

UNIVERZA V LJUBLJANI
FAKULTETA ZA KEMIJO IN KEMIJSKO TEHNOLOGIJO

DIPLOMSKO DELO

Samo Purič

Ljubljana, 2019

UNIVERZA V LJUBLJANI
FAKULTETA ZA KEMIJO IN KEMIJSKO TEHNOLOGIJO
UNIVERZITETNI ŠTUDIJSKI PROGRAM 1. STOPNJE BIOKEMIJA

**Preverjanje napovedi heličnosti
nestrukturiranih peptidov z
eksperimentalnimi podatki**

DIPLOMSKO DELO

Samo Purič

MENTOR: prof. dr. Jurij Lah

Ljubljana, 2019

IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani Samo Purič sem avtor diplomskega dela z naslovom:

Preverjanje napovedi heličnosti nestrukturiranih peptidov z eksperimentalnimi podatki

S svojim podpisom zagotavljam, da:

- je diplomsko delo rezultat mojega raziskovalnega dela pod mentorstvom prof. dr. Jurija Laha;
- 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, dne _____

Podpis avtorja:

Zahvala

Za njuno mentorstvo, vodenje in nasvete pri izdelavi tega dela se zahvaljujem prof. dr. Juriju Lahu in dr. Sanu Hadžiju.

Kazalo

Zahvala	ii
Kazalo	iii
Kazalo slik	v
Seznam uporabljenih kratic in simbolov	vi
Povzetek.....	vii
Abstract.....	viii
1 Uvod	1
1.1 Intrinzično nestrukturirani proteini.....	1
1.2 α -vijačnica	2
1.3 Model prehoda vijačnica-naključni klobčič	4
1.3.1 Lifson-Roig (LR)	4
1.4 AGADIR	5
1.4.1 Približek, ki upošteva enkratno nukleacijo.....	5
1.4.2 Terminalne pozicije ter N1, N2 in N3 preference.....	5
1.4.3 Interakcije stranskih skupin	6
2 Namen dela	7
3 Materiali in metode	8
3.1 Zbiranje podatkov	8
3.1.1 Eksperimentalni podatki iz literature	8
3.1.2 Eksperimentalni podatki pridobljeni v laboratoriju	8
3.2 Določanje deleža α -vijačnice.....	8
3.3 Napoved deleža α -vijačnice – AGADIR.....	9
3.4 Določanje povezanosti med eksperimentalnimi podatki in izračuni s pomočjo algoritma AGADIR.....	10
3.5 Analiza peptidov	10
4 Rezultati.....	11
4.1 Baza podatkov	11
4.2 Primerjava eksperimentalne vrednosti z izračunano.....	12
4.3 Analiza sekvenc	13
4.3.1 Interakcije stranskih skupin	14
5 Razprava	16
6 Sklepne ugotovitve.....	17
7 Literatura.....	18

8	Priloge.....	19
8.1	Priloga 1	19

Kazalo slik

Slika 1: Prikaz proste energije (G) globularnega proteina (A), z izrazitim energijskim minimumom, in IDP-ja (B), z razgibano površino brez izrazite termodinamsko ugodne strukture pri sobni temperaturi. Prirejeno po C. K. Fisher, C. M. Stultz: "Constructing Ensembles for Intrinsically Disordered Proteins".	1
Slika 2: Graf hidrofobnosti v odvisnosti od neto naboja ("Uversky plot"). Na x-osi je nanešena povprečna normalizirana hidrofobnost ak-ostankov polipeptidne verige po Kyte-Doolittle skali (Tabela 1). Na y-osi so nanešeni povprečni neto naboji polipeptidne verige. Z neprekinjeno premico je označena meja med urejeno in neurejeno strukturo z enačbo $H = (R + 1,151)/2,785$ (kjer je H hidrofobnost in R neto naboj). IDP-ji se v tem grafu pojavljajo na levi strani ločnice pri nižjih vrednostih povprečne normalizirane hidrofobnosti, medtem ko so globularni, urejeni proteini na desni strani ločnice pri višjih vrednostih hidrofobnosti. Prirejeno po: V. N. Uversky and A. K. Dunker, "Understanding protein non-folding".	2
Slika 3: Shema α -vijačnice. Z črtkano črto so označene vodikove vezi, ki nastajajo med aaminsko skupino i-tega ostanka in karboksilno skupino i+4 ostanka peptidnega ogrodja. Prav tako je dobro vidna orientacija in pozicija stranskih skupin (črne krogle) aminokislin v prostoru kar omogoča nastanek i, i+3 in i, i+4 interakcij. Prirejeno po: J. Feher, "Protein Structure".	3
Slika 4: Lifson-Roig oznake in uteži za α -vijačnico. S krogi so označeni osnovni elementi modela LR, to so ostanki. v je utež za iniciacijo (nastanek α -vijačnice), w pa je utež za propagacijo (podaljševanje α -vijačnice kot posledica tvorbe novih H-vezi). Z oznako h so označeni ostanki ki ležijo v ozkem območju kombinacije kotov (Φ, ψ) ki se ujemajo z formacijo α -vijačnice, medtem ko so z oznako c označeni preostali, ne-helični ostanki (ti ostanki hkrati služijo tudi kot referenca in so obteženi z statistično utežjo 1). Prirejeno po: A. J. Doig, "Recent advances in helix-coil theory".	4
Slika 5: Primer vnosa sekvence (levo) in eksperimentalnih parametrov (desno) v spletno orodje AGADIR.	9
Slika 6: Primer izhodnih podatkov spletnega orodja AGADIR. Orodje za vsak aminokislinski ostanek posebej izračuna verjetnost za nastanek α -vijačnice in nato poda oceno heličnosti peptida v odstotkih (angl. »Percentage helix«, označeno z rdečo).	9
Slika 7: Graf eksperimentalnih vrednosti heličnosti v odvisnosti od izračunane vrednosti po AGADIR-ju. Z modro barvo je označen set 40ih peptidov iz literature in laboratorijskih meritev. Z oranžno barvo so označeni alaninski peptidi. Na grafu sta zapisani tudi vrednosti enačb premic za posamezen niz podatkov.	12
Slika 8: Graf povprečne normalizirane hidrofobnosti po lestvici Kyte-Doolittle v odvisnosti od povprečnega neto naboja peptidov (»Uversky Plot«). Z oranžno, črtkano črto je narisana teoretično določena ločnica med IDP-ji na eni in globularnimi proteini na drugi strani.	13
Slika 9: Prikaz razporeditve števila ugodnih (levo) in neugodnih (desno) i, i+3 in i, i+4 interakcij med stranskimi skupinami v setu preučevanih peptidov.	14
Slika 10: Tortni diagram razporeditve i, i+3 in i, i+4 interakcij po posameznih kategorijah parov aminokislin.	15

Seznam uporabljenih kratic in simbolov

IDP – Intrinzično nestrukturiran protein (*angl.* »Intrinsically disordered protein«)

IDR – Intrinzično nestrukturirana regija (*angl.* »Intrinsically disordered region«)

CD – Cirkularni dikroizem (*angl.* »Circular dichroism«)

MRE – Srednja molarna eliptičnost aminokislinskih ostankov (*angl.* »Mean residue molar ellipticity«)

ak - aminokislina

NMR – jedrska magnetna resonanca

SAXS – ozkokotno rentgensko sipanje (*angl.* »Small-Angle X-Ray Scattering«)

LR – *Lifson-Roig*

C_{α} – alfa ogljikov atom (C-atom aminokislina na katerega je vezana stranska skupina – R)

ΔG_{Hel} – sprememba Gibbsove proste energije za prehod iz neurejenega zvitja v α -vijačnico

ΔG_{Int} – sprememba Gibbsove proste energije - intrinzična nagnjenost ak-ostankov k tvorbi α -vijačnice

ΔG_{Hbond} – sprememba Gibbsove proste energije, ki opisuje vodikove vezi med atomi polipeptidnega ogrodja

ΔG_{SD} – sprememba Gibbsove proste energije – interakcije stranskih skupin ak-ostankov

ΔG_{dipol} – sprememba Gibbsove proste energije – interakcija dipola z nabitimi skupinami

ΔG_{nonH} – sprememba Gibbsove proste energije – interakcija ostankov, ki niso v helični konformaciji z vijačnico

$\Delta G_{electrostat}$ – sprememba Gibbsove proste energije – interakcije med nabitimi skupinami

f_H – faktor heličnosti

$[\theta]_{OBS}$ – izmerjena eliptičnost peptida

$[\theta]_C$ – eliptičnost peptida s 100-odstotnim naključnim zvitjem

$[\theta]_H$ – eliptičnost peptida s 100-odstotnim heličnim zvitjem

Povzetek

Za intrinzično nestrukturirane proteine (IDP-je, *angl.* »Intrinsically disordered proteins«) je značilna odsotnost enoznačno definirane 3D strukture. Njihova prostorska razporeditev se konstantno spreminja, pri določenih pogojih jih tako najdemo kot nabor različnih konformacijskih oblik, med katerimi, energijsko gledano, nobena zelo ne izstopa. Neurejena je lahko tudi samo regija oziroma domena proteina (IDR, *angl.* »Intrinsically disordered region«). Zaradi svoje funkcionalnosti so IDP-ji udeleženi v signalizaciji, regulaciji in drugih celičnih procesih. Najbolj pogost strukturni element ki ga najdemo v IDP-jih (in tudi ostalih proteinih) je α -vijačnica.

V tem diplomskem delu smo se osredotočili na algoritem AGADIR, ki se uporablja za napovedovanje vsebnosti α -vijačnice v monomernih peptidih na osnovi analize aminokislinskega zaporedja. Zbrali smo eksperimentalne podatke heličnosti peptidov IDP-jev pri katerih zaznamo težnjo po tvorjenju α -vijačnice iz literature (CD spektri) in te vrednosti primerjali z napovedjo algoritma AGADIR. Poleg tega smo opravili tudi analizo aminokislinskih zaporedij peptidov na podlagi lastnosti posameznih ak-ostankov in prešteli interakcije stranskih skupin ak-ostankov. Kot referenčen set podatkov smo uporabili sedem alaninskih peptidov, s pomočjo katerih je algoritem AGADIR umerjen.

Na podlagi rezultatov lahko zaključimo, da AGADIR delež α -vijačnice v peptidih IDP-jev močno podceni, vendar pa z veliko večjo natančnostjo napove vsebnost α -vijačnice v alaninskih peptidih. Rezultati analize sekvenc in interakcij stranskih skupin kažejo na to, da AGADIR najverjetneje podceni ugodne energijske prispevke določenih interakcij, ki pripomorejo k stabilnosti α -vijačnice.

