

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA
ENOTA MEDODDELČNEGA ŠTUDIJA MIKROBIOLOGIJE

Tomaž RAKUŠA

**DOLOČITEV ODPRTIH OKVIRJEV GENSKE
SKUPINE PRI BAKTERIJI *Amycolatopsis* sp. ZA
BIOSINTEZO TETRACIKLINSKEGA ANTIBIOTIKA**

DIPLOMSKO DELO

Univerzitetni študij

Ljubljana, 2010

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA
ENOTA MEDODDELČNEGA ŠTUDIJA MIKROBIOLOGIJE

Tomaž RAKUŠA

**DOLOČITEV ODPRTIH OKVIRJEV GENSKE SKUPINE PRI
BAKTERIJI *Amycolatopsis* sp. ZA BIOSINTEZO
TETRACIKLINSKEGA ANTIBIOTIKA**

DIPLOMSKO DELO

Univerzitetni študij

**DETERMINATION OF OPEN READING FRAMES IN BACTERIA
Amycolatopsis sp. FOR BIOSYNTHESIS OF TETRACYCLINE
ANTIBIOTIC**

GRADUATION THESIS

University studies

Ljubljana, 2010

Diplomsko delo je zaključek univerzitetnega medoddelčnega študija mikrobiologije na Biotehniški fakulteti Univerze v Ljubljani. Opravljeno je bilo v laboratoriju Katedre za biotehnologijo, mikrobiologijo in varnost živil Oddelka za živilstvo Biotehniške fakultete Univerze v Ljubljani.

Po sklepu Študijske komisije univerzitetnega dodiplomskega študija mikrobiologije ter na osnovi Pravilnika o diplomskem delu je bil za mentorja diplomskega dela imenovan doc. dr. Hrvoje Petković, za somentorja prof. dr. Peter Raspor, za recenzenta pa prof. dr. Gregor Anderluh.

Mentor: doc. dr. Hrvoje Petković

Somentor: prof. dr. Peter Raspor

Recenzent: prof. dr. Gregor Anderluh

Predsednica študijske komisije: prof. dr. Darja Žgur Bertok

Komisija za oceno in zagovor:

Predsednik: prof. dr. Darja Žgur Bertok
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Član: doc. dr. Hrvoje Petković
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za živilstvo

Član: prof. dr. Peter Raspor
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za živilstvo

Član: prof. dr. Gregor Anderluh
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Datum zagovora:

Diplomska naloga je del večjega projekta pri katerem sem sodeloval v okviru programa doktorandke Urške Lešnik. Od nje sem dobil nukleotidni zaporedji v FASTA formatu, ki ju je pridobila tekom projekta, kateri sem združil in analiziral. Vsi opisani rezultati v tem delu so rezultat lastnega dela.

Tomaž Rakuša

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

- ŠD Dn
DK UDK 601.4:575.116.4:579.873.7:604.4:615.2/.3(043)=163.6
KG *Amycolatopsis*/kelokardin/antibiotiki/tetraciklin/poliketid sintaze/PKS tipa II/ *in silico* analize
AV RAKUŠA, Tomaž
SA PETKOVIĆ, Hrvoje (mentor)/RASPOR, Peter (somentor)/
ANDERLUH, Gregor (recenzent)
KZ SI-1000 Ljubljana, Jamnikarjeva 101
ZA Univerza v Ljubljani, Biotehniška fakulteta, Enota medoddelčnega študija mikrobiologije
LI 2010
IN DOLOČITEV ODPRTIH OKVIRJEV GENSKE SKUPINE PRI BAKTERIJI *Amycolatopsis sp.* ZA BIOSINTEZO TETRACIKLINSKEGA ANTIBIOTIKA
TD Diplomsko delo (univerzitetni študij)
OP XV, 62 str., 21 pregl., 47 sl., 94 vir.
IJ SI
JI sl/en
AI V diplomskem delu je predstavljena analiza nukleotidnega zaporedja dobljenega iz *Amycolatopsis sulphurea in silico*, za katerega smo predvidevali, da kodira vse potrebne gene za biosintezo kelokardina. S programom FramePlot smo poiskali potencialne odprte bralne okvirje (ORF). Izmed 38 potencialnih ORF smo za 18 ORF predvideli sodelovanje v biosintezi molekule kelokardina. Te smo nadalje analizirali s programi, dostopnimi na spletni strani NCBI in komercialnim programom Vector NTI in jim določili najbližje homologe, motive in ohranjene domene ter jim s pomočjo podatkov teh analiz in literature poskusili določiti vlogo v biosintezi kelokardina. V skupini genov smo našli vse gene, ki smo jih glede na zgradbo molekule kelokardina tudi pričakovali. Med geni tako najdemo gene za t.i. minimalni PKS (poliketid sintaze). To je minimalni nabor genov potrebnih za sintezo poliketidov, ki ga sestavljajo ketosintaza α (KS α), ketosintaza β (KS β) in acilprenašalni protein (angl. ACP), ciklaze/aromataze, oksigenaze, metiltransferazi, ketoreduktazo, aminotransferazo in acil-CoA ligazo. Poleg teh najdemo tudi gen za protein, ki izloča molekulo kelokardina iz celice in gen za predvideni transkripcijski regulator. V zaporedju sta kodirana tudi gena za transpozazo in glikoziltransferazo, ki pa ju nismo pričakovali in jima ne pripisujemo sodelovanja pri biosintezi kelokardina. Skupino genov za biosintezo kelokardina glede na primerjavo z ostalimi skupinami genov uvrščamo med PKS tipa II.

KEY WORDS DOCUMENTATION

DN Dn
DC UDC 601.4:575.116-4:579.873.7:604.4:615.2/.3(043)=163.6
CX *Amycolatopsis*/chelocardin/antibiotics/tetracycline/polyketide synthases/PKS type II
/in silico analysis
AU RAKUŠA, Tomaž
AA PETKOVIĆ Hrvoje (supervisor)/RASPOR, Peter (co-advisor)/
ANDERLUH, Gregor (reviewer)
PP SI-1000 Ljubljana, Jamnikarjeva 101
PB University of Ljubljana, Biotechnical Faculty, Interdepartmental Programme in
Microbiology
PY 2010
TI DETERMINATION OF OPEN READING FRAMES IN BACTERIA
Amycolatopsis spp. FOR BIOSYNTHESIS OF TETRACYCLINE ANTIBIOTIC
DT Graduation thesis (University studies)
NO XV, 62 p., 21 tab., 47 fig., 94 ref.
LA Sl
AL sl/en
AB In this work we describe *in silico* analysis of nucleotide sequence from
Amcolatopsis sulphurea, for which we assumed to have all necessary genes for
chelocardin biosynthesis. We used FramePlot program to identify the potential
ORFs. Out of 38 potential ORFs we identified 18 could participate in the
chelocardin biosynthesis. Then we worked with free on-line NCBI programs as well
with commercial Vector NTI program to search for the closest homologs, motifs and
conserved domains. We found all of the genes that we expected according to
molecular structure of chelocardin. PKS cluster has all the genes for “minimal PKS
(polyketide synthases)”. Those are ketosynthase α (KS α), ketosynthase β (KS β) in
acyl-carrier protein (ACP). There are also cyclase/aromatases, oxygenases,
methyltransferases, ketoreductase, aminotransferase and acyl-CoA ligase. In addition
we also found a gene which most likely encodes a protein responsible for secretion
of chelocardin out of the cell and a putative transcription regulator. The sequence
also contains genes for a transposase and glycosyltransferase to which we don't
ascribe any participation in the chelocardin biosynthesis. Genes associated in the
chelocardin biosynthesis belong to the Type-II PKS family.

KAZALO VSEBINE

1	UVOD.....	1
1.1	CILJ DIPLOMSKE NALOGE	1
1.2	DELOVNA HIPOTEZA	1
2	PREGLED OBJAV	2
2.1	AKTINOMICETE	2
2.1.1	Morfologija, biologija, ekologija, fiziologija in metabolizem aktinomicet	2
2.1.2	Klasifikacija rodu <i>Amycolatopsis</i> in umestitev <i>Amycolatopsis sulphurea</i>	2
2.1.3	Aktinomicete kot producenti sekundarnih metabolitov	3
2.2	SEKUNDARNI METABOLITI.....	3
2.3	BIOSINTEZA POLIKETIDOV	4
2.3.1	Biosinteza poliketidov	4
2.3.2	Poliketid sintaze	5
2.3.2.1	Poliketid sintaze tipa I	5
2.3.2.2	Poliketid sintaze tipa II	6
2.3.2.3	Poliketid sintaze tipa III.....	7
2.3.2.4	Drugi tipi poliketidnih sintaz.....	7
2.4	TETRACIKLINI	8
2.4.1	Odkritje in spekter delovanja.....	8
2.4.2	Kemijske lastnosti tetraciklinov	9

2.4.3	Mehanizem delovanja tetraciklinskih antibiotikov	9
2.4.4	Biosinteza oksitetraciklina pri <i>Streptomyces rimosus</i>	10
2.5	KELOKARDIN	12
2.6	BIOINFORMATIKA	13
2.7	ODPRTI BRALNI OKVIR.....	13
2.8	PODATKOVNE ZBIRKE.....	13
2.8.1	Sistem Entrez	14
2.8.2	Baze proteinskih zaporedij	14
2.8.2.1	UniProt – baza proteinskih zaporedij	15
2.8.3	Sekundarne baze podatkov.....	15
2.9	DOLOČANJE VERJETNIH BRALNIH OKVIRJEV V BAKTERIJSKIH GENOMIH Z VIŠJO VSEBNOSTJO G+C.....	15
2.9.1	FramePlot – program za iskanje verjetnih odprtih bralnih okvirjev ..	16
2.10	METODE ZA PRIMERJAVO ZAPOREDIJ	16
2.10.1	Iskanje homolognih zaporedij v podatkovnih bazah	17
2.10.1.1	Program BLAST.....	17
2.10.2	Primerjava več zaporedij.....	18
2.10.2.1	Program CLUSTAL	18
3	MATERIALI IN METODE.....	20
3.1	SHEMATSKI PRIKAZ POTEKA DIPLOMSKEGA DELA.....	21
3.2	ZDRUŽEVANJE SOSESK	22

3.3	DOLOČANJE VERJETNIH ODPRTIH BRALNIH OKVIRJEV S PROGRAMOM FRAMEPLOT	22
3.4	BLAST IN PSI BLAST – ISKANJE PODOBNIH ZAPOREDIJ	23
3.4.1	Iskanje ohranjenih domen	25
3.5	PORAVNAVA VEČ ZAPOREDIJ	25
3.6	SHEMATSKI PRIKAZ GENSKE SKUPINE	26
4	REZULTATI.....	27
4.1	OPISI GENOV	27
4.1.1	ChdP	27
4.1.2	ChdK.....	28
4.1.2.1	<i>Stopnja podobnosti med ChdP (KSα) in ChdK (KSβ)</i>	29
4.1.3	ChdS.....	30
4.1.4	ChdQI	31
4.1.5	ChdMII.....	33
4.1.6	ChdGIV	34
4.1.7	ChdTn.....	35
4.1.8	ChdR.....	36
4.1.9	ChdA.....	37
4.1.10	ChdN.....	38
4.1.11	ChdOI.....	39
4.1.12	ChdMI	40

4.1.13	ChdOIII	41
4.1.14	ChdL	42
4.1.15	ChdX	43
4.1.16	ChdT	45
4.1.17	ChdQII	46
4.1.18	ChdOII	47
4.2	SHEMA SKUPINE GENOV ZA BIOSINTEZO KELOKARDINA	49
5	RAZPRAVA IN SKLEPI	51
5.1	SKLEPI	52
6	POVZETEK	54
7	VIRI	55
ZAHVALA		

KAZALO PREGLEDNIC

Preglednica 1: Taksonomska opredelitev <i>Amycolatopsis sulphurea</i> (Boone in sod., 2001).	3
Preglednica 2: Seznam najbližjih homologov genskega produkta ChdP, pridobljen z analizo BLASTP.....	28
Preglednica 3: Seznam najbližjih homologov genskega produkta ChdK, pridobljen z analizo BLASTP.....	29
Preglednica 4: Seznam najbližjih homologov genskega produkta ChdS, pridobljen z analizo BLASTP.....	30
Preglednica 5: Seznam najbližjih homologov genskega produkta ChdQI, pridobljen z analizo BLASTP.....	31
Preglednica 6: Seznam najbližjih homologov genskega produkta ChdMII, pridobljen z analizo BLASTP.....	33
Preglednica 7: Seznam najbližjih homologov genskega produkta <i>ChdGIV</i> , pridobljen z analizo BLASTP.....	34
Preglednica 8: Seznam najbližjih homologov genskega produkta ChdTn, pridobljen z analizo BLASTP.....	35
Preglednica 9: Seznam najbližjih homologov genskega produkta ChdR, pridobljen z analizo BLASTP.....	36
Preglednica 10: Seznam najbližjih homologov genskega produkta ChdA, pridobljen z analizo BLASTP.....	37
Preglednica 11: Seznam najbližjih homologov genskega produkta <i>ChdN</i> , pridobljen z analizo BLASTP.....	38
Preglednica 12: Seznam najbližjih homologov genskega produkta ChdOI, pridobljen z analizo BLASTP.....	39

Preglednica 13: Seznam najbližjih homologov genskega produkta ChdMII, pridobljen z analizo BLASTP.....	40
Preglednica 14: Seznam najbližjih homologov genskega produkta ChdOIII, pridobljen z analizo BLASTP.....	41
Preglednica 15: Seznam najbližjih homologov genskega produkta ChdL, pridobljen z analizo BLASTP.....	42
Preglednica 16: Seznam najbližjih homologov genskega produkta ChdX, pridobljen z analizo BLASTP.....	43
Preglednica 17: Seznam najbližjih homologov genskega produkta ChdT, pridobljen z analizo BLASTP.....	45
Preglednica 18: Seznam najbližjih homologov genskega produkta ChdQII, pridobljen z analizo BLASTP.....	46
Preglednica 19: Seznam najbližjih homologov genskega produkta ChdOII, pridobljen z analizo BLASTP.....	47
Preglednica 20: Preglednica z osnovnimi podatki za vsak predviden gen v skupini genov za biosintezo kelokardina.....	50

KAZALO SLIK

Slika 1: Biosintezna pot maščobnih kislin in poliketidov (Hopwood, 1997).....	5
Slika 2: Shematski prikaz biosinteze eritromicina (Khosla in sod., 2007).....	6
Slika 3: Primer PKS tipa II (Shen, 2003)	7
Slika 4: PKS tipa III (Shen, 2003).....	7
Slika 5: Primeri tetraciklinov iz različnih generacij (Zakeri in Wright, 2008).....	8
Slika 6: Kemijska struktura različnih tetraciklinskih antibiotikov	9
Slika 7: Shema skupine genov za biosintezo oksitetraciklina	10
Slika 8: Slika prikazuje reakcije, ki jih katalizirajo encimi minimalnega PKS in vodijo do nastanka poliketidne verige (Zhang in sod., 2006).....	11
Slika 9: Shematski prikaz možnih stopenj pri biosintezni poti oksitetraciklina (Zhang in sod., 2006)	12
Slika 10: Molekula kelokardina. Oštevilčena je po sistemu IUPAC.....	12
Slika 11: Shematski prikaz poteka diplomskega dela	21
Slika 12: Primer rezultata, ki ga prikaže program FramePlot	23
Slika 13: Opis predvidenega genskega produkta ChdP.....	27
Slika 14: Opis predvidenega genskega produkta ChdK	28
Slika 15: Poravnava najbolj homolognih zaporedij s ChdP	29
Slika 16: Ohranjen glutamin (Q ¹⁸⁰), ki je značilen za KSβ (Bisang in sod., 1999).....	30
Slika 17: Opis predvidenega genskega produkta ChdS.....	30
Slika 18: V super-družino fosfopantein vezajočih proteinov spada tudi družina AcpP.....	31

Slika 19: Opis predvidenega genskega produkta ChdQI.....	31
Slika 20: Primerjava ohranjenih domen proteinov CmmQI (AAO65366), OtcD1 (AAD10031) in ChdQI.....	32
Slika 21: Opis predvidenega genskega produkta ChdMII.....	33
Slika 22: Opis predvidenega genskega produkta ChdGIV.....	34
Slika 23: Opis predvidenega genskega produkta ChdTn	35
Slika 24: Ohranjena domena znotraj ChdTn, ki spada med domene DDE transpozaz (pfam01609 (NCBI, 2009c)).	36
Slika 25: Opis predvidenega genskega produkta ChdR	36
Slika 26 Opis predvidenega genskega produkta ChdA	37
Slika 27: Domeni, značilni za TetR regulatorne proteine.	38
Slika 28: Opis predvidenega genskega produkta ChdN	38
Slika 29: Opis predvidenega genskega produkta ChdOI.....	39
Slika 30: Opis predvidenega genskega produkta ChdMI	40
Slika 31: Ohranjena domena (pfam00891), ki jo prikaže BLASTP, je značilna za O-metiltrasferaze.	41
Slika 32: Opis predvidenega genskega produkta ChdOIII	41
Slika 33: Zaporedje ChdOIII z označeno značilno ohranjeno domeno ABM (pfam03992 (NCBI, 2009c)).....	42
Slika 34: Opis predvidenega genskega produkta ChdL	42
Slika 35: Grafični prikaz za nekaj zadetkov BLASTP analize zaporedja ChdL.	43
Slika 36: Opis predvidenega genskega produkta.....	43

Slika 37: grafični prikaz ujemanja z <i>ChdX</i> , kot ga prikaže BLASTP analiza	44
Slika 38: Grafični prikaz ohranjenih domen prikaže tudi domeno, značilno za beta podenoto obroč hidroksilirajoče dioksigenaze (cd00667 (NCBI, 2009c)).....	44
Slika 39: Opis predvidenega genskega produkta ChdT	45
Slika 40: Opis predvidenega genskega produkta ChdQII	46
Slika 41: Analiza BLASTP proteinskega zaporedja ChdQII.	46
Slika 42: Opis predvidenega genskega produkta ChdOII	47
Slika 43: Grafični prikaz ujemanja <i>ChdOII</i> z najbolj podobnimi proteini	47
Slika 44: Shematski prikaz predvidene skupine genov za biosintezo kelokardina.	49

OKRAJŠAVE IN SIMBOLI

Acil-CoA	acilkoencim A
ACP	acilprenašalni protein (angl. Acyl Carrier Protein)
ak	aminokislina
ARO	aromataza
AT	aciltransferaza
BLAST	algoritem za primerjavo zaporedij in iskanje optimalnih lokalnih poravnav (angl. Basic Local Alignment Search Tool)
CTC	klortetraciklin
DDBJ	Japonska baza DNA zaporedij (angl. DNA Database of Japan)
DH	dehidrataza
DNA	deoksiribonukleinska kislina
EBI	Evropski inštitut za bioinformatiko (angl. European Bioinformatics Institute)
EMBL	Zveza evropskih laboratorijev molekularne biologije (angl. European Molecular Biology Laboratory)
ER	enoilreduktaza
FAD	flavin adenin dinukleotid
FAS	sintaza maščobnih kislin (angl. Fatty Acid Synthase)
INSDC	Podatkovna zbirka nukleotidnih zaporedij mednarodnega združenja (angl. International Nucleotide Sequence Database Collaboration)
InterPro	Zbirka podatkov o proteinskih domenah in funkcionalnih mestih (angl. Integrated Resource of Protein domains and functional sites)
kb	kilobaza

KS	ketosintaza
NCBI	Nacionalni center za biotehnološko informacijo (angl. National Center for Biotechnology Information)
nt	nukleotid
ORF	odprti bralni okvir (angl. Open Reading Frame)
OTC	oksitetraciklin
PCR	verižna reakcija s polimerazo (angl. Polymerase Chain Reaction)
PDB	Podatkovna zbirka struktur bioloških makromolekul (angl. Protein Data Bank)
Pfam	Zbirka podatkov o proteinskih družinah
PIR	Zbirka podatkov o proteinih (angl. Protein Information Resource)
PKS	poliketid sintaza
PSI BLAST	položajno specifičen algoritem BLAST s ponavljanjem (angl. Position Specific Iteration BLAST)
PRINTS	Zbirka ohranjenih proteinskih motivov, s pomočjo katerih določamo proteinske družine
ProDom	Celovit nabor proteinskih domenskih družin, ki so avtomatsko anotirane iz baze UniProtKB/Swiss-Prot
PROSITE	Zbirka podatkov o proteinskih domenah, družinah in funkcionalnih skupinah, kot tudi vzorcev in profilov za njihovo identifikacijo
SAM	S-adenozil-L-metionin
TE	tioesteraza
UniProtKB/Swiss-Prot	Podatkovna zbirka dobro anotiranih proteinskih zaporedij
UniProtKB/TrEMBL	Zbirka podatkov, ki vsebuje računalniško anotirana zaporedja iz zbirke vnesenih zaporedij v EMBL, ki še niso del UniProtKB/Swiss-Prot

1 UVOD

Tetraciklini so antibiotiki širokega spektra, ki so zaradi učinkovitega protibakterijskega delovanja in šibkih stranskih učinkov zelo uporabljani predvsem v medicini in veterini, večinoma za zdravljenje raznih infekcij. Množična uporaba tetraciklinov je najverjetneje privedla do pojava bakterijske rezistence, zaradi česar zdravljenje velikokrat ni več učinkovito (Chopra in Roberts, 2001). Pojav rezistence na pogosto uporabljane tetracikline vodi raziskovalce k proučevanju drugih tetraciklinskih molekul s protibakterijskim delovanjem, na katere bakterije še niso odporne in bi zato lahko nadomestili obstoječe, medicinsko pomembne tetraciklinske antibiotike.

Skupine genov za biosintezo tetraciklinov so dovzetne za manipulacije z orodji molekularne biologije (Staunton in Weisman, 2001). S spreminjanjem genov za biosintezo lahko vplivamo na spremembo produkta ali donosa, s tem da lahko npr. odstranimo ali dodamo določeno stransko skupino na molekuli ali povečamo produkcijo le-te. Pri tem si lahko v veliki meri pomagamo z bioinformatičnimi orodji, s katerimi lahko med drugim analiziramo neko zaporedje, predvidimo proteine, ki jih kodira in napovemo njihovo vlogo. Tako nam bioinformatika služi kot zanesljiva opora pri načrtovanju predvidenih manipulacij genskih skupin, za spremljanje pridobljenih rezultatov in načrtovanju nadaljnega dela. V diplomskem delu pridobljeni rezultati bodo zato v veliko pomoč pri nadaljnjem raziskovalnem delu glede biosinteze tetraciklinskega antibiotika kelokardina in pripravi novih analogov kelokardina.

1.1 CILJ DIPLOMSKE NALOGE

Poglavitni namen diplomskega dela je določitev predvidenih odprtih bralnih okvirjev v genomskem nukleotidnem zaporedju iz bakterije *Amycolatopsis sulphurea*. Za analizirano zaporedje namreč predvidevamo, da kodira vse potrebne gene za biosintezo tetraciklinskega antibiotika kelokardina.

1.2 DELOVNA HIPOTEZA

Glede na literaturne podatke lahko pričakujemo v skupini genov za biosintezo kelokardina naslednje:

- v skupini genov bodo zapisani vsi geni, ki sodelujejo pri biosintezi kelokardina,
- v skupini genov bodo zagotovo zapisani geni za t.i. minimalni PKS (minimalni nabor genov, ki sintetizirajo poliketid). To so ketosintaza α ($KS\alpha$), ketosintaza β ($KS\beta$) in acilprenašalni protein (ACP), ki katalizirajo sintezo osnovne poliketidne verige,
- skupino genov bodo sestavljali tudi encimi, ki delujejo na osnovno poliketidno verigo in skrbijo za nastanek osnovnega tetraciklinskega skeleta. Med temi pričakujemo vsaj eno oksigenazo, vsaj eno ciklazo/aromatazo, vsaj eno ketoreduktazo, dve metiltransferazi in eno aminotransferazo,
- skupina genov bo verjetno vsebovala tudi proteine za rezistenco proti delovanju kelokardina, s čimer se bakterija zavaruje pred delovanjem molekule, in regulatorje, ki uravnavajo biosintezo kelokardina.

2 PREGLED OBJAV

V tem poglavju se bomo najprej dotaknili aktinomicet, med katere spada tudi *Amycolatopsis sulphurea*, ki jo bomo tudi podrobneje opisali in taksonomsko opredelili. V nadaljevanju se bomo osredotočili na sekundarne metabolite s poudarkom na poliketidih, njihovi biosintezi, tipih poliketid sintaz in njihovih glavnih značilnosti. Ugotovili bomo, da uvrščamo kelokardin med tetracikline in to med PKS tipa II, zato bomo kot primer uporabili biosintezo oksitetraciklina, ki je v več pogledih podobna biosintezi kelokardina. V drugem delu pregleda objav bomo predstavili bioinformatična orodja, ki jih uporabljamo za iskanje predvidenih odprtih bralnih okvirjev in določitev predvidenih funkcij genskih produktov, ki jih le-ti kodirajo.

2.1 AKTINOMICETE

Aktinomicete spadajo v deblo *Actinobacteria*, red *Actinomycetales*. Aktinobakterije so druga največja skupina po Gramu pozitivnih bakterij. Sestavlja jo več kot 30 različnih taksonomskih družin. Večinoma so primarni aerobi, po obliki so paličasti ali razvejani (Madigan in sod., 2009).

2.1.1 Morfologija, biologija, ekologija, fiziologija in metabolizem aktinomicet

Aktinomicete so velika skupina po Gramu pozitivnih bakterij, ki navadno oblikujejo razvejane filamente. Posledica njihove uspešne rasti in razvejitve je mreža prepletenih filamentov, ki jo imenujemo micelij. Za aktinomicete je značilno, da so v večini sporogene. Morfološke lastnosti in proces nastanka spor se pri aktinomicetah med vrstami razlikujejo in spadajo med pomembne klasifikacijske lastnosti. Pomembna značilnost aktinomicet je tudi visoka vsebnost G+C, ki se v povprečju nahaja med 63% in 78%. Bakterije, ki imajo najvišje razmerje G+C med aktinomicetami, imajo tudi najvišje razmerje med vsemi bakterijami, saj je vsebnost G+C aktinomicet najvišja med vsemi bakterijami (Madigan in sod., 2009).

Aktinomicete so najpogosteje saprofitski organizmi, najdeni v zemlji, kjer pretvarjajo kompleksne biopolimere kot so lignoceluloza, hemiceluloza, pektin ipd. Najdemo pa jih tudi v drugih habitatih, npr. rastlinskih delih in vodi. Proizvajajo veliko pomembnih biološko aktivnih substanc. Iz aktinomicet je bilo izoliranih okoli 61% vseh znanih bioaktivnih snovi. Velika večina bioaktivnih snovi v uporabi danes je iz streptomiset. Zaradi tega dejstva imajo aktinomicete zelo pomembno vlogo v industriji (Vijayakumar, 2007).

2.1.2 Klasifikacija rodu *Amycolatopsis* in umestitev *Amycolatopsis sulphurea*

Bakterije iz rodu *Amycolatopsis* so do 1964 leta pripadale rodu *Nocardia*, takrat pa je Lechevalier s sodelavci zaradi novih dokazov, med drugim odzivom na fage zaradi odsotnosti mikolične kisline v celični steni, nekatere vrste ločil iz tega rodu in jih uvrstil v dva nova rodova, t.j. *Amycolata* in *Amycolatopsis* (Lechevalier in sod., 1986).

Bakterije iz rodu *Amycolatopsis* so aerobne, negibljive in se ne barvajo po Ziehl-Neelsenu. Bakterije iz tega rodu tvorijo razvejan substratni micelij, ki je razdeljen na kvadrataste ali

elipsoidne elemente. Zračne hife so lahko sterilne, ali pa se diferencirajo v dolge verige sporam-podobnih struktur. V celični steni bakterije najdemo mezo-diaminopimelično kislino, galaktozo in arabinozo in so bogate z izo- in anteizo-razvejanimi maščobnimi kisljinami. Vsebnost G+C nukleotidnih parov se giblje med 66-73% (Tseng in sod., 2006; Gibson in sod., 2003; Holt, 1994).

