

**SVEUČILIŠTE U ZAGREBU  
PREHRAMBENO-BIOTEHNOLOŠKI FAKULTET**

# **Diplomski rad**

**Zagreb, 7. listopad, 2010.**

**Dina Rešetar, 4991/BM**

**SVEUČILIŠTE U ZAGREBU  
PREHRAMBENO-BIOTEHNOLOŠKI FAKULTET**

# **Diplomski rad**

**Zagreb, 7. listopad, 2010.**

**Dina Rešetar, 4991/BM**

**RAČUNALOM PODPOMOŽNUTA  
KLASIFIKACIJA  
NERIBOSOMALNIH PEPTID  
SINTETAZA**

Ovaj je rad izrađen u Kabinetu za bioinformatiku, Zavoda za biokemijsko inženjerstvo,  
Prehrambeno-biotehnološkog fakulteta, Sveučilišta u Zagrebu pod vodstvom

dr. sc. DASLAVA HRANUELI, redovitog profesora

uz svesrdnu pomoć

dr. sc. ANTONIA STARČEVIĆA, višeg asistenta

Zahvaljujem se mentoru, prof. dr. sc. Daslavu Hranueli što mi je omogućio izradu ovog diplomskog rada, na stručnom vodstvu i konstruktivnim savjetima.

Također se zahvaljujem dr. sc. Antoniu Starčeviću na nezamjenjivoj pomoći tijekom izrade diplomskog rada te uvijek dobrodošlim savjetima i potpori.

Puno hvala mami i tati na bezuvjetnoj ljubavi i razumijevanju.

Hvala prijateljima i meni dragim osobama na satima i satima smijanja i neprocjenjive zabave.

# TEMELJNA DOKUMENTACIJSKA KARTICA

Diplomski rad

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

Znanstveno područje: Biotehničke znanosti  
Znanstveno polje: Biotehnologija

## RAČUNALOM PODPOMOŠNA KLASIFIKACIJA NERIBOSOMALNIH PEPTID SINTETAZA

*Dina Rešetar, 4991/BM*

**Sažetak:** U posljednje se vrijeme, u javnosti, pojavilo veliko zanimanje za oblikovanje novih lijekova. Biološki aktivne supstancije za njihovo se oblikovanje sve više traže u genomima i metagenomima sekvenciranih mikroorganizama. Razvijen je čitav niz javno dostupnih baza podataka i bioinformatičkih alata za njihovu analizu. To se posebno odnosi na bioinformatičke alate koji analiziraju genske nakupine, čiji produkti sudjeluju u sintezi sekundarnih metabolita u uvjetima *in silico*. U ovome je radu sistematizirano znanje o dosada poznatim i detaljno opisanim sekundarnim metabolitima, neribosomalno sintetiziranih peptida i siderofora, radi njihove bolje klasifikacije. Detaljnija klasifikacija omogućila je izradu vlastitih profila proteina za navedene skupine sekundarnih metabolita kojima je unaprijeđen bioinformatički alat *ClustScan*. Ti su profili testirani na novosekvencioniranom i dosada neopisanom genomu vrste *Streptomyces scabies*. Na temelju sličnosti genoma vrste *S. scabies* i dosad anotiranih genoma *S. coelicolor* i *S. avermitilis* pretpostavlja se da će anotiranje novog genoma dovesti do pronalaska novih genskih nakupina odgovornih za biosintezu neribosomalno sintetiziranih peptida i siderofora.

**Ključne riječi:** neribosomalno sintetizirani peptidi, siderofori, *S. scabies*, *ClustScan*

**Rad sadržava:** 56 stranica, 29 slika, 8 tablica, 51 literaturna navoda, 4 priloga

**Jezik izvornika:** hrvatski

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

**Mentor:** dr. sc. Daslav Hranueli, red. prof.

**Pomoć pri izradi:** dr.sc. Antonio Starčević, viši asistent

**Stručno povjerenstvo za ocjenu i obranu:**

1. Dr. sc. Božidar Šantek, red. prof.
2. Dr. sc. Daslav Hranueli, red. prof.
3. Dr. sc. Blaženka Kos, izv. prof.
4. Dr. sc. Ana Vukelić, izv. prof. (zamjena)

**Datum obrane:** 7. listopada, 2010.

## **BASIC DOCUMENTATION CARD**

**Graduate Thesis**

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

**Scientific area:** Biotechnical Sciences  
**Scientific field:** Biotechnology

### **COMPUTER AIDED CLASIFICATION OF NONRIBOSOMAL PEPTID SYSTEMS**

*Dina Rešetar, 4991/BM*

**Abstract:** During the last several years there has been an increased public interest in the design of novel drugs. Majority of biologically active substances for their design have been largely looked in sequenced microbial genomes and metagenomes. For that reason a number of publically available databases and bioinformatics tolls were developed to analyse gene clusters whose products are involved in the biosynthesis of secondary metabolites *in silico*. In this work the knowledge of already known secondary metabolites, nonribosomally synthesized peptides and siderophores, was systematized for their better classification. More detailed classification allowed the design of the propriety protein profiles for the gene cluster proteins involved and subsequently used for the improvement of the *ClustScan* program package. These profiles have been tested on the newly sequenced genome of *Streptomyces scabies*. Based on the similarity of the *S. scabies* genome to the already annotated genomes of *S. coelicolor* and *S. avermitilis* it was assumed that the annotation of the new genome will lead to the discovery of novel gene clusters responsible for the production of yet unrecognised nonribosomally synthesized peptides and siderophores.

**Keywords:** nonribosomally synthesized peptides, siderophores, *S. scabies*, *ClustScan*

**Thesis contains:** 56 pages, 29 figures, 8 tables, 51 references, 4 supplements

**Original in:** Croatian

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

**Mentor:** Ph. D. Daslav Hranueli, Full Professor

**Technical support and assistance:** Ph. D. Antonio Starčević, Senior Assistant

#### **Reviewers:**

1. Ph. D. Božidar Šantek, Full Professor
2. Ph. D. Daslav Hranueli, Full Professor
3. Ph. D. Blaženka Kos, Assistant Professor
4. Ph. D. Ana Vukelić, Assistant Professor (substitute)

**Thesis defended:** 7<sup>th</sup> October, 2010

# SADRŽAJ RADA

<b>1. U V O D</b>	1
<b>2. T E O R I J S K I D I O</b>	
<b>2.1. SEKUNDARNI METABOLITI</b>	2
<b>2.1.1. Neribosomalno sintetizirani peptidi</b>	4
2.1.1.1. Linearne sintetaze neribosomalnih peptida	7
2.1.1.2. Ponavljajuće sintetaze neribosomalnih peptida	8
2.1.1.3. Nelinearne sintetaze neribosomalnih peptida	9
<b>2.1.2. Poliketidno/peptidni hibridi</b>	10
<b>2.1.3. NRPS neovisni siderofori</b>	10
<b>2.2. BIOINFORMATIČKI ALATI ZA ANOTACIJU GENSKIH NAKUPINA</b>	15
2.2.1. NRSPredictor, NP.searcher, SEARCHPKS i drugi alati	17
2.2.2. Programski paket <i>ClustScan</i>	18
<b>3. E K S P R I M E N T A L N I D I O</b>	
<b>3.1. MATERIJAL</b>	20
<b>3.1.1. Računalna podrška i operativni sustav</b>	20
<b>3.1.2. Baze podataka</b>	20
3.1.2.1. Baza podataka <i>GenBank</i>	20
3.1.2.2. Baza podataka NRPS-PKS	20
3.1.2.3. Baza podataka Pfam	21
<b>3.1.3. Bioinformatički računalni paketi i programi</b>	22
3.1.3.1. Računalni program <i>Glimmer</i>	22
3.1.3.2. Računalni program <i>GeneMark-PS</i>	22
3.1.3.3. Programski paket <i>HMMER</i>	22
3.1.3.4. Programski paket <i>ClustScan</i>	23
<b>3.1.4. Sekvencirani genomi vrsta <i>Actinomyces</i></b>	23
3.1.4.1. Genom bakterije <i>Streptomyces scabies</i>	23
<b>3.2. METODE RADA</b>	24
<b>3.2.1. Prikupljanje literaturnih podataka</b>	24
<b>3.2.2. Prikupljanje sekvencija DNA i proteina</b>	24
<b>3.2.3. Analiza sekvencija proteina programskim paketom <i>HMMER</i></b>	27
<b>3.2.4. Pronalaženje strukturnih gena</b>	30
<b>3.2.5. Preuzimanje gotovih profila proteina iz baze podataka <i>Pfam</i></b>	31
<b>3.2.6. Izrada vlastitih profila proteina pomoću programskog paketa <i>ClustScan</i></b>	32
<b>3.2.7. Anotacija genoma bakterije <i>Streptomyces scabies</i> pomoću programskog paketa <i>ClustScan</i></b>	33



<b>4. R E Z U L T A T I</b>	
4.1. Rezultati analize literaturnih podataka	36
4.2. Rezultati analize sekvencija proteina programskim paketom <i>HMMER</i>	37
4.3. Rezultati anotacije genoma bakterije <i>Streptomyces scabies</i> pomoću programskog paketa <i>ClustScan</i>	40
<b>5. R A S P R A V A</b>	
5.1. Pretraživanje literaturnih podataka	45
5.2. Obrada prikupljenih sekvencija DNA i proteina	46
5.3. Anotacija genoma bakterije <i>Streptomyces scabies</i> pomoću programskog paketa <i>ClustScan</i>	47
5.4. Proširenje funkcionalnosti programskog paketa <i>ClustScan</i>	49
<b>6. Z A K L J U Č C I</b>	50
<b>7. P O P I S L I T E R A T U R E</b>	51
<b>8. P R I L O Z I</b>	
8.1. Popis u radu upotrijebljenih kratica	55
8.2. Sadržaj kompaktnog diska	56

## **1. UVOD**

Sekundarni metaboliti mikroorganizama smatraju se kemijskim oblikom diferencijacije. Strukturno su vrlo raznolika skupina molekula, različitih bioloških aktivnosti pa se stoga primjenjuju u medicini, veterini i agroindustriji. Najveći broj sekundarnih metabolita sintetiziraju vrste roda *Streptomyces* i srodnih rodova aktinomiceta. Na njih otpada gotovo 8.000 poznatih antibiotički aktivnih supstancija. Sam rod *Streptomyces* sintetizira 75 % od svih klinički važnih antibiotika.

Posljednjih nekoliko godina u znanstvenoj javnosti postoji osobito zanimanje za oblikovanje novih farmaceutskih supstancija. Analizom enzimskih sustava, kontrole sinteze i diferencijacije uključene u sintezu sekundarnih metabolita razvile su se nove ideje proizvodnje lijekova manipulacijom genskih nakupina u uvjetima *in vitro*. Napredak na području bioloških i računalnih znanosti omogućio je pohranjivanje velikog broja genoma i proteoma, kao i njihovu organizaciju i detaljnu analizu. Razvijene su prve javno dostupne baze podataka i bioinformatički alati koji omogućuju predviđanje genetičkih elemenata. Razvojem bioinformatičkih alata koji analiziraju genske nakupine, čiji produkti sudjeluju u sintezi sekundarnih metabolita, omogućena je manipulacija genskim nakupinama u uvjetima *in silico*. To znači računalnu analizu i manipulaciju genskim nakupinama čija funkcija i konačni kemijski produkt još nisu poznati, uz pomoć bioinformatičkih alata i programskih paketa. Međutim, da bi se stvorili preduvjeti za metode *in silico*, prethodno je neophodno sekvencionirati nove genome i metagenome mikroorganizama koji žive u tlu ili u vodenom okolišu kako bi se pronašle nove genske nakupine i novi biosintetski putovi sekundarnih metabolita.

Izradom ovog diplomskog rada sistematizirati će se znanje o dosada istraženim i detaljno opisanim sekundarnim metabolitima, neribosomalno sintetiziranih peptida i siderofora, radi njihove bolje klasifikacije. Detaljnija klasifikacija omogućit će izradu vlastitih profila proteina za navedene skupine sekundarnih metabolita kojima ćemo unaprijediti bioinformatički alat *ClustScan* i testirati na novosekvencioniranom i dosada neopisanom genomu vrste *Streptomyces scabies*. Na temelju sličnosti genoma vrste *S. scabies* i dosad anotiranih genoma *S. coelicolor* i *S. avermitilis* pretpostavlja se da će anotiranje novog genoma dovesti do pronalaska novih ili već opisanih genskih nakupina odgovornih za biosintezu neribosomalno sintetiziranih peptida i siderofora.

## **2. TEORIJSKI DIO**

## 2.1. SEKUNDARNI METABOLITI MIKROORGANIZAMA

Sekundarni metaboliti mikroorganizama su strukturno vrlo različita skupina molekula koja se primjenjuje u medicini, veterini i agroindustriji. Gotovo 50 % najvažnijih klinički primijenjenih lijekova kao lijekovite supstancije sadržavaju sekundarne metabolite (Demain, 1999). Osim antibiotika, fungicida, antivirusnih sredstava i citostatika, u kliničkoj se primjeni nalaze i imunosupresori, antihipertenzivna sredstva, antidiabetici, antimalarici, te antikolesterolemi (Grabley i Thiericke, 1999; Hranueli i Cullum, 2001). Komercijalno su važni i proizvodi za agroindustriju, kao što su to antiparazitici, kokcidiostatici, životinjski promotori rasta i prirodni insekticidi. Mikroorganizmi koji sintetiziraju najveći broj sekundarnih metabolita pripadaju vrstama roda *Streptomyces* i srodnih rodova aktinomiceta. Na njih otpada gotovo 8.000 poznatih antibiotički aktivnih supstancija. Sam rod *Streptomyces* sintetizira 75 % od svih klinički važnih antibiotika (Demain, 1999).

Mikroorganizmi, bakterije i plijesni, sintetiziraju sekundarne metabolite nakon eksponencijalne faze rasta, pri brzinama rasta manjim od maksimalne, dakle tijekom idiofaze. Nisu neophodno potrebni za vegetativni rast stanica bakterija i plijesni koje ih proizvode, već su uključeni u različite mehanizme preživljavanja u prirodi, te kao čimbenici u ekološkim natjecanjima i simbiozi. Smatraju se kemijskim oblikom diferencijacije. Ekspresija sekundarnog metabolizma je specifična za soj (to jest različiti sojevi iste vrste mogu proizvoditi različite sekundarne metabolite) (Demain, 1999; Grabley i Thiericke, 1999; Hranueli i Cullum, 2001).

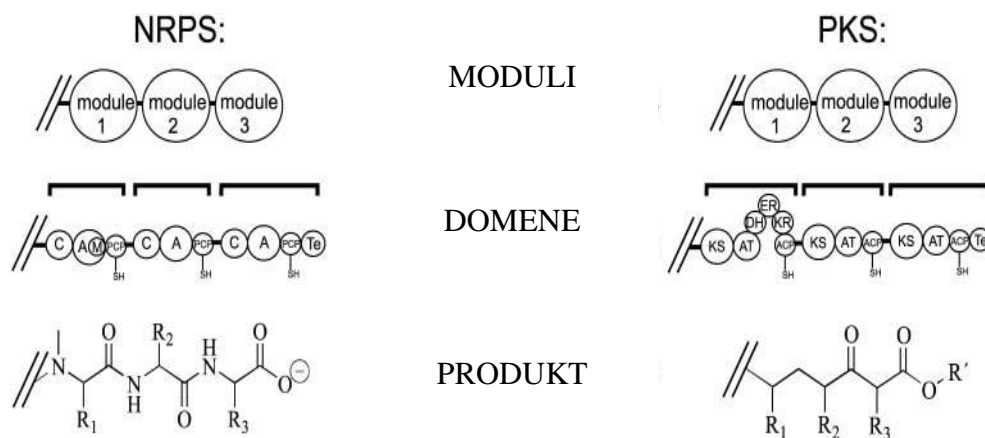
Geni čiji produkti sudjeluju u biosintezi sekundarnih metabolita (npr. antibiotika), kao i geni za otpornost prema vlastitom antibiotiku, obično su u bakterija i nekih gljiva (npr. *Penicillium* i *Aspergillus*) organizirani u jednoj genskoj nakupini smještenoj na kromosomskoj DNA, vrlo rijetko na plazmidima. Međutim u nekih gljiva, kao npr. *Acremonium chrysogenum* i *Cephalosporium acremonium*, geni čiji produkti sudjeluju u biosintezi cefalosporina, nalaze se na različitim kromosomima (Mootz i sur., 2002). Biosintetski put sekundarnog metabolita može uključivati i više od 40-ak gena koji čine DNA sekvenciju dužine i do 100 kb i može biti dvostruko veći od ribosoma (Fischbach i sur., 2008). Veliki su geni (engl. "giant genes") u arheja i eubakterija rijetki, a genske nakupine sekundarnih metabolita predstavljaju međusobno vrlo različite i najbrže evoluirajuće genetičke elemente (Reva i Tümmler, 2008).

Posljednjih nekoliko godina u znanstvenoj javnosti postoji osobito zanimanje za oblikovanje novih farmaceutskih supstancija. Analizom enzimskih sustava, kontrole sinteze i diferencijacije uključene u sintezu sekundarnih metabolita razvile su se nove ideje proizvodnje lijekova manipulacijom genskih nakupina u uvjetima *in vitro*, postupcima kombinatorne biosinteze. Međutim sojevi konstruirani na taj način vrlo često ne sintetiziraju niti jednu novu supstanciju ili daju vrlo nizak prinos nove supstancije. Jedno od mogućih rješenja toga problema bilo bi oblikovanje novih genskih nakupina homolognom rekombinacijom u uvjetima *in silico* primjenom bioinformatičkih alata (Hranueli i Cullum, 2003). Međutim, prethodno je neophodno sekvencioniranje novih genoma kako bi se pronašle nove genske nakupine te analize metagenoma mikroorganizama koji žive u tlu ili u simbiozi s morskim organizmima ne bi li se pronašli novi biosintetski putovi sekundarnih metabolita (Hranueli i sur., 2008).

S obzirom na kemijsku strukturu spojeva, među sekundarnim metabolitima bakterija i plijesni najveću i najbolje istraženu skupinu čine poliketidi (npr. eritromicin A ili rapamicin). Biosintetski im put započinje uzastopnom kondenzacijom jednostavnih građevnih jedinica (karboksilnih kiselina) i vodi ka izgradnji ugljikovih lanaca određenih dužina katalitičkim djelovanjem enzimskog sustava poliketid sintaze (engl. "Polyketide Synthases", PKS). Nakon toga, ravnolančani kondenzati podliježu ciklizaciji i poslije-kondenzacijskim izmjenama. (Hranueli i Cullum, 2001; Yadav i sur., 2003). Druga velika skupina sekundarnih metabolita jesu neribosomalno sintetizirani peptidi u čijoj sintezi sudjeluje enzimski sustav sintetaza neribosomalnih peptida (engl. "Non Ribosomal Peptide Synthetases", NRPS), a treću i najmanje istraženu skupinu sekundarnih metabolita čine NRPS neovisne siderofor sintetaze (engl. "NRPS Independent Siderofores", NIS).

Iako upotrebljavaju različite građevne jedinice, sustavi PKS i NRPS imaju sličnu modularnu građu katalitičkih domena (Slika 1) i slične mehanizme sinteze produkta (Hranueli i sur., 2008). Biosintetski put obje skupine obuhvaća spajanje jednostavnih građevnih jedinica u složene kemijske strukture katalitičkim djelovanjem enzimskih kompleksa poliketid sintaza ili sintetaza neribosomalnih peptida. Na tim enzimskim sustavima biosinteza se odvija kao na proizvodnoj traci, s time da je redosljed kemijskih reakcija uvjetovan redosljedom enzimskih domena unutar modula. Dakle, svakim monomerom upravlja modul, kojeg čini međusobno povezani set enzimskih domena. U većini slučajeva, broj modula odgovara broju monomera koji se ugrađuju u konačan produkt.

Zbog tih se razloga sustavi PKS i NRPS nazivaju modularnim sustavima tio-kalupa (eng. "Thiotemplate Modular Systems", TMS) (Donadio i sur. 2007).



