
# DNA sequencing and PCR

# Resources

- This lecture
- Campbell and Farrell's Biochemistry, pp. 377-380, 384-387

# What is DNA sequencing?

- DNA sequencing is the process of determining the exact order of nucleotides in a genome.
- Importance:
  - Identification of genes and their localization
  - Identification of protein structure and function
  - Identification of DNA mutations
  - Genetic variations among individuals in health and disease
  - Prediction of disease-susceptibility and treatment efficiency
  - Evolutionary conservation among organisms

# DNA sequencing of organism genome

- Viruses and prokaryotes first
- Human mitochondrial DNA
- The first eukaryotic genome sequenced was that of yeast, Saccharomyces cerevisiae.
- The genome of a multicellular organism, the nematode Caenorhabditis elegans.
- Determination of the base sequence in the human genome was initiated in 1990 and completed in May 2006 via the Human Genome Project

| SPECIES | BASE PAIRS (estimated) | GENES (estimated) | CHROMOSOMES |
|---|---|---|---|
| **Human** *(Homo sapiens)* | 3.2 billion | ~ 25,000 | 46 |
| **Mouse** *(Mus musculus)* | 2.6 billion | ~ 25,000 | 40 |
| **Fruit Fly** *(Drosophilia melanogaster)* | 137 million | 13,000 | 8 |
| **Roundworm** *(Caenorhabditis elegans)* | 97 million | 19,000 | 12 |
| **Yeast** *(Saccharomyces cerevisia)* | 12.1 million | 6,000 | 32 |
| **Bacteria** (Escherichia coli) | 4.6 million | 3,200 | 1 |
| **Bacteria** (H. influenzae) | 1.8 million | 1,700 | 1 |

# Nucleotides per genomes

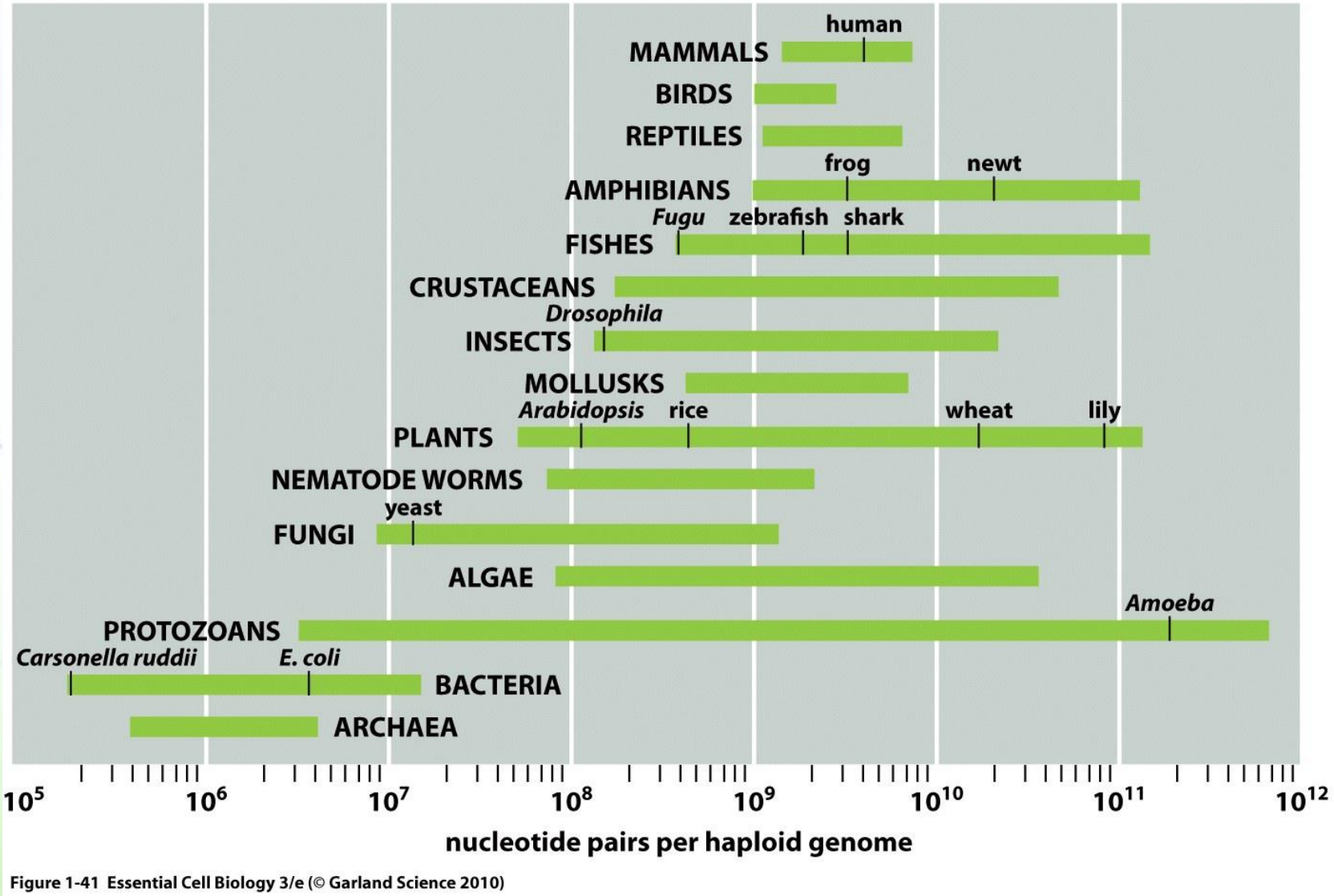
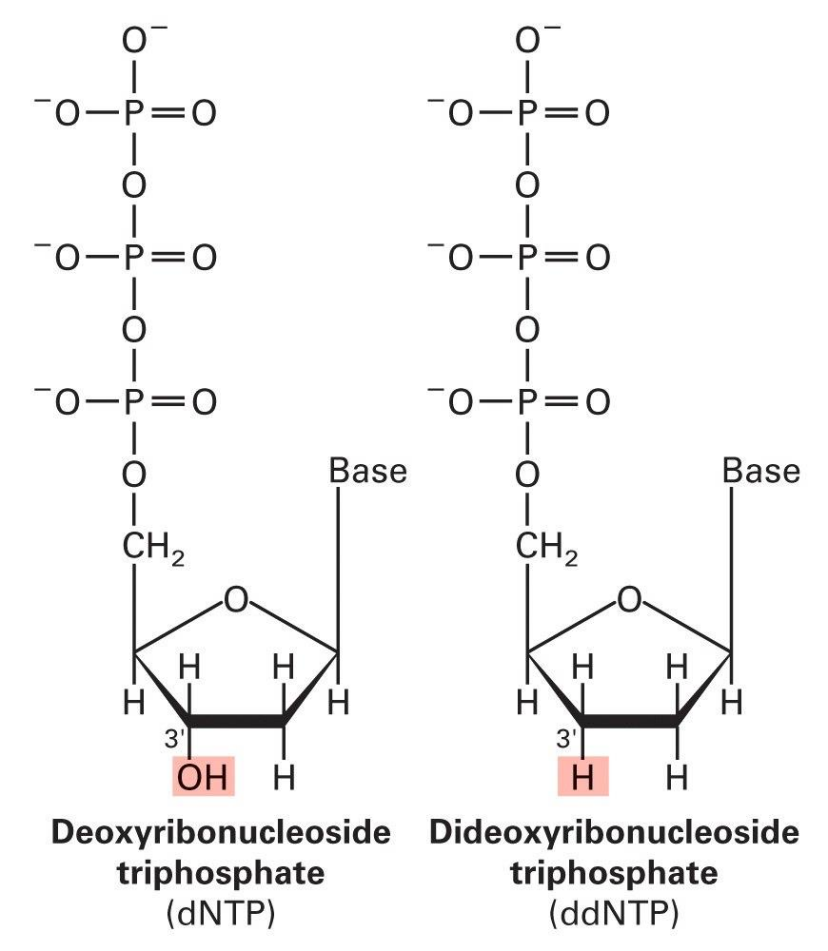



Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)

# Method of DNA sequencing

- Based on premature termination of DNA synthesis by dideoxynucleotides

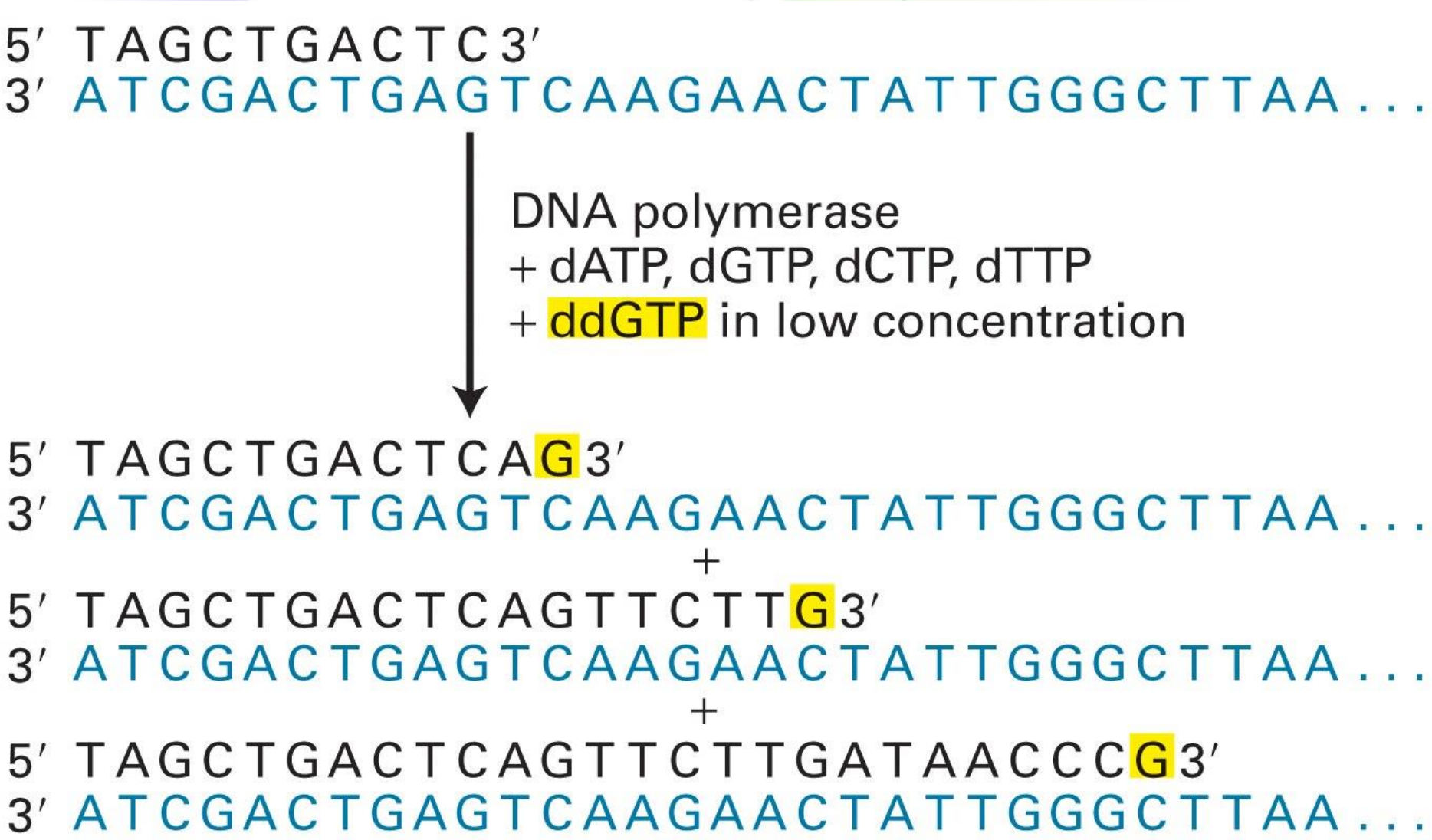


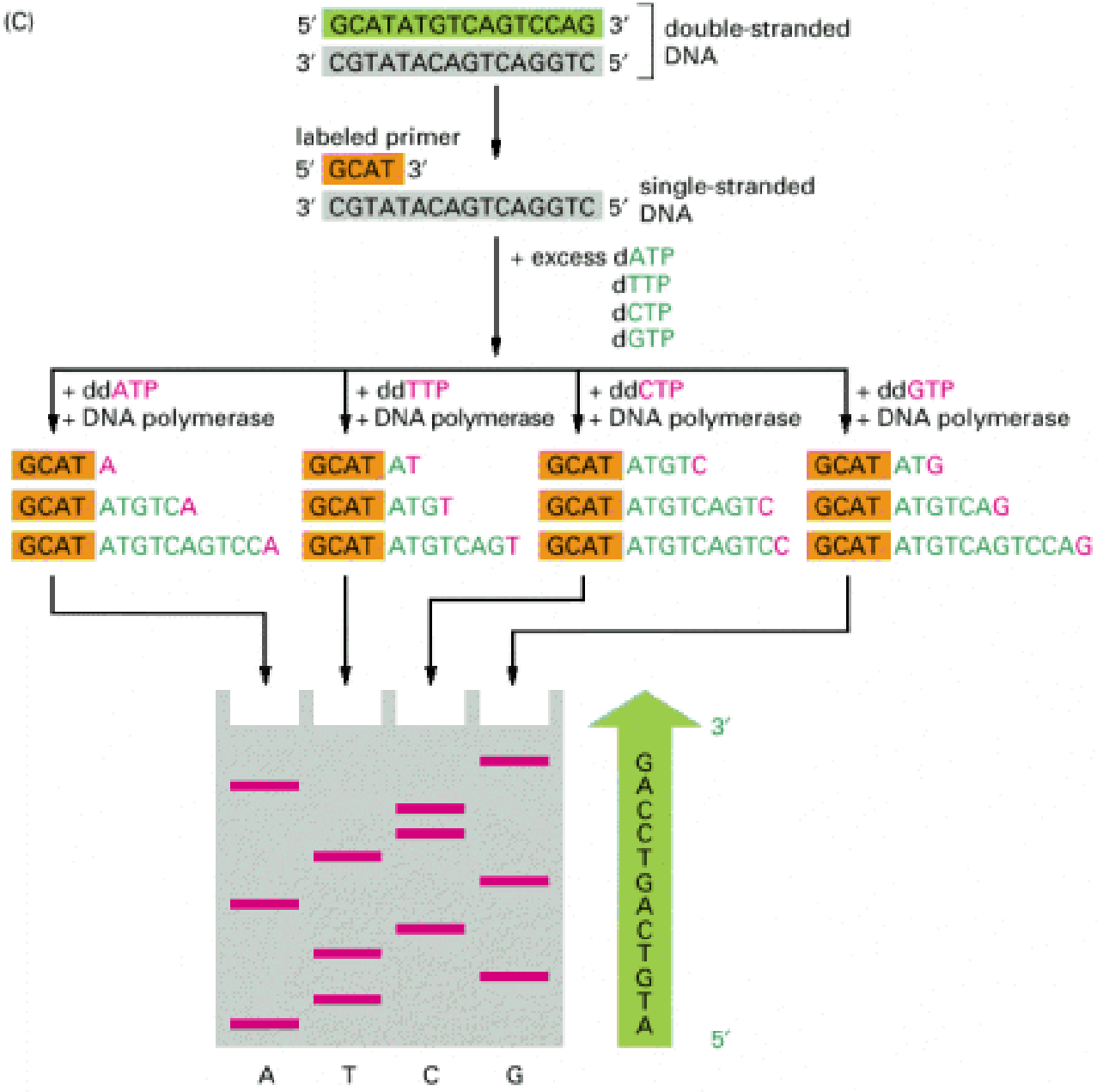
# The process...

- DNA synthesis is initiated from a primer that has been labeled with a radioisotope
- Four separate reactions are run, each including deoxynucleotides plus one dideoxynucleotide (either A, C, G, or T)
- Incorporation of a dideoxynucleotide stops further DNA synthesis because no 3 hydroxyl group is available for addition of the next nucleotide

# Generation of fragments

- A series of labeled DNA molecules are generated, each terminated by the dideoxynucleotide in each reaction
- These fragments of DNA are then separated according to size by gel electrophoresis and detected by exposure of the gel to X-ray film
- The size of each fragment is determined by its terminal dideoxynucleotide, so the DNA sequence corresponds to the order of fragments read from the gel

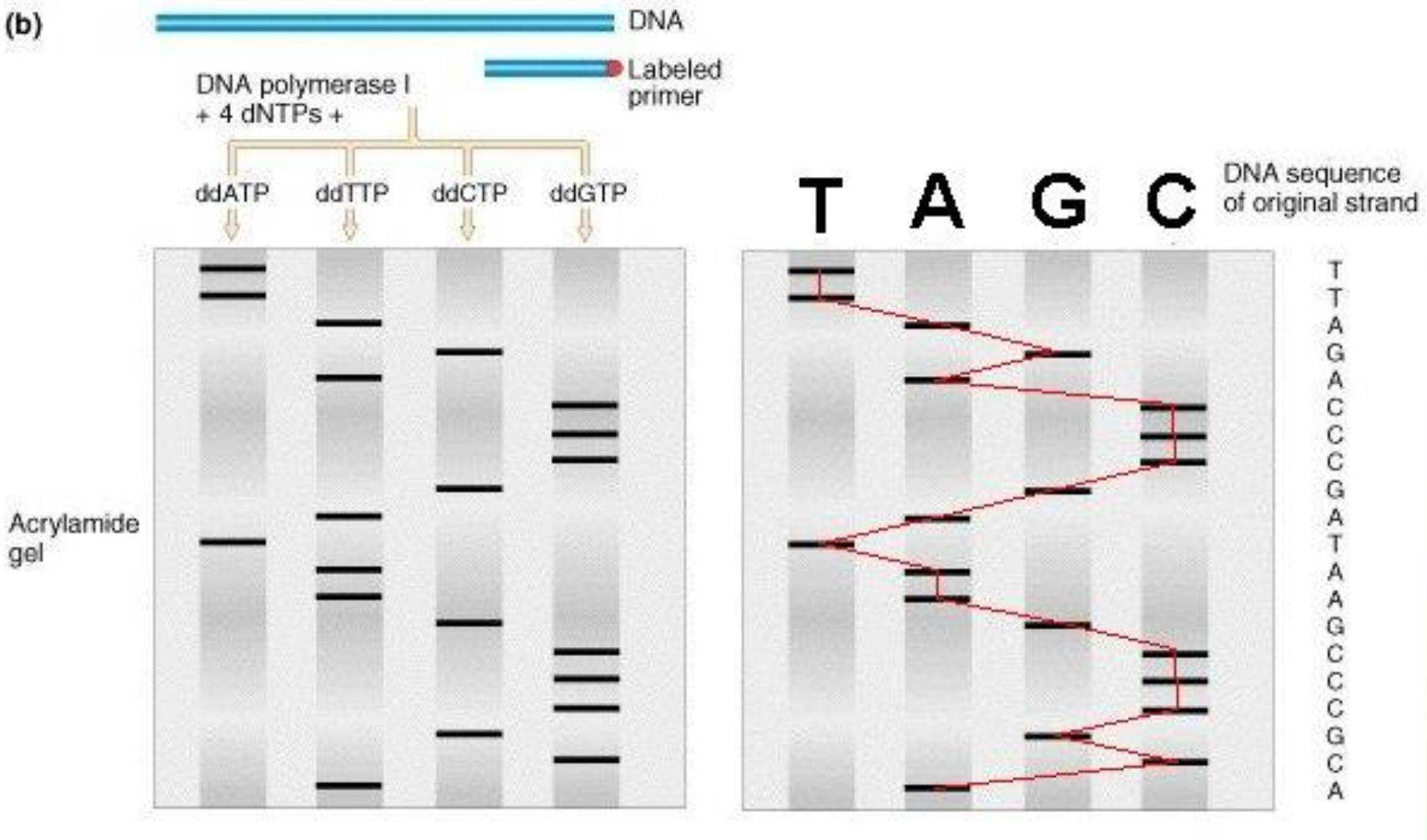

5′ TAGCTGACTC 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
DNA polymerase
+ dATP, dGTP, dCTP, dTTP
+ ddGTP in low concentration
5′ TAGCTGACTCAG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
+
5′ TAGCTGACTCAGTTCTTG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
+
5′ TAGCTGACTCAGTTCTTGATAACCCG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .




DNA sequence reading directly from the bottom of the gel upward, is

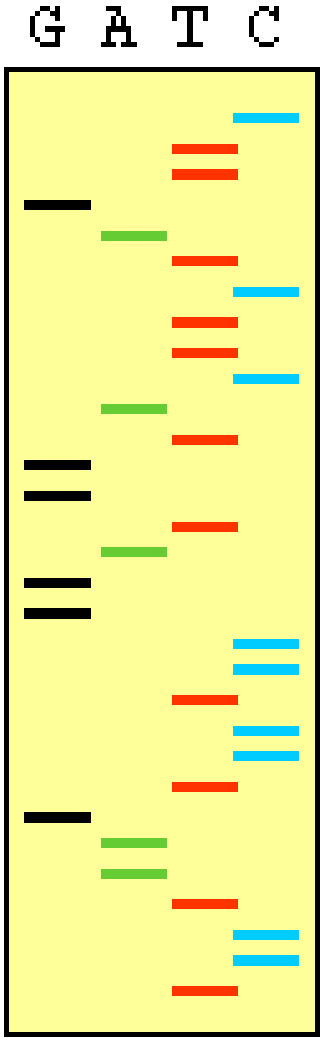
ATGTCAGTCCAG

1 12

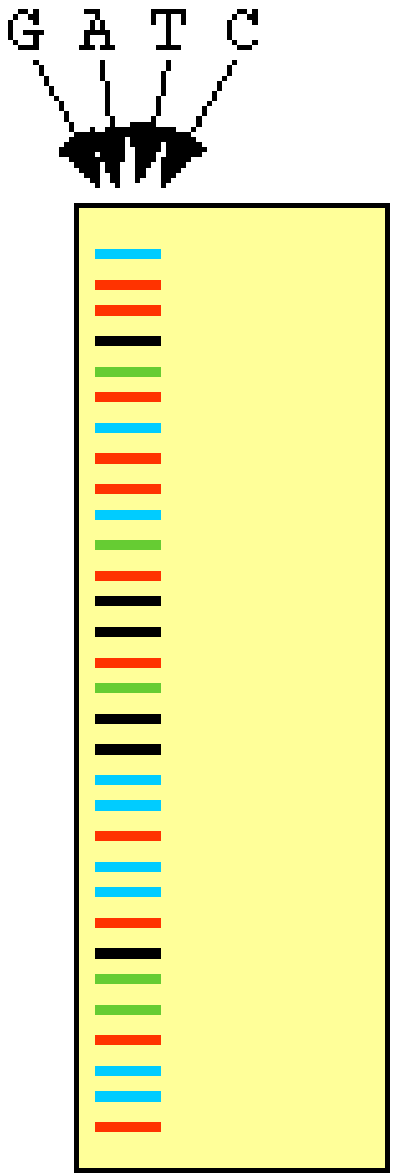

(b)
DNA
Labeled primer
DNA polymerase I
+ 4 dNTPs +
ddATP
ddTTP
ddCTP
ddGTP
Acrylamide gel
T A G C
DNA sequence of original strand
T T A G A C C C G A T A A G C C C G C A

# Fluorescence-based DNA sequencing

- Large-scale DNA sequencing is frequently performed using automated systems using fluorescence-based reactions using labeled ddNTPs
- In this case, all four terminators can then be placed in a single tube, and only one reaction is necessary
- The reactions are run into one lane on a gel and a machine is used to scan the lane with a laser



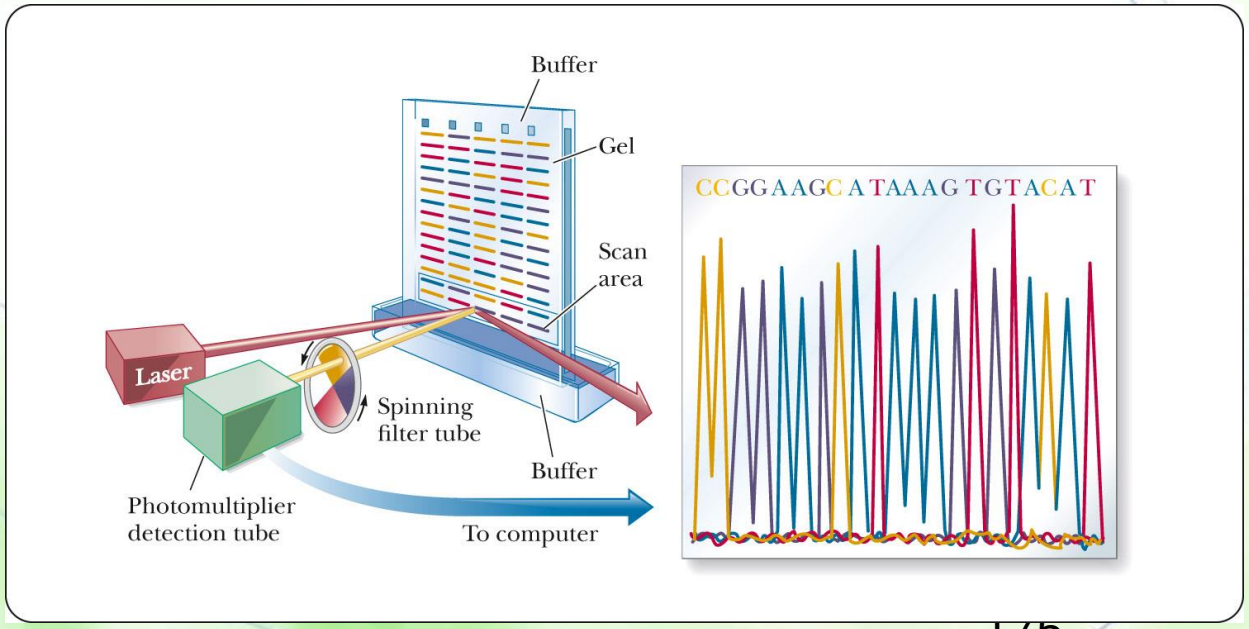

Here's what the products would look like in separate gel lanes.

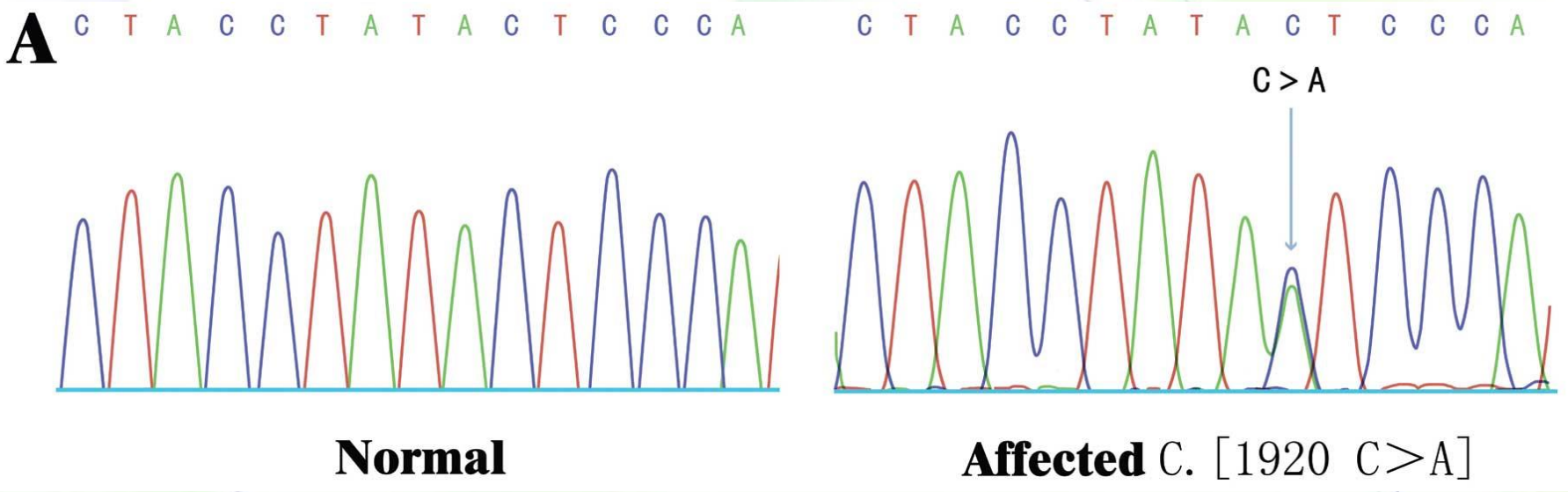



Here's what the products would look like in a single gel lanes.

# Detection of fragments

- The wavelength of fluorescence can be interpreted by the machine as an indication of which reaction (ddG, ddA, ddT, or ddC) the product came from
- The fluorescence output is stored in the form of chromatograms

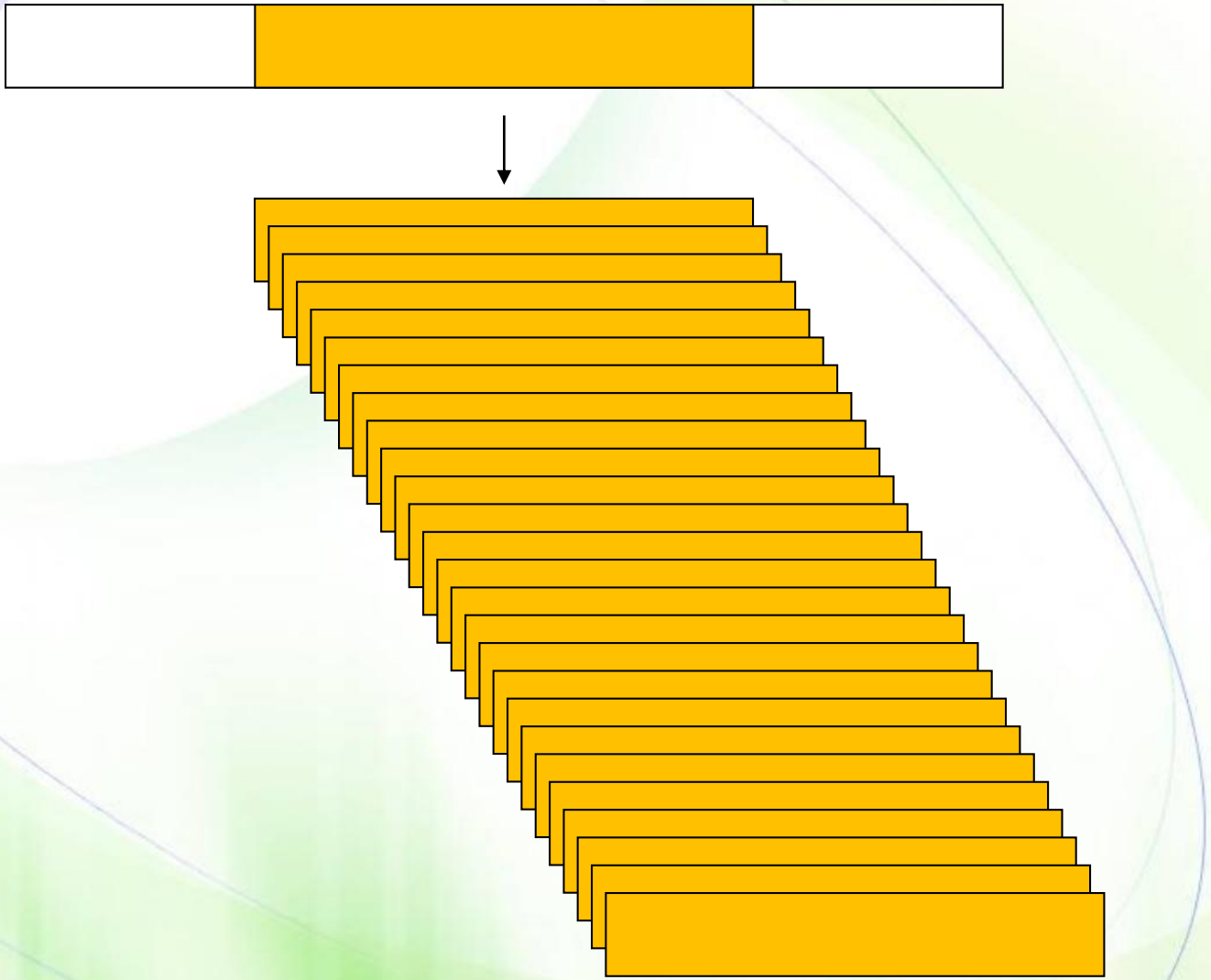





**What does it mean?**

# Polymerase Chain Reaction

- In 1984, Kary Mullis devised a method called the polymerase chain reaction (PCR) for amplifying specific DNA sequences
- PCR allows the DNA from a selected region of a genome to be amplified a billionfold, effectively "purifying" this DNA away from the remainder of the genome
- The PCR method is extremely sensitive; it can detect a single DNA molecule in a sample



# Components of PCR reaction

- A pair of primers that hybridize to the target DNA.
  - These primers should be specific for the target sequence and which are often about 15-25 nucleotides long. The region between the primers is amplified
- All four deoxyribonucleoside triphosphates (dNTPs: dATP, dCTP, dGTP and dTT)
- A heat-stable DNA polymerase

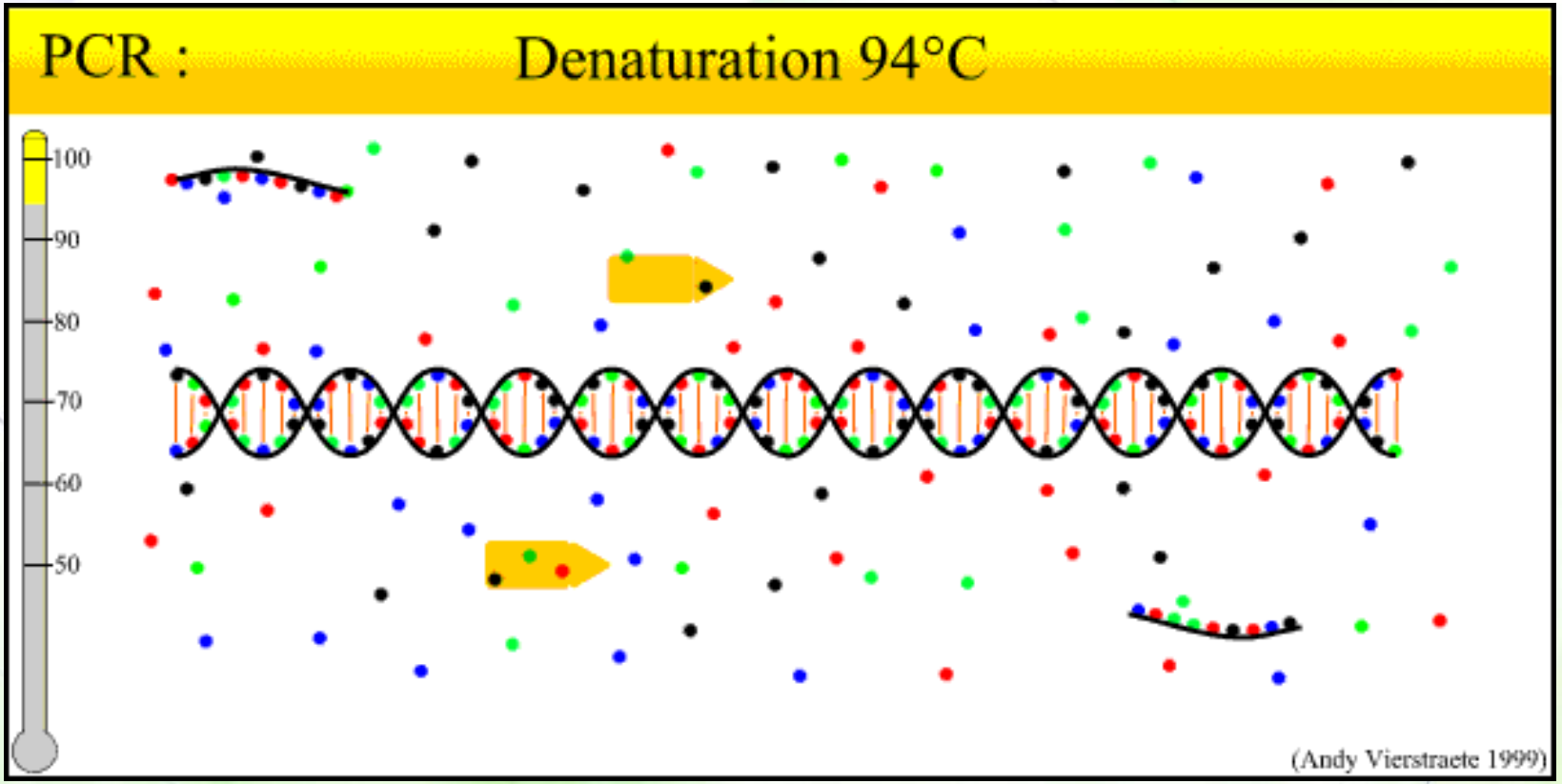
# DNA polymerases

- Suitably heat-stable DNA polymerases have been obtained from microorganisms whose natural habitat is hot springs

- For example, the widely used Taq DNA polymerase is obtained from a thermophilic bacterium, Thermus aquaticus, and is thermostable up to 94°C

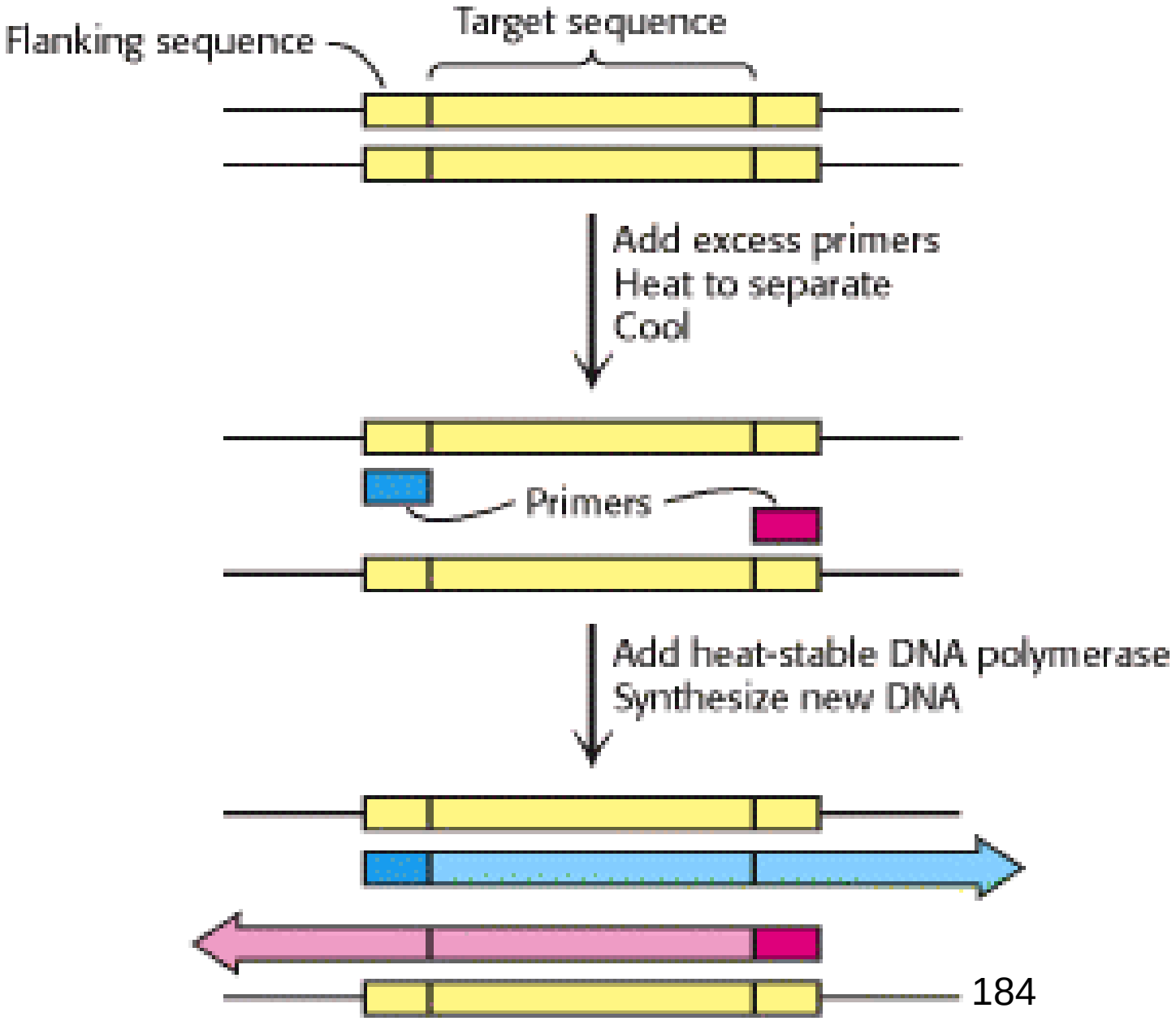
# PCR cycle

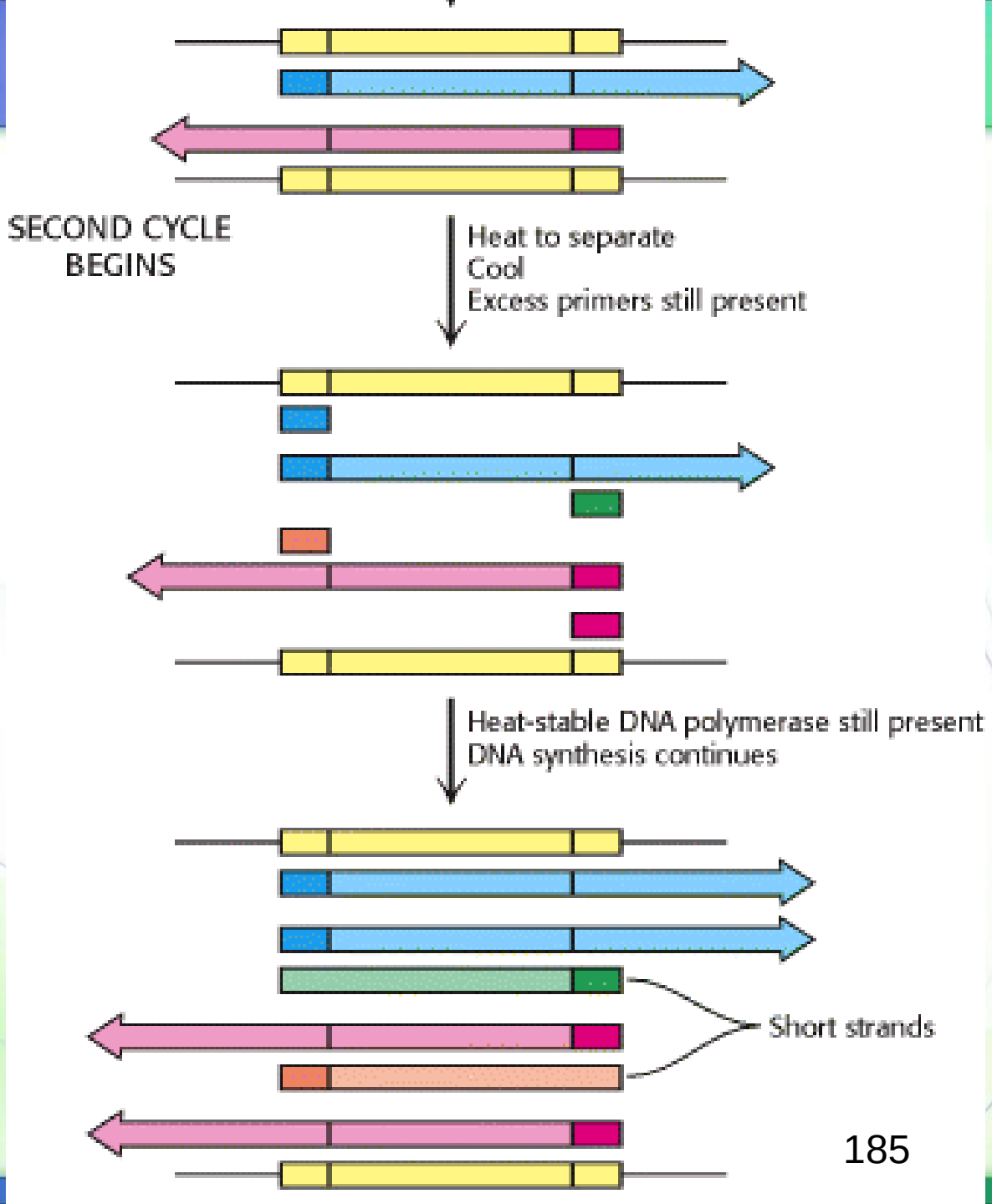
- Denaturation, typically at about 93-95°C. At this temperature the hydrogen bonds that hold together the two polynucleotides of the double helix are broken, so the target DNA becomes denatured into single-stranded molecules
- Reannealing at temperatures usually from about 50°C to 70°C where the primers anneal to the DNA
- DNA synthesis, typically at about 70-75°C, the optimum for Taq polymerase

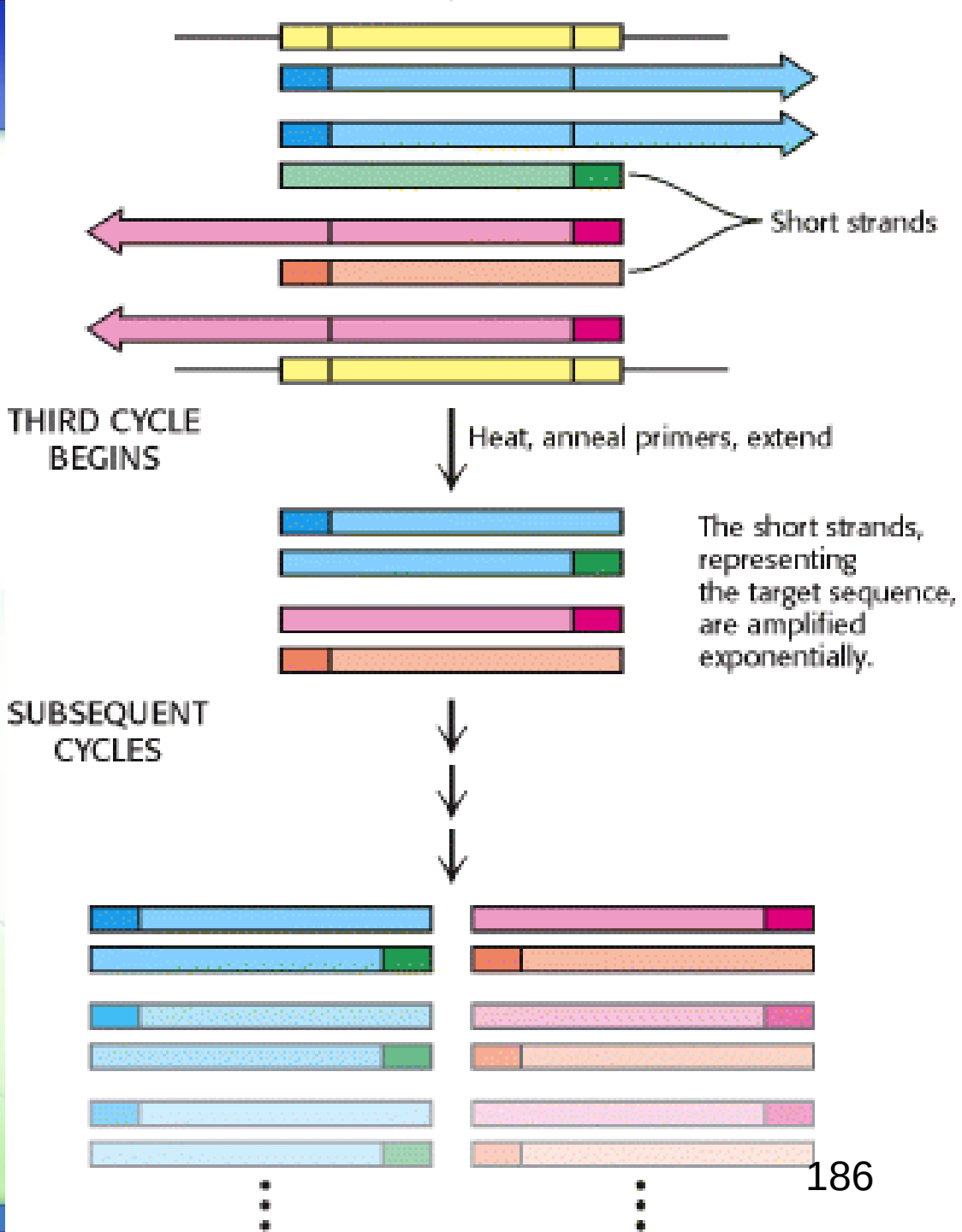

PCR : Denaturation 94°C
100
90
80
70
60
50
(Andy Vierstraete 1999)


A template sequence with 5 ATT repeats.

FIRST CYCLE BEGINS




SECOND CYCLE
BEGINS
Heat to separate
Cool
Excess primers still present
Heat-stable DNA polymerase still present
DNA synthesis continues
Short strands

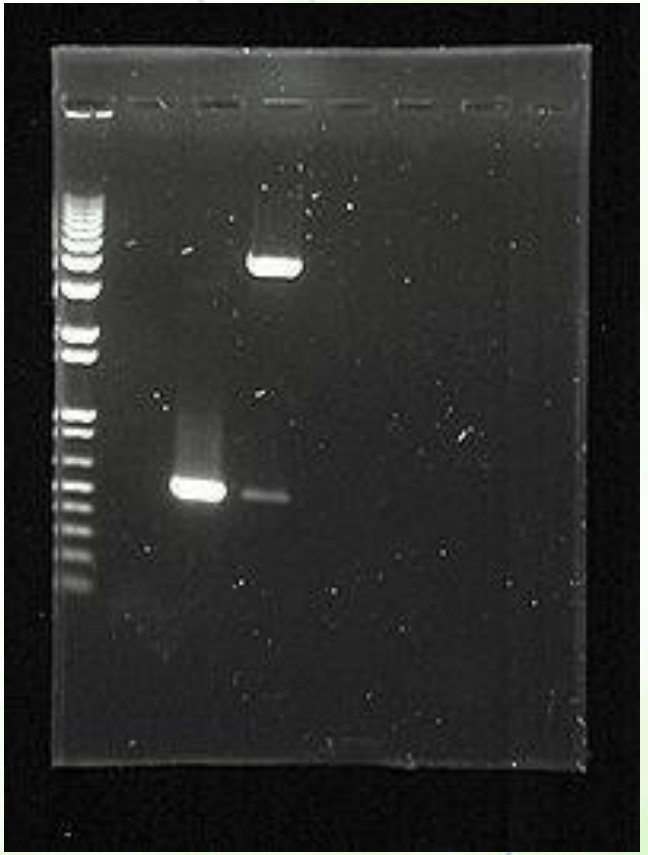

Short strands
THIRD CYCLE BEGINS
Heat, anneal primers, extend
The short strands, representing the target sequence, are amplified exponentially.
SUBSEQUENT CYCLES

# PCR cycles

- 20-30 cycles of reaction are required for DNA amplification,
  - the products of each cycle serving as the DNA templates for the next-hence the term polymerase "chain reaction“
- Every cycle doubles the amount of DNA
- After 30 cycles, there will be over 250 million short products derived from each starting molecule

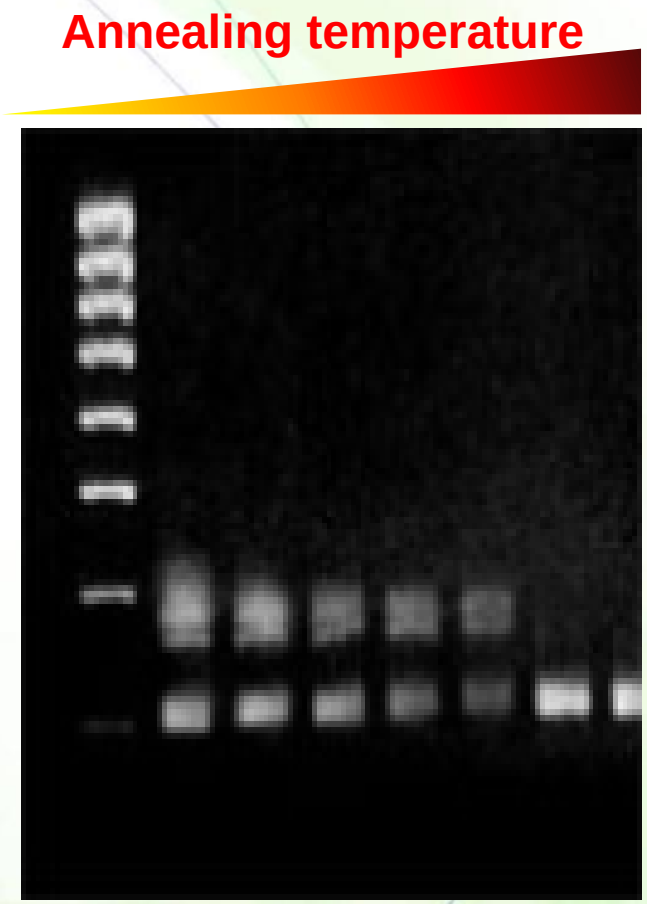
# Detection of DNA fragments

- This DNA fragment can be easily visualized as a discrete band of a specific size by agarose gel electrophoresis



# Importance of primers

- The specificity of amplification depends on the specificity of the primers to not recognize and bind to sequences other than the intended target DNA sequences
- How can you prevent it?
- How can you take advantage of it?



