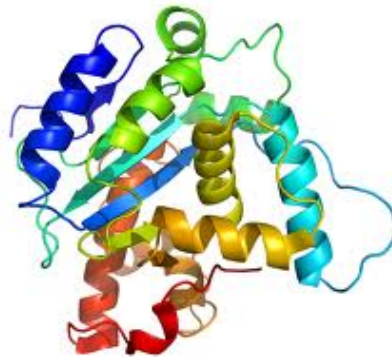


Computational Modeling of Molecular Structure

Jianlin Cheng, PhD

Computer Science Department

Informatics Institute

University of Missouri, Columbia

Spring, 2013

Objectives

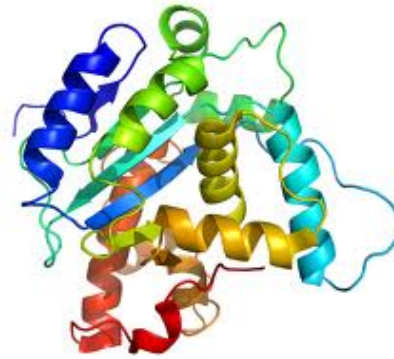
- Properties of molecular structures (proteins, RNA, genome / DNA)
- Computational representation of molecular structures
- Computational modeling of molecular structures
- Application of modeling of molecular structures

Significance of Studying Molecular Structures

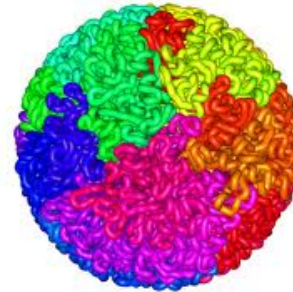
- One foundation of life sciences
- Personal healthcare and medicine
- One major topic of bioinformatics and computational biology – an important field of computer science
- **A great application area of computer algorithms and data structures**
- **A great application area of engineering**
- **A very interdisciplinary field (CS, math, biology, chemistry, physics)**

Three Kinds of Structures

- Protein Structure



- Genome Structure

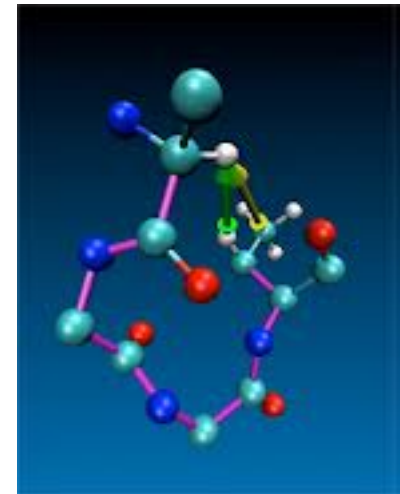
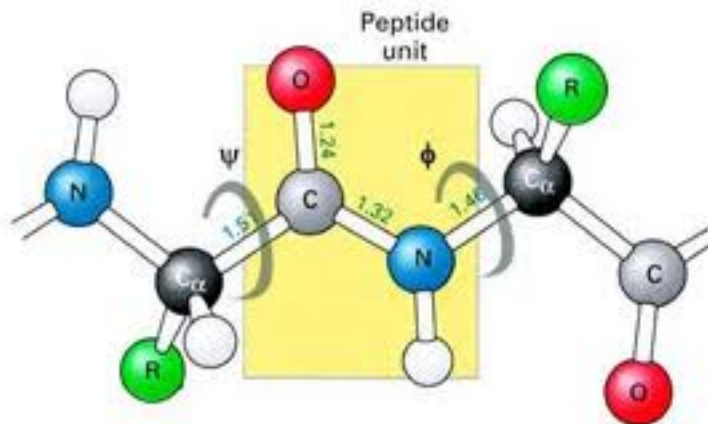
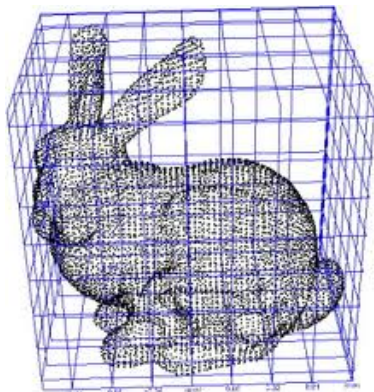
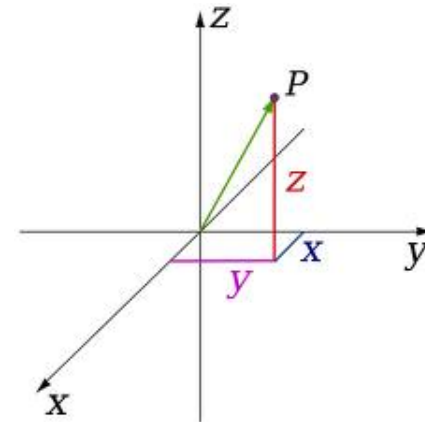


- RNA Structure



Representation of Molecular Structures

- X, Y, Z coordinates
- Euclidean grid
- Vector and angles
- Computer graphics



Algorithms

- Grid-based simulation (random walk)
- Vector-based simulation
- Angular-based simulation
- Gradient descent simulation and variants
- Simulated annealing
- Markov Chain Monte Carlo
- Probabilistic modeling
- Constraint-based optimization

Software Packages

- RasMol, Jmol, PyMol
- Modeller, Rosetta, I-TASSER, IMP, CNS, Tinker, etc
- Your own algorithm, implementation, and practice

Course Format

- Course web site:
<http://people.cs.missouri.edu/~chengji/cscmms/>
- Username and password: **cmms**
- Problem solving
- Active learning by practicing
- Syllabus (see details)

Teaching Format of Each Topic



Group:

4 – 5 students per group

Rotate as topic coordinator

**Each member participates
in every topic**

**All members present
the whole project**

Grading

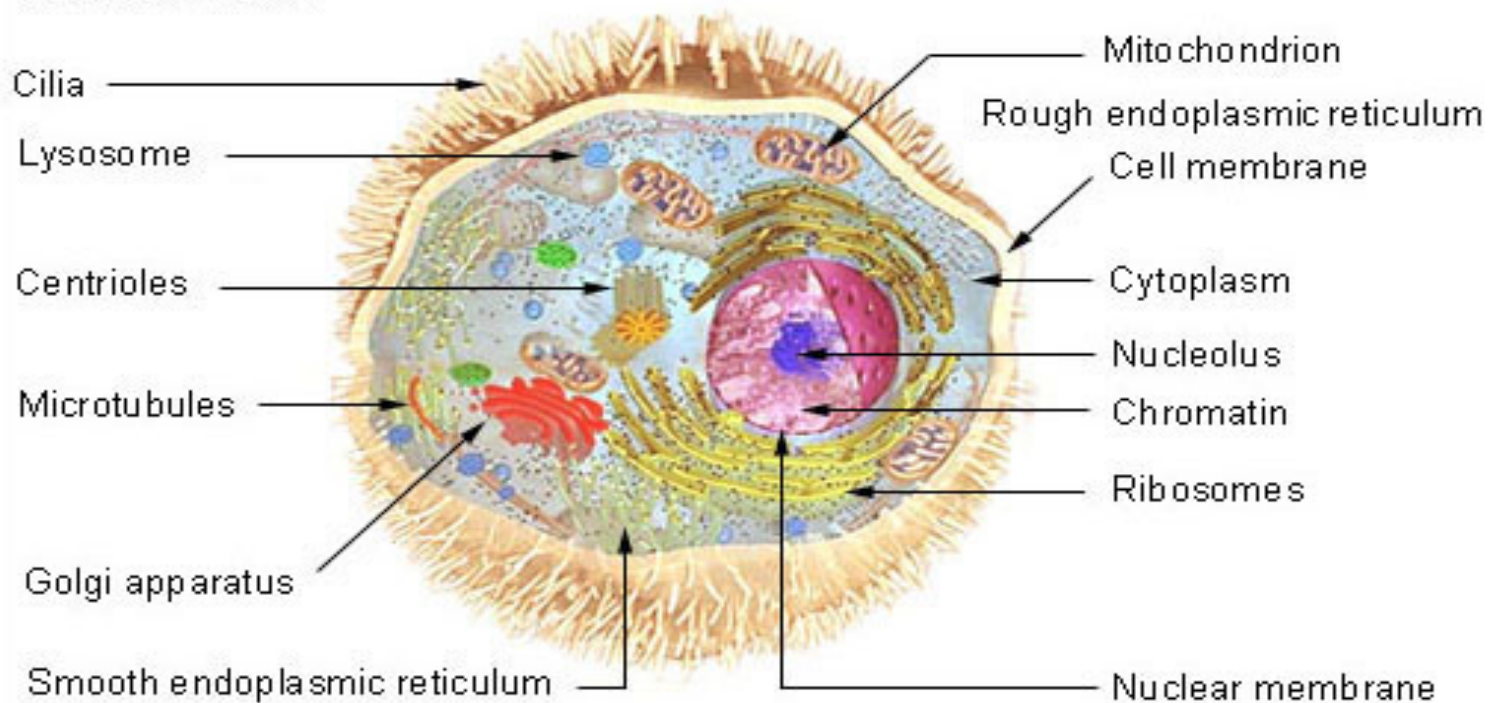
- Class discussions (15%)
- Literature reviews (10%)
- Topic plan presentation (20%, group)
- Topic implementation (25%, group)
- Topic report (20%, group)
- Final presentation (10%, group)
- Grade scale: A+, A, A-, B+, B, B-, C+, C, C-, and F.

