

Course organization

– Introduction (Week 1-2)

- Course introduction
- A brief introduction to molecular biology
- A brief introduction to sequence comparison

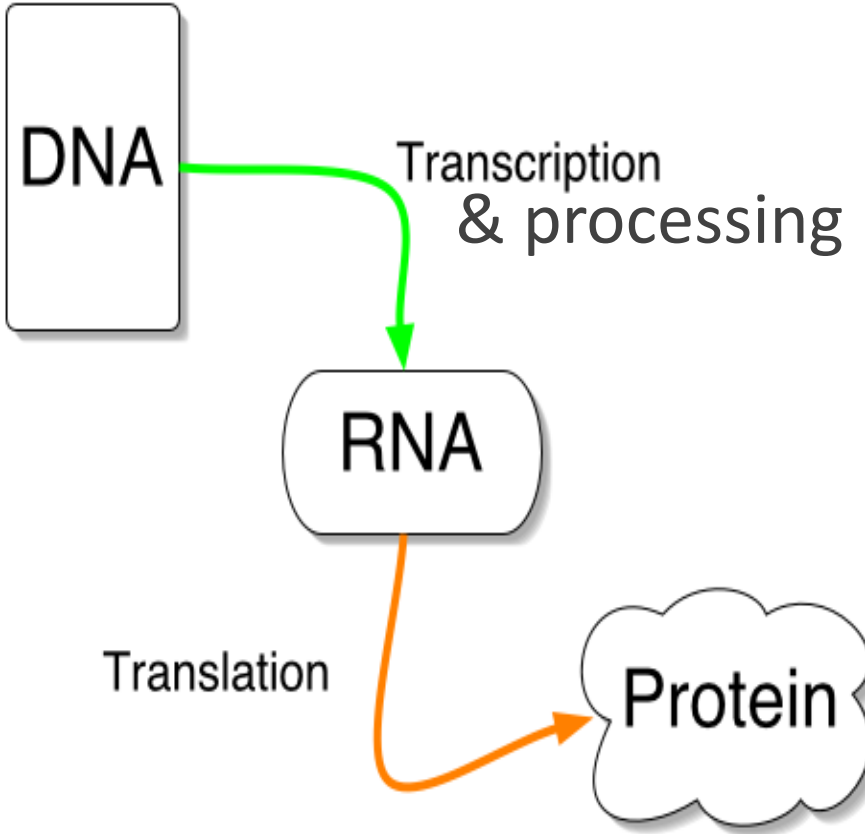
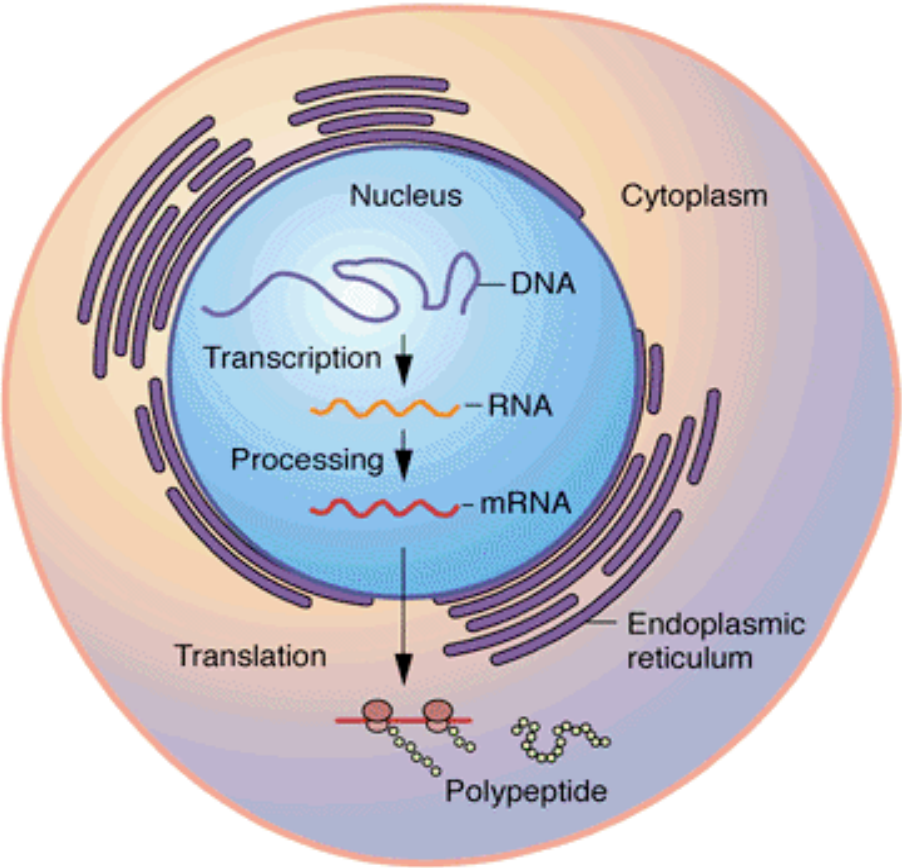
– Part I: Algorithms for Sequence Analysis (Week 3 - 11)

- Chapter 1-3, Models and theories
 - » Probability theory and Statistics (Week 3)
 - » Algorithm complexity analysis (Week 4)
 - » Classic algorithms (Week 5)
 - » Lab: Linux and Perl
- Chapter 4, Sequence alignment (week 6)
- Chapter 5, Hidden Markov Models (week 8)
- Chapter 6. Multiple sequence alignment (week 10)
- Chapter 7. Motif finding (week 11)
- Chapter 8. Sequence binning (week 11)

– Part II: Algorithms for Network Biology (Week 12 - 16)

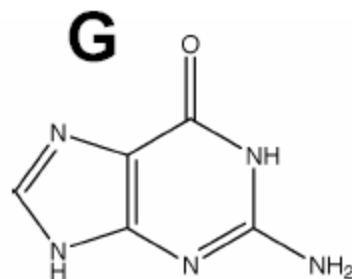
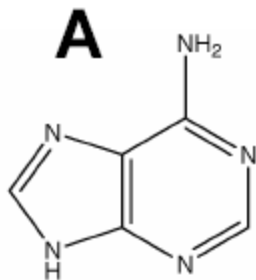
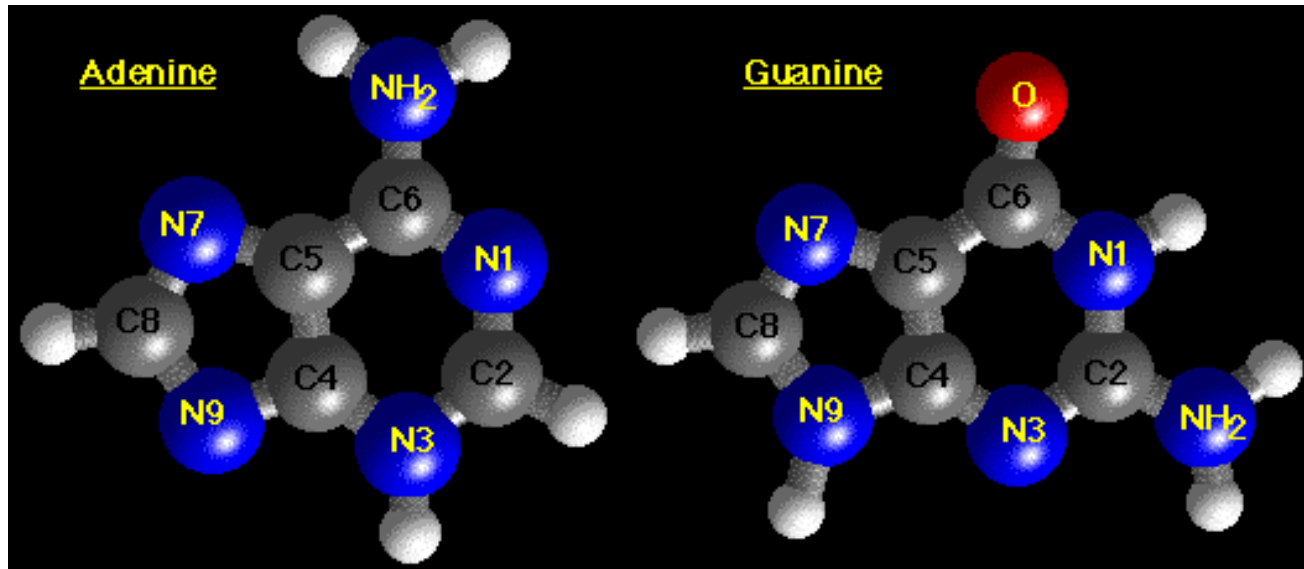
A brief introduction to Molecular Biology