Ključne besede: Intrinzično nestrukturiran protein, α -vijačnica, AGADIR

Comparison of α -helical content prediction for intrinsically disordered peptides with experimental data

Abstract

Intrinsically disordered proteins (IDPs) are proteins that lack a well-defined 3D structure. Their spatial arrangement is constantly changing, and we can only describe them (at a given temperature) as a set of distinct conformational shapes with very similar conformational energies. Apart from IDPs, intrinsically disordered regions (IDRs) are also very common. Because of their functional properties, IDPs are involved in many cellular processes such as signalization and regulation. The most common structural element that is found in IDPs is the α -helix.

In this work, we focus on AGADIR, which is an algorithm for predicting α -helical content in monomeric peptides based on amino acid sequence. We gathered experimental data obtained with CD spectroscopy of peptides from IDPs with a significant helix propensity and compared the values with the ones determined by AGADIR. In addition, we also ran an analysis of amino acid sequences of the peptides used and the possible interactions between functional groups. As a reference set of data, we used seven alanine-based peptides.

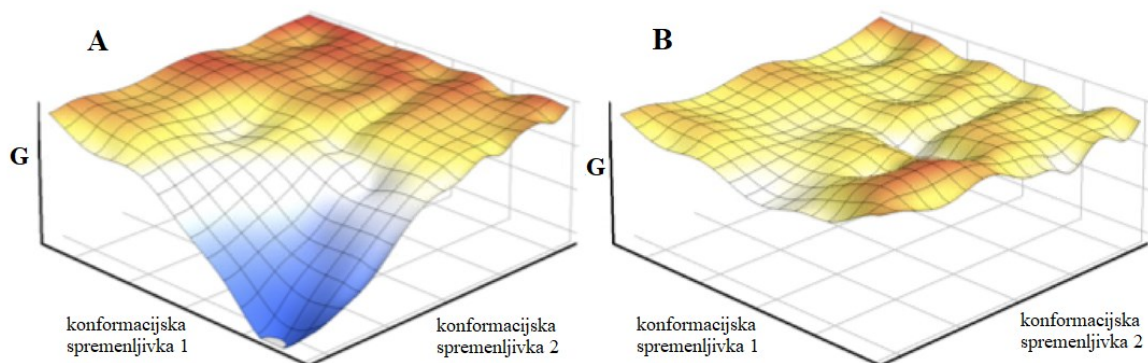
From the obtained data, we can see that AGADIR underestimates the α -helical content of peptides from IDPs, but on the other hand predicts the helical content of alanine-based peptides with a much higher degree of accuracy. The results of the sequence analysis suggest that AGADIR underscores the positive interactions between amino acid functional groups, responsible for stabilizing the α -helix.

Key words: Intrinsically disordered protein, α -helix, AGADIR

1 Uvod

1.1 Intrinzično nestrukturirani proteini

Intrinzično nestrukturirani proteini oziroma IDP-ji (*angl.* »Intrinsically Disordered Proteins«) so proteini brez enoznačno definirane 3D strukture pri fizioloških pogojih. Kljub temu, da niso zviti in da se njihova prostorska razporeditev dinamično spreminja, ne gre za denaturirane proteine, pač pa prav ta nestrukturiranost predstavlja njihovo nativno, funkcionalno stanje. S svojim obstojem nasprotujejo klasični paradigmi, ki povezuje strukturo proteina z njegovo funkcijo (*angl.* »structure-function paradigm«). Za razliko od globularnih proteinov, za IDP-je ni značilna gladka energijska površina z izrazito vdolbino (dobro definiran energijski minimum), temveč razgibana površina, brez izrazite energijsko ugodne strukture (Slika 1). To pomeni da se IDP pri določenih pogojih pojavlja kot nabor različnih konformacijskih oblik [1,2,3].

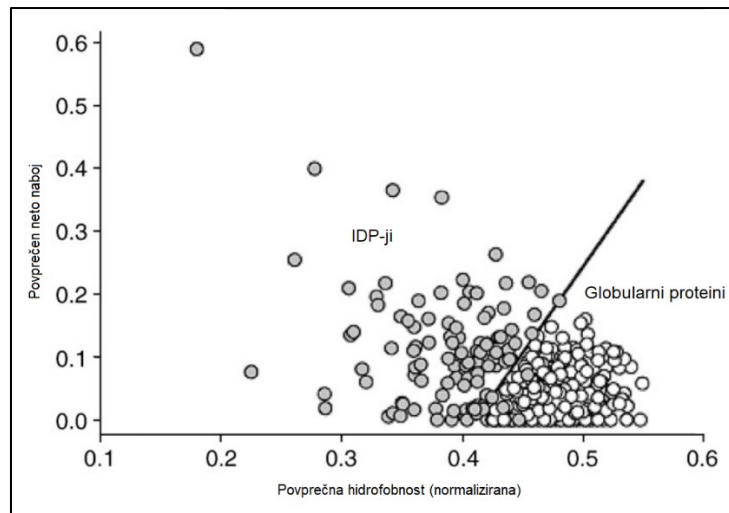


Slika 1: Prikaz proste energije (G) globularnega proteina (A), z izrazitim energijskim minimumom, in IDP-ja (B), z razgibano površino brez izrazite termodinamsko ugodne strukture pri sobni temperaturi. Prirejeno po C. K. Fisher, C. M. Stultz: "Constructing Ensembles for Intrinsically Disordered Proteins".

Zaradi svojih lastnosti, kot so na primer enostavna regulacija s post translacijskimi modifikacijami in visoka specifičnost vezave brez tvorbe močnih vezi, IDP-ji igrajo pomembno vlogo v celični signalizaciji in regulaciji. Zaradi njihove razširjenosti, vloge v celičnih procesih in vpletenosti v razvoj bolezni je zato intrinzično nestrukturirane proteine in regije – IDR (*angl.* »Intrinsically Disordered Regions«) vredno preučevati [2].

Primarne metode za karakterizacijo nestrukturiranosti proteinov in peptidov so NMR, SAXS in CD spektroskopija. Poleg eksperimentalnih tehnik se uporabljajo tudi različna bio-informacijska orodja, ki na osnovi aminokislinskega zaporedja proteina ocenijo delež nestrukturiranosti. Ena od takšnih metod za identifikacijo IDP-jev je tudi graf hidrofobnosti aminokislinskega zaporedja v odvisnosti od neto naboja (t.i. »Uversky plot«, Slika 2). Za izračun se uporablja povprečna normalizirana hidrofobnost ak-ostankov po lestvici *Kyte-Doolittle* (Tabela 1, Materiali in metode) in povprečen neto naboj ak-ostankov. Vrednosti hidrofobnosti oziroma hidrofilnosti za posamezno aminokislino po *Kyte-Doolittle* skali so določene na osnovi treh parametrov, in sicer: i) ΔG prenosa med vodo in parno fazo za

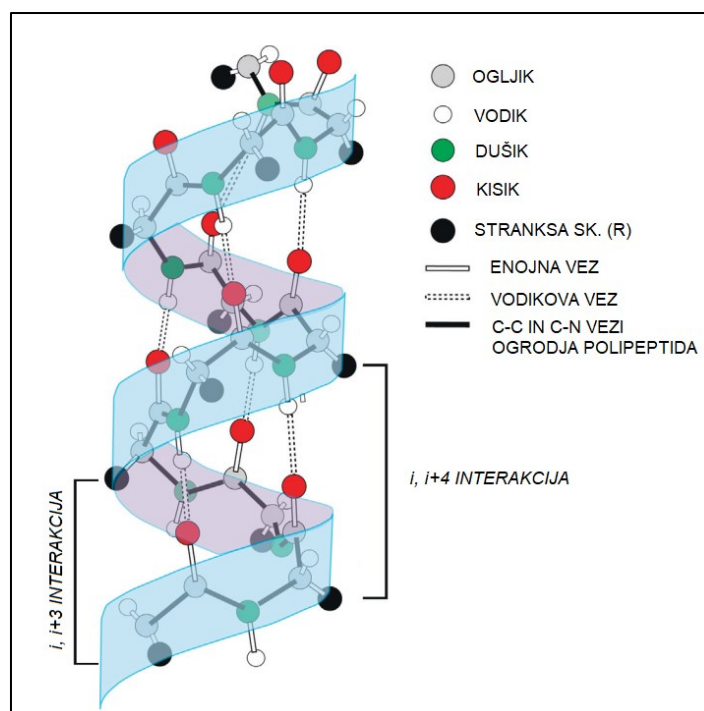
posamezno aminokislino, ii) delež določenih aminokislinskih ostankov, ki so v nativni strukturi več kot 95-odstotno zakopani in iii) delež določenih ak-ostankov, ki so v nativni strukturi 100-odstotno zakopani. Za IDP-je je značilna kombinacija nizke celokupne hidrofobnosti in nekoliko višjega neto naboja. Nasprotno je za globularne proteine značilna višja vrednost celokupne hidrofobnosti in nekoliko nižji celokupni neto naboj [2,4,5].



Slika 2: **Graf hidrofobnosti v odvisnosti od neto naboja ("Uversky plot")**. Na x-osi je nanešena povprečna normalizirana hidrofobnost ak-ostankov polipeptidne verige po *Kyte-Doolittle* skali (Tabela 1). Na y-osi so nanešeni povprečni neto naboji polipeptidne verige. Z neprekinjeno premico je označena meja med urejeno in neurejeno strukturo z enačbo $H = (R + 1,151)/2,785$ (kjer je H hidrofobnost in R neto naboj). IDP-ji se v tem grafu pojavljajo na levi strani ločnice pri nižjih vrednostih povprečne normalizirane hidrofobnosti, medtem ko so globularni, urejeni proteini na desni strani ločnice pri višjih vrednostih hidrofobnosti. Prirejeno po: V. N. Uversky and A. K. Dunker, "Understanding protein non-folding".

