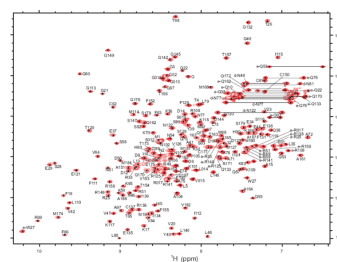
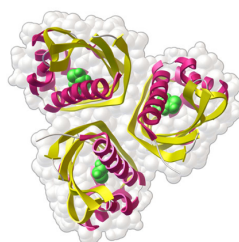
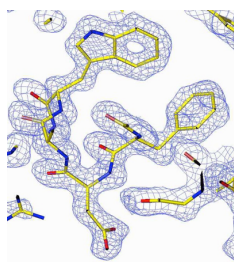
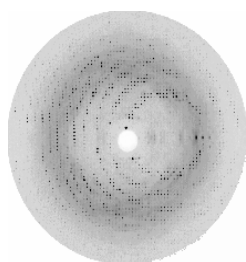


3D struktura makromolekul in vizualizacija struktur

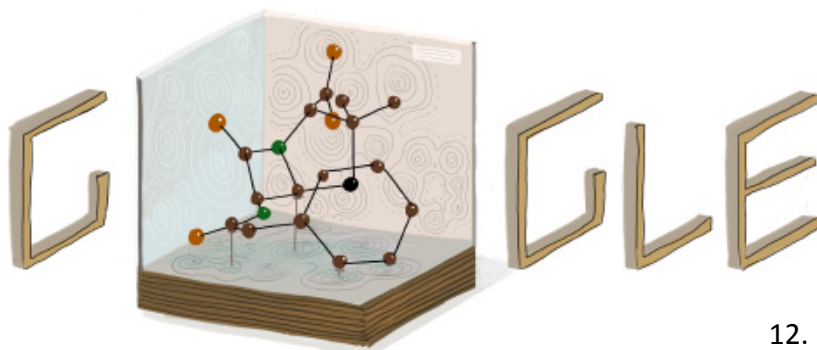


ATOM	1	N	PRO	A	1	8.316	21.206	21.530	1.00	17.44
ATOM	2	CA	PRO	A	1	7.608	20.729	20.336	1.00	17.44
ATOM	3	C	PRO	A	1	8.487	20.707	19.092	1.00	17.44
ATOM	4	O	PRO	A	1	9.466	21.457	19.005	1.00	17.44
ATOM	5	CB	PRO	A	1	6.460	21.723	20.211	1.00	22.26
...										
HETATM	130	C	ACY	401	3.682	22.541	11.236	1.00	21.19	
HETATM	131	O	ACY	401	2.807	23.097	10.553	1.00	21.19	
HETATM	132	OXT	ACY	401	4.306	23.101	12.291	1.00	21.19	

N
C
O
C
C
O
O
O



Miha Pavšič
13. 5. 2014



12. maj 2014



Dorothy Crowfoot Hodgkin
(roj. 12. 5. 1910)

- izpopolnila tehniko rentgenske difrakcije za določevanje strukture molekul
- struktura penicilina in vitamina B₁₂ (Nobelova nagrada za kemijo – 1964)
- struktura inzulina (1969)

3D struktura makromolekul in vizualizacija struktur

Pregled obravnavanih tem:

1. Strukturni podatki (3D), njihova ločljivost in uporabnost

2. 3D struktura bioloških makromolekul

- Proteini
- Nukleinske kisline
- Ogljikovi hidrati

3. Metode za določevanje 3D struktur makromolekul (visoka ločljivost)

- Rentgenska kristalografija (*X-ray*)
- Jedrska magnetna resonanca (NMR)

4. Zapis 3D strukturnih podatkov in podatkovne baze

- Formati PDB, PDBML, mmCIF
- Podatkovna baza Protein Data Bank (PDB)
- Protein Structure Initiative (PSI)

5. Vizualizacija 3D struktur makromolekul

- Osnove
- Načini predstavitev strukturnih podatkov
- Programi

6. Primerjava 3D struktur makromolekul

- Superpozicija in podobnost
- Zvitje in klasifikacija struktur → nadaljevanje naslednjič

1. Strukturni podatki (3D), njihova ločljivost in uporabnost

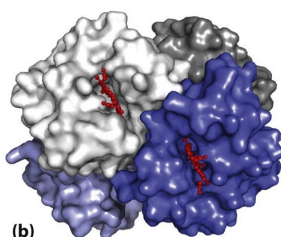
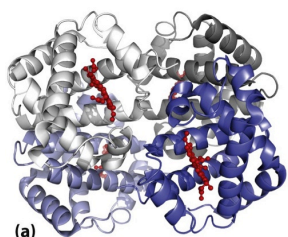
Najenostavnejši zapis strukture: **koordinate (x, y, z) vseh atomov.**

Vendar:

- (makro)molekule so do določene mere fleksibilne/dinamične,
- določene regije (makro)molekul nimajo urejene strukture,
- struktura neke (makro)molekule je lahko v različnih okoljih drugačna (pogoji, interakcija z drugimi molekulami).



Tehnično pravilno je, da uporabljamo izraz **MODEL STRUKTURE**,
a v praksi najpogosteje rečemo kar "struktura".



Na slikah (a) in (b) je **model strukture hemoglobina**, ponavadi pa napišemo kar "**struktura hemoglobina**, prikazana kot trak (a) oz. molekulska površina (b)".

1. Strukturni podatki (3D), njihova ločljivost in uporabnost

Prvi modeli struktur bioloških makromolekul:

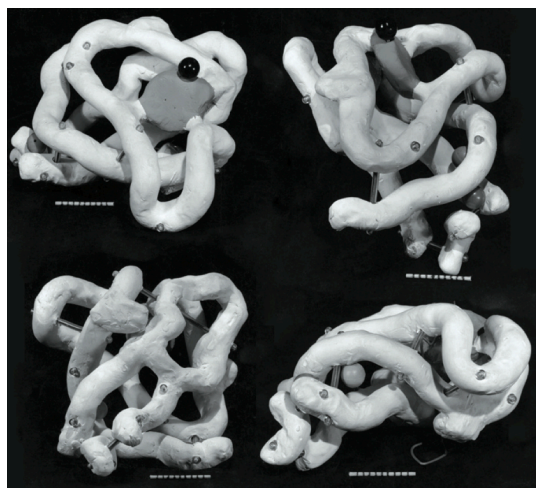


Model strukture DNA

James Watson, Francis Crick & sod.
1953

Relativno visoka ločljivost, ki so jo omogočali:

- poznavanje natančne strukture osnovnih gradnikov,
- enoličnost molekule DNA,
- predpostavljen model podvojevanja ipd.



Model strukture mioglobina

John Kendrew in sod.
1960

Nizka ločljivost!

Struktura gradnikov (aminokislin) je sicer znana, a relativna urejenost atomov in stranskih skupin v prostoru ni tako enolična kot pri DNA.

1. Strukturni podatki (3D), njihova ločljivost in uporabnost

Ločljivost (resolucija, angl. *resolution*)

- S pojmom ločljivost opišemo, katere podrobnosti lahko v modelu strukture zanesljivo razločimo.
- Primer: **ločljivost 2 Å pomeni, da lahko v modelu strukture zanesljivo ločimo pod-strukture, ki so vsaj 2 Å narazen.**

$$1 \text{ \AA} = 10^{-10} \text{ m} = 0,1 \text{ nm}$$

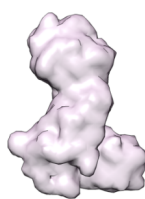
Ångström (po švedskem fiziku - Anders Jonas Ångström)

- Manjša številka – boljša (večja) ločljivost.

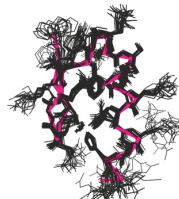
Strukture (in metode, s katerimi jih določamo), delimo na **nizko-** in **visokoločljivostne**:



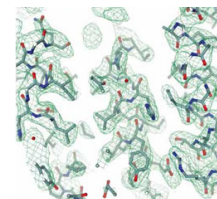
šaperon Mm-cpn
krio-EM tomografija
(cryo EM)



nuclear pore targeting complex
ozkokožno sipanje X-ray žarkov
(SAXS)



jedrsko magnetna
rezonanca
(NMR)



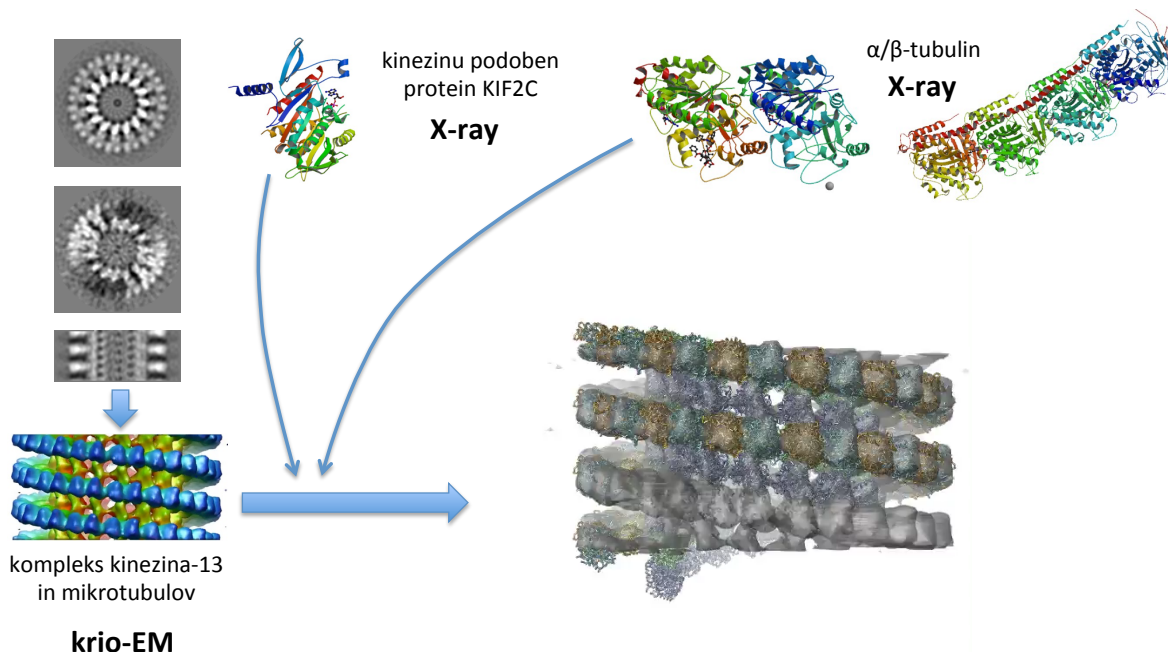
rentgenska
kristalografija
(X-ray crystallography)

+ še nekateri drugi načini pridobivanja strukturnih podatkov (MS, powder diffraction, ...)

1. Strukturni podatki (3D), njihova ločljivost in uporabnost

Pogosto uporabljamo **hibridne modele struktur** (posebej v primeru velikih kompleksov):

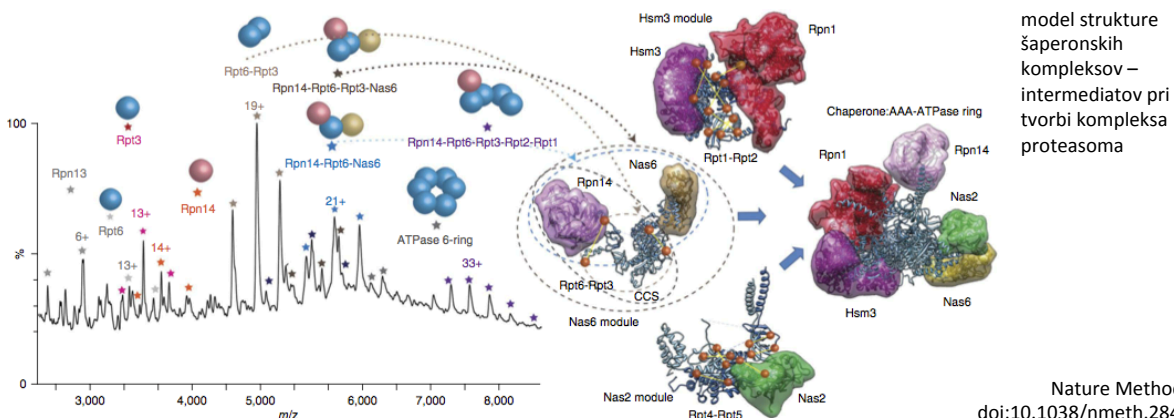
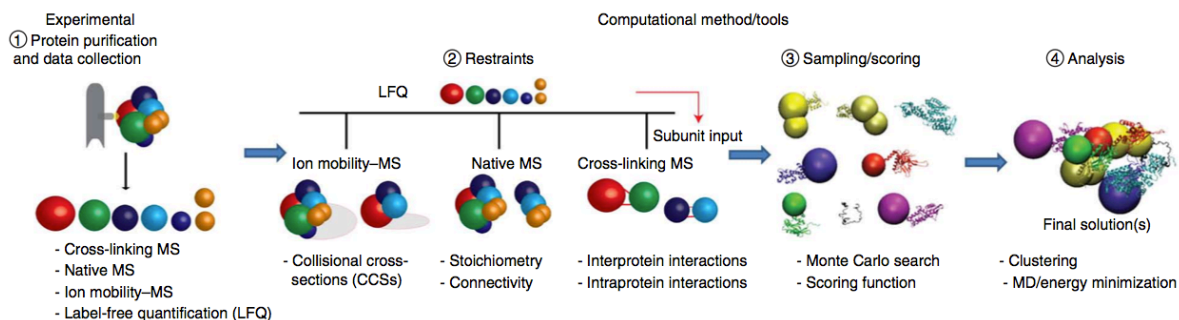
1. strukturo manjših delov (gradnikov kompleksa) določimo pri višji ločljivosti,
2. strukturo večjih kompleksov določimo pri nizki ločljivosti ali jih modeliramo,
3. visokoločljivostne modele gradnikov prilagamo v model celotnih kompleksov.



1. Strukturni podatki (3D), njihova ločljivost in uporabnost

Primer hibridnega pristopa:

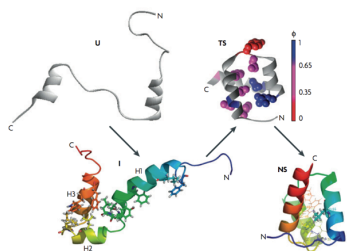
križno povezovanje + MS + LFQ + struktura komponent \rightarrow končen model kompleksa



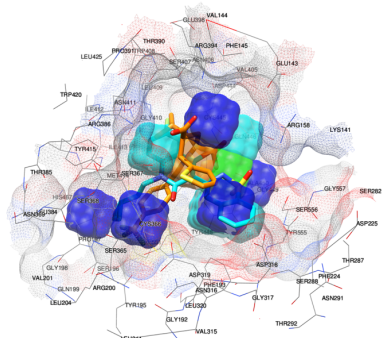
model strukture šaperonskih kompleksov – intermediatov pri tvorbi kompleksa proteasoma

1. Strukturni podatki (3D), njihova ločljivost in uporabnost

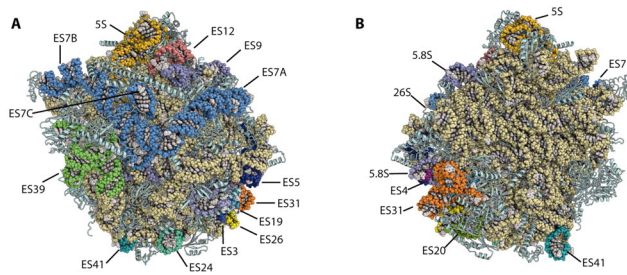
Uporabnost struktur makromolekul:



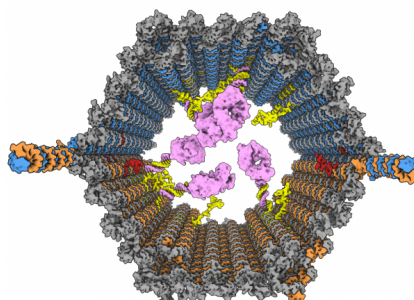
razumevanje zgradbe makromolekul in procesov zvijanja



ciljano načrtovanje novih zdravil in izboljševanje obstoječih



razumevanje delovanja posameznih proteinov (encimi, strukturni proteini, ...), nukleinskih kislin in večjih sistemov (ribosom, jedrna pora, mišični proteini, ...)

