Bioinformatics 1

Biology, Sequences, Phylogenetics

Part 1

Sepp Hochreiter

Master Bioinformatics

➥3 credits (plus 3 credits for exercises)

➥ first basic course: module M6 (required for other modules)

Biological Chemistry

- 2.6 credits (plus 3.2 credits for exercises)

- Module “informatics”

- Examination at the end of the class
- Course manuscript: http://www.bioinf.jku.at/teaching/ws2010/bin1/
- Also videos

Schedule Master Bioinf 2010 1. Sem.

320007/320061 Molekulare Biologie der Zelle I

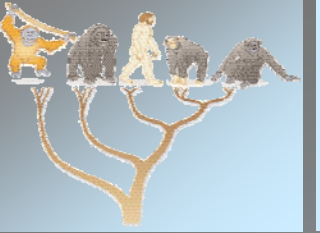
→

Meeting

Mondy 4. Oct. 11:00

in front of the elevator in the 10. Level of the TNF tower

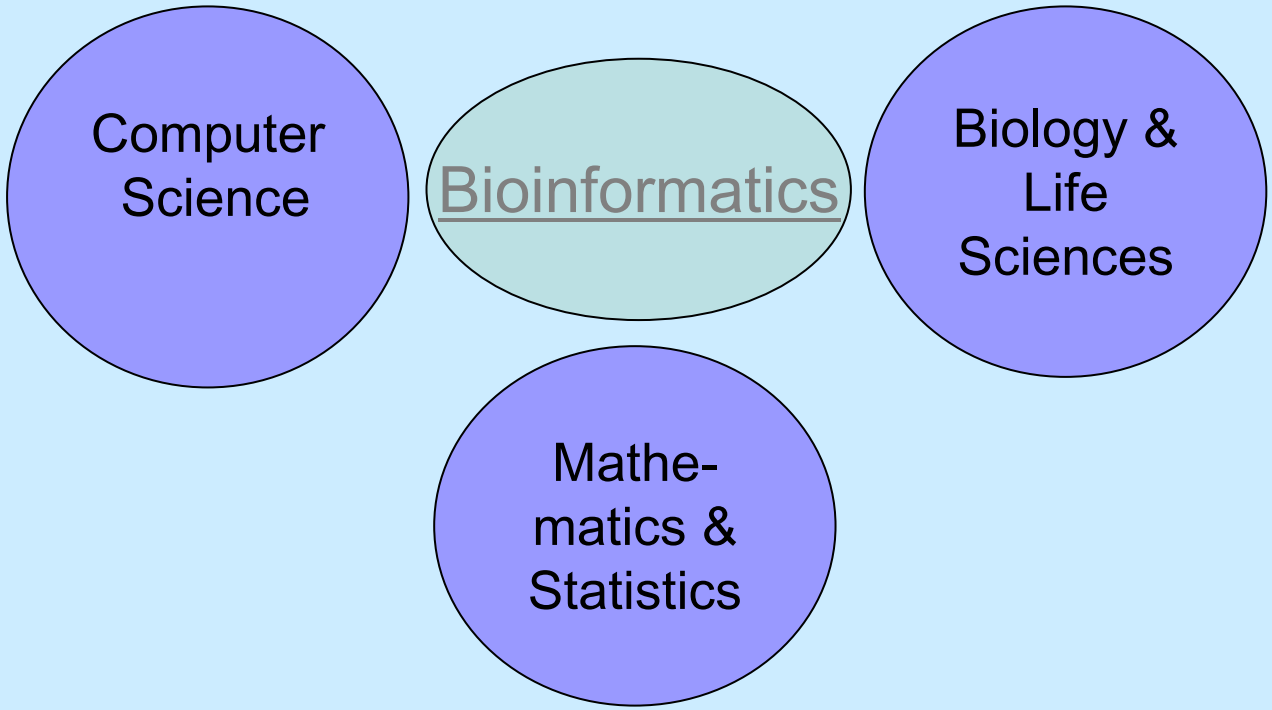
	Mon	Tue	Wed	Thur	Fri
08:30-09:15				Alternative for 340023 (Monday) 326011 Algorithms and data structures 2KV	
09:15-10:00					
10:15-11:00	348408 Introduction I to Gen Chemistry 1VO		365027 Bioinformatics I 2UE	332056+332057 Statistical Signal Processing 2VO+2UE	365046 Numerical and symbolical methods for Bioinformatics 2KV
11:00-11:45					
12:00-12:45	344014 Arti .Intelligence 2VO				
12:45-13:30					
13:45-14:30	340023 Algorithms and Data Structures 2 2VO (Alternative: 326011, Thursday)	351001 Information Systems 1 2VO	347334 Chemistry for Physicist II 2VO	351003/351004 Information Systems 1 2UE	348408 Introduction I to General Chemistry 1VO
14:30-15:15					
15:30-16:15	365027 Bioinformatics I 2VO	347333 Introduction into instrumental analytics for Life-Sciences 3KV	347333 Introduction into instrumental analytics for Life-Sciences 3KV	351002/351005 Information Systems 1 2UE	
16:15-17:00					



What is Bioinformatics?

- Interface of biology and computers
- Analysis of proteins, genes and genomes using computer algorithms and computer databases
- Analysis and storage of the billions of DNA base pairs that are sequenced by genomics projects

What is Bioinformatics?



What is Bioinformatics?

- "Bioinformatics is a new subject of genetic data collection, analysis and dissemination to the research community." *Hwa A. Lim (1987)*

- *"Bioinformatics:* Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data,including those to acquire, store, organize, archive, analyze, or visualize such data." *NIH working definition (2000)*

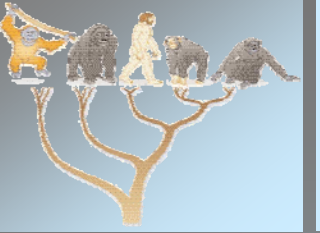
What is Bioinformatics?

http://en.wikipedia.org/wiki/Bioinformatics

- creation and advancement of *databases*, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the *management* and analysis of biological data. →BI IV
- mapping and analyzing DNA and protein sequences, aligning different DNA and protein sequences to compare them and *creating and viewing 3-D models of protein structures*. → BI III
- focus on developing and applying computationally intensive techniques (e.g., *pattern recognition, data mining, machine learning algorithms, and visualization*) →BI II
- Major research efforts in the field include sequence alignment, *gene finding, genome assembly, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, genome-wide association studies and the modeling of evolution*. →BI III

Questions Answerd by Bioinformatics

- From where came the first human?
- Is Anna Anderson the tsar‘s daughter Anastasia?
- Are the neanderthals the ancestors of the humans?
- What are the evolutionary relationships between species



homo erectus

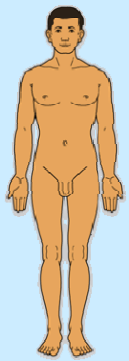
homo sapiens

Is Anna Anderson the tsar‘s daughter Anastasia Romanov?

Anastasia (1909)
and
Anna Anderson (1959)

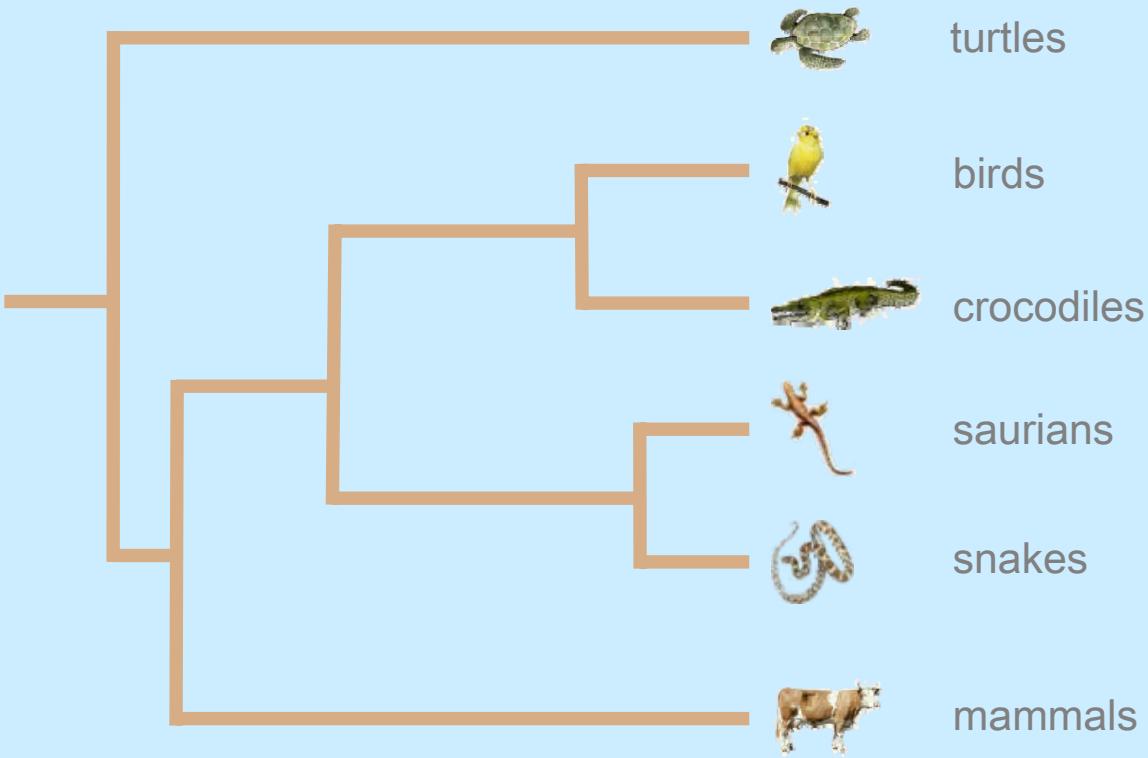
The kids of
the tsar

Are the neanderthals the human ancestors or a different species?



Phylogeny: history of species

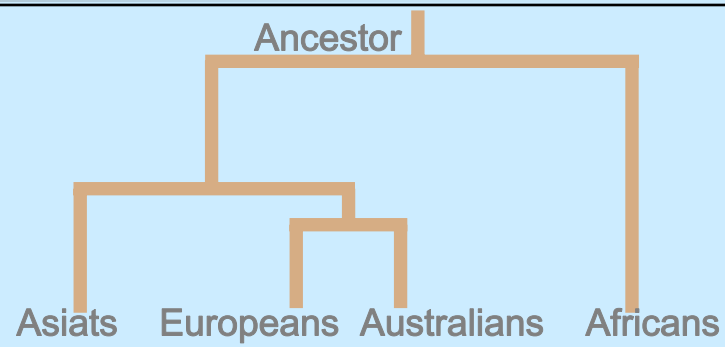
Phylogenetic knowledge: evolutionary trees



Three Anwers

➥From where came the first human?

Africa!



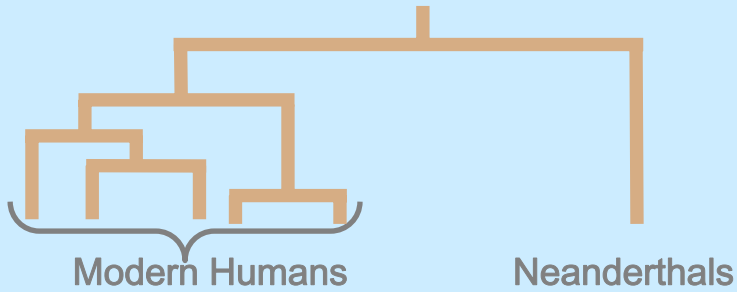
➥Is Anna Anderson the tsar‘s daughter Anastasia?

No!

	91	106	324	337
Anna Anderson	CCACCATGAATATTGC		TAGTCAAATCCCTT	
Carl Maucher (Grand nephew F. Schanzkowska)	CCACCATGAATATTGC		TAGTCAAATCCCTT	
Prince Philip (Grand nephew zar)	TCACCATGAATATTGT		CAGTCAAATCCCTC	

➥Are the neanderthals the ancestors of the humans?

No! Separate Species



Literature

- D. W. Mount, Bioinformatics: Sequences and Genome analysis, CSHL Press, 2001
- D. Gusfield, Algorithms on strings, trees and sequences: computer science and computational biology, Cambridge Univ. Press, 1999
- R. Durbin, S. Eddy, A. Krogh, G. Mitchison, Biological sequence analysis, Cambridge Univ. Press, 1998
- M. Waterman, Introduction to Computational Biology, Chapmann & Hall, 1995
- Setubal and Meidanis, Introduction to Computational Molecualar Biology, PWS Publishing, 1997
- Pevzner, Computational Molecular Biology, MIT Press, 2000
- J. Felsenstein: Inferring phylogenies, Sinauer, 2004
- W. Ewens, G. Grant, Statistical Methods in Bioinformatics, Springer, 2001
- Blast: http://www.ncbi.nlm.nih.gov/BLAST/tutuotial/Altschul-1.html

Contents

1 Biological Basics
1.1 The Cell
1.2 Central Dogma of Molecular Biology
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Introns, Exons, and Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Contents

2 Bioinformatics Rescources
2.1 Data Bases
2.2 Software
2.3 Articles

Contents

3 Pairwise Alignment
3.1 Motivation
3.2 Sequence Similarities and Scoring
3.2.1 Identity Matrix
3.2.2 PAM Matrices
3.2.3 BLOSUM Matrices
3.2.4 Gap Penalties
3.3 Alignment Algorithms
3.3.1 Global Alignment - Needleman-Wunsch
3.3.2 Local Alignment - Smith-Waterman
3.3.3 Fast Approximations: FASTA, BLAST and BLAT
3.4 Alignment Significance
3.4.1 Significance of HSPs
3.4.2 Significance of Perfect Matches

Contents

4 Multiple Alignment
4.1 Motivation
4.2 Multiple Sequence Similarities and Scoring
4.2.1 Consensus and Entropy Score
4.2.2 Tree and Star Score
4.2.3 Weighted Sum of Pairs Score
4.3 Multiple Alignment Algorithms
4.3.1 Exact Methods
4.3.2 Progressive Algorithms
4.3.3 Other Multiple Alignment Algorithms
4.4 Profiles and Position Specific Scoring Matrices

Contents

5 Phylogenetics
5.1 Motivation
5.1.1 Tree of Life
5.1.2 Molecular Phylogenies
5.1.3 Methods
5.2 Maximum Parsimony Methods
5.2.1 Tree Length
5.2.2 Tree Search
5.2.3 Weighted Parsimony and Bootstrapping
5.2.4 Inconsistency of Maximum Parsimony
5.3 Distance-based Methods
5.3.1 UPGMA
5.3.2 Least Squares
5.3.3 Minimum Evolution
5.3.4 Neighbor Joining
5.3.5 Distance Measures
5.4 Maximum Likelihood Methods
5.5 Examples