Central dogma

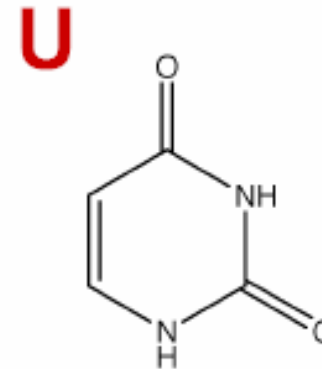
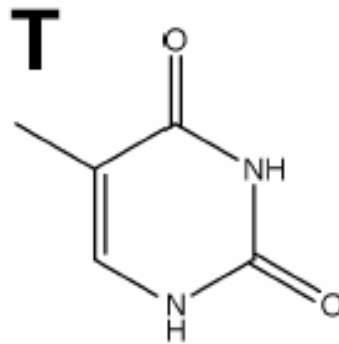
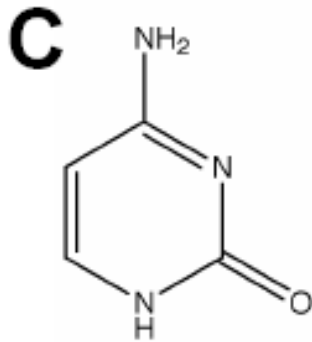
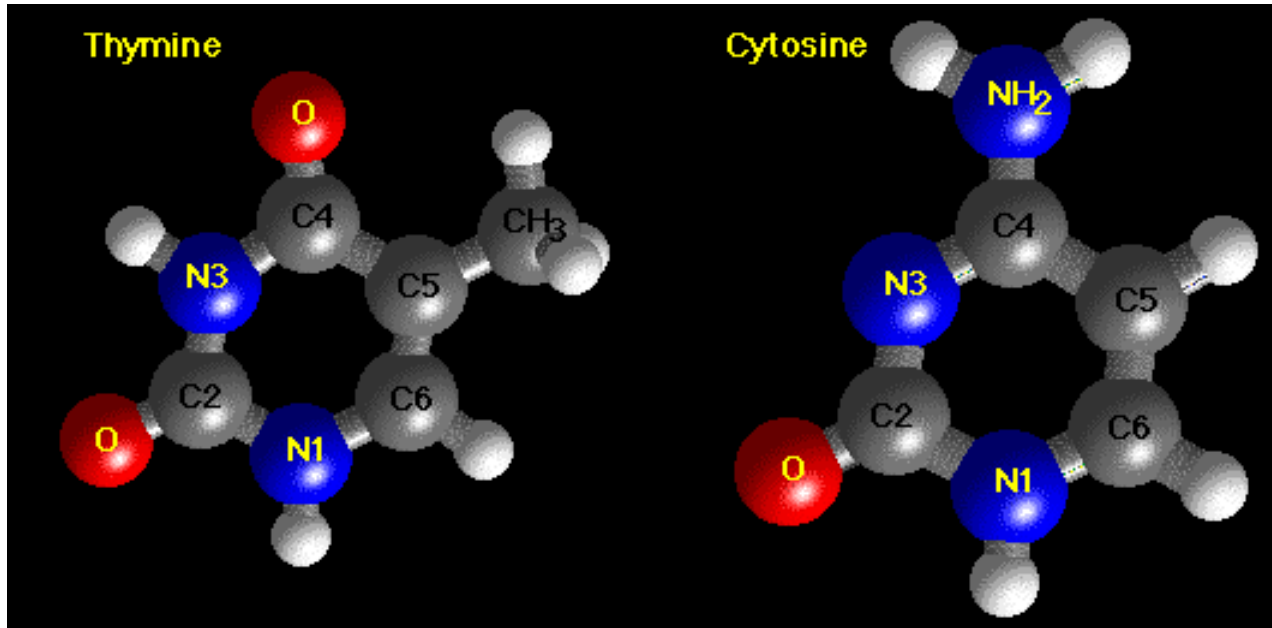


Copyright © 1997, by John Wiley & Sons, Inc. All rights reserved.

