Baze podataka iz područja molekularne biologije – pretraživanje i bibliometrijska analiza znastvenih radova i časopisa

Vujnović, Sandra

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:593339

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-02-06**

Sveučilište u Zagrebu Filozofski fakultet University of Zagreb Faculty of Humanities and Social Sciences

Repository / Repozitorij:

ODRAZ - open repository of the University of Zagreb Faculty of Humanities and Social Sciences

zir.nsk.hr

dabar DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
SMJER BIBLIOTEKARSTVO

Ak. god. 2022./2023.

Sandra Vujnović

Baze podataka iz područja molekularne biologije – pretraživanje i bibliometrijska analiza znastvenih radova i časopisa

Diplomski rad

Mentor: prof. dr. sc. Sonja Špiranec

Zagreb, prosinac 2022.

Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

Sandra Vujnović
(potpis)

Zahvaljujem mentorici prof. dr. sc. Sonji Špiranec na stručnosti, savjetima, razumijevanju i potpori tijekom izrade ovog diplomskog rada.

Zahvaljujem svim profesorima na nesebičnom dijeljenju znanja i korektnosti.

Zahvaljujem svojoj obitelji koja je na ovom putu, putu koji zovemo život, uz mene.

Svima veliko hvala!

Sadržaj

Sadržaj ii
1. Uvod 1
2. Baze podataka 3
2.1. Baze podataka iz područja molekularne biologije 5
2.2. GenBank 8
2.3. UniProtKB 9
2.3.1. Gene Ontology (GO) 11
2.4. InterPro 12
2.5. PubMed 15
2.6. Europe PMC 16
3. Bibliometrija 18
3.1. Bibliometrijska analiza 18
3.2. Autorstvo i koautorstvo 18
3.3. Mjerenje znanstvene produktivnosti 19
3.4. Mjerenje znanstvene suradnje 20
3.5. Znanstvene institucije i zemlje 20
3.6. Časopisi 21
3.6.1. Procjena kvalitete časopisa 22
3.7. WoSCC - JCR 22
3.7.1. Faktor odjeka 23
3.8. Scopus – SCImago Journal & Country Rank 24
3.8.1. SCImago Journal Rank (SJR) 24
3.9. h-indeks 25
4. Osnove molekularne biologije 26
4.1. DNA – deoksiribonukleinska kiselina 26

4.1.1. Gen 26
4.2. Proteini 27
4.3. Projek ljudskog genoma 28
5. Istraživanje 29
5.1. Pretraživanje baze podataka UniProtKB 30
5.2. Bibliometrijska analiza znanstvenih radova 31
5.2.1. Godina publiciranja i područje istraživanja 31
5.2.2. Zemlje 33
5.2.3. Autorstvo, koautorstvo, timski rad 34
5.3. Bibliometrijska analiza časopisa 35
5.3.1. h-indeks 36
5.3.2. SCImago Journal Rank - kvartili (Q) 41
6. Zaključak 43
7. Literatura 46
Sažetak 49
Summary 50

1. Uvod

Otkrivanje novih spoznaja događa se na svakodnevnoj razini. Znanstveni instituti, istraživački centri, sveučilišta različitom metodologijom pridonose izgradnji svjetskog znanja. Zahvaljujući razvoju tehnologije danas se informacije prikupljaju i objedinjuju u različitim digitalnim formama.

Baze podataka predstavljaju takvu jednu riznicu informacija. Objedinjavanjem informacija otvara se mogućnost za kreiranje novih ideja odnosno novih spoznajnih putova, a izbjegava objavljivanje istih radova te skretanje u pogrešnim smjerovima. Zahvaljujući računalima i automatiziranoj obradi podataka više nego ikad informacijska pismenost znanstvenika i znanstvena komunikacija dolaze do izražaja. Jedna od glavnih zadaća znanstvenika postaje prepoznavanje vjerodostojnosti informacija te vrednovanje istih, a otvorenost znanstvene komunikacije unapređuje kvalitetu znanstvenih istraživanja, a time i razvoj znanosti.

Iako je znanja sve više bibliometrijske analize pokušavaju uvesti red. Različitim analizama i statističkim mjerenjima usmjeravaju znanstvenu produktivnost.

Molekularna biologija, u današnje doba najekspanzivniji smjer biologije, s informatikom čini koheziju u obliku bioinformatike. Mnoge molekule, procesi i rekacije postaju "vidljivi" zahvaljujući računalnoj tehnologiji. Više nego ikad ljudska znanja prodiru u samu srž postojanja, u dio koji čini nas i druge organizme onakvima kakvi jesmo. U sam zapis života.

Ovo je vrijeme intenzivnih istraživanja na molekularnoj razini, a znanstveni radovi koji pritom nastaju grade zapise u bazama podataka, čine članke u znanstvenim časopisma i na kraju omogućuju bibliometriji da kroz statističku analizu mjeri važnost, značaj i kvalitetu samog istraživanja. Dobiveni rezultati postaju temelj za neke nove spoznaje, radove, a ciklus se iznova gradi, ali sada na jednom višem spoznajnom nivou.

Ovaj diplomski rad objedinjuje elemente navedenog ciklusa.

Prvi dio rada opisuje što su baze podataka te naglašava važnost znanstvene komunikacije i informacijske pismenosti znanstvenika. Drugi dio prikazuje načine pretraživanja informacija u najznačajnijim biološkim bazama podataka: GenBank, UniProtKB, InterPro, PubMed i Europe PMC. Treći dio posvećen je spoznajama o bibliometriji i bibliometrijskoj analizi znanstvenih radova i časopisa. Četvrti dio opisuje najznačajnije pojmove iz molekularne biologije. Zatim slijedi istraživački dio na temu ljudskog inzulina. Na osnovu informacija o znanstvenim

radovima čija otkrića grade zapis o ljudskom inzulinu u bazi UniProtKB napravljena je bibliometrijska analiza.

Cilj rada je prikazati načine pretraživanja informacija u najvažnijim bazama podataka iz područja molekularne biologije i bibliometrijski analizirati publikacije čija otkrića grade zapis o ljudskom inzulinu u bazi UniProtKB te s bibliometrijskog aspekta vrednovati časopise koji su navedene publikacije objavili.

2. Baze podataka

Prema definiciji baze podataka su organizirane zbirke podataka u strojno čitljivom obliku. U kontekstu pretraživanja literature na internetu predstavljaju organizirane zbirke zapisa, dokumenata ili informacija koje su pohranjene i dostupne korisnicima kroz različite sustave za pretraživanje. Mogu biti komercijalne (naplaćuju korištenje) ili u otvorenom pristupu (besplatne) (Lapiš, 2018).

Lapiš (2018) dalje navodi da se prema sadržaju baze podataka dijele na: baze podataka s cjelovitim tekstom, citatne i bibliografske baze podataka.

Baze podataka s cjelovitim tekstom su zbirke cjelovitih tekstova, odnosno članaka i drugih radova iz različitih publikacija (elektronički časopisi, zbornici, patenti itd.).

Bibliografske baze podataka sadrže podatke o radovima objavljenim u različitim publikacijama, ali ne sadrže cjeloviti tekst već ponekad samo poveznicu prema cjelovitom tekstu. Ove baze pružaju informacije o autoru rada, publikaciji u kojoj je rad objavljen, godini objave i slične podatke koji mogu biti korisni prilikom pretraživanja pojedinih radova ili znanstvenih područja.

Citatne baze, u svojoj osnovi bibliografske, pored samih radova obrađuju i popise korištene literature koje autori navode na kraju svojih radova. Citatne baze često se pretražuju kako bi se pronašle informacije o tome koji su radovi najpopularniji, najčitaniji ili najviše citirani unutar nekog znanstvenog područja (Lapiš, 2018).

Pojam relevantnosti i kvalitete baze podataka za neko područje određen je stupnjem selektivnosti, ažurnosti obrade novih publikacija i pokrivenosti područja za koje je baza podataka deklarirana (Jokić, 2005).

Računala, mreže, unaprijeđeni istraživački instrumenti te velika količina automatizacije doveli su do eksplozije podataka dobivenih znanstvenim istraživanjem. Takav pristup dovodi do potpuno novog načina provođenja znanstvenog istraživanja koji više ne podrazumijeva da istraživač najprije razvije hipotezu koju onda testira prikupljanjem nužnih podataka, već se za stvaranje ili testiranje hipoteze koristi obiljem već postojećih podataka (Stojanovski, 2015).

Više nego ikad očituje se važnost znanstvene komunikacije i informacijske pismenosti znanstvenika.

Zahvaljujući informacijskoj pismenosti znanstvenik pronalazi i prepoznaje informacije potrebne u istraživačke svrhe, analizira ih i vrednuje, koristi ih ili odbacuje te usmjerava istraživanje prema prikupljanju novih informacija ili nadogradnji postojećih.

Informacijska pismenost nije samo funkcionalna vještina već kompetencija koja se temelji na na etičkim vrijednostima, kritičkom mišljenju i razumijevanju informacijskih prostora. O informacijskoj pismenosti znanstvenika ovisi znanstvena komunikacija (Špiranec, 2015).

Znanstvena komunikacija je neizostavna sastavnica moderne znanosti i pokretač znanstvenoistaživačkih procesa. Podrazumijeva cijelu lepezu postupaka, od pristupa znanstvenim informacijama, njihova iskorištavanja, stvaranja, formalnog i neformalnog objavljivanja i diseminacije. Uobičajeno se znanstvena komunikacija određuje kao sustav kojim se istraživanja i znanstveni radovi stvaraju, vrednuju i prenose znanstvenoj zajednici te čuvaju za buduća korištenja (Špiranec, 2015). Autorica Stojanovski dijeljenje informacija opisuje kao srž znanstvene komunikacije (Stojanovski, 2015)

Dijeljenje istraživačkih podataka povećava mogućnost reproduciranja rezultata, njihovom ponovnom primjenom maksimalizira se vrijednost istraživanja, ubrzava napredak znanosti, osigurava transparentnost znanstvenih istraživanja, smanjuje mogućnost pristranosti pri interpretaciji rezultata i povećava kredibilitet objavljenih znanstvenih spoznaja (Stojanovski, 2015).

Ideja o slobodnom i neometanom pristupu znanstvenim publikacijama postaje međunarodnim pokretom i dobiva jasne i prepoznatljive ciljeve te strategiju ostvarivanja početkom 2002. godine, formuliranjem Budimpeštanske inicijative za otvoreni pristup (Budapest Open Access Initiative, 2002) (Melinščak Zlodi, 2018). Kako bi otvoren pristup znanstvenim publikacijama zaista postao pravilo u svijetu znanstvene komunikacije, potrebno je sinergijsko djelovanje i dobra volja svih glavnih dionika znanstveno-istraživačke i izdavačke djelatnosti - znanstvenika/autora, njihovih ustanova, tijela koja financiraju znanstvena istraživanja, izdavača znanstvene literature i država u kojima se sve ove aktivnosti događaju (Macan, 2018a).

Autori Vodopijevac i Kranjec (2018) navode primjer disiplina poput meteorologije, astronomije i genomike (područja genetike) gdje su koristi otvorenog pristupa vidljive u bržem spoznajnom napretku, smanjivanju financijskih troškova i općoj korisiti za zajednicu što i jest krajnja svrha znanosti.

2.1. Baze podataka iz područja molekularne biologije

Prve baze podataka iz područja molekularne biologije bile su specijalizirane, izolirane i s malom količinom podataka. Organizirali su ih i obrađivali stručnjaci iz različitih područja biologije. Kada su se podatci preklapali, primjerice podatci o redoslijedu aminokiselina zajednički su bazama podataka koje sadrže informacije o redoslijedu aminokiselina i bazama podataka koje opisuju strukturu proteina, postojao je relativno mali interes da se koriste kontrolirani rječnici kojima bi se postigla uniformnost. The International Scientific Unions i CODATA (Committee on Data of the International Council of Scientific Unions) dali su značajan doprinos rješavanju tog problema, ali problem i dalje postojao (Lesk, 2019).

Uslijed prepoznavanja važnosti bioinformatike u biološkim istraživanjima, dodatnog povećanja broja bioloških baza podataka te pojave interneta i World Wide Web-a krenulo se u smjeru integracije baza. Veliki financijski zahtjevi i potreba za spajanjem biološkog i računalnog znanja doveli su do razvoja nacionalnih i međunarodnih institucija odgovornih za obradu, organiziranje, skladištenje i čuvanje podataka (Lesk, 2019).

Lesk (2019) nadalje tvrdi da shvaćanje da različite baze podataka nisu zasebna spoznajna područja već područja koja treba međusobno uskladiti i povezati dovelo je do razvoja poveznica između baza. Takav pristup omogućio je:

1. sistematizaciju formata i rječnika kako bi podaci iz različiti baza (ili zbirki) bili ujednačeni
2. olakšan pristup srodnim materijalima putem pokazivača ili poveznica
3. razvoj softvera za pretraživanje informacija čime se pojednostavio put od preuzimanja podataka do njihove analize

Povećanjem broja i količine informacija u bazama podataka iz područja molekularne biologije pojavio se trend njihove integracije u još veće i obuhvatnije baze kako bi se omogućilo kombiniranje različitih kategorija informacija te okupljanja informacija na jednom mjestu (Lesk, 2019).

Neke intitucije održavaju baze podataka koje su fokusirane na pojedinačne organizme. Primjer je baza podataka FlyBase koja sadržava sve genske, molekularne i druge relevantne informacije za porodicu kukaca Drosophilidae s fokusom na modelni organizam Drosophila melanogaster – vinska mušica.

Svaka baze podataka sastoje se od informacija (zapisa), sheme odnosno organizacije tih zapisa i alata za pristup. Autor Lesk (2019) baze podataka iz područja molekularne biologije (biološke baze) dijeli u četiri skupine:

1.) arhivske (*archival)* baze podataka

Uključuju:

- redoslijed nukleotida i redoslijed aminokiselina, uključujući anotacije
- primarne stukture nukleinskih kiselina i proteina, uključujući anotacije
- baze podataka usmjerene na organizme, uključujuću pretraživanje genoma organizma
- baze podataka o ekspresiji proteina
- baze podataka o metaboličkim putevima
- baze podatka o međudjelovanju i regulatornim mehanizmima

2.) izvedene (*derived*) baze podataka

Sadrže informacije dobivene prikupljanjem i analizom podataka iz arhivske baze. Primjerice:

- sekundarna i tercijarna struktura proteina
- klasifikacija i međuodnosi proteina – porodica proteina

3.) bibliografske baze podataka

Sadrže:

- znanstvenu literaturu

4.) mrežne baze podatka

Sadrže:

- baze podataka o bazama podataka

Zbog velike brojnosti baza podataka iz područja molekularne biologije i načina na koji prezentiraju informacije, danas više ne postoji jasna granica između primarnih baza podataka (baze podataka nastale prikupljanjem informacija nastalih izravno eksperimentalnim putem u laboratorijima) i sekundarnih baza podataka (baze podataka koje obrađuju, povezuju i grade nove spoznaje na temelju analize informacija nastalih u primarnim bazama) (Lesk, 2019).

