Introduction to Molecular Biology

Content

- Cells and organisms
- Molecules of life (Biomolecules)
- Central dogma of molecular biology
- Genes and gene expression

@: Most pictures have been freely obtained from:
http://www.accessexcellence.org/

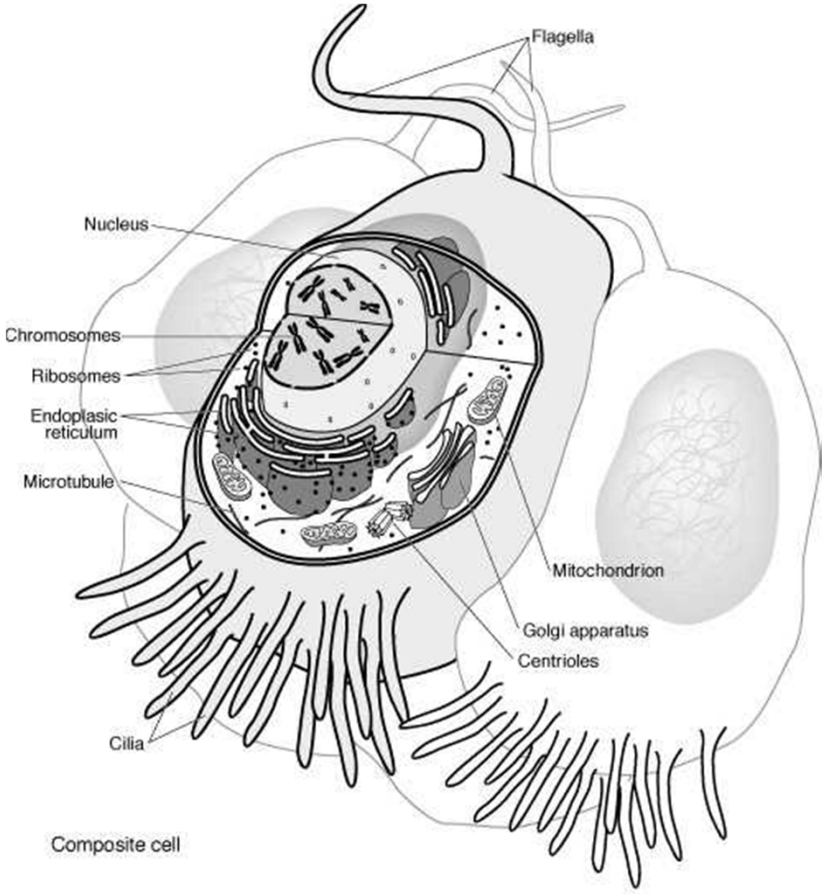
Cells & organisms

Cells and organisms

Organisms:

- Unicellular: One single cell (or simpler)
  - Archaea:
  - Prokaryotes: Bacteria, Yeast
  - Eukaryote: Protozoos
- Pluricellular: Eukaryote cells
  - Different organization levels (e.g. tissues)
  - Diversity in number, type and size
- Viruses are not properly organisms

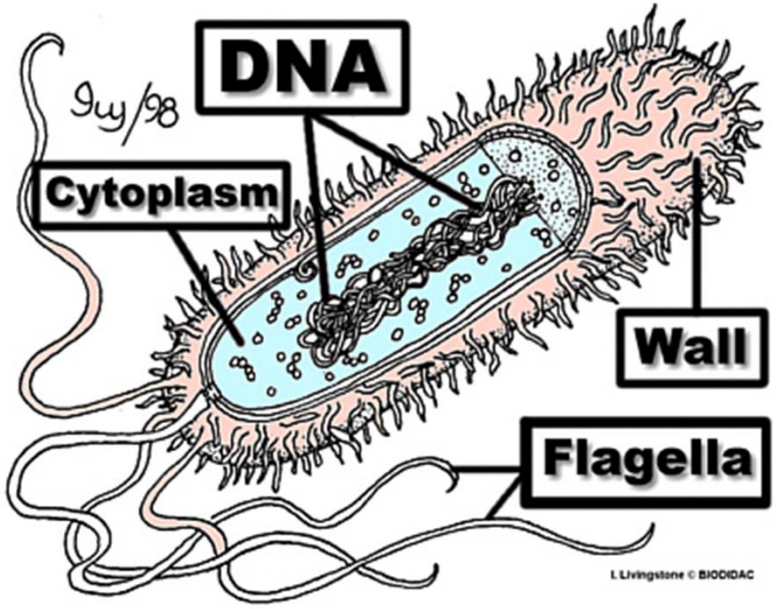
Eukaryotes



"Eu" good,

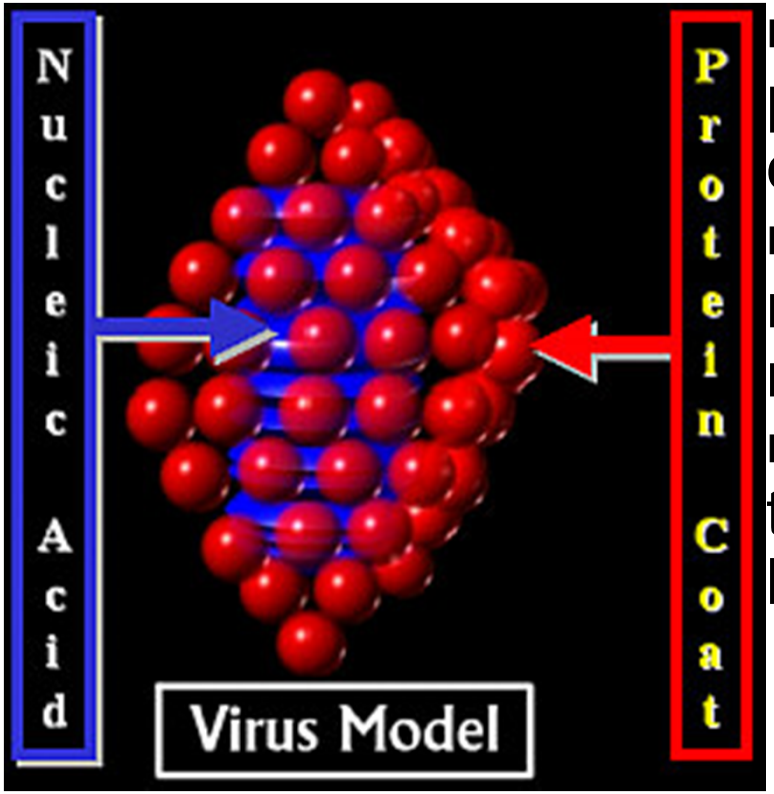
- "Karyo" nut or kernel →
- Presence of nucleous separated from del cytoplasm by the nucleous envelope.
- DNA: double-stranded, it is organized in chromosomes
- The cell contains other membrane-bound organelles
- Sexual reproduction is common

Prokaryotes



- Lack a cell nucleous
- DNA: circular single stranded
- Lack other membrane-bound organelles.
- Without sexual reproduction although there are genetic recombination

Viruses



- Contain nucleic acids and proteins but no other characteristics
- They use the cellular machinery of their hosts to replicate
- NO can replicate outside the host → NOT living beings

Content

- Cells and organisms
- Molecules of life (Biomolecules)
- The central dogma
- Genes and gene expression

@: Most pictures have been freely obtained from:
http://www.accessexcellence.org/

Biomolecules

Key biomolecules

- The basic components of biological systems are
  - Sugars (carbohydrates)
  - Fats (lipids)
  - Nucleic acids
  - Proteins
- Sugars and lipids have no important role regarding "biological information"→ Not discussed here

