

CRISPR-Cas-systeemit ja
Streptomyces-bakteerien
antibioottien biosynteesitutkimus

Pro gradu -tutkielma

Turun yliopisto

Biokemian laitos

Biokemian maisteriohjelma

maaliskuu 2019

Olli-Pekka Virtanen

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

TURUN YLIOPISTO

Biokemian laitos

VIRTANEN, OLLI-PEKKA: CRISPR-Cas-systeemit ja *Streptomyces*-bakteerien antibioottien biosynteesitutkimus

Pro gradu -tutkielma, 60 s.

Biokemia

maaliskuu 2019

Ohjaajat: FM Kaisa Palmu, FT Vilja Siitonen & prof. Mikko Metsä-Ketelä

CRISPR-Cas-systeemi on useaan tyyppiin jaettu bakteerien ja arkkien adaptiivinen immuunipuolustusjärjestelmä, joka toimii kolmeosaisesti: adaptaatio, ilmennys ja häirintä. Tyyppeihin kuuluu esimerkiksi hyvin tunnetut tyypin I- (I-E: *Escherichia coli*) ja tyypin II-systeemi (II-A: *Streptococcus pyogenes*). Tyypin II-A CRISPR-Cas-systeemin modulaarista Cas9-endonukleaasia ja ohjattua häirinnän sekvenssispesifistä DNA:n katkomista voidaan käyttää sekvenssispesifisessä geneettisessä muokkauksessa: Cas9:llä tehdään kohdegeeniin katkos, joka korjataan tämän ja muokatun geenielementin välisellä homologisella rekombinaatiolla. Prosessia kutsutaan CRISPR/Cas9-geenieditoinniksi, jota on alun perin käytetty esimerkiksi eukaryoottien muokkaukseen.

Työn tarkoituksena oli tutkia *Streptomyces*-bakteerien C-glykosidiantibioottien metaboliareittien geenejä delemalla näitä CRISPR/Cas9-geenieditointiplasmideilla ja tulkitsemalla geenien roolia yhdistetuotossa deleetioiden vaikutuksen perusteella. Kohteena oli *Streptomyces showdoensis* showdomysiiniin liittyvä *sdmA*-geeni ja *S. albus* pseudouridimysiiniin (PUM) liittyvät *stuC*- tai *stuG*-geenit. Showdomysiinin on osoitettu estävän syöpäsolujen kasvua, kun taas PUM tehoaa antibioottiresistentteihin bakteereihin inhiboimalla transkriptiokoneiston toimintaa. Kyseisten antibioottien metaboliareittien tuntemusta on mahdollista soveltaa esimerkiksi lääkekehityksessä.

Deleetiokanta *S. showdoensis* ATCC15227- Δ *sdmA* ei tuottanut showdomysiiniä. Tämän lisäksi osoitettiin, että komplementoitu deleetiokanta ja villityppi tuottivat showdomysiiniä, osoittaen *sdmA*-geenin olevan tärkeä showdomysiinin tuotossa. Jatkotutkimuksissa selvisi, että SdMA on tärkeä showdomysiinin C-nukleosidisen sidoksen synteesissä. *S. albus* DSM40763-kannan *stuG*-deleetiota yritettäessä huomattiin, että menetelmä ei todennäköisesti sovellu kyseiseen kantaan; syynä huono rekombinaatiotehokkuus tai kannan resistenssi selektioantibiootteihin.

Avainsanat: Cas9, C-nukleosidi, CRISPR-Cas, geenieditointi, *Streptomyces*

Sisällys

Lyhenteet.....	3
OSA I – KIRJALLINEN OSA.....	5
1. Johdanto	6
1.1 Mikä on CRISPR?.....	6
1.1.1 CRISPR-Cas-puolustussysteemin löytö.....	6
1.1.2 CRISPR-Cas-systeemien luokittelu	7
2. CRISPR-immuunipuolustuksen toiminta.....	8
2.1 Adaptaatio	9
2.1.1 Adaptaatioon tarvittavat Cas-proteiinit	10
2.1.2 Invasiivisen nukleotidisekvenssin valinta	11
2.1.3 Johtajasekvenssin tärkeys ja CRISPR-alueen toistojaksojen synteesi.....	12
2.1.4 Invasiivisen nukleotidisekvenssin integrointi isännän genomiin.....	13
2.1.5 Pohjustettu hankinta tehostaa adaptaatiota.....	14
2.2 CRISPR-Cas alueen ilmentäminen ja crRNA-biogeneesi	17
2.3 Vieraan nukleotidisekvenssin häirintä	20
2.3.1 Tyypin I systeemi.....	20
2.3.2 Tyypin II systeemi.....	21
2.3.3 Tyypin III systeemi	21
3. Cas9-entsyymien rakenne ja toiminta.....	22
3.1 Cas9:n yleinen rakenne	22
3.2 Tunnistuslohko.....	24
3.3 Nukleaasilohko.....	26
4. CRISPR-Cas ja geneettinen muokkaus.....	29
4.1 CRISPR-systeemin löytö geenieditoinnin kannalta.....	29
4.2 CRISPR/Cas bakteerien muokkauksessa: mahdollisuudet ja vaikeudet.....	29
4.3 CRISPR/Cas9:n käyttö <i>Streptomyces</i> -bakteereissa.....	31
4.3.1 Perinteinen vastaan uusi genomienmuokkaus <i>Streptomyceeteissä</i>	31
4.3.2 <i>Streptomyceettien</i> muokkaus CRISPR/Cas9-plasmideilla.....	32
OSA II – KOKEELLINEN OSA	35
5. Tutkimuksen tavoite.....	36
6. Materiaalit ja menetelmät.....	39
6.1 Plasmidikonstruktien valmistaminen ja kloonaukset.....	39

6.1.1 Q5-PCR-reaktio <i>sdmA</i> -deleetioinserttien valmistamiseksi	39
6.1.2 NEBuilder-reaktio <i>sdmA</i> -deleetiokonstruktiin kokoamiseksi	39
6.1.3 Phusion PCR-reaktio <i>stuG</i> -insertin valmistamiseksi	39
6.1.4 Plasmidikloonaus ja eristys <i>E. colista</i>	40
6.1.5 <i>E. coli</i> -tranformaatio	41
6.2 <i>Streptomykeettien</i> kloonaus	42
6.2.1 Konjugaatio <i>E. colin</i> ja <i>Streptomykeettien</i> välillä	42
6.2.2 Plasmidin pudotus <i>Streptomyces</i> -bakteerista	42
6.2.3 <i>Streptomyces</i> -protoplastien valmistus	43
6.2.4 <i>Streptomyces</i> -protoplastitransformaatio	43
6.3 DNA:n eristys <i>Streptomyces</i> -soluista	43
6.3.1 Kromosomaalisen DNA:n eristys	43
6.3.2 <i>Streptomyces</i> -plasmidin eristys	44
7. Tulokset	45
7.1 Showdomysiinin synteesireitin selvittäminen	45
7.1.1 Showdomysiinin <i>sdmA</i> -geenin inaktivoitikonstrukti	45
7.1.2 Showdomysiinin <i>sdmA</i> -geenideleetion komplementaatio	47
7.1.3 Geenideleetion vaikutus showdomysiinin tuottoon	48
7.2 Pseudouridimysiinin synteesireitin selvittäminen	49
7.2.1 Pseudouridimysiinin <i>stuG</i> -geenin inaktivoitikonstrukti	49
7.2.2 Pseudouridimysiinin <i>stuG</i> -geenideleetion komplementaatio	50
7.2.3 Geenideleetion mahdollinen vaikutus <i>stu</i> -yhdisteiden tuottoon	51
8. Tulosten tarkastelu	54
8.1 Showdomysiinimetaboliareitin <i>sdmA</i> :n deleetion tulosten tarkastelu	54
8.2 Pseudouridimysiinimetaboliareitin <i>stuG</i> :n deleetion tulosten tarkastelu	55
9. Päätelmät ja yhteenveto	58
9.1 Showdomysiinimetaboliareitin tieto syöpälääketutkimuksessa	58
9.2 Pseudouridimysiinimetaboliareitin tieto antibioottitutkimuksessa	58
9.3 Bakteerien genomisen muokkaus CRISPR-Cas:lla	60
Lähteet	61

Lyhenteet

<i>act</i>	aktinorodiini (geeniklusteri)
<i>aln</i>	alnumysiini (geeni)
apo	Inaktiivinen entsyymi, josta puuttuu kofaktorit
ATCC	Viittaa ATCC-bakteerikokoelman bakteerikantakoodiin (engl. American Type Culture Collection)
Cas	CRISPR-systeemin proteiini (engl. CRISPR-associated)
CRISPR	Prokaryoottien immuunipuolustuksen DNA-toistojakso (engl. clustered regularly interspaced short palindromic repeats)
crRNA	CRISPR-systeemin lyhyt ribonukleiinihappo
DMSO	dimetyylisulfoksidi
DNA	deoksiribonukleiinihappo
dPUM	deoksipseudouridimysiini
DSB	Nukleiinihapposekvenssin kaksoisjuosteen katkos
DSM	Viittaa DSMZ-bakteerikokoelman bakteerikantakoodiin (saks. Deutsche Sammlung von Mikroorganismen und Zellkulturen)
ET/U	<i>Escherichia coli</i> ET12567/pUZ8002
GAA	guanidinoetikkahappo
HNH	HNH-endonukleaasimainen domeeni Cas9:ssä (nimetty tyypillisen histidiinin ja arginiinin perusteella)
holo	Aktiivinen entsyymi, jossa on kofaktorit
LB	Yleinen kasvatusmediumi (engl. Luria broth)
LC-MS	Nestekromatografi ja massaspektrometri yhdessä
MS	mannoosisoijajauho (agar)
NaAc	natriumasetaatti
NEB	engl. New England Biolabs Inc. (osakeyhtiö)
NMR	ydinmagneettinen resonanssi

