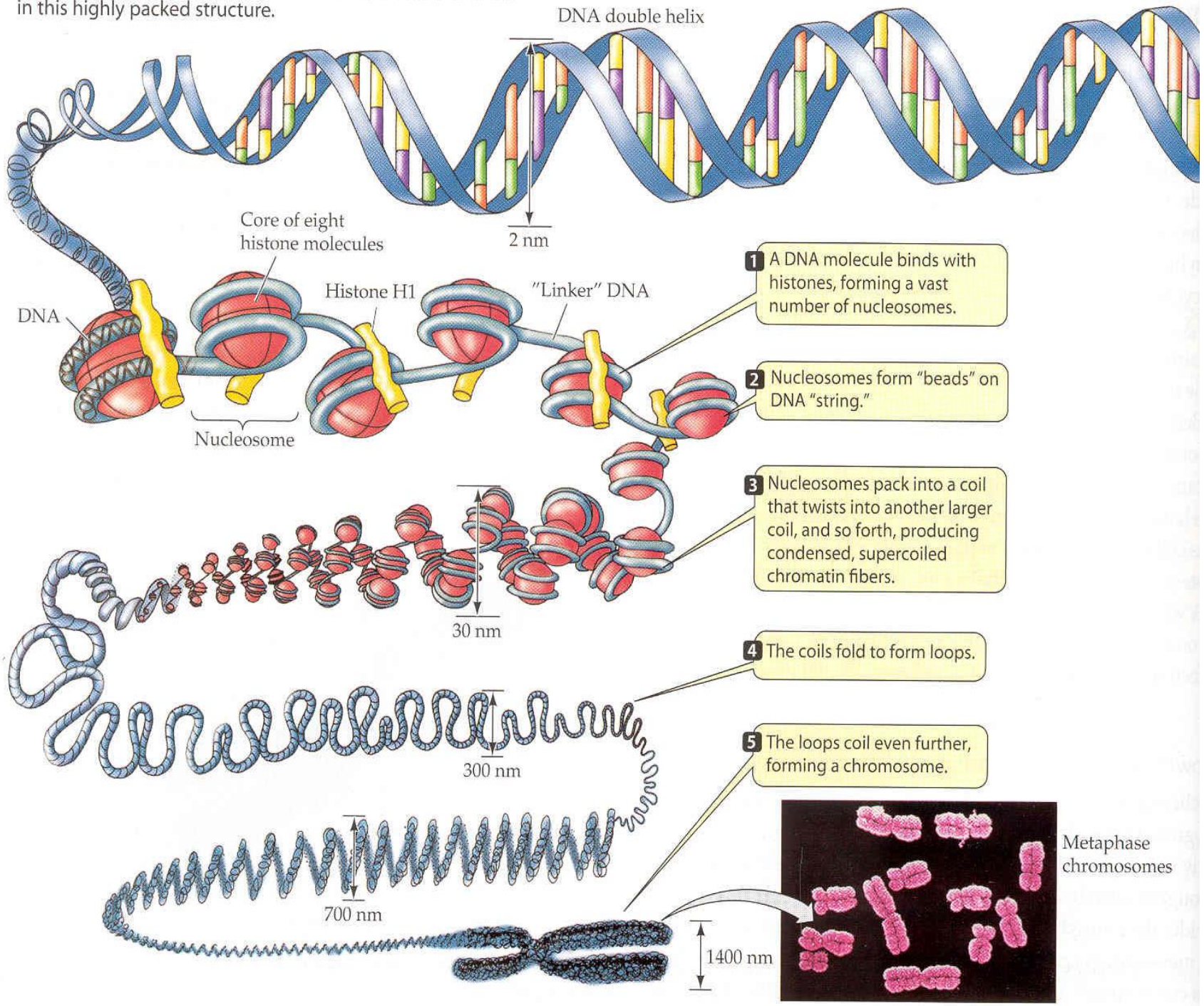
in this highly packed structure.



HISTORY

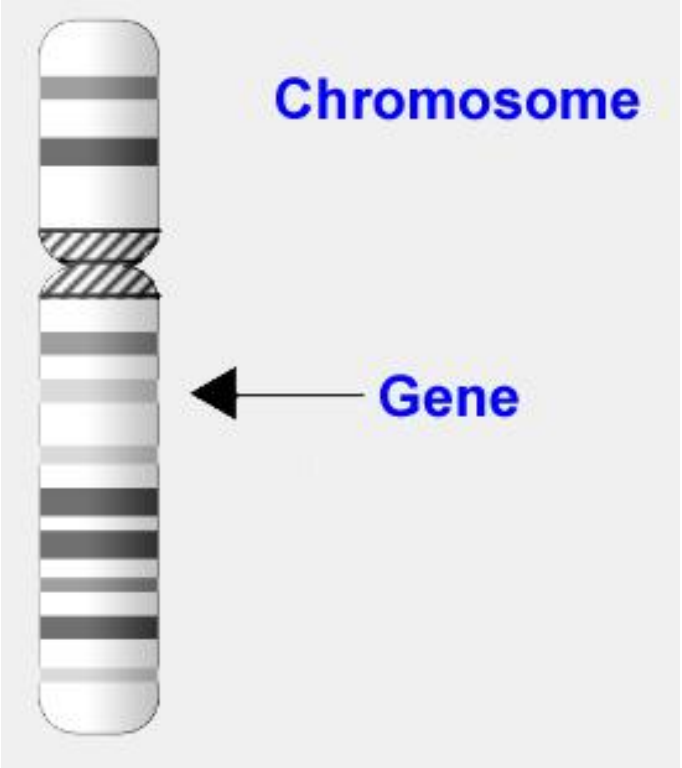
- Term GENE was introduced by JOHANSSEN in 1909 based on Mendelian Factors.
- Gene Concept was given by SUTTON.
- Studied & Elaborated by MORGAN, BRIDGES, and MULLER.

SUMMARY OF EVOLUTION OF GENE CONCEPT

YEAR	SCIENTIST	GENE CONCEPT
1866	G.J. MENDEL	A UNIT FACTOR THAT CONTROLS SPECIFIC PHENOTYPIC TRAIT
1902	SIR A.E.GARROD	ONE GENE –ONE METABOLIC BLOCK THEORY
1940	BEADLE & TATUM	ONE GENE-ONE ENZYME THEORY
1957	U.M.INGRAM	ONE GENE-ONE POLYPEPTIDE THEORY
1960s	C.YANOFSKY & CO-WORKERS	GENE IS A UNIT OF RECOMBINATION
Early 1970s	E.B.LEWIS	COMPLEMENTATION TEST IN *DROSOPHILA*

