

# Nativno neuređeni proteini

---

Tomljanović, Ingrid

Undergraduate thesis / Završni rad

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:560806>

Rights / Prava: [In copyright](#)

Download date / Datum preuzimanja: **2020-10-30**



Repository / Repozitorij:

[Repository of Faculty of Science - University of Zagreb](#)





**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO – MATEMATIČKI FAKULTET**  
**BIOLOŠKI ODSJEK**

**NATIVNO NEUREĐENI PROTEINI**

**INTRINSICALLY DISORDERED PROTEINS**

**SEMINARSKI RAD**

Ingrid Tomljanović

Preddiplomski studij molekularne biologije  
(Undergraduate Study of Molecular Biology)

Mentor: doc. dr. sc. Dubravko Pavoković

Zagreb, 2016.

## SADRŽAJ

1.	UVOD.....	2
1.1.	STRUKTURA GLOBULARNIH PROTEINA.....	2
1.2.	NATIVNO NEUREĐENI PROTEINI (IDP).....	3
2.	VAŽNOST I OTKRIĆE IDP.....	4
3.	BIOINFORMATIČKI ALATI ZA ANALIZU IDP.....	7
4.	EKSPERIMENTALNE METODE ISTRAŽIVANJA IDP.....	9
5.	PODJELA IDP PREMA MEHANIZMU DJELOVANJA I STANIČNOJ ULOZI.....	11
6.	IDP U BILJKAMA PRI NORMALNIM I STRESNIM UVJETIMA.....	13
7.	ZAKLJUČAK.....	14
8.	ZAHVALE.....	14
9.	LITERATURA.....	15
10.	SAŽETAK.....	19
11.	SUMMARY.....	19

## 1. UVOD

Proteini su linearni polimeri aminokiselina koji najčešće posjeduju stabilnu i kompleksnu trodimenzionalnu strukturu. Nativno neuređeni proteini (IDP, *engl.* intrinsically disordered proteins) su proteini koji ne posjeduju stabilnu tercijarnu strukturu pri fiziološkim uvjetima. Istraživanja u posljednja dva desetljeća pokazala su da, usprkos nedostatku strukture, IDP sudjeluju u nizu staničnih procesa, a posebice u signalizaciji i regulaciji. IDP čine značajan udio proteoma mnogobrojnih organizama, a njihov udio raste s kompleksnošću organizma (Tompa, 2012., Dyson, 2016.). Cilj ovog rada opisati je osnovne značajke IDP te najvažnije eksperimentalne i bioinformatičke metode istraživanja IDP. Nadalje, rad obuhvaća podjelu IDP prema mehanizmu djelovanja i staničnoj ulozi te kratki osvrt na IDP u biljkama pri normalnim i stresnim uvjetima.

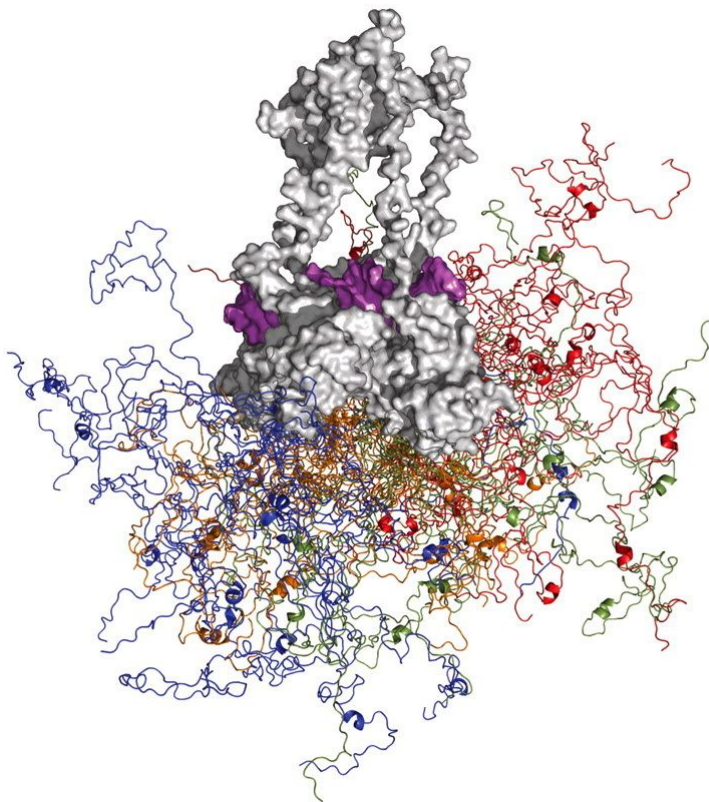
### 1.1. STRUKTURA GLOBULARNIH PROTEINA

Proteini su polimeri građeni od aminokiselina vezanih peptidnom vezom. Broj aminokiselina u proteinima znatno varira, no većina ih sadrži manje od 2000 aminokiselina. U svim organizmima proteine gradi 20 aminokiselina u L-konformaciji. Svaka aminokiselina ima amino i karboksilnu skupinu, atom vodika te bočni ogranak vezan na alfa atom ugljika. Bočni ogranci aminokiselina određuju njihova svojstva te se one međusobno razlikuju po veličini, reaktivnosti, hidrofobnosti i naboju (Nelson i Cox, 2013).

Primarnu strukturu proteina čini linearni slijed aminokiselina, koje svojim svojstvima definiraju sekundarnu, tercijarnu i kvaternu strukturu. Sekundarna struktura odnosi se na određeni segment polipeptidnog lanca te opisuje lokalnu prostornu raspodjelu atoma proteinske okosnice. Neki od najčešćih elemenata sekundarne strukture su  $\alpha$ -zavojnica,  $\beta$ -ploča i  $\beta$ -okret. Tercijarnu strukturu proteina čini sveukupan trodimenzionalni raspored atoma polipeptidnog lanca. Ukoliko je protein sastavljen od više zasebnih polipeptidnih lanaca (podjedinica), pri opisu njihovog položaja u prostoru govorimo o kvaternoj strukturi danog proteina (Nelson i Cox, 2013).

## 1.2. NATIVNO NEUREĐENI PROTEINI (IDP)

Nativno neuređeni proteini i proteinske regije (*engl.* intrinsically disordered proteins/regions, IDP/IDR) pripadaju nedavno priznatoj skupini proteina koji su biološki aktivni usprkos nedostatku jasno definirane trodimenzionalne strukture (Uversky, 2014a). Za razliku od globularnih proteina, IDP u nativnom stanju ne posjeduju jedinstvenu stabilnu strukturu, već fluktuirajući skup strukturnih konformacija (Slika 1., Uversky, 2014b; Csizmok i sur., 2007). Takve skupove konformacija (*engl.* conformational ensembles) odlikuje heterogenost i dinamičnost, tj. izrazita varijabilnost prostornih koordinata atoma polipeptidnog lanca u ovisnosti o vremenu (Uversky, 2014b; Habchi i sur., 2014). IDP najčešće pokazuju manjak stabilne tercijarne, rjeđe i sekundarne strukture, ali nipošto ne i potpuno odsutstvo strukturnih elemenata (Tomba, 2012).



