

UNIVERZITET U BEOGRADU
MATEMATIČKI FAKULTET

MASTER RAD

Razvoj metaprediktora za utvrđivanje neuređenosti proteina

Autor:
Una STANKOVIĆ

Mentor:
dr Jovana KOVAČEVIĆ

ČLANOVI KOMISIJE:

dr Jovana Kovačević
prof. dr Gordana Pavlović-Lažetić
dr Nina Radojčić



Beograd, 2018

Sažetak

Coming soon!

Zahvalnica

Sadržaj

Sažetak	ii
Zahvalnica	iii
1 Uvod	2
2 Biološke osnove	3
2.1 Proteini	3
2.1.1 Funkcije i osobine proteina	4
2.1.2 Struktura proteina	5
2.1.3 Savijanje proteina	7
2.1.4 Denaturacija proteina	7
2.2 Neuređenost proteina	8
2.2.1 Eksperimentalno ispitivanje neuređenosti proteina	10
2.2.2 Računarsko ispitivanje neuređenosti proteina	10
3 Predikcija neuređenosti proteina	12
3.1 Prediktori	12
3.1.1 SPINE-D/ SPOT-D	12
3.1.2 PONDR	12
3.1.3 s2D	12
3.1.4 IUPred	12
3.1.5 ESpritz	12
3.1.6 SEG	12
3.1.7 Disopred2	12
3.2 Baza podataka DisProt	12
4 Aplikacija	13
4.1 Arhitektura	13
4.2 Funkcionalnosti	13
4.3 Korišćenje aplikacije	13
4.4 Primer upotrebe	13
5 Implementacija	14
6 Zaključak	15
Bibliografija	16

Slike

2.1	Prikaz centralne dogme molekularne biologije [4]	4
2.2	Prikaz struktura proteina	6
2.3	Šematski prikaz struktura proteina [6]	7
2.4	Prikaz primarne strukture [6]	8
2.5	Prikaz α -heliksa [6]	9
2.6	Prikaz β -strukture [6]	9
2.7	Prikaz sekundarnih struktura [9]	10
2.8	Prikaz savijanja proteina [7]	11

Tabele

2.1	Spisak esencijalnih i neesencijalnih aminokiselina	4
-----	--	---

Tati, mami i Olgi...

Glava 1

Uvod

Proteini su biološki makromolekuli neophodni za izgradnju i pravilno funkcionisanje ćelija, i igraju mnogobrojne uloge u različitim procesima koji se odvijaju unutar organizma. Struktura proteina zavisi od rasporeda aminokiselina i utiče na njegovu funkciju. Primarna struktura podrazumeva niz aminokiselina koje učestvuju u izgradnji proteina, dok se sekundarna odnosi na oblik koji protein zauzima u prostoru (spirala ili traka). Proteine sa nestabilnom sekundarnom strukturom nazivamo neuređenim. Pored značajne uloge u obavljanju brojnih bioloških funkcija, otkriveno je i da postoji veza između ovih proteina i razvoja savremenih, neizlečivih bolesti i zbog toga su oni u fokusu bioinformatičke zajednice.

Neuređenost proteina se utvrđuje eksperimentalno, laboratorijskim analizama, ili uz pomoć prediktora za automatsko predviđanje neuređenosti proteina. Laboratorijske analize spadaju u spore, veoma skupe metode, koje ne mogu da odgovore na potrebe akademske zajednice i industrije. Iz tog razloga, poslednjih godina, došlo je do razvoja velikog broja alata za automatsko predviđanje neuređenosti proteina. Zbog velike brojnosti ovih alata, razvijaju se metaprediktori koji predstavljaju njihove kombinacije. Specifičan cilj ovog master rada je razvoj jednog metaprediktora za određivanje neuređenosti proteina koji bi konsenzusom objedinio rezultate najnovijih prediktivnih alata na osnovu metodologije na kojoj su zasnovani. Alat će biti testiran na skupu proteina sa eksperimentalno utvrđenom neuređenošću DisProt (eng. *DisProt*).