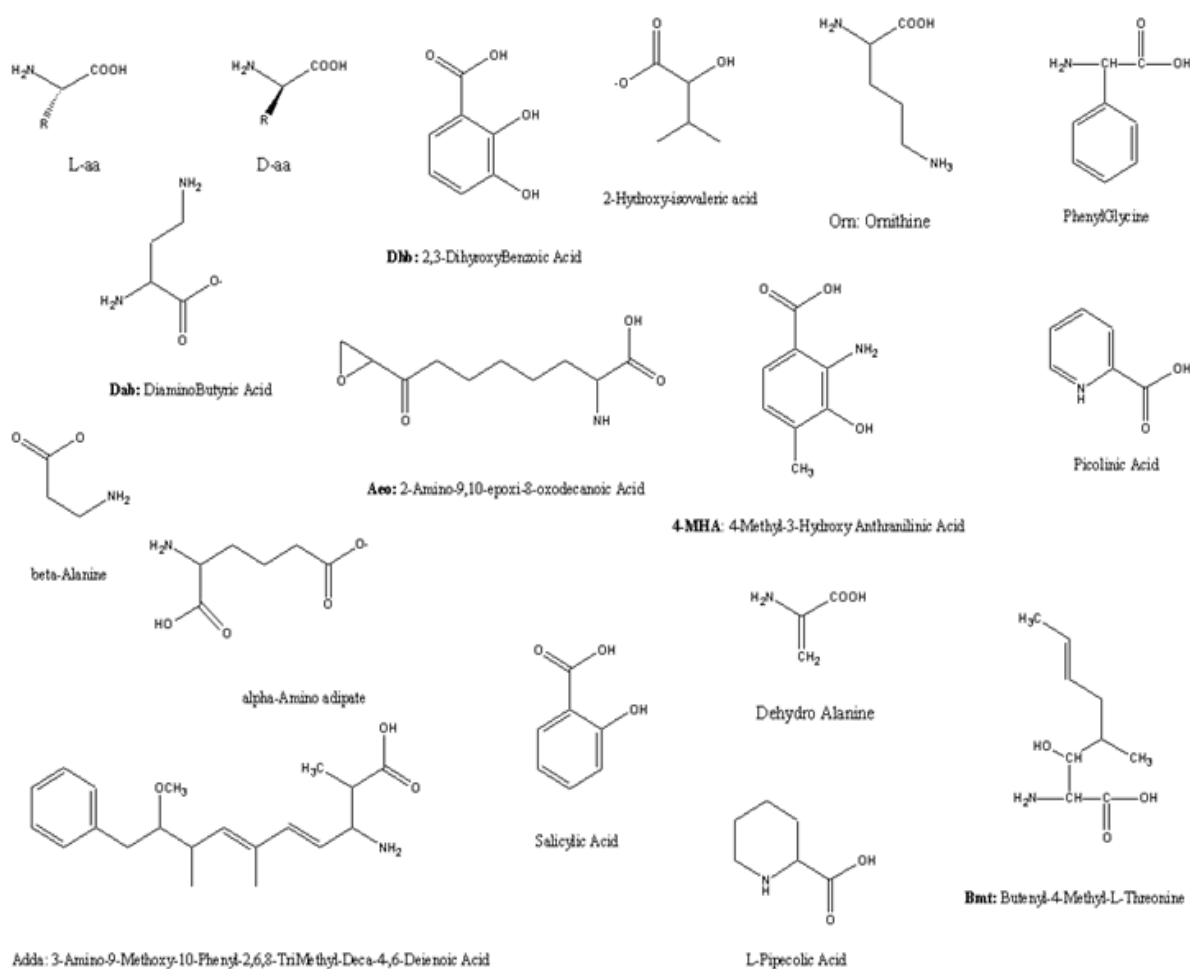
Slika 1. Od modula do produkata: moduli NRPS i PKS podijeljeni su u domene koje kataliziraju pojedinačne enzimske reakcije. Sastav produkata određen je nizom aktivnih domena u pripadajućim modulima.

### 2.1.1. Neribosomalno sintetizirani peptidi

Neribosomalno sintetizirani peptidi su velika skupina sekundarnih metabolita. U prirodi ih sintetiziraju mikroorganizmi, koji većinom prebivaju u zemlji, kao što su to Gram pozitivne bakterije rodova *Actinomyces* i *Bacillus*, te eukariotski mikroorganizmi iz skupine plijesni (Schwarzer i Marahiel, 2001). Mehanizam biosinteze neribosomalno sintetiziranih peptida prvi je puta opisan 1971. godine tijekom proučavanja mehanizama biosinteze antibiotika gramicidina S i tirocidina. Do danas je izolirano oko 700 strukturalno različitih neribosomalno sintetiziranih peptida. Za vrlo je malen broj od njih, međutim, poznat mehanizam biosinteze (Caboche i sur., 2008). Neribosomalno sintetizirani peptidi (obično dugački od 3 do 15 aminokiselina) ne sintetiziraju se translacijom na ribosomima, već uzastopnom kondenzacijom aminokiselina katalitičkim djelovanjem više-funkcionalnih neribosomalnih peptid sintetaza (NRPS).

Sustav NRPS može biti izgrađen od jednog polipeptidnog lanca, kao u plijesni (npr. 1.6 MDa ciklosporin sintetaza), ili od nekoliko podjedinica, kao u bakterijskim sustavima (npr.

tirocudin sintetaza, koja se sastoji od tri podjedinice proteina: TycA, TycB i TycC). Sustavima NRPS se sintetiziraju gotovo svi peptidni antibiotici i dio siderofora. Neribosomalno sintetizirane peptide karakterizira velika strukturalna raznolikost. Mogu biti linearni, ciklički ili razgranato-ciklički, te makrociklički laktami ili laktoni što je posljedica ugradnje uobičajenih (proteinogenih) i neuobičajenih aminokiselina kao što su to alfa-hidroksi i karboksi kiseline povezane vezama koje nisu samo peptidne ili disulfidne (Slika 2) (Wenzel i Müller, 2005).

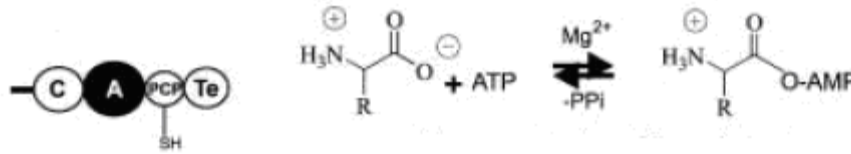


Slika 2. Građevne jedinice za sintezu neribosomalnih peptida.

Osnovni modul sustava NRPS sadržava domenu za kondenzaciju (engl. "Condensation domain", C), domenu za adenilaciju aminokiselina (engl. "Adenylation domain", A) i mali protein nosač peptidila (engl. "Peptidyl Carrier Protein", PCP ili T) (Schwarzer i Marahiel,

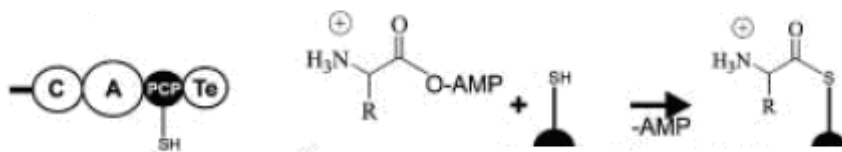


2001; Challis i Naismith, 2004). Domena A bira određenu aminokiselinu kao početnu ili produžnu građevnu jedinicu i aktivira supstrat adenilacijom (Slika 3). Za to je potrebna jedna molekula ATP.



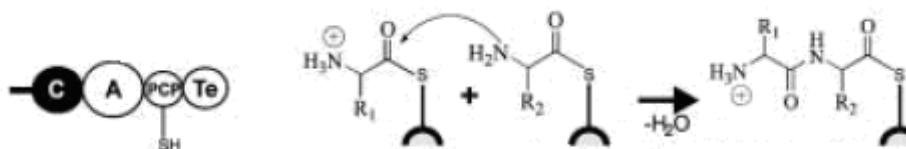
Slika 3. Prepoznavanje i aktivacija supstrata za sintezu neribosomalno sintetiziranog peptida.

Adenilat te rastući peptidni lanac prenose se pomoću fosfopanteteinske ruke na domenu PCP za koji se vežu preko tiolne skupine serina (Slika 4), formirajući aktiviran tioesterski vez.



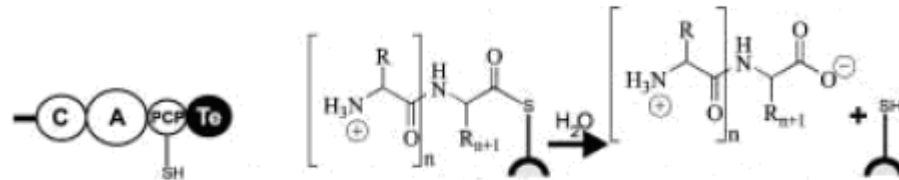
Slika 4. Kovalentno vezivanje aktivirane aminokiseline za PCP domenu preko "pokretne ruke" 4' fosfopanteteina.

Domena C katalizira stvaranje peptidne veze između aminoacil tioestera na domeni PCP istog modula i prethodnog modula (Slika 5) (Challis i Naismith, 2004).



Slika 5. Stvaranje peptidne veze.

Početni (inicijacijski) modul sustava NRPS obično ne sadržava domenu C, dok se na kraju posljednjeg modula nalazi domena tioesteraza (engl. "Thioesterase domain", Te), odgovorna za odvajanje linearnoga peptidnog lanca od enzima i ciklizaciju (Slika 6).



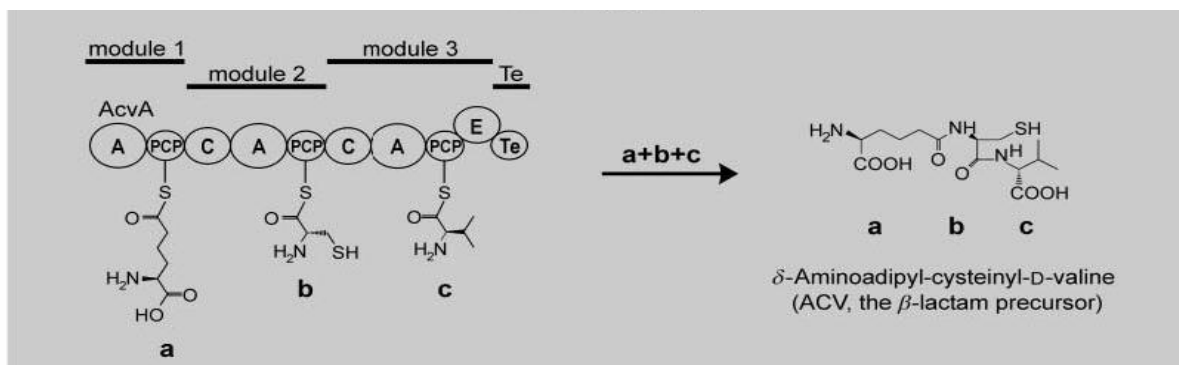
Slika 6. Otpuštanje produkta kidanjem tioesterske veze sa zadnjim modulom.

Prošireni osnovni modul sustava NRPS može još sadržavati i domene: epimerazu (engl. "Epimerisation domain", Ep; obično se nalazi na C terminalnom kraju modula nakon domene PCP i odgovorna je za racemizaciju), metiltransferazu (engl. "Methyltransferase domain", MT; kao dio domene A, katalizira metilaciju dušika u aminima), ciklizacijsku (engl. "Cyclization domain", Cy, umjesto C domene u modulima koji ugrađuju cistein, serin ili treonin, odgovorna je za formiranje heterocikličkih prstenova), oksidacijska (engl. "Oxidation domain", Ox, u kompleksu sa Cy domenom, katalizira oksidaciju tiazolinskog prstena) (Schwarzer i Marahiel, 2001; Sattely i sur., 2008).

Strukturalna raznolikost posljedica je i različite organizacije domena unutar modula, odnosno različitih mehanizama sinteze produkata, zbog čega se sustavi NRPS dijele u tri grupe: linearne, ponavljajuće i nelinearne (Schwarzer i Marahiel, 2001).

### 2.1.1.1. Linearne sintetaze neribosomalnih peptida

U linearnih sustava NRPS svaki se modul upotrebljava samo jednom tijekom sinteze polipeptidnog produkta, a osnovne domene unutar modula su poredane u nizu C-A-PCP. Moduli mogu, ali i ne moraju, sadržavati domene koje nisu osnovne. Mehanizam sinteze prikazan je na primjeru sinteze ACV prekursora (Slika 7).



Slika 7. Mehanizam biosinteze ACV prekursora pomoću linearne sintetaze neribosomalnih peptida.

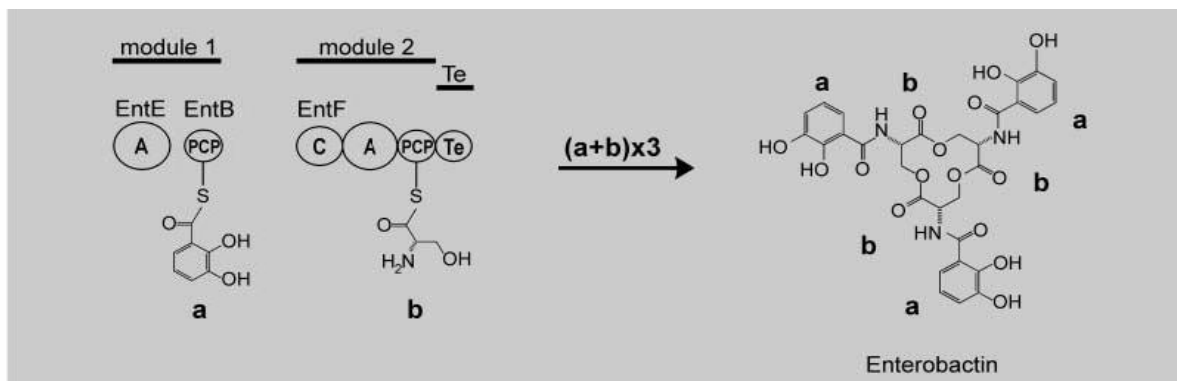
Sekvencija linearnog peptidnog lanca u potpunosti odgovara broju i redosljedu modula. Na temelju te činjenice može se reći da za peptid dužine  $n$  aminokiselina sustav NRPS sadržava  $n$  modula, a organizacija modula je gotovo uvijek A-PCP-(C-A-PCP) $_{n-1}$ -Te. Prvu aminokiselinu rastućeg peptidnog lanca donosi početni modul kojem nedostaje domena C. Na kraju sustava NRPS nalazi se domena Te koja s enzimskog kompleksa oslobađa gotov peptidni lanac. Na Slici 10A prikazani su kemijskim formulama neki od produkata linearnih sustava NRPS). (Mootz i sur., 2002).

### 2.1.1.2. Ponavljajuće sintetaze neribosomalnih peptida

Ponavljajući (tzv. iterativni) sustavi NRPS upotrebljavaju svoje module i domene više puta tijekom sinteze jednog peptidnog produkta. Mehanizam biosinteze peptida enterobaktina tijekom kojeg se moduli a i b upotrebljavaju tri puta, primjer je djelovanja ovog tipa sustava NRPS (Slika 8). Nastali identični međuprodukti vežu se na aktivno mjesto C-terminalnog kraja domene Te sve dok se ne sintetiziraju svi.

Zatim slijedi oligomerizacija u konačni višejedinični peptidni produkt koji se oslobađa nakon ciklizacije (neki od primjera neribosomalnih peptida koji se oslobađaju sa takvih enzimskih sustava prikazan je na Slici 10B). Još uvijek nije poznat mehanizam kontrole koji odlučuje kada sinteza međuprodukta staje i započinje oligomerizacija. Također je nemoguće razlikovati ponavljajuće od linearnih sustava NRPS ako se samo analiziraju primarne

sekvencije DNA ili proteina. Prema dosadašnjim spoznajama ponavljajuće sustave NRPS se može smatrati linearnim sustavima NRPS sa ponavljajućom aktivnošću na krajnjoj domeni PCP (domeni Te ili didomeni PCP-C) (Schwarzer i Marahiel, 2001).



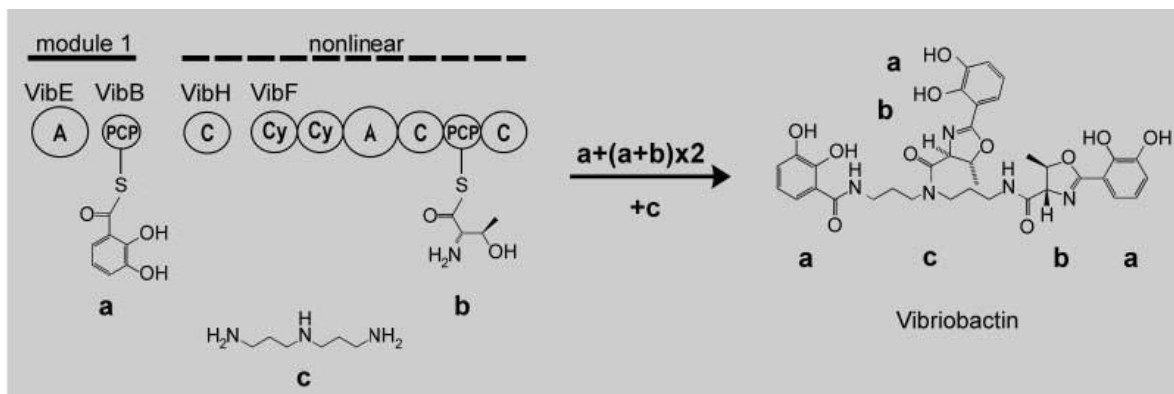
Slika 8. Mehanizam biosinteze enterobaktina pomoću ponavljajuće sintetaze neribosomalnih peptida.

### 2.1.1.3. Nelinearne sintetaze neribosomalnih peptida

Svojstva nelinearnih sustava NRPS jesu: uzastopno ponavljanje pojedinih modula i neobičan raspored domena (odstupanje od organizacije osnovnog modula C, A i PCP) zbog prisustva barem jedne neobične domene unutar modula. Domene koje ne čine osnovni modul mijenjaju strukturu građevnih jedinica, i kao takve potrebne su za nastajanje heterocikličkih prstenova (Schwarzer i Marahiel, 2001). Na primjer, domena Cy (u nekim modulima zamjenjuje domenu C ili pak nakon kondenzacije katalizira povezivanje cisteina, serina ili treonina stvarajući tako tiazolinski ili oksazalinski prsten), te domena Ox (katalizira oksidaciju tiazolinskog prstena u aromatski tiazol upotrebljavajući flavin mononukleotid (FMN) kao kofaktor; ovisna je o domeni Cy) (Challis i Naismith, 2004).

Neobičan raspored domena i prisustvo domene Cy unutar drugog modula prikazan je na Slici 9. Nelinearne sintetaze neribosomalnih peptida kao supstrat također mogu upotrebljavati male, topljive molekule kao što su amini (Schwarzer i Marahiel, 2001). Ovaj tip sustava NRPS vjerojatno je najčešći oblik sustava NRPS u prirodi. Peptidni produkti koji nastaju ovim mehanizmom biosinteze su nelinearni, često s neobičnom, cikliziranom ili

razgranato-cikliziranom strukturom (Slika 10C).



Slika 9. Mehanizam biosinteze vibriobaktina pomoću nelinearne sinteze neribosomalnih peptida.

