Introduction to molecular biology

A6M33BIN

Karel Jalovec

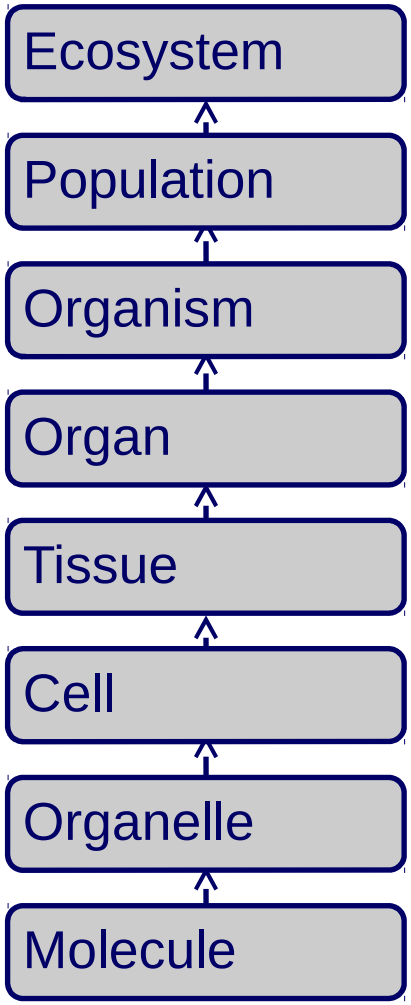
jalovkar@fel.cvut.cz

Spring 2014

Inspired by: Mark Craven (craven@biostat.wisc.edu), BMI/CS 576

www.biostat.wisc.edu/bmi576/

Biological hierarchy



Lake Monona, Mojave desert, …

Central Wisconsin Humans, Monona Muskie

Pine tree, Lizard, Human…

Heart, Lungs, Brain…

Adipose tissue, Blood, Nerve tissue…

Fat cell, Blood cell, Muscle cell…

Nucleus, Mitochondrion…

DNA, RNA, Protein, Lipids…

DNA

- “blueprint” for an organism
- composed of small molecules called *nucleotides*
- four different nucleotides distinguished by the four *bases*: adenine (A), cytosine (C), guanine (G) and thymine (T)
- is a *polymer* : large molecule consisting of similar units (nucleotides in this case)
- nucleotides form a strand of DNA. It can be thought of as a string composed of the four letters: A, C, G, T

  `ctgctggaccgggtgctaggaccctgactgcccggggccgggggtgcg gggcccgctgag`…

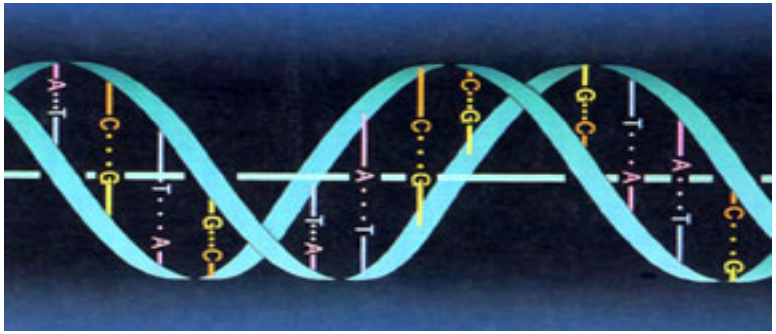
- DNA molecules usually consist of two strands arranged in the double helix

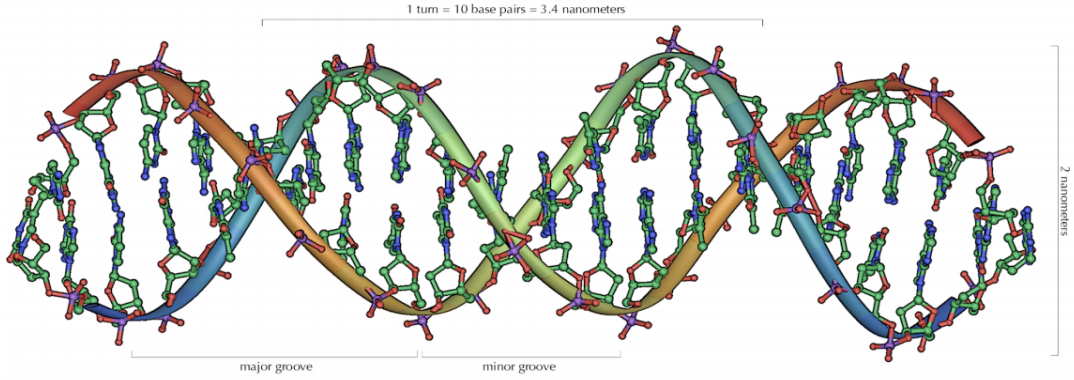
Watson-Crick base pairs

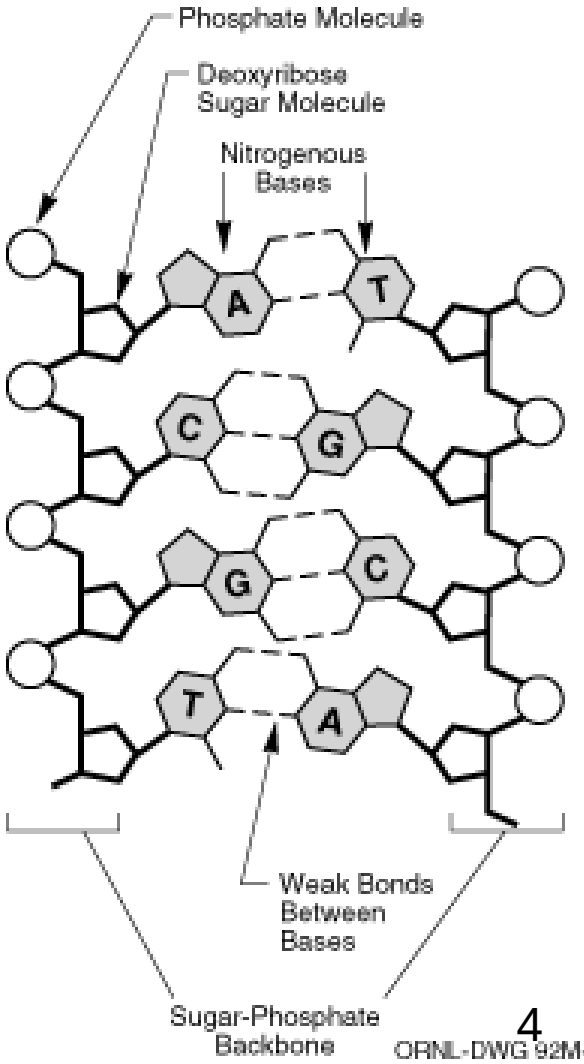
- in double-stranded DNA

  A always bonds to T

  C always bonds to G

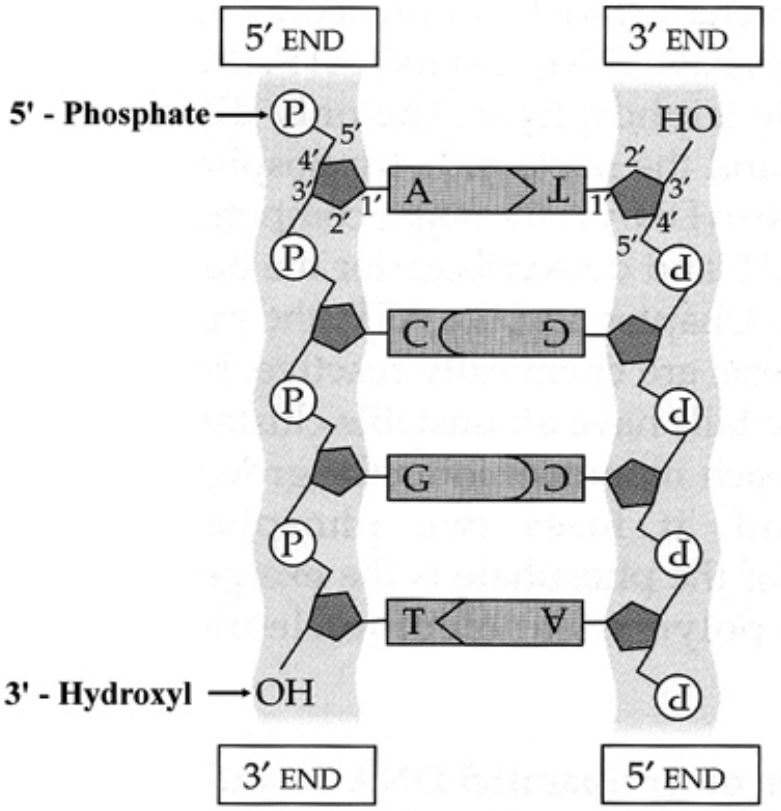






The double helix

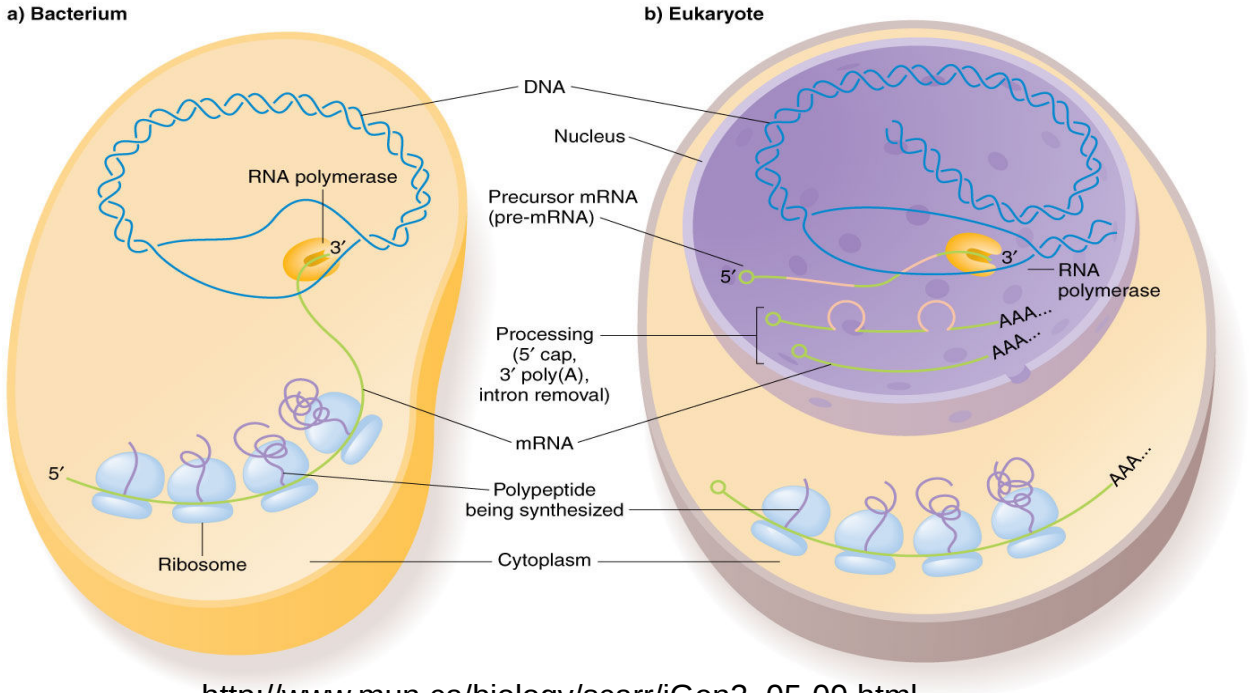
- each strand of DNA has a “direction”
- at one end, the terminal carbon atom in the backbone is the 5’ carbon atom of the terminal sugar
- at the other end, the terminal carbon atom is the 3’ carbon atom of the terminal sugar
- therefore we can talk about the *5’* and the *3’* ends of a DNA strand
- in a double helix, the strands are *antiparallel* (arrows drawn from the 5’ end to the 3’ end go in opposite directions)