U nastavku rada bit će prikazane sljedeće biološke baze: GenBank, UniProtKB, InterPro, PubMed i Europe PMC. Većinu navedenih baza autor A. Lesk (2019) navodi i izdvaja u svom djelu *Introduction to bioinformatics*.

Autor Lesk (2019) navodi da podatke o redoslijedu aminokiselina održavaju baze: 1.) GenBank sa sjedištem u NCBI (National Centar for Biotechnology Information), USA; 2.) European Nucleotide Archive (ENA) sa sjedištem u EBI (European Bioinformatics Institute), UK; 3.) Centar for Information Biology and DNA Data Bank of Japan, sa sjedištem u National Institute of Genetics, Japan (Lesk, 2019). U radu će biti prikazana baza GenBank jer je prva osnovana od navednih tri baza koje predstavljaju arhivu nukleotidnih sekvencija.

Baza United Protein Database (UniProtKB) održava arhivu redoslijeda aminokiselina u proteinu, redoslijed je određen gotovo isključivo translacijom sekvencije nukleotida, a nastala je ujedinjavanjem baza: SWISS-PROT, The Protein Identification Resource (PIR) i Translated EMBL (TrEMBL) (Lesk, 2019).

Baza InterPro je glavna ili „krovna“ sekundarna baza podataka. Objedinjuje sadržaje i anotocije nekoliko individualnih baza podataka koje sadrže podatke o porodici proteina, strukturi ili aktivnim mjestima na proteinu te ujedno sadrži i poveznice prema drugim izvorima (Lesk, 2019).

Lesk (2019) dijeli baze podataka u četiri skupine, a kao primjer bibliografske baze navodi PubMed. Također, prethodno navedene baze nude poveznice prema bazi PubMed stoga je radom obuhvaćena i ova baza.

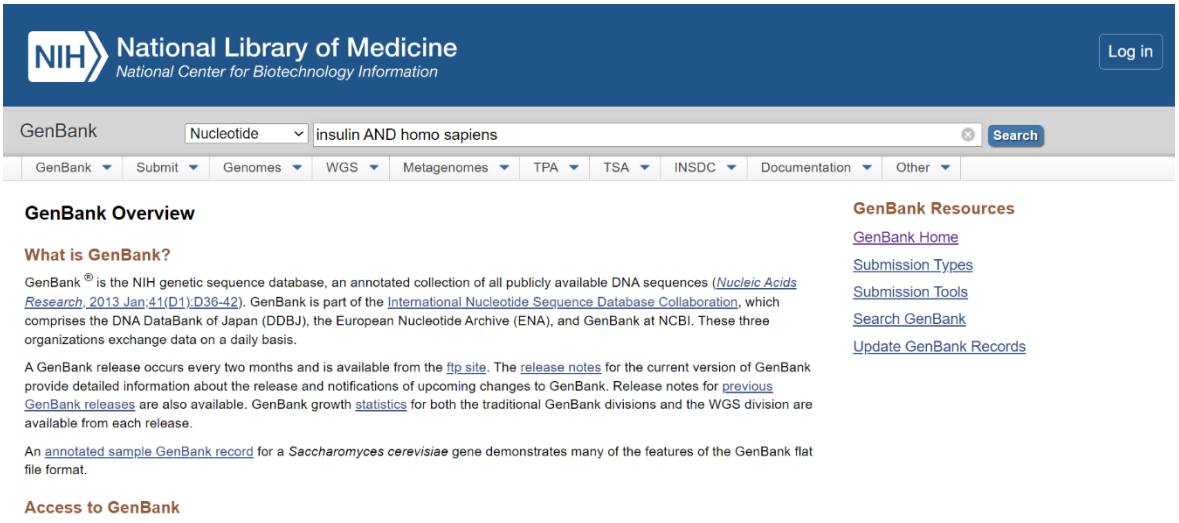
Baze UniProtKB i InterPro nude poveznicu prema bazi cjelovitog teksta iz područja biologije - Europe PMC. Iz tog razloga ovim radom prikazana je i ta baza.

2.2. GenBank

Temeljne baze podataka iz područja molekularne biologije opisuju redoslijed nukleotida i redoslijed aminokiselina jer nukleotidi čine osnovnu jedinicu građe nukleinskih kiselina (DNA i RNA), a aminokiseline čine osnovnu jedinicu građe proteina.

Podatke o redoslijedu nukleotida održavaju: GenBank, European Nucleotide Archive (ENA) i Centar for Information Biology and DNA Data Bank of Japan. Zajedno su ujedinjeni pod nazivom INSDC (International Nucleotide Sequence Database Collaboration). Na dnevnoj razini razmjenjuju pristigle radove o novim nukleotidnim sekvencijama iz različitih laboratorija diljem svijeta. Na taj način baze jednako pokrivaju nukleotidne podatke, ali se međusobno se razlikuju po formatu, anotacijama i poveznicama.

Na naslovnoj strani baza GenBank nudi osnovno pretraživanje. U polje za pretraživanje upisuje se tekst (naziv gena, naziv proteina, naziv bolesti, ime znanstvenika). Za detaljnije opisivanje traženog pojma u polju za pretraživanje moguće je korisiti Booleove operatore AND, OR i NOT. Primjerice za pretraživanje informacija o redoslijedu nuklotida u genu za ljudski inzulin u polje pretraživanja upisujemo: insulin AND Homo sapiens. (Slika 1)



Slika 1. Zaslon naslovne strane biološke baze GenBank

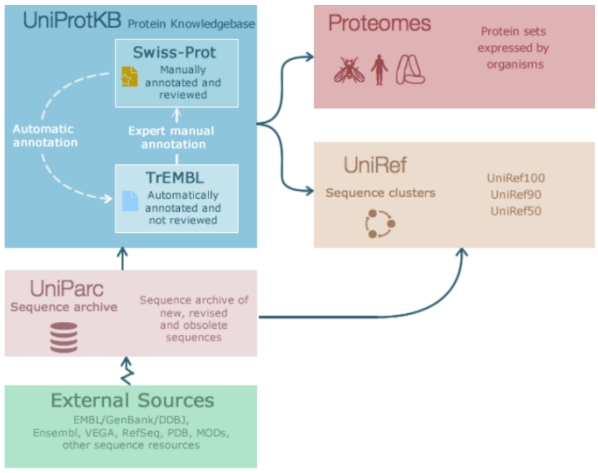
Nakon izlistavanja rezultata pretraživanja, s lijeve strane postoje filteri za dodatno sužavanje izbora. Filteri su: skupina organizama (carstva živih bića + virusi), vrsta nukleinske kiseline, odabir baze podataka, vrsta i duljina sekvencije, odabir staničnih dijelova koji sadrže genski

zapis, podatci o datumu objave podatka. Filteri postoje i s desne strane. Nakon izbora određene skupine organizama koristeći filter na lijevoj strani, desna strana nudi popis najčešće istraživanih organizama koji toj skupini pripadaju.

Odabirom željenog zapisa otvara se zaslon s različitim informacijama i anotacijama vezanim za traženu nukleotidnu sekvenciju. Svaki zapis unutar baze podataka GenBank ima svoj identifikacijski broj, nazvan pristupni kod koji je jedinstven unutar cijelog INSDC-a. Zapis čine tri dijela. U prvom dijelu su opći podatci o željenoj sekvenciji koji uključuju: broj parova baza, vrsta i stuktura nukleinske kiseline na kojoj se nalazi tražena sekvencija, datum objavljivana sekvencije, opis što sekvencija prikazuje i koliko puta je nadopunjena, informacije o publikaciji i poveznica prema radu u PubMed-u. Drugi dio opisuje samu sekvenciju kroz numeričke podatke o duljini kodirajuće regije, eksonu, mRNA, genu i sl. Treći dio je kompletna sekvencija odnosno točan raspored nukleotida u traženom genu. Zapis nudi dvije poveznice - FASTA i Graphics. Većina softvera korisiti program odnosno alat FASTA za analizu preklapanja sekvencija različiti organizama. Graphics daje grafički prikaz sekvencije gena (https://www.youtube.com/watch?v=QCgSYQnmVuc, 2020).

2.3. UniProtKB

UniProt (Universal Protein Resource) je slobodno dostupan izvor podataka (repozitorij) o građi i funkciji proteina. UniProt baze podataka su: UniProtKB (UniProt Knowledgebase), UniRef (UniProt Reference Clusters), UniParc (UniProt Archive) i Proteomes. (Slika 2)



Slika 2. Grafički prikaz UniProt baze podataka

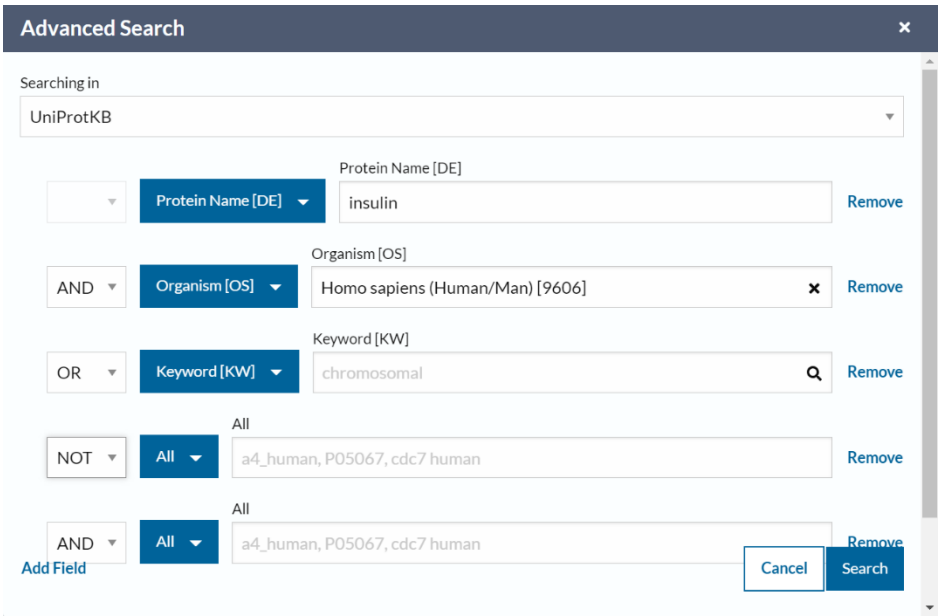
Centralalna baza podataka u UniProt-u je UniProt Knowledgebase (UniProtKB). UniProtKB je biološka baza koja sadrži podatake o strukturi aminokiselina i funkciji proteina. Sastoji se od dva dijela: Swiss-Prot i TrEMBL (https://www.uniprot.org/help/uniprotkb, 2022).

U Swiss-Prot dijelu informacije o svakom zapisu su pregledane i ručno označene od strane stučnjaka. Postupak ručnog označivanja i pregledavanja detaljno je definiran kako bi se osiguralo da se svim ručno obrađenim unosima postupa na jednak način. Proces se sastoji od obveznih koraka koji uključuju: odabir redoslijeda, analiza redoslijeda, odabir literature, navođenje izvora informacija, osiguranje kvalitete i objedinjavanje unosa. Obradu provode biolozi uz pomoć niza alata koji su razvijeni u bliskoj suradnji s računalnim stručnjacima.

U TrEMBL dijelu zapisi su računalno automatski generirani i obogaćeni automatskim bilješkama i klasifikacijama. (https://www.youtube.com/watch?v=VC-tCkfvAHs)

Baza nudi dva načina pretraživanja: osnovno i napredno. U osnovnom pretraživanju u polje za pretraživanje upisuje se tekst (naziv gena, proteina, vrste organizma i sl.) npr. inzulin.

U naprednom pretraživanju odabire se područje pretraživanja (baza podataka, ime proteina, organizam, ključna riječ, ime gena, patologija, unutarstanični smještaj, genska ontologija itd.), upisuje se željeni pojam u odabrano područje pretraživanja te se vrši pretraživanje. Koristeći "add field" otvaraju se dodatna polja koja sužavaju pretragu. Pretraživanja se provodi pomoću Booleovih operatora AND, OR i NOT. (Slika 3)



Slika 3. Prikaz zaslona naprednog pretraživanja u bazi UniProtKB

Nakon izlistavanja rezultata pretraživanja, s lijeve strane postoje filteri za dodatno sužavanje izbora čime se postiže manji odziv i veća preciznost rezultata. Filteri su: rezultati recenzirani od strane stručnjaka (Swiss-Prot), nerecenzirani rezultati (TrEMBL), rezultati dobiveni na „popularnim“ organizmima (do sada genski najviše istraživani organizmi: čovjek, miš, štakor, vinska mušica, riba zebrica), taksonomija, proteini „s“ ili „koji sadrže“, anotacije, duljina redoslijeda. Rezultati se mogu preuzeti na računalo.

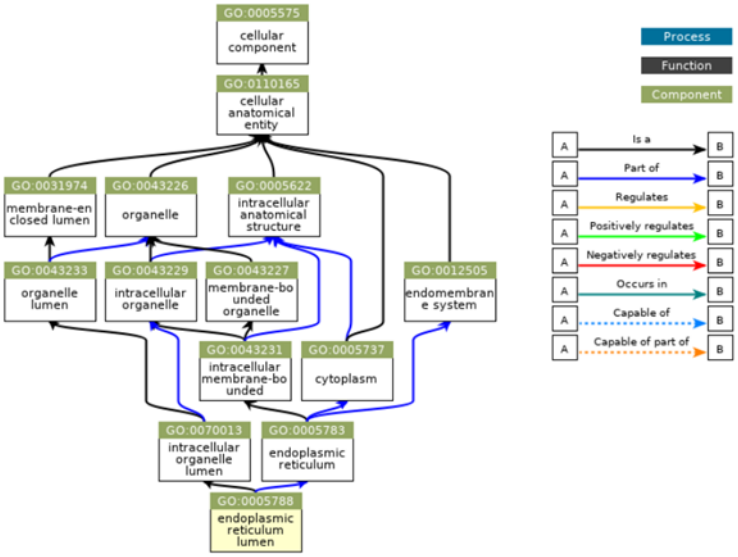
Odabirom željenog zapisa otvara se zaslon s različitim informacijama i anotacijama vezanim za traženi protein. Uz svaki podatak nalazi se poveznica prema: izvoru znanstvene informacije (primjerice prema: znanstvenom članku objavljenom u PubMed-u ili u Europe PMC-u), alatima za analizu redoslijeda aminokiselina (BLAST) i vanjskim bazama podataka uključujući Gene Ontology (GO).

2.3.1. Gene Ontology (GO)

Genska ontologija (GO) je bioinformatička inicijativa koja okuplja podatke u istoimenoj bazi podataka (GO). Sadrži informacije o produktima gena (proteini i nekodirajuće RNA) i njihovim ulogama u različitim organizmima (od bakterija do čovjeka) sa svrhom postizanja dosljednosti u opisima genskih produkata u različitim bazama podataka. Baza se redovito ažurira i nadopunjuje.

Achburner et al. (2000) navode da je „ cilj Gene Ontology Consortiuma proizvesti dinamičan, kontrolirani rječnik koji se može primijeniti na gotovo sva živa bića premda se znanja o ulogama gena i proteina u stanicama na svakodnevnoj razini gomilaju i mijenjaju. U tu svrhu konstruirane su tri neovisne domene: biološki proces (BP - *biological process*), molekularna funkcija (MF- *molecular function*) i stanična komponenta (CC- *cellular component*)“.

