

De novo dizajn enzima primjenom računalnih metoda

Skoko, Dominik

Undergraduate thesis / Završni rad

2023

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:217:550221>

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

Download date / Datum preuzimanja: **2024-07-11**



Repository / Repozitorij:

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





Sveučilište u Zagrebu
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
Kemijski odsjek

Dominik Skoko

Student 3. godine Preddiplomskog sveučilišnog studija KEMIJA

***De novo* dizajn enzima primjenom računalnih metoda**

Završni rad

Rad je izrađen u Zavodu za biokemiju

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

Zagreb, 2023.

Datum predaje prve verzije Završnog rada: 25. rujna 2023.

Datum ocjenjivanja Završnog rada i polaganja Završnog ispita: 27. rujna 2023.

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

Potpis:

Sadržaj

§ SAŽETAK.....	VII
§ 1. UVOD.....	1
1.1. Sklapanje proteina	2
1.2. Sinteza proteina.....	5
1.3. Reakcija Kempove eliminacije.....	6
§ 2. METODE U DIZAJNU PROTEINA	7
2.1. Usmjeren evolucija.....	7
2.2. Racionalni dizajn.....	8
2.3. Poluracionalni dizajn.....	10
§ 3. METODE DE NOVO SINTEZE PROTEINA.....	12
3.1. Stvaranje okosnice	13
3.2. Dizajniranje sekvence	16
3.3. Kompatibilnost između sekvence i strukture	18
3.4. Eksperimentalna karakterizacija i udio uspješnosti dizajna	20
§ 4. DE NOVO SINTEZA KEMPOVE ELIMINAZE	21
4.1. Geometrija prijelaznog stanja	21
4.2. Smještanje prijelaznog stanja i dizajn proteina	22
4.3. Eksperimentalna karakterizacija	23
4.4. Optimizacija enzima usmjerenom evolucijom	25
§ 5. LITERATURNI IZVORI.....	XXVII

§ Sažetak

U biološkom svijetu, proteini predstavljaju jednu od temeljnih jedinica života. Zbog svoje složenosti i visoko prilagodljive strukture mogu imati raznovrsna svojstva i funkcije što ih čini primamljivim za primjenu u industriji. Dobrim poznavanjem procesa kojim nastaju, zakonitosti koji uvjetuju njihovu strukturu i funkciju te razvojem metoda za njihovu sintezu, danas je moguće dizajnirati proteine *de novo* korištenjem računalnih metoda. Kroz ovaj rad razmotriti će se zakonitosti koje uvjetuju sklapanje proteina, koraci kroz koje se razvijalo polje proteinskog dizajna i računalne metode koje se koriste danas u *de novo* dizajnu. Detaljnije će se razmotriti jedan zanimljiv primjer primjene ove tehnologije, dizajniranje enzima koji katalizira Kempovu eliminaciju, reakciju za koju ne postoji niti jedan prirodni enzim.

§ 1. UVOD

Proteini su biološke polimerne makromolekule koje se sastoje od aminokiselinskih podjedinica. Oni igraju ključnu ulogu u normalnom funkcioniranju živih organizama. Najvažnije značajke proteina su regulacija metaboličkih procesa, transport materijala te strukturna uloga. Širok raspon funkcija proteina u biološkim svijetu otvara potencijal za njihovu primjenu u tehnologiji. Danas, proteini se koriste u deterdžentima, obradi hrane, proizvodnji biogoriva, farmaceutskoj industriji, bionanotehnologiji i raznim drugim područjima. S otkrićem kako proteini nastaju i boljim uvidom u fizikalno – kemijske zakonitosti koje definiraju njihovo sklapanje, razvijene su metode za dizajniranje proteina sa željenom funkcijom i strukturom.

Metode dizajna, redizajna i proteinskog inženjerstva se mogu svrstati u tri glavne kategorije: usmjerena evolucija, racionalni i poluracionalni dizajn. Prva od navedenih metoda se temelji na principu evolucije, gdje se mutagenezom generiraju proteini mutanti čija se svojstva karakteriziraju i zatim se probiru najpoželjniji proteini.^{1,2} Za razliku od usmjerene evolucije, racionalni dizajn obuhvaća bioinformatičke i računalne tehnike za modifikaciju postojećih proteina te *de novo* dizajn.^{3,4} Zadnje navedena metoda kombinira usmjerenu evoluciju i racionalni dizajn kako bi iskoristila prednosti svake.⁵

U svojim počecima racionalni dizajn ovisio je o poluempirijskim modelima dobivenim iz eksperimentalnih opažanja i osnovnim poznavanjima iz biofizike proteina. Sa sve detaljnijim razumijevanjem fundamentalnih fizičkih i kemijskih zakonitosti sklapanja proteina i povećanjem računalne moći, postalo je moguće dizajnirati proteine *de novo*. Na taj se način mogu dobiti strukture proteina koje prethodno nisu pronađene u prirodi, s raznim mogućim funkcijama. Ako uzmemo u obzir da tipični protein ima oko 200 aminokiselina, to znači da postoji 20^{200} kombinacija jedinstvenih sekvenci koje mogu dati različite proteine. Evolucija je samo istražila prostor ovih sekvenci koji su od biološkog značaja, što je broj proteina reda veličine 10^{12} . Ostatak prostora je neistražen zbog čega su *de novo* metode sinteze proteina primamljive.⁶

Proteini koji kataliziraju kemijske reakcije u organizmu se nazivaju još i enzimima, oni mogu povećati reakcijske brzine do 17 redova veličine. Kod primjene enzima u industriji postoje ograničenja na kemijske reakcije i uvjete u kojima se mogu katalizirati. Prirodni enzimi su evolucijski optimizirani za katalizu bioloških reakcija u biološkim sustavima i zbog toga ne

mogu katalizirati reakcije koje nisu srodne onima u organizmu. *De novo* dizajn rješava ove probleme otvarajući mogućnost dizajniranja enzima za reakcije za koje ne postoje prirodni enzim, da pri tome da zadrže svoju katalitičku aktivnost u industrijski povoljnim uvjetima. Jedan primjer takvog *de novo* sintetiziranog enzima je Kempova eliminaza, ona je specifično dizajniran da katalizira Kempova eliminaciju. Ovaj primjer će se koristiti za detaljniji pogled u *de novo* način sinteze proteina u nastavku ovog rada.⁵⁻⁷

Postoji razlika u metodama proteinskog inženjerstva i metodama dizajna proteina. Proteinsko inženjerstvo ima svrhu redizajnirati ili modificirati strukturu prirodnih proteina kako bi se bolje razumjele njihove funkcije, izmijenile njihove funkcije ili prenamijenila njihova prirodna struktura za ugradnju novih funkcija, dok u proteinskom dizajnu je generalno prihvaćeno da se početna točka ili početni okvir ne zada, nego se definira topologija proteina kojeg se želi izgraditi i onda se pronade sekvenca ili sekvence koje će stabilizirati željeni sklop.³

Unatoč tome, u daljnjem razmatranju računalnih metoda *de novo* dizajna, kada se bude pričalo o metodama dizajna proteina, pod tim će se podrazumijevati i metode proteinskog redizajna i inženjerstva zato što se te metode koriste suplementarno dizajnu ili su igrale ključnu ulogu u razvoju metoda dizajna proteina.

1.1. Sklapanje proteina

Aminokiselinski slijed koji čini polipeptidni lanac predstavlja primarnu strukturu proteina. Tijekom biosinteze polipeptida na ribosomu često dolazi do kotranslacijskog sklapanja proteina u svoju nativnu strukturu. Termodinamička hipoteza opisuje nativnu strukturu u normalnim fiziološkim uvjetima kao konfiguraciju u kojoj je Gibbsova slobodna energija cijelog sustava najniža, odnosno nativna konformacija predstavlja globalni minimum Gibbsove slobodne energije. Ova hipoteza je primarno potkrijepljena eksperimentima denaturacije i renaturacije, te eksperimentima kemijske sinteze proteina. Kako su proteini dinamičke molekule uslijed konstantnih promjena u okolini i interakciji s drugim spojevima, njihova nativna struktura se zapravo sastoji od ansambla sličnih konformacija koje prelaze jedna u drugu zbog dinamičnosti polja Gibbsove slobodne energije.⁸⁻¹⁰

Glavna sila sklapanja proteina su hidrofobni ogranci koji se ukopaju u jezgru proteina kako ne bi bile izložene polarnom otapalu, minimizira se volumen koji protein zauzima u vodi, a maksimiziraju se povoljne van der Waalove interakcije koje pogoduju gustom pakiranju dok se izbjegavaju steričke smetnje. Polarne grupe se solvativiraju u vodi i zato one koje dopijaju u unutrašnjost hidrofobne jezgre imaju tendenciju tvoriti unutar proteinske vodikove veze kako

bi se kompenzirala energetska zahtjevnost uklanjanja vode iz solvatacijske sfere tijekom prelaska nesklopljene forme u nativnu.⁶

Različite razine proteinske strukture su stabilizirane različitim interakcijama. Primjerice, nastanak α -zavojnica i β -ploča je primarno uvjetovan vodikovim vezama između karbonilnih kisika i amidnih vodika polipeptidne okosnice. Tu, naravno, isto imaju doprinos i interakcije bočnih ogranaka, steričke smetnje i torzijski efekti koji imaju tendenciju prema određenim elementima sekundarne strukture. Interakcije sekundarnih struktura koje dovode da se hidrofobni dijelovi zakopaju u jezgru proteina, a hidrofilni da budu izloženi otapalu odgovorne su za stvaranje tercijarne strukture.⁶

Bitno je naglasiti da sklapanje proteina nije nužno hijerarhijski proces u kojem iz primarne strukture nastaje sekundarna, a iz sekundarne tercijarna. Sekundarne strukture ne moraju nužno biti stabilne u otopini te postoje segmenti aminokiselinske sekvence koji se mogu sklopiti u α -zavojnici ili β -ploču ovisno o nastaloj tercijarnoj strukturi proteina. Četiri klasična mehanizma su: (1) okvirni model (engl. *Framework model*) u kojem nastaju elementi sekundarne strukture koji zatim uslijed međusobnih interakcija stvaraju tercijarnu strukturu; (2) model hidrofobnog urušavanja (engl. *Hydrophobic collapse model*) koji implicira da se protein prvo naglo uruši oko svojih hidrofobnih bočnih ogranaka i onda se presloži u nativnu tercijarnu strukturu; (3) model nukleacijskog propagiranja (engl. *nucleation propagation model*) koji kaže da lokalne interakcije stvaraju male nativne sekundarne strukture koje služe kao jezgra za daljnje stvaranje nativne strukture; (4) model nukleacijske kondenzacije (engl. *nucleation condensation model*) koji predlaže da prisutnost metastabilnih nukleusa neće potaknuti sklapanje sve dok se ne akumulira dovoljan broj dalekometnih stabilizirajućih interakcija.¹¹