Introduction to Molecular Biology for Computer Science and Engineering Students

Introduction to Molecular Biology

- Cell is the unit of structure and function of all living things.

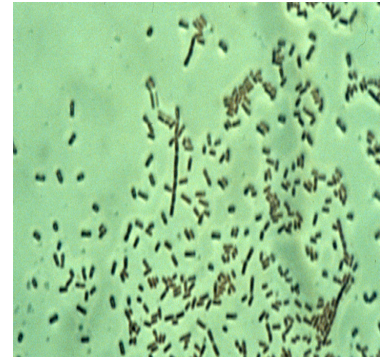
Cell Structure



Two types of cells: eukaryote (higher organisms) and prokaryote (lower organisms)

Central Dogma of Molecular Biology

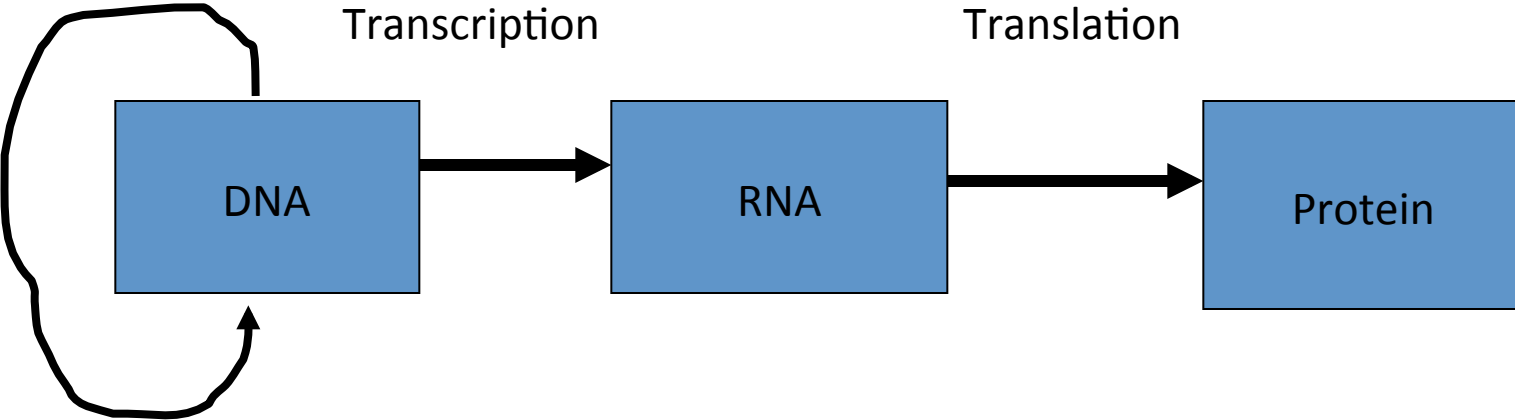
Phenotype



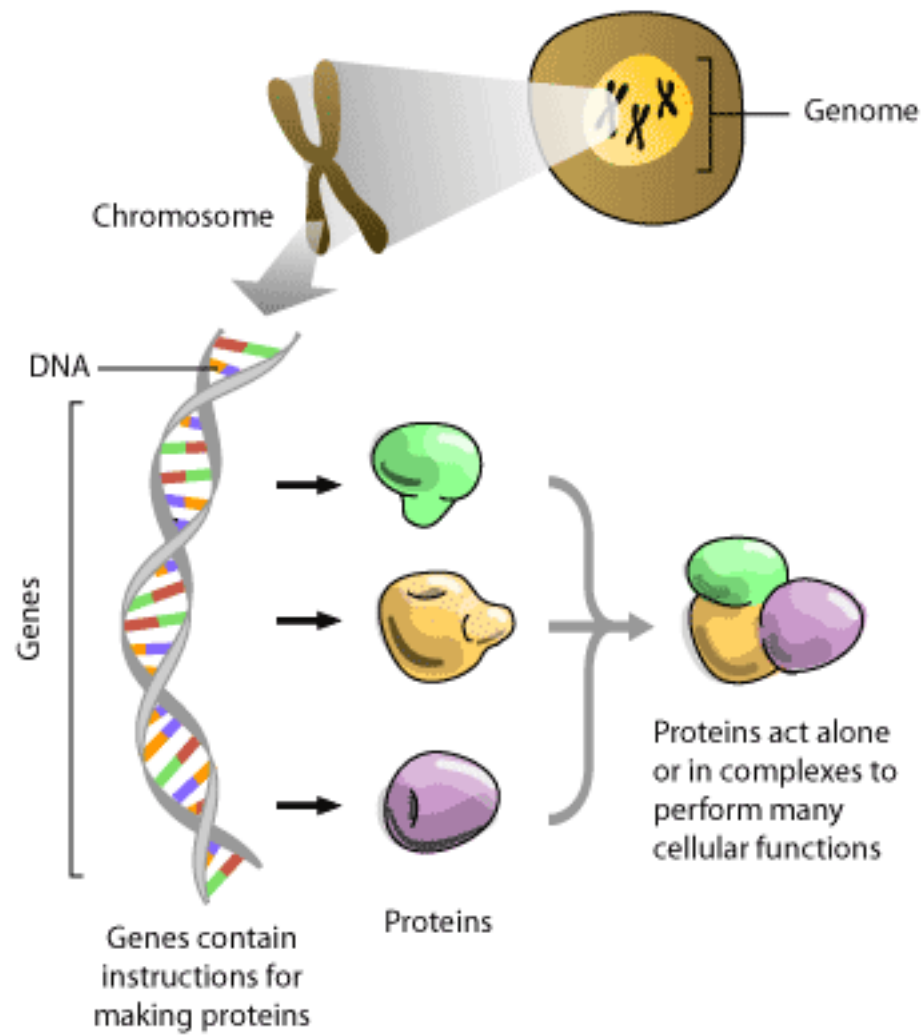
Replication

Transcription

Translation



Genotype

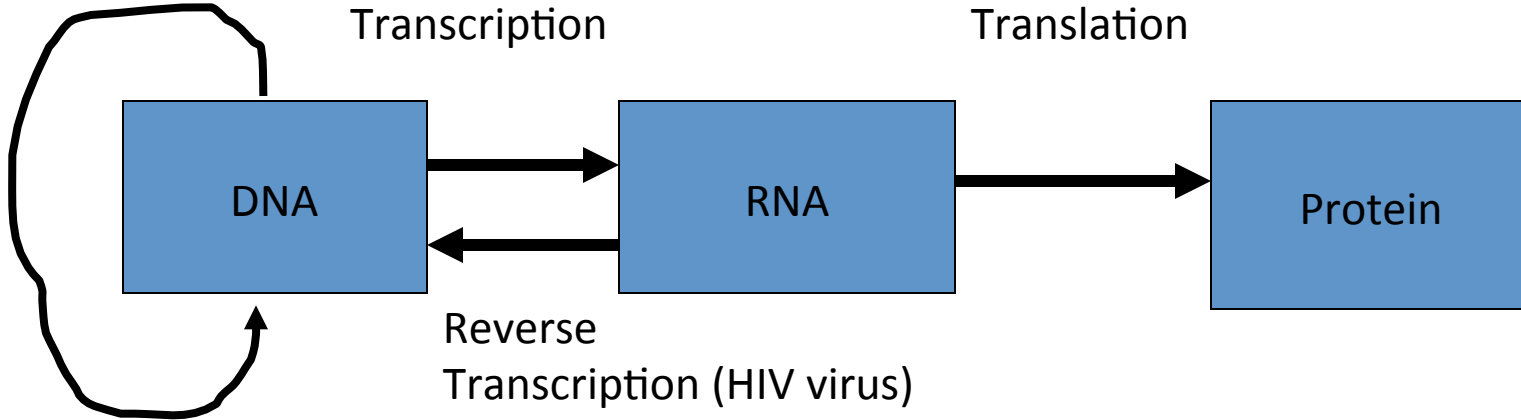


Central Dogma of Molecular Biology

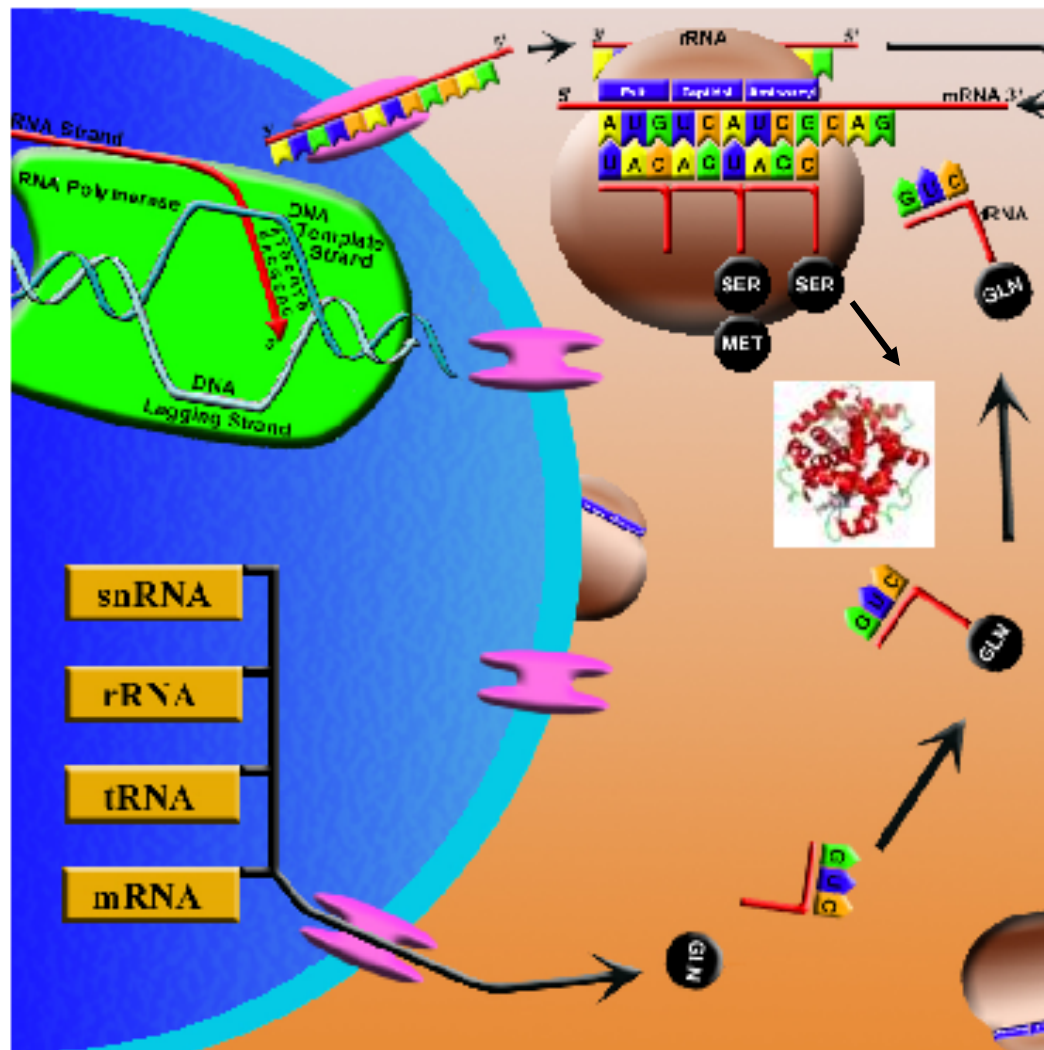
Replication

