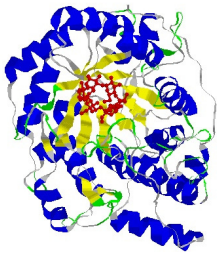
Molecular Biology

Sami Khuri
Department of Computer Science
San José State University
San José, California, USA
sami.khuri@sjsu.edu
www.cs.sjsu.edu/faculty/khuri

Biology Review

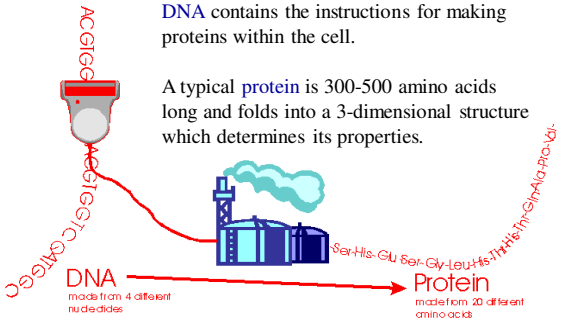


- DNA
- RNA
- Proteins
- Central Dogma
- Transcription
- Translation

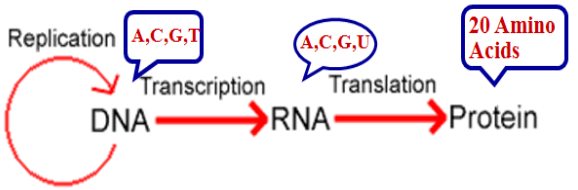
Protein Factory

DNA contains the instructions for making proteins within the cell.

A typical protein is 300-500 amino acids long and folds into a 3-dimensional structure which determines its properties.

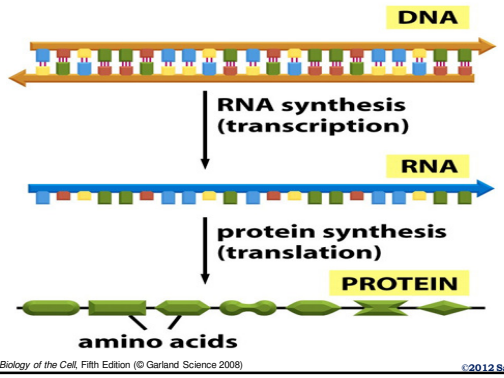


DNA → RNA → Protein



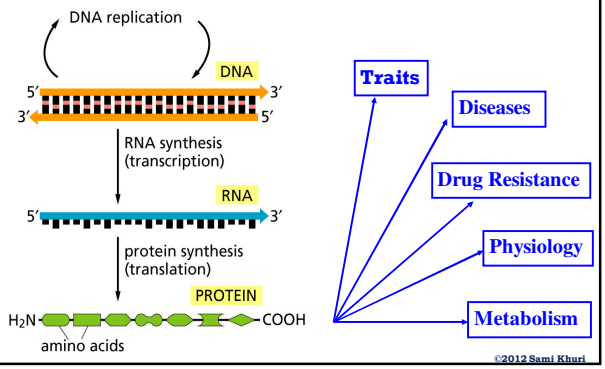
Adenine	(A)	Adenine	(A)	{A-Y}-{BJOUX}
Guanine	(G)	Guanine	(G)	
Cytosine	(C)	Cytosine	(C)	
Thymine	(T)	Uracil	(U)	

Central Dogma of Molecular Biology



Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Central Dogma of Molecular Biology



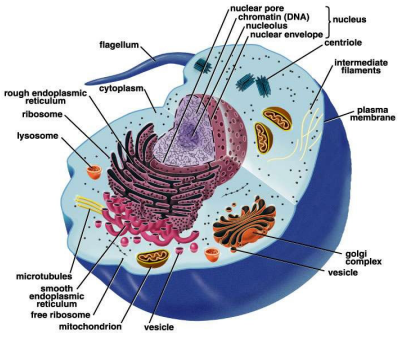
Prokaryotes and Eukaryotes

A **cell** is the fundamental working unit of every living organism.

There are two kinds of cells:

- **prokaryotes**, which are single-celled organisms with no cell nucleus: archea and bacteria.
- **eukaryotes**, which are higher level organisms, and their cells have nuclei: animals and plants.

Eukaryotic Cell



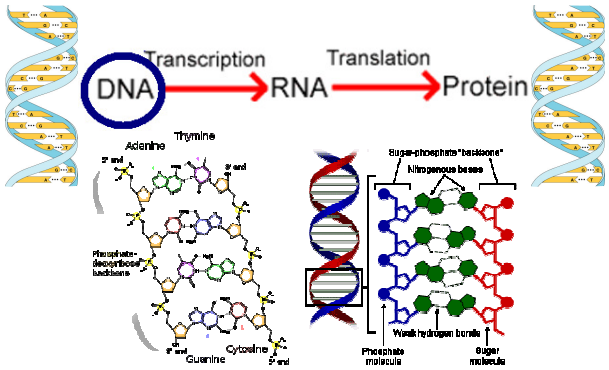
A **cell** carries the entire set of genetic instructions: the genome, that makes an entire organism.
The **instructions** are encoded in DNA as genes and packaged as chromosomes in the nucleus.

