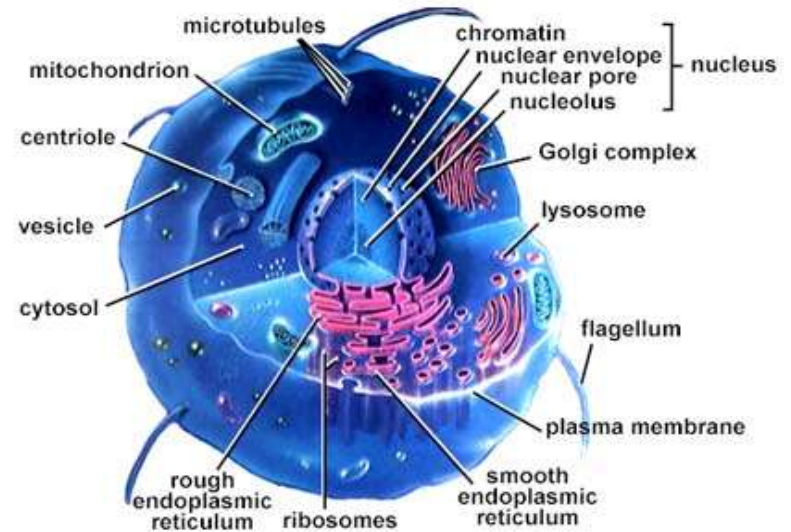


Bioinformatika

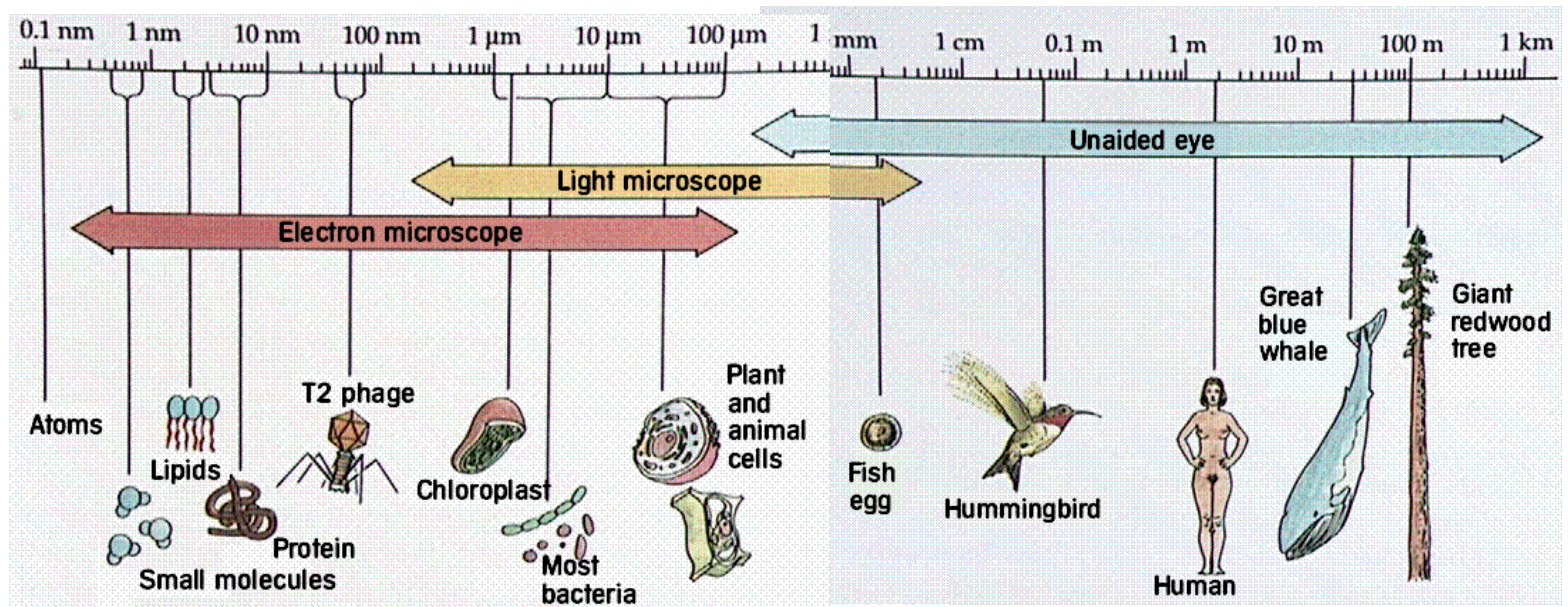
- ✦ zpracování hromadných biologických dat
- ✦ kořeny v 50. a 60. letech – struktura DNA, proteinu, genetický kód
- ✦ data = sekvence, struktury, exprese genu, metabolismus,...
- ✦ analýzy = zarovnávání sekvencí, fylogenetické stromy, identifikace důležitých strukturních a funkčních motivů, předpovězení struktury,...
- ✦ DNA - RNA - PROTEIN ::: PROTEIN - STRUKTURA – FUNKCE
- ✦ rozsah dat u člověka:
 - ✦ 3 miliardy bází DNA (x2) / buňku
 - ✦ 30000 různých genu/proteinů
 - ✦ 100 miliard buněk / jedince

Bunky

- Zakladni forma organizace zive hmoty
- Molekuly/geny/proteiny
- Proteinove komplexy/membrány
- Organely a jine substruktury
- **Bunka**
- Tkan/pletivo
- Organismy



Velikosti elementu



Zaklady genetiky

- Mendel 19.stol.



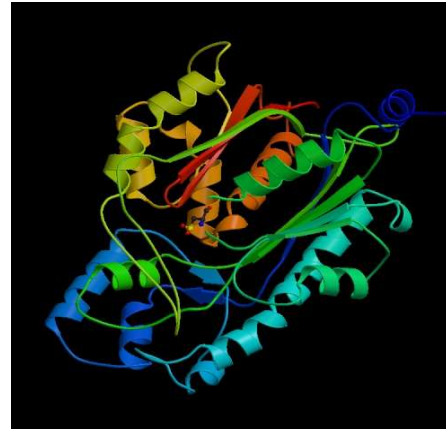
V pokusech s rostlinami si vsiml, ze potomstvi dvou rodicu nezavisi na jejich vzhledu /**fenotyp**/, nybrz na jakychsi symbolicky popsanych faktorech, ktere do značne miry odpovidaji dnesnimu oznaceni gen /**genotyp**/

GENOTYP

FENOTYP

```
>chs1  
atgacagaat  
acaggatgac  
tatgacgtga  
cggcttatat  
gatgacc...
```

```
>chs1  
MFVDDHLA  
VNQNFYLR  
SHRQL...
```



GEN.KOD

STRUKTURA

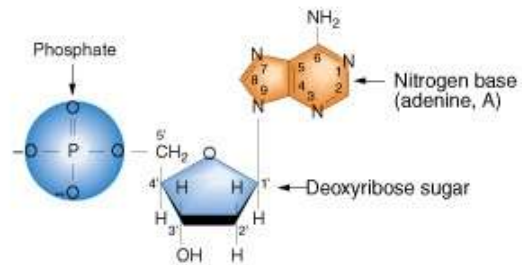
FUNKCE

Struktura DNA

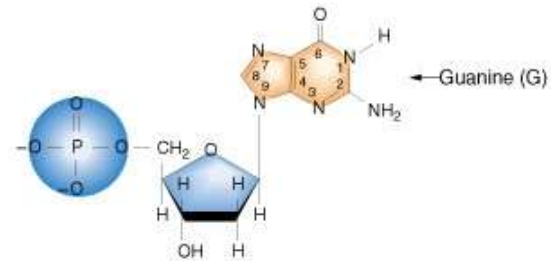
- Objevena v 19.stol
- Považována za jednoduchou
- Polymerická molekula
- Monomerem je
 - Ribonukleotid /RNA/
 - deoxyribonukleotid /DNA/
- Symbolické znacení A,C,G,T,U
- Prostorová struktura: Watson, Crick, Franklin

Struktura DNA

Purine nucleotides

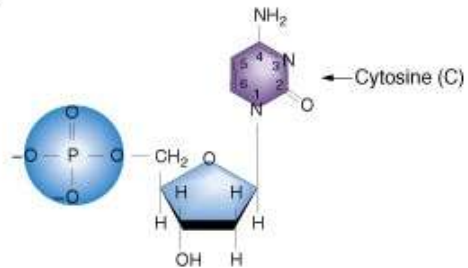


Deoxyadenosine 5'-phosphate (dAMP)

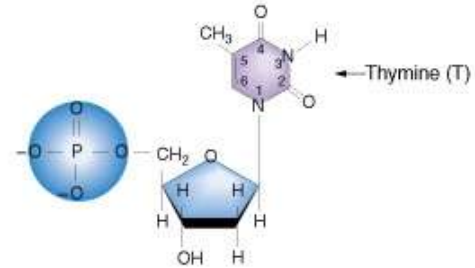


Deoxyguanosine 5'-phosphate (dGMP)

Pyrimidine nucleotides



Deoxycytidine 5'-phosphate (dCMP)



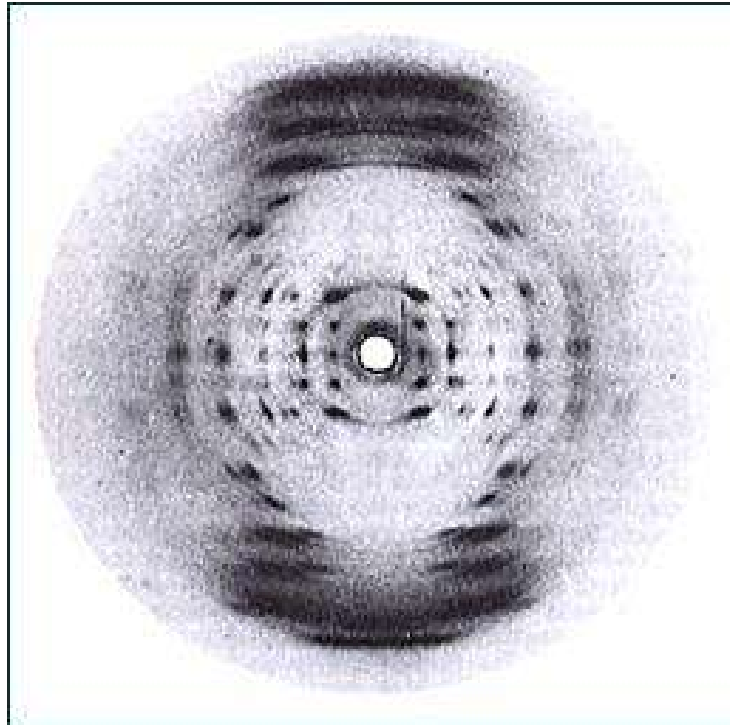
Deoxythymidine 5'-phosphate (dTMP)

Struktura DNA

Organism	%A	%T	%G	%C	$\frac{A+G}{T+C}$	$\frac{A+T}{G+C}$
<i>E. coli</i> bacterium	26.0	23.9	24.9	25.2	1.04	1.00
<i>S. cerevisiae</i> yeast	31.7	32.6	18.3	17.4	1.00	1.80
<i>Z. mays</i> corn	25.6	25.3	24.5	24.6	1.00	1.04
<i>D. melanogaster</i> fly	30.7	29.4	19.6	20.2	1.01	1.51
<i>H. sapiens</i> human	30.2	30.3	19.9	19.6	1.01	1.53