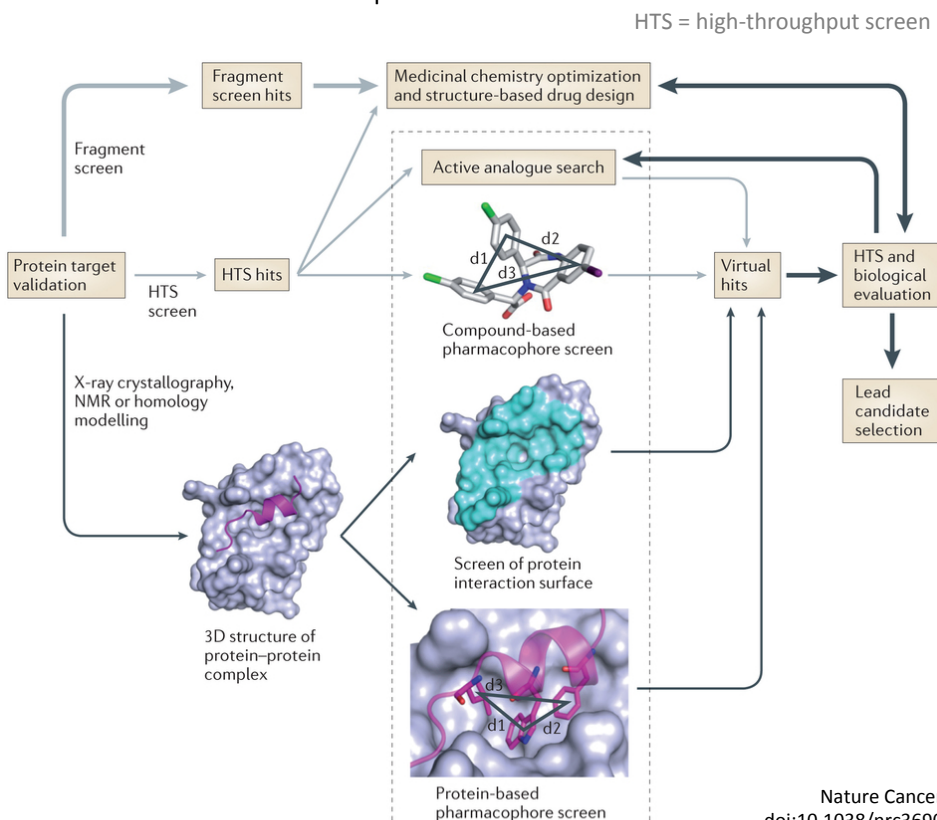
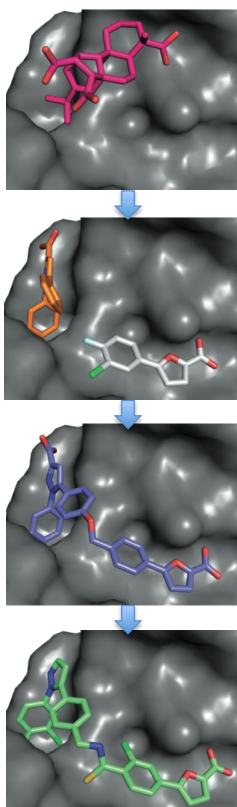


gradnja umetnih encimov, nanostroj/robotov in drugih struktur

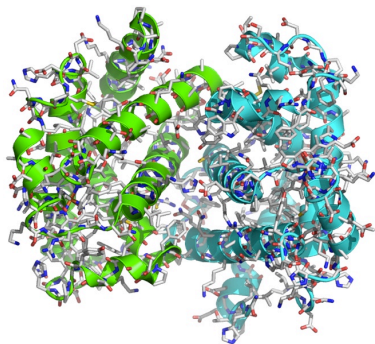
1. Strukturni podatki (3D), njihova ločljivost in uporabnost

Primer: razvoj zdravilnih učinkovin na osnovi strukturnih podatkov.

iskanje najboljprimernejše molekule preko fragmentov (fragment-based screening)



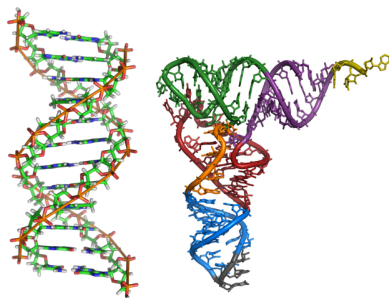
2. 3D struktura bioloških makromolekul



PROTEINI

aminokisliline

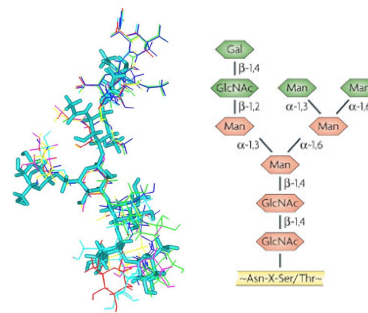
Izredno kompleksna 3D struktura: 20 različnih gradnikov, veliko prostostnih stopenj in možnih medsebojnih interakcij.



NUKLEINSKE KISLINE

nukleotidi

V osnovi zelo enolična zgradba, v kateri najdemo regije sekundarnih struktur → kompleksnejše stabilne 3D strukture (tRNA, rRNA, ...).



POLISAHARIDI

monosaharidi

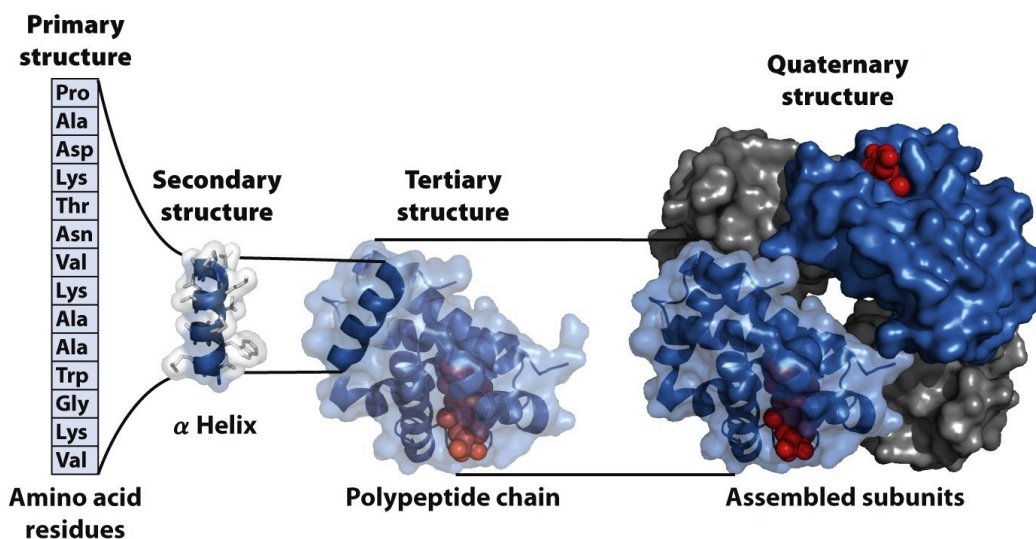
Pogosto zelo fleksibilni, struktura je ponavadi stabilizirana šele ob vezavi/interakciji na npr. protein.

+ sestavljene makromolekule/kompleksi: glikoproteini, proteoglikani, nukleoproteinski kompleksi

2.a. 3D struktura bioloških makromolekul: PROTEINI

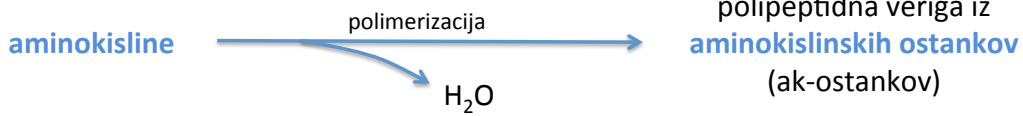
Nivoji strukture proteinov:

- primarna struktura: zaporedje aminokislinskih ostankov
- sekundarna struktura: urejene pod-strukture (vijačnice, nagubane površine)
- terciarna struktura: položaj atomov/podstruktur v prostoru
- kvartarna struktura: podenote oz. kako le-te oblikujejo protein

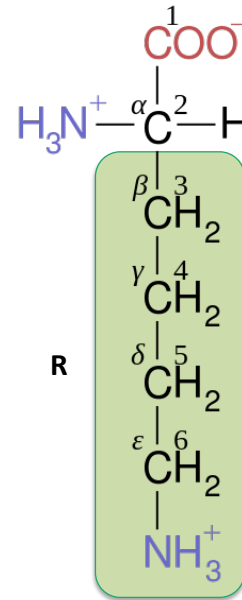
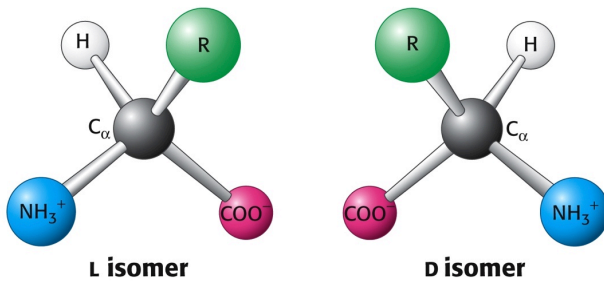


2.a. 3D struktura bioloških makromolekul: PROTEINI

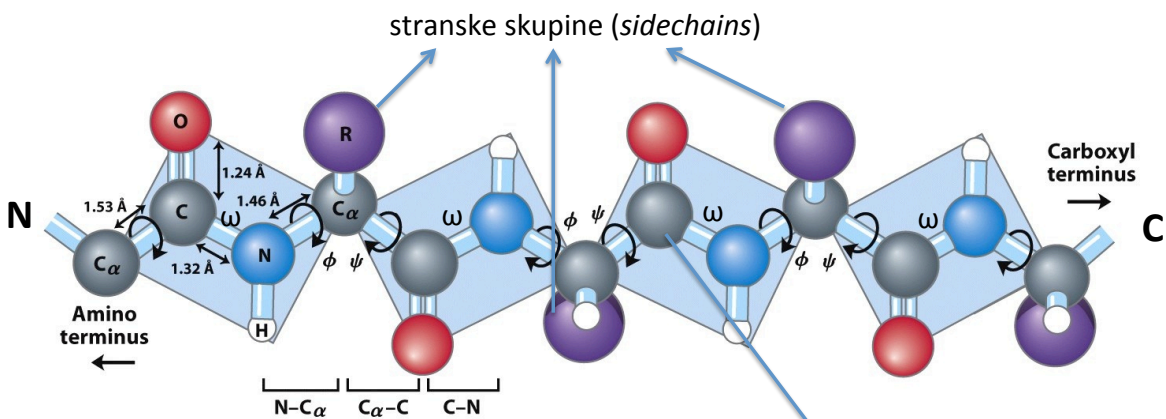
Osnovni gradniki: **aminokisljine**.



V proteinih najdemo L- α -aminokislinske ostanke (D-ak-ostanke le izjemoma, npr. razni mikrobnii peptidi):



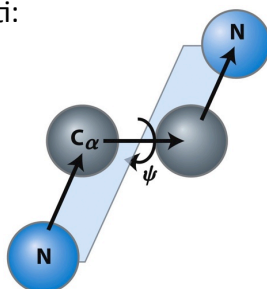
2.a. 3D struktura bioloških makromolekul: PROTEINI



Konformacijo **ogrodja** polipeptidne verige lahko opišemo z **dihedralnimi** koti:

- Φ (fi): C-N-C α -C
- Ψ (psi): N-C α -C-N
- ω (omega): C α -C-N-C α

ω : 180° (*trans* – ponavadi)
ali 0° (*cis* – redko,
predvsem pri Xaa-Pro)



ogrodje, glavna veriga (*backbone, main chain*)
iz atomov C, C α in N

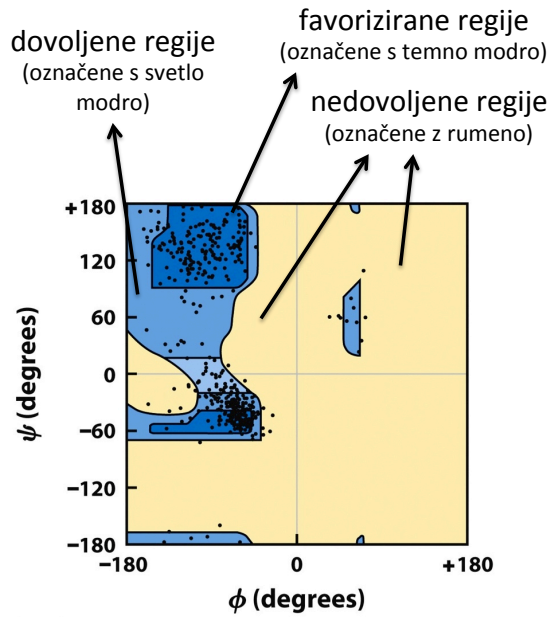
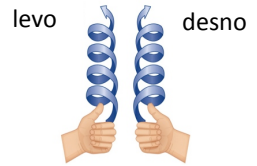
Stranska skupina -
dihedralni koti χ (ksi):
 $\chi_1, \chi_2, \chi_3, \dots$

2.a. 3D struktura bioloških makromolekul: PROTEINI

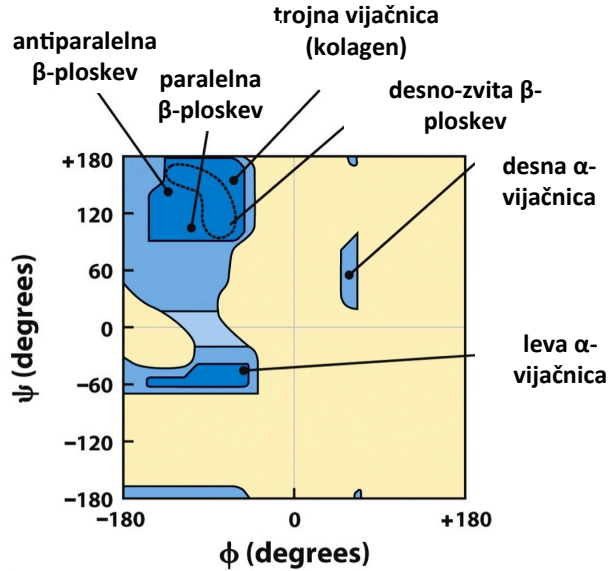
Ramachandranov diagram



Gopalasamudram
Narayana Iyer
Ramachandran
(Indija)



