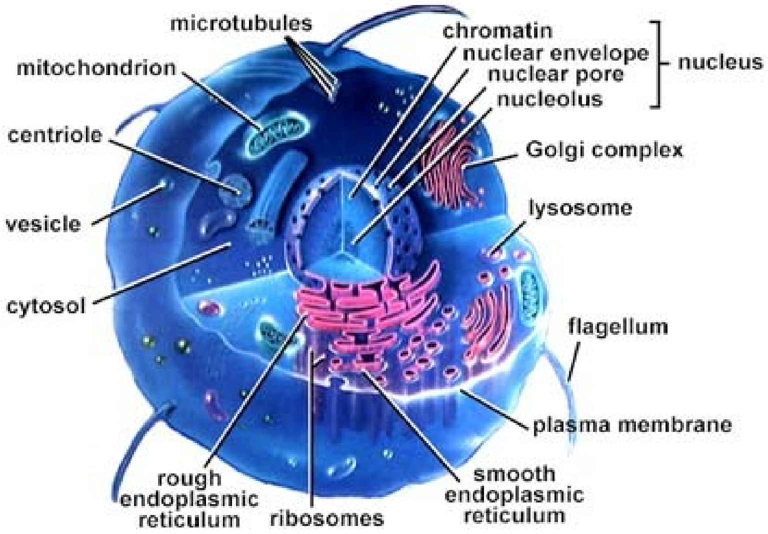
Bioinformatika

- zpracovani hromadnich biologickych dat
- koreny v 50. a 60. letech – struktura DNA, proteinu, geneticky kod
- data = sekvence, struktury, exprese genu, metabolizmus,...
- analyzy = zarovnavani sekvenci, fylogeneticke stromy, identifikace dulezitych strukturnich a funkcnich motivu, predpovidani struktury,...
- DNA - RNA - PROTEIN ::: PROTEIN - STRUKTURA – FUNKCE
- rozsah dat u cloveka:
  - 3 miliardy bazi DNA (x2) / bunku
  - 30000 ruznych genu/proteinu
  - 100 miliard bunek / jedince

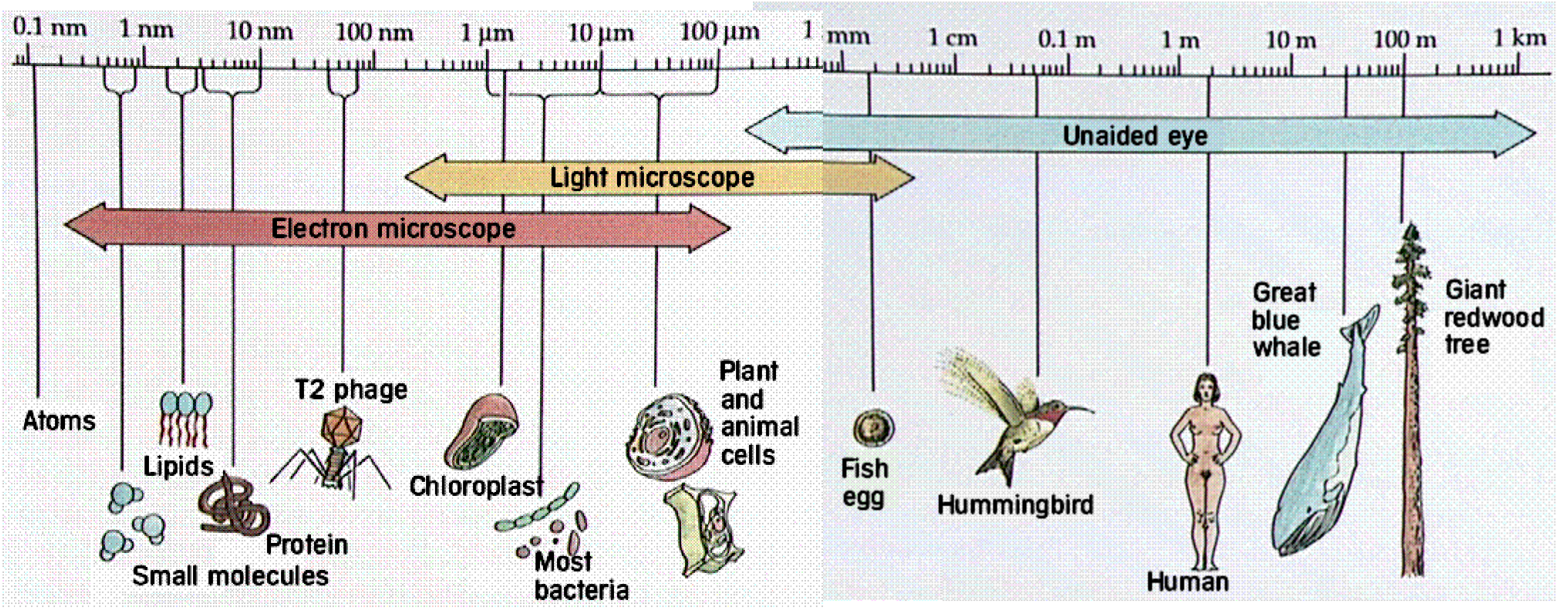
Bunky

- Zakladni forma organizace zive hmoty

- Molekuly/geny/proteiny
- Proteinove komplexy/mem
- Organely a jine substruktu
- **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 znacne miry odpovidaji dnesnimu oznaceni gen /**genotyp**/

GENOTYP

FENOTYP

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

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

GEN.KOD

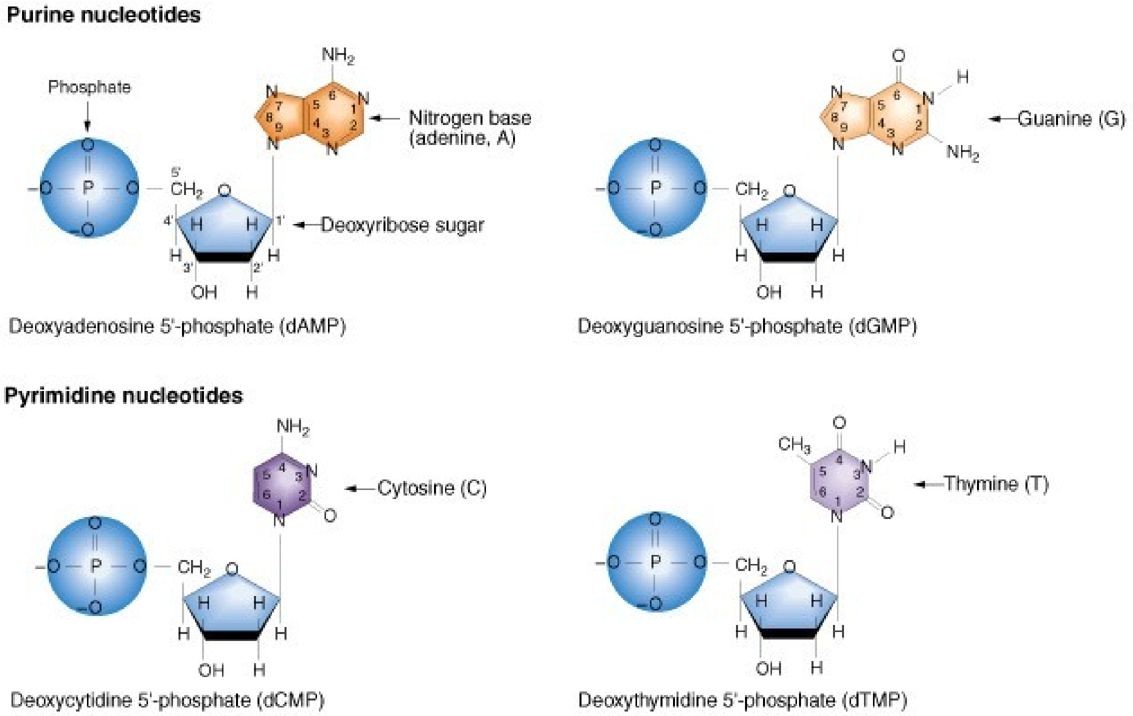
STRUKTURA

FUNKCE

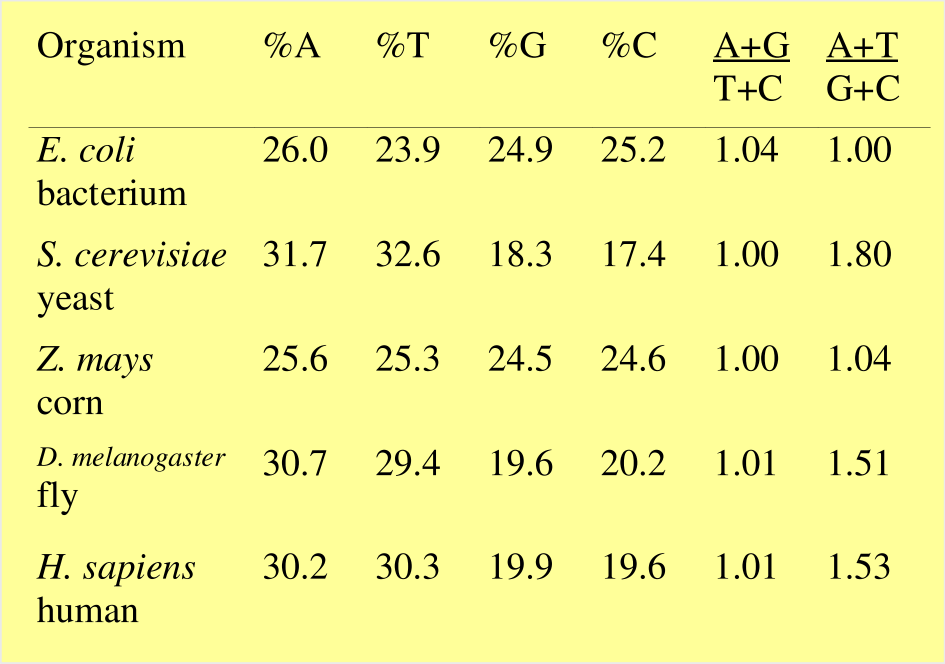
Struktura DNA

- Objevena v 19.stol
- Považována za jednoduchou
- Polymericka molekula
- Monomerem je
  - Ribonukleotid /RNA/
  - deoxyribonukleotid /DNA/
- Symbolicke znaceni A,C,G,T,U
- Prostorova struktura: Watson, Crick, Franklin