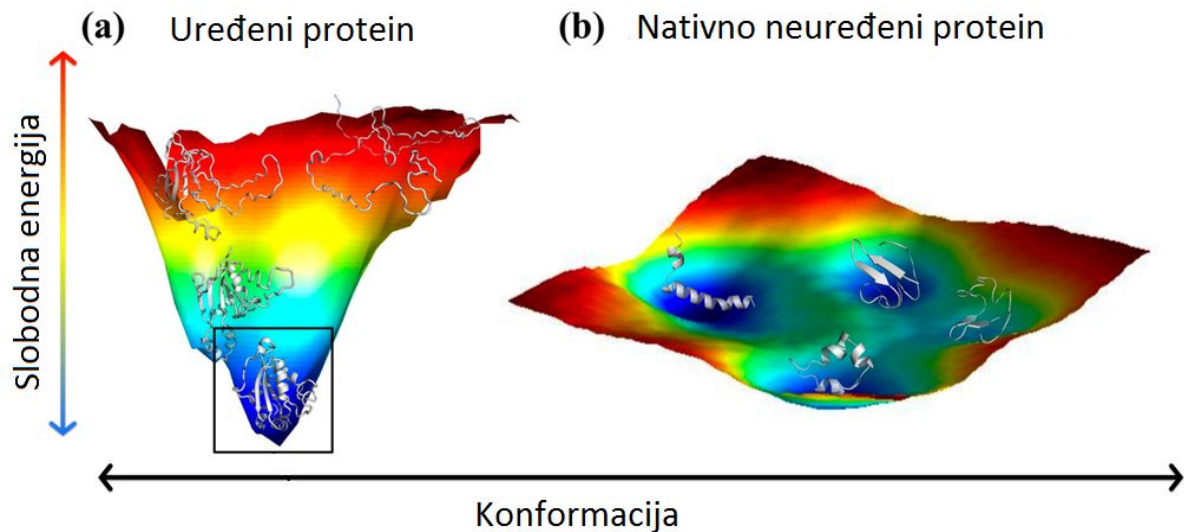
**Slika 1.** Primjer IDP: struktura tumor supresora p53 i njegove nativno neuređene N-terminalne transaktivacijske domene prikazane kao skup od 20 konformacija (Wells i sur., 2008).

## 2. VAŽNOST I OTKRIĆE IDP

Postupno otkrivanje proteina čija konformacijska stabilnost nije izravan preduvjet njihovoj funkcionalnosti dovelo je, u proteklih nekoliko desetljeća, do promjene ustaljenih gledišta o vezi strukture s funkcijom proteina te do osnutka novog područja istraživanja (Tompa, 2012). Upravo je nedostatak strukturnog uređenja element koji se smatra ključnim za funkciju IDP (Pietrosemoli i sur., 2013). Za ulogu mnogih proteina nužno je da istodobno sadrže i uređene i neuređene regije (Uversky, 2014b).

Zahvaljujući unaprijeđenim metodama istraživanja, IDP su postali prihvaćeni kao mnogobrojna klasa proteina s važnim staničnim ulogama. Primjerice, važnost IDP očituje se kroz njihovu ulogu u razvoju neurodegenerativnih poremećaja, tumora i kardiovaskularnih bolesti kod ljudi (Uversky i sur., 2014e; Midic i sur., 2009; Uversky i sur., 2008). IDP ujedno predstavljaju i atraktivne mete za razvoj novih lijekova (Uversky, 2012; Uversky, 2014b), a nedavno je istaknut i niz potencijalnih primjena IDP u biotehnologiji (Uversky, 2014d).

Razlog relativno nedavnom otkriću i priznavanju IDP bilo je prevladavajuće mišljenje da je jasno definirana stabilna struktura nužan preduvjet za postojanje specifične funkcije nekog proteina. Navedeni pogled, poznat kao tzv. “*structure-function paradigm*,” potkrjepljuje činjenica da su izravan uzrok gubitka funkcije proteina uslijed denaturacije upravo deformacije u njegovoj strukturi. Ovaj koncept predstavlja temelj moderne strukturne biologije te se očituje već u Fischerovom ključ-brava modelu iz 1894. Snažno ga potkrjepljuju mnogobrojne proteinske strukture u Protein Data Bank (PDB) bazi podataka, koja sadrži čak 121414 strukture od 09.08.2016. (Bernstein i sur., 1977). Stoga, prva otkrića funkcionalnih proteina koji svojom prirodom odstupaju od ove paradigme bila su smatrana anomalijama i nerijetko ignorirana od strane šire znanstvene zajednice (Tompa, 2012). No, danas je potvrđeno da mnoge strukture u PDB-u sadrže nativno neuređene regije (IDR), koje se smataju u strukturne elemente isključivo pri vezanju liganda (Tompa, 2012). Također, upravo se dinamičnošću IDP i IDR razjašnjavaju pojave kao što su poteškoće u kristalizaciji proteina i vrlo česte regije nedostatka elektronske gustoće (*engl.* missing electron density regions) u strukturama riješenim metodom kristalografije X-zraka (Gall i sur., 2007; Dosztanyi i sur., 2009). Ove regije su u tolikoj mjeri mobilne, tj. nedostaje im stabilna fiksna struktura, da ih se ne može adekvatno opisati samo jednim skupom prostornih koordinata (Slika 2, Uversky, 2014b). Stoga, proteini mogu biti neuređeni na različitim mjestima duž polipeptidnog lanca, u različitim količinama te na različitim razinama strukture.