NUC	Cas9-proteiinin nukleasilohko
PAM	Protovälikkeen viereinen sekvenssi
PCR	polymeraasiketjureaktio
PDB	Proteiinirakenteiden arkisto (engl. Protein Data Bank)
PI	PAM:in kanssa vuorovaikuttava (domeeni)
pre-crRNA	CRISPR-systeemin ribonukleinihapon esiaste
PS	protovälike
PUM	pseudouridimysiini (geeniklusteri pienellä)
<i>red</i>	undekyyliprodigiosiini (geeniklusteri)
RNA	ribonukleinihappo
RNAasi	ribonukleaasi
RuvC	RuvC-resolvaasiproteiinien kaltainen domeeni Cas9:ssä
<i>sdm</i>	showdomysiini (geeni)
sgRNA	Opastaja-RNA, joka on kahden RNA:n yhdistelmä
STU	strepturidimysiini
tracrRNA	transaktivoiva-CRISPR-RNA
TSB	Yleinen kasvatusmediumi (engl. Tryptic soy broth)
UTP	uridiini-5'-trifosfaatin
v/v	Tilavuuden suhde tilavuuteen
w/v	Painon suhde tilavuuteen

OSA I – KIRJALLINEN OSA

1. Johdanto

1.1 Mikä on CRISPR?

Prokaryoottien genomeissa on useanlaisia DNA-toistojaksoja, jotka sijaitsevat geenien alueilla tai näiden välisillä alueilla. Näitä ovat genomien ympäri sirotellut toistojaksot (engl. interspersed repeats), jotka ovat prokaryooteissa useimmiten geenien välisillä alueilla (Treangen *ym.*, 2009). Näihin kuuluvat esimerkiksi transposonit, faagien elementit ja tutkimuksen kannalta tärkeät CRISPR-toistojaksot (engl. clustered regularly interspaced short palindromic repeats) (Treangen *ym.*, 2009). CRISPR-toistojaksoille on ominaista, että lyhyiden säännöllisten toistosekvenssien välissä on tyypillisesti saman pituinen epäsäännöllinen sekvenssi (21 – 37 emäsparia) (Jansen *ym.*, 2002). Kyseistä epäsäännöllistä sekvenssiä kutsutaan nykyään välikkeeksi (engl. spacer) (Barrangou *ym.*, 2007).

1.1.1 CRISPR-Cas-puolustussysteemin löytö

Tyypilliset CRISPR-toistosekvenssit havaittiin ensimmäisen kerran *Escherichia coli*ssa, tämän *iap*-geeniä (alkalinen fosfataasi) tutkittaessa (Ishino *ym.*, 1987). Noin 10 vuotta myöhemmin eukaryoottien ja prokaryoottien lyhyet DNA-toistosekvenssit luokiteltiin SSR:ksi (engl. short-sequence DNA repeat) (van Belkum *ym.*, 1998). Nämä huomattiin myöhemmin uudestaan, kun arkkien ja bakteerien tyypilliset toistosekvenssit luokiteltiin SRSR:ksi (engl. short regularly spaced repeats) (Mojica *ym.*, 2000). Lopulta kyseiset prokaryoottien toistojaksot nimettiin CRISPR:ksi, mutta tämän lisäksi huomattiin CRISPR-sekvenssien yhteydessä olevat Cas-proteiinit (engl. CRISPR-associated) (Jansen *ym.*, 2002).

CRISPR-alueen alkuperä saatiin ensimmäisen kerran selville, kun *Yersinia pestis*en löydettiin integroivan inaktiivista profaagin nukleotidisekvenssiä CRISPR-alueeseensa (Pourcel *ym.*, 2005). Vihiä CRISPR-Cas-systeemin mahdollisesta roolista prokaryoottien immuunipuolustuksessa saatiin, kun huomattiin, että bakteerit lisäsivät CRISPR-alueeseensa faagien ja plasmidien ekstrakromosomaalisesta DNA:sta johdettuja sekvenssin pätkiä, eli välikkeitä (Bolotin *ym.*, 2005; Mojica *ym.*, 2005). Välikkeiden huomattiin olevan homologisia useiden tunnettujen geenien kanssa, ja että ne ovat yhteydessä immunitettiin faageja vastaan (Bolotin *ym.*, 2005; Mojica *ym.*, 2005).

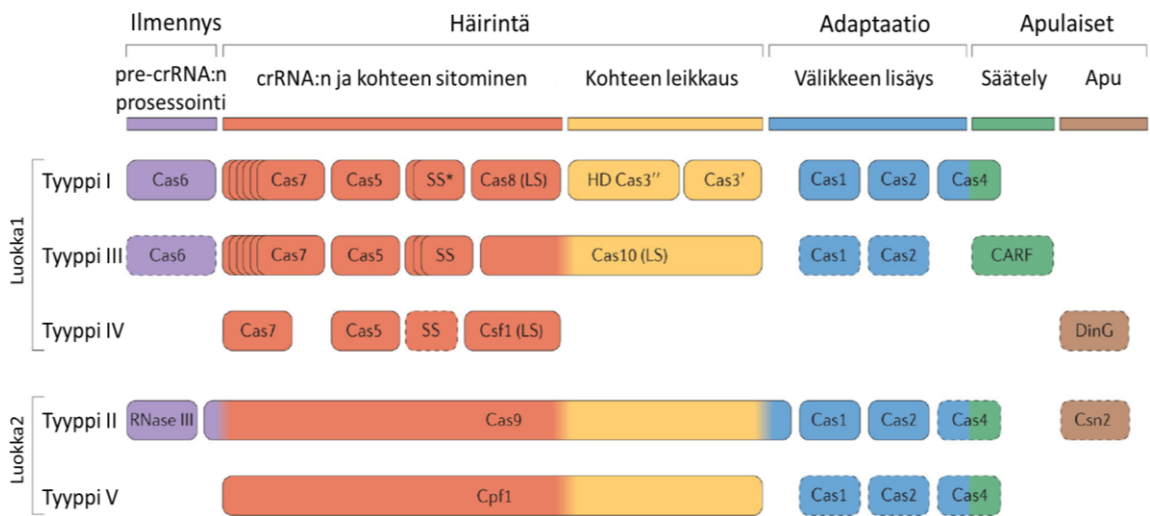
Lopulta varmistettiin, että CRISPR-alueeseen liitetty välike tuo spesifin resistenssin välikkeestä johdettua faagia vastaan, ja että bakteerikanta voi adaptoitua nopeasti faagipopulaatioon lisäämällä uusia välikkeitä (Barrangou *ym.*, 2007). Näin CRISPR-Cas-systeemin varmistettiin olevan prokaryoottien immuunipuolustusjärjestelmä.

1.1.2 CRISPR-Cas-systeemien luokittelu

CRISPR-Cas-systeemit voidaan jakaa kahteen luokkaan: luokassa 1 on moniosaiset vaikuttajaproteiinikompleksit, kun taas luokassa 2 on vain yksi vaikuttajaproteiini. Luokan 1 CRISPR-Cas-systeemeihin kuuluvat tyypin I, III ja lisäksi mahdollisesti tyypin IV systeemit, kun taas luokkaan 2 kuuluu tyypin II systeemit ja mahdollisesti myös tyypin V systeemit. Kyseisillä CRISPR-systeemien tyypeillä ja näiden alatyypeillä on eri määrä eri Cas-proteiineja ja tyyppien kohteena voi olla joko DNA tai RNA, mutta myös tyyppien tapa käsitellä crRNA:ta (CRISPR-RNA) vaihtelee. Eri CRISPR-Cas-luokat, tyypit ja näiden proteiinit ovat esitettynä kuvassa 1. (Luo, Leenay ja Beisel, 2016)

Vaikka CRISPR-luokkien systeemeissä on iso kirjo proteiineja, Cas1 ja Cas2 löytyvät lähes kaikista systeemeistä (Kim *ym.*, 2013). Nämä kuitenkin puuttuvat tyypin IV systeemistä (Koonin ja Krupovic, 2015). Tyypin II CRISPR-Cas-systeemeillä on yleisten Cas1- ja Cas2-proteiinien lisäksi monitoiminen vaikuttajaproteiini Cas9 (ennen vuotta 2012 myös nimellä Cas5, Csn1 tai Csx12), sekä satunnaisesti myös neljäs proteiini Csn2 (tyyppi II-A) tai Cas4 (tyyppi II-B) (Chylinski *ym.*, 2014; Heler *ym.*, 2015; Wei, Terns ja Terns, 2015); kun taas esimerkiksi tyypin I ja III systeemeissä Cas9:n sijasta toimii kaskadiproteiinien kompleksi (engl. cascade: CRISPR-associated complex for antiviral defense), sekä ainakin yksi erillinen proteiini (Makarova *ym.*, 2015). Kuvasta 1 voi saada vain yleisen kuvan systeemeistä, koska siitä puuttuu nykyään löydettyjä proteiineja.

