

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

Kemijski odsjek

Lucija Brtan

**Strategije za poboljšanje prekomjerne proizvodnje proteina**

Kemijski seminar I

Poslijediplomski sveučilišni studij Kemija (smjer Biokemija)

Izrađen prema:

G.L. Rosano, E.S. Morales, E.A. Ceccarelli, *Protein Sci* **8** (2019) 1412-1422.

K.H. Sumida et al, *J. Am. Chem. Soc.* **146(3)** (2024) 2054-2061.

Zagreb, 2025.

Sadržaj

[§ 1. UVOD 1](#_Toc195732098)

[§ 2. Bakterija *E. coli* kao domaćin rekombinantne ekspresije proteina 2](#_Toc195732099)

[2.1. Soj BL21(DE3) kao preferirani domaćin za ekspresiju rekombinantnih proteina 2](#_Toc195732100)

[2.2. Neki od izvedenih laboratorijski sojeva BL21(DE3) i njihova uloga u optimizaciji rekombinantne ekspresije 3](#_Toc195732101)

[§ 3. Inovacije u dizajnu plazmida: modifikacije ključnih funkcionalnih domena plazmida 5](#_Toc195732102)

[3.1. Replikon i broj kopija plazmida 5](#_Toc195732103)

[3.2. Promotori i njihova regulacija 6](#_Toc195732104)

[3.3. Selekcijski markeri i alternativni sustavi selekcije 6](#_Toc195732105)

[3.4. Afinitetni privjesci i fuzijski partneri 6](#_Toc195732106)

[§ 4. Uvjeti kultivacije i praćenje rasta 7](#_Toc195732107)

[§ 5. računalna optimizacija proteinskih sekvenci 8](#_Toc195732108)

[5.1. Računalni alat ProteinMPNN 9](#_Toc195732109)

[5.2. Dizajn varijanti mioglobina s povećanom stabilnošću 10](#_Toc195732110)

[5.3. Dizajn varijanti TEV proteaze s poboljšanom stabilnošću i katalitičkom aktivnošću 12](#_Toc195732111)

[§ 6. ZAKLJUČAK 16](#_Toc195732112)

[§ 7. Literaturni izvori xvii](#_Toc195732113)

1. UVOD

Razvoj rekombinantne ekspresije proteina u mikrobnim sustavima, osobito u bakteriji *Escherichia coli*, predstavlja jednu od najvažnijih prekretnica u suvremenoj biokemiji i biotehnologiji. Tradicionalni pristupi koji su zahtijevali velike količine biljnog ili životinjskog materijala te znatne volumene bioloških tekućina za izolaciju ciljnih proteina, danas su u velikoj mjeri zamijenjeni preciznijim, učinkovitijim i ekonomičnijim metodama prekomjerne ekspresije proteina. Rekombinantna ekspresija značajno je unaprijedila mogućnosti za strukturnu i funkcionalnu karakterizaciju proteina, kao i njihovu primjenu u industriji i razvoju biotehnoloških proizvoda.1

U teoriji, postupak dobivanja rekombinantnog proteina slijedi relativno jednostavan slijed koraka: genska sekvenca od interesa klonira se u odgovarajući ekspresijski vektor, koji se potom transformira u prikladnog domaćina. Nakon indukcije ekspresije slijedi izolacija prekomjerno eksprimiranog proteina te pročišćavanje. Međutim, u praktičnoj primjeni ovaj proces često uključuje niz izazova. Među najčešćim su neadekvatan rast stanica domaćina, toksičnost ciljnog proteina, stvaranje netopljivih inkluzijskih tijela, kao i potpuni izostanak ekspresije.1,2

Zbog navedenih ograničenja, uspješna proizvodnja rekombinantnih proteina zahtijeva detaljno planiranje, optimizaciju i često multidisciplinarni pristup.

Bakterija *Escherichia coli* se smatra najčešće korištenim domaćinom zbog visokih prinosa ekspresije, jednostavne manipulacije te niskih troškova uzgoja. Upravo zbog toga, *E. coli* je dobro istražen i opisan sustav za prekomjernu ekspresiju proteina, a genetskim inženjeringom razvijeni su brojni sojevi modificirani za olakšanu ekspresiju različitih rekombinantnih proteina.2

U ovom seminaru prikazat će se recentna dostignuća u području proizvodnje rekombinantnih proteina s naglaskom na napredak u razvoju sojeva *E. coli,* dizajnu ekspresijskih vektora, optimizaciji uvjeta kultivacije te primjeni računalnih alata za poboljšanje ekspresijskih svojstava i stabilnosti proteina.

1. Bakterija *E. coli* kao domaćin rekombinantne ekspresije proteina

Odabir organizma domaćina, čiji će stanični mehanizmi sinteze proteina služiti za proizvodnju ciljanog proteina, ujedno definira cjelokupni metodološki pristup proizvodnje proteina. Među mikroorganizmima, kao sustavi za ekspresiju najčešće se koriste bakterije, kvasci, vlaknaste gljive i jednostanične alge. Svaki od navedenih sustava ima svoje prednosti i ograničenja, a njihov odabir uvelike ovisi o specifičnostima ciljanog proteina.3,4 Na primjer, ukoliko je za funkcionalnost proteina potrebna eukariotska posttranslacijska modifikacija (poput glikozilacije), prokariotski ekspresijski sustavi poput *Escherichia coli* neće biti prikladni.5

Prednosti korištenja *E. coli* kao organizma domaćina dobro su poznate: *E. coli* posjeduje iznimno brzu kinetiku rasta. U medijima koji sadrže glukozu i soli, uz optimalne uvjete okoline i aeracije, vrijeme udvostručenja populacije iznosi oko 20 minuta.6 To znači da kultura inokulirana 1:100 razrjeđenjem zasićene predkulture može dosegnuti stacionarnu fazu rasta unutar svega nekoliko sati. Međutim, valja imati na umu da sama ekspresija rekombinantnog proteina može predstavljati značajan metabolički stres za stanicu, čime se generacijsko vrijeme znatno produljuje.7