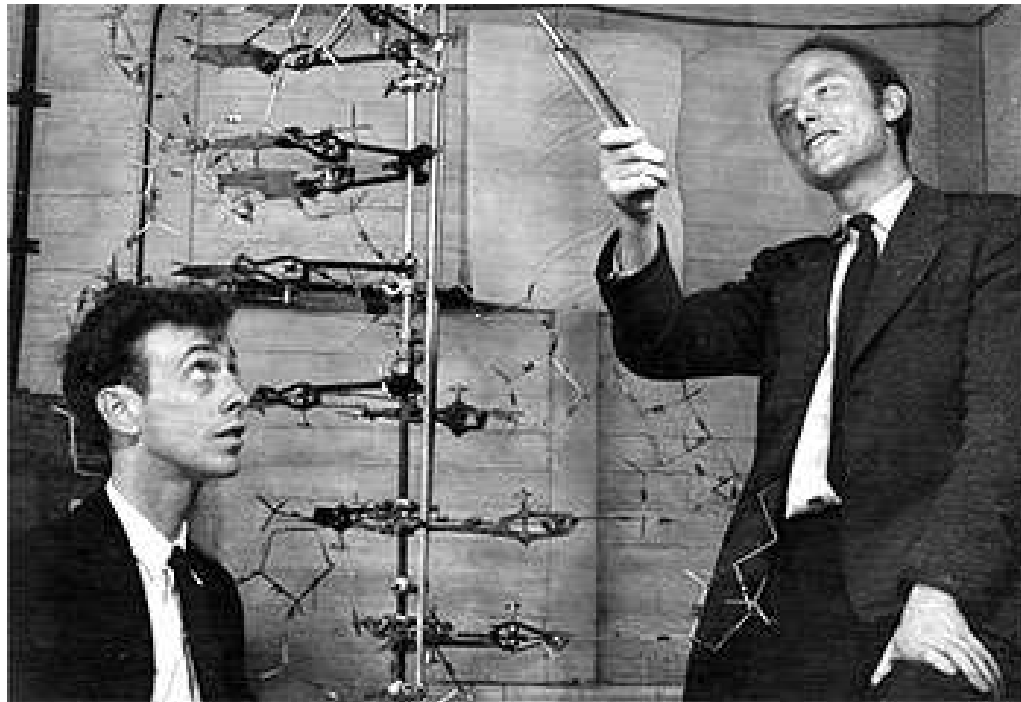
Obsah bazi v DNA ruznych organismu

Struktura DNA



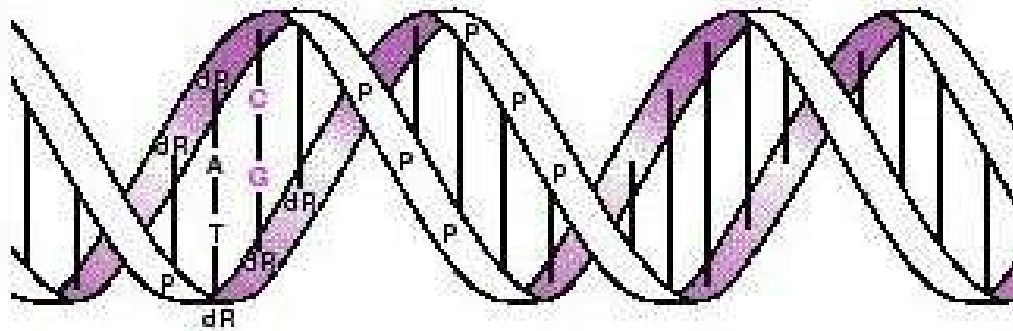
Difrakce krystalické DNA z roku 1952 od Rosalind Franklin
Svedci o periodicite 0.3nm a 3.4nm

Struktura DNA



Struktura DNA

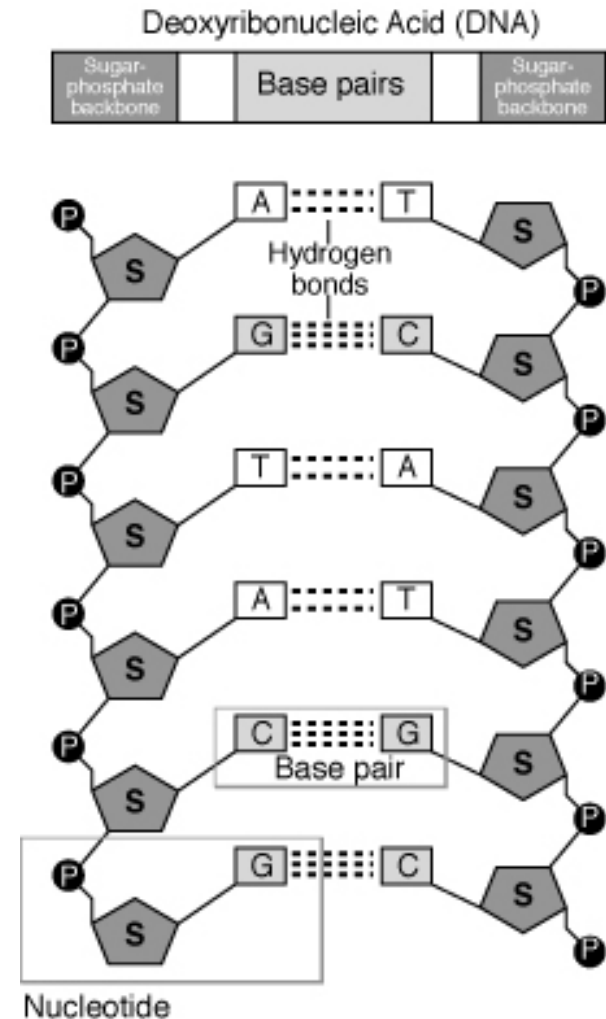
4.10 DOUBLE HELIX



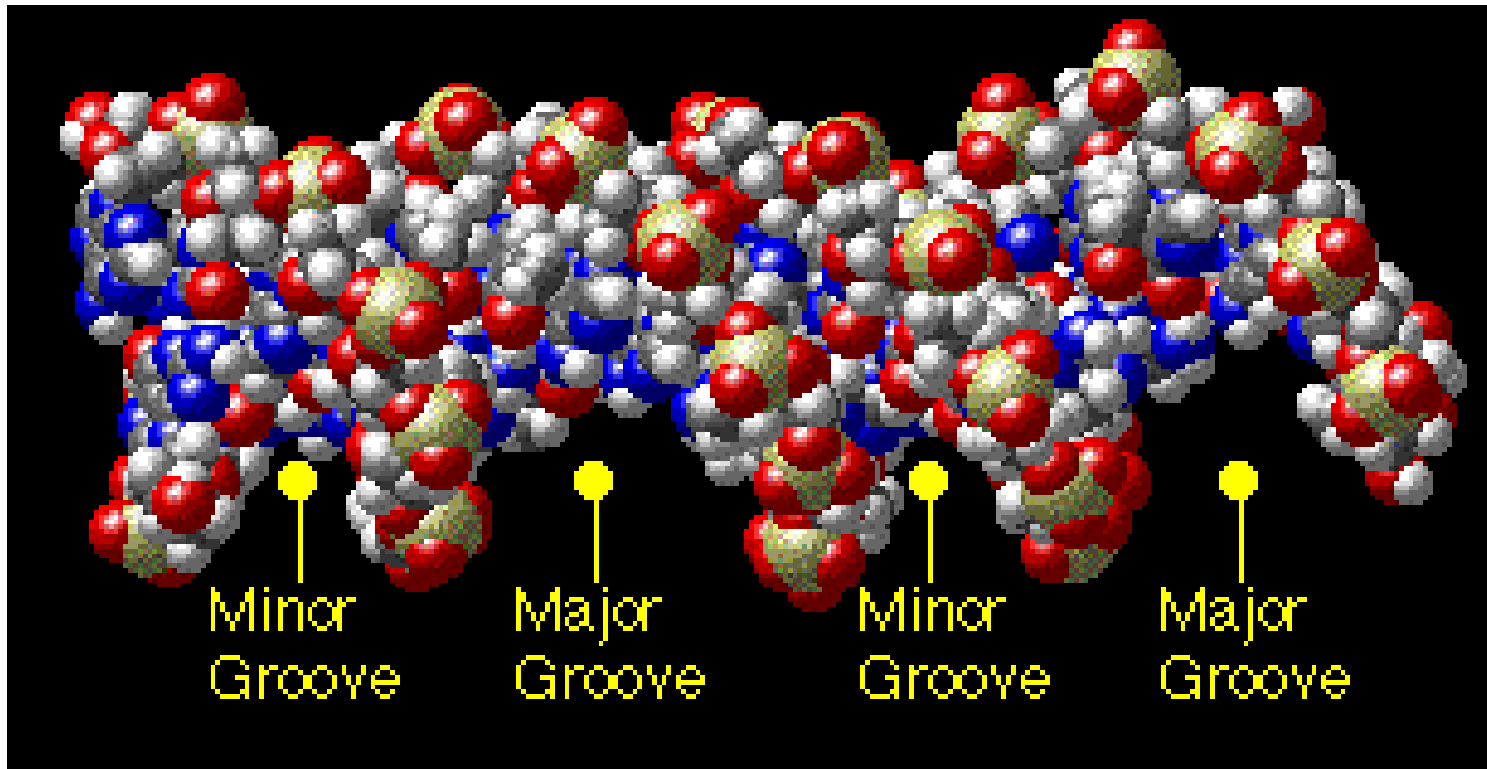
P = phosphate, dR = deoxyribose, A, T, C, G = base pairs

Sekvence DNA

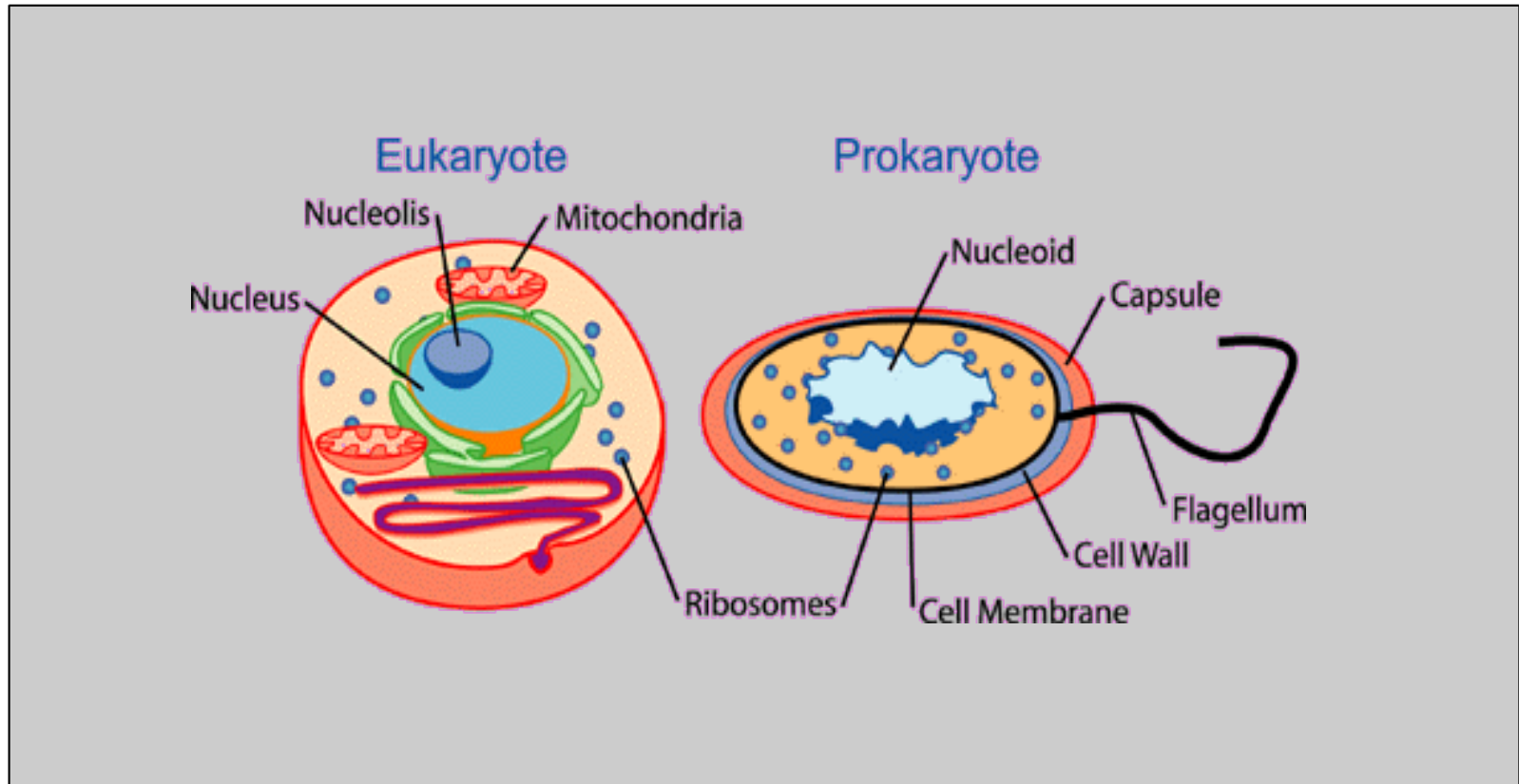
Vazby mezi vlaknem jsou vodíkové můstky, které jsou ve srovnání s kovalentními vazbami slabé. Působením teploty do 100°C může dojít k denaturaci, t.j. rozpojení vláken. Využívá se to v PCR, ale vlastně také v replikaci, transkripci.



