STSs and ESTs

- Sequence-Tagged Site: short, unique sequence
- Expressed Sequence Tag: short, unique sequence from a coding region
  - 1991: 609 ESTs [Adams et al.]
  - June 2000: 4.6 million in dbEST
  - Genome sequencing center at St. Louis produce 20,000 ESTs per week.

What Are ESTs and How Are They Made?

- Small pieces of DNA sequence (usually 200 - 500 nucleotides) of low quality.
- Extract mRNA from cells, tissues, or organs and sequence either end. Reverse transcribe to get cDNA (5' EST and 3'EST) and deposit in EST library.
- Used as "**tags**" or markers for that gene.
- Can be used to identify similar genes from other organisms (Complications: variations among organisms, variations in genome size, presence or absence of **introns**).
- 5' ESTs tend to be more useful (cross-species conservation), 3' EST often in UTR.

DNA Markers

- Uniquely identifiable DNA segments.
- Short, <500 nucleotides.
- Layout of these markers give a map of genome.
- Markers may be polymorphic (variations among individuals). Polymorphism gives rise to alleles.
- Found by PCR assays.

Polymorphisms

- Length polymorphisms
  - Variable # of tandem repeats (VNTR)
  - Microsatellites or short tandem repeats
  - Restriction fragment length polymorphism (RFLP) caused by changes in restriction sites.
- Single nucleotide polymorphism (SNP)
  - Average once every ~100 bases in humans
  - Usually biallelic
  - dbSNP database of SNPs (over 100,000 SNPs)
  - ESTs are a good source of SNPs

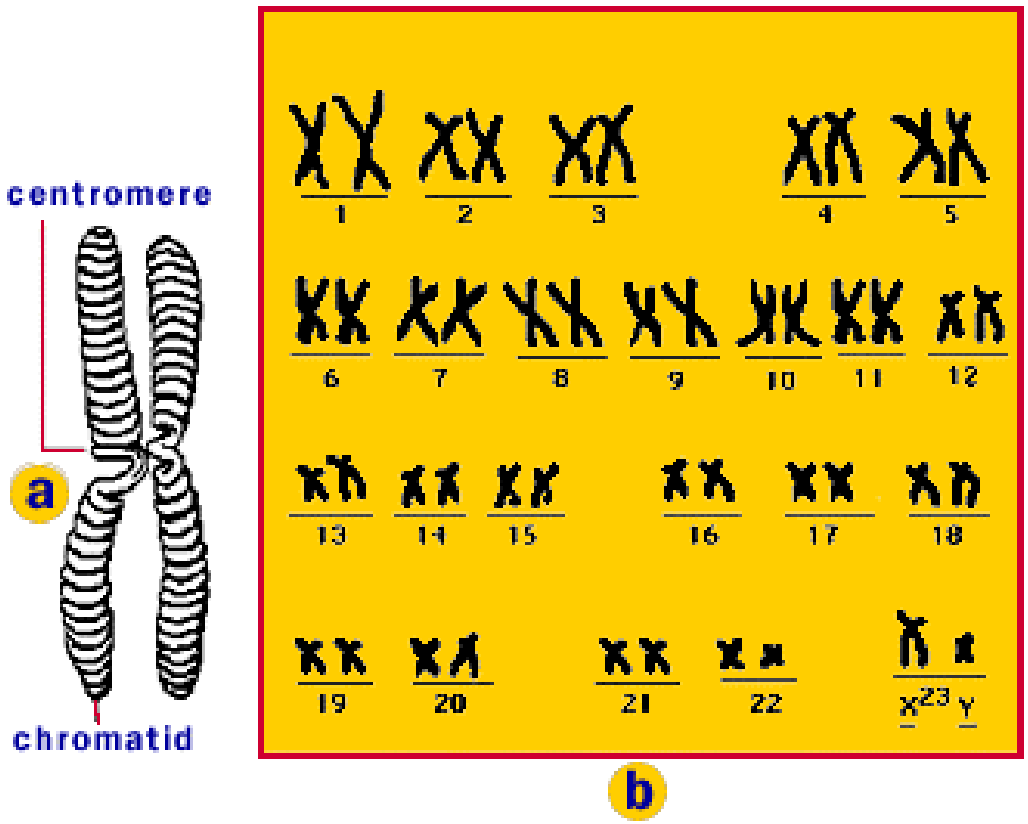
SNPs

- SNPs often act as “disease markers”, and provide “genetic predisposition”.
- SNPs may explain differences in drug response of individuals.
- Association study: study SNP patterns in diseased individuals and compare against SNP patterns in normal individuals.
- Many diseases associated with SNP profile.

