

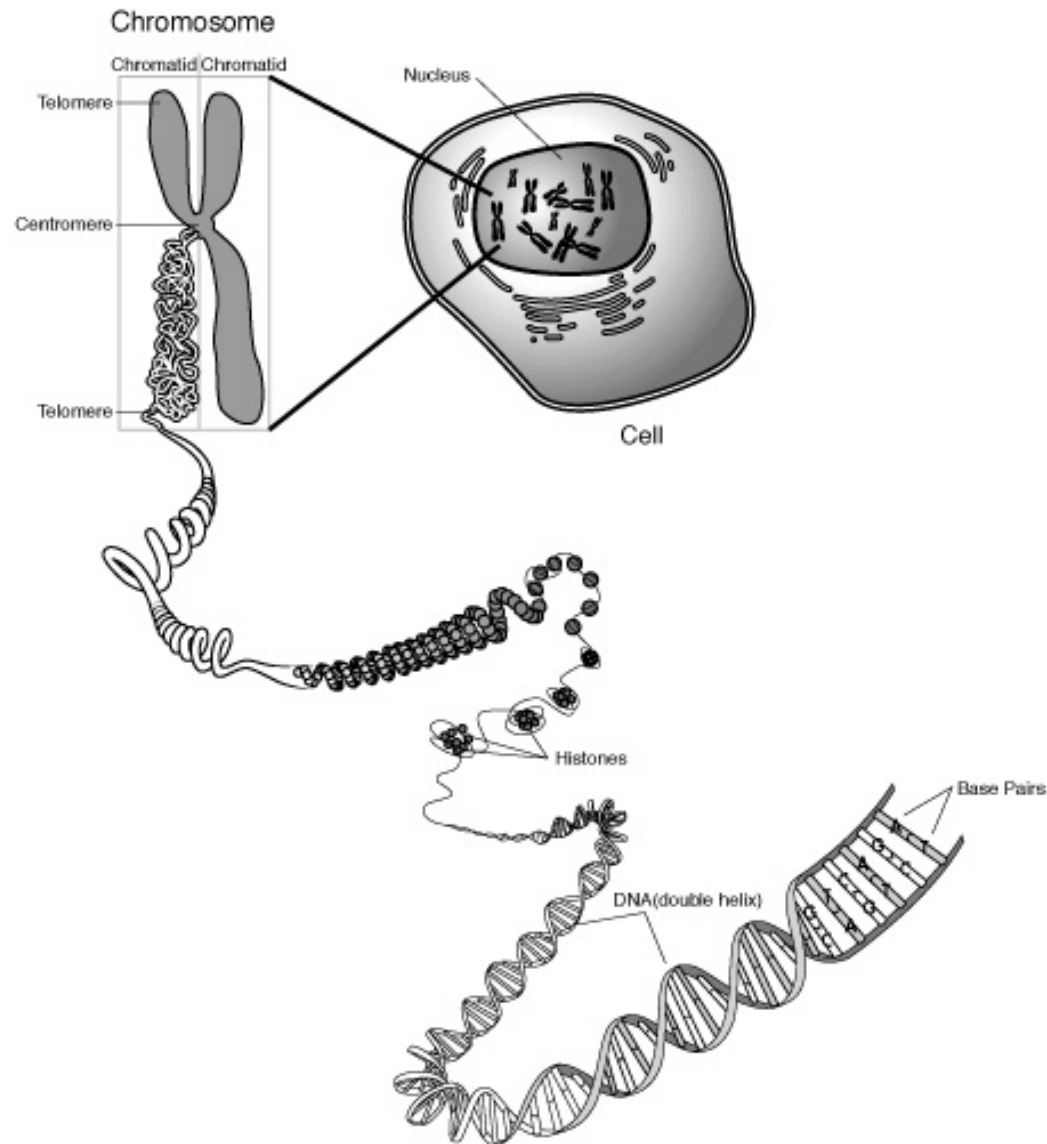
Genetic code, transcription and translation

Adapted from the lesson
“Introduction to genome biology”

S. Dudoit and R. Gentleman

University of Berkeley

Chromosomes and DNA



DNA structure

- A **deoxyribonucleic acid** or **DNA** molecule is a double-stranded polymer composed of four basic molecular units called nucleotides.
- Each **nucleotide** comprises
 - a phosphate group;
 - a deoxyribose sugar;
 - one of four nitrogen bases:
 - purines: **adenine (A)** and **guanine (G)**,
 - pyrimidines: **cytosine (C)** and **thymine (T)**.

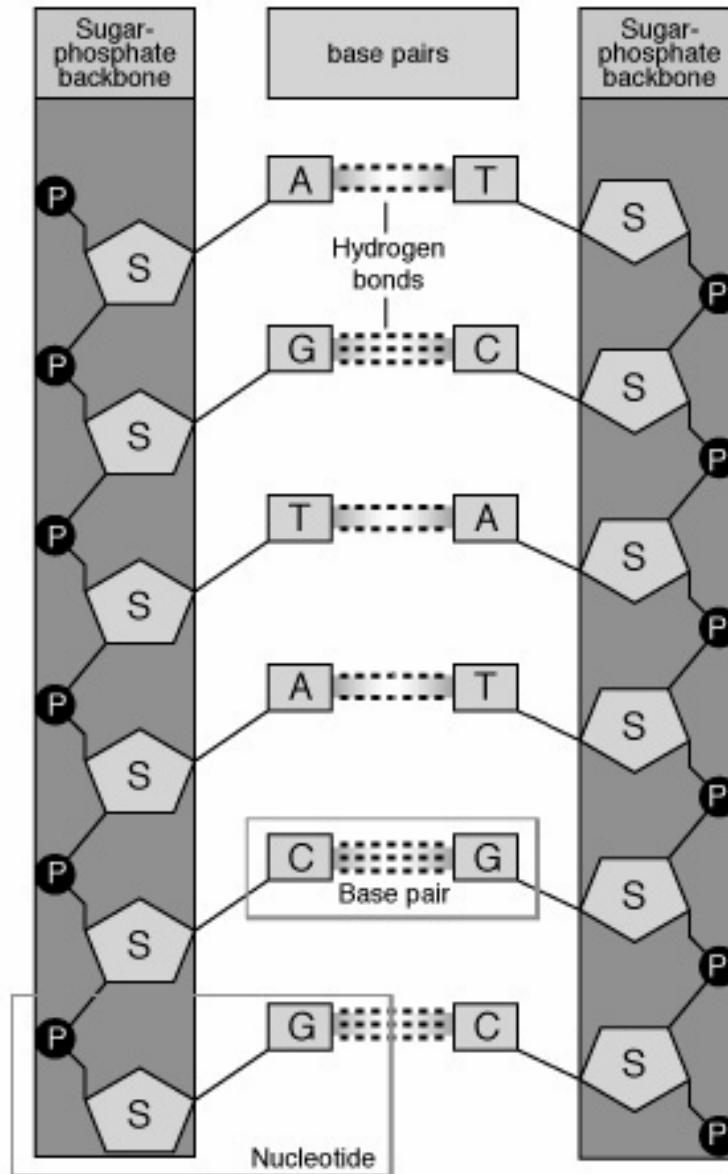
DNA structure

- Base-pairing occurs according to the following rule:
 - **C pairs with G,**
 - **A pairs with T.**
- The two chains are held together by hydrogen bonds between nitrogen bases.

