

UNIVERZA V LJUBLJANI
FAKULTETA ZA KEMIJO IN KEMIJSKO TEHNOLOGIJO

DIPLOMSKO DELO

Aljaž Simonič

Ljubljana, 2023

UNIVERZA V LJUBLJANI
FAKULTETA ZA KEMIJO IN KEMIJSKO TEHNOLOGIJO

UNIVERZITETNI ŠTUDIJSKI PROGRAM 1. STOPNJE
BIOKEMIJA

**Molekulska dinamika proteina EpCAM iz
različnih vrst in njegova interakcija s paralognim
proteinom Trop2**

DIPLOMSKO DELO

Aljaž Simonič

MENTOR: doc. dr. Miha Pavšič

Ljubljana, 2023

IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani Aljaž Simonič sem avtor diplomskega dela z naslovom: *Molekulska dinamika proteina EpCAM iz različnih vrst in njegova interakcija s paralognim proteinom Trop2.*

S svojim podpisom zagotavljam, da:

- je diplomsko delo rezultat mojega raziskovalnega dela pod mentorstvom doc. dr. Mihe Pavšiča;
- sem poskrbel, da so dela in mnenja drugih avtorjev, ki jih uporabljam v predloženem diplomskem delu, navedena oziroma citirana v skladu z navodili;
- se zavedam, da je plagiatorstvo, v katerem so tuje misli oziroma ideje predstavljene kot moje lastne, kaznivo po zakonu (Zakon o avtorski in sorodnih pravicah – uradno prečiščeno besedilo (ZASP-UPB3) (Ur. list RS, št. 16/2007);
- sem poskrbel za slovnično in oblikovno korektnost diplomskega dela;
- je elektronska oblika diplomskega dela identična tiskani obliki diplomskega dela.

V Ljubljani, 28. 6. 2023

Podpis avtorja:

Računalniški del diplomske naloge je bil izveden na Arnesovi gruči. Zahvaljujem se dr. Janu Joni Javoršku in mag. Barbari Krašovec iz Centra za mrežno infrastrukturo Inštituta Jožef Stefan za ureditev dostopa in za tehnično pomoč. Najlepša hvala tudi vsem ostalim, ki vzdržujejo Arnesovo gručo in širše omrežje SLING.

Laboratorijski del diplomske naloge je bil izveden na Katedri za biokemijo, Fakulteta za kemijo in kemijsko tehnologijo. Za možnost tega se zahvaljujem katedri in fakulteti. Hvala vsemu osebju katedre, ki skrbite, da so laboratoriju urejeni, založeni in da je delo v njih mogoče.

Najlepša hvala Joštu Hočevarju, ki je v okviru svojega preteklega dela pripravil himerni protein EpEX-sfGFP in mi pomagal med laboratorijskim delom.

Hvala tudi Neži Koritnik in Luki Kavčiču z Odseka za molekularno biologijo in nanobiotehnologijo Kemijskega inštituta, ki sta me prvič uvajala v raziskovalno delo v laboratoriju. Izkušnje, ki sem jih pridobil pod vajinim vodstvom, so mi pri laboratorijskem delu močno pomagale.

Najlepša hvala doc. dr. Gregorju Gunčarju za strokovni pregled dela in članstvo v komisiji za zagovor.

Seveda se zahvaljujem tudi svojemu mentorju, doc. dr. Mihi Pavšiču, za vodstvo in pomoč pri delu, za izjemno odzivnost in seveda za vso spodbudo.

Zahvaljujem se svoji stari mami, Mariji Kušar, za jezikovni pregled diplomskega dela.

Resnično hvala vsem prijateljem in družini, da mi lepšate dneve.

Na koncu pa še Janezu – hvala za ves čas, ki sva ga lani preživel na vrtu, da te je zanimalo, kaj počnem za diplomsko, in za tvoje navdušenje nad naravo, ki si ga delil z mano in vsemi drugimi okoli sebe. Počivaj v miru.

Molekulska dinamika proteina EpCAM iz različnih vrst in njegova interakcija s paralognim proteinom Trop2

Povzetek

EpCAM in Trop2 sta paralogna homodimerna transmembranska proteina, ki sta bila odkrita kot antigen na površini celic karcinoma debelega črevesja in kot antigen na površini celic trofoblasta. EpCAM se pojavlja pri vseh vretenčarjih, Trop2 pa je nastal z retrotranspozicijo *EPCAM* in je prisoten pri plazilcih, ptičih in sesalcih. Izražena sta v epitelijskih celicah in sta pomembna za razvoj epitelijskih tkiv, regulirata celično adhezijo in sodeljujeta pri proliferativni signalizaciji. Pogosto sta močno izražena v karcinomih. S programom AlphaFold Multimer smo predvideli strukture EpCAM iz hišne miši, domače kokoši, navadne krempljarke in navadne cebrice ter simulirali dinamiko njihovih zunajceličnih delov, skupaj s simulacijo dinamike kristalne strukture zunajceličnega dela človeškega EpCAM. Prav tako smo s programom AlphaFold Multimer predvideli strukture človeškega homodimera EpCAM, homodimera Trop2 in heterodimera EpCAM:Trop2, simulirali njihove zunajcelične dele skupaj s kristalnima strukturama zunajceličnega dela homodimerov EpCAM in Trop2 ter na podlagi teh simulacij z metodo MM/GBSA izračunali spremembo proste energije pri tvorbi dimera. Ugotovili smo, da je N-končna domena žabjega EpCAM fleksibilnejša od ostalih. Interakcijska površina med podenotama je bila v primerjavi z ostalimi manjša v EpCAM iz navadne krempljarke, medtem ko ima tiroglobulinska zanka v EpCAM iz navadne zebrice manjšo interakcijsko površino z nasprotno podenoto. Izračuni proste energije so napovedali najmočnejšo interakcijo med podenotama v zunajceličnem delu EpCAM iz domače kokoši, najšibkejšo pa v EpCAM iz navadne cebrice. Hidrofobni vezavni žep v EpCAM iz navadne cebrice je manjši od ostalih, v EpCAM iz navadne krempljarke pa je rigidnejši – oboje bi lahko vplivalo na vezavne lastnosti žepa. Ugotovili smo tudi, da sta podenoti v heterodimeru povezani močneje kot v homodimeru EpCAM in podobno močno kot v homodimeru Trop2 in da bi posledično tvorba heterodimera lahko bila termodinamsko ugodna. V laboratoriju z nativno elektroforezo in s prečnim povezovanjem heterodimera nismo opazili, vendar zaradi omejitev uporabljenega pristopa na podlagi laboratorijskih rezultatov ne moremo zagotovo izključiti obstoja heterodimera.

Ključne besede: EpCAM, Trop2, molekulska dinamika, proteinske interakcije, prosta energija

Molecular dynamics of EpCAM from different species and its interaction with the paralogous protein Trop2

Abstract

EpCAM and Trop2 are paralogous homodimeric transmembrane proteins that were first discovered as an antigen on the surface of colorectal cancer cells and an antigen on the surface of trophoblast cells, respectively. EpCAM is present in all vertebrates, while Trop2 arose by retrotransposition of *EPCAM* and is present in reptiles, birds and mammals. They are expressed on epithelial cells and participate in epithelial development, regulate cell adhesion and participate in proliferative signalling. Using AlphaFold Multimer, we predicted the structures of EpCAM dimers from house mouse, chicken, African clawed frog and zebrafish and performed long molecular dynamics simulations of their extracellular parts, together with the simulation of the crystal structure of the human extracellular part of EpCAM. Using AlphaFold Multimer we also predicted the structures of human EpCAM homodimer, Trop2 homodimer and EpCAM:Trop2 heterodimer. We performed short molecular dynamics simulations of extracellular parts of these structures, alongside simulations of crystal structures of extracellular parts of human EpCAM and Trop2 homodimers, and calculated the free energy difference upon dimer formation for these structures using MM/GBSA. We determined that the N-terminal domain of African clawed frog EpCAM exhibits greater flexibility compared to EpCAMs from other species. Dimer interface was smaller in African clawed frog EpCAM compared to others, while thyroglobulin loop in zebrafish EpCAM formed smallest interface with the opposite subunit. Interaction between subunits is the strongest in chicken and the weakest in zebrafish. Hydrophobic binding pocket in zebrafish EpCAM is smaller than others, while in African clawed frog it is more rigid – both differences could result in different binding properties. We also discovered that the subunits in the putative heterodimer are more strongly bound than in EpCAM homodimer and approximately as strong as in Trop2 homodimer. Consequently, heterodimer formation could be thermodynamically favourable. In the laboratory, we didn't observe the heterodimer using native electrophoresis and crosslinking, but we couldn't conclusively exclude the possibility of heterodimer formation due to the limitations of our approach.

Keywords: EpCAM, Trop2, molecular dynamics, protein-protein interactions, free energy

Kazalo

<i>1 Uvod.....</i>	1
1.1 Evolucija.....	1
1.2 Struktura.....	2
1.3 Biološka funkcija.....	3
1.4 Uporaba v medicini.....	5
<i>2 Namen dela in hipotezi.....</i>	7
2.1 Namen dela.....	7
2.2 Hipotezi.....	7
2.2.1 EpCAM iz različnih vrst.....	7
2.2.2 Heterodimer EpCAM:Trop2.....	7
<i>3 Materiali in metode.....</i>	9
3.1 EpCAM iz različnih vrst.....	9
3.1.1 Poravnava zaporedij.....	9
3.1.2 Strukturno modeliranje in simulacije molekulske dinamike.....	9
3.1.3 Analiza simulacij.....	10
3.1.3.1 Analiza vezavne energije z metodo Nwat-MM/GBSA.....	11
3.1.3.2 Izračun prostornine hidrofobnega žepa.....	12
3.2 Heterodimer EpCAM:Trop2.....	14
3.2.1 Strukturno modeliranje in simulacije molekulske dinamike.....	14
3.2.2 Analiza simulacij.....	14
3.2.3 Laboratorijsko določanje obstoja heterodimera.....	15
3.2.3.1 Priprava vzorcev.....	16
3.2.3.2 Nativna elektroforeza.....	17
3.2.3.3 Prečno povezovanje.....	17
<i>4 Rezultati in razprava.....</i>	19
4.1 EpCAM iz različnih vrst.....	19
4.1.1 Poravnava.....	19
4.1.2 Modeli struktur.....	19
4.1.3 RMSD in fleksibilnost proteina.....	21
4.1.4 Interakcijska površina med podenotama dimera.....	23
4.1.5 Predvidene vezavne energije.....	24
4.1.6 Prostornina hidrofobnega žepa.....	26
4.2 Heterodimer EpCAM:Trop2.....	28
4.2.1 Predvidene strukture in potek simulacij.....	28
4.2.2 Predvidene vezavne energije.....	29
4.2.3 Laboratorijsko določanje obstoja heterodimera.....	32
4.2.4 Trop2 pri pH 4,5 reverzibilno denaturira.....	32
4.2.4.1 Disociacija z dodatkom soli.....	33
4.2.4.2 Analiza interakcij.....	34
<i>5 Zaključek.....</i>	37
5.1 EpCAM iz različnih vrst.....	37
5.2 Heterodimer EpCAM:Trop2.....	37
<i>6 Literatura.....</i>	39
<i>7 Priloge.....</i>	45

Seznam uporabljenih kratic in simbolov

EpEX	zunajcelični del EpCAM
EpIC	znotrajcelični del EpCAM
ER	endoplazemski retikulum
G	Gibbsova prosta energija
GB	posplošen Born (angl. <i>Generalized Born</i>)
IZ	interval zaupanja
MM/GBSA	molekulska mehanika s posplošeno Bornovo solvacijo in površino (angl. <i>Molecular Mechanics / Generalized Born Surface Area</i>)
MD	molekulska dinamika
PDB	podatkovna baza proteinov (angl. <i>Protein Data Bank</i>)
RIP	regulirana intramembranska proteoliza
RMSD	kvadratno-korenjen povprečen kvadriran odmik
RMSF	kvadratno-korenjena povprečna kvadrirana fluktuacija
SASA	topilu dostopna površina (angl. <i>Solvent Accessible Surface Area</i>)
sfGFP	zeleni fluorescenčni protein, ki se zelo učinkovito zvija (angl. <i>Superfolder Green Fluorescent Protein</i>)
Trop2IC	znotrajcelični del Trop2

1 Uvod

V družini genov TACSTD se nahajata dva transmembranski glikoprotein kodirajoča gena: gen *TACSTD1 (EPCAM, TROP1)*, ki kodira protein EpCAM oziroma Trop1 [1] ter gen *TACSTD2 (TROP2)*, ki kodira protein Trop2 (EpCAM2) [2, 3]. Oba proteina sta evolucijsko ohranjena in sta si medsebojno precej podobna – človeški EpCAM in Trop2 sta si na nivoju aminokislinskega zaporedja 49 % identična in 67 % podobna [3].

EpCAM je bil odkrit leta 1979 kot antigen na površini človeških celic raka debelega črevesja z uporabo monoklonskega protitelesa 1083-17-1A, dobljenega z imunizacijo miši s celicami raka debelega črevesa [4]. Funkcijsko je bil najprej opisan kot homofilna celična adhezijska molekula in bil po tem tudi poimenovan [5], kasneje pa so bile odkrite njegove druge funkcije v različnih signalnih poteh in pri uravnavanju tesnih stikov, vloga homofilne adhezijske molekule pa se je izkazala za vprašljivo [1, 6–8].

Trop2 je bil odkrit s podobnimi metodami kot EpCAM dve leti kasneje, leta 1981, kot površinski antigen celic trofoblasta, v okviru študije, ki je zaznala tudi EpCAM in slednjega poimenovala kot Trop1 [2, 9].

1.1 Evolucija

TACSTD1 (EPCAM) je prisoten le pri vretenčarjih [1], od obloustk do sesalcev. Ima 9 eksonov, od katerih prvi kodira 5'-neprevedeno regijo in signalni peptid, drugi N-končno domeno, tretji tiroglobulinsko tip-1 domeno, eksoni 4–6 C-končno domeno, ekson 7 transmembransko vijačnico, eksona 8 in 9 pa EpIC in 3'-neprevedeno regijo [6]. Gen je najverjetneje nastal s premešanjem eksonov – tiroglobulinska tip-1 domena je prisotna tudi v mnogih drugih proteinih in te domene so homologne tiroglobulinski domeni v EpCAM [10].

Tiroglobulinske tip-1 domene so prisotne le v živalih. Pogosto služijo kot inhibitorji proteaz (ne v EpCAM – le tiroglobulinske domene s kratko tiroglobulinsko zanko se lahko vežejo v aktivno mesto proteaz [8]), lahko pa tudi v vezavi proteinov brez proteazne aktivnosti ali, v primeru tiroglobulina, kot vir jodiranih hormonov. Večinoma jih najdemo v zunajceličnih proteinih [11].

Medtem ko zaporedje C-končne domene ni opazno podobno kateremukoli drugemu zaporedju izven družine genov *TACSTD*, se strukturno ujema s SEA domenami [8, 12]. Te domene so večinoma prisotne v zunajceličnih proteinih (izjema je nukleoporin 54), pogosto pred transmembranskim odsekom, in so vpletene v proteolizo. Najpogosteje so v živalih in njihovih bližnjih sorodnikih, kot so bičkarji ovratničarji (*Choanoflagellata*), najdemo jih pa tudi v drugih evkariontih [12].

Chen in Bell sta s simulacijami molekulske dinamike monomerov EpEX iz različnih vrst ugotovila, da se tiroglobulinska zanka v EpEX iz *Danio rerio*, ki se v zaporedju ne ujema s tiroglobulinskimi zankami EpCAM iz ostalih skupin vretenčarjev, precej

močneje veže na mesto v C-končni domeni, kar je lastnost zaporedja same zanke, ne razlik v preostalih delih proteina [13].

Gen *TACSTD2* (*TROP2*) nima intronov in njegov promotor ni soroden promotorju *TACSTD1*, prisoten pa je le pri plazilcih, ptičih in sesalcih. Najverjetnejše je nastal z retrotranspozicijo *TACSTD1*, pri kateri je bil vstavljen navzdol od nepovezanega promotorja, pred nekaj več kot 300 milijoni let (pred ločitvijo amniotov na ptiče, plazilce in sesalce) [2, 10]. Pri človeku je Trop2 49 % identičen in 67 % podoben EpCAM na nivoju aminokislinskega zaporedja [3].

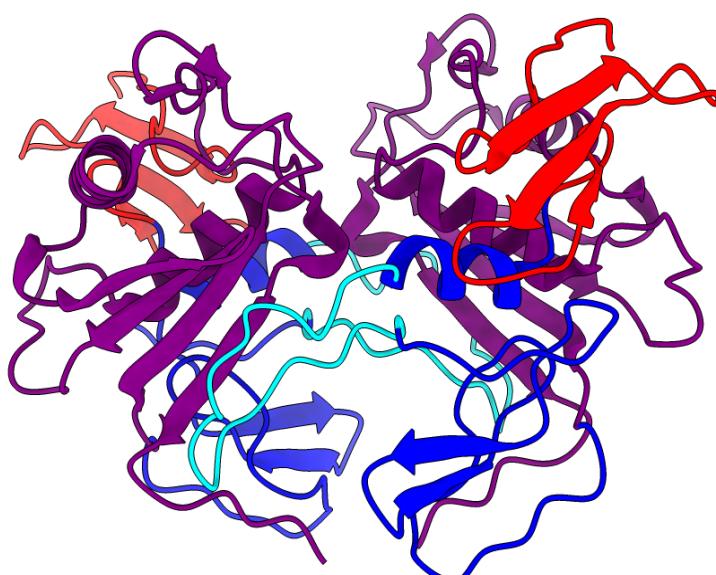
1.2 Struktura

EpCAM in Trop2 sta homodimerna transmembranska proteina tipa I, torej z N-končnim delom izven celice. Večji del proteina predstavlja zunajcelični del, sledita mu enojna transmembranska α -vijačnica in kratek znotrajcelični del [8, 14, 15]. Človeška EpCAM in Trop2 sta homodimera. Za dimerizacijo je zadosten že njun zunajcelični del [16, 17], grobozrnate MD simulacije pa so nakazale, da dimerizirata tudi njuni transmembranski vijačnici [14, 18].