http://hshgp.genome.washington.edu/teacher_resources/modules-view.htm

Chromosomes

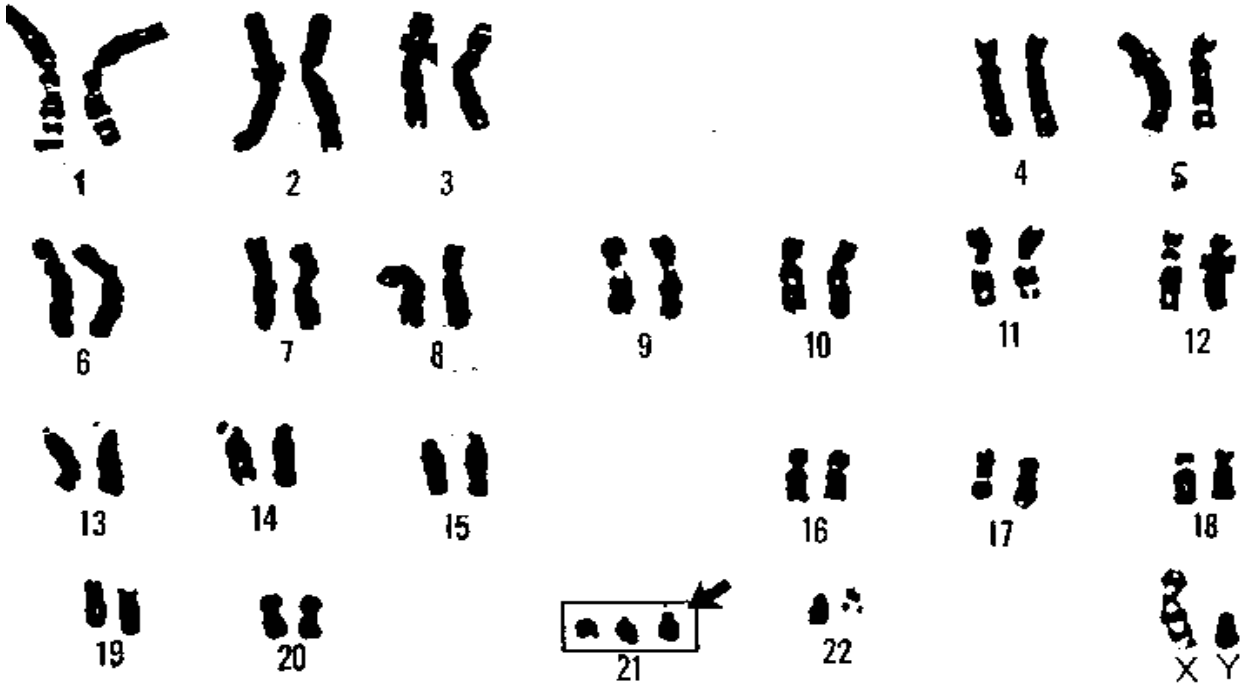
- DNA is packaged into individual *chromosomes* (along with proteins)
- *prokaryotes* (single-celled organisms lacking nuclei) typically have a single circular chromosome
- *eukaryotes* (organisms with nuclei) have a species-specific number of linear chromosomes



http://www.mun.ca/biology/scarr/iGen3_05-09.html

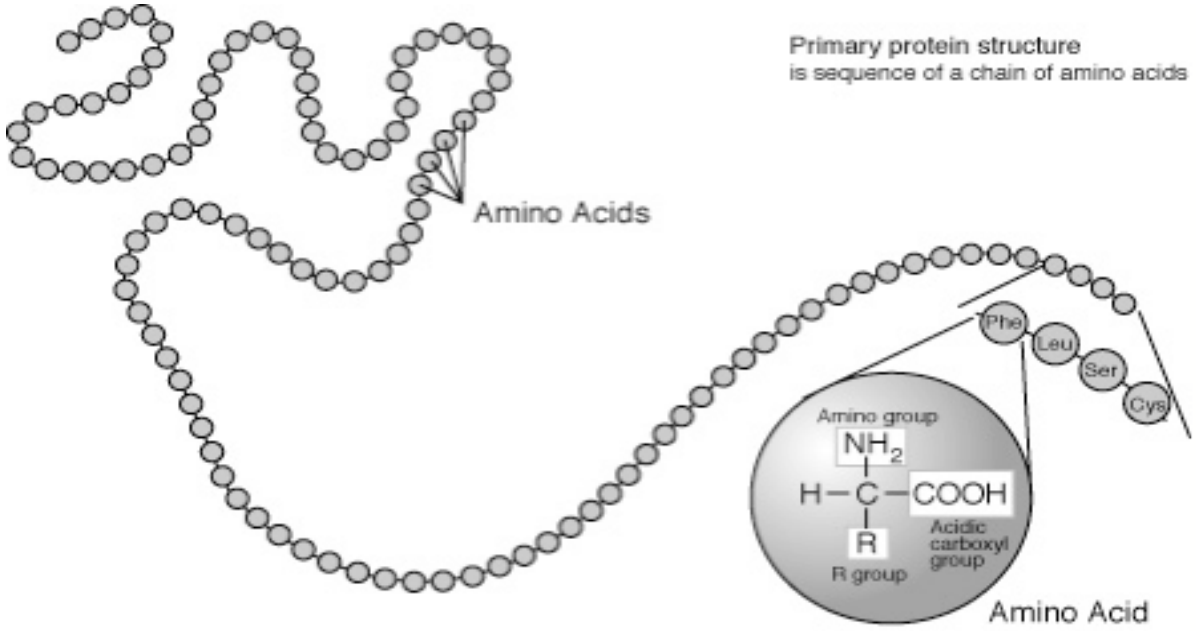
Genome

- the term *genome* refers to the complete complement of DNA for a given species
- the human genome consists of 46 chromosomes
- every cell (except sex cells and mature red blood cells) contains the complete genome of an organism



Proteins

- proteins are molecules composed of one or more *polypeptides*
- a polypeptide is a polymer composed of *amino acids*
- cells build their proteins from 20 different amino acids
- a polypeptide can be thought of as a string composed from a 20-character alphabet



Protein structure

- Primary structure

  amino acid linear sequence of the polypeptide chain

- Secondary structure

  highly regular local sub-structures (alpha helix, beta strand or beta sheets)

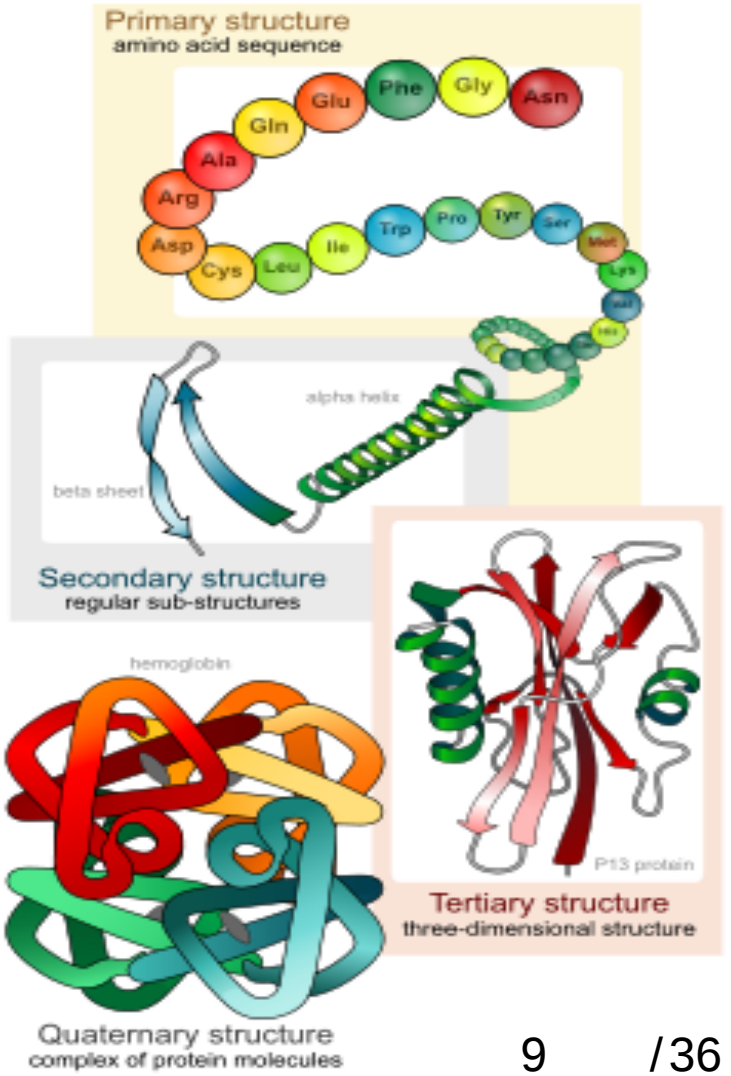
- Tertiary structure

  three-dimensional structure of a protein molecule (given by hydrophilic and hydrophobic properties of molecules)

- Quaternary structure

  three-dimensional structure of a multi-subunit protein and how the subunits fit together

- Protein structure



The alphabet of proteins

Alanine	**Ala**	**A**
Arginine	**Arg**	**R**
Aspartic Acid	**Asp**	**D**
Asparagine	**Asn**	**N**
Cysteine	**Cys**	**C**
Glutamic Acid	**Glu**	**E**
Glutamine	**Gln**	**Q**
Glycine	**Gly**	**G**
Histidine	**His**	**H**
Isoleucine	**Ile**	**I**
Leucine	**Leu**	**L**
Lysine	**Lys**	**K**
Methionine	**Met**	**M**
Phenylalanine	**Phe**	**F**
Proline	**Pro**	**P**
Serine	**Ser**	**S**
Threonine	**Thr**	**T**
Tryptophan	**Trp**	**W**
Tyrosine	**Tyr**	**Y**
Valine	**Val**	**V**

Protein functions

- structural support
- storage of amino acids
- transport of other substances
- coordination of an organism’s activities
- response of cell to chemical stimuli
- movement
- protection against disease
- selective acceleration of chemical reactions
- ATP synthase - powering your cells

DNA variations - mutations

```
ATG ACC CAG CAG CCA ATG AAA
Met Thr Gln Gln Pro Met Lys
```

