

struktura bioloških molekul

andrej perdih • katja lakota • alja prah

famnit textbooks | famnitovi učbeniki | 2

strukture bioloških molekul

famnit textbooks | famnitovi učbeniki | 2



Strukture bioloških molekul

Andrej Perdih
Katja Lakota
Alja Prah



Andrej Perdih, Katja Lakota, Alja Prah
Strukture bioloških molekul
Univerzitetni učbenik z recenzijo in navodila za vaje

Zbirka
Famnit textbooks, Famnitovi učbeniki, 2
ISSN 2738-2966

Recenzenta
Janez Mavri
Snežna Sodin-Šemrl

Jezikovni pregled
Ana Sollner Perdih

Izdala in založila
Založba Univerza na Primorskem (zanjo: Klavdija Kutnar, rektorica)
Titov trg 4, SI-6000 Koper
Glavni urednik
Jonatan Vinkler
Vodja založbe
Alen Ježovnik

Koper, 2020

ISBN 978-961-293-020-2 (spletna izdaja: pdf)
<http://www.hippocampus.si/ISBN/978-961-293-020-2.pdf>
ISBN 978-961-293-021-9 (spletna izdaja: html)
<http://www.hippocampus.si/ISBN/978-961-293-021-9/index.html>

DOI: <https://doi.org/10.26493/978-961-293-020-2>

© 2020 Založba Univerze na Primorskem

Učbenik je namenjen študentom in študentkam *Bioinformatike* in *Varstvene biologije* pri predmetu *Strukture bioloških molekul*.



Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=28974595
ISBN 978-961-293-020-2 (pdf)
ISBN 978-961-293-021-9 (html)

kazalo

- Katja Lakota in Alja Prah
- 9 I. POGLAVJE
BIOLOŠKE MAKROMOLEKULE**
- Andrej Perdih
- 43 II. POGLAVJE
DOLOČEVANJE 3D STRUKTURE BIOLOŠKIH MOLEKUL
S KRISTALOGRAFIJO**
- Andrej Perdih
- 57 III. POGLAVJE
DOLOČEVANJE 3D STRUKTURE BIOLOŠKIH MOLEKUL Z NMR
SPEKTROSKOPIJO**
- Andrej Perdih
- 75 IV. POGLAVJE
MOLEKULSKO MODELIRANJE**
- Andrej Perdih
- 99 V. POGLAVJE
NAPOVEDOVANJE 3D STRUKTURE PROTEINOV Z RAČUNSKIMI
METODAMI**
- Andrej Perdih
- 109 VI. POGLAVJE
UPORABA PROTEINSKIH STRUKTUR PRI NAČRTOVANJU BIOLOŠKO
AKTIVNIH MOLEKUL**
- Katja Lakota in Alja Prah
- 121 VAJA 1
SPLETNA ORODJA IN PODATKOVNE BAZE O BIOLOŠKIH
MAKROMOLEKULAH**
- Andrej Perdih in Alja Prah
- 133 VAJA 2
VIZUALIZACIJA MAPE ELEKTRONSKE GOSTOTE**

Andrej Perdih in Alja Prah

**139 VAJA 3
VIZUALIZACIJA BIOLOŠKIH MOLEKUL IN PRILEGANJE KONFORMACIJ
MAKROMOLEKUL**

Alja Prah

**147 VAJA 4
HOMOLOGNO MODELIRANJE 3D STRUKTURE NEZNANEGA
PROTEINA**

Alja Prah

**157 VAJA 5
ENERGIJSKA PRIMERJAVA KONFORMACIJ CIKLOHEKSANA**

Andrej Perdih

**161 VAJA 6
MOLEKULSKO SIDRANJE INHIBITORJA V AKTIVNO MESTO ENCIMA
MURD S PROGRAMOM ARGUSLAB**

DODATEK

167 VPRAŠANJA ZA PONOVIŠČE IN UTRJEVANJE SNOVI

171 DODATNA LITERATURA

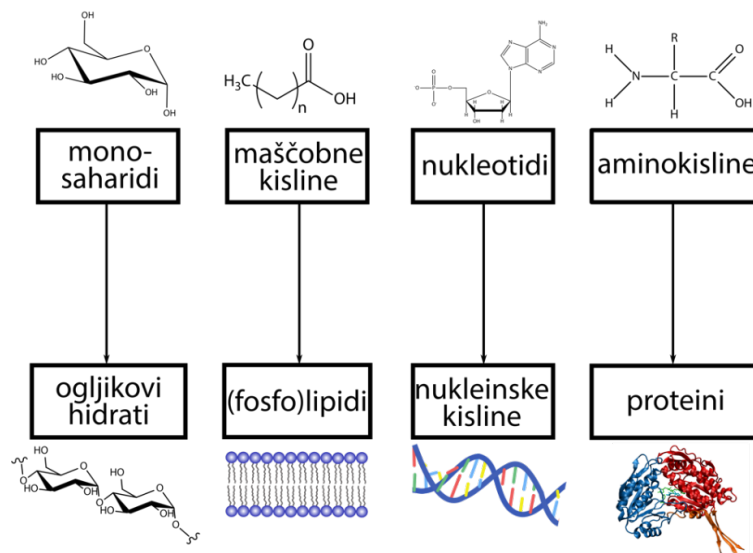
Katja Lakota in Alja Prah

biološke makromolekule

```
function quicksort(array)
  var list less, equal, greater
  if length(array) <= 1
    return array
  select a pivot value pivot
  for each x in array
    if x < pivot then list = less
    if x = pivot then list = equal
    if x > pivot then list = greater
  return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

Biološke makromolekule so ključne tako za strukturo kot tudi za funkcijo celic. Večinoma so biološke makromolekule **polimeri**, torej velike molekule, sestavljene iz več ponavljajočih se manjših podenot, imenovanih **monomeri**. Polimeri omogočajo nešteto možnosti za raznolikost makromolekul, sestavljenih iz omejenega nabora monomerov (p^m , kjer je p dolžina polimera in m število različnih monomernih enot), in enostavno recikliranje. Običajno nastajajo v kondenzacijskih reakcijah (*dehidracija*), kjer se tvori kovalentna vez med dvema monomeroma in se odcepi molekula vode. Razgradnji nazaj na monomerne enote pa rečemo *hidroliza*, pri kateri se ob prisotnosti vode cepi vez med dvema enotama.

Poznamo štiri večje skupine bioloških makromolekul - to so **ogljikovi hidrati**, **lipidi**, **nukleinske kisline** in **proteini** (Slika 1).



Slika 1. Poenostavljen pregled bioloških makromolekul in njihovih gradnikov.

Vse štiri skupine imajo v človeškem telesu ključne funkcije, ki jih bomo predstavili v nadaljevanju. Poleg vode pa te štiri skupine tudi predstavljajo večino mase človeškega telesa (Preglednica 1).

Preglednica 1. Sestava človeškega telesa.

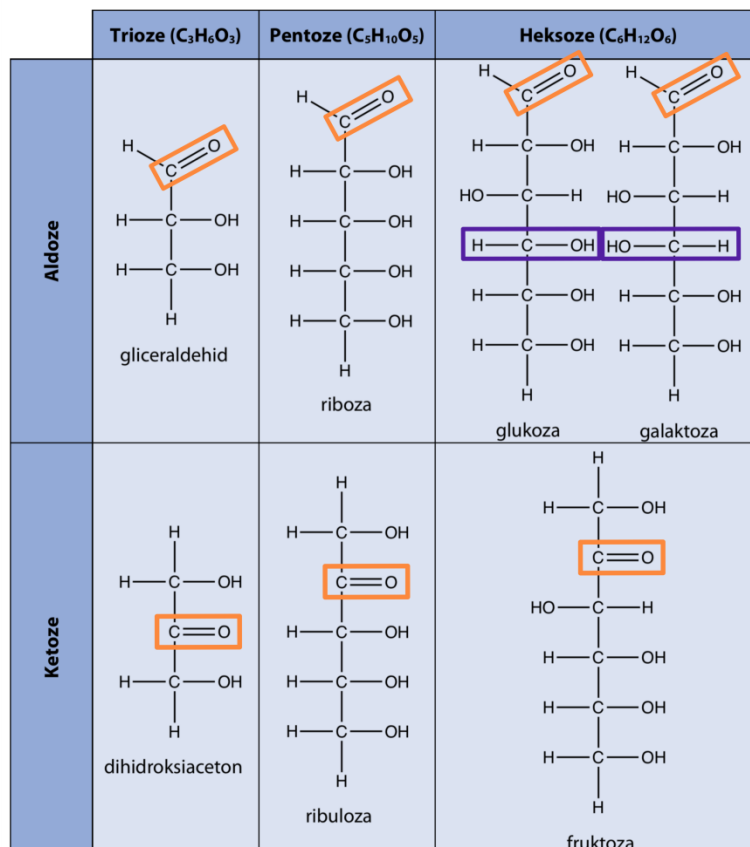
Spojina	% mase	% molekul
voda	65	98
druge anorganske spojine	1.5	0.7
lipidi	12	0.5
druge organske spojine	0.4	0.04
proteini	20	0.01
RNA	1	3×10^{-5}
DNA	0.1	3×10^{-11}

1. OGLJIKOVI HIDRATI (SLADKORJI)

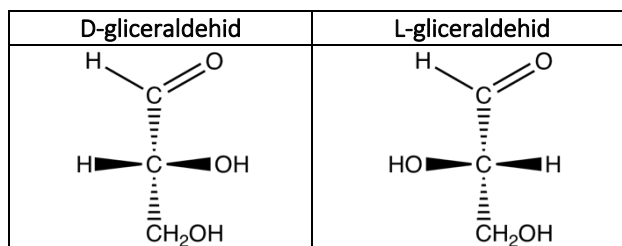
Ogljikovi hidrati oz. sladkorji so spojine, sestavljene iz ogljika, vodika in kisika. Iz tega izhaja tudi njihovo ime, saj sta vodik in kisik tipično prisotna v razmerju 2:1, kot v vodi (hidrati). Sladkorji so zgrajeni iz **monosaharidnih** enot, ki služijo kot vir energije celicam (glukoza, galaktoza, fruktoza), hkrati pa so sestavni deli drugih celičnih struktur (npr. RNA, DNA). **Polisaharidi** so oblika zaloge energije (npr. škrob), služijo pa tudi kot strukturni elementi celice (npr. hitin). **Oligosaharidi** služijo tudi za adhezijo celic, sodelujejo v procesih prepoznavanja med celicami in so označevalci za lokalizacijo proteinov v celici. V prehrani so pomembni kot sladila in kot vir vlaknin, delujejo kot konzervansi v hrani, zagotavljajo teksturo in služijo kot hrana kvasovkam (kvašeni izdelki in alkoholne pijače).

1.1. ENOSTAVNI SLADKORJI - MONOSAHARIDI

Kemijsko so monosaharidi polihidroksi aldehidi ali ketoni, kar pomeni, da imajo več –OH skupin in eno –C=O skupino, ki se bodisi nahaja (*a*) na koncu – **aldehid** (kot pri glukozi) ali (*b*) v sredini molekule – **keton** (kot pri fruktozi). Ker so polihidroksi spojine, so topni v vodi, netopni pa v lipofilnih topilih. Zanje je značilen sladek okus. Njihova osnovna formula je CH_2O_n . Monosaharide poimenujemo glede na število atomov ogljika – trioze, tetraze, pentoze, heksoze. Nekaj predstavnikov je prikazanih na Sliki 2. V kolikor ne obstaja trivialno ime (kot npr. fruktoza), imajo ketoze končnico -uloza.



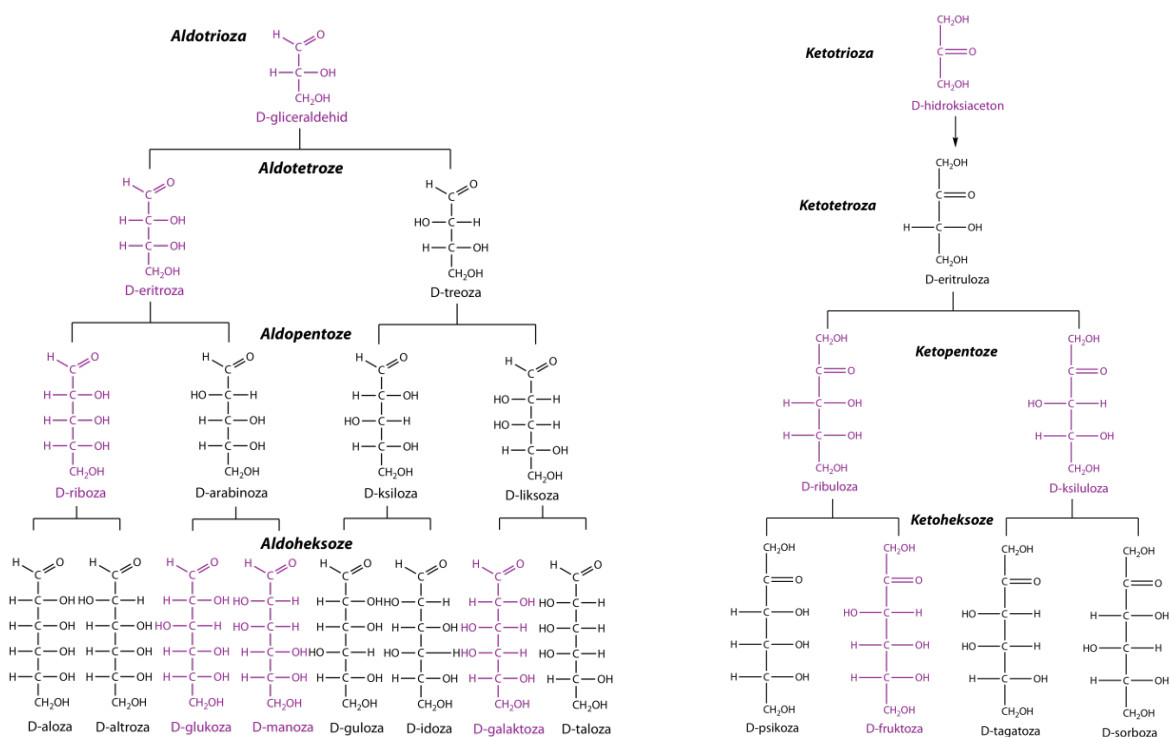
Slika 2. Monosaharide delimo na aldoze in ketoze ter glede na število ogljikovih atomov.



Slika 3. Obe obliki glicerinaldehida.

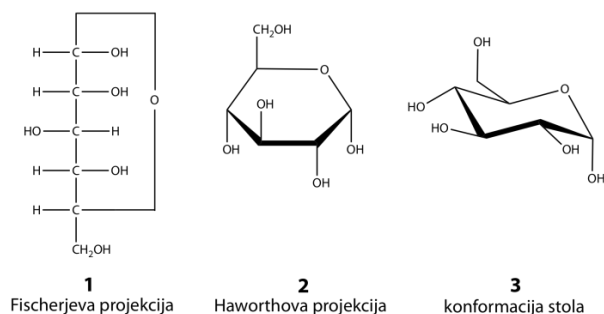
Dva monosaharida z identičnima molekulskima formulama sta še vedno lahko stereoisomera, tj. se razlikujeta v orientiranosti atomov v prostoru. Stereoizomere označujemo s črkama D in L, ki se pri sladkorjih ne nanašata na optično sučnost, ampak se določita glede na pripadnost D ali L vrsti **glicerinaldehida** (Slika 3), iz katerega potem dobimo raznolike sladkorje tako, da podaljšujemo verigo ogljikov (D oz. L oblika pentoze je zato odvisna od pozicije hidroksilne –OH skupine na ogljiku C4, heksoze pa na C5). Tako D- oz. L-obliko poimenujemo glede na položaj –OH skupine na predzadnjem ogljikovem atomu.

Ker ima vsak naslednji ogljikov atom lahko –OH skupino na D ali L strani, število možnih sladkorjev raste z 2^n dodatnih atomov ogljika (Slika 4). *Epimeri* so diastereoizomeri, ki se razlikujejo le na enem stereogenem centru (npr. biološko pomembnejša epimera glukoze sta manoza in galaktoza). V naravi ne najdemo vseh oblik sladkorjev, npr. idoze, taloze, redki so prisotne tudi aloza, altroza in guloza. Prav tako najdemo le nekaj L-oblik sladkorjev, npr. L-arabinozo (v pektinu), L-fukozo (6-deoksi-L-galaktoza) in L-ramnozo (6-deoksi-L-manoza) v rastlinskih glikozidih.



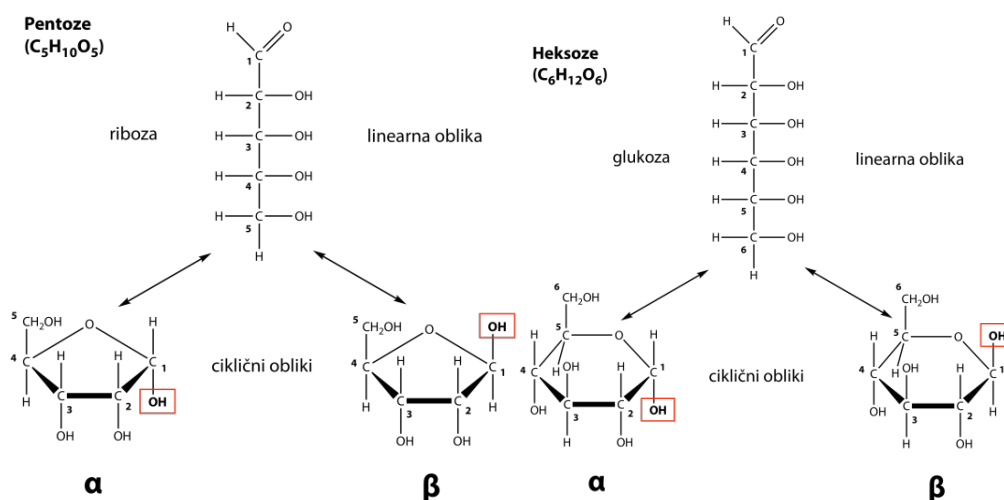
Slika 4. S podaljševanjem verige ogljikov dobimo raznolike monosaharide.

Zaradi bližine aldehydne in hidroksi skupine v strukturi formuli pri sladkorjih s štirimi ogljikovimi atomi ali več se ti dve skupini povežeta. Tako nastane **ciklična oblika** (t. i. polacetal), ki je lahko 5-členski obroč (**furanoza**) ali 6-členski obroč (**piranoza**). Stabilnost obroča pri piranozah je večja, zato tudi heksoze obstajajo pretežno kot piranoze. V resnici 6-členski obroč, kot je narisano ploskovno na Sliki 5.2, energetsko ni ugoden in zato obstaja v konformaciji »stola«, kot vidno na Sliki 5.3. Večinoma se monosaharidi v raztopinah nahajajo v ciklični obliki, položaj –OH skupine glede na obroč pa določa kemične lastnosti molekule.



Slika 5. Prikaz monosaharidne enote v obliki (1) Fischerjeve in (2) Haworthove projekcije ter v t. i. (3) konformaciji stola.

Ob ciklizaciji nastane še en nesimetričen ogljikov atom, možna sta dva anomera, α in β . V Haworthovem načinu pisanja je pri α obliki hidroksilna skupina na C1 ogljikovem atomu napisana navzdol, pri β obliki pa napisana navzgor (Slika 6). V Fischerjevem zapisu pa je pri D sladkorjih v α obliki hidroksilna skupina na C1 atomu napisana na desni, v β obliki pa na levi (pri L sladkorjih je ravno obratno). V raztopini α in β oblika prehajata druga v drugo v procesu, ki mu rečemo **mutarotacija**. Na primer: glukoza v vodni raztopini nastopa 36 % kot α -piranoza, 64 % kot β -piranoza in manj kot 1 % kot neciklična oblika, α - in β -furanoza skupaj. Proces mutarotacije je povezan s cepitvijo in tvorbo kemijskih vezi.



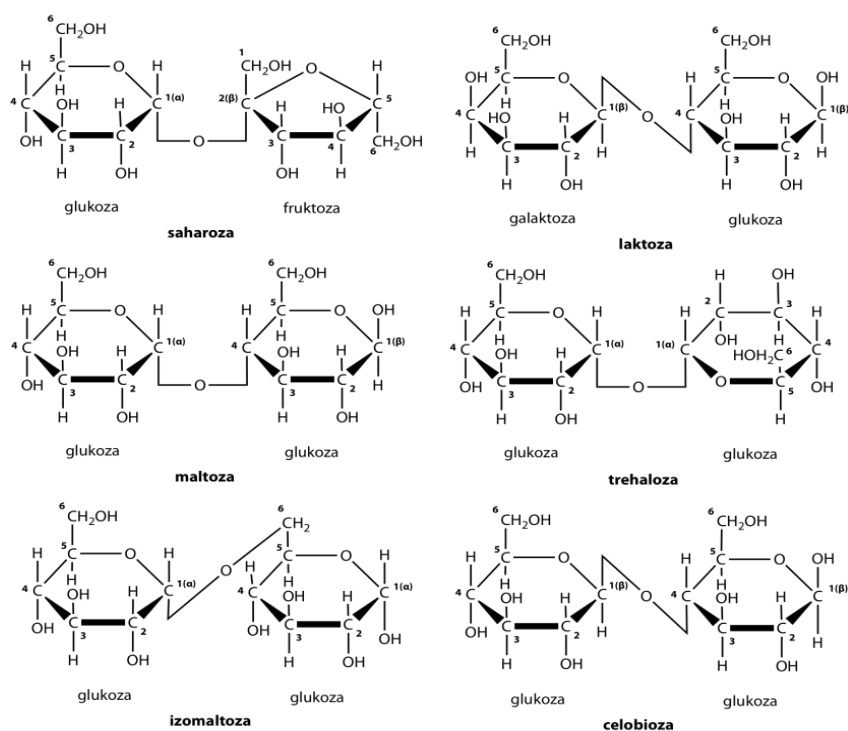
Slika 6. Prikaz α in β oblik pentoz in heksoz v Haworthovi projekciji.

Trioze so pomembne v, na primer mitohondrijih, ob razgradnji glukoze med celičnim dihanjem (piruvat). Pentozri riboza in deksiriboza sta pomembni sestavni enoti RNA oz. DNA molekul. Fruktozo vsebuje molekula, pomembna za prenos energije v celici - **adenozin fosfat** (*adenozin*

trifosfat = ATP, *adenozin monofosfat* = AMP). V nekaterih sadežih je fruktoza glavni sladkor (grozdje, hruške, češnje, fige). Glukoza je glavni vir energije v celici, nastaja s fotosintezo in se razgradi z dihanjem. V telesu dvigne raven inzulina, ki omogoča njen prehod v celico. Med vsebuje pretežno monosaharide (fruktozo in glukozo). Heptoze so pomembne za biosintezo lipida A (to je sestavni del lipopolisaharida (LPS), ki gradi celično steno Gram negativnih bakterij in sproži imunski odziv telesa).

1.2. DISAHARIDI

Disaharidi so sestavljeni iz dveh monosaharidnih enot, ki ju povezuje *glikozidna vez*. Ta nastane ob povezavi C1 ogljikovega atoma prvega monosaharida (orientiranega α ali β) z enim od ogljikovih atomov drugega monosaharida (orientacija te vezi je določena z njegovo strukturo). Od vezi je tudi odvisno, kakšen encim jo lahko cepi ob razgradnji. Biološko pomembnejši disaharidi so *saharoza*, *maltoza*, *trehaloza* in *laktoza* (Slika 7).



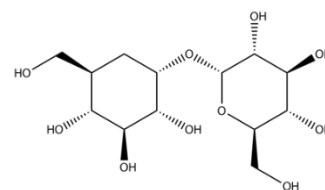
Slika 7. Nekateri pomembnejši disaharidi.

Saharoza oz. α -D-glukopiranozil-(1 \rightarrow 2)- β -D-fruktofuranozid. Ta zapis pomeni, da nastane glikozidna vez med C1 atomom α -D-glukoze (piranozni obroč) in C2 atomom β -D-fruktoze (furanozni obroč). Saharoza je sladkor, ki ga najpogosteje uporabljamo pri kuhanju. V soku sladkorne pese in sladkornega trsa ga je med 10 in 20 %, s kristalizacijo odstranijo nečistote in pridobije kristale čiste saharoze. Disaharid na monosaharidne enote cepita encima invertaza (v kvasu, rastlinah) in sukraza (v tankem črevesju). Za razliko od večine disaharidov, kjer se povežeta reducirajoči in nereducirajoči ogljikov atom molekule, sta tu povezana oba reducirajoča ogljikova atoma, kar onemogoči možnost nadaljnje povezave z monosaharidi.

Maltoza je sestavljena iz dveh enot glukoze, povezanih z α -1,4 glikozidno vezjo. Vsebujejo jo semena, nastane ob cepivi škroba z amilazo. *Izomaltoza* je izomer maltoze, pri kateri sta

glukozni enoti povezani z α -1,6 vezjo. Ob povezavi maltoze z dodatno glukozo nastane *maltotrioza*, ob nadaljnjem daljšanju pa *dekstrini* (maltodekstrini) in škrob.

Trehaloza je sestavljena iz dveh enot glukoze, povezanih z α -1,1 glikozidno vezjo. Je krvni sladkor čebel in metuljev in dobra zaloga energije, saj s cepitvijo ene vezi pridobimo dve glukozi, kar je energetsko ugodnejše kot cepitev ene vezi škroba. Pomembna je pri zaščiti celic pred izsušitvijo, saj tvori gel in zadržuje vodo (Slika 8).



Slika 8. Trehaloza.

Laktoza nastane s tvorbo glikozidne vezi med C1 ogljikovim atomom β -D-galaktoze in C4 ogljikovim atomom α ali β -D-glukoze. Je glavni sladkor v mleku sesalcev, vendar nekateri odrasli nimajo več *laktaze*, encima za razgradnjo laktoze, zato so laktozo intolerantni. Zato v črevesju to laktozo razgradijo bakterije, ki povzročijo pline, nerazgrajena laktoza pa poveča osmotski pritisk v črevesju, kar poveča količino vode v črevesju in povzroča drisko. Laktoza je stranski produkt pri proizvodnji sira in se zaradi svojih lastnosti (stisljivost, nizka cena) pogosto dodaja v farmacevtske oblike (npr. tablete) kot polnilo.

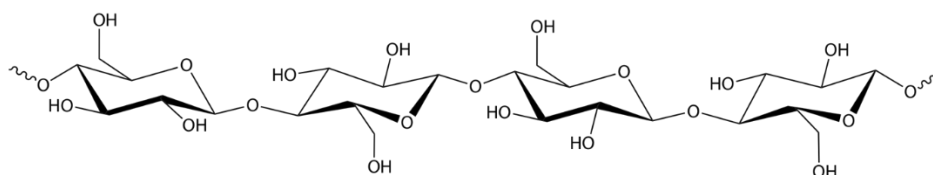
1.3. OLIGOSAHARIDI IN POLISAHARIDI

Oligosaharidi (iz grščine, *oligo* pomeni *malo*) vsebujejo 3-10 monosaharidnih enot in se v glavnem nahajajo v rastlinah. Pri rastlinah (stročnice, zelje) so pogosti takšni oligosaharidi, pri katerih je galaktoza vezana na saharozo. Primer sta rafinoza (sestavljena iz galaktoze, glukoze in fruktoze) in stahioza (sestavljena iz dveh galaktoz, glukoze in fruktoze). Encimi v tankem črevesu ju ne razgradijo, zato ju prebavijo bakterije debelega črevesa, kar povzroča neprijeten nastanek plinov. Oligosaharidi se preko O- ali N-glikozidne vezi pripenjajo na proteine, lipide in aminokislino (npr. na celični membrani, kjer služijo medceličnemu prepoznavanju).

Osnovne enote daljših **polisaharidov** so prav tako monosaharidi. Polisaharide lahko sestavljajo enake (*homopolisaharidi*) ali različne (*heteropolisaharidi*) monosaharidne enote. Razlikujejo se še v tipu vezi in razvejanosti. Imajo različne funkcije – so vir energije (npr. škrob, glikogen), tvorijo strukturne elemente (npr. celuloza, hitin, dekstran) ter celično steno.

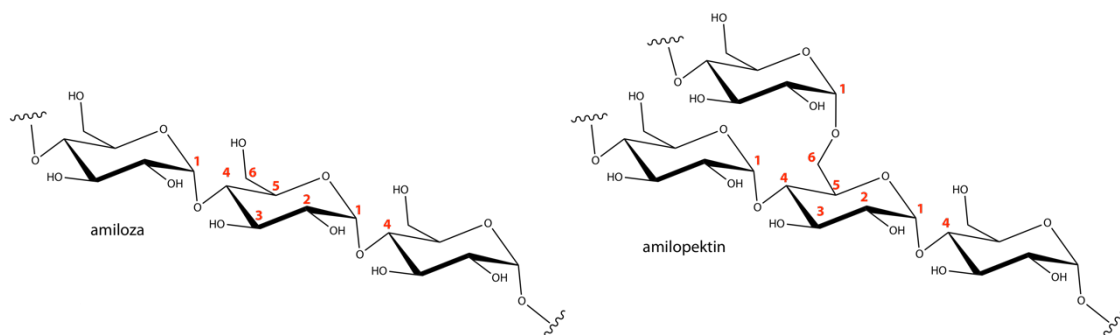
1.3.1. Homopolisaharidi

Celuloza, prikazana na Sliki 9, je linearen polimer, sestavljen iz glukoznih monomerov, povezanih z β -1,4 glikozidno vezjo. Posamezne verige se med seboj povezujejo z vodikovimi vezmi, kar tvori vlakna. Celuloza je glavna komponenta *celične stene* in zagotavlja strukturno oporo celici. Vez β omogoča raztegnjeno strukturo molekule in posledično povezavo več verig v fibrilarno strukturo. Človek in živali nimajo encima, ki bi razgradil β glikozidno vez, vendar se v prebavilih nekaterih živali nahajajo bakterije, ki to lahko storijo.



Slika 9. Celuloza, linearni polimer sestavljen iz glukoznih enot.

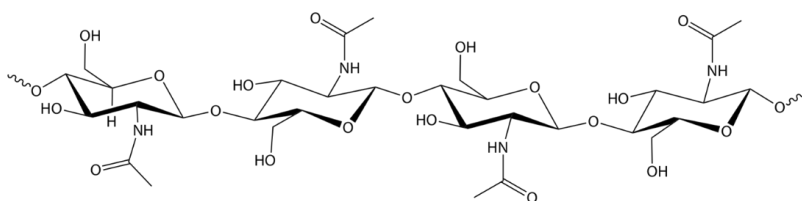
Škrob na Sliki 10 sestavljata nerazvejana *amiloza* (do 25 %) in razvejan *amilopektin* (75 % ali več). Amiloza je sestavljena iz nekaj tisoč α -D-glukoznih monomernih enot, povezanih z α -1,4 glikozidno vezjo, ki so urejene v desnosučno vijačno strukturo. Amilopektin ima podobno osnovno strukturo glukoznih enot, povezanih z α -1,4 glikozidno vezjo, na vsakih 20-30 enot pa se pri amilopektinu za razliko od amiloze pojavi α -1,6 glikozidna vez in pride do bolj razvejane strukture. Pri človeku in živalih α -glikozidne vezi razgradijo encimi *amilaze*.



Slika 10. Škrob je polimer sestavljen iz amiloze in amilopektina.

Glikogen je sestavljen iz glukoznih monomernih enot, ki so povezane z α -1,4 glikozidno vezjo. Na vsakih 12-18 monomerov se pojavi α -1,6 glikozidna vez. Nahaja se v jetrih in mišicah in je namenjen skladiščenju glukoze.

Hitin je polisaharid, sestavljen iz N-acetil-D-glukoza-2-aminskih monomernih enot, povezanih z β -1,4 glikozidno vezjo. Monomerna enota je podobna molekuli glukoze, le da ima namesto ene hidroksilne -OH skupine acetilaminsko skupino, kar omogoči nastajanje vodikovih vezi med verigami in daje strukturi trdnost. Nahaja se v celični steni gliv in v skeletu členonožcev, uporablja pa se tudi kot biorazgradljiva nit za šivanje ran (Slika 11).



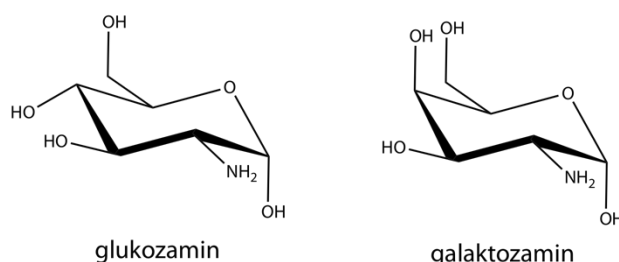
Slika 11. Hitin, polimer iz N-acetilglukozaminskih enot.

Inulin je sestavljen iz fruktoznih enot, povezanih z β -2,1 glikozidno vezjo ter z eno ali več molekulami glukoze na koncu verige. Je rezervna hrana v nekaterih gomoljih. Zaradi organoleptičnih lastnosti se uporablja kot nadomestilo maščob v namazih in kremah, saj zakrije okus po umetnih sladilih. Je neprebavljiv za človeške encime, v črevesju pa povzroča rast koristnih probiotičnih bakterij.

1.3.2. Heteropolisaharidi

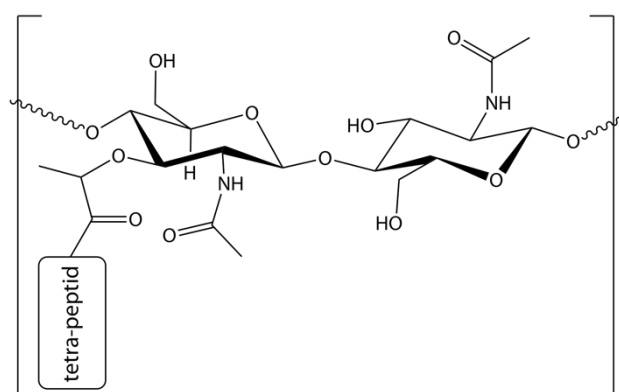
Nekateri polisaharidi poleg enostavnih monosaharidov vsebujejo tudi aminosladkorje ali kisline.

Glukozamin in **galaktozamin** sta pogosti monomerni enoti heteropolisaharidov. Amino skupina (-NH₂) se nahaja na C2 ogljikovem atomu in je pogosto acetilirana. Sta sestavna dela glikolipidov in kompleksnih polisaharidov. Glukozamin je prekursor glikozaminoglikanov, glavnih gradnikov hrustanca, N-acetil galaktozamin pa hitina in hondroitin sulfata (Slika 12).



Slika 12. Glukozamin in galaktozamin, pogosti monomerni enoti heteropolisaharidov.

Peptidoglikan tvori bakterijsko celično steno. Sestavljen je iz enot N-acetilglukozamina in N-acetilmuraminske kisline, ki so povezane z β -1,4 glikozidno vezjo. Na vsaki N-acetilmuraminski kislini najdemo krajšo aminokislinsko verigo, dolgo 4 do 5 aminokislin. Sintezo peptidoglikanske mreže zavirajo nekateri najpomembnejši antibiotiki v klinični praksi (npr. penicilini, cefalosporini) (Slika 13).



Slika 13. Peptidoglikan, osnovni gradnik bakterijske celične stene.

Agar je mešanica agaroze (polisaharid disaharida D-galaktoze in 3,6-anhidro-L galaktopiranoze) in agaropektina. Agar je pomemben za strukturo celične stene nekaterih vrst alg in se uporablja za gojenje bakterij v mikrobioloških laboratorijih.

Glikozaminoglikani so nerazvejani polimeri, sestavljeni iz disaharidnih enot, pri čemer je ena monomerna enota teh disaharidnih enot bodisi N-acetilglukozamin ali N-acetilgalaktozamin. Glikozaminoglikani tvorijo strukturo zunajceličnega matriksa. S povezovanjem glikozaminoglikanov in proteinov nastane **proteoglikan**. Nahajajo se v vezivnem tkivu, hrustancu in kosteh. Predstavnik glikozaminoglikanov sta: *hialuronska kislina*, sestavljena iz disaharidov iz N-acetilglukozamina in D-glikuronske kisline, in *hondroitin sulfat*, ki je sestavljen iz disaharidov iz N-acetilgalaktozamina in D-glikuronske kisline. Pri obeh se izmenjujejo β -1,3 in β -1,4 vezi.

Heparin je snov, ki se v medicini široko uporablja kot *antikoagulant* (tj. snov, ki preprečuje strjevanje krvi). Sestavljen je iz različno sulfatiranih disaharidnih enot iduronske kisline in glukozamina, povezanih z α -1,4 vezjo.

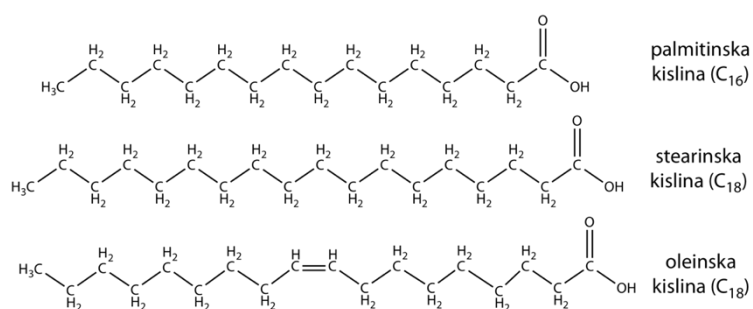
Glikoproteini so krajši oligosaharidi, povezani na proteine prek amidnega dušika na asparaginu (N-glikozidna vez) ali prek kisika na aminokislinah serin oz. treonin (O-glikozidna vez). Postranslacijska modifikacija, pri kateri se glikoproteini s pomočjo encimov pripnejo na proteine, se imenuje *glikozilacija*. Glikoproteini so pomembni za interakcije med celicami, medcelično prepoznavanje in kot označevalci, kam je protein namenjen (celični organeli, transport na površino celice).

2. LIPIDI

Ime *lipidi* se uporablja za v vodi netopne (hidrofobne) biološke molekule oz. molekule, ki so topne v nepolarnih topilih. V to skupino štejemo **maščobe** (masti), **olja**, **voske** pa tudi **trigliceride**, **fosfolipide** in **steroidne**. Imajo različne naloge: zaloga energije, izolacija, oblažitev, celično signaliziranje (hormoni in vitamini), sestavljajo pa tudi celične membrane.

2.1. TRIGLICERIDI

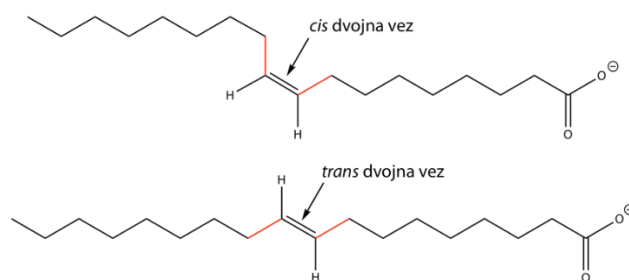
Zaloga energije v celicah je shranjena v obliki maščobnih kapljic v citoplazmi, v katerih se nahajajo **trigliceridi**. Iz enake mase maščobe dobimo dvakrat toliko energije kot iz ogljikovih hidratov ali beljakovin, torej so dobra zaloga energije. Trigliceridi so estri **glicerola** in treh **maščobnih kislin**. Slednje so lahko različne (Slika 14) – razlikujejo se glede na *dolžino verige* ogljikovih atomov (kratko, srednje in dolgoverižne) in *nasičenostjo*, kar se nanaša na prisotnost dvojnih vezi (pri *nasičenih* maščobnih kislinah niso prisotne dvojne C=C vezi, medtem ko pri *nenasičenih* so).



Slika 14. Struktura nekaterih bolj pogostih maščobnih kislin.

Kratkoverižne maščobne kisline (do 6 ogljikovih atomov) najdemo v mleku (*propanojska*, *butanojska*) in nastajajo ob prebavi vlaknin. **Srednjeverižne** maščobne kisline (6 do 12 ogljikovih atomov) najdemo na primer v kokosovem in palmovem olju. **Dolgoverižne** maščobne kisline (več kot 12 ogljikovih atomov) so najpogostejše v prehrani, npr. nasičeni maščobni kislini *stearinska* (C₁₆H₃₂O₂) in *palmitinska* (C₁₈H₃₆O₂), med nenasičenimi je najpogostejša *oleinska* (C₁₈H₃₄O₂) z eno dvojno vezjo. Ker biosinteza poteka iz ostankov očetne kisline, imajo maščobne kisline sodo število ogljikovih atomov (običajno 16 do 18).

Nenasičene maščobne kisline vsebujejo **eno ali več dvojnih vezi**, ki so v položaju *cis* ali *trans* (Slika 15). Naravne nenasičene maščobne kisline so vedno v *cis* konfiguraciji. *Cis* oblika ustvari ukrivljenost, zaradi togosti dvojne vezi je konformacija omejena in take nenasičene maščobne kisline so ukrivljene. To povzroči, da se ne morejo lepo zložiti druga poleg druge, kar posledično vpliva na njihovo tališče. **Nasičene** maščobne kisline **ne vsebujejo dvojnih vezi** (so nasičene z največjim možnim številom vodikov), veriga ostane ravna. Na splošno velja da so maščobe z nenasičenimi maščobnimi kislinami tekoče, tiste z nasičenimi pa trdne – odsotnost dvojnih vezi jim namreč omogoča, da se urejajo v bolj goste strukture. Industrijsko (s hidrogenacijo, npr. margarina) in ob povišani temperaturi (cvrtje) nastajajo nenasičene maščobne kisline v *trans* obliki. Le-te pripomorejo k nastanku srčno-žilnih in drugih obolenj. S kisikom iz zraka maščobe oksidirajo – posebej hitro oksidirajo dvojne vezi, nastanejo krajše maščobne kisline, ki imajo značilen vonj po žarkih maščobah.

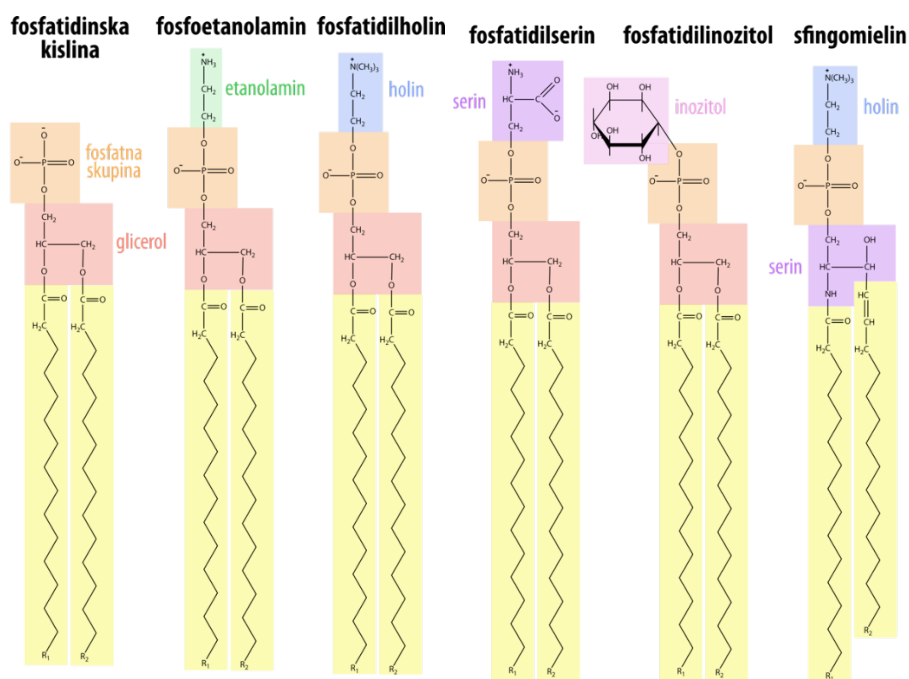


Slika 15. *Cis* oz. *trans* konfiguracija maščobnih kislin.

Esencialne maščobne kisline so večkrat nenasičene maščobne kisline (kar pomeni več dvojnih vezi), ki jih telo ne more tvoriti samo, a so nujno potrebne za njegovo delovanje. Z oznako ω -številka povemo, na katerem ogljikovem atomu, šteto od konca, se nahaja prva dvojna vez. Med esencialne maščobne kisline štejemo npr. *linolensko* (ω -3) in *linolno* (ω -6) maščobno kislino, obe z osemnajstimi ogljikovimi atomi (C18). Človeško telo ju podaljša do dolgoveržnih maščobnih kislin z dvajsetimi ogljikovimi atomi (C20) — iz linolenske nastaneta *eikozapentaenojska* kislina (EPA C20) in *dokozaheksaenojska* kislina (DHA C22), iz linolne pa *arahidonska* kislina (C20). Pomembne so za strukturo celične membrane, prav tako pa se iz njih tvorijo številne signalne molekule (npr. prostaglandini, endogeni kanabinoidi).

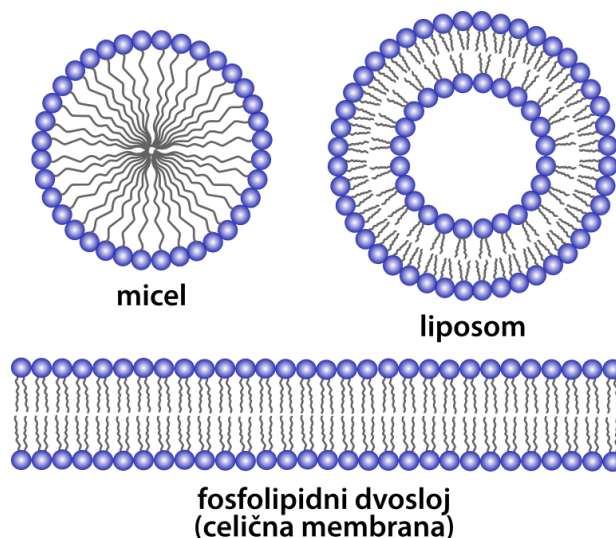
2.2. FOSFOLIPIDI

Fosfolipide sestavljata dve maščobni kislini, povezani s polarno skupino (Slika 16). V kolikor je v fosfolipid vgrajen glicerol, je tretja hidroksilna skupina povezana s fosfatno skupino, ta pa z majhno polarno molekulo, kot so *holin* (nastane **fosfatidilholin** - lecitin), *serin* (**fosfatidilserin**), *inozitol* (**fosfatidilinozitol**) ali *etanolamin* (**fosfoetanolamin**). Edini fosfolipid, ki ne vsebuje glicerola, je **sfigomielin**, v katerem sta maščobni kislini vezani na *serin*, nanj pa je vezan prek fosfatne skupine tudi *holin*.



Slika 16. Struktura nekaterih fosfolipidov.

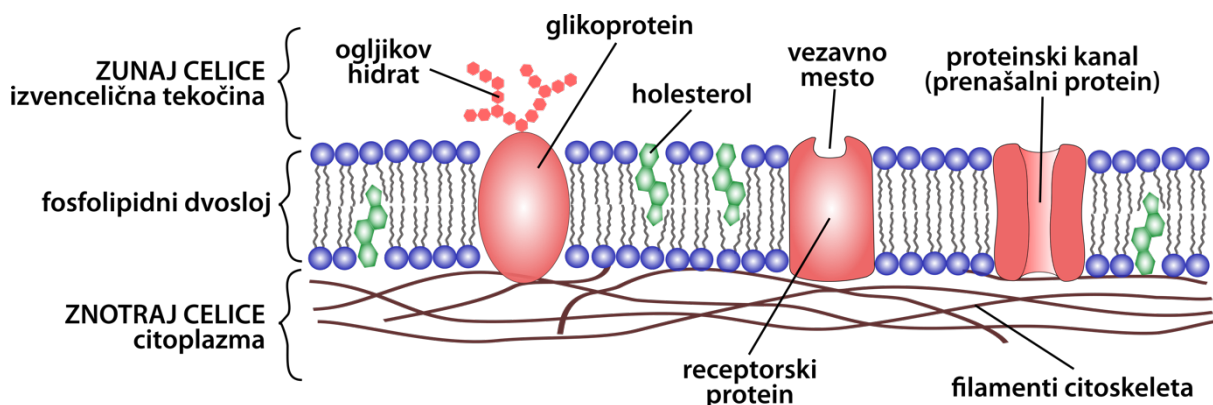
Fosfolipidi so **amfilne** molekule (*hidrofobne* na enem delu in *hidrofilne* na drugem). Fosfolipidi se uporabljajo kot emulgatorji, saj omogočajo tvorbo emulzij vode in olja. V vodi tvorijo dvosloj, tako da so hidrofobni deli obrnjeni v sredino, proti vodi pa polarne oz. hidrofilne »glave«. Tako tvorijo preproste *micele*, *liposome* in *celično membrano* (Slika 17). Liposomi se pogosto uporabljajo za dostavo različnih hidrofobnih/hidrofilnih sestavin v celico, saj se zlijejo s celično membrano. Uporabljajo pa se tudi v kozmetiki za dostavo aktivnih učinkovin v globlje dele kože.



Slika 17. Fosfolipidi lahko tvorijo micele, liposome ali fosfolipidni dvosloj.

Celična membrana je **tekoči mozaik**, tj. fosfolipidni dvosloj debeline od 4 do 10 nm (oz. 40 do 100 Å) s hidrofobnimi deli obrnjenimi v sredino (Slika 18). Poleg fosfolipidov celična membrana vsebuje še *glikolipide* (na zunanji strani membrane – sladkorne komponente služijo imunskemu prepoznavanju, interakcijam med celicami) in *holesterol* ter *proteine*, ki se gibljejo po lipidnem

dvosloju in omogočajo med drugim, prenos signalov, pritrjevanje celice in selektivno prepustnost membrane (npr. prenašalni proteini). Nahajanje fosfolipidov v membrani je različno, npr. fosfatidilserin se nahaja le na notranji (citosolni) strani membrane, na zunanjo stran se premakne le ob apoptozi (tj. celični smrti), ko pride do negativnega naboja in vezave makrofagov.

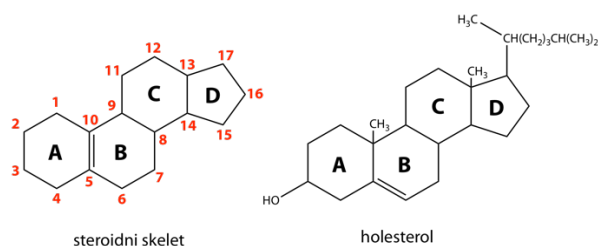


Slika 18. Enostaven shematski prikaz celične membrane.

2.3. STEROIDI

Steroidi so organske molekule, ki imajo štiri obroče (tri cikloheksanske in enega ciklopentanskega) v značilni konfiguraciji (Slika 19). Kljub temu, da strukturno torej niso podobni ostalim predstavnikom lipidov, jih v to skupino uvrščamo zaradi njihove hidrofobnosti in netopnosti v vodi. So pomembne komponente celičnih membran, ki vplivajo na njihovo fluidnost, sodelujejo pa tudi pri signalizaciji. V telesu je največ **holesterola**, iz katerega se tvorijo ostali ženski oz. moški steroidni **hormoni** (*testosteron, estrogeni, progesteron*), **kortikosteroidi**, **vitamin D** in **žolčne kisline**.

V strukturi holesterola so obroči, ki so rigidni in močno hidrofobni, zaradi polarnih hidroksilnih skupin pa je molekula amfifilna. Pomemben je v celičnih membranah, saj spreminja njihovo fluidnost. Celična membrana gliv in praživali namesto holesterola za isti namen vsebuje ergosterol. Ker ergosterola ni v živalski celični membrani, so encimi, s katerimi ga proizvajajo glive, dobra specifična tarča za protiglivična zdravila. Tudi rastline proizvajajo fitosterole, bogata z njimi so olja.



Slika 19. Tipičen steroidni skelet in najbolj razširjen predstavnik steroidov, holesterol.

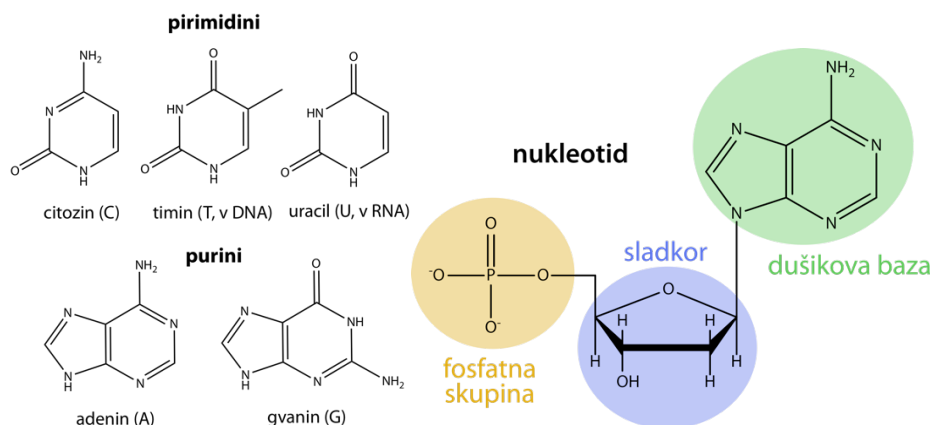
2.4. VOSKI

Voski so zmesi različnih snovi — v glavnem jih sestavljajo estri enovalentnih alkoholov z višjimi maščobnimi kislinami. Proizvajajo jih živali in rastline. Glavna sestavina satovja je **čebelji vosek**, ki ga sestavlja ester palmitinske kisline in miricil alkohola. Vosek, ki ga proizvajajo lojne žleze ovc, **lanolin**, pa je ester sterolov. Ptice iz trtične lojne žleze izločajo vosek za mazanje perja, da se jih manj prime voda. Za človeka je zanimiv **karnauba vosek**, ki ga proizvaja palma *Copernicia prunifera*, in je zaradi razširjene uporabe znan kot ti. kraljica voskov. Iz nafte pridobivajo **parafin**, ki ga prav tako uvrščamo med voske, čeprav namesto estrov vsebuje alkane. Ker so voski netopni v vodi, jih uvrščamo v skupino maščob. Njihovo tališče je nad temperaturo 40 °C. Rastline proizvajajo voske za kontrolo evaporacije in hidracije, sicer pa se voski uporabljajo v kozmetiki, prehranski industriji (npr. ovitki za sire), kot premaz oz. loščilo na lesenih izdelkih, nepremočljiv premaz na čevljih ter za izdelavo voščenk in sveč.

3. NUKLEINSKE KISLINE

Nukleinske kisline so polimeri, sestavljeni iz več monomernih enot *nukleotidov*. Med nukleinske kisline štejemo *DNA (deoksiribonukleinsko kislino)* in *RNA (ribonukleinsko kislino)*. So praktično najpomembnejše biološke makromolekule, saj so nosilke dednih informacij v celici. V evkariontskih celicah se DNA nahaja v jedru in v nekaterih drugih organelih, kot so mitohondriji in tudi kloroplasti v rastlinskih celicah. Zaporedje DNA oz. RNA, ki kodira en genski produkt (RNA oz. protein), je imenovano **gen**, celoten komplet genov organizma pa imenujemo **genom**. Gen je sestavljen iz regulatornih in prepisljivih regij. Haploidni genom človeka vsebuje približno 3×10^9 nukleotidnih parov, ki predstavljajo približno 25.000 genov, razporejenih na 23 parnih kromosomih.

3.1. PRIMARNA STRUKTURA DNA IN RNA



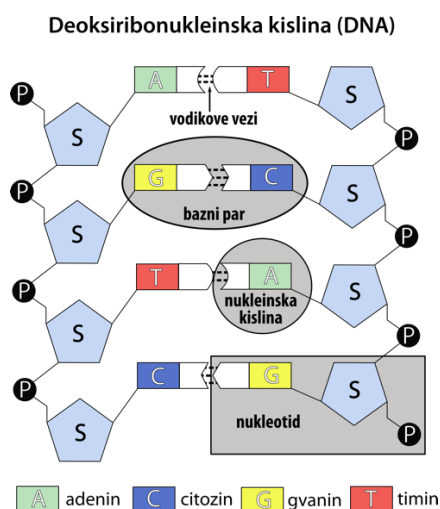
Slika 20. (Levo) Strukturne formule dušikovih baz, ki so lahko pirimidinske ali purinske. (Desno) Nukleotid je sestavljen iz fosfatne skupine, sladkorja in dušikove baze.

Nukleotidi, gradniki nukleinskih kislin, so sestavljeni iz **fosfatne skupine** in **nukleozidov** (glej Sliko 20). Le-ti vsebujejo **sladkorno enoto** (pentozo *ribozo* pri RNA oz. *deoksiribo*zo pri DNA) in nanjo z glikozidno vezjo vezano aromatsko heterociklično **dušikovo bazo**, ki je lahko *pirimidinska* (citozin in timin v DNA oz. uracil v RNA) ali *purinska* (adenin in gvanin). Namesto timina se v RNA molekuli pojavi uracil, ki je timinu strukturno podoben, vendar ne vsebuje ene

metilne skupine na aromatskem obroču. Glikozidna vez je v naravnih oligonukleotidih vedno v obliki β *stereoizomera*, pri čemer je dušikova baza nad ravnino sladkorja v isti smeri kot 5-hidroksi substituent. Zaradi drugačne sladkorne enote je RNA veliko bolj občutljiva na hidrolizo kot DNA in zato manj stabilna.

3.2. SEKUNDARNA IN TERCIARNA STRUKTURA DNA

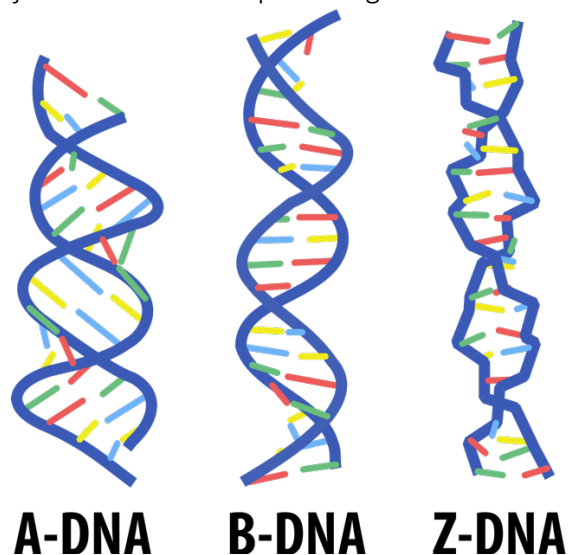
Posledica strukture nukleotidov je, da so nastajajoče polinukleotidne verige usmerjene. Med seboj so namreč nukleotidi tekoče povezani tako, da se fosfatna skupina mononukleotida, vezana na 5' mestu sladkorja, poveže s 3' mestom naslednje sladkorne enote. Na 5' koncu verige najdemo prosto fosfatno skupino, na 3' koncu pa hidroksilno skupino zadnjega dodanega nukleotida. Zaporedje oligonukleotidov običajno podajamo v enočrkovnih oznakah, kot npr. A, T, G, C, U, in ga običajno pišemo od 5' proti 3' koncu. Nastajajoča veriga raste na 3' koncu s pripenjanjem novih nukleotidov preko fosfodiestrskeske vezi. Dve komplementarni verigi DNA tvorita strukturo značilne **dvojne vijačnice**. Pri tem so fosfatne skupine in sladkorne enote na zunanji strani, dušikove baze pa so obrnjene v notranjost (Slika 21). Nukleotidi dveh verig so med seboj prečno povezani z vodikovimi vezmi – adenin se vedno poveže s timinom, gvanin pa s citozinom. To so t. i. **komplementarni bazni pari** (pogosto se uporablja okrajšava *bp*). Bazni par A-T tvori **dve** vodikovi vezi, par G-C pa **tri**.



Slika 21. DNA obstaja v obliki dvojne vijačnice, pri čemer se dve verigi povežeta preko vodikovih vezi med posameznimi nukleotidi. *P* = fosfatna skupina, *S* = sladkor.

Dvojna vijačnica DNA lahko obstaja v treh različnih strukturnih oblikah – **A-DNA**, **B-DNA** ali **Z-DNA** (Slika 22). V naravi najbolj pogosta je **B-DNA**, model dvojne vijačnice, ki sta ga leta 1953 predlagala Watson in Crick. B-DNA je desnosučna vijačnica z 10 baznimi pari na zavoju, ki v višino obsegajo 34 Å, razdalja med baznimi pari pa je 3.4 Å. Pri tem nastaneta vzdolž vijačnice *veliki* in *mali žleb* s približno enako globino, ampak z različno širino, kar vpliva na sposobnost interakcije z ostalimi molekulami. Posamezne baze v nukleinskih kislinah so ploski aromatski obroči, medsebojne povezave z vodikovimi vezmi pa niso rigidne in se lahko fleksibilno spreminjajo glede na pogoje okoliškega topila.

Zaradi neplanarnosti deoksiriboze lahko ob dehidraciji nastane **A-DNA** konformacija, ki je širša. Višina enega obrata je 26 Å, tvori ga 11 baz in ima nagnjene baze, kar rezultira v globokem in ozkem glavnem žlebu in širokem in plitkem malem žlebu, centralno pa je prazen prostor. Pomembna je za zanesljivost podvajanja DNA in je prisotna na aktivnem mestu DNA polimeraze. Občasno se pojavi tudi **Z-DNA** oblika, ki je edina levosučna oblika dvojne vijačnice. Pojavi se ob visokih koncentracijah kationov spermina in spermidina, na mestih metilacije citozina in blizu promotorjev aktivno transkriptiranih genov.



Slika 22. Konformacije A, B in Z oblike dvojne vijačnice DNA.

3.3. KVARTARNA STRUKTURA DNA

DNA se v evkariontski celici s proteini (histonskimi in nehistonskimi) povezuje v **kromatin**. Te povezave so nujne za superzvitje DNA in za regulacijo procesov v celici (aktiviranje/zavrtje transkripcije, popravilo in podvajanje DNA). DNA je v *in vivo* pogojih polianion in veže pozitivno nabite proteine. Med mejozo in mitozo se kromatin poveže še tesneje, vidni so **kromosomi**. Človek ima 23 parov kromosomov. Osnovni del kromatina so **nukleosomi** (približno 146 baznih parov DNA povezanih z 8 histoni) in so povezani s povezovalno DNA.

3.4. SEKUNDARNA IN TERCIARNA STRUKTURA RNA

Za razliko od DNA se RNA ne povezuje v dvojno vijačnico s komplementarno verigo. Vseeno pa se RNA veriga pogosto zvije in komplementarni deli tvorijo urejene sekundarne strukture - helikse. Nekateri virusi vsebujejo celotno genetsko informacijo v obliki RNA genoma. RNA je sestavljena iz štirih nukleotidov (*adenin, citozin, gvanin in uracil*). V procesu zorenja RNA pa lahko nastajajo tudi **modificirani nukleotidi** (čez 100 različnih modifikacij, katerih vloga še ni v celoti pojasnjena – npr. pseudouridin, ribotimidin), največ jih je prisotnih v tRNA. Ker je RNA negativno nabita, so za stabilizacijo sekundarnih in terciarnih struktur potrebni kovinski ioni (Mg^{2+}). Ker RNA zavzame kompleksne 3D strukture, so ugotovili, da poleg prenosa genetskega zapisa RNA lahko opravlja tudi druge funkcije, na primer katalizo reakcij. Te RNA molekule imenujemo ribocimi. Tako RNA sodeluje v *sinтеzi, regulaciji in procesiranju* proteinov.

Poznamo več tipov RNA, vsak tip služi posebni funkciji, naštetih so le nekateri:

a) sinteza proteinov

- **informacijska RNA (mRNA)** — nastane ob prepisovanju DNA in služi kot predloga za translacijo v protein, je linearna.
- **prenašalna RNA (tRNA)** — sekundarna struktura ima obliko lista deteljice in vsebuje antikodon ter na 3' (vedno CAA) delu vezano aminokislino. Ima 3D strukturo, ki se prilaga A in P mestu v ribosomu. Njena naloga je prenos aminokislina do ribosoma. tRNA vsebuje neobičajne nukleotide (pseudouridin, dihidrouridin, inozin).
- **ribosomska RNA (rRNA)** — služi kot ogrodje za ribosome, hkrati ima encimske funkcije za povezavo aminokislina v protein. Zato je 3D struktura med različnimi organizmi ohranjena.

b) postranskripcijska modifikacija RNA in replikacija DNA

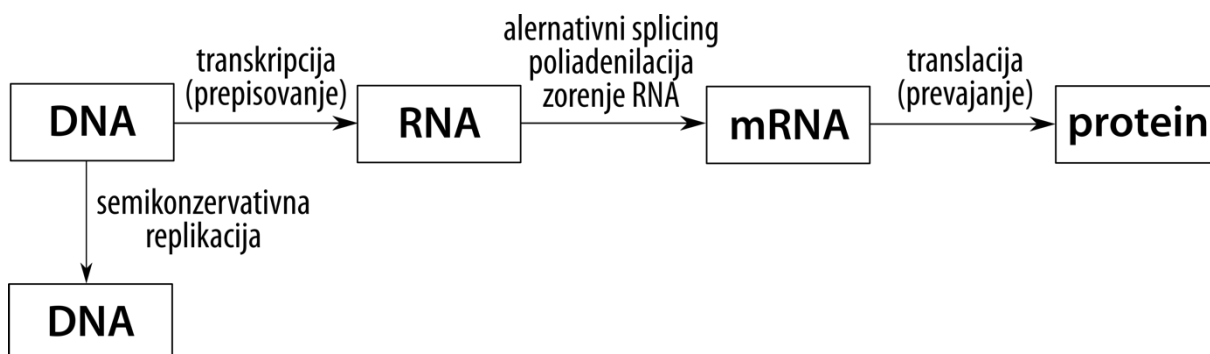
- **mala jedrna RNA (snRNA)** — v jedru celice se povezuje s proteini v *snRNP* (male nukleolarne ribonukleoproteine, ti. "snurps"), ki tvorijo spajalna telesa (v njih se iz pre-mRNA izrezujejo introni).
- **mala nukleolarna RNA (snuRNA)** — odgovorna za modifikacijo nukleotidov RNA.

c) regulatorna RNA

- **dolga nekodirajoča RNA (lncRNA)** — je kodirana tako v »med-genetskem« področju kot tudi prekrivajoče s kodirajočimi zaporedji. Dolga je več kot 200 bp in ima različne biološke funkcije (regulacija genske transkripcije, regulacija mRNA procesiranja, regulacija epigenetike) in povezave z boleznimi. Podatkovna zbirka o lncRNA: lncrna.big.ac.cn.
- **mala interferenčna RNA (siRNA)** — izhaja iz dolge dvojnivijačne RNA, ki jo endoribonukleaza (Dicer) loči na dve verigi, nakar se vodilna veriga v RISC kompleksu (t. i. "RNA induced silencing complex", ki vsebuje še proteine, med drugim encim argonavt, ki je endonukleaza, sposobna razgraditi mRNA) zelo specifično veže na eno tarčno mRNA s komplementarnim zaporedjem, ki jo tako označi za razrez in prepreči translacijo.
- **mikro RNA (miRNA)** — izhaja iz krajših RNA lasnic, ki nastanejo po razrezu pri-miRNA z encimom Drosha v jedru, pri čemer nastane pre-miRNA, ki ga odpotuje v citosol. Po razrezu z endoribonukleazami (Dicer) v kompleksu RISC najde tarče z delno komplementarnostjo (kar je glavna razlika od siRNA) in deluje na zelo različne načine (destabilizacija mRNA v RISC kompleksu, prepreči tvorbo iniciacijskega kompleksa na ribosomu, pospeši de-adenilacijo itd). Podatkovna zbirka o povezavah miRNA z boleznimi: mir2disease.org.

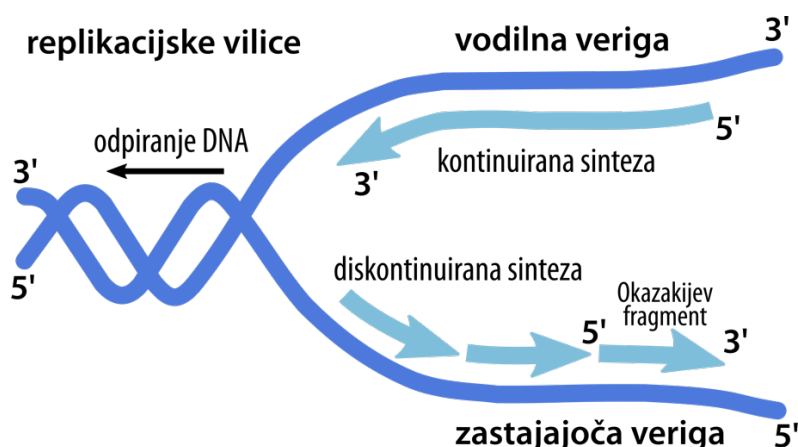
3.5. IZ DNA DO PROTEINOV – OSREDNJA DOGMA MOLEKULARNE BIOLOGIJE

Osrednja dogma molekularne biologije pravi, da se lahko informacija prenaša le v eni smeri — od DNA do proteinov (Slika 23). Pri tem so najpomembnejši procesi **podvojevanje oz. replikacija** DNA, **prepisovanje oz. transkripcija** DNA v mRNA in pa **prevajanje oz. translacija** mRNA v proteine (s pomočjo tRNA). Velja omeniti, da obstajajo posebni primeri, ko tečejo procesi v obratno smer — na primer v primeru retrovirusov iz RNA v DNA.



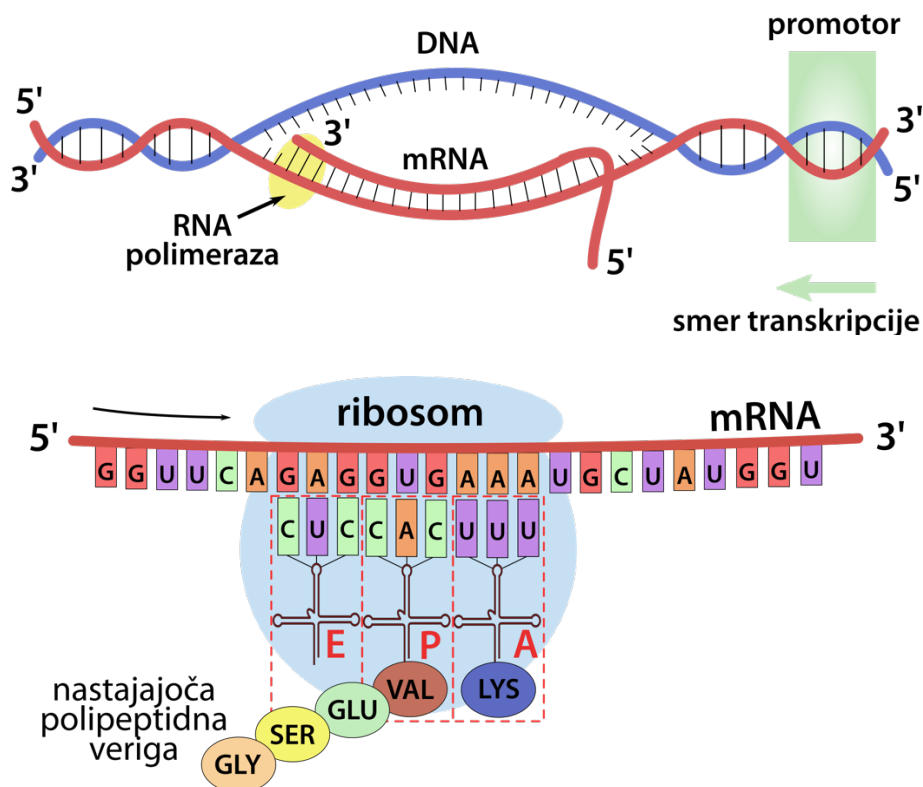
Slika 23. Osrednja dogma molekularne biologije pravi, da lahko informacija potuje samo v eni smeri — iz DNA prek RNA do proteinov.

Podvajanje oz. replikacija DNA je *semikonzervativen* proces, pri čemer se ena vijačnica ohrani, druga pa se sintetizira na novo (Slika 24). Encimi družine topoizomeraze razvijejo DNA, helikaze pa prekinejo vodikove vezi med komplementarnimi baznimi pari (odpiranje DNA). DNA se razvije v Y obliko, ki ji rečemo **replikacijske vilice**. Podvajanje sestoji iz treh stopenj: *iniciacije*, *elongacije* in *terminacije*. Za začetek podvajanja so potrebni oligonukleotidni RNA začetniki, ki se tvorijo s pomočjo encima primaza in se vežejo na verigo DNA. Encim DNA-polimeraza fosfodiesterske vezi tvori samo v smeri od 5' do 3' in zato v replikacijskih vilicah pripenja nove nukleotide na 3' konec obstoječe vodilne verige (to je veriga v 3' do 5' smeri), sinteza poteka kontinuirano. Vodilna veriga ima samo en RNA oligonukleotidni začetnik, medtem ko jih je pri zastajajoči verigi potrebnih več. Le-te imenujemo Okazakijevi fragmenti. Sinteza tako poteka diskontinuirano v več krajših segmentih.



Slika 24. Podvojevanje oz. replikacija molekule DNA.

Nato encim RNAza odstrani oligonukleotidne začetnike, drugi tip DNA-polimeraze zapolni manjkajoče dele, encim ligaza pa dokončno poveže DNA zaporedje v popolno dvojno vijačnico in zaključi podvojevanje DNA. Sledi delitev celice - mitoza. Ker na zastajajoči verigi prepis ne more poteči do konca, imajo evkariontske celice na koncih kromosomov telomere (šest nukleotidna zaporedja, ki obsegajo 10^4 - 2×10^4 baznih parov pri človeku), ki se skrajšajo z vsako celično delitvijo (kar povzroča apoptozo celice in staranje) in imajo lastni mehanizem podaljševanja.

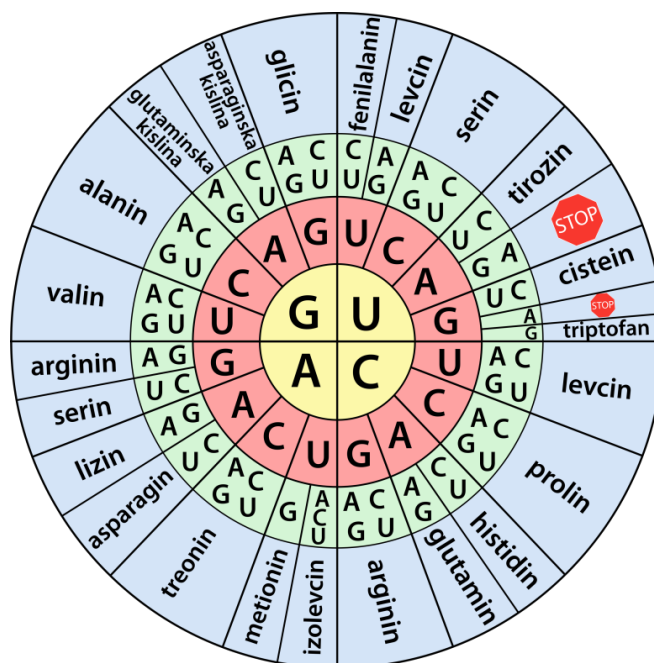


Slika 25. Poenostavljen prikaz transkripcije in translacije. (Zgoraj) DNA se prepíše v mRNA v procesu *transkripcije*. (Spodaj) mRNA se prevede v aminokislinsko zaporedje proteina v procesu *translacije*.

Za **syntezo proteinov** se mora DNA najprej prepisati v mRNA (informacijsko RNA, ki informacijo prenese iz jedra, kjer se nahaja DNA, do ribosomov, kjer se sintetizirajo proteini) s procesom **prepisovanja oz. transkripcije** (Slika 25, zgoraj), nato pa se zaporedje mRNA prevede v aminokislinsko zaporedje proteina v **procesu prevajanja oz. translacije** (Slika 25, spodaj). Najprej se pri procesu **transkripcije** po razvitju določenega dela DNA na regulatorne regije vežejo različni transkripcijski dejavniki in encim RNA-polimeraza, ki ustvari transkripcijski mehurček (10-20 baznih parov ločeni verigi DNA). V **mRNA** se prepisuje le matrična veriga DNA, in sicer v smeri 5'-3', tako da RNA-polimeraza v mRNA dodaja nukleotide na 3' koncu verige. Po transkripciji se pri evkariontih mRNA še procesira v jedru — na 5' strani dobi "*kapo*" (to je metilirani gvanozin, ki zaščiti mRNA pred razgradnjo z eksonukleazami in pomaga pri vezavi na ribosom), na 3' koncu *poliA rep* (nekaj sto adeninov, ki služijo stabilizaciji, transportu in vezavi na ribosome), z alternativnim izrezovanjem (t.i. "*splicingom*") pa se odstranijo nekodirajoči introni. Nuklearne pore prepoznajo kapo na 5' koncu in mRNA preide v citoplazmo, kjer na endoplazmatskem retikulumu pride do vezave na *ribosom*, kjer poteka **translacija**, pri kateri sodelujejo tudi molekule **tRNA**. Le-te so namenjene transportu aminokislin do ribosoma in imajo mesto za vezavo aminokislino ter mesto, ki mu rečemo *antikodon*. To je triplet RNA, ki je komplementaren tripletu mRNA, ki kodira posamezno aminokislino (temu rečemo *kodon*). Aminoacil-tRNA sintetaza sintetizira vezavo med tRNA in aminokislino, ki jo kodira antikodon — produkt take reakcije je aminoacil-tRNA molekula.

Na ribosomu so tri mesta, ki služijo vezavi tRNA — A (aminoacil mesto), P (peptidil mesto) in E (izhodno mesto). Kodon mRNA na A-mestu ribosoma se veže z antikodonom na pripadajoči

aminoacil-tRNA. S pomočjo rRNA (ribosomalne RNA) velike podenote ribosoma, se tvori peptidna vez z aktivirano aminokislino pripeto na tRNA, ki je vezana na P-mestu ribosoma. Aminokislina na P-mestu se odcepi od svoje tRNA, temu pa sledi premik mRNA. Na A-mestu pride nov kodon, prejšnji pa se premakne iz mesta A na P. tRNA brez aminokislino se premakne iz P-mesta na E-mesto. Translacija se zaključi, ko pride na A-mesto eden od *stop kodonov* UAA, UAG, UGA, ki nimajo ustreznega antikodona. Posledično se vežejo sprostitveni dejavniki, translacija se zaključi in podenoti ribosoma razpadeta. Ena mRNA se lahko hkrati prevaja na več ribosomih hkrati — tvori se polisom.



Slika 26. Genetska koda. Posamezne aminokislino so kodirane z zaporedji treh nukleotidov. Znak STOP označuje t. i. STOP kodon, ki povzroči konec procesa translacije.

Aminokislino so torej kodirane s *kodoni*. Vsak kodon je sestavljen iz treh nukleotidov. Poznamo štiri nukleotide (to so *adenin*, *citozin*, *gvadin* in *uracil*), kar pomeni, da obstaja 4^3 oz. 64 različnih kodonov, kar bi pomenilo, da bi morale obstajati tudi 64 različnih tRNA. Temu pa ni tako, zato se je razvila hipoteza o ohlapni bazi, po kateri se tretja baza (oz. nukleotid) posameznega kodona (oz. prva baza antikodona) lahko parira z neobičajnimi pari (npr. G z U, antikodon pogosto vsebuje tudi nukleozid inosin (hipoksantin), ki se lahko pari z A, C ali U), kar zmanjša učinek mutacij (npr. vsi 4 kodoni GGU, GGG, GGA in GGC kodirajo glicin), pa tudi translacija je hitrejša.

Ker obstaja le 20 proteinogenih aminokislino, je genetska koda *degenerirana*, kar pomeni, da lahko več kodonov kodira isto aminokislino (Slika 26). Genetska koda je *univerzalna*, kar pomeni da je enaka med organizmi, le mitohondrijska je nekoliko drugačna.

Tako za vsak posamezen organizem, kot tudi za njegovo potomstvo je pomembno, da DNA zaporedje ostaja nespremenjeno, saj napake lahko vodijo v celično smrt. V vsaki celici v telesu se vsak dan zgodi več tisoč napak DNA. Pri podvajanju DNA prihaja do napak približno na vsakih

10.000 baz, še več sprememb pa povzročijo produkti metabolizma (reaktivne kisikove, dušikove spojine), ksenobiotiki (snovi, ki ne nastajajo v organizmu) in okolje (UV svetloba). Pride lahko do zamenjave baz in prelomov ene ali obeh verig DNA. Zato ima organizem mnogo mehanizmov (encimov), ki zaznajo poškodbe DNA in mnogo poti, da te poškodbe popravi (npr. popravljanje DNA z izrezovanjem nepravilnih nukleotidnih parov in kemično modificiranih baz, popravljanje insercij in delecij, popravljanje prelomov enovijačne in dvovijačne DNA).

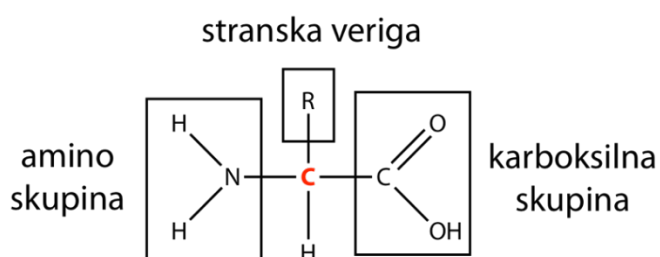
4. PROTEINI - BELJAKOVINE

Zadnja skupina makromolekul so beljakovine oz. proteini. Proteini so zgrajeni iz aminokislin in so zaradi svoje vpletenosti v številne procese znotraj in zunaj celic zelo pomembne molekule v človeškem organizmu. Med drugim sodelujejo pri katalizi reakcij, imunskem sistemu, strukturi celic, transportu molekul in signalizaciji.

Celoten nabor proteinov, ki je v določenem trenutku izražen iz genoma v nekem tkivu oz. celici, imenujemo **proteom**. Proteom je večji od genoma, kar je posledica posttranslacijskih modifikacij in spojivnih variant. Trenutno je strukturno nepoznanih 50 % regij vseh proteinov v evkariontih (za ta del uporabljajo izraz "dark" oz. temen proteom). Pri polovici le-teh gre za proteine, kjer je nepoznana celotna struktura. Veliko teh proteinov namreč ne zavzame stabilne 3D strukture in jim rečemo intrinzično neorganizirani proteini.

4.1. AMINOKISLINE

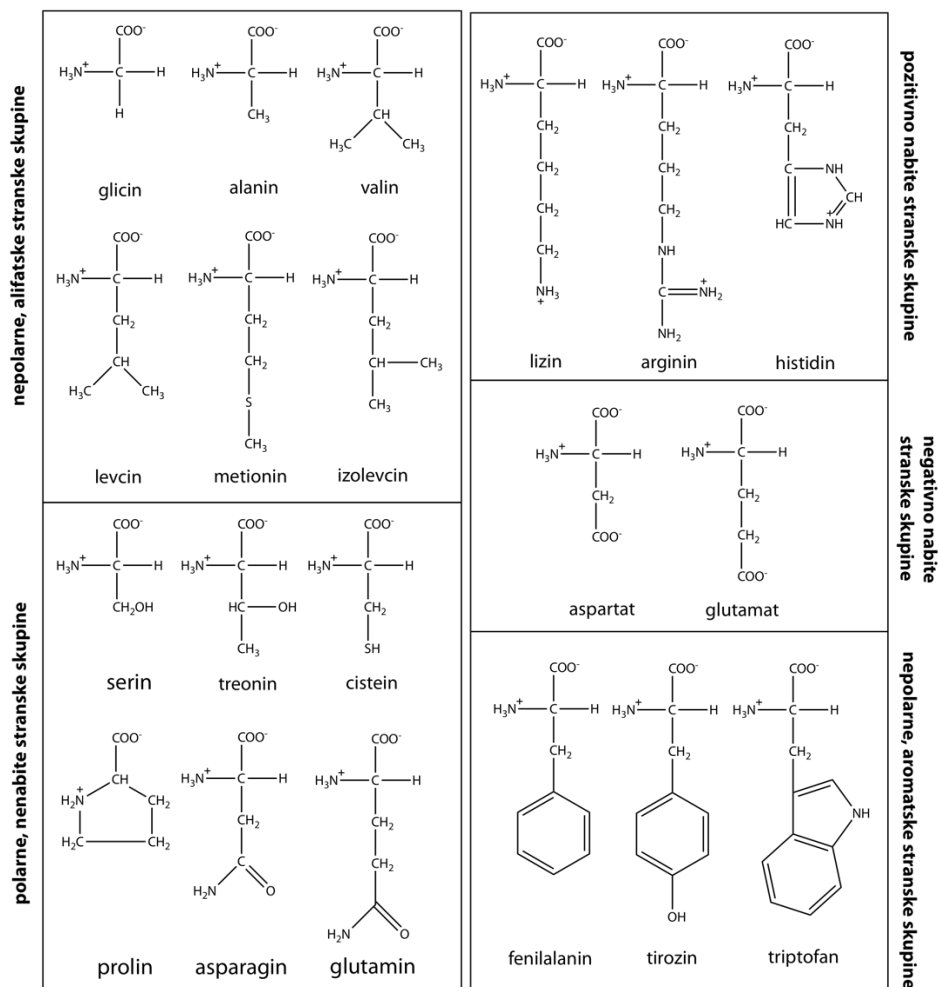
Osnovna gradbena enota proteinov so **aminokislina**. Skupno vsem proteinogenim aminokislinam je, da imajo *karboksilno* (-COOH) in *amino* (-NH₂) skupino. Na osrednji α -C atom je lahko vezana tudi stranska veriga. Stranska veriga določa lastnosti aminokislina (polarna/nepolarna, pozitivno/negativno nabita, aromatska). Večina naravnih amino kislin je v L-obliki (razen glicina, ki je akiralen). D-aminokislina se pojavljajo v bakterijski steni (Slika 27).



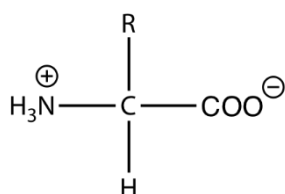
Slika 27. Aminokislina imajo na eni strani amino (N-terminalni del ali Nt), na drugi pa karboksilno skupino (C-terminalni del ali Ct). Med seboj se ločijo po raznolikih stranskih verigah.

V naravi najdemo preko 500 aminokislin. Genetska koda pa kodira 20 aminokislin, iz katerih se tvorijo proteini (Slika 26 in 28). Nekatere aminokislina so za človeka **esencialne**, kar pomeni, da jih moramo v telo vnašati s hrano, saj jih sami ne moremo sintetizirati (to so fenilalanin, valin, treonin, triptofan, izolevcin, metionin, levcin, lizin, histidin, za otroke pa so esencialne še cistein, tirozin in arginin). Aminokislina zapisujemo s *tričrkovnim* ali *enočrkovnim* zapisom npr. alanin = Ala oz. A; cistein = Cys oz. C. Obstajajo tudi ne-proteinogene aminokislina, ki se ne vgrajujejo v polipeptidne verige, kot sta GABA in L-DOPA. Le-ta se sicer vgrajuje v proteine pri

posameznikih s Parkinsonovo boleznijo in je lahko izdaten vir reaktivnih kisikovih zvrsti. Aminokislina se lahko tudi postranslacijsko spremenijo (npr. hidroksilacija prolina). Iz aminokislin v metabolizmu nastajajo druge molekule: iz triptofana in tirozina neurotransmiterji, iz histidina histamin, iz glicina hem, iz aspartata, glicina in glutamina pa nukleotidi.



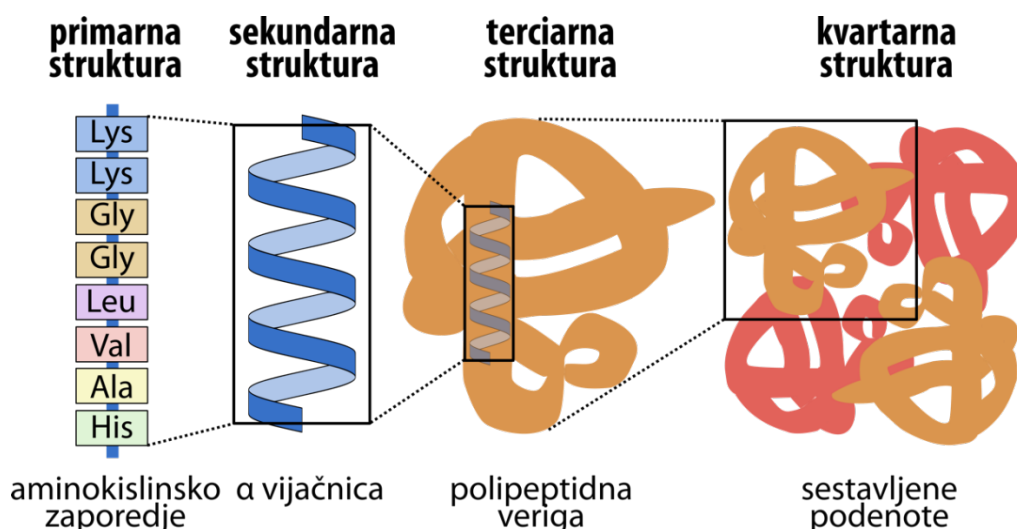
Slika 28. Proteini pri človeku so sestavljeni iz 20 aminokislin.



Slika 29. Zwitterion.

V vodni raztopini so aminokislina v obliki iona dvojčka (**zwitteriona**, Slika 29). Karkosilni ion (COO^-) je glavna oblika, če je pH nad pKa za α -karkosilno kislino (pH 2.2). Amino $-\text{NH}_2$ skupina je protonirana ($-\text{NH}_3^+$), če je pH pod pKa za α -amino skupino (pH 9.4). Z drugimi besedami, med pH 2.2 in 9.4 sta v aminokislinah večinoma prisotna $-\text{COO}^-$ in $-\text{NH}_3^+$ v dinamičnem ravnotežju, kar pomeni da je neto naboj 0. Če upoštevamo še stransko verigo, je natančen neto 0 naboj prisoten pri pH, ki mu rečemo pI (**izoelektrična točka**), pri katerem je topnost aminokislina najmanjša.

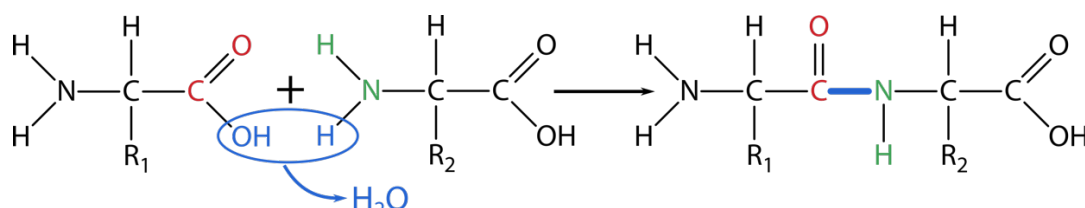
4.2. STRUKTURA PROTEINOV



Slika 30. Strukturo proteinov delimo na primarno, sekundarno, terciarno in kvartarno.

Primarna struktura

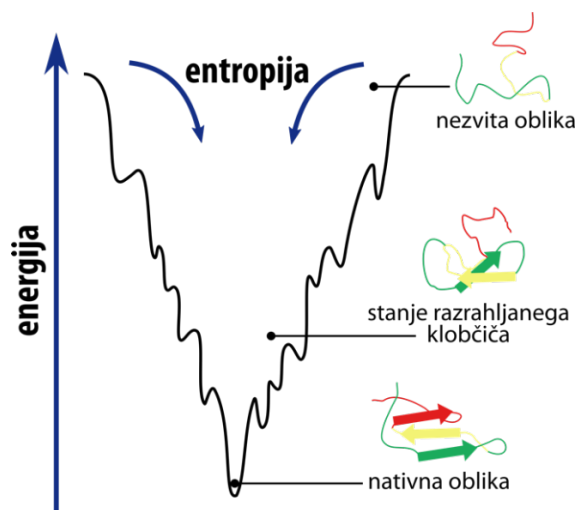
Primarna struktura je **zaporedje aminokislin** v polipeptidu, ki je določeno z zaporedjem nukleotidov v genu. Vsi proteini so sestavljeni iz različnih zaporedij 20 aminokislin, ki so med seboj povezane s *peptidno vezjo* (Slika 31). Podobno kot veriga DNA je tudi polipeptidna veriga zaradi strukture aminokislin usmerjena. Na enem koncu ima veriga prosto amino skupino, zato temu koncu rečemo N-terminalni konec, na drugem pa ima prosto karboksilno skupino oz. C-terminalni konec. Običajen zapis zaporedja vsebuje na levi strani N-terminalno aminokislino, na desni pa C-terminalno. Konformacija proteina je določena z dvema konformacijskima kotoma, o katerih bomo govorili tudi v naslednjem poglavju: phi ϕ (med C_{α} -N) in psi ψ (med C_{α} -C'), zaradi steričnih ovir pa niso vsi koti energetsko ugodni.



Slika 31. Tvorba peptidne vezi med dvema aminokislinama z reakcijo kondenzacije.

Če bi vsak aminokislinski ostanek imel na voljo le dve različni konformaciji, bi polipeptid s 100 ostanki lahko zavzel 2^{100} različnih konformacij. Za naključno iskanje nativne oz. termodinamsko najstabilnejše konformacije, ob predpostavki, da za vsako novo konformacijo porabimo le 10^{-10} s, bi tak protein torej potreboval 10 milijard let, da bi preiskali vse. To imenujemo tudi Levinthalov paradoks. Zato obstaja več hipotez o zvijanju proteinov v nativno obliko, kot sta npr. spontano ob nastanku polipeptida (kotranslacijsko) ali spontani kolaps celega polipeptida v kompaktno stanje (t. i. hidrofobni kolaps, vmes je prisotno stanje razrahljanega klobčiča oz. "molten globule", Slika 32). Nativno zvit protein ima najnižjo prosto energijo, zato so evolucijsko favorizirani tisti proteini, ki najbolj gladko prehajajo iz nezvite v nativno obliko.

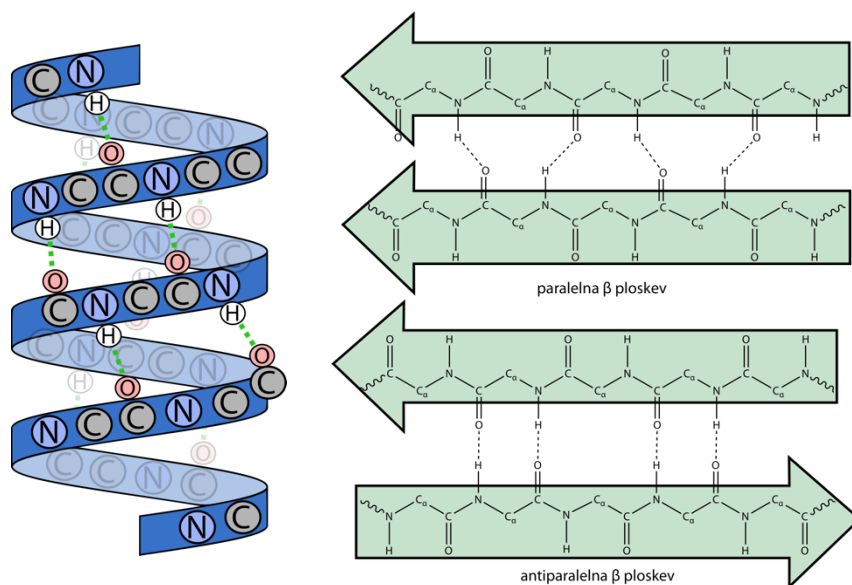
Občasno pa morajo preiti čez lokalne minimume, pri čemer pomagajo molekule imenovane šaperoni.



Slika 32. Nativna oblika proteina je termodinamsko stabilna in določena z zaporedjem aminokislin oz. s primarno strukturo proteina - Afinsenova dogma.

Sekundarna struktura

Sekundarna struktura je **lokalna ureditev** polipeptidne verige, kot posledica tvorbe vodikovih vezi med bližnjimi deli polipeptida, v značilne oblike, kot so **α -vijačnice** in **β -ploskve**. Slednje so lahko paralelne ali antiparalelne (Slika 33).



Slika 33. Posamezni deli polipeptidne verige (stranske skupine aminokislin niso prikazane) lahko tvorijo α -vijačnice (levo) ali β -ploskve (desno). Vodikove vezi, odgovorne za tvorbo teh lokalnih struktur, so označene s črčkastimi črtami.

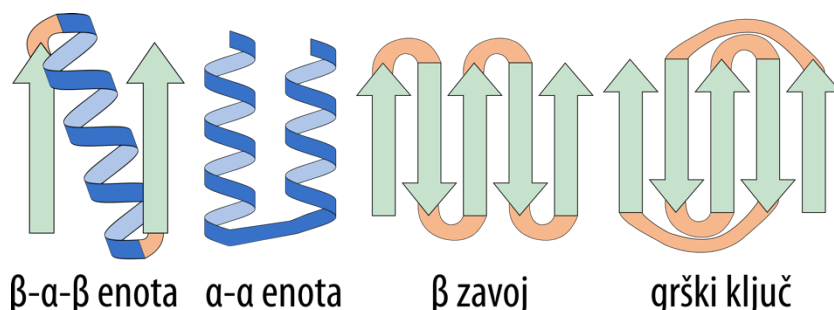
α -vijačnica je paličasta, spiralno zavita struktura, ki jo stabilizirajo vodikove vezi med C=O in N-H skupinami različnih aminokislin. Zaradi teh vezi se tvori vijačnica s 3.6 aminokislinami na

posamezen zavoj. Stranske skupine gledajo v zunanost vijačnice, kjer lahko pride tudi do interakcij med njimi. Take strukture prevladujejo v fibrilarnih proteinih (npr. keratin, kolagen).

Pri **β -ploskvah** se vodikove vezi tvorijo med C=O in N-H skupinami bolj oddaljenih aminokislin dveh vzporednih verig, pri čemer so stranske skupine spet orientirane navzven. V kolikor so aminokisliline obrnjene v isto smer, govorimo o **paralelni** β ploskvi, če pa so obrnjene obratno, govorimo o **antiparalelni** β -ploskvi. Take strukture so prisotne pri svili.

Terciarna struktura

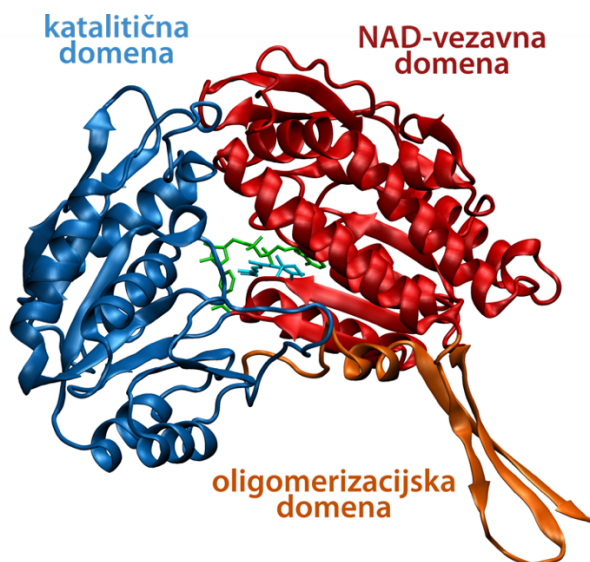
Terciarna struktura predstavlja **zvitje proteina v prostoru**. Protein je zgrajen iz kombinacije sekundarnih struktur, ki so med seboj povezane z relativno gibljivimi deli polipeptidne verige (npr. **zanke, zavoji**). Tvorijo se vodikove, ionske, hidrofobne, disulfidne vezi med različnimi deli polipeptida. Kombinacije sekundarnih struktur imenujemo **motiv** (supersekundarna struktura). Le-ti so podobni med različnimi proteini, ni pa nujno, da označujejo tudi podobno funkcijo. Primeri motivov so: β - α - β - α - β motiv, motiv *omega* (to je katerokoli zaporedje aminokislin, ki tvori stabilno obliko grške črke omega), motiv samih α oblik (*coiled coil*), α/β valjev, *podkev* in β valjev, *grškega ključa* (Slika 34). Podatkovni bazi **CATH** (www.cathdb.info) in **SCOP** (scop2.mrc-lmb.cam.ac.uk) razvrščata proteine glede na njihovo strukturo in strukturne elemente, pa tudi glede na funkcijo in evolucijo.



Slika 34. Nekateri motivi, ki jih najdemo v proteinih: β - α - β motiv, α - α motiv, β zavoj in motiv grškega ključa.

Domena rečemo delu polipeptidne verige, ki se zvije samostojno in tudi opravlja določeno nalogo neodvisno, sestavljen pa je iz več motivov (Slika 35). Velika večina proteinov je sestavljenih iz več domen. Npr. NAD-vezavna domena je sestavljena iz β - α - β - α - β motiva in je prisotna tako v proteinu aldehyd-dehidrogenazi kot tudi v alkohol-dehidrogenazi in laktat-dehidrogenazi.

Porušenje terciarne in sekundarne strukture proteina imenujemo **denaturacija**. Povzročijo jo lahko *spremembe pH* (spremenijo se elektrostatske interakcije med nabitimi stranskimi skupinami aminokislin — pod izoelektrično točko je protein pozitivno nabit, nad izoelektrično točko pa negativno), *višje temperature* (cistein tvori S-S vezi — keratinu v laseh omogoča termostabilnost in netopnost v toplilih) in *organska topila* (porušijo se hidrofobne interakcije, tvorijo se nove vodikove vezi med molekulo topila in proteinom).



Slika 35. Proteini so večinoma sestavljeni iz večih domen, ki imajo svoje naloge. Na sliki je aldehid-dehidrogenaza v kompleksu z NAD⁺ in retinojsko kislino (PDB koda proteina 5FHZ).

Kvartarna struktura

Številni proteini imajo le eno polipeptidno verigo in tako samo tri že opisane nivoje strukture. Kvartarna struktura pa predstavlja medsebojno ureditev več različnih polipeptidnih podenot v prostoru (Slika 30, desno). Primer je hemoglobin, ki ga sestavljajo štiri polipeptidne verige (po dve enaki), pa tudi encim DNA polimeraza, ki ga sestavlja deset podenot.

Posttranslacijske modifikacije

Dokončno funkcijo pa proteini pridobijo s posttranslacijskimi modifikacijami. To so večinoma encimsko katalizirane modifikacije, pri katerih se doda kovalentno vezan del. Pogoste so *fosforilacija* (dodajanje fosforilne skupine), *acetilacija* (dodajanje acetilne skupine), *hidroksilacija prolina* (nastane hidroksiprolin z dodatno -OH skupino, ki predstavlja približno 4 % aminokislin v živalskem tkivu), *glikozilacija* (dodajanje sladkorjev), *lipidacija* (dodajanje hidrofobnih skupin). Med posttranslacijske modifikacije prištevamo tudi strukturne spremembe, kot je proteolitična cepitev verige (npr. proteini odgovorni za strjevanje krvi, inzulin), pa tudi nastanek disulfidne vezi med cisteini in še dodajanje prostetičnih skupin (npr. nekovalentna vezava železa v hem ali kovalentna v citokrom).

4.3. FUNKCIJE PROTEINOV

Protein lahko opravlja svojo nalogo le, če ima pravilno 3D obliko. Že iz oblike lahko pogosto sklepamo na nalogo, ki jo protein opravlja. Funkcija večine proteinov je odvisna od njihove sposobnosti vezave različnih ligandov. S tem sta povezani dve lastnosti, ki karakterizirata interakcije proteinov in ligandov. Ena je **afiniteta vezave**, to je moč vezave, ki jo podajamo s *konstanto ravnotežja* (K_{eq}) ali s *konstanto disociacije* (K_d), druga pa je **specifičnost vezave**. Največja specifičnost in afiniteta vezave se pojavita, ko je struktura liganda komplementarna vezavnemu mestu.

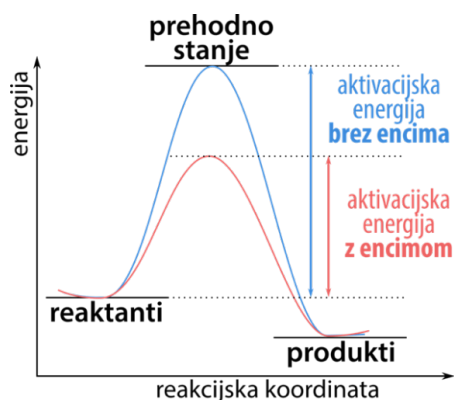
Proteini v telesu opravljajo številne pomembne funkcije (Preglednica 2). Pomembnejše skupine proteinov so (1) encimi, ki katalizirajo kemijske reakcije v človeškem telesu, (2) proteini, ki sodelujejo v celični signalizaciji in vezavi ligandov (sem spadajo razni receptorji, protitelesa in hormoni), ter (3) strukturni proteini, ki dajejo celicam obliko in trdnost.

Preglednica 2. Poznamo različne tipe proteinov, ki v organizmu opravljajo specifične funkcije.

Tip proteina	Funkcija	Primer
Receptorji	Odziv na dražljaje	Dopaminski, adrenergični
Encimi	Pospeševanje kemičnih reakcij	Prebavni encimi
„Obrambni“ proteini	Obramba organizma	Protitelesa, strjevanje krvi
Strukturni proteini	Podpora	Svila, kolagen, keratin
Skladiščni proteini	Shranjevanje amino kislin	Ovalbumin, kazein
Transportni proteini	Transport drugih molekul	Hemoglobin, feritin
Kontraktilni proteini	Gibanje	Miozin, aktin
Hormoni	Ohranjanje homeostaze	Insulin
Regulatorji genske aktivnosti	Transkripcijski faktorji	NF- κ B

4.3.1. Encimi

Encimi so proteini, ki **katalizirajo reakcijo** substrata z **nižanjem aktivacijske proste energije**, pri čemer se sami ne spremenijo (Slika 37). Posledično taka reakcija poteka hitreje oz. pri pogojih (temperatura, tlak), ki so združljivi z življenjem. Vsak encim ne more katalizirati vsake reakcije, ampak so pri svojem delovanju **specifični**, kar je posledica njihove tridimenzionalne strukture. Večina encimov ima t. i. *aktivno mesto*, v katero se veže posamezen substrat. Obliko in sestavo aktivnega mesta lahko pametno izkoristimo tudi za načrtovanje specifičnih zdravilnih učinkovin, saj so inhibitorji encimov analogi prehodnega stanja.



Slika 37. Encimi znižajo aktivacijsko energijo biokemijske reakcije.

Vsak encim lahko predstavimo z njegovim lastnim številom **EC** (Enzyme Comission). Tako mu pripišemo štiri številke, med katerimi prva grobo umešča encim v eno izmed sledečih šestih skupin, glede na tip reakcije, ki jo katalizira:

- *EC 1: oksidoreduktaze, ki katalizirajo reakcije oksidacije oz. redukcije (npr. dehidrogenaze, oksidaze)*
- *EC 2: transferaze, ki katalizirajo prenos funkcionalnih skupin iz ene molekule na drugo (npr. transaminaze, kinaze)*

- *EC 3: hidrolaze, ki katalizirajo hidrolizo raznih vezi (npr. lipaze, amilaze, peptidaze, fosfataze)*
- *EC 4: liaze, ki katalizirajo razcep vezi, ki ne poteka s hidrolizo ali z oksidacijo (npr. dekarboksilaze)*
- *EC 5: izomeraze, ki katalizirajo prenos skupin znotraj molekule (npr. izomeraze, mutaze)*
- *EC 6: ligaze, ki katalizirajo spajanje dveh molekul s kovalentnimi vezmi (npr. sintetaze)*
- *EC 7: translokaze, ki katalizirajo prenos ionov ali molekul preko membrane*

Tako lahko npr. encim laktat dehidrogenazo opišemo z EC številom 1.1.1.27, encim amilazo s številom 3.2.1.1, encim holin acetiltransferazo pa s številom 2.3.1.6.

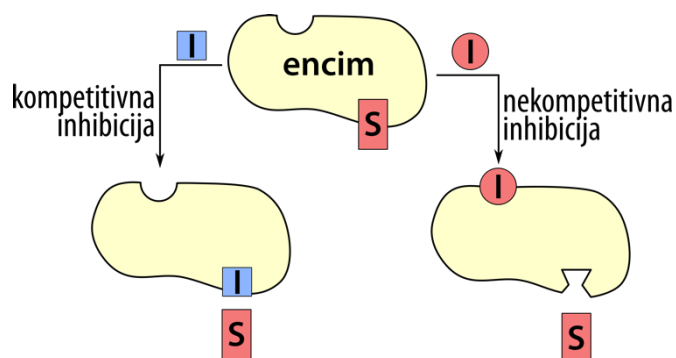
Veliko encimov za svoje delovanje potrebuje **kofaktor**. To je ne-aminokislinska molekula, ki sodeluje v kemijski reakciji, ki ne more poteči s katero od dvajsetih aminokislin. Kofaktorji so lahko *anorganski* (kovine, npr. železo ali cink) ali *organske* molekule (tj. koencimi, kot so FAD = flavin adenin dinukleotid, NAD = nikotin adenin dinukleotid in TPP = tiamin pirofosfat). **Apoencim** imenujemo (neaktivni) encim brez kofaktorja, apoencim skupaj s kofaktorjem pa **holoencim**.

Na delovanje encimov vpliva več različnih dejavnikov, kot so temperatura, pH, koncentracija substrata in prisotnost inhibitorjev. Višja **temperatura** ponavadi pomeni tudi večjo hitrost reakcije, ob prisotnosti encimov ali brez. Potrebno pa je poudariti, da imajo encimi nek temperaturni optimum, pri katerem najbolje delujejo in da tako kot vsi proteini pri višjih temperaturah denaturirajo, tj. izgubijo svojo obliko in s tem tudi aktivnost. Pri običajnih živalskih proteinih se to zgodi nad približno 40 °C. Obstajajo pa tudi encimi, pri katerih je ta meja znatno višja, npr. pri nekaterih termofilnih bakterijah (kot primer lahko navedemo DNA polimerazo I iz *Thermus aquaticus*), ki živijo na geotermalnih območjih in prenesejo temperature tudi do 120 °C. Temperaturno neodvisne oz. manj odvisne so encimske reakcije pri nekaterih psihrofilnih (nizkotemperaturnih) organizmih, pri katerih je glavni prispevek k prosti energiji aktivacije entropijski.

Podobno kot za temperaturo, velja tudi za **pH**, da imajo encimi nek optimalen pH, pri katerem najbolje delujejo. Ta optimalna vrednost se od encima do encima razlikuje. Tako npr. pepsin, ki deluje v želodcu, svojo maksimalno aktivnost dosega pri pH vrednostih okoli 2.0. V bolj bazičnem okolju bo neaktiven, pri pH nad 8.0 bo denaturiral. Ravno obratno velja za tripsin, ki deluje v tankem črevesju, in svojo optimalno aktivnost dosega pri pH vrednostih okoli 8.0.

Z večanjem **koncentracije substrata** se večja hitrost tvorbe kompleksov encim-substrat in s tem tudi aktivnost encima. Seveda le do takrat, ko so vsa aktivna mesta encima zasedena z molekulami substrata, tj. pride do nasičenja encima. Takrat se tudi aktivnost encima ustali.

Encimske reakcije lahko upočasnijo ali celo ustavijo **encimski inhibitorji** oz. **zaviralci** (Slika 38). Le-ti so lahko *kompetitivni* (če tekmujejo s substratom za vezavno mesto) ali *nekompetitivni* (vežejo se na drugo vezavno mesto kot substrat, kar konformacijsko spremeni aktivno mesto encima) in *reverzibilni* (vezava je začasna, nekovalentna) ali *ireverzibilni* (ostanejo kovalentno vezani, tak encim ne deluje več). Poznamo tudi **alosterične aktivatorje**, ki se vežejo na encim in konformacijsko spremenijo aktivno mesto k delovanju.



Slika 38. Inhibitorji so lahko kompetitivni ali nekompetitivni.

Encimi so posebej pomembni v metabolizmu (npr. citokromi, proteaze), pa tudi pri prenašanju signalov in celični regulaciji (npr. kinaze, fosfataze).

Citokrom P450 je družina oksidaznih encimov, ki vsebujejo kofaktor *hem* – pomembna je pri sintezi hormonov, holesterola ter metabolizmu toksičnih endogenih in eksogenih substanc. Tako *Cyp3A4* katalizira pretvorbo lipofilnih molekul v bolj hidrofilne. Njegovi substrati so kofein, teofilin, testosteron, med drugim pa metabolizira tudi 50 % vseh uporabljenih zdravil. Nahaja se v jetrih, kjer omogoča, da se snovi iz krvi lahko izločijo skozi urin, in v tankem črevesju. Resveratrol iz rdečega vina in furanokumarini iz soka grenivke so ireverzibilni, *samomorilski inhibitorji* *Cyp3A4*, kar pomeni, da postanejo reaktivni šele med samo katalitično reakcijo ("se porabijo") in se ireverzibilno vežejo na encim.

Okrog 5 % zdravilnih učinkovin učinkuje na **proteaze**. To so encimi, ki povzročijo *hidrolizo peptidne vezi* v proteinih. Evolucijsko so se razvile večkrat, zato obstaja več tipov proteaz z različnim mehanizmom cepitve peptidne vezi. Največkrat zdravilna učinkovina posnema strukturo substrata in z njim tekmuje za vezavo v aktivno mesto encima. Primer takih zdravilnih učinkovin so zaviralci encima angiotenzin konvertaze (ACE) in HIV-proteaz. Glavni problem pri razvoju novih učinkovin pa je doseganje selektivnosti za vezavo v aktivno mesto encima.

Druga velika družina zdravilnih učinkovin deluje na encime **kinaze**. To so encimi, ki *povzročijo fosforilacijo hidroksilnih skupin na aminokislinah*. Glede na to, katero aminokislino fosforilirajo, jih delimo na **tirozinske kinaze** in **serin/treoninske kinaze**. Kinaze so pomembno vpletene v prenos signalov, povezanih s celičnim ciklom, apoptozo in diferenciacijo, in zato z mnogimi bolezenskimi stanji, še posebej pa z rakom. Kinaze vsebujejo manjšo domeno, na katero se veže ATP, in večjo domeno na katero se veže peptid. V aktivnem mestu se zgodi prenos zadnje fosfatne skupine ATP na peptid — nastaneta ADP in fosforiliran peptid. V glavnem se kot zdravilne učinkovine razvijajo kinazni inhibitorji, ki zavirajo delovanje preveč delujočih kinaz — večina razvitih učinkovin tekmuje z ATP za vezavno mesto (inhibitorji tipa I), medtem ko se drugi vežejo na nekatalitično mesto v neaktivni kinazi (inhibitorji tipa II).

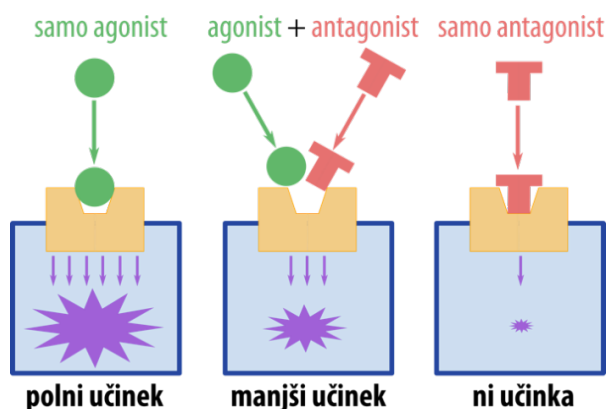
4.3.2. Receptorji

Receptorji so molekule v celični membrani, citosolu ali jedru, ki vežejo ligande in se s tem tako spremenijo, da aktivirajo ali inhibirajo biokemijsko pot ter tako sodelujejo v procesih **celične signalizacije**. V grobem lahko receptorje razdelimo na intracelularne, ki se nahajajo znotraj celice v citoplazmi ali v jedru (npr. jedrni receptorji), in na tiste, ki se nahajajo v celični

membrani in povezujejo celično zunanost z njeno notranostjo (npr. ionski kanali, tirozin kinazni receptorji, z G-proteinom sklopljeni receptorji).

Prvi korak celične signalizacije je vezava liganda na receptor. Podobno kot encimi so tudi receptorji specifični za posamezne ligande. Ligandi so lahko endogeni ali eksogeni, proteini, hormoni, živčni prenašalci, zdravila ali toksini. Za vezavo na intracelularne receptorje morajo ligandi preiti celično membrano, zato so to tipično majhne hidrofobne molekule (npr. steroidi). Za vezavo na površinske receptorje prehod celične membrane ni nujen, zato so taki ligandi lahko tudi polarni ali nabiti.

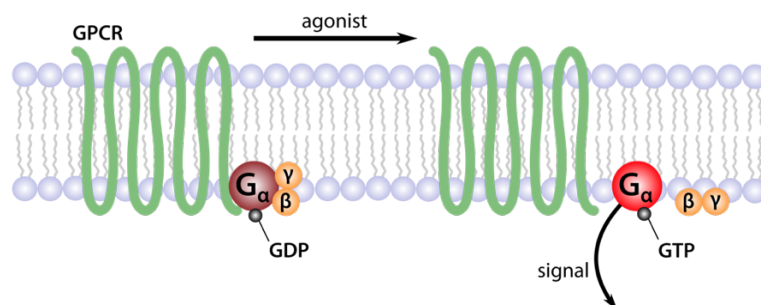
Ligande ločimo na agoniste in na antagoniste (Slika 39). **Agonist** je ligand, ki aktivira receptor in sproži biološki odgovor. Endogeni agonist z največjo učinkovitostjo je definiran kot *polni agonist*. Agonisti so lahko *kompetitivni*, ki tekmujejo za vezavno mesto z endogenim ligandom, ali pa *alosterični regulatorji*, ki se vežejo na drugo mesto kot endogeni ligand in spremenijo vezavno mesto za ligand tako, da je afiniteta vezave večja. *Delni* oz. *parcialni agonisti* sprožijo odziv, vendar tudi če so vsi receptorji zasedeni z delnim agonistom, ne more priti do 100 % učinka. **Antagonist** je ligand, ki se na receptor veže, ne sproži pa nobenega učinka, kar povzroči inhibicijo signalne poti, saj ne more priti do vezave agonista. Antagonisti so lahko *kompetitivni* ali *nekompetitivni* in *reverzibilni* ali *ireverzibilni*. Tipično se antagonisti vežejo močneje kot agonisti in jih lahko izpodrinejo.



Slika 39. Glede na učinek ki ga sprožijo, ločimo ligande na agoniste in na antagoniste.

Z G-proteinom sklopljeni receptorji (GPCR)

Predvideva se, da človeški genom kodira približno 800 z **G-proteinom sklopljenih receptorjev** oz. **GPCR**, ki so najpogostejše tarče zdravilnih učinkovin. Z G-proteinom sklopljeni receptorji (Slika 40) sedemkrat preidejo celično membrano, znotraj celice pa imajo vezan G-protein, ki sestoji iz treh različnih podenot α , β in γ (kar imenujemo heterotrimer). Po vezavi liganda na del receptorja na zunanji strani celice pride do konformacijske spremembe receptorja. Posledično se znotraj celice odcepi α -podenota G-proteina z vezanim GTP, ki naprej aktivira različne poti (npr. *cAMP* ali *fosfatidilinozitol* signalno pot). Po kaskadi reakcij to v končni fazi pripelje do spremenjenega izražanja genov, spremenjene permeabilnosti membrane, spremenjenega znotrajceličnega metabolizma in podobno. V to skupino spadajo npr. beta adrenergični receptorji, rodopsin, dopaminski receptorji, kot tudi nekateri serotoninški receptorji.



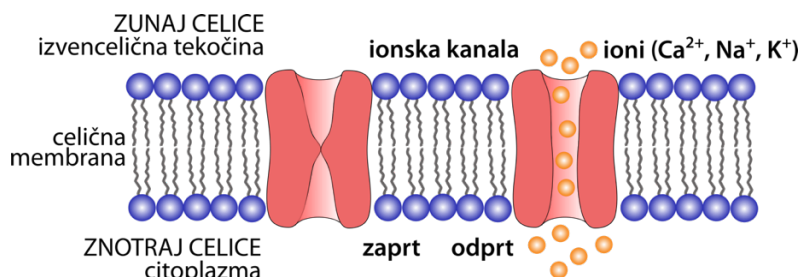
Slika 40. Ob vezavi agonista na z G-proteinom sklopljeni receptor (GPCR) pride do konformacijske spremembe receptorja in aktivacije G-proteina na notranji strani membrane.

Tirozin kinazni receptorji

Tirozin kinazni receptorji spadajo med receptorje, ki so na notranji strani sklopljeni z encimom. Kinaze so namreč encimi, ki katalizirajo prenos fosfatne skupine. Po vezavi liganda na ekstracelularno domeno receptorja tipično pride do foforilacije tirozina na notranji strani, kar omogoči prenos signala naprej. Kot omenjeno, so vpleteni v regulacijo ravninskih dejavnikov, primer tirozin kinaznega receptorja pa je tudi inzulinski receptor. Trenutno zaradi njihove vpletenosti v regulacijo ravninskih dejavnikov razvijajo veliko zdravilnih učinkovin z delovanjem na **tirozin kinazne receptorje**, predvsem kot potencialne protirakave učinkovine.

Ionski kanali

Ionski kanali (Slika 41) so proteini, ki omogočajo **prehod ionov** (npr. kalija, natrija, kalcija, klor) čez celično membrano in s tem prenos signala. Anatomsko se ionski kanali pri mieliniziranih nevronih nahajajo v območju Ranvierjevih zažemkov. Ionskih kanalov je veliko različnih vrst, v kontekstu receptorjev pa govorimo o od liganda odvisnih ionskih kanalih. Prisotni so v vseh celicah v organizmu in sprožajo zelo različne odzive, zaradi česar so zanimive farmakološke tarče. Njihova slaba stran pa je, da se mnogo molekul nanje veže nespecifično, kar povzroči veliko neželenih učinkov. Vezava ligandov povzroči *stabilizacijo določene konformacije* ionskega kanala (odprta/zaprta za prehod ionov) ali *spremembo elektrostatičnih lastnosti* kanala (kar omogoči/prepreči pretok ionov). Primeri ionskih kanalov so tarče hitro delujočih neurotransmiterjev (nikotinski receptor za acetilholin, GABA_A receptor, serotoninški 5-HT₃ receptor) in so vpleteni v prenos živčnega signala. Le-ta ponavadi poteka tako, da presinaptični nevron v sinaptično špranjo sprosti določeno količino neurotransmiterjev, ki se nato vežejo na ionski kanal na postsinaptičnem nevronu. Konformacija receptorja se spremeni in ioni lahko prehajajo čez membrano, kar vodi bodisi do depolarizacije ali do hiperpolarizacije. Posebej pomembni so od napetosti odvisni natrijevi kanali, ki so mesto za vezavo lokalnih anestetikov.

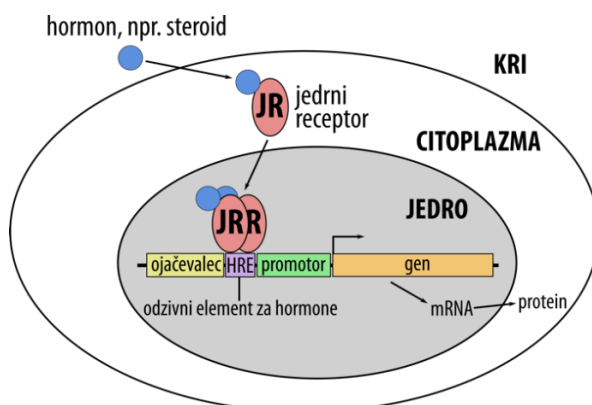


Slika 41. Ionski kanali s spremembo konformacije omogočijo prehod ionov čez celično membrano.

Vezava GABA (γ -aminomaslena kislina) na zunajcelični del GABA_A receptorja povzroči konformacijsko spremembo transmembranskega dela, ki je sestavljen iz petih podenot, vsaka podenota pa iz štirih α -heliksov. To spremeni prehod ionov čez membrano oz. povzroči prehod Cl⁻ ionov čez membrano, kar povzroči hiperpolarizacijo nevrona. S tem se zmanjša možnost za uspešno sprožitve akcijskega potenciala, potrebnega za prenos živčnega signala po nevronu. Tako deluje GABA kot inhibitorni neurotransmiter. Na isti receptor, a na drugo mesto kot GABA, se vežejo tudi etanol in nekatera zdravila.

Jedrni receptorji

Jedrni receptorji so velika družina **transkripcijskih dejavnikov**, ki se nahajajo v citoplazmi (znotraj celice), po vezavi ligandov pa odpotujejo v jedro in se vežejo na DNA ter tako direktno **vplivajo na izražanje genov** (Slika 42). Sem štejemo receptorje za steroidne, kortikoidne in ščitnične hormone ter tudi receptorje za vitamin A in D. Za učinkovine, ki naj bi delovale nanje, je za doseg učinka bistven dober prehod čez membrano celice.

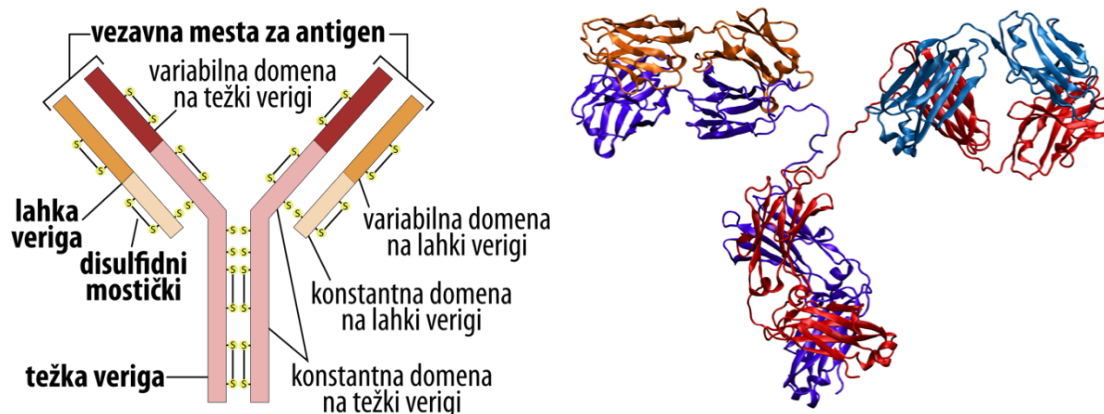


Slika 42. Jedrni receptorji po vezavi ligandov vplivajo na izražanje genov.

4.3.3. Protitelesa

Protitelesa pripadajo družini proteinov, imenovani *imunoglobulini*. Proizvaja jih imunski sistem in so namenjena **obrambi organizma pred škodljivimi tujki** (antigeni). Ko se taki tujki (npr. virusi, bakterije, glive, itd.) znajdejo v telesu, jih bo imunski sistem s pomočjo protiteles "označil" za uničenje.

Protitelesa so sestavljena iz dveh *težkih* (vsaka ima eno variabilno in tri ali več konstantnih domen) in dveh *lahkih verig* (vsaka ima eno konstantno in eno variabilno domeno), ki so med seboj povezane z disulfidnimi mostički (Slika 43). Mesto, ki ga protitelesa na antigenu prepoznajo, imenujemo **epitop**, in je komplementaren **paratopu** na variabilni domeni obeh verig na Fab regiji. Temu delu rečemo komplementarnost determinirajoča regija ali CDR (ang. *complementarity-determining region*). Drugi del, kjer sta povezani težki verigi, se imenuje Fc, in je odgovoren za učinke protiteles (z njim se veže na Fc receptorje, na komplement ipd). Imunoglobulinska domena, ki jo tvori »sendvič« dveh β -ploskev, v katerih je povezanih sedem (konstantna domena) ali devet (variabilna domena) antiparalelnih β -verig v obliki grškega ključa, je prisotna v več kot 200 človeških proteinih (npr. nekateri citokinski receptorji, adhezivne molekule, kostimulatorne molekule).



Slika 43. (Levo) Protitelesa so sestavljena iz lahkih in težkih verig. (Desno) Kristalna struktura imunoglobulina (PDB koda 1IGT).

Po izpostavitvi antigenu aktivirane B celice, imenovane tudi plazmatke, tvorijo protitelesa, kot del zaščitnega odgovora. Protitelesa so poliklonska. To pomeni, da vsaka B celica proizvaja protitelesa proti drugemu delu antigena, zato so vezavna mesta na protitelesih različna. V laboratorijih so s fuzijo B celice in nesmrtni mielomske celice uspeli vzpostaviti hibridne celice, ki omogočajo gojenje monoklonskih protiteles. Vse so potomke ene B celice in imajo isto vezavno mesto. Protitelesa se uporabljajo v **terapevtske** namene kot del bioloških zdravil in v **diagnostične/raziskovalne** namene. Protitelesa so visoko specifična, saj lahko »ločijo« med fosforiliranim in nefosforiliranim proteinom.

Andrej Perdih

določevanje 3d strukture bioloških molekul s kristalografijo