Nucleic acids and proteins

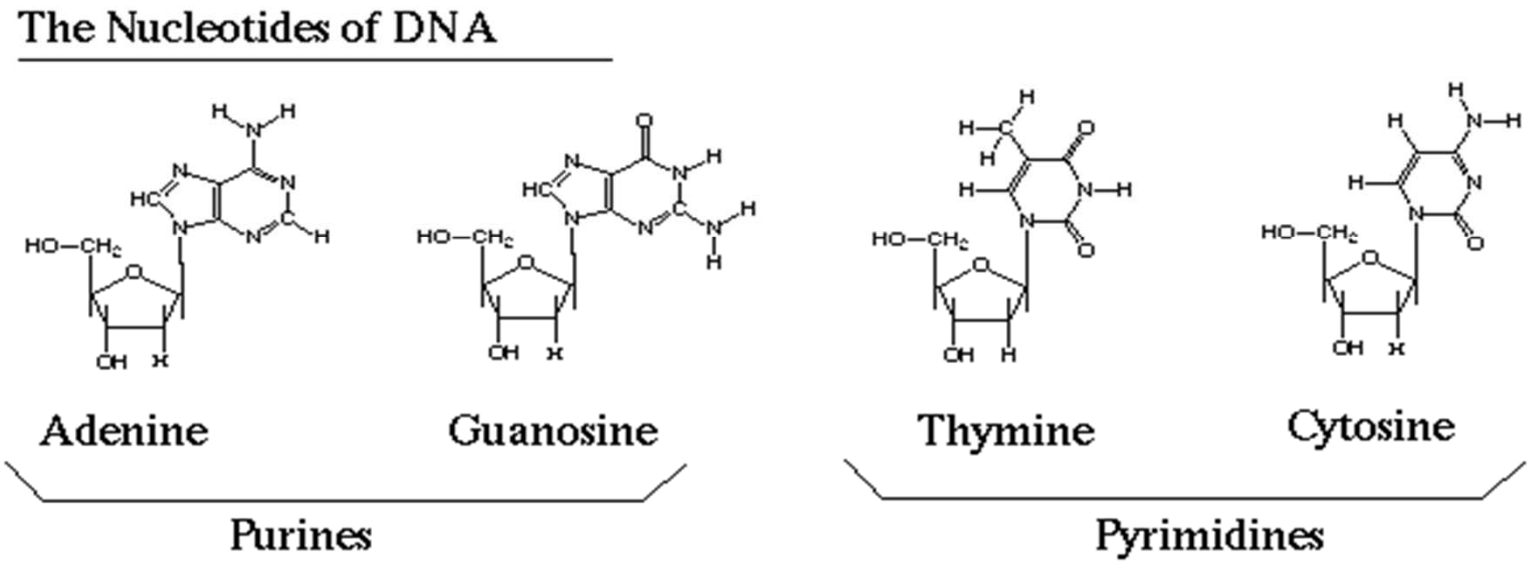
- Molecules that contain and transport information are
  - DNA (4 different nucleotides):Contains encoded biological information
  - RNA (4 different nucleotides):Carries information from DNA to proteins
  - Proteins (20 different amino acids):Function and structure of living beings

Nucleotides

- Basic components of nucleic acids
- Consisting of
  - A sugar (Ribose or Deosyribose)
  - A nitrogen base
  - A phosphate group
- In biological speaking we say "bases" instead of nucleotides to describe a string

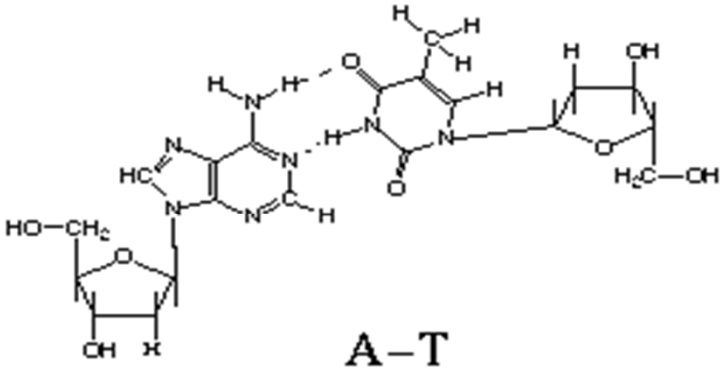
DNA nucleotides

Deoxyribonucleic Acid (DNA) contains four nucleotide bases.



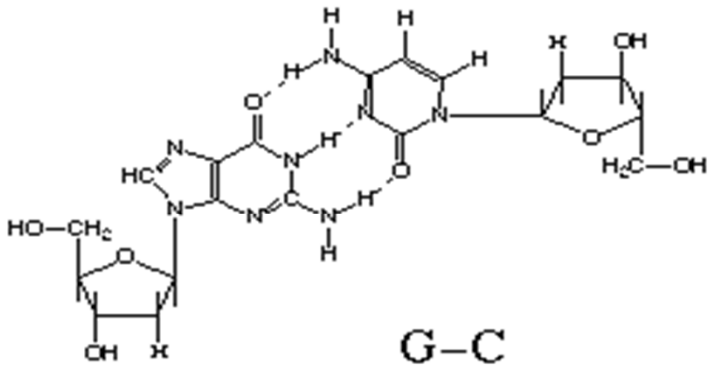
The pairing between complementary bases

DNA Basepairs



A–T

Adenosine–Thymidine
(Adenine–Thymine)



G–C

Guanosine–Cytidine
(Guanine–Cytosine)

The primary structure of DNA

- Sequence of nuceotides
- Forms an unbranched polymer
- Organized in a double-stranded

```
atgaatcgta ggggtttgaa cgctggcaat
acgatgactt ctcaagcgaa cattgacgac
ggcagctgga aggcggtctc cgagggcgga
```

DNA vs RNA

- DNA is organized into a complementary double helix. RNA does not.
- One of the four bases are :different
  - DNA: A,C,G,T
  - RNA: A,C,G,U
- Differ from the nucleotide sugar
  - DNA: Deoxyribose
  - RNA: Ribose

