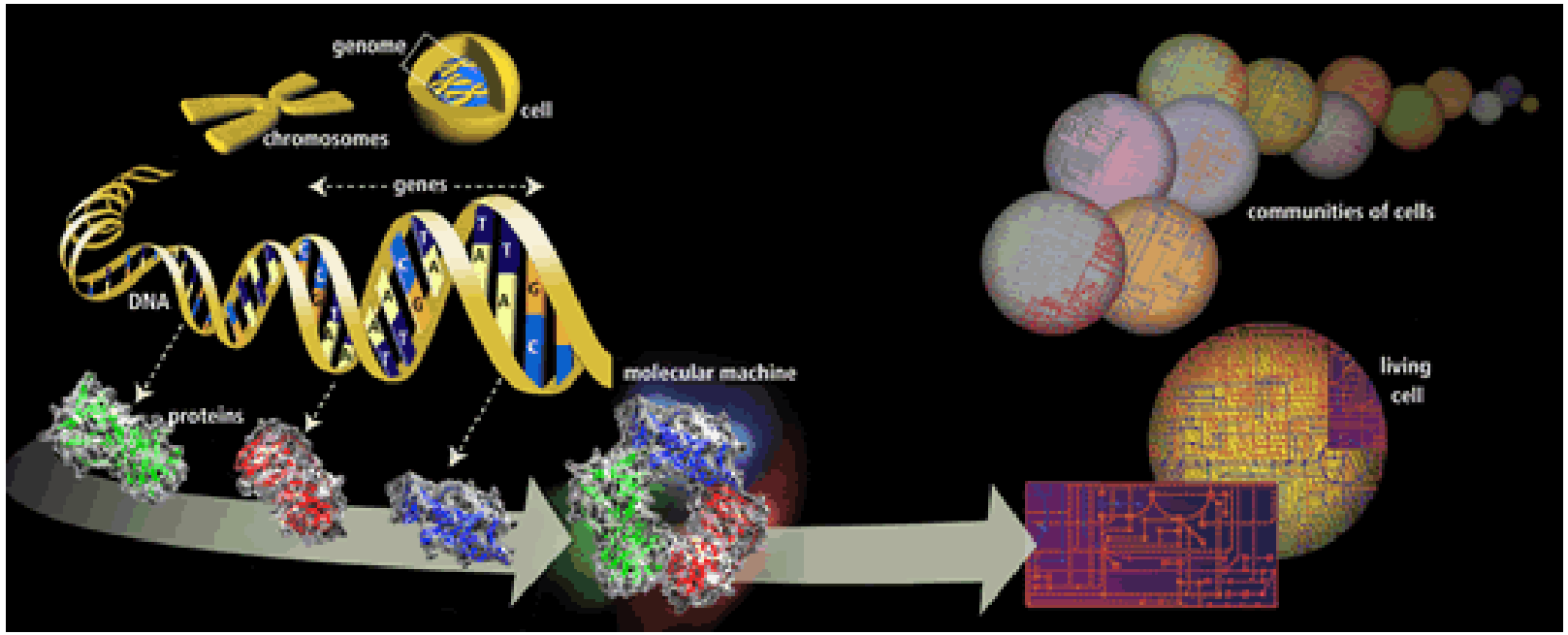
EECS730: Introduction to Bioinformatics

Lecture 1: Molecular Biology Primer



Angela Brooks, Raymond Brown, Calvin Chen, Mike Daly, Hoa Dinh, Erinn Hama, Robert Hinman, Julio Ng, Michael Sneddon, Hoa Troung, Jerry Wang, Che Fung Yung

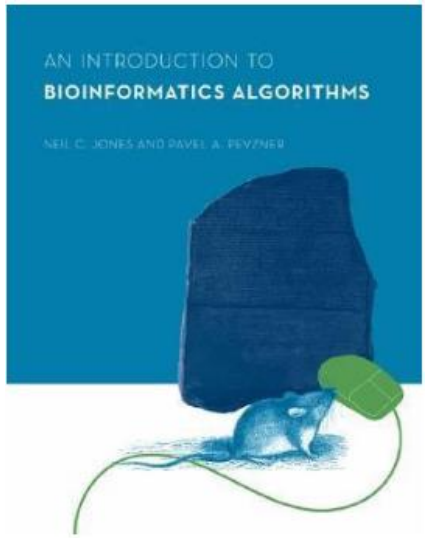
Slides adapted from Dr. Shaojie Zhang (University of Central Florida)

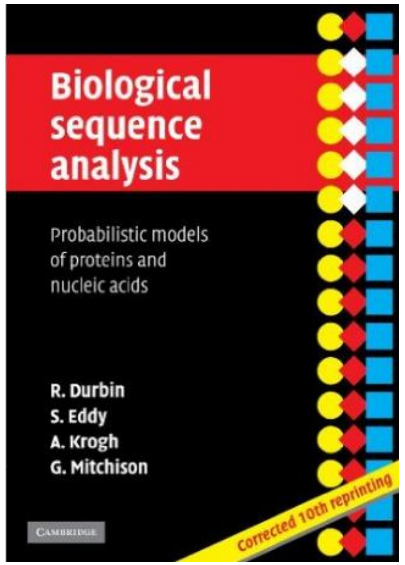
- KUMC visit

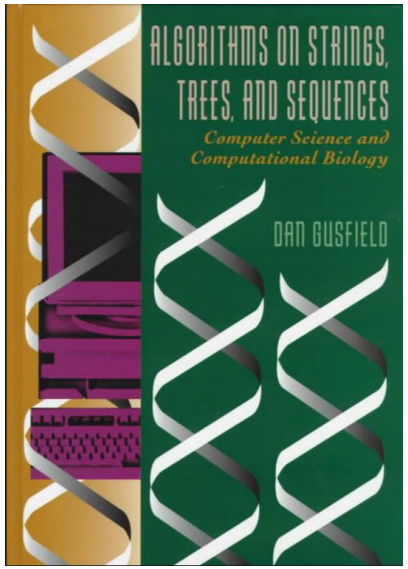
EECS 730 Introduction to Bioinformatics

Instructor: Cuncong Zhong
Office: 2026 Eaton Hall
Class meeting: TR 11:00-12:15 LEA 3153
Email: cczhong@ku.edu
Webpage: http://ittc.ku.edu/~cczhong/teaching.html
Office hours: Tuesday Thursday 9:00AM-10:00AM

- Books



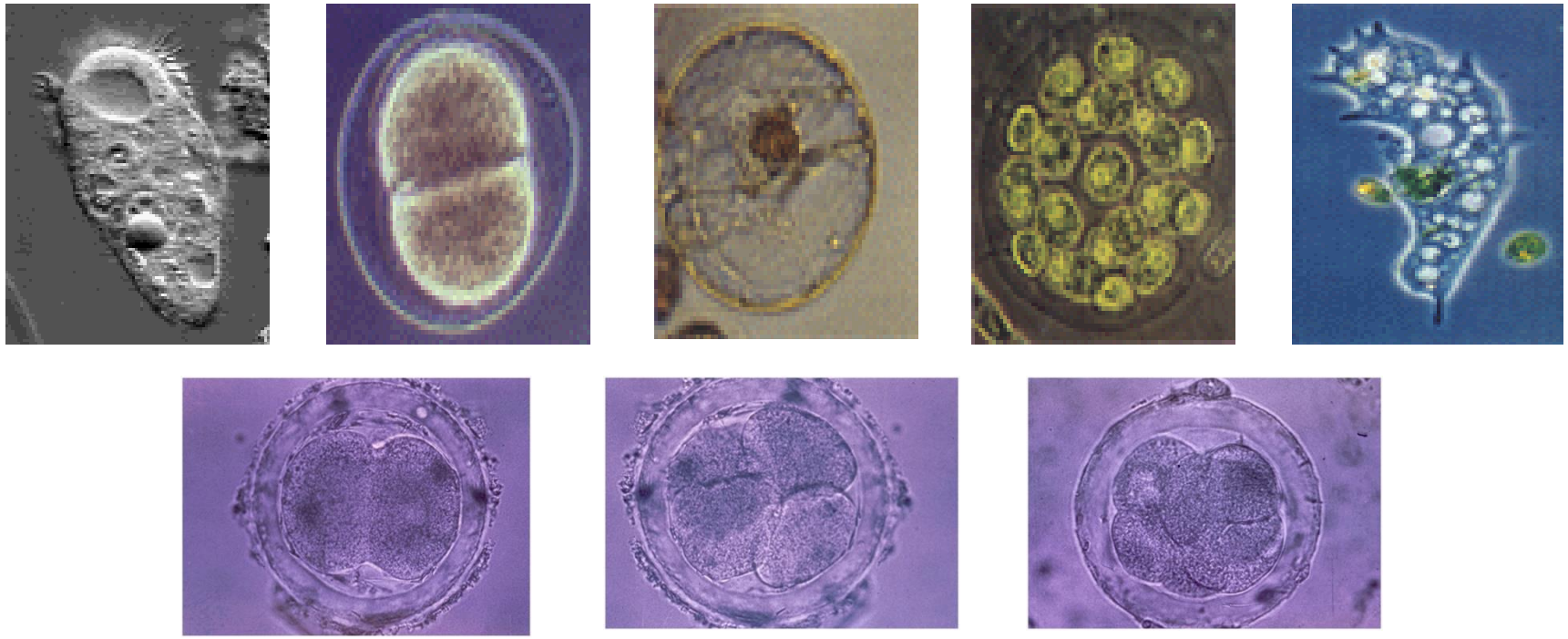




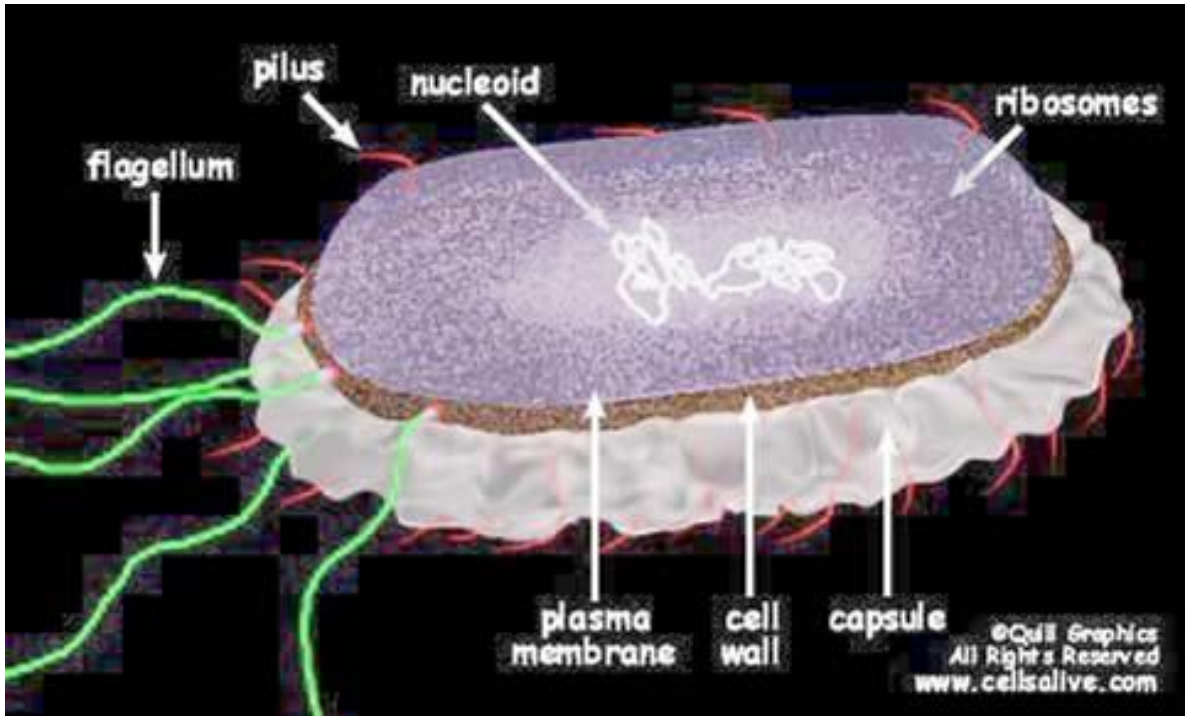
Outline

- Cells
- DNAs
- Transcription and RNAs
- Translation and proteins
- Comparative genomics