DNA structure



DNA structure



DNA structure

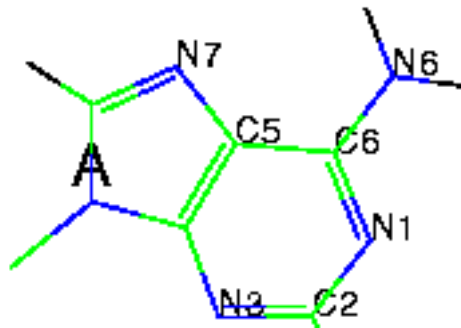


Four nucleotide bases:

- purines: A, G
- pyrimidine: T, C

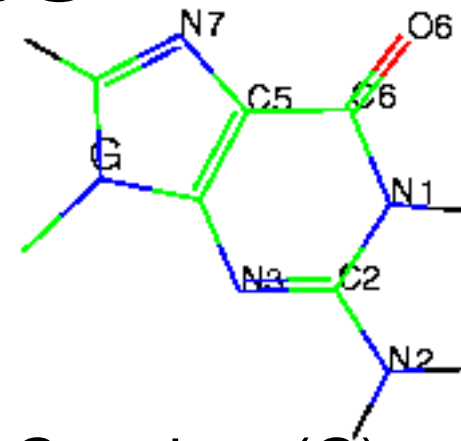
A pairs with T, 2 H bonds
C pairs with G, 3 H bonds

Nucleotide bases



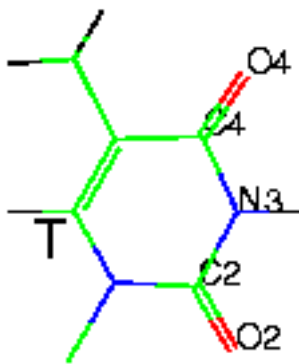
Adenine (A)

Purines

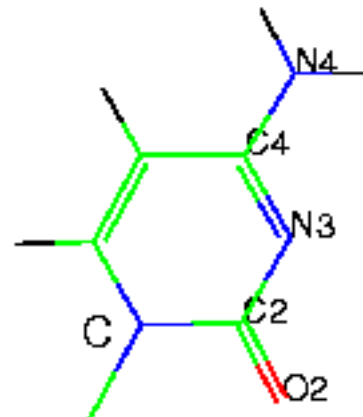


Guanine (G)

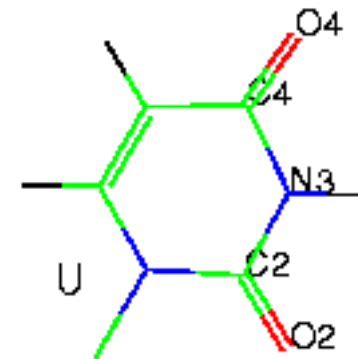
Pyrimidines



Thymine (T)
(DNA)



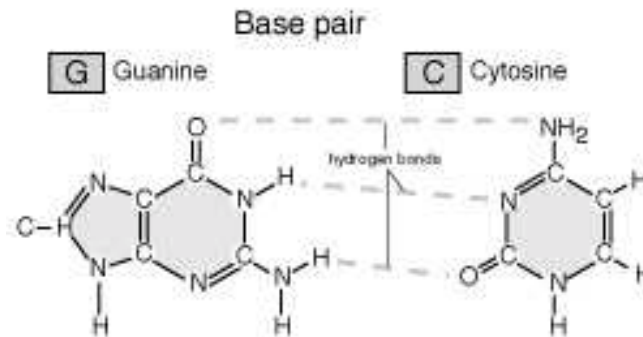
Cytosine (C)



Uracil (U)
(RNA)