1.1.1. Problem sklapanja proteina

Problem sklapanja proteina se može interpretirati na dva načina¹⁰:

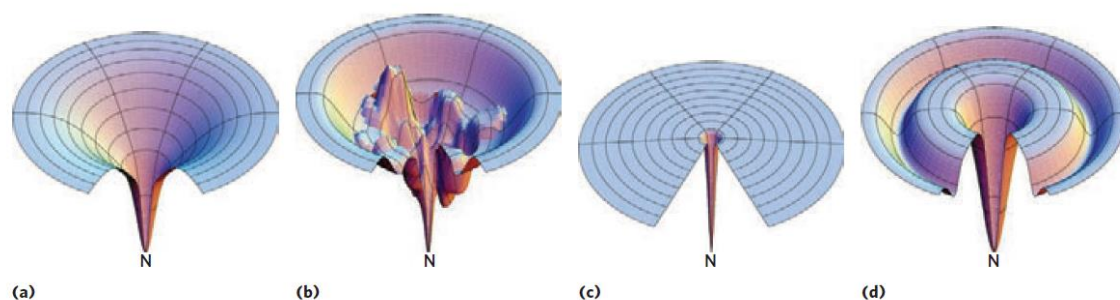
(a) Problem prvog reda: kako protein odabire u minutama svoju jedinstvenu strukturu unatoč astronomsko velikom broju mogućih konformacija.

(b) Problem drugog reda: u koju strukturu će se određena aminokiselinska sekvenca sklopiti

Drugi problem se smatra danas riješenim bioinformatikom uz pomoć neuronskih mreža i umjetne inteligencije. Za uspjeh programa predviđanja strukture iz sekvence je zaslužna ogromna količina baza podataka proteinskih struktura i sekvence koje se prikupljaju nekoliko

desetljeća. Ostaje još samo pitanje kako protein u kratkom vremenskom roku odabere najnižu energetska konformaciju ako mora proći kroz ogromni broj konformacija.¹⁰

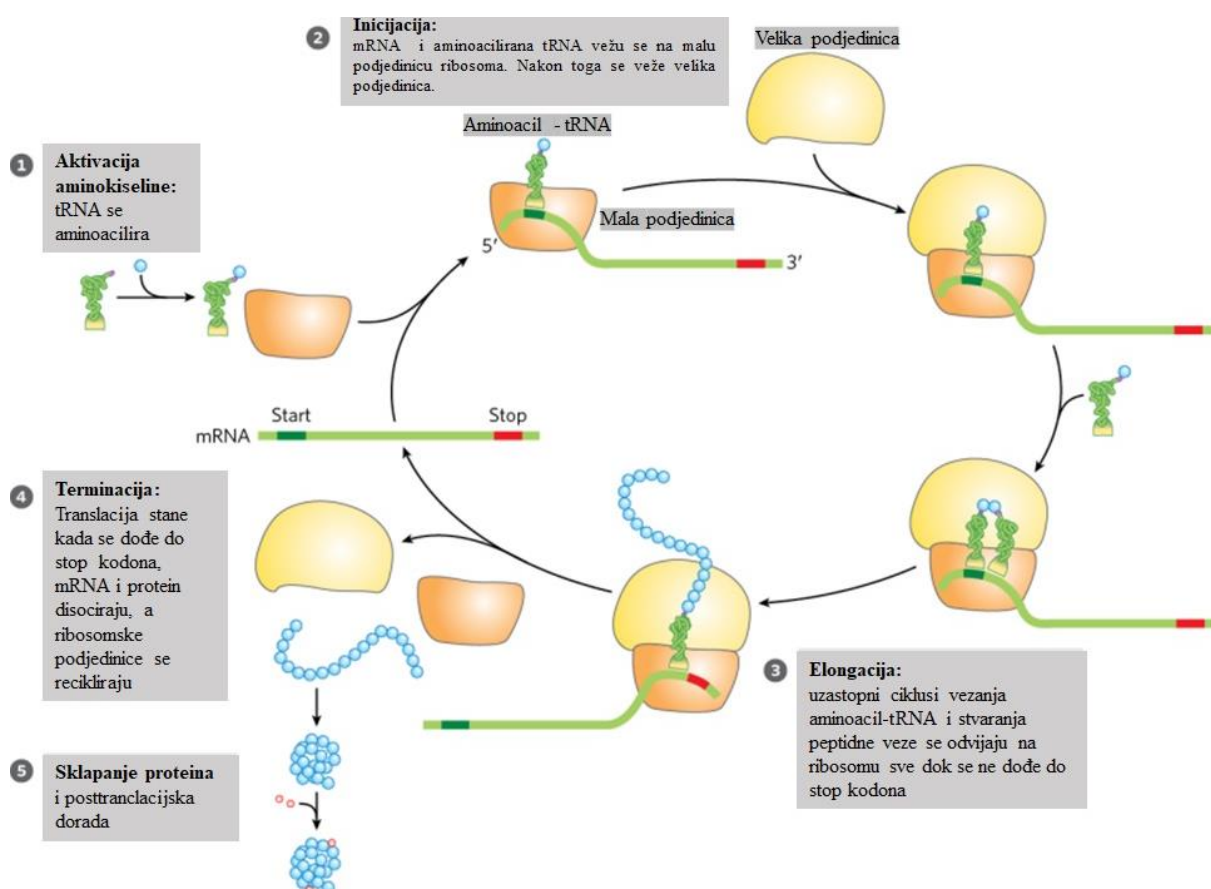
Za proteinski lanac sastavljen od 100 aminokiselinskih ostataka broj konformacija je najmanje 2^{100} , a prelazak iz jedne konformacije u drugu ne može biti brži od pikosekunde (vrijeme jedne termalne vibracije). To bi značilo da bi bilo potrebno oko 10^{10} godina da protein prođe kroz sve konformacije i sklopi se u najstabilniju strukturu, no eksperimentalno znamo da protein pronade svoju nativnu strukturu u minutama (Levinthalov paradoks).^{8,10}



Slika 1. Termodinamika sklapanja proteina prikazana kao lijevci slobodne energije. (a) Jednostavni i glatki lijevak znači da postoji više puteva sklapanja, a da ne postoje međustanja sa znatnom stabilnošću. (b) Tipični lijevak koji opisuje više stabilnijih međustanja na više puteva sklapanja koji vode do nativne strukture proteina. (c) Lijevak proteina sa samo jednom nativnom strukturom, bez međustanja i samo nekoliko puteva sklapanja. (d) Protein s međustanjima znatne stabilnosti na svim putevima sklapanja koji vode do nativnog stanja. Preuzeto i prilagođeno prema: Nelson, D. L., Cox, M. M., *Lehninger Principles of biochemistry*, W.H. Freeman, New York, 2013, str. 146.

Popularni način prikazivanja kako se protein sklapa je termodinamski koncept lijevaka slobodne energije. Lijevak predstavlja multidimenzionalni konformacijski prostor kroz koji protein prolazi tijekom sklapanja. Gornji dio lijevka predstavlja rasklopljeno stanje s visokom konformacijskom entropijom i slobodnom energijom. Kako se protein krene sklapati, lijevak se sužava, smanjuje se veličina konformacijskog prostora i na kraju dolazi se do nativne strukture na dnu lijevka. Lijeenci mogu imati različite oblike i veličine ovisno o kompleksnosti puta sklapanja i stabilnosti međustanja kao što je prikazano na slici 1.¹²

1.2. Sinteza proteina



Slika 2. Shematski pregled pet stadija sinteze proteina preuzeto i prilagođeno prema: Nelson, D. L., Cox, M. M., *Lehninger Principles of biochemistry*, W.H. Freeman, New York, 2013, str. 1114.

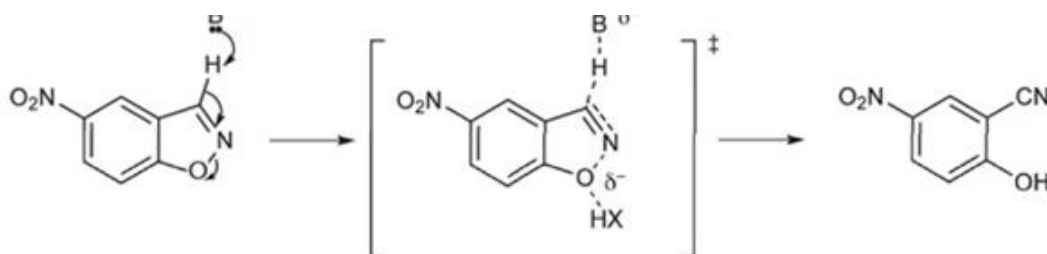
Aminokiselinska sekvenca nekog proteina kodirana je u genima u obliku nukleotida deoksiribonukleinske kiseline (DNA). Biosinteza proteina može se ugrubo razdijeliti na transkripciju i translaciju. Tijekom transkripcije protein polimeraze ribonukleinske kiseline (RNA polimeraza) prevode DNA u predglasničku ribonukleinsku kiselinu (pre-mRNA). Nadalje, pre-mRNA se posttranskripcijski modificira kako bi nastala glasnička RNA (mRNA) koja prenosi informaciju do ribosoma u citoplazmi gdje započinje izgradnja proteina translacijom, proces je prikazan na slici 2. U translaciji, mRNA se veže na ribosom gdje se očitavaju kodoni i uparuju sa odgovarajućim antikodonima. Antikodoni se nalaze na prijenosnoj RNA (tRNA) koja na sebi ima vezanu aminokiselinu za koju kodira kodon, aminoacilirana je, uzastopnim uparivanjem odgovarajućeg antikodona tRNA s kodonom na mRNA dolazi do povezivanja različitih aminokiselina peptidnom vezom. Translacija se

nastavlja sve dok se ne dođe do stop kodona na mRNA lancu čime dolazi do prekida translacije i završetka sinteze proteina.¹²

Pomoću postojeće tehnologije moguće je sintetizirati polipeptide kemijskim putem. Koriste se metode sinteze peptida na krutoj fazi, potpomognute mikrovalovima te metodom native kemijske ligacije, no navedeni načini ograničeni su na polipeptide do oko 300 aminokiselina.¹³

Za razliku od kemijske sinteze proteina, kemijska sinteza DNA je puno razvijenija te je moguće sintetizirati cijele gene ili kromosome *de novo*, a tehnologija rekombinantne DNA omogućava da se sintetička DNA ugradi u biološke sustave. U kontekstu *de novo* dizajna proteina, razvoj metoda sinteze DNA omogućio je eksperimentalno testiranje velikog broja računalno osmišljenih proteina te unaprjeđenje postojećih metodologija dizajna proteina. Poznavanjem načina na koji DNA kodira za određeni polipeptid, slijedi da se dizajnirani protein može kodirati u sintetičku DNA koja ugradnjom u model organizam dovodi do proizvodnje željenog proteina. Jedan od najpoznatijih i najkorištenijih model organizama za ovakvu tehnologiju je *Escherichia coli*, gdje se u plazmidne vektore ugrađuje sintetička DNA, rekombinantni plazmid se zatim transformacijom uvodi u bakteriju te u konačnici eksprimira u željeni protein.⁶

1.3. Reakcija Kempove eliminacije



Slika 3. Mehanizam Kempove eliminacije. Preuzeto i prilagođeno prema: D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChanice, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Nature 453 (2008) 190–195.