Glava 2

Biološke osnove

U ovoj sekciji biće ukratko predstavljene biološke osnove neophodne za razumevanje rada i motivacije koja stoji iza određenih njegovih elemenata. Najpre, biće opisano šta su proteini, koje su njihove osnovne funkcije i kakva im je struktura. Potom, biće opisana svaka od struktura ponaosob, uz priložen grafički prikaz istih. Na kraju, posebno će biti opisani neuređeni proteini, njihova uloga i uzroci koji mogu dovesti do njihove pojave.

2.1 Proteini

Proteini (grč. *protos* - "zausimam prvo mesto") su biološki makromolekuli neophodni za izgradnju i pravilno funkcionisanje ćelija, i igraju mnogobrojne uloge u različitim procesima koji se odvijaju unutar organizma. Predstavljaju najvažniji sastojak žive materije i utiču na brojnost i raznolikost živih bića. Specifičnost proteina je tolika da svaka biljna i životinjska vrsta ima svoje proteine, dok se, kod viših organizama, razlikovanje može uočiti i na individualnom nivou. Broj proteina u živim bićima je ogroman, a, kao primer, uzmimo *E.coli* sa 3000 i čoveka sa 5 miliona proteina [1].

Proteini i peptidi su izgrađeni od 22¹ *L* – aminokiseline² koje se javljaju u prirodi i povezani su peptidnim vezama, koje nastaju između α -karboksilne grupe jedne aminokiseline i α -amino grupe druge aminokiseline, pri čemu se oslobađa molekul vode [2]. Ovim postupkom nastaje nerazgranati polipeptidni lanac izgrađen od glavnog lanca, koji se pravilno ponavlja, i međusobno različitih ogranaka. "Standardna" grupa aminokiselina se može podeliti na esecijalne i neesecijalne, čiji spisak se može videti u tabeli 2.1 [2, 3].

Svaki molekul proteina nastaje u ćeliji živog organizma. Proteinski lanac se sastoji od određenog broja i vrsti aminokiselina poređanih po specifičnom redosledu. Ovaj redosled je unapred određen i zavisi od redosleda nukleotida u dezoksiribonukleinskoj kiselini (DNK), odnosno gena, koji predstavljaju trojke nukleotida. Svakoj takvoj trojci, jedinstveno je pridružena po jedna aminokiselina, na osnovu genetskog koda. Proces, kojim se enkodirana informacija prevodi iz DNK u niz aminokiselina u proteinskom lancu, posredstvom glasnike (eng. *messenger*) ribonukleinske kiseline

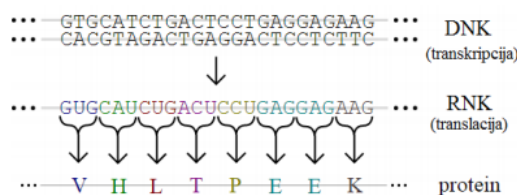
¹Neki proteini u svom sastavu mogu da imaju 22 različite aminokiseline. Pored 20 "standardnih" aminokiselina, postoje i 2 "nestandardne" i to su Selenocistein (eng. *Selenocysteine*, simboli *Sec*, *U*) i Prolizin (eng. *Pyrrolysine*, simboli *Pyl*, *O*). Ove dve aminokiseline se ređe javljaju [2].

²*L* – aminokiseline su one sa levom prostornom konfiguracijom, analogno, postoje i *D* – aminokiseline, sa desnom

Esencijalne	Neesencijalne
Arginin	Alanin
Histidin	Asparagin
Leucin	Asparaginska kiselina
Izoleucin	Cistein
Lizin	Glutaminska kiselina
Metionin	Glutamin
Fenilalanin	Glicin
Treonin	Prolin
Triptofan	Serin
Valin	Tirozin

TABELA 2.1: Spisak esencijalnih i neesencijalnih aminokiselina

(RNK) i transportne (eng. *transfer*) RNK, naziva se genska ekspresija. Proces sinteze proteina predstavlja *centralnu dogmu molekularne biologije*, čiji se prikaz može videti na slici 2.1 [4].