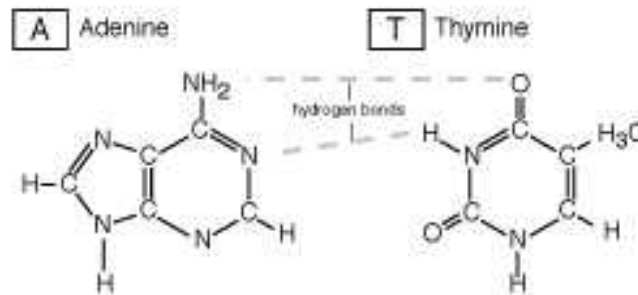
Nucleotide base pairing

G-C pair

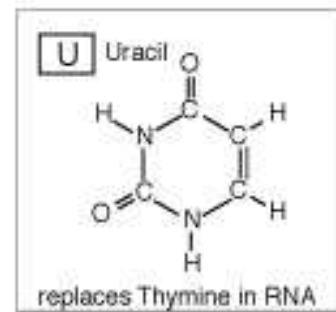


3 H bonds

A-T pair



2 H bonds



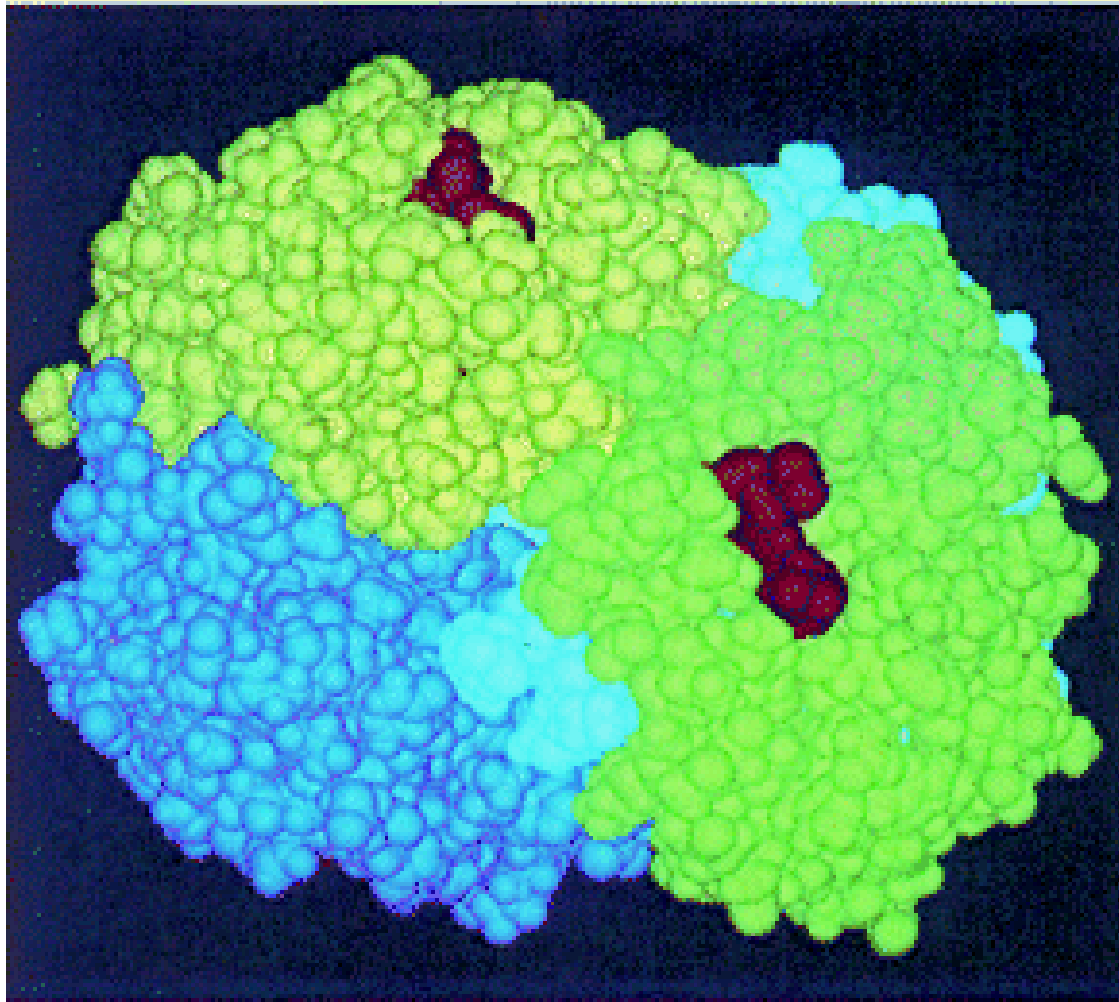
DNA structure

- Polynucleotide chains are **directional** molecules, with slightly different structures marking the two ends of the chains, the so-called **3' end** and **5' end**.
- The 3' and 5' notation refers to the numbering of carbon atoms in the sugar ring.
- The 3' end carries a sugar group and the 5' end carries a phosphate group.
- The two complementary strands of DNA are **antiparallel** (i.e, 5' end to 3' end directions for each strand are opposite)

The human genome in numbers

- 23 pairs of chromosomes;
- 2 meters of DNA;
- 3,000,000,000 bp;
- 35 M (males 27M, females 44M);
- 30,000-40,000 genes.

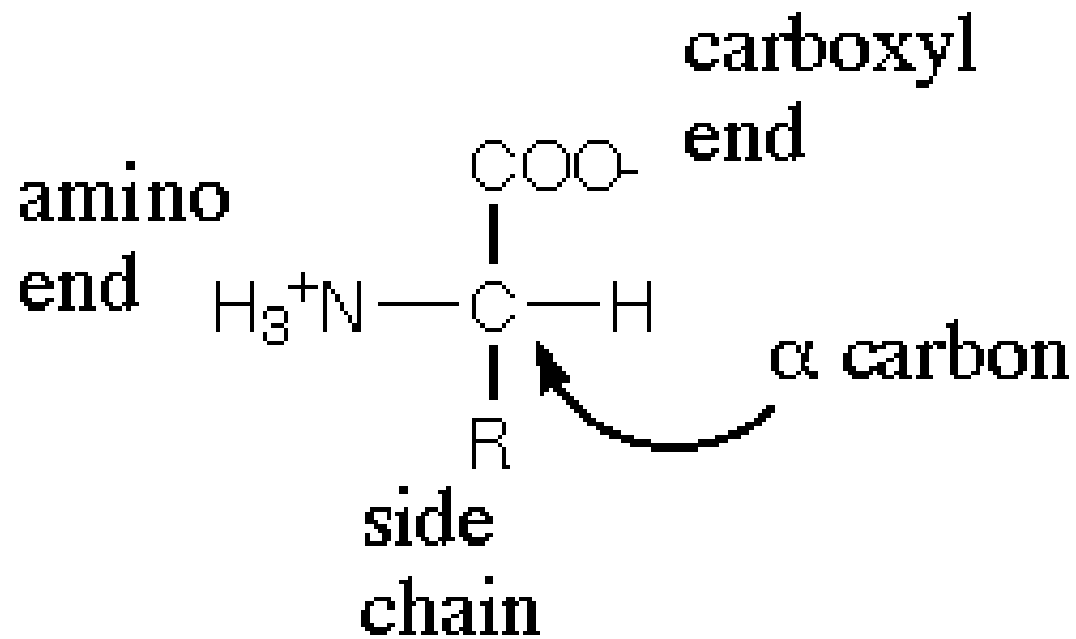
Proteins



Proteins

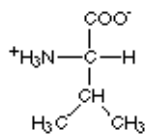
- **Proteins:** large molecules composed of one or more chains of amino acids, **polypeptides**.
- **Amino acids:** class of 20 different organic compounds containing a basic amino group ($-\text{NH}_2$) and an acidic carboxyl group ($-\text{COOH}$).
- The order of the amino acids is determined by the **base sequence** of nucleotides in the **gene** coding for the protein.
- E.g. hormones, enzymes, antibodies.