Transcription

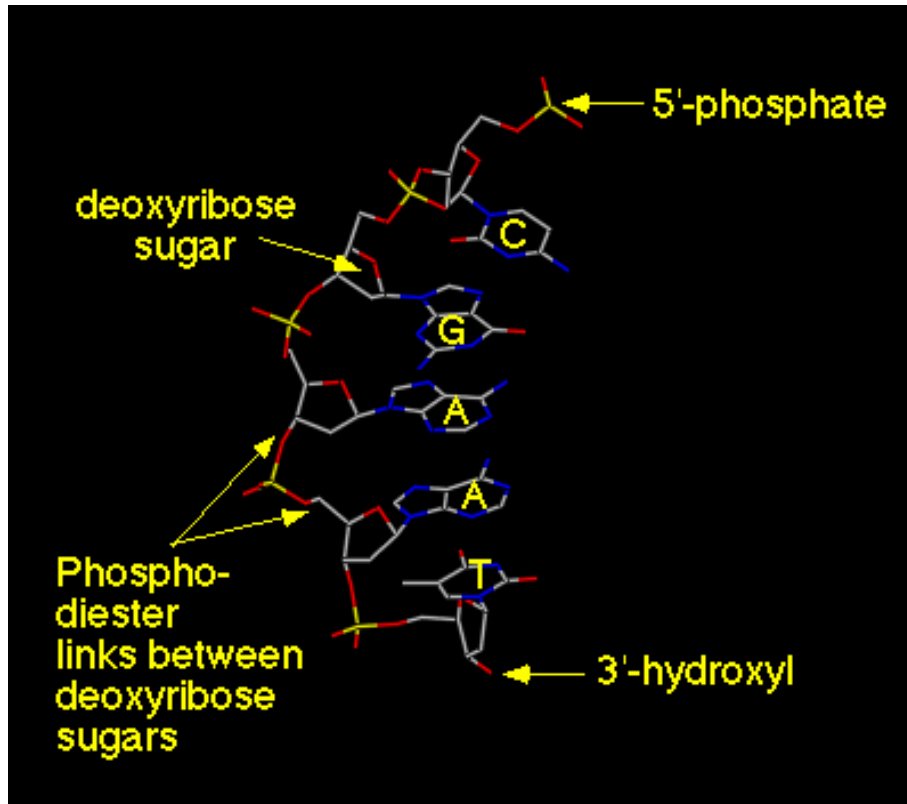
Translation



Information flow



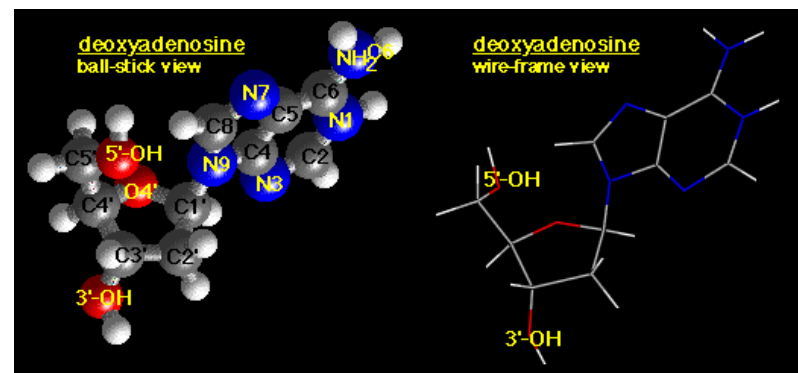
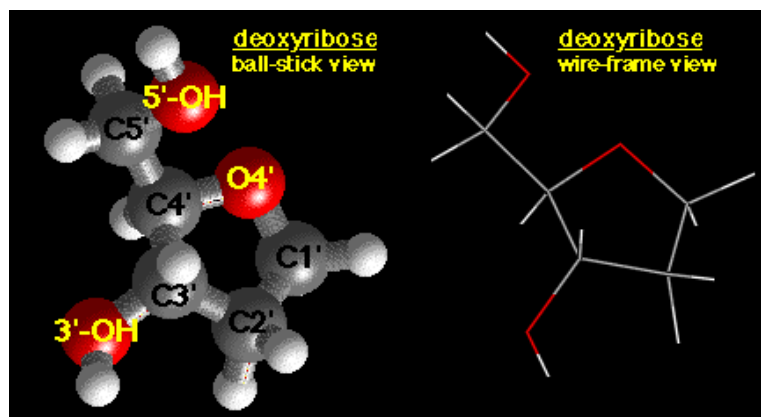
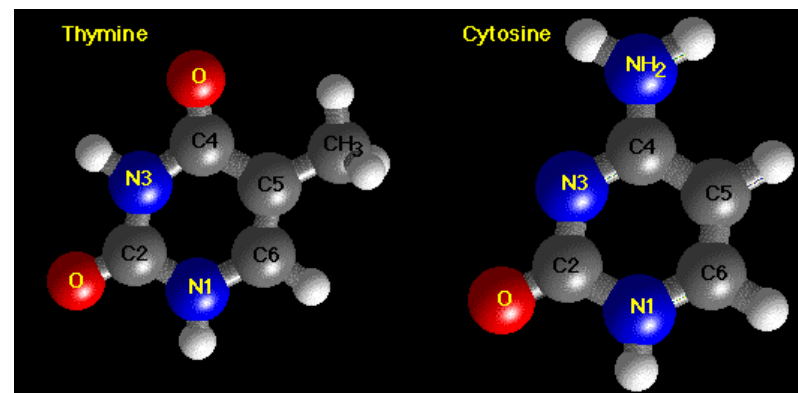
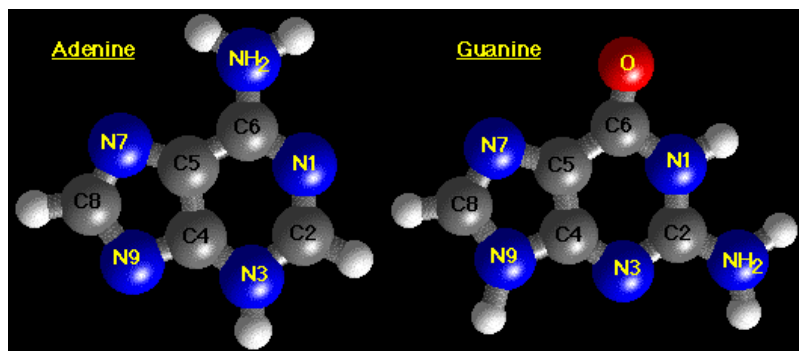
DNA (Deoxyribose Nucleotide Acids)

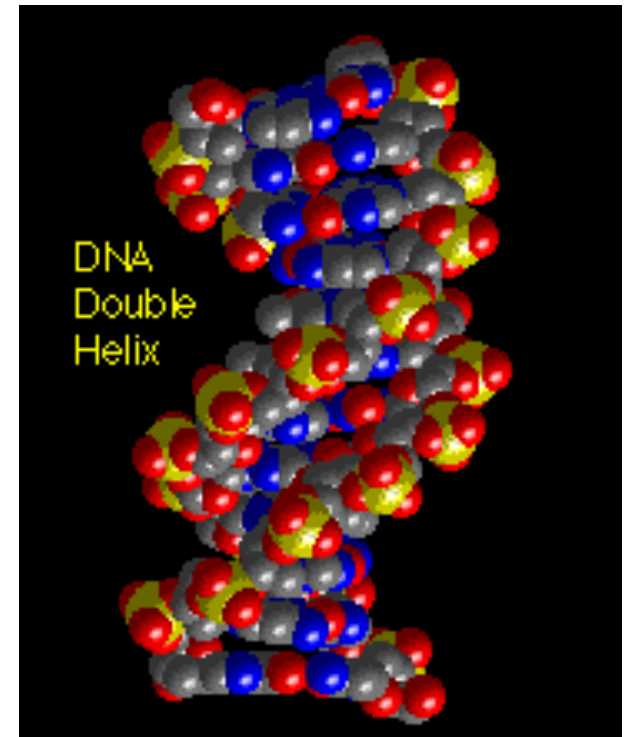
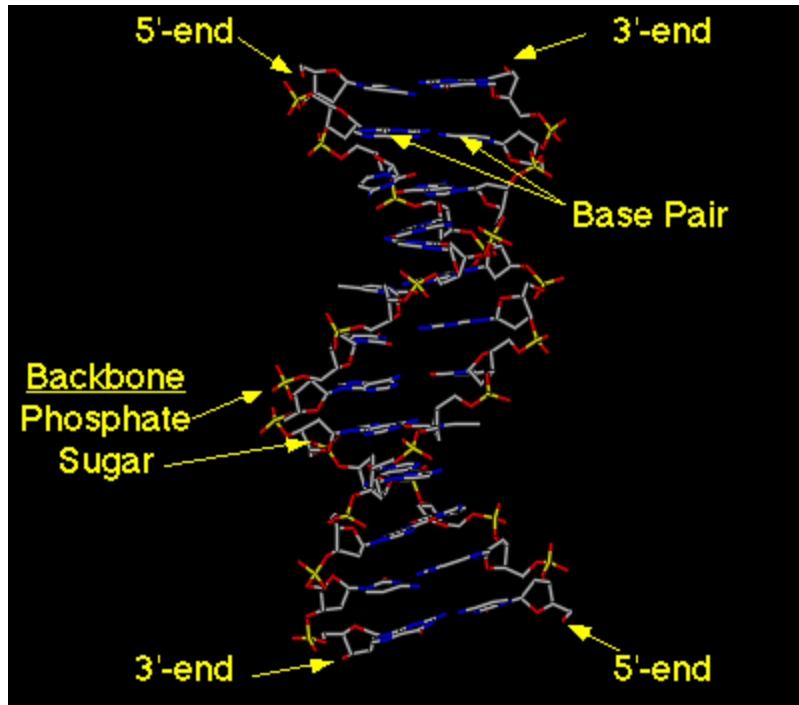


CGAATGGGAAA.....

DNA is a polymer. The monomer units of DNA are nucleotides, and the polymer is known as a "polynucleotide." Each nucleotide consists of a 5-carbon sugar (deoxyribose), a nitrogen containing base attached to the sugar, and a phosphate group.

