)	1	16.5 ± 1.1	151	2.0×10^5	9	2.8×10^{-6}
	2	12.0 ± 1.8	N.D.	N.D.	N.D.	N.D.
Phusion	1	21.0 ± 1.9	175	2.4×10^5	13	2.6×10^{-6}
	2	16.6 ± 1.1	N.D.	N.D.	N.D.	N.D.
Pwo	1	22.5 ± 1.2	170	2.4×10^5	13	2.4×10^{-6}
	2	17.6 ± 0.6	N.D.	N.D.	N.D.	N.D.

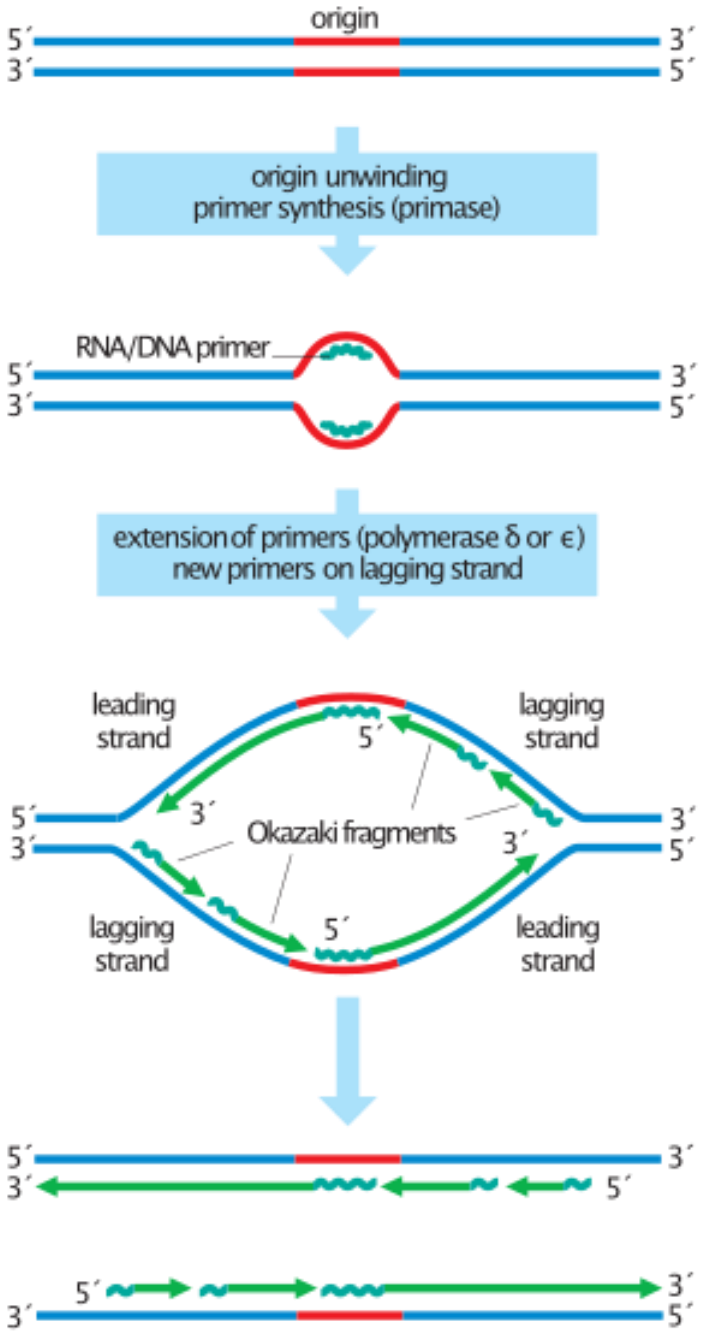
N.D.: not determined.

Prokaryotic DNA replication

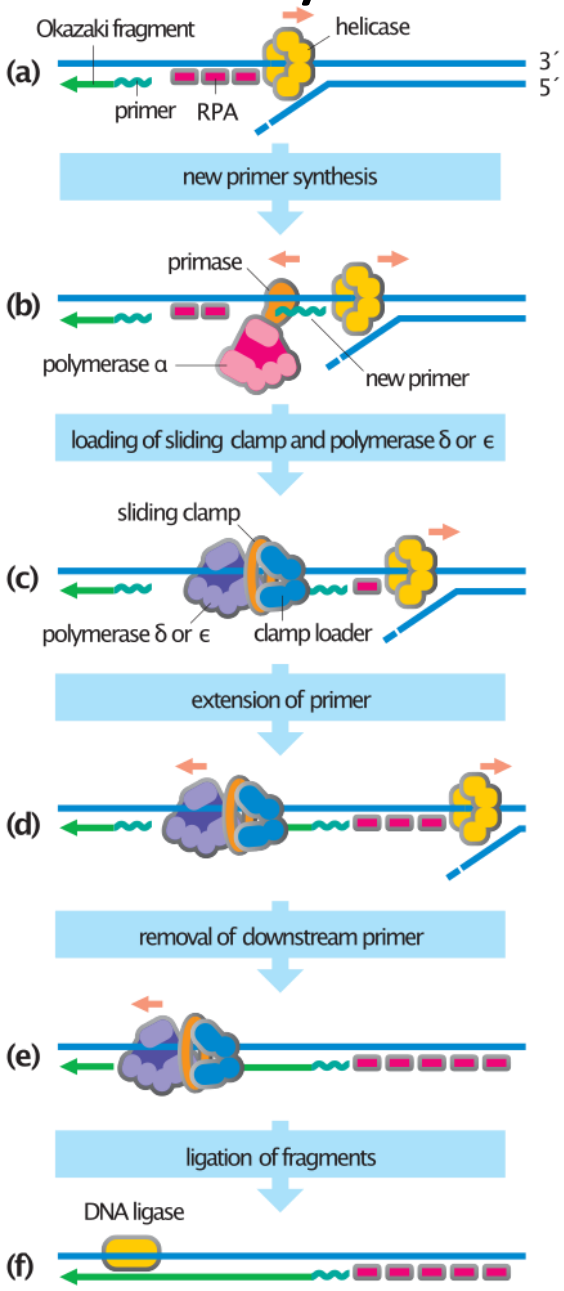
Theta-type replication

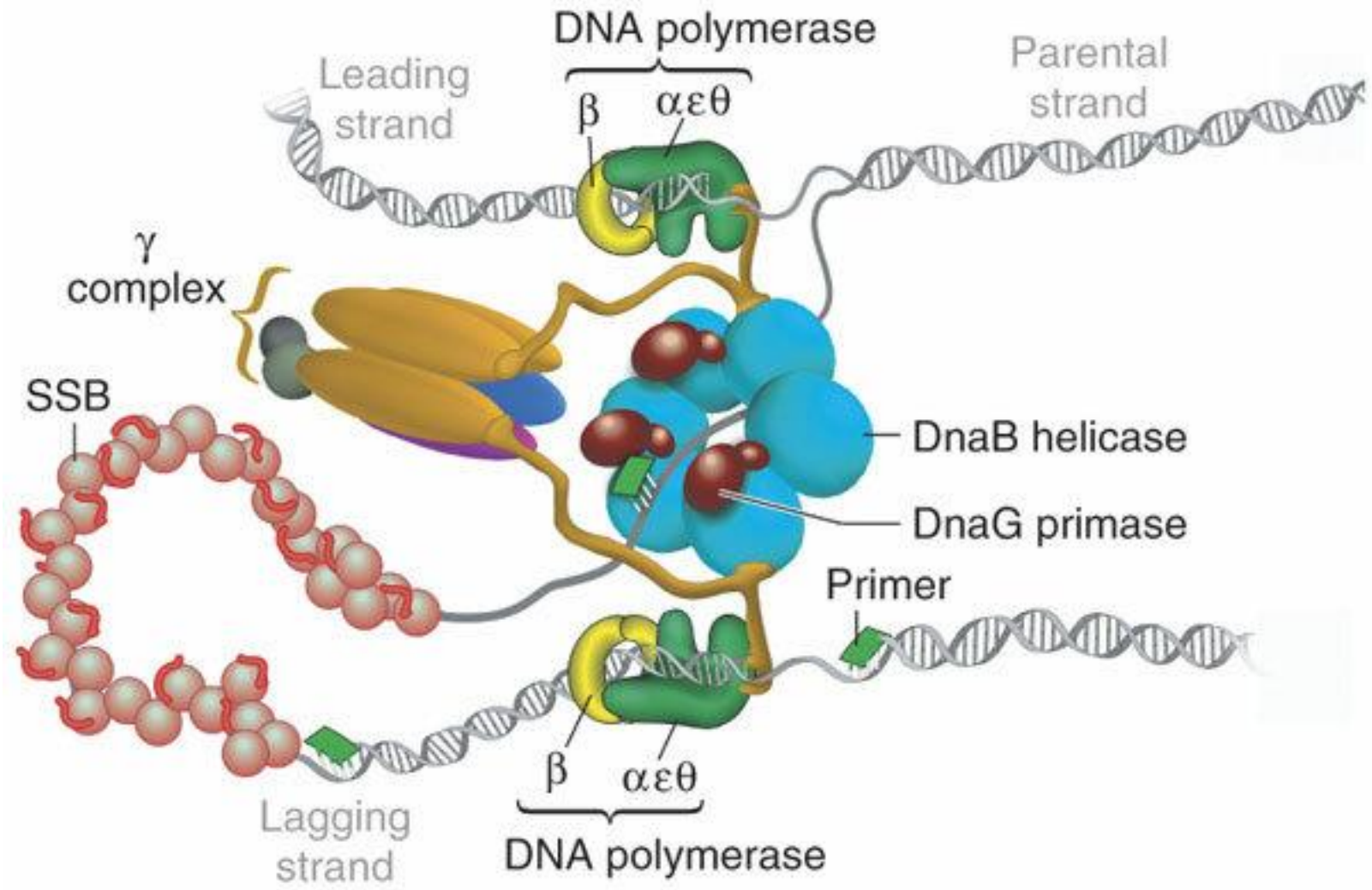






Prokaryotic DNA replication





Prokaryotic DNA replication

oriC:
- origin of replication
- contains binding sites for DnaA

DnaA:
- DnaA binding to oriC leads to strand displacement

DnaC helicase loader:
- interacts with DnaA and recruits DnaB helicase

DnaB helicase:
- unwinds DNA

DnaG primase :
- lays down RNA primer

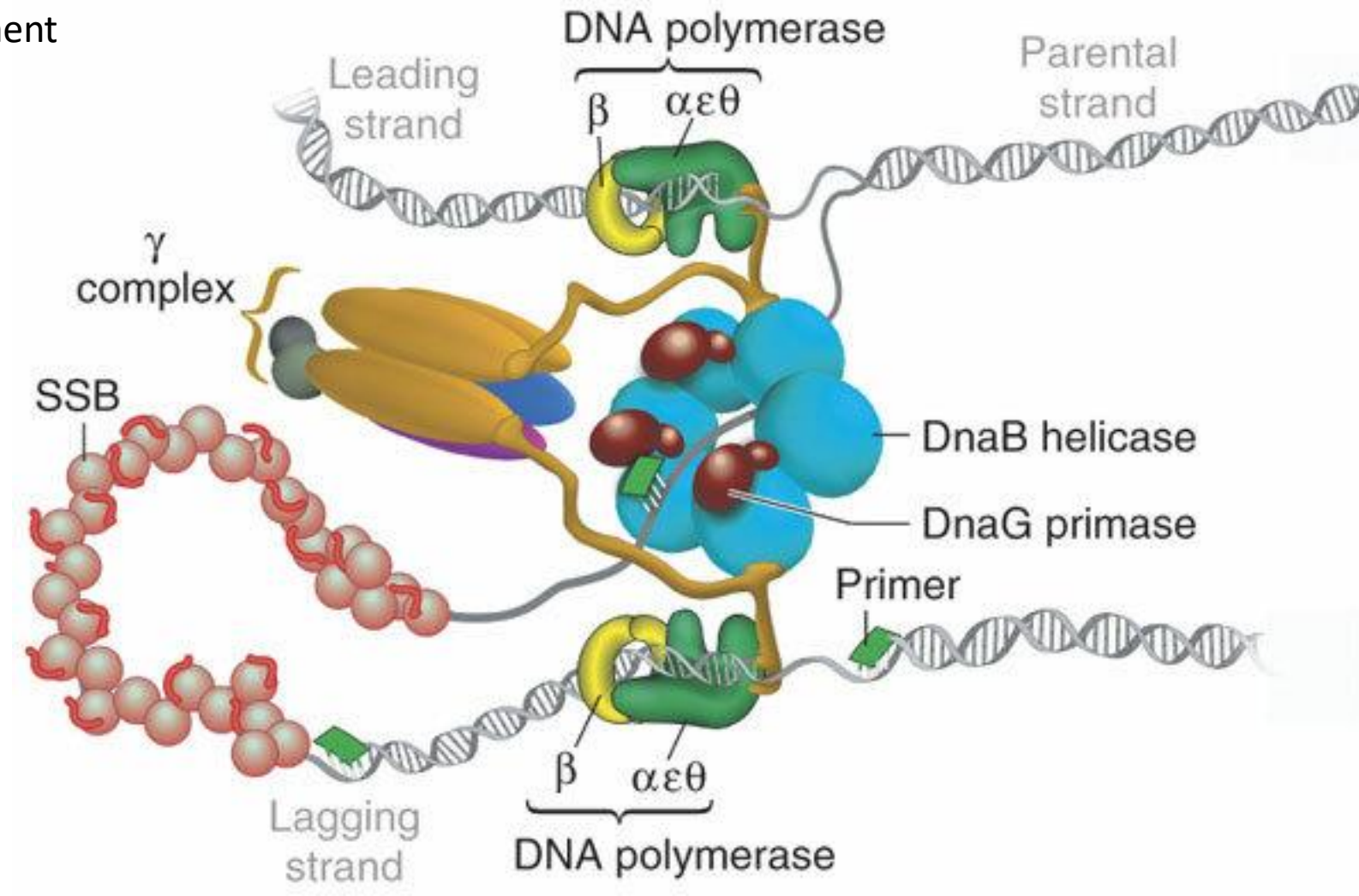
DNA polymerase III holoenzyme:
- synthesizes DNA