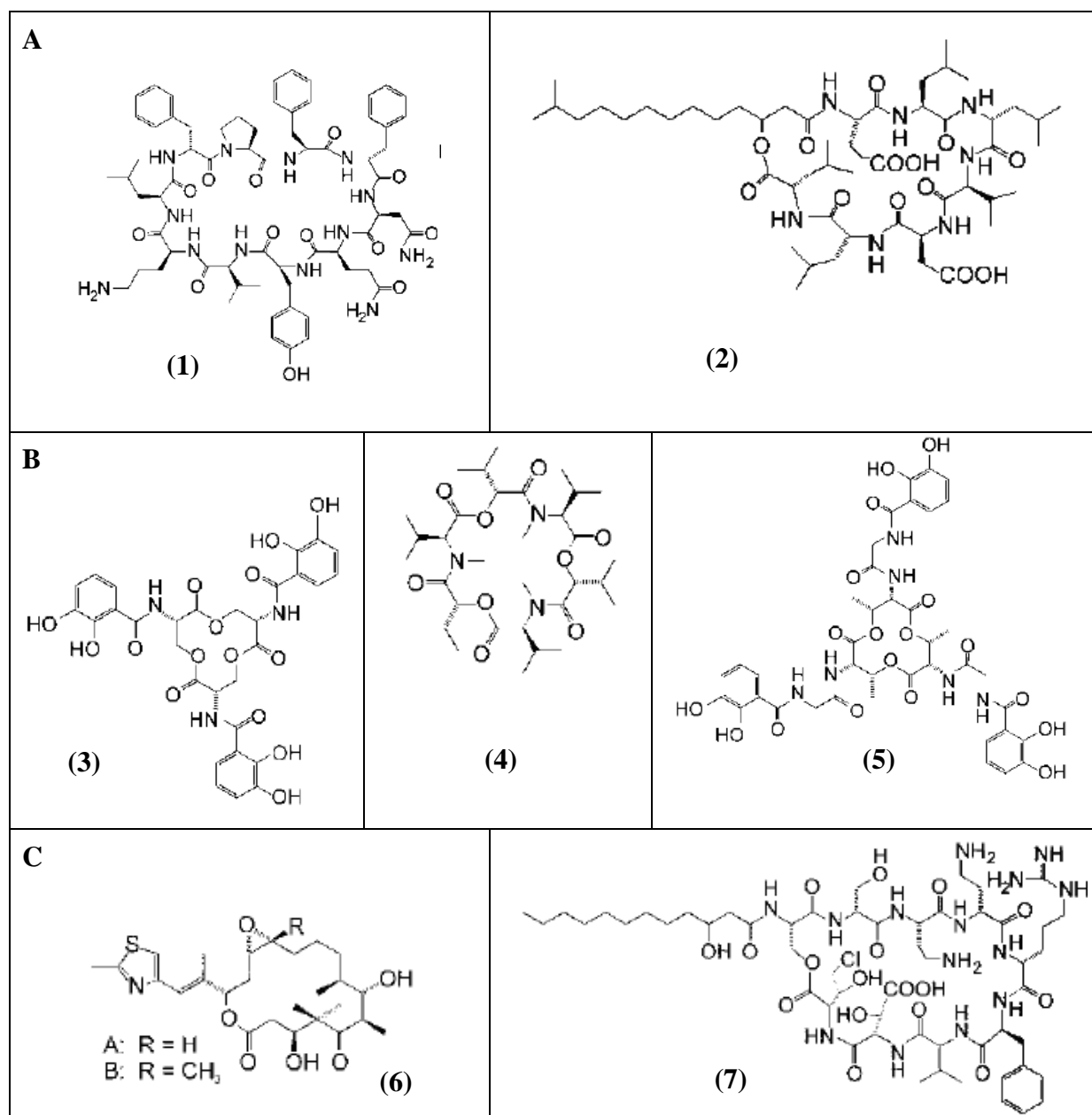
### 2.1.2. Poliketidno/peptidni hibridi

Poliketidno/peptidni sustavi upotrebljavaju module PKS i NRPS prilikom sinteze produkta. U ovu se skupinu prirodnih spojeva ubrajaju i spojevi čiji su moduli cjeloviti sustavi NRPS ili PKS uz pokoju domenu iz drugog tipa modularnih enzima. To je moguće zbog slične modularne strukture, ali i vrlo sličnih reakcija koje kataliziraju pojedine domene. Takav tip organizacije domena otežava mogućnost predviđanja strukture konačnog proizvoda, a dodatan problem predstavlja i izostavljanje (engl. "module skipping") cijelih modula, koji se smatraju neaktivnima, tijekom biosinteze (Wenzel i Müller, 2005). Hibridni sustavi predstavljaju temelj razvoja ideje za manipulaciju modularnim sustavima, međutim zbog složenosti njihove građe neće biti obuhvaćeni u ovom radu.

### 2.1.3. NRPS neovisni siderofori

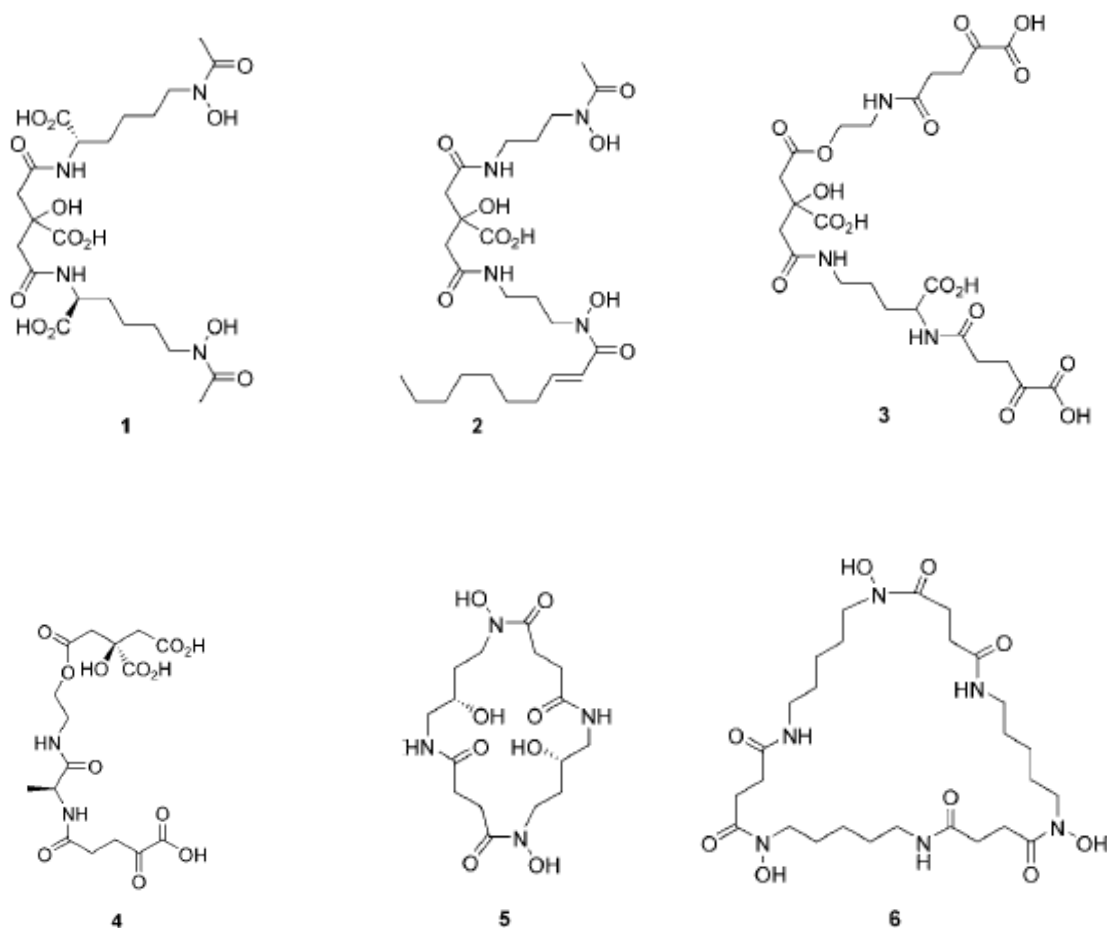
Da bi održale stalne uvjete u unutarnjoj okolini stanica (homeostazu), većina bakterija i plijesni posjeduju različite mehanizme kojima prikupljaju željezo iz okoline. Jedan od vrlo zanimljivih i detaljnije opisanih sustava (s visokim afinitetom za ione željeza) jest asimilacija željeza pomoću siderofora. Mehanizam je prisutan u više od 40 različitih bakterijskih vrsti,

uključujući biljne i životinjske patogene, saprofite i biljne simbiote. Siderofori su molekule male molekularne mase ( $M_r < 1500$ ). Njihova je biosinteza uvjetovana prisutnošću željeza u okolini. Osim što stanicu opskrbljuju izvanstaničnim željezom, upotrebljavaju se i za skladištenje željeza unutar stanice. Istraživanjem ovih biosintetskih mehanizama nastoje se pronaći inhibitori sinteze siderofora, kao potencijalnih antibakterijskih spojeva primjenjivih u medicini i agroindustriji (Barona-Gomez i sur., 2006).



Slika 10. Primjeri neribosomalno sintetiziranih peptida: A) linearni NRPS: tirocidin (1), surfaktin (2); B) ponavljajući NRPS: enterobaktin (3), eniatin (4), bacilibaktin (5); C) nelinearni NRPS: epotilon (6), siringomicin (7).

Mnogi siderofori su polipeptidi sintetizirani pomoću sustava NRPS. Veću skupinu, koja je i mnogo manje istražena predstavljaju siderofori koji nisu polipeptidi. Izgrađeni su od dikarboksi kiselina, diamina ili amino alkohola kao građevnih jedinica povezanih amidnom ili esterskom vezom. Mehanizam njihove sinteze nazvan je "NRPS neovisan biosintetski put siderofora", a takvi siderofori, NRPS neovisni siderofori (engl. "NRPS Independent Siderophore", NIS). Nekoliko takvih primjera siderofora prikazano je na Slici 11. Ovaj mehanizam sinteze siderofora prisutan je u više od 40-ak bakterijskih vrsta, a sve do sada poznate činjenice oko ovog biosintetskog puta temelje se na proučavanju 8 različitih sustava NIS: aerobaktin, rizobaktin, alkaligin, desferioksamin, vibrioferin, stafilobaktin, antrakelin i akromobaktin.



Slika 11. Primjeri NRPS neovisnih siderofora: aerobaktin (1), rizobaktin (2), akromobaktin (3), vibrioferin (4), alkaligin (5) i desferioksoamin E (6).

Mehanizam biosinteze NRPS neovisnog siderofora aerobaktina u bakterije *Escherichia coli* prvi je i najbolje istražen sustav NIS. Kao i u sustavima NRPS i PKS, geni sustava NIS također su organizirani u genske nakupine (Slika 12). Svaka genska nakupina sadržava barem jedan gen koji kodira za NIS siderofor sintetazu (na Slici 12 obojane svijetlo sivo, tamno sivo i crno). NIS siderofor sintetaza katalizira formiranje ključne amidne veze između dvije molekule supstrata, što odgovara C domeni u sustavima NRPS (Challis, 2005). Ishodni supstrat u biosintezi aerobaktina je aminokiselina lizin. Tijekom biosinteze nastaju međuprodukti: N-6-hidroksilizin, N-6-acetil-N-6-hidroksilizin i hidroksamska kiselina (Slika 13). Formiranje ključne amidne veze, koja povezuje dikarboksilnu kiselinu i diamin u konačan produkt kataliziraju enzimi IucA i IucC. Dakle, genska nakupina sustava NIS za sintezu aerobaktina sadržava dvije NIS siderofor sintetaze.

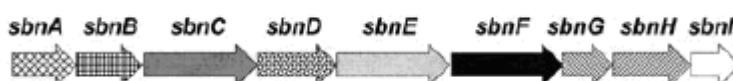
### aerobaktin















### vibrioferin



### stafilobaktin

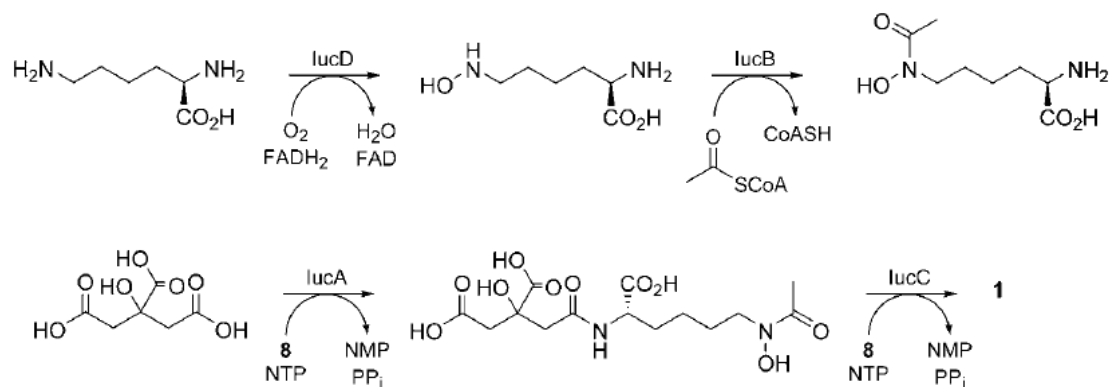


	Tip A siderofor sintetaze		Rieske monooksigenaza
	Tip B siderofor sintetaze		Nepoznata funkcija
	Tip C siderofor sintetaze		Transport/regulacija
	Acil CoA transferaza		Moguća aldolaza
	Flavin-ovisna monooksigenaza		Ornitiin ciklodeaminaza
	PLP-ovisna dekarboksilaza		O-acetil serin sulfhidrolaza
	PLP-ovisna aminotransferaza		Acetil CoA sintetaza/ ACP
	ATP-ovisna ligaza aminokiselina		

Slika 12. Prikaz primjera genskih nakupina uključenih u biosintezu NRPS neovisnih siderofora.



Analizom genskih nakupina drugih sustava NIS utvrđeno je da sadržavaju barem jedan ili više gena koji pokazuju visok stupanj sličnosti sa *iucA* i *iucC*. Višestrukim poravnanjem 88 sekvencija DNA koje kodiraju za NIS siderofor sintetaze zaključeno je da postoje tri skupine NIS siderofor sintetaza (Tablica 1) (Challis, 2005).



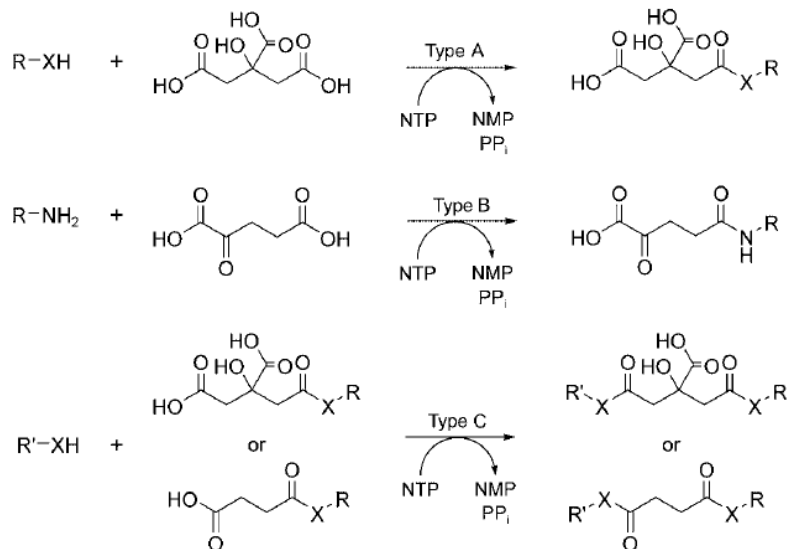
Slika 13. Mehanizam biosinteze NRPS neovisnog siderofora aerobaktina (1).

Tablica 1. Podjela siderofor sintetaza u sustavima NIS.

SKUPINA	HOMOLOGNI ENZIMI
Tip A	<i>IucA</i> , <i>RhbC</i> , <i>PvsD</i> , <i>AcsD</i> , <i>SbnE</i> , <i>AsbA</i>
Tip B	<i>PvsB</i> , <i>AcsA</i> , <i>SbnC</i>
Tip C	<i>IucC</i> , <i>RhbF</i> , <i>AcsC</i> , <i>SbnF</i> , <i>AsbB</i> , <i>AlcC</i> , <i>DesD</i>

Na osnovi 8 analiziranih mehanizama biosinteze NRPS neovisnih siderofora (aerobaktin, rizobaktin, alkaligin, desferioksamin, vibrioferin, stafilobaktin, antrakelin i akromobaktin) predložena su dva modela prema kojima siderofor sintetaze vjerojatno djeluju unutar sustava NIS. Prema prvom modelu (Slika 14) svaka skupina NIS siderofor sintetaza pokazuje specifičan afinitet prema karboksilnoj kiselini (prvom supstratu), međutim, ne i prema drugoj vrsti supstrata (amin ili alkohol). Takve NIS siderofor sintetaze, pretpostavlja se, kataliziraju nastajanje:

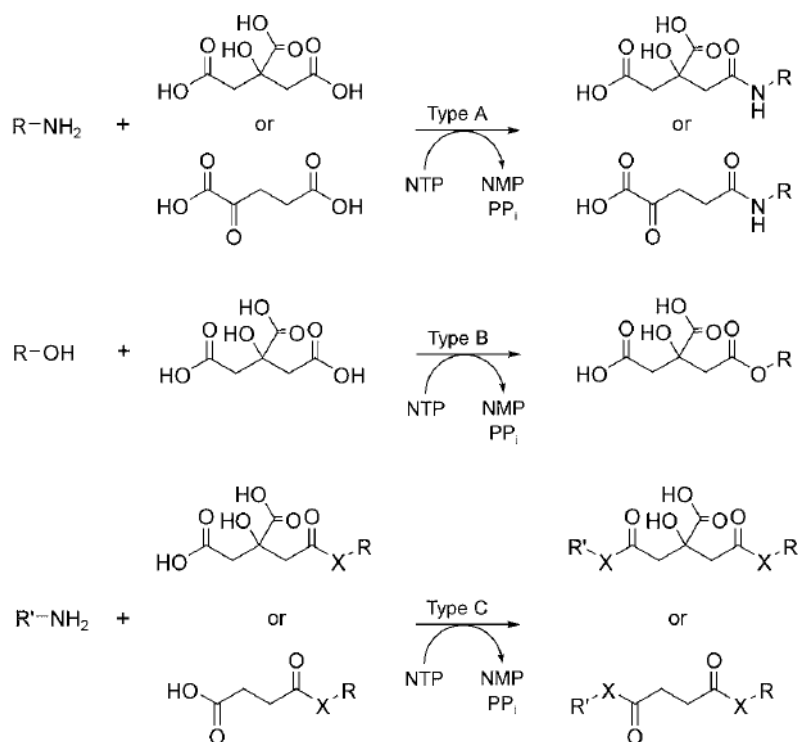
- amidnih ili esterskih veza između amino ili hidroksilne skupine u različitim supstratima i jedne od dvije prokiralne karboksilne skupine u citratnoj kiselini,
- amidne veze između amino ili hidroksilne skupine u različitim supstratima i C5 karboksilne grupe  $\alpha$ -ketoglutarata,
- amidne ili esterske veze između amina ili alkohola i karboksilne skupine unutar monoamida ili monoestera citratne kiseline ili sukcinata.



Slika 14. Pretpostavljeni mehanizam djelovanja NIS siderofor sintetaza unutar sustava NIS, prema prvom modelu (X = NH<sub>2</sub> ili OH).

Prema drugom modelu svaka skupina NIS siderofor sintetaza pokazuje specifičan afinitet prema jednom od supstrata, alkoholu ili nukleofilnom aminu, ali ne i prema karboksilnim kiselinama. Drugi model također pretpostavlja tri mehanizma (Slika 15):

- nastajanje amidne veze između amina i jedne prokiralne karboksilne grupe u citratu ili C5 karboksilne grupe  $\alpha$ -ketoglutarata,
- nastajanje esterske veze između alkohola i jedne prokiralne karboksilne grupe citratne kiseline,
- nastajanje amidne veze između amino grupe različitih supstrata i derivata monoamida ili monoestera citratne kiseline ili sukcinata.



Slika 15. Pretpostavljeni mehanizam djelovanja NIS siderofor sintetaza unutar sustava NIS, prema drugom modelu ( $X = \text{NH}_2$  ili  $\text{OH}$ ).

Trenutno se prvi model čini prihvatljivijim zbog dva osnovna razloga, to su:

- svaki tip enzima mora tolerirati znatne strukturalne varijacije unutar molekula nukleofilnih supstrata, bez obzira da li je supstrat amin ili alkohol, i
- nukleofilni amini se lako protoniraju pri fiziološkom pH. Bez obzira da li je nukleofil amino ili hidroksilna skupina potrebna je baza (proton akceptor) da bi se deprotonirao.