```
function quickSort(array)
var list less, equal,
if length(array) ≤ 1
return array
select a pivot value pivot
for each x in array
if x < pivot then less
if x = pivot then equal
if x > pivot then greater
return concatenate(quickSort(less), equal, quickSort(greater))
```

1. ZGODOVINA IN VLOGA KRISTALOGRAFIJE

Kristalografija je metoda, ki omogoča določitev tridimenzionalne (3D) razporeditve atomov znotraj kristala, ki ga obsevamo z x-žarki. Le-ti se ob interakciji s kristalom razpršijo v specifične smeri, odvisne od strukture kristala in molekule, ki jo vsebuje. S poznavanjem kotov sipanih žarkov ter njihovih faz lahko kristalograf določi tridimenzionalno sliko na osnovi poznane elektronske gostote proteina v kristalu. Metodo so najprej uporabljali za reševanje struktur organskih in anorganskih molekul z manjšo molekulsko maso, kasneje pa se je pričela uporabljati tudi za določevanje struktur bioloških molekul (proteinov, RNA, DNA). V tem poglavju se bomo osredotočili na uporabo te metode za določevanje 3D strukture proteinov – proteinsko kristalografijo.

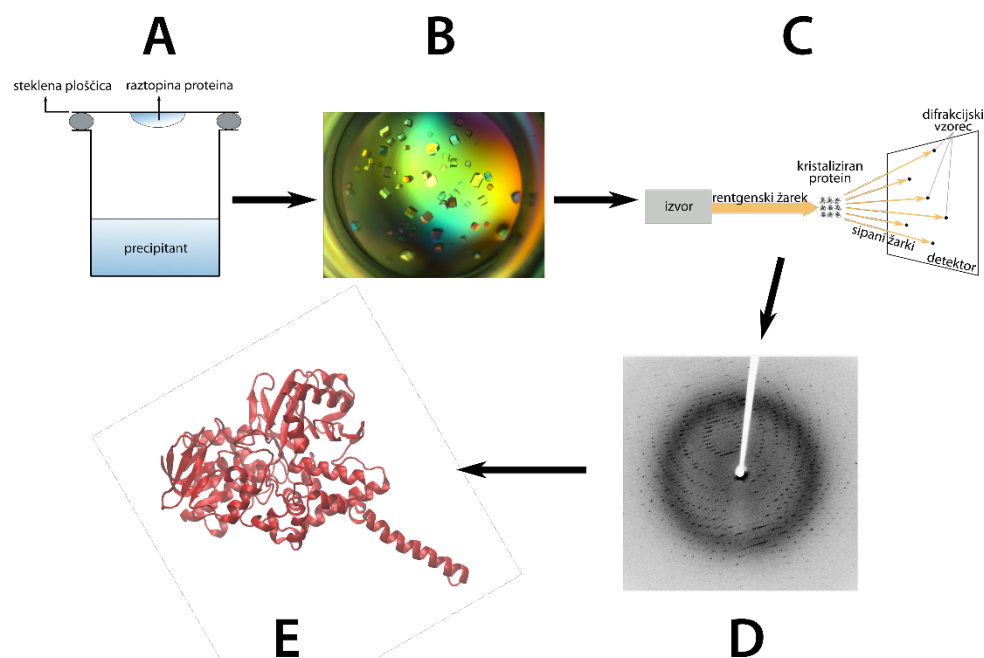
Čeprav so raziskovalci študirali kristalizacijo proteinov že v drugi polovici 19. stoletja, za enega izmed pomembnih korakov začetka razvoja metode lahko štejemo leto 1934, ko sta John D. Bernal in Dorothy Crowfoot prva izmerila difrakcijo kristalov pepsina v raztopini, ki sta jih obsevala z rentgenskimi x-žarki. Prve kristalne strukture proteinov so rešili v poznih petdesetih letih prejšnjega stoletja. Prva struktura proteina mioglobina je bila objavljena leta 1958. Za to odkritje in vpeljavo metode sta leta 1962 **John Cowdery Kendrew** in **Max Perutz** dobila Nobelovo nagrado za kemijo. Od takrat so raziskovalci določili že več kot 150.000 kristalnih struktur proteinov, nukleinskih kislin, pa tudi virusov in drugih zanimivih bioloških molekul, ki so dostopne v proteinski banki struktur PDB. Druga metoda, **NMR spektroskopija**, ki jo bomo spoznali v Poglavju III, je po obsegu določila le kakšnih 10 % toliko struktur bioloških molekul kot kristalografija.

Pomembna prednost kristalografije je, da lahko določi strukturo praktično poljubno velike biološke molekule, medtem ko je NMR spektroskopija omejena na nekoliko manjše biološke sisteme velikosti nekje do 70 kDa. Rentgenska kristalografija se tako danes rutinsko uporablja v raziskavah struktur bioloških molekul. Poleg velike količine pridobljenega bazičnega znanja o strukturnih lastnostih bioloških molekul rentgenska kristalografija omogoča vpogled tudi v vezavo ligandov (npr. fizioloških ligandov, zdravilnih učinkovin), na njihove tarče (npr. receptorje, encime, ionske kanale, itd.). Ti podatki omogočajo strukturno podprto načrtovanje novih zdravilnih učinkovin in več o uporabi proteinskih struktur v ta namen bomo spoznali v Poglavju VI.

Kljub obsežnemu napredku tehnologije, povezane s to metodo, nekatere biološke makromolekule, predvsem membranski proteini, ostajajo velik izziv za kristalizacijo. Le-ti namreč potrebujejo specifične detergente ter druga sredstva za ustrezno solubilizacijo proteina, taki detergenti pa pogosto motijo proces kristalizacije. Membranski proteini pa predstavljajo pomembne skupine proteinov v telesu, kot so na primer ionski kanali in različni receptorji za nevrottransmitterje. V zadnjih letih pridobiva na pomenu tudi **krioelektronska mikroskopija**, ki tudi omogoča določitev 3D struktur velikih bioloških makromolekul, tudi proteinskih kompleksov.

2. EKSPERIMENT PROTEINSKE KRITALOGRAFIJE

Najprej na kratko opišimo metodo proteinske kristalografije, ki je predstavljena na sliki 1. Nato bomo v nadaljevanju natančneje obravnavali njene posamezne korake. Podobno bi pristopali tudi v primeru, če bi določevali 3D strukturo drugih makromolekul (npr. nukleinskih kislin). Delo pričnemo tako, da najprej pridobimo, izoliramo in očistimo protein, katerega 3D strukturo želimo določiti. Nato očiščen protein raztopimo in kristaliziramo z uporabo različnih kristalizacijskih sredstev (A). Kristalizacija proteina predstavlja tudi najbolj nepredvidljiv korak metode (B). Nato kristale proteina obsevamo z rentgenskimi žarki, kar povzroči, da se vpadni x-žarki sipajo na elektronih, ki sestavljajo biološko makromolekulo (C). Dobimo difrakcijsko sliko proteina (D). Z merjenjem kotov in intenzitet teh razpršenih rentgenskih žarkov lahko pridobimo tridimenzionalno sliko gostote elektronov znotraj kristala. Še pred tem moramo rešiti t. i. fazni problem in določiti faze sipanih žarkov v difrakcijski sliki. Zato imamo na voljo več metod, ki jih bomo še spoznali. Pretvorba difrakcijske slike v elektronsko gostoto je možna, zato ker sta prostora odbojev in elektronske gostote povezana s Fourierjevo transformacijo. V dobljeno elektronsko gostoto, ob poznavanju primarne strukture proteina, umestimo njegov 3D model tako, da kar najbolje upošteva eksperimentalne omejitve dobljene elektronske gostote. Tako najprej pridobimo začetni 3D model proteina, ki ga nato v več iteracijah optimizacije izboljšamo (E). Na koncu ocenimo kvaliteto končnega 3D modela z več kvalitativnimi in kvantitativnimi kriteriji, ki nam povedo, kako dober model proteina smo dobili.



Slika 1. Shematski prikaz metode proteinske kristalografije.

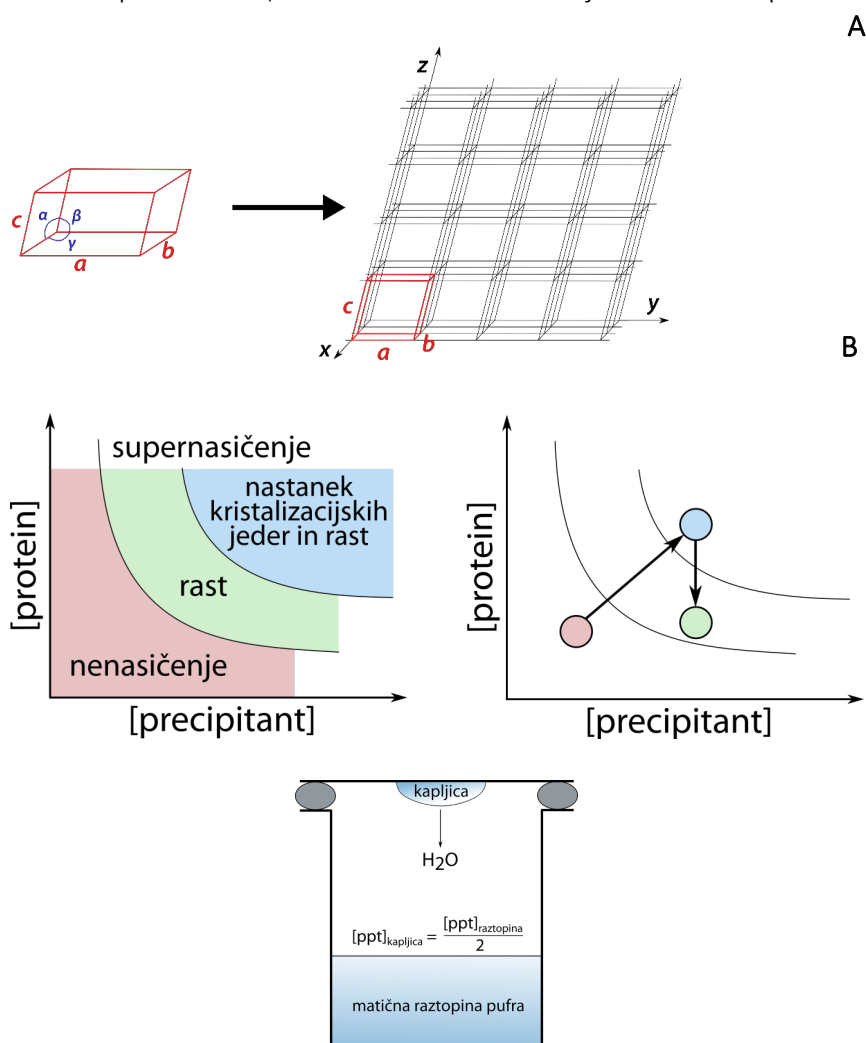
3. PRIPRAVA IN KRISTALIZACIJA PROTEINA

Za pričetek proteinske kristalografije potrebujemo raztopino očiščenega proteina v praviloma nativnem fiziološkem stanju. Protein lahko pridobimo na dva načina. Lahko ga izoliramo iz izvornega vira, kar pa pogosto ne omogoča pridobitve zadostne količine proteina. Največkrat gen za protein vstavimo v primeren ekspresijski sistem (npr. bakterije ali glive), da tako dobimo

večje količine proteina. Dobljeni vzorec proteina ocenimo z naslednjimi kriteriji: *čistost in homogenost, topnost proteina, monodisperznost, funkcionalna aktivnost in stabilnost*.

Zakaj sploh potrebujemo kristale za določitev strukture proteina? Intenziteta dobljene difrakcije le ene molekule je žal premajhna za dobro detekcijo. Urejena **kristalna rešetka** molekul v bolj ali manj isti konformaciji pa signifikantno ojača signal. Že majhen proteinski kristal namreč vsebuje milijarde proteinskih molekul, ki sestavljajo **kristalno mrežo** (Slika 2). Če protein ni uniformno urejen v kristalu, bo difrakcijska slika slaba in iz nje ne bomo dobili uporabnih podatkov za določitev strukture. V primeru visoke urejenosti kristala pa bo difrakcijska slika rentgenskih žarkov uporabna.

Najmanjša enota, ki se ponavlja skozi celoten kristal, se imejuje **osnovna celica** (*ang. unit cell*) (Slika 2). Na tem mestu omenimo še **asimetrično enoto** (*ang. asymetrical unit*), ki je najmanjši del kristalne strukture, s pomočjo katere lahko s simetrijskimi operacijami sestavimo osnovno celico. To je predvsem pomembno, če osnovna celica vsebuje več kot eno proteinsko molekulo.



Slika 2. (A) Osnovna celica in kristalna rešetka. (B) Shema pristopa kristalizacije proteinov in prikaz metode viseče kapljice.

Kljub temu da so to velike biološke molekule, tudi proteini pod določenimi pogoji tvorijo kristale. **Kristalizacija proteina** je pogosto opisana z izrazom „*more art than science*“ in zahteva

veliko napora in izkušenj, včasih pa tudi malo sreče z izborom pogojev. Različne korake kristalizacije, glede na količino obarjalnih spojin/precipitantov - snovi, ki omogočajo kristalizacijo proteina - si lahko ogledamo na diagramu Slike 2B.

Kristalizacijo pričnemo z očiščenim proteinom, ki ga predhodno še skoncentriramo. Osnovna ideja tega koraka je, da vzorec proteina v kristalizacijskih pogojih najprej privedemo do supernasičenja raztopine z dodatkom precipitantov. To nato vodi do nastanka kristalizacijskih jeder proteina in nadaljnje spontane nukleacije in rasti proteinskega kristala (Slika 2B). Supernasičenje dosežemo z variiranjem: *temperature, pH, koncentracije precipitantov* (npr. soli, kot so: amonijev sulfat, amonijev fosfat, natrijev citrat, itd.) ter *organskih topil* (npr. etanol, propanol, izopropanol, itd.).

Za kristalizacijo proteinov pogosto uporabljamo **metodo viseče kapljice** (ang. *hanging drop method*). Aparatura za to metodo je prikazana na spodnjem delu Slike 2B. Puffer za kristalizacijo predstavlja matično raztopino ($V = 0.5 - 1 \text{ mL}$). Na ravno stekleno površino damo kapljico, ki vsebuje $\frac{1}{2}$ raztopine proteina in $\frac{1}{2}$ pufra za kristalizacijo ($V = 6-10 \text{ mL}$). To pomeni, da je koncentracija precipitantov v kapljici $\frac{1}{2}$ koncentracije le-teh v matični raztopini pufra. Kapljico obrnemo v raztopino in sistem neprodušno zapremo. Koncentracija precipitanta oz. precipitantov v kapljici se bo počasi ekvilibrirala s koncentracijo v matični raztopini preko difuzije vodne pare. Tako bo lahko prišlo najprej do supernasičenja raztopine, nato do nukleacije ter na koncu do tvorbe proteinskega kristala.

Ravno določitev optimalnih pogojev, kjer bo kristalizacija uspešna, je tisti korak, ki ga zelo težko napovemo v naprej. Pogosto je potrebno izvesti veliko eksperimentov pri različnih pogojih in imeti tudi malo sreče, da na koncu dobimo proteinski kristal. Zato danes pri tem koraku kristalografije pogosto uporabljamo robotske sisteme, ki lahko v kratkem času preučijo veliko različnih kristalizacijskih pogojev.

4. DIFRAKCIJSKI EKSPERIMENT

Zakaj pri kristalografiji sploh uporabljamo rentgenske žarke? V vseh vrstah mikroskopije je količina informacij, ki jih lahko dobimo, omejena z valovno dolžino uporabljenega elektromagnetnega (EM) valovanja. Pri običajni mikroskopiji, kjer uporabljamo vidno svetlobo (VIS), je najkrajša valovna dolžina okoli 350 nm in to definira tudi najmanjše strukture, ki jih lahko določimo. To omogoča, da na primer lahko z mikroskopom opazujemo organele v celici. Da lahko določimo strukturo proteina z atomistično natančnostjo, potrebujemo elektromagnetno valovanje z valovno dolžino na nivoju velikosti atomov, to je okoli 0.1 nm. V to območje elektromagnetnega valovanja se uvrščajo x-žarki z valovnimi dolžinami med 0.01 do 10 nm.

Najpogosteje uporabljamo dva izvora rentgenske svetlobe:

1. **Anode.** Začetne difrakcijske eksperimente s proteinskimi kristali lahko izvedemo s tem lokalnim virom rentgenskega EM valovanja. Anodni izvori rentgenskih žarkov so (relativno) poceni in enostavni za vzdrževanje ter omogočajo hiter potek difrakcijskih eksperimentov. Ena glavnih omejitev je ozek obseg valovnih dolžin, ki jih lahko

generiramo. Te lastnosti so odvisne od materiala, ki sestavlja anodo – ponavadi je to baker (Cu), ki ga uporabljamo zaradi visoke toplotne prevodnosti, in omogoča učinkovito hlajenje. Prav tako ima ugodno razporeditev energijskih nivojev (močne linije $K\alpha$ in $K\beta$), kar omogoča učinkovitejše generiranje rentgenske svetlobe.

Pri tem pristopu z močnim električnim potencialom pospešimo elektrone iz njihovega izvora na kovino anode. Pri trku z anodo se 99 % kinetične energije elektronov spremeni v toploto (zato moramo anodo med tem korakom vseskozi hladiti), približno 1 % pa se pretvori v rentgenske žarke. Rentgenski žarki nastanejo tudi kot posledica interakcije teh elektronov z atomi kovine na anodi. Namreč, hitri elektroni lahko izbijejo katerega od elektronov na notranjih lupinah atoma, prazno mesto pa zapolni elektron iz višjih lupin. Pri prehodu iz lupine z višjo energijo na lupino z nižjo energijo, se izseva rentgenski foton z energijo, ki je enaka razliki energij obeh lupin.

Nekoliko izpopolnjena verzija tega sistema je **vrteča anoda** (ang. *rotating anode*), kjer je anoda zasnovana tako, da se njena površina vrti v snopu rentgenskih žarkov in tako poveča količino generiranih rentgenskih žarkov ter optimizira postopek hlajenja.

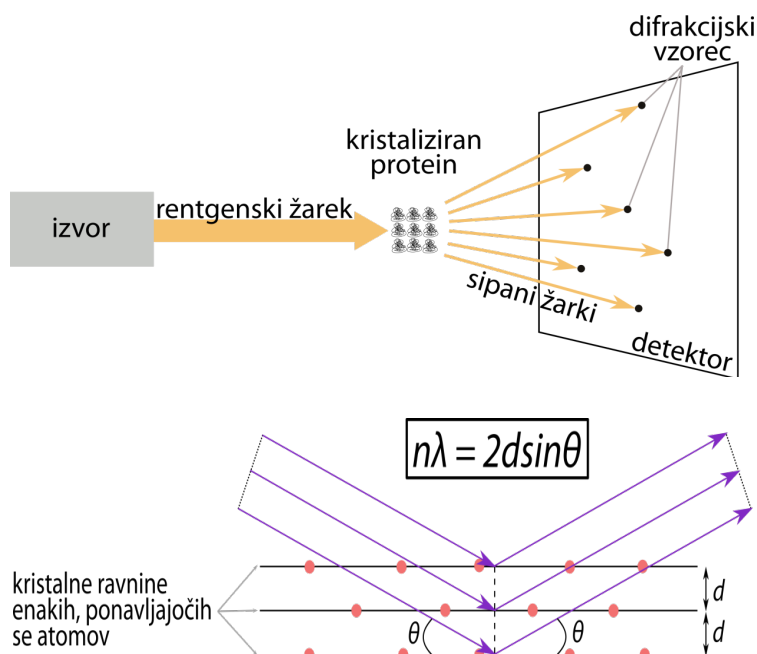
Nastale rentgenske žarke s pomočjo sistema leč ali zrcal in filtrov (oz. monokromatorja) skoncentriramo na ozek spekter valovnih dolžin (lahko tudi le na posamezno valovno dolžino) in jih usmerimo na proteinski kristal. Ta postopek ne samo poenostavi analizo podatkov, ampak tudi odstrani tisti del rentgenskega sevanja, ki kristal le razgradi brez prispevanja uporabnih informacij.

2. **Pospeševalniki delcev - sinhrotroni** so eden najboljših virov rentgenskih žarkov, ki so na voljo kristalografom. Sinhrotroni so veliki instrumenti, ki se zaradi svoje velikosti lahko razprostirajo celo čez več držav in se uporabljajo prvenstveno v raziskavah fizike delcev. Omogočajo pospešitev električno nabitih delcev (npr. elektronov), do hitrosti blizu svetlobne hitrosti. Pospešeni elektroni nato krožijo v posebnih ceveh z radijem lahko tudi več kilometrov. Tako stanje dosežemo z uporabo magnetnih polj. Ker je taka pot hitrih elektronov ukrivljena, nabiti delci oddajajo energijo v več področjih EM spektra valovanj, tudi v obliki rentgenskih žarkov. Velika prednost sinhrotronov je, da lahko izberemo izsevno valovno dolžino x-žarka, kar je izredno pomembno pri nekaterih kristalografskih eksperimentih (npr. MAD eksperimentih, ki jih bomo spoznali). Naj za konec omenimo, da lahko za generiranje x-žarkov poleg obeh virov uporabimo tudi **lasersko svetlobo**.

Pri **difrakcijskem eksperimentu** preiskovani proteinski kristal za meritve namestimo tako, da se lahko vrti v snopu rentgenskih žarkov, kot to prikazuje Slika 3. Obstaja več načinov pritrditve in danes proteinske kristale navečkrat zajamemo v posebno zanko ter jih nato hitro zamrznemo s tekočim dušikom. To storimo zato, da omejimo poškodbe kristala, ki so posledica kasnejše interakcije kristala z visokoenergijskimi rentgenskimi žarki. Če so zamrznjeni prehitro, proteinski kristali pogosto počijo, zato jih ponavadi pred zamrzovanjem namočimo še v raztopino krioprotektivnega sredstva in jih tudi med samim eksperimentom neprestano hladimo.

Zanko z ujetim proteinskim kristalom namestimo na **goniometer**. To je naprava, ki omogoča natančno postavitev kristala znotraj rentgenskega žarka ter vrtenje. Ker sta proteinski kristal in

generiran x-žarek pogosto zelo majhna, je treba kristal natančno centrirati znotraj žarka (natančnost $\sim 25 \mu\text{m}$), pri čemer si lahko pomagamo s posebno kamero, s katero opazujemo kristal. Najpogostejša vrsta goniometra je "*kappa goniometer*", ki omogoča več kotov vrtenja in tako natančno pridobitev difrakcijske slike v vseh možnih orientacijah kristala. Naj še enkrat omenimo, da ves čas poteka difrakcijskega eksperimenta tok tekočega dušika hladi kristal pri približno 100 K, in s tem preprečuje, da bi kristal zaradi visokih obsevanih energij razpadel.



Slika 3. Poenostavljena shema difrakcijskega eksperimenta in grafični prikaz Braggovega zakona.

Med difrakcijskim eksperimentom se obsevani x-žarki odbijejo/sipajo od elektronov. Zadeti elektron začne najprej vibrirati z isto frekvenco, kot jo ima vhodni rentgenski žarek nato pa prične sipati sekundarne x-žarke, ki se širijo v vse smeri. Za rentgensko difrakcijo je pomembno predvsem **elastično sipanje na elektronih**, kjer se valovna dolžina in faza žarka ohranita, žarek pa spremeni smer. Sipani žarki se nato širijo v vse smeri in nekateri izmed njih med seboj interagirajo, saj ima molekula veliko elektronov. Tako nastanejo ojačitve in oslabitve rentgenske svetobe.

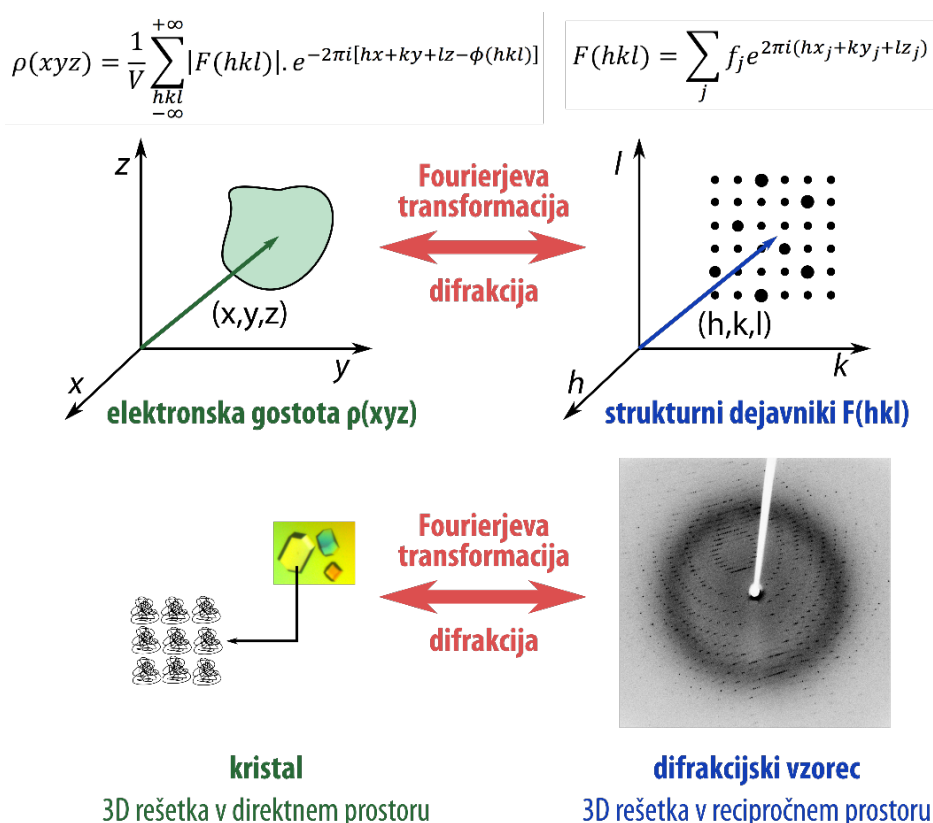
Pogoj za konstruktivno difrakcijo je podan z **Braggovim zakonom**, ki sta ga razvila oče in sin, William Henry Bragg in Lawrence Bragg, ter za svoje delo leta 1915 dobila tudi Nobelovo nagrado. Ojačitev žarka po odboju od elektronov, ki sestavljajo atom, dobimo takrat, ko je razlika poti žarkov, ki se sipajo (odbijejo) od različnih kristalnih ravnin, enaka mnogokratniku valovnih dolžin vpadnega valovanja. Pri čemer je d razdalja med dvema sosednjima ravninama, ϑ je uklonski kot sipanja, n mora biti celo število, λ pa je valovna dolžina vpadnega valovanja (Slika 3).

Iz povedanega sledi, da je končni vzorec sipanih žarkov odvisen od števila in položaja elektronov v molekuli, torej od strukture molekule. Rezultat difrakcijskega eksperimenta je tridimenzionalna mapa odbojev (reflekcij) oz. **difrakcijski vzorec**, ki je direktno povezan s porazdelitvijo elektronov v preiskovani molekuli (Slika 3).

Odbite oz. sipane rentgenske žarke zaznamo z različni **detektorji**. V preteklosti so v ta namen pogosto uporabljali kar fotografski film. Kasneje so v ta namen razvili števec odbitih fotonov t. i. scintilacijske števec, ploščate slikovne detektorje (ang. *image plate detectors*), matrične detektorje (ang. *array detectors*) in CCD naprave (ang. *charge-coupled devices*).

5. IZRAČUN 3D MAPE ELEKTRONSKE GOSTOTE

V naslednjem koraku dobljene dvodimenzionalne difrakcijske vzorce, ki jih dobimo z rotacijo proteinskega kristala v vse smeri, pretvorimo v tridimenzionalni model elektronske gostote molekule. Pri tem uporabljamo matematično tehniko, imenovano **Fourierjeva transformacija**. Le-ta namreč povezuje **direktni** in **recipročni** prostor - to je atomsko strukturo in njeno uklonsko/difrakcijsko sliko. Z uporabo **reverzne Fourierjeve transformacije** tako lahko iz izmerjenih amplitud odbojev in njenih faznih kotov izračunamo elektronsko gostoto v osnovni celici, kot to prikazuje slika 4.



Slika 4. Shema določitve elektronske gostote proteina iz dobljenih odsefov s pomočjo inverzne Fourierjeve transformacije.

Obdelava podatkov se začne z **indeksiranjem** izmerjenih difrakcijskih žarkov. Tako določimo **dimenzije enotne celice** in pozicije žarkov v recipročnem prostoru. Med tem postopkom določimo simetrije kristala, to je njegovo **prostorsko skupino** (ang. *space group*). Prostorske skupine so kombinacije 32 kristalografskih **točkovnih skupin** (simetrijskih operacij, kot so rotacije in zrcaljenja, v katerih ostane pri premiku vsakega atoma kristala na položaj drugega istovrstnega atoma točka, okoli katere poteče operacija, fiksna) in 14 **Bravaisovih mrež** (možnih 3D ureditev v neskončni mreži diskretnih točk). Slednje nadalje spadajo v enega od sedmih

kristalnih razredov (možnih razporeditev atomov in molekul v kristalu). Obstaja **230 edinstvenih prostorskih skupin**, ki opisujejo vse možne kristalne simetrije.

Pri proteinih predvsem zaradi prisotnih kiralnih elementov nekatere prostorske skupine niso dovoljene, tako da ostane na voljo 65 prostorskih skupin od 230. **Indeksiranje odbojev** običajno izvedemo s pomočjo računalnika. Ta pretvori na stotine 2D difrakcijskih slik, ki vsebujejo tisoče odsevov posnetih v različnih orientacijah kristala, v eno datoteko. Odseve v recipročnem prostoru opišemo z **Millerjevimi indeksi**, ki predstavljajo koordinate h, k, l recipročnega prostora, pripišemo pa jim tudi intenzivnost vsakega odboja. Pogosto v obdelavo vključimo tudi oceno napake dobljenega odseva ter podatke uredimo še s primerjavo intenzivnosti difrakcij pri različnih merjenih orientacijah.

Podatki, zbrani iz difrakcijskega eksperimenta, predstavljajo recipročni prostor kristalne rešetke. Koordinata vsake difrakcije je odvisna od velikosti in oblike osnovne celice v direktnem prostoru ter pripadajoče simetrije znotraj kristala. Med eksperimentom izmerjena intenziteta $I(h, k, l)$ vsake "difrakcijske točke" je sorazmerna s kvadratom amplitude **strukturnega faktorja F**.

$$I(h, k, l) = |F(h, k, l)|^2$$

Strukturni faktor je kompleksno število, ki vsebuje informacije, ki se nanašajo tako na amplitudo $|F(h, k, l)|$ kot fazo vala ϕ .

$$F(h, k, l) = |F(h, k, l)|e^{-i\phi}$$

Z uporabo Eulerjevega teorema lahko strukturni faktor zapišemo tudi v obliki:

$$F(h, k, l) = \sum_{n=1}^N f_n e^{2\pi i(hx+ky+lz)}$$

Pri čemer je f_n atomski sipni faktor atoma n , x, y, z pa koordinate tega atoma v enotni celici direktnega prostora. Elektronska gostota molekule $\rho(x, y, z)$ je povezana s strukturnim faktorjem preko reverzne Fourierjeve transformacije. Enačba elektronske gostote molekule, ki povezuje oba prostora, ima spodnjo obliko:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) e^{-2\pi i(hx+ky+lz) + i\alpha(hkl)}$$

h, k, l	koordinate indeksov odbojev v recipročnem prostoru
xyz	koordinate atomov v realnem prostoru
F	amplitude odbojev
α	faze (fazni kot) odbojev
V	volumen osnovne celice

Iz enačbe elektronske gostote vidimo, da za pridobitev elektronske gostote molekule poleg izmerjenih koordinat in intenzitet odbojev potrebujemo poznati tudi njihove fazne kote (faze). Faze odbojev pa med difrakcijskim poskusom ni mogoče neposredno izmeriti, saj detektor meri le intenzitete odbojev. To imenujemo **fazni problem**. Faze (oz. fazne kote) moramo zato določiti s pomočjo drugih metod, ki jih na kratko opisujemo spodaj. Temu koraku pravimo tudi »faziranje« ali reševanje **faznega problema**.

METODA 1: MULTIPLA IZOMORFNA ZAMENJAVA

(ang. *Mutiple Isomorphous Replacement - MIR*)

MIR je najstarejša metoda reševanja faznega problema v kristalografiji in z njo so leta 1958 rešili prvo kristalno strukturo mioglobina. Pri MIR z uvedbo „težkega atoma“ v proteinski kristal spremenimo dobljeno difrakcijsko sliko. V ta namen uporabimo atome kot so: Pt, Pb, Hg. S primerjavo dobljenih difrakcijskih slik proteinskih kristalov z vsebovanimi težkimi atomi in pa brez njih, je mogoče določiti položaj (koordinate) težkih atomov ter tudi njihove faze. Nato ocenimo vpliv uvedenega atoma na intenzitete preostalih difrakcijskih žarkov, ki izhajajo iz proteinskih atomov, kar nam pomaga pri določitvi njihovih faz. Ponavadi pri tem pristopu uporabljamo vsaj dva različna težka atoma (dva ločena eksperimenta), ki ju kristaliziramo skupaj s proteinom. Metoda MIR tudi predpostavlja, da se dimenzija enotne celice ohranja pri vseh eksperimentih in da imajo proteinski kristali definirano mesto vezave. Pri tej metodi pogosto naletimo na težave, saj se proteinski kristali v prisotnosti težkih atomov ne tvorijo.

METODA 2: MULTIPLA ANOMALNA DISPERZIJA

(ang. *Mutiple Anomalous Dispersion - MAD*)

MAD je zelo podobna tehnika kot je MIR, le da tu uporabimo dejstvo, da atomi težkih atomov pri določenih valovnih dolžinah anomalno razpršijo x-žarke s spremembo amplitude in faze. Pri tej metodi lahko poleg prej omenjenih težkih atomov v strukturo proteina vključimo tudi aminokislino selenometionin, ki namesto žvepla vsebuje z elektroni bogatejši atom selena. Za uporabo MAD metode moramo zato biti sposobni proizvesti x-žarke z različnimi valovnimi dolžinami, kar nam omogoča sinhrotron. Z merjenjem difrakcij pri teh izbranih valovnih dolžinah lažje najdemo lokacijo in fazo težkega atoma oz. selenometionina. Analogno kot pri MIR metodi se ti podatki nato uporabijo za določitev faz preostalih difraktiranih žarkov.

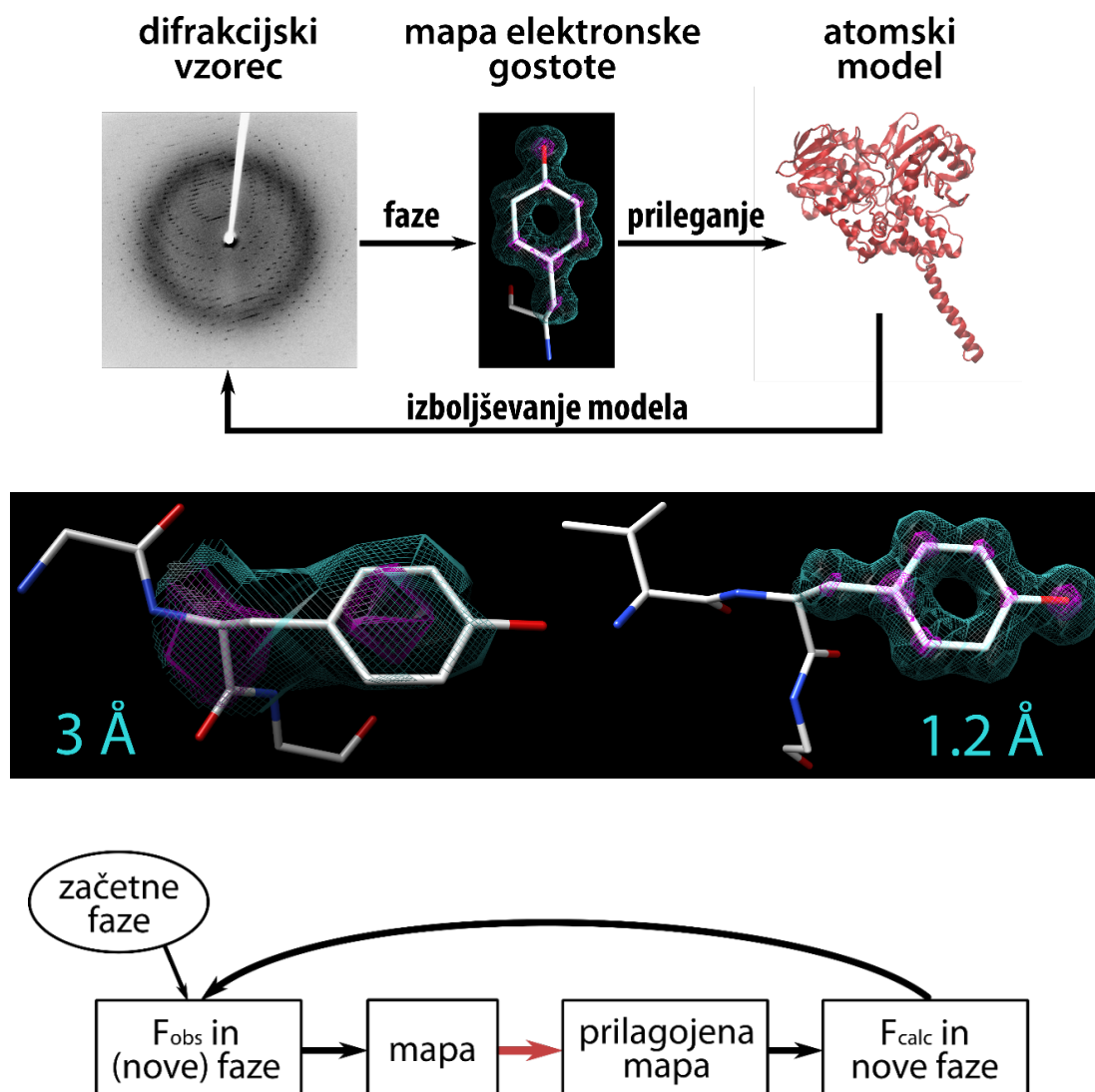
METODA 3: MOLEKULSKA ZAMENJAVA

(ang. *Molecular Replacement - MR*)

MR je najpogosteje uporabljena tehnika proteinske kristalografije, saj pogosto obstaja že rešena 3D struktura homolognega proteina. Tako lahko kot začetne faze pri izračunu elektronske gostote privzamemo kar te, ki so bile dobljene pri homologni strukturi, in jih kombiniramo z dobljenimi intenzitetami odbojev. Metoda molekulske izmenjave je uporabna le, če želimo določiti strukturo proteinu, ki je po primarni sekvenci vsaj v 25 % podoben proteinu, za katerega 3D struktura že obstaja. Preden pričnemo z izračunom elektronske gostote, moramo s pomočjo računalnika prenesti znani 3D proteinski model v umetno zgrajeno enotno celico, ki ima enako orientacijo kot struktura za katero želimo dobiti 3D model. Za to uporabimo različne rotacijske in translacijske funkcije.

6. GRAJENJE IN OCENA KVALITETE 3D MOLEKULSKEGA MODELA

Ko izmerjenim difrakcijskim žarkom določimo faze, lahko izračunamo tudi **3D mapo elektronske gostote** proteina. Zaradi poznavanja primarne sekvence preiskovanega proteina lahko v mapo elektronske gostote vgradimo **3D model proteinske strukture**. Pri tem si pomagamo z različnimi računalniškimi programi, ki nam predlagajo najboljša ujemanja elektronske gostote in zgrajenega modela. Začetni model nato še izboljšamo in optimiziramo, da dobimo najboljše možno ujemanje z eksperimentalno elektronsko gostoto in predlagano 3D konformacijo (Slika 5).



Slika 5. (Zgoraj) Koraki gradnje 3D atomskega modela na osnovi eksperimentalno pridobljenega difrakcijskega vzorca. (Sredina) Prikaz resolucije manj (3 Å) in bolj (1.2 Å) kvalitetne mape elektronske gostote stranske verige tirozina v molekuli proteina. (Spodaj) Shema iterativnega popravljanja začetnega 3D modela do najboljšega ujemanja z dobljenimi eksperimentalnimi podatki.

Pri oceni kvalitete končnega 3D modela je pomembno merilo **resolucija**. Ta je definirana kot vrednost najmanjše razdalje, ki jo še lahko ločimo. Resolucija je odvisna od količine in kvalitete

dobljenih eksperimentalnih odbojev in določenih faz. Seveda dobra resolucija omogoča lažjo in natančnejšo gradnjo 3D molekulskega modela proteina. Navadno velja, da resolucija nižja od 3.0 Å že omogoča sprejemljivo postavitev 3D modela in zadovoljivo interpretacijo elektronske gostote. Pri resoluciji pod 2.0 Å gradnja 3D modela s pomočjo različnih računalniških programov poteka že skoraj popolnoma avtomatizirano. Če je dobljena resolucija nižja od 1.2 Å, pa včasih lahko celo določimo pozicije nekaterih vodikovih atomov v proteinu, ki jih z rentgensko difrakcijo večinoma sploh ne moremo določiti. Za boljšo predstavbo razliko med resolucijo 3.0 Å in 1.2 Å na primeru stranske verige aminokislina tirozina prikazujemo na srednjem delu Slike 6.

Ko pridobimo začetni 3D model, ga potem v več stopnjah poskušamo **optimizirati** in **izboljšati** (ang. *model refinement*), da dobimo kar najboljše možno ujemanje z eksperimentalno elektronsko gostoto. Na osnovi izračunanih koordinat atomov v začetnem modelu in vrednosti njihovih **B-faktorjev** (Debye-Wallerjev faktor), ki predstavljajo, kot bomo to videli še v nadaljevanju, toplotno gibanje atoma, poskušamo njihove koordinate atomov v modelu spremeniti tako, da bolj ustrezajo opazovanim difrakcijskim podatkom, kar lahko rezultira v optimalnejših fazah, kot so tiste, ki smo jih uporabili na začetku. Nato izračunano elektronsko gostoto z novimi fazami prilegamo novemu modelu in ta iterativni postopek nadaljujemo, dokler ne dobimo najboljše možne korelacije med difrakcijskimi podatki in 3D modelom proteina (Slika 5).

Ustreznost ujemanja lahko ovrednotimo z **R faktorjem**, ki primerja eksperimentalno elektronsko gostoto in 3D kristalografski model, preko izračunanih F_{calc} in eksperimentalno dobljenih F_{obs} strukturnih faktorjev F . Uporabimo spodnjo enačbo:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

Vrednosti R faktorja in interpretacija so naslednji:

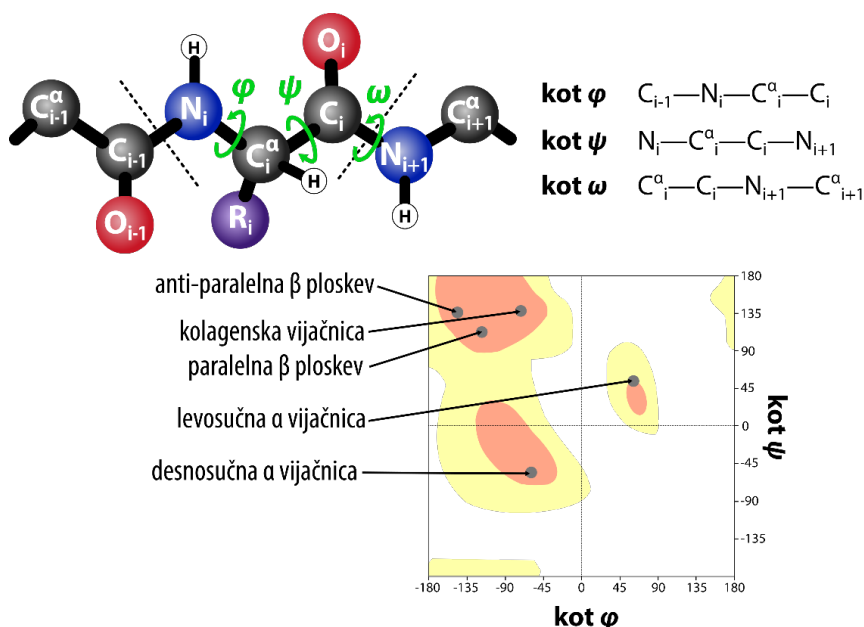
0.6	<i>Zelo slabo ujemanje</i>
0.5	<i>Slabo ujemanje</i>
0.4	<i>Spodnja meja sprejemljivosti ujemanja</i>
0.2	<i>Dobro za proteinske molekule</i>
0.05	<i>Dobro za majhne organske molekule</i>
0	<i>Popolno ujemanje</i>

Kot merilo za kakovost modela se je v zadnjih letih uveljavil **prosti R faktor** (R_{free}). Pri tej metodi določen delež odbojev (5–10 %), ki jih damo v testni niz, ločimo od preostalih odbojev. 3D model proteina zgradimo z upoštevanjem slednjih in na koncu izračunamo še ujemanje modela s testnim setom ter določimo R_{free} faktor.

Kvaliteto dobljenega 3D proteinskega modela potem validiramo še z drugimi geometrijskimi kriteriji. Preverimo, ali so dolžine vezi, valenčni koti in ravnine v skladu z dovoljenimi vrednostmi za molekulske strukture, ki sestavljajo proteine. Prav tako pogledamo, ali imajo aminokislina ustrezno okolico. Za veliko proteinov je na primer značilno, da imajo hidrofobno sredico in polarno površino. Dovoljene konformacije stranskih verig so navadno tudi omejene in obstajajo knjižnice možnih rotamerov, s katerimi lahko preverimo skladnost dobljene strukture.

Ovrednotimo lahko tudi, ali se tvorijo ustrezni pari vodikovih vezi med aminokislinskimi preostanki.

Dovoljene konformacije glavne proteinske verige ovrednotimo z **Ramachandranovim diagramom**. Trije torzijski (dihedralni) koti, fleksibilna ψ (definirajo ga atomi: $N(i) - C\alpha(i) - C(i) - N(i+1)$) in ϕ (definirajo ga atomi: $C(i-1) - N(i) - C\alpha(i) - C(i)$) ter kot ω , ki je zaradi srednje C-N amidne vezi, ki je planarna, navadno fiksen z vrednostjo 180° , namreč določajo potek glavne verige (definicijo torzijskega kota si pogledajte še v *Poglavju IV*). Z Ramachandranovim diagramom vizualiziramo energijsko dovoljene in prepovedane regije za vse možne vrednosti torzijskih kotov ψ in ϕ aminokislinskih preostankov v proteinu. Tako ocenimo pravilnost napovedanega 3D modela. Glede na to, v katerih območjih se nahajajo vrednosti obeh kotov, pa lahko napovemo tudi sekundarno strukturo posameznih predelov proteina (Slika 6).



Slika 6. Torzijski koti glavne proteinske verige in primer Ramachandranovega diagrama.

Prej omenjeni **Debye-Wallerjev faktor** ali **B-faktor** opisuje (ne)zanesljivost položaja atoma v kristalu. Pogojen je z resolucijo, saj so praviloma pri nižji resoluciji B faktorji višji, lahko pa tudi služi kot merilo urejenosti strukture v posameznih delih polipeptidne verige. Vrednosti B faktorja pod 10 označujejo atome, katerih pozicije se med posameznimi molekulami v kristalu le malo razlikujejo. Vrednosti nad 50 pa kažejo, da so ti atomi zaradi gibljivosti v elektronski gostoti slabše vidni.

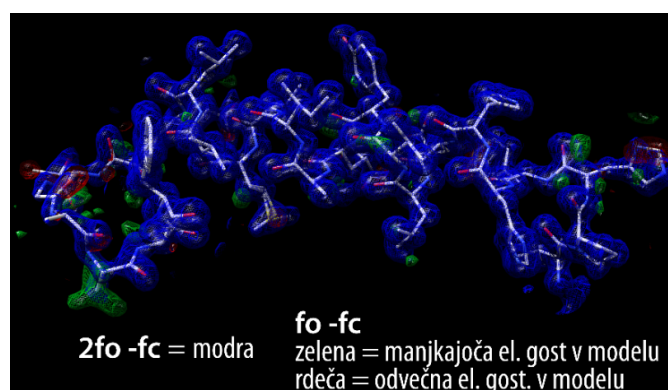
ATOM	92	N	GLY	A	47	20.191	-18.979	14.342	1.00	10.01	N
ATOM	93	CA	GLY	A	47	21.427	-18.690	13.599	1.00	11.79	C
ATOM	94	C	GLY	A	47	22.404	-17.931	14.477	1.00	11.39	C
ATOM	95	O	GLY	A	47	22.629	-18.337	15.636	1.00	12.19	O
ATOM	96	N	ASP	A	48	22.940	-16.834	13.965	1.00	12.03	N
ATOM	97	CA	ASP	A	48	23.937	-16.030	14.714	1.00	12.88	C
ATOM	98	C	ASP	A	48	23.240	-14.894	15.470	1.00	11.36	C
ATOM	99	O	ASP	A	48	23.940	-14.032	15.990	1.00	12.12	O
ATOM	100	CB	ASP	A	48	25.036	-15.510	13.785	1.00	16.65	C
ATOM	101	CG	ASP	A	48	25.867	-16.616	13.153	1.00	25.48	C
ATOM	102	OD1	ASP	A	48	26.002	-17.685	13.788	1.00	30.60	O
ATOM	103	OD2	ASP	A	48	26.369	-16.401	12.027	1.00	33.86	O

Okupanca (ang. *occupancy*) označuje, v kolikšnem deležu molekul je prisoten nek atom.

ATOM	92	N	GLY	A	47	20.191	-18.979	14.342	1.00	10.01	N
ATOM	93	CA	GLY	A	47	21.427	-18.690	13.599	1.00	11.79	C
ATOM	94	C	GLY	A	47	22.404	-17.931	14.477	1.00	11.39	C
ATOM	95	O	GLY	A	47	22.629	-18.337	15.636	1.00	12.19	O
ATOM	96	N	ASP	A	48	22.940	-16.834	13.965	1.00	12.03	N
ATOM	97	CA	ASP	A	48	23.937	-16.030	14.714	1.00	12.88	C
ATOM	98	C	ASP	A	48	23.240	-14.894	15.470	1.00	11.36	C
ATOM	99	O	ASP	A	48	23.940	-14.032	15.990	1.00	12.12	O
ATOM	100	CB	ASP	A	48	25.036	-15.510	13.785	1.00	16.65	C
ATOM	101	CG	ASP	A	48	25.867	-16.616	13.153	1.00	25.48	C
ATOM	102	OD1	ASP	A	48	26.002	-17.685	13.788	1.00	30.60	O
ATOM	103	OD2	ASP	A	48	26.369	-16.401	12.027	1.00	33.86	O

Okupanca za atome proteina je ponavadi 1. Manjša od 1 je lahko v primerih, ko je v proteinu prisotnih več konformacij aminokislinskega preostanka in ko je ligand prisoten v večih konformacijah ali ni prisoten v vseh osnovnih celicah.

Dobljeno elektronsko gostoto proteina lahko prikazujemo v obliki diferenčnih map (Slika 7). Poznamo več vrst map in iz banke PDB lahko pridobimo **F₀-F_c diferenčno (omit) mapo**, ki prikazuje diferenco (razliko) med eksperimentalno mapo elektronske gostote F₀ in mapo F_c, ki jo izračunamo iz dobljenih atomskih koordinat v modelu. Ta mapa nam prikaže vse atome, ki so napačno locirani in manjkajočo še ne interpretirano elektronsko gostoto. Še bolj pomembna je **2F₀-F_c diferenčna mapa**. To je eksperimentalna mapa F₀ skupaj z diferenčno mapo F₀-F_c. Ta mapa izgleda kot popravljeni model in je navadno statistično najmanj pristranska mapa elektronske gostote. Mapo elektronske gostote proteina boste vizualizirali tudi v okviru aktivnosti pri Vaji 2.



Slika 7. Shematski prikaz F₀-F_c in 2F₀-F_c diferenčne mape.

Tako določeno optimizirano 3D strukturo proteina skupaj s preostalimi eksperimentalnimi rezultati (npr. mapa elektronske gostote, diferenčne mape, itd.) nato deponiramo v **PDB** (ang. *Protein Data Bank*) banko struktur bioloških molekul. Koordinate strukture so navadno dostopne v molekularnem formatu **pdb**. Povejmo še, da datoteka **pdb** ne vsebuje le koordinat atomov, ampak še veliko drugih uporabnih podatkov: od podrobnosti kristalografske/difrakcijske meritve, do naštetih aminokislin, ki so bile slabše ali sploh ne zaznane v mapi elektronske gostote, podatkov o osnovni celici itd. Vse te informacije nam omogočajo vizualizacijo proteinske molekule ter kompleksnejše razumevanje pridobljenih podatkov.

Andrej Perdih

določevanje 3d strukture bioloških molekul z nmr spektroskopijo