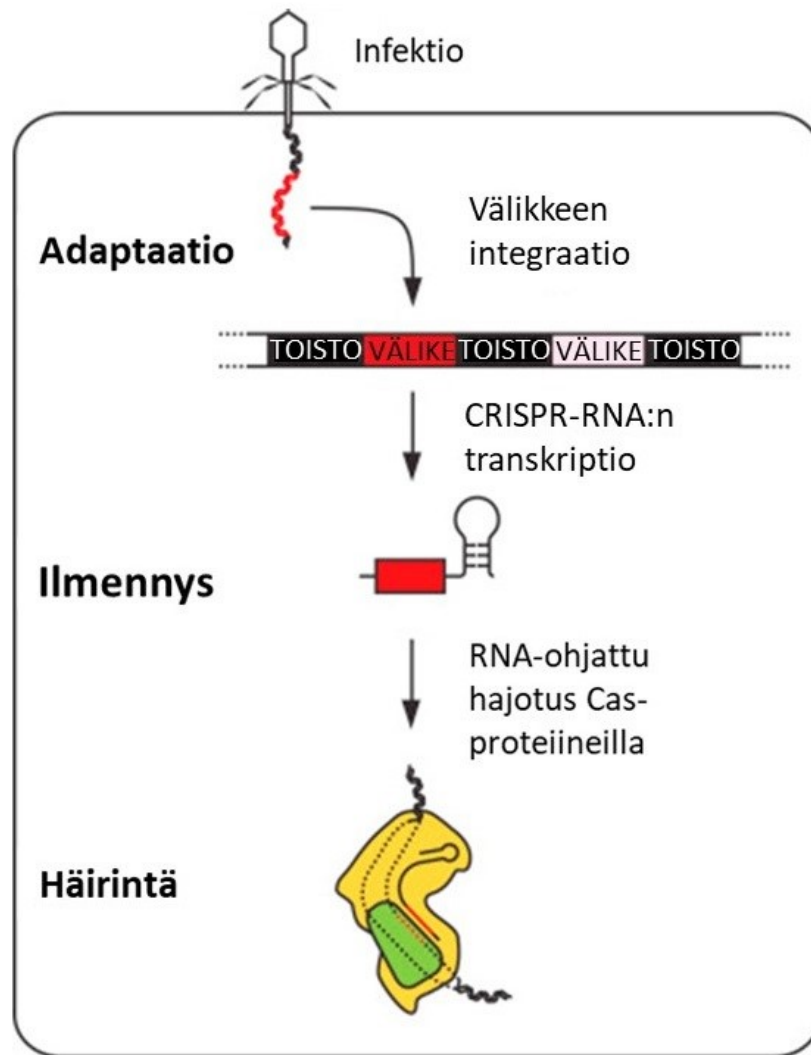
Tutkielmassa perehdytään systeemien proteiinien toimintaan ja näiden soveltuvuuteen geenieditointiin, mutta keskitytään lähinnä tyypin I, II ja III systeemeihin. Tyypin II systeemi on näistä tärkein lähinnä Cas9:n monitoimisuuden takia: kuvasta 1 voi nähdä miten tyypin II systeemin Cas9:llä on monta funktiota, jotka tyypin I, III ja IV systeemeissä tehdään lukuisien proteiinien muodostamalla kompleksilla.



Kuva 1 Tunnetut CRISPR-Cas-systeemit ja näiden proteiinit luokittain – Kuvassa on listattuna löydetty CRISPR-Cas-systeemityypit luokittain, ja näiden proteiinit rooliinsa perustuen. Katkoviivalla piirretyt proteiinit eivät ole systeemin toiminnalle pakollisia. Kuva on muokattu Makarova *ym.* (2015) artikkelista Springer Naturen luvalla.

2. CRISPR-immuunipuolustuksen toiminta

Lyhyesti kuvailtuna CRISPR-Cas-puolustusmekanismi toimii kolmiosaisesti (kuva 2). Ensin tapahtuu adaptaatio, jossa bakteerin genomissa sijaitsevaan CRISPR-alueeseen liitetään invasiiviseen nukleotidisekvenssiin homologisia välikkeitä. Toiseksi tapahtuu transkriptio ja ilmennys. Välikkeestä transkriptoidaan pitkä CRISPR-prekursori-RNA (pre-crRNA), jonka ilmennetyt Cas-proteiinit prosessoivat crRNA:ksi. Viimeisenä tapahtuu häirintä: invasiivinen DNA- tai RNA-sekvenssi tunnistetaan Cas-proteiinin ja crRNA:n kompleksin ohjaamana, ja Cas-proteiinit hajottavat tunnistetun nukleotidisekvenssin. (Rath *ym.*, 2015)



Kuva 2 CRISPR-Cas-systeemi on kolmiosainen – CRISPR-Cas-systeemin toiminta voidaan jakaa kolmeen perusvaiheeseen. Ensiksi adaptaatiossa CRISPR-alueeseen liitetään uusia välikkeitä. Tämän jälkeen ilmennyksessä tapahtuu CRISPR-RNA:n transkriptio ja RNA:n prosessointi. Viimeiseksi häirinnässä tapahtuu invasiivisen nukleotidisekvenssin tunnistus ja Cas-proteiineilla ohjattu hajotus. (Rath *ym.*, 2015) Kuva on muokattu Rath *ym.* (2015) artikkelista.

2.1 Adaptaatio

Adaptaatio on prosessi, jossa bakteeri tai arkki hankkii spesifisesti toimivan geneettisen muistin, jolla ylläpidetään puolustusta toistuvia nukleinihappohyökkäyksiä vastaan (Rath *ym.*, 2015). Adaptaatiossa prokaryootin genomisen DNA:n tai tämän plasmidin CRISPR-alueeseen integroidaan uusi välikesekvenssi, joka on homologinen faagin genomista peräisin olevaan invasiiviseen nukleotidisekvenssiin (Bolotin *ym.*, 2005;

Barrangou *ym.*, 2007). Kyseistä vierasta faagin tai plasmidin nukleotidisekvenssiä kutsutaan protovälikkeeksi (engl. protospacer, PS), jonka vieressä on adaptaation kannalta tärkeä PAM-sekvenssi (engl. protospacer adjacent motif) (Mojica *ym.*, 2009; Marraffini ja Sontheimer, 2010).

Kun välikkeet integroituvat genomisessa DNA:ssa sijaitsevaan CRISPR-alueeseen, kyseinen immunologinen muisti voi periytyä sukupolvelta toiselle (Rath *ym.*, 2015). Samalla resistenssi faageja vastaan periytyy, koska CRISPR-Cas-systeemi tunnistaa välikkeitä hyväksi käyttäen spesifisesti faagin genomin, antaen näin prokaryootille immuniteetin kyseistä faagia vastaan; puolustus siis adaptoituu faagipopulaation mukaan (Barrangou *ym.*, 2007). CRISPR-alueen välikkeiden perusteella suojaus toimii faagin nukleotidisekvenssin lisäksi plasmidi-DNA:ta vastaan (Bolotin *ym.*, 2005; Mojica *ym.*, 2005; Pourcel, Salvagnol ja Vergnaud, 2005).

2.1.1 Adaptaatioon tarvittavat Cas-proteiinit

Adaptaatioissa tapahtuva uusien välikkeiden hankinta on osoitettu usean eri tyypin CRISPR-Cas-systeemeissä (muun muassa I-A, I-B, I-E, I-F ja II-A), mutta adaptaation vaiheiden mekanismit ovat vain osittain tiedossa. Vaiheet ovat: välikkeen valinta, välikemateriaalin tuonti, integraatio ja uusien toistojaksojen synteesi, mutta tämän lisäksi CRISPR-alueen kokoa rajoittaa välikkeiden mahdollinen poistomekanismi. (Rath *ym.*, 2015)

Tyyppin I-E CRISPR-systeemissä (*E. coli*) Cas1 ja Cas2 proteiinit ovat vastuussa välikeintegraatiosta (Yosef *ym.*, 2012). Tämä osoitettiin Cas1:n ja Cas2:n yli-ilmentämisen avulla saadulla välikeintegraatiolla *E. coli* CRISPR-systeemin välikealueeseen (Yosef *ym.*, 2012). Cas1- ja Cas2-proteiinit muodostavat stabiilin kompleksin, jossa on kaksi Cas1:tä yhtä Cas2:ta kohti (nähtävissä kuvassa 5) (Nuñez *ym.*, 2014). Kompleksin muodostusta estävät mutaatiot estävät adaptaation, joten kompleksin muodostuminen on tärkeää välikkeen hankinnassa (Nuñez *ym.*, 2014). Vaikka sekä Cas1 että Cas2 ovat nukleaaseja (Beloglazova *ym.*, 2008; Nam *ym.*, 2012; Kim *ym.*, 2013), Cas2:n on havaittu olevan katalyyttisesti tarpeeton välikkeen integraatiossa, kun taas Cas1:llä on integraatiossa katalyyttisesti tärkeä rooli nukleaasiaktiivisuutensa takia (Nuñez *ym.*, 2014). Cas1-Cas2-kompleksi on myös todennäköisesti vastuussa välikkeiden kuljetuksesta CRISPR-alueelle (Rath *ym.*, 2015).

Eräät systeemit tarvitsevat adaptaatioon lisäkomponentteja: esimerkiksi tyypin II-A CRISPR-systeemi (*Streptococcus pyogenes*) tarvitsee invasiivisen nukleotidisekvenssin valintaan Cas1:n ja Cas2:n lisäksi Cas9:n, mutta myös systeemille ominaisen tracrRNA:n (engl. trans-activating crRNA). Näitä ilman hankinta ei tapahdu, mutta myös Csn2-apuproteiini tarvitaan mahdolliseen kompleksin muodostukseen. (Heler *ym.*, 2015; Wei *ym.*, 2015)

Tyypin I-E systeemin Cas1- ja Cas2-proteiineilla on havaittu taipumus välikkeiden hankintaan vieraasta DNA:sta (Levy *ym.*, 2015), kun taas tyypin II-A systeemillä ei ole huomattu erityistä taipumusta välikkeiden hankintaan vieraasta DNA:sta (Wei *ym.*, 2015). Tämän voisi todennäköisesti selittää tyypin I systeemille ominaisella pohjustetulla hankinnalla (engl. primed acquisition) (Fineran *ym.*, 2014).

2.1.2 Invasiivisen nukleotidisekvenssin valinta

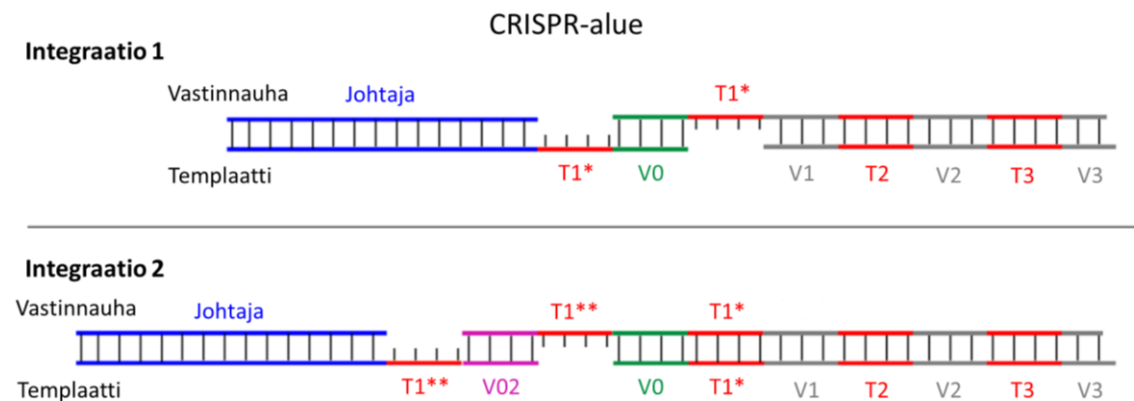
Adaptaatiossa lisätty välikesekvenssi ei ole sattumanvarainen, vaan tämä valitaan sekvenssispesifisesti (Hille ja Charpentier, 2016). Välikkeen hankinnassa integroituva välike tunnistetaan ja valitaan vieraan nukleotidisekvenssin (PS) PAM-alueen perusteella, ja PAM-sekvenssi samalla määrää integroituvan välikkeen suunnan CRISPR-alueessa (kuva 4) (Mojica *ym.*, 2009). Invasiivisessä DNA:ssa sijaitsevien PAM-alueiden tunnistus on tärkeä osa adaptaation välikkeen valinnassa ja hankinnassa, mutta myös häirinnässä (Leon, Mendoza ja Bondy-Denomy, 2018). Tyypin I ja II systeemit tarvitsevat 2 – 5 emäsparin pituisen PAM-sekvenssin, jonka Cas-proteiinit tunnistavat (Leon, Mendoza ja Bondy-Denomy, 2018). Kun taas esimerkiksi tyypin III systeemi on erilainen: toisin kuin tyypin I ja II systeemit, tyypin III systeemi ei etsi kohteestaan PAM-sekvenssiä (Marraffini ja Sontheimer, 2010).