Proteins and Nucleic Acids

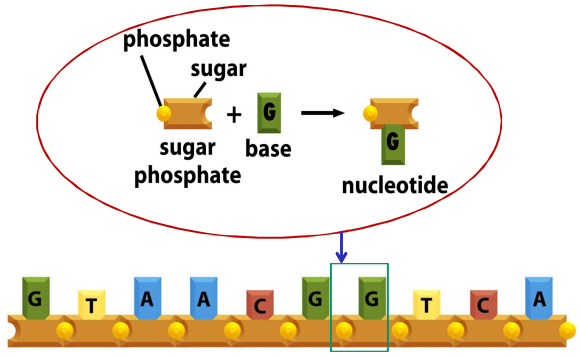
All living organisms have a similar molecular chemistry.
The main actors in the chemistry of life are molecules:

- **proteins**: which are responsible for what a living being is and does in a physical sense.
  "We are our proteins" R. Doolittle.
- **nucleic acids**: which encode the information necessary to produce proteins and are responsible for passing the "recipe" to subsequent generations.
  Living organisms contain 2 kinds of nucleic acids:
- **Ribonucleic acid** (**RNA**)
- **Deoxyribonucleic acid** (**DNA**)

DNA

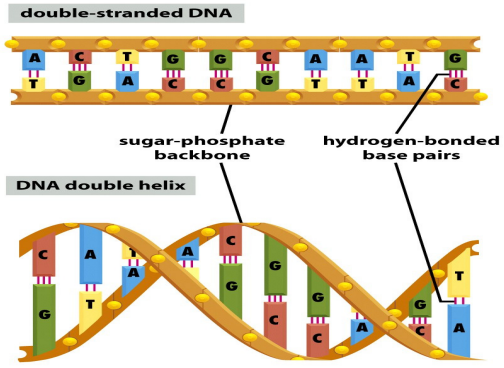


DNA Strand



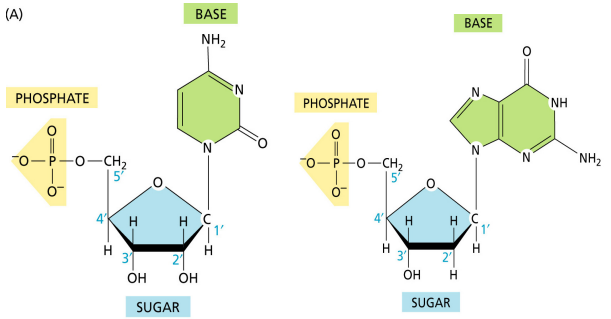
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Double-Stranded DNA



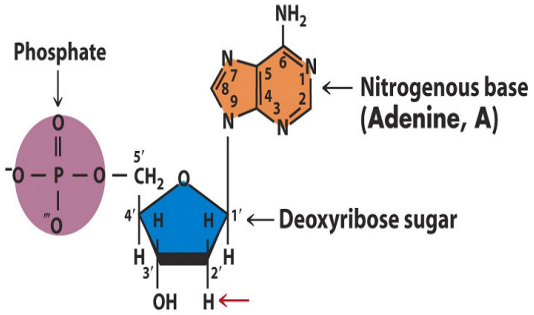
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Sugars Found in Nucleic Acids

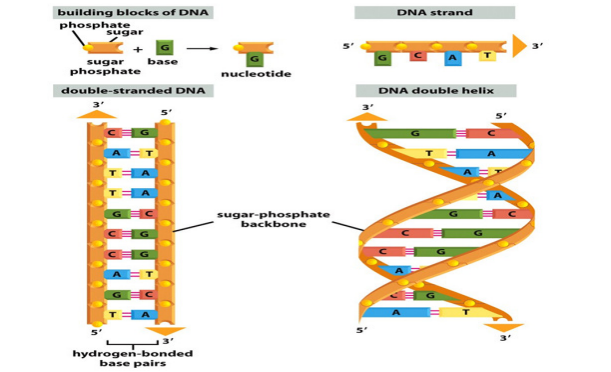


Pentose sugar present in **RNA** Pentose sugar present in **DNA**

A Nucleotide



Double-Stranded DNA

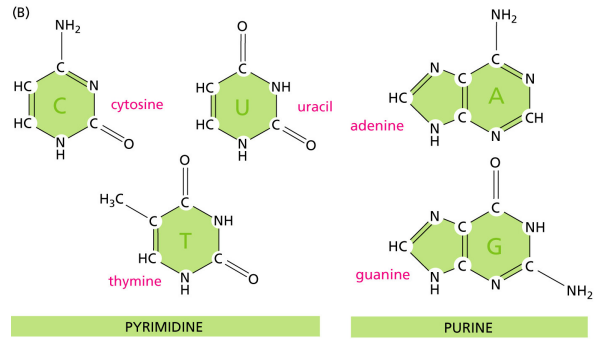


Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

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**).