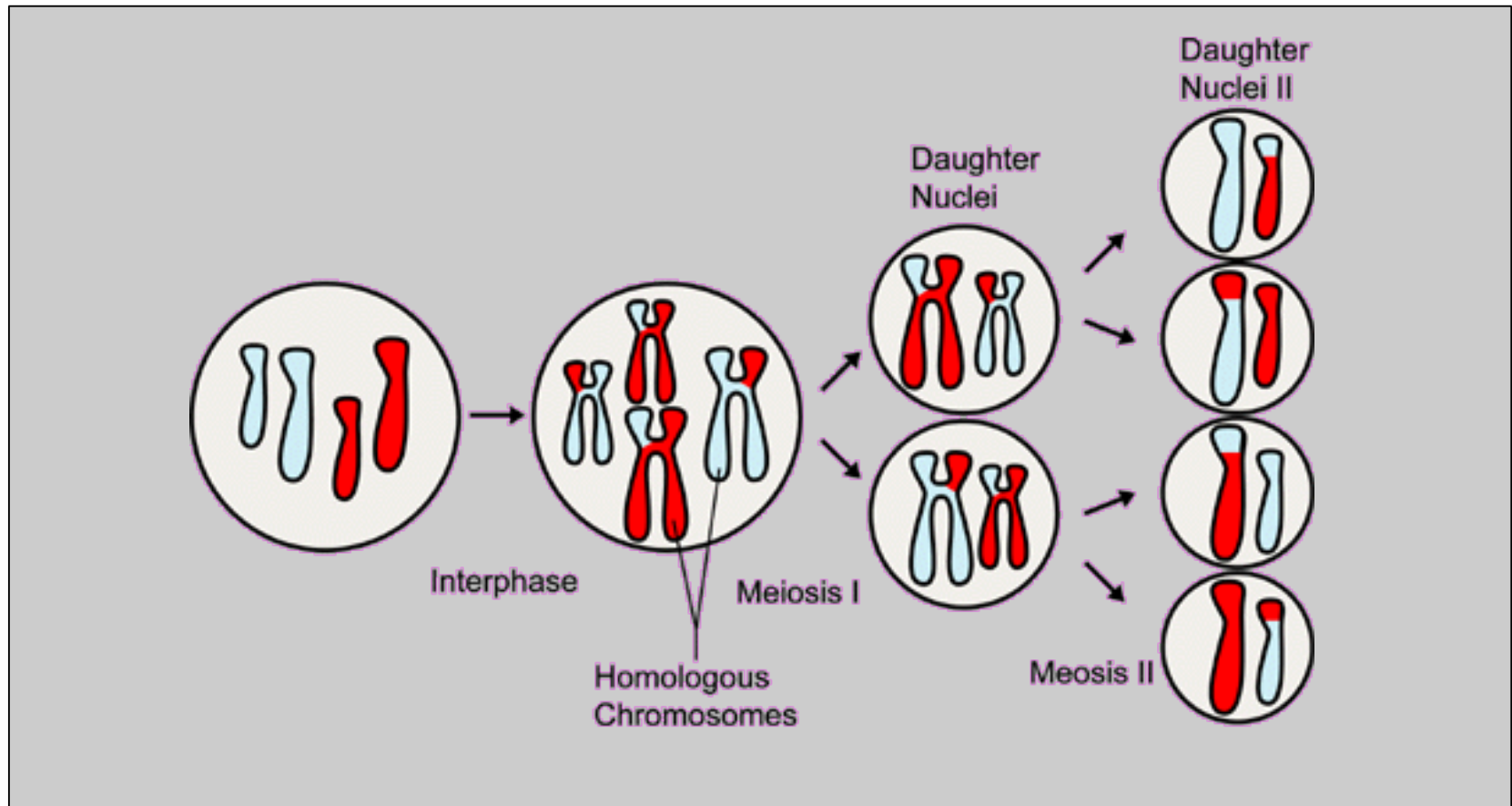
Sekvence DNA



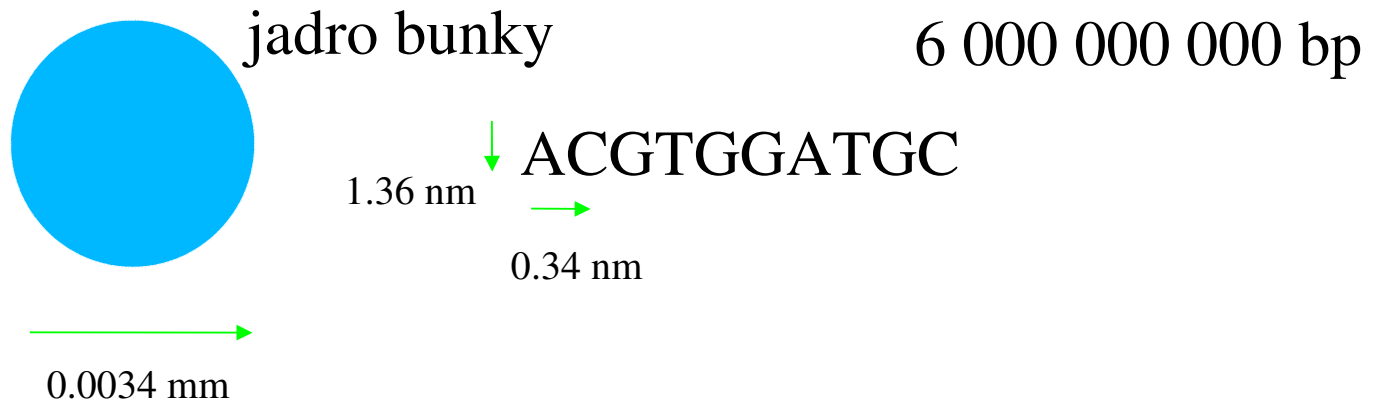
Zaklady genetiky



Zaklady genetiky



Organizace DNA



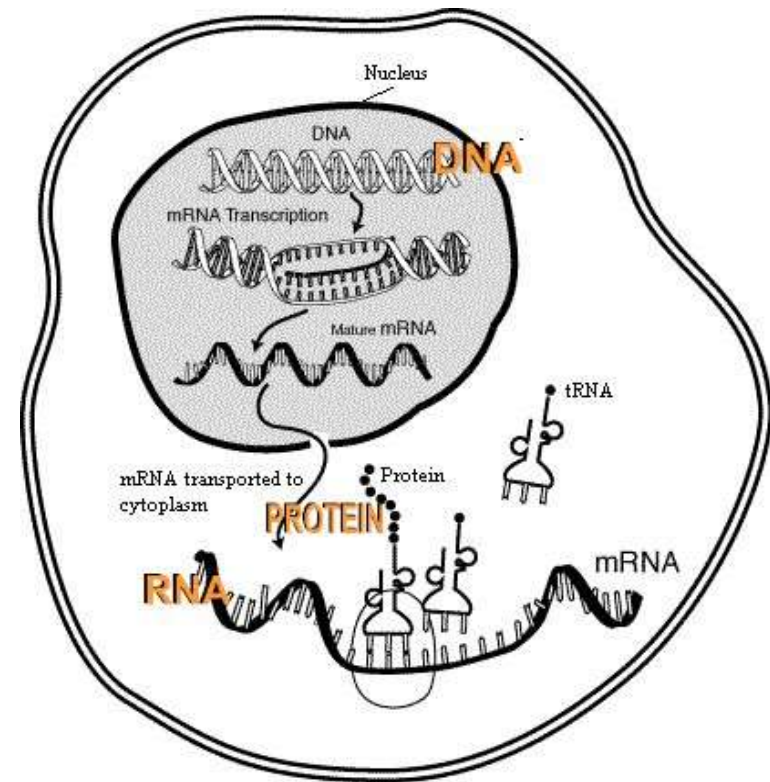
LINEARNE	10000 bp	600 000 x
PLOSNE	25 000 000	240 x
PROSTOROVE	62 500 000 000	~0.1 x

Jeden gen – jeden protein

- Muže za to mechanismus exprese genu/syntezy proteinu
- DNA -> RNA -> PROTEIN
- Pravidlo ktere jak dnes vime nemuze byt chapano doslovne:
 - Alternativni „splicing“
 - Post-translacni modifikace
 - Zadny protein
 - Retroviry

Centralni dogma

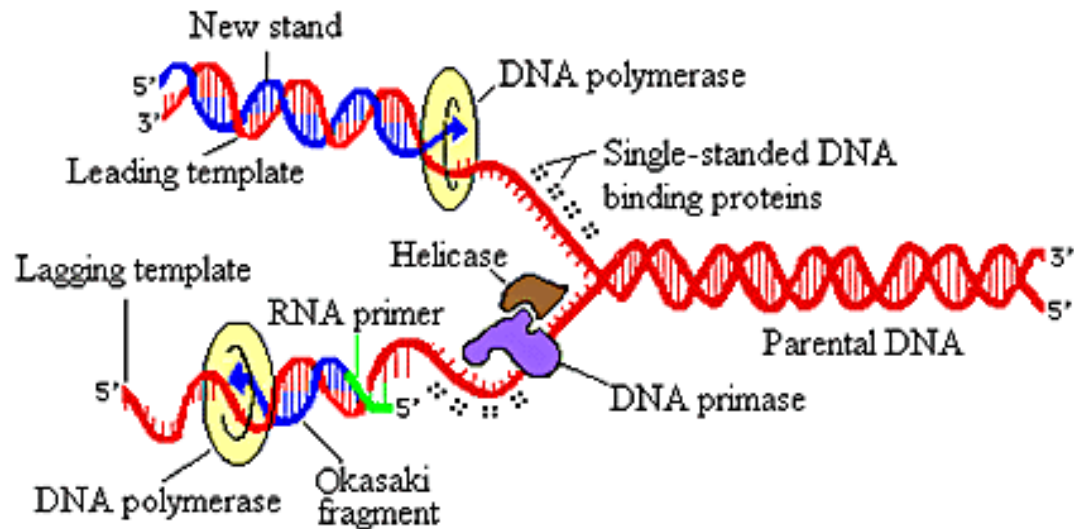
- DNA → RNA → PROTEIN



Replikace DNA

- Hlavní enzym: DNA polymeraza
- Templat: DNA
- Substrat: deoxyribonukleotidy
- Produkt: DNA

Replikace DNA



**Collaboration of Proteins
at the Replication Fork**

Transkripce

- Hlavní enzym: RNA polymeraza
- Transkripční komplex: celá škála proteinů vázaných na specifické sekvence DNA a sebe navzájem
- Templat: DNA
- Substrat: ribonukleotidy
- Produkt: RNA

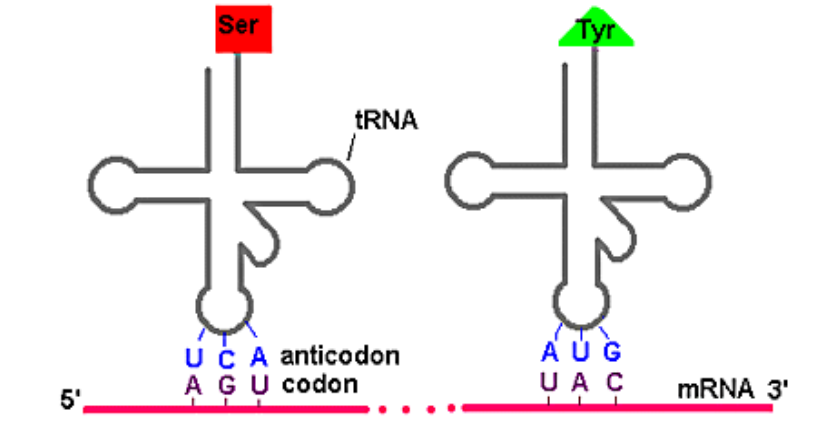
Struktura proteinu

- Polymerická molekula
- Monomerem je aminokyselina
- Symbolické znacení ACEFGHIKLMNPQRSTVWY
- Prostorová struktura se určuje z difrakce röntgenových paprsků nebo magnetickou rezonancí /NMR/