1.2 α -vijačnica

α -vijačnica predstavlja najbolj zastopano obliko sekundarne strukture proteinov. Struktura in pravila za nastanek α -vijačnice so dobro poznana, prav tako so dobro razvite eksperimentalne metode za določanje deleža vijačnice v peptidih (pogosto se uporablja CD spektroskopija). Pri α -vijačnici gre za linearno translacijo polipeptidne verige, ki je pravokotno združena z vrtenjem okoli osi. Glavno interakcijo, ki stabilizira zvitje vijačnice, predstavljajo vodikove vezi, ki se tvorijo med vodikom aminske skupine ene peptidne vezi in kisikom karboksilne skupine na drugi peptidni vezi, 4 aminokislinske ostanke naprej (Slika 3). Vzpostavitev take vodikove vezi je entalpijsko ugoden proces, vrednost entalpije reakcije znaša približno 2 kcal/mol. Poleg tega k stabilizaciji pripomore tudi nastanek intramolekularnih van der Waalsovih interakcij (0,5 – 1 kcal/mol) pri prehodu v strukturo α -vijačnice. Ta prispevka sta entalpijsko ugodna in morata, za nastanek α -vijačnice, preseči neugoden, entropijski prispevek zvitja konformacije polipeptida v urejeno strukturo (konformacijski prostor polipeptida se ob zvitju v vijačnico skrči).



Slika 3: **Shema α -vijačnice.** Z črtkano črto so označene vodikove vezi, ki nastajajo med amino skupino i -tega ostanka in karboksilno skupino $i+4$ ostanka peptidnega ogrodja. Prav tako je dobro vidna orientacija in pozicija stranskih skupin (črne krogle) aminokislin v prostoru kar omogoča nastanek $i, i+3$ in $i, i+4$ interakcij. Prirejeno po: J. Feher, "Protein Structure".

Za vsak zavoje α -vijačnice je v povprečju potrebnih 3,6 aminokislinskih ostankov, dolžina enega zavoja vzdolž osi je 5,4 Å. Takšna razporeditev aminokislinskih ostankov okoli osi vijačnice pomeni, da so njihove stranske skupine, razporejene v naslednjih kombinacijah: $i, i+3$, $i, i+4$ in $i, i+7$, v prostoru blizu skupaj in lahko tvorijo ugodne interakcije, ki stabilizirajo vijačnico. Stranske skupine parov aminokislinskih ostankov, ki so razporejeni v naslednjih kombinacijah: $i, i+2$, $i, i+5$ in $i, i+6$ pa se pojavljajo na nasprotnih straneh vijačnice, posledično ne prihaja do interakcij. Interakcije ki stabilizirajo nastanek vijačnice so, poleg vodikovih vezi, solni mostički (pozitivni-negativni naboj), interakcije hidrofobnih ostankov, bazičnih in aromatskih ostankov ter interakcije med polarnimi in nepolarnimi ak-ostanki. V literaturi so, poleg $i, i+3$ interakcij, najpogosteje obravnavane $i, i+4$ interakcije, ki so tudi energijsko najmočnejše. Energija omenjenih interakcij se giblje med 0,5 in 1,5 kcal/mol [5,6,7].

Urejenost strukture v α -vijačnico je povezana z aminokislinsko sestavo proteina oziroma peptida. Iz kristalnih struktur izhaja, da se določene aminokislinske pogosteje pojavljajo v strukturi α -vijačnice (npr. alanin, levcin in glutamat) kot druge (npr. prolin, glicin in aspartat). Pojav opišemo s t.i. nagnjenostjo aminokislinskih ostankov k tvorjenju vijačnice (angl. »helix propensity«). Gre za lestvico topilu izpostavljenih ak-ostankov v sredinskih položajih α -vijačnice. Najvišjo nagnjenost za tvorbo strukture vijačnice ima alanin, najnižjo pa glicin (če ne upoštevamo prolina, ki je poseben primer zaradi svoje edinstvene geometrije in pomankanja vodikovega atoma na amidni skupini). Za aminokislinske z najnižjo težnjo po tvorbi vijačnice je značilna polarna skupina, ki je od peptidnega ogrodja ločena z eno samo $-\text{CH}_2$ skupino. Lestvica težnje po tvorbi α -vijačnice: Ala > Leu > Arg > Met > Lys > Gln > Glu > Ile > Trp > Ser > Tyr > Phe > Val > His > Asn > Thr > Cys > Asp > Gly [9].

1.3 Model prehoda vijačnica-naključni klobčič

Pri opisu nastanka α -vijačnice v polipeptidu klasični sistemi dveh stanj (od popolnoma helične strukture do popolnoma nestrukturirane) odpovejo. Mehanizem zvijanja v strukturo vijačnice je namreč zapleten proces, rezultat pa je kompleksna mešanica (1) popolnoma heličnih, (2) popolnoma nestrukturiranih in predvsem (3) delno heličnih struktur za katere so značilni nestrukturirani N- in C-konci (*angl.* »frayed coil ends«). Proces nastanka vijačnice je konceptualno razdeljen na dva parametra: (i) nukleacijo – pri kateri pride do združitve treh zaporednih ostankov v stabilen helični segment in (ii) rast vijačnice – pri kateri prihaja do nastanka H-vezi. Za natančne teoretične napovedi heličnosti peptidov se v ta namen uporablja modele prehoda vijačnica-naključni klobčič, ki upoštevajo ravnotežje različnih konformacijskih stanj polipeptida, ki jih le-ta zavzame. Dva glavna tipa takšnih modelov sta (i) *Zimm-Bragg*-ova teorija, ki šteje vodikove vezi in (ii) *Lifson-Roig*-ova teorija, ki upošteva konformacijo aminokislinskih ostankov. Od obeh modelov se je bolj uveljavil model *Lifson-Roig* (LR), ki se je, ob odkritjih dodatnih elementov ki vplivajo na stabilnost α -vijačnice, tudi konstantno dopolnjeval in posodabljal [10].

1.3.1 Lifson-Roig (LR)

Konformacijska enota modela LR je aminokislinski ostanek. Vsakemu ostanku se, glede na to ali je kombinacija kotov (Φ, ψ) okoli C_{α} -atoma helična, določi konformacija vijačnice (h) ali pa konformacija neurejene strukture (c). Vsakemu ostanku pripišemo statistično utež, glede na njegovo konformacijo in konformacijo sosednjih ostankov. Ostanek, ki je v helični konformaciji (h) in ki ga na obeh straneh obdajata helična ostanka, dobi statistično utež w . Ostanki na obeh koncih vijačnice (torej **chh** in **hhc**) imajo statistično utež v , medtem ko dobijo ostanki ki ne kažejo helične strukture (c) utež 1 (c je referenčno stanje, Slika 4). Za nastanek vodikove vezi so potrebni 3 zaporedni ostanki v helični konformaciji (hhh).

C_{α}	CO NH	C_{α}	CO NH	C_{α}	CO NH	C_{α}	CO NH	C_{α}	CO NH	C_{α}	CO NH	C_{α}	LR številčenje		
1		2		3		4		5		6		7		8	LR oznake
c		c		h		h		h		c		c		c	LR utež
1		1		v		w		v		1		1		1	

Slika 4: **Lifson-Roig oznake in uteži za α -vijačnico**. S krogi so označeni osnovni elementi modela LR, to so ostanki. v je utež za iniciacijo (nastanek α -vijačnice), w pa je utež za propagacijo (podaljševanje α -vijačnice kot posledica tvorbe novih H-vezi). Z oznako h so označeni ostanki ki ležijo v ozkem območju kombinacije kotov (Φ, ψ) ki se ujema z formacijo α -vijačnice, medtem ko so z oznako c označeni preostali, ne-helični ostanki (ti ostanki hkrati služijo tudi kot referenca in so obteženi z statistično utežjo 1). Prirejeno po: A. J. Doig, "Recent advances in helix-coil theory".

Ravnotežje vijačnica-naključni klobčič po celotni verigi opišemo s statističnimi utežmi za vsako možno konformacijo ki vsebuje α -vijačnico. Populacija določene konformacije v ravnotežju je tako statistična utež konformacije ulomljena z vsoto statističnih uteži vseh možnih konformacij (t.i. statistična vsota). Večja kot je statistična utež neke konformacije, bolj je ta

stabilna. Helični homopolimer z M ostanki ima tako statistično utež v^2w^{M-2} in populacijo v ravnotežju ki je enaka v^2w^{M-2} deljeno z vsoto uteži vseh konformacij. Zaradi zasnove modela LR na osnovi ostankov se le-ta lažje uporablja tudi v primeru heteropolimerov, saj omogoča zamenjave aminokislin na določenih pozicijah pri čemer se spremenijo samo vrednosti parametrov v in w . Glavna pomanjkljivost LR modela je dejstvo, da peptidne konformacije analizira na osnovi *Flory-jeve* hipoteze izoliranih parov, ki vsako kombinacijo dihedralnih kotov (Φ, ψ) obravnava kot neodvisno od kombinacij kotov sosednjih ostankov. Zaradi tega upošteva tudi helične konformacije ki so sicer sterično onemogočene [7,8].

1.4 AGADIR

AGADIR je spletno orodje oziroma algoritem, ki napove delež α -vijačnice v strukturi peptida. Osnovan je na modelu prehoda vijačnica-naključni klobčič in se uporablja za napovedovanje vsebnosti α -vijačnice v monomerih peptidih, poleg tega izračuna tudi heličnost posameznega aminokislinskega ostanka v polipeptidnem zaporedju. Za izračun potrebujemo vhodne eksperimentalne parametre: pH, temperaturo in ionsko moč, delež vijačnice pa je podan v odstotkih [11].

Algoritem AGADIR je osnovan na modelu LR prehoda vijačnica-naključni klobčič, ki so mu z nizom popravkov dodali različne parametre ki vplivajo na stabilnost strukture α -vijačnice (Enačba 1). V nadaljevanju so opisani najpomembnejši aspekti modela ter njegove prednosti in slabosti, ki so pomembne za razumevanje njegovega delovanja ter služijo kot osnova za analizo natančnosti napovedi pri primerjavi z eksperimentalnimi podatki heličnosti. Minimalna dolžina vijačnice, ki jo AGADIR dovoljuje je 4 ostanke v h konformaciji, kar pomeni da so izključene vse vijačnice z eno samo vodikovo vezjo [10].