Kempova eliminacija^{14,15} je dobro proučena lužnato katalizirana ireverzibilna reakcija uklanjanja protona s benzizoksazolnog prstena dajući 2-hidroksibenzonitrilni produkt. Reakcija prati E2 mehanizam, pri čemu baza deprotonira ugljik na položaju C-3. Nastali 2-cijanofenolatni anion stabiliziran je delokalizacijom negativnog naboja na kisiku preko benzenskog prstena. Iako su za ovu reakciju opažena katalitička antitijela¹⁶, ne postoji niti jedan prirodni enzim koji ju katalizira.⁷

§ 2. METODE U DIZAJNU PROTEINA

Novi proteini se mogu dizajnirati pomoću dva opća načina, usmjerenom evolucijom ili racionalnim dizajnom. Usmjerenom evolucijom je *in vitro* ili *in vivo* metoda koja se temelji na proizvodnji mutacija u organizmu ili stanici kako bi se dobio protein sa željenim karakteristikama. S druge strane, racionalni dizajn se primarno temelji na *in silico* metodi dizajna proteina, gdje se koriste računalne metode za generiranje proteina sa željenim funkcijama i strukturama. Obje metode imaju svoje prednosti i nedostatke te su zbog toga spregnute u metodama poluracionalnog dizajna kako bi se kombinirale prednosti obje tehnika.

2.1. Usmjerenom evolucijom

Usmjerenom evolucijom je metoda proteinskog inženjerstva koja koristi principe evolucije u modeliranju proteina željene stabilnosti ili funkcionalnosti. Temelj ove metode uvođenje je malog broja nasumičnih mutacija u gen koji kodira za protein od interesa. Proteini mutanti, koji nastanu nakon ekspresije mutiranih gena, identificiraju se i sastavlja se biblioteka mutanata. Kroz iterativni postupak odabiru se skupine mutanata koje imaju željena svojstva te ponovno podliježu mutagenezi. Akumulacija željenih mutacija evoluira protein u smjeru željenih svojstava.¹

Prvi korak u usmjerenom evolucijom dobivanje je raznolikih mutanata i izgradnja biblioteke mutanata. Ovaj proces mutageneze može biti nasumičan ili usmjeren. Kod nasumične mutageneze, odabrani gen podliježe mutacijama pomoću lančane reakcije polimeraze sklone greškama (engl. *Error-Prone Polymerase Chain Reaction*, EP-PCR) ili metode rekombinantne DNA kako bi se dobili raznovrsni mutanti. Usmjerenom mutagenezom koriste se sintetičke početnice (engl. *primers*) sa željenom mutacijom kako bi se izazvala mutacija na određenom dijelu gena, na ovaj način se dobivaju mutanti veće specifičnosti. Nakon generiranja biblioteke mutanata, geni mutanata se ukloniraju u vektor i transformiraju u stanice. Nakon što stanice eksprimiraju raznovrsne proteine traže se varijante sa poboljšanim svojstvima. Kada se identificiraju proteini čija svojstva su se poboljšala, geni mutanti se izoliraju te podliježu sljedećem ciklusu mutageneze.²

Novija metoda automatizirane usmjerene evolucije koristi modificirane M13 bakteriofage za kontinuiranu infekciju *E. coli* čime je omogućena kontinuirana evolucija proteina. Iz

bakteriofaga se ukloni gen III (gIII) koji kodira za protein III (pIII), neophodan za njegovo razmnožavanje, te se na njegovo mjesto umetne gen za protein koji želimo mutirati. Nadalje, bakteriofagom se inficiraju stanice posebno dizajnirane *E. coli*, koja ima plazmid s gIII, a čija je ekspresija vezana za aktivnost proteina kojeg mutiramo i ima plazmid s domenom koja kodira za polimerazu podložnu proizvodnji grešaka te čija ekspresija je inducirana prisustvom arabinoze. U ovakvom sustavu, inficirana *E. coli* će proizvoditi željeni protein koji će, ako pokaže ciljane aktivnosti, inducirati gIII na plazmidu da se proizvede pIII i ti bakteriofagi će se moći dalje razmnožavati. Zatim, dodatkom ili uklanjanjem arabinoze regulira se stopa mutacije te dolazi do mutiranja gena koji kodira za ciljani protein. Mutirani proteini čija aktivnost inducira stvaranje pIII će omogućiti daljnju evoluciju tog mutanta kroz stvaranje nove generacije infektivnih bakteriofaga s genom mutantom, dok oni koji ne induciraju pIII će stvarati bakteriofage koji se dalje neće moći razmnožavati i time se prekida prijenos gena proteina koji nema željenu aktivnost u sljedeće generacije.¹⁷

Svaki od navedenih pristupa usmjerenoj evoluciji ima svoje prednosti i nedostatke s obzirom na kombinacije sekvenci koje se proizvedu, no ograničavajući faktor usmjerene evolucije u usporedbi s računalnim metodama je srodnost novih proteina dobivenih navedenom metodom početnim proteinima. Teško je ili gotovo nemoguće dobiti proteine novih funkcija i struktura usmjerenom evolucijom. Dodatno, usmjerena evolucija eksperimentalna je metoda koja zahtjeva puno vremena i korištenje skupe opreme i reagensa. Unatoč tome, ova metoda pokazala se korisnom u inženjerstvu proteina jer se mogu optimizirati postojeći proteini kako bi im bila unaprijeđena stabilnost, specifičnost i reaktivnost.¹

2.2. Racionalni dizajn

Racionalni dizajn proteina je pristup čiji cilj je dobiti protein s željenom strukturom i/ili funkcijom pomoću dobrog razumijevanja odnosa između strukture i aminokiselinske sekvence proteina. Koriste se računalne metode, tehnike strukturne biologije te detaljno razumijevanje fizikalno-kemijskih principa koji uvjetuju sklapanje proteina. Danas, racionalni dizajn obuhvaća mnoštvo metoda i pristupa koji se znatno razlikuju, ali sa zajedničkim ciljem stvaranja novih proteina kroz razumijevanje odnosa strukture i sekvence. Zato najbolji pristup za prikazivanje širine ovih metoda je razmotriti korake koji su kroz povijest doprinijeli mogućnosti da se danas dizajniraju proteini u potpunosti *de novo* računalnim metodama.^{3,18}

Nativna struktura proteina je ona u kojoj je konformacijska slobodna energija lanca minimizirana, zato je potrebno dizajnirati proteinsku sekvencu koja za zadanu strukturu ima

minimum slobodne energije. Predviđanje sklapanja određene aminokiselinske sekvence u nativnu strukturu proteina je ono što smatramo problemom sklapanja proteina. Razmatranjem ovog problema došlo je do pitanja je li moguće, ako znamo neku strukturu, osmisлити sekvencu koja će se sklopiti u nju.⁴

Počeci rješavanja ovog problema, koji se često naziva i inverzni problem sklapanja proteina, su bili pristupi temeljeni na kvalitativnim pravilima i fizikalnim modelima, koji su demonstrirali mogućnost stvaranja *de novo* sekvence koja se sklopi u određenu strukturu. Prve strategije dizajna su bile progresivne, tako što se iterativno povećavala kompleksnost dizajniranih sekvenci. To je doprinijelo fundamentalnom razumijevanju strukture i funkcije proteina. Otkriveno je da specifični motivi hidrofилnih (h) i hidrofobnih (p) aminokiselinskih ostataka omogućuju, uslijed hidrofobnog efekta, stvaranje sekundarnih struktura kao što su α -zavojnice (npr. motiv hpphpp...) i β -ploče (npr. motiv hphp...). Došlo je do pojave koncepta pozitivnog dizajna, uvođenje bočnih ogranaka koji stabiliziraju nativno stanje, i negativnog dizajna, uvođenje ogranaka koji destabiliziraju ne-nativno stanje.^{3,4,18}

Sljedeći korak je došao sa razvojem tehnologije sinteze dužih peptida i sintetičkih gena, razvojem metoda molekularne mehanike i dinamike te boljim razumijevanjem termodinamike i kinetike sklapanja proteina. Generalno je postalo prihvaćeno da je pakiranje hidrofobnih ogranaka aminokiselina u unutrašnjost proteinske jezgre glavna pokretačka sila sklapanja proteina u vodenoj otopini, a da polarne interakcije pomažu u definiranju detaljne geometrije proteinske strukture.¹⁸ Sistematskim eksperimentima i analizama strukture prirodnih proteina, otkriveni su detaljniji odnosi između strukture i sekvence proteina, određene aminokiseline imaju tendenciju prema specifičnim sekundarnim strukturama i rotamerima. Rezultat toga je razvoj modela u kojem su bočni ogranci „pakirani“ u proteinske jezgre kao 3D slagalice. Energetski izrazi za rangiranje poželjnosti aminokiseline sekvence za zadanu okosnicu su optimizirani tako da sadrže samo najvažnije interakcije, no broj sekvenci koje je potrebno razmotriti za pronalaženje one s minimalnom energijom i dalje je predstavljalo problem. Povećanjem računalne snage i primjenom računalnih algoritama kao što su genetički, Monte-Carlo i eliminacija mrtvih krajeva (engl. *dead-end elimination*) je omogućeno brzo pretraživanje najboljih kombinacija bočnih ogranaka i njihovih rotamera za zadanu strukturu. Ovim napretkom je dizajniran protein samo iz zadane okosnice, prvi pravi *de novo* protein.^{3,18,19}