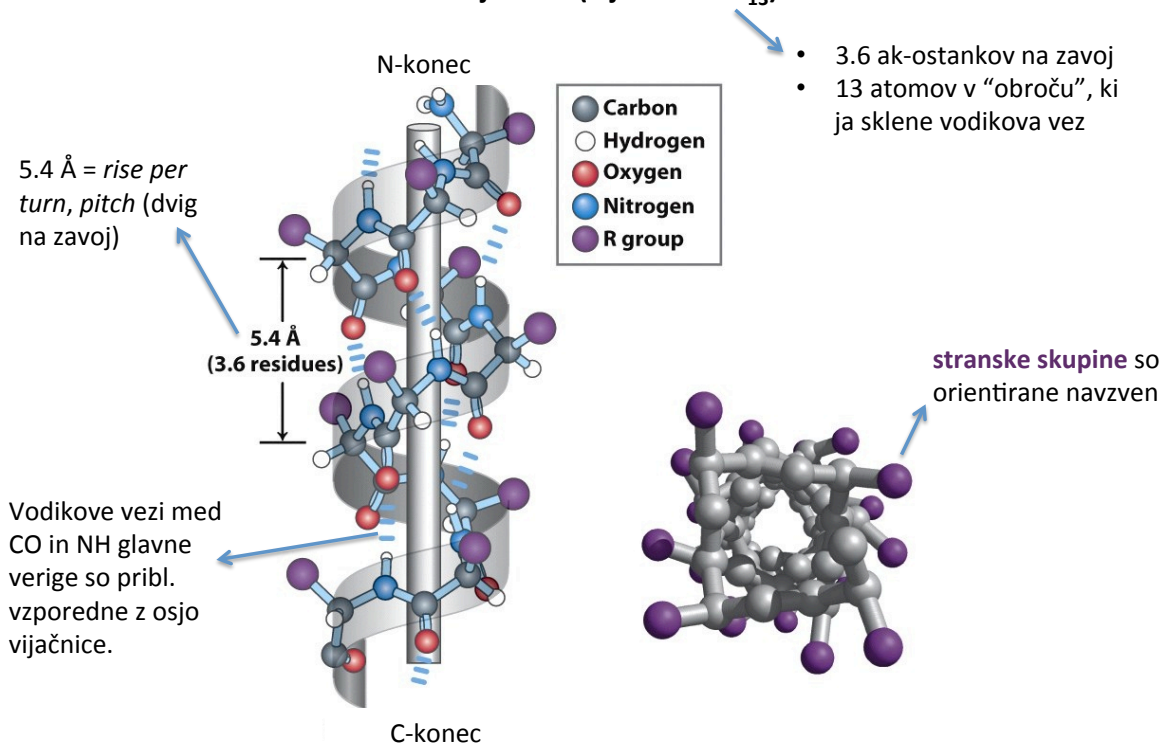
Populacija Φ in Ψ kotov v proteinih.



Za elemente sekundarne strukture je značilna kombinacija določenih vrednosti Φ in Ψ .

2.a. 3D struktura bioloških makromolekul: PROTEINI

Elementi sekundarne strukture: α -vijačnica (vijačnica 3.6_{13})

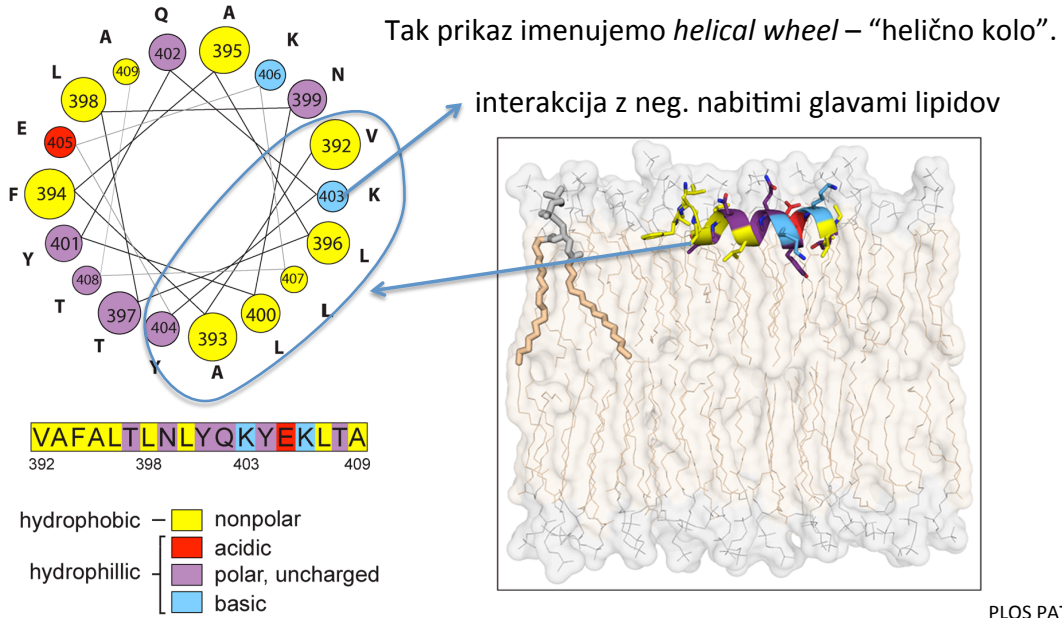


Poznamo še druge tipe vijačnic (npr. vijačnica 3_{10} (ožja kot α) in π (širša kot α)), ki pa se pojavljajo redkeje.

2.a. 3D struktura bioloških makromolekul: PROTEINI

Vijačnice - amfipatičnost

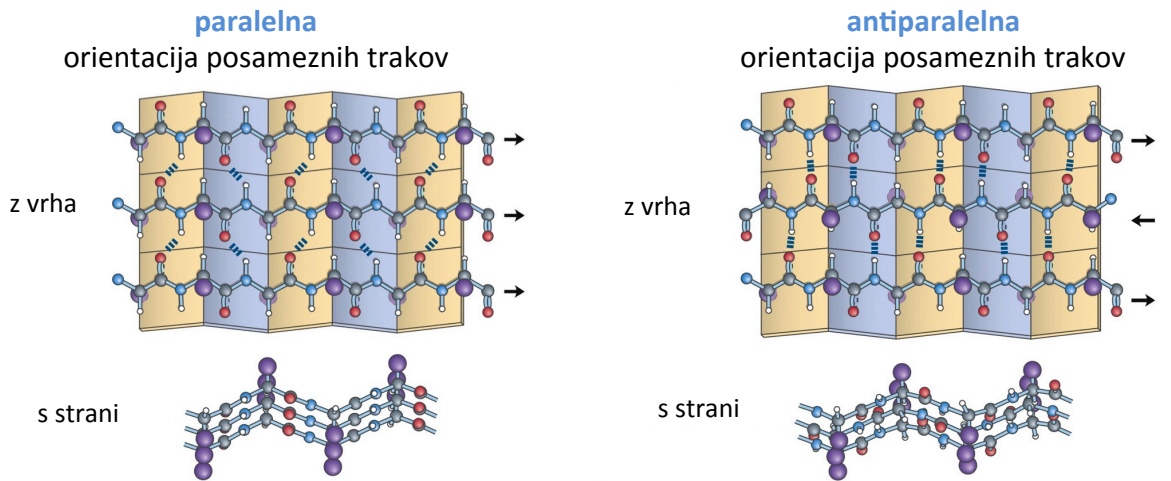
Vijačnice imajo lahko na eni strani izražen hidrofoben značaj (npr. v stiku s hidrofobno notranjostjo proteina ali sredico membrane), na drugi strani pa polarnega (npr. v stiku s topilom, notranjost hidrofилnega kanalčka).



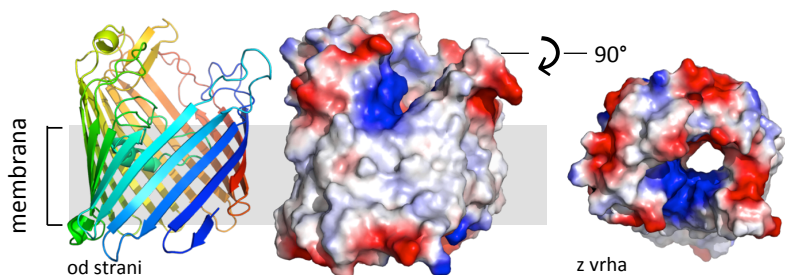
PLOS PATHOGENS
DOI: 10.1371/journal.ppat.1000351

2.a. 3D struktura bioloških makromolekul: PROTEINI

Elementi sekundarne strukture: β -površina



Podobno kot heliksi je tudi β -površina lahko **amfipatična**, primer so porini – izmenjujoči akostanki s polarnimi in hidrofobnimi str. skupinami.



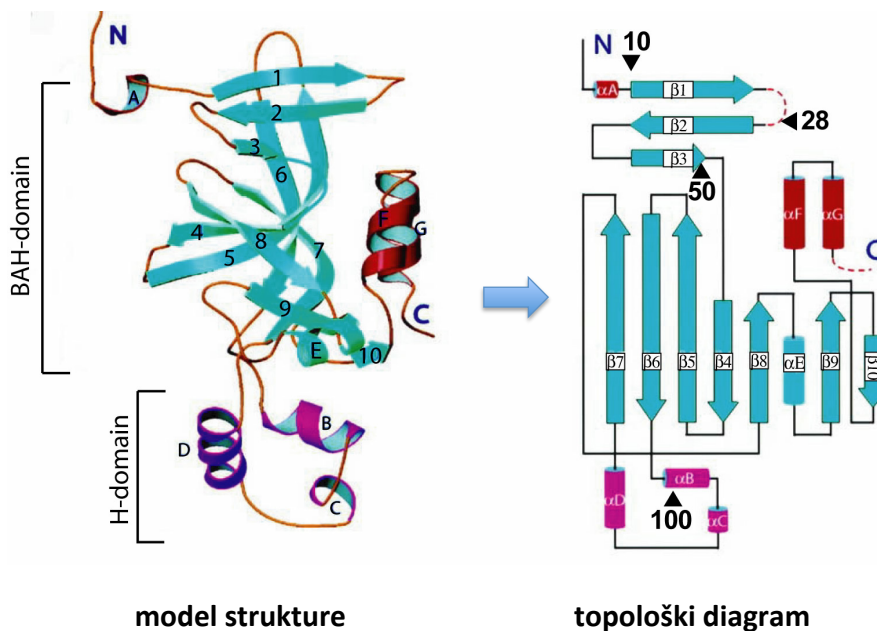
2.a. 3D struktura bioloških makromolekul: PROTEINI

Prikaz elementov sekundarne strukture skupaj z aminokislinskim zaporedjem (analiza 3D strukture):

Strežnik **PDBsum** - analiza struktur iz baze ali prenesene strukture.

2.a. 3D struktura bioloških makromolekul: PROTEINI

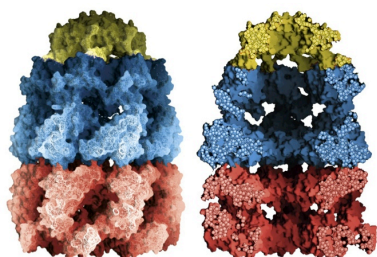
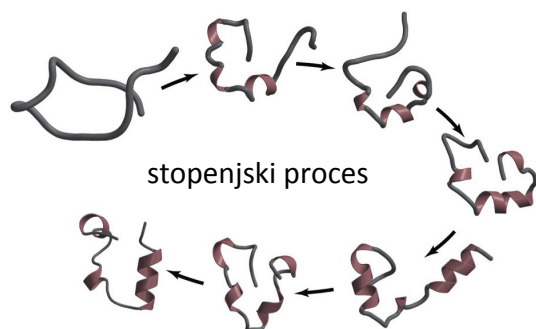
Topologija – prikaz elementov sekundarne strukture – kako si sledijo in kako so organizirani.



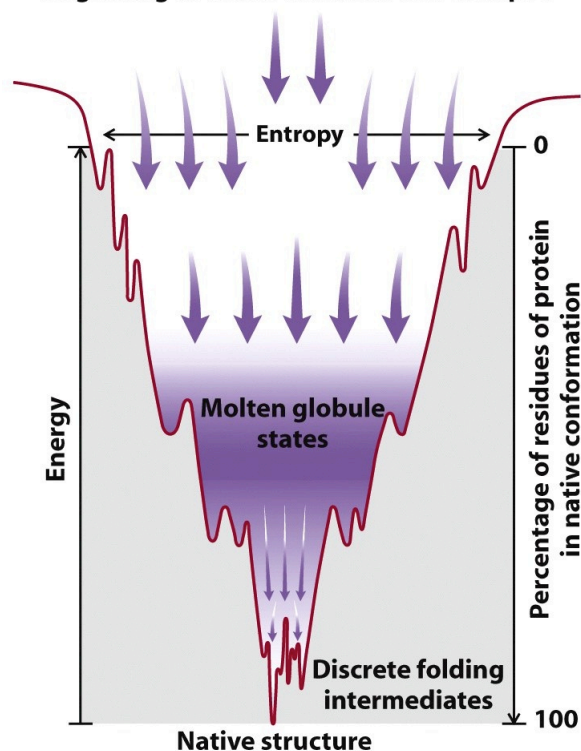
- trakovi kot puščice (smer polipeptidne verige!)
- vijačnice kot cilindri
- dolžina simbolov je sorazmerna št. ak-ostankov v posameznem elementu sekundarne strukture

2.a. 3D struktura bioloških makromolekul: PROTEINI

Zvijanje proteinov (*folding*)



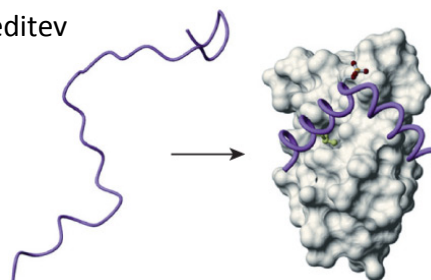
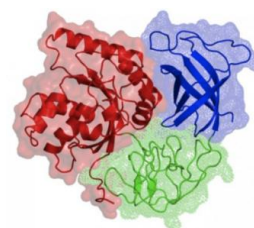
Beginning of helix formation and collapse



2.a. 3D struktura bioloških makromolekul: PROTEINI

Nekatere druge splošne značilnosti proteinov/proteinskih struktur:

- hidrofobna notranjost, hidrofilna zunanost (razen npr. pri delih, ki so v stiku s hidrofobno sredico membran ipd.)
- dodatna stabilizacija strukture z **disulfidnimi vezmi** (oksidativno okolje!)
- polipeptidna lahko tvori več enakih ali različnih **domen** (neodvisno zvijanje domen), stiki med domenami so lahko polarni ali hidrofobni
- nekatere regije polipeptidne verige (še posebej na površini) so lahko **neurejene** (ureditev npr. ob vezavi drugega proteina)
- nekateri proteini so pretežno neurejeni (**intrinzično neurejeni proteini** – *intrinsically disordered proteins, IDPs*); ureditev strukture šele ob vezavi na drugi molekulo



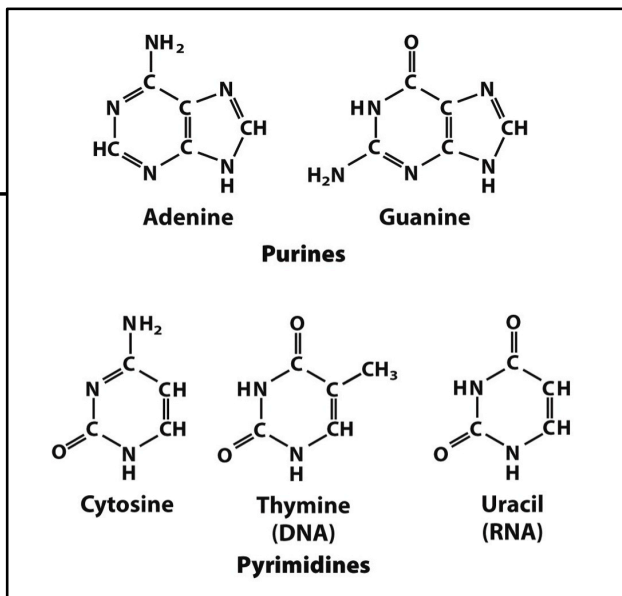
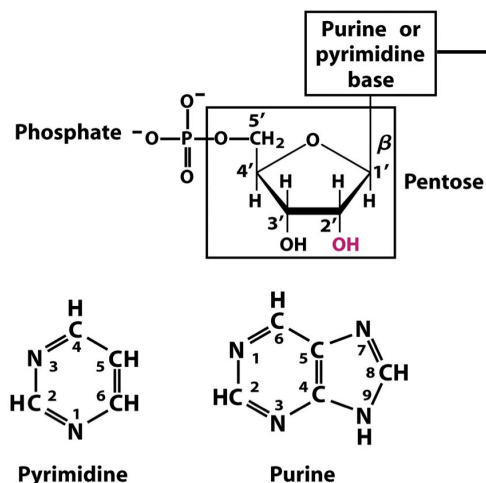
2.b. 3D struktura bioloških makromolekul: NUKLEINSKE KISLINE

Nivoji strukture:

- **primarna** struktura: zaporedje nukleotidov
- **sekundarna** struktura: interakcije med bazami, npr. A=T
- **terciarna** struktura: položaj atomov/podstruktur v prostoru
- **kvartarna** struktura: interakcije med nukleinskimi kislinami (npr. v ribosomu), kromatin, ...