Preglednica 1: Taksonomska opredelitev *Amycolatopsis sulphurea* (Boone in sod., 2001)

domena:	<i>Bacteria</i>
deblo:	<i>Actinobacteria</i>
razred:	<i>Actinobacteria</i>
podrazred:	<i>Actinobacteridae</i>
red:	<i>Actinomycetales</i>
podred:	<i>Pseudonocardineae</i>
družina:	<i>Pseudonocardiaceae</i>
rod:	<i>Amycolatopsis</i>
vrsta:	<i>sulphurea</i>

Rod *Amycolatopsis* pripada družini *Pseudonocardiaceae*, za katere je značilno, da v celični steni ne vsebujejo mikolične kisline, ki je značilna za družino *Nocardiaceae* (Majumdar in sod., 2006).

Nekatere bakterije iz rodu *Amycolatopsis* so komercialno pomembne, saj proizvajajo bioaktivne snovi, kot so antibiotiki (npr. vankomicin in rifamicin), ali razgrajujejo širok nabor aromatičnih spojin (Majumdar in sod., 2006).

2.1.3 Aktinomicete kot producenti sekundarnih metabolitov

Odkritji aktinomicina (Waksman in Woodruff, 1940) in streptomocina leta 1943 (Kingston, 2004) sta prvi nakazali veliko pomembnost aktinomicet kot proizvajalcev antibiotikov in drugih sekundarnih metabolitov z biološko aktivnostjo. Čas od njunega odkritja pa vse do konca 80-tih let pogosto imenujemo tudi zlata doba odkrivanja novih antibiotikov. Ocena števila odkritih antibiotikov med različnimi avtorji variira (Challis in Hopwood, 2003), v splošnem pa jih dve tretjini pripisujejo bakterijam iz redu *Actinomycetales*, kar pomeni okoli 8000 aktivnih substanc (Weissman in Müller, 2009) od tega jih po ocenah 70-80% proizvedejo streptomicete, manjši del pa prispevajo še rodovi *Saccharopolyspora*, *Amycolatopsis*, *Micromonospora* in *Actinoplanes* (Challis in Hopwood, 2003).

2.2 SEKUNDARNI METABOLITI

Poznamo dva tipa mikrobnega metabolizma. Primarni metabolizem je metabolizem v času eksponentne rasti celice, medtem ko se sekundarni metabolizem začne proti koncu (eksponentne) faze rasti in traja precej vzporedno s stacionarno fazo. Sekundarni metaboliti so med najbolj kompleksnimi in pomembnimi metaboliti, ki se uporabljajo v medicinske in ostale industrijske namene (Madigan in sod., 2009).

Sekundarni metaboliti neposredno niso nujni za rast celice. Za proizvodnjo sekundarnih metabolitov bakterije uporabljajo predvsem spojine, ki izhajajo iz primarnega

metabolizma. Sekundarne metabolite pogosto proizvajajo sporogeni organizmi v času sporulacije (Madigan in sod., 2009).

Obseg njihovega biološkega delovanja je širok in vključuje tako inhibicijo rasti drugih mikroorganizmov, kot tudi toksične efekte proti mnogoceličnim organizmom, kot so rastline in nevretenčarji. Znana je tudi njihova vloga hormonom podobnih molekul v celični delitvi, pa tudi vloga v transportu kovin. Poleg teh vlog imajo nekateri sekundarni metaboliti tudi funkcije, ki jih danes verjetno še ne znamo razložiti (Challis in Hopwood, 2003).

2.3 BIOSINTEZA POLIKETIDOV

Poliketidi so ena največjih skupin spojin med sekundarnimi metaboliti. Raznolikost med poliketidi lahko dobro ilustrira razmerje med številom C-atomov med 6-metilsalicilno kislino (6-MSA), ki je z 6 C-atommi eden manjših poliketidov in poliketidom maitotoksinom, ki je sestavljen iz več kot 150 C-atomov (Hranueli in Cullum, 2001). Poliketidi posedujejo različne biološke aktivnosti, kot so protibakterijska, protiglivna, protiparazitska, imunosupresivna in antitumorska aktivnost. Nekateri poliketidi so znani tudi kot pigmenti. (Pfeifer in Khosla, 2001). So tudi bogat vir klinično uporabnih farmacevtikov (Komaki in Harayama, 2006). Komerzialno pomembne poliketide uporabljamo predvsem v medicini in veterini. Eritromicin, doksirubicin, FK506 se tako uporabljajo v medicinske namene, tetraciklini in avermektin se uporabljajo v medicini in v veterini, tilozin, monezin in narazin pa se uporabljajo le v veterini (Hershberger, 1996).

Glede na strukturo poliketide delimo v štiri skupine: aromatske poliketide (npr. tetraciklin in aktinorodin), makrolide (npr. eritromicin in rafamicin), polietre (npr. monezin in salinomycin) in poliene (npr. amfotericin in kandicin) (Shen, 2000).

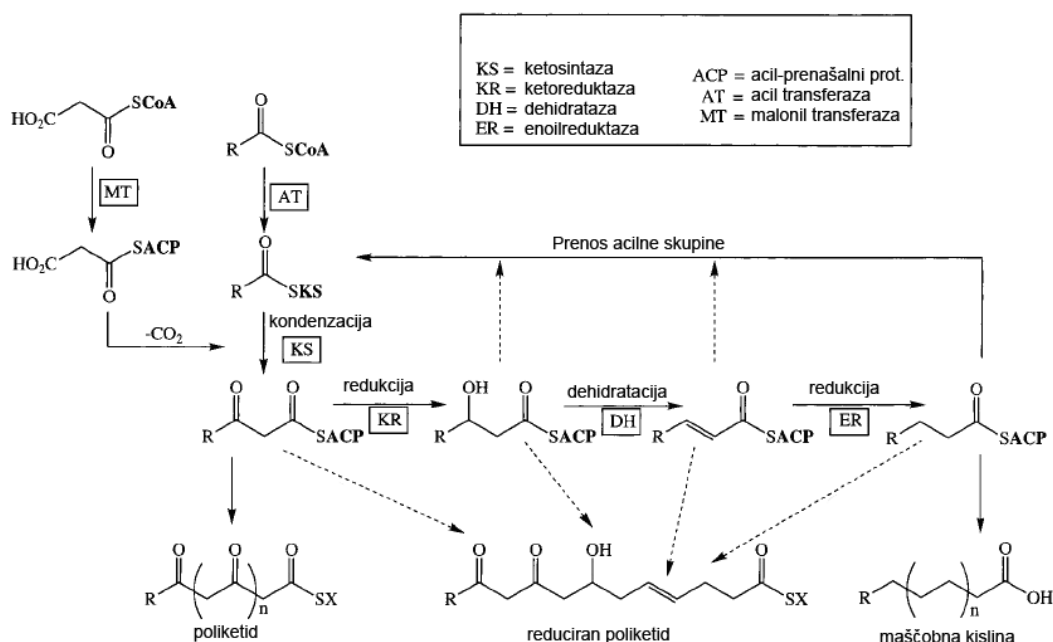
Do sedaj so našli že veliko število biološko aktivnih poliketidov, od katerih so vsem komercialno pomembnim poliketidom že določena zaporedja njihovih genov in so dobro raziskani. Hiter razvoj genomike je omogočil raziskovanje tudi ostalih, komercialno zaenkrat manj pomembnih poliketidov. Tako nastaja ogromna količina zaporedij, ki jih je potrebno anotirati, saj bomo s pomočjo teh podatkov lažje predvideli biosintezo poliketidov in nanjo tudi vplivali (Hopwood, 1997; Komaki in Harayama, 2006).

2.3.1 Biosinteza poliketidov

Kljub temu, da so poliketidi med seboj po strukturi različni, je mehanizem biosinteze poliketidov podoben za vse. Glede na podobnost s sintazami maščobnih kislin sklepajo, da izvirajo PKS prav iz sintaz maščobnih kislin. Osnovno ogljikovo ogrodje poliketidov je podobno kot pri maščobnih kislinah zgrajeno s kondenzacijo kratkih CoA-aktiviranih kislin, ki se zaporedno v procesu dekarboksilativne kondezacije kondezirajo v daljšo verigo, ki se naknadno še dodatno modificira (Slika 1). Kot začetne enote se najpogosteje uporabljajo acetil-CoA ali propionil-CoA, kot podaljševalne enote pa v glavnem malonil-CoA in metilmalonil-CoA (Hopwood, 1997).

Razlika v biosintezi maščobnih kislin in poliketidov je, da je β -ketonska skupina intermediata naraščajoče poliketidne verige večinoma podvržena popolni redukciji v

metilensko skupino, medtem ko je pri poliketidih lahko nereducirana, kot je to značilno za aromatske poliketide, ali pa je podvržena različnim stopnjam redukcije (Hopwood, 1997).



Slika 1: Biosintezna pot maščobnih kislin in poliketidov (Hopwood, 1997)

2.3.2 Poliketid sintaze

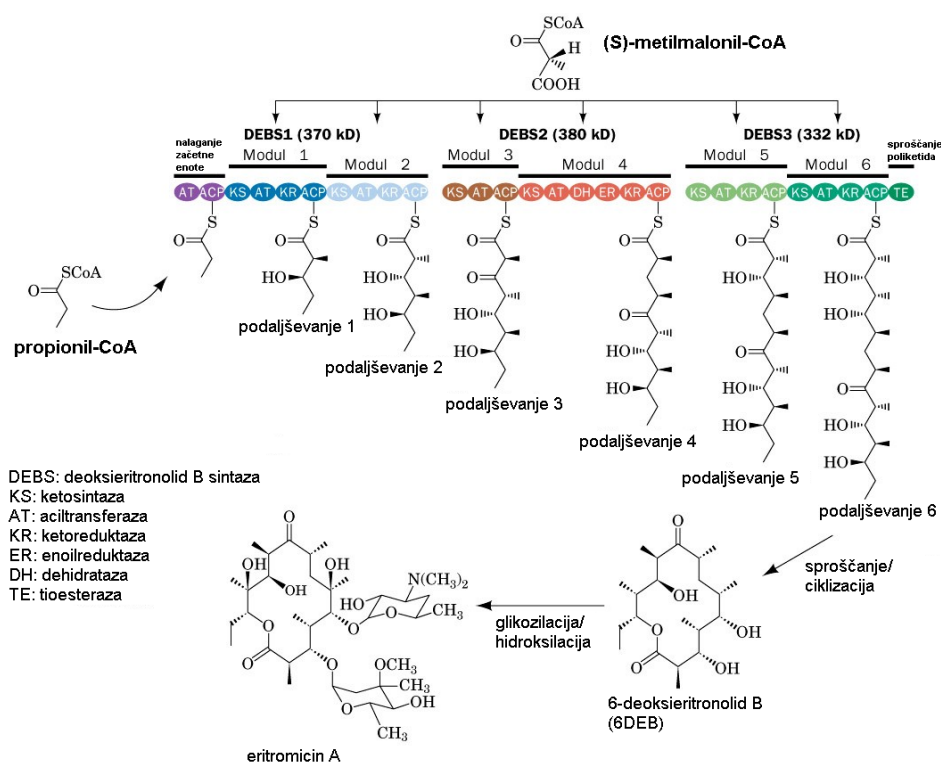
Skupek genov za biosintezo poliketidov imenujemo poliketid sintaze oz. skrajšano PKS. (Staunton in Weissman, 2001). Zaradi podobnosti s sintazami maščobnih kislin so tudi PKS najprej delili v dva tipa: tip I, ki združuje modularne PKS, in tip II, za katerega je značilen multiencimski kompleks. Novejša delitev poleg teh dveh osnovnih tipov dodaja še tretji tip, to so PKS, podobne kalkonskim sintazam (Austin in Noel, 2003; Shen 2003). Po tej definiciji sedaj velja, da obstajajo trije tipi poliketidnih sintaz, ki se med seboj razlikujejo po strukturi in funkciji (Komaki in Harayama, 2006), odkrivamo pa tudi poliketid sintaze, ki imajo lastnosti več skupin in jih zato ne moremo uvrstiti v nobenega izmed teh treh tipov. Kot primer lahko navedemo PKS, ki sodeluje pri biosintezi lovastatina in pripada modularnim sistemom tipa I, a se uporablja iterativno kot v primeru tipa II (Shen, 2003).

2.3.2.1 Poliketid sintaze tipa I

Najdene so bile predvsem pri bakterijah. So multifunkcionalni proteini, sestavljeni iz enega ali več modulov. Ti moduli vsak posebej katalizirajo en cikel v podaljševanju poliketidne verige, torej v nastajanju poliketidne verige navadno sodeluje vsak modul le enkrat. Vsak modul pa je sestavljen iz več domen. Vedno najdemo vsaj tri osnovne domene, to so ketosintaza, aciltransferaza in acilprenašalni protein. Ostale domene vključujejo encimske aktivnosti za redukcijo β -keto skupine, pogosto pa se med domenami pojavljajo ketoreduktaze, enoireduktaze in dehidrataze. Glavni produkti tega tipa PKS so makrolidi,

polietri in polieni. (Staunton in Weissman, 2001; Shen, 2003; Donadio in Sosio, 2003; Austin in Noel, 2003).

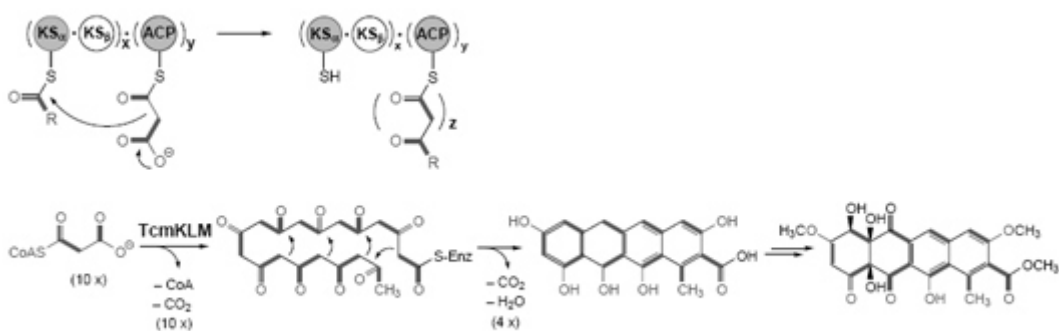
Dober primer PKS tipa I je PKS za biosintezo eritromicina (Slika 2). Celoten sistem je sestavljen iz treh velikih polipeptidov, vsak z dvema podaljševalnima moduloma. Prvi protein poleg podaljševalnih modulov vsebuje tudi dvodomenski (AT in ACP) modul za vezavo začetne enote. Biosinteza se začne z vezavo propionil-CoA na modul za vezavo začetne enote. Temu modulu sledi še šest podaljševalnih modulov, vsak sestavljen iz vsaj treh domen (KS, AT in ACP), ki zaključijo podaljševalni cikel, poleg teh domen pa moduli vsebujejo še dodatne domene, ki največkrat sodelujejo v redukciji podaljševalnih gradnikov. Molekula eritromicina tako nastane s karboksilno kondenzacijo iz propionil-CoA, kot začetne enote, in šestih molekul metilmalonil-CoA, ki služijo kot podaljševalne enote. Po kondenzaciji se molekula sprosti iz modula, nato pa je podvržena še dodatnim post-PKS modifikacijam.



Slika 2: Shematski prikaz biosinteze eritromicina (Khosla in sod., 2007)

2.3.2.2 Poliketid sintaze tipa II

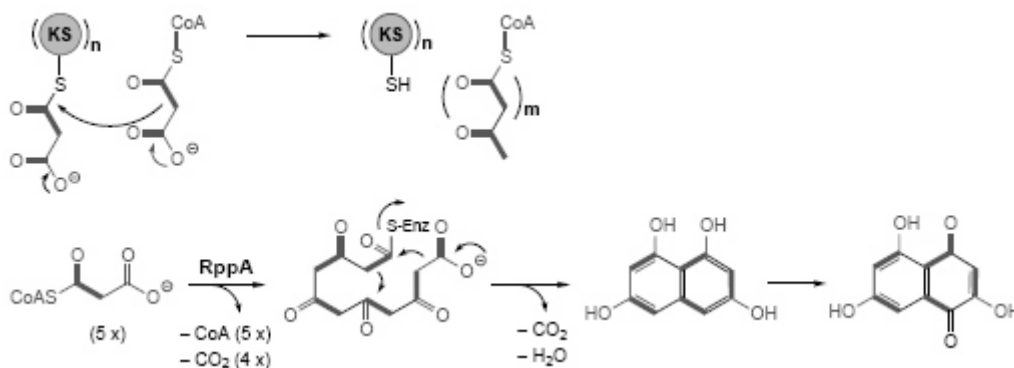
Za razliko od tipa I je za PKS tipa II značilno, da je multiencimski kompleks, pri katerem so proteini monofunkcionalni, torej se navadno vsaka encimsko katalizirana reakcija zgodi na različnem peptidu. Poliketidna veriga se tako sestavi v več ponavljajočih ciklih, kjer encimi sodelujejo večkrat (Slika 3). Tak tip PKS so našli predvsem pri bakterijah. Podrobneje o PKS tipa II piše v poglavju 2.4.4 (Shen, 2003; Donadio in Sosio, 2003; Austin in Noel, 2003).



Slika 3: Primer PKS tipa II (Shen, 2003)

2.3.2.3 Poliketid sintaze tipa III

Gre za t.i. kalcion sintaze, ki jih najdemo predvsem pri rastlinah in nekaterih bakterijah. So homodimerni encimi z enostavno arhitekturo, ki jih naredi veliko enostavnejše za opazovanje in manipulacijo. Za razliko od ostalih dveh tipov za delovanje ne potrebujejo ACP (Slika 4). Eksperimentalni podatki kažejo, da so PKS tipa III strukturno podobni homodimernim KAS tipa III. Encimi PKS tipa III so sposobni katalizirati spreminjajoče se število ponavljajočih kondenzacij in lahko uporabijo širok nabor začetnih molekul (Shen, 2003; Donadio in Sosio, 2003; Austin in Noel, 2003).



Slika 4: PKS tipa III (Shen, 2003)

2.3.2.4 Drugi tipi poliketidnih sintaz

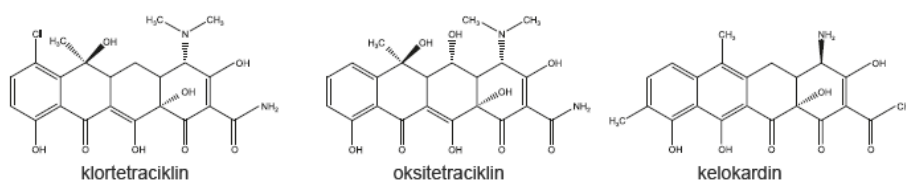
Razcvet raziskav poliketidnih sintaz je pripeljal do odkritja mnogo novih PKS in ugotovitev, da vseh poliketidnih sintaz ne moremo razdeliti enostavno v 3 osnovne tipe, saj obstajajo tudi izjeme, ki imajo značilnosti več tipov. Primer je lahko sintaza 6-metilsaliciklične kisline, ki bi naj pripadala tipu II, saj na iterativni način nastane aromatičen produkt, a spada med PKS tipa I, ker je producirana v glivah. Podobnih primerov je več, obstajajo tudi poliketidne sintaze tipa I, ki nimajo drugače značilnega encima acil transferaze (AT), znani so tudi primeri PKS tipa II, ki ne potrebujejo ACP in ne delujejo iterativno ipd. (Shen, 2003).

2.4 TETRACIKLINI

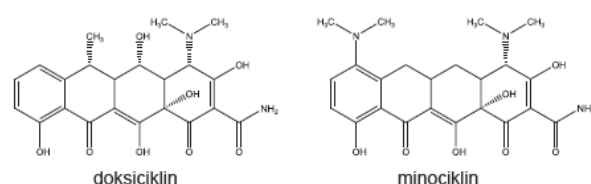
2.4.1 Odkritje in spekter delovanja

Tetraciklini so družina antibiotikov s širokim spektrom delovanja. Prva tetraciklina so odkrili že v 40-tih letih. To sta bila klortetraciklin, ki ga proizvaja *Streptomyces aureofaciens* in oksitetraciklin, ki ga proizvaja *S. rimosus*. Kasneje odkrite tetracikline lahko razdelimo v dve skupini (Slika 5). Lahko so biosintezni proizvod mikroorganizmov, med katerimi jih proizvajajo predvsem bakterije iz rodu *Streptomyces* (npr. demetilklorid), ali pa so pol-sintezni derivati, kot sta npr. doksiciklin in minociklin (Chopra in Roberts, 2001).

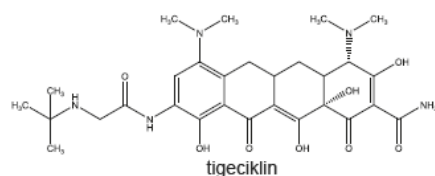
Naravni produkti:



Druga generacija:



Tretja generacija:



Slika 5: Primeri tetraciklinov iz različnih generacij (Zakeri in Wright, 2008)

Molekule tetraciklinov aktivno delujejo tako proti po Gramu pozitivnim, kot po Gramu negativnim bakterijam, atipičnim mikroorganizmom, kot so klamidije, mikoplazme in rikecije, ter tudi protozajem. Njihova največja prednost je odsotnost močnih stranskih učinkov, zato se intenzivno uporabljajo v zdravljenju infekcij tako v medicini, kot tudi veterini. Uporabljajo se tudi kot profilaktiki pred malarijo (doksicilin), v nekaterih državah pa jih še vedno uporabljajo tudi kot rastne faktorje za živali, tako se npr. klortetraciklin uporablja v ribogojnicah. Poleg tega, da so jih uporabljali ob infekcijah, so tetracikline od začetka 50-tih let predpisovali tudi ob neinfektivnih stanjih (Chopra in Roberts, 2001).

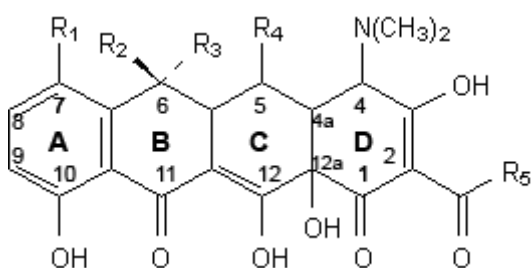
Zaradi velike porabe tetraciklinov se je pojavilo že veliko patogenih organizmov, ki so odporni na tetracikline, zato je sedaj njihova uporaba omejena, še posebej je pod vprašajem uporaba tetraciklinov kot rastnih faktorjev pri živalih (Chopra in Roberts, 2001).

Odgovor na naraščajočo odpornost na tetraciklinske antibiotike je leta 2005 odobren antibiotik tigecklin. Tigecklin je antibiotik širokega spektra, ki spada v tretjo generacijo

tetraciklinov in je zaenkrat edini predstavnik glicilciklinov. Je derivat minociklina, ki mu je dodana glicilamido funkcionalna skupina na 9. C-atom D obroča. Zaradi dodane stranske skupine tige ciklin zaobide različne rezistenčne mehanizme s proteini, ki odstranjujejo molekulo iz celice (angl. efflux proteins), kot tudi mehanizme rezistence zaradi spremenjenega vezavnega mesta na ribosomu, saj se na 30S vezavno mesto veže tudi ob prisotnosti rezistenčnih mehanizmov. Ameriški vladni urad za zdravila in prehrano (FDA; angl. Food and Drug Administration) je tige ciklin odobril za zdravljenje zapletenih infekcij kože in mehkega tkiva ter abdominalnih infekcij (Peterson, 2008; Greer, 2006; Doan in sod., 2006).

2.4.2 Kemijske lastnosti tetraciklinov

Molekula tetraciklina je sestavljena iz štirih linearno povezanih krožnih molekul, na katere so vezane različne funkcionalne skupine. Najenostavnejši tetraciklin s protibakterijsko aktivnostjo je 6-deoksi-6-dimetiltetraciklin (glej Slika 6). Za protimikrobno aktivnost so pomembni: linearno spojeni tetracikel, naravno prisotna stereokemijska konfiguracija na 4a, 12a (povezava med A-B obročema), dimetilamino skupina na 4. C-atomu mestu in ohranjen ketoenolni sistem na pozicijah 11, 12 in 12a. (Chopra in Roberts, 2001).



	R ₁	R ₂	R ₃	R ₄	R ₅
tetraciklin	H	OH	CH ₃	H	NH ₂
6-demiltetraciklin	H	OH	H	H	NH ₂
2-acetil-2-dekarboksamid-tetraciklin	H	OH	CH ₃	H	NH ₂
klortetraciklin	Cl	OH	CH ₃	H	NH ₂
6-demitklortetraciklin	Cl	OH	H	H	NH ₂
oksitetraciklin	H	OH	CH ₃	OH	NH ₂
6-deoksitetraciklin	H	H	CH ₃	H	NH ₂
2-acetil-2-dekarboksamid- oksitetraciklin	H	OH	CH ₃	OH	CH ₃
minociklin	N(CH ₃) ₂	H	H	H	NH ₂

Slika 6: Kemijska struktura različnih tetraciklinskih antibiotikov

2.4.3 Mehanizem delovanja tetraciklinskih antibiotikov

Za tetracikline na splošno velja, da delujejo na bakterijo z inhibicijo sinteze proteinov. Inhibicija je posledica vezave antibiotika na vezavno mesto A na 30 S ribosomalni

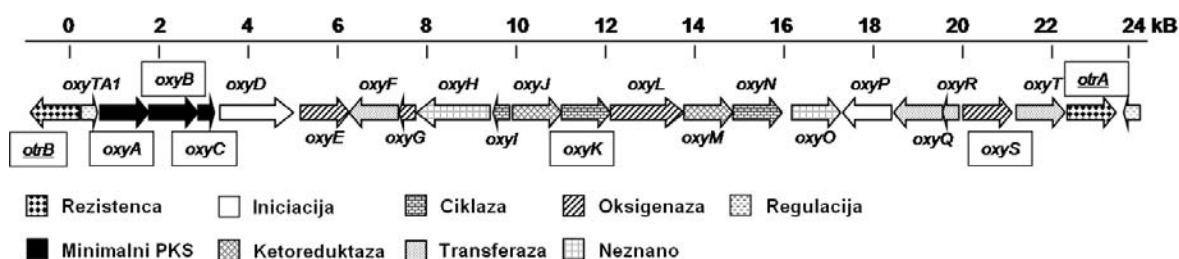
podenoti, kar prepreči vezavo aminoacil tRNA na njeno vezavno mesto na ribosomu, s tem pa ne more priti do začetka translacije (Rasmussen in sod. 1991; Chopra in Roberts, 2001).

Za manjšo skupino t.i. atipičnih tetraciklinov je značilno, da se razlikujejo glede mehanizma delovanja. Ti tetraciklini namreč slabo inhibirajo proteinsko sintezo, a kljub temu močno inhibirajo vstavljanje radioaktivno označenih prekurzorjev v makromolekule, npr. v DNA, RNA in proteine. Drugačen način delovanja atipičnih tetraciklinov so dokazali tudi s poskusi na proteinskih lizatih, kjer ta skupina tetraciklinov skoraj ni imela vpliva na inhibicijo translacije, kar je sicer značilnost večine tetraciklinov (Rasmussen in sod., 1991). Nadaljnji eksperimenti so pokazali, da ta skupina atipičnih tetraciklinov, v katero spadata tudi kelokardin in 6-tiatetraciklin, verjetno primarno deluje na citoplazemsko membrano. Najverjetneje interakcija teh molekul z membrano povzroča lizo celic (Chopra, 1994).