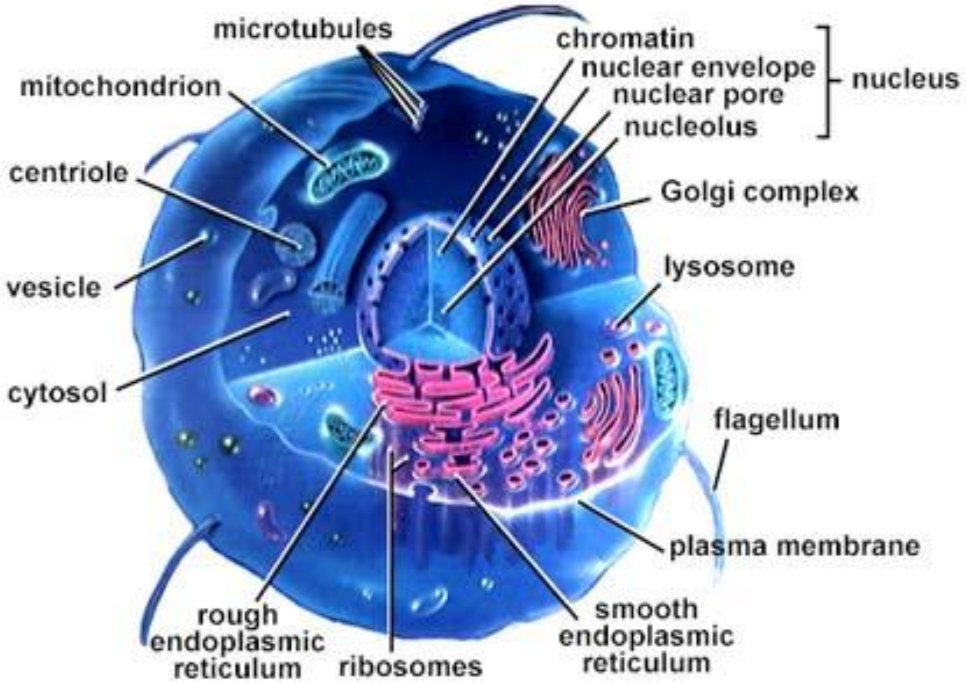
What is life made of?



Prokaryotes and Eukaryotes



Prokaryotic cell



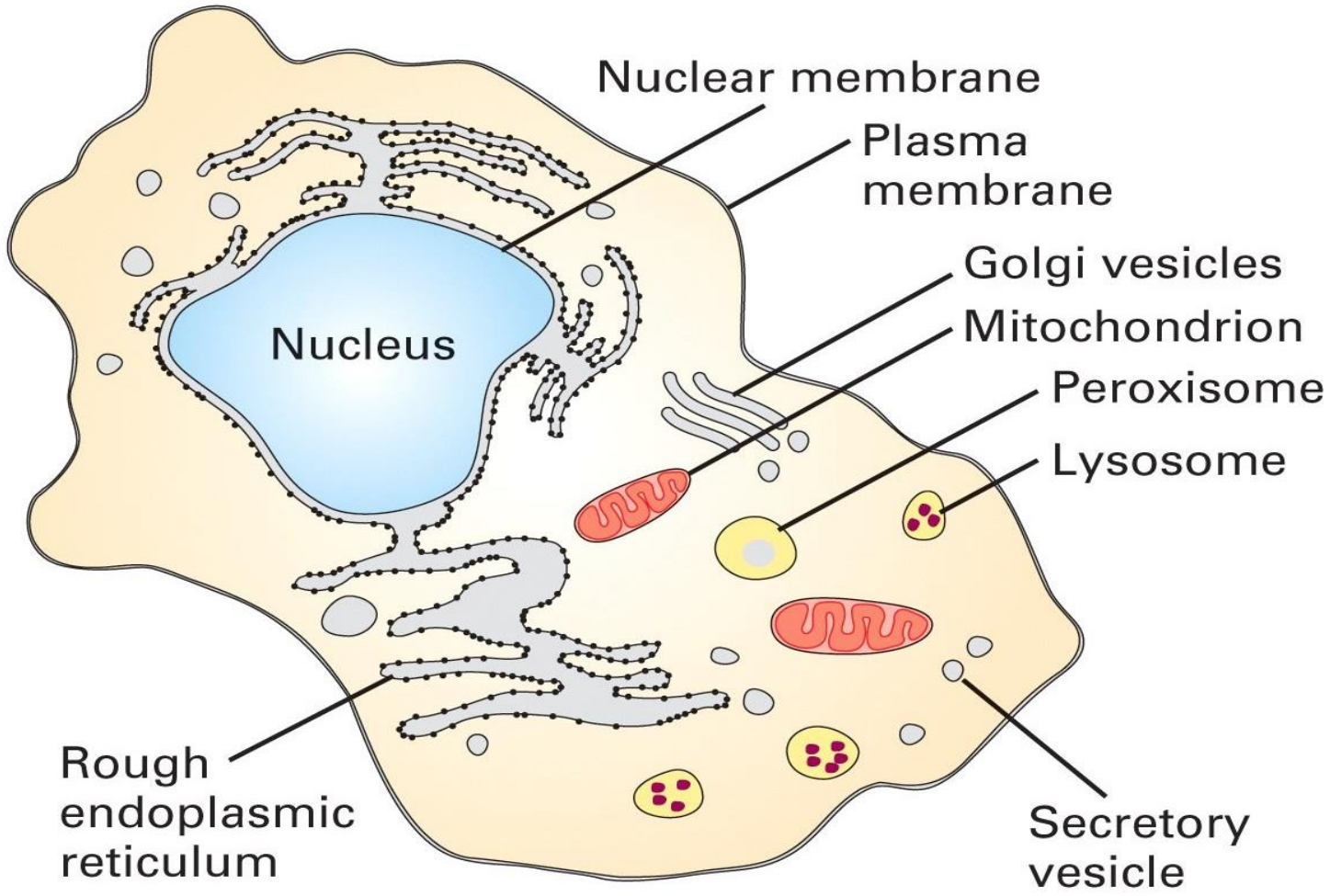
Eukaryotic cell

Prokaryotes and Eukaryotes cont.

Prokaryotes	**Eukaryotes**
Single cell	Single or multi cell
No nucleus	Nucleus
No organelles	Organelles
One piece of circular DNA	Chromosomes
No mRNA post transcriptional modification	Exons/Introns splicing

- A cell is a smallest structural unit of an organism that is capable of independent functioning
- All cells have some common features
- They Born, eat, replicate, and die

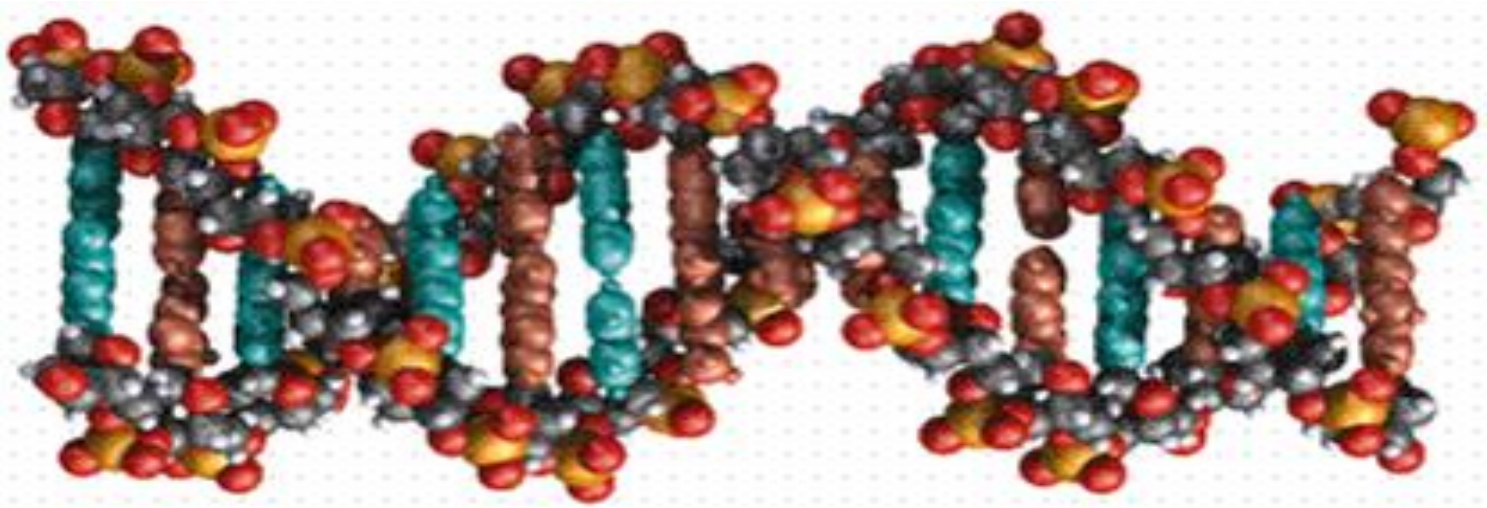
An eukaryotic cell



ORGANELLE	LOCATION	DESCRIPTION	FUNCTION
cell wall	plant, not animal	*outer layer *rigid, strong, stiff *made of cellulose	*support (grow tall) *protection *allows H2O, O2, CO2 to pass into and out of cell
cell membrane	both plant/animal	*plant - inside cell wall *animal - outer layer; cholesterol *selectively permeable	*support *protection *controls movement of materials in/out of cell *barrier between cell and its environment *maintains homeostasis
nucleus	both plant/animal	*large, oval	*controls cell activities
nuclear membrane	both plant/animal	*surrounds nucleus *selectively permeable	*Controls movement of materials in/out of nucleus
cytoplasm	both plant/animal	*clear, thick, jellylike material and organelles found inside cell membrane	*supports /protects cell organelles
endoplasmic reticulum (E.R.)	both plant/animal	*network of tubes or membranes	*carries materials through cell
ribosome	both plant/animal	*small bodies free or attached to E.R.	*produces proteins
mitochondrion	both plant/animal	*bean-shaped with inner membranes	*breaks down sugar molecules into energy
vacuole	plant - few/large animal - small	*fluid-filled sacs	*store food, water, waste (plants need to store large amounts of food)
lysosome	plant - uncommon animal - common	*small, round, with a membrane	*breaks down larger food molecules into smaller molecules *digests old cell parts

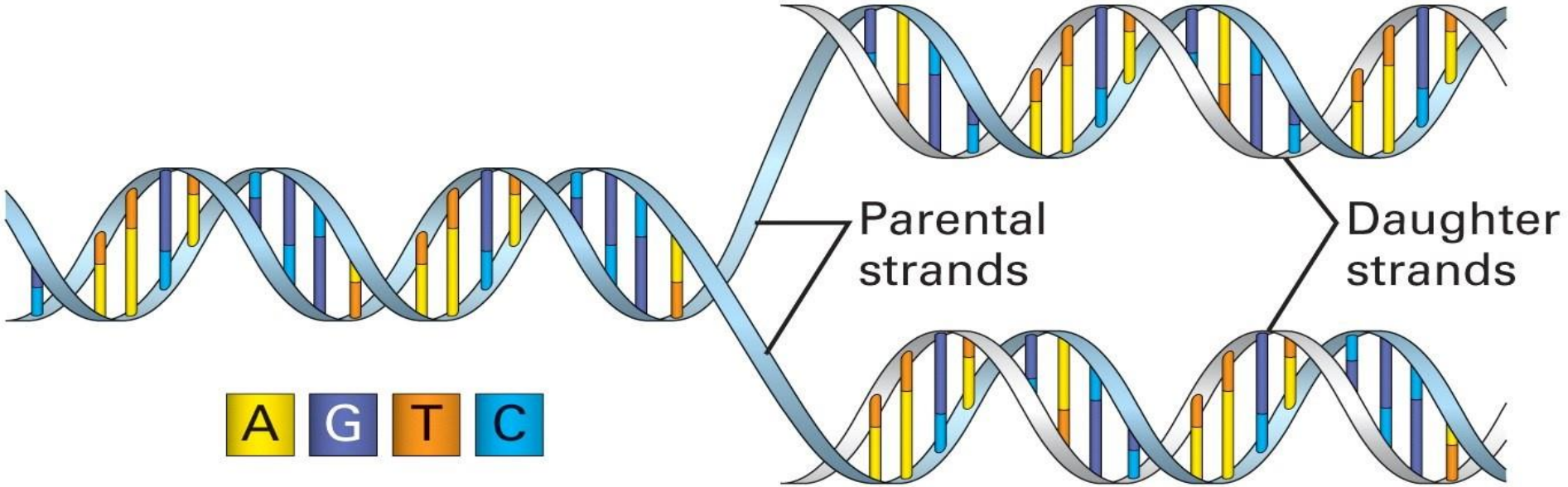
http://utahscience.oremjr.alpine.k12.ut.us/sciber00/7th/cells/sciber/orgtable.htm