Iz literaturnih podataka moguće je zaključiti da se podjela NIS siderofor sintetaza, u tri skupine, temelji na specifičnosti afiniteta enzima prema supstratima (Challis, 2005). Međutim, potrebna su dodatna biokemijska i genetička istraživanja kako bi se potvrdio jedan od dva predložena modela.

## 2.2. BIOINFORMATIČKI ALATI ZA ANOTACIJU GENSKIH NAKUPINA

Napredak na području bioloških i računalnih znanosti omogućio je pohranjivanje velikog broja genoma i proteoma, kao i njihovu organizaciju i detaljnu analizu. Razvijene su

prve javno dostupne baze podataka i bioinformatički alati koji omogućuju predviđanje genetičkih elemenata: strukturnih gena, pseudogena, regulatornih regija, dijelova molekule DNA što ne sadržavaju genetičku uputu i ponovljenih sekvencija. Oblikovanje novih supstancija u proizvodnji novih lijekova manipulacijom genskim nakupinama PKS i NRPS u uvjetima *in vitro*, znatno je olakšana oblikovanjem strukturiranih, javno dostupnih baza podataka (poput GenBank, NRPS-PKS) koje sadrže dobro opisane, to jest anotirane sekvencije genskih nakupina (Hranueli i sur., 2008). Za analizu genskih nakupina poliketida i neribosomalno sintetiziranih peptida razvijeno je nekoliko bioinformatičkih alata koji upotrebljavaju donekle različite strategije i pristupe. Takvi bioinformatički alati i programski paketi omogućuju analizu i manipulaciju genskim nakupinama čija funkcija i konačni kemijski produkt još nisu poznati. Neki od alata navedeni su u tekstu koji slijedi.

### **2.2.1. NRPS predictor, NP.searcher, SBSPKS i drugi alati**

Baza podataka NRPS-PKS omogućuje korisnicima povezivanje kemijskih struktura prirodnih spojeva s domenama i modulima u odgovarajućim sustavima NRPS i PKS (Ansari i sur., 2004). Tvrtka Ecopia BioSciences Inc. razvila je alat *DecipherIT<sup>TM</sup>* koji automatski anotira genske nakupine mikroorganizama iz tla (Zazopoulos i sur., 2003). Sličnu primjenu ima i alat pod nazivom *Biogenerator*, norveške tvrtke Biosergen AS, koji omogućuje i predviđanje biološke aktivnosti kemijskih supstancija (Zotchev i sur., 2006). Računalni program NRSPredictor, predviđa specifičnosti domena adenilacije (domena A), genskih nakupina NRPS, za odabir građevnih jedinica tijekom biosinteze neribosomalno sintetiziranih peptida (Rausch i sur., 2005). Drugi je koristan izvor podataka baza podataka ASMPKS s računalnim programom *MAPSI* (Tae i sur., 2007). Analizu izoliranih genskih nakupina omogućuje računalni program CLUSEAN (Weber i sur., 2008).

Računalni program NP.searcher traži genske nakupine prirodnih produkata kao što su poliketidi, neribosomalni peptidi i hibridne molekule u sekvenciranim genomima. Omogućuje i predviđanje moguće strukture konačnog produkta primjenjujući 20 standardnih aminokiselina i 22 druga intermedijera u biosintezi neribosomalno sintetiziranih peptida, te četiri acil-coA građevne jedinice poliketida (Li i sur., 2009).

Internetski poslužitelj SBSPKS (engl. "Structure Based Sequence Analysis of Polyketide Synthases") omogućuje analizu genskih nakupina sustava PKS i strukturni prikaz

više-enzimskih sustava. Organiziran je u tri glavne funkcionalne cjeline: NRPS\_PKS, Model\_3D\_PKS i Dock\_Dom\_Anal. Baza podataka NRPS-PKS je proširena verzija već prije spomenute istoimene baze podataka. Sadržava 131 sekvenciju DNA genskih nakupina sustava PKS i NRPS te oko 4.397 katalitičkih domena, s poznatim afinitetom prema supstratu, sustava PKS i NRPS. Model\_3D\_PKS omogućuje predviđanje tercijarne i kvaterne strukture više-enzimskog modularnog sustava PKS uz pomoć Dock\_Dom\_Anal funkcije koja predviđa sve moguće kombinacije interakcija između polipeptidnih lanaca, odnosno njihovih poveznica i domena smještenih na terminalnim krajevima (Anand i sur., 2010). Teorija poveznica i domena uključuje formiranje kovalentnih ili nekovalentnih interakcija između poveznica i domena različitih modula koji su dio različitih polipeptidnih lanaca. Kovalentne interakcije omogućuju formiranje jednog polipeptidnog lanca, tj. tercijarnu strukturu, a nekovalentne formiranje modularnog sustava PKS kvaterne strukture (Weissman, 2004).

### 2.2.2. Programski paket *ClustScan*

Programski paket *ClustScan* pruža mogućnost brze, poluautomatske anotacije modularnih biosintetskih genskih nakupina pomoću na znanju utemeljenih predviđanja aktivnosti i specifičnosti njihovih modula i katalitički aktivnih domena. Nakon učitavanja, sekvencija, molekule DNA se automatski prevode u šest okvira čitanja u rezultatne sekvencije proteina pomoću programa *Transeq* (Rice i sur., 2000). Programski paket *ClustScan* omogućuje traženje gena na temelju sekvencije transliranoga proteina pomoću programa *GeneMark-PS* (Besemer i Borodovsky, 2005) ili *Glimmer* (Delcher i sur., 2007). Katalitički aktivne domene mogu se prepoznati unutar pronađenih enzima pomoću programskog paketa *HMMER* (Eddy, 1998) upotrebom proteinskih profila preuzetih iz baze podataka Pfam (Finn i sur., 2008), ili pomoću vlastitih proteinskih profila, strogo definiranim ili relaksiranim parametrima. Nakon obavljene anotacije genske nakupine i biosintetskoga puta, konačnu je anotaciju moguće "eksportirati" u obliku zapisa *GenBank*, *EMBL* ili *XML* i upotrijebiti je u drugim aplikacijama ili za upis u baze podataka GenBank (Anonymous 1, 2010) odnosno EMBL (Anonymous 2, 2010) (Starcevic i sur., 2008).

Starčević i suradnici (2008) usporedili su uspješnost generičkog programskog paketa *ClustScan* (Tablica 2) u odnosu na sustave *SEARCHPKS* i *MAPSI* (Yadav i sur., 2003; Tae i sur., 2007). Za procjenu uspješnosti *ClustScan* programa važna su dva kriterija. To su

funkcionalnost i točnost predviđanja, te brzina i prikladnost programa za anotaciju velikih skupina sekvencija molekula DNA. Točnost predviđanja je demonstrirana na primjeru jedne dobro opisane genske nakupine što sadržava genetičku uputu za biosintezu poliketidnog antibiotika eritromicina (Starcevic i sur., 2008).

Tablica 2. Usporedba generičkog računalnog programskog paketa *ClustScan* sa sustavima *SEARCHPKS* i *MAPSI*.

<b>SVOJSTVA</b>	<b><i>CLUSTSCAN</i></b>	<b><i>SEARCHPKS</i></b>	<b><i>MAPSI</i></b>
Učitavanje sekvencije DNA	Da	Samo protein	Da <sup>1</sup>
Prepoznaje specifičnost domena AT	Da	Da	Da
Prepoznaje stereokemiju domena KR	Da	Ne	Ne
Prepoznaje neaktivne domene KR	Da	Ne	Ne
Prepoznaje neaktivne domene DH	Da	Ne	Ne
Prepoznaje neaktivne domene ER	Da	Ne	Ne
Uređivanje predviđanja	Da	Ne	Ne
Eksport zapisa anotacije	Da	Ne	Ne
Predviđanje kemijskih struktura	Da	Ne	Da <sup>2</sup>
Eksport kemijske strukture u standardnom obliku	Da	Ne	Ne

<sup>1</sup> Zahtjeva dugačke sekvencije DNA (> 200 kb za bakterije visokog G+C-sastava) za preciznu identifikaciju gena

<sup>2</sup> Ograničeno predviđanje kemijskih struktura bez mogućnosti njihova uređivanja

### **3. EKSPERIMENTALNI DIO**

## **3.1. MATERIJAL**

### **3.1.1. Računalna podrška i operativni sustav**

Ovaj rad je izrađen na računalu slijedećeg sklopovlja: prijenosno računalo HP: procesor Intel® Core™ 2Duo CPU T5470 1.60GHz 782MHz, radne memorije 0,99GB. Na čvrstom disku upotrijebljenog računala instaliran je operativni sustav "Microsoft Windows XP Professional".

### **3.1.2 Baze podataka**

#### **3.1.2.1. Baza podataka GenBank**

Baza podataka GenBank osnovana je u listopadu 1992. godine i održava je Nacionalni centar za biotehnoške informacije (engl. "National Center for Biotechnology Information", NCBI). Ona sadržava dobro opisane, tj. anotirane sekvencije DNA. U njoj se može pronaći 85.759.586.764 parova baza iz 82.853.685 analiziranih sekvencija DNA. Cjeloviti se pregled trenutne verzije baze podataka GenBank nalazi na poslužitelju centra NCBI (Anonymous 1, 2010). Baza podataka GenBank dio je međunarodne suradnje (engl. "International Nucleotide Sequence Database Collaboration") koja pruža mogućnost neprestane razmjene i dostupnosti sekvencija DNA s bazama podataka EMBL (engl. "European Molecular Biology Laboratory") i DDBJ (engl. "DNA Data Bank of Japan"). Baza se svakodnevno nadopunjava, a svaka dva mjeseca izlazi novo izdanje.

#### **3.1.2.2. Baza podataka NRPS-PKS**

Baza podataka NRPS-PKS, utemeljena pri Nacionalnom centru za imunologiju (engl. "National Institute of Immunology", NII), New Delhi, Indija, sadržava računalni program za analiziranje velikih više-enzimskih, više-domenskih megasintetaza uključenih u biosintezu farmaceutski važnih prirodnih produkata (Anonymous 3, 2004; Ansari i sur., 2004). Na internetskoj stranici baze podataka NRPS-PKS nudi se mogućnost jednostavne izolacije različitih domena iz vlastite polipeptidne sekvencije i određivanja katalitičkih aktivnosti domena, aktivnih mjesta, specifičnosti za supstrat te omogućuje usporedbu domena neopisanih NRPS/PKS genskih nakupina na temelju njihove sličnosti, specifičnosti za supstrat te motiva aktivnog mjesta. Rezultati tih analiza organizirani su u četiri integrirane



baze podataka za pretraživanje i određivanje organizacije domena i specifičnosti za supstrate neribosomalnih peptid sintetaza i tri tipa poliketid sintaza. Te su baze podataka nazvane NRPSDB (engl. "A Database of Non-Ribosomal Peptide Synthetases"), PKSDB (engl. "A Database of Modular Polyketide Synthases"), ITERDB (engl. "A Database of Type I Iterative PKS Gene Clusters") i CHSDB (engl. "A Database of Chalcone Synthases"), prema tipu prirodnih produkata i mehanizmu biosinteze. Međusobno su povezane kako bi se mogle provoditi analize genskih nakupina odgovornih za biosintezu poliketidno/peptidnih hibridnih produkata. Obradivanjem velikog broja biosintetskih genskih nakupina dokazano je da, uz točno određivanje domena NRPS i PKS sustava, baza podataka NRPS-PKS može također predvidjeti specifičnosti adenilacijskih i aciltransferaznih domena sa relativno velikom točnošću. Ova svojstva bazu podataka NRPS-PKS čine vrijednim izvorom za identifikaciju prirodnih produkata biosintetiziranih pomoću NRPS/PKS genskih nakupina pronađenih u novosintetiziranim genomima. Baza NRPS-PKS putem grafičkog sučelja omogućuje korisnicima povezivanje kemijskih struktura prirodnih spojeva sa domenama i modulima u odgovarajućim neribosomalnim peptid sintetazama ili poliketid sintazama. Također nudi smjernice za zamjenu domena/modula kao i za eksperimente za mjesno specifičnu mutagenezu koji omogućuju biosintezu novih prirodnih produkata. Bazi podataka NRPS-PKS moguće je pristupiti preko internetske stranice <http://www.nii.res.in/nrps-pks.html> (Anonymous 3, 2004; Ansari i sur., 2004).

### 3.1.2.3. Baza podataka Pfam

Baza podataka Pfam je velika zbirka višestruko poravnatih proteinskih domena koje su organizirane u proteinske obitelji i čine skrivene Markovljeve modele. Za svaku proteinsku obitelj unutar baze podataka Pfam moguće je pregledati: višestruko poravnanje, strukturu proteinskih domena, filogenetska stabla, poveznice na druge baze podataka i poznate strukture proteina. Jedan protein može pripadati u nekoliko Pfam obitelji, a 74 % sekvencija proteina ima barem jednu sličnu proteinsku sekvenciju unutar baze podataka Pfam. Taj broj se zove pokrivenost baze. Baza podataka Pfam-A je ručno izrađeni dio baze koji sadržava više od 9.000 unosa. Za svaki unos pohranjena su poravnanja proteinskih sekvencija i skriveni Markovljevi modeli. Postoji još i baza podataka Pfam-B koja sadržava veliki broj malih obitelji slabije kvalitete podataka. Ona se koristi u slučajevima kada nisu pronađene sekvencije sa sličnošću unutar baze podataka Pfam-A (Finn i sur., 2008).

### 3.1.3. Bioinformatički računalni paketi i programi

#### 3.1.3.1. Računalni program *Glimmer*

Računalni program *Glimmer* (engl. "Gene Locator and Interpolated Markov ModelER") služi za pronalaženje gena u DNA mikroorganizama, posebno u genomima bakterija, arheja i virusa. Razvijen je i prvotno upotrijebljen u instituciji "The Institute for Genomic Research" (TIGR) za anotaciju više od 100 različitih bakterijskih vrsta. Programom *Glimmer* može se konstruirati model za dijelove DNA s genetičkom uputom, koristeći druge otvorene okvire čitanja u učitanoj (analiziranoj) sekvenciji DNA kao podatak za uvježbavanje (engl. "training data"). Manje je učinkovit prilikom analize kratkih sekvencija DNA, kao i u sekvencija DNA s visokim G+C sastavom baza. Takvu DNA imaju upravo genomi streptomiceta. To su dugački dijelovi DNA koji ne sadržavaju genetičku uputu pa mogu smanjiti preciznost predviđanja u dijelova DNA s genetičkom uputom (Delcher i sur., 2007).

#### 3.1.3.2. Računalni program *GeneMark-PS*

Računalni program *GeneMark* je razvijen 1993. godine kao prvi bioinformatički program za prepoznavanje gena. Učinkovit je i precizan alat za analizu genoma. Sadržava knjižnicu modela različitih vrsta bakterija. Prilikom pretraživanja željene sekvencije DNA odabire se model koji je najbliži vrsti koja je izvor analizirane DNA (Besemer i Borodovsky, 2005).

#### 3.1.3.3. Programski paket *HMMER*

Programski paket *HMMER* upotrebljava, za analizu pojedinih aminokiselina u sekvenciji proteina, profile proteina u obliku skrivenih Markovljevih modela (engl. "hidden Markov models", HMM). Zbog toga programski paket dobro razlikuje ključne konzervirane aminokiseline (odgovorne za katalitičku aktivnost) od onih manje važnih. Pored toga, programski paket *HMMER* uzima u obzir insercije i delecije te im pridodaje stupanj vjerojatnosti (Eddy, 1998). Funkcija *hmmpfam* (sastavni dio paketa *HMMER*) čita sekvenciju po sekvenciju upita i traži sličnost sa nekim od proteinskih profila u bazi podataka. Profili proteina koji se nalaze u bazi podataka koju funkcija *hmmpfam* čita, organizirani su u obitelji proteina (npr. obitelj proteina globina). Sve sekvencije unutar jedne obitelji proteina pokazuju značajan stupanj sličnosti. Uspješnost određivanja sličnosti točnija je prilikom usporedbe

dviju sekvencija proteina, nego u usporedbi dviju sekvencija DNA. Za promatranu sekvenciju, moguće je, pomoću parametriziranih modela, odrediti koji model najvjerojatnije objašnjava podatke, tj. promatranu sekvenciju te koji je najvjerojatniji put kroz "stanja" za sekvenciju i za model.

#### 3.1.3.4. Programski paket *ClustScan*

Računalni generički programski paket *ClustScan* je razvijen 2008. godine u Kabinetu za bioinformatiku, Prehrambeno-biotehnološkog fakulteta, Sveučilišta u Zagrebu. Generički programski paket *ClustScan* upotrebljava se za anotaciju modularnih genskih nakupina sa na znanju utemeljenim predviđanjima aktivnosti i specifičnosti njihovih modula i katalitički aktivnih domena (Starčević i sur., 2008). Tim se programskim paketom mogu anotirati genske nakupine u novosekvenciranim genomima i metagenomima kao i predvidjeti kemijske strukture produkata biosinteze iz sekvencija DNA genskih nakupina. Računalni generički programski paket *ClustScan* može se preuzeti sa Web stranice <http://bioserv.pbf.hr/cms/index.php?page=clustscan> (Anonymous 4, 2009). Za detaljniji opis vidi podpoglavlje 2.2.2.

### 3.1.4. Sekvencirani genomi vrsta *Actinomyces*

#### 3.1.4.1. Genom bakterije *Streptomyces scabies*

Vrsta *Streptomyces scabies* je biljni patogen koji uzrokuje krastavost krumpira, ali i ostalog korjenastog povrća poput repe, pastrnjaka, mrkve i rotkve (Lambert i Loria, 1989). Ovu bakteriju karakterizira pseudomicelijski rast. Pseudohife su vegetativni oblici stanica koji se lome ili fragmentiraju u spore.

Sekvenciranje genoma vrste *S. scabies* soj 87.22 je završeno i javno dostupno, ali njegova anotacija još nije objavljena. Pošlo se od pretpostavke da je genom vrste *S. scabies* dugačak oko 8,1 Mb te da sadržava 72 % G+C parova baza. Kromosom bakterije sastoji se od 10.148.695 bp (eng. "base pairs", parovi baza) s G+C sadržajem od 71,45 %. Genom je sekvencioniran nasumičnom metodom (eng. "shotgun"), a na stranici su dostupne sve njegove sekvencije. Ukupno ima 158.304 očitavanja (eng. "reads"), što čini 89.897 Mb. Prema tome genom je teoretski pokriven 99,99 % (Anonymous 5, 2010).

## **3.2. METODE RADA**

### **3.2.1. Prikupljanje literaturnih podataka**

Prilikom prikupljanja literaturnih podataka upotrebljavani su mrežni pretraživači literature kao što su: "Google Scholar" (Anonymous 6, 2010) i "Science direct" (Anonymous 7, 2010).

### **3.2.2. Prikupljanje sekvencija DNA i proteina**

Kod prikupljanja primarnih sekvencija DNA i proteina, pristupano je već postojećim bazama podataka GenBank i NRPS-PKS putem Interneta. Za pristup bazama podataka upotrebljavan je Web preglednik "Google Chrome". Baze podataka koje su uključene u "International Nucleotide Sequence Database Collaboration", a u koje spada i baza podataka GenBank, povezuje sustav za pretraživanje i prikupljanje podataka, Entrez. Odabirom "Entrez home" (Anonymous 1, 2010) i upisivanjem ključne riječi u polje "Search across databases" pokreće se željeno pretraživanje. Upisivanjem pristupnog broja (engl. "accession number") u polje "Search across databases" drugi je način na koji je pretraživana baza podataka GenBank pomoću sustava Entrez. Na primjer, upisivanjem NC\_000964.2 (pristupni broj baze podataka GenBank za neribosomalno sintetizirani peptid balhimicin pronađen u prikupljenim literaturnim podacima) i pokretanjem pretraživanja, odmah se učitao traženi zapis GeneBank. Odabirom sučelja "Nucleotide" dobio se ispis svih pronađenih sekvencija DNA. Odabirom svakog pojedinog ispisa otvara se zapis baze podataka GenBank.