2.4.4 Biosinteza oksitetraciklina pri *Streptomyces rimosus*

Študij biosinteze klortetraciklina in oksitetraciklina, ki sta prva odkrita tetraciklina, so začeli izvajati že v 60-tih letih prejšnjega stoletja. Raziskave so večinoma temeljile na analizah mutant, nastalih z blokiranjem genov in poskusih z dohranjevanjem z različnimi substrati. Do večjega preskoka v razumevanju biosinteze je prišlo šele v devetdesetih letih, ko so bile razvite metode molekularne genetike in so določili zaporedje skupini genov za biosintezo oksitetraciklina (Hunter in Hill, 1997; Zhang in sod., 2006) in klortetraciklina (Ryan, 1995). Na podlagi razumevanja biosinteze oksitetraciklina in klortetraciklina lahko ugotovimo, da sta si biosintezi teh dveh tetraciklinov zelo podobni.

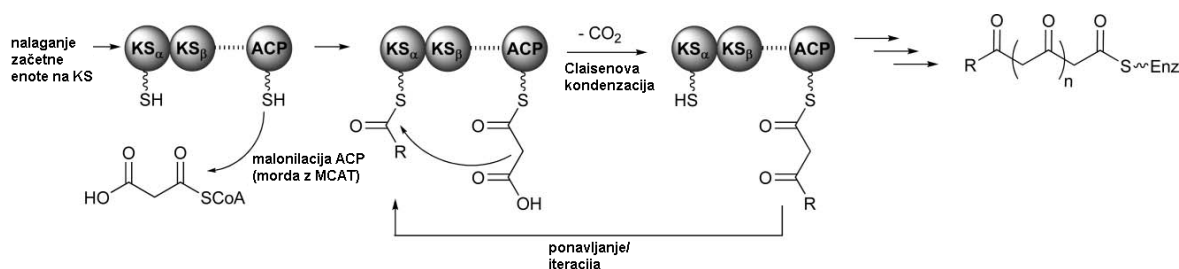
Skupina genov za biosintezo oksitetraciklina vsebuje 21 odprtih bralnih okvirjev (Slika 7), ki se nahajajo med dvema genoma za rezistenco (Petković in sod., 2006; Zhang, 2006) ter nedavno identificiranega regulatornega proteina (Lešnik in sod., 2009). Prvi gen, ki kodira rezistenco, *otcX*, je transporter, ki črpa tetraciklin iz celice (angl. efflux pump), drugi gen za rezistenco pa kodira protein, ki varuje ribosom pred vezavo antibiotika (RPP; angl. Ribosomal Protection Protein). RPP so topni proteini, ki so po zgradbi podobni elongacijskim faktorjem in imajo GTPazno aktivnost. Rezistenčni mehanizmi so pogosto kodirani znotraj skupine genov za biosintezo antibiotikov, saj kodirajo odpornost organizma na delovanje tetraciklina, ki ga sam proizvede (Zakeri in Wright, 2008).



Slika 7: Shema skupine genov za biosintezo oksitetraciklina

Biosinteza osnovnega skeleta tetraciklinskih antibiotikov poteka s pomočjo encimov, ki tvorijo t.i. minimalni PKS (Slika 8). Tega sestavljata dve ketosintazi (KS α in KS β) in acilprenašalni protein (ACP), ki so v genski skupini za biosintezo oksitetraciklina označeni

kot OxyA, OxyB in OxyC. ACP služi kot mesto, kamor se veže naraščajoča veriga, ta pa jo nato prenese na aktivno mesto ketosintaze α ($KS\alpha$). Ta minimalni PKS kompleks katalizira Claisen-podobno reakcijo kondenzacije, ki privede do nastanka C-C vezi. Ta reakcija se potem ponavlja, kar vodi do daljšanja verige, pri tem pa njeno dolžino v večji meri določa ketosintaza β ($KS\beta$) (Slika 8). $KS\alpha$ in $KS\beta$ sta si zelo podobni, razlikujeta se le v aktivnem mestu s cisteinom, ki ga $KS\alpha$ ima, $KS\beta$ pa je katalitično neaktiven, saj ima cistein zamenjan z glutaminom (Zhang in sod., 2007; Zakeri in Wright, 2008).



Slika 8: Slika prikazuje reakcije, ki jih katalizirajo encimi minimalnega PKS in vodijo do nastanka poliketidne verige (Zhang in sod., 2006)

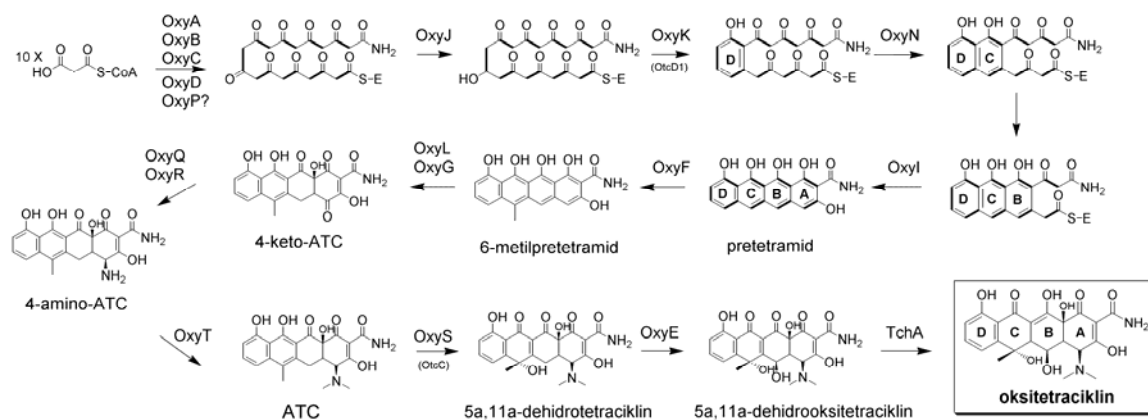
Kot začetna enota v biosintezi oksitetraciklina služi malonamil-CoA, na katerega encimi minimalnega PKS v ponavljajočih reakcijah dekarboksilativne kondenzacije nanizajo še osem ekvivalentov malonil-CoA, da nastane ogrodje iz 19 C-atomov (Slika 8). Za zapiranje obročev sta v nadaljevanju pomembni dve aromatazi/ciklazi (Petković in sod, 1999), OxyK in OxyN, ki sta odgovorni za zapiranje obročev D, C in B, medtem ko zaprtje obroča A verjetno potече spontano (Zhang in sod., 2006).

V skupini genov sta kodirani tudi dve metiltransferazi. OxyF je C-metiltransferaza in predvidoma metilira ogljikov atom na poziciji C-6, medtem ko OxyT metilira amino skupino, ki jo na 4. C-atom predhodno doda aminotransferaza OxyQ. (Petković in sod., 2006; Zhang in sod., 2006).

V skupini genov najdemo tudi dva gena (*oxyJ* in *oxyM*), ki kodirata od NADH-odvisne ketoreduktaze. Zhang in sodelavci (2006) so potrdili, da OxyJ katalizira redukcijo na 9. C-atomu, medtem ko za predvideno ketoreduktazo OxyM točne vloge v biosintezi oksitetraciklina zaenkrat še ne poznamo. Med geni genske skupine najdemo dva gena, ki vplivata na dodajanje dušikovih komponent v molekulo. Gen *oxyD* predvidoma kodira amidotransferazo, ki sodeluje v biosintezi amidne začetne enote, medtem ko *oxyQ* verjetno kodira aminotransferazo, ki aminira ogljikov atom na C-4 mestu pri 4-keto-ATC (Slika 9) (Zhang in sod., 2006).

Skupina genov vsebuje tudi štiri gene, ki katalizirajo oksido-redukcijske reakcije, ki kažejo veliko podobnost z oksigenazami v skupini genov za biosintezo mitramicina, kar nakazuje njuno evolucijsko povezanost. Za MtmOII so Prado in sodelavci (1999) dokazali, da sodeluje pri začetni hidroksilaciji v biosintezi mitramicina, zato tudi za OxyL predvidevajo, da katalizira analogno reakcijo, ki uporablja kot substrat 6-metil-pretetramid. OxyG je majhna oksigenaza, ki tvori kinon. Zanj predvidevajo vpletenost v kinonsko formacijo obroča A v 4-keto-ATC (Slika 9). OxyE je verjetno od FAD odvisna

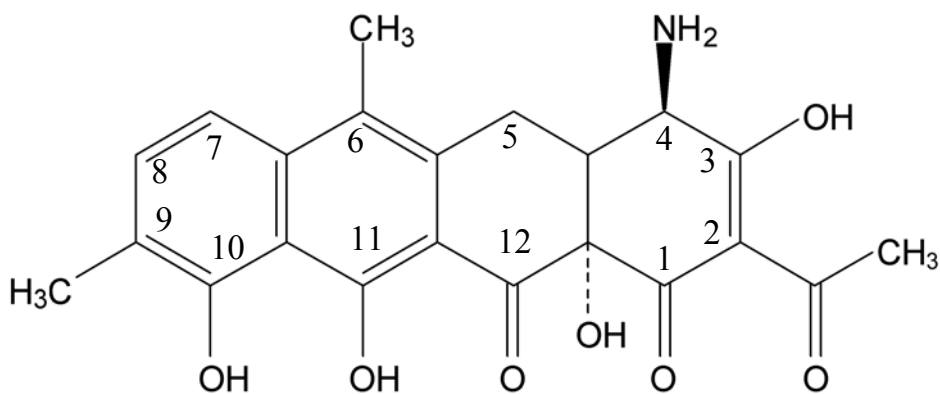
monoooksigenaza, ki je verjetno vpletena v oksidacijo ogljikovega atoma na C-4 mestu (Slika 9) (Wang in sod., 2009; Zhang in sod., 2007; Zakeri in Wright, 2008).



Slika 9: Shematski prikaz možnih stopenj pri biosintezni poti oksitetraciklina (Zhang in sod., 2006)

2.5 KELOKARDIN

Kelokardin je antibiotik širokega spektra, ki ga proizvaja po Gramu pozitivna bakterija *Amycolatopsis sulphurea*. Prvi so ga opisali leta 1962 (Oliver in sod., 1962). Po strukturi pripada družini tetraciklinov, čeprav ima struktura tudi nekaj lastnosti, ki so med tetraciklini manj pogoste. Ena takih je prisotnost druge C-metilne skupine na 9. C-atomu, ki je bila ob odkritju kemijske strukture kelokardina novost v tej skupini antibiotikov (Slika 10)(Mitcher in sod., 1970).



Slika 10: Molekula kelokardina. Oštevilčena je po sistemu IUPAC.

Kelokardin uvrščamo v posebno skupino tetraciklinov, ki ne delujejo na translacijo proteinov, ampak najverjetneje vplivajo na delovanje celične membrane (Chopra, 1994). Nepravilno delovanje membrane je verjetno povezano z relativno planarnostjo B, C in D obroča, zaradi česar je molekula lipofilna in neionizirana. Pri interakciji s celico atipični tetraciklini tako ostanejo ujeti v hidrofobnem okolju citoplazmatske membrane, s čimer povzročijo motnje v njenem delovanju (Chopra in Roberts, 2001).

Glede na verjetni mehanizem delovanja kelokardina, ta ni onemogočen z običajnimi mehanizmi rezistence, ki bodisi odstranjujejo antibiotike iz citoplazme (Tet(B) Tet(K)), bodisi varujejo ribosome pred inhibicijo s tetraciklini (Tet(M)). Prav tako za razliko od večine tetraciklinov, kelokardin in 6-tiatetraciklin delujeta na celice bakteriocidno in ne bakteriostatično (Chopra, 1994).

Učinkovito delovanje 6-tiotetraciklina na bakterije, ki posedujejo odpornost na tetracikline, je vzpodbudilo izvedbo druge faze kliničnih poizkusov na prostovoljcih. Antibiotik se je pri tem izkazal kot učinkovit, a so ga zaradi nezaželenih stranskih učinkov prepovedali za uporabo v medicini (Chopra, 1994).

2.6 BIOINFORMATIKA

Bioinformatika je rezultat interdisciplinarnosti sodobnih bioloških znanosti (Xia, 2007). Kaminski je bioinformatiko definiriral kot uporabo računalnikov, podatkovnih baz in računalniških metod za obvladovanje bioloških informacij (Kaminski, 2000). V zadnjih nekaj desetletjih je napredek na področju molekularne biologije v povezavi z naprednimi genskimi tehnologijami privedel do eksplozivne rasti bioloških informacij, ki jih je pridobila znanstvena skupnost (NCBI, 2004b), zato je uporaba bioinformacijskih orodij nujna v skoraj vseh vidikih upravljanja s podatki, ki jih dobimo s sodobno biologijo (Kaminski, 2000).

2.7 ODPRTI BRALNI OKVIR

Bakterijski geni, ki kodirajo proteine, so navadno dolga in neprekinjena zaporedja, ki jih pogosto imenujemo odprti bralni okvir (ORF; angl. Open Reading Frame). Odprti bralni okvir bi lahko definirali kot zaporedje kodonov, ki se začnejo s startnim kodonom in končajo s enim od stop kodonov. Najenostavnejša pot za iskanje odprtih bralnih okvirjev je prevesti zaporedje v vseh 6 teoretičnih bralnih okvirjev in s primernim programom poiskati vsa možna aminokislinska zaporedja med start-kodonom in stop-kodonom. Pri iskanju je priporočljivo določiti minimalno število nukleotidov v odprtem bralnem okvirju, saj s tem enostavno izločimo prekratke, naključne zadetke, ki imajo podobno zaporedje, a ne kodirajo iste funkcije. Težje je detektirati t.i. »navidezne gene« (angl. shadow genes), ki imajo prekrivajoče odprte bralne okvirje na nasprotni DNA verigi. Za iskanje takih genov so primerni algoritmi, ki uporabljajo t.i. skriti Markov model. Taki algoritmi dobljene rezultate primerjajo z že poznanimi zaporedji genov drugih organizmov, ki so shranjeni v bazah podatkov (Orengo in sod., 2003).

2.8 PODATKOVNE ZBIRKE

Baza bioloških podatkov je velika, organizirana, rastoča zbirka podatkov, ponavadi povezana z računalniškimi programi, ki nadzorujejo dodajanje in podajanje podatkov, shranjenih v sistemu. Baze podatkov so velikokrat medsebojno povezane (NCBI, 2004b).

Večina bioinformacijskih podatkov je shranjena v treh velikih bazah podatkov. To so ameriška NCBI (angl. National Center for Biotechnology Information), evropska EMBL (angl. European Molecular Biology Laboratory) in japonska DDBJ (angl. DNA Database of Japan). Te tri inštitucije so združene v organizacijo INSDC (angl. International

Nucleotide Sequence Database Collaboration), ki je bila ustanovljena pred več kot 18 leti. Eno izmed ustanovnih članic, Los Alamos National Laboratory, je v tem času zamenjala DDBJ (Mizrachi, 2007; INSDC, 2009).

Največ podatkov vse baze v večini pridobijo neposredno od znanstvenikov, ki jih individualno vnašajo preko spletnih obrazcev. Velik vir zaporedij so tudi genomski projekti in patentne prijave. Te podatke vsaka od inštitucij zbira posebej, nato pa jih na dnevni ravni združijo in tako ustvarjajo največjo bazo bioinformatičnih podatkov, imenovano Internacionalna baza nukleotidnih zaporedij (angl. International Nucleotide Sequence Database) (INSDC, 2009).

Vsako dodano zaporedje v to bazo dobi svojo dostopno številko (angl. accession number), ki dokazuje, da je bilo zaporedje res vpisano, poleg tega pa omogoča, da se avtorji v člankih nanjo sklicujejo, saj je enostavno dostopna. Dostopna številka ostane vedno enaka, tudi če avtorji vnesejo v zaporedje kake spremembe. Prav tako je enaka v vseh treh sodelujočih bazah (Benson in sod., 2008).

Podatke iz skupne baze lahko članice na svojih straneh prikazujejo v različnih oblikah in poleg ponujajo svoja orodja za delo z njimi. Kljub temu so z namenom standardizacije zapisov sestavile navodila, ki definirajo obliko zapisa zaporedij. S tem so pridobile zapise, zapisane v standardizirani obliki, ki zato ne povzročajo težav pri obdelavi. Lastnosti zapisov so shranjene v dokumentu z angleškim naslovom »Feature Table Definition« (Mizrachi, 2007).

2.8.1 Sistem Entrez

Entrez je sistem za iskanje po povezanih bazah podatkov, ki ga ureja NCBI. Vsebuje znanstveno literaturo, DNA in proteinske baze, 3D proteinske strukture in baze proteinskih domen, celotne genome, taksonomske informacije in druge podatke, ki so z medsebojnimi povezavami povezani v velik sistem (NCBI, 2009b).

Sistem Entrez sestavlja tudi GenBank, ki je baza vseh javno dostopnih DNA zaporedij in njihovih prevodov v proteinska zaporedja (NCBI, 2004a). Ustvarjena je bila leta 1982, v povprečju se število vnosov vanjo podvoji v 18 mesecih, predvidevanja pa kažejo, da bo v letu 2009 število njenih shranjenih zaporedij preseglo 100 milijonov (NCBI, 2009b).

2.8.2 Baze proteinskih zaporedij

Baze proteinskih podatkov lahko v osnovi razdelimo na dve skupini. V prvo skupino uvrščamo zaporedja, ki so prevedena iz kodirajočih regij, dobljena iz DNA zaporedij, shranjenih v GenBank, EMBL in DDBJ, v drugo skupino pa spadajo proteinska zaporedja, shranjena v bazah PIR (angl. Protein Information Resource), UniProtKB/SWIS-PROT, Protein Research Foundation in PDB (angl. Protein Data Bank). Za zaporedja iz druge skupine je značilno, da so prevedena in preverjena s strani znanstvenikov, medtem ko so zaporedja iz prve skupine prevedena avtomatsko (NCBI, 2006; Mizrachi, 2007).

2.8.2.1 UniProt – baza proteinskih zaporedij

UniProt baza je nastala z ustanovitvijo konzorcija treh inštitucij: Švicarskega inštituta za bioinformatiko (angl. Swiss Institute of Bioinformatics), Evropskega inštituta za bioinformatiko (angl. European Bioinformatics Institute) in skupine PIR, ki je del univerze Georgetown. Pri tem so omenjene institucije združile svoje baze in osnovale univerzalno bazo UniProt.

UniProt (angl. Universal Protein Resource) baza sestoji iz štirih delov, izmed katerih je najpomembnejša UniProt Knowledgebase (UniProt), ki je centralna baza proteinskih zaporedij z natančnimi, konsistentnimi in bogatimi zaporedji in funkcijskimi anotacijami., ki so med seboj navzkrižno povezane. Baza sestoji iz dveh sekcij: UniProt/Swiss-Prot sekcija vsebuje ročno vnesena, klasificirana in preverjena zaporedja in UniProt/TrEMBL sekcijo, katera zaporedja so klasificirana in določena avtomatsko (Bairoch in sod., 2005). Ker vsebuje UniProt/Swiss-Prot baza ročno anotirana zaporedja, vsebuje veliko manj zaporedij kot TrEMBL, zato pa vsebuje veliko dodatnih informacij, ki jih z avtomatsko anotacijo ne moremo pridobiti. Za boljšo kakovost anotacij obstaja težnja, da se tudi zaporedja v TrEMBL ročno obdelajo in potem prenesejo v Swiss-Prot bazo (Mizrachi, 2007).

2.8.3 Sekundarne baze podatkov

Baze podatkov značilnosti, ki so skupne proteinskim družinam, domenam in funkcionalnim skupinam, so postale pomembno orodje pri računalniški funkcionalni klasifikaciji novih zaporedij, ki še nimajo določene biološke funkcije (Apweiler in sod., 2001).

V zadnjih desetih letih so razvili veliko metod za prepoznavanje značilnosti in združevanje zaporedij. Te metode so se razvijale vsaka zase za reševanje različnih analitičnih problemov, zato za reševanje različnih problemov niso vse enako primerne. Najpogosteje uporabljane sekundarne proteinske baze so PROSITE, PRINTS, Pfam in ProDom. Pri tem se Pfam fokusira predvsem na divergentne domene, PROSITE na analizo funkcionalnih skupin, PRINTS je uveljavljen predvsem pri iskanju družin, natančneje pri specializaciji in definiranju hierarhije od super-družin do nivojev pod družino (EMBL-EBI, 2009b; Apwiler in sod., 2001; Hunter in sod., 2009).

Vse omenjene sekundarne baze proteinskih podatkov ne vsebujejo enakega formata zapisa in nomenklature, zato so pri EBI razvili bazo InterPro (angl. Integrated Resource of Protein domains and functional sites). Ta baza združuje vse zgoraj omenjene proteinske sekundarne baze v enem paketu, pri tem pa omogoča dostop do širšega in komplementarnega obsega strani in metod za prepoznavanje domen (EMBL-EBI, 2009b; Apwiler in sod., 2001; Hunter in sod., 2009).

2.9 DOLOČANJE VERJETNIH BRALNIH OKVIRJEV V BAKTERIJSKIH GENOMIH Z VIŠJO VSEBNOSTJO G+C

Bakterijski genom je za razliko od evkariontskih genomov veliko lažji za preučevanje. Odsotnost intronov, krajše zaporedje povprečnih odprtih bralnih okvirjev in velik del

kodirajočega zaporedja naredijo bakterije enostavnejše glede bioinformatičnih raziskav. Tako so že leta 1995 objavili prvi genom bakterije, in sicer za *Haemophilus influenzae* (Orengo in sod., 2003). Do sedaj so določili zaporedja tudi več aktinomicetnih genomov, npr. genom *Mycobacterium tuberculosis*, *M. smegmatis*, *Corynebacterium diphtheriae*, *Streptomyces avermitilis*, *S. coelicolor*, *S. griseus*, *Saccharopolyspora erythraea*, *Frankia alni* itd. (Ishikawa in sod., 2009).

Za analizo aktinomicetnih genomov z visoko vsebnostjo G+C so na Oddelku za bioaktivne molekule na Nacionalnem inštitutu za infekcijske bolezni v Tokiu (Ishikawa in Hotta, 1999) program Frame (Bibb in sod., 1984) nadgradili v program FramePlot (poglavje 2.9.1), ki nam olajša iskanje odprtih bralnih okvirjev na podlagi višje vsebnosti nukleotidov G+C na tretjem mestu v kodonu.

2.9.1 FramePlot – program za iskanje verjetnih odprtih bralnih okvirjev

FramePlot 4.0 beta (v nadaljevanju FramePlot) je brezplačen, na spletu dostopen program za določanje predvidenih odprtih bralnih okvirjev. Program se odlično obnese pri določanju odprtih bralnih okvirjev v zaporedjih organizmov z visokim odstotkom G+C (Ishikawa, 2008), pri čemer program temelji na dejstvu, da imajo streptomicete na tretjem mestu v kodonu v povprečju 92% G+C (Nakamura in sod., 1997).

2.10 METODE ZA PRIMERJAVO ZAPOREDIJ

Najpogostejša metoda za analizo zaporedij je iskanje podobnih zaporedij v bazah podatkov, ki vsebujejo že shranjena in opisana zaporedja. Poleg te metode uporabljamo še primerjavo nekaj zaporedij med sabo, npr. poravnavo dveh zaporedij in iskanje njihovih razlik in skupnih točk. Bolj napredna metoda, ki jo uporabljamo, je iskanje motivov v zaporedju in predvidevanje 3-dimenzionalne strukture (Kaminski, 2000).

Dejstvo, da zaporedja med seboj primerjamo, temelji na predpostavki, da si bolj podobna zaporedja delijo skupnega prednika in/ali pa kodirajo proteine s podobno funkcijo. Najboljši način, da ugotovimo njihovo podobnost je, da jih poravnamo in jim dodamo vrzeli, da dobimo čim boljše ujemanje, če je to potrebno (Kaminski, 2000).

Odkritje homologije zaporedja z že poznanim proteinom ali družino proteinov privede do prvih sklepov o funkciji novega gena (Altschul, 1990). Mnogo funkcionalnih in evolucijsko pomembnih proteinskih podobnosti lahko prepoznamo le s primerjavo 3-dimenzionalnih struktur. Ko teh struktur ni na voljo, si pomagamo s prepoznavanjem ohranjenih vzorcev, ki jih pridobimo s poravnavanjem zaporedij (Altschul in Koonin, 1998).

V polipeptidni verigi se ponavlja 20 različnih aminokislin, zato je še posebej pri krajših polipeptidih velika verjetnost, da bosta imela dva polipeptida podobna zaporedja. Ugotovili so, da je za naključna proteinska zaporedja ta podobnost med 5 in 10%, če pa pri poravnavanju zaporedij dovolimo še vrzeli, se ta verjetnost poveča na 10-20%. V primeru ugotavljanja podobnosti nukleinskih kislin, ki so sestavljene iz le štirih gradbenih elementov, je število po naključju podobnih zaporedij še večje (Orengo in sod., 2003). Z namenom, da zmanjšamo število naključnih ujemanj, ki ne odražajo sorodnosti, je

priporočljivo, da primerjamo med seboj proteinska zaporedja, oz. da nukleotidna zaporedja najprej prevedemo v proteinska in jih šele nato primerjamo. Proteinska zaporedja so namreč precej bolj ohranjena kot nukleotidna zaporedja, ki jih kodirajo. V nukleotidnih zaporedjih lahko pride do mutacij, ki jih v prevedenem, proteinskem zaporedju zaradi degenerativne narave genskega koda ne zaznamo (Dale in von Schantz, 2007; Orengo in sod., 2003).

2.10.1 Iskanje homolognih zaporedij v podatkovnih bazah

Metode za primerjavo proteinskih zaporedij lahko v grobem razdelimo v dve skupini. V eno skupino štejemo natančne metode, ki preverijo vse možne kombinacije, da pridejo do idealne poravnave. Večina teh robustnih algoritmov deluje na principu optimizacijske strategije, poznane kot dinamično programiranje. Slabost teh metod je velika poraba procesorske moči, s tem pa tudi časa. Druga skupina metod so hitre metode, ki ne izračunajo idealne poravnave, je pa ta poravnava dovolj dobra, da s pomočjo poravnave najdemo v bazi podatkov najbolj podobna zaporedja. Te metode so veliko hitrejše in so zato primernejše za iskanje homologij v velikih bazah, kot je npr. GenBank (Orengo in sod., 2003; Xia, 2007).

Osnovni princip za primerjavo zaporedij je pri vseh algoritmih podoben. Kot prvo je potrebno imeti shemo točkovanja podobnih aminokislinskih ostankov, da so podobni aminokislinski ostanki v shemi med seboj zamenljivi. Drug faktor je optimizacijska strategija, ki mora preveriti vse alternativne možnosti poravnave proteinskih ostankov, da lahko tako maksimizira število poravnanih podobnih zaporedij (Orengo in sod., 2003).

Metodi FASTA in BLAST spadata v drugo, hevristično skupino metod za primerjavo zaporedij. Obe temeljita na Smith-Waterman algoritmu (Smith in Waterman, 1981), ki poišče in poravna regije podobnosti med dvema zaporedjema (Xia, 2007). Lahko se zgodi, da bosta zgrešili kako homologno zaporedje, a sta v primerjavi z metodami iz prve skupine veliko hitrejši in zato največkrat uporabljeni za iskanje v velikih bazah podatkov, kjer je hitrost pomembnejši faktor kot občutljivost metode (Xia, 2007; Altschul in Koonin, 1998)

2.10.1.1 Program BLAST

BLAST (angl. Basic Local Alignment Search Tool) je algoritem, ki so ga razvili Altschul in sodelavci (1990). Poleg paketa programov FASTA (Pearson in Lipman, 1988) razvitega v 80-tih letih prejšnjega stoletja, predstavlja BLAST najpomembnejše orodje za hitro primerjavo dveh zaporedij (Orengo in sod., 2003).