* 1. Soj BL21(DE3) kao preferirani domaćin za ekspresiju rekombinantnih proteina

Brzim razvojem biotehnologije krajem 20. stoljeća, brojni sojevi bakterije Escherichia coli istraživani su s ciljem optimizacije proizvodnje rekombinantnih proteina. Među njima se istaknula linija B, čiji je derivat **BL21(DE3)** postao jedan od najčešće korištenih sojeva u eksperimentalnoj i industrijskoj proizvodnji rekombinantnih proteina. Zahvaljujući sekvenciranju genoma ovih sojeva, danas su dobro poznate molekularne osnove fenotipova koji omogućuju visoku učinkovitost heterologne ekspresije. 13-15

Soj BL21(DE3) pokazuje brojne prednosti u usporedbi s drugim laboratorijskim linijama, kao što je K-12. Na primjer, u standardnim uvjetima uzgoja u bogatim medijima, vrijeme udvostručenja stanica je približno 20 minuta, dok u minimalnim medijima linija B također pokazuje superioran rast.13 Osim toga, ove stanice su **nemotilne,** što se pripisuje deleciji u fli genima, ključnima za biosintezu flagela. Izostanak sinteze flagela smanjuje energetski trošak stanice, što može dodatno doprinijeti brzini rasta, čime se smanjuje vrijeme uzgoja, a i same ekspresije proteina.16

Metabolički profil soja BL21 također je povoljniji - primjerice, tijekom rasta na glukozi kao jedinom izvoru ugljika, proizvodi manje količine acetata u usporedbi s K-12 sojem. S obzirom da se pri pH vrijednosti od oko 7,0 acetat nalazi u ravnoteži s octenom kiselinom koja može difuzijom ulaziti u stanicu i narušavati unutarstaničnu homeostazu, ovaj fenomen može utjecati na smanjenje vitalnosti stanica. Međutim, pokazano je da kultivacija pri nešto višim pH vrijednostima (7.5–8.5) smanjuje „acetatni stres“ i time poboljšava prinos rekombinantnih proteina.17-19

Još jedna važna karakteristika ovog soja je **deficijencija proteolitičkih sustava**. Naime, integracija IS186 elementa u promotor lon gena rezultira inaktivacijom jedne od glavnih staničnih proteaza, dok delecija gena za membransku proteazu OmpT smanjuje razgradnju proteina tijekom i nakon stanične lize. Ove modifikacije uvelike doprinose stabilnosti proizvedenih proteina.15

Također je zabilježeno da BL21 stanice imaju **poboljšane sekrecijske sposobnosti**, djelomično zahvaljujući prisutnosti dodatnog sustava tipa II za sekreciju (T2SS), što u određenim slučajevima može biti povoljnije od citoplazmatske ekspresije.8-9

Kao ključna inovacija u razvoju ovog sustava ističe se konstrukcija soja **BL21(DE3)** od strane Studiera i Moffata, koji u genomu sadrži integrirani gen za **T7 RNA polimerazu** pod kontrolom **lacUV5 promotora.** U ovom sustavu, ciljna genska sekvenca klonira se u ekspresijski plazmid pod kontrolom T7 promotora, a ekspresija se inducira dodatkom izopropil -D-1-tiogalaktopiranozida (IPTG-a). Ova konfiguracija omogućuje **preciznu i snažnu kontrolu ekspresije**, što je čini izrazito pogodnom za proizvodnju različitih proteina.9,10,14

* 1. Neki od izvedenih laboratorijski sojeva BL21(DE3) i njihova uloga u optimizaciji rekombinantne ekspresije

Kako je opisano u prethodnom poglavlju, soj Escherichia coli BL21(DE3) postao je nezaobilazna platforma za proizvodnju rekombinantnih proteina, što je potaknulo razvoj brojnih derivata optimiziranih za različite izazove ekspresije. Ovi sojevi rezultat su **racionalnog inženjeringa**, pri čemu su ciljano modificirani poznati molekularni mehanizmi s ciljem povećanja razine, topljivosti i stabilnosti eksprimiranih proteina, kao i preciznije kontrole ekspresije.9

* + 1. **Regulacija bazalne ekspresije**

Jedan od prvih izazova u ekspresiji toksičnih ili osjetljivih proteina jest smanjenje tzv. „leaky” ekspresije, odnosno sinteze proteina i u odsutnosti spoja za indukciju. Soj **BL21(DE3) pLysS** rješava ovaj problem uvođenjem plazmida pLysS s genom za **T7 lizozim** koji jeprirodni inhibitor T7 RNA polimeraze, čime se omogućuje **preciznija regulacija transkripcije i bazalne ekspresije**. **22-24**

* + 1. **Korekcija kodonske pristranosti**

Sojevi **CodonPlus** i **Rosetta** konstruirani su za ekspresiju eukariotskih gena čiji kodoni nisu česti u E. coli. Oni sadržavaju dodatne kopije gena za rijetke tRNA molekule (npr. Arg, Leu, Ile). U novije vrijeme, ovaj je pristup dodatno optimiziran integracijom tih tRNA gena direktno u rRNA operon, čime je nastao soj **SixPack**. Takva modifikacija nije utjecala na uvjete rasta, ali je omogućila značajan porast rijetkih tRNA molekula u eskponencijalnoj fazi rasta.26

* + 1. **Formiranje disulfidnih mostova**

Sojevi poput **Origami**, **SHuffle** i **CyDisCo** razvijeni su za ekspresiju proteina koji sadrže **disulfidne veze**. Oni sadrže mutacije u redoks sustavima (geni trxB, gor), čime se omogućuje **oksidativno okruženje u citoplazmi** i olakšava pravilno smatanje eukariotskih proteina.2

* + 1. **Podešavanje razine ekspresije**

Sojevi **Tuner** i **Lemo21(DE3)** omogućuju **fino podešavanje ekspresije** pomoću varijabilne indukcije (npr. različite koncentracije IPTG-a ili L-ramnoze). U sustavu **RiboTite**, razina ekspresije dodatno je regulirana riboprekidačem (engl. riboswitch), koji omogućuje **ekspresiju samo u prisutnosti dvaju induktora**, IPTG-a i PPDA.2

* + 1. **Sekrecija smotanih proteina**

Kod proteina osjetljivih na citoplazmatske uvjete, sekrecija u periplazmu može značajno poboljšati stabilnost. Soj **TatExpress BL21** koristi modificirani **Tat sekrecijski sustav**, omogućujući sekreciju potpuno smotanih proteina mase do 150 kDa, poput **rekombinantnog ljudskog hormona rasta**, s prinosom do **30 mg/L**.2,20,21