Struktura DNA



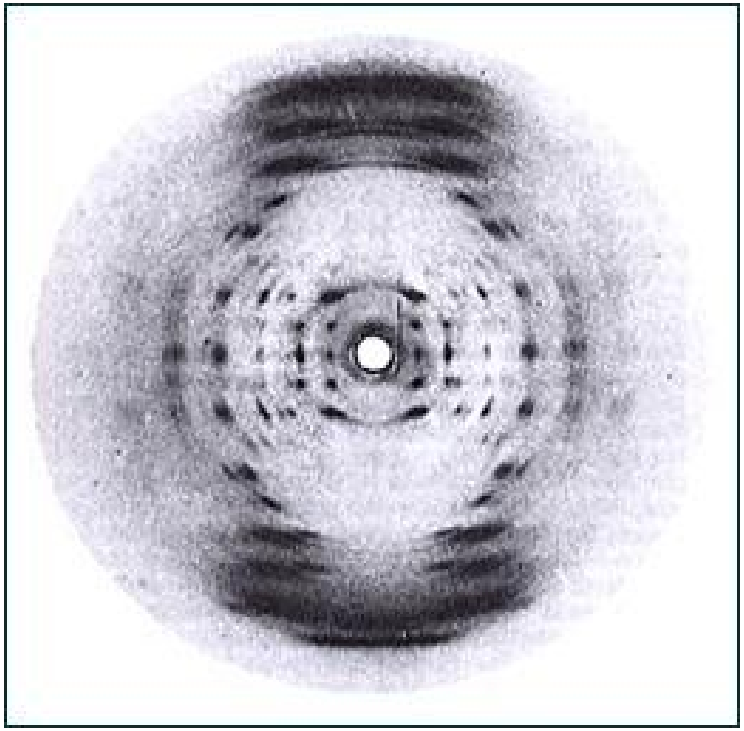
Struktura DNA



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

Obsah bazi v DNA ruznych organismu

Struktura DNA

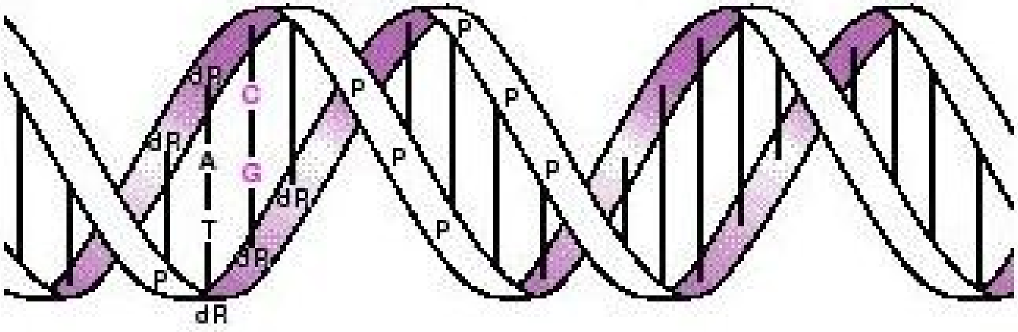


Difrakce krystalicke DNA z roku 1952 od Rosalind Franklin
Svedci o periodicite 0.3nm a 3.4nm

Struktura DNA

Struktura DNA

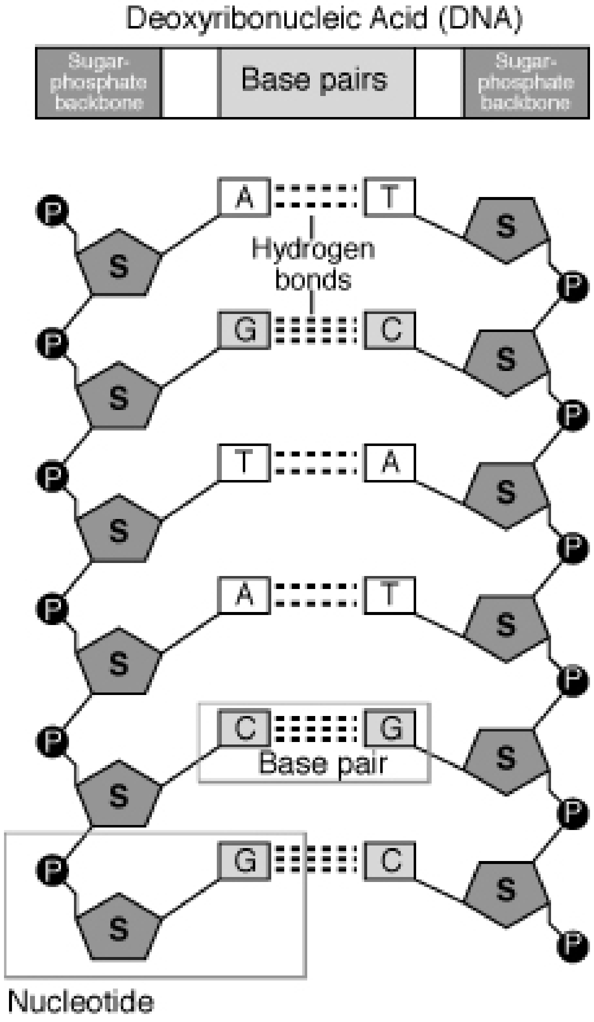
4.10 DOUBLE HELIX



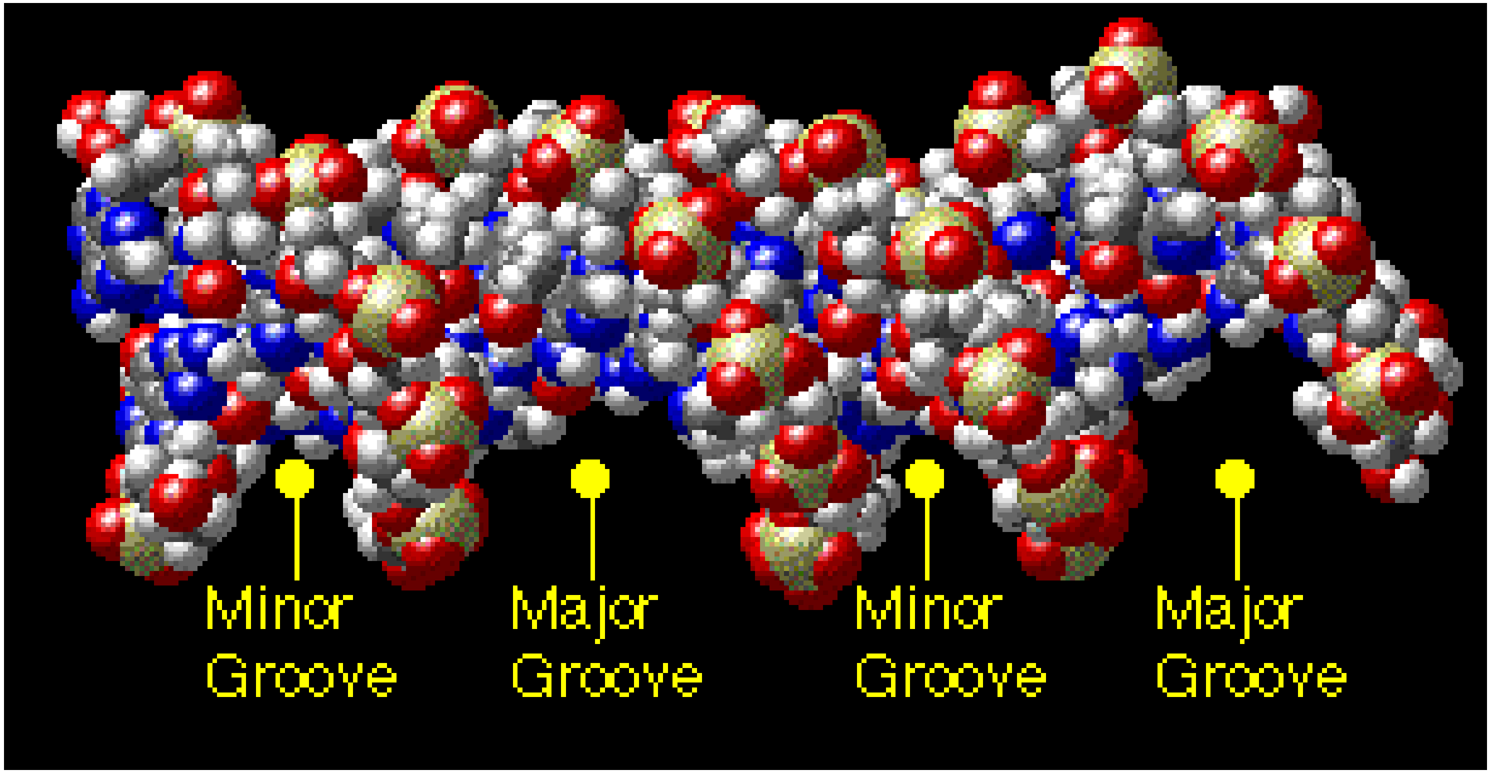
P = phosphate, dR = deoxy ribose, A,T, C, G = base pairs

Sekvence DNA

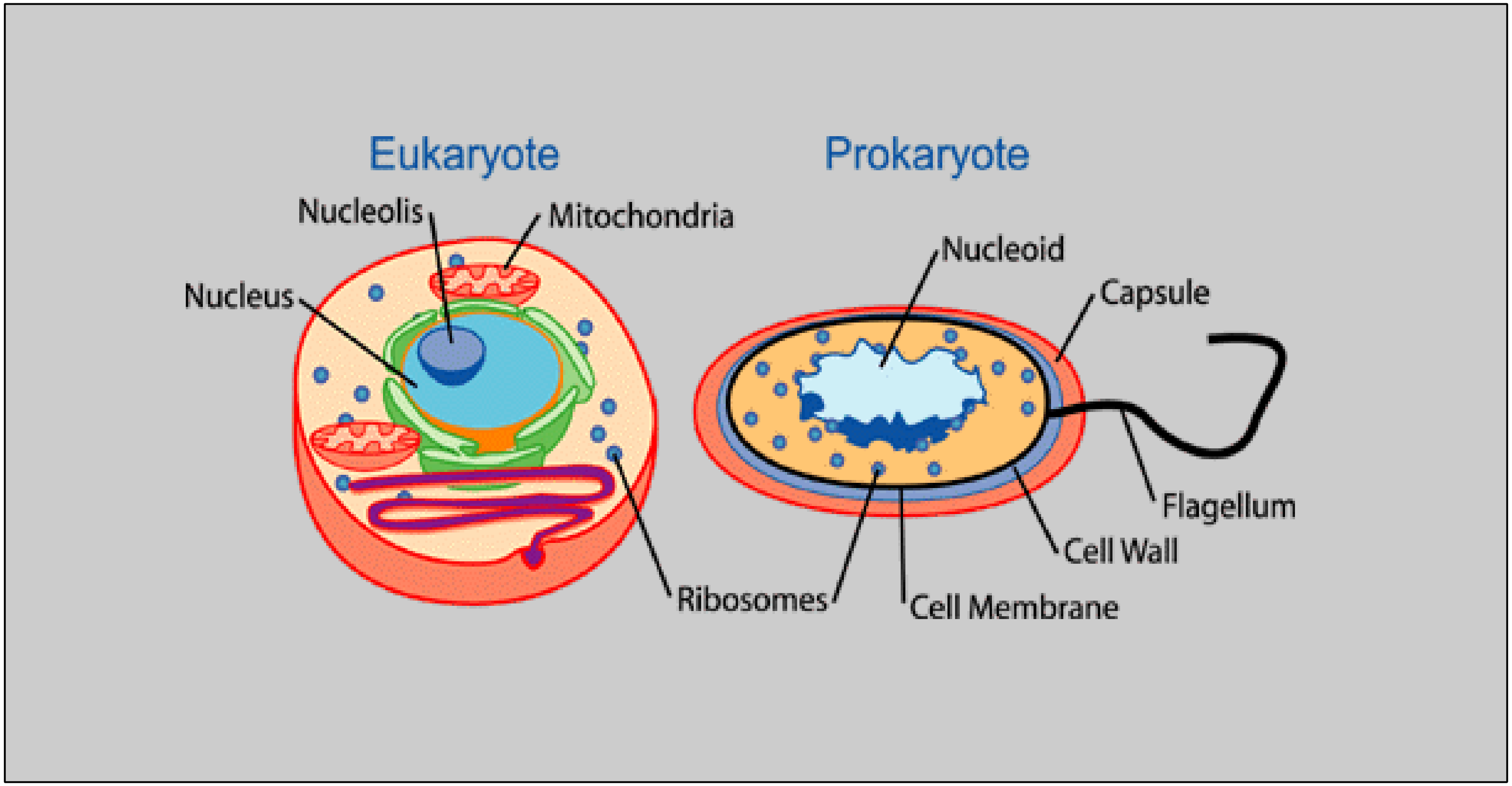
Vazby mezi vlaknem jsou vodikove mustky, ktere jsou ve srovnani s kovalentnimi vazbami slabe. Pusobením teploty do 100oC muze dojit k denaturaci, t.j. rozpojeni vlaken. Vyuziva se to v PCR, ale vlastne taky v replikaci, transkripci.