Rath *ym.* (2015) mukaan tyypin I CRISPR-systeemissä Cas1 ja Cas2 todennäköisesti tunnistavat PAM-sekvenssin (Rath *ym.*, 2015). Tyypin II-A systeemissä taas Cas9 tunnistaa tracrRNA:n kanssa PAM:n Cas1:n ja Cas2:n sijaan, mutta ainakin Cas1 tarvitaan hankintaan nukleasiaktiivisuutensa takia (Heler *ym.*, 2015). Cas9-geenin mutaatio estää PAM-tunnistuksen ja koska tracrRNA:ta tarvitaan Cas9:n kofaktorina crRNA:n prosessointiin, myös deletio tracrRNA:ta ilmentävässä geenialueessa estää PAM-tunnistuksen (Heler *ym.*, 2015). Cas9-mutaation takia PAM-tunnistuksen puuttuessa CRISPR-alueeseen integroidaan välikkeitä ilman PAM-sekvenssin tunnistusta (Heler *ym.*, 2015).

Tunnistettava PAM-sekvenssi vaihtelee bakteerikantojen välillä eri Cas9-proteiinien ortologien mukaan (Hsu *ym.*, 2014). Bakteerikannoista *S. pyogeneksen* tyyppin II CRISPR-Cas-systeemin SpCas9:n (*S. pyogeneksen* Cas9; myös nimellä SpyCas9) sitoma PAM-sekvenssi (5'-NGG-3'; kirjain N on mikä tahansa nukleinihappoemäs) on niin sanottu kanoninen sekvenssi (Anders *ym.*, 2014).

2.1.3 Johtajasekvenssin tärkeys ja CRISPR-alueen toistojaksojen synteesi

Johtajasekvenssi (engl. leader sequence) on CRISPR-alueen konservoitunut sekvenssi, joka sijaitsee transkriptiosuunnan perusteella alueen ylävirrassa, kuten on esitetty kuvassa 4 (Rath *ym.*, 2015). Sekvenssi on tärkeä esimerkiksi tyyppin I-E ja II-A systeemeissä: osa johtajasekvenssistä ja vähintään yksi välikkeiden välinen toistojakso tarvitaan adaptaatioon (Yosef, Goren ja Qimron, 2012; Wei *ym.*, 2015). Uudet välikkeet liitetään CRISPR-alueen johtajasekvenssin alavirtaan (Swarts *ym.*, 2012) ja johtajasekvenssin viereistä ensimmäistä toistojaksosekvenssiä käytetään uusien välikkeiden välisten toistojaksojen templaattina (kuva 3) (Yosef, Goren ja Qimron, 2012). Ensimmäisen toistojakson sekvenssi siis periaatteessa monistuu. Välikkeiden liittämissijaintiin on huomattu poikkeuksia esimerkiksi *Sulfolobus islandicus*ssa (Erdmann ja Garrett, 2012).



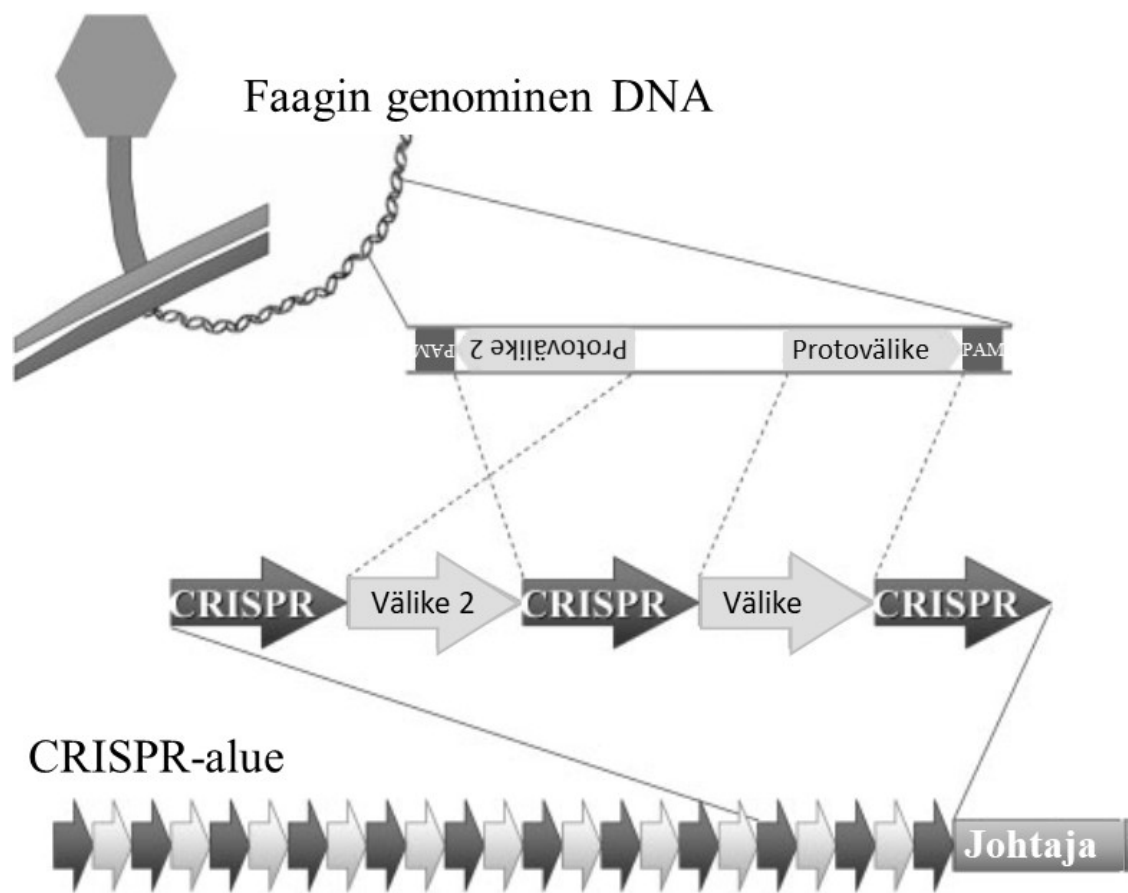
Kuva 3 Välikkeen integraatio CRISPR-alueeseen – Välikkeen lisäyksessä CRISPR-alue leikataan johtajan viereisen toistojakson (T1) kohdalta, jonka juosteiden väliin uusi välike integroidaan. Polymeerasi täyttää katkoskohdat integraatioiden välissä. Kuvassa toistojaksot ovat merkitty T:llä ja välikkeet V:llä. Kuva on muokattu Arslan *ym.* (2014) artikkelista.

2.1.4 Invasiivisen nukleotidisekvenssin integrointi isännän genomiin

Integraatiossa eliön CRISPR-alueen elementit pitää tunnistaa, jotta tiedetään mihin välike integroidaan ja mistä CRISPR-alueen sekvenssi pitää katkaista välikkeen liittämistä varten; integraatiota ennen Cas1-Cas2-kompleksi tunnistaa CRISPR-alueen ja valmistelee tämän välikkeen liittämistä varten (Rath *ym.*, 2015). Välikkeen integraatiomekanismista tiedetään kuitenkin vielä vähän (Hille ja Charpentier, 2016).

Genomissa CRISPR-alueen palindromisekvenssi muodostaa ristinmuotoisen sekundäärirakenteen (engl. cruciform), jolla on tärkeä rooli adaptaatiossa uusien välikkeiden hankinnassa (Nuñez *ym.*, 2015). CRISPR-toistojaksot tunnistetaan kyseisten sekundäärirakenteiden perusteella, mutta tämän lisäksi myös sekvenssispesifisesti (Nuñez *ym.*, 2015). Välikeintegraatiota katalysoivat Cas1-Cas2-kompleksit sitoutuvat ristin muotoiseen sekundäärirakenteeseen spesifisesti (kuva 5) (Nuñez *ym.*, 2015). Sitoutuneen kompleksin Cas1-entsyymi leikkaa metallikeskuksellaan CRISPR-alueen toistojakso-DNA:n spesifisesti ristin muotoisien sekundäärirakenteiden vierestä (Arslan *ym.*, 2014).

Kuten kuvassa 3 on esitetty, uusi CRISPR-alueeseen tuleva aluke liitetään ensimmäisen toistojakson jälkeen Cas1:n tekemien leikkauskohtien väliin (Arslan *ym.*, 2014). CRISPR-alueen templaattinauhan vastinnauha leikataan johtajan ja toistojakson välisestä liitoskohdasta, ja itse templaattinauha leikataan johtajan viereisen toistojakson ja vanhan välikkeen välisestä liitoskohdasta; leikkauksen jälkeen uuden välikkeen 3'-pää ligoidaan templaattinauhan toistojakson irtonaiseen 5'-päähän, kun taas vastinnauhan ligaatiossa uuden välikkeen 3'-pään liitetään leikatun toistojakson 5'-pää (Arslan *ym.*, 2014). Polymeraasi täyttää yksijuosteiset toistojaksot (kuva 5). Cas1 siis toimii esimerkiksi tyypin I-E systeemissä integraasina (Arslan *ym.*, 2014).