Amino acids

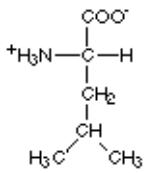


Amino acids

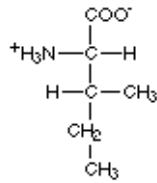
Amino acids with hydrophobic side groups



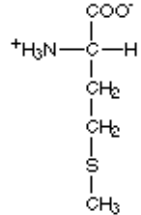
Valine
(val)



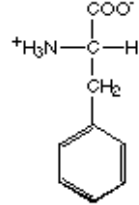
Leucine
(leu)



Isoleucine
(ile)

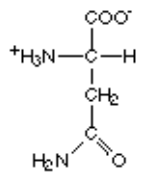


Methionine
(met)

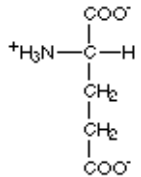


Phenylalanine
(phe)

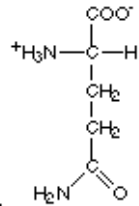
Amino acids with hydrophilic side groups



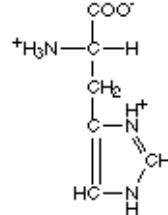
Asparagine
(asn)



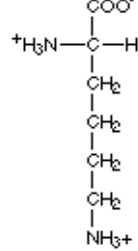
Glutamic acid
(glu)



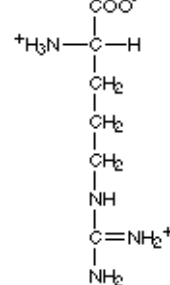
Glutamine
(gln)



Histidine
(his)



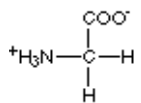
Lysine
(lys)



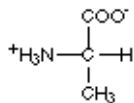
Arginine
(arg)

Aspartic acid
(asp)

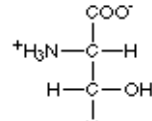
Amino acids that are in between



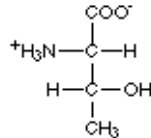
Glycine
(gly)



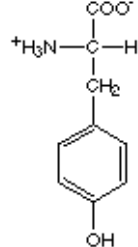
Alanine
(ala)



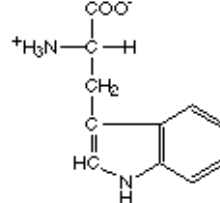
Serine
(ser)



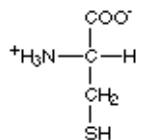
Threonine
(thr)



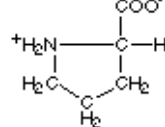
Tyrosine
(tyr)



Tryptophan
(trp)

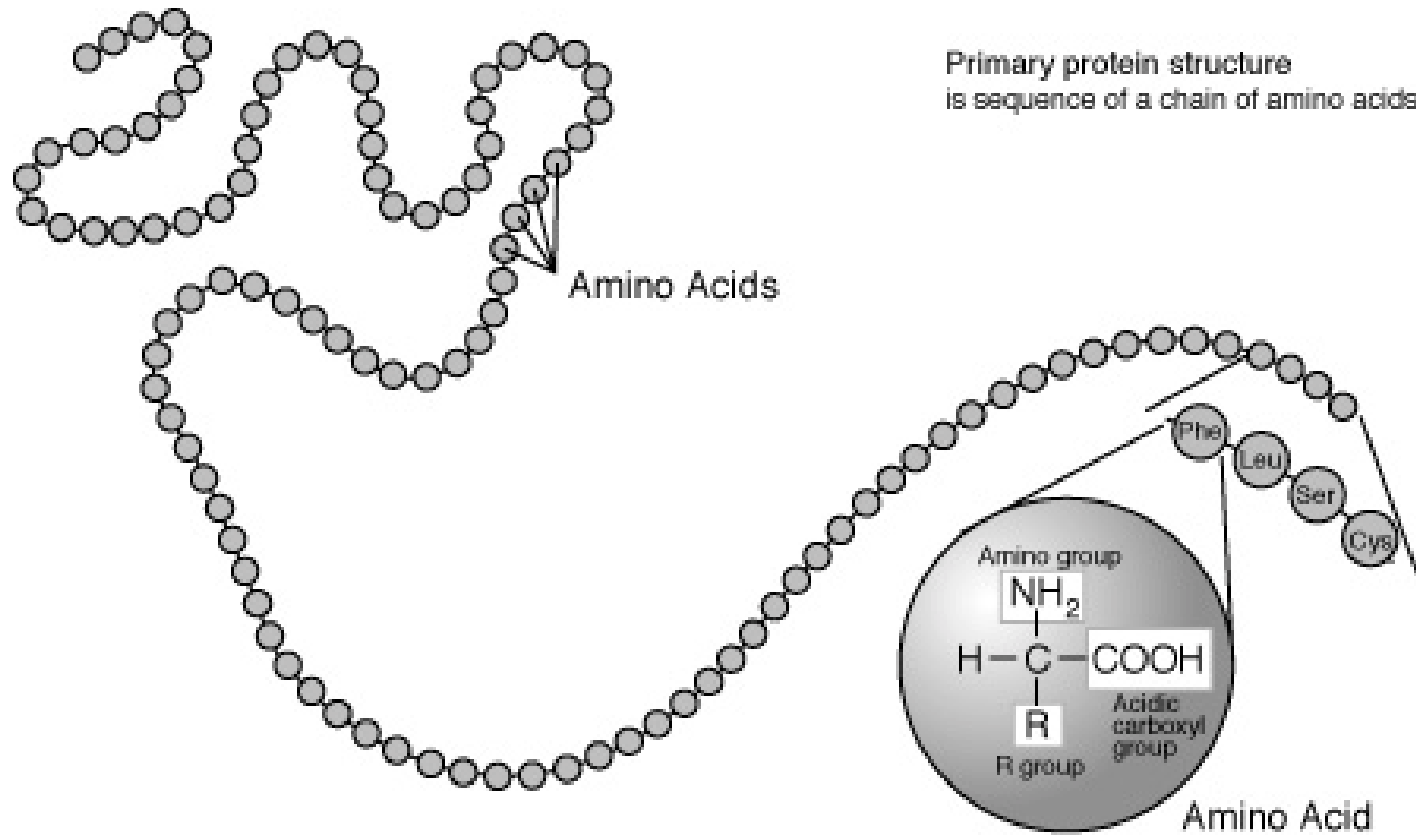


Cysteine
(cys)

