

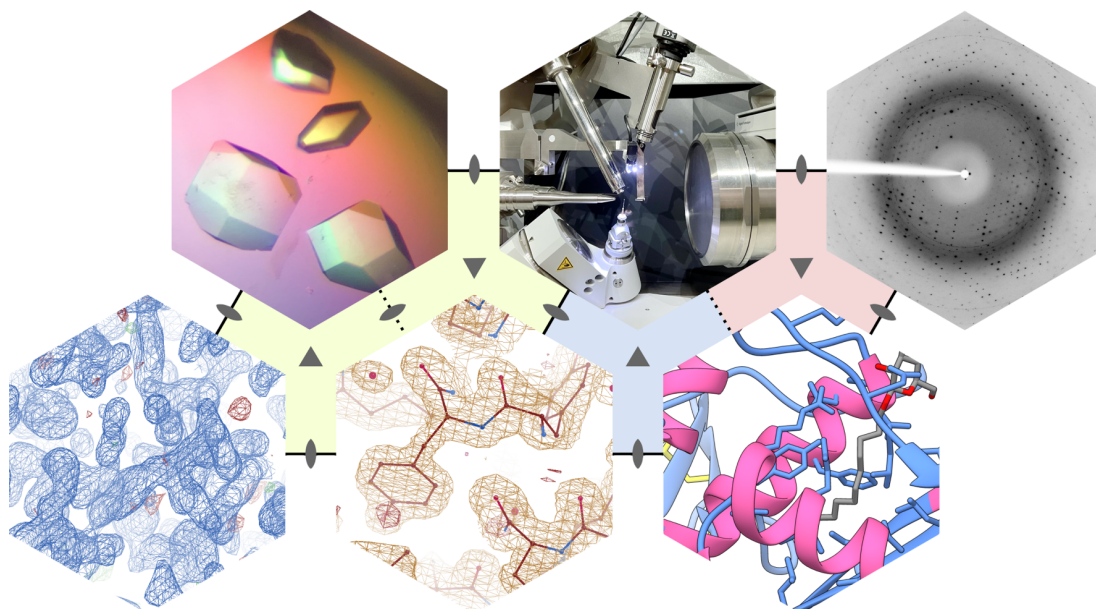
Univerza v Ljubljani
Fakulteta za *kemijo in kemijsko tehnologijo*



Miha Pavšič

Makromolekulska kristalografija

Praktikum iz X-žarkovne kristalografije bioloških makromolekul



Ljubljana, 2021

Makromolekulska kristalografija: Praktikum iz X-žarkovne kristalografije bioloških makromolekul

<i>Napisal</i>	doc. dr. Miha Pavšič
<i>Recenzija</i>	prof. dr. Kristina Djinović Carugo izr. prof. dr. Marjetka Podobnik
<i>Prelom</i>	doc. dr. Miha Pavšič
<i>Slikovni material</i>	doc. dr. Miha Pavšič
<i>Lektoriranje</i>	mag. Alenka Klemenc
<i>Urednica založbe</i>	doc. dr. Barbara Modec
<i>Založila</i>	Univerza v Ljubljani, Fakulteta za kemijo in kemijsko tehnologijo
<i>Za založbo</i>	prof. dr. Jurij Svete

© (2021) UNIVERZA V LJUBLJANI, FAKULTETA ZA KEMIJO IN KEMIJSKO TEHNOLOGIJO

1. spletna izdaja
Ljubljana, 2021

<https://www.fkkt.uni-lj.si/sl/zalozba-ul-fkkt/e-gradiva/>

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=54895875
ISBN 978-961-7078-16-9 (pdf)

Uvodna beseda

Razumevanje delovanja organizma temelji na razumevanju delovanja celic, ki ga sestavljajo, slednje na razumevanju biokemijskih procesov in vanje vključenih bioloških makromolekul, delovanje teh pa je neposredno povezano z njihovo strukturo. Da bi lahko razpoložljive strukturne podatke pravilno in učinkovito uporabljali, in morda celo prispevali v zakladnico strukturne biologije, moramo razumeti, kako do teh podatkov pridemo – torej, kako *določimo* tridimenzionalno strukturo bioloških makromolekul. En pristop – X-žarkovna kristalografija – je predstavljen v tem učbeniku, ki zraven kratkih teoretičnih osnov vsebuje tudi praktične primere v obliki navodil za vaje. Za razumevanje je nujno poznavanje osnov biokemije in molekularne biologije, strukture proteinov in bioinformatike. Vsekakor naj bralec ob poglobljenem študiju v roke vzame še dodatno obsežnejšo literaturo.

Delo je prvenstveno namenjeno študentom magistrskega študijskega programa Biokemija na Fakulteti za kemijo in kemijsko tehnologijo Univerze v Ljubljani, in sicer v okviru predmeta **Metode določanja 3D strukture makromolekul** kot učno gradivo, ki spremlja predavanja s teoretično podlago. Predmet je zasnovan tako, da študenti med vajami spoznajo praktično vse korake, ki vodijo do kristalne strukture biološke makromolekule, in sicer v okviru laboratorijskih in računalniških vaj. Nekatere tematike so obravnavane pri drugih predmetih (na primer priprava rekombinantnih proteinov), zato jih učbenik ne obravnava. Koristen vir informacij bo predstavljal tudi študentom drugih biokemijskih in molekularno-bioloških študijev. Vsem želim prijeten in uspešen študij.

Miha Pavšič

Kazalo

Uvod	7
Kaj je struktura?	7
Določanje 3D strukture makromolekul	7
Navodila za bralca	8
Ostale informacije	8
 Od ideje do strukture	 8
Pregled	9
Pot do kristalne strukture	9
 1 Načrtovanje proteinskih konstruktov	 11
1.1 Pridobivanje proteinov za kristalizacijo	11
1.2 Rekombinantni proteini	12
1.3 Načrtovanje konstruktov in modifikacij	12
1.3.1 Modifikacije na nivoju DNA	12
1.3.2 Modifikacije na nivoju proteina	14
1.4 Ocena kristalizabilnosti	16
1.4.1 Orodje XtalPred-RF	18
 Vaja 1: Načrtovanje proteinskih konstruktov	 21
 2 Kristalizacija	 25
2.1 Proces kristalizacije	25
2.2 Praktična izvedba kristalizacije	26
2.2.1 Kristalizacija po principu viseče kapljice	26
2.3 Kristalizacijski presejalni testi	27
2.3.1 Ocenjevanje rezultatov kristalizacijskih testov	28
2.4 Zanimiva dodatna literatura	29
 Vaja 2: Kristalizacija lizocima	 31
 3 Snemanje in procesiranje difrakcijskih podatkov	 35
3.1 Priprava kristalov za difrakcijsko analizo	35
3.2 Snemanje difrakcijskih podatkov	36
3.3 Procesiranje difrakcijskih podatkov	37
 Vaja 3: Snemanje in procesiranje difrakcijskih podatkov	 38
 4 Reševanje, piljenje in validacija modela strukture	 39
4.1 Povezava med difrakcijskimi podatki in strukturo	39
4.2 Določitev faz sipanih žarkov	39
4.2.1 Molekulska zamenjava	39
4.3 Asimetrična enota in njena sestava	40
4.4 Gradnja modela strukture v praksi	40
4.5 Ocenjevanje kvalitete modela strukture	41
4.6 Validacija modela strukture	41
 Vaja 4a: Ročno piljenje modela strukture v grafičnem okolju	 43

Vaja 4b: Reševanje strukture človeške karbonske anhidraze II	51
Okrajšave in simboli	56
Literatura	58

Uvod

Kaj je struktura?

Ko rečemo “določili bomo strukturo proteina”, po navadi mislimo, da bomo z uporabo enega ali več pristopov prišli do koordinat atomov, ki proteinsko molekulo sestavljajo. S tem ne mislimo, da bomo strukturo samo modelirali, na primer na osnovi znanih struktur sorodnih proteinov, temveč bomo uporabljali eksperimentalne podatke, kar nam zagotavlja, da bo določena struktura realna, prikaz dejanskega stanja. Pa je res tako?

Strukture, deponirane v podatkovnih zbirkah, se vse preveč rado interpretira na način “tako je in nič drugače”. Moramo se zavedati, da so te strukture pravzaprav *modeli struktur*, ki predstavljajo eno ali več stanj, ki jih zavzame molekula, ali neko povprečno stanje pri določenih pogojih¹. Modeli kot taki torej ne zajemajo vseh vidikov strukture (omejeni podatki o dinamiki/fleksibilnosti molekul ter z njo povezanimi različnimi konformacijami, vpliv okolja molekule med določevanjem strukture, ...), prav tako lahko vsebujejo napake. Pri njihovi uporabi moramo biti torej previdni in poznati vsaj osnove njihovega določanja, da se bomo zavedali pasti ter jih znati pravilno interpretirati in uporabljati.

Določanje 3D strukture makromolekul

Rentgenska (X-žarkovna) kristalografija je metoda, s katero je bil do sedaj določen večji delež struktur makromolekul v zbirki [Protein Data Bank \(PDB\)](https://www.rcsb.org/pdb) (1), in sicer dobrih 89 % (tabela 1 in slika 1; podatki so z dne 2. 12. 2020, s strani <https://www.rcsb.org/stats/>).

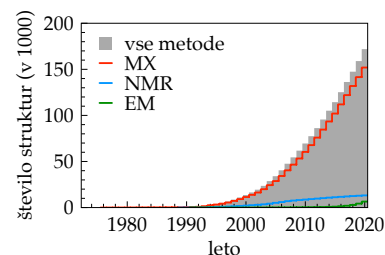
Metoda	Proteini	NK	Proteini/NK	Drugo	Vsota	Delež [%]
MX	134576	2110	7119	8149	151954	88,4
NMR	11523	1309	269	92	13193	7,7
EM	4334	53	1544	555	6486	3,8
HM	162	6	3	6	177	0,1
drugo	99	3	0	4	106	<0,1
Skupaj	150694	3481	8935	8806	171916	

Do nedavnega je bila X-žarkovna kristalografija tudi edina metoda, ki je bila primerna za določitev strukture zelo velikih makromolekul ali njihovih kompleksov pri (skoraj) atomski ločljivosti. Gradnja tovrstnih modelov namreč zahteva ločljivosti, ki so boljše od 4 Å, saj takrat definiranje koordinat posameznih atomov postane možno in smiselno. Ločljivost večine makromolekulskih struktur, deponiranih v zbirko PDB, je boljša kot 4 Å; pri relativno majhnih makromolekulah so se najvišje ločljivosti približale teoretični meji pri približno 0,5 Å (slika 2). V to skupino metod je pred nekaj leti vstopila krio-elektronska mikroskopija, s pomočjo katere je možno doseči ločljivosti, ki so boljše od 1,5 Å, o čemer priča tudi struktura mišjega apoferitina (0,5 MDa) pri ločljivosti 1,25 Å (2).

Ko govorimo o rentgenski kristalografiji v smislu določevanja struktur makromolekul, pogosto uporabljamo besedno zvezo *makromolekulska X-žarkovna*

¹ Vsekakor besedne zveze *model strukture* ne gre zamenjevati s teoretičnimi modeli, npr. modeli na osnovi homologije ali celo modeli *ab initio*.

Tabela 1: Število v zbirko PDB deponiranih modelov struktur makromolekul glede na uporabljen metodo (MX – makromolekulska kristalografija, NMR – jedrska magnetna resonanca, EM – krio-elektronska mikroskopija, HM – hibridne metode). Okrajšava NK pomeni nukleinske kisline.



Slika 1: Rast števila struktur, deponiranih v zbirko PDB. Podatki za leto 2020 niso dokončni.

kristalografija (na kratko: *makromolekulska kristalografija*, okrajšava MX). Še ožji pojem, ki se nanaša na proteine, pa je *proteinska kristalografija* (PX). V praksi se "proteinska kristalografija" uporablja na splošno in zajema vse makromolekule; gre za zgodovinske razloge, saj so bile prve strukture makromolekul strukture proteinov, pa tudi danes strukture proteinov predstavljajo skoraj 93 % vseh struktur, deponiranih v zbirko PDB, strukture nukleinskih kislin, ki niso v kompleksu s proteini, pa zgolj 2 %. Tudi v tem besedilu je beseda "protein" pogosto uporabljena z mislijo na vse makromolekule, tako zaradi razširjene splošne uporabe kot tudi zaradi dejstva, da so opisani koraki med določevanjem strukture proteinov, nukleinskih kislin ter njihovih kompleksov zelo podobni, če ne celo identični.

Navodila za bralca

Povezave

Besedilo vsebuje precej povezav do spletnih virov, vključene so pa na dva načina:

- manj pomembne so prikazane v vijolični barvi in so vidne oz. delujoče le v elektronski obliki, na primer povezava do spletne strani [Univerze v Ljubljani](#), ki posebne predstavitve zagotovo ne potrebuje;
- pomembnejše povezave do spletnih podatkovnih zbirk so prav tako obarvane vijolično, a z drugo pisavo in izpisane, tako da je naslov URL viden v natisnjeni obliki, hkrati pa delujejo kot aktivne povezave tudi v elektronski obliki, na primer povezava do zbirke struktur PDB na naslovu <http://www.rcsb.org/>.

Znotraj datoteke PDF kot povezave delujejo tudi številke slik, tabel in strani ter elementi kazala.

Navodila za računalniške vaje

Velika večina literature na temo makromolekulske kristalografije, vključno s tekočo literaturo (članki), je v angleščini. Za lažje povezovanje pojmov so nekateri pomembnejši ali za kristalografijo specifični izrazi navedeni tudi v angleškem jeziku in sicer v *poševnem tisku* v oklepaju neposredno za izrazom v slovenščini; primer: sejanje (*seeding*).

Navodila za računalniške vaje

V delu navodil, ki se nanaša na računalniške vaje, je za označevanje ukazov ter datotek in map uporabljena naslednja notacija:

mapa V navodilih je za ločevanje med imeni map in datotek uporabljen znak / (poševnica, angl. *slash*), kar velja za UNIX in podobne operacijske sisteme (Linux, macOS), pri okolju MS Windows® pa se uporablja \ (leva poševnica, angl. *backslash*);

menu predstavlja nekaj, kamor kliknemo in s tem odpremo seznam možnosti, vključimo neko določeno nastavitve ali pa kar gumb, ki sproži izvajanje.

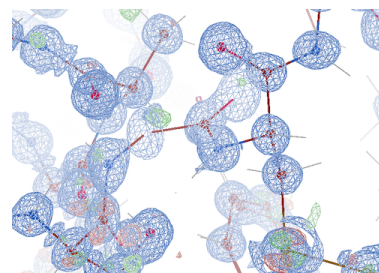
Dostop do materiala

Material v obliki datotek, potrebnih za izvedbo vaj, je na voljo pri avtorju – miha.pavsic@fkkt.uni-lj.si.

Ostale informacije

Vse shematske slike struktur (prikaz kot trak – angl. *ribbon*, molekulska površina ...) so narisane s programom UCSF Chimera (4). Slike struktur, kjer je prikazana elektronska gostota, so bile narisane s programom Coot (5).

V učbeniku je nekaj slik, ki so vzete iz drugih virov; ti so navedeni ob sliki.



Slika 2: Izsek strukture in karte elektronske gostote krabina, določene z X-žarkovni difrakcijo pri ločljivosti 0,54 Å (3). Elektronski oblaki posameznih atomov so jasno vidni, pri tej ločljivosti je bilo moč določiti tudi položaje vodikovih atomov (vezi z njimi so prikazane kot sive paličice). PDB ID [1EJG](#).

Pregled

Pot do kristalne strukture

Potek določanja 3D strukture nekega proteina z rentgensko difrakcijo lahko poenostavljeno razdelimo na naslednje korake, ki so shematsko prikazani na sliki 3 na strani 10:

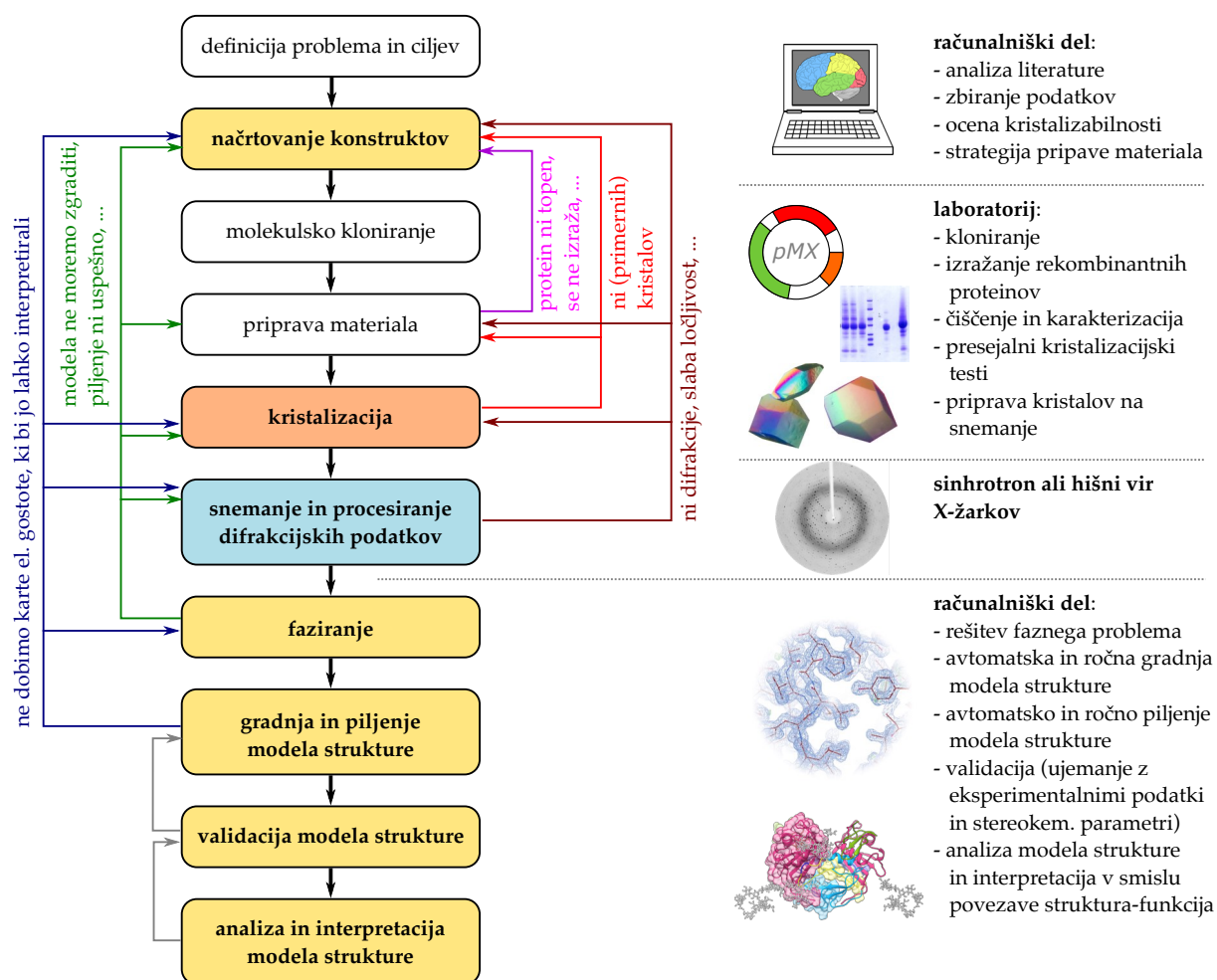
1. **Načrt projekta**, v okviru katerega definiramo problem in identificiramo specifične cilje – slednji so v našem primeru kristalne strukture, katerih interpretacija nam bo po pričakovanju dala odgovore na zastavljena vprašanja;

Primer *problema*: "Kako vezava kalcijevih ionov na α -aktinin-1 vpliva na njegovo prečno povezovanje aktinskih filamentov?" Do odgovora bi se lahko, na primer, dokopali s kristalnimi strukturami α -aktinina-1 oz. relevantnega dela z in brez vezanih ionov Ca^{2+} , eventualno še s strukturo kompleksa z aktinom – te strukture oz. njihova določitev bi bili naši *cilji*.

2. **Priprava in čiščenje makromolekule** – v skladu z načrtom projekta in zastavljenimi cilji razmislimo, strukturo katerih proteinov, njihovih delov ali kompleksov želimo določiti ter kako bomo pripravili material za kristalizacijo. Danes se večinoma uporabljajo rekombinantni proteini, kjer imamo veliko možnosti vplivanja na lastnosti proteinskih molekul – lahko pripravimo nemutirane ali mutirane proteine in njihove dele, odstranimo neželene fleksibilne regije, vplivamo na posttranslacijske modifikacije ipd. Pri tem ponavadi želimo imeti čim manjše konstrukte, ki nam bodo še omogočali odgovoriti na zastavljena vprašanja, hkrati pa načrtno vplivati na njihove lastnosti z namenom povečanja verjetnosti tvorbe kristalov. Ravno ta del, načrtovanje proteinskih konstruktov, je predmet razdelka 1 in spremljajoče računalniške vaje.
3. **Kristalizacija** – iskanje kristalizacijskih pogojev z uporabo presejalnih testov, optimizacija kristalizacije ter v končni fazi pridobitev kristalov, primernih za difrakcijsko analizo. To je ponavadi ozko grlo projekta, saj se sestave kristalizacijskih raztopin, ki bodo omogočale tvorbo kristalov, (še) ne da izračunati oz. napovedati. Kristalizacija je predmet razdelka 2 in spremljajoče laboratorijske vaje.
4. **Difrakcijska analiza** – najprej kristale karakteriziramo (ocena njihove kvalitete in kvalitete difrakcije – ločljivost, defekti ...), nato pa s primernimi kristali posnamemo difrakcijske podatke (na domačih oz. hišnih virih X-žarkov ali sinhrotronu), podatke procesiramo (indeksiranje, določitev in piljenje parametrov osnovne celice, skaliranje, integracija) ter tako pridemo do najverjetnejše prostorske skupine, Millerjevih indeksov hkl ter intenzitete I_{hkl} . Snemanje in procesiranje difrakcijskih podatkov je na kratko opisano v razdelku 3, spremlja ga demonstracijska vaja.
5. **Določitev faze sipanih žarkov in izračun elektronske gostote** – da lahko izračunamo karto elektronske gostote, potrebujemo zraven intenzitete sipanih žarkov (podatek o amplitudi) še njihovo fazo; ta t. i. fazni problem lahko rešimo z različnimi metodami, na primer z molekulsko zamenjavo, prek anomalnega sipanja in izomorfno zamenjavo. Molekulsko zamenjavo obravnava razdelek 4, s spremljajočo računalniško vajo.

6. **Interpretacija karte elektronske gostote ter gradnja in piljenje modela**, v okviru katere v karto elektronske gostote zgradimo model strukture, koraki gradnje pa se ciklično izmenjujejo s koraki piljenja. Naš cilj je model strukture, ki je čim bolj v skladu tako z eksperimentalnimi podatki kot tudi s splošnimi stereokemijskimi parametri (dolžine vezi, koti, ...). Korake obravnava razdelek 4, s spremljajočo računalniško vajo.
7. **Validacija in analiza modela**, kjer dodatno analiziramo realnost našega modela strukture in poskušamo z njim odgovoriti na vprašanja, zastavljena med načrtovanjem projekta.

Proces ni tako enostaven in enosmeren, pogosto je potrebno vračanje en ali več korakov nazaj, in sicer zaradi optimizacije določenih stopenj, pridobivanja dodatnih podatkov ipd. Nekaj možnih in pogostih scenarijev je prikazanih na sliki 3.



Slika 3: Pregledna shema procesa določevanja strukture biološke makromolekule z X-žarkovno kristalografijo. Proces ne poteka enosmerno od problema do modela strukture temveč je pogosto potrebno vračanje en ali več korakov nazaj, kar ponazarjajo krivljene puščice ob straneh. Prikazani so pogosti scenariji, ne pa vsi možni. Koraki z obarvanim ozadjem predstavljajo vaje, opisane v nadaljevanju, ki se izvajajo kot **računalniške** ali **laboratorijske** vaje oziroma kot **demonstracija**.

1

Načrtovanje proteinskih konstruktov

1.1 Pridobivanje proteinov za kristalizacijo

Na splošno lahko proteine oz. biološke makromolekule za namen kristalizacije pridobivamo na tri načine:

- izolacija iz tkiv, organov ali celotnih organizmov,
- kemijska sinteza,
- izražanje z uporabo homo- in heterolognih ekspresijskih sistemov.

