

**Sveučilište u Zagrebu
Prehrambeno-biotehnološki fakultet
Sveučilišni prijediplomski studij Biotehnologija**

Mihovil Pupačić
0058217141

**IN SILICO PREDIKCIJA AKTIVNOSTI PEPTIDA
ZAVRŠNI RAD**

Predmet: Bioinformatika

Mentor: prof. dr. sc. Antonio Starčević

Zagreb, 2024.

TEMELJNA DOKUMENTACIJSKA KARTICA

Završni rad

Sveučilište u Zagrebu
Prehrambeno-biotehnološki fakultet
Sveučilišni prijediplomski studij Biotehnologija

Zavod za biokemijsko inženjerstvo
Laboratorij za bioinformatiku

Znanstveno područje: Biotehničke znanosti
Znanstveno polje: Biotehnologija

In silico predikcija aktivnosti peptida

Mihovil Pupačić, 0058217141

Sažetak:

Bioinformatički alati imaju sve širu primjenu u području analize proteina, njihovih sekvenci te aktivnosti. Na temelju sastava polipeptidnog lanca, međusobnih interakcija aminokiselina u njemu te ostalih parametara, razni bioinformatički alati utemeljeni na elementima strojnog učenja omogućuju pobližu predikciju bioaktivnosti i toksičnosti proteina. U ovom radu nasumično je odabrano 100 proteina iz *ToxProt* baze podataka te je svakom pronađen homolog BLAST pretraživanjem. Izabrani homolozi su proteini koji još nisu klasificirani unutar *ToxProt* baze podataka te pokazuju što nižu E-vrijednost jer to ukazuje na visoki stupanj homologije. Za svaki od ukupno 200 proteina je provedena predikcija bioaktivnosti pomoću bioinformatičkog alata *PeptideRanker* te predikcija toksičnosti pomoću *ToxinPred2*. Proteini sa visokim stupnjem homologije i sličnim rezultatima predikcija su dodatno 3D generirani *AlphaFold* alatom, čime je njihova sličnost potvrđena. S obzirom na homolognost sekvence te 3D strukture proteina i njegovog homologa bilo je moguće predvidjeti aktivnost homologa kojemu je do sada funkcija još bila nepoznata.

Ključne riječi: bioaktivnost, toksin, homologija, ToxProt, BLAST, PeptideRanker, ToxinPred2, AlphaFold

Rad sadrži: 35 stranica, 22 slika, 1 tablica, 34 literaturnih navoda, 0 priloga

Jezik izvornika: hrvatski

Rad je u tiskanom i elektroničkom obliku pohranjen u knjižnici Sveučilišta u Zagrebu Prehrambeno-biotehnološkog fakulteta, Kačićeva 23, 10 000 Zagreb

Mentor: prof. dr. sc. Antonio Starčević

Datum obrane: 16.09.2024.

BASIC DOCUMENTATION CARD

Undergraduate thesis

University of Zagreb
Faculty of Food Technology and Biotechnology
University undergraduate study Biotechnology

Department of biochemical engineering
Laboratory for bioinformatics

Scientific area: Biotechnical Sciences
Scientific field: Biotechnology

In silico prediction of peptide bioactivity

Mihovil Pupačić, 0058217141

Abstract:

Bioinformatics tools are increasingly applied in the analysis of proteins, their sequences, and activities. Based on the composition of polypeptide chains, amino acid interactions, and other parameters, machine learning-based bioinformatics tools enable predictions of protein bioactivity and toxicity. In this study, 100 proteins were randomly selected from the ToxProt database, and homologs were identified using BLAST search. These homologs, not yet classified in ToxProt, exhibited low E-values, indicating a high degree of homology. Bioactivity was predicted for all 200 proteins using PeptideRanker, and toxicity using ToxinPred2. Proteins with high homology and similar prediction results were further 3D-modeled with AlphaFold, confirming their similarity. Based on sequence homology and 3D structure comparison, it was possible to predict the activity of homologs whose functions were previously unknown.

Keywords: Bioactivity, toxin, homology, ToxProt, BLAST, PeptideRanker, ToxinPred2, and AlphaFold.

Thesis contains: 35 pages, 22 figures, 1 tables, 34 references, 0 supplements

Original in: Croatian

Thesis is deposited in printed and electronic form in the Library of the University of Zagreb, Faculty of Food Technology and Biotechnology, Kačićeva 23, 10 000 Zagreb

Mentor: Antonio Starčević, PhD, Full Professor

Thesis defended: 16.09.2024.

SADRŽAJ

1. UVOD	3
2. TEORIJSKI DIO	4
3. EKSPERIMENTALNI DIO	13
3.1. MATERIJALI	13
3.1.1. BAZA PODATAKA - <i>UNIPROT</i>	13
3.1.2. BLAST (<i>BASIC LOCAL ALIGNMENT SEARCH TOOL</i>).....	14
3.1.3. <i>PYTHON</i>	16
3.1.4. <i>SCIKIT-LEARN</i>	17
3.1.5. <i>PANDAS</i>	17
3.1.6. <i>NUMPY</i>	18
3.1.7. <i>ALPHAFOLD</i>	18
3.2. METODE.....	19
3.2.1. ODABIR PODATAKA	19
3.2.2. ODABIR HOMOLOGA	19
3.2.3. PREDIKCIJA BIOAKTIVNOSTI – <i>PEPTIDERANKER</i>	21
3.2.4. PREDIKCIJA TOKSIČNOSTI – <i>TOXINPRED2</i>	22
3.2.5. <i>ALPHAFOLD</i>	23
4. REZULTATI I RASPRAVA	26
4.1. PREDIKCIJA BIOAKTIVNOSTI POMOĆU ALATA <i>PEPTIDERANKER</i>	26
4.1.1. UTJECAJ DULJINE POLIPEPTIDNOG LANCA NA TOČNOST PREDIKCIJE BIOAKTIVNOSTI..	26

4.1.2. UTJECAJ AMINOKISLINSKOG SASTAVA POLIPEPTIDNOG LANCA NA BIOAKTIVNOST PROTEINA.....	27
4.1.3. UTJECAJ E-VRIJEDNOSTI NA USPJEŠNOST PREDIKCIJE BIOAKTIVNOSTI HOMOLOGNIH PROTEINA.....	28
4.1.4. PRIMJERI USPJEŠNE PREDIKCIJE BIOAKTIVNOSTI HOMOLOGNIH PROTEINA.....	29
4.2. PREDIKCIJA TOKSIČNOSTI POMOĆU ALATA <i>TOXINPRED2</i>	30
4.2.1. SASTAV TOKSIČNIH PEPTIDA I UTJECAJ VISOKE E-VRIJEDNOSTI NA TOČNOST PREDIKCIJE TOKSIČNOSTI PROTEINA.....	31
5. ZAKLJUČCI.....	31
6. POPIS LITERATURE	32

1. UVOD

Peptidi su kratki lanci aminokiselina međusobno povezani peptidnim vezama koji imaju ključne uloge u mnogim biološkim procesima. Obavljaju razne uloge, pa tako razlikujemo one koji sudjeluju u staničnoj signalizaciji, imunološkoj obrani organizma, neurotransmisiji, regulaciji apetita, angiogenezi, protuupalnim procesima te antioksidativnim aktivnostima [1]. Uloga peptida u organizmu može se usporediti s ulogom proteina, koji su također polimeri aminokiselina, no proteini su sastavljeni od znatno dužih polipeptidnih lanaca te imaju složenije strukture. Peptidi su obično kraći, s manje od 50 aminokiselina, dok proteini mogu imati stotine do tisuće aminokiselina [2].

Biološki aktivni peptidi oni su peptidi koji imaju specifične fiziološke ili farmakološke učinke u organizmu. Ti peptidi mogu aktivirati ili inhibirati određene biološke funkcije, djelujući kao hormoni, neurotransmiteri ili antimikrobni agensi. Njihova aktivnost može biti korisna u terapijskom kontekstu, kao što je liječenje bolesti, kontrola infekcija i regulacija metaboličkih procesa. Pokazali su se korisnima u procesu dizajniranja lijekova, ponajviše antibiotika, kod kojih je bilo potrebno nadjačati utjecaj rezistencije bakterije na već postojeći antibiotik. Također, peptidi su pokazali visoki potencijal u proizvodnji antiviralnih, antifungalnih te antiparazitskih lijekova jer tako proizvedeni lijekovi imaju veću mogućnost prevladavanja utjecaja bakterijske rezistencije [3].

Koristeći razne bioinformatičke alate, zamisao je ovoga rada predvidjeti do sad još nepoznatu funkciju određenih peptida i proteina. Predikcija strukture i biološke aktivnosti peptida *in silico*, tj. u stvarnom biološkom okruženju, od velike je važnosti jer poznavanje strukture peptida može pomoći pri identifikaciji mjesta aktivnosti ili mehanizama djelovanja tih peptida. Zatim, predikcija biološke aktivnosti omogućava identifikaciju potencijalno korisnih peptida prije njihove sinteze i testiranja, čime se štedi vrijeme i resursi u istraživačkom procesu. Konačno, predikcija može pomoći u dizajnu novih peptida s optimalnim terapijskim svojstvima, pružajući nove mogućnosti za liječenje različitih bolesti. Razvoj novih bioinformatičkih alata i metoda za predikciju biološke aktivnosti peptida predstavlja značajan napredak u biomedicinskom istraživanju. Ovi alati omogućavaju detaljnu analizu sekvenci te stvaranje trodimenzionalnih struktura prema navedenim sekvencama što pruža uvid u aktivnost i funkciju promatranih peptida. Ideja ovog znanstvenog rada upoznavanje je s raznim bioinformatičkim alatima te korištenje istih u svrhu predikcije bioaktivnosti i toksičnosti proteina na temelju slijeda aminokiselina u polipeptidnom lancu.

2. TEORIJSKI DIO

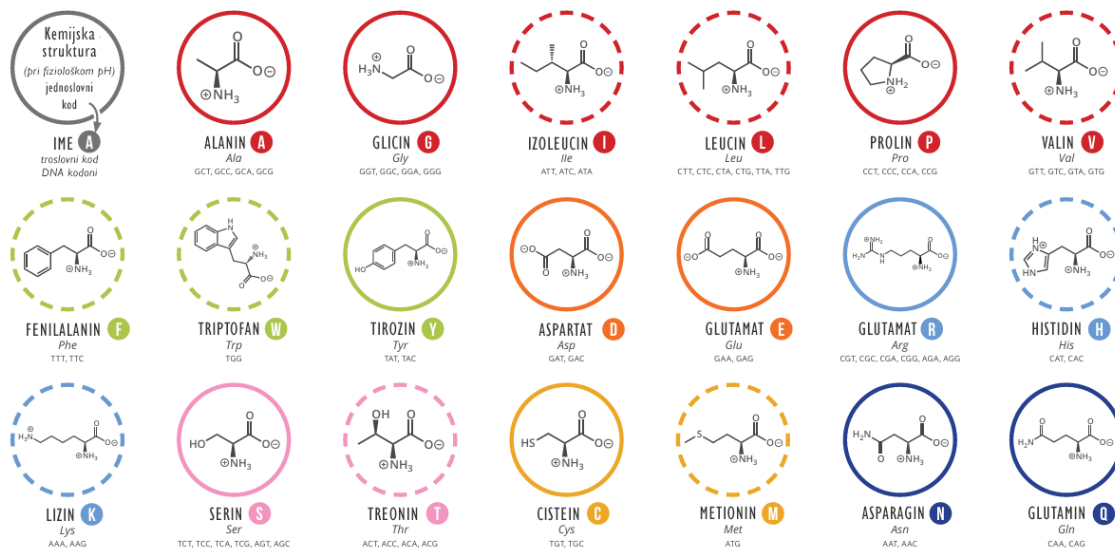
2.1. STRUKTURE PROTEINA

Proteine i peptide možemo nazvati i završnim produktima procesa dekodiranja genetičkog materijala sadržanog u genima [4]. Svaki gen sadrži informaciju za sintezu nekog proteina ili funkcionalne RNA molekule. Peptidi se međusobno razlikuju po aminokiselinama i vrstama veza koje sadrže, a kao rezultat toga razlikuju se i po trodimenzionalnim strukturama koje stvaraju. Nastaju povezivanjem individualnih jedinica od ukupno 20 različitih aminokiselina koje se mogu međusobno uvelike razlikovati, kako po građi tako i po kemijskim svojstvima (slika 1.) Peptidna veza, koja je glavna vrsta veze između aminokiselinskih jedinica u polipeptidnim lancima, nastaje povezivanjem α -karboksilne skupine jedne aminokiseline s α -amino skupinom druge aminokiseline uz izdvajanje molekule vode. U peptidnom lancu razlikujemo pravilno organiziranu okosnicu lanca (čine ju atomi peptidne veze i α -ugljikovi atomi sa vezanim vodikovim atomom) i različite pobočne ogranke. Pobočni ogranci su u prostoru naizmjenično raspoređeni s jedne i druge strane okosnice kako bi se izbjegle steričke smetnje [5].