1.4.1 Približek, ki upošteva enkratno nukleacijo

Zaradi energijske bariere ki jo predstavlja proces nukleacije v začetnem koraku nastanka α -vijačnice, AGADIR uporabi približek, ki upošteva enkratno nukleacijo (*angl.* »Single-helical sequence approximation«). Ta predpostavlja, da so konformacijska stanja z dvema ali več heličnimi segmenti v kratkem peptidu redka in jih zato v kalkulaciji ne upošteva (odstranjena so iz statistične vsote). S povečevanjem dolžine peptida ta posplošitev ne drži več, saj so multipli helični segmenti znotraj peptida v tem primeru lahko dovolj dolgi, da presežejo nukleacijsko bariero. Dodatna omejitev je tudi dejstvo, da lahko pri daljših peptidih (30 ak-ostankov in več) pride do nastanka terciarnih interakcij med α -vijačnicami, ki jih modeli prehoda vijačnica-naključni klobčič in AGADIR ne upoštevajo [10].

1.4.2 Terminalne pozicije ter N1, N2 in N3 preference

Prvi ostanek v helični konformaciji (h) na N-koncu peptida označimo z N1, tistega pred njim, torej zadnji v ne helični konformaciji (c), pa označimo z »N-cap«. Ostanke na N-koncu α -heliksa so zato označeni na takšen način: N'-N-cap-N1-N2-N3 ... pri čemer je ostanek N-cap tisti, katerega kombinacija kotov Φ in ψ ne ustreza helični konformaciji. K nastanku vijačnice

na mestu N-cap najbolj pripomorejo aminokislinski ostanki Asn, Asp, Ser in Thr, ki delujejo kot prejemniki vodikove vezi iz NH-skupin na ostankih N2 in N3 in s tem stabilizirajo nastanek vijačnice. Razlog za to so ostanki na mestih N1, N2 in N3. Ti se od preostalih ostankov razlikujejo po tem, da njihove NH skupine niso udeležene v $i, i+4$ stabilizaciji z vodikovimi vezmi med atomi ogrodja peptidne verige. Podobna situacija je tudi na C-koncu, kjer pa C3, C2 in C1 ostanki (poimenovanje je homologno poimenovanju na N-koncu) v vodikovi vezi ne sodelujejo z karboksilnimi CO skupinami. Te se pogosto povežejo z NH skupinami ogrodja polipeptidne verige ki sledi vijačni sekvenci. V splošnem so, zaradi dipola vijačnice (pozitiven naboj obrnjen proti N-koncu, negativen pa proti C-koncu), negativno nabite skupine energijsko preferirane na N-koncu in pozitivno nabite skupine na C-koncu [5,8,11].

1.4.3 Interakcije stranskih skupin

Kot je bilo omenjeno že na začetku uvoda, se v α -vijačnici zaradi njene strukture stranske skupine α -ostankov na mestih $i, i+3$ in $i, i+4$ v prostoru razporedijo tako, da so blizu skupaj. To omogoča nastanek ugodnih interakcij med stranskimi skupinami (R), ki pripomorejo k stabilizaciji α -vijačnice. Ker za vse mogoče kombinacije interakcij 20 aminokislin ni eksperimentalnih podatkov o vrednosti proste energije, algoritem AGADIR uporablja bazo podatkov, ki je bila razvita s pomočjo statističnih metod iz nekega manjšega začetnega seta eksperimentalnih podatkov. Iz tega razloga je veliko vrednosti interakcij stranskih skupin samo približek prave vrednosti, nekatere pa so popolnoma izvzete iz računanja. To še posebej velja za Pro, His, Cys in Trp [8,11].

$$\Delta G_{Hel} = \Delta G_{Int} + \Delta G_{Hbond} + \Delta G_{SD} + \Delta G_{dipol} + \Delta G_{nonH} + \Delta G_{electrostat} \quad (1)$$

Enačba 1 : Izračun spremembe proste energije za prehod polipeptidne verige iz naključnega zvitja (klobčiča) v α -vijačnico v algoritmu AGADIR. Vključeni so naslednji parametri: ΔG_{Int} – intrinzična težnja aminokislinskih ostankov zaporedja po tvorjenju heliksa; ΔG_{Hbond} – vsota neto prispevka vseh vodikovih vezi med atomi ogrodja polipeptida znotraj območja heliksa; ΔG_{SD} – vsota neto prispevka vseh interakcij med stranskimi skupinami na položajih $i, i+3$ in $i, i+4$, upoštevane so tudi šibke privlačne in odbojne sile med nabitimi ostanki v heliksu na položaju $i, i+1$; ΔG_{dipol} – vsota prispevkov interakcij dipola vijačnice z nabitimi skupinami znotraj in zunaj vijačnice; ΔG_{nonH} – vsota prispevkov k stabilnosti vijačnice ostankov, ki niso v helični konformaciji ($\gg N\text{-cap}$ in $\gg C\text{-cap}$); $\Delta G_{electrostat}$ – vsota prispevkov interakcij med nabitimi skupinami (vključno z interakcijami nabitih stranskih skupin) [10].

2 Namen dela

Namen naloge je preveriti napovedi vsebnosti α -vijačnice v peptidih intrinzično nestrukturiranih proteinov s tendenco po tvorbi vijačnice s pomočjo algoritma AGADIR. Napovedan delež α -vijačne strukture bomo primerjali z eksperimentalnimi podatki iz literature in laboratorijskih meritev. To bi nas vodilo do boljšega razumevanja napovedi heličnosti strukture peptidov. V literaturi smo poiskali eksperimentalne vrednosti 40ih peptidov (CD spektri) in iz njih izračunali eksperimentalne vrednosti vsebnosti α -vijačnice. Set peptidov smo analizirali z algoritmom AGADIR in določili enačbo korelacije med napovedano vsebnostjo α -vijačnice in eksperimentalno določenimi vrednostmi.

3 Materiali in metode

3.1 Zbiranje podatkov

CD spektre peptidov, ki smo jih uporabili za analizo v tem delu, smo pridobili iz literature, poleg tega pa smo uporabili tudi eksperimentalne podatke iz laboratorija.

3.1.1 Eksperimentalni podatki iz literature

Z uporabo spletne baze podatkov »DisProt« smo poiskali raziskovalne članke v katerih je bil določen peptid okarakteriziran kot IDP na osnovi CD spektroskopije. Osredotočili smo se na peptide dolžine od 13 do 60 aminokislinskih ostankov. Iz članka smo pridobili CD spekter peptida ter prepisali eksperimentalne pogoje pri katerih je bil le-ta posnet (temperatura, ionska moč in pH). Vrednost intenzitete signala pri 222 nm smo ocenili iz slike CD spektra s spletnim orodjem »WebPlotDigitizer« [14].

3.1.2 Eksperimentalni podatki pridobljeni v laboratoriju

CD spektri sintetičnih peptidov so bili posneti v laboratoriju katedre za fizikalno kemijo na FKKT. Za izris grafov smo uporabili spletno orodje »Capito« [15], ki na osnovi surovih podatkov nariše krivuljo in jo z uporabo algoritma *Savitzky-Golay* pogladi.

3.2 Določanje deleža α -vijačnice

Delež α -vijačnice v posameznem peptidu smo določali na osnovi vrednosti CD spektra pri 222 nm. Vse surove podatke smo pretvorili v standardne enote – MRE (*angl.* »Mean Residue molar Ellipticity«) [$\text{deg}/\text{cm}^2/\text{dmol}^{-1}$]. Za izračun vsebnosti α -vijačnice iz podatkov pridobljenih s pomočjo CD spektroskopije smo uporabili naslednje enačbe:

$$f_H = ([\theta]_{\text{OBS}} - [\theta]_{\text{C}})/([\theta]_{\text{H}} - [\theta]_{\text{C}});$$

$$[\theta]_{\text{C}} = 2220 - 53T;$$

$$[\theta]_{\text{H}} = (-44000 + 250T) (1-3/N_r);$$

kjer je f_H delež α -vijačnice (od 0 do 1), $[\theta]_{\text{OBS}}$ je odčitani signal pri 222 nm pretvorjen v »MRE«, $[\theta]_{\text{C}}$ je teoretični signal ki bi ga imel peptid s 100-odstotno naključno strukturo (*angl.* »random coil«), $[\theta]_{\text{H}}$ je teoretični signal ki bi ga imel peptid s 100-odstotno α -vijačno strukturo, T je eksperimentalna temperatura v $^{\circ}\text{C}$, N_r pa je število aminokislinskih ostankov v peptidu [12].

3.3 Napoved deleža α -vijačnice – AGADIR

Napovedi vsebnosti α -vijačnice smo opravili s pomočjo algoritma AGADIR. Za analizo je potrebno vnesti aminokislinsko zaporedje in 3 eksperimentalne parametre (pH, temperaturo in ionsko moč). Program nato s pomočjo teh podatkov izračuna delež α -vijačnice in ga poda v odstotkih. Dovoljena je uporaba le standardnih aminokislin, program pa dovoljuje tudi dve modifikaciji na N-koncu (acetilacija ali sukcinilacija) in eno na C-koncu (amidacija).

Calculation

Enter your sequence(s) :

WASLFDTLDSTNKELVNR

WTLDSTNKELVNR

YGSEADEMAKALEAELN

Select some parameters:

Nterm protection

Cterm protection

Parameter screening

Parameter	Value	Allowed Range
pH	<input type="text" value="7"/>	[1.00 - 14.00]
Temperature	<input type="text" value="278"/>	[273K - 400K]
Ionic Strength	<input type="text" value="0.1"/>	[0.0 - 1.0]

Check box to output the prediction at the residue level

Slika 5: Primer vnosa sekvence (levo) in eksperimentalnih parametrov (desno) v spletno orodje AGADIR.