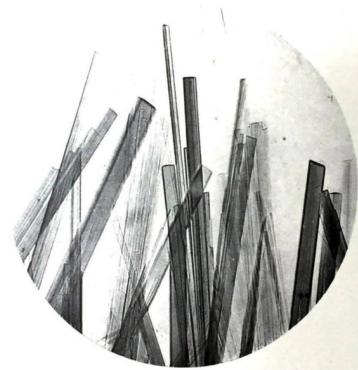
Izolacija proteinov iz tkiv, organov ali celotnih organizmov je bila aktualna predvsem pred razvojem in široko uporabo tehnologije rekombinantne DNA (slika 1.1). S tem pristopom so povezane številne težave:

- **majhna količina** – veliko število proteinov je prisotnih v majhnih količinah, zato je potrebno uporabiti veliko izhodnega materiala, prav tako je izolacija redkih proteinov težja;
- **kemijska nehomogenost** – proteini so velikokrat kemijsko nehomogeni, ker je posledica posttranslacijskih modifikacij (glikozilacija, proteolitsko procesiranje, fosforilacija, ...);
- **težka dostopnost izhodnega materiala** – pogosto je težko priti do zadostnih količin izhodnega materiala za izolacijo proteina, npr. v primeru "eksotičnih" organizmov. Prav tako so s pridobivanjem materiala živalskega in človeškega izvora povezani številni etični vidiki;
- **odsotnost enostavnih označevalnih regij** – na primer, taki proteini npr. ne vsebujejo široko uporabljane heksahistidinske oznake za čiščenje z afinitetno kromatografijo z imobiliziranimi kovinskimi ioni.

Nekaj navedenih težav bi lahko enostavno rešili s **kemijsko sintezo** polipeptidnih verig ter verig nukleinskih kislin, a je to pri večjih molekulah in zahtevanih večjih količinah produkta zahtevno in drago, prav tako pa se pri proteinih pojavijo težave s pravilnim zvijem polipeptidne verige. Tak pristop je torej primeren zgolj za manjše molekule, npr. peptide in kratke segmente nukleinskih kislin, ki jih je najenostavneje naročiti pri podjetjih, specializiranih za njihovo sintezo.

Proteine za kristalizacijo danes najpogosteje pripravljamo z **uporabo homo- in heterolognih ekspresijskih sistemov**, translacijo *in vitro* pa zaradi omejitev pri količini pridobljenega proteina in potencialnimi težavami s pravilnim zvijem uporabljamo redko. Kot prokariotski sistem uporabljamo večinoma bakterije *Escherichia coli* ali *Bacillus subtilis*, kot evkariotski pa kvasovke (*Saccharomyes cerevisiae*, *Pichia pastoris*), insektne celice (organizma *Spodoptera frugiperda* ali *Trichoplusia ni*, zapis najpogosteje vnašamo s pomočjo bakulovirusov) ter sesalske celice (mišje, človeške). Druge organizme, vključno z rastlinami, za pridobivanje proteinov za kristalizacijo uporabljamo redkeje. Podrobnosti in specifične dela s posameznim ekspresijskim sistemom tu niso obravnavane, zato naj se bralec usmeri na drugo relevantno literaturo, ki obravnava značilnosti in izbor primerne ekspresijskega sistema (7).

V smislu proizvodnje rekombinantnih proteinov za uporabi v strukturni biologiji so v novejšem času stopile v ospredje **visoko zmogljive metode**



Slika 1.1: Eden prvih kristaliziranih proteinov je bil hemoglobin. Leta 1909 sta Reichert in Brown objavila več kot 600 fotografij kristalov hemoglobina, izoliranega iz 109 različnih živalskih vrst, ter predpostavila, da sorodnost med temi vrstami korelira z obliko kristalov (6).

(angl. *high-throughput (HT) methods*) proizvodnje proteinov, razvite predvsem v okviru projektov strukturne genomike. Omogočajo hitro in učinkovito pripravo konstruktov za izražanje proteinov, izražanje samo, čiščenje ter nadaljnje korake. Pregled tovrstnih metod je na voljo v znanstveni literaturi (8; 9), prav tako so zanimivi opisi nekaterih delujočih visoko zmogljivih sistemov (10; 11; 12; 13).

1.2 Rekombinantni proteini

Uporaba ekspresijskih sistemov zahteva, da zapis za želeni protein kloniramo s pomočjo **tehnologije rekombinantne DNA**, kar nam nudi veliko fleksibilnosti pri načrtovanju proteinskih konstruktov z želenimi lastnostmi (visok nivo izražanja, topnost, enostavno čiščenje, sposobnost tvorbe kristalov ...). Za te modifikacije se moramo odločiti že v stopnji **načrtovanja konstruktov** na nivoju aminokislinskega oz. nukleotidnega zaporedja, na nekatere lastnosti proteinov pa lahko vplivamo tudi kasneje, ko so ti že očiščeni (na primer odstranitev glikozilacije in metilacija lizinskih ostankov). Po navadi delamo z več konstrukti vzporedno, s tem prihranimo čas ter povečamo možnost priprave proteina v primerni količini in čistosti, prav tako pa pogosto delamo kristalizacijske presejalne teste z več konstrukti oz. oblikami proteina vzporedno. Sicer je nemogoče z gotovostjo napovedati, kateri konstrukt bo dal (uporabne) kristale, kljub temu pa lahko **sposobnost oz. težnjo po tvorbi kristalov, t.i. kristalizabilnost**, vsaj približno ocenimo prek analize aminokislinskega zaporedja konstruktov. To je podrobneje obravnavano v razdelku 1.4 na strani 16. V nadaljevanju sledi opis nekaterih pogostih modifikacij, ki jih uvajamo na nivoju DNA in proteinov, z namenom vplivati na nekatere njihove lastnosti.

1.3 Načrtovanje konstruktov in modifikacij

V okviru načrtovanja strategije priprave proteinskih konstruktov v prvi vrsti spada izbira ekspresijskega sistema, saj so od tega odvisni izbor vektorjev, način dela, možne posttranslacijske modifikacije ipd. Včasih se odločimo za delo z dvema sistemoma vzporedno (npr. bakterijski sistem in insektne celice), s čimer povečamo možnost, da nam bo uspelo pripraviti protein v topni obliki v zadostnih količinah. Pri tem kot enega izmed sistemov pogosto izberemo ravno bakterijskega, saj je čas od konstrukta DNA do izraženega proteina kratek, v primeru neuspeha pa imamo v tem času že pripravljene konstrukte za katerega od kompleksnejših sistemov.

1.3.1 Modifikacije na nivoju DNA

Prek modifikacije zaporedja DNA v zaporedje proteina najpogosteje uvedemo:

- **označevalni zapis** za detekcijo s protitelesi in čiščenje s pomočjo afinitetne kromatografije z imobiliziranimi kovinskimi ioni (npr. nikljeva afinitetna kromatografija), najpogostejše tovrstne oznake pa so heksahistidinska oznaka (His₆-oznaka oz. repek) in oznake STREP (zaporedje WSHPQFEK), HA (YPYDVPDYA; izbira iz proteina hemaglutinina) ter FLAG (zaporedje DYKDDDDK), za detekcijo s protitelesi pa se pogosto uporablja tudi oznaka Myc (zaporedje EQKLISEEDL); pregled in podroben opis označevalnih zapisov naj bralec poišče v drugi literaturi (14; 15; 16);
- **zaporedje, znotraj katerega cepi visokospecifična peptidaza**, s katero lahko med čiščenjem odstranimo vnesen označevalni zapis ali drugi večji fuzijski del, ki bi lahko negativno vplival na sposobnost tvorbe kristalov¹. Pogosto se uporabljajo cepitvena mesta za naslednje peptidaze (↓ označuje mesto cepitve): trombin (LVPR↓), proteaza virusa jedkanja tobaka - TEV (Tobacco Etch Virus; ENLYFQ↓), enterokinaza (DDDDK↓) in faktor Xa (IDGR↓). Bralec lahko obširne preglede encimov najde v drugi literaturi (14; 17).

¹ Na primer, kratke oznake kot je His₆ so pogosto strukturno neurejene in lahko negativno vplivajo na tvorbo kristalnih kontaktov.

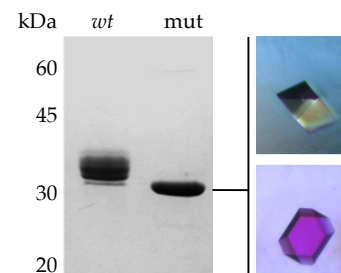
Umestitev označevalnega zaporedja in cepitvenega mesta na N- oz. C-konec proteina zahteva tehten razmislek, pri katerem upoštevamo potencialno motenje pri zvijanju ter dostopnost oznake in cepitvenega mesta. Dostopnost lahko izboljšamo tako, da med želeni protein in oznako/cepitveno mesto vstavimo kratek fleksibilen povezovalni peptid, npr. z zaporedjem GS.

Tehnologija rekombinantne DNA nam omogoča tudi številne druge modifikacije, paziti pa moramo, da z njimi proteina ne spremenimo do te mere, da nam struktura s funkcijskega vidika ne omogoča več pravilne interpretacije :

- **delecija N- in/ali C-končnih regij**, ki so pogosto neurejene in lahko negativno vplivajo na sposobnost tvorbe medmolekulskih kontaktov v kristalu ter posledično na kristalizacijo (nasprotno, posledice modifikacij z namenom promoviranja medmolekulskih kristalnih kontaktov prek uvajanja dodatnih regij so izredno nepredvidljive);
- **izbris N- in O-glikozilacijskih mest** prek uvedbe drugačno smiselnih mutacij na mesta aminokislinskih ostankov, kamor so pripeti ogljikovi hidrati (v primeru N-glikozilacije pogosto $Nx[ST] \rightarrow Qx[ST]$ ali $Sx[ST]$), s čimer omogočimo/olajšamo tvorbo medmolekulskih kontaktov med polipeptidnimi verigami ter hkrati povečamo kemijsko (v primeru heterogene glikozilacije, primer na sliki 1.2) ter konformacijsko homogenost vzorca (ogljikohidratne verige so na površini proteina ter pogosto nimajo stabilne konformacije). Zavedati pa se moramo, da sprememba glikozilacije vpliva tudi na intrinzične lastnosti samega proteina, njegovo dinamiko in stabilnost²;
- **priprava samo določene domene ali več domen**, kar pomeni manjši protein (pogosto lažje delo), večjo topnost in načeloma večjo konformacijsko homogenost, posebej v primeru daljših fleksibilnih meddomenskih povezav; delo s posameznimi domenami vedno ni (enostavno) izvedljivo, saj so lahko stiki med domenami večdomenskega proteina hidrofobni, pri izolirani domeni pa bi to pomenilo izpostavitve teh regij ter povečane težnje k agregaciji. Prav tako pa je lahko prisotnost neke domene oz. meddomenskih stikov ključnega pomena za pravilno zvijte ali konformacijo neke druge domene;
- **fuzija dveh različnih proteinov ali njihovih domen** prek povezovalne regije (20; 21), kar se lahko uporablja tudi z namenom povečanja možnosti tvorbe kristalnih kontaktov, na primer fuzije s proteinom, ki veže maltozo (*maltose binding protein*, MBP);
- **vnos mutacij na površini proteina**, z namenom vplivanja na površinski naboj, na primer povečanje topnosti z uvedbo bolj polarnih ali celo nabitih aminokislinskih ostankov na površino, uvedbo *lepljivih površin*, z namenom promoviranja kristalnih kontaktov (22) ali pa zmanjšanja površinske entropije, kar je v ločenem odstavku opisano v nadaljevanju;
- **vnos drugih mutacij, ki vplivajo na konformacijo proteina in njegovo aktivnost** (namenska inaktivacija ali namenska aktivacija proteina) – pri kristalizaciji proteaz je včasih smiselno mutirati aminokislinski ostanek, ključen za njihovo aktivnost, ter s tem preprečiti samorazgradnjo.

V okviru priprave rekombinantnih proteinov je treba omeniti še možnost zamenjave metionina (Met) s selenometioninom (se-Met), kar pride v poštev za določitev faz sipanih žarkov prek anomalnega sipanja uvedenih selenovih atomov v fazi določanja strukture (na primer pri Se-SAD). Pri tem sicer ne gre za uvedbo modifikacij na nivoju DNA, gre namreč zgolj za to, da za metionin avksotrofen gostiteljski (mikro)organizem gojimo v gojišču, ki kot vir metionina vsebuje selenometionin. Na tem mestu je modifikacija omenjena zgolj zato, ker je uporabna le v primeru rekombinantnih proteinov.

V luči uvedbe modifikacij je zanimiv tudi pristop **usmerjene evolucije**, pri kateri iščemo mutante nekega proteina z želenimi lastnostmi v obsežni knjižnici naključnih mutantov (23) (nekoliko bolj poljuden opis postopka v slovenskem jeziku je v članku (Po)moč evolucije (24)). Drugače povedano, knjižnico mutantov nekega proteina izpostavimo selekcijskemu pritisku, pri



Slika 1.2: Iz primera na sliki je razvidno, kako je uvedba treh mutacij $N \rightarrow Q$ in posledično ugasnitev N-glikozilacije povečala kemijsko homogenost mutiranega proteina (mut) v primerjavi s proteinom divjega tipa (wt); mutirana oblika je kasneje dala kristale (desno), ki so sipali do ločljivosti 1.9 Å (18).

² Zanimivo, Lee in sod. so prek simulacij molekulske dinamike proteinskih struktur z N-pripetimi oligosaharidi pokazali, da N-glikozilacija zmanjša dinamiko in domnevno posledično poveča stabilnost proteinov (19).

čemer je način izbire mutant oz. "ribarjenja" po knjižnici odvisen od lastnosti, ki jo iščemo (topnost, vezava liganda ...). V okviru tega se uporabljajo pristopi, kot sta predstavitev na fagu in predstavitev na ribosomu. Na podoben način lahko pripravimo tudi kombinatorične knjižnice konstruktov, modificiranih na različne načine (skrajšanje, mutacije, fragmentacija) (25).

Pristop, povezan z usmerjeno evolucijo, je uporaba kristalizacijskih šaperonov. To so proteini, ki se vežejo na protein oz. makromolekulo, ki jo želimo kristalizirati, in prek visokega potenciala tvorbe medmolekulskih kontaktov lahko pozitivno vplivajo na kristalizacijo. Primeri pogosto uporabljenih kristalizacijskih šaperonov so nanotelesa (*nanobodies*) in DARPINi (*Designed Ankyrin Repeat ProteiNs*). Znani so številni primeri, kjer je uporaba tovrstnih orodij odločilno prispevala k določitvi strukture nekega proteina; pregled pristopov najdemo v znanstveni literaturi (26; 27; 28).

Zmanjšanje površinske entropije pomeni, da prek uvedbe mutacij zmanjšamo neurejenost stranskih skupin aminokislinskih ostankov na površini proteina, ki je povezana z njihovo fleksibilnostjo (29). Aminokislinski ostanki, katerih stranske skupine so pogosto izpostavljene topilu, so Lys, Glu in Gln; pri omenjenem pristopu jih najpogosteje mutiramo v Ala, Ser ali Thr. Koncept površinske entropije in njene vloge pri kristalizaciji temelji na naslednji enačbi:

$$\Delta G = \Delta H - T(\Delta S_{protein} + \Delta S_{topilo}) \quad (1.1)$$

Glede na to, da so entalpijske vrednosti kristalnih kontaktov relativno majhne (medmolekulski kontakti so šibki), je kristalizacija entropijsko voden proces, ki pa vključuje spremembo entropije tako v smislu proteina kot topila. Prispevek topila (S_{topilo}) gre na račun s površine proteinskih molekul sproščenih molekul vode, do česar pride zaradi udeležbe teh delov površine proteina pri tvorbi medmolekulskih kontaktov, in ima kot tak pozitiven predznak, saj pride do povečanja entropije. Prispevek proteina ($S_{protein}$) izvira iz:

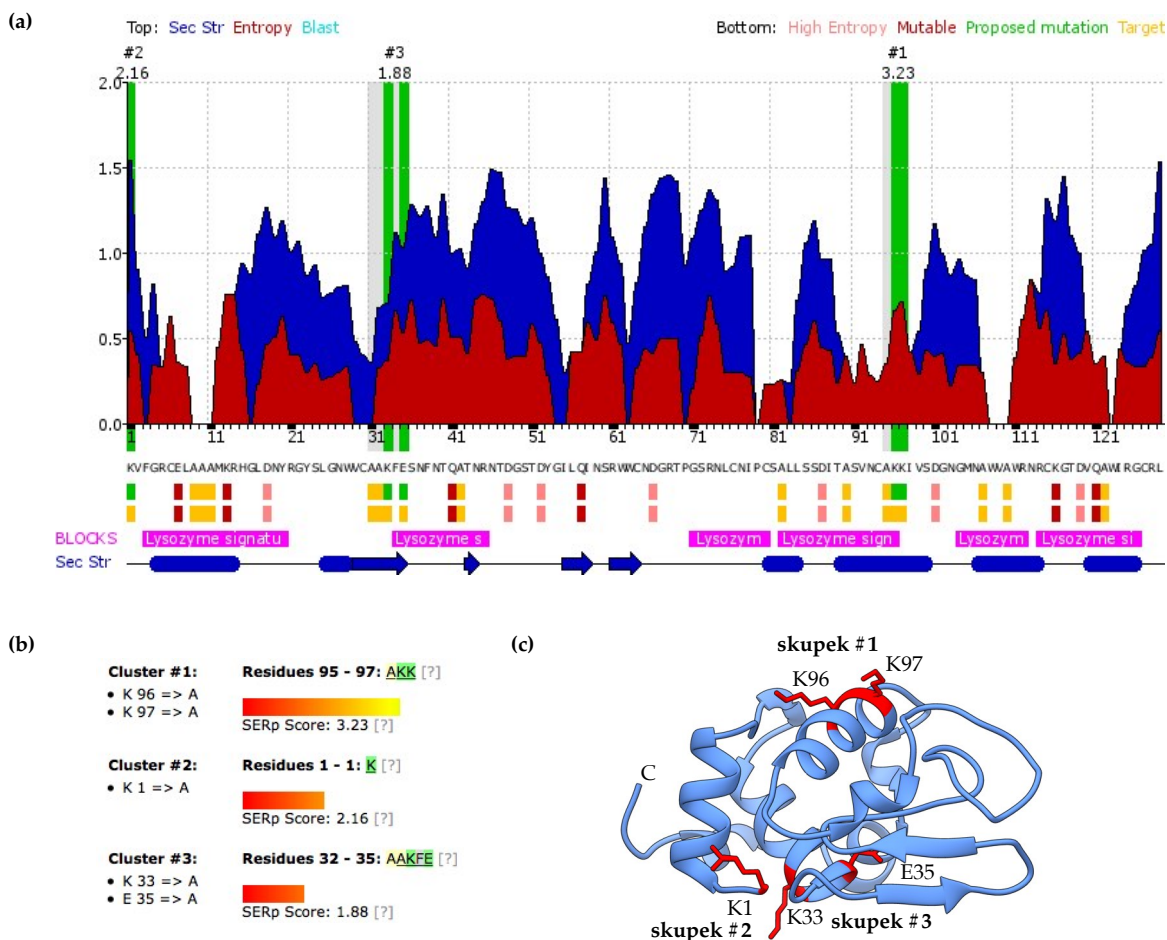
- zmanjšanja entropije zaradi zlaganja proteinskih molekul v urejeno kristalno mrežo,
- ureditve stranskih skupin aminokislinskih ostankov na površini proteina zaradi tvorbe kristalne mreže (to imenujemo tudi *entropijski ščit* pred nastankom medmolekulskih interakcij).

Slednji je predvsem izrazit v primeru aminokislinskih ostankov z velikimi fleksibilnimi stranskimi skupinami, kot sta lizinski in glutamatni ostanek. Po teoriji bi lahko z mutacijo slednjih zmanjšali omenjeni entropijski ščit in s tem prispevali k ugodnejši energetski bilanci nastanka kristala. Ravno na tem temelji pristop zmanjšanja površinske entropije (*surface entropy reduction, SER*), ki je implementiran v programu SERp (30). Slednji prek kompozitne napovedi, sestavljene iz napovedi sekundarne strukture, površinskega entropijskega profila in poravnave z zaporedji homologov z znano strukturo (za oceno izpostavljenosti aminokislinskih ostankov topilu), identificira skupke (*clusters*) aminokislinskih ostankov, ki so na površini proteina ter bi jih bilo smiselno mutirati ter s tem zmanjšati površinsko entropijo. Primer analize aminokislinskega zaporedja kokošjega lizocima, ki sicer precej rad kristalizira, je prikazan na sliki 1.3 na strani 15, ki vsebuje tudi kompozitni diagram posameznih analiz. V tem diagramu so nad aminokislinskim zaporedjem z modro prikazani vrhovi, ki naj bi ustrezali zankam in drugim manj urejenim regijam v strukturi, z rdečo pa regije z visoko entropijo; svetlo modra barva je namenjena prikazu aminokislinskih ostankov, ki so v strukturah homologov drugačni. Najprimernejši skupki za mutiranje so prikazani z zeleno. Pod aminokislinskim zaporedjem so posebej označeni aminokislinski ostanki, ki imajo nizko (rumeno) oz. visoko entropijo (rožnato), so predlagani za mutacije (zeleno) ali pa jih je na splošno smiselno zmutirati z namenom (rdeče). Sekundarna struktura je kot trak na dnu.

1.3.2 Modifikacije na nivoju proteina

Protein lahko modificiramo tudi po izražanju. Najpogostejše modifikacije so:

- **omejena proteoliza**, z namenom priprave manjših fragmentov (cepitev na izpostavljenih regijah z endopeptidazami); klasičen in splošno znan

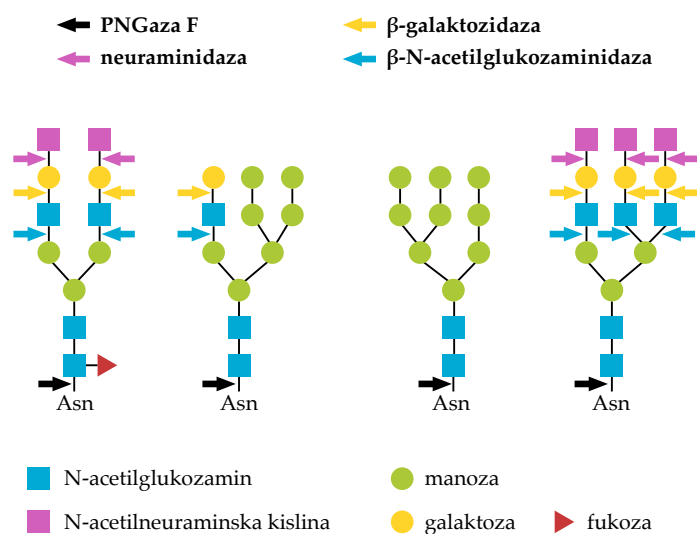


Slika 1.3: Napoved površinske entropije. Primer analize aminokislinskega zaporedja kokošjega lizocima C (Uniprot ID P00698, regija 19–147) z orodjem SERp. (a) Kompozitni diagram rezultatov posameznih analiz, podrobneje opisan na strani 14. (b) Skupki in predlagane mutacije. Ocena SERp ima arbitrarne enote. (c) Prikaz identificiranih skupkov na strukturi lizocima C (PDB ID 193L). Polipeptidna veriga je prikazana kot moder trak, skupki pa kot rdeče paličice (tudi trak je na njihovem mestu rdeče obarvan).

primer je proteolitska fragmentacija protiteles IgG s cisteinsko proteazo papainom, pri kateri pride do cepitev na pregibnih regijah težkih verig (med CH₁ in CH₂) in posledično do ločitve fragmentov F_{ab} in F_c;

- **odstranitev označevalnih zaporedij** (pogosto nimajo urejene strukture) prek vnešenih cepitvenih mest za visoko specifične endopeptidaze (proteaza 3C, TEV, enterokinaza ...);
- **odstranitev neurejenih N- in C-končnih regij**, lahko tudi označevalnih zaporedij, s pomočjo amino- ali karboksi-peptidaz (torej eksopeptidaz), ki aminokislinske ostanke odcepljajo z N- oz. C-konca tako dolgo, dokler ne naletijo na sterično oviro (nezadostno izpostavljen aminokislinski ostanek in s tem nedostopna peptidna vez) ali aminokislinski ostanek, ki ne ustreza njihovi specifičnosti (na primer nekatere peptidaze ne cepijo tik pred/za prolinskimi ostanki);
- **kemijska metilacija lizinskih aminokislinskih ostankov**, kar nevtralizira pozitivni naboj na površini proteinov in promovira hidrofobne medmolekulske interakcije prek metiliranih lizinskih ostankov (31);
- **(delna) odstranitev ogljikovih hidratov** z glikoziliranih proteinov s pomočjo glikozidaz (32), kar zmanjša kemijsko (kot posledica heterogene glikozilacije) in konformacijsko heterogenost ter lahko pozitivno prispeva k tvorbi kristalnih kontaktov; na primer PNGaza F (peptid-N-glikozidaza F) cepi med N-acetilglukozaminom in asparaginskim aminokislinskim ostankom z manozo bogatih, hibridnih in kompleksnih na Asn pritrdjenih ogljikohidratnih verig. Številni proizvajalci prodajajo posamezne glikozidaze in tudi mešanice glikozidaz, ki bolj ali manj popolno odstranijo

sladkorne ostanke, predvsem je težava lahko z dostopnostjo vezi Asn-NAcGln. Mesta cepitve različnih glikozidaz so prikazana na sliki 1.4.