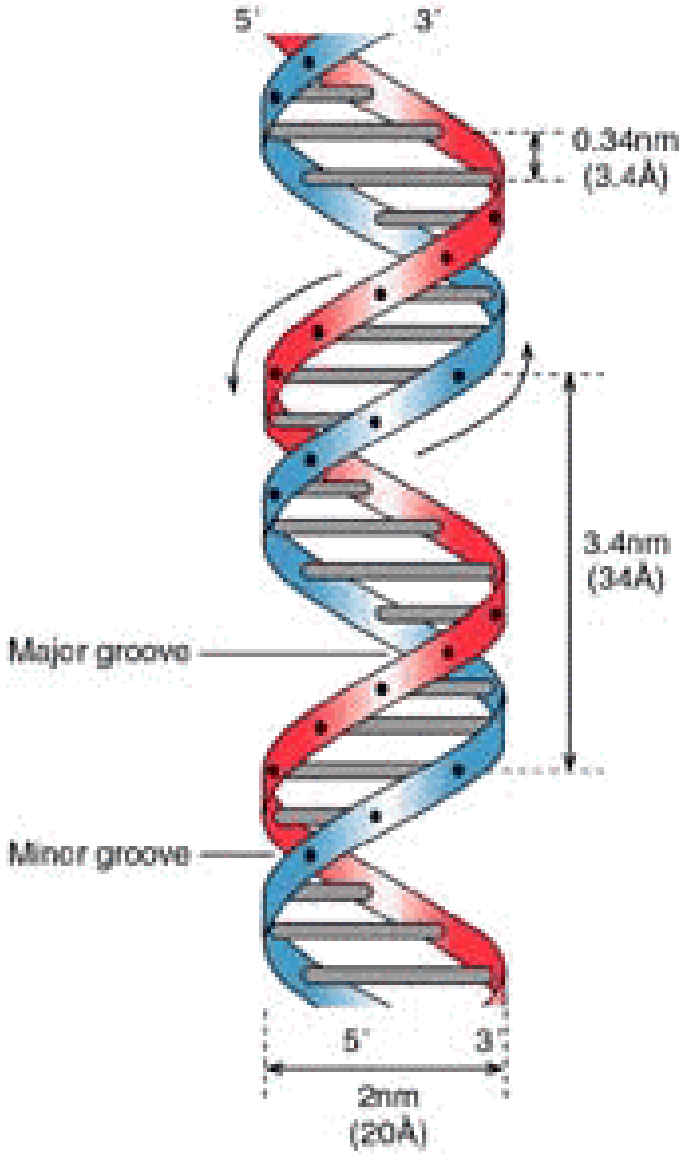
Review of Molecular Biology Background

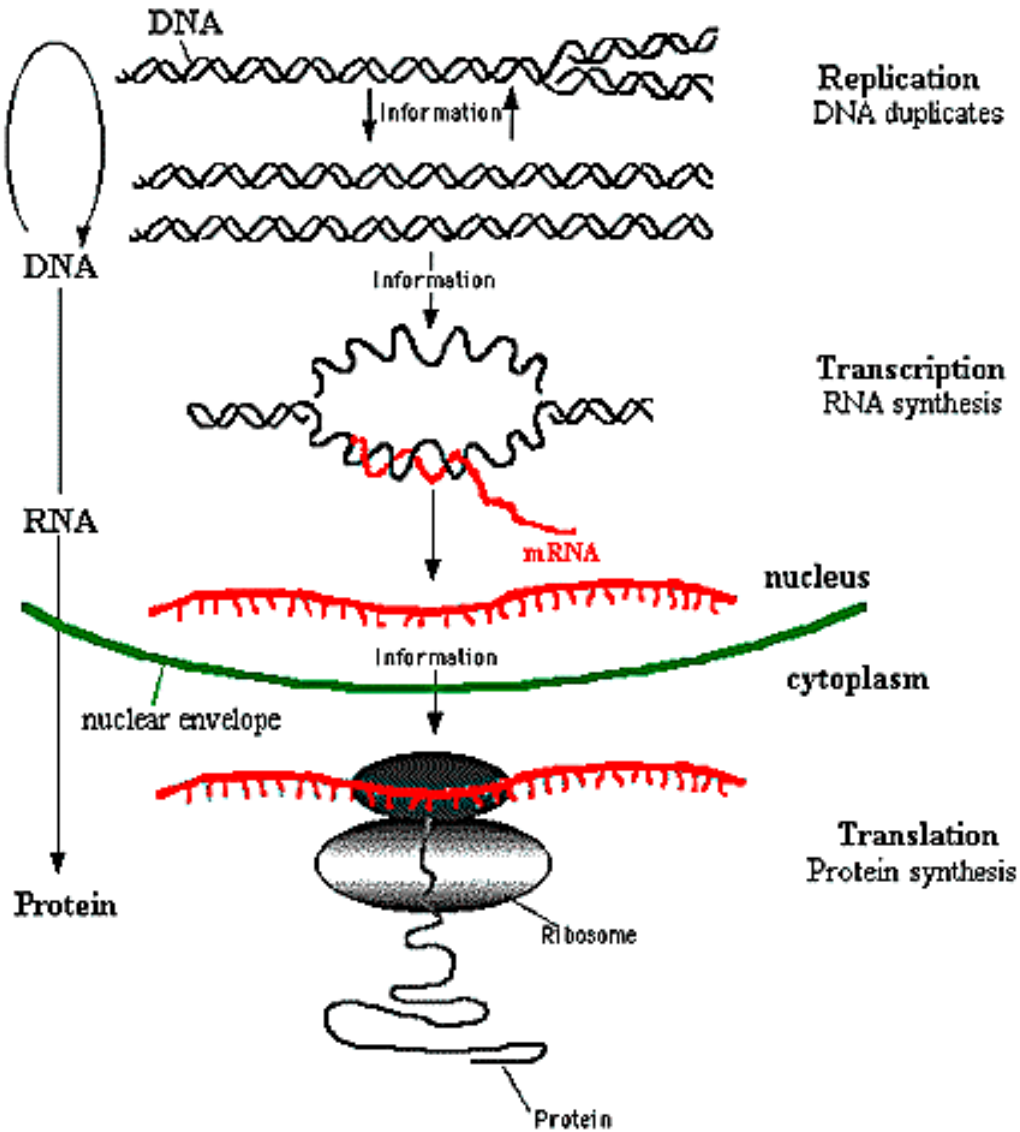
Chromosomes

Human chromosomes!