N-končni glutaminski ostanek (po odstranitvi signalnega peptida) EpCAM je modificiran v piroglutamatni ostanek. Je tudi heterogeno N-glikoziliran, v človeku na mestih N74, N111 in N198 [14]. Tudi v Trop2 je N-končni glutamin modificiran v piroglutamat. Ima štiri N-glikozilirana mesta – N33, N120, N168 in N208 (N111 in

N198 v EpCAM sta ohranjena v Trop2 kot N120 in N208) [2, 15]. S303 in S322 v Trop2IC lahko fosforilira protein kinaza C [3].

Objavljena sta dva eksperimentalna modela strukture neglikozilirane mutante človeškega EpEX, oba dobljena z rentgensko kristalografijo – model s PDB kodo 4MZV, ki so ga leta 2014 objavili Pavšič in sod. [14], in model EpEX v kompleksu z dvema enoverižnima fragmentoma variabilne regije (scFv) anti-EpEX



Slika 1: EpEX (PDB ID 4MZV) z različno pobarvanimi domenami - N-končna domena rdeča, tiroglobulinska tip-1 domena modra (in tiroglobulinska zanka v njej svetlo modra) in C-končna domena vijolična. Slika je bila narejena s programom UCSF ChimeraX 1.5 [31]

protitelesa, ki ima PDB kodo 6I07 in so ga leta 2019 objavili Casaleotto in sod. [19]. Kristalna struktura zunajceličenga dela Trop2 je bila rešena leta 2021 [15].

EpEX in Trop2 tvorijo tri domene, kot je na primeru EpEX prikazano na sliki 1. Prva je majhna N-končna domena (ostanki 24 do 62) brez izrazitega hidrofobnega jedra, ki jo stabilizirajo tri disulfidne vezi. Je strukturno edinstvena in prisotna le v družini TACSTD. Sledita ji tiroglobulinska tip-1A domena (ostanki 63 do 138) in C-končna SEA domena (ostanki 139-265), ki sta pomembni za signalizacijo preko RIP in dimerizacijo - tiroglobulinska domena s tiroglobulinsko zanko veže β -ploskev SEA domene nasprotne podenote [8]. SEA domena vsebuje tudi hidrofoben žep, v katerega se je v kristalu, uporabljenem za določitev strukture EpCAM, vezal detergent decil- β -D-maltopiranozid [14]. V EpCAM ima predvidoma ta žep večjo prostornino kot v Trop2 [20].

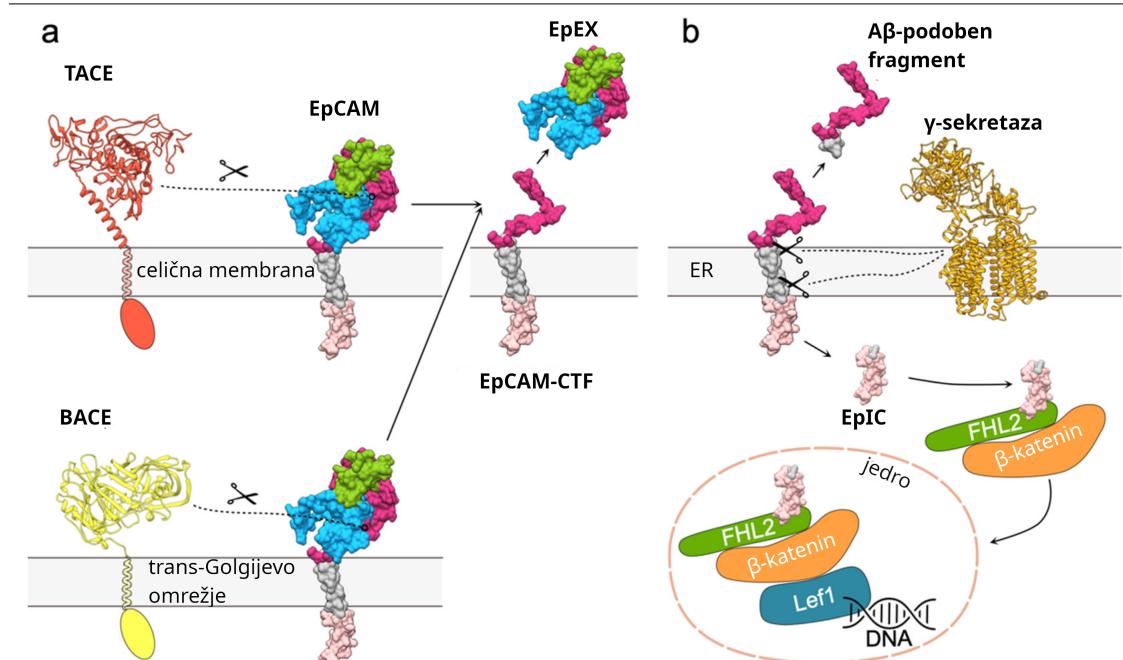
Medtem ko EpIC nima nagnjenosti k tvorbi sekundarne strukture [8], vsebuje Trop2IC α -vijačnico in kratko 3_{10} -vijačnico, njegova konformacija pa se spremeni ob fosforilaciji na S303 [18].

1.3 Biološka funkcija

EpCAM in Trop2 sta izražena v embrionalnih tkivih, placenti in na diferenciranih epitelijskih celicah, kjer se nahajata na bazolateralnem delu [1, 2]. V zdravih razvitih tkivih se pogosto, a ne zmeraj, izražata na istih celicah [21], kar bi zaradi njune podobnosti lahko vodilo v tvorbo heterodimera [1].

EpCAM je bil originalno opisan kot homofilna celična adhezijska molekula, vendar se je skozi čas nabralo mnogo s tem nezdružljivih rezultatov [1, 6]. Gaber in sod. so leta 2018 ugotovili, da EpCAM ne tvori oligomernih struktur, večjih od dimera, in tako ni homofilna adhezijska molekula [7].

EpCAM in Trop2 preko transmembranske vijačnice vežeta klavdin-1 in klavdin-7 in tako regulirata tesne stike, njuno izbitje pa destabilizira ta klavdina in tesne stike [22]. Vežeta tudi številne integrine [1]. Močno izražanje EpCAM ošibi adherentne stike [6].



Slika 2: RIP EpCAM. a) Cepitev EpCAM s TACE ali BACE in sprostitev EpEX, ostane EpCAM-CTF (C-končni fragment). b) γ -sekretaza v ER cepi EpCAM-CTF, EpIC pa sproži tvorbo signalnega kompleksa EpIC–FHL2– β -katenin–Lef1. Prevedeno iz Gaber in sod. [8].

Kot je prikazano na sliki 2 za EpCAM, sta EpCAM in Trop2 tarča več proteaz v procesu regulirane intramembranske proteolize (RIP) in ta vodi do tvorbe signalnega kompleksa, ki aktivira številne celično delitve spodbujajoče gene [2, 6, 8]. RIP je relevantna predvsem v rakavih celicah, v zdravih tkivih pa ne oz. precej manj [6].

Transmembranska proteaza matriptaza cepi EpCAM in Trop2 na istem mestu v tiroglobulinski zanki. Cepitveno mesto je v dimeru zakopano v interakcijski površini; po cepitvi pa kljub temu, da proteina ostaneta povezana z disulfidnimi vezmi, več ne tvorita dimera [8, 15], prav tako pa več ne moreta stabilizirati klavdinov [22].

Transmembranska proteaza TACE oz. ADAM17 cepi EpCAM in Trop2 na različnih mestih v SEA domeni [15]. V EpCAM je cepitveno mesto zakopano v interakcijski površini [8] in mutanta, ki ne dimerizira, se cepi mnogo hitreje, mutanta z disulfidno vezjo med podenotama se ne cepi oziroma se cepi zelo počasi [23]. Nasprotno je cepitveno mesto v Trop2 izpostavljeno na površini dimera [15]. Blizu cepitvenega mesta proteaze TACE EpCAM cepi tudi β -sekretaza 1 (BACE) [8]. Ta je aktivna v kislih veziklih in ima optimalno aktivnost pri pH 4 [6], kjer je EpCAM monomer [6, 16]. Posledično je dejstvo, da je BACE cepitveno mesto v dimeru zakopano, verjetno nepomembno [6].

Cepitvi s TACE ali BACE sprostita zunajcelični del, nato pa γ -sekretaza cepi znotraj transmembranske vijačnice preostanka in sprosti EpIC oz. Trop2IC [6, 8]. Ta vežeta FHL2, ki v signalnem kompleksu z β -kateninom in Lef1 sproži transkripcijo številnih

genov, ki kodirajo za proliferacijo spodbujajoče proteine, kot so ciklini A2, D1 in E, c-myc in drugi [8].

Molekularne funkcije EpCAM in Trop2 so podobne. Szabo in sodelavci so ugotovili, da izbitje EpCAM v miših vodi do motenj v razvoju in funkciji tkiv črevesja (kar v ljudeh ustreza genski bolezni kongenitalna šopasta enteropatija (angl. *congenital tufting enteropathy*) [1, 21]), manjše abnormalnosti pa so tudi v tkivih želodca, trebušne slinavke, ledvic in lasnih mešičkov. Izbitje Trop2 vodi le do nekoliko bolj prepustne kože brez drugih opaženih motenj (v ljudeh in miših sicer inaktivirajoče mutacije Trop2 povzročijo tudi gensko bolezen gelatinozna kaplji-podobna roženična distrofija, ki privede do slepote [2, 21]), izbitje obeh pa močno poslabša motnje razvoja tkiv, ki jih prizadene izguba EpCAM (z izjemo črevesnega epitelija, kjer se Trop2 ne izraža in ostane enako kot v živalih, ki jim je bil izbit le EpCAM). Očitno lahko Trop2 v teh tkivih delno kompenzira izgubo EpCAM. EpCAM in Trop2 sta tudi enako sposobna regulirati stabilnost in lokalizacijo klavdina-7 [21]. Nista pa v miših povsem zamenljiva – v miših, ki so gensko spremenjene tako, da izražajo Trop2 v črevesnih epitelijih, izbitje EpCAM še vedno vodi do motenega razvoja črevesja, čeprav so motnje precej blažje [24].

Ker sta EpCAM in Trop2 pogosto prisotna v istih celicah in sta si v zaporedju in funkciji podobna, sta Fagotto in Aslemarz postavila hipotezo, da tudi tvorita heterodimer EpCAM:Trop2 [1]. Paralogi neredko tvorijo heterodimere, če so prisotni v istih celicah. Primer je heterodimer α -aktinina 1 in 4, ki je v mnogih celičnih linijah najpogostejša oblika α -aktinina [25].

1.4 Uporaba v medicini

EpCAM in Trop2 sta pogosto močno izražena v karcinomih. Močno izražanje kateregakoli od teh dveh proteinov je v številnih oblikah raka povezano s poslabšano prognozo, obstajajo pa izjeme, kjer imata nasproten učinek [3, 6].

Protitelesa proti EpCAM se uporabljajo za izolacijo krožečih tumorskih celic iz krvi. Količina krožečih tumorskih celic je negativno povezana s prognozo [6].

Za namen zdravljenja raka je bilo razvitih in uporabljenih več protiteles proti EpCAM:

- mišje monoklonsko protitelo 1083-17-1A, s katerim je bil odkrit, je bilo v Nemčiji uporabljano pod imenom edrecolomab za zdravljenje raka debelega črevesa [6, 26],
- bispecifično protitelo catumaxomab, ki veže EpCAM in CD3, je bilo v Evropski uniji odobreno za lajšanje z rakom povezanih zapletov [6],
- s kelatorjem konjugirano protitelo nofetumomab merpentan za diagnostično slikanje s ^{99m}Tc [26–28].

*ALJAŽ SIMONIČ: MOLEKULSKA DINAMIKA PROTEINA EPCAM IZ RAZLIČNIH VRST IN NJEGOVA
INTERAKCIJA S PARALOGNIM PROTEINOM TROP2*

Nobeno od navedenih protiteles več ni v uporabi – edrecolomab se je izkazal kot manj učinkovit od kemoterapevtika 5-fluorouracila, catumaxomab pa je bil umaknjen iz prodaje, ker njegova proizvodnja in trženje nista bila finančno smiselna [6, 26].

Trop2 je tarča zdravila sacituzumab govitekan, konjugata protitelesa in zdravila, ki se uporablja za zdravljenje trojno negativnega raka dojke [29].

2 Namen dela in hipotezi

2.1 Namen dela

Namen dela je bil preučiti spremembe strukture, dinamike in posledično funkcije EpCAM skozi evolucijo s strukturnim modeliranjem in simulacijo EpEX iz pet različnih vrst, ki pripadajo štirim različnim razredom vretenčarjev:

- *Homo sapiens* (človek),
- *Mus musculus* (hišna miš),
- *Gallus gallus* (domača kokoš),
- *Xenopus laevis* (navadna kremljarka),
- *Danio rerio* (navadna cebrica).

Tako bi dopolnili delo Chen in Bell [13], v katerem so bili simulirani monomeri.

Namen dela je bil tudi s strukturnim modeliranjem in računalniškimi simulacijami predvideti termodinamsko ugodnost nastanka človeškega heterodimera EpCAM:Trop2 ter dobljene napovedi eksperimentalno preveriti.

2.2 Hipotezi

2.2.1 EpCAM iz različnih vrst

Fleksibilnost tiroglobulinske zanke bo največja pri EpCAM iz *Danio rerio*, saj je ta daljša in se v zaporedju ne ujema s tiroglobulinskimi zankami iz EpCAM iz drugih razredov vretenčarjev.

2.2.2 Heterodimer EpCAM:Trop2

Nastanek človeškega heterodimera EpCAM:Trop2 je ugoden in ga bomo opazili. EpCAM in Trop2 sta paralogna dimerna proteina, pogosto prisotna v istih celicah, in v takšnih primerih neredko pride do heterodimerizacije.

ALJAŽ SIMONIČ: MOLEKULSKA DINAMIKA PROTEINA EPCAM IZ RAZLIČNIH VRST IN NJEGOVA
INTERAKCIJA S PARALOGNIM PROTEINOM TROP2

3 Materiali in metode

Strukture, predvidene z AlphaFoldom, in urejene strukture, primeri konfiguracijskih datotek za simulacije ter skripte za analizo so dostopne na strani Github, povezava je v poglavju Priloge.

Za urejanje struktur in določene analize smo uporabili program UCSF Chimera, ki ga razvija Resource for Biocomputing, Visualization, and Informatics Univerze v Kaliforniji San Francisco s finančno podporo NIH P41-GM103311 [30].

Za pripravo slik struktur smo uporabili program UCSF ChimeraX, ki ga razvija Resource for Biocomputing, Visualization, and Informatics Univerze v Kaliforniji San Francisco s finančno podporo NIH R01-GM129325 in Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases [31].

3.1 EpCAM iz različnih vrst

3.1.1 Poravnava zaporedij

Iz baze podatkov Uniprot [32] smo izbrali zaporedja EpCAM z ID P16422 (*Homo sapiens*), Q99JW5 (*Mus musculus*), Q5F381 (*Gallus gallus*), Q8AWG0 (*Xenopus laevis*) in B8JM11 (*Danio rerio*).

Poravnavo smo izvedli z orodjem Clustal Omega [33] na spletni strani Evropskega inštituta za bioinformatiko (EMBL-EBI). Poravnavo smo vizualizirali z orodjem Jalview [34].

3.1.2 Struktурno modeliranje in simulacije molekulske dinamike

Za človeški EpCAM je bila vzeta struktura biološke enote s PDB kodo 4MZV [14]. Izbrisali smo vse atome, ki niso del proteina, N-končni piroglutamatni ostanek smo spremenili v glutaminski ostanek, Q74, Q111 in Q198, ki so v proteinu, uporabljenem za kristalizacijo, na mestih N-glikoziliranih asparaginskih ostankov, pa smo spremenili nazaj v asparaginske ostanke s programom UCSF Chimera 1.15 [30] ter knjižnico rotamerov Dunbrack 2010 [35]. Ostanke C-končno od S260 smo izbrisali.

Preko spletnega vmesnika za AlphaFold 2.1.0 [36, 37] z manjšimi bazami zaporedij na strežniku Google Colab [38] smo modelirali strukture dimerov EpCAM iz vseh vrst razen človeka, z vnesenimi zaporedji (Uniprot ID navedeni v prejšnjem podpoglavlju) za celoten protein brez signalnega zaporedja (*M. musculus*, *X. laevis*, *D. rerio*) oziroma z odrezanim znotrajceličnim delom (*G. gallus*, do vključno S282). Ostanke C-končno od ostanka, ki se poravna s S260 v človeškem EpCAM, smo iz predvidenih struktur naknadno izbrisali.

S strežnikom JPred4 [39] smo za ta proteinska zaporedja napovedali sekundarno strukturo.

Odsek, ki se v poravnavi nahaja na ekvivalentnem mestu kot odsek M73–G79 v človeškem EpCAM (razen za *D. rerio*, pri katerem se zaporedje tiroglobulinske zanke ne ujema s človeškim in kjer smo izbrali odsek Y65–D72), je bil v vseh z Alphafoldom predvidenih strukturah remodeliran kot zanka s programom MODELLER 10.1 [40], s protokolom »standard« preko vmesnika v programu UCSF Chimera 1.15.