Genetický kod

- Rozlustili Nirenberg, Matthaei, Khorana, Holley
- Život je digitalní
- 3 nukleotidy kodují jednu aminokyselinu v proteinovém řetězci
- $4^3 = 64$, ale jenom 20 aminokyselin

Geneticky kod



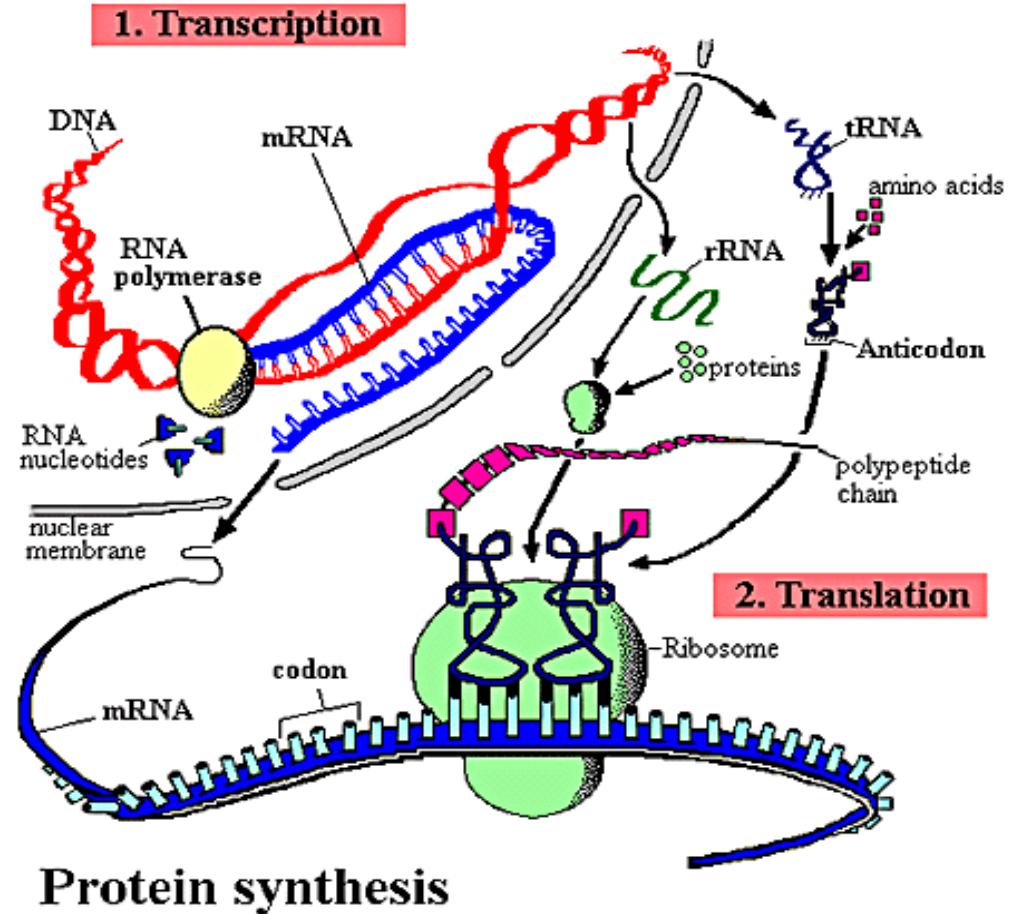
		2nd base in codon					
		U	C	A	G		
1st base in codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G	3rd base in codon
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G	
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G	
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

The Genetic Code

Translace

- Ribozom: cela skala RNA a proteinu vazanych na specificke sekvence RNA a sebe navzajem. t-RNA poskytuje molekularni pojitko mezi kodonem a odpovidajici aminokyselinou
- Templat: RNA
- Substrat: t-RNA s navazanou aminokyselinou
- Produkt: peptid nebo protein, volna t-RNA

Transkripce Translace



Kontrolni otazky

- ✦ DNA urciteho organizmu obsahuje 17% thyminu /T/. Jaky je obsah cytozinu /C/?
- ✦ Jaka je komplementarni sekvence k sekvenci 5'-ACGT-3'? Co je na techto sekvencich zajimave? Jake vlastnosti by mohl mit protein, ktery se vaze na DNA s takovou sekvenci?

Souhrn

- ☀ Centralni dogma =
- ☀ DNA – molekula nachazejici se v bunek, nese geneticke informace v podobe sekvence /bazi/
- ☀ RNA – do formy ...RNA se transkripci prepisuje sekvence konkretnich genu, aby mohla opustit jadro, forma ...RNA napomaha rozpoznavani v procesu translace
- ☀ PROTEIN – polymericka, sekvence usporadanych v prostoru tak, ze molekula plni specifickou funkci v bunce /napr. enzym/
- ☀ LIDSKY GENOM – ... miliardy bazi ve 22+X+Y, asi 30000 ruznych genu
- ☀ Slozite ORGANIZMY jsoubunecne, kazda bunka plni specifickou funkci danou kombinaci informace a, ve kterem se nachazi
- ☀ Vetsina PROCESU v bunkach je kontrolovanych, nekdy vhodne zabudovanych za pomoci membran do bunecnych organel a jinych struktur

Souhrn

- ☀ Centralni dogma = DNA – RNA -PROTEIN
- ☀ DNA – polymericka molekula nachazejici se v jadre bunek, nese geneticke informace v podobe sekvence nukleotidu /bazi/
- ☀ RNA – do formy mRNA se transkripci prepisuje sekvence konkretnich genu, aby mohla opustit jadro, forma tRNA napomaha rozpoznavani kodonu v procesu translace
- ☀ PROTEIN – polymericka molekula , sekvenceaminokyselin usporadanych v prostoru tak, ze molekula plni specifickou funkci v bunce /napr. enzym/
- ☀ LIDSKY GENOM – 3 miliardy bazi ve 22+X+Y chromozomech, asi 30000 ruznych genu
- ☀ Slozite ORGANIZMY jsou mnohobunecne, kazda bunka plni specifickou funkci danou kombinaci geneticke informace a prostredi, ve kterem se nachazi
- ☀ Vetsina PROCESU v bunkach je kontrolovanych proteiny, nekdy vhodne zabudovanych za pomoci membran do bunecnych organel a jinych struktur

Struktura proteinu

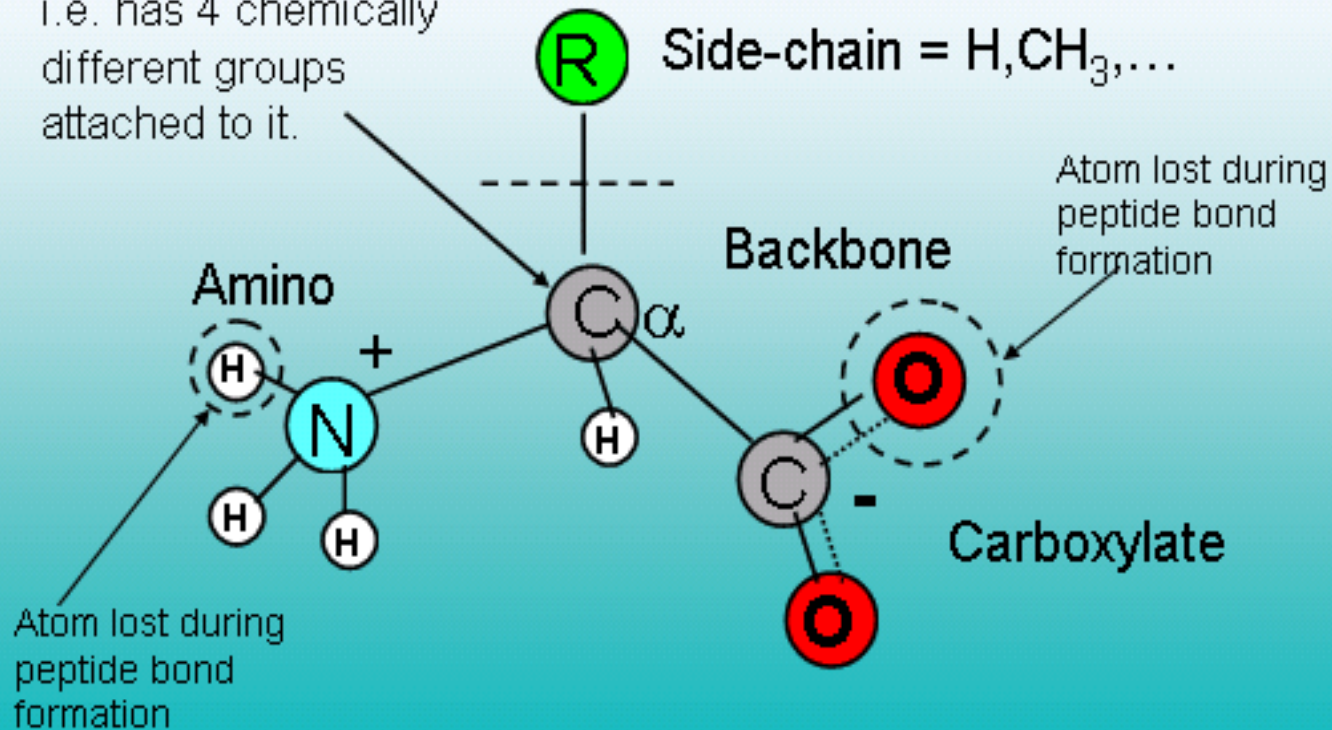
- ✦ Primární = sekvence
- ✦ Sekundární = původně jenom α -helix a beta struktury, dnes i jiné substruktury
- ✦ Terciární = 3-D uspořádání
- ✦ Domény – jedna nebo víc na protein
- ✦ Komplexy – skladající se z více podjednotek

How does a protein fold?

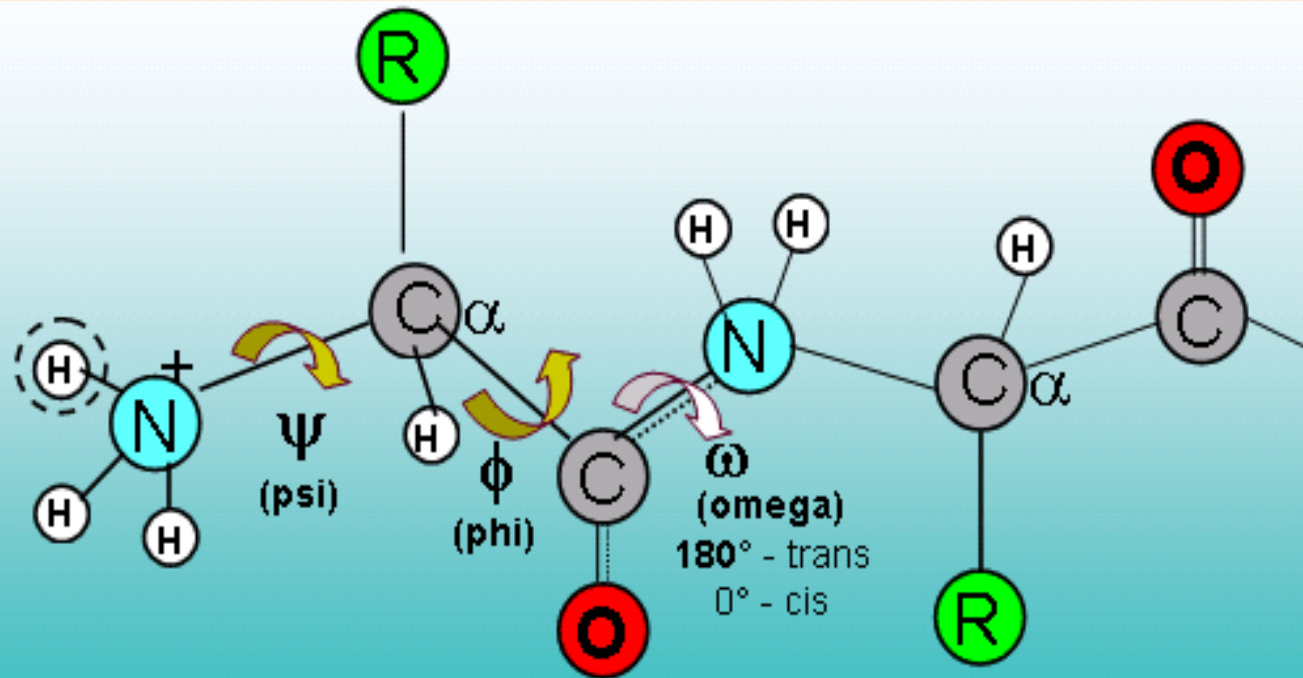
- Most newly synthesized proteins fold without assistance!
 - *Ribonuclease A: denatured protein could refold and recover its activity (C. Anfinsen -1966)*
 - “Structure implies function”
 - *The amino acid sequence encodes the protein’s structural information*