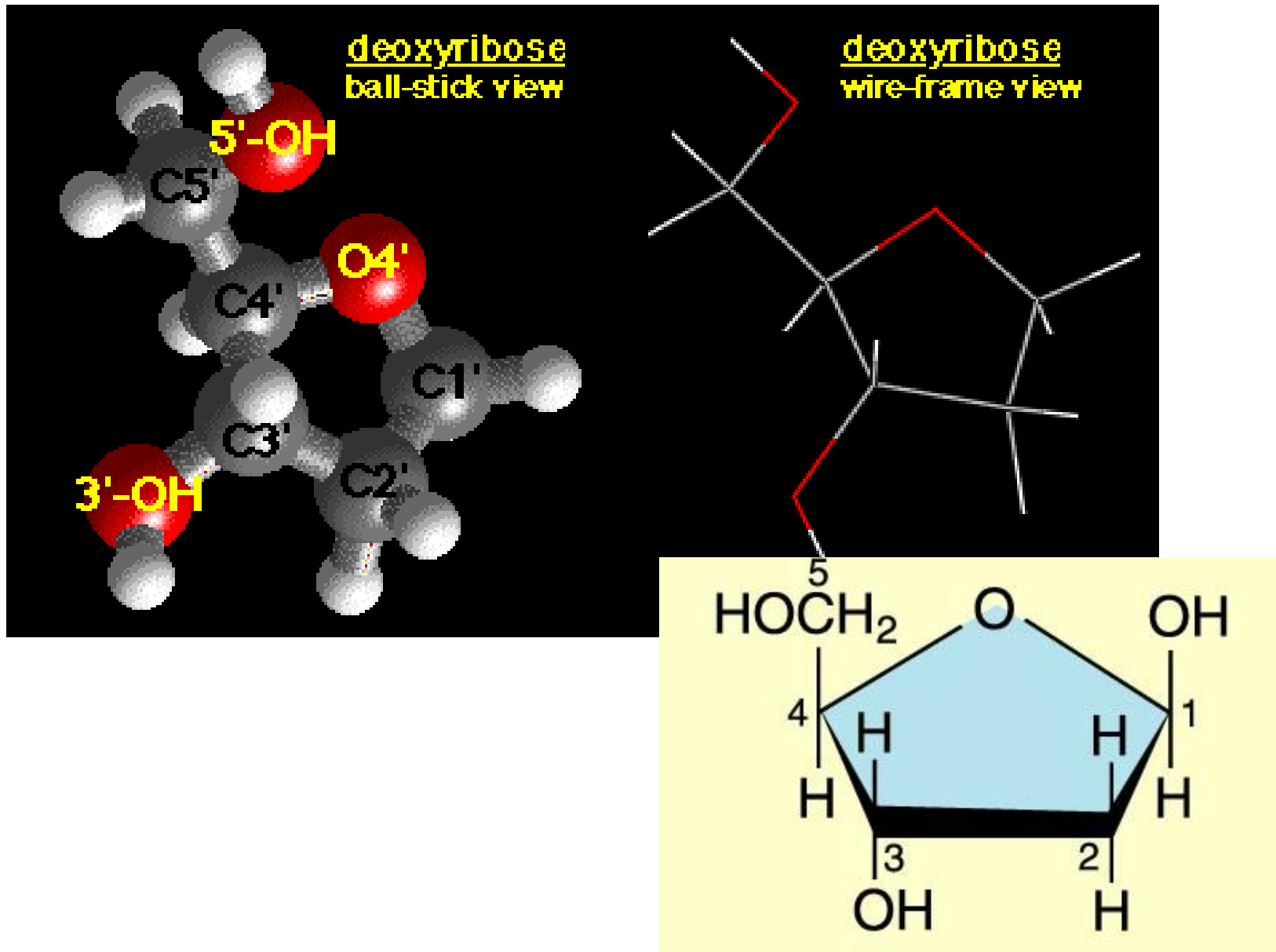
Structure of A and G



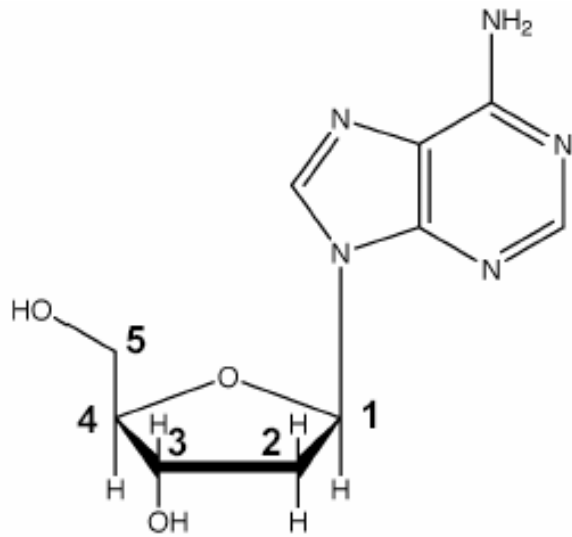
Structure of C and T



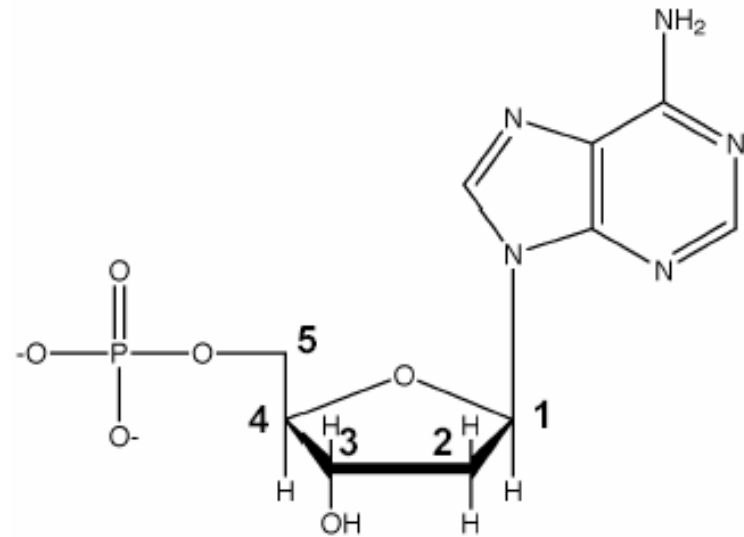
Structure of deoxyribose



Nucleoside and Nucleotide

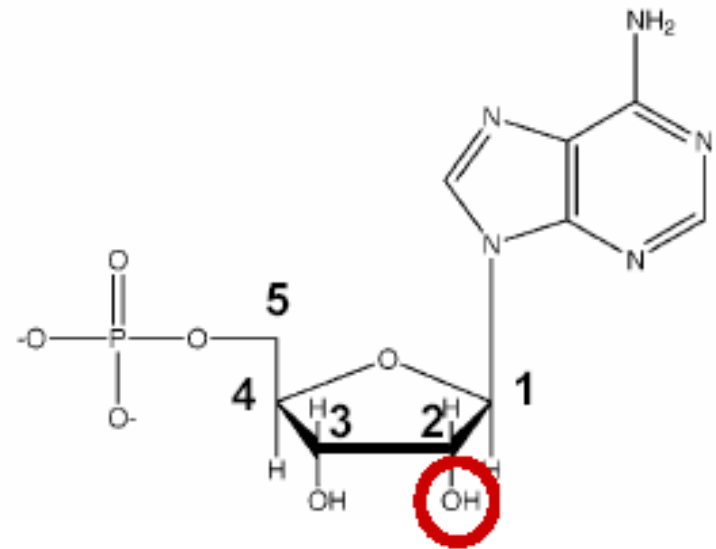
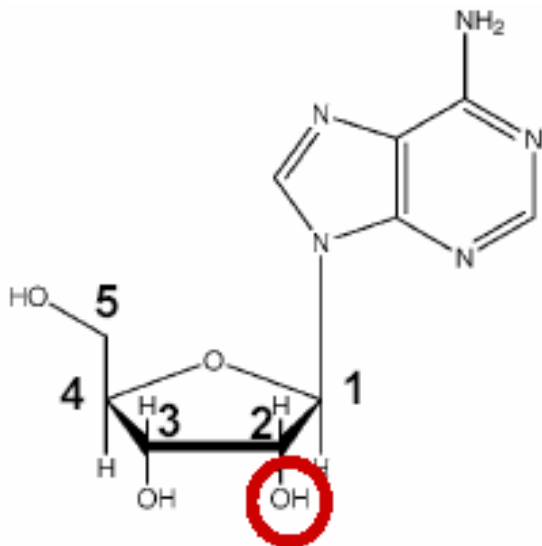