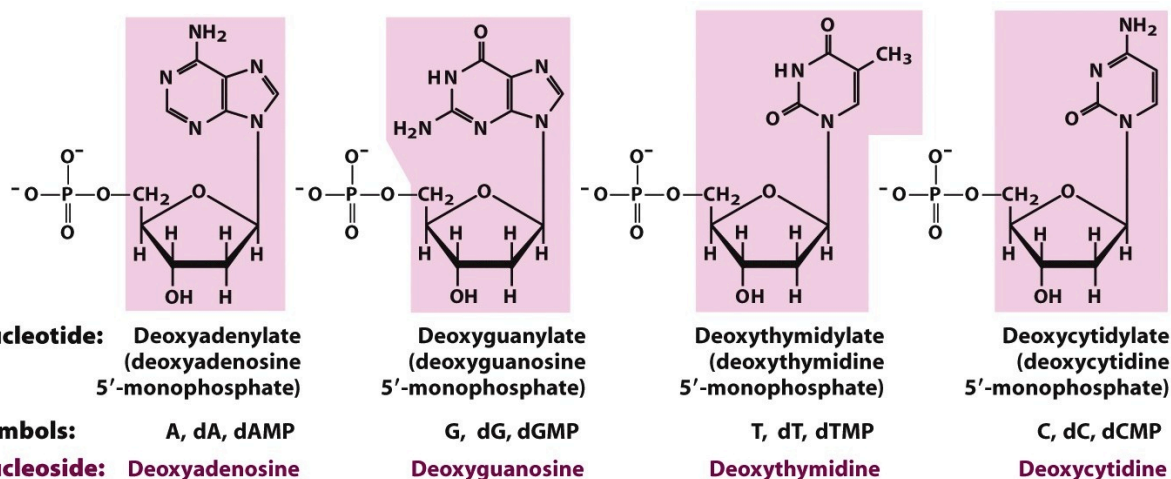
Osnovni gradniki - nukleotidi

nukleozid = baza + sladkorni ostanek



2.b. 3D struktura bioloških makromolekul: NUKLEINSKE KISLINE

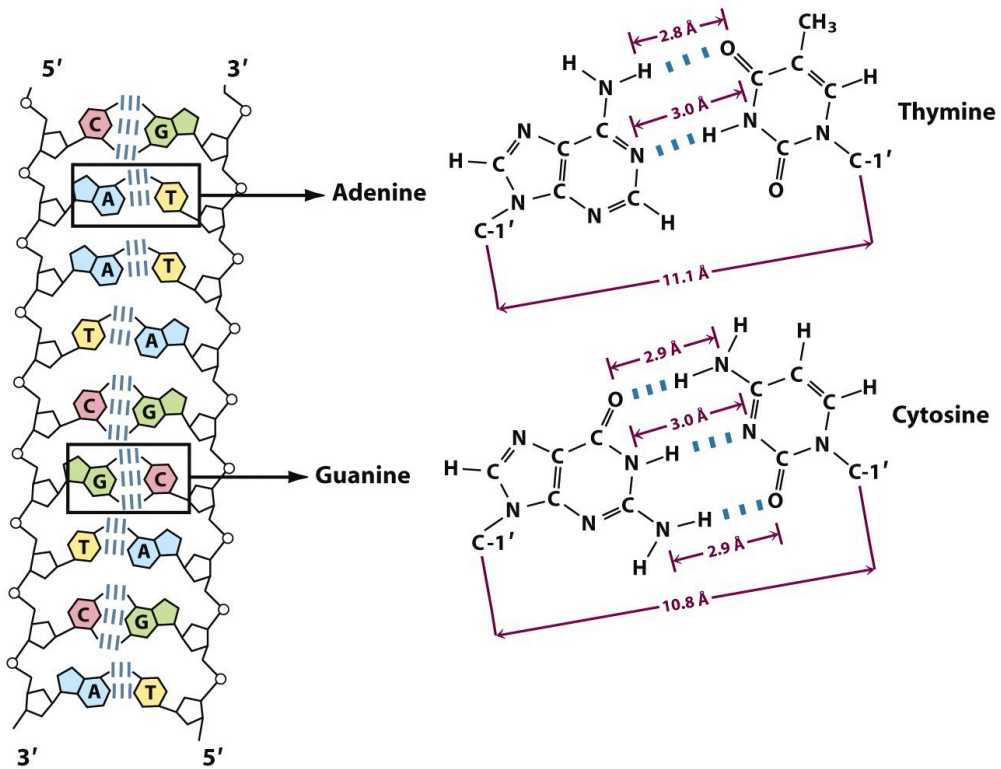
Osnovni gradniki - nukleotidi



Deoxyribonucleotides

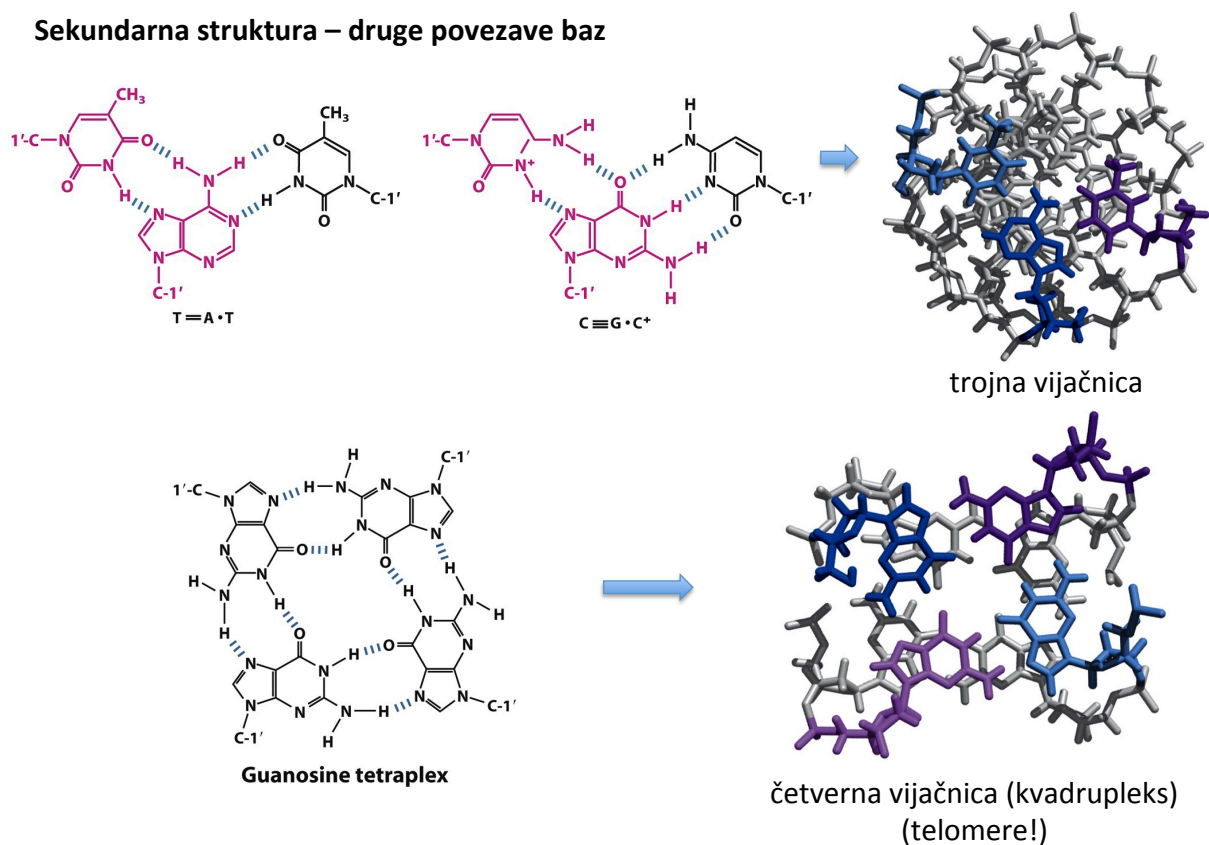
2.b. 3D struktura bioloških makromolekul: NUKLEINSKE KISLINE

Sekundarna struktura – osnovni bazni pari



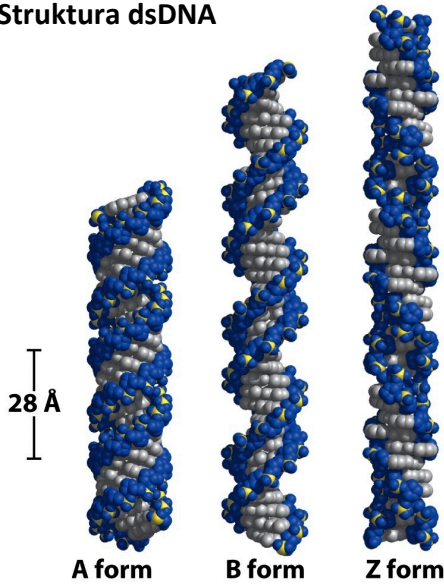
2.b. 3D struktura bioloških makromolekul: NUKLEINSKE KISLINE

Sekundarna struktura – druge povezave baz



2.b. 3D struktura bioloških makromolekul: NUKLEINSKE KISLINE

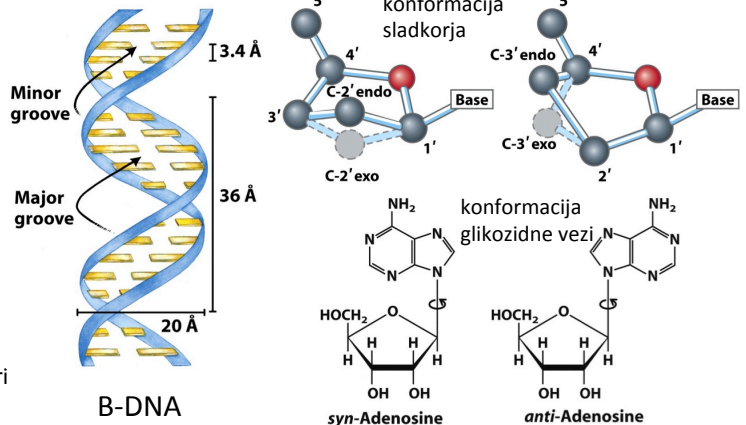
Struktura dsDNA



Glede na okolje in zaporedje nukleotidov:

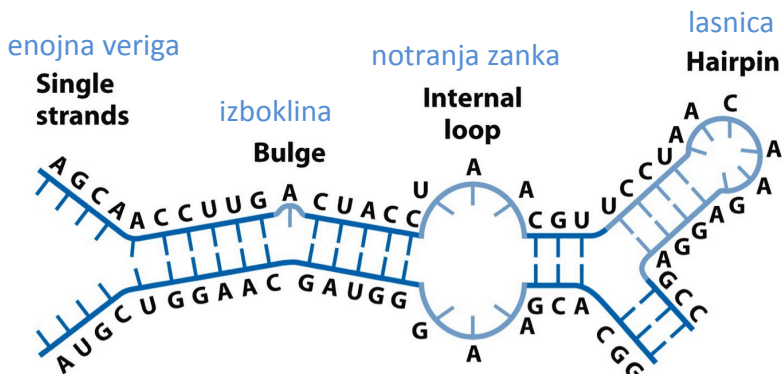
- oblika B je običajna,
- obliko Z zavzame verigi z zaporedjem izmenjujočih purinskih in pirimidinskih nukleotidov, npr. GCGCGCGC,
- oblika A pri RNA-ssDNA in RNA-RNA ter pri dehidrirani dsDNA.

	A form	B form	Z form
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10.5	12
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Sugar pucker conformation	C-3' endo	C-2' endo	C-2' endo for pyrimidines; C-3' endo for purines
Glycosyl bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

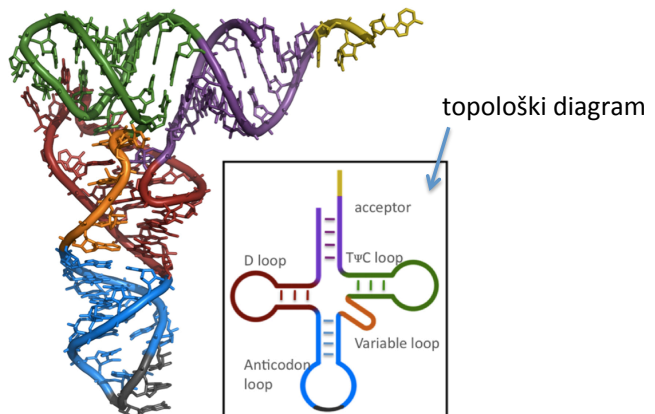


2.b. 3D struktura bioloških makromolekul: NUKLEINSKE KISLINE

Sekundarna struktura RNA



Terciarna struktura RNA – npr. tRNA



2.b. 3D struktura bioloških makromolekul: NUKLEINSKE KISLINE

Struktura evkariontskega ribosoma (60S+40S + eIF1 in eIF6)
(nukleoproteinski kompleks: rRNA + proteini)
Neposreden dokaz, da je ribosom **RIBOCIM**.