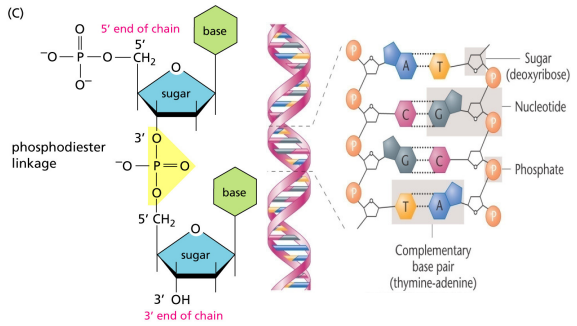
Purines and Pyrimidines



Double Helix

- The binding of two nucleotides forms a base pair.
- The double helix is formed by connecting complementary nucleotides A-T and C-G on two strands with hydrogen bonds.
- Knowledge of the sequence on one strand allows us to infer the sequence of the other strand.
- The bases are arranged along the sugar phosphate backbone in a particular order, known as the DNA sequence, encoding all genetic instructions for an organism.

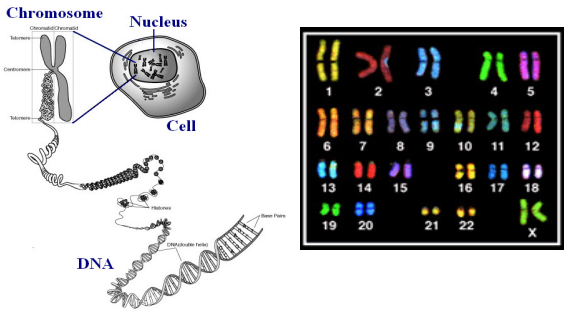
DNA Phosphodiester Backbone



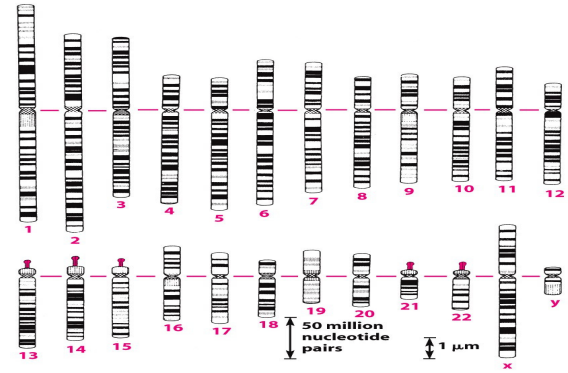
DNA and Chromosomes

- The **genome** is a complete set of instructions for making an organism, consists of tightly coiled threads of **DNA** organized into structures called **chromosomes**.
- Besides the reproductive cell and red blood cell, every single cell in the human body contains the human genome.

Chromosomes and Genome



Human Chromosomes



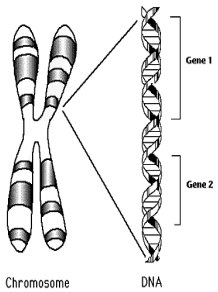
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Pairs of Chromosomes in Species

Table 3-2 Numbers of Pairs of Chromosomes in Different Species of Plants and Animals

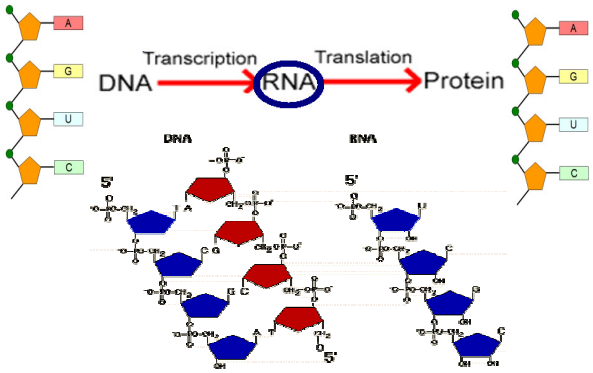
Common name	Scientific name	Number of chromosome pairs	Common name	Scientific name	Number of chromosome pairs
Mosquito	*Culex pipiens*	3	Wheat	*Triticum aestivum*	21
Housefly	*Musca domestica*	6	Human	*Homo sapiens*	23
Garden onion	*Allium cepa*	8	Potato	*Solanum tuberosum*	24
Toad	*Bufo americanus*	11	Cattle	*Bos taurus*	30
Rice	*Oryza sativa*	12	Donkey	*Equus asinus*	31
Frog	*Rana pipiens*	13	Horse	*Equus caballus*	32
Alligator	*Alligator mississipiensis*	16	Dog	*Canis familiaris*	39
Cat	*Felis domesticus*	19	Chicken	*Gallus domesticus*	39
House mouse	*Mus musculus*	20	Carp	*Cyprinus carpio*	52
Rhesus monkey	*Macaca mulatta*	21			

Genes



- A **gene** is a specific sequence of nucleotide bases along a chromosome carrying information for constructing a protein. A gene encodes a protein (or an RNA).
- The distance between **genes** is often much larger than the genes themselves.
- The human genome has around 25,000 genes.