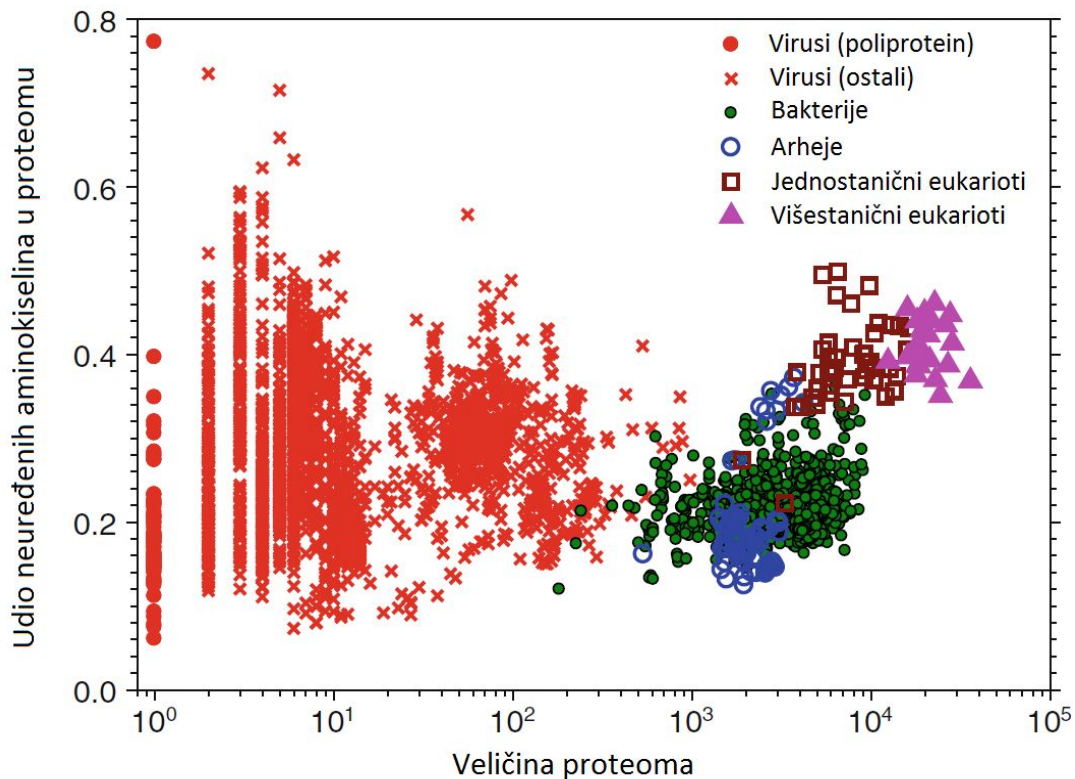
**Slika 2.** Usporedni shematski prikaz strukturnih konformacija i krajobraza slobodne energije (*engl.* free-energy landscape). Naspram jedne konformacije uređenog proteina (a), IDP ima skup od 4 različite konformacije (b). Prilagođeno iz Burger i sur. (2014).

Pomna eksperimentalna karakterizacija IDP bila je isprva onemogućena time što je većina metoda bila prilagođena analizi globularnih proteina, ali je nekolicina značajnih računalnih istraživanja omogućila najraniji uvid u mnoge aspekte IDP koji su kasnije eksperimentalno potvrđeni. Fokus ovih istraživanja bile su mnogobrojnost i uloga IDP u staničnim procesima. Već 1998. godine pokazano je da tadašnje izdanje baze podataka SwissProt sadrži 15,000 proteina s neuređenim regijama od 40 ili više uzastopnih aminokiselina (Romero i sur., 1998).

Kasnije studije ukazale su na činjenicu da su IDP manje brojniji u prokariotskim nego u eukariotskim proteomima. Primjerice, Tompa (2012) navodi da prema konzervativnim procjenama 10-35% prokariotskih i 15-45% eukariotskih proteina sadrži duge neuređene regije, (LDR, *engl.* long disordered regions) od 30 ili više uzastopnih aminokiselina. Proteini se najčešće klasificiraju kao IDP ako sadrže bar jednu ovakvu LDR. Prema Peng i sur. (2014) prokarioti imaju 7%, a eukarioti 25% LDR u svojim proteinima. Ward i sur. (2004a) tvrde da su LDR prisutne u 2.0% arhejskih, 4.2% bakterijskih i 33.0% eukariotskih proteina.

Dosad najopsežnija računalna studija IDP, provedena na ukupno 3484 proteoma, ukazuje na jasnu granicu u količini neuređenosti između prokariota i eukariota, kao i na oštar porast u neuređenosti pri prijelazu iz prokariotskog u eukariotski tip stanice (Slika 3). Autori predlažu da takav porast broja IDP i IDR odražava mehanizam prilagodbe stanica na

povećane potrebe za signalizacijom i regulacijom pri porastu stanične kompleksnosti (Xue i sur., 2012). Postotci neuređenosti, odnosno udjeli IDP po skupinama organizama, u navedenim se studijama razlikuju zbog različitih softverskih alata korištenih za procjenu neuređenosti, no u svima je vidljiv trend porasta količine IDP s kompleksnošću stanica.



**Slika 3.** Sadržaj neuređenosti (*engl.* disorder content: the average fraction of disordered residues) u odnosu na veličinu proteoma. Šest skupina označenih simbolima obuhvaća 3,484 proteoma iz virusa, arheja, bakterija i eukariota. Korišten je procjenitelj neuređenosti PONDR-VSL2B. Prilagođeno iz Xue i sur., 2012.

Iako je proveden niz bioinformatičkih istraživanja proteinske neuređenosti poput prethodno navedenih, ona su pružila uvid u količinu neuređenosti u proteomima mnogih organizama, ali nisu uspostavila deskriptivne statističke parametre koji bi omogućili objektivnu procjenu neuređenosti među proteomima. Stoga, u nedavnom je radu predstavljen set statističkih mjera neuređenosti proteoma te je računalnim procjeniteljima IUPred i DisEMBL detaljno analizirana neuređenost u 10 eukariotskih proteoma. Rezultat ovog kvantitativnog istraživanja je set smjernica koje će olakšati objektivno određivanje statusa neuređenosti određenog proteina (Schnell i sur., 2016).



### 3. BIOINFORMATIČKI ALATI ZA ANALIZU IDP

Primjena bioinformatičkih alata u ranim računalnim studijama uvelike je potakla razvoj područja istraživanja nativno neuređenih proteina (Habchi i sur., 2014). Bioinformatički alati su i danas od iznimne važnosti za izučavanje svojstava IDP, a validacijom rezultata i usmjeravanjem eksperimentalnih istraživanja ubrzavaju otkriće novih IDP i IDR (He i sur., 2009). Dosad je, zahvaljujući specifičnostima proteinskog slijeda IDP, razvijeno više od 50 različitih računalnih alata za procjenu nativne neuređenosti proteina (procjenitelji neuređenosti, *engl.* intrinsic disorder predictors) (He i sur., 2009).

Procjenitelji neuređenosti kao ulazne podatke koriste aminokiselinski slijed proteina. Zatim, na temelju određenog kriterija pružaju informaciju o sklonosti polipeptidnog lanca za neuređenošću (*engl.* disorder propensity) na razini pojedinih aminokiselina (He i sur., 2009). U razvoju prvih algoritama za procjenu neuređenosti, kao kriterij za identifikaciju neuređenih regija iskorištene su posebnosti aminokiselinskog sastava IDP. Aminokiselinski sastav te kompleksnost sekvence IDP statistički se značajno razlikuje od sastava globularnih proteina, što omogućuje predviđanje neuređenosti s velikom točnošću. Analizama primarne strukture pokazano je da su IDP bogati hidrofilnim i polarnim aminokiselinama te glicinom i prolinom, a siromašni aromatskim i hidrofobnim aminokiselinama (Dosztanyi i sur. 2009; Habchi i sur., 2014; ). Nadasve se ističe mali udio triptofana, fenilalanina, valina, izoleucina i tirozina u IDP (Lietaud i sur. 2016.). Ovakav sastav odgovoran je za nemogućnost potpunog smatanja IDP u stabilne tercijarne strukture, a nekad i sekundarne strukture (Habchi i sur., 2014).