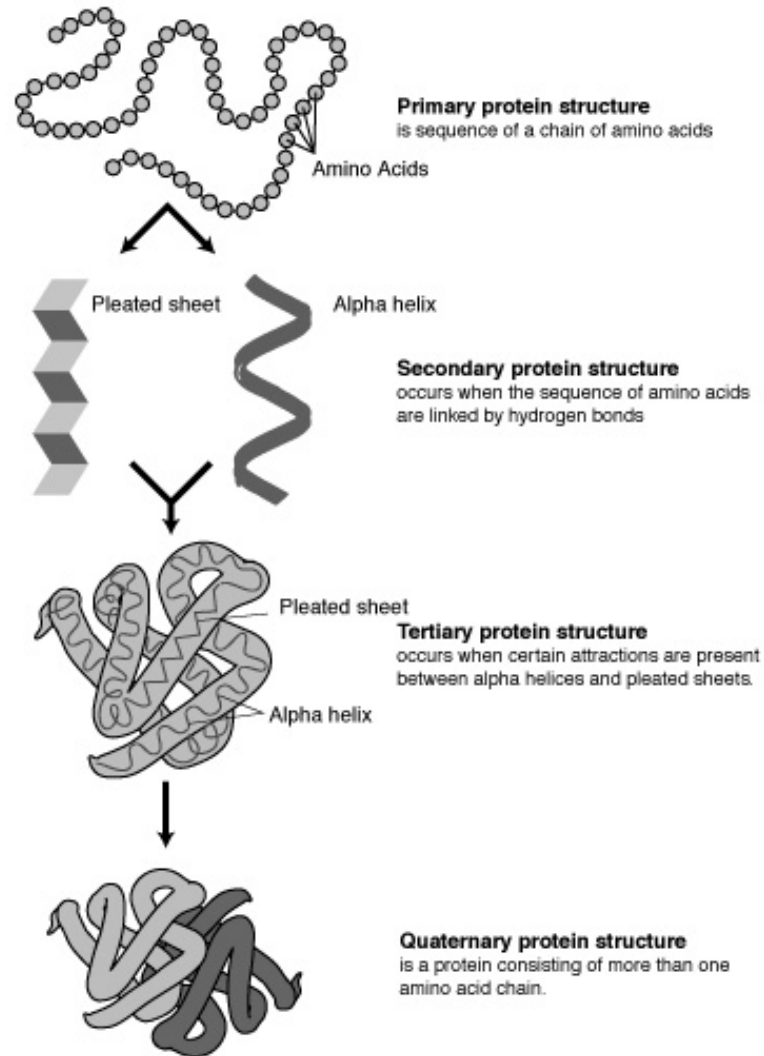


Proline
(pro)

Proteins



Proteins



Differential expression

- Each cell contains a complete copy of the organism's genome.
- Cells are of many different types and states
E.g. blood, nerve, and skin cells, dividing cells, cancerous cells, etc.
- What makes the cells different?
- **Differential gene expression**, i.e., **when**, **where**, and **how much** each gene is expressed.
- On average, 40% of our genes are expressed at any given time.

Central dogma

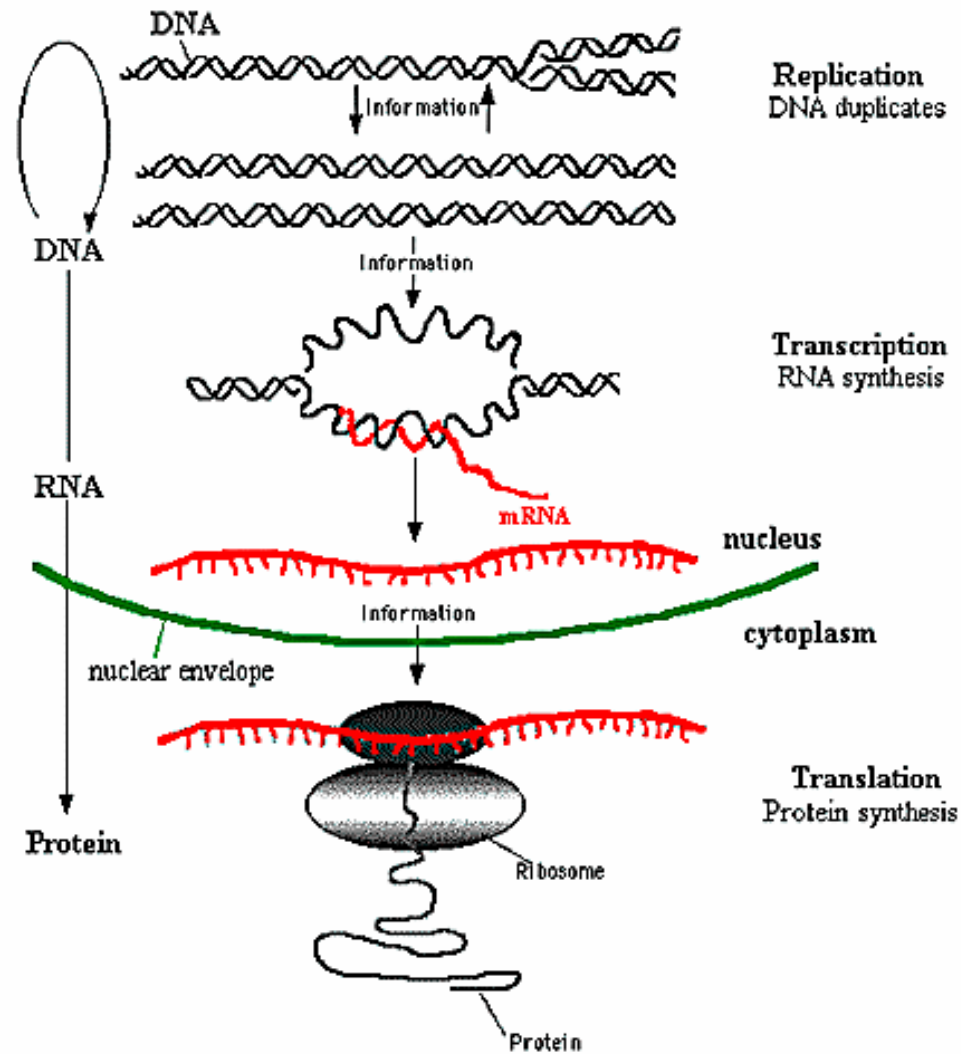
The **expression** of the genetic information stored in the DNA molecule occurs in two stages:

- (i) **transcription**, during which DNA is transcribed into mRNA;
- (ii) **translation**, during which mRNA is translated to produce a protein.

DNA → mRNA → protein

Other important aspects of regulation: methylation, alternative splicing, etc.

Central dogma

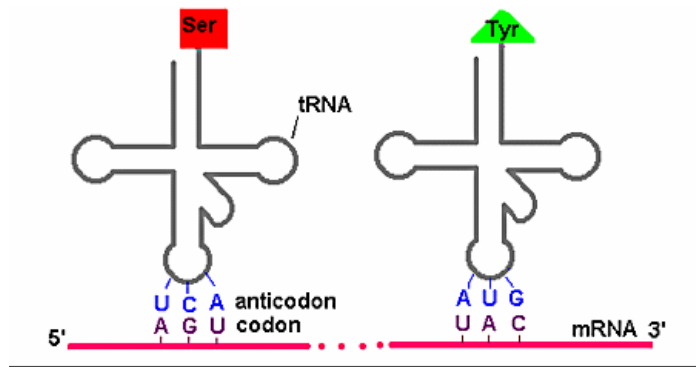


The Central Dogma of Molecular Biology

The genetic code

- **DNA:** sequence of **four** different nucleotides.
- **Proteins:** sequence of **twenty** different amino acids.
- The correspondence between DNA's four-letter alphabet and a protein's twenty-letter alphabet is specified by the **genetic code**, which relates nucleotide triplets or **codons** to **amino acids**.