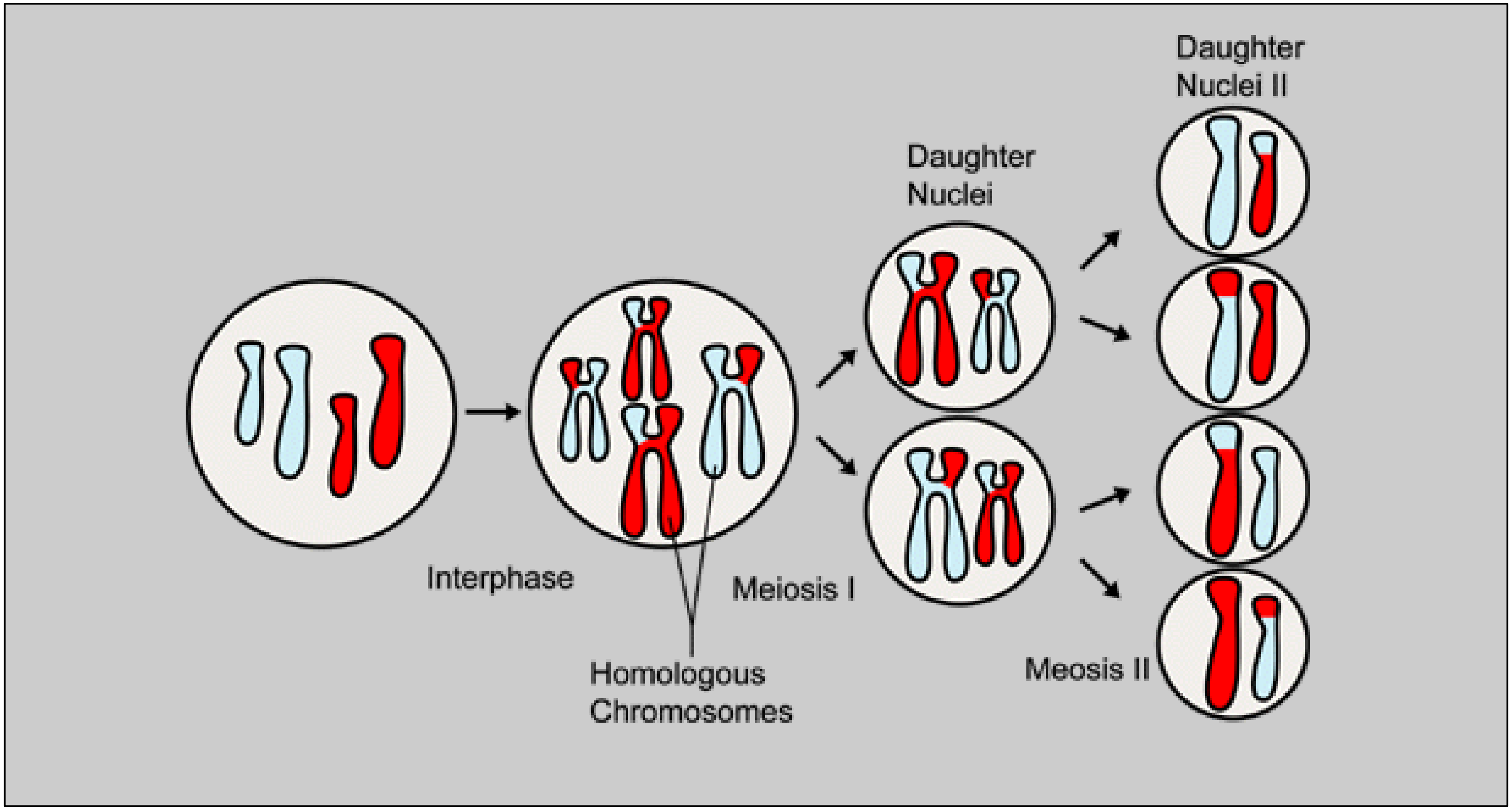
Sekvence DNA



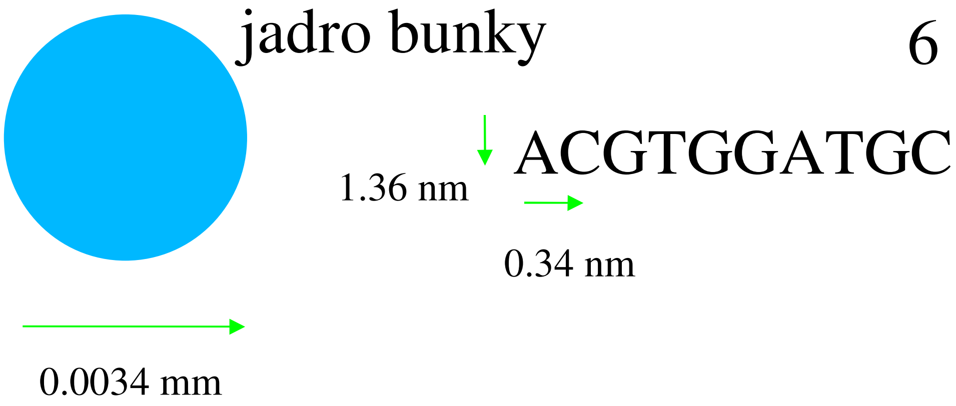
Zaklady genetiky



Zaklady genetiky



Organizace DNA



6 000 000 000 bp

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

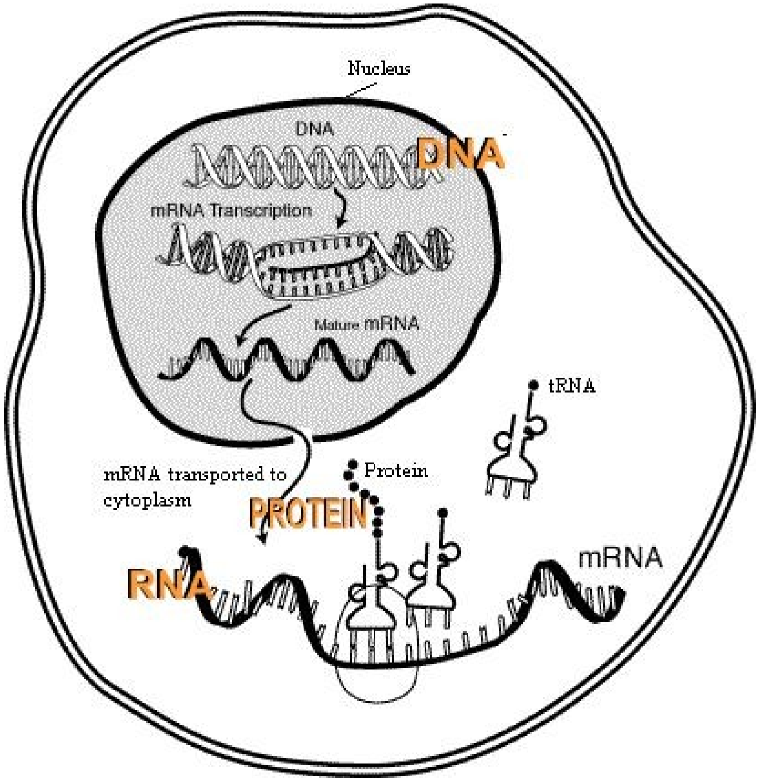
Jeden gen – jeden protein

- 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

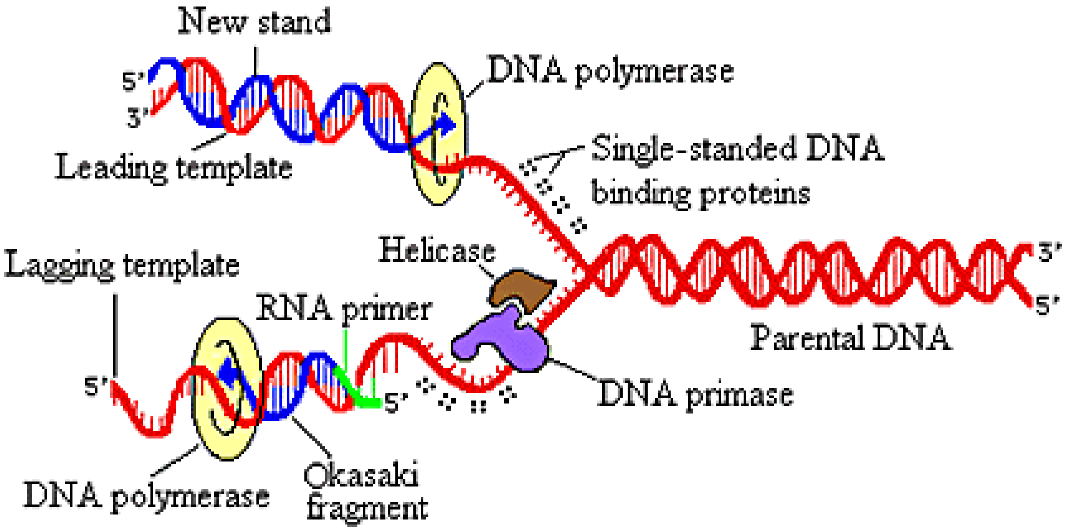
- DNA -> RNA -> PROTEIN



Replikace DNA

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

Replikace DNA



Collaboration of Proteins at the Replication Fork

Transkripce

- Hlavni enzym: RNA polymeraza
- Transkripcní komplex: cela skala proteinu vazanych na specificke sekvence DNA a sebe navzajem
- Templat: DNA
- Substrat: ribonukleotidy
- Produkt: RNA

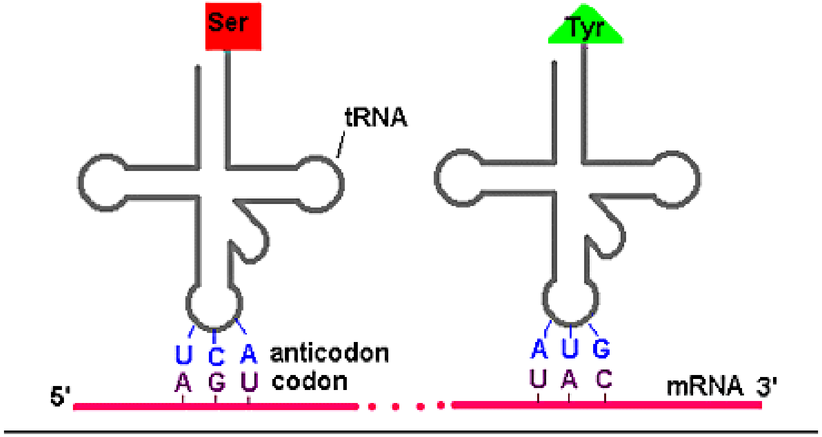
Struktura proteinu

- Polymericka molekula
- Monomerem je aminokyselina
- Symbolicke znaceni ACEFGHIKLMNPQRSTVWY
- Prostorova struktura se urcuje z difrakce rontgenovych paprsku nebo magnetickou rezonanci /NMR/

Geneticky kod

- Rozlustili Nirenberg, Matthaei, Khorana, Holley
- Zivot je digitalni
- 3 nukleotidy koduji jednu aminokyselinu v proteinovem retezci
- 4**3 = 64, ale jenom 20 aminokyselin