SSB:
- single strand binding protein
- stabilizes ssDNA

Leading strand:
- continuously synthesized

Lagging Strand:
- synthesized in short separate fragments

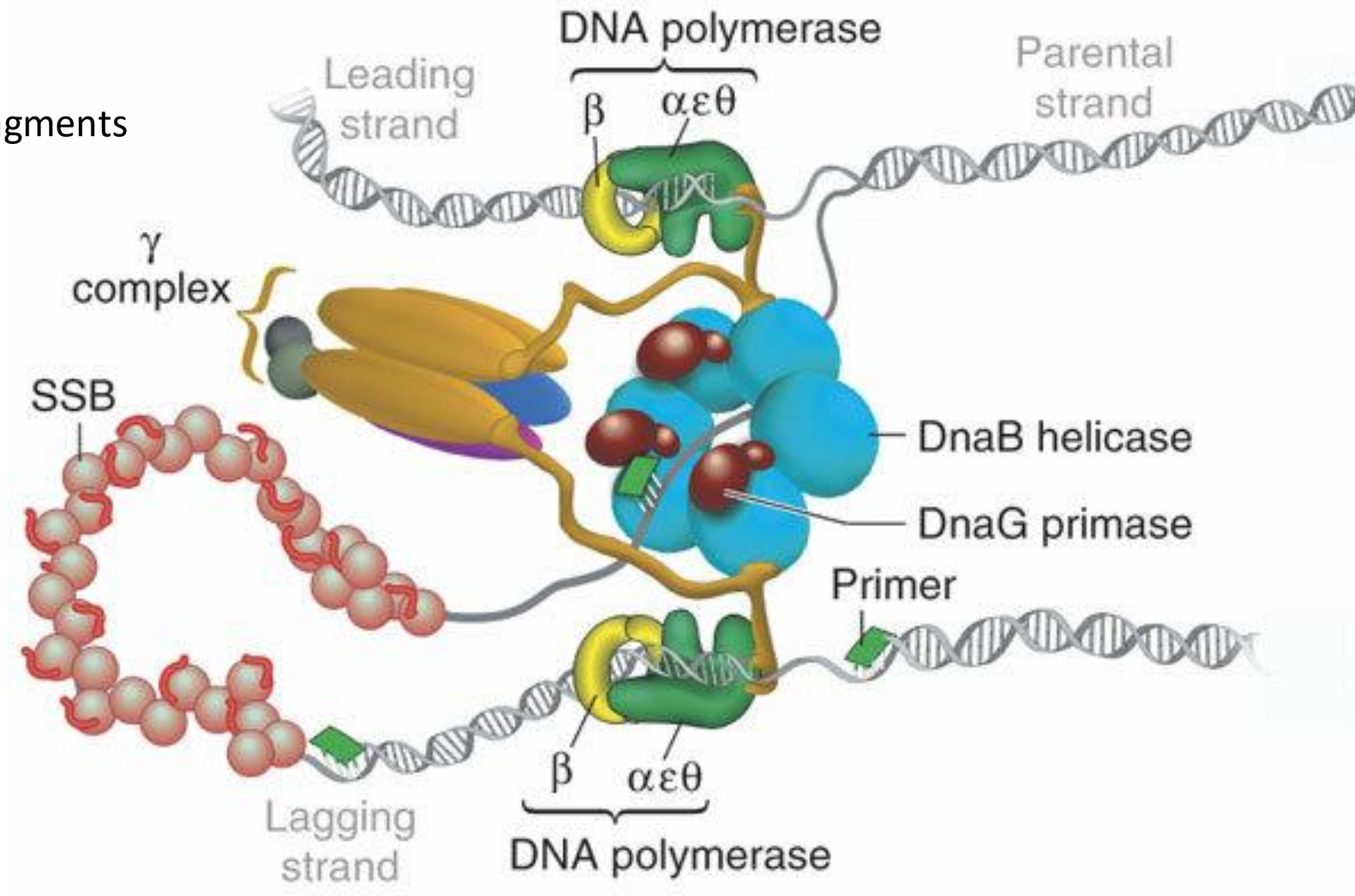
Okazaki fragment



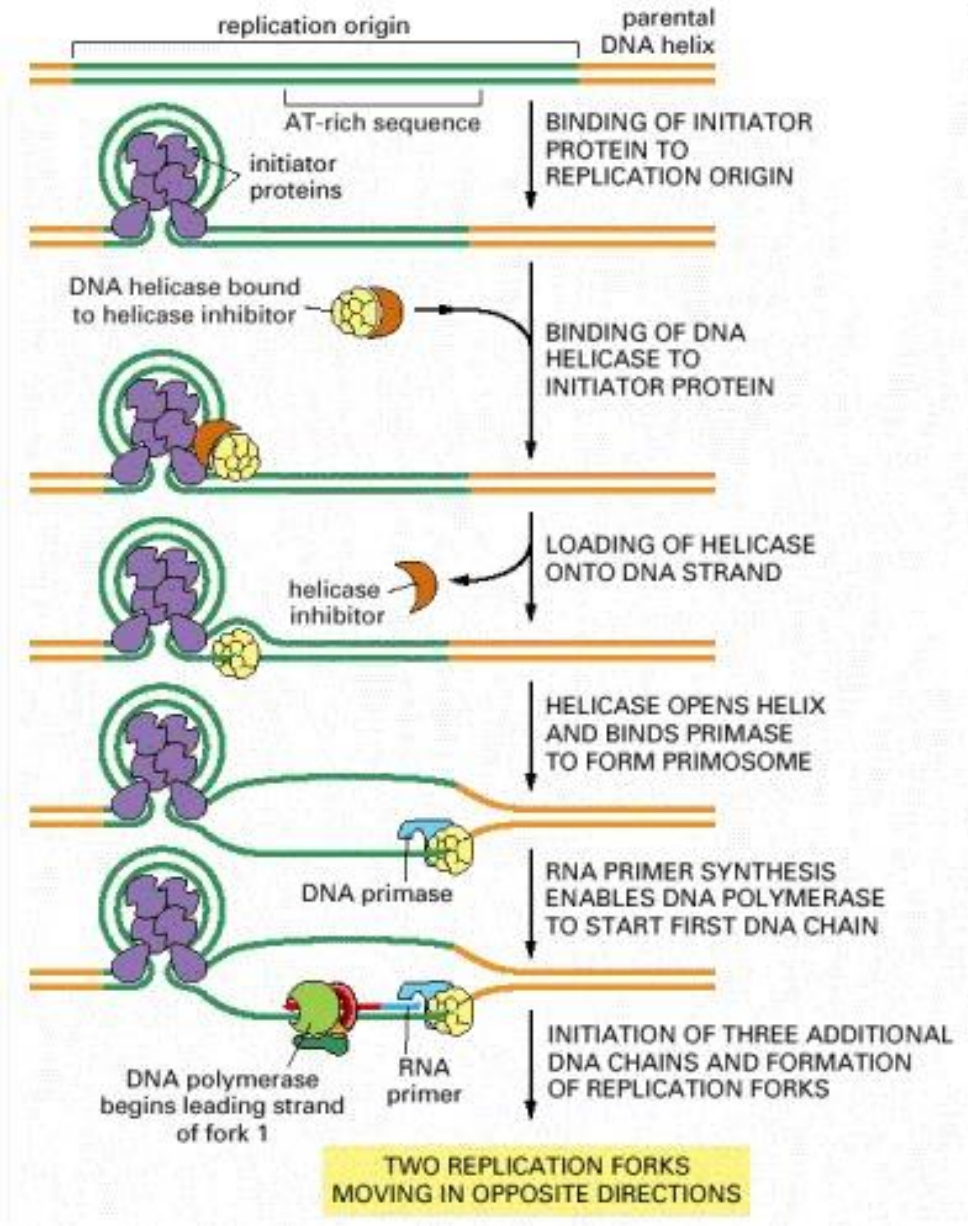
Resolving Okazaki fragments

RNaseH and DNA polymerase I:
- degrade RNA primers
- fill in deoxyribonucleotides

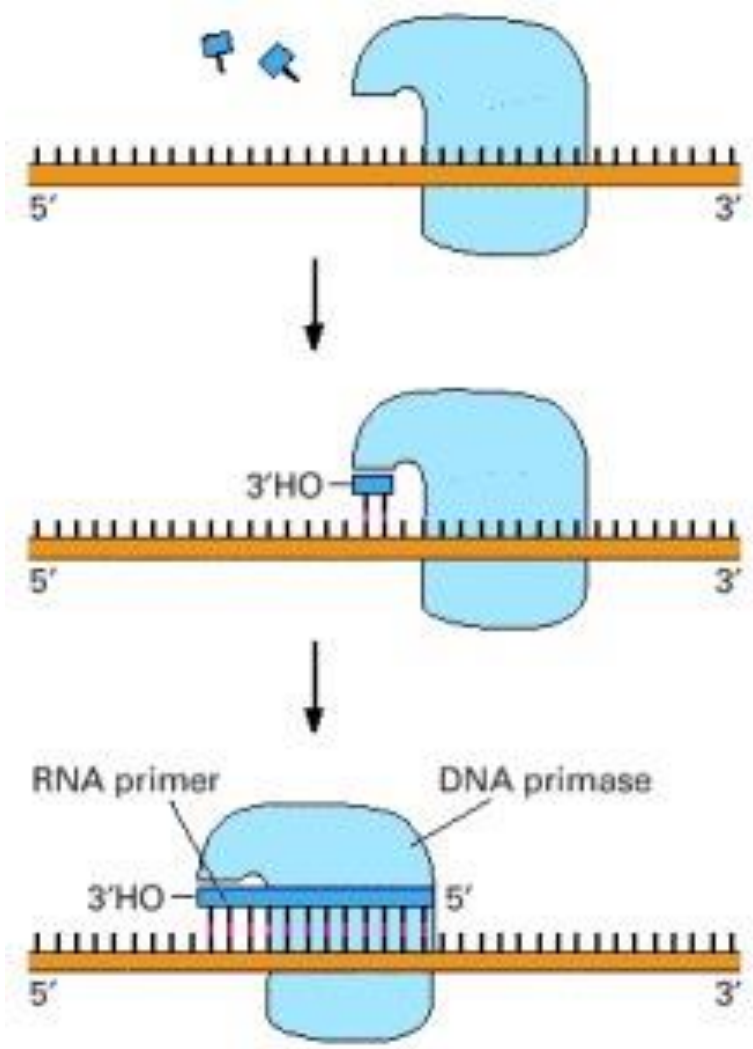
Ligase:
- seals (ligates) the gap between two Okazaki fragments



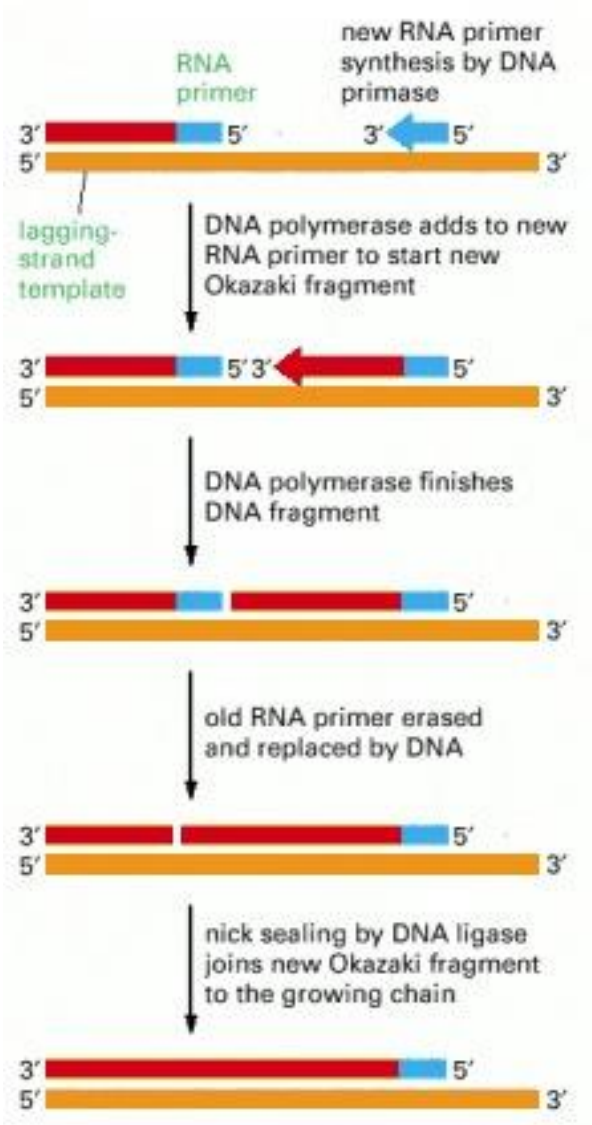
Origin of replication and initiation of DNA synthesis



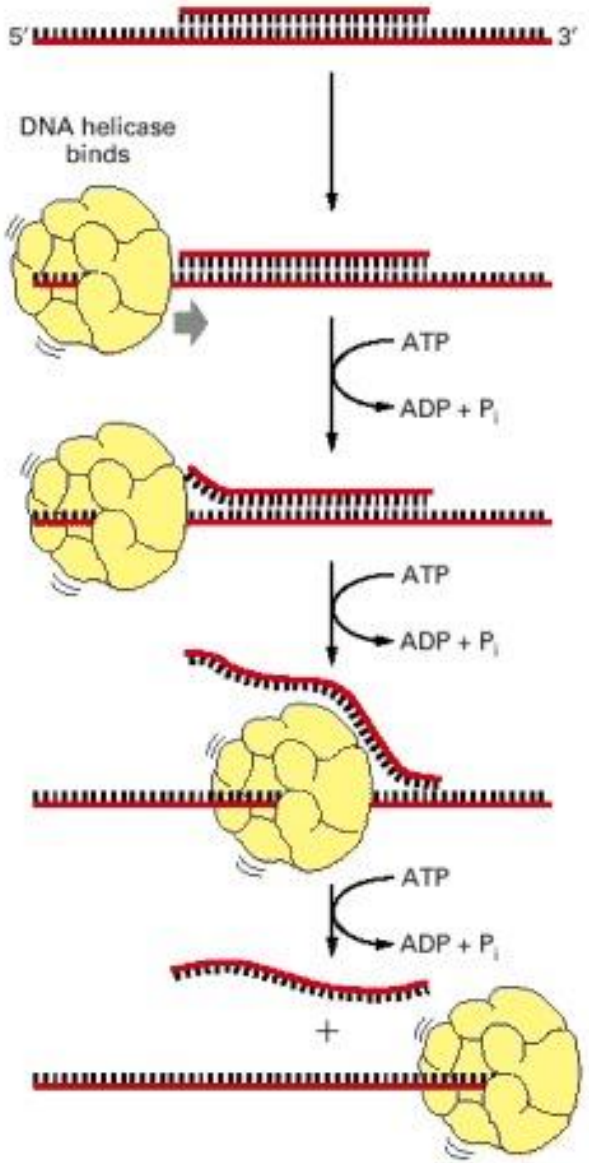
Short RNA primer synthesis by DNA primase



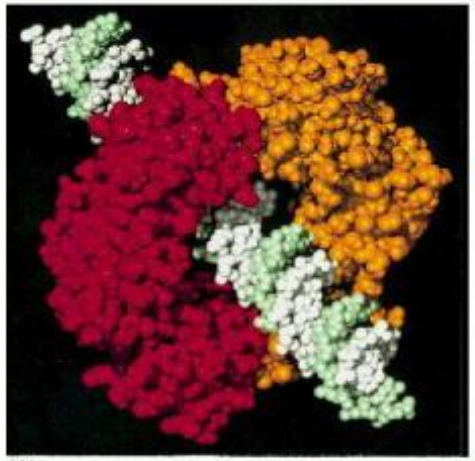
Lagging strand DNA synthesis



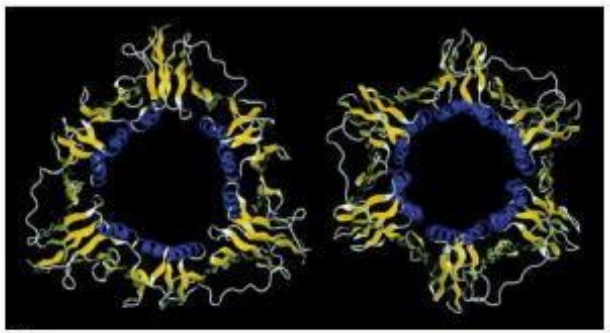
Strand displacement by DNA helicase



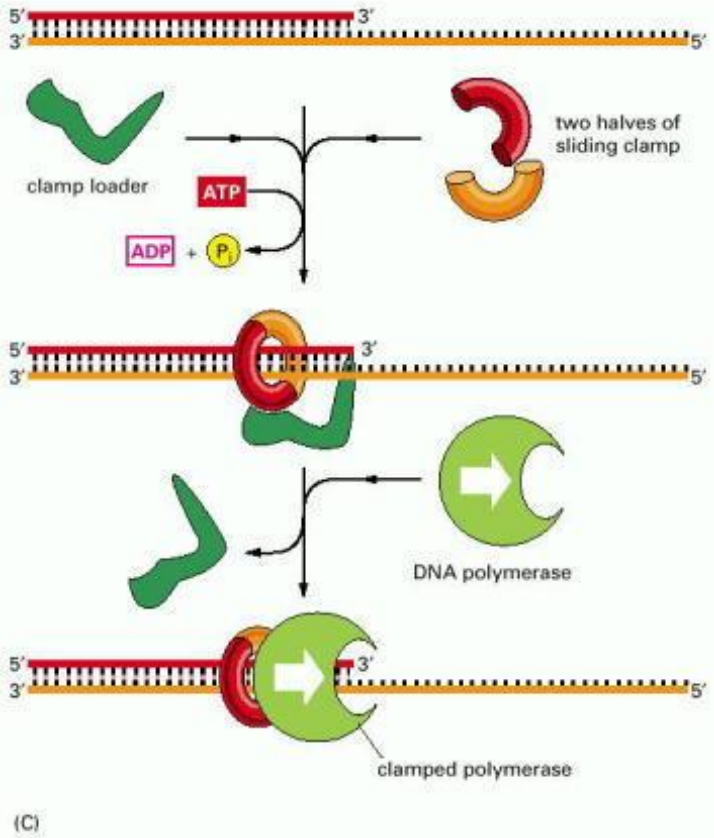
DNA synthesis by DNA polymerase



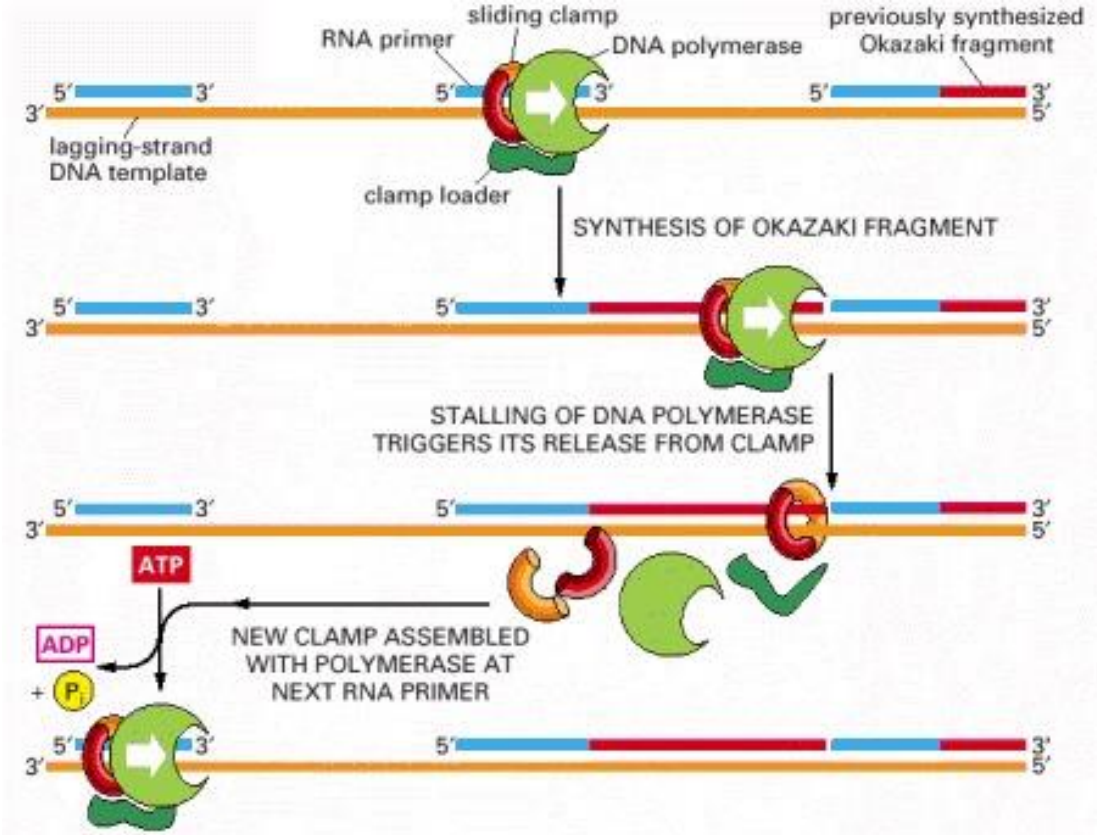
(A)