Biological Basics

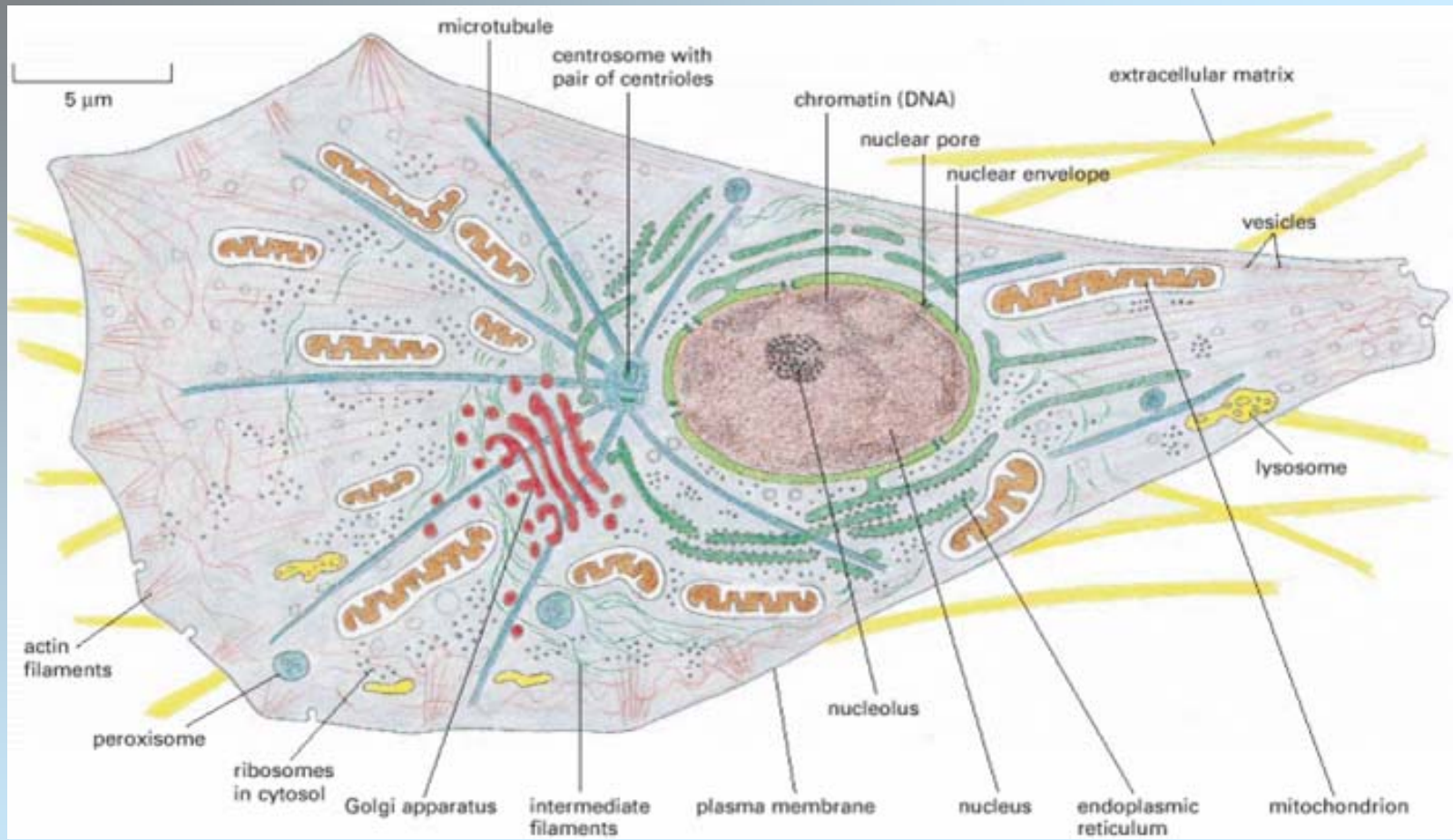
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Bioinformatics processes data from molecular biology

Molecular biology attempts at discovering the principles of the cell which is the largest unit all lifeforms have in common

The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



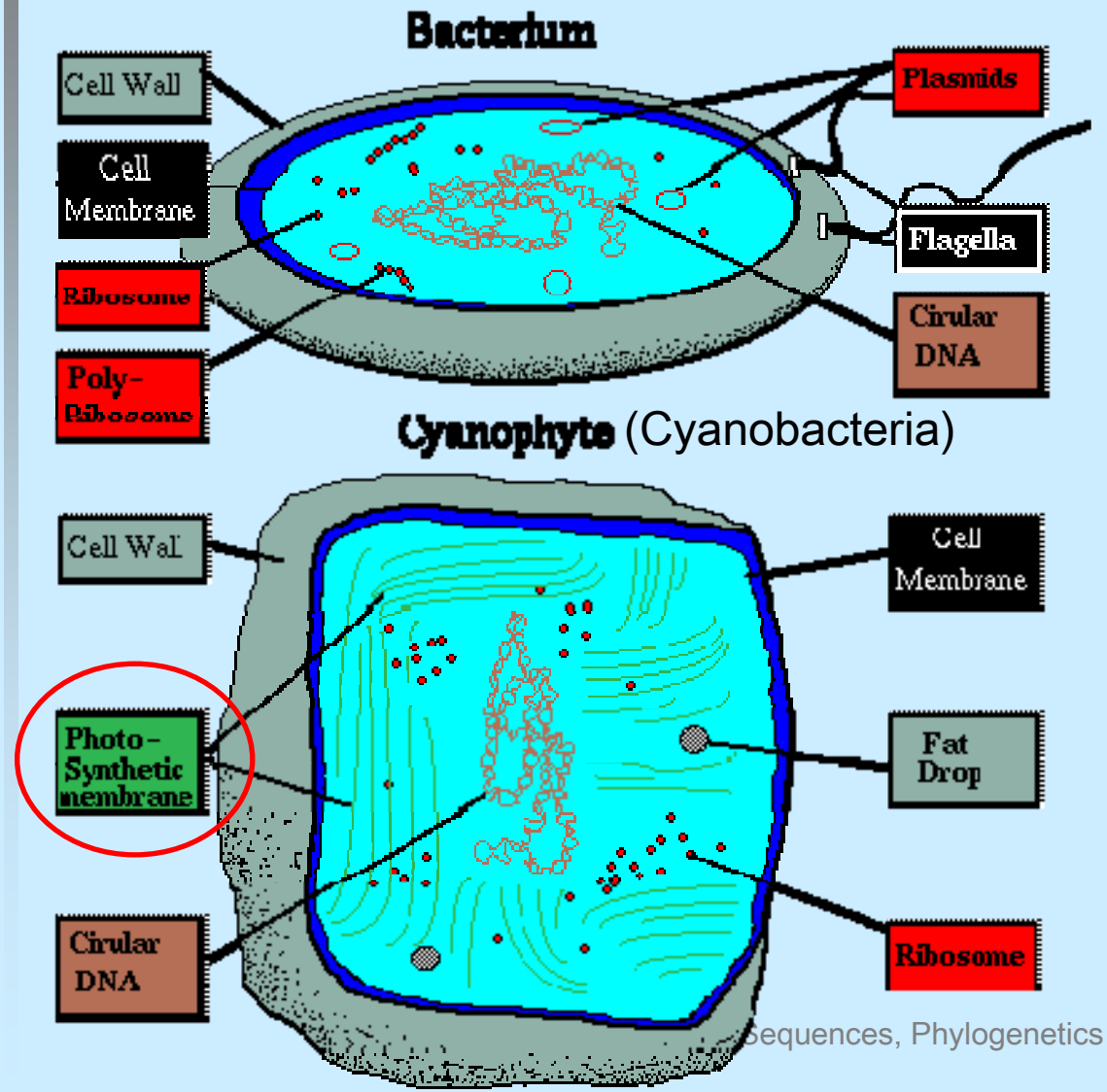
Organelles

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

- **nucleus**: eukaryotes only - a cell's information center →chromosomes, DNA
- **nucleolus**: ribosome subunits assembly
- **mitochondria** and **chloroplasts**: the power generators (oxidative phosphorylation, photosyntesis); contain their own genome form mother
- **endoplasmic reticulum** (ER) – eukaryotes only : transport network for certain modifications and specific destinations; the rough ER, which has ribosomes on its surface and secretes proteins into the cytoplasm
- **golgi apparatus** – eukaryotes only : process and package macromolecules; vesicles
- **ribosomes**: RNA/protein complex that synthesizes proteins from mRNA
- **lysosomes**: enzymes (acid hydrolases) that digest old organelles, food particles, and engulfed viruses or bacteria
- **peroxisomes**: rid the cell of toxic peroxides
- **centrosome**: cytoskeleton, microtubules for directing the transport through ER and golgi apparatus
- **vacuoles**: store food and waste.

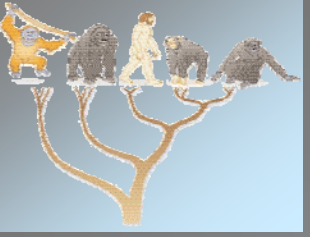
The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



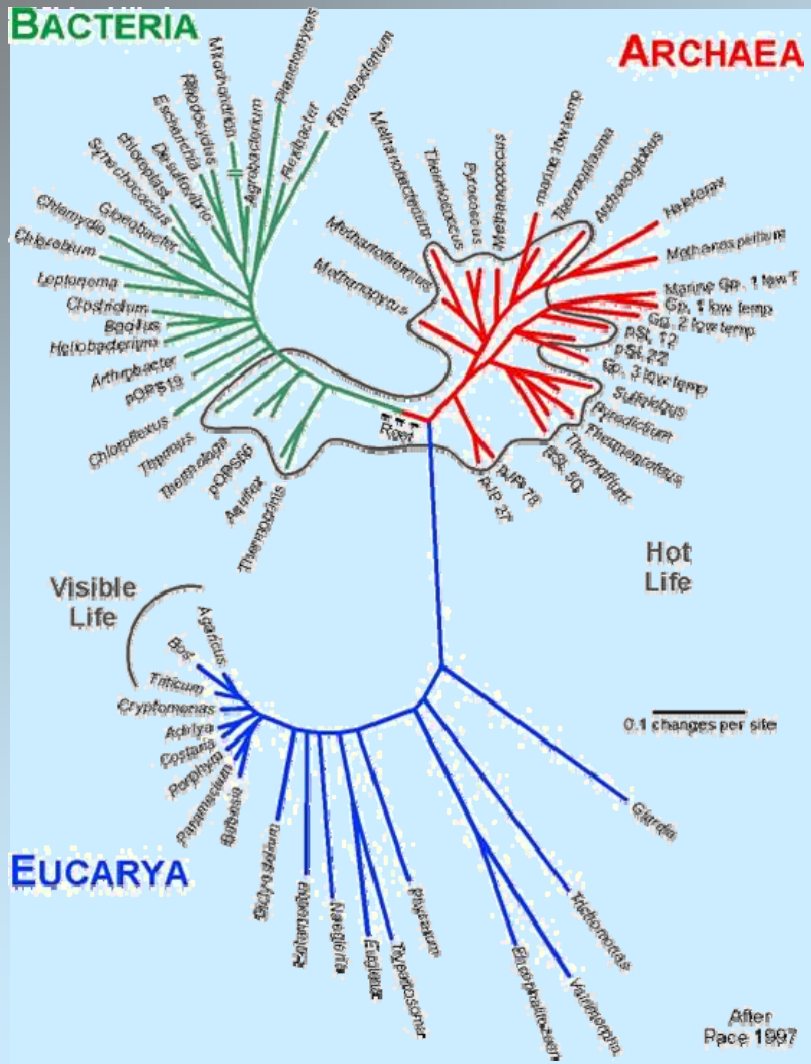
Eukaryotic cells possess a nucleus (plants, vertebrates)

Prokaryotic cells do not possess a nucleus (bacteria, archaea)



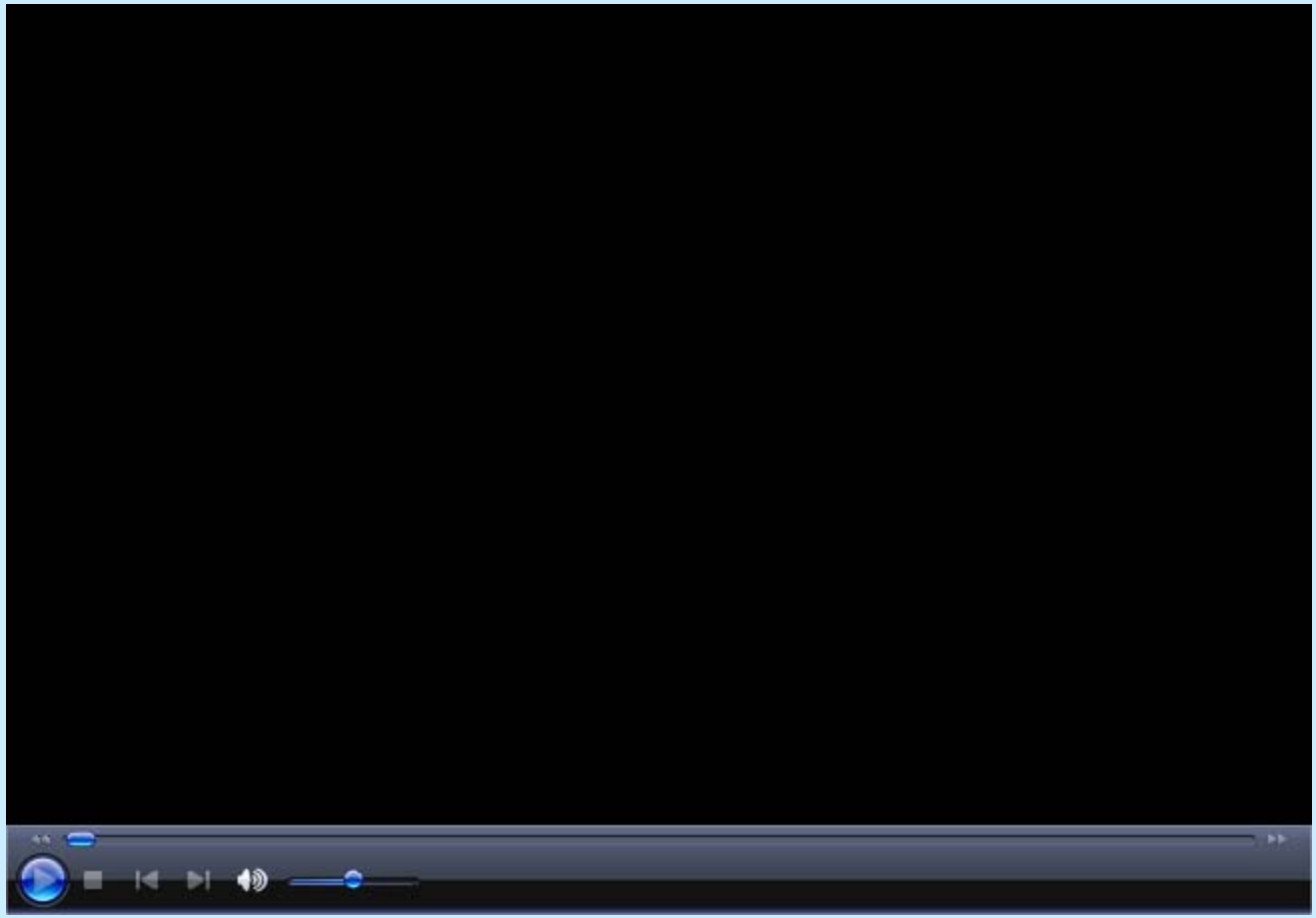
The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



Cell surface

The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Glycolisis

The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Pathway glycogen

The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Pathway hormone

Central Dogma

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

How are the nano-machines in the cell constructed?

↪These machines are proteins or protein-RNA complexes

Where is the information about these machines stored?