Nucleoside

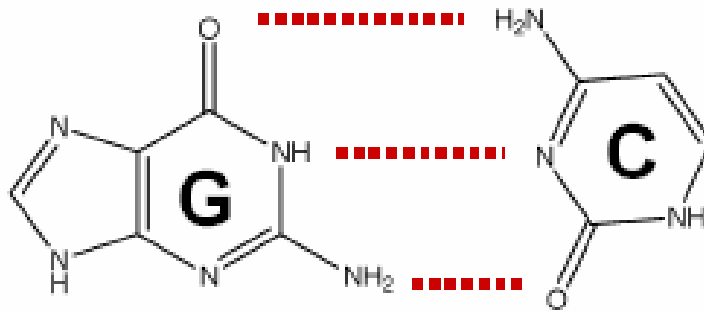
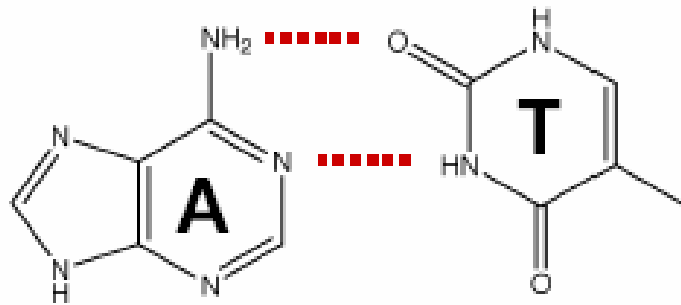


Nucleotide

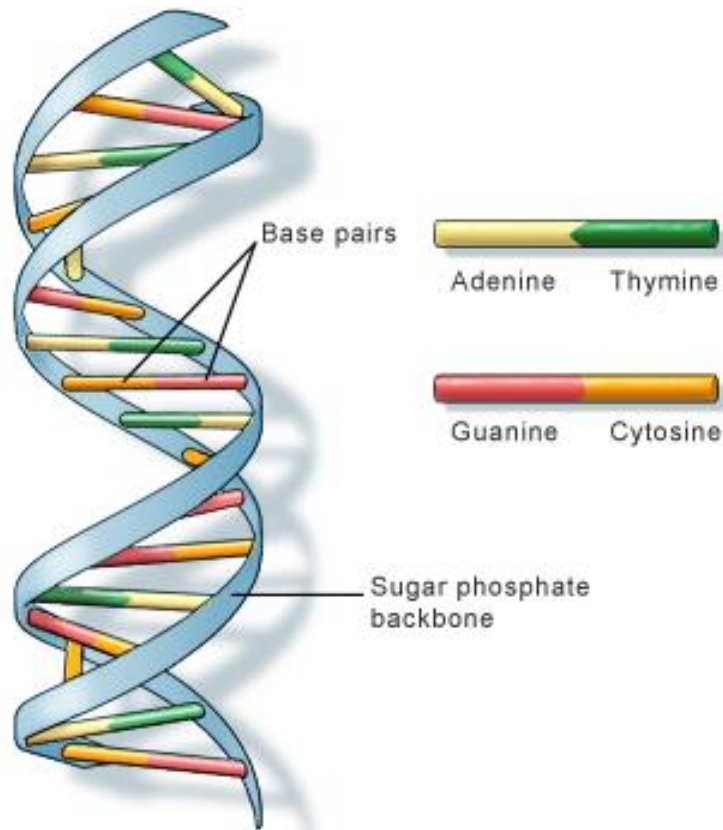
RNA Structure



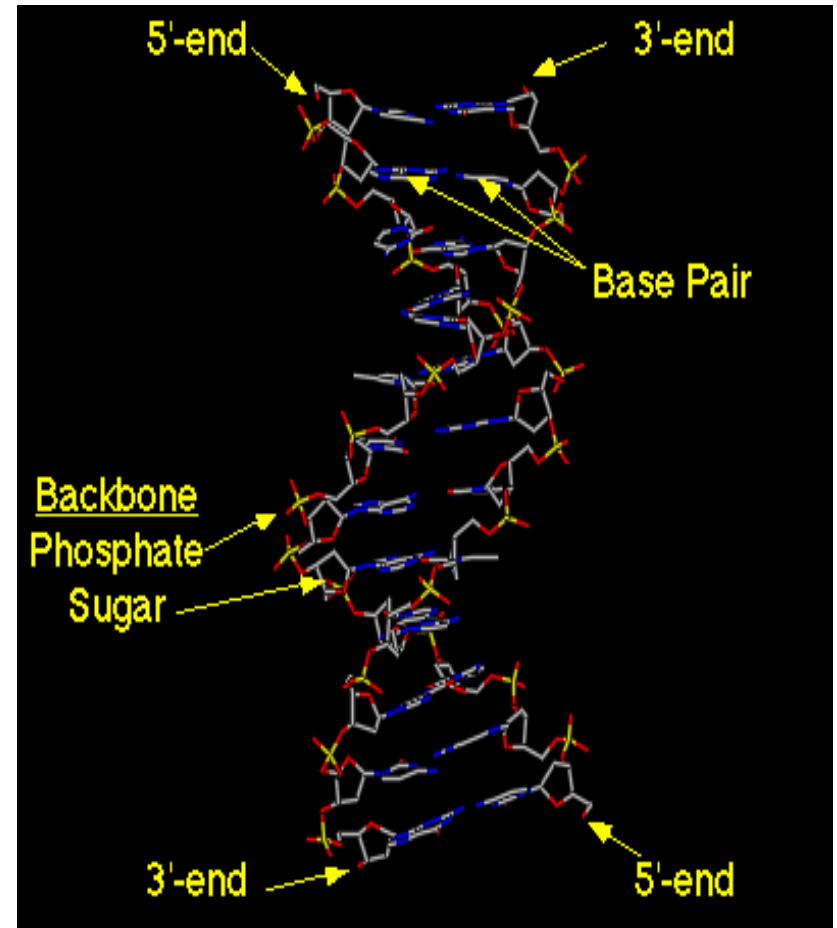
Base pair



Double helix structure



U.S. National Library of Medicine



Quiz

1. A DNA strain with 10 nucleotides can form 4¹⁰ different sequences.

2. In a DNA molecular, the percentage of base A is 38%, then the percentage of base C and G in total is : (C)

A. 76%

B. 62%

C. 24%

D. 12%

3. In a DNA strain, A: C: T: G=1: 2: 3: 4, then in its complimentary strain , A: C: T: G is (B)

A. 1: 2: 3: 4

B. 3: 4: 1: 2

C. 4: 3: 2: 1

D. 1: 3: 2: 4

4、 In a DNA strain, $(A+G)/(T+C)=0.4$. The corresponding percentages in its complimentary strain and the whole DNA molecular are (B)