Tako pripravljene strukture smo pripravili v programu VMD 1.9.3 [41]. Tvorili smo 6 disulfidnih vezi, med C27 in C46, C29 in C59, C38 in C48, C66 in C99, C110 in C116 ter C118 in C135 (številčenje za človeški EpCAM, pri drugih so bili povezani ekvivalentni cisteinski ostani po vrstnem redu), nato nevtralizirali in ionizirali z 0,15 M NaCl ter obdali s škatlo modela vode TIP3P [42] v obliki kvadra in z robovi vsaj 15 Å oddaljenimi od najbližjega atoma proteina. Histidinski ostanki so bili modelirani kot nevtralni, z vodikom na Nε (tričrkovna oznaka HSE).

Simulacije molekulske dinamike (MD) smo izvedli s programom NAMD 3.0-alpha9 [43]. Sisteme, pripravljene v prejšnjem odstavku, smo minimizirali za 20.000 korakov in ekvilibrirali 10 ps s korakom 0,5 fs z vsemi atomi proteina pritrjenimi na mestu in nato minimizirali celoten sistem, pri čemer so se vsi atomi lahko premikali, za 20.000 korakov. Na podlagi tako minimiziranih sistemov smo zagnali 5 neodvisnih simulacij v NPT ansamblu, vsako z 1 ns ekvilibracije in 500 ns produkcijske simulacije, s korakom 2 fs. Vse vezi z vodikom so bile simulirane kot rigidne, za vzdrževanje rigidnosti je bil uporabljen algoritem SHAKE [44] znotraj proteina in SETTLE [45] za molekule vode. Za vzdrževanje temperature pri 310 K je bila uporabljena Langevinova dinamika, za vzdrževanje tlaka pri 1,01325 bar pa modificirana Nosé-Hooverjeva metoda z Langevinovim batom. Nevezne sile so bile zglajene (angl. *switching*) pri 10 Å in odrezane (angl. *cutoff*) pri 12 Å, približek elektrostatike nad to razdaljo je bil izračunan z metodo PME (Ewaldova mreža delcev, angl. *Particle Mesh Ewald*). Uporabili smo polje sil CHARMM36 [46].

Koordinate atomov so bile shranjene v trajektorijo simulacije vsakih 2500 korakov (5 ps).

3.1.3 Analiza simulacij

Za analize, opisane v tem podpoglavlju, smo, če ni drugače navedeno, uporabili program VMD 1.9.4a44. Analizirali smo produkcijske simulacije.

Iz trajektorij smo odstranili molekule vode. Z vtičnikom pbctools 2.8 za program VMD smo obe podenoti proteina premaknili v centralno periodično celico (angl. *wrapping*). Tako obdelane trajektorije so bile uporabljene za vse analize razen za izračun vezavne energije.

RMSD v odvisnosti od časa smo izračunali na osnovi strukturno poravnanih C_α atomov proteina za vsako vmesno stanje^a. Referenca je bila izvorna struktura, iz katere so se zagnale simulacije.

RMSF smo izračunali za C_α za vsako podenoto posebej, na osnovi C_α atomov, ki so bili poravnani za vsako podenoto posebej. Upoštevana so bila vsa vmesna stanja.

SASA smo izračunali za vsako 5. vmesno stanje (vsakih 25 ps), za celoten kompleks in za vsako podenoto posebej. Uporabili smo polmer molekul topila 1,4 Å. Interakcijsko površino smo izračunali po formuli 1.

$$\text{Interakcijska površina} = \frac{\text{SASA}_A + \text{SASA}_B - \text{SASA}_{AB}}{2} \quad (1)$$

Za statistično analizo smo uporabili program LibreOffice Calc in spletni kalkulator statskingdom.com [47]. Za primerjavo vseh skupin podatkov smo uporabili enosmerno ANOVA, nato pa še Tukeyev HSD test (s spletnim kalkulatorjem statskingdom.com).

Za statistično primerjavo koeficientov variacije interakcijske površine smo uporabili zankanje (angl. *bootstrap*). Iz primerjanih podatkov smo naključno izbrali 10 posameznih podenot iz simulacije in nato znotraj podatkov izbranih simuliranih podenot spet izvedli zankanje (naključno izbrali po 20.000 podatkovnih točk iz vsake izbrane simulacije). Za vsako ponovitev smo izračunali razliko koeficientov variacije. Ta postopek smo ponovili 10.000×. P-vrednosti so bile izračunane po formuli

$$P = 2 \times \frac{\#((c_{vA} - c_{vB}) < 0) + 1}{B + 1} \quad (2),$$

kjer je $\#((c_{vA} - c_{vB}) < 0)$ število razlik koeficientov variacije (vrstni red c_{vA} in c_{vB} je bil izbrana tako, da je dobljena P-vrednost ≤ 1), ki so manjši od 0, B=10.000 pa število ponovitev zankanja. Postopek je v preglednem članku opisal Hesterberg [48]. Za ta namen smo napisali skripto *bootstrap_compare.py*. P-vrednosti smo popravili s Holm-Bonferronijevim postopkom.

3.1.3.1 Analiza vezavne energije z metodo Nwat-MM/GBSA

Metoda MM/GBSA temelji na izračunu razlike energije med kompleksom in disociiranimi podenotama (metoda določanja energije vezave na osnovi končnih stanj):

$$\Delta G = G_{AB} - G_A - G_B \quad (3).$$

^a Vsakih 5 ps.

Bolj specifično so energije izračunane na podlagi danega polja sil, energijo interakcije s topilom pa se dobi s kombinacijo implicitnega modela vode GB in izračunom hidrofobne energije z uporabo SASA:

$$G = E_{MM} + G_{GB} + G_{SA} \quad (4),$$

kjer je E_{MM} molekulsko-mehanska energija v vakuumu, G_{GB} energija polarne interakcije z vodo, dobljena z implicitnim modelom vode GB, in G_{SA} energija nepolarne interakcije z vodo, ki je izračunana z uporabo SASA:

$$G_{SA} = \gamma \times SASA \quad (5),$$

kjer je γ površinska napetost [49, 50].

Vključitev molekul vode v interakcijski površini med podenotama lahko znatno izboljša korelacijo rezultatov MM/GBSA in sorodnih metod z eksperimenti [51–53] – takšne molekule vode lahko prispevajo kar 25% ΔG vezave [54]. Zaradi tega smo se odločili upoštevati tudi prispevek vode, ki se nahaja znotraj interakcijske površine.

Naš protokol smo zasnovali po zgledu metode Nwat-MM/GBSA Maffucci in Continija [52], z manjšimi spremembami. Ohranjenih je bilo 25 molekul vode, oddaljenih največ X od polarnih ostankov^b obeh podenot, kjer je X najnižja razdalja, ki da vsaj 25 molekul vode. V primeru, da je znotraj oddaljenosti X bilo več kot 25 molekul vode (dve ali več molekul vode z enako oddaljenostjo od podenot, kar je mogoče zaradi omejene ločljivosti razdalj), smo izmed najbolj oddaljenih izbrali tiste z nižjim indeksom^c kisikovega atoma. Ohranjene molekule vode so pri homodimerih bile obravnavane kot del podenote A (v heterodimerih je to EpEX). Za ta namen smo napisali skripto *extract_numwat_with_wrapping.tcl*, namenjeno zaganjanju v programu VMD.

Energije smo izračunali za vsako 100. vmesno stanje^d s programom NAMD 2.14, poljem sil CHARMM36, implicitno solvacijo po metodi GB s parametri GB^{OB}II [55], površinsko napetostjo 0,00542 kcal mol⁻¹ Å⁻² in koncentracijo monovalentnih ionov 0,15 M. Sile so bile odrezane pri 18 Å in zglajene pri 16 Å. Za ta namen smo napisali skripto *nwat_mmgbsa.sh*.

3.1.3.2 Izračun prostornine hidrofobnega žepa

V programu UCSF Chimera 1.15 smo z orodjem »MatchMaker« poravnali strukture posameznih podenot iz vseh 5 uporabljenih struktur in jih shranili v PDB formatu. Nato smo v programu VMD vsako 100. vmesno stanje poravnali z ustrezajočo strukturo iz

^b Kot definirano s selekcijsko ključno besedo »polar« v programu VMD 1.9.4a44 – Arg, Asp, Asn, Cys, His, Glu, Gln, Lys, Ser, Thr, Tyr.

^c Interno številčenje atomov v programu VMD oziroma trajektoriji.

^d Vsakih 500 ps.

prejšnjega stavka. Za to poravnavo so bili upoševani vsi atomi, poravnana in shranjena je bila vsaka podenota posebej.

Nato smo s skripto POVME_pocket_id.py, ki je del programa POVME2 [56, 57], v začetni strukturi monomera človeškega EpEX določili prisotnost žepov. Nastavili smo parametra ločljivost zaznave žepa (na 2,0 Å) in število sosedov (na 6). Ostali parametri so bili privzeti. Rezultat so bile koordinate in polmeri 5 vključitvenih krogel, ki so imele za 5 Å večji polmer, kot bi bilo potrebno za popolno vključitev hidrofobnega žepa v izvorni strukturi človeškega EpEX. Ker ta skripta s temi parametri v ostalih strukturah ni zaznala prisotnosti žepa v SEA domeni, so bili za izračun prostornine za vse strukture uporabljeni enake nastavitev kot za človeški EpEX.

Pri izračunu prostornin s programom POVME2.py smo uporabili 5 vključitvenih krogel, ki jih je predlagala skripta POVME_pocket_id.py. Izključene so bile točke, bližje od 1,09 Å do kateregakoli atoma in točke izven konveksne površine proteina. Na začetku žepa smo kot kroglo sosednjih točk vzeli koordinate 2. najbolj zunanje vključitvene krogle s polmerom 4 Å. Točke, ki niso zvezno povezane s to kroglo, niso bile obravnavane kot del žepa.

Pri statistični obravnavi smo kot neodvisne podatkovne točke obravnavali posamezno podenoto iz vsake neodvisne simulacije (po 2 podenoti na simulacijo).

Za statistično primerjavo koeficientov variacije prostornine hidrofobnega žepa smo uporabili zankanje.

3.2 Heterodimer EpCAM:Trop2

3.2.1 Struktурno modeliranje in simulacije molekulske dinamike

Za človeški EpCAM, Trop2 in hipotetični heterodimer sta bili pripravljeni po dve strukturi. Za EpCAM sta bili vzeti struktura biološke enote s PDB kodo 4MZV s spremembami Q74N, Q111N in Q198N (enako kot v podpoglavlju 3.1.2) in struktura, z Alphafoldom predvidena kot homodimer iz zaporedja z Uniprot kodo P16422. Za Trop2 sta bili iz strukture s PDB kodo 7PEE [15] vzeti verigi B in C z izbrisanimi vsemi atomi, ki niso del enega izmed standardnih aminokislinskih ostankov, in spremembama Q120N in Q208N, in struktura, z Alphafoldom predvidena kot homodimer iz zaporedja z Uniprot kodo P09758. Za hipotetični heterodimer so bile z Alphafoldom kot heterodimer z zaporedji P16422 in P09758 predvidene tri strukture, za nadaljnje delo sta bili uporabljeni druga in tretja. Vse z Alphafoldom predvidene strukture smo modelirali s celotnim zaporedjem, razen signalnega zaporedja, in jim naknadno odrezali del C-končne regije, od S260 (za EpCAM) oz. S269 naprej (za Trop2). Uporabljena je bila enaka različica Alphafolda kot v podpoglavlju 3.1.2.

Modelirali smo tudi strukturo mutirane oblike EpCAM TYLoop^{mut} [23], ki smo jo pripravili iz strukture EpCAM 4MZV s spremembami Q74N, Q111N in Q198N (enako kot v podpoglavlju 3.1.2) in uvedenimi mutacijami K83D, P84D in L88D s programom UCSF Chimera [30] ter knjižnico rotamerov Dunbrack 2010 [35].

Strukture 4MZV, 7PEE verigi B in C in strukturi heterodimera smo razdelili na monomere in vsak monomer posebej pripravili na simulacije.

Te strukture smo pripravili enako kot v podpoglavlju 3.1.2. Nato smo za vsak sistem zagnali po 5 neodvisnih 5 ns dolgih simulacij z enakimi parametri kot v podpoglavlju 3.1.2, z izjemo uporabljenega polja sil, ki je bilo CHARMM36m [58].

Koordinate atomov so bile shranjene v trajektorijo simulacije vsakih 5 ps (2500 korakov).

3.2.2 Analiza simulacij

Tvorbo vodikovih vezi med atomi skeleta vijačnice sredi tiroglobulinske zanke Trop2 v strukturi heterodimera smo analizirali z vtičnikom Hydrogen Bonds 1.2 za program VMD. Uporabili smo 3,5 Å za največjo dovoljeno razdaljo med donorjem in akceptorjem in 30° za največji dovoljeni odmik od idealnega kota.

Vezavne energije so bile izračunane podobno kot v podpoglavlju 3.1.3.1, le da smo vzeli vsako drugo vmesno stanje^e 4. in 5. nanosekunde simulacije in uporabili polje sil CHARMM36m. Pri statistični analizi je bila vsaka simulacija obravnavana kot ena neodvisna podatkovna točka, primerjali smo rezultate različnih proteinov (torej dve

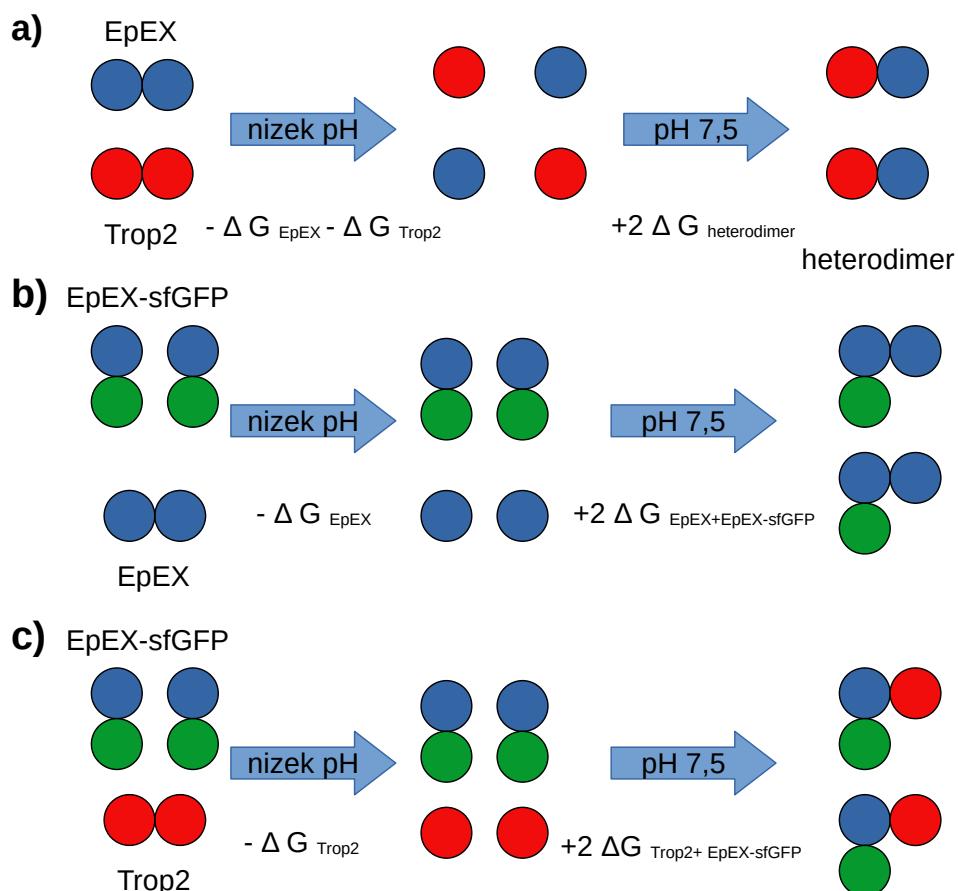
e Vsakih 10 ps.

različni strukturi istega proteina skupaj, razen pri TyLoop^{mut}, kjer je bila simulirana le ena struktura).

Z vtičnikom NAMD Energy 1.4 za program VMD in programom NAMD 2.14 smo analizirali energijo tiroglobulinske zanke Trop2 podenot^f. Analizirana sta bila odseka 81–103 in 83–93. Uporabili smo polje sil CHARMM36m.

Analizirali smo tudi MM/GBSA energijo monomerov, izvlečenih iz simulacij dimerov, in energijo posebej simuliranih monomerov (enaki parametri simulacij kot za dimere) za strukture 4MZV, 7PEE verigi B in C ter obe strukture heterodimera. Vse molekule vode so bile izbrisane, analizirali smo energijo vsakega vmesnega stanja 4. in 5. nanosekunde.

3.2.3 Laboratorijsko določanje obstoja heterodimera



Slika 3: Shema načrtovane priprave heterodimera. **a)** Načrtovana priprava heterodimera EpEX – Trop2. **b)** Načrtovana priprava dimera EpEX – EpEX-sfGFP. **c)** Načrtovana priprava dimera Trop2 – EpEX-sfGFP.

Ker okoli pH 4,7 tako EpEX [16] kot tudi Trop2 [neobjavljeni rezultati] disociirata na monomere, smo se odločili, da bomo heterodimer poskušali pripraviti z zakisanjem in nevtralizacijo mešanice homodimerov, kot je prikazano na sliki 3. Poleg pH 4,5 oziroma

^f V homodimeru Trop2 obe podenoti, v heterodimeru pa samo Trop2.

