Određivanje mikroorganizama pomoću MS/MS analize proteoma

Ujević, Sara

Master's thesis / Diplomski rad

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Food Technology and Biotechnology / Sveučilište u Zagrebu, Prehrambeno-biotehnološki fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:159:992242

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

prehrambeno biotehnološki fakultet

Repository / Repozitorij:

Repository of the Faculty of Food Technology and Biotechnology

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU
PREHRAMBENO-BIOTEHNOLOŠKI FAKULTET

DIPLOMSKI RAD

Zagreb, rujan 2016.

Sara Ujević
580/MB

ODREĐIVANJE MIKROORGANIZAMA POMOĆU MS/MS ANALIZE PROTEOMA

Rad je izrađen u Laboratoriju za sistemsku biomedicinu na Zavodu za molekularnu medicinu, Instituta Ruđer Bošković pod nadzorom dr. sc. Maria Cindrića i u Kabinetu za bioinformatiku na Zavodu za biokemijsko inženjerstvo, Prehrambeno-biotehnološkog fakulteta Sveučilišta u Zagrebu pod mentorstvom dr.sc. Antonia Starčevića, izv.prof. Prehrambeno-biotehnološkog fakulteta Sveučilišta u Zagrebu te uz pomoć pri izradi dr.sc. Damira Orosa

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu **Diplomski rad**
Prehrambeno-biotehnološki fakultet
Zavod za biokemijsko inženjerstvo
Kabinet za bioinformatiku

Znanstveno područje: Biotehničke znanosti
Znanstveno polje: Biotehnologija

ODREĐIVANJE MIKROORGANIZAMA POMOĆU MS/MS ANALIZE PROTEOMA

Sara Ujević, 580/MB

Sažetak: „Shotgun" tehnika u proteomici postala je široko primjenjiva. Zasniva se na digestiji proteina, sekvenciranju dobivenih peptida tandemnom spektrometrijom masa nakon čega slijedi identifikacija proteina pretragom baze podataka. Tehnike tandemne spektrometrije masa postaju sve privlačnije za identifikaciju mikroorganizama zahvaljujući njihovoj brzini i niskoj cijeni. Razvojem tehnološki naprednijih uređaja za spektrometriju masa postignuta je veća preciznost mjerenja čime se povećala količina dobivenih podataka. Unatoč tome, bioinformatička analiza tih podataka i dalje predstavlja veliki problem zbog nedostatka prikladnih bioinformatičkih alata. U ovom radu uspoređeni su bioinformatički alati, ProteinLynx Global Server (PLGS) i novorazvijeni alat koji se zasniva na LSI metodi, s ciljem određivanja vrste odnosno soja mikroorganizma. Rezultati su pokazali da je novorazvijeni bioinformatički alat daleko učinkovitiji te za razliku od PLGS-a ne zahtijeva sužavanje područja pretrage i ne ovisi u tolikoj mjeri o preciznosti snimanja samog uređaja. Ovim alatom uspješno je određen analizirani soj kvasca *Candida albicans* SC5314.

Ključne riječi: *proteomika, digestija, tandemna spektrometrija masa, bioinformatički algoritam, LSI*

Rad sadrži: 49 stranica, 11 slika, 9 tablica, 86 literaturna navoda

Jezik izvornika: hrvatski

Rad je u tiskanom i elektroničkom (pdf format) obliku pohranjen u: Knjižnica Prehrambeno-biotehnološkog fakulteta, Kačićeva 23, Zagreb

Mentor: *Izv.prof.dr.sc. Antonio Starčević*

Pomoć pri izradi: *dr.sc. Damir Oros*

Stručno povjerenstvo za ocjenu i obranu:
1. Prof.dr.sc. *Vladimir Mrša*
2. Izv.prof.dr.sc. *Antonio Starčević*
3. Izv.prof.dr.sc. *Damir Stanzer*
4. Doc.dr.sc. *Jurica Žučko* (zamjena)

Datum obrane: 26. rujna 2016. godine

BASIC DOCUMENTATION CARD

University of Zagreb **Graduate Thesis**
Faculty of Food Technology and Biotechnology
Department of Biochemical Engineering
Section for Bioinformatics

Scientific area: Biotechnical Sciences
Scientific field: Biotechnology

MICROBIAL DETECTION AND IDENTIFICATION USING MASS SPECTROMETRY PROTEOME ANALYSIS

Sara Ujević, 580/MB

Abstract: The shotgun proteomics strategy has become widely adopted. It is based on digesting proteins into peptides followed by peptide sequencing using tandem mass spectrometry (MS/MS) and automated database searching. Tandem mass spectrometry methods are becoming increasingly attractive for identifying microorganisms, mainly due to their speed and low cost. Drastic improvements in accuracy of mass spectrometry devices have led to an increasment of data being generated. Nevertheless, structuring such information is still hard due to lack of appropriate bioinformatics tools. In this work two bioinformatics tools, ProteinLynx Global Server (PLGS) and novel bioinformatic algorithm based on LSI, have been compared in order to identify microbial species i.e. strains. Unlike PLGS, the results showed that this novel algorithm proved to be more efficient because it is not limited by database search area and also not constrained by accuracy of the mass spectrometry devices. *Candida albicans* SC5314 has been successfully detected and identified by this novel algorithm.

Keywords: *proteomics, digestion, tandem mass spectrometry, bioinformatic algorithm, LSI*

Thesis contains: 49 pages, 11 figures, 9 tables, 86 references

Original in: Croatian

Graduate Thesis in printed and electronic (pdf format**) version is deposited in:** Library of the Faculty of Food Technology and Biotechnology, Kačićeva 23, Zagreb

Mentor: *Antonio Starčević, Associate professor, PhD.*

Technical support and assistance: *Damir Oros, PhD.*

Reviewers:
1. PhD. *Vladimir Mrša*, Full professor
2. PhD. *Antonio Starčević*, Associate professor
3. PhD. *Damir Stanzer*, Associate professor
4. PhD. *Jurica Žučko,* Assistant professor (substitute)

Thesis defended: 26th September 2016

Sadržaj

1. UVOD **1**
2. TEORIJSKI DIO **3**
2.1. PROTEOMIKA 3
2.2. TEKUĆINSKA KROMATOGRAFIJA 4
2.2.1. Tekućinska kromatografija visoke djelotvornosti 5
2.2.2. Tekućinska kromatografija ultra visoke djelotvornosti 6
2.3. SPEKTROMETRIJA MASA 6
2.3.1. Ionski izvori 7
2.3.2. Analizatori 9
2.3.3. Tandemna spektrometrija masa 9
2.3.4. MS^E 11
2.4. PRETRAGA BAZE PODATAKA 12
2.4.1. Parametri prilikom pretrage baze podataka 14
2.4.2. ProteinLynx Global Server 15
2.4.3. Proteinske baze podataka 16
3. EKSPERIMENTALNI DIO **18**
3.1. MATERIJAL 18
3.1.1. Kemikalije 18
3.1.2. Laboratorijski pribor 18
3.1.3. Tehnička oprema 19
3.1.4. Osnovne otopine 19
3.1.5. Ispitivani uzorak 20
3.1.6. Programi za upravljanje instrumentima, obradu podataka i programski alati 20
3.2. METODE 21
3.2.1. Razbijanje kvaščevih stanica i izolacija proteina 21
3.2.2. Određivanje koncentracije proteina metodom po Bradford 21
3.2.3. Digestija proteina 22
3.2.4. Analiza peptida UPLC-MS^E pristupom 22
3.2.5. Bioinformatička analiza MS/MS spektara novorazvijenom metodom baziranom na LSI 25
3.2.6. Genska ontologija 26
4. REZULTATI I RASPRAVA **28**
4.1. ODREĐIVANJE VRSTE 29
4.1.1. Određivanje vrste mikroorganizma korištenjem programa ProteinLynx Global Server 29
4.1.2. Određivanje vrste mikroorganizma novorazvijenom metodom baziranom na LSI 31
4.2. ANALIZA PROTEOMA 33
4.3. ANALIZA GENSKE ONTOLOGIJE 37
4.4. DOPRINOS I OČEKIVANA PRIMJENA ISTRAŽIVANJA 39

5. ZAKLJUČCI 41
6. LITERATURA 42
7. PRILOZI

1. UVOD

Posljednjih nekoliko godina analiza proteina spektrometrijom masa načinom odozdol nagore (*engl.* bottom-up) postaje metoda od izbora za identifikaciju proteina u proteomskim istraživanjima širokih razmjera (Washburn i sur., 2001). Takav pristup još se naziva i „shotgun" proteomika analogno „shotgun" tehnici sekvenciranja genoma. Prvi korak je digestija proteina u peptide specifičnim proteazama, najčešće tripsinom, čije je cijepanje uniformno i predvidljivo. Dobivena smjesa tako nastalih peptida može varirati od nekoliko stotina (za protein) do nekoliko tisuća peptida (za proteom), zbog čega se peptidi razdvajaju tehnikama tekućinske kromatografije nakon čega se analiziraju tandemnom spektrometrijom masa koja daje podatke o masi tripsinom dobivenih fragmenata kao i cjelovito ili češće parcijalno očitanje aminokiselinskog slijeda u fragmentu (Link i sur., 1999). Uz pomoć ovih informacija moguće je usporediti očitane fragmente s *in silico* translatiranim genima iz javno dostupnih baza podataka i na posredan način potvrditi vrstu pa čak i soj mikroorganizma čiji se proteom analizira. Upravo zbog ove činjenice, spektrometrija masa odnedavno se sve učestalije rabi u kliničkoj praksi.

Usprkos tome što napretci u području masene spektrometrije čine generiranje rezultata relativno jeftinim i brzim, bioinformatička analiza dobivenih rezultata još uvijek predstavlja problem. Postojeće metode oslanjaju se na usporedbu spektara odnosno očitanih peptida s proteinskim sekvencama pohranjenim u javno dostupnim bazama podataka što je dugotrajan i računalno vrlo zahtjevan postupak. Ovisno o veličini pretraživane baze podataka, upravo je gore spomenuta usporedba ograničavajući faktor u analizi proteoma pomoću spektrometrije masa. Svi dostupni bioinformatički alati pritom se oslanjaju na poboljšanja same tehnologije snimanja spektara, pri čemu uređaji nove generacije postižu veću preciznost prilikom snimanja masa, ili ovise o sužavanju područja pretrage. Sužavanje područja pretrage limitira efikasnost metode, pošto je potrebno imati prethodne informacije o analiziranom uzorku.

Cilj ovog diplomskog rada jest određivanje vrste odnosno soja ispitivanog mikroorganizma. U tu svrhu će se usporediti jedan od vodećih bioinformatičkih alata, ProteinLynx Global Server s novorazvijenim alatom za analizu spektara peptidnih fragmenata koji se temelji na latentnom semantičkom indeksiranju (*engl.* latent semantic indexing) te će se ispitati njihova učinkovitost u detekciji i identifikaciji mikroorganizma. Pritom korišten radni mikroorganizam jest *Candida albicans* SC5314. Također će se provesti i analiza genske ontologije putem statističkih alata dostupnih u PANTHER platformi kako bi se odredila

značajna zastupljenost bioloških procesa i pripadajućih gena u stanicama kvasca u trenutku uzorkovanja.

2. TEORIJSKI DIO

2.1. PROTEOMIKA

Proteini su biološke molekule neophodne za odvijanje procesa unutar stanice. Cjelokupni komplet proteina nekog organizma eksprimiranih unutar pojedine stanice, tkiva ili organizma naziva se proteom, a znanstvena disciplina koja se bavi proučavanjem proteoma naziva se proteomika (Anderson i Anderson, 1998). Sam pojam, zajedno s ključnim inovacijama koje su je omogućile, poznati su nam već više od 10 godina (Wilkins i sur., 2006). Proteomika se vrlo brzo razvila u jedno od zanimljivijih područja proučavanja u prirodnim znanostima (Martens i sur., 2005). Za razliku od genoma koji ostaje konzerviran tijekom cijelog života stanice, proteom je dinamičan i složen sustav podložan kvalitativnim i kvantitativnim promjenama. Zbog toga proučavanje proteoma postaje nepohodno za razumijevanje bioloških sustava.

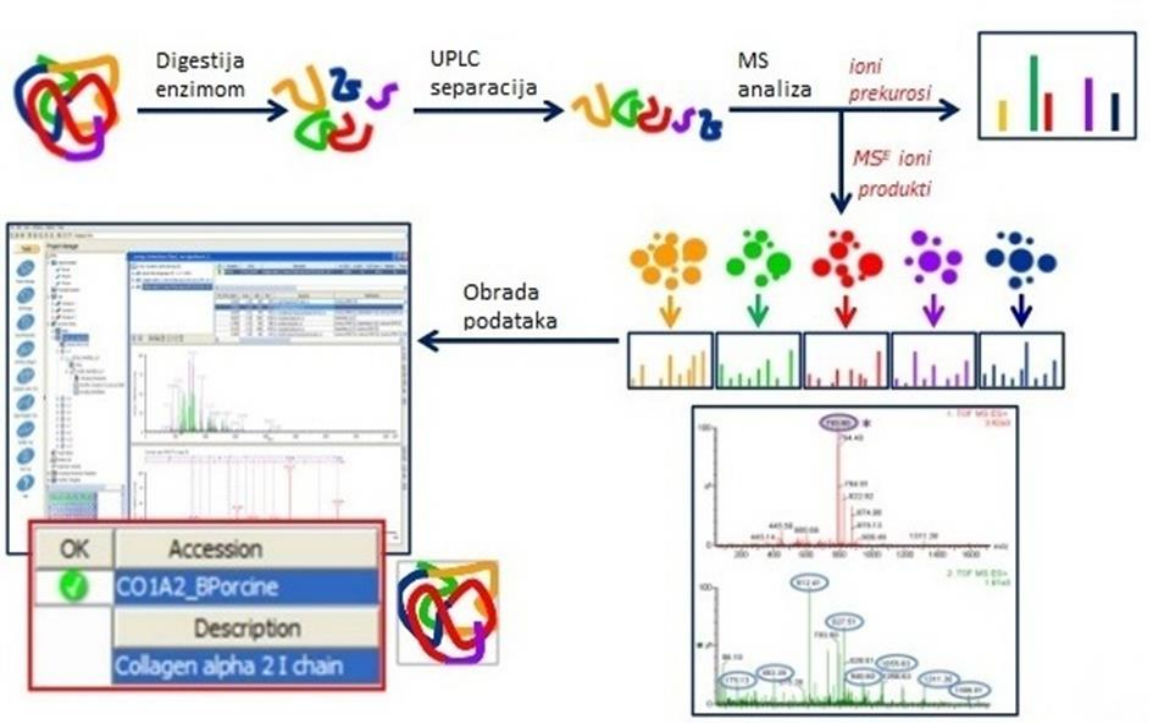
Proteom definira osnovnu razliku među stanicama i tkivima kao i odgovore stanice na okolišne čimbenike (Kinter i Sherman, 2005) te njenu izloženost različitim faktorima i stupnjevima stresa (Armengaud, 2013). Kao što se sekvence nukleotida razlikuju među pojedinim mikroorganizmima tako se razlikuju i sekvence aminokiselina odgovarajućeg proteina. Analiza proteina određenog mikroorganizma može se promatrati kao indirektna analiza kodirajućeg dijela genoma tog mikroorganizma (Karlsson i sur., 2015). Osnovne zadaće proteomike su: identifikacija i kvantifikacija svih proteina eksprimiranih u stanici ili tkivu (Aebersold i Mann, 2003), karakterizacija interaktoma tj. složene proteinske mreže, otkrivanje biomarkera za detekciju i dijagnozu bolesti te identifikacija potencijalnih meta za razvoj novih lijekova (Wilkins i sur., 2006).