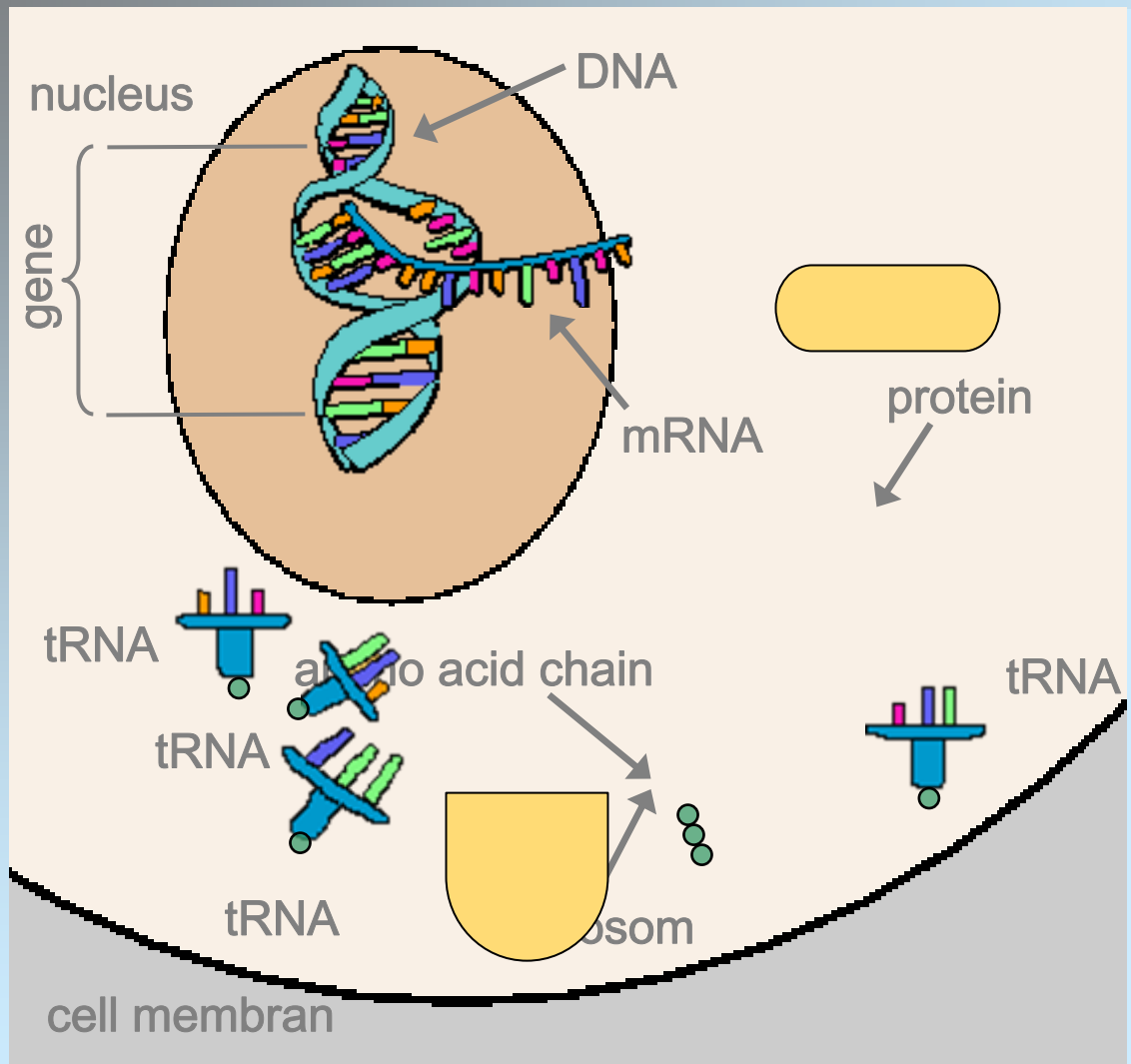
↪Everything is stored in the DNA

How is the information in the DNA used to build proteins?

↪Central dogma: DNA ➔ RNA ➔ Proteins

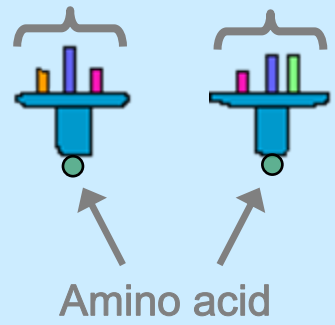
Central Dogma

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



1. transcription (mRNA)
2. transport
3. translation (ribosom, tRNA)
4. folding (protein)

codons/basetriplets



Central Dogma

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Dogma

DNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Deoxyribonucleic acid (DNA) codes all information of life

- ➥double helix as sequence of nucleotides with a deoxyribose
- ➥ends are called 5' and 3‘; DNA is written from 5' to 3‘
- ➥upstream is towards the 5' end downstream towards the 3'

- ➥5 nucleotides (nucleobases, bases): adenine (A), thymine (T), cytosine (C), guanine (G), and uracil (U)
- ➥first 4 in DNA whereas uracil in RNA instead of thymine
- ➥two classes: purines (A, G) / pyrimidines (C, U, T)

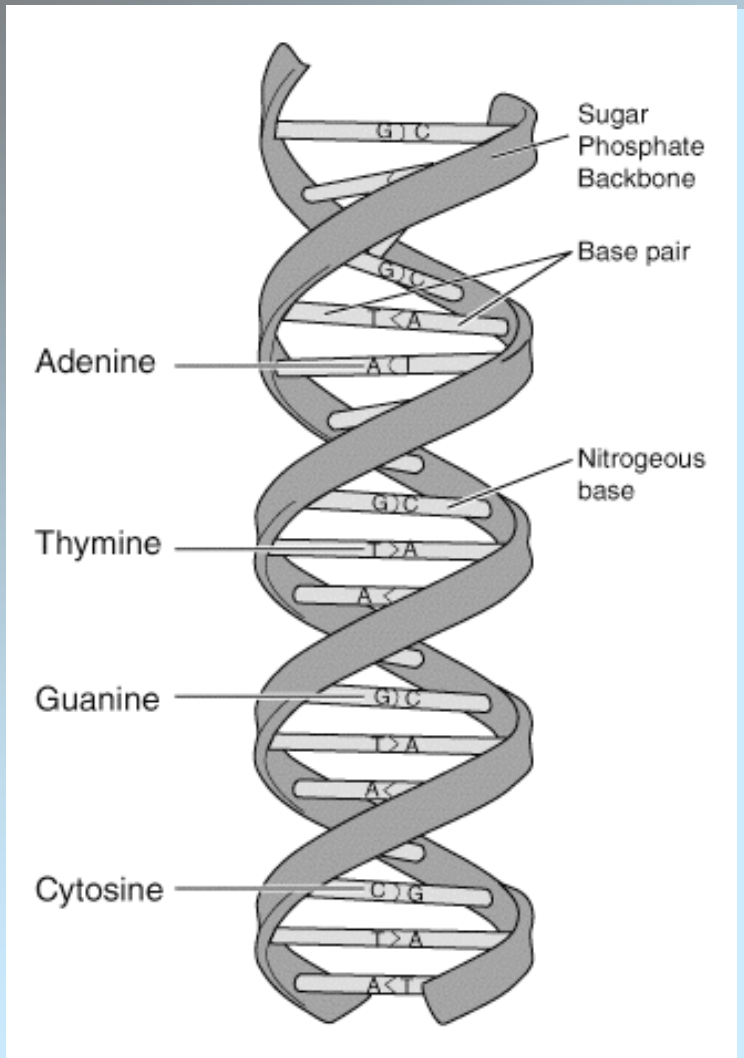
DNA

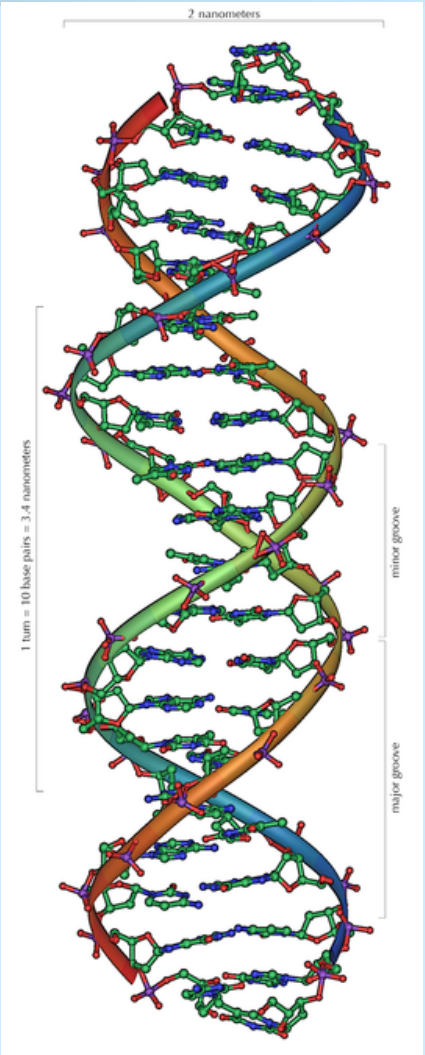
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

- hydrogen bonds between purines and pyrimidines
- base pairs: A—T and C—G
- each helix of the DNA is complementary to the other

DNA

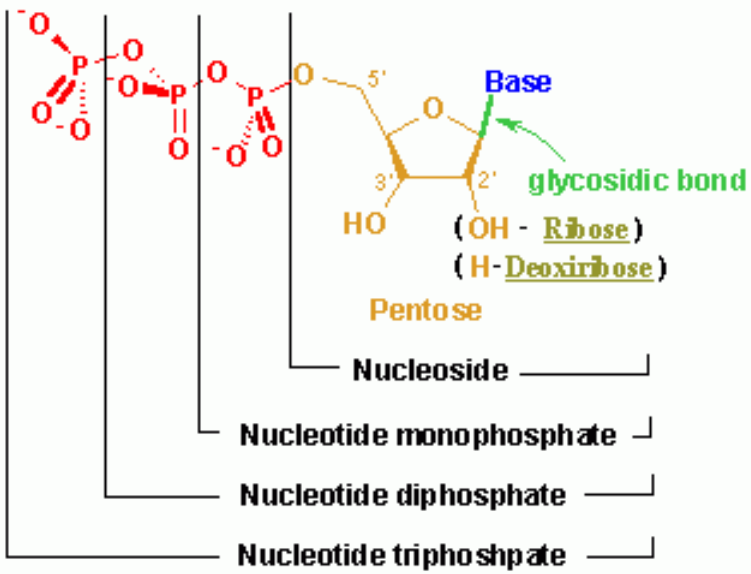
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

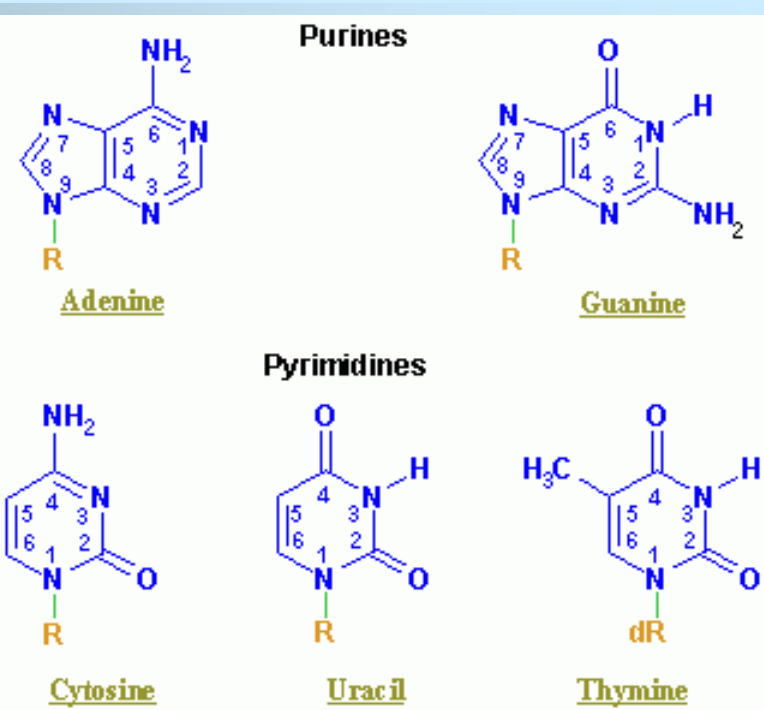




DNA

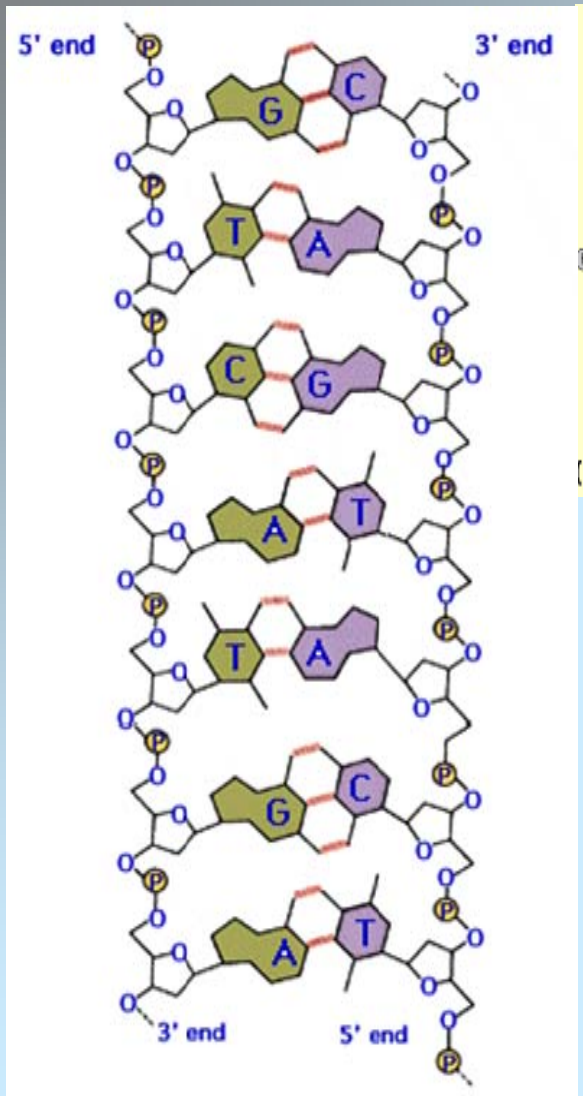
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