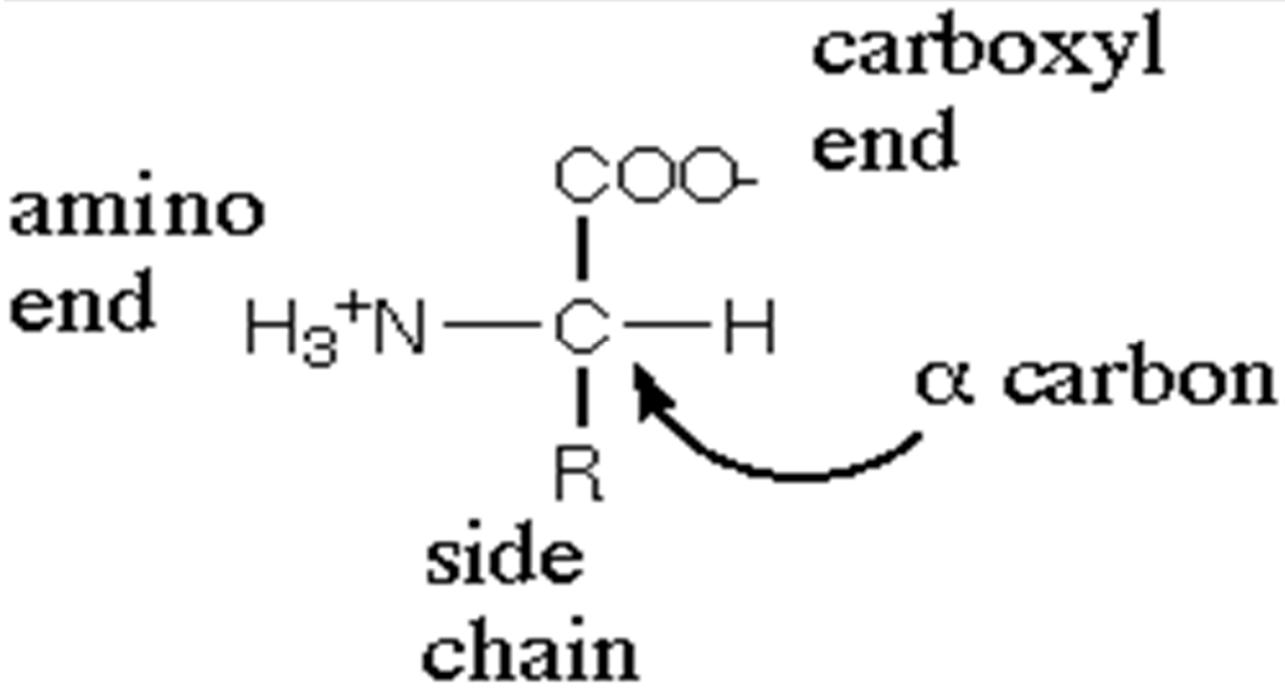
DNA strands sense

- Each DNA strand has a polarity:
  - The strand begins with the 1st nucleotide 5'-hydroxyl (or 5'-phospho) group, and
  - Ends with the last nucleotide 3'-hydroxyl group
  - The strand goes 5' to 3' ("*Five prime to three prime*")
- The two DNA strands are antiparallel
  - One goes 5' → 3' and the other goes 3' → 5'.

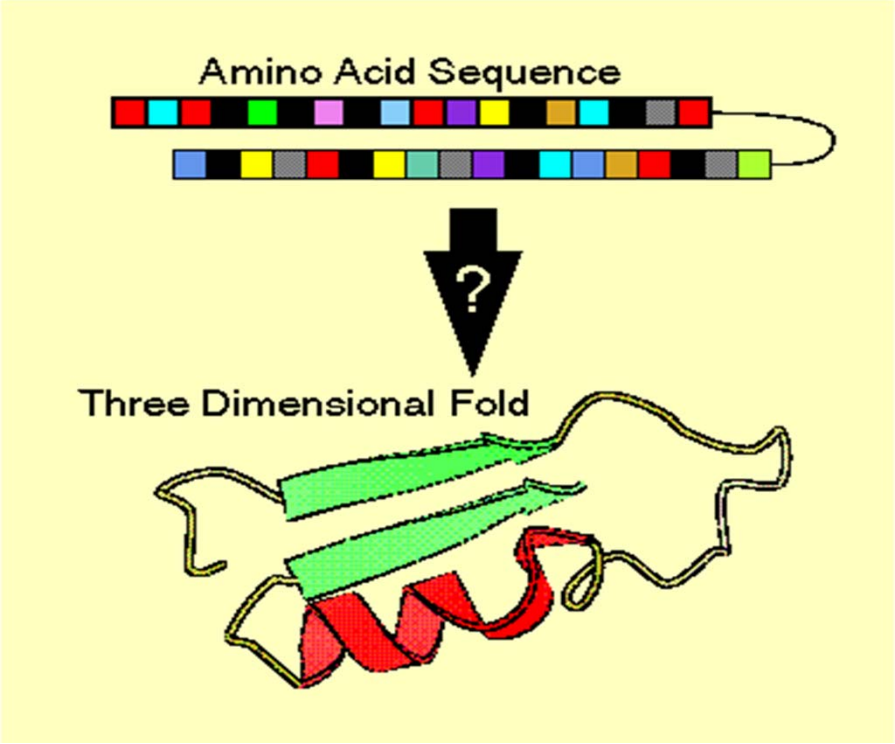
Proteins

- Amino acid sequence
- Forms an unbranched polymer
- There are 20 different amino acids (AA)
- The key function of proteins is in its three dimensional structure

Amino acids



Proteins “fold” into conformational structure

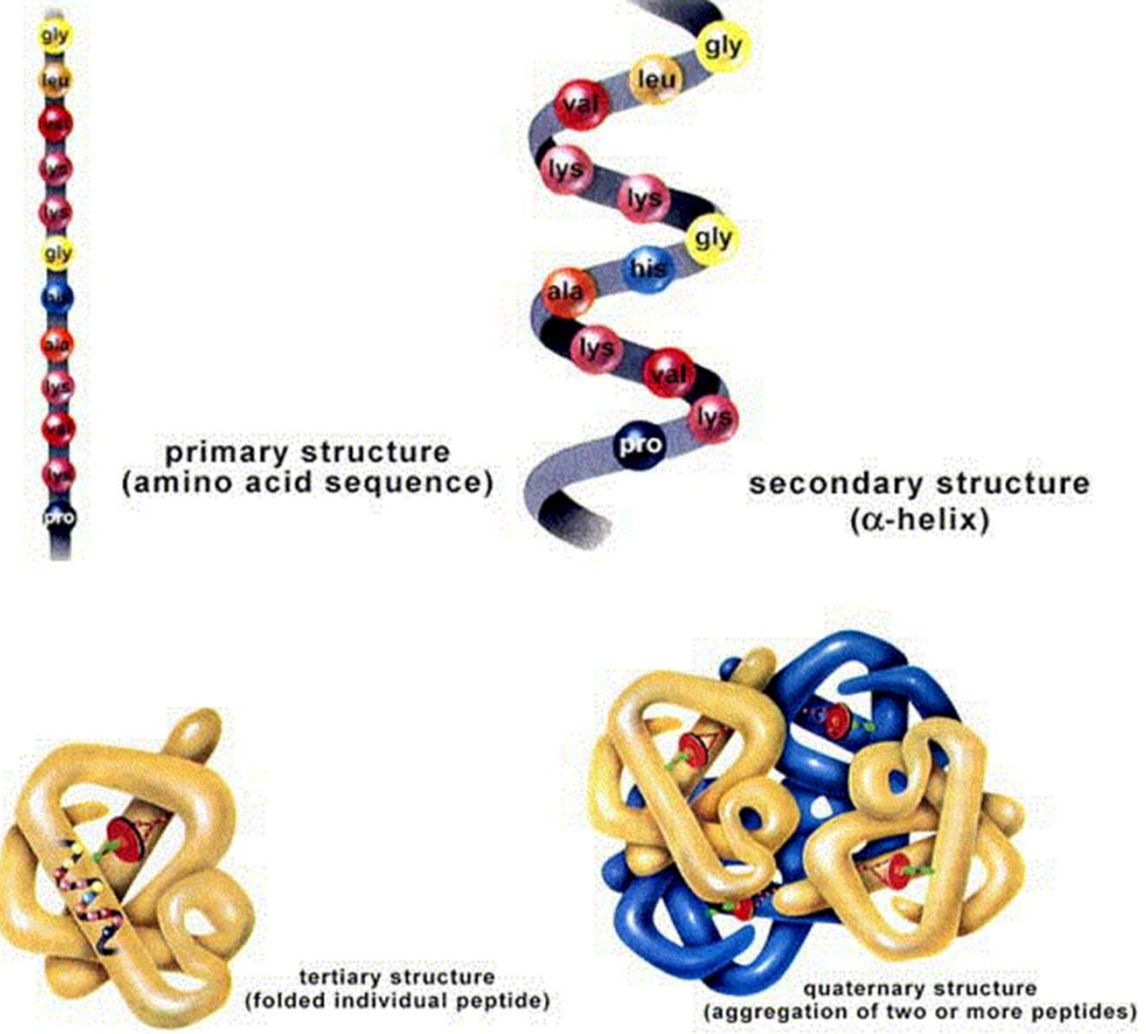


Protein's structure

Proteins fold into lowest to highest complexity level:

- *Primary structure*: amino acid sequence
- *Secondary structure*: regularly repeating local structures stabilized by hydrogen bonds.
- *Tertiary structure*: fold into 3-dimensional structures.
- *Quaternary structure*: structure formed by several protein molecules (protein complex).