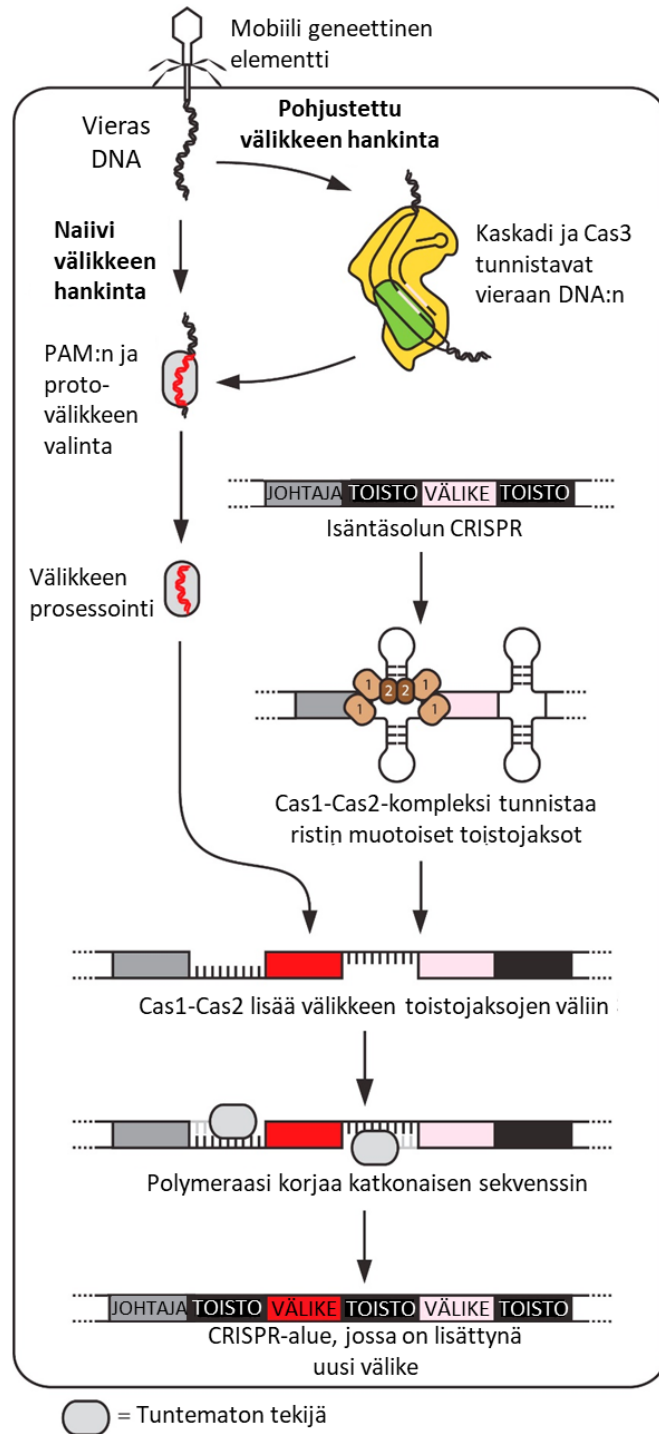
Kuva 4 Välikkeen suunta CRISPR-alueessa – Kuvassa on esitetty, miten CRISPR-alueeseen integroituvan faagin genomista johdetun välikkeen suunta valitaan PAM:n ja protovälikkeen (PS) orientaation perusteella. Välikkeet liitetään säännöllisten CRISPR-toistojen väliin ja johtajasekvenssin alavirtaan. Kuva on muokattu Mojica *ym.* (2009) artikkelista.

2.1.5 Pohjustettu hankinta tehostaa adaptaatiota

Tyypin I systeemin adaptaatio voidaan jakaa pohjustettuun välikkeen hankintaan ja naiiviin hankintaan. Pohjustetussa hankinnassa prokaryootti tunnistaa invasiivisen nukleotidisekvenssin PS:n, jonka tämä on aiemmin tunnistanut ja lisännyt välikkeenä CRISPR-välikealueeseensa. Vaihtoehtoisesti voidaan tunnistaa uusi sekvenssi, joka on homologinen vanhempaan välikkeeseen, kun taas naiivissa hankinnassa tunnistetaan täysin uusi sekvenssi. I-E-tyypin CRISPR-systeemissä naiivin välikkeen hankinnasta on vastuussa Cas1 ja Cas2, kun taas pohjustetussa hankinnassa käytetään Cas1:n ja Cas2:n lisäksi kaskadi-crRNA-kompleksia (engl. cascade-crRNA) ja Cas3-proteiinia. (Datsenko *ym.*, 2012; Fineran ja Charpentier, 2012; Swarts *ym.*, 2012)

Pohjustetussa hankinnassa tyypin I systeemien välikkeiden hankintaa tehostetaan aiemmin integroitujen välikkeiden perusteella; pohjustetussa hankinnassa tapahtuu siis niin sanotusti välikkeen rikastuminen, kun sama aiemmin havaittu PS lisätään välikkeenä useampaan otteeseen CRISPR-alueeseen (Datsenko *ym.*, 2012; Swarts *ym.*, 2012). Rikastuvaa välikettä vastaavan PS:n tunnistus tapahtuu PAM-spesifisesti aikaisemmassa infektiossa CRISPR-alueeseen liitetystä välikkeestä johdetun crRNA:n avulla (Datsenko *ym.*, 2012; Swarts *ym.*, 2012).

Toisaalta pohjustettu hankinta toimii myös kohdesekvensseille, joilla ei ole tunnistuksessa täyttä komplementaarisuutta (Fineran *ym.*, 2014; Hille ja Charpentier, 2016). Puolustus voi siis toimia myös eräitä faagien mutanteja vastaan, mutta ei jos kyseessä on PAM:n mutaatio (Fineran *ym.*, 2014). Pohjustetun hankinnan toiminta ilman täyttä komplementaarisuutta osoittaa, ettei pohjustuksessa välikkeitä valita hajotetusta sekvenssistä, koska itse hajotukseen (häirintään) tarvitaan tarkempi komplementaarisuus (Fineran *ym.*, 2014; Hille ja Charpentier, 2016). On kuitenkin osoitettu, että pohjustetulla hankinnalla on yhteys häirintään (Redding *ym.*, 2015).



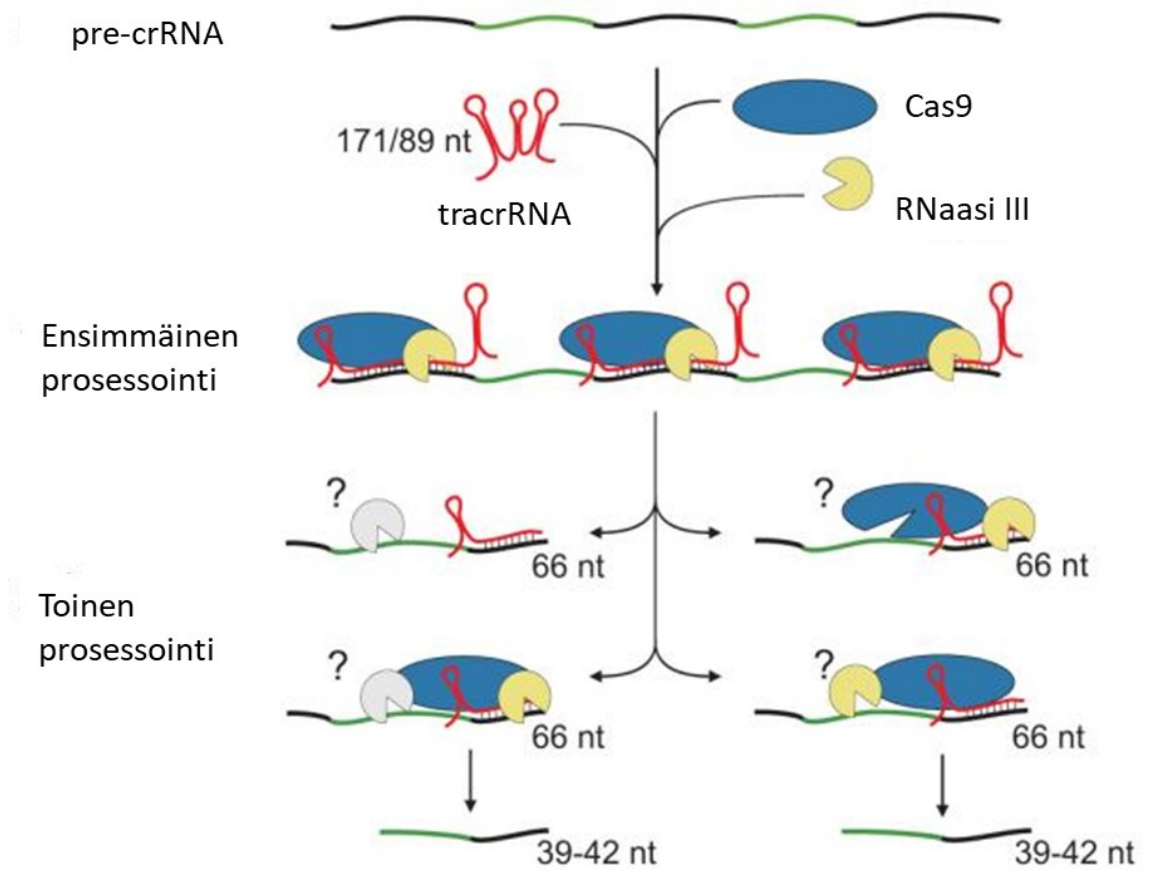
Kuva 5 CRISPR-välikkeiden hankinta, eli adaptaatio – Kuvassa on esitettyä Tyypin I CRISPR-systeemin naivi välikkeen hankinta ja pohjustettu välikkeen hankinta. Molempiin välikkeiden hankintaan tarvitaan Cas1-Cas2-proteiinikompleksi ja kohdesekvenssin viereinen PAM-sekvenssi. Kaskadi ja Cas3 tarvitaan kohde-DNA:n tunnistamiseen pohjustetussa hankinnassa. Kuva on muokattu Rath *ym.* (2015) artikkelista.