Normal sequence
(Reading frame marked with spaces)

```
    ▼
ATG CCC CAG CAG CCA ATG AAA
Met Pro Gln Gln Pro Met Lys
```

Misssense substitution
(threonine replaced by proline)

```
       ▼
ATG ACC TAG CAG CCA ATG AAA
Met Thr STOP -   -   -   -
```

Nonsense substitution
(altered DNA sequence prematurely signals the cell to stop building a protein)

```
     ▼
ATG ACA CAG CAG CCA ATG AAA
Met Thr Gln Gln Pro Met Lys
```

Silent substitution
(threonine encoded by different codon)

```
    ▼▼▼
ATG --- CAG CAG CCA ATG AAA
Met  -  Gln Gln Pro Met Lys
```

Deletion without shift of a reading frame
(deletes one amino acid)

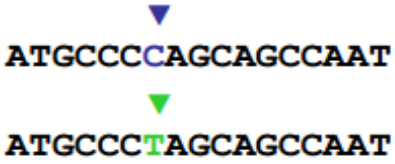
```
    ▼
ATG -CCC AGC AGC CAA TGA AA
Met  Pro Ser Ser Gln STOP -
```

Deletion with shift of a reading frame
(causes changes in following amino acid sequence and causes premature termination)

```
                ▼▼▼ ▼▼▼ ▼▼▼
ATG ACC CAG CAG CAG CAG CAG CCA ATG AAA
Met Thr Gln Gln Gln Gln Gln Pro Met Lys
```

Expansion of a trinucleotide repeats

DNA variations - polymorphism

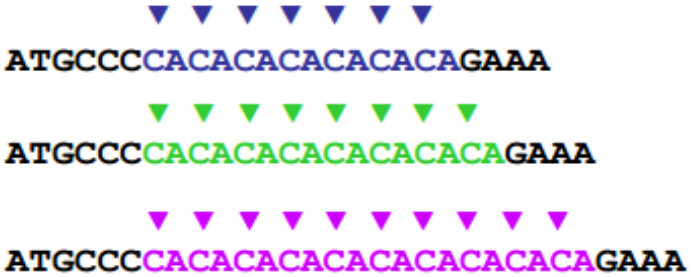


Single nucleotide polymorphism

Chromosome A

Chromosome B

Three possible genotypes



Short tandem repeat

Allele A

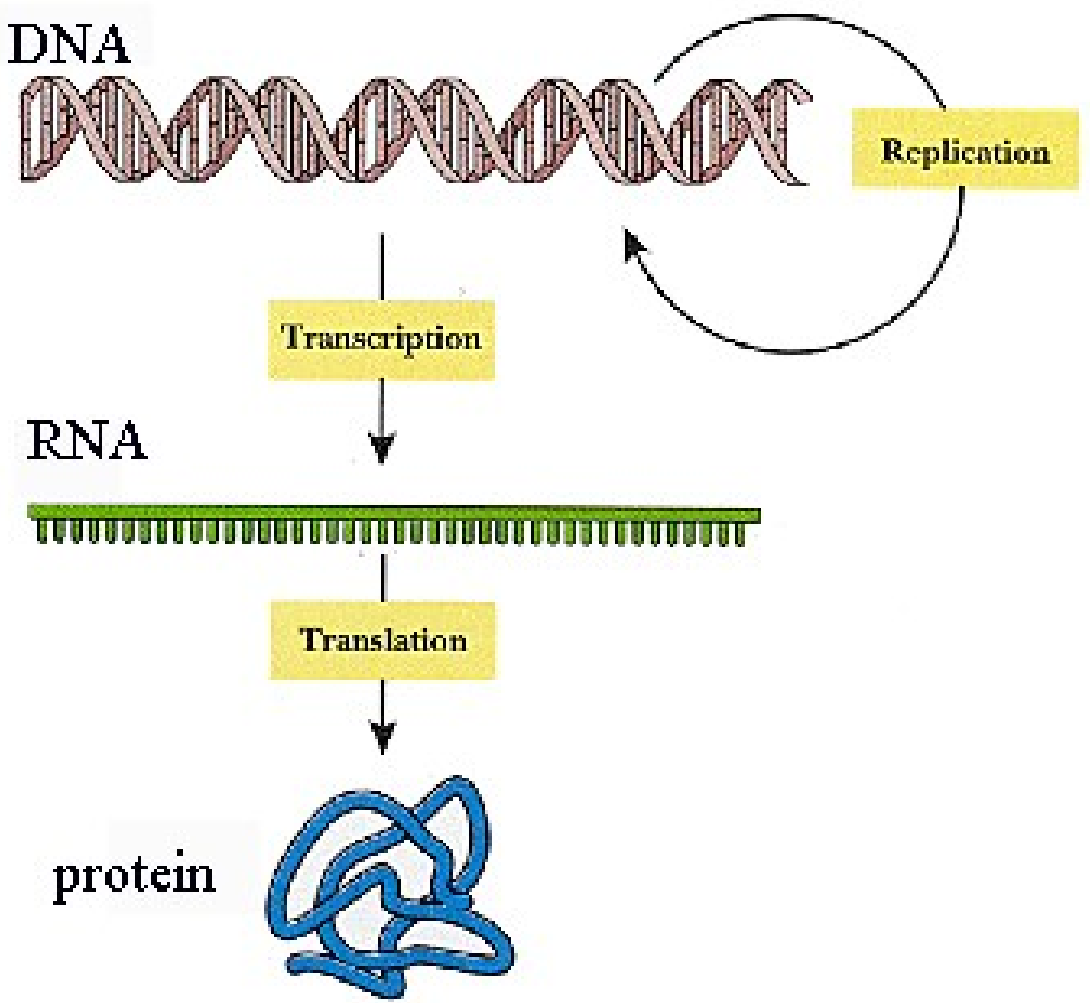
Allele B

Allele C

Many possible genotypes

How to construct a protein from a DNA?

The central dogma



DNA content

- Coding DNA

  genes (gene is a sequence of DNA bases that carries the information required for constructing a particular protein – encodes a protein)

  the human genome comprises ~ 25,000 protein-coding genes

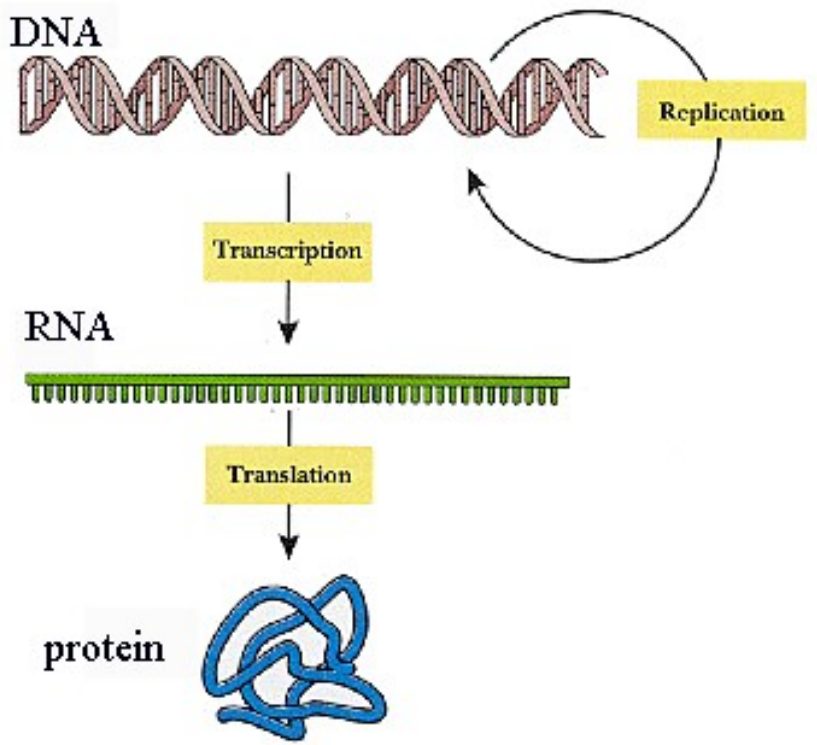
- Non-coding DNA

  transcribed into functional non-coding RNA molecules, give rise to RNA transcripts

  - *ribosomal RNA* (rRNA), which includes major constituents of ribosomes
  - *transfer RNAs* (tRNAs), which carry amino acids to ribosomes
  - *micro RNAs* (miRNAs), which play an important regulatory role in various plants and animals
  - etc.