Slika 1.4: Cepitvena nekaterih glikozidaz, ki se uporabljajo za encimsko deglikozilacijo N-glikoziliranih proteinov.

Očiščene proteine nato karakteriziramo z različnimi metodami (gelska filtracija, statično in dinamično sipanje svetlobe, termična stabilnost, elektrofo-reza pod nativnimi pogoji, izoelektrični fokusiranje, eno- in dvo-dimenzionalni spektri NMR, masna spektrometrija, ...) in tako ocenimo njihovo stabilnost, monodisperznost, zvijete ipd.

Pristop, povezan z v prejšnjem razdelku omenjeno usmerjeno evolucijo, a tokrat na nivoju proteinov, je uporaba kristalizacijskih šaperonov. Gre za proteine, ki se vežejo na protein oz. makromolekulo (izberemo jih iz bazena mutantov), ki jo želimo kristalizirati, in prek visokega potenciala tvorbe medmolekulskih kontaktov lahko pozitivno vplivajo na kristalizacijo. Primeri pogosto uporabljenih kristalizacijskih šaperonov so nanotelesa (*nanobodies*) in DARPINi (*Designed Ankyrin Repeat ProteinNs*). Znani so številni primeri, kjer je uporaba tovrstnih orodij odločilno prispevala k določitvi strukture nekega proteina. Pregled pristopov je na voljo v znanstveni literaturi (26; 27; 28).

1.4 Ocena kristalizabilnosti

Zmožnost nekega proteina oz. proteinskega konstrukta, da bo tvoril kristale, imenujemo **kristalizabilnost**. Na enostaven način jo lahko ocenimo z upoštevanjem nekaterih lastnosti proteina, npr. prisotnost daljših neurejenih regij in obsežna in heterogena glikozilacija, oboje je povezano z zmanjšano kristalizabilnostjo. Na tak način lahko preprosto rangiramo več variant/konstruktov za isti izvorni protein, a ocena je vseeno precej subjektivna in ne vključuje statistične evaluacije.

Bolj relevantno oceno kristalizabilnosti lahko dobimo z uporabo računalniških orodij. V grobem jih lahko razdelimo v dve skupini:

- orodja, ki kristalizabilnost ocenijo na podlagi iz aminokislinskega zaporedja izluščenih/napovedanih fizikalno-kemijskih lastnosti konstrukta;
- orodja, ki ne temeljijo na analizi fizikalno-kemijskih lastnosti konstrukta.

Orodja iz prve skupine analizirajo aminokislinsko zaporedje konstrukta, ki nas zanima, ocenijo oz. napovejo določene karakteristike zaporedja in proteina samega ter to statistično ovrednotijo v luči rezultatov enakih analiz aminokislinskih zaporedij proteinov, za katere so znani tudi rezultati realnih presejalnih kristalizacijskih testov. Pri slednjih gre torej za empirične

podatke o tem, ali je nek protein kristaliziral ali ne; ti podatki pa se pogosto črpajo iz strukturno-genomskih projektov, saj kot taki vsebujejo tako podatke, katere proteine jim je uspelo kristalizirati kot tudi tiste, kateri proteini niso kristalizirali. Vsekakor obetavni rezultati analize zaporedja načrtovanega proteinskega konstrukta ne pomenijo, da bomo v praksi zagotovo hitro dobili kristale – še vedno je ključna izvedba presejalnih (*screening*) testov, pri katerih preizkusimo veliko različnih kristalizacijskih pogojev.

Drugi tip orodij za napoved kristalizabilnosti pa ne temelji na napovedi fizikalno-kemijskih lastnosti konstruktov, s čimer se izognejo težavam z (ne)zanesljivostjo tovrstnih napovedi. Aminokislinsko zaporedje konstrukta zgolj razkosajo na *k*-mere ter jih kot take analizirajo v luči empiričnih podatkov (podoben set kot pri prvi skupini orodij). Gre za strojno učenje, na katerem temeljijo tudi nekatera orodja iz prve skupine, le vhodni parametri se razlikujejo (pri prvi skupini napovedane fizikalno-kemijske lastnosti, pri drugi pa *k*-meri zaporedja). Nekatera od orodij iz druge skupine kljub temu upoštevajo nekatere napovedane fizikalno-kemijske lastnosti, a v omejenem obsegu.

V znanstveni literaturi je opisanih precej orodij za napoved kristalizabilnosti, večino spremlja tudi prosto dostopen strežnik, na katerem program teče. Žal pa je nekaj naslovov strežnikov, predvsem starejšega datuma, neaktivnih. Na tem mestu je navedenih zgolj nekaj najbolj pogosto uporabljenih orodij ter najnovejših objav z delujočimi naslovi:

- **CrystalP2** (<http://biomine.cs.vcu.edu/servers/CRYSTALP2/>), ki pri napovedi uporablja aminokislinsko sestavo, sosledje aminokislinskih ostankov, izoelektrično točko in hidrofobnost (33);
- **PDPredictor** (<https://bioinformatics.anl.gov/cgi-bin/tools/pdpredictor/>) temelji na analizi povprečne hidrofobnosti (indeks GRAVY, *GRand AVerage hydropathY*) ter nekaterih drugih lastnosti, uporablja strojno učenje (metoda podpornih vektorjev, SVM) (34);
- **PPCPred** (<http://biomine.ece.ualberta.ca/servers/PPCPred/>), ki uporablja rezultate analize hidrofobnosti, aminokislinske sestave, napovedi neurejenosti, sekundarne strukture ter dostopnosti posameznih aminokislinskih ostankov topilu (35);
- **XtalPred-RF** (<https://xtalpred.godziklab.org>), ki z aplikacijo metode naključnih gozdov (*random forest*, RF) (36) predstavlja izboljšavo osnovnega strežnika XtalPred, kjer je napoved temeljila na analizi različnih fizikalno-kemijskih parametrov (dolžina zaporedja, izoelektrična točka, indeks GRAVY, indeks nestabilnosti, ovite vijačnice ...) (37). Orodje je podrobneje opisano na strani 18;
- **TargetCrys** (<http://csbio.njust.edu.cn/bioinf/TargetCrys>) za razliko od napovedi fizikalno-kemijskih parametrov uporablja analizo aminokislinske sestave ter sestave dipeptidov in tripeptidov, temelji pa na metodi podpornih vektorjev (SVM) (38);
- **Crysalis** (<http://biotool.xmu.edu.cn/crysalis/>), ki predstavlja izboljšavo starejšega orodja PredPPCrys (39) in temelji na analizi aminokislinskega zaporedja samega, brez napovedi fizikalno kemijskih lastnosti, uporablja metodo podpornih vektorjev (40);
- **Crysf** (<http://nmrcen.xmu.edu.cn/crysf/>), ki za napoved kristalizabilnosti uporablja pripise iz zbirke UniProt (41);
- **fDETECT** (<http://biomine.cs.vcu.edu/servers/fDETECT/>), ki zraven kristalizabilnosti napove tudi obete za uspešno izražanje ter čiščenje proteina, uporablja pa modele, osnovane na logistični regresiji (42);
- **DeepCrystal** (<https://machinelearning-protein.qcri.org/>), ki aplicira globoke nevronske mreže brez potrebe po ekstrakciji oz. napovedi fizikalno-kemijskih lastnosti konstrukta, identificira pa pogoste *k*-mere (43);

- **BCrystal** (<https://machinelearning-protein.qcri.org/>) z vključitvijo analiz nekaterih fizikalno-kemijskih lastnosti predstavlja nadgradnjo orodja DeepCrystal (44).

1.4.1 Orodje XtalPred-RF

Osnovno orodje XtalPred je bilo predstavljeno leta 2007 in sicer kot spletna aplikacija za napoved kristalizabilnosti proteinov (37). Napoved je temeljila na primerjavi mnogih lastnosti proteina z distribucijo teh lastnosti proteinov v zbirki TargetDB³. Gre torej za podatke o tem, kateri izraženi proteini so bili (uspešno) kristalizirani in kateri ne. Lastnosti, ki so bile napovedane oz. izluščene iz aminokislinskega zaporedja proteinskega konstrukta ter uporabljene za oceno kristalizabilnosti, so:

- **dolžina** – večji proteini so kompleksnejši in pogostejše vsebujejo regije, neugodne za kristalizacijo;
- **izoelektrična točka** – ekstremno kisli in bazični proteini zaradi prevladujočega enoznačnega naboja, posledičnega odboja, težje kristalizirajo, prav tako pa je, statistično gledano, znanih več struktur bazičnih kot kislih/nevtralnih proteinov;
- **hidrofobnost** oz. indeks GRAVY⁴ – povezan je s topnostjo proteina, in sicer večji, kot je, bolj je protein hidrofoben in s tem manj topen v vodnih raztopinah;
- **neurejenost**, opisana kot dolžina najdaljše neurejene regije, napovedane z orodjem DISOPRED (47) – daljše, kot so neurejene regije, manjša je kristalizabilnost;
- **nestabilnost** – indeks nestabilnosti se oceni na podlagi frekvence posameznih dipeptidov znotraj analiziranega zaporedja, za katere je znano, da so povezani z metabolno nestabilnostjo proteina;
- **sekundarna struktura** – večji, kot je delež polipeptidne verige, ki ne tvori elementov regularne sekundarne strukture (napovedana z orodjem PSIPRED (48)), večja je verjetnost za prisotnost neurejenih regij in s tem manjša kristalizabilnost;
- **ovije vijačnice** (*coiled coils*), napovedane z orodjem COILS (49);
- **insercije** glede na homologe v neredundančni zbirki aminokislinskih zaporedij – več, kot jih je in daljše kot so, večja je verjetnost za prisotnost neurejenih zank na površini proteina (insercije so po navadi povezane s prisotnostjo zank na površini); homologe identificira prek iskanja z orodjem PSI-BLAST;
- **površinska entropija**;
- **površinska hidrofobnost**;
- **izbokline na površini** (*surface ruggedness*) – bolj, kot je površina "neravna", težje se tvorijo obsežni medmolekulski kontakti in manjša je kristalizabilnost.

Dodatno je orodje upoštevalo še napoved transmembranskih vijačnic (orodje TMHMM (50)), prisotnost signalnega peptida (orodje RPSP (51)) in prisotnost regij z nizko kompleksnostjo zaporedja (orodje SEG (52)). Na osnovi teh analiz je bila kot rezultat podana ocena kristalizabilnosti, in sicer je bil analiziran protein oz. njegovo aminokislinsko zaporedje uvrščeno v enega od razredov od 1 do 5 (1 pomeni najvišjo kristalizabilnost), t.i. *expert pool crystallizability class*. Dodatno je orodje podalo tudi homologe analiziranega zaporedja iz genomov mikroorganizmov, ki jih je prav tako klasificiralo v enega od petih razredov kristalizabilnosti.

V letu je 2014 je bila predstavljena nadgrajena verzija orodja, imenovana XtalPred-RF (36). Vir empiričnih podatkov je bila zbirka TargetTrack⁵, in sicer 4.924 proteinskih struktur (pozitiven set) ter 21.898 proteinov, ki jih niso uspeli kristalizirati (negativen set), pri obeh pa so upoštevali le proteine dolžine 50–800 aminokislinskih ostankov. Pri orodju XtalPred-RF gre v osnovi za enake napovedi kot pri predhodni verziji, je pa napoved nadgrajena z metodo naključnega gozda, ki poda dodatno oceno *random forest crystallizability class*, v razponu 1–10. Ta je izračunana na osnovi metode naključnih gozdov (*random forests*), to je pristopa strojnega učenja, ki v fazi učenja na osnovi seta podatkov izračuna odločitvena drevesa (drevesna struktura, ki ponazarja relacije med

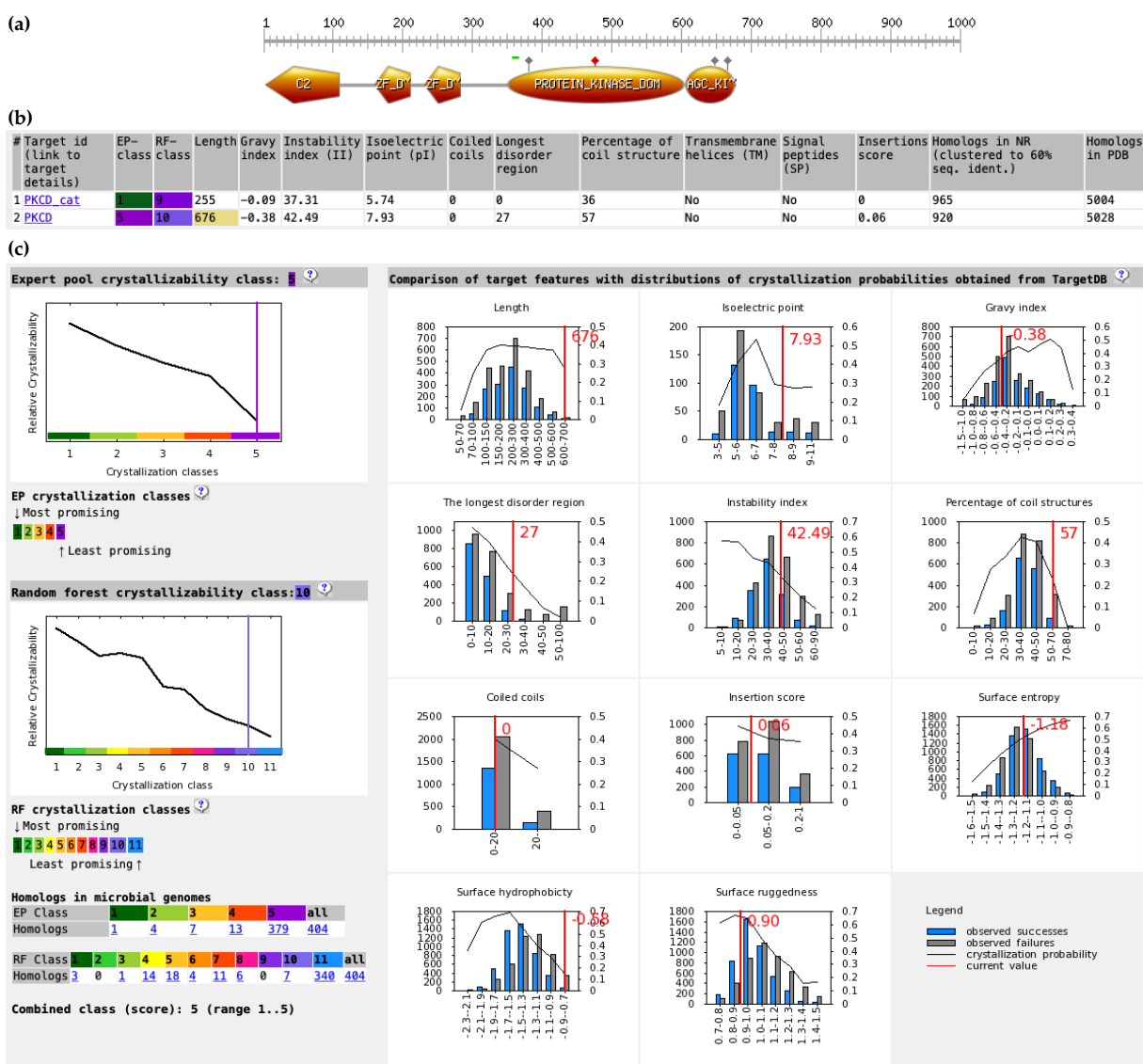
³ Zbirka TargetDB je bila osnovana leta 2004 in je predstavljala centraliziran vir podatkov iz različnih raziskovalnih centrov, ki se ukvarjajo s strukturno genomiko (določanje 3D strukture vseh proteinov, zapisanih v genomu nekega organizma). To so podatki o tem, kateri proteini se kristalizirajo ter o doseženi stopnji poti v cevovodu (*pipeline*) od zapisa DNA do strukture za posamezni protein (45).

⁴ Indeks GRAVY (*G*rand *A*verage *h*ydrophathy) je merilo za hidrofobnost oz. hidrofilitet proteina, enostavno izračunano kot povprečje hidrofobnosti posameznih aminokislinskih ostankov po formuli $i_G = \sum_{j=1}^n \frac{h_{aa_j}}{n}$, kjer je h_{aa} hidrofobnost posameznega aminokislinskega ostanka (vseh v proteinu je n) po lestvici hidrofobnosti Kyte-Doolittle (46). V tej lestvici so aminokislinski ostanki razvrščeni od najbolj (Ile, vrednost 4,5) do najmanj hidrofobnega (Arg, vrednost -4,5).

⁵ Zbirka TargetTrack je naslednica zbirke TargetDB in sicer je nastala z združitvijo podatkov le-te ter zbirke PepcDB. Vse te zbirke je vzdrževala skupnost PSI (*Protein Structure Initiative*), ki je z delovanjem prenehala leta 2017. Zadnja verzija zbirke TargetTrack iz leta 2017 je dostopna v digitalni knjižnici ZENODO.

vhodnimi atributi in izidom), ki jih nato v fazi analize oz. klasifikacije vhodnih podatkov uporabi za izračun modusa razreda.

Primer napovedi kristalizabilnosti za aminokislinsko zaporedje, izračunane z orodjem XtalPred-RF, je prikazan na sliki 1.5. Uporabljeno je bilo aminokislinsko zaporedje človeške protein kinaze C, izooblika δ (PKC δ), in sicer v celotni ter skrajšani obliki (samo katalitična domena). PKC δ je namreč sestavljena iz več regij, nekatere od njih so povezane s fleksibilnimi povezovalnimi regijami. V primeru celotnega zaporedja je orodje XtalPred-RF že samo dolžino prepoznalo kot problematično, tudi izoelektrično točko, k najslabšemu razredu kristalizabilnosti pa sta bistveno prispevali prisotnost neurejenih regij (najdaljša neurejena regija je dolga 27 aminokislinskih ostankov) ter visoka povprečna hidrofobnost. Glede na oba algoritma, klasičnega in RF, je celotno zaporedje uvrščeno v najbolj neobetaven razred. Napoved za skrajšano obliko, ki predstavlja zgolj katalitično domeno, je sicer nekoliko boljša (na primer neurejenih regij naj ne bi bilo več; ni prikazano na sliki) – 1. razred pri klasičnem algoritmu in 9. razred pri RF.



Slika 1.5: Primer napovedi kristalizabilnosti za človeško protein kinazo C δ (PKC δ , Uniprot ID Q05655). (a) Shematski prikaz PKC δ , kot ga prikaže ProSite. (b) Tabelarni prikaz rezultatov analiz celotnega zaporedja (PKCD, ostanki 1–676) in samo katalitične domene (PKCD_cat, ostanki 349–603). (c) Podrobnejši prikaz posameznih analiz za celotno zaporedje PKC δ . Podrobnejši opis je na strani 19.

Vaja 1

Načrtovanje proteinskih konstruktov

Namen vaje

Namen vaje je na primeru enega od proteinov identificirati lastnosti, ki vplivajo na kristalizabilnost, ter na osnovi tega načrtati konstrukt za ta protein s čim višjo kristalizabilnostjo.

Material

Vaja je računalniška, uporabljajo se zgolj prosto dostopne zbirke in spletna orodja. Potrebujete le spletni brskalnik in internetno povezavo.

Postopek dela

Sledite navedenim korakom, rezultate napovedi pa pustite odprte v brskalniku, saj si jih bomo na koncu nekaj skupaj ogledali ter debatirali o njih.

1. **Izberite si poljuben protein.** Lahko gre za protein, ki je bil tema vaše seminarske naloge pri katerem drugem predmetu, je predmet vašega raziskovalnega dela, ...
2. **Poiščite zapis za izbrani protein v zbirki UniProt** (<http://www.uniprot.org>)⁶ ter si ga oglejte. Bodite pozorni na prisotnost domen, povezovalnih regij ipd. V primeru, da v zapisu ni veliko pripisov (anotacij), uporabite orodje **InterPro**, ki vam pomaga pri identifikaciji domen in pomembnih mest (aktivno mesto, motiv za vezavo nekega liganda ipd.). Ti podatki so pomembni za kasnejše načrtovanje krajših konstruktov, saj se izognemo npr. razredu neke domene na polovico in s tem nevarnosti, da se tak protein sploh ne bi zvil.
3. Na osnovi izluščenih značilnosti proteina **pripravite več predlogov konstruktov**, katerih aminokislinska zaporedja si, skupaj s smiselnimi imeni, shranite v običajno tekstno datoteko v formatu FASTA (nekaj osnovnih informacij glede tega formata je na strani 22). Med konstrukti naj bo:
 - celotno aminokislinsko zaporedje izbranega proteina;
 - del tega proteina, ki predstavlja neko zaključeno celoto, na primer katalitični del (pri encimih), zunajcelično regijo (transmembranskih proteinov tipa I) ipd.
4. Vsebinsko datoteke FASTA prekopirajte v okence za vnos zaporedij (slika 1.6), vključite možnost za iskanje bakterijskih homologov, vnesite vaš e-poštni naslov (za vsak primer, če po nesreči zaprete okno, preden so izračuni zaključeni) ter potrdite, da strežnik uporabljate v izobraževalne namene).
5. Počakajte, da se napoved zaključi – to prepoznate tako, da so v tabeli z rezultati (podobno kot na sliki 1.5) prikazana vsa zaporedja, ki ste jih

⁶ Zbirka UniProt je zbirka aminokislinskih zaporedij, ki vsebuje še ogromno dodatnih podatkov v obliki pripisov (funkcija, domene, posttranslacijske modifikacije,...), razdeljena je pa na ročno pregledan del UniProt/SwissProt ter nepregledan del UniProt/TrEMBL, ki vsebuje prevedena nukleotidna zaporedja in zgolj avtomatske pripise.

Slika 1.6: Zaslonski posnetek spletne strani za vnos podatkov v orodje XtalPred-RF.

vnesli. Z klikom na ime konstrukta (enako je imenu zaporedja v datoteki FASTA) pridete do celotne **napovedi kristalizabilnosti** oz. **posameznih molekulskih parametrov**, ki so prispevali k oceni. Oglejte si jih ter jih **povežite z oceno kristalizabilnosti**.

6. V primeru, da so ocene kristalizabilnosti vseh načrtanih konstruktov slabe, pripravite nove, izboljšane konstrukte. Pri tem upoštevajte napovedane molekulске parametre iz prejšnje stopnje, predvsem tiste, ki so povezani z manjšo kristalizabilnostjo, hkrati pa pazite, da bo konstrukt predstavljal neko enoto, ki se bo lahko, vsaj po pričakovanjih, uspešno zvila.
7. Preglejte **bakterijske homologe izbranega proteina**, ki imajo višjo kristalizabilnost, ter jih primerjajte z vašim konstruktom.
8. Napoved kristalizabilnosti naredite še z vsaj **dvema drugima orodjema** iz seznama na strani 17 ter primerjajte rezultate z napovedjo programa XtalPred-RF.

Vprašanja za razmislek

1. Na kateri konec najboljšega konstrukta (v luči kristalizabilnosti) bi bilo najbolj smiselno umestiti heksahistidinsko oznako?
2. Če bi želeli heksahistidinsko oznako po čiščenju odcepiti, katero peptidazo bi uporabili? Načrtajte aminokislinsko zaporedje konstrukta s heksahistidinsko oznako in cepitvenim mestom za visoko specifično peptidazo.