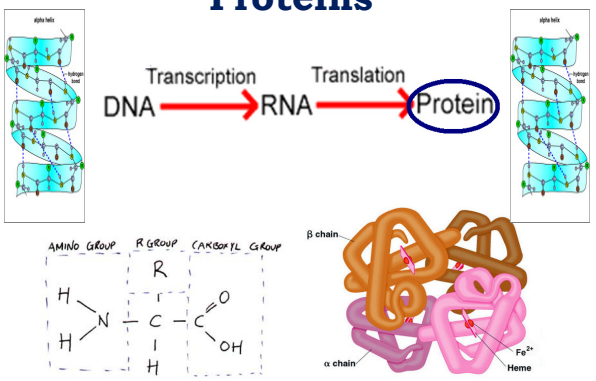
RNA



Ribonucleic Acid - RNA

- **RNA** is found in the cell and can also carry genetic information.
- While DNA is located primarily in the nucleus, **RNA** can also be found in the cytoplasm.
- **RNA** is built from the nucleotides cytosine, guanine, adenine and uracil (U) (instead of thymine).
- **RNA** has its sugar phosphate backbone containing ribose.
- **RNA** forms a single strand.
- **RNA** molecules tend to have a less-regular three-dimensional structure than DNA.

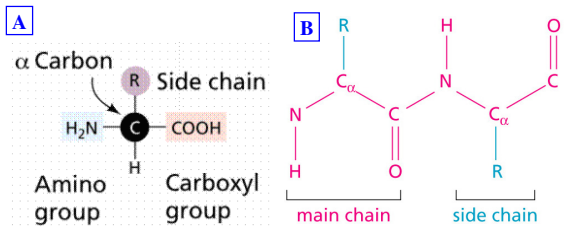
Proteins



Proteins

- 20 different **amino acids** are used to synthesize **proteins**.
- The shape and other properties of each **protein** is dictated by the precise sequence of **amino acids** in it.
- The function of a **protein** is determined by its unique three-dimensional structure.

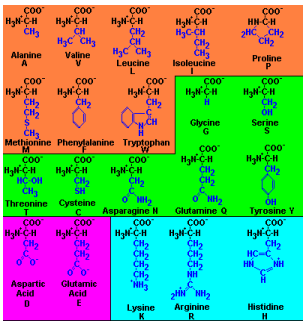
Structure of the Amino Acid



A) The functional properties of proteins are almost entirely due to the side chains in the amino acids.

B) Two amino acids: the main chain is in red and the side chain in blue.

The Twenty Amino Acids



Orange:
nonpolar and hydrophobic.

The other amino acids are: polar and hydrophilic - "water loving".

Magenta:
acidic - "carboxy" group in the side chain.

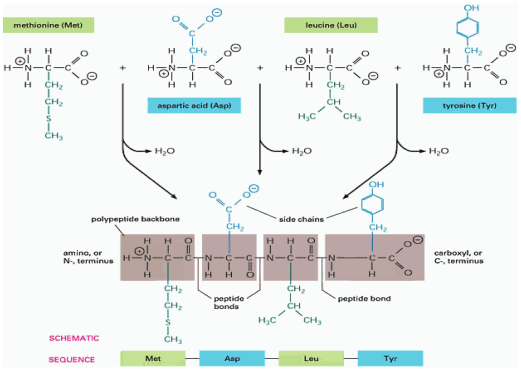
Light blue:
basic - "amine" group in the side chain.

The 20 Amino Acids

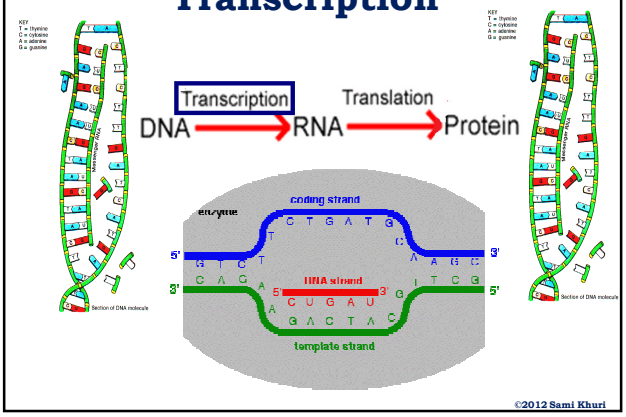
1-letter	3-letter	Amino acid	1-letter	3-letter	Amino Acid
A	Ala	Alanine	M	Met	Methionin
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic Acid	P	Pro	Proline
E	Glu	Glutamic Acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonin
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