Svaki pojam koji se nalazi u sklopu ontološke baze podataka opisan je i postavljen u relacijske odnose. Trenutno je opisano oko 43 000 pojmova u sve tri domene. Relacijskim odnosima („ dio je“, „je“, „regulira“, „sposobnost za“, „događa se u“) stvara se ontološka struktura – DAG (Directed Acyclic Graph) (youtube.com/watch?v=6Am2VMbyTm4&t=2254s,). (Slika 4)



Slika 4. Prikaz ontološke stukture za pojam *lumen endoplazmatskog retikuluma* u bazi GO

2.4. InterPro

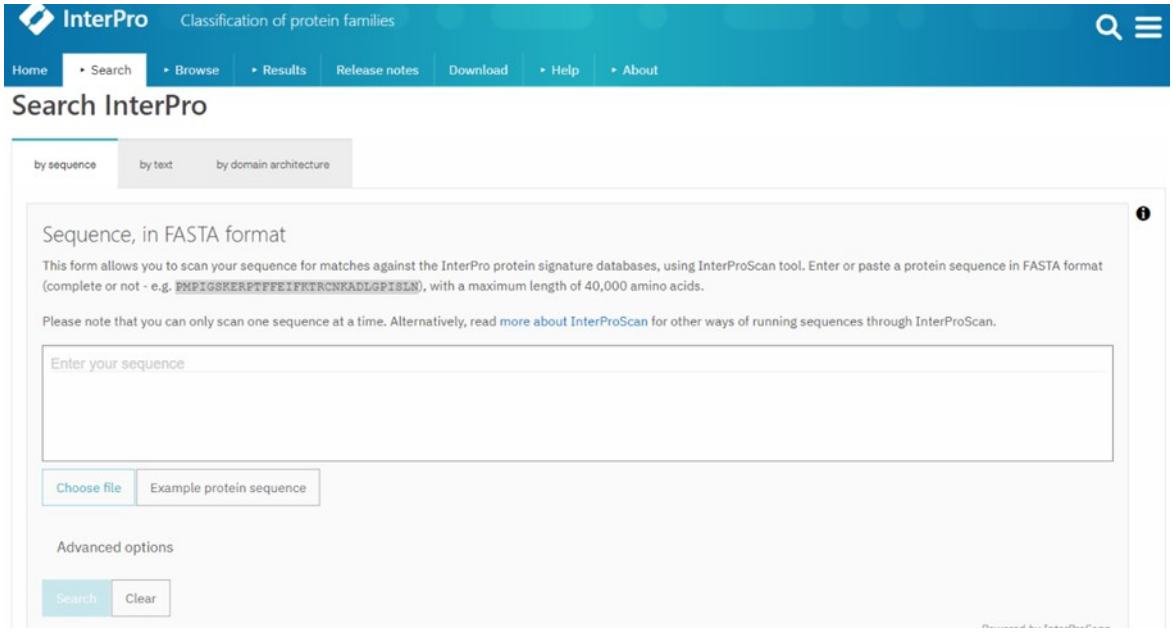
Mnoge sekundarne baze podataka nastale su prikupljanjem i analizom primarnih podataka. Sekundarne biološke baze podataka sadrže informacije o: porodici proteina, strukturi ili aktivnim mjestima na proteinu. Glavna ili „krovna“ sekundarna baza je InterPro. Sve informacije i podatci u bazi InterPro su u slobodnom pristupu i ažuriraju se svakih 8 tjedana. Datum zadnje izmjene vidljiv je na naslovnoj strani korisničkog sučelja.

Na naslovnoj stranici baza InterPro nudi tri različita načina pretraživanja: pretraživanje pomoću: redoslijeda aminokiselina, teksta i strukturne domene proteina (https://www.youtube.com/watch?v=5dw03Bp7_HM, 2020). (Slika 5)

Za pretraživanje pomoću redoslijeda aminokiselina u polje za pretraživanje upisuje se slijed aminokiselina nekog proteina ili peptida. Podatak o slijedu aminokiselina koji gradi neki protein ili peptid nalazi se u primarnim bazama podataka.

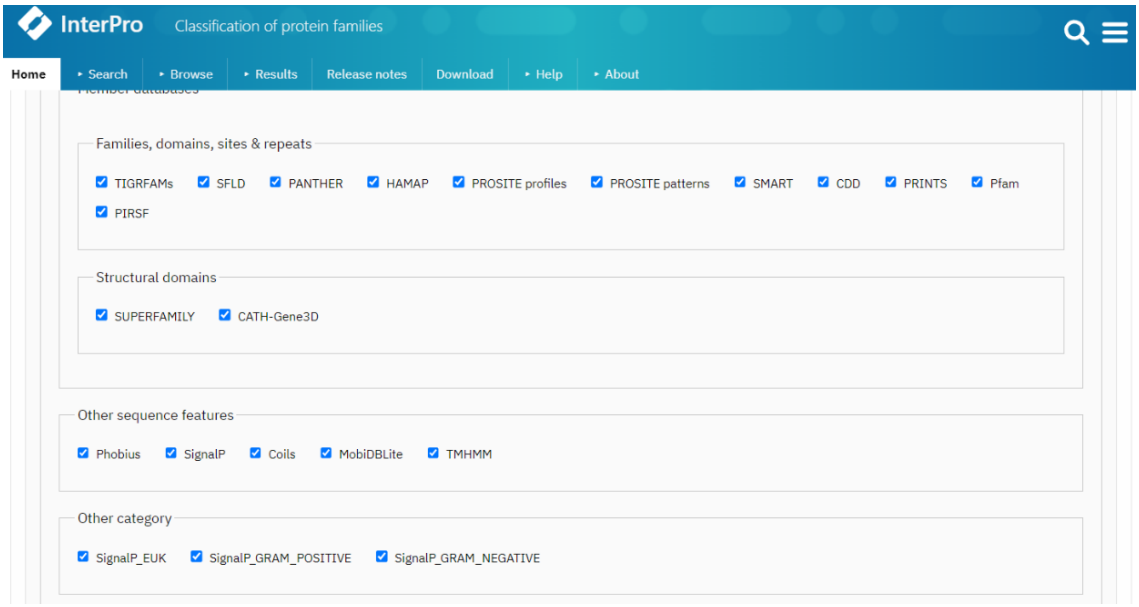
Primjerice za pretraživanje informacija o ljudskom inzulinu u polje pretraživanja upisujemo redoslijed aminokiselina, a ne tekst: insulin AND Homo sapiens. Redoslijed glasi:

„MALWMRLLPL LALLALWGPD PAAAFVNQHL CGSHLVEALY LVCGERGFFY TPKTRREAED LQVGQVELGG GPGAGSLQPL ALEGSLQKRG IVEQCCTSIC SLYQLENYCN“



Slika 5. Prikaz korisničkog sučelja baze InterPro

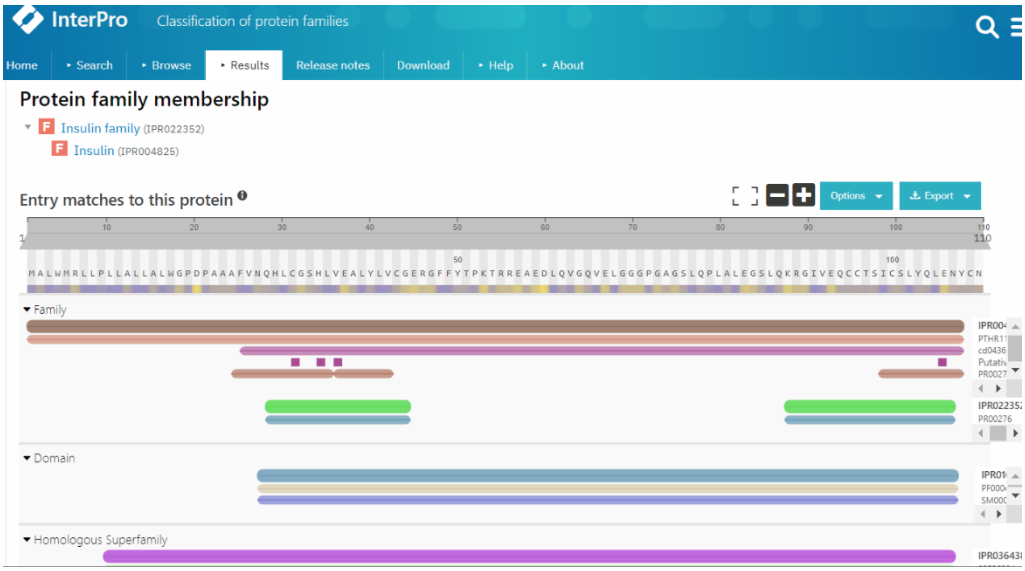
Napredno pretraživanja pomoću redoslijeda nudi opciju pretraživanja u svim ili samo određenim bazama podataka i pretraživanje prema značajkama određenog rasporeda aminokiselina u proteinu sa svrhom predviđanja karakteristika proteina (topljivosti, mjesta cijepanja signalnih peptida, položaj zavojnice i dr.). (Slika 6)



Slika 6. Prikaz izgleda zaslona za napredno pretraživanje u bazi InterPro

Otvaranjem željenog zapisa može se doznati: peptidna (proteinska) struktura, kojoj porodici protein pripada, hijerarhijska strukturu proteina (kojoj nadporodici pripada) te podatke o biološkim procesima u kojima protein sudjeluje, molekularnoj funkciji i lokaciji u stanici. Biološki procesi, funkcija i stanični dio (lokacija) objašnjeni su pojmovno kroz gensku ontologiju (GO).

Informacije su prikazane u obliku sekvencije (rasporeda aminokiselina) (Slika 7).

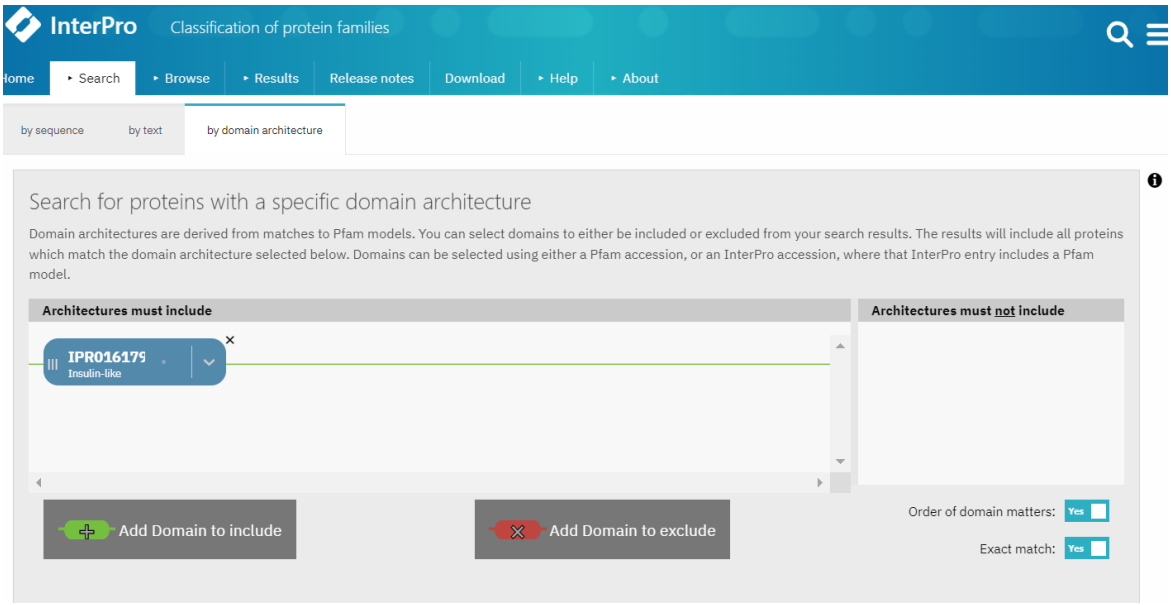


Slika 7. Prikaz rezultata pretraživanja pomoću sekvenije aminokiselina za ljudski inzulin u biološkoj bazi InterPro

Drugi način pretraživanja je pretraživanje pomoću teksta. U polje pretraga upisuje se riječ (npr. „inzulin“) ili pristupna oznaka zapisa. Ova opcija ne nudi mogućnost naprednog pretraživanja. Pretraživanjem pomoću riječi otvara se željeni zapis iz kojeg se može doznati: pristupni broj proteina, pristupni broj/evi srodnih zapisa (npr. pristupni broj prema obitelji tog proteina), opisne informacije o proteinu (uloga, struktura, vrste živih bića u kojima je zastupljen, poznate bolesti vezane za taj protein), popis referenci, unakrsne reference, genska ontologija (GO). Unutar zapisa smješten je glavni izbornik koji nudi pristup dodatnim informacijama o strukturi proteina, kemijskim reakcijama u kojima sudjeluje, skupine živih bića koje također posjeduju taj protein u većoj ili manjoj količini strukturnog preklapanja. Informacije su prikazane u tekstualnom obliku.

Treći način pretraživanja je pretraživanje pomoću domene proteina. Domene su kratke regije (dijelovi) proteina koje imaju određenu funkciju. Ova opcija ne nudi mogućnost naprednog pretraživanja. U polje pretraživanja upisujemo pristupni broj domene koja nas zanima pritiskom gumba „Add domain to include“. Pristupni broj tražene domene doznaje se uz pomoć pretraživanja po sekvenciji gdje se u kategoriji „domena“ može doznati pristupni broj za određenu domenu.

Pretraživnje nudi i filtere: važnost podudaranja redoslijeda domena (u slučaju da se pretražuju više domena jednog proteina) i odabir rezultata s potpuno preklapanjem domena. Rezultati će uključivati sve proteine različitih organizmima koji odgovaraju odabranoj domeni. (Slika 8)



Slika 8. Prikaz rezultata pretraživanja pomoću domene proteina za ljudski inzulin u biološkoj bazi InterPro

2.5. PubMed

PubMed je sustav za pretraživanje, razvijen i održavan od strane Nacionalnog centra za biotehnološke informacije (NCBI). Dostupan je javnosti od 1996. godine. Sadrži više od 34 milijuna citata i sažetaka biomedicinske literature. Ne uključuje cjelovite članke iz časopisa već poveznice na puni tekst dostupan iz drugih izvora, kao što su web stranice izdavača ili PubMed Central (PMC).

PubMed uključuje: Medline, PubMed Central i Bookshelf.

Medline je glavna bibilografska baza podatka Nacionalne medicinske knjižnice SAD-a (US National Library of Medicine). Medline predstavlja najveću komponentu PubMed-a.

PMC je druga po veličini komponenta PubMeda. PMC je arhiva cjelovitog teksta koja uključuje članke iz časopisa koje je pregledala i odabrala za arhiviranje Nacionalna medicinska knjižnica.

Bookshelf je arhiva punog teksta knjiga, izvješća, baza podataka i drugih dokumenata koji se odnose na biomedicinu, zdravlje i znanost o životu (https://pubmed.ncbi.nlm.nih.gov/about/).