PREGLED DVADESET STANDARDNIH AMINOKISELINA

AMINOKISELINE SU GRADIVNI BLOKVI PROTEINA U ŽIVIM ORGANIZMIMA, U PRIRODI SE NALAZI PREKO 500 AMINOKISELINA - MEĐUTIM, LJUDSKI GENETSKI KOD IZRAVNO KODIRA 20. "ESENCIJALNE" AMINOKISELINE MORAJU SE DOBITI IZ PREHRANE, A NESENCIJALNE AMINOKISELINE MOGU SE SINTETIZIRATI U TIJELU

Chart Key: ● ALIFATSKE ● AROMATSKE ● KISELE ● BAZIČNE ● HIDROKSILNE ● SADRŽE SUMPOR ● ● AMIDNE ○ NEESENCIJALNE ○ ESENCIJALNE

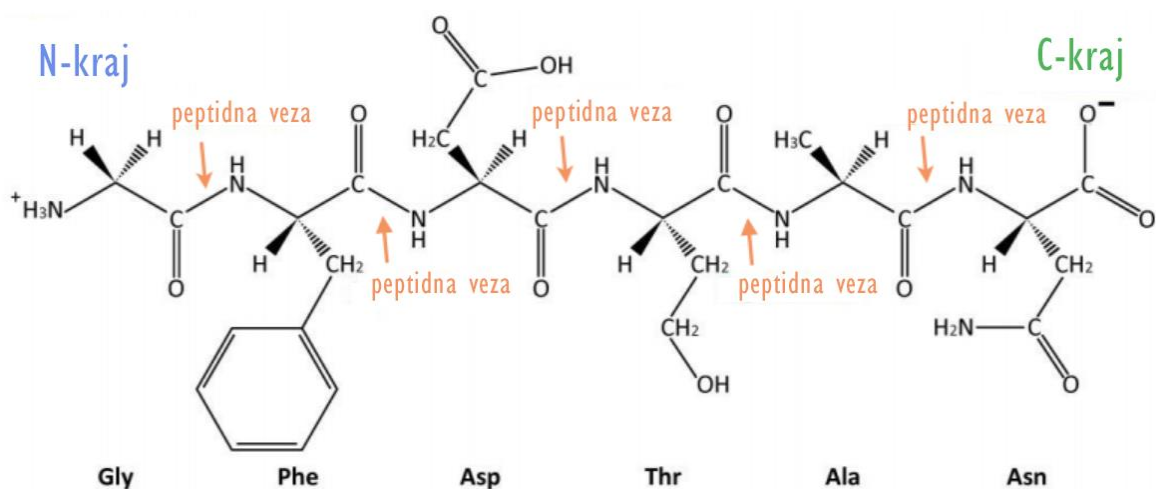


Napomena: Ovaj grafikon prikazuje one aminokiseline za koje ljudski genetski kod izravno kodira. Selenocistein se često naziva 21. aminokiselinom, ali je kodiran na poseban način. U nekim je slučajevima teško razlikovati asparagin/aspartat i glutamin/glutamat. U tim se slučajevima koriste redom oznake asx (B) i glx (Z).

Slika 1. Prikaz 20 aminokiselina koje grade polipeptidne lance proteina i peptida [6]

2.1.1. PRIMARNA STRUKTURA PROTEINA

Konformacija ili trodimenzionalni oblik peptida ili proteina, direktno ovisi o linearnom slijedu aminokiselina u polipeptidnom lancu, tj. primarnoj strukturi proteina (slika 2). Slijed aminokiselina u primarnoj strukturi određen je slijedom nukleotida u polinukleotidnim lancima. O primarnoj strukturi ovisi način smatanja proteina u trodimenzionalni oblik te time i funkcija proteina. Peptidna veza kojom su aminokiseline međusobno povezane je vrlo stabilna, a da bi došlo do njenog cijepanja u staničnim uvjetima potrebno je djelovanje proteolitičkih enzima [5].

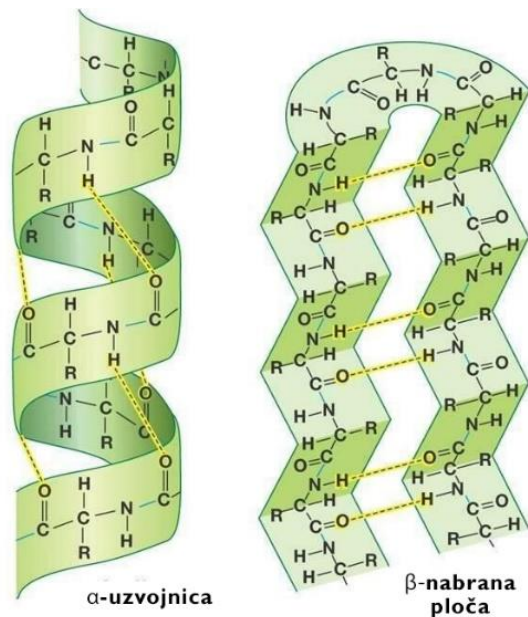


Slika 2. Struktura peptidnog lanca [7]

2.1.2. SEKUNDARNA STRUKTURA PROTEINA

Sekundarna struktura opisuje prostorni odnos aminokiselinskih ostataka nastao formiranjem vodikovih mostova u osnovnom dijelu polipeptidnog lanca. Ova razina proteinske strukture određuje raspored atoma okosnice lanca u prostoru. Konformacija bočnih ogranaka aminokiselina se kod sekundarne strukture ne uzima u obzir.

Unutar ove razine strukture proteina razlikujemo dvije periodične strukture: α -uzvojnica i β -nabranu ploču (slika 3), te strukturni element tzv. β -okret.



Slika 3. Shematski prikaz sekundarnih struktura [8]

α-uzvojnica je čvrsto namotana u štapićastu strukturu čiju unutrašnjost čini tijesno uvijena polipeptidna okosnica, a pobočni se ogranci pružaju prema van u uvojitom rasporedu. Stabilizirana je vodikovim vezama između karbonilne skupine peptidne veze svake aminokiseline i NH-skupine peptidne veze aminokiseline koja se u linearnom obliku nalazi četiri ostatka dalje. Ovisno o slijedu aminokiselina formiraju se različite sekundarne strukture. Aminokiseline u pravilu teže stvaraju α-uzvojnice, ali postoje iznimke poput alanina, glutaminske kiseline i leucina koji češće tvore α-uzvojnici. Valin, treonin i izoleucin sadrže β-C atome koji se dodatno granaju pa time destabiliziraju α-uzvojnici i lakše stvaraju β-nabranu ploču. Serin, asparaginska kiselina i asparagin sadrže pobočne skupine koje su donori ili akceptori vodikovih atoma, a kako se nalaze neposredno uz okosnicu peptidnog lanca, ulaze u interakciju s vezama u α-uzvojnici, čime ju razaraju. Prolin je specifičan po tome što nema slobodnu -NH₂ skupinu, a njegova struktura prstena onemogućuje pravilno stvaranje struktura poput α-uzvojnice i β-nabrane ploče [5]. Jedan u nizu eksperimenata koji potvrđuju prevalenciju određenih aminokiselina u sekundarnim strukturama je i eksperiment turskog znanstvenika Cevdet Nacara koji je izračunao učestalost pojave svake aminokiseline u sekundarnim strukturama, što je prikazano u tablici 1.

Vodikove su veze u α-uzvojnici međusobno paralelne, a paralelne su i s osi pružanja uzvojnice te povezuju grupe koje se nalaze jedna iznad druge u navoju (slika 3). Pobočni ogranci aminokiselina usmjereni su prema van i mogu stupati u interakcije s pobočnim ograncima u drugim dijelovima polipeptidnog lanca, što je bitno za uspostavljanje tercijarne strukture proteina [5].

U strukturi β -nabrane ploče polipeptidni lanac je gotovo sasvim istegnut, a ne tijesno uvijen. Aminokiselinski su ostatci udaljeni 0,35 nm, pri čemu su pobočni ogranci usmjereni ispod i iznad ravnine ploče, α -C atomi se nalaze u bridovima nabora ploče, a polipeptidna okosnica teče u smjeru pružanja ploče. β -nabrana ploča nastaje povezivanjem NH- i CO- skupine različitih dijelova polipeptidnog lanca vodikovim vezama. Veze su međusobno paralelne, ali okomite na smjer pružanja lanca pa ga ne sabijaju, nego naprotiv drže dijelove lanca na određenoj udaljenosti (slika 3).

Tablica 1. Učestalost pojave svake od aminokiselina u sekundarnim strukturama [9]

Amino Acid	α -helix	Reverse turn	β -sheet
Ala	1.41	0.82	0.72
Arg	1.21	0.90	0.84
Asn	0.76	1.34	0.48
Asp	0.99	1.24	0.39
Cys	0.66	0.54	1.40
Gln	1.27	0.84	0.98
Glu	1.59	1.01	0.52
Gly	0.43	1.77	0.58
His	1.05	0.81	0.80
Ile	1.09	0.47	1.67
Leu	1.34	0.57	1.22
Lys	1.23	1.07	0.69
Met	1.30	0.52	1.14
Phe	1.16	0.59	1.33
Pro	0.34	1.32	0.31
Ser	0.57	1.22	0.96
Thr	0.76	0.96	1.17
Trp	1.02	0.65	1.35
Tyr	0.74	0.76	1.45
Val	0.90	0.41	1.87

2.1.3. TERCIJARNA STRUKTURA PROTEINA

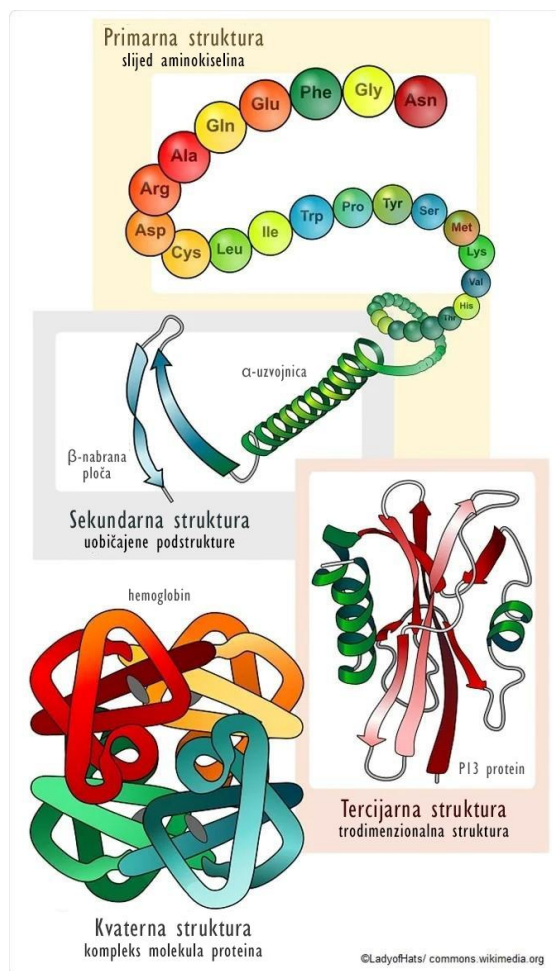
Tercijarna je struktura stabilizirana velikim brojem različitih nekovalentnih veza između odgovarajućih grupa u pobočnim lancima aminokiselinskih ostataka. Veze koje stabiliziraju tercijaru strukturu su ionske, vodikove, van der Waalsove i hidrofobne interakcije te se u nekim slučajevima javljaju i kovalentne disulfidne veze. Svaka od 20 aminokiselina koje tvore polipeptidne lance sadrži različitu pobočnu skupinu. Sastav pobočnih skupina je iznimno bitan za strukturu proteina ili peptida jer često ulaze u međusobne interakcije čime održavaju jedan dio polipeptidnog lanca u određenom obliku ili konformaciji [5].

Aminokiseline s nabijenim pobočnim lancima mogu tvoriti ionske veze, a polarne aminokiseline tvore vodikove veze, dok će hidrofobni pobočni ogranci međusobno ulaziti u interakcije putem van der Waalsovih sila, ujedno i najslabije veze među navedenima. Disulfidne veze nastaju oksidacijom dvaju cisteinskih ostataka koji se pri tome povežu disulfidnim mostom u tzv. cistin. Zbog navedenih interakcija, koje su specifične za pojedine aminokiseline, slijed aminokiselina u polipeptidnom lancu najbitniji je faktor u određivanju trodimenzionalne strukture proteina ili peptida, a samim time i njihovih funkcija [5].

2.1.4. KVATERNA STRUKTURA PROTEINA I UTJECAJ POSTSINTETSKIH MODIFIKACIJA NA AKTIVNOST PROTEINA

Proteini koji sadrže više od jednog proteinskog lanca imaju dodatnu strukturnu razinu - kvaternu strukturu. Kvaterna struktura posljedica je međusobnog povezivanja različitih polipeptidnih lanaca nekovalentnim vezama između pobočnih lanaca njihovih aminokiselinskih ostataka [5].

Nakon završetka sinteze proteini se mogu dodatno modificirati na različite načine tzv. postsintetskim modifikacijama. Najčešće modifikacije su: vezanje acetilne grupe na amino terminalnu grupu; vezanje metilne grupe na aminoskupine lizina ili arginina ili na hidroksilne skupine serina ili treonina; vezanje ugljikohidratnih jedinica na specifične asparaginske, serinske ili treoninske ostatke (glikozilacija); vezanje fosfatne skupine na ostatak serina, treonina ili tirozina (fosforilacija); vezanje lipida na α -amino skupinu ili na $-SH$ skupinu cisteina; hidroksiliranje prolina. Postsintetske modifikacije proteina imaju utjecaj na stabilnost, topljivost i aktivnost proteina [5].



Slika 4. Prikaz odnosa svih stupnjeva strukture proteina [10]

2.2. FUNKCIJE PROTEINA I PEPTIDA U ŽIVOM ORGANIZMU

Male promjene u strukturi proteina mogu imati veliki utjecaj na njihovu funkciju. Na primjer, pri vezanju kisika na hemoglobin dolazi do male strukturne promjene koja omogućuje da se nakon vezanja prvog kisika, poveća afinitet molekule hemoglobina prema novim molekulama kisika.

Od peptida u ljudskom organizmu valja izdvojiti peptide koji smanjuju razinu kolesterola u krvi, zatim peptide koji smanjuju efekt starenja te peptide koji vezanjem s mineralima održavaju odgovarajući pH u crijevima. Povećana razina kolesterola i triglicerida u krvi, odnosno hiperlipidemija, jedan je od najvećih pokazatelja ozbiljnih kardiovaskularnih komplikacija. Peptidi izolirani iz biljnih izvora, osobito oni dobiveni iz soje pokazali su se učinkovitima za smanjenje razine kolesterola u krvi. Samim time su i dijetne baze postale jednim od najčešćih načina borbe protiv hiperlipidemije [3]. Iako sam mehanizam rada tih peptida još nije poznat, smatra se da do smanjenja razine kolesterola u krvi dolazi zato što peptidi cijepaju

kolesterol tijekom njegovog putovanja kroz crijeva, čime se onemogućuje veliki porast njegove koncentracije u krvi. Drugi način na koji peptidi utječu na razinu kolesterola u krvi je putem interakcija između hidrofobnih aminokiselinskih ogranaka peptida i žučne kiseline, čime se povećava izlučivanje masnih kiselina u stolici. Primjer takvih peptida su aminokiselinski sljedovi: *LPYPR* i *IAVPGEVA* koji su izolirani iz soje i pokazuju veliki utjecaj na smanjenje razine kolesterola u krvi [3].