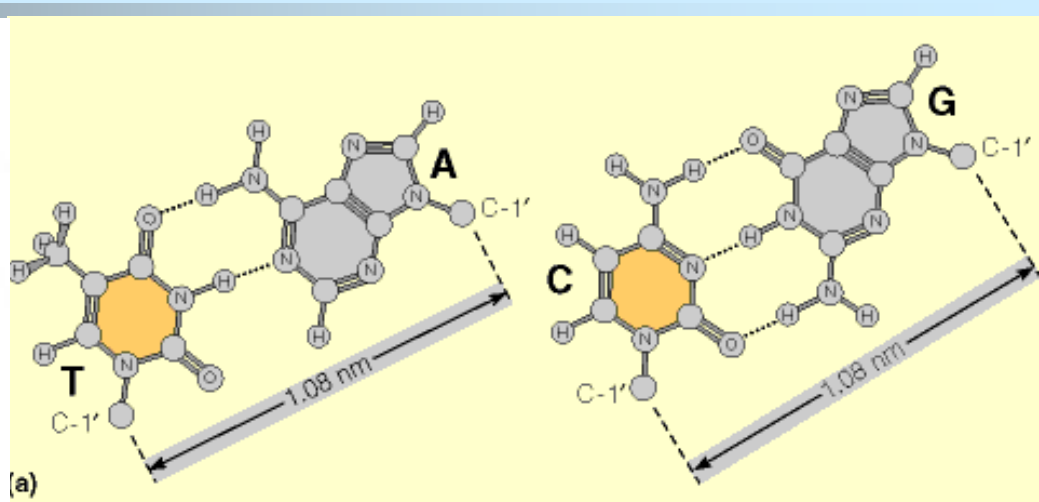


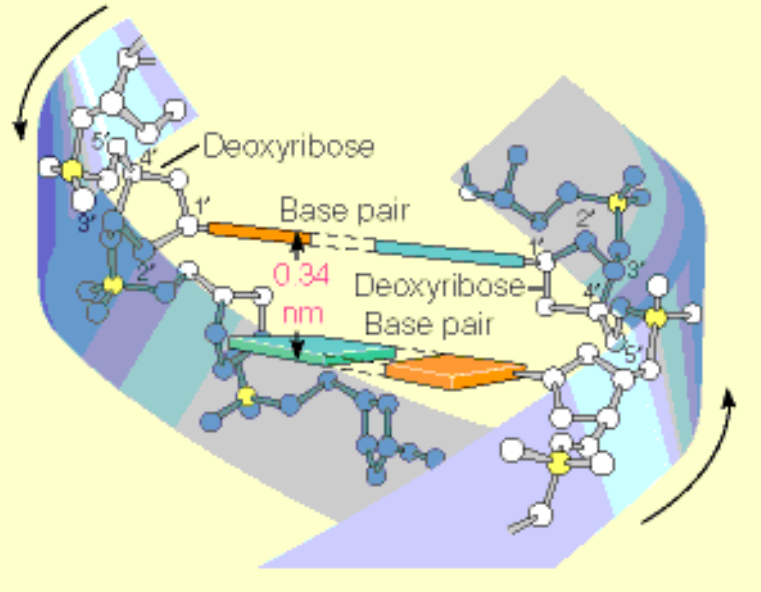


DNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding







DNA

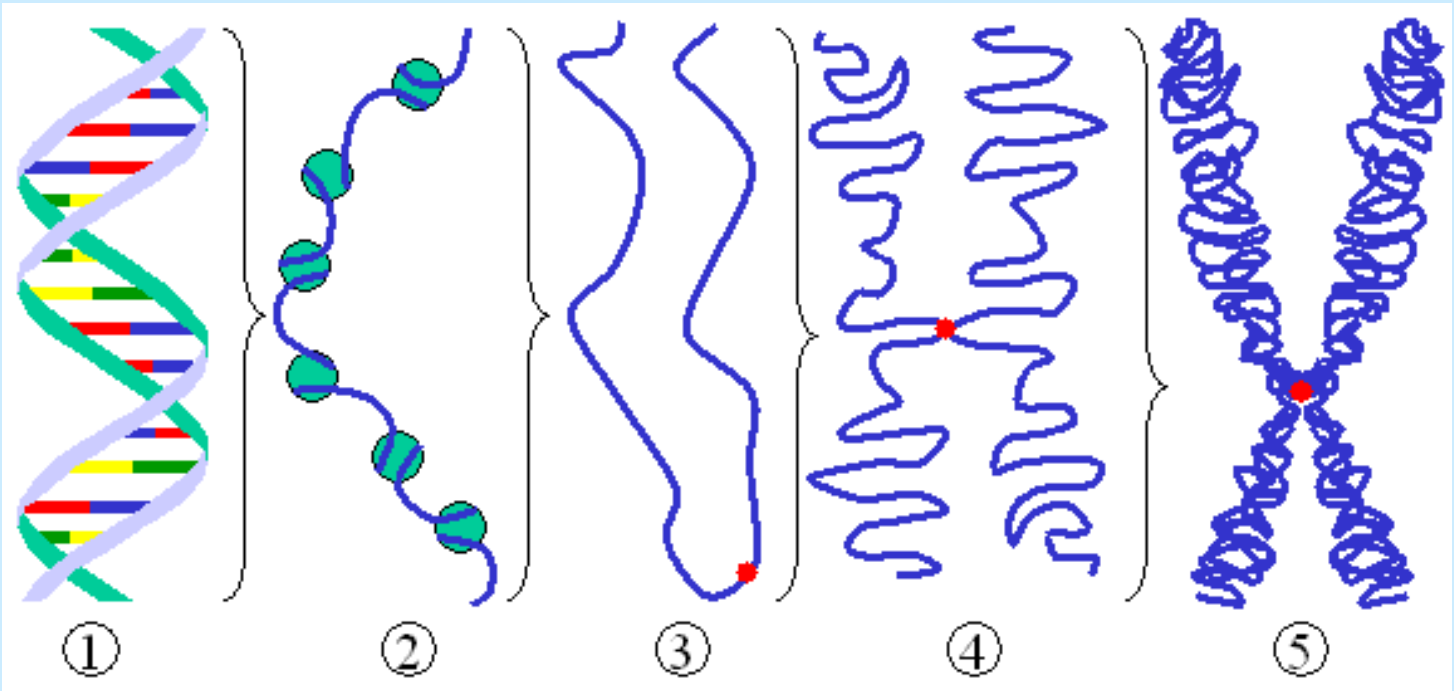
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

DNA

DNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

- The DNA is condensed in the nucleus in the chromosomes
- DNA wraps around histones resulting in chromatin
- Two chromatins linked at the centromere are a chromosome



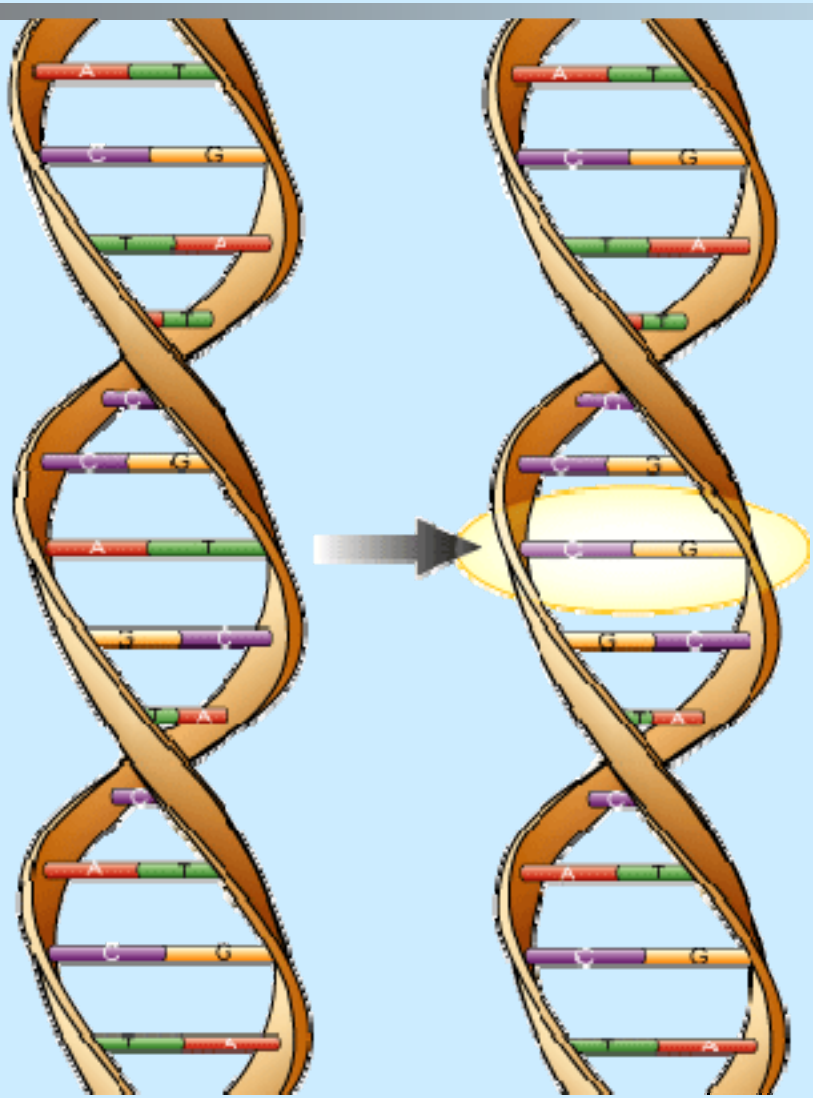
DNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

- Single DNA nucleotides differ at each human
- Small differences are inherited from both parents (except maternal mitochondrial DNA)
- Variation in the DNA at the same position in at least 1% of the population: single nucleotide polymorphism (SNP -- pronounced snip)
- SNPs occur all 100 to 300 base pairs
- Current research relate diseases to SNPs (schizophrenia or alcohol dependence).

DNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

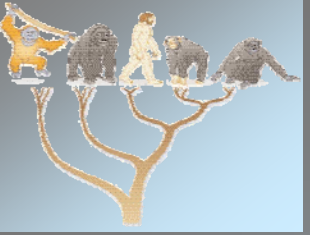


RNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

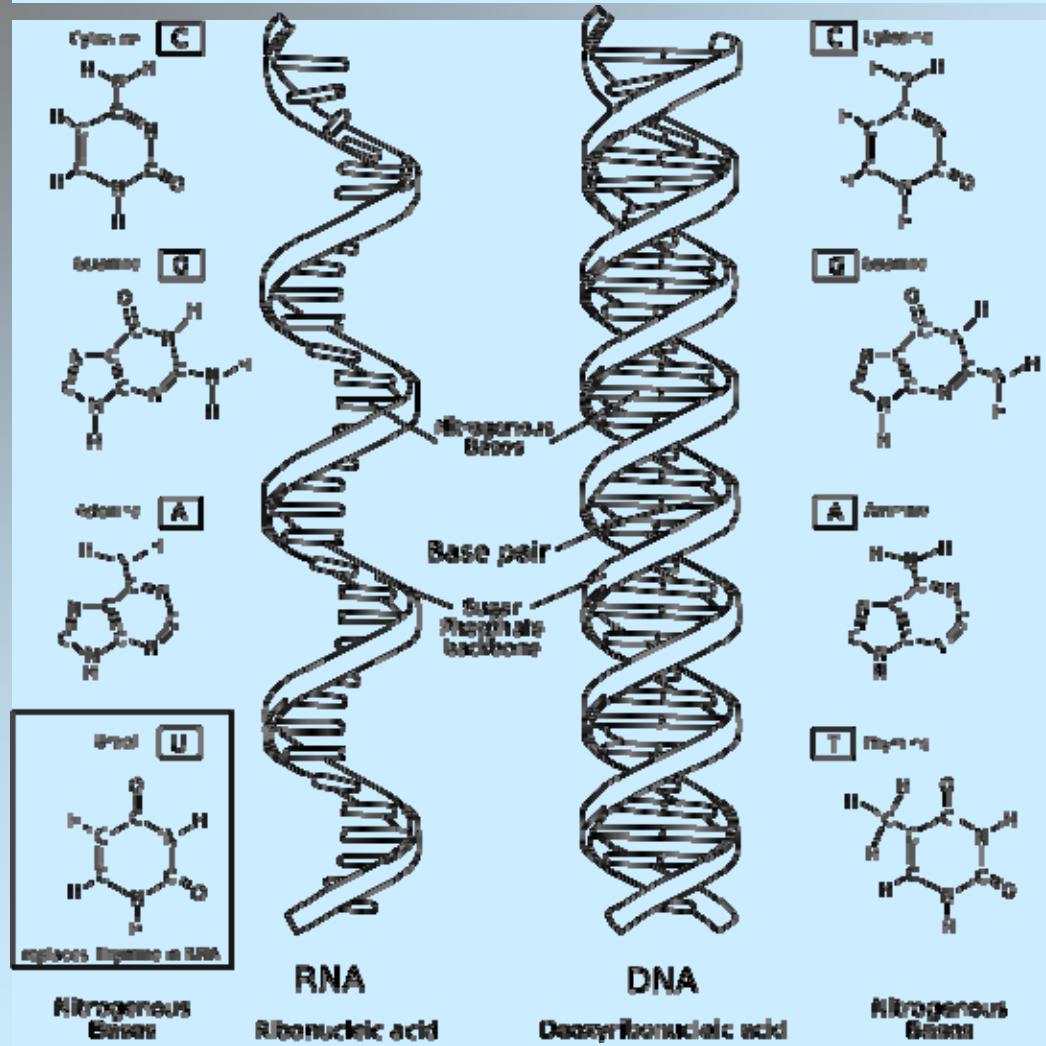
- Ribonucleic acid (RNA): sequence of nucleotides
- Contrast to DNA: ribose rings instead of deoxyribose; uracil instead of thymine
- transcribed from DNA through RNA polymerases
- kind of RNA:

mRNA (messenger),		tRNA (codon coding),
dsRNA (double stranded),		miRNA (micro),
RNAi (interference),		siRNA (small interfering),
ncRNA (non-coding)	like	rRNA (ribosomal)



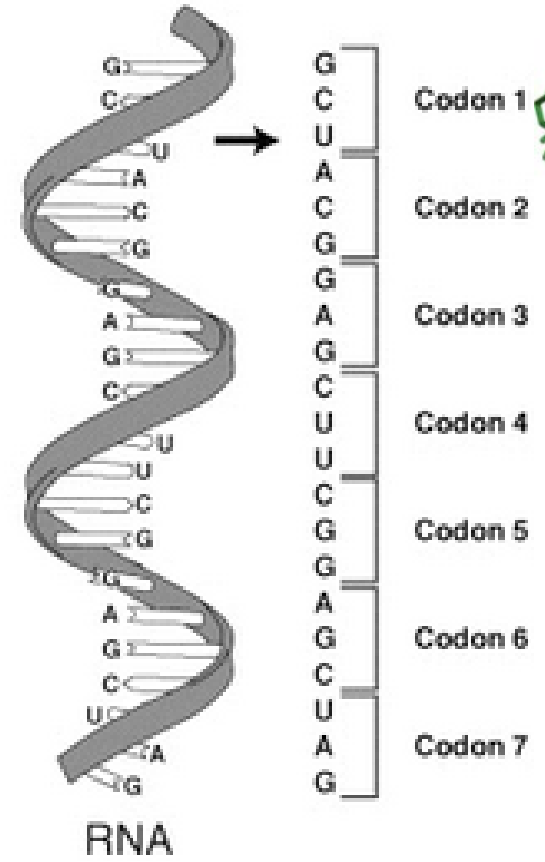
RNA

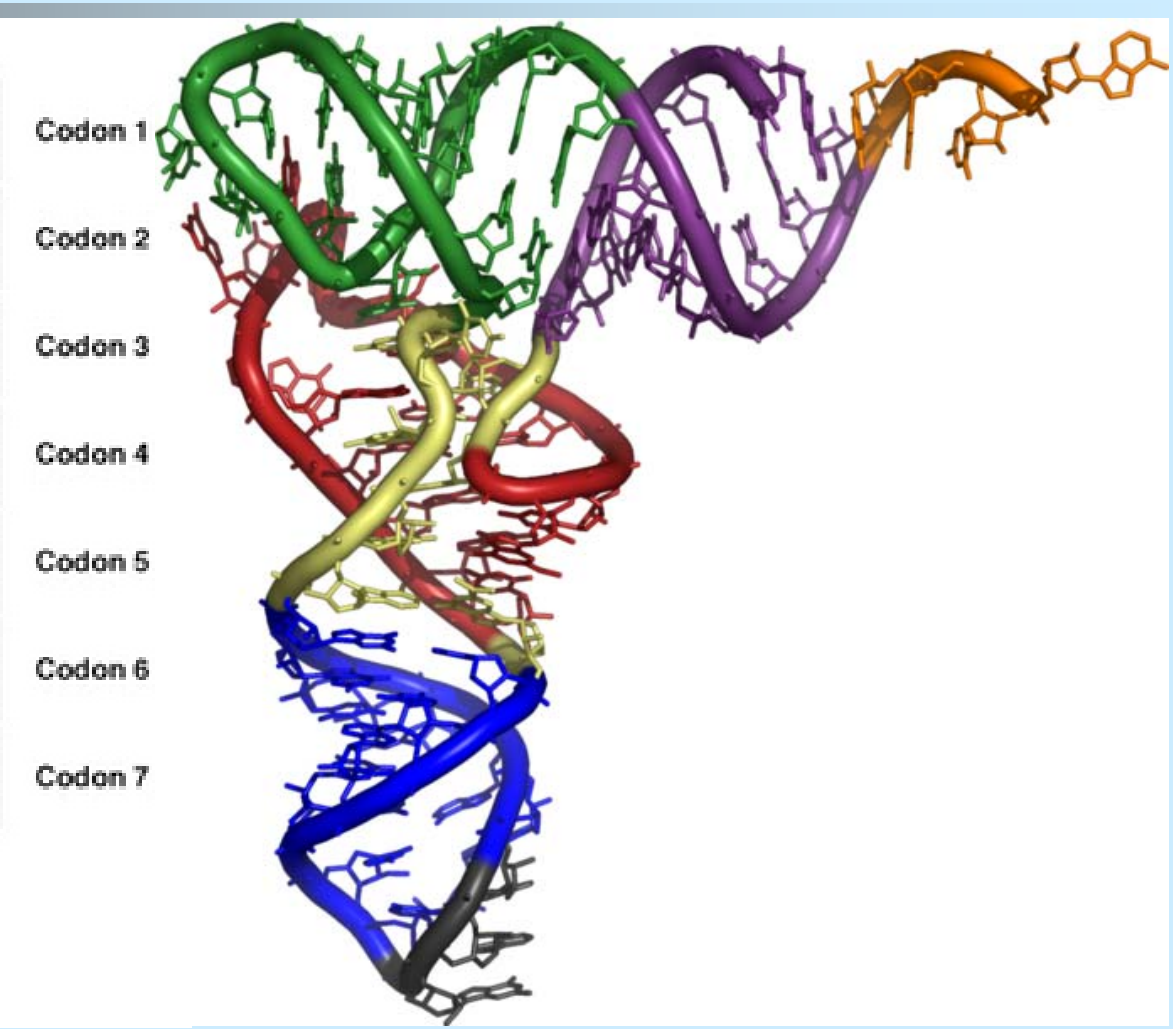
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



RNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding





RNA

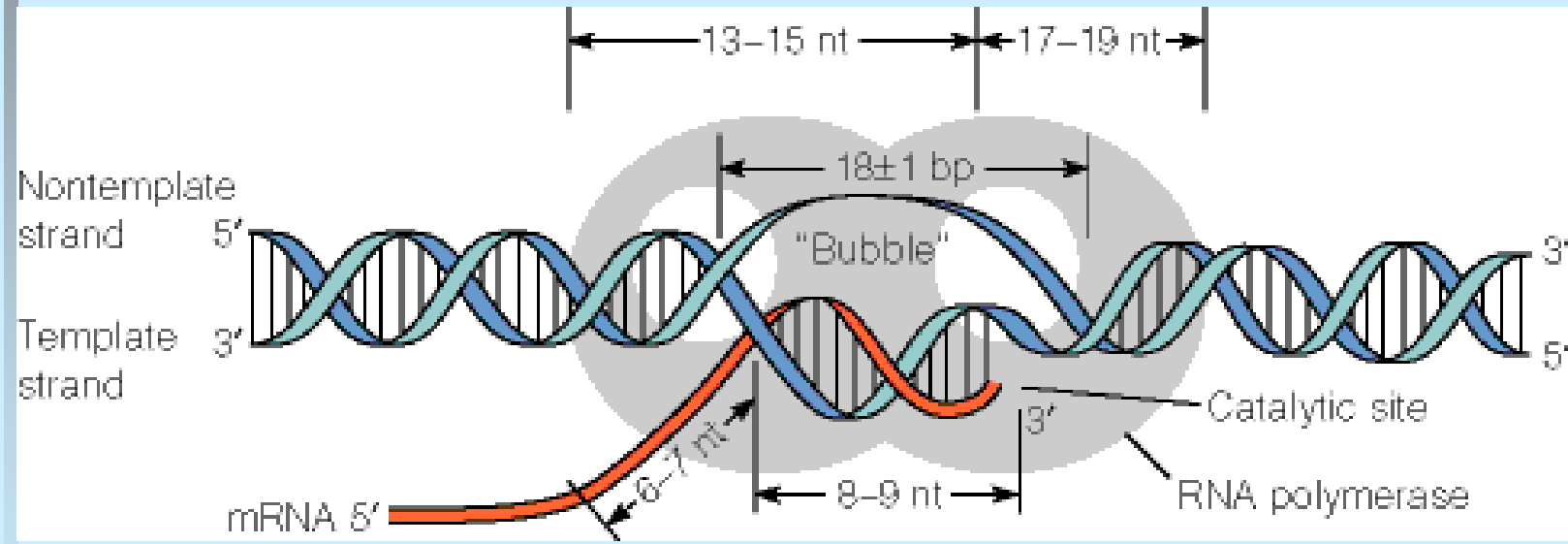
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

tRNA

Transcription

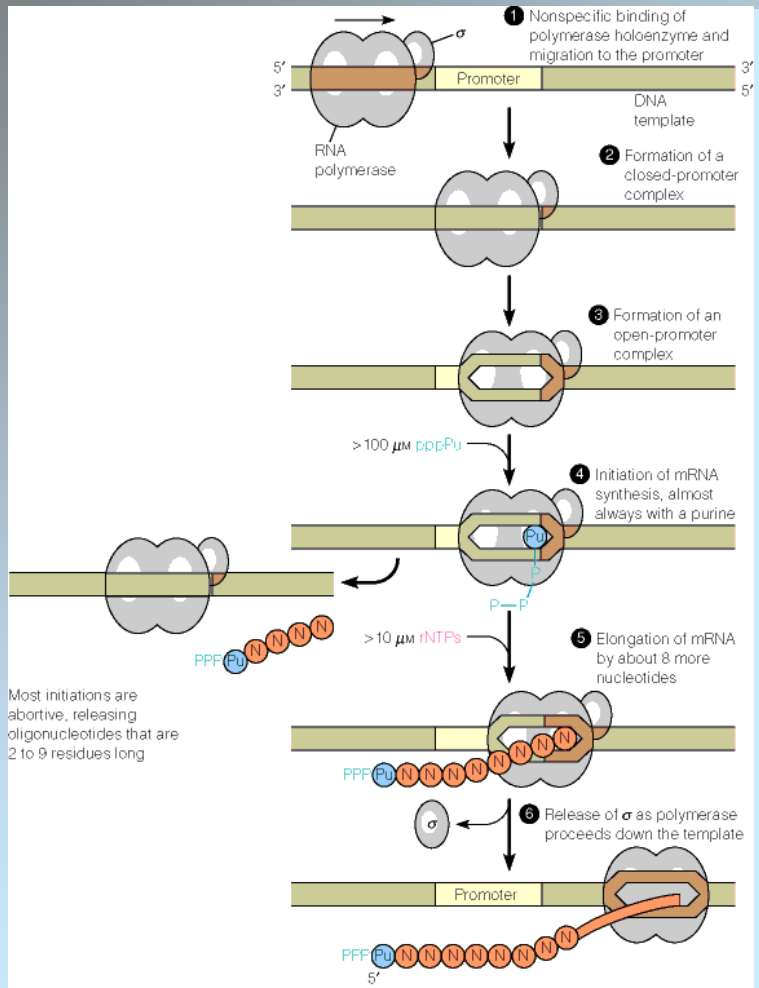
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

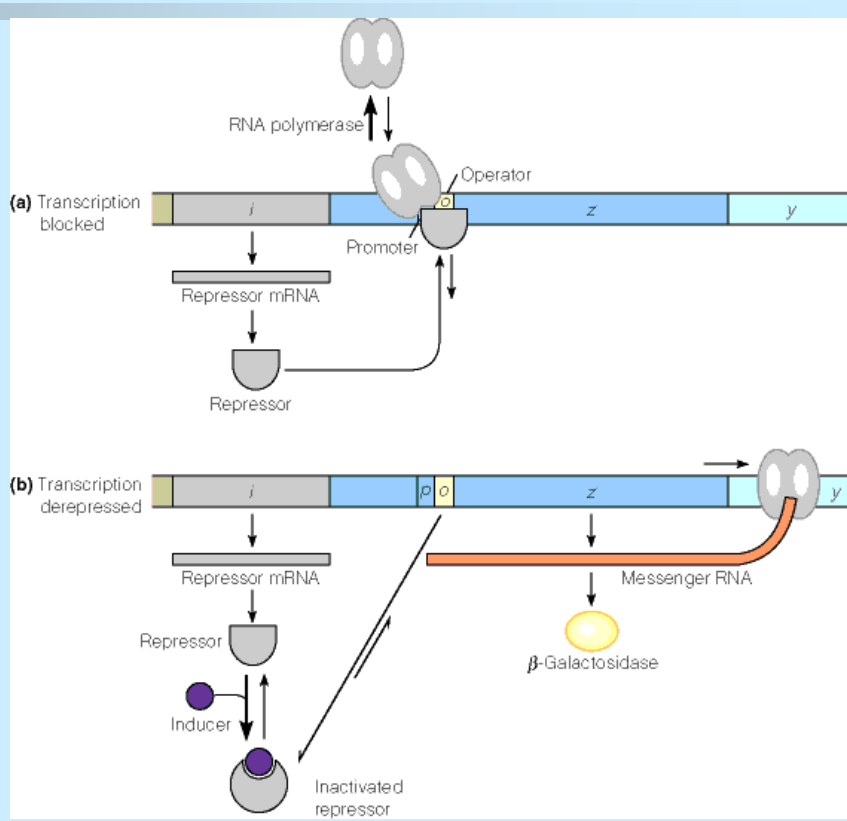
Transcription is the process of reading out a RNA (mRNA) from the DNA



Transcription Initiation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding





Transcription Initiation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Transcription

Transcription Initiation

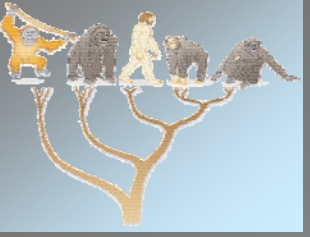
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Transcription inhibition

Transcription Elongation

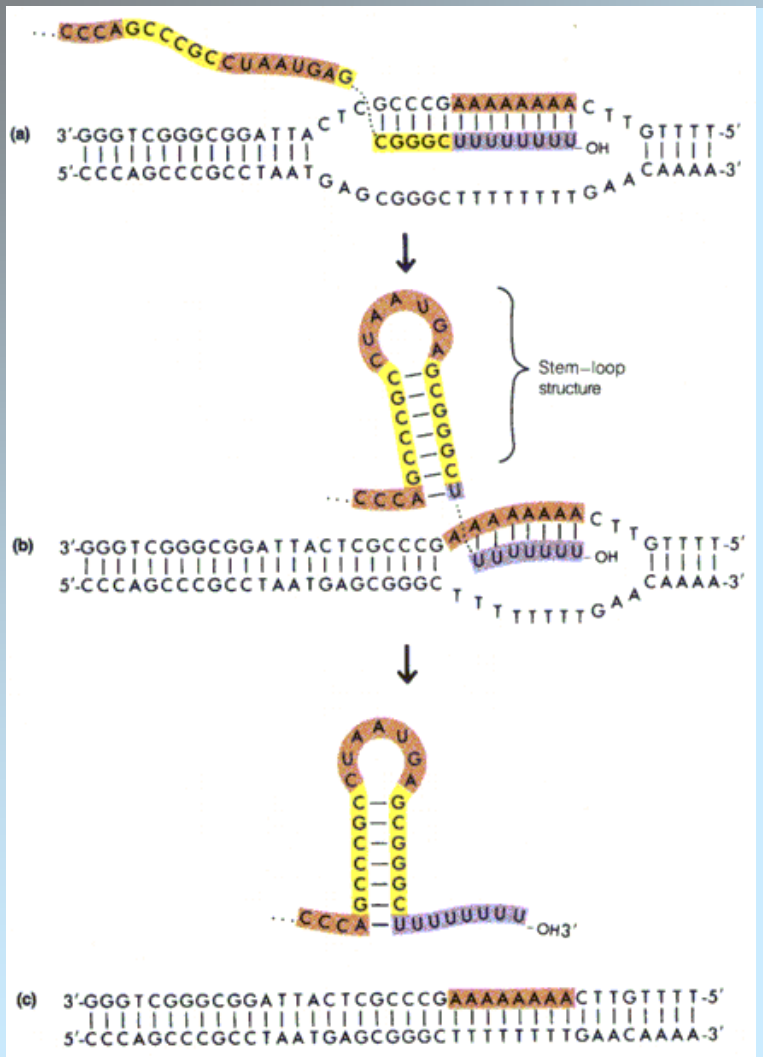
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

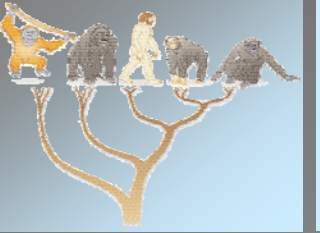
- After 8 nucleotides the sigma-subunit is dissociated from polymerase
- For elongation there exist promoters