Geneticky kod



2nd base in codon

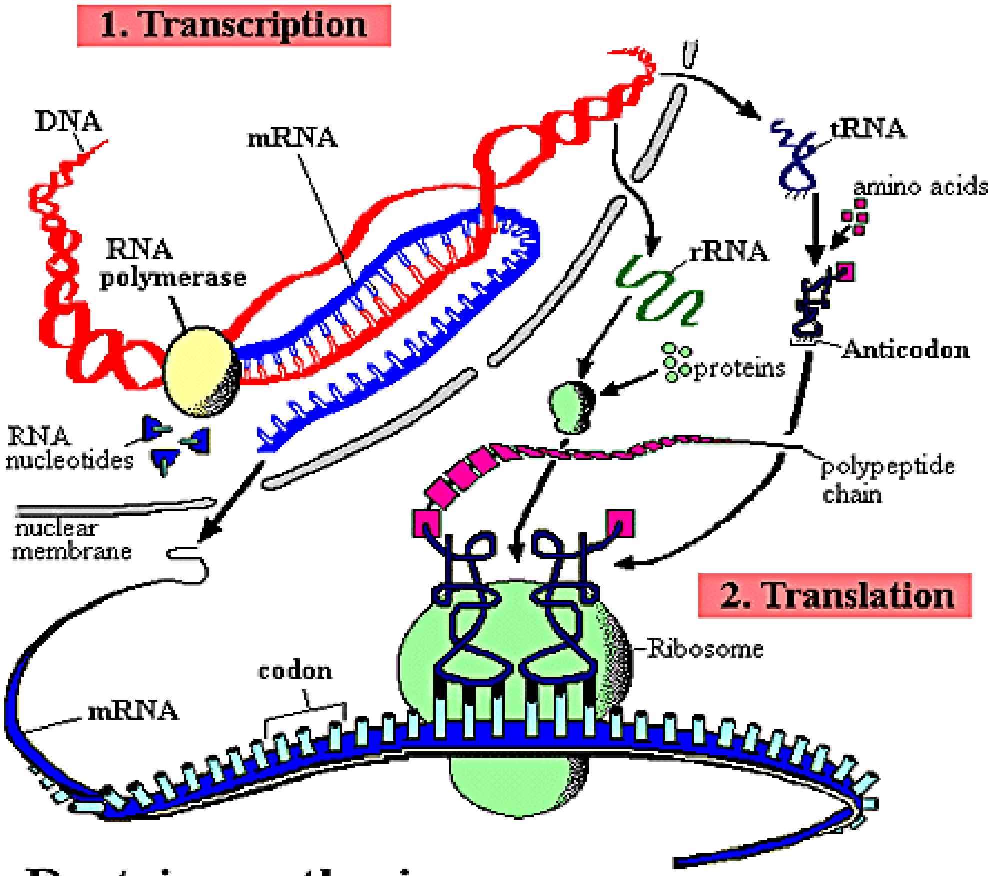
1st base in codon	U	C	A	G	3rd base in codon
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala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



Protein synthesis

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?

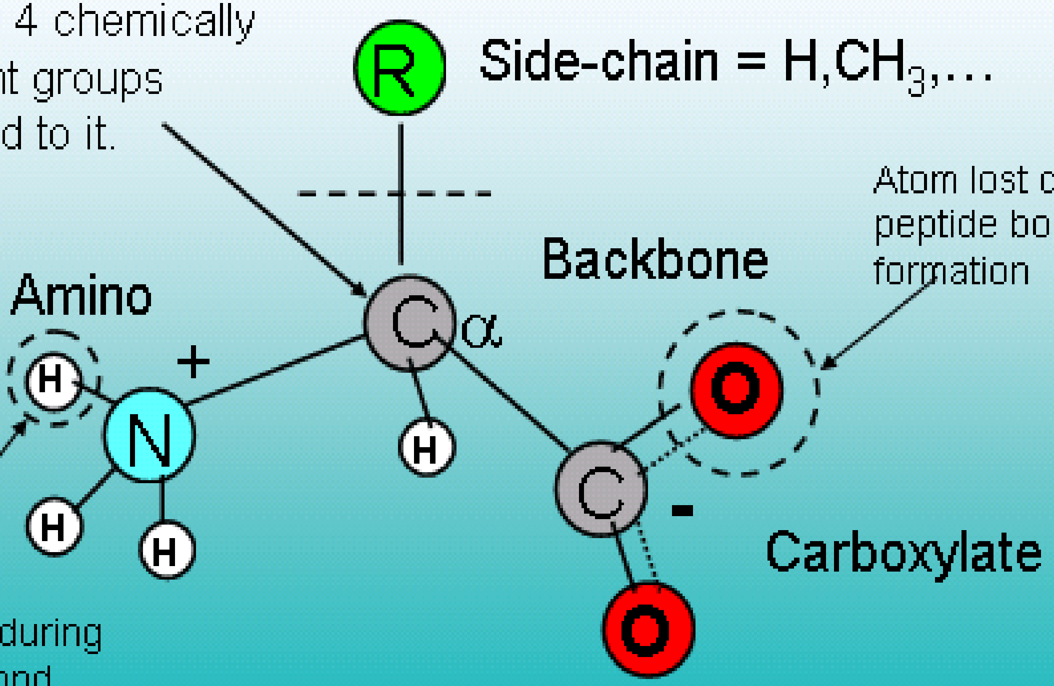
Struktura proteinu

- Primarni = sekvence
- Sekundarni = puvodne jenom a-helix a beta struktury, dnes i jine substruktury
- Terciarní = 3-D usporadáni
- Domeny – jedna nebo vic na protein
- Komplexy – skladajici se z vice 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

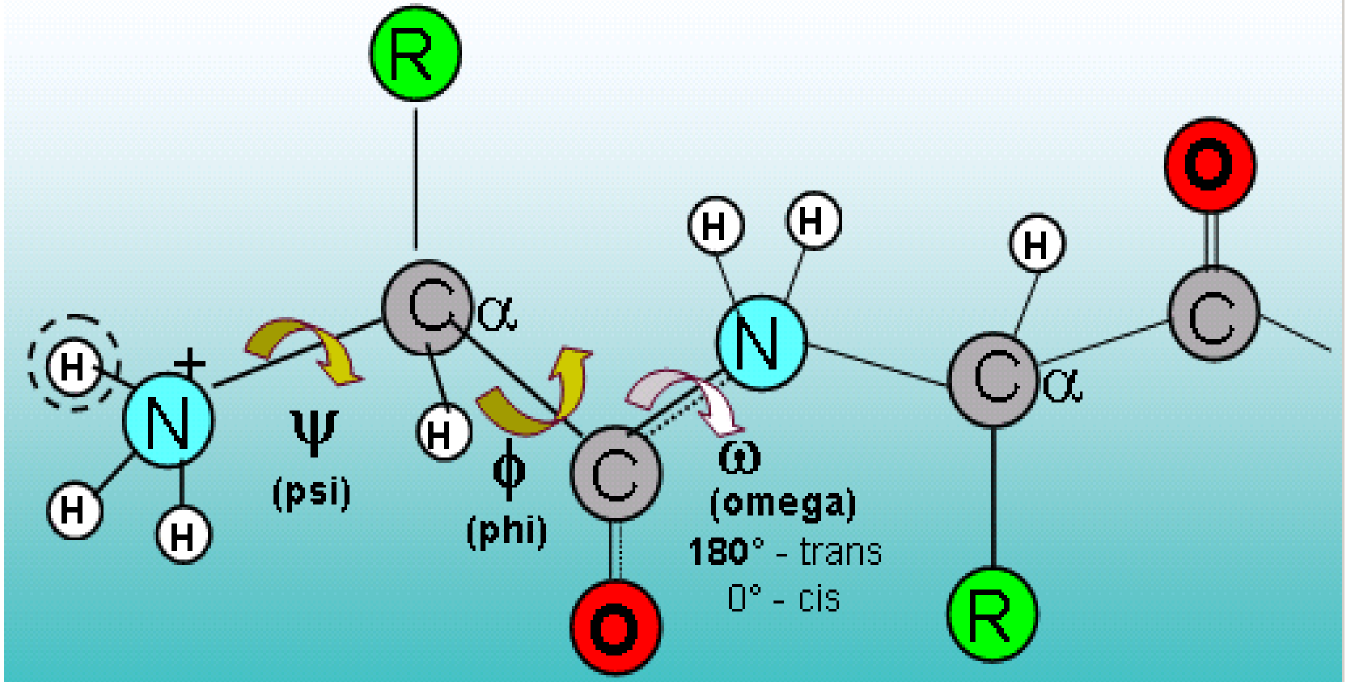


Cα is a chiral centre: i.e. has 4 chemically different groups attached to it.

Atom lost during peptide bond formation

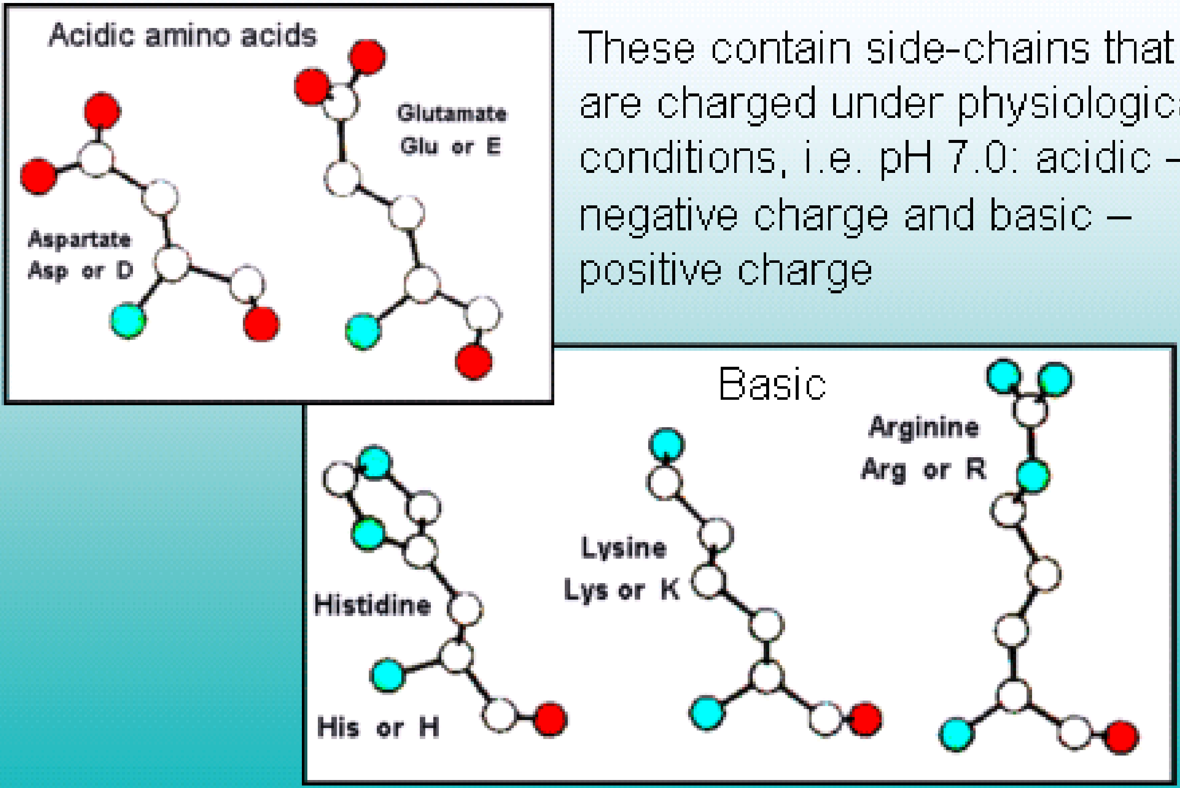
Atom lost during peptide bond formation

Backbone Torsion Angles

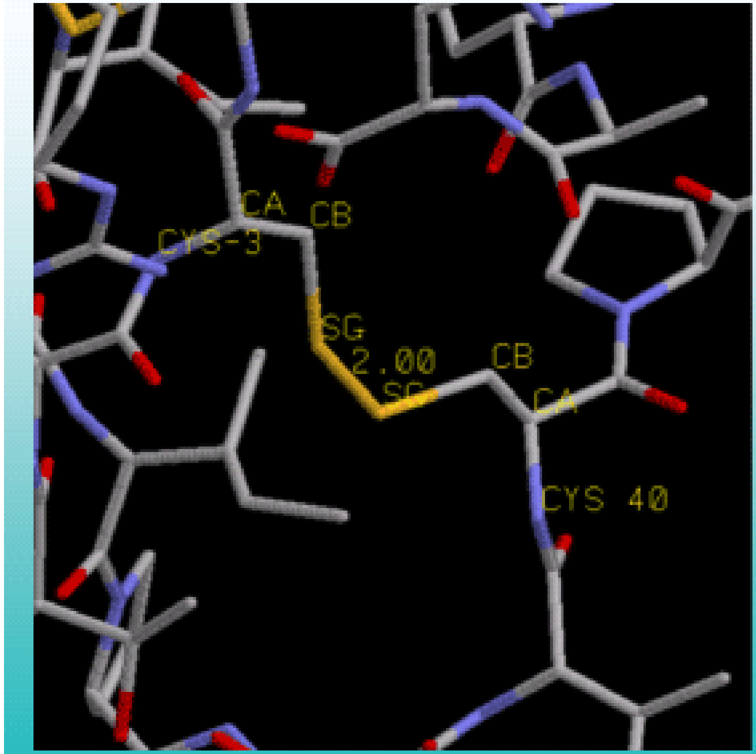


Charged residues

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 Sγ (or ‘SG’) atoms of two cysteine residues