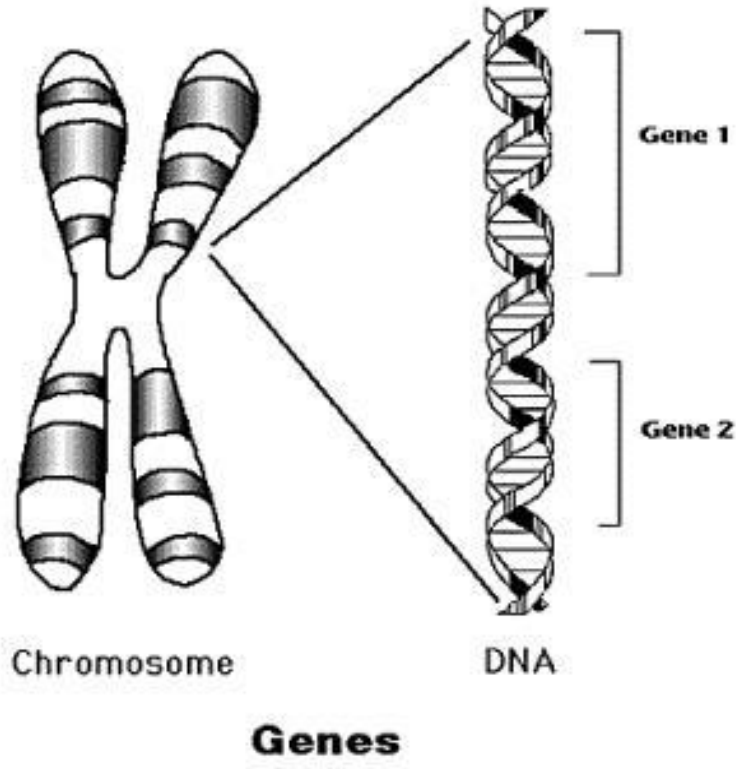
CLASSICAL DEFINITION OF GENE

- Gene is the Unit of Function (one gene specifies one character), Recombination, and Mutation.



MORDERN DEFINITION OF GENE

- Gene is the Unit of Genetic Information, i.e., the sequence of DNA that specifies one polypeptide.
- Includes coding as well as non-coding regulatory sequences.



Recently, the Sequence Ontology Consortium reportedly called the gene a **‘locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions.**

ESSENTIAL FEATURES

- Determines the physical as well as physiological characters.
- Situated in the chromosome.
- Occupies a specific position known as Locus.
- Arranged in single linear order.
- Occur in functional states called Alleles.
- Some have more than 2 alleles known as Multiple Alleles.

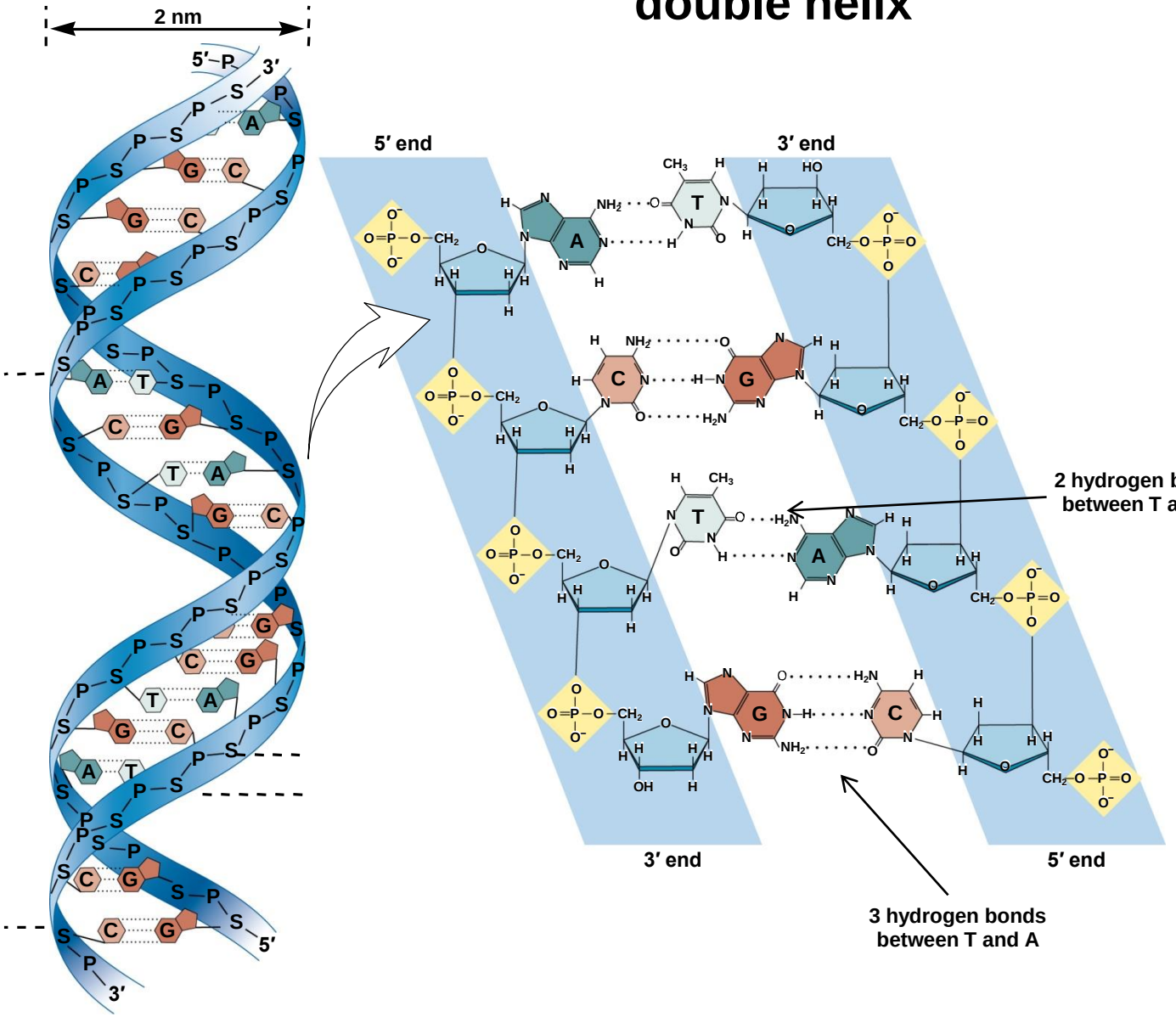
- Some may undergo sudden change in expression called as Mutant Gene (Mutation).
- May be transferred to its homologous (Cross-over) or non-homologous counterpart (Translocation).
- Can duplicate themselves very accurately (Replication).
- Synthesizes a particular Protein.
- Determines the sequence of amino acid in the polypeptide chain (The Genetic Code).

SOME TERMS RELATED TO GENE

BENZER has coined new terms to denote the relationship between DNA molecule and genetic phenomenon.

- RECON - It is the smallest unit of DNA capable of undergoing Crossing Over and Recombination.
- MUTON - It is the smallest unit of DNA which can undergo Mutation.
- CISTRON - It is the unit of Function. It is the Gene in real sense capable of synthesizing a Polypeptide chain of an Enzyme.
- COMPLON - It is the unit of Complementation.