## Advantages

- Easy, fast, sensitive, robust
- Discovery of gene families
- Disease diagnosis

## Disadvantages

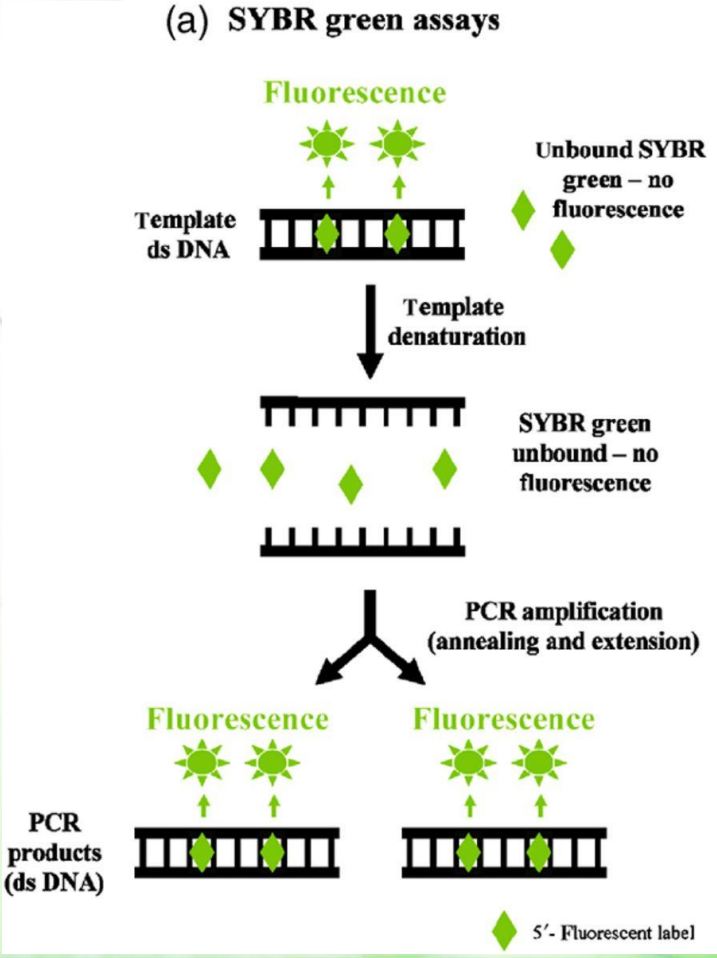
- Primers must be known
- Contamination
- Product length is limited (usually <5 Kb)
- Accuracy is an issue
- Not quantitative

# Forensic medicine

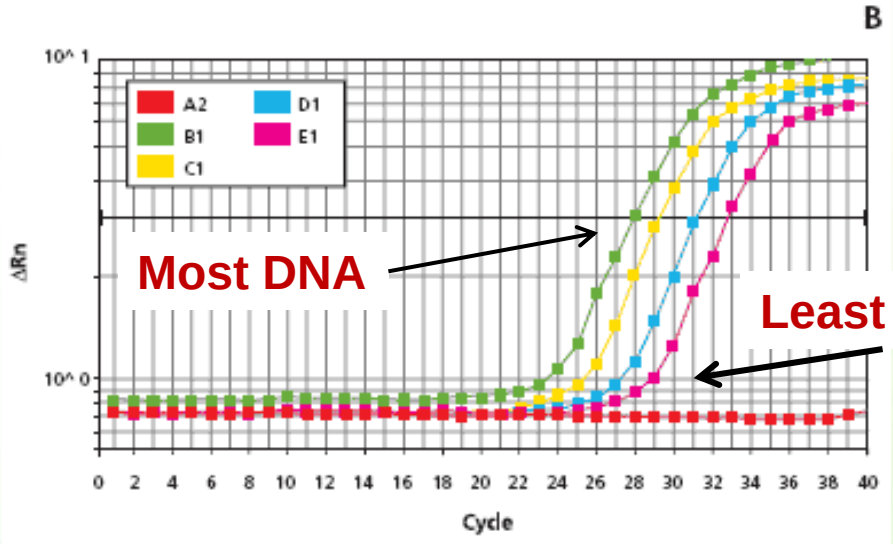
- An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population

- PCR amplification of multiple genes is being used to establish paternity and criminal cases

# Quantitative PCR (qPCR)

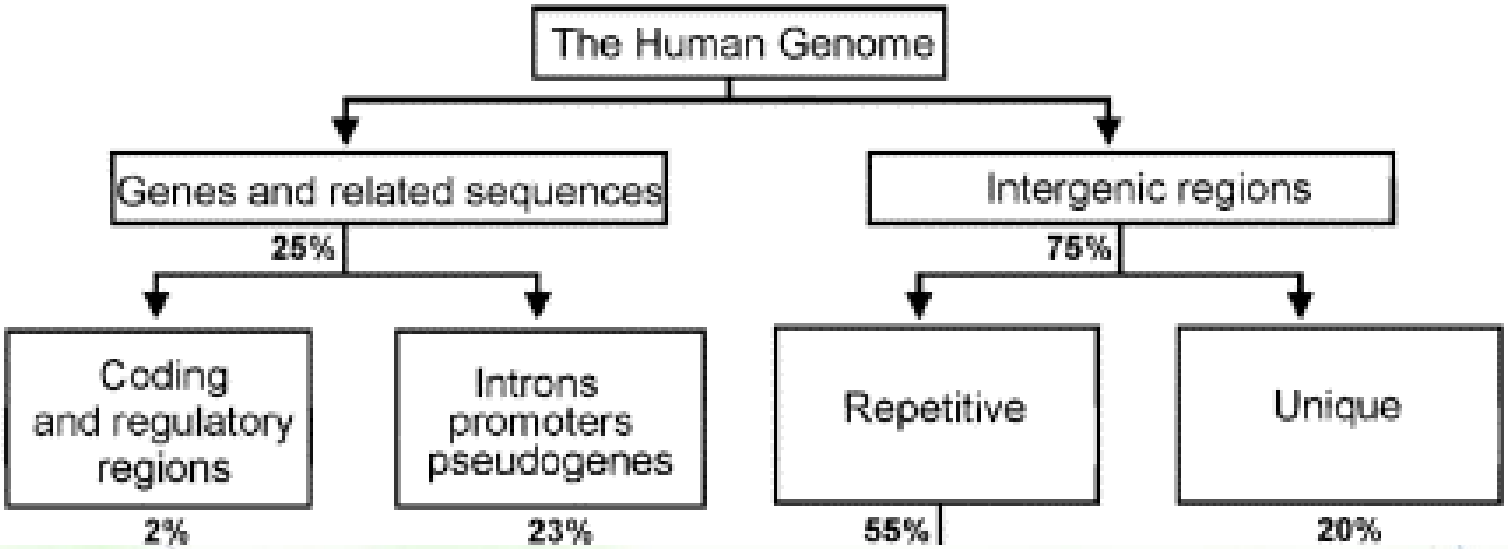



- SYBR green binds to double-stranded DNA and fluoresces only when bound.
- Another way of relative quantitation of amount of DNA in a sample by amplifying it in the presence of SYBR green.
- The higher the amount of DNA, the sooner it is detected.

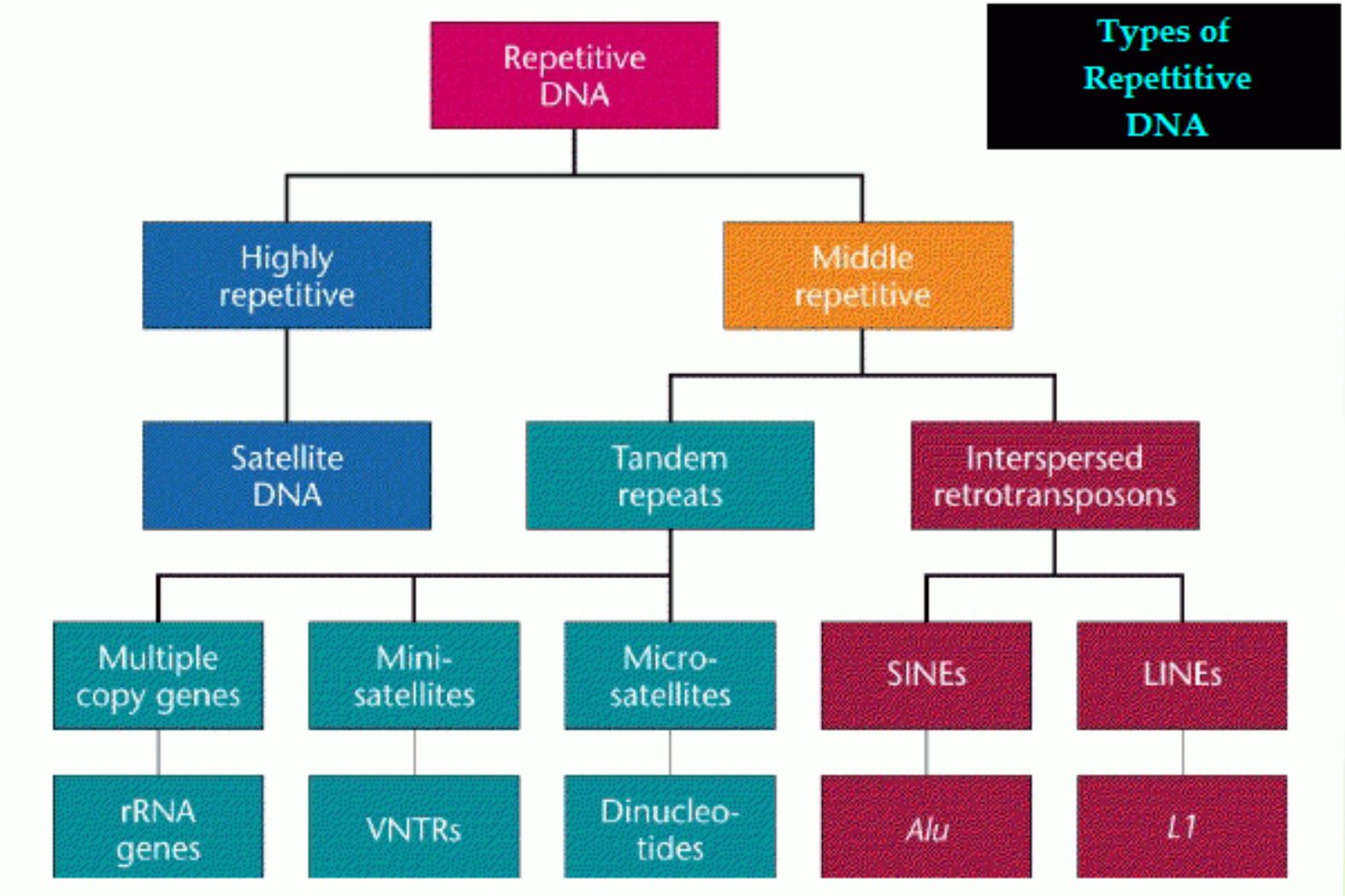



- http://www.youtube.com/watch?v=kvQWKcMdyS4
- http://www.bio.davidson.edu/courses/immunology/flash/rt_pcr.html

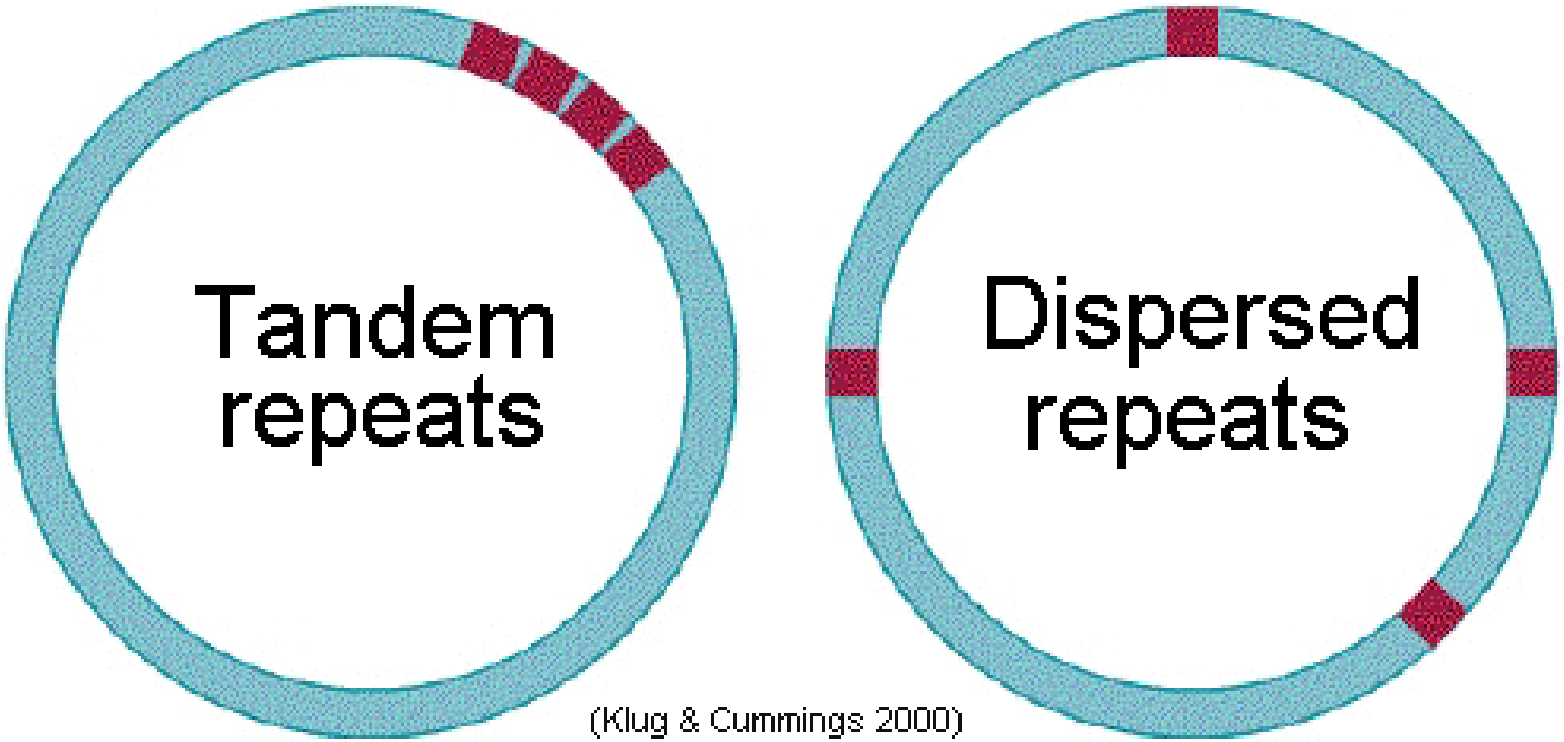
# Components of the human genome



# Repetitive DNA sequences



# Tandem vs. dispersed



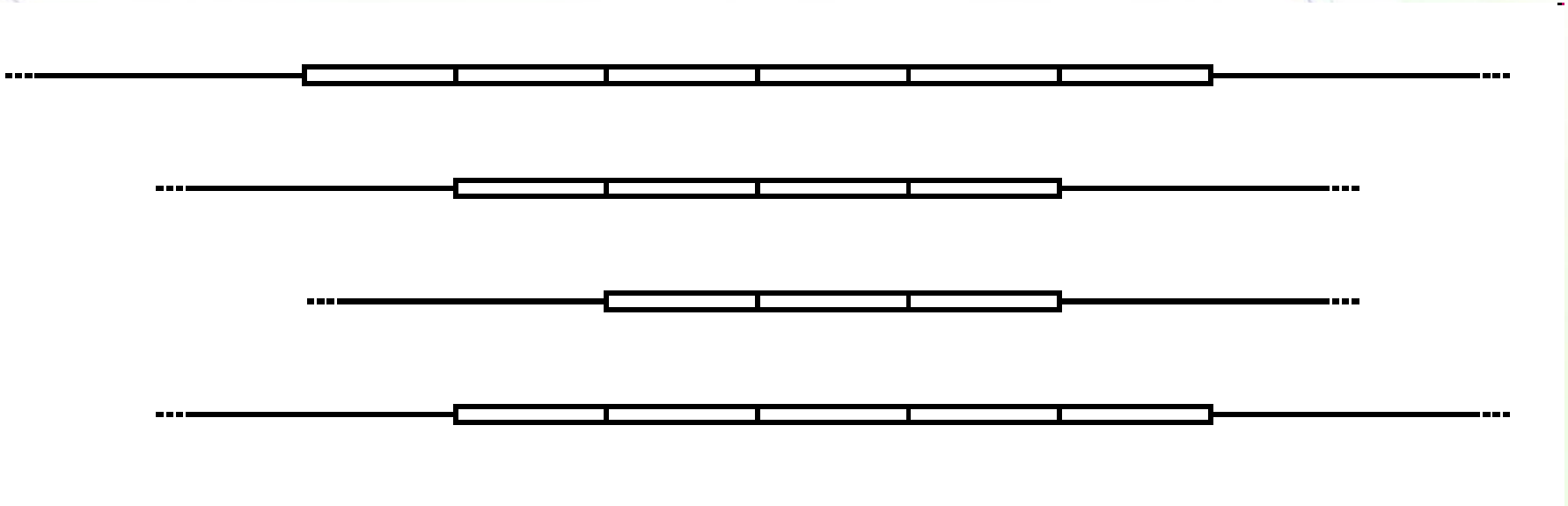

(Klug & Cummings 2000)

# Satellite (macro-satellite) DNA

- Repeats of 100 to 6500 bp
  - Tandemly centromeric repeats (171 bp) unique to each chromosome
    - Each chromosome has its unique sequence
    - possible to make DNA probes specific to each
  - Telomeric repeats

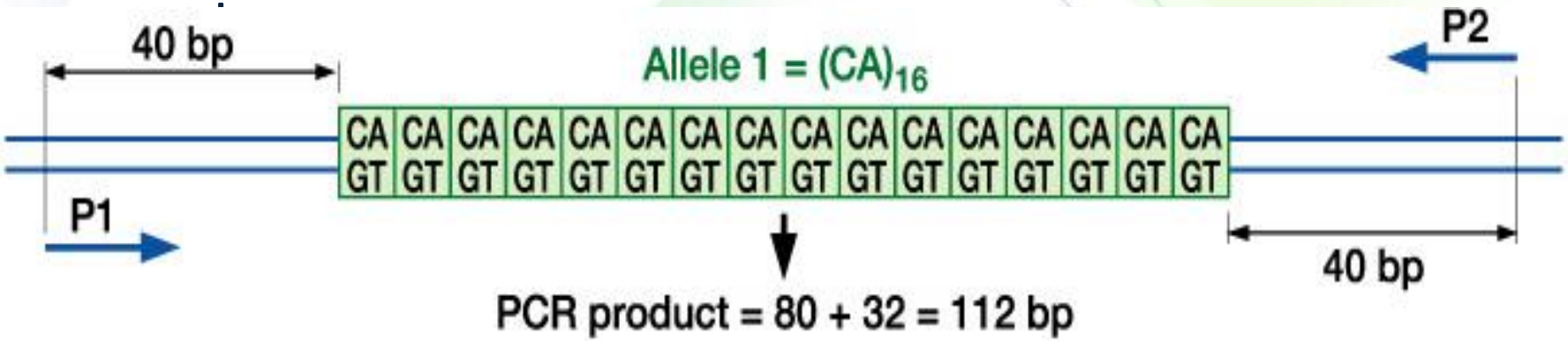
# VNTRs (minisatellite)

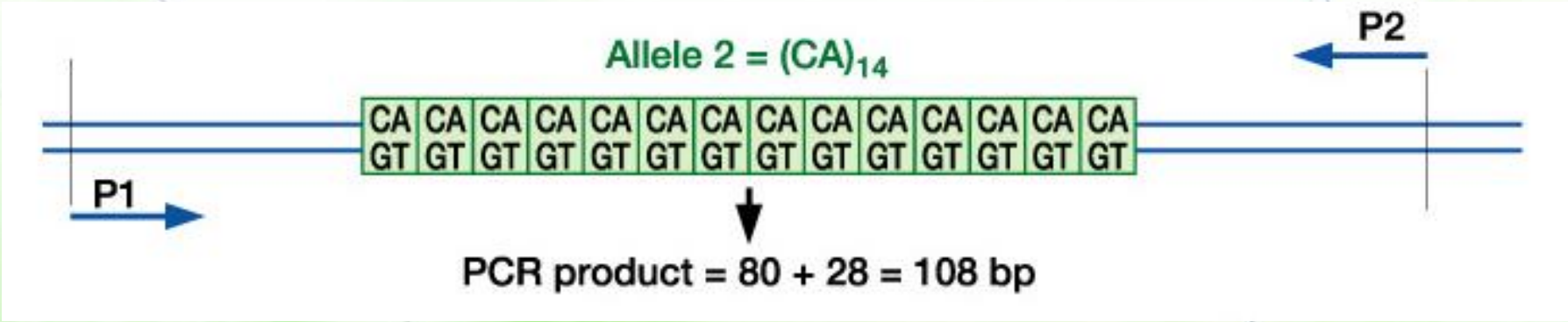
- Mini satellite sequences or VNTRs (variable number of tandem repeats)
- They are composed of 20 to 100 bp repeats



# STRs (microsatellites)

- STRs (short tandem repeats) composed of 2 to 10 bp



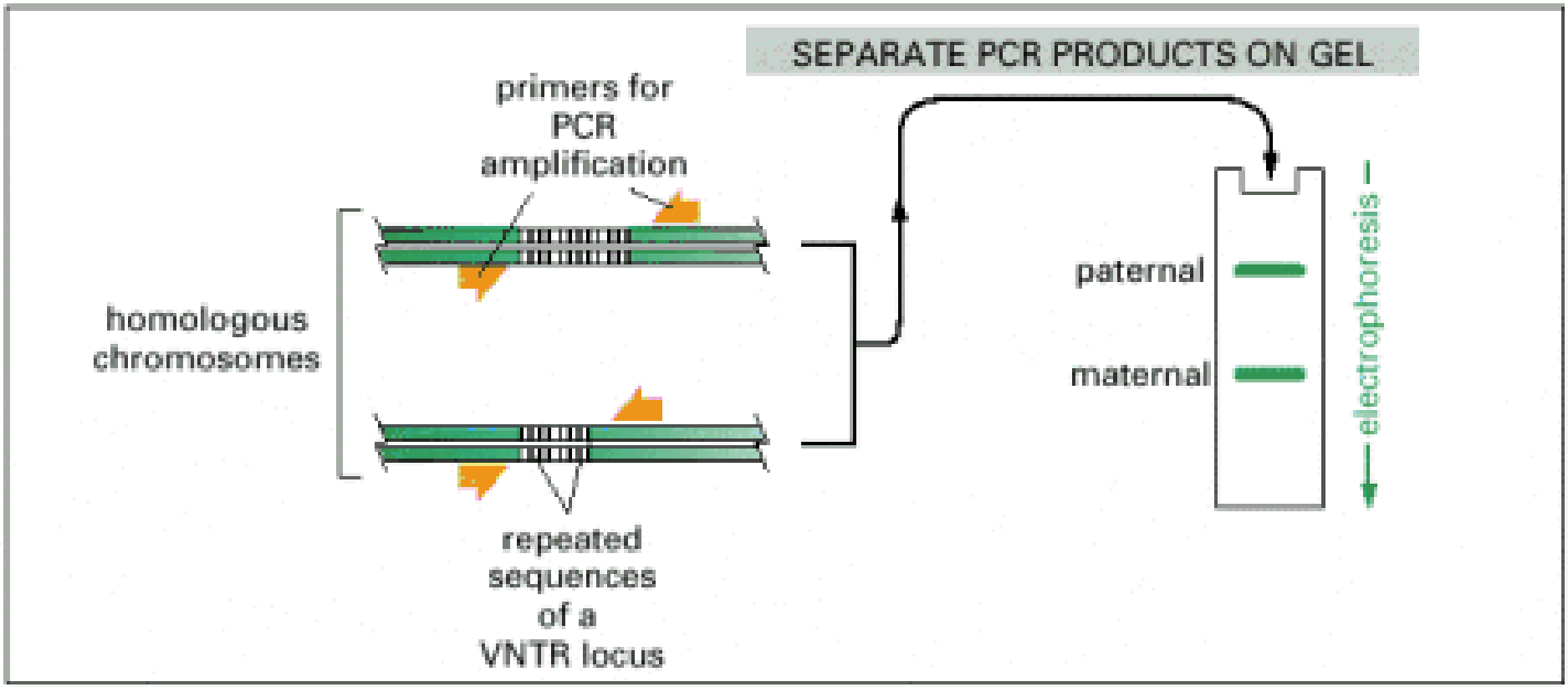
# Polymorphisms of VNTR and STR

- The number of repeats in micro and mini satellites are highly variable (polymorphic)

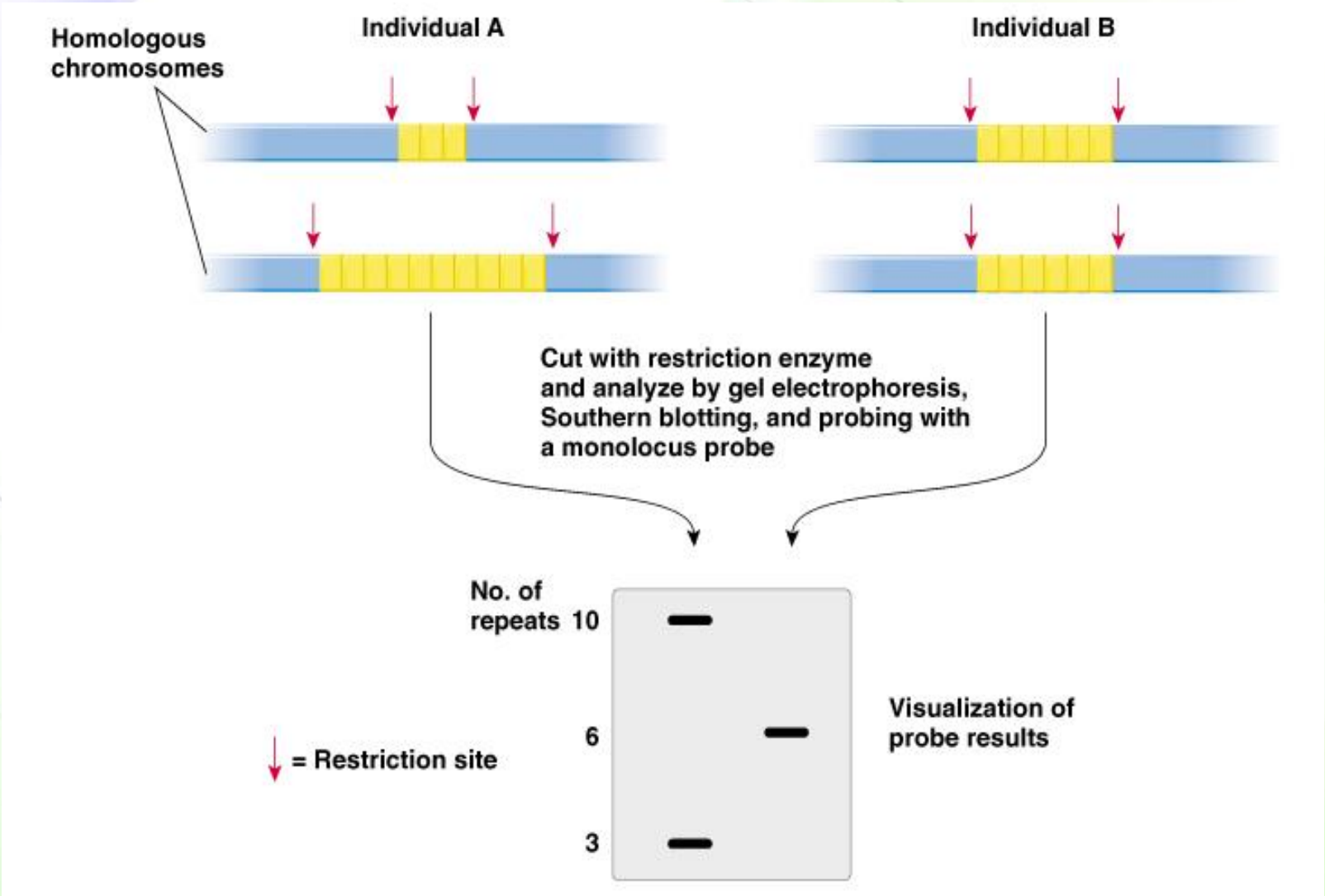
- Useful in:
  - Locating genes in DNA (gene mapping)
  - DNA profiling for paternity testing, forensic testing, confirmation of relatedness and dead body identification

# PCR of VNTRs

- VNTR blocks can be extracted and analyzed by RFLP or PCR and size determined by electrophoresis

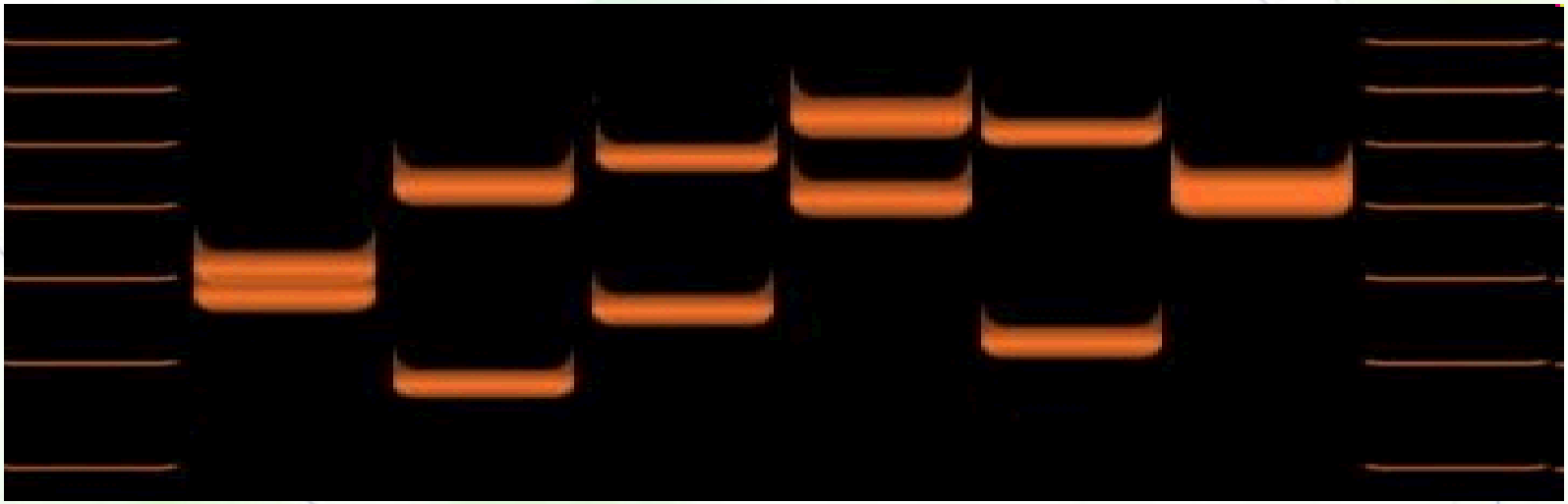


# Microsatellites and VNTRs as DNA Markers



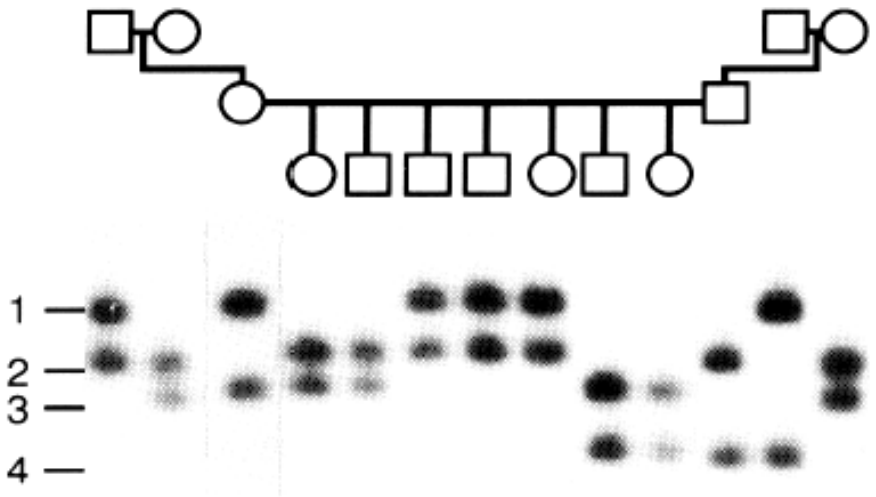
# VNTR in medicine and more

- The picture below illustrates VNTR allelic length variation among 6 individuals.

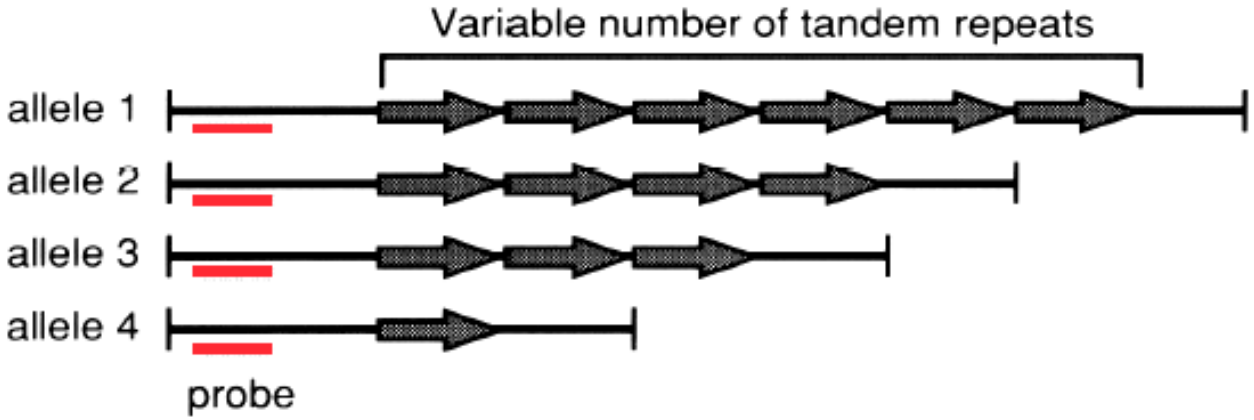


**The likelihood of 2 unrelated individuals having same allelic pattern extremely improbable**

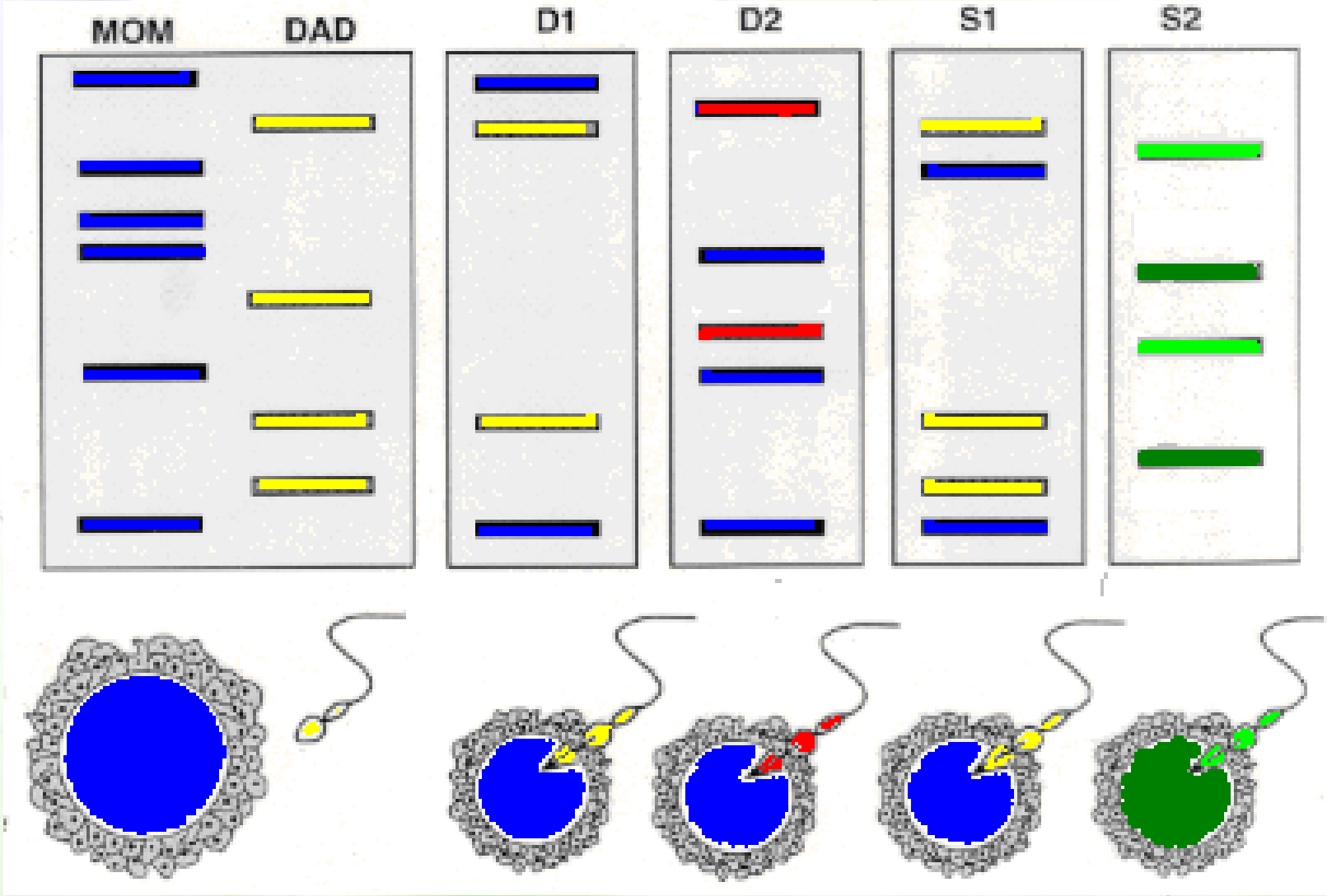
# Real example




single-locus probe but multiple alleles




Thompson & Thompson Genetics in Medicine, p. 130, 1991


MOM
DAD
D1
D2
S1
S2

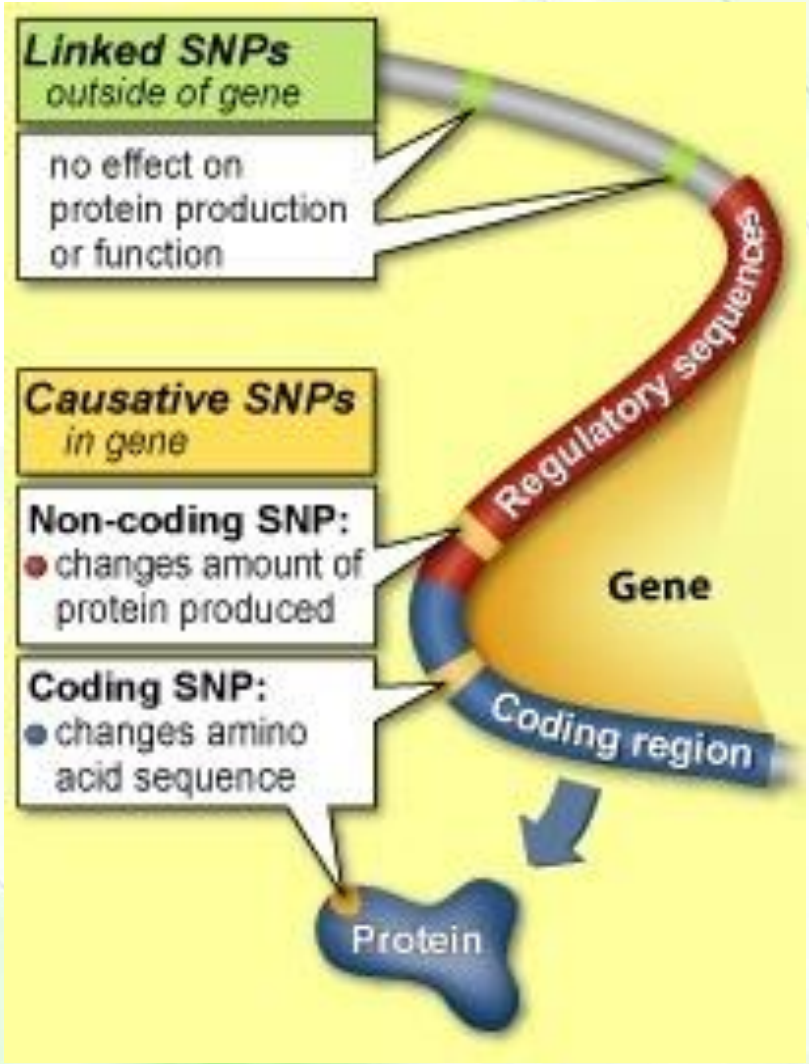
# SINEs and LINEs

- Two thirds (66.7%) of the repetitive non-coding DNA sequences that are dispersed throughout the genome
- Divided into short and long interspersed sequences, SINES and LINES.
- LINES:
  - 7000 bp in length
  - Represent about 4% of our total human genome
- SINES:
  - Shorter interspersed elements 90 to 500 bp in length
    - Example: Alu sequence (~300 bp)
      - Alu sequences are unique to humans (and some apes)
      - 10% of the total human genome

# Single nucleotide polymorphism (SNPs)

- Another source of genetic variation
- Single-nucleotide substitutions of one base for another
- Two or more versions of a sequence must each be present in at least one percent of the general population
- SNPs occur throughout the human genome - about one in every 300 nucleotide base pairs.
  - ~10 million SNPs within the 3-billion-nucleotide human genome

# Categories of SNPs




# Transcription

# Resources

- This lecture
- Campbell and Farrell's Biochemistry, Chapter 11

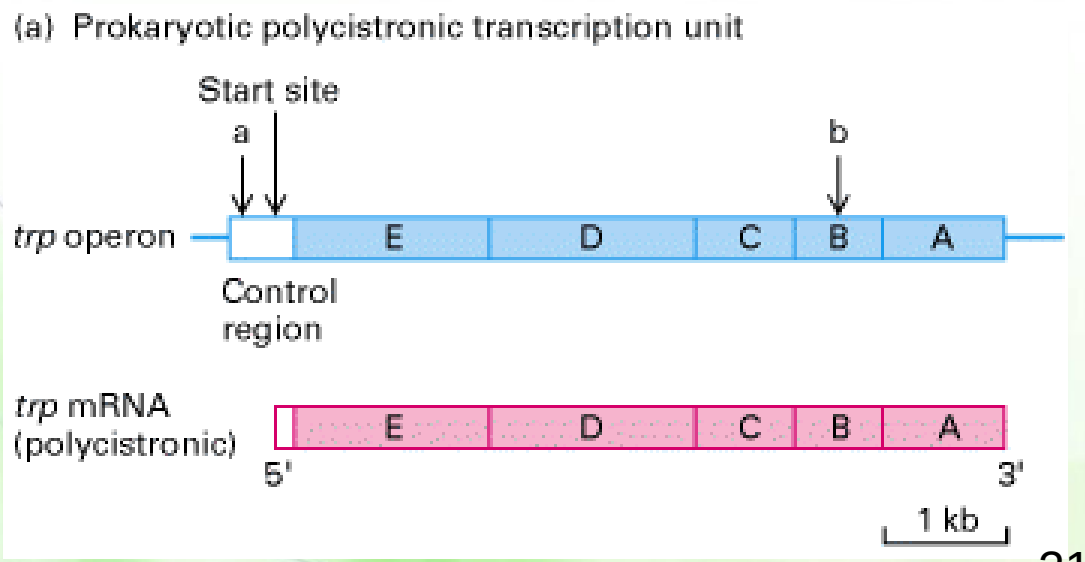
# Definition of a gene

- The entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide

- A cistron: a genetic unit that encodes a polypeptide(s)
  - If it encodes one polypeptide, it is monocistronic
  - If it encodes several polypeptides, it is polycistronic

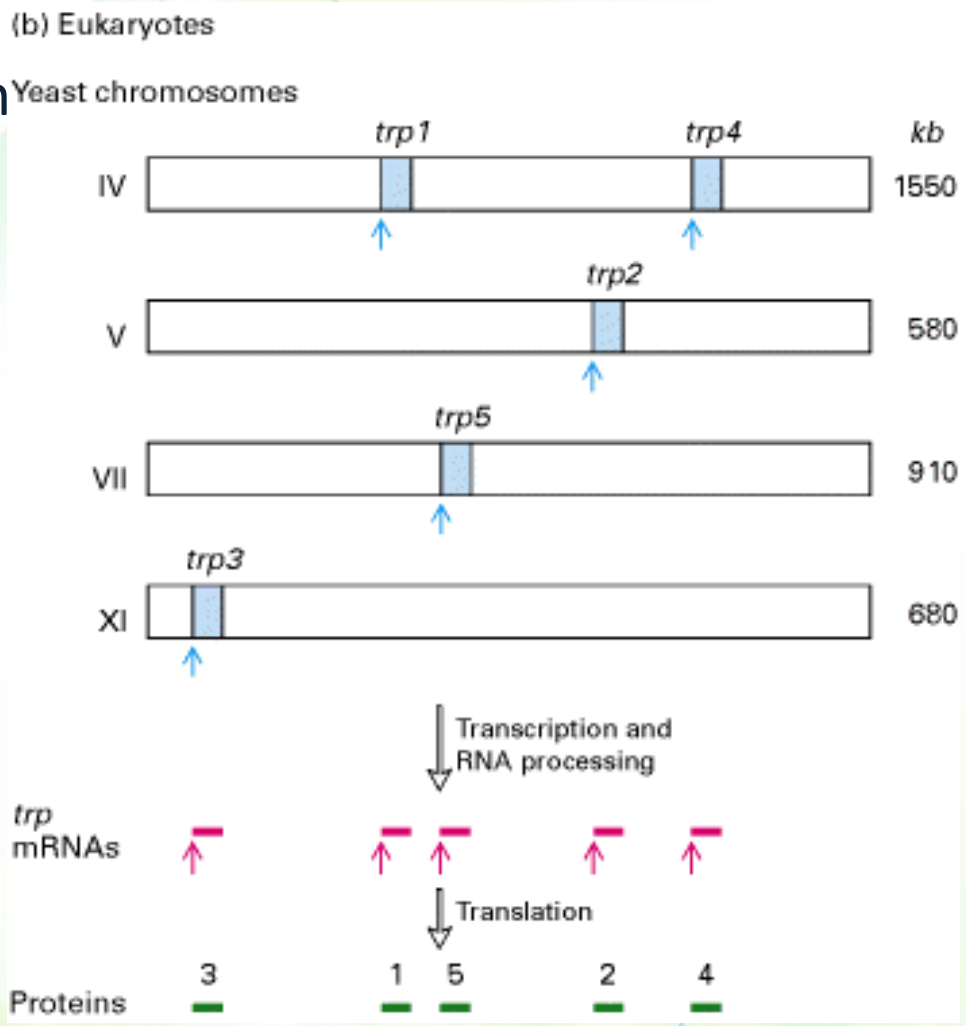
# Prokaryotic genes (operon)

- In bacteria, genes are polycistronic.
- In bacteria, genes that encode enzymes involved in related functions often are located next to each other
  - The genes encoding the enzymes required to synthesize the tryptophan are located in one contiguous stretch
- This cluster of genes comprises a single transcription unit referred to as an operon.



# Eukaryotic genes

- Most eukaryotic transcription units produce mRNAs that encode only one protein, thus termed monocistronic



# Coding DNA in human genome

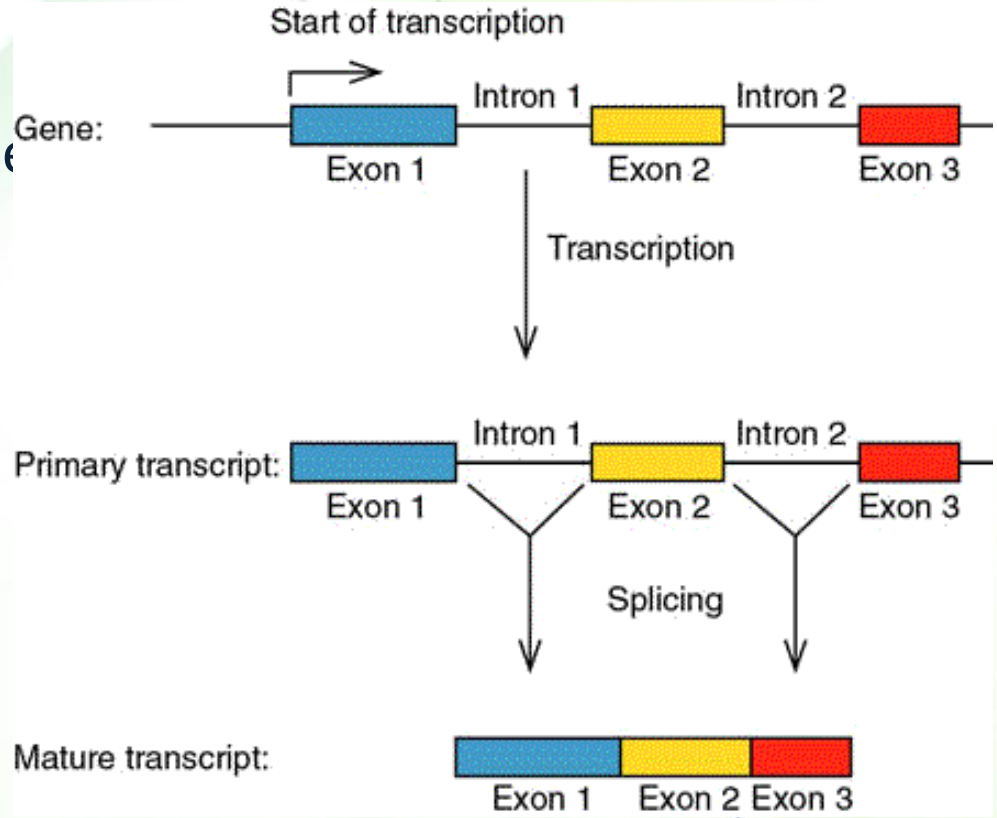
- Most of eukaryotic genomes contain non-protein coding regions
- In humans, for example, only 3% of the human genome contains protein-coding regions

# Introns vs. exons

- The genomes of most eukaryotic cells contain specific DNA sequences that do not code for proteins
  - These pieces of DNA are known as introns
  - The coding regions are known as exons
- When RNA is synthesized, the RNA molecule contains both introns and exons and is known as precursor-mRNA (or pre-mRNA)

# RNA splicing

- The intron sequences are removed from the newly synthesized RNA through the process of RNA splicing
- Now the RNA molecule is known as mRNA

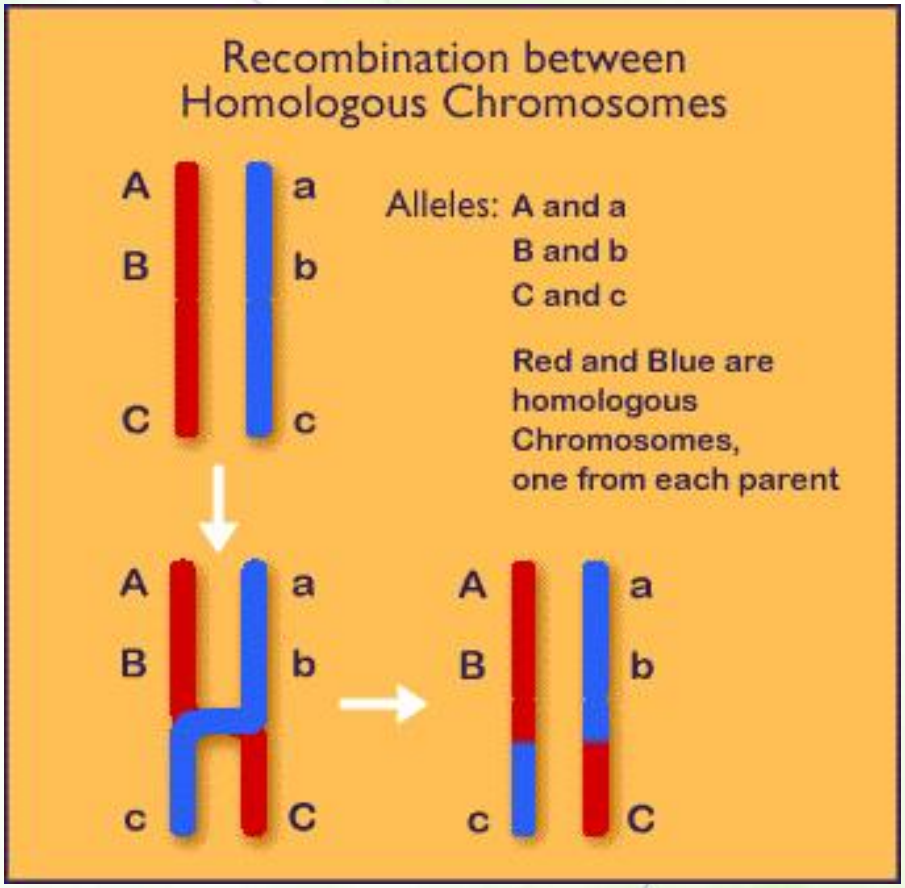


# Importance of introns

- The exon-intron arrangement may facilitate the emergence of new proteins
- How?