Peptide 3		YGSEADEMAKALEAELN								
res,	aa,	Hel,	Ncap,	Ccap,	Hstaple,	Schellm,	CaH,	13Ca,	JaN	
-0.087392										
01,	Y,	0.0,	0.08,	0.00,	0.00,	0.00,	-0.15,	0.00,	7.19	
02,	G,	0.1,	1.65,	0.00,	0.00,	0.00,	-0.06,	0.00,	5.57	
03,	S,	1.7,	6.71,	0.00,	0.00,	0.00,	-0.04,	-0.09,	6.70	
04,	E,	8.5,	0.79,	0.00,	0.00,	0.00,	-0.03,	0.20,	6.02	
05,	A,	9.2,	0.74,	0.00,	0.00,	0.00,	-0.02,	0.18,	5.64	
06,	D,	10.0,	9.31,	0.00,	0.00,	0.00,	-0.01,	0.04,	6.60	
07,	E,	19.2,	2.24,	0.05,	0.00,	0.00,	-0.06,	0.46,	5.99	
08,	M,	21.1,	0.46,	0.32,	0.00,	0.00,	-0.07,	0.41,	6.23	
09,	A,	21.5,	0.19,	0.12,	0.00,	0.00,	-0.04,	0.41,	5.54	
10,	K,	21.0,	0.18,	0.64,	0.00,	0.00,	-0.07,	0.50,	6.19	
11,	A,	19.7,	0.08,	1.56,	0.00,	0.00,	-0.04,	0.40,	5.66	
12,	L,	18.0,	0.04,	1.74,	0.00,	0.00,	-0.07,	0.36,	6.49	
13,	E,	14.9,	0.00,	3.10,	0.00,	0.00,	-0.05,	0.43,	6.17	
14,	A,	12.0,	0.00,	2.95,	0.00,	0.00,	-0.02,	0.28,	5.89	
15,	E,	7.3,	0.00,	4.64,	0.00,	0.00,	-0.02,	0.29,	6.33	
16,	L,	5.3,	0.00,	2.07,	0.00,	0.00,	0.02,	0.13,	6.79	
17,	N,	0.0,	0.00,	5.26,	0.00,	0.00,	-0.15,	0.11,	7.41	
Percentage helix			12.63							

Slika 6: Primer izhodnih podatkov spletnega orodja AGADIR. Orodje za vsak aminokislinski ostanek posebej izračuna verjetnost za nastanek α -vijačnice in nato poda oceno heličnosti peptida v odstotkih (*angl.* »Percentage helix«, označeno z rdečo).

3.4 Določanje povezanosti med eksperimentalnimi podatki in izračuni s pomočjo algoritma AGADIR

Eksperimentalne vrednosti iz strokovne literature in laboratorijskih meritev smo vnesli v tabelo v Excel datoteki in jim dodali vrednosti ki smo jih izračunali z orodjem AGADIR. Na osnovi pridobljenih podatkov smo izrisali graf izračunanih vrednosti po AGADIR-ju v odvisnosti od eksperimentalno določenih vsebnosti α -vijačnice v peptidih.

3.5 Analiza peptidov

Za analizo aminokislinskih zaporedij peptidov smo uporabili programsko opremo »Anaconda« [16], ki deluje na osnovi Python-a. Glede na njihove lastnosti smo posamezne aminokislino razdelili v podkupnine (nabite, hidrofobne, polarne, aromatske itd.) in na osnovi te razdelitve opravili analizo. V prvem delu smo se osredotočili na zastopanost posameznih skupin aminokislinskih v peptidih, v drugem delu pa smo opravili analizo na podlagi $i, i+3$ ter $i, i+4$ interakcij med stranskimi skupinami aminokislinskih ostankov v α -vijačnici [12]. Poleg tega smo izvedli tudi analizo sekvenc glede na njihov neto naboj in hidrofobnost ter izrisali »Uversky plot« [4]. Neto naboj posameznega peptida je izračunan pri $\text{pH} = 7$, glede na vsebnost nabitih stranskih skupin. Hidrofobnost aminokislinskih ostankov smo izračunali na podlagi normalizirane »Kyte-Doolittle« lestvice (Tabela 1), kjer vrednost 1 zavzame najbolj hidrofobna aminokislina – izolevcin, vrednost 0 pa najbolj hidrofilna aminokislina – arginin.

Tabela 1: Kyte-Doolittle lestvica hidrofobnosti aminokislinskih ostankov. V skrajnem desnem stolpcu so prikazane normalizirane vrednosti.

Aminokislina	Hidrofobnost po Kyte-Doolittle	Normalizirana HIDROFOBNOST
izolevcin	4,5	1,00
valin	4,2	0,97
levcin	3,8	0,92
fenilalanin	2,8	0,81
cistein	2,5	0,78
metionin	1,9	0,71
alanin	1,8	0,70
glicin	-0,4	0,46
treonin	-0,7	0,42
serin	-0,8	0,41
triptofan	-0,9	0,40
tirozin	-1,3	0,36
prolin	-1,6	0,32
histidin	-3,2	0,14
glutamat	-3,5	0,11
glutamin	-3,5	0,11
asparta	-3,5	0,11
asparagin	-3,5	0,11
lizin	-3,9	0,07
arginin	-4,5	0,00

4 Rezultati

4.1 Baza podatkov

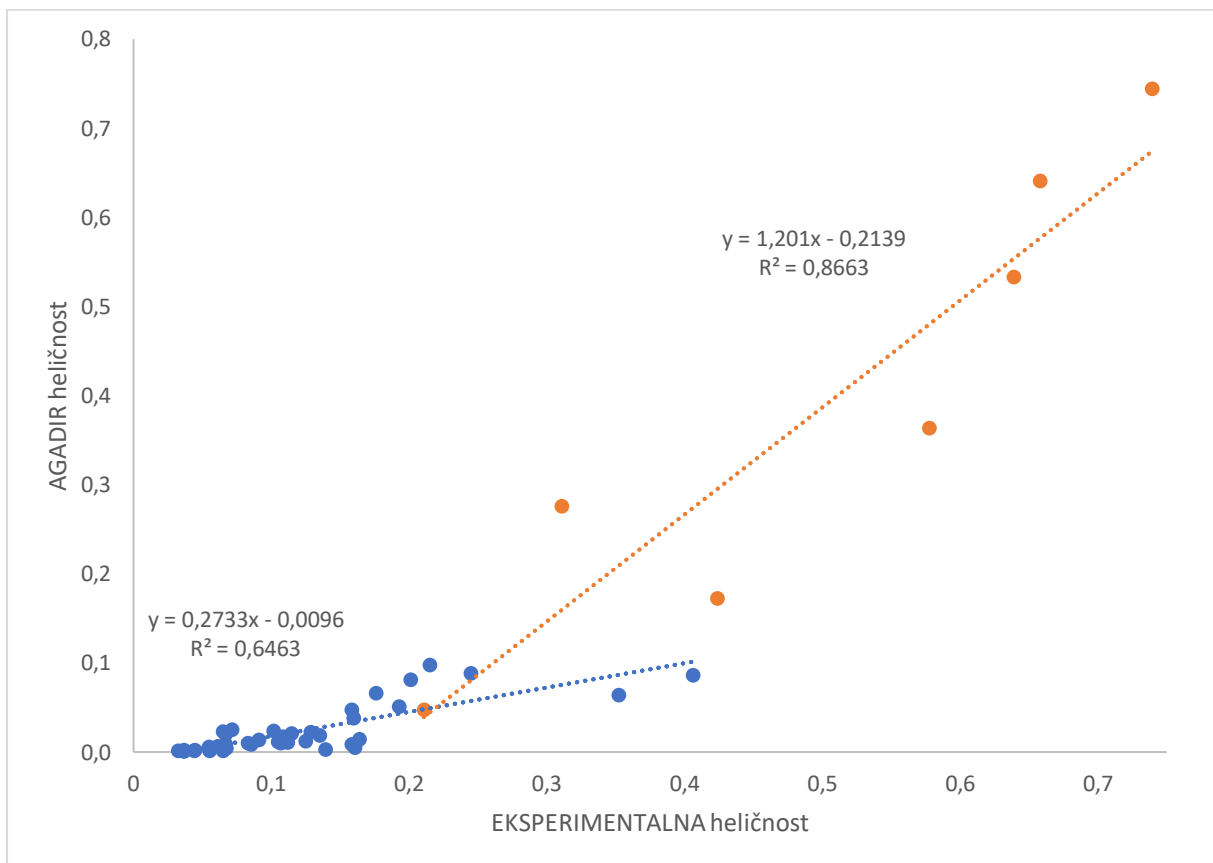
Za analizo natančnosti napovedi algoritma AGADIR smo uporabili eksperimentalne podatke heličnosti 47 peptidov (vključno z sedmimi alaninskimi peptidi, ki služijo kot referenca). Povprečna dolžina peptidov je 29,17 ak-ostankov. Razpon dolžin je od 13 do 50 ak-ostankov. Eksperimentalne heličnosti pridobljene iz literature in laboratorijskih meritev so v razponu od 3,2% do 40,6 % (ob upoštevanju referenčnih alaninskih peptidov je zgornja vrednost 73,9%). Vrednosti izračunane s pomočjo algoritma AGADIR se gibljejo v območju od 0,7% do 9,8% (ob upoštevanju alaninskih peptidov do 74,5%). Povprečna eksperimentalna vrednost vsebnosti heliksa za set 40 peptidov je 12,3%, povprečje izračunanih vrednost po AGADIR-ju pa 2,4%. V spodnji tabeli (Tabela 2) so navedeni eksperimentalni podatki za celoten set peptidov.

Tabela 2: Seznam peptidov z navedenimi eksperimentalnimi vrednostmi, eksperimentalnimi pogoji in izračunanimi vrednostmi po AGADIR-ju.