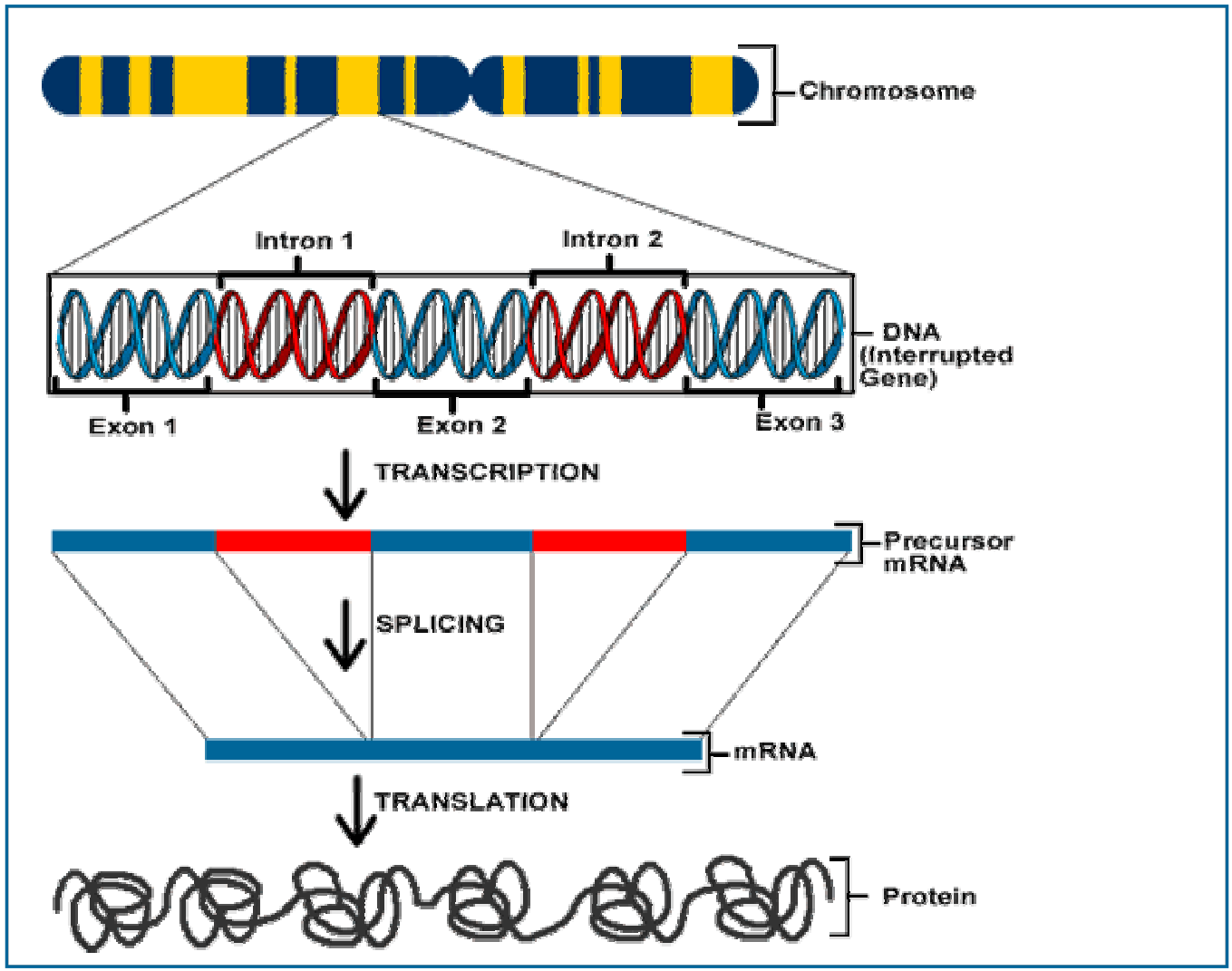


DNA Molecule

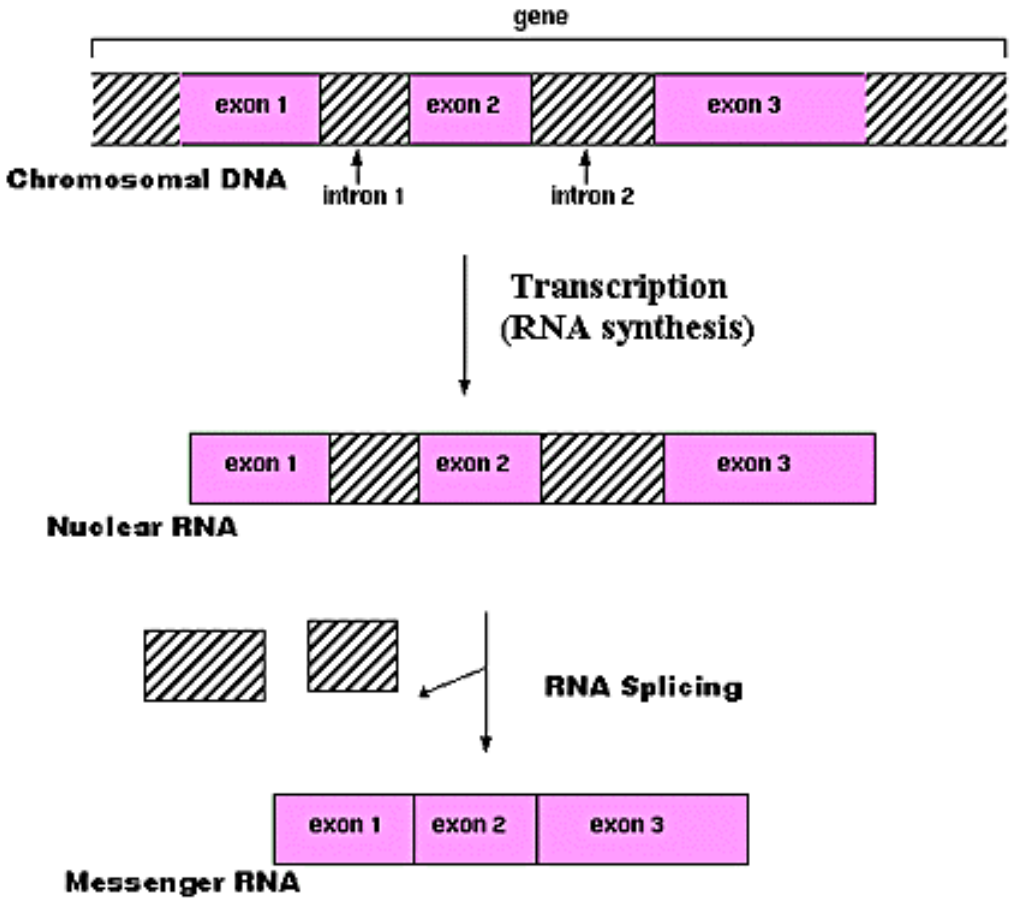




The Central Dogma of Molecular Biology

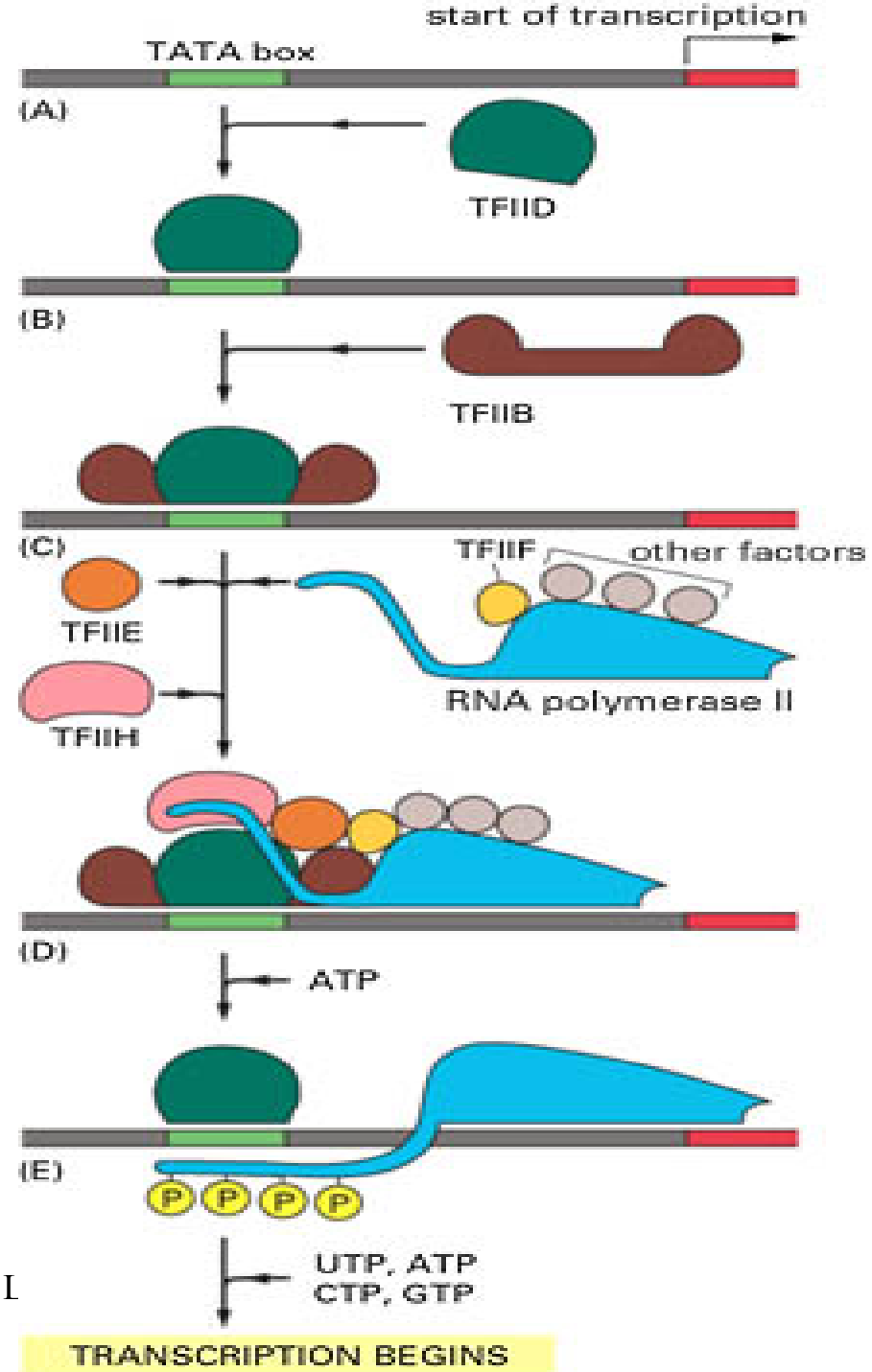

Chromosome
Intron 1
Intron 2
DNA
(Interrupted
Gene)
Exon 1
Exon 2
Exon 3
TRANSCRIPTION
Precursor
mRNA
SPLICING
mRNA
TRANSLATION
Protein

DNA Transcription



RNA synthesis and processing

Transcription Initiation



Transcription

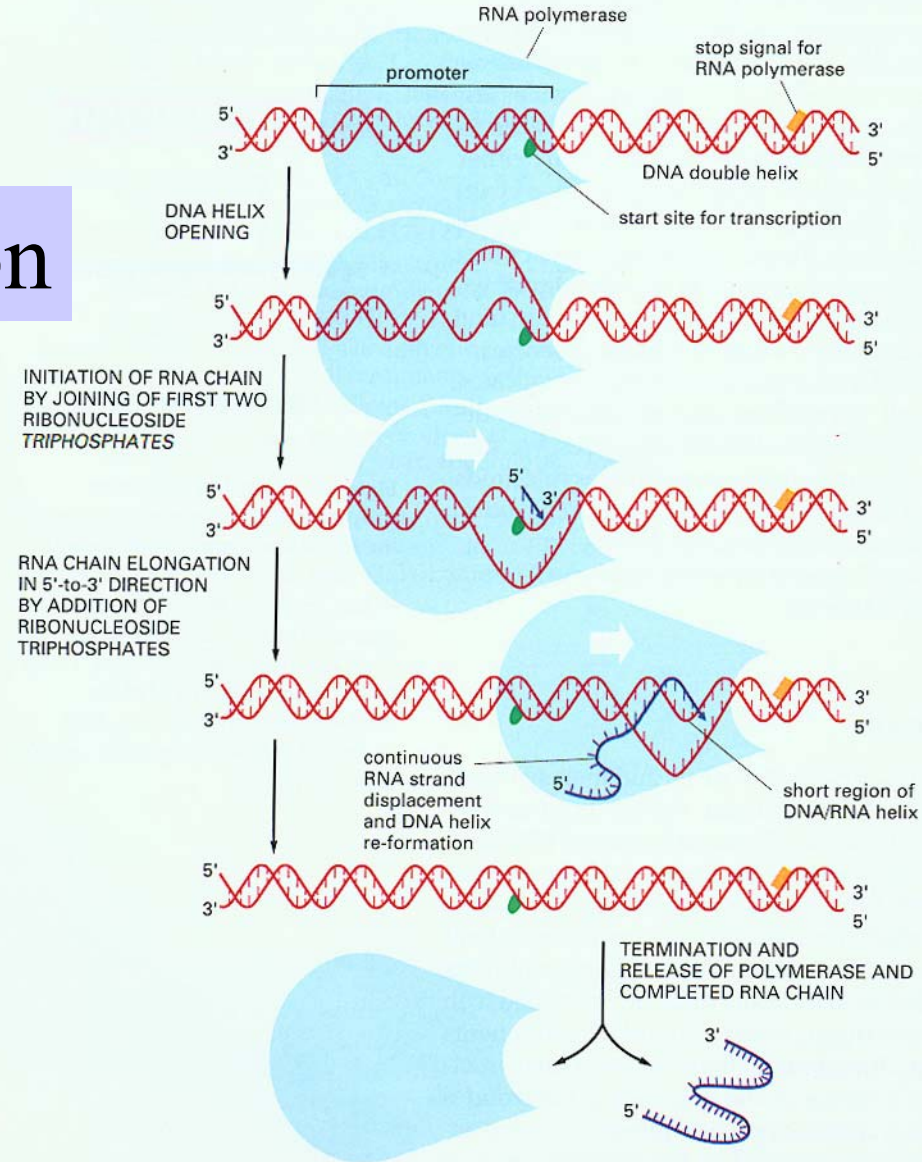


Figure 6–2 The synthesis of an RNA molecule by RNA polymerase. The enzyme binds to the promoter sequence on the DNA and begins its synthesis at a start site within the promoter. It completes its synthesis at a stop (termination) signal, whereupon both the polymerase and its completed RNA chain are released. During RNA chain elongation, polymerization rates average about 30 nucleotides per second at 37°C. Therefore, an RNA chain of 5000 nucleotides takes about 3 minutes to complete.