A、 0.4 & 0.6 B、 2.5 & 1

C、 0.4 & 0.4 D、 0.6 & 1

Genome

- The hereditary info present in every cell

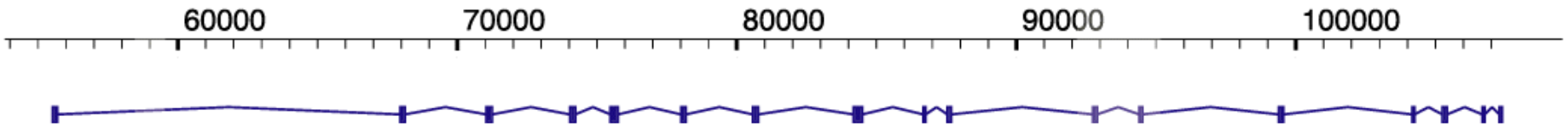
	Base Pairs	Genes
<i>Mycoplasma genitalium</i>	580,073	483
MimiVirus	1,200,000	1,260
<i>E. coli</i>	4,639,221	4,290
<i>Saccharomyces cerevisiae</i>	12,495,682	5,726
<i>Caenorhabditis elegans</i>	95.5×10^6	19,820
<i>Arabidopsis thaliana</i>	115,409,949	25,498
<i>Drosophila melanogaster</i>	122,653,977	13,472
Humans	3.3×10^9	~25,000

In a Mammalian Genome

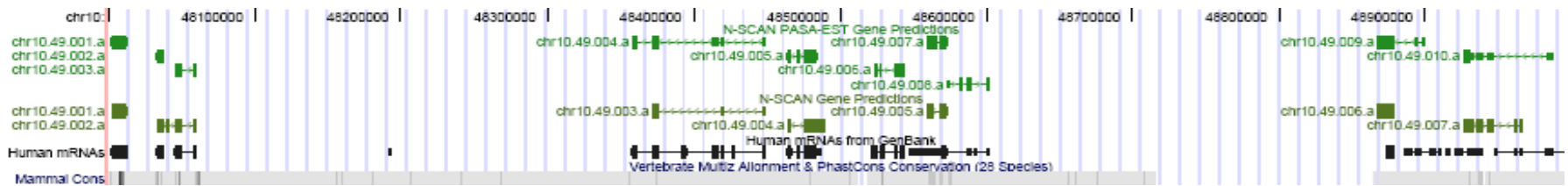
- Only about **1%** for protein coding
- Mammalian genomes are large
 - 8,000 km of 10pt type