Osnovni algoritem BLAST je enostaven in robusten. Uporablja se ga v različne namene in za primerjavo različnih zaporedij. Algoritem se uporablja za primerjavo nukleotidnih in proteinskih zaporedij, za iskanje motivov, identifikacijo genov, iskanje večkrat ponovljenih podobnih regij v dolgem zaporedju in še druge namene (Altschul in sod., 1990).

Algoritem BLAST v obeh zaporedjih najprej poišče t.i. »vroče točke«. To so krajša ujemanja med obema zaporedjema, iz katerih nato nadaljuje iskanje v tisti regiji, dokler točkovanje ne pade pod določeno število točk (Johnson in sod., 2008). S tem postopkom

najdemo identična ali podobna zaporedja s tistimi, ki so že shranjena v bazah nukleotidnih in proteinskih zaporedij, ki so na voljo. Vsako ujemanje zaporedje nad določeno dolžino se shrani in pozneje pojavi v popisu najboljših ujemanj. Algoritem na podlagi dobro definirane statistične interpretacije pravilne zadetke loči od naključnih zadetkov. Zadetke nam razvrsti glede na stopnjo podobnosti od največje podobnosti naprej, kar naredi tako, da jih porazdeli glede na E-vrednost (Altschul in sod., 1990). E-vrednost ali pričakovana vrednost je parameter, ki nam pove verjetnost, da je ujemanje med zaporedjema nastalo po naključju in ni posledica podobnosti. Nižja kot je ta vrednost, manjša je verjetnost naključnega ujemanja med zaporedjema (Altschul in Koonin, 1998; Orengo in sod. 2003).

Obstaja več različic algoritmov BLAST. Najpogosteje sta uporabljana algoritem BLASTP, ki se uporablja za primerjavo proteinskega zaporedja z bazo proteinskih zaporedij in BLASTN, ki se uporablja za primerjavo nukleotidnega zaporedja z bazo nukleotidnih zaporedij. Zelo uporabna različica je BLASTX, pri kateri vnesemo nukleotidno zaporedje, ki se nato v vseh šestih bralnih okvirjih prevede v proteinsko zaporedje in ta zaporedja primerja z bazami proteinskih zaporedij. Algoritem TBLASTX uporabljamo za primerjavo prevedenega DNA zaporedja s prevedenimi zaporedji iz baz nukleotidnih zaporedij (Dale in von Schantz, 2007).

2.10.2 Primerjava več zaporedij

BLAST in FASTA sta najbolj uporabljani iskalni metodi, ki sta nekakšen kompromis med hitrostjo metode in iskanjem optimalne poravnave. Metodi sta v primerjavi z ostalimi metodami hitrejši, a zato poravnava ni nujno optimalna, poleg tega pa lahko poravnava le dve zaporedji naenkrat. Za optimalno poravnavo in poravnavo več zaporedij naenkrat uporabljamo zato druge programe. Najpogosteje v ta namen uporabljamo programe iz družine CLUSTAL (Dale in von Schantz, 2007).

Poravnave več proteinskih zaporedij so pomembno orodje v študiju zaporedij. Z njimi poiščemo konzervativne regije v zaporedju. To nam pomaga pri načrtovanju nadaljnjih poskusov testiranja in spreminjanja funkcije specifičnega proteina, v predvidevanju funkcije in strukture tega proteina in pri identifikaciji novih članov proteinskih družin (EMBL-EBI. 2009a).

Ugotovili so, da si v veliko primerih med seboj sorodni proteini, ki kodirajo homologne superdružine in izvirajo iz skupnega prednika, delijo tudi veliko ohranjenih mest. Ker je homologija celotnega proteina večkrat slab pokazatelj sorodnosti proteinov, za primerjavo uporabljamo konzervativna mesta, ki so značilna za vse člane proteinske družine. Poravnave več zaporedij nam tako omogočajo odkrivanje vzorcev homologije, kot tudi iskanje mest, ki so pomembna za stabilnost proteina in so zato v zaporedju bolj ohranjena (Orengo in sod. 2003).

2.10.2.1 Program CLUSTAL

Programi iz te družine so eni prvih in najpogosteje uporabljenimi programi za analizo zaporedij. Prvi CLUSTAL program sta napisala Higgins in Sharp že leta 1988 (Higgins in Sharp, 1988). Narejen je bil za uporabo na osebnih računalnikih. Skozi leta so program

izboljševali in mu dodajali nove funkcije, tako da je danes CLUSTAL najbolj popularen program za poravnavo zaporedij.

Najbolj uporabljana oblika programa Clustal je strežniški program ClustalW2, ki ga najdemo na spletni strani Evropskega bioinformatičnega inštituta (EBI). ClustalW2 je npr. veliko naprednejši, saj omogoča spreminjanje velikega števila nastavitvev, ima grafični vmesnik in več novih funkcij (Larkin in sod., 2007; Chenna in sod., 2003).

Program ClustalW2 se uporablja za dva namena:

- Simultana poravnava več zaporedij

Za poravnavo zaporedij s ClustalW2 je potrebno vnesti zaporedja v pravilnem formatu v obrazec na spletni strani ali naložiti datoteko z zapisanimi zaporedji. Pri tem je možno tudi spremeniti privzete nastavitve, da si algoritem prilagodimo svojim potrebam. Program nato zaporedja poravna in nam rezultat vrne na dva načina. Rezultat se lahko objavi na spletni strani ali nam ga program pošlje na naš e-poštni račun. Rezultat ostane na strežniku 24 ur, v primeru velikih datotek pa veliko manj. Program poravnava tako nukleotidna kot proteinska zaporedja, pri čimer je potrebno paziti, da zaporedja niso zapisana v komplementarni smeri (EMBL-EBI, 2009a).

Simultano poravnava večih zaporedij program prične s primerjanjem vseh možnih parov zaporedij, da izdela matrico točk parnih poravnav. To lahko stori neposredno iz rezultata, ki ga dobimo z BLASTP. Dve najbolj podobni zaporedji nato poravnava med seboj in tvori konsenzno zaporedje. Vsako od preostalih zaporedij se potem v vsakem krogu glede na podobnost poravnava s konsenznim, tako da se naredi novo konsenzno zaporedje. Ta postopek poteka, dokler ne dobimo poravnave vseh vključenih zaporedij. Seveda se tudi pri teh poravnavah upošteva medsebojno zamenljivost aminokislin in možnost tvorjenja vrzeli (Dale in von Schantz, 2007).

- Risanje filogenetskega drevesa

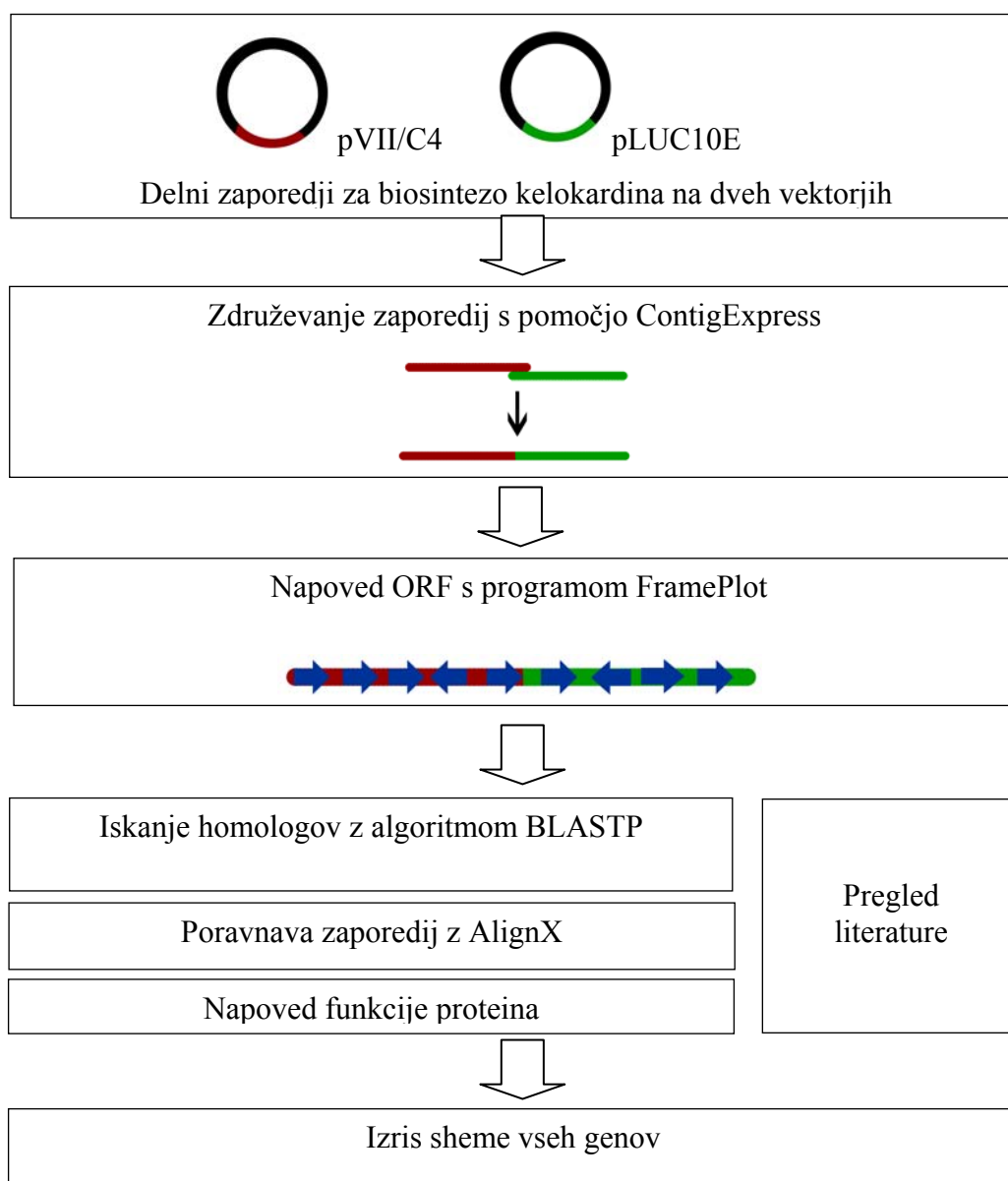
Če hočemo narisati filogenetsko drevo, moramo vnesti zaporedje v obliki poravnanih zaporedij. Nato imamo možnost izbire metode za izris filogenetskega drevesa. Izbiramo lahko med več metodami: sosedsko-povezovalno metodo (angl. Neighbour - Joining), Pyhilip ali Dist. Program nato odvisno od količine poravnanih zaporedij in izbrane metode čez nekaj časa vrne narisano drevo (EMBL-EBI, 2009a).

3 MATERIALI IN METODE

Izhajajoč iz ciljev diplomskega dela, ki načrtujejo določitev odprtih bralnih okvirjev v neobdelanem nukleotidnem zaporedju v elektronski obliki pri *Amycolatopsis sulphurei* (Lešnik, 2009), smo uporabili naslednje bioinformacijske metode dela:

- Za združevanje prekrivajočih zaporedij smo uporabili modul ContigExpress, ki je del programa Vector NTI (Invitrogen Corporation, 2004).
- Za iskanje odprtih bralnih okvirjev smo uporabili program FramePlot 4.0 beta (Ishikawa in Hotta, 1999).
- Primerna mesta za vezavo ribosoma smo iskali v programu Vector NTI (Invitrogen Corporation, 2004).
- Za iskanje motivov smo uporabljali orodja programa Vector NTI in prikaz ohranjenih domen na strani NCBI (Invitrogen Corporation, 2004; NCBI, 2009d).
- Zaporedja smo poravnali s AlignX, ki je modul programa Vector NTI in temelji na programu Clustal (Invitrogen Corporation, 2004).
- Proteinom smo predvideli funkcije na podlagi rezultatov analize BLAST s programom BLASTP, dosegljivim na spletni strani NCBI (NCBI, 2009a).
- Shemo genov smo narisali v programu Vector NTI (Invitrogen Corporation, 2004).

3.1 SHEMATSKI PRIKAZ POTEKA DIPLOMSKEGA DELA



Slika 11: Shematski prikaz poteka diplomskega dela

Za analizo zaporedja, ki kodira gene za biosintezo kelokardina, smo najprej delno prekrivajoči se zaporedji iz vektorjev VIIIC4 in pLUC10E (Lešnik, 2009) združili v eno zaporedje s pomočjo ContigExpress. Nato smo s programom FramePlot 4.0 predvideli vse kodirajoče odprte bralne okvirje. Te smo nato z algoritmom BLASTP primerjali z zaporedji v bazah podatkov (NCBI) in ugotavljali ohranjena zaporedja, iz dobljenih rezultatov pa nato ob pomoči literature poskusili predvideti funkcijo vsakega gena. Na koncu smo narisali shemo vseh genov, za katere predvidevamo, da sodelujejo pri biosintezi kelokardina v *Amycolatopsis sulphurea*.

3.2 ZDRUŽEVANJE SOSESK

Za združevanje sosesk (angl. contig) smo uporabili program Vector NTI oz. njegov modul ContigExpress, ki uporablja CAP3 program za združevanje zaporedij (Huang in Madan, 1999) in omogoča enostavnejše delo. Za sestavljanje sosesk v eno je pogoj, da se soseske, ki jih želimo združiti, na nekem delu prekrivajo (Invitrogen Corporation, 2004). ContigExpress nam je glede na prekrivajoče regije zaporedji združil in izpisal novo, sestavljeno zaporedje. Sestavljeno zaporedja nam je program prikazal tudi grafično.

S programom ContigExpress smo združili delni nukleotidni zaporedji za biosintezo kelokardina s plazmidov pVIIC4 in pLUC10E, ki ju je v okviru doktorskega disertacije pridobila dr. Urška Lešnik (Lešnik, 2009). Plazmid pLUC10E vsebuje 8 kb insert z nekaterimi geni za biosintezo kelokardina. Tudi plazmid pVIIC4 vsebuje gene za biosintezo kelokardina in je velik 35,6 kb. Zaporedji insertov teh dveh plazmidov se na odseku okoli KS α prekrivata, na podlagi česar smo zaporedji lahko združili in dobili 37,5 kb dolgo zaporedje, kateremu smo v nadaljevanju določili predvidene odprte bralne okvirje, katere smo v nadaljevanju analizirali in določili gene, za katere smo predvideli sodelovanje pri biosintezi kelokardina.

3.3 DOLOČANJE VERJETNIH ODPRTIH BRALNIH OKVIRJEV S PROGRAMOM FRAMEPLOT

Za pomoč pri iskanju odprtih bralnih okvirjev smo uporabili program FramePlot 4.0 beta (Ishikawa in Hotta, 1999), saj zaradi visokega G+C razmerja v genomu *Amycolatopsis sulphurea* z njim uspešno določamo predvidene odprte bralne okvirje.

Znano je, da kodoni, ki se med seboj razlikujejo le v nukleotidu na tretjem mestu, velikokrat kodirajo enako aminokislino. Da bakterije dosežejo visok odstotek G+C genoma, morajo na tretjem mestu kodona A ali T zamenjati s G ali C, če le ta sprememba ne vpliva na spremembo aminokislino, ki jo ta kodon kodira. Tretje mesto v kodonu je namreč najbolj variabilno, saj sprememba na tem delu večinoma ne vpliva na proteinsko zaporedje, ki ga nukleotidno zaporedje kodira. Za bakterije z visokim deležem G+C je zato značilen izredno visok delež G+C na tretjem mestu kodona, s katerim ohranjajo visok nivo G+C tudi v kodirajočih predelih nukleotidnega zaporedja.

FramePlot izračuna odstotek G+C na tretjem mestu za vsak kodon v setu, ki ga določimo pod parametrom »velikost okna« (angl. window size). »Velikost okna« je parameter, ki nam pove, koliko kodonov levo in desno upošteva program pri izračunu povprečne G+C vrednosti. Izračunan odstotek G+C program nato izriše na sredi »okna« in medtem, ko se po zaporedju premika z določenim korakom (angl. step size), postopek ponavlja. Če ima nek bralni okvir visok odstotek G+C v kodonu na tretjem mestu, izstopa od preostalih grafov na predelu, kjer predvidoma kodira odprti bralni okvir, zaradi česar je določanje predvidenih bralnih okvirjev lažje. (Ishikawa in Hotta, 1999).

Za določanje potencialnih odprtih bralnih okvirjev smo v okence za vnos zaporedja vnesli zaporedje v .fasta formatu zapisa in uporabili privzete nastavitve, čeprav program ponuja prilagoditev grafičnega prikaza z različnimi parametri. Slikovni prikaz, ki ga dobimo po vnosu DNA zaporedja, omogoča enostavno ločevanje regij, ki kodirajo proteine, od

nekodirajočih regij. Za vsak bralni okvir je graf narisani v drugačni barvi, opcijsko tudi v različnem stilu črte (glej npr. Slika 13). Potencialni ORF je ponavadi enostavno videti iz narisane grafa, saj se ponavadi nahaja na predelu, kjer je graf enega bralnega okvirja bistveno višje od preostalih dveh. Za ORF, ki ga izberemo, nam program izpiše nukleotidno zaporedje in domnevno aminokislinsko zaporedje. Ta zaporedja lahko neposredno iz programa primerjamo z bazo podatkov na NCBI (The National Center for Biotechnology Information), ki se je v zadnjih 20-ih letih skokovito razširila (Hopwood, 1999). Vsakemu stop-kodonu v potencialnem bralnem okvirju ponavadi program najde več start-kodonov, zato se moramo odločiti katerega izmed njih bomo izbrali. Najbolj primeren potencialni ORF iščemo tako, da za vsak start-kodon pridobimo proteinsko zaporedje, nato pa to zaporedje pošljemo v primerjavo na strežnike NCBI. Glede na BLASTP rezultate in primerjavo grafov, ki jih izriše program FramePlot, se nato odločimo med katerim start-kodonom in stop-kodonom je zaporedje najprimernejše za potencialni ORF.

FEATURES

3rd-letter G+C, 90.3%; Base composition, 214 A 417 C 476 G 192 T

20249 <Prev Next> 116

>f2_(86..1381) 432 aa 31-Aug-2009 with RBS
VTGPSDAHRVVTIGIGVAPGDRGTKQFMELITAGRTATRPISLFDASSFRSRVAAECNF
DPIAAGLSQRQIRKMDRITQFCVVAAREAVADSGMLGEQDPLRTGVAIGTACGMTQSLDR
EYAVVSDGSSWLVDPYGVPLQYDYFLPSSMATEIAMLVEAEGPVLVSTGCTSGVDVI
GHAADLIRDGEADINVAGASEAPISPIIVACFDAIKATTARNHEPESASRPFQTRS GFV
LGEAAAVFVLEELRHAKRRGAHIYAEIVGYASRCNAYSMTGLRPGREMADAIDGALNQA
RIDPSSRIGYVNAHGSSTRQNDRHETAAIKTSLGAHAYQVPVSSIKSMVGHSLGAIGSLEV
AACALTIHESVIPPTANLHVDPPECDLDYVPLVAREQEVDPVLSVASGFGGFQSAILLTG
PDGRTGKRVTOR

Search with against format low complexity filter

>-120
TGCCCGCCAACAGCCGGTGGGACCTACCCGGTG

>f2_(86..1384) 1299bp 31-Aug-2009 with RBS
GTGACAGGACCAAGCGATGCCCATCGTGTCTGTGATTACCGGGATTGGGGTGGTCGCCCGG
GGCGACCCGTGTACCAAGCAGTTCTGGGAGCTGATCACGGCGGGCCGTACGGCCACCCGG

Slika 12: Primer rezultata, ki ga prikaže program FramePlot

3.4 BLAST IN PSI BLAST – ISKANJE PODOBNIH ZAPOREDIJ

V programu FramePlot (Ishikawa in Hotta, 1999) smo izbrano zaporedje poslali v BLASTP analizo na strežnike NCBI. Pri tem se je našemu zahtevku dodelila identifikacijska številka (ID) in nas je preusmerilo na spletno stran z vmesnikom, s katerim smo lahko nastavili nekatere parametre za prikaz rezultatov, kot so npr. tip prikaza rezultatov, št. prikazanih rezultatov, omejitev prikaza rezultatov z določenimi E-vrednostmi idr. Imamo pa tudi možnost, da izberemo prikaz rezultatov v obliki za PSI-BLAST. Večinoma je bilo tudi tukaj dovolj, da smo sprejeli prednastavljene vrednosti. S pritiskom na gumb za prikaz poročila (angl. View report) program zažene analizo in nam jo čez nekaj časa prikaže.

Rezultati so v privzeti nastavitvi predstavljeni na več načinov. Najprej je v tabeli prikazan opis zaporedja, preiskanih baz in uporabljenega algoritma. Pod njim sledi grafični prikaz

nekaterih najbolj verjetnih ohranjenih domen. Če hočemo o teh ohranjenih domenah izvedeti več, jih izberemo in program nas preusmeri na (pod)stran CDD (CDD; angl., NCBI Conserved Domain Search), kjer so te domene bolje predstavljene. Pod grafičnim prikazom ohranjenih domen, aktivnih mest ipd. so nato grafično prikazani najbližji zadetki, torej zaporedja, za katere algoritem izračuna, da so najbolj podobna iskanemu. Iz grafičnega prikaza je viden del, v katerem sta si zaporedji podobni, poleg tega pa se že iz barve popisa najbližjih homologov da razbrati, ali ima iskano zaporedje sebi podobna zaporedja ali ne. Rdeča barva prikazuje rezultate, ki so najbolj točkovani in predstavljajo zelo podobna zaporedja, medtem ko je njeno nasprotje črna barva, ki kaže zelo slabo ujemajoča zaporedja.

Pod grafičnim prikazom so prikazani zadetki tudi v obliki tabele. V njej so rezultati razporejeni glede na rezultat, ki jih algoritem izračuna glede na stopnjo podobnosti. Rezultati v tabeli so vsi predstavljeni na enak način. Najprej je zapisana gi številka, nato oznaka baze, dostopna številka in ime lokusa, kjer zaporedje leži. Med sabo so ločeni z navpično črto. Ta zapis služi tudi kot povezava na proteinsko bazo, kjer je shranjen opis tega zaporedja. Temu sledi okrajšan opis zaporedja, naslednji je napisan rezultat poravnave v bitih in kot zadnja je napisana E-vrednost. Število točk je zapisano v obliki hiperpovezave, ki nas prenese po strani navzdol, kjer so vsa prikazana zaporedja prikazana še enkrat v daljši obliki in prikazano poravnavo z iskanim zaporedjem (Madden, 2003).

Na koncu strani so vsi rezultati prikazani tudi v obliki poravnave in njenega neokrajšanega opisa. Daljša oblika zapisa vsebuje celoten opis zaporedja, ki mu sledi tudi organizem, v katerem je bilo zaporedje najdeno. Temu nato sledi izpisana dolžina tega zaporedja v aminokislinah. Za tem sledi opis poravnave. Poravnavo najprej opišejo točke v bitih in E-vrednost. Temu sledijo parametri identičnost (angl. Identity), »pozitivni« (angl. Positives) in vrzeli (angl. Gaps). Identičnost nam pove število identičnih aminokislin v zaporedjih, parameter pozitivni nam pove število ohranjenih substitucij in parameter vrzeli nam pove število vseh vrzeli v poravnavi. Na koncu sledi poravnava opisane zaporedja z iskanim (Madden, 2003).

Zaporedja smo iskali v proteinski bazi nr, ki združuje naslednje baze: non-redundant GenBank CDS translations, PDB, SwissProt, PIR in PRF bazo, in je najobsežnejša baza proteinskih zaporedij (NCBI, 2009a). Ta baza ima veliko zaporedij avtomatsko anotiranih, kar znižuje kvaliteto in zanesljivost zadetkov. Iskanje samo v bazah z zaporedji, ki jih obdelajo znanstveniki in so eksperimentalno dokazana, ni primerno, saj večina genov, ki se uvrščajo med najboljše zadetke, ni shranjenih v teh bazah.

Pri uporabi algoritmov BLAST je potrebno tudi nekaj pazljivosti. V zameno za večjo hitrost delovanja je namreč algoritem žrtvoval nekaj natančnosti, zato ni nujno, da so poravnave, ki jih prikaže BLAST, v resnici tudi najboljše. Za optimalno poravnavo, ali za poravnavo več zaporedij naenkrat, je priporočljivo uporabiti drug program, npr. CLUSTAL (Higgins in Sharp, 1988). Zavedati se moramo, da popis zadetkov, ki jih vrne BLAST ni nujno zelo pravilen, saj se večkrat zgodi, da je ob drugih nastavitvah ali ob uporabi drugih programov ta popis nekoliko drugačen (Dale in von Schantz, 2007).

Poleg BLASTP smo za vsako zaporedje naredili tudi analizo PSI-BLAST in primerjali njene rezultate z rezultati BLASTP. S tem smo preverjali ali so najbolj homologni proteini tudi v najbolj ohranjenih delih zaporedja podobni iskanemu proteinskemu zaporedju.

3.4.1 Iskanje ohranjenih domen

Že zgoraj je bilo omenjeno, da se med rezultati BLASTP analize lahko pojavi tudi slikovni prikaz ohranjenih domen, s izbiro katerega nas preusmeri na (pod)stran z ohranjenimi domenami (CDD). Ohranjene domene definiramo kot ločene funkcionalne in / ali strukturne enote beljakovin. Te vsebujejo vzorce oz. motive, ki jih je program našel v zaporedju. NCBI-jeva baza ohranjenih domen vsebuje preko 12 000 vnosov in je povezana z ostalimi bazami iz skupine Entrez (NCBI, 2009d). Na strani CDD najdemo tudi orodja za delo z ohranjenimi domenami, s katerimi lahko npr. primerjamo domene med sabo, najdemo aktivna mesta, preverimo njihovo predvideno funkcijo idr. Podobno kot na pri vseh programih na NCBI strani, tudi tukaj poteka nenehen razvoj, zato se orodja neprestano izboljšujejo in se uvajajo nova. Za uporabo vseh funkcij je treba zato veliko spremljati razvoj teh orodij.

3.5 PORAVNAVA VEČ ZAPOREDIJ

Poravnave več zaporedij so eno glavnih orodij v molekularni biologiji. Poravnave uporabljamo za iskanje značilnih vzorcev, določanje proteinskih družin, kot tudi za detektiranje in prikazovanje podobnosti novih z že poznanimi zaporedji (Dale in von Schantz, 2007; Larkin in sod., 2007).

S poravnami več zaporedij smo preverjali, na katerih odsekih se je naše preiskovano zaporedje ujemalo s sorodnimi, in na ta način predpostavili, kateri del predvidenega odprtega bralnega okvirja še verjetno kodira predvideni protein. S poravnami smo si pomagali tudi pri iskanju značilnih oz. ohranjenih zaporedij.

V namen poravnave zaporedij smo uporabljali AlignX, ki je eden od modulov programa Vector NTI (Invitrogen Corporation, 2004). Zaporedja smo si v AlignX uvozili direktno iz spletne strani programa BLASTP.

AlignX omogoča veliko uporabnih funkcij in hitro delo. Zaporedja smo si v AlignX uvozili direktno iz strani programa BLASTP. Na strani namreč najdemo opcijo za shranjevanje zaporedja, obkljukamo izbrana zaporedja in zeleno obliko zapisa zaporedja shranimo (večinoma smo shranjevali v datoteke formata .gb, ki poleg zaporedja shrani tudi druge informacije, pridobljene z anotacijo). Shranjevanje zaporedja program Vector zazna in nam ponudi možnost shranjevanja v njegovo bazo.

Z AlignX zaporedja uvozimo in jih poravnamo. Program nam poravnavo prikaže grafično, pod njo pa izpiše konsenzno zaporedje. Med uporabnimi možnostmi programa je tudi opcija enostavne izključitve zelenega zaporedja iz poravnave, ki smo jo tudi večkrat uporabili. Poravnana zaporedja je mogoče tudi ročno popravljati, če ugotovimo kako napako, prav tako pa AlignX omogoča tudi enostavno kopiranje poravnave v druge programe v obliki besedila ali slike.