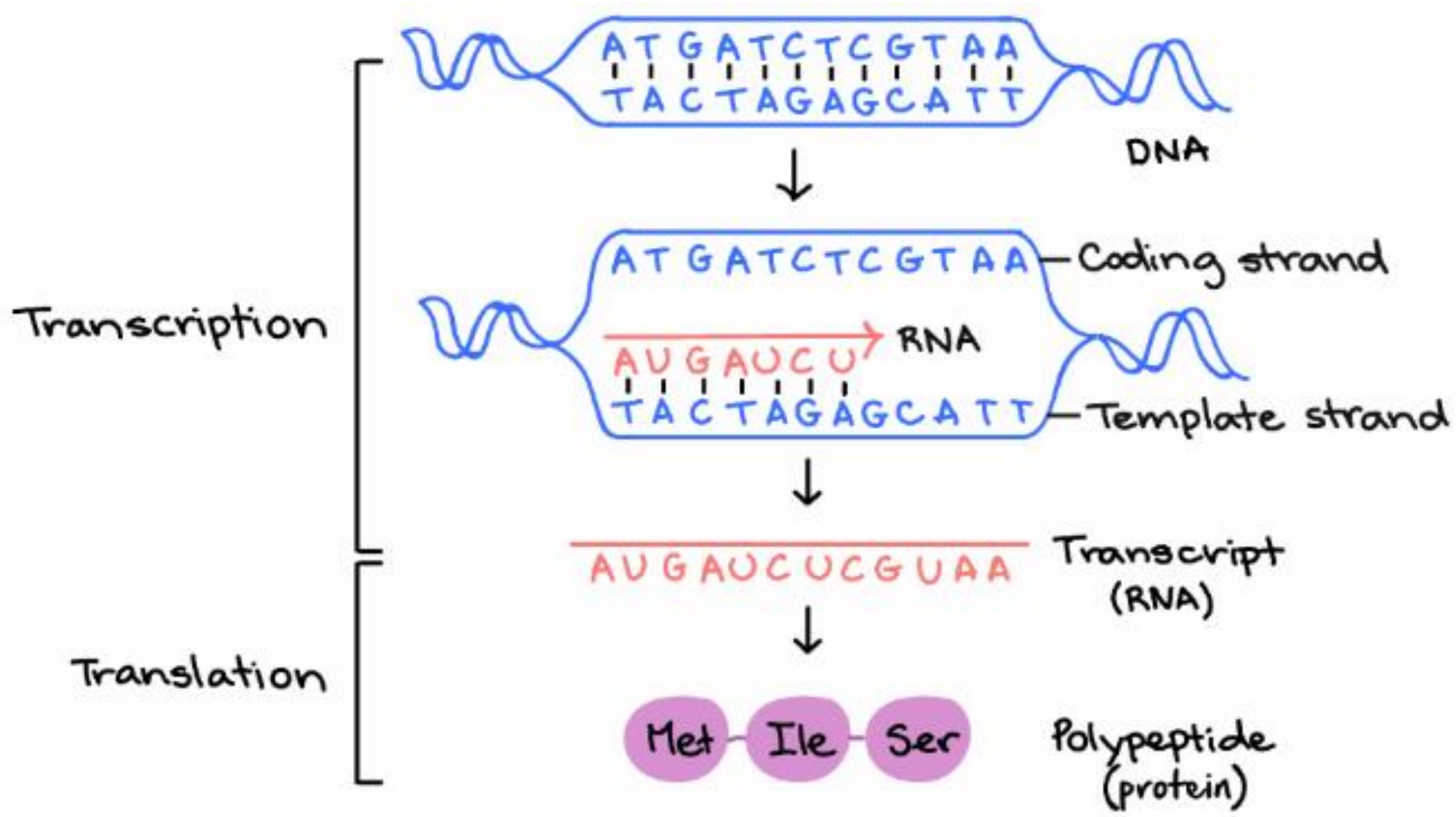
DNA exists in an anti-parallel double helix



Marked Ends of Polynucleotide Chain :

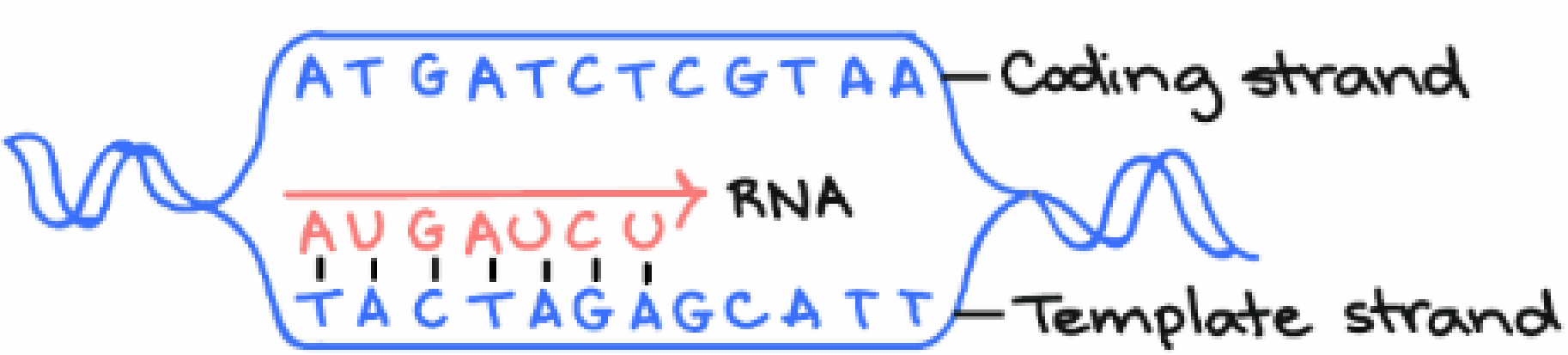
Each polynucleotide chain has marked ends. Its top end has a **sugar residue with free 5' carbon atom** which is not linked to another nucleotide. The **triphosphate is still attached to it.** This end is called the **5' end or 5' –P terminus.**

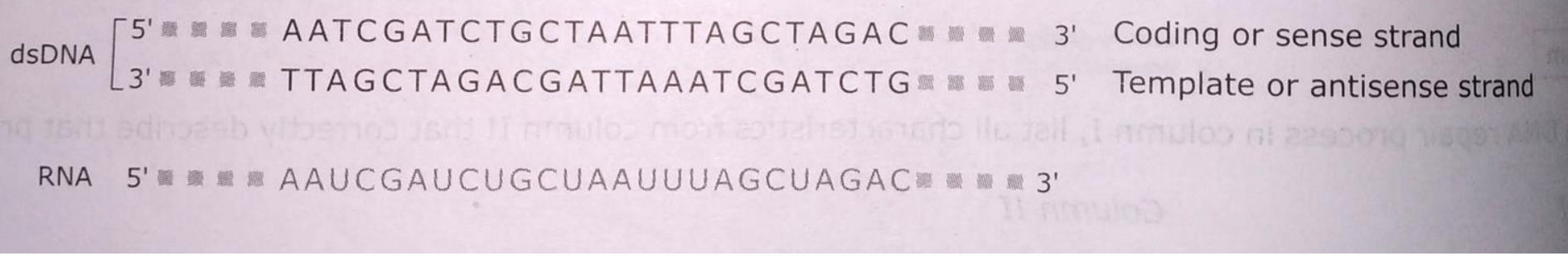
The other end of the chain ends in a **sugar residue with C -3 carbon** atom not linked . It bears **3' – OH group**. This end of polynucleotide chain is called **3' end or 3' –OH terminus**. It means the polynucleotide chains have direction and are marked as **3' – 5'.**