A is for adenine
G is for guanine
C is for cytosine
T is for thymine



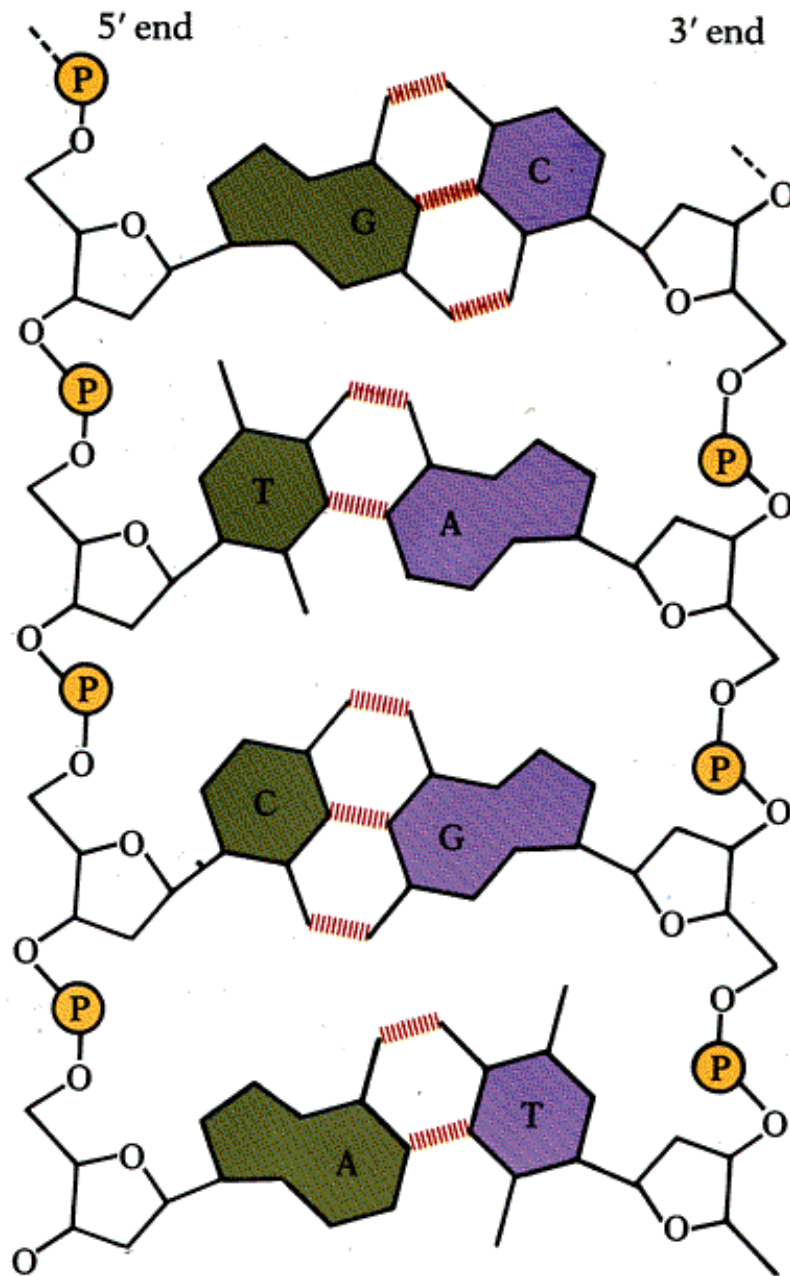


Base Pairs:

A-T (2 H-bonds)

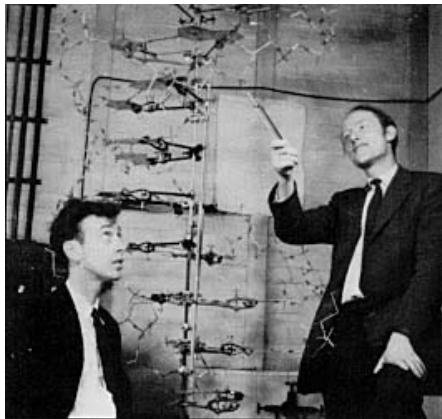
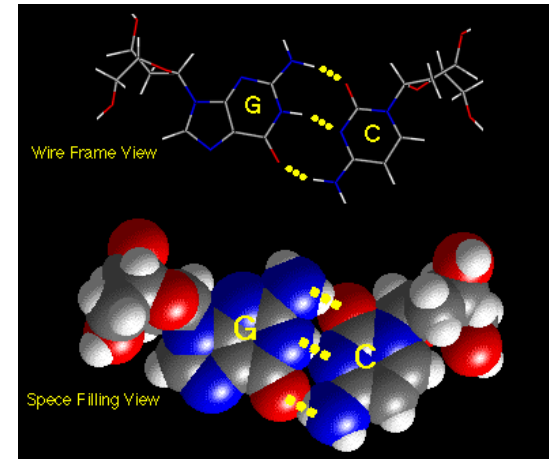
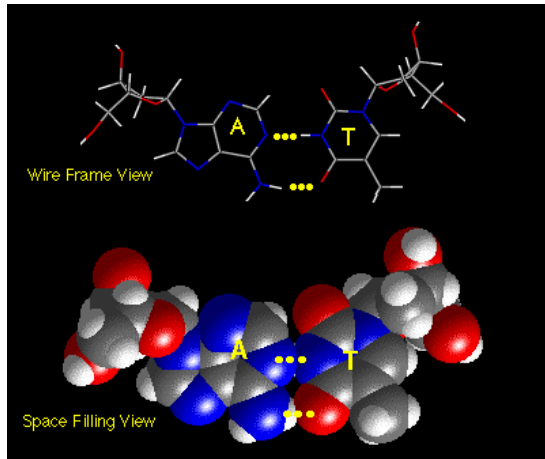
C-G (3 H-bonds)