Unatoč takvom napretku i dalje je problem bio dizajnirati veće proteine koji se kooperativno slažu te proteine koji imaju strukturne elemente koji nikada nisu viđeni u prirodi. Posljednji

korak u oblikovanju dizajna proteina bio je razvoj računalnih metoda temeljenih na sklapanju fragmenata i proteinske bioinformatike.^{3,18}

2.3. Poluracionalni dizajn

Racionalni dizajn zahtjeva dobro razumijevanje strukturnih značajki enzimskog aktivnog mjesta i njihovog doprinosa njegovoj funkciji. Kompleksnost između strukture i funkcije u enzimima pokazalo se ograničavajućim faktorom u primjeni racionalnog dizajna. S druge strane, usmjerena evolucija omogućuje relativno brzo inženjerstvo enzima bez potrebe dubokog razumijevanja odnosa strukture i funkcije, no postoji problem da nasumične mutacije stvaraju velik broj mutanata čija svojstva se moraju ispitati i karakterizirati. Uvođenje nasumičnih mutacija kada informacije o strukturne i funkciji postoje nije najpogodnije, bolji pristup bi bio uzorkovanje samo najefektivnijih mutacija. Analize mutiranih enzima pokazuju da su najkorisnije mutacije koje utječu na enzimska svojstva one u aktivnom mjestu ili u njegovoj blizini. Kombinacijom prednosti racionalnog dizajna i usmjerene evolucije mogu se upotpuniti njihovi nedostaci, tada pričamo o poluracionalnom dizajnu.⁵

Ukoliko su dostupne strukturne informacije proteina, moguće je direktno podvrgnuti mutagenazi aminokiseline koje su u kontaktu sa supstratom ili su blizu šupljine aktivnog mjesta. U većini slučajeva izbor pozicija koje će se podvrgnuti mutagenazi ostaje racionalan, ali i dalje postoji veliki izbor aminokiselina koje se trebaju testirati. Isto tako, nije najpogodnije mutirati samo individualne pozicije jer se pokazalo da istovremeno mutiranje više ciljanih pozicija ili kombinacija ciljanih i nasumičnih mutacije može imati sinergijski učinak, koji se ne postiže mutiranjem samo individualnih pozicija.⁵

U nekim slučajevima strukturne informacije nisu dostupne ili dostatne za racionalni izbor aminokiselinskih ostataka koje će se podvrgnuti nasumičnom mutiranju, tada se molekularno modeliranje pokazalo uspješno u predviđanju rezidua koje će biti u kontaktu s molekulama supstrata.⁵

Još jedan pristup je istovremena nasumična mutagenaza i mutagenaza zasićenja mjesta. Ovim pristupom se mogu povećati šanse dobivanja korisnih mutacija na nepredvidljivim lokacijama u kombinaciji sa mutacijama na pažljivo odabranim mjestima. Na ovaj način dobiva se veći broj poželjnih mutanata u usporedbi s mutagenozom cijelog gena.⁵

U slučaju nedostupnosti strukturnih i funkcijskih informacija te nemogućnosti predviđanja aminokiselinskih ostataka koji se trebaju podvrgnuti mutaciji, uspješno je primijenjena nasumična mutagenaza cijelog gena popraćena mutagenozom zasićenja mjesta. Višestrukim

mutiranjem cijelog gena identificiraju se potencijalni aminokiselinski ostaci koje proizvode povoljne promjene. Zatim se ti aminokiselinski ostaci podvrgnu mutagenezi zasićenja mjesta kako bi se optimizirala ciljana svojstva i promjene proteina.⁵

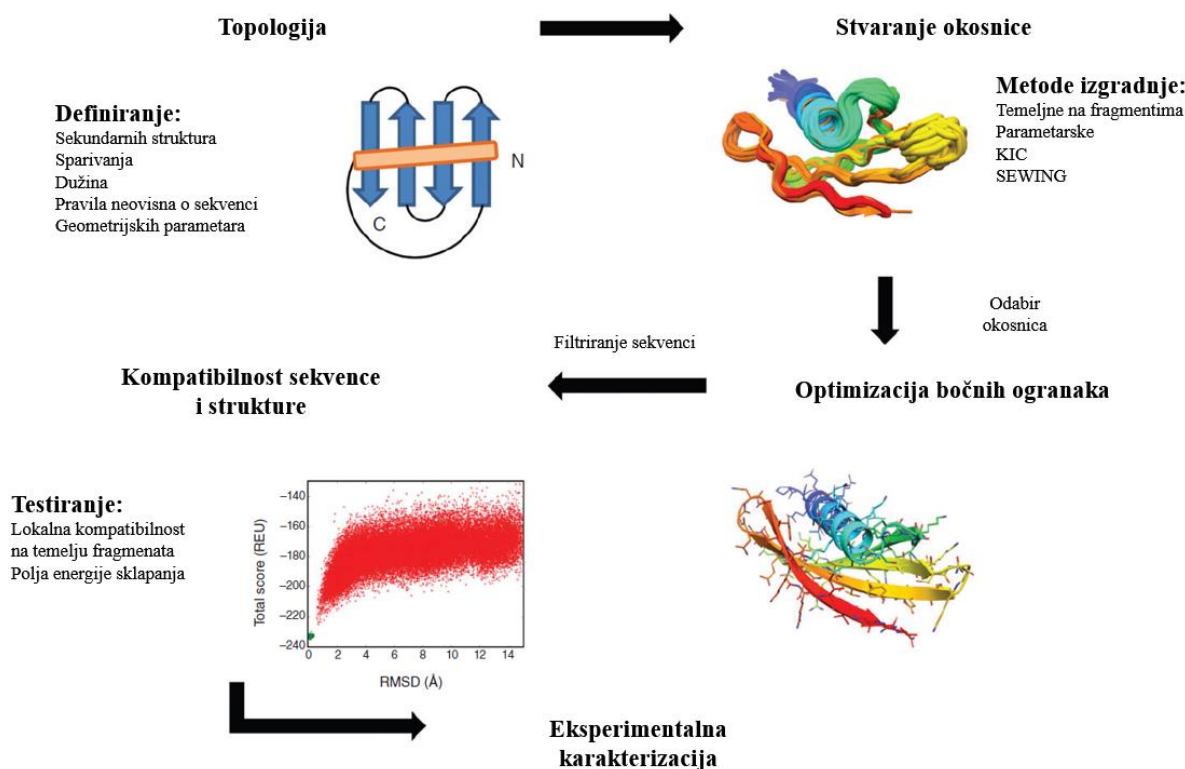
Iako je u prethodno navedenim pristupima ograničen broj generiranih mutanata tako što se usredotočilo na stvaranje varijanata koje će imati poželjne promjene, i dalje je to veliki broj proteina koji se moraju ispitati. Računalnim metodama temeljenim na algoritmima racionalnog proteinskog dizajna moguće je probrati poželjne sekvence mutanata iz virtualne biblioteke ili dizajnirati aktivno mjesto enzima u postojećim proteinima. Prvim načinom mogu se računalno eliminirati mutacije koje nisu u skladu s proteinskim sklopom, koristeći virtualnu biblioteku mogućih mutanata, računalnim metodama se mogu odabrati mutanti koji teoretski proizvode najpoželjnije promjene, takav set mutanata se onda podvrgne ciljanoj mutagenezi i eksperimentalno karakterizira. Drugim pristupom može se izdvojiti katalitičko aktivno mjesto iz postojećeg enzima, a zatim se računalnim metodama odaberu mjesta na nekom drugom proteinu koji nema katalitička svojstva, čijom mutacijom će se dobiti željeno aktivno mjesto, time se nekatalitički protein pretvara u enzim.⁵

§ 3. METODE DE NOVO SINTEZE PROTEINA

Znatni napreci u *ab initio* metodama i metodama predviđanja proteinske strukture prema predlošku, kombinirano sa znatnim povećanjem u računalnoj moći i broju javno dostupnih baza podataka proteinskih sekvenci i strukturnih informacija, dozvoljavaju točno predviđanje proteinske strukture različitog raspona veličine i kompleksnosti. Predviđanje proteinske strukture razvilo se do razine efikasnog rješavanja problema inverznog sklapanja proteina i pronalaženja aminokiselinske sekvence koja se sklapa u željenu proteinsku strukturu.²⁰

Tradicionalno, proteini se dizajniraju pomoću jednostavnijih metoda racionalnog dizajna i usmjerene evolucije. Takav pristup koristio se u rješavanju mnogih biotehnoških problema, kao što je dizajniranje novih katalizatora, terapeutika, biosenzora i samosklapajućih nanomaterijala. Ove metode su se oslanjale na podatke o strukturi i funkciji prirodnih proteina što je ograničilo razvoj proteina s karakteristikama i svojstvima koja nisu nikad bila opažena u prirodi. Proteini su aktivni u specifičnim staničnim uvjetima i temperaturnim rasponima, u puno slučajeva reagiraju na alosteričke signale, a neki se mogu sklopiti samo pomoću šaperona i u konačnici svi proteini se trebaju razgraditi za održavanje stanične homeostaze. Navedena biološka ograničenja su izuzetno nepovoljna u redizajnu i inženjerstvu proteina, često se javljaju problemi u stabilnosti kada se uvode mutacije, dolazi do niske ekspresije rekombinantnih proteina i drugih neočekivanih pojava. Također, unatoč rastućem broju poznatih trodimenzionalnih proteinskih struktura dostupnih u bazi *Protein Data Bank* (PDB), pronalazak strukture optimalne geometrije i stabilnosti za kreiranje željene funkcije nije uvijek moguć.²⁰

De novo računalni dizajn proteina suočava se s ovim ograničenjima te omogućava dizajniranje neograničenog broja proteina sa preciznom geometrijom, željenom funkcijom i optimiziranom stabilnošću. Trenutno jedan od najopsežnijih softverskih alata za *de novo* dizajn proteina je softverski paket *Rosetta*²¹⁻²³, koji obuhvaća metodologiju opisanu u nastavku teksta. Kao nastavak na metode racionalnog dizajna razmotriti će se metode koje se koriste danas u racionalnom dizajnu proteina.²⁰



Slika 4. Tijek računalnog *de novo* dizajna proteina. Preuzeto i prilagođeno prema: E. Marcos, D. Silva, *WIREs Computational Molecular Science* **8**, (2018).