Na proces starenja mogu utjecati vanjski i unutarnji faktori. Vanjski faktori poput ultraljubičastog zračenja, zagađenosti zraka te pušenja, negativno utječu na ljudsku kožu. UV zračenje ometa sintezu kolagena, inducira proizvodnju kolagenaze koja razgrađuje izvanstanične proteine čime se u konačnici uništava i sama koža. Razvijanjem biološki aktivnih peptida otkriveno je da određeni peptidi mogu pomoći pri pomlađivanju kože te omogućavaju stanici veću aktivnost pri popravku kože. Danas se stoga bioaktivni peptidi koriste pri zacjeljivanju rana, poravnavanju bora te kao stimulansi proizvodnje kolagena. Osim toga kozmetički se bioaktivni peptidi danas koriste i zbog svojih antioksidativnih, antibakterijskih efekata te mogućnosti izbjeljivanja kože [11].

2.3. BIOLOŠKI AKTIVNI PROTEINI I PEPTIDI

Bioaktivni peptidi skupina su bioloških molekula koje su uobičajeno skrivene u strukturi matičnih proteina i postaju aktivne nakon cijepanja proteina. Bioaktivni peptidi, pa tako i proteini, važni su sastojci hrane, nutraceutika i farmaceutskih proizvoda [3]. U živim organizmima obavljaju mnoge bitne funkcije koje bez njihovog postojanja ne bi bile moguće (slika 5.)

Od biološki aktivnih peptida bitno je izdvojiti glutation, bradikinin, angiotenzin, vazopresin, oksitocin te TRH, odnosno tireotropin-oslobađajući hormon. Glutation je peptid sastavljen od 3 aminokiseline, što ga čini tripeptidom. Aktivan je pri katalitičkoj aktivnosti mnogih enzima te pri detoksifikaciji H_2O_2 . Bradikinin se formira u plazmi, a nastaje kao posljedica protelitičke aktivnosti tripsina ili drugih enzima prisutnih npr. u zmijskom otrovu. Uzrokuje proširenje krvnih žila, sužavanje bronha i bronhiola odnosno dišnih puteva te povećanje propusnosti krvnih kapilara. Angiotenzin se luči iz bubrega pri niskom dotoku krvi do bubrega. Uzrokuje kontrakciju krvnih žila te povišenje krvnog tlaka [12].

Uloge bioaktivnih proteina i peptida uključuju antioksidativne, antimikrobne, antihipertenzivne, imunomodulatorne i druge biološke aktivnosti. Na primjer, laktotransferin i laktoperoksidaza igraju važnu ulogu u imunološkom obrambenom sustavu tako što pojačavaju prirodnu zaštitu organizma protiv patogena [12]. Biološki aktivne peptide i proteine možemo podijeliti u dvije skupine: endogeni i egzogeni bioaktivni peptidi/proteini. Endogeni peptidi nastaju unutar stanice, a mogu nastajati iz različitih tipova stanica ili žlijezda. Egzogeni peptidi pak u tijelo ulaze putem raznih izvora, poput hrane, dodataka hrani ili lijekova. Uz sve navedene funkcije koje obavljaju biološki aktivni peptidi, uz to su još i sigurni za korištenje u živim organizmima zbog čega su na meti mnogih istraživača i znanstvenih projekata. Biološki aktivni peptidi pokazali su se izvrsnima za liječenje kroničnih bolesti zbog svog visokospecifičnog djelovanja i visoke aktivnosti pri niskim koncentracijama, čime se smanjuje toksičnost takvih molekula. Osim toga, biološki aktivni peptidi imaju kratko vrijeme zadržavanja u ljudskom organizmu, dok se sintetički lijekovi nakupljaju u tijelu što može imati ozbiljne posljedice na organizam [3].



Slika 5. Shematski prikaz različitih funkcija bioaktivnih peptida [13]

2.3.1. IZVORI BIOLOŠKI AKTIVNIH PEPTIDA

Biološki aktivni peptidi nalaze se u svakom živom organizmu, stoga se mogu izolirati iz velikog broja različitih izvora. Pri izolaciji biološki aktivnih peptida bitno je prilagoditi tehnike izolacije organizmu iz kojeg izoliramo. Isto tako, kako bi izolacija i pročišćavanje peptida iz prirodnih izvora bilo isplativo, u soju domaćinu potrebna je velika ekspresija traženog peptida, kako bi se nakon pročišćavanja osigurao dovoljan prinos za daljnje korištenje peptida [14].

Analizom biološki aktivnih peptida izoliranih iz morskih životinjskih izvora ustanovljeno je da veliki broj njih iskazuje antikancerogeno djelovanje jer stimuliraju različite stanične mehanizme koji uzrokuju smrt stanice. Nadalje, proteini u mlijeku imaju važna svojstva, kao što je antibakterijsko, antioksidativno i imunoprotektivno djelovanje. Mnoga istraživanja dokazala su da peptidi imaju mogućnost slanja signala sitosti do mozga, čime se smanjuje apetit osobe, a samim time i potreba za hranom. Također, peptidi su pokazali potencijal i u borbi protiv pretilosti jer utječu na apsorpciju nutrijenata u crijevima, čime se također smanjuje i apetit [5].

2.4. TOKSINI I OTROVI

Dugi niz godina čovječanstvo koristi otrovne životinje i njihove otrove kao moguće medicinske pripravke. Smrtonosne životinje poput tarantula koriste se diljem Srednje i Južne Amerike u borbi protiv astme ili teških bolesti poput raka, dok se otrov smrtonosne zmijske kobre koristi u borbi protiv niza bolesti poput astme, dječje paralize, multiple skleroze te bolova [15]. Osim pauka i zmijske, otrovi se javljaju i u ostalim životinjama poput puževa, gmazova, škorpiona, stonoga, lignji i osa. Otrovi ovih životinja često imaju jako specifično djelovanje te visoku aktivnost pri niskim koncentracijama što ih razlikuje od ostalih tipova molekula koje koristimo pri izradi lijekova. Upravo se zbog ovih svojstava svijet peptida prikazuje kao sljedeći korak u evoluciji visokospecifičnih i visokodjelotvornih lijekova [5].

Tehnologija peptida razvija se velikom brzinom te se svakodnevno pronalaze nove mogućnosti njihovog korištenja. Samim time postoji i velika potreba za analizom novih izvora biološki aktivnih peptida. Jedan od najbogatijih izvora peptida su otrovi i toksini. Aktivnost i jednih i drugih predstavlja jedan od glavnih toksikoloških problema u prirodnim ekosustavima, ali, osim toga, imaju sve veću primjenu u drugim aspektima života. Toksine proizvode brojni mikroorganizmi i beskralješnjaci, kao i više biljke i životinje, dok otrove proizvode mnoge skupine životinja, od mješince, do kraljeznjaka. Toksini su širi pojam za otrovne supstance proizvedene u živim organizmima, dok se pod pojmom otrova smatraju sve supstance koje su unesene u organizam injektiranjem najčešće putem ugriza ili uboda. Na kemijskoj razini otrovi su mješavine bioaktivnih komponenti, kao što su proteini, glikoproteini i peptidi koji čine 90-95% suhe tvari otrova te drugih kemijskih tvari kao što su lipidi, nukleozidi, aminokiseline i metalni ioni. Toksini se mogu uvelike razlikovati po kemijskoj strukturi, ali uglavnom ih grade proteini, peptidi i manje organske molekule. Pošto potječu iz različitih izvora, često se imenuju prema skupinama organizama koji ih proizvode, na primjer, mikotoksini, toksini algi, otrovi insekata [16]. Također, neki od tih peptida su se pokazali dobrim inhibitorima angiotenzin konvertirajućeg enzima, ublaživačima boli, dok neki pokazuju aktivnost sličnu inzulinu, čime je vidljiva njihova uloga u živim organizmima [3] [17].

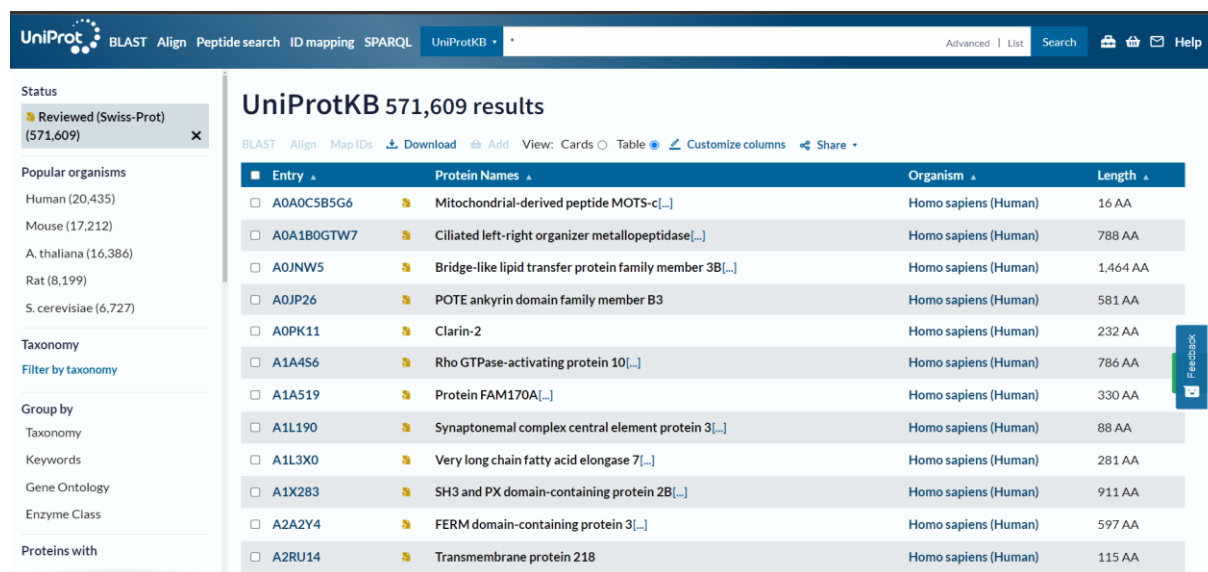
3. EKSPERIMENTALNI DIO

3.1. MATERIJALI

3.1.1. BAZA PODATAKA - UniProt

U ovom je radu napravljena analiza te predikcija aktivnosti 100 različitih peptida i proteina. Podaci su preuzeti s internetske baze podataka *UniProt (release 2024_03)*. *UniProt* je baza podataka dostupna široj javnosti u kojoj je moguće na jednostavan način doći do velikog broja informacija o trenutno dostupnim, uspješno sekvencioniranim proteinima. Budući da je ovim pristupom određeni protein dovoljno jednom analizirati te tako dobivene podatke postaviti na stranicu kako bi bili dostupni za korištenje, takve baze podataka omogućuju lakšu i bržu suradnju znanstvenika diljem svijeta [18].

U vrijeme pisanja ovog znanstvenog rada ukupan broj proteina i peptida u *UniProtKB* bazi podataka je 245 382 527. Biolozi, molekularni biolozi, genetički inženjeri te ostali znanstvenici iz područja biologije svakodnevno manualno analiziraju proteine te uz pomoć podataka dostupnih iz do sada objavljenih znanstvenih radova nadopunjavaju do tad nepoznate informacije. Na ovaj način analizirani proteini su u manjini u skupini svih proteina sadržanih u *UniProtKB* bazi podataka te ih je u vrijeme pisanja ovog znanstvenog rada bilo 571 609. Prilikom pretraživanja ovakve baze podataka vidljivi su nam podaci poput *Naziv unosa (Entry name)*, odnosno *UniProtID*, zatim imena proteina, organizma iz kojeg protein potječe te duljina njegovog polipeptidnog lanca (slika 6). Također, moguće je primijeniti filtere kako bismo povratno dobili samo željene informacije (slika 7).



Entry	Protein Names	Organism	Length
A0A0C5B5G6	Mitochondrial-derived peptide MOT5-c[...]	Homo sapiens (Human)	16 AA
A0A1B0GTW7	Ciliated left-right organizer metalloproteinase[...]	Homo sapiens (Human)	788 AA
A0JNW5	Bridge-like lipid transfer protein family member 3B[...]	Homo sapiens (Human)	1,464 AA
A0JP26	POTE ankyrin domain family member B3	Homo sapiens (Human)	581 AA
A0PK11	Clarin-2	Homo sapiens (Human)	232 AA
A1A456	Rho GTPase-activating protein 10[...]	Homo sapiens (Human)	786 AA
A1A519	Protein FAM170A[...]	Homo sapiens (Human)	330 AA
A1L190	Synaptonemal complex central element protein 3[...]	Homo sapiens (Human)	88 AA
A1L3X0	Very long chain fatty acid elongase 7[...]	Homo sapiens (Human)	281 AA
A1X283	SH3 and PX domain-containing protein 2B[...]	Homo sapiens (Human)	911 AA
A2A2Y4	FERM domain-containing protein 3[...]	Homo sapiens (Human)	597 AA
A2RU14	Transmembrane protein 218	Homo sapiens (Human)	115 AA

Slika 6. UniProtKB baza podataka sa svim manualno provjerenim proteinima i peptidima [19]

UniProt Data	4	^
Names & Taxonomy	2	∨
Sequences	1	∨
Function		∨
Miscellaneous	1	∨
Interaction		∨
Expression		∨
Gene Ontology (GO)		∨
Pathology & Biotech		∨
Subcellular location		∨
PTM / Processing		∨
Structure		∨
Publications		∨
Date of		∨
Family & Domains		∨
External Resources		∨

Slika 7. Mogući filteri pri pretraživanju UniProt baze podataka [20]

U ovom radu se fokusiramo na *ToxProt* bazu podataka koja se sastoji od 7 906 različitih toksina i otrova. Ta je baza podataka napravljena analizom proteina i peptida koji se izlučuju u životinjskom otrovu. U otrovne životinje ubrajamo zmije, pauke, škorpione, puževe stošce (morski puževi roda *Conus*), stonoge, meduze, kukce, guštere te neke ribe i čudnovate kljunaše. U *ToxProt* bazi podataka svi su proteini i peptidi imenovani te je dostupna velika količina korisnih informacija o samom proteinu, poput redoslijeda aminokiselina u njegovom polipeptidnom lancu, funkcije peptida, kofaktora proteina, načina regulacija aktivnosti, znanstvenih radova u kojima se protein navodi i dr. [20].

U svrhu izrade ovog završnog rada postavljeni su određeni filteri pri pretraživanju kako bismo povratno dobili samo željene podatke. Od svih dostupnih podataka izdvojeni su *UniProtID*, organizam kojem protein ili peptid pripada, duljina polipeptidnog lanca te sekvenca polipeptidnog lanca napisana na način da jedno slovo predstavlja jednu aminokiselinu u lancu. Sekvenca proteina bit će korištena za pronalazak homolognih sekvenci.