224 Chapter 6 : Basic Genetic Mechanisms

Transcription Steps

RNA polymerase needs many transcription factors (TFIIA,TFIIB, etc.)

(A) The promoter sequence (TATA box) is located 25 nucleotides away from transcription initiation site.

(B) The TATA box is recognized and bound by transcription factor TFIID, which then enables the adjacent binding of TFIIB. DNA is somewhat distorted in the process.

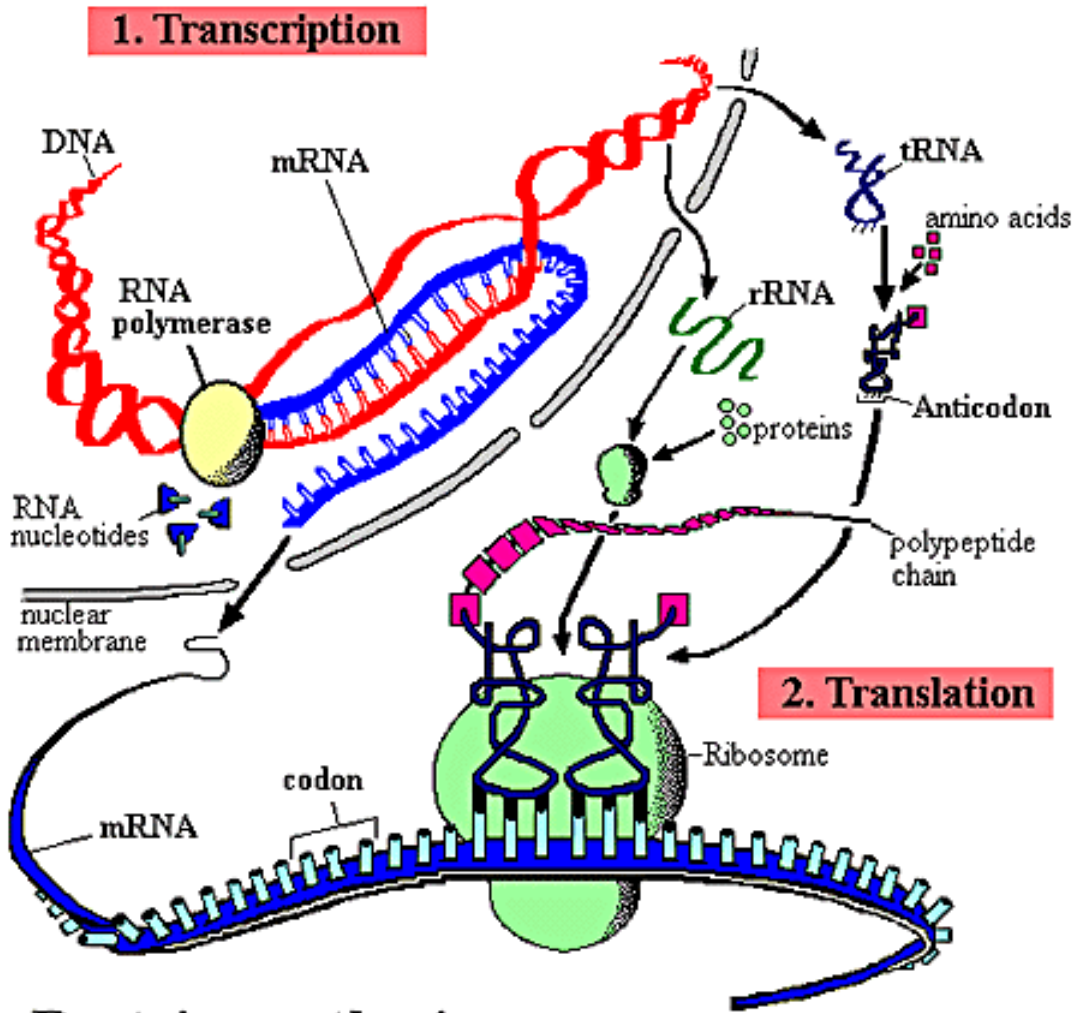
(D) The rest of the general transcription factors as well as the RNA polymerase itself assemble at the promoter. What order?

(E) TFIIH then uses ATP to phosphorylate RNA polymerase II, changing its conformation so that the polymerase is released from the complex and is able to start transcribing. As shown, the site of phosphorylation is a long polypeptide tail that extends from the polymerase molecule.

Transcription Factors

- The general transcription factors have been highly conserved in evolution; some of those from human cells can be replaced in biochemical experiments by the corresponding factors from simple yeasts.