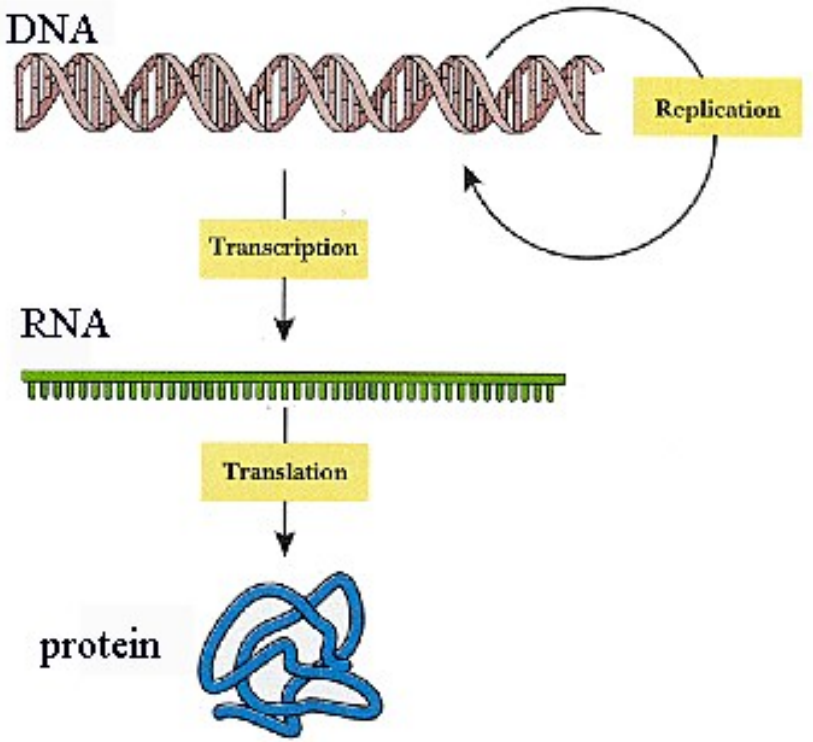
- Junk DNA

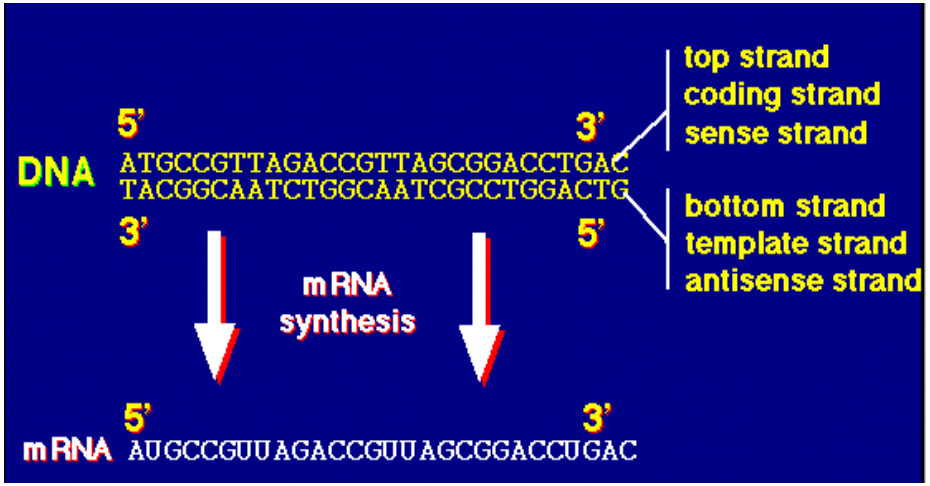
  no known biological function



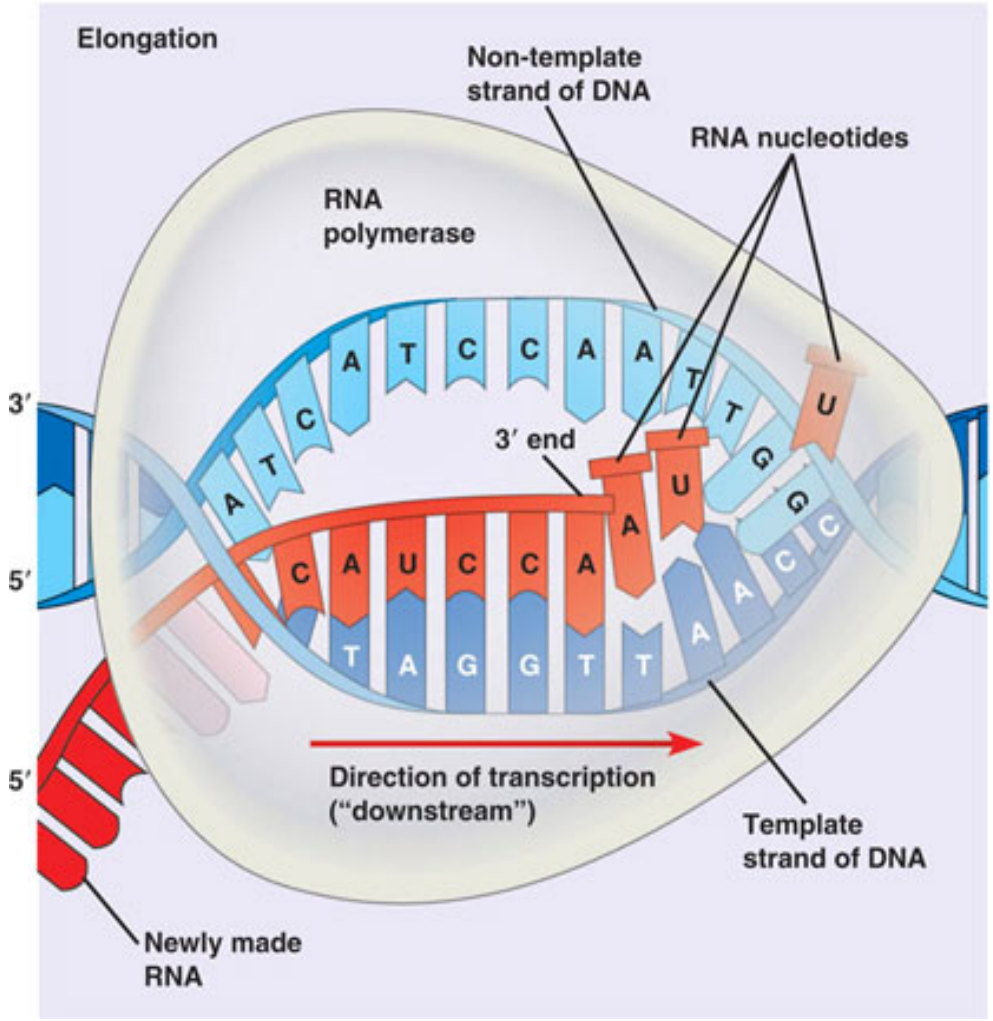
Transcription

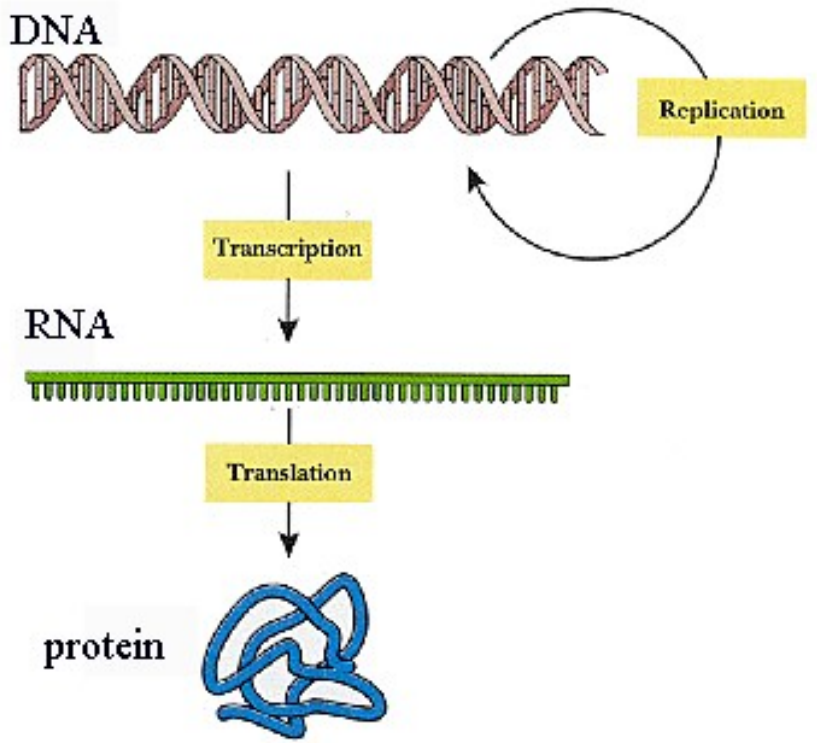
- Transcription is the process of creating a complementary RNA copy of a DNA subsequence
- *RNA polymerase* is the enzyme that builds an RNA strand from a gene
- RNA that is transcribed from a gene is called *messenger RNA* (mRNA)





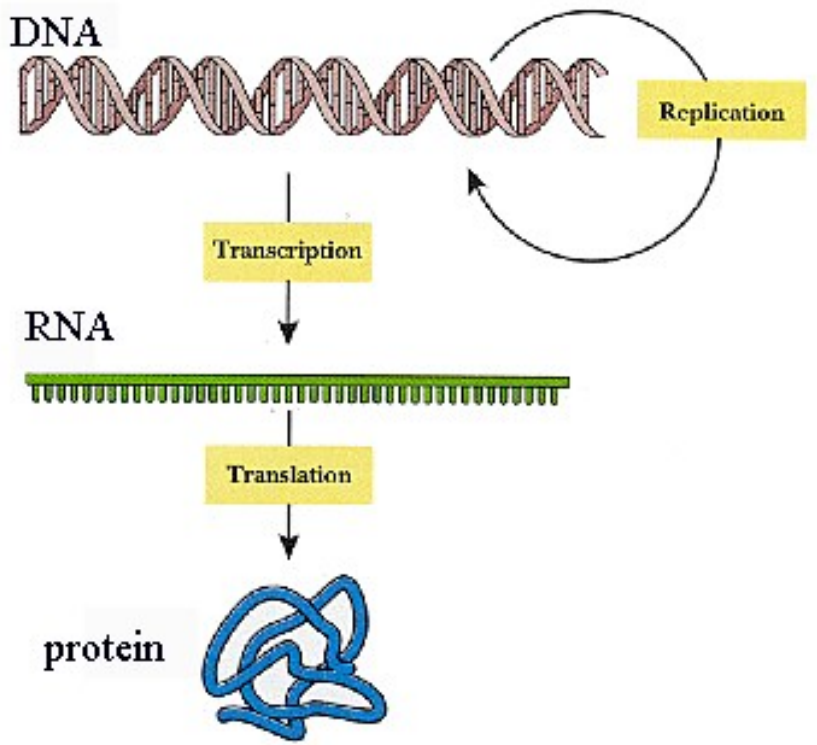
Transcription





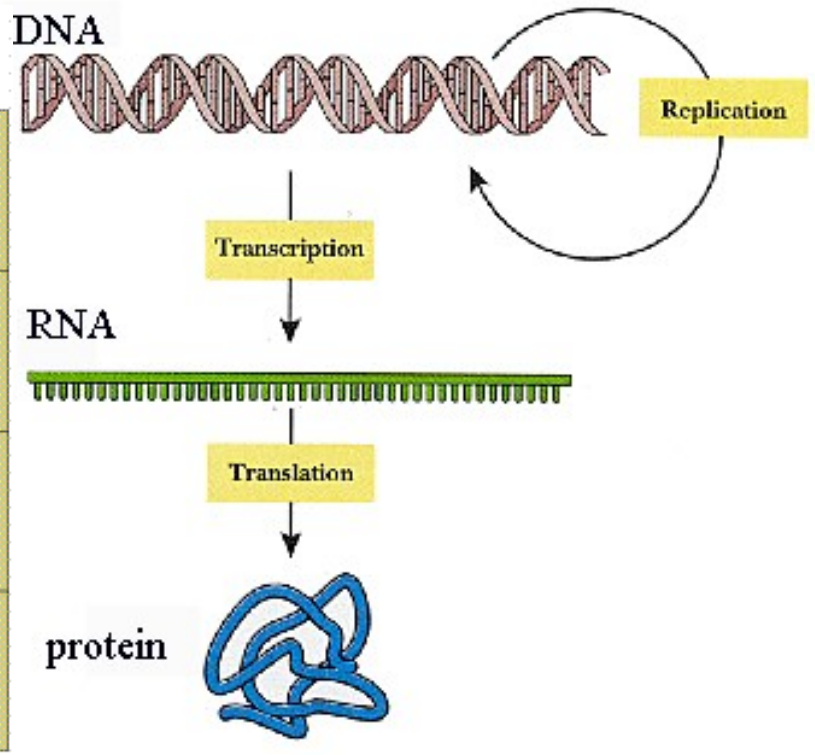
RNA

- RNA is like DNA except:
  - backbone is a little different
  - often single stranded
  - the base uracil (U) is used in place of thymine (T)
  - a strand of RNA can be thought of as a string composed of the four letters: A, C, G, U
- triplets of bases encoding amino acids are called *codons*
- the grouping of codons is called the *reading frame*