2.2 CRISPR-Cas alueen ilmentäminen ja crRNA-biogeneesi

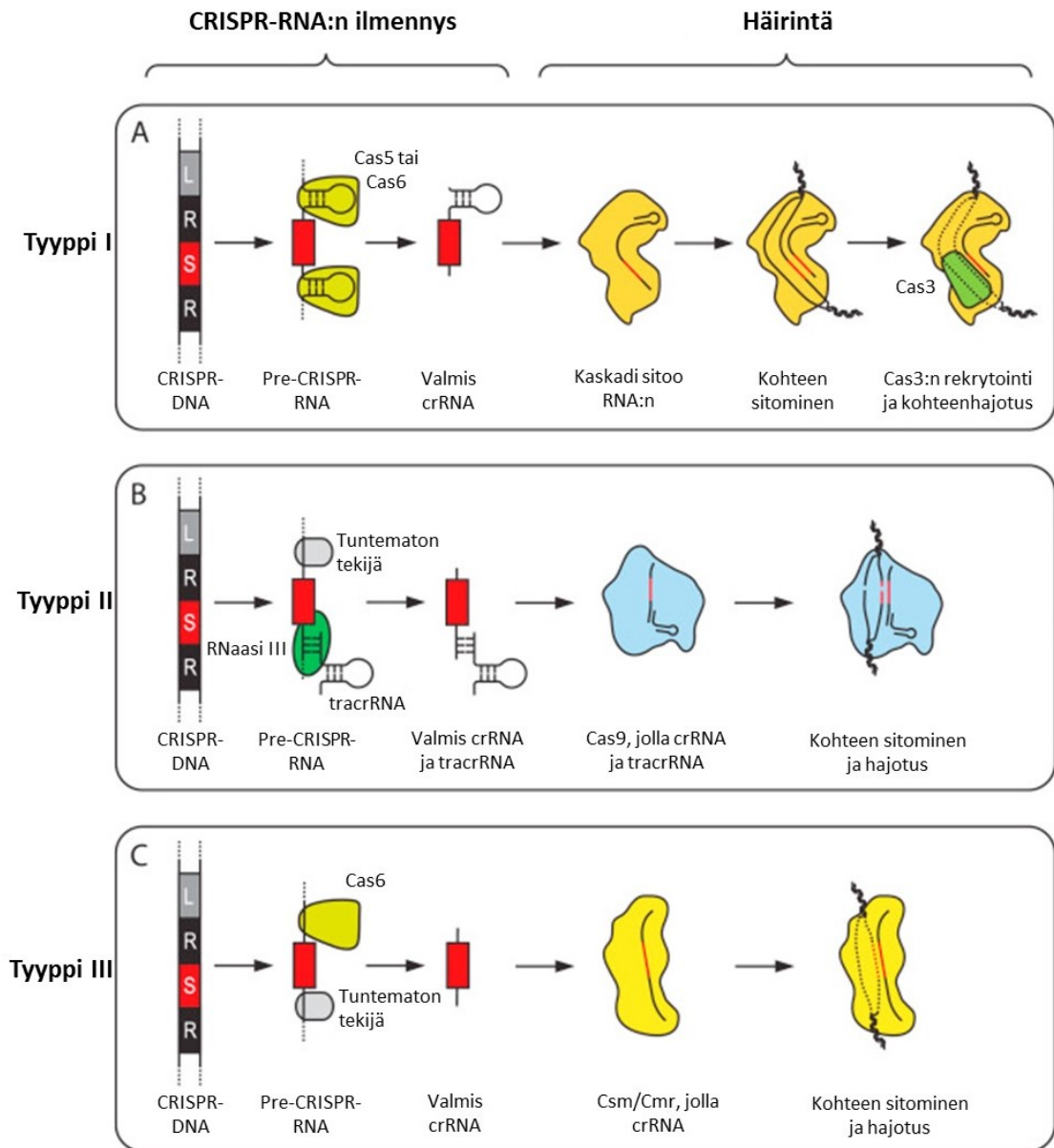
CRISPR-Cas-alueen transkriptio on samankaltainen useissa organismeissa, mutta syntyvissä RNA-proteiinikomplekseissa on eroavaisuuksia tyyppien välillä (kuva 7) (Rath *ym.*, 2015). Johtajasekvenssi on tärkeä CRISPR:n transkriptiossa (kuva 4), koska sekvenssi sisältää useasti CRISPR-systeemin promoottori alueet, ja alueet johon säätelyproteiinit sitoutuvat; useissa kannoissa CRISPR-alueen transkriptio alkaa johtajasekvenssistä, kuten esimerkiksi *E. colissa*, *Pyrococcus furiosuksessa* ja *Sulfolobus solfatarikuksessa* (Lillestøl *ym.*, 2009; Pul *ym.*, 2010; Hale *ym.*, 2012).

CRISPR:n RNA-biogeneesi tapahtuu siten, että CRISPR-alueesta transkriptoidaan pitkä prosessoimaton pre-crRNA, joka pilkkotaan pienemmiksi palasiksi Cas-proteiineilla muodostaen yksittäisiä erillisiä toistojaksojen välisiä välikkeitä (Rath *ym.*, 2015). Esimerkiksi tyyppin I-E systeemissä Cas6e pilkkoo pre-crRNA:n noin 60 nukleotidin pituiseksi crRNA:ksi (Brouns *ym.*, 2008). Palindromista sekvenssiä sisältävät pre-crRNA:t muodostavat mahdollisia prosessoinnin kannalta tärkeitä hiusneularakenteita, jota sitovat esimerkiksi tyyppin I systeemien Cas5- tai Cas6-proteiinit (Charpentier *ym.*, 2015).

Cas6 on vastuussa pre-crRNA:n prosessoinnista tyyppin I lisäksi myös tyyppin III systeemissä (Leon, Mendoza ja Bondy-Denomy, 2018). Tyyppin II systeemin crRNA-biogeneesissä pre-crRNA:n prosessointiin tarvitaan poikkeuksellisesti RNAasi III (Cas-proteiineista poikkeava ribonukleaasi), tracrRNA ja myös Cas9-proteiini (Deltcheva *ym.*, 2011). Tyyppin II systeemeille ominainen tracrRNA on pieni ei-koodaava RNA, joka käynnistää RNAasi III:n vastuulla olevan pre-crRNA-prosessoinnin, mutta ottaa osaa myös häirintään (Jinek *ym.*, 2012). CRISPR:n crRNA-kypsymiseen tarvittavalla tracrRNA:lla on crRNA:n kanssa noin 24 nukleotidin pituinen komplementaarinen osa ja on lähes täydellisesti komplementaarinen prosessoitavan pre-crRNA:n kanssa (Deltcheva *ym.*, 2011). Komplementaarisuutta käytetään hyväksi siten, että RNAasi III tunnistaa ja prosessoi hajottamalla kaksijuosteisessa tracrRNA-pre-crRNA-dupleksissa olevan pre-crRNA:n 66 nukleotidin pituisiksi pätkiksi ensimmäisessä prosessointivaiheessa, mutta RNA-prekursorin prosessointivaihe valmiiksi crRNA:ksi (noin 40 nukleotidin pituinen) on vielä tuntematon (kuva 6) (Deltcheva *ym.*, 2011).



Kuva 6 Prekursori-RNA:n prosessointi Tyypin II systeemissä – Tyypin II systeemissä CRISPR-alueesta ilmennetty pre-crRNA prosessoidaan crRNA:ksi Cas9:n, tracrRNA:n, RNAasi III:n avulla. Toinen prosessointivaihe on vielä tuntematon: prosessoinnissa kuitenkin mahdollisesti toimii myös tuntematon tekijä. Kuva on muokattu Deltcheva *ym.* (2011) artikkelista Springer Naturen luvalla.



Kuva 7 CRISPR-RNA:n ilmentäminen ja kohteen häirintä tyypittäin – A: Tyyppi I:ssä pre-CRISPR-RNA:n prosessoi Cas5 tai Cas6 ja kohdennetun DNA:n häirintään tarvitaan Cas3. B: Tyyppi II:ssä Pre-CRISPR-RNA:n prosessoi RNAasi III ja tracrRNA, mutta myös Cas9 (ei näytetty). Tämän lisäksi tuntematon tekijä katkaisee 3'-pään. Kohteena olevan DNA:n häirintä tapahtuu Cas9:n ja crRNA:n ohjaamana. C: Tyyppi III:ssä Pre-CRISPR RNA:n prosessoi Cas6, ja tässäkin systeemissä tuntematon tekijä katkaisee 3'-pään. Csm- tai Cmr-kompleksi ja crRNA tarvitaan kohteen hajottamiseen (joko DNA tai RNA). (Rath *ym.*, 2015) Kuva on muokattu Rath *ym.* (2015) artikkelista.

2.3 Vieraan nukleotidisekvenssin häirintä

CRISPR-Cas-systeemin ohjaama häirintä on prosessi, jossa invasiivinen kohdesekvenssi tunnistetaan ja hajotetaan tuotetun crRNA:n ja Cas-proteiinien toimesta. Tässä Cas-proteiineihin sitoutuneena crRNA tunnistaa kohdesekvenssin tämän PAM-alueeseen sitoutuessaan, näin johtaen spesifisten Cas-proteiinien katalysoimaan kohdesekvenssin hajotukseen. (Brouns *ym.*, 2008; Garneau *ym.*, 2010)

2.3.1 Tyypin I systeemi

Tyypin I systeemeistä tähän asti on muun muassa näytetty I-E- ja I-F-systeemien häirintämekanismit; *E. coli*n tyypin I CRISPR-systeemien aiheuttamassa häirinnässä nukleaasi-helikaasi Cas3 hajottaa kohdesekvenssinsä usean proteiinin muodostaman tunnistuskompleksin (engl. surveillance complex), eli kaskadiproteiinikompleksin, tunnistamana (Hochstrasser *ym.*, 2014; Leon, Mendoza ja Bondy-Denomy, 2018). Tyypin I-E-systeemille ominaisessa niin sanotussa CasABCDE-kaskadiproteiinikompleksissa tärkeimmät komponentit ovat Cas6e (CasE) ja crRNA, joiden lisäksi kompleksissa on yksi Cas8 (CasA), kaksi Cse2:ta (CasB), kuusi Cas7:ää (CasC), sekä yksi Cas5e (CasD) (Jore *ym.*, 2011; Rath *ym.*, 2015).