5,5 smo se za disociacijo homodimerov pri pripravi heterodimer EpEX – Trop2 (slika 3a) odločili uporabiti tudi 0,5 M NaI.

Na sliki 3bc je prikazana shema tvorbe dimerov z EpEX-sfGFP. EpEX-sfGFP obstaja kot monomer [59], predvidoma zato, ker se sfGFP na C-končno od EpEX sterično ovirata. Z EpEX ali Trop2 s prostim C-koncem takšnega steričnega oviranja predvidoma ne bi bilo ali pa bi bilo precej manjše. Ker se na sliki 3c iz enega dimera tvorita dva, bi lahko tvorba dimera EpEX-sfGFP s Trop2 bila ugodna, četudi tvorba navadnega heterodimera ni. Ker ima EpEX-sfGFP približno dvakrat večjo molekulsko maso kot monomer EpEX ali Trop2, bi s prečnim povezovanjem in NaDS-PAGE opazili tvorbo dimerov himere EpEX-sfGFP z EpEX ali Trop2 pri približno trikratni molekulski masi monomera EpEX ali Trop2.

3.2.3.1 Priprava vzorcev

Neglikozilirana mutanta EpEX (koncentracija 14,7 mg/ml) je bila pripravljena, kot so opisali Pavšič in sod. (2014) [14], delno glikozilirana mutanta Trop2 (konc. 15,8 mg/ml) kot v Pavšič (2021) [15] in himerni protein EpEX-sfGFP (konc. 1,14 mg/ml) kot v Hočevar (2021) [59]. Proteini so bili do začetka dela shranjeni pri -80 °C in nato med delom pri -20 °C.

Vzorce smo zakisali z dodatkom acetatnega pufra do želenega pH (4,5 ali 5,5), počakali 5 minut z vzorci na sobni temperaturi in jih nato nevtralizirali z dodatkom HEPES pufra do pH 7,5. Kontrolni vzroci so bili pripravljeni z redčenjem proteinov z že vnaprej zmešanimi acetatnim in HEPES pufrom v enakih razmerjih, kot se na koncu nahajajo v zakisanih in nevtraliziranih vzorcih. pH mešanic so bili izračunani s pomočjo spletnega kalkulatorja webqc.com [60] in nato v laboratoriju preverjeni z mešanjem čistih pufrov v predvidenih razmerjih.

Za ugotavljanje disociacije v prisotnosti soli smo pripravili mešanice EpEX in Trop2 z 1 M izbrane soli (NaCl, NaI, MgCl₂) v 20 mM HEPES pufru s pH 7,5 in 150 mM NaCl (kontrola je bil vzorec samo v tem pufru). Vzorci so bili pripravljeni brez ali z 0,1 % v/v detergenta Triton X-100. Vse navedene koncentracije so pred prečnim povezovanjem, po dodatku glutaraldehida so bile nižje za 20 %. Pred prečnim povezovanjem z glutaraldehidom smo vzorce na sobni temperaturi pustili 5 minut.

Namesto zakisanja in nevtralizacije smo se odločili preizkusiti tudi odstraniti dejavnik, ki naj bi sprožil disociacijo homodimerov, z dializo. Pripravili smo vzorce s koncentracijami A: 0,7 µg µl⁻¹ EpEX + 150 mM NaCl + 0,5 M NaI, B: 0,7 µg µl⁻¹ Trop2 + 150 mM NaCl + 0,5 M NaI, C: 0,7 µg µl⁻¹ EpEX + 0,7 µg µl⁻¹ Trop2 + 150 mM NaCl + 0,5 M NaI, D: 1 µg µl⁻¹ + EpEX 1 µg µl⁻¹ Trop2 + 150 mM NaCl zakisan na pH 5,5 ter jih pred dializo pustili 5 minut na sobni temperaturi. Vzorec E smo, da bi se izognili denaturaciji EpEX zaradi NaI in denaturaciji Trop2 zaradi nizkega pH, pripravili tako, da smo zmešali prvo polovico – Trop2 z 0,5 M NaI v 72,5 M HEPES pufru pri pH 7,5 – in drugo polovico – EpEX pri pH 4,5 (10 mM HEPES 20 mM acetat) – ter ju pustili 5

minut pri sobni temperaturi in zmešali tik pred dializo. Vzorci A, B in C so imeli končne volumne 27,5 µl, vzorca D in E pa 40 µl. Dializo smo izvedli v Mini GeBAflex-tube (Gene Bio-Application Ltd.) proti 20 mM HEPES 150 mM NaCl pH 7,5 pufru, čas trajanja dialize je bil 3 ure.

3.2.3.2 Nativna elektroforeza

Z elektroforeznim sistemom PhastSystem (Pharmacia) smo izvedli nativno poliakrilamidno elektroforezo s PhastGel Gradient 8–25 (GE Healthcare, 17-0542-01) in PhastGel Buffer Strips – Native (GE Healthcare, 17-0517-01).

Agarozno nativno elektroforezo smo izbrali, ker omogoča opazovanje potovanja proteinov v obe smeri na enem samem gelu. Izvedena je bila po protokolu Kim (2011) [61]. Tris-glicinski pufer je imel pH 7,5 do 8. Kjer je navedeno, je kadička za elektroforezo bila hlajena z ledom.

Poliakrilamidna nativna elektroforeza s fosfatnim pufrom je bila izvedena po protokolu Gallagher (2018) [62], vsak vzorec je bil nanešen na dva 7 % poliakrilamidna gela, od katerih je v enem elektroforeza potekala v smeri katode, v drugem pa v smeri anode. Elektroforezni pufer je imel pH 7,5. Nanašalni pufer se je od navedenega v referenci razlikoval v vsebnosti metilenskega modrila za vizualizacijo elektroforezne fronte v smeri katode. Kadička za elektroforezo je bila hlajena z ledom.

Kot lestvico/standard smo uporabljali Isoelectric Focusing Calibration Kit, Broad pI Kit (Pharmacia Biotech, 17-0471-01), razredčen po navodilih proizvajalca.

Geli so bili pobarvani s Coomassie Brilliant Blue.

Kjer je navedeno, so bili vzorci zamrznjeni z ali brez nanašalnega pufra pri –20 °C in odmrznjeni pred nanosom.

3.2.3.3 Prečno povezovanje

Prečno povezovanje smo izvedli z glutaraldehidom (Merck KgaA, 8.14393.1000) v vzorcih, ki so imeli celokupno koncentracijo proteinov 0,5 mg ml⁻¹ (po dodatku glutaraldehyda, pred tem je koncentracija 0,6 mg ml⁻¹). V vzorec smo dodali 0,5% v/v glutaraldehid, raztopljen v dH₂O, do 0,1% v/v končne koncentracije, pustili 5 minut in nato reakcijo prekinili z dodatkom Tris-vsebujočega 6× NaDS nanašalnega pufra z β-merkaptoetanolom. Vzorci so bili po 5 min denaturacije na 95 °C nanešeni pa NaDS-PAGE gel. Uporabili smo velikostno lestvico Pierce™ Unstained Protein Marker (ThermoFisher Scientific, 26610). Geli so bili pobarvani z barvilm Coomassie Brilliant Blue.

ALJAŽ SIMONIČ: MOLEKULSKA DINAMIKA PROTEINA EPCAM IZ RAZLIČNIH VRST IN NJEGOVA
INTERAKCIJA S PARALOGNIM PROTEINOM TROP2

4 Rezultati in razprava

4.1 EpCAM iz različnih vrst

4.1.1 Poravnava

Homo/24-26	24	- Q EEC V CEN Y K L A V N C F V N N N R O C O C T - - S - V G A Q N T V I C S K L A A K C L V M K A E M N G S K L G R R A	82
Mus/24-261	24	- O R D C V C D N Y K L A T S C S L N E Y G E C O C T - - S - Y G T Q N T V I C S K L A S K C L A M K A E M T H S K S G R R I	82
Gallus/24-25	24	- Q D C T C T C K N K R V T N C R L I D - N V C H C N - - S - I G S S V P V N C E I L T S K C L L M K A E M A N T K S G R R E	81
Xenopus/23-	23	Q N P G C K C K T H Y L G K C D N S G A S S D C O C A L S I - G P A S Q A V D C T K L I P K C W L M K R E S L G T K A G R R V	84
Danio/19-25*	19	- - - Q C A C K T M K W A N C D - - - D S C S C S L R L T E S S T Q T L N C S K L V P V K C F L M Q A E M Y R A R N H Q D T	73
Homo/24-26	83	- - - - - K P E G A L Q N N D G L Y D P D C D E S G L F K A K Q C N G T S M C W C V N T A G V R R T D K - D T E I I T C S E R V	139
Mus/24-261	83	- - - - - K P E G A I Q N N D G L Y D P D C D E S G L F K A K Q C N G T A T C W C V N T A G V R R T D K - D T E I I T C S E R V	139
Gallus/24-25	82	- - - - - K P K D A L Q D T D G L Y D P E C E T N G V F K A R Q C N N T D T C W C V N T A G V R R T D K H D T D L K C N Q L V	138
Xenopus/23-	85	- - - - - K P V Q A L I D N D G L Y D P E C E T N G V F K A R Q C N N T D T C W C V N T A G V R R T D K G D K N W K C P E L V	142
Danio/19-25*	74	R S G G K P V E T A F V D N D G I Y D P V C E S D G K F K A V Q C N N T E V C W C V N S A G V R R S D K K D K N I K C E P - A	135
Homo/24-26	140	R T Y W I I I E L K H K A R E K P Y D S K S L R T A L Q K E I T T R Y Q L D P K F I T S I L Y E N N - - V I T I D L V Q N S S	200
Mus/24-261	140	R T Y W I I I E L K H K A R E S P Y D H O S L Q T A L Q E A F T S R Y K L N Q K F I K N I M Y E N N - - V I T I D L M Q N S S	200
Gallus/24-25	139	R T T W I I I E M R H A E R K T P L N A E S L T R Y L K D T I T S R Y M L D G R Y I S G V V Y E N S - - T I T I D L K Q N S S	199
Xenopus/23-	143	R T N W V I V E M K R N N S D S - V N D D V L I O A L K T T I L N R Y G L P E K Y V S V E L E G S S - - F I Y I D L K Q N G T	202
Danio/19-25*	136	E T Y W V R V E M K H K S V D V P I D A T K L R T G I E N V L Q Q R Y G L D K K L V S E V O Y D K D G R L I V V D V K K D E N	198
Homo/24-26	201	Q K T Q N D D V D I A D V A Y Y F E K D V K G E S L F H S - K K M D L I T V N G E Q L D L D P G Q T L I Y Y V D E K A P E F S	260
Mus/24-261	201	Q K T Q D D D V D I A D V A Y Y F E K D V K G E S L F H S S K S M D L R V N G E P L D L D P G Q T L I Y Y V D E K A P E F S	261
Gallus/24-25	200	D K T P G D D V D I T D V A Y Y F E K D V K D D S I F L N N K - L N M N I D N E E - - L K F D N M M V Y Y V D E V P P E F S	257
Xenopus/23-	203	Q K L P G E V D I T D V G Y Y M E K D I K G D P L F H P D E K F E I I L V N G K N F G - - V K E P V I Y Y V D E K P H E I T	261
Danio/19-25*	199	D - - - R T T D L S L M T Y Y M E K D I K V L P L F W N G Q P F E V D V P G T K V S - - M E N V L I Y Y V D D K A P T F T	254

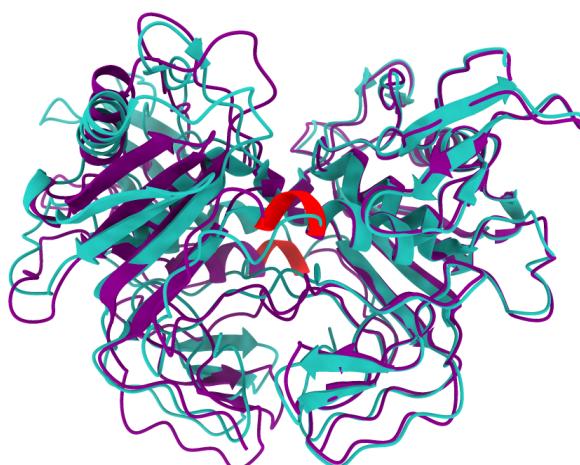
Slika 4: Poravnava zaporedij vseh petih uporabljenih EpEX. Zaporedja so bila poravnana s Clustal Omega [33] in vizualizirana s programom Jalview [34]. Shema barv je »Zapp«.

Poravnava, prikazana na sliki 4, je bila uporabljena za grafe RMSF. Datoteka poravnave je dostopna v poglavju Priloge.

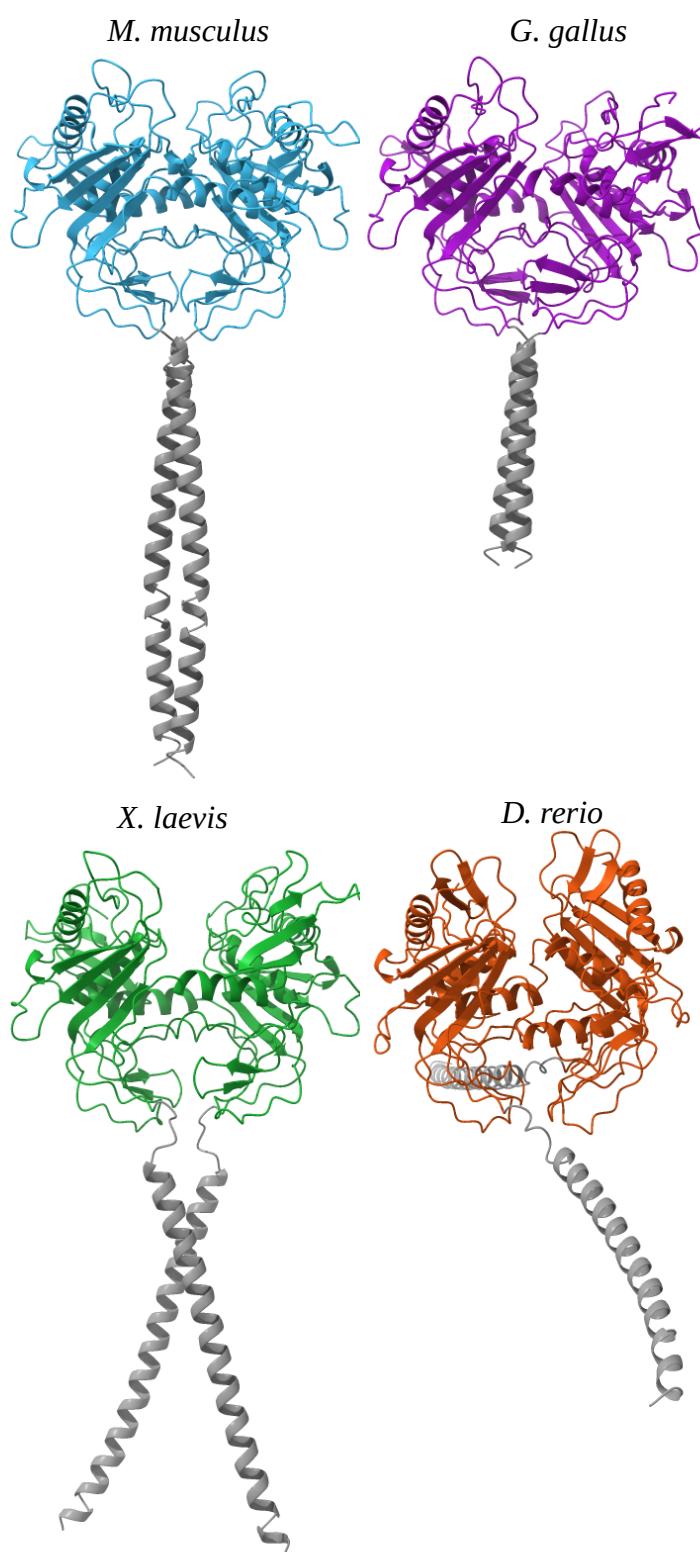
4.1.2 Modeli struktur

Predvideni modeli struktur so bili precej podobni eksperimentalno določeni strukturi 4MZV, terciarna struktura je bila bolj podobna eksperimentalni kot kvartarna struktura.

Glavna razlika med predvidenimi in eksperimentalnimi strukturami je bila za en obrat podaljšana α -vijačnica tiroglobulinske tip-1 domene, kar Alphafold v vseh modelih napove z nižjo oceno zanesljivosti (nižja pLDDT vrednost) kot preostanek α -vijačnice. Ta razlika je bila prisotna tudi v z



Slika 5: Strukturna poravnava 4MZV (moder) in z Alphafoldom predvidene strukture človeškega EpCAM (vijolično). Del tiroglobulinske zanke, ki ga Alphafold generira kot vijačnico, je rdeč.