Iz poravnave zaporedij se da že na prvi pogled ugotoviti, v katerem delu so si zaporedja med seboj bolj in v katerem manj podobna in katera zaporedja so bolj podobna med sabo. Medsebojno podobnost najenostavneje prikažemo z dendrogramom, ki nam ga AlignX izriše.

3.6 SHEMATSKI PRIKAZ GENSKE SKUPINE

Shemo genov smo narisali s programom Vector NTI. V njem lahko grafično poleg predvidenih genov predstavimo tudi veliko drugih lastnosti, npr. restrikcijska mesta, predvidena mesta vezave ribosomov, ohranjena zaporedja, aktivna mesta in drugo. Prav tako program grafično prikaže te lastnosti za zaporedja, ki jih prenesemo iz baz podatkov in imajo te lastnosti shranjene v svojem opisu.

4 REZULTATI

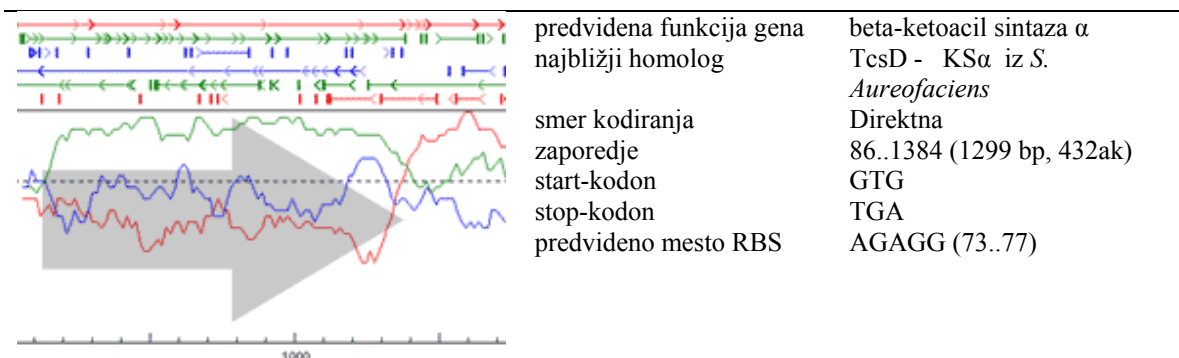
Cilj diplomskega dela je bila analiza *in silico* vseh genov v preiskovanem zaporedju, za katere smo predvideli sodelovanje pri biosintezi antibiotika kelokardina. V nadaljevanju je vsak izmed teh genov tudi podrobneje opisan, in sicer je za vsak gen prikazan izsek iz grafičnega prikaza FramePlot analize, predstavljena je predvidena funkcija njegovega genskega produkta in najbližji proteinski homolog le-tega. Poleg tega je za vsak genski produkt predstavljena tudi preglednica najbolj podobnih proteinskih zaporedij, pridobljena z analizo BLASTP, ki nakazuje predvideno funkcijo vsakega genskega produkta in stopnjo podobnosti z ostalimi proteini s podobno predvideno funkcijo. V primeru, da je za genski produkt značilna tudi ohranjena domena ali motiv, ki še dodatno potrjuje predvideno funkcijo, je predstavljen/a tudi ta. Za vsak gen oz. genski produkt so rezultati analize *in silico* povzeti in interpretirani tudi v pisni obliki.

Na koncu tega poglavja je grafično prikazana shema nukleotidnega zaporedja z vsemi geni, ki so opisani v nadaljevanju in preglednica z osnovnimi podatki za vse gene v zaporedju, za katerega predvidevamo, da kodira vse potrebne gene za biosintezo kelokardina.

Diplomsko delo je bilo opravljeno na Katedri za biotehnologijo, mikrobiologijo in varnost živil (tedaj imenovano Katedra za biotehnologijo) Oddelka za živilstvo, in sicer v času od 23.11.2007 do 26.3.2008.

4.1 OPISI GENOV

4.1.1 ChdP



Slika 13: Opis predvidenega genskega produkta ChdP. Na sliki so prikazane tri krivulje, ki vsaka predstavlja odstotek G+C na tretjem mestu v kodonu za določeno velikost okna za vse tri možne bralne okvirje. Nad krivuljami so za vsak bralni okvir v enaki barvi v obe možni smeri predstavljeni potencialni start-kodoni (puščica v smeri prepisovanja) in stop-kodoni (navpična črta). Visok odstotek G+C na tretjem mestu v kodonu za bralni okvir v zeleni barvi lepo nakazuje potencialni odprti bralni okvir v smeri proti desni, ne pa tudi proti levi, saj se na tem odseku v levo smer pojavlja preveč stop-kodonov, ki preprečujejo, da bi bil genski produkt kodiran na celotnem delu zaporedja, na katerem bralni okvir izstopa od preostalih dveh.

Preglednica 2: Seznam najbližjih homologov genskega produkta ChdP, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>S. aureofaciens</i> BAB12566	427	0.0	77% / 86%	<i>tcsD</i>	ketosintaza	klortetraciklin
<i>S. rimosus</i> AAK61718	425	0.0	77% / 85%	<i>oxyA</i>	ketosintaza α	oksitetraciklin
<i>S. aureofaciens</i> BAA92278	392	3e-177	77% / 86%	/	ketosintaza	/
<i>S. platensis</i> BAA92279	392	3e-175	77% / 85%	<i>Ks</i>	ketosintaza	
<i>S. spiramyceticus</i> AAV48829	419	2e-171	70% / 80%	<i>aspA</i>	Ketoacil sintaza tipa II	spiramicin

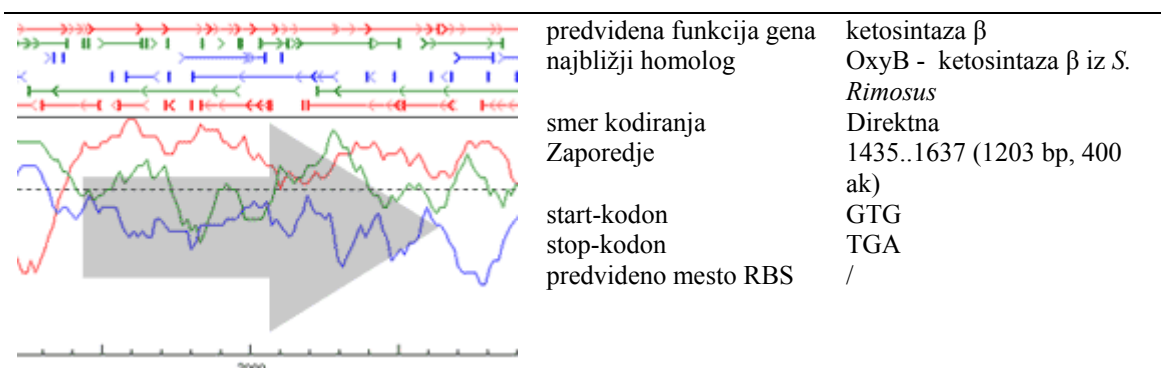
Predvideni gen *chdP* najverjetneje kodira KS α , saj med zaporedji z visoko stopnjo identičnosti najdemo večinoma gene za KS α (Slika 13). Rezultati analize BLASTP so zelo zanesljivi, saj za prvih 50 zadetkov E-vrednost ne pade pod e-130. Podobno je tudi s stopnjo identičnosti, ki je pri 50 najbolj homolognih proteinih vedno nad 60%.

Najboljše ujemanje je s proteinskima zaporedjema, ki ju kodirata gena *tscD* in *oxyA*. E-vrednost je v obeh primerih 0. TscD iz *S. aureofaciens* in OxyA iz *S. rimosus* (Zhang in sod., 2007) sta oba ketosintazi, ki sodelujeta pri biosintezi dveh podobnih tetraciklinskih antibiotikov, klortetraciklina in oksitetraciklina.

Zaporedje CmmP (Menendez in sod., 2004) je po homologiji s ChdP na osmem mestu. Gre za KS α , ki sodeluje v biosintezi kromomicina, ki je kemijsko podoben kelokardinu, zato genski homolog iz skupine genov za biosintezo kromomicina pogosto kažejo visoko stopnjo homologije z predvidenim ORF iz genske skupine za biosintezo kelokardina.

Kot začetek bralnega okvirja za *chdP* bi lahko potencialno označili dve poziciji z enakim DNA zaporedjem pred start kodonom. Na podlagi analize BLASTP in poravnavo zaporedij smo ugotovili, da se daljši bralni okvir dobro ujema s KS α iz *S. aureofaciens* (BAB12566) in *S. rimosus* (AAK61718), zato predvidevamo, da ima odprti bralni okvir za ChdP dolžino 432 aminokislin.

4.1.2 ChdK



Slika 14: Opis predvidenega genskega produkta ChdK

Preglednica 3: Seznam najbližjih homologov genskega produkta ChdK, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>S. rimosus</i> AAZ78324	422	5e-156	73% / 84%	<i>oxyB</i>	CLF	oksitetraciklin
<i>S. aureofaciens</i> BAB12567	425	2e-151	71% / 82%	<i>tcsE</i>	beta-ketoacil sintaza β	klortetraciklin
<i>S. antibioticus</i> BAA92278	405	9e-142	64% / 78%	<i>orf1</i>	keto-acil sintaza β	Oviedomicin
<i>Micromonospora</i> sp. ACP19354	407	6e-141	64% / 77%	<i>saqB</i>	ketosintaza β	galtamicin B
<i>S. fradiae</i> CAA60570	408	3e-139	62% / 77%	<i>urdB</i>	CLF	Urdamicin

Gen *chdK* je dolg 1203 bp in verjetno kodira KS β . Analiza BLASTP zaporedja ChdK (nam kot najboljši rezultat vrne proteinsko zaporedje OxyB iz *Streptomyces rimosus*). Tudi naslednji zadetki so proteini, katerih predvidena funkcija je KS β . Podobno kot pri ChdP so tudi za ChdK rezultati BLASTP zelo verjetni, saj je E-vrednost zelo nizka, stopnja podobnosti pa pri prvih zadetkih ne pade pod 60%. Med najbolj homolognimi zaporedji se pojavljajo zaporedja iz istih skupin genov, kot so se pojavile pri analizi BLASTP predvidenega proteina ChdP, ki kodira predvideni genski produkt KS α , kar smo tudi pričakovali.

4.1.2.1 Stopnja podobnosti med ChdP (KS α) in ChdK (KS β)

BLASTP tako za genske produkte ChdP kot za ChdK prikaže med ohranjenimi aminokislinskimi zaporedji dve domeni z visoko stopnjo podobnosti, značilni za poliketid sintaze tipa II, KS α in KS β . Glavna razlika med KS α in KS β je v aktivnem mestu z ohranjenim cisteinom, značilnim za genski homolog KS α , ki katalizira kondenzacijo intermediatov ketoacil-ACP. Pri homologu KS β pa je namesto ohranjene regije s cisteinom tu značilna regija z ohranjenim glutaminom, za katerega predvidevajo vlogo pri dekarboksilaciji KS β (Bisang. in sod., 1999). Za zanesljivo določitev genov smo zato ti značilni regiji tudi identificirali (Slika 15 in Slika 16).

		168	180
BAA92278	(153)	GVVSTGCTSGIDV	
BAB12566	(168)	GVVSTGCTSGIDV	
BAA92279	(153)	GVVSAGCTSGIDV	
P43678	(167)	GVVSAGCTSGIDV	
ChdP	(167)	GLVSTGCTSGVDV	

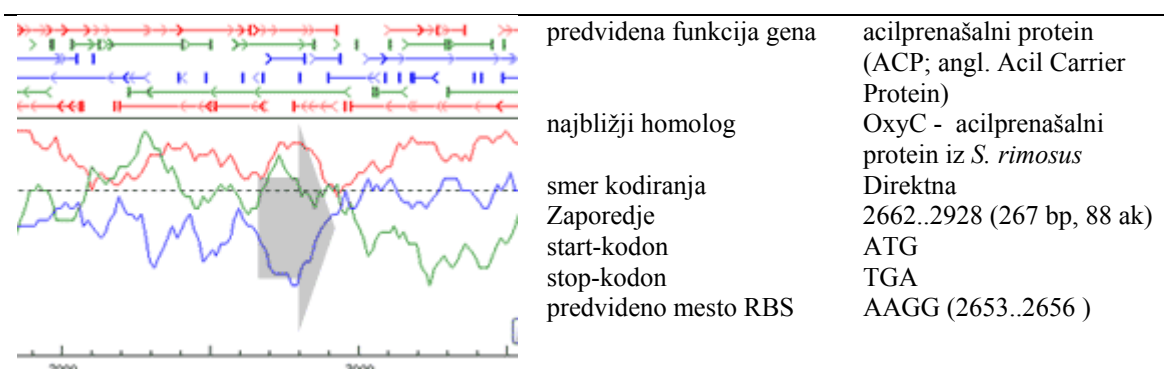
Slika 15: Poravnava najbolj podobnih zaporedij s ChdP, v kateri je označen cistein (C¹⁷⁴), ki služi kot aktivno mesto (Meurer and Hutchinson, 1995). Vidna je tudi ohranjenost regije okoli cisteina.

V zaporedju ChdK smo identificirali ohranjeno regijo z ohranjenim glutaminom (Q), ki pri KS β zamenjuje aktivno mesto s cisteinom (Slika 16). (Bisang. in sod., 1999).

		176	189
AAD13537	(159)	VVSDQAGGLDAIAQ	
BAB12567	(157)	LVAEQAGGLDALAK	
AAZ78326	(175)	LVAEQAGGLDAVAK	
ChdK	(171)	LVSEQAGGLDAVAK	
CAA60570	(162)	VVSDQAGGLDAVAQ	
CAG14966	(162)	VVSEQAGGLDALAQ	

Slika 16: Ohranjen glutamin (Q¹⁸⁰), ki je značilen za KS β (Bisang in sod., 1999).

4.1.3 ChdS



Slika 17: Opis predvidenega genskega produkta ChdS

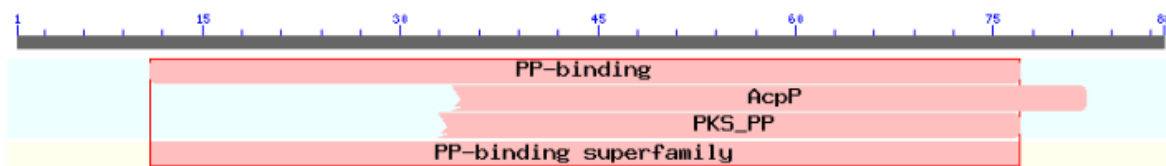
Preglednica 4: Seznam najbližjih homologov genskega produkta ChdS, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>S. rimosus</i> AAZ78324	95	4e-18	61% / 77%	<i>oxyC</i>	ACP	Oksitetraciklin
<i>S. echinatus</i> ABL09957	88	2e-17	58% / 75%	<i>orf9</i>	ACP	Aranciamicin
<i>S. venezuelae</i> AAB36564	89	2e-16	53% / 73%	<i>acp</i>	ACP	Jadomicin
<i>S. arenae</i> CAA60570	86	1e-15	52% / 66%	<i>ncnC</i>	ACP	naptociklinon (angl. naphthocyclinone)
<i>S. fradiae</i> AAD13537	89	1e-15	55% / 75%	<i>urdC</i>	ACP	Urdomicin

Predvideni gen *chdS* najverjetneje kodira acilprenašalni protein (ACP). Z dolžino 88 aminokislin (Slika 17) ga lahko štejemo v skupino manjših proteinov, kar se sklada z velikostjo drugih ACP. E-vrednost je zaradi krajšega zaporedja nekoliko višja, kar pa še ne pomeni, da so dobljeni rezultati manj zanesljivi. Tudi tukaj se med prvimi zadetki analize BLASTP pojavljajo zaporedja iz podobnih skupin genov, kot se pojavljajo v genskih skupinah za biosintezo spojin s pomočjo PKS tipa II. To smo tudi pričakovali, saj je ACP poleg KS α in KS β tretji protein, ki ga uvrščamo v t.i. proteine minimalnega PKS, torej spada med esencialne encime, ki katalizirajo biosintezo naraščajoče poliketidne verige.

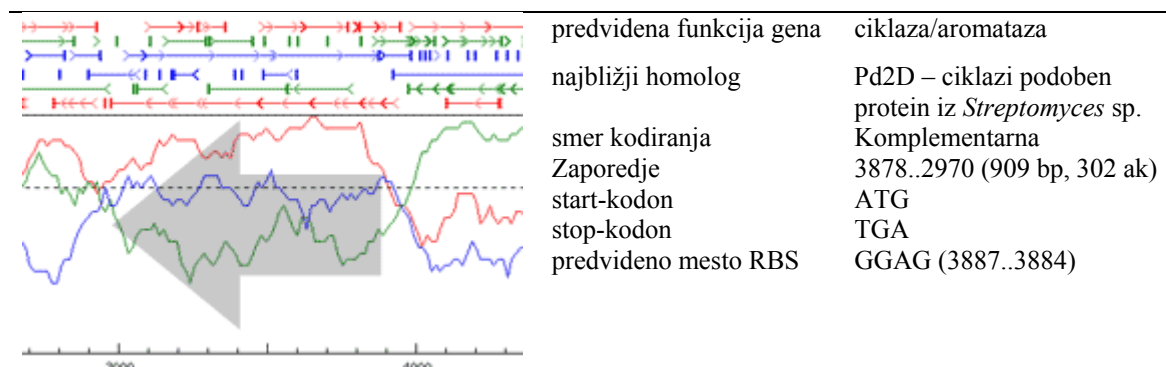
Za večino genov, ki kodirajo minimalni PKS tipa II, je značilno, da je ACP postavljen takoj za KS α in KS β , prepisovanje ACP pa poteka v isto smer kot prepisovanje KS α in KS β , čeprav obstajajo tudi izjeme (Ichinose in sod., 2003).

V zaporedju najdemo tudi ohranjeno domeno AcpP (COG0236 (NCBI, 2009c),) ki je značilna za acilprenašalne proteine (Slika 18).



Slika 18: V super-družino fosfopantein vezajočih proteinov uvrščamo tudi družino AcpP, ki združuje acilprenašalne proteine.

4.1.4 ChdQI



Slika 19: Opis predvidenega genskega produkta ChdQI

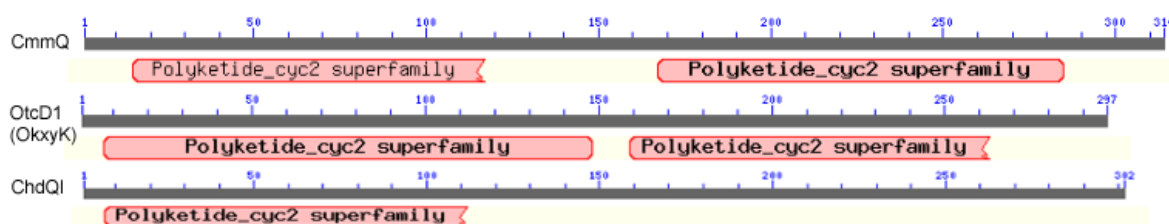
Preglednica 5: Seznam najbližjih homologov genskega produkta ChdQI, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces</i> sp. AAO65366	314	3e-41	33% / 52%	<i>pd2D</i>	JadI ciklazi podoben/protein	
<i>Salinispora tropica</i> YP_001159026	319	1e-37	33% / 51%	lokus Strop_2197	ciklaza/dehidraza	/
<i>Micromonospora</i> sp. YP_289283	314	4e-37	32% / 49%	<i>saqL</i>	ciklaza	galtamicin
<i>Thermobifida fusca</i> YP_289283	314	9e-37	32% / 49%	lokus Tfu_1222	bifunkcionalna ciklaza/dehidrataza	aktinorodin
<i>Catenulispora acidiphila</i> ZP_04371219	316	3e-36	31% / 49%	lokus CaciDRAFT_27720	poliketidna ciklaza/aromataza	in
<i>S. sviveus</i> YP_002207117	306	5e-36	33% / 51%	lokus SSEG_02721	lipidni transport ciklaza	/
<i>S. griseus</i> subsp. <i>griseus</i> CAE17552	318	1e-35	32% / 48%	<i>cmmQ</i>	aromataza	kromomicin

Predvideni gen *chdQI* najverjetneje kodira ciklazo/aromatazo, saj se glede na analizo BLASTP kot najbolj homologni proteini pojavljajo ciklaze, aromataze, dehidraze in

njihove kombinacije (Preglednica 5). Za genski produkt ChdQI smo predvideli enako funkcijo kot jo ima CmmQ, ki služi kot aromataza pri biosintezi kromomicina in predvidoma sodeluje pri zapiranju prvega obroča (Menendez in sod, 2004). Zadetki analize BLASTP zaporedja ChdQI imajo nižji odstotek identičnosti in podobnosti kot zaporedja iz minimalnega PKS. Kljub temu je E-vrednost dovolj visoka, da rezultatom lahko zaupamo. Med prvimi zadetki so večinoma zaporedja, ki pripadajo skupini t.i. »non-redundant« bazo podatkov, ki je avtomatsko anotirana, kar pomeni, da ti rezultati niso tako zanesljivi, saj funkcijo tem zaporedjem določijo le na podlagi analiz *in silico* in rezultatov ne preverijo v laboratoriju (Apwiler in sod., 2001).

Med najbolj homolognimi proteini z ChdQI najdemo tudi bifunkcionalno aromatazo/ciklazo OtcD1(OxyK) iz *S. rimosus*, za katero so s prekinitvijo gena dokazali velik vpliv na pravilno biosintezo oksitetraciklina, najverjetneje pa sodeluje pri nastanku prvega obroča. Glede na podobnost s CmmQI in OtcD1 bi tudi za ChdQI predvidevali, da bo sestavljena iz dveh domen iz super-družine značilne za poliketidne ciklaze (cl110449 (NCBI, 2009c)), kar je značilno tudi za CmmQI in OtcD1, prikaz ohranjenih domen (Slika 20) pa tega ne potrjuje, čeprav se zaporedje ChdQI z OtcD1 in CmmQ ujema v celotni dolžini. Tako ne moremo z gotovostjo trditi, da gre res za dvodomensko aromatazo/ciklazo, predvidimo pa lahko njeno sodelovanje pri ciklizaciji prvega obroča.



Slika 20: Primerjava ohranjenih domen proteinov CmmQI (AAO65366), OtcD1 (AAD10031) in ChdQI.

4.1.5 ChdMII



Slika 21: Opis predvidenega genskega produkta ChdMII

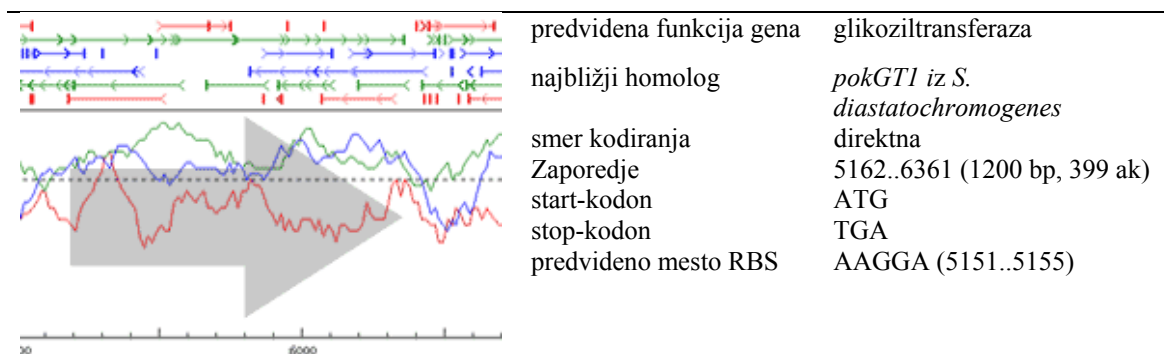
Preglednica 6: Seznam najbližjih homologov genskega produkta ChdMII, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>S. griseus</i> subsp. <i>griseus</i> CAE17532	342	2e-81	44% / 61%	<i>cmmMII</i>	C-metiltransferaza	kromomicin
<i>S. argillaceus</i> CAK50769	326	5e-72	45% / 64%	<i>orf9</i>	C-metiltransferaza	mitramicin
<i>Chthoniobacter flavus</i> ZP_03133286	333	3e-45	36% / 51%	lokus CfE428DRAF T_6454	metiltransferaza tipa 12	/
<i>Verrucomicrobium spinosum</i> ZP_02928000	342	6e-45	33% / 49%	lokus VspiD_010100 015160	od SAM odvisna metiltransferaza	/
<i>Nostoc punctiforme</i> ACC79025	325	1e-40	30% / 48%	lokus Npun_R0239	metiltransferaza tipa 12	/

Predvideni gen *chdMII* verjetno kodira metiltransferazo. Odprti bralni okvir je velik 1032 nukleotidov, ki kodirajo 343 aminokislin dolgo proteinsko zaporedje. Šest nukleotidov pred start-kodonom se začne zaporedje AGGA, ki verjetno služi kot mesto vezave ribosoma (Slika 21).

Analiza BLASTP (Preglednica 6) vrne kot najbolj ujemajoč rezultat C-metiltransferazo (*CmmMII*) iz *Streptomyces griseus* subsp. *griseus*. Identičnost je 44%, E-vrednost je nizka (2^{-81}). Naslednji vrnjeni rezultat je spet C-metiltransferaza (*MtmMII*) iz *S. argillaceus*, ki sodeluje pri biosintezi mitramicina (Lombo in sod., 2000). Za obe metiltransferazi je značilno, da kodirata metiltransferazi, ki metilirata 9. C-atom kromomicina oz. mitramicina (Menendez in sod., 2004; Lombo in sod., 2000). Med rezultati analize BLASTP ni genskih homologov za biosintezo oksitetraciklina (OTC) in klortetraciklina (CTC), saj strukturi OTC in CTC na 9. C-atomu nimata metilne skupine (Zhang, 2006), kar dodatno potrjuje, da gre verjetno za metiltransferazo, ki metilira 9. C-atom štiričlenskega skeleta kelokardina.

4.1.6 ChdGIV



Slika 22: Opis predvidenega genskega produkta ChdGIV

Preglednica 7: Seznam najbližjih homologov genskega produkta *ChdGIV*, pridobljen z analizo BLASTP

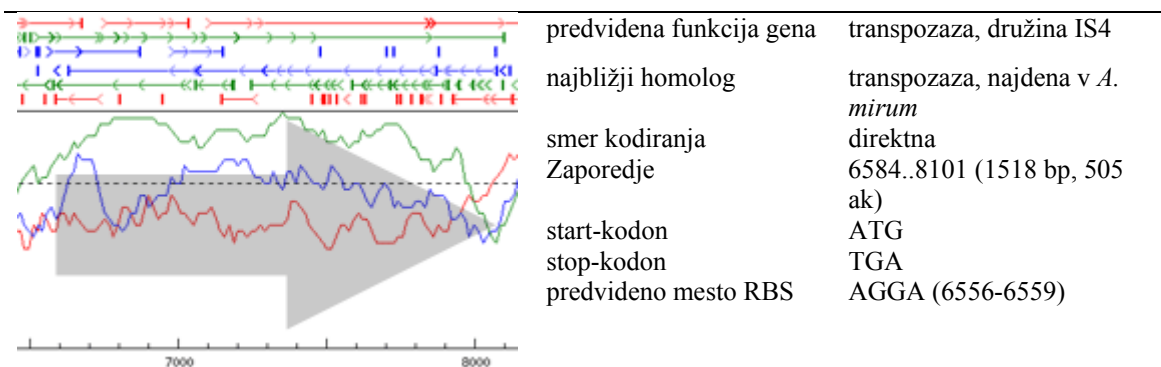
Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces diastatochromogenes</i> ACN64823	397	1e-59	41% / 51%	<i>pokGT1</i>	glikoziltransferaza	poliketomicin
<i>S. griseus</i> subsp. <i>griseus</i> CAE17547	421	1e-58	38% / 51%	<i>cmmGIV</i>	glikoziltransferaza	kromomicin
<i>Streptomyces argillaceus</i> CAK50788	407	9e-53	38% / 50%	<i>mtmGIV</i>	glikoziltransferaza	mitramicin
<i>Actinomadura melliaura</i> ZP_02928000	400	3e-38	35% / 45%	<i>atG1</i>	O-glikoziltransferaza	/
<i>S. antibioticus</i> CAA05642	426	5e-37	31% / 41%	<i>oleG2</i>	glikoziltransferaza	oleandomicin

Za gen *chdGIV* predvidevamo, da kodira protein s funkcijo glikoziltransferaze (Slika 22). Glede na kemijsko strukturo kelokardina, ki ne vsebuje glikozidnih komponent, v skupini genov, ki kodira biosintezo kelokardina nismo pričakovali glikoziltransferaze. Kljub temu rezultati analize BLASTP nakazujejo, da gen verjetno kodira glikoziltransferazo, saj najbolj homologna zaporedja kodirajo glikoziltransferaze (Preglednica 7).