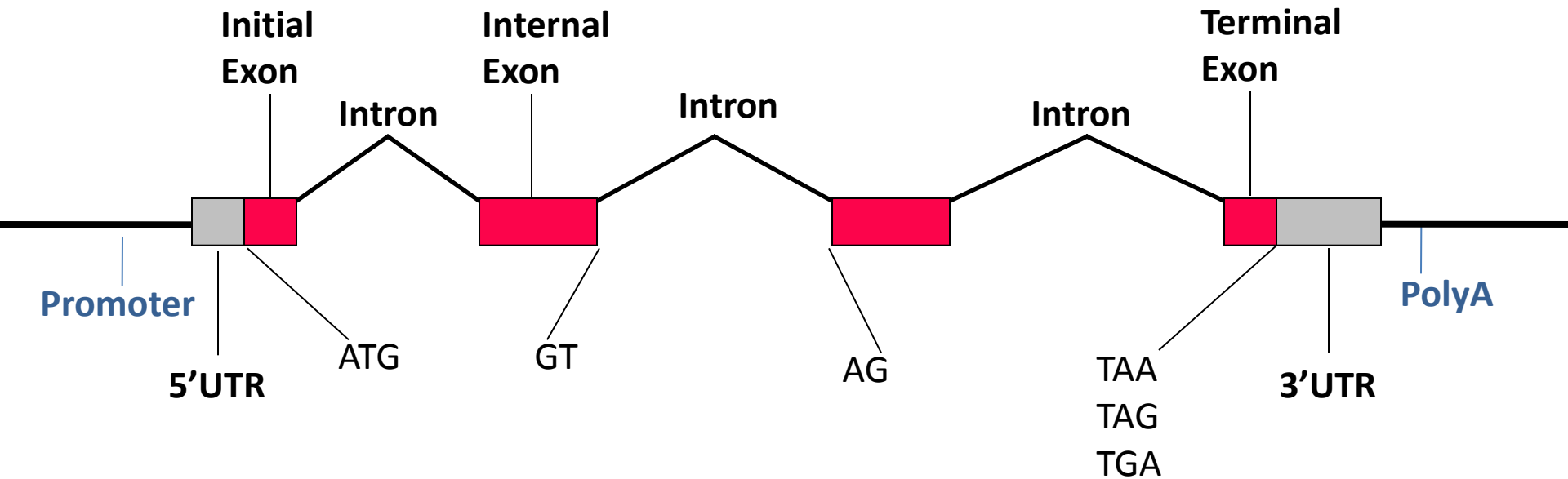
A Typical Human Gene Structure



Genes in a Genome



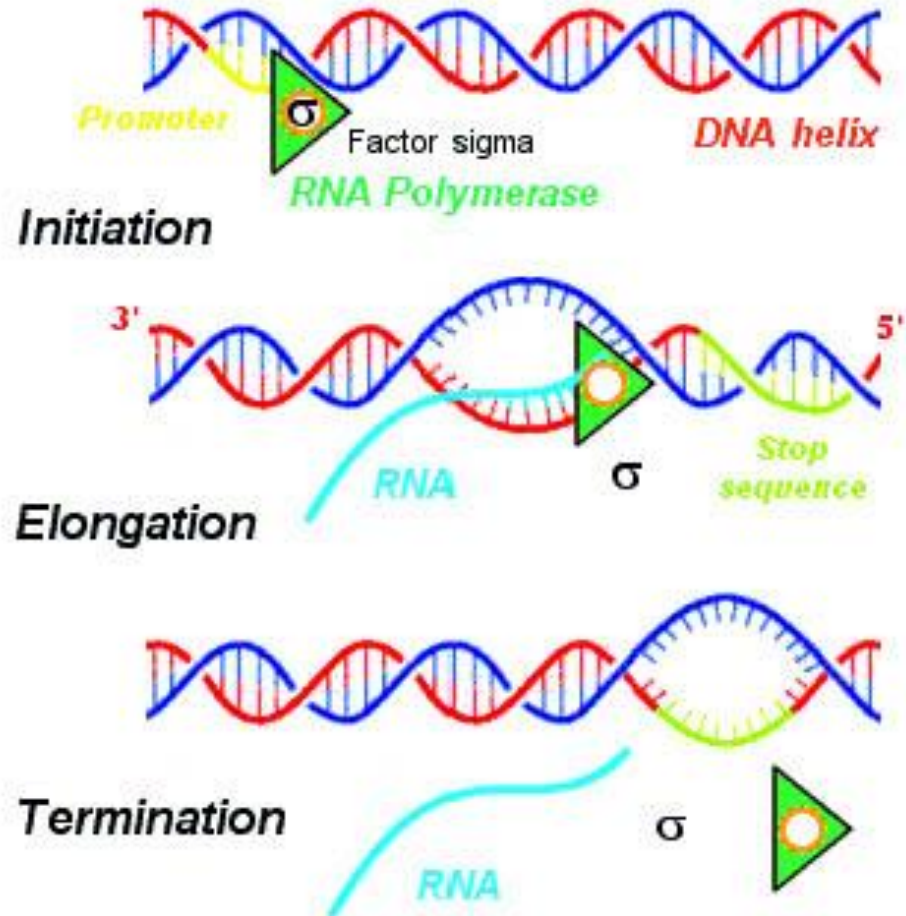
Gene Structure



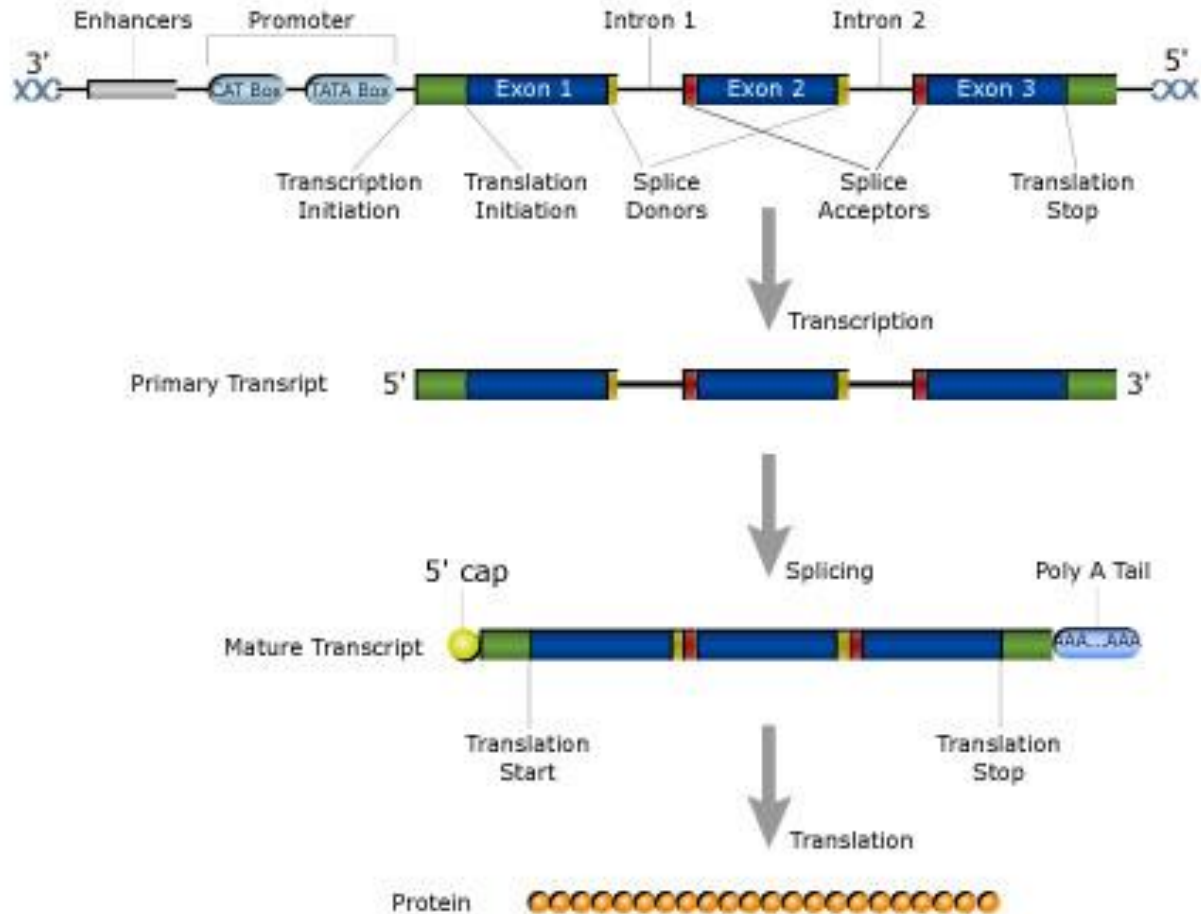
Gene Structure

- Transcribed 5' to 3'
- Promoter region and transcription factor binding sites precede 5'
- Transcribed region includes 5' and 3' untranslated regions
- In eukaryotes, most genes also include introns, spliced out before export from nucleus, hence before translation

Transcription

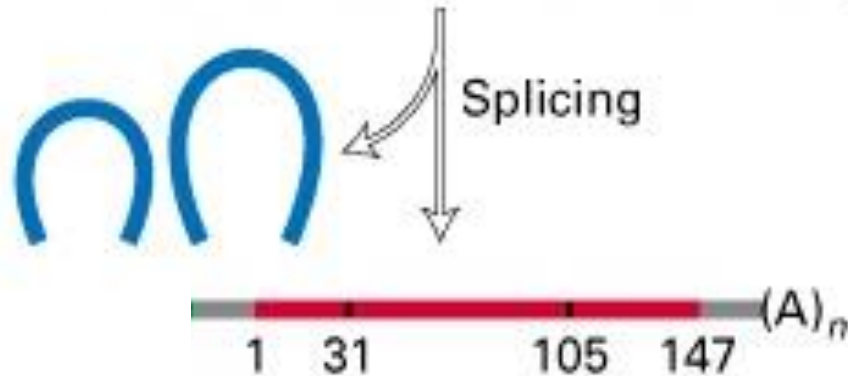
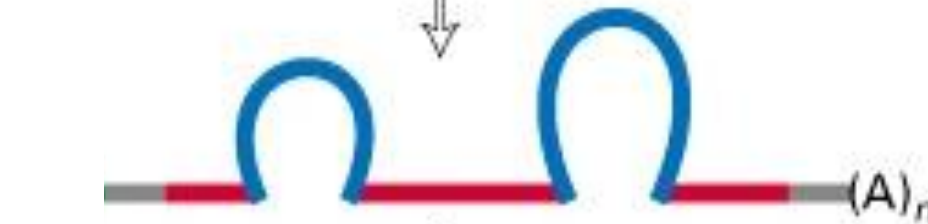