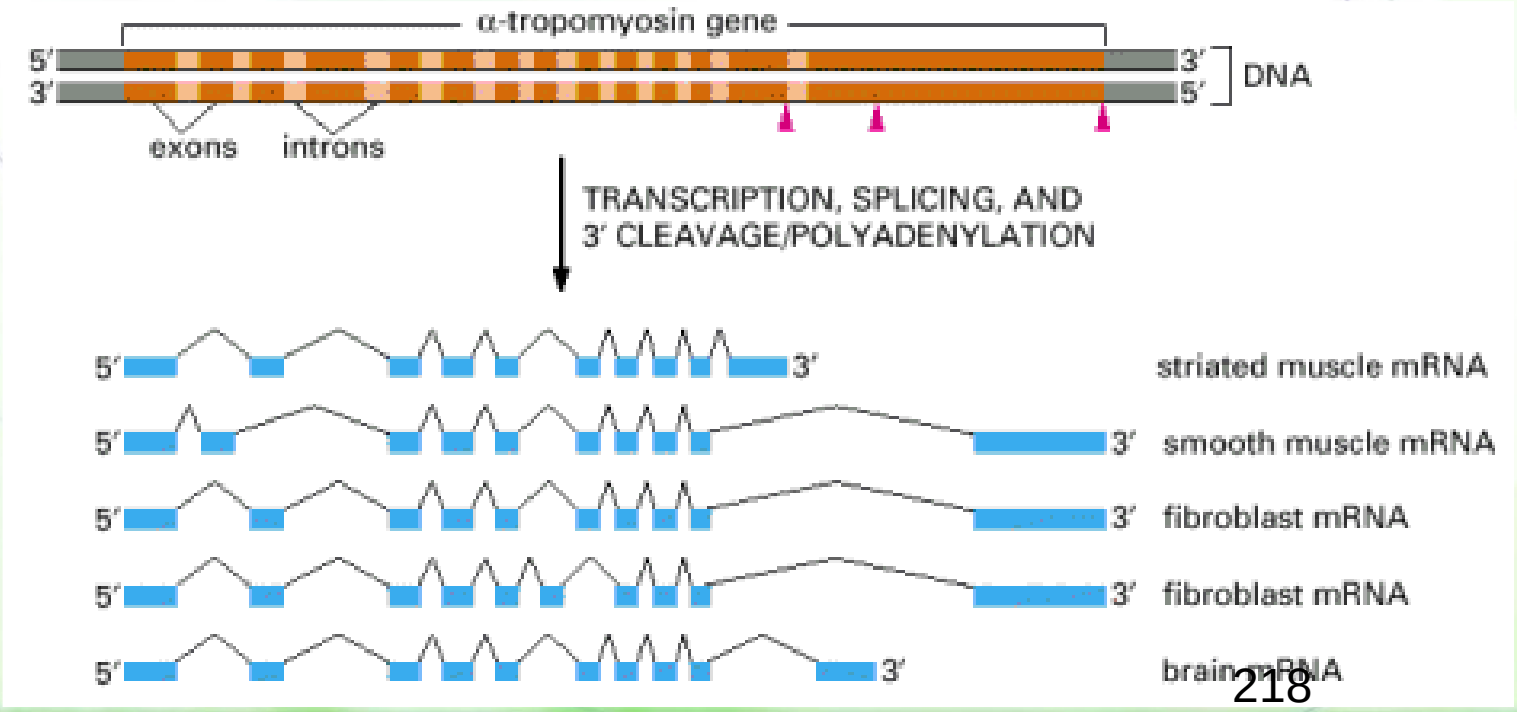
# One...

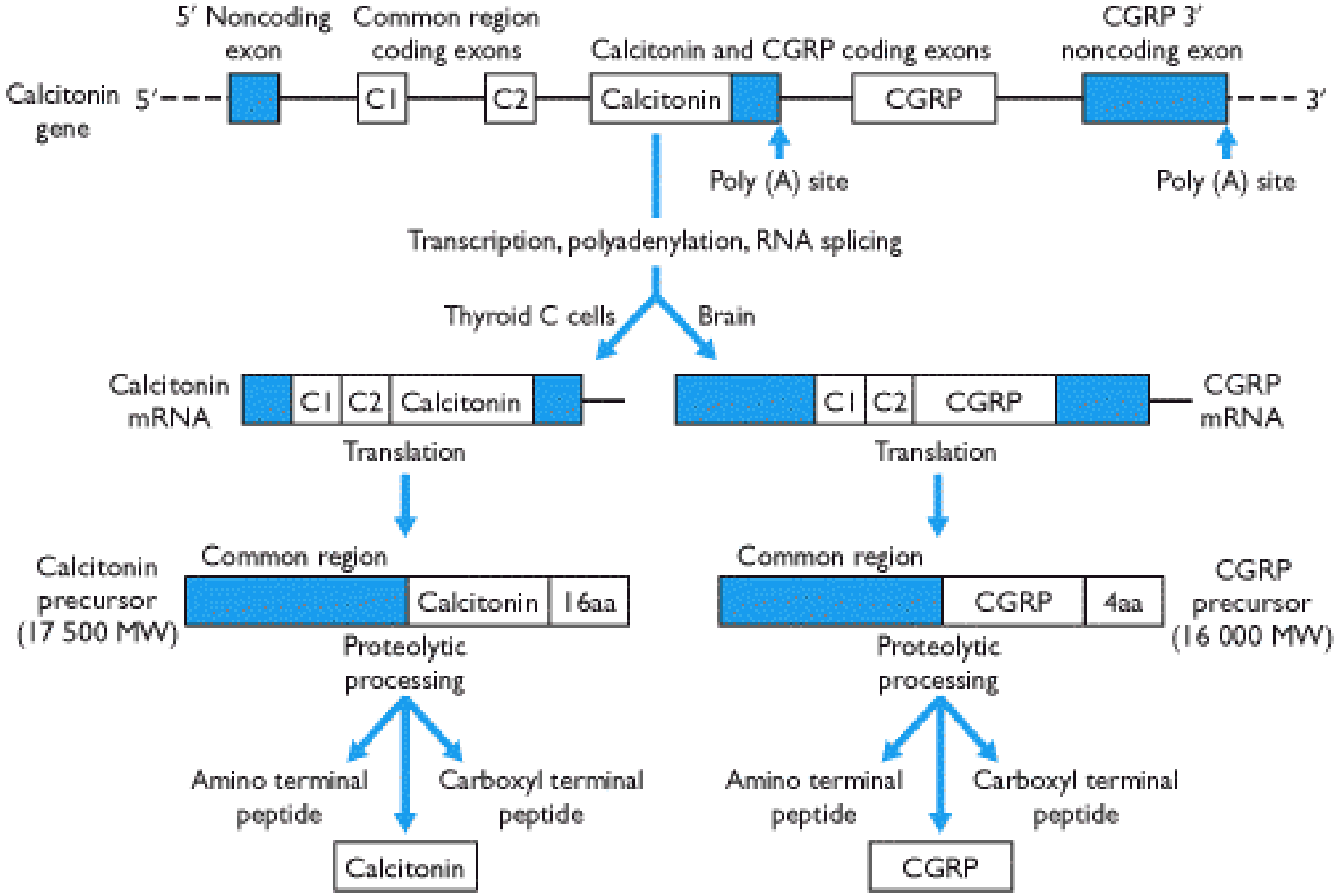
- Introns allow genetic recombination



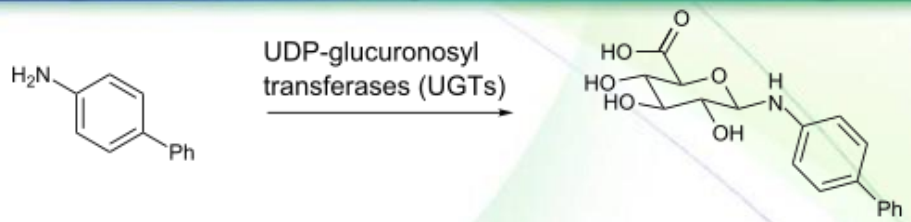
# Alternative splicing

- The transcripts are spliced in different ways to produce different mRNAs and different proteins
  - These are known as protein isoforms
- The process is known as alternative splicing

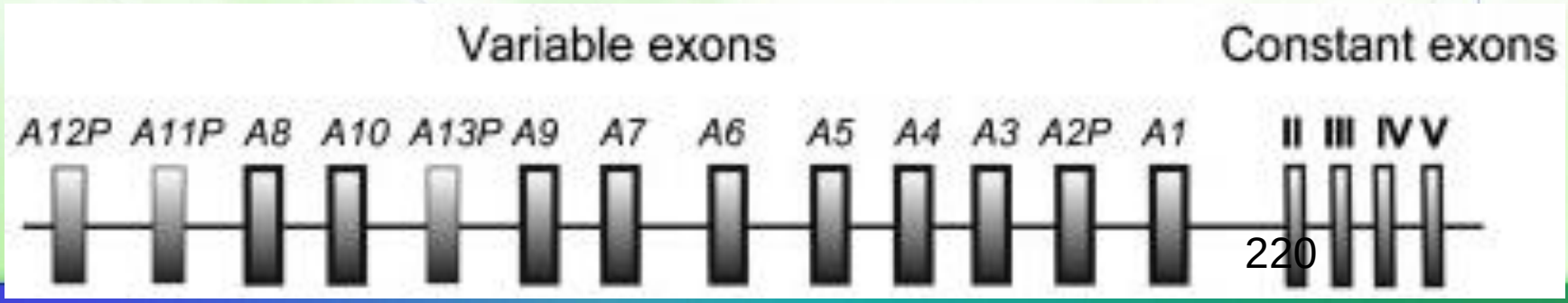



5′ Noncoding exon
Common region coding exons
Calcitonin and CGRP coding exons
CGRP 3′ noncoding exon
Calcitonin gene
5′
C1
C2
Calcitonin
CGRP
3′
Poly (A) site
Poly (A) site
Transcription, polyadenylation, RNA splicing
Thyroid C cells
Brain
Calcitonin mRNA
C1
C2
Calcitonin
C1
C2
CGRP
CGRP mRNA
Translation
Translation
Calcitonin precursor (17 500 MW)
Common region
Calcitonin
16aa
Common region
CGRP
4aa
CGRP precursor (16 000 MW)
Proteolytic processing
Proteolytic processing
Amino terminal peptide
Carboxyl terminal peptide
Amino terminal peptide
Carboxyl terminal peptide
Calcitonin
CGRP

# UDP-glucuronosyltransferase gene




- The 5' region of the UGT1A complex contains 13 tandemly arrayed first exons, (4 pseudo exons and 9 viable)
  - Each first exon has its own promoter element
- Exons 2, 3, 4, and 5 are located in the UGT1A 3' region.
  - All UGT isoforms contain the same C-terminal domain encoded by exons 2 through 5
- The 9 first exons are independently spliced to exon 2 to generate 9 UGT1A transcripts
  - The first exon determines substrate specificity
  - The C-terminal region specifies interactions with UDP-glucuronic acid



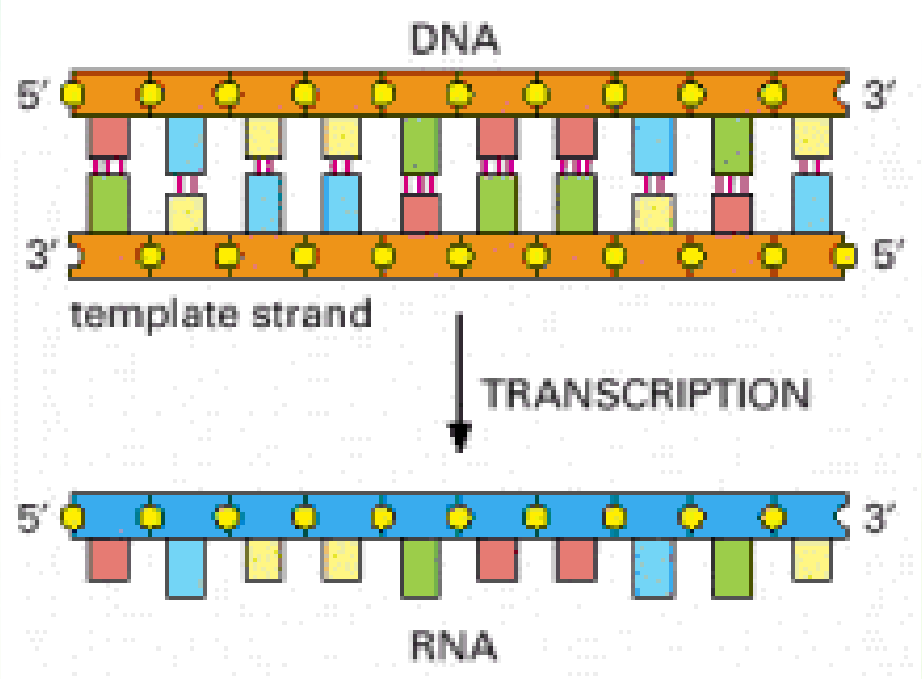
# The general mechanism of transcription

# General description

- Transcription is the process of making RNA from DNA
- One of the two strands of the DNA double helix acts as a template for the synthesis of an RNA molecule

# Complementary sequences

- mRNA is complementary to DNA
- The RNA chain produced by transcription is also known as the transcript

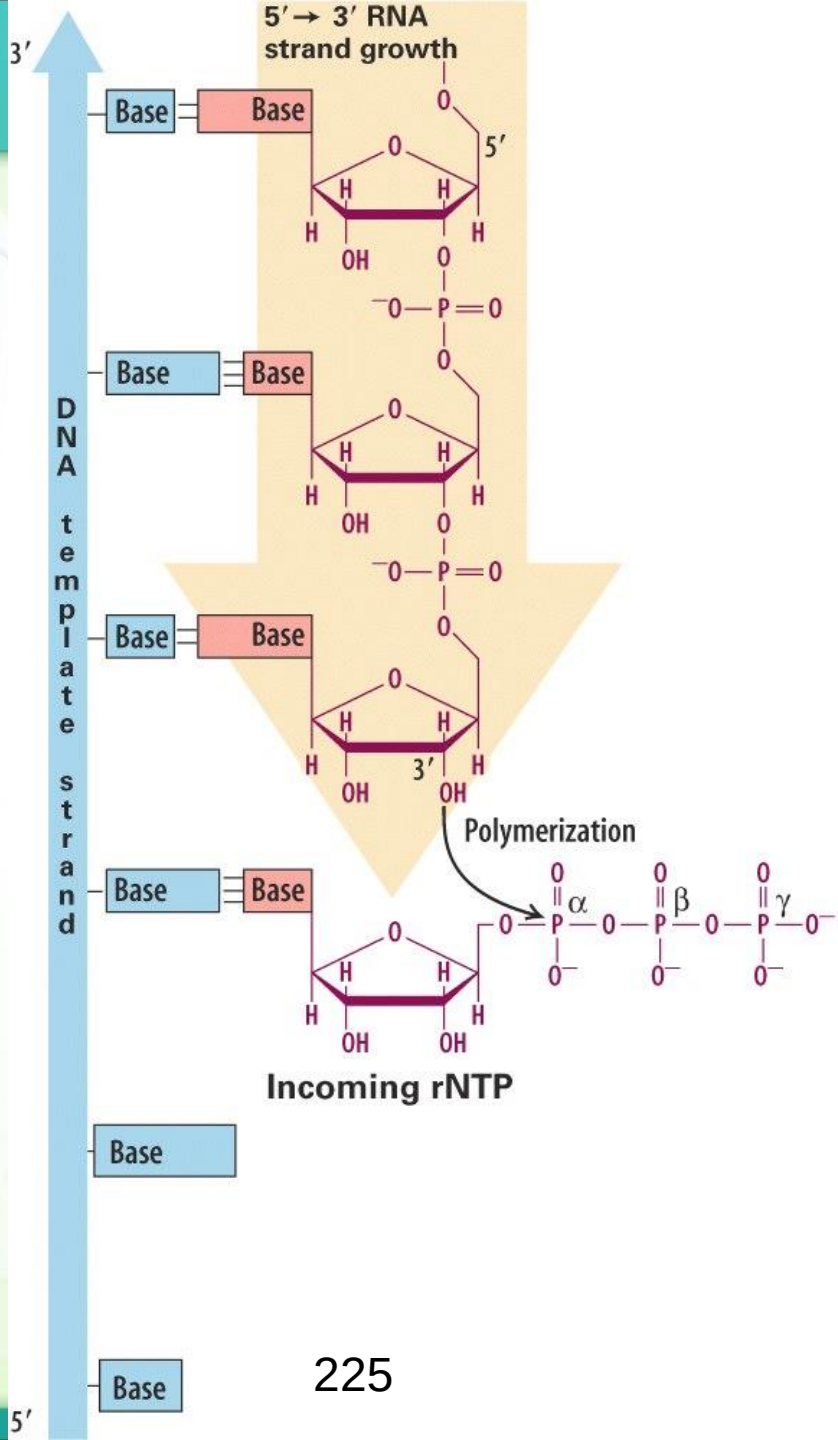


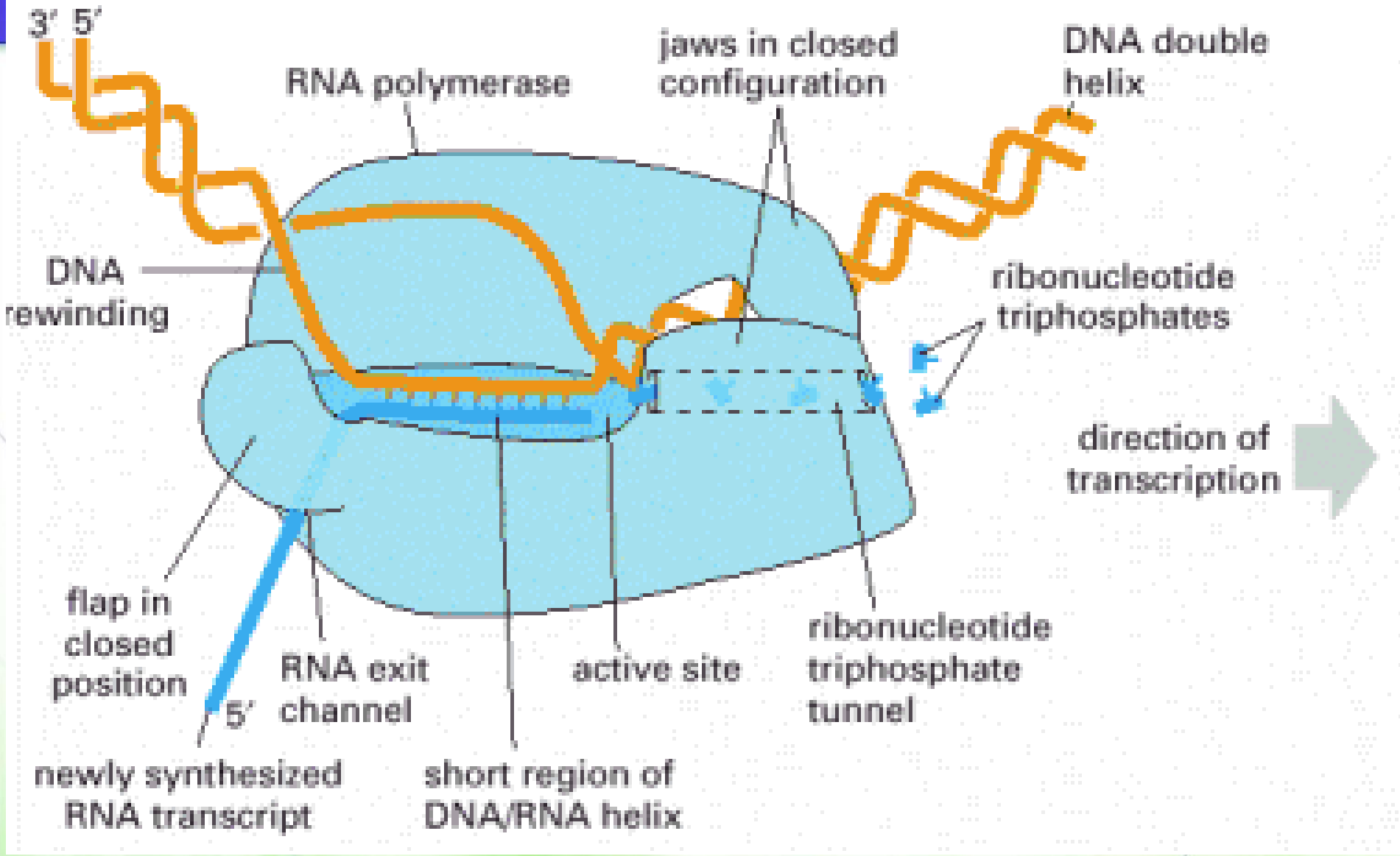
# Enzyme and substrate

- The enzymes that perform transcription are called RNA polymerases
- RNA polymerases catalyze the formation of the phosphodiester bonds between two nucleotides
- The growing RNA chain is extended in the 5-to-3 direction
- The substrates are nucleoside triphosphates (ATP, CTP, UTP, and GTP)

# Energy

- A hydrolysis of high-energy bonds in NTP provides the energy needed to drive the reaction forward

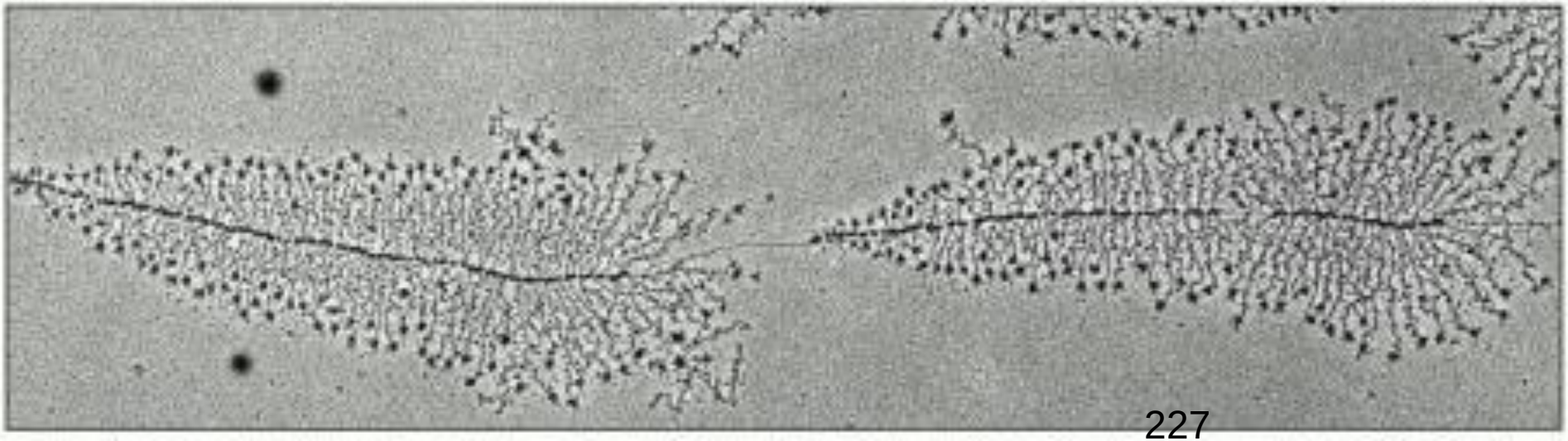



3′ 5′
RNA polymerase
jaws in closed configuration
DNA double helix
DNA rewinding
ribonucleotide triphosphates
direction of transcription
flap in closed position
RNA exit channel
active site
ribonucleotide triphosphate tunnel
5′
newly synthesized RNA transcript
short region of DNA/RNA helix

# Polysomes

- As RNA is synthesized, it is initially bonded to DNA, but after a short distance, the older polymerized RNA nucleotides are separated, and the newer ones become bonded

- This allows the simultaneous synthesis of many RNA chains from the same gene forming structures known as polysomes



# DNA replication vs. transcription

- The RNA strand does not remain hydrogen-bonded to the DNA template strand
- RNA polymerase read the A in DNA and inserts U in the growing chain of RNA rather than T
- RNA molecules are much shorter than DNA molecules
- Unlike DNA, RNA does not store genetic information in cells

# DNA polymerase vs. RNA polymerase

- RNA polymerase catalyzes the linkage of ribonucleotides, not deoxyribonucleotides
- Unlike DNA polymerases, RNA polymerases can start an RNA chain without a primer
- RNA polymerases make about one mistake for every 104 nucleotides copied into
  - the consequences of an error in RNA transcription are much less significant than that in DNA replication

# Exonuclease activity

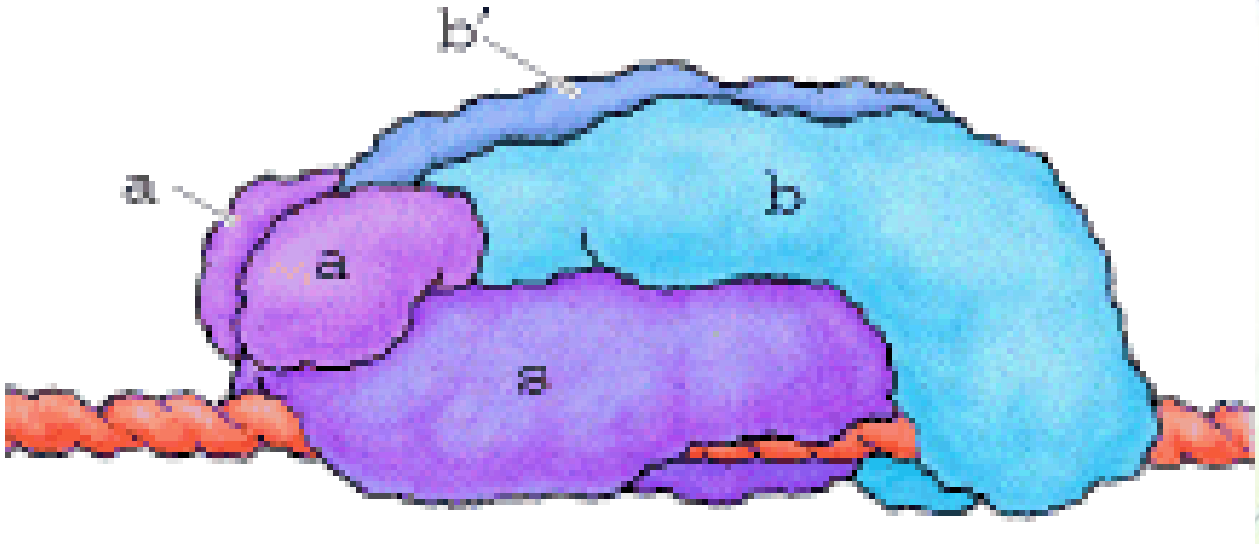
- Although RNA polymerases are not as accurate as the DNA polymerases, they have a modest proofreading mechanism

# Transcription in prokaryotes

# The RNA polymerase

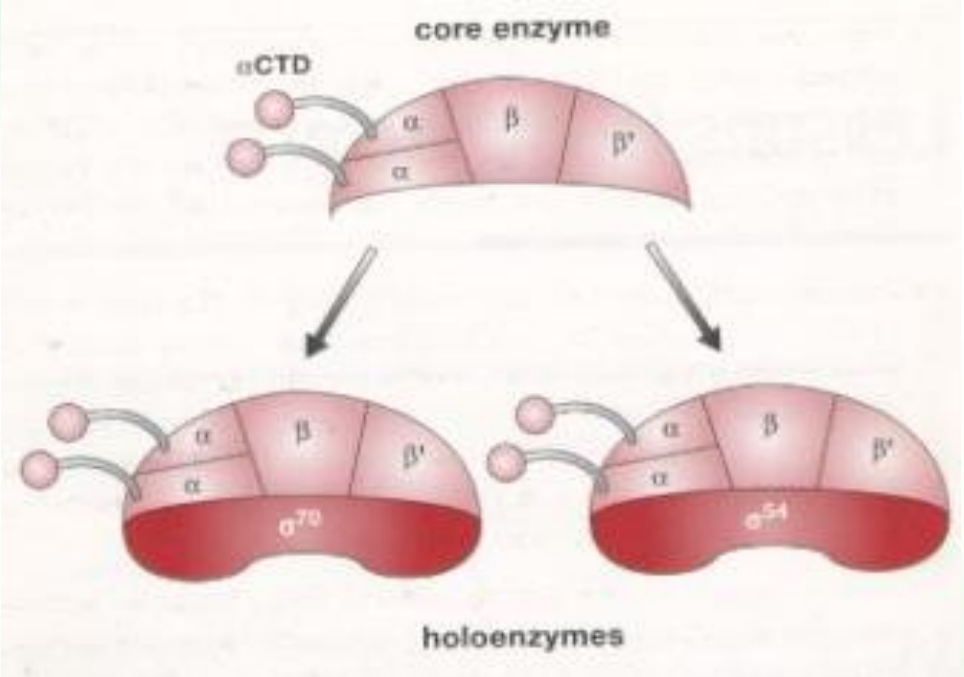
- E. coli RNA polymerase is made up of multiple polypeptide chains

- The intact enzyme consists of four different types of subunits, called $\alpha$, $\beta$, $\beta'$, and $\sigma$



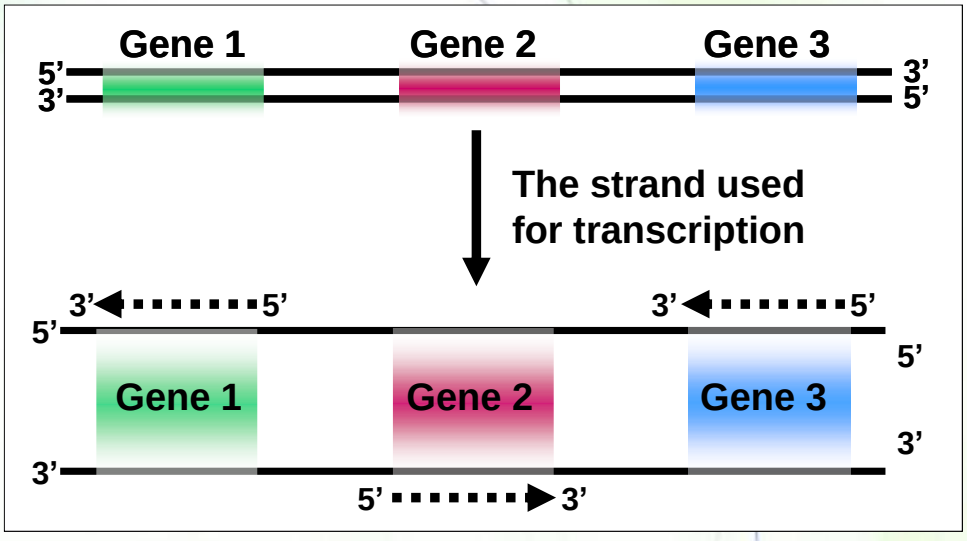
# The core polymerase

- The core polymerase consists of two β, one β, and one β' subunits
- The core polymerase is fully capable of catalyzing the polymerization of NTPs into RNA
  - σ is not required for the basic catalytic activity of the enzyme



# Using DNA strands

- Although both enzymes can read both DNA strands, RNA polymerase uses one strand at a time in order to make a RNA molecule.
- The transcribed DNA strand = template, anti-sense, (-) strand
  - The other strand: sense, (+)strand, coding strand

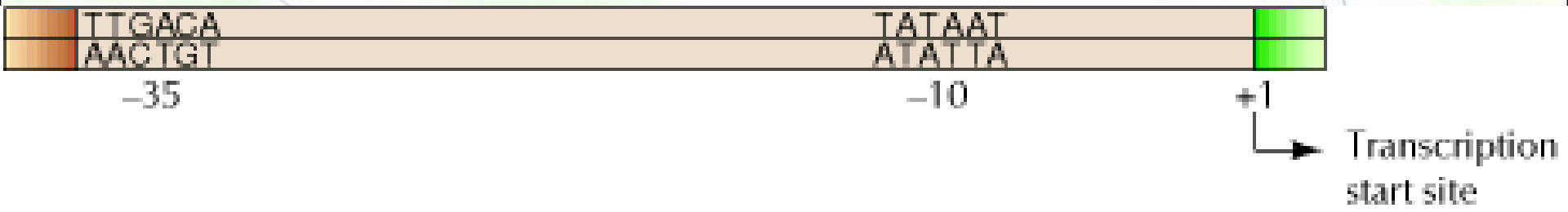


# Consensus sequences (the promoter)

- The DNA sequence to which RNA polymerase binds to initiate transcription of a gene is called the promoter
  - A promoter is "upstream" of the transcription initiation site
- The region upstream of the transcription initiation site contains two sets of sequences that are similar in a variety of genes

# The regions

- These common sequences are located approximately 10 and 35 base pairs upstream of the transcription start site
- They are called the (-10) and (-35) elements
- The transcription initiation site is defined as the +1 position
  - Open reading frame: DNA sequence that can be transcribed into mRNA from first base to last one



# How do we know they are important?

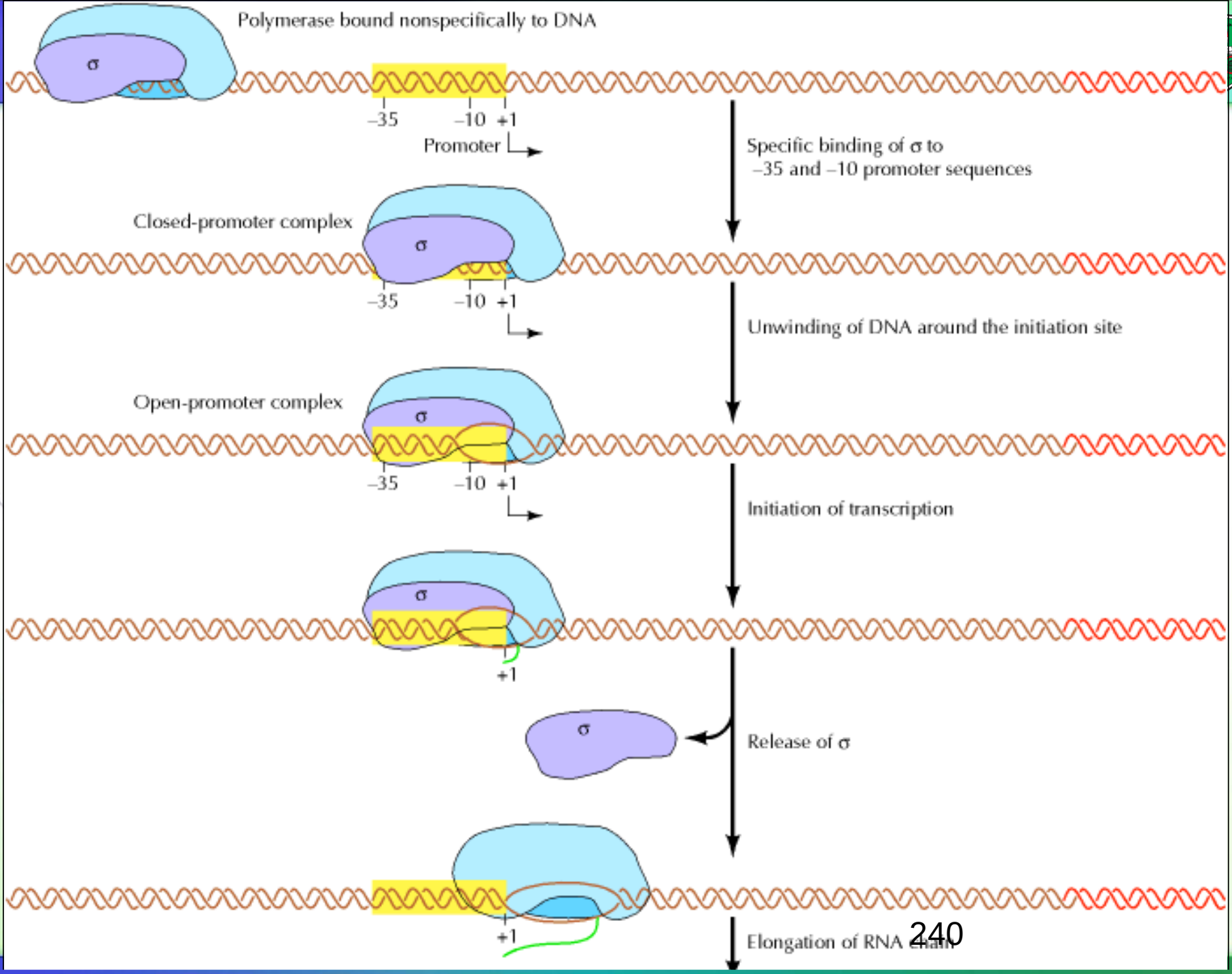
- Genes with promoters that differ from the consensus sequences are transcribed less efficiently than genes whose promoters match the consensus sequences
- Mutations introduced in either the -35 or -10 consensus sequences have strong effects on promoter function
- RNA polymerase generally binds to promoters over approximately a 60-base-pair region, extending from -40 to +20.
- The $\sigma$ subunit binds specifically to sequences in both the -35 and -10 promoter regions

# Role of the σ subunit

- In the absence of σ , RNA polymerase binds to DNA with low affinity and nonspecifically
- The role of σ is to identify the correct sites for transcription initiation and direct the polymerase to promoters by binding specifically to both the -35 and -10 sequences

# Mechanism of transcription (initiation)

- The binding between the polymerase and a promoter is referred to as a closed-promoter complex
- The polymerase unwinds approximately 15 bases of DNA to form an open-promoter complex
- Single-stranded DNA is available as a template
- Transcription is initiated by the joining of two NTPs
- After addition of about the first 10 nucleotides, σ is released from the polymerase

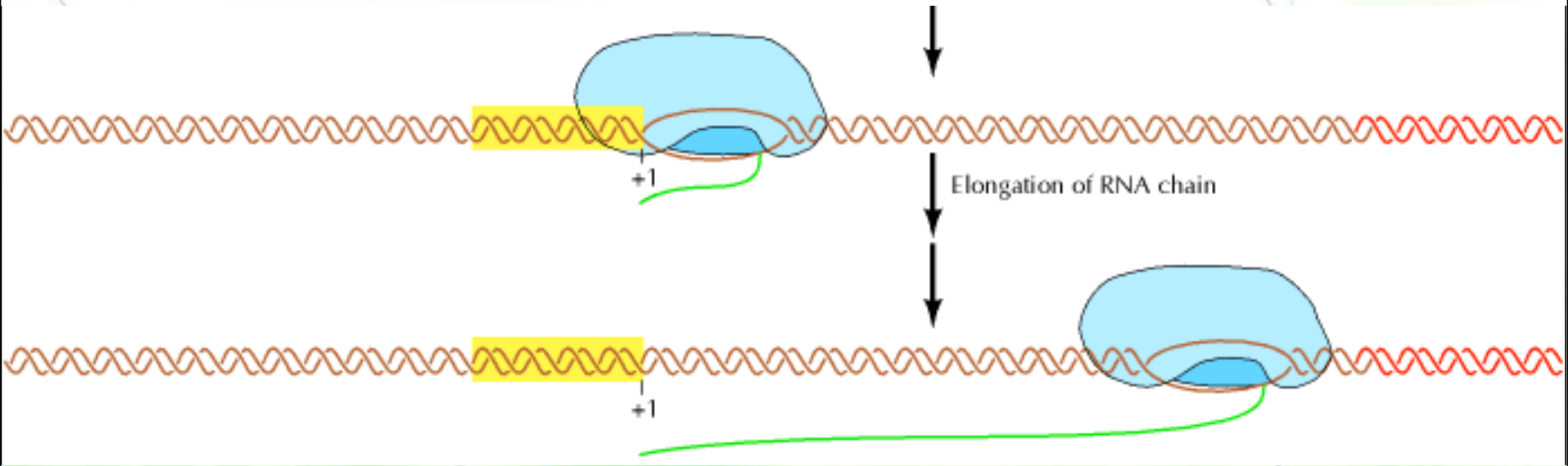

Polymerase bound nonspecifically to DNA
σ
–35
–10
+1
Promoter
Specific binding of σ to
–35 and –10 promoter sequences
Closed-promoter complex
σ
–35
–10
+1
Unwinding of DNA around the initiation site
Open-promoter complex
σ
–35
–10
+1
Initiation of transcription
σ
+1
σ
Release of σ
+1
Elongation of RNA chain

# Mechanism of transcription (elongation)

- As the polymerase moves forward, it
  - unwinds the template DNA ahead of it
  - elongates the RNA
  - rewinds the DNA behind it

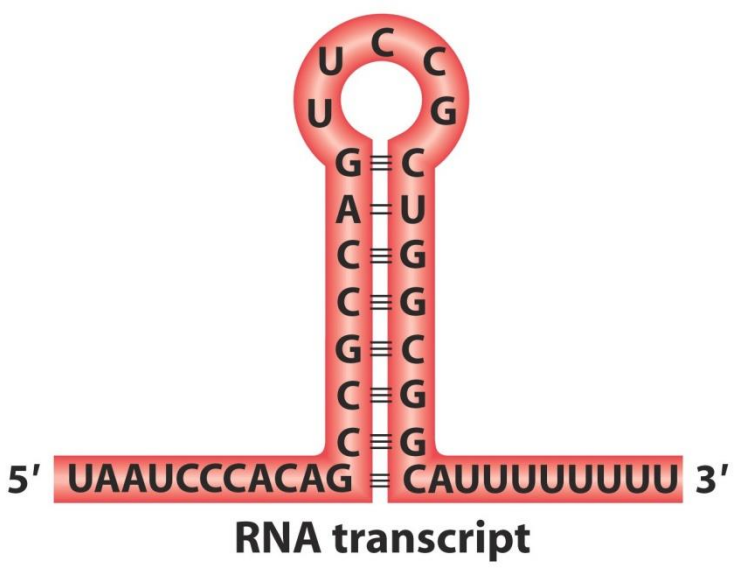
# Mechanism of transcription (termination)

- RNA synthesis continues until the polymerase encounters a termination signal where the RNA is released from the polymerase, and the enzyme dissociates from its DNA template



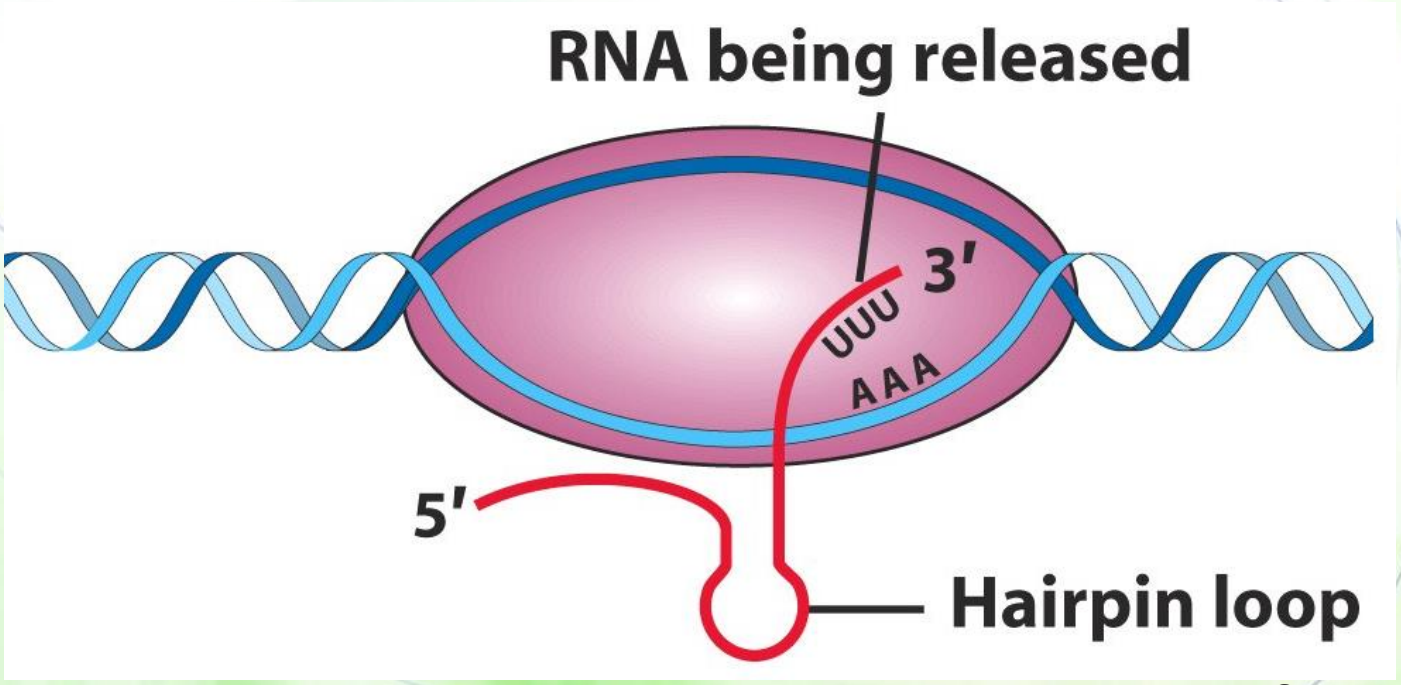
# Termination sequences

- The simplest and most common type of termination signal in E. coli consists of a symmetrical inverted repeat of a GC-rich sequence followed by A residues
- Transcription of the GC-rich inverted repeat results in the formation of a stable stem-loop structure



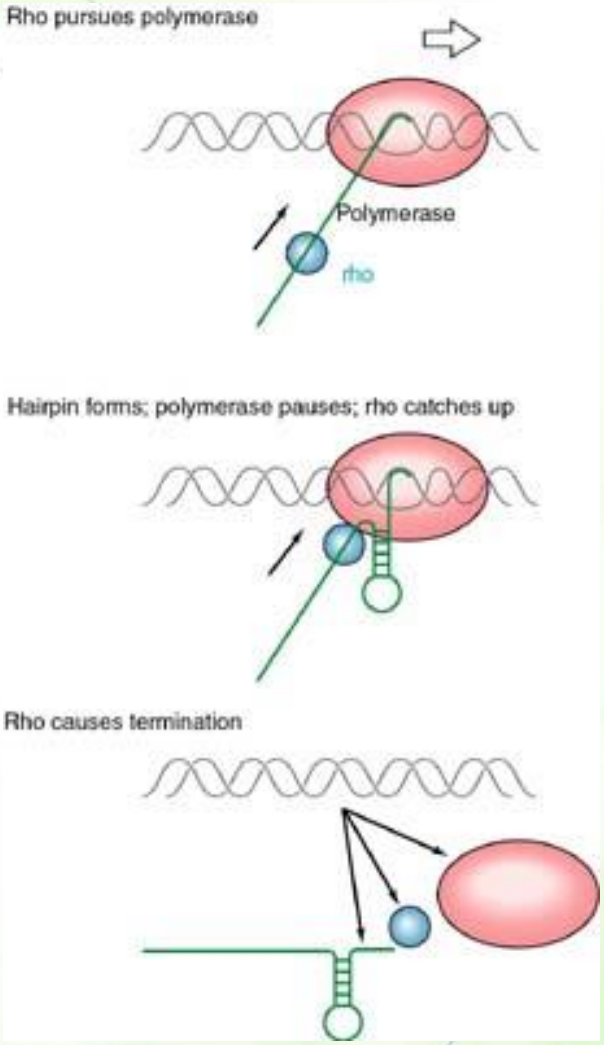
# The effect of the stem loop structure

- The formation of this structure breaks RNA association with the DNA template, destabilizes the RNA polymerase binding to DNA, and terminates transcription



# Rho-dependent terminator

- Rho is a helicase that follows the RNA polymerase along the transcript. When the polymerase stalls at a hairpin, Rho catches up and breaks the RNA-DNA base pairs, releasing the transcript.
- Rho-dependent termination signals do not have the string of U residues at the end of the RNA.