3.1.2. BLAST (*Basic Local Alignment Search Tool*)

BLAST (*Basic Local Alignment Search Tool*) osnovni je alat za pretraživanje homologije, odnosno alat koji traži sličnosti između dane sekvence i sekvence već dostupne u sustavu.

Program uspoređuje sekvence nukleotida ili proteina s već poznatim sekvencama te izračunava statističku značajnost preklapanja. BLAST se bazira na složenim statističkim principima korištenjem kojih je omogućeno da se statistički prikaže rezultat preklapanja ili ne preklapanja određenih sekvenci. Kao najbitniji parametar se pokazao *E-vrijednost*. Niska *E-vrijednost* ukazuje na veliku statističku mogućnost da su dvije analizirane sekvence zapravo homologne [21].

Za razumijevanje rada BLAST-a potrebno je poznavati evoluciju te različite mehanizme unošenja promjena u slijed nukleotida unutar gena. Promjena u slijedu nukleotida se potom očitava kao promjena u slijedu aminokiselina u polipeptidnom lancu, što rezultira promijenjenom strukturom proteina, a time i drugačijom funkcijom. Takva promjena, odnosno mutacija, u esencijalnim genima rezultirat će smrću stanice, stoga za esencijalne gene kažemo da su konzervirani jer kroz evoluciju nisu bili podložni stvaranju mutacija. Dakle, očekuje se da će dvije homologne sekvence iskazivati visoke stupnjeve konzerviranosti, a samim time i velik broj preklapanja pri usporedbi slijeda aminokiselina u polipeptidnom lancu [21].

Prilikom BLAST pretraživanja najprije je potrebno potvrditi da su uspoređivane sekvence homologne. Zatim, pregledom poravnanja utvrditi da visok stupanj preklapanja sekvenci upućuje na to da kodiraju za isti protein, a samim time da ti proteini imaju istu funkciju. Naravno, moguća su odstupanja od ovakvih logičkih zaključaka, u slučaju da je visoka točnost preklapanja sekvenci slučajna ili proteini nemaju istu funkciju unatoč sličnom slijedu aminokiselina u polipeptidnom lancu.

Prilikom svakog BLAST pretraživanja dobiju se rezultati poput prikazanih u nastavku (slika 8). Takav se prozor sastoji od dijela sa sažetim informacijama o unesenoj sekvenci, bazi podataka u kojoj je napravljena pretraga te jedinstveni RID (*request ID*) koji predstavlja direktnu poveznicu s tom specifičnom BLAST pretragom. Osim toga, vidljiv je i prozor u kojemu imamo mogućnost odabira filtera pretrage, odabirom kojih možemo postaviti parametre pretraživanja na samo određene taksonomske grupe ili možemo postaviti željeni raspon varijabli poput *postotni identitet*, *E-vrijednost* ili *pokrivenost upita* čime ćemo dobiti podatke bliže željenima. Najveći dio prozora s prikazanim rezultatima čini *Opisni prozor (Description tab)*, u kojemu možemo vidjeti rezultate BLAST pretrage te sve željene vrijednosti koje će nam pomoći pri odabiru željenog homologa. Osim imena proteina te organizma iz kojeg dolazi taj homolog, možemo vidjeti i druge bitne podatke poput *E-vrijednosti*, *postotnog identiteta* i *duljine homologa*. Rezultati s najnižom *E-vrijednosti* su oni za koje s najvećom sigurnošću možemo reći da su homolozi unesene sekvence.

Job Title **Protein Sequence**

RID [9FJA99HZ016](#) Search expires on 07-18 21:38 pm [Download All](#) ▾

Program BLASTP [Citation](#) ▾

Database nr [See details](#) ▾

Query ID Icl|Query_5859560

Description unnamed protein product

Molecule type amino acid

Query Length 63

Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity **E value** **Query Coverage**

to to to

[Filter](#) [Reset](#)

Compare these results against the new **Clustered nr database** [?](#)
[BLAST](#)
✕

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download ▾ Select columns ▾ Show 100 ▾ [?](#)

select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-Am; AltName: Full=Am975; AltName: Full=Bromocontryphan Ama1055; AltName: ...	Conus amadis	129	129	100%	2e-37	100.00%	63	P0C248.2
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-R; AltName: Full=Bromocontryphan; Contains: RecName: Full=(Des-Gly1)-contry...	Conus radiatus	118	118	100%	5e-33	88.89%	63	P58786.1
<input checked="" type="checkbox"/>	conotoxin precursor O2 [Conus judaeus]	Conus judaeus	118	118	100%	7e-33	90.48%	63	UMA83526.1
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-Mi; Flags: Precursor [Conus miles]	Conus miles	118	118	100%	7e-33	90.48%	63	A0A1P8NVS3.1
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-Sm; Flags: Precursor [Conus stercuscarum]	Conus stercuscarum	117	117	100%	8e-33	90.48%	63	P58787.1
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-Vi; AltName: Full=Contryphan-R-like; Flags: Precursor [Conus virgo]	Conus virgo	117	117	100%	9e-33	88.89%	63	F5C3T9.1
<input checked="" type="checkbox"/>	conotoxin superfamily O2 [Conus magus]	Conus magus	115	115	100%	5e-32	88.89%	63	QFQ61089.1
<input checked="" type="checkbox"/>	TPA: conotoxin superfamily O2 [Conus magus]	Conus magus	115	115	100%	7e-32	87.30%	63	DAC80593.1
<input checked="" type="checkbox"/>	TPA: conotoxin precursor O2 [Conus judaeus]	Conus judaeus	115	115	100%	8e-32	87.30%	64	DAZ86838.1
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-Lo; AltName: Full=Contryphan-Lo1; AltName: Full=Lo959; Contains: RecName: F...	Conus buxus lo...	114	114	100%	1e-31	85.71%	63	P0C250.2
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-P; AltName: Full=Trp-Contryphan-P; Flags: Precursor [Conus purpurascens]	Conus purpurascens	114	114	100%	1e-31	87.30%	63	P58784.1
<input checked="" type="checkbox"/>	conotoxin precursor O2 [Conus judaeus]	Conus judaeus	114	114	100%	2e-31	87.30%	63	UMA83525.1
<input checked="" type="checkbox"/>	RecName: Full=Contryphan-Eb; Flags: Precursor [Conus ebraeus]	Conus ebraeus	114	114	100%	3e-31	87.30%	63	A0A1P8NVS9.1

Slika 8. Rezultati BLAST pretraživanja [22]

3.1.3. Python

Python je programski jezik poznat po svojoj jednostavnosti, čitljivosti i svestranosti. Njegova popularnost uglavnom proizlazi iz lako razumljive i intuitivne sintakse, koja omogućava programerima brzo pisanje i čitanje koda. Pored toga, *Python* podržava različite programerske paradigme, uključujući proceduralno, objektno orijentirano i funkcionalno programiranje, što olakšava razvoj raznih aplikacija [23].

Jedna od velikih prednosti *Pythona* širok je spektar dostupnih biblioteka. *Python* ekosistem nudi bogat izbor specijaliziranih biblioteka za razna područja kao što su *web* razvoj, baze podataka, znanstveno računanje, automatizacija, igre, itd. Te biblioteke omogućuju programerima brzo i jednostavno dodavanje novih funkcionalnosti svojim projektima, što štedi vrijeme i pojednostavljuje razvoj procesa. Od velikog broja dostupnih alata i biblioteka koje se koriste u bioinformatičkim istraživanjima te projektima strojnog učenja valja izdvojiti – *Biopython*, *pyMOL*, *Biskit*, *Scikit-learn*, *NumPy* te *Matplotlib*. U izradi ovog rada koristili smo se bibliotekama *Scikit-learn*, *NumPy*, *Pandas* i *Toxinpred2* o kojima će kasnije biti rečeno više.

3.1.4. Scikit-learn

Scikit-learn je *open-source* biblioteka razvijena 2007. godine, a danas se koristi u znanstvenim radovima koji uključuju strojno učenje pa se samim time često koristi i u bioinformatičari [24]. Uključuje različite algoritme strojnog učenja, alate za prilagodbu modela, obradu podataka i analizu modela. *Scikit-learn* se u bioinformatičari koristi u području klasifikacije gena, predviđanja razvoja bolesti ili pri analizi bioloških podataka. Ima mogućnost brze obrade velike količine podataka što ga čini odličnim bioinformatičkim alatom. Ova je biblioteka napravljena na bazi *SciPy*, odnosno *Scientific Python*, koji mora biti instaliran prije samog *Scikit-learn*a. Alati napravljeni po *SciPy*-u zajedno zovemo *SciKits*, pa je tako i ova biblioteka dobila svoje ime jer omogućuje algoritme strojnog učenja.

3.1.5. Pandas

Pandas je *open source*, BSD-licencirana biblioteka koja pruža visoko učinkovite, jednostavne za korištenje strukture podataka i alate za analizu podataka za programski jezik *Python*. *Pandas* jača *Python* dajući popularnom programskom jeziku mogućnost rada s podacima nalik tablicama, omogućavajući brzo učitavanje, usklađivanje, manipulaciju i spajanje, uz ostale ključne funkcije. *Pandas* i njegove visoko optimizirane performanse najviše dolaze do izražaja kada je pozadinski kod napisan u C-u ili u *Pythonu* [25]. BSD licencirana biblioteka znači da je softver distribuiran pod BSD (*Berkeley Software Distribution*) licencom. Ova licenca je jedna od *open-source* licenci koja omogućava slobodno korištenje, distribuciju i modifikaciju softvera.

Pandas omogućava rad s podacima u različitim formatima kao što su CSV, *Excel* i SQL baze, uz podršku za mnoge druge formate. Ova funkcionalnost je izuzetno korisna za rad s različitim izvorima podataka i integraciju s drugim alatima. *Pandas* je često korišten za istraživanje, analizu i pripremu podataka za strojno učenje te vizualizaciju rezultata analize. Omogućava obradu velikih količina podataka, kao i generiranje statističkih izvještaja i vizualizaciju rezultata. Osim toga, *Pandas* se može kombinirati s drugim poznatim bibliotekama za analizu podataka i strojno učenje, što daljnje proširuje njegove mogućnosti.

3.1.6. **NumPy**

NumPy je jedna od osnovnih biblioteka u *Pythonu* za numeričko računanje po čemu je i dobila ime – *Numerical Python*. Omogućuje rad s velikom količinom podataka i rad na višedimenzionalnim matricama. Također, sadrži i razne matematičke funkcije kojima omogućuje obradu velike količine podataka. *NumPy* omogućuje bogatu statističku i numeričku podlogu na koju se *Scikit-learn* oslanja pri implementaciji algoritama strojnog učenja ili obrade podataka. Samim time je nastanak, pa tako i velika učinkovitost i široka upotreba *Scikit-learn* rezultat postojanja *NumPy*-a i njegovih složenih matematičkih modela.

3.1.7. **AlphaFold**

AlphaFold, razvijen od strane *Google Deep Mind* grupe, inovativni je AI model koji omogućuje precizno predviđanje 3D struktura proteina. Ovaj model koristi elemente dubokog učenja kako bi procijenio oblik proteina temeljen na njihovim sekvencama aminokiselina. Ključ uspjeha *AlphaFold*-a je njegova mogućnost uspješnog predviđanja strukture što se danas pokazuje korisnim u analizi strukturnih varijacija te kao pomoć u razvoju lijekova putem identifikacije potencijalnih mjesta vezanja. U bioinformatici, *AlphaFold* je značajno unaprijedio predviđanje strukture proteina, omogućujući bolje razumijevanje njihove funkcije, interakcija i mehanizama djelovanja. Njegova integracija u baze podataka i istraživačke procese ubrzava otkrića i poboljšava sposobnost dizajniranja ciljanih terapija i istraživanja bioloških procesa na molekularnoj razini. *AlphaFold* je kao takav implementiran u bazu podataka *UniProt*, čime je omogućeno pogledati strukturu od gotovo bilo kojeg proteina jednostavnim pretraživanjem njegovog imena na *UniProt* stranici. Osim što precizno predvidi strukturu proteina na temelju sekvence polipeptidnog lanca, *AlphaFold* nam daje i informaciju o stupnju sigurnosti u točnost predikcije. Zbog raznih razloga program ponekad ne može sa stopostotnom sigurnošću predvidjeti strukturu, ali budući da nam je dostupna informacija za koje je dijelove strukture proteina točnost predikcije veća, moguće je donijeti adekvatne zaključne vezane uz funkciju proteina. Ukoliko program pokazuje visoki postotak sigurnosti u izgled trodimenzionalne strukture, zaključci koji se dobivaju na temelju računalno generirane strukture vrijedit će i u stvarnim uvjetima, čime se omogućuju lakša i brža istraživanja proteina te gena i njihovih funkcija.

3.2. METODE

3.2.1. ODABIR PODATAKA

Za izradu je ovog rada bilo potrebno slučajnim odabirom odabrati 100 proteina/peptida iz *ToxProt* baze podataka. Cjelokupna je baza podataka preuzeta u obliku *Excel* tablice pri čemu se u svakoj ćeliji nalazio po jedan protein/peptid. Zatim je generirano 100 nasumičnih brojeva od 0 do završnog broja ćelija u *Excel* tablici putem internetskog generatora nasumičnih brojeva [26]. Potom je bilo potrebno odabrane proteine označiti, odvojiti ih od neoznačenih te svakom pridodati informacije dostupne u *ToxProt* bazi specifične za taj protein. Takvom metodom odabira slučajnih proteina i peptida dobiveni su podaci (slika 9.) za daljnju obradu.