Genetic material of life



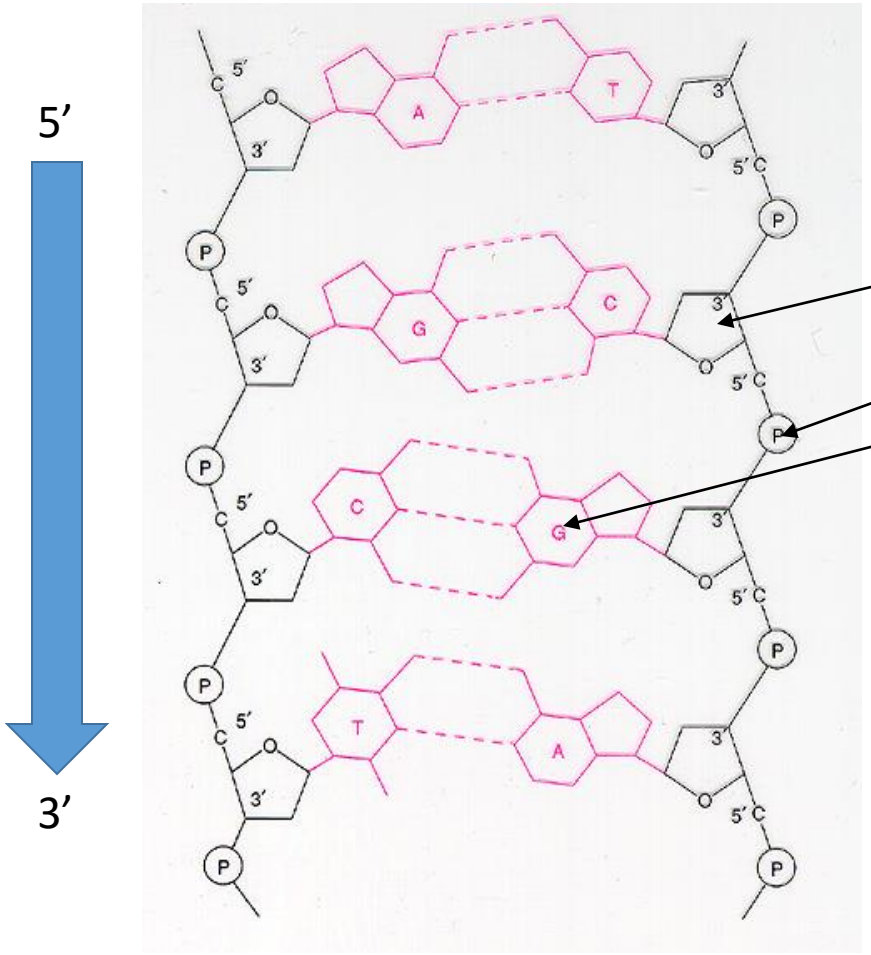
- What we inherit from our parents and what we pass down to our children (why we look like our parents?)
- The “blueprint” of life; easier to pass compared to the whole “building”

DNA: the code of life



- The structure and the four genomic letters code for all living organisms
- Adenine (A), Guanine (G), Thymine (T), and Cytosine (C) which pair A-T and C-G on complimentary strands.

DNA cont.

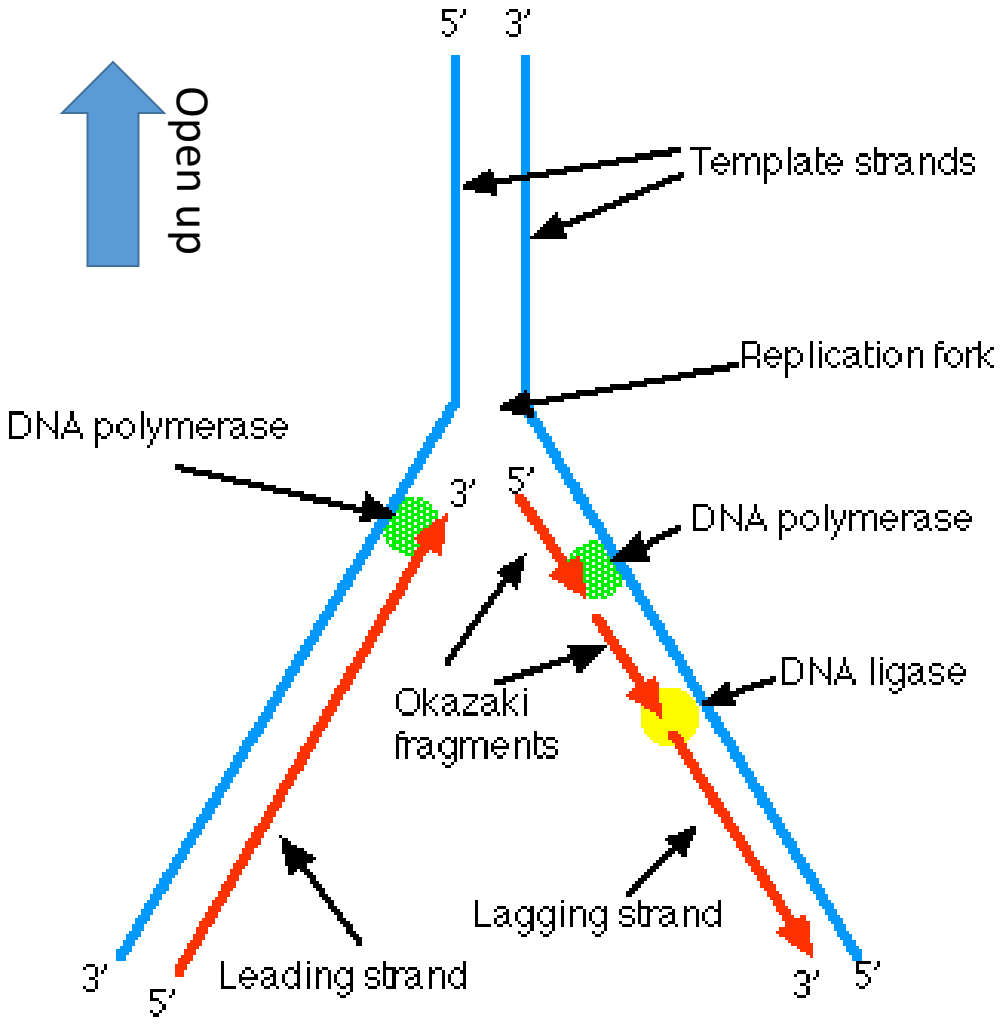


- DNA has a double helix structure which composed of
  - sugar molecule
  - phosphate group
  - and a base (A,C,G,T)

- DNA always reads from 5’ end to 3’ end for transcription replication
  - 5’ ATTTAGGCC 3’
  - 3’ TAAATCCGG 5’

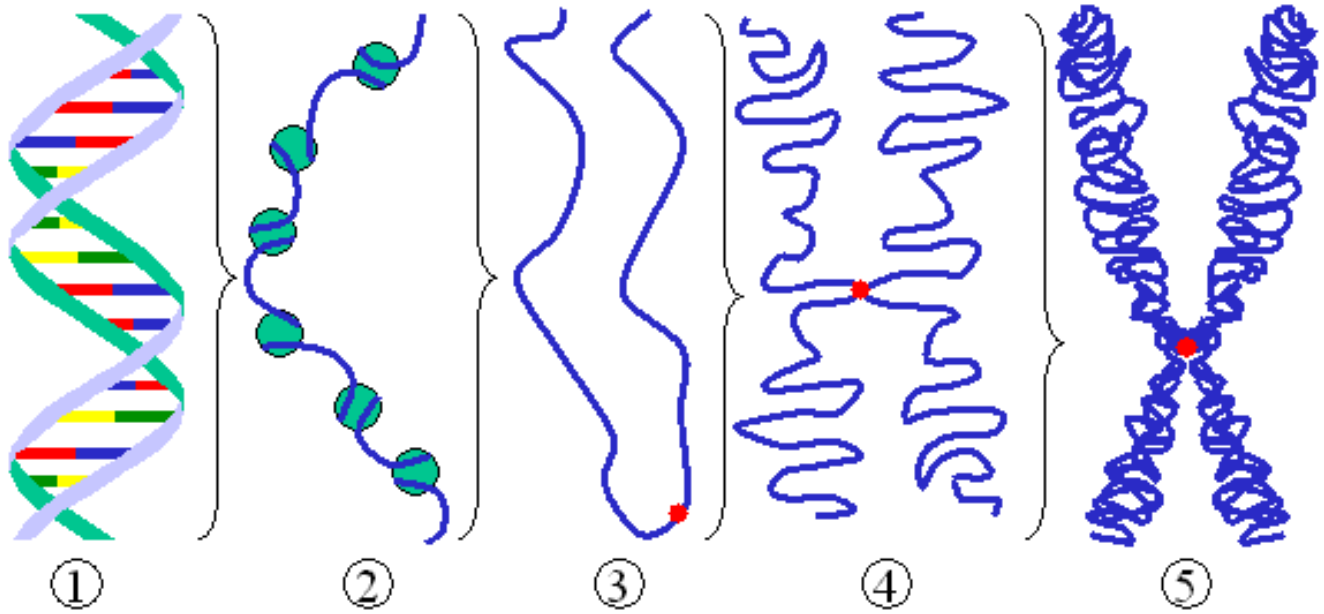
DNA replication

- DNA can replicate by splitting, and rebuilding each strand.
- Note that the rebuilding of each strand uses slightly different mechanisms due to the 5’ 3’ asymmetry, but each daughter strand is an exact replica of the original strand.