* + 1. **Koekspresija više proteina**

Za koekspresiju više rekombinantnih proteina koristi se **Duet sustav plazmida i modificiranog BL21(DE3) soja**, koji omogućava kloniranje dvaju gena pod zasebnim T7 promotorima unutar jednog plazmida. Kombinacijom kompatibilnih Duet plazmida moguće je potaknuti koekspresiju do **osam različitih proteina**. Ipak, budući da svi promotori u sustavu reagiraju na isti induktor, **precizna kontrola stehiometrije pojedinih gena može biti ograničena**.27

* + 1. **Multiparametarska genetska kontrola: Marionette sojevi**

U cilju rješavanja problema stehiometrijske kontrole u kompleksnim biosintetskim putevima, Meyer i sur. razvili su **Marionette sojeve**, dostupne u **BL21, DH10B** i **MG1655** sojevima. Ovi sojevi sadržavaju **12 različitih transkripcijskih faktora**, integriranih u genom, koji odgovaraju na specifične kemijske induktore. Time je omogućena **neovisna i precizna regulacija ekspresije** svakog pojedinog gena.27

1. Inovacije u dizajnu plazmida: modifikacije ključnih funkcionalnih domena plazmida

U sklopu heterologne ekspresije proteina, ciljna sekvenca se najčešće klonira u odgovarajući ekspresijski plazmid. Osnovni elementi takvog plazmida uključuju promotor, regiju za inicijaciju translacije, selekcijski marker te replikacijski element (replikon). Ovisno o svrsi istraživanja, vektor može sadržavati i dodatne domene kao što su afinitetni privjesci fuzijski partneri, kao i sekvence koje poboljšavaju topljivost proteina ili olakšavaju njegovu detekciju i pročišćavanje.

* 1. Replikon i broj kopija plazmida

Replikon je osnovna jedinica autonomne replikacije plazmida te uključuje podrijetlo replikacije i regulatorne sekvence. Broj kopija plazmida u stanici izravno ovisi o tipu replikona, što može značajno utjecati na ukupni prinos proteina, ali i na stabilnost sustava. Plazmidi s visokim brojem kopija, kao što su oni iz pUC serije (500–700 kopija), mogu uzrokovati metaboličko opterećenje, smanjenu stopu rasta i nestabilnost plazmida. S druge strane, plazmidi s niskim brojem kopija, poput onih temeljenih na p15A ori (npr. pACYC i pBAD), omogućuju stabilniji sustav kada je proizvodnja ciljnog proteina toksična za stanicu domaćina. Za višestruku ekspresiju, koriste se kompatibilni replikoni kako bi se izbjegla međusobna kompeticija za replikacijske enzime.28-30

* 1. Promotori i njihova regulacija

Promotor je ključni regulatorni element koji određuje vremenski tijek i intenzitet ekspresije ciljnog gena. Najčešće korišteni promotori u E. coli sustavima su T7, tac, lac i araPBAD. T7 promotor, prisutan u pET seriji vektora, izuzetno je snažan te omogućuje visoke razine transkripcije nakon indukcije IPTG-om. Međutim, visoka bazalna ekspresija može biti problematična, osobito kod toksičnih proteina, što se može ublažiti dodatkom T7 lizozima (npr. korištenjem ranije opisanog soja koji sadrži pLysS plazmid).

Za finiju kontrolu ekspresije koriste se inducibilni promotori poput araPBAD, čija se aktivacija događa u prisutnosti arabinoze, ili temperaturno ovisni promotori kao što su cspA i λcI857. Također, inovativni sustavi poput RiboTite omogućuju dvofaktorsku kontrolu ekspresije pomoću kombinacije IPTG i specifičnog efektora (npr. PPDA), čime se gotovo u potpunosti uklanja neželjena bazalna ekspresija.28,31,32

* 1. Selekcijski markeri i alternativni sustavi selekcije

Za osiguranje prisutnosti plazmida u stanicama domaćina, tradicionalno se koriste geni za otpornost na antibiotike, poput *bla* (ampicilin), *cat* (kloramfenikol) ili *tet* (tetraciklin).33 No, zbog potencijalnog širenja antimikrobne rezistencije i visokih troškova, razvijeni su i alternativni sustavi bez antibiotika, tzv. „plasmid addiction“ sustavi. Oni se temelje na brisanju esencijalnog gena iz kromosoma i njegovom uključivanju u plazmid, čime se osigurava da samo stanice koje zadrže plazmid prežive.34

* 1. Afinitetni privjesci i fuzijski partneri

Za olakšano pročišćavanje i detekciju rekombinantnog proteina, uobičajeno se koriste afinitetni privjesci kao što su His-tag, FLAG-tag, Strep-tag ili c-Myc.35 Ovi peptidni privjesci omogućuju specifično vezanje proteina na kromatografske smole te jednostavnu eluciju. Veći fuzijski partneri poput MBP (maltose-binding protein), NusA, GST ili SUMO poboljšavaju topljivost i stabilnost proteina, ali se nakon pročišćavanja najčešće uklanjaju proteolitičkom razgradnjom. U novije vrijeme razvijeni su i tzv. stimulus-osjetljivi tagovi, poput ELP (engl. elastin-like polipeptidi) ili *β*-roll domena, koji omogućuju selektivnu precipitaciju proteina promjenom temperature ili ionske jakosti, što značajno smanjuje potrebu za klasičnim kromatografskim postupcima.36

1. Uvjeti kultivacije i praćenje rasta

Promjena uvjeta uzgoja najjednostavniji je način za modulaciju rasta E. coli, a time se izravno utječe na volumetrijski prinos rekombinantnog proteina. Iako se često optimizira temperatura kulture nakon indukcije, ostali vanjski parametri često se zanemaruju. Primjerice, Luria-Bertani (LB) medij, iako vrlo popularan, nije najsvrsishodniji za prekomjernu ekspresiju proteina.2 Zamjena LB medija za bogatije medije poput Terrific Broth (TB) rezultira većom gustoćom uzgojenih stanica.