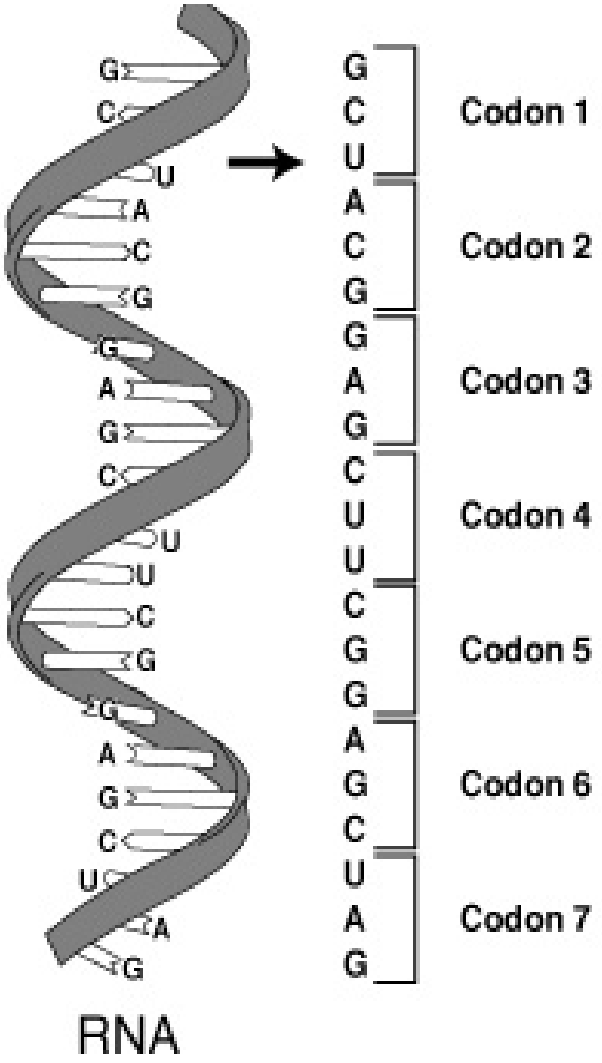


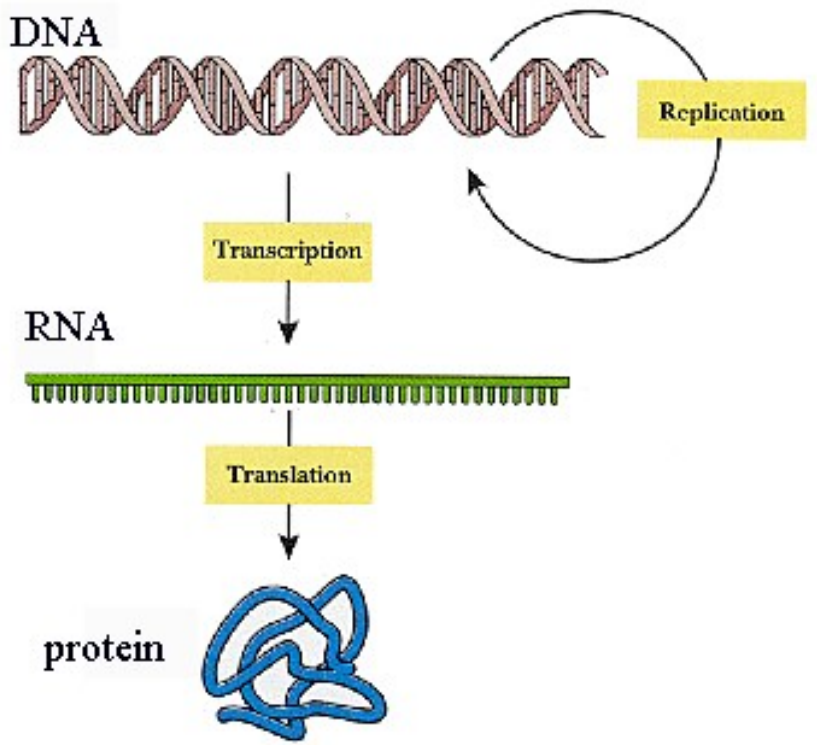
Codons

First letter	Second letter: U		Second letter: C		Second letter: A		Second letter: G		
U	UUU, UUC	Phenyl-alanine	UCU, UCC, UCA, UCG	Serine	UAU, UAC	Tyrosine	UGU, UGC	Cysteine	U C A G
	UUA, UUG	Leucine			UAA, UAG	Stop codon Stop codon	UGA	Stop codon	
							UGG	Tryptophan	
C	CUU, CUC, CUA, CUG	Leucine	CCU, CCC, CCA, CCG	Proline	CAU, CAC	Histidine	CGU, CGC, CGA, CGG	Arginine	U C A G
					CAA, CAG	Glutamine			
A	AUU, AUC, AUA	Isoleucine	ACU, ACC, ACA, ACG	Threonine	AAU, AAC	Asparagine	AGU, AGC	Serine	U C A G
	AUG	Methionine; initiation codon			AAA, AAG	Lysine	AGA, AGG	Arginine	
G	GUU, GUC, GUA, GUG	Valine	GCU, GCC, GCA, GCG	Alanine	GAU, GAC	Aspartic acid	GGU, GGC, GGA, GGG	Glycine	U C A G
					GAA, GAG	Glutamic acid			