Levels of protein structure: 0,1

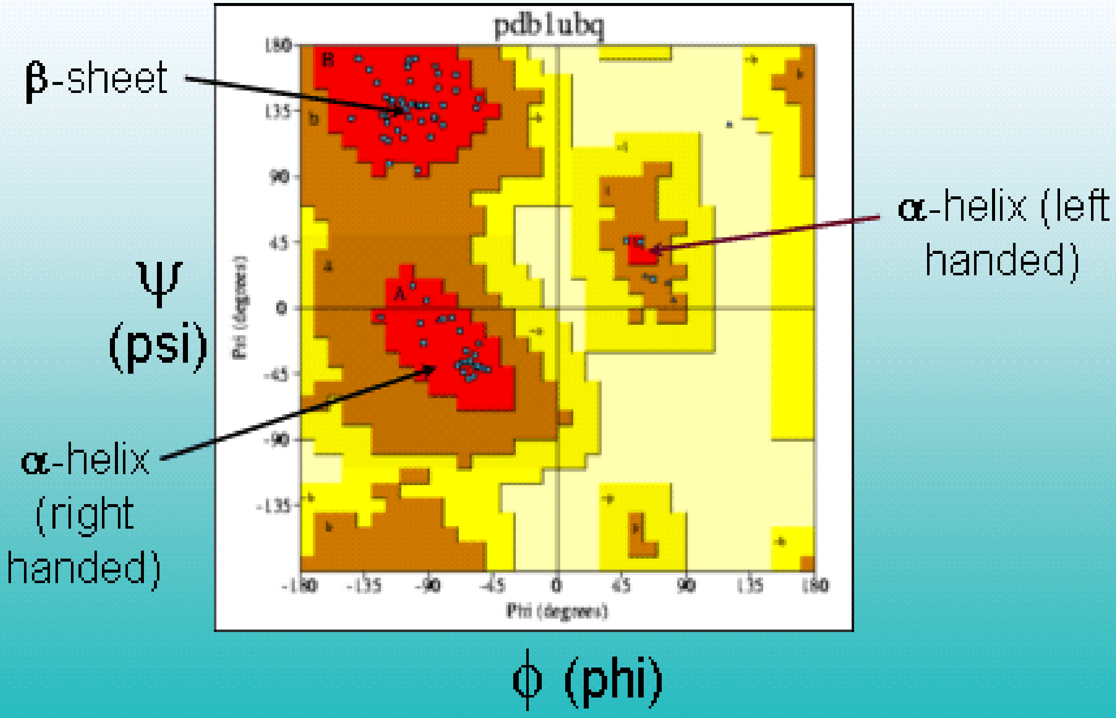
- Zeroth: amino acid composition – no structural information
- Primary
  - This is simply the order of covalent linkages along the polypeptide chain, i.e. the sequence itself

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



β-sheet

ψ (psi)

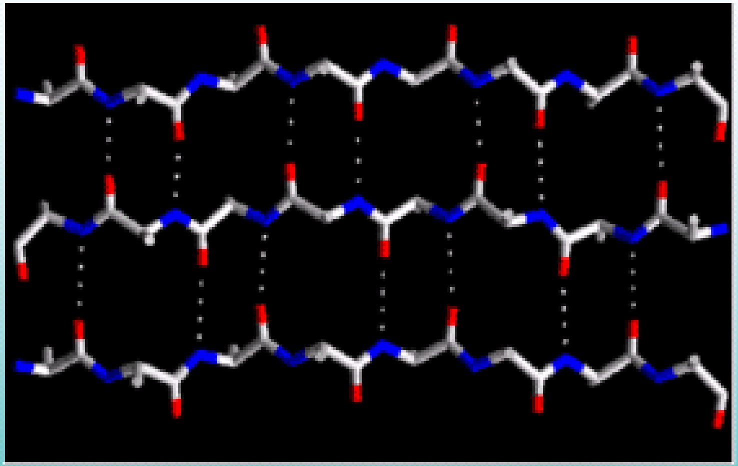
α-helix (right handed)

α-helix (left handed)

ϕ (phi)

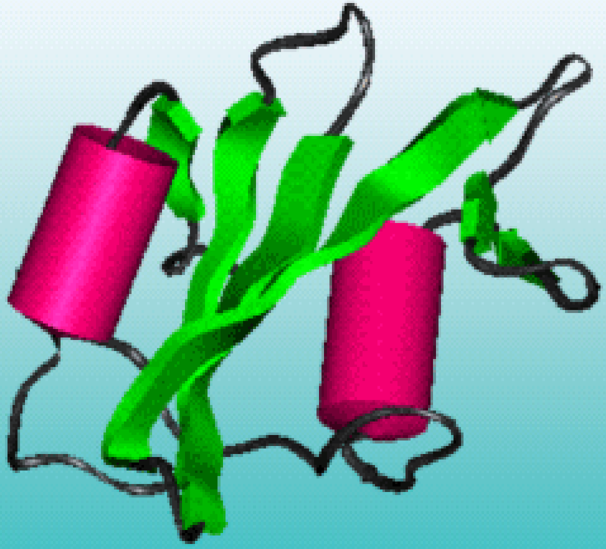
The β-sheet

- 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 lost loosely packed arrangements of residues.

Levels of protein structure: 3



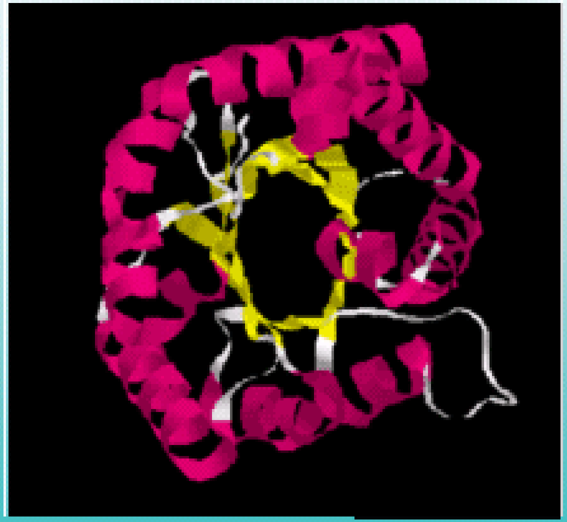
- Tertiary
  - packing of secondary structure elements into a compact spatial unit
  - "Fold" or domain – this is the level to which structure prediction is currently possible

Levels of protein structure: 4

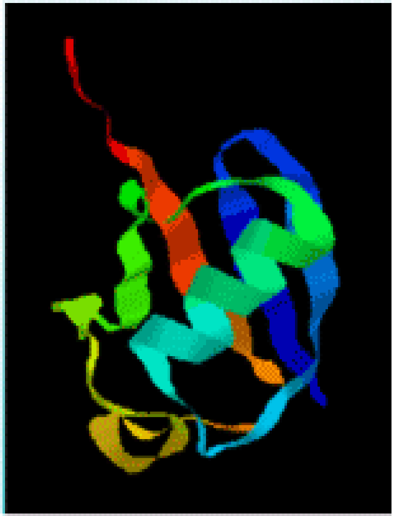
- Quaternary
  - Assembly of homo- or heteromeric protein chains
  - Usually the functional unit of a protein, especially for enzymes

Structural classes: 2

α/β (parallel β-sheet)



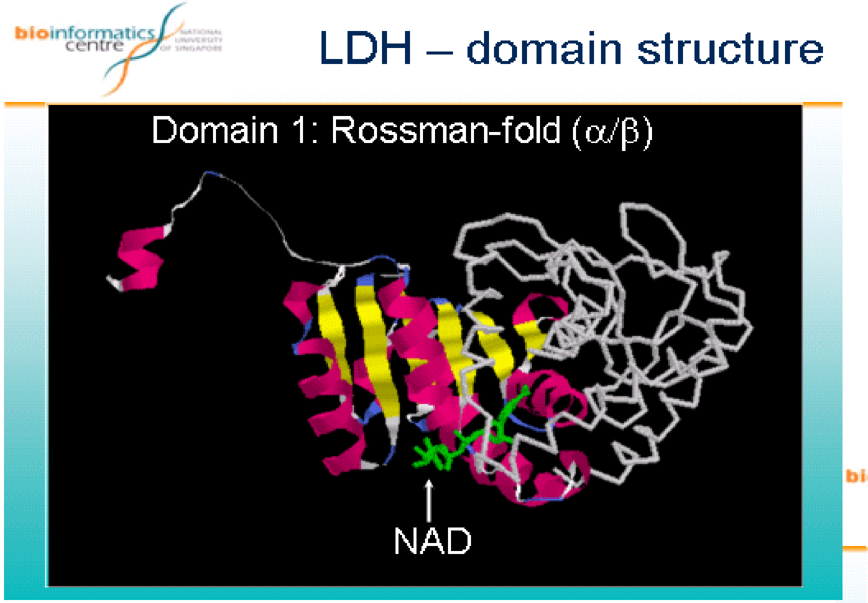
α+β (antiparallel β-sheet)



Most popular class!