Crystal Structure of the Eukaryotic 60S Ribosomal Subunit in Complex with Initiation Factor 6

Sebastian Klinge,* Felix Voigts-Hoffmann,* Marc Leibundgut, Sofia Arpagaus, Nenad Ban†

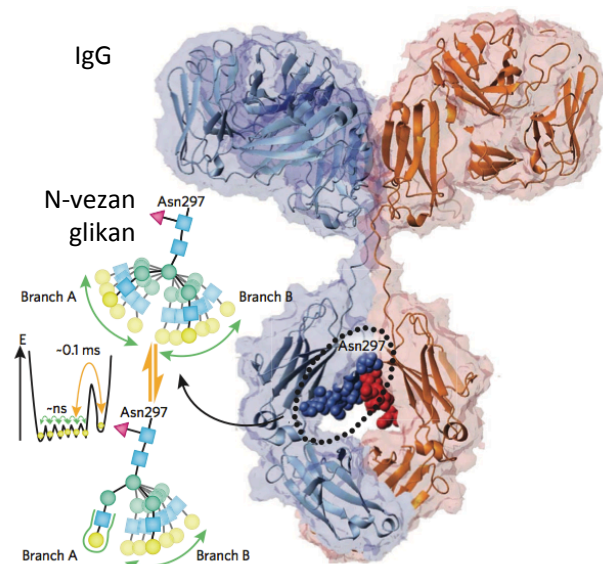
2.c. 3D struktura bioloških makromolekul: OGLJIKOVI HIDRATI

Ogljikovi hidrati (oligo- in polisaharidi) nastopajo kot makromolekule:

- **samostojno** (npr. škrob, celuloza, ...)
- v povezavi s proteini (vedno na površini proteinov, saj se dodajo po zvijanju):
 - **glikoproteini** (krajše verige, navadno vezane na Asn)
 - **proteoglikani** (daljše nabite verige, navadno vezane na Ser – glikozaminoglikani (GAG))

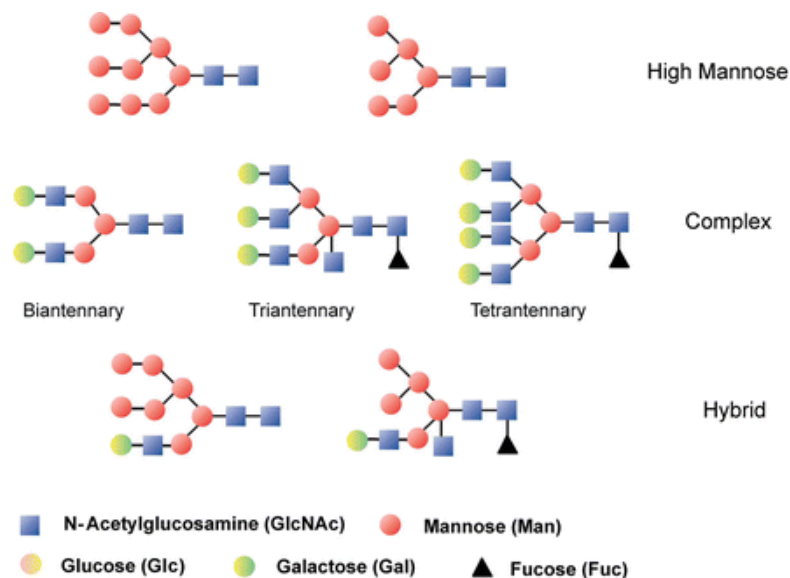
Struktura ogljikovih hidratov je navadno preveč "neurejena", zato je ne moremo določiti oz. podati nekega statičnega modela:

- v eksperimentalno določenih strukturah glikoproteinov in proteoglikanov "vidimo" ponavadi samo prvega ali še nekaj sladkornih ostankov naprej od mesta pritrditve,
- strukturo polisaharidnih verig se lahko uredi ob vezavi npr. na protein.



2.c. 3D struktura bioloških makromolekul: OGLJIKOVI HIDRATI

Ponavadi stukturo opišemo v smislu zaporedja sladkornih ostankov in topologije, ki jo določimo z masno spektroskopijo in cepitvijo s specifičnimi endo- in eksoglikozidazami.

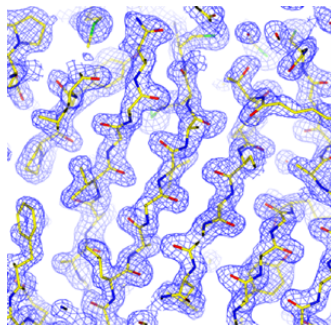


3. Metode za določevanje 3D struktur makromolekul (visoka ločljivost)

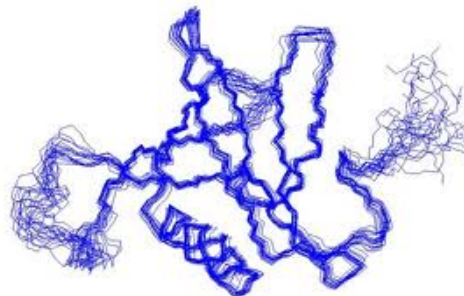
Za določevanje struktur makromolekul pri visoki ločljivosti se trenutno uporabljata dve metodi:

- rentgenska kristalografija (*X-ray crystallography*)
- jedrska magnetna resonanca (*Nuclear Magnetic Resonance, NMR*)

Nekateri med te metode prištevajo tudi krio-EM.



del elektronske gostote proteina, določene z rentgensko kristalografijo

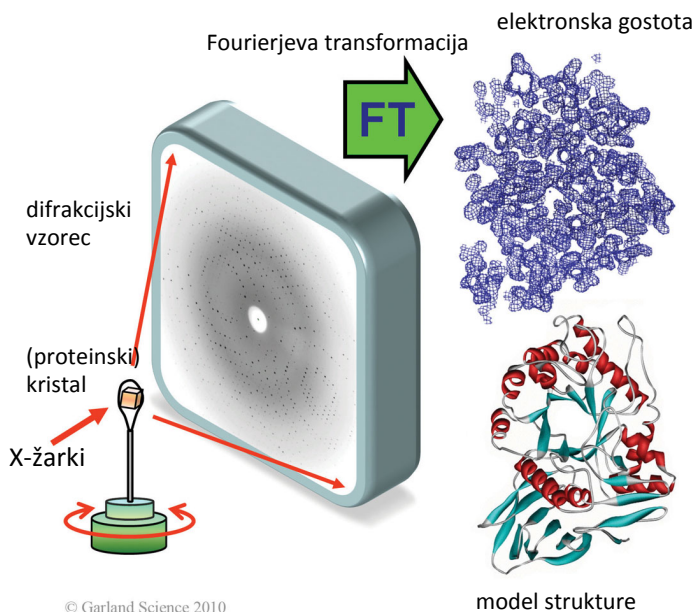


modeli proteina, ki (vsi!) ustrezajo rezultatom eksperimenta z NMR

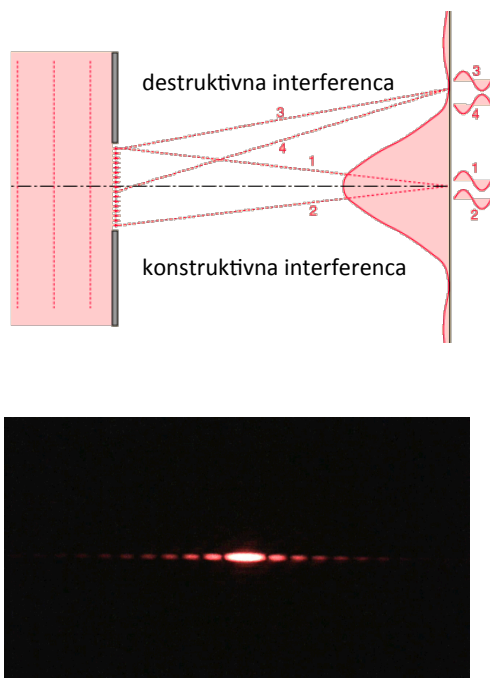
X-ray in NMR sta KOMPLEMENTARNI metodi – vsaka ima svoje prednosti in slabosti, združeni rezultati, pridobljeni z obema metodama, pa dajo popolnoma nov vpogled v strukturo in delovanje makromolekul.

3.a. Metode za določevanje 3D struktur makromolekul (visoka ločljivost): X-ray

Metoda temelji na sipanju rentgenskih žarkov na atomih v kristalu, iz vzorca sipanih žarkov pa lahko rekonstruiramo elektronsko gostoto.



analogija z vidno svetlobo - nastanek interferenčnega vzorca



3.a. Metode za določevanje 3D struktur makromolekul (visoka ločljivost): X-ray

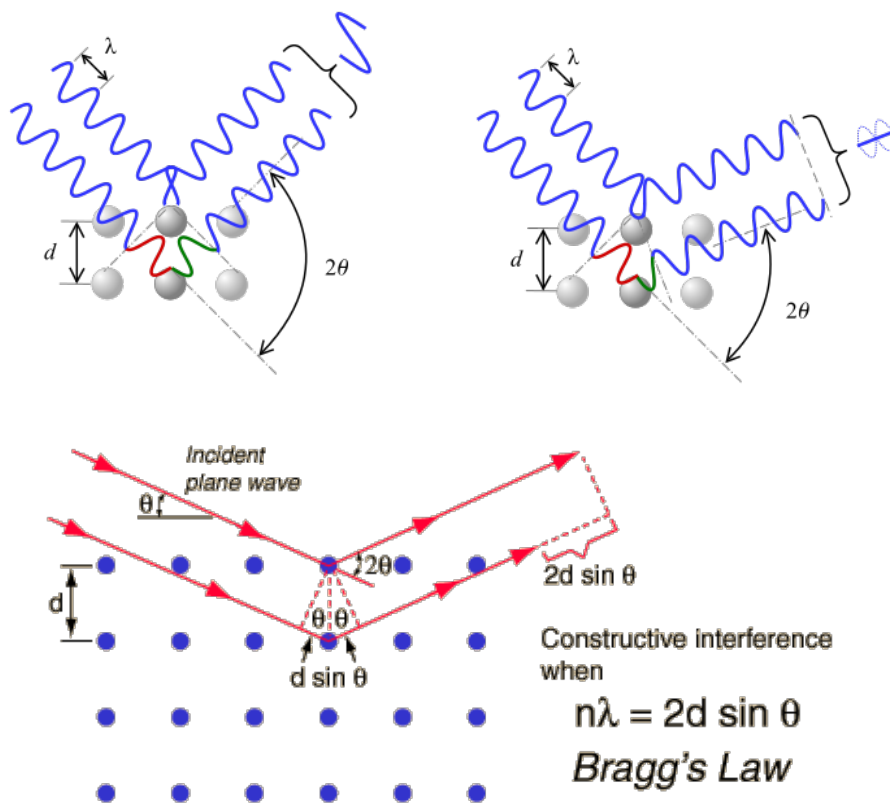
Braggov zakon

Opisuje sipanje svetlobe na kristalni mreži.

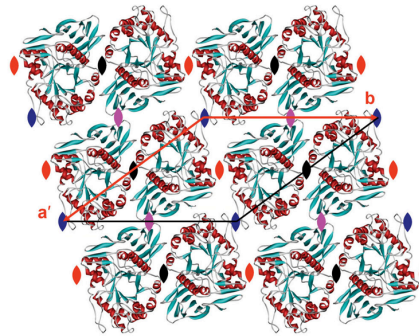
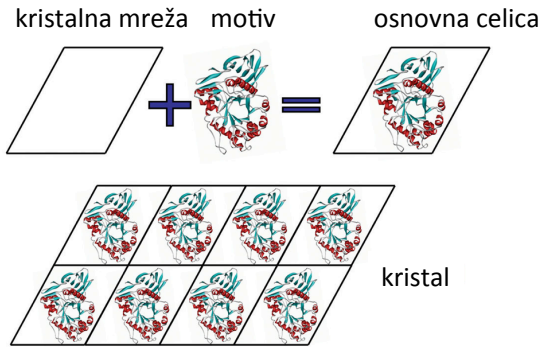
Do sipanja pride, če je valovna dolžina svetlobe (λ) ranga razdalj med ravninami (d).

Če želimo proučevati atome, moramo uporabiti svetlobo valovne dolžine primerljive z razdaljo med atomi (1-2 Å).

žarki X



3.a. Metode za določanje 3D struktur makromolekul (visoka ločljivost): X-ray

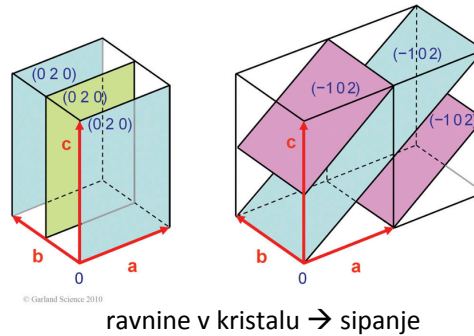
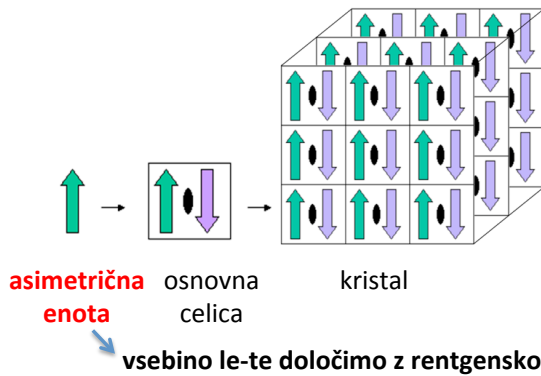


© Garland Science 2010

© Garland Science 2010

zlaganje molekul v kristal

molekule v kristalu so v medsebojnem stiku



© Garland Science 2010

3.a. Metode za določanje 3D struktur makromolekul (visoka ločljivost): X-ray

Rekonstrukcija elektronske gostote iz difrakcijskega vzorca – amplitud in faz

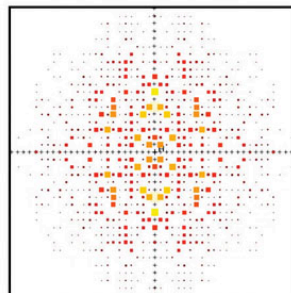
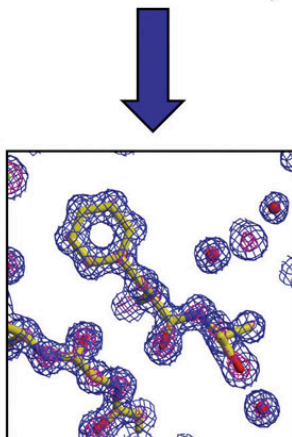
h, k, l – indeksi ravnin

elektronska gostota

$$\rho(x, y, z) = \frac{1}{V} \sum_{-h}^h \sum_{-k}^k \sum_{-l}^l F_{hkl} \cdot \exp[-2\pi i(hx + ky + lz - \alpha_{hkl})]$$

amplituda strukturnih faktorjev (intenziteta lis – povezano s kristalno mrežo)

faza strukturnih faktorjev (povezana s položajem atomov)



FAZNI PROBLEM

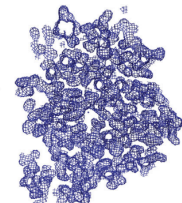
?