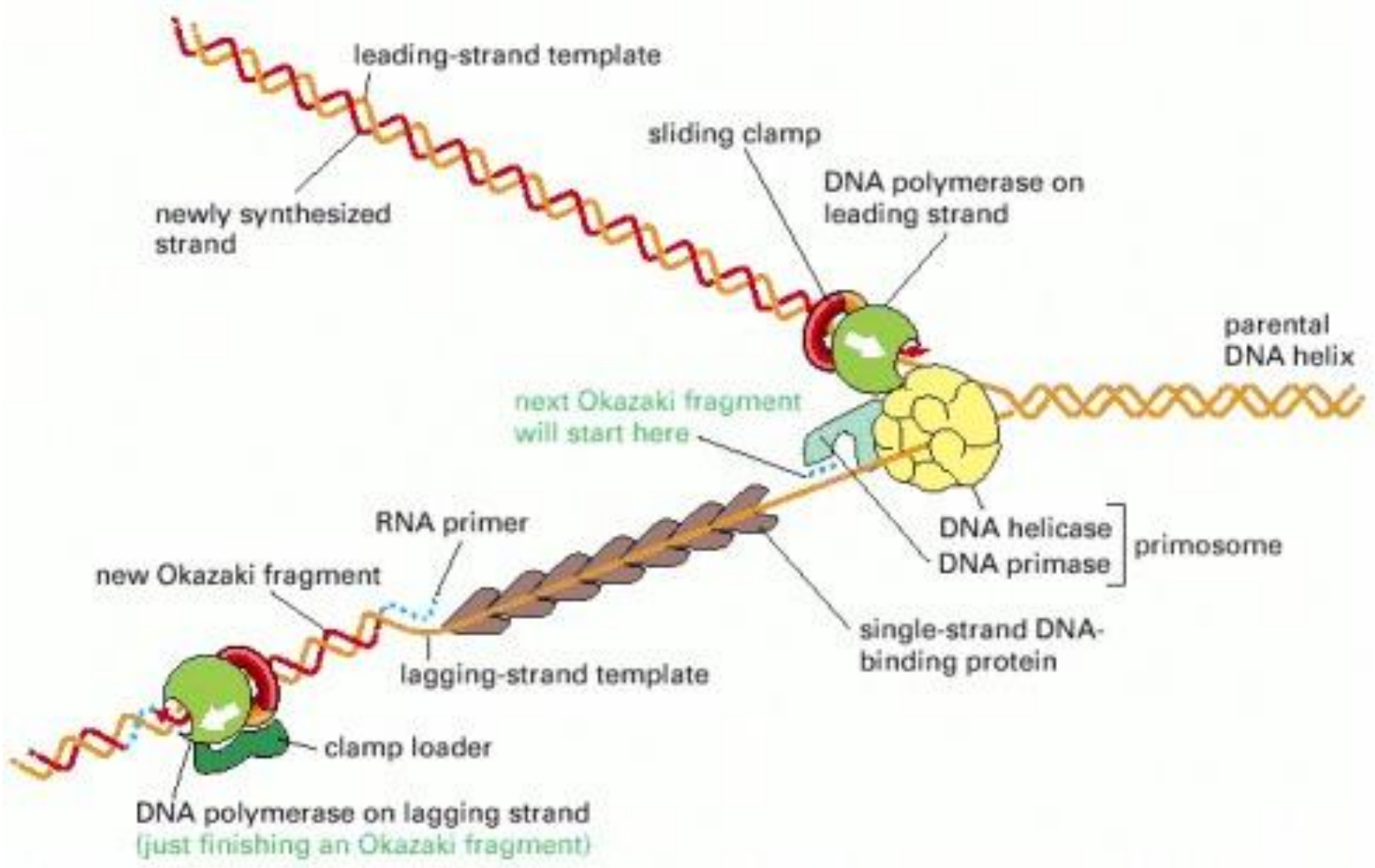
(B)



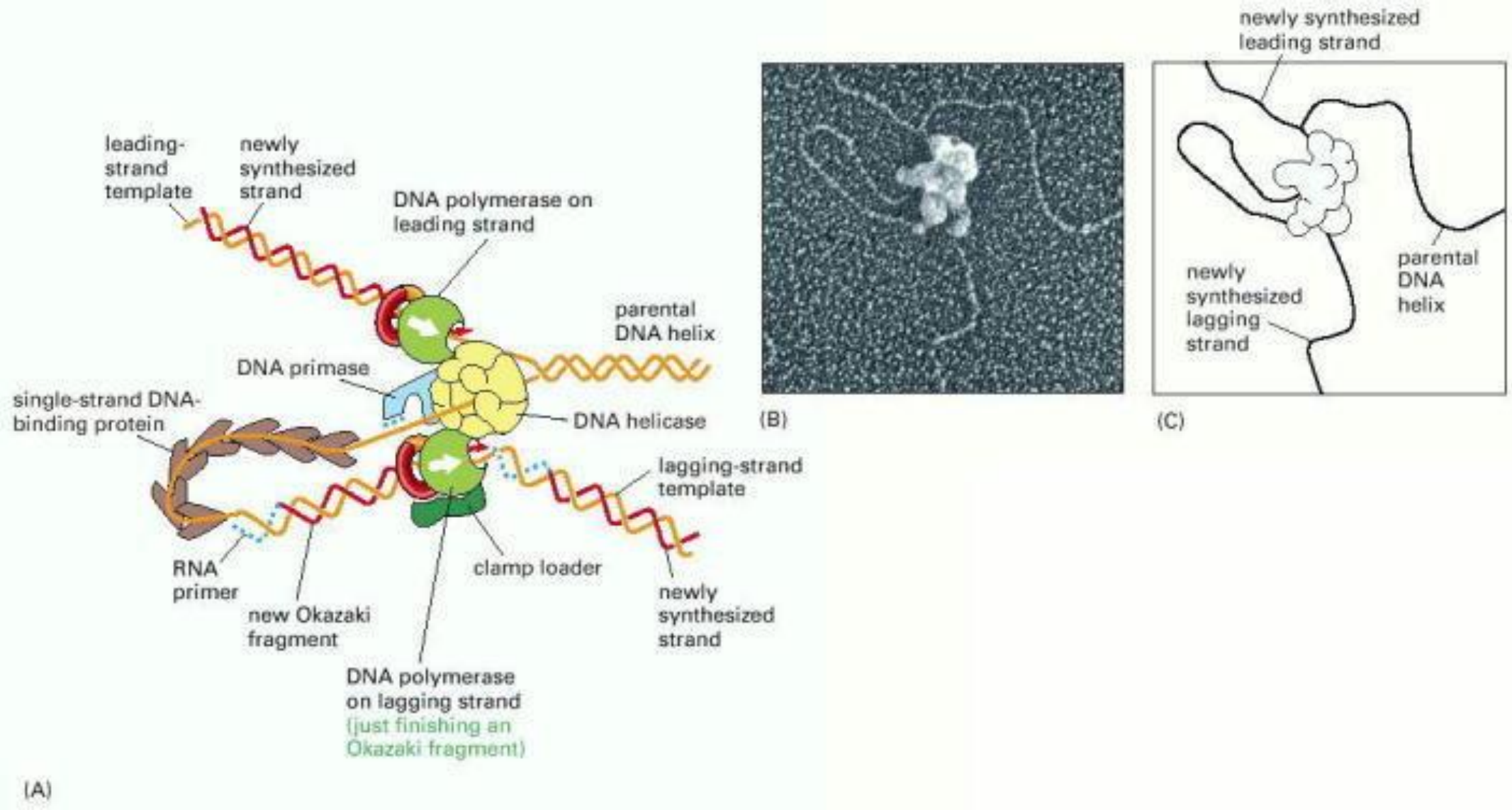
(C)