Kvalitativnom analizom proteina određuje se primarni slijed aminokiselina nekog proteina na temelju kojeg je moguće odrediti gen koji kodira informaciju za sintezu tog proteina, a postupkom kvantitativne analize određuje se količina ekspresije pojedinog proteina ovisno o njegovoj funkciji unutar stanice. Uobičajeni postupak analize proteina uključuje slijedeće korake: izolaciju proteina iz biološkog materijala, separaciju kompleksnog sustava proteoma na što veći broj komponenti, analizu peptida spektrometrijom masa te završni korak koji uključuje računalnu obradu podataka.

Analiza proteina spektrometrijom masa može se provoditi na dva načina: odozgor nadolje i odozdol nagore (*engl.* top-down i bottom-up) (Bogdanov i Smith, 2005). Pristup odozgor nadolje temelji se na analizi intaktnih proteina s ciljem identifikacije proteina (Michalski i

sur., 2012), ispitivanja posttranslacijskih modifikacija (Tian i sur., 2012) i karakterizacije njihovih kompleksa (Zhang i sur., 2010).

Raširenији pristup analize proteina načinom odozdol nagore uključuje digestiju proteina specifičnim peptidazama, separaciju tekućinskom kromatografijom te analizu peptidnih fragmenata pomoću spektrometra masa (slika 1). Dobiveni podaci koriste se za određivanje detaljne strukture proteina i njegovih mogućih posttranslacijskih modifikacija (Galić i Cindrić, 2008).



Slika 1. Analiza proteina načinom odozdol nagore (Waters, 2012)

2.2. TEKUĆINSKA KROMATOGRAFIJA

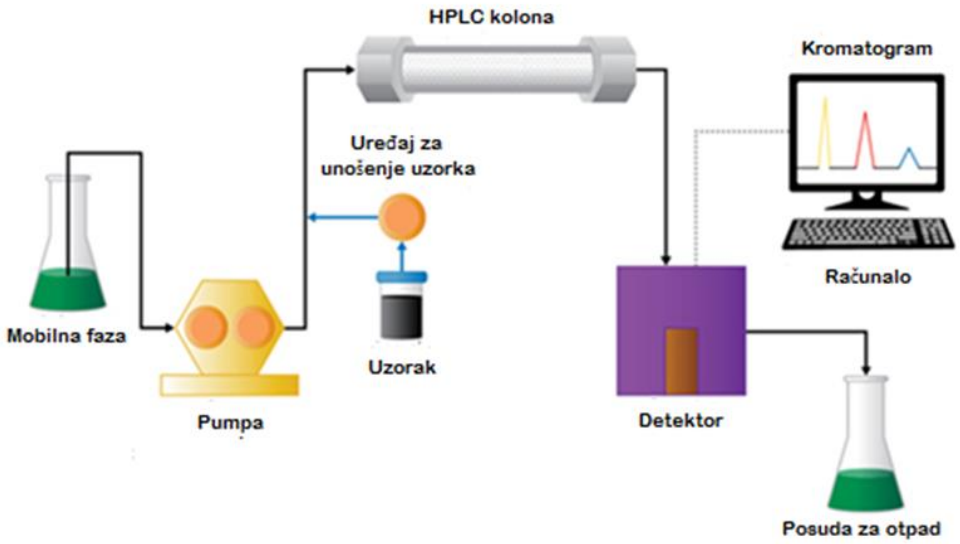
Tehnike tekućinske kromatografije (LC, engl. Liquide Chromatography) pogodne su za razdvajanje proteina i peptida nastalih enzimskom razgradnjom, jer su i proteini i peptidi topivi u vodi ili smjesi vode i pogodnog organskog otapala (npr. acetonitril). Princip na kojem se temelji razdvajanje peptida i proteina tekućinskom kromatografijom jest različita interakcija otopljenog analita s nepokretnom i tekućom pokretnom fazom. Razlika interakcije analita s nepokretnom fazom može nastati djelovanjem adsorpcije, ionske izmjene, razdiobe među fazama ili veličine analita koji se razdvajaju (Cindrić i sur., 2009). Uglavnom se koristi

tekućinska kromatografija obrnutih faza (*engl.* Reverse Phase, RP) koja razdvaja proteine prema njihovoj hidrofobnosti. Nepokretne faze koje se pritom koriste su nepolarni oktadecilsilicijev dioksid (C_{18}), oktilsilicijev dioksid (C_8) i butilsilicijev dioksid (C_4), dok je pokretna faza polarna i gradijentno se primjenjuje.

Tijekom posljednjih nekoliko godina za analizu proteina koristi se tekućinska kromatografija spregnuta sa spektrometrom masa (LC-MS/MS). Tekućinski kromatograf i spektrometar masa povezani su međuspojem (*engl.* Interface), koji ima višestruku ulogu: otparavanje otapala, ionizacija neutralnih molekula analita i uvođenje istih u analizator. Budući da smjesa peptida (nakon digestije proteina) može varirati od nekoliko stotina (za protein) do nekoliko tisuća peptida (za proteom) potreban je korak LC separacije kako bi se smanjila kompleksnost analiziranih peptida koji ulaze u ionizacijski izvor i uvelike poboljšala osjetljivost analize do biološki relevantnih koncentracija.

2.2.1. Tekućinska kromatografija visoke djelotvornosti

Najčešće korišteni tip tekućinske kromatografije u vezanom LC-MS sustavu jest tekućinska kromatografija visoke djelotvornosti (*engl.* High Performance Liquid Chromatography, HPLC). Tekućinska kromatografija visoke djelotvornosti razvijena je u kasnim šezdesetim godinama prošloga stoljeća. Komponente HPLC sustava jesu spremnik za pokretnu fazu, visokotlačna pumpa, sustav za unošenje uzorka (injektor), kolona, detektor, računalo (slika 2).



Slika 2. Osnovne komponente HPLC sustava (Waters, 2015)

Analiza na HPLC-u započinje injektiranjem male količine tekućeg uzorka u mobilnu fazu koja dalje prolazi kroz kolonu s punilom. Eluiranje ili separacija spojeva proizlazi iz različite pokretljivosti sastavnica uzorka, a brzina kojom se uzorak kreće kroz kolonu ovisi o vremenu zadržavanja u pokretnoj fazi. Vrijeme provedeno od trenutka injektiranja uzorka do trenutka izlaska iz kolone i detektiranja naziva se retencijsko vrijeme ili vrijeme zadržavanja (McPolin, 2009), a karakteristično je za svaku pojedinu komponentu i uvelike ovisi o nepokretnoj fazi i sastavu pokretne faze (Skoog i sur., 1999).

2.2.2. Tekućinska kromatografija ultra visoke djelotvornosti

Korporacija Waters predstavila je 2004. godine UPLC Acquity System, tekućinsku kromatografiju ultravisoke djelotvornosti (*engl.* Ultra Performance Liquid Chromatography, UPLC). Uređaj pruža nove mogućnosti u primjeni tekućinske kromatografije posebno po pitanju smanjenja vremena rada i potrošnje otapala (Waters, 2015). UPLC kromatografski sustav osmišljen je na način da može izdržati visoki sistemski povratni tlak (Novakova i sur., 2006).

Tekućinska kromatografija ultravisoke djelotvornosti omogućuje efikasnije razdvajanje komponenti (pojedinačnih ili u smjesi) u kraćem vremenskom periodu jer koristi manje čestice punila kolone (< 2 μm) i omogućuje uporabu viših tlakova (do 15 000 psi) u odnosu na tekućinsku kromatografiju visoke djelotvornosti (Klinčić i Herceg-Romanić, 2011). Nadalje, smanjen je i volumen injektiranja (2 μL) u odnosu na HPLC (5 μL) uz povećanu osjetljivost analize (Swartz, 2005). Stoga UPLC pokazuje bolje razlučivanje, kraće vrijeme analize i poboljšanu osjetljivost mjerenja u usporedbi s HPLC-om (Zhu i sur., 2008), dok je mehanizam razdvajanja ostao isti.

2.3. SPEKTROMETRIJA MASA

Spektrometrija masa (MS) je instrumentalna analitička tehnika koja mjeri omjer mase i naboja (*engl.* mass-to-charge ratio, m/z) prethodno ioniziranih atoma ili molekula. Često se koristi za identifikaciju proteina. Metoda počinje pripremom uzoraka peptida, slijedi ju bioinformatička obrada i na kraju analiza podataka. Osnovni dijelovi spektrometra masa su ionizator (ionizacijska komora), analizator masa i ionski detektor (Gross, 2011). U spektrometru masa uzorak se uvodi u izvor iona u tekućem ili plinovitom stanju gdje se on ionizira, a nastali ioni putuju u analizator masa gdje se razdvajaju prema njihovim omjerima

mase i naboja. Na kraju ioni odlaze u detektor koji registrira njihov broj pri svakoj vrijednosti m/z, što rezultira stvaranjem spektra masa (Vučinić, 2012). Mogućnosti izvedbe su brojne, a razlikuju ih način rada ionizatora, analizatora i detektora.

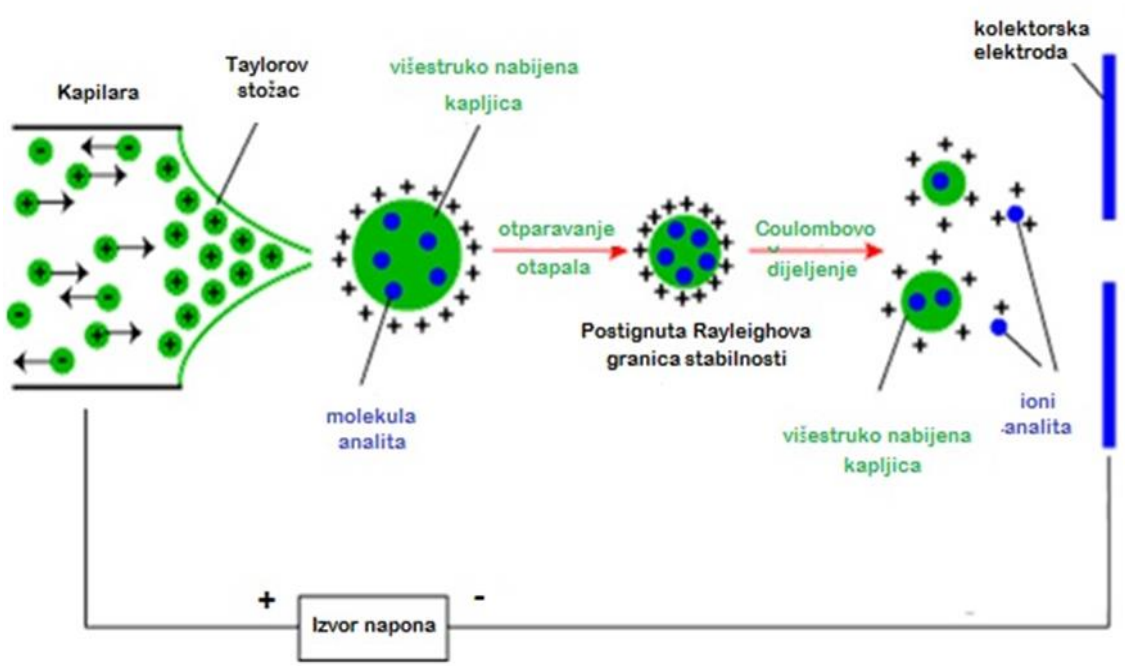
2.3.1. Ionski izvori

Funkcija ionskog izvora u spektrometru masa je stvaranje iona u plinovitoj fazi. Analit se prevodi u plinsku fazu i istovremeno ionizira (Ekman i sur., 2009). U spektrometriji masa izbor ionizacijske tehnike ovisi o unutarnjoj energiji koja se prenosi tijekom ionizacije te o fizikalno-kemijskim svojstvima analita koji će biti ioniziran. Postoji nekoliko različitih metoda ionizacije među kojima se ionizacija elektroraspršenjem (*engl.* Electrospray Ionization, ESI) i matricom potpomognuta ionizacija uz desorpciju laserskim zračenjem (*engl.* Matrix-assisted laser desorption/ionization, MALDI) najčešće koriste u analizi proteina i peptida (Steen i Mann, 2004).

Elektroraspršenje je način ionizacije uzorka direktno iz otopine koje omogućava analizu molekula bez obzira na veličinu i električni naboj. Tehniku su uveli i predložili Dole i sur. (1968), no tek su je Fenn i sur. (1989) uspjeli potpuno prilagoditi za korištenje u istraživanju biomakromolekula. Ukratko se može opisati kao vrsta elektrolitske ćelije u kojoj se dio naboja prenosi ionima u plinskoj fazi (Galić, 2004). Ionizacija se odvija tako da pokretna faza i analit ulaze u ionizator kroz kapilaru (iglu) koja ujedno predstavlja elektrodu pod visokim naponom (2-5 kV, Cindrić i sur., 2009). Pod utjecajem primijenjenog električnog polja dolazi do razdvajanja pozitivnog i negativnog naboja u otopini. Kada je kapilara priključena na pozitivni kraj izvora napona, pozitivno nabijeni ioni putuju prema katodi i akumuliraju se na površini tekućine koja se nalazi na kraju kapilare (Hoffmann i Strobant, 2007). Pri kritičnoj jakosti polja, na vršku kapilare, formira se tzv. Taylorov stožac u kojem se kontinuirano proizvode kapljice obogaćene pozitivno nabijenim ionima (Galić, 2004). Kolektorska elektroda privlači tako nabijene kapljice i daje im dodatno ubrzanje. Početna veličina kapljica ovisi o protoku otopine i svojstvima otapala. Kako otapalo otparava pod utjecajem struje dušika, temperature i električnog potencijala (ioni unutar otapala pokušavaju doći na površinu) tako se smanjuje veličina kapljica, ali naboj ostaje konstantan (Cindrić i sur., 2009). Kada se sila elektrostatskog odbijanja naboja u kapljici izjednači s površinskom napetošću kapljice, postiže se Rayleighova granica stabilnosti (Galić, 2004).

Nakon što su se dovoljno smanjile kapljice tako da se svi mogući ioni nalaze na površini kapljica, sile kulonskog odbijanja postaju veće od sila napetosti površine, kapljice se

otparavaju ili razbiju na manje kapljice (Coulombovo dijeljenje). Analit kristalizira ili prelazi u plinsku fazu. Ioni analita u trenutku napuštanja površine nisu u potpunosti bez otapala već se oko njih stvara sfera uparenog otapala. Nakon otparavanja kapljica, otvor iza kojeg je visoki vakuum i analizator privlači tako formirane kvazimolekulske ione (slika 3, Cindrić i sur., 2009).



Slika 3. Shematski prikaz ESI načina ionizacije (University of Bristol, 2005)

Tehnikom elektroraspršenja nastaju višestruko nabijeni ioni što omogućuje upotrebu spektrometara masa malog m/z mjernog područja (Kellner i sur., 1999) i detekciju kako relativno malih iona poput peptida te organskih i anorganskih spojeva, tako i velikih makromolekula poput proteina, RNA i DNA (Vučinić, 2012). Smanjenjem protoka analizirane otopine (mikroraspršenje, nanoraspršenje) povećana je osjetljivost tehnike, granice detekcije spuštene su na red veličine atomola (Galić, 2004), zbog čega je ESI veliku primjenu našao u povezivanju s tekućinskom kromatografijom visoke djelotvornosti i kapilarnom elektroforezom (Glish i Vachet, 2003). Smatra se vrlo blagom ionizacijskom tehnikom koja ne razara nekovalentno vezane biomolekularne komplekse poput proteina (Talkington i sur., 2005) pa čak ni čitave viruse (Bothner i Siuzdak, 2004).