The (L-)amino acid

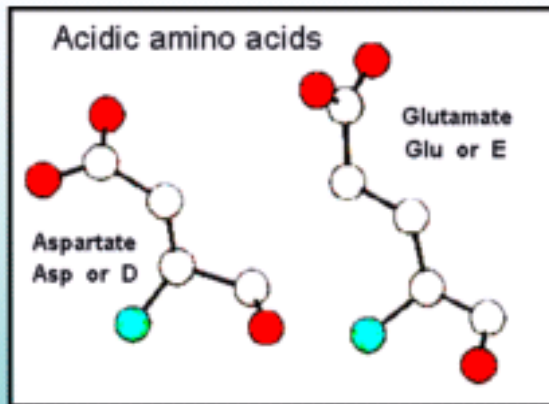
C_{α} is a chiral centre:
i.e. has 4 chemically
different groups
attached to it.



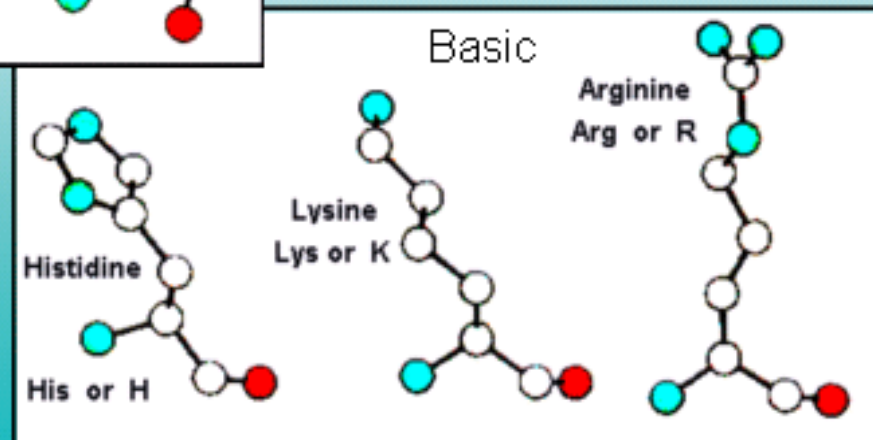
Backbone Torsion Angles



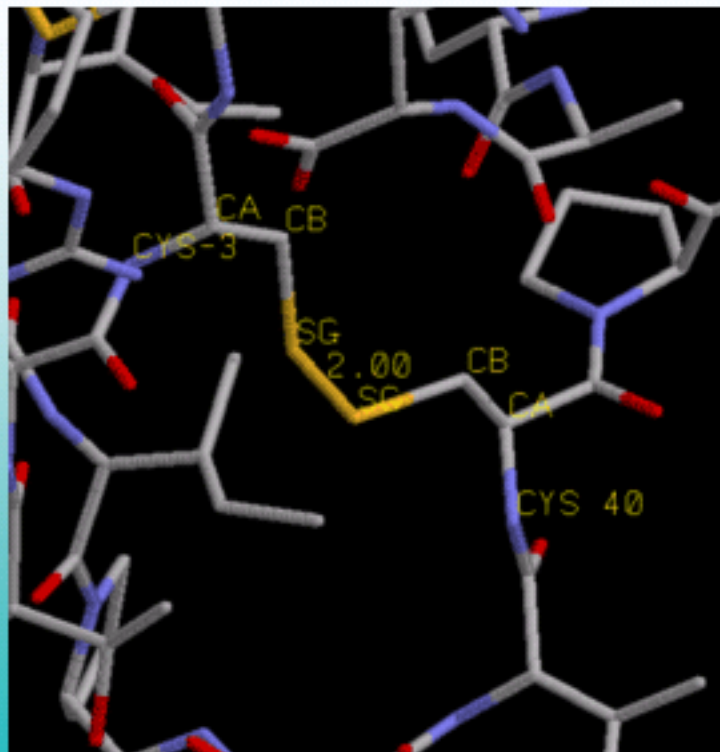
Charged residues



These contain side-chains that are charged under physiological conditions, i.e. pH 7.0: acidic – negative charge and basic – positive charge



The disulfide bond



- = “disulfide bridge”
- Only in extracellular proteins
- Formed by oxidation of the SH (thiol) group of cysteine residues
- Covalent bond between the S γ (or ‘SG’) atoms of two cysteine residues

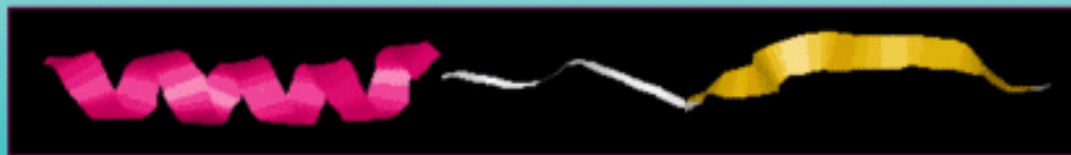
Levels of protein structure: 0,1

- Zeroth: amino acid composition – no structural information
- Primary
 - This is simply the order of covalent linkages along the polypeptide chain, i.e. the sequence itself

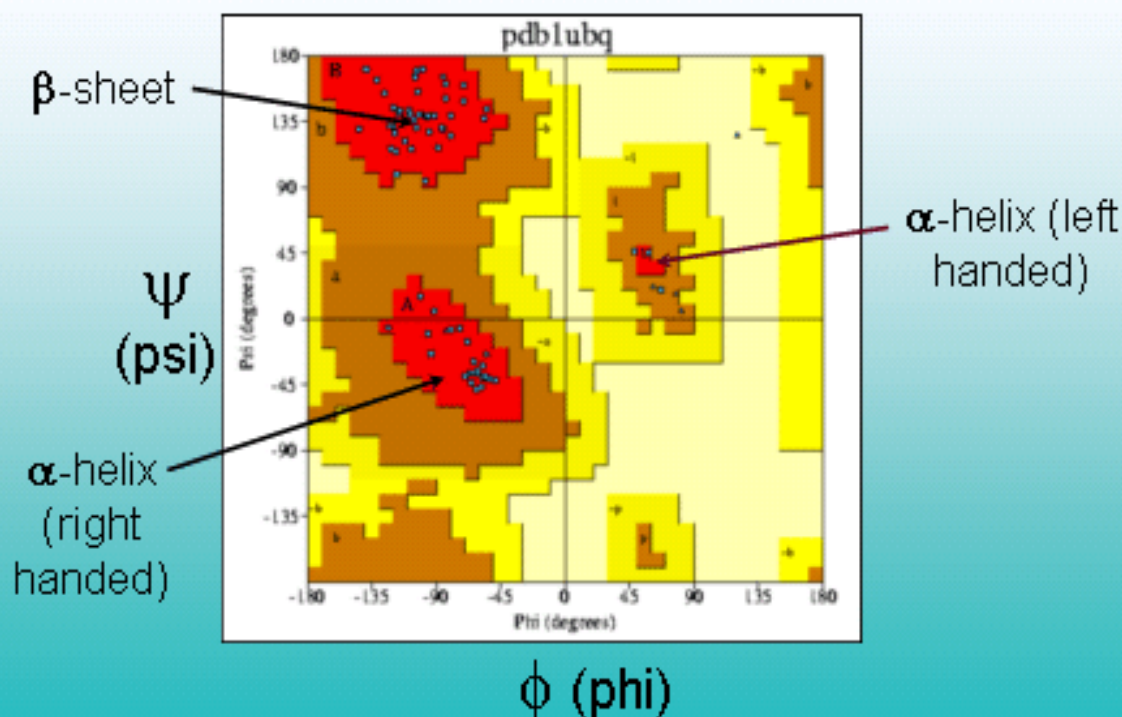
MHGAYRTPRSKTDAYGCQILETRAS

Levels of protein structure: 2

- Secondary
 - Local organization of the protein backbone: α -helix, β -strand (which assemble into β -sheets), turn and interconnecting loop



Ramachandran / phi-psi plot



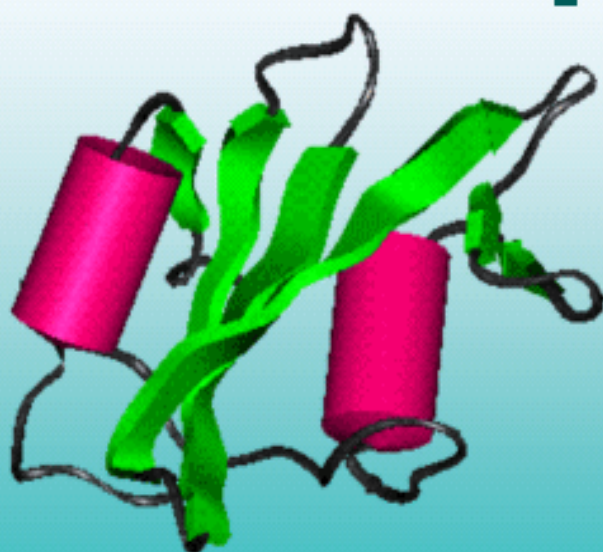
The β -sheet

- Side-chains project alternately up or down



- Amphipathicity if solvent exposed: hydrophilic residues on one face; hydrophobic ones on the other
- Backbone almost fully extended: thus one of the most loosely packed arrangements of residues.

Levels of protein structure: 3



- Tertiary
 - packing of secondary structure elements into a compact spatial unit
 - “Fold” or domain – this is the level to which structure prediction is currently possible

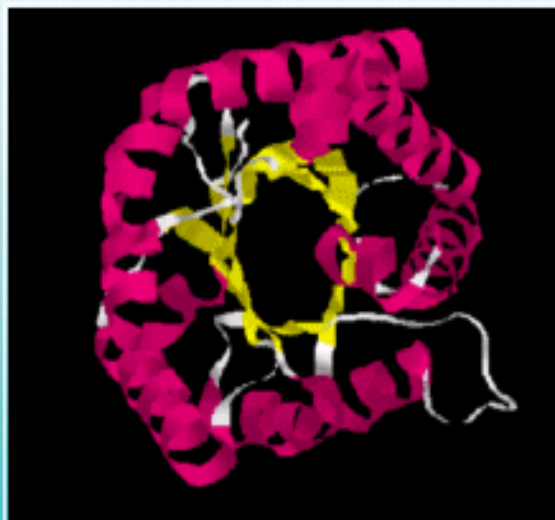
Levels of protein structure: 4



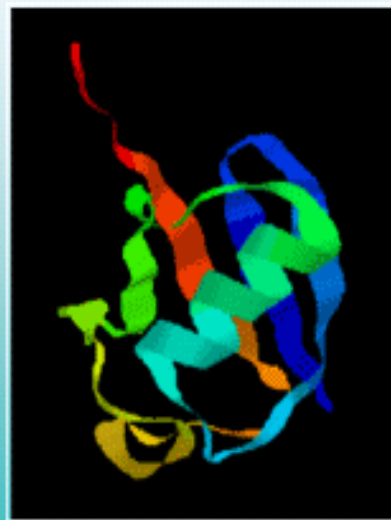
- Quaternary
 - Assembly of homo- or heteromeric protein chains
 - Usually the functional unit of a protein, especially for enzymes

Structural classes: 2

α/β (parallel β -sheet)



$\alpha+\beta$ (antiparallel β -sheet)



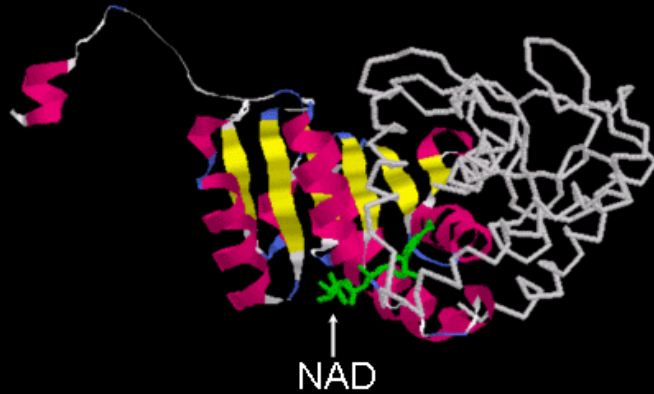
Most popular class!