# Transcription in eukaryotes

# RNA polymerases

- In contrast to bacteria, which contain a single type of RNA polymerase, eukaryotic nuclei have three, called RNA polymerase I, RNA polymerase II, and RNA polymerase III
  - RNA polymerase I transcribes rRNA genes
  - RNA polymerase II transcribes protein-encoding genes
  - RNA polymerase III transcribes tRNA genes and one rRNA gene

# Eukaryotic RNA polymerases

- Eukaryotic transcription initiation must deal with the packing of DNA into nucleosomes
- While bacterial RNA polymerase is able to initiate transcription without the help of additional proteins, eukaryotic RNA polymerases cannot.
  - They require the help of a large set of proteins called general transcription factors

# General transcription factors

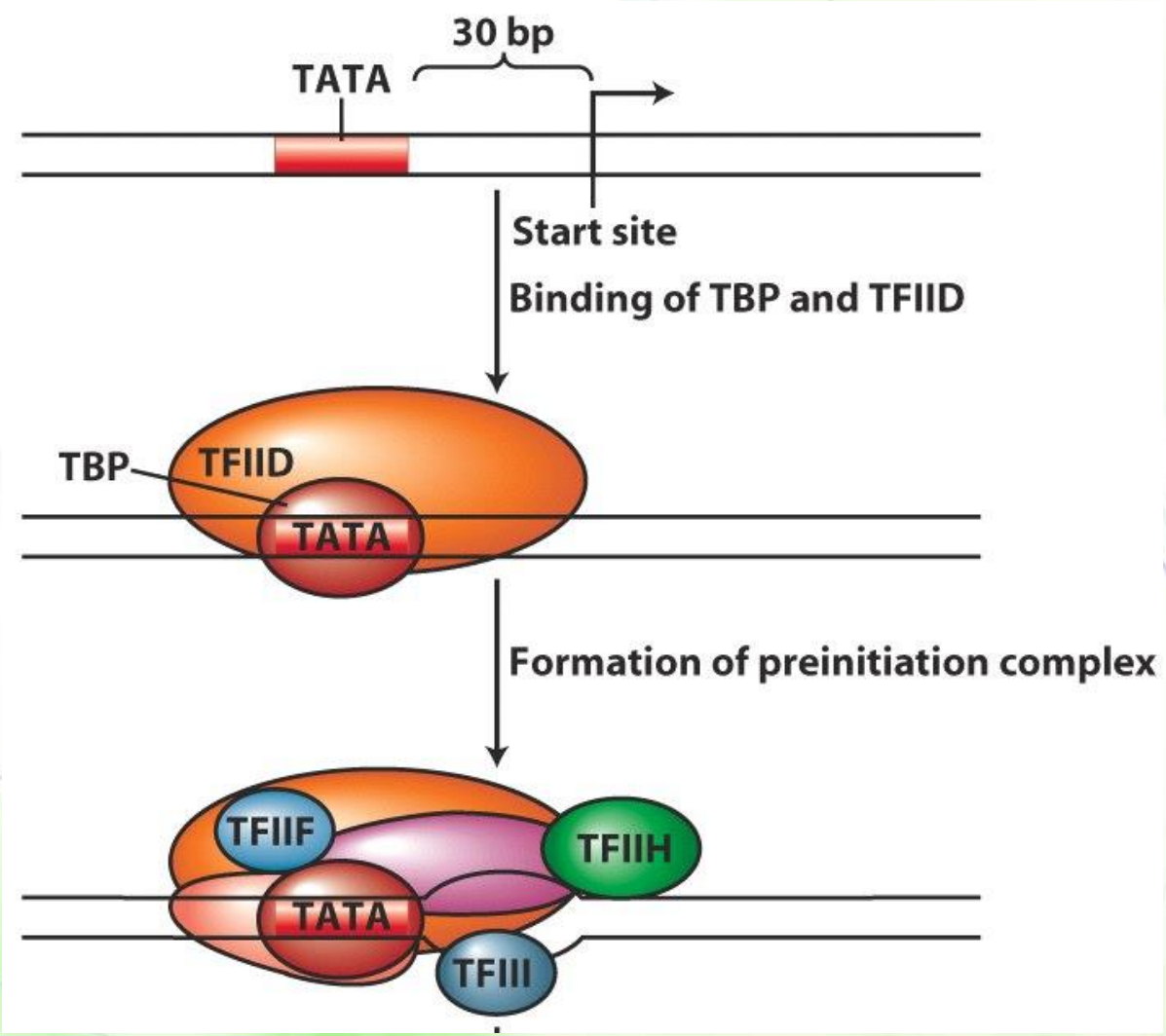
- These general transcription factors
  - help position the RNA polymerase correctly at the promoter
  - aid in pulling apart the two strands of DNA to allow transcription to begin
  - push the RNA polymerase forward to begin transcription

# Why are they general?

- The proteins are "general" because they assemble on all promoters used by RNA polymerase II
- They are designated as TFII (for transcription factor for polymerase II), and listed as TFIIA, TFIIB, and so on

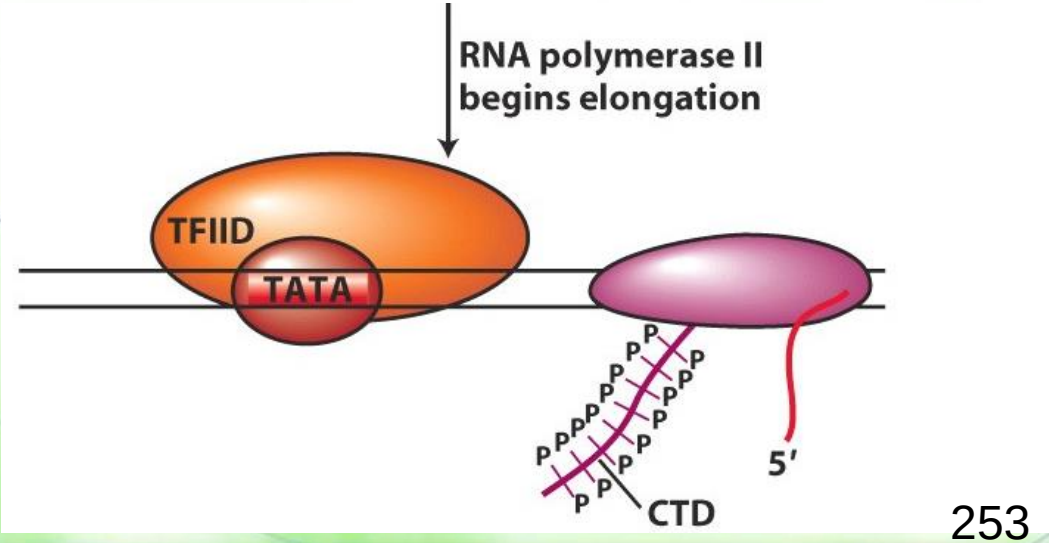
# Mechanism of transcription (elongation)

- TFIID binds to a TATA box located upstream from the transcription start site
  - The binding of TFIID causes a bend in the DNA of the TATA box
  - This bend attracts other proteins to assemble on the promoter
  - Along with RNA polymerase II, these protein factors form a transcription initiation complex
- One of them is TFIIH, which contains a DNA helicase.
  - TFIIH creates an open promoter complex exposing the DNA template to the RNA polyemerase


30 bp
TATA
Start site
Binding of TBP and TFIID
TBP
TFIID
TATA
Formation of preinitiation complex
TFIIF
TFIIH
TATA
TFIII

# Mechanism of transcription (elongation)

- Movement of the polymerase is activated by the addition of phosphate groups to the "tail" of the RNA polymerase.
- This phosphorylation is also catalyzed by TFIIH, which, also contains a protein kinase subunits

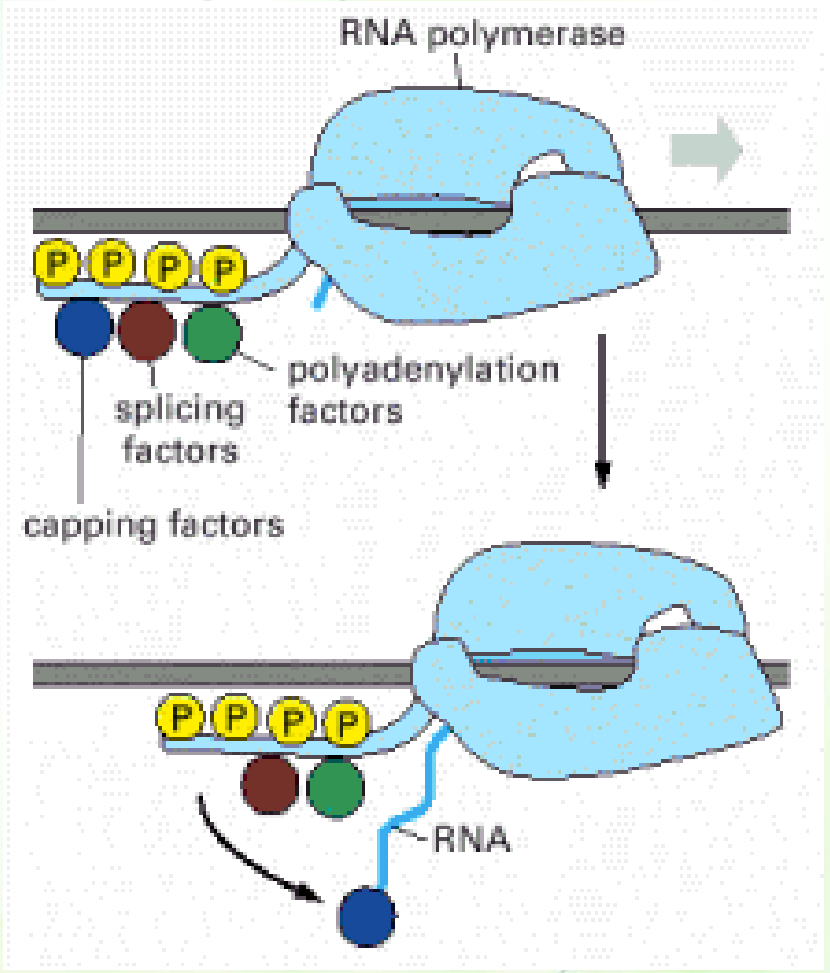


# Mechanism of transcription (termination)

- Termination begins by stopping the RNA polymerase. There is a eukaryotic consensus sequence for termination, which is AAUAAA.
- After termination occurs, the transcript is released, and the phosphorylated Pol II is released from the DNA.
- The phosphates are removed by phosphatases, and Pol II is recycled for another round of transcription.
- Termination is coupled to the process that cleaves and polyadenylates the 3 end of a transcript.

# Phosphorylation of RNA polymerase II

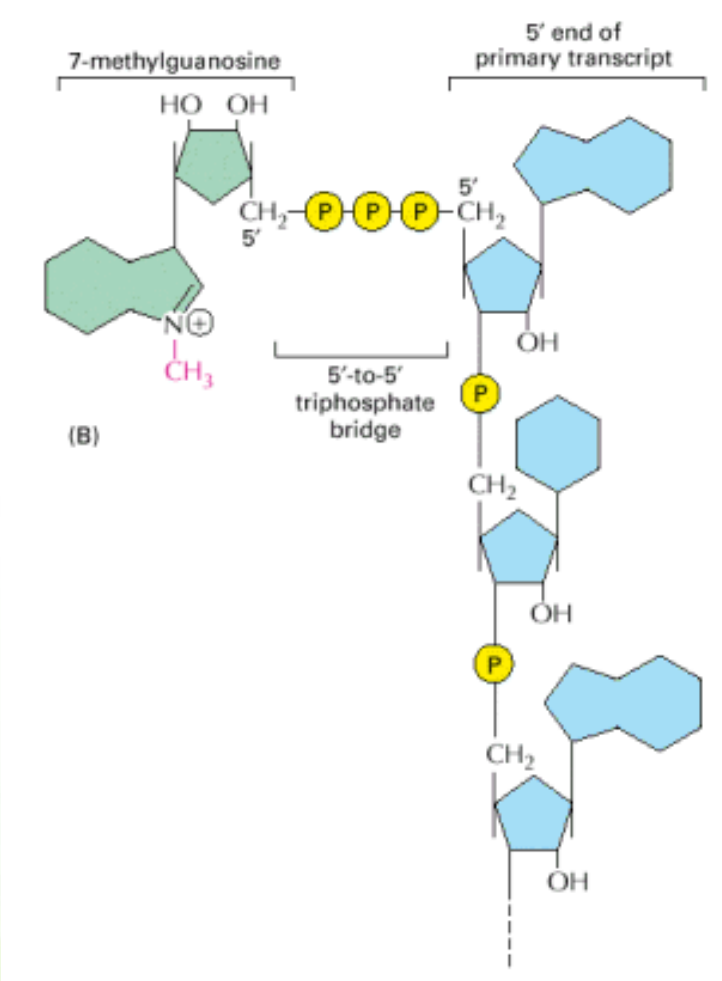
- RNA is processed and modified extensively
  - Capping
  - Splicing
  - Polyadenylation
- Some of these processing proteins are associated with the tail of RNA polymerase II
- These proteins jump from the polymerase tail onto the RNA molecule as it appear



# Addition of a cap

- As soon as RNA polymerase II has produced about 25 nucleotides of RNA, the 5' end of the new RNA molecule is modified by addition of a "cap" that consists of a modified guanine nucleotide

- The guanine is added in a reverse linkage (5’ to 5’ instead of 5’ to 3’)

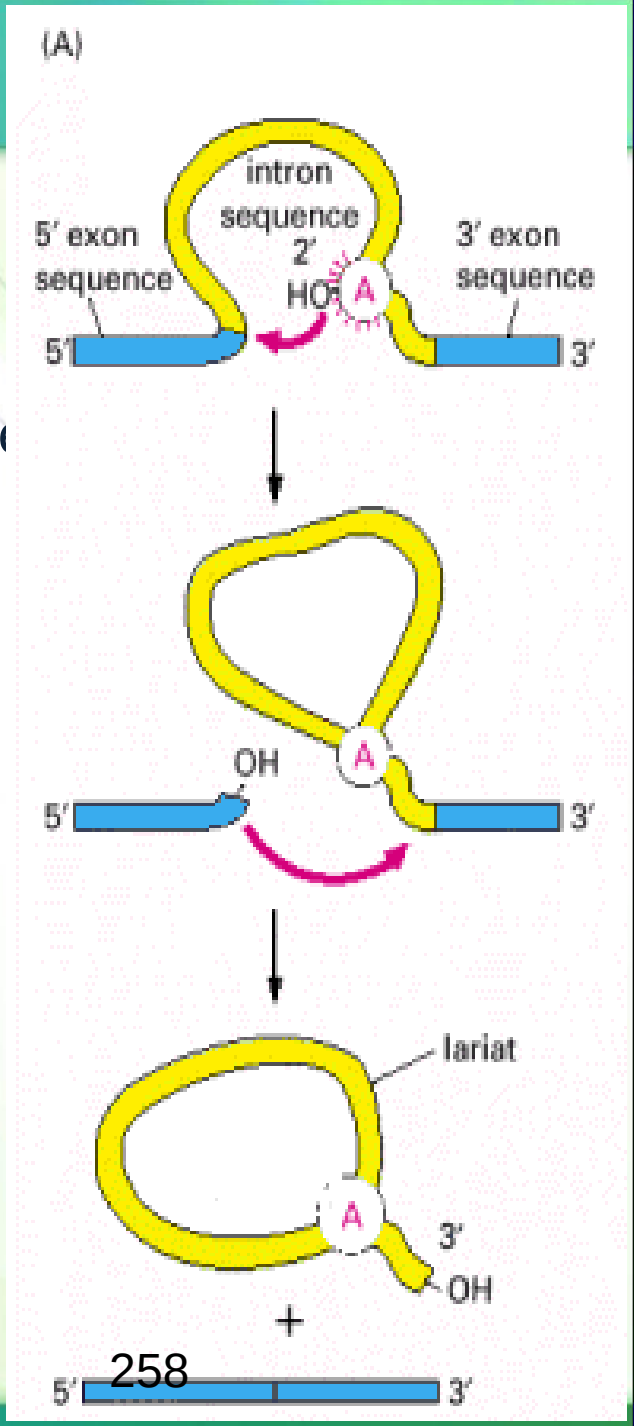


# Importance of capping

- The 5’-methyl cap signals the 5’ end of eukaryotic mRNAs
  - this helps the cell to distinguish mRNAs from the other types of RNA molecules, which are uncapped
- In the nucleus, the cap binds a protein complex called CBC (cap-binding complex), which helps the RNA to be exported into the cytoplasm
- The 5’-methyl cap also has an important role in the translation of mRNAs to proteins

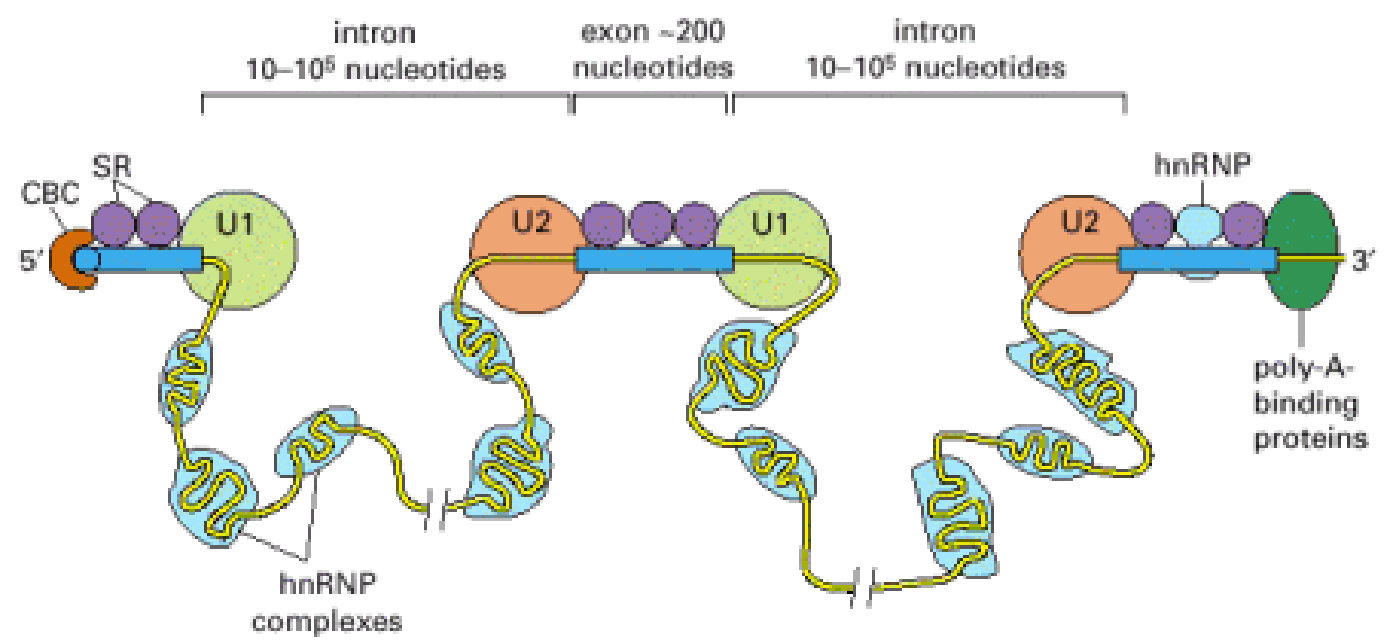
# RNA splicing

- The machinery that catalyzes pre-mRNA splicing consists of 5 RNA molecules and ove 50 proteins.
  - The RNA molecules are known as snRNAs (small nuclear RNAs)
  - Each one of them is complexed with protein subunits to form a snRNP (small nuclear ribonucleoprotein)
  - These snRNPs form the core of the spliceosome, the assembly of RNA and proteins that perform pre-mRNA splicing
  - The catalytic site itself is largely formed by RNA molecules instead of proteins



# hnRNP

- Another class of proteins that assemble on pre-mRNA is hnRNPS (heterogeneous nuclear ribonuclear proteins)
  - hnRNP particles bind to introns
  - They have different functions

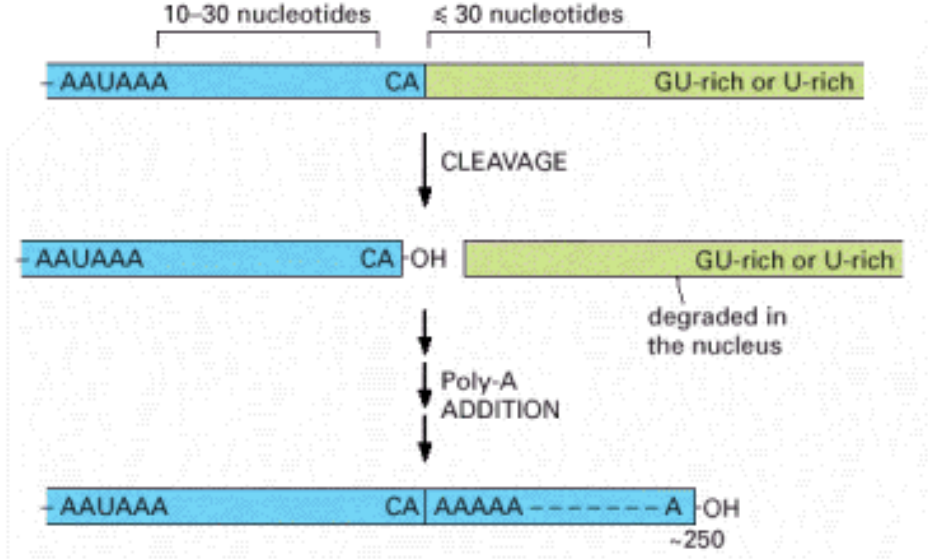


# Accuracy of splicing

- The consistent exon size (more uniform than introns)
- The assembly of the spliceosome occurs as the pre-mRNA emerges from the RNA polymerase II
- As RNA synthesis proceeds, spliceosome components, called the SR proteins, mark the 3’ and 5’ splice site
- hnRNPs define introns
- Spliceosome assembly is co-transcriptional, but splicing occurs post-transcriptionally

# Polyadenylation

- A certain sequence in the mRNA (AATAA) in the 3’ ends of mRNAs is recognized by RNA-binding proteins and RNA-processing enzymes that cleave the RNA.
- Poly-A polymerase adds ~200 A nucleotides to the 3’ end produced by the cleavage.
  - The nucleotide precursor for these additions is ATP



# Poly-A polymerase

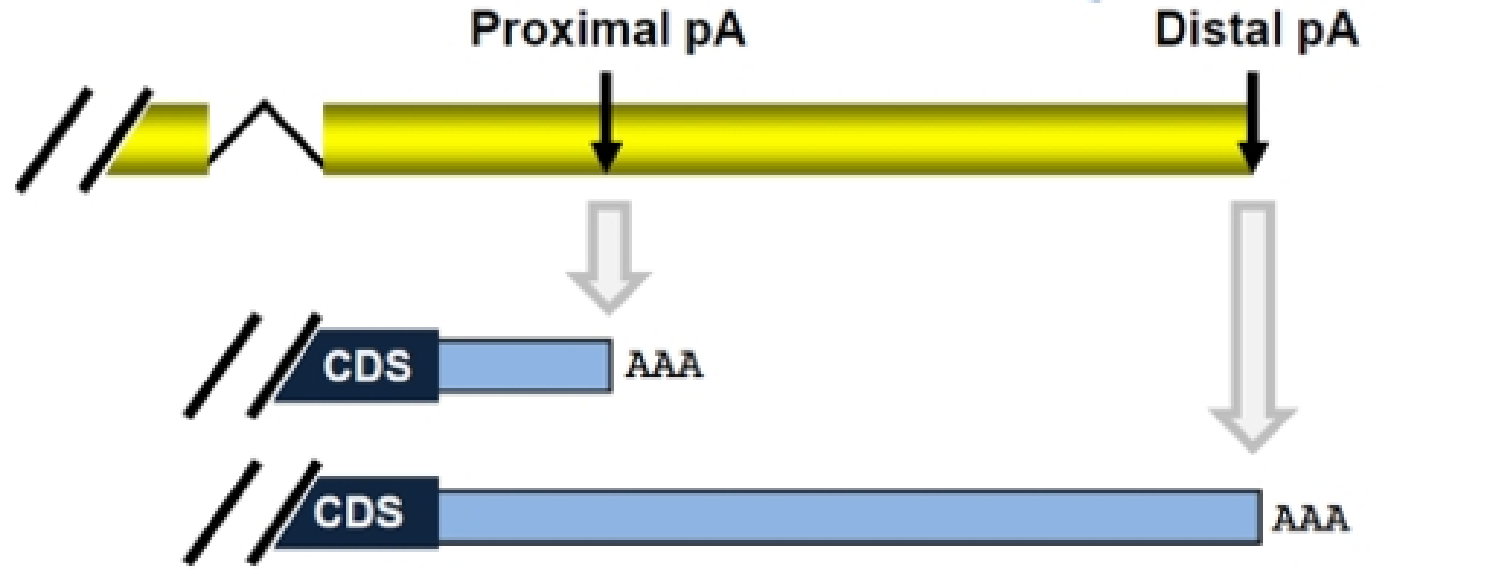
- Poly-A polymerase does not require a template
- hence the poly-A tail of eukaryotic mRNAs is not directly encoded in the genome

# Poly-A-binding proteins

- Poly-A-binding proteins bind to the poly-A tail
  - Help in transporting mRNA from the nucleus to the cytosol
  - Help in protein synthesis
  - Stabilize mRNA

# Alternative polyadenylation

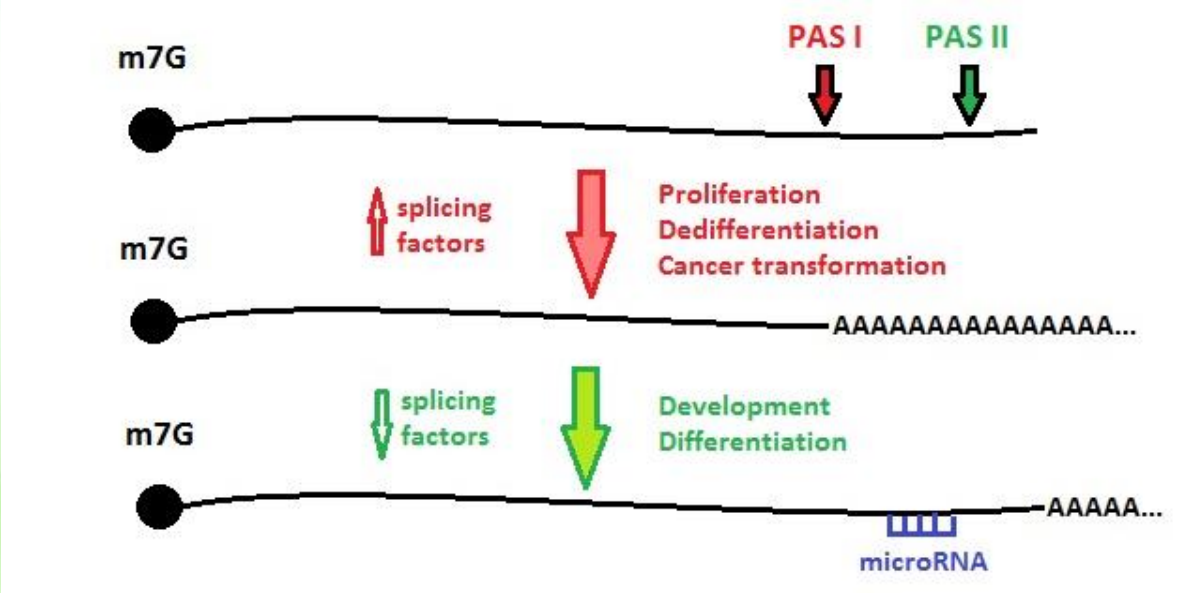
- Many protein-coding genes have more than one polyadenylation site, producing mRNAs with different lengths of a noncoding sequence at the 3’-end called the 3‘-untranslated region (3’-UTR).



# Alternative polyadenylation

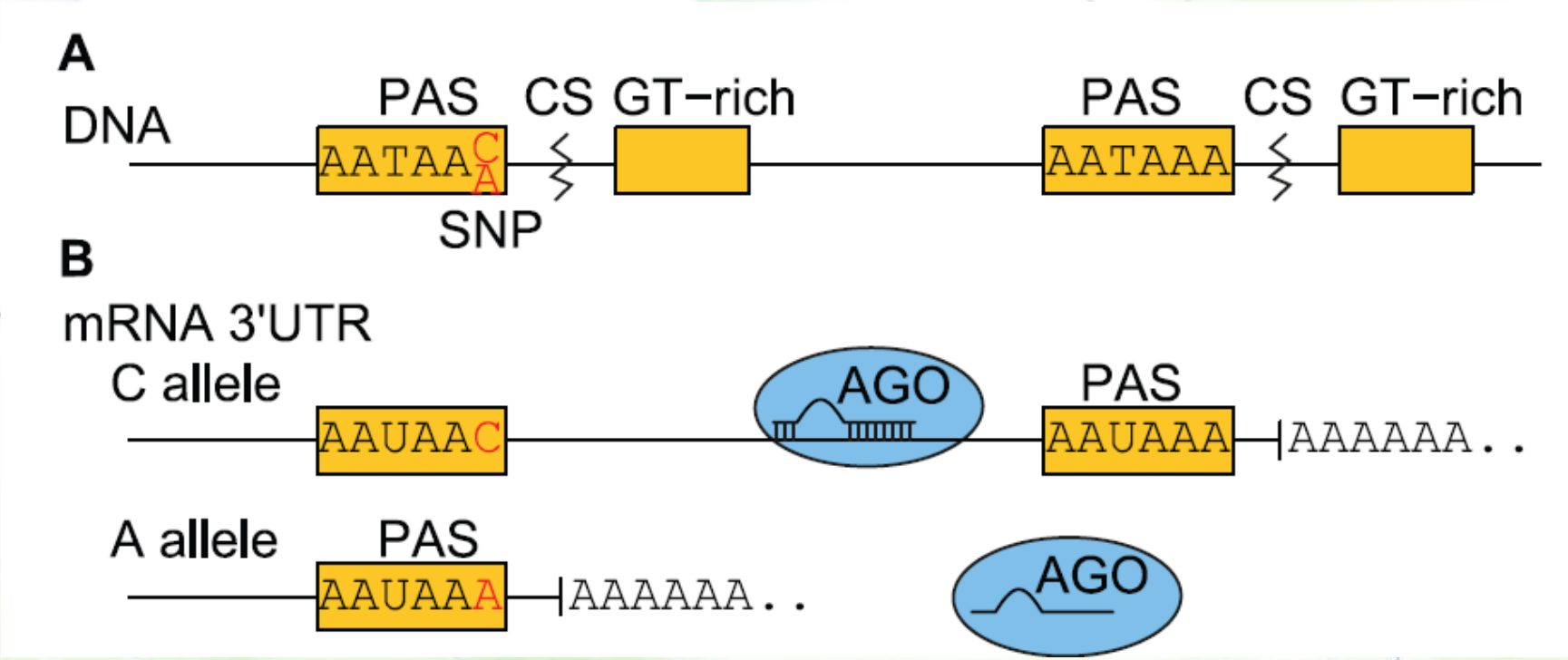
- The choice of poly(A) site can be influenced by extracellular stimuli that regulate the expression of the proteins that take part in polyadenylation.
- Having a shorter transcripts would remove regulatory elements in the 3‘-UTR and influence the half-lives of mRNA and, hence the amount of generated proteins.

- Example: Longer 3’-UTR would contain binding sites for microRNAs at the 3‘-UTR, which tend to repress translation and promote degradation of the mRNAs.



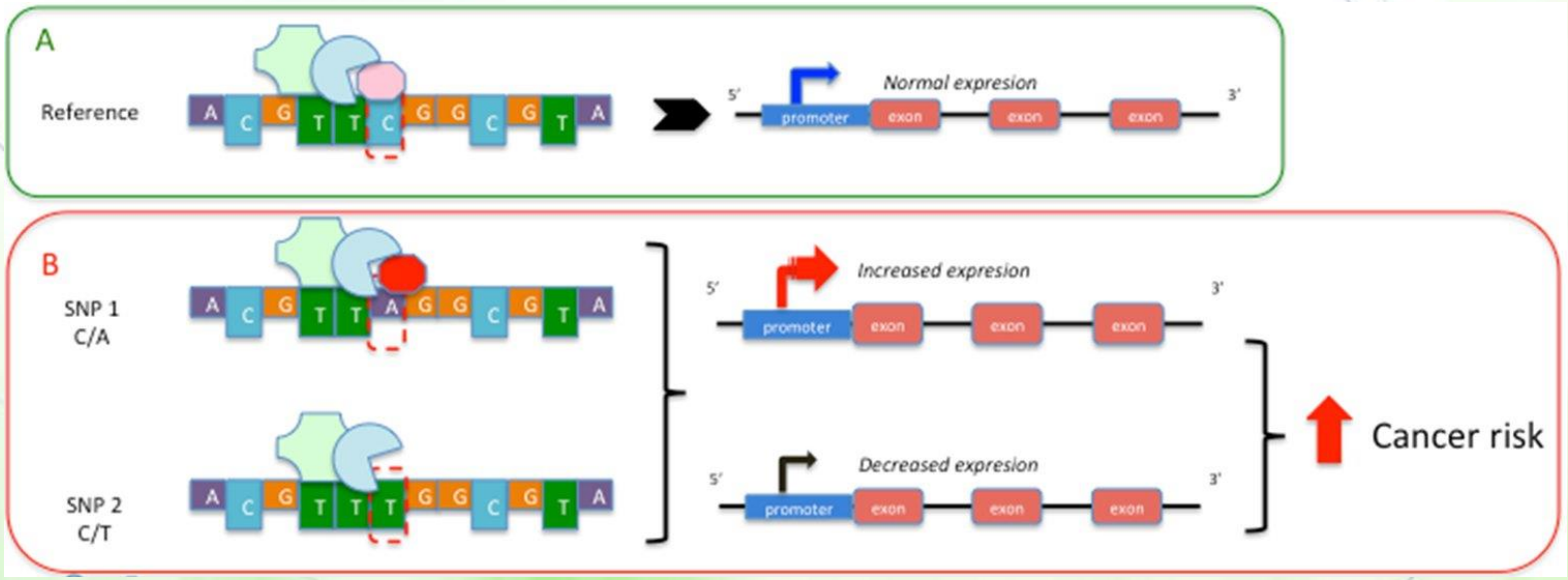
# SNPs and alternative polyadenylation

The presence of SNPs within the polyadenyation signal can also alter the length of the mRNA and , hence, protein amount in cells.



# SNPs in promoter

- Single nucleotide polymorphisms (SNP) in promoter region can alter the binding of transcription factors required for the expression of a gene.
- These variations may increase or decrease the expression of the affected gene, which eventually can influence the risk of developing a disease.



# mRNA transport

- Transport of mRNA from the nucleus to the cytoplasm, where it is translated into protein, is highly selective- and is associated to correct RNA processing

- Defective mRNA molecules like interrupted RNA, mRNA with inaccurate splicing, and so on, are not transported outside the nucleus

# Degradation of mRNAs

- The vast majority of mRNAs in a bacterial cell are very unstable, having a half-life of about 3 minutes
- The mRNAs in eukaryotic cells are more stable (up to 10 hours; average of 30 minutes)
- Exonucleases are responsible for degradation

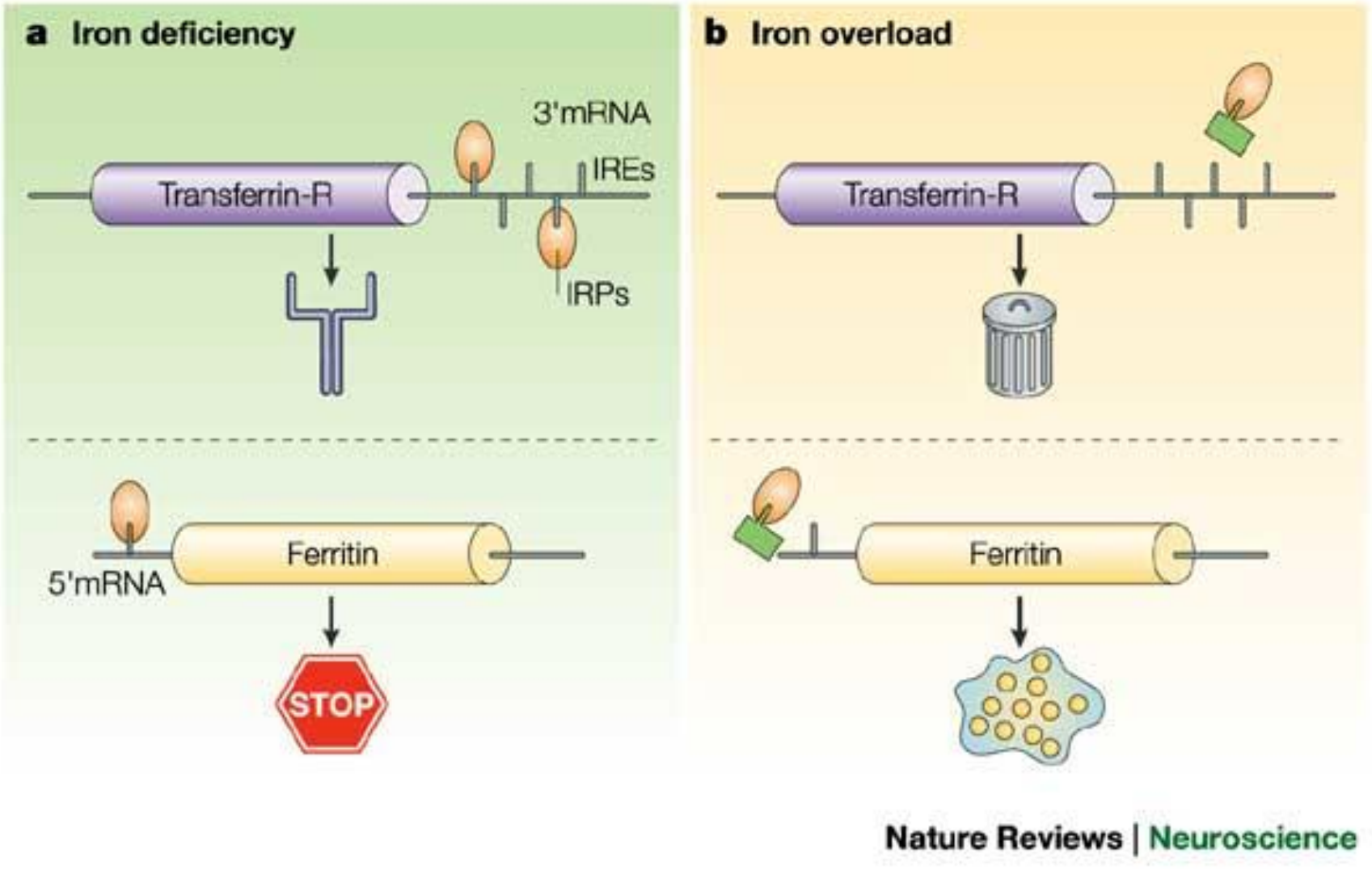
# Regulation of mRNA stability

# Iron-responsive elements

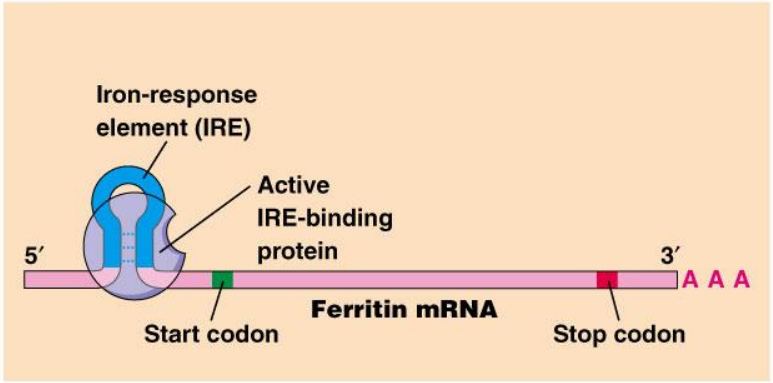
- In human cells, there are regions of mRNA called iron responsive elements (IREs)
- These regions are contained within the mRNA sequences that code for certain proteins that regulate the levels of iron
  - Ferritin, transferrin receptor, ferroportin, and DMT1
- Iron responsive element binding protein (IRE-BP) binds to these mRNA sequences influencing protein expression

# Effect on expression

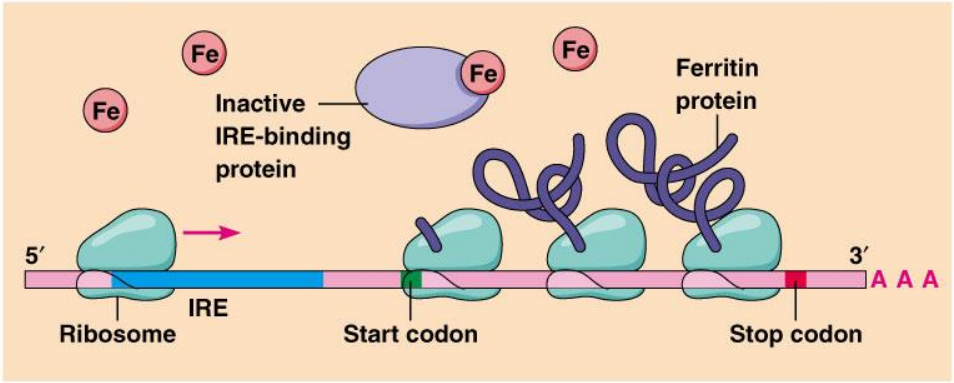
- When iron is abundant, it binds to IRE-BP, disabling the binding of IR-BP to ferritin mRNA
  - This prevents the degradation of the mRNA molecules allowing the production of more ferritin protein
  - Therefore, the iron itself causes the cell to produce more iron storage molecules
- On the other hand, at low iron levels, the IRE-BP will bind to the ferritin mRNA and, thus, the mRNA will be destabilized, making less ferritin protein
- An opposite effect is seen on the stability of transferrin mRNA


a Iron deficiency
3'mRNA
IREs
Transferrin-R
IRPs
5'mRNA
Ferritin
STOP
b Iron overload
Transferrin-R
Ferritin
Nature Reviews | Neuroscience

**(a) Low iron concentration.** IRE-binding protein binds to IRE, so translation of ferritin mRNA is inhibited.

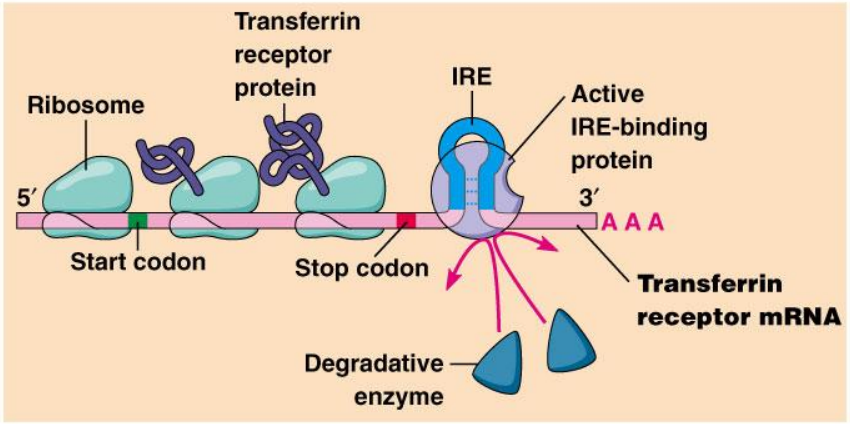



**(b) High iron concentration.** IRE-binding protein cannot bind to IRE, so translation of ferritin mRNA proceeds.

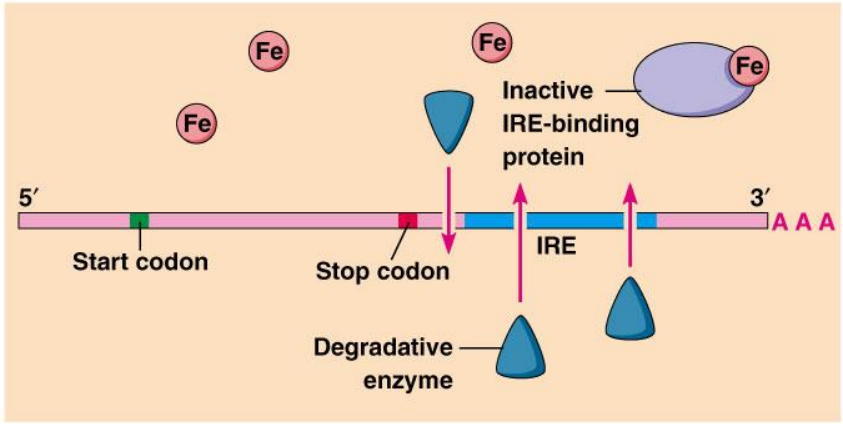





**(a) Low iron concentration.** IRE-binding protein binds to the IRE of transferrin receptor mRNA, thereby protecting the mRNA from degradation. Synthesis of transferrin receptor therefore proceeds.

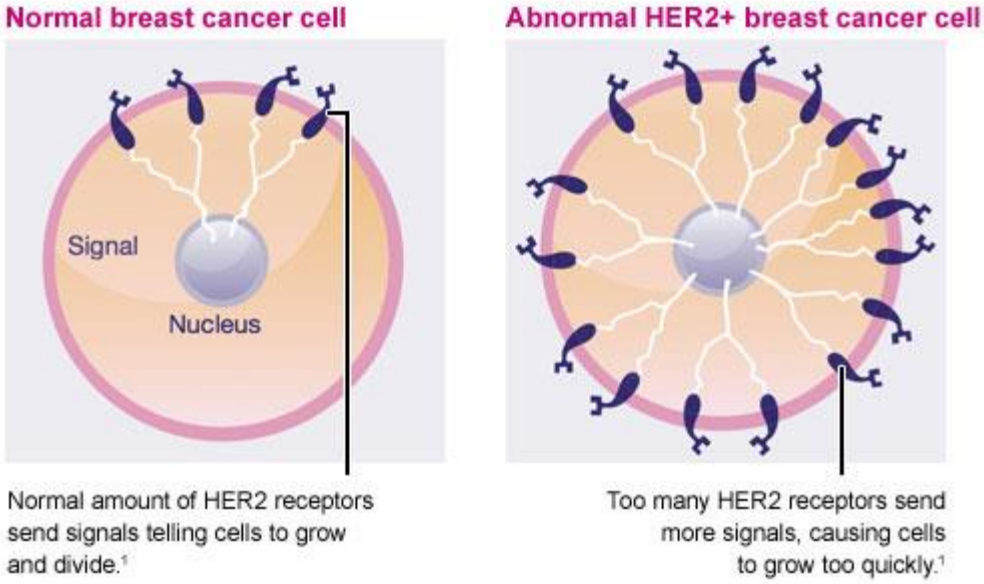



**(b) High iron concentration.** IRE-binding protein cannot bind to IRE, so mRNA is degraded and synthesis of transferrin receptor is thereby inhibited.

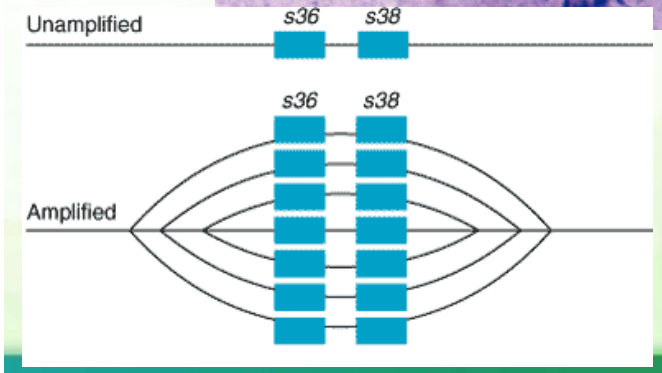


# Gene amplification

- It is an increase in copy number of a restricted region of a chromosome increasing the is the quantity of DNA in these regions.
- It is a mechanism that cancer cells use to escape resistance from methotrexate whereby the target gene, dihydrofolate reductase, is amplified.
- It is also a mechanism by which breast tumor cells progress and become more aggressive whereby they amplify the human epidermal growth factor receptor 2 (HER2), which stimulates cell growth.

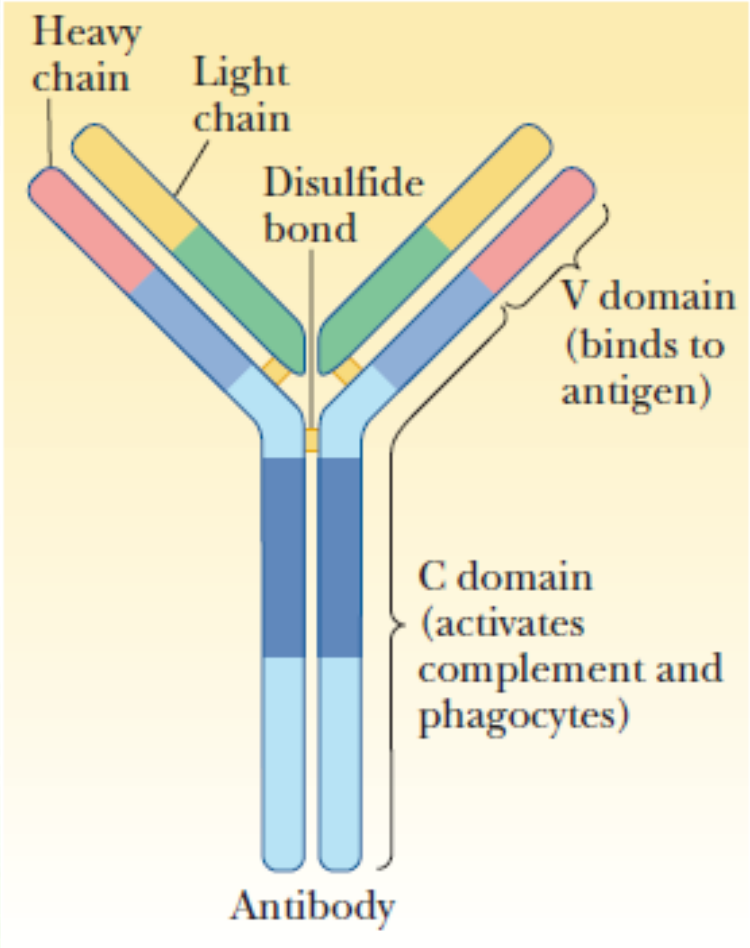






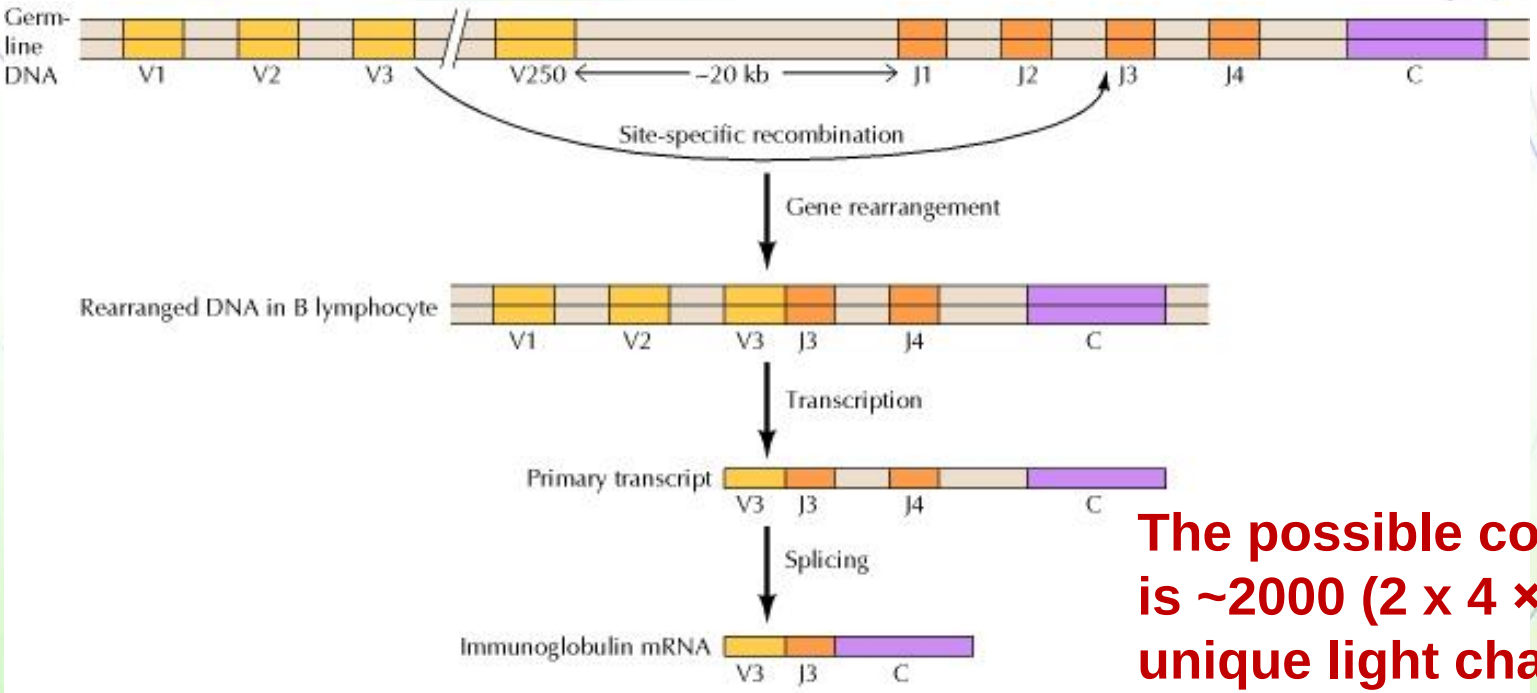
# Structure of antibodies

- Antibodies consist of two identical heavy chains and two identical light chains held together by disulfide bonds.
- Both contain constant and variable regions.
- The variable regions are responsible for recognition of antigens.
- immune system has the ability to produce about $10^{10}$-$10^{11}$ different antibodies.
- **How is diversity generated?**



# Gene rearrangement of the light chain

- There are two types of immunoglobulin light chains: κ and λ
- Each is a product of at least 3 genes:
  - Variable (VL) gene; 250 genes
  - Joining region (J) gene: 4 genes
  - Constant region (CL) gene: 1 gene

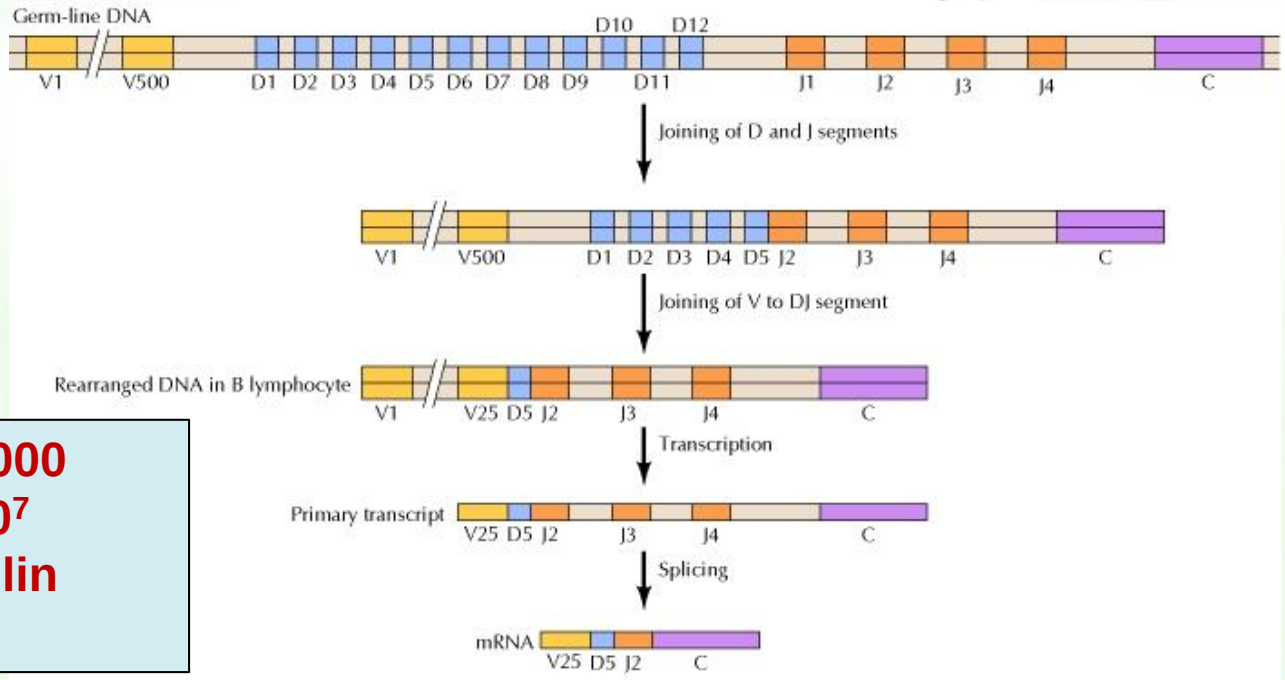



**The possible combinations is ~2000 (2 x 4 × 250) unique light chains**

# Gene rearrangement of the heavy chain

- Heavy chain is a product of at least 4 genes :
  - Variable region (VH) gene: 500 genes
  - Diversity region (D) gene: 12 genes
  - Joining region (J) gene: 4 genes
  - Constant region (CH) gene: 1 gene

**The possible combinations is ~24000 (500 x 12 x 4) unique heavy chains**




**2000 light chains x 24,000 heavy chains = ~ $5 \times 10^7$ different immunoglobulin molecules.**

# Additional mechanisms

- The joining of immunoglobulin gene segments is often imprecise, resulting in the formation of ~$10^5$ different light chains and ~ $2 \times 10^6$ heavy chains, which can then combine to form more than $10^{11}$ distinct antibodies.
- During recombination, nucleotides are added or deleted.
- Further antibody diversity is generated by a process known as somatic hypermutation, which results in the introduction of frequent mutations into the variable regions of both heavy-chain and light-chain genes.