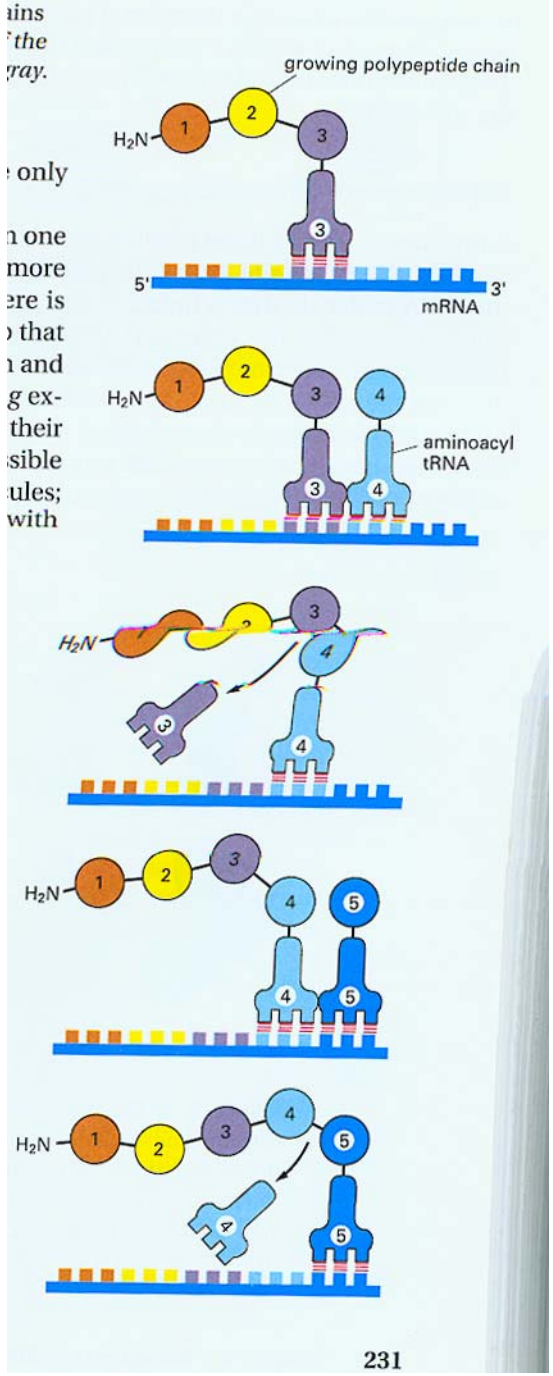
Protein Synthesis

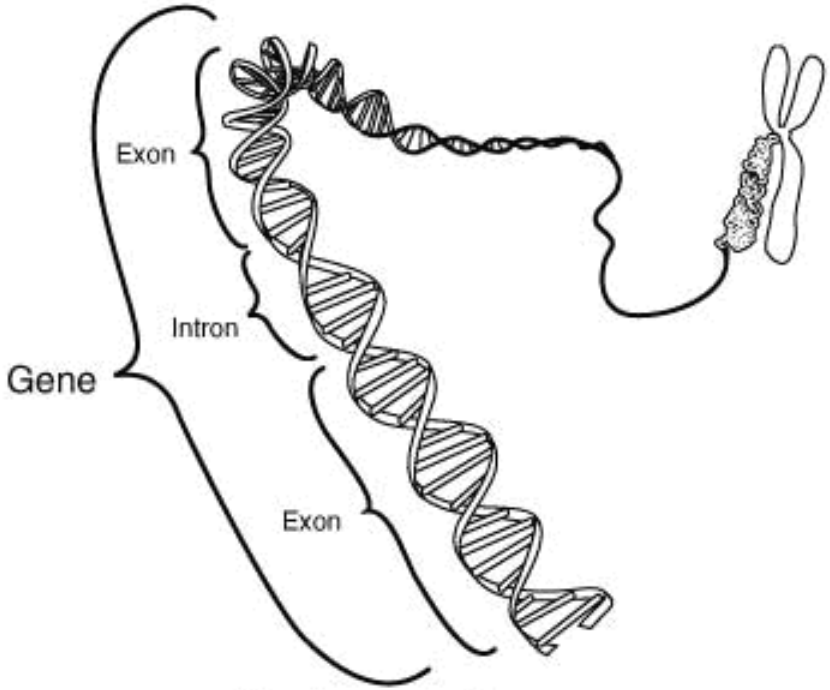


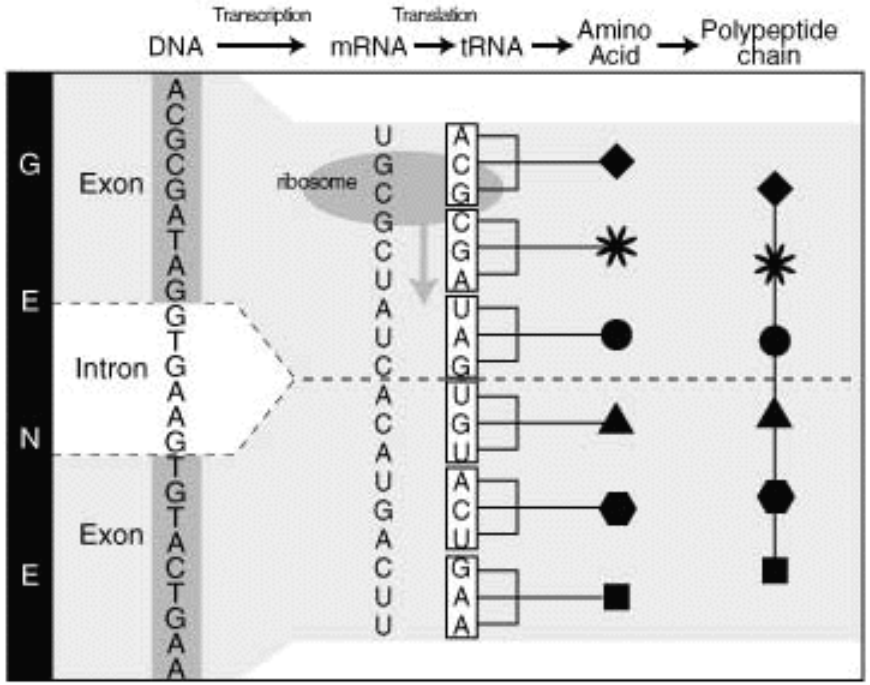
Protein synthesis

Protein Synthesis:

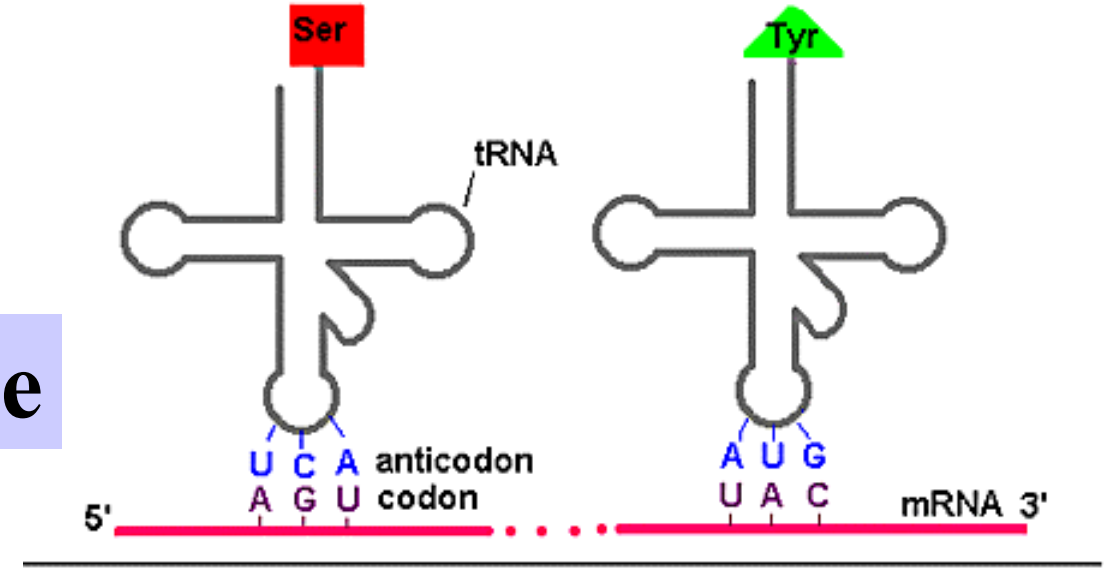
Incorporation of amino acid into protein




Exon
Intron
Gene
Exon


Transcription
Translation
DNA
mRNA
tRNA
Amino Acid
Polypeptide chain
G
E
N
E
Exon
Intron
Exon
ACGCGATAGGTGAAGTGTACTGAA
ribosome
UGCGCUAUCACACAUGACUU
ACG
CGA
UAG
UGU
ACU
GAA

The Genetic Code

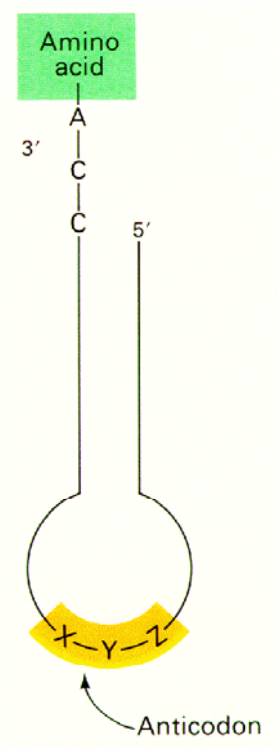


2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The Genetic Code

The Adaptor molecule: tRNA



tRNA is the Adaptor

Small RNA (76 bases)

Has acceptor site for amino acid at 3' end

Has a site for complementary base pairing to the mRNA (Anticodon Loop).

Nomenclature:

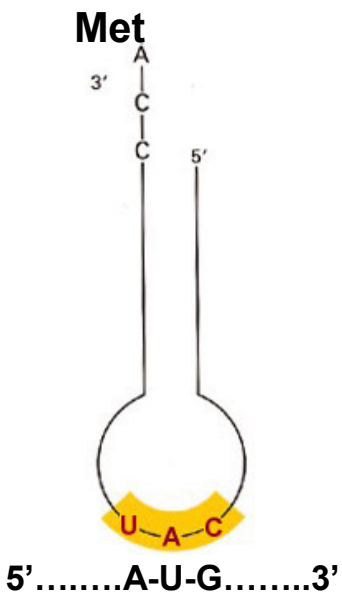
$tRNA^{\text{Amino acid}}$

Example:

tRNA for methionine is $tRNA^{Met}$

tRNA

Codon:Anticodon Interactions



Anticodon of tRNA must base pair to the codon.