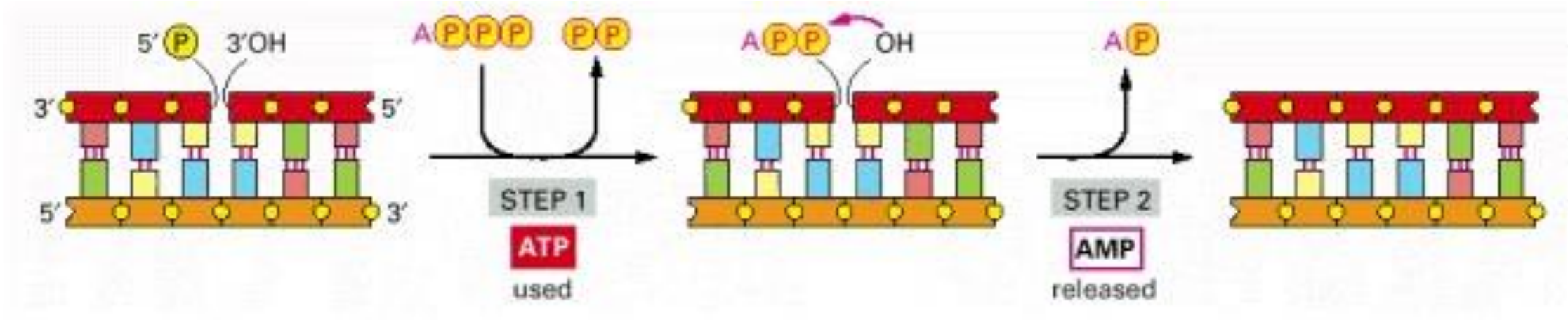
The complete replication fork

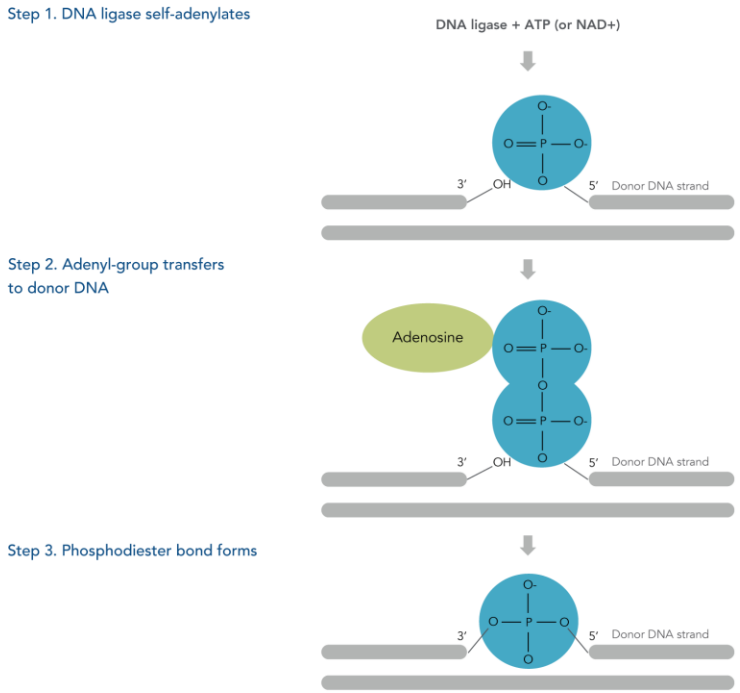


The complete replication fork

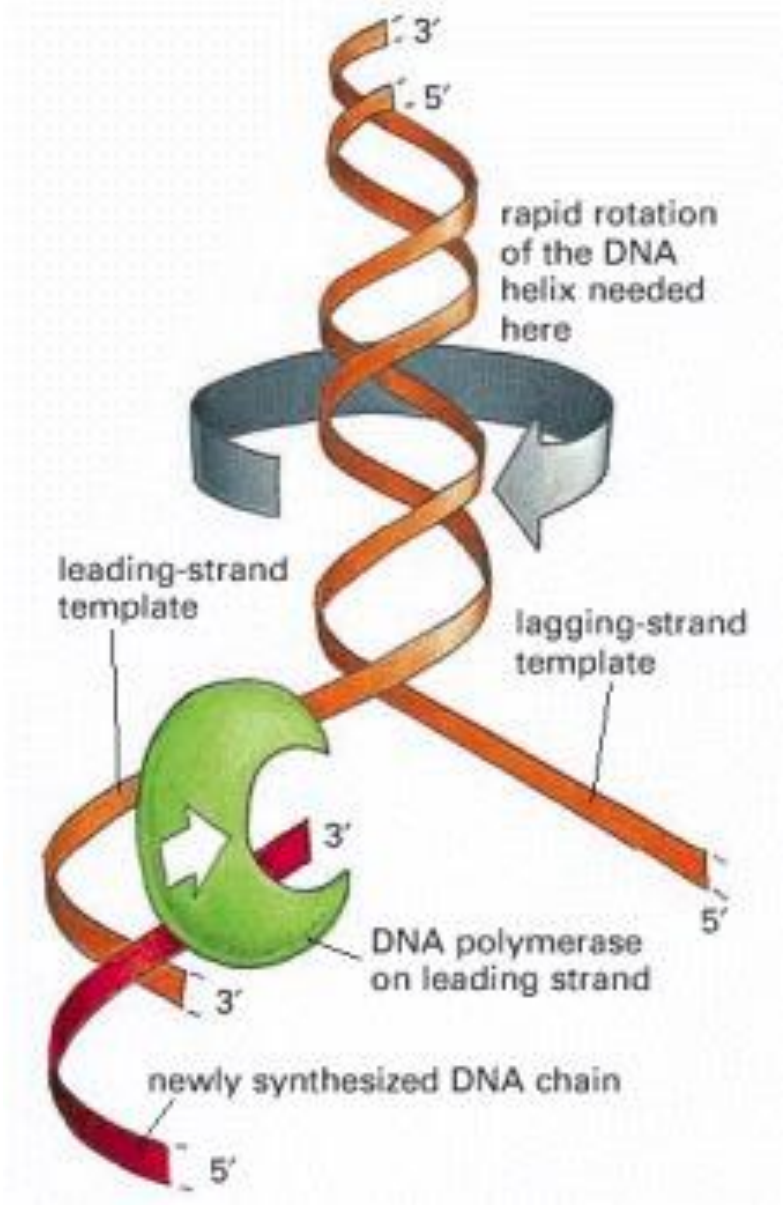


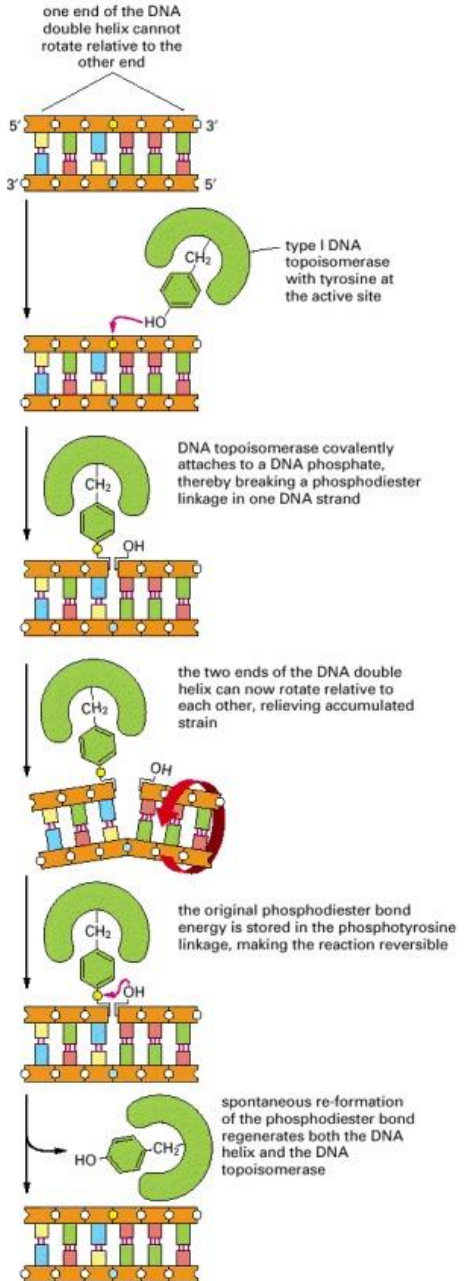
Strand ligation be DNA ligase



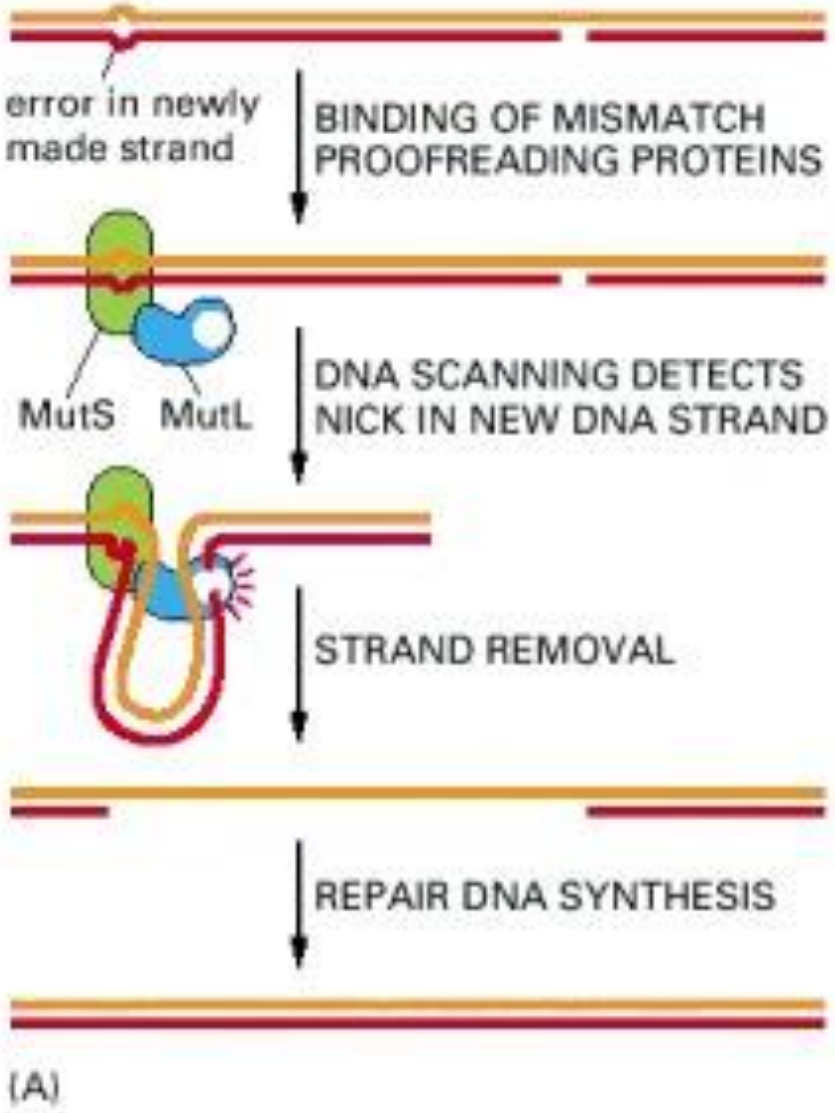


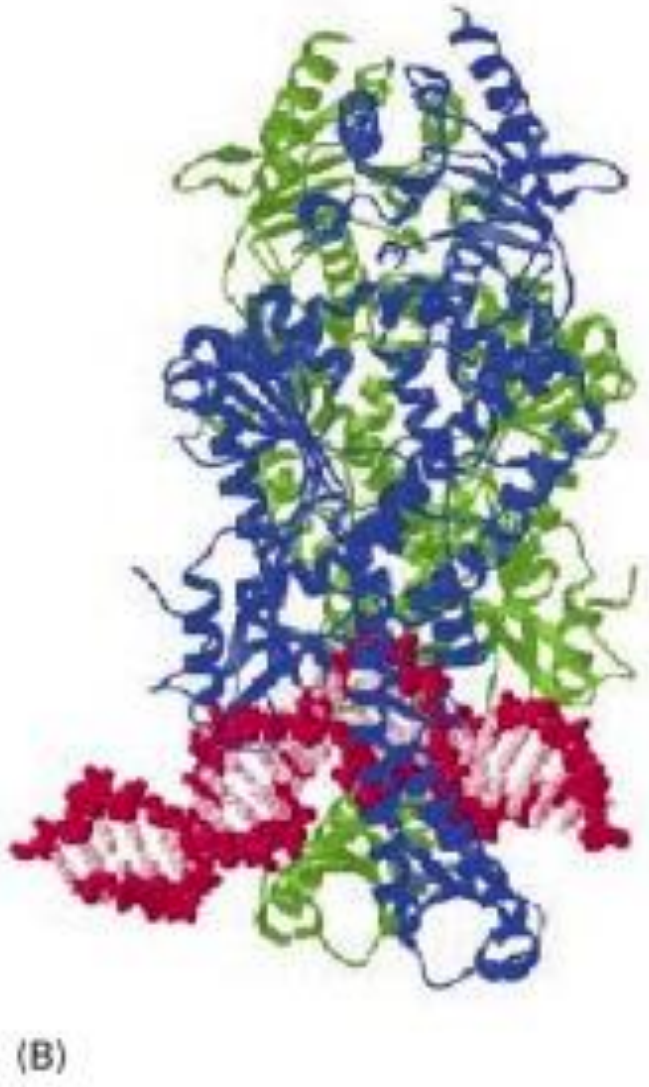
DNA topoisomerases prevent “tangling” of DNA





Strand-directed mismatch repair

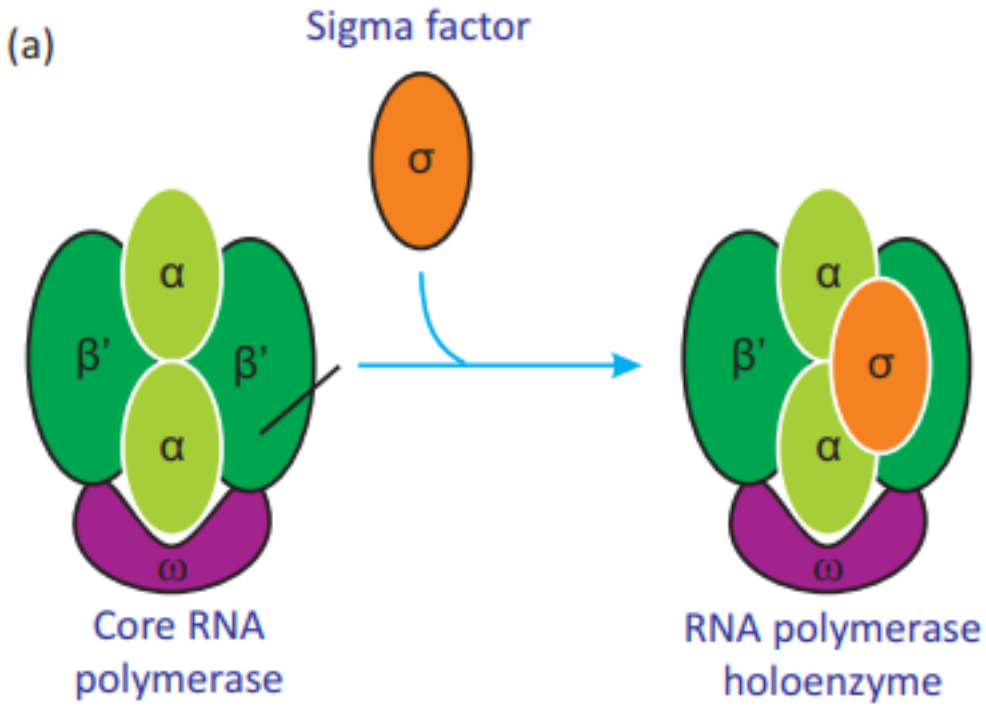




Transcription

Prokaryotic

- RNA is synthesized by a single RNA polymerase enzyme which contains multiple polypeptide subunits.
- In *E. coli*, the RNA polymerase has subunits: two α, one β, one β’ and one ω and σ subunit (α2ββ’ωσ). This complete enzyme is called as the **holoenzyme**.
- The **σ subunit may dissociate from the other subunits to leave a form known as the core enzyme.**

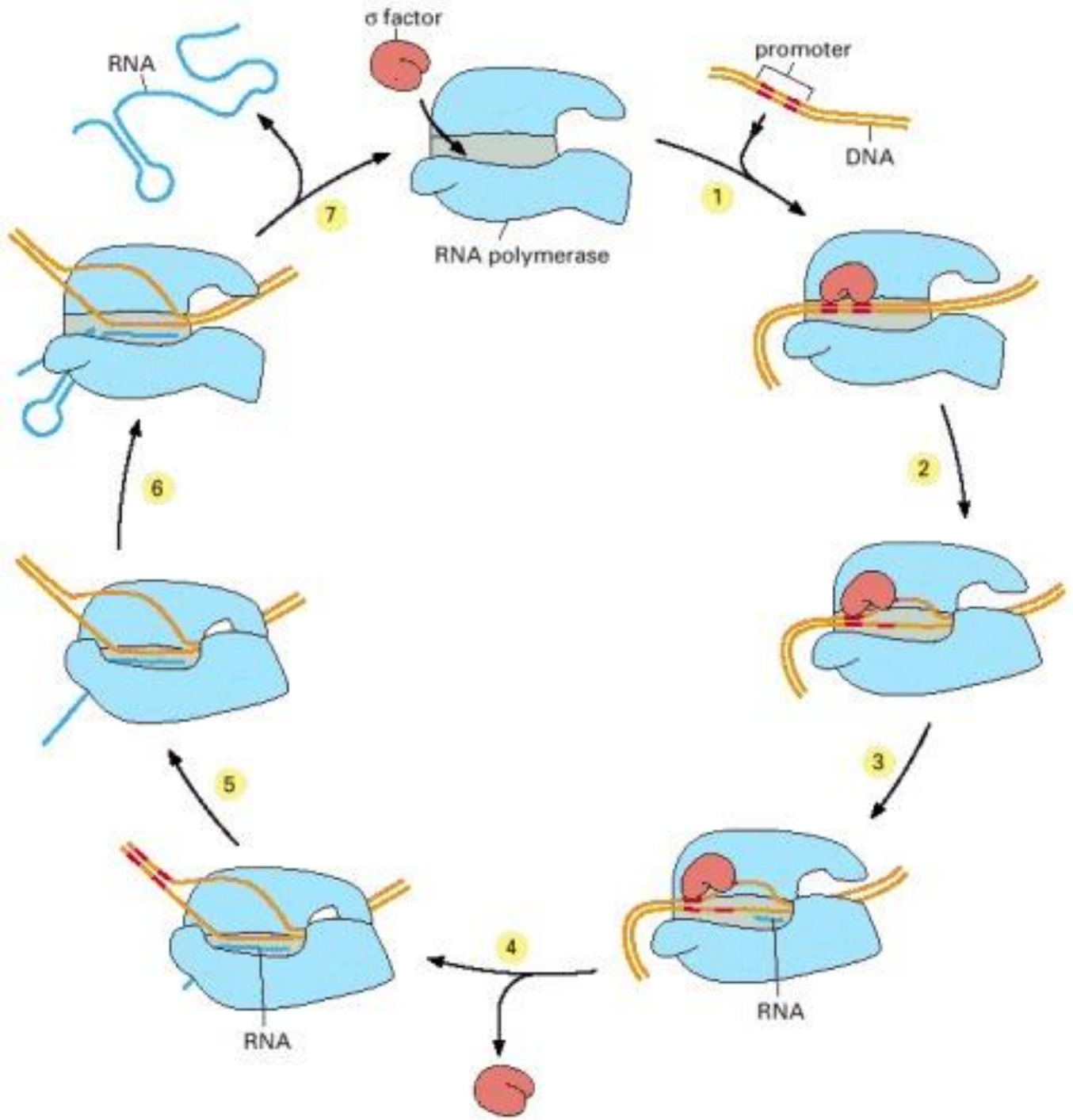


Subunit	Size	#/Molecule	Function
α	36.5 kD	2	chain initiation and interaction with regulatory proteins
β	151 kD	1	chain initiation and elongation
β'	155 kD	1	DNA binding
σ	70 kD	1	promoter recognition

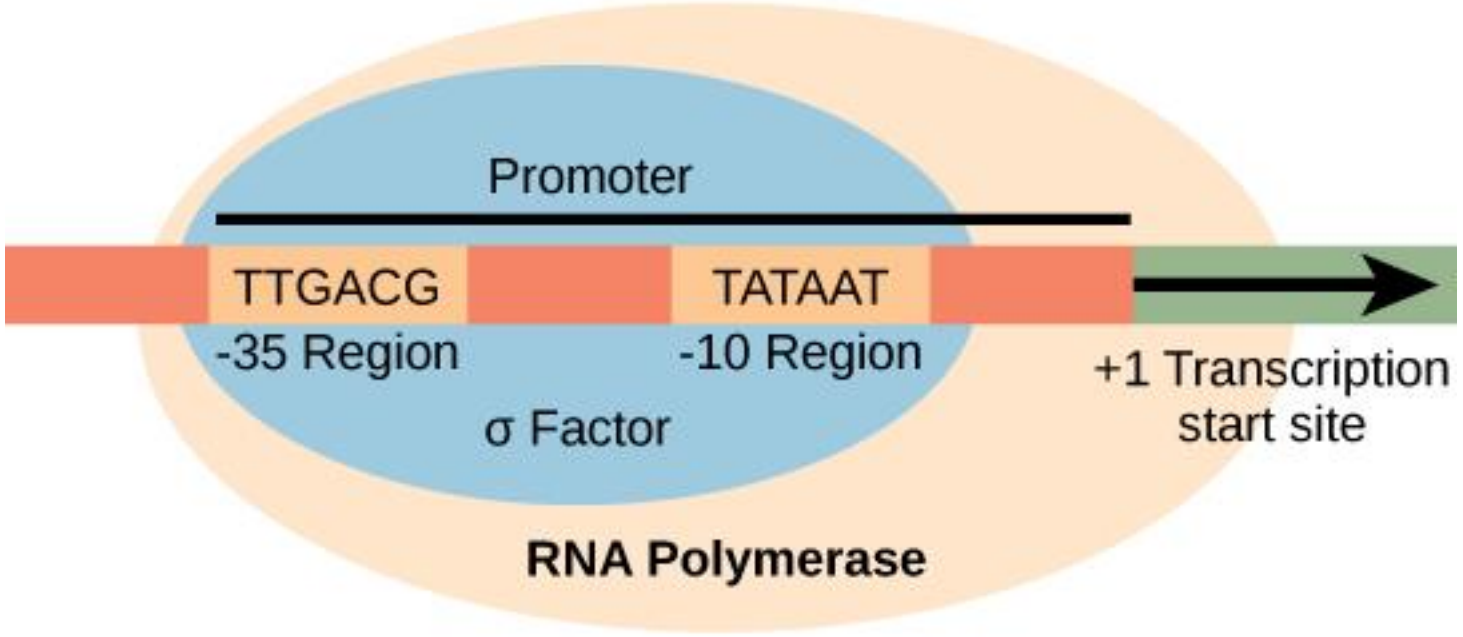
Transcription

Major steps:

- **Initiation (1-3)**
- **Elongation (4-5)**
- **Termination (6-7)**



Initiation



- The holoenzyme binds to a promoter region about 40–60 bp in size and then initiates transcription a short distance downstream (i.e. 3 to the promoter).
- Within the promoter lie two 6 base pair sequences that are particularly important for promoter function.
- They are highly conserved between species.
- Using the convention of calling the first nucleotide of a transcribed sequence as +1, these two promoter elements lie at positions –10 and –35, that is about 10 and 35 bp, respectively, upstream of where transcription will begin.
- The –10 sequence has the consensus Because this element was discovered by Pribnow, it is also known as the Pribnow box. It is an important recognition site that interacts with the σ factor of RNA polymerase.
- The –35 sequence has the consensus **TTGACA** and is important in DNA unwinding during transcriptional initiation.
- RNA polymerase does not need a primer to begin transcription; having bound to the promoter site, the RNA polymerase begins transcription directly.

E. coli sigma factors

Sigma factors recognize promoters by consensus sequences

Gene	Factor	–35 Sequence	Separation	–10 Sequence
rpoD	σ^{70}	TTGACA	16–18 bp	TATAAT
rpoH	σ^{32}	CCCTTGAA	13–15 bp	CCCGATNT
rpoN	σ^{54}	CTGGNA	6 bp	TTGCA
fliA	σ^{28} (σ^{F})	CTAAA	15 bp	GCCGATAA
sigH	σ^{H}	AGGANPuPu	11–12 bp	GCTGAATCA

FIGURE 11.34 *E. coli* sigma factors recognize promoters with different consensus sequences.

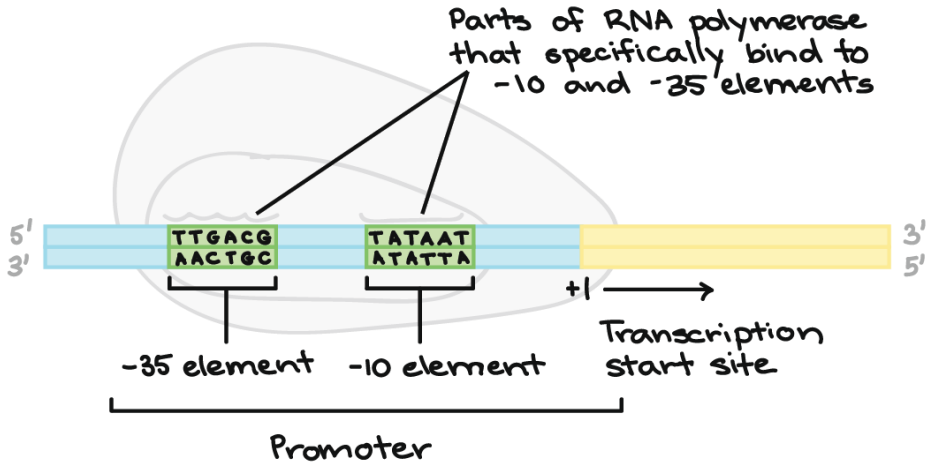
***Sigma factors in* Escherichia coli**

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

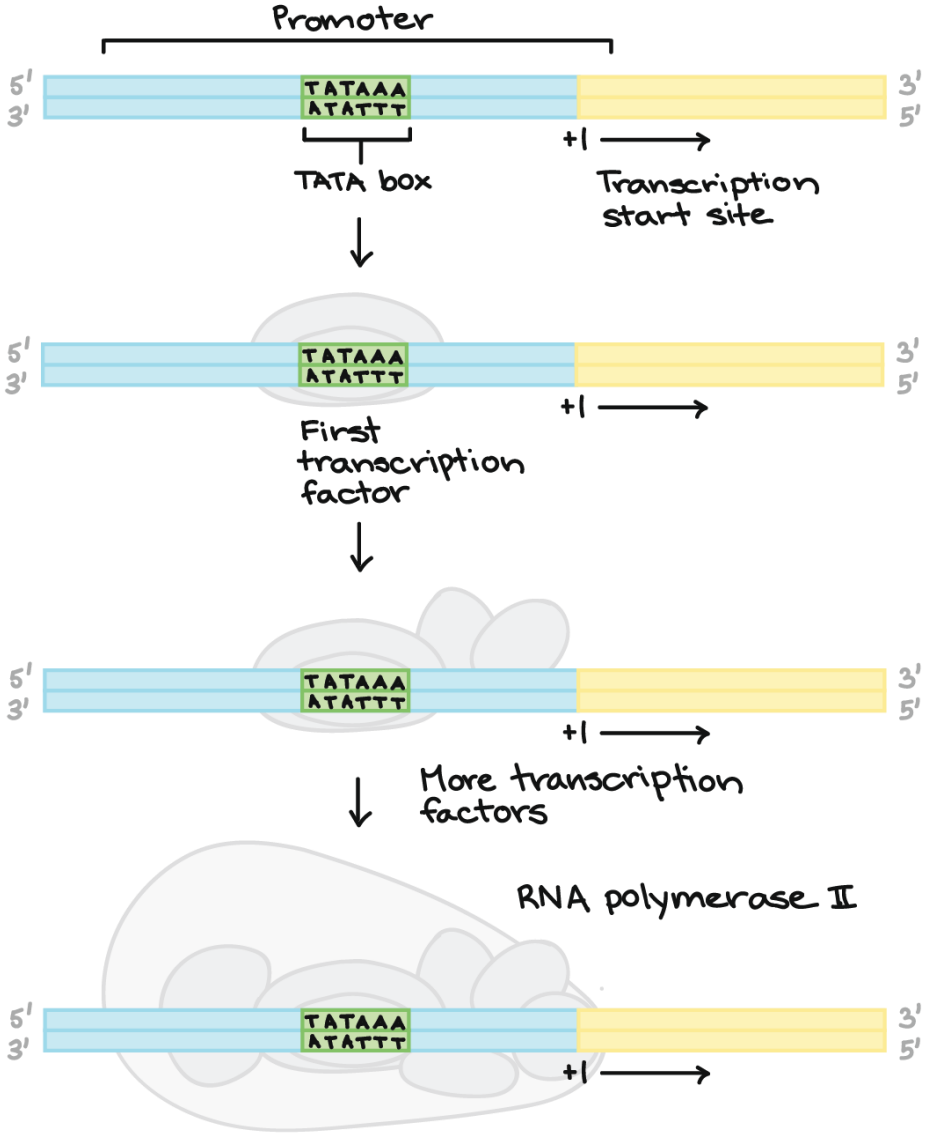
[a]Superscript number indicates size of protein in kilodaltons.