```
function quicksort(array)
var list less, equal, greater
if length(array) <= 1
return array
select a pivot value pivot
for each x in array
if x < pivot then list = less
if x = pivot then list = equal
if x > pivot then list = greater
return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

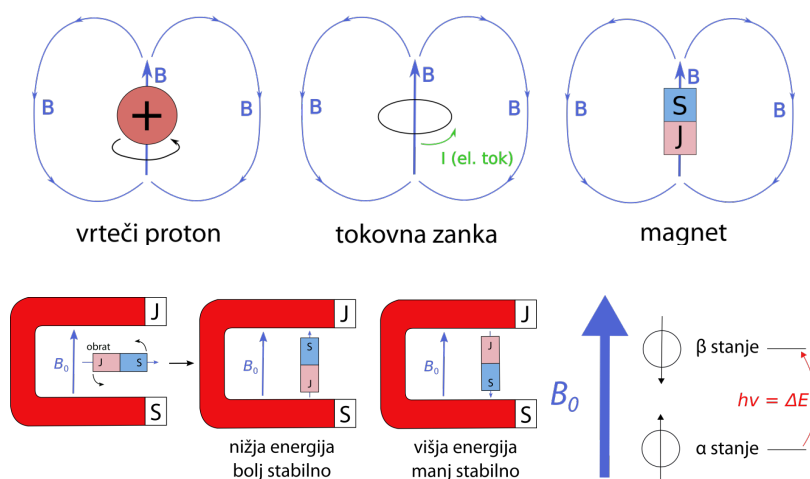
1. JEDRSKA MAGNETNA REZONANCA IN NMR EKSPERIMENT

Jedrska ali nuklearna magnetna resonanca (ang. *Nuclear magnetic resonance* - NMR) je najbolj uporabljana spektroskopska tehnika v organski kemiji, ima pa tudi, kot bomo videli, pomembno vlogo pri določevanju 3D strukture bioloških molekul, predvsem proteinov, pa tudi nukleinskih kislin (npr. DNA, RNA). Najprej bomo spoznali splošne lastnosti NMR pojava, ki sta ga leta 1946 neodvisno odkrila **Felix Bloch** in **Edward Mills Purcell** in si za njegovo odkritje leta 1952 delila Nobelovo nagrado za fiziko.

NMR pojav temelji na opažanju, da imajo atomska jedra pozitiven naboj, nekatera med njimi pa tudi **spin** in zato od nič različen magnetni moment. Takšna **magnetno aktivna jedra** morajo imeti liho število protonov ali nevtronov, kot npr. jedra ^1H , ^2H , ^{13}C , ^{15}N , ^{31}P in ^{19}F . Spin si v klasični fiziki lahko predstavljamo kot vrtenje oz. natančneje precesijo nabitega delca, pri čemer nastane okrog njega majhno magnetno polje, kot to prikazuje Slika 1. Smer takega spina in z njim povezanega magnetnega momenta je poljubna, vse orientacije v prostoru imajo enako energijo. Če pa magnetno aktivno jedro (npr. ^1H proton) postavimo v zunanje magnetno polje z magnetno gostoto B_0 , vse orientacije nimajo več enake energije. Spin se pri interakciji s poljem B_0 orientira tako, da se njegov magnetni moment postavi za večino jeder v smeri vektorja B_0 (energijsko ugodnejše stanje α), manjša populacija jeder pa v nasprotno smer (energijsko manj ugodno stanje β) (Slika 1). V principu je možnih še več stanj, katerih število je odvisno od spinskega kvantnega števila I . Energijska razlika med obema stanjema je podana z enačbo:

$$\Delta E = h\nu = \gamma \frac{h}{2\pi} B_0$$

kjer je h Planckova konstanta, B_0 jakost zunanjega magnetnega polja in γ giromagnetno razmerje, ki je za vsako jedro konstantno ($26.753 \text{ s}^{-1}\text{gauss}^{-1}$ za vodik).



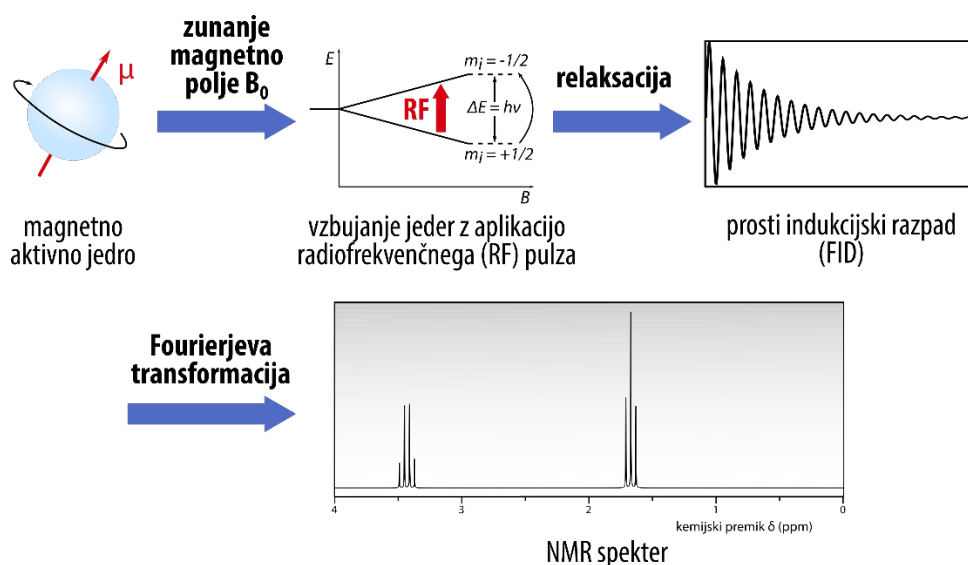
Slika 1. (Zgoraj) Prikaz magnetno aktivnega jedra ^1H kot vrtečega nabitega delca, ki se obnaša kot majhen magnet, in primerjava s podobnimi sistemi, tokovno zanko in magnetom. (Spodaj) Obnašanje jedra v zunanjem magnetnem polju.

Poglejmo si potek enodimenzionalnega **NMR eksperimenta**, kjer obravnavamo ^1H protone kot **magnetno aktivna jedra**. Vzorec s preiskovano molekulo, ki vsebuje magnetno aktivna jedra ^1H , damo v zunanje magnetno polje B_0 in ga obsevamo z elektromagnetnim valovanjem primerne frekvence. To so navadno radijski valovi s frekvencami med 60 in 1000 MHz. S tem kratkim

pulzom radiofrekvenčnega (RF) valovanja vzbudimo jedra v vzorcu, da le-ta preidejo iz osnovnega stanja v vzbujena stanja, npr. iz energijsko ugodnejšega stanja α v energijsko višje stanje β , kot to prikazuje Slika 2. To se zgodi, ko je zunanja frekvenca RF valovanja enaka lastni frekvenci magnetno aktivnega jedra. Jedra lahko vzbujamo pod določenim kotom glede na magnetno polje spektrometra in obseg vzbujanja lahko nadzorujemo tudi s časom trajanja pulza (cca med 3-8 μ s).

Po določenem času se vzbujena jedra vrnejo v osnovno stanje, kar imenujemo **relaksacija**, in ob tem izsevajo elektromagnetno valovanje določenih frekvenc. Temu pravimo tudi »prosti indukcijski razpad« (ang. *free induction decay* FID). Relaksacijski čas je odvisen od vsakega magnetno aktivnega jedra. Gre pa vedno za zelo šibek signal, zato za njegovo detekcijo potrebujemo občutljive detektorje (t. i. sprejemna tuljava). Navadno vsak NMR eksperiment za ojačitev signala in lažjo detekcijo večkrat ponovimo (npr. pri ^1H NMR eksperimentih naredimo cca. 16 ponovitev). Pri atomskih jedrih, ki niso vodik, je relaksacijski čas precej dolg (cca. 8 sekund za atom ^{13}C), zato je pridobivanje NMR spektrov za ta jedra precej zamudno in lahko traja tudi več ur, saj eksperiment lahko ponovimo le po koncu relaksacije jedra.

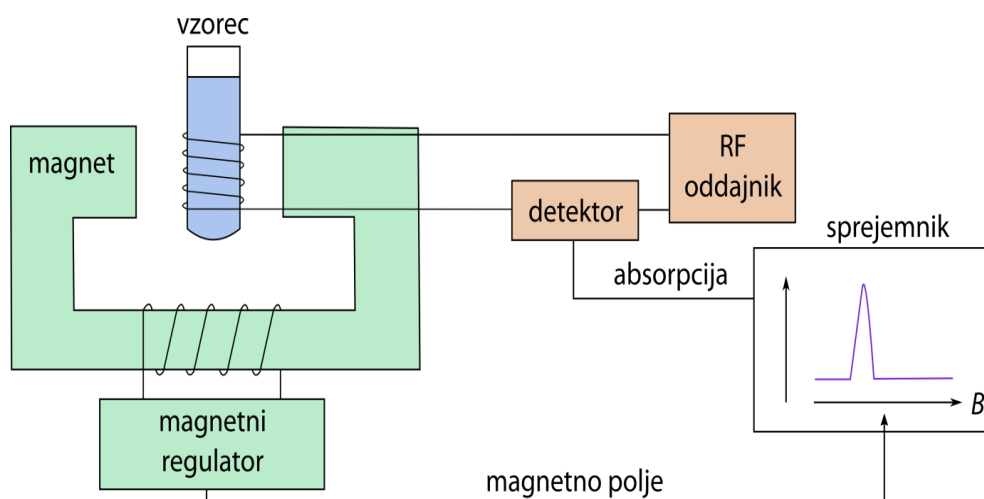
Zaznani signal relaksacije vsebuje vse frekvence, ki jih oddajajo magnetno aktivna jedra v preiskovani molekuli. Dobljene signale vseh jeder zapišemo in obdelamo s **Fourierovo transformacijo**. Zapis nihanja prestavimo iz časovne domene v frekvenčno. Tako iz zapisa vseh nihanj magnetno aktivnih jeder po času dobimo zapis nihanj magnetno aktivnih jeder pri posameznih frekvencah - to je **1D NMR spekter** (Slika 2).



Slika 2. Poenostavljena shema enodimenzionalnega ^1H NMR eskperimenta.

Pri NMR eksperimentih ne moremo uporabljati večine standardnih topil, kot so voda, aceton, tetrahidrofuran in podobna, saj vsebujejo magnetno aktivna jedra. Ta bi predstavljala praktično vse molekule, ki sestavljajo raztopino in tako tvorila večino dobljenega signala. Da se izognemo zaznavanju samo teh signalov, uporabljamo magnetno neaktivna devterirana topila, pri katerih je več kot 99 % vodikovih atomov zamenjanih z devterijem D. Najpogosteje uporabljeno devterirano topilo pri NMR eksperimentih je devteriran kloroform (CDCl_3).

NMR spektrometer, ki ga prikazuje Slika 3 sestoji iz vrtečega se držala za preiskovani vzorec, ki je postavljen znotraj močnega magneta. Magnet ima tudi poseben magnetni regulator za natančno produkcijo apliciranega magnetnega polja. Radiofrekvenčni oddajnik aplicira EM RF pulz, ki vzbudi magnetno aktivna jedra in nato detektor zazna izsevano EM valovanje, ki je posledica relaksacije jeter. Sprejemnik s pripadajočo računalniško opremo omogoča vizualizacijo in obdelavo pridobljenih podatkov.



Slika 3. Shema NMR spektrometra.

2. MAGNETNO SENČENJE JEDER IN KEMIJSKI PREMIK

Poglejmo si nekaj lastnosti NMR eksperimenta, ki jih s pridom izkoriščamo za lažje določanje strukture. Vsaj atomska jedra molekule obdajajo različno gosti oblaki elektronov. Elektronska gostota okrog posameznega atoma je odvisna od polarizacije posameznih vezi in lokalne molekularne strukture. Elektroni so tudi delci s svojim lastnim spinom in se zato odzivajo na zunanje magnetno polje, ki ga apliciramo med NMR eksperimentom. V tem polju se orientirajo, podobno kot atomska jedra, večinoma nasproti zunanjemu polju in ga tako zmanjšajo. Pravimo tudi, da elektronski oblak **senči jedra** pred zunanjim magnetnim poljem, kot to shematsko prikazuje Slika 4. Čim večja je elektronska gostota okrog atomskega jedra, tem močnejše je senčenje.

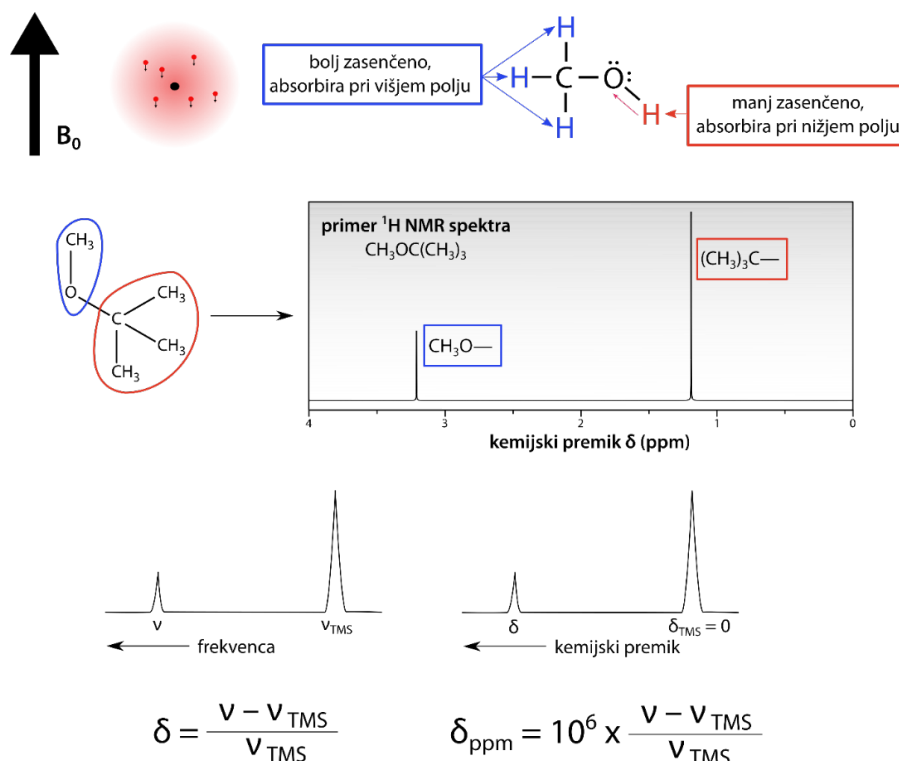
Lokalna struktura torej vpliva na to, kakšno magnetno polje čuti posamezno atomsko jedro in zato ima vsako jedro v specifičnem okolju nekoliko drugačno resonančno frekvenco, pri kateri absorbira dano RF energijo. Na primeru metanola (Slika 4) lahko vidimo, da je vodik hidroksilne skupine manj zasenčen zaradi polarizacije OH vezi in zato absorbira energijo pri nižjem magnetnem polju. Vsi trije metilni CH₃ protoni so bolj zasenčeni, saj imajo okrog sebe večjo gostoto elektronov in zato absorbirajo pri višjem magnetnem polju. To se odraža tudi v različnih resonančnih frekvencah obeh različnih tipov protonov (magnetno aktivnih jeter) v dobljenem NMR spektru.

Pri senčenju je razlika v resonančnih frekvencah jeter v različnih okoljih relativno majhna. Zato za lažjo predstavbo uvedemo pojem **kemijskega premika** (ang. *chemical shift*), ki je resonančna frekvenca magnetnega jedra, podana glede na resonančno frekvenco standarda, ki je navadno TMS (tetrametilsilan) v enotah deli na milijon oz. ppm (ang. *parts per million*). S tako

predstavitvijo premikov lažje identificiramo in ločimo posamezna preiskovana magnetno aktivna jedra molekule (Slika 4). Vrednosti kemijskih premikov za različne funkcionalne skupine in atome so podane v tabelah, ki so dostopne tudi na spletu, in nam olajšajo asignacijo posameznih signalov jeder v spektru.

Magnetno aktivna jedra (npr. ^1H), ki se nahajajo v kemijsko enakem okolju, so kemijsko in ponavadi tudi magnetno ekvivalentna. Ti protoni vstopajo v resonanco pri istem kemijskem premiku in dajo enoten signal v dobljenem NMR spektru. Tako na NMR spektru molekule *tert*-butilmetiletra na Sliki 4, trije vodikovi atomi metilne CH_3 skupine absorbirajo energijo pri isti frekvenci RF valovanja in signal dobimo pri določenih vrednostih kemijskega premika. Preostalih devet atomov *tert*-butilne (*tert*-Bu) skupine pa absorbira pri drugi frekvenci in signal zaznamo pri drugi vrednosti kemijskega premika.

Pri interpretaciji NMR spektra nas zanima tudi **intenziteta posameznega signala**, to je intenziteta izsevane energije pri frekvenci relaksacije posameznega jedra, ki je podana s površino pod vsakim signalom in je sorazmerna s številom posameznih magnetno ekvivalentnih jeder. Tako lahko na primeru *tert*-butilmetiletra vidimo, da je inteziteta signalov v razmerju 1: 3, kar sovpada s številom magnetno ekvivalentnih protonov v molekuli: trije ekvivalentni protoni metilne skupine in devet ekvivalentnih protonov *tert*-butilne skupine.

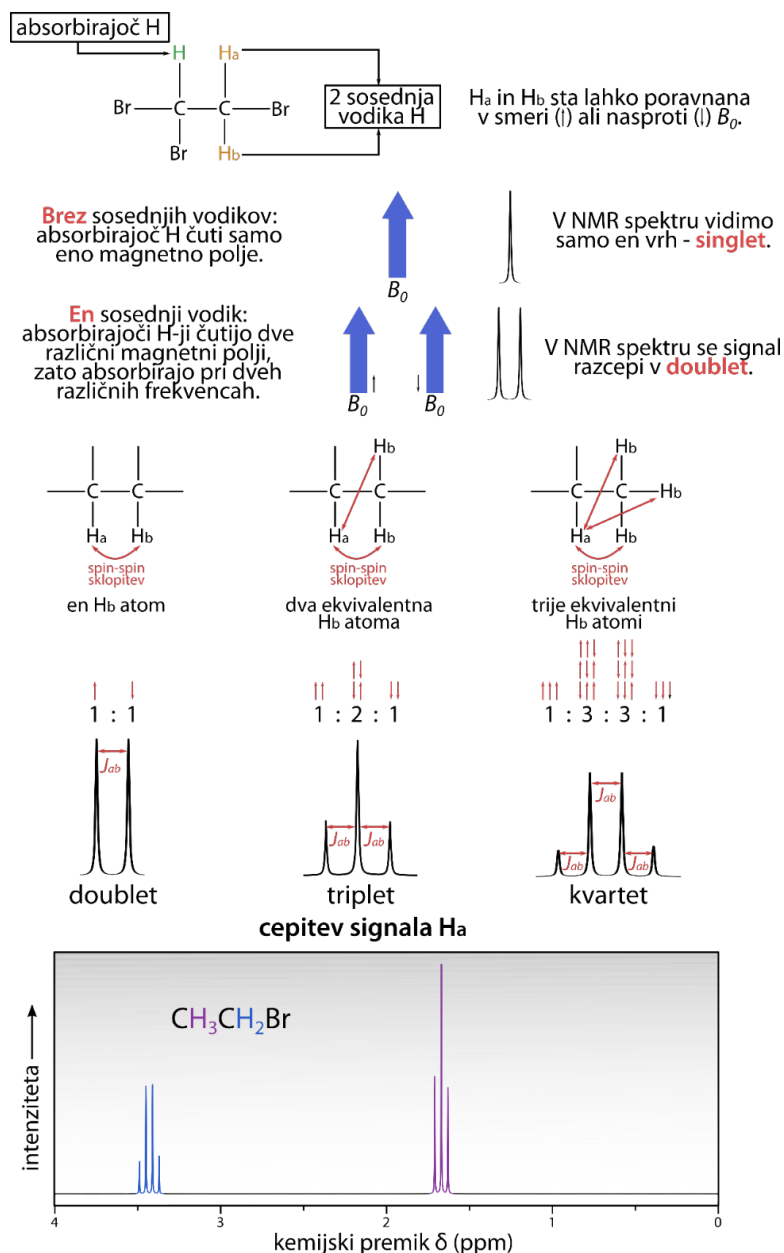


Slika 4. Različna okolica, v kateri se magnetna jedra nahajajo, pomeni tudi različno senčenje, ki ga opišemo s kemijskim premikom in je uporabno za asignacijo signalov.