The genetic code



		2nd base in codon					
		U	C	A	G		
1st base in codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G	3rd base in codon
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G	
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G	
	G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G	

The Genetic Code

Start codon: initiation of translation (AUG, Met).

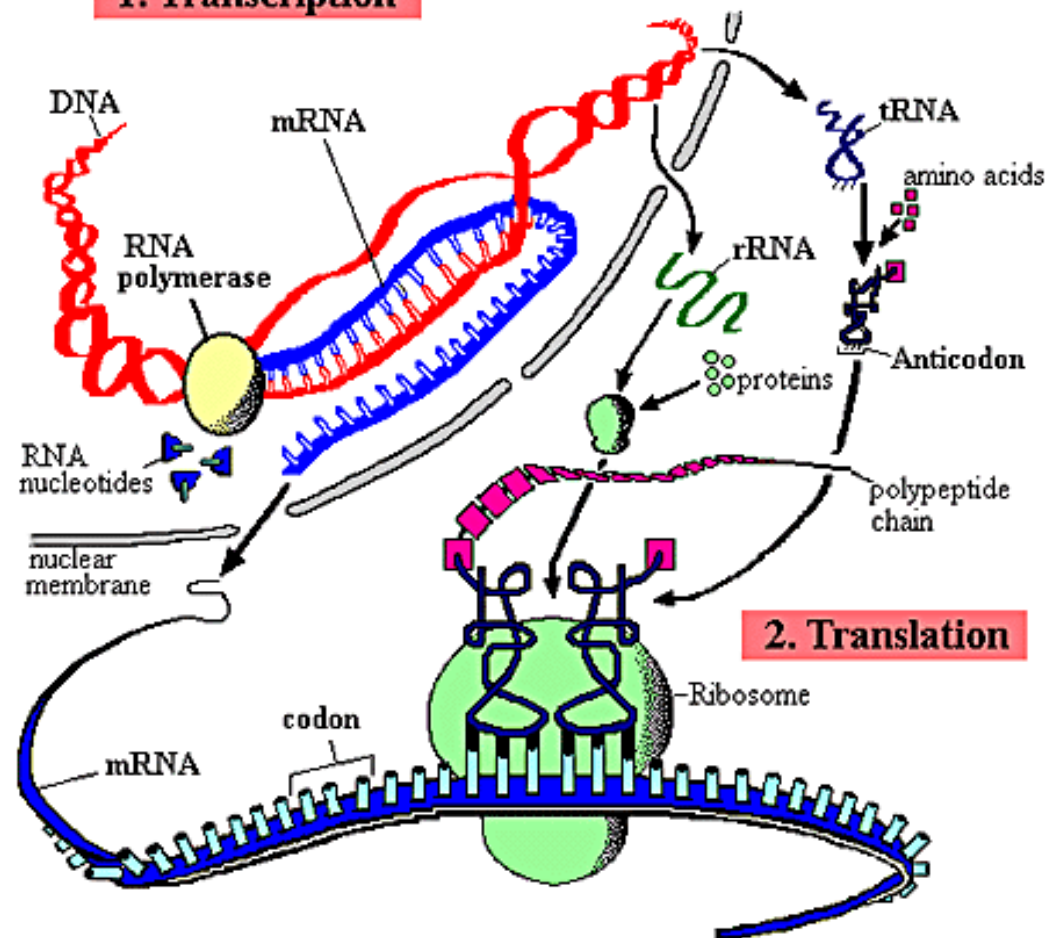
Stop codons: termination of translation.

Mapping between codons and amino acids is **many-to-one**: 64 codons but only 20 a.a..

Third base in codon is often redundant, e.g., stop codons.

Protein synthesis

1. Transcription



Protein synthesis

Transcription

- Analogous to DNA replication: several steps and many enzymes.
- **RNA polymerase** synthesizes an RNA strand complementary to one of the two DNA strands.
- The RNA polymerase recruits **rNTPs** (ribonucleotide triphosphate) in the same way that DNA polymerase recruits dNTPs (deoxynucleotide triphosphate).
- However, synthesis is **single stranded** and only proceeds in the 5' to 3' direction of mRNA (no Okazaki fragments).

Transcription

- The strand being transcribed is called the **template** or **antisense** strand; it contains **anticodons**.
- The other strand is called the **sense** or **coding** strand; it contains **codons**.
- The RNA strand newly synthesized from and complementary to the template contains the same information as the coding strand.

Transcription

5' ...A T G G C C T G G A C T T C A... 3' Sense strand of DNA
3' ...T A C C G G A C C T G A A G T... 5' Antisense strand of DNA



Transcription of antisense strand (5->3 direction)