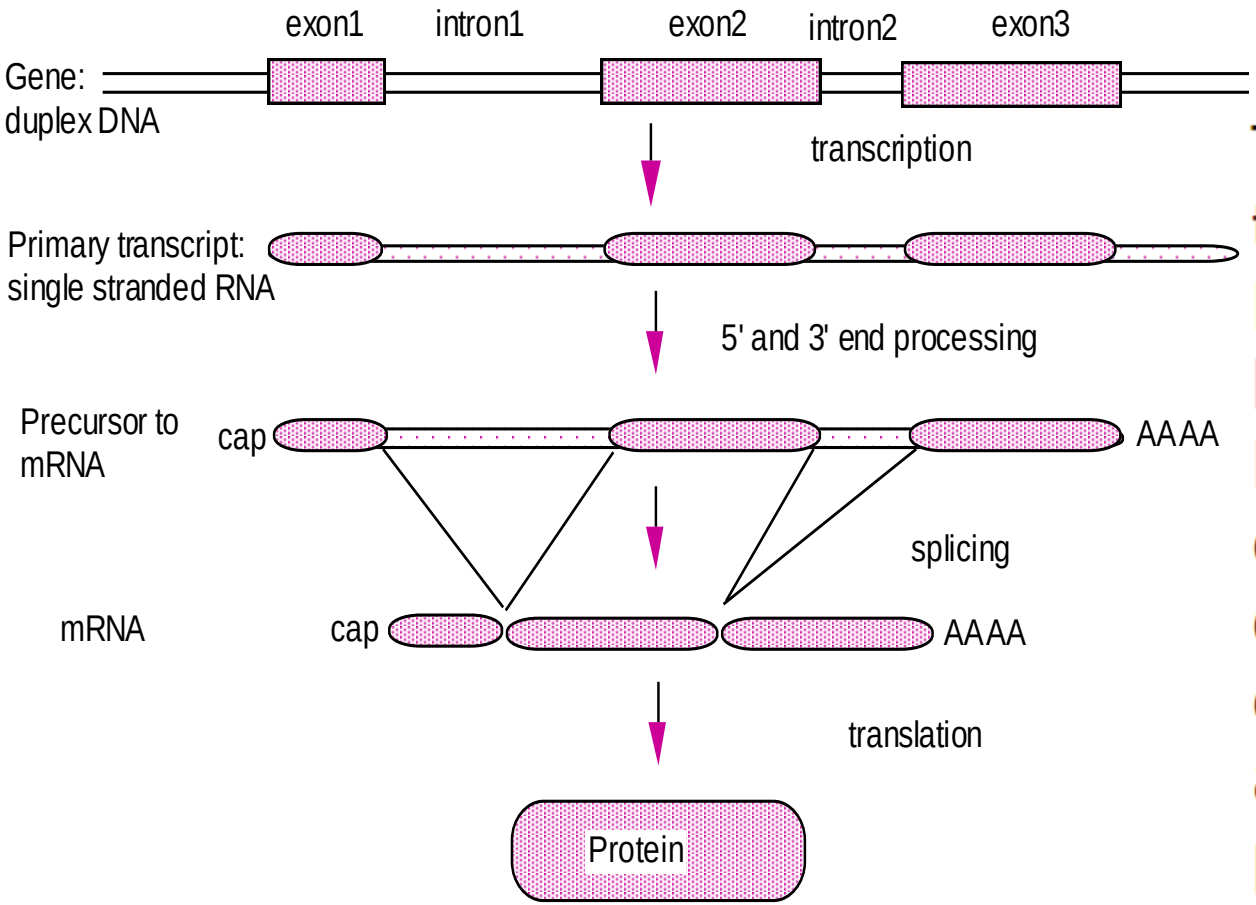

ATGATCTCGTAA
TACTAGAGCATT
DNA
ATGATCTCGTAA
Coding strand
RNA
AUGAUCU
TACTAGAGCATT
Template strand
Transcription
AUGAUCUCGUAA
Transcript (RNA)
Translation
Met
Ile
Ser
Polypeptide (protein)

Materials Required for the synthesis of RNA

- The enzyme RNA Polymerase.
- DNA template.
- All four types of ribonucleoside triphosphates (ATP, CTP, GTP, and UTP).
- Divalent metal ions Mg^{2+} or Mn^{2+} as a cofactor.




dsDNA
5' AATCGATCTGCTAATTTAGCTAGAC 3'
Coding or sense strand
3' TTAGCTAGACGATTAAATCGATCTG 5'
Template or antisense strand
RNA
5' AAUCGAUCUGCUAAUUUAGCUAGAC 3'