Transcription Termination

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

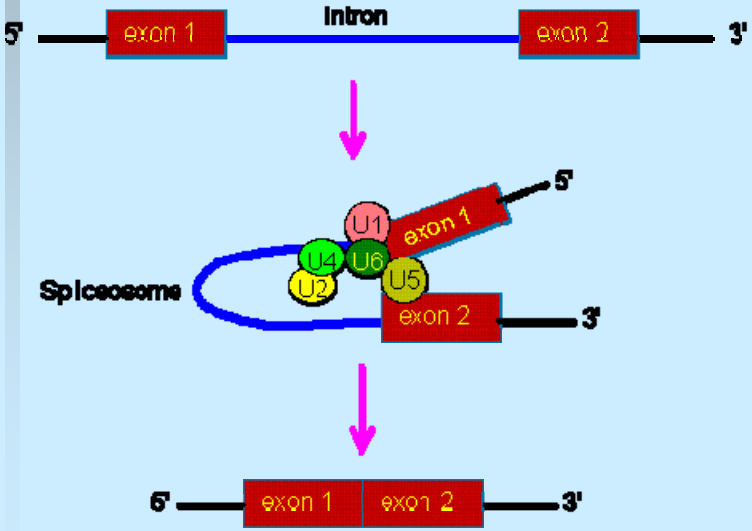


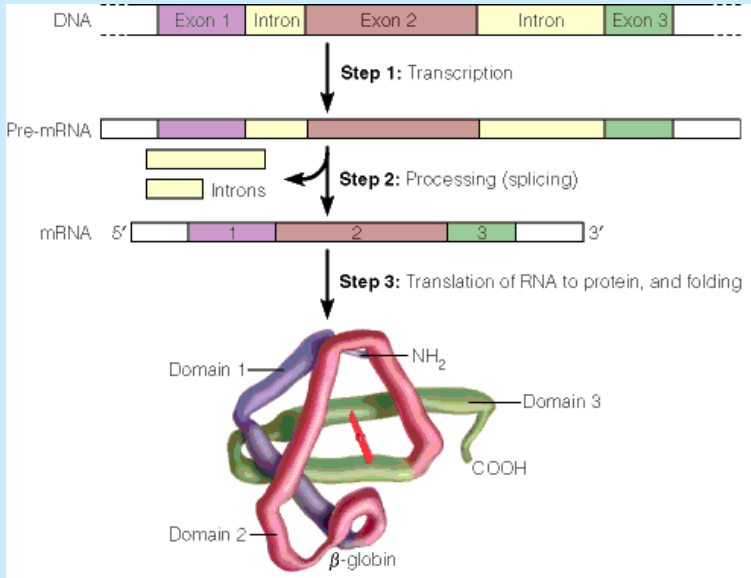


Splicing, Exons and Introns

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

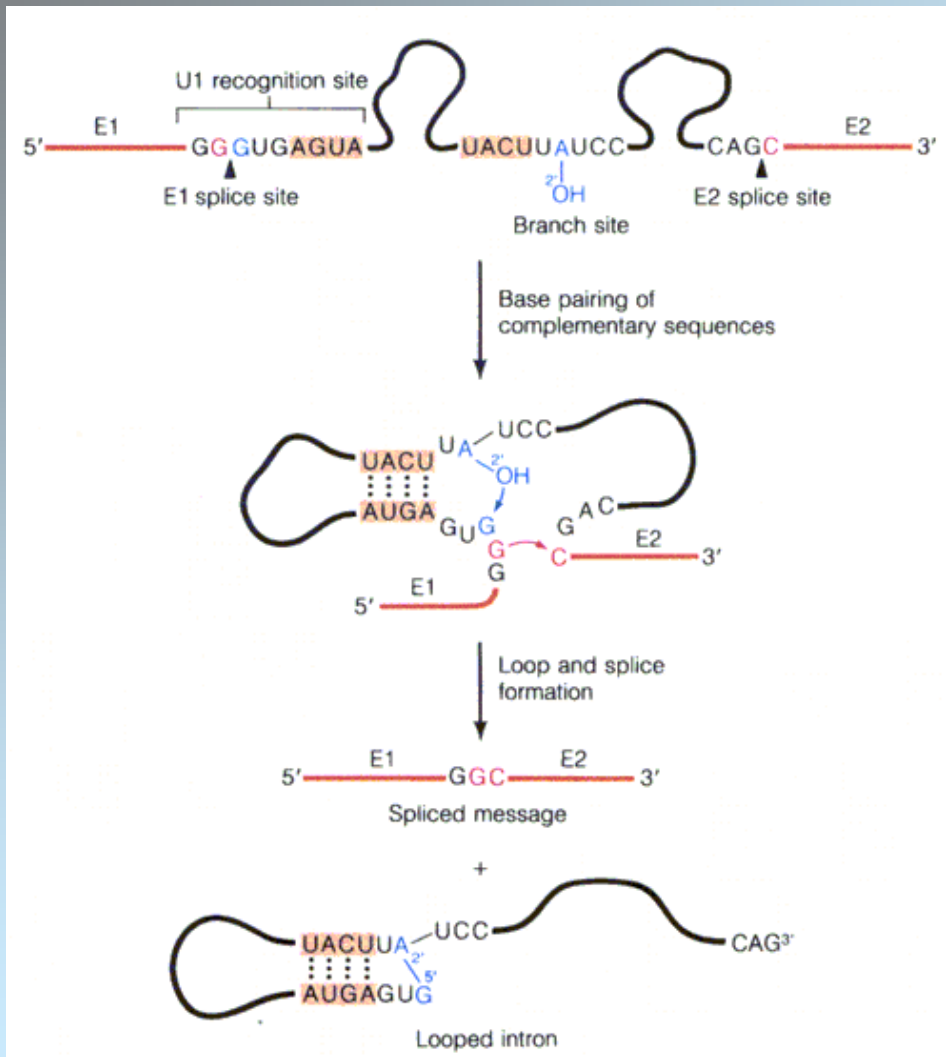
- Splicing modifies pre-mRNA released after transcription
- Non-coding sequences: introns (intragenic regions)
- coding sequences: exons are glued together
- A snRNA complex, the spliceosome, performs the splicing but some RNA sequences can perform autonomous splicing





Splicing, Exons and Introns

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

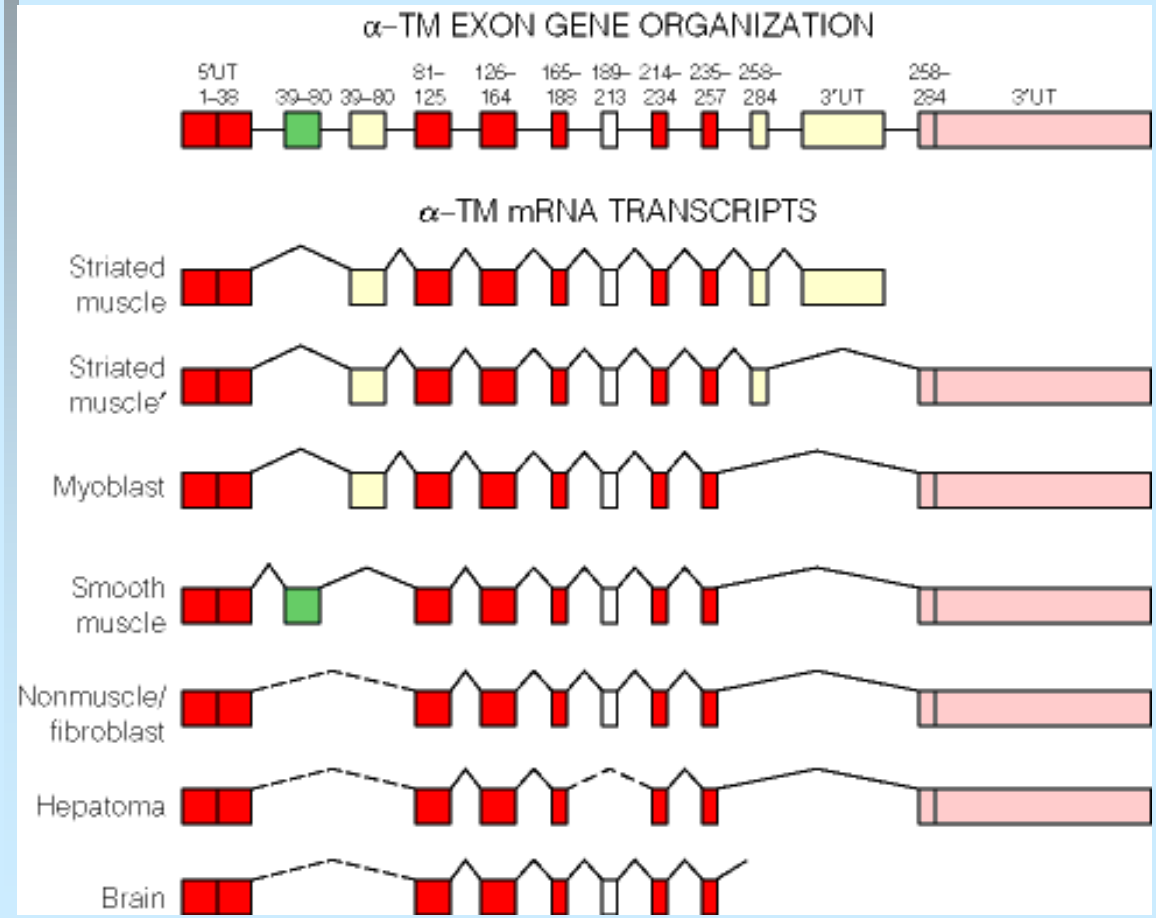


Self-splicing

Splicing, Exons and Introns

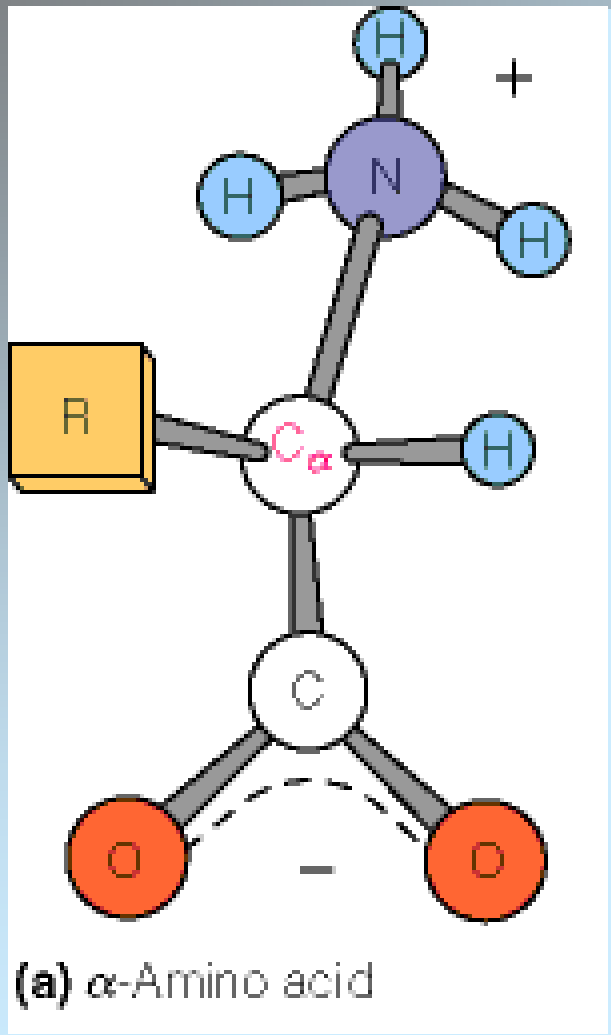
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

- pre-mRNA can be spliced in different ways: alternative splicing, therefore a gene can code different proteins
- Alternative splicing is controlled by signalling molecules