Zapis baze podataka GeneBank (Slika 16) sadržava mnoštvo informacija, počevši od raznih opisa sekvencije, pa do same sekvencije nukleotida. Ovaj oblik zapisa uz podatke sadržava i komentare, tj. pojašnjenja, koja pružaju mogućnost jednostavnijeg čitanja, identifikacije i upotrebe različitih tipova podataka iz baze. Odabirom zapisa FASTA u izborniku "Format" (Slika 16, zaokruženo crveno) promijenjen je oblik zapisa (Slika 17). Zapis FASTA je jednostavan, minimalistički zapis koji se sastoji od prvog retka zvanog zaglavlje (engl. "header"). Zaglavlje započinje sa znakom ">" nakon kojeg slijedi kratki opis sekvencije. Nakon toga, u novom retku slijedi sama sekvencija DNA ili proteina. Zapis FASTA je računalni standard, stoga su sekvencije sačuvane u takvom obliku. Pretraživač Entrez također pruža mogućnost preuzimanja sekvencija proteina vezanih za svaku sekvenciju

DNA (Slika 16, zaokruženo zeleno) tako da su i one pohranjene.

Format: [GenBank](#) [FASTA](#) [Graphics](#) [More Formats](#) ▼

GenBank: Y16952.3

## Amycolatopsis balhimycina biosynthetic gene cluster for balhimycin, strain DSM 5908

[Comment](#) [Features](#) [Sequence](#)

LOCUS Y16952 66669 bp DNA linear BCT 14-DEC-2008  
DEFINITION Amycolatopsis balhimycina biosynthetic gene cluster for balhimycin, strain DSM 5908.  
ACCESSION Y16952  
VERSION Y16952.3 GI:46275283  
KEYWORDS abc gene; bbr gene; bgtfA gene; bgtfB gene; bgtfC gene; bhaA gene; bhp gene; bmt gene; bpsA gene; bpsB gene; bpsC gene; bpsD gene; dihydroxyphenylacetic acid synthase; dpgA gene; dpgB gene; dpgC gene; dpgD gene; enoyl-CoA hydratase; enoyl-CoA-isomerase; glycosyl transferase; halogenase; hydrolase; hydroxyacyl-dehydrogenase; ORF1; ORF10; ORF11; ORF2; ORF3; ORF5; ORF6; ORF7; ORF8; ORF9; orfX; oxyA gene; oxyB gene; oxyC gene; oxyD gene; P450 monooxygenase; pdh gene; peptide synthetase; pgat gene; phenylglycine amino transferase; putative ABC transporter ATP-binding protein; putative prephenate dehydrogenase; putative two-component system respons; putative two-component system sensor kinase; putative VanY-type carboxypeptidase; StrR family transcriptional regulator; vanR gene; vanS gene; vanY gene.  
SOURCE Amycolatopsis balhimycina  
ORGANISM [Amycolatopsis balhimycina](#)  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.  
REFERENCE 1  
AUTHORS Pelzer,S., Sussmuth,R., Heckmann,D., Recktenwald,J., Huber,P., Jung,G. and Wohlleben,W.  
TITLE Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908  
  
COMMENT On Apr 7, 2004 this sequence version replaced gi:[15131491](#).  
FEATURES Location/Qualifiers  
source 1..66669  
/organism="Amycolatopsis balhimycina"  
/mol\_type="genomic DNA"  
/strain="DSM 5908"  
/db\_xref="taxon:[208443](#)"  
[gene](#) complement (<1..759)  
/gene="vanS"  
[CDS](#) complement (<1..759)  
/gene="vanS"  
/codon\_start=1  
/transl\_table=[11](#)  
/product="putative two-component system sensor kinase"  
/protein\_id="[CAG25751.1](#)"  
/db\_xref="GI:46275283"  
/translation="MDRAAGMSVRLKLTLSYACFLVLAGVLLLASVWLFLLRDVPDVLAKPPPGVRLERSVLRNFLPAAGSVLFFLLLFGLLGGWILAGRMLAPLTRITDAARMA"

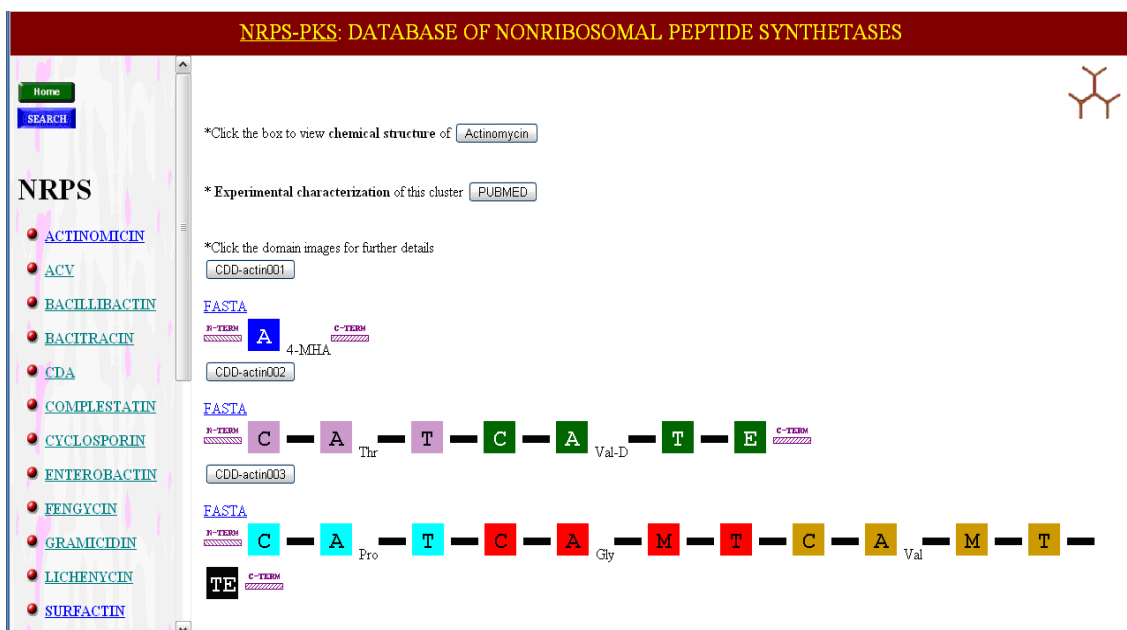
Slika 16. Dio zapisa baze podataka GeneBank za sekvenciju DNA balhimicina.

GenBank: Y16952.3

### Amycolatopsis balhimycina biosynthetic gene cluster for balhimycin, strain DSM 5908

```
>gi|46275283|emb|Y16952.3| Amycolatopsis balhimycina biosynthetic gene cluster
for balhimycin, strain DSM 5908
CTGCAGCAGCAGCGTCGCGGACCCACCACCCGGCTGATGTGACCGGAGGGCATGATGACGACACGGCGT
TTCTCGGCGATCGGGAGCAGCGTTTCGATGGCCTCTTCCACAAGCAGGGACAGGTGCGACCGGTTCCCGGG
TGAACCGCGCGCTGATCGGCGCGGGCTGAGCACCAGCAGCGCCTCGGTGAGGTGATCGCCGGGCGTTCAC
CGCGTGGAGGCGGTCGAAGACCAGCAGCGGGTCTTTCGCCGGATCGTTGCGGGCCACTTCGAGAAGCGCC
TGCCTGATCGCCAGCGGGGTGCGCAGCTCGTGCAGGGCGTTCGCGGGCGAACCTCCGCTGCGCGGCGACGT
GGGCTTCGAGCCGGGCGAGCATGGCGTCGAAGGCATCGGGCGAGTTTCGGGAACTCGTCTTCGGTGCCTTC
CAGCCCGGATCCGGTGGGAGAGCGACCCGTTTCGCGGCCATGCGCGCGCGTTCGGTATCCGGGTCAGCGGG
GCGAGCATGCGGGCGGGGAGGATCCAGCCGCCAGCAGGCCGAACAGCAGCAGGAAGAAGCAGCACTGAGC
CCGCGGCCCGCAGGAAGTTTGGCGACAGGACCGGCTCCAGCACCCCGCCGGGGGGCGGTTTGGCCAG
CACGTCGGGGACATCACGCAGCAGGAACAGCCACACGGAAGCGAGCAGCAGTACACCGGCGAGCAGGAGG
```

Slika 17. Dio zapisa FASTA baze podataka GeneBank za sekvenciju DNA balhimicina.



Slika 18. Prikaz modula, domena i njihovih poveznica za aktinomycin u NRPSDB. Geni su prikazani odvojeno sa modularnom organizacijom i specifično obojenim imenovanim domenama. Poveznice između domena i modula su prikazane punim crnim crtama.

Baza podataka NRPS-PKS poslužila je kao znanstveno utemeljen izvor podataka za analizu neribosomalno sintetiziranih peptida. Primjerice, pretraživanje baze NRPSDB, koja se nalazi u sklopu baze NRPS-PKS, započinje odabirom jedne od ponuđenih biološki aktivnih supstancija s lijeve strane početne stranice. Moguće je pregledati kemijsku strukturu svake

supstancije te literaturne podatke (iz baze podataka PubMed), grafički prikaz modula i domena poliketid sintaza i neribosomalnih peptid sintetaza odgovornih za nastanak spoja (Slika 18). Za svaki se odabrani element (gen, modul, domenu i poveznicu) može dobiti sekvencija proteina u obliku zapisa FASTA.

Sve prikupljene sekvencije DNA i proteina u obliku zapisa FASTA nalaze se u prilogu (vidi: podpoglavlje: 8.2.1.).

### **3.2.3. Analiza sekvencija proteina programskim paketom *HMMER***

Sve prikupljene sekvencije proteina koje pripadaju istoj klasifikacijskoj grupi (unutar klasifikacijske tablice izrađene prema prethodno obrađenim literaturnim podacima, (vidi: podpoglavlje 4.1., Tablica 3 i 4, podpoglavlje 8.2.1., tablica 3P i 4P) pohranjene su u jedan zapis FASTA sa zajedničkim zaglavljem, spajanjem pojedinačnih sekvencija pomoću alata "WordPad". Nakon spajanja, svaki je zapis FASTA obrađen računalnim programom *Readseq* (Anonymous 8, 2010). Program *Readseq* upotrijebljen je za validiranje zapisa FASTA. Sve se datoteke nalaze u prilogu (vidi: podpoglavlje 8.2.2.). Kao izlazni oblik sekvencije (eng. "Output sequence format") odabran je Pearson|Fasta|fa zapis te opcija uklanjanja praznina (engl. "Remove gap symbols"). Sekvencije se nakon obrade neposredno spremaju na čvrsti disk u obliku ispravnog FASTA zapisa.

Spremljene proteinske sekvencije u obliku zapisa FASTA analizirane su korištenjem programskog paketa *HMMER* (Anonymous 9, 2010), točnije njegove funkcije *hmmpfam*. Pretraživanje je izvršeno čitavom bazom podataka profila proteina Pfam. Bazi podataka Pfam pristupljeno je putem Interneta (Anonymous 10, 2010). Baza podataka je preuzeta u obliku jedne datoteke sa nekoliko stotina profila proteina HMM poznatih proteinskih domena. Upute za preuzimanje baze nalaze se na Internetskoj stranici:

[ftp://ftp.sanger.ac.uk/pub/databases/Pfam/current\\_release/](ftp://ftp.sanger.ac.uk/pub/databases/Pfam/current_release/) (Anonymous 11, 2010)

Rezultati analize funkcijom *hmmpfam* sastoje se od nekoliko dijelova:

- zaglavlje,
- klasifikacijska lista analiziranih sekvencija,
- detaljan prikaz analiziranih domena redom kako se pojavljuju unutar sekvencije proteina,

- izlazno poravnanje.

Zaglavlje donosi nekoliko osnovnih podataka o upotrijebljenom programu i sekvencijama proteina (engl. "sequence file"). Sekvencije su poredane na temelju vrijednosti E (engl. "E-value", Slika 19, označeno crveno) i uspjehu pogotka (engl. "Score", Slika 19, označeno zeleno). Što je uspjeh pogotka viši, a vrijednost E manja, znači da je pronađen visok stupanj sličnosti između analizirane sekvencije i profila proteina, odnosno jedna ili više domena iz analizirane sekvencije pripadaju domenama iz obitelji proteina. Prikazuje se i broj pojavljivanja "N" (Slika 19, označeno plavo) pojedine domene unutar sekvencije.

```

hmmpfam - search one or more sequences against HMM database
HMMER 2.3.2 (Oct 2003)
Copyright (C) 1992-2003 HHMI/washington university school of Medicine
Freely distributed under the GNU General Public License (GPL)
-----
HMM file:                database\pfam_fs
Sequence file:           sequences\all_NIS_NRPS
-----

Query sequence: NIS
Accession:    [none]
Description:   Type A

Scores for sequence family classification (score includes all domains):
Model          Description                                     Score      E-value     N
-----
IuCA_IuCC      IuCA / IuCC family                                12970.4     0           32
DUF1997        Protein of unknown function (DUF1997)             31.6       9.7e-009    5
MP             Viral movement protein (MP)                       23.5       8.3e-007    6
TraS           Plasmid conjugative transfer entry ex             19.8       1.6e-006    4
TSNR_N         Thiostrepton-resistance methylase, N              20.5       2.8e-005    4
YuaN           YuaN-like family                                  19.6       8.8e-005    4

```

Slika 19. Prikaz rezultata programskog paketa *HMMER*, zaglavlje i klasifikacijska lista analiziranih sekvencija.

Za svaku se domenu može saznati:

- broj domene (Slika 20, označeno crveno) npr. 3/32 za domenu IucA\_IucC znači da je to treća po redu domena od 32 ukupno pronađene,
- početak (seq-f) i kraj (seq-t) domene unutar analizirane sekvencije, izraženo u parovima baza (Slika 20, označeno plavo),
- da li sekvencija pokazuje potpunu ili djelomičnu sličnost, na što ukazuju točkice (poklapanje sekvencija nije u cijelosti) ili uglate zagrade (poklapanje sekvencija je potpuno) (Slika 20, označeno zeleno). Ukoliko se pojave dvije točkice znači da je



poklapanje lokalno, negdje unutar sekvencija. Ukoliko se pojave dvije uglate zagrade poklapanje sekvencija je globalno, odnosno obuhvaćena je čitava sekvencija domene,

- početak, odnosno kraj poklapanja analizirane sekvencije sa sekvencijom u bazi podataka, odnosno modelom, što nam govore vrijednosti *hmm-f* i *hmm-t*, izražene u parovima baza (Slika 20, označeno smeđe).

Parsed for domains:									
Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value		
Yhft	1/1	5	15	1	11	-.	-0.2	9.1	
IGF2_C	1/3	20	27	50	57	.]	2.1	5.2	
DUF2403	1/3	60	70	44	54	..	0.7	23	
Phage_Coat_B	1/4	65	96	1	33	[.	3.2	1.2	
BssC_TutF	1/4	119	127	51	60	.]	0.6	13	
zf-AD	1/4	121	138	58	75	.]	3.3	4.1	
Iuca_Iucc	1/32	151	474	1	350	[.	445.1	1e-130	
CHRD	1/4	157	175	1	21	[.	1.5	9.6	
NOC3p	1/5	159	169	95	105	.]	0.3	20	
MR_MLE	1/6	175	199	58	81	..	0.5	15	
MotA_activ	1/4	222	232	87	97	.]	1.0	9.6	
LMP	1/4	226	237	1	12	[.	0.3	13	
HIPIP	1/3	255	261	68	74	.]	0.6	8.2	
Herpes_UL87	1/3	265	289	152	175	..	0.0	4.7	
Radical_SAM_N	1/5	292	297	349	354	.]	1.5	2.4	
MP	1/6	294	318	182	206	.]	4.4	0.48	
Exonuc_VII_S	1/4	310	326	25	41	..	1.9	12	
GTP_EFTU_D2	1/5	311	321	63	75	.]	0.5	29	
IDEAL	1/4	357	373	1	17	[.	2.4	12	
DUF1116	1/5	386	395	210	219	.]	2.6	2.9	
DUF1997	1/5	392	421	160	189	.]	8.2	0.057	

Slika 20. Prikaz rezultata programskog paketa *HMMER* i analize domena (engl. "domain parse") redom kako se domene pojavljuju unutar sekvencije proteina.

Prvu liniju čini konsenzus model HMM (Slika 21, označeno crveno). Aminokiselina prikazana u konsenzus sekvenciji modela ima najvišu vjerojatnost da će se pojaviti na toj poziciji sudeći prema modelu HMM (nije nužno da je to aminokiselina s najvišim uspjehom pogotka). Velika štampana slova označavaju visoko ušćuvan dio čija je vjerojatnost > 0,5 za modele proteina odnosno > 0,9 za modele DNA. Središnja linija (Slika 21, označeno plavo) pokazuje samo aminokiseline u kojima se sekvencije poklapaju, ili kada pogodak ima pozitivan uspjeh pogotka i stoga se smatra ušćuvanim. Treća linija (Slika 21, označeno zeleno) pokazuje čitavu analiziranu sekvenciju. Izvješće o poravnanju može sadržavati dodatnu liniju "CS" (engl. "consensus structure", Slika 21, označeno smeđe) koja pokazuje da li se radi o strukturi zavojnice (engl. "helix", H), uzvojnice (engl. "coil", C) ili ploče (engl. "sheet", E).

```

Alignments of top-scoring domains:
Yhft: domain 1 of 1, from 5 to 15: score -0.2, E = 9.1
      *->vkiVLiAlGG<-*
      ++i++i+ll+G
NIS   5   IDIIVIILLCG   15

IGF2_C: domain 1 of 3, from 20 to 27: score 2.1, E = 5.2
      *->iTLPseDP<-*
      +tLPse+P
NIS   20  MTLPSEKP    27

DUF2403: domain 1 of 3, from 60 to 70: score 0.7, E = 23
      *->PlnEeLSvHFR<-*
      Pl+E S+HFR
NIS   60  PLDEQKSLHFR   70

Phage_Coat_B: domain 1 of 4, from 65 to 96: score 3.2, E = 1.2
      CS   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
      *->mKAmKqRIAKFSPvASFRnLCiAGsvTAAtSLP<-*
      K + R+A FSP R A vTA S P
NIS   65  -KSLHFRVAYFSPTQHRRFAFPARLVTASGSYP   96

BssC_TutF: domain 1 of 4, from 119 to 127: score 0.6, E = 13
      *->psCeaFktr<-*
      p Ce+F+ +r
NIS   119 PLCETFH-QR   127

zf-AD: domain 1 of 4, from 121 to 138: score 3.3, E = 4.1
      CS   HHHHHHHHHHHHHHHHHHH-
      *->aykFrerCeEsqkkLrel<-*

```

Slika 21. Prikaz rezultata *HMMER* programskog paketa, izlazno poravnanje (engl. "alignment output").

Analizom dobivenih rezultata izdvojene su domene koje su zadovoljavale postavljene uvjete: vrijednost  $E < 10^{-5}$  i uspjeh pogotka  $> 0$ , bez obzira da li je obuhvaćena domena u cijelosti ili samo djelomično. Na taj su način klasifikacijske tablice nadopunjene specifičnim domenama za svaku klasifikacijsku grupu i izrađeni profili proteina koji će kasnije poslužiti pri anotaciji genoma (vidi: podpoglavlje 4.2., Tablica 5; podpoglavlje 8.2.3.).

### 3.2.4. Pronalaženje strukturalnih gena

Sekvencija genoma vrste *Streptomyces scabies* preuzeta je sa "Blast Servera" Sangerova Instituta, tj. sa njihove "FTP" stranice (Anonymous 12, 2010). Programski paket *ClustScan* dostupan je putem Interneta (Anonymous 3, 2010) nakon registracije korisnika.