5' ...A U G G C C U G G A C U U C A... 3' mRNA



Translation of mRNA

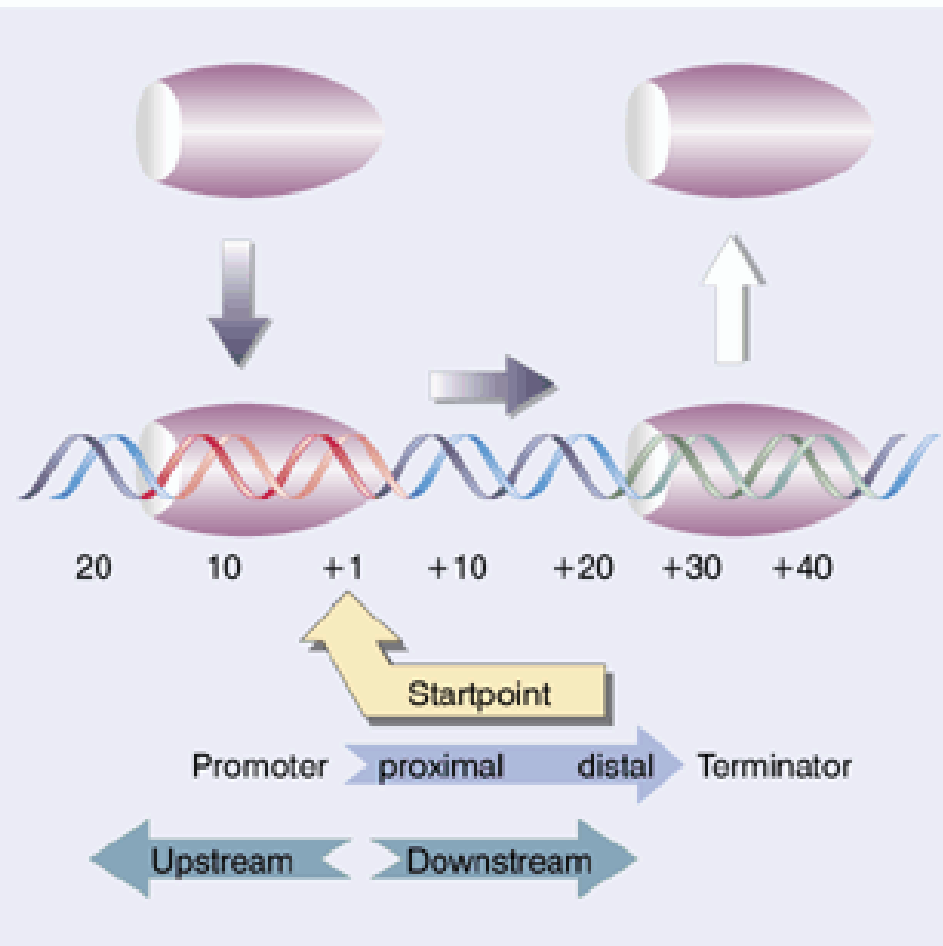
Met — Ala — Trp — Thr — Ser — Peptide

Transcription

- **Promoter.** Unidirectional sequence upstream of the coding region (i.e., at 5' end on sense strand) that tells the RNA polymerase both **where** to start and on **which strand** to continue synthesis. E.g. TATA box.
- **Terminator.** Regulatory DNA region signaling end of transcription, at 3' end .
- **Transcription factor.** A protein needed to initiate the transcription of a gene, binds either to specific DNA sequences (e.g. promoters) or to other transcription factors.

Transcription

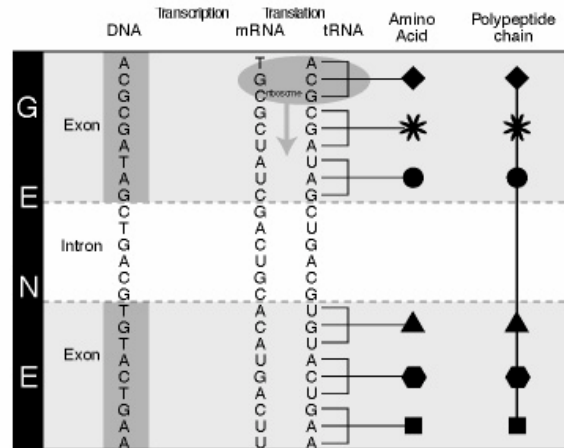
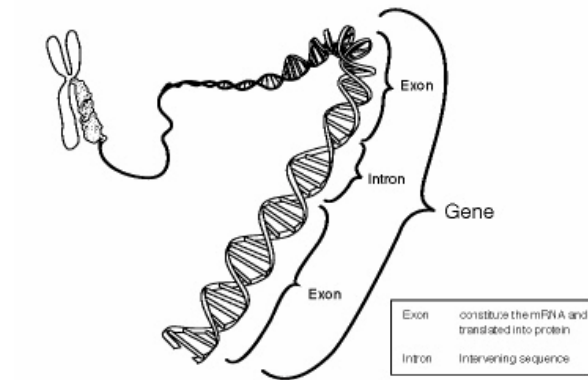
Figure 9.2 Overview: a transcription unit is a sequence of DNA transcribed into a single RNA, starting at the promoter and ending at the terminator.



Exons and introns

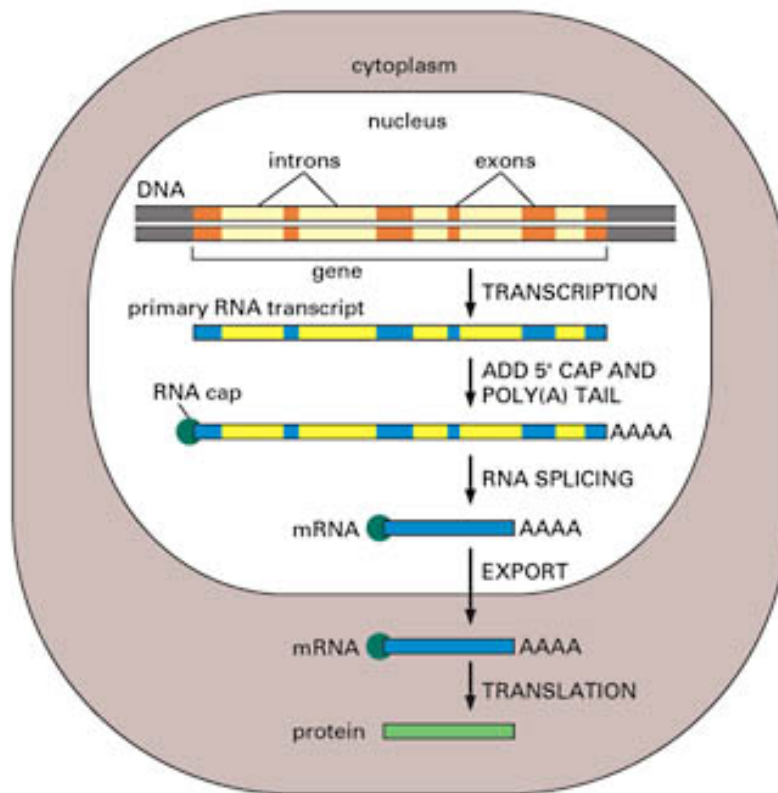
- Genes comprise only about 2% of the human genome.
- The rest consists of **non-coding** regions
 - chromosomal structural integrity,
 - cell division (e.g. centromere)
 - regulatory regions: regulating when, where, and in what quantity proteins are made .
- The terms **exon** and **intron** refer to coding (translated into a protein) and non-coding DNA, respectively.