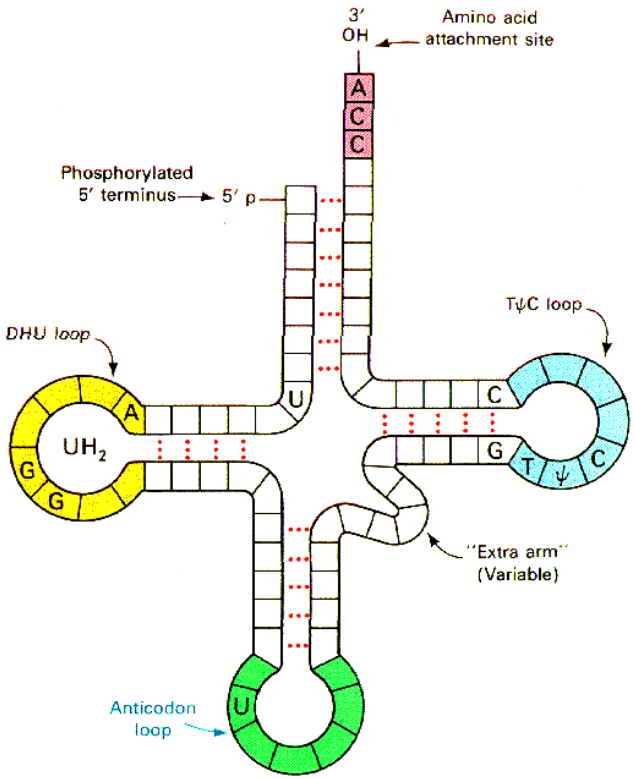
Convention dictates writing nucleotide sequence in the 5'-3'.

Thus, the anticodon of 5'-AUG-3' is 5'-CAU-3'.

Be cautious because at first glance this appears backwards.

tRNA

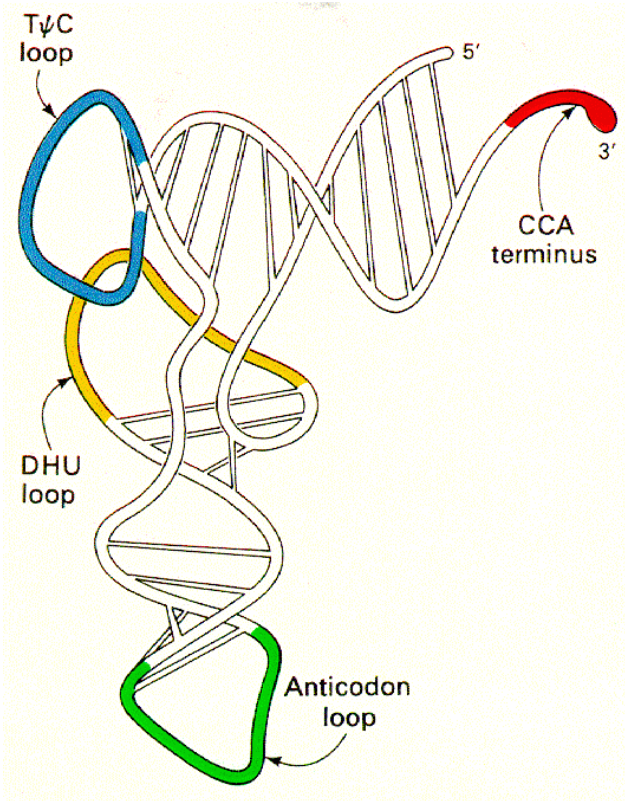
tRNA Sequence Proposed Structure



- **First nucleic acid ever sequenced**
- **The TΨC loop (thymidine-pseudouridine-cytidine) has these characteristic bases**
- **Same with the D loop and stem (dihydrouridine)**
- **Anticodon loop and anticodon stem**
- **Variable loop**
- **tRNAs <u>DO NOT</u> look like this in 3D**

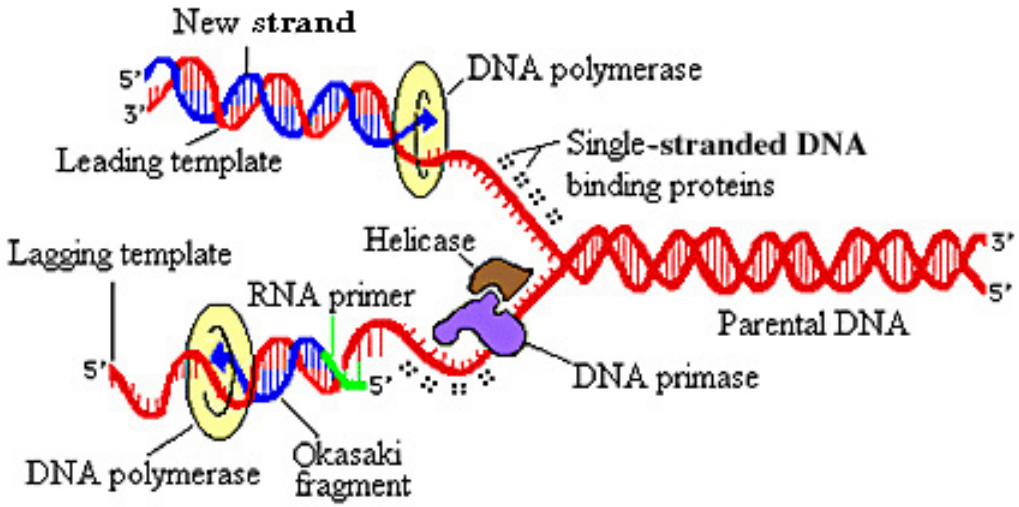
tRNA

tRNA Structure Solved



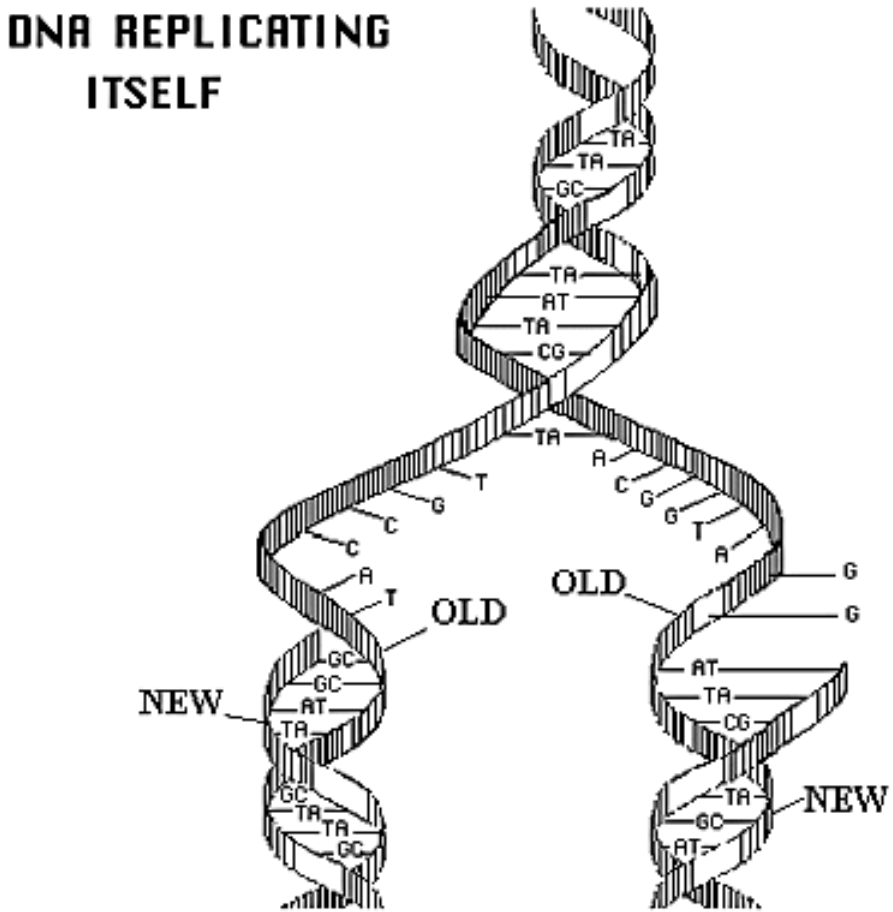
- **Solved by X-ray crystallography independently by Alexander Rich and Aaron Klug**
- **First RNA tertiary structure solved**
- **Binding between stems involves non-Watson-Crick interactions**