Općenito, proces *de novo* dizajna proteina odvija se u četiri koraka²⁰ koja su prikazana na slici 4:

- Definiranje ciljne strukture ili topologije
- Generiranje prikladnih proteinskih okosnica
- Dizajniranje podudarajućih aminokiselinskih sekvenci
- Procjena kompatibilnosti dizajnirane sekvence i ciljne strukture.

3.1. Stvaranje okosnice

De novo dizajn proteina započinje sa definiranjem strukture ili topologijom ciljnog proteina, a zatim se generiraju kompatibilne proteinske okosnice. Dizajn se događa u dva koraka, prvo se izgrađuju okosnice, zatim se odabiru ili filtriraju samo one koje će najbolje odgovarati zadanom cilju. Svrha je dobiti okosnice koja zadovoljavaju niz specifikacija koje je dizajner unaprijed definirao i prebrati samo one koje će dalje u toku dizajna pogodovati stvaranju sekvenci sa niskoenergetskim interakcijama bočnih ogranaka.²⁰

Mnoge metode izgradnje okosnice temeljene su na prethodno dobivenim eksperimentalnim podacima, tipično izoliranim iz poznatih trodimenzionalnih proteinskih struktura. Stoga,

metode izgradnje proteinske okosnice moguće je klasificirati prema stupnju korištenja prijašnjeg znanja. Na jednu ruku, postoje *ab initio* metode koje koriste vrlo malo ili uopće ne koriste eksperimentalne podatke, primjer je metoda kinematičkog zatvaranja (engl. *Kinematic closure*, KIC). Na drugu ruku, postoje metode koje se temelje isključivo na prethodnom znanju o strukturama, kao što je metoda produljivanja strukture pomoću grafova nativnih podstruktura (engl. *Structure extension with native-substructure graphs*, SEWING).²⁰

Nadalje, postoji podjela metode prema stvaranju kontinuiranih ili diskontinuiranih okosnice. Primjerice, metoda sklapanja fragmenata (engl. *Fragment assembly*) proizvede kontinuirane okosnice sastavljene od elemenata sekundarne strukturne spojene petljama. Dok parametarske metode (engl. *Parametric design*) daju okosnice sekundarnih strukturnih elemenata koje nisu međusobno povezane petljom, ona se dodaje u dodatnom koraku.²⁰

3.1.1. Sklapanje fragmenata

Sklapanje fragmenata kombinatorna je metoda temeljena na postojećim strukturnim podacima. Okosnica se dizajnira tako što se koriste fragmenti iz baza podataka proteinskih struktura koji se mogu primijeniti u dizajnu bilo kojeg proteinskog sklopa. Iako je pogodna za dizajn proteina različitih razina kompleksnosti, porast broja elemenata u okosnici eksponencijalno povećava kompleksnost metode.²⁰

U primjeni metode sklapanja fragmenata, prvi korak je definirati željenu topologiju proteina u obliku nacрта u kojem se navodi redoslijed, dužina i način spajanja sekundarnih struktura. Zatim se koristi softverski paket, kao što je *Rosetta* koja pomoću baze podataka riješenih proteinskih struktura odabire proteinske fragmente ovisno o kompatibilnosti sa definiranim nacrtom. Ovi fragmenti se zatim uklapaju u okosnicu korištenjem Monte Carlo algoritma za umetanje fragmenata.^{20,24}

U ovoj fazi protein je definiran u niskoj rezoluciji, koriste se pojednostavljeni oblici bočnih ogranaka u obliku sfere te se koristi pojednostavljena energetska funkcija koja uzima u obzir samo vodikove veze na okosnici, Van der Waalsove interakcije za izbjegavanje steričkih smetnji i ima član koji pogoduje kompaktnoj strukturi. Često se koriste i centriodi koji imitiraju velike bočne ogranke kako bi se ostvario adekvatan razmak između elemenata sekundarne strukture i time omogućila akomodacija velikih bočnih ogranaka tijekom dizajniranja sekvence.

Glavna prednost metode sklapanja fragmenata leži u njenoj mogućnosti brzog stvaranja raznih okosnica željene sekundarne strukture, ali nije učinkovita u stvaranju detaljnijih

strukturnih elemenata. Iz tog razloga, koriste se geometrijska ograničenja i funkcije pristranosti kako bi se usmjerilo uzorkovanje prema strukturama koje sadrže ključne strukturne značajke.²⁰

3.1.2. Parametarski dizajn

U konstrukciji jednostavnijih elemenata sekundarne strukture, kao što su β -bačve, koriste se matematičke jednačbe, poželjno, s malim brojem geometrijskih parametara. Izgradnja okosnica na ovaj način vrlo je brza, no moguće je generiranje ogromnog broja strukturno mogućih okosnica što otežava adekvatan izbor u daljnjem tijeku dizajna. Iz tog razloga postoji više pristupa u odabiru poželjnih okosnica, a izbor ovisi o problemu, odnosno strukturi koja se želi izgraditi. U dizajnu primjerice jednolančanih snopova zavojnica potrebno je zasebno modelirati petlje koje spajaju pojedine zavojnice dobivene parametarskim dizajnom.²⁰

3.1.3. Kinematičko zatvaranje

Cilj je odrediti konfiguracije koje zadovoljavaju inicijalno definirana ograničenja kao što su duljina veze, kutovi veze i torzijski kutovi unutar kovalentnog vezanog lanca. Ako je zadan segment s nekim brojem, N , ϕ/ψ torzijskih stupnjeva slobode i fiksnim N- i C-krajem, KIC uzorkuje $N - 6$ torzijskih kutova prema Ramachandranovim vjerojatnostima i analitički rješava kinematičku jednačbu s preostalim 6 torzijskih stupnjeva slobode kako bi zatvorio lanac bez uzrokovanja promjene u ostatku strukture. Kinematička metoda razvijena za predviđanje strukture petlji i peptida gotovo uopće ne zahtijeva podatke o strukturama postojećih proteina, ali zato sa porastom dužine peptida raste računalna zahtjevnost.²⁰

3.1.4. Produljivanje strukture pomoću grafova nativnih podstruktura

Metoda je temeljena na konceptu genetičke rekombinacije, na različite načine se kombiniraju kompatibilni mali strukturni motivi dobiveni iz postojećih proteinskih struktura, uglavnom iz elemenata sekundarne strukture. Prednost metode je automatsko generiranje struktura koje nalikuju nativnim. Dodatno, za razliku od drugih metoda izrade okosnica, ovdje nije potrebno prethodno definiranje topologije proteina.²⁰

Počinje se stvaranjem grafa u kojem svaki čvor predstavlja mali strukturni motiv ili podstrukturu izdvojenu iz postojećih proteinskih struktura. Zatim se ti strukturni motivi uspoređuju, računa se korijen srednjeg kvadratnog odstupanja (engl. *root-mean-square deviation*, RMSD). Podstrukture se mogu kombinirati ako N-terminalna regija jedne podstrukture je slična C-terminalnoj druge, njihovi čvorovi se povezuju rubovima koji

predstavljaju tu sličnost. Nove proteinske strukture se onda generiraju tako što se ide kroz graf od čvora do čvora preko rubova koji ih povezuju i kombiniraju se slične podstrukture pazeći da ne dolazi do steričkih smetnji.^{20,25}

3.1.5. Konstrukcija petlji

Neke metode ne stvaraju automatski petlje koje povezuju proteinske strukture, pa se koriste razne metode rekonstrukcije petlji. Postoje algoritmi koji brzo pretražuju prirodne petlje iz baza podataka fragmenata koji se javljaju u prirodi, a također se može koristiti prethodno navedena metoda KIC, koja može graditi petlje bez korištenja informacija o nativnim strukturama.²⁰

3.2. Dizajniranje sekvence

Cilj dizajna sekvence je identificiranje aminokiselinskog slijeda, koji minimizira energetske funkcije, za fiksnu proteinsku okosnicu. Za efikasnu identifikaciju takvih sljedova, koriste se algoritmi poput onih u *Rosetta* softverskom paketu. Ovi algoritmi se oslanjaju na razlaganje određenih članova energetske funkcije tako da se interakcije između bliskih parova rotamera aminokiselinskih ostataka mogu računati zasebno, a zatim se suma tih interakcija koristi za procjenu ukupnog energetske stanja proteina. Time je omogućeno da se interakcije između svih mogućih parova rotamera izračunaju unaprijed, što značajno povećava računalnu učinkovitost u dizajniranju sekvenci proteina. Najopćenitiji i široko primijenjen algoritam u takvom pristupu, unutar softverskog paketa *Rosetta* je *Packer*, koji koristi optimizaciju simuliranog žarenja (engl. *simulated annealing*) u istraživanju različitih kombinacija aminokiselina i pronalasku one sa najnižom energijom za fiksiranu okosnicu.²⁰

Kako bi se dobili bolji rezultati iterativno se alternira između optimizacije fiksne okosnice i malih promjena u strukturi okosnice, to je dizajn sekvence fleksibilne okosnice (engl. *flexible-backbone sequence design*). Ovaj pristup je pogotovo koristan u *de novo* dizajnu, gdje se proteinska okosnica generira ni iz čega i potrebno ju je dodatno usavršiti, ovisno o postavljenim bočnim ograncima. Dizajn sekvence za fiksne okosnice nije uvijek pogodan jer se dobije uski set sekvenci koje previše ovise o početnoj strukturi okosnice, što ograničava pronalazak najoptimalnijih niskoenergetskih proteinskih dizajna. *FastDesign* metoda u *Rosetti* koja koristi *FastRelax* minimizacijski algoritam jedan je od najmoćnijih protokola za fleksibilni dizajn. Unutar ciklusa optimizacije bočnih ogranaka i strukture okosnice omogućava se pronalazak bočnih ogranaka koji inicijalno imaju steričke smetnje, ali nakon malih promjena u strukturi

okosnice mogu proizvesti dobro pakiranje čime se dobiju strukture s manje unutrašnjih šupljina i nižom ukupnom energijom.²⁰