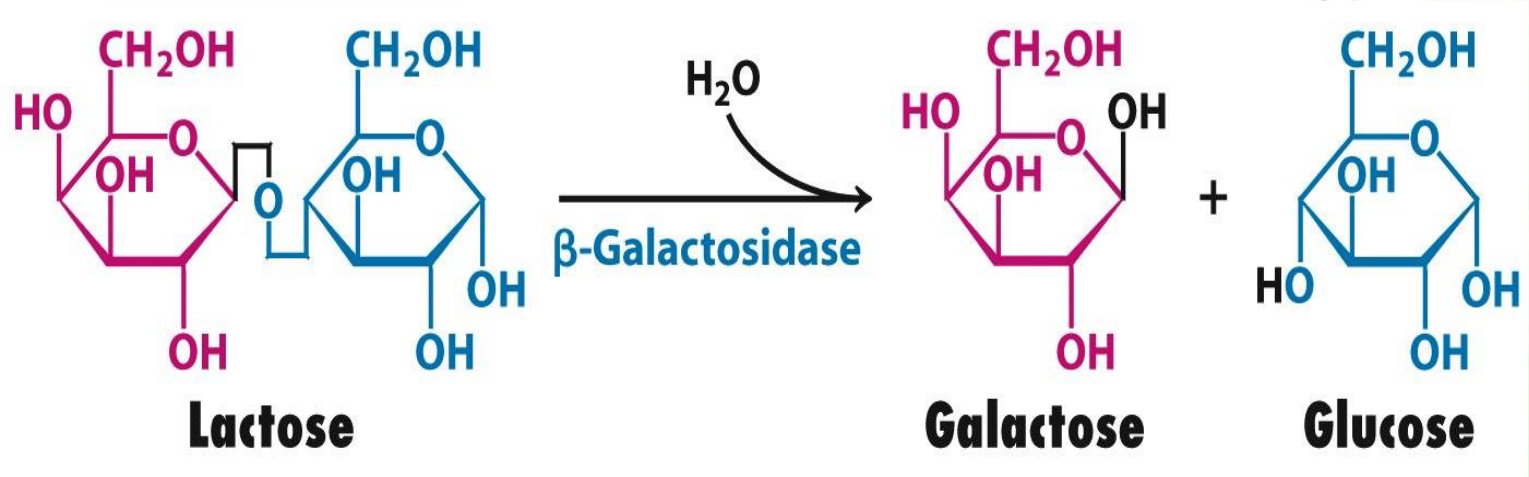

# Transcription-regulation

# Regulation of transcription in prokaryotes

The lac operon

# Metabolism of lactose

- In the 1950s, pioneering experiments were carried out by François Jacob and Jacques Monod who studied regulation of gene transcription in E. coli by analyzing the expression of enzymes involved in the metabolism of lactose

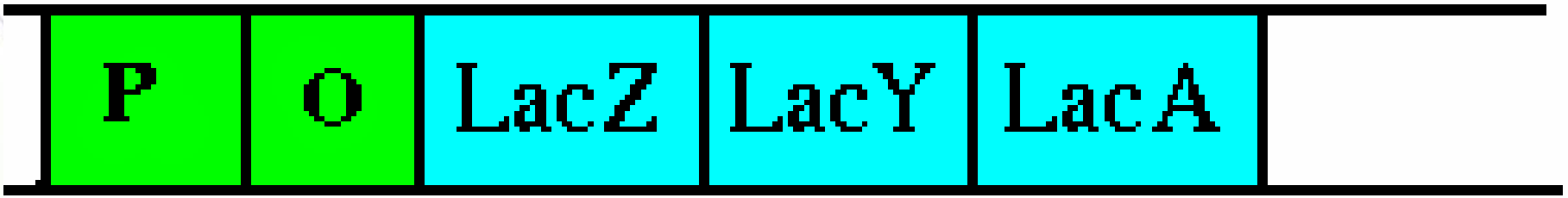


# Components of the lac operon

- Lactose induces the synthesis of enzymes involved in its own metabolism including:
  - $\beta$-galactosidase: catalyzes the cleavage of lactose
  - lactose permease: transports lactose into the cell
  - a transacetylase: acetylates $\beta$-galactosides
- These genes are located in one operon known as the lac operon

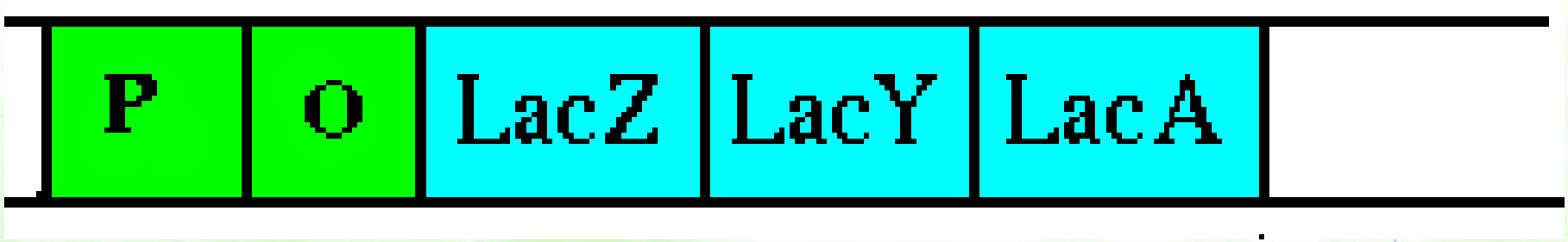
# What is an operon?

- A cluster of genes transcribed from one promoter producing a polycistronic mRNA



# The operator

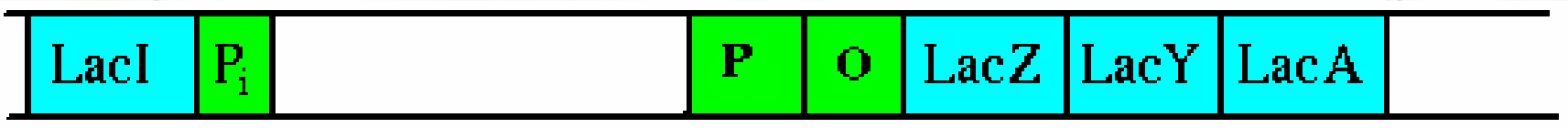
- The DNA region that regulates gene expression (transcription) is called a promoter. Usually this region is localized right before the start site of transcription
- It includes the RNA polymerase binding site
- The promoter also includes a region known as the operator region that also regulates transcription



# The i protein (lac repressor)

- Transcription of the lac operon is also controlled by a protein expressed by the *i* gene

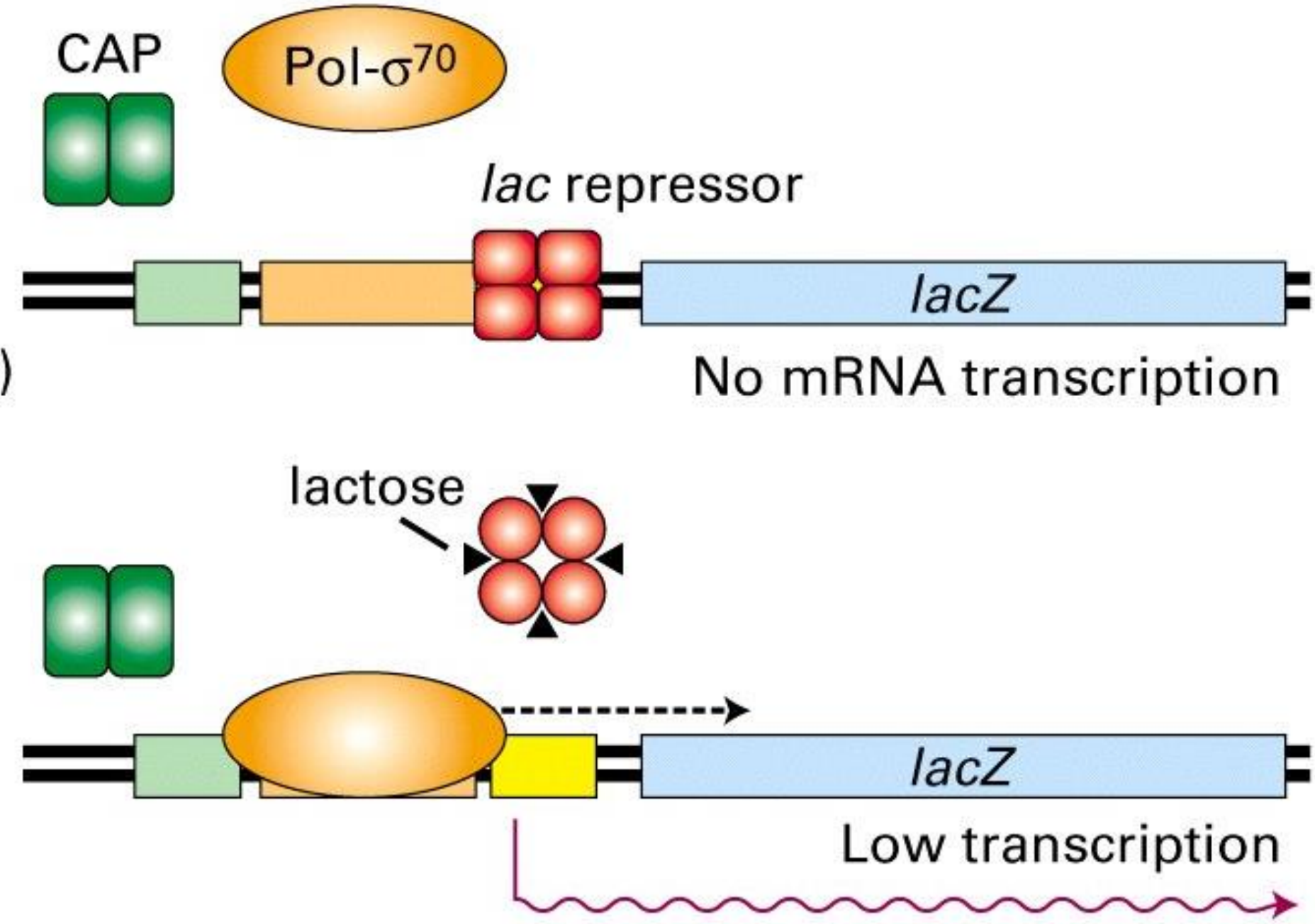
- The *i* protein (lac repressor) blocks transcription by binding to the operator preventing the RNA polymerase from biding to the promoter



# Regulation by lactose (positive)

- The addition of lactose leads to induction of the operon because lactose binds to the repressor, thereby preventing it from binding to the operator DNA

- This is known as positive regulation


CAP
Pol-σ70
lac repressor
lacZ
No mRNA transcription
lactose
lacZ
Low transcription

# Cis vs. trans regulatory elements

- Regulatory sequences like the operator are called cis-acting control elements, because they affect the expression of only linked genes on the same DNA molecule

- Proteins like the repressor are called transacting factors because they can affect the expression of genes located on other chromosomes within the cell

# Effect of mutations

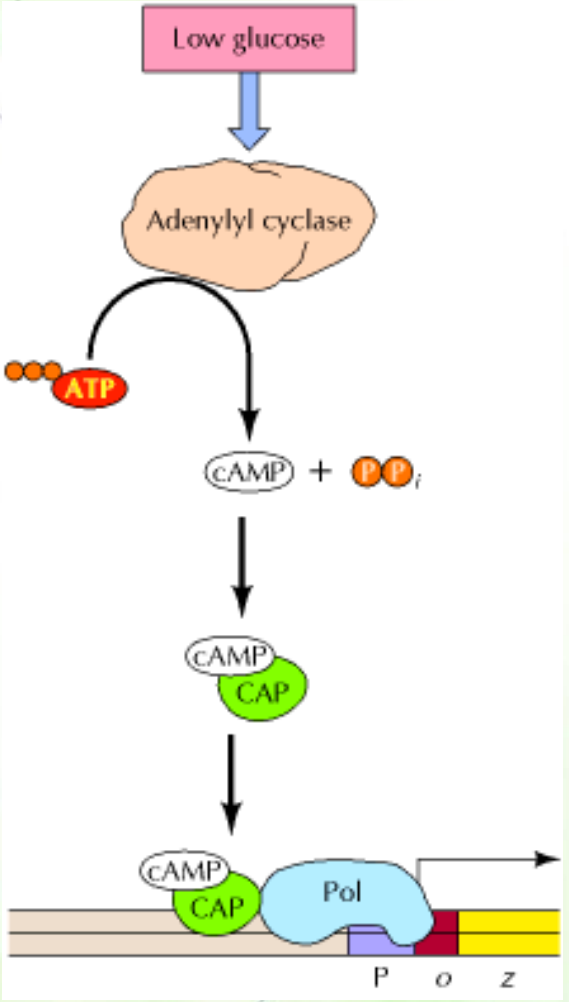
- Mutations affecting o result in constitutive expression (always on) since these mutations prevent i from binding to the operator
- Mutants of i are either constitutive or noninducible (always off)
- In constitutive i  mutants, i always binds lactose, so expression of the operon is always induced
- In noninducible i mutants, the repressor binds to the operator very tightly even in the presence of lactose

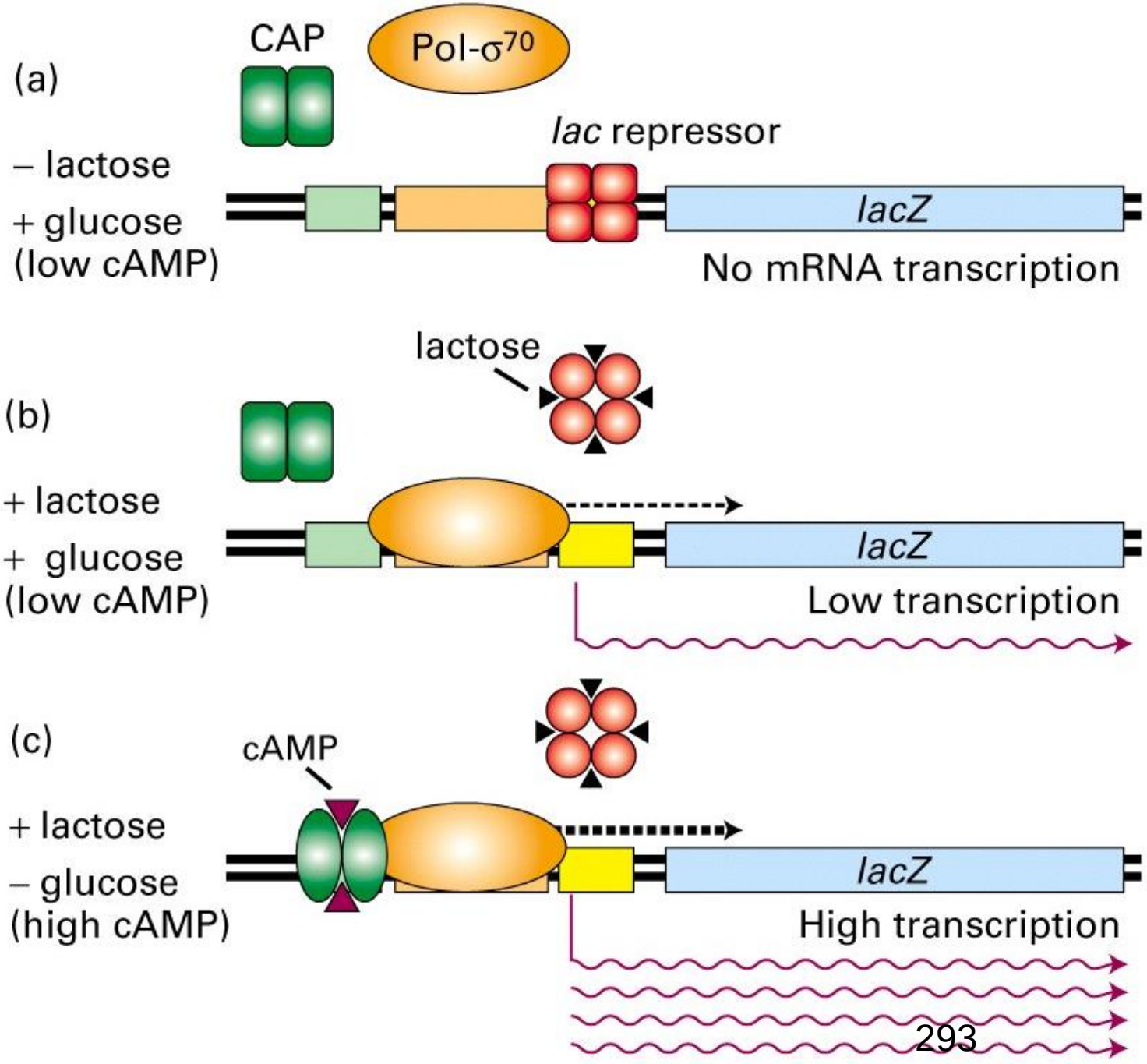
# Regulation by glucose (negative)

- Glucose is preferentially utilized by bacterial cells
- If E. coli are grown in medium containing both glucose and lactose, the lac operon is not induced and only glucose is used by the bacteria
- Glucose represses the lac operon even in the presence of the normal inducer (lactose)
- This is known as negative regulation

# How does glucose repress the expression of the lac operon?

- Low glucose activates the enzyme adenylyl cyclase, which converts ATP to cAMP
- cAMP then binds to catabolite activator protein (CAP)
- cAMP stimulates the binding of CAP to DNA upstream of the promoter
- CAP then interacts with the RNA polymerase, facilitating the binding of polymerase to the promoter and activating transcription

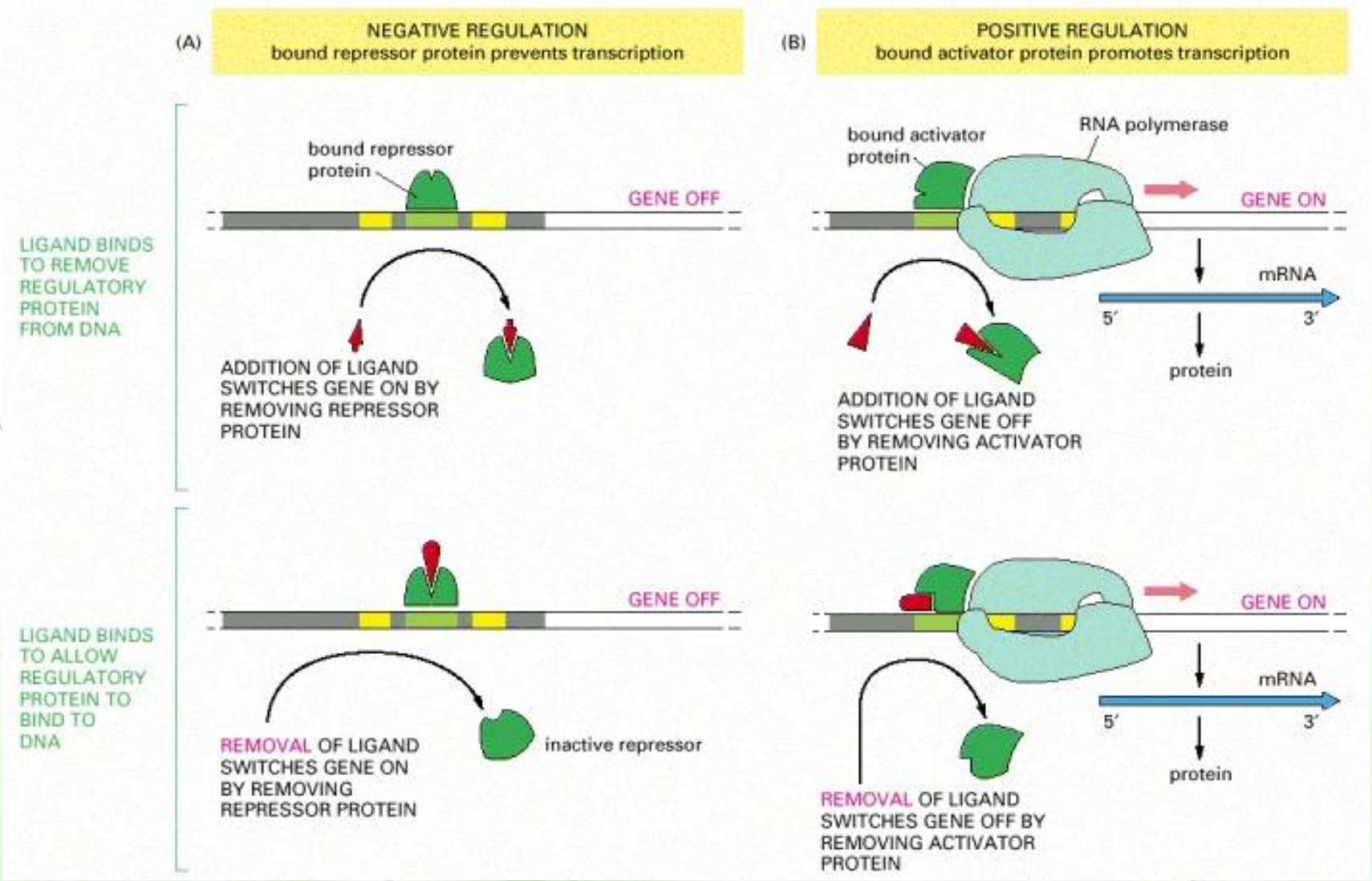



CAP
Pol-σ70
(a)
lac repressor
– lactose
+ glucose
(low cAMP)
lacZ
No mRNA transcription
lactose
(b)
+ lactose
+ glucose
(low cAMP)
lacZ
Low transcription
(c)
cAMP
+ lactose
– glucose
(high cAMP)
lacZ
High transcription

# Resources

- http://www.sumanasinc.com/webcontent/animations/content/lacoperon.html

# Positive vs. negative regulation



# Regulation of transcription in eukaryotes

# Regulatory mechanisms

- Although the control of gene expression is far more complex in eukaryotes than in bacteria, the same basic principles apply
- Transcription in eukaryotic cells is controlled by:
  - Cis-acting DNA sequences
  - Transcriptional regulatory proteins
  - Repressor proteins
  - Modification of DNA and its packaging into chromatin

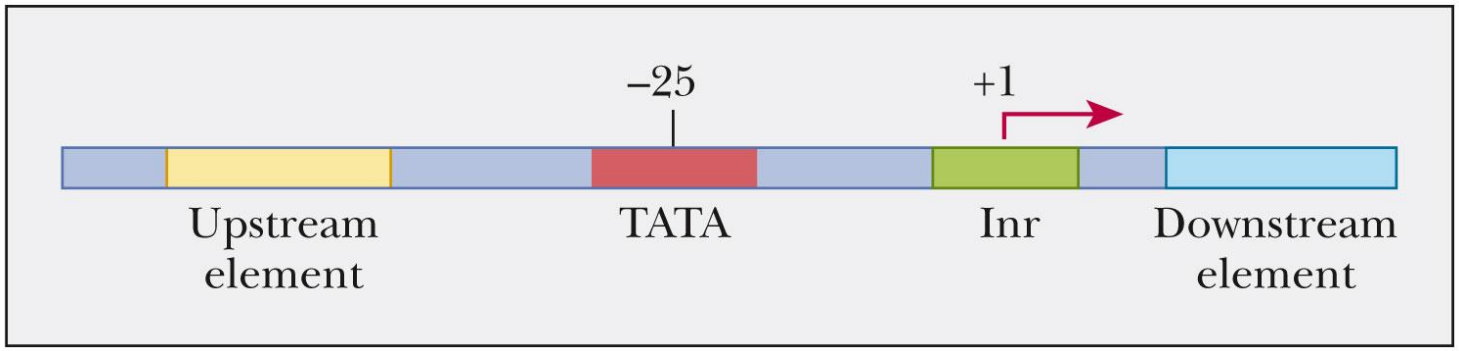
# Regulatory sequences

## Promoters and enhancers

- As already discussed, transcription in bacteria is regulated by the binding of proteins to cis-acting sequences (e.g., the lac operator)
- Similar cis-acting sequences regulate the expression of eukaryotic genes:
  - promoters
  - enhancers

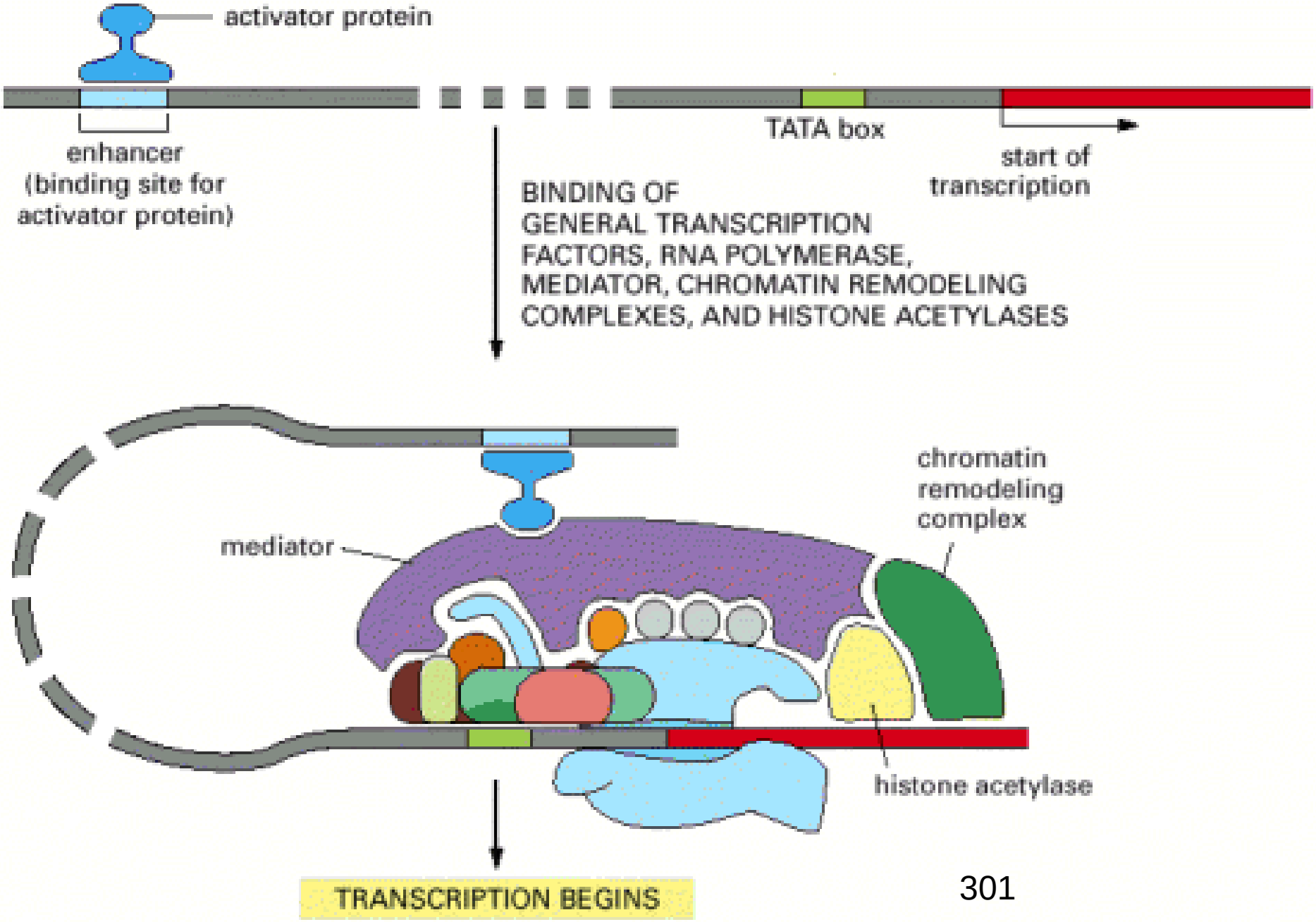
# General components of promoters

- An upstream elements, which act as enhancers and silencers.
- The TATA box, which has a consensus sequence of TATAA(T/A).
- The initiator element (*Inr)*, which surrounds the +1 site.
- A possible downstream regulator



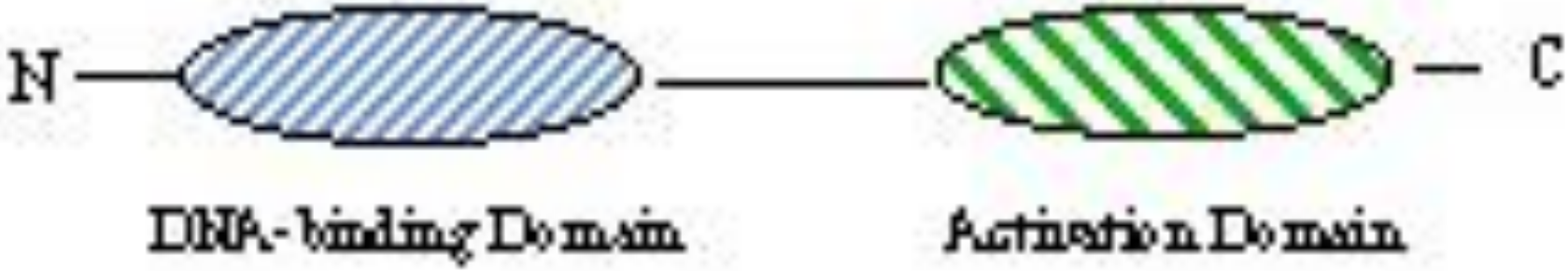
# Enhancers

- Genes that are regulated solely by the consensus elements in the promoter region are constitutively expressed.
- Gene-specific transcription factors (also called *transactivators or activators)* are located farther away from the transcription start site and *bind to gene-specific regulatory sequences and interact with mediator proteins* such as coactivators.
- These cis-acting regulatory sequences, also called enhancers, function by binding transcription factors that then regulate RNA polymerase.
- They can stimulate transcription when placed either upstream or downstream of the promoter, in either a forward or backward orientation because of DNA looping


activator protein
enhancer
(binding site for
activator protein)
TATA box
start of
transcription
BINDING OF
GENERAL TRANSCRIPTION
FACTORS, RNA POLYMERASE,
MEDIATOR, CHROMATIN REMODELING
COMPLEXES, AND HISTONE ACETYLASES
chromatin
remodeling
complex
mediator
histone acetylase
TRANSCRIPTION BEGINS

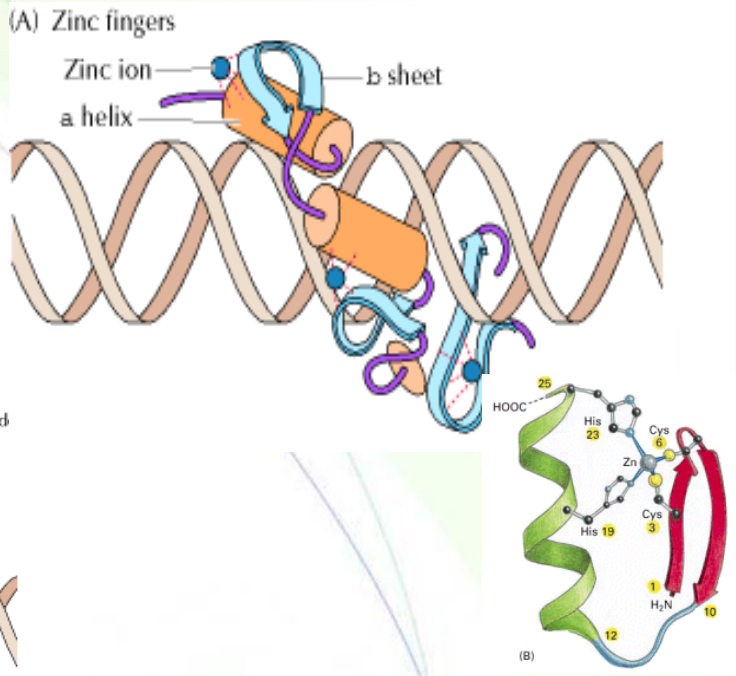
# Transcriptional regulatory proteins

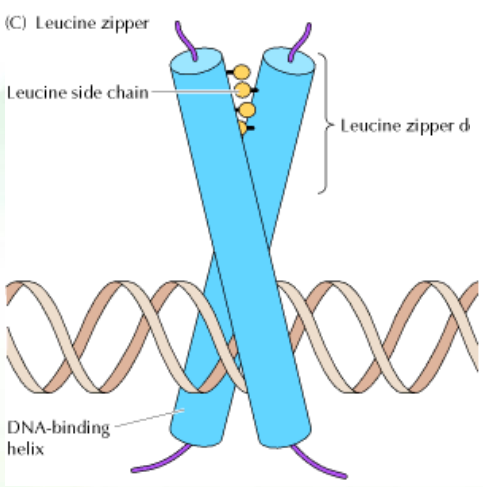
- These proteins to consist of two domains:
  - One region of the protein specifically binds DNA (DNA-binding domain)
  - the other activates transcription by interacting with other components of the transcriptional machinery (regulatory or activation domain)
- Both activities are independent and can be separated from each other

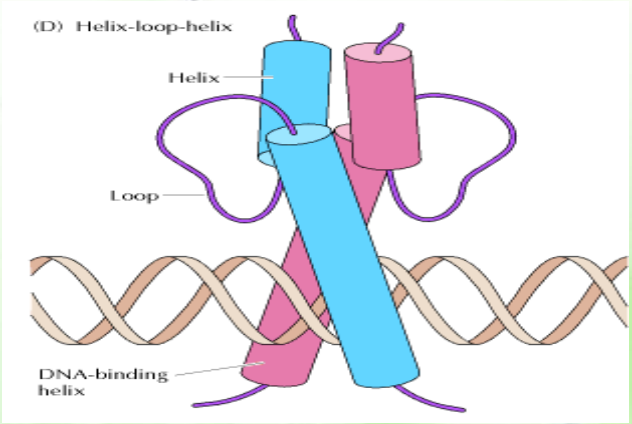


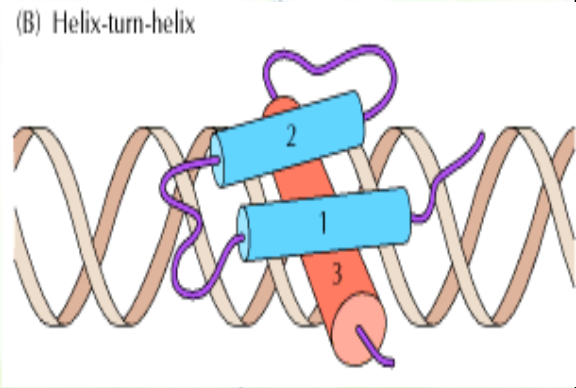
# DNA-binding domains

- Zinc finger domains (Steroid receptors)
- helix-turn-helix motif
- leucine zipper (CREB)
- helix-loop-helix



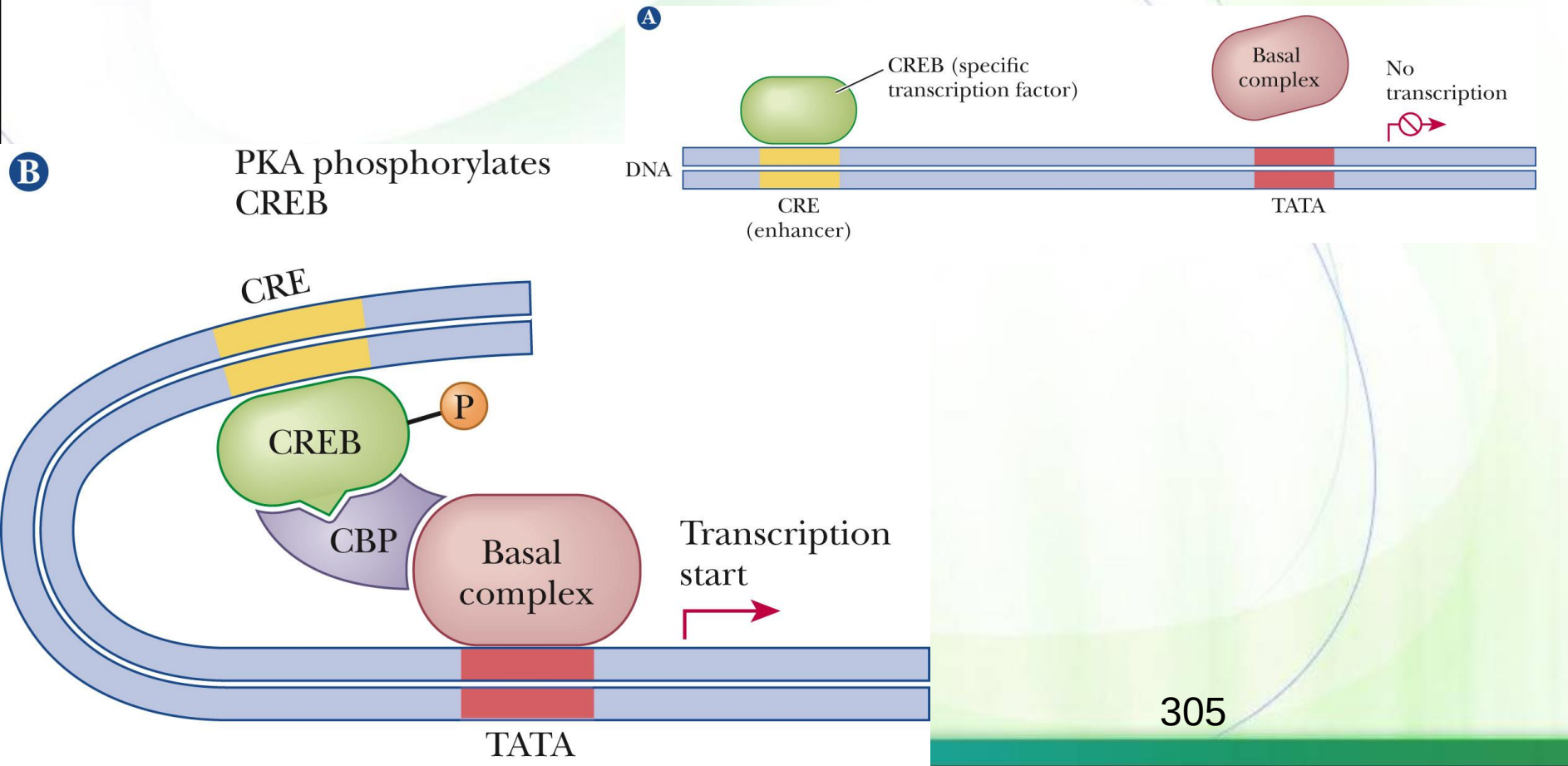



# The activation domains

- These activation domains are thought to stimulate transcription by interacting with general transcription factors, such as TFIIB or TFIID, thereby facilitating the assembly of a transcription complex on the promoter
  - Acidic domains
  - Glutamine-rich domains
  - Proline-rich domains

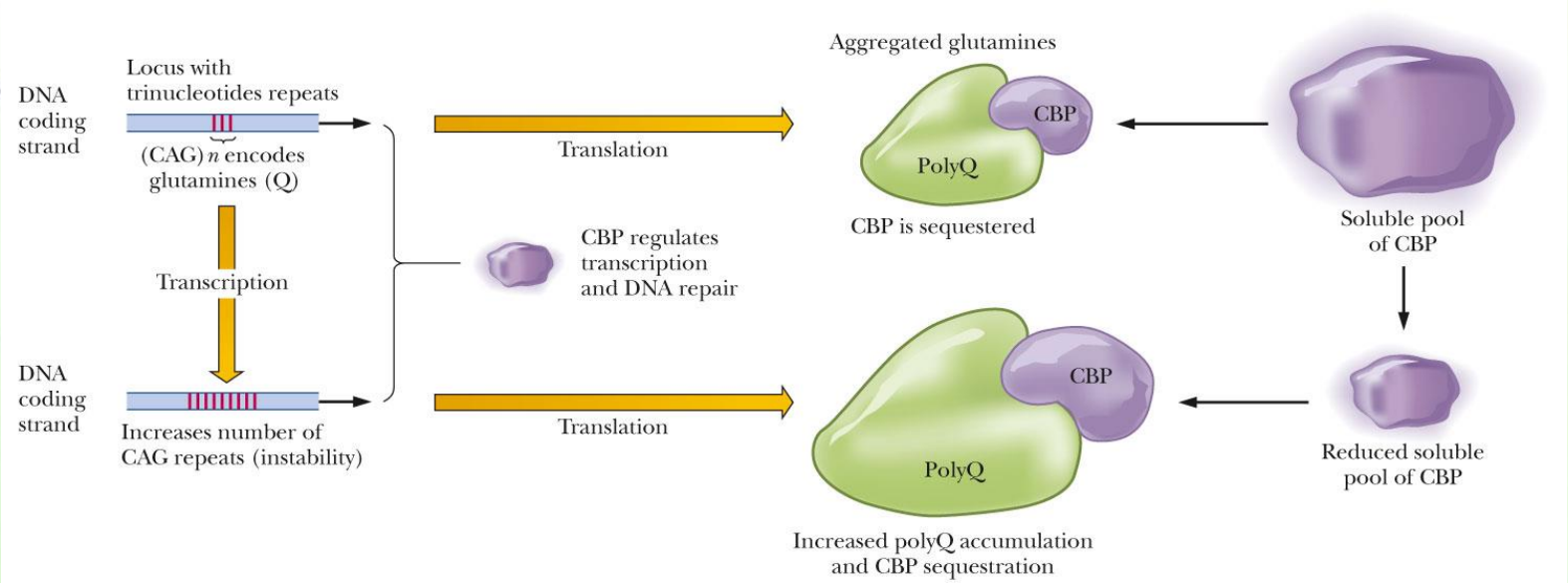
# cAMP-response element (CRE)

- Unphosphorylated CREB does not bind to CREB binding protein, and no
- transcription occurs.
- In the presence of cAMP, protein kinase A is activated phosphorylating CREB, which binds to CBP. The dimer can then form a complex with the RNA polymerase and, thereby, activating transcription.



# Huntington's disease

- It is a disease caused by a mutation in a transcription protein known as huntingtin. The mutation is an increase in the number of a trinucleotide repeat of CAG (it encodes polyglutamine)
- The polyglutamine product sequesters CBP, making less of it available for molecular processes, such as transcription and DNA repair.
- The loss of the DNA repair leads to a propagation of the CAG repeats and leads to the disease becoming worse in successive generations.



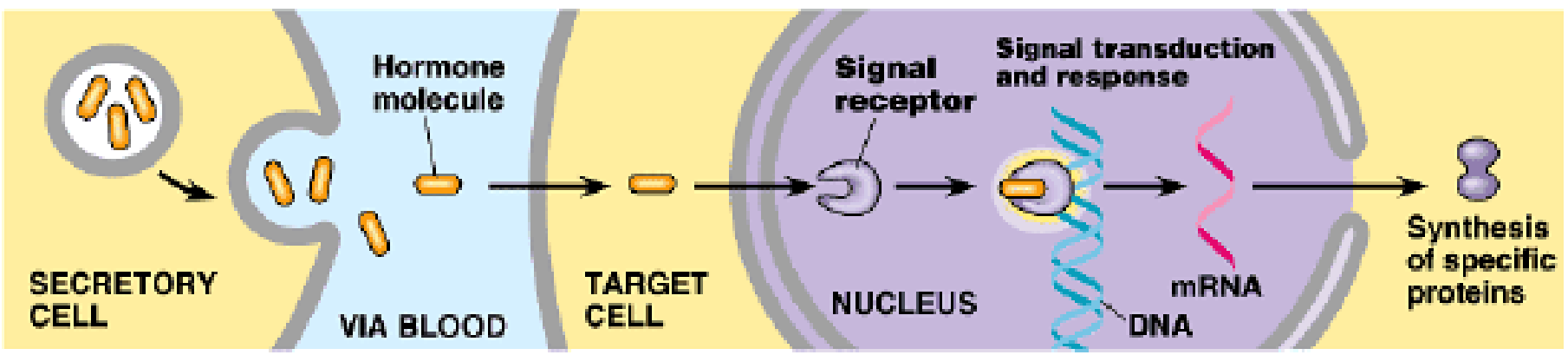
# Eukaryotic Repressors

- Repressors bind to specific DNA sequences and inhibit transcription

- Repressors may have
  - both DNA-binding and protein-binding domains
  - DNA-binding domains, but not protein-interaction domains
  - protein-interacting domains, but not DNA-binding domains

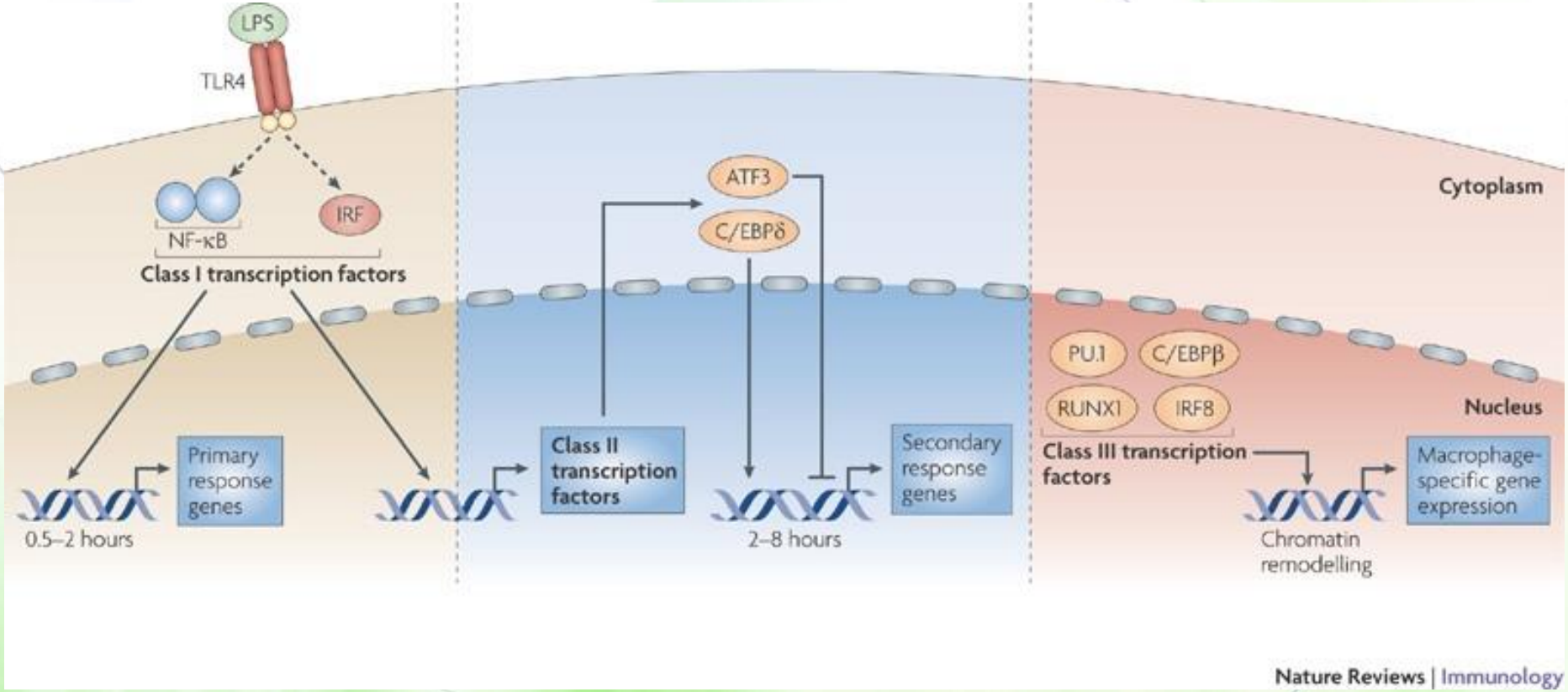
# Steroid receptors

- The receptors to which lipophilic steroid hormones bind are ligand-activated proteins that regulate transcription of selected genes
- They are found in the cytosol and the nucleus
- Upon hormonal binding, the hormone-receptor complex bind to specific DNA promoter/enhancer sequences



# Trancriptional regulatory network

- primary, secondary,…etc transcription regulation

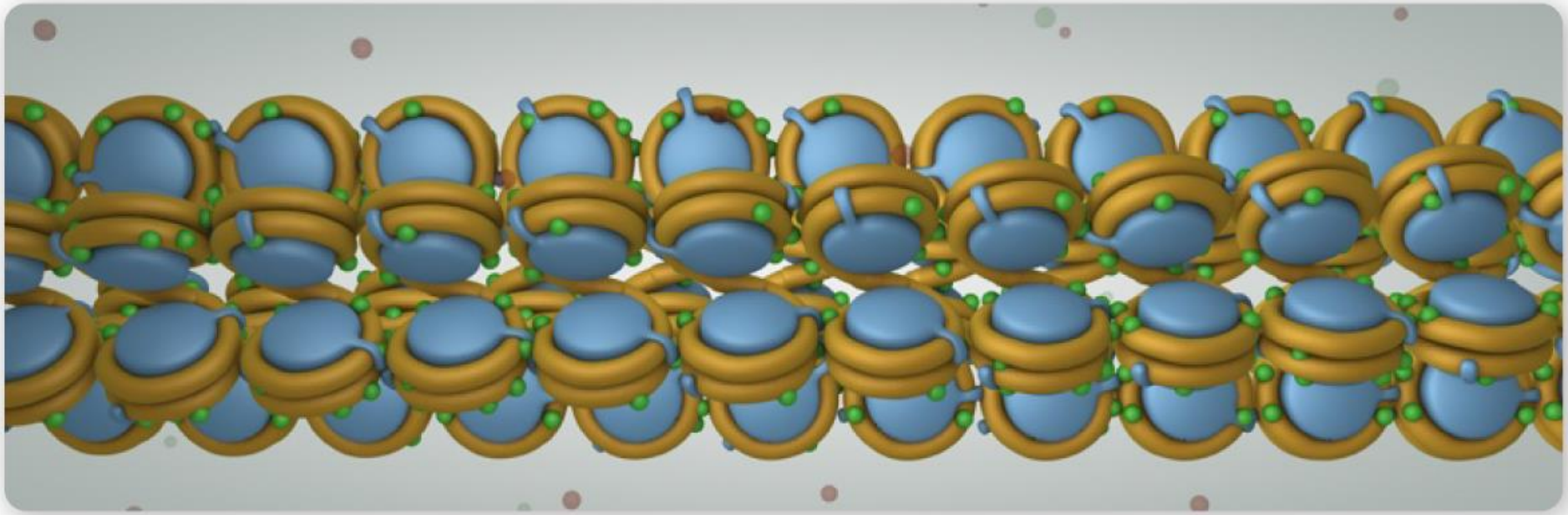


# Epigenomics

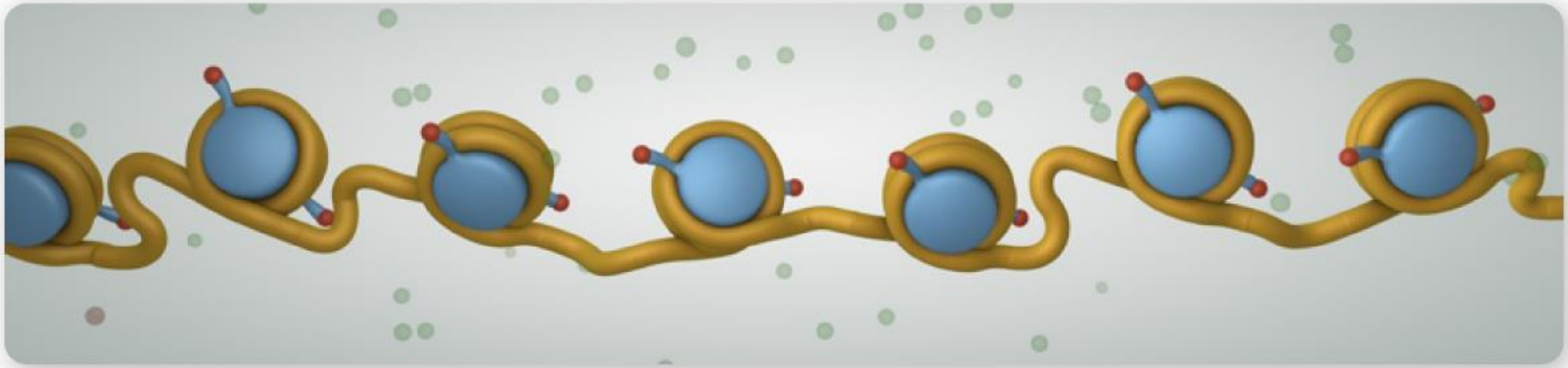
- Epi: “above” or “in addition to”
- It indicates genetic alterations in gene expression without a change in DNA sequence
  - Can be caused by the pattern of chromosomal packaging and modification (e.g. methylation)

# Modulation of chromosomal structure

- The packaging of eukaryotic DNA in chromatin has important consequences in terms of its availability as a template for transcription
  - Actively transcribed genes are found in loose chromatin, and vice versa
  - Even in loose chromatin, the tight winding of DNA around the nucleosome core particle can prevent transcription factors from binding DNA and the RNA polymerase to transcribe through a chromatin template



The epigenome tightly wraps **inactive** genes, making them unreadable.