Replication



Collaboration of Proteins at the Replication Fork

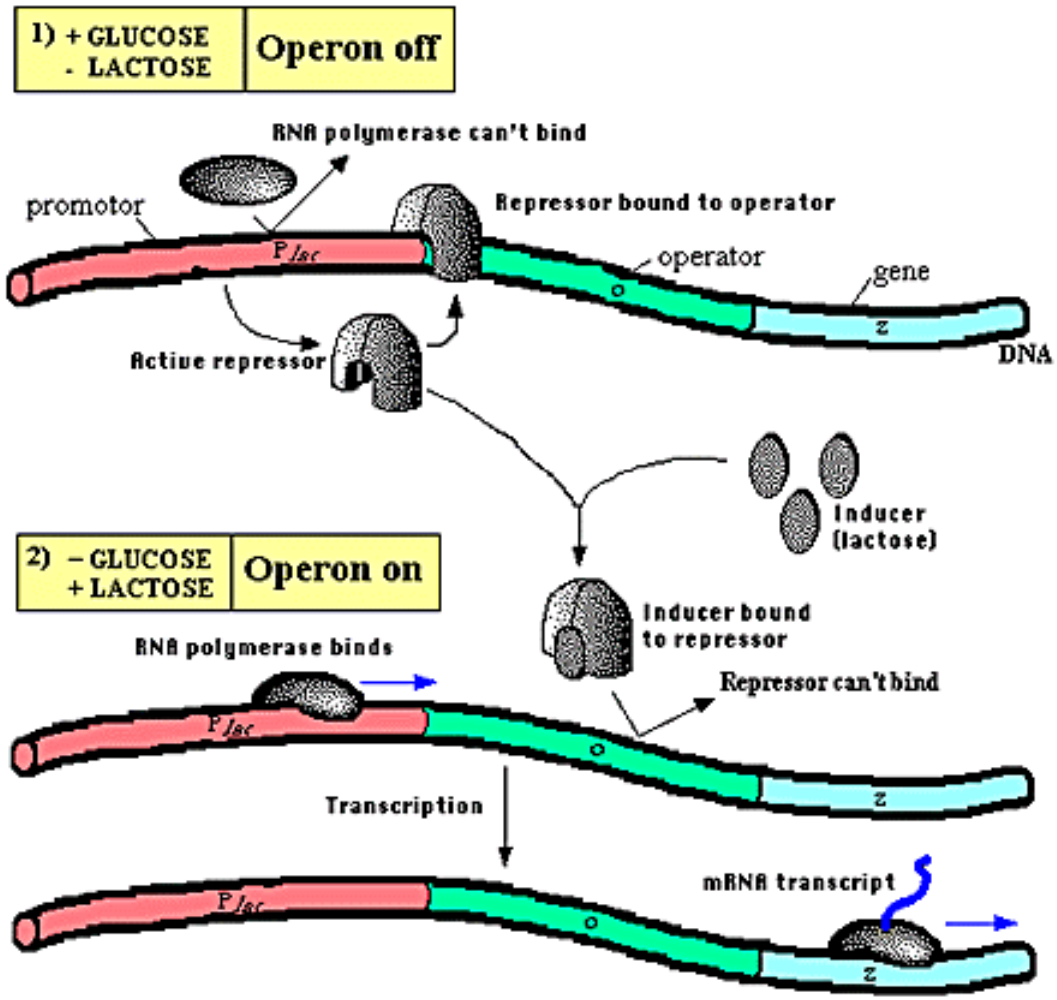
DNA Replication



Steps in Replication

1. Parent strands unwound by **DNA helicases**.
2. Single stranded **DNA binding proteins** attach to prevent strands from winding back together.
3. **DNA polymerase** catalyzes elongation of leading & lagging strands & checks accuracy of own work!).
4. RNA primer needed repeatedly on lagging strand to facilitate synthesis of Okazaki fragments. **DNA primase** (polypeptide bound together as part of a group - primosomes) helps to build the primer.
5. Finally, each new **Okazaki fragment** is attached to the completed portion of the lagging strand in a reaction catalyzed by **DNA ligase**.

lac Operon Induction

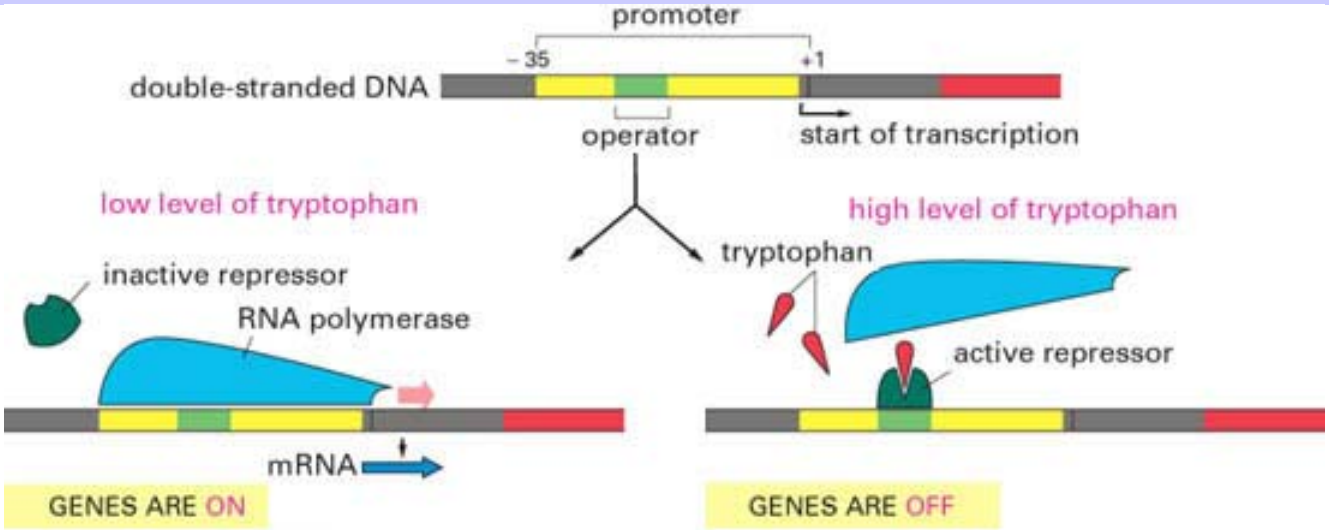


Induction of the *lac* Operon

Genetic Switches

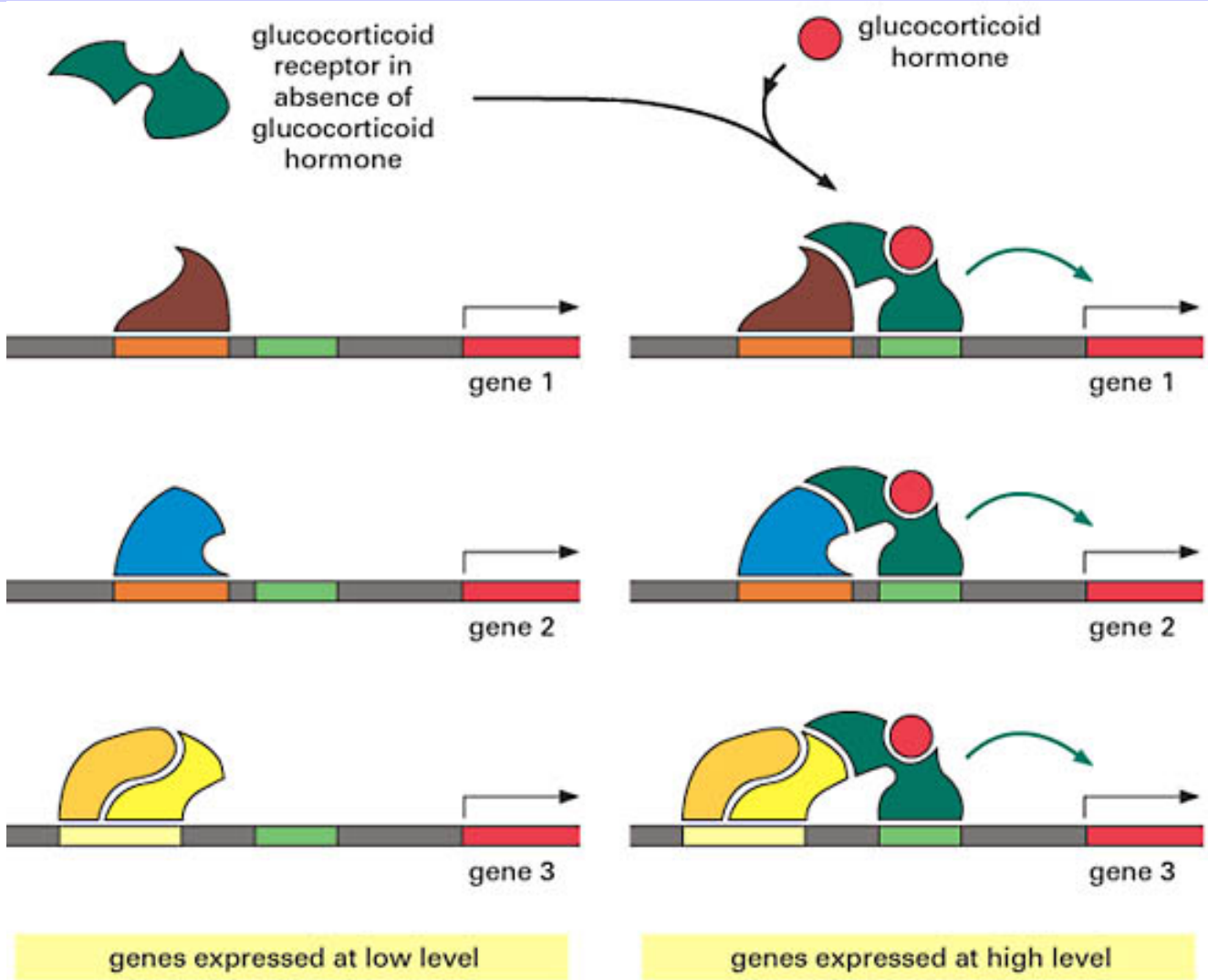
- Gene transcription can be switched on and off by gene regulation proteins, e.g. *lac* operon in *E.Coli*. Glucose and lactose levels control transcription initiation of *lac* operon, i.e., whether the lac operon is switched "ON" or "OFF".
- If **lactose** is absent, repressor protein binds to operator, prevents RNA polymerase from transcribing lac operon's genes. **Operon is OFF.**
- When inducer, **lactose**, is added, lactose (and its relatives) binds to repressor and changes the repressor's shape, which then leaves the operator. As long as the operator remains free of the repressor, RNA polymerase can transcribe operon's structural genes into mRNA. **Operon is ON.**