Patrice Koehl

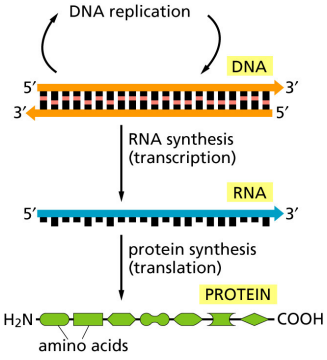
Protein Structure



Transcription

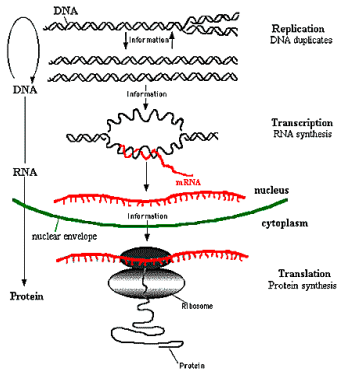


Central Dogma of Molecular Biology



According to the **central dogma of molecular biology**, there is a single direction of flow of genetic information from the **DNA**, which acts as the information store, through **RNA** molecules from which the information is translated into **proteins**.

Steps of the Central Dogma

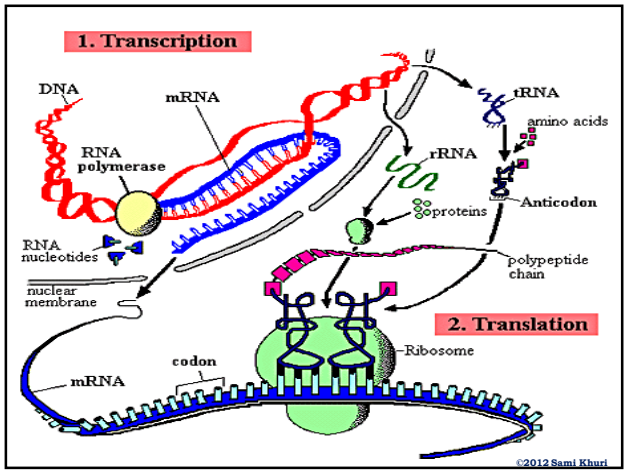


Genetic information embodied in DNA is **replicated** into more DNA

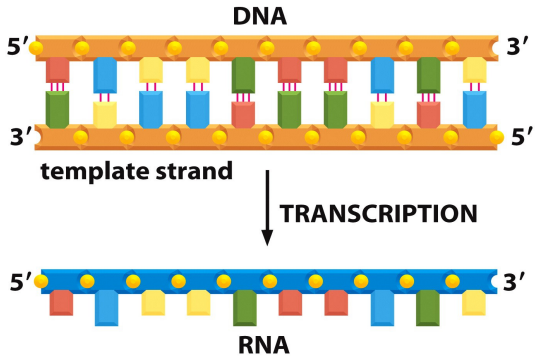
The synthesis of an RNA from a sequence of DNA. The resulting RNA is **mRNA**.

In eukaryotic cells, the mRNA is **spliced** and it migrates from the nucleus to the cytoplasm.

Messenger RNA carries coded information to ribosomes that "read"and use it for **protein synthesis**.

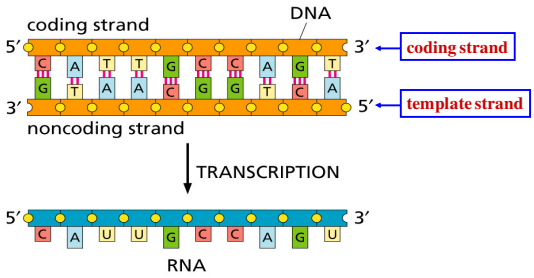


Transcription



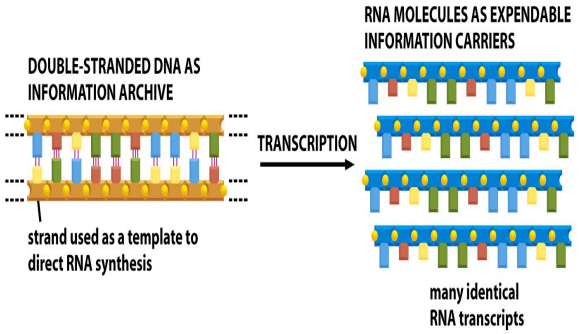
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Transcription



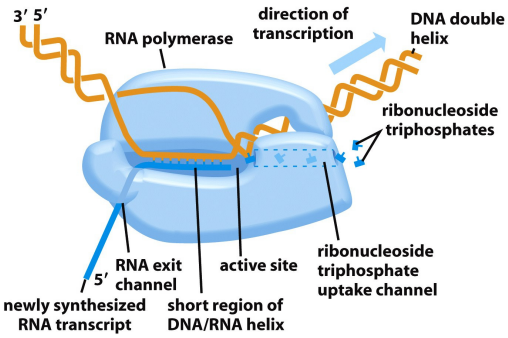
Transcription is the process in which one DNA strand: the **template strand**, is used to synthesize a complementary RNA.