splice



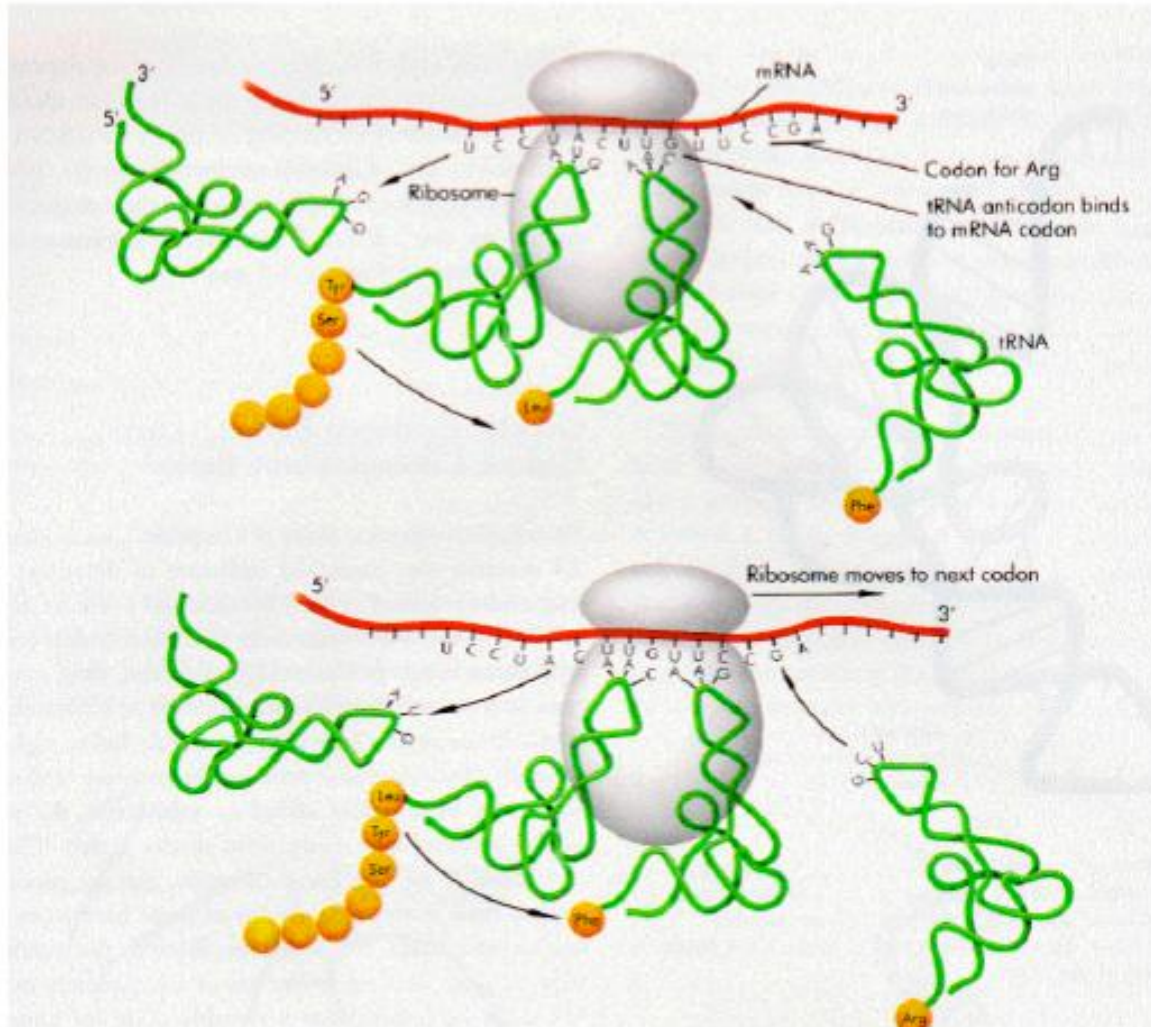
RNA Processing

Primary
mRNA



β-Globin
mRNA

Translation



<http://bioweb.uwlax.edu/GenWeb/Molecular/Theory/Translation/translation.htm>

Genetic code (Codons)

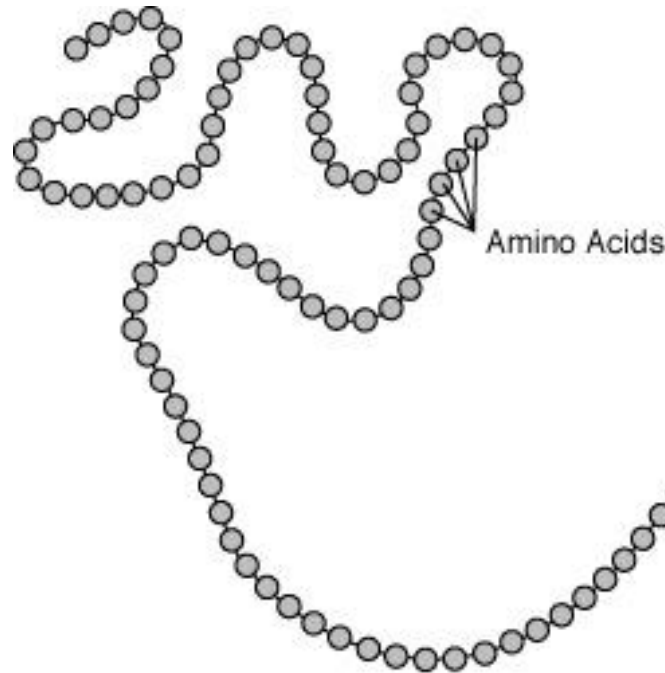
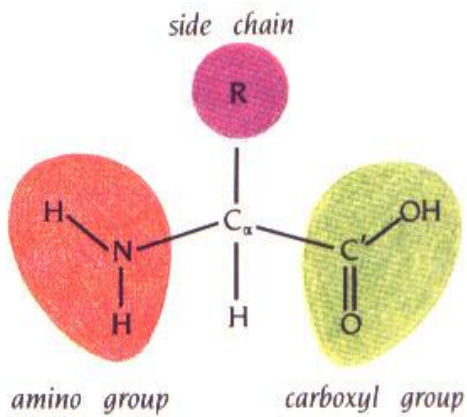
First Position	Second Position				Third Position
	U	C	A	G	
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
	Leu (L)	Ser (S)	Stop	Stop	A
	Leu (L)	Ser (S)	Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
	Leu (L)	Pro (P)	His (H)	Arg (R)	C
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

Six reading frames

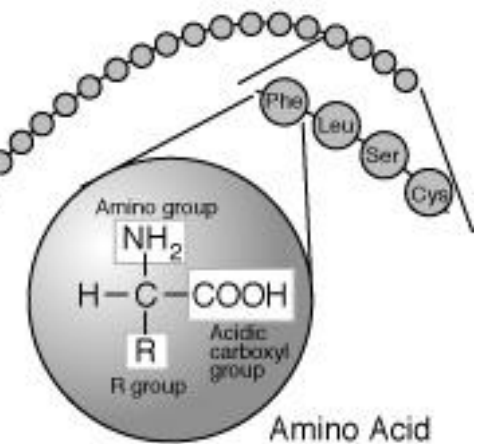
- 5' 3'
atgccaagctgaatagcgtagaggggttttcat
cattgaggacgatgta~~taa~~
 - Atg..
 - Tgc..
 - Gcc..
- 3 more reading frames on the reverse complement strand

Protein

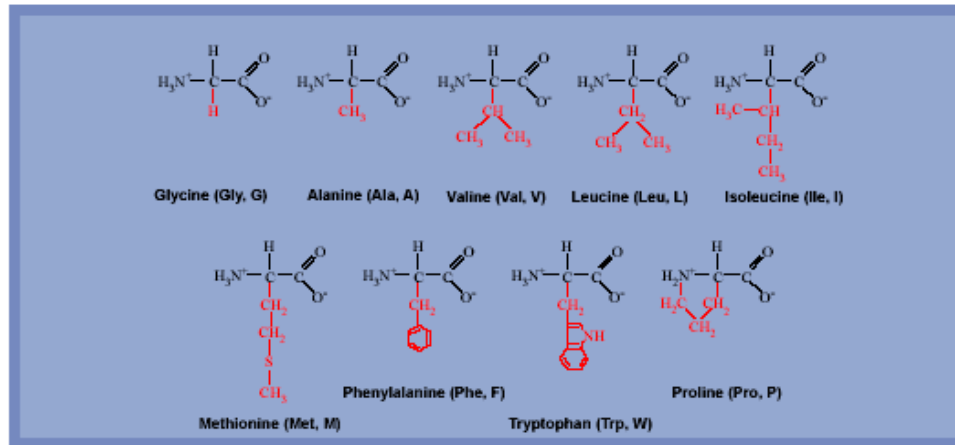
- 20 amino acids



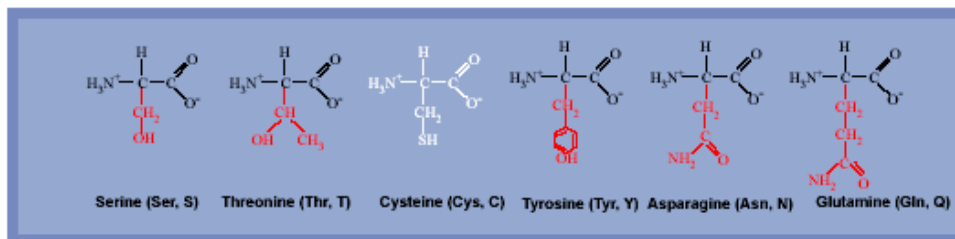
Primary protein structure is sequence of a chain of amino acids