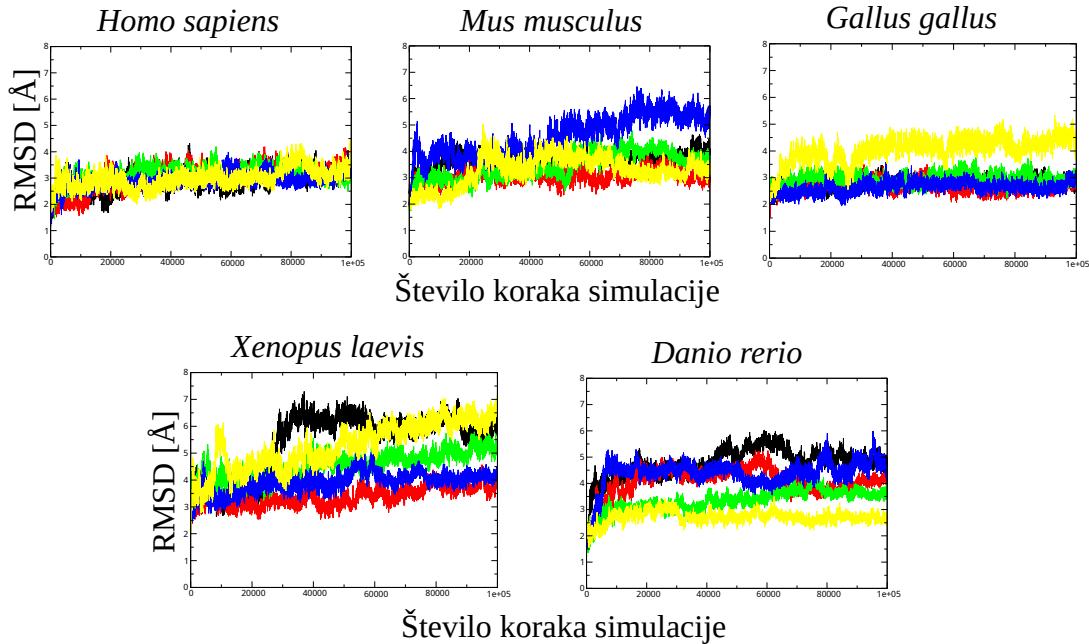


Slika 6: Z AlphaFoldom predvidene strukture. Siva sta transmembranski in citosolni del (citosolni ni bil vključen pri *G. gallus*), ki sta bila pred simulacijami odrezana. Imena vrst so nad slikami.

AlphaFoldom generirani strukturi človeškega EpCAM (pripravljen v delu »Heterodimer EpCAM:Trop2«), kar je prikazano na sliki 5. Strežnik JPred4 [39] je za EpCAM iz vseh preučevanih vrst predvidel konec α -vijačnic na mestu, homolognemu koncu α -vijačnic v kristalni strukturi človeškega EpEX. Ta del α -vijačnice v človeškem in *X. laevis* EpCAM vsebuje tudi glicinski ostanek, ki je v α -vijačnicah neugoden [63]. Zaradi vsega tega smo sklepali, da gre pri vseh modelih verjetno za napako in smo ta del spremenili v zanko.

Strukture, kot jih je predvidel AlphaFold, so prikazane na sliki 6. Zanimivo je, da AlphaFold nekatere transmembranske vijačnice (v *M. musculus* in *G. gallus*) predvidi kot dimerizirane, drugih pa ne (v *X. laevis* in *D. rerio*). Glede na visoko podobnost transmembranskih regij verjetno gre za artefakt.

4.1.3 RMSD in fleksibilnost proteina

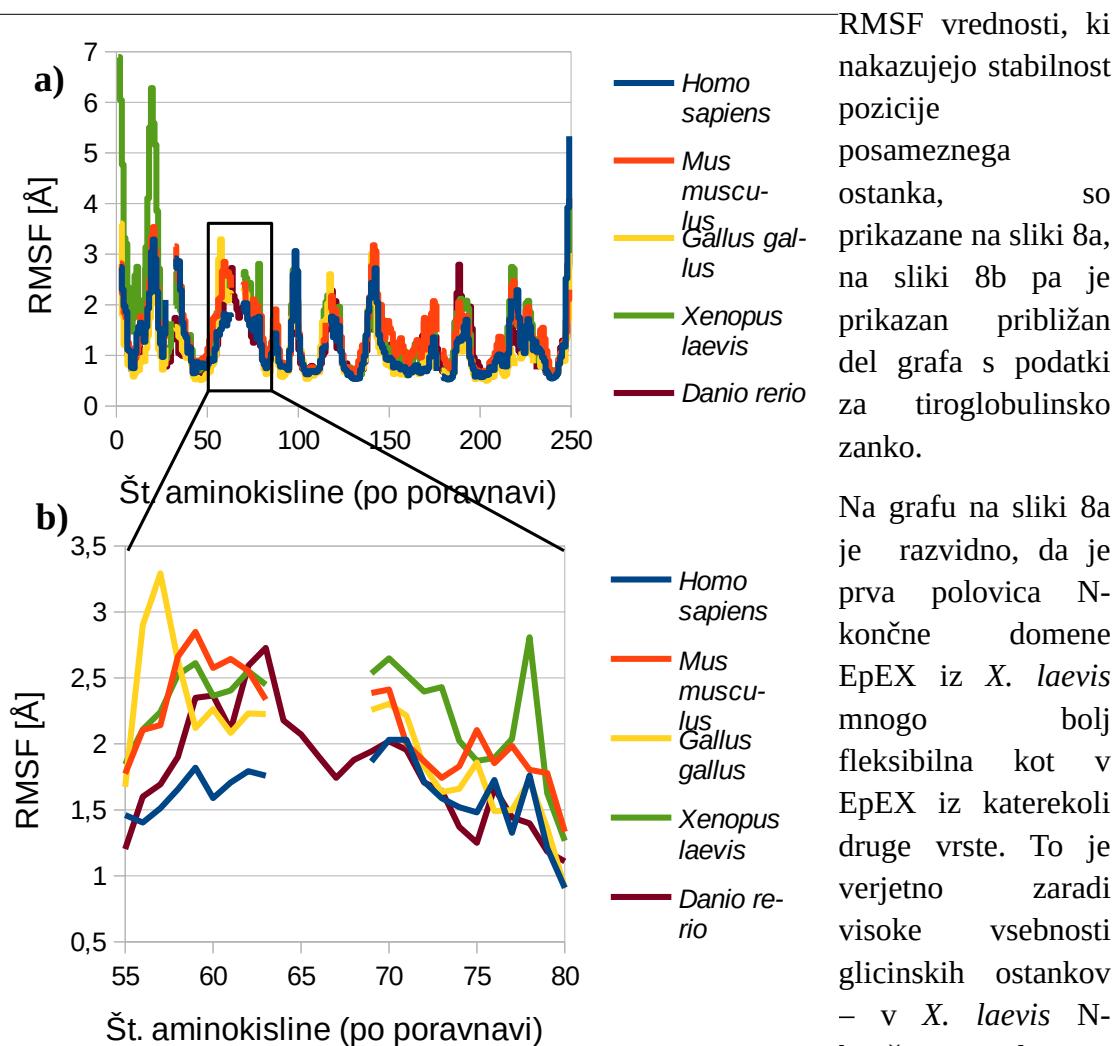


Slika 7: RMSD $C\alpha$ atomov v odvisnosti od koraka simulacije za po pet različnih produkcijskih simulacij EpEX iz vseh petih vrst. X os se začne pri 0 in konča pri 100.000 (kar ustreza času 500 ns), Y os pa se začne pri 0 Å in konča pri 8 Å. Slika je bila ustvarjena s programom xmgrace.

RMSD produkcijskih simulacij v primerjavi z začetno strukturo se začnejo okoli 2 Å, kot je razvidno na sliki 7. Prav tako se pri večjem številu simulacij RMSD ne poveča drastično. To nakazuje, da je bil simulirani sistem na začetku produkcijskih simulacij zadosti ekvilibriran – ne opazujemo več začetne kristalne ali predvidene strukture.

Nekaj simulacij je doseglo višji RMSD, pri teh višjih vrednostih pa so tudi stabilno ostajale. To bi lahko nakazovalo na obstoj več stabilnih konformacijskih stanj. V takšna stanja se posamezne simulacije, tako kratke [64] kot tudi dolge [65], pogosto ujamejo. Pri *M. musculus* in *G. gallus* EpEX izstopata le po ena simulacija, pri *X. laevis* in *D. rerio* pa strukture iz različnih simulacij pokrivajo večji razpon RMSD vrednosti.

Ker se vrednosti RMSD med simulacijami razlikujejo in bolj ali manj stabilno ostajajo pri določeni vrednosti (pri *M. musculus*, *X. laevis* in *D. rerio* je opaznih nekaj prehodov med različnimi stanji, a RMSD večine simulacij ostaja stabilen od začetka do konca), smo kot neodvisne podatkovne točke pri statističnih analizah namesto posameznih delov simulacij vzeli povprečja celotne simulacije. Pri analizah lastnosti, za katere ne pričakujemo, da bi nanje močno vplivala konformacija nasprotne podenote, smo kot neodvisne podatkovne točke vzeli povprečja lastnosti za posamezno podenoto v eni simulaciji.



Slika 8: RMSF Ca atomov v EpEX. Vrednosti so povprečje vseh simuliranih podenot. X os je mesto ostanka v poravnavi, vrzeli v poravnavi so prikazane kot prekinitev črte. **a)** RMSF celotnega simuliranega dela EpEX. **b)** Približan del grafa a, ki prikazuje RMSF tiroglobulinske zanke.

RMSF vrednosti, ki nakazujejo stabilnost pozicije posameznega ostanka, so prikazane na sliki 8a, na sliki 8b pa je prikazan približan del grafa s podatki za tiroglobulinsko zanko.

Na grafu na sliki 8a je razvidno, da je prva polovica N-končne domene EpEX iz *X. laevis* mnogo bolj fleksibilna kot v EpEX iz katerekoli druge vrste. To je verjetno zaradi visoke vsebnosti glicinskih ostankov – v *X. laevis* N-končna domena znotraj zavojev in zank vsebuje štiri glicinske ostanke, v *D. rerio* nobenega, v *H. sapiens* in *G. gallus* enega ter v *M. musculus* dva glicinska ostanka.

RMSF tiroglobulinske zanke so podobni, v začetnem delu pa s približno 1 Å nižjim RMSF izstopa človeški EpEX. To bi morebiti lahko bila posledica spremembe tega odseka v z Alphafoldom predvidenih strukturah v zanko. V človeškem EpEX je razvidno omrežje vodikovih vezi M50–S53–R57. Če v EpEX iz drugih vrst obstajajo podobne strukture, jih pri spremenjanju v zanko ne bi dobili.

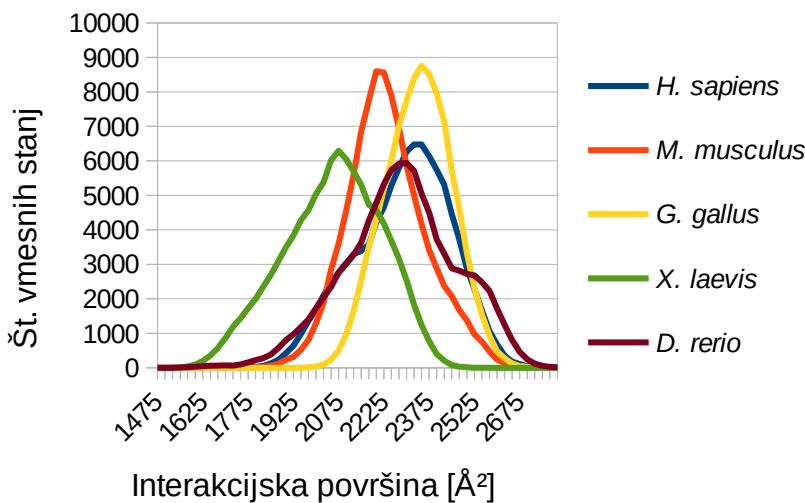
4.1.4 Interakcijska površina med podenotama dimera

Tabela 1: Interakcijske površine in njihovi 95% intervali zaupanja, dobljeni iz Studentove t porazdelitve (povprečje posamezne simulacije kot neodvisna podatkovna točka).

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>	<i>D. rerio</i>
Interakcijska površina [\AA^2]	2300	2250	2350	2070	2290
95% IZ [\AA^2]	(2188; 2411)	(2140; 2357)	(2272; 2433)	(1988; 2148)	(2153; 2431)
Koef. variacije	6,9%	5,8%	4,7%	7,9%	8,1%

Najnižjo interakcijsko površino ima EpEX iz *X. laevis* (primerjava vseh interakcijskih površin med seboj s Tukeyevim HSD testom da statistično pomembno razliko vseh ostalih EpEX z EpEX iz *X. laevis* – p-vrednosti so v poglavju Priloge, tabela P1).

Pri primerjavi koeficientov variacije z zankanjem je razlika po popravku za večkratno primerjanje statistično pomembna le pri paru *G. gallus* – *X. laevis* ($p=0,0020$). Nepopravljene p-vrednosti so v tabeli P2, poglavje Priloge.



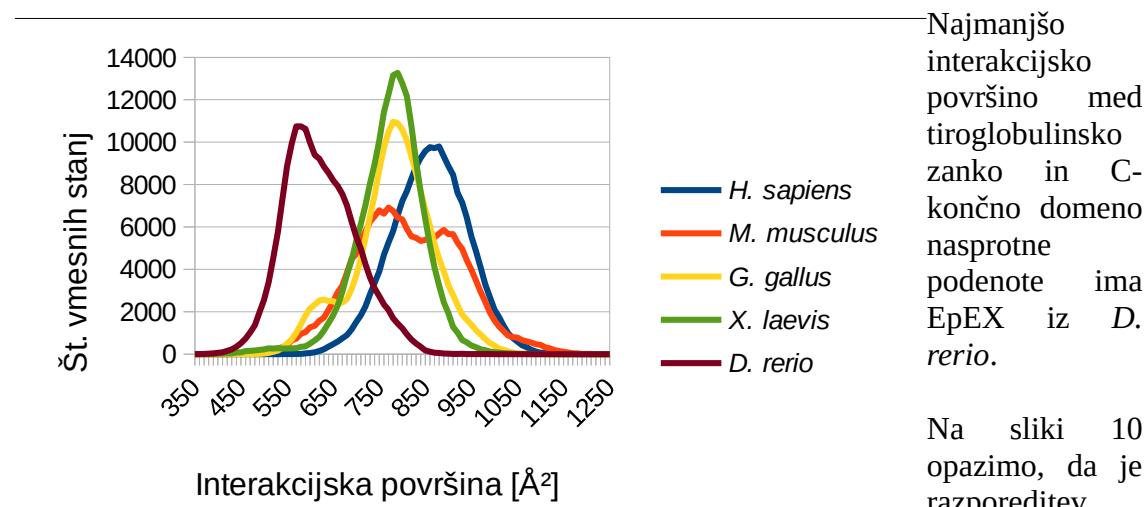
Slika 9: Razporeditev interakcijskih površin EpEX iz različnih vrst. Slika je bila ustvarjena s programom LibreOffice Calc.

Tabela 2: Interakcijske površine tiroglobulinskih zank s C-končno domeno in njihovi 95% intervali zaupanja, dobljeni iz Studentove t porazdelitve (povprečje posamezne simulacije kot neodvisna podatkovna točka).

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>	<i>D. rerio</i>
Interakcijska površina [\AA^2]	865	818	786	780	627
95% IZ [\AA^2]	(827; 902)	(749; 887)	(738; 834)	(753; 807)	(589; 665)
Koeficient variacije	9,4%	14,0%	11,3%	9,4%	12,4%

Tabela 3: P-vrednosti, dobljene pri medsebojnih primerjavah interakcijskih površin tiroglobulinske zanke s C-končno domeno nasprotne podenote s Tukeyevim HSD testom. Vrednosti, ki nakazujejo statistično pomembnost, so odebunjene. Kriterij za statistično pomembnost je $p < 0,05$.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>
<i>M. musculus</i>	0,49			
<i>G. gallus</i>	0,066	0,80		
<i>X. laevis</i>	0,040	0,68	0,99	
<i>D. rerio</i>	1,7e-9	4,0e-7	1,7e-5	3,4e-5



Slika 10: Razporeditev interakcijskih površin tiroglobulinske zanke s C-končno domeno nasprotne podenote.

zanke s C-končno domeno pri *M. musculus* in *G. gallus* bimodalna, pri *M. musculus* bolj izrazito kot pri *G. gallus*. Pri *G. gallus* izstopata le dve simulirani podenoti, pri *M. musculus* pa so enakomerno porazdeljene med oba vrhova. To bi potencialno lahko bila posledica remodeliranja začetnega dela tiroglobulinske zanke v z Alphafoldom predvidenih strukturah.

4.1.5 Predvidene vezavne energije

Tabela 4: Predvidene vezavne energije, dobljene z metodo Nwat-MM/GBSA in njihovi 95% intervali zaupanja (po Studentovi t-porazdelitvi). Energije so podane v enotah kcal mol⁻¹.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>	<i>D. rerio</i>
Povprečje [kcal mol ⁻¹]	-216	-207	-240	-196	-174
95% IZ [kcal mol ⁻¹]	(-239; -194)	(-226; -188)	(-256; -225)	(-204; -188)	(-191; -157)

Tabela 5: P-vrednosti, dobljene pri medsebojnih primerjavah vezavnih energij različnih proteinov s Tukeyevim HSD testom. Vrednosti, ki nakazujejo statistično pomembnost, so odebujene. Kriterij za statistično pomembnost je $p < 0,05$.