Hydrogen bonds: non-covalent bonds mediated by hydrogen atoms



Uncoiled DNA Molecule

Source: Dr. Gary Stormo, 2002



James Watson & Francis Crick



Maurice
Wilkins



Rosalind
Franklin



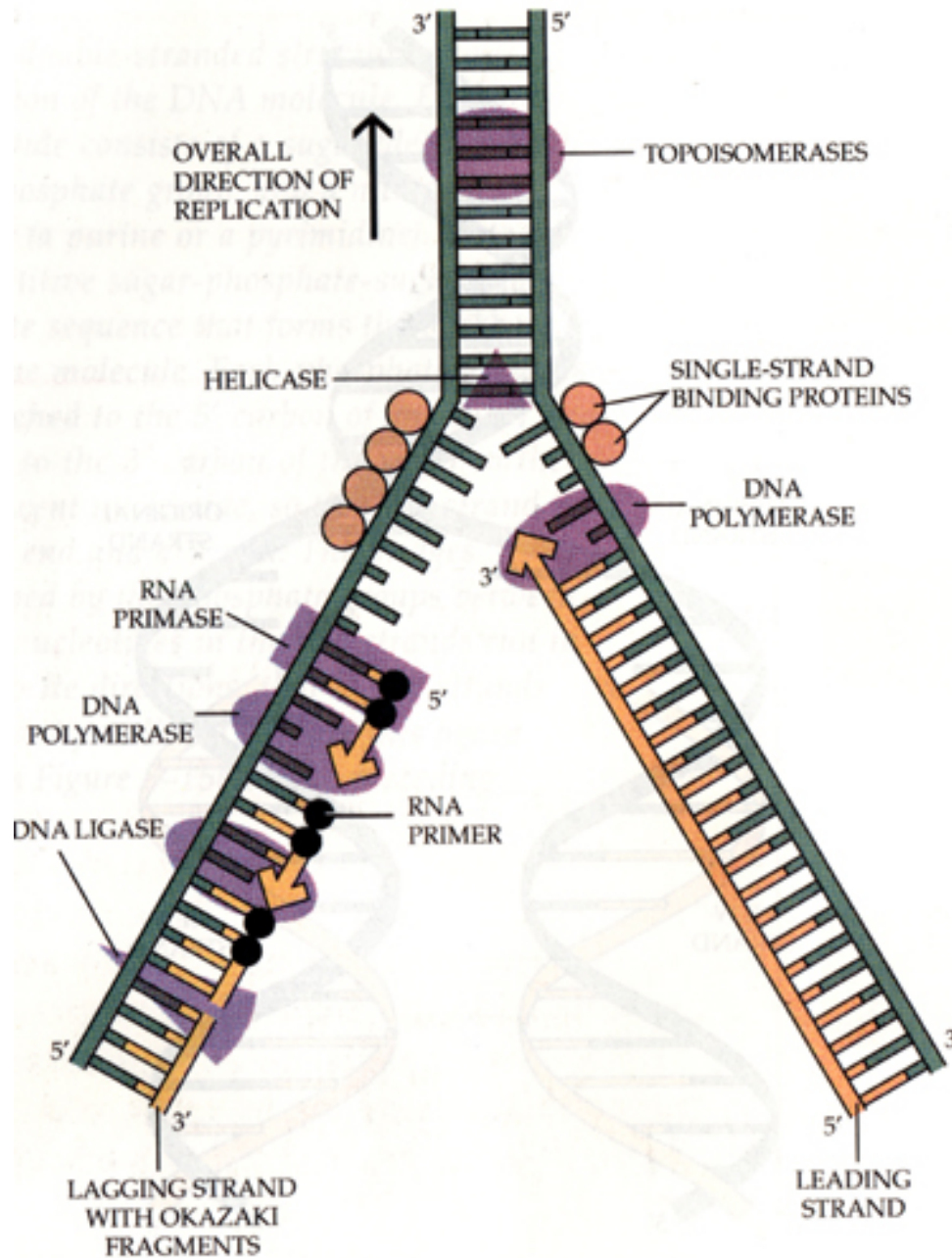
Linus
Pauling



Erwin
Chargaff

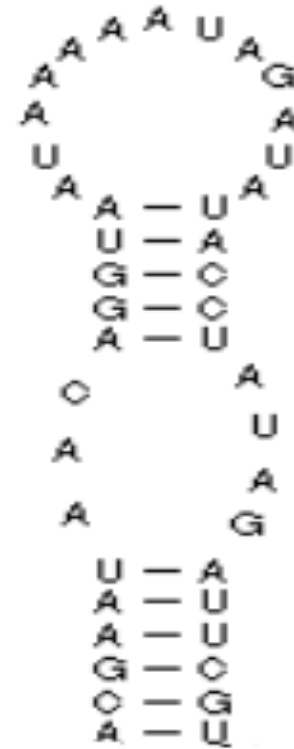
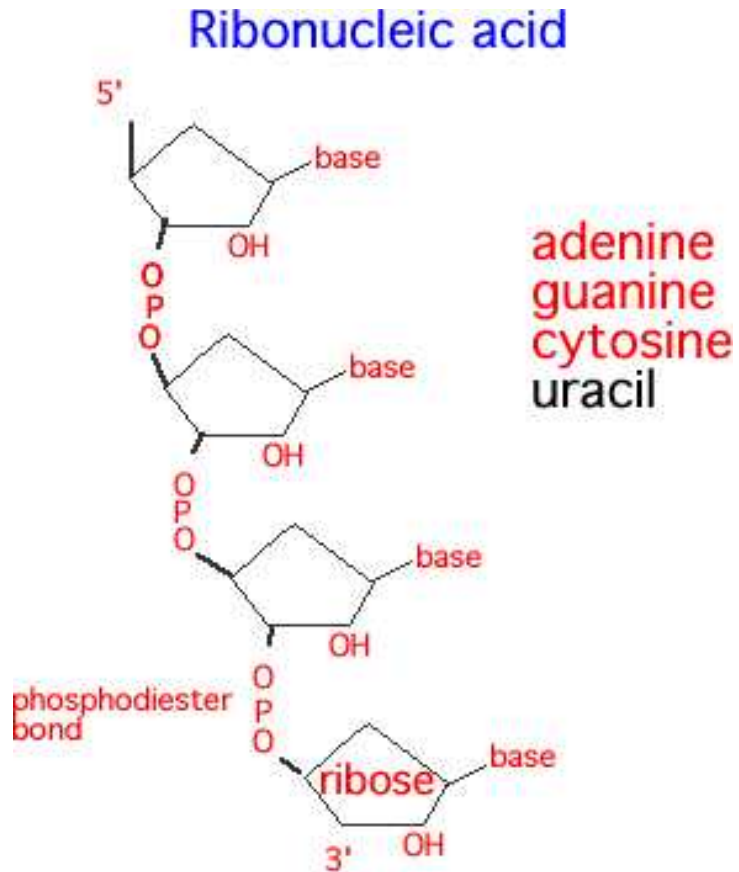
Fundamental Problems: How genetic information pass from one cell to another and from one generation to next generation

DNA Replication



DNA
Polymerase

RNA (Ribose Nucleotide Acids)