3. SKLOPITVE MED JEDRNIMI SPINI IN SKLOPITVENE KONSTANTE

Poleg kemijskega premika in intenzitete signala pri večini NMR spektrov dobimo dodatne informacije o strukturi molekule še zaradi **medsebojnega delovanja magnetno aktivnih jeder**.

Interakcija magnetnega jedra s sosednjim magnetno aktivnim jedrom povzroči sklopitve med njunimi spini t. i. *spin-spin sklopitve* (ang. *spin-spin coupling*). To se v NMR spektru pokaže kot **cepitve signala** za posamezno magnetno jedro na več črt. Ta pojav sicer zaplete izgled NMR spektra, vendar v kombinaciji s prej omenjenima parametroma omogoča pridobitev mnogih pomembnih informacij o okolici, v kateri se posamezna jedra nahajajo, ter o medsebojnih interakcijah med magnetno aktivnimi jedri v molekuli, kar nam pomaga pri reševanju strukture.



Slika 5. Primer cepitve signalov v NMR spektru, sklopitvene konstante in NMR spekter bromoetana.

Oglejmo si ozadje *spin-spin* sklopitve nekoliko bolj podrobno. Da bomo lažje razumeli cepitve signalov v NMR spektru, bomo uporabili molekulo 1,1,2-tribromoetana, predstavljeno na Sliki 5 ter dva derivata: 1,1,2,2,2-pentabromoetan in 1,1,2,2-tetrabromoetan. Izhodna molekula

ima magnetno aktiven atom **H**, na sosednjem C atomu pa ima vezana dva ekvivalentna magnetno aktivna protona **Ha** in **Hb**.

Začnimo z derivatom 1,1,2,2,2-pentabromoetanom, kjer atom **H** nima sosednjih vodikov, ampak še dve dodatni molekuli Br. V primeru take molekule bi proton **H** čutil le eno magnetno polje B_0 , korigirano za senčenje elektronov, in bi v NMR spektru opazili le en signal – t. i. **singlet** z določeno vrednostjo kemijskega premika.

Situacija se spremeni, če na sosednji C atom uvedemo proton **Ha** in tako dobimo molekulo 1,1,2,2-tetrabromoetan. Ta proton **Ha** ima spin, ki je lahko poravnan v smeri magnetnega polja $\uparrow B_0$ ali pa v nasprotni smeri \downarrow . Distribucija obeh spinov je približno enakovredna - 50 % : 50 %. Zaradi sklopitve spinov sosednjih jeder **H** in **Ha**, tako proton **H** absorbira pri dveh različnih frekvencah RF valovanja, saj nanj vpliva tudi spin sosednjega magnetno aktivnega jedra, ki enkrat polje ojača, drugič pa zmanjša. Zato **H** absorbira enkrat pri višji, drugič pa pri nižji frekvenci RF valovanja. Poleg tega moramo vedeti, da na frekvenco, pri kateri absorbira **H**, tako kot v prejšnjem primeru, vpliva še senčenje elektronov. Zato za eno magnetno jedro **H** v izvedenem NMR spektru dobimo dva signala – t. i. **dublet**, z enako intenziteto, razmerje intenzitet je 1 : 1. Razdalja med signaloma dubleta, ki je razlika vrednosti kemijskega premika, je **sklopitvena konstanta J** (ang. *coupling constant*) in je merilo moči interakcije med sosednjima magnetno aktivnima jedroma. Vrednost sklopitvene konstante J podajamo v enotah frekvence (Hz) in vrednosti le-teh za različne strukturne elemente v molekulah so, podobno kot za kemijski premik, podane v različnih tabelah. Ti podatki so dostopni tudi na spletu.

Za konec si pogledajmo še primer molekule 1,1,2-tribromoetana, kjer imamo za atom **H** na sosednji C atom vezana dva ekvivalentna protona **Ha** in **Hb**. Ker sta tudi tu za vsak proton možni dve enako verjetni orientaciji spinov (\uparrow ali \downarrow), dobimo štiri možne razporeditve: $Ha/Hb \uparrow\uparrow$, $Ha/Hb \uparrow\downarrow$, $Ha/Hb \downarrow\uparrow$, $Ha/Hb \downarrow\downarrow$, pri čemer sta dve razporeditvi, $Ha/Hb \uparrow\downarrow$ in $Ha/Hb \downarrow\uparrow$, ekvivalentni. Zato proton **H** zaradi spin-spin sklopitve z dvema sosednjima magnetno aktivnima jedroma **Ha** in **Hb** absorbira pri treh različnih RF frekvencah. Signal protona **H** se v NMR spektru zato loči na tri črte, nastane **triplet** z razmerjem intenzitet signalov 1 : 2 : 1. Tudi za triplet lahko izračunamo pripadajočo sklopitveno konstanto J.

Za vajo lahko bralec poskuša razložiti še nastanek **kvarteta**, kjer je razmerje intenzitet dobljenih signalov 1 : 3 : 3 : 1, kot posledice sklopitve protona s tremi sosednjimi ekvivalentnimi magnetno aktivnimi jedri. Prav tako naj bralec preveri svoje razumevanje razloženega z analizo enostavnega NMR spektra bromoetana, na Sliki 5.

Za konec naj omenimo, da so sklopitve med spini magnetno aktivnih jeder, ki so oddaljene za več kot 3 kovalentne vezi (npr. proton oddaljen za dva C atoma stran), ponavadi prešibke, da bi jih lahko zaznali z enostavnimi NMR eksperimenti. Kot bomo videli v nadaljevanju, to omogočajo nekatere druge tehnike. Zanimivo je tudi, da v NMR spektrih nekaterih cikličnih in aromatskih sistemov opazimo tudi kompleksnejše vzorce sklopitve, ki so uporabni za lažjo interpretacijo podatkov in določitev strukture.

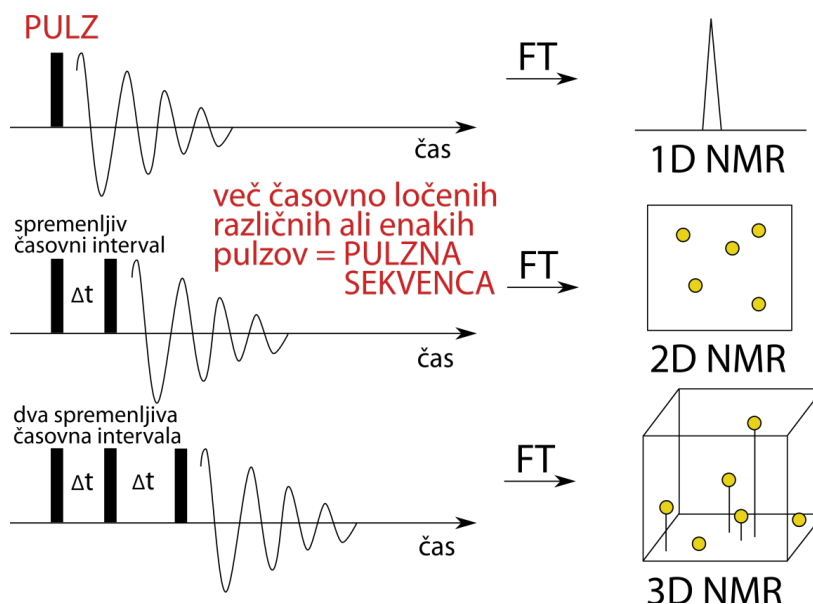
4. VEČDIMENZIONALNA NMR SPEKTROSKOPIJA

Pri večjih molekulah ter še posebej pri biomolekulah, kot so proteini in nukleinske kisline RNA in DNA, velika količina informacij v enostavnih 1D NMR spektrih, ki smo jih spoznali do sedaj, vodi do nepreglednega prekrivanja dobljenih signalov. Množica superponiranih podatkov praktično onemogoča asignacijo signalov magnetno aktivnih jeder k posameznim elementom preiskovane biološke molekule. Posamezne signale magnetno aktivnih jeder moramo zato še dodatno ločiti, da jih lahko pripišemo posameznim atomom.

Rešitev je uporaba **večdimenzionalnih NMR spektrov**, ki omogočajo, da podatke bolj učinkovito ločimo ter jih tako tudi lažje identificiramo in asigniramo posamezne signale jeder, kot to prikazuje Slika 6. Osnovna ideja večdimenzionalnih NMR eksperimentov je aplikacija več časovno ločenih radiofrekvenčnih RF pulzov (t. i. **pulzne sekvence**), ki omogoča pridobitev NMR spektrov v več dimezijah. Različni večdimenzionalni NMR eksperimenti se med seboj razlikujejo po času, uporabljeni frekvenci RF valovanja in jakosti teh impulzov.

Dvodimenzionalna NMR spektroskopija (2D NMR) je sklop NMR metod, ki dajo NMR spekter, prikazan v prostoru, ki je določen z dvema apliciranimi RF frekvencama v uporabljeni pulzni sekvenci (Slika 6). Prvi 2D NMR eksperiment je leta 1971 izvedel raziskovalec Jean Jeener. Še bolj komplicirana pa je **tridimenzionalna NMR spektroskopija (3D NMR)**, kjer apliciramo tri časovno ločene RF pulze pri različnih frekvencah v pulzni sekvenci.

Večdimenzionalni NMR eksperimenti so nadalje lahko **homonuklearni** ali **heteronuklearni**, glede na to ali v enem eksperimentu vzbujamo le eno vrsto magnetno aktivnih jeder, najpogosteje protone ^1H , ali pa jih vzbujamo več. Za določanje struktur bioloških molekul sta predvsem pomembni jedri **izotopa dušika ^{15}N** in **ogljika ^{13}C** .



Slika 6. Poenostavljen princip večdimenzionalne NMR spektroskopije in uporabe različnih pulznih sekvenc.

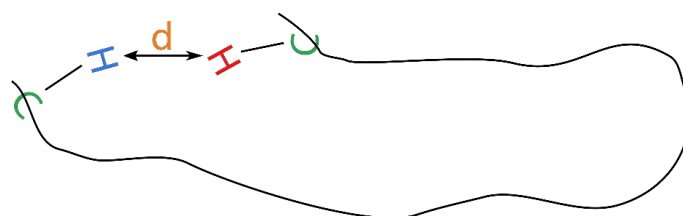
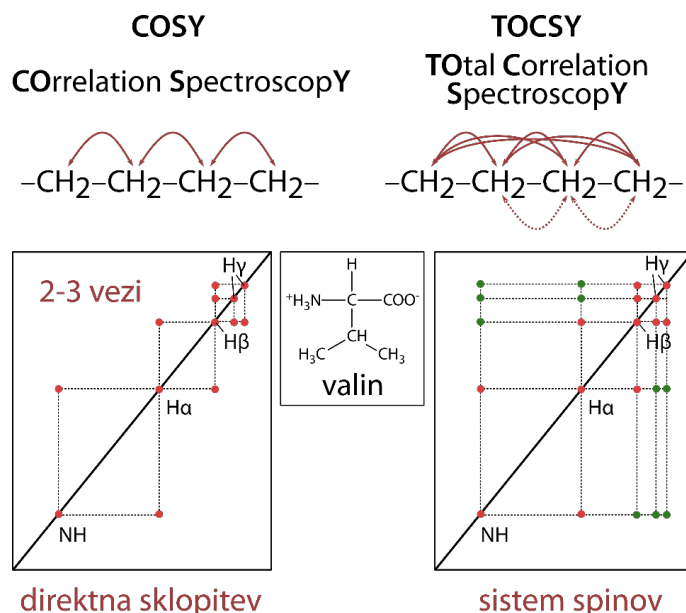
Pri enodimenzionalnih NMR spektrih smo imeli dve fazi eksperimenta - fazo *vzbujanja* z aplikacijo RF valovanja in fazo *FID*, kjer smo detektirali izsevano energijo in generirali NMR signal. Pri vseh večdimenzionalnih NMR eksperimentih pa je pulzna sekvenca nekoliko drugačna. Pri 2D NMR eksperimentih je, kot smo omenili, pulzna sekvenca sestavljena iz zaporedja vsaj dveh radiofrekvenčnih (RF) pulzov, ki sta ločena s časovnim intervalom. Skoraj vsi 2D poskusi imajo štiri stopnje: (1) pripravljalna stopnja (ang. *preparation period*), kjer vzbudimo jedra z aplikacijo prvega RF impulza. Tej sledi (2) evolucijska stopnja (ang. *evolution period*), določeno časovno obdobje, v katerem pustimo magnetno aktivnim jedrom, da prosto precesirajo kot posledico aplikacije začetnega impulza. V tretji (3) stopnji mešanja (ang. *mixing period*) z drugo serijo RF impulzov ponovno vplivamo na magnetno aktivna jedra in nato v (4) stopnji detekcije (ang. *detection period*) opazujemo »prosti indukcijski razpad« (ang. *free induction decay* FID) vzorca kot funkcijo časa, na način, ki je enak enodimenzionalnemu FT-NMR.

Po obdelavi podatkov 2D NMR spekter predstavimo z dvema frekvenčnima osema v obliki kemijskega premika, kot smo to prikazali za enodimenzionalni primer. Vsaka frekvenčna os je povezana z eno od dveh časovnih spremenljivk - to sta dolžina obdobja evolucije (čas evolucije) in čas, ki preteče v obdobju zaznavanja (čas zaznavanja). Za pridobitev obeh moramo pridobljen signal obdelati s Fourierjevo transformacijo. Intenzivnosti vrhov v 2D NMR spektru lahko predstavimo z uvedbo tretje dimenzije, veliko bolj uveljavljen prikaz pa je s konturami.

Popularen 2D ^1H NMR eksperiment je eksperiment **homonuklearne korelacijske spektroskopije (CORrelation SpectroscopY - COSY)**. Sestavljen je iz enega RF impulza ($p1$), ki mu sledi določen čas evolucije ($t1$). Temu sledi mešanje z uvedbo drugega impulza ($p2$) ter obdobje merjenja ($t2$). 2D COSY ^1H NMR eksperiment omogoča prepoznavanje interakcij sosednjih protonov, ki se nahajajo ponavadi največ 3 kovalentne vezi narazen in predstavljajo interakcijo jeder skozi vezi (Slika 7). Na primeru aminokislina valina si lahko pogledamo, da signali za isti proton (npr. NH, Ha, Hb, Hg) v 2D COSY spektru ležijo na diagonalni. V primeru, da pride do interakcije protonov, ki so oddaljeni za tri vezi (npr. NH : Ha in Ha : Hb), dobimo signal, ki ne leži na diagonalni in kaže na skopljenost teh dveh magnetno aktivnih jeder, ker sta oddaljena med seboj za 3 kovalentne vezi.

Ker nas NMR spektroskopija tu zanima predvsem zaradi določanja 3D strukture proteinov, je pomembno vedeti, da ima več aminokislina zelo podobne 2D NMR COSY vzorce, ki jih med seboj težko ločimo. Zato pogosto uporabljamo še drug homonuklearen NMR eksperiment, kjer lahko detektiramo interakcijo protonov tudi preko več kot treh kovalentnih vezi. Ta 2D NMR tehnika se imenuje homonuklearna **popolna korelacijska NMR spektroskopija** oz. **2D TOCSY (Total Correlation Spectroscopy)**. 2D TOCSY NMR spekter je nekakšen generaliziran 2D COSY spekter. V TOCSY spektru navzkrižne korelacijske vrhove opazimo ne le za jedra (protone), ki so neposredno povezana preko nosilne C-C vezi (sklopitev preko 3 vezi), temveč tudi med bolj oddaljenimi jedri na skupni nosilni verigi. To si lahko ogledamo na 2D TOCSY spektru na Sliki 7. To opažanje s pridom uporabimo za prepoznavanje večjih medsebojno povezanih mrež magnetno aktivnih protonov na stranskih verigah aminokislina. Da lahko opazimo tak efekt, v pulzni sekvenci apliciramo več ponavljajočih se RF impulzov, ki med fazo mešanja povzročajo izotropno mešanje jeder. Daljši časi mešanja povzročijo, da se polarizacija širi preko večjega števila vezi kot pri COSY eksperimentu. Na primeru aminokislina valina na Sliki 7 vidimo, da

poleg 2D COSY korelacijskih vrhov opazimo za protone še dodatne vrhove, npr. tudi med protonom NH in med protonoma H β in H γ .

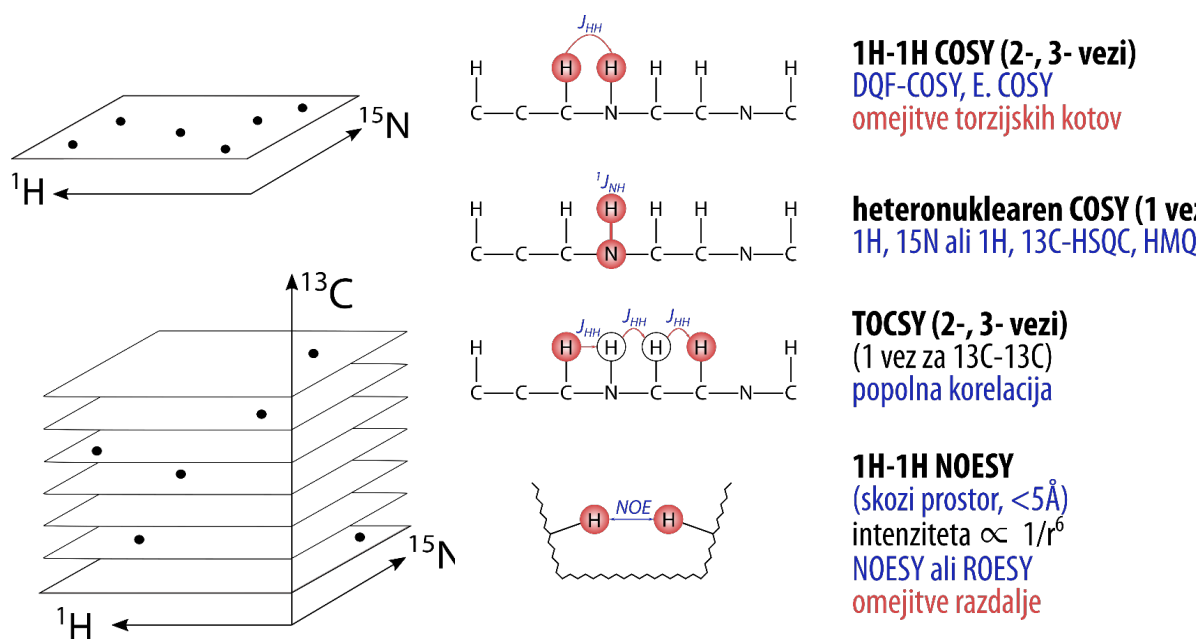


Slika 7. (Zgoraj) Sklopitve, ki jih zaznamo z 2D COSY in TOCSY NMR, ter pripadajoča spektra valina posneta z obema tehnikama. (Spodaj) Prikaz dveh neskopljenih protonov, ki sta v prostoru blizu skupaj. Njuno interakcijo lahko zaznamo z izvedbo 2D NOESY NMR eksperimenta.

Za določitev 3D strukture proteina je posebej pomemben pojav, ki ga imenujemo **jedni Overhauserjev efekt** (ang. *nuclear Overhauser effect* - NOE). NOE je po definiciji *prenos polarizacije jedrnega spina iz ene spinske populacije jedra na drugo v procesu „navzkrižne relaksacije“*. Poenostavljeno je NOE interakcija magnetno aktivnih jeder oz. sklapljanje dipolov, ki poteka preko prostora in ne preko kemijskih vezi, kot to velja za 2D COSY in TOCSY poskuse. NOE dosežemo z aplikacijo specifične pulzne sekvence in tako detektiramo jedra, ki so prostorsko blizu skupaj, vendar niso povezana preko kovalentnih vezi. **2D NOESY (Nuclear Overhauser Effect Spectroscopy) NMR spektroskopija** omogoča detekcijo interakcije med atomi, ki so oddaljeni do približno 5-6 Å. Dobljeni 2D NOESY spekter je na prvi pogled podoben 2D COSY spektru, s signali resonanc istih jeder na diagonali. Signali, ki ne ležijo na diagonali, pa nastanejo zaradi NOE in povezujejo resonanco tistih jeder, ki so prostorsko blizu. Med 2D NMR poskusi, kjer opazujemo interakcijo preko prostora, omenimo še različico 2D NOESY eksperimenta **2D ROESY (Rotating frame nuclear Overhauser Effect Spectroscopy)**, ki je uporabna pri molekulah velikosti med 1000 in 2000 Da, saj je pri njih NOE zelo šibak. Pri 2D

ROESY NMR spektroskopiji določene spine v fazi mešanja zaklenemo in s tem omogočimo lažjo detekcijo interakcij jader preko prostora.

Pri manjših proteinih in peptidih prej omenjeni homonuklearni 2D NMR spektri pogosto omogočajo določitev 3D strukture. Pri večjih proteinih pa pri homonuklearnih 2D NMR ^1H eksperimentih lahko pride do prevelikega prekrivanja dobljenih signalov, saj imajo različni protoni enake ali zelo podobne kemijske premike. Zato za večje proteine kombinacija ^1H 2D NOESY in 2D COSY spektrov pogosto ne zadostuje več za določitev 3D strukture. V ta namen pričnemo izvajati še **heteronuklearne 2D NMR eksperimente**, kjer dobljeni signali temeljijo na spajanju magnetno aktivnih jader dveh različnih vrst. Pogosto sta ti dve jedri proton ^1H in neko drugo magnetno aktivno jedro. Najpogosteje uporabimo magnetno aktivna jedra izotopa dušika ^{15}N in ogljika ^{13}C . Ker je le-teh v naravi malo, moramo protein, ki ga uporabimo za NMR eksperiment, pridobiti tako, da ga obogatimo s temi magnetno aktivnimi jedri. To je večinoma drag in dolgotrajen postopek, vendar je pri nekaterih proteinih nujen za določitev 3D strukture. Izotopsko označene proteine dobimo tako, da bakterije razmnožujemo v okolju z izotopsko označenimi hranili.



Slika 8. (Levo) ^1H - ^{15}N HSQC NMR in primer na rezine razdeljenega NMR spektra, ki ga dobimo s trojnimi resonančnimi eksperimenti. (Desno) Shematski prikaz različnih 2D NMR eksperimentov, ki jih uporabljamo pri določitvi 3D strukture proteinov.

Najpogosteje uporabljeni heteronuklearni eksperiment je ^1H - ^{15}N HSQC (**Heteronuclear Single-Quantum Correlation**) NMR spektroskopija. Ta 2D NMR poskus zazna interakcije med magnetno aktivnimi jedri dveh različnih vrst, ki so ločena z eno kovalentno vezjo, konkretno med dušikom ^{15}N in protonom ^1H . Kot rezultat dobimo v 2D NMR spektru signal za par sklopljenih jader, katerih koordinati sta kemijska premika za oba atoma (Slika 8). To je posebej uporabno pri določanju 3D strukture proteinov, saj imajo vse aminokisliline, razen prolina amidni proton H vezan na dušikov atom N kot del peptidne vezi in v HSQC NMR spektru to lahko zaznamo. V spektru dobimo dodatne signale za vse N-H elemente, prisotne v stranskih

verigah. V praksi pogosto zaradi prisotnih številnih drugih dejavnikov v NMR spektru niso vidni vsi vrhovi.

HSQC 2D NMR poskus je zelo občutljiv in ga je mogoče izvesti razmeroma hitro. Gre za enega standardnih eksperimentov, ki se uporabljajo za določanje 3D strukture proteina. Pogosto se uporablja že za preverjanje primernosti proteina za določanje strukture z uporabo NMR. HSQC NMR obstaja še v dodatnih različicah, kot sta ^1H - ^{15}N TOCSY-HSQC in ^1H - ^{15}N NOESY-HSQC. Poleg 2D HSQC eksperimenta obstaja tudi **HMBC (Heteronuclear Multiple-Bond Correlation) NMR spektroskopija**, ki omogoča detekcijo korelacije med dvema različnima magnetno aktivnima jedroma preko daljše razdalje, ponavadi preko 2 do 4 kovalentnih vezi.

Za konec omenimo še **heteronuklearne 3D NMR eksperimente**, kjer opazujemo in detektiramo prenos magnetizacije preko treh različnih magnetno aktivnih jeder. Imenujemo jih tudi **trojni resonančni eksperimenti** (ang. *triple resonance experiments*). Jedra, ki jih uporabljamo so v ta namen so ^1H , ^{15}N and ^{13}C jeder. Zaradi več prisotnih „dimenzij“ in sklapljanj prek različnih jeder je asignacija signalov lažja.

Pri trojnih resonančnih eksperimentih vzamemo za izhodišče peptidno ogrodje, ki ga sestavljajo atomi **CA-C(O)-NH**, ki so vsi magnetno aktivni. Z uporabo specifični pulznih sekvenc lahko nato selektivno opazujemo prenos magnetizacije med temi atomi peptidnega ogrodja. Ime trojnega resonančnega eksperimenta je odvisno od tega, kako se magnetizacija, glede na uporabljeno pulzno sekvenco, prenaša preko peptidne vezi Tako poznamo **(1) HNCO**, **(2) HN (CA) CO**, **(3) HNCA**, **(4) HN (CO) CA**, **(5) HNCACB** in **(6) CBCA (CO) NH** trojne resonančne eksperimente. Magnetna jedra v imenu posameznega eksperimenta so razvrščena v istem zaporedju kot poteka prenos magnetizacije. Če so jedra navedena v oklepajih, to pomeni, da so vključena v pot prenosa magnetizacije, vendar za njih ne merimo signalov. Pri vseh šestih izvedenkah imamo v 2D spektru spodaj ^1H - ^{15}N ravnino, ki je podobna kot 2D spekter HSQC, ki jo v 3D razširimo s ^{13}C dimenzijo (Slika 8). Velikokrat te 3D spektre vizualiziramo kot »rezine« različnih 2D NMR spektrov, ki jih za lažjo interpretacijo podatkov izrežemo iz celotnega 3D spektra pri konstantnem kemijskem premiku enega izmed magnetnih jeder.

5. DOLOČANJE 3D STRUKTURE PROTEINOV Z NMR SPEKTROSKOPIJO

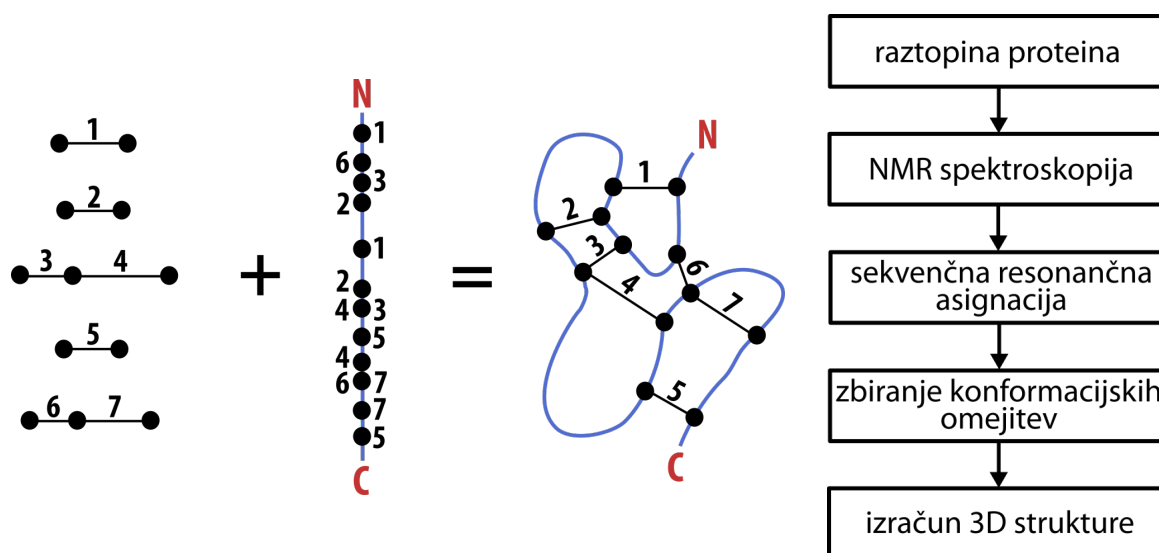
NMR spektroskopija se uporablja tudi za določanje 3D strukture proteinov. Prednost pred rentgensko kristalografijo je tudi v tem, da poleg strukturnih podatkov NMR poda tudi veliko podatkov o dinamiki proteinov in njihovih kompleksov. Poleg tega dobimo podatke o strukturi v raztopini, kar ustreza fiziološkim pogojem.

Tehniko so v drugi polovici 20. stoletja razvili raziskovalci **Richard R. Ernst** in **Kurt Wüthrich** na ETH v Švici ter **Ad Bax**, **Marius Clore** in **Angela Gronenborn** na National Institute of Health (NIH) v ZDA. Prvi mejnik se je zgodil leta 1957, ko je bil objavljen prvi NMR spekter proteina. Leta 1980 je Kurt Wüthrich opisal metodo določanja strukture proteina z NMR spektroskopijo, ki so jo kasneje prej omenjeni raziskovalci optimizirali in izboljšali. Leta 1986 je bila določena 3D struktura proteina tako z NMR spektroskopijo kot tudi s proteinsko kristalografijo, kar je omogočilo primerjavo rezultatov. Leta 2002 pa je Kurt Wüthrich prejel Nobelovo nagrado za kemijo za dosežke na področju določanja proteinske strukture z NMR.

Najprej predstavimo osnovni ideji NMR spektroskopije proteinov, kot to prikazuje Slika 9. Posamezne korake metode, ki so prav tako predstavljeni na tej sliki, bomo nato natančneje razložili v nadaljevanju.

1. IDEJA: S pomočjo različnih homonuklearnih in heteronuklearnih NMR eksperimentov pridobimo 2D in 3D NMR spektre, ki bodo z metodo **sekvenčne asignacije** omogočili **določitev položajev** vseh preiskovanih magnetno aktivnih jeder v preiskovani biološki molekuli ter identifikacijo **geometrijskih omejitev - vezne razdalje in vrednosti torzijskih kotov** med njimi.

2. IDEJA: S pomočjo računskih metod izračunamo 3D strukturo proteina, ki ustreza dobljenim geometrijskim omejitvam.



Slika 9. Osnovna ideja proteinske NMR spektroskopije in koraki proteinske NMR spektroskopije.

A) PRIPRAVA VZORCA

Za pričetek NMR eksperimenta potrebujemo raztopino dobro očiščenega proteina. Običajno je volumen vzorca med 300 in 600 μL , koncentracija proteina pa med 0.1 in 3 mmol. Podobno kot pri rentgenski kristalografiji lahko protein pridobimo iz naravnih virov ali s pomočjo rekombinantnih DNA tehnik z genskim inženiringom. Slednje omogoča lažjo produkcijo zadostne količine proteina in tudi izotopsko označevanje proteina, npr. izotopa dušika ^{15}N in ogljika ^{13}C za heteronuklearne NMR eksperimente. Očiščen pripravljen protein nato raztopimo v ustrezni raztopini pufru. Vzorec za NMR meritev pripravimo v posebni stekleni cevki s tanko steno.

B) IZVEDBA NMR EKSPERIMENTOV

Za določitev 3D strukture proteina je potrebno izvesti več večdimenzionalnih NMR eksperimentov. Za to uporabimo različne kombinacije pulznih sekvenc, ki omogočajo generiranje takih NMR spektrov. Kot smo omenili, ponavadi posnamemo dva tipa homonuklearnih NMR eksperimentov. Prvi tip je tisti, pri katerem detektiramo prenos magnetizacije med kemijskimi vezmi (npr. 2D COSY in TOSCY eksperimenta). Pri drugem tipu

pa prenos magnetizacije poteka skozi prostor in ne preko vezi: to sta predvsem 2D NOESY in 2D ROESY NMR eksperimenta. Prvi tip NMR eksperimentov uporabljamo predvsem za asignacijo kemijskih premikov določenemu jedru (protonu), z drugim tipom (deloma pa tudi s prvim) pa pridobimo ustrezne prostorske omejitve, ki jih potrebujemo za izračun 3D strukture.

Če imamo izotopsko označen protein, navadno z izotopom ^{15}N poskusimo najprej posneti 2D ^1H - ^{15}N HSQC NMR spekter. Ta NMR spekter prikaže korelacije med ^{15}N dušikom in nanj kovalentno vezanim protonom. Tako dobimo v spektru, kot smo že omenili, signal za vse amidne vezi glavne proteinske verige brez prolina. Poleg tega vse aminokislino, ki imajo v svoji stranski verigi N-H segment (aminokislino Trp, Asn, Gln, Arg, His in Lys), dajo še dodatne signale. ^1H - ^{15}N HSQC NMR spekter se pogosto imenuje kar »*prstni odtis proteina*«, saj ima vsak protein edinstven vzorec. Analiza ^1H - ^{15}N HSQC spektra raziskovalcem omogoča, da ocenijo, ali je v njem prisotno pričakovano število signalov in s tem prepoznajo morebitne težave zaradi različnih konformacij ali heterogenosti vzorca. Ta sorazmerno hiter heteronuklearni NMR eksperiment pomaga pri odločitvi ali izvedemo nadaljnje zahtevnejše večdimenzionalne heteronuklearne NMR eksperimente.

Če smo protein označili s ^{13}C izotopom in imamo zadosti zmogljiv NMR spektrometer, izvedemo tudi trojne resonančne eksperimente, ki smo jih že predstavili, kot so (1) HNC0, (2) HN (CA) CO, (3) HNCA, (4) HN (CO) CA, (5) HNCACB in (6) CBCA (CO) NH NMR eksperimenti. Pridobljeni kompleksni 3D NMR spektri bodo omogočili lažjo asignacijo signalov magnetno aktivnim jedrom.

Ovisno od koncentracije vzorca, moči magnetnega polja NMR spektrometra ter vrste eksperimentov lahko izvedba večdimenzionalnih NMR poskusov traja več ur ali celo več dni, da dobimo dovolj močan signal za generiranje uporabnega NMR spektra. Več dimenzij vključenih v eksperiment (2D v primerjavi s 3D NMR) le še podaljša čas meritev.

C) SEKVENČNA ASIGNACIJA NMR SIGNALOV

V naslednjem koraku analiziramo pridobljene NMR spektre in asigniramo vse nastale signale k prisotnim magnetno aktivnim jedrom, ki sestavljajo protein. To običajno dosežemo s t. i. sekvenčno asignacijo (ang. *sequential assignation*) z uporabo informacij, pridobljenih iz NMR eksperimentov. Postopek asignacije je odvisen tudi od tega, ali smo protein izotopsko označili, saj nam heteronuklearni NMR spektri z uporabo izotopov ^{15}N in ^{13}C omogočijo lažjo identifikacijo posameznih signalov.

Če protein ni bil izotopsko označen, poskusimo z uporabo 2D eksperimentov asignirati vse protone. To tehniko je prvi uporabil Kurt Wüthrich. Najprej uporabimo homonuklearne 2D COSY in 2D TOCSY spektre, ki omogočajo zaznavo tistih protonov, ki so povezani preko kemijskih vezi. Kot že omenjeno, COSY spekter prikaže interakcijo med protoni, vezanimi na sosednje C atome, TOCSY pa omogoča zaznavo interakcije protonov, ki so oddaljeni tudi za več C atomov. Tako na primer v 2D COSY spektru α proton aminokislino zaradi prenosa magnetizacije da signal z β protoni. Analogno β protoni dajo signal z α in γ protoni, če so v preiskovani aminokislini seveda prisotni, ter γ protoni z β in δ protoni. V 2D TOCSY NMR spektih α proton in posledično vsi drugi protoni magnetizirajo vse preostale prisotne protone (npr. β , γ , δ), če so le-ti prisotni in povezani z neprekinjeno stransko verigo posamezne aminokislino.

2D COSY in 2D TOCSY NMR eksperimenta omogočata izgradnjo **spin sistemov** (ang. *spin system*), za posamezne aminokislino, ki jih sestavlja seznam karakterističnih kemijskih premikov peptidnega protona, α protona in preostalih protonov v stranskih verigah. Ko identificiramo spinske sisteme posameznih aminokislinskih, jih poskušamo povezati med seboj z uporabo 2D NOESY NMR spektrov. Kot sedaj že vemo, NOESY eksperimenti omogočajo interakcijo protonov preko prostora in zato bomo v spektru opazili signale za protone, ki so prostorsko blizu skupaj, ne glede na to, ali so del istega spinskega sistema ali ne. S kombinacijo obeh pristopov lahko za manjše proteine (nekje do velikosti 10 kDa), ob poznavanju primarne sekvence, asigniramo vsak signal posamezni aminokislini. Pri večjih proteinih pri homonuklearnih 2D NMR eksperimentih opazimo močno prekrivanje signalov, saj imajo določeni protoni zelo podobne ali pa celo identične kemijske premike in jih ne moremo ločiti. Z večanjem velikosti proteina ta težava postaja le še bolj izrazita.

Pri večjih proteinih moramo zato protein obogatiti z drugimi magnetno aktivnimi jedri, predvsem ^{15}N in ^{13}C , da lahko s pomočjo heteronuklearnih 2D in 3D NMR eksperimentov asigniramo vse signale. Najprej navadno posnamemo »prstni odtis proteina« oz. ^1H - ^{15}N HSQC NMR spekter. Ta sorazmerno hiter heteronuklearni NMR eksperiment pomaga pri odločitvi, ali izvedemo nadaljnje zahtevnejše večdimenzionalne heteronuklearne NMR eksperimente. Prav tako podatke iz spektra uporabimo za lažjo asignacijo amidnih protonov in protonov stranske verige proteina, ki so vezani na dušik, saj oboji dajejo signal pri tem eksperimentu. Nadalje se lahko poslužimo še ^{15}N NOESY-HSQC in ^{15}N TOCSY-HSQC različice NMR eksperimentov, ki podajo še natančnejše informacije tako skapljanja protona z jedrom ^{15}N preko vezi kot tudi preko prostora.

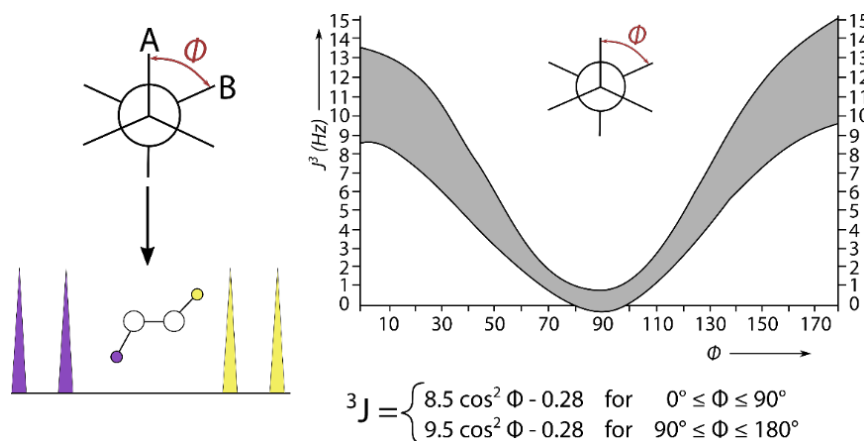
Če je protein označen še z jedrom ^{13}C , izvedemo še različne trojne resonančne NMR eksperimente, ki nam pokažejo, kako se prenaša magnetizacija preko amidne (oz. peptidne vezi) v proteinih. Dobimo več 3D NMR spektrov, ki nam omogočajo, da signale ločimo in tako lažje pripišemo posamezni signal specifičnemu magnetnemu jedru.

D) ZBIRANJE KONFORMACIJSKIH OMEJITEV

Ko imamo enkrat vse NMR signale asignirane k posameznim magnetno aktivnim jedrom preiskovanega proteina, nas zanimajo prostorske omejitve med njimi. Glavna parametra, ki jih potrebujemo, sta omejitvi razdalj (ang. *distance restraints*) in omejitve kotov (ang. *angle restraints*) med njimi. Ti podatki nam bodo v pomoč pri izračunu končne 3D konformacije proteina.

Največ prostorskih omejitev dobimo iz 2D NOESY eksperimentov. Dobljeni NOESY signal med dvema jedroma nakazuje, da sta le-ti prostorsko blizu skupaj. Navadno je ta razdalja v intervalu med 1.8 in 6 Å, kjer lahko opazimo NOE. Intenzivnost posameznega NOESY signala je sorazmerna z $1/r^6$ razdalje med jedroma, in tako, glede na intenziteto vrha, lahko ocenimo razdaljo med jedroma. Razmerje med intenziteto in razdaljo s to zvezo je le približno, zato običajno določimo interval razdalje za vsako indetificirano interakcijo med jedri. Proces identifikacije in ocene razdalj je navadno časovno zelo zahteven in pri tem uporabljamo več računalniških programov, kot so *PASD*, *XPLOR-NIH*, *UNIO*, *CYANA*, *ARIA* in drugi. Uporaba ^{15}N in ^{13}N NOESY NMR eksperimentov še poveča kvaliteto pridobljenih podatkov, saj omogoča boljšo ločitev in asignacijo dobljenih signalov.

Poleg prostorskih omejitev in identifikacije jeder, ki so prostorsko blizu skupaj, je mogoče iz eksperimentov določiti tudi nekatere **omejitve torzijskih kotov** vezi, predvsem ψ in ϕ torzijskih kotov glavne proteinske verige. Eden od pristopov je uporaba **Karplusove enačbe**, poimenovane po Nobelovem nagrajencu za kemijo Martinu Karplusu, ki opisuje odvisnost vrednosti sklopitvene konstante $^3J_{H,H}$, (npr. med protonoma A in B, ki sta 3 vezi narazen) in njunim torzijskim kotom (npr. ϕ ali ψ kotoma) (Slika 10). Omejitve torzijskih kotov lahko alternativno ocenimo tudi direktno iz vrednosti kemijskih premikov σ . Oba pristopa temeljita na opažanju, da geometrija torzijskega kota vpliva na vrednost sklopitvene konstante J in kemijskega premika σ med magnetno aktivnima jedroma A in B.



Slika 10. Sklopitvene konstante in torzijski koti. Prikaz Karplusove enačbe, ki opisuje odvisnost vrednosti sklopitvene konstante $^3J_{H,H}$ in torzijskega kota.

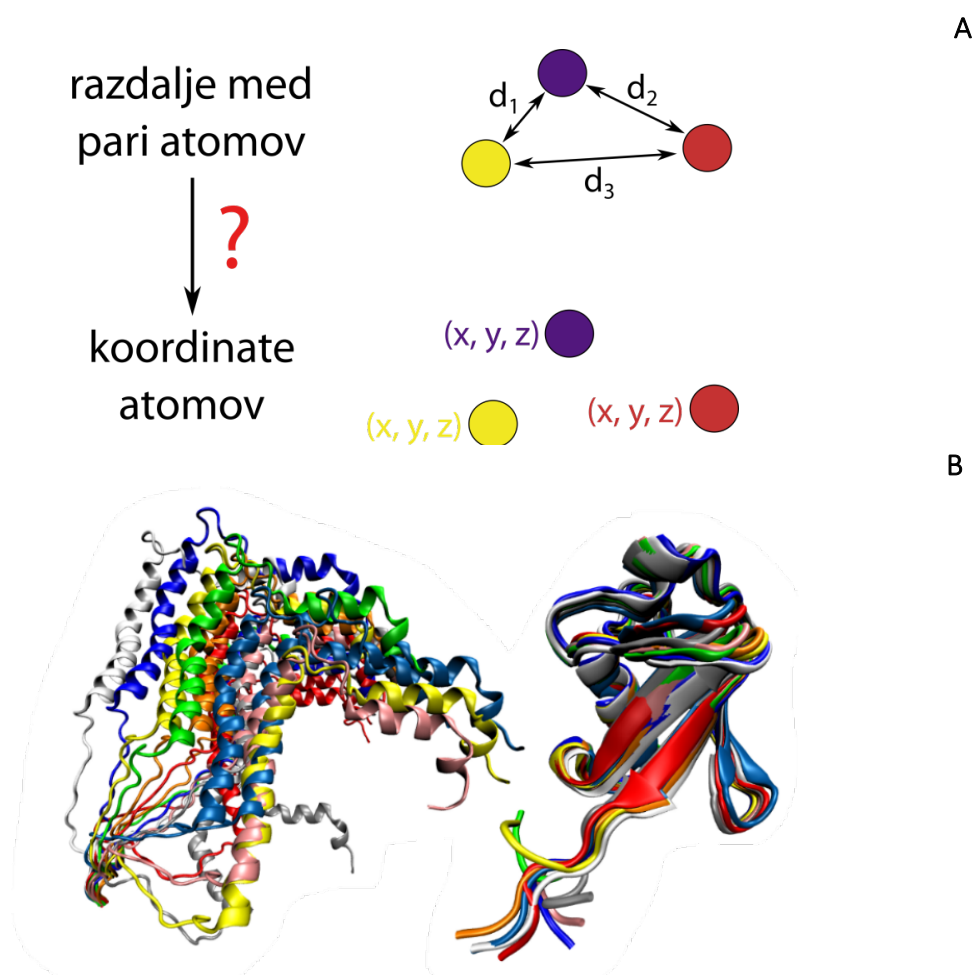
E) IZRAČUN IN VALIDACIJA 3D STRUKTURE PROTEINA

Na koncu asignacije signalov z uporabo različnih NMR spektrov kot so 2D NOESY, 2D COSY, 1H - ^{15}N HSQC NMR ter različni 3D NMR spektri, ki jih dobimo s trojnimi resonančnimi eksperimenti vemo, kateremu atomu pripada posamezen NMR signal. Za vsako aminokislino dobimo iz različnih NMR spektrov tudi od 10 do 20 **omejitev razdalj** med jedri na posamezno aminokislino. Na osnovi kemijskih premikov in vrednosti sklopitvenih konstant pridobimo tudi **omejitve torzijskih kotov**. Nadalje lahko z dodatnimi **izmenjevalnimi eksperimenti** šibkeje vezane protone zamenjamo z devterijem (ang. *deuterium exchange experiments*) in tako identificiramo **donorje vodikove vezi**.

S pomočjo računalniške simulacije izračunamo 3D strukturo proteina, ki najbolj zadosti dobljenim strukturnim omejitvam. Uporabljamo različne pristope, kot so **distančna geometrija**, **molekulska dinamika** in **metoda simuliranega ohlajanja** (ang. *simulated annealing*). Pri tem se poslužimo že prej omenjenih računalniških programov, kot so *XPLOR-NIH*, *CYANA* ali *GeNMR*, ki poleg dobljenih omejitev pri izračunu 3D strukture upoštevajo tudi splošne lastnosti proteinskih molekul.

Kot rezultat računov dobimo več možnih rešitev, družino 3D molekulskih modelov proteina (ang. *ensemble of structures*), ki ustrezajo začetnim robnim pogojem. Ko z upoštevanjem geometrijskih omejitev ter splošnih lastnosti proteinskih molekul programi izračunajo 3D model, temu sledi izračun njegove energije. Strukturo nadalje še geometrijsko optimiziramo z

uporabo različnih empiričnih potencialnih funkcij in optimizacijskih algoritmov, ki jih bomo spoznali v nadaljevanju. Tudi končni rezultat vedno predstavimo kot družino možnih konformacij proteinske strukture (Slika 11B).



Slika 11. (A) Ideja za gradnjo 3D molekulskega modela z upoštevanjem pridobljenih geometrijskih omejitev. **(B)** Z NMR spektroskopijo pridobljena 3D struktura proteina je vedno predstavljena kot družina možnih 3D struktur. Na levi je bolj fleksibilen protein, ki lahko obstaja v več različnih konformacijah, na desni pa protein z bolj rigidno strukturo.

Kvaliteto dobljenih 3D strukture proteina moramo na koncu še ovrednotiti in validirati. Na splošno bo kvaliteta modela odvisna od količine in kakovosti eksperimentalnih NMR podatkov, ki smo jih uporabili za njegovo generiranje, ter pravilne interpretacije teh podatkov. V praksi ni "standardnega proteina", s katerim bi primerjali dobljeni model. Zato je natančnost podana glede na stopnjo ujemanja modela in nabora eksperimentalnih podatkov.

Za oceno kvalitete dobljene 3D strukture lahko izračunamo RMSD (ang. *root-mean square distance*) vrednosti za glavno proteinsko verigo ter tudi za njene stranske verige za vse dobljene rešitve. Kot bomo to podrobneje spoznali v prihodnjem poglavju, je RMSD povezan s standardno deviacijo atomov, ki jih primerjamo, in predstavlja parameter za 3D primerjavo konformacij. Večinoma velja kriterij, da mora biti RMSD glavne verige manjši od 0.5 Å pri vseh konformacijah, pri čemer so vrednosti za stranske verige lahko precej večje (več kot 1 Å). Prav tako kot pri proteinski kristalografiji lahko tudi tu izračunamo Ramachandranov diagram, ki

ovrednoti, ali so konformacije strukture dovoljene in identificira elemente sekundarne strukture.

V preteklosti so bile strukture, določene z NMR, na splošno slabše kakovosti od tistih, ki so bile določene z rentgensko difrakcijo. Deloma je to posledica manjše količine informacij, ki jih vsebujejo podatki, ki jih pridobimo z NMR spektroskopijo. Zaradi tega je postala običajna praksa ugotavljanja kakovosti družin proteinskih konformacij pridobljenih z NMR primerjava s 3D strukturo tega proteina, ki je določena z rentgensko difrakcijo. Žal za mnoge proteine struktura, rešena s to metodo, še ne obstaja. Prav tako so proteini v raztopini fleksibilne molekule, zato lahko kristalna struktura proteina, ki je statična, premalo natančno opiše to fleksibilnost, ki jo lahko detektiramo z NMR eksperimenti. Pri primerjavi struktur določenih s kristalografijo in tistih, določenih z NMR spektroskopijo velja poudariti, da kristalno polje zmanjša fleksibilnost proteina. Pri **validaciji proteinskih struktur** uporabljamo več programov, kot sta *PROCHECK* in *WHAT IF*, ki temeljita na statističnih metodah, medtem ko drugi programi temeljijo na fizikalnih principih, kot program *CheShift* ali pa na mešanici obeh - program *PSVS*.

Z NMR eksperimenti lahko študiramo tudi **dinamiko proteina**. Med NMR eksperimenti se, tako kot pri drugih spektroskopskih tehnikah, zgodijo tudi procesi prenosa energije, in zato lahko merimo relaksacijske čase, ki jih posamezno jedro potrebuje, da se vrne v svojo izhodno stanje. NMR relaksacija se dogaja preko dveh procesov: to sta spin-spin relaksacija in spin-mrežna relaksacija, ki sta povezani z dvema hitrostnima konstantama T_1 in T_2 . Če te konstante izmerimo, lahko pridobimo podatke o internih gibanjih znotraj molekule in identificiramo fleksibilne regije, ki so pomembne za funkcijo proteina.

NMR metodo lahko uporabljamo tudi za **študij interakcij proteinov z drugimi molekulami**, kot so npr. proteini, nukleinske kisline, peptidi in organske molekule-ligandi. Pri teh študijah primerjamo spremembe kemijskih premikov in relaksacijskih časov pred in po interakciji z drugo molekulo. S tem lahko npr. identificiramo interakcijske površine med (bio)molekulama. Prav tako lahko s spremembami kemijskih premikov pri vezavi majhnih organskih molekul identificiramo nove ligande ali fragmente, ki so primerni, za razvoj novih zdravilnih učinkovin, če je protein npr. pomembna terapevtska tarča.