Promoters

Prokaryotic



Eukaryotic



The promoter of bacteria consists of 3 – 4 consensus sequences:

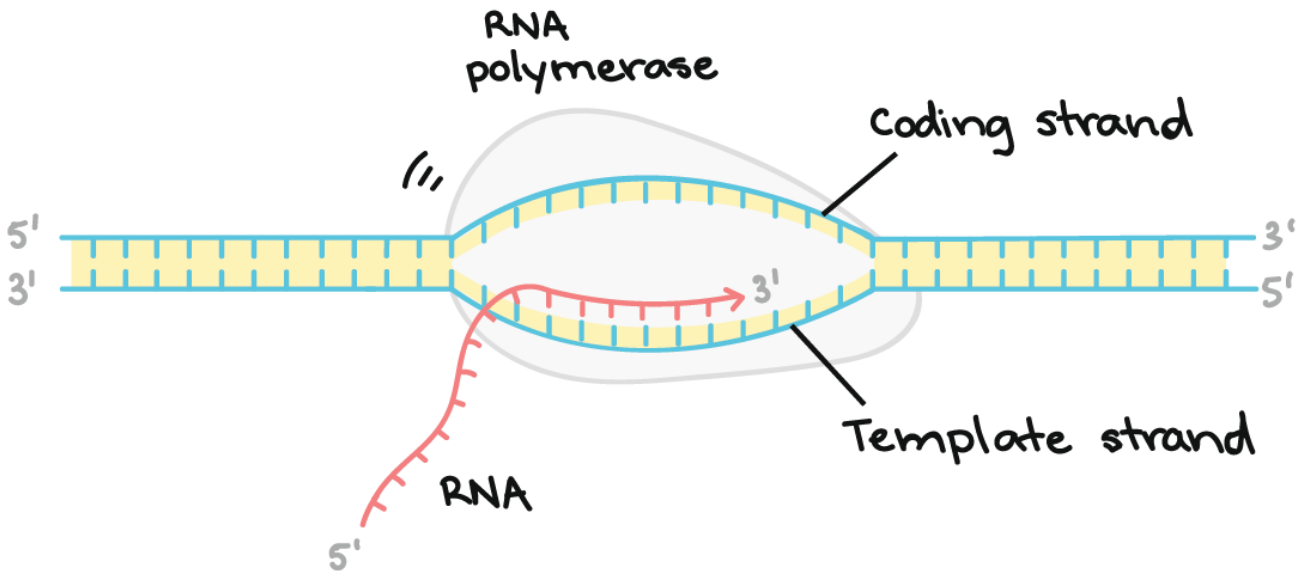
(I) Pribnow box (–10 sequence)

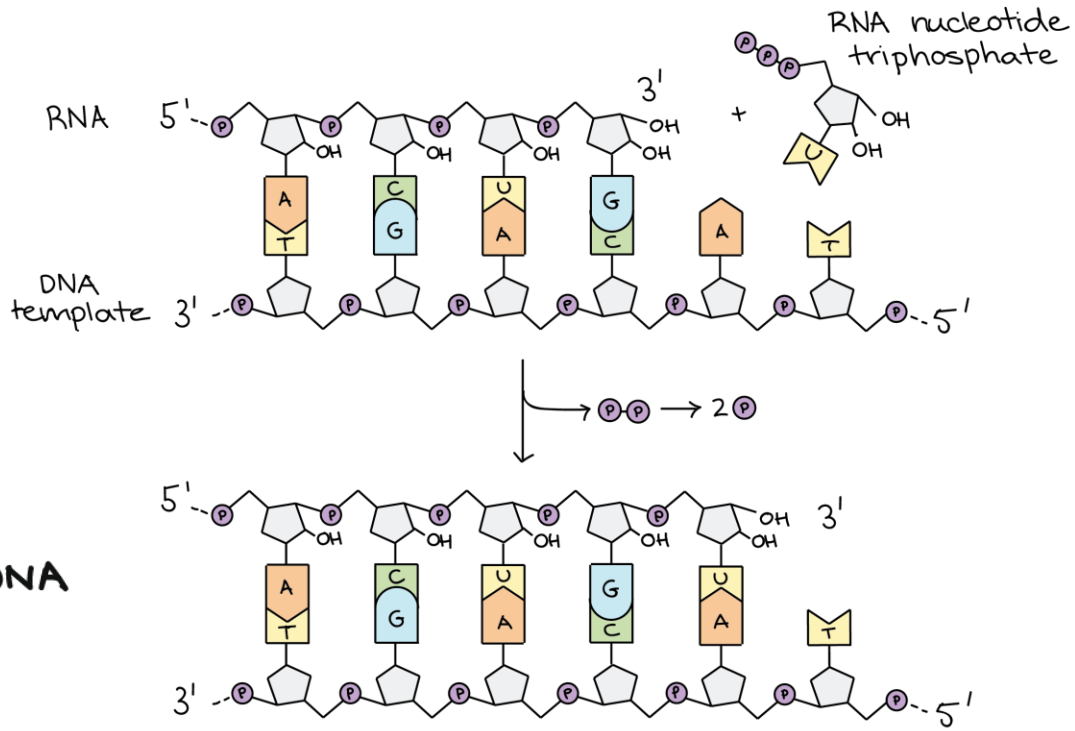
(ii) Recognition box (–35 sequence)

(iii) UP element.

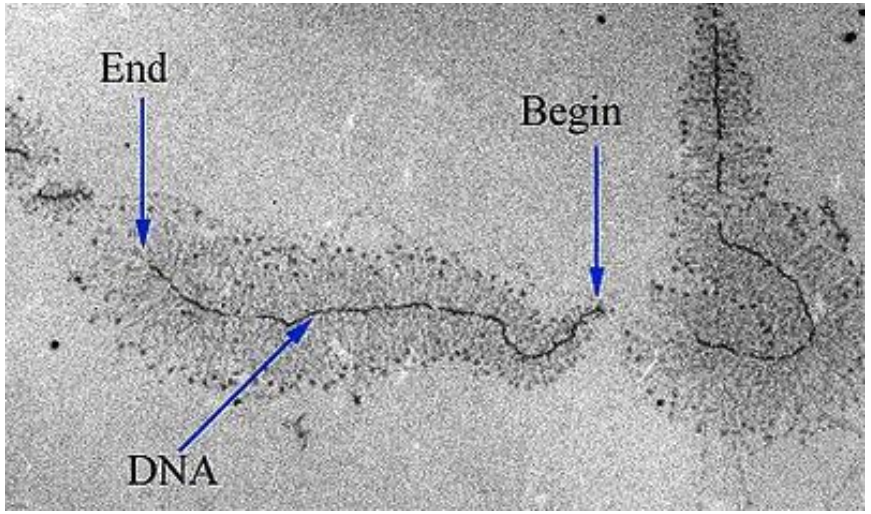
(iv) –10 extended box

Elongation



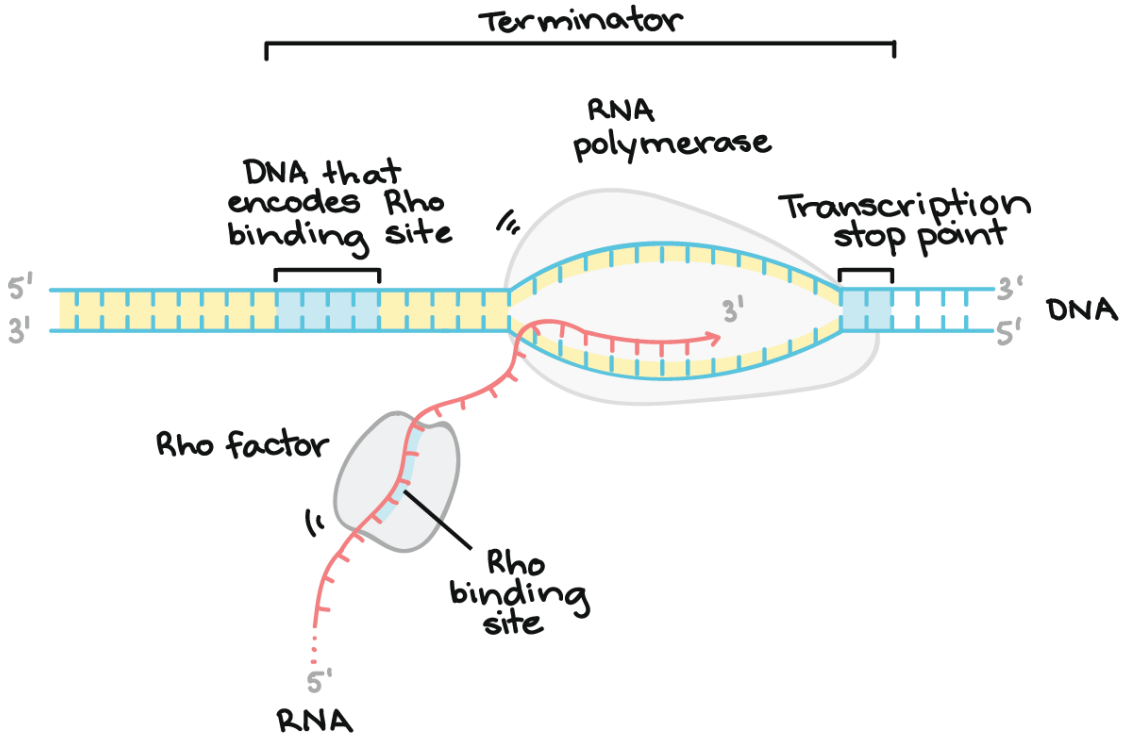


Coding strand	5'	A T G A T C T C G T A A 3'
RNA	5'	A U G A U C 3'
Template strand	3'	T A C T A G A G C A T T 5'



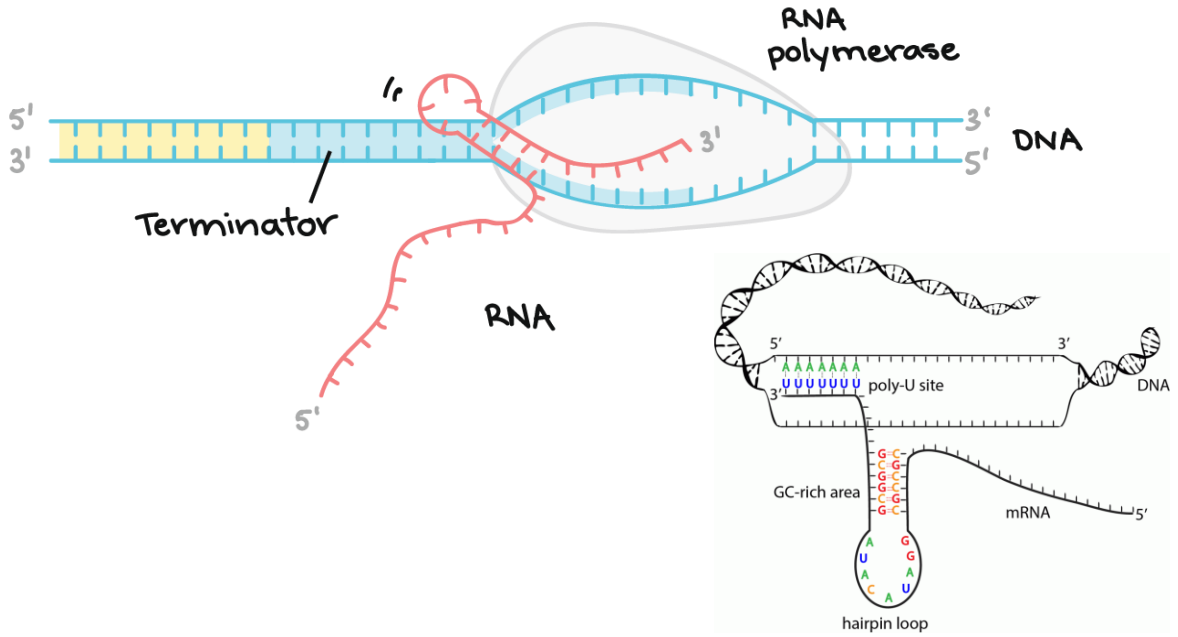
Termination

Rho-dependent



In **Rho-dependent termination**, the RNA contains a binding site for a protein called Rho factor. Rho factor binds to this sequence and starts "climbing" up the transcript towards RNA polymerase.
When it catches up with the polymerase at the transcription bubble, Rho pulls the RNA transcript and the template DNA strand apart, releasing the RNA molecule and ending transcription. Another sequence found later in the DNA, called the transcription stop point, causes RNA polymerase to pause and thus helps Rho catch up.