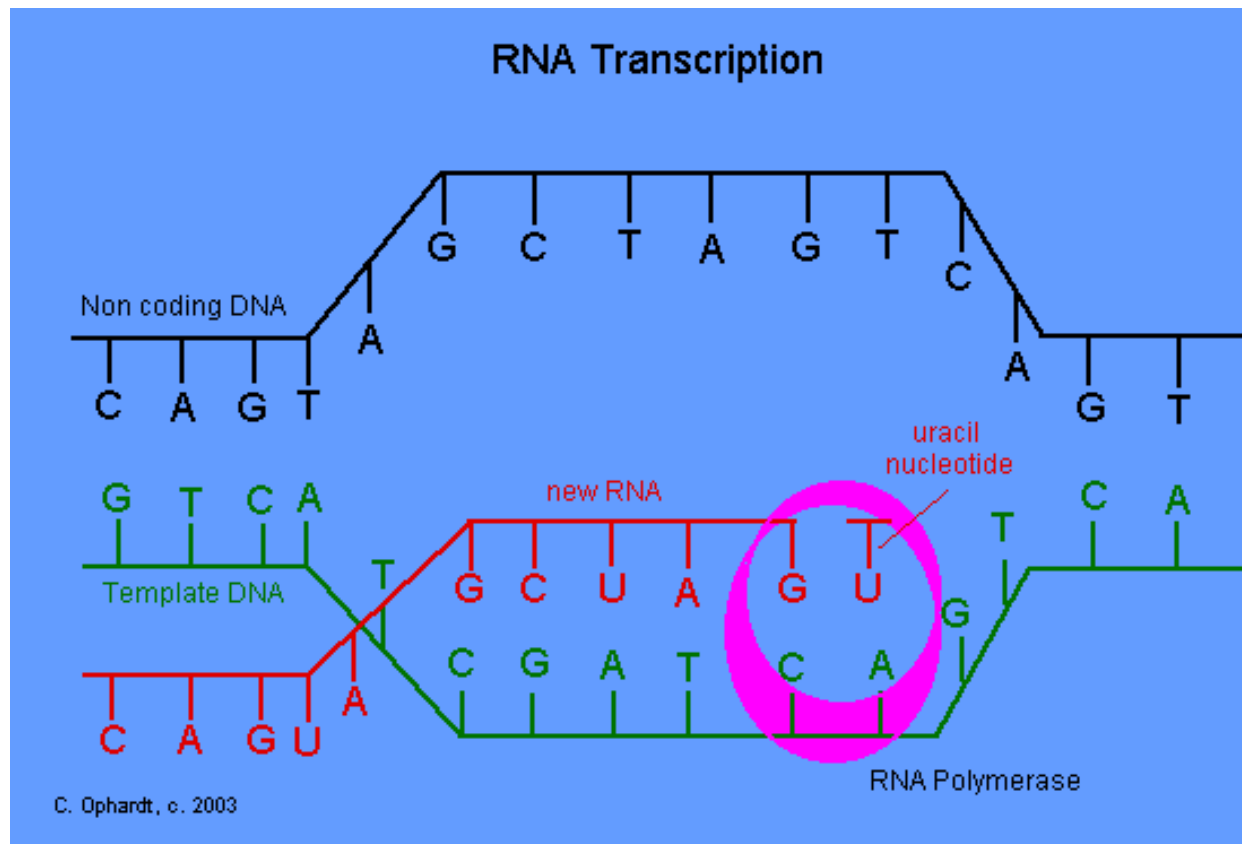


ACGAAUACAGGUAAUAAAAUAGAUAUACCUAUAGAUUCGU

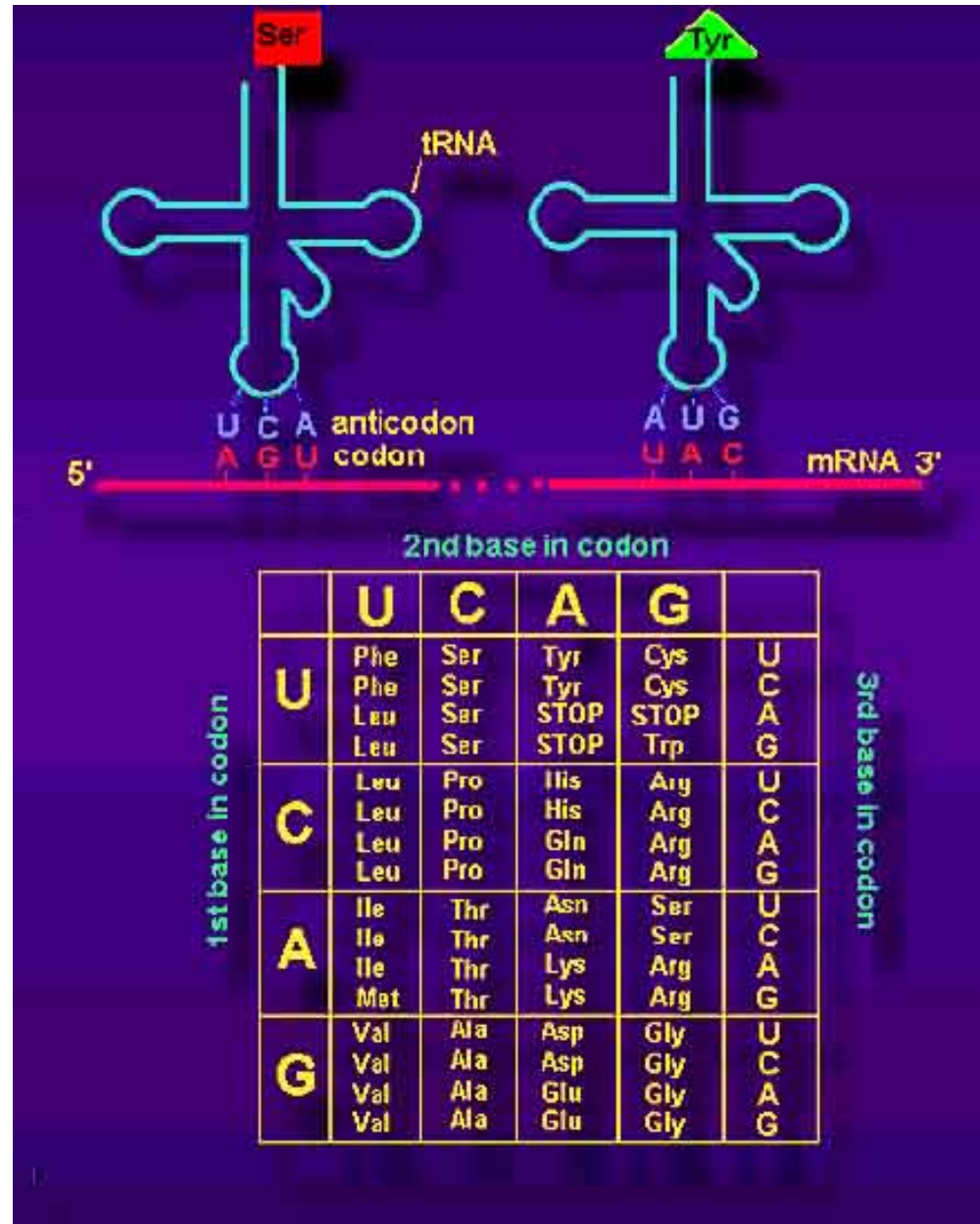
Different Kinds of RNA

- mRNA: messenger RNA
carry genetic information out of nucleus for protein synthesis
(transcription process: RNA polymerase)
- rRNA: ribosomal RNA
constitute 50% of ribosome, which is a molecular assembly for protein synthesis
- tRNA: transfer RNA