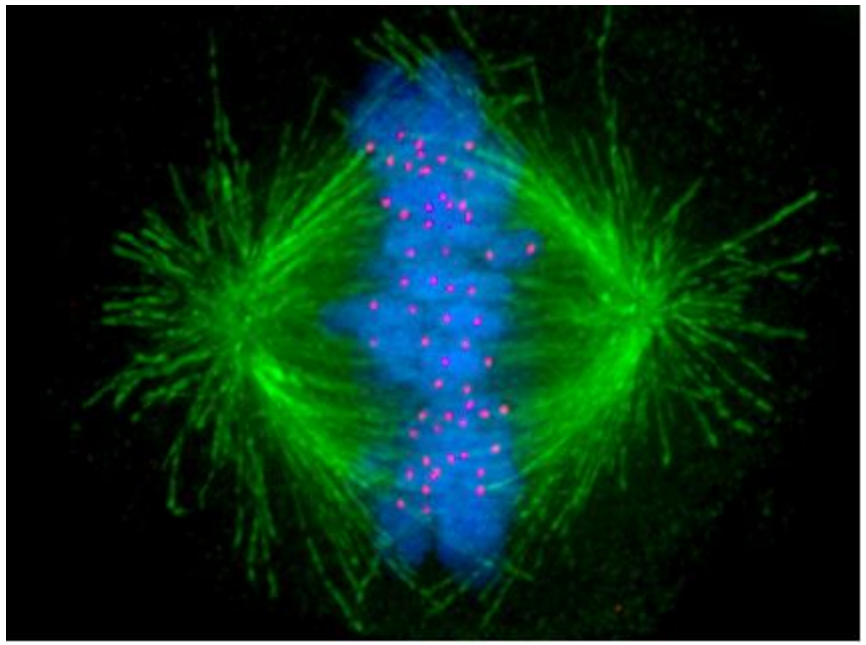


http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNAReplication.html

Packed DNA: Chromosomes



- (1) Double helix DNA strand.
- (2) Chromatin strand (DNA with histones)
- (3) Condensed chromatin during interphase with centromere.
- (4) Condensed chromatin during prophase
- (5) Chromosome during metaphase



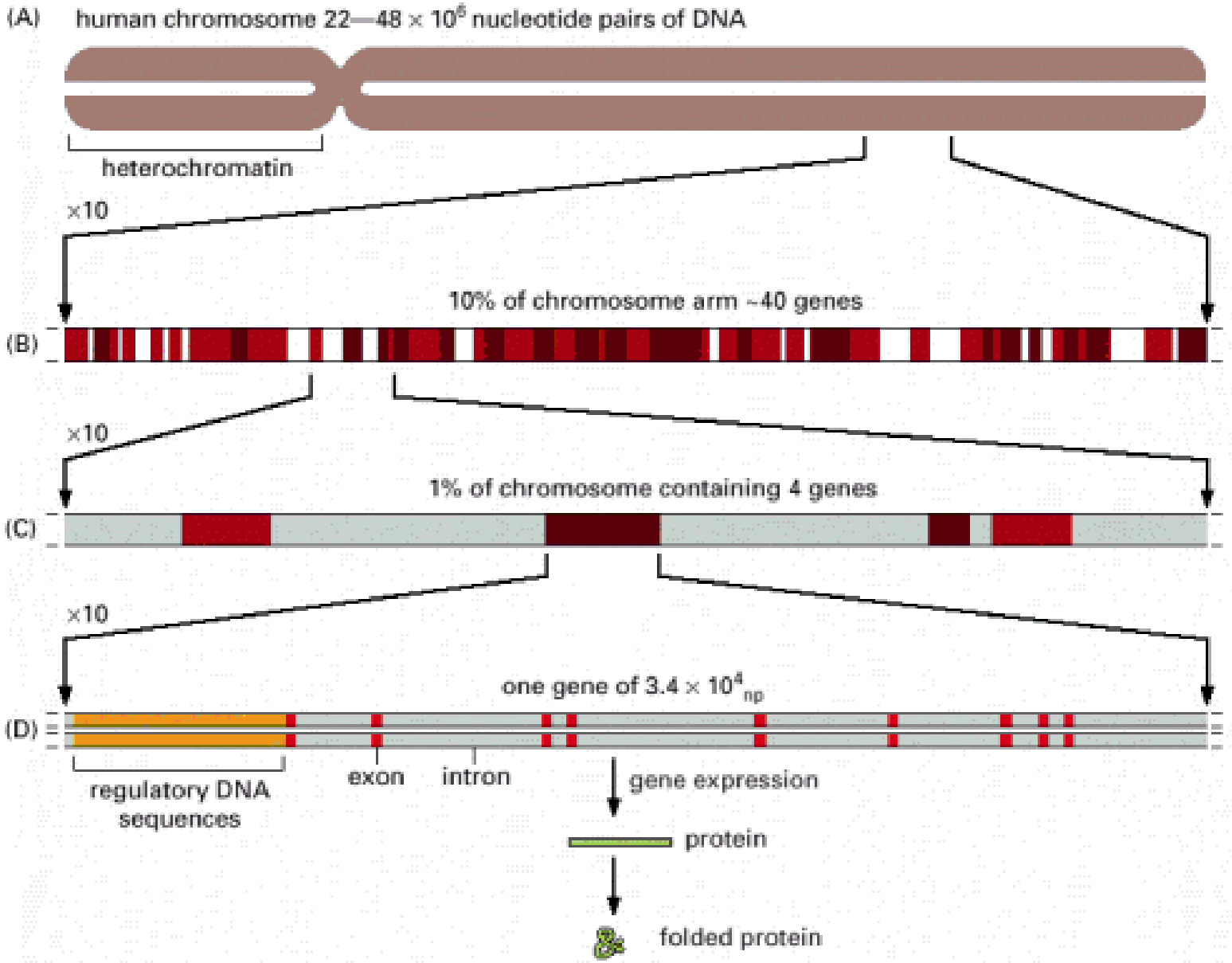
spindle fibers attach to the centromere via the kinetochore during mitosis

https://en.wikipedia.org/wiki/Centromere

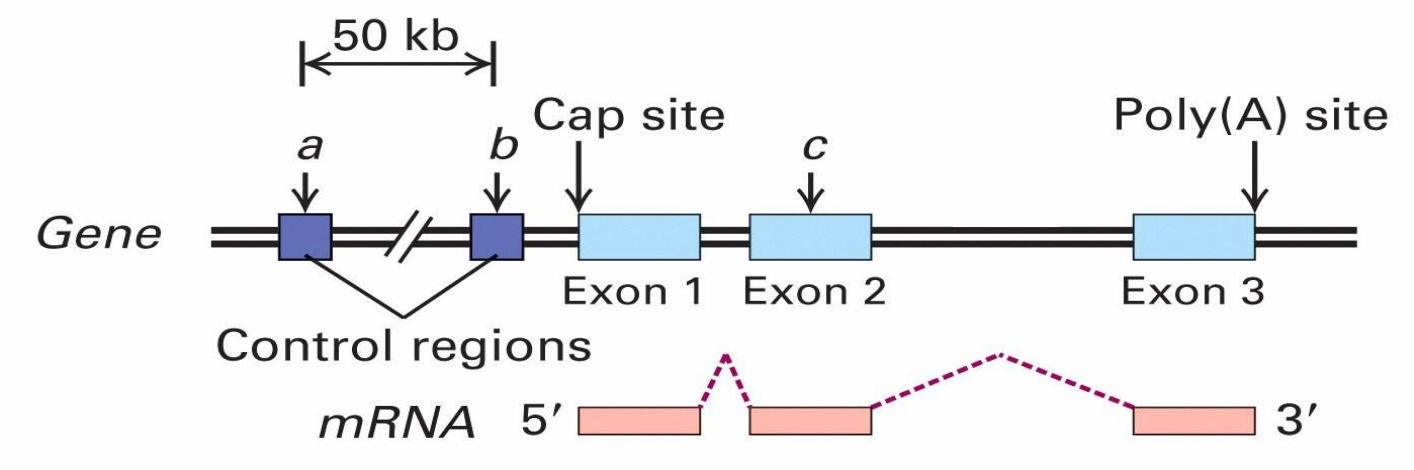
Chromosome

Organism	Number of base pair	number of Chromosomes
Prokayotic		
Escherichia coli (bacterium)	**$4x10^6$**	**1**
Eukaryotic		
Saccharomyces cerevisiae(yeast)	**$1.35x10^7$**	**17**
Drosophila melanogaster(insect)	**$1.65x10^8$**	**4**
Homo sapiens(human)	**$2.9x10^9$**	**23**
Zea mays(corn)	**$5.0x10^9$**	**10**