Andrej Perdih

molekulsko modeliranje

```
function quicksort(array)
  var list less, equal, greater
  if length(array) <= 1
    return array
  select a pivot value pivot
  for each x in array
    if x < pivot then list = less
    if x = pivot then list = equal
    if x > pivot then list = greater
  return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

Velikokrat tridimenzionalne strukture bioloških molekul obravnavamo tudi z računskimi metodami, da lahko pridobimo vpogled v njihovo strukturo in funkcijo, ki je povezana na primer s konformacijskimi spremembami ali z vezavo različnih ligandov. To področje teoretičnih raziskav imenujemo tudi **računska kemija** (ang. *computational chemistry*), kjer raziskujemo s pomočjo računalnika. Pogosto za tak tip raziskav uporabljamo tudi izraz **molekulsko modeliranje** (ang. *molecular modelling*) kot krovni izraz za teoretične metode in tehnike, ki omogočajo opisovanje in simuliranje strukture in obnašanja atomov in molekul. Da bomo lahko razumeli, kako uporabljamo te metode pri študiju 3D stukture bioloških molekul, se bomo najprej seznanili s osnovami molekulskega modeliranja, ki jih bomo potrebovali za poglobljeno razumevanje.

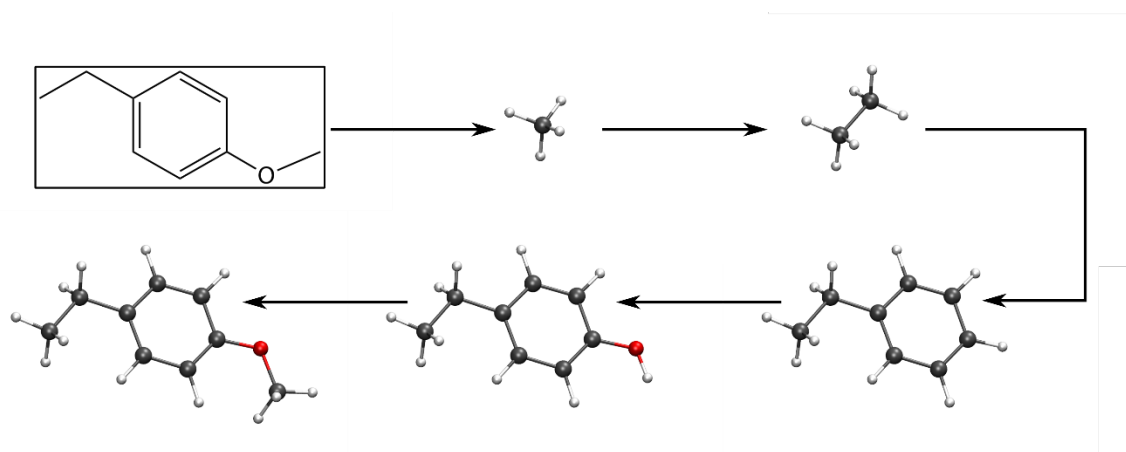
1. VIZUALIZACIJA MOLEKUL IN BIOMOLEKUL

Molekulska grafika se ukvarja z grafično predstavitvijo molekulskega modela. Omogoča prikaz oz. *vizualizacijo* 3D modelov atomov in molekul, uporabljamo pa jo tudi, kot bomo videli, za predstavitev drugih rezultatov računov z molekulami, kot so simulacije molekulske dinamike, kvantnokemijskih računov, molekulskega sidranja, farmakofornega modeliranja itd.

Molekulsko modeliranje se prične z izgradnjo tridimenzionalnega molekulskega modela, ki predstavlja atomistično predstavitev strukture obravnavane molekule. Za gradnjo 3D molekulskega modela uporabimo gradnike, tj. atome, ki jih s pomočjo grafičnega vmesnika povežemo med seboj s standardnimi geometrijami, znanimi medatomske razdaljami, tipi hibridizacije in valenčnimi koti. Pri takšnem pristopu izgradnje tridimenzionalnih modelov moramo previdno definirati vse parametre strukture, kot so povezanost atomov, stereokemija ter tautomerija, saj vse nadaljnje računske operacije temeljijo na kvaliteti teh vhodnih podatkov. Spodaj je predstavljen primer take gradnje molekule.

Primer

Gradnja 3D konformacije 4-etil-metoksifenila s pomočjo grafičnega vmesnika.



Druga možnost pridobitve začetnih koordinat 3D molekulskega modela je uporaba eksperimentalnih struktur, ki jih pridobimo s prej obravnavanimi metodami rentgenske difrakcije (X-ray) ali pa NMR spektroskopije. Ta način pogosto uporabljamo takrat, ko z metodami molekulskega modeliranja preučujemo biološke makromolekule - encime, receptorje, DNA, RNA, včasih pa tudi pri raziskavah majhnih organskih molekul.

Eksperimentalni podatki o določenih strukturah bioloških makromolekul so dostopni v javni bazi proteinskih struktur **Protein Data Bank (PDB)**. Druga pomembna baza struktur je **Cambridge Structural Database (CSD)**, ki omogoča dostop do strukturnih podatkov o majhnih molekulah. Koordinate lahko pridobimo tudi z rezultati homolognega modeliranja in drugih metod napovedi 3D proteinskih struktur (glej Poglavje V). Večina grafičnih programov uporablja standardno označevanje oz. barvanje atomov, kot ga prikazuje Preglednica 1, čeprav lahko tudi med programi obstajajo manjše variacije.

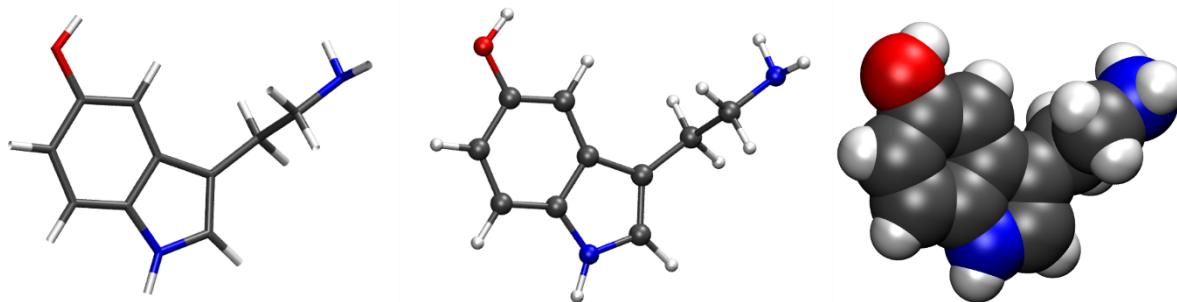
Preglednica 1. Standardne barvne oznake atomov.

Atom	Barva atoma
vodik	<i>bela</i>
ogljik	<i>črna/siva</i>
kisik	<i>rdeča</i>
dušik	<i>modra</i>
žveplo	<i>rumena</i>
fosfor	<i>oranžna</i>
fluor	<i>zelena</i>
klor	<i>zelena</i>
brom	<i>rjava</i>
jod	<i>vijolična</i>

Pri grafični predstavitvi konformacije organske molekule, atome najpogosteje predstavimo s kroglicami, kemijske vezi pa s paličicami, ki jih povezujejo. Take modele imenujemo **modeli kroglic in paličic** (ang. *ball-and-stick models*). Zaradi preglednosti lahko atome predstavimo tudi kot točke, povezane s kemijskimi vezmi. Dobimo enostavne **žičnate** (*wire*) ali **paličaste** (*stick*) **modele**.

Omenjene predstavitve molekul ne podajo veliko informacij o prostoru (volumnu), ki ga zavzema molekula. V ta namen uporabljamo **CPK** (kalotne) **modele**, poimenovane po začetnicah imen njihovih avtorjev: **Corey, Pauling in Koltun**. V CPK modelu atomi predstavimo kot sfere, ki se med seboj prekrivajo in tvorijo molekulo. Velikost sfere je določena z **van der Waalsovimi radijem** posameznega atoma, ki predstavlja približek prostora, ki ga zavzema atom. Tako je van der Waalsov radij atoma vodika 1.2 Å, ogljika 1.7 Å, dušika 1.55 Å in kisika 1.52 Å, ter ga imenujemo tudi **atomski** ali **molekulski volumen**, celotno površino, ki jo očrta, pa **van der Waalsova površina**.

Na Sliki 1 je z različnimi modeli predstavljena mala organska molekula - neurotransmitter serotonin. Molekulske modele lahko predstavimo tudi brez prikazanih vodikovih atomov. To storimo predvsem takrat, ko bi celotna mreža vodikov ovirala vizualizacijo preučevanih strukturnih elementov.



Slika 1. Predstavitev organske molekule neurotransmiterja serotonina z različnimi modeli: palični, kroglice in paličice in CPK model.

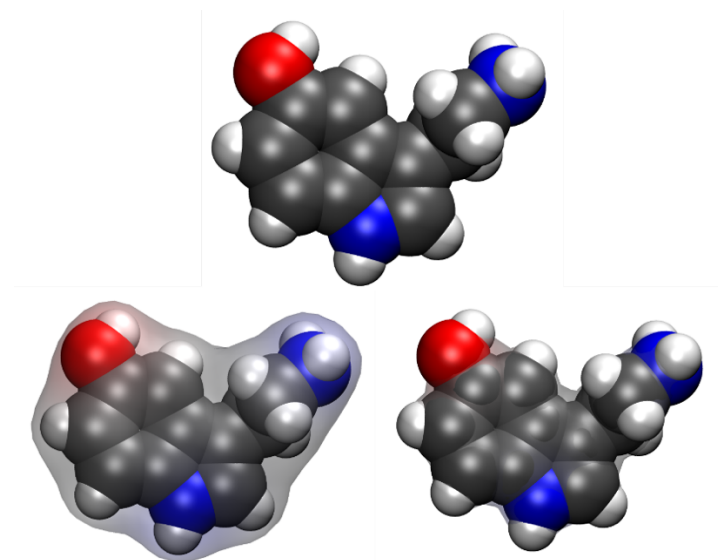
Podatke o molekuli, kot so koordinate atomov, delni naboji, povezanost atomov, topologija in še druge lastnosti molekule, ki jih bomo spoznali, zapišemo v različnih **standardnih molekulskih formatih** (ang. *molecular file formats*), ki jih programi za vizualizacijo lahko prikažejo. Obstaja več različnih formatov in večina programov za modeliranje omogoča branje in zapisovanje molekul v več zapisih. Pri bioloških makromolekulah najpogosteje uporabljamo format **pdb**, v katerem so zapisane strukture v proteinski bazi PDB. Molekule organskih spojin pogosto zapisujemo v drugih standardnih zapisih. Nekateri izmed uporabljenih formatov za zapis molekul so: **cdx** (ChemDraw eXchange file), **mol** (MDL molfile), **mol2** (SYBYL format), **sdf** (structure data file), obstaja pa še precej drugih.

Primer

Zapis organske molekule adrenalina v mol2 formatu.

<pre> @<TRIPOS>MOLECULE ADRENALIN 26 26 1 0 0 SMALL @<TRIPOS>ATOM 1 C 0.0000 0.0000 0.0000 C.3 2 C 0.0000 0.0000 1.5000 C.3 3 H 1.0267 0.0000 -0.3630 H 4 C -0.7071 -1.2247 -0.5000 C.ar 5 O -0.6505 1.1268 -0.4600 O.3 6 C -0.8350 -1.4463 -1.8764 C.ar 7 C -1.4950 -2.5894 -2.3431 C.ar 8 C -2.0270 -3.5109 -1.4333 C.ar 9 C -1.8991 -3.2894 -0.0569 C.ar 10 C -1.2392 -2.1463 0.4098 C.ar 11 H -0.4212 -0.7295 -2.5841 H 12 O -1.6211 -2.8078 -3.6999 O.3 13 O -2.6776 -4.6377 -1.8933 O.3 14 H -2.3130 -4.0062 0.6508 H 15 H -1.1397 -1.9740 1.4804 H 16 H -2.0888 -3.6179 -3.8471 H 17 H -2.9752 -5.1532 -1.1568 H 18 H -1.0088 0.0000 1.8567 H 19 H 0.5044 -0.8737 1.8567 H 20 N 0.6883 1.1921 1.9867 N.3 21 H -0.6505 1.1268 -1.4070 H 22 C 0.6883 1.1921 3.4467 C.3 23 H 1.6593 1.1921 1.6433 H 24 H 1.1927 2.0657 3.8033 H 25 H -0.3206 1.1921 3.8033 H 26 H 1.1927 0.3184 3.8033 H @<TRIPOS>BOND 1 16 12 1 2 12 7 1 3 11 6 1 4 7 6 ar 5 7 8 ar 6 13 8 1 7 13 17 1 8 6 4 ar 9 8 9 ar 10 21 5 1 11 4 1 1 12 4 10 ar 13 5 1 1 14 3 1 1 15 9 10 ar 16 9 14 1 17 1 2 1 18 10 15 1 19 2 18 1 20 2 19 1 21 2 20 1 22 23 20 1 23 20 22 1 24 22 24 1 25 22 25 1 26 22 26 1 </pre>	<p>osnovni podatki o molekuli</p> <p>koordinata atomov (x, y, z) in atomski tipi</p> <p>povezanost atomov in tip vezi</p> <p>npr. kovalentna enojna vez med atomoma 22 in 25</p>	<p>adrenalin</p>
--	---	-------------------------

Organske molekule in biološke molekule lahko prikažemo tudi s površinami. Večina razvitih metodologij za konstrukcijo molekulskih površin uporabi kroglico, ki ji lahko izberemo polmer. Če je topilo voda, je npr. polmer kroglice enak 1.4 Å. Kroglica nato potuje po van der Waalovi površini molekule. Očrt, ki ga med kotaljenjem opiše središče kroglice, predstavlja **površino, ki je dostopna topilu** (ang. *solvent-accessible surface area* - SASA). Narišemo lahko tudi od **topila izključeno površino** (ang. *solvent-excluded surface*). To površino lahko imenujemo tudi **molekulska** ali **Connollyjeva površina** (po njenem avtorju), saj je inverzna površini, dostopni topilu.



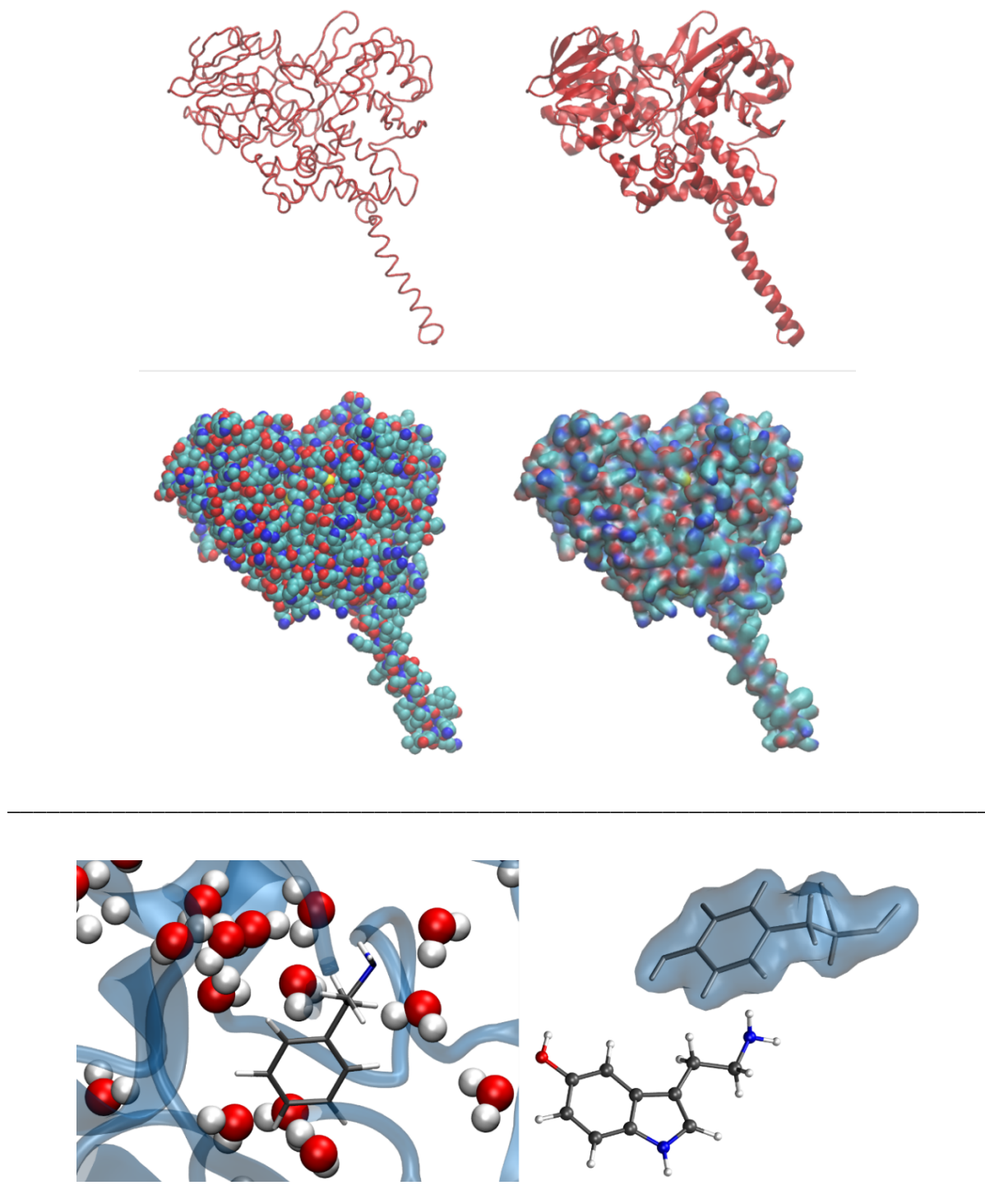
Slika 2. Površina, dostopna topilu (levo), in Connollyjeva površina (desno) za molekulo serotonina. Na vrhu je prikazan izhodni CPK model molekule.

Poleg bolj celostne predstavitve molekule lahko molekulske površine uporabimo tudi za bolj nazorno predstavo nekaterih izračunanih lastnosti molekule. Prikažemo lahko **izopovršine** ali **konture**; tridimenzionalne obrise, kjer ima preiskovana količina konstantno vrednost neke lastnosti, npr. molekulskih orbital (npr. HOMO: *Highest Occupied Molecular Orbital* in LUMO: *Lowest Unoccupied Molecular Orbital*), elektronske gostote ter elektrostatskega potenciala.

Pri vizualizaciji strukture bioloških makromolekul, tu imamo v mislih predvsem proteine, lahko shematsko predstavimo njihovo celotno 3D strukturo. To pomeni, da prikažemo potek glavne proteinske verige, na kateri lahko še dodatno poudarimo elemente sekundarne strukture (npr. α -vijačnice, β -strukture). Prav tako lahko protein vizualiziramo z njegovo površino ali pa njen volumen predstavimo s CPK modelom. Za vizualizacijo je na voljo več programov, kot so *VMD*, *PyMOL*, *Chimera*, *Hermes* in drugi. Na Sliki 3 je encim monoamin-oksidaza (MAO), ki sodeluje v procesih razgradnje bioloških aminov, prikazan z opisanimi pristopi. Natančneje si boste tehnike vizualizacije bioloških makromolekul, ogledali tudi pri eni izmed vaj pri predmetu (Vaja 2 in Vaja 3).

V uporabi so tudi kombinacije opisanih vizualizacijskih tehnik. Tako lahko na primer del proteina, ki predstavlja aktivno mesto, predstavimo shematsko s potekom glavne verige, atome liganda ter vodnih molekul pa vizualiziramo s modelom paličic in CPK modelom. Lahko prikažemo tudi interakcije liganda le z izbranimi aminokislinami. Vse to nam omogoča, da podrobno spoznamo ključne lastnosti proteina ter preučimo interakcije, ki so odgovorne za

medmolekulsko prepoznavanje ligandov. Na spodnjem delu Slike 3 si nekaj takih vizualizacijskih prijemov lahko ogledate za prej omenjeni encim monoamin-oksidozo.

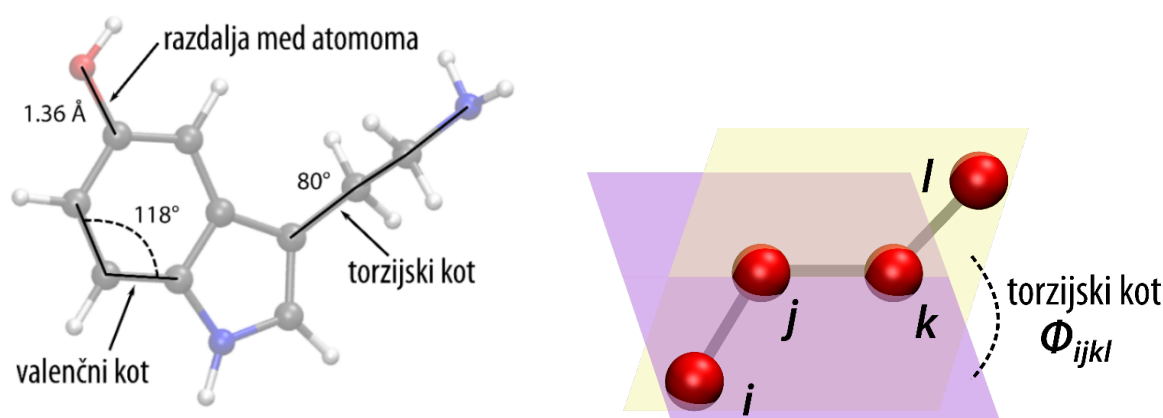


Slika 3. Vizualizacija makromolekul na primeru encima monoamin-oksidoza MAO. (Zgoraj) Prikaz poteka glavne verige proteina in dodatni prikaz elementov sekundarne strukture (alfa vijačnice, beta strukture in proteinskih zank). (Sredina) CPK model in površina, dostopna topilu encima MAO. (Spodaj) Primeri kombinacije opisanih vizualizacijskih tehnik.

Pri molekulah pogosto ovrednotimo nekaj osnovnih geometrijskih parametrov, ki definirajo **konformacijo molekule**. **Konformacija molekule** je definirana s prostorsko (3D) razporeditvijo atomov, ki sestavljajo molekulo. Shematsko si te parametre lahko ogledamo na primeru molekule serotonina na Sliki 4.

V molekuli merimo tako **vezne kot nevezne razdalje** in te navadno navajamo v Å ($1 \text{ Å} = 10^{-10} \text{ m}$). Na Sliki 4 je tako označena kovalentna C-O vez, ki je dolga 1.36 Å. Prav tako nas zanimajo tudi vrednosti kotov med tremi zaporedno vezanimi atomi, ki jih imenujemo **valenčni koti**. Na primer vrednost valenčnega kota treh označenih atomov ogljika v molekuli serotonina je 118° . Zadnji pomemben geometrijski parameter je **torzijski kot** molekule. Torzijski kot ϕ med zaporednimi atomi $ijkl$ definira kot, ki ga oklepata ravnini, ki potekata skozi atome ijk in jkl (Slika 5). Na sliki je tako definiran torzijski kot medaminskim dušikovim atomom in ogljikovim atomom indolnega skeleta, ki meri 80°

Poudarimo, da so vrednosti prvih dveh geometrijskih parametrov za dano molekulo praviloma bolj ali manj fiksne in se ne pri različnih konformacijah spreminjajo veliko oz. je za spremembo le-teh potreben velik vložek energije. Tako so spremembe valenčnih kotov stran od ravnotežnih vrednosti, ki jih definira njihova hibridizacija, povezane s steričnimi napetostmi v molekuli. Spremembe torzijskih kotov pa so pri sobni temperaturi enostavnejše. Če okrog osi rotacije niso prisotni preveliki substituenti in če je le-ta sp^3 hibridizirana (npr. C-C vez), prav obseg njihovih dostopnih vrednosti definira konformacijski prostor molekule. Več o tem, kako analiziramo konformacije molekule prek analize torzijskih kotov, bomo izvedeli v nadaljevanju.



Slika 4. (Levo) Pomembni geometrijski parametri na primeru molekule serotonina. (Desno) Definicija torzijskega kota med zaporednimi atomi $ijkl$.

2. PRILEGANJE KONFORMACIJ MOLEKUL

V molekulskem modeliranju pogosto primerjamo med seboj različne konformacije organskih molekul ter bioloških makromolekul ali pa konformacije delov dveh različnih molekul, ki imajo podobno ali enako funkcijo. Za kvantitativno oceno ujemanja uporabljamo parameter **RMSD (ang. root-mean square distance)**, ki se poleg uporabe v molekulskem modeliranju pojavlja še na drugih področjih (statistiki, računalništvu, ekonomiji itd). **RMSD omogoča določitev standardne deviacije primerjanih atomov v molekularnih strukturah**. Matematično ga definiramo kot:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i,j=1}^N d_{i,j}^2}$$

kjer je $d_{i,j}$ razdalja med atomoma i in j , ki sta bodisi identična, pri primerjavi konformacij iste molekule, ali pa imata primerljivi funkciji v dveh različnih molekulah, in N je število primerjanih razdalj. Razdaljo definiramo v ångströmih [Å]. Vsako razdaljo med primerjanima atomoma v prostoru izračunamo po formuli:

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

kjer so (x_i, y_i, z_i) in (x_j, y_j, z_j) koordinate primerjanih atomov i in j .

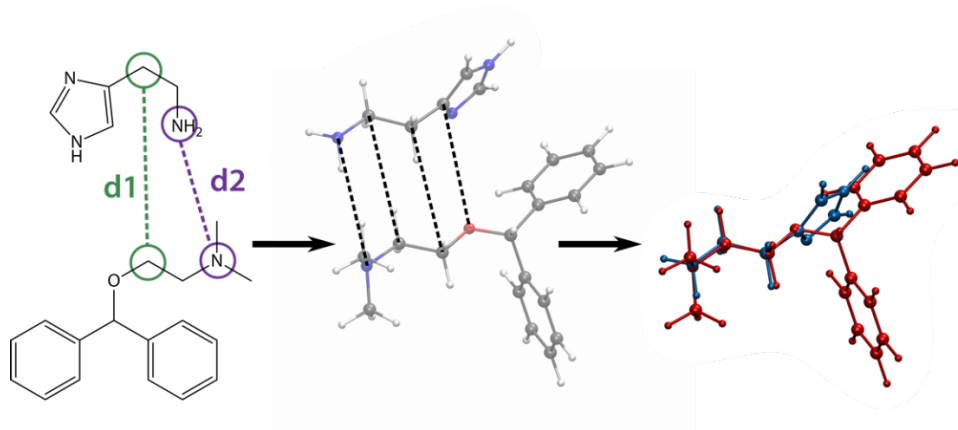
Ko izračunamo vse razdalje in jih vnesemo v zgornjo enačbo, dobimo RMSD vrednost, ki predstavlja kvantitativni parameter za ujemanje konformacij. Z optimizacijskim (imenovanim tudi minimizacijskim) algoritmom lahko poiščemo minimalno vrednost RMSD. S takim optimizacijskim postopkom eno molekulo kot togo telo, ne glede na njeno začetno lego v prostoru, **prilegamo** (ang. *alignment*) na položaj atomov druge molekule.

Ta operacija omogoča kvalitativno primerjavo dveh prileganih konformacij izračunani RMSD pa je kvantitativno merilo za podobnost le-teh. Vrednost RMSD parametra odraža standardno deviacijo le tistih parov atomov, ki smo jih primerjali. Če je vrednost RMSD med prileganimi atomi majhna, potem sta primerjana dela konformacij v obeh molekulah zelo podobna. Kadar ne primerjamo konformacij identičnih molekul, moramo za pridobitev popolne slike primerjanih sistemov obravnavati tudi strukturni in funkcijski pomen različnih delov molekul. Za lažjo ilustracijo si pogledajmo dva primera.

Primeri

3D prileganje konformacij dveh malih organskih molekul

Na sliki sta narisana dva liganda, nevrotransmitter histamin in zdravilna učinkovina difenhidramin, ki je antagonist na histaminskih H1 receptorjih. Za izračun RMSD moramo najprej določiti pare atomov, ki jih bomo prilegali. Slika prikazuje dva izbrana para: ogljika na fleksibilnih verigah molekul definirata razdaljo d_1 , aminska dušika pa razdaljo d_2 . Bralec naj za vajo označi še preostale možne pare atomov, ki imajo podobno funkcijo. Vrednost RMSD dobimo, če seštejemo vse razdalje $d_{i,j}$ med izbranimi pari atomov po zgornji formuli. Nato RMSD minimiziramo po vseh izbranih razdaljah $d_{i,j}$ in tako molekuli prilegamo - eno konformacijo prilegamo kot togo telo na drugo. Rezultat RMSD prileganja konformacij omenjenih molekul je predstavljen na desni strani. Bralec lahko za vajo razmisli, kateri deli v zdravilni učinkovini so glede na rezultat prileganja odgovorni za vezavo in kateri preprečijo fiziološko aktivacijo H1 receptorja.



3D prileganje aktivnih mest več sorodnih proteinov

V bioinformatiki imamo pri študiju 3D struktur proteinov pogosto opraviti z več proteini, ki so sorodni po funkciji ter tudi po strukturi. Ena izmed takih družin je družina protein kinaz, za katere velja, da imajo vsi predstavniki vezavno mesto za energijsko molekulo ATP. S 3D prileganjem struktur teh proteinov ali pa samo njihovih aktivnih mest lahko ugotovimo, kako podobna oz. različna so študirana vezavna mesta posameznih proteinov. Prav tako lahko identificiramo posamezne aminokisline, ki prispevajo k tem razlikam. Primer takega prileganja nekaterih protein kinaz z vezano molekulo ATP lahko vidite na spodnji sliki. Pri protein kinazah smo označili le potek glavne verige proteina, ki definira aktivno mesto.



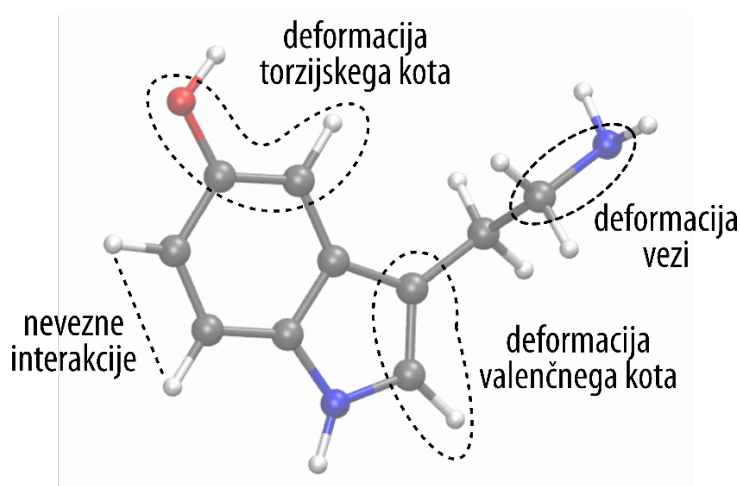
3. ENERGIJA MOLEKUL

Do sedaj smo govorili predvsem o tem, kako predstavimo in analiziramo 3D **geometrijo** organskih molekul in bioloških molekul. Drugi pomemben (in z geometrijo tesno povezan) parameter, ki ga tudi želimo določiti za vsako molekulo, pa je njihova **energija**. Zato obstajata dva pristopa. Bolj pogosto je v uporabi *empirični* pristop k računanju energije struktur, ki ga predstavlja **molekulska mehanika** (ang. *Molecular Mechanics* - MM), in ga bomo podrobneje spoznali v tem razdelku. Drugi pristop je natančnejša in fizikalno osnovana **kvantna mehanika** (ang. *Quantum Mechanics* - QM) in spoznali bomo le njene osnove.

3.1. MOLEKULSKA MEHANIKA

Molekulska mehanika (MM) je poenostavljen opis obnašanja atomov in molekul, ki temelji na predpostavki, da lahko atome opišemo kot nabite kroglice ustreznih mas, ki so med seboj povezane s kovalentnimi vezmi, ki jih opišemo kot prožne vzmeti. Formalno je MM definirana kot *teoretični opis atomov in molekul, ki omogoča računsko obravnavo strukture in izračun lastnosti atomov in molekul z uporabo metod klasične mehanike in elektrostatike*.

Struktura male organske molekule in shematično prikazane geometrijske spremembe, ki jih pri MM opišemo kot *deformacijo kovalentne vezi, deformacijo valenčnega kota, spremembe torzijskih kotov in neveznih interakcij* in imajo vpliv na končno energijo, so prikazane na Sliki 5. Kovalentno vez si v klasični molekularni mehaniki predstavljamo kot prožno vzmet, ki povezuje dva atoma v molekuli. Poleg **veznih interakcij** atomi v molekuli interagirajo tudi preko **neveznih interakcij**, ki jih opišemo z vsoto **Coulombovega potenciala** za opis elektrostatskih interakcij in **Lennard-Jonesovega potenciala** za opis van der Waalsovih interakcij.



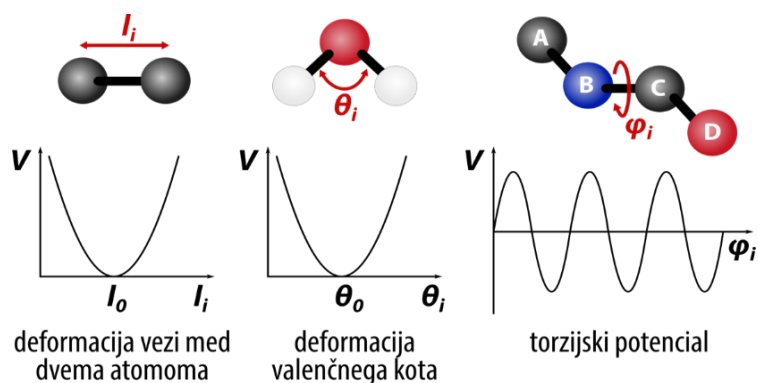
Slika 5. Struktura male organske molekule serotonina in shematično prikazane geometrijske spremembe, ki jih pri MM ločeno opišemo in imajo vpliv na končno energijo.

Empirično potencialno energijsko funkcijo (polje sil ali ang. *force field*), ki v molekularni mehaniki opisuje celoten sistem molekule s splošnimi koordinatami v obliki vektorja r^N , zapišemo s spodnjo splošno funkcijsko obliko:

$$V(r^N) = \sum_{\text{vezi}} \frac{k_{\text{vezi}}}{2} (l_i - l_{i,0})^2 + \sum_{\text{koti}} \frac{k_{\text{kota}}}{2} (\theta_i - \theta_{i,0})^2 + \sum_{\text{torzije}} \frac{k_{\text{torzije}}}{2} (1 + \cos(n\varphi - \varphi_0)) + \sum_{i=1}^N \sum_{j=i+1}^N \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

V molekularnem modeliranju se še vedno tradicionalno uporabljajo nekatere enote izven mednarodnega SI sistema enot. Energije molekul so velikokrat podane v kilokalorijah (1 kcal = 4.184 kJ) in jih normiramo na 1 mol snovi. Tako dobimo enoto kcal/mol.

- 1. Vezni členi empirične potencialne funkcije:** deformacije kovalentnih vezi, valenčnih kotov in torzijskih kotov opisujejo obnašanje med atomi, ki so kovalentno povezani (Slika 6).

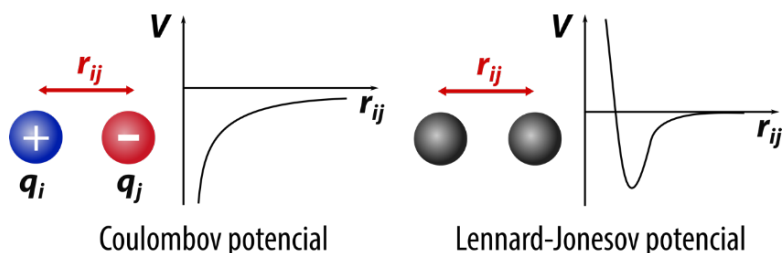


Slika 6. Prikaz opisa deformacije vezi med dvema atomoma (levo) ter deformacije valenčnega kota (na sredini) in pripadajoča energijska diagrama. Torzijski kot, definiran z zaporednimi atomi ABCD, z označeno osjo rotacije ter pripadajoči energijski diagram - torzijski potencial (desno).

Energijo **deformacije kovalentne vezi med dvema atomoma** ter **deformacije valenčnega kota**, kot prva dva člena v zgornji enačbi, opišemo z obnašanjem prožne vzmeti, za katero velja Hookov zakon. Po Hookovem zakonu je energija sorazmerna kvadratu odmika od ravnotežne razdalje med atomoma oz. kvadratu deformacije ravnotežne vrednosti valenčnega kota. V enačbi količini l_i in q_i predstavljata vezno razdaljo in valenčni kot, $l_{i,0}$ in $q_{i,0}$ pa idealno ravnotežno vezno razdaljo ter ustrezni ravnotežni valenčni kot. Ravnotežni vrednosti privzamemo iz rezultatov kvantnokemijskih (QM) računov ali iz dostopnih eksperimentalnih podatkov. Na analogen način pridobimo vrednosti za empirični konstanti k_{vezi} in k_{kota} . Pri teh dveh členih empirične potencialne funkcije so večji odmiki od idealnih leg energetsko zelo zahtevni in se zato redkeje zgodijo. Konformacijski prostor, ki je dostopen molekuli, je povezan s spremembami torzijskih kotov in razdalj, ki so odvisne od neveznih interakcij.

Enačba, ki opisuje vpliv **torzijskega kota**, odraža periodične lastnosti tega geometrijskega elementa. $k_{torzije}$ predstavlja torzijsko bariero, n pa periodičnost, ki pove, kolikokrat se posamezni minimum oz. maksimum ponovi med rotacijo za 360° . ϕ_0 je referenčni torzijski kot, ki smo ga definirali že prej. Pogosto moramo za pravilni zapis **torzijskega potenciala** - odvisnosti energije od vrednosti torzijskega kota, le-tega zapisati kot vsoto več enostavnih trigonometričnih funkcij.

2. Nevezne interakcije med atomi opišemo z dvema členoma - Coulombovim potencialom in Lennard-Jonesovim potencialom (Slika 7).

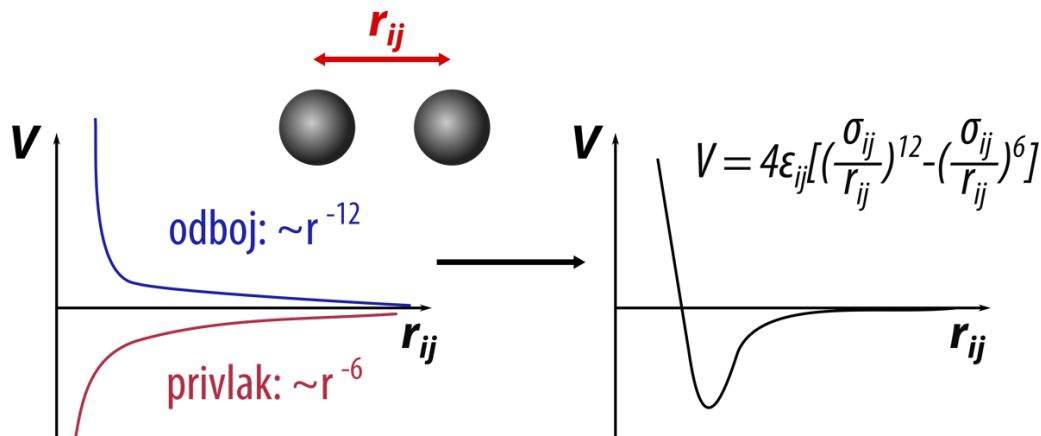


Slika 7. Prikaz elektrostatske interakcije (Coulombov potencial) med dvema točkastima naboje (q_i in q_j) in pripadajoči energijski diagram, ki prikazuje odvisnost elektrostatske energije od razdalje med njima (levo). Prikaz van der Waalsovih interakcij med dvema atomoma in pripadajoči energijski diagram, Lennard-Jonesov potencial (desno).

Prvi člen, **Coulombov potencial**, služi za opis elektrostatskih interakcij, kjer sta q_i in q_j vrednosti **delnega (parcialnega) naboja** atomov, r_{ij} pa je razdalja med atomoma. Delni naboj q je naboj, ki je po absolutni vrednosti manjši od osnovne enote naboja elektrona ($e_0 = 1.602 \times 10^{-19}$ As), in opiše asimetrično porazdelitev elektronske gostote v kemijskih vezeh. Seštevek vseh delnih nabojev predstavlja celokupni naboj molekule, ki je lahko nabita (kation ali anion) ali pa nevtralna.

Pozitivni delni naboj pomeni nižjo elektronsko gostoto, negativni pa večjo na izbranem atomu. Atomski naboji so ključni element polja sil. Žal niso kvantne observable, tako da obstaja več arbitrarnih postopkov, kako jih dobiti iz elektronske gostote. Grafično jih predstavimo tako, da izbrano molekulsko površino pobarvamo glede na vrednost delnega naboja atoma, ki se nahaja pod površino posameznega atoma. Delne naboje atomov s presežkom elektronske gostote (negativna vrednost delnega naboja) pobarvamo rdeče, tiste s primanjkljajem (pozitivna vrednost delnega naboja) pa modro. Čim višja je vrednost (v pozitivno ali negativno smer), tem intenzivnejša je barva.

Z **Lennard-Jonesovim potencialom**, drugim členom neveznih interakcij v empirični potencialni funkciji, poenostavljeno opišemo vpliv napolarnih **van der Waalsovih interakcij** (Slika 8). Van der Waalsove interakcije predstavimo kot vsoto odbojnih in privlačnih členov. Vrednost r_{ij} je razdalja med atomoma, ki vstopata v interakcijo, A_{ij} in B_{ij} sta parametra, ki ju določimo s pomočjo znanih vrednosti za posamezne atome. Člen r^{-12} opisuje **kratkosežne (Paulijeve) odbojne sile**, ki nastanejo zaradi odboja pri kontaktu elektronskih orbital, r^{-6} člen pa opisuje **privlačne daljnosežne interakcije**: van der Waalsove ali disperzijske sile.

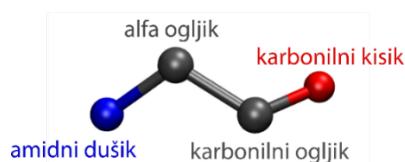


Slika 8. Potek privlačnega in odbojnega člena, ki skupaj sestavljata Lennard-Jonesov potencial.

Za natančen opis molekule z molekulsko mehaniko potrebujemo kvalitetne **parametre empirične potencialne funkcije** torej ravnotežne konstante, ravnotežne razdalje, delne naboje in van der Waalsove parametre, ki opisujejo obnašanje vsakega posameznega atoma. Ker ima isti atom v molekuli lahko različne lastnosti - npr.aminski dušik, amidni dušik, kvarterni dušik, mora imeti vsak tak atom tudi različen niz parametrov. Zato uvedemo pojem **atomskega tipa** in teh je seveda precej več kot atomov. Od kakovosti uporabljenih parametrov je odvisna tudi kakovost rezultatov, ki jih dobimo z uporabo polja sil. Vsak atomski tip ima tako svoj niz parametrov za uporabo v polju sil in postopek pridobivanja le-teh imenujemo **parametrizacija**. Za lažjo predstavbo si pogledjmo v Preglednici 2 nekaj vrednosti veznih in kotnih parametrov polja

sil atomov, ki sestavljajo ogrodne atome vsake aminokisljine. Ko si boste ogledali preglednico, premislite tudi, za katere geometrijske spremembe amidne vezi je potreben večji energijski vložek.

Preglednica 2. Vrednosti parametrov empirične potencialne funkcije (ravnotežne konstante, ravnotežne razdalje), ki ga uporabljamo pri računski obravnavi atomov, ki sestavljajo glavno proteinsko verigo.



Parametri člena MM, ki opisuje deformacije valenčne vezi	k_{vezi} [kcal/mol Å ²]	$l_{i,0}$ [Å]
<i>karbonilni kisik (O) - karbonilni ogljik (CO)</i>	570.0	1.229
<i>karbonilni ogljik (CO) - alfa ogljik (CA)</i>	317.0	1.522
<i>alfa ogljik (CA) - amidni dušik (N)</i>	337.0	1.449
Parametri člena MM, ki opisuje deformacije valenčnega kota	k_{kota} [kcal/mol Å ²]	$q_{i,0}$ [°]
<i>karbonilni kisik (O) - karbonilni ogljik (CO) - alfa ogljik (CA)</i>	80.0	120.40
<i>karbonilni ogljik (CO) - alfa ogljik (CA) - amidni dušik (N)</i>	63.0	110.10

Kot zanimivost omenimo, da poleg opisanih parametrov in energijskih členov, ki so del praktično vseh empiričnih potencialnih funkcij, so v nekaterih poljih sil vgrajeni še dodatni členi, ki omogočajo natančnejše opisovanje atomske strukture. Tako osnovnim členom MM dodajo energijski člen, ki skrbi za **preprečevanje nepravilnih torzij** (ang. *improper torsions*), npr. za lažje ohranjanje planarnosti aromatičnih elementov v molekuli. **Sklopljeni členi** (ang. *cross terms*) pa sočasno obravnavajo več gibanj v molekuli, kot je na primer sprememba valenčnega kota in ene od kovalentnih vezi tega kota skupaj.

Izračunana potencialna energija molekule ima lahko celokupen pozitiven ali negativen predznak, saj v empirični potencialni funkciji MM nastopajo členi obeh predznakov. Člena za deformacijo vezi in valenčnega kota, prikazana na Sliki 6, sta lahko pozitivna, v idealnem primeru pa 0, ko je vrednost parametra enaka ravnotežni vrednosti. Vrednost torzijskega potenciala je navadno pozitivna. Pri členih na Sliki 7, ki opisujejo prispevek neveznih interakcij, pa celokupna vsota tako privlačnih kot odbojnih prispevkov med atomi določi končni predznak energije molekule.

Vrednosti potencialne energije, pridobljene z molekulsko mehaniko, lahko direktno primerjamo le med konformacijami iste molekule. Velja pravilo, da nižja vrednost dobljene potencialne energije sovpada z energijsko ugodnejšo konformacijo. Pri primerjavi izračunanih energij med različnimi molekulami je potrebno biti previdnejši, saj izračunane vrednosti niso direktno primerljive in iz razlike med njimi še ne moremo takoj sklepati o stabilnosti struktur. V teh primerih analiziramo tudi vrednosti energije posameznih členov potencialne funkcije. Poudariti moramo tudi, da dobljene energije MM ne predstavljajo termodinamskih količin, kot

je prosta energija ali entalpija, ampak služijo kot prva ocena energije stabilnosti posamezne konformacije molekule.

V mnogih dostopnih programskih paketih so vgrajena različna polja sil MM. Omenimo nekaj programov, ki se uporabljajo za modeliranje tako malih molekul kot tudi bioloških makromolekul:

AMBER - *Assisted Model Building with Energy Refinement*

CHARMM - *Chemistry at HARvard Macromolecular Mechanics*

GROMOS - *GROningen MOlecular Simulation package*

OPLS - *Optimized Potential for Liquid Simulations*

3.2. KVANTNA KEMIJA

Na kratko se pomudimo še pri bolj natančnem opisu atomske strukture. Poleg **molekulske mehanike**, lahko energijo molekul izračunamo tudi drugače. To je predvsem pomembno kadar želimo preučevati lastnosti, ki jih z molekulsko mehaniko ne moremo zadovoljivo opisati. Taki so na primer tvorbe in razcepi kovalentnih vezi, ki se zgodijo pri kemijski reakciji. Tu iz empiričnega opisa preidemo v natančnejši fizikalni opis atomske strukture.

Fizikalno osnovo za računanje energetike molekulske strukture podaja **Schrödingerjeva enačba**, ki obravnava molekule kot zbirko jeder in elektronov brez predhodnega poznavanja kemijskih vezi. Predstavlja eno temeljnih enačb **kvantne mehanike** (ang. *quantum mechanics - QM*) in je nekakšen analog II. Newtonovega zakona iz klasične mehanike. Schrödingerjeva enačba je analitično rešljiva le v zelo redkih primerih, kot je vodikov atom, kjer so rešitve poznane tudi kot atomske orbitale. Stacionarno obliko Schrödingerjeve enačbe zapišemo kot:

$$\hat{H} \cdot \vec{\Psi}(x, y, z) = E \cdot \vec{\Psi}(x, y, z)$$

\hat{H} je Hamiltonov operator energije in je vsota kinetične in potencialne energije molekule, Ψ je valovna funkcija molekule, E pa je energija obravnavanega molekularnega sistema.

Kadar računamo energijo molekule z reševanjem Schrödingerjeve enačbe, izvajamo **ab initio** (lat. iz prvih principov) **račune**. Zato je na voljo več razvitih programskih paketov, npr. *Gaussian*, *GAMESS*, *Q-Chem*, *Spartan*, *ArgusLab* in drugi. Pri QM računih ne uporabljamo eksperimentalnih podatkov in parametrov, pomemben je le nabor atomov, ki definira število protonov in elektronov molekule. Čas QM računa izjemno hitro raste z večanjem števila obravnavanih delcev. Približno se računski čas poveča za desetkrat, če podvojimo število atomov v obravnavanem molekularnem sistemu, čeprav je dejanski čas odvisen še od mnogih faktorjev, tako od dostopne računalniške tehnologije kot uporabljene QM računske metode. QM računi so zato zahtevni in dolgotrajni že za manjše molekule. Trenutno v sprejemljivih računskih časih s QM računi obravnavamo molekule z nekaj 100 atomi.

Za reševanje Schrödingerjeve enačbe uporabljamo več numeričnih metod. Pogosto uporabljamo tudi izraz modelna kemija (ang. *model chemistry*). Eno izmed teh metod predstavljajo **Hartree-Fockove (HF) enačbe**. V zadnjih letih se je uveljavila tudi **metoda gostotnih funkcionalov** (ang. *Density Functional Theory - DFT*), ki pri računih uporablja

elektronsko gostoto in s tem bolje obravnava medsebojni vpliv elektronov t. i. medelektronske korelacije, kar Hartree-Fockove enačbe ne upoštevajo.

Pri QM računih je pomemben izbor matematičnih funkcij, ki jih uporabimo kot začetni nastavek za konstruiranje valovne funkcije molekule ψ . Ta matematični nastavek je **bazni niz** (ang. *basis set*). **3-21G, 6-31G, 6-31G(d)** so oznake nekaterih pogosteje uporabljanih baznih niz, ki temeljijo na uporabi poznanih Gaussovih funkcij.

Pri QM računih velikokrat poslužujemo **semiempiričnih metod**, ki za hitrejše reševanje Schrödingerjeve enačbe uvajajo različne približke. Obstaja več semiempiričnih metod in **MNDO, AM1, CNDO, EHT** ter **PM3** so le nekatere izmed njih. Semiempirične metode nekatere zahtevnejše računске korake (npr. računanje večatomskih integralov) ocenijo na osnovi predhodno znanih eksperimentalnih podatkov. V okviru vaj boste izvedli tudi kvantnomehanski račun, ko boste primerjali energije dveh različnih konformacij izbrane molekule (Vaja 5).

Molekulsko modeliranje bioloških molekul, kot so encimi, receptorji in nukleinske kisline z več tisoč atomi, je z uporabo izključno QM metod trenutno še neizvedljivo. Po drugi strani pa modeliranje encimskih mehanizmov, kjer pride do cepitve in tvorbe kemijskih vezi, ni mogoče le z uporabo empiričnih polj sil MM, saj polja sil lahko uporabimo le za račune, kjer se konfiguracija sistema med računanjem ohranja.

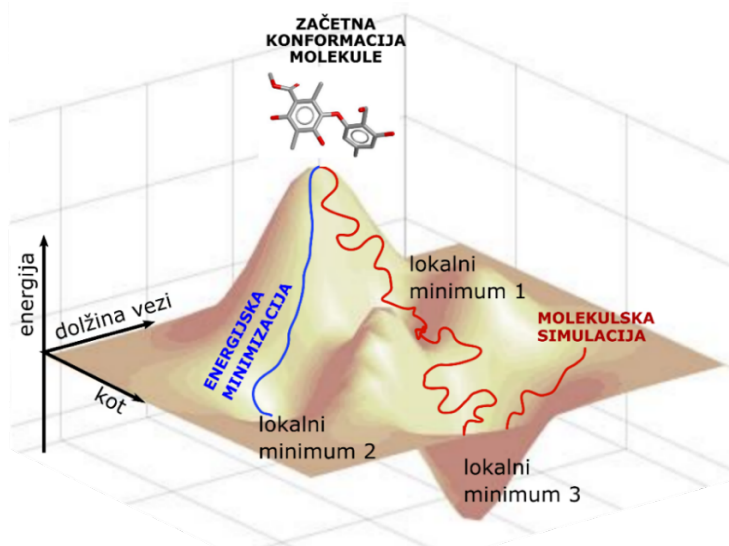
Rešitev je prišla z razvojem hibridnih (oz. mešanih ali multiskalnih) tehnik, ki združujejo klasični in kvantni opis atomov in molekul. Pri teh **večnivojskih (multiskalnih) kvantnomehansko-molekulskomehanskih (QM/MM) pristopih** makromolekulo razdelimo na dva dela. Manjši del, kjer se zgodijo najpomembnejše spremembe, ki so navadno razcepi in tvorbe kovalentnih vezi, obravnavamo s QM opisom in zanj rešujemo Schrödingerjevo enačbo. Preostali večji del pa opišemo s poljubno empirično potencialno funkcijo (MM). Energijo sistema izračunamo kot vsoto QM prispevka, MM prispevka ter t. i. mešanega energijskega prispevka, ki obravnava interakcijo med obema sistemoma.

4. RAČUNSKO PREISKOVANJE POVRŠINE POTENCIALNE ENERGIJE MOLEKUL

Začetni tridimenzionalni model molekule, zgrajen s pomočjo grafičnega programa ali pa določen z eksperimentalnimi metodami, se pogosto nahaja v neugodni konformaciji, ki ima visoko vrednost potencialne energije in ne predstavlja eksperimentalno relevantne konformacije molekule. Večdimenzionalno površino, ki ponazarja energijo molekule kot funkcijo vseh možnih razporeditev – tj. koordinat njenih atomov, imenujemo **površina potencialne energije** (ang. *potential energy surface* - PES).

Na PES želimo najprej identificirati **stacionarne točke**, tako **minimume (lokalne in globalnega)**, kjer je molekula v najbolj ugodnih energijskih legah, kot tudi **sedla**, v primeru iskanja geometrije prehodnega stanja med reakcijo. Včasih želimo ovrednotiti konformacijski prostor, ki je molekuli na voljo pri danih pogojih, na primer pri določeni temperaturi ali tlaku. Ključni računski metodi, ki ju uporabljamo v ta namen in ju bomo spoznali v tem odseku, sta **geometrijska optimizacija** in **molekulska dinamika**. Shematsko sta obe metodi prikazani na Sliki 7. Na koncu razdelka si bomo na kratko pogledali še **konformacijsko analizo**, kjer bomo študirali vpliv

spremembe vrednosti torzijskega kota (rotacijo okrog izbrane kovalentne vezi) na energijo molekule.



Slika 7. Shematski prikaz poenostavljene površine potencialne energije (PES) in dveh računskih metod študija konformacijskega prostora: geometrijske optimizacije (modra črta) in simulacije molekulske dinamike (rdeča črta) iz začetne konformacije na dvodimenzionalni ploskvi potencialne energije.

4.1. ENERGIJSKA MINIMIZACIJA - GEOMETRIJSKA OPTIMIZACIJA

Geometrijska optimizacija (ang. *geometry optimization*) je računski postopek iskanja konformacij molekul v energijskih minimumih, v določenih pogojih pa tudi v sedlih, na površini potencialne energije molekule. Če na PES iščemo minime, postopek imenujemo tudi **energijska minimizacija**. V literaturi se pogosto oba postopka obravnavata kot sinonim.

Navadno geometrijsko optimizacijo poženemo iz začetne kemijsko smiselne konformacije molekule \mathbf{x}_k . Nato z uporabo numeričnega **optimizacijskega algoritma** v več korakih (oz. iteracijah) poiščemo tako konformacijo molekule, ki ima v lokalni okolici najnižjo vrednost potencialne energije - lokalni minimum.

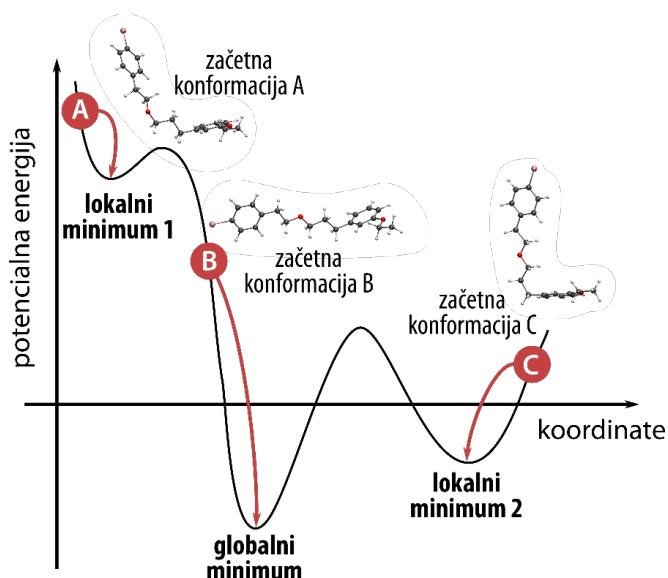
Optimizacijski algoritmi so numerične tehnike iskanja minimumov na PES. Potencialno funkcijo $V(\mathbf{x})$ navadno razvijemo v Taylorjevo vrsto okrog izhodne konfiguracije \mathbf{x}_k , kot to kaže spodnja enačba. Prvi odvod potencialne funkcije $V'(\mathbf{x}_k)$ glede na odvajano koordinato imenujemo tudi **gradient**. $3N \times 3N$ matrika drugih odvodov $V''(\mathbf{x}_k)$ pa je **Hesse-jeva matrika**.

$$V(\mathbf{x}) = V(\mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)V'(\mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)^2V''(\mathbf{x}_k) + \dots$$

Optimizacijski algoritem kot kriterij iskanja stacionarne točke uporablja **uporablja prvi odvod** (oz. gradient), ki se mu včasih pridruži tudi **drugi odvod potencialne energijske funkcije** (oz. Hessian). V minimumu velja, da so prvi odvodi po vseh koordinatah enaki nič, drugi odvod pa večji od 0.

V skupino metod, ki temeljijo na prvem odvodu, uvrščamo metode **najhitrejšega spusta** (ang. *steepest descent*) in **konjugiranih gradientov** (ang. *conjugate gradients*), v drugi skupini pa so najpomembnejše **Newton-Raphsonove metode**.

Kompleksnost večdimenzionalne površine potencialne energije molekule je vzrok, da izhodiščna konformacija molekule vpliva na končni rezultat geometrijske optimizacije. Optimizacijski algoritmi uspejo najti le stacionarno točko, ki je blizu začetni strukturi. Pri tem niso sposobni prečkati področij z višjo energijo, lokalnih energijskih maksimumov, in doseči morebitnih konformacij na PES s še nižjo potencialno energijo. Zato je še pred začetkom geometrijske optimizacije potrebno zelo premišljeno zgraditi začetno stanje in uporabljati dostopne eksperimentalne podatke. Če želimo ovrednotiti, kako blizu globalnega minimuma smo, moramo narediti več energijskih minimizacij, izhajajoč iz različnih začetnih konformacij sistema, kot to prikazuje Slika 8. Pri računanju z biološkimi makromolekulami, kjer imamo podatke o eksperimentalno določeni 3D strukturi, navadno privzamemo, da je struktura že v bližini energijsko najugodnejše in najbolj reprezentativne konformacije.



Slika 8. Shematska predstavitev geometrijske optimizacije na 1D ploskvi potencialne energije. Iz različnih začetnih konformacij molekule optimizacijski algoritem poišče različne stacionarne točke, ki so lokalni minimumi ali pa globalni minimum.

4.2. MOLEKULSKA DINAMIKA

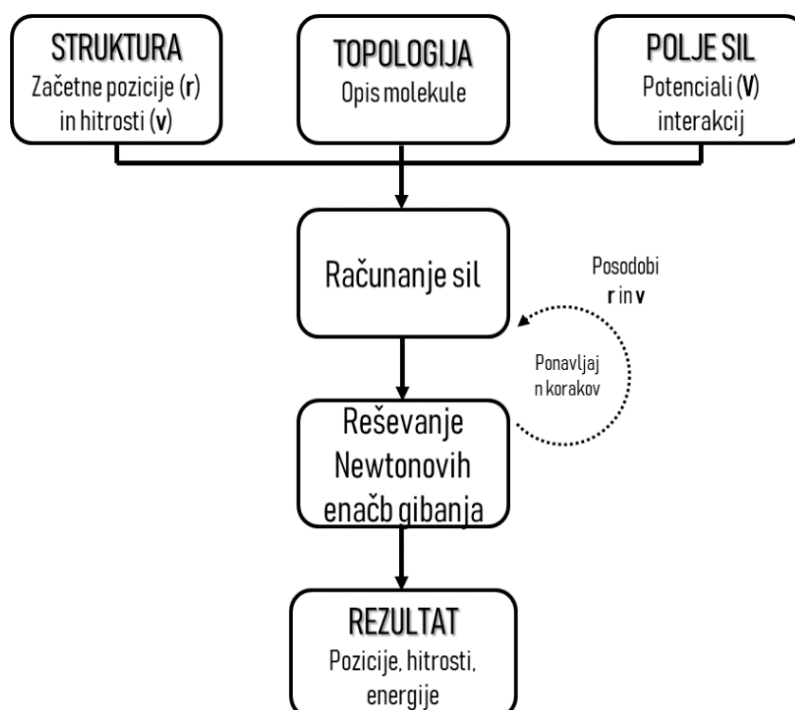
Geometrijska optimizacija prikaže statično idealizirano sliko molekularnega sistema v eni izmed njegovih energijsko najugodnejših konformacij pri absolutni temperaturi 0 K. Samo poznavanje energijsko ugodnih konformacij ne omogoča preučevanja dinamičnega vidika obnašanja molekul ter računanja makroskopskih lastnosti, kot so na primer termodinamske količine. V ta namen uporabljamo metode **molekulskih simulacij** (ang. *molecular simulations*), ki omogočajo preiskovanje površine potencialne energije in generiranje konformacijskega prostora, ki je pri danih pogojih simulacije dostopen posameznemu molekularnemu sistemu (Slika 7).

Tako lahko iz molekulskih simulacij izračunamo makroskopsko merljive količine in primerjamo rezultate simulacij z eksperimentalnimi podatki. Teoretično osnovo za to povezavo med strukturo molekul in njihovimi makroskopskimi lastnostmi omogoča **statistična mehanika**. Najpogosteje uporabljani metodi molekulskih simulacij sta **molekulska dinamika** (ang. *molecular dynamics*) (**MD**) in metoda **Monte Carlo**. Tu bomo podrobneje spoznali molekulska dinamiko.

Pri metodi molekulske dinamike nove konformacije pridobimo z integriranjem klasičnih *Newtonovih* enačb gibanja. Na nivoju atomov in molekul tako rešujemo II. Newtonov zakon, ki ga za i -ti delec (atom) zapišemo kot:

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i(t)}{m_i} = - \frac{1}{m_i} \frac{\partial V(r(t))}{\partial r_i(t)}$$

pri tem velja, da je sila F_i na i -ti delec (atom) z maso m_i in a_i pospešek delca i . F_i na i -ti delec je enaka negativnemu gradientu potencialno energijske funkcije, ki je enaka produktu mase i -tega delca in njegovega pospeška. Pomembnejše korake MD metode predstavlja Slika 9.



Slika 9. Shematski prikaz korakov metode molekulske dinamike (MD).

Pri MD simulaciji energijo konformacije izbrane molekule izračunamo z **empirično potencialno funkcijo MM**, ki smo jo spoznali v prejšnjem poglavju. Za vse atome, ki sestavljajo simulirani sistem, moramo poznati tudi njihove atomske tipe in definirati parametre empirične potencialne funkcije. To imenujemo **topologija** molekule.

Ko imamo vse omenjeno pripravljeno, izračunamo **rezultante sil** na posamezne atome. Po posebnem postopku sistemu na začetku simulacije pripišemo začetne hitrosti v_{i0} , ki tudi definirajo še temperaturo sistema. Nato za kratek časovni interval t imenovan **integracijski korak**, pustimo, da se vsak atom giblje v smeri dobljene rezultante sil, kot to opisuje II. Newtonov zakon. Izbrani časovni interval gibanja mora biti krajši od najhitrejših gibanj v molekuli in je navadno 1 ali 2 fs ($1 \text{ fs} = 10^{-15} \text{ s}$). Z uporabo **integracijskega algoritma** sočasno tudi rešimo Newtonovo gibalno enačbo. Po tem času določimo nove koordinate in izračunamo nove sile na atome ter postopek ponovimo za poljubno število korakov, ki določijo celotni čas

simulacije. Ta je navadno med 10-100 ns, lahko pa simuliramo sistem tudi več μs . Rezultat simulacije MD je časovno zaporedje različnih konformacij sistema - **trajektorija**.

Za boljšo predstavo MD metode si pogledjmo enostaven primer reševanja II. Newtonovega zakona za delec i z maso m_i , začetno lego x_{i0} ter začetno hitrostjo v_{i0} v eni dimenziji, kot to že poznamo iz fizike. Najprej II. Newtonov zakon zapišemo kot:

$$F_i = m_i \cdot a_i = m_i \cdot \frac{dv}{dt} = m_i \frac{d^2x}{dt^2}$$

Če predpostavimo, da je pospešek a_i za delec i v času t konstanten, lahko nove koordinate delca x_i dobimo z enačbo:

$$x_i = x_i^0 + v_i^0 t + \frac{at^2}{2}$$

Pospešek a_i na delec i izračunamo iz odvoda energijske potencialne funkcije V v začetni koordinati x_0 :

$$a_i = -\frac{1}{m_i} \frac{dV}{dx_0}$$

Pri večatomskih molekularnih sistemih, ki se premikajo v tridimenzionalnem prostoru, ne moremo uporabljati analitičnih rešitev, kot so prikazane zgoraj za enostavno enodimenzionalno gibanje, ampak uporabljamo različne *numerične metode*, ki smo jih že omenili, imenovane tudi **integracijski algoritmi**.

Ena izmed takih numeričnih tehnik je **metoda končnih razlik** (ang. *finite difference methods*), kjer izvedemo integracijo v seriji kratkih časovnih korakov z dolžino Δt . Ta premisa je osnova za množico integracijskih algoritmov, med drugimi tudi za **Verletov algoritem**, ki predpostavlja, da se položaj r delca i v časovnih intervalih $t+\Delta t$ in $t-\Delta t$ da zapisati v obliki Taylorjeve vrste:

$$r_i(t + \Delta t) = r_i(t) + \frac{dr_i}{dt} \Delta t + \frac{d^2r_i}{dt^2} \frac{\Delta t^2}{2!} + \frac{d^3r_i}{dt^3} \frac{\Delta t^3}{3!} + \dots$$

$$r_i(t - \Delta t) = r_i(t) - \frac{dr_i}{dt} \Delta t + \frac{d^2r_i}{dt^2} \frac{\Delta t^2}{2!} - \frac{d^3r_i}{dt^3} \frac{\Delta t^3}{3!} + \dots$$

Kombinacija zgornjih enačb da končno obliko Verletovega algoritma za izračun nove pozicije delca:

$$r_i(t + h) = 2r_i(t) - r_i(t - h) + \frac{d^2r_i}{dt^2} \frac{\Delta t^2}{2!} + O(\Delta t^4)$$

Ker se v Verletovem integracijskem algoritmu hitrosti ne pojavijo eksplicitno, jih lahko izračunamo na več različnih načinov, med drugim tudi iz pozicij v času $t+\Delta t$ in $t-\Delta t$:

$$\frac{dr_i}{dt} = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} + O(\Delta t^3)$$

V literaturi najdemo še mnoge integracijske algoritme, kot so **Leap-frog algoritem**, **velocity Verlet algoritem** ter mnogi drugi. Z vsemi opisanimi algoritmi ohranimo celotno energijo sistema konstantno. V primeru, da želimo izvajati simulacijo molekulske dinamike še pri konstantni temperaturi, uporabimo ustrezen termostat oz. barostat v primeru konstantnega tlaka (najpogosteje uporabljen je Berendsenov termostat oz. barostat).

Trajektorijo, pridobljeno z metodo molekulske dinamike za izbran (bio)molekularni sistem (npr. protein ali nukleinsko kislino), nato analiziramo. Tako izvemo več o obnašanju sistema, ki ga študiramo. Poglejmo si dva geometrijska parametra za analizo MD trajektorij, ki sta uporabna za začetno evalvacijo simulacij bioloških makromolekul.

1. RMSD parameter

RMSD parameter smo spoznali že pri analizi statičnih konformacij molekul v drugem razdelku tega poglavja in predstavlja standardno deviacijo atomskih razdalj med ekvivalentnimi atomi glede na referenčno strukturo - ta je pogosto prva konformacija simuliranega atomskega sistema v MD. Definiran je tako kot prej:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i,j=1}^N d_{i,j}^2}$$

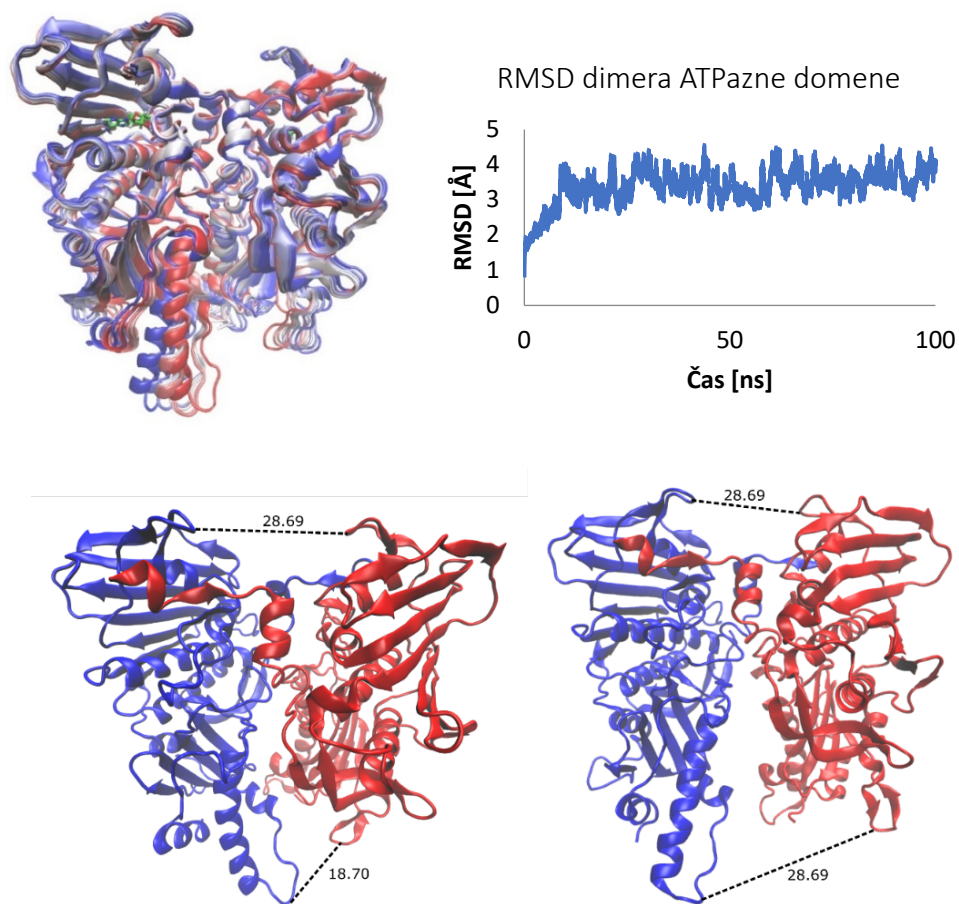
kjer je $d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$ razdalja med atomoma i in j , s koordinatami x , y , z in N število primerjanih razdalj.

Primer

Izračun RMSD parametra pri analizi molekulske dinamike ATPazne domene človeške topoizomeraze II α

Da bi boljše razumeli obnašanje proteina ATPazne domene človeške topoizomeraze II α , smo izvedli 100 ns dolgo simulacijo molekulske dinamike dostopnega dimera kristalne strukture te domene v kompleksu z molekulo AMP-PNP. Glede na izhodiščno geometrijo je izračunana RMSD vrednost znašala 3.39 ± 0.50 Å, kar nakazuje spremembo konformacije glede na začetno strukturo. Časovni potek RMSD parametra, spet relativno glede na začetno kristalno strukturo, kot tudi nekaj reprezentativnih konformacij si lahko ogledate na naslednji sliki.

Ko vizualiziramo rezultate, ugotovimo, da je začetna razdalja med izbranimi atomoma na zgornjih zankah 28.69 Å. Po 100 ns simulacije se le-ta zmanjša na 23.42 Å. Po drugi strani se razdalja med atomoma na spodnjih zankah z 18.70 Å poveča na 57.63 Å. Glede na funkcijo ATPazne domene, ki veže eno izmed dveh DNA molekul, dobljen rezultat najverjetneje kaže, da s to fleksibilnostjo ATPazna domena olajša vezavo DNA.



2. RMSF parameter

Ko študiramo konformacijski prostor proteinov z molekulsko dinamiko, nas pogosto zanima tudi njihova interna dinamika. S tem imamo v mislih lokalno premikanje proteinskih zank, vrtenje domen in druga gibanja. Le-ta so lahko ključna za podrobno razumevanje funkcije bioloških molekul. V takšnih primerih poleg RMSD parametra izračunamo še **RMSF parameter** (ang. *Root-Mean-Square Fluctuations*).

Z RMSF parametrom lahko študiramo lokalno premikanje posameznega atoma ali pa večje enote, kot je aminokislina proteina, v izvedenem simulacijskem času. Z računanjem teh lokalnih sprememb v proteinu tako ocenimo fleksibilnost njegovih delov. RMSF parameter je definiran kot:

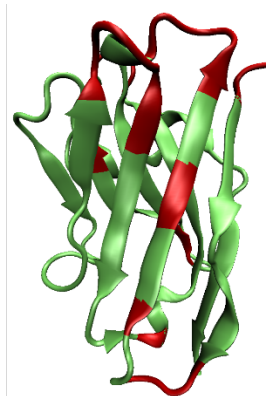
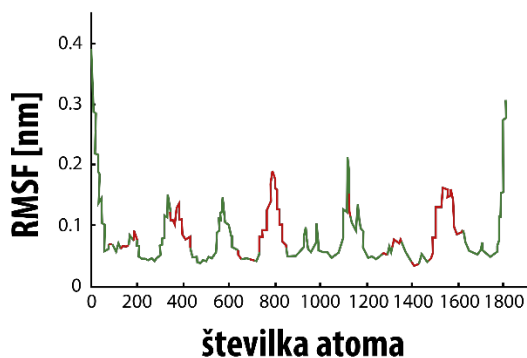
$$RMSF(a) = \sqrt{\frac{1}{T} \sum_{t=0}^T (a_t - \bar{a})^2}$$

kjer je T čas MD simulacije, na katerem je RMSF izračunan, t izbrani časovni korak, a_t pozicija atoma (ali aminokislina). $(a_t - \bar{a})$ pa predstavlja odmik atoma (oz. izbrane enote) a v času t , in sicer glede na njeno povprečno pozicijo \bar{a} .

Primer

Izračun RMSF parametra pri analizi molekulske dinamike nanotelesa

S pomočjo RMSF parametra lahko določimo interno fleksibilnost nanotelesa. To so posebna enodomenska protitelesa, ki kot del imunskega odziva lahko prepoznajo različne antigene. Spodaj si lahko ogledate dobljeni RMSF graf po izvedeni simulaciji molekulske dinamike. Na 3D strukturi nanotelesa so z rdečo pobarvani bolj fleksibilni deli proteinske strukture, kjer so vrednosti RMSF parametra višje, z zeleno pa bolj rigidni deli simuliranega nanotelesa. Predvidevamo, da so bolj fleksibilni deli tisti, ki omogočajo prepoznavo in vezavo antigenov.

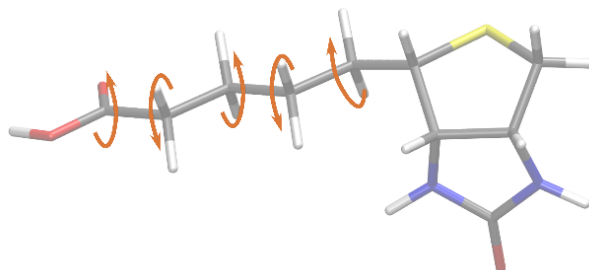


4.3. ANALIZA LOKALNEGA KONFORMACIJSKEGA PROSTORA MOLEKULE - TORZIJSKI POTENCIAL

Poglejmo si še računsko metodo, ki jo uporabljamo za boljše razumevanje 3D strukture, predvsem manjših organskih molekul. Molekule, ki se med seboj razlikujejo le po 3D razporeditvi atomov, so **konformacijski izomeri** ali **konformeri**. Okrog vsake enojne sp^3 vezi, ki je σ -vez, je namreč možna rotacija in take konformere imenujemo **rotameri**. Konformeri lahko, če ni prisotnih steričnih omejitev, že pri sobni temperaturi prosto prehajajo eden v drugega, saj so energijske bariere med njimi velikosti le nekaj kcal/mol.

Primer

Molekula biotina ima zaradi možnih rotacij okrog označenih sp^3 hibridiziranih atomov, ki sestavljajo njegovo verigo, σ -C-C vezi veliko možnih konformacij.



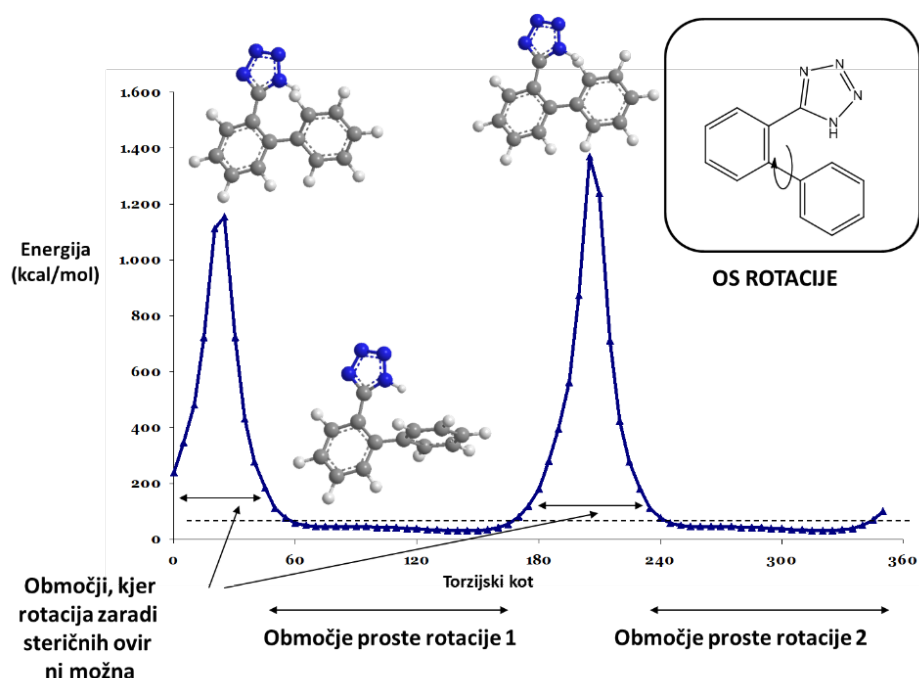
Poznavanje konformacijskega prostora je včasih zelo pomembno za razumevanje obnašanja biološko aktivnih molekul in njenega medmolekulskega prepoznavanja z biološko tarčo. Analizo energijskih sprememb in obnašanja konformerov ter iskanje lokalnih minimumov molekul in preostalih možnih konformacij imenujemo tudi **konformacijska analiza**. To si bomo ogledali na posebnem primeru take analize z uporabo **torzijskega potenciala**.

Primer

Torzijski potencial sterično ovirane C-C vezi organske molekule

Poglejmo si primer konformacijske analize okrog označene C-C vezi, ki povezuje oba fenilna obroča spodaj prikazane geometrijsko optimizirane konformacije organske molekule. Po rotaciji okrog izbrane C-C vezi narišemo graf energije posameznih konformacij pri različnih vrednostih torzijskega kota, ki ga poleg obeh ogljikovih atomov tvorita še oba sosednja ogljika. Dobimo torzijski potencial okrog izbranega torzijskega kota. Iz grafa opazimo, da konformacije, kjer je vrednost torzijskega kota med 0° in 60° ter med 170° in 240° , tej molekuli niso dostopne, saj imajo le-te visoko vrednost potencialne energije. Za ostale vrednosti torzijskega kota pa je rotacija okrog izbrane C-C vezi prosta.

Med rotacijo okrog torzijskega kota lahko z merjenjem razdalj med poljubnima atomoma tudi ugotovimo, kako ugodna je energija v posamezni poziciji. Ta informacija je pomembna v primerih, ko je določena razdalja med funkcionalnima skupinama v molekuli ključna za uspešno medmolekulsko prepoznavo in biološko delovanje, in tako lahko ocenimo, kako ugodna je ta lega za preiskovano molekulo.

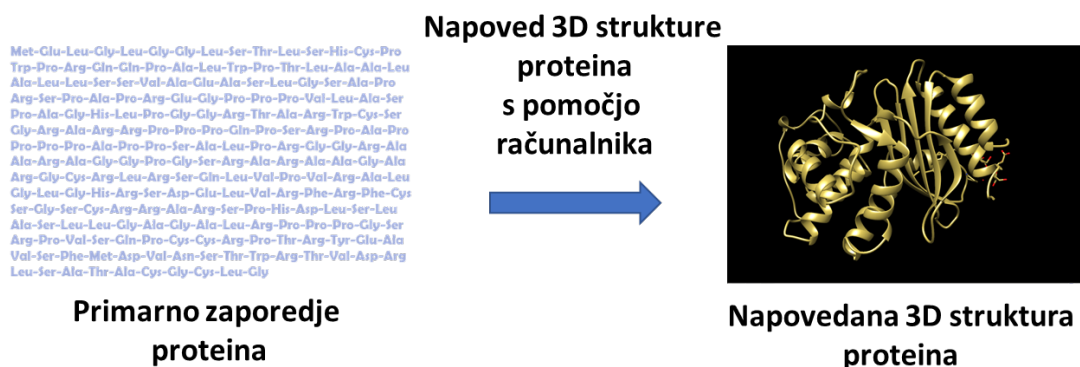


Andrej Perdih

napovedovanje 3d strukture proteinov z računskimi metodami