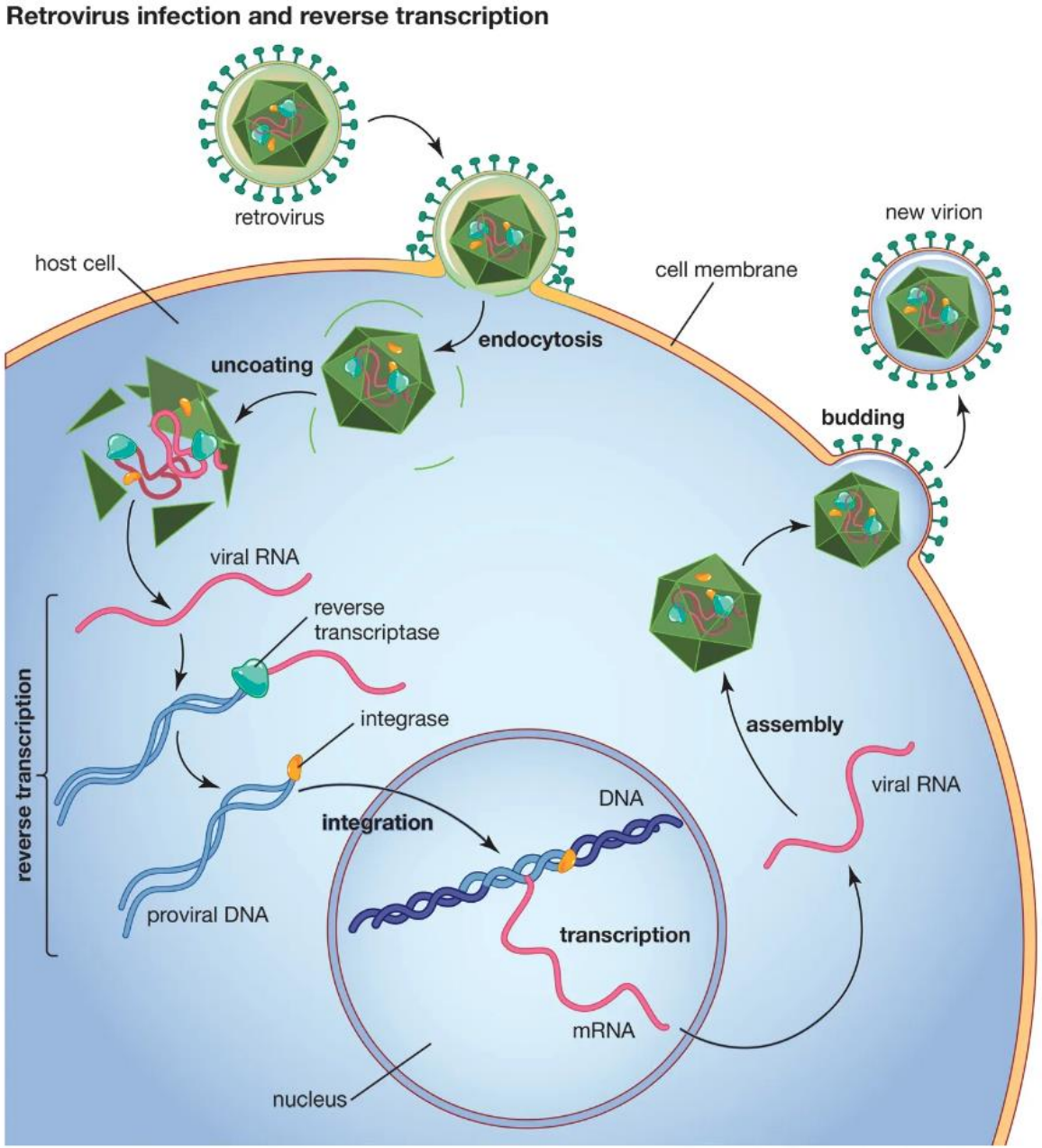
Rho-independent

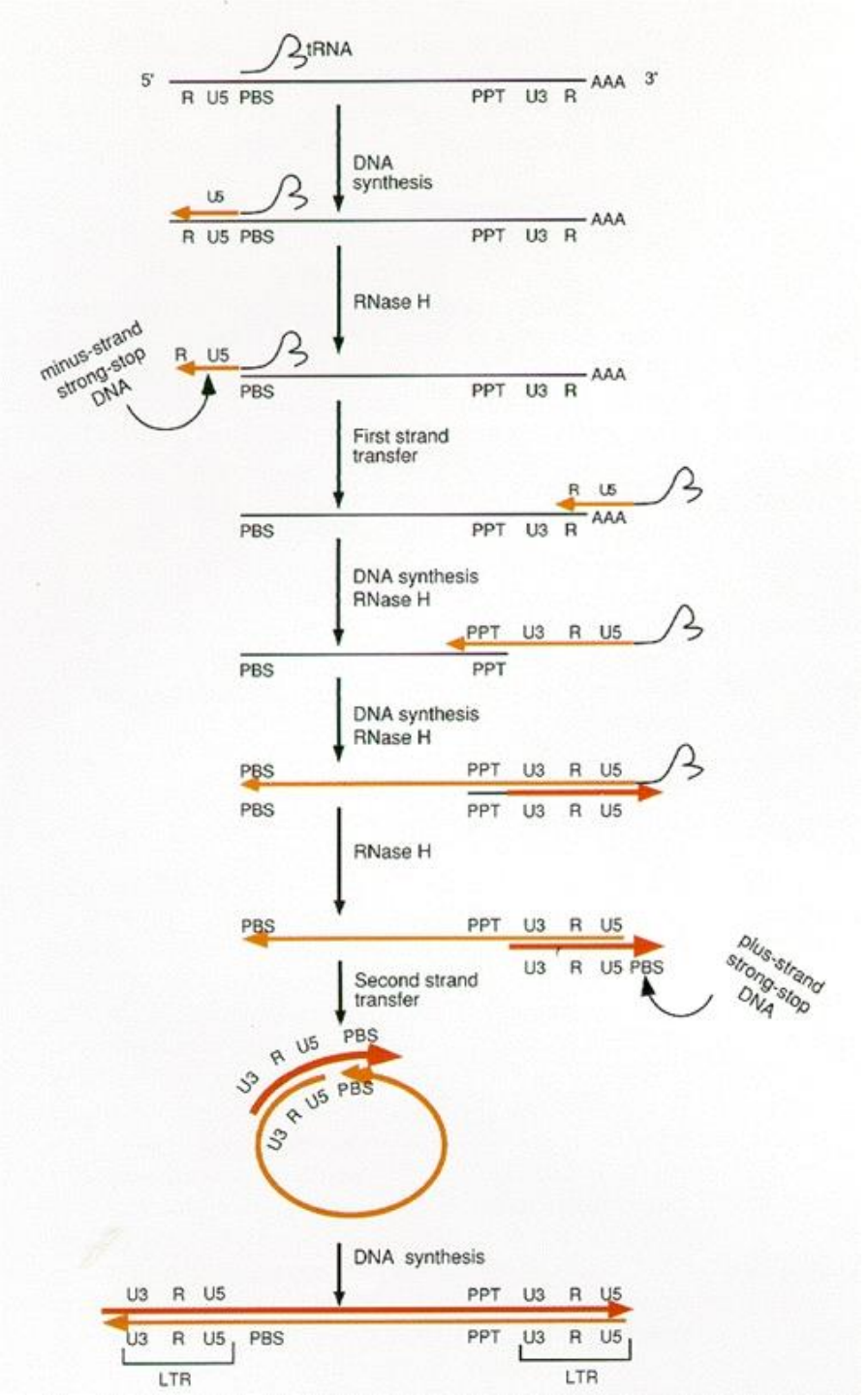


Rho-independent termination depends on specific sequences in the DNA template strand. As the RNA polymerase approaches the end of the gene being transcribed, it hits a region rich in C and G nucleotides. The RNA transcribed from this region folds back on itself, and the complementary C and G nucleotides bind together. The result is a stable hairpin that causes the polymerase to stall.
In a terminator, the hairpin is followed by a stretch of U nucleotides in the RNA, which match up with A nucleotides in the template DNA. The complementary U-A region of the RNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, produces enough instability for the enzyme to fall off and liberate the new RNA transcript.

Reverse-Transcription

Reverse-transcription





RT-PCR

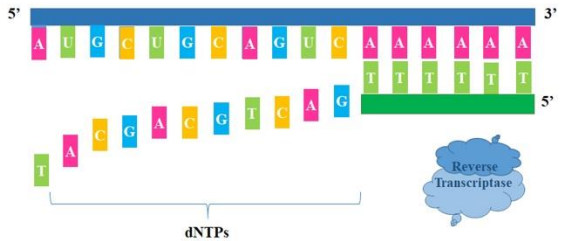
4.8 Reverse transcription polymerase chain reaction (RT-PCR)

In RT-PCR, The RNA population is converted to cDNA by reverse transcription (RT), and then the cDNA is amplified by the polymerase chain reaction .The cDNA amplification step provides opportunities to further study the original RNA species, even when they are limited in amount or expressed in low abundance. Common applications of RT-PCR include detection of expressed genes, examination of transcript variants, and generation of cDNA templates for cloning and sequencing.

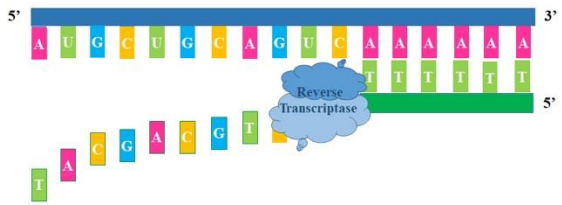
1 a. **RNA**
RNA consist of Start codon AUG and ends with poly A tail

b. **Oligo dT Primer**
Oligo dT Primer is binding to RNA poly A tail

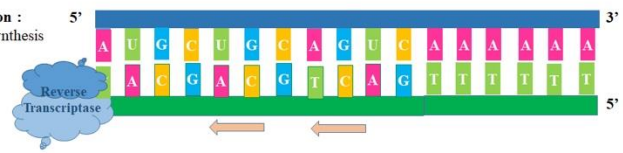
c. **Reverse Transcriptase and dNTPs**



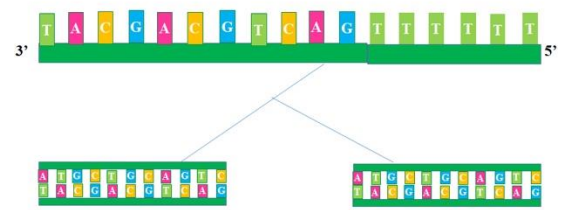
d. **Reverse Transcriptase** is an enzymes binds to oligo dT primer and synthesises the cDNA by adding dNTPs



e. **RNA hybrid formation :**
First - strand cDNA synthesis



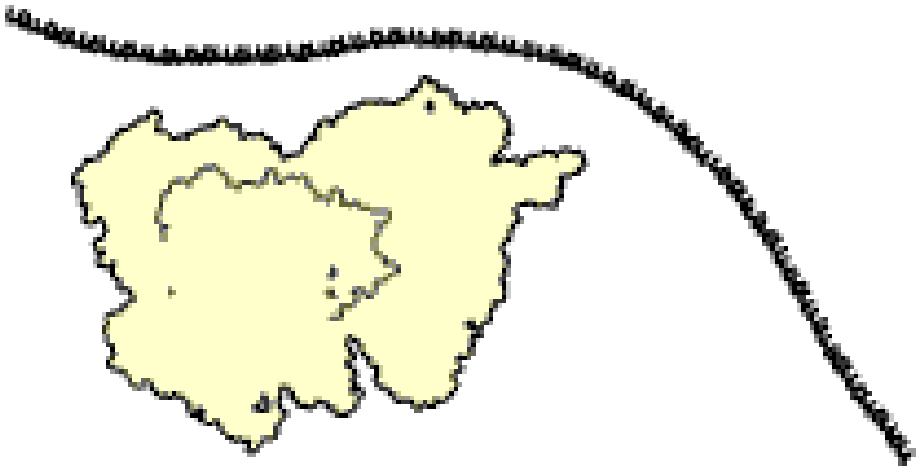
f. **complimentary DNA**

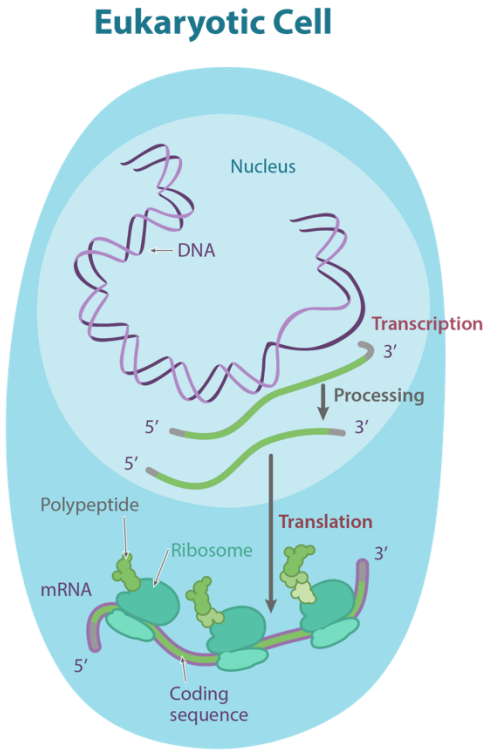


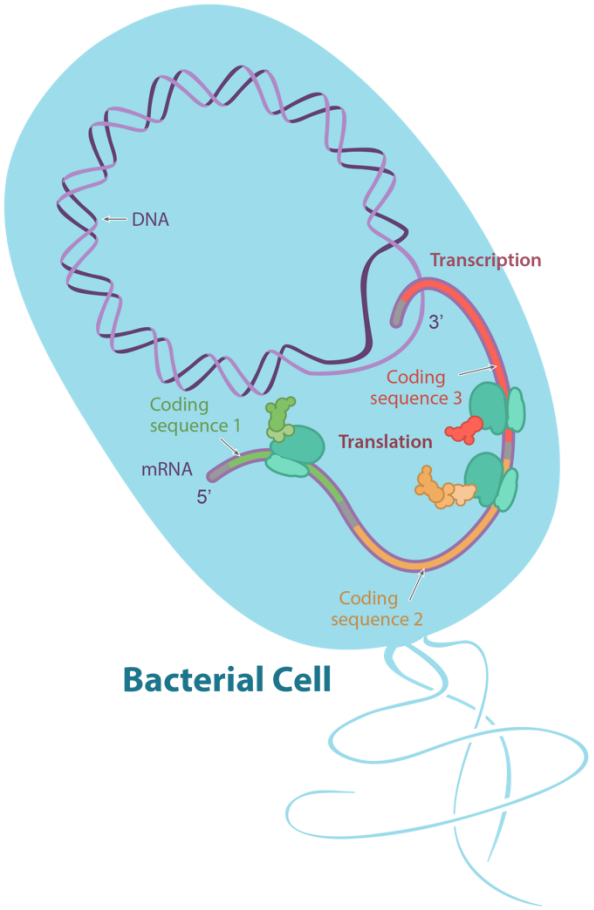
2 a. **Amplification of cDNA** with Specific Primers and Taq Polymerase

Translation

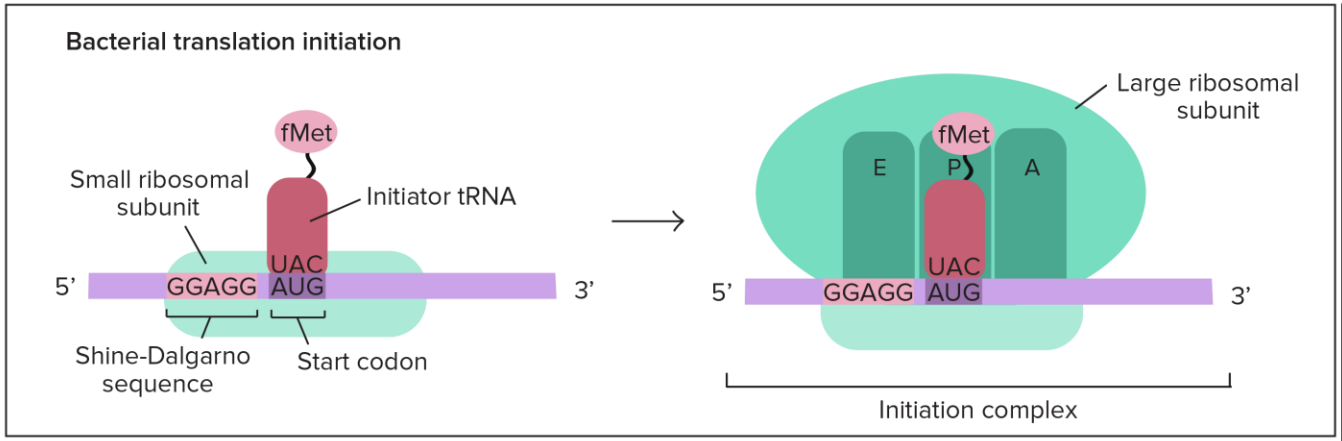
Translation

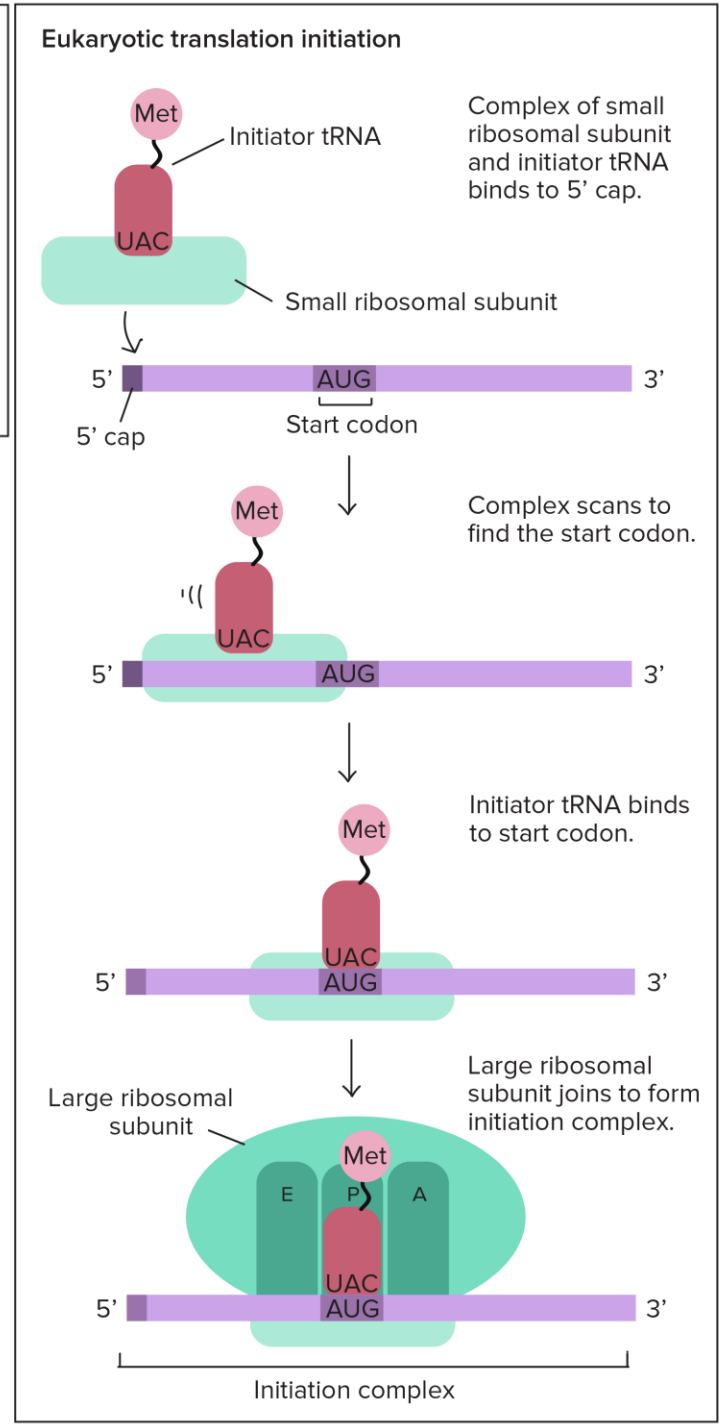






Initiation





In order for translation to start, we need a few key ingredients:

- A ribosome (which comes in two pieces, large and small)
- An mRNA with instructions for the protein we'll build
- An "initiator" tRNA carrying the first amino acid in the protein, which is almost always methionine (Met)

During initiation, these pieces must come together in just the right way. Together, they form the **initiation complex**, the molecular setup needed to start making a new protein.