Za rešitev je na voljo več metod.

h	k	l	F(hkl)	φ(hkl)
2	0	0	228.0	180.0
1	0	1	10.4	90.0
2	0	1	901.8	270.0
1	1	1	367.0	332.1
1	2	3	149.3	37.8
8	9	1	97.9	255.1
7	7	2	111.5	139.7

FT →

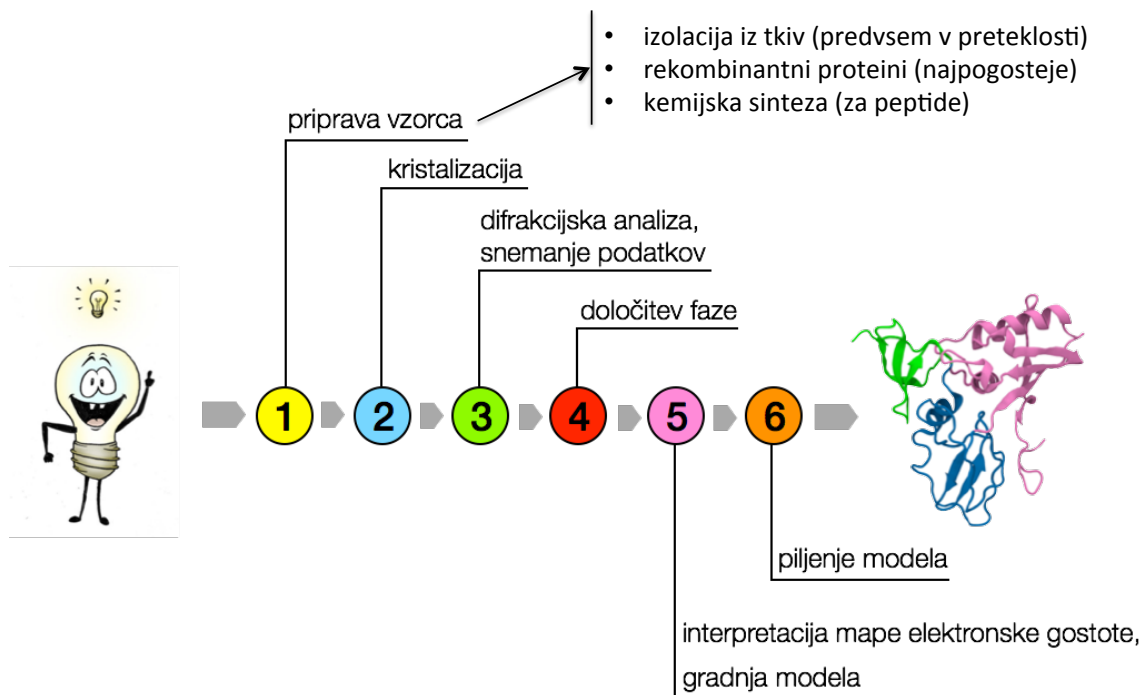
← FT



© Garland Science 2010

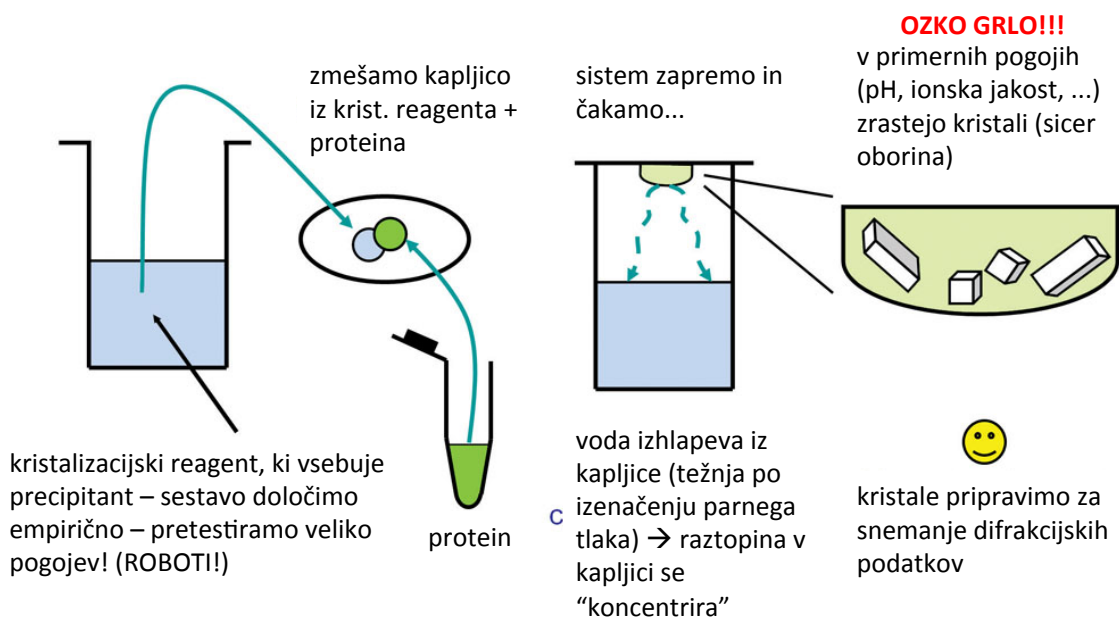
3.a. Metode za določevanje 3D struktur makromolekul (visoka ločljivost): X-ray

Pregled določevanja strukture z rentgensko kristalografijo v praksi



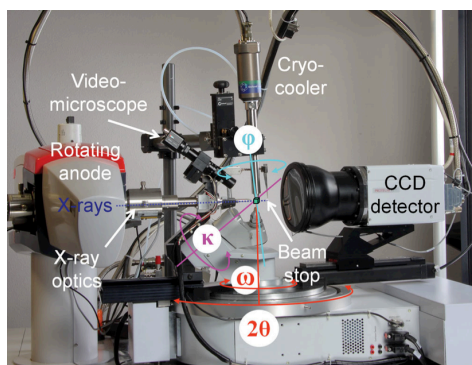
3.a. Metode za določevanje 3D struktur makromolekul (visoka ločljivost): X-ray

Kristalizacija – ponavadi izkoriščamo učinek izsoljevanja, za kar uporabimo metodo kristalizacije z difuzijo vodne pare (*vapour diffusion*).

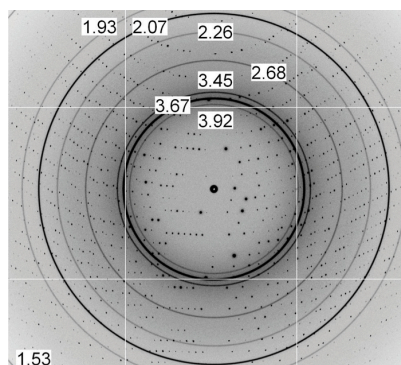


3.a. Metode za določanje 3D struktur makromolekul (visoka ločljivost): X-ray

Snemanje difrakcijskih podatkov – “hišni” viri (manjša jakost) in močnejši (sinhrotroni).



snemanje na hišnem viru

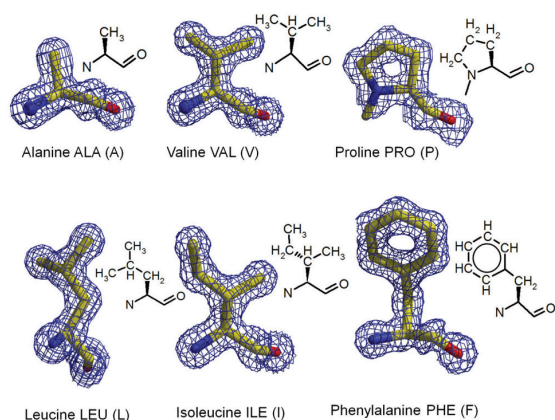


difrakcijski vzorec

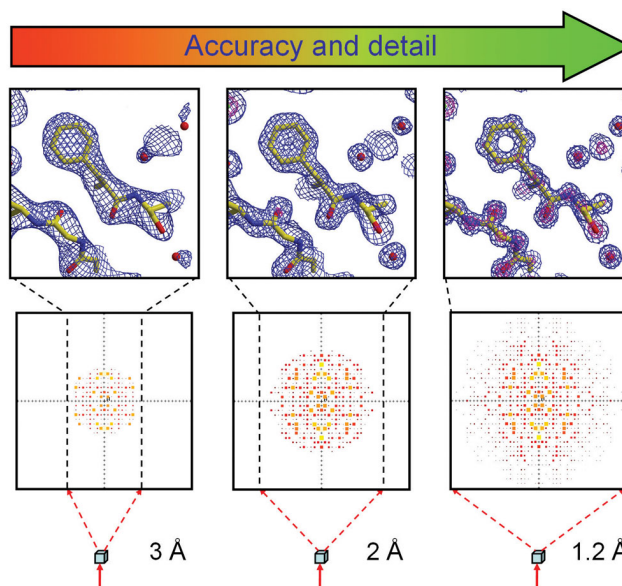
Sledi:

- procesiranje podatkov (lega in intenziteta difrakcijskih “lis”) → določitev osnovne celice,
- faziranje (reševanje faznega problema),
- gradnja modela v dano elektronsko gostoto (poznamo ak-zaporedje in vemo, kako izgledajo posamezni ak-ostanki!)

3.a. Metode za določanje 3D struktur makromolekul (visoka ločljivost): X-ray



Primer, kako izgledajo posamezni ak-ostanki.



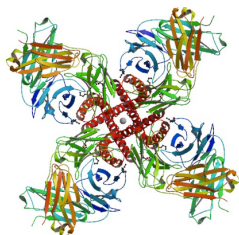
Ločljivost – pri boljši ločljivosti lahko razločimo podrobnosti strukture in model je bolj natančen.

3.a. Metode za določevanje 3D struktur makromolekul (visoka ločljivost): X-ray

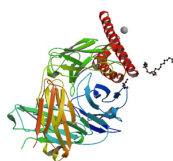
Pomembno!

ASIMETRIČNA ENOTA

- določimo strukturo ASIMETRIČNE enote, ki ni nujno (ponavadi ni!) enaka biološko aktivni obliki kristalizirane molekule
- **kristalnih kontaktov ne smemo enostavno enačiti s biološko relevantnimi medmolekulskimi kontakti** (razen v določenih primerih)
- v asimetrični enosti sta lahko tudi dve ali več enakih molekul, ki med seboj niso v stiku



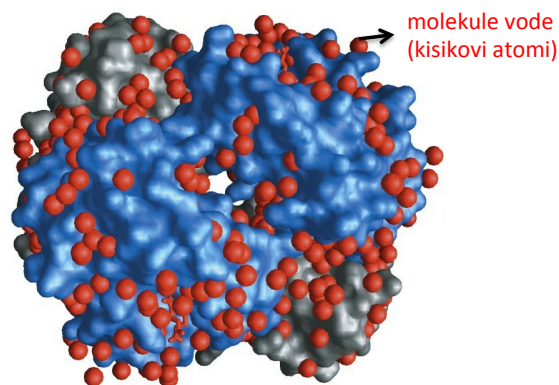
kalijev kanalček – biološka enota je tetramer



asimetrična enota v kristalu kalijevega kanalčka – ena podenota (ostale so simetrijsko povezane s to)

KRISTALNA VODA

- ~50% kristala predstavlja voda (strukture so zato do določene mere podobne tistim v raztopini)
- večina te vode je “neurejene” – molekule so v asimetričnih enotah na različnih mestih
- nekatere molekule vode (tik ob površini) so pri vseh molekulah na istem mestu – te molekule “vidimo” v strukturi!



3.a. Metode za določevanje 3D struktur makromolekul (visoka ločljivost): X-ray

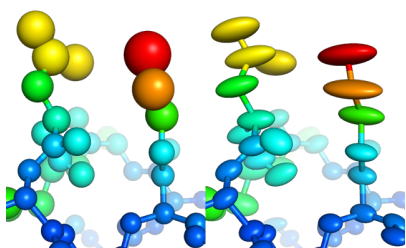
Pomembno!

NEUREJENI DELI

- tistih delov molekule, ki so neurejeni, navadno ne vidimo (ponavadi gre za N- in C-konce ter razne zanke na površini) – v modelu strukture manjkajo
- nekatere dele, ki so deloma neurejeni, “vidimo”, a je elektronska gostota slabše definirana
- to relativno vibracijsko gibanje atomov opišemo s t.i. temperaturnim ali B-faktorjem

$$B_j = 8\pi^2 \times \bar{u}_j^2$$

kvadrat povprečnega radija vibracije

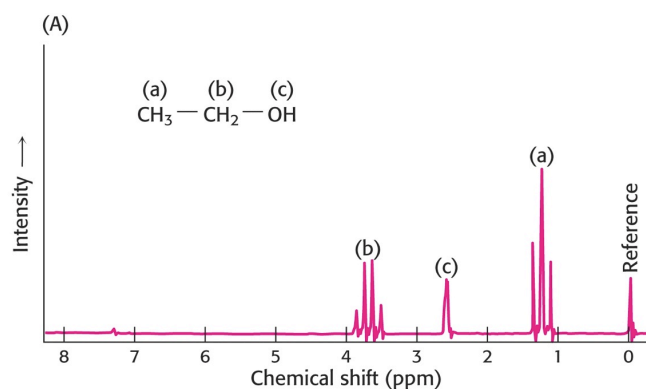
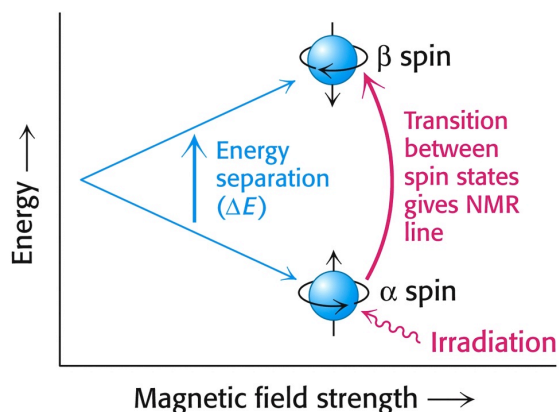


VODIKOVI ATOMI

- so premajni in imajo premalo elektronov (slabo sipajo), zato jih v strukturah, določenih z rentgensko difrakcijo, ne vidimo
- (najbolje sipajo atomi z več elektroni)
- včasih jih za namen interpretacije strukture dodamo naknadno (računalniško)

3.b. Metode za določanje 3D struktur makromolekul (visoka ločljivost): NMR

Metoda temelji na absorpciji in re-emitiranju EM valov atomskih jeder v magnetnem polju.

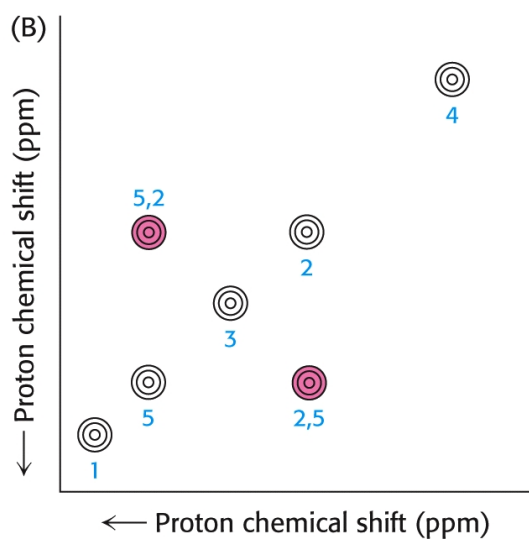
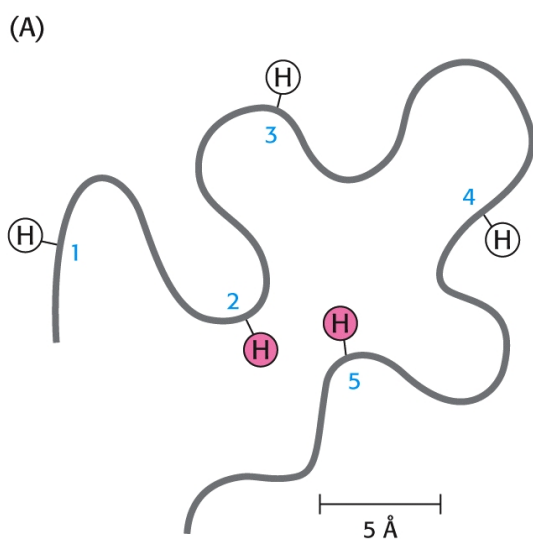


Molekula je pri snemanju v raztopini (torej odpadejo težave s kristalizacijo).