Procjenitelji neuređenosti međusobno se jako razlikuju te je zato pri detaljnoj interpretaciji neuređenosti nekog proteina nužno koristiti više njih ili kvalitetan metaprocjenitelj (metaprediktor), te poznavati za što je softver specijaliziran i obratiti pažnju na postojanje konsenzusnih regija u izlaznim podacima. Metaprocjenitelji su često kvalitetniji u procjeni neuređenosti od samostalnih alata. Oni istovremeno kombiniraju nekoliko procjenitelja te traže konsenzus njihovih procjena. Prema načelima i metodama na kojima počivaju, procjenitelje neuređenosti može se svrstati u tri široke kategorije: algoritme bazirane na fizikalno-kemijskim svojstvima aminokiselina, algoritme bazirane na strojnom učenju te algoritme koji mjere mogućnost ostvarivanja stabilizirajućih intramolekularnih kontakata. Od 2016. godine dostupan je i web server Mollack koji konstruira modele neuređenih proteina na osnovi eksperimentalnih i simuliranih podataka (<http://cmstultz-mollack.mit.edu>, 08.09.2016., Stultz i sur. 2016).

Procjenitelji koji pripadaju prvoj širokoj skupini, tzv. “propensity-based predictors” oslanjaju se na fizikalno-kemijska svojstva aminokiselina te na sklonost aminokiselina za formiranje određenih strukturnih elemenata, uzimajući pritom u obzir načelo da IDP posjeduju visoki neto naboj i nisku hidrofobnost (Uversky, 2014b). Ovakvi alati su CH plot (Uversky i sur., 2000), FoldIndex (Prilusky i sur., 2005) i GlobPlot (Linding, 2003).

Druga velika skupina alata za procjenu neuređenosti obuhvaća algoritme bazirane na strojnom učenju. Ona uključuje prve razvijene procjenitelje neuređenosti, tj. PONDR skupinu (<http://www.pondr.com/>, 09.08.2016.) (Romero i sur., 1997; Li i sur., 1997; Romero i sur., 2000; Peng i sur., 2006). U ovu skupinu također pripadaju i algoritmi DisEMBL te DISOPRED2 (Linding i sur., 2003; Ward i sur., 2004b). Nedavno su predstavljene i procjenitelji DisPredict te DisoMCS (Hoque i sur. 2015., Cong i sur. 2015.) te DFLpred, prvi algoritam za identifikaciju neuređenih proteina fleksibilnih linkera (Kurgan i sur., 2016.).

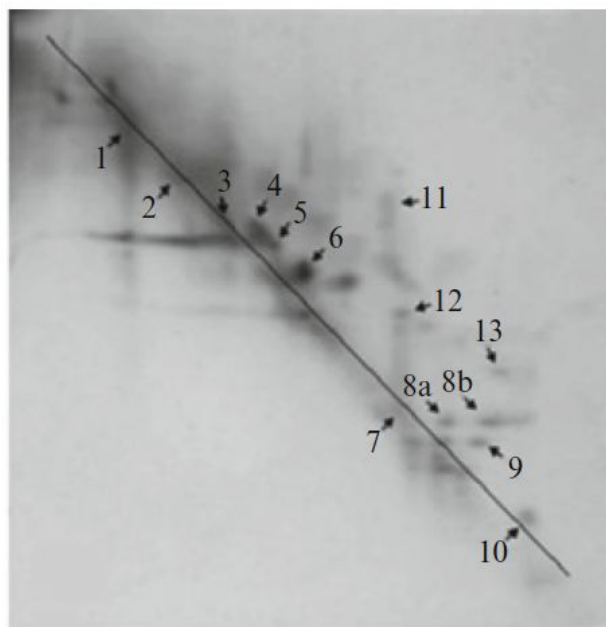
Ovakvi procjenitelji trenirani su na podacima iz proteinskih baza podataka, primjerice na kratkim regijama nedostatka elektronske gustoće u PDB-u ili na neuređenim proteinima i regijama iz baze DisProt (<http://www.disprot.org/>, 09.08.2016.). DisProt je baza eksperimentalno potvrđenih IDP, čija najnovija verzija sadrži 694 IDP i 1539 IDR (Sickmeier i sur., 2007). Broj zapisa u bazi DisProt je relativno malen naspram rezultata računalnih predikcija brojnosti IDP. Ta činjenica jasno ocrta veliku potrebu za novim visokoprotočnim studijama fokusiranim na eksperimentalnu identifikaciju IDP na proteomskoj razini (Csizmek i sur., 2007). Uz DisProt, potrebno je istaknuti postojanje još tri baze podataka o IDP: MobiDB, IDEAL i D2P2 (Tossato i sur., 2014, Ota i sur. 2014, Gough i sur. 2013.).

Treća skupina alata vodi se idejom da se IDP ne smataju potpuno jer ne mogu ostvariti dovoljan broj stabilizirajućih kontakata između svojih konstitutivnih aminokiselina. Primjer ovakvog alata je IUPred (Dosztanyi i sur., 2005). Za razliku od prethodne skupine, ovakvi procjenitelji su neovisni o skupovima podataka za strojno učenje koji ne sadrže ravnomjerne količine svih vrsta IDP, a posebno većinski neuređenih IDP. Stoga se IUPred i sl. smatraju najboljima za identifikaciju većinski neuređenih proteina ili dugih neuređenih regija (Dosztanyi i sur., 2009). Važno je naglasiti da su procjenitelji neuređenosti uključeni u proces nepristrane evaluacije kvalitete softvera za predviđanje strukture proteina, tj. CASP (Critical Assessment of Structure Prediction, <http://www.predictioncenter.org/index.cgi>, 09.08.2016.). Prema rezultatima analiza provedenih u sklopu CASP10 trenutno najbolji procjenitelji neuređenosti su DISOPRED3, PrDOS i MFDp (Monastyrskyy i sur., 2013).

#### 4. EKSPERIMENTALNE METODE ISTRAŽIVANJA IDP

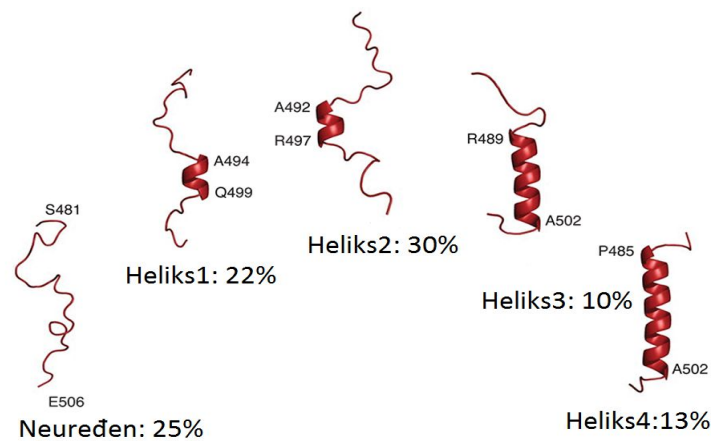
Metode iz Proteomike za izolaciju IDP temelje se na neobičnim svojstvima IDP kao što su otpornost na denaturaciju pri visokim temperaturama i jakim kiselinama (Habchi i sur., 2014; Tompa, 2010). U ovakvim uvjetima većina globularnih (uređenih) proteina denaturira i precipitira, dok IDP ostaju u otopini (Csizmok i sur., 2007; Tompa, 2010).