IDP	sekvenca	eksp. helicnost	eksp. pogoji (T/pH/I)	AGADIR helicnost
C peptid	KETAAAKFERQHS	0,201	1.7/6.88/0.2	0,081
S peptid	KETAAAKFERQHMDSSSTA	0,176	1.0/4.03/0.2	0,066
SPR2 (HD)	SYQQQCKQPCQPPVPCPT	0,037	20.0/7.5/0.2	0,001
AP7-1	DDNGNYGNGMASVRTQGNTYDDLASLISYL	0,161	25/7.4/0.1	0,005
AP24-1	ADDEDEDASSGLCNQYNQNVTTTRPNKPKMF	0,139	25/7.4/0.2	0,003
pKID	DSVTDSQKRREILSRPSYRKILNDLSSDAP	0,193	25/7.5/0.2	0,051
PUMA	RVEEEEWAREIGAQLRRAADDLNAQYERRRQEEQH	0,215	5/7.0/0.3	0,098
Chromogranin A	WELESLSAIEAELEKVAHQALRRG	0,245	22/7.5/0.2	0,089
TatH2 (44-89)	SDDEPKQDKTSQDADFTAKTIADKQADTNQEQAKTEDAKRHDKEQV	0,065	25/7.4/0.2	0,023
P1 (pilosulin)	GLGSVFGRLARILGRVIPKV-NH2	0,115	20/7/0.1	0,021
P2 (pilosulin)	GLLSKFGRLARILGRVIPKV-NH2	0,125	20/7/0.1	0,012
Rev ARM	GAMATRQARRNRRRWRERQRAAAA	0,159	25/7.4/0.2	0,048
L B1 (95-115)	VITKANLIFANGSTQTAEFKG	0,085	5/4.1/0.2	0,008
L B1 (114-138)	KGTFEKATSEAYAYADTLKKDNGEY	0,083	5/2.4/0.2	0,010
L B1 (136-155)	GEYTVDVADKGYTLNLIKAFGD	0,055	22/6.1/0.2	0,006
SN peptid	IMEKADSNKTRIDEANQRATKMLGSG	0,129	25/7.4/0.1	0,023
SB peptid	DEVVDMRVNVKVLERDQKLSLDDRA	0,109	25/7.4/0.2	0,017
ST peptid	ALSEIETRHSEIHKLENSIRELHDMFMD	0,067	25/7.4/0.3	0,009
T peptid (HMGB1)	KEEEDDEEDEDDEEEEEDEEEDDDDD	0,044	25/7/0.1	0,001
β-Cat (17-48)	DRKAAVSHWQQSYLDSGHSGATTTAPSLSG	0,055	20/6.2/0.1	0,002
Nogo-33	YERHQAQIDHYLGLANKNKVDAMAKIQAKIPGL	0,091	20/6.8/0.1	0,014
Peptid A	TSAQKAKAEERKRKMSRGL	0,406	5/4.8/0.1	0,086
LL-37	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES	0,067	22/6/0.1	0,020
mAriN37	MGRRLVTVRIQRAGRPLQERVLVLFVRSRRPRTAS	0,061	25/7/0.1	0,006
CD4(CYTO)	RCRHRRRQAERMSQIKRLSEKTCQCPHRFQKTCSPI	0,071	25/7.4/0.1	0,025
KorB ΔA297	PDPDKLKKAIQVEHDERPARLILNRRPPAEGYAWLKYEDDGQEFANLADVKLVALIEG	0,102	25/7/0.1	0,024
Thymosin beta-4	MSDKPDMAEIEKFDKSKLKTETQENPLPSKETIEQEQAGES	0,158	20/5.5/0.1	0,009
ID (MDBP 2)	PSRREQPPKPKSPKAPGTGRGRGPKGSGTTRPKAATSEGVQVK	0,065	20/7/0.1	0,001
LigBCen2NR	NKTDTIAPTQSVVSLPTTIQVVSYSEINKEALDLSNYKIINSSN	0,111	20/7.4/0.1	0,016
Myosin IIC tpA	LTFITRTVRQVFRLEEGVASDEEEAEGAEPGAPGQPEAPPATPQ	0,037	25/7.2/0.12	0,002
MSR C17A peptid	MAQWDDFPDQEDTDSCTESVKFDARSVTALLPHPKNGPTLQERMKSYKGC	0,107	25/7/0.1	0,010
HigA2-P1	NRDLFAELSSALVEAKQHSEGW	0,112	25/7.4/0.17	0,011
ccda37-62	RRLRAERWKAENQEGMAEVARFIEMN	0,135	25/7.4/0.17	0,019
phd4 w56-73	WASLFDTLDSTNKLNVNR	0,105	25/7.4/0.17	0,012
phd5 W61-73	WTLDSNKLNVNR	0,164	25/7.4/0.17	0,015
RelA_NLS (293-316)	DDRHRIEEKRRKTYETFKSIMKKS	0,160	25/7.4/0.17	0,038
FCP1 (YG(944-958))	YGSEADEMAKALEALN	0,352	25/7.4/0.17	0,064
2G	NRDLFGELSSALGEAKQHSEGW	0,068	25/7.4/0.17	0,004
4G	NRDLFGELSGALGEAKQHSEGW	0,044	25/7.4/0.17	0,002
6G	NRDLFGELGGALGEAKQHSEGW	0,032	25/7.4/0.17	0,002
AK17	Ac-(AAKAA)(AAKAA)(AAKAA)GY-NH2	0,311	25/7.4/0.17	0,276
Scholtz 14	YAEAAKAAEAAKAF	0,211	25/7/0.1	0,048
Scholtz 20	YAEAAKAAEAAKAAEAAKAF	0,424	25/7/0.1	0,173
Scholtz 26	YAEAAKAAEAAKAAEAAKAAEAAKAF	0,578	25/7/0.1	0,364
Scholtz 32	YAEAAKAAEAAKAAEAAKAAEAAKAAEAAKAF	0,639	25/7/0.1	0,533
Scholtz 38	YAEAAKAAEAAKAAEAAKAAEAAKAAEAAKAAEAAKAF	0,658	25/7/0.1	0,641
Scholtz 50	YAEAAKAAEAAKAAEAAKAAEAAKAAEAAKAAEAAKAAEAAKAF	0,739	25/7/0.1	0,745

4.2 Primerjava eksperimentalne vrednosti z izračunano

Za primerjavo napovedi algoritma AGADIR smo pripravili graf izračunanih vrednosti v odvisnosti od eksperimentalno določenih vsebnosti vijačnice v peptidih. Kot referenco smo uporabili tudi set sedmih alaninskih peptidov (na osnovi takšnih peptidov je bil algoritem namreč umerjen). Pri alaninskih peptidih se tako izračunana vrednost po AGADIR-ju ne bi smela bistveno razlikovati od eksperimentalno določene.



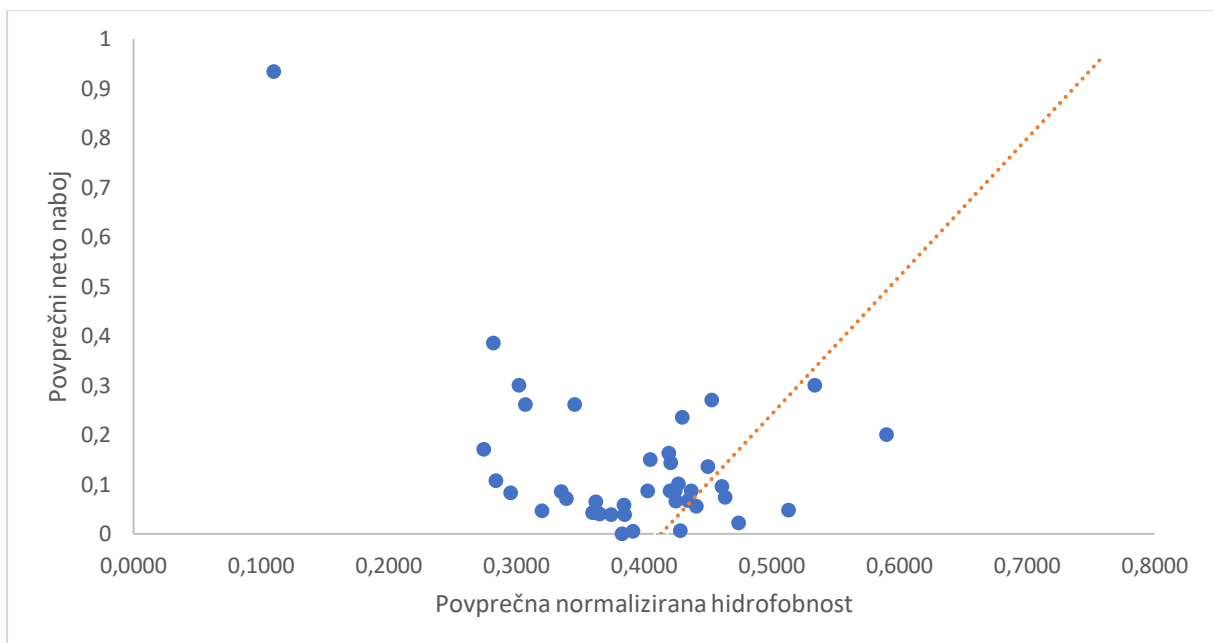
Slika 7: Graf eksperimentalnih vrednosti heličnosti v odvisnosti od izračunane vrednosti po AGADIR-ju. Z modro barvo je označen set 40ih peptidov iz literature in laboratorijskih meritev. Z oranžno barvo so označeni alaninski peptidi. Na grafu sta zapisani tudi vrednosti enačb premic za posamezen niz podatkov.

Na podlagi grafa (Slika 7) je jasno vidno, da algoritem AGADIR pri svoji napovedi vsebnosti α -vijačnice v peptidih, ki izhajajo iz IDP-jev, le-to močno podceni. Vrednost naklona premice 1 bi kazala na popolno ujemanje napovedi z eksperimentalnimi podatki, vrednost 0,27 pa kaže na skoraj štirikratno podcenenje vrednosti parametra s strani algoritma. Če primerjamo povprečno eksperimentalno vrednost z povprečno vrednost izračunano po AGADIR-ju je to razmerje enako 0,20 (12,3/2,4). Za primerjavo natančnosti algoritma AGADIR smo izračunali tudi vrednosti za sedem alaninskih peptidov, ki jih uporabljamo kot referenco (oranžen niz na Sliki 7). Tudi v tem primeru bi idealno napoved predstavljal naklon premice, ki bi bil enak 1,

je pa iz naklona premice in same razpršenosti podatkov vidno, da je v primeru alaninskih peptidov AGADIR pri napovedi veliko bolj natančen.

4.3 Analiza sekvenc

Pri analizi sekvenc preučevanih peptidov smo se osredotočili na aminokislinsko zastopanost v posameznih sekvencah. Eden od načinov klasificiranja IDP-jev je tudi t.i. »Uversky Plot«, na katerem je narisana povprečna normalizirana hidrofobnost peptida v odvisnosti od povprečnega neto naboja. Za IDP-je je značilna nizka povprečna vrednost hidrofobnosti in nekoliko višji neto naboj kot za globularne proteine.