2.3.2. Analizatori

Nakon ionizacije ioni se razdvajaju u analizatoru na temelju omjera mase i naboja (El-Aneed i sur., 2009). Analizatori masa predstavljaju ključnu točku u MS analizi. Svi analizatori masa koriste statičko ili dinamičko električno polje ili pak magnetsko polje koja mogu biti samostalna ili u kombinaciji. Osnovne karakteristike prema kojima se razlikuju analizatori su osjetljivost, gornja granica m/z, razlučivanje i točnost mjerenja mase. Osjetljivost je mjera odziva detektora dobivena analizom određenog analita, a gornja granica mase predstavlja najveću vrijednost m/z koju uređaj može izmjeriti. Razlučivanje je sposobnost razdvajanja iona približno istih m/z vrijednosti, dok se točnost mjerenja mase odnosi na sposobnost određivanja molekulske mase što bliže točnoj masi (izotopna masa).

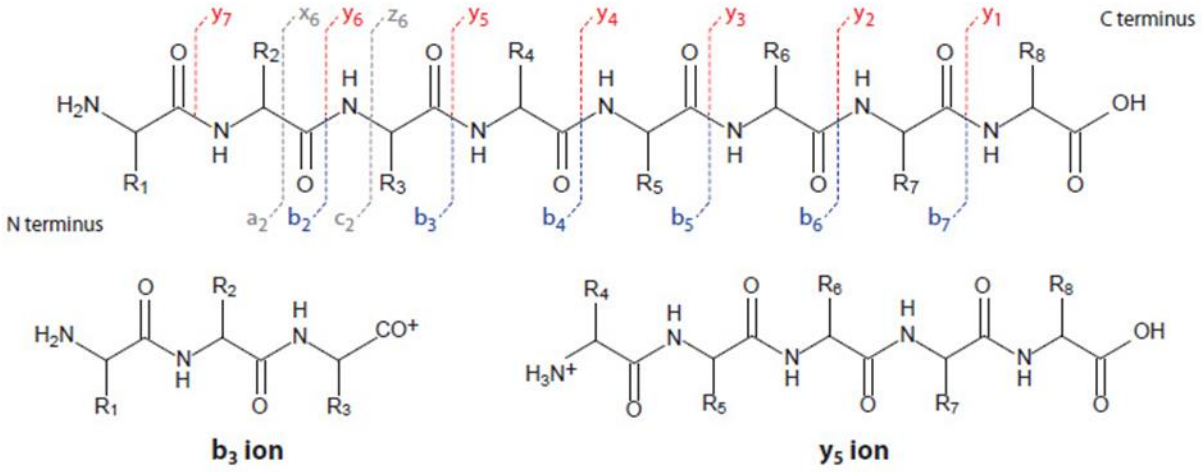
Najčešće korišteni analizatori u proteomskim istraživanjima jesu kvadrupol (*engl*. quadrupole), analizator vremena leta (*engl*. time-of-flight, TOF), ionska stupica (*engl*. ion trap) i Fourier transformirana ionska ciklotronska rezonancija (*engl*. Fourier transform ion cyclotron resonance, FTICR). Ioni nastali ESI ionizacijom najčešće se razdvajaju kvadrupolnim analizatorima te analizatorima ionskih stupica (Lane, 2005). Upotrebom hibridnih analizatora nastalih kao kombinacija istih (TOF-TOF) ili različitih analizatora (Q-TOF) mogu se poboljšati svojstva spektrometra masa. Prvi analizator masa služi za odabir iona specifičnog omjera m/z, dok drugi i/ili treći se pri takvoj izvedbi spektrometra masa koristi kao krajnji analizator (Galić, 2004). Njihovom upotrebom povećava se razlučivanje, osjetljivost i točnost mjerenja mase prilikom MS analize. Takvi spektrometri masa s hibridnim analizatorima ističu se teoretski neograničenim rasponom masa, jako visokim brzinama pregleda (do 10^6 u/s) te visokom snagom razlučivanja koja rezultira točnošću masa u rasponu ppm (Seger i Griesmacher, 2007). Točnost mjerenja mase i razlučivanje masenog analizatora imaju značajan utjecaj na informacije sadržane u spektru što je jako bitno u koraku identifikacije peptida (Cox i Mann, 2009).

2.3.3. Tandemna spektrometrija masa

MS analiza podrazumijeva mjerenje omjera mase i naboja analita koji može biti peptid najčešće dobiven cijepanjem proteinskih uzoraka odgovarajućom proteazom. Peptidni ioni analizirani u prvom stupnju rada spektrometra masa dalje se fragmentiraju tehnikama tandemne spektrometrije masa (MS/MS) koja se pak može odvijati u vremenu ili prostoru pa način fragmentacije ovisi o uvjetima ionizacije i vrsti instrumenta. U slučaju ionizacije

elektroraspršenjem i uporabom spektrometra tipa Q-TOF, u prvom analizatoru masa (kvadrupol) odabiru se ioni određene vrijednosti m/z (ioni prekursori) kojima se zatim dovodi dodatna energija u posebnoj kolizijskoj ćeliji kako bi se izazvala fragmentacija, nakon čega nastaju fragmenti iona prekursora (ioni produkti) koji se analiziraju u drugom analizatoru masa (TOF). Na taj način tandemna spektrometrija masa pruža uvid u primarnu strukturu proteina (Hunt i sur., 1986). Nakon detekcije iona produkata nastaju MS/MS spektri koji se lako mogu povezati s MS spektrima u kojima se nalaze ioni prekursori.

Kako bi nastali ioni produkti, najčešće se koristi kolizijom inducirana disocijacija (*engl.* collison induced dissociation, CID) koja se zasniva na sudaru molekula iona prekursora s inertnim plinom (dušikom, helijem ili argonom) pri čemu dolazi do fragmentacije i stvaranja iona produkata (Wells i McLuckey, 2005).



Slika 4. Nomenklatura fragmentnih iona nastalih disocijacijskim mehanizmom CID. Ioni serije a_m, b_m i c_m sadržavaju N-kraj, a ioni serije x_n, y_n i z_n sadržavaju peptidni C-kraj. Indeksi *m* i *n* označavaju redni broj aminokiseline u nizu (Steen i Mann, 2004)

Prilikom fragmentacije peptidnih iona prekursora, kovalentna veza puca unutar peptidnih iona na različitim mjestima pa tako mogu nastati ioni produkti koje dijelimo na: x, y, z i a, b, c ione u ovisnosti o mjestu fragmentacije u peptidu (slika 4). Nomenklaturu nastalih fragmentnih iona predložili su Roepstorff i Fohlman (1984), a kasnije su je modificirali Johnson i suradnici (1987). Koji tip iona produkata će nastati ovisi o načinu fragmentacije pa tako disocijacijskim mehanizmom CID većinom nastaju fragmentni ioni serije b i y. Usprkos djelomičnom pojednostavljenju MS/MS spektara (očitavanjem samo b- odnosno y- iona unutar jednog MS spektra), interpretacija MS/MS spektara nije jednostavna jer se trebaju

uzeti u obzir ioni koji gubitkom -OH, $-NH_2$ ili CH_3 skupina pridonose kompleksnosti spektara.

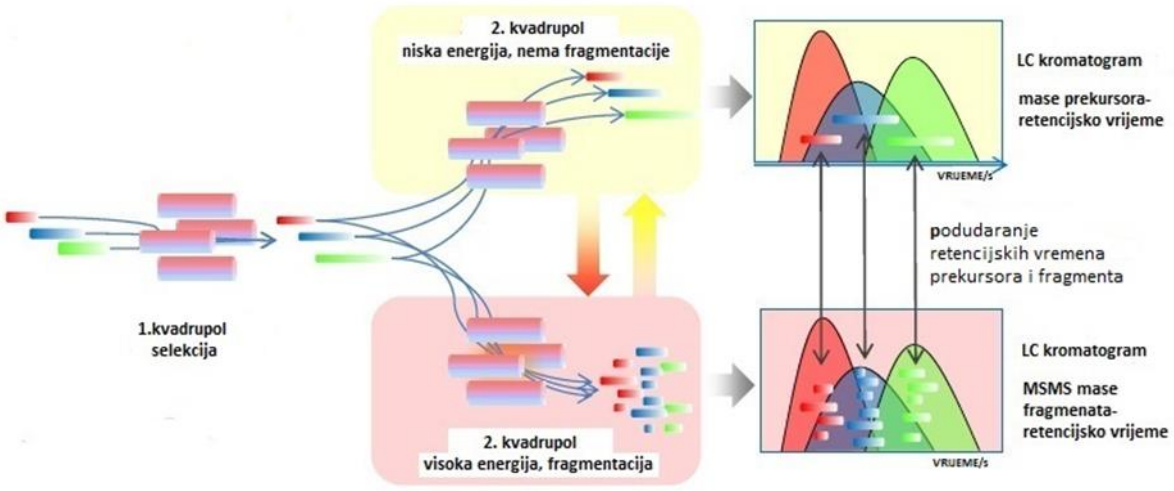
U pojedinim izvedbama spektrometrija masa moguća je i daljnja MS^3 analiza gdje ioni produkti nastali u MS/MS analizi postaju ioni prekursori koji nakon selekcije u analizatoru ponovno prolaze postupak fragmentacije.

2.3.4. MS^E

U standardnoj MS/MS analizi, svi peptidi ulaze u spektrometar masa gdje se odabiru ioni prekursori za fragmentaciju na temelju njihovog intenziteta (DDA metoda, *engl.* Data dependent acquisiton). Spomenuta analiza ograničena je fragmentacijom malog broja peptidnih iona tijekom jednog radnog ciklusa (Michalski i sur., 2011). Procijenjeno je da se potpunom tripsinskom digestijom oko 5000-6000 kvaščevih proteina može dobiti do 300 000 različitih uzoraka peptida (Cagney i sur., 2003). Iako Q-TOF spektrometri masa mogu fragmentirati do osam peptida u sekundi ili manje, većina od 300 000 peptida neće biti analizirana pa na taj način dolazi do gubitka dijela informacija u sekvenci (Geromanos i sur., 2009). Da bi se nadišlo ovo ograničenje, razvijena je metoda podatkovno neovisno prikupljanje podataka (DIA, *engl.* Data independent acquisition) nazvana MS^E gdje se ^{E} odnosi na upotrebu povišene energije (*engl.* elevated energy) tijekom procesa prikupljanja podataka (Slade, 2013). Komercijalno je dostupna još od Q-TOF Premier spektrometra masa, iako su tehnike slične MS^E primijenjene i kod ranijih generacija Q-TOF sustava od strane različitih proizvođača (Bateman i sur., 2002, Silva i sur., 2006) i drugih istraživačkih grupa (Purvine i sur., 2003, Niggeweg i sur., 2006).

Tijekom MS^E analize peptidni ioni prekursori ulaze u prvi analizator koji propušta sve ione do kolizijske ćelije u kojoj dolazi do kontinuirane i brze izmjene energije. Energija naizmjenično prelazi iz stanja niže energije pri kojoj peptidi ostaju intaktni, u stanje povišene energije pri kojoj dolazi do fragmentacije svih propuštenih peptida (Law i Lim, 2013). Dakle, u prvom radnom koraku MS^E analize snimaju se svi ioni prekursori eluirani s LC kolone u danom vremenu, dok u drugom dolazi do njihove fragmentacije i snimanja složenog seta dobivenih iona produkata. Ponavljanjem ovog ciklusa snimanja prekursora i produkata duž cijelog LC gradijenta, dobivaju se kromatogrami prekursora i produkata fragmentacije. Prilikom obrade podataka odgovarajući algoritam usklađuje vrijeme zadržavanja (retencijsko vrijeme) prekursora s vremenom zadržavanja njihovih produkata fragmentacije i kada nađe

savršeno podudaranje između njihovih vremena zadržavanja, pridružuje ione produkata odgovarajućim prekursorima (slika 5).



Slika 5. Princip MS^E metode u Q-TOF tipu spektrometra masa (Parviainen, 2014)

Ovakav pristup omogućuje detekciju značajno većeg broja peptida, u odnosu na metodu podatkovnog zavisnog prikupljanja (DDA, *engl.* Data dependent acquisition), gdje nije potrebno prethodno znanje o komponentama uzorka i omogućeno je paralelno prikupljanje podataka na nepristran način. Drugim riječima, nema prethodne selekcije peptidnih iona prekursora (na temelju njihovih intenziteta) prije procesa fragmentacije. Podaci se ne gube što rezultira znatno boljom identifikacijom proteina.

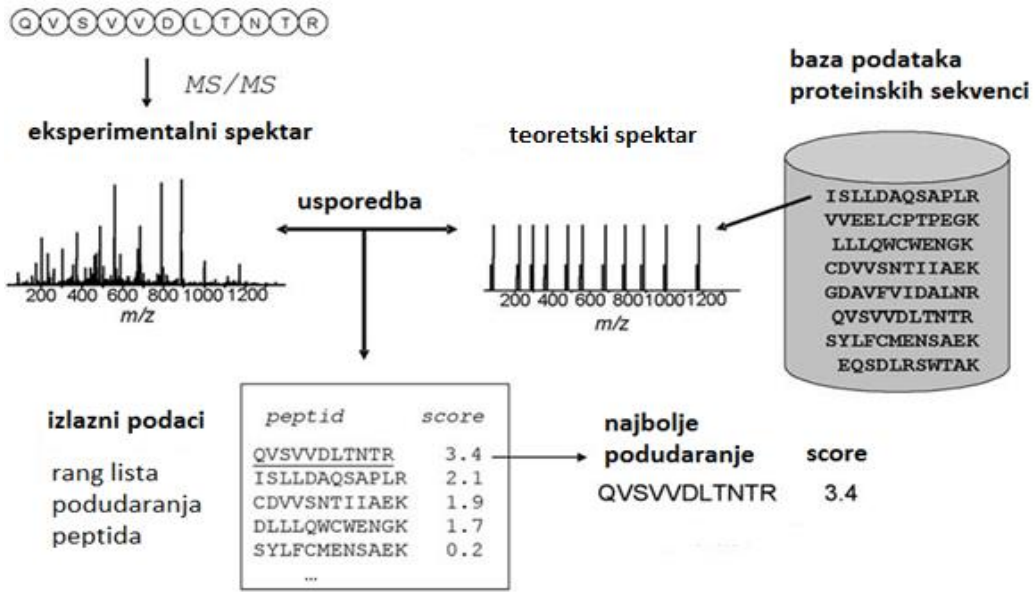
Budući da se gubi poveznica između masa prekursora i masa njihovih produkata, kvaliteta identifikacije proteina uvelike ovisi o uspješnom povezivanju iona prekursora s njihovim fragmentima pa je odabir programa za obradu MS^E podataka jako bitan.

2.4. PRETRAGA BAZE PODATAKA

Nakon analize spektrometrijom masa proteini se mogu identificirati pretragom baza podataka i sekvenciranjem peptida *de novo* (Reinders i sur., 2004). Iako su se pretražni programi za sekvenciranje peptida *de novo* značajno poboljšali u zadnje vrijeme, pretraga proteinskih baza podataka još uvijek se koristi kao glavna metoda u proteomskim istraživanjima. Uspoređuju se mase peptidnih iona dobivenih MS analizom i mase iona nastalih MS/MS analizom s podacima koji su pohranjeni u proteinskim bazama podataka (Nesvizhskii, 2007).