Domain: a LEGO piece

- A domain is a compact folding unit of protein structure, usually associated with a function.
 - It is usually a “fold” - in the case of monomeric soluble proteins.
 - Comprises normally only one protein chain: rare examples involving 2 chains are known.
 - Domains can be shared between different proteins.

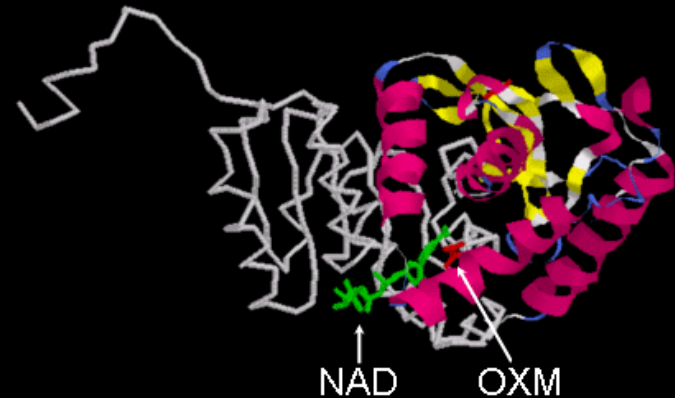
LDH – domain structure

Domain 1: Rossmann-fold (α/β)



LDH – domain structure - 2

Domain 2: substrate-binding ($\alpha+\beta$):



Protein architectures

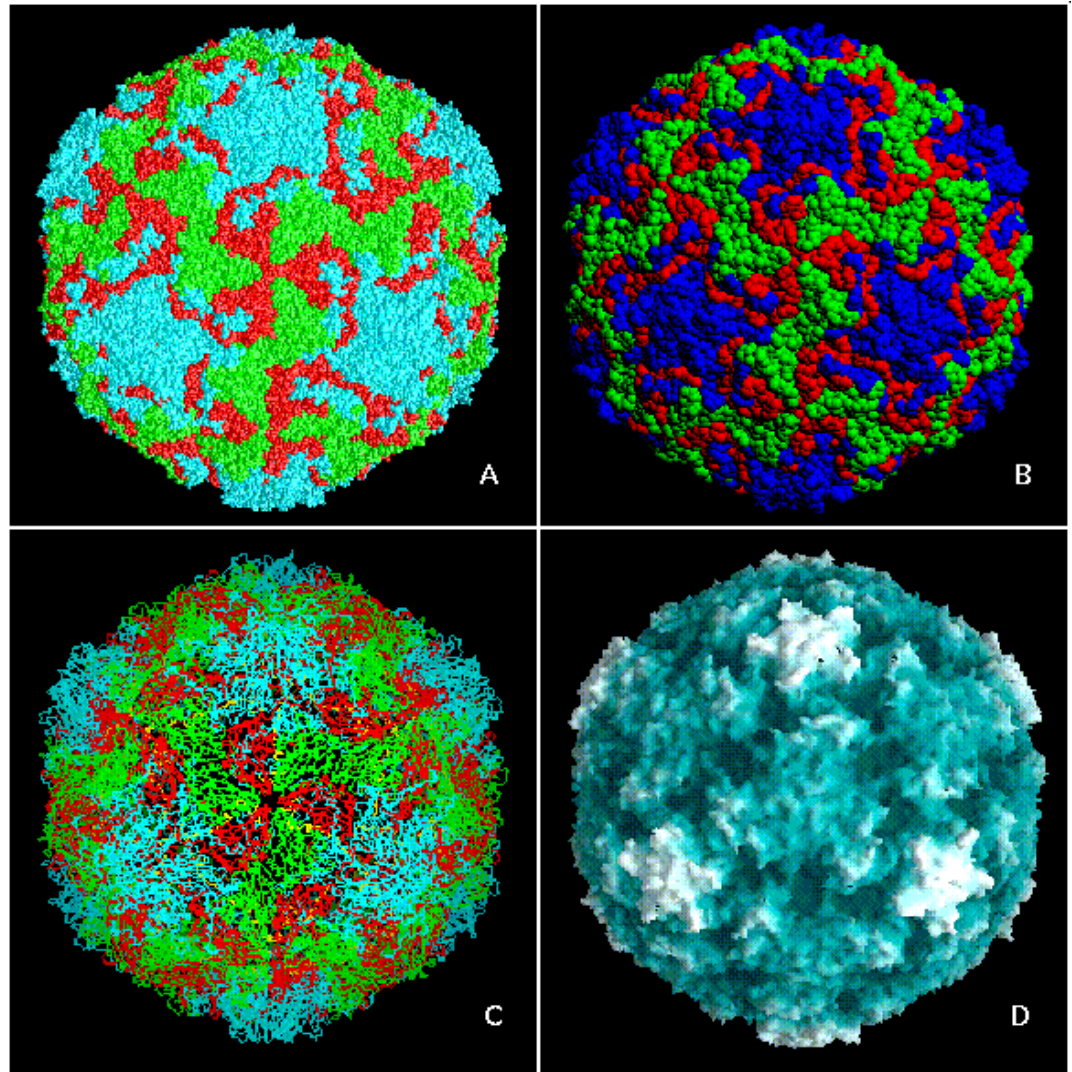
- Beads-on-a-string: sequential location: tyrosine-protein kinase receptor TIE-1 (immunoglobulin, EGF, fibronectin type-3 and protein kinase)

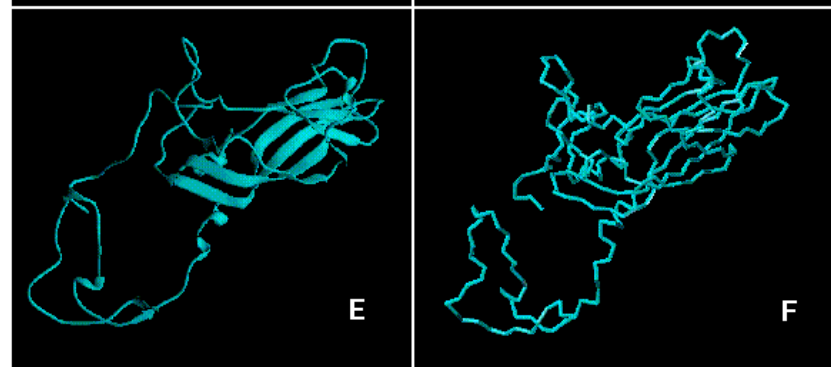
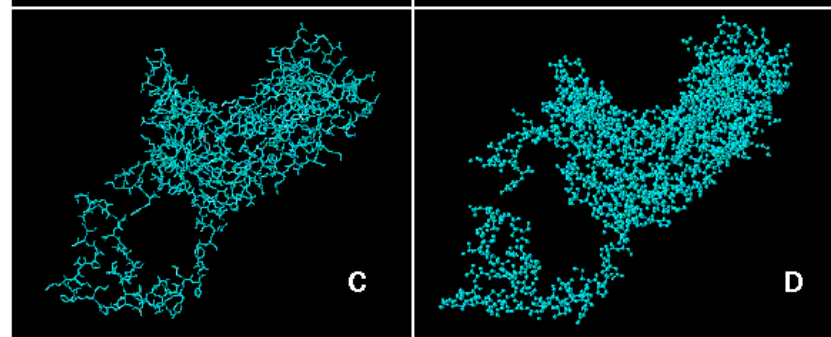
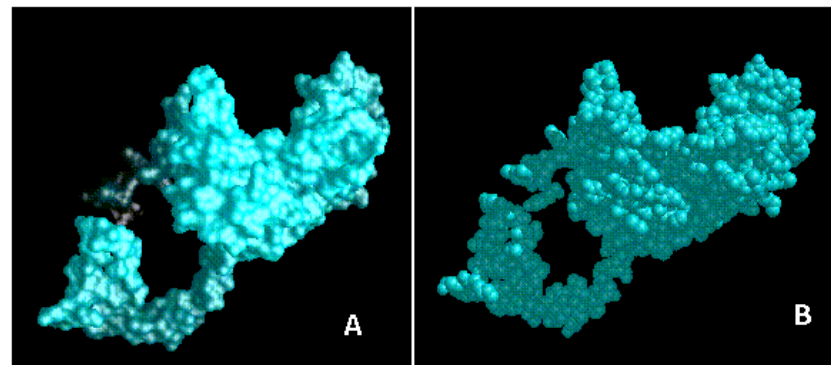
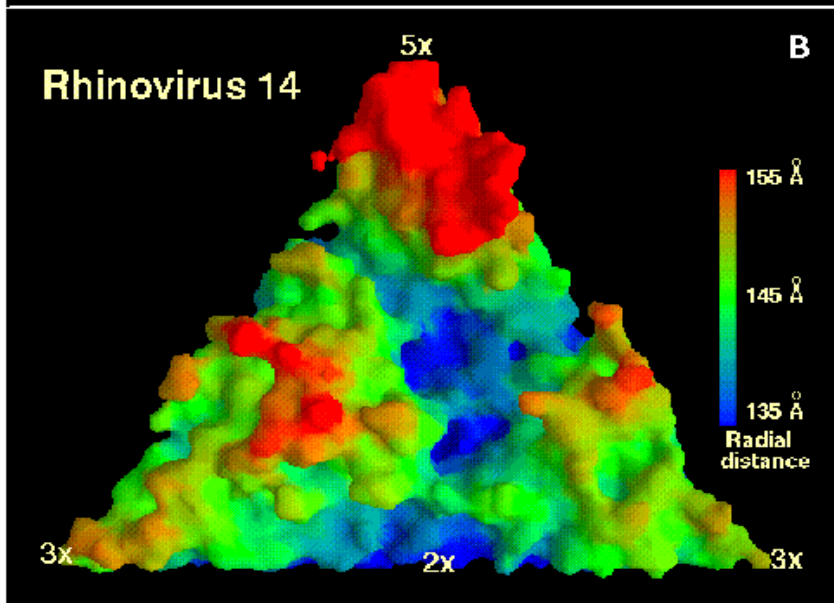
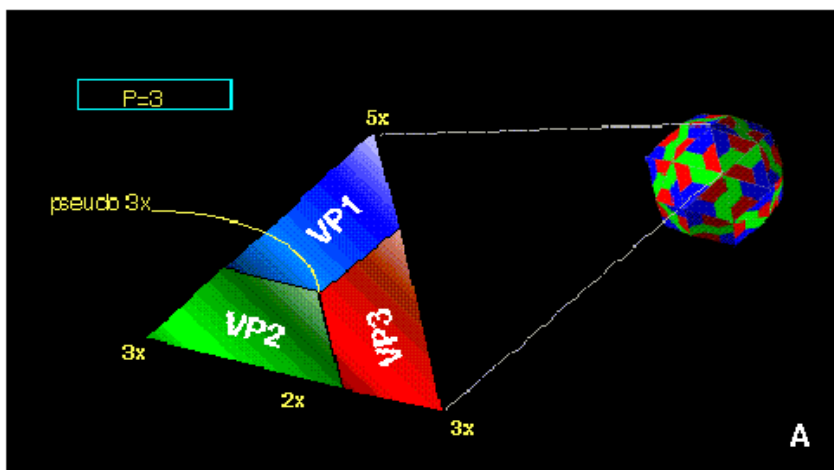