Bibliografskoj bazi Medline može se pristupiti putem korisničkog sučelja Ovid i Web of Science.

Ovid sučelje nudi osnovno pretraživanja, pretraživanje citata, pretraživanja unutar polja pretraživanja, napredno pretraživanja te pretraživanje više različitih polja pretraživanja. Pretraživanje u Ovid sučelju može se svesti na tri osnovne strategije: korištenje ključnih riječi i /ili kontroliranog rječnika MeSH, korištenje povijesti pretraživanja u svrhu kombiniranja rezultata pretraživanja pomoću Booleovih operatora AND i OR te korištenje graničnika u svrhu smanjenja rezultata pretrage.

Web of Science sučelje nudi osnovno i napredno pretraživanje. Kod osnovnog pretraživanja može se pretraživati više pojmova u velikom broju polja pretraživanja (tema, naslov, autor, naziv publikacije, godina objave, MeSH rječnik, jezik, vrsta organizma, PubMed ID, vrsta publikacije itd.) te koristiti vremenski graničnik. Pojmovi se povezuju Booleovim operatorima AND, OR i NOT. Napredno pretraživanje nudi također polja za pretraživanja s mogućnošću dodatnog indeksiranja. Korištenjem povijesti pretraživanja mogu se kombinirati rezultati pretraživanja pomoću Booleovih operatora AND i OR te koristiti vremenski graničnik. Željeni rezultati mogu se označiti i time kreirati vlastitu bazu dokumenata.

Baš kao i u Ovid sučelju pretraživanje u Web of Science se može svesti na tri osnovne strategije: korištenje ključnih riječi i/ili kontroliranog rječnika MeSH, korištenje povijesti pretraživanja u svrhu kombiniranja rezultata pretraživanja pomoću Booleovih operatora AND i OR te korištenje graničnika u svrhu smanjenja rezultata pretrage.

2.6. Europe PMC

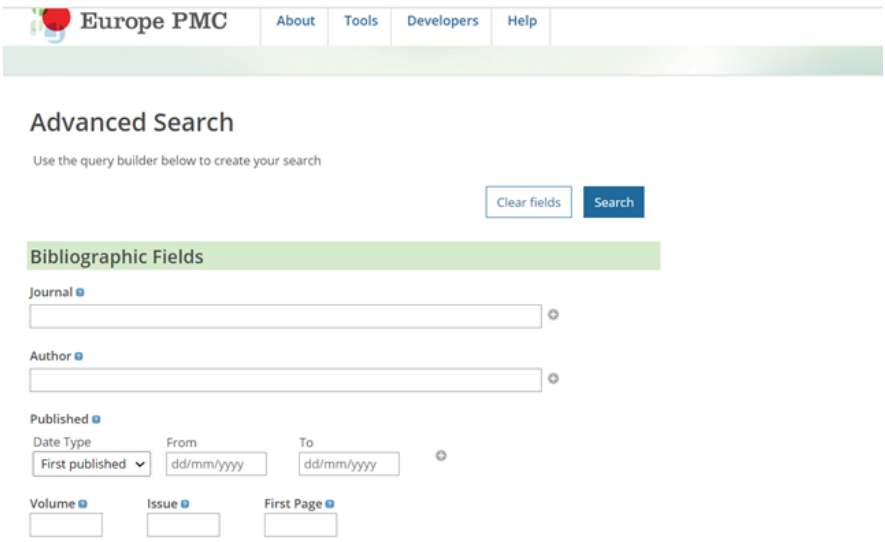
Europe PMC je baza podataka koja sadrži 41,3 milijuna sažetaka i 8,3 milijuna članaka s punim tekstom iz područja biologije. Objedinjuje sadržaje PubMed-a, PubMed Central (PMC),

Agricole i još 25 poslužitelja koji objavljuju biološke podatke. Europe PMC održava EBI (Europski bioinformatički institut) - međunarodna, inovativna i interdisciplinarna istraživačka organizacija čiji je cilj učiniti svjetske javne biološke podatke besplatno dostupnima znanstvenoj zajednici. EBI je u sastavu EMBL-a (European Molecular Biology Laboratory) (https://europepmc.org/About, 2022).

Europe PMC nudi dva načina pretraživanja: osnovno i napredno.

U osnovnom pretraživanju u polje za pretraživanje upisuje se tekst npr. inzulin. Nakon izlistavanja rezultata pretraživanja, na sredini korisničkog sučelja nalazi se opcija sortiranja. Rezultati se mogu sortirati prema: broju dobivenih citata, datumu unosa u bazu, relevantnosti i datumu objave u časopisu. S lijeve strane postoje filteri za dodatno sužavanje izbora: vrsta publikacije, besplatna dostupnosti punog teksta i vremenski period.

U naprednom pretraživanju ispunjavaju se područja pretraživanja: bibliografski podatci, poveznice prema drugim bazama podataka, anotacije. (Slika 9)



Slika 9. Prikaz izgleda zaslona za napredno pretraživanje u bazi EuropePMC

Odabirom željenog zapisa otvara se zaslon s informacijama o članku. Odmah ispod naslova članka, na sredini zaslona, smještene su osnovne informacije: autor(i), naziv časopisa koji je objavio rad, datum objave, DOI te poveznice za dijeljenje znanstvenog članka putem društvenih mreža. S lijeve strane postoje poveznice prema: sažetku, punom tekstu (ako postoji opcija), popisu i broju referenci, citatima i utjecaju (broj citata, grafički prikaz citiranja kroz vrijeme, poveznica prema publikacijama koje su citirale rad, altmetrijski pokazatelji) te poveznica prema sličnim člancima.

3. Bibliometrija

Pojam bibliometrija sadrži dva korijena: “biblio” i “metric”. Riječ “biblio” dolazi od grčkog biblion, što je slično riječi Byble (os), a znači knjiga. Riječ “metric” dolazi od latinskog ili grčkog “metricus” ili “metrikos” što znači znanost o mjerenju odnosno mjerenje (Danesh i Mardani-Nejad, 2021).

Pojam prvi rabi Alan Pritchard 1969. godine umjesto, po njegovom mišljenju, nespretnog pojma stataističke bibliografije. Definirao je bibliometriju kao primjenu matematičkih i statističkih metoda na knjige i ostale medije u zanstvenom komuniciranju sa svrhom određivanje prirode i smjera razvoja određenog znanstvenog područja (Hebrang- Grgić, 2017).

Autori Danesh i Mardani-Nejad (2021) navode da je prvi rad o bibliometrijskoj analizi objavio francusko-švicarski botaničar Alphonse de Candolle 1873. godine. U radu je proučavao faktore koji utječu na znanstvenu uspješnost neke zemlje.

Nakon drugog svjetskog rata bibliometrijske studije doživjele su značajan razvoj uslijed razvoja računalne tehnologije.

3.1. Bibliometrijska analiza

Bibliometrijskim analizama može se vrednovati znanstvena produkcija autora, institucije, zemlje, razvoj neke discipline prema broju časopisa, problematika koju obrađuju i sl. Kao izvor podataka za navedena istraživanja koriste se baze podataka. Rezultati istraživanja korisni su znanstvenicima, knjižničarima odnosno informacijskim stručnjacima, ali i ljudima koji donose odluke u znanstvenoj politici zemlje (Jokić, 2005).

Kreatori znanstvene publikacije prema autorici Jokić (2005) su autori, institucije i zemlje stoga se bibliometrijska istraživanja mogu provesti analizirajući navedene podatke u bibliografskim i citatnim bazama podataka.

3.2. Autorstvo i koautorstvo

Standardni model autorstva u znanstvenom publiciranju podrazumijeva objavljivanje rada kojega potpisuje jedan ili više autora. Tendencija suvremenog autorstva u većini znanstvenih

disciplina i područja je postupno i sve veće smanjenje broja jednoautorskih radova. U nekim područjima gotovo da više niti ne postoje jednoautorski radovi (Jokić, 2005).

Autorica Jokić u djelu *Bibliometrijski aspekti vrednovanja znanstvenog rada* (2005) navodi veliki broj primjera istraživanja temeljena na broju autora po radu u različitim znanstvenim područjima i u različitim vremenskim razdobljima. Sva ta istraživanja potvrđuju porast broja autora po radu i smanjenje broja jednoautorskih radova.

Autorstvo postaje kolektivna aktivnost posebno u STEM području zbog potrebe istraživanja u velikim specijaliziranim laboratorijskim timovima i korištenjem skupih visokospecijaliziranih laboratorijskih instrumenata.

Ovisno o smjeru istraživanja suradnja među autorima omogućuje povezivanje interdisciplinarni područja, razvoj novih poddisciplina te potiče na komunikaciju i otvorenost u znanosti.

Bibliometrijsko istraživanje autorstva je interesantno i s aspekta vrednovanja doprinosa svakog pojedinog autora u koautorstvu. Naročito se korisnim čini prijedlog vezan za procjenu autorstva odnosno autorskog udjela u radu navođenjem uloge svakog pojedinog koautora. Time bi se značajno olakšale procjene u procesu vrednovanja doprisnosa svakog autora u radu (Jokić, 2005).

3.3. Mjerenje znanstvene produktivnosti

Znanstvena produktivnost autora kao pokazatelja aktivnosti u istraživanjima i potencijalnom doprinosu razvoja, uobičajeno se mjeri kroz broj objavljenih radova, članaka, knjiga, inovacija kroz određeno vremensko razdoblje (Jokić, 2005).

Kao orijentir u znanstvenoj produktivnosti nezaobilazan je Lotkin zakon. Godine 1926. kemičar, matematičar i statističar Alfred James Lotka objavljuje članak o znanstvenoj produktivnosti autora. Lotka smatra da mali broj autora objavljuje veliki broj radova, a najveći broj autora objavi jedan ili dva rada unutar određenog područja (Jokić, 2005).

Prema Lotkinom zakonu ako se zna broj autora koji su napisali samo jedan članak, može se predvidjeti broj autora koji su napisali dva, tri, četiri i više članaka. Broj autora koji stvaraju n članaka aproksimativan je odnosu I: n^2. Prema Lotkinom zakonu ako je 100 autora napisalo po jedan članak, može se znati da je 25 autora napisalo po dva članka (25= 100: 2^2), 11 autora po

tri članka (11= 100: 3^2) itd. Takva distribucija karakteristična je za područje prirodnih znanosti, a u drugim područjima omjer je drugačiji (Hebrang Grgić, 2016).

Mjerenje znanstvene produktivnosti kada je riječ o jednom ili dva autora u radu ne predstavlja problem. Veći je problem vrednovanja doprinosa pojedinog autora u višeautorskim radovima. U analizi znanstvene produkcije tada se najčešće koriste dva metodološka pristupa. Prvi, gdje se jedan rad računa kao jedan rad za svakog autora, bez obzira koliko je autora na radu. Drugi pristup temelji se na mjerenju doprisnosa svakog pojedinog autora (Jokić, 2005).

3.4. Mjerenje znanstvene suradnje

Suradnja među znanstvenicima intenzivnije se počinje istraživati krajem sedamdesetih godina. Istraživani su različiti oblici suradnje i koautorstva. Veći dio istraživanja posvećen je institucionalnoj suradnji koja se može pratiti kroz međunarodne bibliografske i citatne baze podataka (Jokić, 2005).

Jedna od najpoznatijih mjera vidljivosti suradnje je utvrđivanje koautorstva na publikacijama i podatci o adresama institucija navedenih autora. Bibliometrijska analiza višeautorskih radova može se koristiti samo kao parcijalni indikator suradnje u znanstvenim istraživanjima, a kao specifični indikator može se koristiti u slučajevima kada je jasno istaknut doprinos svakog autora. Ujedno odnosi među autorima koji zajedno surađuju na istraživanjima znatno doprinose razumijevanju komunikacije u znanosti (Jokić, 2005).

Međunarodna suradnja ovisi o puno faktora, a jedan od osnovnih je veličina zemlje odnosno njen znanstveni potencijal. Također suradnja može biti uvjetovana "blizinom" među zemljama: fizičkom, geografskom, kulturološkom, jezičnom, povijesnom odnosno socioekonomskim faktorima (Jokić, 2005).

3.5. Znanstvene institucije i zemlje

Znanstveni instituti, fakulteti, sveučilišta, kao i pojedine zemlje ili regije koriste se bibliometrijskim metodama kako bi mjerile svoju znanstvenu produkciju i svoj znanstveni utjecaj u znanost.

Podatci dobiveni istraživanjem znanstvene produktivnosti sveučilišta i znanstvenih instituta, analizom citiranosti i mjerenjem međunarodne suradnje, pridonose planiranju znanstvene

politike određene zemlje. Planiranje znanstvene politike može značajno utjecati na promjenu statusa zemlje u svjetskom znanstvenom okružju. U znanstvenoj produkciji na svjetskoj razini ističu se: Francuska, Njemačka, Velika Britanija te Sjedinjene Američke Države i Japan (Jokić 2005).

3.6. Časopisi

Prvi znanstveni časopisi počeli su izlaziti 1665. godine (Journal des Sçavans i Philosophical Transactions of the Royal Society) i od tada do danas čine osnovu znanstvene komunikacije i jedan od glavnih izvora novih znanstvenih ideja. Uloga časopisa u znanstvenoj komunikaciji odvija se kroz produkciju, konzumaciju i diseminaciju znanstvenih i stručnih informacija relevantnoj znanstvenoj javnosti (Romić i Mitrović, 2017).

Iako je znanstvena komunikacija uvjetovanima i prirodom same znanosti i tehničkim dostignućima, ključna uloga u objavljivanju rezultata znanstvenih i stručnih istraživanja i dalje pripada časopisima, bilo u tiskanom ili digitalnom obliku (Krajna i Zabjan Bogut 2011).

Rezultati znanstvenih i stručnih istraživanja objavljuju se u časopisima u obliku znanstvenih članaka. Znanstveni članci prolaze kroz postupak vrednovanja koji provode recenzenti časopisa čime se vrši provjera: ispravne metodologije, interpretacije i zaključaka rezultata istraživanja opisanih u člancima (Lock, 1985 navedeno u Macan i Petrak, 2015).

Objavljivanjem rada u časopisu autoru je omogućeno predstavljanje znanstvenoj zajednici, a time i mogućnost stjecanja profesionalnog priznanja i napredovanja. Časopisi također štite intelektualno vlasništvo autora članka (Macan i Petrak, 2015).

Znanstveni časopisi imaju znatnu ulogu i u provođenju znanstvene politike. Vrijednosna prosudba časopisa može utjecati na odluke o financijskoj potpori znanstvenim projektima, na rangiranje akademske i znanstvene ustanove, napredovanje pojedinaca u znanstvenome i akademskom sustavu te na određivanje smjera u kojem će se znanstveno istraživanje i znanstveno otkriće nastaviti.

Od 1960-ih godina uočava se porast kvantitativnih istraživanja obilježja časopisa i radova/članaka koje oni objavljuju kao osnovnih jedinica na kojima se provode tzv. bibliometrijske analize (Macan i Petrak, 2015).