Protein's structure



Storing the information: the databases

- Storing information efficiently is not a trivial procedure. Take as an example the structure of a protein

- Some important information to take into account after finding the structure :
  - tissue, organelle, author and data of the experiment
  - primary sequence
  - coordinates of the atoms 3D to describe the tertiary structure
- Example: prion human

- LEU GLY GLY TYR MET LEU GLY SER ALA MET SER ARG PRO ILE ILE HIS PHE GLY SER
ASP TYR GLU ASP ARG TYR TYR ARG GLU ASN MET HIS ARG TYR PRO ASN GLN VAL TYR
TYR ARG PRO MET ASP GLU TYR SER ASN GLN ASN ASN PHE VAL HIS ASP CYS VAL ASN
ILE THR ILE LYS GLN HIS THR VAL THR THR THR THR LYS GLY GLU ASN PHE THR GLU
THR ASP VAL LYS MET MET GLU ARG VAL VAL GLU GLN MET CYS ILE THR GLN TYR GLU
ARG GLU SER GLN ALA TYR TYR GLN ARG

Storing information (2)

- Coordinates of the 1692 atoms

```
ATOM      1  N   LEU A 125       2.863 -15.219  -0.703  1.00  0.00           N
ATOM      2  CA  LEU A 125       3.920 -14.209  -0.705  1.00  0.00           C
ATOM      3  C   LEU A 125       5.265 -14.836  -1.065  1.00  0.00           C
ATOM      4  O   LEU A 125       5.419 -15.343  -2.175  1.00  0.00           O
ATOM      5  CB  LEU A 125       3.934 -13.422   0.620  1.00  0.00           C
ATOM      6  CG  LEU A 125       2.735 -12.479   0.802  1.00  0.00           C
ATOM      7  CD1 LEU A 125       2.918 -11.733   2.120  1.00  0.00           C
ATOM      8  CD2 LEU A 125       2.639 -11.428  -0.307  1.00  0.00           C
ATOM      9  H   LEU A 125       2.597 -15.618   0.186  1.00  0.00           H
ATOM     10  HA  LEU A 125       3.725 -13.494  -1.493  1.00  0.00           H
ATOM     11  HB2 LEU A 125       3.969 -14.127   1.453  1.00  0.00           H
ATOM     12  HB3 LEU A 125       4.832 -12.802   0.651  1.00  0.00           H
ATOM     13  HG  LEU A 125       1.809 -13.054   0.839  1.00  0.00           H
ATOM     14 HD11 LEU A 125       2.964 -12.448   2.937  1.00  0.00           H
ATOM     15 HD12 LEU A 125       3.840 -11.151   2.090  1.00  0.00           H
ATOM     16 HD13 LEU A 125       2.078 -11.061   2.284  1.00  0.00           H
ATOM     17 HD21 LEU A 125       3.583 -10.893  -0.412  1.00  0.00           H
ATOM     18 HD22 LEU A 125       2.386 -11.889  -1.259  1.00  0.00           H
ATOM     19 HD23 LEU A 125       1.850 -10.720  -0.066  1.00  0.00           H
ATOM     20  N   GLY A 126       6.251 -14.760  -0.176  1.00  0.00           N
ATOM     21  CA  GLY A 126       7.629 -15.134  -0.436  1.00  0.00           C
ATOM     22  C   GLY A 126       8.550 -13.932  -0.294  1.00  0.00           C
```

Visualization of a protein

- Go to pdb data base to use the application jmol.
- The program reads the information in the previous format (pdb) and displays a 3D representation of the protein.

Go to “View in 3D (Jmol)”
Accept the installation of the Java applet
You will see a 3D representation of a protein.
To represent the human prion:
Open the Jmol console
Paste the information of the human prion in the editor fields and “load”.

Content

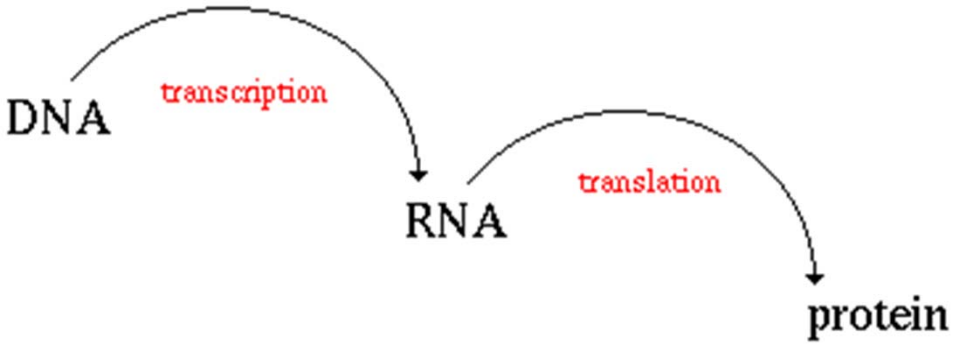
- Cells and organisms
- Molecules of life (Biomolecules)
- Central dogma of molecular biology
- Genes and gene expression

@: Most pictures have been freely obtained from:
http://www.accessexcellence.org/