Dodatne informacije

Format FASTA

Format FASTA je format datotek za nukleotidna ali aminokislinska zaporedja. Razvila sta ja David J. Lipman in William R. Pearson leta 1985, in sicer kot del programskega paketa FASTA, ki je program za iskanje podobnih zaporedij. Format se še danes splošno in veliko uporablja, saj je enostavno berljiv za računalnike in hkrati tudi za ljudi.

Gre za običajno tekstno datoteko, ki ima lahko običajno končnico (*file extension*) *txt*, pogosto pa se za označevanje vsebine (nukleotidna/aminokislinska zaporedja) uporablja ena od naslednjih končnic: *fasta*, *fna*, *ffn*, *ffa*, *frn* ali kar *seq*. Datoteke lahko oblikujemo s katerim od programov, ki omogočajo delo s tekstnimi datotekami, na primer Notepad (okolje Windows) ali TextEdit (okolje macOS). Tovrstnih datotek ne pripravljamo v kompleksnejših urejevalnikih besedil, kot je, na primer Microsoft Word, saj slednji v datoteko shranijo še veliko drugih ukazov, namenjenih oblikovanju (sicer lahko v teh programih datoteke shranimo kot običajne neoblikovane tekstne datoteke, a to samo zaplete delo).

V datoteko lahko shranimo eno ali več zaporedij. V primeru enega zaporedja datoteko oblikujemo tako, da se prva vrstnica začne z znakom $>$, ki mu sledi ime zaporedja, samo zaporedje pa je v eni ali več vrsticah, ki sledijo. Kot primer si oglejmo zaporedje za človeški interferon γ :

```
>hIFNg
MKYTSYILAFQLCIVLGS LGCYCQDPYVKEAENLK KYFNAGHSDVADNGTLF
LGILKNWKEESDRKIMQSQIVSFYFKLFKNFKDDQSIQKSVETIKEDMNVKF
FNSNKKKRDDFEKLTNYSVTDLNVQRKAIHELIQVMAELSPA AKTGKRKRSQ
MLFRGRRASQ
```

V zgornjem primeru je *hIFNg* ime zaporedja, samo zaporedje pa je izpisano v naslednjih vrsticah. V isto datoteko formata FASTA lahko shranimo tudi več nukleotidnih ali aminokislinskih zaporedij. Tak primer, ki vsebuje zaporedje iz zgornjega primera ter še zaporedje prvih 50 aminokislinskih ostankov istega proteina ločeno (kot ime za slednje je uporabljeno *hIFNg_1-50*), je prikazan tukaj:

```
>hIFNg
MKYTSYILAFQLCIVLGS LGCYCQDPYVKEAENLK KYFNAGHSDVADNGTLF
LGILKNWKEESDRKIMQSQIVSFYFKLFKNFKDDQSIQKSVETIKEDMNVKF
FNSNKKKRDDFEKLTNYSVTDLNVQRKAIHELIQVMAELSPA AKTGKRKRSQ
MLFRGRRASQ
>hIFNg_1-50
MKYTSYILAFQLCIVLGS LGCYCQDPYVKEAENLK KYFNAGHSDVADNGT
```


2

Kristalizacija

Priprava kvalitetnih kristalov (velikost, sposobnost sipanja X-žarkov do čim višje ločljivosti, nizka stopnja mozaičnosti, stabilnost, nezraščenost) pogosto predstavlja ozko grlo v procesu določanja kristalne strukture neke makromolekule. K uspehu lahko veliko pripomore smotrno načrtovanje proteinskih konstruktov, kar je podrobneje opisano v razdelku 1 na strani 11, prav tako pa tudi čim bolj obsežni **presejalni kristalizacijski testi** (*crystallization screens*). Čeprav se zdi, da včasih kristalizacija tako velikih in kompleksnih molekul, kot proteini so, meji na magijo, temu vsekakor ni tako – gre le za to, da vseh procesov (še) ne znamo natančno razložiti, njihova obsežnost, kompleksnost in medsebojna povezanost pa so tolikšne, da je napovedovanje ustreznih kristalizacijskih pogojev zaenkrat nemogoče.

Proteinski oz. makromolekulski kristali na splošno se od kristalov soli in majhnih organskih molekul precej razlikujejo. Nekaj posebnosti:

- tipična velikost kristalov, ki se uporabljajo v rentgenski kristalografiji makromolekul, je ponavadi od 50 μm do 0,5 mm;
- povezava med videzom proteinskih kristalov in njihovo difrakcijsko kvaliteto ni enoznačna (na pogled grd kristal lahko dobro sipa, lep pa sploh ne);
- proteinske molekule v kristalu tvorijo medmolekulske kontakte in omogočajo urejeno zlaganje novih molekul v kristalno mrežo;
- medmolekulske interakcije v kristalu so navadno redke in šibke, posledica je velika občutljivost makromolekulskih kristalov (mehanska in temperaturna občutljivost);
- makromolekulski kristali so ponavadi mehki in bolj podobni gostemu skupku želeja kot trdnim delcem,
- prostor med makromolekulami v kristalu zapolnjuje topilo (voda in topilenci), ki v povprečju predstavlja 40–60 % kristala;
- medmolekulske interakcije v kristalu so specifične za posamezen protein in pogosto vključujejo nabite aminokislinske ostanke na površini (pozor pri površinski mutagenezi!);
- za isto makromolekulo/protein lahko pri različnih pogojih dobimo različne kristalne oblike (te se lahko pojavijo celo v isti kristalizacijski kapljici), ki imajo navadno različno difrakcijsko kvaliteto;
- ista kristalna oblika se lahko pojavlja v različnih kristalnih načinih (s tem pojmom opisujemo videz kristala, kjer so pri enaki kristalni struktura posamezne kristalne stranice različno izražene).

2.1 Proces kristalizacije

Da dobimo kristale katerekoli spojine, tudi najpreprostejše, moramo načeloma doseči stanje prenasíčenosti (supernasičenost, supersaturacija, *supersaturation*). To je termodinamsko nestabilno stanje (metastabilno stanje), ob vračanju v stabilno stanje pa del topljenca tvori medmolekulske kontakte, kar se na makroskopskem nivoju kaže kot nastanek kristalinične (urejeno stanje – kristali) ali amorfne faze (neurejeno stanje – oborina). Potek oz. končen rezultat tega procesa je odvisen od lastnosti vzorca in pogojev: temperatura,

pH, topljenci (pufri, soli, ...) in njihova koncentracija, gostota in viskoznost raztopine, tip površine, električna in magnetna polja, vibracije ... Ko govorimo o proteinih, lahko ti parametri vplivajo na njegovo topnost, zmožnost tvorbe ugodnih medmolekulskih kontaktov ter konformacijo proteina in njegovih topilu izpostavljenih regij. Kristalizacija je tako izrazito multiparametričen proces (nanj vpliva več parametrov), ki ga v osnovi lahko razdelimo na tri faze:

1. **nukleacija**, to je spontan nastanek prvih urejenih skupkov. Nukleacija zahteva višji nivo supersaturacije kot rast kristala. Včasih je pogoje, pri katerih pride do nukleacije, težko reproducirati, lahko pa so celo takšni, da nadaljnja rast kristala ni optimalna – v teh primerih uporabljamo vnos kristalnih jeder (sejanje, *seeding*);
2. **rast kristala**, to je urejeno nalaganje makromolekul na površino nukleacijskega jedra oz. nastajajočega kristala. Višji kot je nivo supersaturacije, večja je hitrost rasti kristala, hkrati pa je večja tudi verjetnost, da bo med rastjo prišlo do napak v kristalni mreži (slabša difrakcijska kvaliteta kristalov!);
3. **ustavitev rasti**, za katero so lahko vzrok pomanjkanje makromolekul v raztopini, napake med rastjo kristala, "zastropitev" površine kristala z drugimi molekulami in/ali razgradnja makromolekul.

Stanje prenasíčenost dosežemo z večanjem koncentracije biološke makromolekule (zaradi izhlapevanja topila) in/ali z večanjem koncentracije ene ali več spojin, ki zmanjšajo topnost proteina – takim spojinam rečemo **obarjalna sredstva** ali **precipitanti**, opisana so v razdelku 2.3, skupaj s kristalizacijskimi testi, katerih komponenta so. V praksi lahko prenasíčenost dosežemo na več načinov, kar je na kratko opisano v naslednjem razdelku, opis dogodkov na molekularnem nivoju pa je v primeru ene od teh izvedb opisan v razdelku 2.2.1.

2.2 Praktična izvedba kristalizacije

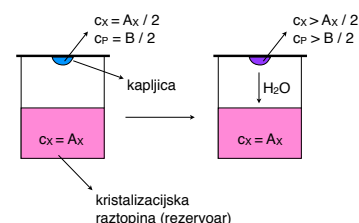
Stanje, pri katerem se vrednosti parametrov (pH, topljenci in njihova koncentracija ...) kažejo v prenasíčenosti raztopine, lahko dosežemo na več osnovnih načinov:

- **difuzija prek hlapov** (*vapour diffusion*), kjer kapljico, sestavljeno iz raztopine proteina in kristalizacijskega reagenta, v zaprtem sistemu pustimo počasi ekvilibrirati proti večji količini enakega kristalizacijskega reagenta;
- **dializa v majhnem merilu** (mikrodializa), ki omogoča počasno višanje koncentracije precipitanta in drugih komponent kristalizacijskega reagenta v raztopini proteina;
- **šaržne metode** (*batch methods*), pri katerih raztopino proteina zmešamo s kristalizacijsko raztopino, kapljico pa nato prekrijemo s plastjo parafinskega olja;
- **difuzija prek stične površine** (*interface diffusion*), kjer gradient koncentracije proteina in kristalizacijskega reagenta ustvarimo s pomočjo počasne difuzije komponent obeh raztopin preko njune stične površine v kapilari.

Najpogostejše se uporablja metoda difuzije preko hlapov, ki obstaja v treh izvedbah: **viseča kapljica** (*hanging drop*), **sedeča kapljica** (*sitting drop*) in **sendvič** (*sandwich drop*). Način sedeče kapljice je shematsko prikazan na sliki 2.1. Vsem trem izvedbam je skupna počasna vzpostavitev ravnotežja med kapljico in raztopino rezervoarja.

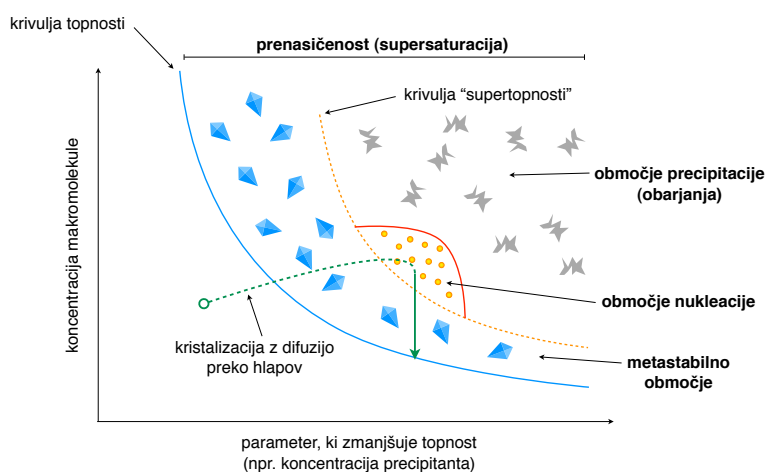
2.2.1 Kristalizacija po principu viseče kapljice

Na poenostavljen način lahko potek kristalizacijskega eksperimenta ponazorimo s t. i. **faznim diagramom** (*phase diagram*), imenovanim tudi diagram topnosti (slika 2.2). V tovrstnih diagramih koncentracijo proteina v zmesi



Slika 2.1: Shematski prikaz kristalizacije z metodo difuzije preko hlapov na način viseče kapljice. Na levi je prikazana kapljica na kristalizacijski plošči tkoj po nastavitvi eksperimenta, pri čemer je bila za pripravo kapljice uporabljena enaka volumna raztopine proteina (s koncentracijo B) in kristalizacijske raztopine komponente X (s koncentracijo A_x). Topilo (voda) počasi izhlapeva iz kapljice v smeri vzpostavitve ravnotežja med kapljico in rezervoarjem, pri čemer v kapljici naraščata tako koncentracija proteina kot koncentracija ostalih komponent raztopine (desno).

nanesemo na ordinatno os (y -os), na absciso (x -os) pa na primer koncentracijo precipitanta. Predstavljajmo si hipotetičen primer, ko med kristalizacijskim eksperimentom koncentracija precipitanta v kristalizacijski kapljici narašča do točke, kjer je koncentracija proteina na meji topnosti v novih (spremenjenih) pogojih. Ob nadaljnjem naraščanju koncentracije precipitanta pridemo v t. i. metastabilno območje, kjer lahko ob primernih pogojih ($pH \dots$) pride do heterogene in homogene spontane nukleacije. Če do tega torej dejansko pride in se začnejo molekule proteina nalagati na urejen način, dobimo kristal, ki raste na račun dodajanja novih molekul iz supernasičene raztopine, dokler ne pridemo nazaj na krivuljo topnosti (ravnatežno stanje). V tej točki s površine kristala oddisociira enako število molekul, kot se jih na novo doda, kristal torej navzven ne raste in se ne raztaplja. V primeru, da pa pogoji v sistemu ne omogočajo nukleacije, lahko pride bodisi do obarjanja proteina (v skrajnem primeru pride do denaturacije), ločitve faz (supernasičena raztopina proteina ter raztopina proteina na meji topnosti) ... lahko pa se tudi ne zgodi nič navzven opaznega – ostanemo v metastabilnem stanju ali celo pod krivuljo topnosti (kapljica ostane "prazna").



Slika 2.2: Poenostavljen prikaz faznega diagrama za kristalizacijo preko difuzije hlapov. Majhen krog predstavlja začetno stanje, konec puščice pa končno stanje (ravnatežno stanje na krivulji topnosti).

2.3 Kristalizacijski presejalni testi

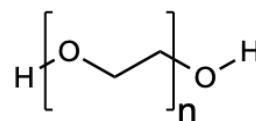
Kot uvodoma omenjeno je izredna raznolikost in kompleksnost makromolekul ter procesov, ki potekajo med kristalizacijo, razlog, da je trenutno nemogoče napovedati ustrezne pogoje, ki bi omogočali nastanek nukleacijskih jeder ter rast kristalov proteinov. Problem iskanja takšnih pogojev rešujemo pretežno empirično, in sicer s preizkušanjem večjega števila pogojev oz. presejanjem (*crystallization screening*), pri katerem uporabljamo mešanice (**kristalizacijske raztopine**¹), ki ponavadi vsebujejo:

- **precipitant** – spojina, ki zmanjša topnost makromolekul;
- **pufer** – določa vrednost pH ter s tem vpliva na naboj na površini makromolekul;
- **sol** – ioni soli vplivajo na topnost makromolekul in tvorbo medmolekulskih kontaktov.

Kot precipitanti se najpogosteje uporabljajo polietilenglikoli (PEG, slika 2.3), sledijo pa spojine: amonijev sulfat (klasično obarjalno sredstvo za proteine), 2,4-metilpentandiol (MPD), etilenglikol, glicerol in drugi alkoholi (2-propanol, 1,4-butandiol ...), natrijev citrat, kalij-natrijev tartrat ter še številne druge.

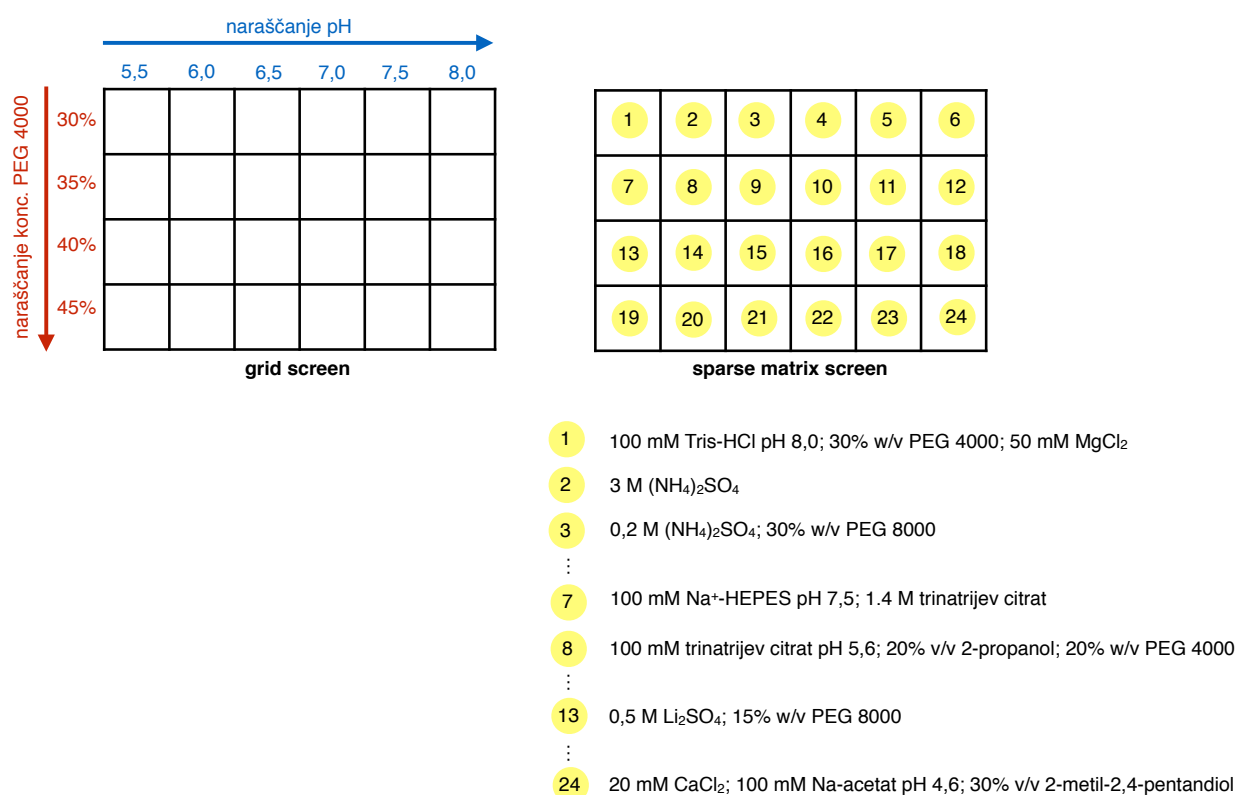
V zvezi s sestavo kristalizacijskih raztopin pri presejalnih testih obstaja več strategij:

¹ Pogosto jim v žargonu rečemo kar kristalizacijski pufri (*crystallization buffers*), kar tehnično ni pravilno, saj vsebujejo še kaj drugega kot pufre, nekatere od teh raztopin pa pufrske komponente sploh ne vsebujejo.



Slika 2.3: Polietilenglikol (PEG) je polietar, pri katerem se ponavlja osnovna enota glikola. Dolžino verige ponavadi opišemo kot povprečno molekulsko maso n , npr. PEG 200, PEG 8000 ipd. Nizkomolekularni PEG (npr. PEG 200) so tekoči, visokomolekularni PEG (npr. PEG 2000 in višji) pa so vodotopne trdne snovi bele barve, njihove vodne raztopine pa so brezbarvne.

- **mrežni testi** (*grid screens*) – sistematična variacija manjšega števila parametrov, npr. % PEG/pH testi, ki kljub omejenosti pogosto nudijo uporabne informacije o kristalizabilnosti proteina (primer je na sliki 2.4);
- **testi na osnovi redkih (raztresenih) matrik** (*sparse matrix screens*) – ti testi so osnovani na pogojih oz. kristalizacijskih raztopinah, ki so se v preteklosti (torej na empirični osnovi) že izkazali kot primerni za številne proteine, tipičen primer takšnega testa je Crystal Screen avtorjev Jancarik in Kim (53) (primer je na sliki 2.4);
- **polni / nepopolni faktorialni testi** (*full / incomplete factorial design screens*) – ti testi nudijo podrobno analizo (medsebojnih!) vplivov različnih spremenljivk na kristalizacijo, zaradi preobsežnosti kompletnih testov pa pogosto uporabljamo nepopolne, pri katerih analizo omejimo na vplive prvega reda ter ustrezno uravnotežimo kristalizacijski načrt (posamezni vplivi – faktorji se pojavljajo z enako frekvenco po celotnem testu).



Slika 2.4: Primer *grid* (levo) in *sparse matrix* (desno) kristalizacijskega testa v merilu 6×4 (24 pogojev).

Kristalizacijske teste lahko nastavljammo "na roko", kar zahteva precej časa, prav tako potrebujemo večje količine proteina (približno 0,5-1 μ l za posamezni pogoj oz. kapljico). Tak način po navadi uporabljamo v kasnejših stopnjah optimizacije kristalizacijskega pogoja, ki ga identificiramo s pomočjo obširnih kristalizacijskih testov, nastavljenih s pomočjo **robotskih sistemov**. Ti ob uporabi že pripravljenih komercialno dostopnih kristalizacijskih testov omogočajo testiranje ogromnega števila pogojev (npr. 1000) ob minimalni porabi vzorca (kapljice velikosti 100 nl ali celo manjše). Pri tem gre v veliki večini za način sedeče kapljice, saj je to najlažje avtomatizirati prek robotskega pipetiranja raztopin.

2.3.1 Ocenjevanje rezultatov kristalizacijskih testov

Po nastavitvi presejalnih kristalizacijskih testov moramo plošče redno pregledovati. Frekvenca pregledovanja je kmalu po nastavitvi višja (vsak dan), kasneje jo zmanjšamo. Pregledovanje lahko izvajamo ročno, kar je, sploh v

primeru obširnih kristalizacijskih testov, zelo zamudno, zato se danes pogosto uporabljajo avtomatizirani robotski sistemi, ki z zeleno frekvenco poslikajo vse kristalizacijske kapljice. Slike nato pregledamo sami, je pa napredna programska oprema že sposobna prepoznati, ali je v neki kapljici kristal, kar nam prihrani zamudno pregledovanje tisočih slik.

Za ocenjevanje kristalizacijskih testov se, tako pri ročnem kot avtomatskem pregledovanju, pogosto uporablja ocenjevalni sistem, kjer so ocene v območju 0–9. Enomestna števila so izbrana zato, da jih lahko med seboj brez dvomnosti kombiniramo. Pomen oznak:

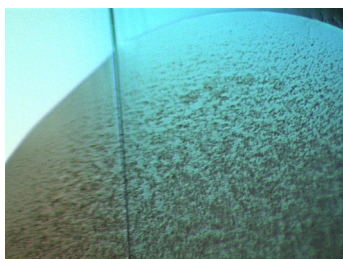
- **0** pomeni prazno kapljico, kar pomeni, da stanja prenasíčenosti nismo dosegli;
- **1** pomeni prisotnost delcev v raztopini, ki niso kristali, lahko gre za kakšne drobne smeti, ki so med nastavljanjem padle v kapljico;
- **2–9** pomenijo prisotnost oborine, ločitev faz in kristale, primeri in podrobna razdelitev pa je prikazana na sliki 2.5 na strani 30.

Oborine so lahko kristalinične (gre pravzaprav za izredno drobne kristale oz. strukture, ki spominjajo na kristale – pri teh ne gre za denaturacijo proteina) ali pa amorfne, pri slednjih ponavadi pride do denaturacije proteina. Tretji tip oborine je t.i. gelatinozna oborina, ki jo je zaradi prosojnosti nekoliko težje opaziti.

V primeru, da po daljšem času po nastavitvi kristalizacijskega testa (na primer, 2–3 tedne) ostane večina kapljic praznih oz. v njih ni prisotne oborine, je to znak, da prenasíčenosti večinoma nismo dosegli. V tem primeru je priporočeno test ponoviti pri višji koncentraciji proteina. Obratno pa, če je v veliki večini kapljic prisotna oborina, še posebej, če se ta pojavi že kmalu po nastavitvi testa, to pomeni, da smo uporabili raztopino proteina s previsoko koncentracijo, zato velja test ponoviti z nižjo koncentracijo proteina.