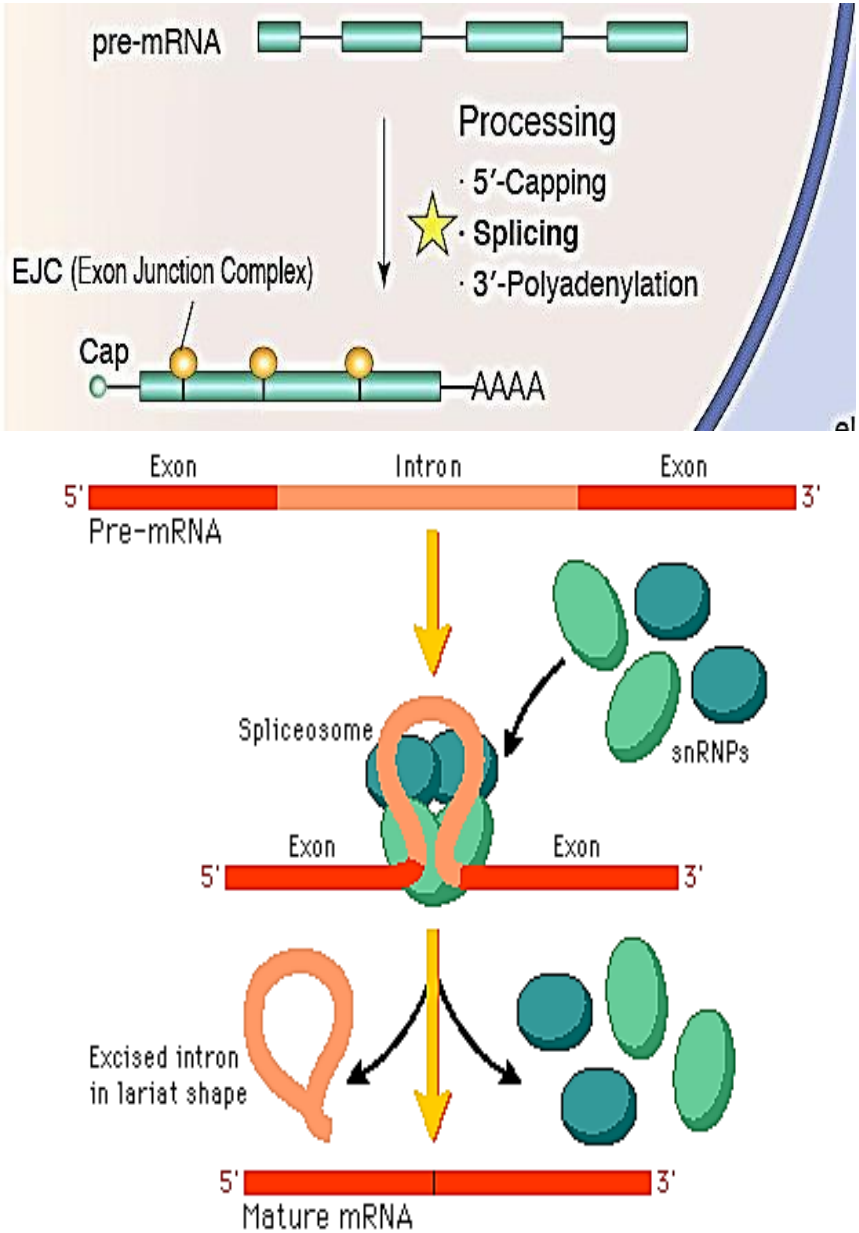


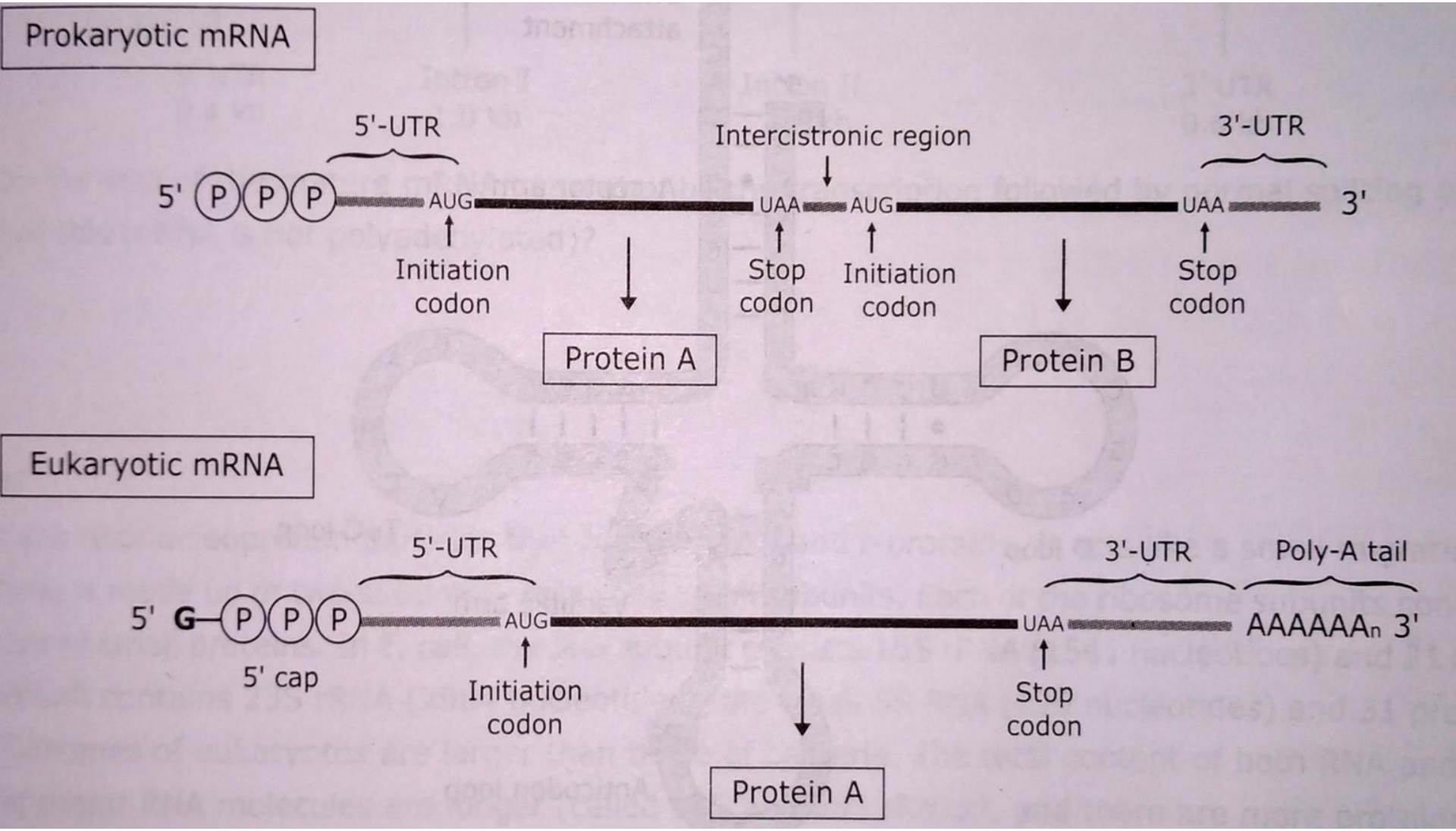
The original transcript from the DNA is called **heavy nuclear RNA (hnRNA)**. It contains transcripts of both introns and exons. The introns are removed by a process called **splicing** to produce **messenger RNA**

Regulation of Gene Action at Post-transcription level (in eukaryotes)

Splicing:

Splicing removes the introns, noncoding regions that are transcribed into RNA, in order to make the mRNA able to create proteins. Cells do this by spliceosomes (composed of small nuclear ribonucleoproteins, snPNPs) binding on either side of an intron, looping the intron into a circle and then cleaving it off. The two ends of the exons are then joined together.




Prokaryotic mRNA
5'-UTR
Intercistronic region
3'-UTR
5'
P
P
P
AUG
UAA
AUG
UAA
3'
Initiation codon
Stop codon
Initiation codon
Stop codon
Protein A
Protein B
Eukaryotic mRNA
5'-UTR
3'-UTR
Poly-A tail
5'
G
P
P
P
AUG
UAA
AAAAAAn 3'
5' cap
Initiation codon
Stop codon
Protein A

Human Genome Project

- The HGP was a collaborative international, government and Private **sponsored effort to map and sequence the entire human genome.**

- HGP was formulated by the **U.S. Department of Energy and NIH.**

- After years of multi-billion-dollar research, the **HGP and Celera Genomics** (a non-governmental biotechnology company) jointly announced drafts of the human genome sequence in 2000.

Salient Features of Human Genome

1. The **human genome** contains **3164.7 million nucleotide bases.**
2. The **average gene consists of 3000 base pairs** (gene codes for the protein **dystrophin** consists of **2.4 million base pairs).**
3. The total number of genes is about **30,000.**
4. The **size of chromosome** is in between about **55 Mb to 250 Mb.**
5. **Chromosome 1** has maximum **number of genes (2968),** and the **Y** has the fewest **(231).**
6. About **98.5% genome is non-coding** (only about **1.5%** is **the protein coding sequence)**

Table 1.15 Features of human genome

Size of DNA	3.2×10^9 bp
Number of genes	approximately 30,000
Largest gene	2.4×10^6 nucleotide pairs
Mean gene size	27,000 nucleotide pairs
Smallest number of exons per gene	1
Largest number of exons per gene	178
Largest exon size	17,106 nucleotide pairs
Mean exon size	145 nucleotide pairs
Percentage of DNA sequence in exons	~1.5%
Percentage of DNA in high-copy repetitive elements	approximately 50%
Average percentage of GC content	41%
Average number of genes per chromosome	1400