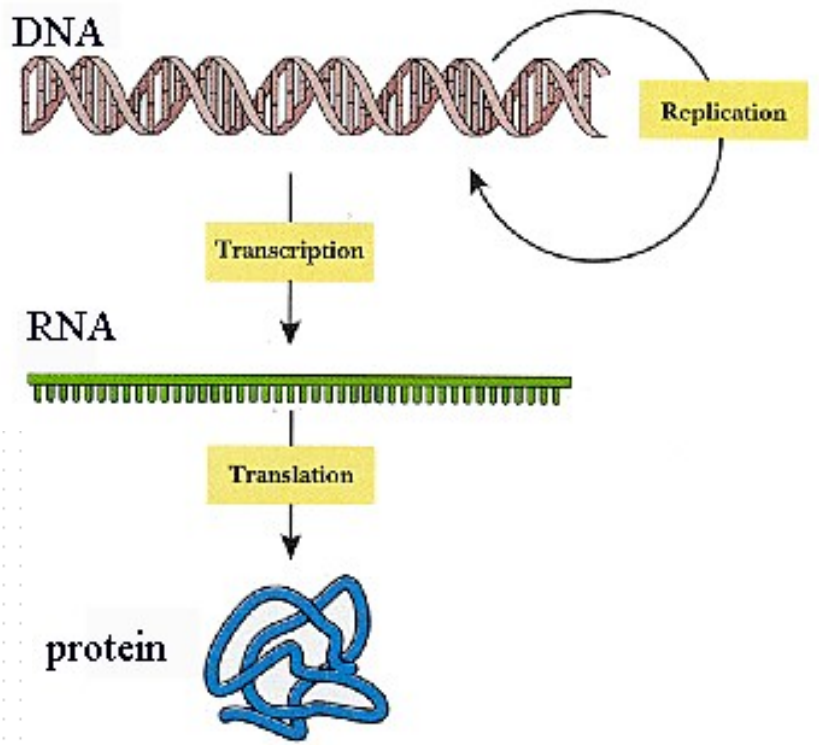
Codons and reading frames

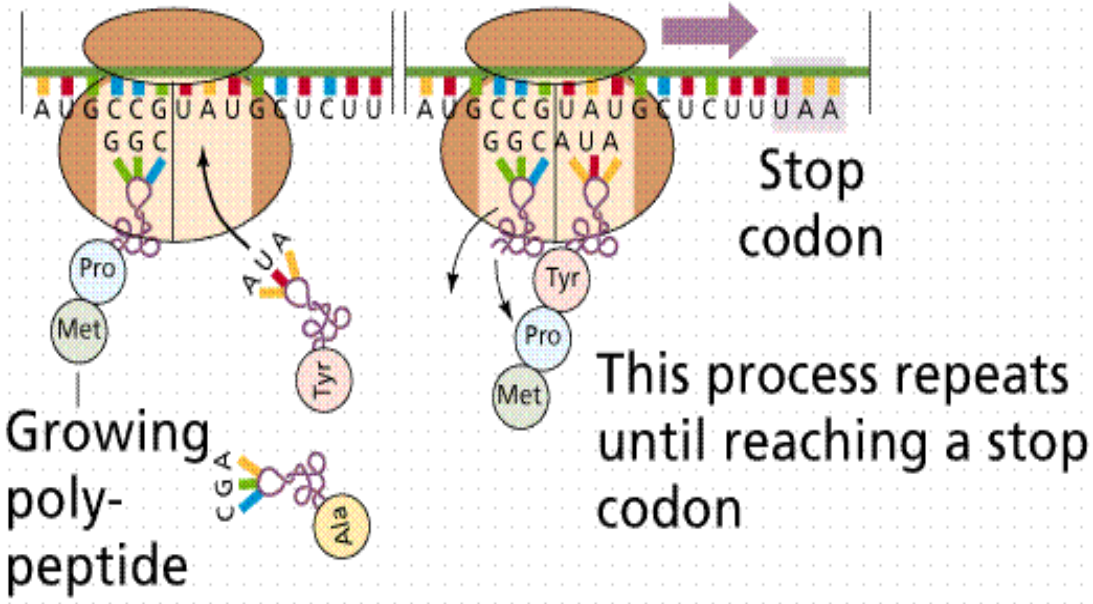




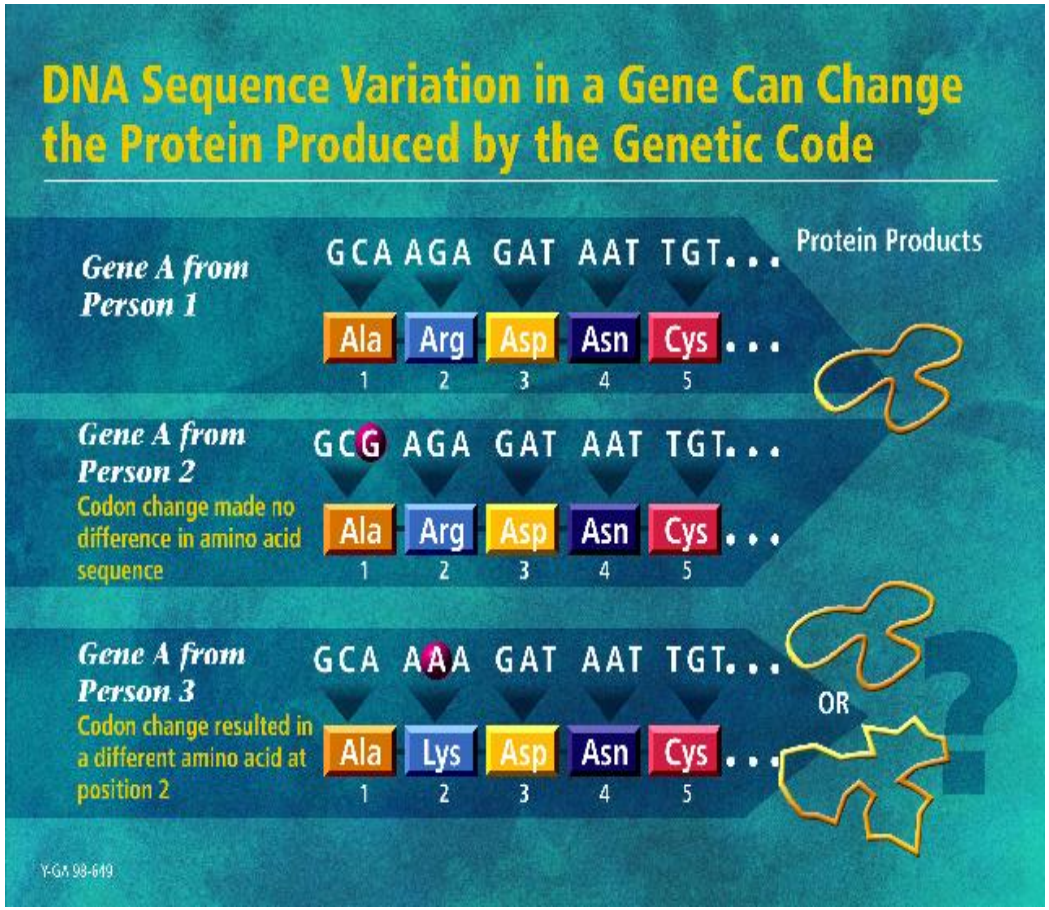
Translation

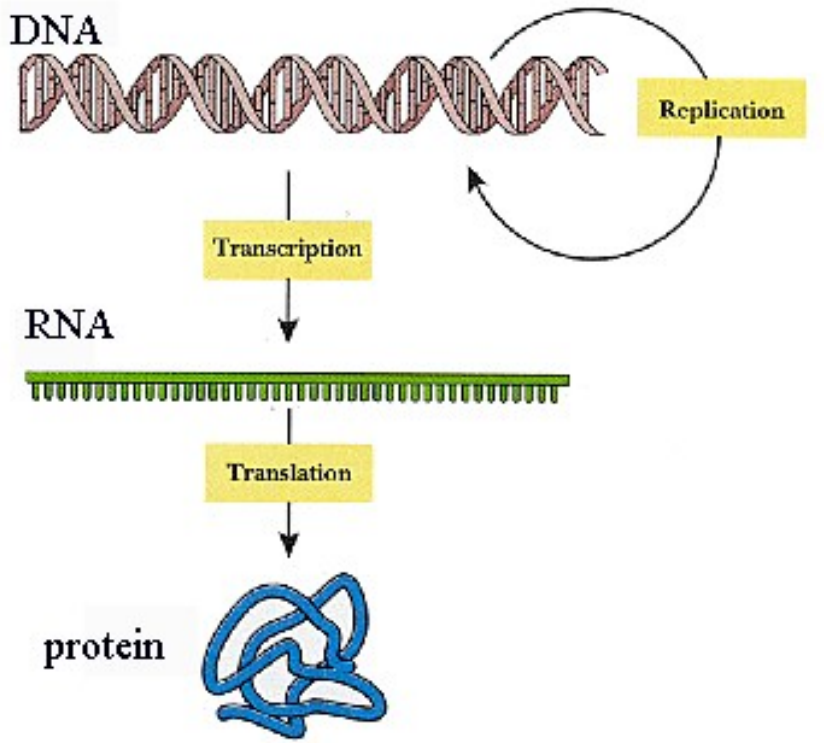
- *ribosomes* are the machines that synthesize proteins from mRNA
- translation begins with the *start codon and* ends with the *stop codon*