1	Entry	Organism	Protein names	Sequence	Length [AA]	Bioactivity prediction
209	P0C248	Conus amadis (Amadis cone)	Contryphan-Am (Am975)	I MGKLTILVLAALLSTQVMVC	63	0.999676
333	P23026	Pseudonaja textilis (Eastern brown snake)	Basic phospholipase A2	te MHPAHLVLLGVCVSLGASE	145	0.999898
349	P33589	Lachesis muta muta (Bushmaster)	Thrombin-like enzyme	gyr VIGGDECNINEHRFLVALYDGI	228	0.561676
394	P60514	Psalmopoeus cambridgei (Trinidad chevron tara	Psalmotoxin-1 (PcTx1) (Pi-	EDCIPKWKGCVNRHGDCCEC	40	0.990315
455	P83644	Leiurus hebraeus (Deathstalker scorpion) (Leir	Toxin Lqh4 (Lqh IV) (LqhIV	GVRDAYIADDKNCVYTCGANE	65	0.999431
464	P84674	Daboia russelii (Russel's viper) (Vipera russelii)	Basic phospholipase A2	VI SLLEFGMMILEETGKLAIVPFYS	121	0.999856
565	Q6EER4	Bothriechis schlegelii (Eyelash palm pitviper)	Basic phospholipase A2	B: MRTLWIVAVLLVGVENLLQF	138	0.999186
591	Q7T1T4	Bothrops jararacussu (Jararacussu)	Snake venom metalloprot	TELPKGAVQPKYEDAMQYEFK	369	0.364379
632	Q95P90	Mesobuthus martensii (Manchurian scorpion) (E	HMG-CoA reductase inhibi	MVKMQVIFIAFIAVIACSMVYG	94	0.999761
657	Q9TWL9	Conus magus (Magical cone)	Conodipine-M alpha chain	QXPSTAELOCKINSNACSVFFS)	77	0.99705
830	C0HK16	Daboia russelii (Russel's viper) (Vipera russelii)	Basic phospholipase A2	d: SLLEFGKMMILEETGKLAIPSYSS	121	0.999877
843	C0HKQ6	Xylocopa appendiculata circumvolans (Japanes	Antimicrobial peptide Xac-	GFVALLKKLPLILKHLF	17	0.89807
851	C0HLF2	Centruroides ornatus (Scorpion) (Centruroides	Beta-mammal toxin Co1	KEGYLVNHSTGCKYECFKLGD	66	0.991547
909	C0JAV7	Loxosceles apachea (Apache recluse spider)	Dermonecrotic toxin LapS	WIMGHMVNIAIQIDEFVNLG/	273	0.289975
988	C0JB41	Loxosceles spinulosa (Recluse spider)	Dermonecrotic toxin LspiS	FALAHMVNDFDIMKSYLDEG/	278	0.0553742
1079	D2Y2L3	Cyriopogopus hainanus (Chinese bird spider) (H	U4-theraphotoxin-Hhn1a	I MKMTLIAILTCAAVLVLHTTAAE	85	0.999846
1080	D2YZQ3	ActinERIA villosa (Okinawan sea anemone)	DELTA-thalatoxin-Av1b	(C MRHFVFLYMLALSIPTAFAK	226	0.240049
1083	D6MKR0	Crotalus horridus (Timber rattlesnake)	Acidic phospholipase A2	C MRTLWIVAVLLGVEGSLVQFI	138	0.999745
1106	F8S114	Crotalus adamanteus (Eastern diamondback rat	Thrombin-like enzyme cro	MVLIRVLANLLILQLSYAQKSS	262	0.397747
1482	P16354	Heloderma suspectum (Gila monster)	Phospholipase A2 isozyme	GAFIMPGTLWCGAGNAASDY/	143	0.934154
1579	P58810	Conus marmoreus (Marble cone)	Chi-conotoxin MriB (Cono	GVGCCGYKLCCHPC	13	0.894789
1595	P59354	Buthus occitanus mardochei (Moroccan scorp	Alpha-like toxin Bom4 (Bo	GRDAYIAQPENCYECAKNSY	65	0.997265
1620	P61408	Grammostola rosea (Chilean rose tarantula) (Gr	Beta-theraphotoxin-Gr1b	(YCQKWLWTCDSERKCCEDM	29	0.999115
1831	Q45Z26	Tropidechis carinatus (Australian rough-scaled	s Acidic phospholipase A2	5 MYPALLVLLAVCVSLGAAASI	151	0.999842
1843	Q4VDB5	Loxosceles gaucho (Spider)	Dermonecrotic toxin LgSic	AGNRRPIWIMGHMVNIAIGQIE	280	0.201148
1870	Q68KK0	Solenopsis invicta (Red imported fire ant) (Soler	Phospholipase A1 (PLA1)	(MRKFAAIFVFFVQCTHLYSLA	346	0.458923
2149	A0A3G3C7U4	Conus amadis (Amadis cone)	Alpha-conotoxine-like Am	MGMRMMFTVLLVLLATTVVS	61	0.998894

Slika 9. Podaci korišteni u ovom radu – prikazan je samo dio podataka, dok je cijela tablica sa informacijama o svim proteinima dostupna u prilogu

3.2.2. ODABIR HOMOLOGA

Za svaki je od 100 odabiranih proteina bilo potrebno pronaći po jednog homologa. Homologe smo tražili pomoću drugog, često korištenog bioinformatičkog alata, BLAST-a [22]. BLAST (*Basic Local Alignment Search Tool*) je koristan iz razloga što omogućuje brzi pronalazak homolognih sekvenci.

BLAST se može koristiti pri donošenju zaključaka o funkcionalnim i evolucijskim odnosima između sekvenci, a u ovom je radu korišten u svrhu pronalaska sekvenci koje su homologne sekvencama slučajno odabranih proteina. Nakon pronalaska homologije, potrebno je provjeriti postoji li homologna sekvenca unutar nekog proteina koji se već nalazi u *ToxProt* bazi

podataka i ukoliko se ne nalazi, potrebno je dodati tu homolognu sekvencu u *Excel* tablicu zajedno sa svim dostupnim informacijama na *UniProt* stranici. Takvim smo načinom prikupljanja podataka dobili homologe prvih 100 slučajno odabranih proteina (slika 10)

1	Entry	Organism	Protein names	Sequence	Length	Bioactivity prediction	Identities	Positives	Gaps	Evalue
209	UM483526	Conus jadaeus	Conotoxin precursor O2	MGKLTILVLVAALLSTQVMV	63	0.999299	57/63(90%)	58/63(92%)	0/63(0%)	6E-33
333	UNW14934	Pseudonaja nuchalis	Phospholipase A2 protein chain A	HPAHLVLLGVCVSLLGASDI	144	0.999888	144/144(100%)	144/144(100)	0/144(0%)	5E-101
349	AEJ31999	Crotalus adamanteus	Serine proteinase 5	MVLIRVLANLLILQLSYAQKSE	262	0.420362	181/233(78%)	199/233(85)	7/233(3%)	2E-130
394	COHKD0.1	Heteroscodra maculata	Pi-theraphotoxin-Hm3a	EPICIPKWKSCVNRHGDCCA	37	0.994916	32/37(86%)	33/37(89%)	0/37(0%)	2E-14
455	ACJ23088	Buthus occitanus israelis	putative alpha toxin Tx2	MNYLIMISLALLMTGVESGT	87	0.999765	60/65(92%)	62/65(95%)	0/65(0%)	6E-38
464	AAZ53186	Daboia russelii limitis	Phospholipase A2	MRTLWIVAVCLLIGVEGSLLEF	137	0.999889	112/121(93%)	115/121(95)	0/121(0%)	3E-76
565	AUS82464	Crotalus mitchellii	Secretory phospholipase A2	MRTFWIVAVLLGVEGNLLQF	138	0.999831	122/138(88%)	129/138(93)	0/138(0%)	3E-87
591	P83512	Bothrops asper	Snake venom metalloproteinase BaP1	TELPKGAVQPKVEDAMQYEF	408	0.288756	326/369(88%)	341/369(92)	0/369(0%)	0
632	AMX81477	Mesobuthus eupeus	Venom peptide meuPep28	MVKLQVILAFIAVACSYYVE	94	0.999622	71/94(76%)	86/94(91%)	0/94(0%)	3E-50
657	UBT01703	Conus ammiralis	conodipine	MKMLESAWLIALAMPVIAE	182	0.971729	60/75(80%)	64/75(85%)	0/75(0%)	9E-35
830	ABC684	Daboia siamensis	Full-Basic phospholipase A2 DsM-S1	MRTLWIVAVCLIGVEGSLLEF	137	0.999893	116/121(96%)	118/121(97)	0/121(0%)	9E-79
843	UTO68649	Xylocopa aruana	XYTX-Xa1a	MILIFAFGERIVHITLHLENLL	156	0.123785	14/17(82%)	14/17(82%)	0/17(0%)	7.00E-04
851	XP_023228690	Centruroides sculpturatus	beta-mammal toxin Css4	MNSLLHITACLALVGTWAKEG	87	0.999849	62/66(94%)	65/66(98%)	0/66(0%)	2E-38
909	COJAW8	Loxosceles deserta	Phospholipase D	WIMGHMVNAIQIDEFVNLGI	273	0.166472	262/273(96%)	265/273(97)	3/273(1%)	0
988	AAI66074	Loxosceles boneti	sphingomyelinase D-like protein 3	RPKPIWVDAVHMVNDLELVD	276	0.0601684	120/272(44%)	178/272(65)	3/272(1%)	6E-80
1079	ABY77731	Cyrtopogopus schmidti	HWTX-VIc precursor	MKVTLIAITCAAVLVLHTTAAE	84	0.999598	75/85(88%)	81/85(95%)	1/85(1%)	3E-44
1080	APQ32126	Triactis producta	actinoporin	LRKMKHFLVFLYMCLETISM	229	0.252638	158/226(70%)	189/226(83)	1/226(0%)	1E-101
1083	AUS82451	Crotalus adamanteus	secretory phospholipase A2	MRTLWIVAVLLGVEGSLQVF	138	0.999796	132/138(96%)	136/138(98)	0/138(0%)	3E-94
1106	XP_039181661	Crotalus tigris	thrombin-like enzyme crotalase	MVLIRVLANLLILQLSYAQKSE	262	0.271351	250/262(95%)	255/262(97)	0/262(0%)	0
1482	1604193A	Heloderma suspectum	phospholipase A2 Pa5	GAFIMPGLWCGAGNAASDY	142	0.907369	142/142(100%)	142/142(100)	0/142(0%)	3E-103
1579	QFQ61111	Conus magus	conotoxin superfamily T	MRCILPVLIIIIITASAPGVIV	61	0.99983	12/12(100%)	12/12(100%)	0/12(0%)	0.0006
1595	O804800B	Scorpiones	neurotoxin BOT I	GRDAYIAQPCNVYCAQNS	56	0.994927	61/65(94%)	64/65(98%)	0/65(0%)	4E-38
1620	6MK4_A	Thrixopelma pruriens	Chain A, Beta/omega-theraphotoxin-Tp2a	YQCQKMMWTCDSERKCCCKG	30	0.999762	25/29(86%)	26/29(89%)	0/29(0%)	2E-22
1831	AAZ22652	Notechis scutatus	PLA-2 precursor	MYPALHLLVLLAVCVSLLGAS	151	0.999825	144/151(95%)	147/151(97)	0/151(0%)	5E-103
1843	AAQ16123	Loxosceles intermedia	dermonecrotic protein 1	ARVVLGCWVSLVQAAQTDDE	302	0.694975	276/280(99%)	276/280(98)	0/280(0%)	0
1870	XP_012534382	Monomorium pharaonis	phospholipase A1	MKIVATLILVLVQCIYLRAPNF	327	0.455456	171/346(49%)	232/346(67)	23/346(6%)	7E-113
2149	AGK23165	Conus textile	A superfamily conotoxin Tx1.3 precursor	MGMRRMFTVFLVLLVATAV	61	0.99952	50/60(83%)	51/60(85%)	0/60(0%)	1E-26

Slika 10. Lista homologa dobivena pretraživanjem pomoću BLAST-a – prikazan je samo dio podataka, dok je cijela tablica sa informacijama o svim homolozima dostupna u prilogu

Za pravilni odabir homologa potrebno je poznavati način rada sustava te značenje različitih vrijednosti unutar BLAST programa. Nakon uspješno obavljenog pretraživanja od strane BLAST-a dobije se veliki broj informacija koje je potrebno dodatno obraditi (slika 11). Za odabir odgovarajućih homologa koristi se varijabla *E-vrijednost*. Niska *E-vrijednost* (npr. od 0.01 do 1^{-10}) govori da je statistički jako mala vjerojatnost da je do poklapanja tražene sekvence s nađenom sekvencom došlo zbog slučajnosti, tj. ukazuje da je velika vjerojatnost da je protein koji pokazuje nisku *E-vrijednost* traženi homolog onog proteina čiji smo redoslijed aminokiselina u polipeptidnom lancu unijeli u BLAST. Visoka *E-vrijednost* (npr. > 0.1) sugerira da je do preklapanja došlo zbog slučajnosti [27].

U ovome je radu veliki broj homologa tražen upravo po ovom principu, odnosno zahtjevu za niskom *E-vrijednosti*, dok je za manji broj njih uzet rezultat pretraživanja sa nešto višom *E-vrijednosti* kako bi se mogao provjeriti utjecaj tog parametra na završne rezultate. Kada je određeni homolog nađen, skinuti su svi dostupni potrebni podaci o tom homologu te su ti podaci također unešeni u *Excel* tablicu u kojoj se nalazi 100 originalnih proteina. Zajedno s podacima poput imena proteina, organizma iz kojeg potječe, sekvence i duljine u aminokiselinama, pretraživanjem pomoću BLAST-a dobiveni su i podaci poput *Identities*, *Positives*, *Gaps* (slika 8.).

Ti su parametri vidljivi na *Alignments* prozoru rezultata BLAST pretraživanja (slika 11.). *Gaps* je omjer koji pokazuje na koliko se mjesta, odnosno na koliko se aminokiselina, razlikuju dvije

sekvence. Na ovom prozoru su praznine u homolognoj sekvenci prikazane minusom (-). *Identities* je omjer aminokiselina koje se nalaze na istom mjestu u obje sekvence s ukupnom količinom aminokiselina u originalnoj sekvenci unešenoj u BLAST [28]. Ukoliko su aminokiselinski ostaci na nekom mjestu u genomu slični, ali ne isti, to će se na *Alignments* prozoru moći vizualizirati kao plus (+) te će tada vrijednost *Positives* činiti vrijednost *Identities* uključujući sva mjesta u genomu označena plusom [29].