Csizmok i sur. (2006) razvili su metodu otkrivanja IDP koja uključuje toplinsku denaturaciju kao prvi korak u pročišćavanju IDP. Daljnje razdvajanje IDP i preostalih termostabilnih globularnih proteina postiže se dvodimenzionalnom poliakrilamidnom gel elektroforezom s 8M urejom. U prvoj dimenziji u nativnom gelu dolazi do razdvajanja proteina prema omjeru naboja i mase. U drugoj dimenziji zbog prisutnosti 8M ureje dolazi do denaturacije i usporavanja globularnih proteina u odnosu na prvu dimenziju, dok IDP ostaju većinom nepromijenjeni te jednako mobilni. Zbog toga se očekuje da će IDP, prešavši jednaku udaljenost u oba gela, biti pozicionirani na dijagonali denaturirajućeg gela (Slika 4). Usporeni globularni proteini smještaju se u područje iznad dijagonale. Identifikacija proteina iz gela vrši se metodom masene spektrometrije. Primjenom proteomskih metoda IDP su dosad uspješno identificirani u proteomima *S. cerevisiae*, *E. coli*, *D. melanogaster* i fibroblastima NIH3T3 vrste *M. musculus* (Galea i sur., 2009; Csizmok i sur., 2006; Szöllösi i sur., 2008). Sličan je princip korišten u identifikaciji fosforiliranih proteina iz sjemena *A. thaliana* (Irar i sur., 2006).



**Slika 4.** Razdvajanje i identifikacija IDP u uzorcima *S. cerevisiae* tretiranim toplinom. Proteini su identificirani masenom spektrometrijom. (Csizmok i sur., 2006).

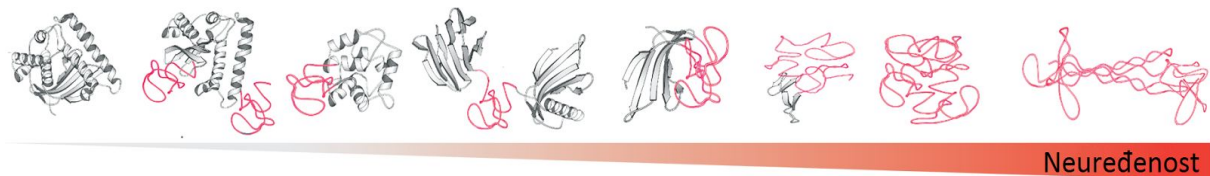
Mnoge metode strukturne biofizike primjenjive su u detekciji nedostatka jedinstvene stabilne strukture proteina. Najistaknutija metoda koja omogućava kvantitativni opis IDP je spektroskopija NMR. NMR pruža informacije o dinamici IDP te heterogenom skupu konformacija kojim se opisuje struktura IDP (Slika 5.). No, zbog ograničenja NMR kao što je visoka količina smetnji (*engl.* noise) naspram signala i uprosječnjavanja signala brzoizmjenjujućih događaja, nužno je upotpuniti analize spektroskopskim metodama, simulacijama molekularne dinamike i dr. metodama računalne biofizike (Habchi i sur., 2014; Burger i sur., 2014).



Slika 5. Skup konformacija dijela neuređene C-terminalne domene virusnog  $N_{TAIL}$  proteina prema mjerenjima NMR (Jensen i sur., 2011).

## 5. PODJELA IDP PREMA MEHANIZMU DJELOVANJA I STANIČNOJ ULOZI

IDP su isprva bili dijeljeni u diskretne klase po stupnju neuređenosti, no danas se smatra da postoji kontinuirani spektar proteinske strukture koji obuhvaća sve mogućnosti od većinskih IDP do većinski uređenih (globularnih) proteina (van der Lee i sur., 2014; Dyson i Wright, 2005). IDP najčešće dijelimo prema prema načinu djelovanja na molekularnoj razini (*engl.* molecular function) i ulozi u staničnim procesima (*engl.* biological process).



**Slika 6.** Mogućnosti neuređenosti proteinske strukture. Prilagođeno iz Habchi i sur., 2014.

Prema načinu djelovanja IDP svrstavamo u pet skupina (Habchi i sur., 2014). Efektori, pretraživači, sastavljači, predstavljači funkcionalnih mjesta (*engl.* effectors, scavengers, assemblers, display sites) posjeduju funkcije koje se zasnivaju na molekularnom prepoznavanju i vezanju IDP i liganda pri čemu vezanje liganda izaziva ispravno smatanje IDP (*engl.* folding upon binding) (van der Lee i sur., 2014). No, neki IDP ostaju neuređeni i nakon vezanja, tvoreći tzv. “fuzzy complexes,” kao u slučaju vezanja transkripcijskog faktora CREB na koaktivator CBP (Fuxreiter i Tompa, 2012; Tompa, 2010; van der Lee i sur., 2014). Primjer “effector” proteina je p21<sup>Cip1</sup>, inhibitor proteina Cdk2 (Habchi i sur., 2014). Funkcija pete skupine, tj. dinamičnih veznih elemenata (*engl.* flexible linkers, entropic chains), ovisi isključivo o fleksibilnosti polipeptidnog lanca, a ne o interakcijama s ligandima (Habchi i sur., 2014).

IDP su najviše uključeni u staničnu regulaciju i signalizaciju (Wright i Dyson, 2014). Iakoucheva i sur. (2002) navode da 70% svih signalnih proteina sadrži IDR. Također, IDP sudjeluju u kontroli staničnog ciklusa te imaju šaperonske uloge (Tompa, 2012; Pietrosemoli i sur., 2013; Kovacs i Tompa, 2012). Uloga IDP u složenim mrežama proteinskih interakcija kao što su signalne kaskade i mreže transkripcijskih faktora objašnjava se prednostima koje potječu od neuređenosti, tj. konformacijske fleksibilnosti (Pazos i sur., 2013). One uključuju

veću brzinu interakcija, kratkotrajno visokospecifično vezanje više partnera s niskim afinitetom te veće interakcijske površine u kompleksima (Uversky, 2013).

Funkcionalna mjesta u IDP koja određuju specifičnost vezanja su oznake molekularnog prepoznavanja (*engl.* MoRF, molecular recognition features) i kratki linearni motivi (SLiM) (Pietrosemoli i sur., 2013). Funkcija IDP često je regulirana posttranslacijskim modifikacijama (van der Lee i sur., 2014).