Translation

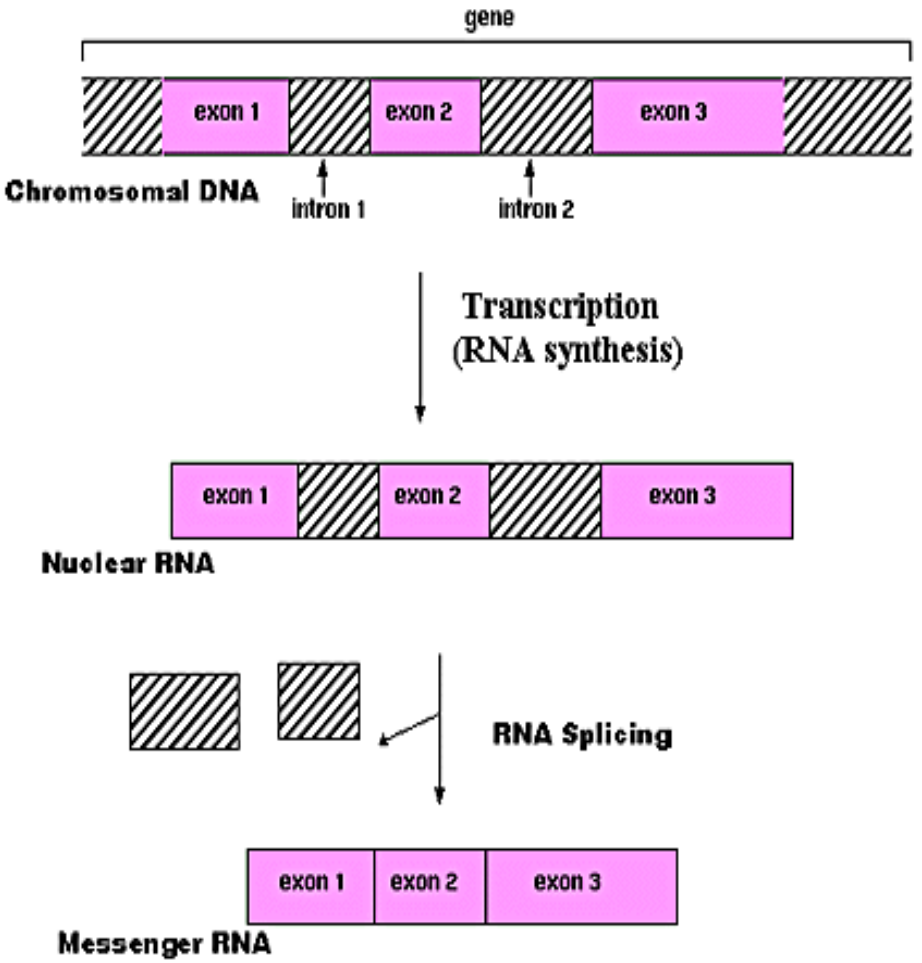




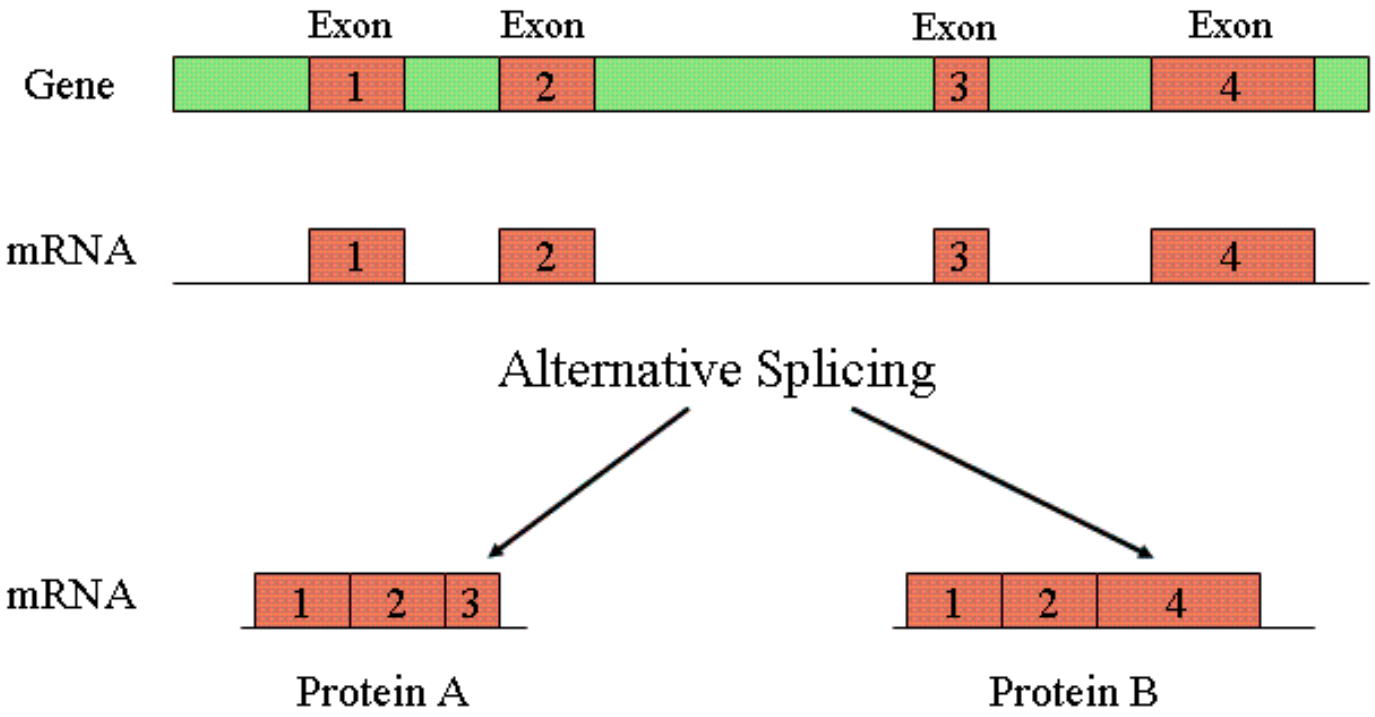
http://www.ornl.gov/hgmis

RNA processing in eukaryotes

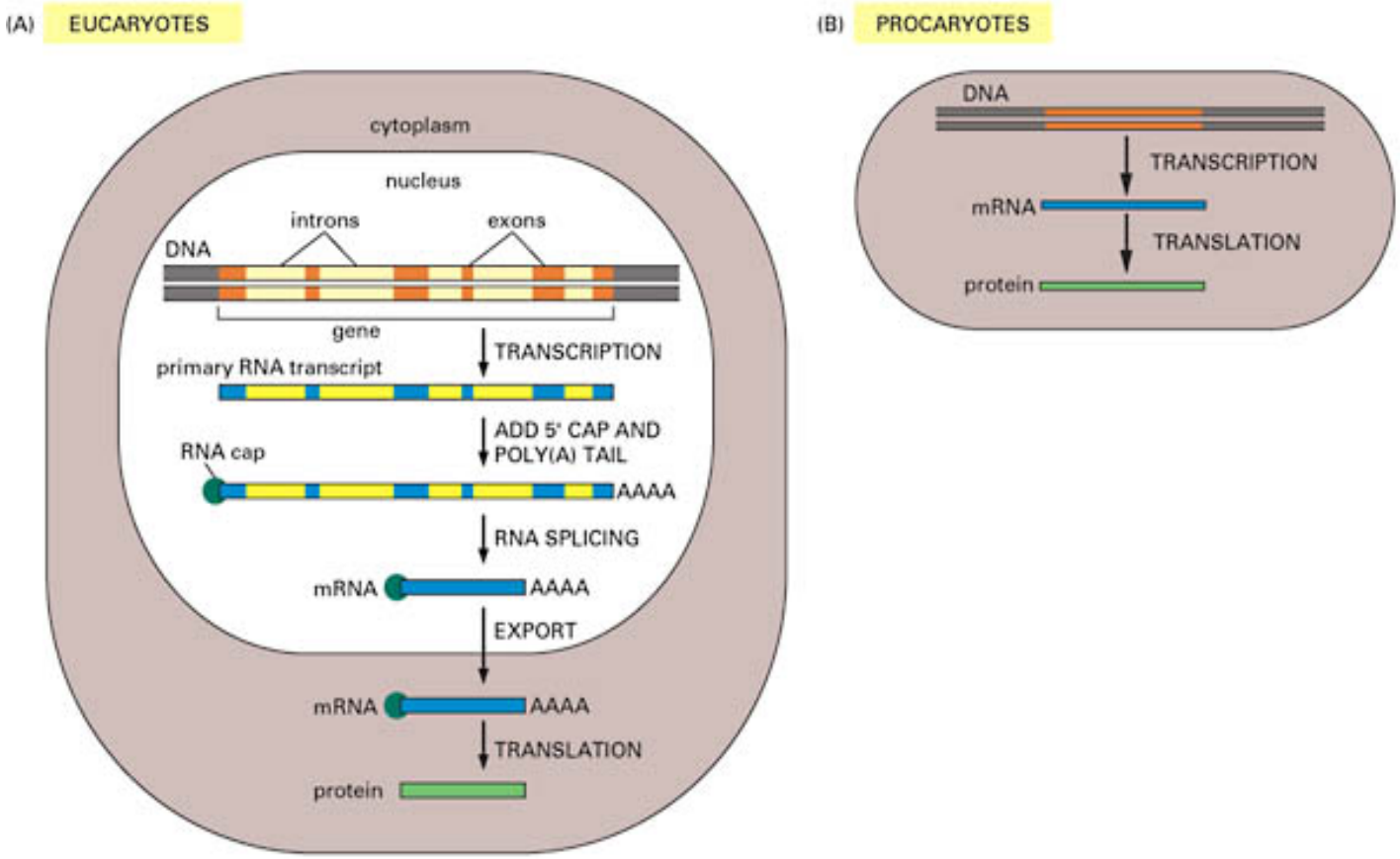
- *eukaryotes* are organisms that have enclosed nuclei in their cells
  - in many eukaryotes, genes/mRNAs consist of alternating *exon/intron* segments
  - exons are the coding parts
  - introns are spliced out before translation



Alternative splicing



Protein synthesis in eukaryotes vs. prokaryotes



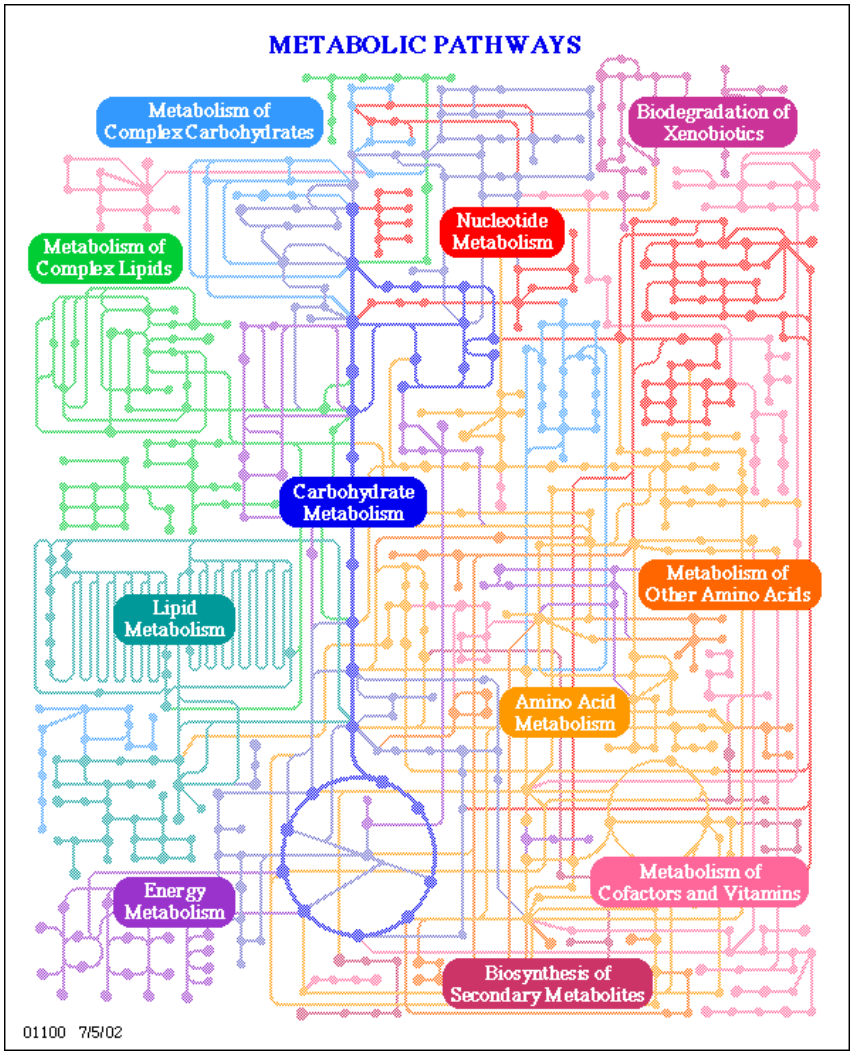
From DNA to protein

All cells in an organism have the same genomic data. How is possible to create different kind of cells and tissue?

The dynamics of cells

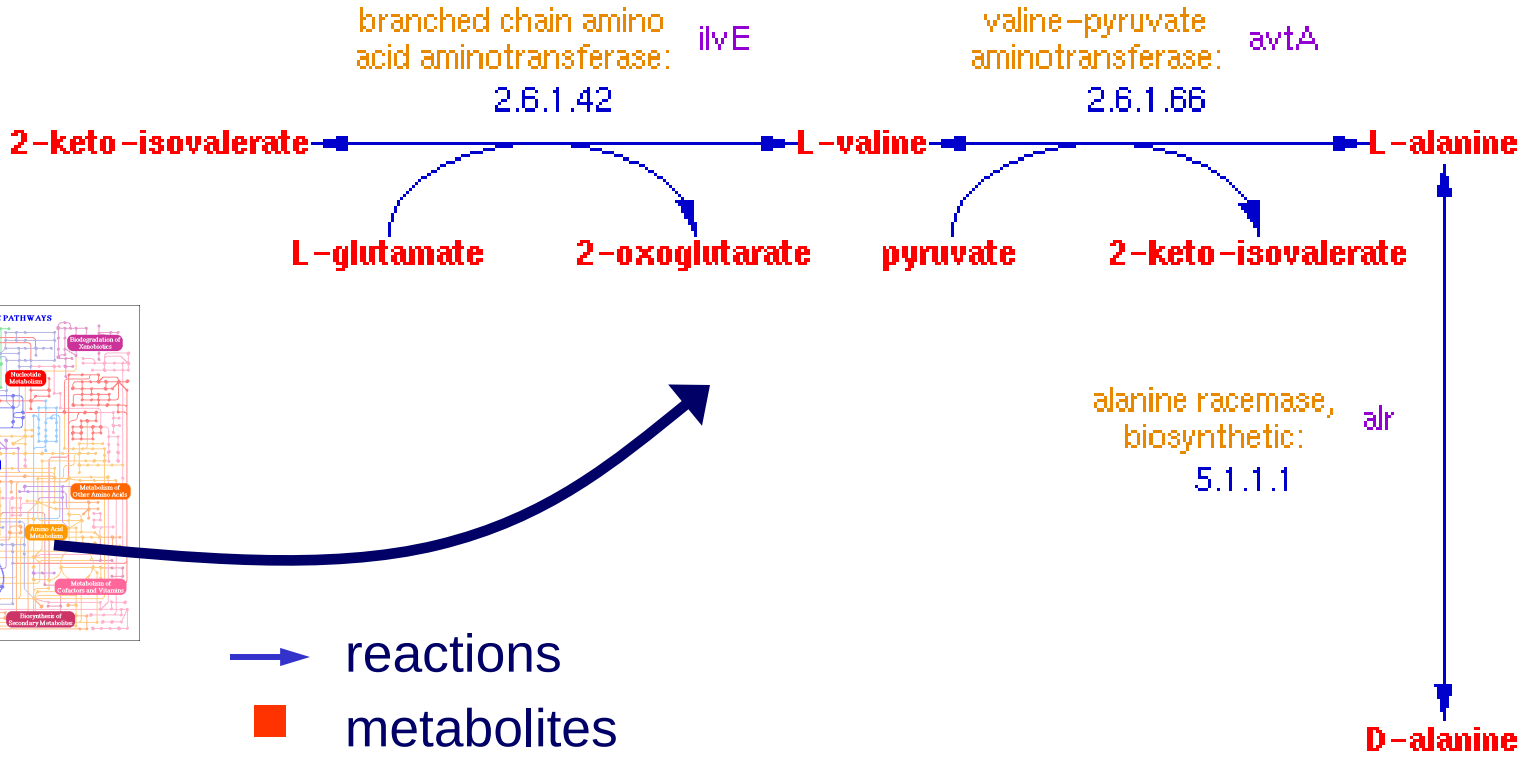
- all cells in an organism have the same genomic data, but the genes expressed in each vary according to cell type, time, and environmental factors
- there are networks of interactions among various biochemical entities in a cell (DNA, RNA, protein, small molecules) that carry out processes such as
  - metabolism
  - intra-cellular and inter-cellular signaling
  - regulation of gene expression

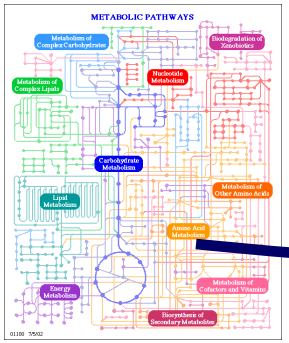
E. coli metabolic pathway map



- circles represent metabolites (small molecules)
- lines represent reactions

The metabolic pathway for synthesizing the amino acid alanine





- reactions
- metabolites
- enzymes (proteins that catalyze reactions)
- genes encoding the enzymes