3.6.1. Procjena kvalitete časopisa

Zbog velikog broja časopisa i naglog razvoja informacijsko – komunikacijske tehnologije jedan od glavnih izazova znanstvenika je gdje objaviti rad, s obzirom na sve teže razlikovanje pouzdanog od nepouzdanog izvora te nestanka postavljenih granica formalne i neformalne komunikacije (Golenko, 2019).

Autorica Golenko (2019) navodi da se problem procjene časopisa za objavu znanstvenog rada može se riješiti na tri načina. Prvi način je procjena kvalitete časopisa na temelju njegovih karakteristika. Drugi način za odabir časopisa je vrednovanje njegove kvalitete uz pomoć bibliometrijske analize. Treće rješenje je odnosi se na utvrđivanje propisnih uvjeta za izbor u znanstvena zvanja u Republici Hrvatskoj, u kojima su utvrđeni kriteriji zastupljenosti (indeksiranosti) časopisa u bazama podataka propisanim *Pravilnikom o uvjetima za izbor u znanstvena zvanja* (NN, br. 28/2017, 72/2019 i 111/2022).

Prema *Pravilniku o uvjetima za izbor u znanstvena zvanja* (2017) u znanstvene radove ubrajaju se radovi objavljeni u znanstvenim časopisima zastupljenima u bazama podataka WoSCC (Web of Science Core Collection) i Scopus. To pravilo vrijedi za većinu znanstvenih područja u RH.

Web of Science Core Collection (WoSCC) i Scopus nude znastvenoj zajednici dva alata koja omogućuju vrednovanje znanstvenih časopisa putem bibliometrijskih karakteristika časopisa i rangiranje časopisa unutar određene skupine prema kvartilima (Q1, Q2, Q3 i Q4) (Golenko, 2019).

3.7. WoSCC - JCR

Journal Citation Reports (JCR) je baza podataka koja izlazi u dva izdanja: Science Edition (za prirodne znanosti) i Social Sciences Esition (za društvene znanosti).

JCR je snažan alat za analizu znanstvenih časopisa. Omogućuje više načina rangiranja, evaluacije, kategorizacije i komparacije znanstvenih časopisa putem citatnih pokazatelja koji se izračunavaju na temelju podataka o citiranosti radova iz WoSCC (Web of Science Core Collection) (Macan i Petrak, 2015).

Zbirka citatnih indeksa WoSCC dostupna je korisnicima od početka 2014. godine putem nove platforme -Web of Science (WoS).

Najvažniji citatni pokazatelji su: faktor odjeka (za razdoblje od dvije i pet godina), indeks brzine citiranja (engl. immediacy index), poluvijek citiranosti (engl. cited half-life), ukupan broj citata (engl. total cites), Eigenfactor score i pokazatelj odjeka članka (engl. article influence score, AI) (Macan i Petrak, 2015).

Časopisi uključeni u JCR raspodijeljeni su u određene predmetne kategorije ovisno o tematici koju pokrivaju. Razlog tomu je postojanje velike razlike u citatnom potencijalu različitih znanstvenih područja. Veliki citatni potencijal imaju časopisi iz područja biomedicine, fizike i kemije stoga bi časopisi iz tih područja bili stalno na vrhu popisa najcitiranijih časopisa (Macan i Petrak, 2015).

JCR na temelju podatka o citiranosti u bazi (WoSCC) časopise svrstane u određene predmetne kategorije dijeli u kvartile (Q1 – Q4) prema visini faktora odjeka. Pripadnost časopisa prvom kvartilu (Q1) znači da prema visini faktora odjeka pripada među 25 % časopisa s najvišim faktorom odjeka unutar skupine časopisa za određeno predmetno područje (Macan i Petrak, 2015). Ako časopis pripada drugom kvartilu (Q2) znači da je prema visini faktora odjeka rangiran od 25% do 50% časopisa unutar predmetnog područja itd.

3.7.1. Faktor odjeka

Faktorom odjeka mjeri se učestalost kojom se prosječni članak objavljen u nekom časopisu navodi/citira u određenom razdoblju. Faktor odjeka se objavljuje u bazi podataka JCR, a na temelju podataka o citiranosti iz časopisa što ih redovito obrađuje citatna baza podataka WoSCC (Macan i Petrak, 2015).

Faktor odjeka je zapravo omjer broja citata koje su u nekoj godini dobili radovi objavljeni u prethodne dvije godine i broja radova objavljenih u te dvije godine. Formula za njegov izračun glasi:

IF (2022) = A/B

pri čemu je: IF (2022) – faktor odjeka časopisa za 2022. godinu

A – broj citata koje su u 2022. godini dobili radovi objavljeni u časopisu za koji se računa IF u 2021. i 2020. godini

B – broj radova objavljenih u 2021. i 2020. godini.

Drugim riječima, faktor odjeka je aritmetička sredina raspodjele citata po svim radovima koje je časopis objavio.

Zahvaljujući svojoj jednostavnosti, postojanosti i razumljivosti vrlo brzo je prihvaćen u znanstvenoj zajednici te je ušao u široku upotrebu kao izravan odraz ugleda i kvalitete nekoga znanstvenog časopisa iako se uspješnost nekog časopisa ne može se gledati kroz prizmu samo jednog pokazatelja (Macan i Petrak, 2015).

3.8. Scopus – SCImago Journal & Country Rank

Besplatno dostupan portal SCImago Journal & Country Rank na temelju podataka o citiranosti radova iz citatne baze Scopus prikazuje niz bibliometrijski pokazatelji o znanstvenim časopisima i znanstvenoj produkciji na razini pojedinih država i regija (Macan i Petrak, 2015).

Najznačajniji bibliometrijski pokazatelj o časopisima je SCImago Journal Rank (SJR).

3.8.1. SCImago Journal Rank (SJR)

Bibliometrijski pokazatelj SCImago Journal Rank (SJR) je mjera znanstvenog utjecaja znanstvenih časopisa koji uzima u obzir citiranost časopisa i važnosti ili prestiž časopisa iz kojih citati dolaze.

Za razliku od faktora odjeka koji jednako rangira sve citate koje neki časopis primi, bez obzira na status časopisa iz kojih citati potječu, SJR mjeri prestiž pa citat iz časopisa koji ima relativno visoku vrijednost SJR-a vrijedi više nego citat iz časopisa s relativno malom vrijednošću SJR-a. Algoritam koji se primjenjuje za izračun sličan je Googleovu Page Ranku algoritmu koji se upotrebljava za rangiranje mrežnih stranica (Macan i Petrak, 2015).

SJR pokazatelj temelji se na citatima koje je neki časopis dobio u tekućoj godini za radove objavljene u prethodne tri godine.

Časopisi, serije knjiga i zbornici s konferencija su svrstani u 27 Scopusovih predmetnih područja (Subject Area) i 309 predmetnih kategorija (Subject Category) (https://www.scimagojr.com/aboutus.php, 2022).

Unutar svake predmetne kategorije serijske publikacije (časopisi i serije knjiga) se po visini SJR-a rangiraju u kvartile – od Q1 do Q4 (Macan, 2018b).

3.9. h-indeks

h–indeks je pokazatelj znanstvene uspješnosti (časopisa, znanstvenika, institucije i zemlje) koji istovremeno uzima u obzir znanstvenu produktivnost, mjerenu brojem objavljenih radova i znanstveni utjecaj, mjeren brojem citata tih radova u drugima znastvenim radovima.

Dva aspekta znanstvenog doprinosa (broj radova i citata) h-indeks prikazuje kao jednu brojčanu vrijednost. h-indeks nekog skupa radova je najveći broj radova h koji je primio barem h citata (Zauder, 2014).

Fizičar J. E. Hirsch predstavio je h- indeks u članku objavljenom 2005. godine i u relativno kratko vrijeme h – indeks postaje popularna mjera za vrednovanje znanstvenog rada.

Kao scientometrijski indikator služi za usporedbu zanstvenika samo iz istog područja i približno istog radnog iskustva, a isto se odnosi i na časopise.

Autorica Jokić u svom radu „*H - indeks kao novi scientometrijski indikator*“ (2009) navodi primjer da prilikom usporedbe dvaju znanstvenika približno istog radnog iskustva, sa sličnim brojem objavljenih radova i/ili sličnim ukupnim brojem citata, ali različitim h-indeksima govori u prilog veće prepoznatljivosti znanstvenika s većim h-indeksom.

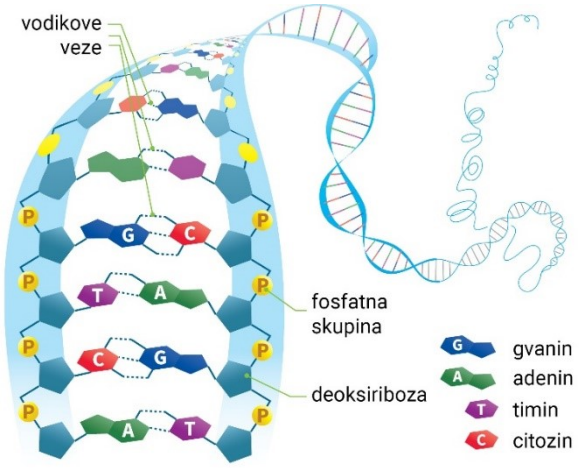
Autorica u istom radu dalje navodi da nisu rijetke situacije da znanstvenik objavi nekoliko značajnih radova s ekstremno visokim brojem citata, međutim njegov h-indeks ne bude naročito visok. Na jedan način h-indeks omogućuje ravnotežu između kvalitete i kvantitete.

h-indeks prisutan je u obje svjetske citatne baze - Web of Science i Scopus.

4. Osnove molekularne biologije

4.1. DNA – deoksiribonukleinska kiselina

Sva živa bića imaju zajedničke osnovne procese poput rasta, razmnožavanja i metabolizma. Ono što im je dodatno zajednička značajka na molekularnoj razini je genom. Genom svih živih bića predstavljaju molekule DNA. Molekule DNA se nalaze u gotovo svim stanicima svakog organizma i nose ukupnu nasljednu uputu za svako živo biće. Nasljedne upute zapisane su u redoslijedu nukleotida pa je sekvencija nukleotida ono što predstavlja ukupnu nasljednu informaciju prisutnu u svakoj stanici (Korać, Pongrac Štimac i Begić 2021). Nukleotidi se sastoje od šećera deoksiriboze, fosfatne skupine i određene dušične baze. U molekuli DNA razlikuju se četiri dušične baze: adenin (A), gvanin (G), citozin (C) i timin (T). Prilikom prikaza sekvencije nukleotida nekog gena odnosno genoma prikazuje se isključivo redoslijed dušičnih baza. Primjer sekvencije nukleotida: TTTGCATGTTA...



Slika 10. Prikaz građe molekule DNA

4.1.1. Gen

Slijed nukleotida nekog genoma organiziran je u različite funkcionalne jednice od kojih je napoznatija ona koju nazivamo gen. Gen je sekvencija DNA, dio genoma koji nosi uputu za neki produkt: molekulu RNA ili protein (Korać, Pongrac Štimac i Begić 2021).

Promjene koje se događaju u sekvenciji DNA zovu se mutacije. Mutacije u genu, tj. u uputi za neki protein mogu stvoriti varijantu proteina /svojstava koja će za organizam biti povoljna ili neutralna, ali i onu koja je za jedinku nepovoljna ili pogubna (Korać, Pongrac Štimac i Begić 2021).

4.2. Proteini

Proteini su najsvestranije molekule u živim sustavima i obavljaju bitne funkcije u svim biološkim procesima. (Berg, Tymoczko i Stryer 2013).

Proteini su izgrađeni od 20 različitih aminokiselina vezanih peptidnom vezom jedne uz druge. Svaka aminokiselina ima svoj naziv i kraticu. (Slika11) Redoslijed aminokiselina unutar određenog proteina zove se sekvencija aminokiselina.

Amino acid	Three-letter abbreviation	One-letter abbreviation	Amino acid	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A	Methionine	Met	M
Arginine	Arg	R	Phenylalanine	Phe	F
Asparagine	Asn	N	Proline	Pro	P
Aspartic acid	Asp	D	Serine	Ser	S
Cysteine	Cys	C	Threonine	Thr	T
Glutamine	Gln	Q	Tryptophan	Trp	W
Glutamic acid	Glu	E	Tyrosine	Tyr	Y
Glycine	Gly	G	Valine	Val	V
Histidine	His	H	Asparagine or aspartic acid	Asx	B
Isoleucine	Ile	I			
Leucine	Leu	L	Glutamine or glutamic acid	Glx	Z
Lysine	Lys	K			

Slika 11. Nazivi aminokiselina i njihovih troslovnih i jednoslovnih međunarodnih kratica

Prilikom prikazivanje redoslijeda aminokiselina u proteinu koriste se međunarodne jednoslovne kratice. Primjer prikaza sekvencije aminokiselina: MLAHHK...

Aminokiseline su vrlo raznolike. Širok je doseg njihovih strukturnih i kemijskih svojstava što proteinima daje svestranost koja im omogućuje različite uloge.

Postoje četiri razine strukture proteina: primarna, sekundarna, tercijarna struktura, a neki proteini imaju i kvartarnu strukturu. Sekvencija aminokiselina odnosno primarna struktura određuje trodimenzionalnu strukturu, a time i sva ostala svojstva proteina (Berg, Tymoczko i Stryer 2013).

4.3. Projek ljudskog genoma

Projekt ljudskog genoma je međunarodni projekt koji je za ciljeve imao određivanje slijeda nukleotida cijelog ljudskog genoma i određivanje slijeda nukleotida u genomu laboratorijskih organizama (bakterija E. coli, kvasac, vinska mušica, miš). Projekt je trajao od 1990. do 2003. godine te se smatra jednim od najambicioznijih i najvažnijih znanstvenih poduhvata u ljudskoj povijesti (https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project).

Znanstvenici *Projekta ljudskog genoma* učinili su javno dostupnim svaki dio nacrta sekvencije ljudskog genoma ubrzo nakon proizvodnje. Ova rutina proizašla je iz dvaju sastanaka na Bermudima na kojima su istraživači projekta pristali na "Bermudska načela" koja postavljaju pravila za brzo javno objavljivanje podataka. Ovaj značajan sporazum zaslužan je za uspostavljanje veće svijesti i otvorenosti za razmjenu podataka u biomedicinskim istraživanjima, što ga čini jednim od najvažnijih naslijeđa Projekta ljudskog genoma (https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project).

5. Istraživanje

Za temu istraživanja odabran je inzulin jer:

1.) 2021. godine obilježena je 100-godišnjica jednog od najvećih pronalazaka u povijesti medicine - otkriće i izolacija inzulina

2.) prije točno 100 godina (1922.) inzulin je po prvi puta korišten u terapijske svrhe odnosno za liječenje pacijenta s dijabetesom, bolesti koja je do tada imala smrtni ishod

3.) inzulin je prvi peptid (protein) kojemu je određen točan redoslijed (sekvencija) aminokiselina. Za to je otkriće Frederick Sanger 1958. godine dobio Nobelovu nagradu.

4.) inzulin je prvi odobreni proizvod, za uporabu u terapijskim postupcima, koji je dobiven korištenjem tehnologije rekombinantne DNA odnosno genetičkog inženjerstva (gen za ljudski inzulin ugrađen je u bakteriju *E. coli*). Prvi rekombinantni humani inzulin dostupan je od 1982.g.