Jedna od metoda za pretragu baze podataka jest metoda „otiska prsta“ mase peptida (*engl.* Peptide Mass Fingerprinting, PMF) (Pappin i sur., 1993). Proteini se zbog svoje kompleksnosti analiziraju spektrometrijom masa kao pocijepani fragmenti odnosno peptidi nastali prethodnim djelovanjem proteaza (najčešće tripsina) čije je cijepanje uniformno i predvidljivo (Steen i Mann, 2004). Potom se načini *in silico* digestija proteina koji se nalaze u bazi podataka. Tako dobivene izračunate mase peptida uspoređuju s eksperimentalno dobivenim peptidima te se određuje stupanj podudarnosti (Ruvinsky i Marshall Graves, 2005). Metoda otiska prsta će dati dobre rezultate samo ako je u uzorku prisutan jedan protein (najviše dva) te ako se ispitivani protein (ili njegov homolog) nalazi u proteinskoj bazi podataka. Također ovakav pristup zahtjeva visoku točnost mjerenja masa (manje od 20 ppm) (Kapp i Schütz, 2007).

Puno obuhvatniji pristup za identifikaciju proteina daje pretraga MS/MS podataka (slika 6) gdje se peptidi dobiveni nakon digestije tripsinom analiziraju tandemnom spektrometrijom masa nakon čega se uspoređuju tako dobiveni MS/MS spektri s *in silico* generiranim MS/MS spektrima peptida u bazi podataka (Ruvinsky i Marshall Graves, 2005).



Slika 6. Pretraživanje MS/MS podataka (Nesvizhskii, 2007)

Veliki je broj pretražnih programa, komercijalnih i javno dostupnih, koji provode pretragu MS/MS podataka. Neki od njih zasnivaju se na heurističnom pristupu, dok drugi koriste probabilistički pristup. Pretražni programi koji koriste heuristični pristup uspoređuju

eksperimentalno dobiveni MS/MS spektar s teoretskim MS/MS spektrom te računaju uspješnost pogotka (*engl.* score) na osnovu sličnosti između ta dva spektra, odnosno računaju broj pikova koji su zajednički za ta dva spektra. Primjeri takvih pretražnih programa jesu SEQUEST (Eng i sur, 1994), Spectrum Mill (Zhang i sur., 2002) i X!-tandem (Craig i Beavis, 2004). S druge strane, probabilistički pretražni programi računaju vjerojatnost da je podudaranje između spektara slučajno. Što je veća ta vjerojatnost, to je uspjeh pogotka manji. Među takve pretražne programe spadaju Mascot (Perkins i sur., 1999), OMSSA, Phenyx, Protein Pilot (Shilov i sur., 2007) i ProteinLynx Global Server.

2.4.1. Parametri prilikom pretrage baze podataka

Prvo što je potrebno za pretraživanje baze podataka jest lista pikova. Maseni detektor daje neobrađene podatke koje je potrebno konvertirati u listu pikova kako bi se moglo izvoditi pretraživanje. Liste pikova su zapravo tekstualni podaci koji dolaze u različitim formatima, obično sadrže liste vrijednosti m/z i intenziteta pikova. Proces u kojem se neobrađeni podaci prevode u listu pikova naziva se „peak detection“ ili „peak picking“, a izvode ga računalni programi koji najčešće dolaze zajedno sa spektrometrima masa (Rognvaldsson i sur., 2004).

Osim liste pikova, prilikom pretrage baze podataka još nekoliko parametara potrebno je uzeti u obzir kako bi identifikacija proteina bila uspješna. Jedan od njih je odabir prikladne baze podataka proteinskih sekvenci koja predstavlja balans između osjetljivosti i brzine pretrage (Edwards, 2011). Pretragom velikih baza podataka produljuje se vrijeme pretrage i smanjuje osjetljivost identifikacije peptida. Povećava se vjerojatnost dobivanja slučajnih podudaranja masa peptida (Nesvizhskii, 2010). Pretragom manjih baza podataka kraće je vrijeme pretrage, ali se mogu propustiti neki značajni peptidi. Odabir prikladne baze podataka ovisi o cilju istraživanja. Za značajnija pretraživanja preporuča se pretraga visokokvalitetnih i održavanih proteinskih baza podataka (Duncan, 2010). Ako je analizirani protein iz organizma čiji je genom u potpunosti sekvenciran, može se izvoditi pretraga proteoma samo tog organizma (Eriksson i Fenyo, 2004).

Jako bitan parametar je tolerancija mase peptida (*engl.* mass tolerance). Točnost mjerenja mase pokazuje odstupanje izmjerene vrijednosti m/z od prave vrijednosti. Određuje koliko mali raspon pretraživanja mora biti odnosno kolika se pogreška mjerenja mase može tolerirati pri usporedbi teoretskih i eksperimentalno dobivenih vrijednosti masa. Ako se postavi mali raspon odstupanja, tek nekoliko proteina trebalo bi biti pogođeno pošto je manji broj slučajnih podudaranja u odnosu na veće odstupanje masa. S druge strane, preusko postavljenim

rasponom riskira se propuštanje određenog broja podudaranja peptida. Odabir vrijednosti tolerancije mase ovisi o tipu spektrometra masa koji je korišten prilikom analize. Sa spektrometrima masa visoke razlučivosti moguće je postići točnost mjerenja mase manju od 0.1 Da.

Jedan od parametara koji otežavaju konačnu identifikaciju proizlazi iz nesavršenog djelovanja proteolitičkih enzima. Ako enzim ne pocijepa protein na točno određenom mjestu pojavljuje se peptid koji veličinom prelazi očekivanu veličinu u bazi podataka. Takva se pojava naziva izmješteno cijepanje (*engl.* missed cleavage site) i potrebno ju je uzeti u obzir budući da broj kombinacija mogućih peptida postaje prevelik pa pretražna analiza može trajati i do nekoliko dana.

Dvije su vrste posttranslacijskih modifikacija koje treba uzeti u obzir prilikom pretrage baze podataka. To mogu biti fiksne modifikacije koje su uvijek prisutne te varijabilne modifikacije koje mogu i ne moraju biti prisutne. Primjer fiksnih modifikacija jest alkilacija cisteina. Pošto dolazi do alkilacije svih cisteina u peptidu, takva promjena će utjecati samo na promjenu mase cisteina, ali ne i na brzinu i specifičnost pretrage. Problem predstavljaju varijabilne modifikacije koje se ne pojavljuju se na svim aminokiselinskim ostatcima. Primjer je fosforilacija koja se može pojaviti na samo jednom serinu u peptidu koji sadrži više serina. Takve modifikacije produljuju vrijeme pretrage i smanjuju njenu specifičnost (Matthiesen i sur., 2011). Zato, ako je cilj pretrage identifikacija proteina, bolje je postaviti što manji broj varijabilnih modifikacija prilikom pretrage.

2.4.2. ProteinLynx Global Server

ProteinLynx Global Server (PLGS) komercijalni je programski paket kojeg razvija i prodaje korporacija Waters. Omogućuje kvalitativna i kvantitativna proteomska istraživanja koja se izvode isključivo na Waters spektrometrima masa. Hijerarhijski pristup pretrazi baze podataka primijenjen je s mogućnosti odabira da se spektri peptida analiziraju za nepoznate modifikacije (AutoMod) i/ili se pretražuju putem BLAST alata. Korisniku nudi izbor istovremeno započeti MASCOT i/ili PLGS pretraživanja iz zajedničkog sučelja pružajući tako mogućnost usporedbe rezultata pretrage. Spektri se prethodno obrađuju u odgovarajućem MaxEnt algoritmu, koji se koristi za svođenje višestrukih izotopa i naboja u monoizotopne pikove (m/z vrijednosti). Ovaj korak pruža bolju detekciju pikova, a stoga i bolje vrijednosti koje definiraju uspjeh pogotka peptida (*engl.* score) jer se na taj način iz ulaznog spektra uklanjaju dodatni pikovi izotopa. Zasniva se na visokim točnostima mjerenja masa (1-5 ppm)

što omogućava pouzdanu identifikaciju peptida, dok za računanje vjerojatnosti podudaranja peptida koristi Bayesovske metode poput Markov lanac Monte Karlo metode (MCMC, Markov Chain Monte Carlo)(Kapp i Schütz, 2007).

2.4.3. Proteinske baze podataka

Proteinske baze podataka sadrže informacije o eksperimentalno određenim strukturama proteina čime pružaju znanstvenoj zajednici pristup korisnim informacijama. Među najpoznatije javno dostupne proteinske baze podataka spadaju UniProtKB (http://www.uniprot.org/) i NCBInr (http://www.ncbi.nlm.nih.gov/protein/) baze podataka.

UniprotKB predstavlja centralno mjesto dostupnih informacija o proteinima s referencama iz više različitih izvora (Magrane i Uniprot Consortium, 2011). Dijeli se na SwissProt i TrEMBL.

SwissProt je visokokvalitetna, ručno anotirana i neredundantna baza proteinskih sekvenci koja obuhvaća informacije dobivene iz znanstvene literature i bioinformatičkih analiza. Anotacije se obnavljaju periodično kako bi baza bila u koraku s najnovijim otkrićima. Manualna anotacija omogućava daljnju provjeru proteinske sekvence i literaturnih izvora. SwissProt baza putem korisničkog sučelja nastoji pružiti korisniku što je moguće više informacija poput funkcije proteina, postrranslacijskih modifikacija, domena i veznih mjesta, homolognih proteina, uključenosti u bolesti i mnogih drugih. Pogodna je za pretraživanje strukturalne analize proteina, odnosno metodu otiska prsta (O' Donovan i sur., 2002). TrEMBL (*engl.* Translated EMBL Nucleotide Sequence Dana Library) obuhvaća sve automatski anotirane unose. Unosi su uglavnom hipotetski proteini koji se generiraju pretpostavljenom translacijom identificiranih ili pretpostavljenih gena u genomskim bazama podataka. Manualnom anotacijom TrEMBL unosa, sekvenca se obrađuje i provjerava te se unosi u SwissProt bazu.

Neredundantna NCBInr (*engl.* NCBI non-redundant, NCBInr) baza podataka predstavlja opsežnu neponavljajuću proteinsku bazu podataka koju razvija NCBI (*engl.* National center for Biotechnology Information). Sadrži prikupljene zapise iz nekoliko baza podataka poput translatiranih kodirajućih regija GenBank baze podataka (Benson i sur., 2011), kao i baza podataka SwissProt (Boeckmann i sur., 2003), PDB (Berman, 2008), PIR (*engl.* Protein information resource, PIR)(Wu i Nebert, 2004) i NCBI RefSeq (Pruitt i sur., 2005). Genbank sadrži detaljno opisane sekvence DNA anotirane nakon sekvenciranja genoma pojedine biološke vrste. NCBInr baza podataka dolazi u obliku komprimiranog FASTA formata zapisa

koji je pogodan za računalno čitanje ili u indeksiranom i aplikativnom obliku prilagođenome za pretraživanje BLAST alatom. Ne sadrži duplikate sekvenci proteina, a svaki protein u FASTA zaglavlju sadrži svoje ime, ime organizma te jedinstveni GI broj (*engl.* gene identifier, GI) pomoću kojeg se može povezati sa drugim bazama podataka na web sjedištu NCBI.

3. EKSPERIMENTALNI DIO

3.1. MATERIJAL

3.1.1. Kemikalije

- Acetonitril, CH_3CN, 99,9%, *Merck, Njemačka*
- Albumin goveđeg seruma, BSA, *Sigma-Aldrich, St. Louis, MO, SAD*
- Amonijev hidrogenkarbonat, *Sigma-Aldrich, St. Louis, MO, SAD*
- Mravlja kiselina, HCOOH, 98-100%, *Kemika, Zagreb, Hrvatska*
- Natrijev hidrogenkarbonat, $NaHCO_3$, *Kemika, Zagreb, Hrvatska*
- Tripsin, izoliran iz svinjske gušterače, pročišćen imunoafinitetnom kromatografijom, *Merck, Njemačka*
- Triton X-100, *Sigma-Aldrich, St. Louis, MO, SAD*

3.1.2. Laboratorijski pribor

- Automatske pipete
- Erlenmeyerove tikvice različitih volumena
- Filter papir, veličina pora 0,45 nm
- Kivete za spektrofotometar
- Mikrobiološka ušica
- Pinceta
- Plastične mikroepruvete
- Propipeta
- Staklene epruvete
- Staklene kuglice
- Staklene i plastične čaše
- Staklene i plastične menzure
- Staklene pipete različitih volumena
- Stakleni lijevak

3.1.3. Tehnička oprema

- Centrifuga 5415 R, *Eppendorf, Njemačka*
- Centrifuga Centric 400, *Tehtnica, Slovenija*
- Densimat, *BioMerieux, Francuska*
- ESI SYNAPT G2-Si, *Waters, Milford, MA, SAD*
- kolona nanoAcquity UPLC 1,7 µm BEH130 C18, 100 µm x 100 mm, *Waters, Milford, MA, SAD*
- nanoACQUITY™ UPLC™ System, *Waters, Milford, MA, SAD*
- predkolona nanoAcquity UPLC 2G-V/M Trap 5 µm Symmetry C18, 180 µm x 20 mm, *Waters, Milford, MA, SAD*
- Spektrofotometar BioPhotometer, *Eppendorf, Njemačka*
- Termomiješalica Comfort, *Eppendorf, Njemačka*
- Tresilica IKA MS 3 basic, *IKA, Njemačka*
- Vakuum centrifuga Concentrator 5301, *Eppendorf, Njemačka*

3.1.4. Osnovne otopine

Otopina amonijevog hidrogenkarbonata

Otopina amonijevog hidrogenkarbonata priprema se otapanjem 3,95 g amonijevog hidrogenkarbonata u 1 L destilirane vode. Potom se pH otopine podesi na pH 7.8. Konačna koncentracija otopine iznosi 50 mM.

Otopina tripsina

Otopina tripsina dobiva se razrjeđivanjem matične otopine s 25 mM otopinom amonijevog hidrogenkarbonata na koncentraciju 20 µg mL^{-1}. Matična otopina se dobiva otapanjem jedne tablete tripsina u 1 mL destilirane vode. Njena konačna koncentracija iznosi 1 mg mL^{-1}.

Bradford reagens

Bradford reagens dobiva se otapanjem 10 mg boje Coomasie brilliant blue G-250 u 3 mL vode. Potom se dodaje 5 mL 95% etanola i 10 mL 85% fosforne kiseline te se dopuni s destiliranom vodom do 100 mL. Otopina se dobro izmiješa i profiltrira kroz filtar papir širine pora 0,45 nm.

3.1.5. Ispitivani uzorak

U eksperimentalnom dijelu rada kao radni mikroorganizam korišten je kvasac *Candida albicans* SC5314 koji pripada zbirci mikrobnih sojeva Farmaceutsko-biokemijskog fakulteta, uzgajan na kompletnoj podlozi koja sadrži Sabouradov bujon.

Podloga je sterilizirana tijekom 15 minuta pri 121 °C. Sastav Sabouraudovog bujona prikazan je u tablici 1. Sabouraudov bujon volumena od 20 mL inokuliran je čistom kulturom kvaščevih stanica. Nakon inokulacije kvaščeva suspenzija uzgajana je 24 sata na tresilici pri 37 °C.