Nakon učitavanja DNA genoma vrste *S. scabies* u programski paket *ClustScan*

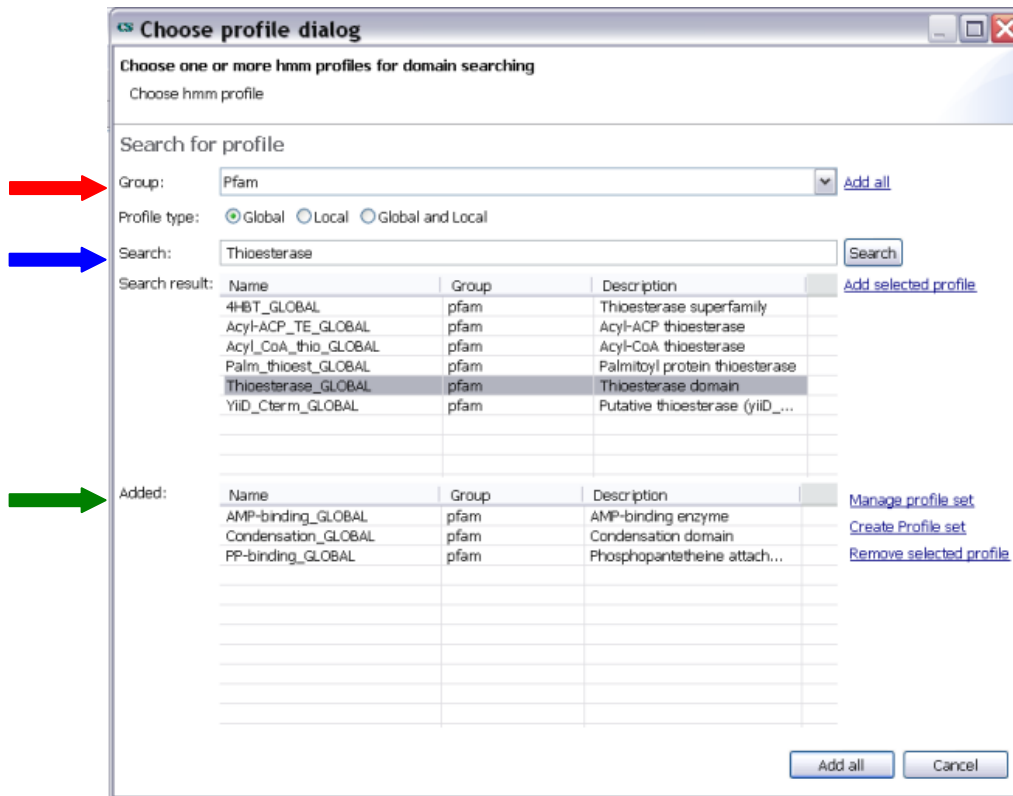
odabirom opcije "*File/Import DNA*", DNA je automatski prevedena u šest otvorenih okvira čitanja pomoću programa *Transeq* (Rice i sur., 2000). Sama analiza sekvencije DNA odvija se na Linux serveru (svaki korisnik ima svoju lozinku (engl. "password"). Lozinka korisniku omogućuje: pristup vlastitoj radnoj površini (engl. "workspace"), učitavanje sekvencije DNA i provođenje analize, dok se sami rezultati pohranjuju na serveru ili na čvrstom disku. Stoga je svaku analizu potrebno provesti samo jednom (to je važno jer pretraživanje jednog genoma, poput vrste *S. scabies*, može trajati nekoliko sati).

Genom je zatim obrađen bioinformatičkim programom *Glimmer*, koji je dio programskog paketa *ClustScan*, odabirom opcije "*Tools/Search for genes with custom model*". Nakon analize rezultata koje je ponudio *Glimmer*, kao model je odabrana bakterija *S. avermitilis* i provedena je analiza programom *GeneMark-PS*. Analiza genoma pomoću programa *GeneMark* pokrenuta je odabirom opcije "*Tools/Search for a genes with a model/Species: Streptomyces avermitilis*".

### **3.2.5. Preuzimanje gotovih profila iz baze podataka Pfam**

Baza profila proteina za pretraživanje genoma vrste *S. scabies* izrađena je upotrebom baze podataka Pfam pomoću programskog paketa *ClustScan* odabirom opcije "*Tools/Search for domains*". Na učitanoj prozoru izabrana je DNA (bakterije *S. scabies*) i opcija "*Add profiles/Search*" (Slika 22).

Za pretraživanje je odabrana baza podataka Pfam (globalni profili proteina, Slika 22 označeno crveno), a pod "search" su upisana imena traženih domena (Slika 22, označeno plavo). Nakon pretraživanja baze podataka Pfam rezultati upita prikazani su pod "Search result". Ako se prikazana domena prihvaća, označava se klikom miša i odabire se opcija "Add selected profile" te automatski prikazuje pod izbornikom "Added" (Slika 22, označeno zeleno). Nakon učitavanja svih domena, odabrana je opcija kreiranja profila proteina (engl. "create profile set"). Profil proteina zatim je učitao odabirom opcije "Manage profile set". Podešeni su parametri programskog paketa *HMMER* (opcija "stringent") i pokrenuto je pretraživanje.



Slika 22. Prozor programskog paketa *ClustScan* za odabir domena i izradu proteinskih profila.

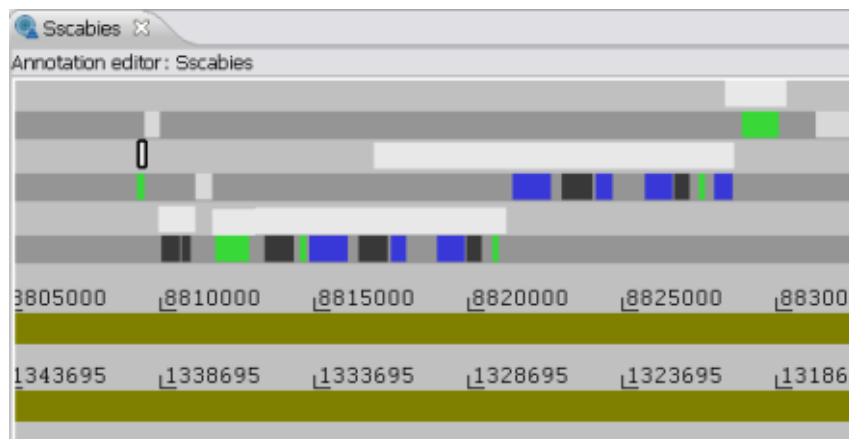
### 3.2.6. Izrada vlastitih profila proteina pomoću programskog paketa *ClustScan*

Odabirom opcije "*Tools/Search for domains/Create new/I want to create an alignment*" moguće je izraditi bazu profila proteina i bez upotrebe baze podataka Pfam. Kako bi se to napravilo, najprije je potrebno prikupiti po nekoliko sekvencija DNA za svaku od domena iz tablica profila proteina koristeći se bazom podataka NRPS-PKS te ih spojiti u jednu datoteku. Tako je za svaku vrstu domena izrađena zasebna datoteka koja je sadržavala nekoliko primjeraka sličnih sekvencija. Sekvencije koje čine jednu datoteku zalijepljene su u za to predviđeni prozor i odabrana je opcija provjere zapisa (sekvencije moraju biti u točnom zapisu FASTA). Temeljni zadatak pri izradi vlastitih profila proteina HMM je odabir opcije "Start Align" čime se pokreće bioinformatički program za višestruko poravnavanje sekvencija DNA ili proteina, *ClustalW* (Anonymous 13, 2009). Nakon poravnanja program sam izrađuje profil proteina i daje mogućnost njegovog imenovanja.

### 3.2.7. Anotacija genoma bakterije *Streptomyces scabies* pomoću programskog paketa *ClustScan*

Nakon što su učitani profili proteina i podešeni parametri programskog paketa *HMMER* pokrenuto je pretraživanje genoma. Programski paket *ClustScan* prikazuje rezultate u obliku:

- liste ili stabla, tj. prozor radne površine (engl. "Workspace window") prikazuje sve pronađene proteine, strukturirane u tri okvira čitanja s lijeva na desno (engl. "forward") i tri okvira čitanja s desna na lijevo (engl. "reverse"). Za svaki otvoreni okvir čitanja prikazan je ukupan broj pronađenih proteinskih domena;
- grafičkom obliku, tj. prozor za uređivanje rezultata anotacije (engl. "Annotation editor") prikazuje sve pronađene gene (Slika 23, bijeli pravokutnici) i proteinske domene (Slika 23, svaka proteinska domena prikazana je drugom bojom pravokutnika).

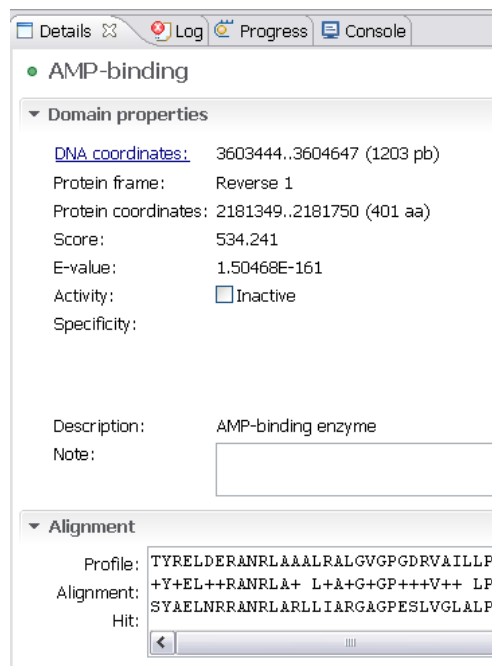


Slika 23. Prozor programskog paketa *ClustScan* za uređivanje rezultata anotacije.

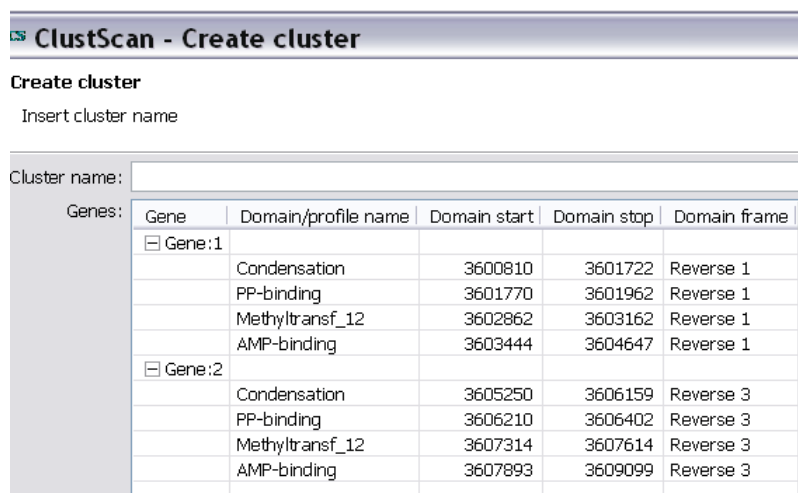
Detaljne informacije (poput koordinata proteinske domene u DNA i proteinu, vrijednosti parametara dobivenih programom *HMMER*, kao i samo poravnanje sekvencija) o svakoj domeni možemo saznati klikom lijeve tipke miša na domenu, čime pokrećemo učitavanje prozora detalja (Slika 24).

Sve domene čiji parametri programskog paketa *HMMER* ne zadovoljavaju postavljane uvjete (vrijednost E, uspjeh pogotka) su izbrisane, a zatim su definirane genske nakupine. Postupak definiranja genske nakupine je vrlo jednostavan. Najprije se označi lijevom tipkom

miša prvi gen, zatim pritisne tipka "ctrl" i označi zadnji gen u genskoj nakupini. Klikom desne tipke miša na zadnji gen pojavljuje se opcija "Create cluster" (Slika 25). U prozoru (Slika 25) su ispisani svi geni obuhvaćeni u gensku nakupinu, proteinske domene unutar njih te položaj unutar sekvencije DNA. Kreiranje genske nakupine završava njezinim imenovanjem (upisom imena u polje "Cluster name").

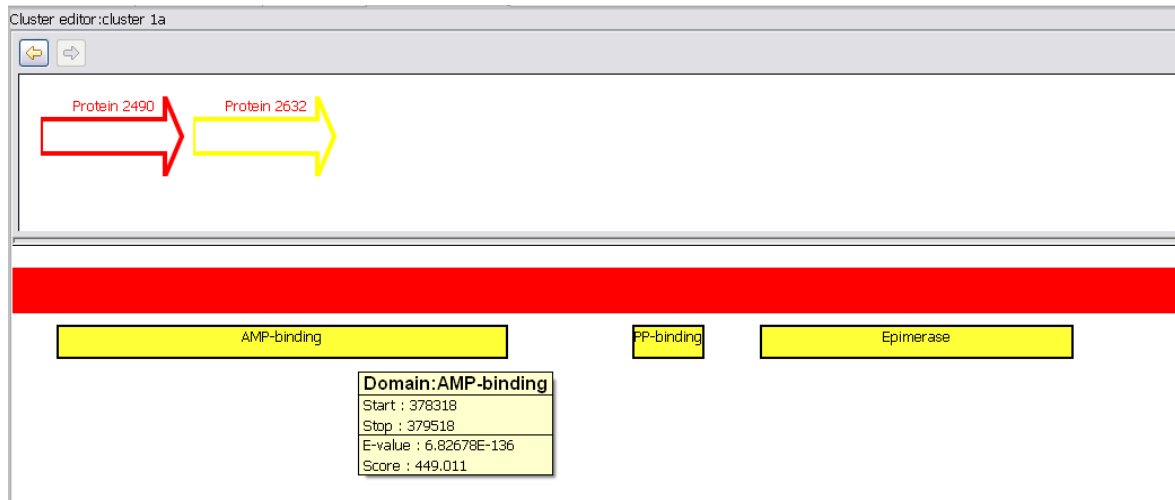


Slika 24. Prozor programskog paketa *ClustScan* s detaljnim informacijama.



Slika 25. Prozor programskog paketa *ClustScan* s detaljnim prikazom gena koji čine jednu gensku nakupinu.

Nakon kreiranja genske nakupine učitava se prozor za uređivanje genskih nakupina "Cluster editor" (Slika 26).



Slika 26. *ClustScan* prozor za uređivanje genskih nakupina.

Svaki gen (Slika 26) prikazan je crvenom strelicom koja ujedno i pokazuje njegov smjer čitanja, odabirom pojedinog gena učitava se dodatni prozor koji prikazuje domene unutar označenog gena. Nakon anotacije genoma bakterije *S. scabies*, "Workspace" sa pronađenim genskim nakupinama sačuvan je na čvrstom disku (vidi: podpoglavlje 8.2.4.).

## **4. REZULTATI**



#### 4.1. REZULTATI ANALIZE LITERATURNIH PODATAKA

Pretraživanjem literature te baza podataka prikupljeno je ukupno 102 sekvencije DNA dobro opisanih sustava NRPS i NRPS neovisnih siderofora. U Tablici 3 prikazani su neki tipični primjeri. Cjelovit popis prikupljenih sekvencija nalazi se u Tablici 3P (vidi: podpoglavlje 8.2.1.) zajedno sa sekvencijama DNA u obliku zapisa FASTA (vidi: podpoglavlje 8.2.2.). Za sustave NRPS prikupljena su 33 primjera, 14 sekvencija DNA linearnih sustava NRPS, 4 sekvencije DNA ponavljajućih sustava NRPS i 15 sekvencija DNA nelinearnih sustava NRPS. Ostatak čini 69 sekvencija sustava NIS, a primjeri su prikazani u Tablici 4. Prikupljeno je 29 sekvencija DNA tipa A, 16 tipa B te 24 tipa C. Cjelovit popis prikupljenih sekvencija DNA nalazi se u Tablici 4P (vidi: podpoglavlje 8.2.1.), sekvencije DNA pohranjene su u obliku zapisa FASTA (vidi: podpoglavlje 8.2.2.).

Tablica 3. Primjer klasifikacije neribosomalno sintetiziranih peptida.

NRPS	SEKUNDARNI METABOLIT	PROIZVODNI MIKROORGANIZAM	GENBANK PRISTUPNI BROJ	BIOLOŠKA AKTIVNOST
<b>Linearne sintetaze</b>				
	SURFAKTIN	<i>Bacillus subtilis</i> ; <i>Staphylococcus aureus</i> subsp. <i>aureus</i> Mu50	NC_000964.2 NC_002758.2	antibiotik
	VANKOMICIN	<i>Amycolatopsis orientalis</i>	AF486630.1	antibiotik
	PIOVERDIN	<i>Pseudomonas fluorescens</i>	U69893.1	siderofor
<b>Ponavljajuće sintetaze</b>				
	GRAMICIDIN S	<i>Brevibacillus brevis</i> <i>Bacillus brevis</i>	BAA00778.1 BAA00777.1 BAA06146.1	antibiotik
	ENNIATIN	<i>Fusarium scirpi</i>	EU128207.1 EU128208.1 ABV54873.1	antibiotik
	BACILLIBAKTIN	<i>Bacillus subtilis</i>	AF184977.1 AY138812.1	siderofor
<b>Nelinearne sintetaze</b>				
	MYXOCHELIN	<i>Stigmatella aurantica</i>	AF299336.1	siderofora
	SYRINGOMYCIN	<i>Pseudomonas syringae</i>	U25130.2	detergent za lizu stanica
	BLEOMICIN	<i>Streptomyces verticillus</i>	AF210249	antitumorski antibiotik

Tablica 4. Primjer klasifikacije NRPS neovisnih siderofora.

NIS	PROIZVODNI MIKROORGANIZAM	GENEBANK PRISTUPNI BROJ	BIOLOŠKA AKTIVNOST
<b>Tip A</b>			
	<i>Escherichia fergusonii</i>	AAL01546.1	IucA protein
	<i>Rhodopseudomonas palustris CGA009</i>	NP_947735.1	biosintetski protein siderofora rizobaktina RhbC
	<i>Sinorhizobium meliloti 1021</i>	NP_436506.1	biosintetski protein rizobaktina RhbC
	<i>Streptomyces avermitilis MA-4680</i>	NP_823642.1	protein transporter željeza
<b>Tip B</b>			
	<i>Vibrio parahaemolyticus</i>	BAC16545.1	biosintetski protein vibrioferina PvsB
	<i>Staphylococcus aureus</i>	AAP82065.1	SbnC protein
	<i>Pectobacterium chrysanthemi</i>	AAL14565.1	AcsA protein
	<i>Rhizobium sp. NGR234</i>	AAQ87282.1	biosintetski protein siderofora
<b>Tip C</b>			
	<i>Shigella boydii</i>	AAK71633.1	IucC protein
	<i>Staphylococcus aureus</i>	AAP82068.1	SbnF protein
	<i>Pectobacterium chrysanthemi</i>	AAL14563.1	AcsC protein
	<i>Sinorhizobium meliloti 1021</i>	NP_436509.1	RhsF biosintetski protein siderofora rizobaktina RhbF

#### 4.2. REZULTATI ANALIZE SEKVENCIJA PROTEINA PROGRAMSKIM PAKETOM HMMER

Tablica profila proteina, Tablica 5, izrađena je na temelju rezultata dobivenih analizom sekvencija proteina pomoću programskog paketa *HMMER*. Potpuni rezultati analize sekvencija proteina programskim paketom *HMMER* nalaze se u prilogima (vidi: podpoglavlje 8.2.3.), djelomični prikaz rezultata prikazan je na Slici 26.