U istraživačkom dijelu pretraživan je ljudski inzulin koji pripada porodici inzulina.

Za istraživanje je odabrana baza UniProtKB jer nudi poveznice prema izvorima informacija odnosno prema publikacijama u bazama PubMed i Europe PMC čime je omogućena bibliometrijska analiza.

Ciljevi istraživanja su: 1. utvrditi rezultate pretraživanja o ljudskom inzulinu u bazi UniProtKB; 2. s bibliometrijskog aspekta analizirati publikacije koje grade zapis o ljudskom inzulinu u bazi UniProtKB; 3. s bibliometrijskog aspekta vrednovati časopise u kojima su navedene publikacije objavljene.

Hipoteze istraživanja glase:

H1: Baza UniProtKB sadrži zapis o ljudskom inzulinu koji sadrži informacije dodatno pregledane, odabrane i ručno označene od strane stučnjaka.

H2: U publikacijama koje grade zapis o ljudskom inzulinu u bazi UniProtKB s godinama raste broj autora u koautorstvu.

H3. Časopisi *Nature* i *Science* objavili su barem jedan znanstveni rad o ljudskom inzulinu jer pripadaju među najcitiranije svjetske znanstvene časopise. Ujedno, to su časopisi s značajnim ugledom i dugogodišnjom tradicijom izlaženja.

Sukladno navedenim ciljevima i postavljenim hipotezama, istraživanje je podijeljeno na tri dijela: pretraživanje baze UniProtKB o ljudskom inzulinu, bibliometrijska analiza publikacija koje grade zapis o ljudskom inzulinu u bazi UniProtKB i bibliometrijska analiza časopisa u kojima su navedene publikacije objavljene.

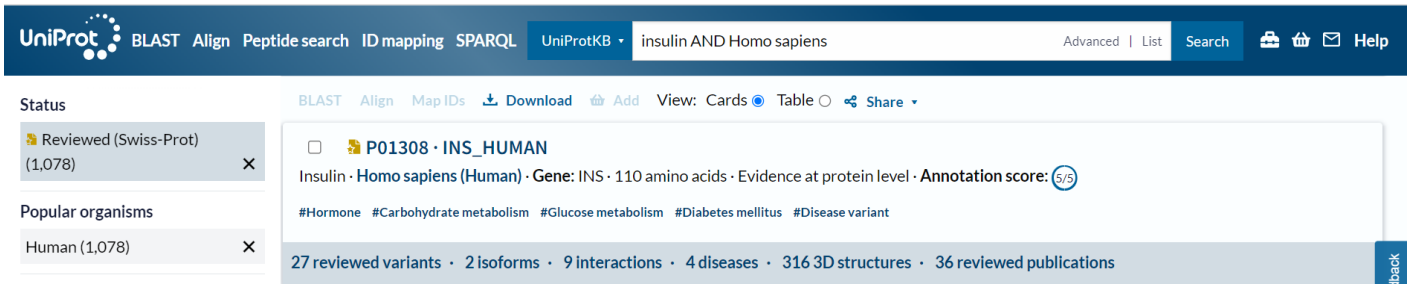
5.1. Pretraživanje baze podataka UniProtKB

Na početku istraživanja bilo je potrebno pronaći zapis o ljudskom inzulinu u bazi UniProtKB. Za pronalazak zapisa o ljudskom inzulinu korišteno je napredno pretraživanje. Izabrana su područja pretraživanja: ime peptida (proteina) - *insulin* i naziv organizama - *Homo sapiens*. Pretraživanja se provelo pomoću Booleovog operatora AND.

Ukupno je dobiveno 4,711 rezutata.

Osim zapisa o ljudskom inzulinu rezultati nude i zapise o: inzulinu sličnim peptidima, receptorima inzulina, enzimima vezanim za metabolizam inzulina (primjerice za njegovu razgradnju) itd.

Korištenjem filtera na lijevoj strani sučelja: pregledane i stručno označene rezultate (Reviewed (Swiss-Prot)) provedeno je dodatno sužavanje izbora na ukupno 1,078. Traženi rezultat prikazan je na slici 12.



Slika 12. Prikaz rezlutata pretrage o ljudskom inzulinu u bazi UniProtKB

Otvaranjem željenog rezultata otvara se zapis. Unutar svakog zapisa nalazi se popis publikacija. Popis publikacija sastoji se od dva dijela: pregledanog i stručno označenog dijela o ljudskom inzulinu - reviewed publications i nepregledanog dijela (računalno mapiranog iz drugih izvora) - unreviewed publications. Za potrebe ovog rada korištene su pregledane i stručno označene publikacije jer one grade zapis o ljudskom inzulinu u bazi.

Do popisa pregledane publikacije može se doći na dva načina.

Zapis u svom gornjem dijelu, ispod osnovnih informacija o proteinu, ima poveznicu prema publikacijama. Odabirom poveznice otvara se popis publikacija. Korištenjem filtera UniProtKB reviewed (Swiss-Prot), koji se nalazi na lijevoj strani zaslona, odabiru se isključivo pregledani i stručno označeni članci.

Do navedenih publikacija može se doći i kraćim putem odabirom poveznice "reviewed publications" naznačene kod ispisa rezultata pretraživanja (Slika 12). Daljnji postupak je isti.

U bazi UniProtKB broj pregledanih i stručno označenih publikacije o ljudskom inzulinu je 36.

U istraživanju će biti korištene 33 publikacije. Tri publikacije nisu navedene u bazama PubMed i Europe PMC jer njihov sadržaj direktno predan bazama podataka nukleotidnih sekvencija (INSDC, tj. EMBL/GenBank/DDBJ) i nije objavljen u recenziranom časopisu. U takvim slučajevima UniProtKB navodi samo naslov i popis autora iz podneska. Ponekad takvi podnesci nemaju naslove.

5.2. Bibliometrijska analiza znanstvenih radova

Za bibliometrijsku analizu odabrane su publikacije koje grade zapis o ljudskom inzulinu u biološkoj bazi UNiProtKB. UniProtKB nudi poveznice prema bazama PubMed i Europe PMC koje daju uvid u osnovne informacijama o publikaciji (autorstvo, godina publiciranja, zemlja publiciranja, popis referenci itd.). Podatke o molekularnom području istraživanja nudi biološka baza UniProtKB.

5.2.1. Godina publiciranja i područje istraživanja

Godina publiciranja znanstvenog rada, čije otkriće gradi zapis o ljudskom inzulinu u bazi UniProtKB, prikazana je u kombinaciji s molekularnim područjem koji znanstveni rad istražuje (Tablica 1). Zbog preglednosti prikaza korišteni su vremenski periodi od jednog desetljeća.

Istraživanje proteina na molekularnoj razini može imati više područja ovisno o tome istražuje li rad: redoslijed nukleotida, redoslijed aminokiselina, nastajanje (sintezu) proteina, varijante redoslijeda, proteinsku strukturu te vezano za strukturu funkciju odnosno procese u kojima protein sudjeluje.

Kategorije	Vremenski period						
	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2019	Ukupno
	Br. Publikacija	Br. Publikacija	Br. Publikacija	Br. Publikacija	Br. publikacija	Br. publikacija	
	1	7	8	9	6	2	33
redoslijed nukleotida		1	3	1	3		8
redoslijed aminokiselina	1	2					3
Sinteza		4					4
Varijante			5	2	3	2	12
struktura, molekularna dinamika				6			6

Tablica 1. Prikaz područja istraživanja o ljudskom inzulinu kroz vremenski period od 1960. do 2019.

Vremenski period i područje istraživanja prikazuju kako su se proširivale i gradile spoznaje o ljudskom inzulinu.

Iz tablice je vidljivo da su znanstvenici slijed aminokiselina u ljudskom inzulinu otkrili prije nukleotidog slijeda. Najvažnije spoznaje o nastajanju (sintezi) inzulina na molekularnoj razini nastale su 70- ih godina.

Otkriće nukleotidnog slijeda pružilo je mogućnost prepoznavanja odstupanja (mutacija) od poznatog slijeda odnosno prepoznavanje varijanti. Neke od tih varijanti uzrokuju genske bolesti. Veliki broj istraživanja proučava i donosi spoznaje o mutiranim oblicima inzulina.

Nuklearna magnetska rezonance korištena 90-tih godina donosi znanja o molekularnoj dinamici inzulina i njegovih varijanti u različitim otopalima.

5.2.2. Zemlje

Zemlje koje su prodonijele istraživanju o ljudskom inzulinu, a rezultati njihova istraživanja uvršteni su u zapis o ljudskom inzulini u bazi UniProtKB, prikazane su u tablici 2.

Istraživanje inzulina na molekularnoj razini započelo je u Europi, u Velikoj Britaniji. Na Sveučilištu u Cambridgeu otkriven je redoslijed aminokiselina.

U narednom preriodu Velikoj Britaniji se pridružuju Njemačka i Švicarska. Velika Britanija među prvima ostvaruje suradnje.

Krajem 70-tih i u 80-tim godinama dominaciju nad istraživanjima preuzimaju SAD i Japan što rezultira njihovom međusobnom suradnjom 90-tih godina.

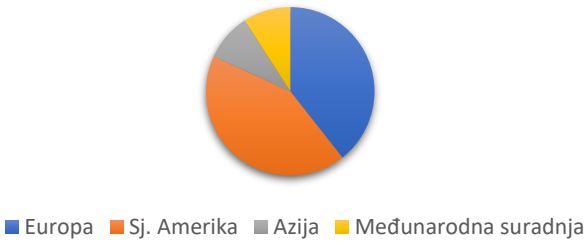
Razdoblje 21. stoljeća obilježeno je međunarodnim suradnjama u kojima sudjeluju tri ili više država.

Grafičkim prikazom prikazan je doprinos pojedinih regija (kontinenta) zapisu o ljudskom inzulinu u bazi UniProtKB od 1960. do 2019. godine. (Graf 1)

Kontinenti	Države	Vremenski period						
		1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2019	Ukupno
		Broj publikacija						33
Europa	UK	1				1		2
	Švicarska		1					1
	Njemačka		3					3
	Danska				2			2
	Norveška					1		1
	Češka						1	1
	UK, Danska		1					1
	UK, Francuska				1			1
	Danska, Italija, Češka						1	1
Sjeverna Amerika	SAD		2	6	4	1		13

	SAD, Japan				2			2
Azija	Japan			2				2
	Međunarodna suradnja (više od 3 države)					3		3

Tablica 2. Prikaz država koje istražuju ljudski inzulin na molekularnoj razini u vremenskom periodu od 1960. do 2019. godine, a njihova otkrića grade zapis o ljudskom inzulinu u biološkoj bazi UniProtKB



Graf 1.: Doprinos pojedinih regija (kontinenta) zapisu o ljudskom inzulinu u bazi UniProtKB od 1960. do 2019.

SAD i Japan surađivali su na dvije publikacije. Suradnja je ostvarena na razini sveučilišta, laboratorija i instituta navedenih zemalja. S obzirom da nije naveden doprinos svakog autora u istraživanju, publikacija je pripisana onoj državi koja je imala više autora. Stoga je jedna suradnja pripisana za potrebe grafa Japanu, a druga SAD-u.

5.2.3. Autorstvo, koautorstvo, timski rad

Bibliometrijskom analizom autora publikacija dobiven je podatak koliko je autora sudjelovalo u realizaciji znanstvenog rada. Rezultati su prikazani kroz periode od jednog desetljeća u svrhu stjecanja uvida kako se mijenja timska suradnja s protokom vremena. (Tablica 3)

Broj autora	Vremenski period						
	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2019	Ukupno
1 autor		1					1
2 autora	1	1		3			5
3 autora					1		1
4 autora		3	1	3			7
od 5 do 10 autora		2	7	3	1	1	14
od 10 do 20 autora						1	1
Projekti, grupe					4		4

Tablica 3. Prikaz broja autora na publikacijama o ljudskom inzulinu koje grade zapis u bazi UniProtKB

Iz navedenih rezultata vidljivo je da znanstveni radovi imaju najčešće od 4 do 10 autora. Veći broj autora (više od 10) tipično je za razdoblje nakon 2000- te kada se ostvaruju međunarodne suradnje i projekti. Manji broj autora karakterističan je za razdoblje od 1960. do 1979. Iznimka je 1991. kada dvije autorice Hua Q.-K. i Weiss M. A. koriste nuklearnu magnetsku rezonancu i dolaze do spoznaja koje opisuju u 3 znanstvena rada. Iz tog razloga u vremenskom period od 1990. do 1999. postoje tri publikacije s dva autora.

Tri publikacije rezultat su međunarodnog, multinstitucionalnog projekta i suradnje: MGC - Mammalian Gene Collection Project te Neonatal Diabetes International Collaborative Group s više od 100 odnosno 40 autora i suradnika koji su navedni u radu.

Jedna publikacija rezultat je studijske grupe unutar jedne države - Norwegian Childhood Diabetes Study Group s preko 40 autora odnosno suradnika navedenih na publikaciji.

5.3. Bibliometrijska analiza časopisa

Za analizu časopisa koji su objavili znanstvene radove čija istraživanja i otkrića grade zapis o ljudskom inzulinu u bazi UniProtKB korišten je portal SCImago Journal & Country Rank. Portal SCImago Journal & Country Rank na temelju podataka o citiranosti radova iz citatne baze Scopus prikazuje niz bibliometrijskih pokazatelja o znanstvenim časopisima.

U ovom radu istraživani su:

1.) h- indeks časopisa
2.) kvartil kojem časopis pripada zadnjih 5 godina

Ukupno 33 znanstvena rada objavljeno je u 19 časopisa. Od njih 19 mali broj časopisa (3) promijenilo je naziv tijekom godina. Razlozi su: dogovor unutar redakcije o promjeni naziva, udruživanje s ostalim časopisima iz istog područja u jedan časopis ili obrnuto, podjela jednog časopisa u specijalizirane rubrike. Novi naziv časopisa naznačen je u zagradi pored imena koji je časopis imao u trenutku publiciranja znanstvenog rada. Analiza za te časopise napravljena je pod njihovim novim nazivom.

5.3.1. h-indeks

Usporedba h- indeksa provedena je samo za časopise unutar iste predmetne kategorije čime se postiže uniformnost.

Kod uočenih većih odstupanja unutar iste predmetne kategorije analizirana je i "starost" časopisa (podatak dostupan na portalu Portal SCImago Journal & Country Rank u rubrici – "*coverage*") te broj objavljenih radova unutar časopisa koji se mogu citirati iz perioda 2020., 2019. i 2018. tzv. "*citable docs*". Navedni podatci (duljina izlaženja časopis i broj objavljenih članaka koji se mogu citirati) utječu na h-indeks časopisa.

Časopisi koji su objavili znanstvene radove o ljudskom inzulinu i čije spoznaje danas grade zapis u bazi UniProtKB pripadaju u četiri predmetna područja:

1.) Multidisciplinarno
2.) Biokemija, genetika i molekularna biologija
3.) Medicina
4.) Kemija

5.3.1.1. Multidisciplinarno područje

Časopisi koji pripadaju u multidisciplinarno područje prikazani su u tablici 4. Ukupno 10 radova objavljeno je u 4 časopisa.