Pred genom *chdGIV* smo našli tudi močno ohranjeno Shine-Dalgarno zaporedje, kar nakazuje, da se glikoziltransferaza verjetno prepisuje sama, saj je naslednji gen v smeri njenega prepisovanja metiltransferaza, ki pa se prepisuje v nasprotno smer (Slika 44).

Rezultati kažejo, da bi glikoziltransferaza lahko bila aktiven encim, vendar ni jasno kakšna bi naj bila njena vloga.

4.1.7 ChdTn



Slika 23: Opis predvidenega genskega produkta ChdTn

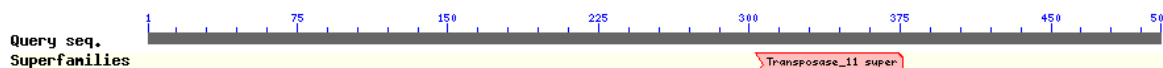
Preglednica 8: Seznam najbližjih homologov genskega produkta ChdTn, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Actinosynnema mirum</i> ZP_03815003	406	4e-52	37% / 51%	lokus AmirDRAFT_06280	transpozaza, družina IS4	/
<i>Salinispora arenicola</i> YP_001537907	451	1e-48	33% / 48%	lokus Sare_3106	transpozaza, družina IS4	/
<i>Frankia</i> sp. YP_001507737	421	2e-46	38% / 51%	lokus Franean1_3429	transpozaza, družina IS4	/
<i>Mycobacterium gilvum</i> YP_001133458	401	2e-46	34% / 47%	lokus Mflv_2192	transpozaza, družina IS4	/
<i>Mycobacterium gilvum</i> ABP45137	401	3e-46	34% / 47%	lokus Mflv_2660	transpozaza, družina IS4	/

Predvideni gen *chdTn* z dolžino 1518 bp verjetno kodira transpozazo iz družine IS4 (Slika 23). Pred zaporedjem se večkrat ponovi zaporedje GGTTG, ki ga FramePlot označi kot možno mesto vezave ribosoma. Poleg tega pa na mejni razdalji 25 nukleotidov nad zaporedjem leži AGGA zaporedje, ki bi prav tako lahko služilo kot mesto vezave ribosoma (Slika 23).

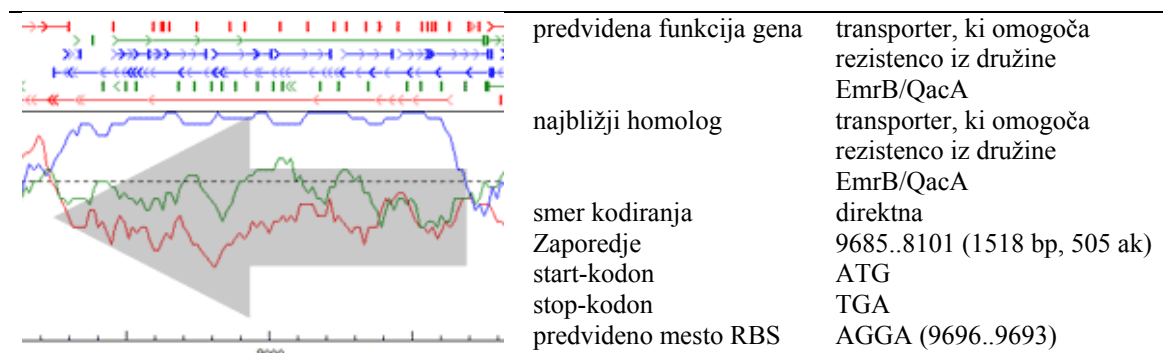
Najbolj podobna proteinska zaporedja so bila najdena v različnih bakterijah in najverjetneje niso vezana na biosintezo poliketidov, se pa transpozaze pogosto nahajajo v skupinah genov, ki kodirajo biosintezo sekundarnih metabolitov, kar povezujejo s prenosom celotnih genskih skupih, tudi s pomočjo horizontalnega prenosa (Palmeaner in sod. 2008). Glede na to, da so vsi najboljši zadetki transpozaze iz družine IS4 in je E-vrednost dovolj nizka, lahko z veliko verjetnostjo trdimo, da je tudi *chdTn* gen za transpozazo družine IS4.

BLASTP prikaže tudi ohranjeno domeno, značilno za super-družino DDE transpozaz (Slika 24). Za to super-družino je značilen ohranjen motiv DDE, z značilnima aspartatom (D) in glutamatom (E) (Rezsohazy in sod., 1993), ki bi naj bili odgovorni za pravilno delovanje kovinskih ionov, potrebnih za kataliziranje reakcije rezanja DNA in njenega prenosa na specifično lokacijo. V super-družino DDE spadajo IS4, IS5, IS421, IS5377 in podobne transpozaze (pfam01609 (NCBI, 2009c)).



Slika 24: Ohranjena domena znotraj ChdTn, ki spada med domene DDE transpozaz (pfam01609 (NCBI, 2009c)).

4.1.8 ChdR



Slika 25: Opis predvidenega genskega produkta ChdR

Preglednica 9: Seznam najbližjih homologov genskega produkta ChdR, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Thermomonospora curvata</i> ZP_04032177	510	3e-87	37% / 57%	lokus TcurDRAFT_32230	transporter, ki omogoča rezistenco, iz družine EmrB/QacA	/
<i>Frankia sp.</i> YP_481117	523	6e-87	39% / 58%	lokus Francci3_2014	transporter, ki omogoča rezistenco, iz družine EmrB/QacA	/
<i>Nakamurella multipartita</i> ZP_04346236	541	1e-85	36% / 51%	lokus NamuDRAFT_03150	transporter, ki omogoča rezistenco, iz družine EmrB/QacA	/
<i>Klebsiella pneumoniae</i> ACI11465	480	2e-83	40% / 59%	lokus KPK_3283	transporter, ki omogoča rezistenco, H ⁺ antiporter-1 (DHA2)	/
Candidatus <i>Methanosphaerula palustris</i> YP_002465120	479	2e-83	34% / 60%	lokus Mpal_0874	transporter, ki omogoča rezistenco, iz družine EmrB/QacA	/

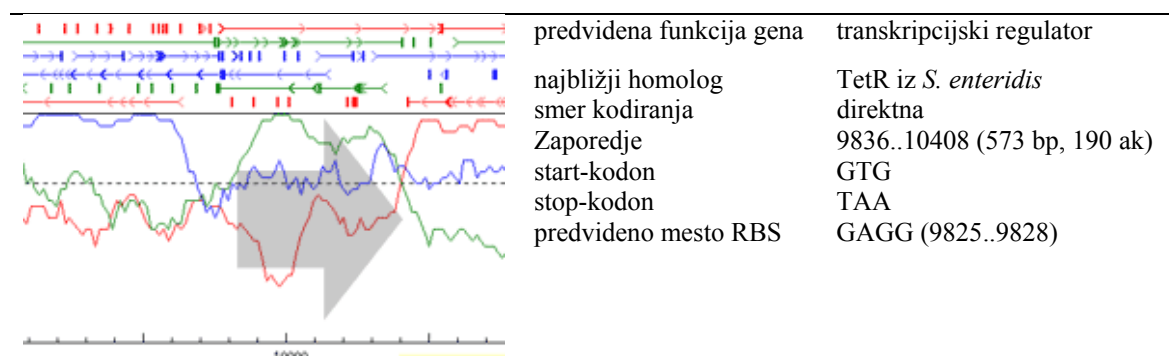
Predvideni gen *chdR* verjetno kodira transporter, odgovoren za rezistenco bakterije na njen produkt. Pred njegovim odprtim bralnim okvirjem smo našli ohranjeno zaporedje za vezavo ribosoma (Slika 25).

Glede na dolžino zaporedja homolognih zaporedij težko ocenimo, kakšna bi naj bila dolžina ChdR, saj se podobna zaporedja po dolžini med seboj kar precej razlikujejo. Tudi podobnost proteina ChdR z njemu najbolj homolognimi je v primerjavi z ostalimi geni iz njegove skupine genov nekoliko nižja (Preglednica 9).

Med organizmi, kjer so našli najbolj homologne proteine, spadajo bakterije iz rodu *Frankia*, *Streptomyces*, *Nocardia* in *Salinospora*, pri katerih pogosto najdemo tudi druge homologe, sorodne proteinom za biosintezo kelokardina (Preglednica 9).

Transporterji iz poddružine EmrB/QacA (TIGR00711 (NCBI, 2009c)) pripadajo skupini transporterjev, ki iz bakterije črpajo molekule z zanjo toksičnim delovanjem (angl. efflux transporters). Predvidoma imajo 14 transmembranskih regij (TIGR00711 (NCBI, 2009c)).

4.1.9 ChdA



Slika 26 Opis predvidenega genskega produkta ChdA

Preglednica 10: Seznam najbližjih homologov genskega produkta ChdA, pridobljen z analizo BLASTP.

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Salmonella enteritidis</i> AAN40999	217	8e-31	44% / 57%	<i>tetR</i>	okrnjen represor tetraciklinske rezistence	/
<i>Klebsiella pneumoniae</i> ABS19067	213	2e-30	41% / 55%	<i>tetR</i>	represor tetraciklinske rezistence	/
<i>Burkholderia cenocepacia</i> YP_620166	210	5e-30	50% / 66%	lokus Bcen_0279	transkripcijski regulator družine TetR	/
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Kentucky ACH85842	216	5e-30	41% / 51%	<i>tetR</i>	represor tetraciklinske rezistence	/
<i>Burkholderia ambifaria</i> ZP_02889404	211	2e-29	49% / 64%	lokus BamIOP4010 DRAFT_1466	transkripcijski regulator družine TetR	/

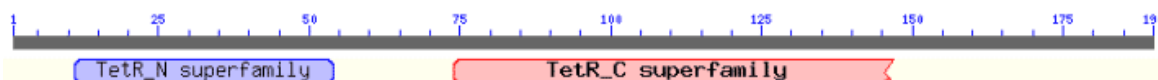
* Med rezultati se večkrat pojavi gen za TetR iz *B. cenocepacia*, ki je v preglednico zapisan le enkrat.

Gen *chdA* najverjetneje kodira transkripcijski regulator iz družine TetR z velikostjo 190 aminokislin (Slika 26). TetR proteini so represorji, ki jim pripisujejo regulacijo biosinteze antibiotikov, črpalk za iznos iz celice (angl. efflux pump), različnih katabolnih poti, osmotskega stresa idr. (Ramos in sod., 2005).

Analiza BLASTP je pokazala, da so genskemu produktu ChdA najbolj podobni proteini najdeni v različnih bakterijah. Med temi so se največ pojavljale bakterije iz rodu *Shigella*, pri kateri je bil najden tudi najbolj podoben homolog genskega produkta ChdA (identičnost 44%, podobnost 57%). Med bakterijami s z ChdA homolognimi proteinskimi produkti se pojavljajo tudi rodovi *Burkholderia*, *Shigella*, *Escherichia*, *Vibrio* idr. Stopnja

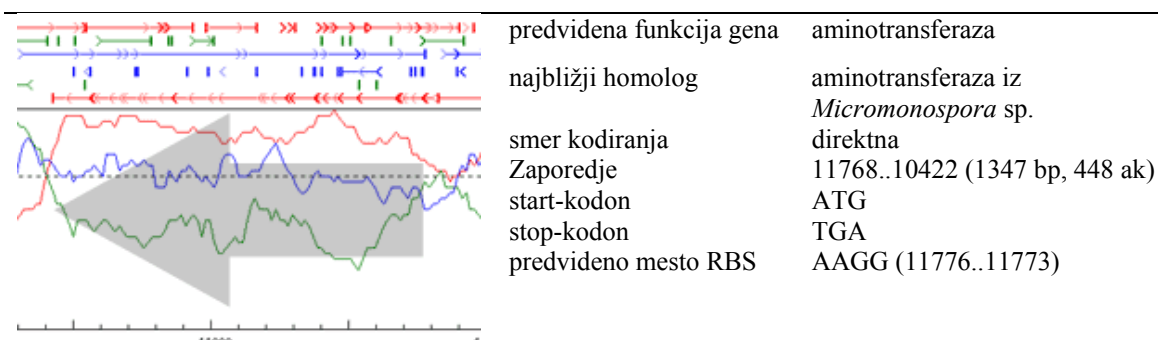
identičnosti/podobnosti je za te homologe navadno nad 30%/45%, kar nakazuje ohranjenost genov za transkripcijske regulatorje v teh bakterijah.

Na proteinskem zaporedju smo našli dve domeni: TetR in TetR_C, ki sta oba značilni za TetR družino represorjev (Slika 27).



Slika 27: Domene, značilni za TetR regulatorne proteine.

4.1.10 ChdN



Slika 28: Opis predvidenega genskega produkta ChdN

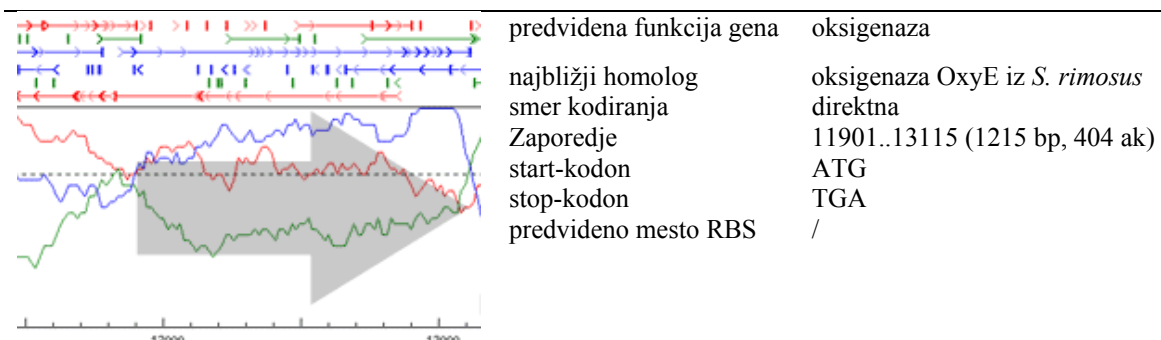
Preglednica 11: Seznam najbližjih homologov genskega produkta *ChdN*, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Micromonospora</i> sp. ZP_04604097	441	6e-118	53% / 67%	lokus MCAG_00354	aminotransferaza razreda / I in II	
<i>Catenulispora acidiphila</i> ZP_04376652	407	2e-117	54% / 67%	lokus CaciDRAFT_82 530	transkripcijski regulator / z HTH domeno in aminotransferazno domeno	
<i>Kribbella flavida</i> ZP_03859946	427	3e-116	52% / 66%	lokus KflaDRAFT_06 21	transkripcijski regulator / z HTH domeno in aminotransferazno domeno	
<i>Nocardia uniformis</i> subsp. <i>tsuyamanensis</i> AAT09795	431	2e-114	50% / 65%	<i>nocG</i>	p-hidroksifenilglicin transaminaza	nokardicin A
<i>Streptomyces coelicolor</i> A3(2) NP_733596	444	1e-113	52% / 63%	lokus SCO3227	aminotransferaza	/

Predvideni odprti bralni okvir, ki kodira gen *chdN*, ima dolžino 1347 nt in kodira 448 aminokislin dolg protein. Pred odprtim bralnim okvirjem je zaporedje AAGG, kamor se predvidoma veže ribosom (Slika 28). Gen *chdN* je kodiran na komplementarni verigi, torej v obratno smer kot sosednja genska homologa *chdA* in *chdOI* (Slika 44). Proteini, ki jih algoritem BLASTP ocenjuje kot najbolj podobne, imajo zelo nizko E-vrednost, obenem pa visoko stopnjo identičnosti, ki za prvih 10 zadetkov ne pade pod 50% (Preglednica 11).

Najbolj podoben homolog genskega produkta ChdN je bil najden v bakteriji iz rodu *Micromonospora* (53% identičnost, 67% podobnost). Med podobnimi zadetki se pojavi tudi genski produkt gena *nocG*, ki služi kot aminotransferaza v *Nocardia uniformis* subsp. *Tsuyamanensis*. Glede na amino skupino na 2. C-atomu kelokardina in homologijo z aminotransferazami predvidevamo, da gen *chdN* kodira genski produkt, ki sodeluje pri prenosu amino skupine na 2. C-atom kelokardina.

4.1.11 ChdOI



Slika 29: Opis predvidenega genskega produkta ChdOI

Preglednica 12: Seznam najbližjih homologov genskega produkta ChdOI, pridobljen z analizo BLASTP.

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces rimosus</i> AAZ78329	418	7e-141	66% / 78%	<i>oxyE</i>	oksigenaza	oksitetraciklin
<i>Streptomyces griseus</i> subsp. <i>griseus</i> CAE17524	415	5e-114	56% / 68%	<i>cmmOI</i>	oksigenaza	kromomicin
<i>Streptomyces diastatochromogenes</i> ACN64816	409	6e-110	54% / 68%	<i>pokOI</i>	monooksigenaza	poliketomicin
<i>Streptomyces argillaceus</i> CAK50769	436	2e-84	48% / 65%	<i>mtmOI</i>	oksigenaza	mitramicin
<i>Streptomyces pristinaespiralis</i> YP_002196477	359	3e-76	45% / 61%	lokus SSDG_05097	oksigenaza	/

Predvideni gen *chdOI* je dolg 1215 bp in najverjetneje kodira protein z oksigenazno aktivnostjo. Pred genom nismo našli ohranjenega mesta za vezavo ribosoma.

Rezultati analize BLASTP (Preglednica 12) zaporedja ChdOI imajo nizko E-vrednost. Najbolj podobno zaporedje je OxyE, ki kodira oksigenazo v *Streptomyces rimosus* (Zhang in sod. 2006). Stopnji identičnosti in podobnosti aminokislinskega zaporedja OxyE in ChdOI sta zelo visoki in sta za 10% višji kot pri drugem zadetku. Med rezultati se poleg OxyE pojavljajo tudi oksigenaze, ki sodelujejo pri biosintezi podobnih tetraciklinov, kot so kromomicin (CmmOI), poliketomicin (PokOI) in mitramicin (MtmOI), ki so tudi ročno anotirane. Natančne funkcije genskega produkta ChdOI pri biosintezi kelokardina z primerjavo z najbolj homolognimi proteini nismo določili, saj najbližji homolog OxyE verjetno oksidira 5. C-atom oksitetraciklina (Zhang in sod. 2006), ki pri molekuli kelokardina sploh ni oksidiran, med zadetki pa se visoko pojavi tudi oksigenaza MtmOI iz

S. argillaceus, za katero so s prekinitvijo gena dokazali, da ni nujno potrebna za biosintezo (Lombo in sod, 1999).

4.1.12 ChdMI



Slika 30: Opis predvidenega genskega produkta ChdMI

Preglednica 13: Seznam najbližjih homologov genskega produkta ChdMII, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces rimosus</i> AAZ78330	345	4e-111	65% / 76%	<i>oxyF</i>	C6-metiltransferaza	oksitetraciklin
<i>Streptomyces sviveus</i> EDY57712	342	1e-108	60% / 73%	lokus SSEG_04292	O-metiltransferaza	/
<i>Catenulispora acidiphila</i> ZP_04375151	352	1e-107	59% / 74%	lokus CaciDRAFT_67410	O-metiltransferaza	/
<i>Streptomyces coelicolor</i> NP_631500	342	7e-107	59% / 73%	lokus SCO7452	O-metiltransferaza	/
<i>Streptomyces aureofaciens</i> CAJ89661	342	2e-106	59% / 73%	lokus SAML0675	O-metiltransferaza	/

Predvideni odprti bralni okvir *cmmMI* kodira 341 aminokislin dolgo proteinsko zaporedje. Pred njim smo našli zaporedje AGGT, ki ga program FramePlot označi kot predvideno mesto vezave ribosoma (Slika 30).

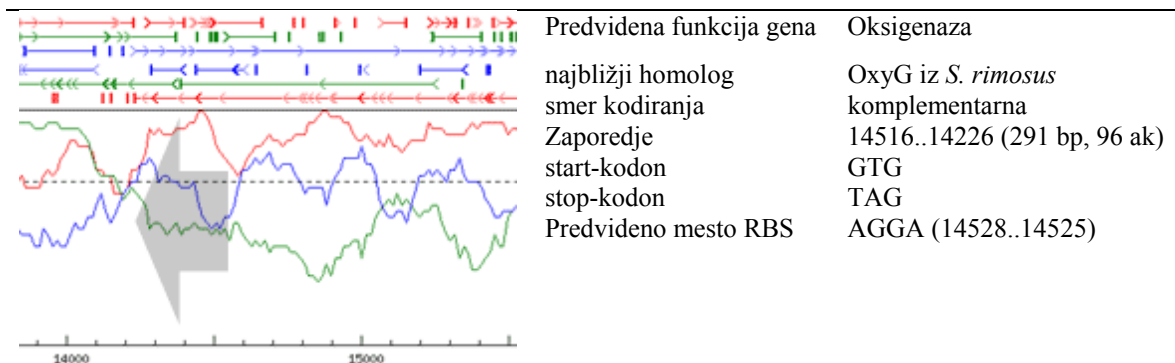
Prvih 7 najboljših zadetkov (genskih homologov) BLASTP analize ima E-vrednost zelo nizko. Algoritem BLASTP kot najboljši zadetek vrne C-metiltransferazo OxyF iz *S. rimosus*, medtem ko naslednji zadetki predvidoma kodirajo O-metiltransferaze. Med zadetki se kot vir najpogosteje pojavljajo bakterije iz rodov *Streptomyces*, med ostalimi zadetki pa prednjačita *Burkholderia* in *Mycobacterium*.

BLASTP pokaže tudi ohranjeno domeno (Slika 31), značilno za O-metiltransferaze (pfam00891 (NCBI, 2009c)). Ker v kelokardinu ne najdemo metilne skupine, vezane preko molekule kisika, in ker je nablížji homolog proteina ChdMI metiltransferaza iz oksitetraciklina, predvidevamo, da gre za metiltransferazo, ki metilira 6. C-atom molekule kelokardina.



Slika 31: Ohranjena domena (pfam00891), ki jo prikaže BLASTP, je značilna za O-metiltransferaze.

4.1.13 ChdOIII



Slika 32: Opis predvidenega genskega produkta ChdOIII

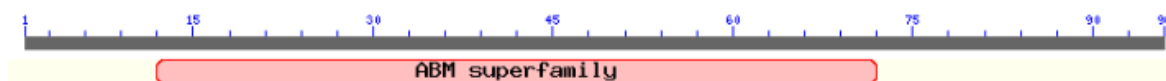
Preglednica 14: Seznam najbližjih homologov genskega produkta ChdOIII, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces rimosus</i> AAZ78331	95	6e-25	61% / 76%	<i>oxyG</i>	oksisgenaza	oksitetraciklin
<i>Streptomyces diastatochromogenes</i> ACN64849	109	1e-12	44% / 63%	<i>Pok03</i>	monooksigenaza	poliketomicin
<i>Streptomyces pristinaespiralis</i> EDY67139	103	1e-11	42% / 59%	lokus SSDG_05475	oksisgenaza	/
<i>Streptomyces argillaceus</i> CAK50778	99	2e-09	36% / 52%	<i>mtmOIII</i>	oksisgenaza	mitramicin
<i>Streptomyces chartreusis</i> CAH10167	106	5e-09	37% / 54%	<i>chaH</i>	oksisgenaza	karteuzin

Odprti bralni okvir *chdOIII* predvidoma kodira oksigenazo. Dolžina nukleotidnega zaporedja je 291 nt in kodira 96 aminokislin dolgo proteinsko zaporedje (Slika 32). Zaradi kratkega zaporedja je E-vrednost rezultatov nekoliko nižja kot običajno, kljub temu pa imajo rezultati dovolj visok odstotek identičnosti in podobnosti, da so lahko relevantni za vsaj prvih 15 zadetkov (Preglednica 14). Najboljše ujemanje z ChdOIII ima OxyG oksigenaza, ki jo najdemo v skupini genov za biosintezo oksitetraciklina v bakteriji *Streptomyces rimosus* (Zhang in sod. 2007, Zhang 2009). Tudi ostali proteini, ki so med najpogostejšimi zadetki imajo vlogo oksigenaze. Med najbolj podobnimi proteini najdemo tudi oksigenazo MtmOIII iz skupine genov za biosintezo mitramicina (Menendez in sod. 2004).

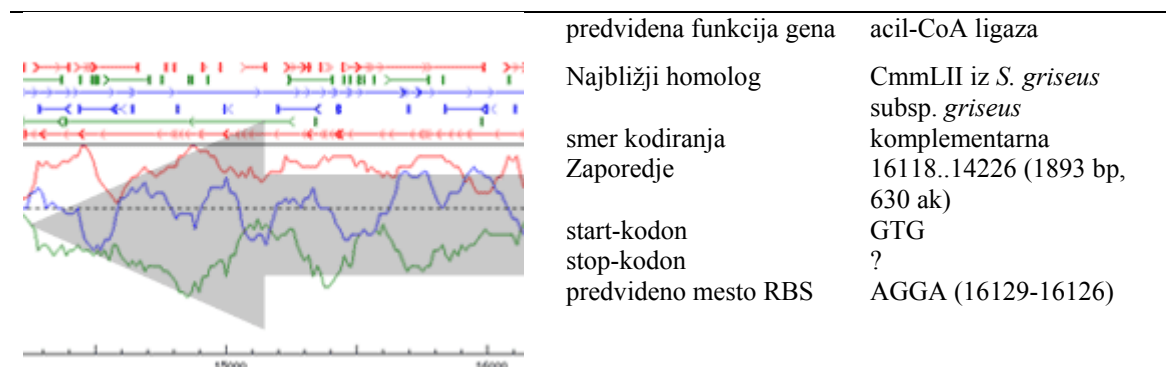
V zaporedju smo našli tudi ohranjeno domeno, ki spada v super-družino domen ABM (Slika 33), ki jih najdemo pri monooksigenazah, ki sodelujejo pri biosintezi antibiotikov (pfam03992 (NCBI, 2009c)). Glede na najdeno ohranjeno domeno lahko manj zanesljivo predvidevamo, da genski produkt ChdOIII spada med monooksigenaze, glede na najvišjo homologijo z OtcG pa lahko predvidevamo sodelovanje pri oksidaciji prvega obroča

kelokardina, saj tudi za protein OtcG predvidevajo sodelovanje pri oksidaciji prvega obroča oksitetraciklina (Zhang in sod., 2006).



Slika 33: Zaporedje ChdOIII z označeno značilno ohranjeno domeno ABM (pfam03992 (NCBI, 2009c)).