RNA: Carrier of Information



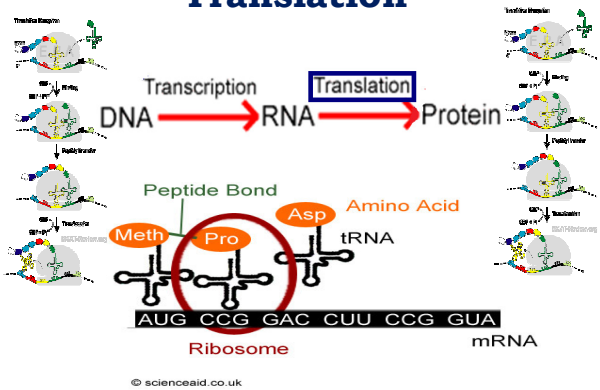
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Synthesizing RNA from 5' to 3'



Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Translation



The Genetic Code

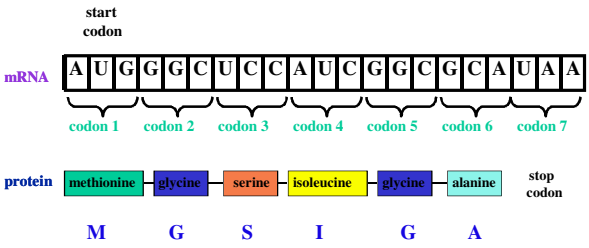
FIRST BASE	SECOND BASE U	SECOND BASE C	SECOND BASE A	SECOND BASE G	THIRD BASE
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

The Genetic Code

The Genetic Code

First Codon Position (5' End)	Second Codon Position U	Second Codon Position C	Second Codon Position A	Second Codon Position G	Third Codon Position (3' End)
U	UUU Phe (F)	UCU Ser (S)	UAU Tyr (Y)	UGU Cys (C)	U
	UUC Phe (F)	UCC Ser (S)	UAC Tyr (Y)	UGC Cys (C)	C
	UUA Leu (L)	UCA Ser (S)	UAA *Stop*	UGA *Stop*	A
	UUG Leu (L)	UCG Ser (S)	UAG *Stop*	UGG Trp (W)	G
C	CUU Leu (L)	CCU Pro (P)	CAU His (H)	CGU Arg (R)	U
	CUC Leu (L)	CCC Pro (P)	CAC His (H)	CGC Arg (R)	C
	CUA Leu (L)	CCA Pro (P)	CAA Gln (Q)	CGA Arg (R)	A
	CUG Leu (L)	CCG Pro (P)	CAG Gln (Q)	CGG Arg (R)	G
A	AUU Ile (I)	ACU Thr (T)	AAU Asn (N)	AGU Ser (S)	U
	AUC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)	C
	AUA Ile (I)	ACA Thr (T)	AAA Lys (K)	AGA Arg (R)	A
	AUG Met (M)	ACG Thr (T)	AAG Lys (K)	AGG Arg (R)	G
G	GUU Val (V)	GCU Ala (A)	GAU Asp (D)	GGU Gly (G)	U
	GUC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)	C
	GUA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)	A
	GUG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)	G

Translation: An Example



Transfer RNA and Translation

- The translation from nucleotides to amino acid is done by means of **transfer RNA** (**tRNA**) molecules, each specific for one amino acid and for a particular **triplet** of nucleotides in mRNA called a **codon**.
- The family of tRNA molecules enables the codons in a mRNA molecule to be **translated** into the sequence of amino acids in the protein.

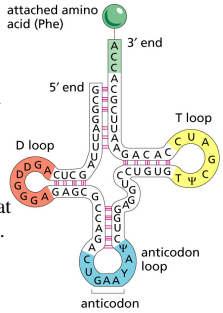
Codons and Anticodons

At least one kind of tRNA is present for each of the 20 amino acids used in protein synthesis.

Each kind of tRNA has a sequence of 3 unpaired nucleotides - the anticodon - which can bind to the complementary triplet of nucleotides - the codon - in an mRNA molecule.

The reading of codons in mRNA requires that the anticodons bind in the opposite direction.

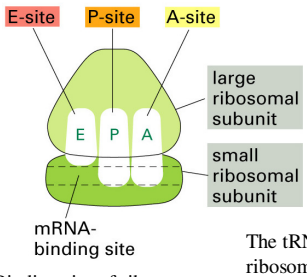
Anticodon: 3' AAG 5'
Codon: 5' UUC 3'



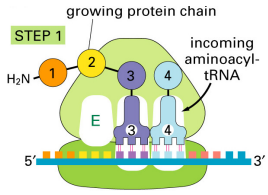
Start and Stop Codons

- The codon AUG serves two related functions
  - It begins most messages; that is, it signals the start of translation placing the amino acid methionine at the amino terminal of the polypeptide to be synthesized.
  - When it occurs within the message, it guides the incorporation of methionine.
- Three **codons**, UAA, UAG, and UGA, act as signals to terminate translation. They are called **STOP codons**.

Translation



Binding site of ribosome for the mRNA and the three tRNA binding sites.