Organelle Genome

- Mitochondrial Genome – **mtDNA**
- Chloroplast Genome - **ctDNA**

Table 1.16 Features of human mitochondrial genomes

Size	16.6 kbp
Nature of DNA molecule	circular dsDNA molecule
Number of DNA molecules	More than one
Number of genes	37
Repetitive DNA	Very little
Introns	Absent
% of coding DNA	93%
Recombination	Not evident
Inheritance	Maternal

Numbers and Size of Genes

Table 1.12 Genome size and number of protein coding genes

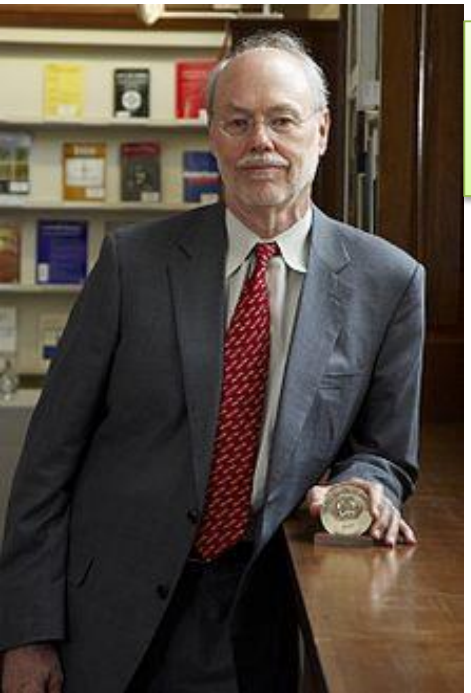
Species	*Genes*	*Genome size* (approx.)
M. genitalium	~470	0.58 Mb
H. influenzae	1727	1.8 Mb
E. coli	4288	4.6 Mb
S. cerevisiae	6275	12 Mb
D. melanogaster	~12000	123 Mb
C. elegans	~20000	100 Mb
Human	~30000	3300 Mb

✓ Size of genes in higher eukaryotes varies greatly. Some genes contain more than **2 million nucleotide pairs.** However, genes that are more than **100,000 nucleotide pairs in length are common.**

✓ Most of the portion of a gene in higher eukaryotes consists of **noncoding DNA that interrupts the relatively short segments of coding DNA.** **Discontinuous or Split or Mosaic Genes**

✓ The coding sequences are called exons, the intervening (noncoding) Sequences are called introns.

Phillip Allen Sharp is an American geneticist and molecular biologist who co-discovered **RNA splicing**. He shared the 1993 **Nobel Prize in Physiology or Medicine** with **Richard J. Roberts** for "the discovery that genes in eukaryotes are not contiguous strings but contain **introns,** and that the splicing of messenger RNA to delete those introns can occur in different ways, yielding different proteins from the same DNA sequence". He has been selected to receive the 2015 **Othmer Gold Medal.**



Discovery of introns in eukaryotic DNA and the mechanism of gene-splicing.

INTRONS

✓ An intron is any nucleotide sequence within a gene which is represented in the primary transcript of the gene, but not present in the final processed form.

✓ The overall length of a gene is determined largely by its introns.

✓ Introns range in size from about 50 nucleotides to > 100,000 nucleotides.

✓ Exons are usually short, typically on the order of 150 nucleotides.

Table 1.13 Average sizes of exons and introns in human genes

Gene product	*Size of gene (kb)*	*Number of exons*	*Average size of intron (bp)*
Insulin	1.4	3	480
β-Globin	1.6	3	490
Serum albumin	18	14	1100
CFTR (cystic fibrosis)	250	27	9100
Titin	283	363	466
Dystrophin	2400	79	30770

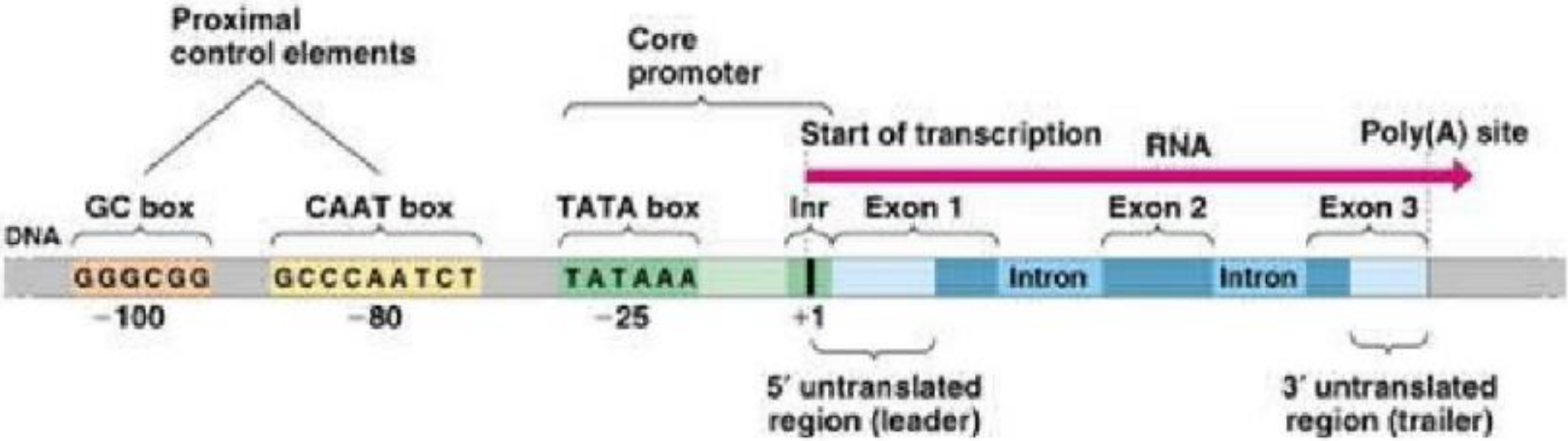
Table 1.14 Types of most common introns

Intron type	*Where found*
GU-AG introns	Eukaryotic nuclear pre-mRNA
AU-AC introns	Eukaryotic nuclear pre-mRNA
Group I introns	Eukaryotic nuclear pre-rRNA, organelle RNAs, some prokaryotic RNAs
Group II introns	Organelle RNAs, some prokaryotic RNAs
Pre-tRNA introns	Eukaryotic nuclear pre-tRNA
Archaeal introns	Various RNAs