U posljednje vrijeme sve veću popularnost dobivaju **autoindukcijski mediji**, a brojna istraživanja potvrđuju njihovu učinkovitost u proizvodnji rekombinantnih proteina. Ovi mediji sadrže barem dva izvora ugljika - glukozu i laktozu - a često i dodatni glicerol koji doprinosi većim prinosima. Glukoza se troši prva tijekom eksponencijalnog rasta, dok se nakon njenog iscrpljenja aktivira potrošnja laktoze, koja zatim inducira transkripciju gena pod kontrolom *lac* promotora. Prednosti autoindukcije uključuju postizanje viših gustoća stanica, visoku ponovljivost trenutka indukcije te uklanjanje potrebe za ručnim dodatkom IPTG-a.37

Složeni (kompleksni) mediji osiguravaju obilje hranjivih tvari, no ne omogućuju preciznu kontrolu metaboličkog statusa stanica. Kako bi se povećala učinkovitost proizvodnje, poželjno je usporiti rast i time omogućiti bolji balans između sinteze i pravilnog smatanja proteina. Najjednostavniji način kontrole rasta je manipulacija temperaturom. Međutim, u velikim bioreaktorima, precizna kontrola temperature značajno povećava troškove pa se razvijaju alternativne metode.37

Jedan od ključnih izazova je razlika u strategijama između **fed-batch** kultivacija **(odnosno uzgoj u bioreaktoru)** i **shake flask** kultivacija (uzgoj tresenjem tikvica u inkubatoru). Dok se u fed-batch sustavima pažljivo prate aeracija i dotok glukoze,37 uzgojem u tikvicama te parametre često nije moguće kontrolirati. No, posljednjih godina razvijene su metode koje omogućuju preciznije upravljanje uzgojem i u laboratorijskim tikvicama. Primjerice, postignuta je **kontrolirano otpuštanje glukoze** pomoću silikonskih kuglica38 ili enzimatskom razgradnjom polisaharida ***in situ***.39 Dostupnost kisika može se dodatno podešavati dizajnom tikvica40 ili dodatkom nemješajućih kisikom obogaćenih ulja.41 Osim toga, razvijeni su uređaji koji omogućuju **praćenje parametara kulture** u mikropločicama i shake flaskovima, čime se ubrzava pronalazak optimalnih uvjeta za ekspresiju.42 Unatoč tome, potreba za specijaliziranom opremom i dalje ograničava širu primjenu ovih pristupa.

Zanimljiv novitet u ovom području je **Opto-T7RNAP** sustav koji koristi plavu svjetlost za indukciju ekspresije.43 U ovom sustavu T7 RNA polimeraza je podijeljena na dvije podjedinice, od kojih je svaka spojena na domene koje dimeriziraju pod utjecajem plave LED svjetlosti. Po osvjetljavanju dolazi do njihovog spajanja i nastanka funkcionalne T7RNAP – čime plava svjetlost služi kao induktor. U mraku dolazi do disocijacije i prestanka ekspresije, čime se omogućuje **precizna, reverzibilna i vremenski kontrolirana ekspresija** proteina. Ovaj sustav predstavlja održivu alternativu IPTG-u, osobito za primjene u industrijskim fermentacijama.

1. računalna optimizacija proteinskih sekvenci

Svi ranije opisani pristupi u (re)dizajnu plazmida i bakterijskih sojeva omogućuju preciznu kontrolu svih aspekata rekombinantne ekspresije, od kontrole transkripcije i translacije, preko stabilnosti sustava, do efikasnog pročišćavanja čime predstavljaju temelj suvremenih pristupa u biotehnologiji i proteinomici. Ponekad, uz sve načinjene modifikacije, prekomjerna ekspresija ciljnog proteina i daje rezultira niskim prinosom. Razvojem neuronskih mreža, dubokog učenja (engl. *Deep Learning*) te umjetne inteligencije, sve više znanstvenika se okreće optimizaciji sekvenci samog proteina navedenim alatima, a sve s ciljem poboljšanja topljivosti i općenitih performansi proteina poput povećanja termičke stabilnosti ili kinetike.44

Jedan od takvih alata je **PROSS** (engl. *protein repair one-stop shop*), koji kombinira informacije iz evolucije proteina i fizikalno-energetske izračune u sklopu programa Rosetta te na temelju 3D strukture ciljnog proteina generira optimizirane sekvence. Ovakv pristup se pokazao učinkovitim u poboljšanju topljivosti i termostabilnosti više prirodnih proteina, što je zainteresiralo brojne znanstvenike da nastave istraživanja u smjeru razvoja i optimizacije novih algoritama koji bi olakšali ovakve procese. Paralelno, rapidnim razvojem dubokog učenja, kreirani su različiti modeli poput jezičnih modela (engl. Large Language Models, skraćeno LLM), koji generiraju nove sekvence za određene enzimske obitelji ili funkcije, konvolucijskih neuronskih mreža koje potom koriste strukturne informacije za predviđanje mutacija s poboljšanom funkcijom, te plićih neuronskih mreža za usmjeravanje kombinatorne evolucije.45-47

Među najistaknutijijim alatima nove generacije je **ProteinMPNN**, koji generira vrlo stabilne sekvence za dizajnirane proteine, ali i optimizira sekvence prirodnih proteina na način da se predviđa sigurnije smatanje u željenu strukturu u usporedbi s izvornom sekvencom.

Kao modelni sustavi u istraživanju i testiranju takvog alata odabrani su mioglobin, jedan od prvih proteina čija je struktura riješena, te široko korištena proteaza iz Tobacco Etch Virusa (TEV).47

* 1. Računalni alat ProteinMPNN

ProteinMPNN je računalni alat razvijen za dizajniranje aminokiselinskih sekvenci koji s visokom pouzdanošću predviđa da će se smotati u zadanu trodimenzionalnu (3D) strukturu. Ova metoda temelji se isključivo na strukturnim informacijama i ne koristi podatke o funkcionalnosti proteina. Iz tog razloga, vrlo je važno prilikom samog dizajna nadopuniti neuronsku mrežu dodatnim informacijama o funkcionalnosti kako bi se očuvala funkcija proteina. Za te svrhe, ispitan je niz pristupa. U svim slučajevima, kako bi se očuvala katalitička aktivnost i vezno mjesto za supstrat, fiksiraju se izvorne kordinate aminokiselinskih ostatataka u neposrednoj blizini supstrata (definirani unutar 7 Å od vezanog liganda u kristalnoj strukturi kompleksa). Te pozicije nazvane su funkcionalne pozicije prve ljuske i one nisu bile podložne redizajnu.47

Za redizajn TEV proteaze, bilo je potrebno dodatno koristiti podatke iz evolucijske analize kako bi se identificirali aminokiselinski ostatci koji su ključni za katalitičku aktivnost. Kod proteina mioglobina, provedene su i ograničene promjene u strukturi proteinske okosnice (engl. *backbone redesign*) kako bi se dodatno stabilizirao protein.