- Domain insertions: "plugged-in" - pyruvate kinase (1pkn): 3 domains - split domain 1



Virova castice R14
je prakticky
sfericky krystal
skladajici se z
identickych
podjednotek

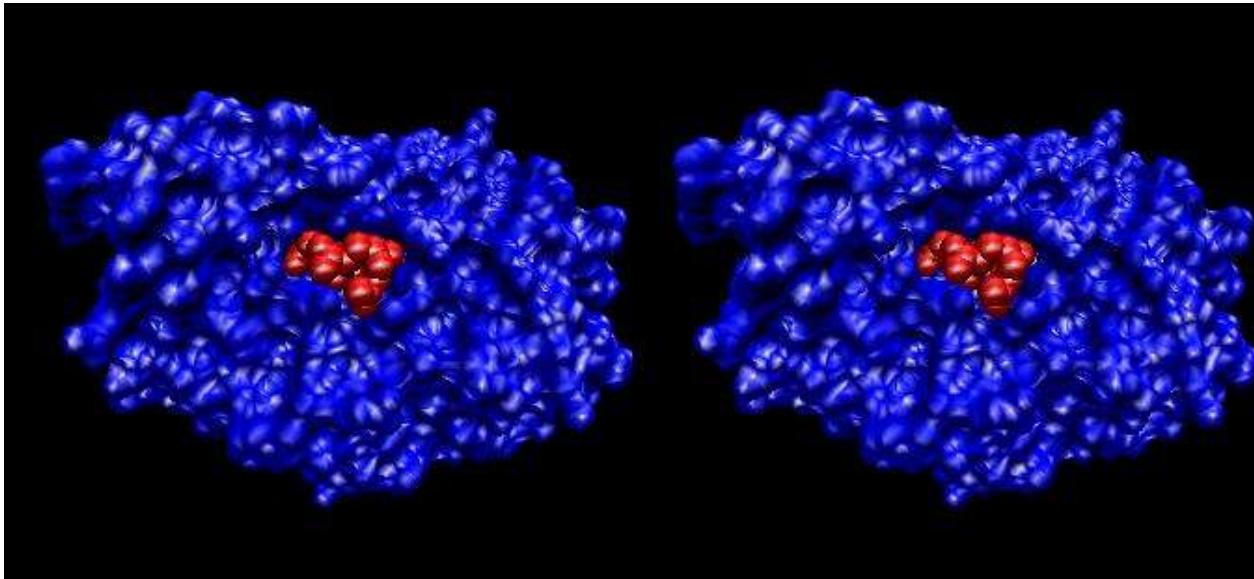




Dilci funkce proteinu

- ✦ Enzymy (katalyzatory, substrat se proměňuje v produkt, aktivní místo)
- ✦ Interakce protein-protein
- ✦ Interakce protein-DNA
- ✦ Interakce protein-ligand
- ✦ Transdukce signalu, regulace
- ✦ Strukturální proteiny (vlákna, glykoproteiny)
- ✦ Motory

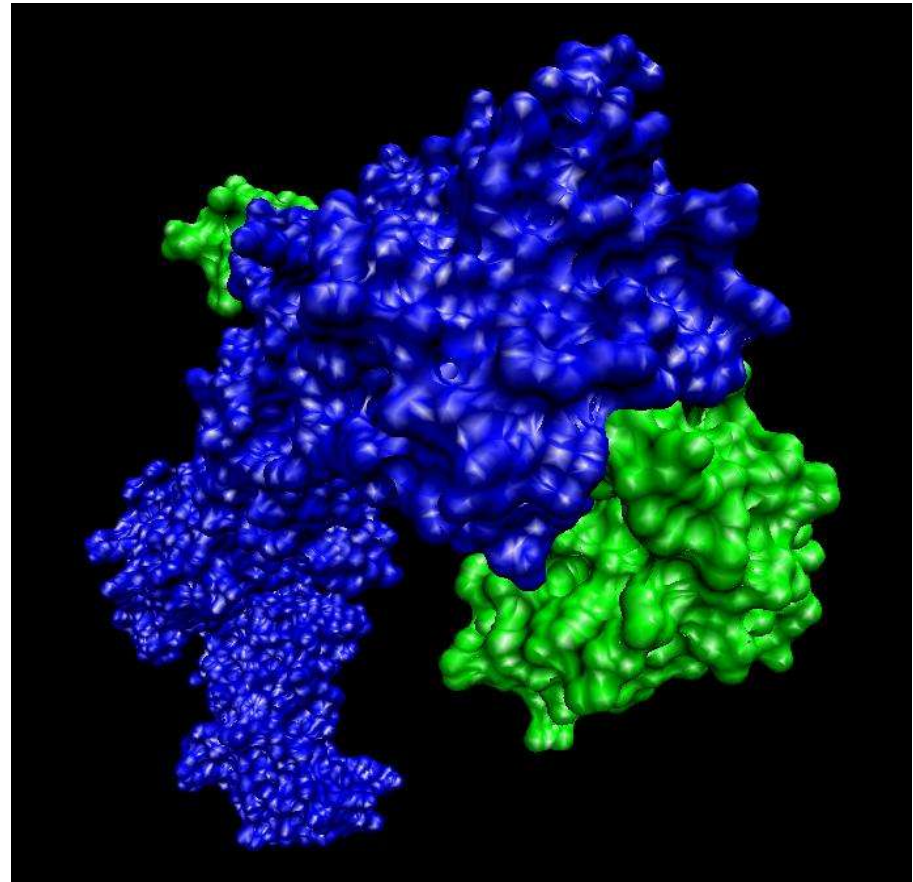
Enzymy (substrat - produkt)



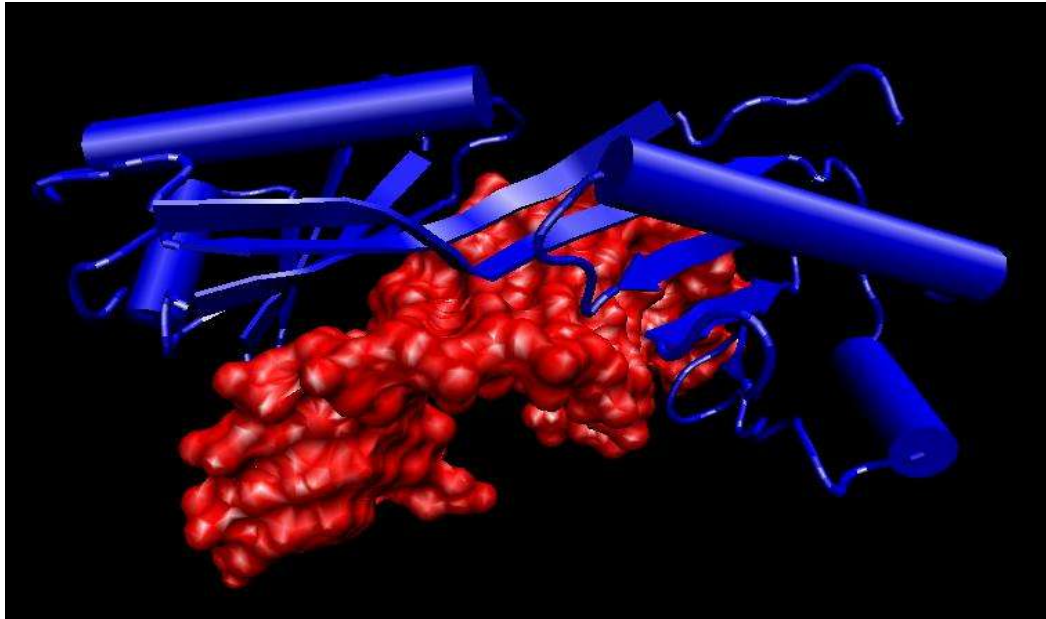
Inhibitor HIV proteazy, která katalyzuje důležité
krok v syntéze funkčního viru způsobujícího AIDS

Interakce protein - protein

Interakce mezi protilátkou a antigenem. Takové interakce jsou základem humorální imunity. Díky proteinové paměti může organismus likvidovat cizí látky.



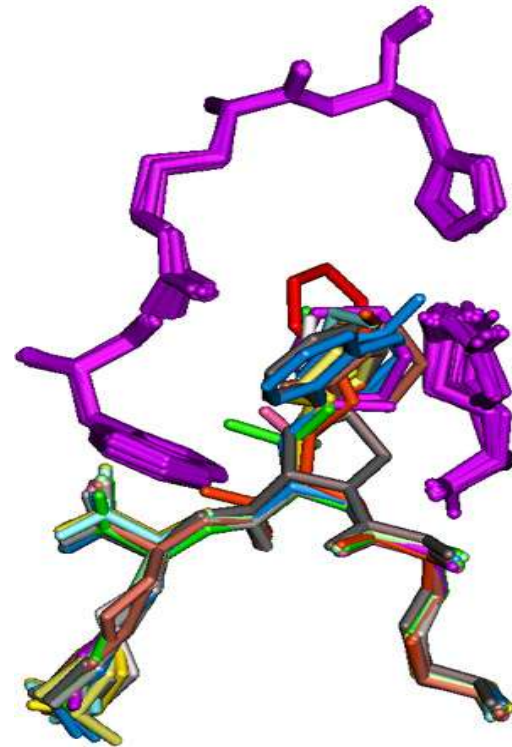
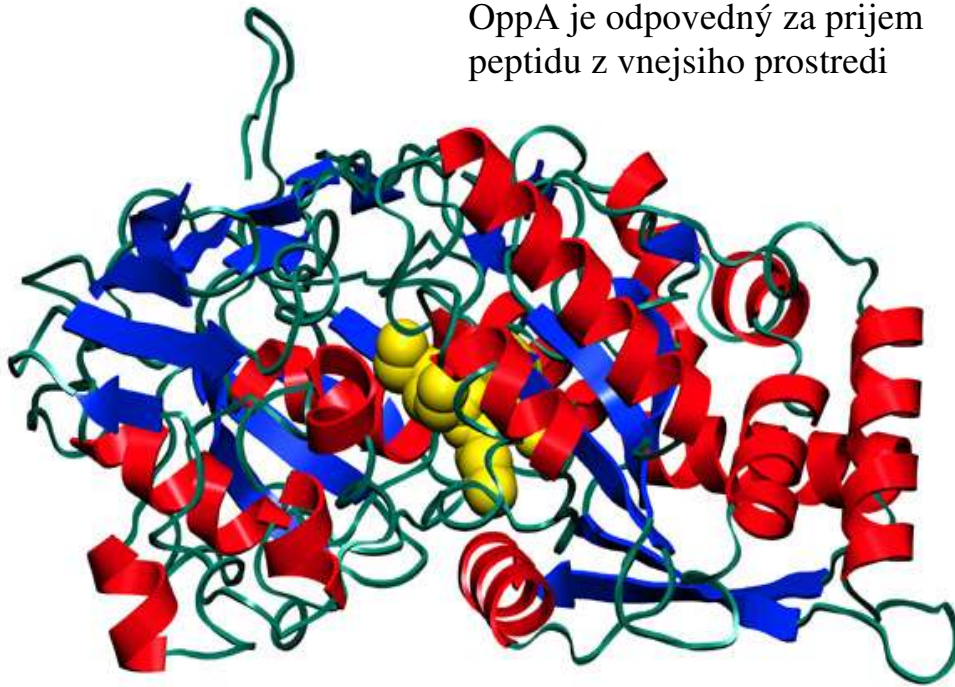
Interakce protein - DNA