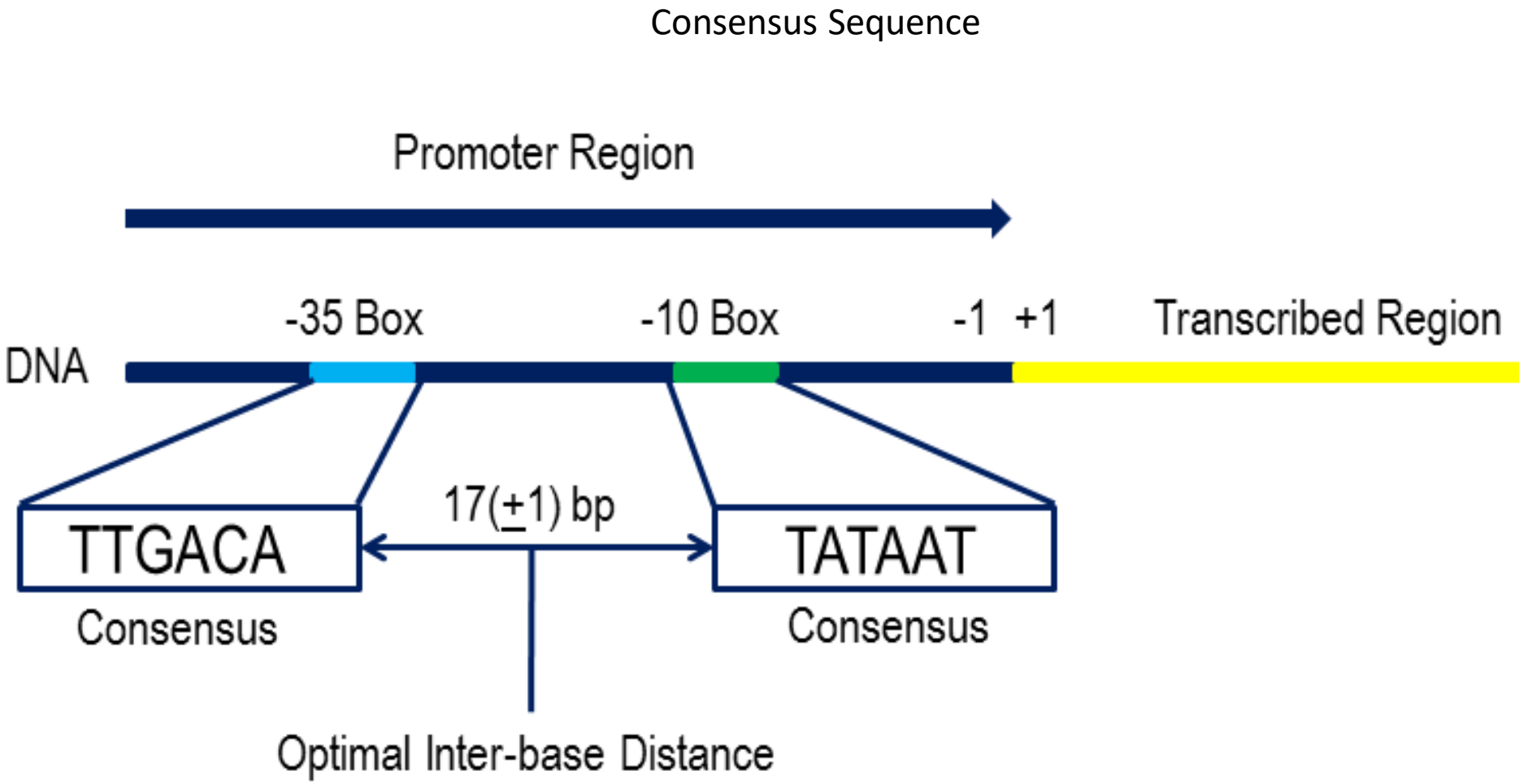
Source: Adapted from T.A.Brown, Genomes 3, Garland Science

Introns usually do not code for proteins. However, **certain introns of group I and II Classes contain open reading frames** that are translated into proteins.

Eukaryotic Gene Structure:



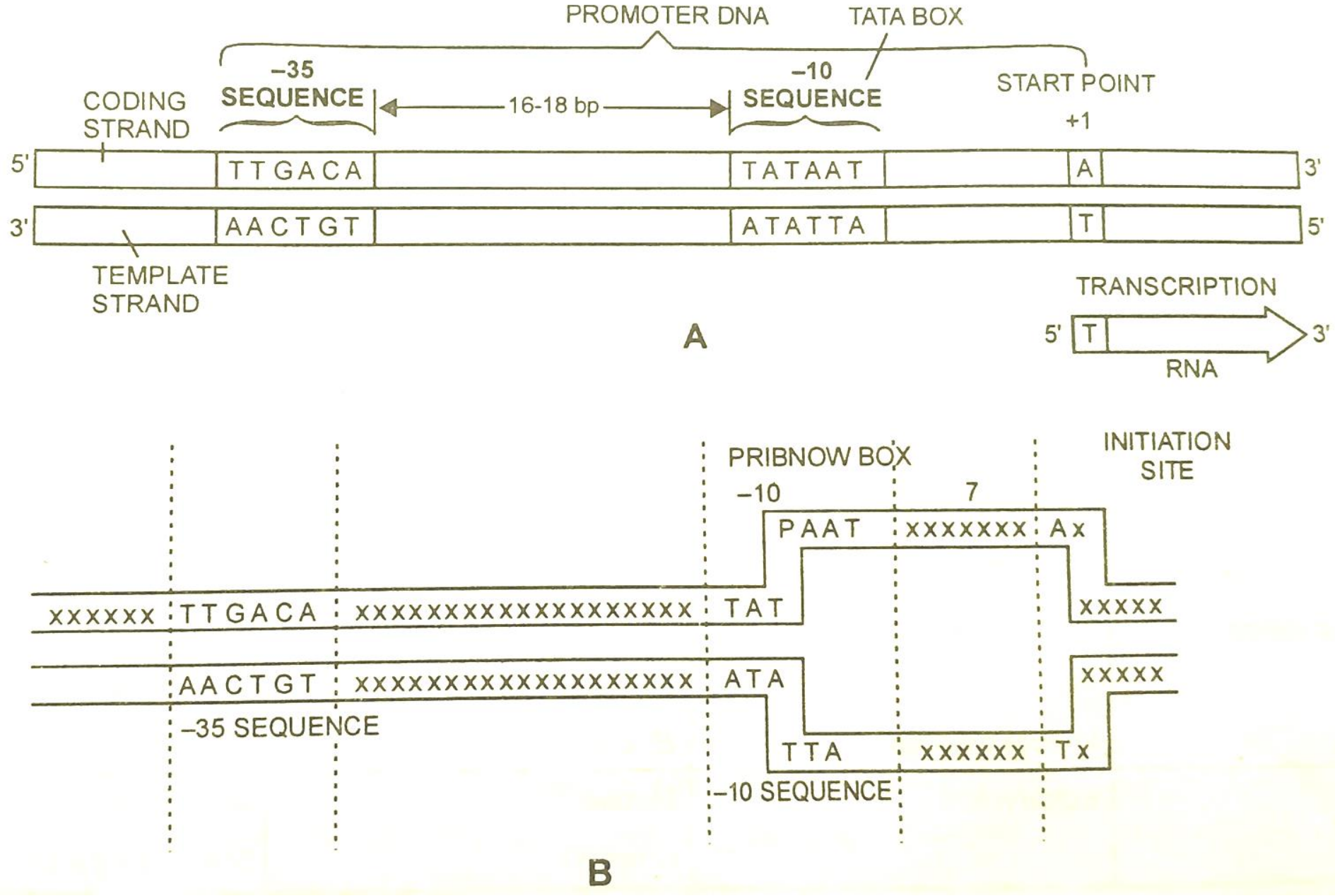
Functional Sequences in Promoter Region in Prokaryotes



Functional Sequences in Promoter Region in Prokaryotes

The promoter region in prokaryotes has two or three important functional sequences. These are **consensus sequences,** i.e., these sequences from promoters of different cistrons differ only by one or two bases.