Zatim su generirane sekvence pomoću alata ProteinMPNN, a strukture tih sekvenci potom su predviđene korištenjem AlphaFold2 algoritma. Dobivene varijante filtrirane su prema pLDDT (engl. *predicted Local Distance Difference Test*) vrijednostima te prema srednjem kvadratnom odstupanju C*α* atoma (C*α* RMSD), čime su odabrani najstabilniji i najvjerojatniji dizajni za daljnju eksperimentalnu validaciju.47,48

* 1. Dizajn varijanti mioglobina s povećanom stabilnošću

Mioglobin je hem-vezujući protein odgovoran za skladištenje kisika u mišićnom tkivu sisavaca, a osim biološke uloge, nalazi primjenu u kliničkoj dijagnostici kao biomarker, te sve više se primjenjuje u industriji biokatalizatora i prehrambenoj industriji.

U provedenom istraživanju korištena je struktura ljudskog mioglobina (nMb, PDB: 3RGK) kao polazišna struktura za redizajn. Kako bi se sačuvala njegova funkcija vezanja kisika, 17 aminokiselina u neposrednoj blizini hema fiksirano je tijekom dizajna. Ukupno je generirano 60 sekvenci pomoću algoritma ProteinMPNN, a zatim su strukture predviđene koristeći AlphaFold2. Odabrano je osam dizajna s visokom strukturnom podudarnošću s nativnim modelom (pLDDT > 85.0 i Cα RMSD < 1.0 Å). Četiri dizajna s najvišom podudarnošću u regiji vezanja hema odabrana su za eksperimentalnu validaciju.

Nadalje, provedena je ograničena rekonstrukcija slabije strukturiranih regija mioglobina kako bi se dodatno povećala stabilnost. Mioglobin, kao globularni protein, karakterizira osam *α*-zavojnica s dvije fleksibilne petlje oko mjesta vezanja hema. Regije između zavojnica E i F te petlje CD remodelirane su uz pomoć metode RoseTTAFold joint inpainting. Iz tih preuređenih modela generirane su nove sekvence uz pomoć alata ProteinMPNN, pri čemu je regija vezanja hema ostala netaknuta. Dodatnih 16 sekvenci, koje su zadovoljile kriterije pLDDT i RMSD predikcije, odabrano je za testiranje.

Ukupno 20 dizajniranih sekvenci sintetizirano je i eksprimirano u *E. coli*. Holo-proteini s vezanim hemom pročišćeni su pomoću kromatografije s imobiliziranim metalnim ionima (*engl. Immobilized Metal Affinity Chromatography,* IMAC) i kromatografijom isključenjem po veličini (*engl. Size-Exclusion Chromatography*, SEC) metoda. Svi dizajni su uspješno eksprimirani i pokazali monomernu prirodu proteina u SEC analizi. Trinaest dizajna imalo je veći prinos topljivog proteina (do 4.1 puta više) u odnosu na nativni mioglobin. Svi su dizajni pokazali slična spektroskopska svojstva vezanja hema kao i nativni protein, potvrđujući očuvanu funkcionalnost.

Termalna stabilnost osam najboljih dizajna ispitana je metodom cirkularnog dikroizma (CD). Svi dizajni pokazali su višu temperaturu denaturacije od nativnog mioglobina (80 °C), pri čemu je šest uzoraka ostalo stabilno i pri 95 °C. Također, sposobnost vezanja hema očuvana je pri povišenim temperaturama, pri čemu je pet dizajna zadržalo značajno vezanje hema pri 95 °C.

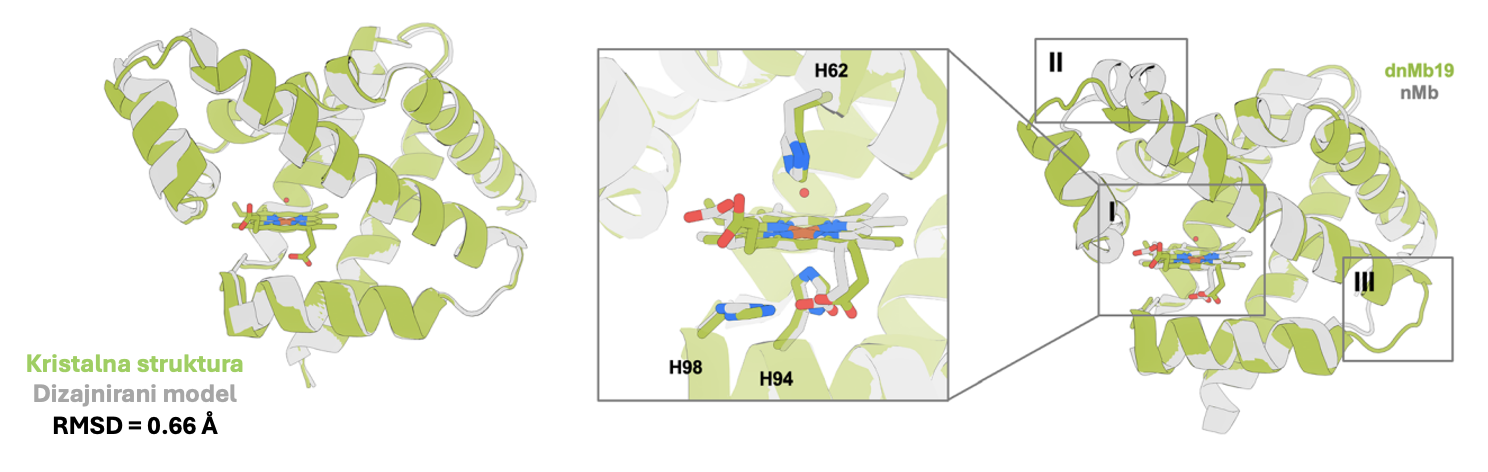
A collage of graphs and diagrams

AI-generated content may be incorrect.