The organization of genes on a human chromosome

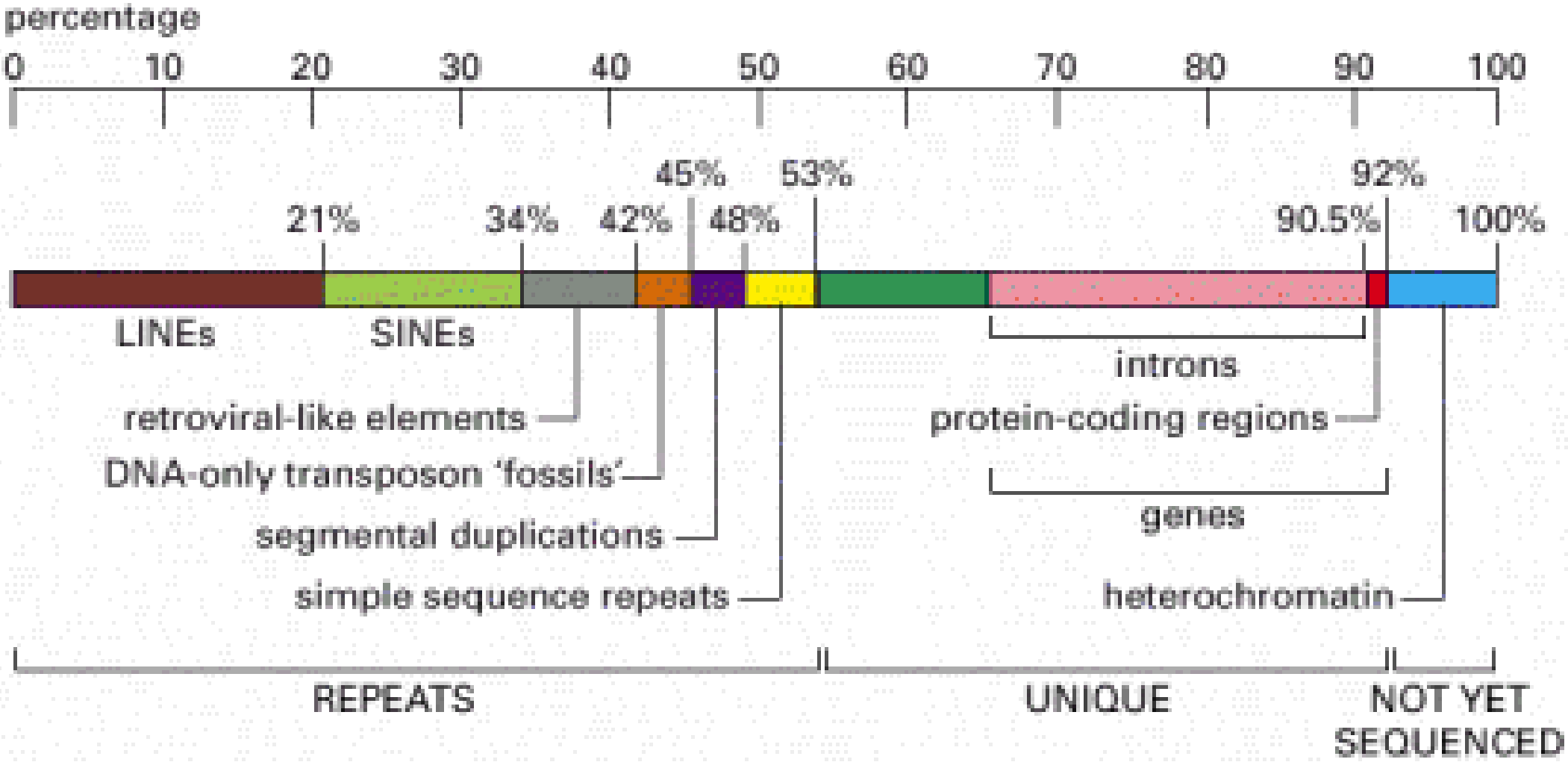


A closer look at the eukaryotic gene structure

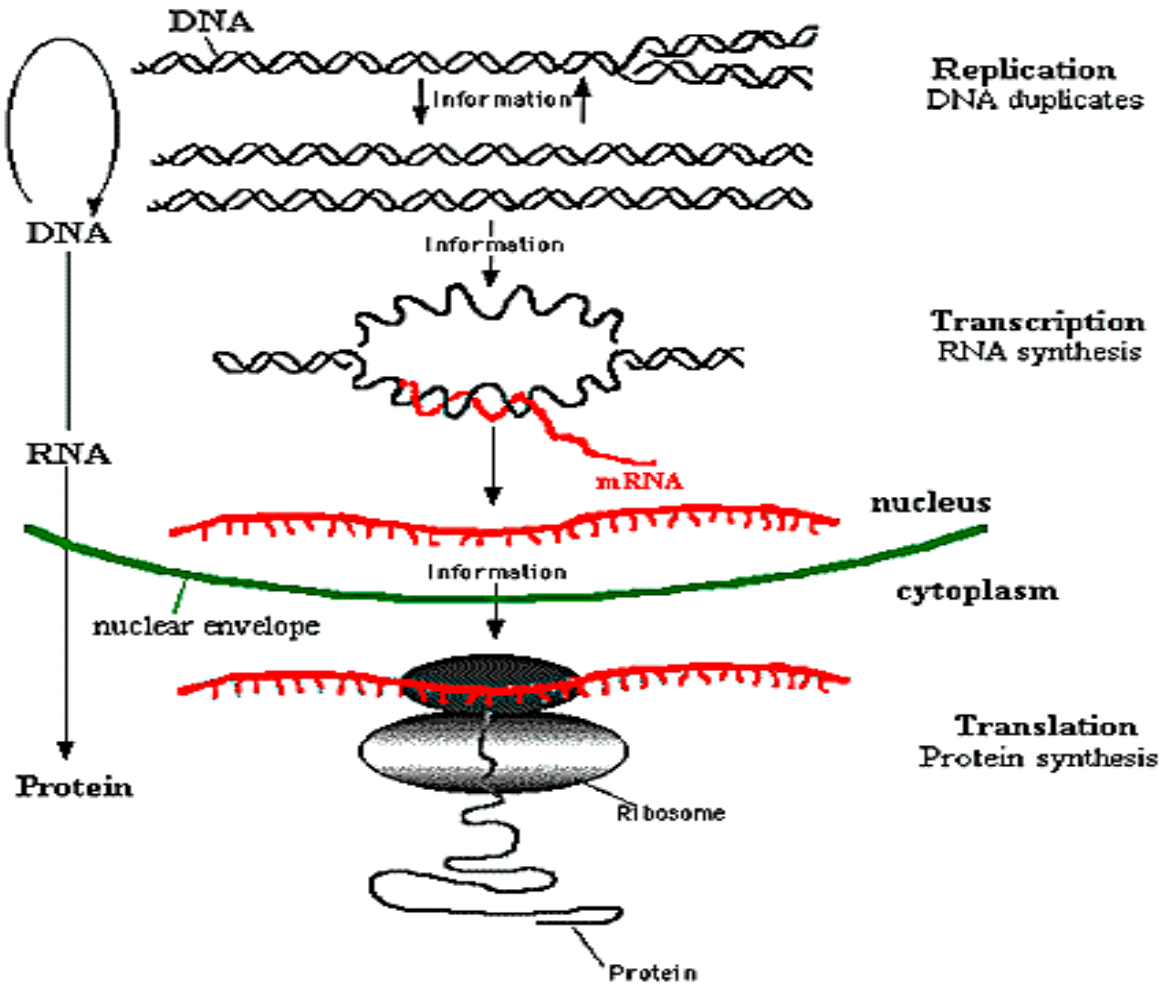


- **Regulatory regions: up to 50 kb upstream of +1 site**
- **Exons: protein coding and untranslated regions (UTR)**
  1 to 178 exons per gene (mean 8.8)
  8 bp to 17 kb per exon (mean 145 bp)
- **Introns: splice acceptor and donor sites, junk DNA?**
  average 1 kb – 50 kb per intron
- **Gene size: Largest – 2.4 Mb (Dystrophin). Mean – 27 kb.**

The human genome



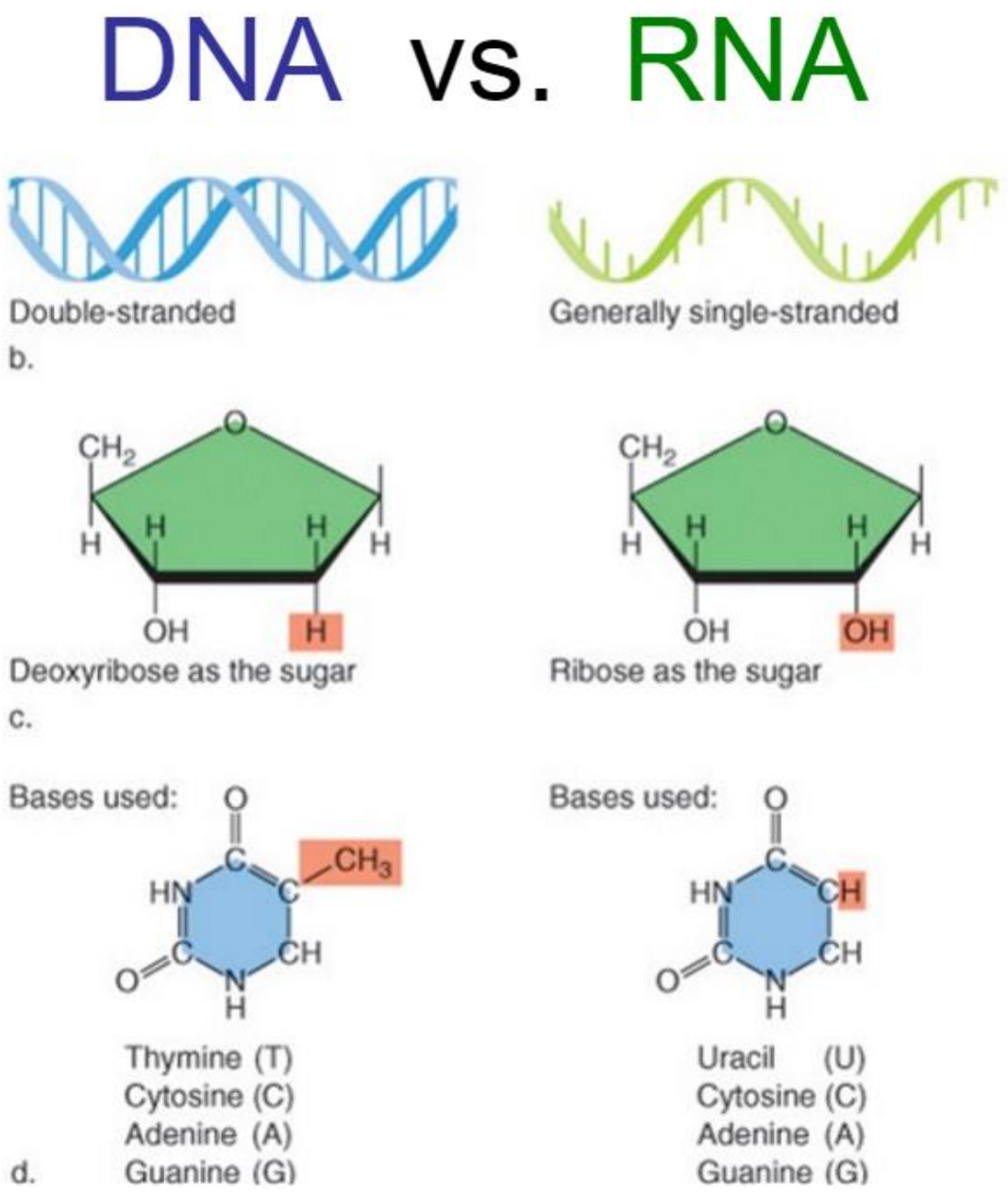
Central dogma



The Central Dogma of Molecular Biology

DNA → RNA: Transcription

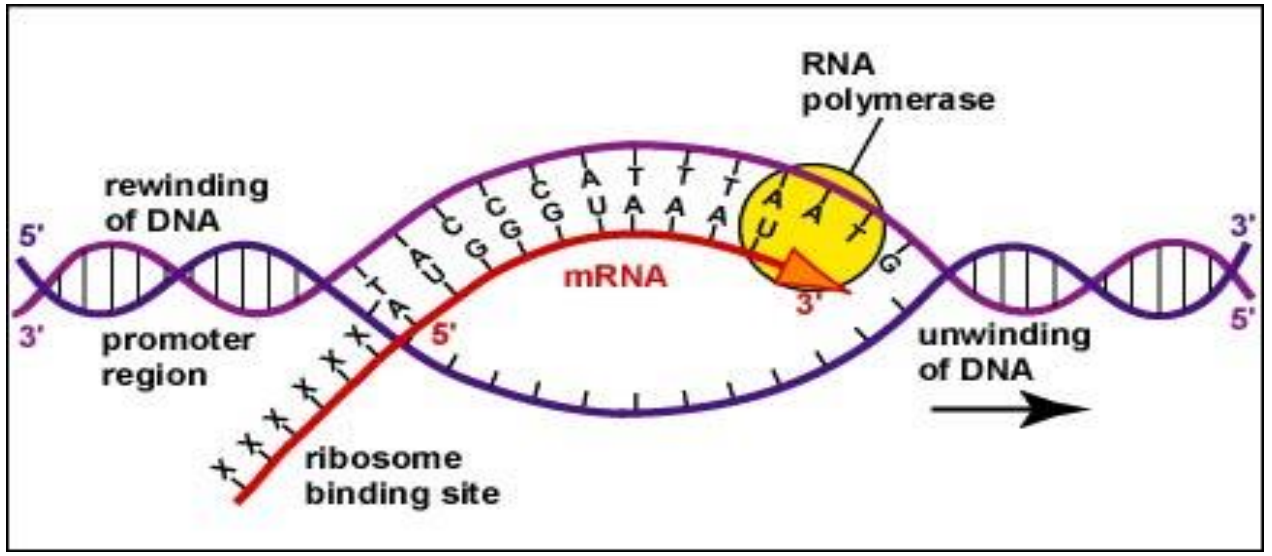
- DNA gets transcribed by a protein known as *RNA-polymerase*
- This process builds a chain of bases that will become mRNA (message RNA)
- RNA and DNA are similar, except that RNA is single stranded and thus less stable than DNA
  - Also, in RNA, the base uracil (U) is used instead of thymine (T), the DNA counterpart