- **Pribnow Box - 10 Region**
- **A Consensus Sequence in – 35 Region -**
- **Upstream Promoter Element**


PROMOTER DNA
TATA BOX
–35 SEQUENCE
–10 SEQUENCE
START POINT
+1
CODING STRAND
16-18 bp
5'
TTGACA
TATAAT
A
3'
3'
AACTGT
ATATTA
T
5'
TEMPLATE STRAND
TRANSCRIPTION
5'
T
3'
RNA
A
PRIBNOW BOX
INITIATION SITE
–10
7
PAAT
xxxxxxx
Ax
xxxxxx
TTGACA
xxxxxxxxxxxxxxxxxx
TAT
xxxxx
AACTGT
xxxxxxxxxxxxxxxxxx
ATA
xxxxx
–35 SEQUENCE
TTA
xxxxxx
Tx
–10 SEQUENCE
B

Pribnow box sequence orients RNA Polymerase on promoter and helps it in locating the precise bases at which transcription will begin.

A Consensus Sequence in – 35 Region - This hexanucleotide sequence is recognized by sigma factor of RNA Polymerase . It is presumed that **sigma subunit of RNA Polymerase binds to – 35 sequence.**

Upstream Promoter Element – AT rich recognition element in the promoter region. It is present in the promoters of certain highly expressed genes and is located between **– 40 and – 60.** this sequence is **bound by the alpha subunit of RNA Polymerase**.

Functional Sequences in Promoter Region in Eukaryotes

Sequences of nitrogenous bases similar to that found in Prokaryotes are present in the promoter region of eukaryotes also.

TATA Box or Goldberg Hognes Box

Hexanucleotide sequence consisting of **(5') TATAAA (3')** it is usually represented by **TATXAX,** where **X** may be **T or A.** It is the site of assembly of a **preinitiation complex (transcription factors and Polymerase which is essential for initiation of transcription.**

In addition eukaryotic genome has **proximal control elements**, named as **GC box** with base sequence **GGGCGG at – 100 site** and a **CAAT box** with base sequence **GCCCAATCT.**

TRANSCRIPTION IN PROKARYOTIC CELL

In prokaryotes, single RNA *polymerase* catalyses synthesis of all three types of RNAs. In *E. coli*, ***RNA polymerase*** enzyme is a large protein with a molecular weight of 4,80,000. It is a holoenzyme and consists of following parts :

1. Core Enzyme : It is formed of five tightly associated protein chains namely β, β′, α, α and ω with molecular weights of 1,60,000, 1,50,000, 90,000, 4,000 and 1,000.

2. Sigma Factor (σ) : It is formed of a single protein chain, loosely attached to core enzyme. It helps in the recognition of start signal on DNA molecule and directs RNA *polymerase* enzyme to bind.

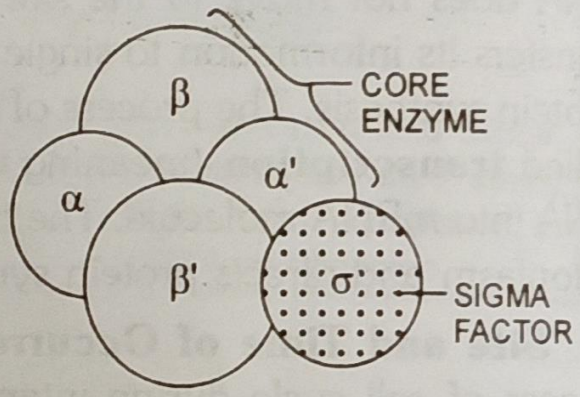


Fig. 15. RNA polymerase enzyme of bacterium *E. coli* showing association of its 5 polypeptides ααββ′ and σ.

The core enzyme without sigma factor can bind to any region in the double stranded DNA and can transcribe DNA. But it does not recognise specific promoters on DNA template. It is unable to discriminate between the two strands of DNA helix to be used as a template. **Sigma factor** recognises specific transcription initiation sites, where RNA synthesis can begin. These sites are called **promoter sites.** Thus, there are different sigma factors, one each for every promoter site.

Function : RNA *polymerase* basically catalyses the formation of phosphodiester bonds between successive nucleotides of a polynucleotide chain during synthesis of both DNA and RNA. With sigma factor, it contributes to the synthesis of *m*RNA, because sigma factor recognises promoter site.

RNA polymerases lack proof reading 3′ → 5′ *exonuclease* activity. Therefore, one error for every 10^4 to 10^5 ribonucleotides incorporated is introduced during RNA transcription. But the mistake during RNA transcription is not serious because of its high turnover.

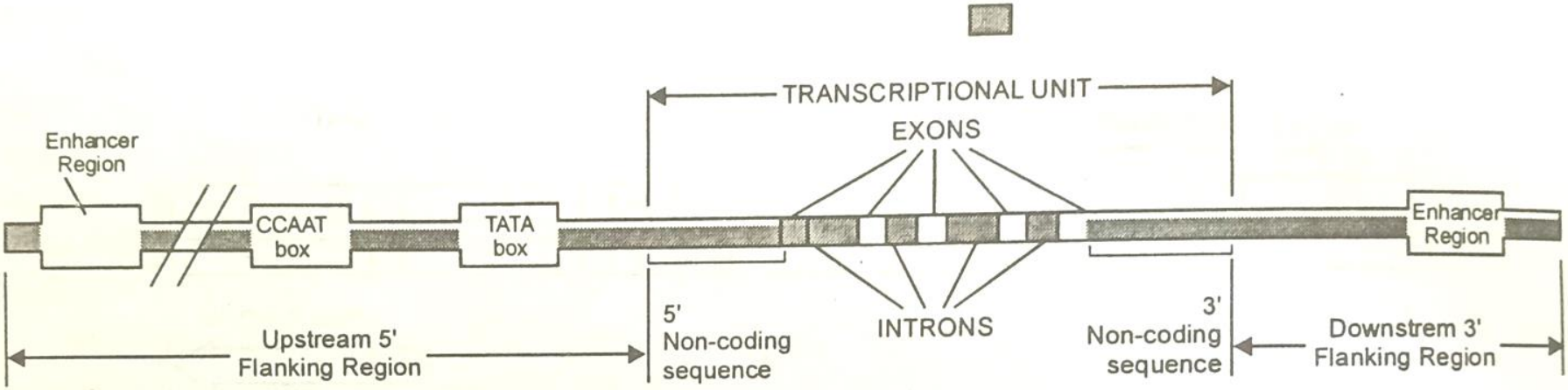


Fig. 18. A eukaryotic gene with its promoter region.

Table 4. Similarity between Pribnow and TATA box sequences

Type of Sequence	General Group	Organism	Sequence
Pribnow Sequence	In Viruses	$\phi \times 174$	5′- - - TACAGTA - - - 3′
		SV40	5′ - - - TATAATG - - - 3′
	In Prokaryotes	*E. coli*	5′ - - - TATAATG - - - 3′
TATA Box	Eukaryotes	Mouse	5′ - - - TATAAAG - - - 3′
		Chicken	5′ - - - TATATAT - - - 3′