[Download](#) [GenPept](#) [Graphics](#)

U-theraphotoxin-Pv7a, partial [Pamphobeteus verdolaga]
 Sequence ID: [CAH7646043.1](#) Length: 93 Number of Matches: 1

Range 1: 51 to 93 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
56.6 bits(135)	2e-08	Compositional matrix adjust.	31/48(65%)	36/48(75%)	6/48(12%)
Query 1	IIECFEFSCEIEKDGKSKGKPKGKPKGDKDKKCC-GGWRCIKMCIKI			47	
	I EC FSC+IE+ EGKPKCKPKG K KKC GGW+CKIK+C+KI				
Sbjct 51	IFECVFSCEIEE-----EGKPKCKPKGKDKKCCNGGWKCKIKLCLKI			93	

Slika 11. *Alignments* prozor proteina *B3A0P0* i njegovog homologa *CAH7646043* [22]

3.2.3. Predikcija bioaktivnosti – *PeptideRanker*

Nakon pronalaska homologa, bilo je potrebno za obje skupine proteina (ukupno 200) napraviti predikciju bioaktivnosti pomoću bioinformatičkog alata *PeptideRanker*. Kako bi se putem alata *PeptideRanker* napravila predikcija, potrebno je sekvencu polipeptidnog lanca napisati tako da se svaka aminokiselina zamijeni slovom abecede koje predstavlja tu aminokiselinu. Takav se zapis naziva FASTA format. FASTA format odnosi se na tekstualne datoteke koje predstavljaju sekvence nukleotida ili aminokiselina u kojima je svaka aminokiselina ili nukleotid zamijenjen jedinstvenim slovom. Sekvence u ovom radu analiziranih proteina i homologa bile su dostupne u FASTA formatu na *UniProt* stranici te su ovako zapisane sekvence preuzete s drugim bitnim podacima o proteinima (slike 10 i 11). U *PeptideRanker* sekvence peptida unosimo na način da će sekvencama svakog peptida biti u zasebnom redu (slika 12).

PeptideRanker: ranks peptide(s) by the predicted probability that the peptide will be bioactive.

Peptides (up to 64kbytes)
(1-letter code, each peptide on a separate line)

```

HGKLTILVLVAALLSTQVIVQGDGDPADRDVAVPRDDNPSGHSKGFHIVLRRAGCPHDPWCG
MHPAHLVLLGCVSLLGASDIPPLPLNLVQFSYLIRKANKYKRPQHNYANYGCYCGSGRGTPVDVDR
CQAHDKCYEDAELKGCYKWTYYCYCGANGPYCKTRTKCQRFVNCNDVVAADCFASYPNRRRYWYFSN
KKRCR
VIGGDECNINEHRFLVALYDGLSGTFLCGGTLINQEWVLTAQHCNRSMLNIZYLGHINKNVCFDDEQRRYP
KKKYFFRCNKNFTKWDEDIRLNRVRFSAHIEPLSLPNPSEDSVCRVHGWGITSPPELTPDVPHCAN
INLFNYTVCRGAYPRMPTKVLCAVLEGGIDTCNRDSGGPLICNGQFQGIIVFNGPDPACAPDQKPGVYTKV
FDYLDWIQSVIAGNTTCS
  
```

Predict Reset

Slika 12. Prikaz sučelja *PeptideRankera* sa unešenim sekvencama originalnih 100 proteina u FASTA formatu [30]

PeptideRanker služi se elementima strojnog učenja kako bi svakoj sekvenci, odnosno peptidu, dodijelio broj u rasponu od 0 do 1 kojim opisuje vjerojatnost da će određeni peptid pokazati bioaktivna svojstva. Bitno je znati da ovaj program ne omogućuje uvid u razinu bioaktivnosti peptida, nego samo daje informaciju o tome koja je vjerojatnost da je taj peptid bioaktivan. Pokretanjem programa dobivaju se rezultati u kojima su vidljive sve unešene sekvence peptida, počnajući od onih koji pokazuju najveću vjerojatnost da su bioaktivni pa do onih koji najvjerojatnije nisu bioaktivni (slika 13). Dio rezultata za promatrane proteine prikazan je na slici 13. Isti je postupak ponovljen sa 100 homologa te su rezultati unešeni u zajedničku *Excel* tablicu.

PeptideRanker

Your query (ID 17218318898469) has been processed

Response:

```

0.999916      MKLVLAIVVILMLLSLSTGAEMSDNHSRSATALRDRLLSPKASICYGTGGRCTDKHCCGWLCCGGPSVGCVVSVAPCK
0.999898      MHPAHLVLLVGVCSLLGASDIPPLPLNLVQFSYLIRCANIKYKRPQWYANYGCGYCGSGGRGTPVDDVDRCCQAHDKCYEDAELGCGYKWTYYCYCGANGPYCKTRTKCQRFCVNCDDVVAAT
0.999897      MKTLLLSPVVVIVCLDLGYTMTCCNQSSQPKTTTNCAGNSCYKKTWSDHRGTIIERGCGCPQVKSIGIKLECHTNECINN
0.999895      MKGMIMLISCLMLIDVWVESKNGYIIEPKGCKYSCSWGSSSTWCNRECKFKKGGSSGYCAWPAWCHCYGLPDNVKIFDYNNKCGK
0.999894      MIAFIVLLSLAVALQSSGTVDFASESSNKKDYRKEIVDKHNALRRSVKPTARNMLRMEWNSRAAQNAKRWADRCTFAHSPPHTRTVGKLRCCGENIFMSTQPFAMSGVWQAWYDEVKKFVYGI
0.999888      MKLTCVLIITVFLFASQLITADYSRDQRQYRAVRLGDENRNFKGARDCCGGQGGCYTQPCCPGLRCRGGGTGGGVCQP
0.999887      MMSKLVLLITICMLFPLTALPLDGDQPADRPAERMDDFISEQHPLFNPIKRCCDWPCTIGCVPCCK
0.999884      MKTLLLSLVVLTIACLDLGYTKCFNDLLTNPKTTELCHRSMYFCFKNSWIAGGVERIERGCSLTCPDIKYNGKIYCCTRDNCNA
0.999877      SLLEFGMILEETGKLAIPSYSSYGCYCGWGGKTPKDATDRCCFVHDCCYGNLPDCNNSKRYRYKVNVAIVCEKGTSCENRICECDKAAAIICFRQNLNTYSKKYMLYPDFLCKGELVC
0.999869      MKTLLLTVMVVIVCLDLGYTLKCNKLVPLFYKTCFAGKNCYKMFVATPKVPVVRGCDVCPKNSALVKYVCCNTDRCN
0.999869      MKLTCMVIITVFLFASQLITADYSRDQRQYRAVRLGDENRNFKGARDCCGGQGGCYTQPCCPGLRCRGGGTGGGVCQP
0.999859      MKLTCVLIIVAVLILTACQFTAADDMEYKWLRLGLSTDXSERGCWLCGLPNACCRGSVCHDYCPR
0.999856      SLLEFGMILEETGKLAIPSYSSYGCYCGWGGKTPKDATDRCCFVHDCCYGNLPDCTPKPDRYKRYKVNVAIVCEQGTSCENRICECDKAAAIICFTKNLNTYSKIYMLYPDFLCKGELVC
0.999853      MKTLLLTVMVVIVCLDLGYTLKCNKLVPLFYKTCFAGKNCYKMFVATPKVPVVRGCDVCPKSSLLVKYVCCNTDRCN
0.999847      MKTLLLTVMVVIVCLDLGYTLKCHNTQLPFIYKTCPEGKNCFKATLKKFPLKFPVVRGCDVCPKNSALLKYVCCSSDKCN
0.999846      MKMTLIAILTCAAVLVLHTTAAEELEAESQLMEIGMPDTELEAVDEERLFECSVSCEIEKEGNKDKKKCKGGWCKCFNMCKVKV
0.999842      MYPAHLVLLAVCVSLGAAASIPARPLNLYQFGNMIQCANHGRPRTHWYMDYGCYCGKGGSGTPVDELDRCCQTHDDCYGAEKLPACNYMMSGPYNYNTSYECNDGELTCKDNDECKAFIC
0.999838      MKLTCVIVAVLFLTACQLITAEYTSRGEQKRALSSDKNKSLTRCNTPTQYCTLHRHCCSLYCHKTIHACA
0.999828      MSSGGLLLGLLTLWAEVTPISGQDRPKKPLGCPRRPQKPCVKECKNDWSCPGQKCCNYGCIDECDRPIFVN
0.999816      MHLSLARSAGLMLLLFAVGNFVGVQPGQITRDVDNGQLADNRRNLQSLRKPMTLFLKSLNKRVSCEYCGYGDPCSSCPTCTSNLLKCM
0.999811      MRCVPVFIILLLLSPSAPSVAHPMTKDDVPOASLHDDAKRTLQVPMNMRGKCCAMLTCCVGR

```

Slika 13. Rezultati *PeptideRanker* predikcije bioaktivnosti [30]

3.2.4. Predikcija toksičnosti – *ToxinPred2*

Osim bioaktivnosti, u sklopu ovog rada napravljena je i predikcija toksičnosti svakog peptida i njegovog homologa. Toksičnost je predviđana pomoću bioinformatičkog alata, *ToxinPred2*. *ToxinPred2* je potrebno pokrenuti pomoću *Pythona*, a nakon brojnih pokušaja i pogreška zaključeno je da je također bitno predikciju raditi na računalu s operativnim sustavom Linux, u ovom slučaju je to bio laptop na *Ubuntu OS*. Za pokretanje sustava je bilo potrebno instalirati brojne biblioteke, što smo napravili pomoću sustava *Pip*. *Pip* je sustav za upravljanje paketima napisan u *Pythonu*, a koristi se za instaliranje i upravljanje softverskim paketima. Instalacija potrebnih datoteka napravljena je sljedećim linijama koda:

```

~$ pip install toxinpred2
~$ pip install numpy
~$ pip install pandas
~$ pip install scikit-learn

```

Nakon instalacije svih potrebnih paketa potrebno je napraviti odgovarajući format datoteke koja će poslužiti kao *input* programu. *Input* će biti tekstualna datoteka FASTA formata u kojoj će biti napisani *UniProtID* peptida te njihova sekvenca u redu ispod. Korištene *input* datoteke su *Originalni_peptidi.fa* i *Homolozi.fa*. Taj format je optimalan za obradu unutar programa *ToxinPred2*. Osim toga, potrebno je naznačiti i *output* odnosno, navesti datoteku u koju će podaci biti spremljeni. *Output* datoteka je .csv formata i stvara se automatski nakon što program obradi podatke. U našem slučaju to su bile datoteke *Originalni_peptidi.csv* i *Homolozi.csv*.

Također, za uspješni rad u *Terminalu* unutar *Linuxa* nije potrebno samo napisati linije koda, nego je potrebno i pozicionirati se u direktorij u kojem se nalaze sve potrebne datoteke koje će program koristiti pri obradi podataka. Dakle, ostale potrebne linije koda su sljedeće:

```
~$ cd toxinpred2-main/
~$ toxinpred2 -i Originalni_peptidi.fa -o Originalni_peptidi.csv -d 2
~$ toxinpred2 -i Homolozi.fa -o Homolozi.csv -d 2
```

Nakon obrade podataka program prikaže parametre pretraživanja te budu vidljiva imena *input* i *output* datoteka (slika 14)

```
helena@helena-Vostro-3520:~$ cd toxinpred2-main/
helena@helena-Vostro-3520:~/toxinpred2-main$ toxinpred2 -i Originalni_peptidi.fa -o Originalni_peptidi.csv -d 2
#####
# The program ToxinPred2.0 is developed for predicting Toxin and non toxin #
# protein from their primary sequence, developed by Prof G. P. S. Raghava's group. #
# #####
Summary of Parameters:
Input File: Originalni_peptidi.fa ; Model: 1 ; Threshold: 0.6
Output File: Originalni_peptidi.csv ; Display: 2
```

Slika 14. Prikaz rezultata predikcije toksičnosti pomoću *ToxinPred2*

3.2.5. *AlphaFold*

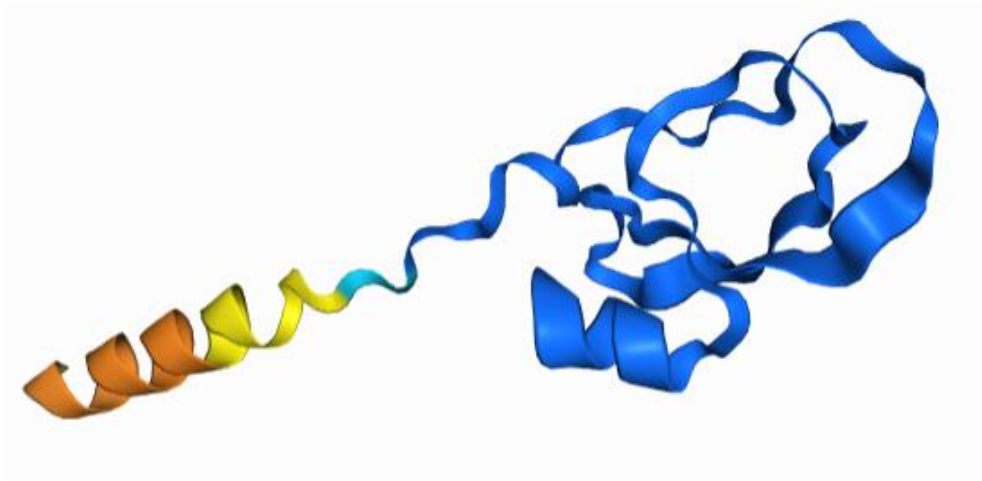
Nakon određivanja bioaktivnosti i toksičnosti, potrebno je bilo napraviti predikciju trodimenzionalne strukture proteina. Strukture proteina su generirane pomoću alata *AlphaFold* koji na temelju polinukleotidne sekvence stvara strukturu te ju obilježava bojama ovisno o tome koliko je siguran da je pravi izgled tog proteina upravo takav kakav je generiran (Slika 15).



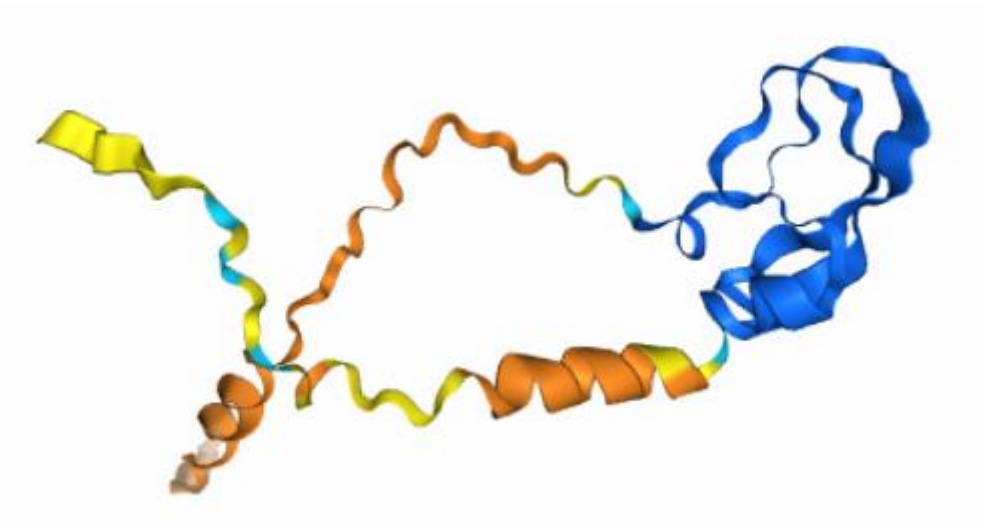
Slika 15. Boje kojima je obojana struktura proteina u programu *AlphaFold* [31]

Generirano je više struktura proteina, a od njih su izdvojeni protein C8YJ95 (slika 16.) i njegov homolog XP_019623731 (slika 17). Originalni protein je dugačak 76 aminokiselina, dok je njegov homolog dugačak 133 aminokiseline čime oboje pripadaju skupini dugih proteina kada

je riječ o proteinima na kojima je program *PeptideRanker* treniran. Na *UniProt* bazi podataka je dostupna informacija o toksičnosti tog proteina (C8YJ95) te je navedeno da je to toksičan protein koji sudjeluje u nekim biološkim putevima stoga zaključujemo da je i bioaktivan. Alatima *PeptideRanker* i *ToxinPred2* smo to i dokazali. Istim alatima smo radili predikciju i za njegov homolog čiji su rezultati, iako sa nižom vrijednosti tj. sigurnošću, isto ukazivali na toksičnost i bioaktivnost tog homolognog proteina.