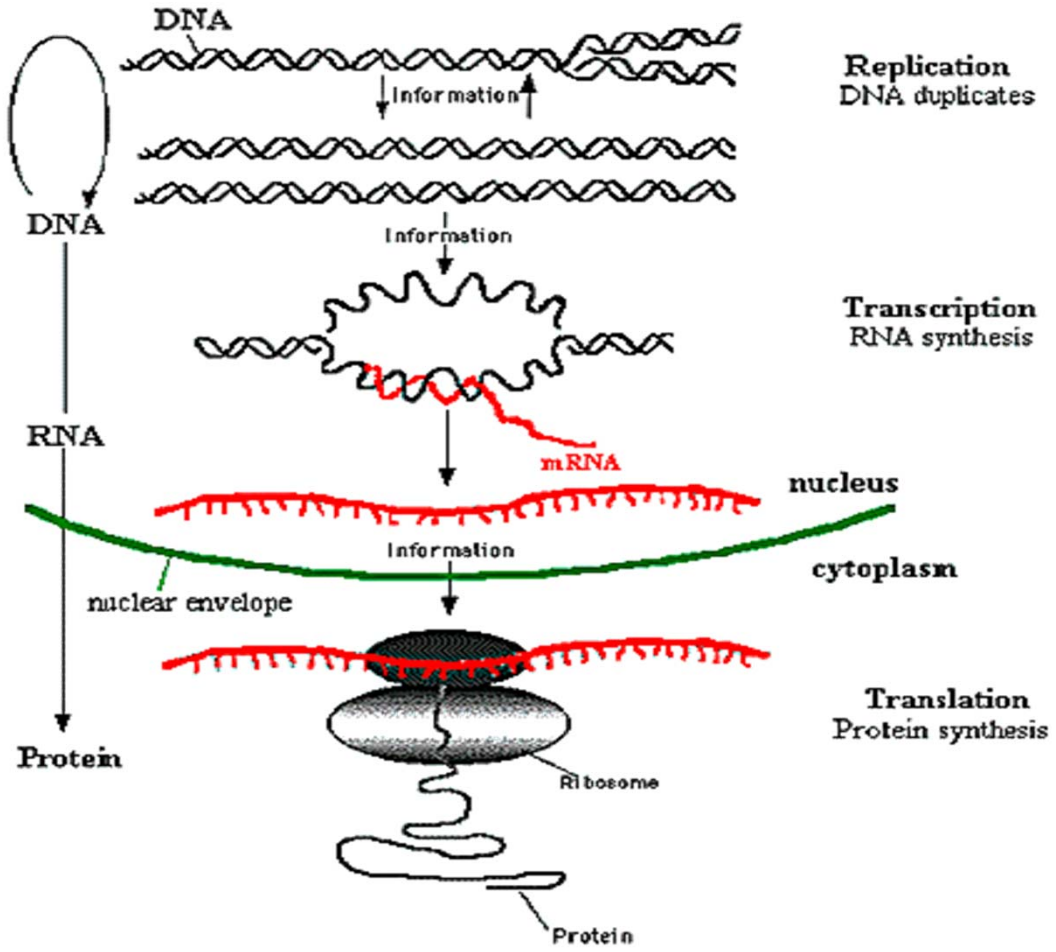
Central dogma of molecular biology

Central dogma of molecular biology

- *Central dogma of molecular biology* states that information encoded in DNA is transferred to proteins through RNA.



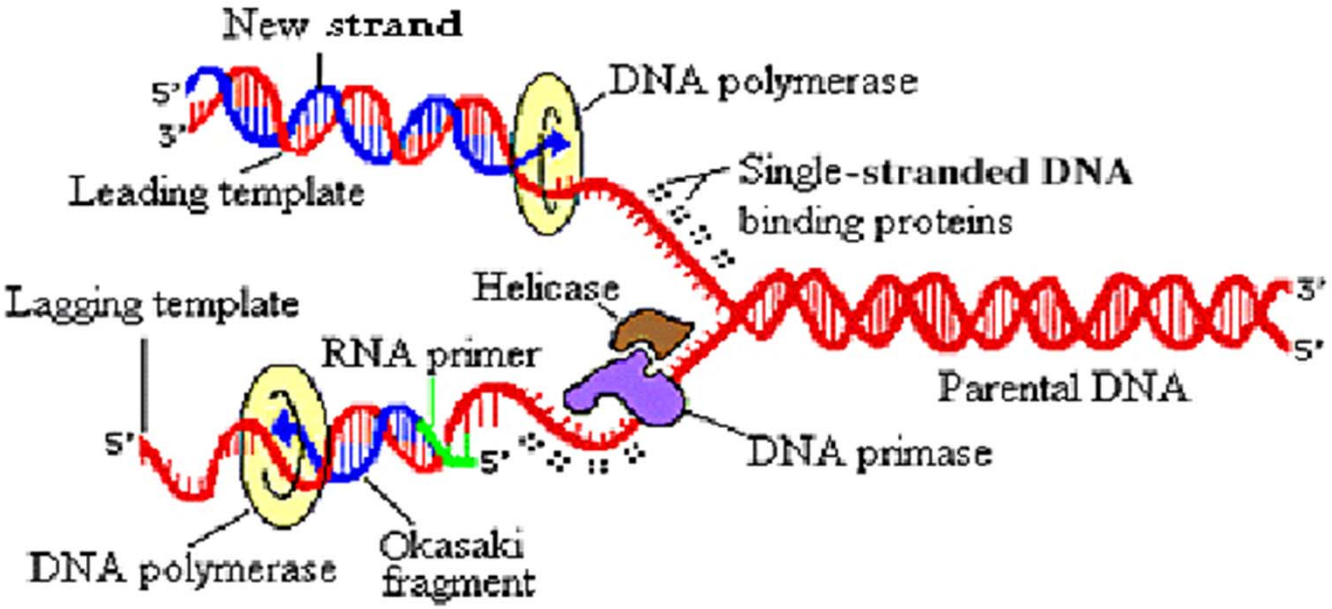
Central dogma of molecular biology



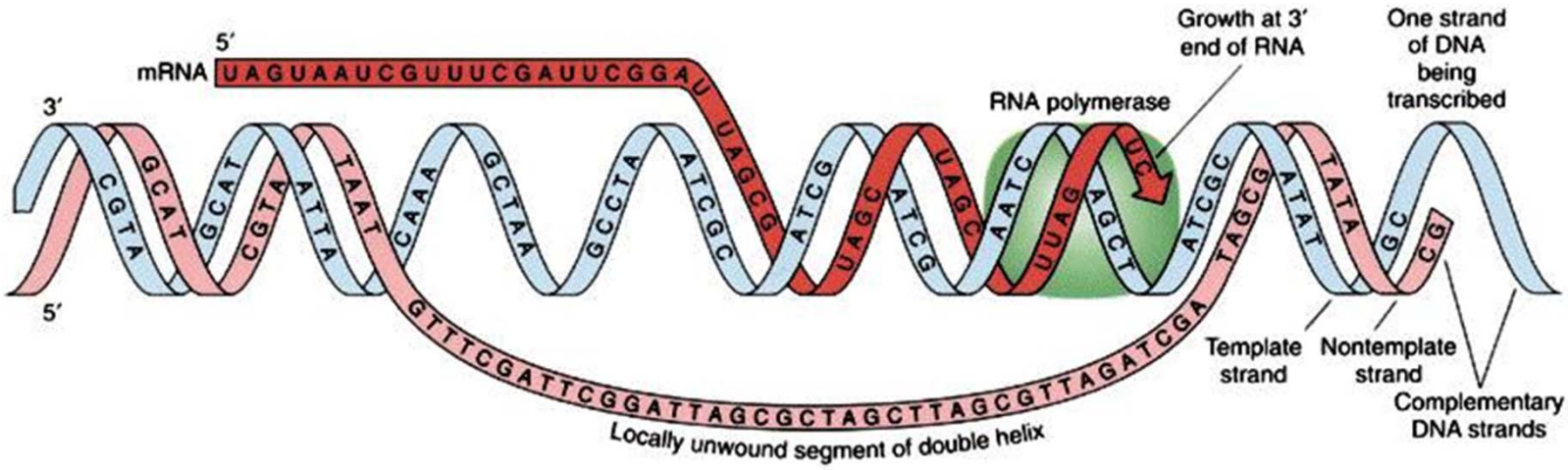
From DNA to proteins

- DNA is replicated in a complex process involving many enzymes: replication **replicación**
- DNA is copied in a string of complementary messenger RNA (mRNA): transcription transcripción
- In eukaryotic cells, the mRNA is processed procesa eliminating coding fragments ("splicing") and migrates from the nucleous to the cytoplasm.
- The mRNA carries coded information to ribosomes (ribosomal RNA) that "read" and perform protein synthesis: translation **traslación**