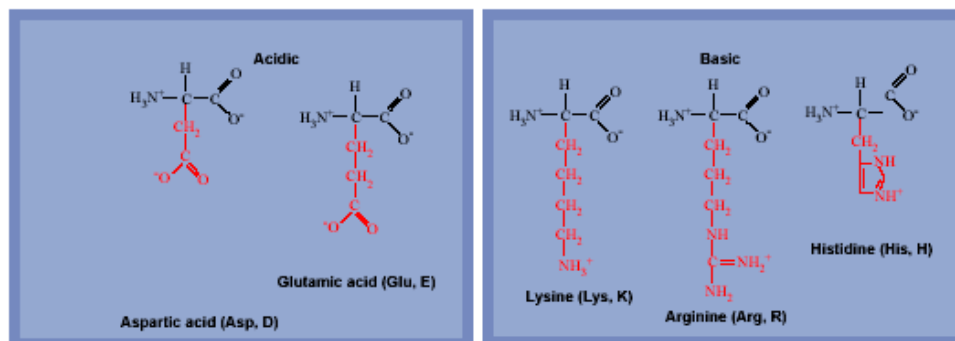
Nonpolar, Hydrophobic R-groups



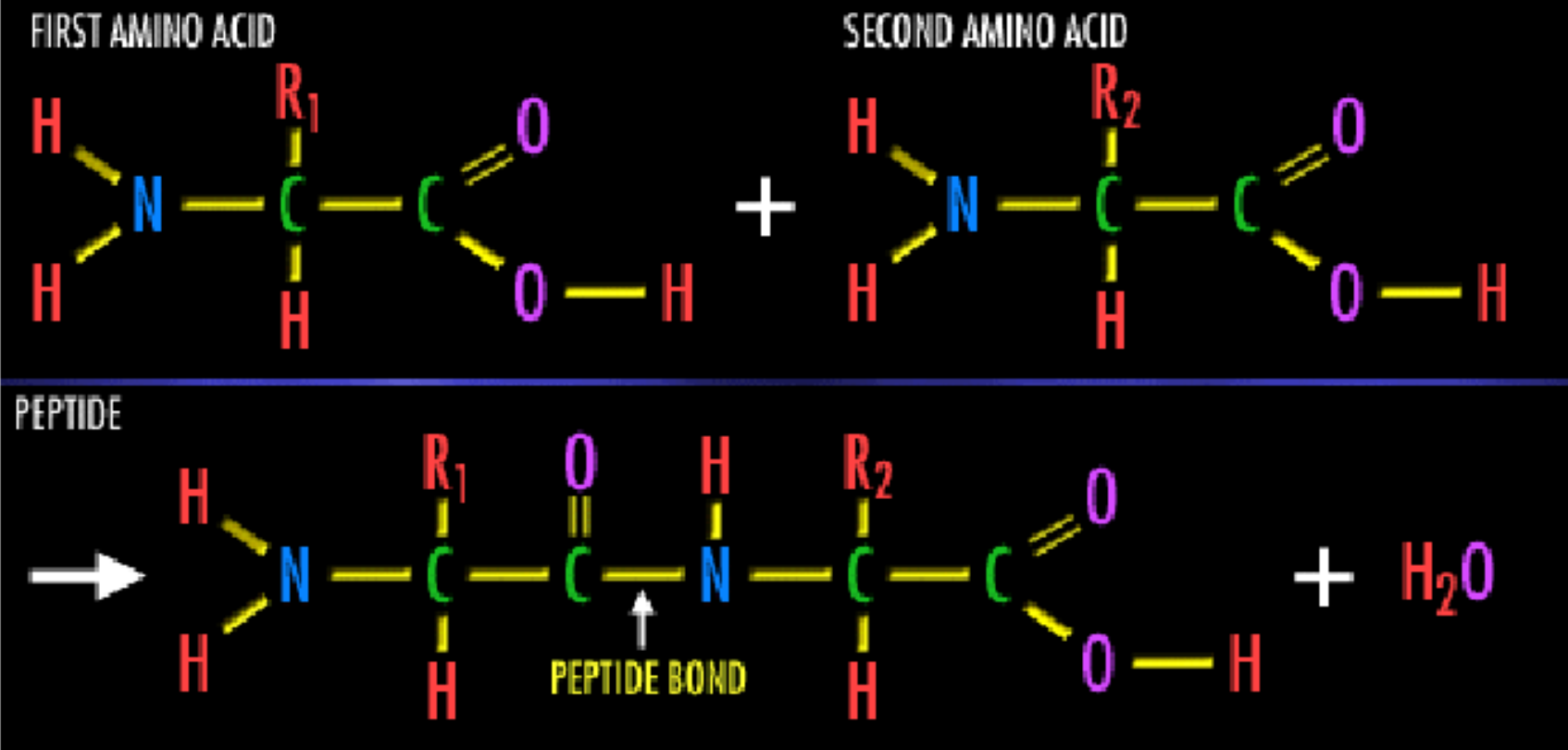
Polar, Hydrophilic R-groups



Electrically charged



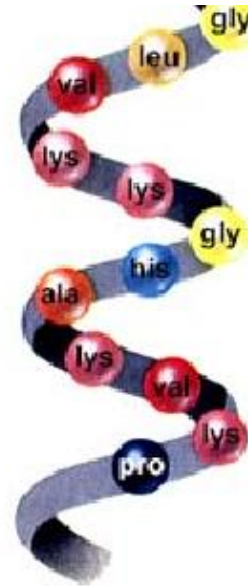
Peptide



Protein Structure



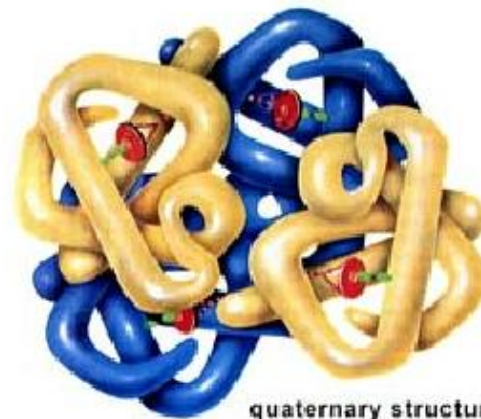
primary structure
(amino acid sequence)



secondary structure
(α -helix)



tertiary structure
(folded individual peptide)



quaternary structure
(aggregation of two or more peptides)

Acknowledgement

- Most of the slides were from Dr. Qi Liu's course materials.