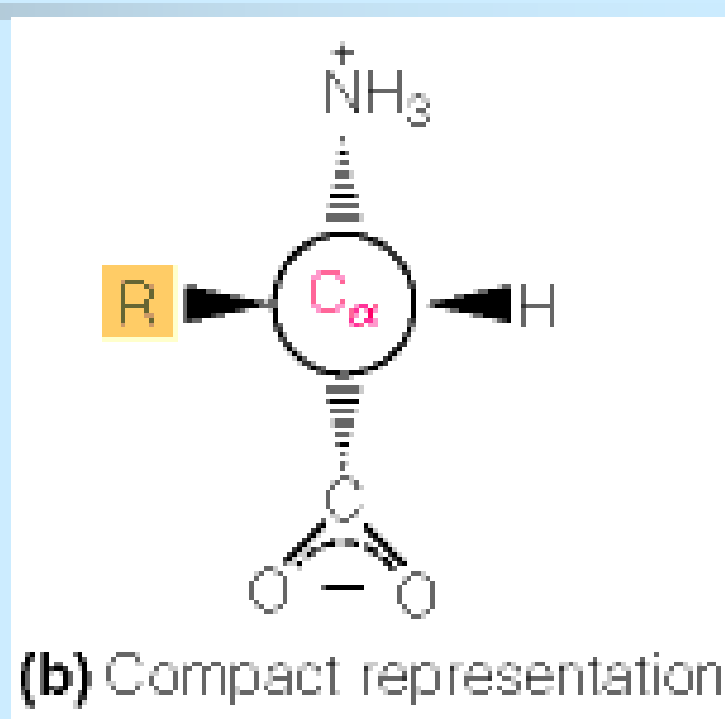


Amino Acids

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



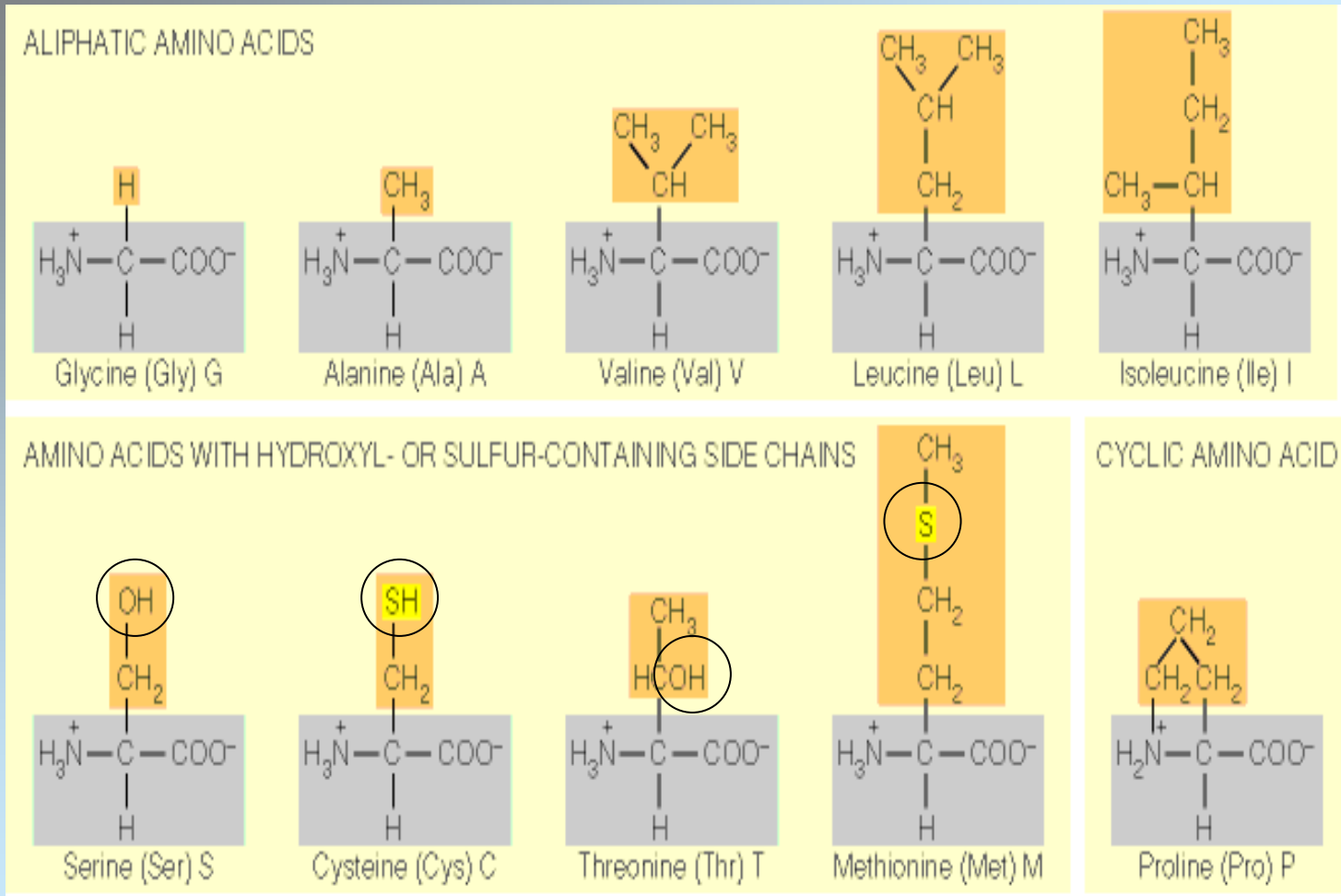
(a) α-Amino acid



(b) Compact representation

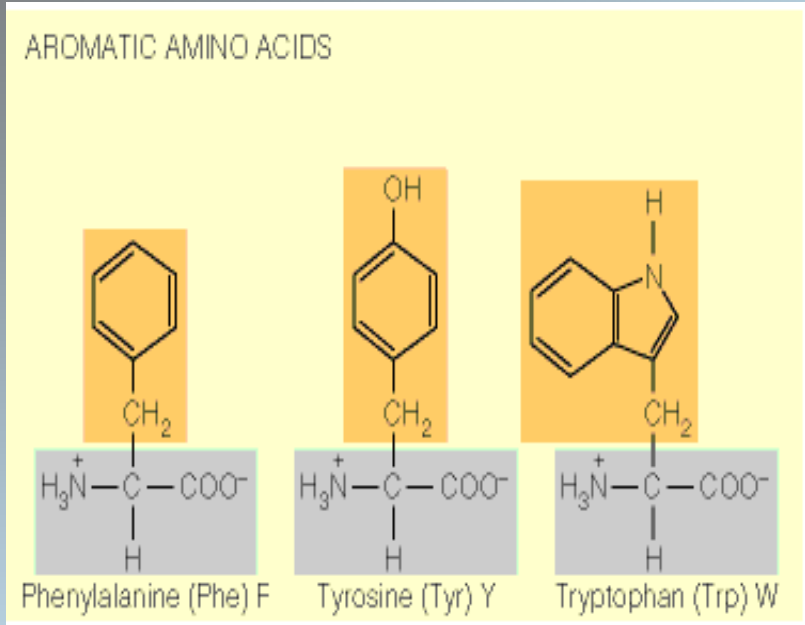
Amino Acids

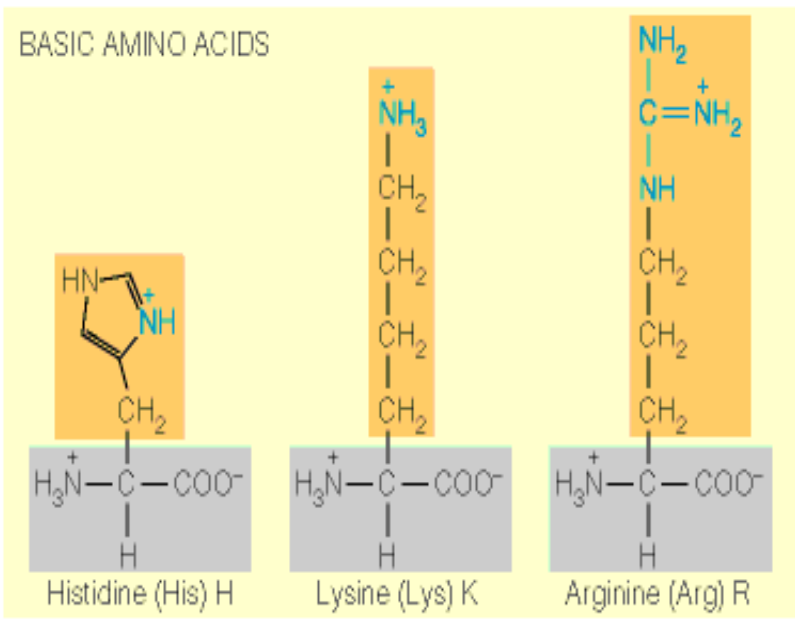
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

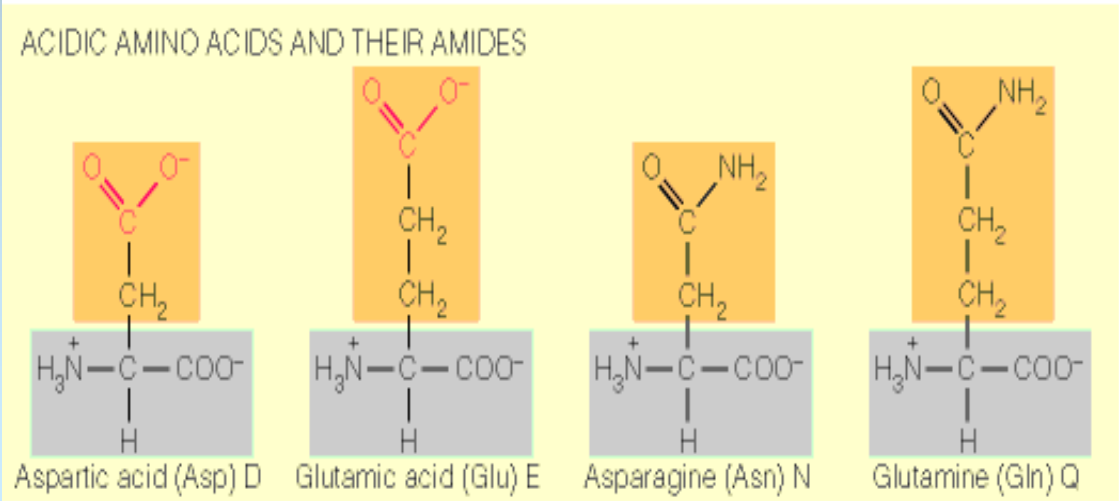


Amino Acids

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding







Amino Acids

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Hydrophobic (nonpolar):

glycine	Gly	G	methionine	Met	M
alanine	Ala	A	phenylalanine	Phe	F
valine	Val	V	tryptophan	Trp	W
leucine	Leu	L	proline	Pro	P
isoleucine	Ile	I			

Hydrophilic (polar)

serine	Ser	S	tyrosine	Tyr	Y
threonine	Thr	T	asparagine	Asn	N
cysteine	Cys	C	glutamine	Gln	Q

acidic (-,hydrophilic)

aspartic acid	Asp	D	glutamic acid	Glu	E

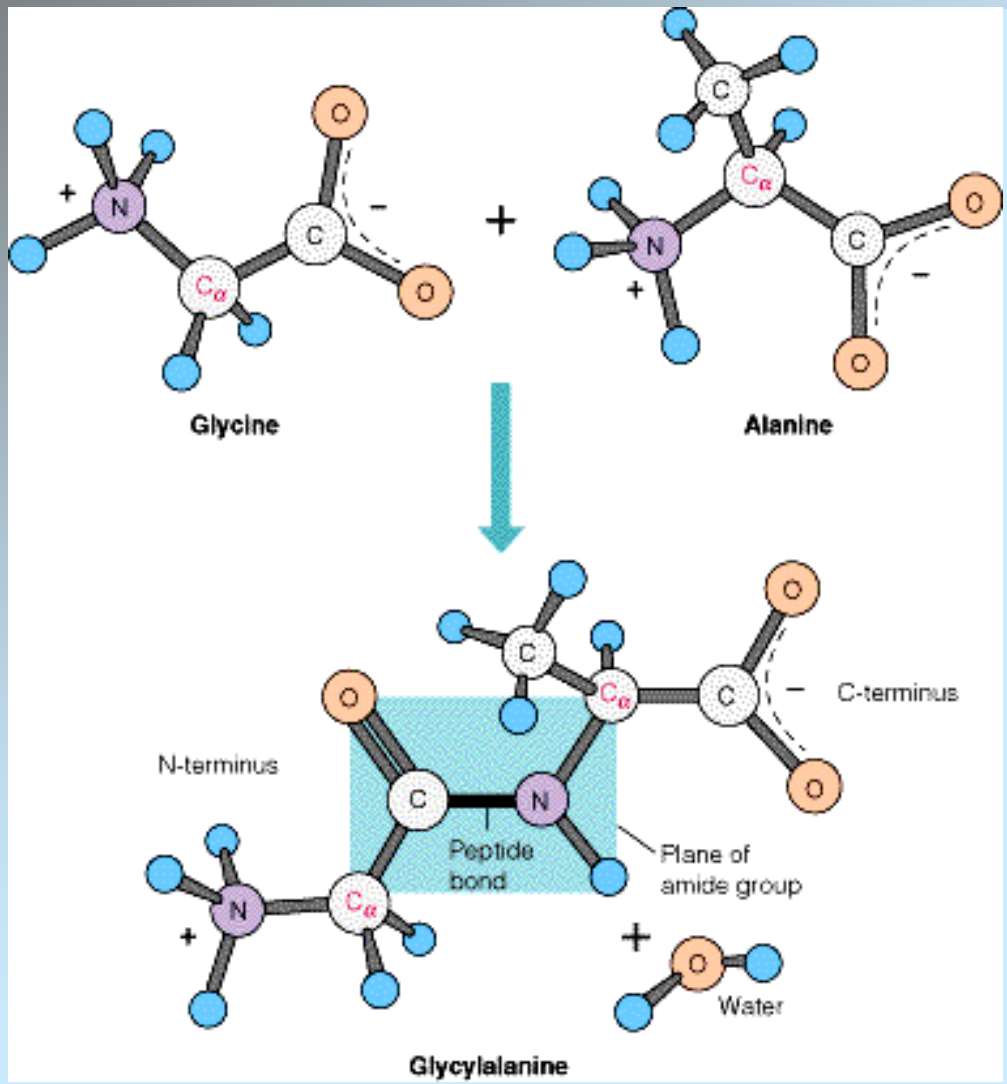
basic (+,hydrophilic)

lysine	Lys	K	arginine	Arg	R
histidine	His	H			

Cysteine and methionine: disulfide bonds

Amino Acids

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



Genetic Code

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

- all proteins consist of these 20 amino acids
- 3D interactions of the amino acids results in nano-machines
- genetic code: instructions for producing proteins from DNA
- protein in coded through a gene which is transcribed into mRNA and then translated into an amino acid sequence which automatically configures into a protein
- genetic code gives the rules for translation
- rules are simple: 3 nucleotides (codon) = one amino acid
  AUG and CUG: start codon

Genetic Code

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

	U		C		A		G		
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC		UCC		UAC		UGC		C
	UUA	leu	UCA		UAA	stop	UGA	stop	A
	UUG		UCG		UAG		UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lys	AGA	arg	A
	AUG	met	ACG		AAG		AGG		G
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	glu	GGA		A
	GUG		GCG		GAG		GGG		G

C = Cytosin, U = Uracil, A = Adenin, G = Guanin

Base pairs DNA: A-T and C-G (T = Thymin)

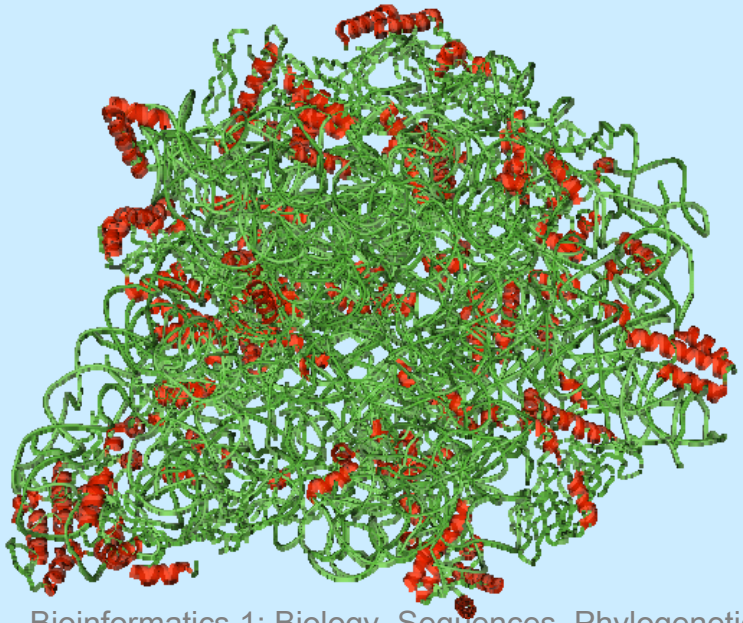
Translation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

After transcription the pre-mRNA is spliced, edited, transported out of the nucelus into the cytosol (eukaryotes)

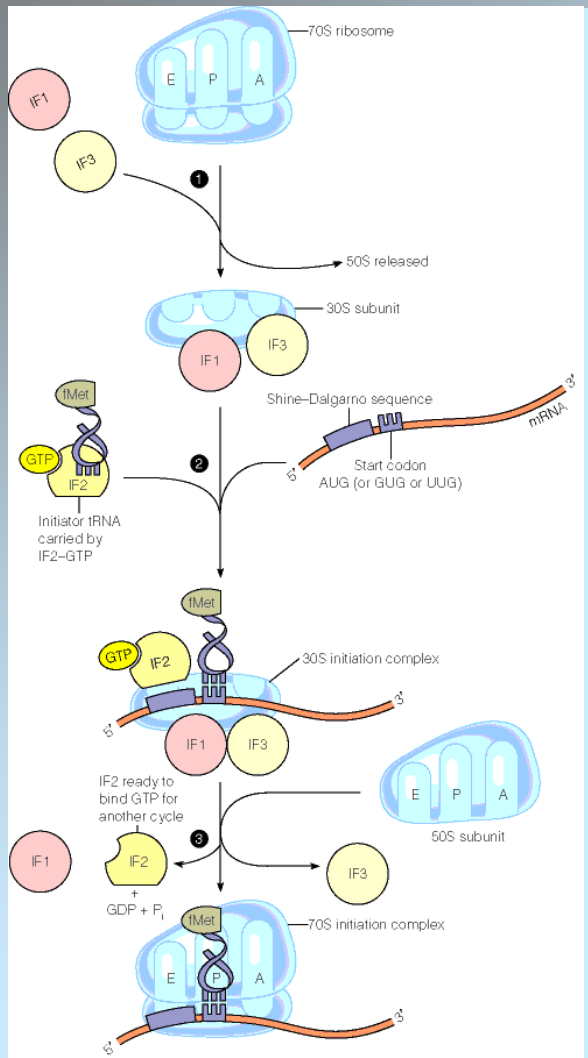
➥The ribosome (protein production machinery) assembles the amina acid sequences out of the mRNA

➥Ribosome consists of two subunits 60S and 40S in eukaryotes and 50S and 30S in bacteria



Translation Initiation

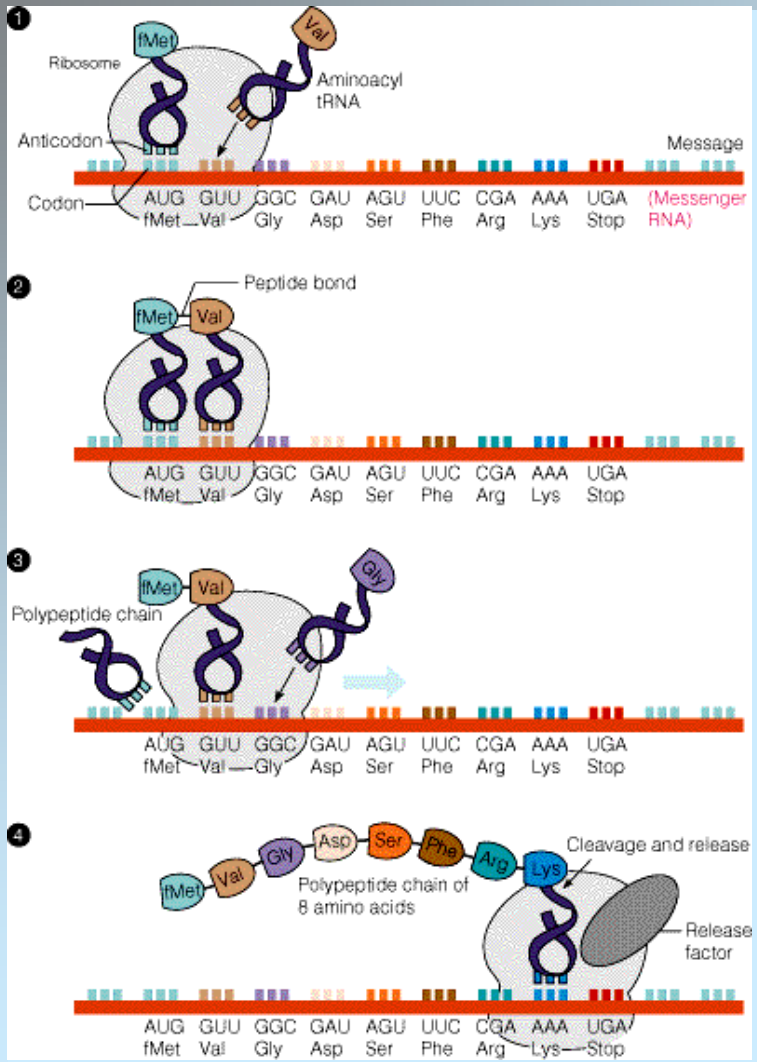
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