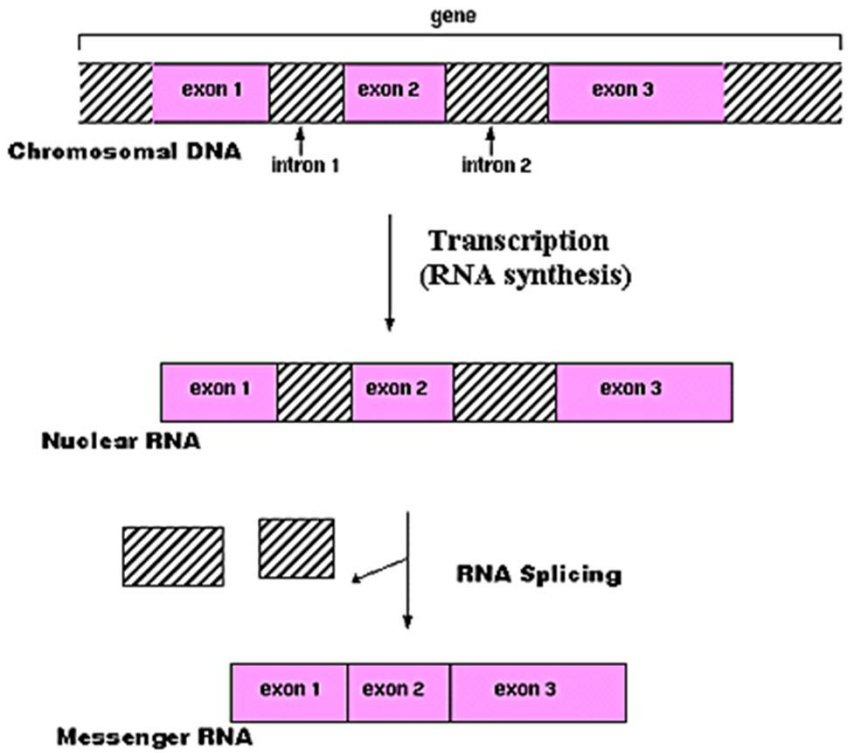
1. Replication



2. Transcription

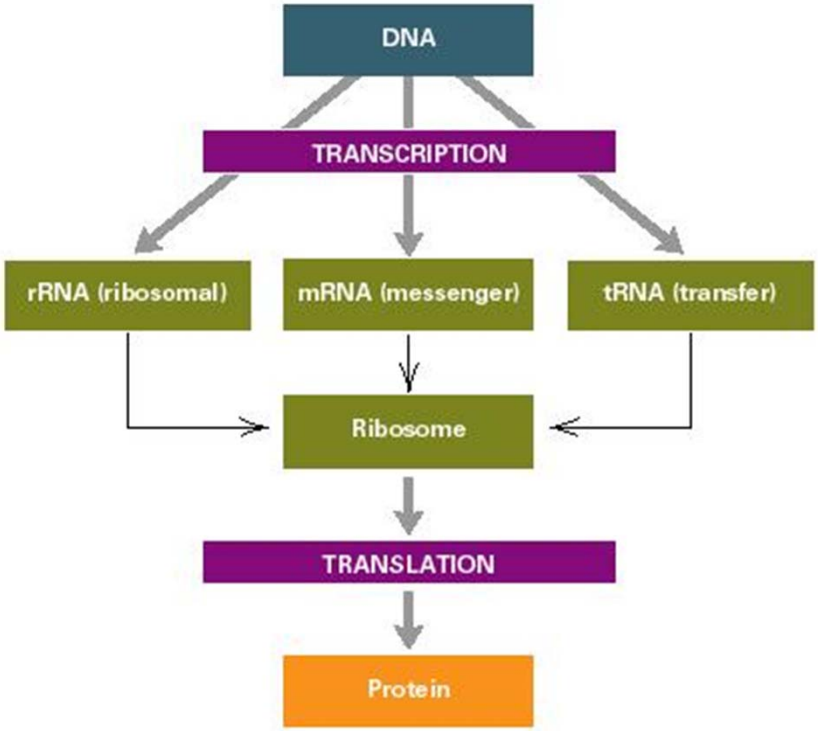


3. RNA processing or splicing

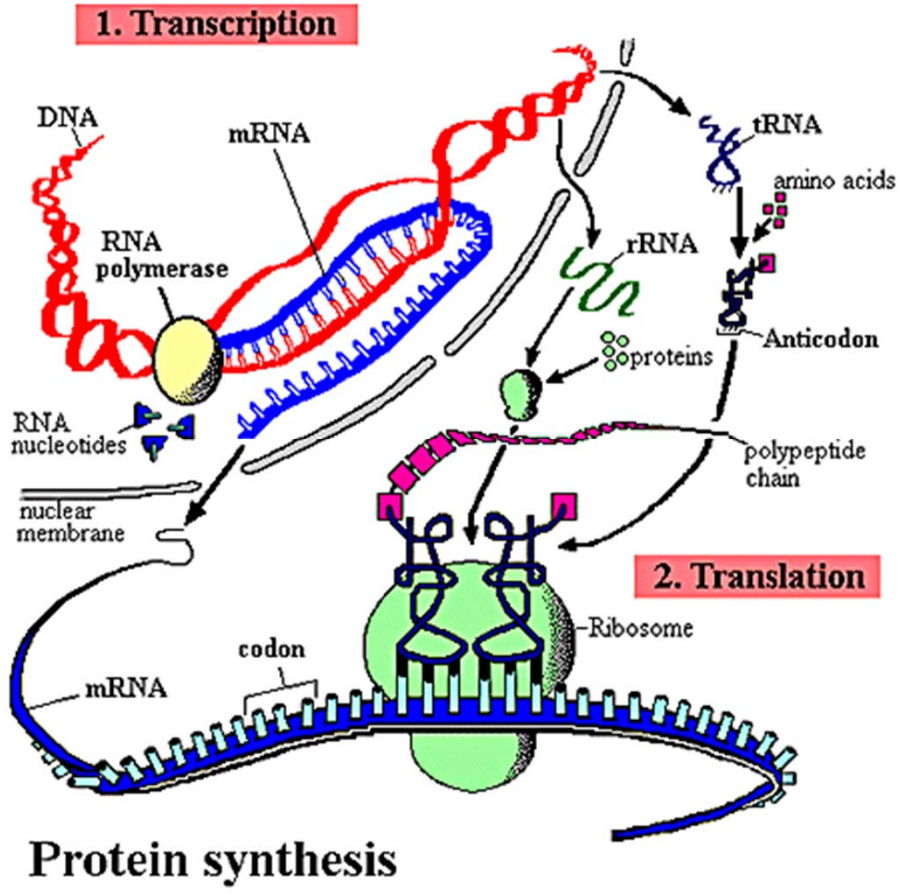


RNA synthesis and processing

1DNA vs. 3 RNA's

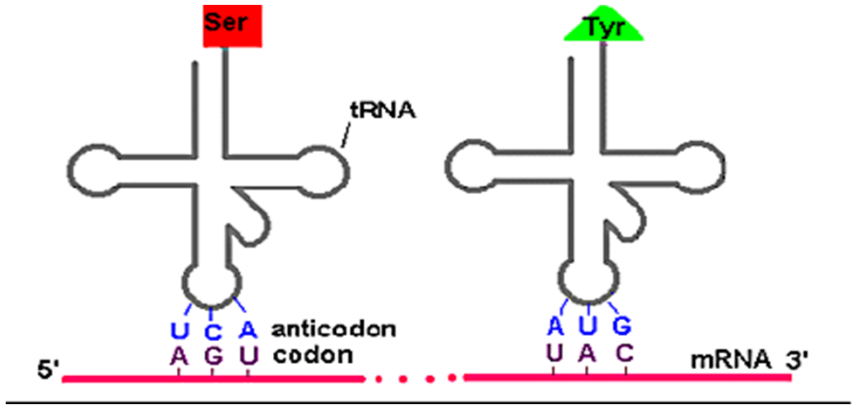


4. Protein synthesis



- The mRNA goes to the cytoplasm where it binds to ribosomes.
- Codon: mRNA information unit.
- The tRNA brings the complementary AA tRNA.
- The AA are bound to the protein to complete the sequence.
- Animations (1), (2)