Exons and introns



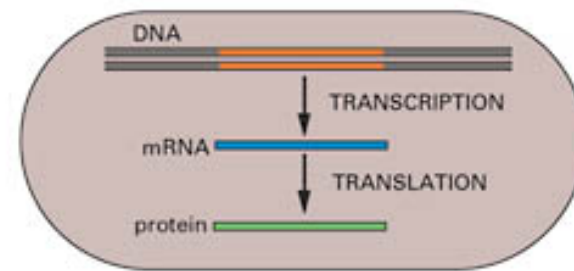
Splicing

(A) EUCARYOTES



Splicing

(B) PROCARYOTES



No splicing

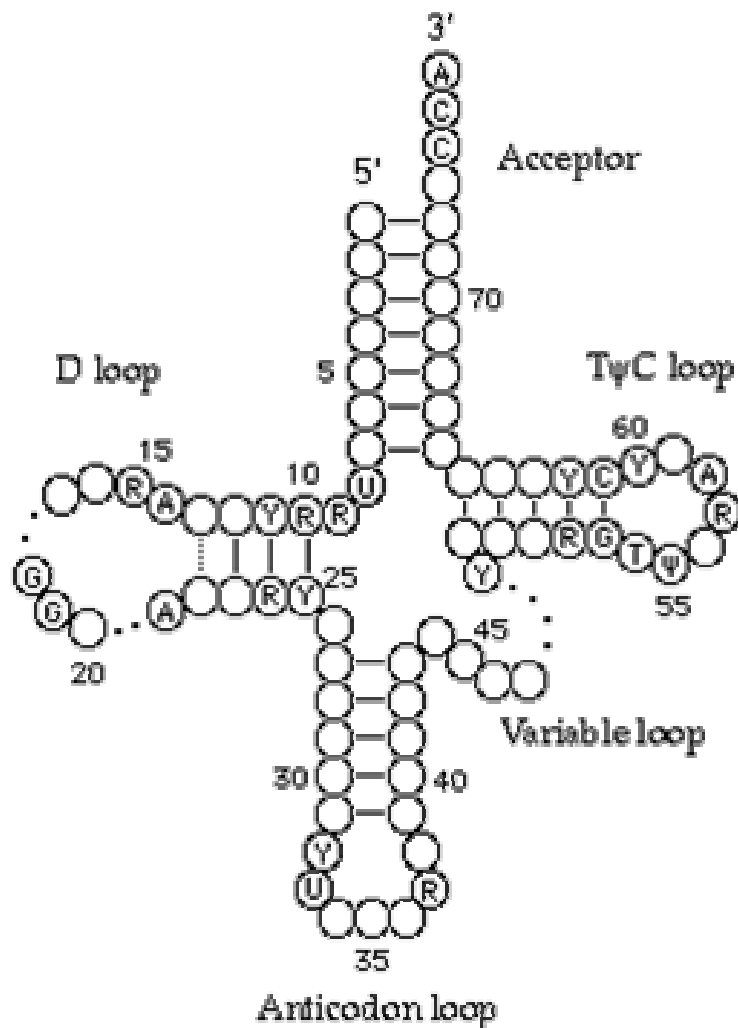
Translation

- **Ribosome:**
 - cellular factory responsible for protein synthesis;
 - a large subunit and a small subunit;
 - structural RNA and about 80 different proteins.
- **transfer RNA (tRNA):**
 - adaptor molecule, between mRNA and protein;
 - specific **anticodon** and **acceptor site**;
 - specific **charger protein**, can only bind to that particular tRNA and attach the correct amino acid to the acceptor site.

Translation

- Initiation
 - **Start codon AUG**, which codes for **methionine, Met**.
 - Not every protein necessarily starts with methionine. Often this first amino acid will be removed in post-translational processing of the protein.
- Termination:
 - **stop codon (UAA, UAG, UGA)** ,
 - ribosome breaks into its large and small subunits, releasing the new protein and the mRNA.

tRNA



- The tRNA has an **anticodon** on its mRNA-binding end that is complementary to the codon on the mRNA.
- Each tRNA only binds the appropriate amino acid for its anticodon.

Alternative splicing

- There are more than 1,000,000 different human antibodies. How is this possible with only ~30,000 genes?
- **Alternative splicing** refers to the different ways of combining a gene's exons. This can produce different forms of a protein for the same gene.
- Alternative pre-mRNA splicing is an important mechanism for regulating gene expression in higher eukaryotes.
- E.g. in humans, it is estimated that approximately 30% of the genes are subject to alternative splicing.

Alternative splicing



Primary isoform



Cryptic exon



Exon extension
(5' or 3')



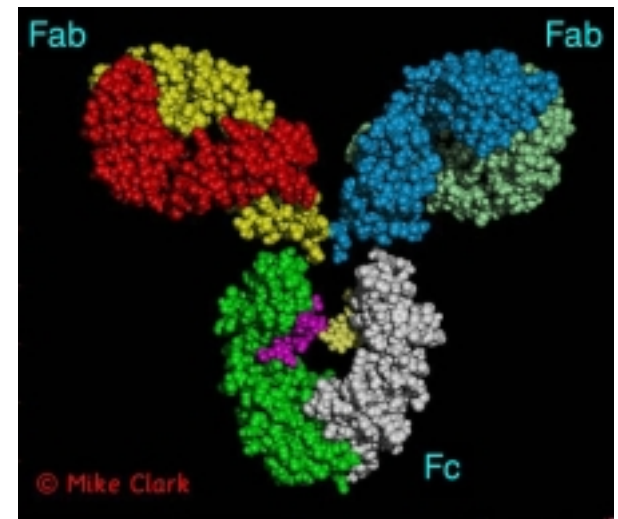
Exon skipping



Exon truncation

Immunoglobulin

- B cells produce antibody molecules called immunoglobulins (Ig) which fall in five broad classes.
- Diversity of Ig molecules
 - DNA sequence: recombination, mutation.
 - mRNA sequence: alternative splicing.
 - Protein structure: post-translational proteolysis, glycosylation.



IgG1

Post-translational processing

- Folding.
- Cleavage by a proteolytic (protein-cutting) enzyme.
- Alteration of amino acid residues
 - phosphorylation, e.g. of a tyrosine residue.
 - glycosylation, carbohydrates covalently attached to asparagine residue.
 - methylation, e.g. of arginine.
- Lipid conjugation.

Functional genomics

- The various **genome projects** have yielded the complete DNA sequences of many organisms.

E.g. human, mouse, yeast, fruitfly, etc.

Human: 3 billion base-pairs, 30-40 thousand genes.

- Challenge: **go from sequence to function**, i.e., define the role of each gene and understand how the genome functions as a whole.

WWW resources

- **Access Excellence**
<http://www.accessexcellence.com/AB/GG/>
- **Genes VII**
<http://www.oup.co.uk/best.textbooks/biochemistry/genesvii/>
- **Human Genome Project Education Resources**
<http://www.ornl.gov/hgmis/education/education.html>
- **Kimball's Biology Pages**
<http://www.ultranet.com/~jkimball/BiologyPages/>
- **MIT Biology Hypertextbook**
<http://esg-www.mit.edu:8001/>