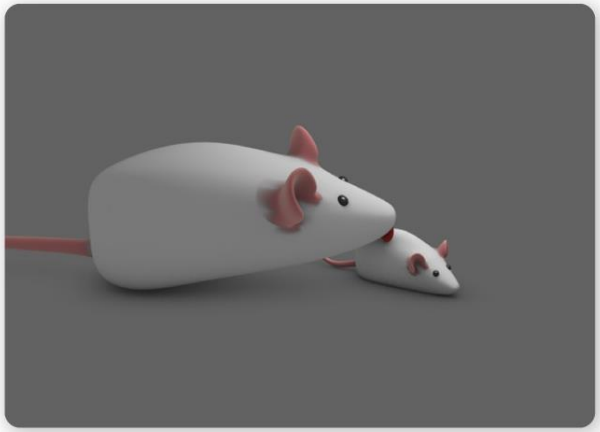
The epigenome relaxes **active** genes, making them easily accessible.

# Identical twins have the exact same genetic information

**But their epigenomes become increasingly different over time**

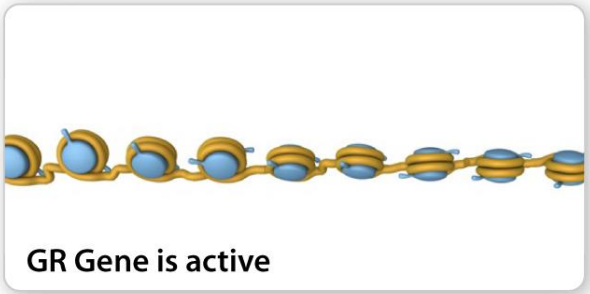
- Epigenetic changes can cause dramatic differences between twins, including many cases where one twin develops a disease and the other does not.

# Maternal care affects the epigenome

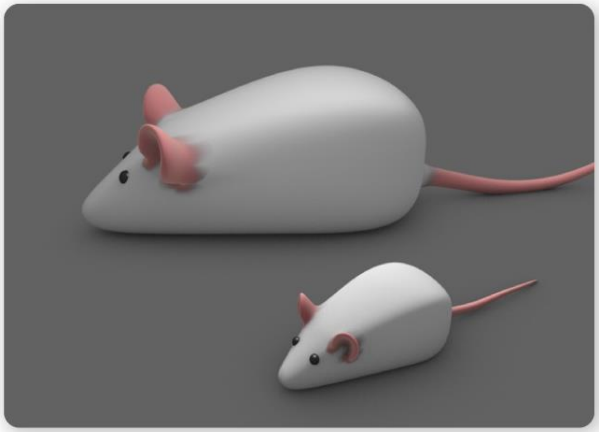


Highly nurtured rat pups tend to grow up to be calm adults.

**High Nurtured**

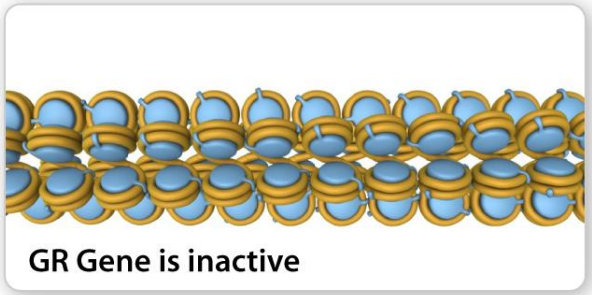



In high nurtured rat pups, the GR gene is active. These rats have an easy time relaxing after stress.



Rat pups who receive little nurturing tend to grow up to be anxious adults.

**Low Nurtured**



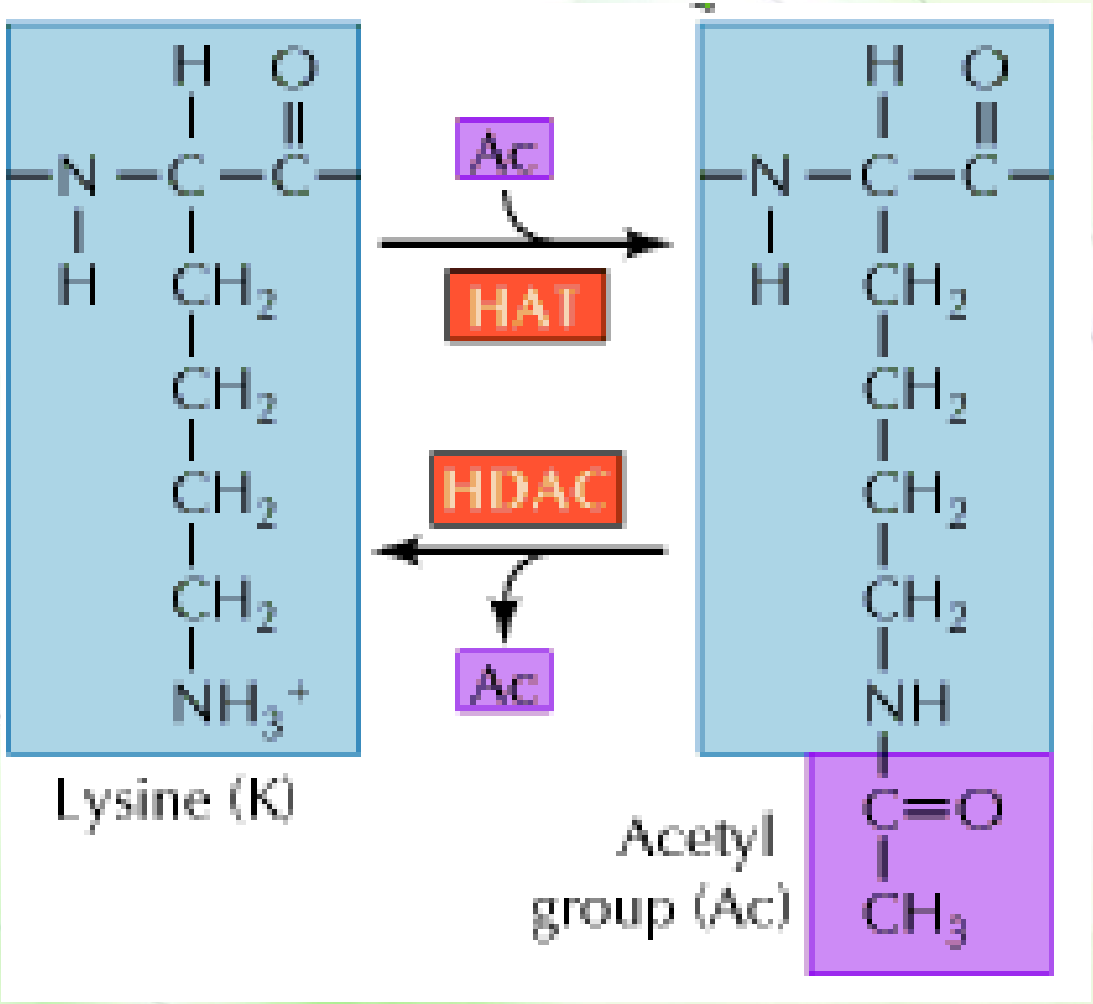

In low nurtured rat pups, the GR gene is epigenetically silenced. These rats have a hard time relaxing after stress.

*When it's active, the GR gene produces a protein that helps the body relax after stress. Mom's nurturing during the first week of life shapes her pups' epigenomes.*

# How are chromosomal structures altered?

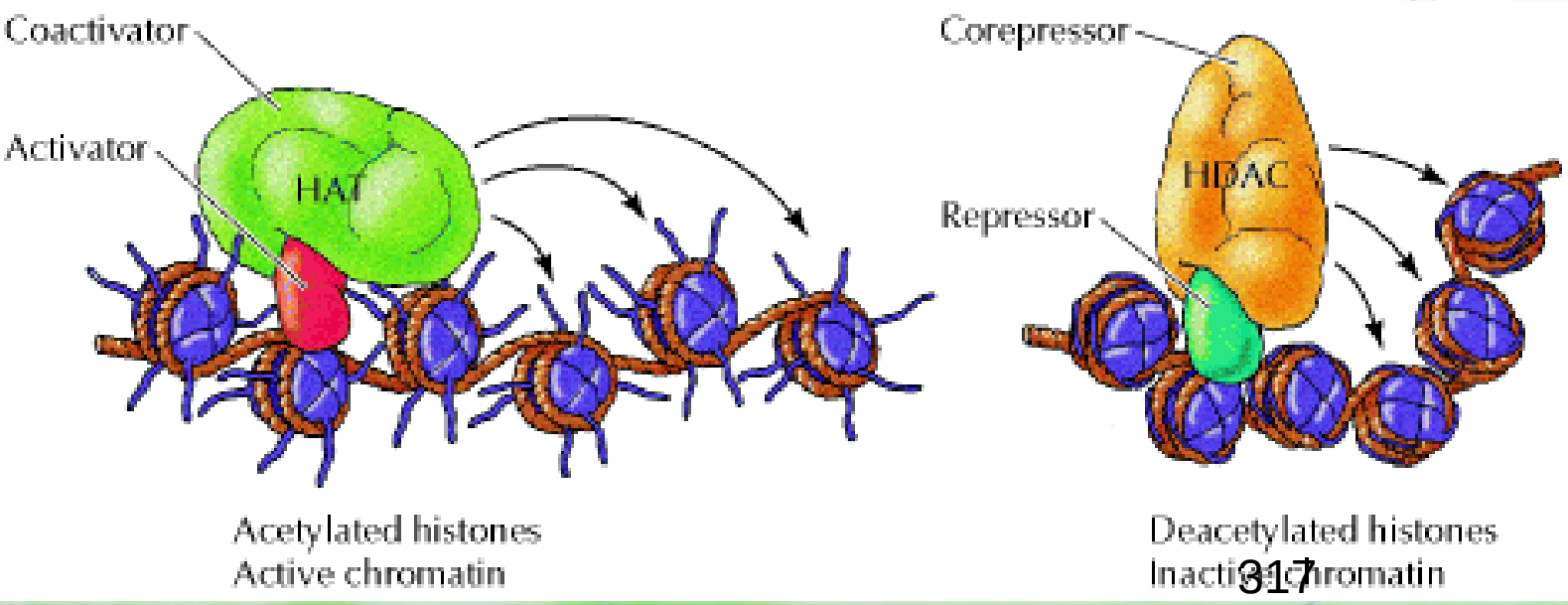
- Decrease in the compactness of the chromatin by:
  - Acetylation of histones
  - Binding of specialized nonhistone chromosomal proteins to nucleosomes of actively transcribed genes to change their conformation.
  - Complexing of transcription factors with nucleosome remodeling factors.

# Histone acetylation



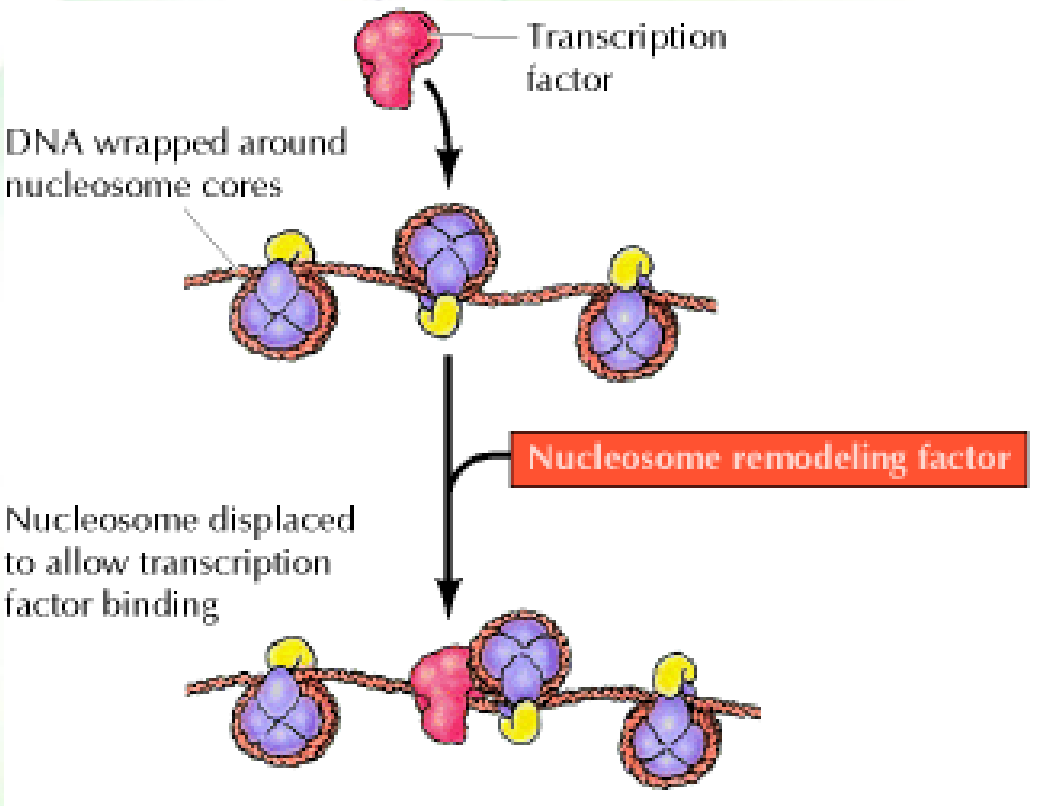
# Enzymatic association

- Transcriptional activators and repressors are associated with histone acetyltransferases and deacetylases, respectively
  - A component of TFIID has been found to be a histone acetyltransferases



# Nucleosome remodeling factors

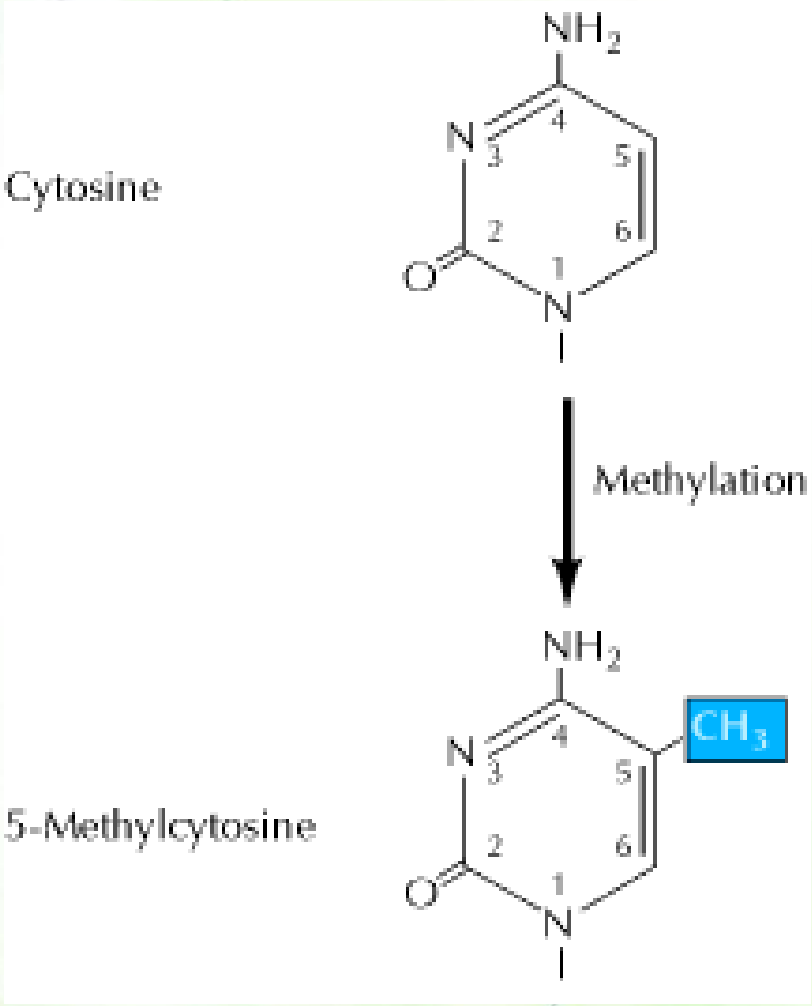
- These are protein complexes that facilitate the binding of transcription factors by altering nucleosome structure by the accessibility of nucleosomal DNA to transcription factors and other proteins without removing the histones



# DNA Methylation

- Cytosine residues in DNA can be modified by the addition of methyl groups at the 5-carbon position

- DNA methylation is correlated with reduced transcriptional activity of genes



# Mechanism of inhibition

- Methylation inhibits transcription of genes via the action of a protein, MeCP2, that specifically binds to methylated DNA and represses transcription
  - Interestingly, MeCP2 functions as a complex with histone deacetylase, linking DNA methylation to histone acetylation and nucleosome structure

# Significance of DNA methylation

- DNA methylation has been established in two important phenomena:
  - X chromosome inactivation
    - In females, one of the X chromosomes is inactivated in every cells.
  - genomic imprinting
    - Either the paternal gene or the maternal gene is active.

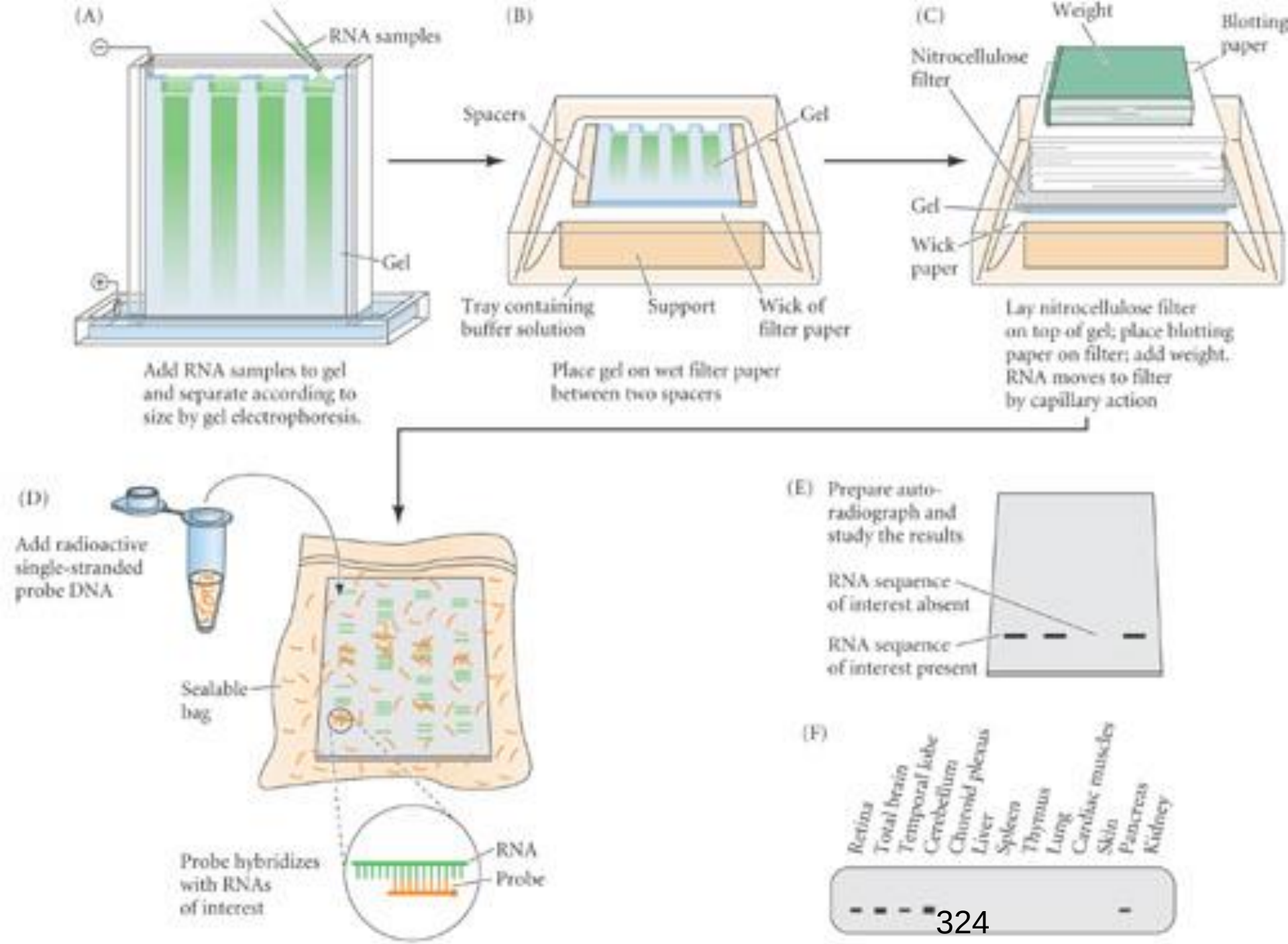
# How do different cells appear?

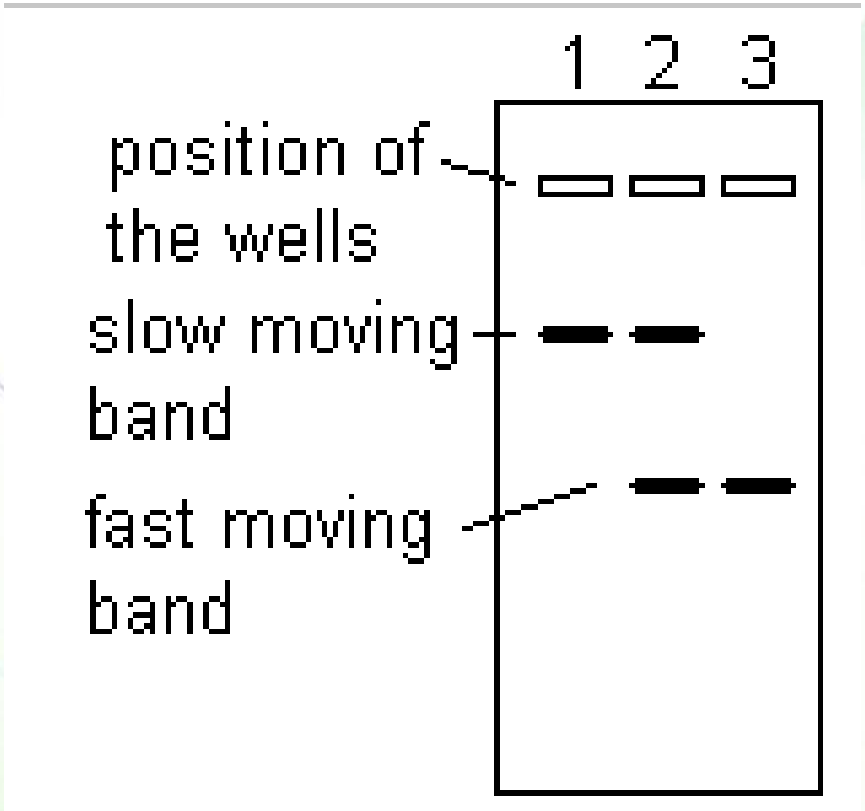
- All cells in one body have the same genome, but cell types in a multicellular organism become different from one another because they synthesize different sets of RNA and protein molecules.
- The patterns of mRNA can differentiate cell types from each other.

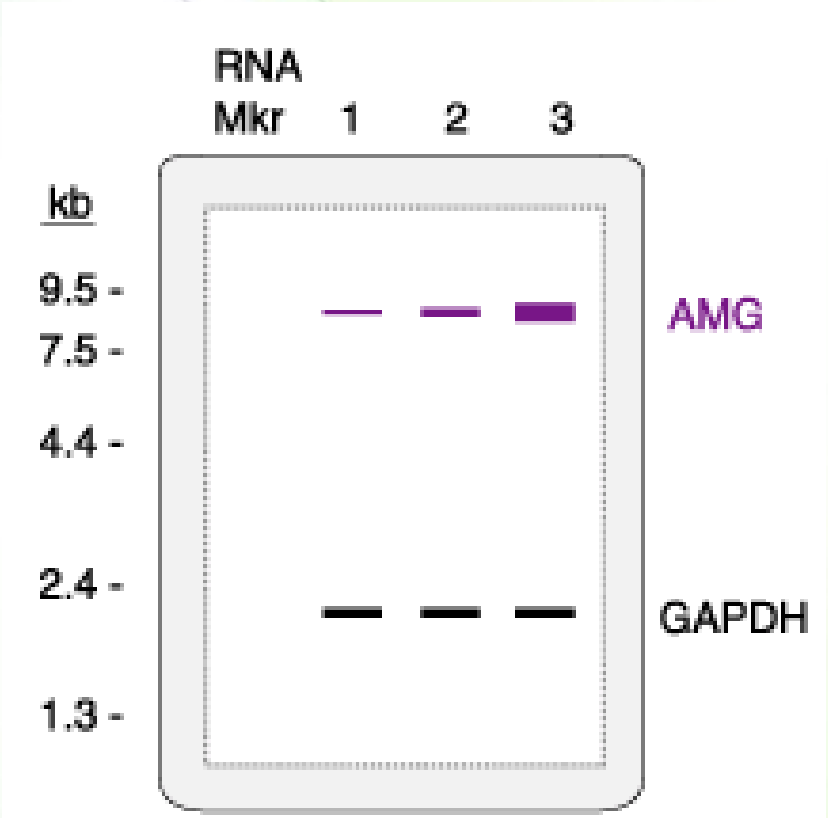
- The transcriptome (the total collection of RNA transcripts in a cell)

# Northern blotting

- This is done exactly like Southern blotting except that RNA from cells is isolated instead of DNA.
- Then RNA molecules are fractionated based on sizes by gel electrophoresis.
- The fractionated RNA molecules are transferred to a membrane.
- The RNA molecules on the membrane are targeted by a labeled DNA probe whose sequence is complementary to a specific RNA molecule.


(A)
RNA samples
Gel
Add RNA samples to gel and separate according to size by gel electrophoresis.
(B)
Spacers
Gel
Tray containing buffer solution
Support
Wick of filter paper
Place gel on wet filter paper between two spacers
(C)
Weight
Blotting paper
Nitrocellulose filter
Gel
Wick paper
Lay nitrocellulose filter on top of gel; place blotting paper on filter; add weight. RNA moves to filter by capillary action
(D)
Add radioactive single-stranded probe DNA
Sealable bag
Probe hybridizes with RNAs of interest
RNA
Probe
(E) Prepare auto-radiograph and study the results
RNA sequence of interest absent
RNA sequence of interest present
(F)
Retina
Total brain
Temporal lobe
Cerebellum
Choroid plexus
Liver
Spleen
Thymus
Lung
Cardiac muscles
Skin
Pancreas
Kidney

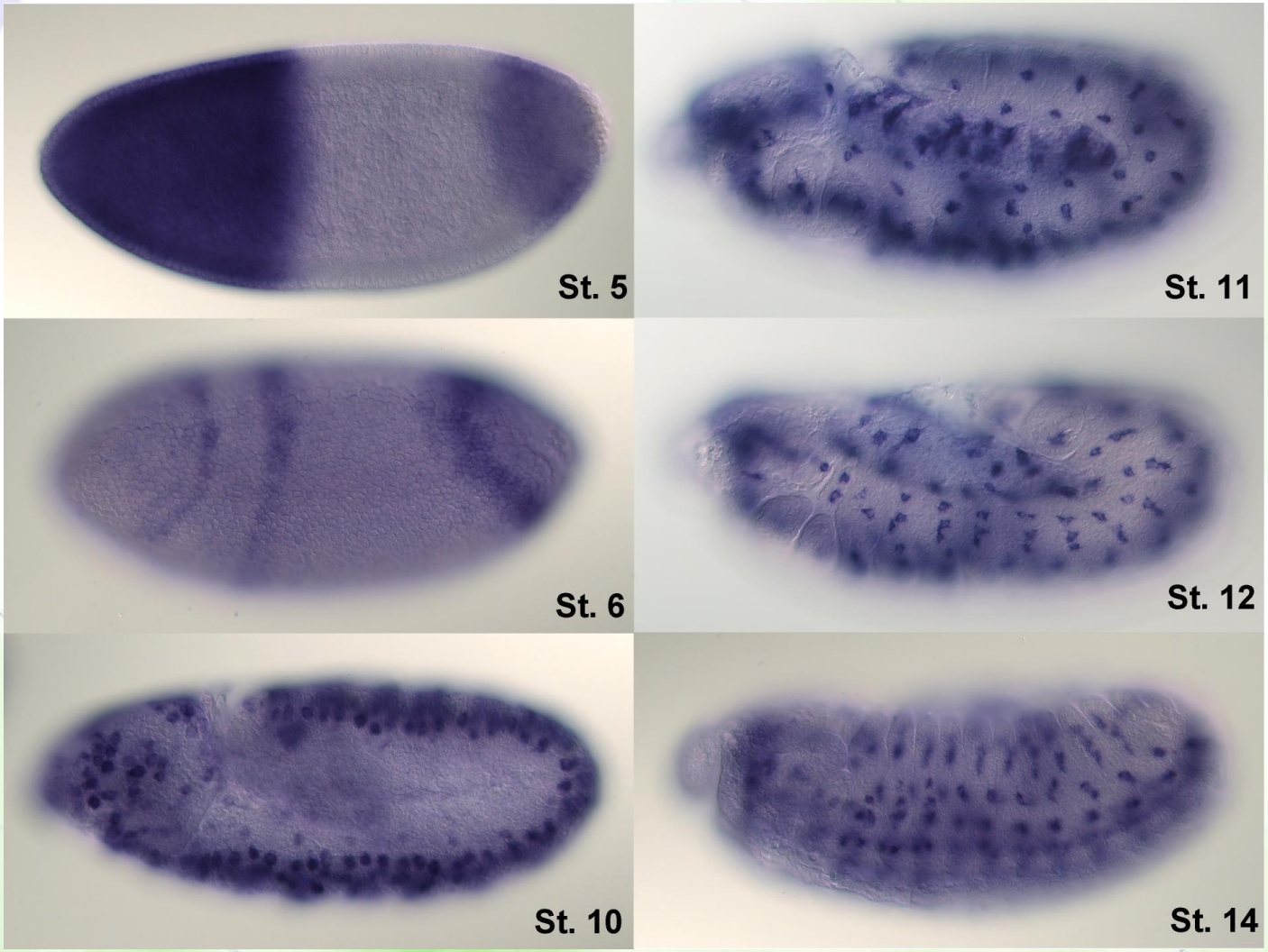

1 2 3
position of the wells
slow moving band
fast moving band

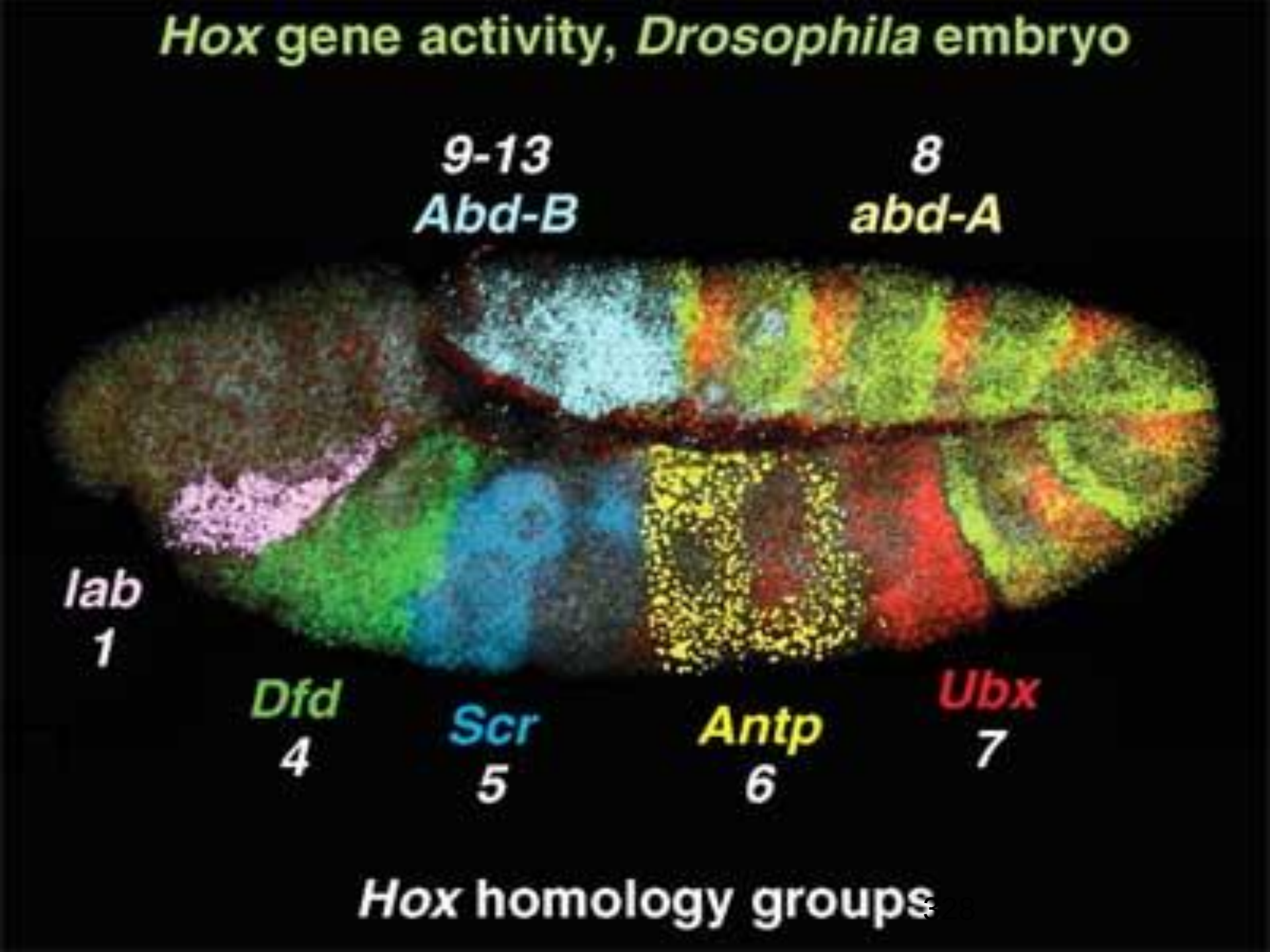

RNA
Mkr 1 2 3
kb
9.5 -
7.5 -
4.4 -
2.4 -
1.3 -
AMG
GAPDH

# In situ hybridization

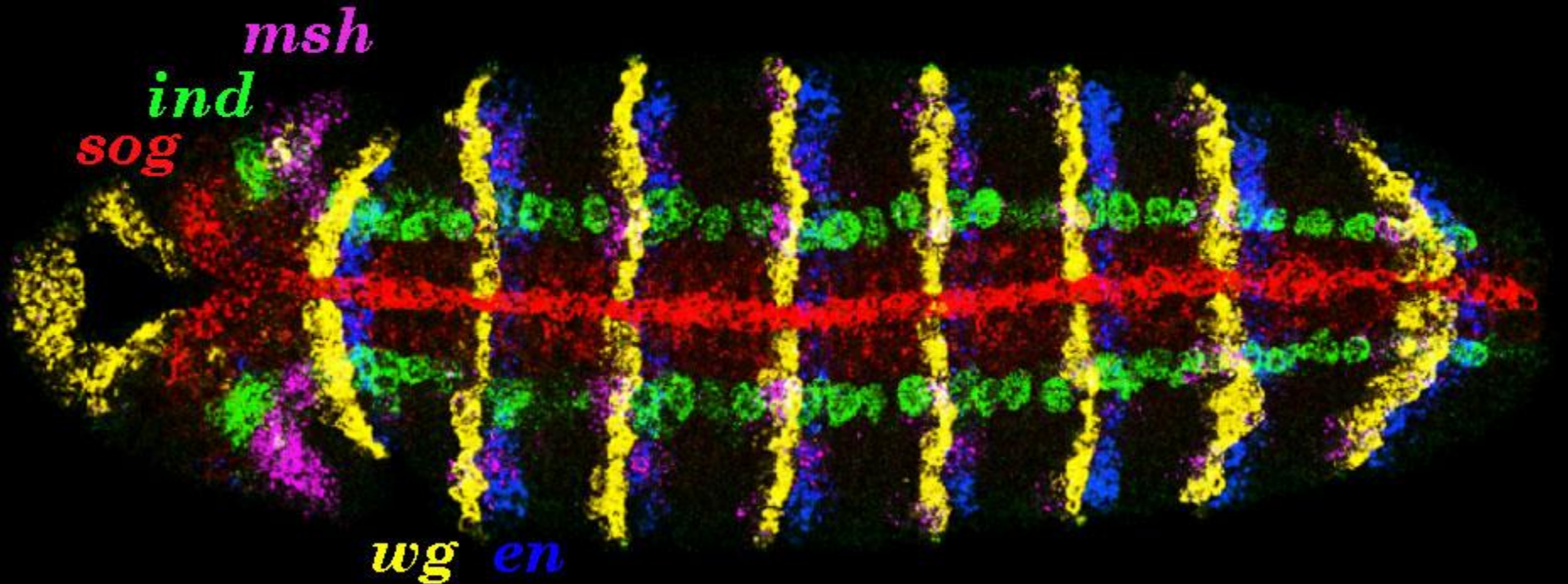
- In situ hybridization methods reveals the distribution of specific RNA molecules in cells in tissues.
- RNA molecules can hybridize when the tissue is incubated with a complementary DNA or RNA probe.
- In this way the patterns of differential gene expression can be observed in tissues, and the location of specific RNAs can be determined in cells.

# RNA expression in drosophila




Hox gene activity, Drosophila embryo
9-13
Abd-B
8
abd-A
lab
1
Dfd
4
Scr
5
Antp
6
Ubx
7
Hox homology groups

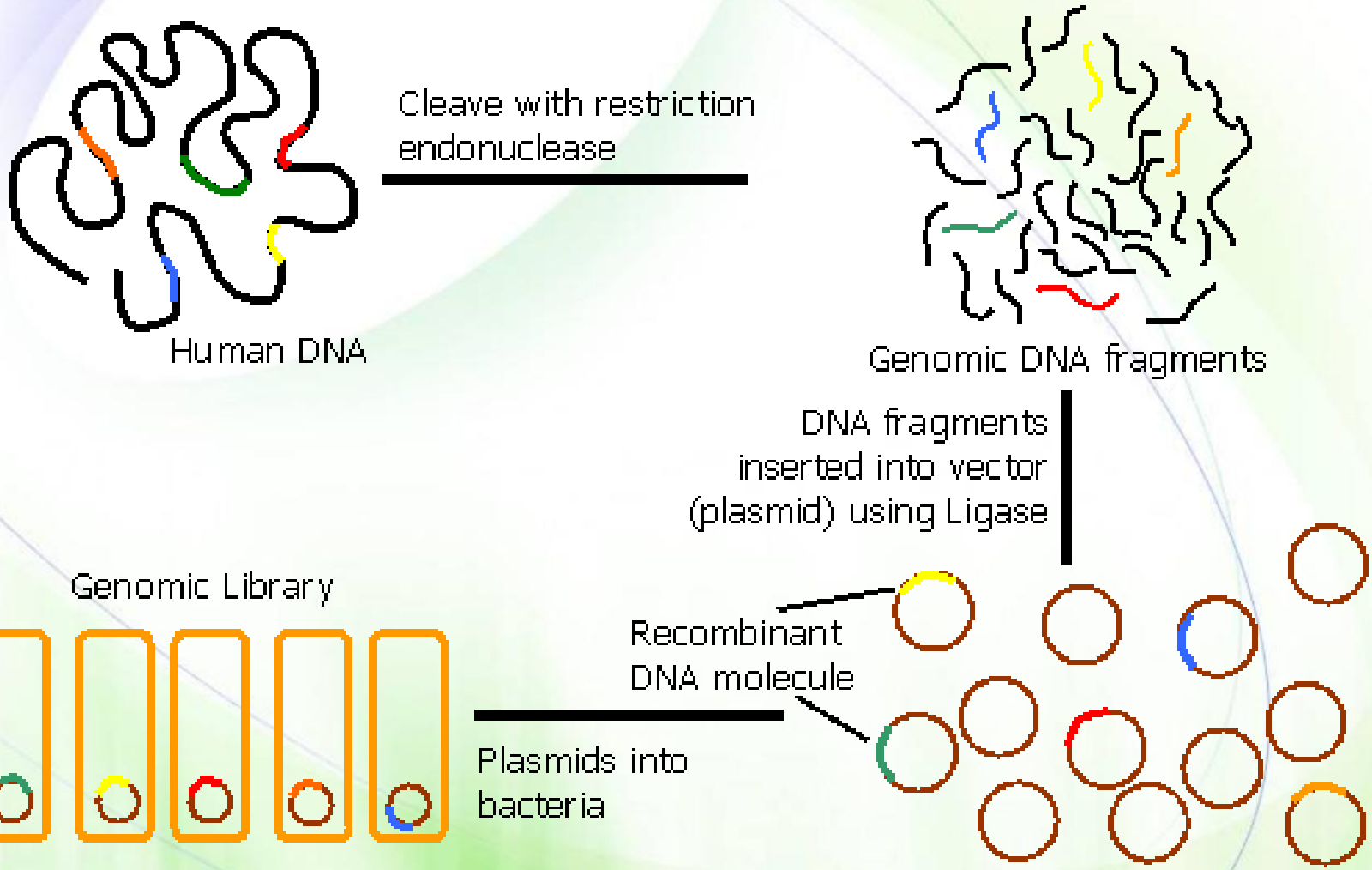
# *Multiplex detection of 5 different transcripts in a single embryo*



# What is a DNA library?

- A library can be created for DNA fragments just like book libraries.
- You can have clones of bacteria each containing a specific piece of DNA.
- You can save these clones in the freezer and take whichever clone you want to study.
- http://www.sumanasinc.com/webcontent/animations/content/dnalibrary.html

# Genomic DNA library



# cDNA library

- This library contains only those DNA sequences that are transcribed into mRNA.
- This is done by extracting the mRNA from cells and then making a complementary DNA (cDNA) copy of each mRNA molecule present.
- cDNA is made by retroviral reverse transcruptase.

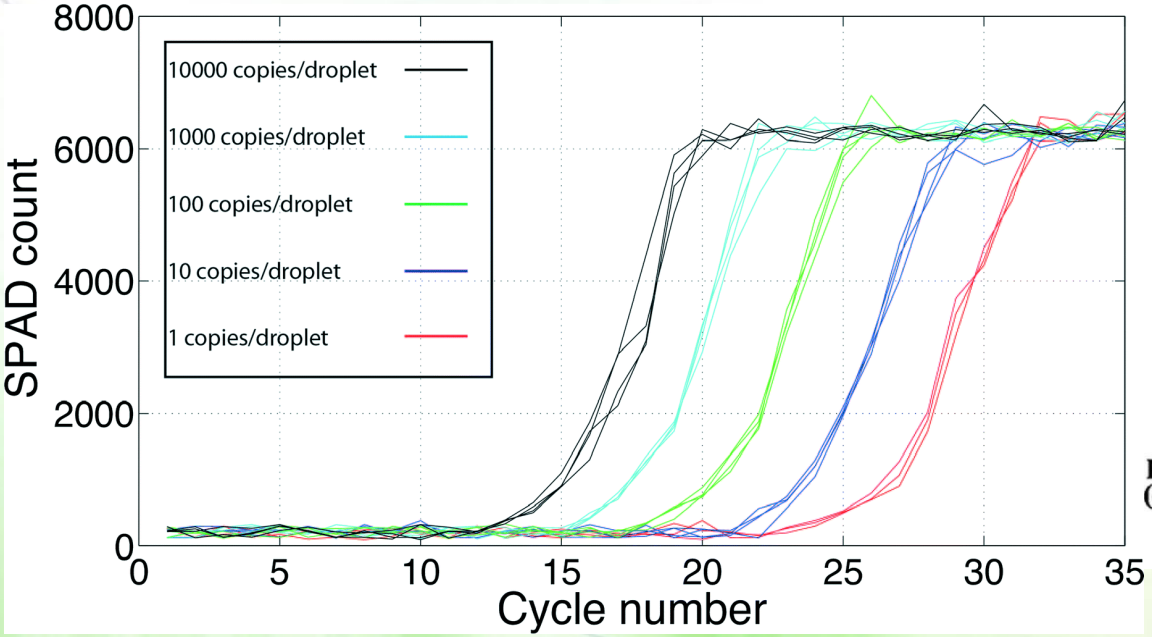
# Genomic vs. cDNA libraries

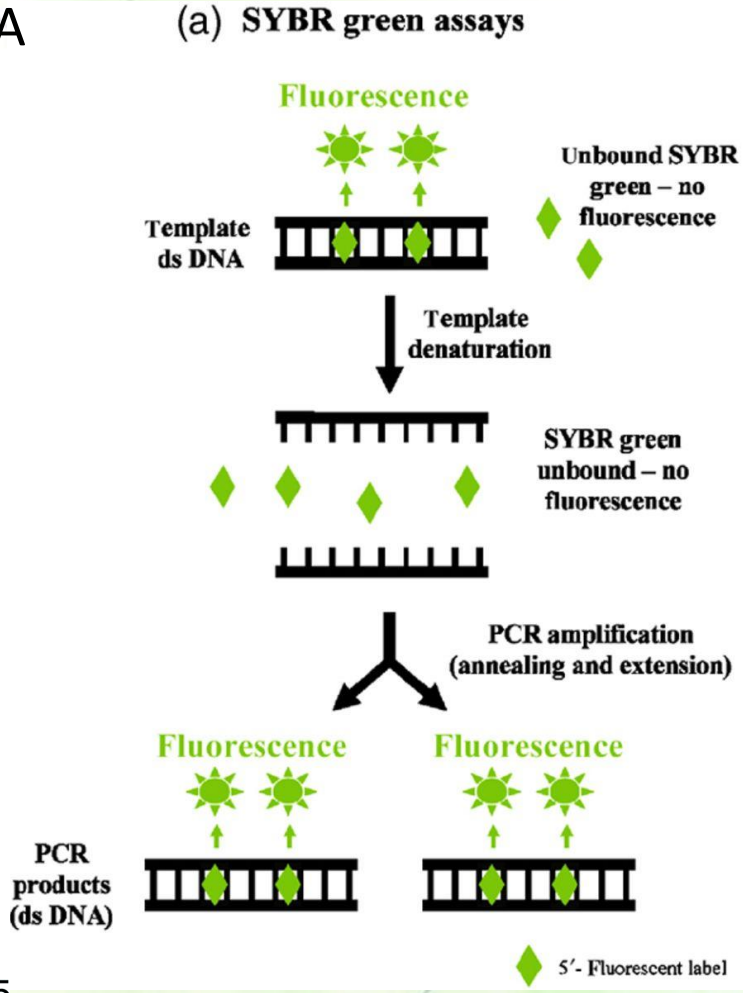
- Genomic clones represent a random sample of all of the DNA sequences in an organism. By contrast, cDNA clones contain only those regions of the genome that have been transcribed into mRNA.

- Because the cells of different tissues produce distinct sets of mRNA molecules, a distinct cDNA library is obtained for each type of cell used to prepare the library.

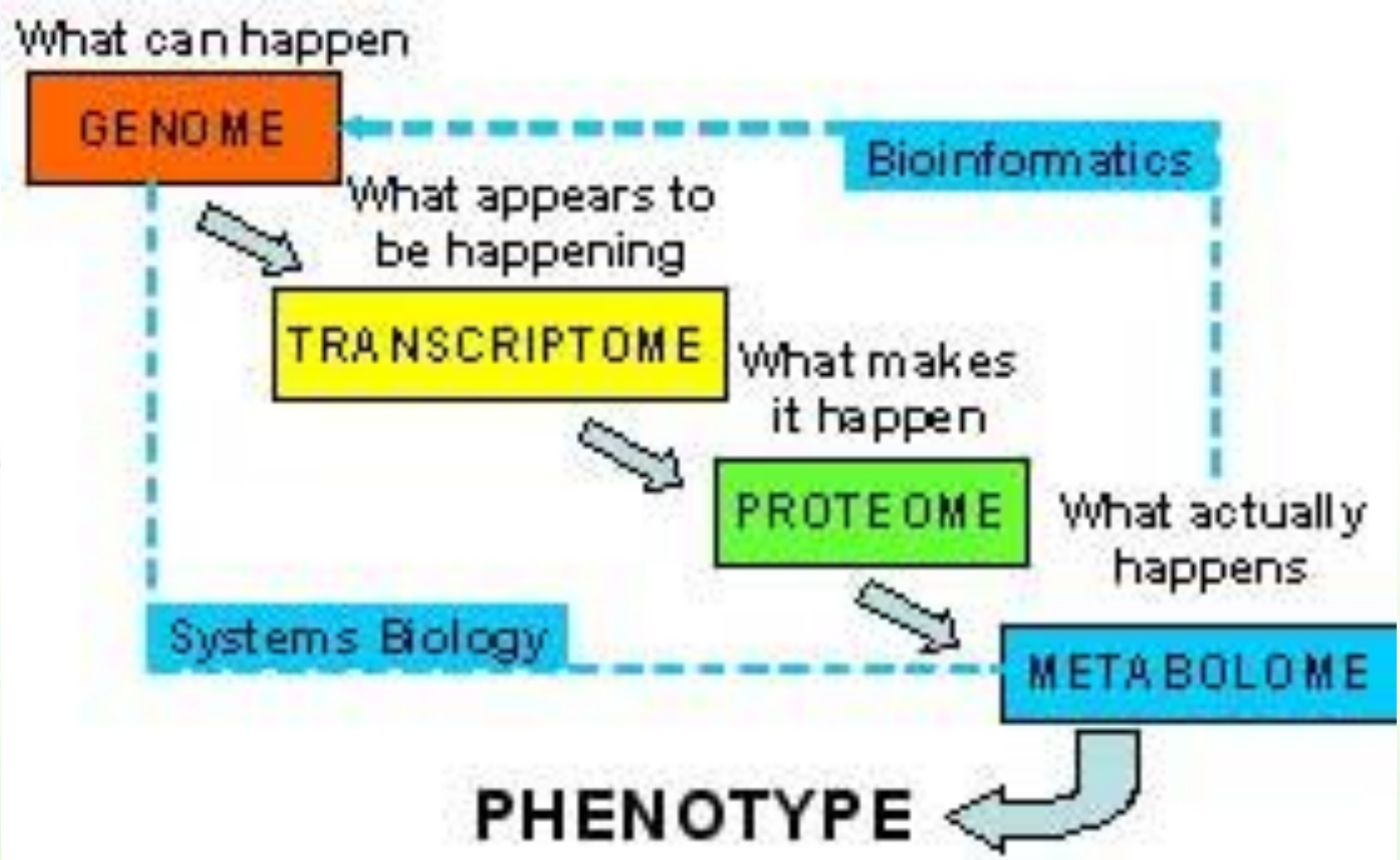
# RT-PCR of mRNA

- Another way of relative quantitation of RNA expression is by converting RNA into cDNA followed by PCR in the presence of SYBR green.
- The higher the amount of RNA (cDNA), the sooner it is detected.



# The science of -omics



# Studying the transcriptome

- One such method in studying transcriptomes is DNA microarrays, which allow the analysis of the RNA products of thousands of genes all at once.

- By examining the expression of so many genes simultaneously, we can understand gene expression patterns in physiological and pathological states.