SLIKA 2.1: Prikaz centralne dogme molekularne biologije [4]

2.1.1 Funkcije i osobine proteina

Pri istraživanju bioloških procesa, neophodno je znati i dobro razumeti funkcije proteina, što se posebno može uočiti kod proučavanja oboljenja ljudi, ako se u obzir uzme činjenica da se mnoga oboljenja pojavljuju kao posledica funkcionalnih mutacija. Proteini su biološki najaktivniji molekuli sa velikim brojem esencijalnih funkcija koje se dele na:

- dinamičke, od kojih su najvažnije:
 1. transportna - prenos molekula (poput kiseonika, gvožđa, lipida) i hormona od mesta sinteze do mesta delovanja,
 2. biološka - regulacija metaboličkih procesa u ćeliji, kontrola i regulacija transkripcije gena i translacija,
 3. katalizatorska - biološka katalizacija ³,
 4. zaštitna - keratin, koagulacija krvi,
 5. održavanje zapremine tečnosti u organizmu,
- strukturne, od kojih su najvažnije:

³Katalizacija predstavlja proces povećavanja brzina reakcija

1. obezbeđivanje čvrstine i elastičnosti organa,
2. davanje oblika organizmu,
3. izgradnja strukturnih elemenata ćelije i
4. bitna uloga u kontraktilnim i pokretnim elementima organizma.

Postoji nekoliko karakteristika proteina:

- grade kompleksna jedinjenja sa različitim supstancama po principu strukturne komplementarnosti i
- poseduju visoku osetljivost na različite agense koji ih denaturišu ⁴. Neki od najčešćih agenasa su: visoka temperatura, pritisak, mehaničko tretiranje, dejstvo kiselina, baza, organskih rastvarača, materija, itd. [1, 4].

2.1.2 Struktura proteina

Struktura proteina zavisi od rasporeda aminokiselina i utiče na njegovu funkciju. Trodimenzionalna struktura, koja se smatra najstabilnijom, formira se presavijanjem polipeptidnog lanca na različite načine. Unutrašnjost takve strukture ima visoku gustinu, pa polipeptidni lanac ne dopušta promene u sastavu i zahteva prisustvo aminokiselina tačno određene veličine. Uobičajena raspodela aminokiselina u proteinima je daleko od ravnomerne. Neke aminokiseline se javljaju mnogo češće od ostalih, tako se, na primer, leucin pojavljuje devet puta više od triptofana. Proteinsku strukturu održavaju različite vrste kovalentnih i nekovalentnih interakcija između hemijskih jedinjenja, npr. vodonične, jonske, elektrostatičke, dipolne, itd.. Nabiranjem i uvijanjem lanaca kreiraju se različiti oblici proteina: vlaknasti, globularni ili eliptični. Ako mutacija dovede do toga da aminokiselina sa malim bočnim lancem bude zamenjena aminokiselinom sa velikim, pojaviće se problem sa formiranjem trodimenzionalne strukture. Ako bi se, pak, velika aminokiselina zamenila sa malom, pojavio bi se prazan prostor, što bi moglo dovesti do destabilizacije molekula proteina [1, 3, 5].

U molekulima proteina postoji hijerarhijska strukturalna organizacija u četiri nivoa:

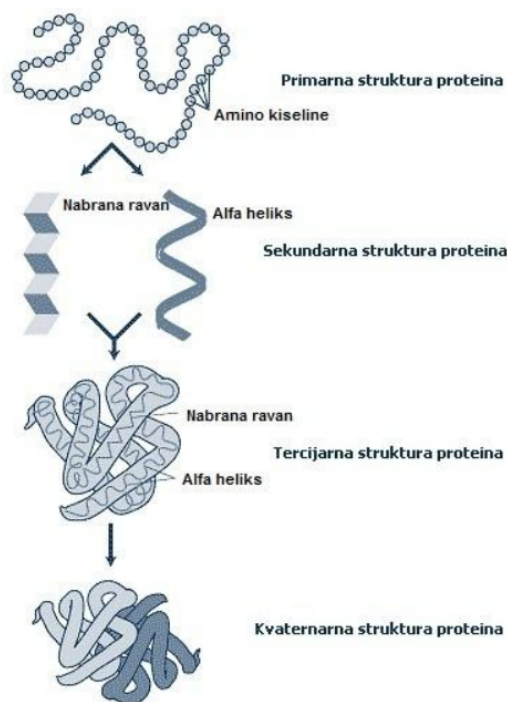
1. primarna,
2. sekundarna,
3. tercijarna i
4. kvaternarna [1].