I-E-systeemissä kaskadiproteiinikompleksin alayksikkö Cas8 kutsuu Cas3:n paikalle kohdesekvenssispesifisesti; Cas8 muuttaa konformaatiotaan sitoessaan kohdesekvenssinsä PAM-sekvenssin, näin kutsuen paikalla kohdesekvenssin hajottavan Cas3:n (kuva 7) (Hochstrasser *ym.*, 2014; Xue, Whitis ja Sashital, 2016). Kohdesekvenssin hajotuksessa Cas3 hajottaa eksonukleaasina kohdesekvenssin 3' → 5'-suuntaan (Sinkunas *ym.*, 2011). Tunnistettu kohdesekvenssi on lähes kokonaan komplementaarinen kaskadiin sidotun crRNA:n kanssa (Leon, Mendoza ja Bondy-Denomy, 2018).

Cas8:lla on aukinainen ja suljettu konformaatio, jotka ovat dynaamisessa reaktiotasapainossa. Suljetussa konformaatiossa Cas8 kutsuu Cas3:n paikalle DNA:n pilkkomista varten, kun taas avoimessa konformaatiossa Cas8 ajaa Cas3:n liittymistä kompleksiksi Cas1:n ja Cas2:n kanssa. Kyseinen kompleksin ajaa todennäköisesti myös adaptaatiossa uusien välikkeiden pohjustettua hankintaa. (Redding *ym.*, 2015; Xue, Whitis ja Sashital, 2016)

Koska tunnistus sijoittuu PAM-sekvenssin sisältämään nukleotidisekvenssiin (Semenova *ym.*, 2011; Wiedenheft *ym.*, 2011) ja PAM-sekvenssit puuttuvat tyypin I CRISPR-systeemien CRISPR-alueen sekvensseistä, eivät häirinnästä vastuussa olevat Cas-proteiinit hyökkäävät näin bakteerin genomissa (tai plasmidissa) sijaitsevaan omaan CRISPR-alueeseen (Hsu *ym.*, 2014).

2.3.2 Tyypin II systeemi

Tyypin II CRISPR-systeemin tekemä häirintä on geneettisen tutkimuksen kannalta tärkeä prosessi, koska systeemi on yksinkertainen käyttää geenieditoinnissa esimerkiksi tyypin I systeemiin verrattuna (Rath *ym.*, 2015). Tyypin II systeemien yksinkertaisuus johtuu siitä, että systeemit eivät tarvitse kaskadiproteiinikompleksia kohdesekvenssin häirintään vaan pelkästään Cas9-proteiinin kofaktoreineen (Deltcheva *ym.*, 2011). Häirinnässä Cas9 tarvitsee kohteensa tunnistuksessa ja hajotuksessa kofaktoreikseen opastaja-RNA:t (engl. guide RNA), eli crRNA:n, sekä muista tyypeistä puuttuvan tracrRNA:n (Deltcheva *ym.*, 2011). Nämä pariutuvat häirinnässä Cas9:n kanssa RNA-dupleksina (kuva 7) (Jinek *ym.*, 2012). Geenieditoinnin helpottamiseksi Cas9:n ohjaamista varten on kehitetty sgRNA (engl. single guide RNA), joka on tracrRNA:n ja crRNA:n kimeerinen fuusio-RNA (Jinek *ym.*, 2012).

Häirinnässä Cas9 tunnistaa kohde-DNA:n opastaja-RNA:ta käyttäen tunnistuslohkollaan ja katkaisee tämän molemmat juosteet (engl. double-stranded break, DSB) nukleaasilohkon domeeneillaan (Gasiunas *ym.*, 2012; Jinek *ym.*, 2012; Nishimasu *ym.*, 2014). Kohdesekvenssissä pitää olla Cas-proteiinin (Cas9) tunnistama PAM-sekvenssi, kuten tyypin I systeemissä (Deveau *ym.*, 2008; Sternberg *ym.*, 2014). Tämän lisäksi tyypin I systeemin tavoin PAM-sekvenssit eivät integroidu CRISPR-alueeseen toksisuuden estämiseksi, jotta Cas9 ei hyökkäisi bakteerin omaan genomiseen DNA:han (Hsu *ym.*, 2014). Häirinnästä vastuussa olevan Cas9-proteiinin ominaiseen rakenteeseen ja tämän toimintaan perehdytään tarkemmin omassa luvussaan.

2.3.3 Tyypin III systeemi

Tyypin III CRISPR-systeemillä on ainakin kolme erilaista nukleaasia faageja vastaan puolustautumiseen: DNAasi Cas10, sekä Csm- ja Cmr-RNAasit; proteiinit aktivoituvat vain invasiivisen nukleotidisekvenssin ollessa läsnä, joka havaitaan spesifisillä tunnistusmekanismeilla (Leon, Mendoza ja Bondy-Denomy, 2018). Tyypin III systeemi tekee kohdennuksen RNA:n emäspariutumisella DNA:n sijaan, mutta hajottaa RNA:n

lisäksi DNA:ta (Osawa *ym.*, 2015; Kazlauskienė *ym.*, 2016). Systeemien Csm- tai Cmr-kompleksi katalysoi RNA:n hajotuksen (kuva 7), ja Cas10 yksijuosteisen DNA:n hajotuksen (Osawa *ym.*, 2015; Kazlauskienė *ym.*, 2016). Tyypin III-B systeemissä RNA:n hajotuksesta on vastuussa Cmr-kompleksi (proteiinit Cmr1 – Cmr6) (Osawa *ym.*, 2015). Tyypissä III-A RNA:n hajotuksesta on taas vastuussa Csm-proteiinikompleksi (koostuu Csm1 – Csm5 proteiineista), mutta kompleksin on osoitettu myös tunnistavan kohde-RNA:n crRNA:n avulla, sekä kutsuvan paikalle Cas10:n (Kazlauskienė *ym.*, 2016).

Kuten edellä on mainittu, tyypin III systeemi ei käytä PAM-sekvenssiä, joten prokaryootin oman genomien kohdentaminen pitää estää toisella tavalla; tyypin III systeemillä on pidentynyt crRNA, joka inaktivoi hajotuksen tunnistessaan genomien oman sekvenssin (kokonaan komplementaarinen kohde-RNA) (Marraffini ja Sontheimer, 2010; Leon, Mendoza ja Bondy-Denomy, 2018). Sekvenssiä valittaessa oikean kohdesekvenssin transkriptoidun RNA:n on siis oltava komplementaarinen crRNA:n kanssa, mutta DNA:n hajotuksessa kohde-RNA:lla ei saa olla komplementaarisuutta pidentyneen crRNA:n 5'-pään kanssa (Marraffini ja Sontheimer, 2010; Leon, Mendoza ja Bondy-Denomy, 2018). Esimerkiksi tyypin III-A systeemissä kyseinen 5'-pään komplementaarisuus saa Csm-kompleksin inhiboimaan Cas10:n, näin estäen DNA:n pilkkomisen (Kazlauskienė *ym.*, 2016). RNA-kohdennuksen takia häirinnässä olevan kohde-DNA:n transkriptioaktiivisuuden pitää olla läsnä (Samai *ym.*, 2015).