2.4 Zanimiva dodatna literatura

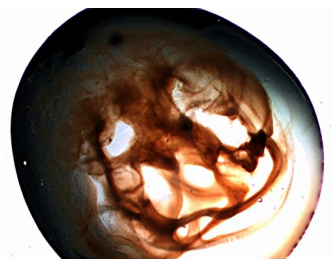
Zgodovinski pregled kristalizacije proteinov od leta 1840 do današnjega časa: Richard Giegé, A Historical Perspective on Protein Crystallization from 1840 to the Present Day, *The FEBS journal*, 280(24), 2013, str. 6456–6497



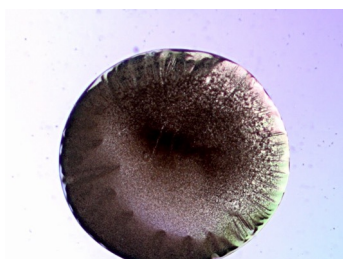
2 - majhna količina oborine



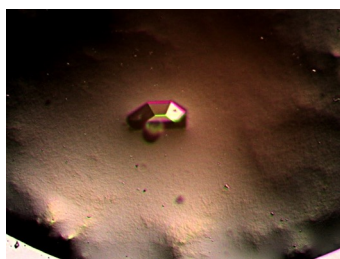
3 - veliko oborine



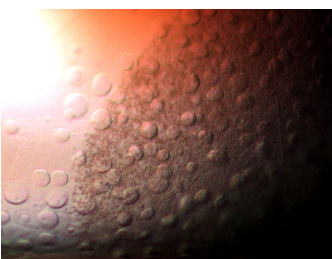
3 - veliko oborine



3 - veliko oborine in kožica na površini kapljice, ki ponavadi predstavlja denaturiran protein



4 - gelatinozna oborina



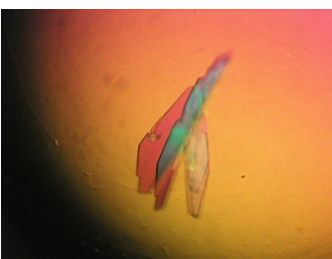
5 - ločitev faz



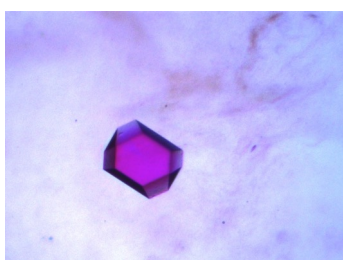
6 - sferuliti



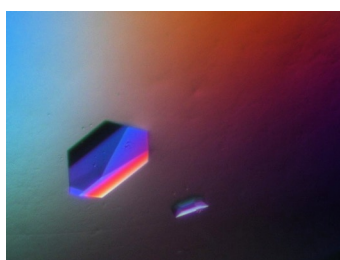
7 - paličasti (1D) kristali , združeni v snope



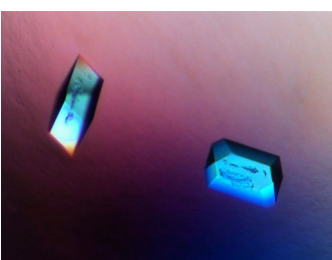
8 - ploščati (2D) kristali



9 - (3D) kristal



9 - (3D) kristala



9 - (3D) kristala

Slika 2.5: Primeri tipičnih rezultatov kristalizacijskih testov, ki bi jih lahko ocenili v območju med 2-9.

Vaja 2

Kristalizacija lizocima

Namen vaje

Namen vaje je na primeru kristalizacije lizocima v praksi spoznati pripravo presejalnega kristalizacijskega testa z omejenim številom pogojev (24) z metodo difuzije preko hlapov (način viseče kapljice) ter analizirati, kako sprememba pogojev v kristalizacijski raztopini vpliva na nastanek in rast kristalov.

Material

Seznam kemikalij

- **založne raztopine precipitantov:**
 - 4 M NaCl,
 - 60 % (w/v) PEG 3350,
 - 60 % (w/v) PEG 4000,
 - 60 % (w/v) PEG 6000;
- **založne raztopine precipitantov:**
 - 1 M natrijev acetat, pH 4,5,
 - 1,5 M TRIS/HCl, pH 8,0;
- **deionizirana voda;**
- **raztopina proteina** – 60 µl raztopine kokošjega lizocima s koncentracijo 100 mg/ml (v 50 mM natrijevem acetatu, pH 4,5).

Uporabljali boste še naslednji drobn material:

- **kristalizacijske plošče** s 24 vdolbinicami (konfiguracija 6×4) za pripravo kristalizacijskih nastavkov po principu viseče kapljice,
 - **silikonizirana krovna stekelca**²,
 - silikonsko mast za zatesnitev stikov med robom vdolbinic in krovnim stekelcem,
 - pipete in nastavke,
 - mikrocentrifugirke.
-

² Pri navadnih, nesilikoniziranih stekelcih, se kapljica ponavadi razleže po površini, pri silikoniziranih pa zaradi hidrofobne površine stekelca ostane večinoma bolj kompaktne izbočene oblike. Razmerje med površino kapljice in njenim volumnom vpliva na proces kristalizacije prek spremenjene hitrosti izhlapevanja vode oz. hitrosti ekvibracije kapljice z okolico.

Postopek dela

Priprava načrta za presejalni kristalizacijski test

Izdelajte načrt za kristalizacijski eksperiment/test, v okviru katerega boste preizkusili 24 različnih pogojev (kristalizacijskih raztopin). Posamezen pogoj naj sestavlja set vrednosti naslednjih parametrov:

- sestava kristalizacijske raztopine (raztopina v rezervoarju):
 - koncentracija in tip precipitanta (PEG ali natrijev klorid),
 - ionska jakost (koncentracija NaCl),
 - pH (4,5 ali 8,0, vmesne vrednosti lahko dobite z mešanjem obeh pufrov, a točne vrednosti pH ne boste poznali);

- koncentracija lizocima v kapljici (lizocim lahko nekoliko razredčite, lahko kar z vodo, vsekakor pa ne preveč, saj v tem primeru kristali ne bodo zrasli v času izvajanja vaje);
- razmerje med volumnom raztopine lizocima in kristalizacijskega reagenta v viseči kapljici.

Sestavo raztopin oz. razporeditev vrednost spremenljivk vpišite v tabelo

2.1. Primer enega pogoja oz. set vrednosti spremenljivk:

- koncentracija lizocima 100 mg/ml,
- sestava kapljice: 2 μ l raztopine lizocima + 2 μ l kristalizacijske raztopine,
- sestava kristalizacijske raztopine: 50 mM natrijev acetat, pH 4,5, 1 M NaCl, 25 % (w/v) PEG 4000.

Pripravite si tudi načrt pipetiranja, pri čemer upoštevajte, da naj bo končni volumen kristalizacijske raztopine v posamezni luknjici kristalizacijske plošče v območju 0,5–1 ml. Za pripravo kristalizacijske raztopine iz zgornjega primera v volumnu 1 ml bi zmešali:

- 50 μ l 1 M raztopine natrijevega acetata pH 4,5,
- 416 μ l 60 % (w/v) PEG 4000,
- 250 μ l 4 M NaCl
- ter dopolnili z deionizirano vodo do 1 ml (torej 284 μ l vode).

	1	2	3	4	5	6
A						
B						
C						
D						

Tabela 2.1: Tabela za načrtovanje presejalnega kristalizacijskega testa.

Priprava in nastavitev kristalizacijskega testa

1. Označite 24 mikrocentrifugirk (posamezna mikrocentrifugirka ustreza eni luknjici na plošči s 24 luknjicami).
2. V označenih mikrocentrifugirkah pripravite kristalizacijske raztopine (po 0,5–1 ml) v skladu z izdelanim načrtom. Pri pipetiranju raztopin PEG bodite pozorni, saj so zelo viskozne. Potrebno je počasno pipetiranje, prav

tako pa je priporočljivo, da odrežete konico nastavka za pipeto in s tem naredite večjo odprtino.

3. Raztopine dobro premešajte z vibracijskim mešalom, mikrocentrifugirke pa na koncu kratko centrifugirajte, da vsebino zberete na dnu.
4. Na robove vseh luknjic na kristalizacijski plošči nanesite silikonsko mast.
5. Pripravite raztopino (ali več raztopin raztopin) lizocima z želeno koncentracijo (lizocim redčite s 50 mM Na-acetat pH 4,5).
6. Raztopine iz posameznih mikrocentrifugirk odpipetirajte v ustrezne luknjice na kristalizacijski ploščici.
7. Pripravite kapljice: na sredino krovnih stekelc odpipetirajte raztopino lizocima (npr. 2 μ l) ter ji dodajte enak volumen raztopine iz posamezne luknjice.
8. Krovna stekelca obrnite nad luknjice ter jih dobro zatesnite s pomočjo nanesene silikonske masti.
9. Kapljice preglejte pod mikroskopom. Kristali lizocima se lahko pojavijo že v prvi minuti!
10. Kapljice preglejte še po 20 minutah ter si zabeležite opažanja.
11. Na podlagi rezultatov vašega testa, ki vsebuje 24 pogojev, ter rezultatov kolegov v vaši skupini analizirajte vpliv posameznih faktorjev na izid kristalizacijskega eksperimenta ter določite najprimernejši pogoj oz. več pogojev, ki so dali velike in posamične kristale.

Vprašanja za razmislek

1. Kakšna je pravzaprav vloga natrijevega klorida in kako je povezana z njegovo koncentracijo v kristalizacijski raztopini?
2. Kako koncentracija precipitanta vpliva na število kristalizacijskih jeder? Menite, da je smiselno imeti ogromno število majhnih kristalov ali majhno število zelo velikih kristalov? Kaj so prednosti enih in drugih?
3. Lahko pH kristalizacijske raztopine ter izoelektično točko (pI) lizocima povežete z izidom kristalizacije pri posameznih pogojih?
4. V nekaterih kapljicah so se kristali začeli pojavljati že skoraj takoj po nastavitvi. Kaj to pomeni v luči faznega diagrama, prikazanega na sliki [2.2](#) na strani [27](#)?

Snemanje in procesiranje difrakcijskih podatkov

Pridobitev kristala proteina oz. biološke makromolekule je, predvsem v smislu ozkega grla v procesu določanja strukture, vsekakor spodbuden rezultat, za nadaljevanje projekta pa je ključnega pomena, da dobljeni kristali dejansko sipajo X-žarke in to do čim višje (ali vsaj zadostne) ločljivosti. Edini način, da to preizkusimo, je difrakcijska analiza, v kateri kristal pripravimo (namestimo na ustrezno podporo) ter z njim posnamemo nekaj difrakcijskih slik. Že iz nekaj slik lahko precej izvemo o njegovi kvaliteti, prav tako pa lahko približno ocenimo parametre, kot so dimenzije in koti osnovne celice ter prostorsko skupino (to naredimo v okviru t. i. procesiranja difrakcijskih podatkov). V primeru, da je kvaliteta difrakcije zadostna in obeta prispevek k določitvi strukture, posnamemo popoln set difrakcijskih podatkov, kar pomeni, da zajamemo signale oz. vse reflekse znotraj asimetrične enote recipročnega prostora v luči prisotne simetrije (povezana s prostorsko skupino). Procesiran set teh podatkov, ki vsebuje Millerjeve indekse refleksov, intenzitete ter oceno nedoločenosti, nato služi kot osnova za reševanje strukture.

3.1 Priprava kristalov za difrakcijsko analizo

Kristal moramo pred snemanjem pripraviti¹. To pomeni, da ga vzamemo iz kapljice, v kateri je zrasel, namestimo na neko podporo ter ga primerno zaščitimo. Običajno uporabljamo dva pristopa, in sicer nameščanje (*mounting*) kristalov v:

- **kapilaro** (hramba in snemanje pri sobni temperaturi),
- **zanko** (hramba pri temperaturi tekočega dušika, snemanje pa po navadi pri temperaturi 100 K).

Nameščanje kristalov v kapilaro je starejši od obeh pristopov. Kristal z majhno količino matične raztopine posrkamo v kapilaro, odvečno tekočino popivnemo, da se kristal nalepi na steno kapilare (sicer se lahko kristal med premikanjem zanke prosto premika), v kapilaro na enem ali obeh koncih posrkamo še malo matične raztopine (da se kristal ne izsuši) ter kapilaro na obeh straneh zatesnimo. Kristal je tako zaščiten pred izsušitvijo, kapilaro pa lahko namestimo neposredno na nastavek goniometriške glave in posnamemo difrakcijski vzorec (pri tem pazimo, da je kristal v snopu X-žarkov). Ta pristop ima sicer kar nekaj slabosti, od katerih je najbolj očitna hitrejša nastajanje obsežnih radiacijskih poškodb (radikali, ki nastajajo med obsevanjem, hitro difundirajo in reagirajo s proteinom). S takim načinom namestitve kristala je povezana tudi večja termična gibljivost molekul v kristalu, posledično se to lahko kaže na nižji ločljivosti in večji nedoločenosti nekoliko bolj fleksibilnih regij/zank makromolekul. Prav tako ta način ni primeren za občutljive in nestabilne kristale.

Danes se v veliki večini primerov uporablja pristop zamrzovanja kristalov in snemanje difrakcijskih podatkov pod krio-pogoji, in sicer pri temperaturi približno 100 K. Bistvena prednost je močno zmanjšan obseg radiacijskih poškodb, kar omogoča snemanje podatkov z daljšimi časi ekspozicije. Pri tem načinu kristal ujamemo v majhno zanko ali mrežasto podporo, nato pa

¹ Obstaja pristop t.i. difrakcijske analize *in situ*, kjer kristal izpostavimo X-žarkom kar v kapljici, v kateri je zrasel. Pristop je bolj ali manj omejen na začetno karakterizacijo kristala, za kvaliteten set difrakcijskih podatkov je potrebno izvesti snemanje na običajen način

zanko potopimo v tekoči dušik (temperatura -196°C) in s tem zamrznemo kristal ter majhno količino matične raztopine, ki ga obkroža. Ko je kristal enkrat zamrznjen, ga moramo ves čas hraniti pri podobni nizki temperaturi, najbolje tako, da je potopljen v tekoči dušik. Lahko ga tudi hranimo v parah nad tekočim dušikom (pri tem načinu moramo biti pozorni na to, da na zanki in kristalu ne pride do kondenzacije in tvorbe kristalov vode iz zraka, saj tudi slednji sipajo X-žarke in nam njihov difrakcijski vzorec delno zakrije našega). Prav tako snemanje difrakcijskih podatkov poteka na način, ki omogoča neprestano hlajenje kristala s curkom par tekočega dušika, ki ima temperaturo 100 K. Primeri zank, ki se uporabljajo za t.i. krio-nameščanje kristalov, so prikazani na sliki 3.1. Zanke so sicer nameščene na podporo v obliki igle, slednja pa je nameščena v kovinsko osnovo, ki omogoča magnetno pritrdjevanje na nastavek goniometriške glave med snemanjem difrakcijskih podatkov.

Pri namestitvi kristalov v krio-zanke in zamrzovanju moramo biti pozorni, da je kristal v raztopini, ki ob hitrem zamrzovanju (*flash-cooling* oz. *flash-freezing*) ne tvori kristalov vode, ki poškodujejo sicer fragilne makromolekulske kristale. To dosežemo tako, da uporabimo krio-protektante oz. krio-zaščitne spojine (*cryo-protectants*) kot komponento raztopine, v kateri je kristal. Najpogostejše uporabljane spojine s takim učinkom so glicerol, etilenglikol ter nizkomolekularni polietilenglikoli (običajno v koncentracijskem območju od 10–30 (w/v) %). Poznamo dva načina njihove uvedbe, in sicer glede na to, če so ali niso prisotne v osnovni kristalizacijski raztopini oz. matični raztopini kristala:

- če so krio-zaščitne spojine tam prisotne lahko kristal enostavno ulovimo v zanko in zamrznemo;
- če krio-zaščitne spojine v matični raztopini niso prisotne, moramo pripraviti kristalizacijsko raztopino s tako sestavo, kot je bila uporabljena za kristalizacijo (lahko nekoliko povišamo koncentracijo precipitanta), kot dodatno sestavino pa uvedemo eno od krio-zaščitnih raztopin. Nato kristal v matični raztopini ujamemo v zanko, ga prenesemo v majhno kapljico kristalizacijske raztopine z dodano krio-zaščitno komponento in šele nato zamrznemo.

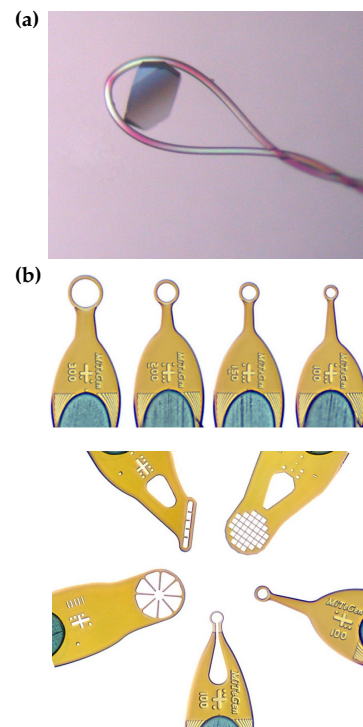
Izbira primerne krio-zaščitne raztopine je sicer predmet optimizacije oz. poizkušanja. Kristale, zamrznjene v zankah, hranimo v ohlajenih Dewar-posodah, in sicer jih naložimo v posebne nastavke, ki preprečujejo njihove fizične poškodbe ter omogočajo robotizirano nameščanje v snop X-žarkov na sinhrotronskih žarkovnih linijah.

3.2 Snemanje difrakcijskih podatkov

Snemanje difrakcijskih podatkov z X-žarki običajno poteka na sinhrotronih ali na t. i. hišnih difraktometrih (*in-house diffractometers*). Njihova prednost je sicer enostavna dostopnost, a sinhrotroni jih prekašajo v marsičem: izredno fokusiran žarek, visok fluks fotonov ter možnost izbire valovne dolžine. Dodatno prednost predstavlja možnost eksperimentov s spremljanjem časovnega poteka procesov.

Danes difrakcijske podatke večinoma snemamo pod krio-pogoji, pri tem pa je sam potek snemanja na sinhrotronu in hišnem difraktometru v osnovi podoben:

1. Kristal je treba iz tekočega dušika prenesti do nastavka goniometriške glave, kar na hišnih virih naredimo ročno, na sinhrotronih pa za to poskrbi robotiziran sistem.
2. Sledi centriranje kristala, pri čemer se z nastavitvijo vijakov goniometriške glave (hišni difraktometri) oz. prilagajanjem položaja goniometriške glave kristal namesti v tak položaj, da bo ob vrtenju ostal v snopu X-žarkov.

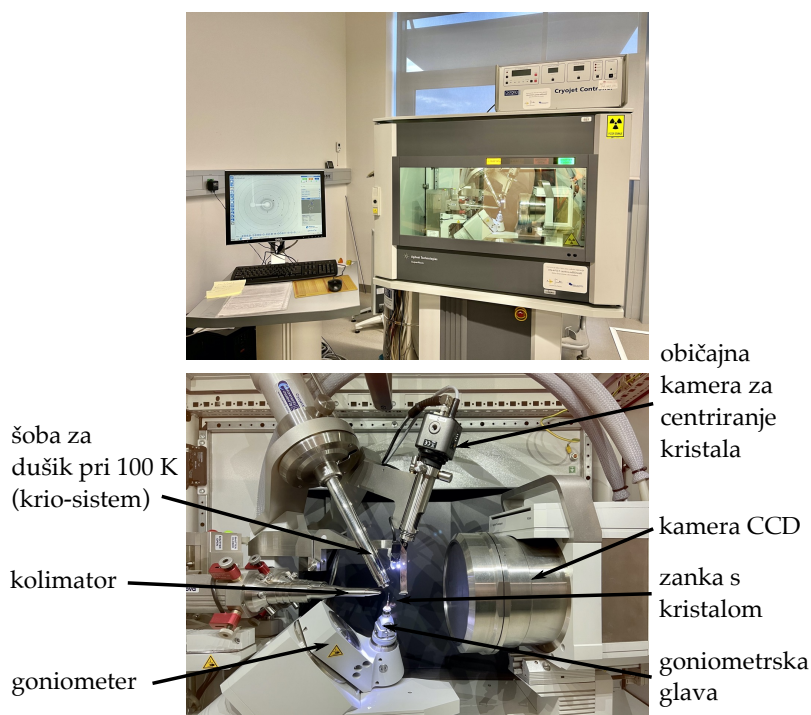


Slika 3.1: Primeri zank za nameščanje makromolekulskih kristalov. (a) Proteinski kristal, ujet v zanko klasične oblike proizvajalca HamptonResearch. (b) Alternativne zanke, mrežice in podobne strukture za nameščanje kristalov proizvajalca Mitegen. Vir slik: spletne strani proizvajalcev.

3. Izpostavitve (ekspozicija) kristala X-žarkom in zajemanje difrakcijskih podatkov prek detektorja (danes ponavadi kamera CCD ali detektorji na osnovi štetja fotonov), ob hkratni rotaciji kristala okoli izbrane osi. Natančno rotacijo omogoča naprava, imenovana goniometer. Po navadi v okviru posamezne ekspozicije zavrtimo kristal za $0,5\text{--}1^\circ$.
4. Po končanem snemanju difrakcijskih podatkov lahko kristal ponovno shranimo v tekočem dušiku za kasnejše ponovno snemanje.

Sinhrotron nam prek nastavitve valovne dolžine X-žarkov zraven snemanja t. i. nativnih podatkov omogoča tudi podatke, kjer je prišlo do anomalnega sipanja. Tega hišni viri načeloma ne omogočajo, razen ko imamo v kristalu prisoten element, ki pri določeni valovni dolžini anomalno sipa (na primer pri valovni dolžini $1,54 \text{ \AA}$ (bakrova anoda) lahko izkoristimo šibek anomalni signal žveplovih atomov). Anomalne podatke lahko uporabimo za rešitev faznega problema, na primer v okviru anomalne disperzije pri eni valovni dolžini (*Single-wavelength Anomalous Dispersion*, SAD).

Na UL FKKT je nameščen hišni difraktometer Agilent SuperNova s CCD-detektorjem Atlas. Ne temelji na rotirajoči anodi, temveč gre za mikrofokusni vir, kjer je snop elektronov usmerjen na zelo majhno območje statične anode. Difraktometer ima dve anodi, in sicer molibdenovo ($\lambda = 0,71 \text{ \AA}$) in bakrovo ($\lambda = 1,54 \text{ \AA}$); za proteinsko kristalografijo se uporablja slednja. Sam sistem je sicer prvenstveno namenjen difrakcijski analizi kristalov manjših molekul in anorganskih spojin, v primeru dovolj velikih makromolekulskih kristalov pa je v omejenem obsegu uporabljen tudi za te. Difraktometer je prikazan na sliki 3.2.



Slika 3.2: Difraktometer Agilent SuperNova na UL FKKT. Na zgornji fotografiji je na desni vidna eksperimentalna komora, na vrhu katere je krmilnik za krio-sistem, levo je kontrolni računalnik. Na spodnji fotografiji je približan del komore s krio-sistemom, kolimatorjem X-žarkov, goniometrom ter kamero CCD.

3.3 Procesiranje difrakcijskih podatkov

Prvo oceno o difrakcijski kvaliteti dobimo že ob vizualni analizi difrakcijskih slik. Želimo visoko ločljivost, odsotnost obročev zaradi difrakcije na

morebitnih kristalih ledu, visoko ostrino difrakcijskih pik ter nizko mozaičnost. Več nam razkrije procesiranje difrakcijskih podatkov, ki so v obliki slik z dodanimi meta-podatki (geometrija, čas ekspozicije ipd.) – procesiranje izvedemo takoj, ko imamo posnetih nekaj difrakcijskih slik. To nam omogoča, da določimo najverjetnejšo prostorsko skupino ter pripravimo strategijo za nadaljnje snemanje. Območje rotacije kristala, za katero moramo posneti difrakcijske podatke, da bomo imeli popoln podatkovni set, je namreč povezano s simetrijo oz. prostorsko skupino.