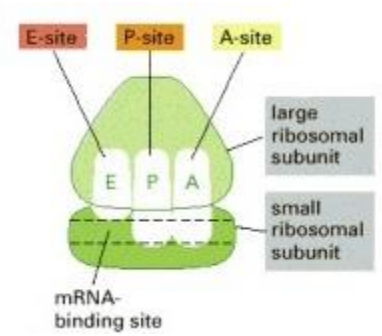
Eukaryotes: first, the tRNA carrying methionine attaches to the small ribosomal subunit. Together, they bind to the 5' end of the mRNA by recognizing the 5' GTP cap (added during processing in the nucleus). Then, they "walk" along the mRNA in the 3' direction, stopping when they reach the start codon (often, but not always, the first AUG).

Prokaryotes: the small ribosomal subunit doesn't start at the 5' end of the mRNA and travel toward the 3' end. Instead, it attaches directly to certain sequences in the mRNA. These **Shine-Dalgarno** sequences come just before start codons and "point them out" to the ribosome.

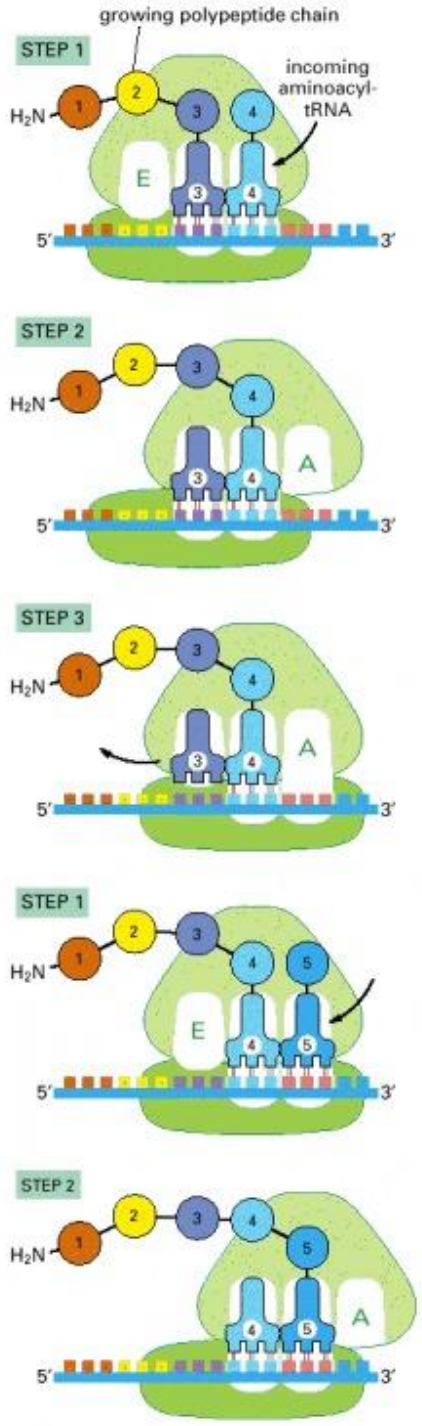
Elongation

An aminoacyl-tRNA molecule binds to a vacant A-site on the ribosome in step 1, a new peptide bond is formed in step 2, and the mRNA moves a distance of three nucleotides through the small-subunit chain in step 3, ejecting the spent tRNA molecule and "resetting" the ribosome so that the next aminoacyl-tRNA molecule can bind. Although the figure shows a large movement of the small ribosome subunit relative to the large subunit, the conformational changes that actually take place in the ribosome during **translation** are more subtle. It is likely that they involve a series of small rearrangements within each subunit as well as several small shifts between the two subunits. As indicated, the mRNA is translated in the 5′-to-3′ direction, and the N-terminal end of a **protein** is made first, with each cycle adding one amino acid to the C-terminus of the polypeptide chain. The position at which the growing peptide chain is attached to a tRNA does not change during the elongation cycle: it is always linked to the tRNA present in the P site of the large subunit.

Synthesis rate: 20 AAs / s (in bacteria)



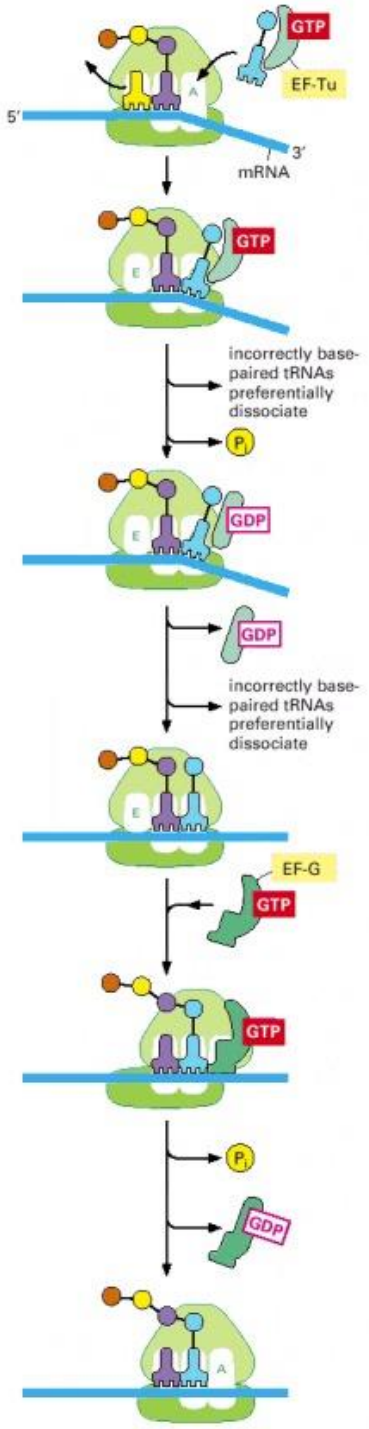
aminoacyl site (abbreviated A)
peptidyl site (abbreviated P)
exit site (abbreviated E)



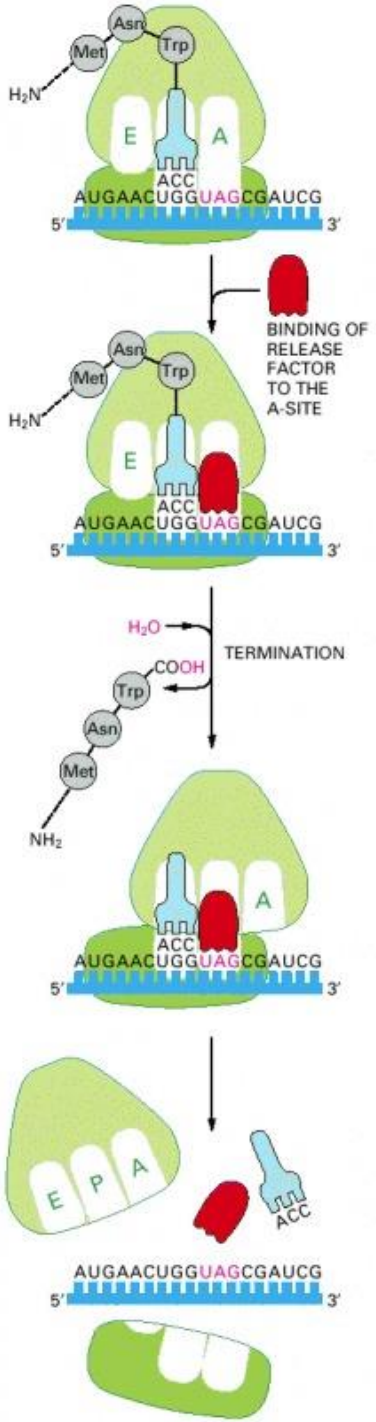
Elongation Factors

Elongation factors help move translation forwards.

EF-Tu delivers tRNAs to the ribosome.
EF-Tu is also thought to increase accuracy of translation.



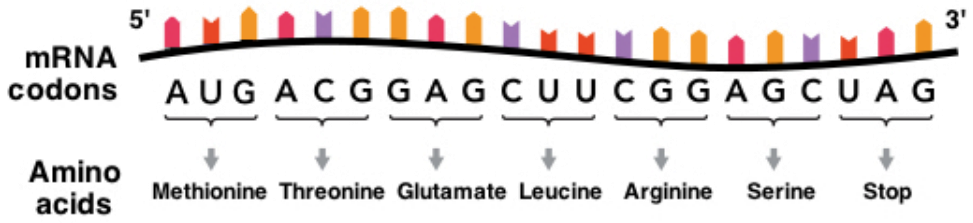
Termination

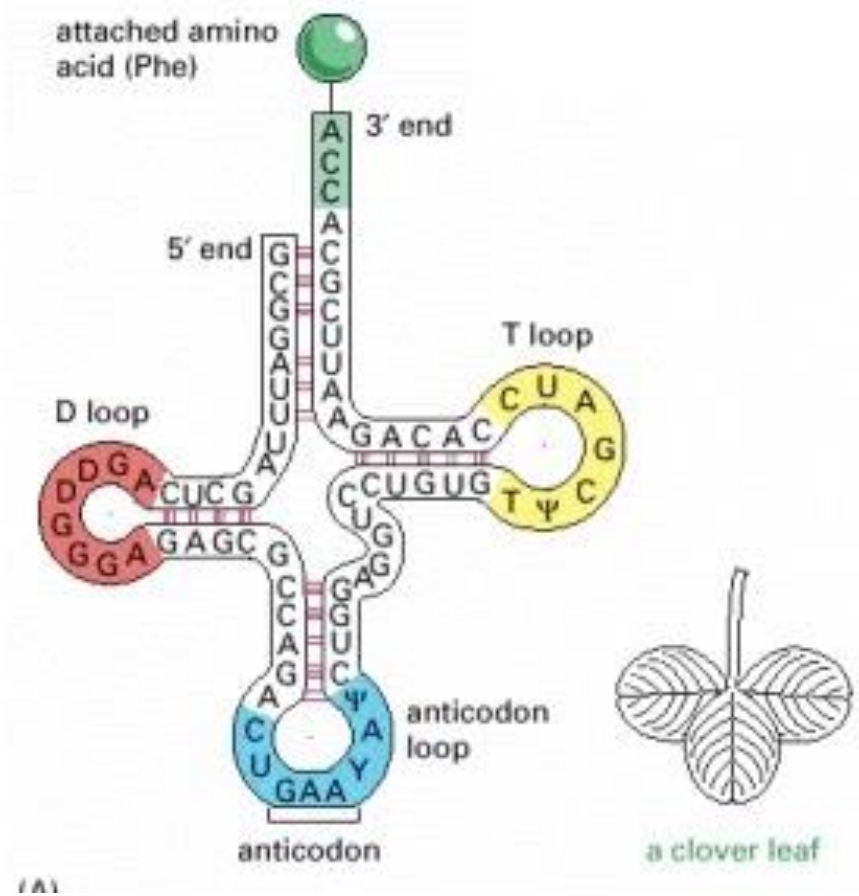


The Genetic Code

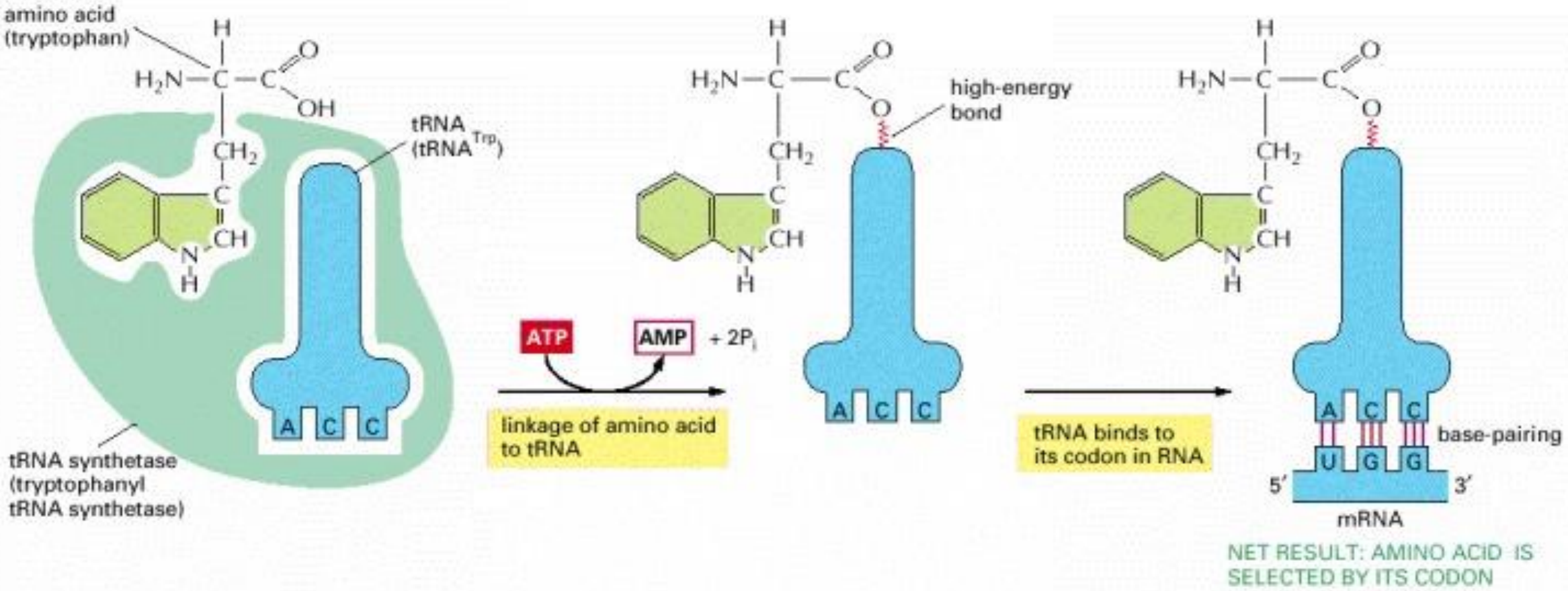
Second letter

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

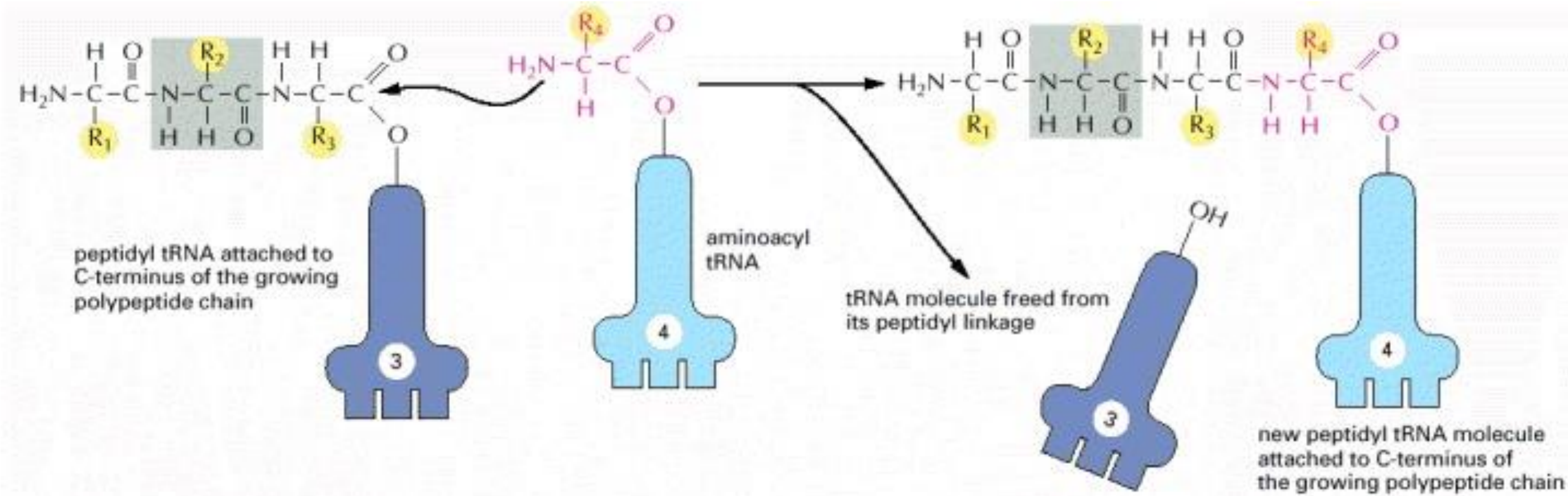




The Genetic Code (aaRS)



Elongation


peptidyl tRNA attached to C-terminus of the growing polypeptide chain
aminoacyl tRNA
tRNA molecule freed from its peptidyl linkage
new peptidyl tRNA molecule attached to C-terminus of the growing polypeptide chain
3
4
3
4
OH

Start and Stop Codons

Start Codon

AUG (ATG on DNA)

Codes for:
- Methionine (eukaryotes)
- Formylmethionine (bacteria, archea, mitochondria)