decode information (map 3 nucleotides to amino acid);
transfer amino acid
- snRNA: small RNA molecules found in nucleus
involve RNA splicing

Transcription of Gene into RNA



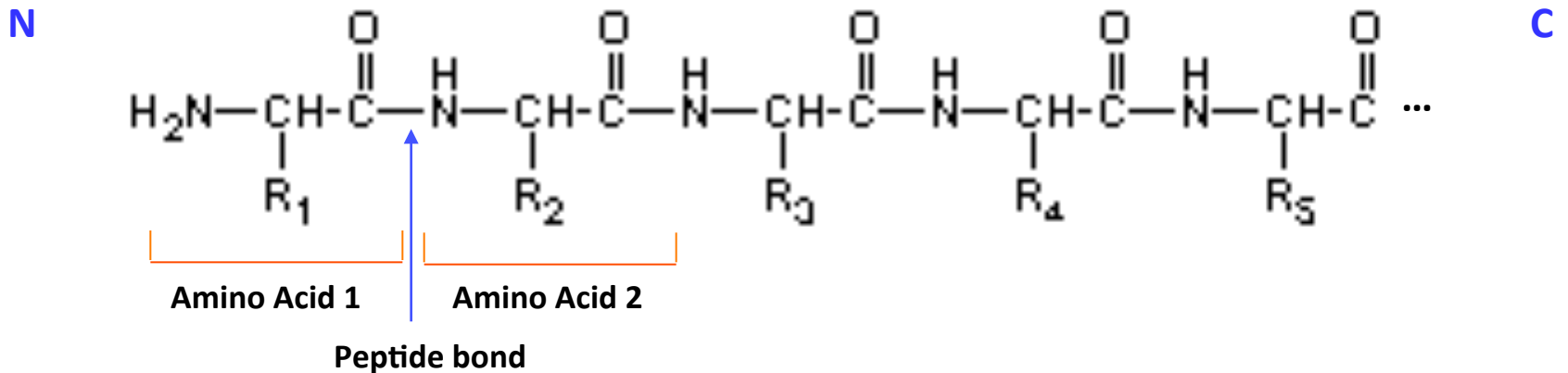
Genetic Code and Translation



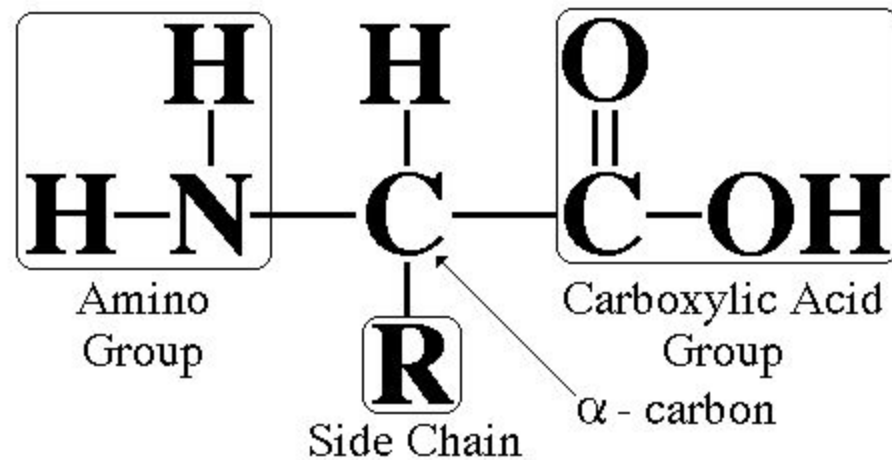
Three Nucleotides
is called a codon.

Protein Sequence

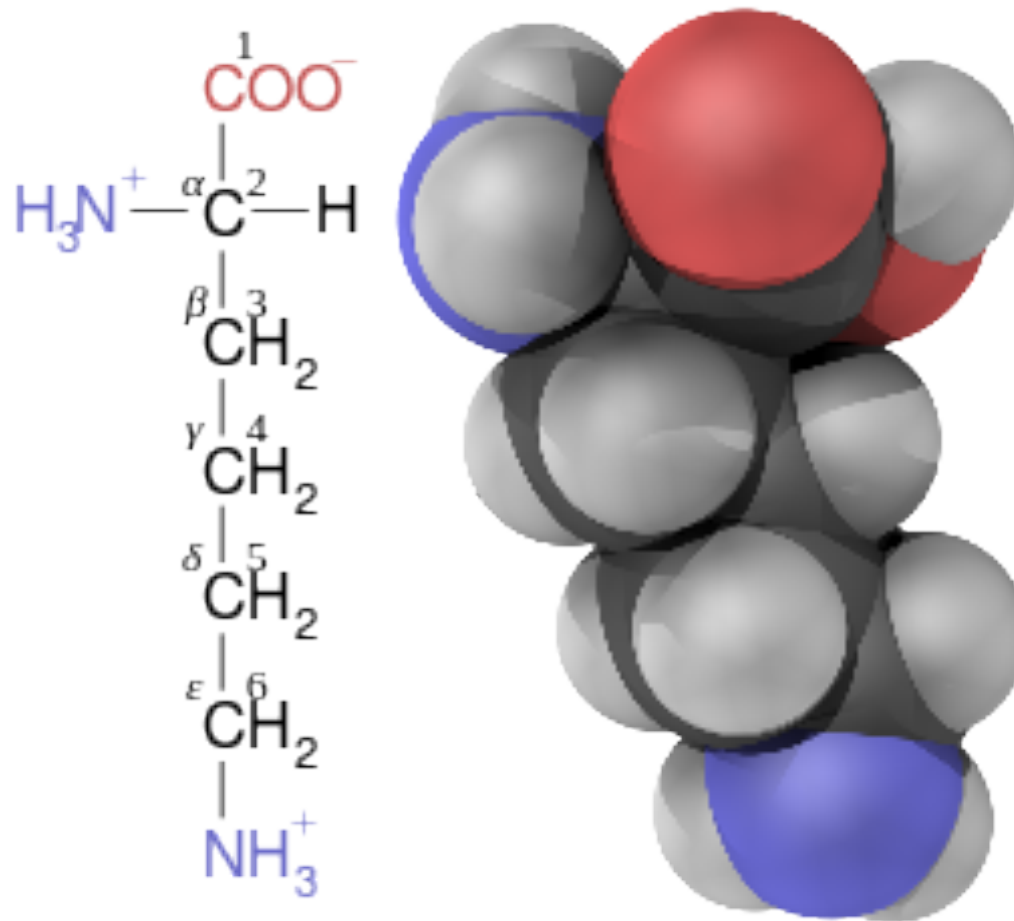
A directional sequence of amino acids/residues