Procesiranje izvajamo s specializiranimi programi, ki lahko že z uporabo avtomatskih postopkov sami hitro in dobro analizirajo difrakcijske podatke, večinoma pa kljub temu izvedemo reprocesiranje ob prilagoditvi oz. ročni izbiri parametrov. Na splošno ločimo naslednje stopnje procesiranja:

1. **indeksiranje**, kar pomeni lociranje difrakcijskih pik in določitev kristalne mreže ter približnih parametrov osnovne celice;
2. **ocena mozaičnosti**,
3. **odstranitev neželenih regij** difrakcijskih slik, kot so sence in obroči zaradi difrakcije kristalov ledu;
4. **piljenje** parametrov osnovne celice;
5. **integracija**, v okviru katere se določijo intenzitete (I) sipanih žarkov in ozadje (σI);
6. **analiza** integriranih podatkov z namenom določitve prostorske skupine;
7. **skaliranje in zlivanje** (*scaling, merging*), imenovano tudi redukcija podatkov, kjer integrirane vrednosti z različnih difrakcijskih slik združimo v en set strukturnih faktorjev.

V okviru procesiranja programi izračunajo nekaj parametrov, ki so kazalniki kvalitete difrakcijskih podatkov. Zraven podatkov o samem kristalu jih navadno navedemo v t.i. **Tabeli 1**, ki je obvezen del članka o na novo določeni makromolekulski strukturi in vsebuje:

- prostorsko skupino, parametre osnovne celice ter valovno dolžino uporabljenih X-žarkov,
- vsebino osnovne celice oz. asimetrične enote (to dejansko določimo med reševanjem same strukture),
- najvišjo ločljivost,
- Wilsonov B-faktor, ki predstavlja splošen izotropni B-faktor,
- vrednosti faktorjev R pri zlivanju podatkov,
- redundantnost (število neodvisnih opažanj) in popolnost podatkov,
- jakost signala, izražena kot razmerje $I/\sigma I$.

Te parametre navedemo tako za celotno območje ločljivosti kot tudi za lupino z najvišjo ločljivostjo (*highest resolution shell*).

Vaja 3

Snemanje in procesiranje difrakcijskih podatkov

Vaja, ki obravnava te korake v procesu določanja strukture, je demonstracijska. Med vajo bo predstavljeno:

- nameščanje kristalov lizocima, pripravljenih pri vaji 1, v zanko in zamrzovanje v tekočem dušiku,
- snemanje difrakcijskih podatkov kristalov lizocima na hišnem viru X-žarkov,
- procesiranje nekaj začetnih difrakcijskih slik kristala lizocima, določitev parametrov osnovne celice in prostorske skupine, pregled kazalnikov kvalitete difrakcijskih podatkov.

Reševanje, piljenje in validacija modela strukture

4.1 Povezava med difrakcijskimi podatki in strukturo

V prejšnjem poglavju je opisano, da s kristalom primerne kvalitete posnamemo set difrakcijskih slik, ki ga nato sprocesiramo – določimo parametre osnovne celice (stranice a, b, c in kote α, β, γ), določimo prostorsko skupino¹ ter izračunamo Millerjeve indekse hkl reflektov in njihove intenzitete I_{hkl} skupaj z oceno nedoločenosti intenzitete σ_{hkl} . Postopek je precej avtomatiziran. Za rešitev strukture potrebujemo elektronsko gostoto. Če strukturne faktorje in elektronsko gostoto zapišemo kot Fourierjevo serijo, ugotovimo, da med njimi obstaja tesna povezava. Elektronska gostota ni nič drugega kot Fourierjeva transformacija strukturnih faktorjev, z drugimi besedami, obstaja enačba, s katero lahko difrakcijske podatke oz. iz njih določene strukturne faktorje pretvorimo v sliko osnovne celice in njene vsebine. Vendar nam pri tem za vsak strukturni faktor manjka ena komponenta, saj lahko izmerimo zgolj intenziteto posameznih reflektov, ne pa tudi popolnega strukturnega faktorja. Amplituda slednjega je proporcionalna kvadratnemu korenu izmerjene intenzitete. Povezava med amplitudo strukturnega faktorja in intenziteto je naslednja:

$$|F_{hkl}| \propto \sqrt{I_{hkl}} \quad (4.1)$$

Povezava med elektronsko gostoto $\rho(x, y, z)$ in difrakcijskim vzorcem (V pomeni volumen osnovne celice, F_{hkl} posamezni strukturni faktor) je:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx + ky + lz)} \quad (4.2)$$

Če torej poznamo intenzitete, poznamo tudi amplitude strukturnih faktorjev, nimamo pa podatka o njihovih fazah.

4.2 Določitev faz sipanih žarkov

Za reševanje t. i. **faznega problema** (določitev faz) je na voljo več metod. V grobem jih delimo na metode **eksperimentalnega faziranja**, kjer faze izračunamo prek anomalnega sipanja (na primer na selenovih atomih), večkratne izoformne zamenjave ipd., ter **metodo molekulske zamenjave**. Metode faziranja oz. njihovo teoretično ozadje so podrobneje predstavljene v drugi literaturi², zato jih tu ne bomo posebej obravnavali. V nadaljevanju sledi zgolj kratek opis pristopa molekulske zamenjave, ki je predmet računalniške vaje.

4.2.1 Molekulska zamenjava

Pri metodi molekulske zamenjave uporabimo že znano podobno strukturo, ki jo preko prileganj aminokislinskih zaporedij poiščemo v zbirki že določenih struktur. Pattersonovo funkcijo strukture, ki jo rešujemo, sistematično primerjamo s Pattersonovo funkcijo modela (znana struktura). Na ta način določimo rotacijsko (intramolekulski vektorji) in translacijsko komponento (intermolekulski vektorji), na osnovi katere lahko izračunamo približne faze, na osnovi teh pa karto elektronske gostote, ki ponavadi ni najboljša, saj so

¹ V praksi velja, da smo lahko prepričani o pravilnosti protorske skupine šele takrat, ko granja in piljenje modela strukture potekata uspešno.

² Dober, izčrpen in moderen pregled je na voljo v učbeniku Biomolecular Crystallography avtorja Bernharda Ruppja Bernhard Rupp, *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, New York, 2010.

faze zgolj približne. Sledi gradnja modela in piljenje, v času katerega natančneje izračunamo strukturne faktorje ter posledično dobimo boljšo karto elektronske gostote.

4.3 Asimetrična enota in njena sestava

Pri določanju strukture z X-žarkovno difrakcijo določamo strukturo asimetrične enote (*asymmetric unit, asu*), to je najmanjši del kristala, na osnovi katerega lahko z aplikacijo simetrijskih operacij prostorske skupine³ generiramo osnovno celico in s translacijo slednje po prostoru celoten kristal. Asimetrična enota lahko vsebuje biološko relevanten skupek⁴ molekul (monomer ali oligomer), le del njega ali pa celo več biološko relevantnih skupkov (spet, lahko gre za monomer ali oligomer). Tehnično gledano nas slednje med določanjem strukture niti ne zanima toliko, je pa pomembno, koliko kopij neke polipeptidne verige je prisotnih v asimetrični enoti. Pri molekularni zamenjavi nas tako npr. zanima, koliko kopij določene polipeptidne verige moramo poiskati in umestiti v asimetrično enoto.

Pri določanju števila molekul oz. kopij neke polipeptidne verige (ali nukleinske kisline) si pomagamo s t. i. **Matthewsovim koeficientom** (imenovan tudi Matthewsov volumen, V_M). Gre za količnik med volumnom osnovne celice (v Å) in v njej vsebovane mase proteina (Da); Matthewsov koeficient ima tako enoto Å³ Da⁻¹. Lahko ga tudi obravnavamo kot količnik med volumnom asimetrične enote in v njej vsebovane mase proteina, kar je ekvivalentno prej napisanemu. Volumen osnovne celice je na podlagi parametrov osnovne celice (stranice in koti) enostavno izračunati, podobno velja za asimetrično enoto. Če poznamo prostorsko skupino, vemo, koliko asimetričnih enot vsebuje osnovna celica, od tukaj do volumna asimetrične enote je zgolj preprost račun. Prav tako poznamo maso proteina (saj poznamo tudi njegovo aminokislinsko zaporedje), preostane nam le, da za različno število kopij proteina na asimetrično enoto ($N = 1, 2, 3 \dots$ kopije) izračunamo Matthewsov koeficient in izračunane vrednosti evaluiramo na osnovi možnih. Namreč, Matthewsov koeficient se ponavlja giblje med 1,62 do 3,53 Å³ Da⁻¹ in je direktno povezan z deležem topila v asimetrični enoti (osnovni celici, kristalu), ki se giblje v območju od 28 % do 78 % (to so ekstremne vrednosti). Na podlagi tega program predlaga najbolj verjetno razmerje N/asu .

³ V primeru proteinskih kristalov govorimo o rotacijah, translacijah in vijačnih oseh.

⁴ V zbirki PDB se za to uporablja termin *biological assembly*.

4.4 Gradnja modela strukture v praksi

Po določitvi faz sipanih žarkov, kar nam omogoča izračun karte elektronske gostote, gradnja modela v praksi poteka navadno tako:

1. **avtomatska gradnja**⁵, med katero računalniški program umesti dele ali celo večino polipeptidne verige/verig v elektronsko gostoto, pri tem pa kot dodaten vhodni podatek uporabi aminokislinsko zaporedje ter (v primeru molekulske zamenjave) lahko tudi strukturo, uporabljeno za samo molekulsko zamenjavo;
2. izmenjevanje med:
 - ročno gradnjo in piljenjem (*refinement*) modela v grafičnem okolju,
 - avtomatskim piljenjem⁶, med katerim pride tudi do izračuna novih kart elektronske gostote ter izračuna faktorjev R in R_{free} .

⁵ Moderni programi za avtomatsko gradnjo modela so precej dobri in nam lahko ob kvalitnih eksperimentalnih podatkih (visoki ločljivosti) že sami zgradijo zelo dober model strukture. Vsekakor je potrebno take modele ročno pregledati, analizirati in izpiliti.

⁶ Pregled o ozadju in teoriji avtomatskega piljenja je na voljo v drugi literaturi (56).

Postopek gradnje in piljenja je torej iterativen, saj nam ustrezno delno prilagojen model omogoča izračun novih, natančnejših faz, na osnovi katerih izračunamo natančnejšo elektronsko gostoto itd. Med piljenjem upoštevamo tudi splošne fizikalno-kemijske/stereokemijske parametre (dolžine vezi in koti med njimi ...), v model dodamo molekule vode in druge molekule (ligande) in/ali ione. Značilnost procesa piljenja je tudi to, da nismo osredotočeni zgolj na posamezni aminokislinski ostanek in elektronsko gostoto, ki mu ustreza, temveč gledamo širše – analiziramo neposredno okolico. V primeru aminokislinskih ostankov glutamin in asparagin npr. ni vseeno, kako je konec

stranske skupine (karbonilna in amino skupina) obrnjen, saj ima posamezna stran različen potencial oz. preference do interakcij z drugimi skupinami v neposredni bližini. Povzeto, naš cilj je, da model strukture ustreza tako eksperimentalnim podatkom kot tudi stereokemijskim parametrom, kar bo najbolje ilustrirano v okviru dveh računalniških vaj:

- **vaje 4a**, ki obravnava ročno piljenje sicer že zgrajene strukture ribonukleaze Sa, z namenom seznanitve s programskim orodjem Coot za delo z modelom strukture in karto elektronske gostote;
- **vaje 4b**, ki vsebuje celoten postopek reševanja strukture človeške karbonske anhidraze od sprocesiranih difrakcijskih podatkov do končnega modela strukture, pri vaji pa se uporabljajo veščine, pridobljene na primeru iz vaje 4R1.

4.5 Ocenjevanje kvalitete modela strukture

Pogosto uporabljano merilo kvalitete modela, ki ga izračunamo med piljenjem, je faktor R , predstavlja pa statistično merilo ujemanja opaženih - eksperimentalno izmerjenih (*observed*, F_{obs} ali F_o) in iz modela izračunanih (*calculated*, F_{calc} ali F_c) amplitud strukturnih faktorjev:

$$R = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|} \quad (4.3)$$

Teoretična vrednost faktorja R sega od 0 (popolno ujemanje med izračunanimi in izmerjenimi intenzitetami) do približno 0,6 (set izmerjenih intenzitet, ki jih primerjamo s setom naključnih intenzitet – takšna struktura bi bila *za v smeti*). V praksi na splošno velja, da je pri strukturah s slabšo ločljivostjo (>3 Å) dopustna vrednost faktorja R nad 0,2, v primeru struktur visoke ločljivosti (< 2 Å) pa pod 0,2. Med piljenjem je pomemben tudi izračun faktorja R_{free} , saj faktor R ne zajema pristranskosti, ki jo vnesemo med piljenjem. Razlika med izračunom vrednosti R in R_{free} je v tem, da pri slednjem upoštevamo le naključno izbran set reflektov (5–10 % vseh reflektov), ki smo jih v postopku piljenja izpustili. V idealnem primeru bi na osnovi piljenega modela izračunali enak F_{calc} , tako za izpuščeni kot pri piljenju uporabljeni set reflektov; v tem primeru bi bila vrednost R_{free} enaka vrednosti R . V primeru prekomernega piljenja (t. i. *overfitting*⁷), kjer bi upoštevali preveč šuma in bi atome na silo umeščali v položaje brez neke osnove (brez definirane elektronske gostote), bi piljen model za izpuščen set napovedal vrednosti F_{calc} precej slabše kot za piljen set. V praksi se smatra za sprejemljivo, če je razlika med R_{free} in R manjša od 0,05–0,1.

⁷ Pri tem gre pravzaprav za to, da imamo v našem modelu preveč parametrov kot jih lahko utemeljimo z eksperimentalnimi podatki.

4.6 Validacija modela strukture

Validacija modela strukture pomeni, da ocenimo zanesljivost modela in sicer tako v smislu eksperimentalnih podatkov (difrakcijski podatki) kot tudi znanih fizikalnih in kemijskih parametrov. Pravzaprav validacijo modela strukture (oz. vsaj nekaj vidikov) delamo že med piljenjem, kjer spremljamo veliko število parametrov, na primer:

- faktorja R in R_{free} kot ujemanje med iz modela izračunano elektronsko gostoto ter elektronsko gostoto, izračunano na podlagi eksperimentalnih podatkov;
- dolžine vezi in kote med njimi, prav tako spremljamo dihedralne kote φ in ψ (naš cilj je, da se ne nahajajo v nedovoljenih regijah Ramachandranovega diagrama);
- sterične ovire (*sterical clashes*), kjer gre za to, da so si posamezni atomi ali skupine atomov preblizu;
- nekovalentne interakcije, kjer preverimo, ali so stranske skupine amino-kislinskih ostankov orientirane tako, da lahko tvorijo ugodne interakcije, če so seveda na primerni razdalji. Podobno velja tudi za interakcije med aminokislinskimi ostanki in ligandi⁸.

⁸ Tipični primer je orientacija distalnega dela stranske skupine asparagina in glutamina, kjer ima vsaka stran (CO in NH_2) specifično preferenco po tvorbi nekovalentnih interakcij z drugimi, sosednjimi skupinami).

Veliko od zgoraj navedenega lahko moderni programi naredijo sami, tudi med avtomatskim piljenjem, a vsak avtomatsko generiran model strukture je treba ročno pregledati. Velja, da je uporaba programov po principu črne škatle lahko izredno nevarna in zavajajoča. Vsekakor je validacija ključen korak, ki ga moramo narediti pred funkcijsko analizo modela strukture, saj model strukture, ki ni skladen tako z eksperimentalnimi podatki kot fizikalno-kemijskimi omejitvami, nima praktične vrednosti.

Vidiki validacije so, tudi v matematičnem smislu, obdelani v številni znanstveni literaturi, priporočeno branje je zgoščen pregledni članek *Validation of protein crystal structures*, avtorja G. J. Kyewegta (57).

Vaja 4a

Ročno piljenje modela strukture v grafičnem okolju

Vaja je osnovana na priročniku za delo s programom Coot, avtorja Paula Emsleya, ki je prosto dostopen na spletu: <https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/cool/web/tutorial/tutorial.pdf>.

Namen vaje

Namen vaje je na primeru ročnega piljenja modela strukture ribonukleaze Sa spoznati principe piljenja strukture v grafičnem okolju v realnem prostoru. Veščine, pridobljene pri tej vaji, so potrebne za vajo 4R2 (stran 51), ki obravnava celoten postopek reševanja strukture, od sprocesiranih difrakcijskih podatkov do končnega modela, vključno z ročnim piljenjem.

Material

Vaja je računalniška. Za izvedbo so potrebne datoteke z vhodnimi podatki, ki jih dobite pri asistentu:

- [rnaSa.pdb](#) z modelom strukture ribonukleaze Sa, ki pa ni popoln,
- [rnaSa.mtz](#), ki vsebuje sprocesirane difrakcijske podatke ter strukturne faktorje ter omogoča izračun karte elektronske gostote.

Nekaj osnovnih informacij o ribonukleazi Sa je v nadaljevanju, sledi pa opis programa Coot, ki je potreben za izvedbo vaje.

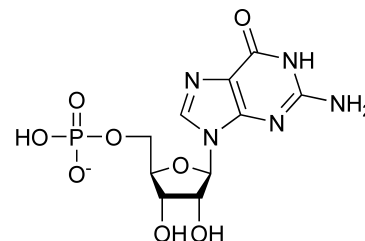
Uvodne informacije o ribonukleazi Sa

Ribonukleaza Sa je iz bakterije *Streptomyces aureofaciens*, predstavniki rodu *Streptomyces* pa so vir številnih tetraciklinskih antibiotikov. Te bakterije so po Gramu pozitivne, imajo visok delež parov GC v genomu, živijo pa pretežno v prsti in razpadajočih rastlinskih tkivih. Imajo značilen vonj po zemlji, ki ga daje hlapen metabolit geozmin.

Delali boste z modelom strukture ribonukleaze Sa (RNaza Sa) iz omenjenega organizma. Ta encim je pravzaprav za gvanil-specifična ribonukleaza, pri cepitvi nastane nukleozid 3'-fosfat in 3'-fosfooligonukleotid, ki se konča z G-P. Ločljivost difrakcijskih podatkov, uporabljenih pri določevanju strukture, je približno 1,8 Å. RNaza Sa je bila kristalizirana v pufru, ki je med drugim vseboval sulfatne ione (SO_4^{2-}) in 3'-gvanozin monofosfat (3'-GMP; slika 4.1).

Programska oprema

Od programske opreme je potreben program **Coot**⁹, ki je trenutno eden najpogostejše uporabljenih programov za gradnjo in piljenje modelov strukture bioloških makromolekul na svetu. Avtor programa je Paul Emsley (58; 5), ki program velikodušno daje v brezplačno uporabo. Spletna stran programa

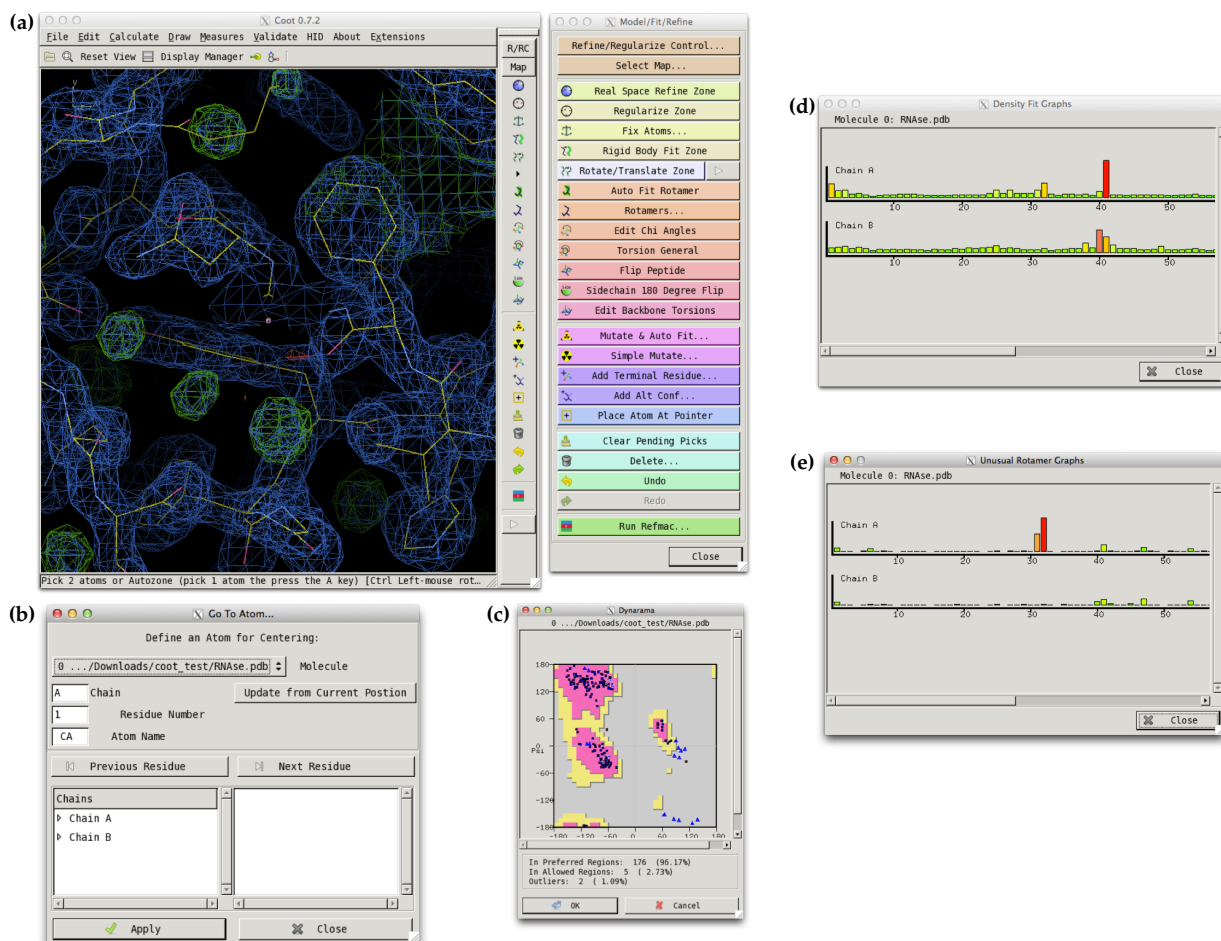


Slika 4.1: Struktura 3'-gvanozin monofosfata (3'-GMP).

⁹ Kratica Coot pomeni *Crystallographic Object-Oriented Toolkit*, hkrati pa je to angleško ime za ptico, ki živi tudi v naših krajih in jo poznamo po imenu liska. To je tudi razlog, da je liska maskota programa.

je dostopna na naslovu <https://www2.mrc-lmb.cam.ac.uk/personal/pemley/coot/>; na tej strani je program na voljo za prenos. Teče na vseh pogosto uporabljenih operacijskih sistemih, instalacija pa včasih ni popolnoma enostavna. Če si želite program namestiti na domači računalnik, vam bo pomagal asistent.

Grafični vmesnik programa Coot je sestavljen iz več oken, ki so ločena, zato je pri delu zelo pripravnno uporabljati dva zaslona. V primeru dela na zgolj enem, pa moramo upoštevati, da del ekrana namenimo oknom programa, namenjenim za specifične analize ali manipulacije (slika 4.2).