Tablica 1. Sastav Sabouraudovog bujona

SASTOJAK	KOLIČINA
Glukoza	20 g
Pepton	10 g
dH_2O	1 L

3.1.6. Programi za upravljanje instrumentima, obradu podataka i programski alati

- Mass Lynx software verzija 4.1. SCN902, *Waters, Milford, MA, SAD*
- Protein Lynx Global Server verzija, PLGS, v. 3.0.1, *Waters, Milford, MA, SAD*
- Program za detekciju vrsta razvijen u Kabinetu za Bioinformatiku pisan u programsku jeziku Python, verzija 2.7.
- Protein Reader software za spektrometriju masa razvijen u Kabinetu za Bioinformatiku, pisan u programskom jeziku Java
- PANTHER „online resource“ platforma, verzija 10.0

3.2. METODE

3.2.1. Razbijanje kvaščevih stanica i izolacija proteina

Uzgojena stanična suspenzija prebaci se u odgovarajuću epruvetu od 15 mL za centrifugiranje i centrifugira 15 min na 2 000 o min^{-1} na +4 °C. Supernatant se baci, a talog stanica resuspendira se u 1 mL 25 mM NH_4HCO_3 puferu (pH 7.8). Potom se talog stanica prebaci u plastičnu mikroepruvetu od 1,5 mL i centrifugira 20 min pri 4 500 o min^{-1} na +4 °C. Taj postupak ispiranja ponovi se još 2 puta.

Dobiveni talog resuspendira se u 300 µL 25 mM NH_4HCO_3 pufera (pH 7.8) koji sadrži 0.1% Triton-X 100. Resuspendirane stanice liziraju se staklenim kuglicama promjera od 0,4 do 0,6 mm tako da visina kuglica u epruveti bude jednaka visini suspenzije stanica. Stanice se miješaju s kuglicama 30 s uz pomoć stolne vibromiješalice te se hlade na ledu 1 min, ciklus miješanja i hlađenja ponovi se 5 puta.

Nakon što se kuglice istalože, talog se baci, a supernatant se prokuha u vodenoj kupelji 3 min. Potom se centrifugira 20 min na 7 000 o min^{-1} pri +4 °C. Nakon centrifugiranja, talog se baci, odvoji se supernatant i prebaci u plastičnu mikroepruvetu od 1,5 mL te se odredi koncentracija proteina metodom po Bradford.

3.2.2. Određivanje koncentracije proteina metodom po Bradford

Metoda po Bradford (Bradford, 1976) kolorimetrijska je metoda za mjerenje koncentracije proteina koja se temelji na reakciji proteina s bojom *Coomassie Brilliant Blue* G-250 (CBB) u kiselom mediju. CBB reagira s pobočnim grupama Arg te u manjoj mjeri s pobočnim grupama His, Lys, Tyr, Trp i Phe. Anionski oblik boje se hidrofobnim interakcijama i ionskim vezama veže sa spomenutim ograncima proteina pri čemu dolazi do stvaranja kompleksa protein – boja i dovodi do vidljive promjene boje iz smeđe u plavu. Promjena boje se određuje spektrofotometrijski pri valnoj duljini od 595 nm i to prema slijepoj probi. Koncentracija proteina se očitava iz baždarne krivulje koja se dobiva mjerenjem apsorbancija niza otopina goveđeg serumskog albumina (*engl.* bovine serum albumin, BSA) poznatih koncentracija (1 µg mL^{-1}, 2 µg mL^{-1}, 4 µg mL^{-1}, 6 µg mL^{-1}, 8 µg mL^{-1}, 10 µg mL^{-1}, 20 µg mL^{-1}, 30 µg mL^{-1}, 40 µg mL^{-1}, 50 µg mL^{-1}).

Uzorci se priprave na način da se u 100 µL uzorka poznate koncentracije BSA doda 1 mL Bradford reagensa, suspenzija se promiješa i ostavi na sobnoj temperaturi 10 minuta.

Uzorci nepoznate koncentracije razrijede se s destiliranom vodom u omjeru 1:100 te se doda 1 mL Bradford reagensa, promiješa i također ostavi 10 min na sobnoj temperaturi. Slijepa proba pripravi se tako da se u 100 µL destilirane vode doda 1 mL Bradford reagensa nakon čega se mjeri apsorbancija pri 595 nm.

3.2.3. Digestija proteina

Slijedeći korak u pripremi uzorka za analizu spektrometrom masa je digestija proteina uz pomoć specifične proteaze tripsina. Nakon što je određena početna koncentracija proteina u uzorcima, konačna koncentracija proteina u uzorcima ujednači se na koncentraciju od 1,5 mg mL^{-1} (razrijeđenije s otopinom tripsinskog pufera) te se na volumen od 100 µL dodaje 5 µL tripsina koncentracije 1 mg mL^{-1}. Uzorci se inkubiraju 18 sati pri 37 °C i 400 o min^{-1} na termomiješalici.

3.2.4. Analiza peptida UPLC-MS^E pristupom

Peptidni uzorci analizirani su nanoUPLC-MS^E pristupom. Program Mass Lynx verzija 4.1 (Waters, Milford, MA, SAD) korišten je za postavljanje parametara korištenog instrumenta. Peptidi su separirani uz pomoć nanoAcquity UPLC instrumenta na koloni nanoAcquity UPLC 1,7 µm BEH130 C18 (tablica 2) pri temperaturi 40 °C i protoku 1 µL min^{-1} . Volumen injektiranja iznosio je 1 µL. Kao pokretna faza A korištena je vodena otopina mravlje kiseline (0,1%; ϕ), dok je pokretna faza B bila vodena otopina koja sadrži 0,1% mravlje kiseline i 95% acetonitrila (ϕ). Metoda koja se koristi traje 60 minuta s gradijentnom elucijom koja raste do 52. minute pri čemu udio faze B raste od 1% do 99%, a potom se uvjeti na koloni vraćaju na početno stanje. Predkolona nanoAcquity UPLC 2G-V/M Trap 5 µm Symmetry C18 korištena je za uklanjanje nečistoća iz uzorka prije separacije na samoj koloni (tablica 2).

Separirani peptidi detektirani su pomoću instrumenta ESI SYNAPT G2-Si. Analiza peptida provedena je u pozitivnom načinu rada instrumenta s parametrima navedenim u tablici 3. Tijekom analize uzoraka korišten je stalni protok leucin-enkefalina (1 ng/µL, protok 0,4 µL/min, $[M + H]^+$ = 556.2771 Da) s ciljem korigiranja točnosti masa. Za procesiranje dobivenih podataka korišten je Protein Lynx Global Server verzija 3.0.1 (Waters, Milford, MA, SAD) s parametrima navedenim u tablici 4.

Tablica 2. Parametri predkolone i kolone

Vrijeme (min)	**Protok (µL min^{-1})**	**% A**	**% B**
PREDKOLONA			
0 – 60.00	1.000	99.0	1.0
KOLONA			
početak	1.000	99.0	1.0
0.10	1.000	99.0	1.0
40.00	1.000	50.0	50.0
52.00	1.000	1.0	99.0
54.00	1.000	99.0	1.0

Tablica 3. Parametri separacije

Mode	MS^{E}, positive
Mass	50 – 3000 Da
Lock Spray Configuration:	
Reference Scan Frequency	60.000 s
Reference Cone Voltage	40.000 V
Reference Trap Collision Energy	4.000
Experimental Instrument Parameters:	
Capillary	3.5000 kV
Source Temperature	50°C
Sampling Cone	40.0000
Desolvation Temperature	150°C
Nanoflow Gas Pressure	1.0 bar
Nebuliser Gas Flow	6.5 bar
Trap Collision Energy	4.0
Transfer Collision Energy	4.0
MS/MS Scan Time	1.0
Run Time	60.0

Tablica 4. Parametri korišteni za procesiranje dobivenih podakata pomoću ProteinLynx Global Server

Databank	SwissProt *C.albicans* Uniprot *C.albicans* SwissProt
Peptide Tolerance	Automatic
Fragment Tolerance	Automatic
Min Fragment Ion Matches per Peptide	3
Min Fragment Ion Matches per Protein	7
Min Peptide Matches per Protein	1
Maximum Protein Mass	250000 Da
Primary Digest Reagent	Trypsin
Secondary Digest Reagent	None
Missed Cleavages	1
Variable Modifier Reagents	Deamidation N Deamidation Q Dehydration ST
False Discovery Rate	4 %
Monoisotopic or Average	Monoisotopic
Peptide Charge	1+
Instrument Type	ESI-QUAD-TOF
ProteinLynx – processing parameters	
Chromatographic Peak Width	Automatic
MS TOF Resolution	Automatic
Lock Mass for Charge 1	556.2771 Da/e
Lock Mass Window	0.25 Da
Low Energy Threshold	135.0 counts
Elevated Energy Threshold	30.0 counts
Intensity Treshold	750.0 counts

3.2.5. Bioinformatička analiza MS/MS spektara novorazvijenom metodom baziranom na LSI

U ovom radu provedeno je testiranje brze, jeftine, selektivne i reproducibilne analitičke metode za određivanje aminokiselinskog slijeda (triptičkih) peptida u proteomskim analizama, koja u konačnici ima aplikaciju kao brze, jednoznačne i učinkovite metode identifikacije mikroorganizama bazirane na spektrometriji masa. Kako upravo interpretacija spektara ima jednu od ključnih uloga, glavni je naglasak na razvoju novih bioinformatičkih algoritama za pretraživanje stalno rastućih bioloških baza podataka.

Klasične medicinsko mikrobiološke metode danas su najzastupljenije u identifikaciji mikroorganizama, međutim te metode imaju i svojih ograničenja. Njihovo značajno ograničenje leži u uzgoju mikroorganizama i posljedično tome vremenu koje je potrebno za njihovu identifikaciju. Kombinacijom visoko osjetljivog tandemnog spektrometra masa s, na Prehrambeno biotehnološkom fakultetu razvijenim inovativnim ekspertnim sustavom za bržu i točniju analizu aminokiselinskog slijeda peptida, moguće je detektirati vrstu/soj mikroorganizma analizom peptidnog „otiska prsta“ bez prethodnog uzgoja. Sve postojeće metode, uključujući ProteinLynx Global Server obrađuju svaki snimljeni spektar zasebno, bez obzira na to što potječu od jednog organizma. Ta činjenica uz konstantan porast baza podataka dovodi do poteškoća. Prva poteškoća leži u samoj veličini baza podataka proteina, dok drugu predstavlja proces sparivanja i usporedbe brojnih zabilježenih pogodaka sa ciljem predikcije organizma. Slične poteškoće nalazimo u domeni obrade pisanog teksta. U ovoj je domeni takav problem elegantno riješiv metodama kao što je Latentno semantičko indeksiranje (LSI).

LSI (*engl.* Latent Semantic Indexing, LSI) je metoda razvijena prvenstveno za analizu nestrukturiranih zbirki tekstova (Deerwester i sur., 1990). U okviru Kabineta za Bioinformatiku PBF-a, ova je metoda prilagođena za analizu mikrobnih proteoma. U okviru prilagodbe, proteomi su prikazani kao tekstualni dokumenti, a metoda ima za cilj pomoću iščitanih fragmenata proteoma dobivenih MS2 analizom (MS/MS analizom) identificirati odgovarajući mikroorganizam. Ključna točka u tom procesu je reprezentacija proteina tj. pronalazak prikladne funkcije koja bi bila u stanju translatirati proteinsku sekvencu u nešto nalik riječi. U nedavnim istraživanjima proteini su uspješno prikazani kao riječi tj. kolekcije tripeptida (Couto i sur., 2007) i iskorišteni za klasifikaciju proteina prema odgovarajućim proteinskim porodicama. LSI je također uspješno primjenjen za predikciju supstratne specifičnosti kod enzima (Baranašić i sur., 2014). Metoda translacije proteina zasniva se na kliznom prozoru duljine 3 aminokiselinska ostatka. Ovom metodom jedan protein daje tri

znakovna slijeda koja se spajaju u jedinstvenu „rečenicu“. Primjerice peptid AIDGEKKL prikazujemo kroz tri slijeda, a to su: [AID, GEK], [IDG, EKK] i [DGE, KKL] čime se koriste sve raspoložive kombinacije. Drugi parametri koji utječu na uspješnost metode su redom: raspon masa koje uzimamo u analizu, broj opaženih iona visokog intenziteta, područje spektra u kojem se nalazi kandidat za -b odnosno -y seriju iona, dozvoljeno odstupanje od teoretski izračunate mase i značajna vrijednost kosinusa kao mjere poklapanja sa proteinom iz pretražive baze podataka (tablica 5).

Tablica 5. Parametri korišteni za bioinformatičku analizu na bazi LSI

PARAMETAR	VRIJEDNOST
Broj slijedova	3
Broj aminokiselina u prozoru	3
Minimalna masa	800.0 Da
Min Maksimalna masa	4 200.0 Da
Ukupan broj b i y iona	20
Analizirana područja spektra	5
Dozvoljeno odstupanje	0.1 Da
Vrijednost kosinusa	0.35

3.2.6. Genska ontologija

Projekt genske ontologije (*engl.* gene ontology, GO) omogućuje upotrebu kontroliranih rječnika precizno definiranih pojmova, koji prezentiraju svojstva gena i njihovih produkata (Gene ontology Consortium, 2001). U okviru ovog projekta razvijena su tri rječnika odnosno ontologije u kojima su definirani pojmovi koji opisuju kojoj staničnoj komponenti pripada genski produkt (*engl.* cellular component), u kojem biološkom procesu sudjeluje (*engl.* biological process) i koja mu je molekularna funkcija (*engl.* molecular function). Svi pojmovi koji se koriste u rječnicima mogu se primijeniti na sve biološke organizme te se neprestano nadograđuju. Svaki od tih pojmova povezan je s jednim ili više drugih termina u istom ili u

drugom rječniku. Cilj ontologije gena je poboljšati anotacije za sve gene i točno odrediti sve poznate informacije od sekvence DNA, proteinskog produkta pa do potencijalne funkcije pojedinog gena.

Dio projekta genske ontologije (GO projekt) je i PANTHER klasifikacijski sustav (*engl.* Panther ANalysis Through Evolutionary Relationships, PANTHER), složeni sustav koji objedinjuje funkcije gena, ontologiju, metaboličke puteve i alate za provođenje statističkih analiza koji omogućuju publici istraživanje gena u širokim mjerilima, istraživanja u proteomici i ekspresiji gena (Mi i sur., 2013). Sastavljen je od tri funkcionalna modula. Prvi modul jest baza podataka koja sadrži sve gene, koji kodiraju za proteine, iz 82 organizma. Ti geni su organizirani prvo u porodice na temelju homologije sekvenci, a potom u potporodice na temelju zajedničkih funkcija. Drugi modul su PANTHER putevi. Modul sadrži 176 puteva, održavanih od strane stručnjaka, koji su povezani s pojedinim proteinima u proteinskim bazama podataka te statističkim modelima i informacijama o filogenetici. Treći modul čine bioinformatički alati koji pružaju korisnicima vizualizaciju, analizu i interpretaciju eksperimentalnih podataka (Thomas i sur., 2006).

Za analizu genske ontologije korišten je jedan od PANTHER alata za statističku analizu podataka uz parametre navedene u tablici 6.