Amino Acid Structure



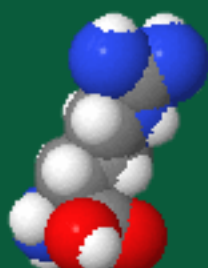
Lysine



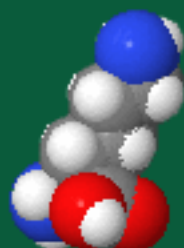
Amino Acids

Amino acid	Abbrev.	Side chain	Hydrophobic	Polar	Charged	Small	Tiny	Aromatic or Aliphatic	van der Waals volume	Codon	Occurrence in proteins (%)
Alanine	Ala, A	-CH ₃	X	-	-	X	X	-	67	GCU, GCC, GCA, GCG	7.8
Cysteine	Cys, C	-CH ₂ SH	X	-	-	X	-	-	86	UGU, UGC	1.9
Aspartate	Asp, D	-CH ₂ COOH	-	X	negative	X	-	-	91	GAU, GAC	5.3
Glutamate	Glu, E	-CH ₂ CH ₂ COOH	-	X	negative	-	-	-	109	GAA, GAG	6.3
Phenylalanine	Phe, F	-CH ₂ C ₆ H ₅	X	-	-	-	-	Aromatic	135	UUU, UUC	3.9
Glycine	Gly, G	-H	X	-	-	X	X	-	48	GGU, GGC, GGA, GGG	7.2
Histidine	His, H	-CH ₂ -C ₃ H ₃ N ₂	-	X	positive	-	-	Aromatic	118	CAU, CAC	2.3
Isoleucine	Ile, I	-CH(CH ₃)CH ₂ CH ₃	X	-	-	-	-	Aliphatic	124	AUU, AUC, AUA	5.3
Lysine	Lys, K	-(CH ₂) ₄ NH ₂	-	X	positive	-	-	-	135	AAA, AAG	5.9
Leucine	Leu, L	-CH ₂ CH(CH ₃) ₂	X	-	-	-	-	Aliphatic	124	UUA, UUG, CUU, CUC, CUA, CUG	9.1
Methionine	Met, M	-CH ₂ CH ₂ SCH ₃	X	-	-	-	-	-	124	AUG	2.3
Asparagine	Asn, N	-CH ₂ CONH ₂	-	X	-	X	-	-	96	AAU, AAC	4.3
Proline	Pro, P	-CH ₂ CH ₂ CH ₂ -	X	-	-	X	-	-	90	CCU, CCC, CCA, CCG	5.2
Glutamine	Gln, Q	-CH ₂ CH ₂ CONH ₂	-	X	-	-	-	-	114	CAA, CAG	4.2
Arginine	Arg, R	-(CH ₂) ₃ NH-C(NH) NH ₂	-	X	positive	-	-	-	148	CGU, CGC, CGA, CGG, AGA, AGG	5.1
Serine	Ser, S	-CH ₂ OH	-	X	-	X	X	-	73	UCU, UCC, UCA, UCG, AGU, AGC	6.8
Threonine	Thr, T	-CH(OH)CH ₃	X	X	-	X	-	-	93	ACU, ACC, ACA, ACG	5.9
Valine	Val, V	-CH(CH ₃) ₂	X	-	-	X	-	Aliphatic	105	GUU, GUC, GUA, GUG	6.6
Tryptophan	Trp, W	-CH ₂ C ₈ H ₆ N	X	-	-	-	-	Aromatic	163	UGG	1.4
Tyrosine	Tyr, Y	-CH ₂ -C ₆ H ₄ OH	X	X	-	-	-	Aromatic	141	UAU, UAC	3.2

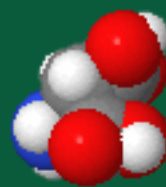
Hydrophilic



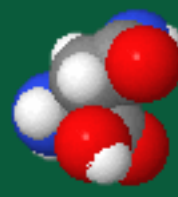
Arg



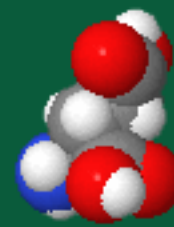
Lys



Asp



Asn



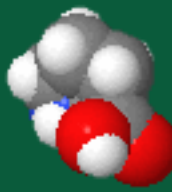
Glu



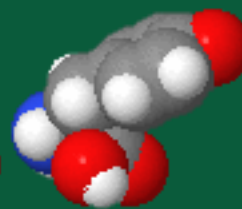
Gln



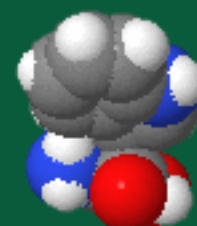
His



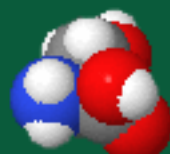
Pro



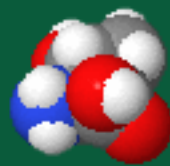
Tyr



Trp



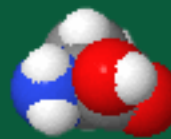
Ser



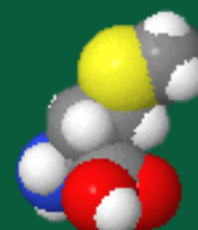
Thr



Gly



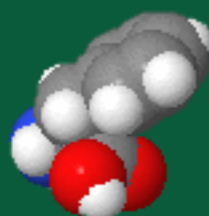
Ala



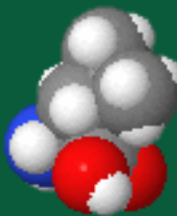
Met



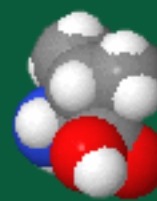
Cys



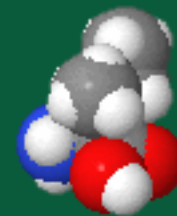
Phe



Leu



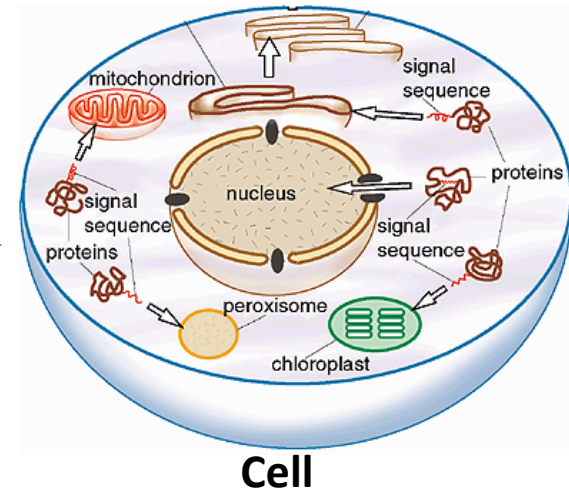
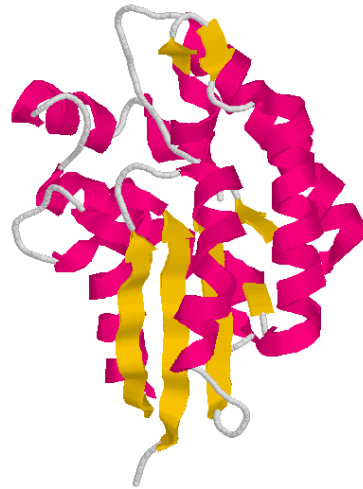
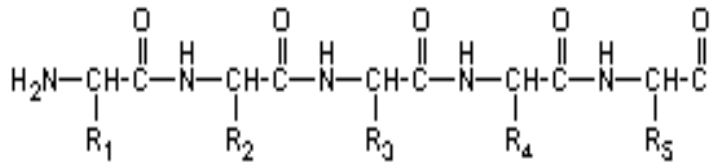
Val



Ile

Central Dogma of Proteomics

AGCWY.....



Sequence

Structure

Function