**Slika 1.** Poboljšanje ekspresije i termostabilnosti mioglobina pomoću ProteinMPNN algoritma.  
**(a)** Prikaz kristalne strukture mioglobina s označenim pozicijama koje su fiksirane tijekom dizajna (plavo) – aminokiseline u neposrednoj blizini hema. Žute regije predstavljaju nekonzervirane dijelove proteinske okosnice podložne strukturnom remodeliranju (engl. backbone redesign).  
**(b)** Kromatogrami dobiveni nakon gel-filtracije 20 dizajniranih varijanti mioglobina, gdej se vidi kako svi analizirani proteini egzistiraju kao monomeri. **(c)** Prinos topljivog proteina za sve dizajne u usporedbi s nativnim mioglobinom (nMb, crvena isprekidana linija). **(d)** Krivulje dobivene CD spektroskopijom za dnMb19 i nativni mioglobin prikazuju poboljšanu termostabilnost dizajnirane varijante (signal izražen kao molarna rezidualna eliptičnost – MRE). **(e)** UV/Vis apsorpcijski spektar za dnMb19 i nativni mioglobin, s prikazom temperaturne stabilnosti. Preuzeto i prilagođeno prema referenci 47.

Posebno se ističe dizajn dnMb19, koji je pokazao najvišu točku topljivosti, uz očuvanu strukturu prema kristalografskim podacima (2.0 Å rezolucija, PDB: 8U5A).

Kristalna struktura dnMb19 pokazuje visoku podudarnost s dizajniranim modelom (C*α* RMSD = 0.66 Å), uključujući regije koje su bile podložne rekonstrukciji. Značajne promjene zabilježene su u konformacijama zavojnica C i E te petlje između E i F zavojnica, dok su interakcije sa hemom u velikoj mjeri očuvane.



**Slika 2.** *Lijevo:* Usporedba predviđene strukture dizajna dnMb19 pomoću AlphaFold2 (sivo) i eksperimentalno određene kristalne strukture (zeleno), ukazuje na visoku strukturnu podudarnost. Desno: Preklapanje kristalne strukture nativnog mioglobina (sivo) i dizajnirane varijante dnMb19 (zeleno, PDB: 8U5A). Petlje II i III prikazuju strukturno redizajnirane regije koje ne sudjeluju u vezanju hema, uključujući produljenje heliksa i zamjenu fleksibilnih petlji. Preuzeto i prilagođeno prema referenci 47.

Ovime je potvrđeno da su metode ProteinMPNN i RoseTTAFold joint inpainting učinkovite u redizajnu nativnih proteina s ciljem povećanja njihove topljivosti, stabilnosti i funkcionalnosti.47

* 1. Dizajn varijanti TEV proteaze s poboljšanom stabilnošću i katalitičkom aktivnošću

Kako bi se ispitao potencijal alata ProteinMPNN u stabilizaciji enzima, kao model uzeta je TEV proteaza. Ova proteaza ima široku primjenu u biotehnologiji, posebice za precizno uklanjanje afinitetnih privjesaka s rekombinantnih proteina putem specifičnog cijepanja prepoznatljivog slijeda ENLYFQ/S. Međutim, njezina upotreba je često ograničena zbog niske topljivosti, termostabilnosti i katalitičke aktivnosti, što zahtijeva dugotrajne inkubacije i često rezultira nepotpunim cijepanjem.

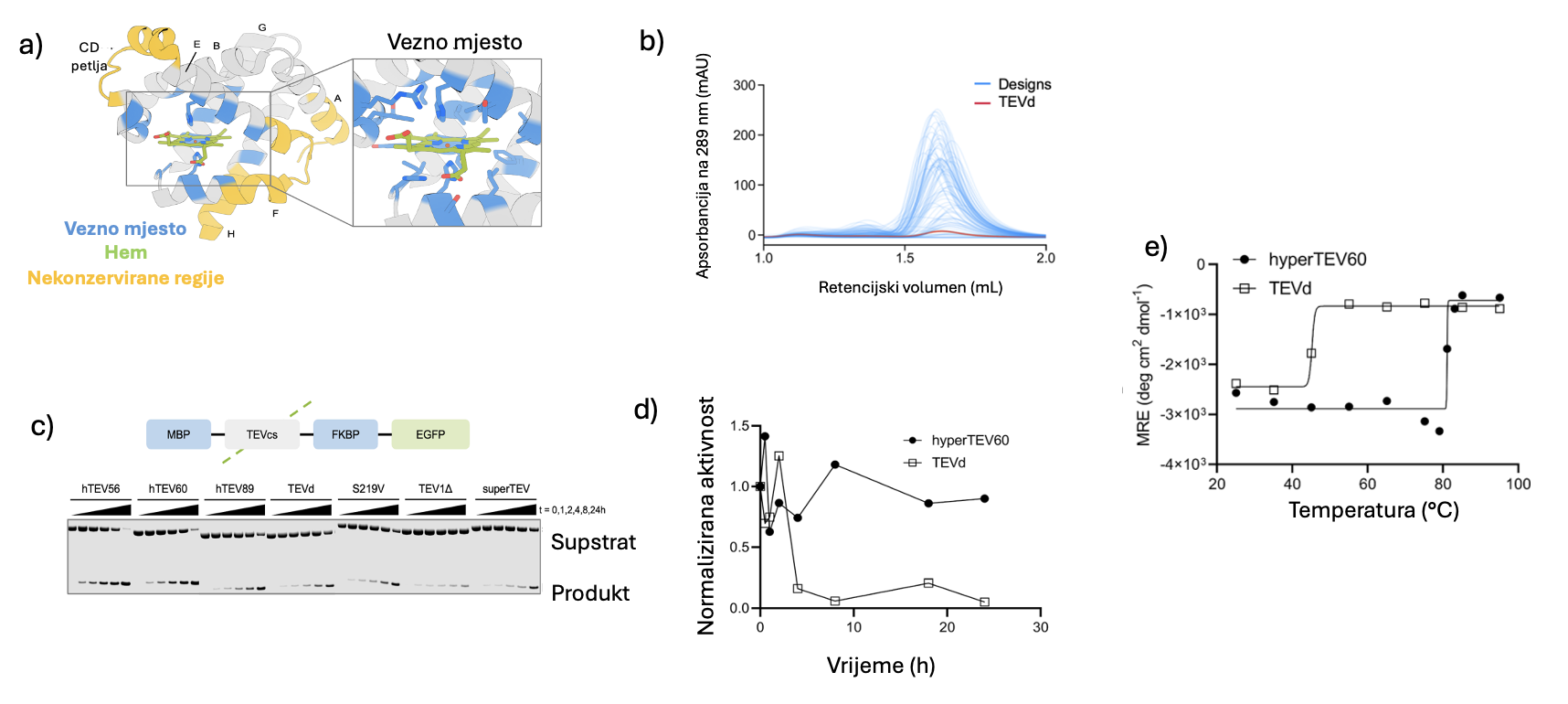
Korišten je oblik TEV proteaze (TEVd, PDB: 1LVM) otporan na autolizu, pri čemu su ostavljena nepromijenjena funkcionalno važna mjesta, uključujući aktivno mjesto i visoko konzervirane aminokiselinske ostatke definirane višestrukim poravnanjem sekvenci unutar TEV obitelji. Kroz četiri razine ograničenja (0%, 30%, 50% i 70% konzerviranih ostataka), generirane su ukupno 144 varijante proteaze koristeći ProteinMPNN. Sve sekvence predviđene su kao stabilne i slične izvornom obliku koristeći AlphaFold2 (pLDDT > 87.5), te su odabrane za eksperimentalnu validaciju.