Prilikom analize rezultata u obzir su uzete samo domene čiji su parametri zadovoljavali zadane uvjete: vrijednost  $E < 10^{-5}$  i uspjeh pogotka  $> 0$ . Tablica 5 sadržava specifične domene za svaku klasifikacijsku skupinu, podatke o tome koliko se puta određena sekvencija proteina ponavlja te da li je promatrana domena u dobivenim rezultatima cjelovita ili djelomična. Primjer za više puta ponovljenu domenu je domena C ponovljena 97 puta, što znači da je

programski paket *HMMER* u spojenoj sekvenciji proteina, sastavljenoj od svih prikupljenih sekvencija proteina za linearni sustav NRPS pronašao upravo toliko domena C (Slika 26). Sekvencije proteina ponovljene samo jednom čine jedinstvene domene, kao na primjer NRPS domena kod linearnog sustava NRPS (Slika 26). Cjelovite sekvencije se podudaraju sa profilima u bazi podataka Pfam i obuhvaćaju čitavu duljinu sekvencije. Parcijalne sekvencije su sekvencije koje se djelomično preklapaju sa profilima u bazi podataka Pfam te ne obuhvaćaju cijelu duljinu sekvencije.

```

Query sequence: NRPS
Accession:      [none]
Description:    Type A 90414 bp

Scores for sequence family classification (score includes all domains):
Model          Description                               Score      E-value     N
-----
AMP-binding    AMP-binding enzyme                         35935.6    0           75
Condensation   Condensation domain                       21566.2    0           97
PP-binding     Phosphopantetheine attachment site       5064.0     0           82
Thioesterase   Thioesterase domain                       1030.9     0           10
Methyltransf_12 Methyltransferase domain                  426.6     3.9e-125   10
Methyltransf_11 Methyltransferase domain                  287.5     2.9e-083   13
KU_N           Ku70/Ku80 N-terminal alpha/beta domain   146.6     7.7e-041   2
KU             Ku70/Ku80 beta-barrel domain              135.3     1.9e-037   1
NAD_binding_4 Male sterility protein                     82.2      4.4e-029   1
RHH_1         Ribbon-helix-helix protein, copG fami    100.2     4e-027     20
NRPS          Nonribosomal peptide synthase            70.3      9.9e-027   1
LRV           Leucine rich repeat variant              71.0      3.8e-018   32
DUF937        Bacterial protein of unknown function    54.7      2.9e-017   21
Alpha-amyl_C2 Alpha-amylase C-terminal beta-sheet d    57.7      1.2e-014   19
Abhydrolase_1 alpha/beta hydrolase fold                  51.3      4.3e-014   5
KU_C          Ku70/Ku80 C-terminal arm                  54.7      1.3e-013   2
CesT          Tir chaperone protein (CesT)              49.6      9.2e-013   16
Cobw_C        Cobalamin synthesis protein cobw C-te    47.7      2e-012     21
Strep_SA_rep  Streptococcal surface antigen repeat     41.6      6.9e-011   19
HicB          HicB family                               39.3      2.8e-010   8
WD40          WD domain, G-beta repeat                  31.6      4.1e-010   10
Drf_DAD       DRF Autoregulatory Domain                 38.6      2.4e-009   16
Clp_N         Clp amino terminal domain                 33.6      4.5e-009   14
HD            HD domain                                  36.5      9.3e-009   13
Epimerase     NAD dependent epimerase/dehydratase f    35.5      1.3e-008   2
PAS_4         PAS fold                                   33.5      1.4e-008   10

```

Slika 26. Djelomični prikaz rezultata programskog paketa *HMMER* za linearni sustav NRPS. Enzimski sastav analiziranih sekvencija proteina ukazuje na postojanje jasno definiranih katalitički aktivnih domena.

Iz baze podataka Pfam ekstrahirani su profili proteina za sve katalitički aktivne domene svih opisanih tipova sustava NRPS i NRPS neovisnih siderofora navedenih u Tablici 5.

Tablica 5. Prikaz domena sustava NRPS i NRPS neovisnih siderofora za izradu vlastitih profila proteina (**Napomena:** nazivi domena preuzeti su neposredno iz rezultata programskog paketa *HMMER*).

	VIŠE PUTA PONOVLJENE DOMENE, CIJELA SEKVENCIJA	JEDINSTVENE DOMENE, CIJELA SEKVENCIJA	JEDINSTVENE ILI VIŠESTRUKO PONOVLJENE DOMENE, DJELOMIČNA SEKVENCIJA
<b>NRPS</b>			
<b>Linearne sintetaze</b>	AMP-binding Condensation PP-binding Thioesterase Methyltransf_12 Epimerase	NRPS	
<b>Ponavljajuće sintetaze</b>	AMP-binding Condensation Thioesterase PP-binding		
<b>Nelinearne sintetaze</b>	PP-binding Condensation NRPS AMP-binding NAD_binding_4 ketoacyl-synt Ketoacyl-synt_C KR adh_short Acyl_transf_1 Methyltransf_12 Thioesterase ADH_zinc_N ECH Methyltransf_11 E1_dh	PhyH ADH_N Acyl-CoA_dh_1 3HCDH_N 2-Hacid_dh_C Acyl-CoA_dh_N Amidohydro_2 GFO_IDH_MocA Beta-lactamase Acyl-CoA_dh_M ABC_tran Lactamase_B HMG_CoA_synt_N GlycosTransf	Isochorismatase 3HCDH 2-Hacid_dh Glycos_transf_2 Epimerase Abhydrolase_1 Aminotran_3 p450 ABC_membrane_2 Pyridoxal_deC Amino_oxidase DUF2089 FA_desaturase HMG_CoA_synt_C
<b>NIS</b>			
<b>TIP A</b>	IucA_IucC		
<b>TIP B</b>	IucA_Iuc_C		
<b>TIP C</b>	IucA_IucC AlcB		

#### 4.3. REZULTATI ANOTACIJE GENOMA BAKTERIJE *Streptomyces scabies* POMOĆU PROGRAMSKOG PAKETA *ClustScan*

Obradom genoma bakterije *S. scabies* bioinformatičkim alatom za predviđanje gena *Glimmer*, pronađeno je 9.462 gena, a upotrebom programa *GenMark-PS* 8.446 gena. Korištenjem domena, vlastitih profila proteina, u 6 otvorenih okvira čitanja izdvojen je 321 protein (Tablica 6) kao potencijalan kandidat za promatrane sustave NRPS i NIS.

Tablica 6. Broj proteina pronađenih u 6 okvira čitanja.

TIP OTVORENOG OKVIRA ČITANJA	BROJ ODREĐENIH PROTEINA
S DESNA NA LJEVO 1	60
S DESNA NA LJEVO 2	47
S DESNA NA LJEVO 3	57
S LIJEVA NA DESNO 1	64
S LIJEVA NA DESNO 2	50
S LIJEVA NA DESNO 3	43

Anotirano je ukupno 8 sustava, od toga je:

- 5 pretpostavljenih neribosomalnih peptid sintetaza,
- 3 pretpostavljene sintetaze NRPS neovisnih siderofora.

Detaljan prikaz nalazi se u Tablici 7.

U nastavku je prikazan po jedan primjer od svake skupine anotiranih sustava (Slike 27, 28, 29) dok se potpuni rezultati anotacije genoma bakterije *S. scabies* pomoću programskog paketa *ClustScan* nalaze u priložima (vidi: podpoglavlje 8.2.4.). Na Slikama 27, 28 i 29 bijeli pravokutnici predstavljaju gene unutar kojih su smještene proteinske domene označene različitim bojama, a crvene linije označavaju module. Slika 29, prikazuje pretpostavljenu gensku nakupinu sustava NIS koja sadržava dva gena, svaki sa po jednom NIS siderofor sintetazom (označeno crvenim pravokutnikom).

Tablica 7. Raspodjela genskih nakupina sekundarnih metabolita u genomu *S. scabies*.

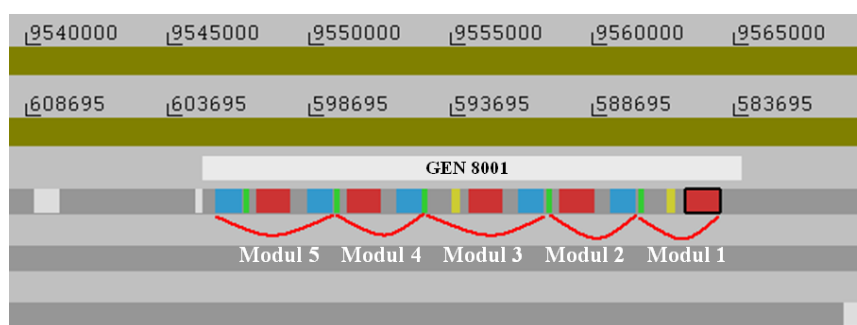
GENSKA NAKUPINA	START	KRAJ	BROJ GENA	DULJINA GENSKE NAKUPINE (BP)
-----------------	-------	------	--------------	------------------------------------

#### NERIBOSOMALNO SINTETIZIRANI PEPTIDI

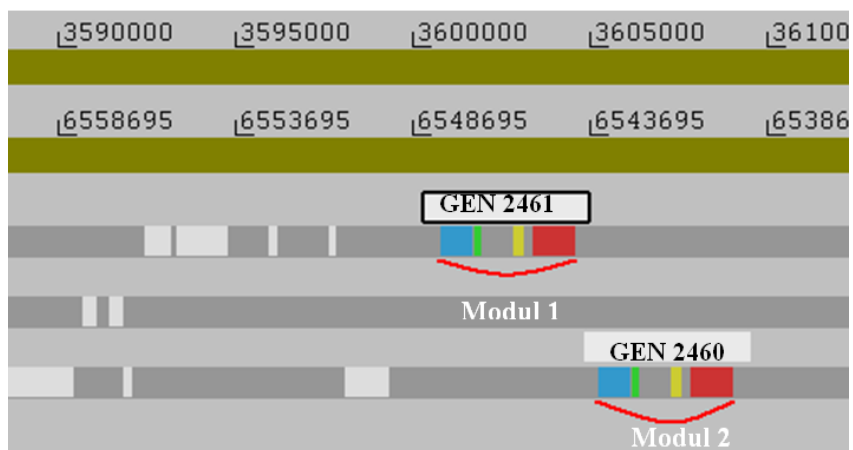
1.	Pretpostavljeni linearni sustav NRPS	364406	381309	2	16903
2.	Pretpostavljeni linearni sustav NRPS	9546250	9565324	1	19074
3.	Pretpostavljeni linearni ili ponavljajući sustav NRPS	3600343	3609567	2	9224
4.	Pretpostavljeni linearni ili ponavljajući sustav NRPS	158013	168113	2	10100
5.	Pretpostavljeni nelinearni sustav NRPS	9477547	9488003	5	10456

#### NRPS NEOVISNI SIDEROFORI (NIS)

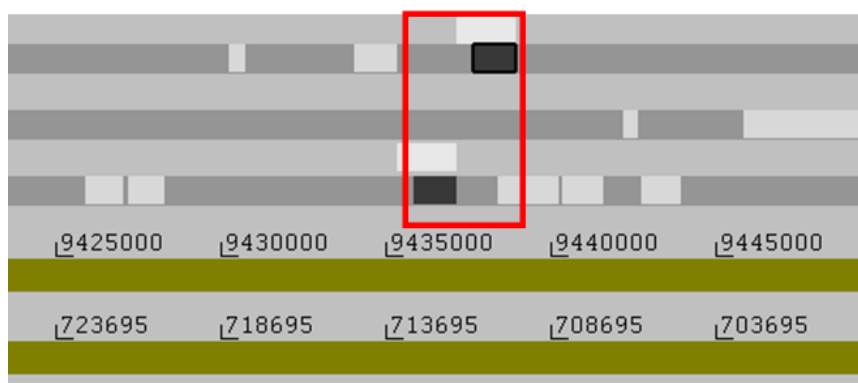
6.	Pretpostavljena sintetaza NIS	6445872	6448193	2	2321
7.	Pretpostavljena sintetaza NIS	9435375	9438981	2	3606
8.	Pretpostavljena sintetaza NIS	2071419	2074787	2	3368



Slika 27. Primjer anotirane genske nakupine (br. 2 iz Tablice 7).



Slika 28. Primjer anotirane genske nakupine (br. 3 iz Tablice 7).



Slika 29. Primjer anotirane genske nakupine (br. 7 iz Tablice 7).

U nastavku se nalaze shematski prikazi (Tablica 8) koji detaljnije opisuju pretpostavljene genske nakupine analiziranih sekundarnih metabolita. Domene unutar istog modula označene su istom bojom, a moduli koji se nalaze na jednome genu, nalaze se u istom retku pod zajedničkim brojem gena. Potrebno je istaknuti da su u shematskom prikazu navedeni samo geni koji sadržavaju module i domene, dok "prazni" geni bez modula i domena, koji su također u nekim slučajevima sastavni dio genskih nakupina, nisu posebno navedeni. Problematična domena obojana je žutom bojom, a potencijalni uzrok uočenog problema navodi se u nastavku (vidi: podpoglavlje 5.3.) Ispod shematskog prikaza pronađenih genskih nakupina nalazi se legenda sa kraticama naziva domena (Tablica 9).

Tablica 8. Shematski prikaz pronađenih genskih nakupina.

---

**Genska nakupina br. 1: Pretpostavljeni NRPS tip A**

GEN 2490: -PP(PCP)-

GEN 2632: -C-A-PCP-C- -C-PCP- -C-A-PCP- -C-A-PCP- -C-A-PCP-ER-

---

**Genska nakupina br. 2: Pretpostavljeni NRPS tip A**

GEN 8001: -A-M12-PCP- -C-A-PCP- -C-A-M12-PCP- -C-A-PCP- -C-A-PCP-C-

---

**Genska nakupina br. 3: Pretpostavljeni NRPS tip A ili B**

GEN 2461: -A-M12-PCP-C-

GEN 2460: -A-M11-PCP-C-

---

**Genska nakupina br. 4: Pretpostavljeni NRPS tip A ili B**

GEN 719: -PCP-C-NRPS-A-M12-PCP-

GEN 675: -C-NRPS-A-M12-PCP-

---

**Genska nakupina br. 5: Pretpostavljeni NRPS tip C**

GEN 7933: -A-

GEN 7934: -M12-

GEN 7937: -A-

GEN 7938: -Abh1-

GEN 7939: -A-

---

**Genska nakupina br. 6: Pretpostavljena sintetaza NIS**

GEN 5117: -AlcB-

GEN 5116: -IucA\_IucC-

---

**Genska nakupina br. 7: Pretpostavljena sintetaza NIS**

GEN 7897: -IucA\_IucC-

GEN 7898: -IucA\_IucC-

---

**Genska nakupina br. 8: Pretpostavljena sintetaza NIS**

GEN 1105: -IucA\_IucC-

GEN 1106: -IucA\_IucC-

---



Tablica 9. Legenda kratika domena za Tablicu 8.

A = AMP-binding	Abh1 = Abhydrolase_1
C = Condensation	M11 = Methyltransf_11
PCP = PP-binding	M12 = Methyltransf_12
NRPS = NRPS	AlcB = AlcB
	IucA_IucC = IucA_IucC

## **5.RASPRAVA**

## 5.1. PRETRAŽIVANJE LITERATURNIH PODATAKA

Pretraživanjem literaturnih podataka pronađeno je vrlo malo znanstvenih radova u kojima se detaljnije obrađuje tematika klasifikacije sustava NRPS i NIS. To potvrđuje da su ovi biosintetski mehanizmi nedovoljno istraženi (Hranueli i sur., 2008). Dostupna literatura uglavnom obrađuje metode izolacije sekundarnog metabolita, iz hranjive podloge na kojoj raste mikroorganizam producent ili iz stanice producenta, te identifikaciju genske nakupine koja kodira za enzim. Vrlo rijetko se detaljno opisuju mehanizmi sinteze navedenih sekundarnih metabolita. Literaturni podaci često nude po nekoliko mikroorganizama producenata za isti sekundarni metabolit, međutim za ovaj rad su prikupljeni samo pojedinačni primjeri, oslanjajući se na svojstvo sličnosti.

Hibridni sustavi NRPS/PKS u literaturnim podacima svrstavaju se u nelinearne sustave NRPS (Challis, 2005) ili u zasebnu skupinu (Wenzel i Müller, 2005). Zbog nedovoljno literaturnih podataka, precizne klasifikaciji i kompleksnosti njihove građe, ti sustavi nisu obuhvaćeni u ovom radu. Klasifikacija sustava NIS do sada se temeljila na svojstvu sličnosti sekvencija NRPS neovisnih sintetaza, koje su dio genskih nakupina takvih sustava. Međutim ovaj oblik klasifikacije nije dostatan za klasifikaciju samih NRPS neovisnih siderofora jer genska nakupina koja nosi informaciju za sintezu NRPS neovisnog siderofora može sadržavati i više od jedne NRPS neovisne sintetaze. Stoga se smatra da je potrebna daljnja analiza ovih sustava (Challis, 2005).

Prilikom prikupljanja sekvencija DNA i proteina sustava NRPS neovisnih siderofora nisu prikupljene cijele genske nakupine već samo izdvojene NRPS sintetaze (vidi: Tablica 4 i 4P). Posljedica je to nedovoljne istraženosti genskih nakupina, pa baze podataka ne sadrže sekvencije cijele genske nakupine NRPS neovisnih siderofora već samo do sada izolirane i proučavane sekvencije NRPS neovisnih sintetaza. Prilikom prikupljanja sekvencija iz baza podataka, pretragom pomoću pristupnog broja, uočeno je da za neke upite postoji samo jedna sekvencija DNA i jedna sekvencija proteina. Međutim, pronađeni su i slučajevi kada se unatoč jednoj sekvenciji DNA, koja pokriva čitavu gensku nakupinu, pojavljuje više proteinskih sekvencija kojima se pristupa preko različitih pristupnih brojeva (na primjer, vankomicin, GenBank pristupni broj DNA sekvencije: AF486630.1) (Anonymous 1, 2010). To je posljedica različitog pristupa sekvencioniranju, tj. u prvom slučaju je sekvencionirana cijela genska nakupina u komadu, a u drugom slučaju gen po gen.

## 5.2. OBRADA REZULTATA ANALIZE SEKVENCIJA PROTEINA PROGRAMSKIM PAKETOM *HMMER*

Iz rezultata dobivenih programskim paketom *HMMER* (Eddy, 1998) izdvajane su domene čije su vrijednosti parametara zadovoljavale uvjete: vrijednost  $E < 10^{-5}$  i uspjeh pogotka  $> 0$ . Te vrijednosti odabrane su kao granične zbog pretpostavke da domene čija je vrijednost  $E$  viša od  $10^{-5}$  i čiji je pogodak blizu nule ili negativan ne zadovoljavaju traženo svojstvo sličnosti. Izdvajane su sve domene koje su zadovoljavale postavljene uvjete parametara bez obzira da li je domena cjelovita ili djelomična, zbog opravdane bojazni da bi isključivanje jedne vrste domena moglo dovesti do gubitka dragocjenih informacija. Ovisno o tome da li su domene cjelovite ili djelomične, te da li se ponavljaju više puta ili su jedinstvene, u tablici profila proteina (vidi: Tablica 5) smještene su u odvojene kolone. Višestruko su se ponavljale domene koje čine okosnicu modula sustava NRPS. Jedinstvene domene su specifične za pojedini tip sustava NRPS. Pojavljivanje djelomičnih domena u rezultatima dobivenim programskim paketom *HMMER* moguće je objasniti na dva načina:

- prilikom prikupljanja sekvencija nije obuhvaćena čitava genska nakupinu; ili
- analizirana sekvencija i profili iz Pfam baze podataka pokazuju djelomičnu sličnost.

Analizom tablice profila (vidi: Tablica 5), uočene su domene svojstvene pojedinim tipovima sustava NRPS, npr. domena Methyltransf\_12 specifična je za linearni sustav NRPS, dok su domene PP-binding, Condensation, Thioesterase i AMP-binding zajedničke svim sustavima NRPS (nazivi domena preuzeti su neposredno iz programskog paketa *HMMER* pa su zbog toga napisani engleskim jezikom). U slučajevima kada su se domene preklapale izdvojena je domena koja je imala bolje parametre. Domena Abhydrolase\_1, prema rezultatima dobivenim programskim paketom *HMMER* (vrijednost  $E$ :  $4,3 \times 10^{-014}$ , uspjeh pogotka: 52.3), klasificirana je kao jedinstvena domena nelinearnih sustava NRPS (vidi: Tablica 5). Ista domena zapažena je u ponavljajućih sustava NRPS (vrijednost  $E$ :  $9,1 \times 10^{-005}$ , uspjeh pogotka: 18.1) i linearnih sustava NRPS (vrijednost  $E$ :  $2,2 \times 10^{-008}$ , uspjeh pogotka: 30.9), međutim zbog slabijih vrijednosti parametara odbačene su.