Kemijski premik (chemical shift) opisuje, kako se resonančna frekvenca nekega jedra razlikuje od standardnega (zaradi različnega okolja).

3.b. Metode za določanje 3D struktur makromolekul (visoka ločljivost): NMR

Okolica vpliva na kemijski premik.

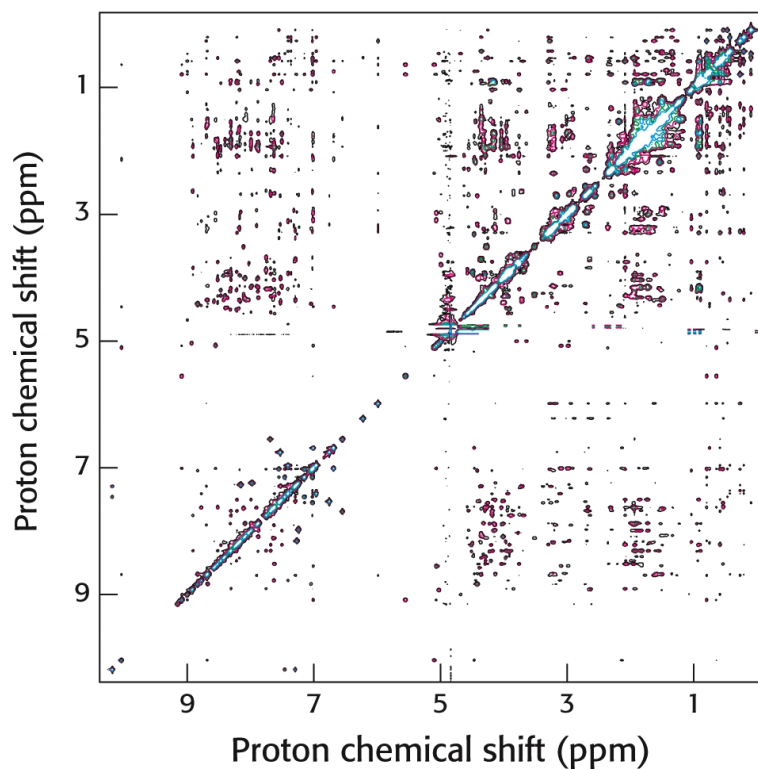


3.b. Metode za določanje 3D struktur makromolekul (visoka ločljivost): NMR

2D spektri NMR:

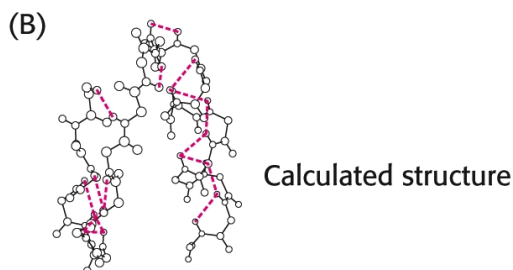
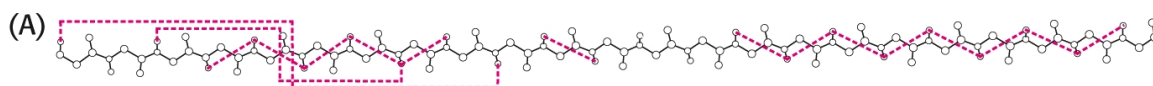
Asignacija spektra – posameznemu signalu pripišemo atom.

Pri velikem številu atomov se signali prekrivajo – od tod izvira omejitev NMR, sja lahko s to metodo določamo strukturo le manjših molekul (npr. do 20 kDa).

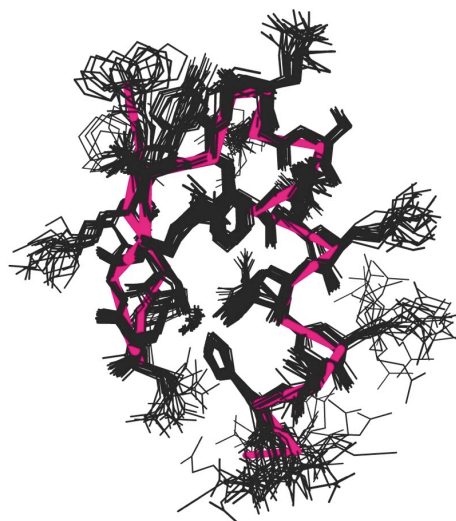


3.b. Metode za določanje 3D struktur makromolekul (visoka ločljivost): NMR

Iz kemijskih premikov izračunamo razdalje, ki nam nato služijo kot omejitve pri računanju strukture.



- Omejitvam ustreza več izračunanih modelov, ki se med seboj bolj razlikujejo na fleksibilnih regijah (N- in C-konec, površinske zanke).
- Vidimo vodikove atome!
- Ponavadi podamo eno **povprečno strukturo** ter več posameznih modelov.



4.a. Zapis 3D strukturnih podatkov in podatkovne baze: FORMAT PDB

Strukturne podatke ponavadi zapišemo kar kot kombinacijo:

- koordinat atomov (x, y, z), ki jim pripišemo
 - njihovo ime (C_α , C_β , C_γ , ...) ter
 - aminokislinski ostanek v zaporedju.
- enolična
rekonstrukcija
strukture

Te podatke lahko zapišemo v datotekah v različnih formatih:

- **PDB (običajni format; Protein Data Bank)**
- PDBML (XML format; *extended markup language*)
- mmCIF (macromolecular Crystallographic Information File)

```
(a) Abbreviated example PDB exchange data file
data_EXAMPLE
_entity poly.entity_id 1
_entity poly.type polypeptide(L)
_entity poly.nstd_linkage no
_entity poly.nstd_monomer no
_entity poly.pdb_seq_one_letter_code
;DIVLTQSPASLSASVGETVITTCRASGNIHNYLAWYQQKQKSPQLLVYITTLADG
VPSRFRSGSGGTQYSLKINSIQPEDFGSYQCQHFWSWTRPTFGGGTKLEIK
;
_entity poly.pdb_seq_one_letter_code can
;DIVLTQSPASLSASVGETVITTCRASGNIHNYLAWYQQKQKSPQLLVYITTLADG
VPSRFRSGSGGTQYSLKINSIQPEDFGSYQCQHFWSWTRPTFGGGTKLEIK
;

(b) Abbreviated example XML data file
<?xml version="1.0" encoding="UTF-8" ?>
<PDBx:datablock datablockName="EXAMPLE"
  xmlns:PDBx="http://deposit.pdb.org/pdbML/pdbx-v1.000.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://deposit.pdb.org/pdbML/pdbx-v1.000.xsd
    pdbx-v1.000.xsd">
  <PDBx:entity_polyCategory>
    <PDBx:entity_poly.entity_id="1">
      <PDBx:type>polypeptide(L)</PDBx:type>
      <PDBx:nstd_linkage>no</PDBx:nstd_linkage>
      <PDBx:nstd_monomer>no</PDBx:nstd_monomer>
      <PDBx:pdb_seq_one_letter_code>
        DIVLTQSPASLSASVGETVITTCRASGNIHNYLAWYQQKQKSPQLLVYITTLADG
        VPSRFRSGSGGTQYSLKINSIQPEDFGSYQCQHFWSWTRPTFGGGTKLEIK
      </PDBx:pdb_seq_one_letter_code>
      <PDBx:pdb_seq_one_letter_code_can>
        DIVLTQSPASLSASVGETVITTCRASGNIHNYLAWYQQKQKSPQLLVYITTLADG
        VPSRFRSGSGGTQYSLKINSIQPEDFGSYQCQHFWSWTRPTFGGGTKLEIK
      </PDBx:pdb_seq_one_letter_code_can>
    </PDBx:entity_poly>
  </PDBx:entity_polyCategory>
</PDBx:datablock>
```

PDBML

```
data_PDR025
_symmetry.space_group_name_H-M 'C 2 2 21'
_cell.length_a 137.290
_cell.length_b 153.100
_cell.length_c 76.200
_cell.angle_alpha 90.00
_cell.angle_beta 90.00
_cell.angle_gamma 90.00
#
loop_
_atom_site.id
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_seq_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
1 O5 ? G A 1 -3.897 61.994 -24.841
2 C5 ? G A 1 -5.016 62.932 -24.760
3 C4 ? G A 1 -4.695 64.310 -25.301
4 O4 ? G A 1 -4.459 65.240 -24.222
5 C3 ? G A 1 -3.413 64.316 -26.115
6 O3 ? G A 1 -3.497 65.336 -27.093
7 C2 ? G A 1 -2.354 64.634 -25.084
8 C1 ? G A 1 -3.091 65.629 -24.212
9 N9 ? G A 1 -2.631 65.588 -22.836
10 C8 ? G A 1 -2.125 64.484 -22.203
11 N7 ? G A 1 -1.781 64.713 -20.965
12 C5 ? G A 1 -2.077 66.050 -20.769
```

mmCIF

4.a. Zapis 3D strukturnih podatkov in podatkovne baze: FORMAT PDB

Format PDB

Običajna tekstna datoteka (odpremo jo lahko z npr. Notepad-om), ki vsebuje strukturne in druge podatke:

- **HEADER:** kategorija proteina, datum, koda zapisa v bazo PDB
- **TITLE:** naslov
- **COMPND:** ime spojin(e)
- **SOURCE:** oranizem
- **KEYWDS:** ključne besede
- **EXPDTA:** podatki o eksperimentalnem določanju strukture
- **AUTHOR:** avtorji
- **JRNL:** v katerem članku je struktura objavljena
- **REMARK:** razne opombe glede določanja strukture, elementov sekundarne strukture, biološko aktivna oligomerna enota, ...
- **SEQRES:** ak-zaporedje ali nt-zaporedje
- **HELIX, SHEET:** elementi sekundarne strukture
- **ATOM:** koordinate atomov aminokislinskih ostankov in nukleotidov
- **CONNECT:** povezave med ostalimi atomi (ki niso del ak oz. nt)

4.a. Zapis 3D strukturnih podatkov in podatkovne baze: FORMAT PDB

Format PDB

kategorija	podatki	koda PDB
HEADER	TOXIN, TRANSFERASE	28-JAN-14
TITLE	CRYSTAL STRUCTURE OF DIPHTHERIA TOXIN AT ACIDIC PH	
COMPND	MOL_ID: 1;	4OW6
COMPND	2 MOLECULE: DIPHTHERIA TOXIN;	
COMPND	3 CHAIN: A, B;	
COMPND	4 FRAGMENT: UNP RESIDUES 26-560;	
COMPND	5 EC: 2.4.2.36	
SOURCE	MOL_ID: 1;	
SOURCE	2 ORGANISM_SCIENTIFIC: CORYNEBACTERIUM DIPHTHERIAE;	
SOURCE	3 ORGANISM_TAXID: 698963;	
SOURCE	4 STRAIN: C7 (BETA);	
SOURCE	5 ATCC: 27012/C7 (BETA)	
KEYWDS	DIPHTHERIA TOXIN TRANSLOCATION, MEMBRANE INSERTION, BICELLES,	
KEYWDS	2 MEMBRANE CHANNELS	
EXPDTA	X-RAY DIFFRACTION	
AUTHOR	O.LEKA,F.VALLESE,M.PIRAZZINI,P.BERTO,C.MONTECUCCO,G.ZANOTTI	
REVDAT	1	02-APR-14 4OW6 0
JRNL	AUTH O.LEKA,F.VALLESE,M.PIRAZZINI,P.BERTO,C.MONTECUCCO,G.ZANOTTI	
JRNL	TITL DIPHTHERIA TOXIN CONFORMATIONAL SWITCHING AT ACIDIC PH.	
JRNL	REF	FEBS J. 2014
JRNL	REFN	ISSN 1742-4658
JRNL	PMID	24628974
JRNL	DOI	10.1111/FEBS.12783

4.a. Zapis 3D strukturnih podatkov in podatkovne baze: FORMAT PDB

Format PDB

ak-zaporedje

SEQRES	1	A	535	GLY	ALA	ASP	ASP	VAL	VAL	ASP	SER	SER	LYS	SER	PHE	VAL
SEQRES	2	A	535	MET	GLU	ASN	PHE	SER	SER	TYR	HIS	GLY	THR	LYS	PRO	GLY
SEQRES	3	A	535	TYR	VAL	ASP	SER	ILE	GLN	LYS	GLY	ILE	GLN	LYS	PRO	LYS
SEQRES	4	A	535	SER	GLY	THR	GLN	GLY	ASN	TYR	ASP	ASP	ASP	TRP	LYS	GLY
SEQRES	5	A	535	PHE	TYR	SER	THR	ASP	ASN	LYS	TYR	ASP	ALA	ALA	GLY	TYR
SEQRES	6	A	535	SER	VAL	ASP	ASN	GLU	ASN	PRO	LEU	SER	GLY	LYS	ALA	GLY
SEQRES	7	A	535	GLY	VAL	VAL	LYS	VAL	THR	TYR	PRO	GLY	LEU	THR	LYS	VAL
SEQRES	8	A	535	LEU	ALA	LEU	LYS	VAL	ASP	ASN	ALA	GLU	THR	ILE	LYS	LYS
SEQRES	9	A	535	GLU	LEU	GLY	LEU	SER	LEU	THR	GLU	PRO	LEU	MET	GLU	GLN

zap. št.	atom	ak	veriga	št. ak	x	y	z	"okupanca" (0 ≤ X ≤ 1)	B	tip atoma	
ATOM	1	N	GLY	A	1	-26.163	-28.379	4.729	1.00	40.15	N
ATOM	2	CA	GLY	A	1	-25.108	-29.325	4.427	1.00	36.36	C
ATOM	3	C	GLY	A	1	-23.915	-28.612	3.830	1.00	40.41	C
ATOM	4	O	GLY	A	1	-23.679	-27.445	4.143	1.00	40.12	O
ATOM	5	N	ALA	A	2	-23.167	-29.296	2.966	1.00	39.58	N
ATOM	6	CA	ALA	A	2	-21.946	-28.718	2.403	1.00	35.51	C
ATOM	7	C	ALA	A	2	-22.168	-27.326	1.802	1.00	36.62	C
ATOM	8	O	ALA	A	2	-21.273	-26.481	1.830	1.00	36.03	O
ATOM	9	CB	ALA	A	2	-21.362	-29.649	1.371	1.00	24.06	C
ATOM	10	N	ASP	A	3	-23.391	-27.065	1.347	1.00	41.58	N
ATOM	11	CA	ASP	A	3	-23.713	-25.832	0.631	1.00	39.84	C

koordinate atomov

4.a. Zapis 3D strukturnih podatkov in podatkovne baze: **FORMAT PDB**

Format PDB

Podatki o biološko-ativni oligomerni obliki – matrica transformacij BIOMT, preko katere lahko sestavimo nativno oligomerno obliko.

```
REMARK 350 BIOMOLECULE: 1
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: TRIMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: TRIMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 TOTAL BURIED SURFACE AREA: 4000 ANGSTROM**2
REMARK 350 SURFACE AREA OF THE COMPLEX: 18930 ANGSTROM**2
REMARK 350 CHANGE IN SOLVENT FREE ENERGY: -27.0 KCAL/MOL
REMARK 350 APPLY THE FOLLOWING TO CHAINS: L, H, P
REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.000000 0.000000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.000000 0.000000
REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.000000 0.000000
```

Biološko aktiven oligomer lahko dobimo:

- neposredno iz podatkovne baze,
- z “ročno” aplikacijo transformacijske matrike,
- s katerim od programom, ki prebere zapis v polju BIOMT.