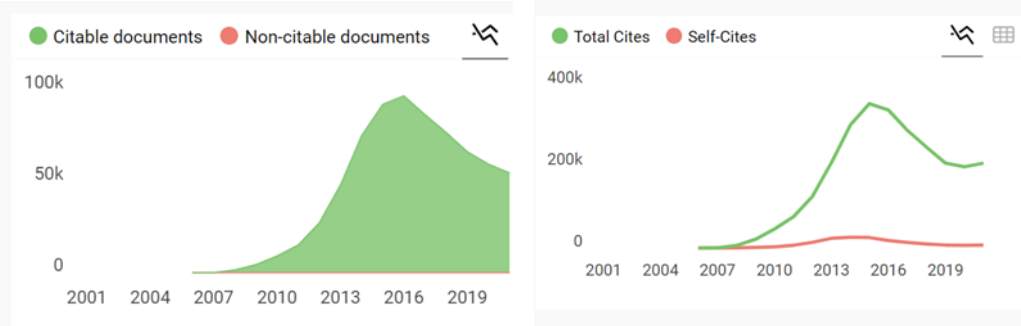
Predmetno područje	Naziv časopisa	h- indeks časopisa	Citable doc. (3 years)
multidisiplinarno	Nature	1276	3834
	Science	1229	4215
	Proceedings of the National Academy of Sciences of the United States of America	805	10594
	Plos One	367	49720

Tablica 4. Tablični prikaz časopisa multidisplinarnog područja

Časopisi *Nature* i *Science* svojim utjecajem i znanstvenim ugledom daleko su ispred ostalih časopisa. Publikacije o ljudskom inzulinu koje su objavljene u ovim časopisima vezane su za redoslijed aminokiselina u inzulinu (1960.) i redoslijed nukleotida u genu koji kodira za inzulin (1979. i 1980.). To su bila prijelomna otkrića koja su stvorila temelj za sve ostale istraživačke radove na temu ljudskog inzulina. Iz toga se može zaključiti da ta dva časopisa objavljuju "cutting- edge" spoznaje određenog područja.

U multidisciplinarno područje osim časopisa *Nature* i *Science* pripadaju još: *Proceedings of the National Academy of Sciences of the United States of America* i *Plos One.* Najviše objavljenjih radova koji se mogu citirati u periodu od tri godine ima časopis Plos One (49 720).

Časopis *Plos One* pokazuje veliki broj objavljenih radova na trogodišnjoj razini (Slika 13). Godine 2016. objavio je čak 88 176 članaka koji se mogu citirati za period od tri godine. Iako se radi o novom časopisu (prvi radovi objavljen su 2006. godine) h-indeks mu je visokih 367. Daljnom analizom ustanovljeno je da se radi o časopisu koji izdaje knjižnica *Public library of science* (SAD).



Slika 13. Podatci o objavljenim publikacijama i citiranosti za časopis *Plos One*

Kod časopisa koji objavljuju velik broj radova može očekivati veći h-indeks jer maksimalni h indeks jednak je ukupnom broju radova što ih je određeni časopis objavio.

5.3.1.2. *Biokemija, genetika i molekularna biologija*

Područje biokemija, genetika i molekularna biologija broji najviše časopisa koji su objavili radove o ljudskom inzulinu, a danas čine zapis u bazi UniProtKB. U 10 časopisa objavljeno je 13 radova. (Tablica 5)

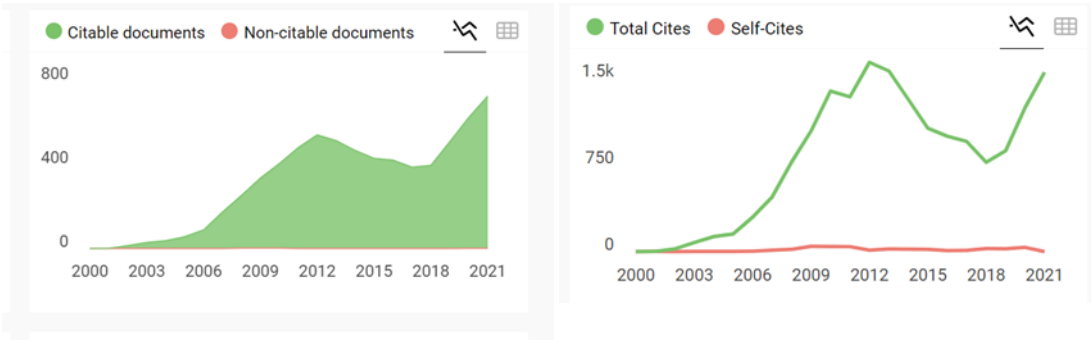
Najmanji h-indeks (72) ima *BMC Medical Genetics* iako mu je broj objavljenih radova u periodu od tri godine u prosjeku s ostalim časopisima (660).

Predmetno područje	Naziv časopisa	h- indeks časopisa	Citable doc. (3 years)
Biokemija, genetika i molekularna biologija	Nature genetics	597	503
	Genome research	305	507
	BMC Medical Genetics	72	660
	Journal of Biological Chemistry	528	4838
	European Jurnal of Biochemistry (FEBS Journal)	212	908

	Helvetica chimica acta	85	192
	Hoppe-Seyler's Zeitschrift fur Physiologische Chemie (Biological Chemistry	123	333
	Biochemistry	260	1718
	Biochimica et Biophysica Acta	153	393
	Journal of Molecular Biology	277	1112

Tablica 5. Tablični prikaz časopisa područja Biokemija, genetika i molekularna biologija

U suprotnosti od časopisa *PlusOne*, *BMC Medical Genetics* ima drukčiji put razvoja. Prvo izdanje časopis *BMC Medical Genetics* imao je 2000.g. Tada je publicirao jedan znastveni rad. Tijekom godina broj publikacija je rastao kao i broj citiranih radova iz tog časopisa. Godine 2020. objavio je najveći broj radova njih 660. h- indeks mu je 72 (Slika 14).



Slika 14. Podatci o objavljenim publikacijama i citiranosti za časopis *BMC Medical Genetics*

Na primjeru ovog časopisa jasno je vidljivo da h-indeks nije pogodan pokazatelj za nove časopise koji se postepeno probijaju na tržište jer h-indeks ovisi o duljini izlaženja nekog časopisa, odjeku članaka što ih objavljuje, broju objavljenih članaka, parametri koji se stječu godinama.

5.3.1.3. Medicina

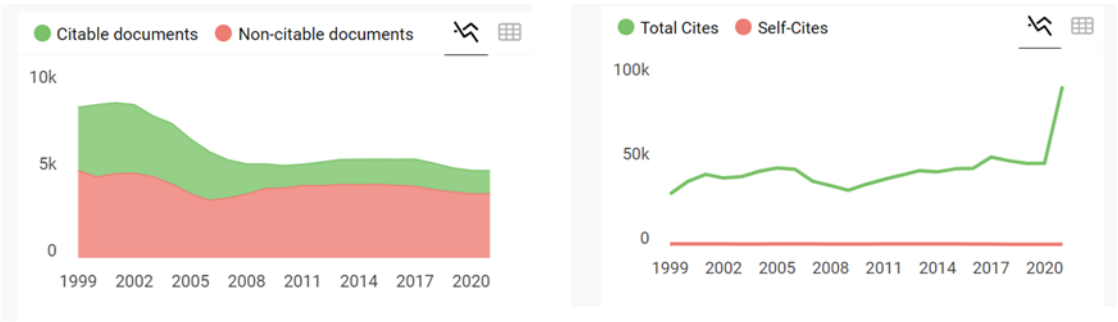
U području medicine ukupno 7 radova objavljeno je u 4 časopisa. (Tablica 6)

Predmetno područje	Naziv časopisa	h- indeks časopisa	Citable doc. (3 years)
Medicina	The Lancet	807	1226
	Journal of Clinical Investigation	505	1583
	Journal of Clinical Endocrinology and Metabolisam	363	1857
	Diabetes	345	683

Tablica 6. Tablični prikaz časopisa iz područja Medicine

Manji broj objavljenih radova u odnosu na druge časopise iz tog područja ima časopis *Diabetes* (683) koji ujedno ima i najmanji h–indeks (345). Najveći h-indeks (807) ima časopis *The Lancet* iako broj publikacija (1226) nije najveći u toj kategoriji unutar tri godine.

U period od 1999. do 2021. broj objavljenih radova koji se mogu citirati u časopisu *The Lancet* nije velik (do 3797 članaka) i kroz razdoblje od zadnjih 20 godina ima tendenciju smanjivanja. Istovremeno broj citiranih radova iz ovog časopisa raste što govori o njegovoj uspješnosti u navedenom području. (Slika 15)



Slika 15. Podatci o objavljenim publikacijama i citiranosti za časopis *The Lancet*

Na primjeru časopisa *The Lancet* jasno je vidljivo da je visoki h-indeks rezultat odjeka članaka što ih časopis objavljuje. Pretraživanjem informacija utvrđeno je da je *The Lancet* jedan od najstarijih, svjetski poznatih medicinskih časopisa, osnovan 1823. godine. Stoga je visok h-indeks rezultat prvenstveno ugleda časopisa što je usko povezano sa: duljinom izlaženja časopisa, odjeka članaka i broja objavljenih članaka tijekom godina.

5.3.1.4. Kemija

U području kemije objavljena su dva rada u jednom časopisu, *European Journal of Inorganic Chemistry*. Časopis je tijekom godina promijenio naziv u *Chemische Berichte* (h-indeks 139, a citable doc. (3 years) -1673). Usporedba s drugim časopisima iz tog područja nije moguća.

Navedeni rezultati istraživanja naglašavaju da prilikom procjene časopisa prema h-indeksu nije dovoljno samo uzeti broj kao relevantnu činjenicu. Treba napraviti dublju analizu koja objašnjava vrijednost h-indeksa.

5.3.2. SCImago Journal Rank - kvartili (Q)

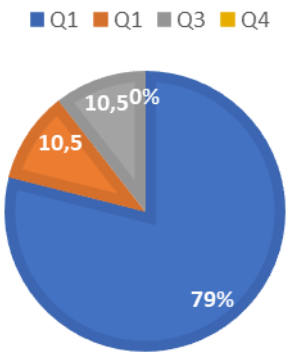
Za većinu znanstvenih područja prema *Pravilniku o uvjetima za izbor u znanstvena zvanja* (2017) u RH u znanstvene radove ubrajaju se radovi objavljeni u znanstvenim časopisima zastupljenima u bazama podataka WoSCC (Web *of Science Core Collection),* Scopus i Medline.

Pravilnik dalje navodi da podjela znanstvenih časopisa u kvartile (Q1, Q2, Q3, Q4) utvrđuje se prema izvješću o citiranosti JCR (Journal *Citation Report)* ili SJR (SCImago *Jouranal Rank indicator)* u godini objave ili prema najnovijem izvješću objavljenom do dana podnošenja zahtjeva.

Istraženo je kojem kvartilu pripadaju časopisi prema bibliometrijskom pokazatelju SJR (SCImago *Jouranal Rank indicator)* za razdoblje od 2017. do 2021. godine.

Za potrebe ovog istraživačkog rada u slučaju da časopis unutar navedenog perioda pripada u različite kvartile (primjerice 2 godine u prvom kvartilu i 3 godine u drugom) časopis se svrstava u onaj kvartil koji prevladava u navedenom vremenskom periodu.

Rezultati su prikazani grafički. (Graf 2)



Graf 2. Kvartili časopisa posljednjih pet godina (2017-2021) prema bibliometrijskom pokazatelju SJR

Od ukupno 19 časopisa, koji su objavili znanstvene radove čija istraživanja i otkrića grade zapis o ljudskom inzulinu u bazi UniProtKB, njih 15 odnosno 79% pripada prvom kvartilu (Q1) u svojoj predmetnoj kategoriji. Prvom kvartilu pripadaju svi časopisi iz Multidisiplinarnog i Medicinskog područja.

Dva časopisa odnosno 10,5% pripadaju drugom kvartilu (Q2) unutar svoje predmetne kategorije. To su: *Biochimica et Biophysica Acta* iz predmetnog područja Biokemija, genetika i molekularna biologija te *Chemische Berichte* iz predmetnog područja Kemija.

Trećem kvartilu (Q3) pripadaju također dva časopisa (10,5%) unutar svoje predmetne kategorije. To su časopisi: *BMC Medical Genetics* i *Helvetica chimica acta* iz predmetnog područja Biokemija, genetika i molekularna biologija.

Časopisi koji pripadaju četvrtom kvartilu (Q4) nisu zastupljeni.

Pravilnik o uvjetima za izbor u znanstvena zvanja (2017) navodi da u području „Prirodnih znanosti“ barem dvije trećine radova potrebnih za izbor mora biti objavljeno u časopisima čiji se čimbenik odjeka nalazi u prva tri kvartila (Q1, Q2 i Q3), a najmanje jedna trećina radova potrebnih za izbor mora biti objavljena u časopisima čiji se čimbenik odjeka nalaze u prvom (Q1) ili drugom kvartilu (Q2) predmetnih kategorija (Subject *Category List)* u koje je razvrstan časopis (https://narodne-novine.nn.hr/clanci/sluzbeni/2017_03_28_652.html).

Navedeni rezultati daju uvid koje časopise iz prirodnog područja odabrati u svrhu napredovanja u znanstvenom i akademskom sustavu.

6. Zaključak

Računala, mreže, unaprijeđeni istraživački instrumenti te velika količina automatizacije doveli su do eksplozije podataka dobivenih znanstvenim istraživanjem. Više nego ikad očituje se važnost znanstvene komunikacije i informacijske pismenosti znanstvenika.

Jedna od glavnih zadaća znanstvenika postaje prepoznavanje vjerodostojnosti informacija te vrednovanje istih, a otvorenost znanstvene komunikacije unapređuje kvalitetu znanstvenih istraživanja, time i razvoj znanosti.

Baza podataka iz područja molekularne biologije ima izrazito puno. Postoji tendencija okupljanja baza, a time i informacija na jednom mjestu. Navedeno će pridonijeti boljoj preglednosti znanja što može izravno utjecati na kvalitetu istraživačkih radova.

Svaka se baza podataka, pa tako i biološka, sastoji od: informacija (zapisa), sheme odnosno organizacije tih zapisa i alata za pristup. Međutim biološke baze podataka slijede neka svoja pravila što je rezultat specifičnosti područja koje obrađuju. U njima nalazimo redoslijed nukleotida, redoslijed aminokiselina, 3D slike proteina, BLAST analizu koja, na osnovu komparacije genetskih sekvenci različitih organizama, omogućava izradu filogenetskog stabla.

Na dnevnoj razini iz različitih laboratorija dolaze otkrića o novim sekvencijama genoma nekog organizma, o otkrivenim funkcijama nekog proteina, o povezanosti nekih procesa. Bibliografska baza PubMed redovito se nadopunjuje novim člancima iz biomedicinskog i molekularnog područja. Podaci koje danas vidimo u bazi Genske ontologije za tri godine će biti drukčiji. Neki podaci će postati irelevantni u odnosu na nove spoznaje, neki će se nadopuniti, a neki proširiti. Zato je bitno pratiti promjene i znanstvena istraživanja temeljiti na nanovijim spoznajama.