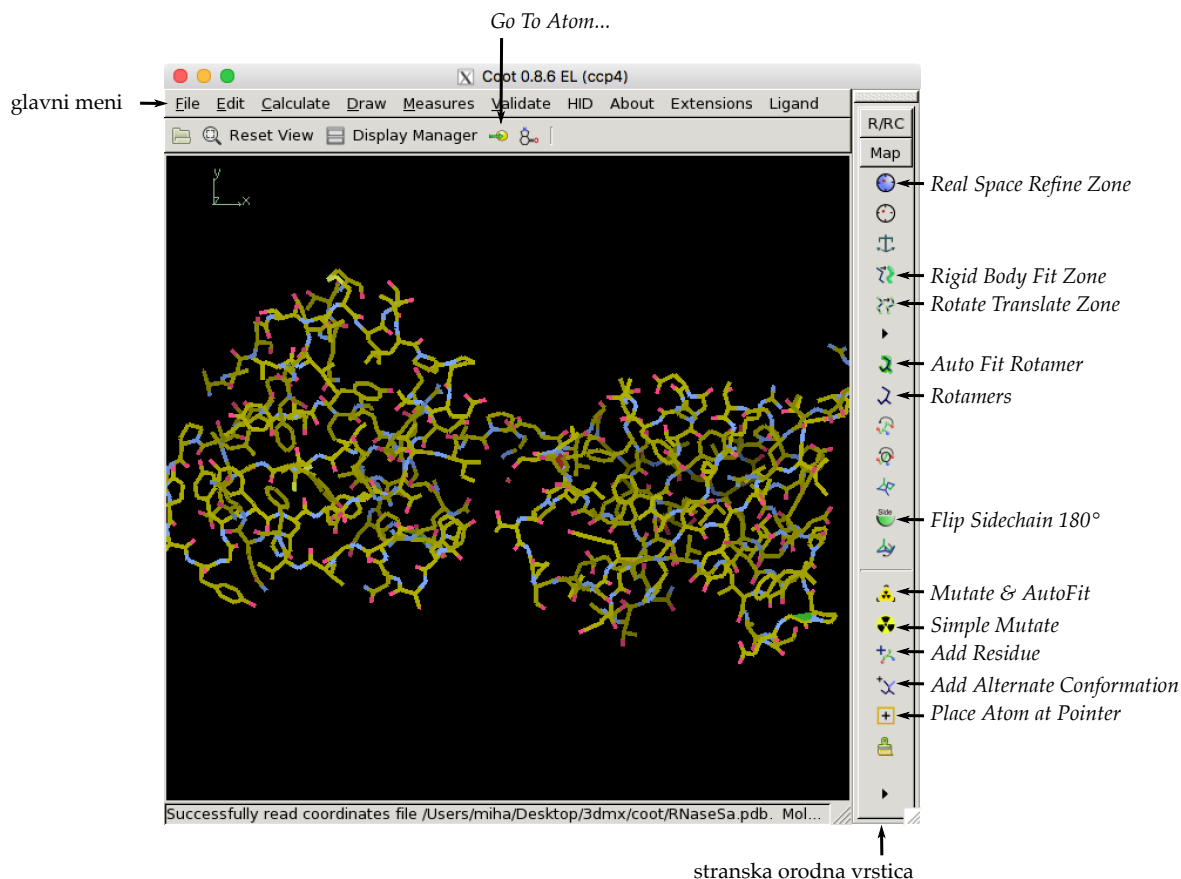


Slika 4.2: Grafični vmesnik programa Coot. Prikazana sta okni za ogled in manipulacijo modela strukture ter karte elektronske gostote ter okno z najpogostejše uporabljanimi orodji (a), okno za pomikanje po aminokislinskem ali nukleotidnem zaporedju oz. verigi/verigah (b), Ramachandranov diagram (c), ter okni za analizo ujemanja modela strukture z elektronsko gostoto (d) in analizo rotamerov (e).

Postopek dela

Odpiranje datotek in osnovne manipulacije

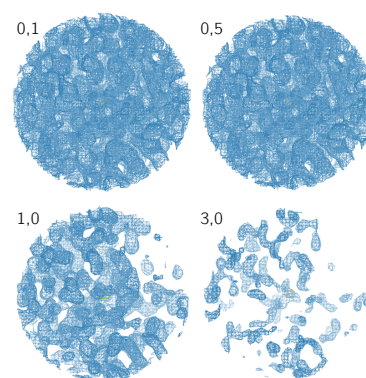
1. Zaženite program Coot in naložite koordinate atomov modela strukture ribonukleaze Sa: **File** ► **Open Coordinates ...** → navigirajte do datoteke **rnaSa.pdb** ter potrdite izbiro. V glavnem oknu programa se odpre model strukture (slika 4.3).
2. Preizkusite načine manipuliranja modela strukture v prostoru:
 - rotacija – držimo levi gumb miške in jo premikamo;
 - translacija – držimo levi gumb miške in jo premikamo, hkrati pa na tipkovnici držimo tipko Alt;



Slika 4.3: Glavno okno programa Coot z odprtim modelom strukture ribonukleaze Sa. Nekatera orodja, ki so omenjena v posameznih korakih vaje, so označena.

- približevanje/oddaljevanje – držimo levi gumb miške in jo premikamo (za oddaljevanje v desno ali proti sebi, za odmikanje pa v levo ali stran od sebe).
3. Razmislite, koliko polipeptidnih verig je v asimetrični enoti ter ali predstavljajo biološko relevanten oligomer. Pomagajte si s pripisi v zbirki UniProt.
 4. Center rotacije (označen z rožnato kockico) prenesemo na posamezni atom tako, da bodisi kliknemo z levim gumbom miške na ta atom ob hkratnem pritisku na tipko Alt, ali pa na ta atom kliknemo s sredinskim gumbom miške. Ob tem se nam izpišeta oznaki atoma in aminokislinskega ostanka, npr. CA /43 VAL/A, kar pomeni atom C α aminokislinskega ostanka Val43 v verigi A.
 5. Po aminokislinskih ostankih se lahko na enostaven način premikamo tudi tako, da v orodni vrstici neposredno nad prikazom modela strukture kliknemo na gumb **Go To Atom ...** → odpre se novo okno, v katerem lahko izberemo posamezni aminokislinski ostank v odprti datoteki PDB (mimogrede, hkrati imamo lahko odprtih več datotek PDB, v tem primeru moramo v tem oknu izbrati zeleno).
 6. Odprite datoteko, ki vsebuje strukturne faktorje za prikaz elektronske gostote: **File > Auto Open MTZ** → navigirajte do datoteke **rnaSa.mtz** ter potrdite izbiro. Odpreta se dve karti elektronske gostote, in sicer:
 - kompozitna karta elektronske gostote $2F_o - F_c$, ki je obarvana modro (privzet način barvanja prve karte elektronske gostote, ki jo odpremo v programu);

- diferenčna karta elektronske gostote $F_o - F_c$, ki jo uporabljamo za detekcijo napak v modelu, in sicer ločimo dvoje različnih načinov prikaza:
 - pozitivne regije so obarvano zeleno, to pomeni, da v našem modelu (uporabljen za izračun F_c) nekaj manjka, a eksperimentalni podatki kažejo, da je nekaj vseeno tam (F_o); $|F_o| > |F_c|$;
 - negativne regije so obarvane rdeče, to pomeni, da je v našem modelu (uporabljen za izračun F_c) nekaj, za kar ni eksperimentalne podlage (F_o) oz. tam ni ničesar, kar bi prispevalo k sipanju žarkov; $|F_o| < |F_c|$.
7. Karta elektronske gostote je prikazana v nekem radiju od središča rotacije (rožnata kockica). Ta radij lahko spremenite v izbirniku **Edit > Map Parameters**, kjer povečate **Map Radius**. Pozor, povečevanje obsega znatno zmanjša hitrost prikazovanja! Identificirajte oblake elektronske gostote, ki ustrezajo proteinu (dobro definirana elektronska gostota) ter kanale topila v kristalu (ni definirane elektronske gostote).
8. Pri karti elektronske gostote lahko spreminjate konturiranje (najenostavneje je s premikanjem miškega kolesčka, lahko tudi s pritiskom na tipki + oz. – na tipkovnici), pri čemer dejansko nastavljamo r.m.s.d. vrednost, pri kateri je prikazana elektronska gostota (ponavadi prikazana kot mreža). Na primer pri 1 bi bila mreža narisana na mestih funkcije $\rho(x, y, z)$ v prostoru, kjer so vrednosti za eno standardno deviacijo večje od povprečne. Z večanjem vrednosti za konturiranje (npr. na 3 ali celi 4) bomo tako videli samo tiste regije, kjer je elektronska gostota res velika (npr. kovinski ioni); z večanjem najprej izginejo tiste regije, kjer je elektronska gostota manjša, npr. (delno) fleksibilne regije. Po drugi strani, npr. z manjšanjem pod npr. 0,5, bo šum postajal bolj izrazit in bomo posledično težko ločili, kateri deli prikazane elektronske gostote predstavljajo molekulo, kateri pa so zgolj šum. Primer karte elektronske gostote $2F_o - F_c$ ribonukleaze Sa pri različnih konturiranjih je prikazan na sliki 4.4.



Slika 4.4: Karta elektronske gostote ribonukleaze Sa, konturirane pri različnih nivojih (r.m.s.d.). Prikazana je elektronska gostota v radiju 12,6 Å od centra rotacije.

Izbira rotamer

Eden od načinov identifikacije problematičnih delov strukture je analiza ujemanja modela strukture z elektronsko gostoto. Ta analiza je dostopna preko menija: **Validate > Density Fit Analysis** → izberite datoteko pdb. Oglejte si histogram:

- Višje vrednosti in rdeča barva pomenita slabo ujemanje oz. slabo geometrijo. V verigi A posebej izstopata dve regiji in sicer v bližini aminokislinskih ostankov 41A in 89A.
- Oglejte si tudi regijo 40B–41B (kliknite na histogram in opazujte model strukture in karto elektronske gostote). Kako razlagate slabo ujemanje z elektronsko gostoto?

V nadaljevanju bomo neujemanje med modelom strukture in karto elektronske gostote popravili:

1. Kliknite na stolpič histograma, ki predstavlja 89A. Coot nacentrira prikaz na atom C α tega aminokislinskega ostanka.
2. Oglejte si situacijo. V stranskem izbirniku kliknite na gumb **Rotamers**.
3. V grafičnem delu okna kliknite (levi klik) na katerikoli atom ostanka 89A, po kliku pa Coot prikaže okno možnih rotamerov stranske skupine. Trenutni je v modelu strukture prikazan z običajno barvo, ostali možni pa s sivo. Izberite najustreznejšega (tistega, ki se najbolje prilega elektronski gostoti) ter izbiro potrdite s klikom na OK.
4. Dodatno izpilite konformacijo tega aminokislinskega ostanka, in sicer z uporabo avtomatskega piljenja v realnem prostoru, kar dosežete s klikom na gumb **Real Space Refine Zone** v stranskem izbirniku. Nato morate aktivirati regijo, ki ji želite izpiliti:

- če gre za samo en aminokislinski ostanek, potem dvakrat zaporedoma kliknete na katerikoli atom tega aminokislinskega ostanka (ni nujno, da sta prvi in drugi klik narejena na isti atom tega aminokislinskega ostanka),
 - če gre za regijo (od ... do ... aminokislinskega ostanka), pa prvič kliknete na katerikoli atom aminokislinskega ostanka na eni strani te regije, drugi klik pa naj bo na katerikoli atom aminokislinskega ostanka na drugi strani te regije (splošen primer regije bi bil lahko od 80. do 85. aminokislinskega ostanka).
5. Coot prikaže nove, izpiljene koordinate s sivo barvo (stare koordinate so rumene). Potrdite s klikom na OK. Atomi aminokislinskega ostanka 89A sedaj lepo ležijo v karti elektronske gostote.
 6. Podobno bi lahko dosegli z uporabo zgolj možnosti **Real Space Refine Zone**. Dvakrat kliknete na *Undo* ter ponovite piljenje z uporabo **Real Space Refine Zone** ter atome stranske skupine povlecite v oblak elektronske gostote.
 7. Oglejte si še regijo okrog aminokislinskega ostanka 41A. Kaj je narobe?
 8. Uporabite možnost **Real Space Refine Zone**, pri čemer izberite celotno regijo polipeptidne verige, od 40A do 42A (najprej kliknite na en atom ostanka 40A, nato pa na en atom ostanka 42A). Povlecite atome v oblak elektronske gostote ter njihove položaje prilagajajte tako dolgo, dokler ne ležijo v oblaku elektronske gostote oz. je ujemanje z elektronsko gostoto in modelom strukture zadovoljivo.

Mehurčkologija

Mehurčki (v angleščini oz. programu Coot se zanje uporablja izraz *blob*) so oblaki elektronske gostote, ki so preveliki, da bi ustrezali molekulam vode. Z njimi se med gradnjo modela strukture pogosto srečamo; ustrezajo lahko ligandom, komponentam kristalizacijskega pufra ... Kot mehurčki so tudi majhni oblaki elektronske gostote, ki ustrezajo molekulam vode, a te zaradi majhnosti obravnavamo posebej. Mehurčke lahko poiščemo tako, da izberemo **Validate > Unmodelled blobs ...** in kliknemo na **Find Blobs** (nastavitve pustite na privzetih vrednostih). Odpre se novo okno s seznamom najdenih mehurčkov. Med delom ga pusite odprtega, saj se lahko s klikanjem na imena mehurčkov pomikate med njimi. Poglejmo si jih:

1. Mehurček 2 (**Blob 2**):

- (a) Kakšne oblike je ta mehurček? Kaj menite, da bi vanj glede na obliko najbolje prilegalo? Razmislite, kaj bi lahko vanj vgradili (glede na sestavo kristalizacijskega pufra).
- (b) V stranskem izbirniku kliknite na gumb **Place Atom At Pointer ...** V oknu, ki se odpre, izberite SO₄, ter izbiro potrdite s klikom na OK.
- (c) Z uporabo možnosti **Real Space Refine Zone** izpilite koordinate sulfatnega iona.

2. Mehurček 3 (**Blob 3**):

- (a) Kakšne oblike je ta mehurček? Kaj menite, da bi se vanj najbolje prilegalo (glede na sestavo kristalizacijskega pufra)? Namig: slika 4.1.
- (b) Koordinate tega liganda so že v zbirki znanih ligandov¹⁰ pod kodo 3GP. Dobimo jih tako: **File > Get Monomer**. V polje vpišite 3GP.
- (c) Z uporabo **Rotate/Translate Zone** (gumb v stranskem izbirniku) premaknite 3'-GMP tako, da se bosta obliki liganda in bloba 3 ujemali: po izbiri tega orodja kliknite dvakrat kliknite na isti atom (ali po enkrat na dva različna atoma) v molekuli 3'-GMP, nato pa z vlečenjem premaknite 3'-GMP. Celoten sistem ali zgolj 3'-GMP lahko med tem tudi vrtite. Končen položaj potrdite s klikom na OK.

¹⁰ Zbirka znanih ligandov se imenuje **Ligand Expo** in je del zbirke PDB. Dostopna je na naslovu <https://ligand-expo.rcsb.org>. Gre za ligande, ki so že bili identificirani v drugih strukturah makromolekul in so dobili svojo identifikacijsko oznako, sestavljeno iz kombinacije črk in števil (skupno trije znaki).

- (d) Prileganje elektronski gostoti optimizirajte z uporabo [Real Space Refine Zone](#). Analizirajte kontakte, ki jih ligand tvori z bližnjimi aminokislinskimi ostanki. To je tudi eden izmed načinov, da ugotovite njegovo pravilno orientacijo!
- (e) Model strukture ribonukleaze Sa in 3'-GMP združite: [Calculate > Merge Molecules](#).
- (f) V isto datoteko PDB zlijte še sulfatni ion iz prejšnjega primera.

3. Mehurček 1 (Blob 1):

- (a) Oglejte si situacijo. Kaj menite, da manjka? Primerjajte zaporedje proteina v strukturi z zaporedjem ribonukleaze Sa v zbirki UniProt. Primerjajte pa tudi zaporedji verig A in B v vašem modelu strukture.
- (b) Razmislite, kaj bi lahko bil razlog, da manjkajočega dela strukture v našem primeru ne moremo identificirati prek karte diferenčne elektronske gostote.
- (c) Manjkajoči del polipeptidne verige dodate tako: [Calculate > Fit Loop](#) → območje zaporedja ustrezno nastavite (novo dodani aminokislinski ostanki bodo imeli te številčne oznake!) ter v spodnje polje vpišite zaporedje aminokislinskih ostankov, ki manjkajo v strukturi (enočrkovne kode). Kliknite [Fit Loop](#) in opazujte.
- (d) Oglejte si C-končni aminokislinski ostanek in njegovo bližino. Bi kaj spremenili? Uporabite [Real Space Refine Zone](#).
- (e) Na C-konec dodajte tudi kisikov atom (OXT, t.j. *terminal oxygen*¹¹): [Calculate > Other Modelling Tools > Add OXT to Residue](#).

¹¹ Tak atom ima samo C-končni aminokislinski ostanek polipeptidne verige in kot tak označuje njen konec.

Voda

Oglejte si majhne okrogle oblačke elektronske gostote v neposredni bližini proteina. To je t. i. kristalna voda. Te molekule so v stiku s proteinom in so zato urejene, v karti elektronske gostote pa jih vidimo kot majhne okrogle oblačke. S podrobnejšo analizo njihove bližnje okolice lahko ugotovimo, kateri atomi katerih aminokislinskih ostankov sodelujejo pri njihovi vezavi. V povprečju lahko v strukturi proteina pri ločljivosti ≈ 2 Å zaznamo toliko molekul vode kot je aminokislinskih ostankov v strukturi¹².

Molekule vode na avtomatiziran način lahko poiščemo tako: [Calculate > Other Modelling Tools > Find Waters](#).

¹² Za zanimivo analizo koliko molekul vode lahko identificiramo v kristalnih strukturah je zanimivo branje članek *How many water molecules can be detected by protein crystallography?* (59).

Alternativna konformacija aminokislinskih ostankov

Oglejte si aminokislinski ostanek 72A. Za kateri aminokislinski ostanek gre? Je s strukturo tega dela ribonukleaze Sa vse v redu?

1. Izbrišite ostanek 72A: [Calculate > Model/Fit/Refine > Delete > Residue/Monomer](#) ali kliknite [Delete Item ...](#) v stranskem izbirniku.
2. Na to mesto dodajte alaninski ostanek: [Calculate > Model/Fit/Refine > Add Terminal Residue](#), lahko pa tudi uporabite gumb [Add Residue](#) v stranskem izbirniku (kliknite na 71A, na katerega želite dodati končen aminokislinski ostanek).
3. Na novo dodani aminokislinski ostanek¹³ Ala72 mutirajte v Cys: kliknite na gumb [Mutate & Auto Fit](#) v stranskem izbirniku, nato kliknite na ostanek 72A (Ala), v oknu pa izberite Cys.
4. Konformacijo dela polipeptidne verige 71A–73A izpilite z uporabo [Real Space Refine Zone](#).
5. Ponovno si oglejte ostanek 72A. Kaj opazite?
6. Aminokislinskemu ostanku Cys72A dodajte alternativno konformacijo: v stranskem izbirniku kliknite na [Add Alternate Conformation](#) ter nato kliknite na ta aminokislinski ostanek. Razmislite, kakšen je najprimernejši delež oz. zasedenost (*occupancy*).
7. Alternativno konformacijo 72A izpilite z uporabo [Real Space Refine Zone](#).

¹³ Kot nov aminokislinski ostanek se ponavlja v fačetni fazi dodajajo alaninski ostanki. Tak ostanek je namreč najmanjši ostanek, ki ima neko stransko skupino z možnostjo usmeritve, hkrati pa nudi dovolj konformacijske svobode, da ga lahko umestimo skoraj kamorkoli.

Ramachandranov diagram

Oglejte si še Ramachandranov diagram: [Validate](#) ▷ [Ramachandran Plot](#) ▷ [RNAseSa.pdb](#):

- So vsi aminokislinski ostanki v dovoljenih regijah? (Samo aminokislinske ostanke, ki so v nedovoljenih regijah, lahko prikažete s klikom na gumb [Outliers Only](#)).
- Se lega posameznih aminokislinskih ostankov v diagramu ujema z njihovim položajem znotraj elementov sekundarne strukture?

Vaja 4b

Reševanje strukture HCAII in validacija

Namen vaje

Namen vaje je na primeru reševanja strukture človeške karbonske anhidraze II spoznati naslednje stopnje: faziranje z uporabo molekulske zamenjave, avtomatska gradnja modela strukture, ročna gradnja modela strukture in piljenje (avtomatsko in ročno) ter validacija modela.

Za izvedbo vaje so potrebne veščine, pridobljene v predhodni vaji 4a (stran 43).

Material

Vaja je računalniška. Za izvedbo so potrebne datoteke z vhodnimi podatki, ki jih dobite pri asistentu:

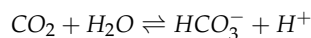
- [hca2.mtz](#), ki vsebuje sprocesirane difrakcijske podatke¹⁴ kristala človeške karbonske anhidraze II, a brez faz, tako da izračun kart elektronske gostote ni možen,
- [hca3.pdb](#) s strukturo homologa - karbonske človeške anhidraze III.

¹⁴ Podatki je posnel dr. Bjoern Sjöblom iz skupine prof. dr. Kristine Djinović Carugo. Podatki so bili posneti pod krio-pogoji.

Nekaj osnovnih informacij o karbonski anhidrazi II je v nadaljevanju, sledi pa opis programske opreme.

Uvodne informacije o karbonski anhidrazi II

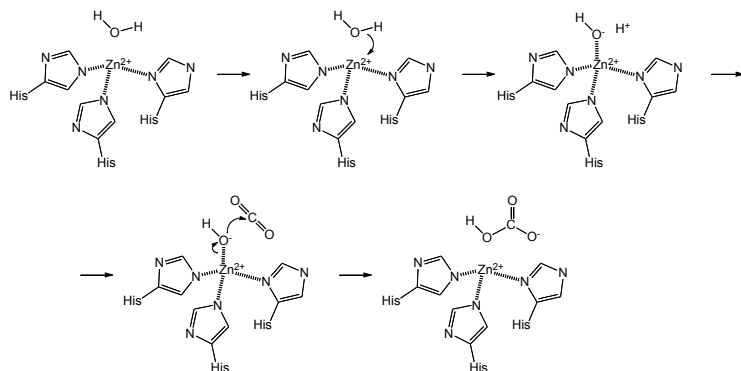
Karbonske anhidraze katalizirajo reverzibilno hidratacijo ogljikovega dioksida:



Nahajajo se v vseh živalih in fotosintetizirajočih organizmih ter nekaterih nefotosintetizirajočih bakterijah. Obstajajo tri evolucijsko nesorodne družine karbonskih anhidraz (α – živali, β – rastline in γ – bakterije), med katerimi ni sekvenčne podobnosti, so pa vse cinkovi metaloencimi. Družine karbonskih anhidraz tako predstavljajo jasen primer konvergentne evolucije katalitske funkcije. Pri človeku poznamo 7 gensko različnih α -izoencimov (CA I–VII). Karbonska anhidraza II (CA II) je najbolj preučena oblika, nahaja se v številnih tipih celic ter je znana kot eden najbolj učinkovitih encimov z maksimalno pretvorbena hitrostjo 10^6 s^{-1} (pri pH 9,0 ter temperaturi 25 °C). Izooblika CAIII ima precej nižjo pretvorbena število ($8 \times 10^3 \text{ s}^{-1}$), nahaja pa se pretežno v mišičnih vlaknih, kjer sodeluje pri facilitirani difuziji CO_2 do tkivnih kapilar. Katalitski mehanizem človeških karbonskih anhidraz je prikazan na sliki 4.5.

Programska oprema

Del vaje, ki se nanaša na gradnjo in piljenje modela v grafičnem okolju, poteka z uporabo programa Coot, ki je opisan v vaji 4a na strani 43.



Slika 4.5: Katalitski mehanizem človeških karbonskih anhidraz.

Za ostale posamezne korake vaje je na voljo več ločenih programov, ki se danes najpogosteje uporabljajo kot del večjih programskih paketov. Najpopularnejša sta dva, in sicer **CCP4** (60), dostopen na naslovu <https://www.ccp4.ac.uk>, in **Phenix** (61), dostopen na naslovu <https://www.phenix-online.org>. Oba sta prosto dostopna in tečeta na vseh pogostih operacijskih sistemih, pri obeh pa gre pravzaprav za skupek velikega števila programov, ki so namenjeni posameznim opravilom, s skupnim in bolj ali manj enotnim grafičnim vmesnikom. Tako pri obeh paketih najdemo programe za analizo difrakcijskih podatkov, faziranje z eksperimentalnimi metodami ter molekulsko zamenjavo, avtomatsko gradnjo modela ter piljenje, validacijo ...

V novejšem času so orodja paketa CCP4 na voljo brezplačno tudi prek spletnega vmesnika **CCP4Cloud**: <https://cloud.ccp4.ac.uk>. To bistveno poenostavi uporabo, sploh za začetniške uporabnike, saj ni potrebe po instalaciji velikega programskega paketa, prav tako pa vsi procesi (nekateri so računsko zelo zahtevni) tečejo na skupku oddaljenih računalnikov. Omejeni smo le s prostorom in številom opravil oz. procesorskim časom, ki nam je na voljo. Ravno ta, spletni vmesnik, je opisan v nadaljevanju, je pa kot *sistem v razvoju* seveda podvržen spremembam, zato se lahko vmesniki oz. način prikaza rezultatov analiz nekoliko razlikuje od tega, prikazanega na slikah.

Postopek dela

Molekulska zamenjava

Najprej analizirajte podobnost med karbonsko anhidrazo II in III prek poravnave aminokislinskih zaporedij. Razmislite, ali je uporaba strukture karbonske anhidraze III dejansko smiselna za molekulsko zamenjavo pri izoobliki II.

V nadaljevanju je orisan postopek molekulske zamenjave s ključnimi koraki, podrobne informacije in navodila

1. Prijavimo se v sistem CCP4Cloud in ustvarimo nov projekt.
2. Dodamo novo opravilo (*job*), prek klika na velik zelen znak plus (+), in sicer uvozimo podatke (**Data Import**). Naložimo datoteko [hca2.mtz](#). Po uvozu se nam prikaže analiza podatkov, ki jih vsebuje, med drugim kratek povzetek (slika 4.6). Vidimo, da je bila valovna dolžina uporabljenih X-žarkov 1,54 Å, stranice osnovne celice imajo dimenzije $a = 42,07$ Å, $b = 41,387$ Å in $c = 72,067$ Å, koti so $\alpha = \gamma = 90^\circ$, $\beta = 104,141^\circ$. Najvišja ločljivost podatkov je 1,59 Å, prostorska skupina je P 1 21 1.
3. Uvozite (**Import Sequence(s) by Copy-Paste**) še aminokislinsko zaporedje karbonske anhidraze II, ki ga najdete v zbirki UniProt. Zaporedje naj bo v formatu FASTA.

File name	0001-01.mtz
Dataset name	proteum/hca2/hca2
Assigned name	[0001-01] hca2 [proteum/hca2/hca2] /hkl/anom/
Wavelength	1.54
Space group	P 1 21 1
Cell	42.07 41.387 72.067 90.0 104.141 90.0
Resolution low	69.88
Resolution high	1.59
Anomalous scattering	Present
Original columns	F_hca2 SIGF_hca2 F_hca2(+) SIGF_hca2(+) F_hca2(-) SIGF_hca2(-) FreeR_flag
Truncation	Truncated dataset will be used instead of the original one.
Columns to be used	F SIGF F(+) SIGF(+) F(-) SIGF(-) FreeR_flag

Slika 4.6: Prikaz povzetka po uvozu difrakcijskih podatkov.

- Prek opravila [Asymmetric Unit Contents](#) določite najbolj verjetno število kopij polipeptidne verige na asimetrično enoto. Izračun poteka prek izračuna Matthewsovega koeficienta, kar je podrobneje opisano na strani 40.
- Sedaj bi že lahko začeli z molekulsko zamenjavo, lahko pa model (strukturo karbonske anhidraze III) nekoliko obdelamo, in sicer odstranimo atome, ki se med strukturo zagotovo razlikujejo. Eden izmed pristopov, ki to naredi, je Chainsaw, do katerega pridemo prek [Prepare MR Model\(s\) from Coordinate data](#) in uporabe datoteke [hca3.pdb](#). Uporaba takega pristopa je posebej smiselna, če se aminokislinsko zaporedje strukture za molekulsko zamenjavo bolj razlikuje od zaporedja proteina v kristalu.
- Molekulsko zamenjavo s programom Phaser začnemo prek izbire [Molecular Replacement with Phaser](#). Kot vhodne podatke izberemo difrakcijske podatke iz datoteke [hca2.mtz](#) ter model Chainsaw, pripravljen v prejšnjem koraku. Pozorni pa moramo biti tudi na to, koliko kopij modela v asimetrični enoti bomo iskali (parameter *Look for X copies in ASU*).
- Po zaključku nam Phaser sporoči nekaj statističnih parametrov, iz katerih lahko izluščimo, ali je bila molekulska zamenjava uspešna. To sta vrednosti LLG (*Log Likelihood Gain*, ki je povezana z ujemanjem modela s podatki (višja kot je, bolje je; vrednosti nad 120 se smatrajo kot dobre) ter TFZ (*Translation Function Z-score*), ki predstavlja število standardnih deviacij od srednje vrednosti. Velja naslednje pravilo: $TFZ < 5 \rightarrow$ molekulska zamenjava ni uspešna; $TFZ > 8 \rightarrow$ molekulska zamenjava je zagotovo uspešna; vmesne vrednosti predstavljajo različne nivoje uspešnosti.
- Po molekulski zamenjavi je smiselno narediti nekaj ciklov avtomatskega piljenja: [Refinement with Refmac](#). Kot vhodne podatke izberemo izhodne podatke prejšnje stopnje.
- Po končanem avtomatskem piljenju si oglejmo, kako so se vrednosti faktorjev R in R_{free} spreminjale s cikli – padanje vrednosti je dober znak, da je bila molekulska zamenjava dejansko uspešna.

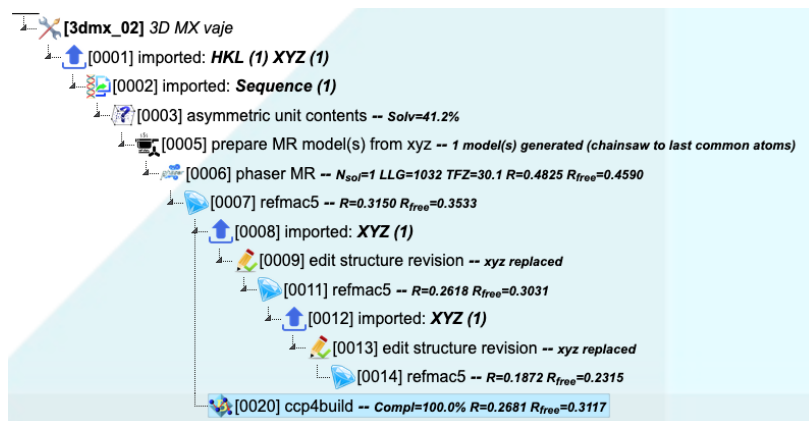
Ročna gradnja in piljenje modela

S CCP4Cloud prenesite datoteke avtomatskega piljenja po molekulski zamenjavi (datoteki mtz in pdb), jih odprite v programu Coot ter po vzoru iz vaje 4a (stran 43) nadaljujte z ročnim piljenjem modela strukture karbonske anhidraze II. Upoštevajte in razmislite:

- Kot rezultat molekulske zamenjave dobimo datoteko pdb. Kaj pravzaprav vsebuje? Je to že model strukture, v našem primeru karbonske anhidraze II? Če ne, kaj moramo s tem modelom najprej narediti?
- Po večjih spremembah oz. popravkih modela strukture ga ponovno prenesite (uvazite koordinate) v CCP4Cloud ter naredite nekaj ciklov avtomatskega piljenja z orodjem Refmac. Tako boste prišli do natančnejših faz oz. strukturnih faktorjev ter posledično do boljše (bolj detaljne) karte elektronske gostote.

- Kje je aktivno mesto karbonske anhidraze II? Kateri aminokislinski ostanki sodelujejo pri koordiniranju osrednjega kovinskega iona?

Ročno in avtomatsko piljenje izmenjujete, dokler vrednosti faktorjev R in R_{free} ne padajo več. Razmislite, ali je model strukture popoln oz. dobro izpiljen ter v skladu tako z eksperimentalnimi podatki kot stereokemijskimi parametri. Primer oz. zaslonska slika iz CCP4Cloud, na kateri so razvidne posamezne stopnje, je prikazan na sliki 4.7. Model strukture še ni popolnoma izpiljen!



Slika 4.7: Koraki do skoraj izpiljenega modela strukture karbonske anhidraze II v CCP4Cloud. Zadnji prikazan korak (ccp4build) vsebuje avtomatsko gradnjo modela strukture po molekularni zamenjavi in prvem avtomatskem piljenju – avtomatsko generiran model vsekakor ni tako dober kot ročno izpiljen.

Avtomatska gradnja modela

Za primerjavo z vašo, ročno grajeno in izpiljeno, strukturo zaženite po prvem avtomatskem piljenju po molekularni zamenjavi avtomatsko gradnjo modela strukture [Automatic Model Building with CCP4Build](#). Kako dober je algoritem v primerjavi z vašo, človeško logiko?

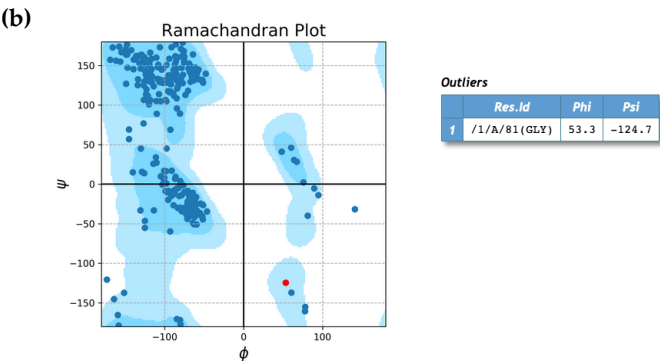
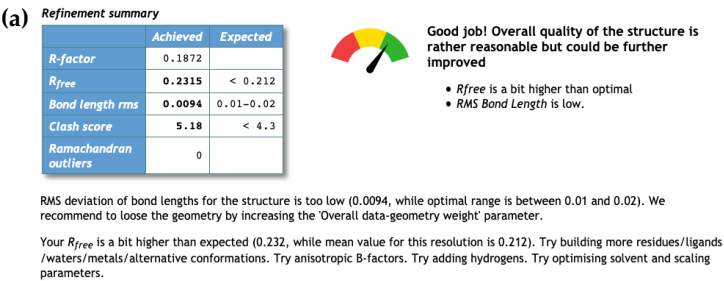
Validacija modela strukture

Validacija modela strukture je, kot je opisano v razdelku 4.6, integralen del piljenja. Številna orodja za ročno validacijo modela strukture vsebuje Coot, vrednosti parametrov, s katerimi ovrednotimo model strukture, nam izračuna tudi Refmac5 med avtomatskim piljenjem (primer nekaj parametrov za nepopolno izpiljen model strukture je na sliki 4.8).

Statistični parametri, izračunani med validacijo modela strukture, so skupaj s podatki o procesiranju difrakcijskih podatkov del t. i. **Tabele 1**. Slednja je dobila ime po tem, da je bila po navadi prva tabela v člankih, ki so opisovali na novo določene strukture makromolekul.

Dodatek za entuziaste

Za tiste, ki jih piljenje in analiza modelov strukture bioloških makromolekul posebej zanimata, bodo vsekakor zanimive karte elektronske gostote, ki jih moramo ob deponiranju samega modela strukture naložiti v zbirko PDB. V tej zbirki sta tako ob kristalnih strukturah na voljo klasična kompozitna karta elektronske gostote ter karta diferenčne elektronske gostote, na voljo so tudi koeficienti karte elektronske gostote oz. strukturni faktorji.



Slika 4.8: Nekaj glavnih parametrov, s katerimi ovrednotimo model strukture, je prikazanih na izhodu programa Refmac5: (a) faktorja *R* in *R*_{free}, r.m.s. dolžine vezi, sterične ovire (*clash score*) ter število aminokislinskih ostankov v nedovoljenih regijah Ramachandranovega diagrama. (b) Ramachandranov diagram. Glicin kot najmanjši in posledično najmanj sterično omejen aminokislinski ostank ni problematičen.

Okrajšave in simboli

Okrajšave

asu	asimetrična enota (angl. <i>asymmetric unit</i>)
krio-EM	krio-elektronska mikroskopija (angl. <i>cryo-electron microscopy</i> , <i>cryo-EM</i>)
LLG	logaritem pridobljenega verjetja (angl. <i>log likelihood gain</i>)
MR	molekulska zamenjava (angl. <i>molecular replacement</i>)
MX	makromolekulska kristalografija (angl. <i>macromolecular crystallography</i>)
NMR	jedrska magnetna resonanca (angl. <i>nuclear magnetic resonance</i>)
PDB	zbirka makromolekulskih struktur, tudi format datoteke (angl. <i>Protein Data Bank</i>)
PX	proteinska kristalografija (angl. <i>protein crystallography</i>)
RF	rotacijska funkcija (angl. <i>rotational function</i>)
TF	translacijska funkcija (angl. <i>translation function</i>)
TFZ	Z-vrednost translacijske funkcije (angl. <i>translation function Z-score</i>)

Simboli

α, β, γ	koti osnovne celice
λ	valovna dolžina
ν	frekvenca
a, b, c	dolžine stranic osnovne celice
F	strukturni faktor (angl. <i>structure factor</i>)
F_c, F_{calc}	strukturni faktor, izračunan na osnovi modela
F_o, F_{obs}	strukturni faktor, izračunan na osnovi eksperimentalnih (opaženih) podatkov
h	Planckova konstanta ($6,62607004 \times 10^{-34} \text{ m}^2\text{kg/s}$)
I	intenziteta
R, R_{free}	statistični faktor, merilo ujemanja med modelom strukture in eksperimentalnimi podatki
V_M	Matthewsov volumen

Literatura

- [1] H. M. Berman, "The Protein Data Bank", *Nucleic Acids Research*, 28(1), 2000, str. 235–242.
- [2] Ka Man Yip, in sod., "Atomic-Resolution Protein Structure Determination by Cryo-EM", *Nature*, 587(7832), 2020, str. 157–161.
- [3] C. Jelsch, in sod., "Accurate Protein Crystallography at Ultra-High Resolution: Valence Electron Distribution in Crambin", *Proceedings of the National Academy of Sciences*, 97(7), 2000, str. 3171–3176.
- [4] Eric F Pettersen, in sod., "UCSF Chimera—a Visualization System for Exploratory Research and Analysis", *J Comput Chem*, 25(13), 2004, str. 1605–12.
- [5] P. Emsley, in sod., "Features and Development of Coot", *Acta Crystallographica Section D Biological Crystallography*, 66(4), 2010, str. 486–501.
- [6] Edward Tyson Reichert in Amos Peaslee Brown, *The Differentiation and Specificity of Corresponding Proteins and Other Vital Substances in Relation to Biological Classification and Organic Evolution : The Crystallography of Hemoglobins*, Carnegie Institute of Washington, Washington, DC, 1909.
- [7] William H. Brondyk, "Chapter 11 Selecting an Appropriate Method for Expressing a Recombinant Protein", in *Methods in Enzymology*, Elsevier, vol. 463, 2009, str. 131–147.
- [8] Raymond C Stevens, "Design of High-Throughput Methods of Protein Production for Structural Biology", *Structure*, 8(9), 2000, str. R177–R185.
- [9] Youngchang Kim, in sod., "High-Throughput Protein Purification and Quality Assessment for Crystallization", *Methods*, 55(1), 2011, str. 12–28.
- [10] Youngchang Kim, in sod., "Automation of Protein Purification for Structural Genomics", *Journal of Structural and Functional Genomics*, 5(1/2), 2004, str. 111–118.
- [11] Rong Xiao, in sod., "The High-Throughput Protein Sample Production Platform of the Northeast Structural Genomics Consortium", *Journal of Structural Biology*, 172(1), 2010, str. 21–33.
- [12] Georg Mlynek, in sod., "The Center for Optimized Structural Studies (COSS) Platform for Automation in Cloning, Expression, and Purification of Single Proteins and Protein–Protein Complexes", *Amino Acids*, 46(6), 2014, str. 1565–1582.
- [13] Joanne E. Nettleship, in sod., "Overview of a High-Throughput Pipeline for Streamlining the Production of Recombinant Proteins", in *High-Throughput Protein Production and Purification*, rednik Renaud Vincentelli, Springer New York, New York, NY, vol. 2025, 2019, str. 33–49.
- [14] José Arnau, in sod., "Current Strategies for the Use of Affinity Tags and Tag Removal for the Purification of Recombinant Proteins", *Protein Expression and Purification*, 48(1), 2006, str. 1–13.
- [15] Michelle E. Kimple, in sod., "Overview of Affinity Tags for Protein Purification", *Current Protocols in Protein Science*, 73(1), 2013.
- [16] David W Wood, "New Trends and Affinity Tag Designs for Recombinant Protein Purification", *Current Opinion in Structural Biology*, 26, 2014, str. 54–61.
- [17] David S. Waugh, "An Overview of Enzymatic Reagents for the Removal of Affinity Tags", *Protein Expression and Purification*, 80(2), 2011, str. 283–293.
- [18] Miha Pavšič, in sod., "Crystal Structure and Its Bearing towards an Understanding of Key Biological Functions of EpCAM", *Nature Communications*, 5, 2014, str. 4764.
- [19] Hui Sun Lee, in sod., "Effects of N-Glycosylation on Protein Conformation and Dynamics: Protein Data Bank Analysis and Molecular Dynamics Simulation Study", *Scientific Reports*, 5(1), 2015, str. 8926.
- [20] Vishnu Priyanka Reddy Chichili, in sod., "Linkers in the Structural Biology of Protein-Protein Interactions.", *Protein Science*, 22(2), 2013, str. 153–167.
- [21] Xiaoying Chen, in sod., "Fusion Protein Linkers: Property, Design and Functionality", *Adv Drug Deliv Rev*, 65(10), 2013, str. 1357–69.

- [22] Zygmunt S. Derewenda in Adam Godzik, "The "Sticky Patch" Model of Crystallization and Modification of Proteins for Enhanced Crystallizability.", *Methods in molecular biology (Clifton, NJ)*, 1607, 2017, str. 77–115.
- [23] Michael S. Packer in David R. Liu, "Methods for the Directed Evolution of Proteins", *Nature Reviews Genetics*, 16(7), 2015, str. 379–394.
- [24] Miha Pavšič, "(Po)Moč Evolucije. Nobelova Nagrada Za Kemijo", *Proteus*, 81(2), 2018, str. 102–108.
- [25] Darren J. Hart in Franck Tarendeau, "Combinatorial Library Approaches for Improving Soluble Protein Expression in *Escherichia Coli*", *Acta Crystallographica Section D Biological Crystallography*, 62(1), 2006, str. 19–26.
- [26] Shohei Koide, "Engineering of Recombinant Crystallization Chaperones", *Current Opinion in Structural Biology*, 19(4), 2009, str. 449–457.
- [27] Magdalena A Bukowska in Markus G Grütter, "New Concepts and Aids to Facilitate Crystallization", *Current Opinion in Structural Biology*, 23(3), 2013, str. 409–416.
- [28] Peer RE Mittl, in sod., "Chaperone-Assisted Structure Elucidation with DARPins", *Current Opinion in Structural Biology*, 60, 2020, str. 93–100.
- [29] Zygmunt S. Derewenda, "Rational Protein Crystallization by Mutational Surface Engineering", *Structure*, 12(4), 2004, str. 529–535.
- [30] Lukasz Goldschmidt, in sod., "Toward Rational Protein Crystallization: A Web Server for the Design of Crystallizable Protein Variants.", *Protein Science*, 16(8), 2007, str. 1569–1576.
- [31] Thomas S Walter, in sod., "Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization.", *Structure (London, England : 1993)*, 14(11), 2006, str. 1617–1622.
- [32] Min-Sung Kim in Dan Leahy, "Enzymatic Deglycosylation of Glycoproteins", in *Methods in Enzymology*, Elsevier, vol. 533, 2013, str. 259–263.
- [33] Lukasz Kurgan, in sod., "CRYSTALP2: Sequence-Based Protein Crystallization Propensity Prediction", *BMC Structural Biology*, 9(1), 2009, str. 50.
- [34] György Babnigg in Andrzej Joachimiak, "Predicting Protein Crystallization Propensity from Protein Sequence.", *Journal of structural and functional genomics*, 11(1), 2010, str. 71–80.
- [35] Marcin J. Mizianty in Lukasz Kurgan, "Sequence-Based Prediction of Protein Crystallization, Purification and Production Propensity", *Bioinformatics*, 27(13), 2011, str. i24–i33.
- [36] Samad Jahandideh, in sod., "Improving the Chances of Successful Protein Structure Determination with a Random Forest Classifier", *Acta Crystallographica Section D Biological Crystallography*, 70(3), 2014, str. 627–635.
- [37] Lukasz Slabinski, in sod., "XtalPred: A Web Server for Prediction of Protein Crystallizability.", *Bioinformatics*, 23(24), 2007, str. 3403–3405.
- [38] Jun Hu, in sod., "TargetCrys: Protein Crystallization Prediction by Fusing Multi-View Features with Two-Layered SVM", *Amino Acids*, 48(11), 2016, str. 2533–2547.
- [39] Huilin Wang, in sod., "PredPPCrys: Accurate Prediction of Sequence Cloning, Protein Production, Purification and Crystallization Propensity from Protein Sequences Using Multi-Step Heterogeneous Feature Fusion and Selection", *PLoS ONE*, 9(8), 2014, str. e105902.
- [40] Huilin Wang, in sod., "Crysalis: An Integrated Server for Computational Analysis and Design of Protein Crystallization", *Sci Rep*, 6, 2016, str. 21383.
- [41] Huilin Wang, in sod., "Critical Evaluation of Bioinformatics Tools for the Prediction of Protein Crystallization Propensity.", *Briefings in Bioinformatics*, 19(5), 2018, str. 838–852.
- [42] Fanchi Meng, in sod., "fDETECT Webserver: Fast Predictor of Propensity for Protein Production, Purification, and Crystallization", *BMC Bioinformatics*, 18(1), 2017, str. 580.
- [43] Abdurrahman Elbasir, in sod., "DeepCrystal: A Deep Learning Framework for Sequence-Based Protein Crystallization Prediction.", *Bioinformatics (Oxford, England)*, 35(13), 2019, str. 2216–2225.
- [44] Abdurrahman Elbasir, in sod., "BCrystal: An Interpretable Sequence-Based Protein Crystallization Predictor", *Bioinformatics*, 36(5), 2020, str. 1429–1438.
- [45] L. Chen, in sod., "TargetDB: A Target Registration Database for Structural Genomics Projects", *Bioinformatics*, 20(16), 2004, str. 2860–2862.
- [46] Jack Kyte in Russell F. Doolittle, "A Simple Method for Displaying the Hydropathic Character of a Protein", *Journal of Molecular Biology*, 157(1), 1982, str. 105–132.
- [47] J.J. Ward, in sod., "Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life", *Journal of Molecular Biology*, 337(3), 2004, str. 635–645.
- [48] David T Jones, "Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices 1 Edited by G. Von Heijne", *Journal of Molecular Biology*, 292(2), 1999, str. 195–202.

- [49] A. Lupas, in sod., "Predicting Coiled Coils from Protein Sequences", *Science*, 252(5009), 1991, str. 1162–1164.
- [50] Anders Krogh, in sod., "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete genomes¹¹Edited by F. Cohen", *Journal of Molecular Biology*, 305(3), 2001, str. 567–580.
- [51] Dariusz Plewczynski, in sod., "The RPSP: Web Server for Prediction of Signal Peptides", *Polymer*, 48(19), 2007, str. 5493–5496.
- [52] John C. Wootton, "Non-Globular Domains in Protein Sequences: Automated Segmentation Using Complexity Measures", *Computers & Chemistry*, 18(3), 1994, str. 269–285.
- [53] J. Jancarik in S. H. Kim, "Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins", *Journal of Applied Crystallography*, 24(4), 1991, str. 409–411.
- [54] Richard Giegé, "A Historical Perspective on Protein Crystallization from 1840 to the Present Day.", *The FEBS journal*, 280(24), 2013, str. 6456–6497.
- [55] Bernhard Rupp, *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, New York, 2010.
- [56] Dale E Tronrud, "Introduction to Macromolecular Refinement.", *Acta crystallographica Section D, Biological crystallography*, 60(Pt 12 Pt 1), 2004, str. 2156–2168.
- [57] Gerard J. Kleywegt, "Validation of Protein Crystal Structures", *Acta Crystallographica Section D Biological Crystallography*, 56(3), 2000, str. 249–265.
- [58] Paul Emsley in Kevin Cowtan, "Coot: Model-Building Tools for Molecular Graphics.", *Acta crystallographica Section D, Biological crystallography*, 60(Pt 12 Pt 1), 2004, str. 2126–2132.
- [59] O Carugo in D Bordo, "How Many Water Molecules Can Be Detected by Protein Crystallography?", *Acta crystallographica Section D, Biological crystallography*, 55(Pt 2), 1999, str. 479–483.
- [60] Martyn D. Winn, in sod., "Overview of the CCP4 Suite and Current Developments", *Acta Crystallographica Section D Biological Crystallography*, 67(4), 2011, str. 235–242.
- [61] Dorothee Liebschner, in sod., "Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix", *Acta Crystallographica Section D Structural Biology*, 75(10), 2019, str. 861–877.