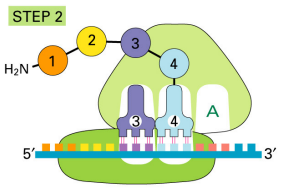
The tRNA molecules bind to the ribosome and are the physical link between the mRNA and the growing protein chain.

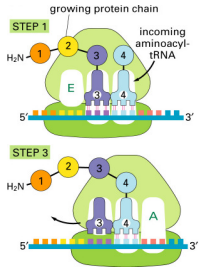
Steps of Translation: Initiation

- The small subunit of the ribosome binds to a site "upstream" of the start of the message.
- It proceeds downstream until it encounters the start codon AUG.
- It is then joined by the large subunit and a special initiator tRNA. The initiator tRNA binds to the P site on the ribosome.
- In eukaryotes, initiator tRNA generally carries methionine (Met).

Steps of Translation: Elongation

An aminoacyl-tRNA able to base pair with the next codon on the mRNA arrives at the A site.



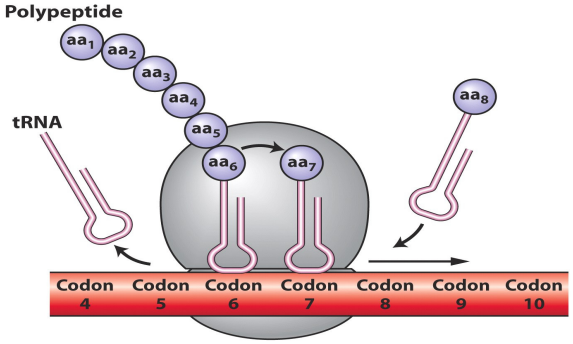


The preceding amino acid is linked to the incoming amino acid with a peptide bond.

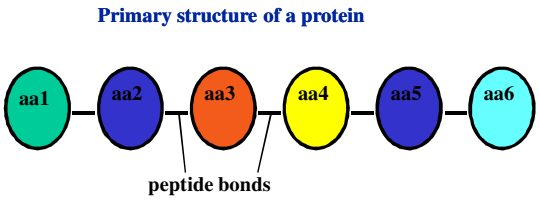
Steps of Translation: Termination

- The end of the message is marked by a STOP codon: UAA, UAG, UGA.
- No tRNA molecules have anticodons for STOP codons. A protein release factor recognizes these codons when they arrive at the A site.
- Binding of this protein releases the polypeptide from the ribosome.
- The ribosome splits into its subunits, which can later be reassembled for another round of protein synthesis.

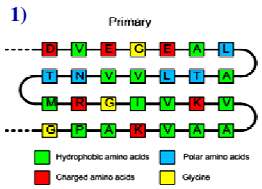
Chain of Amino Acids

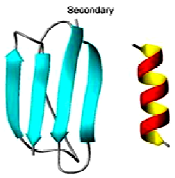


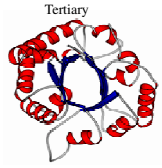
Primary Structure of Protein

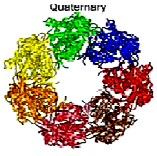


Four Structures of Proteins

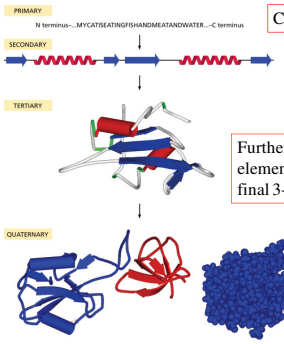








Four Levels of Protein Structure



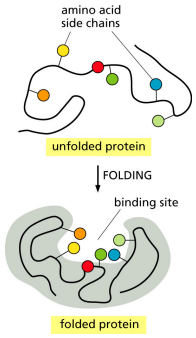
Chain of amino acids

First level of protein folding. Parts of the chain fold to form genetic structures found in all proteins

Further folding and packing together of the elements of secondary structure to produce the final 3-D conformation unique to the protein

Multisubunit protein formed of more than one protein chain. Each protein chain is called a subunit

Folded Protein



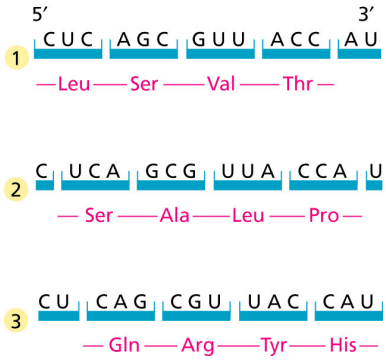
Distant residues (in the primary structure of a protein) can come close in the folded structure.

When a polypeptide chain (primary structure) folds into a tertiary structure, residues that are far apart from each other in the sequence can come close together to form a binding site.