image from the Ecocyc database
www.biocyc.org

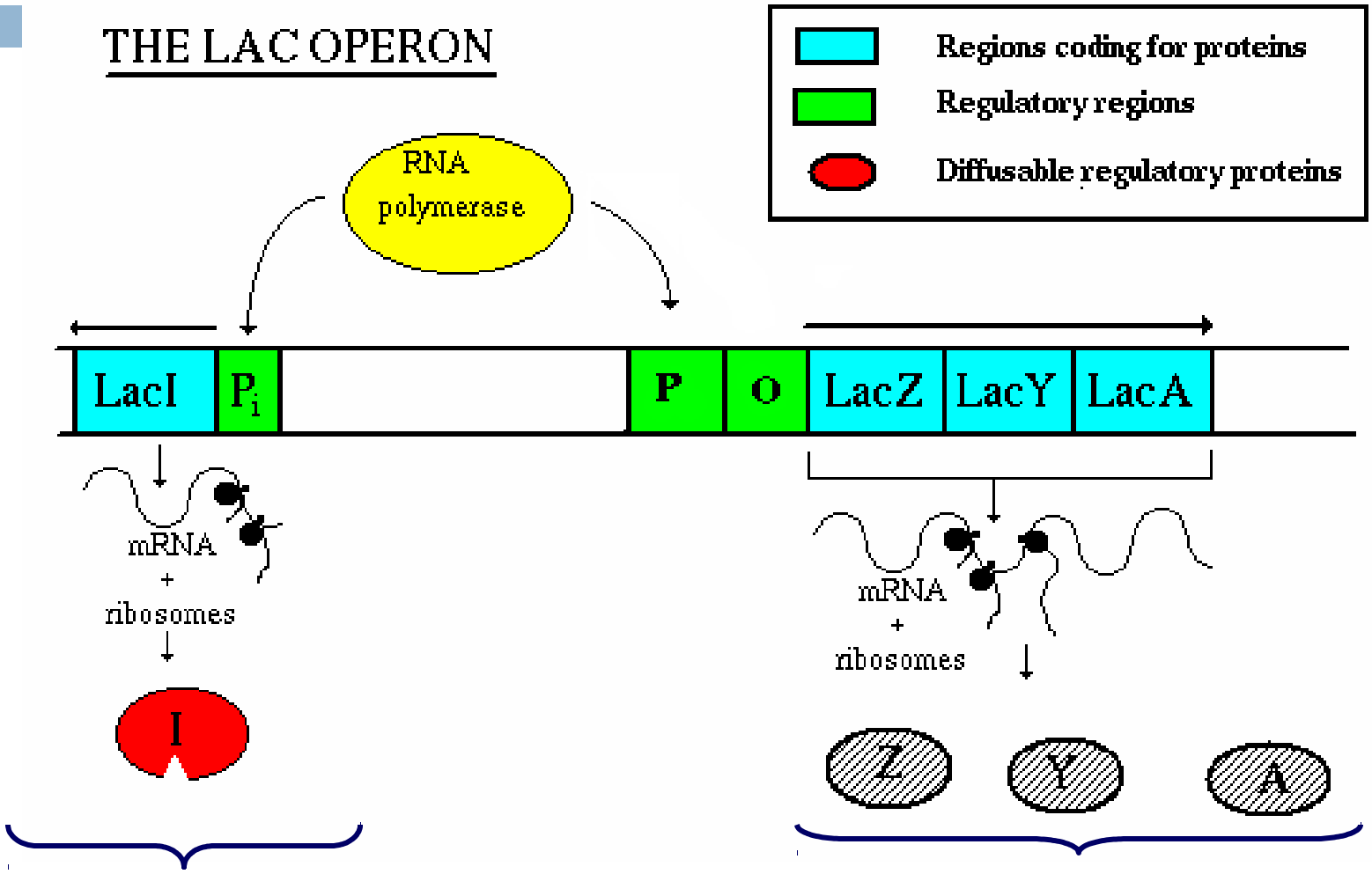
Operon

- Operon is a part of the DNA molecule of the prokaryotic organism that contains two or more genes, but operates as a single transcriptional unit.
- The genes in an operon are transcribed together (and often regulated together).

Gene expression

Lac operon

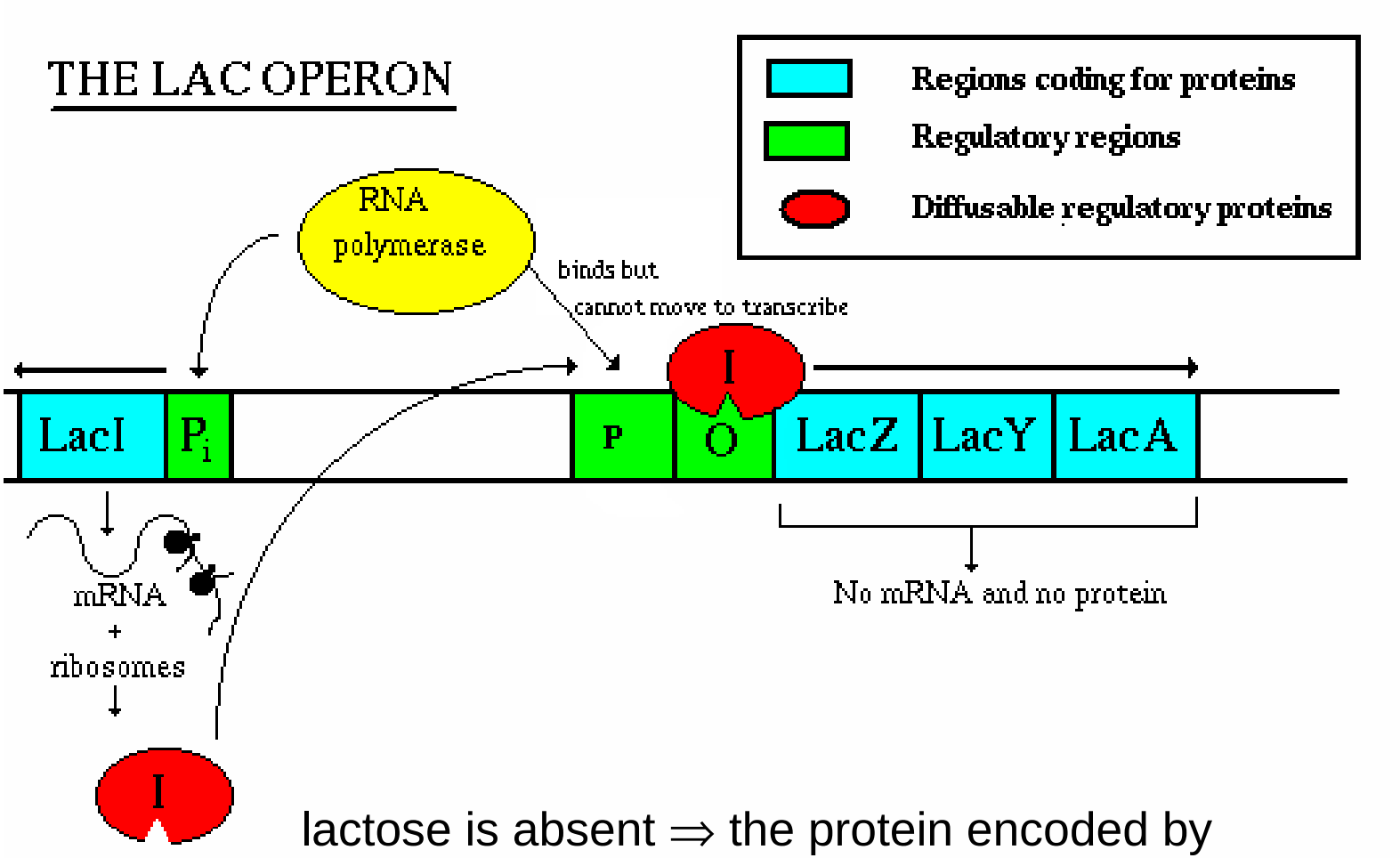
Gene regulation example: the lac operon



this protein regulates the transcription of LacZ, LacY, LacA

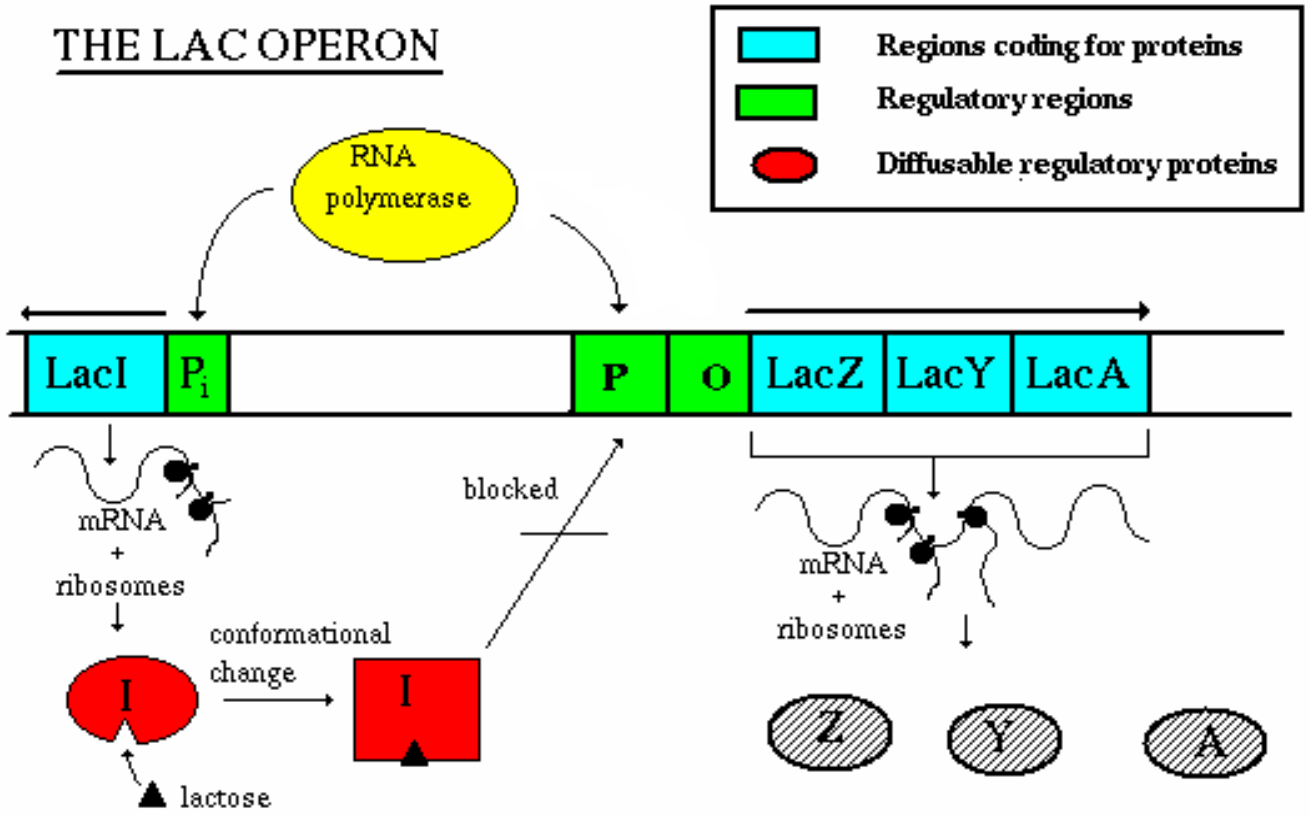
these proteins metabolize lactose

Gene regulation example: the lac operon



lactose is absent ⇒ the protein encoded by lacI represses transcription of the lac operon

Gene regulation example: the lac operon



lactose is present ⇒ it binds to the protein encoded by lacI changing its shape; in this state, the protein doesn't bind upstream ofthe lac operon; therefore the lac operon can be transcribed

Gene regulation example: the lac operon

- this example provides a simple illustration of how a cell can regulate (turn on/off) certain genes in response to the state of its environment
- an *operon* is a sequence of genes transcribed as a unit
- the lac operon is involved in metabolizing lactose
- it is “turned on” when lactose is present in the cell
- the lac operon is regulated at the transcription level

- the depiction here is incomplete; for example, the level of glucose in the cell also influences transcription of the lac operon

Questions