Na slici 2.2 se može videti opšti prikaz mogućih struktura proteina, a na drugoj 2.3 šematski prikaz.

Primarna struktura Predstavlja samu sekvencu aminokiselina⁵ koje učestvuju u izgradnji proteina. Ova struktura ima ključni značaj za određivanje funkcije proteina zbog interakcija koje se javljaju između bočnih lanaca aminokiselina, a koji utiču na trodimenzionalnu strukturu. Proteini koji poseduju sličnu sekvencu aminokiselina nazivaju se *homologi*, a poređenje sekvenci među takvim proteinima može ukazati na

⁴Denaturacija proteina je proces koji izaziva promene u strukturi proteina, čime se menja i njihov fiziološki uticaj.

⁵Redosled kojim su aminokiseline poredane u nekom polipeptidu se zove sekvenca aminokiselina [1].



SLIKA 2.2: Prikaz struktura proteina

genetsku relaciju između različitih vrsta. Prikaz izgleda primarne strukture na primeru insulina kod čoveka se vidi na slici 2.4 [1].

Mnoge genetske bolesti rezultuju u proteinima sa abnormalnim redosledom aminokiselina, što uzrokuje nepravilno presavijanje i gubitak ili nemogućnost normalnog funkcionisanja. Ukoliko su nam poznate strukture normalnih i mutiranih proteina, te informacije možemo iskoristiti za dijagnostikovanje ili proučavanje bolesti. Primarna struktura će sa mutacijama, u najmanju ruku, izmeniti unos, brisanje ili menjanje aminokiselina. Promene u primarnoj strukturi mogu imati uticaja i na više nivoa proteinskih struktura. Takve promene često dovode do lošeg presavijanja proteina i mogu dovesti do njegovog gubitka funkcije [8, 7].

Sekundarna struktura Odnosi se na oblik koji protein zauzima u prostoru i označava pravilno pojavljivanje ponavljano prostornog rasporeda primarne strukture, u jednoj dimenziji. Ovu strukturu čini nekoliko različitih oblika, od kojih su najčešći α -heliks i β -presavijena traka (ili β -struktura), a čest je i tzv. β -okret [1, 5].

α -heliks - tip sekundarne strukture kod kog se gusto pakovani polipeptidni lanac spiralno uvrće. Karakteriše se brojem peptidnih jedinica po okretu i rastojanjem između dva okreta. Spada pod energetski najsiromašniji, a time, i najstabilniji strukture proteina. Heliks mogu obrazovati i *L*- i *D*- aminokiseline, pa postoje i dva tipa heliksa: levostrani i desnostrani. Prikaz izgleda α -heliksa se vidi na slici 2.5 [1].

β -struktura - Za razliku od α -heliksa, sastoji se od dva ili više peptidnih lanaca, ili segmenata polipeptidnih lanaca, a obrazuje se kada se ovakvi tipovi lanca povežu uzdužno. Postoje dva tipa β -strukture: paralelna i antiparalelna. Prikaz izgleda β -strukture se vidi na slici 2.6 [1].



SLIKA 2.3: Šematski prikaz struktura proteina [6]

β -okreti - obrću pravac polipeptidnog lanca praveći kompaktan globularan oblik [7]. Prikaz izgleda sekundarnih struktura se nalazi na slici 2.7.

Tercijarna struktura Podrazumeva unutarmolekularno slaganje polipeptidnog lanca u kompaktnu trodimenzionalnu strukturu specifičnog oblika, koja nastaje prostornim organizovanjem polipeptidnog lanca, koji već poseduje sekundarnu strukturu. Na taj način, približavaju se ostaci aminokiselina koji su udaljeni u primarnoj strukturi. Proteini koji imaju ovakvu strukturu su globularni i kompaktni sa velikom gustinom u središtu [1, 5].

Kvaternarna struktura Predstavlja agregaciju više peptidnih lanaca u molekulu proteina. Mnogi proteini, posebno oni velike mase, izgrađeni su od nekoliko polipeptidnih lanaca. Svaka takva komponenta naziva se *podjedinica* ili *protomer*. Oni mogu biti identični⁶ ili se razlikovati prema strukturi. Ovakav raspored dovodi do brzog i efikasnog transfera substrata od jednog aktivnog centra enzima do drugog [1, 5].