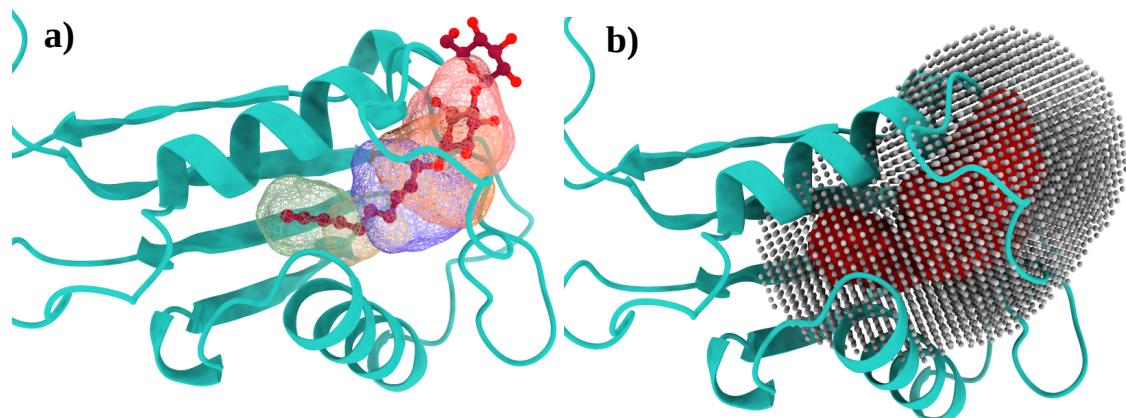
	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>
<i>M. musculus</i>	0,82			
<i>G. gallus</i>	0,083	0,0085		
<i>X. laevis</i>	0,16	0,70	4,6e-4	
<i>D. rerio</i>	8,8e-4	0,010	2,5e-6	0,14

Izračuni po metodi Nwat-MM/GBSA, ki so prikazani v tabeli 4, so kot najmočneje povezani dimer napovedali kokošji EpEX, sledili pa so mu človeški, mišji in EpEX iz *X. laevis*. Najšibkeje je povezan EpEX iz *D. rerio*. Statistično pomembne razlike so prikazane v tabeli 5. Opazimo trend višanja vezavne energije od rib do ptičev, z dvoživkami in sesalci nekje vmes.

Dobljene energije se ujemajo s koeficienti variacije interakcijske površine – dimeri z nižjimi vezavnimi energijami imajo nižji koeficient variacije interakcijske površine (izjema sta *H. sapiens* in *M. musculus* EpEX, kjer ima *H. sapiens* EpEX nižjo energijo vezave in višji koeficient variacije interakcijske površine).

Dobljene absolutne vezavne energije so zelo visoke, kar je pri metodi MM/GBSA zaradi zanemarjanja entropije pričakovano in zaradi tega se s to metodo dobljene energije praviloma obravnavajo relativno [49, 66]. Kljub zanemarjanju entropije so korelacije z eksperimentom tudi brez entropije večinoma zadovoljive [49, 53, 66]. Entropije nismo izračunali, saj metode ali zahtevajo veliko računsko moč za majhno izboljšanje korelacije z eksperimentom ali pa sta metodi (metoda interakcijske entropije in metoda C2) primerni le za sisteme, pri katerih je standardni odklon entalpije interakcije med molekulama v vakuumu večkrat manjši kot pri proteinskih kompleksih in za proteinske komplekse korelacijo z eksperimentom le poslabšata [50, 53, 67].

4.1.6 Prostornina hidrofobnega žepa



Slika 11: a) SEA domena človeškega EpEX (svetlomodra) s hidrofobnim žepom, kot ga zazna POVME (raznobarvna površina) in decil- β -D-maltopiranoidom v žepu (rdeč, kroglični model). b) SEA domena človeškega EpEX, hidrofobni žep je rdeča površina, vključitveno območje, uporabljen za izračune prostornine, pa je prikazano s sivimi kroglicami.

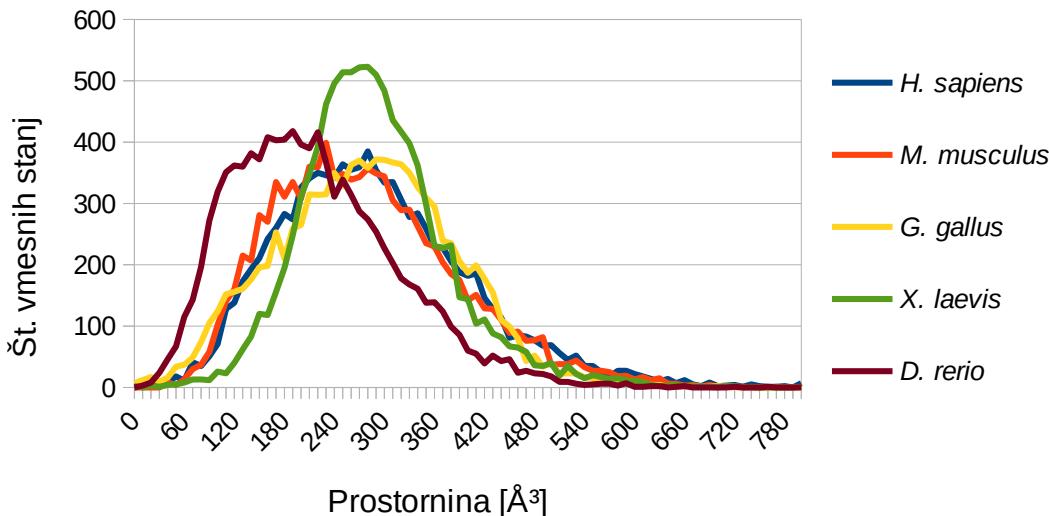
Na sliki 11 sta prikazana prepoznan žep, ki lepo zaobjame celoten zakopan del molekule detergenta v kristalni strukturi 4MZV. Skripta POVME_pocket_id.py je prepoznała prisotnost žepa le v izhodni strukturi človeškega EpEX, kar je verjetno posledica tega, da je žep zaradi vezanega detergenta bolj odprt kot v z Alphafoldom predvidenih strukturah EpEX iz drugih vrst.

Tabela 6: Prostornine hidrofobnega žepa in njihovi 95% intervali zaupanja (po Studentovi t-porazdelitvi). Interval zaupanja je izračunan iz povprečij posameznih simuliranih podenot, koeficient variacije je bil izračunan za vse vrednosti.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>	<i>D. rerio</i>
Povprečje prostornine [\AA^3]	294	284	284	295	221
95% interval zaupanja [\AA^3]	(255; 332)	(243; 324)	(235; 322)	(261; 328)	(185; 258)
Koeficient variacije	39%	39%	37%	30%	44%

Tabela 7: Medsebojna primerjava povprečij prostornin hidrofobnega žepa. Uporabljen je bil Tukeyev HSD test, kriterij za statistično pomembnost je $p < 0,05$.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>
<i>M. musculus</i>	0,99			
<i>G. gallus</i>	0,99	1,00		
<i>X. laevis</i>	1,00	0,99	0,99	
<i>D. rerio</i>	0,042	0,10	0,10	0,039



Slika 12: Razporeditev prostornin hidrofobnega žepa v EpEX iz različnih vrst. Slika je bila ustvarjena s programom LibreOffice Calc.

Rezultati, ki jih je dal program POVME2, so prikazani na sliki 12 in v tabeli 6, skupaj z njihovimi koeficienti variacije. Medsebojna primerjava povprečij prostornin je prikazana v tabeli 7. Najmanjšo prostornino ima hidrofobni žep v EpEX iz *D. rerio*, ostale vrednosti so blizu druge druge. EpCAM iz *D. rerio* bi zaradi tega lahko manj ugodno vezal večje ligande.

Tabela 8: Medsebojna primerjava koeficientov variacije prostornine hidrofobnega žepa. P-vrednosti so bile dobljene z zankanjem po formuli 2 in popravljene s Holm-Bonferronijevim postopkom. Odebeljene so vrednosti, ki ustrezajo izbranemu kriteriju za statistično pomembnost ($p < 0,05$). Nepopravljene p-vrednosti so v tabeli P3, poglavje Priloge.

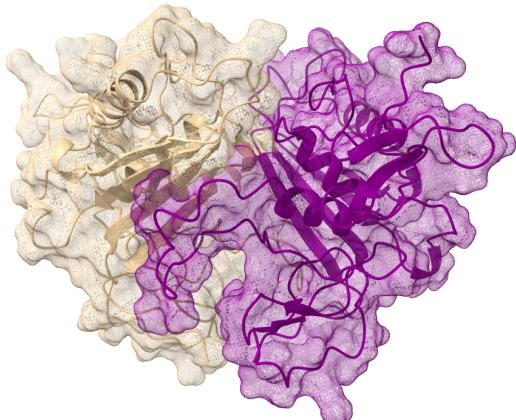
	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>
<i>M. musculus</i>	1,000			
<i>G. gallus</i>	1,000	1,000		
<i>X. laevis</i>	0,024	0,014	0,780	
<i>D. rerio</i>	0,410	0,370	0,780	0,002

Na grafu na sliki 12 je razvidno tudi, da ima hidrofobni žep v *X. laevis* (zeleni stolpci) ožjo porazdelitev vrednosti. To potrdi tudi koeficient variacije (tabela 6, tabela 8), ki je pri *X. laevis* najnižji. To nakazuje večjo rigidnost žepa v *X. laevis*. Po teoriji konformacijske selekcije so rigidnejši receptorji in ligandi bolj specifični in imajo zaradi manjše izgube entropije pri vezavi višjo afiniteto [68–70].

Datoteke z neobdelanimi in delno obdelanimi rezultati tukaj predstavljenih analiz so dostopne na strani Github, povezava je v poglavju Priloge.

4.2 Heterodimer EpCAM:Trop2

4.2.1 Predvidene strukture in potek simulacij



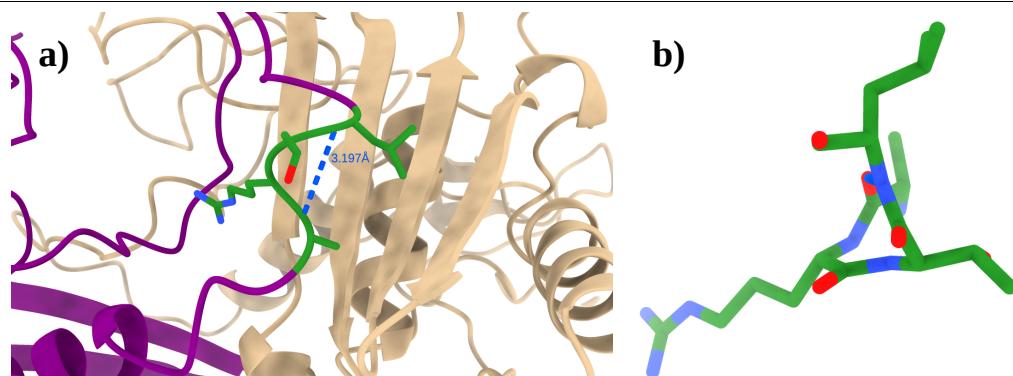
Slika 13: Prikaz strukture *hetdim2*. Bež podenota na levi je EpCAM, vijolična podenota na desni pa Trop2.
Transmembranska in citosolna dela sta bila odrezana.

Za heterodimer EpCAM:Trop2 smo v začetku načrtovali dve z Alphafoldom predvideni strukturi, vendar prvi poskus napovedi strukture ni bil uspešen (struktura *hetdim1*, rezultat prvega poskusa, je imela napačno terciarno strukturo obeh podenot), zato smo napovedali še dve dodatni strukturi (*hetdim2*, ki je prikazan na sliki 13, in *hetdim3*). V nadaljevanju bomo omenjali le *hetdim2* in *hetdim3*.

Vse podenote EpCAM je Alphafold predvidel s podaljšano α -vijačnico tiroglobulinske tip-1 domene, vendar namen tega dela ni bil opazovati obnašanja tiroglobulinske zanke, temveč le analiza

energije dimerizacije, zato jih nismo spremenjali.

V obeh strukturah heterodimera je v tiroglobulinski zanki Trop2 prisotna en obrat dolga 3_{10} -vijačnica (vodikova vez med karbonilom A86 in amidnim vodikom L89), kot je prikazano na sliki 14. Takšne strukture ni v z Alphafoldom predvideni strukturi niti v kristalni strukturi Trop2 homodimera.



Slika 14: 3_{10} -vijačnica v predvideni strukturi *hetdim2*. a) 3_{10} -vijačnica s prikazano vodikovo vezjo. EpEX je bež, Trop2 je vijoličen, 3_{10} -heliks (A86-L89) v Trop2 pa je zelen. b) Pogled skozi 3_{10} -vijačnico, od C- proti N-koncu. Slika je bila ustvarjena z UCSF ChimeraX.

Med potekom simulacij se spremeni v dober obrat dolgo α -vijačnico (vodikove vezi med A86 in V90, v nekaterih simulacijah tudi N85 in L89 ter K84 in T88), ki se v številnih simulacijah na N-koncu nadaljuje v 3_{10} -vijačnico. Vodikove vezi so prisotne med ogrodjem K84–R87 in N85–T88, do neke mere se poleg za α -vijačnico značilnega

para A86–V90 na koncu helične strukture še vedno pojavlja originalni par A86–L89, kar tudi nakazuje na rahlo deformacijo α -heličnega dela in prehajanje v 3_{10} -vijačnico – prehajanje, ki je za 3_{10} -vijačnice in kratke α -vijačnice značilno [71].

4.2.2 Predvidene vezavne energije

Tabela 9: Vezavna energija med podenotama - rezultati Nwat-MM/GBSA in njihove standardne napake. Energije so podane v enotah kcal mol⁻¹.

	EpEX 4MZV	EpEX AlphaFold	Trop2 7PEE verigi BC	Trop2 AlphaFold	hetdim2	hetdim3	TyLoop ^{mut}
Povprečje [kcal mol ⁻¹]	-168	-183	0	0	0	0	-129
Std. napaka [kcal mol ⁻¹]	6	3	6	7	8	11	6

Tabela 10: Povprečja za vsak protein (povprečje simulacij dveh različnih simuliranih struktur, razen za TyLoop^{mut}, kjer je bila simulirana le ena struktura) in njihovi 95% intervali zaupanja (po Studentovi-t porazdelitvi).

	EpEX	Trop2	Heterodimer	TyLoop ^{mut}
Povprečje [kcal mol ⁻¹]	-176	-202	-205	-129
95% IZ [kcal mol ⁻¹]	(-185; -166)	(-213; -192)	(-221; -190)	(-147; -112)

Tabela 11: P-vrednosti, dobljene pri medsebojnih primerjavah vezavnih energij različnih proteinov s Tukeyevim HSD testom. Vrednosti, ki nakazujejo statistično pomembnost ($p < 0,05$), so odebujene.

	EpEX	Trop2	Heterodimer
Trop2	0,0050		
Heterodimer	0,0018	0,98	
TyLoop ^{mut}	7,5e-5	2,0e-8	8,6e-9

Kot je prikazano v tabelah 9 in 10, so rezultati Nwat-MM/GBSA nakazali, da je najmanj ugodna interakcija med podenotama TyLoop^{mut}, kar je bilo pričakovano, saj za razliko od divjega tipa EpEX različica TyLoop^{mut} ne tvori dimera pod pogoji, ki so jih preizkusili Žagar in sod. [23]. Drugačen rezultat bi nakazoval, da uporabljeni metoda ni primerna za naš primer. Sledi divji tip EpEX, Trop2 in heterodimer pa imata zelo podobno energijo vezave. Edino razlika med Trop2 in heterodimerom ni statistično pomembna, kot je prikazano v tabeli 11.

Razlika sprememb proste energije ob tvorbi heterodimera, izračunana po enačbi

$$\Delta \Delta G_{\text{heterodimer}} = \Delta G_{\text{Heterodimer}} - \frac{1}{2} (\Delta G_{\text{EpEX}} + \Delta G_{\text{Trop}_2}) \quad (6),$$

je $-16 \text{ kcal mol}^{-1}$, kar nakazuje, da je heterodimer ugodnejši od homodimerov. Še vedno pa gre za majhno razliko, ki bi se lahko pojavila zaradi netočnosti same metode Nwatt-MM/GBSA (ne upošteva niti razlik v interni strukturi monomera, kot je na primer pojav $\alpha/3_{10}$ -vijačnice v tiroglobulinski zanki heterodimera, niti razlik v konformacijski entropiji) ali polja sil. Nakazuje pa, da je heterodimer mogoč in da je, če se tvori, interakcija med podenotama močna.

Energija, dobljena za EpEX 4MZV, ni primerljiva z vrednostjo iz podpoglavja 4.1.5, kljub temu, da so bile simulacije začete iz iste strukture – uporabili smo namreč nekoliko drugačno verzijo polja sil CHARMM. CHARMM36m, uporabljen v tem delu, solne mostičke, v katerih je pozitivno nabit ostanek arginin, obravnava kot šibkejše v primerjavi s CHARMM36 [58]. Interakcijska površina človeškega EpEX dimera vsebuje štiri arginine, ki tvorijo solne mostičke z nasprotno podenoto, in zato znatna razlika med rezultatoma ni presenetljiva. Tudi stokrat daljši čas simulacije v delu, ki se ukvarja z evolucijo dinamike EpEX, bi lahko prispeval k razlikam, saj bi se v daljšem času simulacije struktura lahko bolj oddaljila od izvorne. To potrjujejo tudi vrednosti RMSD – kot je prikazano na sliki 7, kljub temu da RMSD po ekvilibraciji več ne narašča hitro, še vedno raste, kar je do neke mere opazno tudi pri človeškem EpEX.