Rekombinantni geni za sve dizajne eksprimirani su u *E. coli*, a proteini su pročišćeni koristeći IMAC i SEC metode. Više od 90% varijanti (134/144) uspješno se eksprimiralo u topljivom obliku, pri čemu je 129 dizajna imalo viši prinos od izvornog TEVd soja (prosječni prinos dizajna: 20.1 mg/L vs. TEVd: 1 mg/L).

Katalitička aktivnost analizirana je pomoću fluorogenog peptidnog supstrata konjugiranog s 7-amino-4-trifluormetilkumarinom. Od 144 dizajna, 64 su pokazala aktivnost prema supstratu. Najvišu aktivnost pokazali su dizajni u kojima je fiksirano 50% najkonzerviranijih ostataka. Tri visokoaktivne varijante, hyperTEV56, hyperTEV60 i hyperTEV89, podvrgnute su detaljnoj kinetičkoj analizi, gdje su pokazale i do 26 puta veću katalitičku učinkovitost (*k*cat/*K*m) u odnosu na TEVd.

Nadalje, dizajni su testirani i na fuzijskom proteinskom supstratu MBP-TEVcs-FKBP-EGFP. Varijante hyperTEV56 i hyperTEV60 su u roku od 4 sata ostvarile 50% cijepanje, dok je TEVd za isti učinak zahtijevao 24 sata. Osim toga, dizajnirani enzimi nadmašili su prethodno opisane varijante poput superTEV i TEV1Δ.

Korištenjem CD spektroskopije utvrđeno je da hyperTEV60 ima temperaturu mekšanja oko 84 °C, što je čak 40 °C više od TEVd te predstavlja najstabilniju TEV varijantu dosad opisanu. Također, nakon 4 sata inkubacije na 30°C, hyperTEV60 je zadržao 90% svoje aktivnosti, dok je TEVd zadržao svega 15%.



**Slika 3.** Poboljšanje topljivosti, termostabilnosti i katalitičke učinkovitosti TEV proteaze dizajnirane pomoću ProteinMPNN algoritma. (a) Kristalna struktura TEVd proteaze (PDB: 1LVM) korištena je kao početni model za dizajn. Obojena su fiksirana područja: katalitički ostatci (ružičasto), aktivno mjesto u blizini supstrata (plavo), te 50% najočuvanijih aminokiselina unutar TEV obitelji (žuto). (b) Kromatogrami (SEC) pokazuju monomernu prirodu i topljivu ekspresiju dizajniranih varijanti. (c) Dijagram TEV supstrata (ENLYFQ/S) i rezultati SDS-PAGE gela koji prikazuju kinetiku cijepanja fuzijskog proteina od strane dizajniranih TEV varijanti tijekom vremena. (d) Stabilnost enzimske aktivnosti na sobnoj temperaturi (30 °C) uspoređena između nativne TEVd proteaze i dizajniranih varijanti; prikazana je zadržana aktivnost u vremenu prije početka enzimskog testa. (e) Termostabilnost ispitanih varijanti određena kružnom dikroizmom (CD) – prikazana je promjena molarne ostatkovne eliptičnosti (MRE) u ovisnosti o temperaturi. Preuzeto i prilagođeno prema referenci 47.

Iako su koordinate katalitički važnih ostataka fiksirani tijekom dizajna, uočene su znatne promjene u *k*cat vrijednostima. Kako bi se razumio mehanizam ovog poboljšanja, provedene su mikrosekundne MD simulacije koje su otkrile rigidifikaciju petlji u dizajniranim varijantama, posebice u regiji 115–124, što je koreliralo s povećanom aktivnošću.

Ovi rezultati pokazuju da ProteinMPNN, čak bez eksplicitnog uključivanja funkcionalnih kriterija, može generirati proteine s poboljšanom topljivošću, stabilnošću i funkcionalnom učinkovitošću. Dizajn mutacija izvan aktivnog mjesta očito ima potencijal za alosterno poboljšanje enzimske aktivnosti putem stabilizacije funkcionalno kompetentnih konformacija.47

1. ZAKLJUČAK

Prirodna evolucija najčešće favorizira funkcionalnost naspram stabilnosti, zbog čega mnogi prirodni proteini pokazuju nisku topljivost, ograničenu termostabilnost i lošu ekspresiju u heterolognim sustavima, što u konačnici dovodi do niskog prinosa funkcionalnog proteina.

Unatoč prividnoj jednostavnosti procesa prekomjerne ekspresije rekombinantnih proteina, njegovo uspješno provođenje često zahtijeva sofisticirane strategije, temeljite optimizacije i multidisciplinarni pristup.

Modificirani sojevi izvedeni iz BL21(DE3), poput Rosetta, SHuffle, Lemo21, TatExpress i Marionette, danas omogućuju finu kontrolu ekspresije, poboljšanu topljivost proteina, pravilno formiranje disulfidnih mostova, bolju sekreciju te koekspresiju višestrukih komponenti u metaboličkim putevima. Paralelno s tim, razvoj sofisticiranih ekspresijskih vektora, poput Duet sustava, te implementacija afinitetetnih privjesaka i fuzijskih partnera osjetljivih na stimulus omogućili su olakšanu detekciju, pročišćavanje i stabilizaciju ciljnih proteina.