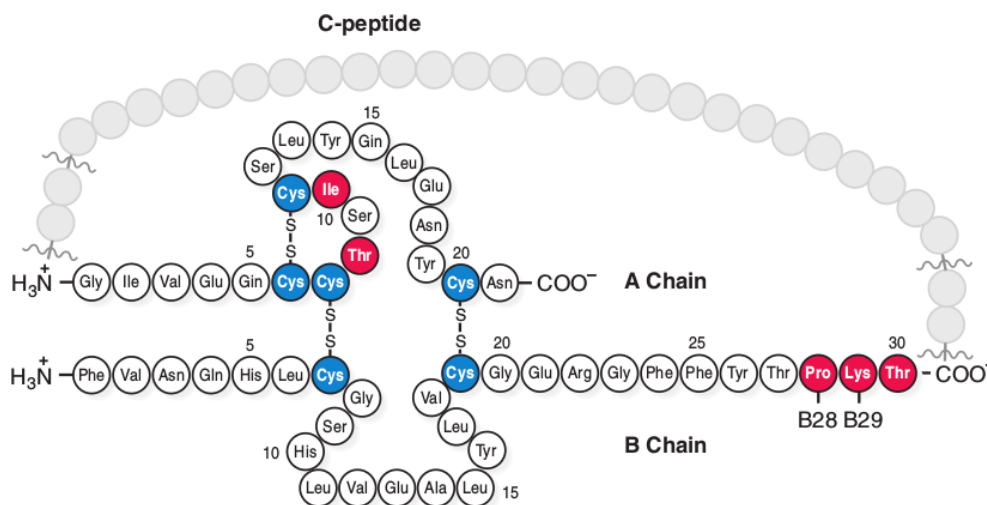
2.1.3 Savijanje proteina

Interakcije između lanaca aminokiselina, koji se nalaze sa strane, određuju kako se dugački polipeptidni lanac presavija u trodimenzionalni oblik funkcionalnog proteina. Presavijanje proteina koje se događa u ćeliji traje od nekoliko sekundi do nekoliko minuta. Na slici 2.8 se može videti opšti prikaz savijanja proteina.

2.1.4 Denaturacija proteina

Denaturisanje proteina rezultuje u odvijanju i dezorganizaciji proteinske sekundarne i tercijarne strukture. Pod idealnim uslovima, denaturisanje proteina može biti *reverzibilno*. To znači da bi se protein, pri prestanku delovanja agenasa, vratio u

⁶Tada takve proteine nazivamo *oligomerima*



SLIKA 2.4: Prikaz primarne strukture [6]

normalno stanje. Međutim, većina proteina ostaje trajno neuređena. O neuređenosti proteina biće više reči u nastavku.

Jedno od objašnjenja zašto se protein ne vraća u originalno stanje se sastoji u tome da protein počinje sa savijanjem pre nego što se izvrši sinteza celog lanca. Osim toga, specijalizovana grupa pomoćnih proteina (engl. *chaperones*) je neophodna za pravilno savijanje mnogih vrsta proteina. Ovi pomoćni proteini interaguju sa polipeptidima u nekoliko faza tokom procesa savijanja, imaju ulogu u tome da održavaju protein nesavijenim dok sinteza nije gotova, ili imaju ulogu katalizatora. Loše savijanje proteina može dovesti do različitih bolesti kao što su: amiloidna bolest ili Prionova bolest [7].

2.2 Neuređenost proteina

Eksperimentalnim utvrđivanjem sekundarne strukture proteina (koje će biti detaljnije opisano u 2.2.1) uočeno je da se neretko, pod određenim fiziološkim uslovima, javljaju proteini sa trodimenzionalnom strukturom koja nije dobro definisana. Usled velikog broja termina koji se koriste za opisivanje ovakvih proteina: prirodno/suštinski neuređeni, nesavijeni, denaturisani ili reomorfni proteini (eng. *intrinsically disordered/unfolded/unstructured*), u ovom radu, biće korišćen samo kraći termin - neuređeni proteini.