Slika 16. Prikaz proteina C8YJ95



Slika 17. Prikaz proteina XP_019623731 (homolog proteina C8YJ95)

E-vrijednost homologa u sklopu poravnanja je $4E-19$, što je niska vrijednost odnosno upućuje na visoki stupanj homologije između sekvenci originalnog proteina i njegovog homologa. Sličnosti među sekvencama je sada moguće i vizualizirati preko rezultata *AlphaFold* predikcije na kojima je vidljiva tamno plava regija koja je jako slična uspoređujući te regije na originalnom i homolognom proteinu (Slike 16. i 17.). Iako BLAST daje dobru procjenu homologije, za uspješno određivanje aktivnosti homologa ili nekog nepoznatog proteina je potrebno analizirati i njegovu 3D strukturu koju je potrebno koristiti kao završnu potvrdu da dva proteina imaju istu ili sličnu funkciju.

Osim plavih regija, na 3D prikazima proteina su vidljive i narančaste te žute regije koje se javljaju na regijama za koje *AlphaFold* nije siguran kako pravilno generirati. Takve regije su najčešće razgranate strukture koje nastaju kao rezultat jakog međusobnog odbijanja čestica uzrokovanim visokim neto nabojem te nedostatka hidrofobnih aminokiselina čiji bi aminokiselinski ostaci pružali viši stupanj zbijenosti. Te regije su zapravo regije u kojima nedostaje fiksna tj. uređena 3D struktura zbog čega takve proteine nazivamo intrinzično poremećenim proteinima (IDP). U takvim proteinima se neke aminokiseline javljaju češće od drugih. One poput Ala, Arg, Gly, Gln, Glu, Lys, Pro i Ser su česte dok se ostale aminokiseline rjeđe javljaju u tim regijama [32]. Određivanje strukture takvih proteina se pokazalo težim nego kod običnih proteina upravo zbog njihove raznolike strukture. Često sudjeluju u raznim regulatornim mehanizmima u organizmu, ali mogu biti i uzročnici mnogih bolesti poput raka ili kardiovaskularnih bolesti.

4. REZULTATI I RASPRAVA

Tijekom izrade ovog rada bilo je prvo potrebno odabrati 100 proteina koji su homologni istom broju proteina odabranih iz *ToxProt* baze podataka. Za proteine *ToxProt* baze podataka je poznato da su toksični i bioaktivni, dok je za homologe to tek bilo potrebno otkriti.

4.1. Predikcija bioaktivnosti pomoću alata *PeptideRanker*

Predikcija bioaktivnosti rađena je pomoću bioinformatičkog alata *PeptideRanker* čiji se rad bazira na elementima strojnog učenja. Izvor podataka koji je korišten za učenje *PeptideRankera* raspoznavanju bioaktivnih peptida sastojao se većinom od kraćih peptida odnosno onih koji su kraći od 20 aminokiselina [33]. Posljedično, najveća prednost ovog programa bila je predikcija bioaktivnosti kraćih peptida, dok nam kod dužih peptida program ne daje rezultate jednako visoke točnosti. Nadalje, budući da iznimno veliki peptidi/proteini mogu sadržavati veći broj funkcionalnih domena, dužim je peptidima teže odrediti bioaktivnost zbog međusobno različitih aktivnosti tih domena [33].

4.1.1. Utjecaj duljine polipeptidnog lanca na točnost predikcije bioaktivnosti

Kako bi provjerili je li *PeptideRanker* uistinu točniji pri određivanju bioaktivnosti kraćih peptida, prvo je bilo potrebno odrediti koliko je dugih, a koliko kratkih proteina u našem uzorku. Od 100 nasumično odabranih peptida i proteina, njih samo 14 sadrži manje od 30 aminokiselina u polipeptidnom lancu, a ukoliko se vodimo radom Mooney C. i sur. (2012.) koji poručuje da su kratki peptidi oni koji sadrže manje od 20 aminokiselina u lancu, onda je njih samo 11. Samim time analizirani podaci nisu idealni za predikciju bioaktivnosti pomoću *PeptideRankera*, ali pomoću ovih podataka je svejedno bilo moguće dokazati prethodnu tvrdnju koja navodi da je prava snaga ovog alata predikcija proteina kraćih od 20 aminokiselina.

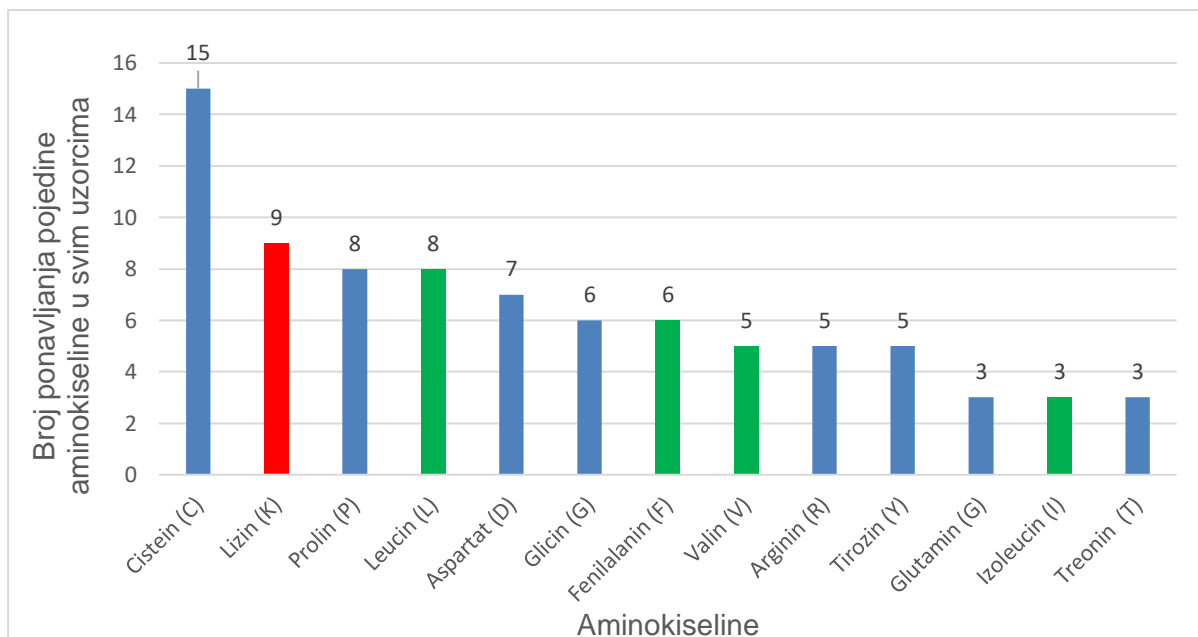
Tu tvrdnju podupire i činjenica da je za veliki broj originalnih proteina koje smo nasumično odabrali iz *ToxProt* baze podataka, *PeptideRanker* predvidio da nisu biološki aktivni, iako znamo da to nije istina. Od 100 nasumično odabranih peptida, za samo njih 20 je bila poznata biološka funkcija te je navedena u *ToxProt* bazi podataka. Od tih 20 bioaktivnih proteina, njih je čak 18 kategorizirano kao dugački peptidi. S obzirom da znamo da su ti peptidi bioaktivni, provjereno je u koliko će slučajeva program kvalitetno odrediti bioaktivnost tih peptida. Od 20 peptida koji pokazuju biološku aktivnost, program je njih 13 označio kao biološki aktivne, tj. imao je točnost od 65% (13/20). Ono što je zanimljivije je to da je program za čak 7 bioaktivnih proteina pretpostavio da nisu bioaktivni, a pošto je od 100 nasumično odabranih proteina za njih 15 pretpostavio da nisu bioaktivni, tih 7 proteina zapravo čini 46.67% svih proteina u uzorku označenih kao neaktivni.

Budući da se program pokazao nepouzdanim pri određivanju bioaktivnosti dugačkih peptida iz baze podataka ovog znanstvenog rada, sve buduće zaključke vezane uz bioaktivnost ćemo donositi na temelju peptida kraćih od 20 aminokiselina jer je u tom području *PeptideRanker* ipak najpouzdaniji.

4.1.2. Utjecaj aminokislinskog sastava polipeptidnog lanca na bioaktivnost proteina

Također, promatran je i utjecaj aminokislinskog sastava polipeptidnog lanca na bioaktivnost proteina ili peptida. Poznato je da se u bioaktivnim peptidima često javljaju hidrofobni aminokiselinski ostaci poput leucina, izoleucina, fenilalanina i valina. Njihova prisutnost povećava bioaktivnost peptida tako što omogućuje i potiče interakcije sa staničnom membranom što je jedna od primarnih aktivnosti bioaktivnih peptida. Osim hidrofobnih aminokislinskih ostataka, u bioaktivnim peptidima često nailazimo i na pozitivno nabijene aminokiselinske ostatke (arginin i lizin) te na aromatske aminokiseline poput triptofana ili tirozina. Pozitivno nabijeni aminokiselinski ostaci ulaze u interakcije s negativno nabijenom membranom te ostalim negativno nabijenim dijelovima stanice poput nukleinskih kiselina. Aromatski aminokiselinski ostaci povećavaju afinitet vezanja peptida na odgovarajuću molekulu čime isto povećavaju biološki učinak peptida. Ove su se aktivnosti pokazale iznimno bitnima kod antimikrobnih i protukancerogenih peptida [33].

Kako bi se ustanovilo javljaju li se često prethodno navedene hidrofobne i pozitivno nabijene aminokiseline u promatranim bioaktivnim proteinima, potrebno je analizirati njihove sekvence. Analizirali smo samo kratke peptide, odnosno one kraće od 20 aminokiselina i to samo one koje je *PeptideRanker* klasificirao kao bioaktivne. Takvih proteina je 7 - *C0HKQ6*, *P58810*, *C0HKZ3*, *P0C642*, *P0C8J2*, *P86309*, *P86820*.



Slika 18. Učestalost pojedinih aminokiselina u polipeptidnim lancima kratkih i bioaktivnih proteina iz uzorka 100 nasumično odabranih proteina iz *ToxProt* baze podataka Crvenom su označene pozitivno nabijene, a zelenom hidrofobne aminokiseline.

Iz grafičkog prikaza rezultata analize aminokiselinskog sastava polipeptidnih lanaca (slika 18) vidljivo je da je cistein najčešća sastavnica polipeptidnog lanca bioaktivnog peptida ili proteina iz razloga što dva pobočna ogranka cisteina mogu međusobno stvarati disulfidne mostove i na taj način stabilizirati peptid, a time i osigurati određenu aktivnost, primjer čega je inzulin. Nadalje, ovim rezultatima analize je dokazana prethodna tvrdnja o utjecaju aminokiselinskog sastava na bioaktivnost proteina. Od hidrofobnih aminokiselina, u ovim primjerima prevladavaju leucin i fenilalanin, valin i izoleucin, a od pozitivno nabijenih lizin. Time se može zaključiti da se kod promatranih kratkih bioaktivnih peptida javljaju upravo one aminokiseline koje je očekivano susresti pri analizi sekvence bioaktivnih peptida.

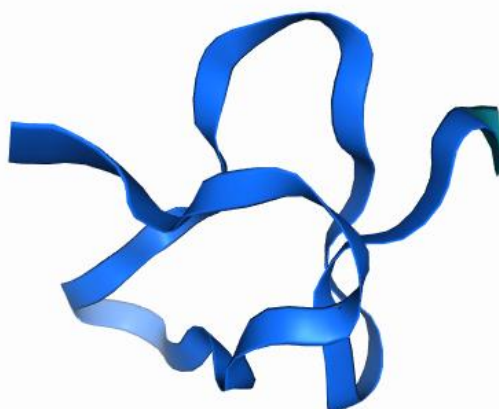
4.1.3. Utjecaj E-vrijednosti na uspješnost predikcije bioaktivnosti homolognih proteina

Od promatranih homologa, za njih 17 je *PeptideRanker* pretpostavio da nisu bioaktivni. 7 homologa se razlikuje po bioaktivnosti od originalnih proteina, a samo 1 od takvih je homolog s velikom E-vrijednošću izabran u svrhu istraživanja učinka visoke vrijednosti parametra *E-vrijednosti*. Pošto je samo 1 od ukupno 11 homologa s visokom *E-vrijednošću* pokazao različitu bioaktivnost u odnosu na originalni protein, može se zaključiti da *E-vrijednost* u ovome slučaju nije bio presudni faktor za određivanje bioaktivnosti peptida. Razlog tome bi mogao biti slučajan jer je većina takvih peptida bila duža od 20 aminokiselina pa je i razina točnosti određivanja bioaktivnosti programa *PeptideRanker* niska.