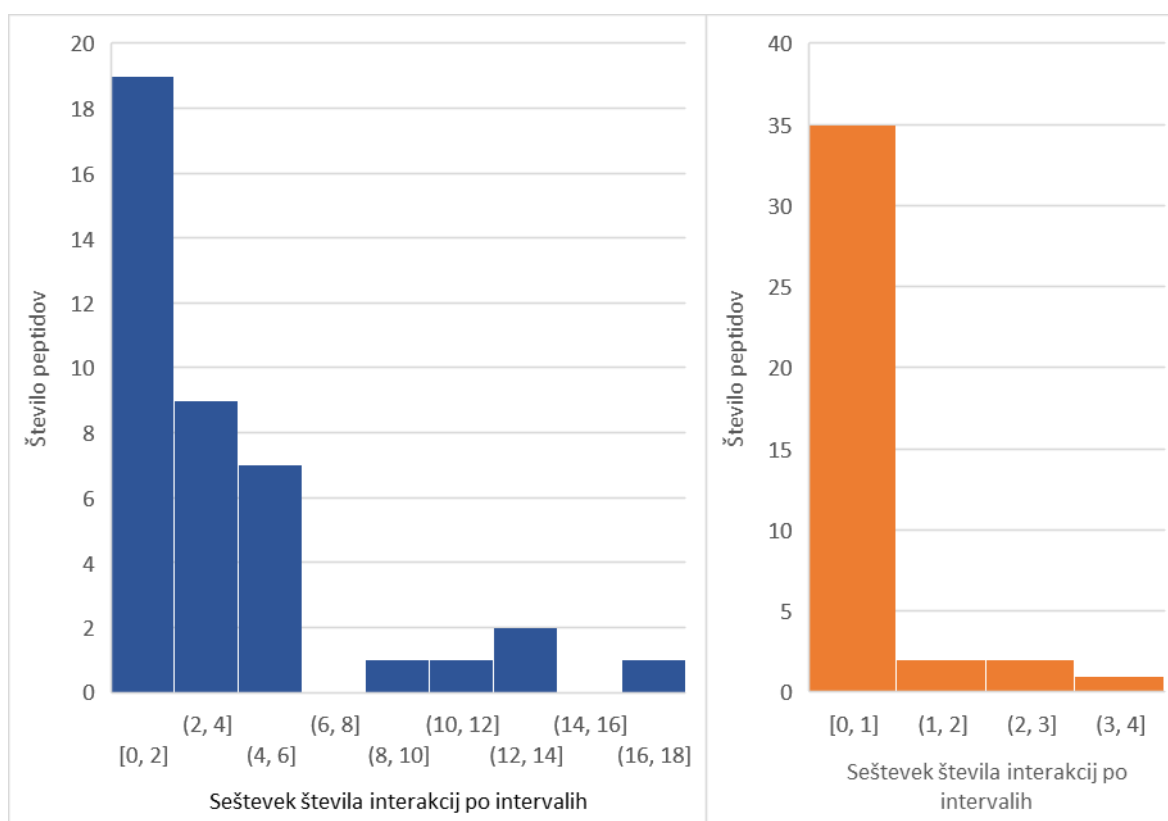


Slika 8: Graf povprečne normalizirane hidrofobnosti po lestvici Kyte-Doolittle v odvisnosti od povprečnega neto naboja peptidov (»Uversky Plot«). Z oranžno, črtkano črto je narisana teoretično določena ločnica med IDP-ji na eni in globularnimi proteini na drugi strani.

Večina preučevanih peptidov (Slika 10) se nahaja na levi strani ločnice v predelu, kjer, zaradi njihovih lastnosti, po navadi najdemo IDP-je. Nekaj peptidov je sicer vidnih tudi na nasprotni strani, kjer pričakujemo globularne peptide z urejeno strukturo. Za IDP-je, ki se zvijejo ob vezavi na tarčo, je značilno, da se nahajajo zelo blizu ločnice, saj izkazujejo dvojni značaj. Gre za neke vrste pogojno zvitje - v osnovnem stanju so nezviti, ob pristnosti tarče pa preidejo v zvito konformacijo.

4.3.1 Interakcije stranskih skupin

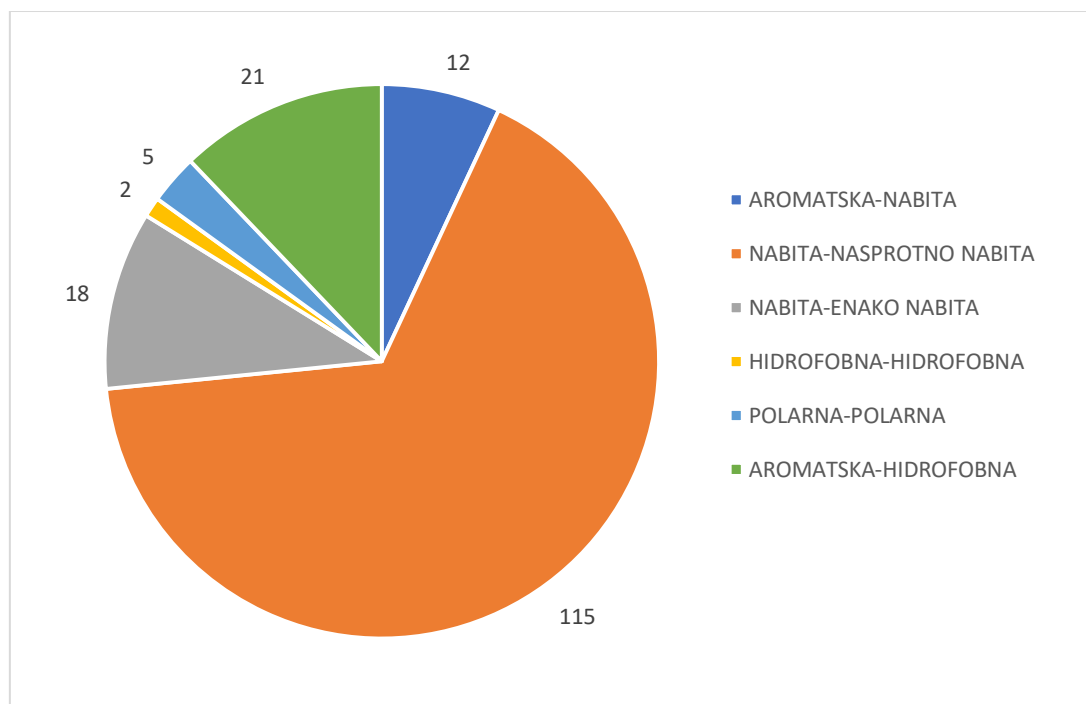
Poleg analize hidrofobnosti in neto naboja smo opravili tudi analizo $i, i+3$ in $i, i+4$ interakcij med posameznimi stranskimi skupinami v peptidih. Osnova za analizo so eksperimentalno določene interakcije iz literature [5,15]. Vse interakcije so ugodne, z izjemo interakcije Lys-Lys, ki je odbojna (tako $i, i+3$, kot tudi $i, i+4$). Povprečno število ugodnih interakcij za posamezen peptid je 4, pri tem je povprečno število interakcij na aminokislinski ostanek 0,14. Povprečno število neugodnih interakcij pa je 0,45.



Slika 9: Prikaz razporeditve števila ugodnih (levo) in neugodnih (desno) $i, i+3$ in $i, i+4$ interakcij med stranskimi skupinami v setu preučevanih peptidov.

Iz histograma na Sliki 9 je razvidno, da se v setu peptidov najpogosteje pojavlja manjše število interakcij med stranskimi skupinami (od 1 do 6, prevladujejo predvsem peptidi z eno ali dvema interakcijama), medtem ko je primerov z večjim številom interakcij (do 18) razmeroma malo. Pri tem je potrebno upoštevati tudi dejstvo, da se večje število interakcij pojavlja v peptidih z daljšim zaporedjem (vsaj 35-40 ak-ostankov in več).

Na Sliki 10 je prikazan delež interakcij glede na lastnost para aminokislinskih ostankov, ki tvorita interakcijo. Daleč največji delež (66%) predstavljajo interakcije med nasprotno nabitimi stranskimi skupinami aminokislinskih ostankov. Preostalo tretjino predstavljajo interakcije aromatskih ak-ostankov z nabitimi ter hidrofobnimi ak-ostanki, interakcije enako nabitih ostankov, hidrofobne in polarne interakcije.



Slika 10: Tortni diagram razporeditve i , $i+3$ in $i+4$ interakcij po posameznih kategorijah parov aminokisljin.

5 Razprava

Rezultati primerjave napovedi heličnosti z eksperimentalnimi podatki jasno kažejo, da algoritem AGADIR močno podceni vsebnost α -vijačnice v preučevanih peptidih intrinzično nestrukturiranih proteinov. Po drugi strani pa z veliko večjo natančnostjo napove vsebnost α -vijačnice v alaninskih peptidih. Rezultat ni presenetljiv, saj podatki iz alaninskih peptidov in peptidov izpeljanih iz alaninskih peptidov predstavljajo večino podatkovne baze, na podlagi katere so ocenjeni parametri modela AGADIR.

Možnih razlogov, zakaj pridemo do takšnega rezultata, je mnogo in težko z gotovostjo pokažemo na pravega oziroma na tistega ki je najbolj pomemben. Kljub temu pa lahko pridemo do določenih zaključkov na podlagi analize delovanja samega algoritma. Eden od možnih vzrokov, zakaj pride do podcenitve deleža α -vijačnice, je lahko povezan z napačnimi vrednostmi energijskih prispevkov interakcij stranskih skupin. Glede na to, da je delež α -vijačnice podcenjen v primerjavi z eksperimentalno določenim deležem, je možno da: (1) so ugodni energijski prispevki interakcij podcenjeni ali (2) da so energijski prispevki neugodnih interakcij precenjeni. Algoritem prav tako ne vsebuje podatkov za interakcije prolina in cisteina z drugimi ak-ostanki zaradi premajhnega števila eksperimentalnih podatkov ki so na voljo.

Iz analize sekvenc izhaja, da so najpogostejše ugodne interakcije, interakcije med nasprotno nabitimi stranskimi skupinami. Energijski prispevek $i, i+3$ interakcije stranskih skupin glutamata in lizina (ExxK) na primer, je v AGADIR-ju ovrednoten z vrednostjo -0,15 kcal/mol [13], medtem ko je vrednost za to isto interakcijo v Schmidler et al. [8] navedena kot -0,9 kcal/mol. Podobno je tudi v primeru $i, i+4$ interakcije med argininom in glutamatom (RxxxE), kjer je AGADIR-jeva vrednost -0,5 kcal/mol, vrednost po Schmidler et al. pa -0,85 kcal/mol. Za interakcijo lizina z glutamatom (KxxxE) je vrednost po AGADIR-ju celo pozitivna: +0,1 kcal/mol, medtem ko je vrednost interakcije po Schmidler et al. -0,1 kcal/mol. Tudi v primeru interakcij, ki ne spadajo v kategorijo nasprotno nabitih stranskih skupin najdemo podobne primere. Eden takšnih je $i, i+3$ interakcija stranskih skupin aromatskih aminokislinskih ostankov fenilalanina in triptofana (FxxW) za katero je AGADIR-jeva vrednost -0,2 kcal/mol, vrednost po Schmidler et al. pa -0,9 kcal/mol. Podobno je tudi v primeru tirozina in fenilalanina (YxxF) – AGADIR: -0,2 kcal/mol, Schmidler: -1,3 kcal/mol. Vsi zgoraj naštet primeri kažejo na to, da AGADIR najverjetneje podceni ugodne energijske prispevke določenih interakcij.