http://slideplayer.com/slide/5670311/

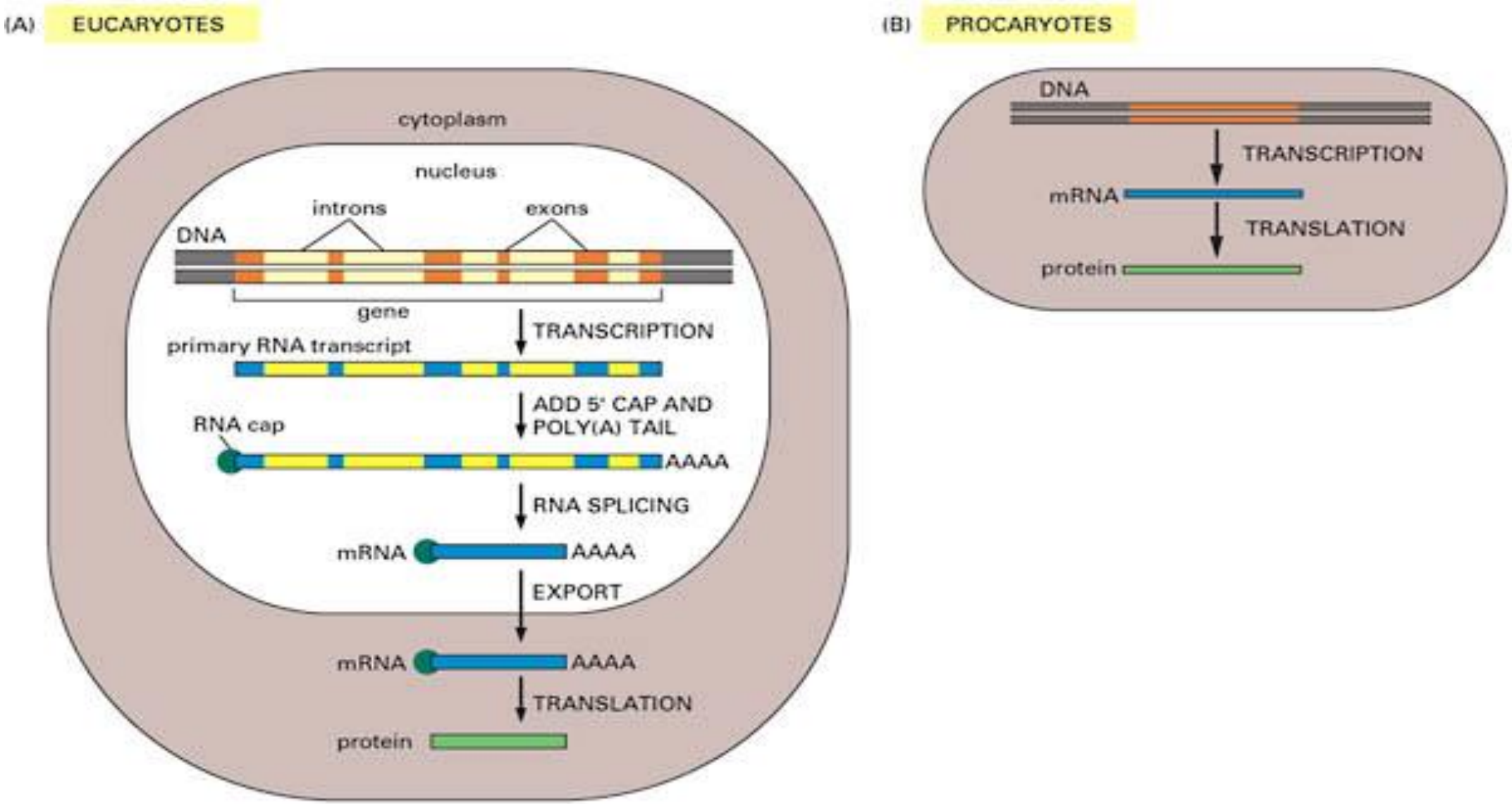
Transcription: DNA to pre-mRNA

- Transcription occurs in the nucleus.
- σ factor from RNA polymerase reads the promoter sequence and opens a small portion of the double helix exposing the DNA bases.

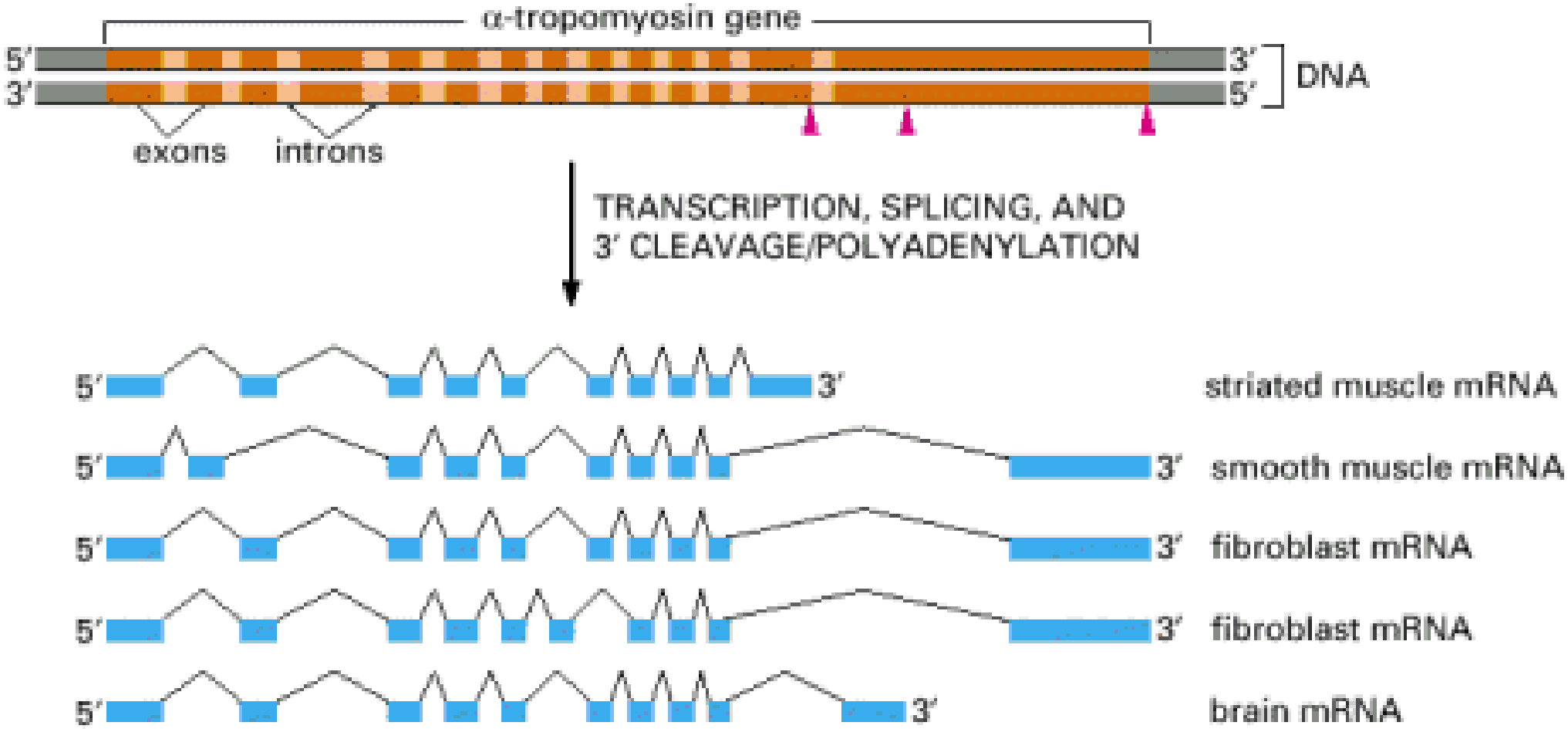


- RNA polymerase II catalyzes the formation of phosphodiester bond that link nucleotides together to form a linear chain from 5’ to 3’ by unwinding the helix just ahead of the active site for polymerization of complementary base pairs.
- The hydrolysis of high energy bonds of the substrates (nucleoside triphosphates ATP, CTP, GTP, and UTP) provides energy to drive the reaction.
- During transcription, the DNA helix reforms as RNA forms.
- When the terminator sequence is met, polymerase halts and releases both the DNA template and the RNA.

pre-mRNA to mature mRNA

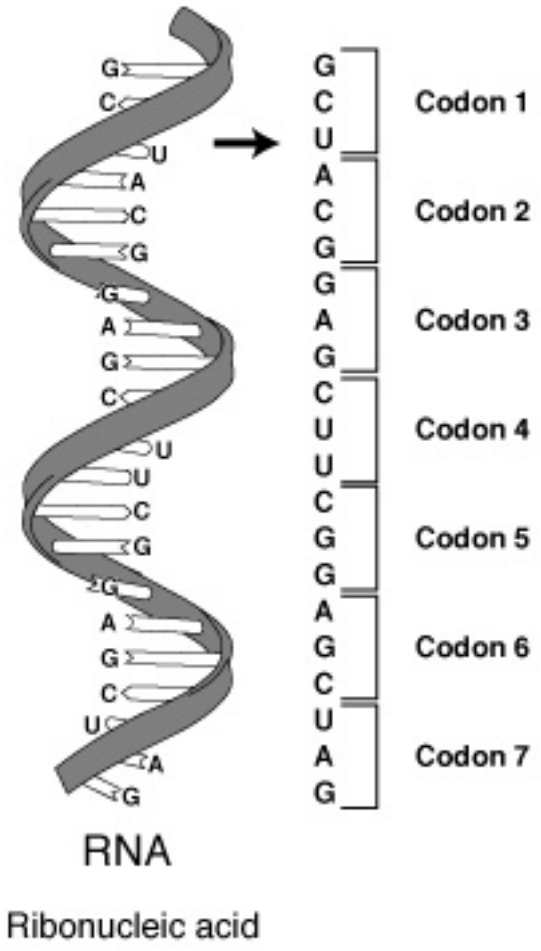


Alternative splicing



translation

- Happened in ribosome
- The process of going from RNA to polypeptide.
- Three base pairs of RNA (called a codon) correspond to one amino acid based on a fixed table.
- Always starts with Methionine and ends with a stop codon