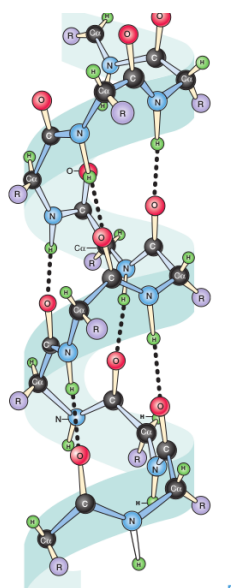
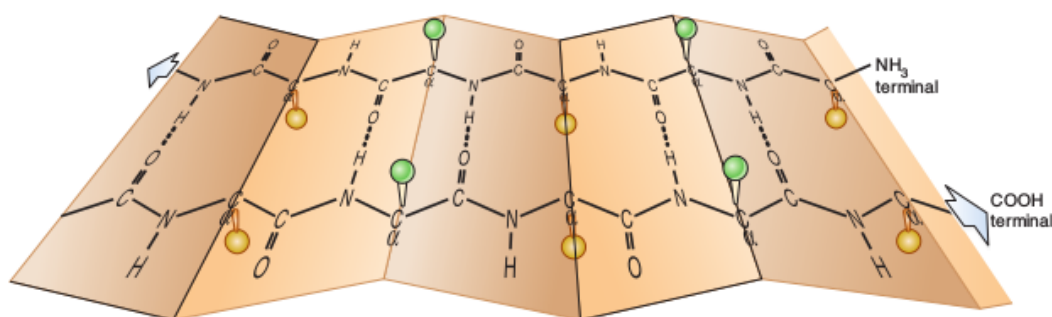
Neuređenost predstavlja inherentno⁷ svojstvo sekvence. Neuređen može biti ceo protein, a mogu biti neuređeni određeni regioni proteina različitih dužina. Kao posledica, ovakve proteine nazivamo inherentno neuređenim proteinima, skraćeno IDP⁸, a ako su u pitanju neuređeni, ali funkcionalni, regioni, onda je skraćenica IDPr⁹. Statističkom analizom došlo se do zaključka da se aminokiseline mogu klasterovati na dve grupe:

1. aminokiseline koje promovišu uređenost (eng. *order promoting*) i
2. aminokiseline koje promovišu neuređenost (eng. *disorder promoting*).

⁷Inherentno = nasleđeno

⁸eng. *Intrinsically Disordered Proteins*

⁹eng. *Intrinsically Disordered Protein Regions*

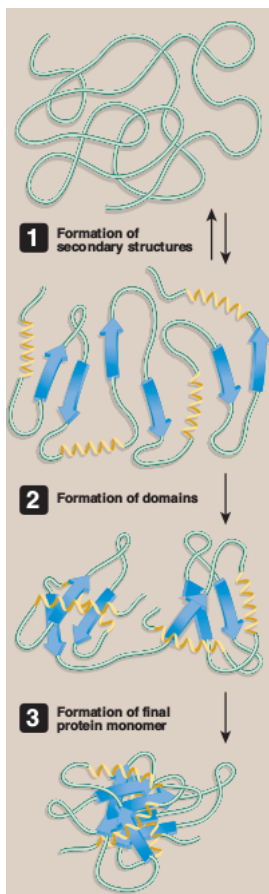
SLIKA 2.5: Prikaz α -heliksa [6]SLIKA 2.6: Prikaz β -strukture [6]

Neuređene proteine ili neuređene regione je teško kategorizovati, a jedan od opštih opisa strukture dat je kao kombinacija više tipova foldona¹⁰:

- foldon (eng. *foldon*) je nezavisno organizujuća jedinica(region) proteina,
- indukativni foldon (eng. *inducible foldon*) je neuređeni region proteina koji savijanje lanca postiže barem delom vezivajući se za partnera,
- ne-foldon (eng. *non-foldon*) je neuređeni region proteina koji nikada ne postiže uređenost,
- polu-foldon (eng. *semi-foldon*) je neuređeni region proteina koji ostaje polovično neuređen i nakon vezivanja za partnera, i
- anti-foldon (eng. *unfoldon*) je region proteina koji iz uredenog prelazi u neuređeno stanje u cilju izvršavanja neke funkcije.