IDP u tkivu modelnog organizma, biljke uročnjak (*Arabidopsis thaliana*) nisu eksperimentalno proučavani na razini proteoma. Irar i sur. (2006) koristili su metodu pročišćavanja proteina na bazi termostabilnosti za identifikaciju fosforiliranih LEA proteina u sjemenkama *A. thaliana*, no cilj tog istraživanja nije bila identifikacija IDP. Prema računalnoj studiji Ward i sur. (2004a) *A. thaliana* sadrži 33.8% LDR od 30 ili više uzastopnih aminokiselina. Nedavna računalna genomska analiza otkrila je prve pojedinosti o ulogama i sadržaju IDP u *A. thaliana* (Pietrosemoli i sur., 2013). Korištenjem procjenitelja neuređenosti DISOPRED pokazano je da čak 57.2% proteina u *A. thaliana* sadrži LDR.

Proteini koji su prema bioinformatičkim procjenama klasificirani kao IDP pokazali su značajnu korelaciju sa biološkim procesima unutar četiri široke kategorije: signalizacija, stanični ciklus, razvoj i odgovor na stres. U analizi bioloških procesa u koje su proteini uključeni koristila se klasifikacija prema pojmovima iz kontroliranog rječnika Gene Ontology (*engl.* Gene Ontology terms, <http://geneontology.org/>). Gene Ontology (GO) je centralni izvor bioloških informacija nastao iz potrebe za konzistentnim načinom opisivanja genskih produkata (Ashburner i sur., 2000). GO sadrži tri ontologije, tj. kontrolirana rječnika kojima se sustavno opisuju tri glavna aspekta genskih produkata: stanična lokalizacija, molekularna funkcija i biološki proces u koji su uključeni (*engl.* cellular component (CC), molecular function (MF), biological process (BP)). Uz razvijene ontologije, GO na osnovi ovih ontologija također sadrži opširne anotacije genskih produkata iz mnogih baza podataka.

## 6. IDP U BILJKAMA PRI NORMALNIM I STRESNIM UVJETIMA

Biljke odlikuje visok stupanj fenotipske plastičnosti, tj. sposobnosti za prilagodbu na promjenjive okolišne uvjete. Smatra se da IDP u biljkama pridonose fenotipskoj plastičnosti jer omogućavaju odvijanje vrlo kompleksnih proteinskih interakcija unutar okolišnog odgovora (Pazos i sur., 2013). IDP se u biljaka dijele u 5 velikih skupina: LEA proteini, transkripcijski faktori NAC i bZIP, te proteini obitelji GRAS i CRY (*Cryptochrome*) (Sun i sur., 2013, Pazos i sur., 2013).

Unatoč brojnim bioinformatički predviđenim biljnim IDP, rad na eksperimentalnoj potvrdi je u zaostatku i potrebna su daljnja istraživanja (Marin i Ott, 2014.)

LEA proteini i transkripcijski faktori NAC i bZIP su tri glavne skupine IDP koje pomažu biljkama da podnesu razne vrste abiotskog stresa (Sun i sur., 2013). NAC proteini su mnogobrojna skupina biljnih proteina koji sudjeluju u razvoju, senescenciji, obrani i odgovoru na abiotski stres (Ricachenevsky i sur., 2013, Jensen i sur., 2010). Ovi proteini imaju kraću, uređenu N-terminalnu domenu za vezanje DNA i dulju neuređenu C-terminalnu domenu varijabilnije sekvence s čestom pojavom aminokiselina Pro, Ser i Thr. Ova transkripcijska regulatorna domena veže nekoliko različitih partnera pri čemu se lokalno smata u strukturu alfa-zavojnice. Transkripcijski faktori bZIP također sadrže DNA-vezujuću i neuređenu domenu, a reguliraju gene cirkadijskog ritma, razvoja i odgovora na stres i zračenje (Sun i sur., 2013.).

LEA proteini su najproučavaniji i najbrojniji biljni IDP (Marin i Ott, 2014). Ekspimiraju se u kasnim fazama razvoja sjemenke i konstitutivno u meristemskim tkivima. LEA pomažu u toleranciji biljke na abiotski stres izazvan sušom i niskim temperaturama (Hincha i Thalhammer, 2012, Kovacs i sur., 2008, Candat i sur., 2014). Preko 700 LEA podijeljeno je u tri grupe, a najveća grupa su dehidrini (Sun i sur., 2013). Osim u biljkama, homolozi LEA proteina nalaze se u nekih mikroorganizama i beskralješnjaka (Hincha i Thalhammer, 2012). Klasifikacija LEA nalazi se u LEAPdb bazi podataka (<http://forge.info.univ-angers.fr/~gh/Leaddb/index.php>, 25.4.2015., Hunault i sur., 2010). Mehanizmi djelovanja LEA još su uvelike nerazjašnjeni. Primjerice, LEA proteini COR15A i COR15B u *A. thaliana* se pri uvjetima suše iz neuređenog stanja smataju u alfa-zavojnicu, no način djelovanja nije im poznat (Marin i Ott, 2014.). Prema istraživanjima u uvjetima *in vitro*, uloge LEA uključuju šaperonske aktivnosti, vezanje metalnih iona, vode i vezikula te

stabilizaciju membrana i proteina u stresnim uvjetima (Boucher i sur., 2010, Sun i sur., 2013, Tolleter i sur., 2010).

GRAS proteini imaju ulogu u rastu i razvoju biljaka te sudjeluju u provođenju signala. Oni djeluju kao integratori signala, primjerice onih od različitih hormona. Integraciju omogućava njihova neuređena N-terminalna regija koja interagira s nekoliko veznih partnera uz pomoć oznaka molekularnog prepoznavanja, tj. MoRF. C-terminalna regija je uređena i evolucijski visoko očuvana (Pazos i sur., 2013, Sun i sur., 2012). CRY proteini su flavoproteini koji imaju esencijalnu ulogu u svjetlosnoj signalizaciji te utječu na otvaranje supki i produkciju antocijanina u biljaka. Pokazano je da CRY1 protein interagira s negativnim regulatorom fotomorfogeneze, COP1 (*constitutive photomorphogenic 1*) (Yang i sur., 2001). Eksperimentalno je dokazano da su rekombinantne C-terminalne domene biljnih CRY proteina neuređene (Sun i sur., 2013).

## **7. ZAKLJUČAK**

Dobivanje podataka o strukturi i funkciji IDP i danas predstavlja izazov zbog njihove iznimne strukturne mobilnosti. No, napredak u razvoju bioinformatičkih i eksperimentalnih metoda u protekla dva desetljeća omogućio je značajan uvid u posebnosti aminokiselinskog slijeda IDP, njihovu strukturnu heterogenost, rasprostranjenost u organizmima, mnoge aspekte njihovih staničnih funkcija i lokalizacije, te potencijalne primjene u biomedicini i biotehnologiji. Iako je postignut velik napredak, računalne studije ukazuju na veliku brojnost i sveprisutnost IDP te ostavljaju prostor za mnoga buduća istraživanja.