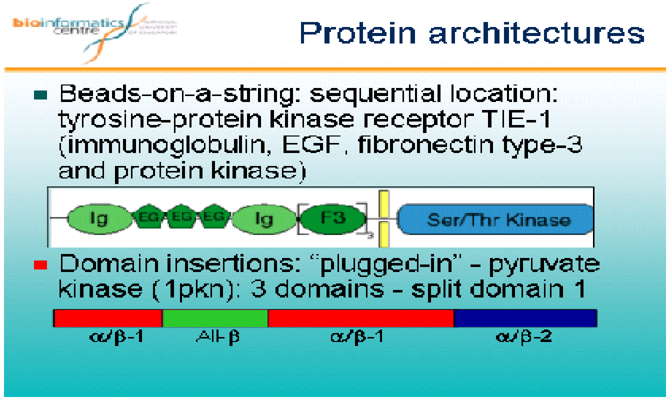
Domain: a LEGO piece

- 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

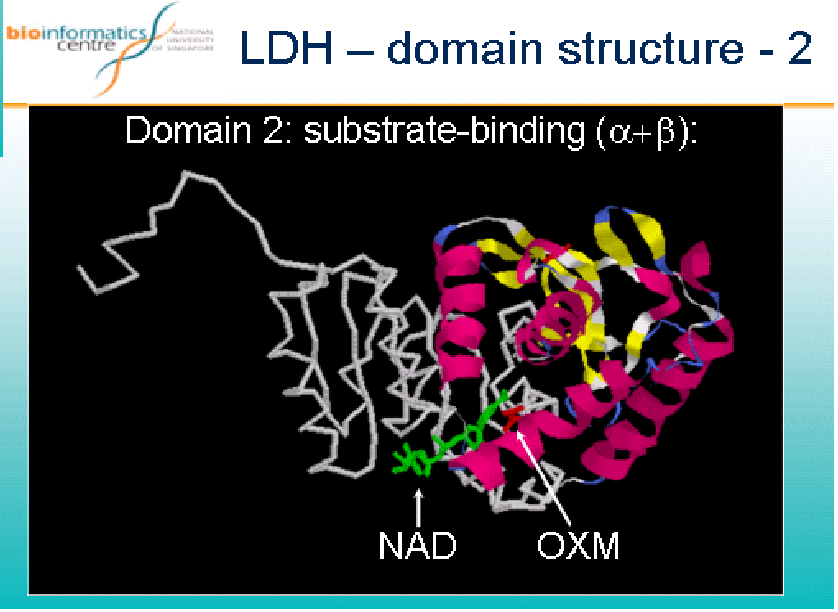
Domain 1: Rossman-fold (α/β)



Protein architectures

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

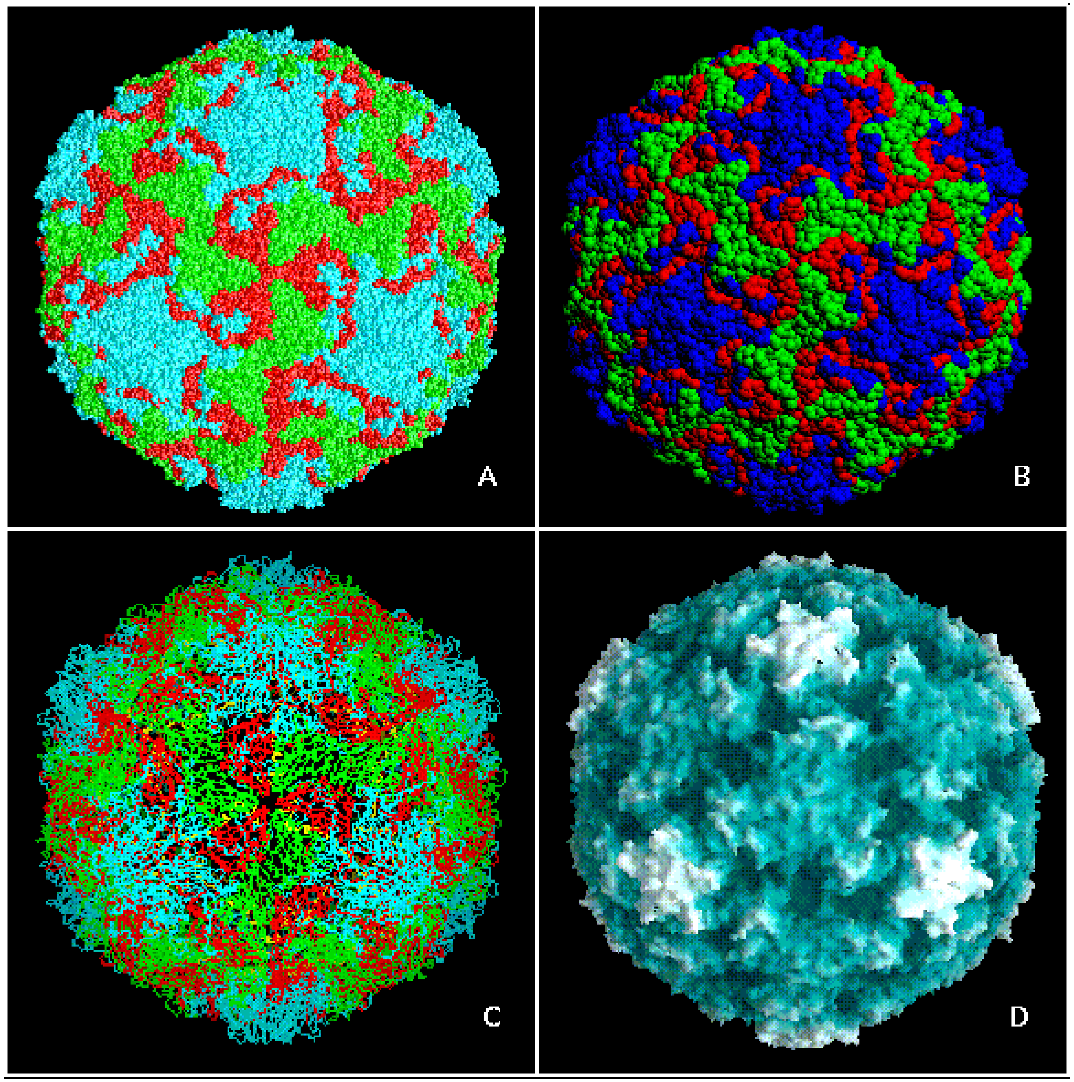
- Domain insertions: "plugged-in" – pyruvate kinase (1pkn): 3 domains – split domain 1