Tryptophan Operon



If level of tryptophan inside cell is low, the tryptophan repressor protein does not bind tryptophan and thus cannot bind to the operator within the promoter. RNA polymerase can thus bind to the promoter and transcribe the five genes of the tryptophan operon (left). If the level of tryptophan is high, however, the repressor binds tryptophan, in which state it can bind to operator, where it blocks the binding of RNA polymerase to the promoter (right). Whenever the level of intracellular tryptophan drops, the repressor releases its tryptophan and is released from the DNA, allowing the polymerase to again transcribe the operon.

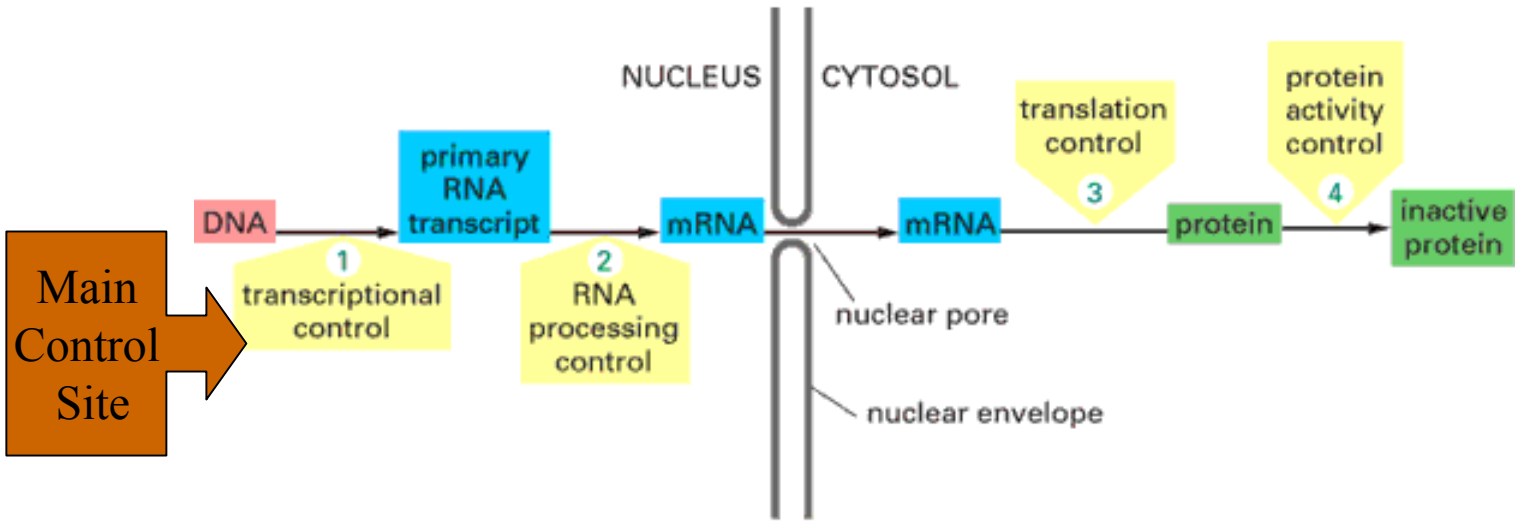
Gene Expression Regulation



Gene Expression Regulation

- The action of the glucocorticoid receptor is illustrated. On the left is shown a series of genes, each of which has various gene activator proteins bound to its regulatory region. However, these bound proteins are not sufficient on their own to activate transcription efficiently. On the right is shown the effects of adding an additional gene regulatory protein "the glucocorticoid receptor in a complex with glucocorticoid hormone" that can bind to the regulatory region of each gene. The glucocorticoid receptor completes the combination of gene regulatory proteins required for efficient initiation of transcription, and the genes are now switched on as a set.

Control of Gene Expression



Polymerase Chain Reaction

- Method to make large quantities of a piece of DNA. Kary Mullis was awarded the Nobel Prize for this **technique**
  - Heat to 95° C for 30 seconds to break bonds between the double strands of DNA.
  - Cool to 37° C for 30 seconds to allow single strand primers to bind to complementary ends of each strand.
  - Heat to 72° C for 60-120 seconds to allow DNA polymerase to extend the primers.
  - Repeat 20-30 cycles.

PCR

```
5′-ACATTTACGCGGATCCATAGGATCAGGATCAGCATTACGATCCGGAATTCCGGGAATTCGAATTCCAA-3′
3′-TGTAAATGCGCCTAGGTATCCTAGTCCTAGTCGTAATGCTAGGCCTTAAGGCCCTTAAGCTTAAGGTT-5′

5′-ACATTTACGCGGATCCAT----------------------------------------------------3′
3′----------------------------------------------GCCCTTAAGCTTAAGGTT-5′
```

PCR

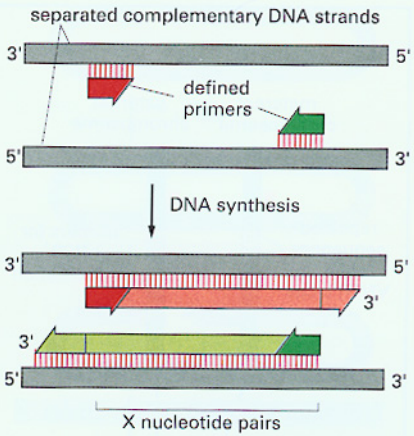


Figure 7–31 The start of the polymerase chain reaction (PCR) for amplifying specific nucleotide sequences *in vitro*. DNA isolated from cells is heated to separate its complementary strands. These strands are then annealed with an excess of two DNA oligonucleotides (each 15 to 20 nucleotides long) that have been chemically synthesized to match sequences separated by X nucleotides (where X is generally between 50 and 2000). The two oligonucleotides serve as specific primers for *in vitro* DNA synthesis catalyzed by DNA polymerase, which copies the DNA between the sequences corresponding to the two oligonucleotides.

Figure 7–32 PCR amplification. PCR produces an amount of DNA that doubles in each cycle of DNA synthesis and includes a uniquely sized DNA species. Three steps constitute each cycle, as described in the text. After many cycles of reaction, the population of DNA molecules becomes dominated by a single DNA fragment, X nucleotides long, provided that the original DNA sample contains the DNA sequence that was anticipated when the two oligonucleotides were designed. In the example illustrated, three cycles of reaction produce 16 DNA chains, 8 of which have this unique length (*yellow*); but after three more cycles, 240 of the 256 DNA chains would be X nucleotides long.

sequences corresponding to the two oligonucleotides.