Zaradi konformacijskih razlik med Trop2 v homo- in heterodimeru smo izračunali tudi energijo (v vakuumu – za razliko od MM/GBSA tu ni upoštevana solvacija) same tiroglobulinske zanke in sicer za odsek P83–S93, kjer se v heterodimeru nahaja $\alpha/3_{10}$ -vijačnica in se najbolj razlikuje od konformacije v homodimeru Trop2, in za celotno tiroglobulinsko zanko (S81–L103). Krajši odsek P83–S93 ima za 25 kcal mol^{-1} višjo energijo v heterodimeru kot v homodimeru ($-112 \text{ kcal mol}^{-1}$ in $-85 \text{ kcal mol}^{-1}$, $p=0,024$ po Studentovem t-testu). Takšna razlika bi izničila bolj ugodno vezavo obeh podenot v heterodimeru, vendar analiza celotne tiroglobulinske zanke ne pokaže razlik (32 kcal mol^{-1} za heterodimer in 33 kcal mol^{-1} za homodimer, $p=0,93$). Očitno tvorba vijačnice v heterodimeru omogoči ugodnejšo konformacijo preostalih delov zanke, kar odtehta neugodnost vijačnice.

Ker bi se potencialne neugodnejše strukture v heterodimeru lahko sprostile v monomerni obliki, smo pripravili še simulacije posameznih monomerov iz struktur 4MZV, 7PEE verigi B in C ter struktur heterodimera in njihove MM/GBSA energije iz 4. in 5. nanosekunde primerjali z energijo monomerov, simuliranih kot podenote kompleksa. Ker se postopek, ki smo ga uporabili za izbiro molekul vode, zanaša na prisotnost obeh podenot dimera, smo v tem delu pobrisali vse vode. Energije posebej

simuliranih monomerov so bile nižje kot energije podenot dimera, vendar se razlike med energijami monomerov in podenot dimera iz struktur homodimerov in iz struktur heterodimerov praktično ne razlikujejo, kot je prikazano v tabeli 12.

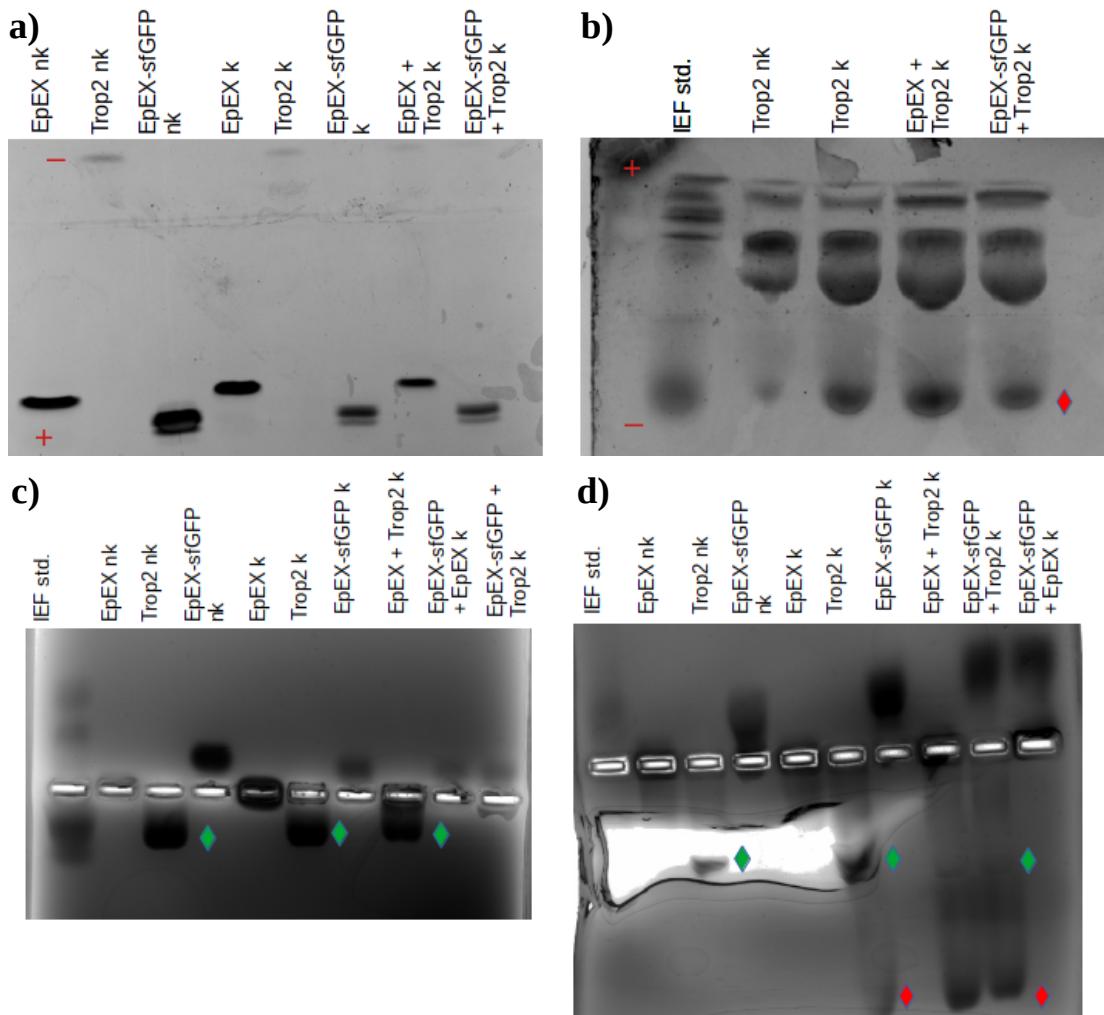
Tabela 12: Popravek iz MM/GBSA s tremi trajektorijami. Prikazani sta razlika energije med monomeri, simuliranimi kot del dimera, in neodvisno simuliranimi monomeri. Vrednosti za homo/heterodimer so vsote povprečja obeh podenot.

	EpEX 4MZV	EpEX <i>hetdim2</i>	EpEX <i>hetdim3</i>	Trop2 7PEE	Trop2 <i>hetdim2</i>	Trop2 <i>hetdim3</i>	Homo- dimer (vsota)	Hetero- dimer (vsota)
ΔG [kcal mol ⁻¹]	27,3	12,8	30,8	25,8	20,5	41,3	53	52,7

Poleg termodinamike je pri tekmovanju različnih proteinov za isto interakcijsko površino pomembna tudi kinetika. Ni nujno, da bodo razmerja med različnimi kompleksi doseгла ravnotežje [50],[72]. Kinetični nadzor interakcije je še posebej pogost v primerih, ko je disociacija počasna [72].

Datoteke z neobdelanimi in delno obdelanimi rezultati tukaj predstavljenih analiz so dostopne na strani Github, povezava je v poglavju Priloge.

4.2.3 Laboratorijsko določanje obstoja heterodimera



Slika 15: **a)** in **b)**: Nativna poliakrilamidna elektroforeza pri pH 8,8 s PhastSystem, z geloma PhastGel Gradient 8–25. **a)** Gel, na katerem je bila elektroforeza izvedena v smeri anode in **b)** v smeri katode (rdeča + in -). Nk pomeni nekisan vzorec, k pa kisan vzorec. Rdeč diamant označuje denaturiran Trop2. **c)** in **d)**: Agarozna nativna elektroforeza, zelen diamant je nativen Trop2. Vzorci na sliki **c)** so bili zamrznjeni in odtajani, na sliki **d)** pa ne.

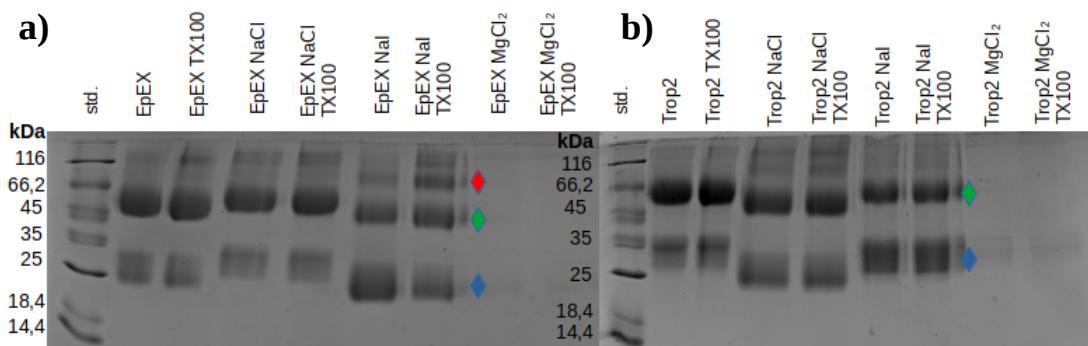
4.2.4 Trop2 pri pH 4,5 reverzibilno denaturira

Na sliki 15a je vidno, da EpEX po zakisanju in nevtralizaciji potuje enako kot nekisan protein. EpEX-sfGFP tvori dve lisici, kar je verjetno posledica heterogene glikozilacije (EpEX-sfGFP nima mutiranih N-glikozilacijskih mest [59], EpEX ima mutirane vse tri [14], Trop2 pa dve od štirih [15]). Na sliki 15b je gel, na katerega so bili nanešeni le vzorci, za katere smo pričakovali, da bodo potovali proti katodi. Pri zakisanih in nevtraliziranih vzorcih se je pojavila nova lisa, ki je najverjetneje denaturiran Trop2.

Te lise nismo opazili na agarozni nativni elektroforezi na sliki 15c. Glavna razlika je bila, da smo vzorec pred nanosom na gel dvakrat zamrznili pri -20 °C in odtajali. Na sliki 15d se ta lisa spet pojavi. Iz tega smo sklepali, da Trop2 pri pH 4,5 denaturirira in da je ta denaturacija s časom in/ali zamrzovanjem reverzibilna.

Na sliki 17a je še dodaten dokaz za denaturacijo Trop2 pri pH 4,5. Pri prečnem povezovanju se izkaže, da med zakisanjem denaturiran Trop2, ki ima sicer na nativni elektroforezi višjo mobilnost, nespecifično agregira in se prečno poveže v skupke, ki so preveliki, da bi vstopili v zbiralni ali ločevalni gel. Pri zakisanju na pH 5,5 na sliki 17b tega ne opazimo.

4.2.4.1 Disociacija z dodatkom soli



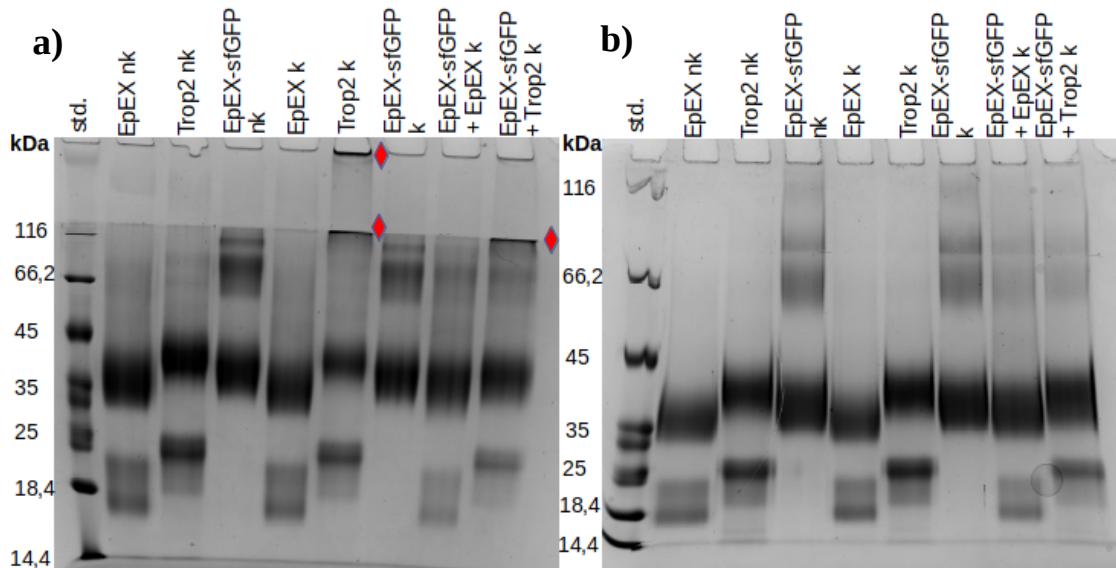
Slika 16: NaDS-PAGE produktov prečnega povezovanja v prisotnosti različnih soli.

Moder diamant je (ne prečno-povezan) monomer, zelen diamant prečno povezan dimer, rdeč pa je prečno povezan trimer. TX100 je detergent Triton X-100. **a)** EpEX **b)** Trop2

Na sliki 16 je videti, da razmerje med intenzivnostmi dimera in monomera praktično enako pri različnih koncentracijah NaCl, pri MgCl₂ pa se je protein obarjal in so lise komajda vidne. Pri NaI (žepi 4A, 4B, 9A in 9B) pa je lisa, pripadajoča (ne prečno-povezanemu) monomeru, precej intenzivnejša, tako za EpEX kot tudi za Trop2. To bi lahko bilo ali posledica tega, da NaI kot kaotropna sol sproži disociacijo dimera, ali tega, da NaI inhibira reakcijo prečnega povezovanja.

Pri EpEX z 1M NaI je opazna tudi lisa, ki ustreza velikosti trimera. Ta lisa je mnogo močnejša pri prečnem povezovanju v prisotnosti detergenta Triton X-100 (rdeč diamant na sliki 16). V prisotnosti detergenta so vidni tudi prečno povezani višji oligomeri. Ker ti niso vidni niti pri kontroli brez dodatne soli niti pri Trop2 v prisotnosti NaI, sklepamo, da ne gre za nespecifično prečno povezovanje, temveč da EpEX pod temi pogojimi denaturira in nespecifično agregira, kar je še precej bolj očitno v prisotnosti detergenta.

4.2.4.2 Analiza interakcij

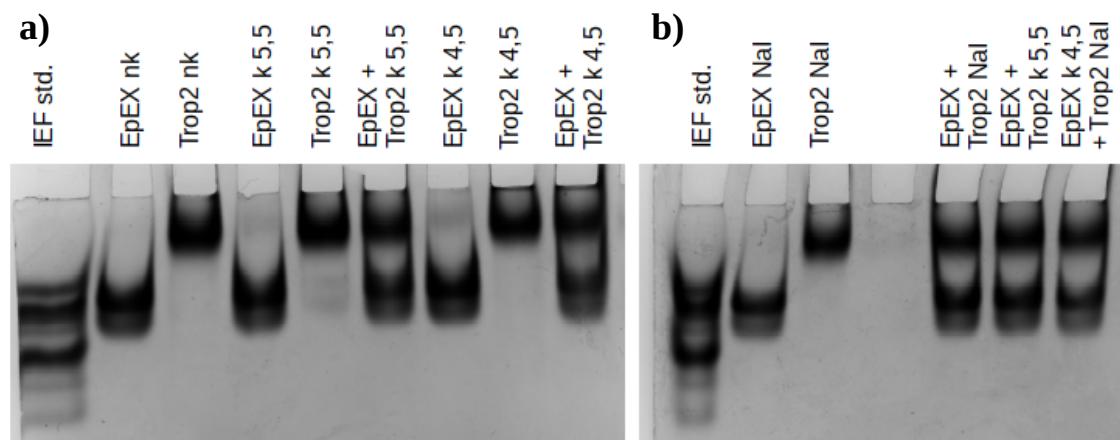


Slika 17: NaDS-PAGE produktov prečnega povezovanja z EpEX-sfGFP. a) Vzorci, zakisani na pH 4,5. Rdeč diamant je denaturiran in agregiran Trop2, ki je po prečnem povezovanju prevelik, da bi vstopil iz iz žepka v 5% zbiralni gel oziroma iz zbiralnega v 10% ločevalni gel. b) Vzorci, zakisani na pH 5,5. Ločevalni gel je 7%.

Pri prečnem povezovanju smo opazili tudi dimer EpEX-sfGFP, ki naj sicer ne bi dimeriziral. To bi lahko bila posledica prisotnosti sfGFP, če interagira preko sfGFP, ki je šibek dimer [73], oziroma če se sfGFP lažje nespecifično prečno povezuje, ali pa šibke interakcije med dvema EpEX-sfGFP preko EpEX, če ga prisotnost sfGFP ne ovira preveč.

Nismo pa dobili lise, ki bi ustrezala velikosti dimera EpEX ali Trop2 z EpEX-sfGFP, torej ne moremo potrditi, da naš pristop z zakisanjem in neutralizacijo dejansko deluje. Do tega bi lahko prišlo bodisi, ker sfGFP sterično ovira tvorbo dimera tudi z EpEX s prostim koncem, bodisi ker pri našem postopku dimera EpEX in/ali Trop2 ne disociirata.

Na sliki 18 ni pri mešanicah EpEX in Trop2 razvidne nobene lise med lisama EpEX in Trop2. EpEX in Trop2 sta podobne velikosti in oblike, imata pa različna naboja in predvidoma bi se heterodimer nahajal med njima. Iz tega sklepamo, da se heterodimer po naših postopkih priprave v mešanicah ne tvori.



Slika 18: Nativna PAGE v smeri anode pri pH 7,5 po protokolu Gallagherja [62]. Vsi vzorci so bili zmešani z nanašalnim pufrom in nato zamrznjeni pri -20 °C za dva dni. a) Zakisani in nevtralizirani vzorci. Številka pomeni pH, na katerega je bil vzorec zakisan. b) Dializirani vzorci.

Vsi vzorci pri nativni elektroforezi po protokolu Gallagherja [62] so potovali v smeri anode, kljub temu da ima Trop2 izoelektrično točko nad 8,8 (slika 15ab, kjer je bila nativna elektroforeza izvedena pri pH 8,8 in je Trop2 potoval proti katodi). To bi lahko bilo zaradi vezave bromofenol modrega na proteine – negativno nabito bromofenol modro se preko pretežno elektrostatskih interakcij veže na proteine [74]. Možno je, da se je zaradi dolgega časa stika (zamrzovanje vzorcev z nanašalnim pufom in odtajanje pred nanosom na gel) barvilo vezalo na proteine (na to nakazuje tudi to, da je bila med potekom elektroforeze fronta bromofenol modrega zelo šibka), z večjim številom vezanih molekul na proteinih z več pozitivno nabitim mestom, in jim dalo neto negativen naboj – podobno kot na elektroforetsko mobilnost deluje vezava barvila Coomassie Brilliant Blue pri modri nativni PAGE.