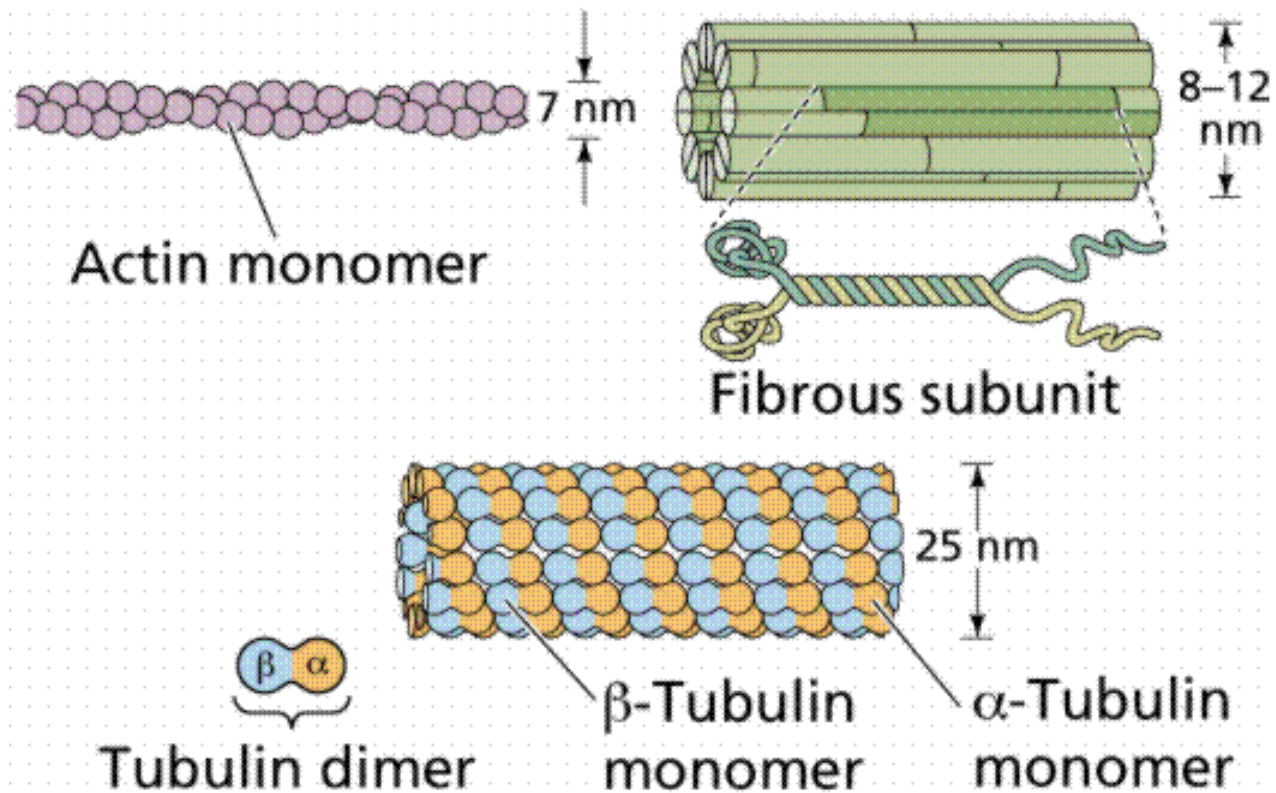
TBP (TATA box binding protein) – důležitá součást transkripčního komplexu, kontrolující polohu, ve které transkripce proběhne

Interakce protein (OppA) – ligand (KXXK)

OppA je odpovědný za příjem
peptidu z vnějšího prostředí

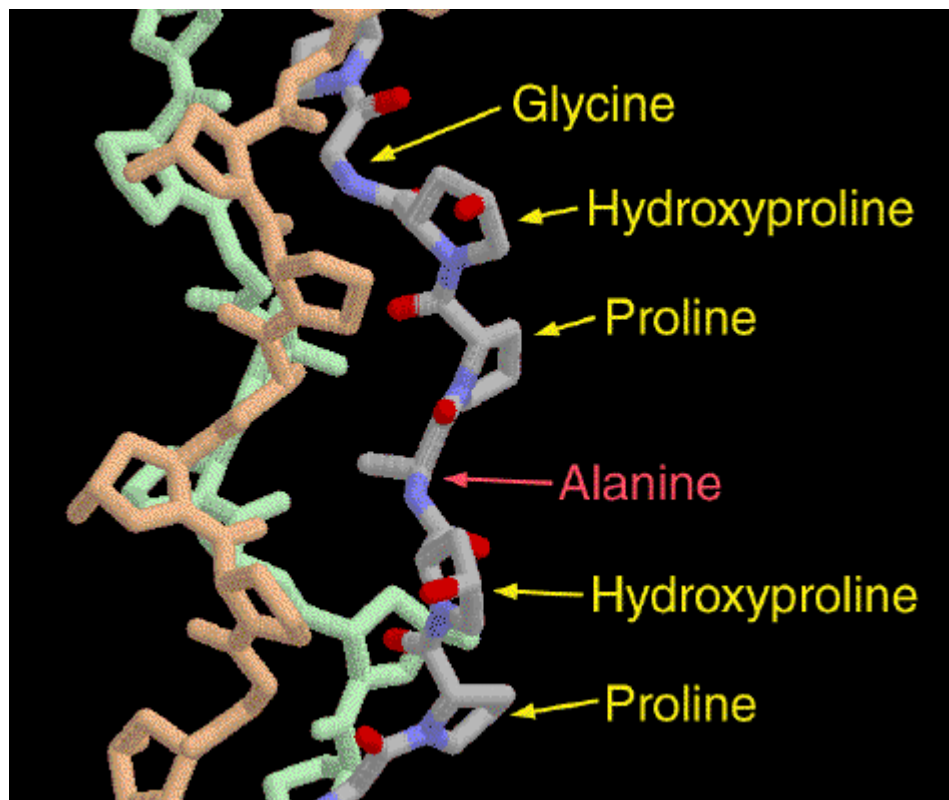


Strukturni proteiny

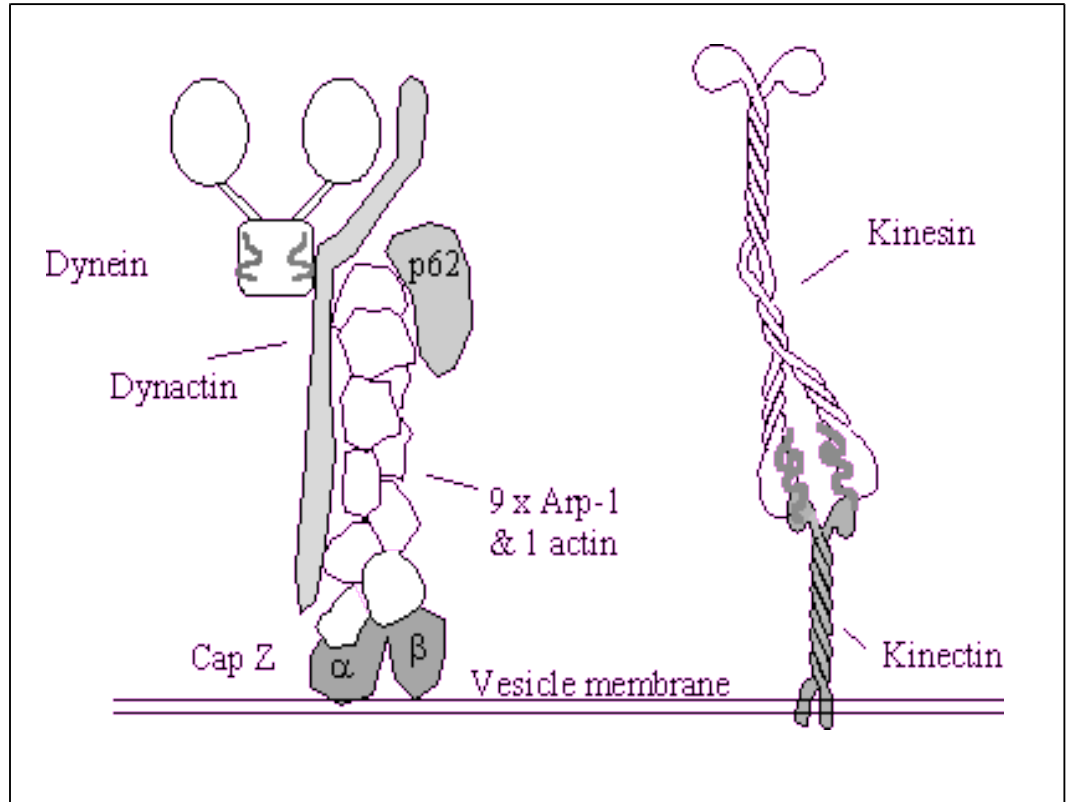
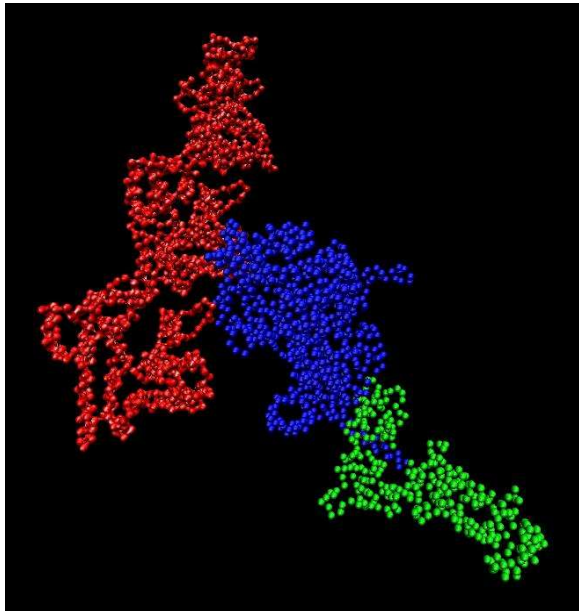


Strukturní proteiny

Kolagen – jeho tvorba vyzaduje vitamin C
Tvorí kolem 30% proteinu v lidském tele. Je periodický (GPP nebo GPK)

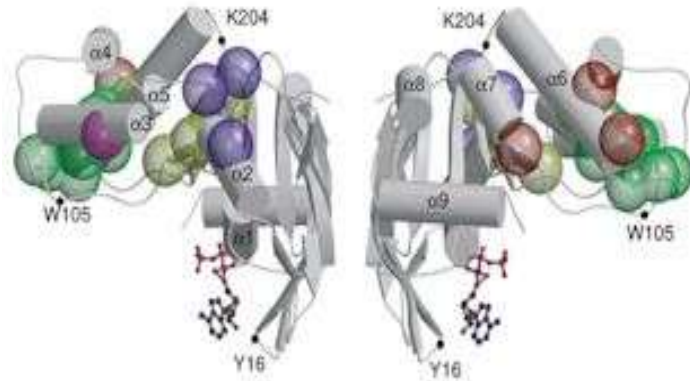


actin - myosin

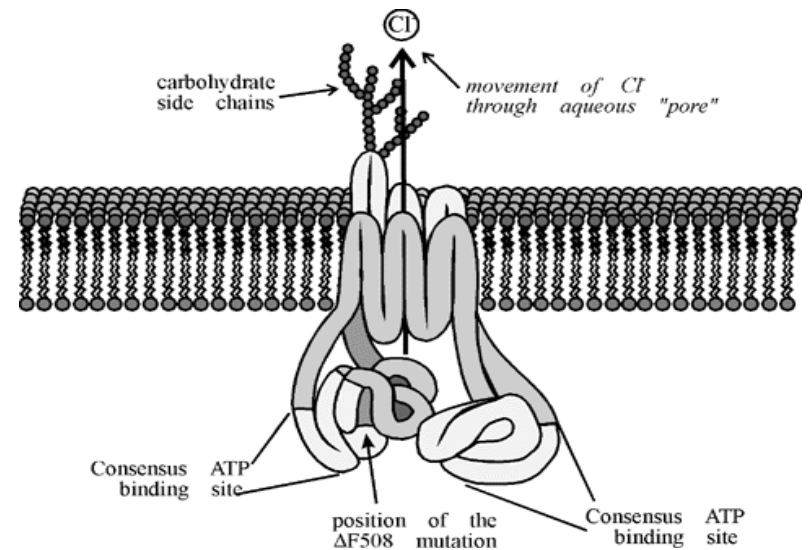


Nemoci způsobene nespravnym skladanim proteinu

Cysticka fibroza	CFTR	transport Na ⁺ a Cl ⁻
Alzheimerova nemoc	APP+LP E	agregace
BSE	PrP	agregace



CFTR, respektive jeho deficit (zpusoben nespravnim skladanim a destrukci v ER) je pricinou cysticke fibrozy. Ovlivnuje predevsim plice, jatra a potni zlazy.



PrPc a PrPSc