Iako je *Packer* snažan algoritam za optimizaciju bočnih ogranaka, znatno je ograničen energetske funkcijama i algoritmima uzorkovanja sekvence, što može uzrokovati pojavu nepoželjnih i neoptimalnih značajki u dizajnu. Kako bi se to izbjeglo koriste se razne strategije koje stvaraju pristranost prema pronalasku optimalnog rješenja. Jedna od najčešćih i najuspješnijih strategija je dizajn pomoću slojeva. Pozicijama aminokiselinskih ostataka se asignira jedan od tri sloja: jezgra, granica ili površina ovisno o okolini u kojoj se nalaze. Zatim se u dizajnu za svaki sloj koriste aminokiseline koje su kompatibilne s tim okruženjem. Ovaj pristup ima dvije prednosti: osigurava da se hidrofobne i polarne aminokiseline smjeste na točne položaje i da se koristi manja skupina mogućih aminokiselina za uzorkovanje svakog položaj što dodatno ubrzava računicu. Strategije koje se koriste za usmjeravanje *Packer* algoritma prema optimalnom rješenju se biraju prema strukturi i svojstvima koje želimo da ciljani protein ima.²⁰

Pravilnim kontroliranjem *Packer* algoritma generiraju se dizajni sa željenim svojstvima, ali identificiranje onih koji imaju veće šanse da budu stabilni i da se sklope u ciljanu strukturu zahtjeva filtriranje i rangiranje pomoću dodatnih kriterija. Neki od najčešćih filtera su²⁰:

1. Ukupna energija: odražava energiju sustava kao težinsku sumu veznih i ne veznih interakcija. Što je ova energija manja, to je veća vjerojatnost da je struktura stabilnija u odnosu na druge konformacije. Na ovu energiju najviše utječe dizajn okosnice i gustoća pakiranja bočnih ogranaka. Strukture najniže energije su generalno najbolji kandidati za dobivanje funkcionalnog dizajna.
2. Pakiranje bočnih ogranaka: odnosi se na gustoću kontakata bočnih ogranaka u proteinskoj jezgri, komplementarnost oblika između sekundarnih struktura i volumen unutrašnjih praznina. Ovo maksimizira specifičnost sekvence za dizajniranu strukturu i površinu zakopanih hidrofobnih aminokiselina koje su ključne za sklapanje proteina. Specifične metrike su razvijene za procjenu kvalitete pakiranja bočnih ogranaka.
3. Kompatibilnost između slojeva i aminokiselinskih tipova: Za stabilnost i topljivost proteina bitno je da su točni tipovi aminokiselina u točnim slojevima. Dodjela bočnih ogranaka u inicijalnom sloju temeljenom samo na strukturi okosnice može proizvesti probleme kao zakopane polarne aminokiseline koje ne tvore vodikove veze ili

hidrofobne aminokiseline koje su izložene na površini. Kako bi se poboljšao dizajn potrebno je napraviti bolje podjele aminokiselina po slojevima.

4. Predviđanje sekundarne strukture: Razmatranjem sličnosti između sekundarne strukture dobivene dizajnom i one koje je predviđena metodama sklapanja zadane aminokiselinske sekvence moguće je eliminirati sekvence koje kodiraju za netočnu strukturu.

3.3. Kompatibilnost između sekvence i strukture

Prethodnim strategijama su kroz pozitivni dizajn dobivene niskoenergetske sekvence optimizirane da stabiliziraju konformaciju zadane okosnice. Do neke mjere, korištene metode su doprinijele i negativnom dizajnu, ali nisu optimizirale sekvencu tako da eksplicitno destabiliziraju druge nepoželjne konformacije. Zato je nemoguće biti siguran da se dizajnirana sekvenca sklapa u ciljanu strukturu. Kako bi se ovaj problem zaobišao, studije *de novo* dizajna dodatno određuju jesu li dizajnirane sekvence kompatibilne s ciljanom strukturom na lokalnoj i na nelokalnoj razini tako da se direktno razmatraju alternativne konformacije. Iako najtemeljitiji test kompatibilnosti sekvence i strukture je *ab initio* test predviđanja strukture, on je često računalno prezahtjevan.²⁰

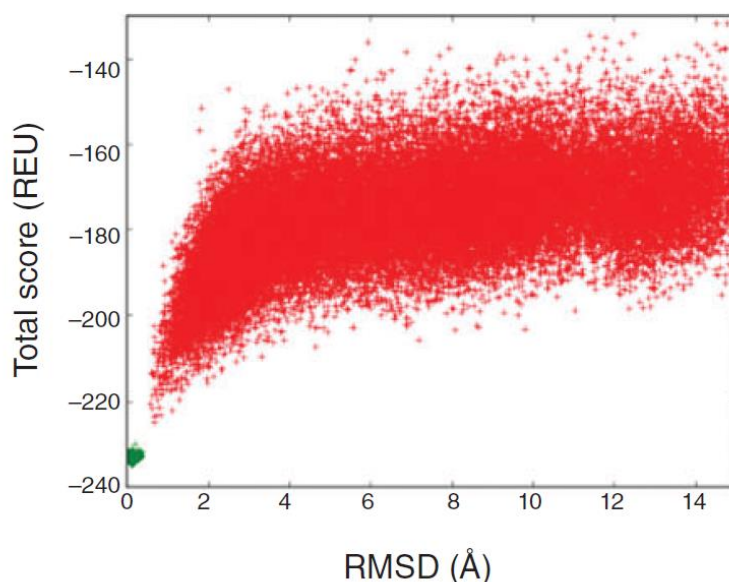
3.3.1. Kompatibilnost strukture i sekvence na lokalnoj razini

Ispitivanja lokalne kompatibilnost dizajna razmatra fragmente sekvence u kojoj su aminokiselinski ostaci poslagani u redosljed. Lokalna kompatibilnost je korisna za filtriranje dizajna prije nego što se krene sa računalno zahtjevnijim nelokalnim testovima kompatibilnosti. Mjerenja lokalne kompatibilnosti variraju u broju uzastopnih aminokiselinskih ostataka koji će se razmatrati i njihovoj definiciji.²⁰

Vrijedi napomenuti da dizajniranje i odabir sekvence već uračunava kompatibilnost na razini pojedinih aminokiselinskih ostataka. Članovi Rosetta energetske funkcije imaju dvije lokalne mjere koje potječu od Ramachandranovih vjerojatnosti: (a) rama (vjerojatnost ϕ/ψ torzijski kutova okosnice ako je zadan tip amino kiseline) i (b) p_aa_p (vjerojatnost vrste aminokiseline ako je zadan ϕ/ψ torzijski kut okosnice).²⁰

Još jedna mjera koja se često koristi u procesu dizajna i koja nije vezana uz Rosetta energetske funkcije je podudaranje sekundarne strukture dizajna i one koja je predviđena računalno iz linearne sekvence koju smo dizajnirali. Mjeri se kao omjer aminokiselinskih ostataka s točnom predviđenom sekundarnom. Postoje još metode u kojima se fragmenti

različitih duljina uspoređuju s ansamblom fragmenata dobivenih iz baza podataka struktura prirodnih proteina.²⁰



Slika 5. Simulacije sklapanja krećući od ravnog lanca. Crvene točke predstavljaju strukture najniže energije dobivene prateći svaku putanju sklapanja. Zelene točke predstavljaju strukture najniže energije dobivene relaksacijom dizajniranih modela. Preuzeto i prilagođeno prema: E. Marcos, D. Silva, *WIREs Computational Molecular Science* **8**, (2018).

3.3.2. Kompatibilnost sekvence i strukture na nelokalnoj razini

Kompatibilnost između sekvence i strukture na nelokalnoj (ili tercijarnoj) razini se najčešće procjenjuju s računicama energetskeg polja. Zlatni standard je *Rosettino ab initio* predviđanje strukture, koristeći Monte Carlo algoritam za umetanje fragmenata. Kreće se od ravnog lanca aminokiselinske sekvence i zatim se nasumično umeću strukturni fragmenti poznatih trodimenzionalnih struktura kako bi se dobile raznovrsne konformacije s različitim kombinacijama fragmenata. Nakon toga se radi visokorezolucijska optimizacija bočnih ogranaka i minimizacija. U tipičnom scenariju, tisuće neovisnih puteva sklapanja se simuliraju da bi se izgradilo energetskeg polje sklapanja. Korijen srednjeg kvadratnog odstupanja se računa između dizajnirane strukture i modela svake simulirane putanje sklapanja kako bi se izgradilo energetskeg polje na osnovi odnosa između energije i korijen srednjeg kvadratnog odstupanja, takav odnos je prikazan na slici 5. Energetskeg polje u obliku lijevka ukazuje na dobar dizajn zato što implicira da je dizajnirana struktura globalni minimum i da postoji znatna energetska razlika između dizajnirane strukture i drugih konformacija.

Sklapanje proteina *ab initio* metodama je računalno zahtjevna tehnika i zato je ograničena na ocjenjivanje malog skupa dizajna. Ako je broj dizajniranih sekvenci i dalje velik nakon korištenja čestih filtera koristi se simulacija poznata kao napredujuće pristrano sklapanje (engl. *biased forward folding*). Ova metoda se razlikuje od *ab initio* metode tako što se koristi manji niz fragmenata u modeliranju sklapanja, uobičajeno tri koja predstavljaju fragmente s najmanjim korijenom srednjeg kvadratnog odstupanja na svakoj poziciji. Ovom metodom dolazi do stvaranja pristranosti prema uzorkovanju konformacija prema dizajniranoj konformaciji. Smanjenje broja fragmenata znatno ubrzava računicu jer se smanji prostor koji se mora uzorkovati, a isto tako ako ishod putanje sklapanja nije nativna struktura pod ovim uvjetima, vrlo je mala vjerojatnost da će se to dogoditi koristeći *ab initio* metode. Ovom strategijom se može brzo procijeniti kvaliteta stotina ili tisuća dizajna, pri čemu se oni koji pokazuju najčešće uzorkovanje native konformacije odaberu za punu *ab initio* simulaciju.²⁰