Analizom rezultata programskog paketa *HMMER* za NRPS neovisne sintetaze primijećeno je da se IucA i IucC sintetaze svrstavaju u istu proteinsku obitelj unutar baze podataka Pfam, odnosno izgubile su specifičnost. Parametri programskog paketa *HMMER* za tip A (vrijednost  $E$ : 0, uspjeh pogotka: 12970.4) i tip C (vrijednost  $E$ : 0, uspjeh pogotka:

12498.3) pokazuju visok stupanj sličnosti sa profilom u bazi podataka Pfam. Stoga nije bilo moguće izdvojiti posebno domene za tip A i tip C NRPS neovisnih sintetaza. Tip C sintetaza, međutim, kao što su pokazali i rezultati, pokazuje visok stupanj sličnosti sa sekvencijom (domenom) AlcB. Provjerom u bazi podataka Pfam pronađeno je da i AlcB pripada obitelji proteinskih domena koje sudjeluje u biosintezi siderofora (alcaligin u vrste *Bordetella* i aerobactin u vrste *E. coli*). Prema literaturnim podacima (Challis, 2005) AlcB je acil-CoA-transferaza, no ipak pokazuje dobre vrijednosti parametara programskog paketa *HMMER* (vrijednost E:  $2.2 \times 10^{-014}$ , uspjeh pogotka: 57.0), odnosno sličnost sa NRPS neovisnim sintetazama tipa C.

NRPS neovisnih sintetaza tipa B nije još dovoljno istražen pa se pretpostavlja da programski paket *HMMER* neće pronaći slične sekvencije unutar baze podataka Pfam. Međutim, rezultati programskog paketa *HMMER* pokazali su da sekvencije NRPS neovisnih sintetaza tipa B pokazuju određen stupanj sličnosti sa IucA\_IucC proteinskom obitelji. To je vidljivo iz vrijednosti parametara (vrijednost E: 0, uspjeh pogotka: 5426.5). No vrijednosti parametara za tip A i tip C NRPS neovisnih siderofora pokazuju jače podudaranje.

### **5.3. OBRADA REZULTATA ANOTACIJA GENOMA BAKTERIJE *Streptomyces scabies* POMOĆU PROGRAMSKOG PAKETA *ClustScan***

Prilikom odabira modela za određivanje gena u genomu *Streptomyces scabies* izabran je model vrste *S. avermitilis*. Iz literaturnih podataka (Ōmura i sur., 2001; Challis i Hopwood, 2003) preuzet je broj gena genoma vrste *S. avermitilis* (8.446 gena; Anonymous 14, 2010) i vrste *S. coelicolor* (8.350 gena; Anonymous 15, 2010). Broj gena u genomu vrste *S. scabies* određen je pomoću bioinformatičkog alata *Glimmer* (Delcher i sur., 2007). Iznosio je 9.462 gena. Model vrste *S. avermitilis* sličniji je po broju gena, stoga je odabran kao model za analizu genoma *S. scabies* pomoću programa *GeneMark* (Besemer i Borodovsky, 2005).

Analizom rezultata uočeni su slučajevi u kojima su domene bile izvan gena. Pretpostavlja se da je to posljedica pomaka okvira čitanja što je najvjerojatnije pogreška nastala prilikom sekvencioniranja. Pošto programski paket *ClustScan* (Starcevic i sur., 2008) omogućuje prilagođavanje koordinata, u nekoliko slučajeva geni su pomaknuti iz jednog okvira čitanja u drugi, kako bi se domene smjestile unutar gena. Razlog tome jest

nemogućnost programskog paketa *ClustScan* da domenu koja nije unutar obuhvati u gensku nakupinu. Prilikom anotiranja uočeno je nekoliko problema koji su onemogućili preciznu klasifikaciju svakog tipa genske nakupine. Na primjer, u genskoj nakupini br. 8, domena IucA\_IucC iako se nalazi unutar gena 1106, nakon kreiranja genske nakupine nije prikazana u prozoru za uređivanje genskih nakupina programskog paketa *ClustScan*. Pretpostavlja se da gen djelomično pokriva domenu pa ona nije prikazana kao njegov sastavni dio.

U mnogim genskim nakupinama uočena je neuobičajena organizacija domena unutar modula. Primjerice, u genskoj nakupini br. 1, u genu 2632 domena C je smještena na kraju modula, dok u slijedećem modulu domena A nedostaje, što je neobično za linearne sustave NRPS. Ovaj neobični smještaj domena C može biti posljedica delecije dijelova modula ili uzastopnih povezivanja više djelomičnih modula (Challis, 2005). Genske nakupine br. 3 i br. 4 zbog nedostatka preciznih rezultata nisu smještene niti u linearne sustave NRPS niti u ponavljajuće sustave NRPS. Obje genske nakupine sadrže po dva modula što je premalo za linearne sustave NRPS, a zbog nepoznavanja kemijske strukture produkta nemoguće je sa sigurnošću tvrditi da se radi o ponavljajućem sustavu NRPS (Schwarzer i Marahiel, 2001). Analizom genske nakupine br. 5 uočeno je da se sastoji od 5 gena, svaki sa po jednom domenom, a zbog specifične domene Abh1 (vidi: Tablica 5), smještene u četvrtom genu, genska nakupina klasificirana je kao nelinearni sustav NRPS.

Analizom genske nakupine br. 8 radi potvrđivanja rezultata, analizirali smo sekvenciju gena 1105 (koji sadržava IucA\_IucC domenu) pomoću programa *BLAST* (Altschul i sur., 1990) tražeći sekvenciju koja će pokazivati značajniji stupanj sličnosti. Takva sekvencija nije pronađena iako su rezultati programskog paketa HMMER, vrijednost E i uspjeh pogotka, bili visoki. Slijedeći gen uključen u ovu gensku nakupinu, gen 1106, sadržava IucA\_IucC domenu, ali ona nije prikazana u prozoru za uređivanje genskih nakupina nakon kreiranja genske nakupine. Pretpostavlja se da gen djelomično pokriva domenu pa ju ne prikazuje kao njegov sastavni dio. Međutim, unatoč tome program *BLAST* (Altschul i sur., 1990) pokazao je da sekvencija pokazuje značajnu sličnost sa genom čiji produkt sudjeluje u biosintezi siderofora. Time su dodatno potvrđeni rezultati, a i pretpostavka da su "prazni geni", koji su također dio genskih nakupina vjerojatno uključeni u određene poslije sintetske modifikacije ili sadržavaju neke nove i dosada nepoznate funkcionalnosti (Donadio i sur., 2007).

#### **5.4. PROŠIRENJE FUNKCIONALNOSTI PROGRAMSKOG PAKETA *ClustScan***

Iz rezultata pretraživanja genoma vrste *S. scabies* pomoću vlastitih profila proteina koji su učitani u programski paket *ClustScan* može se sa sigurnošću klasificirati genske nakupine do određene razine. Do sada je programski paket *ClustScan* omogućavao da se sa visokom sigurnošću ustanovi da li je genska nakupina sustav NRPS, PKS ili NIS, a izradom klasifikacijskih profila proteina (Tablica 7) učitanih u programski paket *ClustScan* omogućena je detaljnija analiza genskih nakupina. Međutim, zbog velikog odstupanja u organizaciji domena unutar modula i zbog nedostatka literaturnih podataka, koje bi eksperimentalno potvrdile mehanizme sinteze, vrlo je teško precizno odrediti kojoj klasifikacijskoj podgrupi pronađene genske nakupine pripadaju.

## **6. ZAKLJUČCI**



Iz dobivenih rezultata i provedene rasprave mogu se izvesti slijedeći zaključci:

1. Na temelju prikupljene i analizirane literature izrađena je detaljna klasifikacija sustava NRPS i NIS, upotpunjena profilima proteina pripadajućih domena.
2. Razdvajanjem obitelji proteina IucA\_IucC u zasebne obitelji, riješio bi se problem gubitka specifičnosti ovih profila proteina i omogućilo točnije razumijevanje mehanizama sinteze NRPS neovisnih siderofora.
3. Programskim paketom *ClustScan* u genomu bakterije *S. scabies* opisano je 8 genskih nakupina, i to:
  - pet genskih nakupina koje sintetiziraju neribosomalno sintetizirane peptide
  - tri genske nakupine koje sintetiziraju NRPS neovisne siderofore.
4. Daljnja istraživanja genskih nakupina, koje sadržavaju genetičku uputu za sustave NRPS, osigurala bi precizniju anotaciju programskim paketom *ClustScan* i na razini podgrupa, a ne samo glavnih tipova.

## **7. POPIS LITERATURE**

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403-410.
2. Anand, S., Prasad, M.V., Yadav, G., Kumar, N., Shehara, J., Ansari, Z., Mahanty, D. (2010) SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res.* **38**, Suppl:W487-496.
3. Anonymous 1 (2010) National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/genbank/>. Pristupljeno 15.05.2010.
4. Anonymous 2 (2010) European Bioinformatics Institute – European Molecular Biology Laboratory EMBL, <http://www.ebi.ac.uk/embl/>. Pristupljeno 30.04.2010.
5. Anonymous 3 (2004) A knowledge based resource for analysis of Non-ribosomal Peptide Synthetases and Polyketide Synthetases, <http://www.nii.res.in/nrps-pks.html>. Pristupljeno 15.05.2010.
6. Anonymous 4 (2009) Thiotemplate Modular Systems Studies, <http://bioser.pbf.hr/cms/index.php?page=clustscan>. Pristupljeno 20.04.2010.
7. Anonymous 5 (2010) The Wellcome Trust Sanger Institute 1: Genome project of *S. scabies*, [http://www.sanger.ac.uk/Projects/S\\_scabies/](http://www.sanger.ac.uk/Projects/S_scabies/). Pristupljeno 20.05.2010.
8. Anonymous 6 (2010) Google Scholar, <http://scholar.google.hr/>. Pristupljeno 27.04.2010.
9. Anonymous 7 (2010) Science direct, <http://www.sciencedirect.com/>. Pristupljeno 28.04.2010.
10. Anonymous 8 (2010) European Bioinformatics Institute – Readseq, <http://www.ebi.ac.uk/cgi-bin/readseq.cgi>. Pristupljeno 10.05.2010.
11. Anonymous 9 (2010) HMMER – biosequence analysis using profile hidden Markov models, <http://hmmer.janelia.org/>. Pristupljeno 18.05.2010.
12. Anonymous 10 (2010) The Wellcome Trust Sanger Institute 2: Pfam, <http://www.sanger.ac.uk/Pfam/>. Pristupljeno 15.05.2010.
13. Anonymous 11 (2010) The Wellcome Trust Sanger Institute 3: <http://www.sanger.ac.uk/>. Pristupljeno 10.05.2010.
14. Anonymous 11 (2010) The Wellcome Trust Sanger Institute 4: Pfam/current\_release [ftp://ftp.sanger.ac.uk/pub/databases/Pfam/current\\_release/](ftp://ftp.sanger.ac.uk/pub/databases/Pfam/current_release/). Pristupljeno 15.05.2010.
15. Anonymous 12 (2009) Clustal: Multiple Sequence Alignment, <http://www.clustal.org/>. Pristupljeno 18.05.2010.
16. Anonymous 13 (2010) Genome project of *Streptomyces avermitilis*, <http://avermitilis.ls.kitasato-u.ac.jp/>. Pristupljeno 28.04.2010.

17. Anonymous 14 (2010) The Wellcome Trust Sanger Institute 5: Genome project of *Streptomyces coelicolor*, [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/). Pristupljeno 14. 05.2010.
18. Ansari, M.Z., Yadav, G., Gokhale, R.S., Mohanty, D. (2004) NRPS-PKS: a Knowledge-based Resource of NRPS/PKS Megasyntases. *Nucleic Acids Res.* **32**, 405-413.
19. Barona-Gomez, F., Lautru, S., Francou, F.X., Leblond, P., Pernodet, J.J., Challis, G.L. (2006) Multiple biosynthetic and uptake system mediate siderophore – dependent iron acquisition in *Streptomyces coelicolor* A3(2) and *Streptomyces amboficus* ATCC 23877. *Microbiolog.* **152**, 3355-3366.
20. Besemer, J., Borodovsky, M. (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* **33**, 451-454.
21. Caboche, S., Pupin, M., Leclère, V., Fontaine, A., Jacques, P., Kucherov, G. (2008) NORINE: a database of nonribosomal peptides. *Nucleic Acids Res.* **36**, 326-331.
22. Challis, L.G. (2005) A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *ChemBioChem*, **6**, 601-611.
23. Challis, L.G., Hopwood, D.A. (2003) Synergy and Contingency as Driving Forces for the Evolution of Multiple Secondary Metabolite Production by *Streptomyces* Species. *Proc. Natl. Acad. Sci. USA*, **100**, 14555-14561.
24. Challis, L.G., Naismith, J.H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr. Opin. Chem. Biol.* **14**, 748-756.
25. Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, **23**, 673-679.
26. Demain, A. (1999) Pharmaceutically active secondary metabolites of microorganisms. *Appl. Microbiol. Biotechnol.* **52**, 455-463.
27. Donadio, S., Monciardini, P., Sosio M. (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat. Prod. Rep.* **24**, 1073-1109.
28. Eddy, S.R. (1998) Profile Hidden Markov models. *Bioinformatics*. **14**, 755-763.
29. Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K. (2008) The Pfam protein families database. *Nucleic Acids Res.* **36**, 281-288.
30. Fischbach, M.A., Walsh, C.T., Clardy, J. (2008) The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. USA*, **105**, 4601-4608.

31. Grabley, S., Thiericke, R. (1999) The impact of natural products on drug discovery. U: Drug Discovery from Nature (Grabley, S., Thiericke, R., ured.), Springer, Berlin, str. 3-37.
32. Hranueli, D., Cullum, J. (2001) Novi hibridni poliketidi dobiveni kombinatornom biosintezom. *Kem. Ind.* **50**, 381-411.
33. Hranuleli, D., Cullum, J. (2003) Bioinformatics of *Streptomyces* species and food production. U: Current Studies of Biotechnology - Vol. III. Food. Croatian Society of Biotechnology (Kniewald, Z. *i sur.* ured.), Zagreb, Croatia, str. 333-340.
34. Hranueli, D., Starčević, A., Žučko, J., Diminić, J., Škunca, N., Željeznak, V., Kovaček, D., Pavlinušić, D., Šimunković, J., Long, P.,F., Cullum, J. (2008) Oblikovanje novih prirodnih spojeva u uvjetima *in silico*. *Kem. Ind.* **57**, 245–256.
35. Lambert, D.H., Loria, R. (1989) *Streptomyces scabies* sp. nov., norn. rev. *Int. J. Syst. Bacteriol.* **39**, 387-392.
36. Li, M.H., Ung, P.M., Zajkowski, J., Garneau-Tsodikova, S., Sherman, D.H. (2009) Automated genome mining for natural products. *BMC Bioinformatics*, **10**, 185.
37. Mootz, H.D., Schwarzer, D., Marahiel, M.A. (2002) Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. *ChemBioChem*, **3**, 490-504.
38. Ōmura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M. (2001) Genome Sequence an Industrial Microorganism *Streptomyces avermitilis*: *Streptomyces avermitilis*: Deducing the Ability of Producing Secondary Metabolites. *Proc. Natl. Acad. Sci. USA*, **98**, 12215-12220.
39. Rausch, C., Weber T., Kohlbacher, O., Wohlleben W., Huson, D.H. (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799-5808.
40. Reva, O., Tümmler, B. (2008) Think big – giant genes in bacteria. *Environ. Microbiol.* **10**, 768-777.
41. Rice, P., Longden, I., Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276-277.
42. Sattely, E.S., Fischbach, M.A., Walsh, C.T. (2008) Total biosynthesis: *in vitro* reconstitution of polyketide and nonribosomal peptide pathways. *Nat. Prod. Rep.* **25**, 757-793.
43. Schwarzer, D., Marahiel, M.A. (2001) Multimodular biocatalysts for natural product assembly. *Naturwissenschaften*, **88**, 93-101.

44. Starčević, A., Žučko, J., Šimunković, J., Long, P.F., Cullum, J., Hranueli, D. (2008) *ClustScan*: an integrated program package for semi-automatic annotation of modular biosynthetic gene cluster and *in silico* prediction of novel chemical structures. *Nucleic Acids Res.* **10**, 6882-6892.
45. Tae, H., Kong, E.B., Park, K. (2007) ASMPKS: an analysis system for modular polyketide synthases. *BMC Bioinformatics*, **8**, 327.
46. Weissman, K.J. (2004) Polyketide biosynthesis: understanding and exploiting modularity. *Philos. Transact. A Math. Phys. Eng. Sci.* **362**, 2671-2690.
47. Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D.H., Wohlleben, W. (2008) CLUSEAN: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Biotechnology*, **140**, 13-17.
48. Wenzel S.C., Müller, R. (2005) Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. *Curr. Opin. Chem. Biol.* **9**, 447-458.
49. Yadav, G., Gokhele, R.S., Mohanty, D. (2003) SEARCHPKS: a program for detection and analysis of polyketide synthase domain. *Nucleic Acids Res.* **31**, 3654-3658.
50. Zazopoulous, E., Haung, K., Staffa, A., Liu, W., Bachmann, B.O., Nonaka, K., Ahlert, J., Thorson, J.S., Shen, B., Farnat, C.M. (2003) A genomics – guided approach for discovering and expressing cryptic metabolic pathways. *Nat. Biotechnol.* **21**, 187-190.
51. Zotchev, S.B., Stepanichikova, A.V., Sergeyko, A.P., Sobolev, B.N., Filimonov, D.A., Poroikov, V.V. (2006) Rational design of macrolides by virtual screening of combinatorial libraries generated through *in silico* manipulation of polyketide synthases. *J. Med. Chem.* **49**, 2077-2087.

## **8. PRILOZI**

## 8.1. Popis u radu upotrijebljenih kratica

A	- domena za adenilaciju aminokiselina (NRPS)
ACP	- mali polipeptid nosač acila (PKS)
AT	- domena aciltransferaze (PKS)
bp	- parovi baza
C	- domena za kondenzaciju (NRPS)
Cy	- domena za formiranje heterocikličkih prstenova (NRPS)
DH	- domena dehidrataze (PKS)
DNA	- deoksiribonukleinska kiselina
E	- očekivana vrijednost
Ep	- domena za epimerizaciju (NRPS)
ER	- domena enoilreduktaze (PKS)
KR	- domena ketoreduktaze (PKS)
KS	- domena ketosintaze (PKS)
MT	- domena za <i>N</i> -, <i>C</i> - ili <i>O</i> -metilaciju (NRPS)
NIS	- sintetaze NRPS-neovisnih siderofora
NRPS	- sintetaza neribosomalno sintetiziranih peptida
Ox	- domena oksidacije (NRPS)
PCP	- mali polipeptid nosač peptidila (NRPS)
PKS	- poliketid sintaza
PKS/NRPS	- mješovita poliketid sintaza i sintetaza neribosomalno sintetiziranih peptida
Te	- domena tioesteraze (NRPS)



## **8.2. Sadržaj kompaktnog diska**

Svi navedeni prilozi, uključujući i cjelovit tekst Diplomskog rada (Diplomski rad DR.pdf), nalaze se na kompaktnom disku (CD-R) nazvanom "Diplomski rad DR: Prilozi".

**8.2.1. Sekvencije DNA cjelovitih genskih nakupina, te sekvencije DNA i proteina pojedinačnih gena, neribosomalno sintetiziranih peptida i NRPS neovisnih siderofora u obliku zapisa FASTA**

**8.2.2. Sekvencije proteina iz priloga 8.2.1., u obliku zapisa FASTA, pripremljene za višestruko poravnavanje sekvencija pomoću programskog paketa *HMMER***

**8.2.3. Rezultati višestrukih poravnanja sekvencija, iz priloga 8.2.2., pomoću programskog paketa *HMMER***

**8.2.4. Rezultat anotacije genoma bakterije *Streptomyces scabies* pomoću programskog paketa *ClustScan***

Priložena je radna površina programskog paketa *ClustScan* (*S. scabies* rezultat anotacije\_Dina). Napomena: za njeno je pregledavanje potreban klijent programskog paketa na osobnom računalu. Kabinet za bioinformatiku će na zahtjev instalirati programski paket *ClustScan* na računalu korisnika.