Iz opisa na spletni strani programa AGADIR [11] je razvidno tudi, da poleg podatkov za Pro in Cys manjkajo nekateri podatki za interakcije histidina (His) in triptofana (Trp). Vendar pa to dejstvo ne bi smelo imeti vpliva na dobljene rezultate saj je, sploh v primeru triptofana, njegova zastopanost v setu peptidov zelo majhna (po en ostanek se namreč pojavi v le 1/3 peptidov, primerov z dvema ostankoma ali več pa sploh ni). V primeru histidina se le-ta pojavi v približno polovici peptidov, največje število ostankov pa je 2, kar pa bi že lahko kazalo na enega od vzrokov za podcenitev deleža vijačne strukture.

6 Sklepne ugotovitve

AGADIR pri napovedi heličnosti peptidov intrinzično nestrukturiranih proteinov močno podceni delež α -vijačnice. Razlog za takšno oceno je najverjetneje premajhna teža energijskega prispevka ugodnih interakcij stranskih skupin. Po drugi strani pa, po pričakovanjih, z veliko večjo mero natančnosti napove vsebnost α -vijačnice v setu alaninskih peptidov.

Možnost izboljšave algoritma zato obstaja predvsem na področju razširitve knjižnice interakcij, da bi le-ta vsebovala eksperimentalno določene vrednosti proste energije prav za vse možne kombinacije dvajsetih ak-ostankov. Poleg i , $i+3$ in $i+4$ bi veljalo v izračune vnesti tudi bolj daljnosežne interakcije (npr. i , $i+7$).

7 Literatura

- [1] V. N. Uversky and A. K. Dunker, “Understanding protein non-folding,” *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1804, no. 6. pp. 1231–1264, Jun-2010.
- [2] M. Guharoy, K. Pauwels, and P. Tompa, “SnapShot: Intrinsic Structural Disorder.,” *Cell*, vol. 161, no. 5, pp. 1230-1230.e1, May 2015.
- [3] C. K. Fisher and C. M. Stultz, “Constructing ensembles for intrinsically disordered proteins,” *Current Opinion in Structural Biology*, vol. 21, no. 3. pp. 426–431, Jun-2011.
- [4] V. N. Uversky, J. R. Gillespie, and A. L. Fink, “Why are ‘natively unfolded’ proteins unstructured under physiologic conditions?,” *Proteins Struct. Funct. Genet.*, vol. 41, no. 3, pp. 415–427, Nov. 2000.
- [5] J. Kyte and R. F. Doolittle, “A simple method for displaying the hydropathic character of a protein,” *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, May 1982.
- [6] H. Qian and J. A. Schellman, “Helix-coil theories: A comparative study for finite length polypeptides,” *J. Phys. Chem.*, vol. 96, no. 10, pp. 3987–3994, 1992.
- [7] J. Feher, “Protein Structure,” in *Quantitative Human Physiology*, Elsevier, 2017, pp. 130–141.
- [8] S. C. Schmidler, J. E. Lucas, and T. G. Oas, “Statistical estimation of statistical mechanical models: helix-coil theory and peptide helicity prediction.,” *J. Comput. Biol.*, vol. 14, no. 10, pp. 1287–310, Dec. 2007.
- [9] C. N. Pace and J. M. Scholtz, “A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins,” 1998.
- [10] A. J. Doig, “Recent advances in helix-coil theory,” *Biophys. Chem.*, vol. 101–102, pp. 281–293, Dec. 2002.
- [11] [Http://agadir.crg.es/](http://agadir.crg.es/), “(16.8.2019).”
- [12] N. Errington, T. Iqbalsyah, and A. J. Doig, “Structure and stability of the alpha-helix: lessons for design.,” *Methods Mol. Biol.*, vol. 340, pp. 3–26, 2006.
- [13] V. Muñoz and L. Serrano, “Elucidating the folding problem of helical peptides using empirical parameters,” *Nat. Struct. Biol.*, vol. 1, no. 6, pp. 399–409, 1994.
- [14] [Https://automeris.io/WebPlotDigitizer/](https://automeris.io/WebPlotDigitizer/), “(20.6.2019).”
- [15] [Https://capito.nmr.leibniz-fli.de/](https://capito.nmr.leibniz-fli.de/), “(28.6.2019).”
- [16] [Https://www.anaconda.com/](https://www.anaconda.com/), “(16.8.2019).” .

8 Priloge

8.1 Priloga 1

Viri eksperimentalnih podatkov (CD spektrov)

PEPTID	AVTORJI	NASLOV	LET O IZD AJE
C peptid:	Bierzynski A, Kim PS, Baldwin RL.	A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A.	1982.
S peptid:	Kim PS, Bierzynski A, Baldwin RL.	A competing salt-bridge suppresses helix formation by the isolated C-peptide carboxylate of ribonuclease A.	1982.
SPR2 (HD):	Tarcsa E, Candi E, Kartasova T, Idler WW, Marekov LN, Steinert PM.	Structural and transglutaminase substrate properties of the small proline-rich 2 family of cornified cell envelope proteins.	1998.
AP7-1, AP24-1:	Michenfelder M, Fu G, Lawrence C, Weaver JC, Wustman BA, Taranto L, Evans JS, Morse DE.	Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains.	2003.
pKID:	Dahal L, Kwan TOC, Shammas SL, Clarke J.	pKID Binds to KIX via an Unstructured Transition State with Nonnative Interactions.	2017.
PUMA:	Rogers JM, Wong CT, Clarke J.	Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure.	2014.
Chromogranin A:	Yoo SH, Ferretti JA.	Nature of the pH-induced conformational changes and exposure of the C-terminal region of chromogranin A.	1993.

TatH2 (44-89):	Chan CS, Haney EF, Vogel HJ, Turner RJ.	Towards understanding the Tat translocation mechanism through structural and biophysical studies of the amphipathic region of TatA from Escherichia coli.	2011.
P1, P2 (pilosulin):	Zelezetsky I, Pag U, Antcheva N, Sahl HG, Tossi A.	Identification and optimization of an antimicrobial peptide from the ant venom toxin pilosulin.	2005.
Rev ARM:	Casu F, Duggan BM, Hennig M.	The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding.	2013.
L B1 (95-115), L B1 (114-138), L B1 (136-155):	Ramírez-Alvarado M, Serrano L, Blanco FJ.	Conformational analysis of peptides corresponding to all the secondary structure elements of protein L B1 domain: secondary structure propensities are not conserved in proteins with the same fold	1997.
SN, SB, ST peptid:	Cánaves JM, Montal M.	Assembly of a ternary complex by the predicted minimal coiled-coil-forming domains of syntaxin. SNAP-25. and synaptobrevin. A circular dichroism study.	1998.
T peptid (HMGB1):	Knapp S, Müller S, Digilio G, Bonaldi T, Bianchi ME, Musco G.	The long acidic tail of high mobility group box 1 (HMGB1) protein forms an extended and flexible structure that interacts with specific residues within and between. the HMG boxes	2004.
β -Cat (17-48):	Megy S, Bertho G, Gharbi-Benarous J, Baleux F, Benarous R, Girault JP.	Solution structure of a peptide derived from the oncogenic protein beta-Catenin in its phosphorylated and nonphosphorylated states.	2005.
Nogo-33:	Li M, Song J.	The N- and C-termini of the human Nogo molecules are intrinsically unstructured: bioinformatics, CD, NMR characterization, and functional implications.	2007.
Peptid A :	Haarmann CS, Green D, Casarotto MG,	The random-coil 'C' fragment of the dihydropyridine receptor II-III	2003.

	Laver DR, Dulhunty AF.	loop can activate or inhibit native skeletal ryanodine receptors.	
LL-37:	Johansson J, Gudmundsson GH, Rottenberg ME, Berndt KD, Agerberth B.	Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37.	1998.
mArfN37:	Bothner B, Lewis WS, DiGiammarino EL, Weber JD, Bothner SJ, Kriwacki RW.	Defining the molecular basis of Arf and Hdm2 interactions.	2001.
CD4(CYTO):	Wray V, Mertins D, Kiess M, Henklein P, Trowitzsch-Kienast W, Schubert U.	Solution structure of the cytoplasmic domain of the human CD4 glycoprotein by CD and 1H NMR spectroscopy: implications for biological functions.	1998.
KorB NΔ297 (298-358):	Rajasekar K, Muntaha ST, Tame JR, Kommareddy S, Morris G, Wharton CW, Thomas CM, White SA, Hyde EI, Scott DJ.	Order and disorder in the domain organization of the plasmid partition protein KorB.	2010.
Thymosin beta-4:	Czisch M, Schleicher M, Hörger S, Voelter W, Holak TA.	Conformation of thymosin beta 4 in water determined by NMR spectroscopy.	1993.
ID (MDBP 2):	Ghosh RP, Nikitina T, Horowitz-Scherer RA, Gierasch LM, Uversky VN, Hite K, Hansen JC, Woodcock CL.	Unique physical properties and interactions of the domains of methylated DNA binding protein 2.	2010.
LigBCen2NR:	Lin YP, Greenwood A, Nicholson LK, Sharma Y, McDonough SP, Chang YF.	Fibronectin binds to and induces conformational change in a disordered region of leptospiral immunoglobulin-like protein B.	2009.
Myosin IIC tpA (1954-2000):	Ronen D, Rosenberg MM, Shalev DE, Rosenberg M, Rotem	The positively charged region of the myosin IIC non-helical tailpiece promotes filament assembly.	2010.

	S, Friedler A, Ravid S.		
MSR C17A peptid:	Nakamura T, Hinagata J, Tanaka T, Imanishi T, Wada Y, Kodama T, Doi T.	HSP90, HSP70, and GAPDH directly interact with the cytoplasmic domain of macrophage scavenger receptors.	2002.
Scholtz 14, 20, 26, 32, 38, 50:	Scholtz JM, Qian H, York EJ, Stewart JM, Baldwin RL.	Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water.	1991.
HigA2-P1, ccda37-62, phd4 w56-73, phd5 W61-73, RelA_NLS (293-316), FCP1 (YG(944-958), 2G, 4G, 6G, AK17:	<i>Laboratorijske meritve</i>		2019