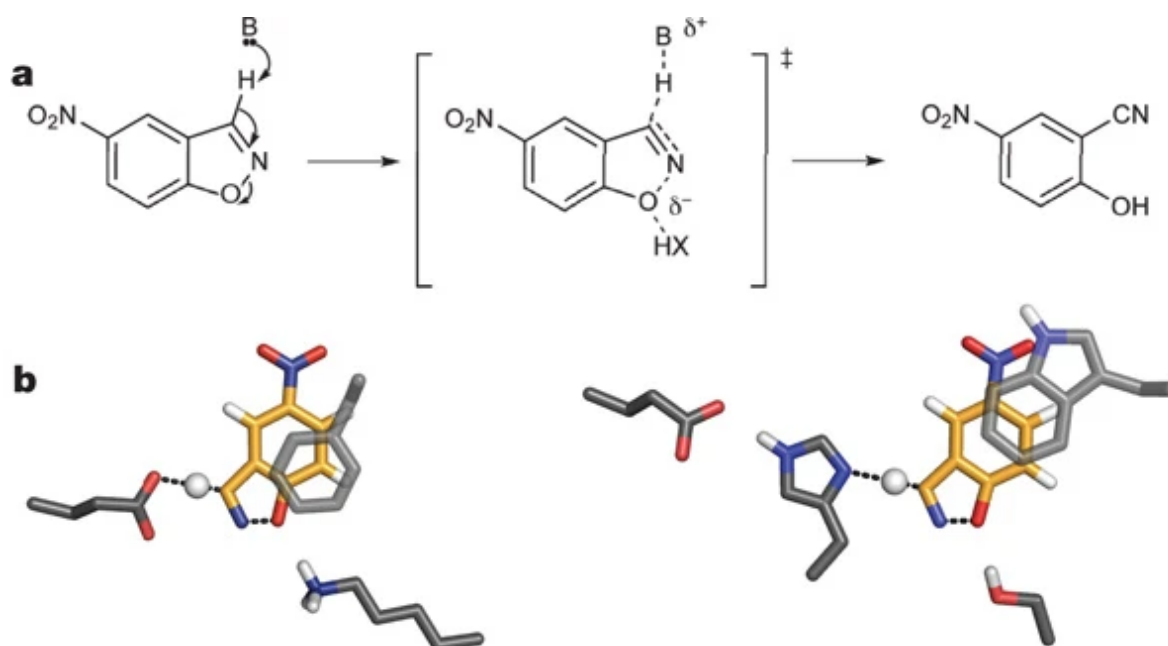
3.4. Eksperimentalna karakterizacija i udio uspješnosti dizajna

U računalnom dizajnu proteina, visoka kompatibilnost sekvence i strukture je potrebna za uspješniju eksperimentalnu karakterizaciju, kroz koju se onda određuje stabilnost, struktura i funkcionalnost dizajniranog proteina. Velika većina dizajniranih proteina pokazuje dobru razinu topljivosti, a manji udio njih usvoji ciljano monomerno stanje i strukturu. Problem često leži u netočnom identificiranju globalnog energetskeg minimuma što je prouzročeno ograničenim funkcijama energije i nedovoljnim uzorkovanjem u energetskom polju. Molekularno dinamičke simulacije su korisne kao dodatni kvalitativni test za stabilnost predviđene strukture.²⁰

§ 4. DE NOVO SINTEZA KEMPOVE ELIMINAZE

Dizajniran je enzim za katalizu Kempove reakcije⁷, za koju nije otkriven niti jedan prirodni enzim. Pristup se sastojao od kvantno-mehaničkog modeliranja idealiziranog aktivnog mjesta enzima, odabira aminokiselinskih bočnih ogranaka koji stabiliziraju prijelazno stanje, pretraživanja i odabira okosnica koje će podržati katalitičku geometriju, karakterizacije dobivenih *de novo* proteina i optimizacije pomoću usmjerene evolucije.

4.1. Geometrija prijelaznog stanja



Slika 6. (a) Kempova eliminacija 5-nitrobenzizoksazola. (b) Primjeri motiva aktivnog mjesta: lijevi koristi karboksilat za deprotonaciju, a desni His – Glu par. Aromatski aminokiselinski ostaci stabiliziraju prijelazno stanje π -stacking interakcijama u oba motiva. Preuzeto i prilagođeno prema: D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChanice, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Nature 453 (2008) 190–195.

U dizajnu enzima koji katalizira Kemp – eliminaciju, slika 6.a, potrebno je prvo odabrati katalitički mehanizam. Deprotonacija ugljika lužinom, ključni je korak Kempove eliminacije pa su odabrane dvije opcije, odnosno dva katalitička motiva, prikazana na slici 6.b;

1. Karboksilna grupa bočnog ogranka aspartata ili glutamata

2. Imidazolni prsten histidina pozicioniran i polariziran karboksilnom skupinom aspartata ili glutamata (imidazol – karboksilat par)

Prema autorima⁷, prednost karboksilata kao baze je u deprotoniranom, tj. bazičnom obliku bočnog ogranka. No, problem ovakvog pristupa je što dolazi do djelomične desolvatacije nabijene skupine u nepolarnim uvjetima, koja može destabilizirati protein te daljnjom desolvatacijom spriječiti vezanje. Korištenje imidazol-karboksilat para kao baze pogodnije je s obzirom na bolja lužnata svojstva histidina naspram karboksilata, no potrebna je regulacija njegovog pK_a i tautomernog stanja. Pri visokim vrijednostima pK_a histidin se može dvostruko protonirati čime postaje nedjelotvorna baza. Sparivanjem histidina sa aspartatom ili glutamatom pomaže u njegovom pozicioniranju i povećanju lužnatosti.

Idealizirano aktivno mjesto koje maksimizira stabilizaciju prijelaznog stanja izračunato je kvantno mehaničkim putem teorije funkcionala gustoće (engl. *Density functional theory*, DFT). Strukture prijelaznog stanja određene su za acetat i imidazol/acetat katalizirane reakcije u plinovitom stanju. Dodatno, korištene su funkcionalne skupine lizina, serina, treonina i tirozina kao donori vodikove veze za stabilizaciju nastalog negativnog naboja na fenolnom kisiku prijelaznog stanja. Katalitički motivi u kojima ne postoje aminokiseline koje nemaju donor vodikove veze su isto korišteni kod dizajna jer je naboj koji nastaje na fenolom kisiku tijekom Kempove eliminacije relativno mali i može se solvatirati vodom. Aromatski bočni ogranci fenilalanina, tirozina i triptofana modelirani su kako bi stabilizirali delokalizirani naboj prijelaznog stanja, te da osiguraju povoljne π -stacking interakcije.

Nakon dizajna aktivnog mjesta, pretraživani su setovi proteinskih okosnica kako bi se dobili kandidati koji će moći podržati prijelazno stanje. Kriteriji za odabir su: mogućnost dobivanja visokorezolucijskih kristalnih struktura, eksprimiranje u *E. Coli*, stabilnost proteina, postojanje džepa za smještanje prijelaznog stanja i raznolikost proteinske strukture.

4.2. Smještanje prijelaznog stanja i dizajn proteina

Veliki ansambl geometrija aktivnog mjesta je generiran tako što su se za svaki katalitički motiv varirale orijentacije katalitičkih bočnih ogranka u odnosu na prijelazno stanje. Ovo je omogućilo da se aktivno mjesto može smjestiti u raznolike proteinske strukture.

RosettaMatch algoritam korišten je u pozicioniranju prijelaznog stanja i katalitičkih aminokiselina u prethodno odabrane proteinske strukture, na način da se zadovolje sva katalitička ograničenja bez steričkih smetnji. Set dobivenih struktura obuhvaća veliki raspon proteinskih sklopova, TIM bačve, β -propelere, žele-role, Rossmanove nabore i lipokaline. U

tipičnom pretraživanju više od 100 000 mogućnosti za smještanje idealiziranog aktivnog mjesta pronađene su u odabranim strukturama. Za svako podudaranje korištena je metoda minimalizacije na bazi gradijenta kako bi se optimizirala orijentacija rigidnog tijela prijelaznog stanja i torzijskih kutova bočnih ogranaka katalitičkih aminokiselina, te da bi se najbolje zadovoljila sva katalitička geometrijska ograničenja.

Aminokiseline koje okružuju katalitičke bočne ogranke i prijelazno stanje redizajnirane su kako bi se optimizirale steričke, Coulombove i vodikove interakcije s prijelaznim stanjem i bočnim ograncima katalitičkih aminokiselina. Na ovaj način maksimizirana je stabilnost konformacije aktivnog mjesta, afinitet prema prijelaznom stanju te stabilnost proteina.

Dobiveni dizajni testirani su za kompatibilnost sa supstratom i produktom tako što su rangirani na osnovu katalitičke geometrije i izračunane energije vezanja prijelaznog stanja. Kroz korake procesa dizajniranja enzima zapažen je porast udjela dizajna koji sadrže TIM bačve. Prvo su predstavljale 25% svih proteina u ulaznom setu, zatim 43% kod inicijalnog podudaranja i 71% u nisko energetske dizajnim. Proučavanje dizajna sugerira da su vezni džepovi u TIM bačvama pogodniji zbog velikog broja pozicija koje su usmjerene prema šupljini i pokazuju kompatibilnost za katalitičke aminokiseline i stabilizaciju prijelaznog stanja. TIM bačva je jedna od najrasprostranjenijih i katalitički raznolikih struktura u prirodnim enzimima, stoga je poželjna struktura u *in silico* dizajnu Kempove eliminaze.

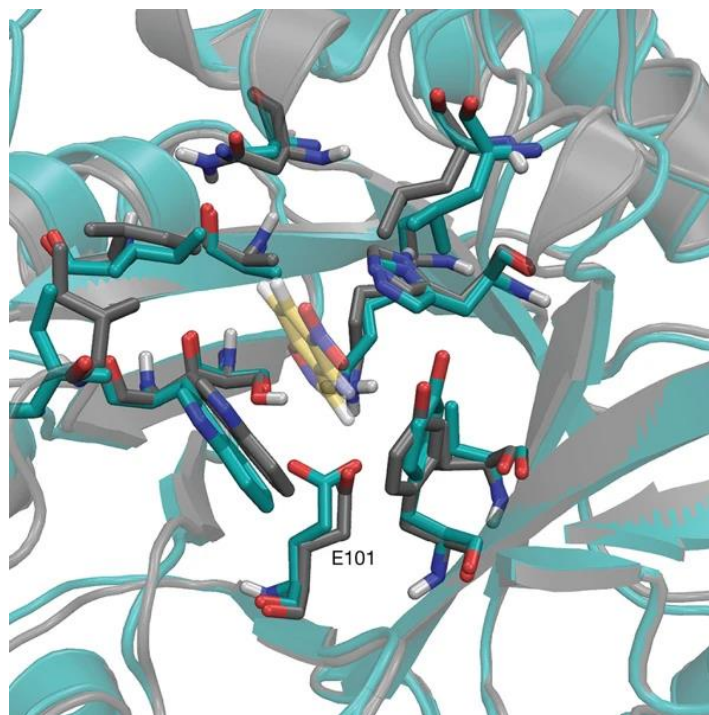
4.3. Eksperimentalna karakterizacija

Završetkom računalnog dizajna aktivnog mjesta, za eksperimentalnu karakterizaciju odabrano je 59 različitih proteinskih dizajna koji obuhvaćaju 17 različitih proteinskih strukturnih okvira. Njih 39 je koristilo karboksilat kao bazu (aspartat ili glutamat), a 20 histidin-karboksilat par (histidin-aspartat ili histidin-glutamat). Proteini su eksprimirani u *E. coli* BL21(DE3) soju koristeći pET29b vektor, zatim pročišćeni afinitetnom kromatografijom na Ni-NTA koloni, te su određeni kinetički parametri proteina. Osam dizajna je pokazalo mjerljivu aktivnost u katalizi Kempove eliminacije. Dokazi da je dizajnirano aktivno mjesto odgovorno za enzimsku aktivnost, dobiveni su mutiranjem katalitičke baza u alanin, odnosno glutamin/asparagin čime je zapažen nagli pad u aktivnosti ili potpuni nestanak katalitičkog svojstva. Dva najaktivnija dizajna su KE59 i KE70.