4.1.14 ChdL



Slika 34: Opis predvidenega genskega produkta ChdL

Preglednica 15: Seznam najbližjih homologov genskega produkta ChdL, pridobljen z analizo BLASTP

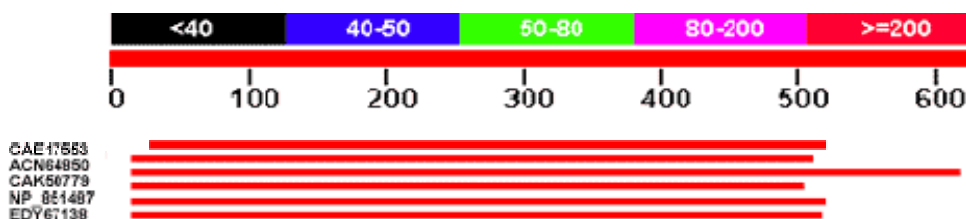
Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces griseus</i> subsp. <i>griseus</i> CAE17553	528	1e-117	47% / 62%	<i>cmmLII</i>	acil-CoA ligaza	kromomicin
<i>Streptomyces diastatochromogenes</i> ACN64850	495	2e-112	48% / 61%	<i>pokL</i>	acil-CoA ligaza	poliketomicin
<i>Streptomyces argillaceus</i> CAK50779	514	1e-95	47% / 58%	<i>mtmL</i>	acil-CoA ligaza	mitramicin
<i>Streptomyces rochei</i> NP_851487	536	2e-87	41% / 54%	lokus pSLA2-L_p065	acil-CoA ligaza	/
<i>Streptomyces pristinaespiralis</i> EDY67138	436	9e-76	43% / 56%	lokus SSDG_05474	acil-CoA ligaza	/

Predvideni gen *chdL* verjetno kodira acil-CoA ligazo. FramePlot nam kot možen odprti bralni okvir izpiše 1893 nt dolgo zaporedje, ki kodira 630 aminokislinsko dolgo proteinsko zaporedje (Slika 34). Pred start-kodonom najdemo zaporedje AGGA, ki verjetno služi za mesto vezave ribosoma. Rezultati BLASTP imajo nizko E-vrednost in nekoliko nižjo stopnjo podobnosti od preostalih genov v genski skupini. Najbolj podobno proteinsko zaporedje kodira CmmLII, ki ima 47% stopnjo podobnosti (Preglednica 15).

Grafični prikaz analize BLASTP (Slika 35) kaže, da se rezultati ujemajo le do dolžine približno 530 aminokislinsko (1590 nt), kar povzroči nekoliko nižji odstotek identičnosti in podobnosti (Slika 35). Večina zadetkov analize BLASTP so acil-CoA ligaze ali imajo njej

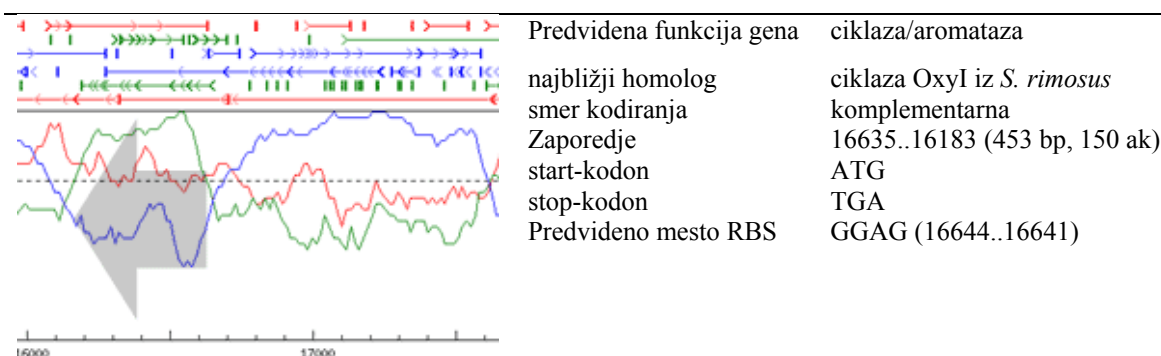
podobno funkcijo. Med temi zadetki najdemo acil-CoA ligaze iz različnih bakterijskih genomov, večinoma pa prevladujejo acil-CoA ligaze iz rodu *Streptomyces* (Preglednica 15).

Stop-kodon smo v nukleotidnem zaporedju identificirali šele pri dolžini 1893 nt (630 ak), glede na analizo BLASTP pa predvidevamo, da se stop-kodon nahaja že po približno 530 ak, kar se bolj sklada tudi z dolžinami homolognih acil-CoA ligaz, obenem pa razloži, zakaj sta ChdL in ChdOIII kodirana v istembralnem okvirju.



Slika 35: Grafični prikaz za nekaj zadetkov BLASTP analize zaporedja ChdL.

4.1.15 ChdX



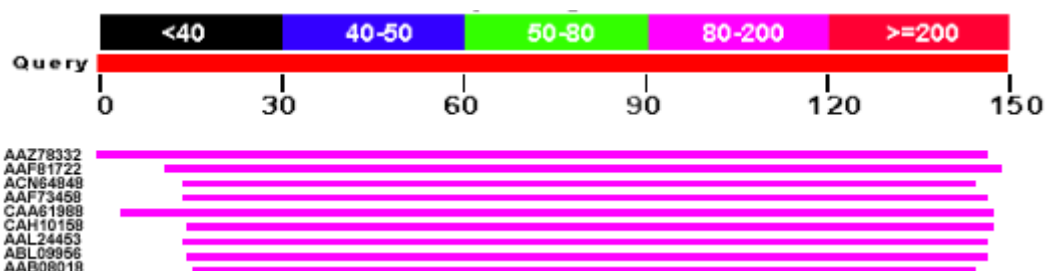
Slika 36: Opis predvidenega genskega produkta

Preglednica 16: Seznam najbližjih homologov genskega produkta ChdX, pridobljen z analizo BLASTP.

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces rimosus</i> AAZ78332	150	9e-40	63% / 74%	<i>oxyI</i>	ciklaza četrtega obroča	oksitetraciklin
<i>Streptomyces maritimus</i> AAF81722	163	1e-32	47% / 65%	<i>encG</i>	neznana	enterocin
<i>Streptomyces diastatochromogenes</i> ACN64848	150	1e-32	55% / 63%	<i>pokC1</i>	ciklaza	poliketomicin
<i>Streptomyces galilaeus</i> AAF73458	144	5e-31	55% / 64%	<i>aknV</i>	mogoče sodeluje pri glikozilaciji	aklarubicin
<i>Streptomyces argillaceus</i> CAA61988	150	2e-30	49% / 59%	<i>mtmX</i>	ciklaza	mitramicin

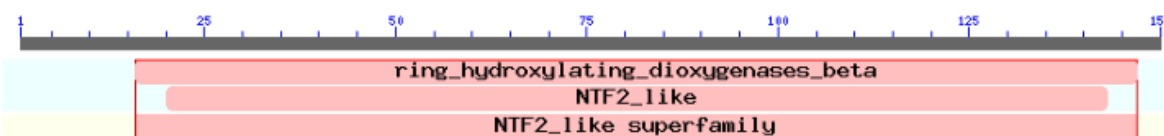
Za ChdX predvidevamo, da kodira protein s funkcijo ciklaze/aromataze in verjetno sodeluje pri ciklizaciji osnovne nascentne poliketidne verige kelokardina (Slika 36). Pred zaporedjem smo našli ohranjeno mesto za vezavo ribosoma (GGAG), kar smo tudi pričakovali, saj je sosednje zaporedje *chdT* kodirano na komplementarni verigi. Med

rezultati najdemo proteine z različnimi predvidenimi funkcijami. Predvsem so to ciklaze, aromataze in dioksigenaze, ki hidrolizirajo aromatske obročje (Preglednica 16).



Slika 37: grafični prikaz ujemanja z *ChdX*, kot ga prikaže BLASTP analiza.

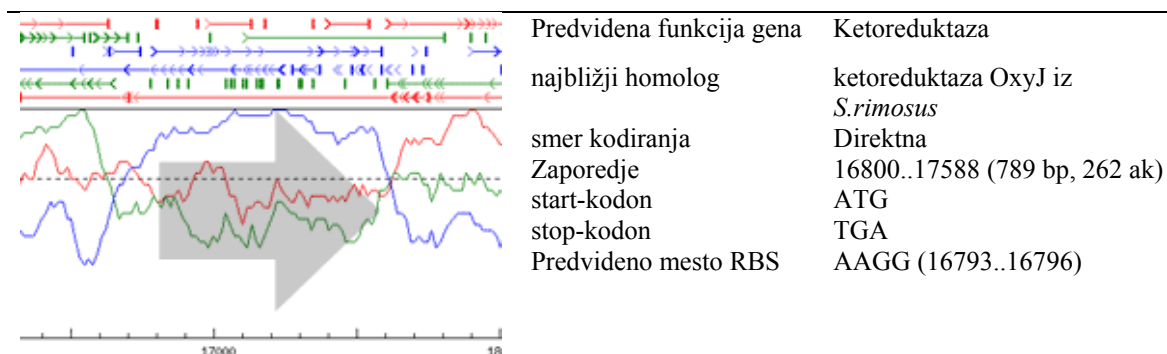
Analiza BLASTP prikaže tudi ohranjeno domeno (Slika 37 in Slika 38), značilno za monodomenske ciklaze (cd00667 (NCBI, 2009c)). Glede na podobnost z OxyJ (Zhang in sod., 2007) in MtmX (Lombo in sod., 1999), ki sta monodomenski ciklazi in verjetno sodelujeta pri zapiranju četrtega obroča, tudi za ChdX predvidevamo, da je monodomenska ciklaza in verjetno sodeluje pri zapiranju četrtega obroča.



Slika 38: Grafični prikaz ohranjenih domen prikaže tudi domeno, značilno za beta podenoto obroč hidroksilirajoče dioksigenaze (cd00667 (NCBI, 2009c)).

Iz dobljenih rezultatov težko določimo natančno biološko vlogo tega proteina, lahko pa na podlagi homologije z OxyJ (Zhang in sod., 2007) in MtmX (Menendez in sod., 2004) predvidevamo njegovo vlogo pri ciklizaciji obročev tetraciklina.

4.1.16 ChdT



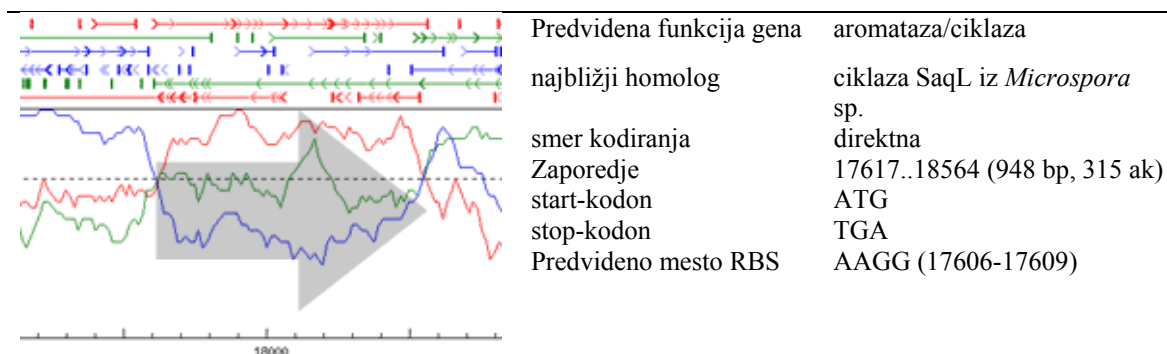
Slika 39: Opis predvidenega genskega produkta ChdT

Preglednica 17: Seznam najbližjih homologov genskega produkta ChdT, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces rimosus</i> AAZ78333	263	1e-108	76% / 85%	<i>oxyJ</i>	C9 ketoreduktaza	oksitetraciklin
<i>Streptomyces antibioticus</i> AAK06787	261	2e-102	71% / 83%	<i>simA6</i>	ketoreduktaza	simociklinon
<i>Streptomyces sp.</i> AAK57528	261	9e-102	71% / 81%	<i>pgaD</i>	ketoreduktaza	/
<i>Streptomyces antibioticus</i> CAG14968	260	4e-101	71% / 82%	<i>ovmT</i>	ketoreduktaza	oviedomicin
<i>Streptomyces galilaeus</i> BAB72043	261	1e-100	72% / 82%	<i>aknA</i>	ketoreduktaza	aklocinomicin

Gen *chdT* z dolžino 789 bp kodira 262 ak dolgo proteinsko zaporedje. Pred genom smo našli RBS mesto AGGA. Dolžina genskega produkta se ujema z dolžinami najbolj homolognih proteinskih zaporedij. Med zadetki je najvišje ketoreduktaza OxyJ iz *S.rimosus* (identičnost 76%, podobnost 85%), ki predvidoma reducira keto-skupino na 8. C-atomu tetraciklinske molekule (Zhang in sod., 2007), zato tudi predvidenemu proteinu ChdT pripisujemo to funkcijo. Tudi za ostale najbližje homologe je značilna visoka stopnja identičnosti/podobnosti z genskim produktom ChdT. Najbližji homologi večinoma kodirajo ketoreduktaze in so bili najdeni predvsem pri bakterijah iz rodu *Streptomyces*.

4.1.17 ChdQII

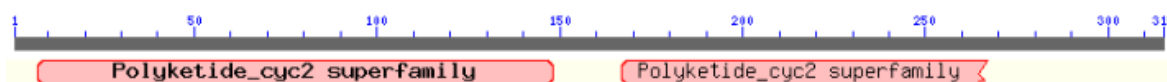


Slika 40: Opis predvidenega genskega produkta ChdQII

Preglednica 18: Seznam najbližjih homologov genskega produkta ChdQII, pridobljen z analizo BLASTP

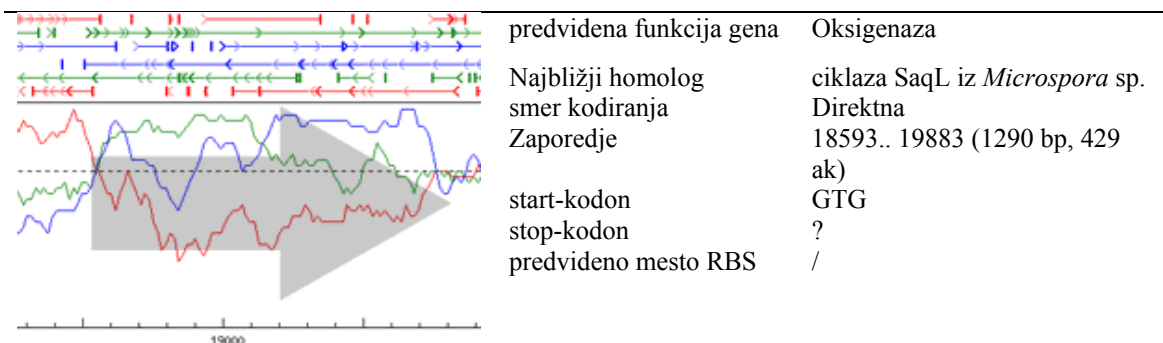
Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Micromonospora</i> sp. ACP19357	314	1e-88	54% / 64%	<i>saqL</i>	ciklaza	sakvajamicin
<i>Streptomyces rimosus</i> AAZ78334	317	8e-88	58% / 67%	<i>oxyK</i>	aromataza	oksitetraciklin
<i>Streptomyces echinatus</i> ABL09954	314	7e-86	54% / 65%	<i>Orf6</i>	ciklaza	aranciamicin
<i>Frankia</i> sp. YP_483220	318	1e-85	53% / 65%	lokus Francci3_4143	bifunkcionalna ciklaza/dehidraza	/
<i>Streptomyces antibioticus</i> AAK06788	315	2e-85	52% / 64%	<i>simA5</i>	ciklaza/dehidraza	simociklinon

ChdQII je tretji izmed proteinov, za katerega sklepamo, da sodeluje pri zapiranju rastoče poliketidne verige v končno strukturo štirih obročev molekule kelokardina. Analiza BLASTP pokaže, da ima zaporedje dve ohranjeni domeni (Slika 41), kar je značilno tudi za proteinsko zaporedje OtcD1(OxyK), ki kodira aromatazo/ciklazo (Petković in sod., 1999; Zhang in sod., 2007). Ta protein najdemo na drugem mestu med najboljšimi zadetki analize BLASTP in ima tudi najvišji odstotek identičnosti z ChdQ izmed vseh zadetkov. Za z ChdQII najbolj homologne proteine je na splošno značilna nizka E-vrednost poravnave proteinskega zaporedja. Med njimi je največ takih s ciklazno, aromatazno in ciklazno/hidroksilazno predvideno funkcijo (Preglednica 18). Natančno biološko vlogo je takim proteinom težko določiti, glede na rezultate analize BLASTP pa predvidevamo, da zaporedje ChdQII kodira dvokomponentno ciklazo/aromatazo.



Slika 41: Analiza BLASTP proteinskega zaporedja ChdQII grafično prikaže dve domeni, značilni za poliketidne ciklaze/aromataze.

4.1.18 ChdOII



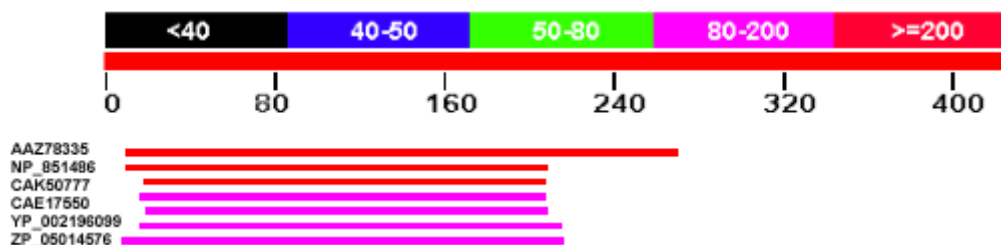
Slika 42: Opis predvidenega genskega produkta ChdOII

Preglednica 19: Seznam najbližjih homologov genskega produkta ChdOII, pridobljen z analizo BLASTP

Organizem	Dolžina prot. (ak)	E-vrednost	Identičnost/ podobnost	Gen/lokus	Funkcija	Gen za biosintezo antibiotika
<i>Streptomyces rimosus</i> AAZ78335	557	7e-59	65% / 75%	<i>oxyL</i>	oksisgenaza	oksitetraciklin
<i>Streptomyces rochei</i> NP_851486	583	1e-55	56% / 72%	lokus pSLA2-L_p064	oksisgenaza	/
<i>Streptomyces argillaceus</i> CAK50777	531	7e-52	57% / 71%	<i>mtmOII</i>	oksisgenaza	mitramicin
<i>Streptomyces griseus</i> subsp. <i>griseus</i> CAE17550	557	4e-48	54% / 69%	<i>cmmOII</i>	oksisgenaza	kromomicin
<i>Streptomyces pristinaespiralis</i> YP_002196099	584	9e-48	54% / 68%	lokus SSDG_05476	oksisgenaza	/

Analiza BLASTP zaporedja ChdOII nam kot najbolj homologne vrne proteine, katerih povprečna velikost je okoli 550 aminokislin in v večini kodirajo oksigenaze. Med najboljše zadetke se uvrščajo tudi oksigenaze, ki sodelujejo pri biosintezi oksitetraciklina (OxyL), mitramicina (MtmOII) in kromomicina (CmmOII) (Preglednica 19).

Za nukleotidno zaporedje *chdOII* ne moremo nedvoumno določiti stop-kodona, glede na poravnave s pomočjo algoritma BLASTP pa predvidevamo, da se nahaja pri dolžini približno 603 bp (200 ak) (Slika 43). Kljub temu je kot odprti bralni okvir (Slika 42) prikazan tisti, za katerega stop-kodon je bil v resnici najden.



Slika 43: Grafični prikaz ujemanja *ChdOII* z najbolj podobnimi proteini

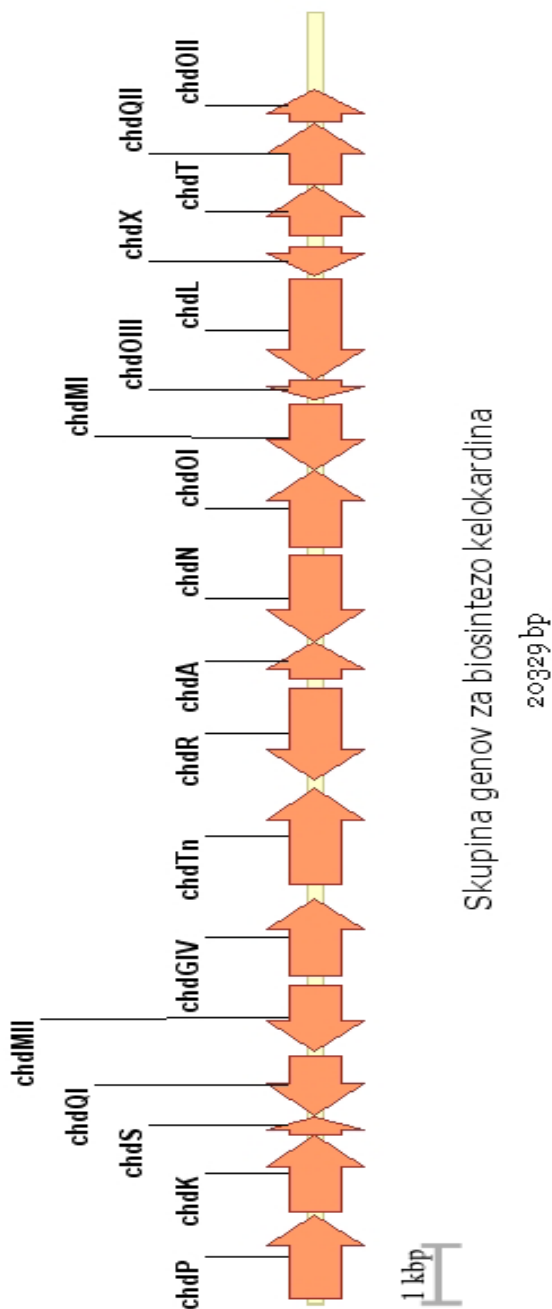
Podobnost proteinskega zaporedja ChdOII z OxyL je 65%, identičnost pa 75%, kar je v primerjavi z proteinskimi zaporedji, ki jih kodirajo ostali geni iz genske skupine za

biosintezo kelokardina precej visoko, a je ta podobnost visoka le na manjšem delu proteinskega zaporedja ChdOII, kar nam prikaže tudi grafični prikaz analize BLASTP (Slika 43). ChdOII je z drugimi podobnimi proteini homogen le na okoli 210 aminokislin dolgem delu zaporedja, v preostalem delu pa skoraj ni več zaznati njegove podobnosti z ostalimi zaporedji. To smo dodatno potrdili tudi s poravnami proteinskih zaporedij najboljše točkovanih zadetkov v programu AlignX. Tudi na grafu, ki ga izriše FramePlot (Slika 43), je vidno, da ima ta odprti bralni okvir visok odstotek G-C v tretji bazi kodona le do dolžine okoli 600 nukleotidov, kar se ujema tudi z rezultati analize BLASTP. Glede na te rezultate lahko predvidevamo, da se translacija verjetno pri dolžini okoli 200 aminokislin ustavi, zato podobno kot pri *chdL* predvidevamo na dolžini okoli 600 nt stop-kodon, ki ga sicer v zaporedju ne najdemo.

Glede na visoko stopnjo podobnosti z OxyL (Zhang in sod., 2006) predvidevamo, da podobno kot omenjeni protein verjetno tudi ChdOII sodeluje pri hidroksilaciji atoma C-12a molekule kelokardina.

4.2 SHEMA SKUPINE GENOV ZA BIOSINTEZO KELOKARDINA

Slika 44 shematsko prikazuje predvidene genske homologe, ki najverjetneje sodelujejo pri biosintezi kelokardina. Na shemi so geni prikazani v obliki puščic, ki se v velikosti in usmeritvi ujemajo z dejansko pozicijo gena glede na preostale gene v genski skupini.



Slika 44: Shematski prikaz predvidene skupine genov za biosintezo kelokardina.

Preglednica 20: Preglednica z osnovnimi podatki za vsak predviden gen v skupini genov za biosintezo kelokardina

Gen	Predvidena funkcija	Prvi nukleotid	Zadnji nukleotid	Dolžina v nt	Dolžina v ak
<i>chdP</i>	ketosintaza α	86	1384	1299	432
<i>chdK</i>	ketosintaza β	1435	2637	1203	400
<i>chdS</i>	ACP	2662	2928	267	88
<i>chdQI</i>	aromataza/ciklaza	3878	2970	909	302
<i>chdMII</i>	metiltransferaza	3998	2967	1032	343
<i>chdGIV</i>	glikoziltransferaza	5162	6361	1200	399
<i>chdTn</i>	transpozaza, <i>is4</i>	6584	8101	1518	505
<i>chdR</i>	izvozni protein	9685	8240	1446	481
<i>chdA</i>	transkripcijski regulator	9836	10408	573	190
<i>chdN</i>	amidotransferaza	11768	10422	1347	448
<i>chdOI</i>	oksidogenaza	11901	13115	1215	404
<i>chdMI</i>	metiltransferaza	14148	13123	1026	341
<i>chdOIII</i>	oksidogenaza	14516	14226	291	96
<i>chdL</i>	acil-coa ligaza	16118	14226	1893	630
<i>chdX</i>	ciklaza/aromataza	16635	16183	453	150
<i>chdT</i>	ketoreduktaza	16800	17588	789	262
<i>chdQII</i>	ciklaza/aromataza	17617	18564	948	315
<i>chdOII</i>	oksidogenaza	18593	19822	1230	409

5 RAZPRAVA IN SKLEPI

Nukleotidno zaporedje, ki verjetno kodira gene, potrebne za biosintezo kelokardina, je v okviru doktorske disertacije pridobila dr. Urška Lešnik (Lešnik, 2009). Zaporedje smo analizirali s pomočjo več računalniških programov, s katerimi smo predvideli potencialne ORF in funkcije podobnih proteinov, ohranjene domene in motive, kar nam je ob pomoči literature omogočilo predvideti verjetno vlogo posameznih proteinov in njihovo sodelovanje pri biosintezi kelokardina.

Program FramePlot 4.0 beta (Ishikawa in Hotta, 1999), ki smo ga uporabili za iskanje odprtih bralnih okvirjev, je za tovrstno delo zelo uporaben, saj so grafi na predelih odprtih bralnih okvirjev vidno nakazovali, da gre za regije DNA zaporedij, ki prav gotovo kodirajo proteinske produkte. Prednost programa je tudi enostavna nadaljna analiza, saj nas FramePlot iz rezultatov direktno preusmeri na BLAST analizo na spletni strani NCBI. Analiza BLASTP nam je podala najbolj podobna zaporedja iskanemu zaporedju, katerim je preko E-vrednosti tudi ocenila verjetnost, da so podani rezultati pravilni. V iskanem zaporedju je program prikazal tudi ohranjene domene, značilne za proteine z določeno funkcijo ali družine proteinov.

Analiza odprtih bralnih okvirjev kloniranega zaporedja je pokazala 38 predvidenih odprtih bralnih okvirjev, od tega smo na podlagi analiz *in silico* za 18 genov predvideli njihovo funkcijo v biosintezi kelokardina (Preglednica 20). Ostali ORF najverjetneje ne sodelujejo pri biosintezi kelokardina.

Rezultati bioinformatičkih analiz in predvidene vloge verjetnih genov se glede na kemijsko strukturo kelokardina večinoma ujemajo, kar nam še dodatno potrjuje, da odsek z določenim zaporedjem res kodira skupino genov, ki je odgovorna za biosintezo kelokardina. Glede na primerjavo s podobnimi skupinami genov lahko z veliko verjetnostjo trdimo, da skupina genov pripada PKS tipa II. V tip II spadajo tudi PKS za biosintezo klortetraciklina (Ryan, 1995), oksitetraciklina (Zhang in sod., 2006), mitramicina (Lombo in sod., 2000) in kromomicina (Menendez in sod., 2004), katerih geni so se pogosto pojavljali med najbolj podobnimi zaporedji z geni za biosintezo kelokardina.