Postoji nekoliko mogućih stanja (oblika) u kojima se protein može naći. Ova stanja i prelazi između njih(neki proteini mogu prelaziti iz neuređenog u uređeno stanje, i obratno), prema *hipotezi proteinskog trojstva*, utiču na funkciju proteina. Svaki od

¹⁰Foldon ostaje u originalnom nazivu, kao posledica manjka literature. [9]



SLIKA 2.8: Prikaz savijanja proteina [7]

време се ради и на kreiranju metaprediktora, koji predviđanje vrše kombinovanjem više tehnika. O ovoj vrsti predikcije biće više reči u narednom poglavlju.

Glava 3

Predikcija neuređenosti proteina

Kao što je navedeno u prethodnom poglavlju, neuređenost proteina se može, osim eksperimentalnog, određivati i računarski. Upravo o tom vidu određivanja, odnosno, predikcije, neuređenosti proteina govori ovo poglavlje. Najpre, biće detaljnije opisan računarski postupak. Nakon toga, biće priče o prediktorima, od kojih će pojedini biti detaljnije objašnjeni. Na kraju, ukratko, će biti predstavljena baza podataka DisProt i njen značaj u ovom radu.

3.1 Prediktori

3.1.1 SPINE-D/ SPOT-D

3.1.2 PONDR

3.1.3 s2D

3.1.4 IUPred

3.1.5 ESpritz

3.1.6 SEG

3.1.7 Disopred2

3.2 Baza podataka DisProt

Glava 4

Aplikacija

4.1 Arhitektura

4.2 Funkcionalnosti

4.3 Korišćenje aplikacije

4.4 Primer upotrebe

Glava 5

Implementacija

Glava 6

Zaključak

Bibliografija

- [1] Vesna Spasojević-Kalimanovska Slavica Spasić Zorana Jelić-Ivanović. *Opšta biohemija*. 2002.
- [2] Marija Jeličić. *Povezanost dužine epitopa i uređenostidelova proteina*. on-line na: <http://elibrary.matf.bg.ac.rs/bitstream/handle/123456789/2428/Marijapdf?sequence=1>. 2012.
- [3] Gerhard Michal Dietmar Schomburg. *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2012. ISBN: 9780470146842.
- [4] Jovana Kovačević. *Strukturna predikcija funkcije proteina i odnos funkcionalnih kategorija proteina i njihove neuređenosti*. on-line na: <http://www.math.rs/files/DoktoratJK2015.pdf>. 2015.
- [5] Ivana Čepelak Dubravka Čvorišćec. *Štrausova medicinska biokemija*. Medicinska naklada, 2009.
- [6] Michael Lieberman. *Biochemistry, molecular biology, and genetics. — 6th edition*. 351 West Camden Street, Baltimore, MD 21201, Two Commerce Square, 2001 Market Street, Philadelphia, PA 19103: Lippincott Williams & Wilkins, a Wolters Kluwer business, 2014. ISBN: 978-1-4511-7536-3.
- [7] Denise R. Ferrier Richard A. Harvey. *Lippincott Illustrated Reviews: Biochemistry, 5th edition*. 351 West Camden Street, Baltimore, MD 21201, Two Commerce Square, 2001 Market Street, Philadelphia, PA 19103: Lippincott Williams & Wilkins, a Wolters Kluwer business, 2011. ISBN: 978-1-60831-412-6.
- [8] Bradford A. Jameson Denise R. Ferrier. *Lippincott's Illustrated Reviews Flash Cards*. 2001 Market Street, Philadelphia, PA 19103: Wolters Kluwer Health, 2015. ISBN: 978-1-4511-9111-0.
- [9] Goran Vinterhalter. *Bioinformatička analiza povezanosti funkcije i neuređenosti proteina*. on-line na: http://www.racunarstvo.matf.bg.ac.rs/MasterRadovi/2017_08_23_Goran_Vinterhalter/rad.pdf. 2018.
- [10] A.Keith Dunker et al. *Intrinsically disordered protein*. on-line na: [https://doi.org/10.1016/s1093-3263\(00\)00138-8](https://doi.org/10.1016/s1093-3263(00)00138-8). 2001.
- [11] A. Keith Dunker Christopher J. Oldfield. *Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions*. on-line na: <https://doi.org/10.1146/annurev-biochem-072711-164947>. 2014.
- [12] Vladimir N. Uversky. *Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins*. on-line na: <https://doi.org/10.1074/jbc.r115.685859>. 2016.