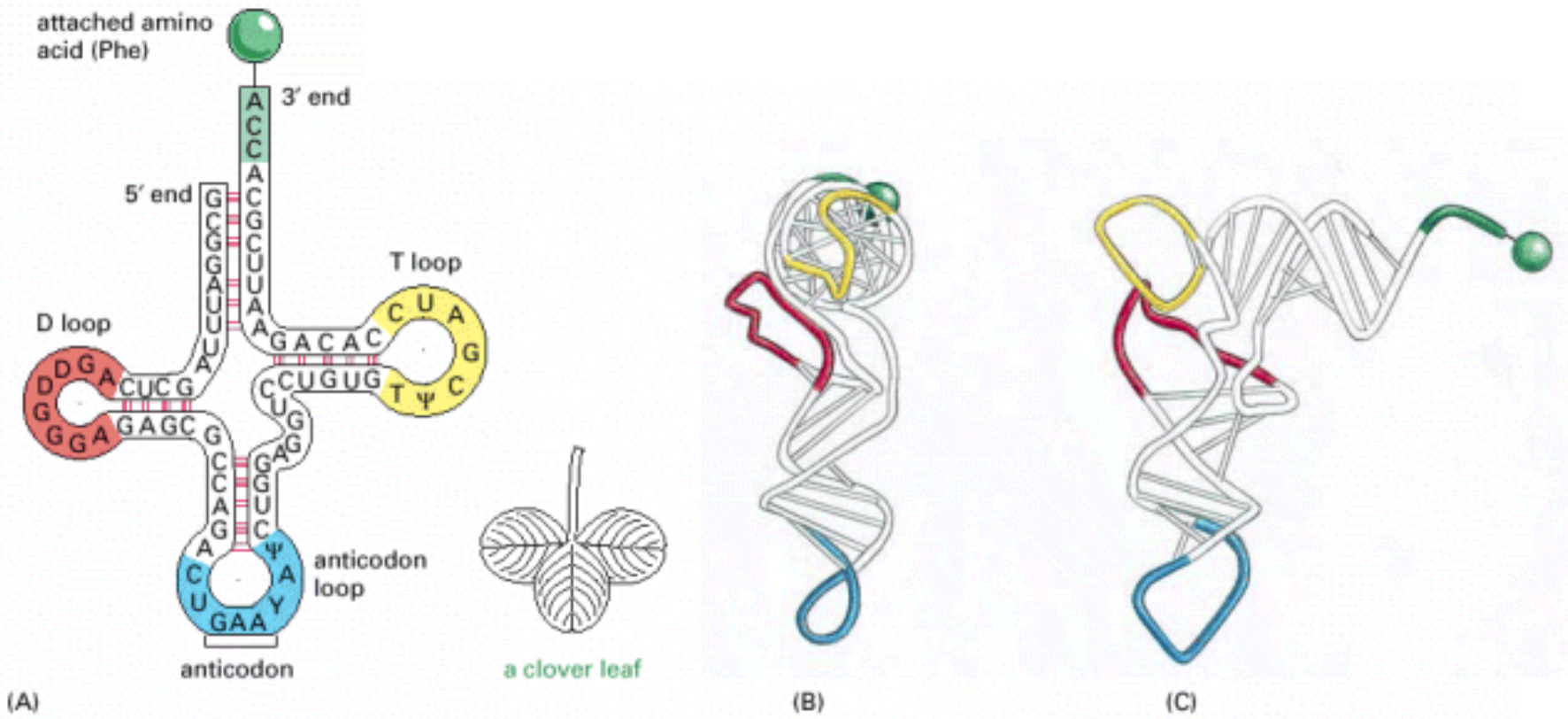


SECOND POSITION

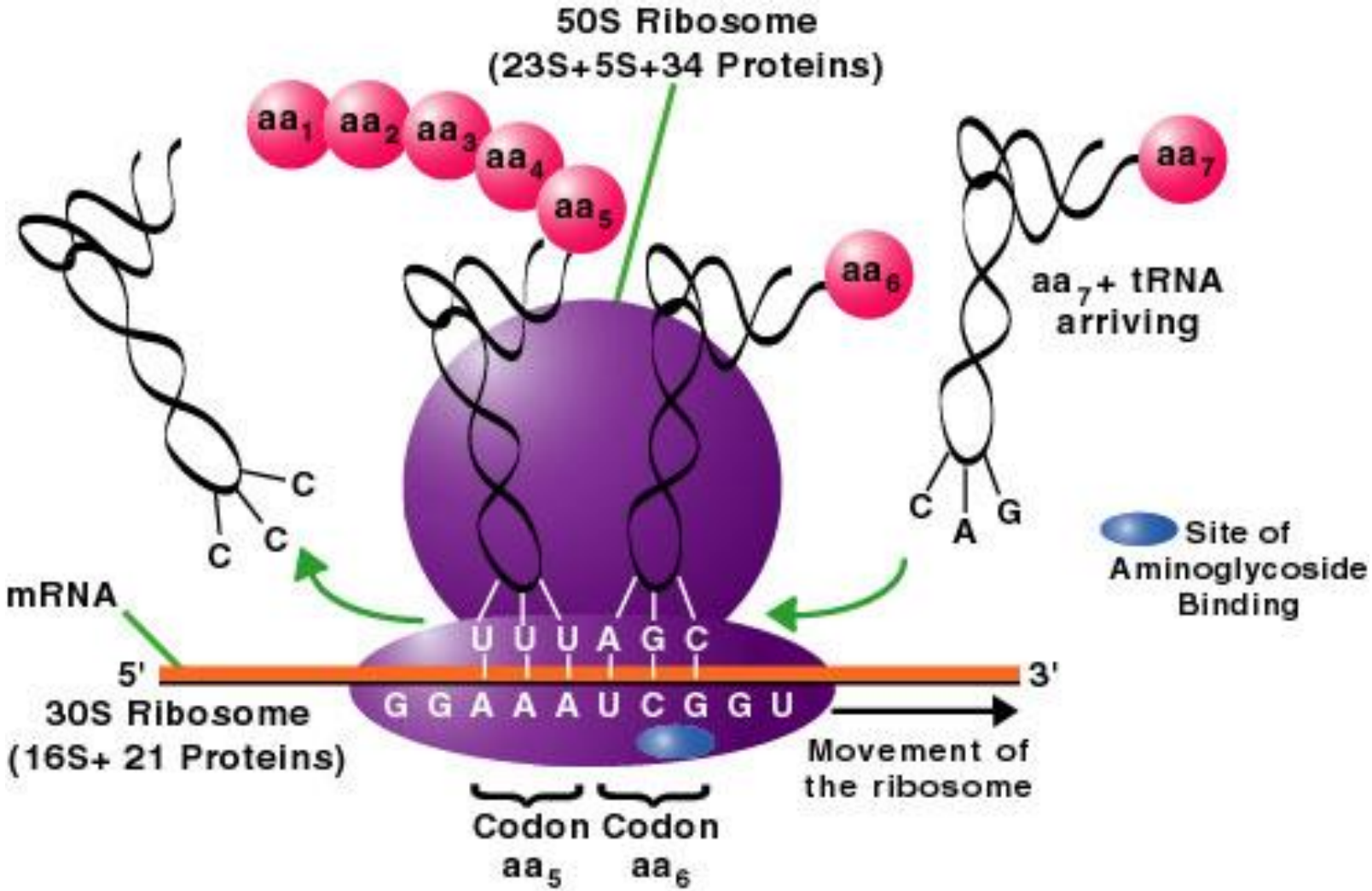
FIRST POSITION	U	C	A	G	THIRD POSITION
U	phenyl-alanine	serine	tyrosine	cysteine	U
					C
	leucine		stop	stop	A
			stop	tryptophan	G
C	leucine	proline	histidine	arginine	U
					C
			glutamine		A
					G
A	isoleucine	threonine	asparagine	serine	U
					C
			lysine	arginine	A
	* methionine				G
G	valine	alanine	aspartic acid	glycine	U
					C
			glutamic acid		A
					G

* and start

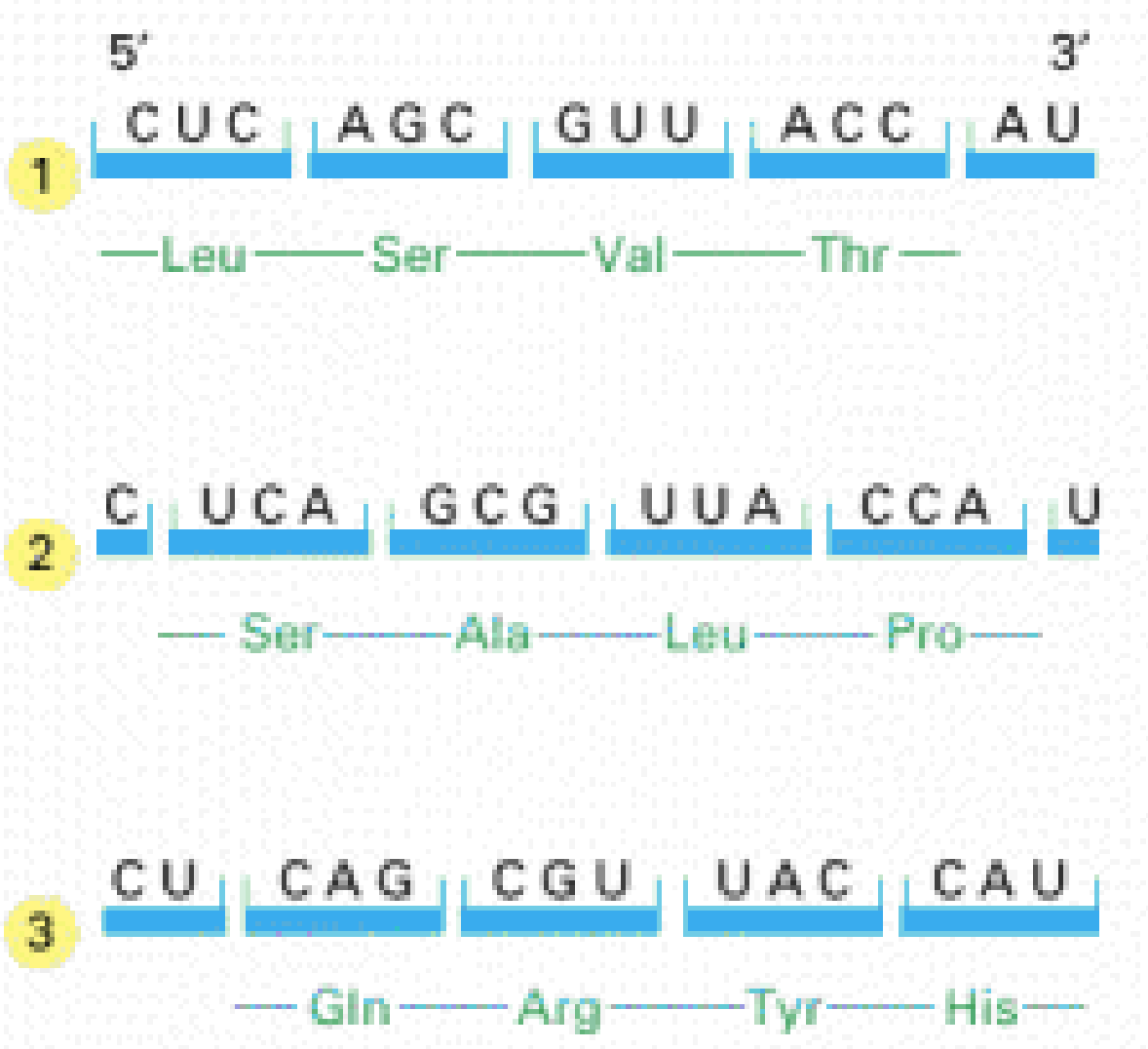
transfer RNA (tRNA)



Ribosome: the protein factory



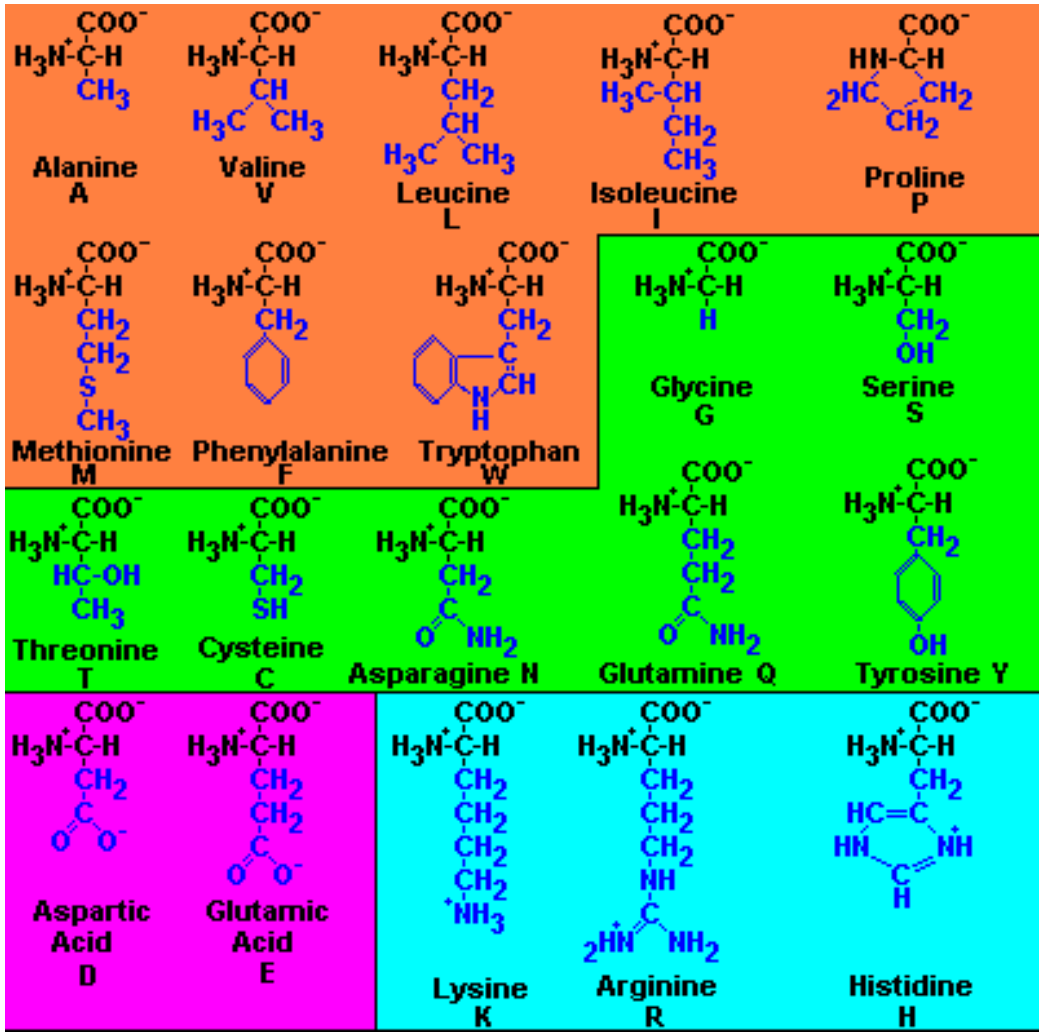
Open Reading frames (ORFs)

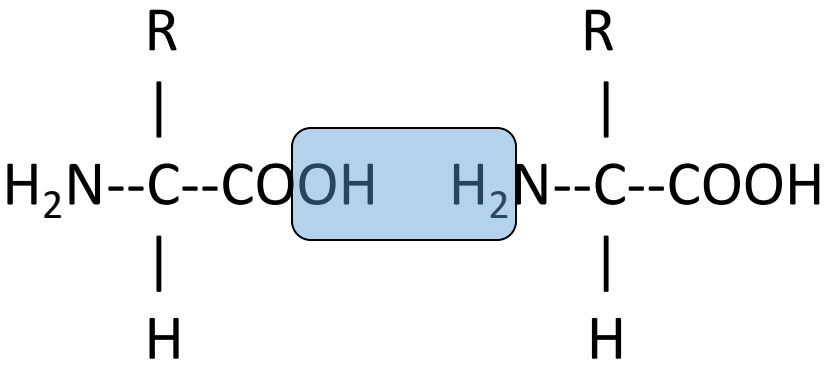


In fact we have 6 open reading frames!