## **8. ZAHVALE**

*Najveće hvala doc. dr. sc. Dubravku Pavokoviću na višegodišnjem mentorstvu te svom uloženom trudu, vremenu i strpljenju.*



## 9. LITERATURA

- Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112(3), pp.535-542.
- Burger, V., Gurry, T. and Stultz, C. (2014). Intrinsically Disordered Proteins: Where Computation Meets Experiment. *Polymers*, 6(10), pp.2684-2719.
- Csizmok, V., Dosztanyi, Z., Simon, I. and Tompa, P. (2007). Towards Proteomic Approaches for the Identification of Structural Disorder. *CPPS*, 8(2), pp.173-179.
- Csizmok, V., Szöllösi, E., Friedrich, P. and Tompa, P. (2006). A Novel Two-dimensional Electrophoresis Technique for the Identification of Intrinsically Unstructured Proteins. *Molecular & Cellular Proteomics*, 5(2), pp.265-273.
- Dosztanyi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21(16), pp.3433-3434.
- Dosztanyi, Z., Meszaros, B. and Simon, I. (2009). Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. *Briefings in Bioinformatics*, 11(2), pp.225-243.
- Dyson, H. and Wright, P. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3), pp.197-208.
- Dyson, H. Jane. "Making Sense Of Intrinsically Disordered Proteins". *Biophysical Journal* 110.5 (2016): 1013-1016. Web. 10 Aug. 2016.
- Fukuchi, Satoshi et al. "IDEAL In 2014 Illustrates Interaction Networks Composed Of Intrinsically Disordered Proteins And Their Binding Partners". *Nucleic Acids Research* 42.D1 (2013): D320-D325. Web. 10 Aug. 2016.
- Fuxreiter, M. and Tompa, P. (2012). Fuzzy Complexes: A More Stochastic View of Protein Function. *Fuzziness*, pp.1-14.
- Galea, C., High, A., Obenauer, J., Mishra, A., Park, C., Punta, M., Schlessinger, A., Ma, J., Rost, B., Slaughter, C. and Kriwacki, R. (2009). Large-Scale Analysis of Thermostable, Mammalian Proteins Provides Insights into the Intrinsically Disordered Proteome. *J. Proteome Res.*, 8(1), pp.211-226.
- Gall, Tanguy Le et al. "Intrinsic Disorder In The Protein Data Bank". *Journal of Biomolecular Structure and Dynamics* 24.4 (2007): 325-341. Web.
- Habchi, Johnny et al. "Introducing Protein Intrinsic Disorder". *Chemical Reviews* 114.13 (2014): 6561-6588. Web.
- He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. and Dunker, A. (2009). Predicting intrinsic disorder in proteins: an overview. *Cell Res*, 19(8), pp.929-949.
- Hincha, D. and Thalhammer, A. (2012). LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochemical Society Transactions*, 40(5), pp.1000-1003.
- Hunault, G. and Jaspard, E. (2010). LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics*, 11(1), p.221.
- Iakoucheva, L., Brown, C., Lawson, J., Obradović, Z. and Dunker, A. (2002). Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *Journal of Molecular Biology*, 323(3), pp.573-584.

- Irar, S., Oliveira, E., Pagès, M. and Goday, A. (2006). Towards the identification of late-embryogenic-abundant phosphoproteome in Arabidopsis by 2-DE and MS. *PROTEOMICS*, 6(S1), pp.S175-S185.
- Iqbal, Sumaiya and Md Tamjidul Hoque. "Dispredict: A Predictor Of Disordered Protein Using Optimized RBF Kernel". *PLOS ONE* 10.10 (2015): e0141551. Web
- Jensen, M., Kjaersgaard, T., Nielsen, M., Galberg, P., Petersen, K., O'Shea, C. and Skriver, K. (2010). The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. *Biochem. J.*, 426(2), pp.183-196.
- Lieutaud, Philippe, François Ferron, and Sonia Longhi. "Predicting Conformational Disorder". *Methods in Molecular Biology* (2016): 265-299. Web. 10 Aug. 2016.
- Linding, R. (2003). GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Research*, 31(13), pp.3701-3708.
- Marín, M. and Ott, T. (2014). Intrinsic Disorder in Plant Proteins and Phytopathogenic Bacterial Effectors. *Chem. Rev.*, 114(13), pp.6912-6932.
- Meng, Fanchi and Lukasz Kurgan. "DFLpred: High-Throughput Prediction Of Disordered Flexible Linker Regions In Protein Sequences". *Bioinformatics* 32.12 (2016): i341-i350. Web. 10 Aug. 2016.
- Midic, Uros et al. "Unfoldomics Of Human Genetic Diseases: Illustrative Examples Of Ordered And Intrinsically Disordered Members Of The Human Diseaseome". *Protein & Peptide Letters* 16.12 (2009): 1533-1547. Web.
- Monastyrskyy, B., Kryshchak, A., Moulton, J., Tramontano, A. and Fidelis, K. (2013). Assessment of protein disorder region predictions in CASP10. *Proteins*, 82, pp.127-137.
- Nelson, David L, Michael M Cox, and Albert L Lehninger. *Lehninger Principles Of Biochemistry*. New York: W.H. Freeman and Company, 2013. Print.
- Oates, M. E. et al. "D2P2: Database Of Disordered Protein Predictions". *Nucleic Acids Research* 41.D1 (2012): D508-D516. Web. 10 Aug. 2016.
- Pazos, F., Pietrosemoli, N., García-Martín, J. and Solano, R. (2013). Protein intrinsic disorder in plants. *Front. Plant Sci.*, 4.
- Peng, K., Radivojac, P., Vucetic, S., Dunker, A. and Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, 7(1), p.208.
- Peng, Z., Yan, J., Fan, X., Mizianty, M., Xue, B., Wang, K., Hu, G., Uversky, V. and Kurgan, L. (2014). Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cellular and Molecular Life Sciences*, 72(1), pp.137-151.
- Pietrosemoli, N., García-Martín, J., Solano, R. and Pazos, F. (2013). Genome-Wide Analysis of Protein Disorder in Arabidopsis thaliana: Implications for Plant Environmental Adaptation. *PLoS ONE*, 8(2), p.e55524.
- Potenza, E. et al. "Mobidb 2.0: An Improved Database Of Intrinsically Disordered And Mobile Proteins". *Nucleic Acids Research* 43.D1 (2014): D315-D320. Web. 10 Aug. 2016.
- Prilusky, J., Felder, C., Zeev-Ben-Mordehai, T., Rydberg, E., Man, O., Beckmann, J., Silman, I. and Sussman, J. (2005). FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, 21(16), pp.3435-3438.
- Ricachenevsky, F., Menguer, P. and Sperotto, R. (2013). kNACKing on heaven's door: how important are NAC transcription factors for leaf senescence and Fe/Zn remobilization to seeds?. *Front. Plant Sci.*, 4.

- Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. and Dunker, A. (1997). Identifying disordered regions in proteins from amino acid sequence. *Proceedings of International Conference on Neural Networks (ICNN'97)*.
- Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E., Guilliot, S., Garner, E., and Dunker, A.K. (1998) Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing* 3:435-446.
- Romero, P., Obradovic, Z., Li, X., Garner, E., Brown, C. and Dunker, A. (2000). Sequence complexity of disordered protein. *Proteins: Structure, Function, and Genetics*, 42(1), pp.38-48.
- Sickmeier, M., Hamilton, J., LeGall, T., Vacic, V., Cortese, M., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V., Obradovic, Z. and Dunker, A. (2007). DisProt: the Database of Disordered Proteins. *Nucleic Acids Research*, 35(Database), pp.D786-D793.
- Sun, X., Jones, W. and Rikkerink, E. (2012). GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. *Biochem. J.*, 442(1), pp.1-12.
- Sun, X., Rikkerink, E., Jones, W. and Uversky, V. (2013). Multifarious Roles of Intrinsic Disorder in Proteins Illustrate Its Broad Impact on Plant Biology. *The Plant Cell*, 25(1), pp.38-55.
- Szöllösi, E., Bokor, M., Bodor, A., Perczel, A., Klement, E., Medzihradzsky, K., Tompa, K. and Tompa, P. (2008). Intrinsic Structural Disorder of DF31, a Drosophila Protein of Chromatin Decondensation and Remodeling Activities. *J. Proteome Res.*, 7(6), pp.2291-2299.
- Tolletier, D., Hinch, D. and Macherel, D. (2010). A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1798(10), pp.1926-1933.
- Tompa, P. (2010). *Structure and function of intrinsically disordered proteins*. Boca Raton: Chapman & Hall/CRC Press.
- Tompa, Peter. "Intrinsically Disordered Proteins: A 10-Year Recap". *Trends in Biochemical Sciences* 37.12 (2012): 509-516. Web.
- Uversky, V. (2012). Intrinsically disordered proteins and novel strategies for drug discovery. *Expert Opin. Drug Discov.*, 7(6), pp.475-488.
- Uversky, V. (2013). A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science*, 22(6), pp.693-724.
- Uversky, V. (2014a). Introduction to Intrinsically Disordered Proteins (IDPs). *Chem. Rev.*, 114(13), pp.6557-6560.
- Uversky, V. (2014b). Intrinsically Disordered Proteins. *SpringerBriefs in Molecular Science*.
- Uversky, V. (2014d). Proteins without unique 3D structures: Biotechnological applications of intrinsically unstable/disordered proteins. *Biotechnology Journal*, 10(3), pp.356-366.
- Uversky, V. (2014e). The triple power of D<sup>3</sup>: Protein intrinsic disorder in degenerative diseases. *Front Biosci*, 19(2), p.181.
- Uversky, V., Davé, V., Iakoucheva, L., Malaney, P., Metallo, S., Pathak, R. and Joerger, A. (2014c). Pathological Unfoldomics of Uncontrolled Chaos: Intrinsically Disordered Proteins and Human Diseases. *Chem. Rev.*, 114(13), pp.6844-6879.
- Uversky, V., Gillespie, J. and Fink, A. (2000). Why are natively unfolded proteins unstructured under physiologic conditions?. *Proteins: Structure, Function, and Genetics*, 41(3), pp.415-427.

Uversky, V., Oldfield, C. and Dunker, A. (2008). Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept. *Annual Review of Biophysics*, 37(1), pp.215-246.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R., Daughdrill, G., Dunker, A., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D., Kim, P., Kriwacki, R., Oldfield, C., Pappu, R., Tompa, P., Uversky, V., Wright, P. and Babu, M. (2014). Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.*, 114(13), pp.6589-6631.

Vincent, Michael, Mark Whidden, and Santiago Schnell. "Quantitative Proteome-Based Guidelines For Intrinsic Disorder Characterization". *Biophysical Chemistry* 213 (2016): 6-16. Web. 10 Aug. 2016.

Wang, Zhiheng et al. "Disomcs: Accurately Predicting Protein Intrinsically Disordered Regions Using A Multi-Class Conservative Score Approach". *PLOS ONE* 10.6 (2015): e0128334. Web.

Ward, J., McGuffin, L., Bryson, K., Buxton, B. and Jones, D. (2004b). The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, 20(13), pp.2138-2139.

Ward, J., Sodhi, J., McGuffin, L., Buxton, B. and Jones, D. (2004a). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *Journal of Molecular Biology*, 337(3), pp.635-645.

Wells, M. et al. "Structure Of Tumor Suppressor P53 And Its Intrinsically Disordered N-Terminal Transactivation Domain". *Proceedings of the National Academy of Sciences* 105.15 (2008): 5762-5767. Web.

Xue, B., Dunker, A. and Uversky, V. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure and Dynamics*, 30(2), pp.137-149.

Yang, H. (2001). The Signaling Mechanism of Arabidopsis CRY1 Involves Direct Interaction with COP1. *THE PLANT CELL ONLINE*, 13(12), pp.2573-2587.

Ziegler, Zachary et al. "Mollack: A Web Server For The Automated Creation Of Conformational Ensembles For Intrinsically Disordered Proteins". *Bioinformatics* 32.16 (2016): 2545-2547. Web.

## 10. SAŽETAK

Nativno neuređeni proteini i proteinske regije (*engl.* intrinsically disordered proteins/regions, IDP/IDR) pripadaju nedavno priznatoj skupini proteina koji su biološki aktivni usprkos nedostatku jasno definirane trodimenzionalne strukture. Za razliku od globularnih proteina, IDP u nativnom stanju ne posjeduju jedinstvenu stabilnu strukturu, već fluktuirajući skup strukturnih konformacija pogodan u staničnoj signalizaciji, regulaciji, šaperonskim aktivnostima, patogenezi bolesti i drugim raznovrsnim funkcijama. Cilj ovog seminara bio je opisati osnovne značajke IDP te najvažnije eksperimentalne i bioinformatičke metode istraživanja IDP. Nadalje, rad obuhvaća podjelu IDP prema mehanizmu djelovanja i staničnoj ulozi te kratki osvrt na IDP u biljkama pri normalnim i stresnim uvjetima. U zaključku, područje IDP svjedočilo je strelovitom rastu u protekla dva desetljeća te se na temelju brojnosti i rasprostranjenosti ovih proteina predviđa velik potencijal za buduća istraživanja unutar ovog područja.

## 11. SUMMARY

Intrinsically disordered proteins and protein regions (IDP/IDR) are a recently recognized group of proteins which are biologically active despite their inherent lack of a well-defined three-dimensional structure. Unlike globular proteins, IDPs lack a unique and stable 3D structure in their native state and exist as fluctuating ensembles of conformations which are well-suited for cellular signaling, regulation, chaperone activity, disease pathogenesis and other various functions. The aim of this Bachelor's thesis is to describe the main characteristics of IDPs and the most important experimental and bioinformatic methods used in IDP research. Moreover, the thesis includes a classification of IDPs according to functional mechanism and cellular role, as well as a brief section on IDPs in plants in normal and stressful conditions. In conclusion, the IDP field has witnessed rapid growth in the past two decades and, based on the abundance and wide spread of these proteins, there is significant potential for future research within this field.