Naposljetku, razvoj računalnih alata, poput ProteinMPNN i AlphaFold2, otvorio je novo poglavlje u području proteinskog inženjeringa, omogućujući optimizaciju stabilnosti i funkcionalnosti proteina izravno na razini sekvence. Primjeri poput dizajna mioglobina i TEV proteaze demonstriraju nevjerojatan potencijal dubokog učenja u racionalnom redizajnu proteina.

Sinergija između molekularne biologije, računalne bioinformatike i kemijskog inženjerstva čini temelj budućeg napretka u području rekombinantne ekspresije proteina. Sustavi koji su danas na raspolaganju omogućuju dosad nezamislive razine kontrole nad sintezom proteina te pružaju snažan alat za temeljna istraživanja, biotehnološke primjene i razvoj novih terapijskih i industrijskih biomolekula.

1. Literaturni izvori
2. G.L. Rosano, E.A. Ceccarelli, Front. *Microbiol.* **5** (2014) 172.
3. G.L. Rosano, E.S. Morales, E.A. Ceccarelli, *Protein Sci.* **8** (2019) 1412–1422.
4. J.L. Adrio, A.L. Demain, *Bioeng. Bugs* **1(2)** (2010) 116–131.
5. A.L. Demain, P. Vaishnav, *Biotechnol. Adv.* **27(3)** (2009) 297–306.
6. S. Sahdev, S.K. Khattar, K.S. Saini, *Mol. Cell. Biochem.* **307** (2008) 249–264.
7. G. Sezonov, D. Joseleau-Petit, R. D’Ari, *J. Bacteriol.* **189** (2007) 8746–8749.
8. W.E. Bentley, N. Mirjalili, D.C. Andersen, R.H. Davis, D.S. Kompala, *Biotechnol. Bioeng.* **35** (1990) 668–681.
9. P. Daegelen, F.W. Studier, R.E. Lenski, S. Cure, J.F. Kim, *J. Mol. Biol.* **394** (2009) 634–643.
10. B.A. Moffatt, F.W. Studier, *Cell* **49** (1987) 221–227.
11. F.W. Studier, B.A. Moffatt, *J. Mol. Biol*. **189** (1986) 113–130.
12. N.M. Stano, S.S. Patel, *J. Biol. Chem.* **279** (2004) 16136–16143.
13. J.W. Dubendorff, F.W. Studier, *J. Mol. Biol.* **219** (1991) 45–59.
14. H. Jeong et al., *J. Mol. Biol.* **394** (2009) 644–652.
15. F.W. Studier et al., *J. Mol. Biol.* **394** (2009) 653–680.
16. S. Kim et al., *Nucleic Acids Res.* **45** (2017) 5285–5293.
17. S.H. Yoon et al., in: S.Y. Lee (Ed.), *Systems Biology and Biotechnology of Escherichia coli, Springer, Dordrecht,* 2009, pp. 1–17.
18. Y.J. Son et al., *Microb. Cell Fact*. **10** (2011) 52.
19. S. Leone et al., *Microb. Cell Fact.* **14** (2015) 106.
20. H. Wang et al., *PLoS One* **9** (2014) e112777.
21. M.P. DeLisa, D. Tullman, G. Georgiou, *Proc. Natl. Acad. Sci. U.S.A.* **100** (2003) 6115–6120.
22. D.F. Browning et al., *Biotechnol. Bioeng.* **114** (2017) 2828–2836.
23. F.W. Studier, *J. Mol. Biol.* **219** (1991) 37–44.
24. L.G. Horga et al., *Microb. Cell Fact.* **17** (2018) 199.
25. R. Morra et al., *Nucleic Acids Res.* **44** (2016) e21.
26. C.J. Robinson et al., *J. Am. Chem. Soc.* **136** (2014) 10615–10624.
27. Z. Lipinszki et al., *ACS Synth. Biol.* **7** (2018) 2656–2664.
28. A.J. Meyer et al., *Nat. Chem. Biol.* **15** (2019) 196–204.
29. N. Gul et al., *J. Mol. Biol.* **426** (2014) 136–149.
30. A.E. Silverstone, R.R. Arditti, B. Magasanik, *Proc. Natl. Acad. Sci. U.S.A*. **66** (1970) 773–779.
31. G. del Solar, M. Espinosa, *Mol. Microbiol.* **37** (2000) 492–500.
32. B. Müller-Hill, L. Crapo, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **59** (1968) 1259–1264.
33. W.V. Shaw, *CRC Crit. Rev. Biochem.* **14** (1983) 1–46.
34. M.C. Roberts, *FEMS Microbiol. Rev.* **19** (1996) 1–24.
35. J. Nilsson et al., *Protein Expr. Purif.* **11** (1997) 1–16.
36. K. Terpe, Appl. *Microbiol. Biotechnol*. **60** (2003) 523–533.
37. F.W. Studier, *Protein Expr. Purif.* **41** (2005) 207–234. https://doi.org/10.1016/j.pep.2005.01.016
38. L. Briand et al., *Biotechnol. Bioeng.* **114** (2017) 2828–2836. https://doi.org/10.1002/bit.26497
39. J. Panula-Perälä et al., *J. Biotechnol.* **150** (2010) 61–69. https://doi.org/10.1016/j.jbiotec.2010.08.003
40. M. Pilarek et al., *Appl. Microbiol. Biotechnol.* **104** (2020) 1375–1386. https://doi.org/10.1007/s00253-019-10288-4
41. J. Hemmerich et al., *Biotechnol. J.* **13** (2018) 1700141. https://doi.org/10.1002/biot.201700141
42. J. Hemmerich et al., *J. Biotechnol.* **283** (2018) 1–10. https://doi.org/10.1016/j.jbiotec.2018.06.348
43. M. Jahn et al., *Eng. Life Sci.* **17** (2017) 960–972. https://doi.org/10.1002/elsc.201700081
44. A. Baumschlager, S.K. Aoki, M. Khammash, *Nat. Chem. Biol.* **13** (2017) 1020–1025. https://doi.org/10.1038/nchembio.2473
45. A. Madani et al., *Nat. Biotechnol.* **41** (2023) 1099–1106.
46. J. Dauparas et al., *Science* **378** (2022) 49–56.
47. B.I.M. Wicky et al., *Science* **378** (2022) 56–61.
48. K.H. Sumida et al., *J. Am. Chem. Soc.* **146(3)** (2024) 2054–2061.
49. J. Jumper et al., *Nature* **596** (2021) 583–589.