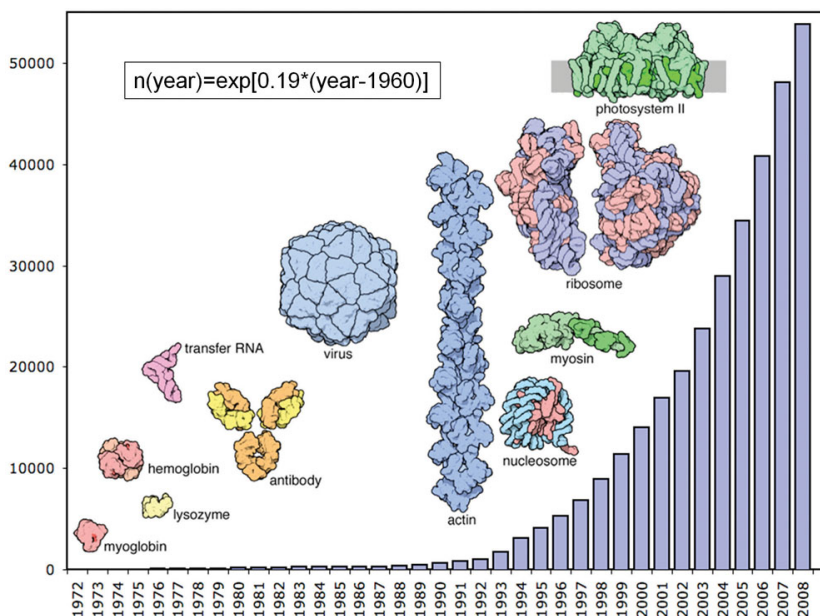
4.b. Zapis 3D strukturnih podatkov in podatkovne baze: **PODATKOVNE BAZE**

Podatki o eksperimentalno določenih strukturah bioloških makromolekul so v bazi **PDB** (Protein Data Bank) – podobno kot pri zaporedjih imamo tudi tu bazo v ZDA (**RCSB PDB**), Evropi (**PDBe**) in na Japonskem (**PDBj**).

RCSB **PDB**
PROTEIN DATA BANK

PDBe
Protein Data Bank in Europe

PDBj



4.b. Zapis 3D strukturnih podatkov in podatkovne baze: PODATKOVNE BAZE

Vsaka od baz ima nekoliko drugačen vmesnik, na slikah je prikazan vmesnik RCSB PDB.



The screenshot shows the RCSB PDB homepage. At the top, there is a navigation bar with the PDB logo and the text 'An Information Portal to Biological Macromolecular Structures'. Below this is a search bar with a dropdown menu set to 'Everything'. The search bar contains the text 'e.g., PDB ID, molecule name, author'. To the right of the search bar is a magnifying glass icon. Below the search bar is a 'Search History, Previous Results' link. The main content area is titled 'Biological Macromolecular Resource' and features a 'Full Description' section with 'Learn: Featured Molecules' and 'Structural View of Biology'. There are also sections for 'Molecule of the Month Aquaporin' and 'Protein Structure Initiative Featured System New NTF2-like Domains'. On the right side, there are 'New Features' and 'RCSB PDB News' sections. The bottom of the page has a 'Explore Archive' section with filters for 'Organism', 'Taxonomy', and 'Disease'.

enostavno iskanje

4.b. Zapis 3D strukturnih podatkov in podatkovne baze: PODATKOVNE BAZE

Napredno iskanje:

- sestavljanje kompleksnih iskanj
- lahko iščemo npr. po številu disulfidnih vezi, št. ak-ostankov, ligandih, ...

The screenshot shows the RCSB PDB homepage with the search bar set to 'Advanced'. The search bar contains the text 'e.g., PDB ID, molecule name, author'. Below the search bar is a 'Search History, Previous Results' link. The main content area is titled 'Advanced Search Interface' and features a 'Choose a Query Type:' dropdown menu. Below this is a search input field and a 'Result Count' button. There is also an 'Add Search Criteria' button. At the bottom, there is a checkbox for 'Retrieve only representatives at 90% sequence identity' and a 'Match all of the above conditions.' section. There are also 'Clear All Parameters' and 'Submit Query' buttons.

4.b. Zapis 3D strukturnih podatkov in podatkovne baze: PODATKOVNE BAZE

Zapis v bazo PDB vsebuje nekatere dodatne elemente, ki jih ni v sami datoteki PDB:

- grafični prikaz zaporedja in strukture,
- poročilo z validacije modela (relativno novo),
- povezave do orodij za analizo strukture,
- povezave do drugih podatkovnih baz,
- ...

Vsak zapis ima unikatno štirimestno PDB-kodo.

koda PDB

The screenshot shows the PDB entry page for 3ENI. The title is 'Crystal structure of the Fenna-Matthews-Olson Protein from Chlorobaculum Tepidum'. The PDB ID is 3ENI. The page includes a 3D ribbon diagram of the protein structure. The Primary Citation section mentions the authors Tronrud, D.E., Wen, J., Gay, L., and Blankenship, R.E. The Molecular Description section provides details about the protein, including its classification as a bacteriochlorophyll a protein and its length of 365 amino acids. The Deposition Summary section lists the authors and the date of deposition.

4.c. Zapis 3D strukturnih podatkov in podatkovne baze: PSI

Protein Structure Initiative (PSI)



- začetki v letu 2000
- cilj: zmanjšanje stroškov in skrajšanje časa do določitve 3D struktur
- povezava raziskovalnih (strukturnih) centrov
- visoko zmogljive (*high-throughput*, HT) metode, visoka stopnja avtomatizacije
- strukturna genomika: določevanje strukture vseh/večine proteinov nekega organizma
- faze PSI:
 1. faza – PSI-1 (2000-2005): analiza metod visoke zmogljivosti, ocena izvedljivosti
 2. faza – PSI-2 (2005-2010): implementacija metod iz 1. faze
 3. faza – PSI-3 (2010→): mreže raziskovalnih institucij, ki uporabljajo visoko zmogljive pristope za določevanje struktur

PSI Nature Structural Biology Knowledgebase: SBKB
<http://sbkb.org>

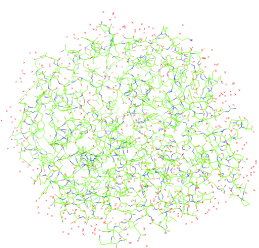
The screenshot shows the PSI Nature Structural Biology Knowledgebase (SBKB) website. The page has a dark red header with the PSI logo and navigation menus. The main content area includes a search bar, a 'Current Focus' section with a featured system, and several news items and articles. The footer contains links to 'Latest PSI Results', 'Latest Structures', and 'See all recent structures'.

5. Vizualizacija 3D struktur makromolekul

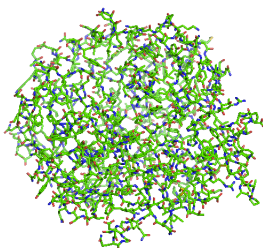
Na voljo več programov (večinoma dostopni brezplačno), najpogosteje se uporabljajo:

- PyMOL,
- VMD (Visual Molecular Dynamics),
- Chimera.

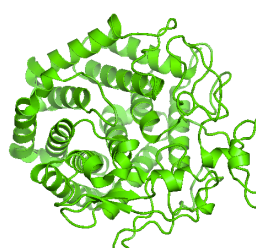
V vseh programih lahko model strukture prikažemo na različne načine:



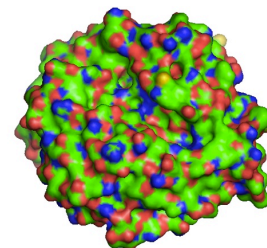
črte (*lines*)



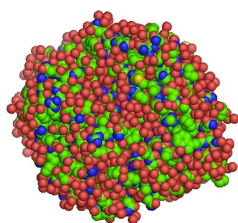
paličice (*sticks*)



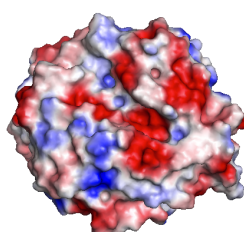
trak (*ribbon*)



molekulska površina (*molecular surface*)



kroglice (*spheres*)



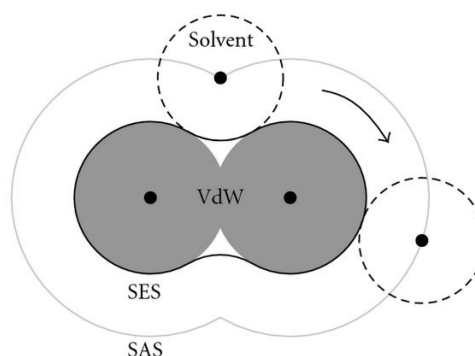
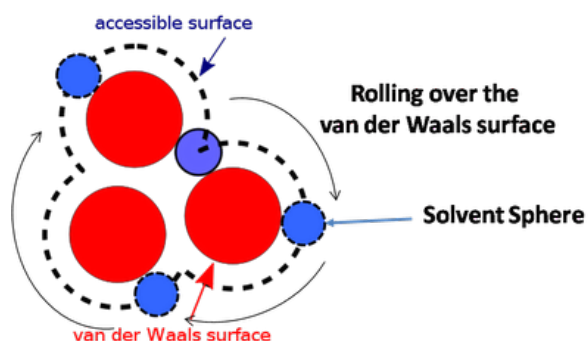
molekulska površina z elektrostatskim potencialom (*molecular surface color-coded according to electrostatic potential*)

modro: + naboj (**DUŠIK**)
rdeče: - naboj (**KISIK**)
belo: nepolarno

5. Vizualizacija 3D struktur makromolekul

Molekulska površina:

- van der Waalsova (vdW) površina – glede na vdW radij atomov,
- topilo dostopna površina – zgeneriramo jo tako, da preko modela strukture z vdW površino "kotalimo" kroglico z radijem molekule vode.

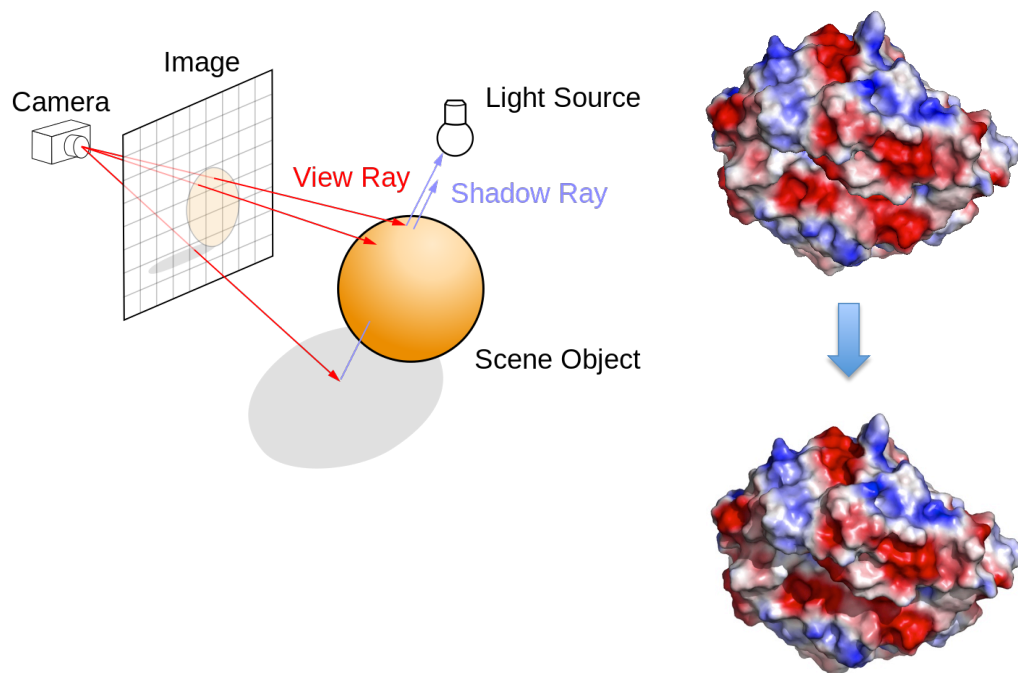


SAS – solvent accessible surface
(površina, dostopna topilu)

SES – solvent excluded surface
(površina, ki ni dostopna topilu)

5. Vizualizacija 3D struktur makromolekul

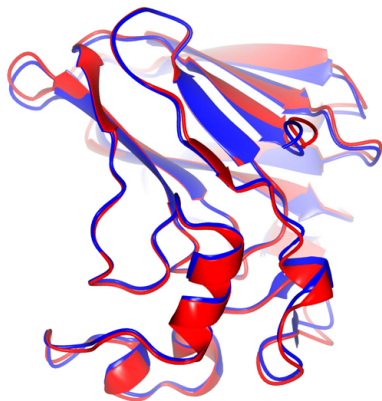
Modele naredimo "privlačnejše na oko" s postopkom, imenovanim *raytracing*, pri katerem simuliramo pot žarkov, njihovo lomljenje, sence ... na različnih površinah...



6.a. Primerjava 3D struktur makromolekul: SUPERPOZICIJA IN PODOBNOST

- Podobnost med dvema ali večimi strukturami lahko ugotovimo z njuno **superpozicijo**.
- **Superpozicija struktur** je analogna prileganju ak-zaporedij.
- Superpozicijo imenujemo tudi strukturno prileganje (*structural alignment*).
- Lahko upoštevamo vse atome (pri zelo podobnih strukturah) ali le npr. atome glavne verige (pri manj podobnih strukturah).
- Prileganje lahko vodimo s prileganjem ak-zaporedij.

Razliko opišemo kot vrednost **RMSD**:



$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

$$\begin{aligned} \text{RMSD}(\mathbf{v}, \mathbf{w}) &= \sqrt{\frac{1}{n} \sum_{i=1}^n \|v_i - w_i\|^2} \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)} \end{aligned}$$

root-mean-square deviation
(koren povprečne vrednosti kvadratov razlik)

6.b. Primerjava 3D struktur makromolekul: ZVITJE IN KLASIFIKACIJA STRUKTUR

Polipeptidne verige s podobnim ak-zaporedjem zavzamejo podobno 3D strukturo – značilno zvitje (*fold*), ki ga opišemo z zaporedjem in relativno orientacijo elementov sekundarne strukture.

Dve osnovni klasifikaciji:

- **SCOP** - Structural Classification of Proteins (nekoliko zastarela in neosvežena; nova verzija **SCOP2**)
- **CATH** – opis nivojev strukture:
 - CLASS - razred (alfa, beta, mešano alfa/beta, ali pa z malo sekundarne strukture)
 - ARCHITECTURE - arhitektura (celotna oblika glede na orientacijo elementov sekundarne strukture)
 - TOPOLOGY - topologija (družina zvitja)
 - HOMOLOGY - homologna superdružina (združuje proteinske domene)