Tablica 6. Parametri korišteni za analizu genske ontologije

Analysis Type	PANTHER Overrepresentation Test
Annotation Version	PANTHER version 10.0
Analyzed List	candida_ESI_proteins.txt (*Candida albicans*)
Reference List	*Candida albicans* (all genes in database)
Annotation Data Set	PANTHER GO-Slim Biological Process
Bonferroni correction	True

4. REZULTATI I RASPRAVA

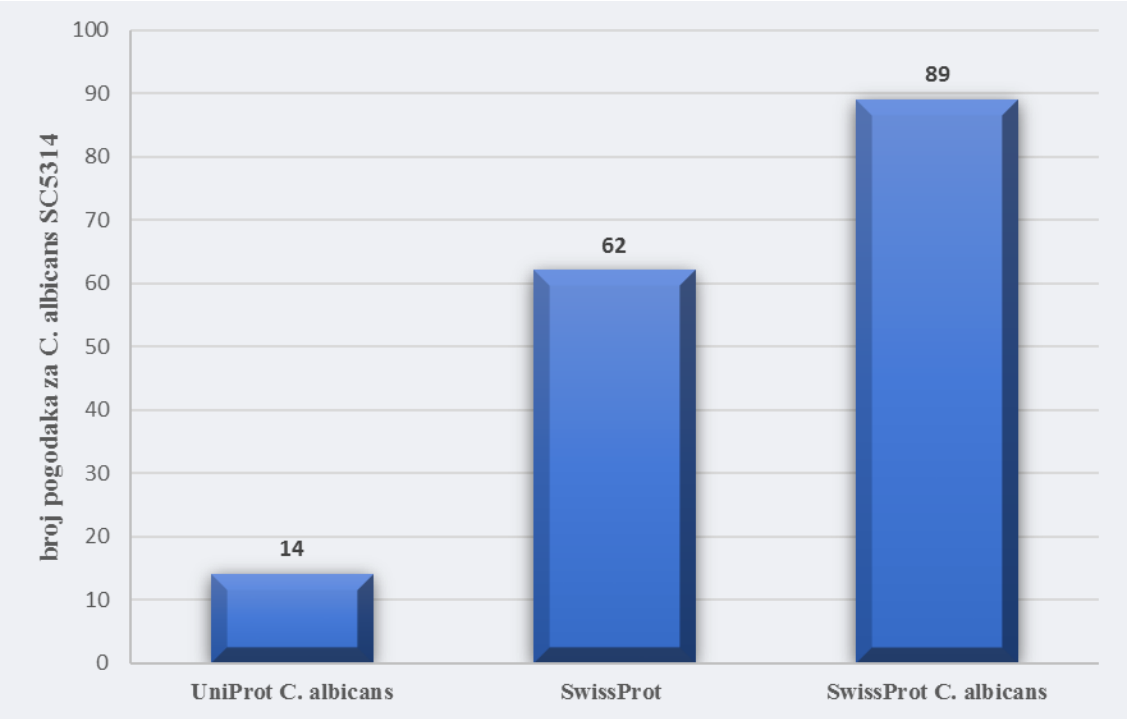
U ovom radu uspoređen je novorazvijeni alat za analizu spektara peptidnih fragmenata, koji se oslanja na LSI, s komercijalnim alatom ProteinLynx Global Server s ciljem identifikacije vrste odnosno soja mikroorganizma. Pritom je, kao radni mikroorganizam, korišten kvasac *Candida albicans* SC5314 uzgajan na Sabouraudovom bujonu. Rezultati određivanja vrste prikazani su slikama 7-9. Rezultati pretrage proteoma kvasca *Candida albicans* SC5314 prikazani su u tablici 9.

Provedena je i analiza ontologije gena putem alata javno dostupne platforme PANTHER s ciljem određivanja zastupljenosti pojedinih GO pojmova unutar skupina gena koji su bili eksprimirani u trenutku uzorkovanja stanica ispitivanog mikroorganizma. Rezultati analize genske ontologije prikazani su na slikama 10 i 11.

4.1. ODREĐIVANJE VRSTE

4.1.1. Određivanje vrste mikroorganizma korištenjem programa ProteinLynx Global Server

Nakon digestije proteina tripsinom, dobiveni peptidi analizirani su UPLC-MS^E pristupom pri čemu je identifikacija proteina provedena pretragom baza podataka. Pretražni program pritom korišten je ProteinLynx Global Server (PLGS). Pretražena je SwissProt baza podataka za kvasac *Candida albicans* te je identificirano 137 proteina. Kako bi se ispitala učinkovitost ovog programa za određivanje vrste, pretražene su još UniProt baza podataka za kvasac *Candida albicans* te cijela SwissProt proteinska baza podataka (tablice u prilozima). Uspoređen je broj identificiranih proteina pronađenih za ispitivani mikroorganizam u pretraživanim bazama podataka te su rezultati prikazani na slici 7.



Slika 7. Broj pronađenih proteina za *Candida albicans* SC5314 u pretraženim bazama podataka

Iz slike je vidljivo da je najveći broj proteina koji pripadaju kvascu *Candida albicans* SC5314 dobiven prilikom pretrage SwissProt baze podataka za *Candida albicans*. Pritom je pronađeno 89 proteina koji pripadaju ispitivanom mikroorganizmu dok su, unutar SwissProt

baze podataka, pronađena 62 proteina koji pripadaju kvascu *Candida albicans* SC5314. Najmanje proteina pronađeno je u UniProt *Candida albicans* bazi podataka.

S obzirom na područja pretrage tih baza podataka (tablica 7), rezultati nisu u skladu s očekivanjima jer je program rezultirao manjim brojem proteina u Uniprot *Candida albicans* bazi podataka koja ima uže područje pretrage u odnosu na Swissprot bazu podataka. Unutar Swissprot baze podataka pretražene su proteinske sekvence unutar 5534 vrste, dok je područje pretrage u Uniprot *Candida albicans* puno uže; pretražuju se proteinske sekvence samo unutar baze podataka za kvasac *Candida albicans* čije je područje pretrage 92 soja.

Tablica 7. Broj proteinskih sekvenci u bazi podataka (pristupljeno travanj, 2016)

Baza podataka	Broj proteinskih sekvenci	Broj vrsta/sojeva
SwissProt *Candida albicans*	1 402	92 soja
UniProt *Candida albicans*	159 068	92 soja
Swissprot	550 960	5 534 vrste

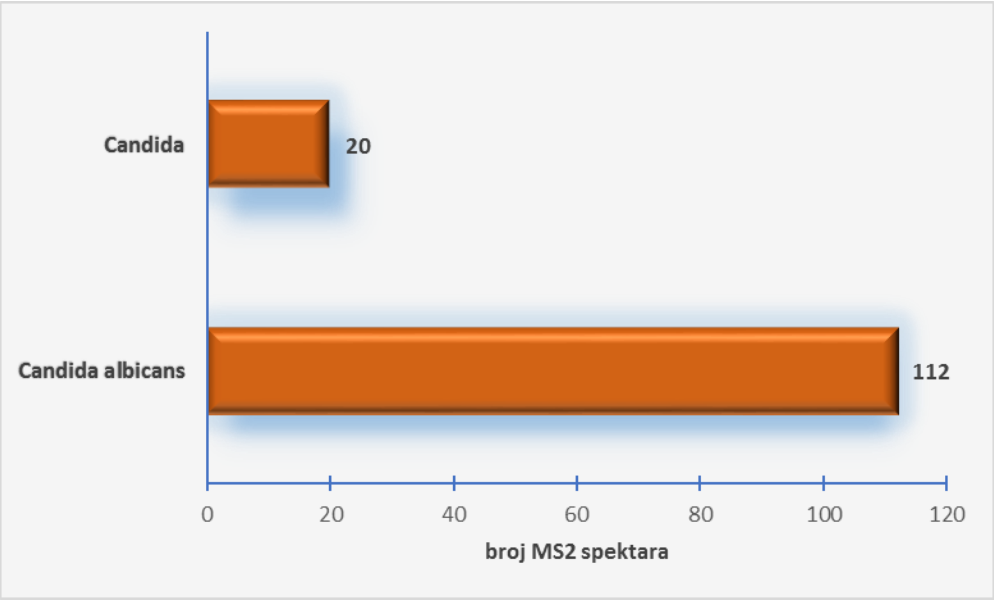
Rezultati su pokazali da takav programski alat mora znati točno što pretražuje te nije prikladan alat za određivanje vrste odnosno soja mikroorganizma. Prilikom pretrage Swissprot *Candida albicans* proteinske baze podataka dao je pozitivne rezultate jer je pronašao najveći broj proteina za analizirani mikroorganizam *Candida albicans* SC5314. Uspješnost identifikacije mikroorganizama limitirana je veličinom pretraživane baze podataka proteinskih sekvenci.

4.1.2. Određivanje vrste mikroorganizma novorazvijenom metodom baziranom na LSI

Na slikama 8 i 9 prikazani su rezultati određivanja vrste na temelju analize MS/MS spektara (MS2 spektara) dobiveni korištenjem programskog alata razvijenog u Kabinetu za Bioinformatiku.

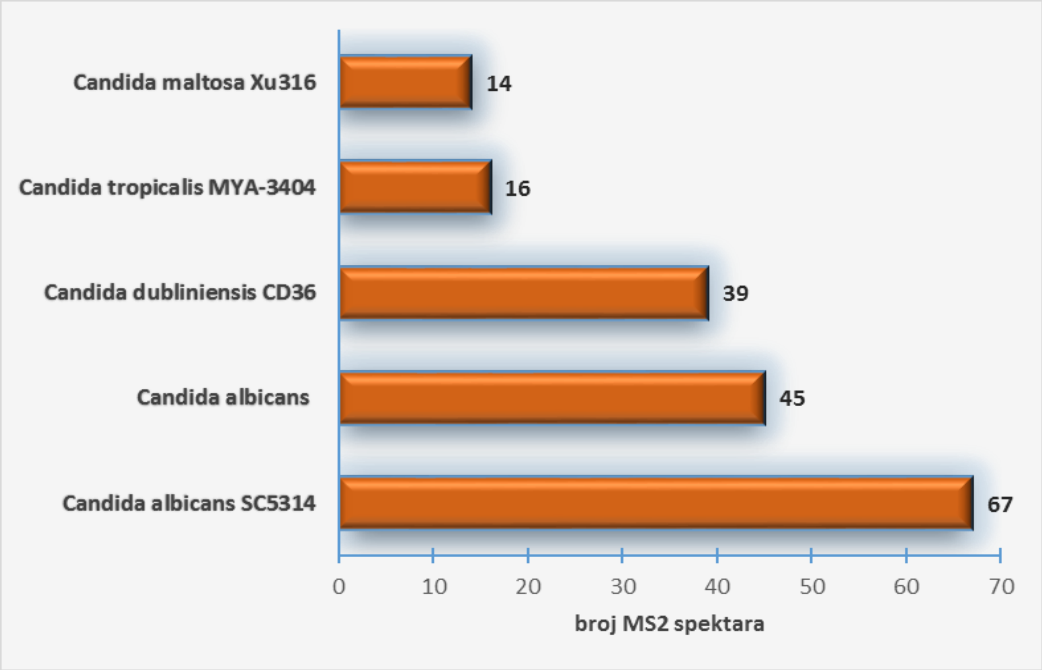
Nakon analize proteina UPLC-MS^E pristupom, dobiveno je 124 268 MS2 spektara. Od tih 124 268 spektara, 3000 MS2 spektara odabrano je na temelju intenziteta MS1 spektara. Intenzitet predstavlja funkciju koncentracije i sposobnosti peptida da primi naboj, pokazuje koliko je MS2 spektar relevantan. Zbog toga se za daljnju analizu odabiru samo oni peptidi koji su najbolje primili naboj i koji su značajni za stanicu te je iščitan njihov slijed aminokiselina.

Nakon toga pretražena je Genbank nr baza podataka (*engl.* non redundant) koja sadrži sve tripsin digeste razvrstane po proteinima i vrstama. Pretragom tako kreirane interne baze podataka dohvaćaju se sve taksonomske grupe kojima ti slijedovi pripadaju, inicira se takozvani „Brojač vrsta“. Kada se analizom MS2 spektra dobije proteinski slijed koji pripada određenoj vrsti, pripadajućem brojaču za tu vrstu povećava se vrijednost za 1. Ona vrsta čija je vrijednost brojača bila najveća odabire se kao vrsta koja se određuje.



Slika 8. Određivanje *Candida albicans* vrste pomoću MS2 analize i pretrage „nr“ baze podataka

Na slici 8 vidljivo je da je daleko najveći broj spektara pronađen za vrstu *Candida albicans*. Na nivou soja, 67 spektara odgovaralo je soju kvasca *Candida albicans* SC5314 (slika 9). Rezultati upućuju na to da je ispitivani soj bio *Candida albicans* SC5314.



Slika 9. Određivanje soja *Candida albicans* SC5314 pomoću MS2 analize i pretrage „nr“ baze podataka

Dodatna provjera identifikacije vrste provedena je pretragom MS1 spektara. Ako je na osnovu MS2 spektara prethodno točno određen soj, pretraga MS1 spektara trebala bi rezultirati većim brojem potvrđenih MS1 spektara za kvasac *Candida albicans* SC5314 u odnosu na broj MS1 spektara za kvasac *Candida dubliniensis*.

Tablica 8. Rezultati pretrage MS1 spektara

Vrsta	**Veličina proteoma**	**Broj potvrđenih MS1 spektara**	**MS1 proteom postotak (%)**
Candida albicans SC5314	43210	7251	16,8
Candida dubliniensis	11899	44	0,37

U tablici 8 prikazan je broj potvrđenih MS1 spektara za kvasac *Candida albicans* SC5314 i *Candida dubliniensis* te je uspoređen postotak dobivenog proteoma na osnovi MS1 spektara. Vidljivo je da je broj potvrđenih MS1 spektara za kvasac *Candida albicans* značajno veći u odnosu na kvasac *Candida dubliniensis* i to je dodatna potvrda da je ovom metodom točno određen soj.

4.2. ANALIZA PROTEOMA

Nakon što je metodom, koju koristi „Brojač vrsta“, potvrđeno da je u uzorku bio prisutan kvasac *Candida albicans* SC5314 sužava se područje pretrage te je pretražen proteom samo te vrste čijom je pretragom dobiveno 54 proteina potvrđena MS2 spektrima, s najvećim vrijednostima kosinusa, u usporedbi s peptidima iz pretraživane baze podataka (GenBank nr), odnosno 757 proteina potvrđena MS2 spektrima, nakon utvrđivanja vrste i spuštanjem vrijednosti kosinusa na još uvijek relevantnih 0,35.

U tablici 9 prikazani su proteini potvrđeni pretragom MS2 spektrima. Prikaz dobivenih proteina potvrđenih MS2 spektrima, nakon spuštanja vrijednosti kosinusa na 0,35, nalazi se u prilozima.