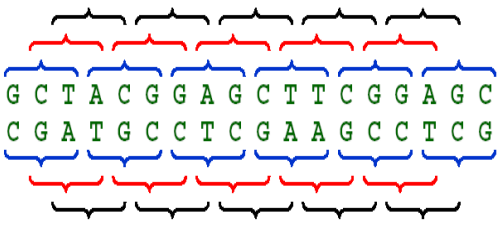
Structure and Function Relationship

- Knowing the relationship between a protein's structure and its function provides a greater understanding of how modifying the structure will affect the function.
- As the vast majority of currently marketed pharmaceuticals act by interacting with proteins, structure-function studies are vital to the design of new drugs, and bioinformatics has an important role in speeding up this process and enabling computer modeling of these interactions.

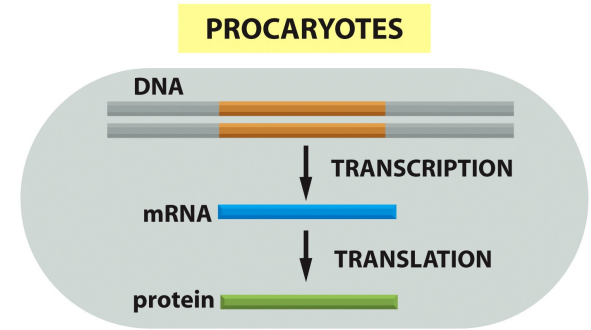
3 Reading Frames of mRNA



Six Reading Frames

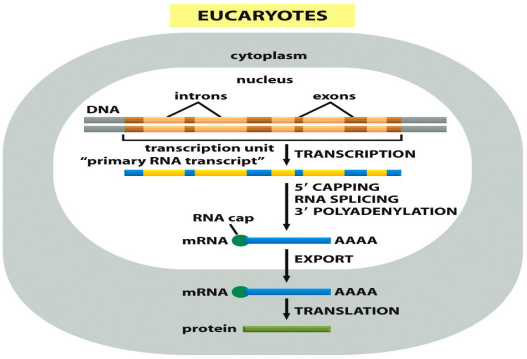


Dogma Revisited: Prokaryotes



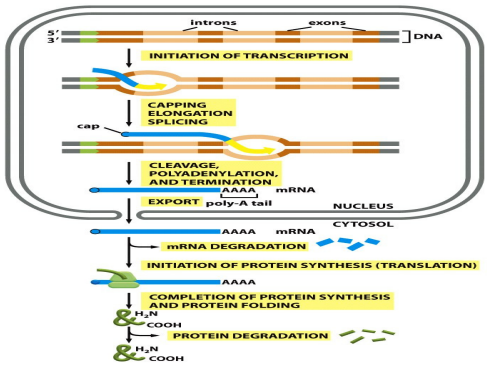
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008)

Dogma Revisited: Eukaryotes



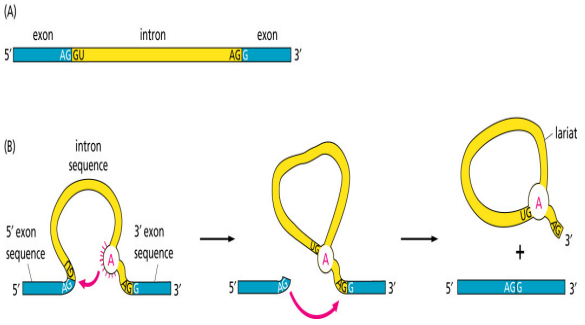
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008) ©2012 Sami Khuri

Dogma Re-revisited



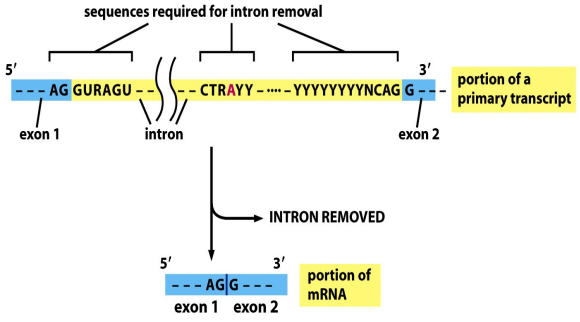
Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008) ©2012 Sami Khuri

The Splicing of an Intron



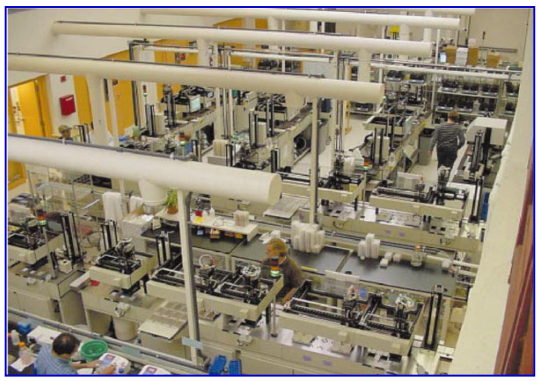
©2012 Sami Khuri

Splicing in Eukaryotes



Molecular Biology of the Cell, Fifth Edition (© Garland Science 2008) ©2012 Sami Khuri

The New Face of Biology



©2012 Sami Khuri

Sequencing SARS

http://www.bcgsc.ca/bioinfo/SARS ©2012 Sami Khuri