ALJAŽ SIMONIČ: MOLEKULSKA DINAMIKA PROTEINA EPCAM IZ RAZLIČNIH VRST IN NJEGOVA
INTERAKCIJA S PARALOGNIM PROTEINOM TROP2

5 Zaključek

5.1 EpCAM iz različnih vrst

Hipotezo, da bo tiroglobulinska zanka v EpCAM iz *D. rerio* fleksibilnejša, smo zavrnili, saj niti RMSF aminokislin v tiroglobulinski zanki niti koeficient variacije interakcijske površine tiroglobulinske zanke v EpEX iz *D. rerio* ni višji od EpEX iz ostalih vrst.

EpEX iz *X. laevis* ima najmanjšo interakcijsko površino, EpEX iz *D. rerio* pa najmanjšo kontaktno površino tiroglobulinske zanke s C-končno domeno nasprotne podenote.

Poleg tega smo ugotovili, da sta v EpEX iz *G. gallus* podenoti najmočneje povezani, v EpEX iz *D. rerio* pa najšibkeje.

Ugotovili smo, da ima hidrofobni žep v EpCAM iz *D. rerio* manjšo prostornino od ostalih, hidrofobni žep v *X. laevis* pa je rigidnejši.

Delo bi bilo mogoče izboljšati z uporabo več nekoliko krajših simulacij, na primer 10 simulacij dolžine 250 ns. Posamezne simulacije, tako kratke [64] kot tudi dolge [65], se lahko ujamejo znotraj lokalnih energijskih minimumov, kar smo pri naših simulacijah opazili tudi mi. Več simulacij z različnimi naključno določenimi začetnimi hitrostmi bolje raziskuje konformacijski prostor proteina, dokler so vsaj zadosti dolge, da se v času simulacije želen dogodek sploh lahko zgodi (daljše za velike konformacijske spremembe in vezavo liganda, mnogo kraje za izračun energij vezave) [65]. Prav tako bi večje število krajših simulacij ne samo bolje preiskalo dostopni konformacijski prostor, temveč bi tudi omogočilo večjo statistično zanesljivost.

5.2 Heterodimer EpCAM:Trop2

S simulacijami in eksperimentalnimi metodami smo preučevali možnost tvorbe heterodimera EpCAM:Trop2.

Energije, dobljene iz simulacij molekulske dinamike, nakazujejo, da je povezava med podenotama v hipotetičnem heterodimeru EpCAM:Trop2 močna in bi lahko bila njegova tvorba energijsko ugodna, v laboratoriju pa zaradi omejitev uporabljenega pristopa nismo dobili rezultatov, na podlagi katerih bi lahko zagotovo zaključili, ali se heterodimer tvori ali ne.

ALJAŽ SIMONIČ: MOLEKULSKA DINAMIKA PROTEINA EPCAM IZ RAZLIČNIH VRST IN NJEGOVA
INTERAKCIJA S PARALOGNIM PROTEINOM TROP2

6 Literatura

- [1] F. Fagotto, A. Aslemarz: EpCAM cellular functions in adhesion and migration, and potential impact on invasion: A critical review. *Biochim. Biophys. Acta BBA - Rev. Cancer* **2020**, 1874, 188436.
- [2] A. R. A. McDougall, M. Tolcos, S. B. Hooper, T. J. Cole, M. J. Wallace: Trop2: From development to disease. *Dev. Dyn.* **2015**, 244, 99–109.
- [3] S. Lenárt, P. Lenárt, J. Šmarda, J. Remšík, K. Souček, P. Beneš: Trop2: Jack of All Trades, Master of None. *Cancers* **2020**, 12, 3328.
- [4] M. Herlyn, Z. Steplewski, D. Herlyn, H. Koprowski: Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proc. Natl. Acad. Sci.* **1979**, 76, 1438–1442.
- [5] S. V. Litvinov, M. P. Velders, H. A. Bakker, G. J. Fleuren, S. O. Warnaar: EpCAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J. Cell Biol.* **1994**, 125, 437–446.
- [6] O. Gires, M. Pan, H. Schinke, M. Canis, P. A. Baeuerle: Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years? *Cancer Metastasis Rev.* **2020**, 39, 969–987.
- [7] A. Gaber, S. J. Kim, R. M. Kaake, M. Benčina, N. Krogan, A. Šali, M. Pavšič, B. Lenarčič: EpCAM homo-oligomerization is not the basis for its role in cell-cell adhesion. *Sci. Rep.* **2018**, 8, 13269.
- [8] A. Gaber, B. Lenarčič, M. Pavšič: Current View on EpCAM Structural Biology. *Cells* **2020**, 9, 1361.
- [9] M. Lipinski, D. R. Parks, R. V. Rouse, L. A. Herzenberg: Human trophoblast cell-surface antigens defined by monoclonal antibodies. *Proc. Natl. Acad. Sci.* **1981**, 78, 5147–5150.
- [10] A. J. Linnenbach, B. A. Seng, S. Wu, S. Robbins, M. Scollon, J. J. Pyrc, T. Druck, K. Huebner: Retroposition in a family of carcinoma-associated antigen genes. *Mol. Cell. Biol.* **1993**, 13, 1507–1515.
- [11] M. Novinec, D. Kordiš, V. Turk, B. Lenarčič: Diversity and Evolution of the Thyroglobulin Type-1 Domain Superfamily. *Mol. Biol. Evol.* **2006**, 23, 744–755.
- [12] J. Pei, N. V. Grishin: Expansion of divergent SEA domains in cell surface proteins and nucleoporin 54. *Protein Sci.* **2017**, 26, 617–630.
- [13] S. H. Chen, D. R. Bell: Evolution of Thyroglobulin Loop Kinetics in EpCAM. *Life* **2021**, 11, 915.
- [14] M. Pavšič, G. Gunčar, K. Djinović-Carugo, B. Lenarčič: Crystal structure and its bearing towards an understanding of key biological functions of EpCAM. *Nat. Commun.* **2014**, 5, 4764.
- [15] M. Pavšič: Trop2 Forms a Stable Dimer with Significant Structural Differences within the Membrane-Distal Region as Compared to EpCAM. *Int. J. Mol. Sci.* **2021**, 22, 10640.
- [16] M. Pavšič, B. Lenarčič: Expression, crystallization and preliminary X-ray characterization of the human epithelial cell-adhesion molecule ectodomain. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **2011**, 67, 1363–1366.
- [17] T. Vidmar, M. Pavšič, B. Lenarčič: Biochemical and preliminary X-ray characterization of the tumor-associated calcium signal transducer 2 (Trop2) ectodomain. *Protein Expr. Purif.* **2013**, 91, 69–76.

- [18] M. Pavšič, G. Ilc, T. Vidmar, J. Plavec, B. Lenarčič: The cytosolic tail of the tumor marker protein Trop2 - a structural switch triggered by phosphorylation. *Sci. Rep.* **2015**, *5*, 10324.
- [19] J. B. Casaletto, M. L. Geddie, A. O. Abu-Yousif, K. Masson, A. Fulgham, A. Boudot, T. Maiwald, J. D. Kearns, N. Kohli, S. Su, et al.: MM-131, a bispecific anti-Met/EpCAM mAb, inhibits HGF-dependent and HGF-independent Met signaling through concurrent binding to EpCAM. *Proc. Natl. Acad. Sci.* **2019**, *116*, 7533–7542.
- [20] J. Hočevar: Vloga hidrofobnega žepa EpCAM in Trop2 pri signaliziranju preko regulirane intramembranske cepitve. Diplomsko delo, Fakulteta za kemijo in kemijsko tehnologijo, Univerza v Ljubljani, Ljubljana, 2018.
- [21] R. Szabo, J. M. Ward, F. Artunc, T. H. Bugge: EPCAM and TROP2 share a role in claudin stabilization and development of intestinal and extraintestinal epithelia in mice. *Biol. Open* **2022**, *11*, bio059403.
- [22] C.-J. Wu, M. Lu, X. Feng, G. Nakato, M. C. Udey: Matriptase Cleaves EpCAM and TROP2 in Keratinocytes, Destabilizing Both Proteins and Associated Claudins. *Cells* **2020**, *9*, 1027.
- [23] T. Žagar, M. Pavšič, A. Gaber: Destabilization of EpCAM dimer is associated with increased susceptibility towards cleavage by TACE. *PeerJ* **2021**, *9*, e11484.
- [24] G. Nakato, S. Morimura, M. Lu, X. Feng, C. Wu, M. C. Udey: Amelioration of Congenital Tufting Enteropathy in EpCAM (TROP1)-Deficient Mice via Heterotopic Expression of TROP2 in Intestinal Epithelial Cells. *Cells* **2020**, *9*, 1847.
- [25] K. S. Foley, P. W. Young: An analysis of splicing, actin-binding properties, heterodimerization and molecular interactions of the non-muscle α -actinin. *Biochem. J.* **2013**, *452*, 477–488.
- [26] C. Menzel, S. Dbel: Abciximab, Arcitumomab, Basiliximab, Capromab, Cotara, Daclizumab, Edrecolomab, Ibritumomab, Igovomab, Nofetumomab, Satumomab, Sulesomab, Tositumomab, and Votumumab. In *Handbook of Therapeutic Antibodies*; Dbel, S., Ed.; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2007; pp 1131–1147.
- [27] J. Garousi, A. Orlova, F. Y. Frejd, V. Tolmachev: Imaging using radiolabelled targeted proteins: radioimmunodetection and beyond. *EJNMMI Radiopharm. Chem.* **2020**, *5*, 16.
- [28] Nofetumomab, Dr. Karl Thomae GmbH
https://www.accessdata.fda.gov/drugsatfda_docs/label/1996/nofedrk082096lb.pdf (pridobljeno 22. 10. 2022).
- [29] Y. Y. Syed: Sacituzumab Govitecan: First Approval. *Drugs* **2020**, *80*, 1019–1025.
- [30] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin: UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- [31] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin: UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82.
- [32] The UniProt Consortium: UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489.
- [33] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, et al.: Fast, scalable generation of high-

- quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.
- [34] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton: Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191.
- [35] M. V. Shapovalov, R. L. Dunbrack: A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19*, 844–858.
- [36] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al.: Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- [37] R. Evans, M. O’Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, et al.: Protein complex prediction with AlphaFold-Multimer. *bioRxiv* October 4, 2021, p 2021.10.04.463034.
- [38] Google Colaboratory
<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb> (pridobljeno 10. 10. 2022).
- [39] A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton: JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **2015**, *43*, W389–W394.
- [40] A. Šali, T. L. Blundell: Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- [41] W. Humphrey, A. Dalke, K. Schulten: VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
- [42] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein: Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- [43] J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, et al.: Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
- [44] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen: Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- [45] S. Miyamoto, P. A. Kollman: Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- [46] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell: Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- [47] ANOVA Calculator - One Way ANOVA and Tukey HSD test
<https://www.statskingdom.com/180Anova1way.html> (pridobljeno 6. 4. 2023).
- [48] T. Hesterberg: Bootstrap. *WIREs Comput. Stat.* **2011**, *3*, 497–526.
- [49] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang, T. Hou: End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119*, 9478–9508.
- [50] T. Siebenmorgen, M. Zacharias: Computational prediction of protein–protein binding affinities. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1448.

- [51] S. Wong, R. E. Amaro, J. A. McCammon: MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein–Protein Interface. *J. Chem. Theory Comput.* **2009**, 5, 422–429.
- [52] I. Maffucci, A. Contini: Improved Computation of Protein–Protein Relative Binding Energies with the Nwat-MMGBSA Method. *J. Chem. Inf. Model.* **2016**, 56, 1692–1704.
- [53] R. M. Crean, C. R. Pudney, D. K. Cole, M. W. van der Kamp: Reliable In Silico Ranking of Engineered Therapeutic TCR Binding Affinities with MMPB/GBSA. *J. Chem. Inf. Model.* **2022**, 62, 577–590.
- [54] D. G. Covell, A. Wallqvist: Analysis of protein-protein interactions and the effects of amino acid mutations on their energetics. The importance of water molecules in the binding epitope. *J. Mol. Biol.* **1997**, 269, 281–297.
- [55] A. Onufriev, D. Bashford, D. A. Case: Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Struct. Funct. Bioinforma.* **2004**, 55, 383–394.
- [56] J. D. Durrant, C. A. F. de Oliveira, J. A. McCammon: POVME: An algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model.* **2011**, 29, 773–776.
- [57] J. D. Durrant, L. Votapka, J. Sørensen, R. E. Amaro: POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **2014**, 10, 5047–5056.
- [58] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, A. D. MacKerell: CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, 14, 71–73.
- [59] J. Hočvar: *Ligand-vezavne lastnosti hidrofobnega žepa EpCAM in njegov vpliv na regulirano intramembransko cepitev*; Magistrsko delo; Univerza v Ljubljani, Fakulteta za kemijo in kemijsko tehnologijo, 2021.
- [60] pH Calculator - Calculates pH of a Solution <http://www.webqc.org/phsolver.php> (pridobljeno 17. 4. 2023).
- [61] R. Kim: Native Agarose Gel Electrophoresis of Multiprotein Complexes. *Cold Spring Harb. Protoc.* **2011**, 2011, pdb.prot4558.
- [62] S. R. Gallagher: One-Dimensional Electrophoresis Using Nondenaturing Conditions. *Curr. Protoc. Protein Sci.* **2018**, 94, e73.
- [63] C. Nick Pace, J. Martin Scholtz: A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998**, 75, 422–427.
- [64] L. S. D. Caves, J. D. Evanseck, M. Karplus: Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci.* **1998**, 7, 649–666.
- [65] P. V. Coveney, S. Wan: On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys. Chem. Chem. Phys.* **2016**, 18, 30236–30240.
- [66] S. Genheden, U. Ryde: The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, 10, 449–461.
- [67] V. Ekberg, U. Ryde: On the Use of Interaction Entropy and Related Methods to Estimate Binding Entropies. *J. Chem. Theory Comput.* **2021**, 17, 5379–5391.
- [68] D. Alvarez-Garcia, X. Barril: Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *J. Chem. Theory Comput.* **2014**, 10, 2608–2614.

- [69] I. F. Thorpe, C. L. Brooks: Molecular evolution of affinity and flexibility in the immune system. *Proc. Natl. Acad. Sci.* **2007**, *104*, 8821–8826.
- [70] M. Münz, J. Hein, P. C. Biggin: The Role of Flexibility and Conformational Selection in the Binding Promiscuity of PDZ Domains. *PLOS Comput. Biol.* **2012**, *8*, e1002749.
- [71] G. L. Millhauser: Views of Helical Peptides: A Proposal for the Position of 310-Helix along the Thermodynamic Folding Pathway. *Biochemistry* **1995**, *34*, 3873–3877.
- [72] H.-X. Zhou: Rate theories for biologists. *Q. Rev. Biophys.* **2010**, *43*, 219–293.
- [73] D. J. Scott, N. J. Gunn, K. J. Yong, V. C. Wimmer, N. A. Veldhuis, L. M. Challis, M. Haidar, S. Petrou, R. A. D. Bathgate, M. D. W. Griffin: A Novel Ultra-Stable, Monomeric Green Fluorescent Protein For Direct Volumetric Imaging of Whole Organs Using CLARITY. *Sci. Rep.* **2018**, *8*, 667.
- [74] M. Subramaman, B. S. Sheshadri, M. P. Venkatappa: Interaction of Lysozyme with Dyes: II. Binding of Bromophenol Blue. *J. Biochem. (Tokyo)* **1984**, *96*, 245–252.

ALJAŽ SIMONIČ: MOLEKULSKA DINAMIKA PROTEINA EPCAM IZ RAZLIČNIH VRST IN NJEGOVA
INTERAKCIJA S PARALOGNIM PROTEINOM TROP2

7 Priloge

Poravnavo aminokislinskih zaporedij, skripte, strukture in rezultati analiz simulacij so dostopni na spletni strani <https://github.com/AljazSimonic/Diplomsko-delov/tree/main/>

Tabela P1: P-vrednosti, dobljene pri medsebojnih primerjavah interakcijskih površin različnih proteinov s Tukeyevim HSD testom. Vrednosti, ki nakazujejo statistično pomembnost, so odebujene. Kriterij za statistično pomembnost je $p < 0,05$.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>
<i>M. musculus</i>	0,8740			
<i>G. gallus</i>	0,8660	0,3420		
<i>X. laevis</i>	0,0030	0,0246	0,0003	
<i>D. rerio</i>	0,9990	0,9250	0,8010	0,0041

Tabela P2: Nepopravljene P-vrednosti za primerjavo koeficientov variacije interakcijske površine, dobljene z zankanjem.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>
<i>M. musculus</i>	0,2940			
<i>G. gallus</i>	0,0530	0,3830		
<i>X. laevis</i>	0,3150	0,0766	0,0002	
<i>D. rerio</i>	0,4240	0,0866	0,0108	0,9230

Tabela P3: Nepopravljene P-vrednosti za primerjavo koeficientov variacije prostornine hidrofobnega žepa, dobljene z zankanjem.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>G. gallus</i>	<i>X. laevis</i>
<i>M. musculus</i>	0,9920			
<i>G. gallus</i>	0,6620	0,6600		
<i>X. laevis</i>	0,0030	0,0016	0,1570	
<i>D. rerio</i>	0,0683	0,0524	0,1560	0,0002