Tablica 9. Proteini dobiveni pretragom proteoma *Candida albicans* SC5314 i potvrđeni MS2 spektrima

Ime proteina	identificirani tripsinski digesti	Vrijednost kosinusa
conserved hypothetical protein	RDDFFDDFWK	0,655023
glyceraldehyde-3-phosphate dehydrogenase, type I, partial	AGILLSPTFVK	0,506038
glyceraldehyde 3-phosphate dehydrogenase	VPTTDVSVVDLTVR	0,58183
RecName: Full=Glyceraldehyde-3-phosphate dehydrogenase	LISWYDNEYGYSTR	0,479309
conserved hypothetical protein	YFVGFDDNVK	0,465357
glyceraldehyde-3-phosphate dehydrogenase	IISNASCTTNCLAPLAK	0,539423
RecName: Full=Glyceraldehyde-3-phosphate dehydrogenase	YKGEVTASGDDLVIDGHK	0,467038
glyceraldehyde-3-phosphate dehydrogenase type 2	TASGNIIPSSTGAAK	0,624896
glyceraldehyde-3-phosphate dehydrogenase	SGVDYVIESTGVFTK	0,648411
glyceraldehyde-3-phosphate dehydrogenase	AASYEEIAQAIK	0,569182
DEHA2G14058p	EALDLIMDAIDK	0,678204
RecName: Full=Enolase; AltName: Full=2-phospho-D-glycerate hydro-lyase; AltName: Full=2-phosphoglycerate dehydratase	AVANVNDIIAPALIK	0,581644
translation elongation factor 1a	TLLEAIDAIEPPTRPTDKPLR	0,535668
glyceraldehyde-3-phosphate dehydrogenase, type I, partial	DIEVVAVNDPFIAPDYAAYMFK	0,682683
2-phospho-D-glycerate hydro-lyase, putative; 2-phosphoglycerate dehydratase, putative; enolase (1), putative	LGANAILGVSLAAANAAAAAQGIPLYK	0,538769
hypothetical protein CaO19.3997	IVGLSDLPEVFK	0,623176
enolase 1	IGSEVYHNLK	0,510705
conserved hypothetical protein	TLAETAQEYVEVAK	0,479796
conserved hypothetical protein	SSTFHSNLK	0,465335

Tablica 9. Proteini dobiveni pretragom proteoma *Candida albicans* SC5314 i potvrđeni MS2 spektrima (nastavak)

Ime proteina	Identificirani tripsinski digesti	Vriijednost kosinusa
fructose-bisphosphate aldolase	DALYTSPETVFAVYESLHK	0,537782
enolase	IQIVGDDLTVTNPTR	0,661362
protein SNO4	TGVFVVEALHPFEVFR	0,585723
hypothetical protein CaO19.8442	DDFFDDFWK	0,578446
conserved hypothetical protein	AAEYVSGVVTGATEGAK	0,549642
pyruvate decarboxylase	LIEETQFPVFTTPMGK	0,606749
hypothetical protein CaO19.10395	IDLSLHPNDPESQTEVIETVEK	0,581439
hypothetical protein CaO19.11135	DGEIFLLENLR	0,4666
alcohol dehydrogenase 1	IPAGTDLANVAPILCAGVTVYK	0,603762
heat shock protein 31, putative; probable chaperone protein Hsp31 homologue	GGVVSAVCHGPAIFENLNDPK	0,499338
Pgk1p	TIVWNGPPGVFEFEK	0,588924
hypothetical protein CaO19.10395	ISTVGELNDLFADK	0,593518
phosphoglycerate kinase	MPIGDSLFDEAGAK	0,565836
conserved hypothetical protein	RDDFFDDFWK	0,527553
hypothetical protein CaO19.3160	INEAITPESEK	0,584664
hypothetical protein CaO19.8442	YGAGAGHPHR	0,541627
YALI0D22352p	TTPSFVAFTDTER	0,517202
Heat shock protein SSA1, partial	SINPDEAVAYGAAVQAAILTGDTSSK	0,457149
29 kDa IgE-binging protein	ITGFTDIGEDILGVTDIMK	0,704386
RecName: Full=Elongation factor 1-beta; Short=EF-1-beta	SIVTLDVKPWDDETDLDELLTNVK	0,509135
enolase	HIANISNAK	0,661487
Phosphoglycerate mutase 1	YADVDPAVVPLTESLALVIDR	0,687526
Malate dehydrogenase, mitochondrial precursor	VAVLGAGGGIGQPLSLLLK	0,49644
Hypothetical protein SPAPDRAFT_67925	GISELGIYPAVDPLDSK	0,517796
Heat shock protein 60, mitochondrial precursor	GEFTDMIAAGIIDPFK	0,507723
Retrotransposon Tca2 gag protein	YFHVAYPDVLEFLLDYNPK	0,560077

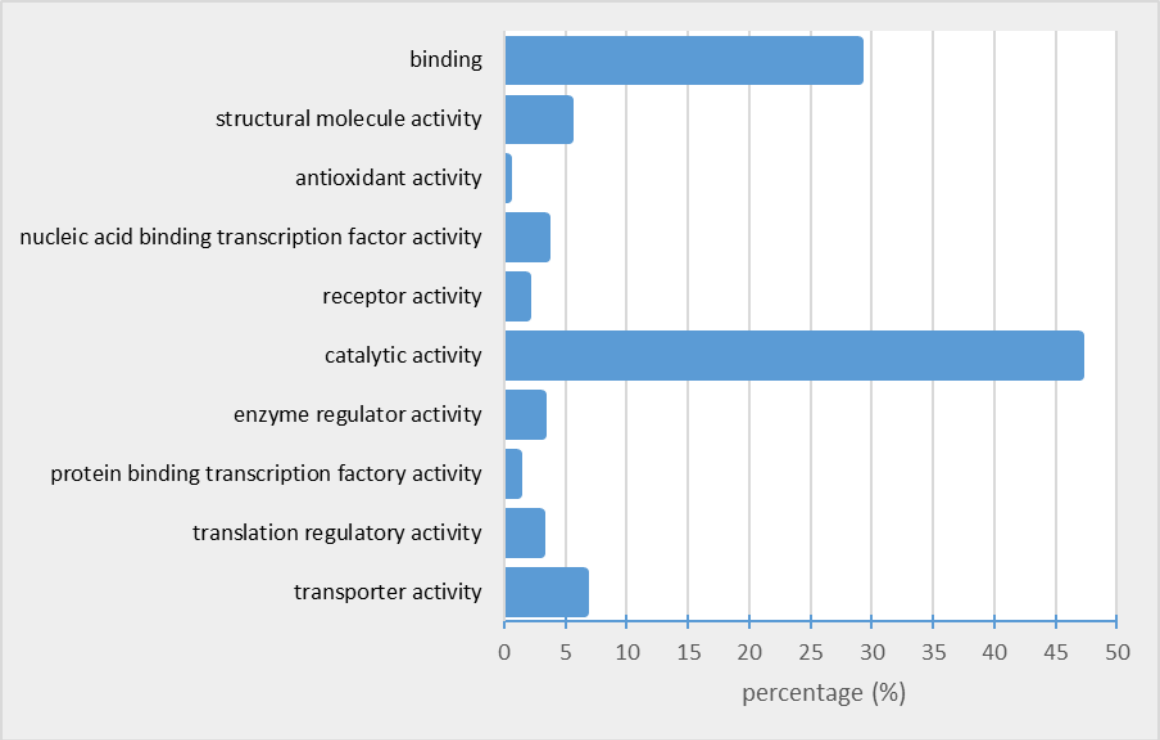
Tablica 9. Proteini dobiveni pretragom proteoma *Candida albicans* SC5314 i potvrđeni MS2 spektrima (nastavak)

Ime proteina	Identificirani tripsinski digesti	Vrijednost kosinusa
Elongation factor 3	DVQTAASDALLAIASAITPTAVK	0,488324
hypothetical protein	SAADIVFLAPGLSAIIDALK	0,494499
conserved hypothetical protein	AVGSNVPDNQK	0,498885
hypothetical protein CaO19.12088	APVGNPEGADKPNKK	0,454749
proteinase	DGYEFDIIPGEDYSDSDDDEDDSDTR	0,517342
conserved hypothetical protein	IPQRPPK	0,495066
potential GPI-protein transamidase complex subunit	LGLVPKAIAILPR	0,585619
hypothetical protein CaO19.8766	LLLSLLLSLLLLLLLKK	0,504549
potential Pip2-like fungal zinc cluster transcription factor	VIWLLIK	0,600669

4.3. ANALIZA GENSKE ONTOLOGIJE

Kao što je prethodno rečeno, svrha genske ontologije je omogućiti što bolju analizu i uvid u generalne biološke mehanizme u koje su pojedini geni uključeni te omogućiti istraživački pristup na razini sustavne biologije.

Na slici 10 prikazani su rezultati analize genske ontologije proteomski identificiranih proteina prema kriteriju molekularne funkcije, dobiveni putem PANTHER alata za funkcionalnu klasifikaciju.

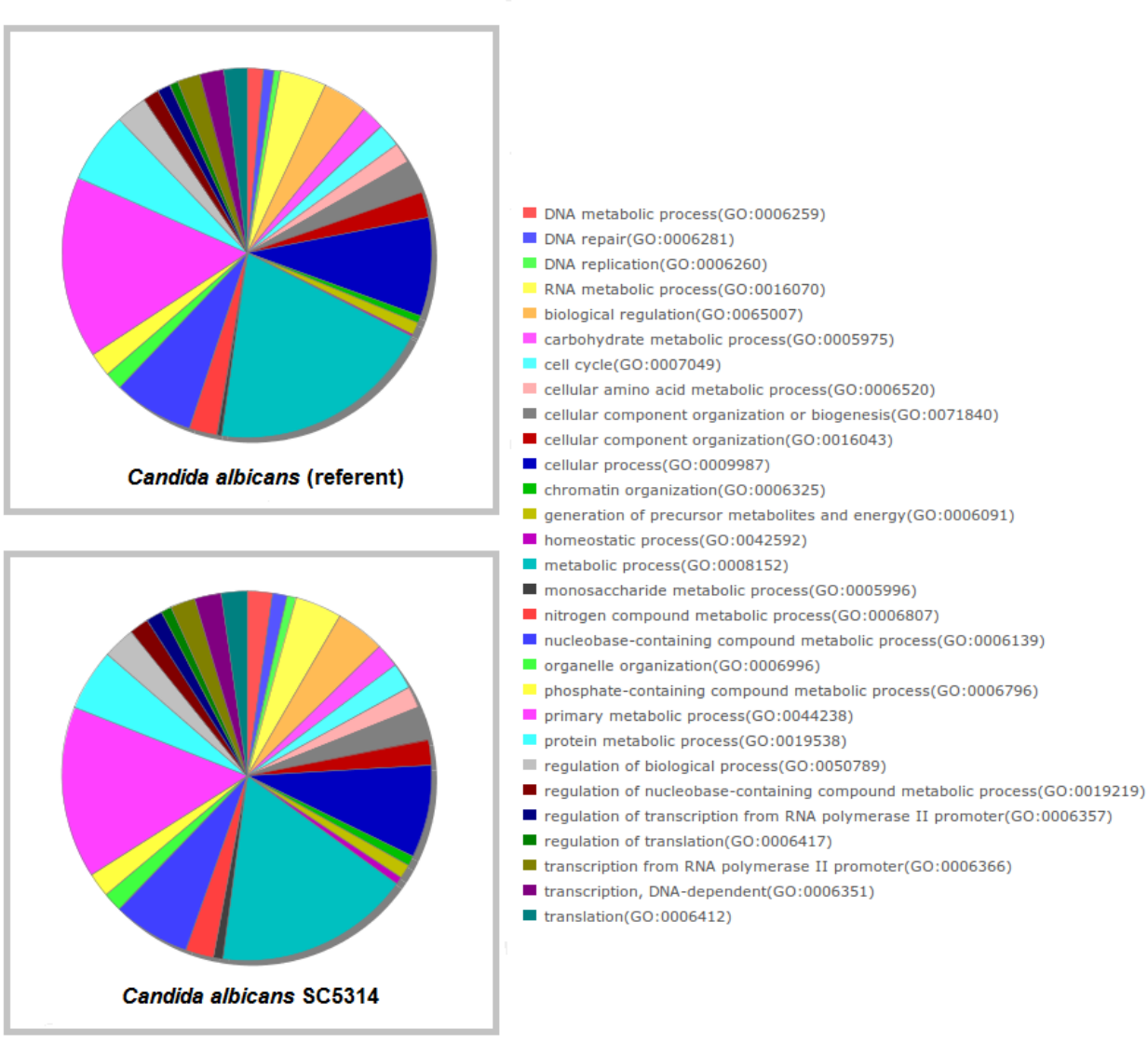


Slika 10. Analiza genske ontologije proteomski identificiranih proteina prema kriteriju molekularne funkcije

Analizom MS2 spektara identificirano je 757 proteina. Genska ontologija pokazuje raspodjelu tih proteina prema njihovoj molekularnoj funkciji. Proteini koji obavljaju katalitičke funkcije u stanici poput hidrolaza, oksidoreduktaza i transferaza pokazali su se, prema kriteriju molekularne funkcije, kao najviše zastupljena skupina proteina.

Također su bili visoko zastupljeni i proteini koji sudjeluju u procesima vezivanja u stanici poput DNA i RNA *binding* proteina, transkripcijskih faktora te nukleotidno vezanih proteina.

Za određivanje značajne zastupljenosti pojedinih funkcionalnih pojmova (GO pojmova), a time i zastupljenosti skupina gena i njihovih produkata u analiziranim stanicama kvasca *Candida albicans* SC5314 korišten je „overrepresentation“ test temeljen na binomnoj razdiobi. Ulazna lista koju su činili GI brojevi (*engl*. gene identifier, GI) proteina identificiranih MS/MS analizom uspoređena je s referentnom listom koju je činio referentni genom kvasca *Candida albicans* (ukupan broj gena u PANTHER bazi podataka za kvasac *Candida albicans*).



Slika 11. Zastupljenost pojedinih bioloških procesa u stanici kvasca *Candida albicans* SC5314 (u odnosu na referentni genom kvasca *Candida albicans*)

Rezultati analize genske ontologije pokazali su značajnu zastupljenost GO pojmova, prikazanih na slici 11, podijeljenih prema biološkoj funkciji. U tim prikazanim značajno zastupljenim biološkim procesima vidljiva je i veća zastupljenost detektiranih gena u odnosu na broj gena koji bi se očekivao za taj biološki proces (*engl*. expected value, računa se u odnosu na referentnu listu).

Najveću zastupljenost pokazali su geni koji kodiraju za proteine uključene u homeostatske procese, metabolizam monosaharida, popravak i replikaciju DNA, stanični ciklus te organizaciju kromatina.

Iz rezultata se može zaključiti da su, u trenutku uzorkovanja, stanice kvasca bile u eksponencijalnoj fazi rasta i nisu bile izlagane nikakvim stresnim uvjetima. Stanice su rasle na bogatoj hranjivoj podlozi koja je sadržavala glukozu kao izvor ugljika te peptone kao izvor dušika. Kvasac je metabolizirao glukozu kao primarni izvor ugljika i energije, zbog čega je vidljiva veća zastupljenost gena koji kodiraju za proteine uključene u metabolizam monosaharida odnosno u proces glikolize.

Eksponencijalna faza rasta karakterizirana je konstantnim rastom stanica i visokim stupnjem dobivanja energije. Stanice se prilikom faze rasta dijele, dolazi do replikacije DNA, transkripcije i reorganizacije kromatina pa je potrebna veća ekspresija gena koji kodiraju za proteine koji sudjeluju u tim procesima (DNA vezujući proteini, histoni, transkripcijski faktori).

4.4. DOPRINOS I OČEKIVANA PRIMJENA ISTRAŽIVANJA

Glavno istraživačko pitanje na koje ovaj diplomski rad nastoji dati odgovor jest postoji li mogućnost jednoznačnog određivanja vrste/soja mikroorganizama na temelju analize proteoma putem spektrometrije masa i računalne obrade snimljenih spektara. Rezultati proizišli iz ovog rada ukazuju na to da postoji.

Konkretno, u okviru ovog diplomskog rada, provjeravana je potpuno nova metoda detekcije vrste pa čak i sojeva mikroorganizama pomoću tandemne spektrometrije masa i novorazvijenog računalnog sustava. Ciljani krajnji korisnici ove metode nalaze se u klinici, veterini i svim ostalim ustanovama i laboratorijima za koje je važno analizirati mikrobni sastav bilo da je riječ o prehrambenoj namirnici ili tkivu. Dosadašnji postupci uglavnom se oslanjaju na uzgoj na selektivnim podlogama i u odnosu na ovu tehnologiju znatno su sporiji

i/ili nedovoljno precizni, stoga bi ova metoda mogla dovesti do poboljšanja u ustaljenoj praksi kao i značajnih ušteda.