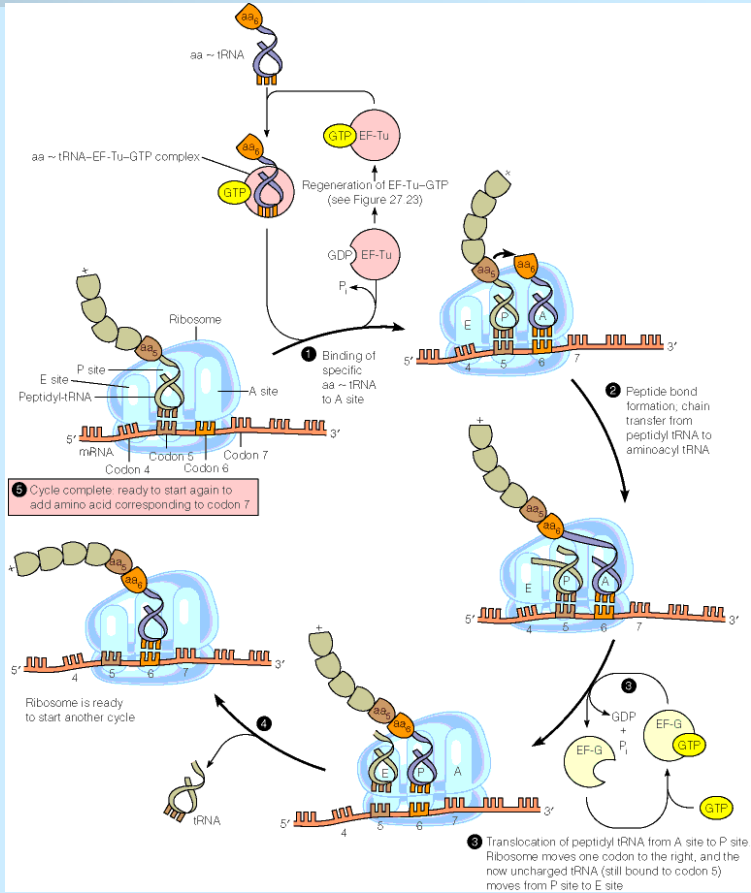


- Inactive ribosomes have dissociated subunits
- Ribosome binds to site at mRNA marked by AGGAGGU (Shine-Dalgarno)
- At this site the initiation factors IF1, IF2, IF3 and 30S ribosomal subunit bind
- The initiator tRNA binds to the start codon
- Then the 50S subunit binds to the complex and translation can start

Translation Elongation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding





Translation Elongation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Ribosome

Translation Termination

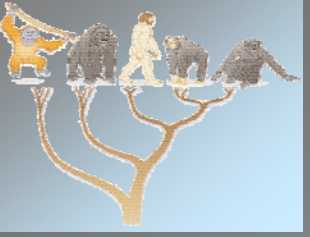
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

➥Termination by a stop codon (UAA, UAG, UGA) which enters the A-site

➥ tRNAs cannot bind, however release factors bind at or near

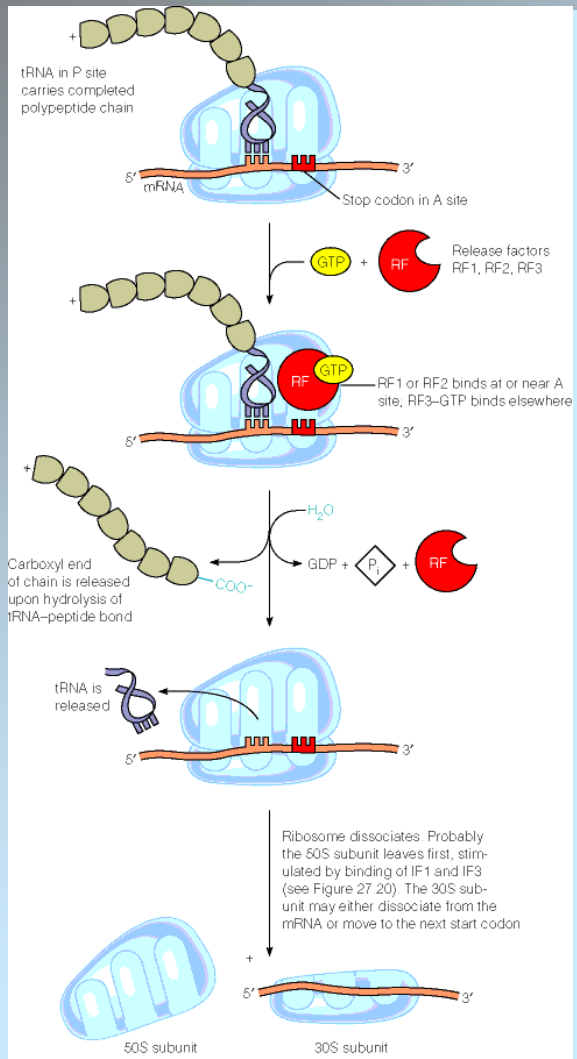
➥amino acid chain is released and the 70S ribosome dissociates

➥30S subunit remains attached to the mRNA and searching for the next Shine-Dalgarno pattern



Translation Termination

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding



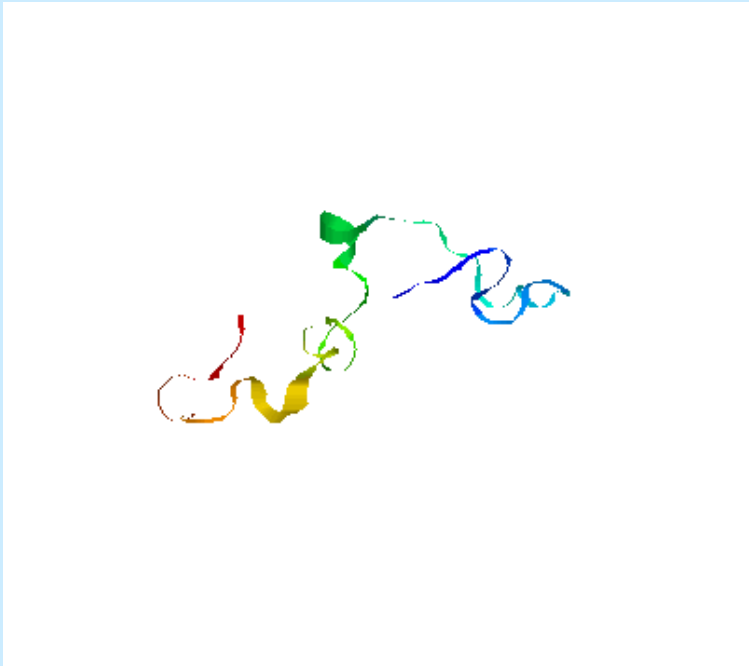
Folding of the Protein

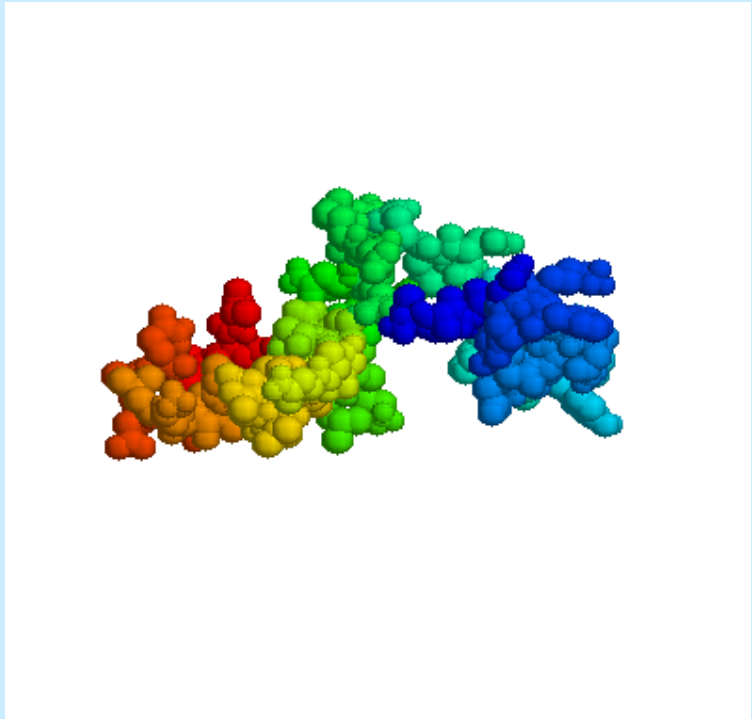
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

- Only the correct folded protein functions correctly (cf Creutzfeld-Jacob, Altzheimer, BSE, Parkinson)
- proteins always fold into their specific 3D structure
- complicated procedure with lots of interactions
- folding pathways are not unique and have intermediate states
- folding is assisted by special chaperones (hide the hydrophobic regions or act as containers)
- Folding time: milliseconds up to minutes or hours
- major tasks in bioinformatics is the prediction of the 3D structure to guess the function or to design new proteins

Folding of the Protein

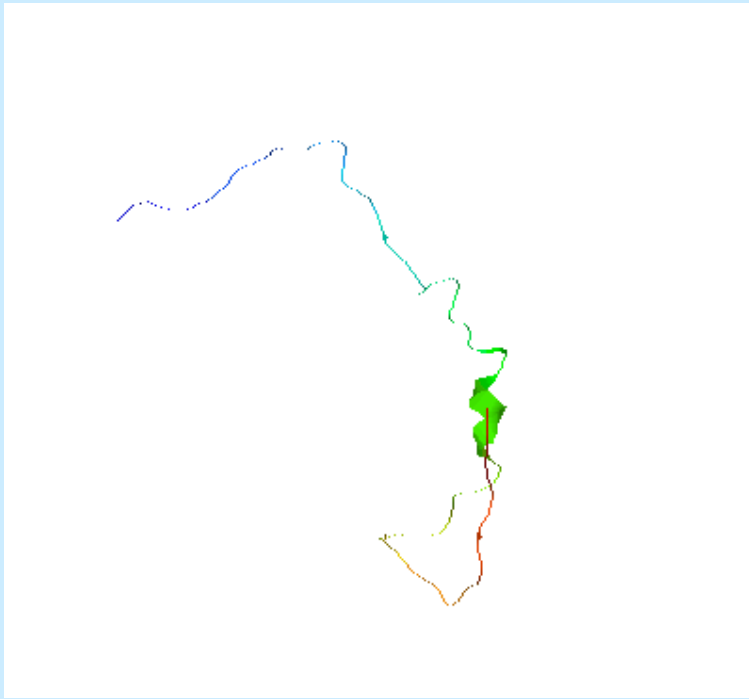
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

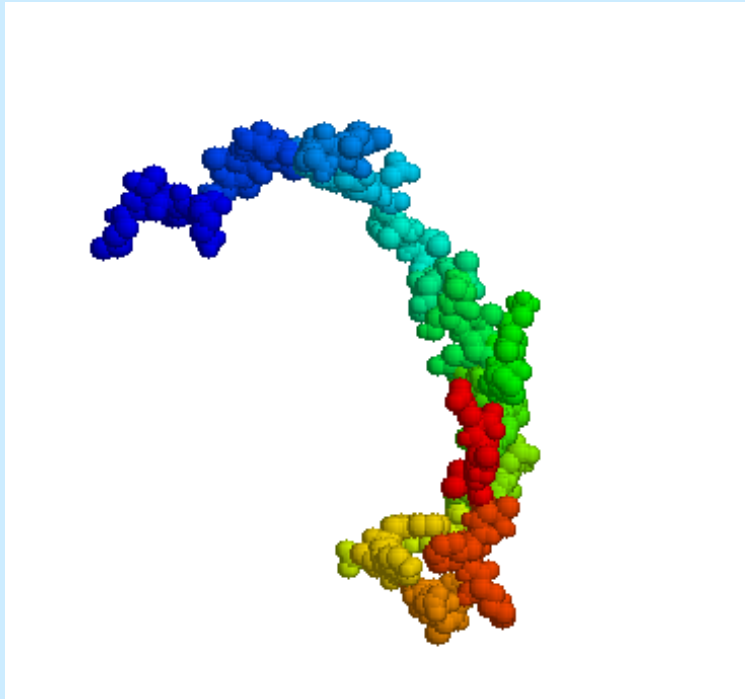


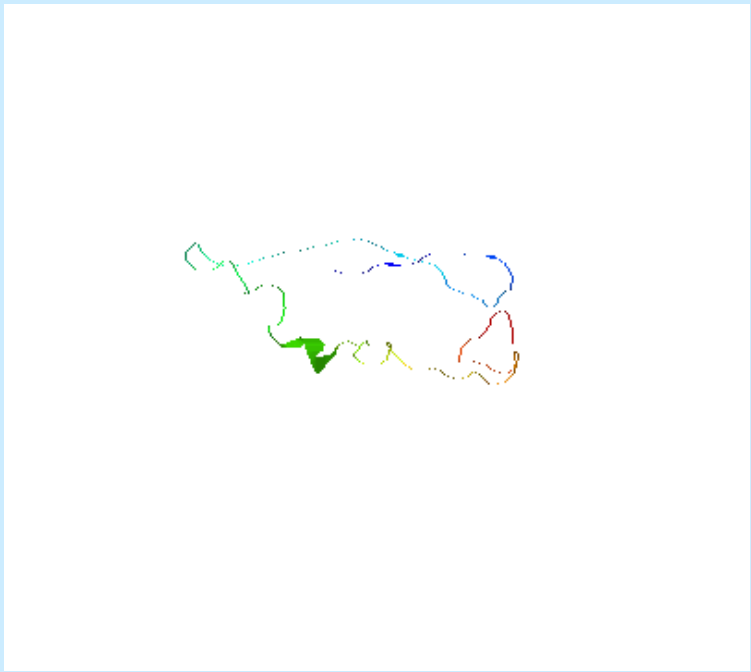
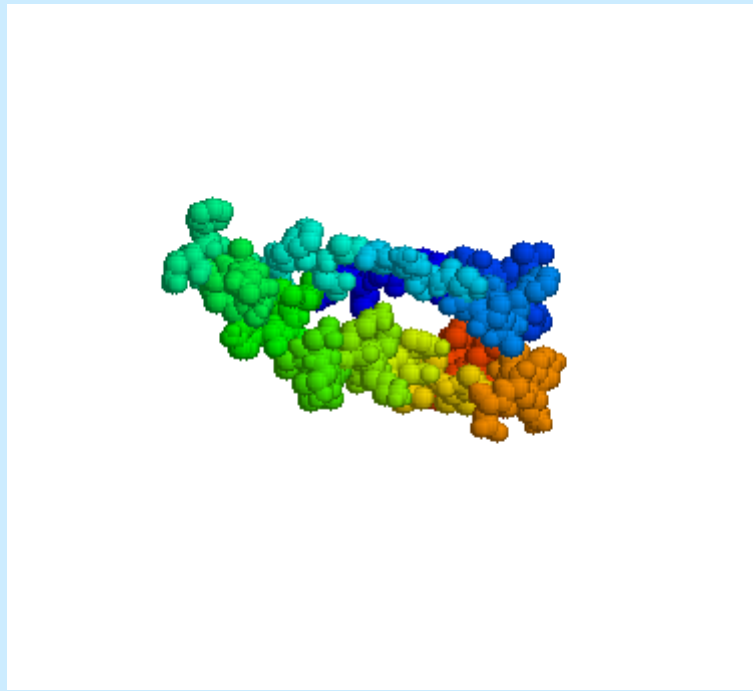


Folding of the Protein

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding





Folding of the Protein

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding