

Pristupi za funkcionalizaciju proteina i proširenje enzimске katalize

Močinić, Antonio

Undergraduate thesis / Završni rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:071356>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-13**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)





Sveučilište u Zagrebu
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
Kemijski odsjek

Antonio Močinić

Student 3. godine Preddiplomskog sveučilišnog studija KEMIJA

PRISTUPI ZA FUNKCIONALIZACIJU PROTEINA I PROŠIRENJE ENZIMSKE KATALIZE

Završni rad

Rad je izrađen u Zavodu za biokemiju

Mentor rada: izv. prof. dr. sc. Aleksandra Maršavelski

Zagreb, 2024.

Datum predaje prve verzije Završnog rada:

19. lipnja 2024.

Datum ocjenjivanja Završnog rada i polaganja Završnog ispita:

12. srpnja 2024.

Mentor rada: izv. prof. dr. sc. Aleksandra Maršavelski

Potpis:

Ovaj rad je napravljen u sklopu projekta „Enzimsko inženjerstvo za održivo recikliranje bioplastike“ na Zavodu za biokemiju Prirodoslovno-matematičkog fakulteta u Zagrebu, pod mentorstvom izv. prof. dr. sc. Aleksandre Maršavelski.

Sadržaj

§ SAŽETAK.....	VII
§ 1. UVOD.....	1
1.1. Proteini	1
1.2. Proteinsko inženjerstvo.....	1
§ 2. ENERGETSKA OPTIMIZACIJA STRUKTURA.....	3
2.1. PELE – Protein Energy Landscape Exploration.....	3
2.1.1. Lokalne perturbacije	3
2.1.2. Predviđanje bočnih ogranaka	4
2.1.3. Minimizacioni algoritmi.....	4
2.2. Nanoreaktori za razgradnju čestica mikroplastike.....	5
2.2.1. Mikroplastika.....	5
2.2.2. Stvaranje nanopora za katalitičku razgradnju mikroplastike	5
2.2.3. Katalitičke mogućnosti nanopora	8
2.3. Unaprjeđenje i proširenje katalitičkih mogućnosti.....	11
2.3.1. Uvođenje dodatnog aktivnog mjesta u svrhu poboljšanja katalitičkih mogućnosti	11
2.3.2. Kataliza kaskadne reakcije.....	13
§ 3. MODELI PROTEINSKOG JEZIKA	16
3.1. ProteinMPNN	16
3.1.1. Opis modela.....	16
3.1.2. Proteini koji sadrže kvarternu strukturu	17
3.1.3. Usporedba modela in silico.....	18
3.1.4. Eksperimentalna potvrda	19
§ 4. LITERATURNI IZVORI.....	XXII

§ Sažetak

Budući da su proteini u svom nativnom obliku odgovorni za čitav niz složenih funkcija koje je jako teško ostvariti u ne-biološkom okruženju, funkcionalizacija proteina pokazala se kao izvrstan pristup u rješavanju mnogih problema suvremenog društva. Modifikacijom postojećih proteina ili modeliranjem novih mogu se riješiti problemi koje bi inače bilo iznimno teško riješiti klasičnim kemijskim putem. Uz pomoć softvera za energetske optimizacije, iz strukturnog proteina FraC, dobivena je nanopora za razgradnju čestica mikroplastike, koja ima veliki potencijal u rješavanju globalnog problema zagađenja plastikom. Sličnim je pristupom poboljšana i katalitička aktivnost jedne serinske esterske hidrolaze, te joj je dodano i dodatno aktivno mjesto, tako da može katalizirati i kaskadne reakcije. Nedavnom pojavom modela proteinskog jezika (engl. *protein language models*), metode koja se temelji na dubokom učenju, znanstvenicima je omogućeno da analiziraju i predviđaju svojstva proteina na temelju njihovih aminokiselinskih sekvenci. *ProteinMPNN* (engl. *Protein Mechanics by Predicted Neural Networks*) je primjer računalnog alata osmišljenog za optimizaciju proteinskih sekvenci kako bi se dobile željene strukture ili funkcije.

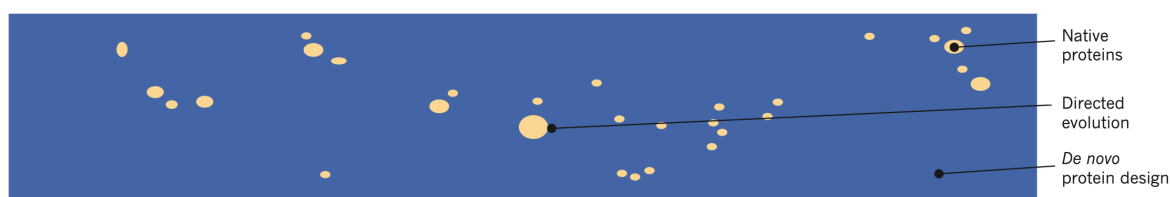
§ 1. UVOD

1.1. Proteini

Proteini su biopolimeri izgrađeni iz 20 proteingoenih aminokiselina, zbog svojih trenutnih i potencijalnih funkcija u fokusu su biomedicinskih i drugih istraživanja. Ne samo da imaju gradivnu ulogu kod većine živih bića, nego su i njihove funkcije iznimno složene. Naprimjer, proteini korištenjem solarne energije mogu stvoriti kompleksne molekula, drugi mogu detektirati male molekula u iznimno malome prisustvu (olfatkorni receptori), detektirati svjetlo (rodopsin), a neki mogu i transformirati kemijsku energiju u kemijske veze (ATP sintetaza) ili mehanički rad (aktin i miozin).¹

1.2. Proteinsko inženjerstvo

Nativni proteini zauzimaju samo mali dio svih mogućih kombinacija aminokiselina. Naprimjer, za slijed od 200 aminokiselina moguće je 20^{200} različitih polipeptidnih sljedova, dok je za sada identificirano oko 10^{12} proteina. Sve ostalo ostavlja prostor za proteinski inženjering.¹ (slika 1.) Dvije su glavne strategije koje se koriste u proteinskom inženjrstvu¹, jedna od njih je usmjerena evolucija koja predstavlja strategiju za poboljšanje već postojećih proteina uvođenjem nasumičnih mutacija.² Druga je *de novo* dizajn gdje se generiraju proteini sa sekvencama nepoznatima u prirodi, ali se i dalje slijede fizikalni principi.¹



Slika 1. Prikaz svih mogućih kombinacija aminokiselinskih sljedova označen je plavom, a žutom bojom prikazani su nativni proteini. Usmjerena evolucija ograničena je samo za područje oko nativnih proteina, dok je u dizajnu *de novo* na raspolaganju čitava plava površina. Slika je preuzeta iz P. S. Huang, *et. al.*, *Nature* **537** (2016) 320–327.

U ovome radu objasnit će se način rada softvera engl. *Protein Energy Landscape Exploration* (PELE) koji svoju primjenu nalazi u određivanju energetski najstabilnijih konformacija pri *de novo* dizajnu proteina. Upotreba softvera pokazat će se na dvaju primjera. Prvi od njih je supstitucija nekoliko aminokiselina na proteinu FraC, koja dovodi do mogućnost razgradnje nPET čestica i time postaje nanoreaktor za razgradnju, po zdravlje iznimno opasnih, čestica mikroplastike. Svoju potencijalnu primjenu nalazi u centrima za gospodarenje otpadnih voda.⁵ Drugi je primjer poboljšanje katalitičkih mogućnosti serinske esterske hidrolaze ugradnjom dodatnog katalitičkog mjesta, što rezultira i mogućnošću katalize kaskadne reakcije.³

Nedavnim razvojem tehnologije stvoreni su i modeli proteinskog jezika (engl. *protein language models*) koji se oslanjaju na metode dubokog učenja. Modeli proteinskog jezika uče iz postojećih kombinacija sekvenci i struktura proteina. Postali su moćno sredstvo u generiranju sekvenci, predviđanju sklapanja proteina, ali i prevođenju željene strukture proteina u aminokiselinski slijed.⁹ U ovome radu, dodatno, je prikazana neuronska mreža ProteinMPNN koja se pokazala kao iznimno uspješan alat za određivanje proteinske sekvence iz stukture.¹⁰

Ovaj rad oslikava samo odabrane aspekte proteinskog inženjeringa, kao dinamičnog područja punog potencijalnih rješenja za mnoge probleme suvremenog društva.

§ 2. ENERGETSKA OPTIMIZACIJA STRUKTURA

Uvođenje određenih promjena u proteinskoj strukturi može dovesti i do novih svojstava koja nativni protein inače ne posjeduje. Na primjeru proteina FraC je pokazano kako uz svega nekoliko izmijenjenih aminokiselina, protein FraC postaje katalitička nanopora za razgradnju čestica mikroplastike.⁵ Osim toga, moguće je poboljšati svojstva već postojećih enzima, što se dogodilo ugradnjom dodatnog katalitičkog mjesta u enzim serinske esterske hidrolaze. Osim poboljšanih katalitičkih svojstva, omogućena je i kataliza kaskadne reakcije.³ U oba primjera korišten je softver PELE (engl. *Protein Energy Landscape Exploration*) koji računa potencijalnu energiju različitih struktura i konformacija u svrhu pronalaska one energetski najstabilnije. Odnosno, prilikom racionalnog dizajna proteina PELE ima važnu ulogu u probiru onih struktura i konformacija koje zadovoljavaju energetski kriterij.⁶

2.1. PELE – Protein Energy Landscape Exploration

Sama metoda mapiranja sastoji se od triju koraka:

- (1) lokalnih perturbacija,
- (2) predviđanja bočnih ogranaka i
- (3) minimizacijskih algoritama.⁶

2.1.1. Lokalne perturbacije

Procedura započinje stvaranjem lokalnih perturbacija liganda. Perturbacije su u ovome slučaju sitni pomaci u položaju liganda. Unatoč tome što se mnogi ligandi promatraju kao kruta tijela koja imaju tri rotacijska i tri translacijska stupnja slobode, neki ligandi, poput palmitata, imaju dodatne stupnjeve slobode koje su rezultat promjene torzijskih kutova veza u kojima je moguća rotacija. Sve veze u kojima je prisutna torzija potrebno je tako i deklarirati prije izvođenja računa. Duljine veza i kutovi među njima smatraju se konstantnima. Ukoliko postoji sterička smetnja između liganda i proteina ili između liganda samoga (ako se radi o fleksibilnom ligandu), perturbacija se odbacuje. Između 400 perturbacija odabire se ona s najboljim rezultatom. Perturbacije se boduju po OPLS-AA polju sila koje u obzir uzima samo interakcije liganda i proteina. To jest, interakcije otapala su zanemarene.⁶ U novijim istraživanjima, dodane su i perturbacije proteina koje koriste reducirani harmonički model.⁷

2.1.2. *Predviđanje bočnih ogranaka*

U sljedećem koraku algoritam optimizira bočne ogranke aminokiselina u blizini perturbiranih atoma. Isprobavaju se mogući rotameri bočnih ogranaka aminokiselina, odnosno promjena torzijskih kutova veza u bočnom ogranaku sa rezolucijom od 10° .⁶ Ukoliko se rotameri ispituju na takav način, njihov je broj iznimno velik, naprimjer za 2086 rotamera za lizin. Veliki broj rotamera pokušava se smanjiti tako da se ispituju samo energetske najstabilniji rotameri.⁸ Ukoliko imamo i perturbaciju proteina, u ovom koraku odabiru se oni ogranci koji su imali najveću promjenu u energiji. Tako je moguće pronaći one bočne ogranke u kojima se dobilo najveće povećanje energije i odrediti poziciju koja bolje odgovara u smislu smanjenja energije.⁷

2.1.3. *Minimizacijski algoritmi*

Posljednji korak je energetska minimizacija svih dijelova strukture koji su sudjelovali u prethodna dva koraka. Ovaj korak osigurava prilagodbu proteinske okosnice na napravljene perturbacije. Detaljnije, koristi se pretpostavka da će bočni ogranci aminokiselina koji se nalaze u lancu odgovoriti na promjene koje su se dogodile u položaju liganda ili promjene u interakcijama u samom proteinu pa će i okosnica proteina slijediti taj odgovor.⁶

Ova tri koraka kumulativno se vrednuju po kriteriju Metropolis. Rezultat može biti prihvaćanje, odnosno definiranje novoga minimuma ili odbijanje. Konformacijske promjene zatim se šire izvan lokalnog područja. Korisnik sam odabire radijus unutar kojeg će se događati strukturne promjene, bilo lokalne ili, kasnije, izvan lokalnog područja.⁶

2.2. Nanoreaktori za razgradnju čestica mikroplastike

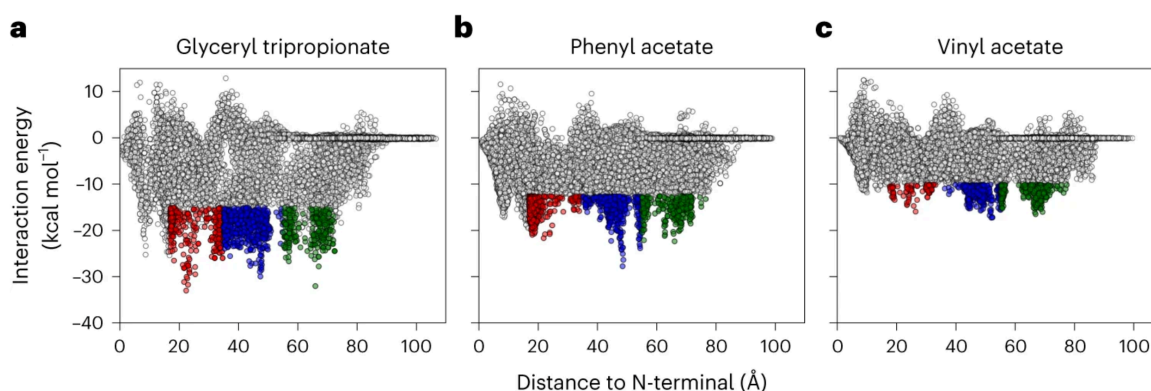
2.2.1. Mikroplastika

Akumulacija plastike u okolišu predstavlja sve veći globalni problem. Jedan od razloga je njezina stabilnost, budući da je plastika polimer sintetičkog podrijetla. Mehanička degradacija plastike u okolišu dovodi do nastajanja čestica koje se nazivaju mikroplastika (5 mm – 1 μ m), sub-mikroplastika (100 nm – 1 μ m) i nanoplastika (1 nm – 100 nm). Istraživanja su pokazala da PET čestice (polietilentereftalat) čine više od 5% identificiranih plastičnih čestica⁵, koje se mogu pronaći u zraku nekih gradova, pitkoj vodi i ledu s planina, čak i Antarktike. Sljedeći odlomci pokazuju kako su znanstvenici funkcionalizirali protein oblika pore fragaceatoxin C (FraC) ugradnjom dva PET hidrolitička mjesta alatima za računalno strukturno modeliranje, a potom i eksperimentalno potvrdili mogućnost razgradnje nanoPET (nPET) čestica.

2.2.2. Stvaranje nanopora za katalitičku razgradnju mikroplastike

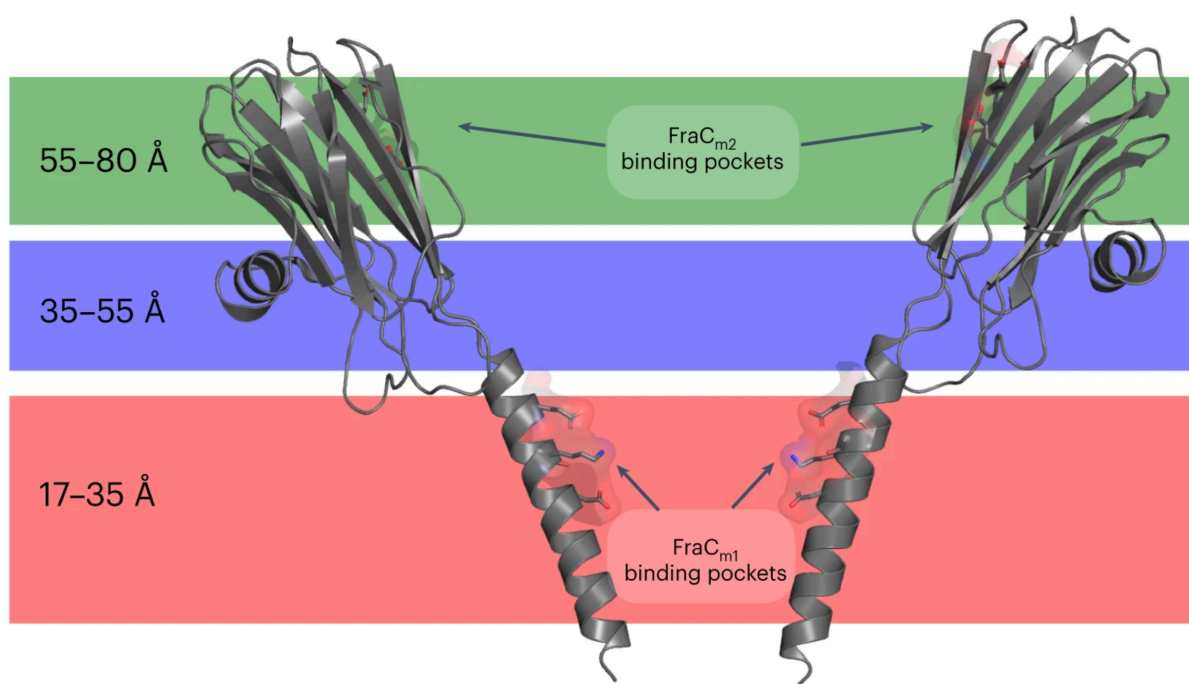
Protein fragaceatoxin C (FraC) je oktamer koji ima izgled pore u obliku slova V. Duljina mu je 7 nm, veličina *cis*-otvora je 6,7 nm, a *trans*-otvora 1,9 nm. FraC sam po sebi nije enzim i nema katalitičku sposobnost, stoga su dva aktivna mjesta nadodana računalnim strukturnim modeliranjem, odnosno proteinska pora je funkcionalizirana kako bi postala enzim i dobila katalitičku mogućnost hidrolize estera.

Unutarnja površina divljeg tipa proteina FraC (FraC_{WT}) pretražena je softverom PELE u svrhu pronalazjenja optimalnog mjesta za vezanje supstrata koje će biti pretvoreno u aktivno mjesto budućeg enzima. Za ispitivanje su upotrijebljena tri estera: gliceril tripropionat, fenil acetat i vinil acetat. Oni su odabrani jer ih većina esteraza uspješno hidrolizira. Rezultati simulacija programa PELE su energijski profili unutarnje površine proteina FraC_{WT} (slika 2).



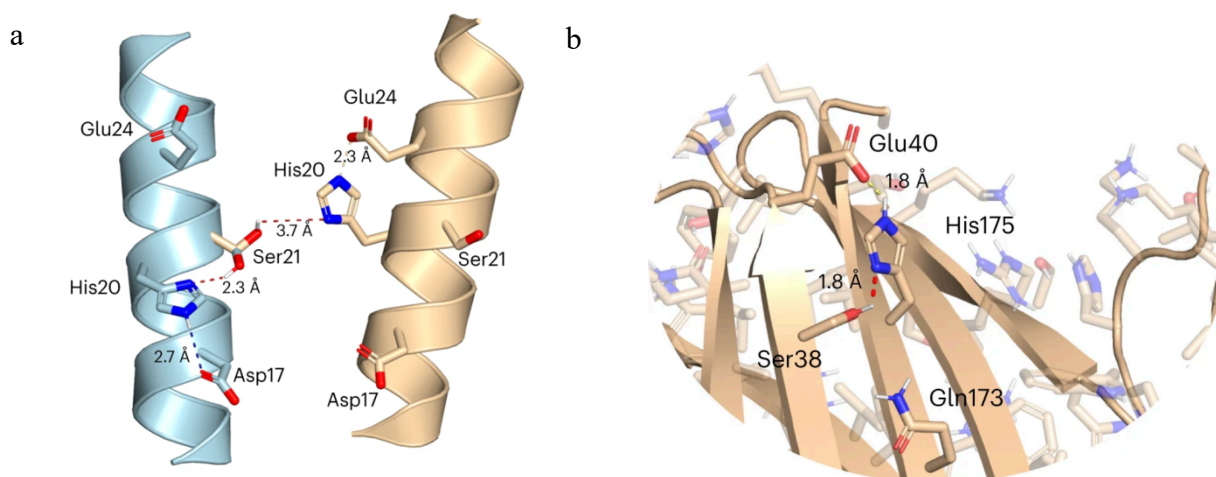
Slika 2. Energetski profili pore dobiveni softverom PELE za tri različita supstrata: (a) gliceril tripropionat, (b) fenil acetat i (c) vinil acetat. Obojena su mjesta ako je energija interakcija supstrata i proteina manja od (a) 15 kcal/mol (b) 12,5 kcal/mol i (c) 10 kcal/mol. Slika je preuzeta iz iz A. Robles-Martín, *et. al.*, *Nat. Catal.* **6** (2023) 1174–1185.

Za sva tri supstrata potencijalna mjesta za dizajn aktivnog mjesta pronađena su na sličnim udaljenostima (slika 3). Poru je moguće podijeliti u tri domene. Pri *trans*-otvoru nalazi se globularna domena, dok se pri *cis*-otvoru nalazi transmembranska regija sačinjena od N-krajeva alfa zavojnica. Između te dvije domene smještena je regija koja povezuje monomere i karakterizirana je prisustvom očuvanih aminokiselina koje su uključene u konformacijske promjene u svrhu stvaranja nanopore. Ona nije odabrana za mutacije zbog gubitka mogućnosti sklapanja monomera u poru. Regije u pori pri otvorima pokazale su se pogodnima za ugradnju aktivnih mjesta.



Slika 3. Prikaz proteina FraC i njegova zrcalna preslika, zajedno s lokacijama aktivnih mjesta: FraC_{m1} i FraC_{m2}. Slika je preuzeta iz iz A. Robles-Martín, *et. al.*, *Nat. Catal.* **6** (2023) 1174–1185.

Odabrana su dva mjesta koja svojim oblikom i veličinom pogoduju zadržavanju i tuneliranju supstrata. Prvo od dva dizajnirana (FraC_{m1}) katalitička mjesta generirano je zamjenama Lys20His i Thr21Ser, a drugo (FraC_{m2}) zamjenom Asp38Ser pri čemu je zamjena Glu173Gln uvedena radi stvaranja oksianionske šupljine (slika 4). Simulacijama je pokazano kako ova aktivna mjesta potencijalno mogu katalizirati reakciju razgradnje esterskih supstrata, uključujući TE (mono(2-hidroksietil)tereftalna kiselina) i/ili ETE (bis(2-hidroksietil)tereftalat). Nadalje, potrebno je ustanoviti mogu li PET čestice uopće dospjeti do aktivnih mjesta u pori. U tu svrhu napravljene su simulacije fleksibilnosti PET lanca duljine 200 monomera. Pokazalo se da su PET lanci dinamični i da se njihov oblik i konformacija mijenjaju. U molekulskom dokiranju pronađeno je više konformacija koje odgovaraju veličini i obliku pore.

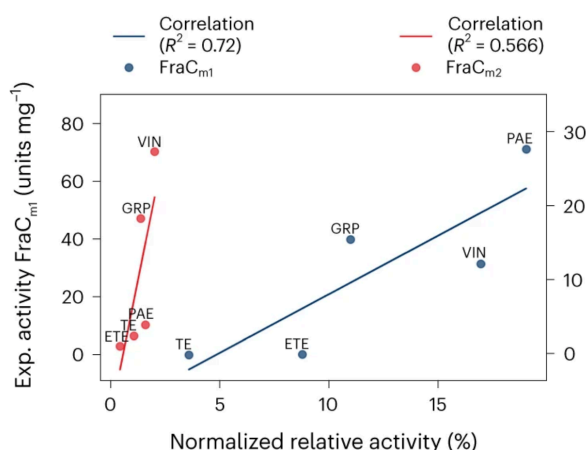


Slika 4. (a) Aktivno mjesto FraC_{m1} gdje je katalitička trijada formirana od Ser21, His20 i Glu24/Asp17. (b) Aktivno mjesto FraC_{m2} gdje je katalitička trijada formirana od Ser38, His175 i Glu40 i Gln173 koji ovdje ima ulogu u oksianionskoj šupljini. Slika je preuzeta iz A. Robles-Martín, et. al., Nat. Catal. 6 (2023) 1174–1185.

Zatim su proteini FraC_{WT} , FraC_{m1} i FraC_{m2} nadeksprimirani u bakteriji *Escherichia coli* i pročišćeni. Oligomerizacija dobivenih proteina u oktamere odvijala se umetanjem proteina na prazne nanodiskove koji su modeli membrane kako bi se preslikalo prirodno stanje akvaporina. I tako su dobiveni oktameri: np FraC_{m1} i np FraC_{m2} .

2.2.3. Katalitičke mogućnosti nanopora

U eksperimentalnim uvjetima oba ugrađena proteina su aktivna i uspješno hidroliziraju TE, ETE i sva tri estera koja su bila korištena u računalnim simulacijama. Također je uočena i korelacija između izračunate PELE-normalizirane relativne katalitičke aktivnosti i eksperimentalne aktivnosti (slika 5). Ukoliko je supstrat *p*-nitrofenol propionat, npFraC_{m1} i npFraC_{m2} pokazuju maksimalnu katalitičku aktivnost pri pH 9 i temperaturi između 35°C i 45°C.

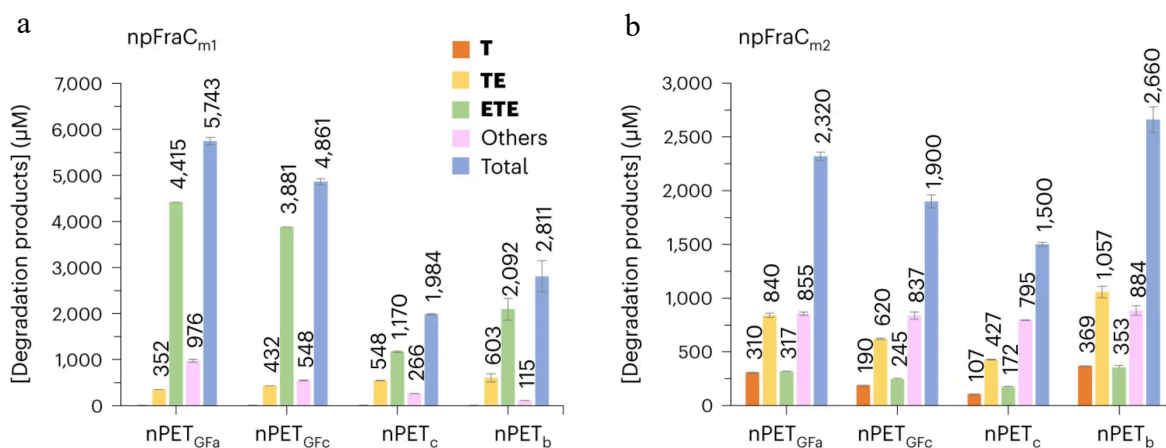


Slika 5. Korelacija između računске PELE normalizirane aktivnosti i eksperimentalne katalitičke aktivnosti za hidrolizu estera gliceril tripropionata (GRP), vinil-acetata (VIN), fenil-acetata (PAE), TE i ETE. Eksperiment se odvijao pri $t = 40^\circ\text{C}$ i pH 8,0. Slika je preuzeta iz A. Robles-Martín, *et. al.*, *Nat. Catal.* **6** (2023) 1174–1185.

Do sad je pokazano kako nanopore s ugrađenim aktivnim mjestima (npFraC_{m1} i npFraC_{m2}) mogu hidrolizirati TE i ETE estere, te se sada postavilo pitanje mogu li hidolizirati i nPET čestice. Za provođenje eksperimenta odabrano je proizvesti nPET čestice *in vitro* kako bi se dobio uvid u fizikalno-kemijska svojstva, te razumio proces razgradnje te produkti razgradnje. Odabrana su četiri različita PET proizvoda iz kojih su dobivene četiri vrste nPET čestica (nPET_b dobivene su iz PET boca za svakodnevnu upotrebu, nPET_c iz PET peleta koji služe kao sirovina za proizvodnju ambalaže, nPET_{GF}, iz kristaličnog PET filma nPET_{GFb}, iz amorfnog PET filma.) Dobivene su nPET razlikuju se po promjeru i njihove su veličine između 53,0 i 108,1 nm. npFraC_{m1} ili npFraC_{m2} dodani su u koncentraciji od 76 nmol/L u otopinu pH 7 koja je sadržavala $2,23 \pm 0,06$ mg/mL nPET čestica, a produkti razgradnje razdvojeni su HPLC-om. Eksperiment

je trajao 48 sati i odvijao se na temperaturi od 40°C jer je se ona prethodno pokazala kao optimalnom za hidrolitičku aktivnost supstrata *p*-nitrofenil propionata. Nakon provedbe eksperimenta u trajanju od 48 sati npFraC_{m1} i npFraC_{m2} zadržali su redom 54,7 ± 2,1% i 32,7 ± 6,3% početne aktivnosti. Identitet dobivenih produkata razgradnje potvrđen je masenom spektrometrijom.

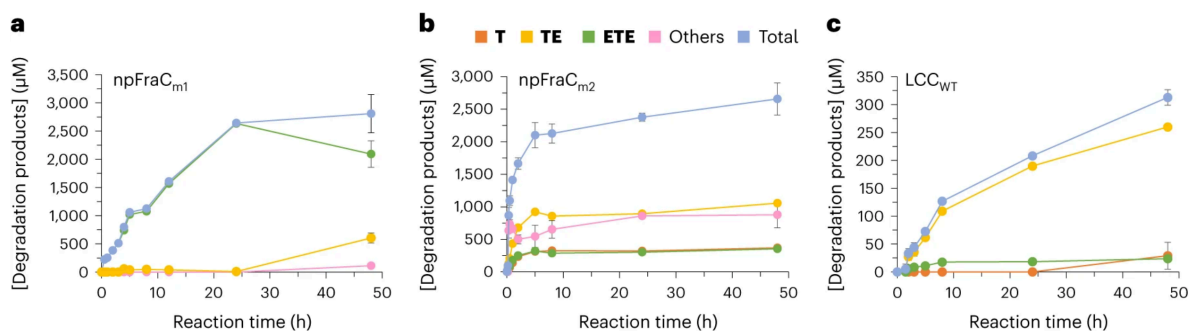
Glavni produkt razgradnje nPET čestica sa npFraC_{m1} je ETE, a detektirani su i ETETE, TE i TETETE. Kod npFraC_{m2} uočila se veća raznolikost oligomera TET, TETE, TETET i TETETE, uz naravno tereftalnu kiselinu T, TE i ETE. Generalno tereftalna kiselina, uz etilen glikol, finalni je produkt razgradnje polimera. Pod istim eksperimentalnim uvjetima enzimi *IsPET*aza, LCC_{WT} i LCC_{WCCG} proizveli su T, TE, ETE, zatim oligomere TET, TETE i/ili TETETE, pri čemu su T i TE su bili glavni produkti razgradnje (slika 6). npFraC_{m1} pokazuje prednost nad npFraC_{m2}, *IsPET*azom, LCC_{WT} i LCC_{WCCG} zbog toga što T nije produkt njegove razgradnje.



Slika 6. Koncentracije produkata degradacije enzima (a) FraC_{m1} i (b) FraC_{m2}. Kvantitativni rezultati dobiveni su za produkte od T do ETETE. Određeno metodama semipreparativnog HPLC-a, a zatim masenom spektrometrijom. Slika je preuzeta iz A. Robles-Martín, *et. al.*, *Nat. Catal.* **6** (2023) 1174–1185.

Kako bi se dobio bolji uvid razgradnju nPET čestica, provedeni su eksperimenti u kojima se prati tijek reakcija. npFraC_{m1} i npFraC_{m2} uspoređeni su sa LCC_{WT}, iako je *IsPET*aza najučinkovitiji enzim divljeg tipa za razgradnju PET, u eksperimentalnim uvjetima stabilniji je LCC_{WT}. Kao supstrat odabrane su nPET_b čestice, jer su najbližnje onima u prirodi. Pokus se provodio pri 40°C. U početku reakcije s npFraC_{m2} vidljiva je puno veća brzina hidrolize, nego

kod npFraCm1. Dok je npFraCm2 značajno aktivna samo 5-6 sati, aktivnost npFraCm1 seže i do 24 sata kada se očituje maksimalna aktivnost. Za usporedbu, aktivnost LCC_{WT} nastavlja se iza 24 sata (slika 7).



Slika 7. Praćenje kinetike reakcije razgradnje nPET_b čestica enzimima (a) npFraC_{m1} (b) npFraC_{m2} i (c) LCC_{WT}. Pri početnoj koncentraciji nPET_b od 1,1 mg/mL, enzima 1,5 g/mL, $t = 40^{\circ}\text{C}$, pH 7,0. Slika je preuzeta iz A. Robles-Martín, *et. al.*, *Nat. Catal.* **6** (2023) 1174–1185.

Još je preostalo ispitati kako veličina čestica utječe na uspješnost katalitičke reakcije. Pripremljene su nPET_b čestice različitih veličina između 69,1 nm i 153,8 nm. Dobiveno je kako se manje čestice bolje razgrađuju u FraC_{m1}, a FraC_{m2} preferira dimenzije između 85,4 i 108 nm.

Navedeno istraživanje⁵ pokazuje kako se dizajnirane katalitičke nanopore mogu jednostavno proizvesti i koristiti u pročišćavanju voda. Dizajnirane katalitičke nanopore su stabilne i predstavljaju alat za razgradnju PET čestica sub-mikro i nano veličina, pri temperaturama oko 40°C.

2.3. Unaprjeđenje i proširenje katalitičkih mogućnosti

Potreba za povećanom učinkovitosti enzima i proširenje njihovih katalitičkih mogućnosti u sve je većem porastu. Ugradnja dodatnog broja aktivnih mjesta u jedinični enzim, u svrhu sinergističkog djelovanja aktivnih mjesta kako bi se povećala učinkovitost katalize ili kako bi se proširio kemijski aspekt katalize, nameće se kao logičan pristup. Enzim koji ima ugrađeno dodatno aktivno mjesto nazvan je engl. *plurizyme*. Znanstvenici³ su na primjeru serinskih esterskih hidrolaza (PDB kod: 5JD4) pokazali kako ugradnja dodatnog aktivnog mjesta ne samo da poboljšava katalitički učinak navedenog enzima, već i uvodi dodatna svojstva koja nije moguće postići u više komponentnom okolišu gdje se dva aktivna mjesta nalaze u dva odvojena enzima.

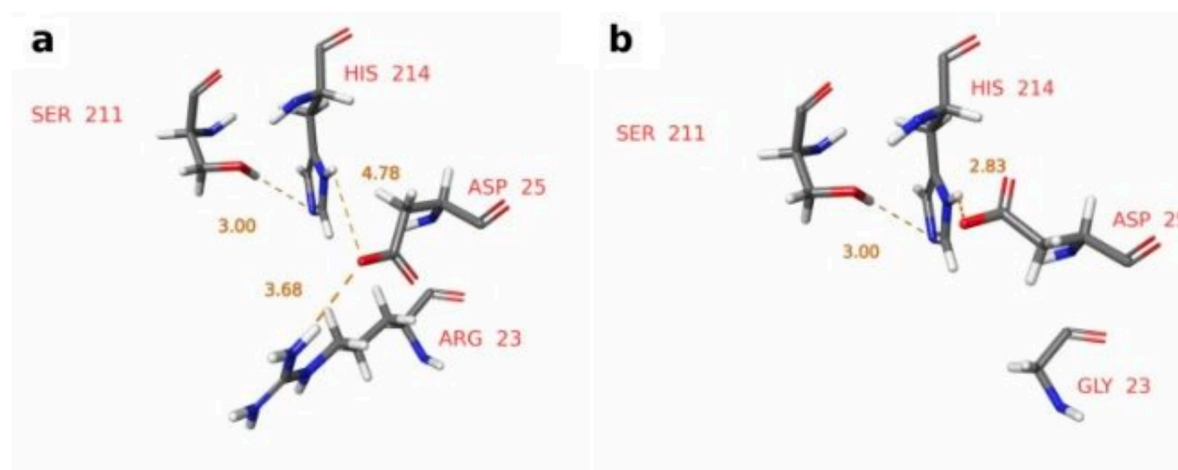
Tekst u nastavku pokazuje kako su znanstvenici³, uvođenjem preinaka u divljem tipu serinskih esterskih hidrolaza, poboljšali, zatim i promijenili njegova katalitička svojstva.

2.3.1. Uvođenje dodatnog aktivnog mjesta u svrhu poboljšanja katalitičkih mogućnosti

U svrhu pretrage enzima za potencijalna mjesta novih katalitičkih trijada upotrijebljen je softver engl. *Protein Energy Landscape Exploration (PELE)*³. Nativna katalitička trijada (u divljem tipu enzima koji je nazvan EH1A) sačinjena je od Ser161, Asp256 i His286. Računima izvedenim u softveru PELE, pronađena je odgovarajuća pozicija za novo katalitičko mjesto u enzimu. Generiran je novi enzim (EH1B) koji sadrži sljedeće izmjene: Glu25Asp, Leu214His, pri čemu je Ser211 zadržan kako bi se stvorilo novo aktivno mjesto, dok izmjena Ser161Ala služi u svrhu inaktivacije nativnog katalitičkog mjesta. Tako je dobivena nova katalitička trijada sastavljena od Ser211, Asp25 i His214. Konačno, generiran je i enzim EH1AB isključivo s dvije mutacije sekvence divljeg tipa (EH1A): Glu25Asp, Leu214His. EH1AB sadrži aktivno mjesto koje je inače prisutno u divljem tipu EH1A (Ser161, Asp256 i His286) i ono aktivno mjesto koje je prisutno u umjetno generiranom enzimu EH1B (Ser211, Asp25 i His214). Dobiveni enzim EH1AB nije pokazao katalitičku prednost nad divljim tipom, iako je imao dvije aktivne katalitičke trijade.

Zbog nedostatne prednosti enzima EH1AB koji sadrži dva katalitička mjesta, nad divljim tipom EH1A koji sadrži jedno katalitičko mjesto, pokrenut je novi set mapiranja površine uz pomoć

prethodno spomenutog softvera PELE na enzimu EA1B. Otkriveno je kako Arg23 ometa ulogu Asp25 u nadodanoj katalitičkoj trijadi, tako što tvori elektrostatske interakcije, onemogućavajući sudjelovanje aspartata u katalitičkoj trijadi (Slika 8a). Ova smetnja može se zaobići uvođenjem izmjene Arg23Gly (Slika 8b), koja je potom učinjena na sva tri enzima: EH1A, EH1B i EH1AB, te je stvoren novi slijed enzima: EH1A1, EH1B1, EH1AB1. Pokazalo se da uvođenjem ove izmjene, novom plurizimu EH1AB1 vrijednosti k_{cat} i k_{cat}/K_M značajno su povećane u katalitičkoj reakciji hidrolize estera, dok je temperaturni raspon u kojem enzim zadržava više od 80% svoje aktivnosti proširen za 20°C. Također, pokazalo se da EH1AB1 hidrolizira svaki od 78 isprobanih supstrata.



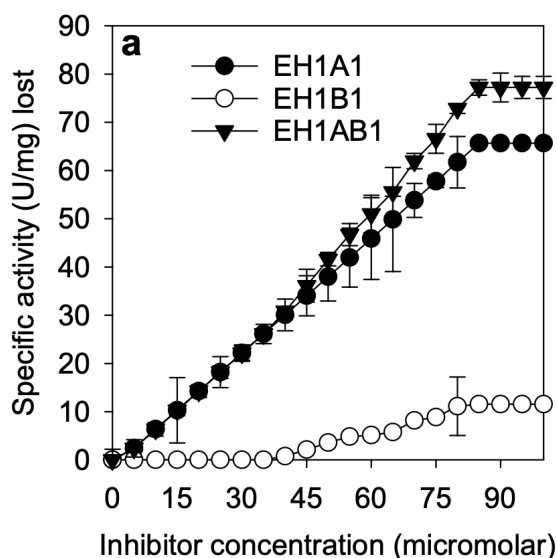
Slika 8. Struktura generirane katalitičke trijade: (a) EH1B i (b) EH1B1. Generiranje je izvršeno simulacijama molekulske dinamike. Mutacija Arg23Gly uzrokuje smanjenje distance His214 – Asp25, što bi trebalo poboljšati katalitičku sposobnost ovog aktivnog mjesta. Sve su udaljenosti izražene u Å. Slika je preuzeta iz S. Alonso, *et. al.*, *Nat. Catal.* **3** (2020) 319–328.

Kako bi se provjerilo jesu li oba katalitička mjesta enzima EH1AB1 aktivna provedena je strukturalna analiza. Priređeni su kokristali enzima s pripadajućim inhibitorom (4-nitrofenil heksilfosfonat). 4-nitrofenil heksilfosfonat pripada skupini suicidalnih inhibitora. To su spojevi koji se vežu za aktivno mjesto enzima i umjesto pretvaranja u produkt, bivaju čvrsto vezani za aktivno mjesto i tako onemogućuju daljnje katalitičko djelovanje.⁴ Pregledom kristalne strukture dobivene difrakcijom X-zraka na kristalu pokazalo se kako se inhibitor veže za Ser161 i za Ser211, odnosno uzrokuje kovalentnu modifikaciju koja inhibira enzim u potpunosti. Time je dokazana katalitička aktivnost oba aktivna mjesta.

2.3.2. Kataliza kaskadne reakcije

S ciljem dizajniranja enzima s biološkim i nebiološkim aktivnim mjestom znanstvenici su proveli daljnje istraživanje. Za početak, uzet je enzim EH1AB1 i odabrana je strategija suicidalnog inhibitora. Ovaj pristup temelji se na odabiru analoga supstrata kojeg će enzim kovalentno vezati u reakciji koja sliča redovnoj katalitičkoj reakciji tog enzima. U ovom slučaju, dobro je poznato da se fosfonatni esteri stehiometrijski vežu za serin katalitičke trijade. Kao inhibitor odabran je 3'-hidroksi-2,2'-bipiridin-3-il metil heksilfosfonat, koji se dodatno može obogatiti ionom prijelaznog metala koji ima tendenciju stvaranja kelata. Cilj je postići specifičnu ugradnju inhibitora u jedno od dva aktivna mjesta. Pokazalo se kako uvođenje prethodno navedenog suicidalnog inhibitora može poslužiti kao baza za uvođenje koordiniranog Cu^{2+} iona u enzim. Unatoč tome, nije postignuta željena selektivnost pri odabiru aktivnog mjesta, te je CuSI (kompleks bakra i suicidalnog inhibitora) vezan za oba aktivna mjesta. Kako bi se približilo rješenju tog problema, ispitana je selektivnost dvaju aktivnih mjesta pri različitim koncentracijama inhibitora. To je učinjeno ispitivanjem enzimske aktivnosti sva tri enzima iz serije (EH1A1, EH1B1 i EH1AB1) pri različitim koncentracijama inhibitora.

U koncentracijama inhibitora ispod $35 \mu\text{mol/L}$, nije zabilježen gubitak aktivnosti enzima EH1B1. Dok je kod EH1A1 i EH1AB1 prisutan gubitak enzimske aktivnosti pri koncentracijama inhibitora manjima od $35 \mu\text{mol/L}$. Nadalje, pri koncentracijama inhibitora iznad $35 \mu\text{mol/L}$ prisutan je značajniji gubitak enzimske aktivnosti kod EH1AB1 radi sumarnog efekta dvaju aktivnih mjesta: Ser161 i Ser211 (slika 9).



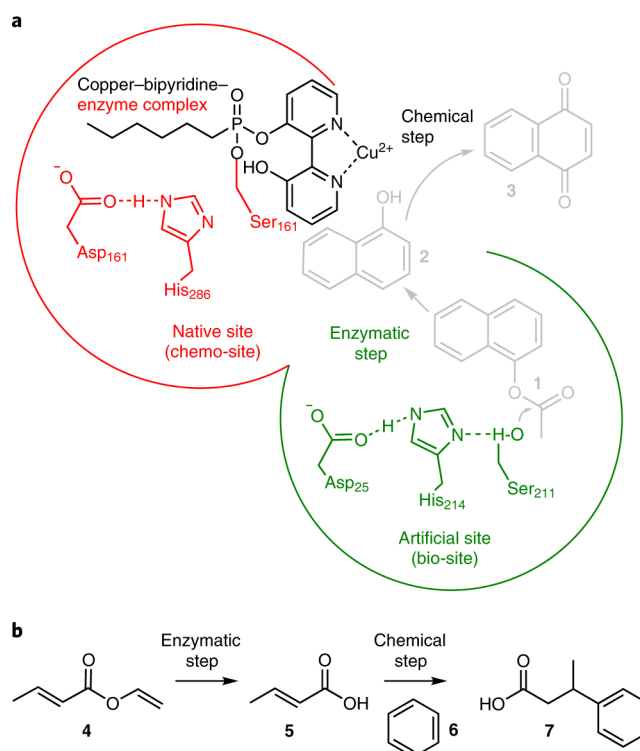
Slika 9. Prikaz gubitka enzimске aktivnosti EH1A1, EH1B1 i EH1AB1 pri različitim koncentracijama inhibitora (hidroksi-2,2'-bipiridin-3-il metil heksilfosfonat). [enzim] = 85 $\mu\text{mol/L}$, $t = 25^\circ\text{C}$, u 50 mmol/L K_2HPO_4 puferu pH = 6,5. Slika je preuzeta iz S. Alonso, *et. al.*, *Nat. Catal.* **3** (2020) 319–328.

Zahvaljujući različitom afinitetu aktivnog mjesta prema inhibitoru pri koncentracijama inhibitora ispod $35\mu\text{M}$, omogućeno je vezanje organo-metalnog kompleksa na Ser161, ali ne i na Ser211. Aktivnost novoga enzima EH1AB1C-B pokazana je na dva primjera: (a) hidroliza estera praćena oksidacijom i (b) Fidel-Castrovo alkiliranje (slika 10).

(a) Polazni supstrat je ester, koji hidrolizira u generiranom aktivnom mjestu (Ser211) do alkohola. Slijedeći korak obavlja nativno aktivno mjesto (Ser161) koje je kovalentno povezano sa CuSI i ima ulogu kemijskog katalizatora koji alkoholnu skupinu oksidira do keto skupine. Preciznije, prvo aktivno mjesto prevodi 1-naftolacetat u 1-naftol, a u slijedećem koraku događa se oksidacija do 1,4-naftokinona. Iskorištenje reakcije od 100% nakon 2 sata.

U reakciji (b) polazni supstrat je također ester, koji u generiranom aktivnom mjestu (Ser161) hidrolizira do masne kiseline. Nadalje, uz dodatak benzena, odvija se Fidel-Castrovo alkiliranje u kemijskom katalitičkom mjestu i konačni produkt reakcije je 3-fenilbutanska kiselina.

Konkretno, u enzimsku reakciju ulazi vinil krotonat, koji se hidrolizira do krotonske kiseline, a zatim, uz dodatak benzena, događa se alkiliranje i dobivanje 3-fenilbutanske kiseline. Iskorištenje reakcije od 83% nakon tri dana.³



Slika 10. Prikaz dvaju primjera reakcija koje se odvijaju u enzimu EH1AB1CB. (a) Zelenom bojom prikazana je reakcija u generiranom aktivnom mjestu, dok je crvenom prikazana reakcija u nativnom aktivnom mjestu koje je nadograđeno sa CuSI. Reakcija se odvijala pri $\text{pH} = 8$, $t = 25^\circ\text{C}$, $[\text{supstrat}] = 10 \text{ mmol/L}$. (b) Prikaz reakcije hidrolize popraćene Fidel-Castrovim alkiliranjem. Reakcija je izvedena pri $t = 4^\circ\text{C}$, u trajanju od 3 dana. Slika je preuzeta iz S. Alonso, *et. al.*, *Nat. Catal.* **3** (2020) 319–328.

§ 3. MODELI PROTEINSKOG JEZIKA

Već je prije spomenuto da se proteini sastoje, većinom, od 20 proteinogenih aminokiselina. Primarna struktura proteina određuje tercijarnu strukturu, a samim time i funkciju proteina. Razumijevanje odnosa između strukture i funkcije u fokusu je mnogih istraživanja. Alati koji spadaju u skupinu modela proteinskog jezika koriste metode dubokog učenja kako bi znanstvenicima pomogli u razumijevanju složenih odnosa između sekvence, strukture i funkcije proteina.⁹

3.1. ProteinMPNN

U prethodnim primjerima proteini su funkcionalizirani korištenjem tehnika koje se temelje na energetske optimizaciji, odnosno rađene su male lokalne promjene u vrsti aminokiselina i njezinim konformacijama, zatim je određivano koja je od njih energetski najpovoljnija.

Nedavno je otkriveno da je prilikom određivanja sekvence iz proteinske strukture moguće upotrijebiti modele dubokog učenja. Za razliku od prethodno opisanih tehnika pronalaska energetskih minimuma kod brojnih rotamera, modeli dubokog učenja ne pretražuju rotamere, čime su izbjegnute kalkulacije koje mogu značajno trošiti računalne resurse.¹⁰

3.1.1. Opis modela

U ovom primjeru znanstvenici su koristili vrstu neuronske mreže koja se naziva engl. *Message-passing neural network* (MPNN). Model sadrži tri sloja enkodera, tri sloja dekodera i skriveni sloj dimenzije 128. Mreža predviđa slijed aminokiselina na autoregresivan način, od N-terminalnog prema C-terminalnom kraju. Pri tome se koristi udaljenosti između $C\alpha$ - $C\alpha$, relativne orijentacije i rotacije okvira $C\alpha$ - $C\alpha$ - $C\alpha$, kao i torzijski kutovi okosnice.

Model je prvo unaprijeđen pretvaranjem strukture u aminokiselinski slijed koristeći se poznatim, divljim tipovima jednolančanih proteina. Odabrano je 19700 proteina iz *Protein Data Bank* (PDB) na kojima je model treniran, validiran i testiran.

Različitim dodacima početnom modelu došlo je do bolje mogućnosti obnavljanja sekvence.¹⁰ Termin obnova sekvence u ovome kontekstu označava pretvorbu proteinske strukture u aminokiselinski slijed. Što je uspješnost u obnovi sekvence veća, to je model bolje odredio aminokiselinski slijed iz strukture, u odnosu na referentni aminokiselinski slijed.¹¹

Prvo je uključivanjem udaljenosti između N, C α , C, O i zamišljenih C β , čiji su položaji određeni u ovisnosti o N, C α i C, uspješnost obnavljanja sekvence porasla je s 41,2% (u početnom modelu) na 49,0%. Zatim nadogradnjom bridova, kao i čvorova u enkoderu okosnice uspješnost obnavljanja sekvence povećana je na 43,1%. Kombiniranjem promjena iz prethodna dva eksperimenta dovelo je do mogućnosti obnavljanja sekvence od 50,5%.

Model je do sada dekodirao aminokiseline redom, s lijeva na desno, odnosno od N-terminalnog kraja prema C-terminalnom kraju. U ovome koraku nadogradnje uvedeno je dekodiranje neovisno o redosljedju, što je također, u kombinaciji s prethodnim unaprijeđenima, pridonijelo poboljšanju u obnavljanju sekvence. Dekodiranje neovisno o redosljedju olakšava slučajeve kada je sekvenca u sredini poznata, a potrebno je odrediti ostatak. U samo dekodiranja poznatih regija se preskače, ali ih se uključuje u dekodiranje ostatka sekvence.¹⁰

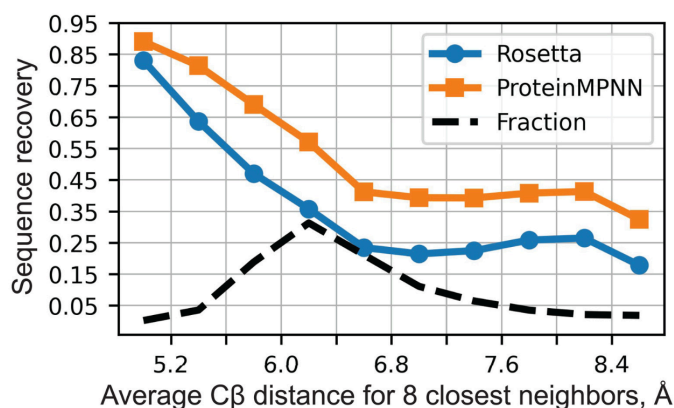
3.1.2. Proteini koji sadrže kvarternu strukturu

Ukoliko se pokušava odrediti primarna struktura proteina koji posjeduju više lanaca, potrebno je doći do istoga rješenja, neovisno o tome koji se lanac prvo određuje. Zato je u model dodana i binarna funkcija koja određuje jesu li dva bočna ogranka, koji ostvaruju interakcije, iz istoga ili drugih lanca. Također, ukoliko se radi o homodimeru, aminokiseline na istim pozicijama moraju biti jednake. To svojstvo umetnuto je u model tako da se određuju dvije razdiobe, za isto mjesto na oba lanca, zatim se iz te dvije razdiobe određuje normalna razdioba iz koje se konačno određuje i aminokiselina.

Za treniranje, validiranje i testiranje modela koji obrađuje proteine s više podjedinica uzeto je više od 23000 proteinskih struktura iz Protein Data Bank (PDB) čije su strukture određene difrakcijom rendgenskog zračenja na kristalu ili kriogenom elektronskom mikroskopijom, a rezolucija je veća od 3,5 Å. Proteini su sadržavali manje od 10000 aminokiselina u svom sastavu i podijeljeni u set za treniranje, potvrđivanje i testiranje.¹⁰

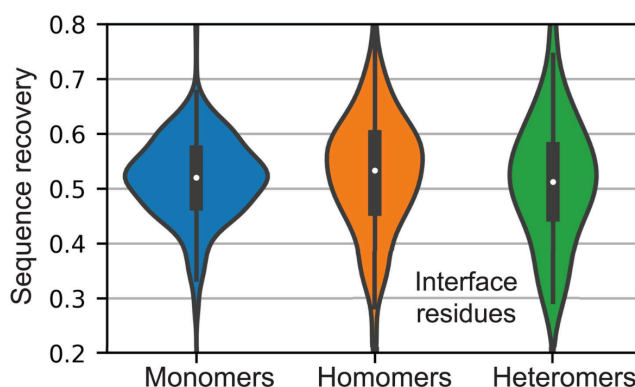
3.1.3. Usporedba modela *in silico*

Za usporedbu modela odabrana su 402 monomera, koji su prevedeni u aminokiselinske sljedove pomoću alata *Rosetta* (koristi pristup pomoću energetske optimizacije) i *ProteinMPNN*-a. *ProteinMPNN* pokazao se efektivnijim, u prosjeku 52,4% nasuprot 32,9% (slika 11).



Slika 11. Usporedba alata *Rosetta* i *ProteinMPNN* u određivanju aminokiselinskog slijeda iz strukture monomera u ovisnosti o prosječnoj udaljenosti C β za 8 najbližih susjeda, odnosno različitoj zakrivljenosti. Slika je preuzeta iz J. Dauparas, *et. al.*, *Science* **378** (2022) 49–56.

Sljedeća evaluacija *ProteinMPNN*-a izvršena je na setu od 690 monomera, 732 homodimera (ispod 2000 AK) i 98 heteromera. Model se pokazao uspješnim za sve tri vrste proteina (slika 12).



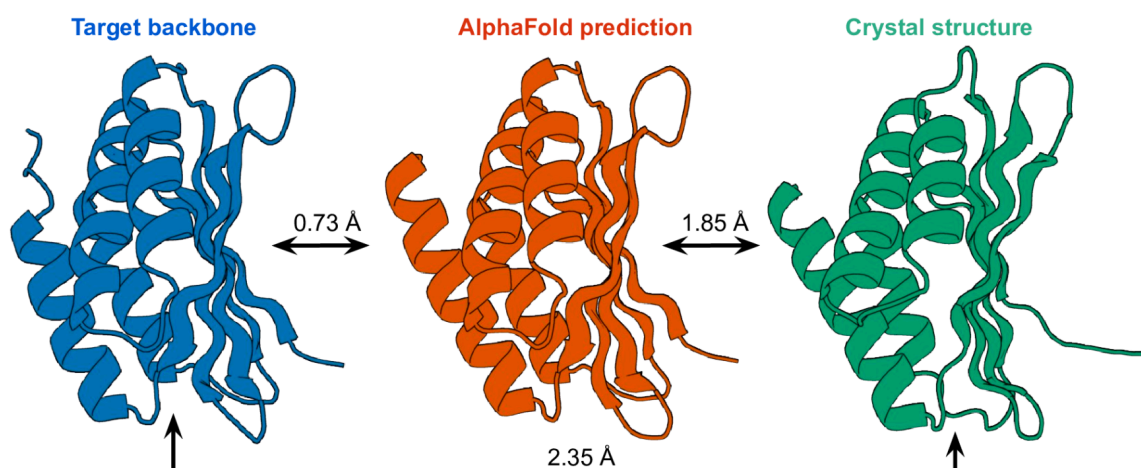
Slika 12. Uspješnost određivanja sekvence za monomere, homodimere i heteromere u alatu *ProteinMPNN*. Slika je preuzeta iz J. Dauparas, *et. al.*, *Science* **378** (2022) 49–56.

AlphaFold i *RoseTTAFold* su alati koji imaju mogućnost predviđanja 3D strukture proteina iz njegove sekvence. Mogu predvidjeti strukturu dosta dobro ako je prisutno poravnanje višestrukih sekvenci, koje generalno sadrži dodatne evolucijske podatke, koji pomažu alatima u točnom predviđanju strukture. Manje su precizni kada na raspolaganju imaju samo jednu sekvencu koju je potrebno sklopiti u 3D strukturu. *ProteinMPNN* ovdje pronalazi ulogu u generiranju sekvenci koje će *AlphaFold* puno bolje i točnije sklopiti nego native sekvence.¹⁰

3.1.4. Eksperimentalna potvrda

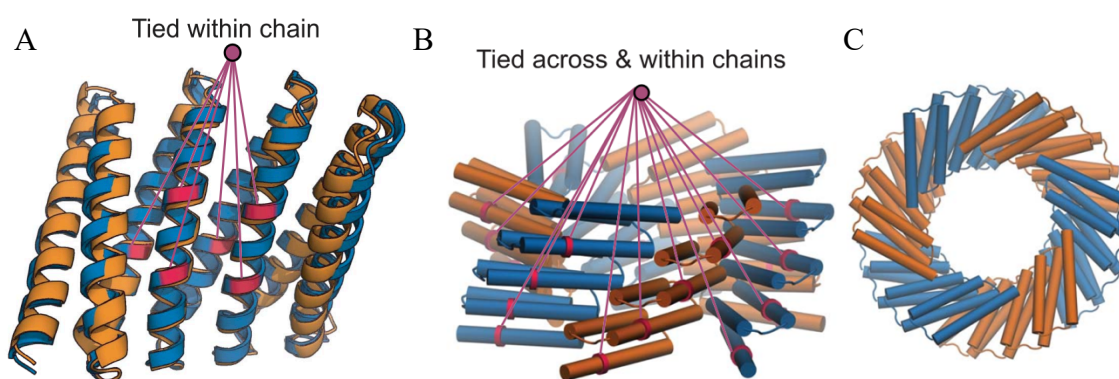
Unatoč uspješnom obnavljanju sekvence *in silico*, pravi test alata *ProteinMPNN* je u eksperimentalnom okruženju, gdje se ispituje sklapanje sekvence i njezina funkcija. Kako bi se evaluirao *ProteinMPNN* odabran je set monomera, čiji dizajn nije uspio pomoću programa *AlphaFold* ili *Rosetta*. Protokol ide tako da su u eksperimentima strukture ostale iste, ali su generirane nove sekvence pomoću *ProteinMPNN*-a. Proteini su zatim eksprimirani u bakteriji *Escherichia coli*, a potom im je određena struktura i biokemijski su karakterizirani.

U prvome eksperimentu znanstvenici¹⁰ su, nasumično u programu *AlphaFold*, generirali ciljni set monomera i oligomera koji su se značajno razlikovali od nativnih proteina. Ti proteini su eksprimirani u *E. coli* i određena im je topivost, koja je iznosila svega 9 mg/L. U sljedećem koraku, sekvence ciljnih struktura ponovo su generirane pomoću *ProteinMPNN*-a. Od 96 pokušaja, 73 je bilo topivo u vodi, sa značajno većim medijanom topivosti koji je iznosio 247 mg/L. Dodatno, 50 njih je imalo zadani oblik monomera ili oligomera što je utvrđeno gel-filtracijom. Mnogi od njih bili su termički stabilni, preciznije sekundarna struktura im je bila očuvana sve do 95°C. Za jednog monomera dobivenog uz pomoć *ProteinMPNN*-a određena je kristalna struktura difrakcijom rendgenskog zračenja na jediničnim kristalu (slika 13). Struktura dobivenog proteina složenija je od većine *de novo* dobivenih proteina. Sadrži pet β -ploča i četiri α -zavojnice. Ovi eksperimentalni rezultati dokaz su da *ProteinMPNN* uspješno pretvara strukture u sekvence. Također, određena je i struktura deset cikličkih homooligomera, gdje se pokazalo veliko podudaranje željenih i eksperimentalno dobivenih struktura, a strukture su eksperimentalno određene ili difrakcijom rendgenskog zračenja na kristalnom uzorku ili elektronskom mikroskopijom.



Slika 13. Plavom bojom prikazana je ciljna struktura željenog proteina (PDB ID: 8CYK), crvenom bojom prikazano je kako je *AlphaFold* predvidio sklapanje aminokiselinskog slijeda dobivenog pomoću *ProteinMPNN*, a zelenom kristalna struktura dobivena difrakcijom rengenog zračenja na kristalu eksperimentalno dobivenog proteina. Slika je preuzeta iz J. Dauparas, *et. al.*, *Science* **378** (2022) 49–56.

Uspješnost *ProteinMPNN*-a eksperimentalno je potvrđena na još složenijim primjerima proteina koji sadrže ponavljajuće dijelove sekvence, a koji su prethodno neuspješno dizajnirani u programu *Rosetta* (slika 14).

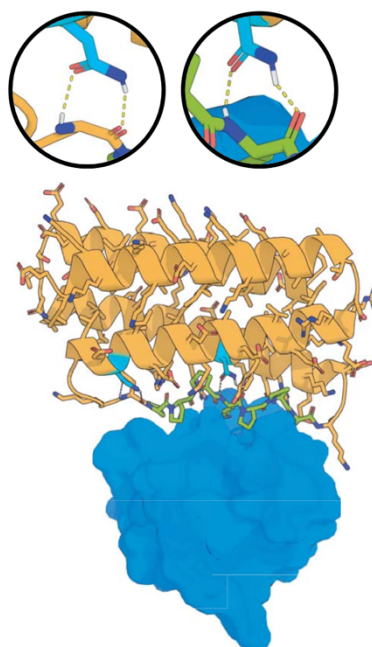


Slika 14. Prikaz složenijih struktura generiranih s *ProteinMPNN*-om. (A) Žuto je označen model, a plavo *AlphaFold*-ovo sklapanje sekvence dobivene u *ProteinMPNN*-u. Crveno su označeni ponavljajući dijelovi sekvence. (B) Crveno su označeni ponavljajući dijelovi sekvence. (C) Prikaz strukture s ponavljajućim dijelovima sekvence.

sekvence, u lancima i između samih lanaca. (C) Pogled na strukturu (B) odozgo. Slika je preuzeta iz J. Dauparas, *et. al.*, *Science* **378** (2022) 49–56.

Kao zadnji i najkompleksniji eksperimentalni test, opet je odabran slučaj u kojem dizajn nije uspješno napravljen uz pomoć programa *Rosetta*. Cilj je napraviti nosač za poliprolinsku zavojnicu, koja ima visoki afinitet za spajanje s SH3 domenom. SH3 domena obično se veže za motive bogate prolinom. Takav motiv sadrži aminokiselinski slijed Pro-Pro-Pro-Arg-Pro-Pro-Lys.

Nosač za navedeni motiv dizajniran je u programu *Rosetta*, no neuspješno, jer se dizajnirana sekvenca ne sklapa u željenu strukturu. Sada se otvara prostor za ProteinMPNN koji uspješno stvara sekvencu, s istim motivom koji je bogat prolinom, koja je također eksprimirana u *E. coli*. Ona se uspješno smotala i eksperimentalno je potvrđena interakcija s domenom SH3 (slika 15). Ovo je još jedan u nizu eksperimentalnih dokaza učinkovitosti *ProteinMPNN*-a, kada tradicionalne metode, kao što je *Rosetta*, zakažu.¹⁰



Slika 15. Povezivanje Grb2 SH3 podjedinice (plavo) sa motivom bogatim prolinom (zeleno) i nosačem (žuto). Slika je preuzeta iz J. Dauparas, *et. al.*, *Science* **378** (2022) 49–56.

§ 4. LITERATURNI IZVORI

1. P. S. Huang, *et. al.*, *Nature* **537** (2016) 320–327.
2. N. Turner, *Nat Chem Biol* **5** (2009) 567–573.
3. S. Alonso, *et. al.*, *Nat. Catal.* **3** (2020) 319–328.
4. D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry*, W. H. Freeman and Company, New York, 2021, str. 200.
5. A. Robles-Martín, *et. al.*, *Nat. Catal.* **6** (2023) 1174–1185.
6. K. W. Borelli, *et. al.*, *J. Chem Theory Comput.* **1** (2005) 1304–1311.
7. B. P. Cossins, *et. al.*, *J. Chem Theory Comput.* **8** (2012) 959–965.
8. M. P. Jacobson, *et. al.*, *Phys. Chem. B* **106** (2002) 11673–11680.
9. J. A. Ruffolo, A. Madani, *Nat Biotechnol* **42** (2024) 200–202.
10. J. Dauparas, *et. al.*, *Science* **378** (2022) 49–56.
11. B. Kuhlman, P. Bradley, *Nat Rev Mol Cell Biol* **20** (2019) 681–697.