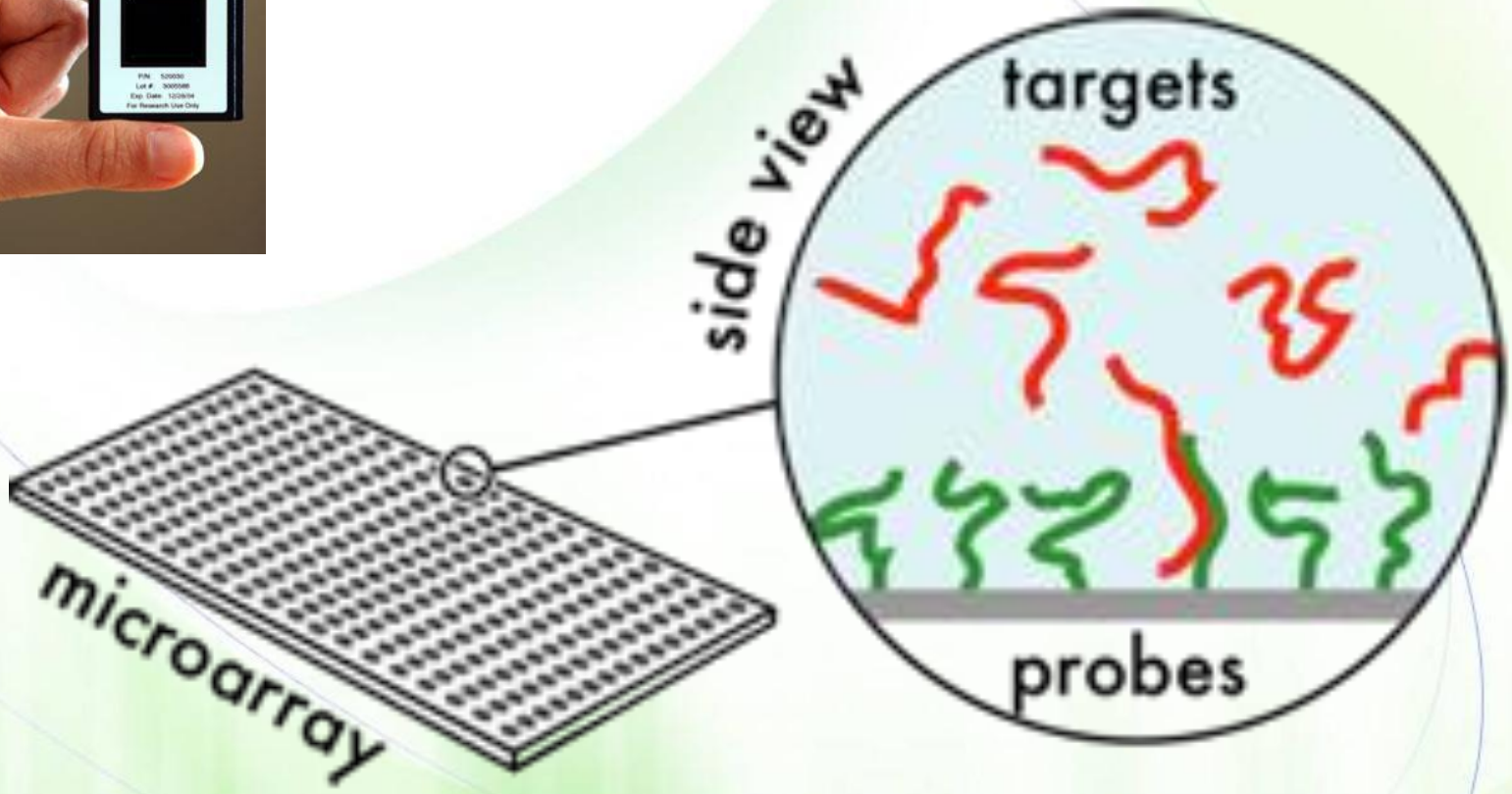
# DNA microarrays

- DNA microarrays are glass microscope slides spotted with up to tens of thousands of DNA fragments in an area the size of a fingernail.

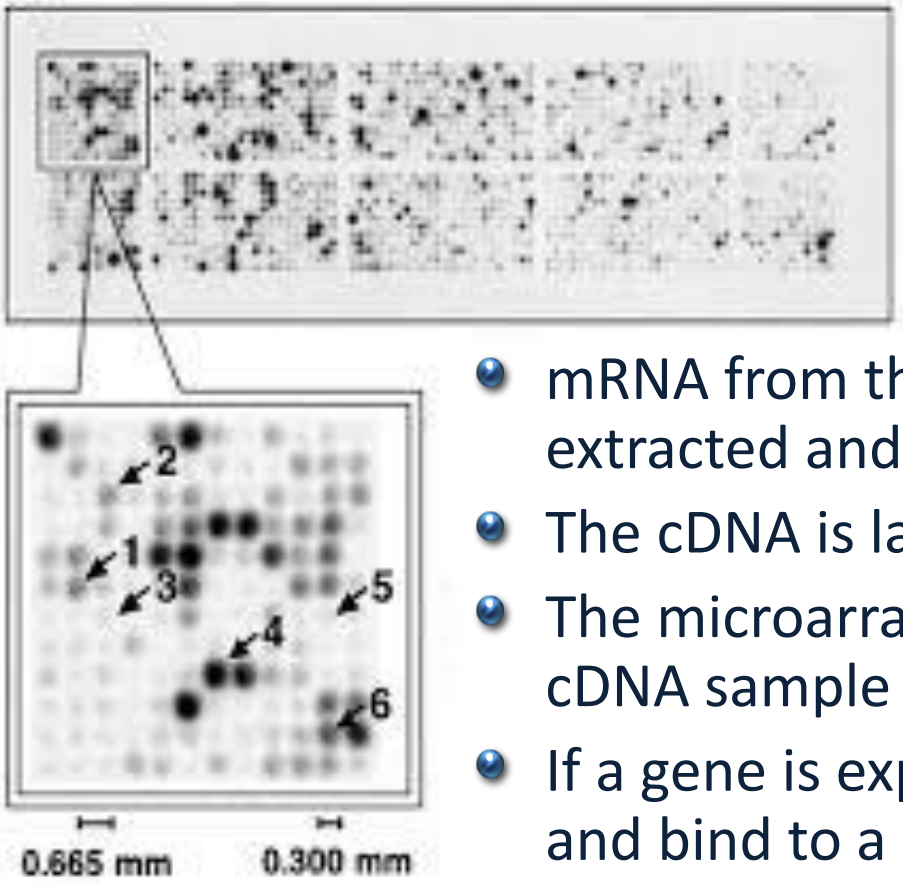
- The exact sequence and position of every DNA fragment on the array is known.

- http://learn.genetics.utah.edu/content/labs/microarray/
- http://www.sumanasinc.com/webcontent/animations/content/dnachips.html


AFFYMETRIX
GeneChip®
Xenopus laevis
Genome Array


side view
targets
probes
microarray

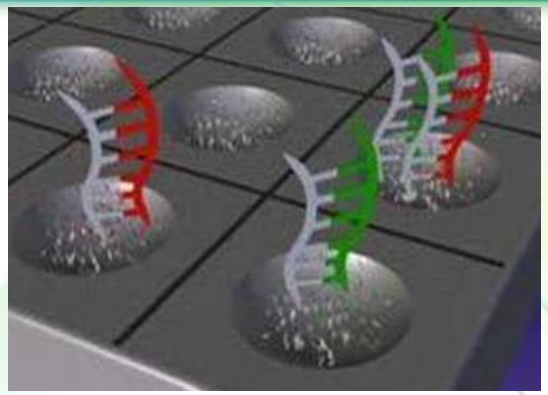
# A DNA microarray



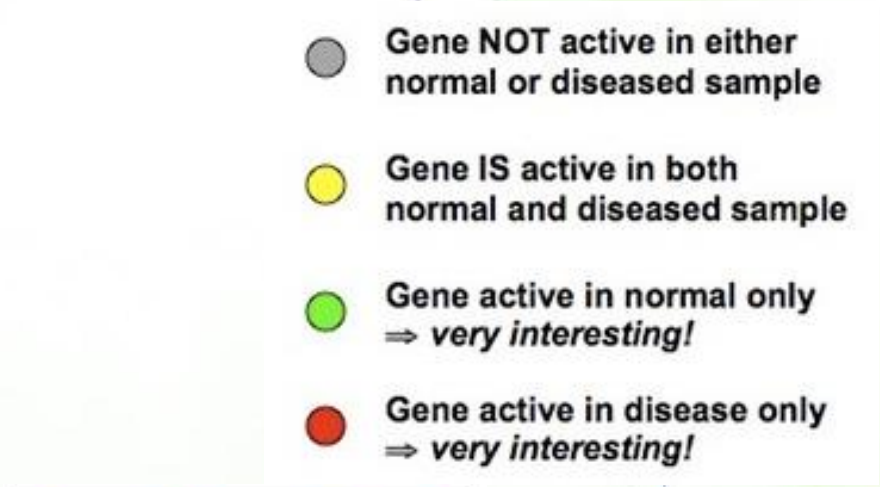

**This is done for a single sample using radioactively labeled cDNA.**

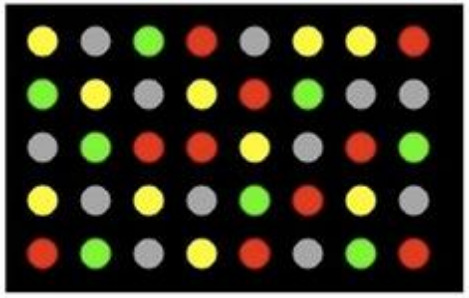
- mRNA from the cells being studied is first extracted and converted to cDNA.
- The cDNA is labeled with a radioactive probe.
- The microarray is incubated with the labeled cDNA sample for hybridization to occur.
- If a gene is expressed, then the cDNA will exist and bind to a specific complementary DNA fragment on the microarray.
- Binding can be detected since the cDNA is labeled and expression is determined.

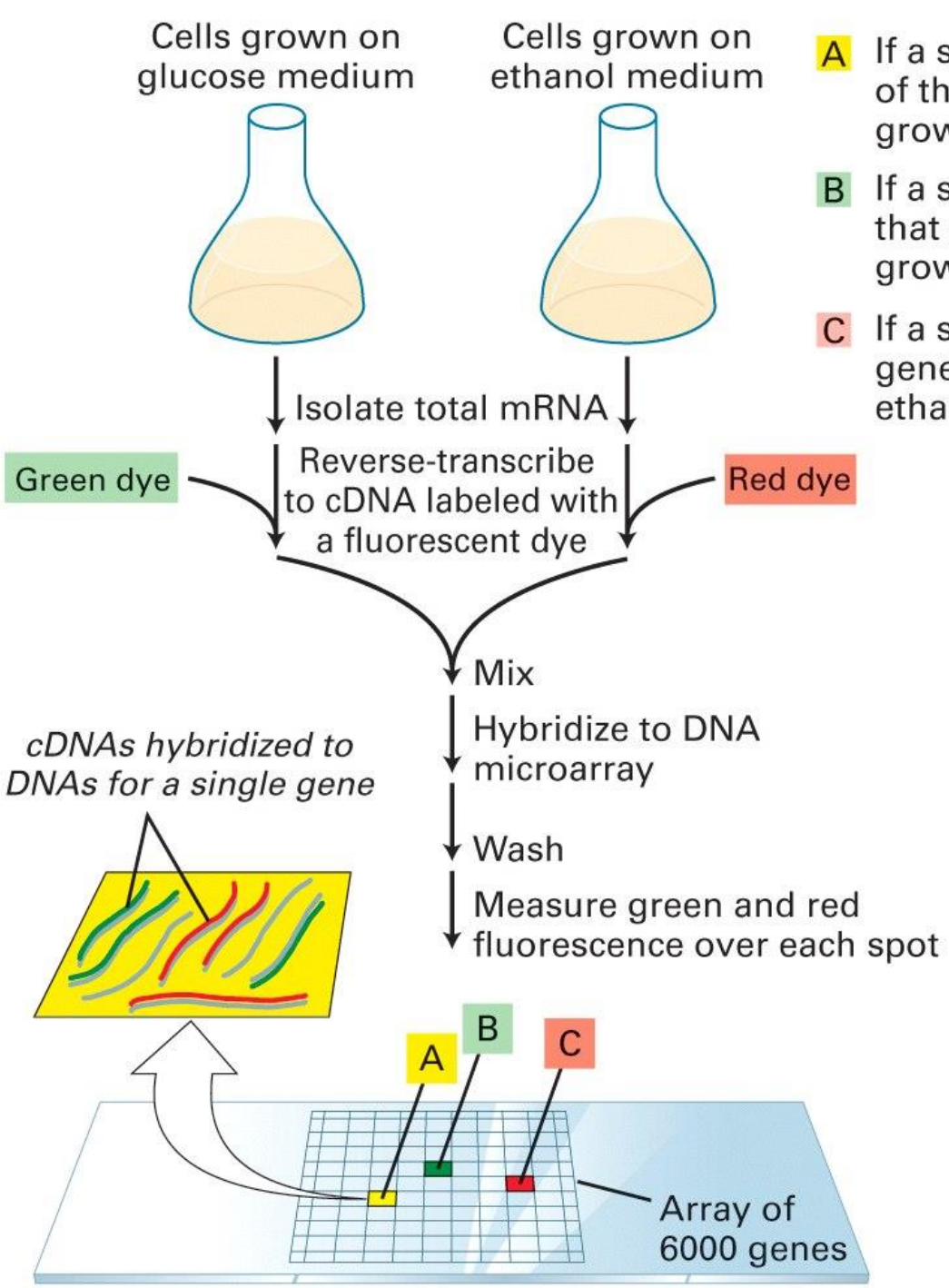
# Comparative expression



- In order to compare expression of genes two different samples, the cDNA molecules are fluorescently labeled with different colors (green and red) and added to the array.
- An increases in the amount of a RNA molecule in one sample versus the other is reflected by an increase the amount of produced cDNA and an increase in fluorescence in the bound spot.



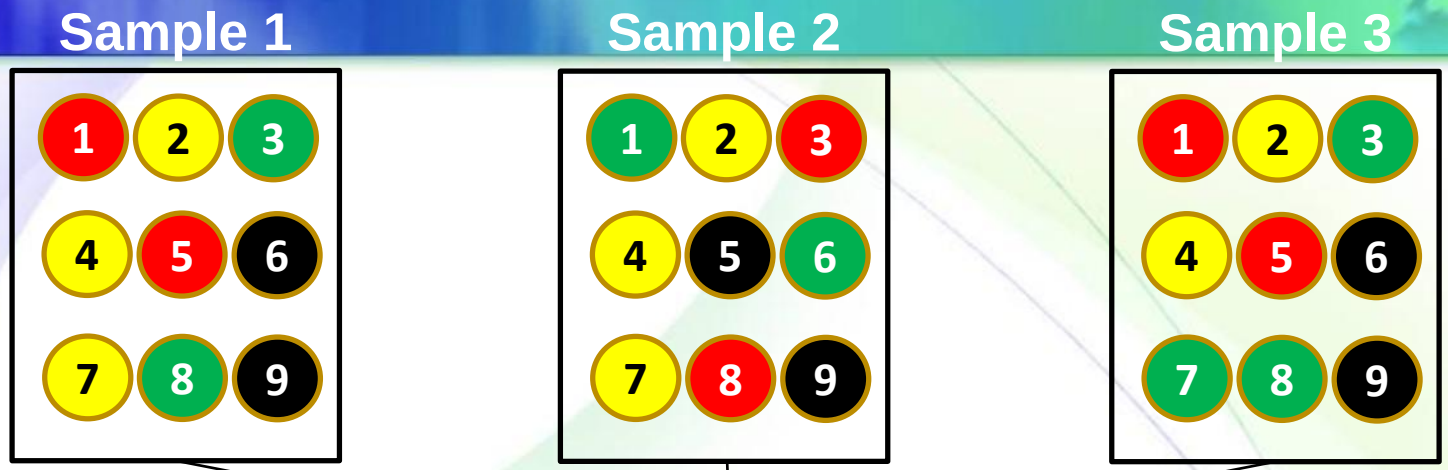





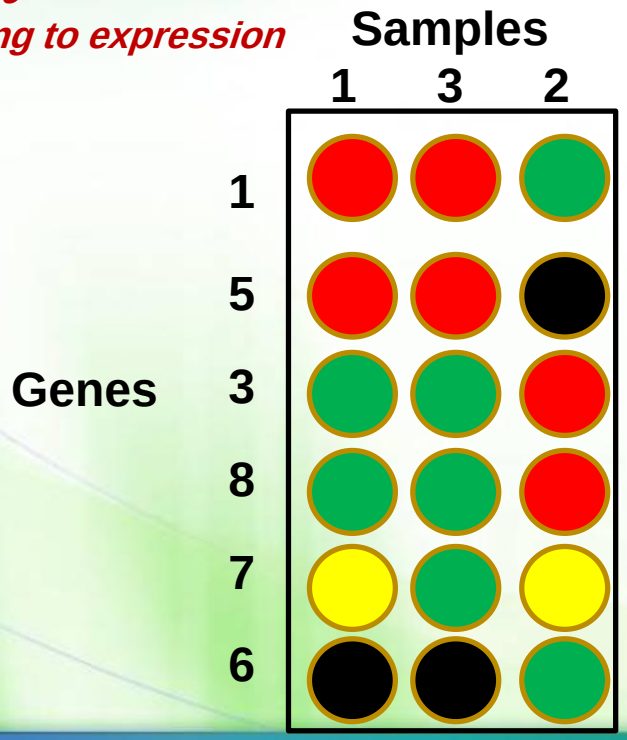

A If a spot is yellow, expression of that gene is the same in cells grown either on glucose or ethanol

B If a spot is green, expression of that gene is greater in cells grown in glucose

C If a spot is red, expression of that gene is greater in cells grown in ethanol

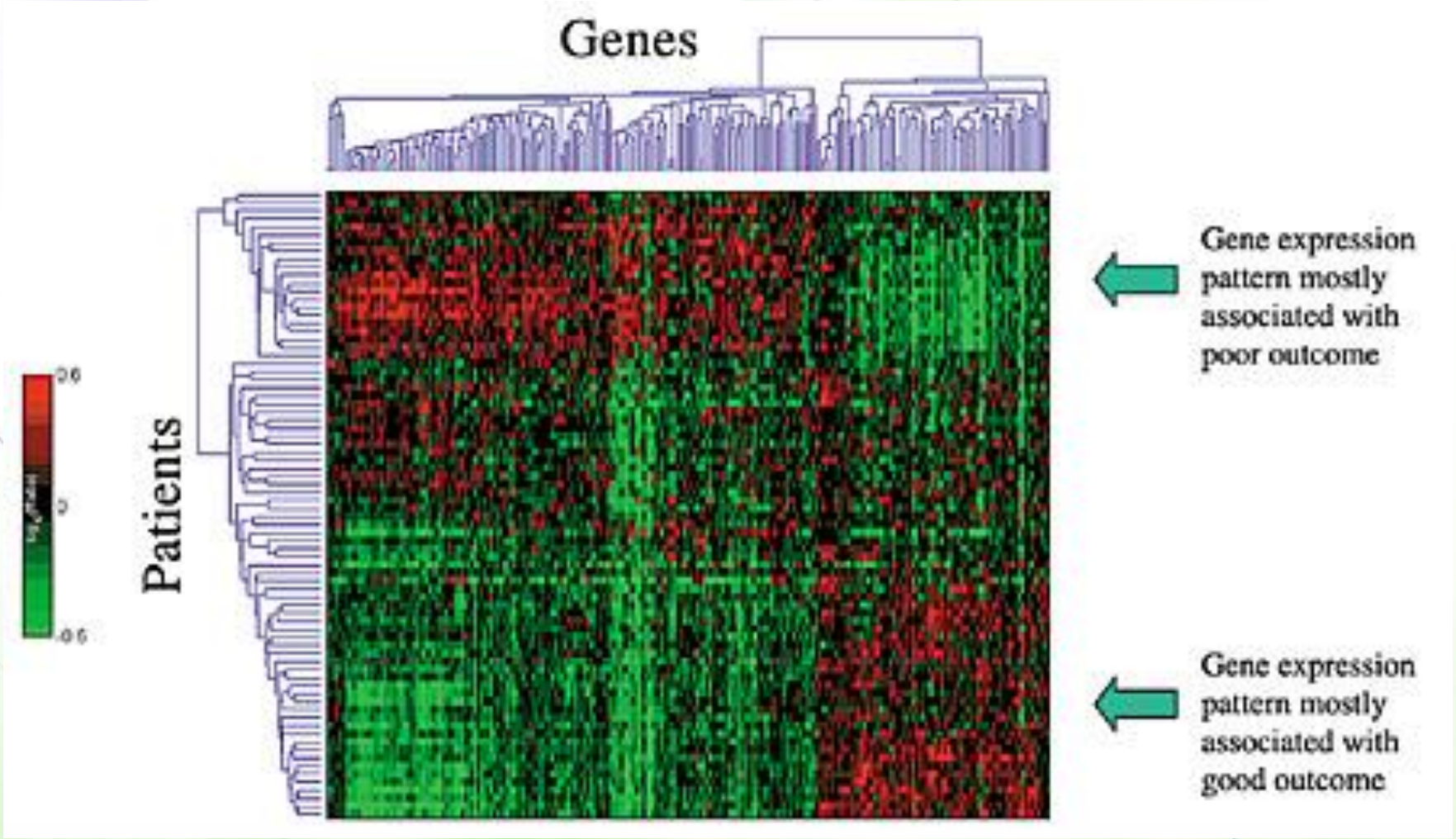



*Combine results*
*Eliminate samples 2, 4,& 9*
*Cluster samples according to expression*




**This can be done at a larger scale whereby samples are compared to the same control sample and a computer program combines all data illustrating differences in expression among the samples classifying them into different groups.**

# DNA microarrays and breast cancer



# Translation

# Resources

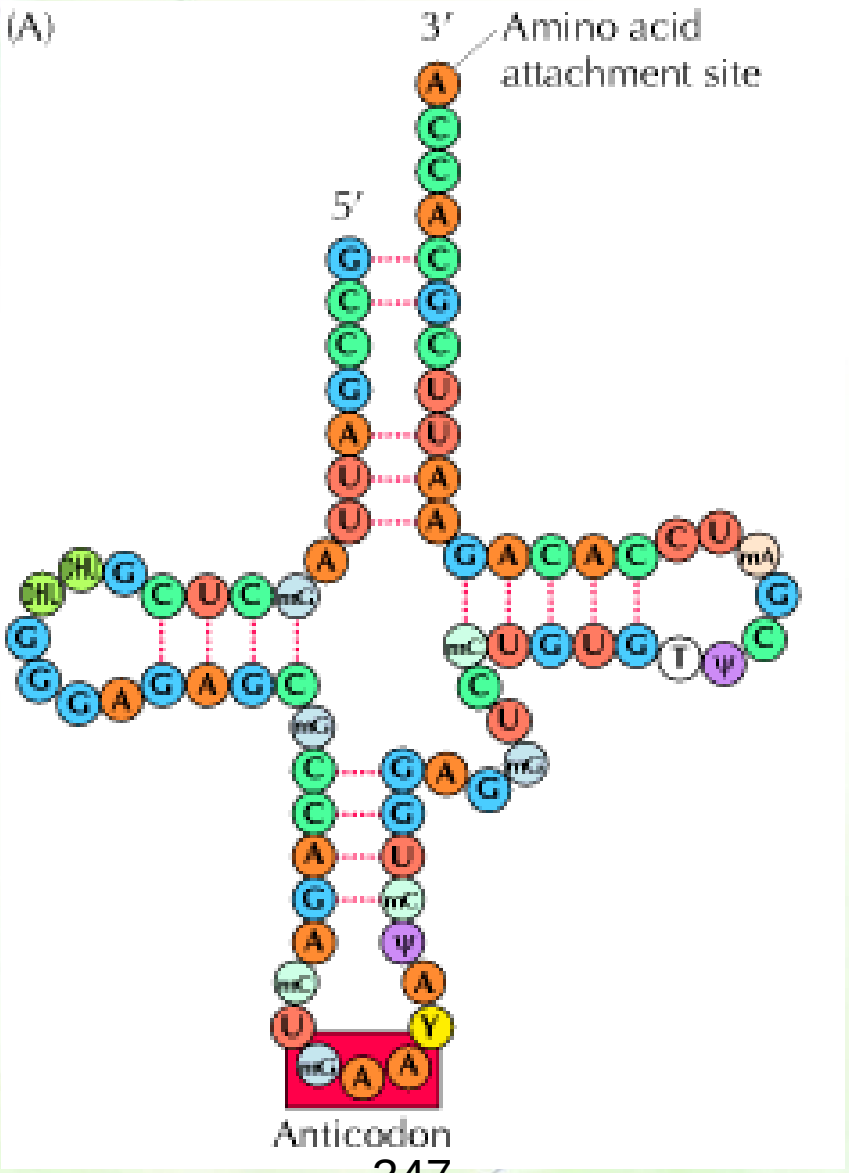
- This lecture
- Campbell and Farrell's Biochemistry, Chapter 12

# General information

- Protein synthesis involves interactions between three types of RNA molecules:
  - tRNAs
  - rRNAs
  - mRNA templates

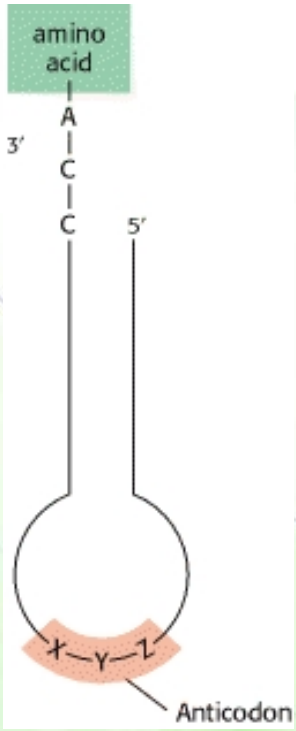
# tRNA structure

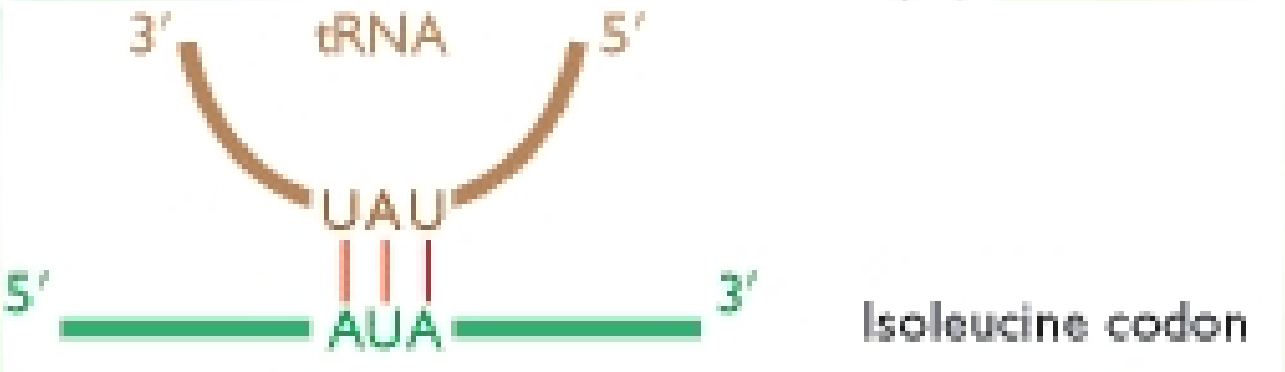
- tRNAs are short RNA molecules (80 bases long)
- An amino acid is covalently attached to the ribose of the terminal adenosine
- "charged" or "activated" tRNA carries one amino acid
- Aminoacyl-tRNA synthetases



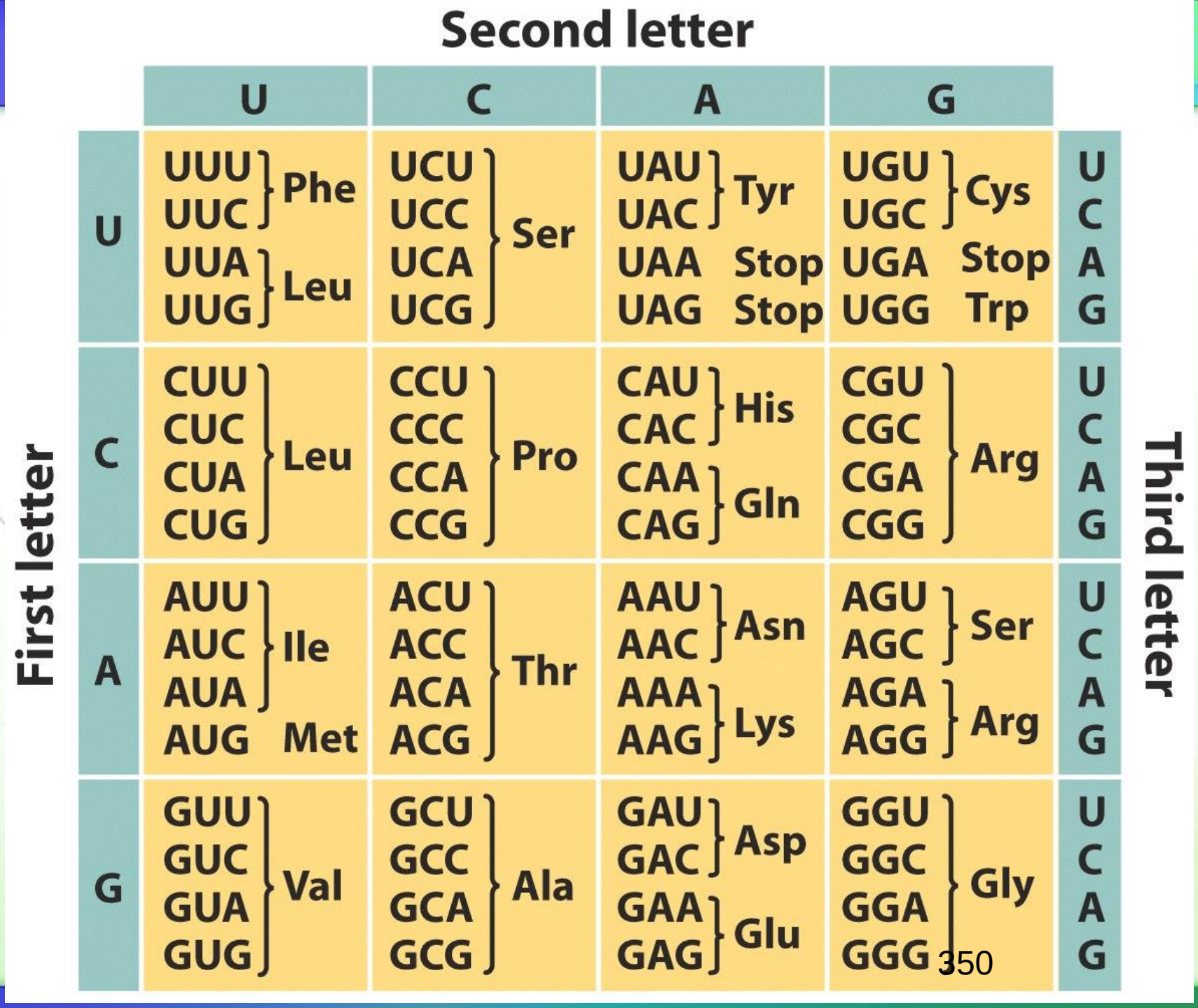
# Codon vs. anticodon

- tRNAs also contain a three-nucleotide sequence known as “anticodon” that pairs with the “codon’ or “triplet” mRNA molecules



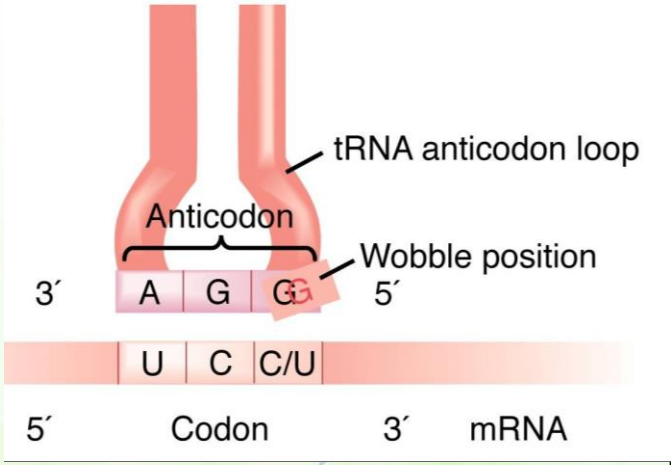
# The genetic code


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

# Features of the genetic

- Not universal
  - Example: AUA in mitochondria (methionine) in cytosol (isoleucine)
- Wobble base pairing (degenrate codon)
  - The bases that are common to several codons are usually the first and second bases, with more room for variation in the third base, which is called the “wobble” base.
  - The degeneracy of the code acts as a buffer against deleterious mutations.

| | U | C | A | G | |
|---|---|---|---|---|---|
| U | UUU, UUC } Phe<br>UUA, UUG } Leu | UCU, UCC, UCA, UCG } Ser | UAU, UAC } Tyr<br>UAA Stop<br>UAG Stop | UGU, UGC } Cys<br>UGA Stop<br>UGG Trp | U<br>C<br>A<br>G |



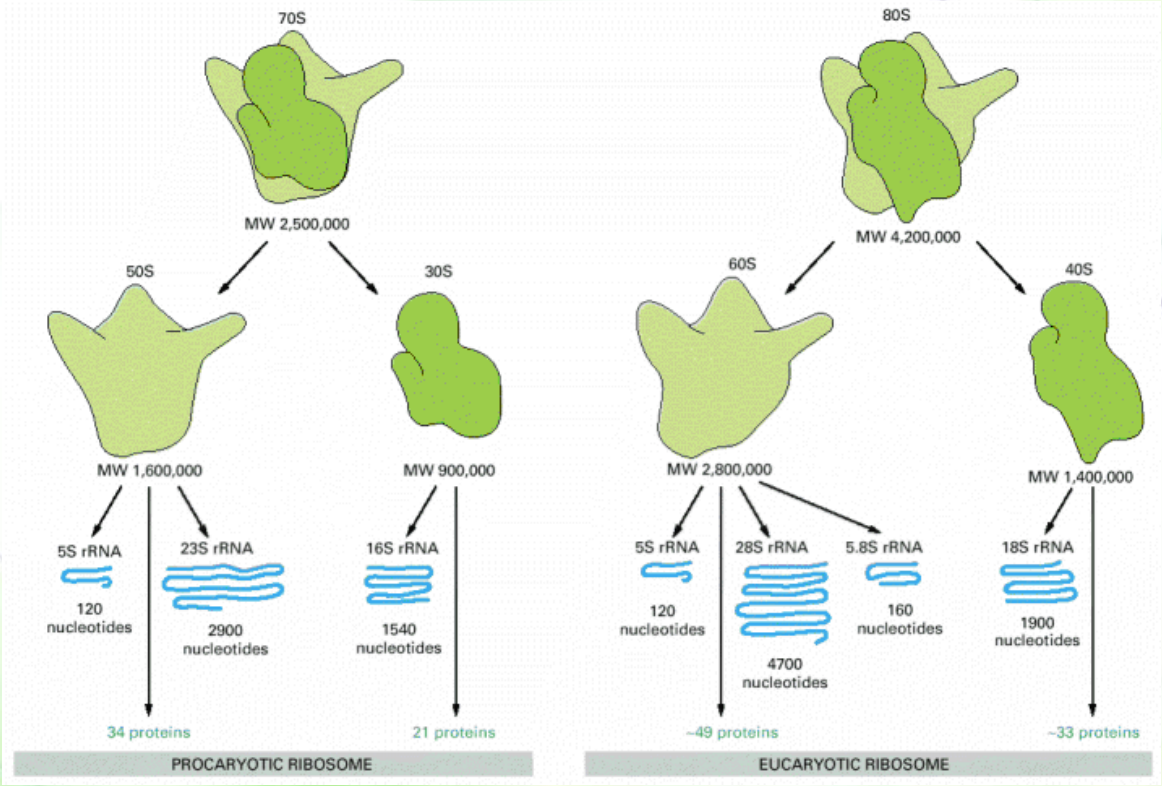
| Third-Base Relationship | Third Bases with Same Meaning | Number of Codons |
|---|---|---|
| Third-base irrelevant | U, C, A, G | 32 (8 families) |
| Purines | A or G | 12 (6 pairs) |
| Pyrimidines | U or C | 14 (7 pairs) |
| Three out of four | U, C, A | 3 (AUX = Ile) |
| Unique definitions | G only | 2 (AUG = Met) |
| | | (UGG = Trp) |
| Unique definition | A only | 1 (UGA = Stop) |

# Other features of the genetic codon

- In addition, the second base of the codon also appears to be very important for determining the type of amino acid.
  - Example: when the second base is U, all the amino acids generated are hydrophobic.
- Thus, if the second or third base were mutated, the mutation would not be silent, but the damage would not be as great because one hydrophobic amino acid would be replaced with another.

# Ribosomes

- The large ribosomal subunit catalyzes the peptidyl transferase reaction.
  - The formation of a peptide bond is an RNA-catalyzed reaction.

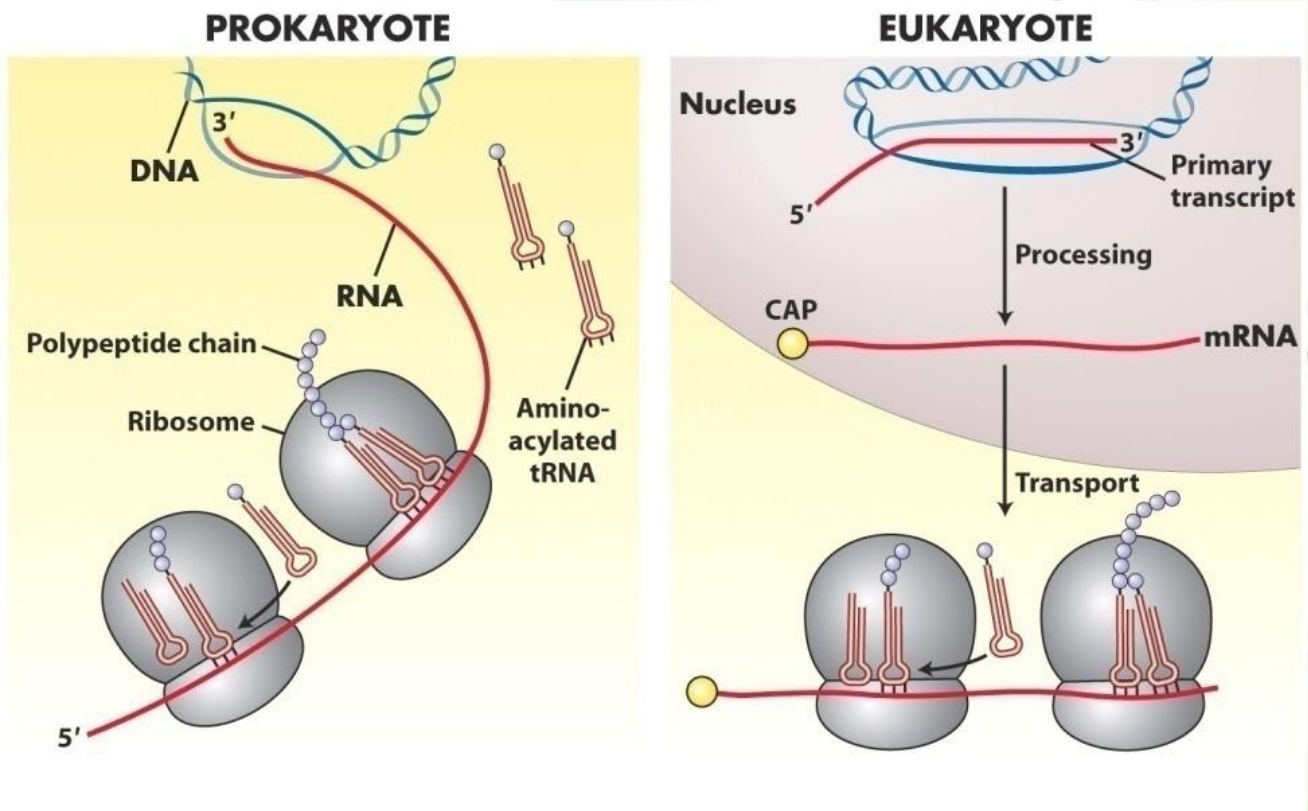


# The general mechanism of translation

- Three stages: initiation, elongation, and termination.
- The direction is 5’ $\rightarrow$ 3’.
- Protein synthesis begins at the amino terminus and extends toward the carboxyl terminus.

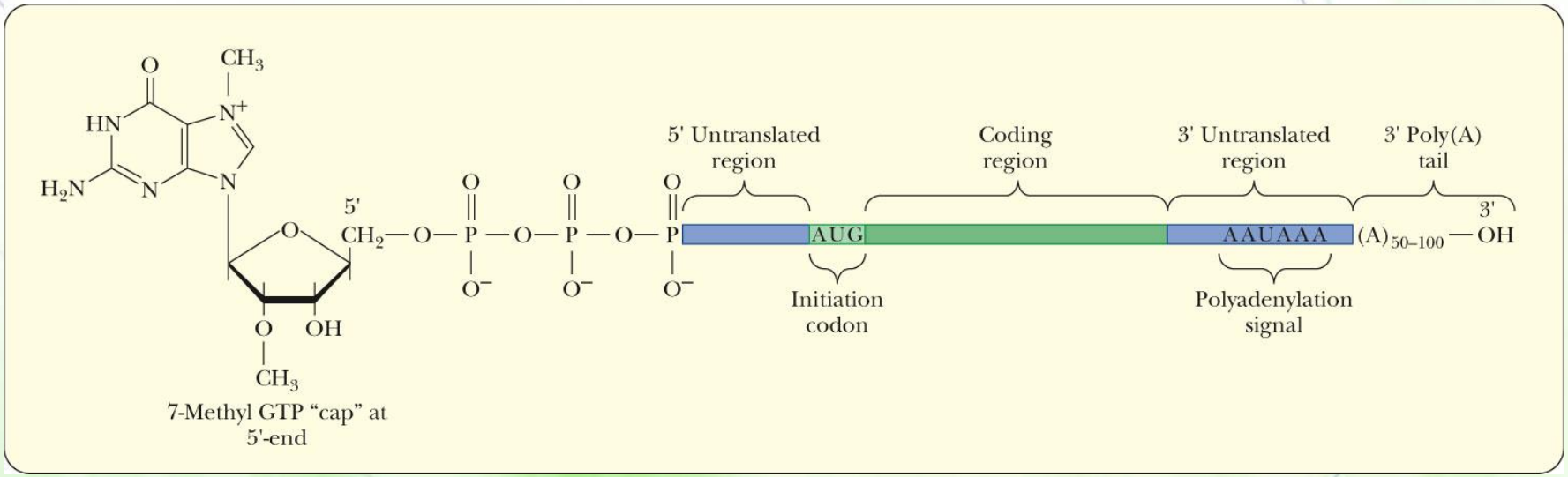
# Transcription/translation Coupling

- Translation and transcription are coupled in space and time in prokaryotes.



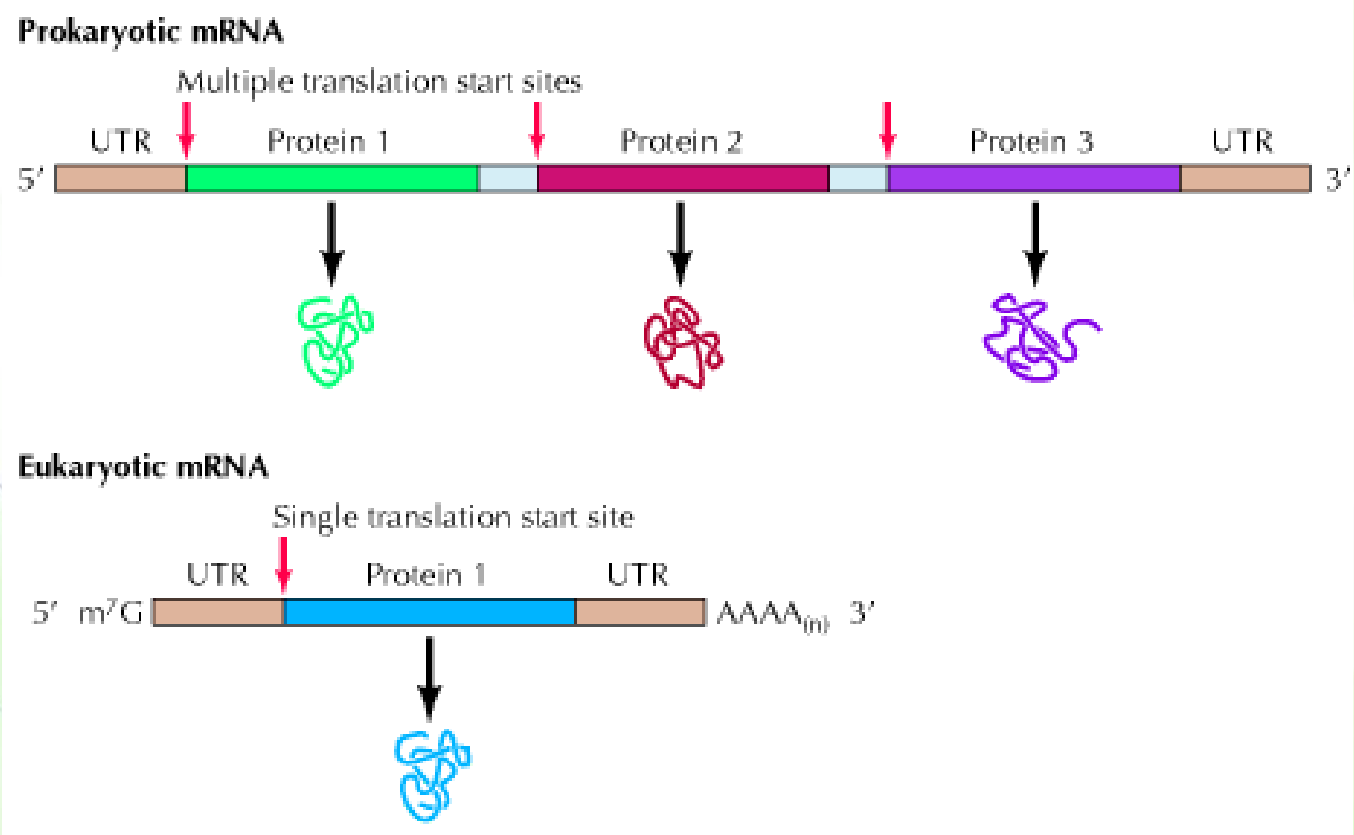
# Start of translation

- In both prokaryotes and eukaryotes, translation starts at specific initiation sites, and not from the first codon of the mRNA.
- The 5´ terminal portions upstream of the initiation sites of both prokaryotic and eukaryotic mRNAs contain noncoding sequences, referred to as 5´ untranslated regions (UTRs).
- There is also a 3'-untranslated region.