Na 5' začetku nukleotidnega zaporedja predvidene skupine genov najdemo nujno potrebne gene za tvorbo poliketidne verige, ki sestavljajo t.i. minimalni PKS, ki katalizira kondenzacijo naraščajoče poliketidne verige. To so ChdP (KS α), ChdK (KS β) in ChdS (ACP). Nastala poliketidna veriga se nato še dodatno v več stopnjah modificira do končne molekule kelokardina. Pri tem imajo pomembno vlogo prav geni, ki smo jih uspeli identificirati. Za nastanek za tetracikline značilnih štirih obročev imajo pomembno funkcijo tri ciklaze/aromataze (ChdQI, ChdQII in ChdX), ki skrbijo za pravilno ciklizacijo obročev. V skupini genov smo našli tudi tri oksigenaze (ChdOI, ChdOII, ChdOIII), ki oksidirajo molekulo na različnih mestih, s čimer prav tako sodelujejo pri pravilem zlaganju obročev. Med geni v genski skupini se nahajata tudi dve metiltransferazi, ki smo ju glede na prisotnost metilnih skupin tudi pričakovali. Glede na podobnost z metiltransferazami iz drugih genskih skupin sklepamo, da metiltransferaza ChdMII prenaša metilno skupino na 9. C-atom in metiltransferaza ChdMI prenaša metilno skupino na 6. C-atom molekule kelokardina (Slika 10). Za prenos amino skupine na 4. C-atom (Slika 10) je najverjetneje odgovorna aminotransferaza ChdN. Genskemu produktu ChdL glede na rezultate

pripisujemo vlogo acil-CoA ligaze, katere funkcije v biosintezi kelokardina ne poznamo. Med geni v genski skupini za biosintezo kelokardina najdemo tudi gen za ketoreduktazo *chdT*. Genski produkt ChdR služi verjetno kot izvozni protein, ki molekulo kelokardina črpa iz bakterije, ki ga proizvaja, zaradi česar je bakterija odporna na lastni antibiotik. Pomembno funkcijo ima verjetno tudi ChdA, za katerega glede na rezultate naših analiz predvidevamo, da sodeluje pri regulaciji transkripcije.

Vsi zgoraj opisani geni oz. genski produkti so bili glede na zgradbo molekule kelokardina in primerjavo z ostalimi podobnimi tetraciklinskimi antibiotiki v skupini genov pričakovani. Med geni pa najdemo tudi dva gena, ki najverjetneje nimata vpliva na biosintezo kelokardina. To sta gena za glikoziltransferazo (*chdGIV*) in transpozazo iz družine IS4 (*chdTn*). Molekula kelokardina nima nobenih sladkornih komponent, zato je prisotnost glikoziltransferaze nepričakovana. Glede na to, da sta gena za glikoziltransferazo in transpozazo kodirana zaporedno (Slika 44), pričakujemo, da sta v skupino genov oba gena prenesla po naključju in ne vplivata na biosintezo. Njunjo prisotnost v genski skupini si lahko razlagamo na več načinov. Gen za glikoziltransferazo je značilen za nekatere podobne tetraciklinske antibiotike, kot sta mitramicin (Lombo in sod., 2000) in kromomicin (Menendez in sod., 2004), zato obstaja možnost, da se je glikoziltransferaza prenesla s pomočjo homologne rekombinacije, medtem ko se je transpozaza v skupini genov verjetno znašla po naključju. Druga možna razlaga bi bila, da se je glikoziltransferaza prenesla skupaj s transpozazo, ki so znane po naključnem premikanju po genomu, pri čimer lahko prenese tudi naključno DNA (Lewin, 2004).

V času analize DNA zaporedja kelokardina smo demonstrirali, da je bioinformatika močno orodje, s katerim lahko napovedujemo funkcijo genov. Kljub temu pa se je potrebno zavedati, da rezultati, ki jih dobimo z analizami *in silico*, niso zadosten dokaz funkcije genov. Za bolj izčrpne potrditve funkcije genov je potrebno njihovo delovanje še dodatno eksperimentalno dokazati. Pri tem se uporabljata predvsem dve metodi: gen lahko z molekularnimi metodami prekinemo in spremljamo njegov vpliv na biosintezo. Ta način so npr. uporabili Petković in sodelavci (1999), ko so s prekinitvijo gena dokazali, da aromataza/ciklaza OtcD1 (OxyK) pomembno vpliva na pravilno biosintezo oksitetraciklina. Pri prekinitvah genov pogosto prihaja do popolne prekinitve biosinteze ali nastanka nepredvidenih spojin, zaradi česar s prekinitvijo gena ne moremo vedno dokazati njegovega vpliva na biosintezo. Sodelovanje celotne skupine genov pri biosintezi pa največkrat dokazujemo s heterologno ekspresijo v nadomestnem gostitelju, ki teh genov nima. Pri takem gostitelju po prenosu genov opazujemo nastanek zelenega proizvoda. Če se bo tvoril zelen produkt, smo nedvoumno dokazali vlogo preučevanih genov.

5.1 SKLEPI

- Glede na kemijsko strukturo kelokardina lahko predvidevamo, da so v genski skupini, ki smo jo določili, vsi potrebni geni za biosintezo kelokardina, s čimer smo potrdili ključno hipotezo. Ključne stopnje v biosintezi kelokardina so:
 - nastanek poliketidne verige z encimi t.i. minimalnega PKS (ChdP, ChdK in ChdS),
 - ketoredukcija na mestu C-8 (ChdT),
 - nastanek štirih obročev (ChdQI, ChdQII in ChdX),

- oksidacija molekule (ChdOI, ChdOII in ChdOIII),
 - prenos amino skupine na mesto C-4 (ChdN),
 - metilacija na mestih C-6 in C-9 (ChdMI in ChdMII).
- S primerjavo analizirane genske skupine s podobnimi genskimi skupinami lahko trdimo, da gre za PKS tipa II, ker se genska skupina glede na organizacijo in zastopanost genov ujema z genskimi skupinami tega tipa.
- V genski skupini najdemo tudi gen *chdR*, ki kodira protein za odpornost na antibiotike in gen *chdA*, katerega genski produkt je podoben transkripcijskim regulatorjem iz družine tetraciklinskih represorjev TetR in verjetno regulira biosintezo kelokardina.
 - Skupina genov vsebuje tudi dva gena, t.j. glikoziltransferaza in transpozaza, za katera ne moremo ugotoviti funkcije v biosintezi kelokardina.
 - Z izvedenimi analizami *in silico* lahko funkcije genov predvidimo zgolj z določeno mero teoretične verjetnosti. Za zanesljivejšo potrditev je potrebno izvesti dodatne eksperimentalne analize.

6 POVZETEK

Namen dela je bil opraviti analizo *in silico* nukleotidnega zaporedja, ki predvidoma kodira gene, ki sodelujejo pri biosintezi kelokardina. Zaporedju smo s pomočjo programa FramePlot predvideli odprte bralne okvirje. S programom smo predvideli 38 odprtih bralnih okvirjev. Tem smo potem z algoritmom BLASTP poiskali najbolj homologne proteine in glede na le-te predvideli funkcijo, ki jo kodirajo. Ugotovili smo, da od 38 predvidenih bralnih okvirjev verjetno le 18 sodeluje pri biosintezi kelokardina. To smo predvideli s primerjavo genov z geni iz skupin genov za biosintezo strukturno podobnih tetraciklinov. Skupina genov za biosintezo kelokardina se začne z geni za t.i. minimalni PKS, ki ga sestavljajo geni *chdP*, *chdK* in *chdS*, ki verjetno kodirajo KS α , KS β in ACP. To so encimi, ki so nujno potrebni za sestavljanje poliketidne verige. Poleg teh najdemo v skupini genov tudi ostale gene, ki poliketidno verigo dodatno spremenijo in pomagajo, da se na koncu sestavi v molekulo kelokardina z za tetracikline značilnimi štirimi povezanimi obroči. Med predvidene gene, ki prav tako sodelujejo pri biosintezi kelokardina, tako spadajo trije geni za oksigenaze (*chdOI*, *chdOII*, *chdOIII*), trije geni, ki verjetno sodelujejo pri ciklizaciji obročev (*chdQI*, *chdQII*, *chdX*), dva gena za metiltransferazo (*chdMI* in *chdMII*), en gen za ketoreduktazo (*chdT*), gen za aminotransferazo (*chdN*) in gen za acil-CoA ligazo *chdL*, katere vpliv na biosintezo kelokardina še ni dokočno poznan. Poleg teh genov najdemo tudi gen *chdA*, ki kodira transkripcijski regulator. Pred njim najdemo gen *chdR*, ki verjetno kodira protein, ki izloča molekulo kelokardina iz celice in tako varuje bakterijo pred njenim delovanjem. Med geni pa nismo pričakovali gena *chdGIV*, ki kodira glikoziltransferazo in gena *chdTn*, ki kodira transpozazo, za katera tudi predvidevamo, da nimata nobene funkcije pri biosintezi kelokardina. Glede na podobnost genov genske skupine z drugimi genskimi skupinami za biosintezo antibiotikov lahko z veliko gotovostjo trdimo, da skupina genov kodira PKS tipa II. Še več, glede na kemijsko strukturo kelokardina so v skupini genov najverjetneje kodirani vsi geni, potrebni za biosintezo kelokardina. Uporaba bioinformacijskih orodij je v veliko pomoč pri analizi zaporedij in predvidevanju genov, ki jih zaporedje kodira ter njihovih funkcij. Kljub temu pa je potrebno funkcionalnost skupine genov še dodatno eksperimentalno potrditi.

7 VIRI

- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215, 3: 403-410.
- Altshul S.F., Koonin E.V. 1998. Iterated profile searches with PSI-BLAST – a tool for discovery in protein databases. *Trends in Biochemical Sciences*, 23, 11: 2444-447.
- Apweiler R., Attwood T.K., Bairoch A., Bateman A., Birney E., Biswas M., Buch P., Cerutti L., Corpet F., Croning M.D.R., Durbin R., Falquet L., Fleischmann W., Gouzy J., Hermjakob J., Hulo N., Jonnasen I., Kahn D., Kanapin A., Karavidopoulou Y., Lopez R., Marx B., Mulder N. J., Oinn T.M., Pagni M., Servant F., Sigrist C.J.A., Zdobnov E. M. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Research*: 29, 1: 37-40.
- Austin M.B., Noel J.P. 2003. The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports*, 20: 79-110.
- Bairoch A., Apwiler R., Wu C.H., Barker W.C., Boeckemann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J., Natale D.A., O'Donovan C., Redaschi N., Yeh L.-S. L. 2005. The Universal Protein Resource (UniProt). *Nucleic Acids Research*, 33: D154-D159.
- Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J, Wheeler D.L. 2008. GenBank. *Nucleic Acids Research*, 36: D25-D30.
- Bibb M.J., Findlay P.R. Johanson M.W. 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene*, 30: 157-166.
- Bisang C., Long P.F., Cortés J., Westcott J., Crosby J., Matharu A.L., Cox R.J., Simpson T.J., Staunton J., Leadlay P.F. 1999. A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature*, 401, 6752: 502-505.
- Boone D.R. Castenholz R.W., Garrity G.M. 2001. *Bergey's manual of systematic biology*. 2nd ed. Vol 1.: The *Archeae* and the deeply branching and phototrophic bacteria. New York, Springer: 721 str.
- Challis G. L., Hopwood D. A. 2003. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 2: 14555-14561.
- Chenna R. Sugawara H., Koike T., Lopez R., Gibson T.J., Higgins D.G., Thompson J.D. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, 31, 13: 3497-3500.
- Chopra I. 1994. Tetracycline analogs whose primary target is not the bacterial ribosome. *Antimicrobial Agents and Chemotherapy*, 38, 4: 637-640.

- Chopra I., Roberts M. 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews*, 65, 2: 232-236.
- Dale W. J., von Schantz M. 2007. From genes to genomes: Concepts and applications of DNA technology. 2nd ed. Chichester, John Wiley & Sons Ltd: 384 str.
- Doan T.L., Fung H.B., Mehta D., and Riska P.F. 2006. Tigecycline: A Glycylcycline antimicrobial agent. *Clinical Therapeutics*, 28,8:1079-1106.
- Donadio S., Sosio M. 2003. Strategies for combinatorial biosynthesis with modular polyketide synthases. *Combinatorial Chemistry & High Throughput Screening*, 6,6: 489-500.
- EMBL-EBI. 2009a. Frequently asked questions about ClustalW2. Cambridge, EMBL-EBI –European Bioinformatics Institute
<http://www.ebi.ac.uk/Tools/clustalw2/faq.html> (avgust, 2009): 1-1.
- EMBL-EBI. 2009b. Protein databases. Cambridge, EMBL-EBI - European Bioinformatics Institute
<http://www.ebi.ac.uk/2can/databases/protein8.html> (avgust, 2009): 1-1.
- Gibson K.J.C., Gilerron M., Constant P., Puzo G., Nigou J., Besra G.S. 2003. Identification of a novel mannose-capped lipoarabinomannan from *Amycolatopsis ulphurea*. *Biochemical Journal*, 372: 821-829.
- Greer N.D. 2006. Tigecycline (Tygacil): the first in the glycylcycline class of antibiotics. *Baylor University Medical Center Proceedings*, 19:155–161.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1426172/> (januar, 2010)
- Hunter S., Apweiler R., Attwood T.K., Bairoch A., Bateman A., Binns D., Bork P., Das U., Daugherty L., Duquenne L., Finn R. D., Gough J., Haft D., Hulo N., Kahn D., Kelly E., Laugraud A., Letunic I., Lonsdale D., Lopez R., Madera M., Maslen J., McAnulla C., McDowall J., Mistry J., Mitchell A., Mulder N., Natale D., Orengo C., Quinn A.F., Selengut J.D., Sigrist C.J., Thimma M., Thomas P.D., Valentin F., Wilson D., Wu C.H., Yeats C. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Research*, 37: D211-D215.
- Hershberger C.L. 1996. Metabolic engineering of polyketide biosynthesis. *Current Opinion in Biotechnology*, 7:560-562.
- Hertweck C., Luzheskyy A., Rebets Y., Bechthold A., 2007. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports*, 24,1: 162-190.
- Higgins D.G., Sharp P.M. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73: 237-244.

- Holt J.G. 1994. Bergey's manual of determinative bacteriology. 9th ed. Baltimore, Williams&Wilkins: 787 str.
- Hopwood D.A. 1997. Genetic contributions to understanding polyketide synthases. *Chemical Reviews*, 97, 7: 2465-2497.
- Hopwood D.A. 1999. Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. *Microbiology*, 145, 9: 2183-2202.
- Hranueli D., Cullum J. 2001. Production of hybrid polyketides by combinatorial biosynthesis. *Kemija u Industriji*, 50: 381-411.
- Huang X., Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Research*, 9: 868-877.
- Hunter I.S., Hill R.A. 1997. Tetracyclines. V: *Biotechnology of antibiotics*. 2nd ed. Strohl W.E. (ed.) New York, USA, Marcel Decker: str. 659-682.
- Ichinose K., Ozawa M., Itou K., Kunieda K., Ebizuka Y. 2003. Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of *Streptomyces* sp. AM-7161: towards comparative analysis of the benzoisochromanquinone gene clusters. *Microbiology*, 149: 1633-1645.
- INSDC.2009. International Nucleotide Sequence Database Collaboration (INSDC). Mishima, Cambridge, Bethesda. INSDC - International Nucleotide Sequence Database Collaboration
<http://www.insdc.org/> (avgust, 2009): 1-1 str.
- Invitrogen Corporation. 2004. Vector NTI Advance™ 10th user's manual. Carisbad, Invitrogen Corporation: 758 str.
- Ishikawa J., Hotta K. 1999. FramePlot: a new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiology Letters*, 174: 251-253.
- Ishikawa J. 2008. Genome analysis system for *Actinomycetes*: Development and Application. *Actinomycetologica*, 22: 46-49.
- Ishikawa J. Hoshino Y. Ishino K., Kurito H., Chiba K., Fujii S., Shibuya K., Hattori M. Yamashita A., Mikami Y., Yazawa K., Takeda K. 2009. Tokio, *Nocardia farcinica* Genome Project Page
<http://nocardia.nih.go.jp/> (avgust, 2009): 1-1.
- Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S., Madden T.L. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36 (Web Server issue): W5-9. doi:10.1093/nar/gkn201
- Kaminski N. 2000. Bioinformatics.: A user's perspective. *American Journal of Respiratory Cell and Molecular Biology*, 23: 705-711.

- Khosla C., Tang Y., Chen A.Y., Schnarr N.A., Cane D.E. 2007. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annual Reviews of Biochemistry*, 76: 195-221.
- Kingston W. 2004. Streptomycin, Schatz vs. Waksman, and the balance of credit for discovery. *Journal of the History of Medicine and Allied Sciences*, 59, 3:441-462.
- Komaki H., Harayama S. 2006. Sequence diversity of type-II polyketide synthase genes in *Streptomyces*. *Actinomycetologica*, 20: 42-48.
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 21: 2947-2948.
- Lechevalier M.P., Prauser H., Labeda D.P., Ruan J.S. 1986. Two new genera of nocardioform *Actinomycetes*: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. *Journal of Systematic and Evolutionary Microbiology*, 36,1: 29-37.
- Lešnik U., Gormand A. M., Magdevska V., Fujs Š., Raspor P., Hunter I., Petković H. 2009. Regulatory elements in tetracycline-encoding gene clusters : the otcG gene positively regulates the production of oxytetracycline in *Streptomyces rimosus*. *Food Technology and Biotechnology*, 1, 9: 45-51.
- Lešnik U. 2009. Kloniranje genov za biosintezo tetraciklinov izbranih bakterij redu Actinomycetales. Doktorska disertacija. Ljubljana, Biotehniška fakulteta, Podiplomski študij bioloških in biotehniških znanosti: 184 str.
- Lewin B. 2004. *Genes VIII*. 8th ed. Oxford, Oxford University Press: 1056 str.
- Lombo F., Künzel E., Prado L., Braun A.F., Bindseil K.U., Frevart J., Bearden D., Mendez C., Salas J. A. 2000. The novel hybrid antitumor compound premithramycinone H provides indirect evidence for a tricyclic intermediate of the biosynthesis of the aureolic acid antibiotic mithramycin. *Angewandte Chemie*, 39, 4: 796-799.
- Madden T. 2003. The BLAST sequence analysis tool. V: *The NCBI handbook*. Bethesda, NCBI - National Center for Biotechnology Information (avgust, 2003)
<http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/handbook/ch16.pdf> (avgust, 2009): 17 str.
- Madigan T., Martinko J.H., Dunlap P.V., Clark D.P. 2009. *Brock biology of microorganisms*, 12th ed. Upper Saddle River, Prentice Hall: 1168 str.
- Magnuson K., Jackowski S., Rock C.O., Cronan J.E. 1993. Regulation of fatty acid biosynthesis in *Escherichia coli*. *Microbiological Reviews*, 57: 522-542.
- Majumdar S., Prabhakaran S.R., Shivaji S., Lal R., 2006. Reclassification of *Amycolatopsis orientalis* DSM 43387 as *Amycolatopsis benzoatilytica* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 56: 199-204.

- Menendez N., Nur-e-Alam M., Brana A. F., Rohr J., Salas J. A., Mendez C. 2004. Biosynthesis of the antitumor chromomycin A3 in *Streptomyces griseus*. *Chemistry & Biology*, 11, 121-32.
- Meurer G., Hutchinson C.R. 1995. Functional analysis of putative β -ketoacyl: acyl carrier protein synthase and acyltransferase active site motifs in a type II polyketide synthase of *Streptomyces glaucescens*. *Journal of Bacteriology*, 177: 477-481.
- Mitcher L.A., Juvarkar J.V., Rosenbrook W., Andres W.W., Schenck J.R., Egan R.S. 1970. Structure of chelocardin, a novel tetracycline antibiotic. *Journal of the American Chemical Society*, 92, 20: 6070-6071.
- Mizrachi I. 2007. GenBank: The Nucleotide Sequence Database. Bethesda, , NCBI - National Center for Biotechnology Information (avgust, 2007)
<http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/handbook/ch1.pdf> (avgust, 2009) : 17 str.
- Nakamura Y., Gojobori T., Ikemura T. 1997. Codon usage tabulated from the international DNA sequence databases. *Nucleic Acids Research*, 25: 244-245.
- NCBI. 2004a. GenBank overview. Bethesda, NCBI - National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html> (avgust, 2009): 1-1.
- NCBI. 2004b. Just the facts: A basic introduction to the science underlying NCBI resources. Bethesda, NCBI - National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html> (avgust, 2009): 1-6.
- NCBI. 2006. Entrez Help: The databases. Bethesda, NCBI - National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/bookshelf/picrender.fcgi?book=helpentrez&part=EntrezHelp&blobtype=pdf> (avgust, 2009): 36 str.
- NCBI. 2009a. BLAST basic local alignment search tool, Blast program selection guide. Bethesda, NCBI - National Center for Biotechnology Information
http://www.ncbi.nlm.nih.gov/blast/BLAST_guide.pdf (avgust, 2009): 20 str.
- NCBI. 2009b. GenBank growth. Bethesda, NCBI - National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html> (avgust, 2009): 1-1.
- NCBI. 2009c. Conserved Domains Database (CDD) and resource group. Bethesda, NCBI - National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml> (avgust, 2009): 1-1.
- NCBI. 2009d. Conserved Domain Database (CDD) help. Bethesda, NCBI - National Center for Biotechnology Information
http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd_help.shtml (avgust, 2009): 1-1.

- Oliva B., Gordon G., McNicholas P., Ellestad G., Chopra I. 1992. Evidence that tetracycline analogs whose primary target is not the bacterial ribosome cause lysis of *Escherichia coli*. *Microbial Agents and Chemotherapy*, 36, 5: 913-919.
- Oliver T.J., Prokop J.F., Bower R.R., Otto R.H., 1962. Chelocardin, a new broad-spectrum antibiotic. I. Discovery and biological properties, *Antimicrobial Agents in Chemoterapy*, 583-591.
- Orengo C., Jones D., Thorton J. 2003. *Bioinformatics: Genes, proteins & computers*. New York, Taylor & Francis Group: 298 str.
- Palmeaner D., Siguier P., Mahillon J. 2008. IS4 family goes genomic. *BMC Evolutionary Biology*, 8:18:1-15.
- Prado L., Lombo F., Brana A.F., Mendez C., Rohr J., Salas J.A. 1999. Analysis of two chromosomal regions adjacent to genes for a type II polyketide synthase involved in the biosynthesis of the antitumor polyketide mitramycin in *Streptomyces argillaceus*. *Molecular and General Genetics*, 261, 2: 216-225.
- Pearson W.R, Lipman D.J. 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, 85, 8:2444-2448.
- Peterson L.R. 2008. A review of tigecycline the first glycylycycline. *International Journal of Antimicrobial Agents*, 32, S4: S215 S222.
- Petković H., Thamchaipenet A., Zhou L.H., Hranueli D., Raspor P., Waterman P.G., Hunter I.S. 1999. Disruption of an aromatase/cyclase from the oxytetracycline gene cluster of *Streptomyces rimosus* results in production of novel polyketides with shorter chain lengths. *Journal of Biological Chemistry*; 274, 46: 32829-32834.
- Petković H., Cullum J., Hranueli D., Hunter I. S., Perić-Concha N., Pigac J., Thamchaipenet A., Vujaklija D., Long P. F. 2006. Genetics of *Streptomyces rimosus*, the oxytetracycline producer. *Microbiology and Molecular Biology Reviews*, 70, 3: 704-728.
- Pfeifer B. A., Khosla C. 2001. Biosynthesis of polyketides in heterologous hosts. *Microbiology and Molecular Biology Reviews*, 65, 1: 106-118.
- Ramos J. L., Martínez-Bueno M., Molina-Henares A.J., Terán W., Watanabe K., Zhang X., Trinidad G. M., Brennan R., Tobes R. 2005. The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews*, 69, 2: 326–356.
- Rasmussen B., Noller H.F., Daubresse G., Oliva B., Misulovin Z., Rothstein D.M., Ellestad G.A., Gluzman Y., Tally F.P., Chopra I. 1991. Molecular basis of tetracycline action: identification of analogs whose primary target is not the bacterial ribosome. *Antimicrobial Agents and Chemotherapy*, 35, 11: 2306-2311.

- Rezsöhazy R, Hallet B, Delcour J, Mahillon J. 1993. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Molecular Microbiology*, 9, 6 :1283-95.
- Ryan M.J., Lotvin J.A., Strathy N. S., Fantini L.K. 1995. Cloning of the biosynthetic pathway for chlortetracycline and tetracycline formation and cosmids useful therein. United States Patent US5866410: 39 str.
- Shen B. 2000. Biosynthesis of aromatic polyketides. *Topics in Current Chemistry*, 209: 1-51.
- Shen B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, 7: 285–295.
- Staunton J. Weismann K. 2001. Polyketide biosynthesis: a millennium review. *Natural Product Reports*, 18: 380-416.
- Smith T.F, Waterman M.S. 1981. Identification of common molecular subsequences. *Journal of Molecular Biology*, 14: 195-197.
- The UniProt Consortium. 2009. The universal protein resource (UniProt) 2009. *Nucleic Acids Research*, 37, D169-D174.
- Tseng M., Yang S.F., Li W.J., Jiang C.L. 2006. *Amycolatopsis taiwanensis* sp. nov., from soil. *International Journal of Systematic and Evolutionary Microbiology*, 58: 1811-1815.
- Vinci V.A., Conder M.J., Mcada P.C., Reeves C.D., Rambosek J., Davis C.R., Hendrickson L.E. 2001. DNA encoding triol polyketide synthase. United States Patent 6174706 B1: 66 str.
- Vijayakumar R., Muthukumar C., Thajuddin N., Panneerselvam A., Saravanamuthu R. 2007. Studies on the diversity of actinomycetes in the Palk Strait region of Bay of Bengal, India. *Actinomycetologica*, 21: 59-65.
- Wang P., Zhang W., Zhan J., Tang Y. 2009. Identification of OxyE as an ancillary oxygenase during tetracycline biosynthesis. *ChemBioChem*, 10, 9: 1544-1550.
- Waksman S.A., Woodruff H.B. 1940. Bacteriostatic and bacteriocidal substances produced by soil actinomycetes. *Proceedings of the Society for Experimental Biology and Medicine*, 45: 609–614.
- Weissmann K., Müller R. 2009. A brief tour of myxobacterial secondary metabolism. *Bioorganic & Medicinal Chemistry*, 17: 2121–2136.
- wwPDB-World Wide Protein Databank. 2009. Protein Data Bank Contents Guide. Version 3.20. wwPDB
<http://www.wwpdb.org/documentation/format32/v3.2.html> (avgust, 2009): 1-1 str.

- Xia X. 2007. Bioinformatics and the cell. Modern computational approaches in genomics, proteomics and transcriptomics. Ontario, Springer Science + Business Media: 349 str.
- Yoo A., Demirev A.V., Lee J.S., Kim S.D., Nam D.H. 2006. Cloning and analysis of type II polyketide synthase gene cluster from *Streptomyces toxytricini* NRRL 15,443. *Journal of Microbiology*, 44, 6: 649-654.
- Zakeri B., Wright G.D. 2008. Chemical biology of tetracycline antibiotics. *Biochemistry and Cell Biology*, 86: 124-136.
- Zhang W., Ames B.D., Tsai S.C., Tang Y. 2006. Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase. *Applied and Environmental Microbiology*: 72, 4, 2573-2580.
- Zhang W., Watanabe K., Wang C.C.C., Tang Y. 2007. Investigation of early tailoring reactions in oxytetracycline biosynthetic pathway. *Journal of Biological Chemistry*, 282, 35: 25717-25725.

ZAHVALA