```
function quicksort(array)
var list less, equal,
if length(array) ≤ 1
return array
select a pivot value piv
for each x in array
if x < pivot then less
if x = pivot then equ
if x > pivot then gre
return concatenate(quicksort
```

Velikokrat omejitve dostopnih eksperimentalnih metod proteinske kristalografije in NMR spektroskopije ne omogočajo, da bi enostavno določili tridimenzionalno strukturo izbranega proteina. Takrat se lahko poslužimo *računskih in teoretičnih metod, s pomočjo katerih lahko s poznavanjem primarne strukture (tj. zaporedja aminokislin) napovemo sekundarno in 3D strukturo tarčnega proteina*, kot to prikazuje Slika 1.



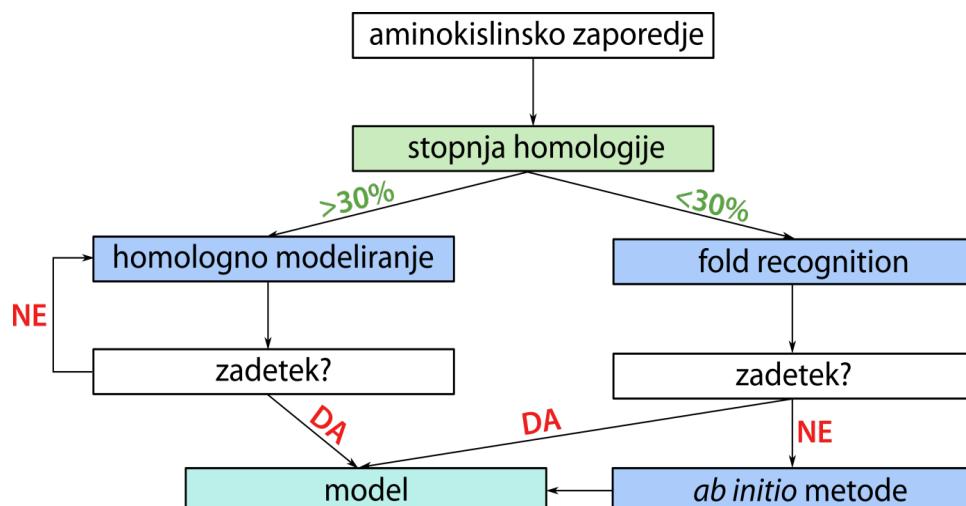
Slika 1. Osnovna ideja napovedovanja tridimenzionalnih struktur proteinov. Na osnovi poznavanja primarnega zaporedja z različnimi teoretičnimi pristopi napovemo 3D strukturo proteina.

Napovedovanje tridimenzionalne strukture proteinov je pomemben del raziskav v bioinformatiki in računski biokemiji, saj so rezultati, napovedane 3D strukture proteinov, uporabni v biomedicinskih raziskavah (npr. pri načrtovanju zdravilnih učinkovin) ter v biotehnologiji (npr. pri načrtovanju novih učinkovitejših encimov za industrijsko proizvodnjo). Pristopi napovedovanja 3D struktur proteinov so pomembni tudi zato, ker obstaja precejšnja vrzel med eksperimentalno določenimi strukturami in znanimi protenskimi primarnimi zaporedji. Le okoli 1 % znanih primarnih zaporedij proteinov ima že eksperimentalno določeno tudi tridimenzionalno strukturo.

Metode napovedovanja struktur proteinov so navkljub kompleksnim pristopom in intenzivnemu razvoju področja še vedno precej omejene in trenutno nobena metoda ne ponudi popolne rešitve in v vseh okoliščinah omogoča zanesljivo določitev 3D strukture tarčne biomolekule. Prav tako obstajajo določeni razredi proteinov, na primer transmembranski proteini, ki so pomembni pri mnogih bioloških procesih. Izpostavimo lahko družino receptorjev sklopljenih z G-proteinom. Določitev 3D strukture teh pomembnih transmembranskih proteinov je še vedno precejšen izziv.

V principu bi morala metoda napovedovanja 3D strukture raziskati celoten konformacijski prostor vseh možnih struktur proteina in jih rangirati po energijski stabilnosti, da bi izbrala tisto konformacijo, ki najbolje predstavlja strukturo. Ta naj bi predstavljala globalni minimum obravnavanega sistema. Žal je konformacijski prostor izredno velik in ga tudi z najmodernejšimi superračunalniki ne moremo raziskati niti za organske molekule ali manjše proteine. Dodaten problem prestavlja tudi ustrezno računanje energije posamezne napovedane konformacije in s tem učinkovito rangiranje. Pomembno vlogo pri določitvi končne biomolekule ima tudi topilo in učinki solvatacije na končno konformacijo.

Spoznali bomo tri skupine metod napovedovanja proteinske strukture, ki so predstavljene na Sliki 2. Za uporabo vsake izmed teh metod moramo že poznati primarno strukturo proteina. V primeru, da obstaja že rešena 3D struktura sorodnega proteina, ki ima 30 ali več procentno stopnjo homologije, tj. ujemanje primeranih sekvenc, lahko za napoved uporabimo **homologno modeliranje** (ang. *homology modeling*), ki je najpogosteje uporabljena napovedna metoda. V primeru, da je homologija manjša kot 30 %, se lahko poslužimo dveh pristopov. V primeru, da obstaja 3D struktura proteina, ki ima vsaj 20 % homologijo in zelo podobno zvitje (tj. podobna oblika in orientacija elementov sekundarne strukture, pri čemer ne upoštevamo povezanosti med posameznimi strukturami), se lahko poslužimo metode **prepoznave zvitij** (ang. *protein threading*). Obema metodama je skupno, da uporabljata t. i. *predloge* oz. *šablone* (ang. *template*) pri napovedi 3D strukture proteina (ang. *template-based methods*). V primeru, da je homologija z znanimi proteini zelo nizka ali pa da nismo bili uspešni s prejšnjimi metodami, lahko **napovemo 3D strukturo proteina tudi iz prvih principov** (lat. *ab initio*). Pri tem pristopu ne uporabljamo eksperimentalnih predlog, kot je to značilno za prejšnja dva pristopa. V nadaljevanju si bomo podrobneje pogledali vsako izmed omenjenih napovednih metod.

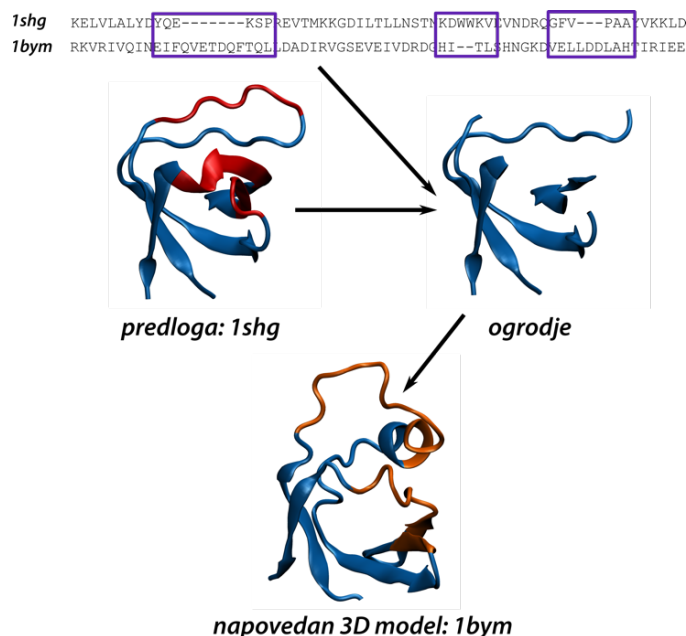


Slika 2. Razdelitev metod za napovedovanje 3D struktur proteinov.

1. HOMOLOGNO MODELIRANJE

Homologno modeliranje (ang. *homology modeling*) omogoča, da na osnovi poznavanja primarnega zaporedja proteina in poznane eksperimentalno določene strukture njemu sorodnega *homolognega* proteina, ki ga uporabimo za predlogo, zgradimo 3D model tarčnega proteina. *Homologno modeliranje torej temelji na predpostavki, da se terciarne strukture sorodnih proteinov/homologov ne bodo pomembno razlikovale navkljub razlikam v njihovem primarnem zaporedju.* 3D strukture proteinov med homologi so bolj ohranjene kot proteinske primarne sekvence, vendar mora biti stopnja homologije vsaj okoli 30 %, da lahko z dovolj visokim nivojem zaupanja uporabimo to metodo. Splošno shemo homolognega modeliranja prikazuje Slika 3.

andrej perdih, katja lakota, alja prah



Slika 3. Metoda homolognega modeliranja za napoved 3D strukture proteinov.

Koraki homolognega modeliranja:

1. Iskanje aminokislinskega zaporedja tarčnega proteina, za katerega želimo zgraditi model
2. Iskanje in identifikacija sorodnih struktur proteina (predlog oz. šablon)
3. Poravnava tarčnega primarnega zaporedja s primarnim zaporedjem predloge
4. Izdelava 3D modela proteina (dodajanje stranskih verig in proteinskih zank)
5. Optimizacija 3D modela proteina
6. Evaluacija končnega 3D modela proteina

Prvi korak homolognega modeliranja je poznavanje primarne strukture tarčnega proteina, za katerega želimo zgraditi 3D model. Nato moramo poiskati primerno predlogo (ang. *template*) z znano 3D strukturo, ki se čimbolj ujema z našim neznanim proteinom, na podlagi katere bomo lahko zgradili naš tarčni model in ga poravnali s tarčnim proteinom. Ta dva koraka se po navadi izvajata skupaj. PDB je spletna baza, v kateri so dostopne 3D strukture biomakromolekul, zato le-ta predstavlja najboljši vir za iskanje dobre predloge. Eno izmed najbolj popularnih programskih orodij, ki nam omogoča primerjavo primarnih aminokislinskih zaporedij s sočasnim iskanjem dostopnih 3D struktur po bazi PDB, je BLAST (*Basic Local Alignment Search Tool*). Spletna orodja in podatkovne baze, ki vsebujejo podatke o bioloških molekul boste natančneje spoznali tudi pri Vaji 1.

Ko identificiramo primerne 3D strukture homolognih proteinov, ki bi jih lahko uporabili kot predlogo, jih natančno ocenimo (tipično nam pri tem pomaga izbrana programska oprema). Pregledamo, koliko aminokislin in atomov manjka, kakšna je resolucija proteina, itd. Velikokrat je tudi pomembno, ali ima protein vezane ligande oz. kofaktorje ali pa je v t. i. prosti apo obliki, saj vezava ligandov včasih bistveno vpliva na 3D konformacijo. Priporočljivo je, da si 3D strukturo izbranega proteina, ki ga bomo uporabili kot predlogo, podrobno ogledamo s kakšnim od prej omenjenih programov za vizualizacijo (npr. VMD, PyMOL, Chimera ipd.). Prav tako je priporočljivo, da pridobimo še originalni znanstveni članek, ki vsebuje podrobnejše

informacije o ozadju določitve 3D strukture tega proteina. Primarno zaporedje strukture proteina, ki bo predstavljal predlogo za homologno modeliranje, nato poravnano s primarnim zaporedjem tarčnega proteina. Včasih to poravnavo pred gradnjo 3D modela poskušamo še izboljšati, tudi z vključitvijo drugih modelov. Ko imamo predlogo in poravnani primarni sekvenci, se lotimo izgradnje našega modela. Za to lahko uporabimo različna razvita programska orodja, kot so *Modeller*, *I-TASSER Swiss-Model*, *MOE*, *Phyre2* in *Prime*. Omenjeni programi za gradnjo 3D modela uporabljajo različne pristope.

Ena možnost je **fragmentno sestavljanje** (ang. *fragment assembly*), kjer celoten model sestavimo iz ohranjenih strukturnih fragmentov, ki jih identificiramo v sorodnih strukturah. Najprej konstruiramo ohranjeno jedro nato pa dodamo še t. i. »spremenljive« regije, ponavadi iz knjižnic takšnih fragmentov. **Ujemanje segmentov** (ang. *segment matching*) je drug možen pristop, kjer protein razdelimo na niz kratkih segmentov, in 3D strukturo vsakega izmed njih nato generiramo ločeno glede na najboljšo dostopno predlogo, ki jo identificiramo v PDB banki. Tretja, najpogosteje uporabljena metoda, poskuša generirati 3D strukturo proteina tako, da leta **zadosti definiranim prostorskim omejitvam** (ang. *satisfaction of spatial restraints*). Metoda je dobila navdih po iz Poglavlja III že poznanih geometrijskih omejitvah, ki jih pridobimo v NMR eksperimentih. Le-te nato uporabimo za izračun in napoved 3D strukture. Najprej izvedemo več 3D prileganj med tarčnim proteinom in predlogo in tako pridobimo več nizov geometrijskih omejitev, kot so razdalje med ogrodnimi atomi, vrednosti torzijskih kotov itd. Te omejitve nato uporabimo v globalnem optimizacijskem postopku, kjer s pomočjo energijske minimizacije oz. geometrijske optimizacije ob upoštevanju teh omejitev iterativno najdemo najboljše pozicije vseh atomov v tarčnem proteinu.

Posebno pozornost moramo posvetiti generiranju 3D strukture tistih delov tarčne proteinske sekvence, ki niso poravnani z izbrano predlogo. Temu navadno pravimo **modeliranje zank** (ang. *loop modeling*), saj le-teh pogosto ne najdemo v izhodni proteinski strukturi. Zaradi manjše količine znanih podatkov uporabljenih pri napovedi 3D konformacij zank, je ta del končne 3D strukture slabše napovedan. V splošnem velja, da so napovedi za zanke, ki vsebujejo do 8 aminokislin, praviloma zadovoljive. Pri tem koraku velikokrat uporabljamo tudi knjižnice že znanih 3D struktur proteinskih zank, ali pa izvedemo lokalne geometrijske optimizacije z različnimi nastavki strukture zanke.

Dobljena začetna 3D konformacija proteina ima pogosto visoko vrednost potencialne energije, saj so določeni deli v konformacijsko in sterično neugodnih konformacijah. Zato dobljeni **3D model optimiziramo** tako, da izvedemo geometrijsko optimizacijo. Tako dobimo realnejšo in stabilnejšo napovedano 3D konformacijo proteina.

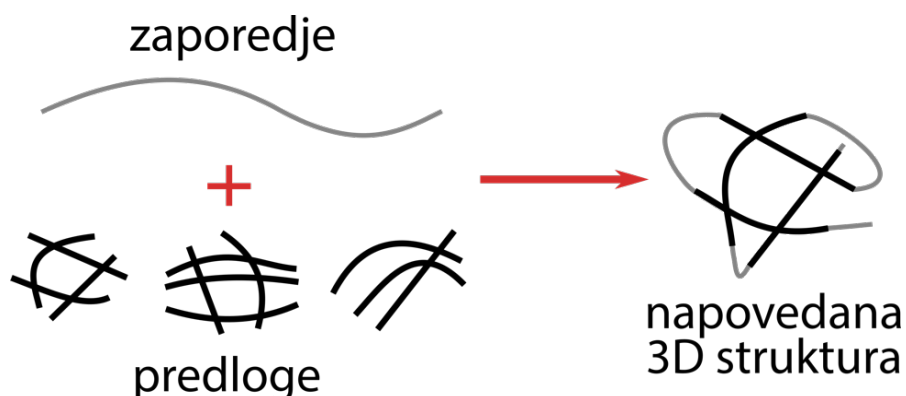
Kot zadnji korak homolognega modeliranja izvedemo še **validacijo dobljenega 3D modela**. Obstaja več pristopov in npr. lahko izvedemo statistično analizo, kjer pregledamo prisotnost in frekvenco kontaktov med posameznimi aminokislinskimi preostanki v napovedani 3D strukturi, v primerjavi z eksperimentalno rešenimi strukturami proteinov v proteinski bazi PDB. Razvili so tudi več globalnih in lokalnih indeksov, ki glede na njihovo vrednost kvantitativno pokažejo, kako dober je končni napovedan 3D model. Pri Vaji 4 boste pridobljeno znanje o homolognem modeliranju preiskusili tudi v praksi, ko boste poskusili napovedati 3D strukturo neznanega proteina.

2. METODA PREPOZNAVE PROTEINSKIH ZVITIJ

Metodo prepoznavanja oz. identifikacije zvitij (ang. *protein threading*, oz. *molecular fold recognition method*) uporabimo pri tistih proteinih, ki imajo podobno zvitje (ang. *fold*) kot proteini z znano 3D strukturo, ni pa še na voljo eksperimentalne strukture homolognega proteina, ki bi služil kot predloga za izgradnjo homolognega modela.

Uporabljamo jo kot metodo izbire za napovedovanje 3D struktur proteinov z nizko homologijo s tistimi proteini, za katere je 3D struktura že znana. Da to metodo lahko uporabimo, po navadi že obstaja protein z vsaj 20 % homologijo, ki ima določeno 3D strukturo. Tako kot homologno modeliranje tudi ta skupina metod spada med tiste, ki uporabljajo predloge oz. šablone delov proteinske strukture. Temeljna ideja za to metodo je, *da obstaja v naravi končno število možnih proteinskih zvitij ter da poznavanje preferenc posameznih aminokislinskih ostankov, da se nahajajo v določenih strukturnih okoljih, definira, katero zvitje bo prisotno*. Trenutno poznamo okoli 1300 različnih proteinskih zvitij. Obstaja veliko programov, ki z metodo prepoznavanja zvitja omogočajo napovedovanje 3D struktur, med drugimi: *Phyre, RAPTOR, PROSPECT, NovaFold in I-TASSER*.

Prej omenjen princip predstavlja tudi glavno razliko v primerjavi s homolognim modeliranjem, saj tu ne iščemo znane predloge med homolognimi proteini, ampak med proteini, ki imajo podobna zvitja. Definirajmo še dva izraza: *motif* = majhna specifična kombinacija elementov sekundarne strukture (glej Poglavlje I), *fold* = splošna oblika in orientacija elementov sekundarne strukture, pri čemer ne upoštevamo povezanosti med posameznimi strukturami. Slika 4 shematsko prikazuje korake napovedi 3D strukture proteina z metodo prepoznavanja zvitij.



Slika 4. Napoved 3D strukture proteina z metodo prepoznavanja zvitij.

Koraki metode prepoznavanja zvitij:

1. *Izdelava knjižnice različnih predlogov zvitij*
2. *Razvoj cenilne funkcije*
3. *Prileganje predlog in tarčnega primarnega zaporedja*
4. *Izgradnja 3D strukture z izbranim optimalnim predlogom*
5. *Optimizacija 3D modela*
6. *Validacija dobljenega 3D modela*

Pri tej metodi iz dostopnih baz 3D proteinskih struktur (npr. PDB, SCOP in CATH) izberemo proteine, ki imajo sorodna zvitja, in bi jih lahko uporabili kot možne predloge za izgradnjo 3D strukture tarčnega proteina. Nato je potrebno za oceno ujemanja med tarčnim primarnim zaporedjem in predlogi zvitij v knjižnici razviti ali uporabiti že razvito cenilno funkcijo (ang. *scoring function*). Dobro zasnovana cenilna funkcija omogoča optimalni izbor najboljših predlog zvitij. Nato prilegamo tarčno zaporedje z vsakim izmed strukturnih predlogov zvitij in sočasno s cenilno funkcijo ocenimo in izberemo najboljše predloge. To je tudi ključen korak te metode in tipično izberemo tisto prileganje, ki je bilo najboljše ocenjeno s cenilno funkcijo. V naslednjem koraku zgradimo 3D model tarčnega proteina tako, da postavimo atome proteinske verige tarčne primarne sekvence na poravnane 3D koordinate izbranega predloga proteinskega zvitja. Nato z različnimi pristopi, podobnimi tistim, ki smo jih opisali pri homolognem modeliranju, zgradimo končni 3D model proteina. Začetni model še optimiziramo z energijsko minimizacijo in ga evalviramo z uporabo različnih kvantitativnih kriterijev, podobnim kot smo opisali pri metodi homolognega modeliranja.

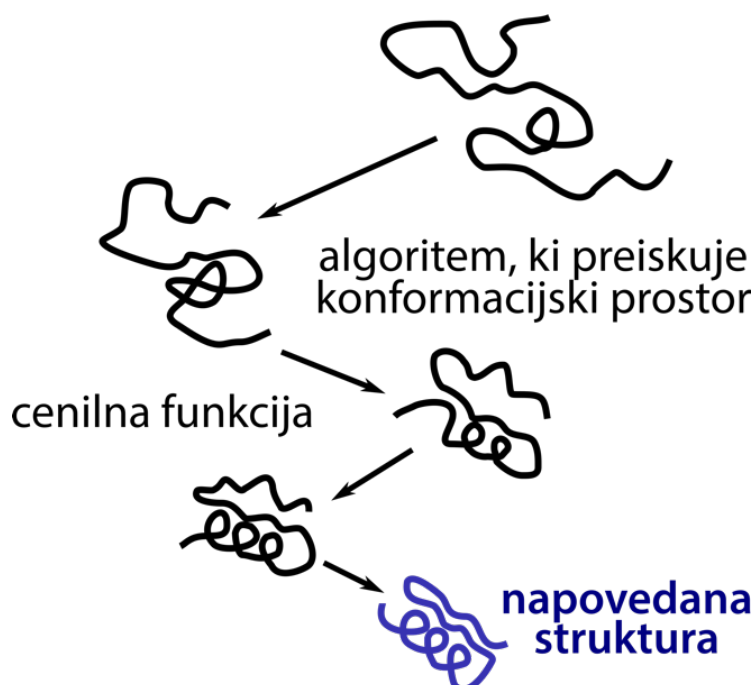
3. NAPOVEDOVANJE 3D STRUKTURE IZ PRVIH PRINCIPOV

Metoda napovedovanja strukture iz prvih principov (lat. *ab initio*) ne uporablja nobenih eksperimentalnih predlog, kot je to značilno za prejšnja dva pristopa. To metodo imenujemo tudi *de novo napoved 3D strukture proteinov* oz. modeliranje na osnovi fizikalnih zakonitosti (ang. *physics-based modeling*) oz. modeliranje brez šablon/predlog (ang. *free modeling*).

To metodo napovedi uporabimo takrat, ko ni znanega proteina z rešeno 3D strukturo, ki bi imel tako homologno primarno zaporedje, da bi ga lahko uporabili kot predlogo/šablono za gradnjo 3D strukture proteina. Temelj teh metod je t. i. **termodinamska hipoteza**, ki jo je postavil Nobelov nagrajenec Christian Anfinsen, in pravi, da nativna konformacija proteina ustreza globalnemu minimumu proste energije hidratiranega proteina.

Pri napovedovanju 3D strukture proteina iz prvih principov njegovo terciarno strukturo napovemo direktno iz njegovega primarnega zaporedja aminokislin. Veliko razvitih algoritmov poskuša oponašati principe, ki se predvidoma dogajajo med zvitjem proteina (t. i. *protein folding*) v končno fiziološko aktivno nativno konformacijo. Zaradi obsežnega konformacijskega prostora proteina, ki ga moramo preiskati, zahtevajo te metode velike računske kapacitete, zato je njihova uporaba primerna za manjše proteine, velikosti nekje do 150 aminokislinskih ostankov. Dostopnih je veliko programov, ki omogočajo *ab initio* napovedovanje 3D struktur, med drugimi: *ROSETTA*, *TASSER/I-TASSER* in *QUARK*.

Večinoma so napovedi proteinskih struktur, ki jih pridobimo s temi metodami, manj natančne. Povprečna napaka napovedi, če tako napovedano 3D konformacijo proteina prilegamo z eksperimentalno konformacijo (glede na dostopne konformacije), je med 5 in 10 Å. To je tudi pričakovano, saj pri generiranju 3D strukture ne vključimo nobenih predhodno znanih podatkov. Navkljub temu je pomembna prednost metode, da lahko identificiramo nova zvitja, ki do sedaj še niso bila poznana, ter prikažemo tudi vpogled v možen mehanizem proteinskega zvitja.



Slika 5. Napoved 3D strukture proteina iz prvih principov (*ab initio*).

Splošna paradigma praktično vseh *ab initio* metod napovedovanja 3D strukture proteinov je dejstvo, da vsebujejo **algoritem, ki preiskuje dostopen konformacijski prostor proteina** ter tako ustvari veliko možnih 3D struktur makromolekule. Ta algoritem mora preiskati tako možne konformacije glavne proteinske verige kot tudi stranskih verig aminokislinskih preostankov. Tako generirane proteinske konformacije ocenimo s pomočjo **cenilne funkcije**, ki izbere najprimernejše 3D strukture na podlagi njihove termodinamične stabilnosti, ki jo ocenimo kar z vrednostjo izračunane potencialne energije z molekulske mehanike.

Cenilne funkcije razdelimo v dve skupini. Fizikalno osnovane cenilne funkcije temeljijo na uporabi fizikalnih zakonov, ki opisujejo (med)molekulske interakcije. Empirične pa temeljijo na predhodnem poznavanju možnih konformacij proteinov in tako ocenijo posamezne generirane konformacije. Zelo pomembno je, da v procesu napovedi upoštevamo tudi vpliv topila na konformacijo proteina. V procesu grajenja kot tudi ocenjevanja možnih proteinskih konformacij pogosto uporabimo tudi energijsko minimizacijo ter simulacije molekulske dinamike.

Uspešne *ab initio/de novo metode* napovedi morajo zadostiti nekaterim dejavnikom:

1. *Natančna cenilna funkcija, ki pripiše termodinamično najstabilnejše stanje tisti napovedani 3D strukturi, ki najbolj ustreza nativni konformaciji proteina*
2. *Učinkovita metoda preiskovanja konformacijskega prostora, ki omogoča prepoznavanje nizkoenergijskih konformacij proteina*
3. *Sposobnost identifikacije energijsko ugodnih konformacij proteina, ki so blizu nativni konformaciji*

Kot zanimivost na koncu še omenimo, da v primeru proteinov, ki jih sestavlja več kot 150 aminokislinskih ostankov, lahko *de novo metode* napovedovanja strukture uporabimo tudi tako, da protein razdelimo na proteinske domene in njihove 3D strukture napovemo ločeno.

Proteinska domena je del proteina, ki se zvije v terciarno strukturo neodvisno od preostalega proteina. Razvili so tudi pristop, kjer se protein razdeli na majhne peptidne fragmente velikosti med 3 do 9 aminokislinskih preostankov in njihove napovedane 3D strukture nato združimo.

Andrej Perdih

uporaba

proteinskih struktur

pri načrtovanju biološko

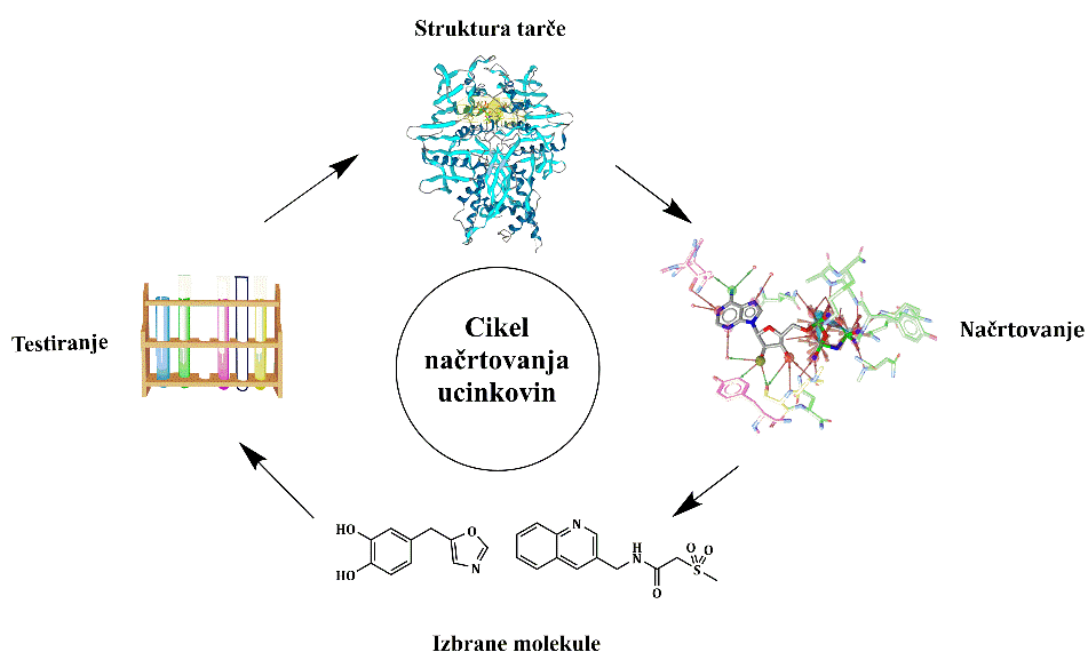
aktivnih molekul

```
function quicksort(array)
var list less, equal,
if length(array) ≤ 1
return array
select a pivot value pivot
for each x in array
if x < pivot then less
if x = pivot then equal
if x > pivot then greater
return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

1. CIKEL RACIONALNEGA NAČRTOVANJA UČINKOVIN

Molekularno ozadje delovanja biološko aktivnih spojin si lahko predstavljamo kot ugodno komplementarnost v geometrijskih in elektrostatskih karakteristikah takega **liganda**, ki je navadno majhna molekula, in njegovega vezavnega mesta na **tarčni biološki makromolekuli**, navadno proteinu, ki vodi do spremembe fiziološke funkcije tarče in potencialno terapevtskega učinka.

Odkrivanje in načrtovanje zdravilnih učinkovin (ang. *drug discovery oz. development*) sta že dolgo prisotna in sta se pričela kot naključno iskanje molekul z zdravilnimi lastnostmi. Omenjeni način danes vedno bolj nadomešča usmerjeno **racionalno načrtovanje učinkovin**, ki temelji na razumevanju strukture in funkcije biološke tarče. Za uspešnost tega pristopa so še posebej pomembne metode določanja 3D struktur bioloških makromolekul, ki smo jih spoznali v prejšnjih poglavjih, kot sta proteinska kristalografija in NMR spektroskopija, saj dobljene strukture lahko uporabimo pri racionalnem načrtovanju zdravilnih učinkovin. Prav tako so v ta namen uporabne tudi 3D proteinske strukture, ki jih dobimo z računskimi metodami.

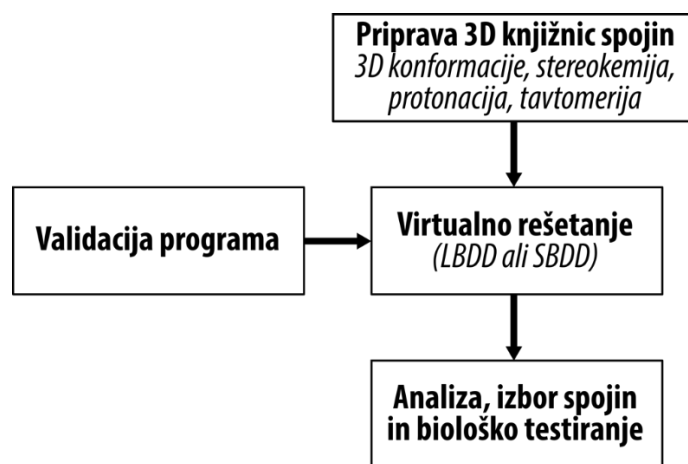


Slika 1. Poenostavljena shema cikla načrtovanja novih zdravilnih učinkovin.

Načrtovanje biološko aktivnih molekul z določenimi makromolekulskimi interakcijami je kompleksen izziv. Proces racionalnega načrtovanja učinkovin najpogosteje opišemo s cikličnim procesom, imenovanim tudi **cikel načrtovanja učinkovin** (ang. *drug design cycle*), kot to prikazuje Slika 1. V njem uporabimo *strukturne podatke* ali podatke o znanih ligandih za *načrtovanje* in *iskanje/reševanje* novih biološko aktivnih spojin t. i. **zadetkov** (ang. *hit molecules*). Sledi ovrednotenje biološke aktivnosti *izbranih molekul* v različnih *eksperimentih/testiranjih* in v optimalnem scenariju medtem pridobimo še *nove strukturne podatke* o njihovi vezavi. S temi podatki pričnemo nov krog **optimizacije** začetnih spojin, tako z vidika biološke aktivnosti kot fizikalno-kemijskih parametrov. V več iteracijah lahko na koncu dobimo spojino z želenimi lastnostmi, **predkliničnega kandidata**, ki vstopi v predklinična in nato klinična testiranja do zdravilne učinkovine. V tem poglavju bomo na kratko spoznali korake načrtovanja v tem ciklu in kako v ta namen uporabimo 3D strukture proteinskih molekul.

2. RAČUNALNIŠKO PODPRTO NAČRTOVANJE BIOLOŠKO AKTIVNIH MOLEKUL

Tehnike računske kemije, ki jih uporabljamo kot pomoč pri načrtovanju in iskanju novih biološko aktivnih molekul, imenujemo **virtualno rešetanje** (ang. *virtual screening* - VS). VS je definirano kot avtomatizirano preiskovanje knjižnic virtualnih molekul z uporabo računalniške tehnologije, z namenom preiskati in zožiti obstoječi kemijski prostor na nekaj razredov spojin, ki so z veliko verjetnostjo aktivne na preiskovani tarči. Metoda virtualnega rešetanja je sorodna svoji eksperimentalni različici, **rešetanju visokih zmogljivosti** (ang. *high-throughput screening* - HTS), le da pri VS različici »rešujemo« in iščemo spojine z računalnikom (*in silico*). Slika 2 shematsko prikazuje splošen postopek virtualnega rešetanja.



Slika 2. Shema postopka virtualnega rešetanja.

Pred izvedbo virtualnega rešetanja je zelo pomembna priprava **knjižnice 3D molekul**, kjer vsako molekulo opišemo z ustrežno konformacijo, včasih celo z nizom več konformacij. Zato uporabimo **generatorje tridimenzionalnih struktur**, ki zgradijo smiselne 3D modele molekul. Previdno moramo ovrednotiti kemijsko smiselnost dobljenih konformacij, pravilnost protonacijskih vzorcev, aromatičnost, stereokemijo in upoštevati možne tautomerne oblike molekule. V primeru slabo in nepremišljeno pripravljenih virtualnih knjižnic je tudi uspeh virtualnega rešetanja znatno manjši.

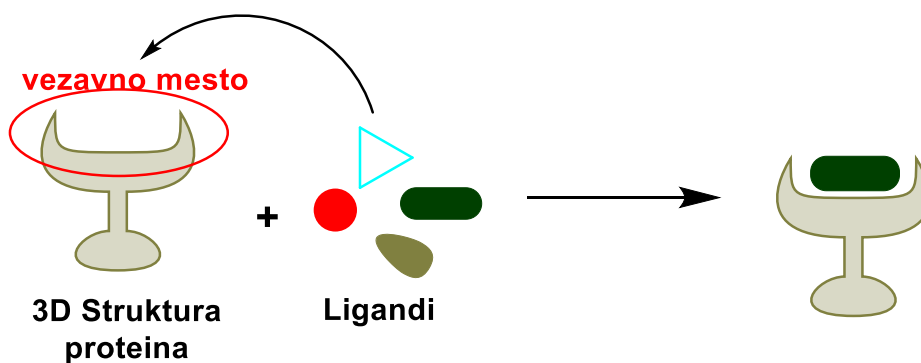
Knjižnice spojin, v katerih iščemo potencialno zanimive spojine, so **stvarne** (komercialne ali korporativne knjižnice dostopnih spojin) ali pa **virtualne**. Pomembne so tudi **knjižnice majhnih fragmentov** s spojinami, ki imajo molekulska masa do okoli 200 Da. Pri načrtovanju in izgradnji virtualnih knjižnic je potrebno paziti, da je načrtovane spojine možno sintetizirati, kar je ob poznavanju ustreznih sinteznih poti z novjšimi programi za načrtovanje knjižnic že možno predhodno ovrednotiti.

Glede na namen ločimo še **splošne knjižnice spojin**, ki so namenjene rešetanju različnih tarč in vsebujejo veliko število strukturno zelo raznolikih spojin, ki jih je lahko tudi več milijonov. **Ciljane** (ang. *targeted*) **knjižnice spojin** so namenjene za raziskave na družinah sorodnih tarč (npr. na protein kinazah) ali pa na specifični tarči (npr. na CDK protein kinazi). **Usmerjene** (ang. *focused*) **knjižnice spojin** poskušajo, glede na identificirano spojino vodnico, čim bolje zapolniti kemijski prostor njenih možnih derivatov. V zadnjih letih se poleg knjižnic sinteznih molekul pojavljajo tako javno dostopni viri kot komercialni ponudniki **knjižnic spojin naravnega izvora**.

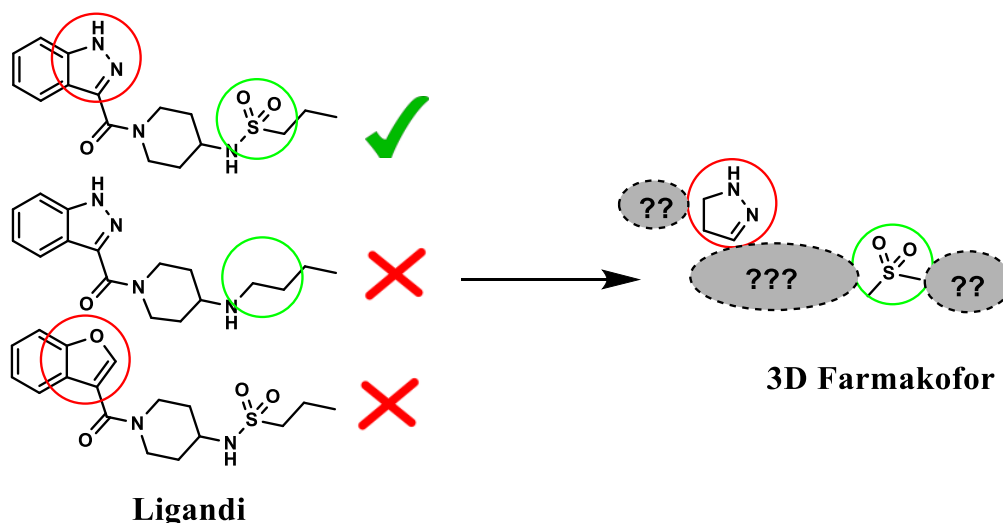
Naj kot zanimivost omenimo, da obstajajo tudi metode virtualnega reševanja, kjer pri reševanju knjižnic ne uporabljamo tridimenzionalnih struktur, ampak npr. topološko dvodimenzionalno predstavitev molekul, molekulska interakcijska polja ter še druge pristope.

V drugem koraku virtualnega reševanja opredelimo **robne pogoje**, ki jim mora virtualna molekula zadostiti, da bo v procesu reševanja identificirana kot potencialni zadetek. Poznamo dve skupini metod virtualnega reševanja.

1. Pri načrtovanju in iskanju novih biološko aktivnih spojin lahko za načrtovanje uporabimo dosegljive podatke o tridimenzionalni strukturi tarče ali kompleksa tarče z ligandom. Takemu pristopu pravimo **načrtovanje na osnovi strukture tarče** (ang. *Structure-Based Drug Design - SBDD*).



2. Druga možnost pa je, da robne pogoje postavimo na osnovi strukture in lastnosti poznanih aktivnih ligandov. To je **načrtovanje na osnovi aktivnih ligandov** (ang. *Ligand-Based Drug Design - LBDD*).

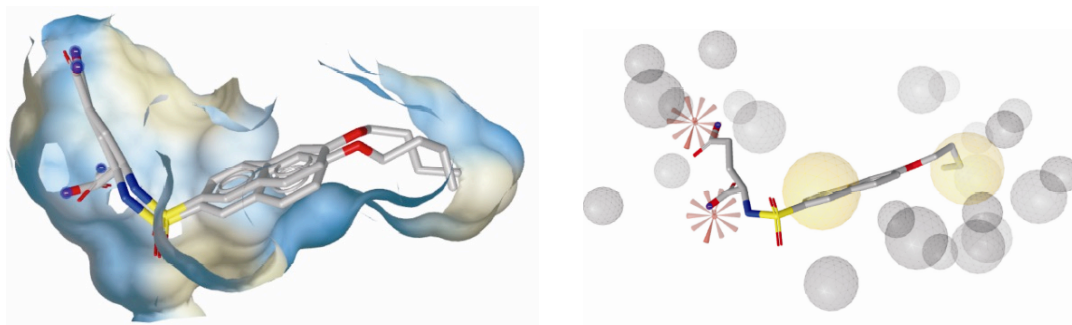


Oba pristopa se med seboj lahko tudi dopolnjujeta ter uporabljata sočasno. Kateri pristop izberemo, je odvisno od konkretnega obravnavanega primera in eksperimentalnih podatkov, ki jih imamo na voljo. Glavna metoda načrtovanja na osnovi strukture tarče je **molekulska sidranje**, pri načrtovanju na osnovi aktivnih ligandov pa se najpogosteje uporabljajo **tridimenzionalni farmakoforni modeli**.

Uporabljeno metodo virtualnega rešanja moramo tudi validirati. Med **validacijo metode VS** preverimo, kako se pod postavljenimi robnimi pogoji obnašajo molekule, za katere imamo tudi eksperimentalne podatke, npr. kristalno strukturo kompleksa te molekule s proteinom. Tako lahko direktno primerjamo eksperimentalne in izračunane vrednosti in s tem pridobimo nove informacije, ki omogočajo nadaljnjo optimizacijo robnih pogojev pri rešanju knjižnic virtualnih molekul.

Primer

Bakterijski encim MurD ligaza, ki sodeluje v biosintezi peptidoglikana, ima določeno 3D kristalno strukturo proteina z vezanim inhibitorjem, ki so jo določili z metodo proteinske kristalografije. V banki PDB jo najdemo pod oznako 2JFF. Metodo načrtovanja na osnovi strukture tarče, molekulske sidranje, smo uporabili za virtualno rešanje novih inhibitorjev. Metodo smo validirali tako, da smo sidrali v aktivno mesto vezani inhibitor in primerjali dobljeno vezavno pozo. Izkazalo se je, da pri danih pogojih sidranja lahko popolnoma reproduciramo vezavno geometrijo. Nato smo validirali še uporabljen metodo načrtovanja na osnovi aktivnih ligandov 3D farmakofornih modelov. Z uporabo kristalne strukture smo postavili na strukturi temelječ 3D farmakoforni model. Nato smo prilegali na ta model molekulo inhibitorja, za katero smo neodvisno zgradili več možnih konformacij. Rezultat prikazan spodaj pokaže, da smo z 3D farmakofornim modelom uspeli reproducirati orientacijo, kot jo najdemo v aktivnem mestu v kristalni strukturi.



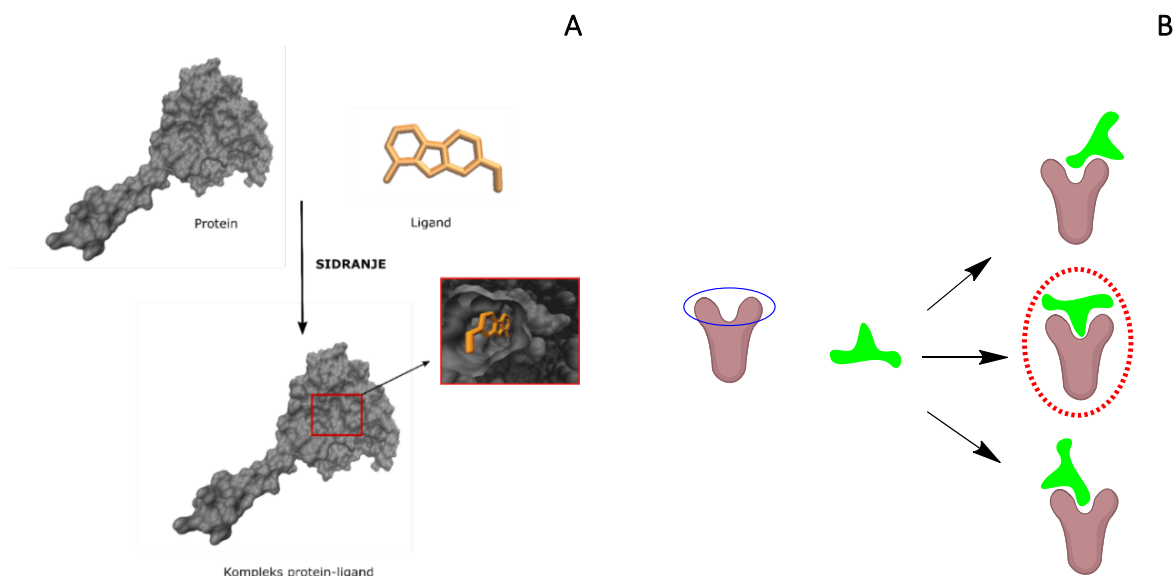
Na koncu postopka virtualnega rešanja **analiziramo** dobljene rezultate in **izberemo spojine iz preiskovane knjižnice**, ki jih ovrednotimo s **testi vrednotenja biološke aktivnosti** na izbrano tarčo, V primeru, da so bile spojine v knjižnici le virtualne, le-te predhodno še sintetiziramo.

Postopek virtualnega rešanja je iterativen in del prej omenjenega cikla načrtovanja učinkovin, saj novi eksperimentalni podatki, ki jih dobimo po testiranju izbranih spojin, vodijo v postavitev novih pogojev rešanja, dokler ne dobimo spojin z optimalno biološko aktivnostjo in želenimi fizikalno kemijskimi lastnostmi.

3. MOLEKULSKO SIDRANJE

Molekulsko sidranje (ang. *molecular docking*) je metoda računalniško podprtega načrtovanja, ki izračuna preferenčno konformacijo izbrane molekule, navadno majhne molekule liganda, lahko pa tudi makromolekule v izbranem aktivnem mestu biološke makromolekule (tarče), ob predpostavki, da le-ti tvorita stabilen kompleks (Slika 3A).

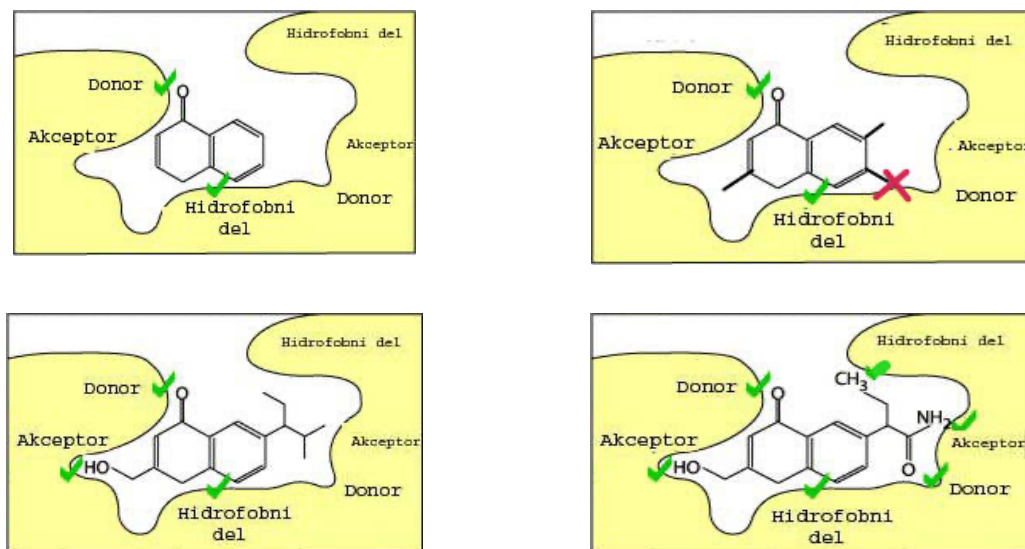
Programi molekulskega sidranja sestavljata dve skupini algoritmov. **Iskalni algoritmi** (ang. *search algorithm*) skrbijo za nastajanje vedno novih veznih konformacij, **cenilne funkcije** (ang. *scoring function*) pa nastali vezavni geometriji liganda izračunajo pripadajočo energijo afinitete, ki predstavlja oceno intenzitete interakcij z makromolekulo. Med računanjem geometrije kompleksa se algoritma dopolnjujeta, tako da cenilna funkcija služi tudi kot kvantitativno merilo, katere konformacije se bodo prenesle v nov korak procesa sidranja. Rezultat molekulskega sidranja vedno predstavlja niz po afiniteti razvrščenih (rangiranih) izračunanih kompleksov ligand/makromolekula (Slika 3B).



Slika 3. (A) Shematski prikaz metode molekulskega sidranja. Iskalni algoritem izračuna možne orientacije liganda v aktivnem mestu tarče, cenilna funkcija pa oceni ugodnost medmolekulskih interakcij. **(B)** Shematski prikaz rezultata molekulskega sidranja - več možnih konformacij, rangiranih po vrednosti cenilne funkcije.

Večina programov za molekulskega sidranje, *AutoDock*, *e-HITS*, *DOCK*, *Glide*, *GOLD*, in drugi, med računom konformacije kompleksa obravnava **ligand fleksibilno** in spreminja njegovo konformacijo, medtem ko je **makromolekula statična**. To seveda vodi do napak pri določanju vezavnih poz, ki se jim nekateri programi vsaj delno izognejo z upoštevanjem fleksibilnosti stranskih verig ali pa dela aktivnega mesta. Zato program za molekulskega sidranje, kot smo omenili prej, pred reševanjem knjižnic molekul validiramo in določimo, pri kakšnih pogojih uspešno reproducira vezavne konformacije ligandov, za katere imamo tudi eksperimentalne podatke - kristalno strukturo kompleksa. Validacije širokega spektra programov za molekulskega sidranje so pokazale, da ti programi večinoma uspešno reproducirajo eksperimentalno določene vezavne geometrije ligandov.

Programi za molekulskega sidranje pri gradnji molekulskega kompleksa uporabljajo predvsem dva iskalna algoritma. Prvi temelji na konceptu **inkrementne konstrukcije**, kjer v aktivno mesto najprej sidramo večji fragment molekule, nato pa z uporabo različnih omejitvenih kriterijev, bodisi energijskih ali geometrijskih, do konca zgradimo konformacijo molekule v aktivnem mestu. To si shematsko lahko ogledamo na Sliki 4. Druga pogosta iskalna metoda so **genetski algoritmi**, ki poskušajo pri ustvarjanju novih konformacij oponašati principe evolucije in selekcije, kot so mutacije, cross-over, itd.



Slika 4: Prikaz iskalnega algoritma inkrementne konstrukcije. **(1)** Program vstavi bazni fragment (črno) v vezavno mesto na proteinu (rumeno). Pozicija liganda je odvisna od energijsko najugodnejše interakcije s proteinom. **(2)** Algoritem nato po korakih dodaja atome baznemu fragmentu; tiste, ki so energijsko ugodni, obdrži (zeleno), tiste, ki niso, odstrani (rdeče). **(3)** in **(4)** Predhodna koraka se ponovita tolikokrat, dokler ne sestavimo liganda, ki je v vezavnem mestu v energijsko najugodnejši konformaciji.

Drugi korak molekulskega sidranja je določitev **afinitete nastalega kompleksa** med ligandom in proteinom. Ko med seboj interagirata majhna molekula in njen makromolekularni partner, lahko nastane nekovalentno, včasih pa tudi kovalentno vezan supramolekularni kompleks, kot je to prikazano spodaj. Molekulsko sidranje na nivoju molekulske mehanike deluje le za nekovalentne vezave.



Ligand

Protein

Kompleks

Osnovna zveza, ki omogoča ocenitev afinitete kompleksov ligand-receptor, izhaja iz termodinamike in predstavlja povezavo med količinama ΔG , ki je sprememba Gibbsove proste energije, in K_d – disociacijska/vezavna konstanta.

$$\Delta G = -RT \ln K_d$$

V enačbi je ΔG razlika med $G^v - G^n$, torej prosto energijo v veznem G^v in v neveznem stanju G^n , K_d pa predstavlja konstanto kemijskega ravnotežja za reakcijo razpada kompleksa ligand-receptor. Vrednost RT pri sobni temperaturi je $0.59 \text{ kcal mol}^{-1}$.

$$K_d = \frac{[R] \cdot [L]}{[RL]}$$

Statistična termodinamika je del fizikalne kemije, ki omogoča matematično/fizikalno opisovanje termodinamskih lastnosti sistema na osnovi obnašanja delcev, ki sestavljajo

obravnavani molekularni sistem. Predstavlja povezavo med fizikalnimi teorijami, ki opisujejo obnašanje atomov in molekul - *kvantno mehaniko in molekulska mehaniko* ter *klasično termodinamiko*. Temelji na **ergodijski hipotezi**, ki trdi, da je *časovno povprečje poljubne količine enako ansambelskemu povprečju te količine*.

Enačbe statistične mehanike predstavljajo teoretično osnovo za razvoj **cenilnih (ang. scoring) funkcij**, ki omogočajo hitrejše in enostavnejše napovedovanje vezavne konstante molekul, in so enostavnejše od popolnoma fizikalno osnovanega pristopa izračuna. Cenilne funkcije so uspešne le pod pogojem, da le-ta zelo dobro opredeli fizikalne sile, udeležene v interakciji, in da je tridimenzionalna struktura receptorja znana.

Cenilne funkcije, ki ocenijo moč interakcij med ligandom in biološko makromolekulo, poskušajo oceniti prosto energijo vezave. Najprej so razvili metode, ki so težile k čim bolj natančni simulaciji fizikalno-kemijskega procesa vezave liganda na protein. Te metode uporabljajo simulacije molekulske dinamike za generiranje ustreznih molekulskih ansamblov ligand-tarča in omogočajo natančen izračun proste energije vezave z uporabo enačb statistične mehanike. Metode, kot so **perturbacija proste energije, termodinamska integracija in metoda počasne rasti**, so sicer zelo natančne, vendar žal tudi počasne in računsko zahtevne. V procesu načrtovanja učinkovin jih zato uporabljamo redkeje.

Kasneje so razvili metode, ki se poslužujejo mnogih približkov in pri računu zanemarijo določene aspekte vezave. Vsaka od skupin znanih cenilnih funkcij ima svoje prednosti in slabosti in trenutno nobena ne izpolnjuje vseh pogojev idealne funkcije, ki naj bi bila zanesljiva, natančna, robustna, splošno uporabna in fizikalno osnovana.

Cenilne funkcije delimo na tri skupine:

1. **Cenilne funkcije, ki temeljijo na polju sil**
2. **Empirične cenilne funkcije**
3. **Cenilne funkcije, dobljene s pomočjo statistične mehanike**

1. Cenilne funkcije, ki temeljijo na polju sil (ang. *force-field scoring function*) ocenijo interakcijo kot vsoto neveznih van der Waalsovih in elektrostatskih členov – t. i. *interakcijsko energijo*, ki ji dodajo še empirično izračunane solvatacijske prispevke. Te cenilne funkcije za izračun uporabljajo empirična polja sil MM, ki smo jih že spoznali, kot sta npr. AMBER in CHARMM. Izgradnja funkcije poteka navadno po dveh metodologijah.

Prva poizkušnja dosledno opisati termodinamski cikel, ki opisuje nastanek kompleksa med proteinom in ligandom, kar je zelo dolgotrajno. Drugi pristop pa predpostavlja, da struktura protein-ligand, pridobljena s proteinsko kristalografijo, ustreza strukturi kompleksa v ravnotežju. Interakcijska energija se nato izračuna na osnovi 3D koordinat kompleksa protein-ligand. Algoritem izračuna posamezne prispevke interakcijske energije, ki se zaradi aditivnosti termodinamskih funkcij dajo sešteti in dajo oceno celotne energije interakcije. Dobljeno enačbo so poimenovali temeljna (ang. *master*) enačba.

$$\Delta G_{vezave} = \Delta G_{interakcije} + \Delta G_{raztapljanja} + \Delta G_{konformacije} + \Delta G_{gibanja}$$

$\Delta G_{interakcije}$ predstavlja spremembo proste energije zaradi specifičnih elektrostatskih in van der Waalsovih sil. $\Delta G_{raztapljanja}$ je solvatacijska prosta energija, $\Delta G_{konformacije}$ predstavlja spremembo proste energije proteina in liganda pri nastanku kompleksa in posledično manjšo fleksibilnost obeh interakcijskih partnerjev. $\Delta G_{gibanja}$ je vsota členov $\Delta G_{rotacija}$, $\Delta G_{traslacija/rotacija}$ in $\Delta G_{vibracije}$, ki se spremenijo med tvorbo kompleksa.

2. Empirične cenilne funkcije (ang. *empirical scoring function*) temeljijo na eksperimentalno določenih kompleksih protein-ligand. Vezavna energija ΔG_{vezave} kompleksa je s koeficienti utežena vsota številnih empiričnih členov (npr. člena za H-vezi, hidrofobne interakcije, interakcije s kovinami ipd.).

$$\Delta G_{vezave} = \sum (\Delta G_{if_i}(l, r))$$

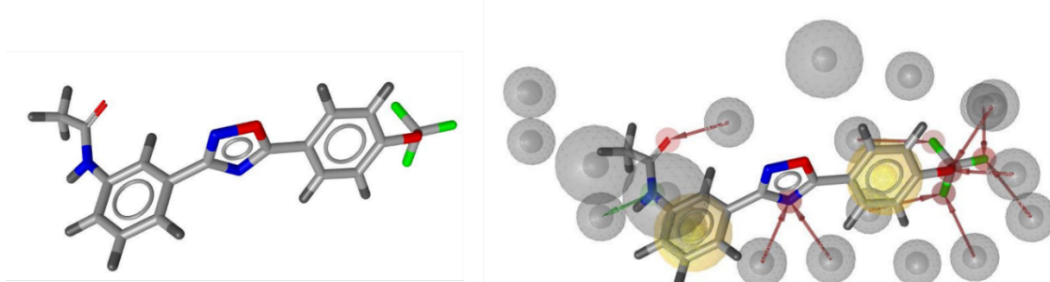
Utežne koeficiente ΔG_i pridobijo tako, da eksperimentalno strukturo korelirajo z eksperimentalnimi vrednostmi konstant (K_i) z eno od statističnih metod (npr. multivariantno analizo). Njihova glavna prednost je hitrost, saj potrebujejo za izračun le delček sekunde. Preciznost napovedi, ki jo te funkcije dosežejo, je 1-1.5 velikostnega razreda eksperimentalne K_i . Slabost empiričnih cenilnih funkcij je v dejstvu, da ni povsem jasno, do katere mere jih lahko uporabljamo za napovedi vezavnih konstant ligandov, ki se strukturno zelo razlikujejo od tistih, s pomočjo katerih so prvotno izračunali utežne koeficiente.

3. Cenilne funkcije, dobljene s pomočjo statistične mehanike (ang. *knowledge-based scoring function*), prav tako temeljijo na eksperimentalno določenih 3D strukturah kompleksov ligand-receptor. Osnovane so na predpostavki, da so medatomske razdalje, ki se pogosteje pojavljajo v kompleksih, ugodnejše od drugih razdalj. S statističnim vrednotenjem razdalj med atomi liganda in proteina dobimo preproste potenciale atomskih interakcij, ki nato služijo za oceno celokupne interakcije.

Zelo pogosta je uporaba več cenilnih funkcij skupaj, **ocenjevanje na osnovi konsenza** (ang. *consensus scoring*), ki v nekaterih primerih daje boljše rezultate.

4. TRIDIMENZIONALNI FARMAKOFORNI MODELI

Farmakofor je po IUPAC definiciji *zbirka steričnih in elektronskih lastnosti, ki so potrebne za zagotovitev optimalnih medmolekulskih interakcij s specifično biološko tarčo, in vodijo do indukcije ali inhibicije biološkega odziva* (Slika 5).











Slika 5. Mala molekula in njen 3D farmakoforni model, ki ustreza njenim steričnim in elektrostatskim lastnostim. Za definicijo farmakofornih elementov glej Preglednico 1.

Posledice, ki sledijo iz te definicije, so:

1. Farmakofor opisuje nujne sterične in elektronske točke v prostoru, ki so potrebne za optimalno interakcijo liganda s preiskovano farmakološko relevantno tarčo.
2. Farmakofor nikoli ne predstavlja realne molekule/liganda ali zbirke funkcionalnih skupin in je najvišji nivo abstrakcije le-teh, ki še vedno upošteva sposobnosti funkcionalnih skupin za interakcijo s preiskovano tarčo.
3. Farmakofori niso specifične funkcionalne skupine ali deli molekul.

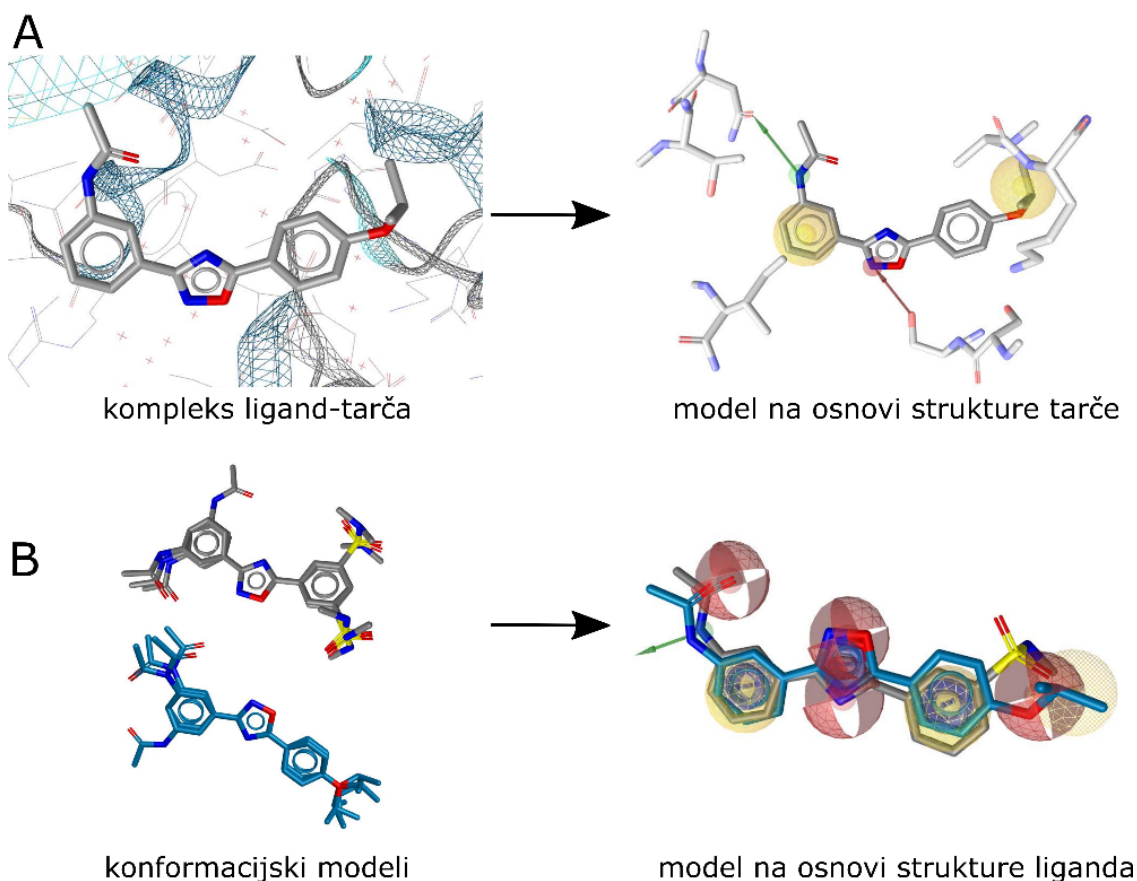
V Preglednici 1 so navedeni pomembnejši **farmakoforni elementi** in njihove shematske oznake, ki jih uporabljamo za postavitve 3D farmakofornega modela. Predstavljajo točke v prostoru, obdane z določenim *tolerančnim radijem*, v katerem mora ležati fragment preiskovane molekule, ki izkazuje enake lastnosti kot posamezni farmakoforni element (npr. *ionski center*, *hidrofobna interakcija*, *sterična omejitev*, *aromatski obroč* itd.). Pri farmakofornih elementih, ki opisujejo *donorje in akceptorje vodikove vezi*, moramo poleg *tolerančnega radija* upoštevati še *usmerjenost H-vezi v prostoru*, zato jim moramo dodati še geometrijsko omejitev, ki omogoča tvorbo H-vezi le pri določenih vrednostih interakcijskega kota.

Preglednica 1. Pregled pomembnejših farmakofornih elementov.

Farmakoforni element	Funkcija	Farmakoforni element	Funkcija
	Kationski center		Hidrofobna interakcija
	Anionski center		Aromatski obroč
	Donor vodikove vezi		Interakcija s kovinskim ionom
	Akceptor vodikove vezi		Sterična (prostorska) omejitev

Obstajata dva temeljna pristopa gradnje farmakofornega modela (Slika 6).

1. Farmakofor lahko zgradimo s pomočjo znane tridimenzionalne strukture kompleksa protein/ligand, ki jo pridobimo s kristalografskimi ali NMR eksperimenti. To so **na strukturi temelječi farmakofori** (ang. *structure-based pharmacophores*).
2. Druga možnost je, da farmakofor postavimo na osnovi niza več ligandov z znano biološko aktivnostjo, za katere predpostavljamo, da se vežejo v isto aktivno mesto z identičnim vzorcem medmolekulskega prepoznavanja. To so **na ligandih temelječi farmakofori** (ang. *ligand-based pharmacophore*). Ključni parameter postavljanja farmakofornih modelov predstavlja tudi pravilno obravnavanje fleksibilnosti molekul.



Slika 6. Postopek grajenja (A) na strukturi temelječih farmakofornih modelov in (B) na ligandih temelječih farmakofornih modelov.

Pri virtualnem rešetanju zgrajeni farmakoforni model predstavlja robne pogoje, ki jih mora virtualna molekula izpolniti, da bo upoštevana kot potencialni zadetek. Največkrat ocenjujemo ujemanje molekul z različnimi metodami prileganja. Prileganje spojine na postavljeni farmakofor je lahko točkovno, v tem primeru je kriterij kakovosti prileganja kar tudi kar RMSD parameter, lahko pa temelji na osnovi prileganja lastnosti, kot so npr. elektrostatski potencial, elektronska gostota, volumen, oblika molekule itd. Razvite pa so bile tudi farmakoforne cenilne funkcije, ki pomagajo razvrstiti spojine, ki jih posamezni farmakofor prepozna kot ugodne.

Farmakoforne modele validiramo z nizi spojin z znanimi podatki o biološki aktivnosti ali pa poskušamo z njim reproducirati biološko konformacijo, če je le-ta znana. Rezultate virtualnega rešetanja na koncu še ovrednotimo z različnimi farmakofornimi cenilnimi funkcijami, od RMSD kriterija do različnih kompleksnejših geometrijskih kriterijev. Grafični prikaz in kritična analiza rezultatov pomagata izbrati najbolj obetavne spojine, ki jim tudi eksperimentalno ovrednotimo biološko aktivnost.