KE59 dizajn koji ima strukturu TIM bačve posjeduje Glu231 kao katalitičku bazu i Trp110 doprinosi delokalizaciji naboja pomoću π -stacking interakcija s prijelaznim stanjem. Dodatno, Leu108, Ile133, Ile178, Val159 i Ala210 stvaraju gusto pakirani hidrofobni džep koji okružuje

nepolarni substrat. Polarni ogranci Ser180 i Ser211 pridonose interakcijama vodikovih veza s nitro-skupinom prijelaznog stanja. Mutacija bočnog ogranka Glu231 u Gln na aktivnom mjestu je poništilo katalitičku aktivnost enzima. Pokušaji mutacije Gly131 u Ser kako bi se dodao donor vodikove veze koji bi stabilizirao negativni naboj na fenolnom kisiku je smanjilo katalitičku aktivnost, to se može interpretirati kao rezultat nepovoljnih elektrostatskih interakcije između kisikovih atoma na serinu i supstratu.

KE70 dizajn koristi mehanizam histidin-aspartat para (His-Asp par). Asp44 se pozicionira i polarizira His16 koji onda optimalno deprotonira supstrat. Tyr47 stvara π -stacking interakcije s prijelaznim stanjem i zajedno s Ile201, Ile139, Val167, Ala18, Ala102 i Trp71 stvara uski hidrofobni džep oko prijelaznog stanja. Aktivno mjesto je opet u okviru TIM bačve s His-Asp parom blizu dna mjesta. Mutacija katalitičke baze His16 u Ala je narušilo katalitičku aktivnost, dok mutacija Asp4 u Asn je smanjilo skoro trostruko katalitičku aktivnost.



Slika 7. Riješena kristalna struktura (cijan) dizajniranog proteina KE07 u ne vezanom stanju je uspoređena s strukturom dobivenom tijekom njegovog dizajna (sivo) u prisustvu prijelaznog stanja (žuto). Preuzeto i prilagođeno prema: D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChanice, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Nature 453 (2008) 190–195.

Uspješno je kristaliziran jedan od ranijih enzima u dizajnu, KE07, čiji mehanizam je temeljen na Glu bočnom ogranku. Na slici 7 se vidi da rješenja kristalna struktura i struktura modelirana dizajniranjem se jako dobro preklapaju. Sličnost između dizajniranog modela i

dobivene kristalne strukture sugerira da aktivna mjesta dobivena u novim enzimima slične onim u odgovarajućim dizajniranim modelima. Isto tako kristalna struktura je otkrila da Lys222 radi solni most sa katalitičkim Glu101 u odsustvu supstrata, dok kod dizajniranog modela amino skupina lizina stabilizira stvaranje fenoksida u prijelaznom stanju. Prema tome stvaranje kompleksa prijelaznog stanja zahtijeva kidanje solnog mosta, pa se očekuje da uklanjanje solnog mosta u nevezanom stanju poboljšava katalizu. Ova pretpostavka je testirana tako što se Lys222 supstituirao sa alaninom što je rezultiralo u poboljšanju efikasnosti enzima, k_{cat}/K_m je porastao za 2.5 puta.

4.4. Optimizacija enzima usmjerenom evolucijom

In vitro usmjerena evolucija pokazala je znatna poboljšanja u stabilnosti, ekspresiji i aktivnosti enzima, te je jedna od boljih metoda za optimizaciju biokatalizatora. No, *in vitro* evolucija enzima zahtijeva početnu točku s dovoljno visokom razinom željene katalitičke aktivnosti. Dizajniranjem enzima osigurava se točno pozicioniranje ključnih katalitičkih funkcionalnih skupina oko supstrata u prijelaznom stanju, stvarajući željenu katalitičku aktivnost. Računalno stvoren dizajn može se koristiti za stvaranje početne točke za usmjerenu evoluciju, unatoč tome što ne modelira sve moguće interakcije, pa je vrijedan alat za unaprjeđenje dizajna katalize i u razjašnjavanju njegovih nedostataka.

Kako bi se istražila razina do koje *in vitro* evolucijske metode mogu popraviti računalni dizajn enzima početni evolucijski eksperimenti provedeni su na KE07 – rani dizajn za koji je kristalna struktura određena. Genska biblioteka KE07 stvorena je kroz nasumičnu mutagenezu koristeći PCR podložan greškama. U svakom ciklusu su lizati 800 – 1600 individualnih kolonija testirani na hidrolizu 5-nitrobenzizoksazola, prateći stvaranje produkta spektrometrijski pri $\lambda = 380\text{nm}$. Najaktivniji klonovi su subklonirani i sekvencirani, a kodirajući plazmidi su korišteni kao početne točke u sljedećim ciklusima mutageneze. Kroz sedam ciklusa nasumične mutageneze i miješanja DNA dobivene su varijante s 4-8 mutacija u odnosu na KE07 i došlo je do poboljšanja u efikasnosti enzima, k_{cat}/K_m se povećao za preko 200 puta. Zabilježeno je kako ključni aspekti računalnog dizajna, uključujući identitete katalitičkih bočnih ogranaka, nisu izmijenjeni evolucijskim procesom već su mutacije uglavnom zapažene u reziduama koje se nalaze uz dizajnirane pozicije (npr. Val12, Ile102, Gly202), a fino su podesile dizajn enzima. Neke mutacije, kao Gly202 u Arg, vjerojatno su povećale fleksibilnost regije susjedne aktivnom mjestu. Hidrofobne rezidue Ile7 i Ile199 na dnu aktivnog mjesta često su mutirane u polarne ili nabijene rezidue (najčešća mutacija Ile7 u Asp), koje mogu držati Lys222 u položaju koji

stabilizira stvaranje negativnog naboja u prijelaznom stanju, te sprječava interakciju Lys222 s Glu101. Konzistentno s ovom idejom pK_a katalitičke rezidue Glu101 promijenilo se od 4.5 do 5.9 u varijanti gdje je Ile7 supstituiran s Asp. Iako Lys222 u Ala mutacija povećava aktivnost originalnog KE07, ona značajno smanjuje aktivnost varijanti dobivenih usmjerenom evolucijom uslijed postojanja nekompenziranog dodatnog naboja.

§ 5. LITERATURNI IZVORI

1. C. Jäckel, P. Kast, D. Hilvert, *Annu Rev Biophys* **37**, 153–173 (2008).
2. M. S. Packer & D. R. Liu, *Nat Rev Genet* **16**, 379–394 (2015).
3. D. N. Woolfson, *J Mol Biol* **433**, 167160 (2021).
4. H. W. Hellinga, *Proceedings of the National Academy of Sciences* **94**, 10015–10017 (1997).
5. R. A. Chica, N. Doucet, J. N. Pelletier, *Curr. Opin. Biotechnol.* **16** (2005) 378-384.
6. P. Huang, S. E. Boyken, D. Baker, *Nature* **537** (2016) 320–327.
7. D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChanice, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, *Nature* **453** (2008) 190–195.
8. K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, *Annu. Rev. Biophys.* **37** (2008) 289-316.
9. C. B. Anfinsen, *Science* **181**, 223–230 (1973).
10. A. V. Finkelstein, N. S. Bogatyreva, D. N. Ivankov, S. O. Garbuzynskiy, *Biophys Rev* **14**, 1255–1272 (2022).
11. A. A. Nickson & J. Clarke, *Methods* **52**, 38–50 (2010).
12. D. L. Nelson & M. M. Cox, *Lehninger Principles of Biochemistry*, W.H. Freeman, 2013, str. 146, 1114.
13. Y. Tan, H. Wu, T. Wei, X. Li, *J Am Chem Soc* **142**, 20288–20298 (2020).
14. M. L. Casey, D. S. Kemp, K. G. Paul, D. D. Cox, *J. Org. Chem.* **13** (1973) 2294-2301.
15. D.S. Kemp, M. L. Casey, *J. Am. Chem. Soc.* **95** (1973) 6670-6680.
16. A. Genre-Grandpierre, C. Tellier, M. Loirat, D. Blanchard, D. R.W. Hodgson, F. Hollfelder, A. J. Kirby, *Med Chem Lett* **7**, 2497–2502 (1997).
17. S. C. Popa, I. Inamoto, B. W. Thuronyi, J. A. Shin, *ACS Omega* **5**, 26957–26966 (2020).
18. I. V. Korendovych & W. F. DeGrado, *Q Rev Biophys* **53**, e3 (2020).
19. A. G. Street & S. L. Mayo, *Structure* **7**, R105–R109 (1999).
20. E. Marcos, & D. Silva, *WIREs Computational Molecular Science* **8**, (2018).
21. S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler, D. Baker, *PLoS One* **6**, e20161 (2011).

22. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, *Methods in Enzymology* **487**, (2011).
23. S. Chaudhury, S. Lyskov, J. J. Gray, *Bioinformatics* **26**, 689–691 (2010).
24. C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, D. Baker, *Methods in Enzymology* **383**, (2004).
25. T. M. Jacobs, B. Williams, T. Williams, X. Xu, A. Eletsy, J. F. Federizon, T. Szyperski, B. Kuhlman, *Science* **352**, 687–690 (2016).