Stop Codons

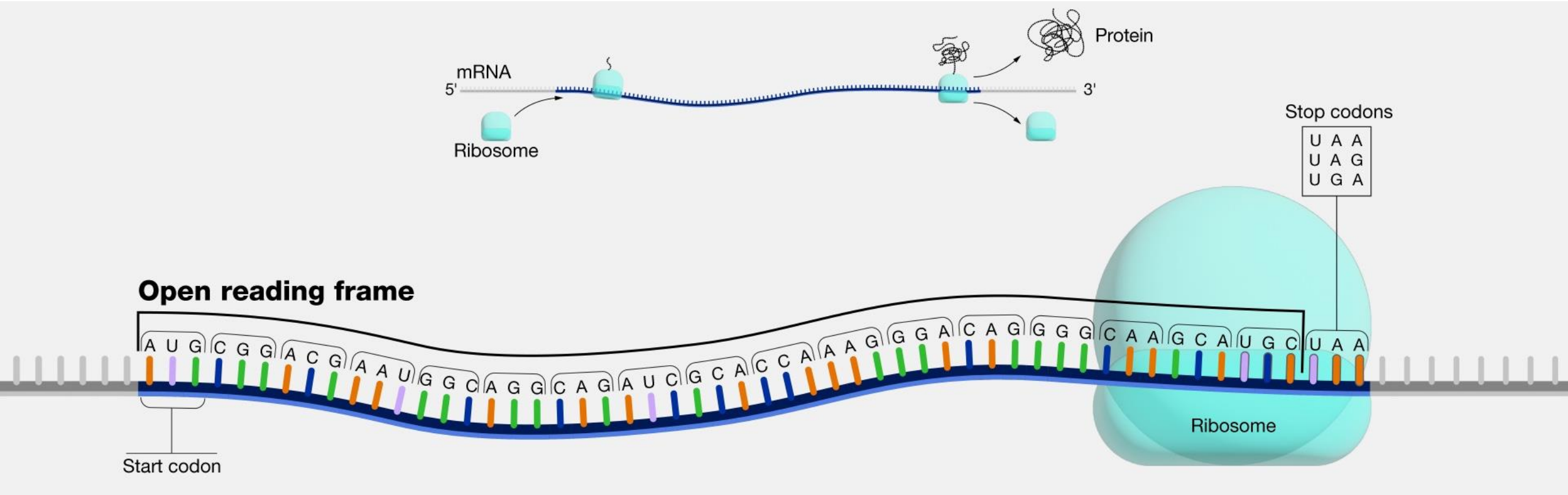
UAG: amber
UGA: opal
UAA: ochre

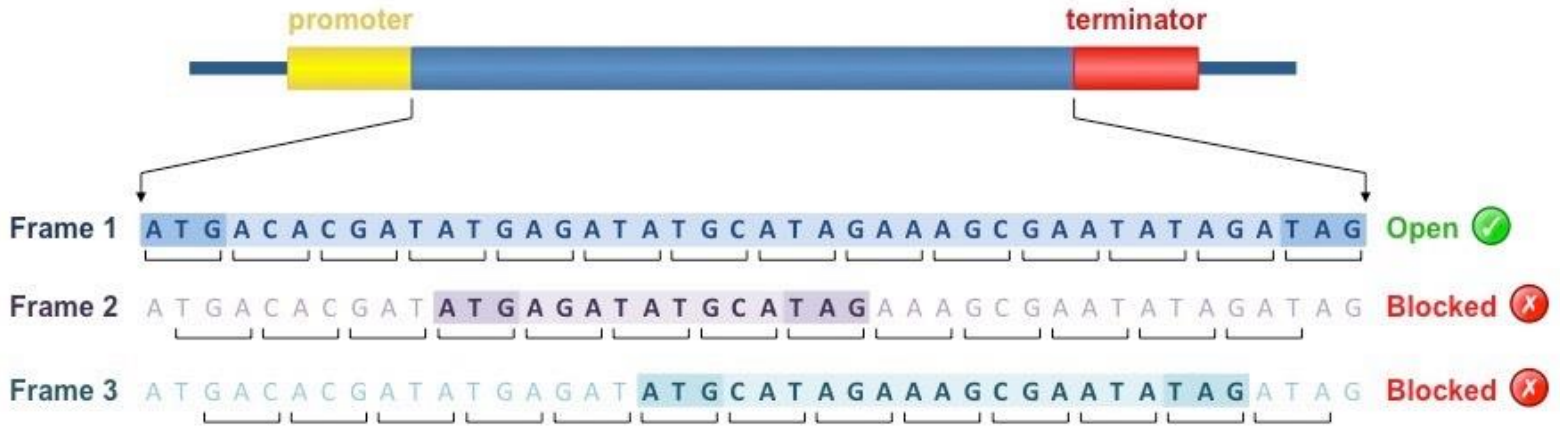
Not recognized by a tRNA but by proteins called release factors.

Bacterial release factors
Class 1:
RF1 -> UAA and UAG
RF2 -> UAA

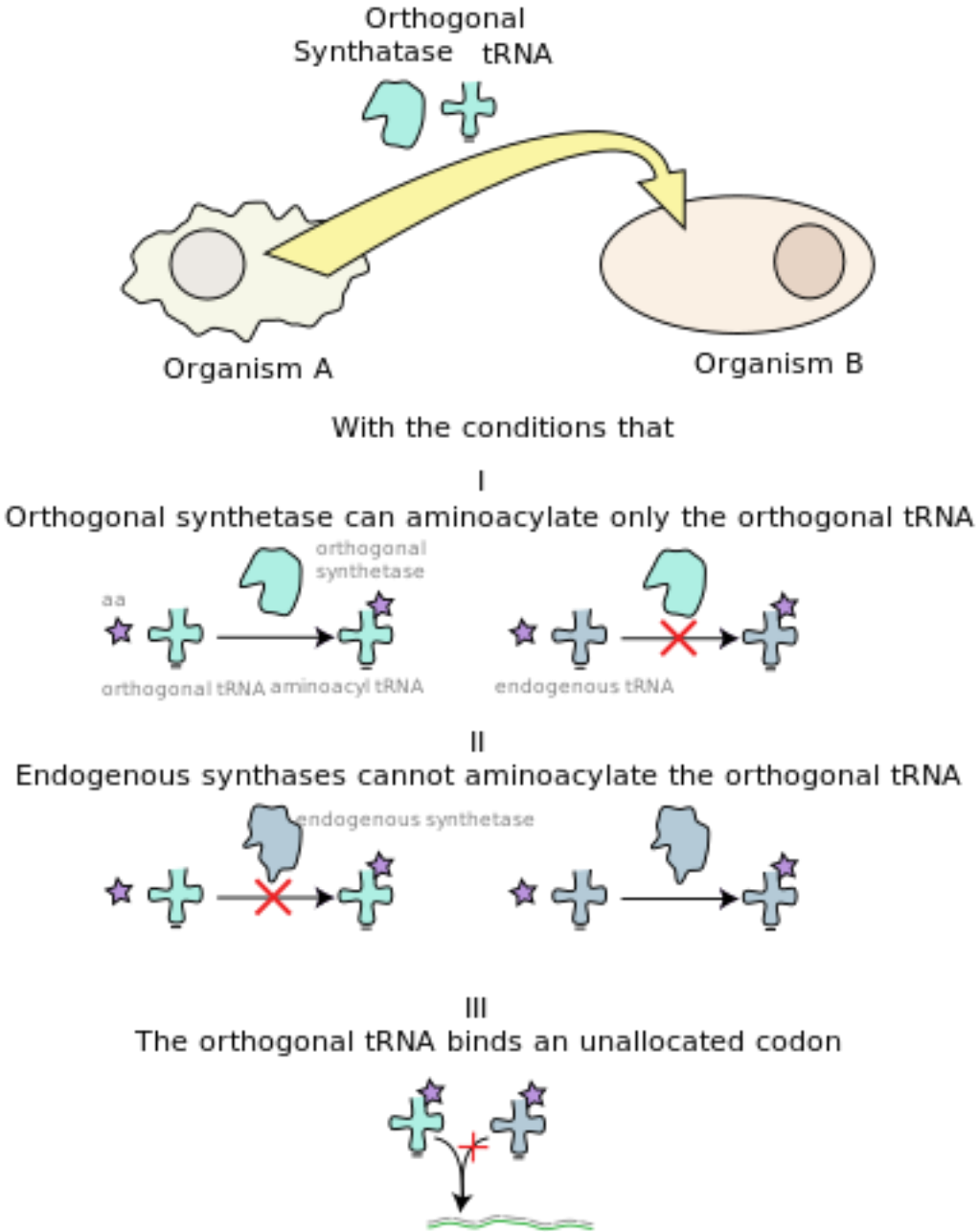
Class 2:
RF3 -> class 2 release factor: enhances activity of class 1

Reading Frame





Genetic Code Expansion



Stop Codons

UAG: amber
UGA: opal
UAA: ochre

Not recognized by a tRNA but by proteins called release factors.

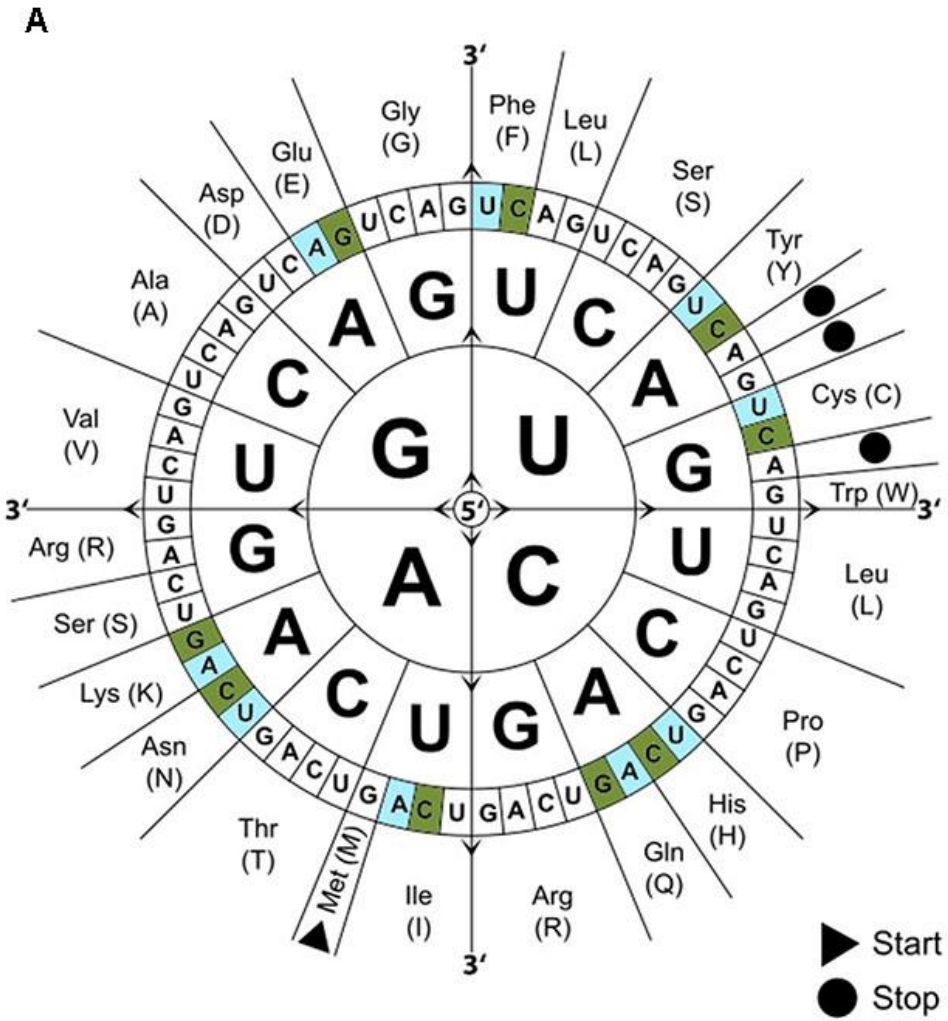
Bacterial release factors
Class 1:
RF1 -> UAA and UAG
RF2 -> UAA and UGA

The key prerequisites to expand the genetic code are:

- the non-standard amino acid to encode,
- an unused codon to adopt,
- a tRNA that recognises this codon, and
- a tRNA synthetase that recognises only that tRNA and only the non-standard amino acid.

Expanding the genetic code is an area of research of synthetic biology, an applied biological discipline whose goal is to engineer living systems for useful purposes. The genetic code expansion enriches the repertoire of useful tools available to science.

Codon Usage Bias



AA	Codon	Freq	AA	Codon	Freq
Ala	GCA	0.28	**Leu**	CTT	0.05
Ala	GCC	0.07	**Leu**	TTA	0.03
Ala	GCG	0.21	**Leu**	TTG	0.02
Ala	GCT	0.45	**Lys**	AAA	0.81
Arg	AGA	0.02	**Lys**	AAG	0.19
(Arg)	**AGG**	**0**	**Met**	ATG	1
(Arg)	**CGA**	**0**	**Phe**	TTC	0.79
Arg	CGC	0.24	**Phe**	TTT	0.21
(Arg)	**CGG**	**0.01**	**Pro**	CCA	0.08
Arg	CGT	0.73	**(Pro)**	**CCC**	**0.01**
Asn	AAC	0.91	**Pro**	CCG	0.82
Asn	AAT	0.09	**Pro**	CCT	0.08
Asp	GAC	0.72	**Ser**	AGC	0.15
Asp	GAT	0.28	**(Ser)**	**AGT**	**0.01**
Cys	TGC	0.8	**Ser**	TCA	0.02
Cys	TGT	0.2	**Ser**	TCC	0.39
Gln	CAA	0.14	**Ser**	TCG	0.04
Gln	CAG	0.86	**Ser**	TCT	0.39
Glu	GAA	0.83	**Stop**	TAA	0.83
Glu	GAG	0.17	**Stop**	TAG	0.17
(Gly)	**GGA**	**0**	**Stop**	TGA	0
Gly	GGC	0.5	**Thr**	ACA	0.02
(Gly)	**GGG**	**0.01**	**Thr**	ACC	0.56
Gly	GGT	0.48	**Thr**	ACG	0.05
His	CAC	0.83	**Thr**	ACT	0.36
His	CAT	0.17	**Trp**	TGG	1
Ile	ATA	0.02	**Tyr**	TAC	0.8
Ile	ATC	0.86	**Tyr**	TAT	0.2
Ile	ATT	0.12	**Val**	GTA	0.21
(Leu)	**CTA**	**0.01**	**Val**	GTC	0.07
Leu	CTC	0.06	**Val**	GTG	0.15
Leu	CTG	0.83	**Val**	GTT	0.57

Frequency refers to the percentage occurrence of synonymous codons encoding amino acids in *E. coli* highly expressed proteins. The nucleic acid sequences of the following genes were combined into a single pseudo-gene and then used in the Kazusa Countcodon program http://www.kazusa.or.jp/codon with eubacterial translation exceptions to generate a codon usage table for that pseudo-gene: ompA (V00307),

Mutations

Point mutations

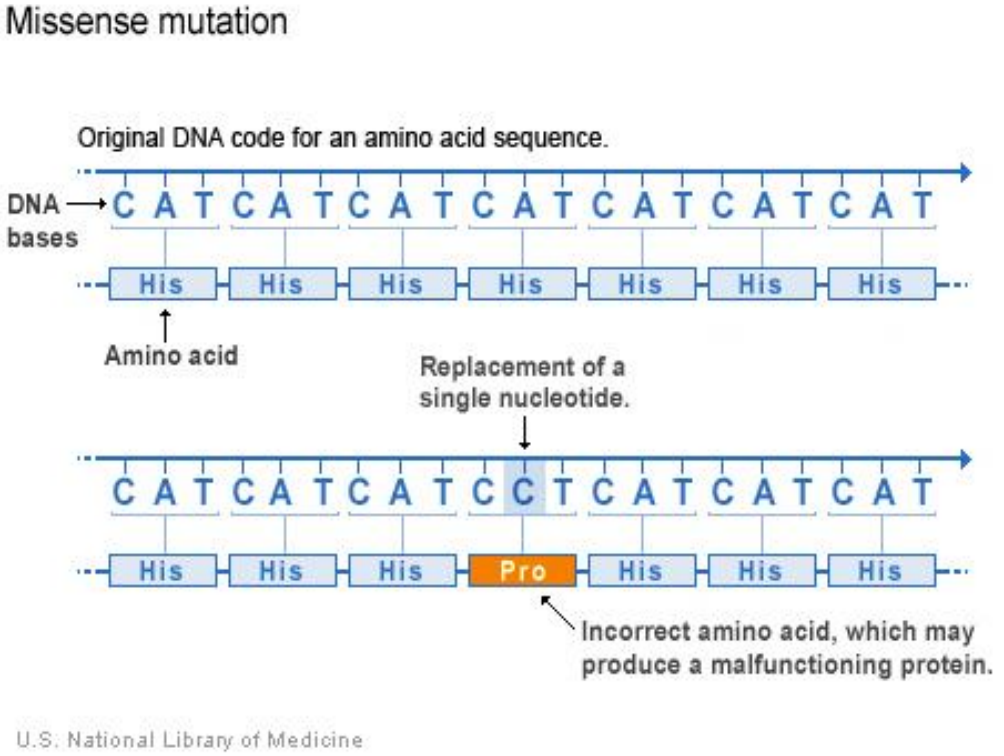
Missense mutation:
- Results in a different amino acid

Nonsense mutation:
- Results in a stop codon

Synonymous substitution (silent mutation):
- Is a mutation that doesn't change the AA

Insertions / Deletions (indels)

- Are frameshift mutations (if not multiple of 3)




```
   DNA: 5' – ATG ACT CAC CGA GCG CGA AGC TGA – 3'
        3' – TAC TGA GTG GCT CGC GCT TCG ACT – 5'
  mRNA: 5' – AUG ACU CAC CGA GCG CGA AGC UGA – 3'
Protein:     Met Thr His Arg Ala Arg Ser Stop

   DNA: 5' – ATG ACT CAC TGA GCG CGA AGC TGA – 3'
        3' – TAC TGA GTG ACT CGC GCT TCG ACT – 5'
  mRNA: 5' – AUG ACU CAC UGA GCG CGU AGC UGA – 3'
Protein:     Met Thr His Stop
```

The End