U usporedbi s postojećim bioinformatičkim platformama, uočeno je da ne postoji adekvatno rješenje koje bi omogućavalo korištenje bez prethodnog sužavanja područja pretrage, dok se rješenje razvijeno u Kabinetu za bioinformatiku može koristiti i na najvećim bazama kao što je Genbank nr i bez ikakvog prethodnog znanja o analiziranom uzorku.

5. ZAKLJUČCI

1. ProteinLynx Global Server nije prikladan alat za određivanje vrste mikroorganizama; uspješnost identifikacije mikroorganizama uvelike ovisi o veličini pretraživane baze podataka proteinskih sekvenci.

2. Metodom koja je razvijena u Kabinetu za Bionformatiku uspješno je određena vrsta odnosno soj analiziranog mikroorganizma.

3. Analiza proteina rezultirala je 54 proteina potvrđena MS2 spektrima uz pomoć kojih je nedvojbeno potvrđeno da je kod analiziranog uzorka bilo riječ o kvascu *Candida albicans* SC5314. Nakon dobivene potvrde, spuštanjem vrijednosti kosinusa na 0,35 dobiveno je ukupno 757 proteina kvasca *Candida albicans* SC5314 potvrđenih MS2 spektrima.

4. Analiza genske ontologije na temelju potvrđenih 757 proteina i usporedbe s referentnim genomom kvasca *Candida albicans* ukazuje na zastupljenost bioloških procesa i pripadajućih gena koji ukazuju na normalne metaboličke aktivnosti (kvasac nije bio izlagan stresu i analiziran je u fazi rasta).

6. LITERATURA

Aebersold, R., Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198-207.

Anderson, N. L., Anderson, N. G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**, 1853–1861.

Armengaud, J. (2013) Microbiology and proteomics, getting the best of both worlds. *Environ. Microbiol.* **15**, 12-23.

Baranašić, D., Žučko, J., Dimnić, J., Gacesa, R., Long, P. F., Cullum, J., Hranueli, D., Starčević, A. (2014) Predicting substrate specifity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *J. Ind. Microbiol. Biot.* **41**, 461-467.

Bateman, R. H., Carruthers, R., Hoyes, J. B., Jones, C., Langridge J. I., Millar, A., Vissers, J. P. (2002) A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J. Am. Soc. Mass Spectr*. **13**, 792–803.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Sayers, E. W. (2011) GenBank. *Nucleic Acids Res.* **39**, 32–37.

Berman, H. M. (2008) The Protein Data Bank: a historical perspective. Acta crystallographica. Section A. *Found. Crystallogr*. **64**, 88–95.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. **31**, 365–370.

Bogdanov, B., Smith, R. D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev*. **24**, 168-200.

Bothner, B., Siuzdak, G. (2004) Electrospray ionization of a whole virus: analyzing mass, structure and viability. *ChemBioChem*. **5**, 258-260.

Bradford, M. M. (1976) A rapid and sensitive for the quantitation of microgram quantitites of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254.

Cagney, G., Amiri, S., Premawaradena, T., Lindo, M., Emili, A. (2003) In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Sci*. **1**, 1-15.

Cindrić M., Marković A., Horvatić A. (2009) Spregnute tehnike tekućinski kromatograf-spektrometar masa: osnove metodologije i primjene. *Medicina* **45**, 218-232.

Couto, B. R., Ladeira, A. P, Santos, M. A. (2007) Application of latent semantic indexing to evaluate the similarity of sets of sequences without multiple alignments character-by-character. *Genet. Mol. Res.* **6**, 983–999.

Cox, J., Mann, M. (2009) Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. *J. Am. Soc. Mass Spectr.* **20**, 1477-1485.

Craig, R., Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., Harshman, R. (1990) Indexing by latent semantic analysis. *J. Am. Soc. Inform. Sci.* **41**, 391-407.

Dole, M., Mach, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. P., Alice, M. B. (1968) Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240-2249.

Duncan, M. V., Aebersold, R., Caprioli, R. M. (2010) The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.* **28**, 659-664.

Edwards, N. J. (2011) Protein identification from tandem mass spectra by database searching. *Methods Mol. Biol.* **694**, 119-138.

Ekman, R., Silberring, J., Westman-Brinkmalm, A., Kraj, A. (2009) Mass Spectrometry: Instrumentation, Interpretation, and Applications, John Wiley & Sons, Inc., New York, SAD.

El-Aneed, A., Cohen, A., Banoub, J. (2009) Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Appl. Spectrosc. Rev.* **44**, 210-230. doi: 10.1080/05704920902717872

Eng, J. K., McCormack, A. L., Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectr*. **5**, 976–989.

Eriksson, J., Fenyo, D. (2004) The statistical significance of protein identification results as a function of the number of protein sequences searched. *J. Proteome Res.* **3**, 979–982.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science*. **246**, 64-71.

Galić, N. (2004) Elektroraspršenje - ionizacija u masenoj spektrometriji. *Kem. Ind.* **53**, 117–123.

Galić, N., Cindrić, M. (2008) Analiza proteina spektrometrijom masa. *Kem. Ind.* **57,** 231-243.

Gene Ontology Consortium (2001) Creating the gene ontology resource: design and implementation. *Genome Res*. **11**, 1425-1433.

Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G. Z., Gorenstein, M. V., Bateman, R. H., Langridge, J. I. (2009) The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683-1695.

Glish, G. L., Vachet, R. W. (2003) The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discov*. **2**, 140-150.

Gross, J. H. (2011) Mass spectrometry: a textbook, 2. izd., Springer-Verlag, Berlin Heidelberg.

Hoffmann, E., Stroobant, V. (2007) Mass spectrometry: Principles and applications, John Wiley & Sons Ltd, Chichester.

Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, *S*., Hauer, C. R. (1986) Protein sequencing by tandem mass spectrometry. *P. Natl. Acad. Sci.USA.* **83**, 6233-6237.

Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., Watson, J. T. (1987) Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem*. **59**, 2621–2625.

Kapp, E., Schütz, F. (2007) Overview of Tandem Mass Spectrometry (MS/MS) Database Search Algorithms. *Curr. Protoc. Protein Sci*. **49**, 1-19.

Karlsson, R., Gonzales-Siles, L., Boulund, F., Svensson-Stadler, L., Skovbjerg, S., Karlsson, A., Davidson, M., Hulth, S., Kristiansson, E., Moore, E. R. B. (2015) Proteotyping: Proteomic characterization, classification and identification of microorganisms - A prospectus. *Syst. Appl. Microbiol.* **38**, 246-257.

Kellner, R., Lottspeich, F., Meyer, H. E. (1999) Microcharacterization of proteins, 2. izd., Wiley-VCH Verlag GmbH, Weinham, str. 213-230.

Kinter, M., Sherman, N. E. (2005) Protein Sequencing and Identification Using Tandem Mass Spectrometry, John Wiley & Sons, New York.

Klinčić, D., Herceg Romanić, S. (2011) Kemijske metode određivanja hidroksiliranih metabolita policikličkih aromatskih ugljikovodika i poliklorbifenila u biološkome materijalu. *Arth. Hig. Rada. Toksiko*. **62**, 77-89.

Lane, C. S. (2005) Mass spectrometry-based proteomics in the life sciences. *Cell. Mol. Life Sci*. **62**, 848–869.

Law, K. P., Lim, Y. P. (2013) Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Rev. Proteomic.* **10**, 551–566.

Levander, F., Rognvaldsson, T., Samuelsson, J., James, P. (2004) Automated methods for improved protein identification by peptide mass fingerprinting. *Proteomics* **4**, 2594–2601.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R. (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682.

Magrane, M., UniProt Consortium (2011) UniProt Knowledgebase: a hub of integrated dana. *Database* **2011**, 1-13.

Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerchkhove, J., Apweiler, R. (2005) PRIDE: The proteomics identifications database. *Proteomics* **5**, 3537-3545.

Matthiesen, R., Azevedo, L., Amorim, A., Carvalho, A. S. (2011) Discussion on common data analysis strategies used in ms-based proteomics. *Proteomics* **11**, 604-619.

McPolin, O. (2009) An introduction to HPLC for Pharmaceutical Analysis, Mourne Training services, Warennpoint, str. 3-20.

Mi, H., Muruganujan, A., Casagrande, J. T., Thomas, P. D. (2013) Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*. **8**, 1551-1566.

Michalski, A., Cox, J., Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10**, 1785–1793.

Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J. J., Cox, J., Horning, S., Mann, M., Makarov, A. (2012) Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol. Cell. Proteomics* **11**, 1-28.

Nesvizhskii, A. I. (2007) Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol. Biol*. **367**, 87-119.

Nesvizhskii, A. I. (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics*. **73**, 2092–2123.

Niggeweg, R., Kocher, T., Gentzel, M., Buscaino, A., Taipale, M., Akhtar, A., Wilm, M. (2006) A general precursor ion-like scanning mode on quadrupole-TOF instruments compatible with chromatographic separation. *Proteomics* ***6***, 41–53.

Novakova, L., Matysova, L., Solich, P. (2006) Advantages of application of UPLC in pharmaceutical analysis. *Talanta* **68,** 908-918.

O'Donovan C., Martin M. J., Gattiker A., Gasteiger E., Bairoch A., Apweiler R. (2002) High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform*. **3**, 275-284.

Pappin, D. J., Hojrup, P., Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. biol.* **3**, 327–332.

Parviainen, V. (2014) Mass spectrometry in clinical protein biomarker discovery. Academic disertation. University of Helsinki.

Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry dana. *Electrophoresis* **20**, 3551-3567.

Pruitt, K. D., Tatusova, T., Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. **33**, 501–504.

Purvine, S., Eppel, J. T., Yi, E. C., Goodlett, D. R. (2003) Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **3**, 847–850.

Reinders, J., Lewandrowski, U., Moebius, J., Wagner, Y., Sickmann, A. (2004) Challenges in mass spectrometry-based proteomics. *Proteomics* **4**, 3686-3703.

Roepstorff, P., Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601.

Rognvaldsson, T., Hakkinen, J., Lindberg, C., Marko-Varga, G., Potthast, F., Samuelsson, J. (2004) Improving automatic peptide mass fingerprint protein identification by combining many peak sets. *J. Chromatogr. B.* **807**, 209–215.

Ruvinsky, A., Marshall Graves, J. A. (2005) Mammalian genomics, CABI Publishing, UK.

Seger, C., Griesmacher, A. (2007) Some important aspects of implementing tandem mass spectrometry in a routine clinical laboratory environment. *Biochem. Medica.* **17**, 29-51. doi: http://dx.doi.org/10.11613/BM.2007.004

Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., Schaeffer, D. A. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics*. **6**, 1638-1655.

Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V., Li, G. Z., Richardson, K., Wall, D., Geromanos, S. J. (2006) Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: A sweet tale. *Mol. Cell. Proteomics* **5**, 589–607.

Skoog, D. A., West, D. M., Holler, J. J. (1999) Osnove analitičke kemije. Školska knjiga, Zagreb.

Slade, S. E. (2013) Application of label-free mass spectrometry-based proteomics to biomarker discovery. Doktorska disertacija, University of Warwick.

Steen, H., Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Mol. Cell Biol*. **5**, 699-711.

Swartz, M. E. (2005) UPLC™: An Introduction and Review. *J. Liq. Chromatogr. R. T.* **28**, 1253-1263.

Talkington, M. T., Siuzdak, G., Williamson, J. R. (2005) An assembly landscape for the 30S ribosomal subunit. *Nature* **438**, 628-632.

Thomas, P. D, Kejariwal, A., Guo , N., Mi, H., Campbell, M. J., Muruganujan, A., Lazareva-Ulitsky, B. (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res*. **34**, 645-650.

Tian, Z., Tolić, N., Zhao, R., Moore, R. J., Hengel, S. M, Robinson, E. W., Stenoien, D. L., Wu, S., Smith, R. D., Paša-Tolić, Lj. (2012) Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* **13**, 2-9.

University of Bristol (2005) High Performance Liquid Chromatography Mass Spectrometry (HPLC/MS), <http://www.bris.ac.uk/nerclsmsf/techniques/hplcms.html> Pristupljeno 15. prosinca, 2015.

Vučinić, S. (2012) Proteomski pristup istraživanju molekularnih mehanizama razvoja makularne degeneracije na modelu štakora *in vivo*. Doktorska disertacija, Sveučilište Josipa Jurja Strossmayera u Osijeku.

Waters (2012) Analysis of Milk and Egg Allergens in Wine Using UPLC-MS, <https://www.waters.com/webassets/cms/library/docs/local_seminar_presentations/NMSS_44_FoodEnv_Gledhill.pdf.>. Pristupljeno 8. prosinca, 2015.

Waters (2015) Beginners Guide to UPLC, <http://www.waters.com/waters/en_HR/UPLC---Ultra-Performance-Liquid-Chromatography-Beginner%27s-Guide/nav.htm?cid=134803622&locale=en_HR>. Pristupljeno 4. prosinca 2015.

Waters (2015) How Does High Performance Liquid Chromatography Work?, < http://www.waters.com/waters/en_HR/How-Does-High-Performance-Liquid-Chromatography-Work%3F/nav.htm?cid=10049055&locale=en_HR>. Pristupljeno 12. prosinca 2015.

Wells, J. M., McLuckey, S. A. (2005) Collision-induced dissociation (CID) of peptides and proteins. *Method. Enzymol*. **402**, 148–185.

Wilkins, M. R., Appel R. D., Van Eyk, J. E., Chung, M. C. M., Görg, A., Hecker, M., Huber L. A., Langen, H., Link, A. J., Paik, Y. K., Patterson, S. D., Pennington, S. R., Rabilloud, T.,

Simpson, R. J., Weiss, W., Dunn, M. J. (2006) Guidelines for the next 10 years of proteomics. *Proteomics* **6**, 4-8.

Washburn, M. P., Wolters, D., Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol*. **19**, 242–247.

Wu, C., Nebert, D. W. (2004) Update on genome completion and annotations: Protein Information Resource. *Hum. Genomics*. **1**, 229–233.

Zhang, H., Cui, W., Wen, J., Blankenship, R. E., Gross, M. L. (2010) Native electrospray and electron-capture dissociation in FTICR mass spectrometry provide top-down sequencing of a protein component in an intact protein assembly. *J. Am. Soc. Mass. Spectrom.* **21**, 1966-1968.

Zhang, N., Aebersold, R., Schwikowski, B. (2002) ProbID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* **2**, 406–1412.

Zhu, S., Li, L., Thornton, C., Carvalho, P., Avery, B. A., Willett, K. L. (2008) Simultaneous determination of benzo[a]pyrene and eight of its metabolites in Fundulus heteroclitus bile using ultraperformance liquid chromatography with mass spectrometry. *J. Chromatogr. B.* **863**, 141-149.

7. PRILOZI

Svi navedeni prilozi uključujući i cjelovit tekst Diplomskog rada (Diplomski rad SU.pdf) nalaze se na presnimljivom kompaktnom disku (CD-R) pod nazivom „Diplomski rad SU: PRILOZI“.

Na disku se nalaze slijedeći podaci:

1. Rezultati identifikacije proteina putem PLGS alata
   - Candida albicans_PLGS.xlsx
     - Sheet 1: SwissProt
     - Sheet 2: SwissProt *Candida albicans*
     - Sheet 3: Uniprot *Candida albicans*

2. Proteini dobiveni pretragom proteoma *Candida albicans* SC5314 i potvrđeni MS1 spektrima
   - Candida_proteins_new.xlsl

3. Cjeloviti rezultati analize genske ontologije („overrepresentation test“) i ulazna lista pritom korištena
   - PANTHER overrepresentation test.xlsx
   - Candida_ESI_proteins.txt