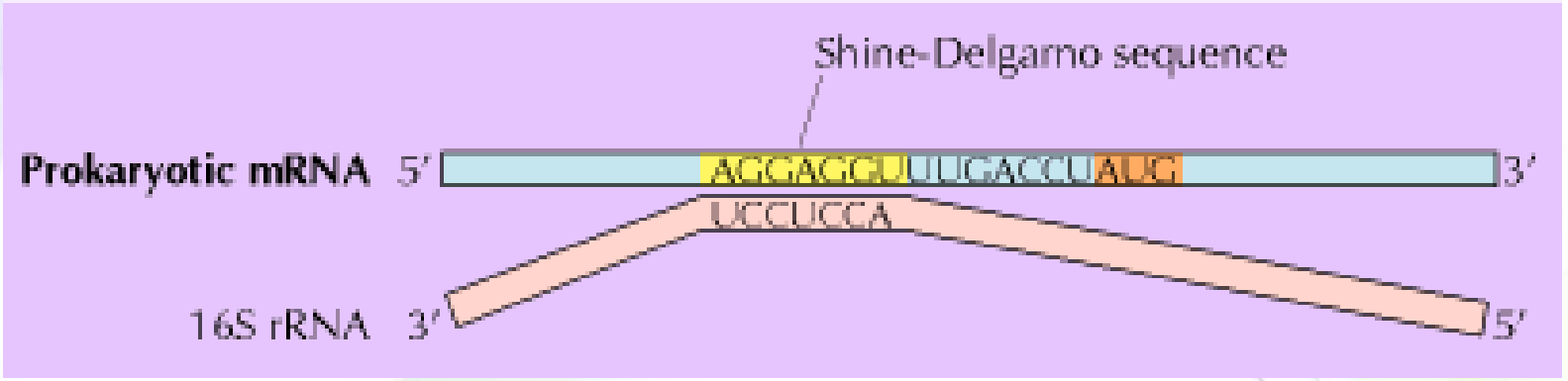


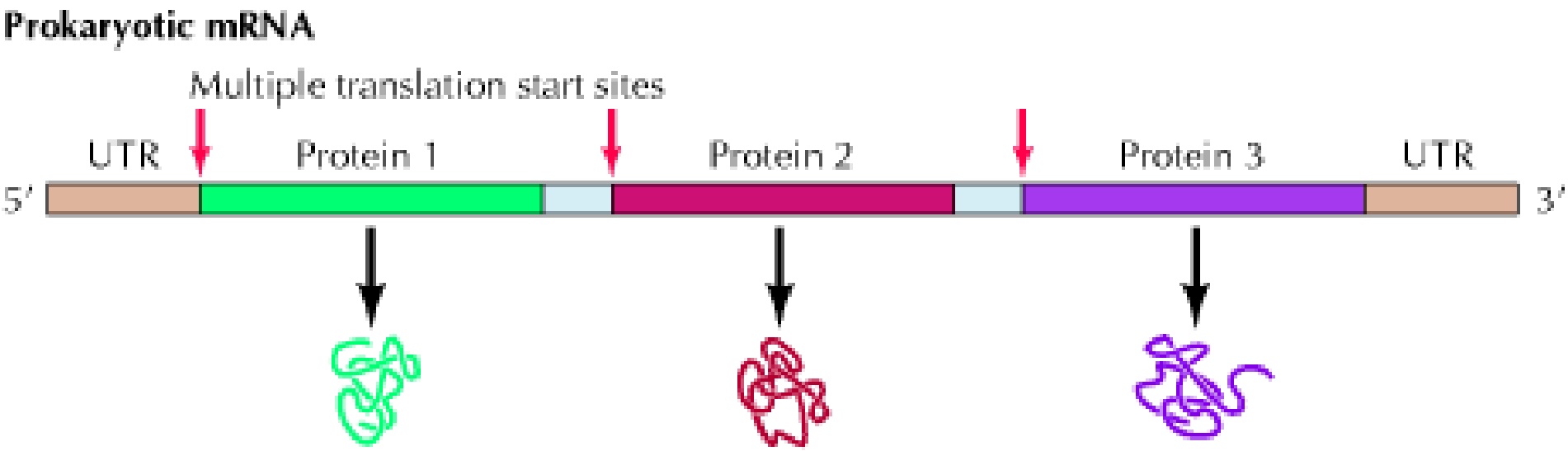
# Remember…

- Bacterial mRNA is polycistronic
- Eukaryotic mRNA is monocistronic



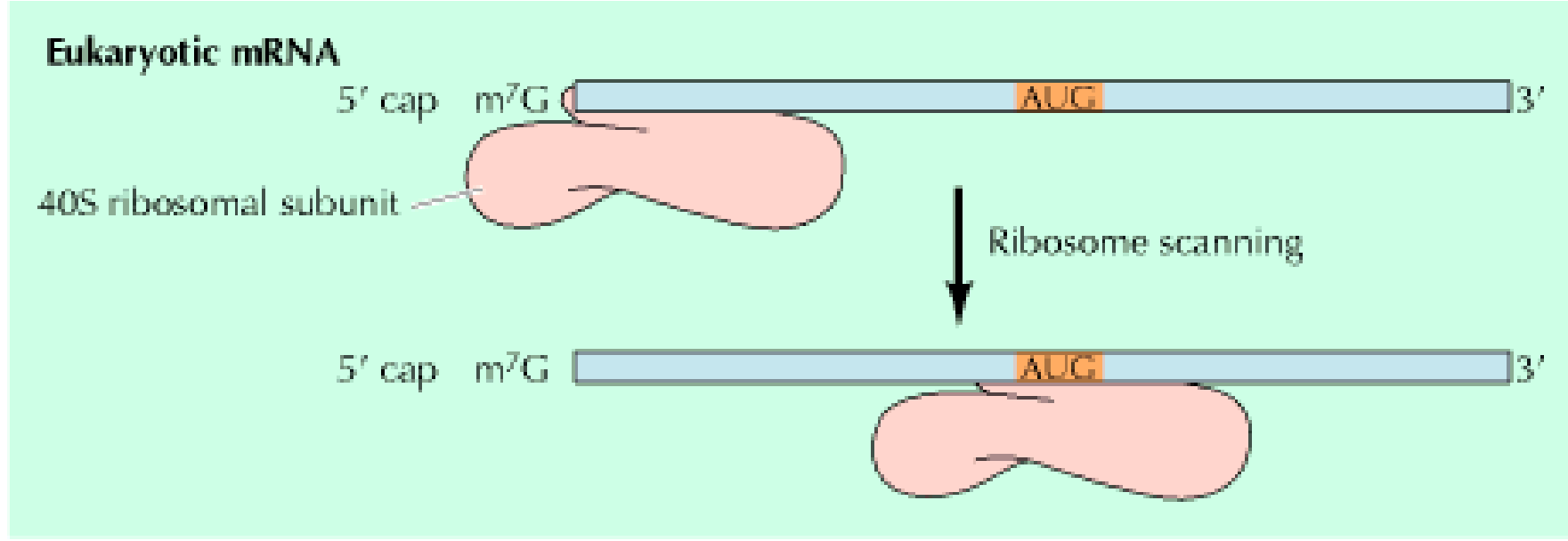
# Shine-Dalgarno sequence



# But in eukaryotes...

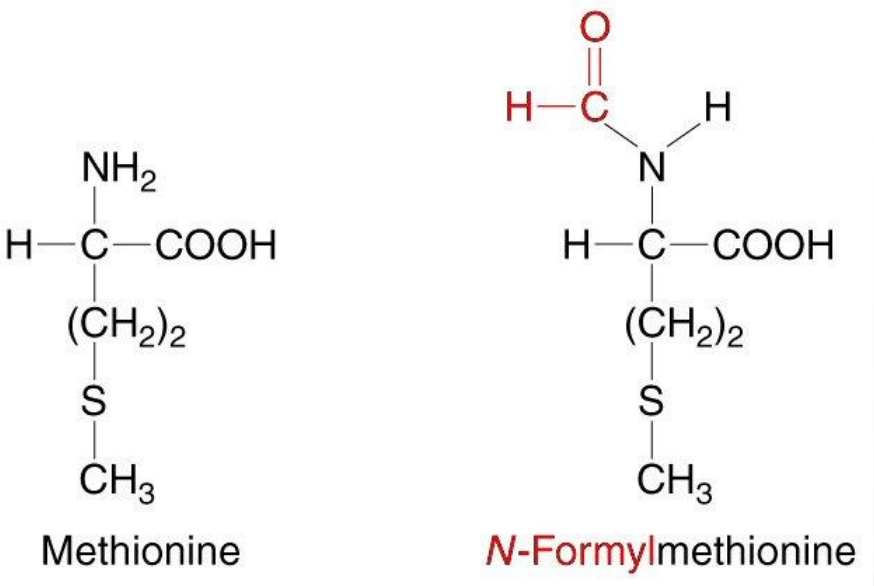
- Eukaryotic ribosomes recognize mRNAs by binding to the 7-methylguanosine cap at their 5´ terminus
- The ribosomes then scan downstream for the AUG initiation codon (Kozak sequence)



# The first amino acid

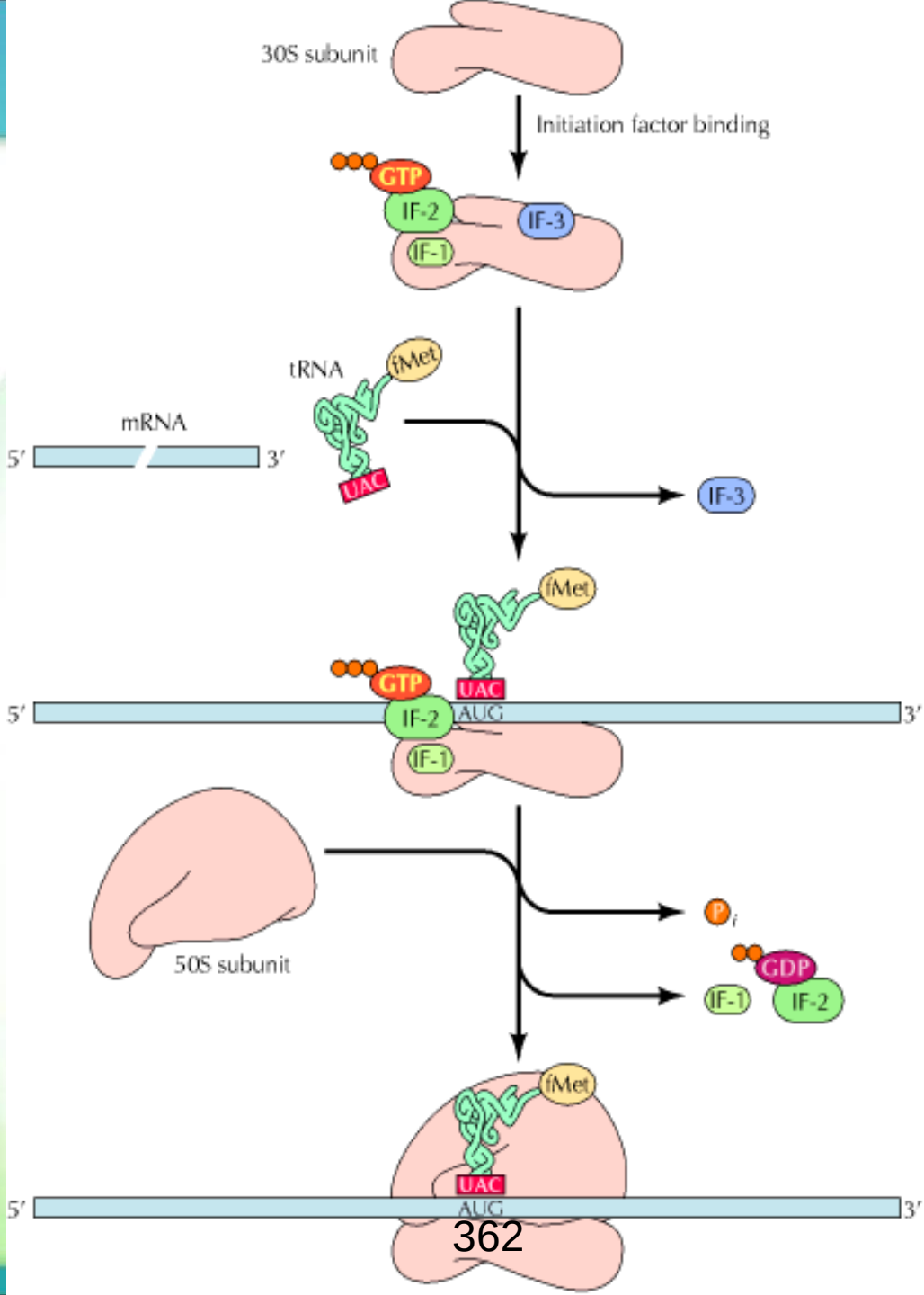
- Translation always initiates with the amino acid methionine, usually encoded by AUG.

- In most bacteria, it is N-formylmethionine.



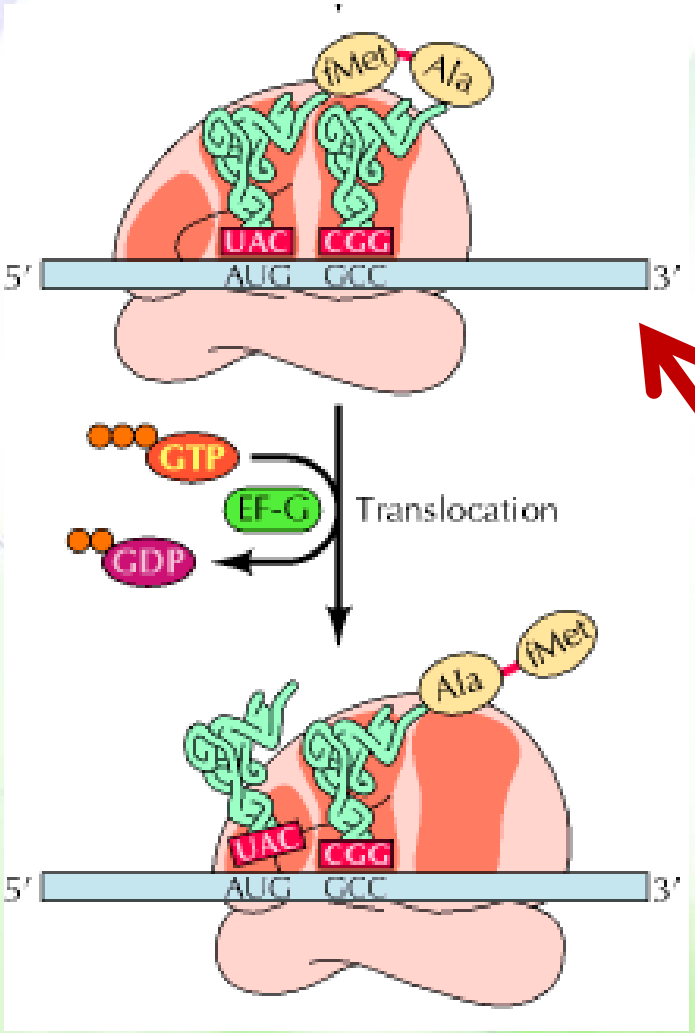
# Translation initiation

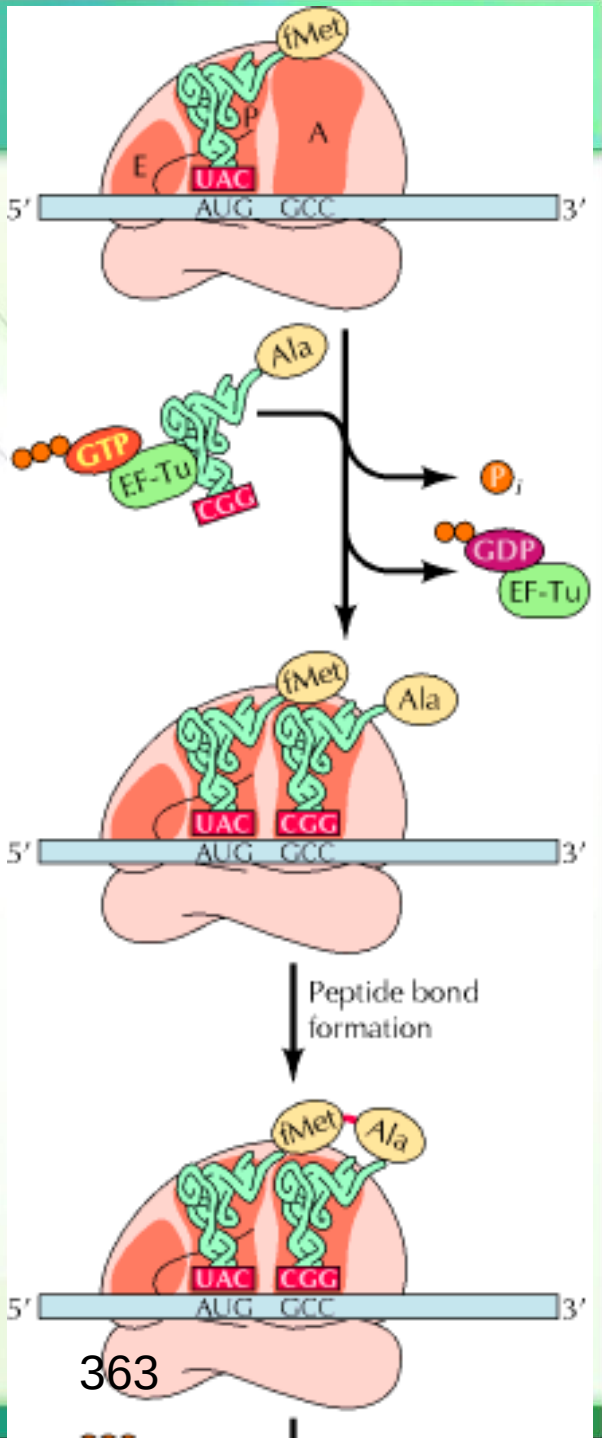
- The 30S ribosomal subunit binds to mRNA and fmet-tRNA in the presence of GTP and the three initiation factors, IF-1, IF-2, and IF-3, forming the 30S initiation complex.
- The 50S ribosomal subunit is added, forming the 70S initiation complex.



# Translation elongation I

Three steps: aminoacyl-tRNA binding, peptide bond formation, and translocation.



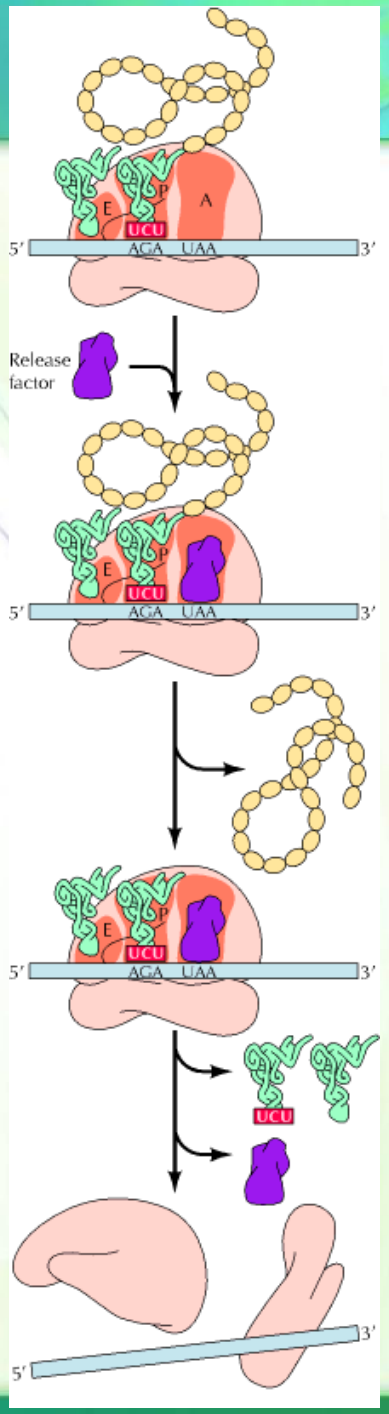



# Details of elongation

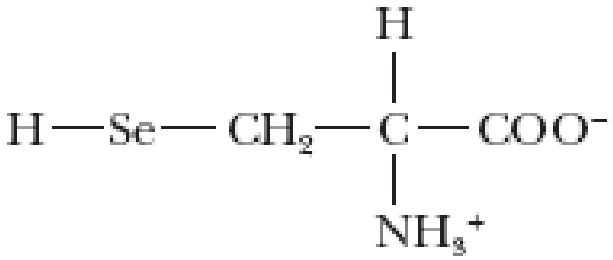
- Step 1: An aminoacyl-tRNA is bound to the A site on the ribosome. Elongation factor EF-Tu (Tu) and GTP are required. The P site on the ribosome is already occupied.
- Step 2: Elongation factor EF-Tu is released from the ribosome and regenerated
- Step 3: The peptide bond is formed, leaving an uncharged tRNA at the P site.
- Step 4: the uncharged tRNA is released. The peptidyl-tRNA is translocated to the P site, leaving an empty A site. The uncharged tRNA is translocated to the E site and subsequently released.

# Translation termination

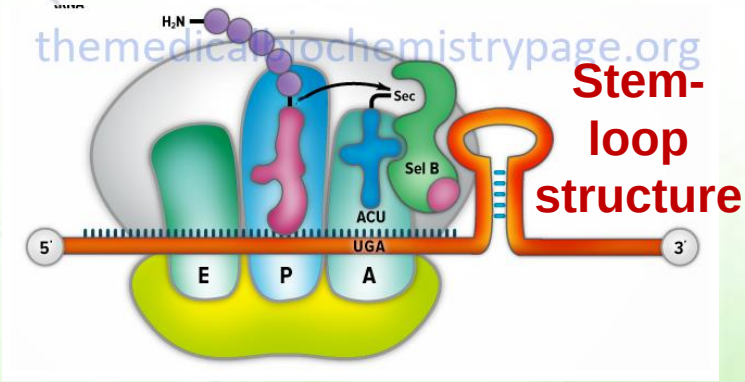
- A stop signal is required for the termination of protein synthesis. The codons UAA, UAG, and UGA are the stop signals. These codons are not recognized by any tRNAs, but they are recognized by proteins called release factors.
- The release factor blocks the binding of a new aminoacyltRNA and facilitates the hydrolysis of the bond between the carboxyl end of the peptide and the tRNA.
- Then, the whole complex dissociates.

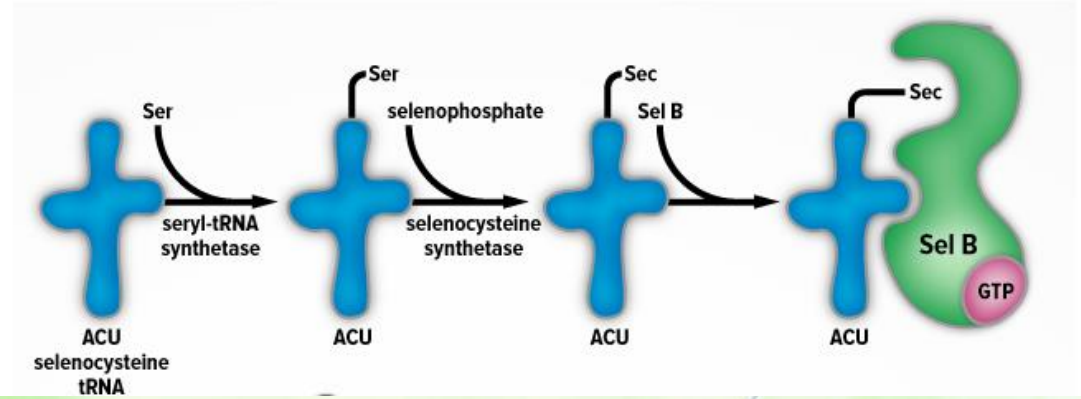


# Selenocysteine



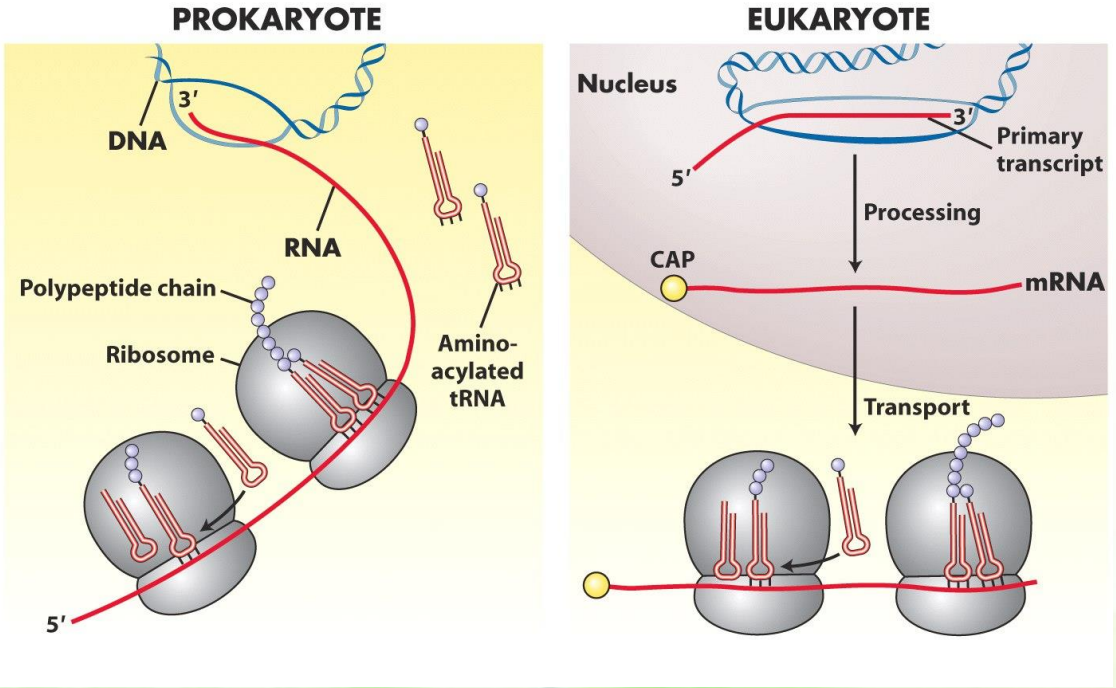

- The oxygen of a serine, which is bound to a special tRNA molecule called $tRNA^{sec}$, is replaced by selenium.
- This tRNA molecule has an anticodon that matches the UGA stop codon. In special cases, the UGA is not read as a stop; rather, the selenocysteinetRNAis loaded into the A site and translation continues.

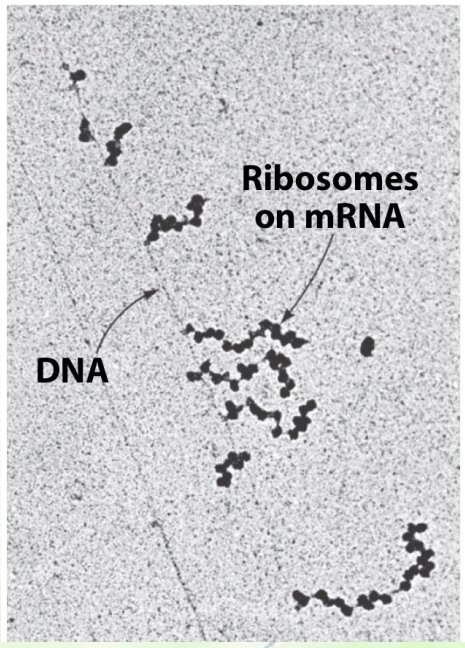


# Polyribosomes (polysomes)

- A single mRNA molecule is translated by several ribosomes simultaneously. Each ribosome produces one copy of the polypeptide chain specified by the mRNA. When the protein has been completed, the ribosome dissociates into subunits that are used in further rounds of protein synthesis.



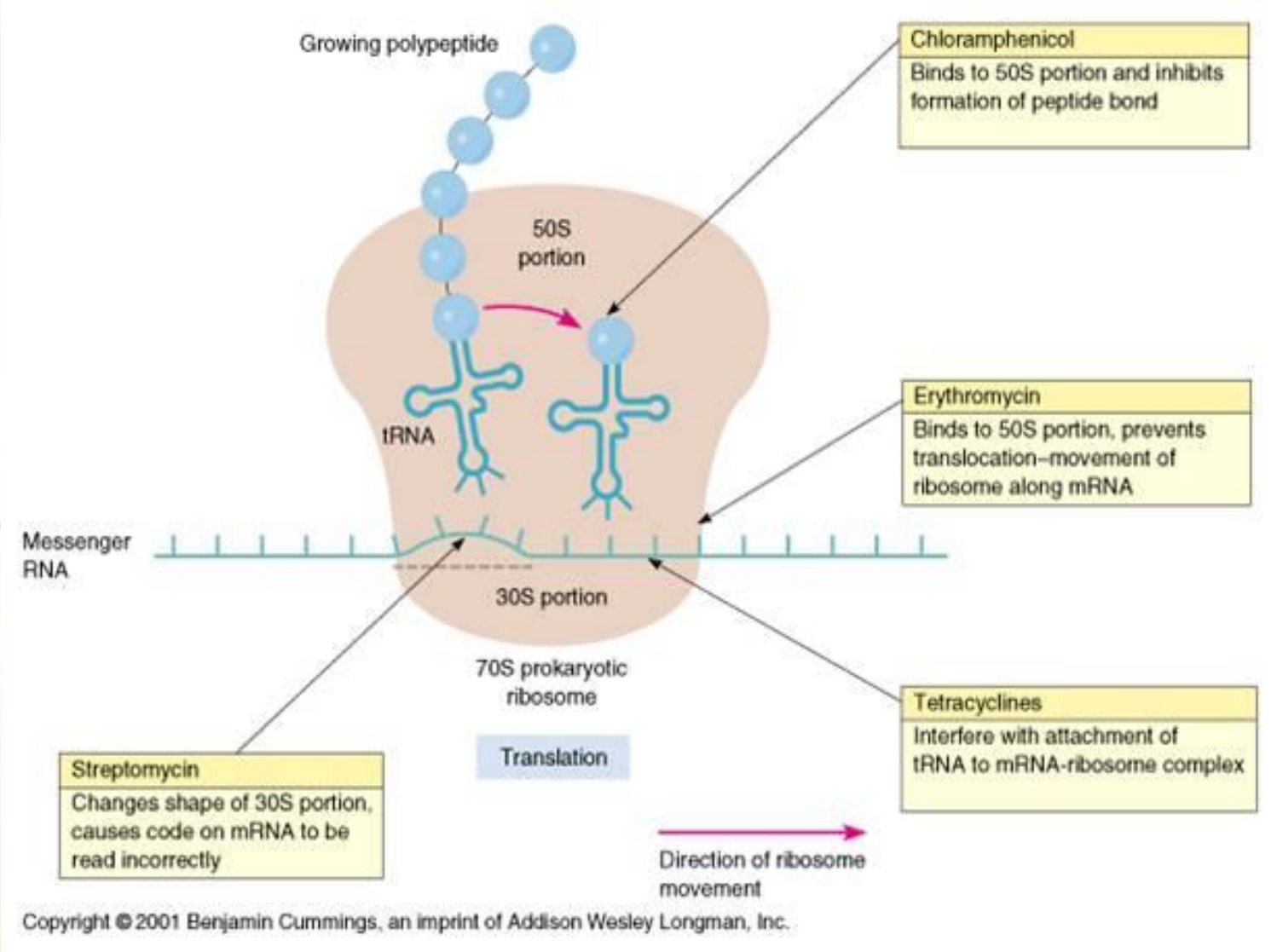



# Inhibitors of translation

| INHIBITOR | SPECIFIC EFFECT |
|---|---|
| Tetracycline | blocks binding of aminoacyl-tRNA to A-site of ribosome |
| Streptomycin | Induces binding of wrong t-RNA-AA complexes resulting in false proteins |
| Chloramphenicol | blocks the peptidyl transferase reaction on ribosomes |
| Erythromycin | blocks the translocation reaction on ribosomes |

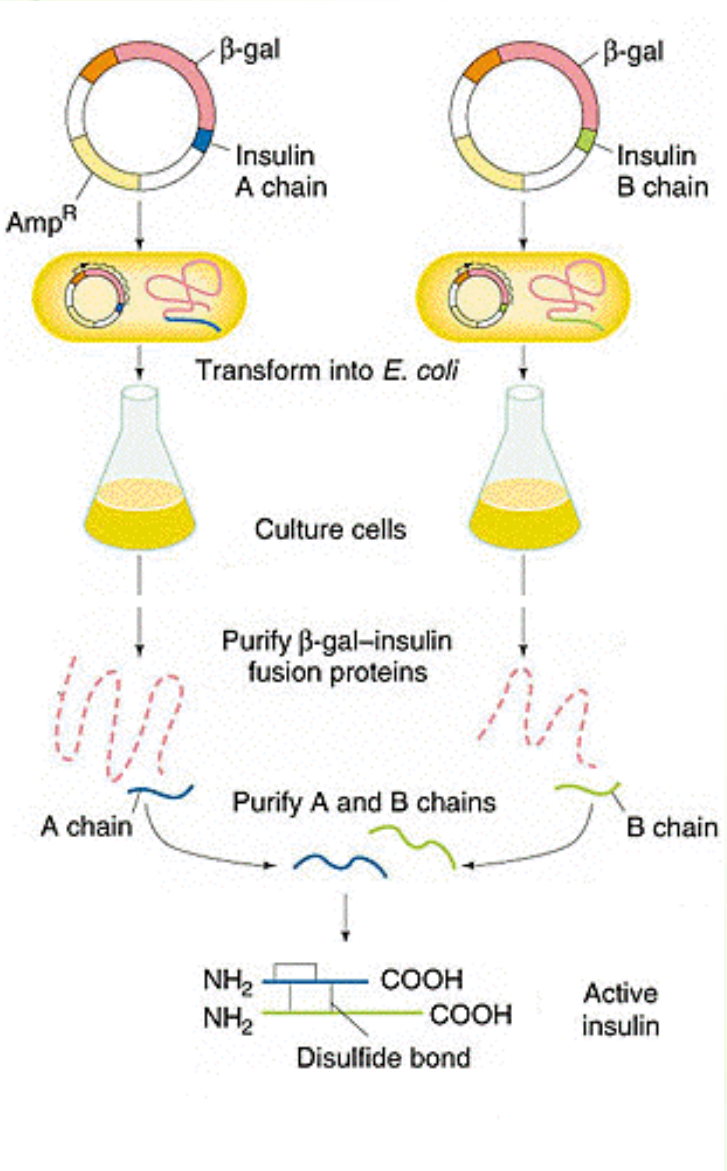
**In eukaryotes, diphtheria toxin is a protein that interferes with protein synthesis by decreasing the activity of the eukaryotic elongation factor eEF2.**

# Inhibitors of translation



# A benefit of cloning

- Production of eukaryotic proteins in bacteria (example: Insulin)
- Challenges: insulin is a dimer lined by disulfide bonds and produced from genes containing introns.
- Solution: synthetic DNA is made for each polypeptide, inserted into bacteria separately. The polypeptides are purified from each bacterial batch and mixed to form the mature insulin protein.

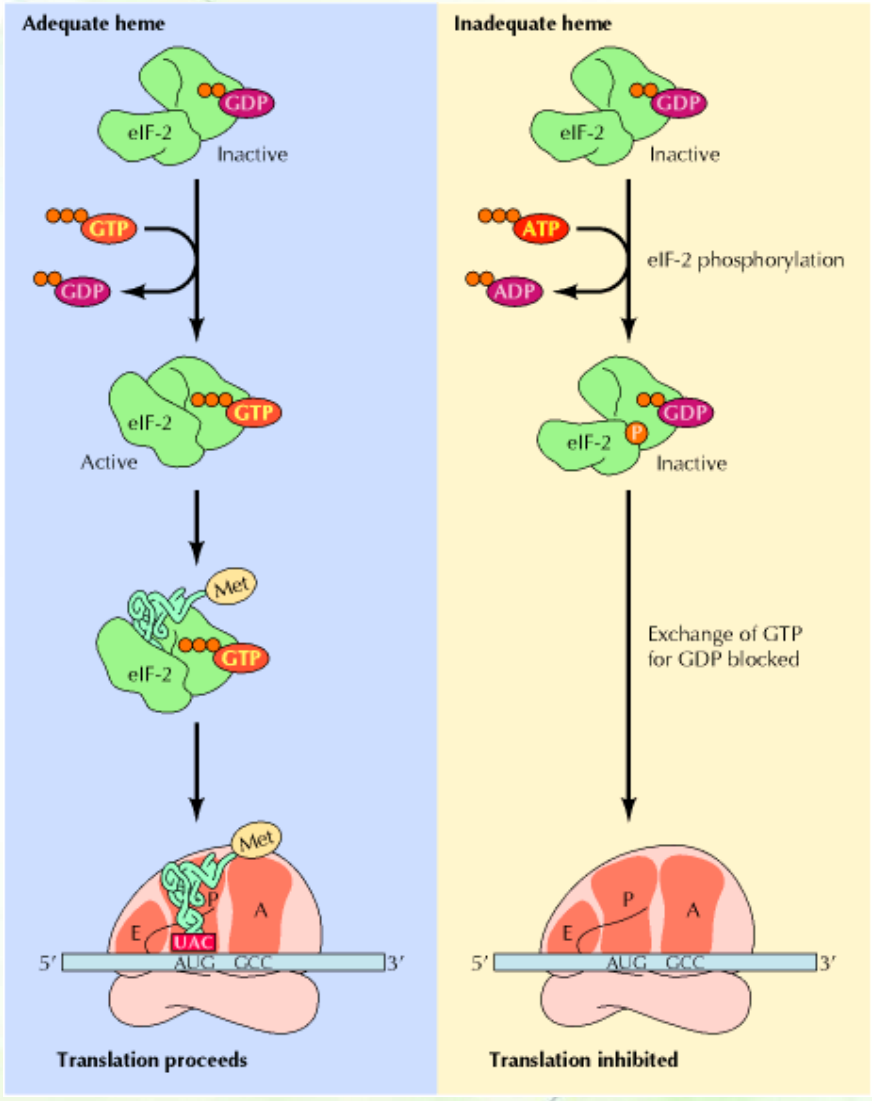


# Heme and protein synthesis

- In reticulocytes (immature erythrocytes), heme stimulates protein synthesis.
- The mRNA is translated only if adequate heme is available to form functional hemoglobin molecules.
- This is done via regulating the activity of eIF-2, which is responsible for escorting initiator methionyl tRNA to the ribosome.
- eIF-2 must be bound to GTP to be active. When it is released from the ribosome, GTP is hydrolyzed to GDP, which must be exchanged with GTP for eIF-2 to be active again.

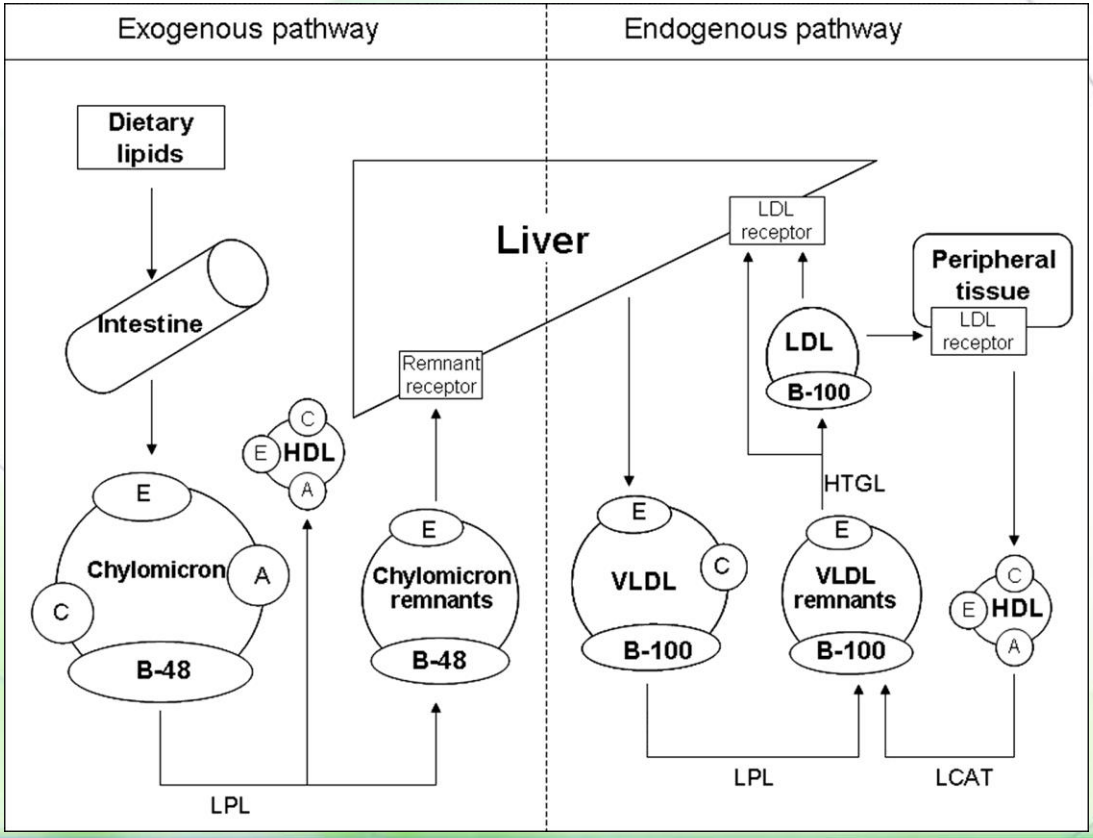
# Regulation

- If adequate heme is available, GDP-GTP exchange occurs and translation is able to proceed.
- If heme supplies are inadequate, a protein kinase that phosphorylates eIF-2 is activated. Phosphorylation of eIF-2 blocks the exchange of GTP for GDP, so eIF-2/GTP cannot be regenerated and translation is inhibited.

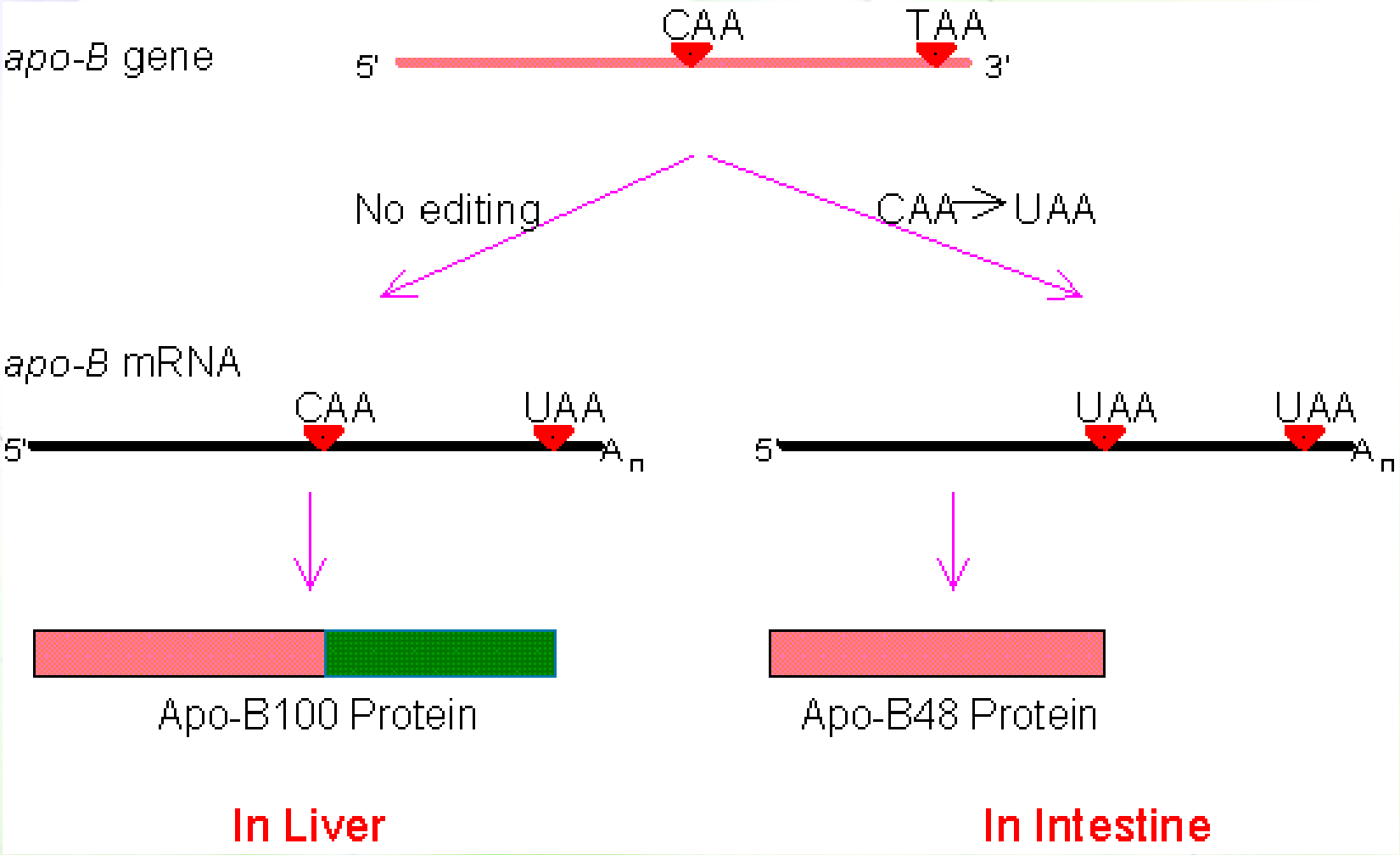


# ApoB-100 vs. apoB-48

- These proteins make up specific lipoprotiens that are responsible for lipid transport.
  - ApoB-100 is a liver proteins that is part of low-density lipoproteins
  - ApoB-48 is an intestinal proteins that is part of chylomicrons
- Both proteins are synthesized from the same gene.

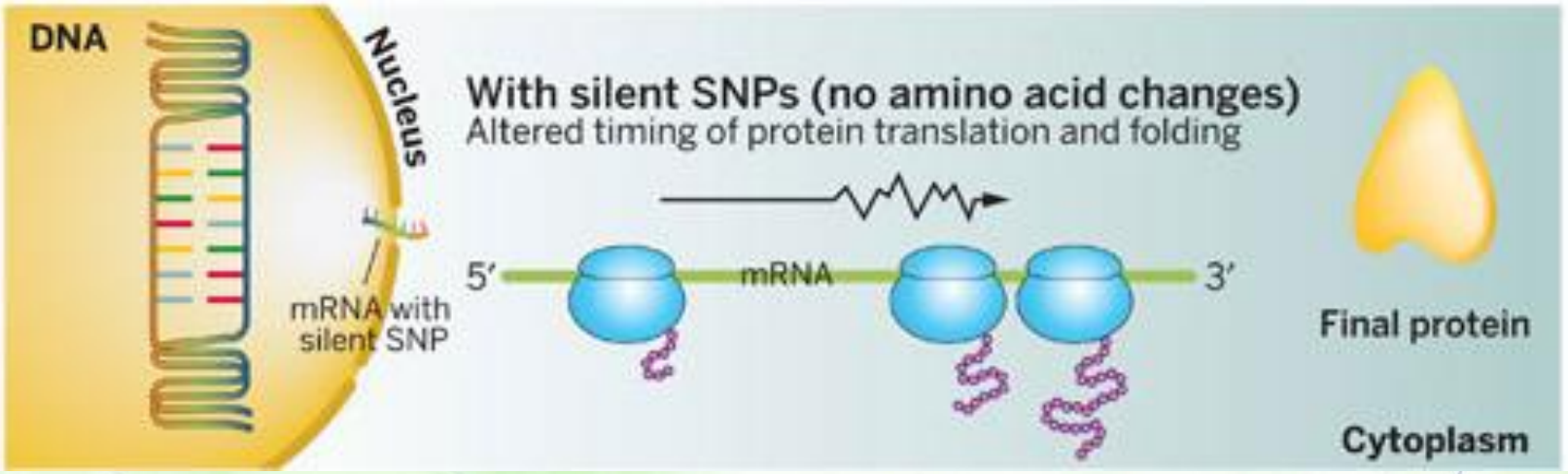


# Gene editing



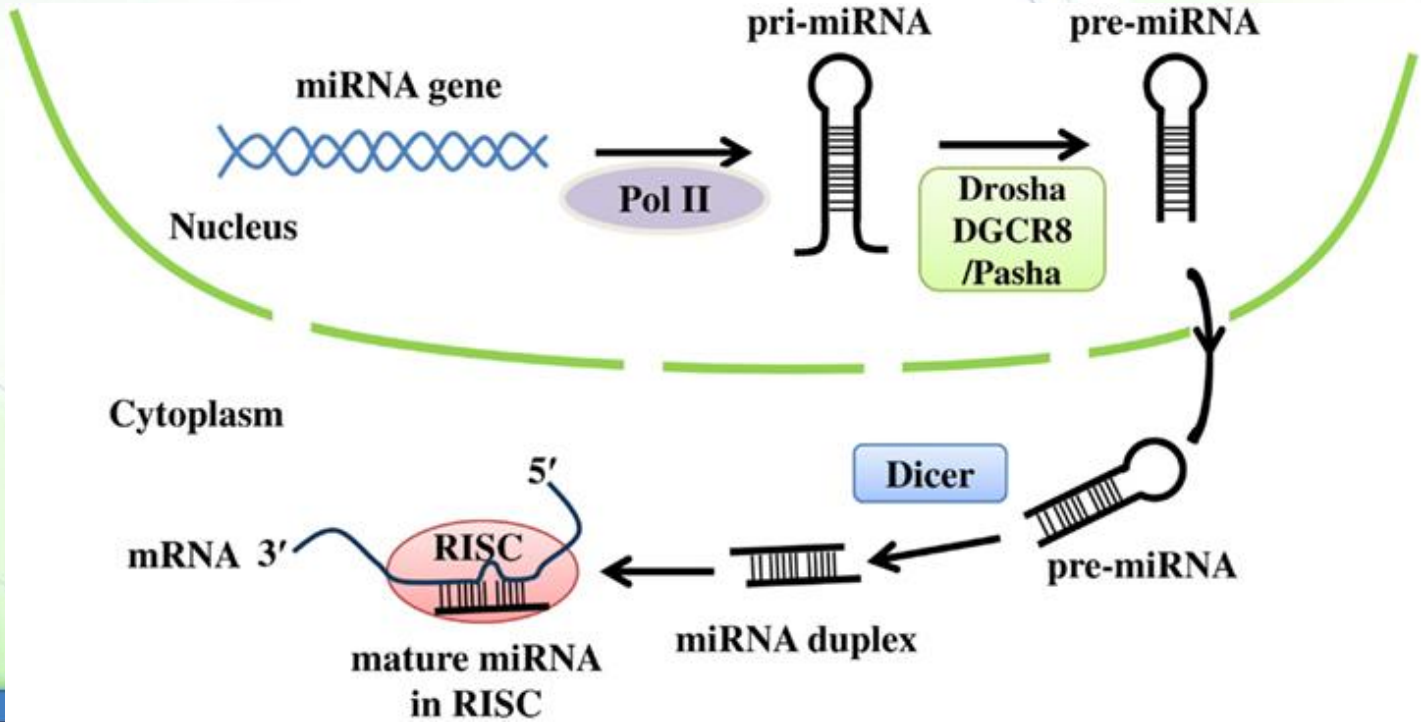
# Role of SNPs in protein synthesis kinetics

- As mRNA is translated at ribosomes into proteins, "silent" SNPs can lead to proteins with identical amino acid sequences but different shapes and activities. The shape-shifting occurs because the silent SNPs in the mRNA alter the timing and mechanical details by which the protein molecules fold.

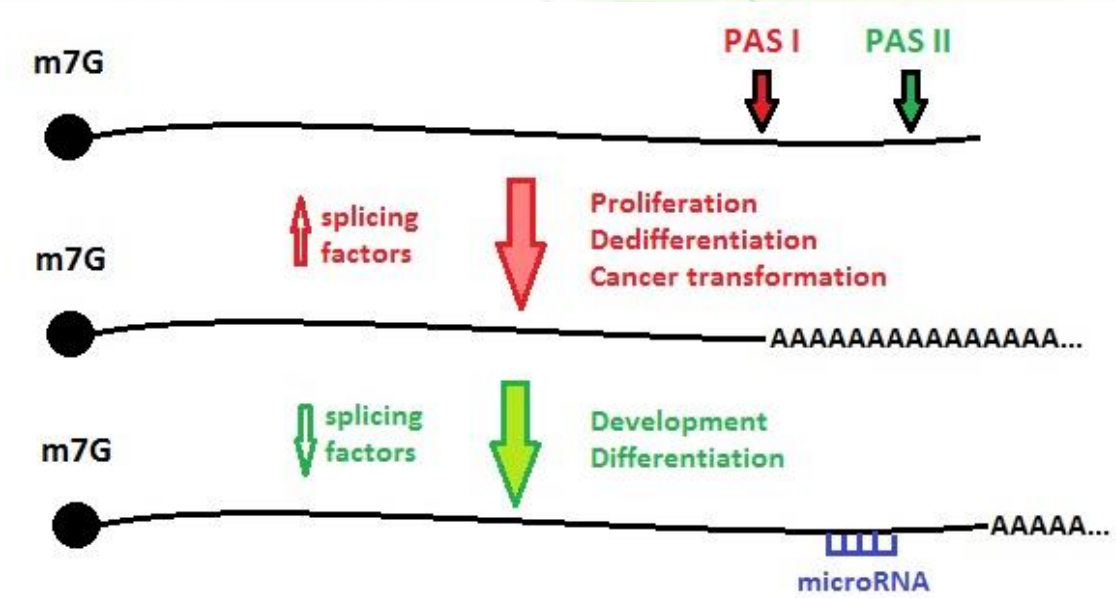


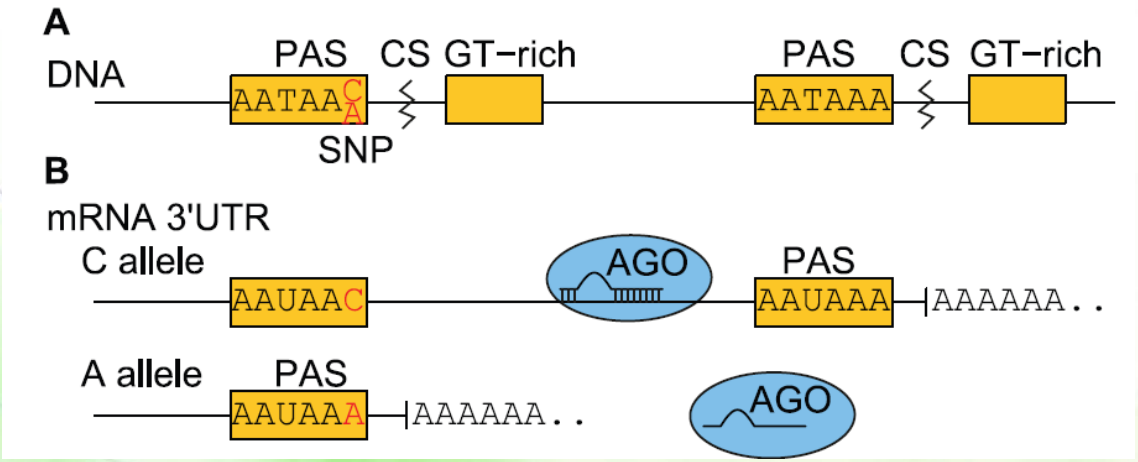
# Regulation by microRNA (miRNA)

- MicroRNA is synthesized by RNA Pol II into primary miRNA (pri-miRNA) transcript .
- Pri-miRNA is processed in the nucleus by Drosha and exported to the cytoplasm by Exportin 5, modified by an endonuclease complex containing Dicer to generate a mature miRNA duplex.
- One strand is loaded onto RISC complex where miRNA is targeted to mRNA resulting in either translation repression of mRNA degradation.



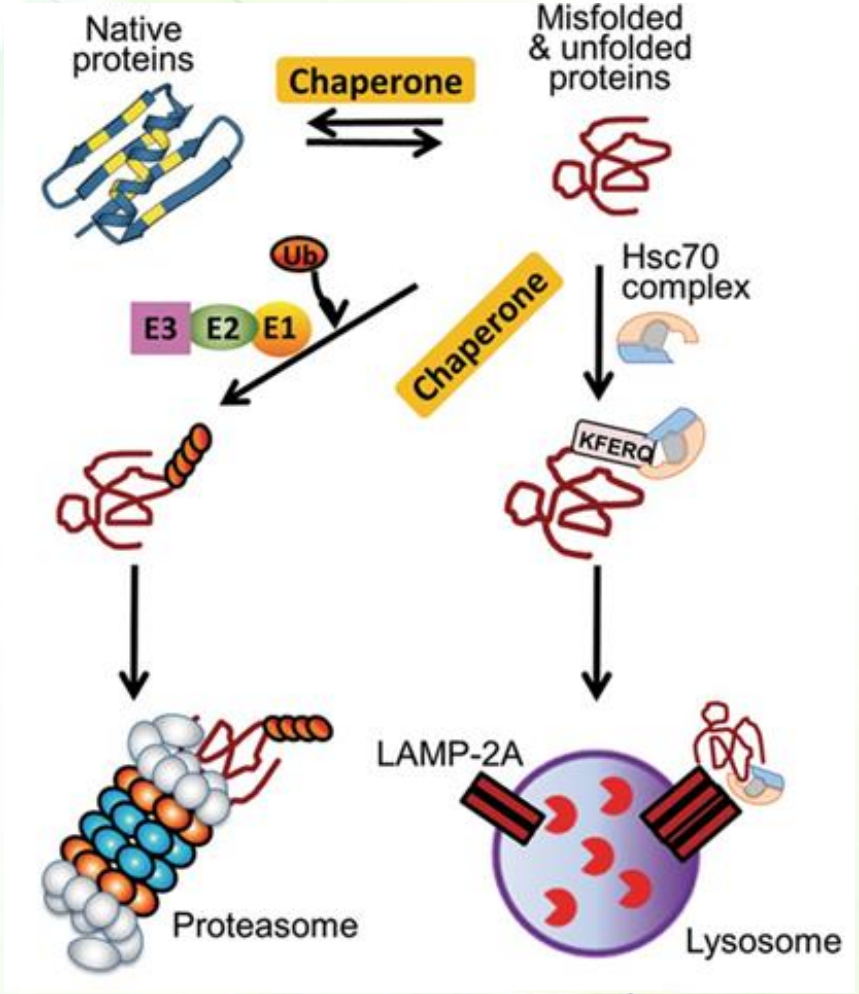
# Alternative polyadenylation



# Fate of (mis)- and (un)-folded proteins

- Proteins are degraded either in degradative subcellular organelles like lysosomes or by the macromolecular **proteasomes**.
- Proteins are targeted for destruction in a proteasome by **ubiquitinylation** which involves labeleing by small polypeptides known as ubiquitin.



# Levels of regulation

- Transcription
- RNA processing
- RNA transport
- mRNA stability
- Translation
- Post-translational modification
- Protein activity
- Protein degradation