Danas je na internetu moguće pronaći brojne informacije čija kvaliteta nije uvijek vjerodostojna ni provjerena. Iz tog razloga organizirane i sustavno pripremljene baze podataka predstavljaju nezaobilazan izvor pouzdanih znanstvenih i stručnih informacija.

Ovaj diplomski rad prikazuje načine pretraživanja informacija u najznačajnijim biološkim bazama podataka: GenBank, UniProtKB, InterPro, PubMed i Europe PMC.

Za temu istraživanja odabran je ljudski inzulin koji pripada porodici inzulina. Mnoga otkrića i istraživanja vezana su za ovu porodicu proteina: inzulin je prvi peptid (protein) kojemu je određen točan redoslijed (sekvencija) aminokiselina za što je znanstvenik F. Sanger dobio

Nobelovu nagradu, genetičkim inženjerstvom gen za ljudski inzulin je ugrađen u bakteriju E. coli što je dovelo do stvaranja prvog rekombinantnog inzulina, ove godine (2022.) slavi se 100-godišnjica korištenja inzulina u terapijske svrhe, prošle godine (2021.) obilježeno je 100 godina od otkrića i izolacije inzulina.

Za pretraživanje o ljudskom inzulinu odabrana je baza UniProtKB jer nudi poveznice prema izvorima informacija odnosno prema publikacijama u bazama PubMed i Europe PMC čime je omogućeno postizanje ciljeva istraživanja koji uključuju bibliometrijsku analizu publikacija o ljudskom inzulinu te s bibliometrijskog aspekta vrednovanje časopisa u kojima su navedene publikacije objavljane.

Pretraživanjem baze UniProtKB dobiven je podatak da ukupan broj pregledanih i stručno označenih publikacije koje grade zapis o ljudskom inzulinu u bazi UniProtKB iznosi 36. U istraživanju su korištene 33 publikacije. Tri publikacije nisu navedene u bazama PubMed i Europe PMC jer njihov sadržaj direktno predan bazama podataka nukleotidnih sekvencija (INSDC, tj. EMBL/GenBank/DDBJ) i nije objavljen u recenziranom časopisu.

Radom je ustanovljeno da su prva istraživanja vezana za molekularnu građu odnose na slijed aminokiselina u ljudskom inzulinu napravljena su Velikoj Britaniji. Nakon toga slijede istraživanja odnosno radovi o nukleotidoj sekvenciji, sintezi, a zatim i o varijantama tj. mutacijama. Europa, Sjeverna Amerika i Azija najviše su pridonijele istraživanju o ljudskom inzulinu. Razdoblje 21. stoljeća obilježeno je međunarodnim suradnjama u kojima sudjeluju tri ili više država.

Bibliometrijskom analizom autora publikacija dobiven je podatak da znanstveni radovi imaju najčešće od 4 do 10 autora. Veći broj autora tipično je za razdoblje nakon 2000- te kada se ostvaruju međunarodne suradnje i projekti. Manji broj autora karakterističan je za razdoblje od 1960. do 1979.

Za analizu časopisa koji su objavili znanstvene radove čija istraživanja i otkrića grade zapis o ljudskom inzulinu u bazi UniProtKB korišten je portal SCImago Journal & Country Rank. Časopisi koji su objavili znanstvene radove o ljudskom inzulinu i čije spoznaje danas grade zapis u bazi UniProtKB pripadaju u četiri predmetna područja: 1. Multidisciplinarno; 2. Biokemija, genetika i molekularna biologija; 3. Medicina; 4. Kemija. Ukupno 33 znanstvena rada objavljeno je u 19 časopisa.

U radu su istraživani: h-indeks časopisa i kvartil kojem časopis pripada zadnjih 5 godina.

Najveći h-indeks imaju časopisi *Nature* i *Science*. *Nature i Science* objavili su prijelomna otkrića vezana za ljudski inzulin (sekvencija aminokiselina i sekvencija nukleotida) što ih svrstava u kategoriju časopisa s najvećim utjecajem, ali i časopisa koji objavljuju „cutting edge“ spoznaje. Značajan utjecaj ima i časopis *The Lancet* iz područja medicine. Istraživanjem je uočeno kako novi časopisi (osnovani od 2000-te) imaju različit put razvoja odnosno strategiju opstanka na tržištu što izravno utječe na njihov h-indeks. Rezultati bibliometrijskog istraživanja dokazuju da prilikom procjene časopisa prema h- indeksu nije dovoljno samo uzeti numeričku vrijednost kao relevantnu činjenicu. Treba napraviti dublju analizu koja objašnjava vrijednost h-indeksa.

Istraživanjem kojem kvartilu pripadaju časopisi prema bibliometrijskom pokazatelju SJR (SCImago *Jouranal Rank indicator)* za razdoblje od 2017. do 2021. godine dobiveni je uvid koje časopise iz prirodnog područja odabrati u svrhu napredovanja u znanstvenom i akademskom sustavu.

7. Literatura

About Europe PMC (2022) Dostupno na: https://europepmc.org/About (25. listopada 2022.)

About UniProt (2022) Dostupno na: https://www.uniprot.org/help/about (20. Svibanj 2022.)

About us SCImago Journal & Country Rank (2022) Dostupno na: https://www.scimagojr.com/aboutus.php (12. listopada 2022.)

Achburner, M. et al., (2000) Gene Ontology: tool for the unification of biology. Nat. Genet (online), 25 (1), 25-29. Dostupno na: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3037419/ (10. srpnja 2022.)

Berg, J.M., Tymoczko, J. L. i Stryer, L. (2013) Biokemija (prev. Weygand-Đurašević, I., Jernej, B. i Kućan, Ž.). 6. izdanje (englesko). Zagreb: Školska knjiga.

Danesh, F. i Mardani-Nejad, A. (2021) A Historical Overview of Bibliometrics. U: Ball, R. (ur.), Handbook Bibliometrics. Berlin; Boston: De Gruyter Saur, str. 7-18.

Exploring Gene Ontology GO annotations tools and resources (2022) https://www.youtube.com/watch?v=6Am2VMbyTm4&t=2254s (5. srpanj 2022.)

Exploring protein function and sequences using UniProtKB – 2018 (2018) Dostupno na: https://www.youtube.com/watch?v=VC-tCkfvAHs (22. svibanj 2022.)

GenBank database Tutorial: A beginners Guide (2020) Dostupno na: https://www.youtube.com/watch?v=QCgSYQnmVuc (9. Studeni 2021.)

Golenko, D. (2019) Vodič za objavljivanje radova u europskom istraživačkom prostoru (online). Zagreb: Pravni fakultet u Rijeci. Dostupno na: https://pravri.uniri.hr/files/ridoc/Vodi_za_objavljivanje_radova_u_EIP2019-12.pdf (2. studenog 2022.)

Hebrang – Grgić, I. (2016) Časopisi i znanstvena komunikacija. Zagreb: Naklada Ljevak.

Human Genome Project. NIH. Dostupno na: https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project (10. listopada 2022.)

Jokić, M. (2005) Bibliometrijski aspekti vrednovanja znanstvenog rada. Zagreb: Sveučilišna knižara.

Jokić, M. (2009) H- indeks kao novi scientometrijski indikator. Biochemia Medica (online), 19 (1), 5-9. Dostupno na: https://hrcak.srce.hr/32249 (2. rujna 2022.)

Korać, P., Pongrac Štimac, Z. i Begić, V. (2021) Biologija 4: udžbenik iz bilogije za četvrti razred gimnazije. Zagreb: Alfa.

Krajna, T. i Zabjan Bogut, A. (2011) Trideset godina časopisa *Polimeri*: bibliometrijske značajke kategoriziranih radova. Polimeri: časopis za plastiku i gumu (online), 32 (2), 81-87. Dostupno na: https://hrcak.srce.hr/clanak/112463 (20. kolovoza 2022.)

Lapiš, S. (2018) Baze podataka u geografiji i strategija njihovog pretraživanja. Geografski horizont (online), 64 (1), 39-45. Dostupno na: https://hrcak.srce.hr/file/322006 (21.rujna.2022.)

Lesk, A. (2019) Introduction to bioinformatics. Fifth edition. United Kingdom: Oxford University Press.

Macan, B. i Petrak, J. (2015) Bibliometrijski pokazatelji za procjenu kvalitete znanstvenih časopisa. U: Hebrang Grgić, I. (ur.), Hrvatski znanstveni časopisi: iskustva, gledišta, mogućnosti. Zagreb: Školska knjiga, str. 37-53.

Macan, B. (2018a) Osiguravanje otvorenog pristupa znanstvenim publikacijama: tko, što i kako?. U: Hebrang Grgić, I. (ur.), Otvorenost u znanosti i visokom obrazovanju. Zagreb: Školska knjiga, str. 59-79.

Macan, B. (2018b) Znanstveni časopsi u kontekstu vrednovanja i kakve veze bibliometrija ima s tim (online). Zagreb: Institut Ruđer Bošković. Dostupno na: http://fulir.irb.hr/4089/1/944198.B_Macan-Bibliometrija_Physiotherapia_Croatica-2018.pdf (2. studenog 2022.)

Melinščak Zlodi, I. (2018) Razvoj otovrenog pristupa u Hrvatskoj. U: Hebrang Grgić, I. (ur.), Otvorenost u znanosti i visokom obrazovanju. Zagreb: Školska knjiga, str. 43 - 58.

Pravilnik o uvjetima za izbor u znanstvena zvanja (2017) Narodne novine 28, Zagreb.

PubMed Overview Dostupno na: https://pubmed.ncbi.nlm.nih.gov/about/ (13. rujan 2022.)

Romić, K. i Mitrović, G. (2017) Bibliometrijske značajke kategoriziranih radova na primjeru časopisa Medijska istraživanja (1995. -2015). Vjesnik bibliotekara Hrvatske (online), 60 (4), 197-220. Dostupno na: https://hrcak.srce.hr/195927 (15. rujna 2022.)

Stojanovski, J. (2015) R(evolucija) znanstvenih časopisa. U: Hebrang Grgić, I. (ur.), Hrvatski znanstveni časopisi: iskustva, gledišta, mogućnosti. Zagreb: Školska knjiga, str. 55-76.

Špiranec, S. (2015) Informacijska pismenost kao oslonac znanstvene komunikacije: argumentacijski i primijenjeni okvir. U: Hebrang Grgić, I. (ur.), Hrvatski znanstveni časopisi: iskustva, gledišta, mogućnosti. Zagreb: Školska knjiga, str. 147-158.

UniprotKB (2022) Dostupno na: https://www.uniprot.org/help/uniprotkb (20. svibanj 2022.)

Using the InterPro website in your research (2020) Dostupno na: https://www.youtube.com/watch?v=5dw03Bp7_HM (12. svibanj 2022.)

Vodopijevac, A i Kranjec. I. (2018) Otvoreni istraživački podatci. U: Hebrang Grgić, I. (ur.), Otvorenost u znanosti i visokom obrazovanju. Zagreb: Školska knjiga, str. 93 - 112.

Zauder, K. (2014) Razvoj scientometrije praćen kroz časopis *Scientometrics* od početka izlaženja 1978. do 2010. godine. Doktorski rad. Zagreb: Sveučilište u Zagrebu, Filozofski fakultet.

Webinar: Retrieving Exon and Coding Region Sequences for Genes (2016) Dostupno na: https://www.youtube.com/watch?v=Xac7anudTD0&t=1488s (12. studeni 2021.)

Baze podataka iz područja molekularne biologije – pretraživanje i bibliometrijska analiza znanstvenih radova i časopisa

Sažetak

Baza podataka iz područja molekularne biologije ima izrazito puno. Postoji tendencija okupljanja baza, a time i informacija na jednom mjestu. Navedeno će pridonijeti boljoj preglednosti znanja što može izravno utjecati na kvalitetu istraživačkih radova. Također, iz različitih laboratorija diljem svijeta dolaze podatci o najnovijim otkrićima. Iz tog razloga izrazito je važno pratiti promjene i znanstvena istraživanja temeljiti na najnovijim spoznajama.

Rad prikazuje načine pretraživanja informacija u najznačajnijim biološkim bazama podataka: GenBank, UniProtKB, InterPro, PubMed i Europe PMC. Za temu istraživanja odabran je ljudski inzulin koji pripada porodici inzulina, porodici za koju su vezana mnoga znanstvena otkrića i istraživanja. Za pretraživanje o ljudskom inzulinu odabrana je baza UniProtKB jer na jednom mjestu nudi poveznice prema izvorima informacija odnosno prema publikacijama u bazama PubMed i Europe PMC čime je omogućeno postizanje ciljeva istraživanja koji uključuju: bibliometrijsku analizu publikacija čija otkrića grade zapis o ljudskom inzulinu u bazi UniProtKB te s bibliometrijskog aspekta vrednovanje časopisa u kojima su navedene publikacije objavljane.

Analizama na razini publikacija dobiven je uvid o: doprinosu pojedinih zemalja i regija u izgradnji spoznaja o ljudskom inzulinu na molekularnoj razini; autorstvu, koautorstvu i suradnji u znanstvenim radovima o ljudskom inzulinu; provedenim istraživanjima o ljudskom inzulinu na molekularnoj razini. Bibliometrijskom analizom znanstvenih časopisa prikazano je na koji način napraviti dublju analizu numeričke vrijednosti h-indeksa te je naglašena važnost iste. Istraživanjem su ujedno dobivene informacije koje časopise iz prirodnog područja odabrati u svrhu napredovanja u znanstvenom i akademskom sustavu.

Ključne riječi: GenBank, UniProtKB, InterPro, PubMed, Europe PMC, ljudski inzulin, bibliometrijska analiza

Databases in the field of molecular biology - search and bibliometric analysis of scientific works and journals

Summary

There are many bases in the field of molecular biology. There is a tendency to gather databases and information in one place. This will contribute to a better transparency of knowledge, which can directly affect the quality of research papers. Also, data on the latest discoveries comes from different laboratories around the world. For this reason, it is extremely important to monitor changes and base scientific research on the latest knowledge.

The paper shows ways to search for information in the most important biological databases: GenBank, UniProtKB, InterPro, PubMed and Europe PMC. Human insulin, which belongs to the insulin family, was chosen for the research topic, because many discoveries and researches are related to this family of proteins. For the search on human insulin, the UniProtKB database was chosen because it offers links to sources of information in one place, i.e. to publications in the PubMed and Europe PMC databases, which enables the achievement of the research goals, which include bibliometric analysis of publications on human insulin in the UniProtKB database and, from a bibliometric aspect, the evaluation of journals in which mentioned publications published.

Analyzes at the level of publications provided insight into: the contribution of individual countries and regions in the development of knowledge about human insulin at the molecular level; authorship, co-authorship and collaboration in scientific papers on human insulin; conducted research on human insulin at the molecular level. The bibliometric analysis of scientific journals shows how to do a deeper analysis of the numerical value of the h-index, and its importance is emphasized. The research also provided information on which journals from the natural field to choose in order to progress in the scientific and academic system.

Key words: GenBank, UniProtKB, InterPro, PubMed, Europe PMC, human insulin, bibliometric analysis