Katja Lakota in Alja Prah

vaja 1: spletna orodja in podatkovne baze o bioloških makromolekulah

```
function quicksort(array)
  var list less, equal, greater
  if length(array) <= 1
    return array
  select a pivot value pivot
  for each x in array
    if x < pivot then list = less
    if x = pivot then list = equal
    if x > pivot then list = greater
  return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

V grobem lahko razdelimo podatke, ki so zbrani v podatkovnih bazah, na:

1. Podatke o zaporedjih (1D) iz katerih lahko delamo analizo poravnav zaporedij pri različnih organizmih, iščemo motive ipd.
2. Strukturne podatke (3D podatki, uporabljajo računalniško grafiko) – iz njih napovedujejo funkcijo, vezavna mesta, pomagajo klasificirati podatke, določiti/razložiti lastnosti molekule, narediti napovedi in simulacije vezav, molekulskega gibanja ipd. S tem poskušamo razumeti funkcijo posamezne molekule in učinke modifikacij nativne oblike. V kolikor 3D oblika ni znana, lahko s strukturno bioinformatiko napovemo nativno obliko molekul.
3. Prisotnost/izražanje RNA, metabolitov v posamezni celici, celičnem modelu, organu (analize izražanja na mikročipih, masna spektrometrija) - nam pomaga razumeti, kako potekajo posamezni celični procesi, kako bolezenski procesi ali posamezne spojine spremenijo ravnotežje v celotnem biološkem sistemu, primerjati množico odzivov (signatures) in na podlagi tega napovedovati/razložiti učinke.

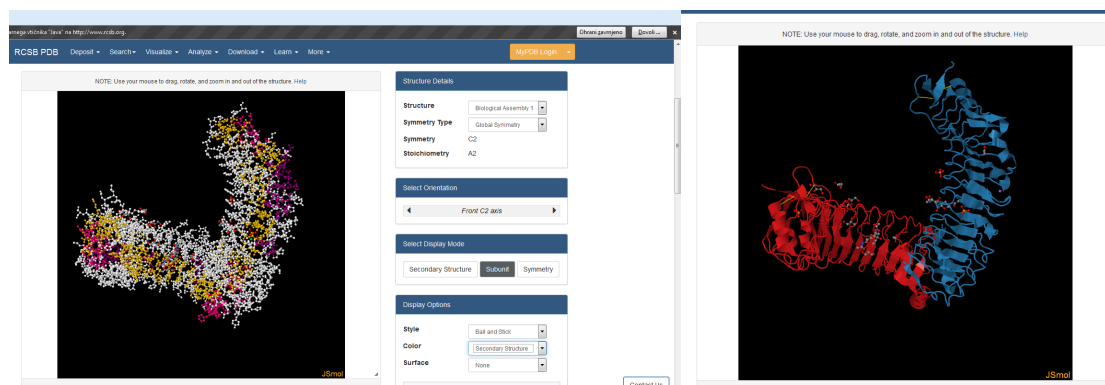
1. SPLETNE PODATKOVNE BAZE PROTEINSKE STRUKTURE

PDB (Protein Data Bank) je podatkovna baza 3D struktur proteinov in njihovih kompleksov in je bila ustanovljena leta 1971. Vsebuje več kot 160.000 struktur bioloških makromolekul. Podatke "deponirajo" raziskovalci in so prosto dostopni na spletu. Večina revij, ki objavlja strukture bioloških makromolekul zahteva, da so strukture deponirane v PDB.

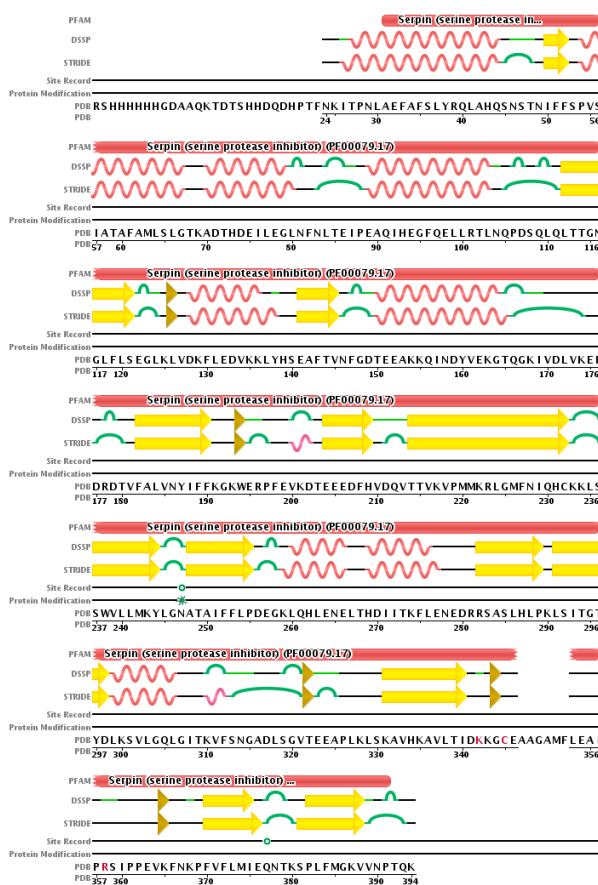
V iskalnik vpišite pojem 'TLR3' in izberite enega od zadetkov. Na prvi strani so podatki o tem kdo, kdaj in s kakšno metodo je določil 3D strukturo izbranega proteina.

The screenshot shows the RCSB PDB website interface. At the top, there is a navigation bar with options like 'Deposit', 'Search', 'Visualize', 'Analyze', 'Download', 'Learn', and 'More'. Below this is the PDB logo and a search bar. The main content area displays the entry for 1ZIW, titled 'Human Toll-like Receptor 3 extracellular domain structure'. It includes a 3D visualization of the protein structure on the left and a detailed information panel on the right. The information panel contains fields for Classification (IMMUNE SYSTEM), Deposited (2005-04-27), Released (2005-06-28), Deposition author(s) (Wilson, L.A., Choe, J.), Organism (Homo sapiens), Expression System (Spodoptera frugiperda), and Structural Biology Knowledgebase (1ZIW) with 15 annotations. Below this, there is an 'Experimental Data Snapshot' section with details: Method: X-RAY DIFFRACTION, Resolution: 2.1 Å, R-Value Free: 0.248, and R-Value Work: 0.208. To the right of this is a 'wwPDB Validation' section with a bar chart showing metrics like Metrics, Percentile Ranks, and Value for various categories such as Clashes, Ramachandran outliers, Sidechain outliers, and RSRZ outliers.

Če izberete zavihek *3D view*, lahko protein rotirate z miško, povečujete/zmanjšujete s kolesčkom, v stranskih možnostih lahko izberete prikaz aminokislin, sekundarnih struktur, vodikovih vezi, topilo dostopne površine.



Pod zavihkom *Sequence* se nahajajo opisi sekundarnih struktur (določene/predpostavljene po različnih podatkovnih bazah in modelih – Pfam, DSSP, STRIDE). Našli pa boste tudi zavihke s podobnimi sekvencami in strukturami pri drugih organizmih, ligande itd.



- Protein Modification Legend**
- * N-acetyl-D-glucosamine - N4-(N-acetyl(amino)glucosyl)-L-asparagine
- Site Record Legend**
- binding site for Mono-Saccharide NAG A 401 bound to ASN A 247 (Software)
- DSSP Legend**
- T: turn
 - E: beta strand
 - empty: no secondary structure assigned
 - B: beta bridge
 - S: bend
 - H: alpha helix


Vpišite v PDB receptor TLR3 (človeška ektodomena; PDB koda 1ZIW) in poiščite odgovore na naslednja vprašanja.

- *Kdaj, s katero metodo in s kakšno resolucijo je bila struktura določena?*
- *Kakšen je % vijačnic in beta ravnin in koliko aminokislinskih ostankov obsega?*
- *Koliko strukturam je podoben po sekvenci in koliko po strukturi, kateri sekvenci (PDB koda) je najbolj podoben in koliko je RMSD prileganja?*

2. SPLETNE PODATKOVNE BAZE INTERAKCIJ MED PROTEINI

Napovedovanje interakcij med proteini je področje, ki vključuje tako bioinformatiko kot tudi strukturno biologijo za identifikacijo interakcij med pari ali skupinami proteinov. Razumevanje interakcij med proteini je pomembno pri preučevanju medceličnega prepoznavanja, modeliranju proteinskih kompleksov in posledično za razumevanje številnih biokemijskih procesov.

Poiskali bomo interakcije tollu podobnega receptorja 3 (»Toll-like receptor 3« – TLR3), pomembnemu proteinu imunskega sistema, ki igra ključno vlogo pri prepoznavanju patogenov in aktivaciji imunskega sistema. V spletno **podatkovno bazo interakcij proteinov MINT** (»*Molecular Interaction Database*«: <http://mint.bio.uniroma2.it/>) v iskalno polje vtipkamo kratico proteina TLR3 in kliknemo *Search*. Prikažejo se nam podatki o interakcijah proteina TLR3.



The Molecular INTERaction Database
An ELIXIR Core Resource

Home Stats Download API Advanced Search Contacts About

Proteins, genes, public.

You searched for: tlr3

10 results for you query

Evidence List | Interaction Network

SHOW ENTRIES

FILTER RESULTS:

Gene A	Gene B	Interaction Type	Detection Method	PubMed	Details
CHEK2	TLR3	Physical association	Two hybrid array	25640309	+
Homo sapiens	Homo sapiens				
MBL2	TLR3	Physical association	Anti bait coimmunoprecipitation	24530528	+
Homo sapiens	Homo sapiens				
MBL2	TLR3	Colocalization	Fluorescence microscopy	24530528	+
Homo sapiens	Homo sapiens				
SRC	TLR3	Physical association	Anti tag coimmunoprecipitation	16858407	+
Homo sapiens	Homo sapiens				
TLR3	SRC	Colocalization	Fluorescence microscopy	16858407	+
Homo sapiens	Homo sapiens				
TLR3	PIK3R1	Physical association	Anti tag coimmunoprecipitation	16858407	+
Homo sapiens	Homo sapiens				
TLR3	PIK3R1	Physical association	Anti bait coimmunoprecipitation	16858407	+
Homo sapiens	Homo sapiens				
TLR3	TLR3	Direct interaction	Nuclear magnetic resonance	25217833	+
Homo sapiens	Homo sapiens				
TLR3	TLR3	Direct interaction	Nuclear magnetic resonance	25217833	+
Homo sapiens	Homo sapiens				
Ywhae	Tlr3	Colocalization	Cosedimentation	16615898	+
Mus musculus	Mus musculus				

Spletni program je našel interakcije TLR3 proteina, zraven lahko vidimo vrsto interakcije in eksperimentalno metodo, s katero je bila interakcija določena. S klikom na PubMed kodo dobimo tudi povezavo do članka, kjer je bila eksperimentalna študija o interakcijah objavljena.

Podobno orodje je tudi **STRING** (<https://string-db.org/>), ki podaja **funkcijske povezave med proteini**. Vnesemo lahko en protein ali skupino, izberemo organizem in kliknemo *Search*. Rezultati pod zavihkom *Analysis* ponudijo podatke o GO razvrstitvi prikazane mreže in o KEGG ter Reactome poteh, v katere je vključena prikazana družina proteinov. V zavihku *Settings* lahko nastavimo ali želimo samo eksperimentalno potrjene interakcije, take ki jih predlaga ko-ekspresija ipd. Prav tako lahko za povezovalne črte med posameznimi proteini tukaj nastavimo pomen (tip interakcije, stopnja zaupanja, da interakcija obstaja).



V MINT in STRING vpišite TLR3.

- *Koliko različnih interakcijskih partnerjev najde vsak program?*
- *Ali je katera od KEGG/REACTOME poti pomembno povezana z prikazanimi proteini?*
- *Kje v celici se nahajajo proteini iz funkcijske mreže?*

3. SPLETNA ORODJA ZA DELO Z ZAPOREDJI

NCBI je podatkovna baza ameriške medicinske knjižnice (US NLM) in je pomemben vir bioinformatičnih podatkov in orodij. Delo z zaporedji si bomo ogledali na primeru njenega spletno dostopnega vmesnika **BLAST** (*Basic Local Alignment Search Tool*). BLAST je algoritem za primerjanje primarnih bioloških zaporedji, kot so aminokislinska zaporedja proteinov in nukleotidna zaporedja nukleinskih kislin. BLAST služi tudi kot iskalno orodje za primerjavo preiskovanih zaporedij in identifikacijo podobnih zaporedij v knjižnicah ali podatkovnih bazah zaporedij.

V spletni podatkovni bazi NCBI Protein (<https://www.ncbi.nlm.nih.gov/protein>) v iskalnik najprej vpišemo TLR3 in kliknemo na »FASTA« v prvem zadetku:

The screenshot shows the NCBI Protein database search results for the query 'TLR3'. The search bar at the top contains 'TLR3' and the search button is highlighted. Below the search bar, there are options for 'Create alert' and 'Advanced'. The results are displayed in a list format, with the first item, 'TLR3 partial [Hydrobates leucorhous]', circled in red. The 'FASTA' link for this item is also circled in red. The page shows 'Items: 1 to 20 of 1953' and various filters and options.

Species: Animals (1,840), Bacteria (4), Viruses (3), Customize ...

Source databases: PDB (78), RefSeq (551), UniProtKB / Swiss-Prot (34), Customize ...

Genetic compartments: Mitochondrion (1)

Sequence length: Custom range...

Molecular weight: Custom range...

Release date: Custom range...

Revision date: Custom range...

Clear all

Show additional filters

Summary: 20 per page, Sort by Default order, Send to: Filters: Manage Filters

See [TLR3](#) to [like receptor 3](#) in the Gene database
[tlr3](#) reference sequences [Genomic \(1\)](#) [Transcript \(2\)](#) [Protein \(2\)](#)

Find related data
 Database: Select
 Find items

Search details
 TLR3[All Fields]
 Search See more...

Recent activity
 Turn Off Clear

Q TLR3 (1953) Protein

rps16, partial (chloroplast) [Centaurea triumfettii subsp. semidecurrans] Protein

TLR3, partial [Hydrobates leucorhous] Protein

MINT: the Molecular INTERaction database

Q 0036-8075 (170754) PubMed

See more...

Items: 1 to 20 of 1953
 << First < Prev Page 1 of 98 Next > Last >>

- [TLR3 partial \[Hydrobates leucorhous\]](#)
 1. 391 aa protein
 Accession: ADV90560.1 GI: 319958765
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [TLR3 partial \[Picoides pubescens\]](#)
 2. 390 aa protein
 Accession: ADV90559.1 GI: 319958778
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [TLR3 partial \[Haemorrhous mexicanus\]](#)
 3. 390 aa protein
 Accession: ADV90558.1 GI: 319958776
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [TLR3 partial \[Falco naumanni\]](#)
 4. 390 aa protein
 Accession: ADV90557.1 GI: 319958774
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [TLR3 partial \[Dromaius novaehollandiae\]](#)
 5. 390 aa protein
 Accession: ADV90556.1 GI: 319958772
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [TLR3 partial \[Falco naumanni\]](#)
 6. 390 aa protein
 Accession: ADV75025.1 GI: 319881143
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [TLR3 partial \[Falco naumanni\]](#)
 7. 390 aa protein
 Accession: ADV75024.1 GI: 319881141
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)
- [TLR3 partial \[Haemorrhous mexicanus\]](#)
 8. 317 aa protein

Odpre se nam okno s aminokislinskim zaporedjem TLR3 proteina.

FASTA ▾

TLR3, partial [Hydrobates leucorhous]

GenBank: ADV90560.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

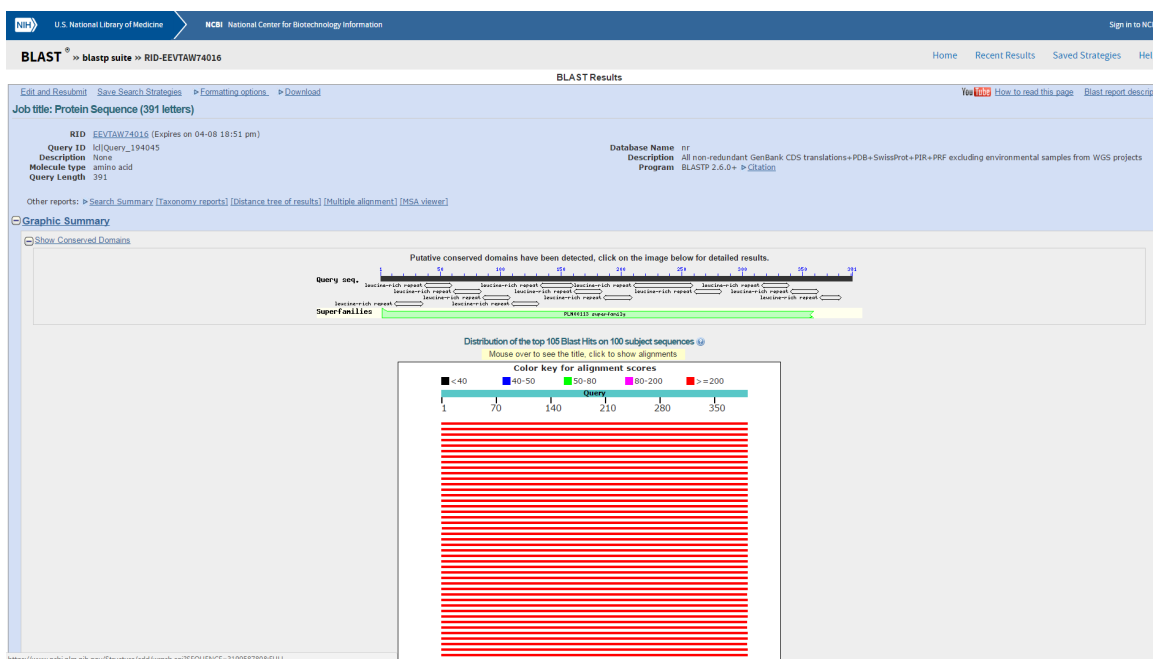
>ADV90560.1 TLR3, partial [Hydrobates leucorhous]

```
KLCTELSNITAIQNLSSLVHVKLSYIGKSTFQGLQGTNLTVLNLSRNSLSLIEDDSFQWLSLQYLNLKHNH
ITHVSSRLFYGLSSLKHLNLSLQNIEDFSFHLYHLEYLIMDNNHFGITANMFTGLNNLKYLKSLCN
CNINLQRITNKTFSSLANSSLQVLNLTTRISTIESGAFSSLGHLKILDGLNEISQELTGHEFKGLNNV
QDIYLSYNNKLNLRSESFIFVPSLRKLMRLKVGCSNLALSPSPFHPQLNLTVDLISNNNIANI KEDLFDG
LHKLDLIDLQHNMLARLWKHANPGGPVFLKDLPLNHLHLNLSKNSGLDEIPVQVFKGLFQLKHLHLSNNL
NLLPATLFDQASLHSLNQLNLTISVEEKVFGPAFKSLRI
```

Nato v iskalnik BLAST protein (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) v prazen prostor pod *Enter accession number(s), gi(s), or FASTA sequence(s)* skopiramo TLR3 aminokislinsko zaporedje, ki smo ga našli na NCBI spletni strani, in kliknemo *BLAST*.

The screenshot shows the NCBI BLAST search page. At the top, it says 'BLAST >> blastp suite' and 'Standard Protein BLAST'. The 'Enter Query Sequence' section has a text area containing the protein sequence: 'KLCTELSNITAIQNLSSLVHVKLSYIGKSTFQGLQGTNLTVLNLSRNSLSLIEDDSFQWLSLQYLNLKHNH...'. Below this are options to 'Or, upload file' and 'Job Title'. The 'Choose Search Set' section includes 'Database' (Non-redundant protein sequences (nr)), 'Organism' (Optional), 'Exclude' (Optional), and 'Entrez Query' (Optional). The 'Program Selection' section has 'Algorithm' options: 'blastp (protein-protein BLAST)', 'PSI-BLAST (Position-Specific Iterated BLAST)', 'PHI-BLAST (Pattern Hit Initiated BLAST)', and 'DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)'. At the bottom, the 'BLAST' button is highlighted with a red circle, and the text 'Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)' is visible.

Počakamo nekaj trenutkov in dobimo rezultate iskanja.



V našem primeru smo dobili 105 zadetkov, barvna lestvica predstavlja delež podobnosti iskanega aminokislinskega zaporedja z zadetki. Rdeča barva predstavlja najboljše ujemanje zaporedji. Več informacij o posameznih ujemanjih zaporedij dobimo v razdelku *Description*, kjer so podane E vrednosti, pri čemer nižja E vrednost pomeni boljše ujemanje zaporedji. BLAST poročilo rezultatov vsebuje še povezave do zunanjih virov, poravnave zadetkov s preiskovanim zaporedjem, taksonomsko poročilo in ostalo.

4. DRUGE UPORABNE BAZE S PODATKI O POSAMEZNIH GENIH/PROTEINIH

KEGG (*Kyoto Encyclopedia of Genes and Genomes*) je podatkovna baza za **razumevanje delovanja bioloških sistemov**, kot so celice, organizmi in ekosistemi. Omogoča razumevanje podatkov, pridobljenih s sekvenciranjem genomov in ostalimi visokozmogljivostnimi tehnologijami. Vsebuje 16 podatkovnih baz (biološke poti, genom, funkcionalni ortologi itd.).

Pogledali si bomo podbazo **KEGG PATHWAYS**, ki je zbirka ročno zgrajenih map bioloških poti. V njej so opisane molekularne interakcije in reakcijske poti metabolizma, procesiranja genetskih informacij, procesiranja okoljskih informacij, celičnih procesov organskih sistemov in človeških bolezni. V iskalnik KEGG PATHWAYS (<http://www.genome.jp/kegg/pathway.html>) vpišemo TLR3 in kliknemo *GO*.



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

Menu PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Select prefix: map Organism Enter keywords: TLR3 Go Help

[New pathway maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for:

- 1. Metabolism**
Global overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
- 2. Genetic Information Processing**
- 3. Environmental Information Processing**
- 4. Cellular Processes**
- 5. Organismal Systems**
- 6. Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

- 7. Drug Development**

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretation of higher-level systemic functions.

- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool
- Color Pathway - selected pathway map coloring tool

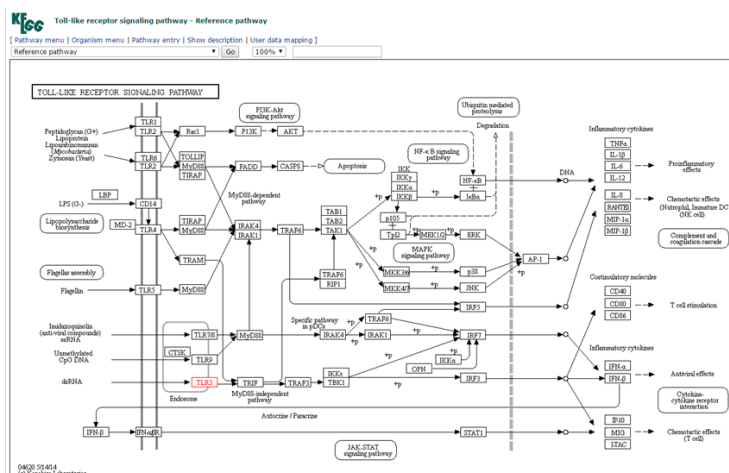
Kliknemo na sliko prvega zadetka.

Pathway Text Search

Number of entries in a page: 20 Hide thumbnail

Items: 1 - 5 of 5

Entry	Thumbnail
map04620	
map05160	
map05161	



Odpre se TLR3 signalna pot, TLR3 protein je označen z rdečo barvo. Slika prikazuje, kako se virusna dsRNA veže na receptor TLR3 in s tem s kaskado dogodkov sproži aktivacijo IRF3, ki v nadaljevanju aktivira interferon beta (IFN-β). Rezultat je protivirusni odziv imunskega sistema. S klikom na element v signalni poti dobimo podrobnejši opis. Če npr. kliknemo na *IFN-β* se odpre okno z dodatnimi informacijami.

KEGG ORTHOLOGY: K05415 Help

Entry	K05415	KO
Name	IFNB	
Definition	interferon beta	
Pathway	ko04060 Cytokine-cytokine receptor interaction ko04151 PI3K-Akt signaling pathway ko04380 Osteoclast differentiation ko04620 Toll-like receptor signaling pathway ko04621 NOD-like receptor signaling pathway ko04622 RIG-I-like receptor signaling pathway ko04623 Cytosolic DNA-sensing pathway ko04630 Jak-STAT signaling pathway ko04650 Natural killer cell mediated cytotoxicity ko05142 Chagas disease (American trypanosomiasis) ko05152 Tuberculosis ko05160 Hepatitis C ko05161 Hepatitis B ko05162 Measles ko05164 Influenza A ko05168 Herpes simplex infection	
Brite	KEGG Orthology (KO) [BR:ko00001] Environmental Information Processing Signal transduction 04630 Jak-STAT signaling pathway	

GeneCards (<https://www.genecards.org/>) je podatkovna baza o humanih genih/proteinih, ki združuje podatke več podatkovnih baz o genomu, proteomu, transkriptomu in povezavah s kliniko, obenem pa ima povezave/podatke o posameznih proizvajalcev za detekcijo/izvedbo bioloških eksperimentov. Vpišite v GeneCards encim GAPDH in poiščite odgovore na naslednja vprašanja.

- Na katerem kromosomu se nahaja gen?
- Kakšna je velikost proteina in kvartarna struktura?
- S katerimi proteini interagira?
- Katera zdravila vplivajo na izražanje GAPDH?
- Kateri SNPi so znani pri Homo sapiens?
- Poiščite povezavo na evlucijsko drevo (Evolution tree na ENSEMBL).
- Poiščite ortologe/paraloge/homologe vašega gena.

The screenshot shows the GeneCards website interface for the GAPDH gene. The main header includes the GeneCards logo and navigation options. The gene name 'GAPDH Gene (Protein Coding)' is prominently displayed, along with its full name 'Glyceraldehyde-3-Phosphate Dehydrogenase'. Below this, there are several tabs for different types of information: Jump to section, Research Products, and various biological categories like Disorders, Domains, and Expression. A table of aliases is provided, listing various identifiers such as 'Glyceraldehyde-3-Phosphate Dehydrogenase' and 'EC 1.2.1.12'. The page also features a 'Summaries' section with an 'Entrez Gene Summary' and a 'GeneCards Summary'. The Entrez summary describes the gene's function in carbohydrate metabolism and its role in various cellular processes. The GeneCards summary mentions associated diseases like Fmr1-Related Disorders and Schistosomiasis. At the bottom, there is a cookie consent banner.

Podobna stran Evropskega bioinformatičnega inštituta in Švicarskega inštituta za bioinformatiko je tudi UniProt (<https://www.uniprot.org/>). Poskušajte za encim GAPDH na tej strani najti naslednje podatke.

- Kateri družini proteinov pripada, kakšne domene in motive vsebuje?
- Kakšne mutacije so znane?
- Katere aminokisljine so modificirane?

strukture bioloških molekul

The screenshot shows the UniProtKB entry for F2YBL9 (F2YBL9_ANOGA). The page is organized into several sections:

- Names & Taxonomy:** Lists the protein name (APL18), gene name (APL18), and organism (Anopheles gambiae).
- PTM / Processing:** Contains two tables:

Feature key	Position(s)	Length	Description	Graphical view	Feature Identifier	Actions
Signal peptide	1 - 20	20	# Sequence analysis			Go to sequence
Chain	21 - 338	318	# Sequence analysis		PRO_000289721	Go to sequence

Position(s)	Length	Description	Graphical view	Feature Identifier	Actions
166 - 166	1	# Combined sources		CAA_300388976	Go to sequence
327 - 361	35	# Combined sources			Go to sequence
348 - 360	13	# Combined sources			Go to sequence
- Structure:** Provides information about the protein's structure, including the UniProt ID (F2YBL9), the protein name (APL18), and the chain length (338).

Podobna zbirka podatkovnih baz in orodij je še **Expasy** (*Expert Protein Analysis System*, <https://www.expasy.org/>), kjer so zbrani podatki in orodja, razdeljene po proteomiki, genomiki, evoluciji, transkriptomiki. Deluje v okviru Švicarskega inštituta za bioinformatiko.

Za encim GAPDH pod *Proteomics tools* poiščite orodje *ProtParam*. Vnesite UniProt/SwissProt identifier za GAPDH (klikniti je potrebno spodaj *Compute parameters* in nato *Submit*). Program poda sestavo (amnokislinsko, atome) in pl za vaš protein.

Isti protein vnesite še v *PeptideCutter* program in ugotovite, kateri encimi ga razgradijo in na katerih mestih.

The screenshot shows the Expasy Bioinformatics Resource Portal search results for the query 'tr3'. The search bar at the top indicates 'Query all databases' and 'tr3'. The search results are displayed in a table:

Resource	Hits	Category	Comment
EPD	3 hits	ge	3 promoters found in EPDnew that match 'tr3'
OMA	32 hits	ev, ge	
OrthoDB	171 hits	ev, ge	
PROSITE	0 hits	pr	PROSITE documentation entries
STRING	68 hits	pr	
SWISS-MODEL Repository	1 hit	bi, pr, st	
UniProtKB	686 hits	pr	UniProtKB entries
ViralZone	0 hits	pr	ViralZone pages (for given virus name/family)
Cellosaurus	4 hits	sy	cell lines
ENZYME	1 hit	pr	ENZYME entries
GPSDB	11 hits	ge, pr	gene synonyms
HAMAP	0 hits	pr	HAMAP family annotation rules
MetaNetX	0 hits	pr, sy	search results for MetaNetX
miROrtho	0 hits	ev, ge	orthologous group(s) returned by your request
MyHits	1 hit	pr	Protein found: 1, Motif found: 0, Other node found: 0,
OpenFlu	0 hits	ge	
Protein Spotlight	0 hits	pr	Spotlight/Prolume articles
Selectome	48 hits	ev	results in Selectome
SWISS-2DPAGE	0 hits	pr	SWISS-2DPAGE entries
SwissLipids	0 hits	pr, sy	No valid response from server.
SwissVar	1 hit	pr	Swiss-Prot protein - variant - disease groups
Vamozone	0 hits	pr	Vamozone pages for given uniprotk entry/family

Andrej Perdih in Alja Prah

vaja 2: vizualizacija mape elektronske gostote

```
function quicksort(array)
  var list less, equal, greater
  if length(array) <= 1
    return array
  select a pivot value pivot
  for each x in array
    if x < pivot then list = less
    if x = pivot then list = equal
    if x > pivot then list = greater
  return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

Elektronska gostota je merilo za *verjetnost nahajanja elektrona na določeni lokaciji*. V proteinski kristalografiji so mape elektronske gostote povprečje molekul v kristalu, ki kristalografom omogoča izgradnjo modela proteina. Kot smo omenili v Poglavlju II, moramo poznati tako amplitude kot tudi faze izmerjenih odbojev, da lahko izračunamo mapo elektronske gostote, ki jo je mogoče interpretirati. Med difrakcijskim eksperimentom ne moremo direktno določiti faze – to imenujemo tudi *fazni problem*. Fazo lahko posredno dobimo na več načinov, kot so molekulske zamenjave in obogatitve proteinskega kristala s težkimi atomi. Rešen 3D model strukture proteina kristalografiji nadalje izboljšujejo in ga nato deponirajo v spletne podatkovne baze struktur. Nekateri avtorji deponirajo tudi datoteke map elektronske gostote ali strukturnih faktorjev, ki jih lahko uporabimo za izračun map elektronske gostote.

Spletna podatkovna zbirka struktur PDB je primarni vir atomističnih 3D modelov. Zdaj je povezana tudi s spletno podatkovno bazo »*Electron Density Server*« (EDS), v primeru ko so poleg struktur dostopne tudi mape. EDS omogoča dostop do map elektronskih gostot in statističnih podatkov v zvezi s prileganjem kristalnih struktur in njihovih map.

V tej vaji bomo v programu **PyMOL** prikazali vizualizacijo 3D mape elektronske gostote na primeru proteina MurD (PDB koda *2JFF*), ki ga bomo uporabili tudi kot modelni sistem pri nalogi molekulskega sidranja pri Vaji 6.

1. NAMESTITEV PROGRAMA IN PRENOS USTREZNIH STRUKTUR

Na spletnem naslovu <https://pymol.org/edu/?q=educational/> se registriramo in naložimo prosto dostopno verzijo programa **PyMOL**. Nato v spletnem brskalniku v iskalnik vpišemo Protein Brookhaven Databank (<https://www.rcsb.org/>) in v iskalno polje vnesemo *2JFF*. Prenesemo strukturo proteina kot **.pdb* datoteko in si jo shranimo na računalnik, enako kot v navodilih za molekulsko sidranje (Vaja 6). Zaradi večje preglednosti v nadaljevanju preneseno datoteko *2jff.pdb* preimenujemo v velike tiskane črke *2JFF.pdb*. Nadalje na naslovu *Protein Data Bank in Europe* (<https://www.ebi.ac.uk/pdbe/>) v iskalno polje vpišemo *2JFF* in kliknemo na Download Files pri prvem zadetku.

2jff Crystal structure of MurD ligase in complex with D-Glu containing sulfonamide inhibitor

Kotnik M, Humljan J, Contreras-Martel C, Oblak M, Kristan K, Herve M, Blanot D, Urleb U, Gobec S, Dessen A, Solmajer T

J. Mol. Biol. (2007) [PMID: [17507028](https://pubmed.ncbi.nlm.nih.gov/17507028/)]

Source organism: [Escherichia coli](https://www.ncbi.nlm.nih.gov/taxonomy/Escherichia%20coli)

Assembly composition: protein only structure

Bound ligands: [LK2](#) [SO4](#)

Modified residues: [KCX](#)

PDB ID: [P14900](#)

[3D Visualisation](#) [Download files](#)

X-ray diffraction
1.89Å resolution
Released: 15 May 2007
 Model geometry
 Fit model/data

Nato izberemo Map in kliknemo na download ob EDS map.

Download ✕

Files for entry 2jff:

- + PDB**
- Map**
 - EDS map ([download](#))
 - EDS difference map ([download](#))
- + Assembly**
- + Molecule**
- + Validation**
- + SIFTS**

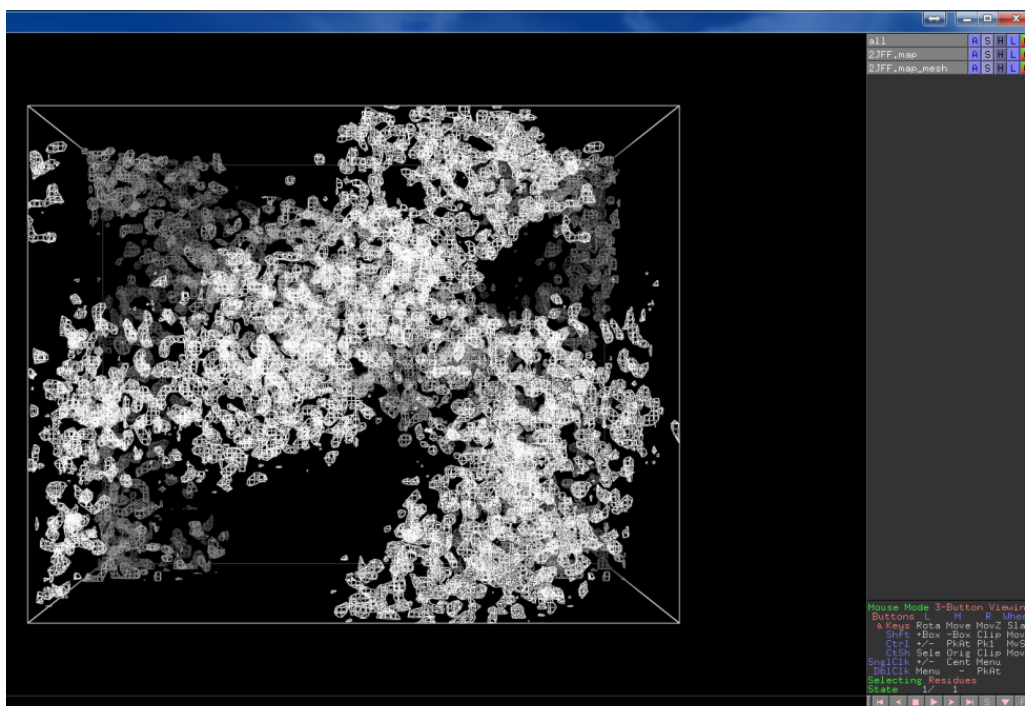
Preneseno datoteko 2jff.ccp4 preimenujemo v 2JFF.map.ccp4.

2. PRIKAZ MAPE ELEKTRONSKE GOSTOTE

Odpremo program PyMOL in v meniju z ukazom »**File/Open**« naložimo preneseno datoteko 2JFF.map.ccp4. Najprej se v grafičnem oknu nič ne spremeni. V nadaljevanju v oknu imen struktur na desni strani kliknemo na gumb »**2JFF.map**«, pojavi se oris prazne kocke. Nato v vrstici gumba **2JFF.map** izberemo »**A**« (»Action«) in »**mesh**« ter »@ level 3.0«:



S tem ukazom ustvarimo nov zaznamek »**2JFF.map_mesh**« in v grafičnem oknu se v kocki pojavi elektronska mapa gostote nivoja **3.0**.



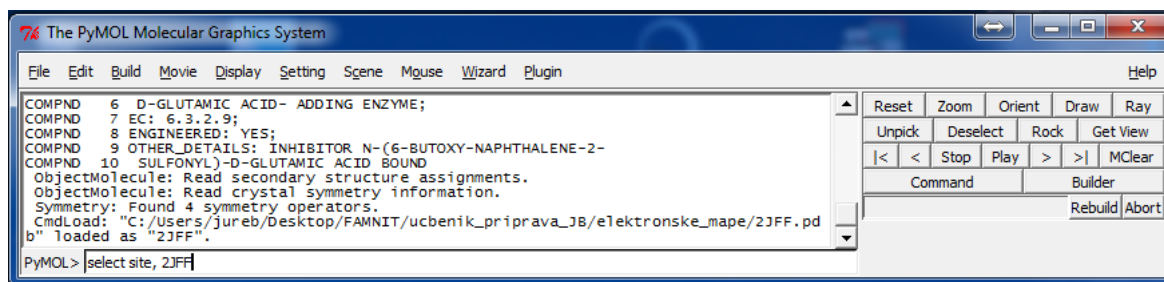
Mapa vsebuje veliko informacij, zato jo bomo v nadaljevanju zmanjšali, preizkusili bomo tudi različne nivoje gostote mape.

Navadno se vrednosti nivojev kontur predstavi v absolutni vrednosti elektronov na kubični Å (\AA^3). Kliknemo na »A« zraven imena **2JFF.map.mesh** in izberemo »level 2.0«. S tem se bo spremenila gostota v grafičnem oknu. Po želji lahko za naš izbrani primer spreminjamo vrednosti nivojev gostote od -2 do +5, največjo gostoto vidimo pri nivoju 0. Za potrebe nadaljnjih nalog bomo nastavili nivo na vrednost 2.

3. ZMANJŠANJE OBMOČJA PRIKAZA MAPE ELEKTRONSKE GOSTOTE

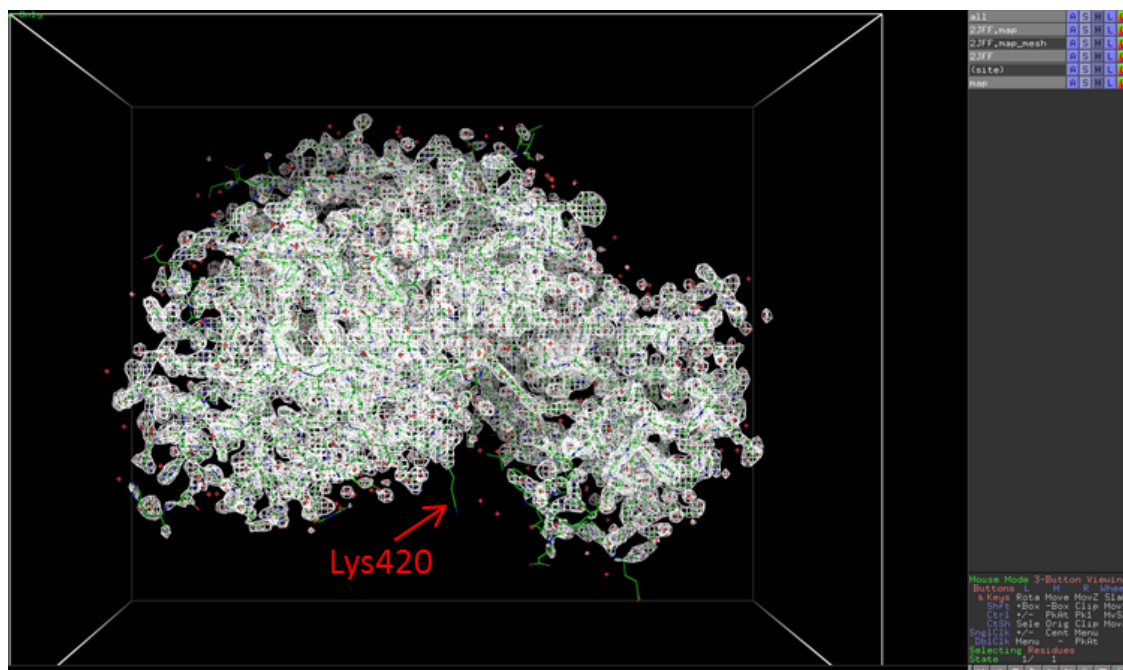
V meniju »File/Open« odpremo 2JFF.pdb. PyMOL avtomatsko poveča lokacijo naloženih atomskih koordinat strukture. Struktura je delno prekrita z mapo elektronske gostote, ki jo bomo v nadaljevanju delno odstranili.

V ukazno vrstico vtipkamo ukaz »select site, 2JFF«.



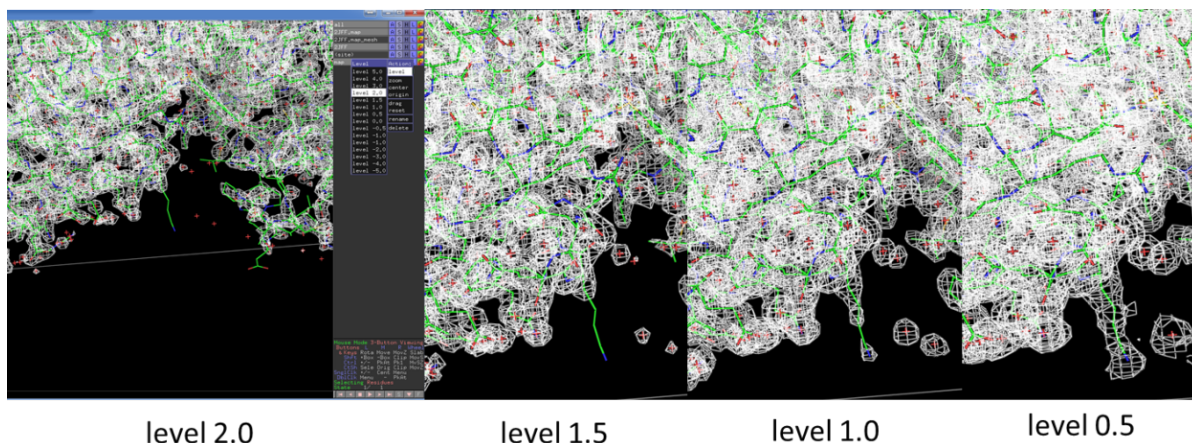
In nato še ukaz »isomesh map, 2JFF.map, 2.0, site, carve=1.6«. S tema dvema ukazoma smo najprej izbrali področje prikaza (»site« se nanaša na celotni protein), prikazali mapo okoli

proteina in ustvarili gumb »map«. V oknu z gumbi na desni strani kliknemo na »2JFF.map.mesh« in s tem skrijemo večji del mape, ostane nam še mapa okoli proteina:



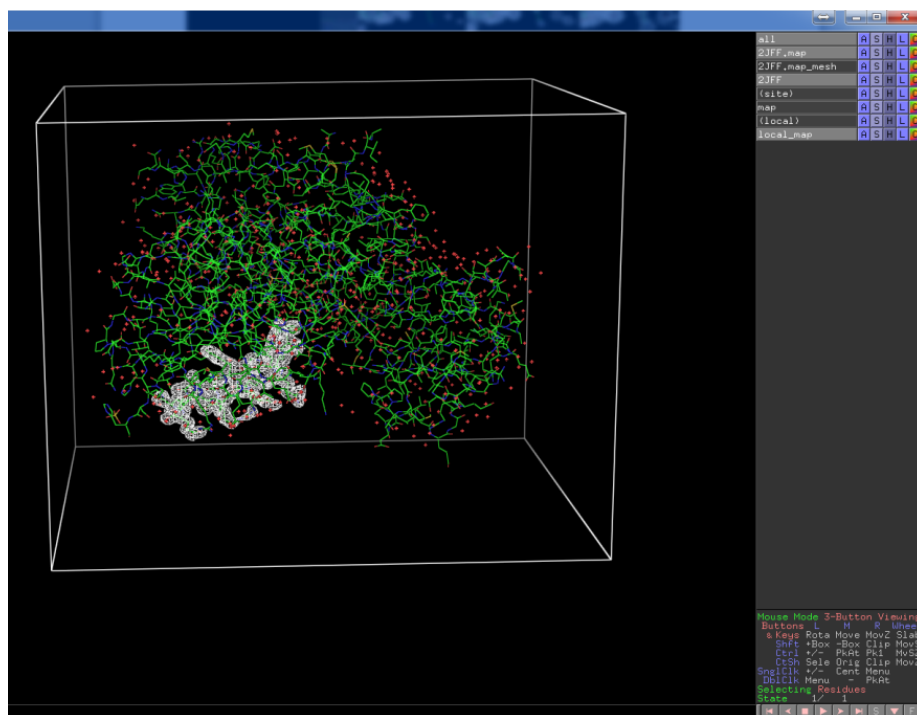
Na sliki je označen lizin 420 - Lys420.

Z ukazom zraven gumba »map« »A/level/level« spremenimo nivoje gostote od 1,5 do 0,5 in približamo aminokislino lizin 420. S tem prikažemo različne mape elektronske gostote v odvisnosti od nivoja mape:

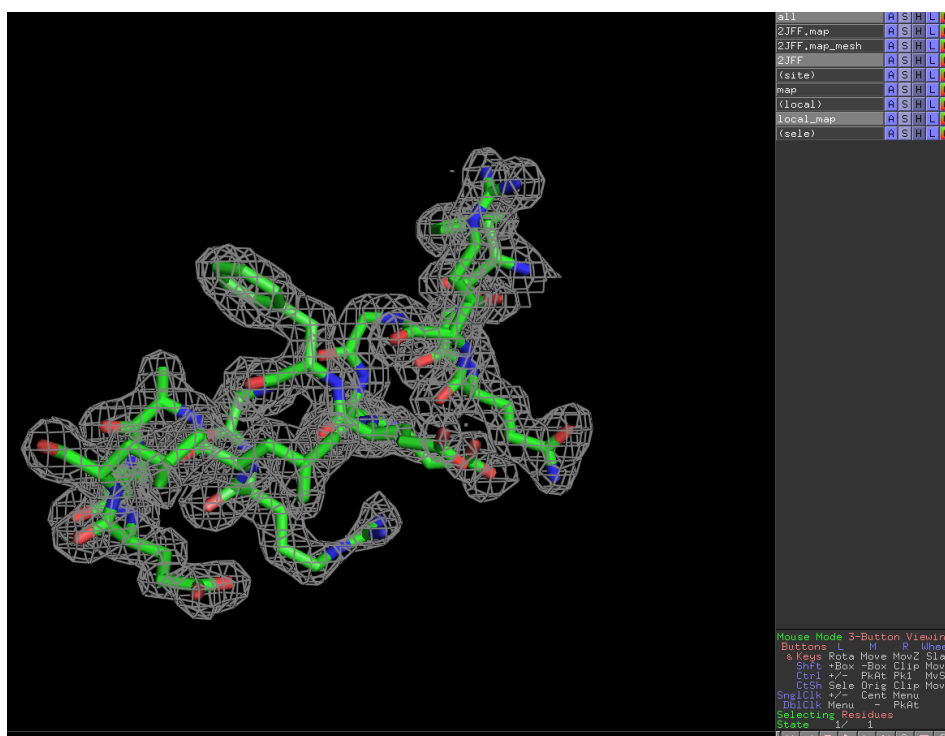


4. PRIKAZ LOKALNE POVRŠINE

Nivo mape nastavimo na 0.5 z zraven gumba »map« »A/level/level 0.5«. Izberemo segment α -vijačnice z ukazom v ukazni vrstici »select local, resi 422-436« in vtipkamo še »isomesh local_map, 2JFF.map, 1.0, local, carve=1.6«. Ustvarita se dva nova gumba »local« in »local_map«. S klikom na gumb »map« skrijemo map in ostane le elektronska mapa okrog izbranega segmenta α -vijačnice,



V nadaljevanju najprej skrijemo žičnati prikaz proteina s klikom na »H/everything« zraven gumba »2JFF«. Nato zraven gumba »local« kliknemo na »S/sticks« in povečamo izbrani segment. Za razločnejši prikaz obarvamo mapo elektronske gostote v siv odtenek z ukazom v ukazni vrstici »color gray50, local_map« in dobimo nekaj podobnega kar prikazuje spodnja slika:



Andrej Perdih in Alja Prah

vaja 3: vizualizacija bioloških molekul in prilagajanje konformacij makromolekul

```
function quicksort(array)
var list less, equal,
if length(array) ≤ 1
return array
select a pivot value pivot
for each x in array
if x < pivot then
if x = pivot then
if x > pivot then
return concatenate(quickso
```

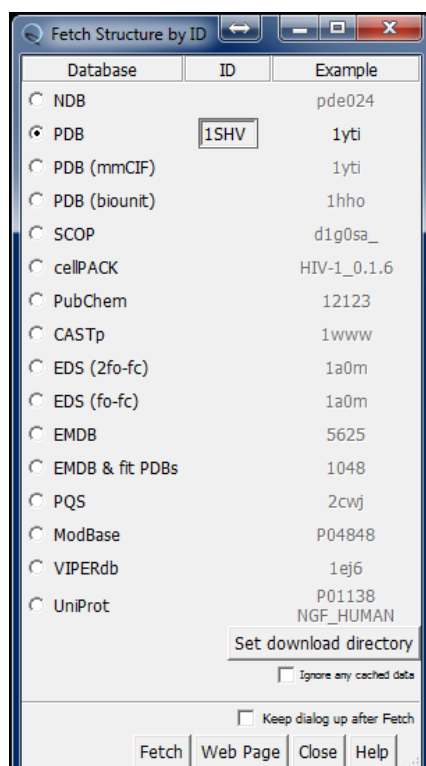
V tej vaji si bomo na praktičnem primeru proteina beta-laktamaze (β -laktamaze) ogledali na kakšne načine lahko vizualiziramo 3D proteinske strukture. β -laktamaze so bakterijski encimi, ki omogočajo bakterijam odpornost pred β -laktamskimi antibiotiki kot so penicilini, cefalosporini in karbapenemi. Delujejo tako, da razgradijo β -laktamsko strukturo antibiotikov. V procesu hidrolize razcepijo štiričlenski β -laktamski obroč in tako izničijo antibakterijske lastnosti antibiotika. Kristalna struktura *1SHV* proteina, ki jo bomo uporabili kot modelni sistem v pričujočih navodilih, je prosto dostopna na spletni strani PDB in vključuje kokristaliziran ligand, na katerem bomo prikazali vizualizacijo medmolekulskih interakcij. Na praktičnem primeru si bomo ogledali tudi prileganje različnih 3D struktur proteinov. Vizualizacijo in prileganje makromolekule bomo predstavili v programu **Chimera**.

1. NALAGANJE PDB STRUKTURNE DATOTEKE *1SHV* v Chimero

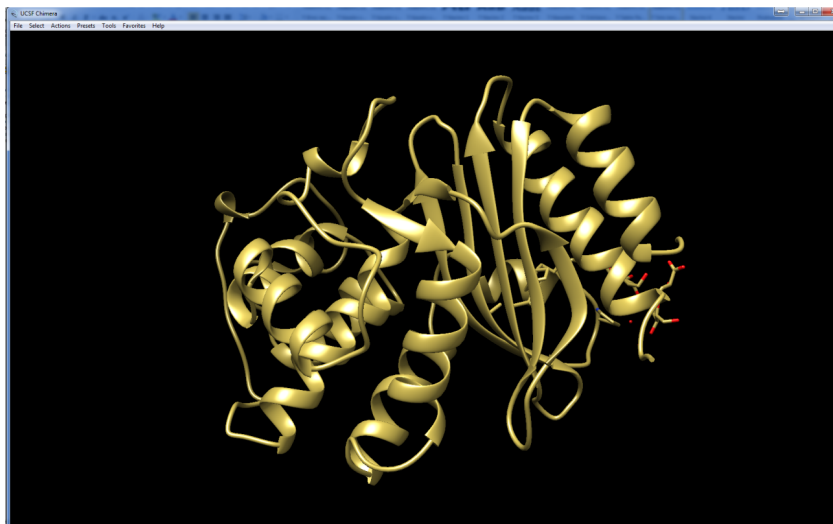
Najprej naložimo program **Chimera** iz spletnega naslova:

<http://www.cgl.ucsf.edu/chimera/download.html>.

Nato v meniju Chimere kliknemo na »**File/Fetch by ID**«. V nadaljevanju izberemo PDB podatkovno bazo, vnesemo kodo »*1SHV*« in kliknemo »Fetch«. Program bo iz podatkovne baze PDB avtomatsko naložil zeleno strukturo. Na zaslonu se bo prikazal spodnji izborni meni.



In nato vidimo naloženo strukturo v Chimeri kot to prikazuje spodnja slika.

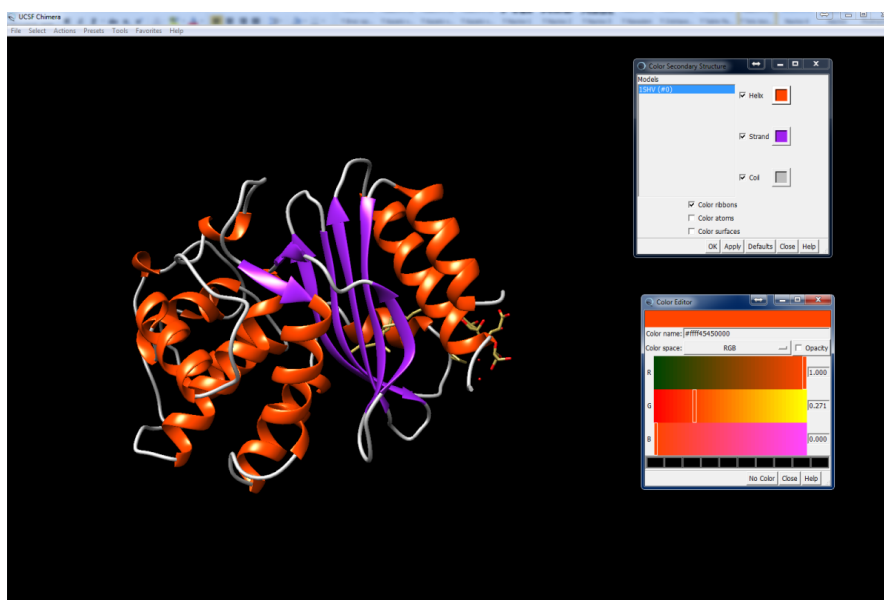


2. PREMIKANJE IN PRIBLIŽEVANJE STRUKTURE

Protein lahko vrtimo s klikom na levo miškino tipko in povlekom kazalca nad strukturo. Na podoben način z uporabo srednjega miškinega gumba dosežemo translacijsko premikanje. S potegom desne miškine tipke lahko približujemo/oddaljujemo strukturo.

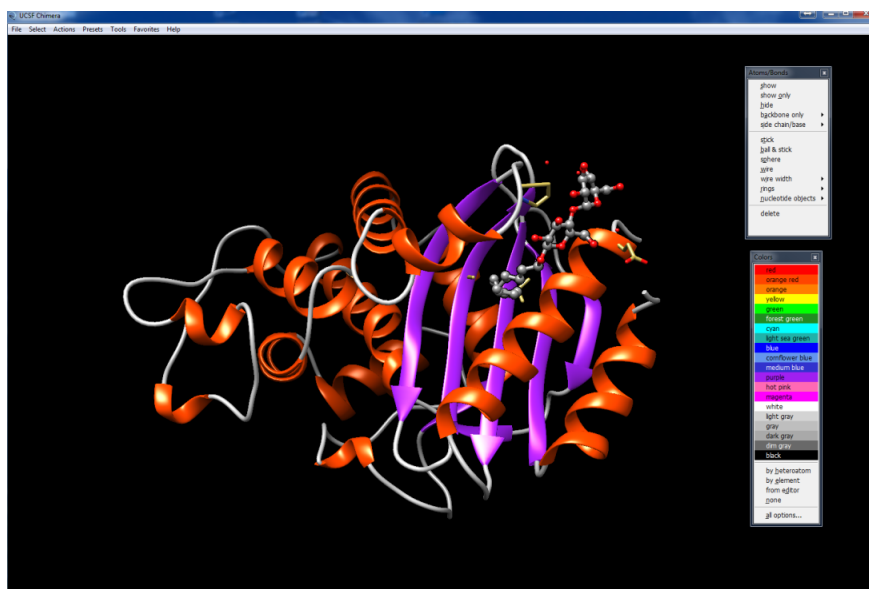
3. PRIKAZOVANJE ELEMENTOV SEKUNDARNE STRUKTURE

Struktura proteina β -laktamaze je prikazana v obliki trakov. Trakove lahko obarvamo glede na elemente sekundarne strukture, kot so vijavnice, ploskve, zanke. Odpremo modelno ploščo z izbiro menija »Tools/Depiction/Color Secondary Structure«. Odprlo se je okno »Color Secondary Structure« v katerem so s kljukico označeni »Helix«, »Strand« and »Coil«. Za vsako izmed njih lahko poljubno spreminjamo barvni prikaz v »Color Editor« oknu. Ko izberemo zelene barve, kliknemo »Apply«. Obarvane sekundarne strukture bi morali videti nekako takole, kot je prikazano na spodnji sliki.

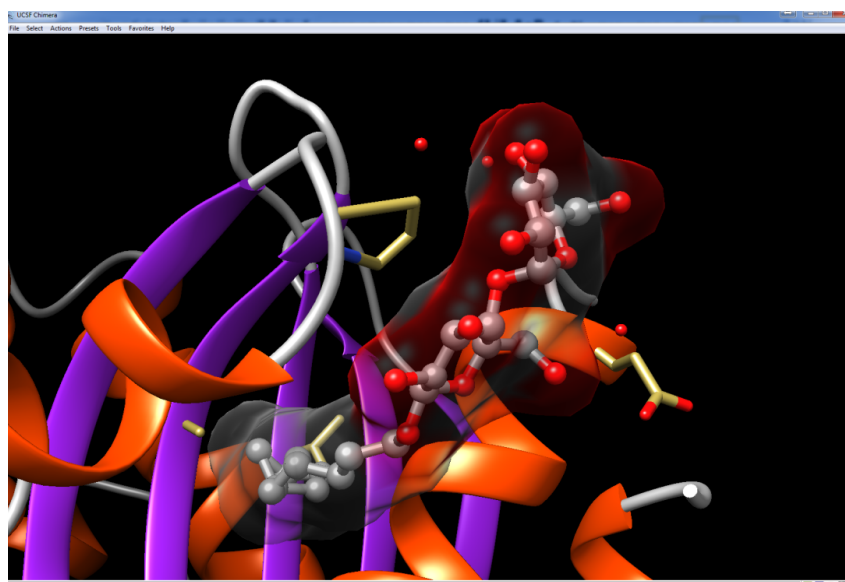


4. SPREMINJANJE PRIKAZA VEZI IN ATOMOV LIGANDA

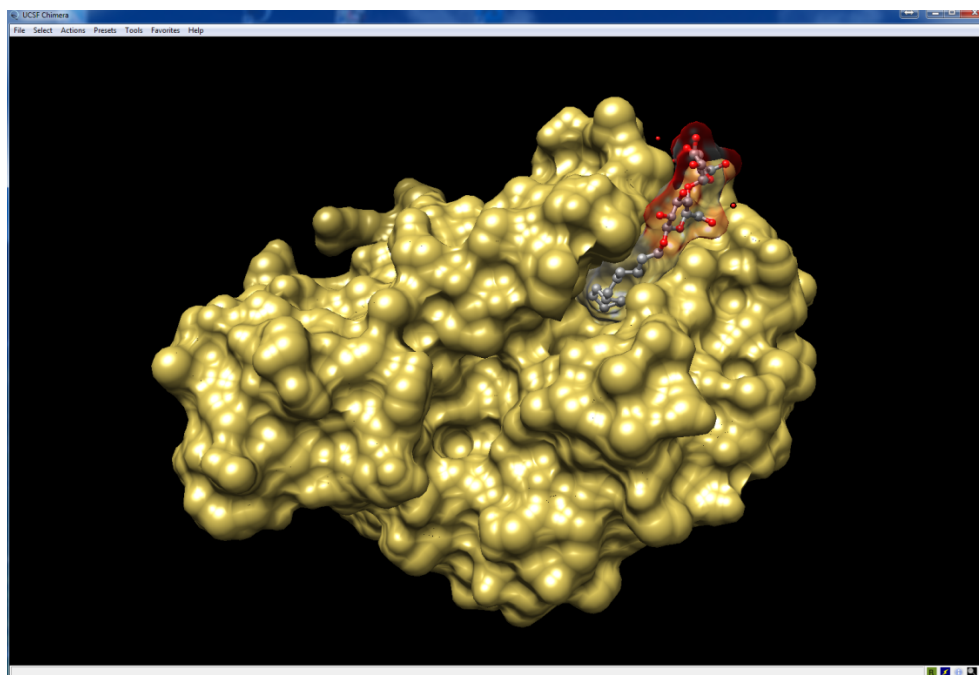
Najprej z levo miškino tipko in potegom zavrtimo molekulo tako, da se bo ligand najbolj videl, predstavljen s palčkami («sticks»). Okoli liganda lahko vidimo tri molekule vode, predstavljene z rdečimi kroglicami. Nato z ukazom v meniju »**Select/Residue/MA4**« izberemo ligand. Za spreminjanje vrste prikaza vezi v meniju odpremo »**Actions/Atoms/Bonds**« in odklopimo od glavnega menija s klikom za črtkano črto. V tem oknu lahko izbiramo med različnim prikazom vezi in atomov, a tokrat bomo izbrali prikaz kroglic in palčk s klikom na »**ball & stick**«. Če želimo spremeniti barvo liganda, izberemo v meniju »**Actions/Color**« in ponovno kliknemo na črtkano črto, da dobimo pojavno okno. Izberemo obarvanje atomov »**by element**«. Grafično okno bi moralo izgledati nekako takole, kot je prikazano spodaj.



V naslednjem koraku bomo prikazali prosojno površino liganda, ki omogoča vpogled v prostor, ki ga zavzema ligand, in posledično prileganje liganda v vezavno mesto. V meniju izberemo »**Action/Surface**« in kliknemo »**Show**«. Prosojnost površine nastavimo z ukazom »**Action/Surface/Transparency**« in izberemo 80 %. Ko približamo ligand, se pojavi situacija na spodnji sliki.

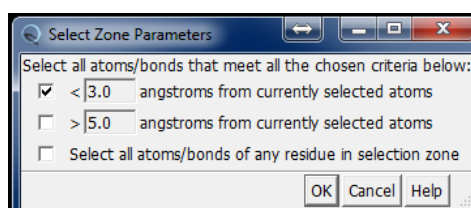


Prikazali bomo tudi površino proteina. Najprej vse odznačimo z ukazom v meniju »**Select/Clear Selection**« in z ukazom »**Select/Chain/A**« izberemo samo protein. V meniju »**Action/Surface/Show**« proteinu dodamo površino. Morali bi dobiti nekaj podobnega spodnji sliki.

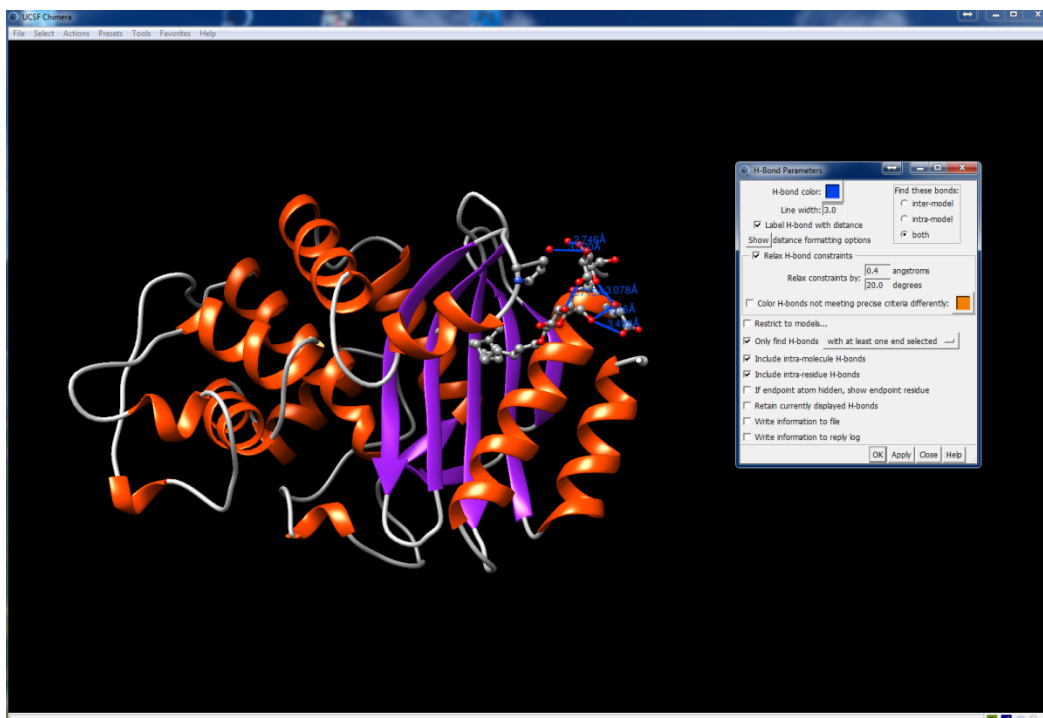


5. VIZUALIZACIJA INTERAKCIJ MED PROTEINOM IN LIGANDOM: VODIKOVE VEZI

V programu Chimera lahko tudi prikažemo vodikove vezi med izbranimi atomi. Z ukazom v meniju »**Select/Residue/MA4**« najprej izberemo ligand in nato še vezavno mesto v okolici liganda »**Select/Zone**«.