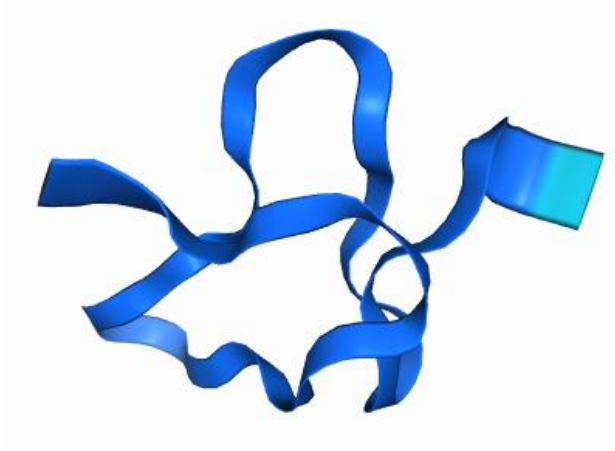
4.1.4. Primjeri uspješne predikcije bioaktivnosti homolognih proteina

Samo 3 homologa su bila kraća od 30 aminokiselina, ali kako je poznato da je *PeptideRanker* precizniji pri predikciji kraćih peptida, njima će se moći s najvećom sigurnošću pretpostaviti aktivnost. Ta 3 peptida su *P0DQO2*, *P0DQT3* i *P86820*, a redom i njihovi homologzi *7A64_A*, *6Q08_A* i *C0HKA7*.

7A64_A (slika 19.) je potpuno homologan s *P0DQO2* (slika 20.) što se može vidjeti po iznosima *Identiteti*, *Pozitivni* i *Praznine* koji ukazuju na to da su sekvence identične, samo se u homologu na C-terminalnom dijelu polipeptidnog lanca nalazi još jedna dodatna aminokiselina. Aktivnost ovoga homologa trebala bi po svim saznanjima biti ista kao i od originalnog proteina, stoga se može zaključiti da isto kao i originalni protein sudjeluje u regulaciji ionskih kanala te njihove propusnosti. Ionski kanali imaju ključnu ulogu u međustaničnoj signalizaciji u živčanom sustavu. Aktivnost ionskih kanala može se promijeniti pod utjecajem metaboličkih reakcija (npr. fosforilacije kanalnog proteina), drugih iona što djeluju kao blokatori kanala, te toksina, otrova, droga i lijekova.

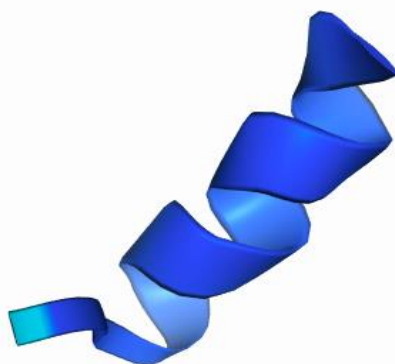


Slika 19. AlphaFold struktura proteina P0DQO2

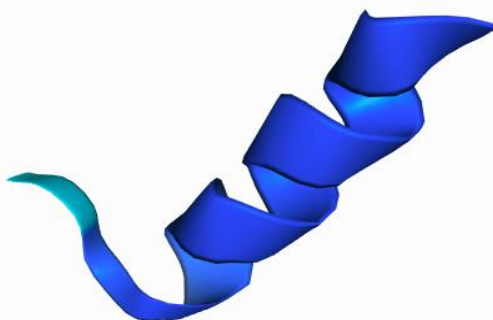


Slika 20. AlphaFold struktura proteina 7A64_A (homolog proteina P0DQO2)

Drugi homolog, 6Q08_A (slika 21.), pokazuje nešto niže vrijednosti *Identiteti*, *Pozitivni*, *Praznine*. Analizirajući sekvencu vidljivo je da je na N-terminalnom dijelu lanca 8 aminokiselina identično, dok se prema C-terminalnom dijelu lanca aminokiseline češće razlikuju. Zbog jako sličnih rezultata bioaktivnosti i toksičnosti može se s većom sigurnošću zaključiti da će homolog imati istu aktivnosti kao i originalni protein. Stoga, jednako kao i originalni protein *P0DQT3* (slika 22), homolog ima antimikrobnu i antibiotsku aktivnost. Prema strukturama (slike 21. i 22.) generiranim programom *AlphaFold*, vidljivo je da su i po strukturi ovi proteini slični, stoga se pretpostavlja da bi im i funkcija trebala biti ista. Također, vidljiva je i struktura α -uzvojnice što je očekivano za antimikrobne i antibiotske peptide.



Slika 21. AlphaFold struktura *P0DQT3* peptida



Slika 22. AlphaFold struktura 6Q08_A peptida (homolog proteina *P0DQT3*)

4.2. Predikcija toksičnosti pomoću alata *ToxinPred2*

U ovome je radu napravljena i predikcija toksičnosti pomoću drugog alata – *ToxinPred2*. *ToxinPred2* je također baziran na elementima strojnog učenja. Program je učen na bazi podataka bogatoj raznim molekulama stoga je preciznost njegove predikcije iznimno visoka na uzorcima različite duljine i sastava polipeptidnog lanca.

4.2.1. Sastav toksičnih peptida i utjecaj visoke E-vrijednosti na točnost predikcije toksičnosti proteina

Prema Sharma N. i sur. (2022.) [34], znanstvenicima koji su razvili ovaj program, postoje određene aminokiseline koje se češće javljaju u toksičnim te one koje se češće javljaju u netoksičnim peptidima. Za toksične peptide to su cistein, lizin, glicin i triptofan, dok se u peptidima koji ne pokazuju toksične indikacije češće javljaju alanin, glutamat, izoleucin, leucin, serin.

Od 100 nasumično odabranih proteina, njih samo 8 nije toksično. Takvi rezultati su bili i očekivani s obzirom na to da se radi o bazi podataka sastavljenoj od otrova i toksina. Uspoređujući aminokiselinski sastav lanaca proteina nasumično odabranih iz *ToxProt* baze podataka s informacijama istraživanja Sharma N. i sur. (2022.) [34] vidljivo je da su i u ovom radu promatrani toksični peptidi sadržavali velike udjele aminokiselina specifičnih za toksične peptide. Za analizu učestalosti aminokiselinskih ostataka uzeta su u obzir 92 peptida klasificirana kao toksična. Analizom su dobiveni rezultati koji govore da se cistein pojavio čak 663 puta u lancima toksičnih peptida, glicin 662, lizin 621, dok se triptofan pojavljivao najrjeđe - 150 puta. Od homologa, njih 18 je klasificirano netoksičnim, iz čega je vidljivo da se rezultati predikcije peptida i njihovih homologa ne poklapaju u potpunosti. Na kraju se utjecaj izbora homologa koji iskazuju veću E-vrijednost nije pokazao utjecajnim na rezultate. Gotovo svi homolozi sa većom E-vrijednosti su klasificirani jednako kao i proteini s kojima dijele redoslijed aminokiselina u polipeptidnim lancima. Od 11 homologa s E-vrijednosti vrijednosti većom od 0, samo 2 su klasificirana drugačije od njihovih originalnih peptida.

5. ZAKLJUČCI

1. Uspješno su pronađeni homolozi 100 nasumično odabranih proteina iz *ToxProt* baze podataka
2. Napravljena je predikcija bioaktivnosti i toksičnosti proteina na temelju njihove polipeptidne sekvence pomoću bioinformatičkih alata *PeptideRanker* i *ToxinPred2*
3. Homolozi za koje su rezultati predikcija bili najtočniji su 3D generirani pomoću AlphaFold-a kako bi se potvrdila njihova sličnost sa originalnim proteinima, a time i odredila aktivnost

6. POPIS LITERATURE

- [1] Zaky AA, Simal-Gandara J, Eun JB, Shim JH, Abd El-Aty AM (2022) Bioactivities, Applications, Safety, and Health Benefits of Bioactive Peptides From Food and By-Products: A Review. *Frontiers in nutrition* 8, 815640. <https://doi.org/10.3389/fnut.2021.815640>
- [2] Rogers, K. (2016) What Is the Difference Between a Peptide and a Protein?. Encyclopedia Britannica. <https://www.britannica.com/story/what-is-the-difference-between-a-peptide-and-a-protein>. Pristupljeno 16. lipnja 2024.
- [3] Akbarian M, Khani A, Eghbalpour S, Uversky VN (2022) Bioactive Peptides: Synthesis, Sources, Applications, and Proposed Mechanisms of Action. *International journal of molecular sciences* 23(3), 1445. <https://doi.org/10.3390/ijms23031445>
- [4] Nature Education (2014) Essentials of Cell Biology. Unit 2: How Do Cells Decode Genetic Information into Functional Proteins, <https://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/122996920/>. Pristupljeno 21. srpnja 2024.
- [5] Teparić, R (2021) Interna skripta iz kolegija Biokemija 1. Prehrambeno-biotehnološki fakultet, Sveučilište u Zagrebu, Zagreb.
- [6] Brunning A (2014) Compound Interest: A Brief Guide to the Twenty Common Amino Acids. <https://www.compoundchem.com/2014/09/16/aminoacids/>. Pristupljeno 21. srpnja 2024.
- [7] Ahern K, Rajagopal I, Tan T (2024) Structure & Function- Proteins I – Biology Libretexts. <https://bio.libretexts.org/@go/page/7810>. Pristupljeno 22. srpnja 2024.
- [8] Faiz M, Ibrahim F, Nasim S, Shakeel CS, Saifullah B (2023) Structural Revelation of Human Transmembrane Proteins by SVM. *Pakistan Journal of Biotechnology* 20, 33-41. <https://doi.org/10.34016/pjbt.2023.20.01.667>
- [9] Nacar C (2023) The Frequencies of Amino Acids in Secondary Structural Elements of Globular Proteins. *Clinical and Experimental Health Sciences*, 13(2), 261-266. <https://doi.org/10.33808/clinexphealthsci.1239176>
- [10] Plata C (2019) From proteins to grains: a journey through simple models. https://www.researchgate.net/publication/331034573_From_proteins_to_grains_a_journey_through_simple_models. Pristupljeno 1. kolovoza 2024.
- [11] Podolskiy DI, Gladyshev VN (2016) Intrinsic Versus Extrinsic Cancer Risk Factors and Aging. *Trends in molecular medicine*, 22(10), 833–834. <https://doi.org/10.1016/j.molmed.2016.08.001>
- [12] Gupta RC (2017) Peptides and Proteins: Structure and Functions PPT. <https://www.slideshare.net/slideshow/peptides-amp-proteins-structure-amp-functions/78159370>. Pristupljeno 1. kolovoza 2024.
- [13] Fadimu GJ, Le TT, Gill H, Farahnaky A, Olatunde OO, Truong T (2022) Enhancing the Biological Activities of Food Protein-Derived Peptides Using Non-Thermal Technologies: A Review. *Foods (Basel, Switzerland)*, 11(13), 1823. <https://doi.org/10.3390/foods11131823>

- [14] Sánchez A, Vázquez A (2017) Bioactive peptides: A review. *Food Qual. Saf.* 1:29–46. <https://doi.org/10.1093/fqsafe/fyx006>
- [15] King G (2013) Venoms to Drugs: Translating Venom Peptides into Therapeutics. *Australian Biochemist* 44(3), 13-15.
- [16] Hodgson E (2012) Toxins and Venoms. *Progress in Molecular Biology and Translational Science* 112, 373–415. <https://doi.org/10.1016/B978-0-12-415813-9.00014-3>
- [17] Xiong X, Menting JG, Disotuar MM *et al.* (2020) A structurally minimized yet fully active insulin based on cone-snail venom insulin principles. *Nat Struct Mol Biol* 27, 615–624. <https://doi.org/10.1038/s41594-020-0430-8>
- [18] The UniProt Consortium (2023) UniProt: the Universal Protein Knowledgebase in 2023, *Nucleic Acids Research* 51(1), 523-531. <https://doi.org/10.1093/nar/gkac1052>
- [19] UniProt. <https://www.uniprot.org/>. Pristupljeno 08. lipnja 2024.
- [20] UniProt (2023) Animal toxin annotation project. <https://www.uniprot.org/help/Toxins>. Pristupljeno 08. lipnja 2024.
- [21] Cordonnier T, Shaffer C, Leung W, Igin SCR (2023) BLAST Exercise: Detecting and Interpreting Genetic Homology. Adaptation of „Detecting and Interpreting Genetic Homology by dr. Jeremy Buhler“. [https://community.gep.wustl.edu/repository/course_materials_WU/annotation/Detecting Interpreting Genetic Homology.pdf](https://community.gep.wustl.edu/repository/course_materials_WU/annotation/Detecting_Interpreting_Genetic_Homology.pdf). Pristupljeno 28. srpnja 2024.
- [22] BLAST: Basic Local Alignment Search Tool. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Pristupljeno 08. lipnja 2024.
- [23] Python Software Foundation. <https://www.python.org/>. Pristupljeno 10. kolovoza 2024.
- [24] scikit-learn: machine learning in Python. <https://scikit-learn.org/stable/index.html>. Pristupljeno 24. srpnja 2024.
- [25] Pandas Python- What Is It and Why Does It Matter?. <https://www.nvidia.com/en-us/glossary/pandas-python/>. Pristupljeno 24. srpnja 2024.
- [26] Number Generator. <https://numbergenerator.org/>. Pristupljeno 24. srpnja 2024.
- [27] Kerfeld CA, Scott KM (2011) Using BLAST to teach "E-value-tionary" concepts. *PLoS biology*, 9(2), e1001014. <https://doi.org/10.1371/journal.pbio.1001014>
- [28] KA-05225 – NLM Customer Support Center. <https://support.nlm.nih.gov/knowledgebase/article/KA-05225/en-us>. Pristupljeno 08. srpnja 2024.
- [29] Biostars – Bioinformatics Explained. <https://www.biostars.org/p/9486511/> Pristupljeno 18. srpnja 2024.
- [30] PeptideRanker. <http://distilldeep.ucd.ie/PeptideRanker/>. Pristupljeno 24. lipnja 2024.
- [31] AlphaFold Server. <https://alphafoldserver.com/>. Pristupljeno 22. lipnja 2024.

[32] Trivedi R, Nagarajaram HA (2022) Intrinsically Disordered Proteins: An Overview. *International journal of molecular sciences* 23(22), 14050.
<https://doi.org/10.3390/ijms232214050>

[33] Mooney C, Haslam NJ, Pollastri G, Shields DC (2012) Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity. *PLoS one* 7(10), e45012.
<https://doi.org/10.1371/journal.pone.0045012>

[34] Neelam S, Leimarembi DN, Shipra J, Gajendra PSR (2022) ToxinPred2: an improved method for predicting toxicity of proteins, *Briefings in Bioinformatics* 23(5).
<https://doi.org/10.1093/bib/bbac174>

PLAGIRANJE

Izjava o izvornosti

Ja Mihovil Pupačić izjavljujem da je ovaj završni rad izvorni rezultat mojeg rada te da se u njegovoj izradi nisam koristio/la drugim izvorima, osim onih koji su u njemu navedeni.

Vlastoručni potpis