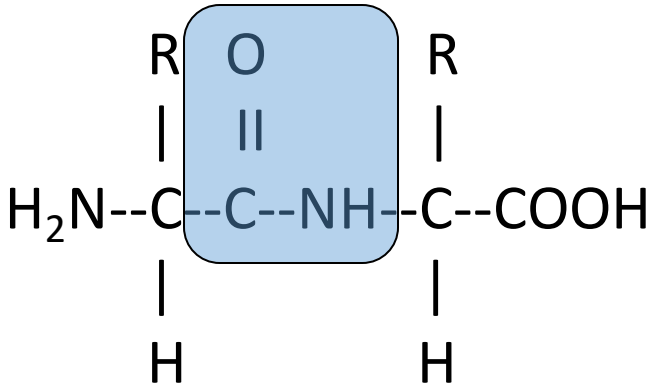
Remember the reverse complement!

Protein molecules



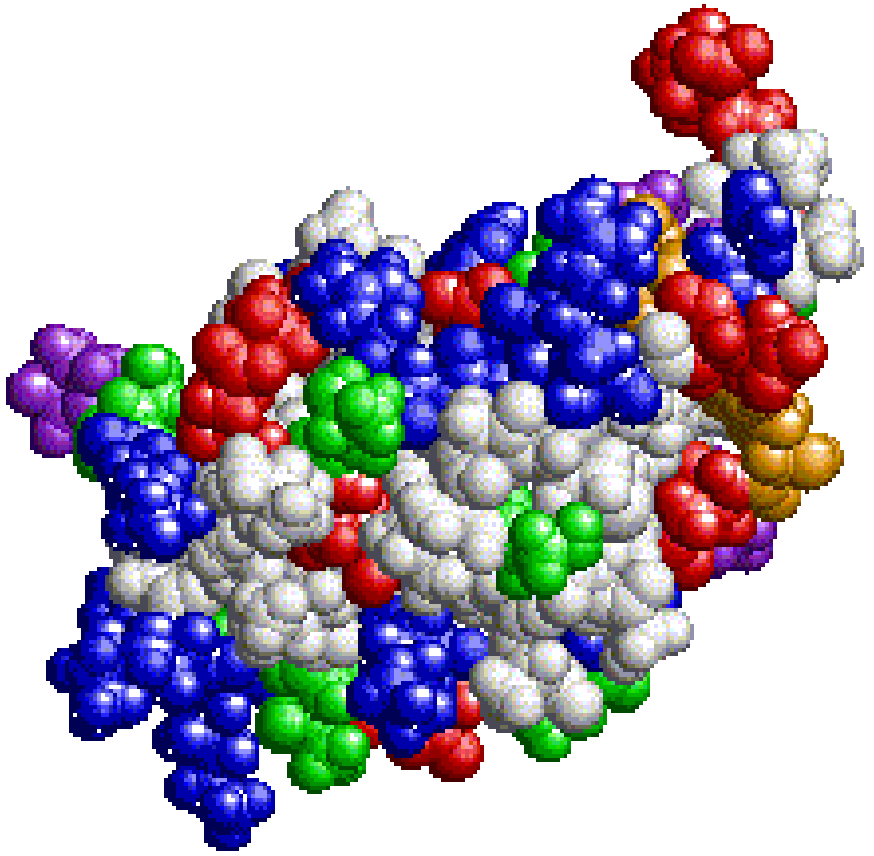


Peptide bond

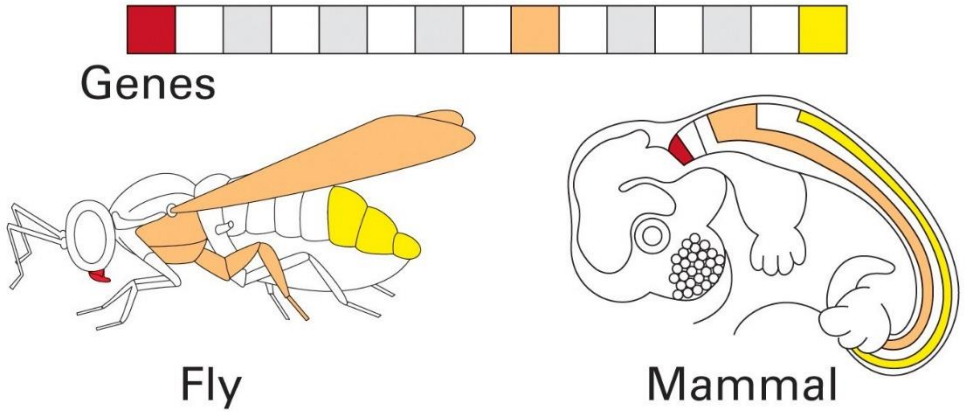


Protein structure-function relationship

- Linear sequence of amino acids folds to form a complex 3-D structure.
- The structure of a protein is intimately connected to its function.



DNA the Genetics Makeup



- Genes are inherited and are expressed
  - **genotype** (genetic makeup)
  - **phenotype** (physical expression)

- On the left, is the eye's phenotypes of green and black eye genes.

Comparative genomics (mammal vs mammal)

- Beta globin chains of closely related species are highly similar:
- Observe simple alignments below:

Human β chain: MVHLT**PE**EK**S**AV**TA**LWGKV N**V**D**E**VGGEALGRLL

Mouse β chain: MVHLT**DA**EK**A**AV**NG**LWGKVN**P**D**D**VGGEALGRLL

Human β chain: VVYPWTQR**F**F**E**SFGDLS**TPD**A**V**MGNPKVKAHGKKV**LG**

Mouse β chain: VVYPWTQR**Y**F**D**SFGDLS**SAS**A**I**MGNPKVKAHGKK V**IN**

Human β chain: AF**S**DGL**A**HLDNLKGTFA**T**LSELHCDKLHVDPENFRLLGN

Mouse β chain: AF**N**DGL**K**HLDNLKGTFA**H**LSELHCDKLHVDPENFRLLGN

Human β chain: **VL**V**C**VL**A**HH**F**GKEFTP**PV**QAA**Y**QKVVAGVA**N**ALAHKYH

Mouse β chain: **MI** V**I** VL**G**HH**L**GKEFTP**CA**QAA**F**QKVVAGVA**S**ALAHKYH

There are a total of 27 mismatches, or (147 – 27) / 147 = 81.7 % identical

Comparative genomics cont. (mammal vs aves)

Human β chain: MVH L TPEEKSAVTALWGKVNVDEVGGEALGRLL

Chicken β chain: MVHWTAEEKQL I TGLWGKVNVAECGAEALARLL

Human β chain: VVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG

Chicken β chain: IVYPWTQRFF ASFGNLSSPTA I LGNPMVRAHGKKVLT

Human β chain: AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGN

Chicken β chain: SFGDAVKNLDNIK NTFSQLSELHCDKLHVDPENFRLLGD

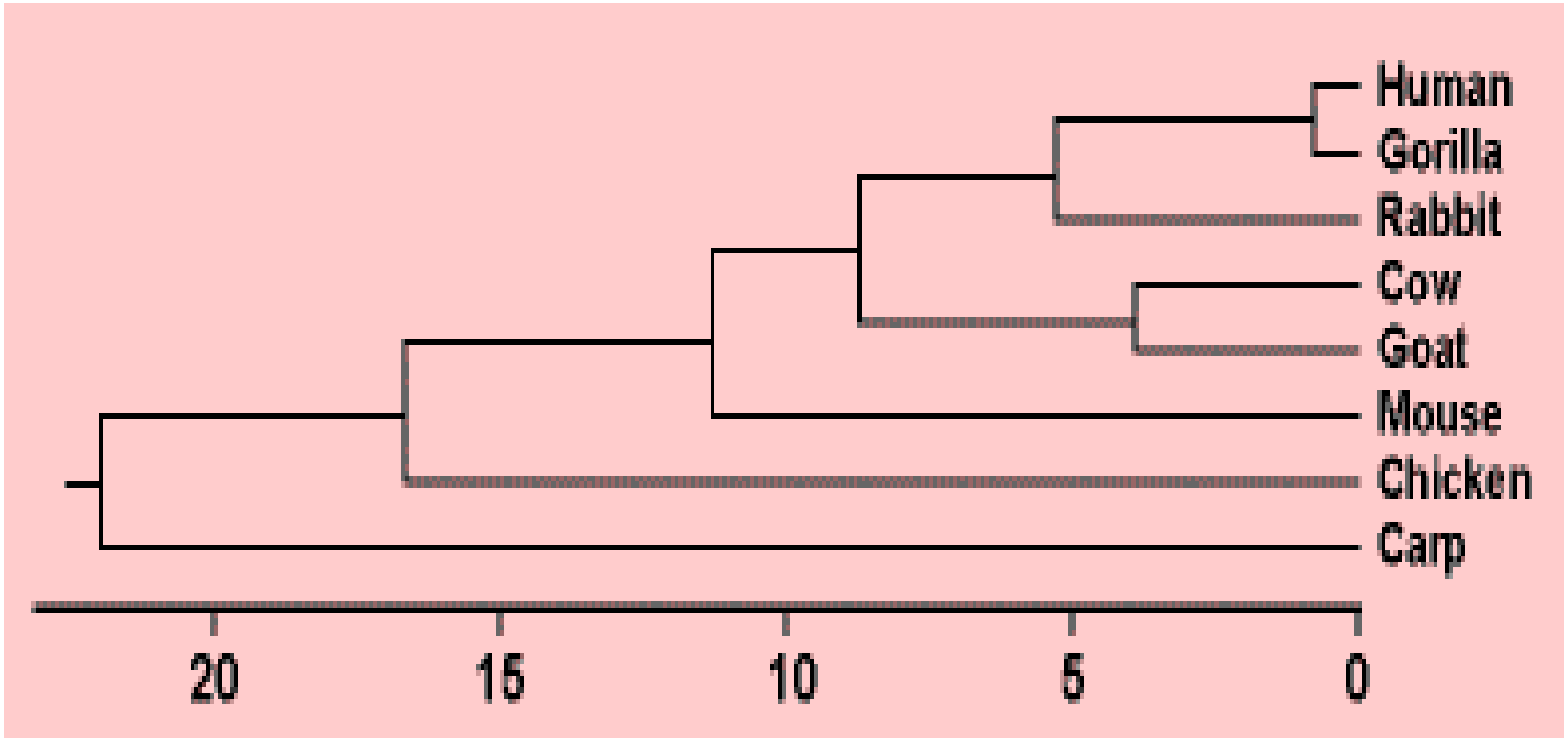
Human β chain: VLVCVLAHHFGKEFTPPVQAAY QKVVAGVANALAHKYH

Chicken β chain: I L I I VLAAHFSKDFTPECQAAWQKLVRVVAHALARKYH

-There are a total of 44 mismatches, or (147 – 44) / 147 = 70.1 % identical

- As expected, mouse β chain is *'closer'* to that of human than chicken's.

Molecular evolution



Phylogenetic tree of Beta globin (Aligned using Clustal, PAM250)

Comparative genomics

- **Which part of the gene/genome has been changed?**
- **How do we tie that to phenotypic change?**

- **Can we compare human and chimp genomes to find what gives us intelligence? Yes!!!**
- **Can we compare healthy human and diseased human to find what genomic changes lead to the disease? Yes!!!**

- **Well, we have to know how to compare sequences first!!!**