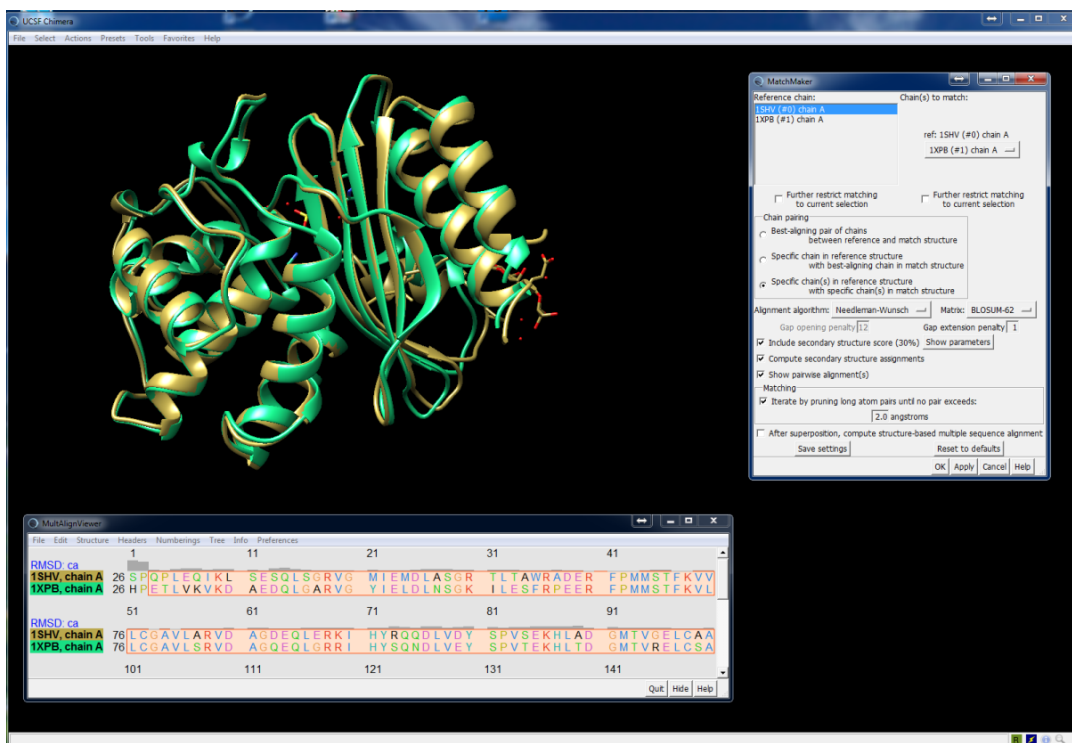
V odrtem pojavnem oknu nastavimo okolico okoli izbranega liganda na 3Å in kliknemo »OK«. Nato v meniju izberemo »**Tools/Structure Analysis/FindHBond**«. Obkljukamo »**Only find H-bonds with**«, da bo program prikazal samo vodikove vezi liganda z njegovo predizbrano okolico. Odkljukamo še »**Label H-bonds with distances**« da bo program izpisal dolžino vodikovih vezi in kliknemo »**Apply**«. Dolžina vodikove vezi je razdalja med donorjem in akceptorjem protona. Dobimo situacijo na spodnji sliki.



6. PRILEGANJE 3D STRUKTUR

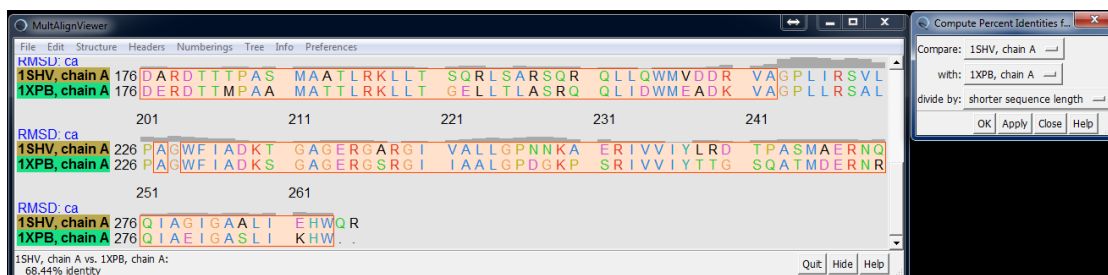
S programom Chimera lahko tudi prilegamo in primerjamo 3D strukture proteinov. Strukturno bomo primerjali dva podtipa β -laktamaz, in sicer že obravnavani tip SHV (PDB koda *1SHV*) in tip TEM (PDB koda *1XPB*).

Najprej z ukazom »File/Close Session« zapremo vse odprte strukture. Nato v meniju Chimere izberemo »File/Fetch by ID«, kliknemo PDB podatkovno bazo in vnesemo kodo »*1SHV*« ter kliknemo »Fetch«. Isti postopek ponovimo še enkrat, samo da tokrat vnesemo kot PDB kodo *1XPB*. Ko imamo obe strukturi naloženi v grafično okno v meniju, izberemo »Tools/Structure Comparison/MatchMaker«. Nato označimo »Specific chain(s) in reference structure with specific chain(s) in match structure« in izberemo pod »Reference chain« *1SHV* in pod »Chain(s) to match« *1XPB*, odključamo »Show pairwise alignment(s)« ter na koncu kliknemo »Apply«. Dobiti bi morali podobno situacijo kot na sliki spodaj.



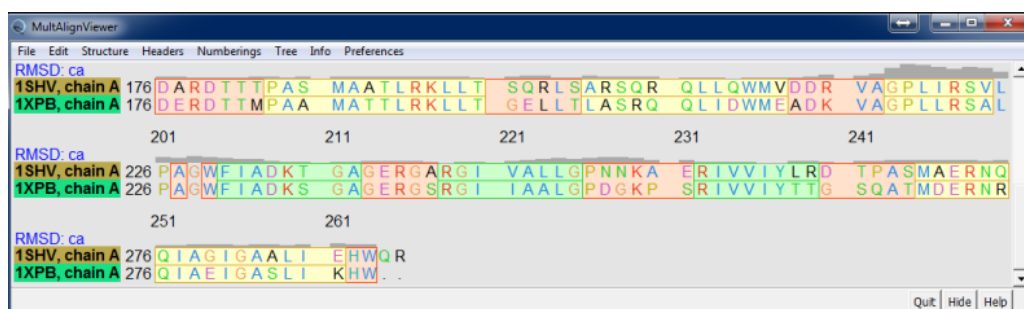
Poleg poravnanih struktur *1SHV* in *1XPB* se je odprlo še pojavno okno »MultiAlignView«, ki prikazuje poravnavo sekvenc proteinov.

Iz poravnave zaporedij lahko dobimo tudi delež enakih delov zaporedij obeh proteinov, in sicer z ukazom »Info/Percent identity...« v pojavnem oknu »MultiAlignView«. Pod razdelkom »Compare« označimo *1SHV* in pod »with« *1XPB* ter kliknemo »apply« in »OK«. Rezultat se nam izpiše v spodnji, komandni vrstici.

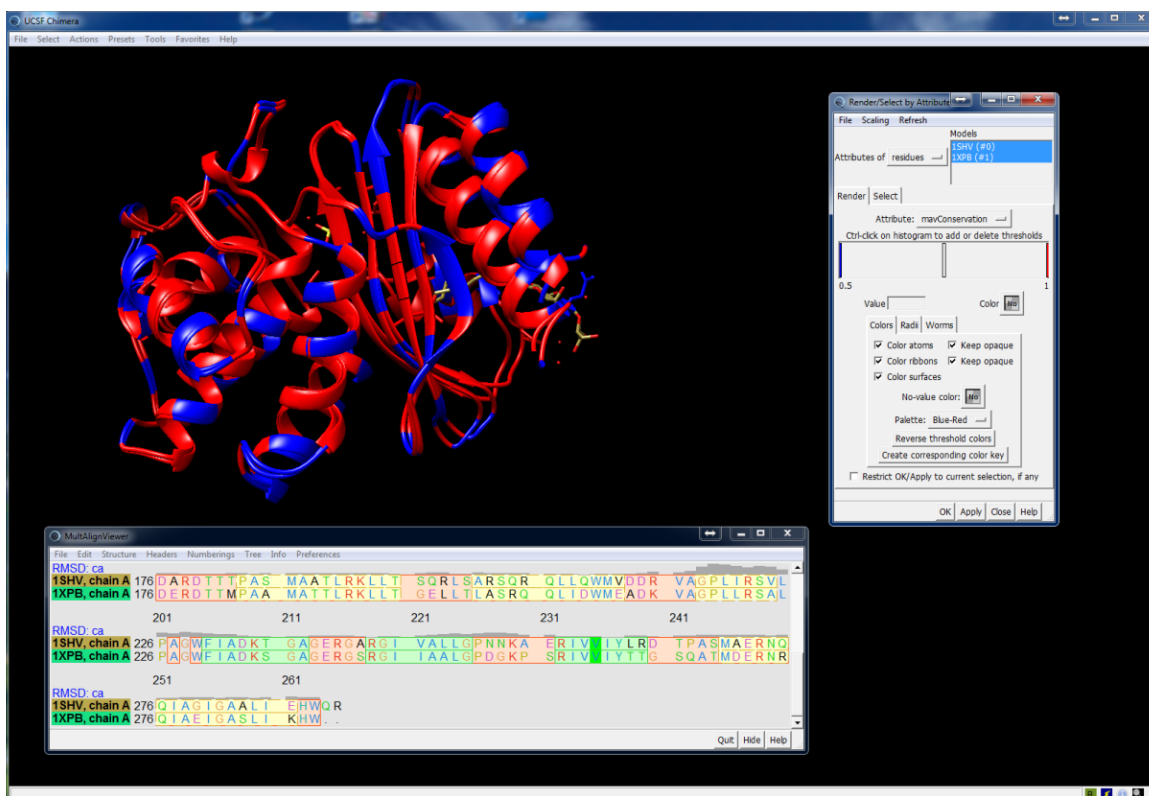


V našem primeru je identičnost zaporedij med obema proteinoma 68.44 %.

Elemente sekundarne strukture lahko pobarvamo tudi glede na njihov tip z izbiro ukaza »Structure/Secondary Structure/Show actual« v pojavnem oknu »MultiAlignView«. Amino kisline β -ploskev se obarvajo zeleno, α -vijačnic pa rumeno:



Na koncu bomo na strukturi proteinov prikazali še delež ohranjenosti njihovih zaporedji. V pojavnem oknu »MultiAlignView« v meniju izberemo »Structure/Render by Conservation« in kliknemo »Apply«.



V našem primeru so se bolj ohranjene regije pobarvale z rdečo, manj ohranjene pa z modro barvo.

Alja Prah

vaja 4: homologo modeliranje 3d strukture neznanega proteina

```
function quicksort(array)
var list less, equal,
if length(array) ≤ 1
return array
select a pivot value pivot
for each x in array
if x < pivot then less
if x = pivot then equal
if x > pivot then greater
return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

V kolikor eksperimentalna struktura proteina ni znana, se lahko za 3D vizualizacijo proteinskega zaporedja poslužimo tehnike, ki jo imenujemo *homologno modeliranje*. Homologno modeliranje nam omogoča, da na osnovi 1) aminokislinskega zaporedja proteina in na osnovi 2) že znane eksperimentalno določene strukture njemu sorodnega *homolognega* proteina, ki ga uporabimo za predlogo, zgradimo 3D model našega tarčnega proteina. Homologno modeliranje torej temelji na predpostavki, da se strukture proteinov s sorodnim aminokislinskim zaporedjem ne bodo pomembno razlikovale. Kvaliteta zgrajenega modela pa je v veliki meri odvisna predvsem od stopnje sorodnosti med našim tarčnim proteinom in tistim, ki ga uporabimo kot predlogo.

Proces homolognega modeliranja poteka v več korakih (Poglavje 5):

1. Iskanje aminokislinskega zaporedja našega tarčnega proteina, za katerega želimo zgraditi model
2. Iskanje in identifikacija sorodnih struktur (predlog oz. matrik)
3. Poravnava tarčnega zaporedja z zaporedjem predloge
4. Izdelava modela (dodajanje stranskih verig, loop-ov)
5. Optimizacija modela
6. Končni 3D model
7. Evaluacija 3D modela

Predstavlajte si, da je vaš kolega v okviru eksperimentalnega dela, ki ga izvaja za diplomsko nalogo, pridobil sledeče aminokislinsko zaporedje neznanega proteina. Rad bi ga podrobneje proučil, vendar eksperimentalno pridobivanje strukture trenutno ni izvedljivo. Zato se je obrnil na vas in na vaše široko znanje bioinformatike. Bi mu lahko pomagali in s pomočjo homolognega modeliranja zgradili 3D model tega proteina?

Zaporedje vašega tarčnega proteina je sledeče:

>Neznani protein:

```
MENQEKASIAGHMFDDVVINGGISGLSAAKLLTEYGVSVLVLEARDRVGGRTYAIRNEHV  
DYVDVGGAYVGPQTQNRILRLSKELGIETYKVVNSERLVQTVKGTKYPFRGAFPPVWNP  
IAYLDYNNLWRTIDNMGKEIPTDAPWEAQHADKWDKMTMKELIEKICWTKTARRFAYLFVNI  
NVTSEPEHVSALWFLWYVKQCGGTTTLFVSVTNGGQERKQVGGSGQVSERIMDLLGDQVKL  
NHPVTHVDQSSDNIIEALNHEHYECKYVINAIPTTLTAKIHFRPELPAERNQLIQRLPM  
GAVIKCMMYYKEAAWKKDYCGCMIIEDEDAPISITLDDTKPDGSLPAIMGFILARKADR  
LAKLHKEIRKKKICELYAKVLGSQEQALHPIHYEKNWCEEQYSGGCYTAYFPPGIMTQYG  
RVIRQPVGRIFFAGTETATKWSGYMEGAVEAGEKAAREVLNGLGKVTEKDIWVQEPESKD  
VPAVEITHTFHERNLPSVSGLLKIIGFSTSVTALGFVLYKYKLLPRS
```

Prva vrstica ponavadi služi podajanju informacij o zaporedju. Preostanek predstavlja aminokislinsko zaporedje v enočrkovnem zapisu. Slovar enočrkovnega zapisa je sledeč:

A alanin	G glicin	I izolevcin	L levcin	P prolin
V valin	F fenilalanin	W triptofan	Y tirozin	D asparaginska kislina
E glutaminska kislina	R arginin	H histidin	K lizin	S serin
T treonin	C cistein	M metionin	N asparagin	Q glutamin

Najprej moramo poiskati primerno predlogo z znano 3D strukturo, ki se čimbolj ujema z našim neznanim proteinom, na podlagi katere bomo lahko zgradili naš model. Eno izmed najbolj popularnih orodij, ki nam omogoča primerjavo aminokislinskih zaporedij, je že omenjeni BLAST (*Basic Local Alignment Search Tool*). Pojdite na BLAST spletno stran (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) in kliknite na Protein BLAST.

V zgornji kvadrantek prilepite zaporedje vašega neznanega proteina. Ker želimo naše zaporedje primerjati le z zaporedji, ki že imajo znano 3D strukturo, pod razdelkom Database izberite možnost Protein Data Bank proteins (pdb). Protein Data Bank je namreč največja baza eksperimentalno določenih proteinskih struktur. Ostalo lahko ostane nespremenjeno. Pritisnite tipko BLAST.

Po nekaj trenutkih boste dobili seznam proteinov, ki se z vašim najboljše ujemajo - vsak je identificiran s svojo 4-znakovno PDB kodo. Razvrščeni so po E (expectation) vrednosti. Bliže kot je ta vrednost 0, večja je stopnja ujemanja med vašim tarčnim proteinom in posameznim zadetkom. Kot omenjeno, je kvaliteta zgrajenega modela izjemno odvisna od stopnje ujemanja med tarčnim zaporedjem in predlogo, zato moramo pri izbiri primerne predloge biti izjemno pazljivi. Opazili boste, da se vaše tarčno zaporedje zelo dobro ujema s kar nekaj najdenimi zaporedji, zato bo v tem primeru uporaba homolognega modeliranja neproblematična.

V praksi je tako visoka stopnja ujemanja redka. Praviloma za izgradnjo modela naj ne bi uporabljali predlog, ki imajo E vrednost večjo od 1 oz. predlog, ki imajo manj kot 30 % stopnje ujemanja z našim tarčnim proteinom.

Poleg tega moramo biti pozorni še na sledeče:

- *Ali je bila struktura določena s proteinsko kristalografijo? Strukture določene z NMR spektroskopijo ali s krioelektronsko mikroskopijo imajo namreč ponavadi znatno manjšo resolucijo.*
- *So v izbrani strukturi določene pozicije vseh atomov?*
- *Ali ima izbrana struktura najboljšo resolucijo med najdenimi zadetki? Dobra resolucija za membranske proteine je okoli 2 Å ali manj, za ostale pa okoli 1 Å ali manj.*
- *Je struktura holoprotein (protein skupaj s kofaktorjem) ali apoprotein (protein brez kofaktorja)?*
- *Če želimo uporabljati model za sidranje zdravilnih učinkovin - ali je vezavno mesto dobro ohranjeno?*

Najbolje je, da si strukturo izbranega proteina podrobno ogledamo s kakšnim od prej omenjenih programov za vizualizacijo (npr. PyMOL, Chimera, VMD, ipd.). Hkrati je vedno, ko delamo s kakšno eksperimentalno določeno strukturo, potrebno prebrati tudi pripadajoči objavljeni raziskovalni članek, da pridobimo več informacij.

Izberite si pet zadetkov in jih poiščite v RCSB PDB bazi (<https://www.rcsb.org/>). Ne zaprite strani z BLAST rezultati.

1A14 PDB koda

COMPLEX BETWEEN NC10 ANTI-INFLUENZA VIRUS NEURAMINIDASE SINGLE CHAIN ANTIBODY WITH A 5 RESIDUE LINKER AND INFLUENZA VIRUS NEURAMINIDASE

DOI: 10.2210/pdb1A14/pdb

Classification: **COMPLEX (ANTIBODY/ANTIGEN)**

Organism(s): [Mus musculus](#), [Influenza A virus \(strain A/Tern/Australia/G70C/1975 H11N9\)](#)

Expression System: [Escherichia coli](#)

Deposited: 1997-12-21 Released: 1998-05-13

Deposition Author(s): [Malby, R.L.](#), [Mccoy, A.J.](#), [Kortt, A.A.](#), [Hudson, P.J.](#), [Colman, P.M.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.5 Å

R-Value Free: 0.270

R-Value Work: 0.200

metoda resolucija

Metric	Percentile Ranks	Value
Clashscore	22	22
Ramachandran outliers	1.0%	1.0%
Sidechain outliers	9.8%	9.8%

This is version 1.2 of the entry. See complete history.

S pomočjo pridobljenih informacij si izberite enega izmed zadetkov, na podlagi katerega boste zgradili svoj model. Shranite njegovo pdb datoteko kot **predloga.pdb**.

Structure Summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment

1A14

COMPLEX BETWEEN NC10 ANTI-INFLUENZA VIRUS NEURAMINIDASE SINGLE CHAIN ANTIBODY WITH A 5 RESIDUE LINKER AND INFLUENZA VIRUS NEURAMINIDASE

DOI: 10.2210/pdb1A14/pdb

Classification: **COMPLEX (ANTIBODY/ANTIGEN)**

Organism(s): [Mus musculus](#), [Influenza A virus \(strain A/Tern/Australia/G70C/1975 H11N9\)](#)

Expression System: [Escherichia coli](#)

Deposited: 1997-12-21 Released: 1998-05-13

Deposition Author(s): [Malby, R.L.](#), [Mccoy, A.J.](#), [Kortt, A.A.](#), [Hudson, P.J.](#), [Colman, P.M.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.5 Å

R-Value Free: 0.270

R-Value Work: 0.200

wwPDB Validation

Metric	Percentile Ranks	Value
Clashscore	22	22
Ramachandran outliers	1.0%	1.0%
Sidechain outliers	9.8%	9.8%

This is version 1.2 of the entry. See complete history.

FASTA Sequence

PDB Format

PDB Format (gz)

PDBx/mmCIF Format

PDBx/mmCIF Format (gz)

PDBML/XML Format (gz)

Biological Assembly 1

Nato se vrnite na stran z BLAST rezultati in kliknite nanj.

Descriptions							Graphic Summary		Alignments		Taxonomy	
Sequences producing significant alignments							Download		Manage Columns		Show 100	
<input checked="" type="checkbox"/> select all 89 sequences selected							GenPept		Graphics		Distance tree of results	
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession					
<input checked="" type="checkbox"/>	Chain N, N9 Neuraminidase [Influenza A virus]	915	915	100%	0.0	93.19%	1NMB_N					
<input checked="" type="checkbox"/>	Chain A, Neuraminidase [Influenza A virus (A/Shanghai/02/2013(H7N9))]	788	788	82%	0.0	96.91%	5L14_A					
<input checked="" type="checkbox"/>	Chain A, Neuraminidase [Influenza A virus]	787	787	82%	0.0	96.91%	4MWJ_A					
<input checked="" type="checkbox"/>	Chain A, Neuraminidase [Influenza A virus]	786	786	82%	0.0	96.65%	4MWL_A					
<input checked="" type="checkbox"/>	Chain A, Tetra-brachion Neuraminidase [Staphylothermus marinus]	773	773	82%	0.0	93.57%	6CRD_A					
<input checked="" type="checkbox"/>	Chain N, Influenza A Subtype N9 Neuraminidase [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	767	767	82%	0.0	93.81%	1NCA_N					
<input checked="" type="checkbox"/>	Chain N, Neuraminidase [Influenza A virus]	767	767	82%	0.0	93.81%	1A14_N	klik				
<input checked="" type="checkbox"/>	Chain A, Neuraminidase N9 [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	766	766	82%	0.0	93.81%	4NN9_A					
<input checked="" type="checkbox"/>	Chain N, Influenza A Subtype N9 Neuraminidase [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	766	766	82%	0.0	93.81%	1NCC_N					
<input checked="" type="checkbox"/>	Chain A, Neuraminidase N9 [unidentified influenza virus]	766	766	82%	0.0	93.56%	3NN9_A					

Shranite si njegovo FASTA zaporedje (klik na Download - FASTA (complete sequence) kot predloga.txt.

Download GenPept Graphics Next Previous Descriptions

Chain N, Neuraminidase [Influenza A virus]
 Sequence ID: **1A14_N** Length: **388** Number of Matches: **1**
 See 27 more title(s)

Range 1: 1 to 388 Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
767 bits(1980)	0.0	Compositional matrix adjust.	364/388(94%)	376/388(96%)	0/388(0%)
Query 83		RGFNNTKGLCTINSWHIYKDNNAVRIGENS... 142			
Sbjct 1		R FNNTKGLCTINSWHIYKDNNAVRIGE+SDV... 60			
Query 143		GKHSNGTIHDRSQYRALISWPLSPPTVYNS... 202			
Sbjct 61		GKHSNGTIHDRSQYRALISWPLSPPTVYNS... 120			
Query 203		ASAVVWYNNRRPVAEINTWARNILRTQES... 262			
Sbjct 121		ASAVIWNRRPVEINTWARNILRTQES... 180			
Query 263		KILRWESLGTAKWIEECSCYGERGTICT... 322			
Sbjct 181		KILKWEPLAGTAKHIEECSCYGERAEIT... 240			
Query 323		VLTDNPRPNDPNVGNKNDPPYGNNGNGV... 382			
Sbjct 241		VLTDNPRPNDPTVGNKNDPPYGNNGNGV... 300			
Query 383		ALTDKSKPIQGQTIIVLNTDWSYSGSF... 442			
Sbjct 301		ALTDKSKPTQGQTIIVLNTDWSYSGSF... 360			
Query 443		NSIVSMCSSTEFIQGNWPDGAKIEYFL 470			
Sbjct 361		NSIVSMCSSTEFIQGNWPDGAKIEYFL 388			

Related Information
[Structure](#) - 3D structure displays
[Identical Proteins](#) - Identical proteins to 1A14_N

BLAST vam že ponudi poravnavo tarčnega zaporedja (Query, 1. vrstica) s predlogo (Sbjct oz. Subject, 3. vrstica). Sredinska vrstica je informacija o kvaliteti ujemanja med obema zaporedjema. Izpisana črka pomeni popolno ujemanje med zaporedjema, znak + pomeni, da se aminokislini ne ujemata, vendar sta istega tipa, presledek pa pomeni, da ujemanja na tistem mestu ni.

Na podoben način si poravnavo lahko ogledamo s programom ClustalOmega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). V prazen kvadratik najprej nalepite zaporedje svojega neznanega proteina skupaj z identifikacijsko vrstico, ki se prične z >. Takoj za tem prilepite še zaporedje svoje izbrane predloge, prav tako skupaj z identifikacijsko vrstico. Pod Output Format naj bo izbran ClustalW with character counts. Kliknite na Submit.

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:

```
>Neznan protein
MENQEKASIAGHMFDDVVINGGISLSAAKLLTEYGVSVLVLEAR
...
VPAVEITHTFHERNLPSVGLLKIIIGFSTSVTALGFVLYKYKLLPRS
>Izbrano zaporedje tarčnega proteina
EVSALWFLWYVKQCGGTTTLFVSVTNGGQERKFGVGGSGQVSERI
...
NHPVTHVDQSSDNIIEALNHEHYECKYVINAIPPTLTAKIHFPEL
```

zaporedje vašega neznanega proteina in izbranega tarčnega proteina

Or, [upload a file](#): No file chosen [Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Ko se račun zaključi, kliknite na Show Colors. Podobno kot pri BLAST-u, so tukaj popolna ujemanja označena z *, z znakom : so označene aminokislino, ki se sicer ne ujemanjo popolno, pripadajo pa istemu tipu, z znakom . pa so označene aminokislino, ki so si nekoliko podobne.

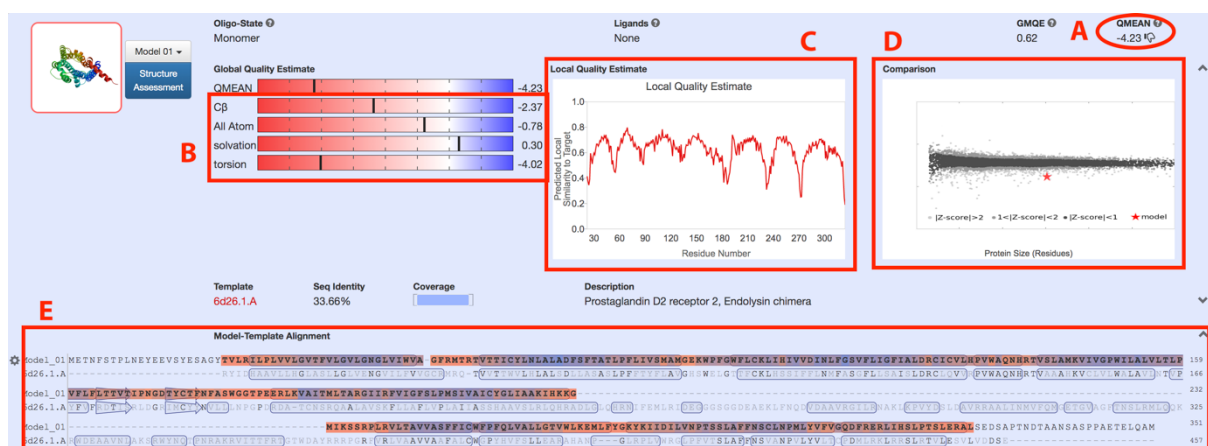
Legenda barv je sledeča:

Rdeča	MAIFVLWP	majhne in hidrofobne aminokislino
Modra	ED	kisle aminoskislino
Zelena	NQSGHTYC	aminokislino s hidroksi, sulfhidrilno, amino skupino + glicin
Roza	KR	bazične aminokislino
Siva	drugo	nenavadne aminokislino

Kliknite na Download Alignment File in skopirajte rezultat v .txt dokument z imenom **poravnava.txt**.

Sedaj se lahko lotimo izgradnje našega 3D modela proteina. Za to bi lahko uporabili veliko različnih orodij, kot so *Modeller*, *Swiss-Model*, *MOE*, *Phyre2*, *Prime*, itd. Pri vaji bomo uporabili Swiss-Model (<https://swissmodel.expasy.org/>). Pojdite na spletno stran in kliknite Start Modelling. Na desni izberite Target-Template Alignment. (Swiss-Model ima tudi možnost avtomatske izgradnje modela, samo na podlagi zaporedja tarčne strukture z enim klikom. Ta opcija bi na vašem primeru verjetno delovala, v praksi pa je pametnejša previdnost na vsakem koraku.)


V zgornji kvadratale prilepite celotno vsebino vašega dokumenta poravnava.txt. Po želji lahko vnesete tudi ime projekta in svoj email naslov. Kliknite na Build Model. Po nekaj trenutkih boste dobili vaš homologni model. Na desni strani si lahko ogledate zgrajeno 3D strukturo, na levi pa povzetek rezultatov.



QMEAN Z-vrednost (na zgornji sliki označena s črko A) predstavlja splošno oceno kvalitete modela in jo lahko podamo za cel protein ali za posamezne aminokislinske preostanke. Osnovana je na globalni primerjavi nekaterih lastnosti (zbranih na sliki *Global Quality Estimate*, označeno s črko B) eksperimentalno določenih proteinskih struktur z našim modelom. QMEAN vrednost pod -4 praviloma pomeni, da je celoten model izjemno nezanesljiv. Ponavadi pa je še bolj pomembno, kako zanesljivi so določeni predeli modela.

Graf *Local Quality Estimate* (označeno s črko C) ima na abscisi navedene aminokislinske preostanke, na ordinati pa lokalno podobnost z eksperimentalnimi strukturami. Bliže kot je ta vrednost 1, boljše je. Deli našega modela, kjer y vrednost pade po 0,6 nakazujejo regije z nizko zanesljivostjo. Tako lahko npr. kljub nezanesljivosti celotnega modela, njegovo vezavno mesto modeliramo z visoko zanesljivostjo.

Pri zadnjem grafu (*Comparison*, označeno s črko D) na x osi najdemo dolžino proteina (skupno število aminokislinskih preostankov), na y osi pa normalizirano QMEAN vrednost. Vsaka točka prikazuje eno eksperimentalno proteinsko strukturo. Najtemnejše točke so eksperimentalne strukture z normalizirano QMEAN vrednostjo znotraj ene standardne deviacije od povprečja. Naš model je označen z rdečo zvezdico.

Čisto spodaj vidimo poravnavo našega tarčnega zaporedja z izbrano predlogo (označeno s črko E). Če kliknete na majhno ikono na  levi strani, lahko zaporedje poljubno obarvate, med drugim tudi z opcijo Clustal, kot smo naredili že prej. Barvna shema QMEAN pa obarva zaporedje in tudi 3D reprezentacijo na desni strani glede na izračunano lokalno QMEAN



vrednost – področja, obarvana modro, predstavljajo regije z višjo zanesljivostjo, medtem ko so rdeče obarvana področja manj zanesljiva. Ponavadi so rdeče obarvane oz. manj zanesljivo napovedane proteinske zanke oz. loopi.


Shranite pdb datoteko vašega modela (klik na Model 01 -> PDB Format) kot **model.pdb**.



Alja Prah


vaja 5: energijska primerjava konformacij cikloheksana

```
function quicksort(array)
var list less, equal, greater
if length(array) <= 1
return array
select a pivot value pivot
for each x in array
if x < pivot then list = less
if x = pivot then list = equal
if x > pivot then list = greater
return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

V prvem koraku prenesemo in namestimo program ArgusLab, ki je dostopen na spletnem naslovu: <http://www.arguslab.com/arguslab.com/ArgusLab.html>. Program ima za osnovo molekulska mehanika in semiempirične metode molekulskih orbital. Odprite program in pritisnite ikono za *Create New Molecule* (ali Ctrl + N). Kliknite *Show the Builder Toolkit* , izberite zavihek *Rings* in poiščite cikloheksan v konformaciji stola.  Kliknite nanj in z desno tipko kliknite v okno.

Kliknite na *Set mode to Selection for right mouse button*  in potem kjerkoli v črno ozadje, da se znebite manipulatorja. Shranite molekulo kot **Naloga2a.agl**.


V meniju *Calculation* kliknite na *Optimize Geometry...* V razdelku *Hamiltonian* izberite UFF in kliknite OK. Izvedla se bo optimizacija geometrije z empiričnim poljem sil imenovanim *Universal Force Field* (če se ne zgodi nič, kliknite na *Run a calculation* ). Shranite , to bo prepisalo prejšnjo datoteko z optimizirano verzijo.

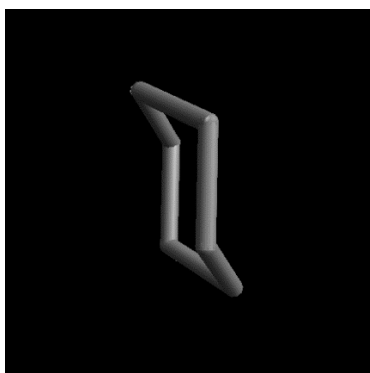
Oglejte si output s klikom na *View latest calculation results* . Zapišite končno energijo v Hartree-jih (majhna vrednost v a. u. pod razdelkom *Final Energy Evaluation*, zabeležite vsa decimalna mesta!).




Zaprte Beležnico. Sedaj shranite datoteko (*File - Save As...*) kot **Naloga2b.agl**, sledila bo optimizacija s *semiempirično kvantnomehansko metodo PM3*. V meniju *Calculation* kliknite na *Optimize Geometry...* V razdelku *Hamiltonian* izberite PM3 in kliknite OK. Shranite rezultat!

Spet si zabeležite končno energijo (tokrat je nekoliko večja, negativna vrednost v a. u., pod razdelkom *Geometry Optimization - Final Geom Energy*).

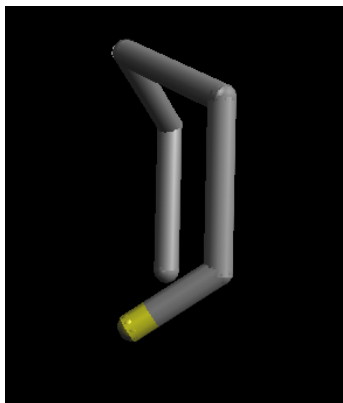
IZ KONFORMACIJE STOLA V KONFORMACIJO KADI




Odprite datoteko **Naloga2a.agl**. Shranite jo kot **Naloga2c.agl**. Odstranite vodikove atome z *Delete Hydrogens* . Prikažite molekulo kot Sticks (*Settings – Display – Cylinder = Normal, OK*). Obrnite molekulo tako kot to prikazuje spodnja slika:



V načinu *Select* , kliknite na ogljikov atom desno spodaj – moral bi se obarvati rumeno. Potem pritisnite tipko Delete, da ga odstranite. Kliknite na *Set mode to Add atoms for right mouse button* . Kliknite na C pod zavihkom Atoms (če ni C (tetrahedral) avtomatsko nastavljen). Prepričajte se, da je *Automatic bonds* nastavljen na ON!  Z levim miškinim

gumbom kliknite na enega izmed zdaj terminalnih ogljikov, obarvati se mora rumeno. Namestite kurzor na mesto, kjer mislite, da bi se moral nahajati ogljik za konformacijo kadi (kot to prikazuje spodnja slika), pridržite tipko Shift in tja kliknite z desno miškino tipko, da namestite ogljikov atom.



Spet pridržite Shift in z desno tipko kliknite na preostali terminalni ogljikov atom, da zaprete obroč. Preklopite na *Select*  način. Pritisnite tipko za dodajanje vodikov . Shranite! 

Spet naredite geometrijsko optimizacijo z UFF poljem sil, s tem da spremenite maksimalno število korakov (Maximum Steps Taken) iz 100 na 200, saj sicer verjetno ne boste dobili optimizirane strukture. Shranite rezultat!

Oglejte si dobljeno konformacijo. Tokrat ne smete videti konformacije stola, ampak nekakšno zavito ladjo, konformacijo kadi. Zabeležite z UFF metodo izračunano energijo tako dobljene konformacije v a. u. enotah.

Ponovite postopek za izračun še s semiempirično metodo PM3 in shranite rezultat pod **Naloga2d.agl** in zabeležite dobljeno energijo v a.u.

Energijsko primerjajte obe konformaciji. Katera je bolj stabilna (nižja energija tako po UFF kot po PM3)?

Pretvorite Hartree-je (a. u.) v enote kJ/mol, tako da množite s faktorjem 2625,515 in podajte razliko med energijo bolj stabilne in manj stabilne konformacije po obeh metodah:

- UFF Δ (manj stabilna – bolj stabilna) [kJ/mol] = _____
- PM3 Δ (manj stabilna – bolj stabilna) [kJ/mol] = _____

Eksperimentalno določena energijska razlika je 23 kJ/mol. Katera metoda se je tej vrednosti bolj približala?

Andrej Perdih


vaja 6: molekularno sidranje inhibitorja v aktivno mesto encima MurD s programom arguslab

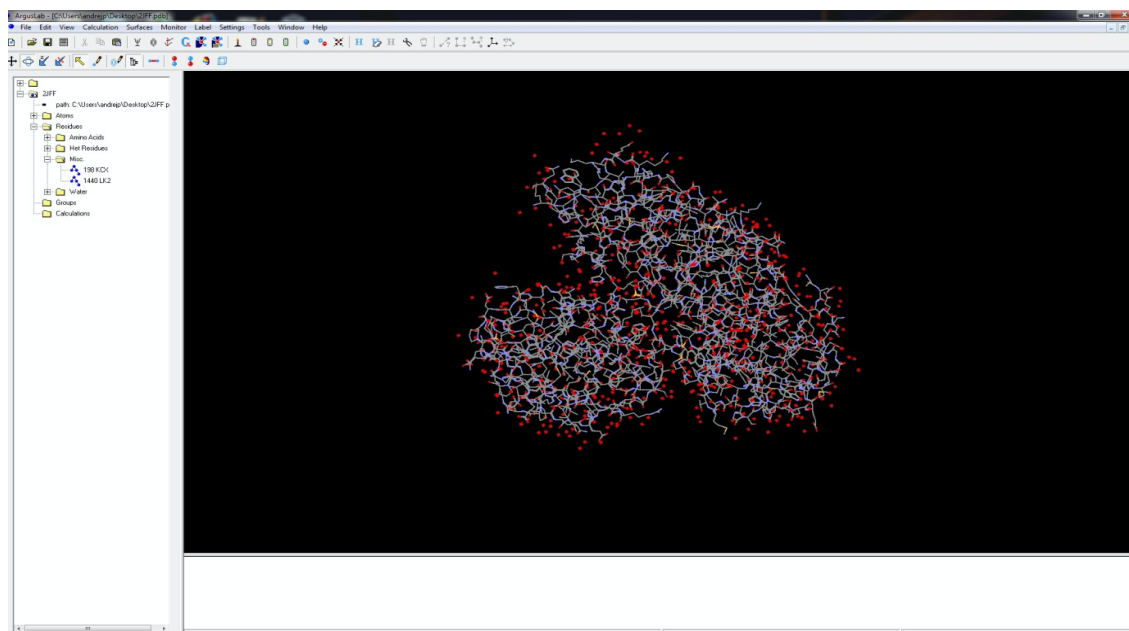
```
function quicksort(array)
var list less, equal, greater
if length(array) <= 1
return array
select a pivot value pivot
for each x in array
if x < pivot then list = append(list, x)
if x = pivot then equal = append(equal, x)
if x > pivot then list = append(list, x)
return concatenate(quicksort(list), equal, quicksort(greater))
```

V tej vaji si bomo ogledali, kako pripraviti in izvesti molekulsko sidranje. Najprej bomo izvlekli eksperimentalno določeno vezavno konformacijo D-Glu sulfonamidnega zaviralca bakterijskega encima MurD in jo sidrali v vezavno mesto MurD. Nato bomo primerjali sidrano strukturo zaviralca z eksperimentalno, da bomo ocenili njuno ujemanje. Ta postopek se pogosto uporablja pri validaciji programov za molekulsko sidranje pred samim virtualnim sidranjem večjih knjižnic spojin, kot smo obravnavali v Poglavlju IV. Preden pričnemo s sidranjem molekule, moramo najprej definirati atome liganda in vezavno mesto na proteinu, kamor se ligand veže. Kristalna struktura *2JFF* proteina, ki jo bomo uporabili kot modelni sistem in jo poznamo že iz Vaje 2, je prosto dostopna na spletni strani Protein Brookhaven Databank (<http://www.rcsb.org>). Obravnavana struktura proteina je rezultat antibakterijskega raziskovalnega projekta, pri katerem so bili udeleženi številni slovenski raziskovalci.

1. PRENOS PDB STRUKTURNE DATOTEKE Z OZNAKO *2JFF*

Nato v spletnem brskalniku v iskalnik vpišemo Protein Brookhaven Databank (<http://www.rcsb.org>) in v iskalno polje vnesemo *2JFF*. 3D strukturo proteina prenesemo kot pdb datoteko. Odpremo *2JFF.pdb* v programu ArgusLab, ki smo ga uporabljali že v prejšnji Vaji 5. To je PDB datoteka, zato poskrbimo, da pri odpiranju izberemo pravo vrsto datoteke (PDB tip).


Najprej poskrbimo, da je orodje **Molecule Tree View tool** vidno. Izberemo *Tools/ Molecule Tree View* ali kliknemo ikono  v orodni vrstici za preklon na drevesni pregled. Nato razširimo drevesni pogled *2JFF* in odpremo mapo *Residues/Misc* za prikaz sulfonamidnega zaviralca (*LK2*). Grafični vmesnik v ArgusLabu bi moral izgledati nekako takole.




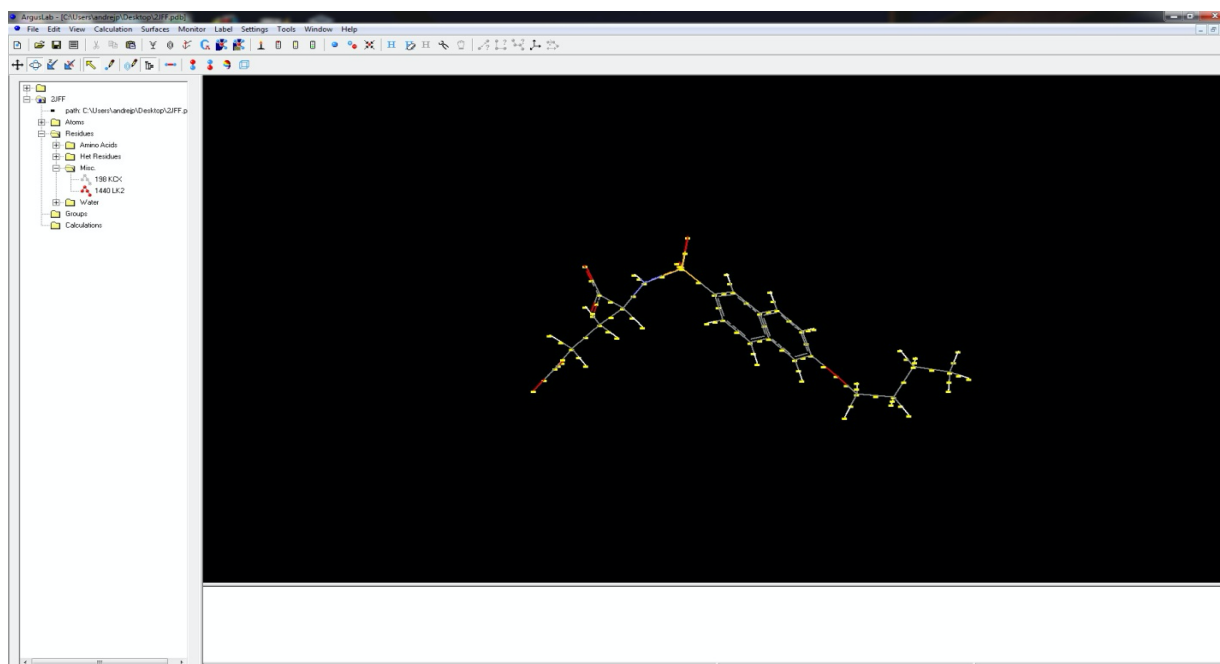
2. PRIPRAVA LIGANDA IN VEZAVNEGA MESTA

Z levo miškino tipko v drevesnem pregledu kliknemo na »1440 LK2« in izberemo sulfonamidni zaviralec tako, da se obarva rumeno.

Nato izberemo *Edit/Hide Unselected* možnost, da skrijemo vse atome, ki niso izbrani. Edini atomi, ki so prikazani na zaslonu, so atomi D-Glu sulfonamidnega zaviralca.

Sulfonamidni zaviralec v oknu centriramo z izbiro *View / Center Molecule* v meniju *Window* ali s pritiskom na gumb  v orodni vrstici.

Medtem ko imamo v drevesnem pregledu izbran zaviralec, dodamo atome vodika s pritiskom na gumb  v orodni vrstici. Potem bi morali spojino videti nekako takole.



Z desno miškino tipko kliknemo na »1440 LK2« v *Tree View* in izberemo možnost »**Make a Ligand Group from this Residue**«. Tako bo ArgusLab zgradil skupino pod mapo **Groups** z istim imenom "1 LK2", ki je tipa »Ligand«.

Z desno miškino tipko kliknemo na »1440 LK2« v *Residues/Misc folder*. Tako bomo spet izbrali atome liganda na zaslonu.

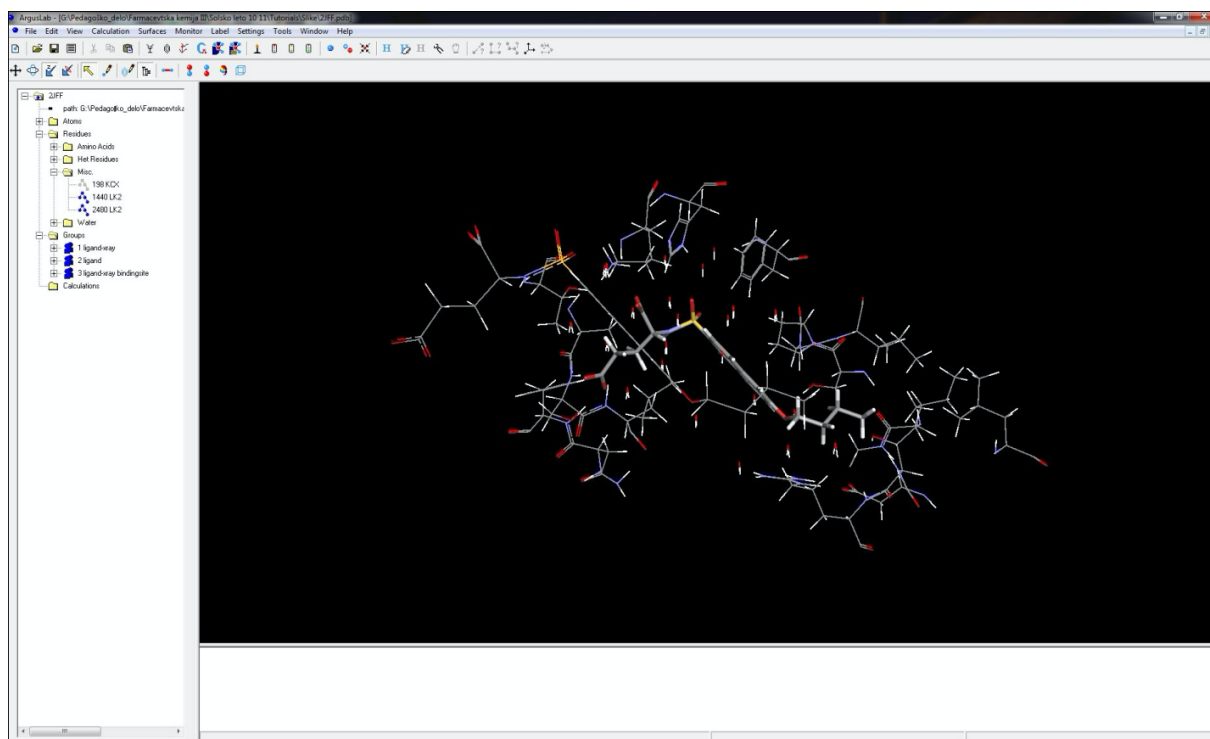
V naslednjih dveh korakih kopiramo in prilepimo izbrani ligand (Copy/Paste ukaza). Če pogledamo v mapo *Residues/Misc* v drevesnem pregledu, bomo videli nov označen ligand z imenom »2480 LK22« (možno je tudi različno število oznake).

Na enak način kot zgoraj naredimo novo skupino z ligandom. Z desno miškino tipko kliknemo na »2480 LK2« in izberemo »**Make a Ligand Group from this Residue**«. Sedaj bi morali imeti dve vrsti ligandov v mapi skupine z imenoma "1 LK2" in "2 LK2".

V nadaljevanju bomo preimenovali skupino ligandov: z desno miškino tipko kliknemo na »1 LK2« v mapi *Group* in izberemo možnost »**Modify Group**«. V polju *Modify Group* vnesemo novo ime. Imenujmo jo »ligand-eksperimentalni«. Pazimo, da ne spremenimo vrste mape *Group*. Enako naredimo za »2 LK2«, poimenujemo ga »ligand«.


Za lažje razlikovanje obeh ligandov na grafičnem vmesniku si lahko pomagamo na več načinov. Lahko npr. spremenimo vizualizacijo enega od ligandov tako, da z desno miškino tipko kliknemo na skupino z imenom »ligand-eksperimentalni« in izberemo »**Set Render Mode**« ter nastavimo možnost »**Cylinder med**«.

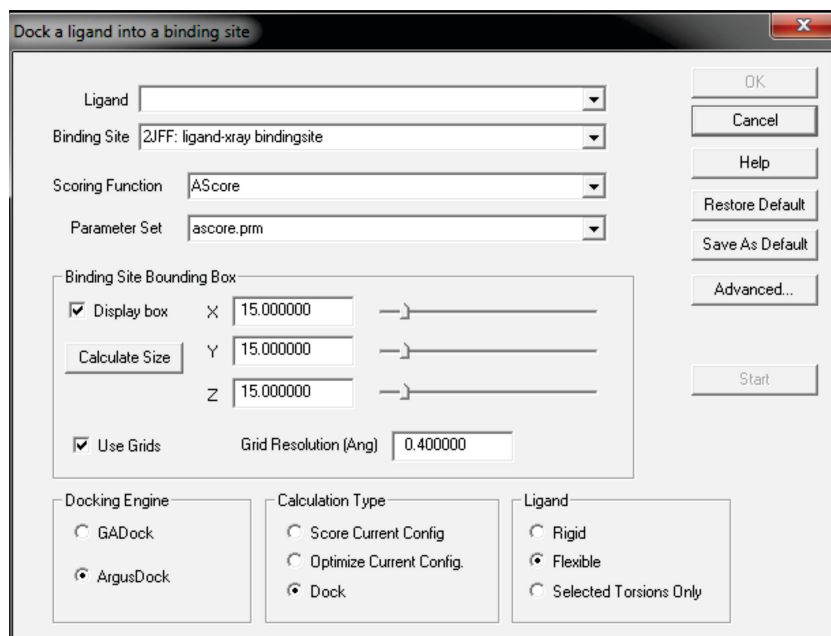
V naslednjem koraku bomo definirali vezavno mesto. Z desno miškino tipko kliknemo na skupino »ligand-eksperimentalni« v skupini *Groups* in nato izberemo možnost »**Make a BindingSite Group for this Group**«. Tako bomo ustvarili vezavno mesto »*BindingSite*«, ki je sestavljeno iz vseh aminokislinskih ostankov, ki imajo vsaj en atom znotraj polmera 3,5 Å od kateregakoli atoma liganda iz skupine »ligand-eksperimentalni«. To nam običajno omogoči dober prikaz pomembnih aminokislinskih ostankov v vezavnem mestu tarčnega proteina. Po centriranju molekule dobimo stanje, ki je prikazano na spodnji sliki.



3. SIDRANJE LIGANDA V VEZAVNO MESTO MurD

Ko bomo sidrali ligand iz skupine »Ligand«, bomo avtomatično zanemarili vse druge ligande/atome, ki so še prisotni v definiranem aktivnem mestu. Opozorimo samo, da so v primeru, če pozabimo ustvariti skupino »Ligand« iz molekule, ki je že v vezavnem mestu, rezultati sidranja neveljavni.

Najprej odpremo okno z nastavitvami za sidranje »*Dock Settings*« z izbiro možnosti »*Calculation/Dock a Ligand...*« ali s klikom na ikono  v orodni vrstici. Tako izgleda okno z nastavitvami za sidranje:



V razdelku »*Ligand*« izberemo ligand, ki ga bomo sidrali. Pazimo, da izberemo skupino »ligand« in ne skupino »ligand-eksperimentalni«.

S klikom na gumb »**Calculate Size**« program izračuna in prikaže prostorske koordinate »škatle« okoli vezavnega mesta. Kot program za sidranje nastavimo »**ArgusDock**«, kot tip Izračun »**Dock**« in fleksibilen ligand »**Flexible**«.

Nato s klikom na gumb »**Start**« pričnemo z molekulskim sidranjem.

V vrstici stanja se na dnu grafičnega okna začnejo pojavljati obvestila. V prvem obvestilu lahko vidimo, da ArgusLab ustvari točkovno mrežo, ki jo program uporabi med sidranjem. Nato se začnejo izpisovati različne faze iskanja konformacij, ki mu sledi obdelava možnih vezavnih geometrij liganda, s čimer se zaključi proces sidranja. Pri tem se spreminjajo smo koordinate liganda iz skupine »ligand«.

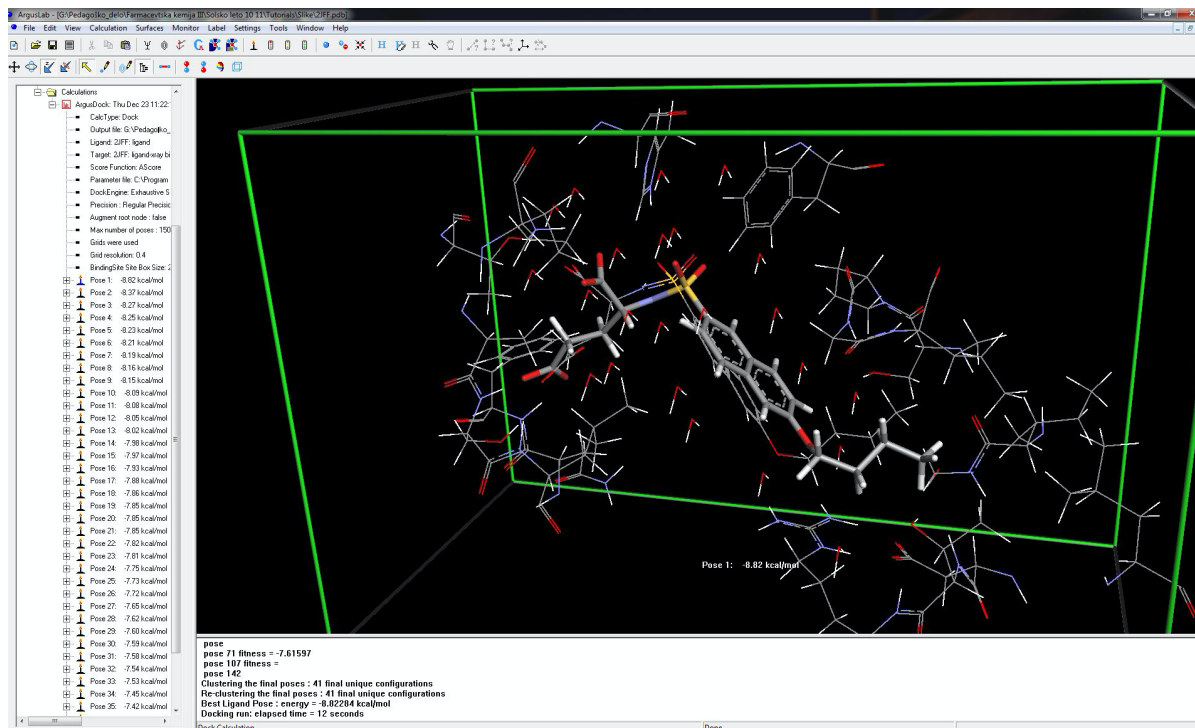
4. ANALIZA REZULTATOV

V zadnjem koraku bomo analizirali ujemanje konformacije sidranega liganda, z eksperimentalno določeno vezavno pozo liganda. V drevesnem pregledu izberemo tako sidrani ligand kakor tudi »eksperimentalni« ligand tako, da hkrati držimo tipko *Ctrl* in levo miškino tipko. Obe skupini se obarvata rumeno. Nato z desno miškino tipko kliknemo na »*Groups*« mapo v drevesnem pregledu in izberemo »*Calc RMSD position between two similar Groups*«. Odpre se pogovorno okno, ki nam prikaže koren povprečja kvadratov odklonov (»root mean square deviation« - RMSD) med obema ligandoma.

Da bi lahko zaključili, da sta eksperimentalna in izračunana konformacija liganda primerljive vrednosti, mora biti RMSD vrednost praviloma manjša od 2,5 Å. Kako je bilo v vašem primeru?

Podatke o izračunih sidranja lahko pridobimo na dveh mestih: »log« datoteka in orodje drevesnega pregleda. Najprej bomo razširili ArgusDock izračun v mapi »*Calculations*« v drevesnem pregledu. Vidimo lahko seznam nastavitev sidranja in več vezavnih poz liganda, pri

čemer je najbolj stabilna vezavna geometrija pod številko 1 (izraz "poza" se običajno uporablja za poimenovanje niza koordinat sidranega liganda). Z desno miškino tipko kliknemo na pozo 2 in izberemo možnost »Display«. S tem ukazom se bodo koordinate liganda spremenile tako, da bodo ustrezale 2. pozi in na zaslonu se bo izpisal rezultat cenilne funkcije. RMSD med pozama sidranega in eksperimentalno določenega liganda lahko določimo na enak način kot v prejšnjem primeru. Na zaslonu bi morali videti naslednjo sliko.



Z desno miškino tipko kliknemo na vezavno pozo 2 in izberemo »Hide« ali »Hide all Props«, da izklopimo prikaz 2. vezavne poze liganda. Grafični prikaz na zaslonu se povrne v osnovni prikaz poze z najnižjo energijo.

Na koncu rezultate še shranimo v ArgusLab *.agl datoteko. Izberemo možnost »File/Save As...« in določimo vrsto datoteke ArgusLab ter shranimo. Tako bodo rezultati sidranja dostopni tudi v bodoče, če oz. ko jih bomo rabili.

vprašanja za ponovitev in utrjevanje snovi

```
function quicksort(array)
  var list less, equal, greater
  if length(array) <= 1
    return array
  select a pivot value pivot
  for each x in array
    if x < pivot then list = less
    if x = pivot then list = equal
    if x > pivot then list = greater
  return concatenate(quicksort(less), quicksort(equal), quicksort(greater))
```

I. BIOLOŠKE MAKROMOLEKULE

- Iz katerih pomembnejših molekul je sestavljeno človeško telo?
- Opišite ogljikove hidrate.
- Naštejte nekatere pomembnejše polisaharide.
- Kakšna je razlika med nasičenimi in nenasičenimi maščobnimi kislinami?
- Kaj so fosfolipidi in zakaj so pomembni?
- Kaj so nukleinske kisline in kaj nukleotidi?
- Kaj so bazni pari in kaj pomeni, da so komplementarni?
- Naštejte in opišite vlogo nekaj tipov RNA.
- Kaj je osrednja dogma molekularne biologije?
- Opišite replikacijo DNA.
- Opišite transkripcijo in translacijo DNA.
- Kaj pomeni, da je genska koda degenerirana?
- Opišite proteine. Kaj je domena?
- Kaj so aminokisliline? Narišite osnovno formulo in naštejte nekaj primerov.
- Opišite primarno/sekundarno/terciarno/kvartarno strukturo proteinov.
- Naštejte nekaj pomembnejših funkcij proteinov v telesu.
- Kaj so agonisti in kaj antagonisti?
- Kaj so encimi in kako delujejo? Kako delimo inhibitorje encimov?
- Kaj so protitelesa?

II. DOLOČEVANJE 3D STRUKTURE BIOLOŠKIH MOLEKUL S KRISTALOGRAFIJO

- Opišite metodo proteinske kristalografije.
- Zakaj pri tej metodi uporabljamo rentgenske žarke?
- Opišite metodo viseče kapljice.
- Narišite shematsko difrakcijski eksperiment in aparaturo.
- Kaj je to Fouriejeva transformacija in zakaj jo uporabljamo v proteinski kristalografiji?
- Kako v kristalografiji rešujemo fazni problem?
- Kako iz elektronske gostote dobimo molekulski model proteina?
- Opišite nekatere kriterije, s katerimi opredelimo kvaliteto strukture, določene z metodo proteinske kristalografije.

III. DOLOČEVANJE 3D STRUKTURE BIOLOŠKIH MOLEKUL Z NMR SPEKTROSKOPIJO

- Opišite 1D NMR eksperiment.
- Katera jedra lahko uporabljamo pri NMR eksperimentih?
- Kako Fouriejevo transformacijo uporabljamo v NMR spektroskopiji?
- Kaj je kemijski premik in zakaj do njega pride?
- Opišite razlog za 1H-1H sklopitev jeder (primer dubleta in tripleta).
- Zakaj pri določevanju struktur proteinov potrebujemo večdimenzionalne NMR spektre?
- Katere podatke dobimo iz 2D NOESY in 2D COSY NMR spektrov?
- Kaj je jedrni Overhauserjev efekt NOE in zakaj je pomemben v NMR spektroskopiji?

- Zakaj je pomemben ^1H - ^{15}N HSQC (Heteronuclear Single-Quantum Correlation) NMR spekter?
- Na kratko opišite trojne resonančne eksperimente.
- Opišite metodo določanja 3D strukture proteinov z NMR spektroskopijo.
- Primerjajte obe metodi (rentgensko kristalografijo in NMR spektroskopijo) in izpostavite prednosti in slabosti obeh za določanje 3D struktur proteinov.

IV. MOLEKULSKO MODELIRANJE

- Definicija molekulskega modeliranja.
- Kaj je molekulska grafika in kako vizualiziramo biološke makromolekule?
- Kaj je konformacija molekule in kaj torzijski kot?
- Kako dobimo površino dostopno molekuli (Solvent accessible surface area).
- Kako primerjamo konformacije molekul in opišite RMSD parameter.
- Osnovne fizikalne enote, ki jih uporabljamo v molekulskem modeliranju (za merjenje razdalj, energije, naboja).
- Opišite molekulsko mehaniko in empirično potencialno funkcijo.
- Opišite osnovne principe kvantne mehanike in zakaj jo uporabljamo pri študiju molekul.
- Kaj je površina potencialne energije molekule in kako jo lahko preiskujemo?
- Kaj je geometrijska optimizacija molekulske strukture?
- Kaj so molekulske simulacije?
- Opišite metodo molekulske dinamike (MD).
- Kako analiziramo rezultate molekulske dinamike?
- Razložite razliko med molekulsko dinamiko in geometrijsko optimizacijo.
- Opišite kaj razumemo pod pojmom konformacijska analiza in navedite primer.

V. NAPOVEDOVANJE 3D STRUKTURE PROTEINOV Z RAČUNSKIMI METODAMI

- Zakaj potrebujemo metode za napovedovanje struktur molekul?
- Opišite metodo homolognega modeliranja.
- Opišite metodo prepoznavne zvitij.
- Kako napovedujemo strukturo proteinov iz prvih principov?

VI. UPORABA PROTEINSKIH STRUKTUR PRI NAČRTOVANJU BIOLOŠKO AKTIVNIH MOLEKUL

- Opišite cikel načrtovanja zdravilnih učinkovin
- Metoda virtualnega rešetanja (definicija, uporaba).
- Kako pripravimo kemijske knjižnice in kakšne tipe le-teh poznamo?
- Osnovi pristopi pri virtualnem rešetanju (npr. strukturno podprto načrtovanje in načrtovanje na osnovi strukture ligandov).
- Opišite metodo molekulskega sidranja.
- Kaj je to farmakofor in farmakoforno modeliranje?
- Kako validiramo metodo virtualnega rešetanja?

VII. PROBLEMSKI NALOGI

S pridobljenim znanjem o uporabi metod molekulskega modeliranja in eksperimentalnih metod določanja 3D struktur proteinov predlagajte strategijo teoretične/eksperimentalne študije, s katero bi z uporabo teh metod pridobili nove uporabne informacije za nadaljnje raziskave. Opredelite, katere podatke potrebujemo, katere metode, ki smo jih spoznali na predavanjih, bi uporabili, kakšne rezultate lahko pričakujemo ter kako jih interpretiramo.

1. Stabilnost in preiskovanje sekundarne strukture peptida

S pomočjo kemometričnih metod domnevamo, da se sekvenca 18 aminokislin zviže v strukturo β -prepognjene verige. Opišite študijo in uporabljene metode, s katero bi lahko teoretično obravnavali stabilnost take strukture, ter kako bi jo lahko kasneje tudi eksperimentalno potrdili.

2. Virtualno reševanje za identifikacijo novih potencialnih zdravilnih učinkovin

MurE ligaza je protein, ki sodeluje v sintezi bakterijskega peptidoglikana, ki je ključna komponenta bakterijske celične stene in bakteriji omogoča preživetje. Zaradi vedno večje rezistence bakterij na znane antibiotike želimo uporabiti določeno kristalno strukturo MurE ligaze za načrtovanje novih protibakterijskih zdravilnih učinkovin. Predlagajte protokol virtualnega reševanja in možne pristope, s katerim bi lahko identificirali nove biološko aktivne spojine – inhibitorje na tej protibakterijski tarči.

dodatna literatura

```
function quicksort(array)
  var list less, equal, greater
  if length(array) <= 1
    return array
  select a pivot value pivot
  for each x in array
    if x < pivot then list = less + [x]
    if x = pivot then list = equal + [x]
    if x > pivot then list = greater + [x]
  return concatenate(quicksort(list less), quicksort(list equal), quicksort(list greater))
```

I. **BIOLOŠKE MAKROMOLEKULE**

- Voet D, Voet JG. **Biochemistry**, 4th Edition, John Wiley & Sons, Inc., **2010**.
- Berg JM, Tymoczko JT, Stryer L. **Biochemistry**, 6th Edition, New York: W H Freeman; **2006**.
- Alberts et al. **Molecular Biology of the Cell**, Sixth Edition, W. W. Norton & Company, **2014**.

II. **DOLOČEVANJE 3D STRUKTURE BIOLOŠKIH MOLEKUL S KRISTALOGRAFIJO**

- Rhodes G. **Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models** 3rd Edition, Academic Press, **2006**.

III. **DOLOČEVANJE 3D STRUKTURE BIOLOŠKIH MOLEKUL Z NMR SPEKTROSKOPIJO**

- Cavanagh J, Skelton NJ, Fairbrother WJ, Rance M, Palmer AG **Protein NMR Spectroscopy: Principles and Practice**. 2th Edition, Academic Press, **2010**.

IV. **MOLEKULSKO MODELIRANJE**

- Leach, A. R. **Molecular Modelling: Principles and Applications**, 2th Edition, Pearson **2001**.

V. **NAPOVEDOVANJE 3D STRUKTURE PROTEINOV Z RAČUNSKIMI METODAMI**

- Cavasotto CN, Phatak SS. **Homology modeling in drug discovery: current trends and applications**. *Drug Discov Today*. **2009**:676-683.
- Friesner RA, Abel R, Goldfeld DA, Miller EB, Murrett CS. **Computational methods for high resolution prediction and refinement of protein structures**. *Curr Opin Struct Biol*. **2013**:177-84.
- Schmidt T, Bergner A, Schwede T. **Modelling three-dimensional protein structures for applications in drug design**. *Drug Discov Today*. **2014**:890-897.

VI. **UPORABA PROTEINSKIH STRUKTUR PRI NAČRTOVANJU BIOLOŠKO AKTIVNIH MOLEKUL**

- Schneider G, Baringhaus K-H, Kubinyi H. **Molecular Design: Concepts and Applications**, Wiley-VCH Verlag, Weinheim **2008**.
- Perdih A, Kotnik M, Oblak M et al. **Uporaba računalniške kemije pri načrtovanju in iskanju novih spojin vodnic**. *Farm. vestn*. **2010**; 61: 195-202.
- Jukič M, Perdih A, Šolmajer T. **Integriran pristop iskanja spojin vodnic naravnega izvora z uporabo eksperimentalnih in računalniško podprtih metod**. *Farm. Vestn*. **2012**; 63: 54-63.
- Valjavec K, Perdih A. **Razvoj zdravilnih učinkovin in metode računalniško podprtega načrtovanja** *Proteus*, **2019**, 81, 350-358.