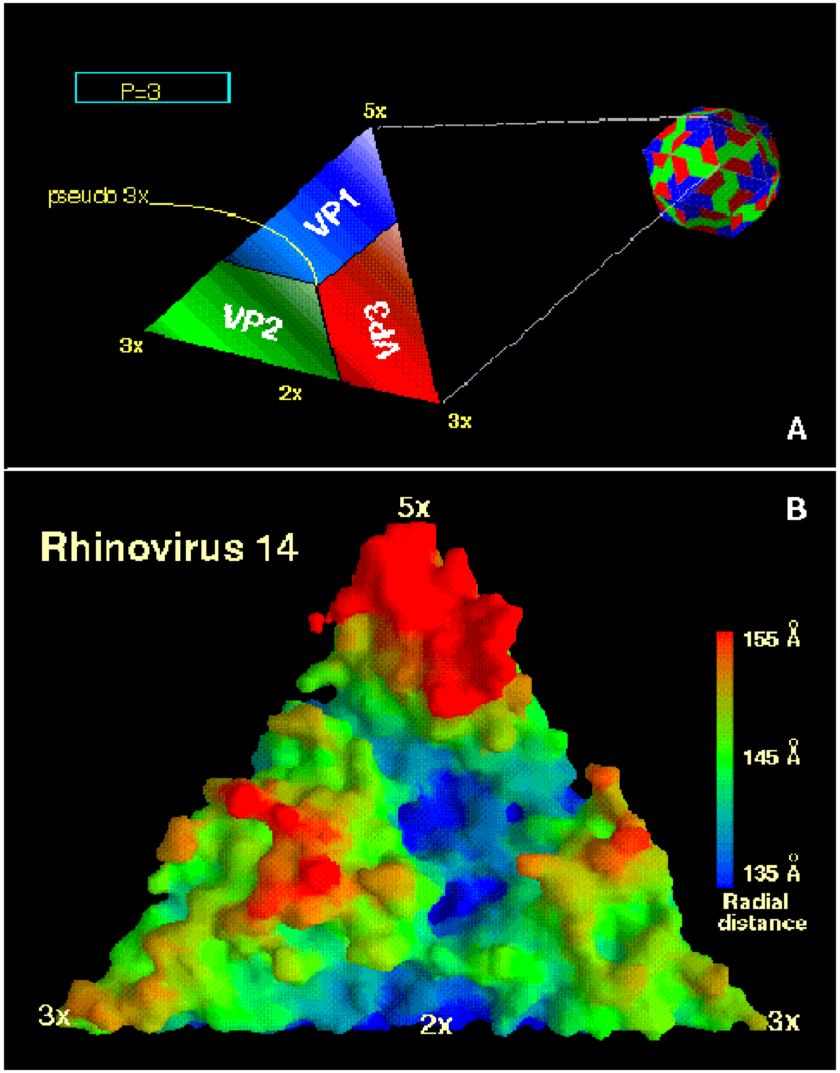


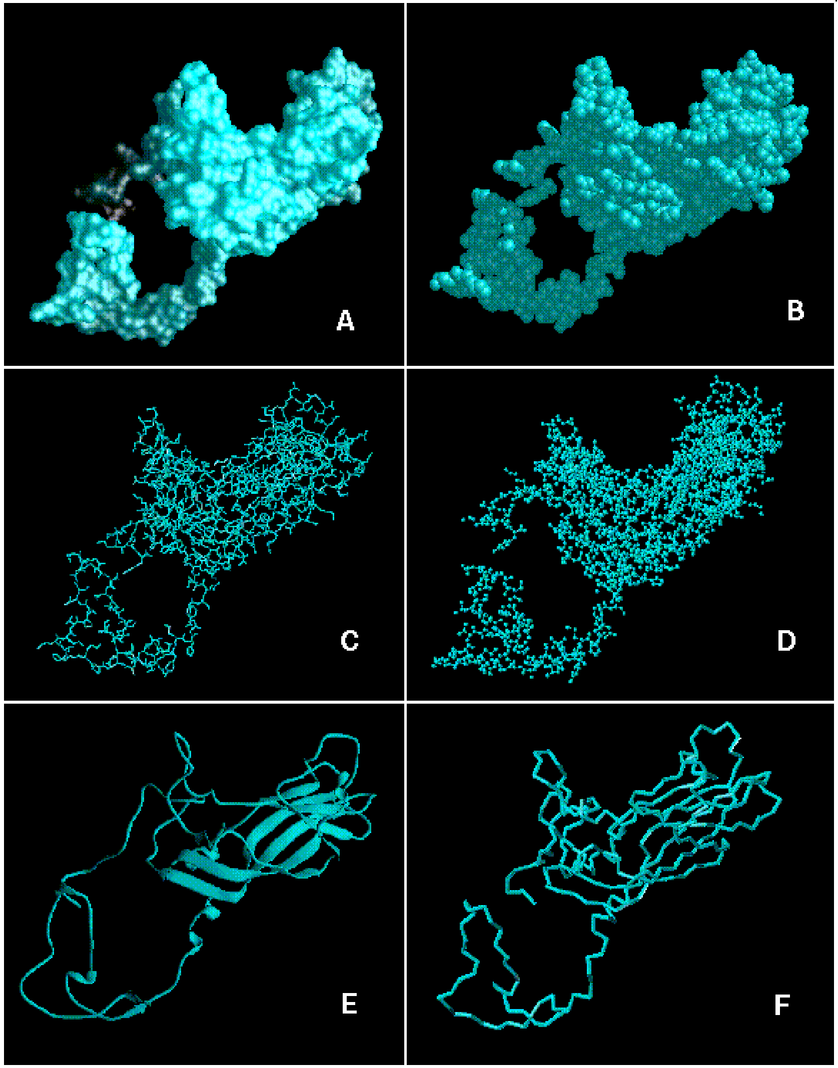
LDH – domain structure - 2

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

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



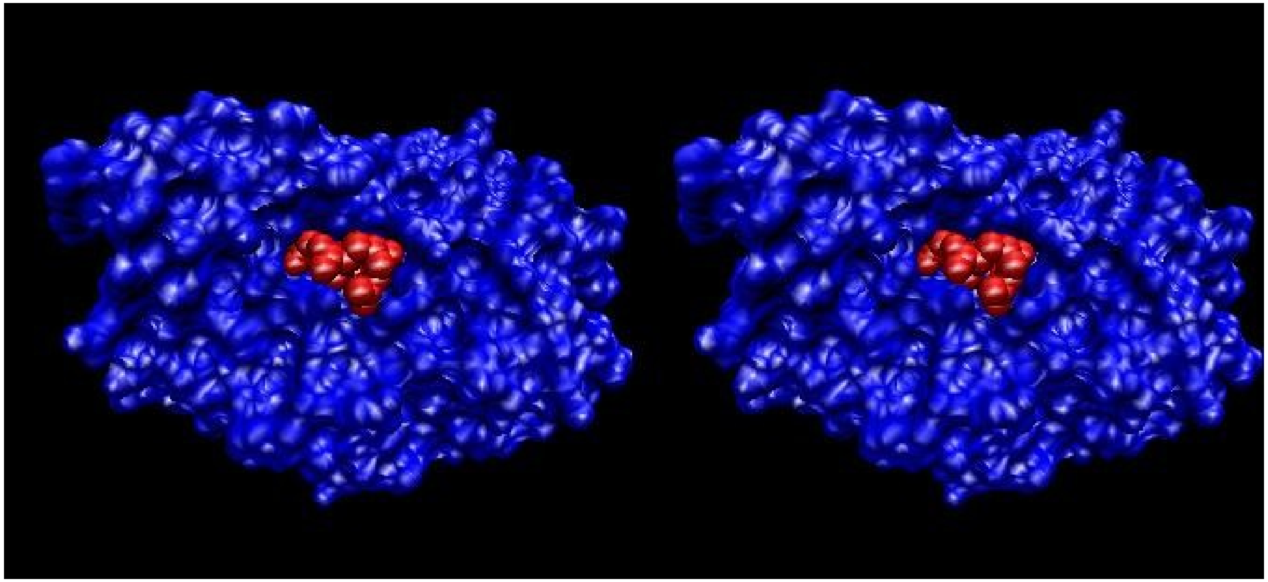

P=3
5x
pseudo 3x
VP1
VP2
VP3
3x
2x
3x
A
B
5x
Rhinovirus 14
155 Å
145 Å
135 Å
Radial distance
3x
2x
3x


A
B
C
D
E
F

Dilci funkce proteinu

- Enzymy (katalyzatory, substrat se promenuje v produkt, aktivni misto)
- Interakce protein-protein
- Interakce protein-DNA
- Interakce protein-ligand
- Transdukce signalu, regulace
- Strukturni proteiny (vlakna, glykoproteiny)
- Motory

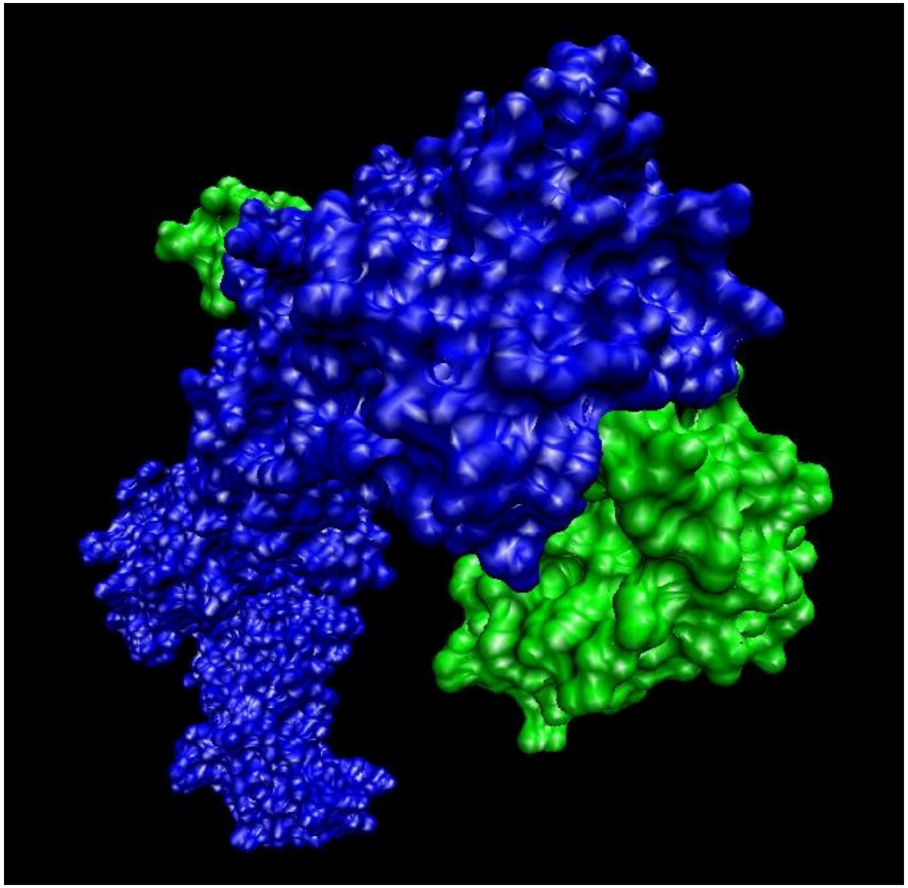
Enzymy (substrat - produkt)



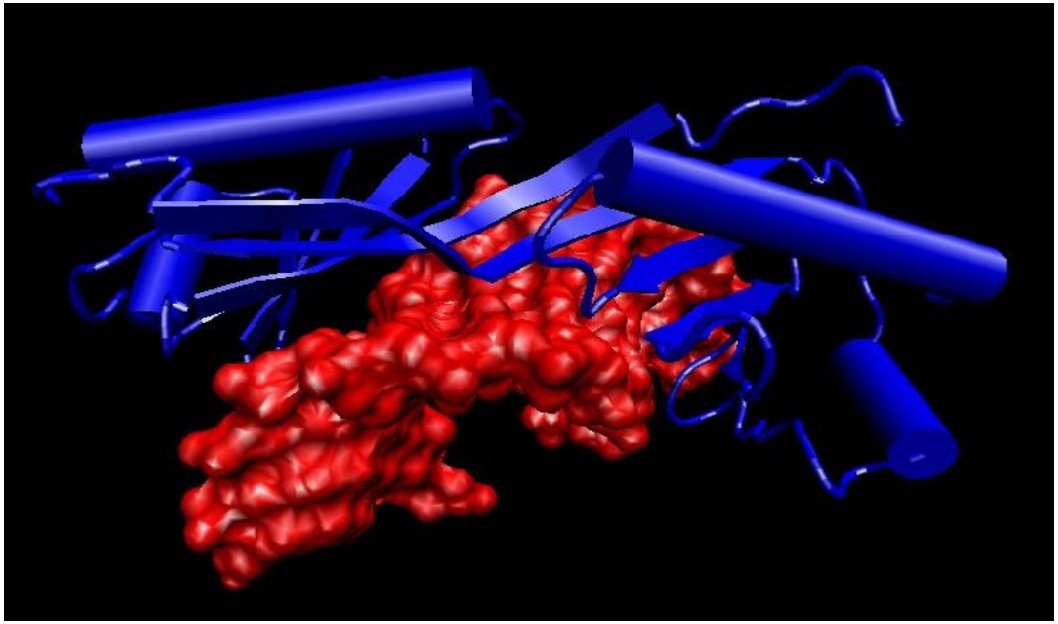
Inhibitor HIV proteazy, která katalyzuje dulezity krok v synteze funkcniho viru zpusobujiciho AIDS

Interakce protein - protein

Interakce mezi protilatkou a antigenem. Takove interakce jsou zakladem humoralni imunity. Diky proteinove pameti muze organizmus likvidovat cizi latky.



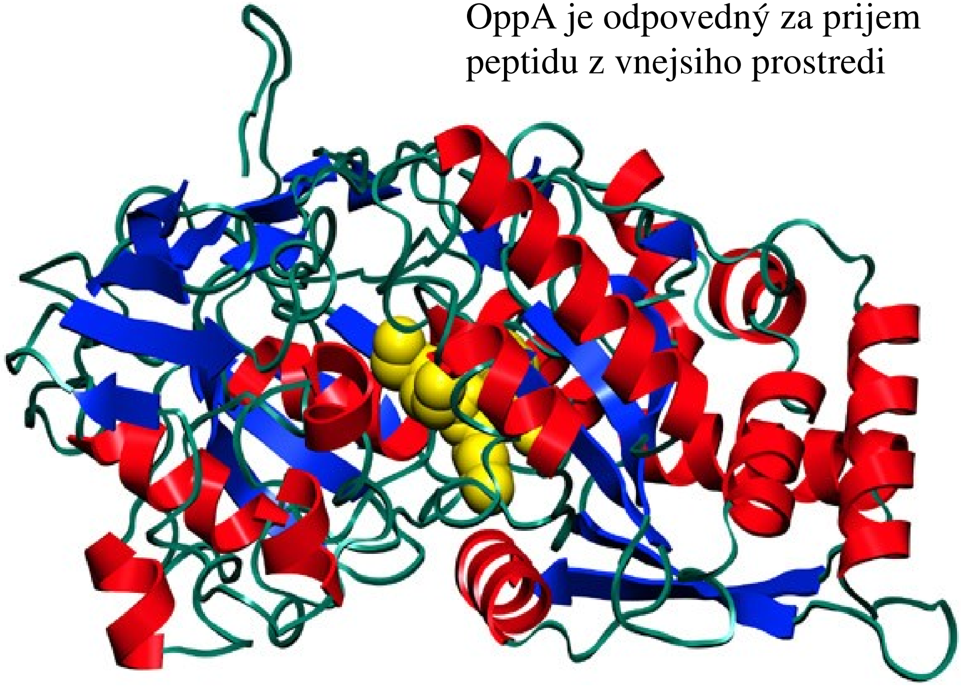
Interakce protein - DNA

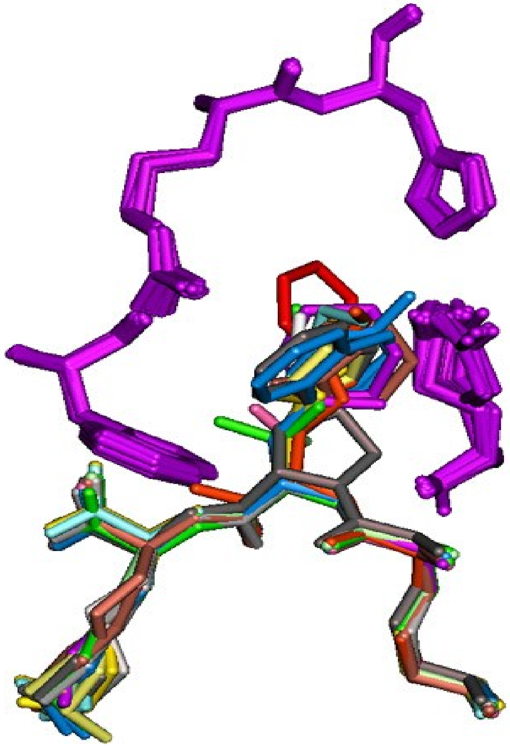


TBP (TATA box binding protein) – dulezita soucast transkripcniho komplexu, kontrolujici polohu, ve ktere transkripce probehne