4’ 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

Content

- Cells and organisms
- Molecules of life (Biomolecules)
- Central dogma of molecular biology
- Genes and gene expression

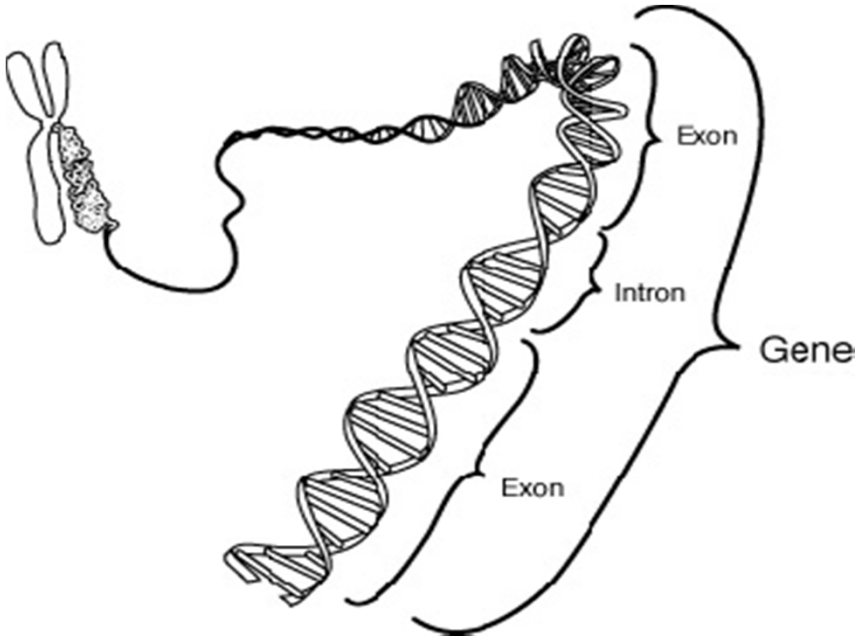
@: Most pictures have been freely obtained from:
http://www.accessexcellence.org/

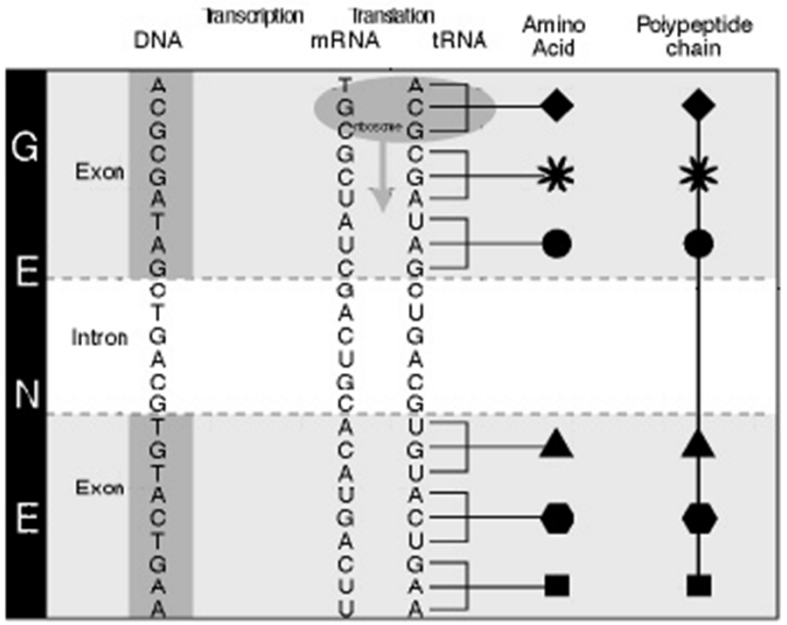
Genes and gene expression

What is a gene?

- It is the functional and physical unit of heredity transmitted from one generation to their offsprings.
- Genes are DNA fragments
- Most of the genes contain the necessary information for the synthesis of a specific protein.

Gene components





How can find a gene in a sequence?

- First guidelines:
  - There are 6 possible frame translation from a sequence
    - 3 from 5' – 3' and 3 from 3'- 5'
  - You have to find open reading frames (ORF) of enough length.
    - An ORF is a continuous stretch of codons that do not contain a stop codon (usually UAA, UAG or UGA)

Six-Frame Translation

ATGACTTGTCATGATCGCGCCAAAATTCAACTCGCCGGAAGGGCCCGACGGGCGACGACC

5'3' Frame 1

```
atgacttgtcatgatcgcgccaaaattcaactcgccggaagggcccgacgggcgacgacc
 M  T  C  H  D  R  A  K  I  Q  L  A  G  R  A  R  R  A  T  T
```

5'3' Frame 2

```
atgacttgtcatgatcgcgccaaaattcaactcgccggaagggcccgacgggcgacgacc
  -  L  V  M  I  A  P  K  F  N  S  P  E  G  P  D  G  R  R
```

5'3' Frame 3

```
atgacttgtcatgatcgcgccaaaattcaactcgccggaagggcccgacgggcgacgacc
   D  L  S  -  S  R  Q  N  S  T  R  R  K  G  P  T  G  D  D
```

3'5' Frame 1

```
ggtcgtcgcccgtcgggcccttccggcgagttgaattttggcgcgatcatgacaagtcat
 G  R  R  P  S  G  P  S  G  E  L  N  F  G  A  I  M  T  S  H
```