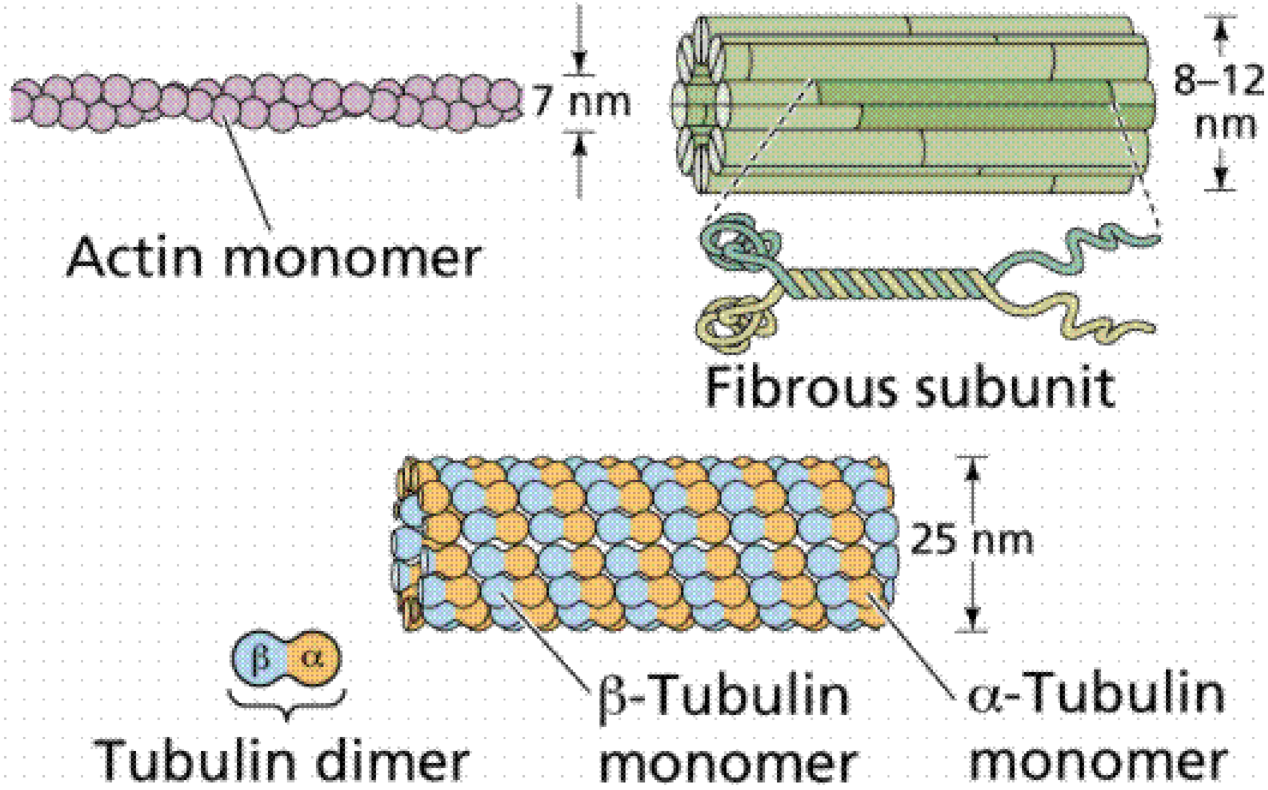
Interakce
protein (OppA) – ligand (KXK)

OppA je odpovedný za prijem
peptidu z vnejsiho prostredi



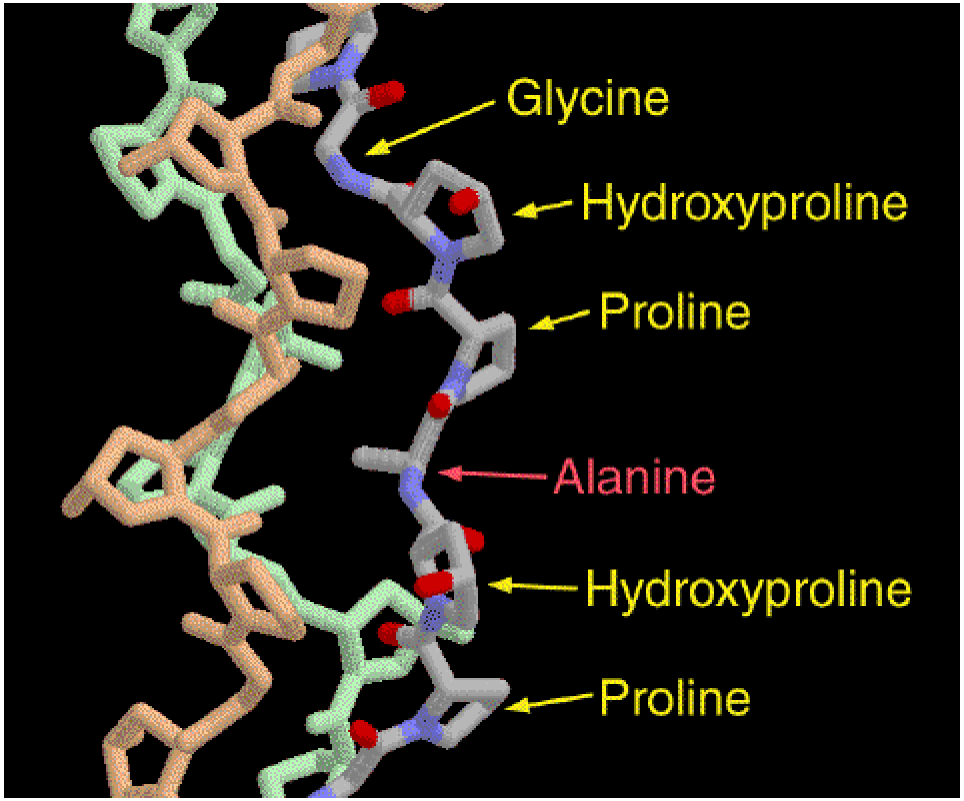


Strukturni proteiny

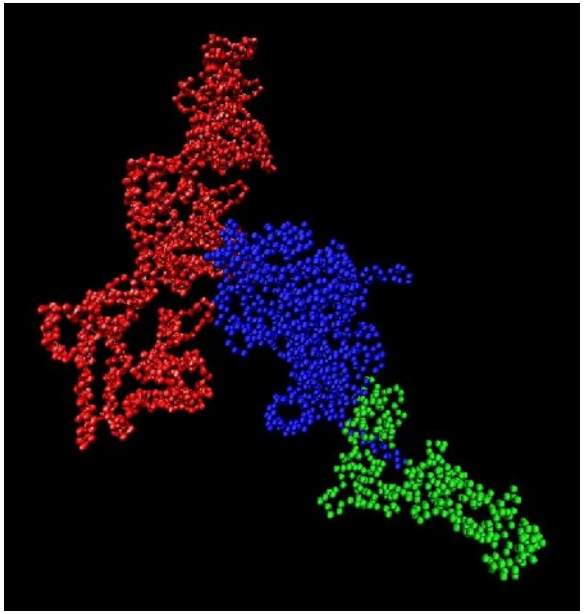


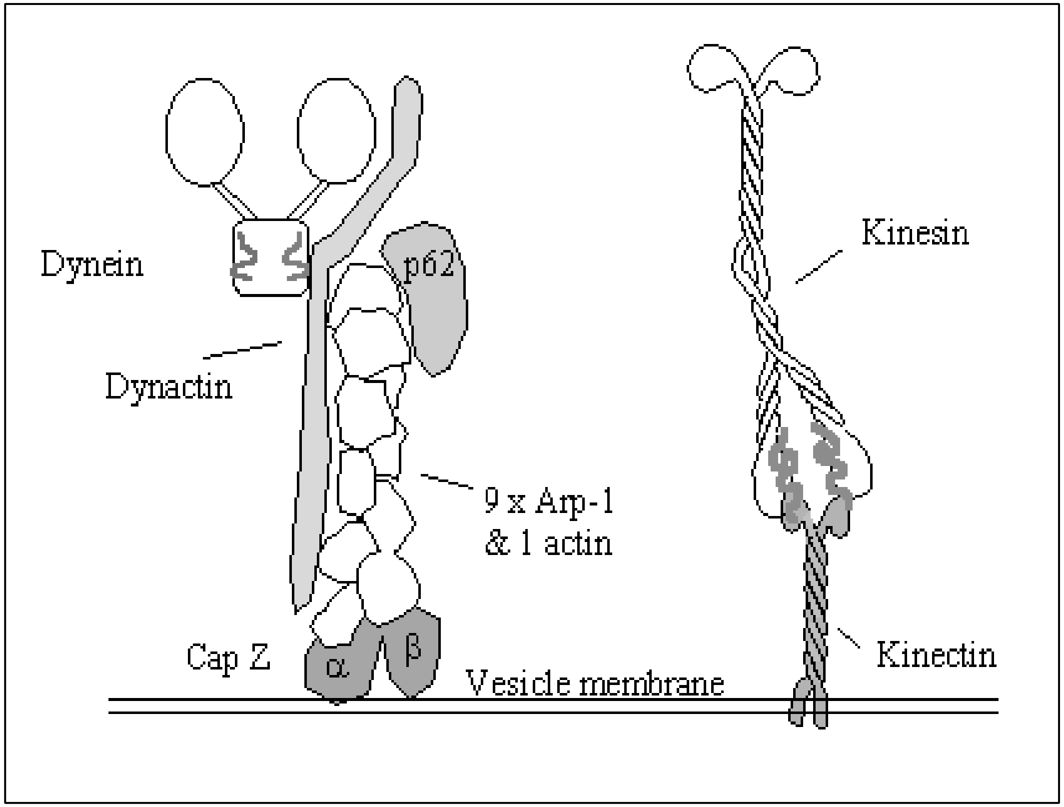
Strukturni proteiny

Kolagen – jeho tvorba vyzaduje vitamin C Tvori kolem 30% proteinu v lidskem tele. Je periodicky (GPP nebo GPK)



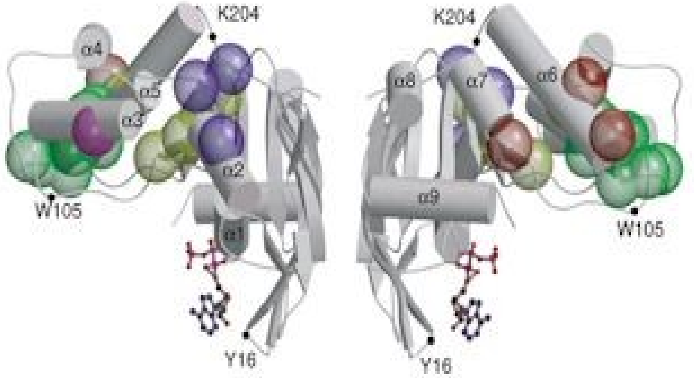
actin - myosin



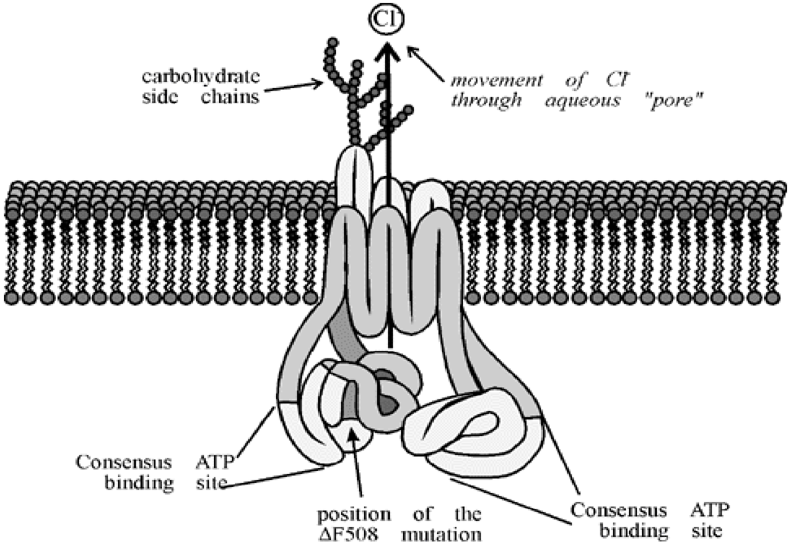


Nemoci zpusobene nespravnym skladanim proteinu

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