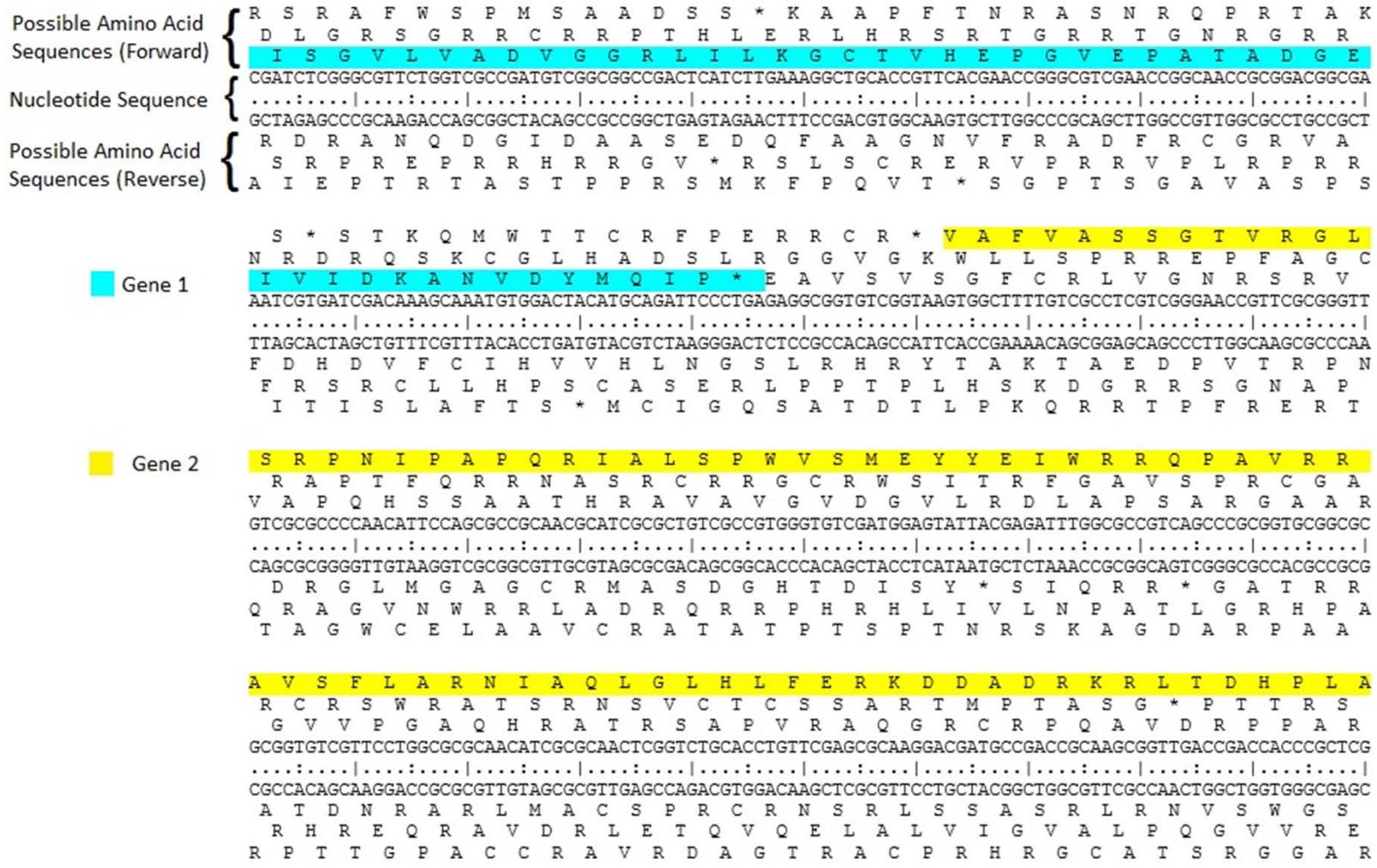
3'5' Frame 2

```
ggtcgtcgcccgtcgggcccttccggcgagttgaattttggcgcgatcatgacaagtcat
  V  V  A  R  R  A  L  P  A  S  -  I  L  A  R  S  -  Q  V
```

3'5' Frame 3

```
ggtcgtcgcccgtcgggcccttccggcgagttgaattttggcgcgatcatgacaagtcat
   S  S  P  V  G  P  F  R  R  V  E  F  W  R  D  H  D  K  S
```

Open reading frame



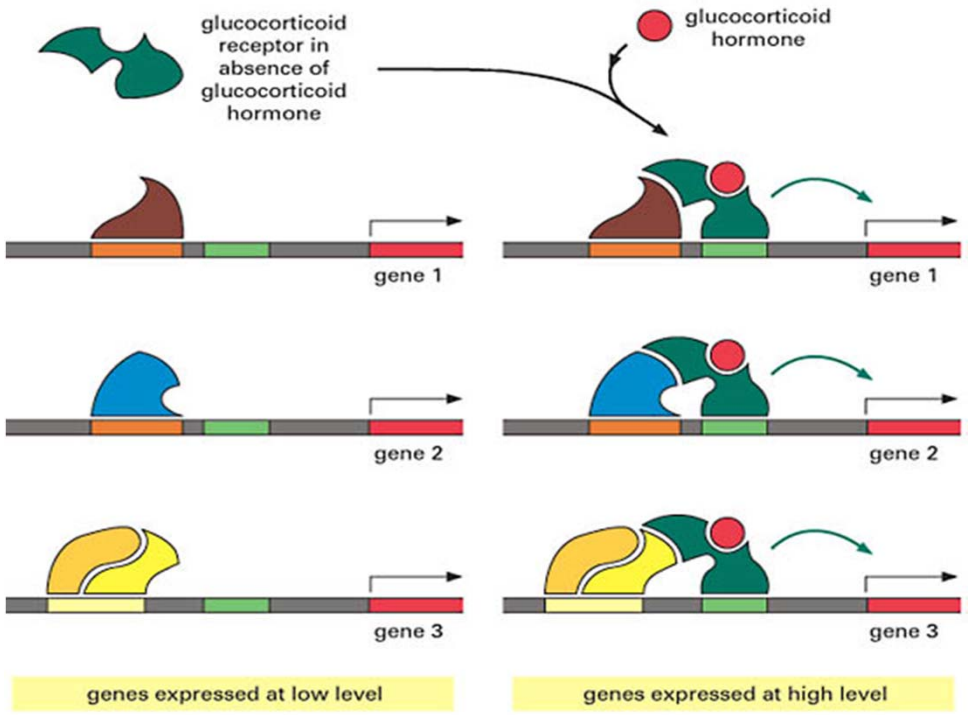
Regulation of gene expression

- Genes can be “turned on” or “off”.
- Each cell expresses (or active or "on") only a fraction of their genes.
- Remaining genes are repressed ("off").
- The process consisting in activate genes and suppress others is *gene regulation*.
- Gene regulation determines:
  - The appearance and different function of different cells types
  - The ability of some cells to react quickly to environmental changes

How genes are regulated?

- Gene regulation can occur at any point in the process of expression but often occurs during transcription.
- Environmental signs or other cells activate proteins called *transcription factor*.
- They bind to the *regulatory regions* of genes, increasing or decreasing the level of transcription→ They control the amount of gene product produced by the gene in every moment.

Example: Genes activation



- With no HGC (glucocorticoid hormone) genes are inactive
- In presence of HGC genes are activated and expressed (in block)

Mutations

- Mutations are genetic changes randomly produced or by the action of mutagens (chemicals, X rays, UV, etc.)
- Most are lethal because the original AA and nucleotides sequence is the product of millions of years of evolution
  - → Product of natural selection (or not...)

Mutations at the molecular level

- A mutation results in a change in the order of nucleotides in the genes
- Consider for example the peptide

```
Nucleotides          TAC TTA {C}GA TAA TGC ATT
Codons mRNA-         AUG AAU {G}CU AUU ACG UAA
Sequence AA-         met asn  ala  ile thr stop
```

- We can change a nucleotide (eg. “C”→”G”) for another or delete it.

Point mutations

- Point mutation
  - Substitution of one nucleotide for another
  - It can be lethal or harmless (due to the degeneracy of the code)

```
Nucleotides          TAC TTA {C}GA TAA TGC ATT
Substitution C→G     TAC TTA {G}GA TAA TGC ATT
Codons mRNA-         AUG AAU {C}CU AUU ACG UAA
New seq.             met asn  pro  ile thr stop
Original seq.        met asn  ala  ile thr stop
```

“Frameshift” mutation

- Change the reading frame
  - Deleting a nucleotide→ change in the grouping of codons

Nucleotides	TAC TTA {C}GA TAA TGC ATT
C supression	TAC TTA GA(T) AAT GCA TT?
Codons mRNA-	AUG AAU CU(A) UUA CGU UA?
New seq.	met asn **leu** **arg** **lys** ???
Original seq.	met asn **ala** ile thr stop