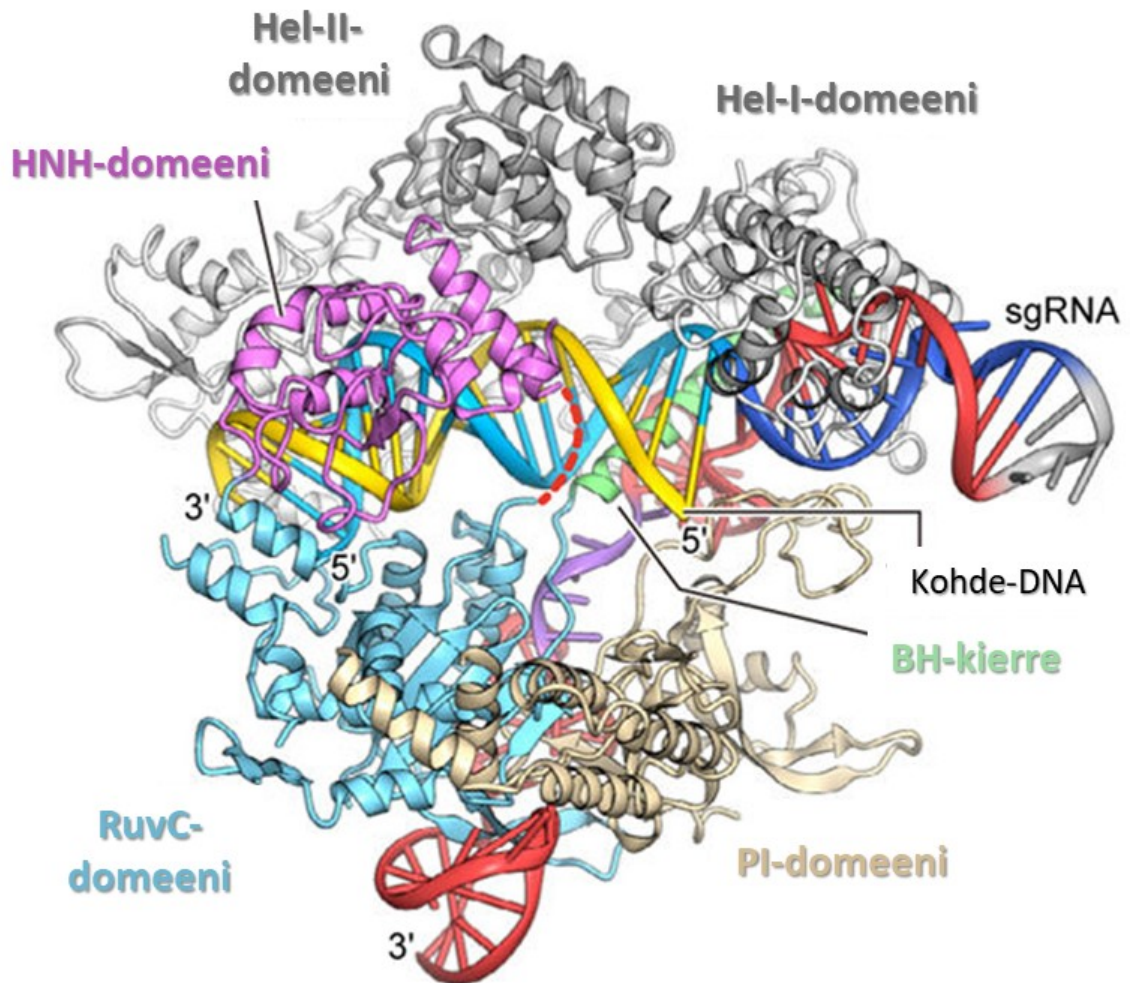
3. Cas9-entsyymin rakenne ja toiminta

3.1 Cas9:n yleinen rakenne

*S. pyogenes*in Cas9-entsyymin aktiivisen rakenteen kuvaamiseen on käytetty kahta eri opastaja-RNA:ta: Jinek *ym.* (2014) käyttivät luonnollisesti Cas9:ää ohjaavia crRNA:ta ja tracrRNA:ta, kun taas Nishimasu *ym.* (2014) käyttivät synteettistä sgRNA:ta havainnollistamaan miten geenieditoinnissa Cas9-sgRNA-kompleksi leikkaa kohde-DNA:n. Luvussa perehdytään Jinek *ym.* (2014) julkaisemaan, mutta Jiangin ja Doudnan (2017) tulkitsemaan proteiinirakenteeseen (kuva 9), sekä Nishimasu *ym.* (2014) rakenteeseen (kuva 8), mutta jälkimmäisestä rakenteesta puuttuu domeeni (REC-lohkosta).

Nishimasu *ym.* (2014) kuvaamassaan proteiinikompleksissa SpCas9 sitoo sgRNA:n ja tämän kohde-DNA:n. Cas9-entsyymillä on kaksi havaittua konformaatiota: aktiivinen holo-muoto, jossa crRNA ja tracrRNA (tai sgRNA) ovat sidottuna, sekä apo-muoto (Jinek *ym.*, 2014; Nishimasu *ym.*, 2014). SpCas9:n apo-muoto on niin sanotussa inaktiivisessa autoinhiboidussa tilassa, johon edellä mainitut opastaja-RNA-kofaktorit eivät ole sidottuna; Cas9 saa opastajien sitoutumisen yhteydessä aktiivisen DNA:ta kohdentavan konformaation (Jinek *ym.*, 2014).

Cas9:llä on kaksiosainen rakenne: tunnistuslohko (REC) ja nukleaasilohko (NUC; kuva 9), ja näitä yhdistää pitkä siltamainen α -kierre (engl. bridge helix, BH; kuva 8 ja 9) ja rakenteeton linkkeri (engl. disordered linker; kuva 8). Cas9:n yleinen rakenne on kuvattuna kuvassa 9 (apo-muoto). Proteiinin REC-lohko voidaan jakaa kolmeen domeeniin: Hel-I, Hel-II ja Hel-III (Hel, eli helikaasi; kuva 9). REC:n ja NUC:n välisen kolon positiivinen pinta sitoo sgRNA-kohde-DNA-heterodupleksin (kuva 8). NUC-lohko voidaan jakaa RuvC-domeeniin, HNH-domeeniin ja PAM:n kanssa vuorovaikuttavaan PI-domeeniin (engl. PAM-interacting; myös nimellä C-terminaalinen domeeni, CTD). RuvC-domeeni taas voidaan jakaa kolmeen motiiviin (I-III). RuvC:n interaktio PI-domeenin kanssa muodostaa positiivisesti varautuneen pinnan, johon negatiivisesti varautuneen sgRNA:n 3'-pää sitoutuu. HNH-domeeni on RuvC II:n ja III:n välissä. (Nishimasu *ym.*, 2014; Jiang ja Doudna, 2017)

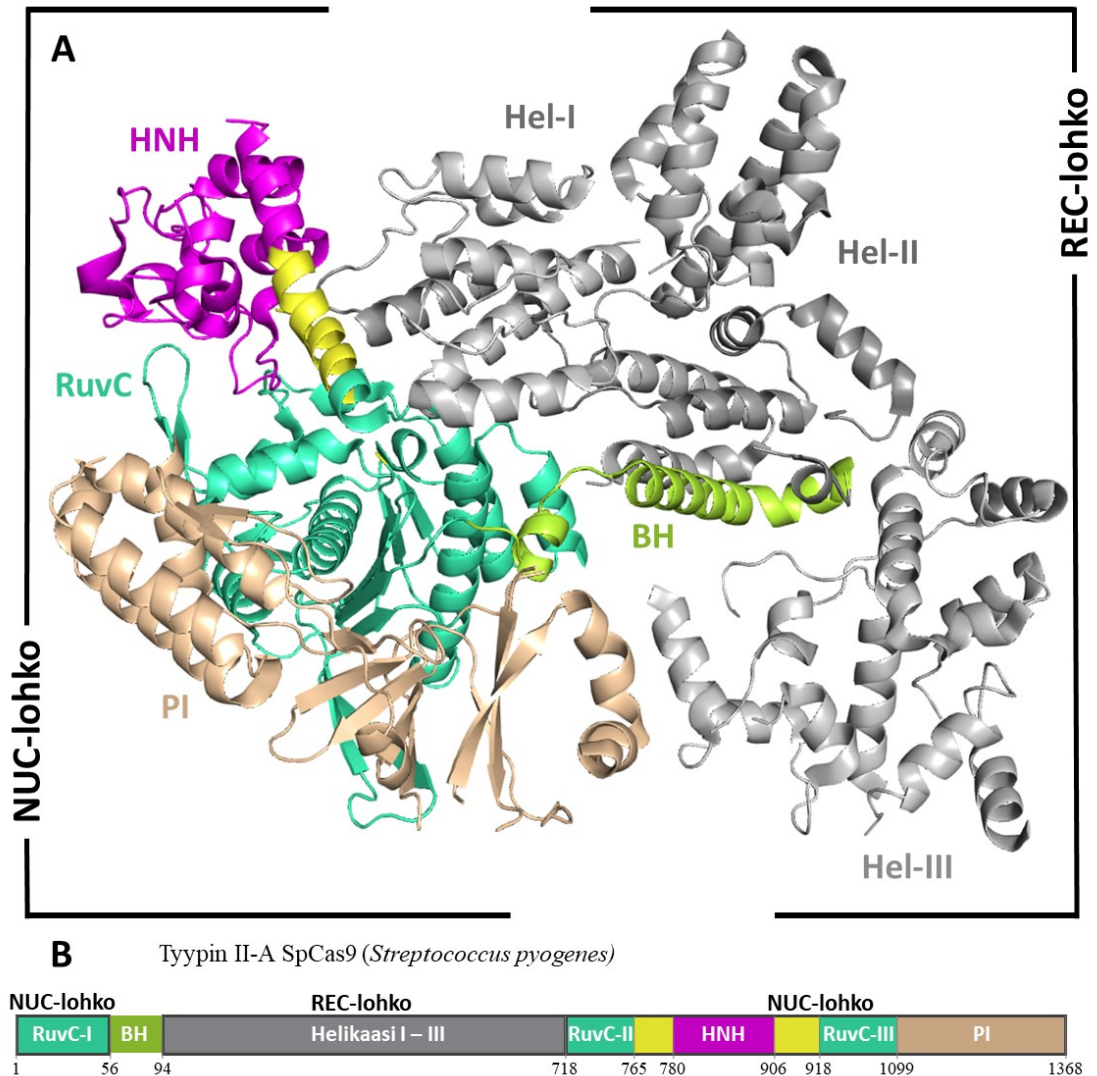


Kuva 8 SpCas9-proteiinin holo-muoto – Kuvassa on SpCas9:n kristallisaatiossa Nishimasu *ym.*, (2014) saama holo-muodon 3D rakenne, jossa domeenit ja rakenteet ovat merkitty eri väreillä. Rakenteessa Cas9 sitoo kohde-DNA:sgRNA-dupleksin. Lohkoja yhdistävä rakenteeton linkkeri on merkitty punaisella katkoviivalla ja yhdistävä BH-kierre on vaalean vihreä. Rakenteesta puuttuu Hel-III-domeeni. Kuva on muokattu Nishimasu *ym.* (2014) artikkelista.

3.2 Tunnistuslohko

Tunnistuslohkolla, eli REC-lohkolla on kolme domeenia (Hel-I – Hel-III), ja Hel-I:llä on pitkä α -kierteistä muodostuva rakenne, joka koostuu 25:stä α -kierteestä ja kahdesta β -laskoksesta. Hel-II:lla taas on kimpun kaltainen kuusikierteinen rakenne. Nishimasu *ym.* tekemän Dali-palvelinhaun (Holm *ym.*, 1993) mukaan REC-lohkon rakenne ei ole samankaltainen muiden tunnettujen proteiinien kanssa, jonka perusteella kyseessä on Cas9:n toiminnalle spesifinen rakenne. Lohkon alussa oleva Hel-I sitoo opastaja-RNA:kohde-DNA-heterodupleksin (engl. guide:target heteroduplex), mutta Hel-II ei ota osaa sitomiseen. (Nishimasu *ym.*, 2014; Jiang ja Doudna, 2017)

REC-lohkon Hel-III-domeeni siirtyy NUC-lohkon HNH-domeenia kohti sgRNA:n sitomisessa ison konformaatiomuutoksen yhteydessä, kun taas PAM:in ja DNA:n sitominen aiheuttaa vain pienen Cas9:n konformaatiomuutoksen. Tämä viittaa siihen, että ensin sitoutuva opastaja-RNA (sgRNA tai crRNA:tracrRNA-dupleksi) toimii Cas9:n säätelijänä. (Jinek *ym.*, 2014; Jiang ja Doudna, 2017)



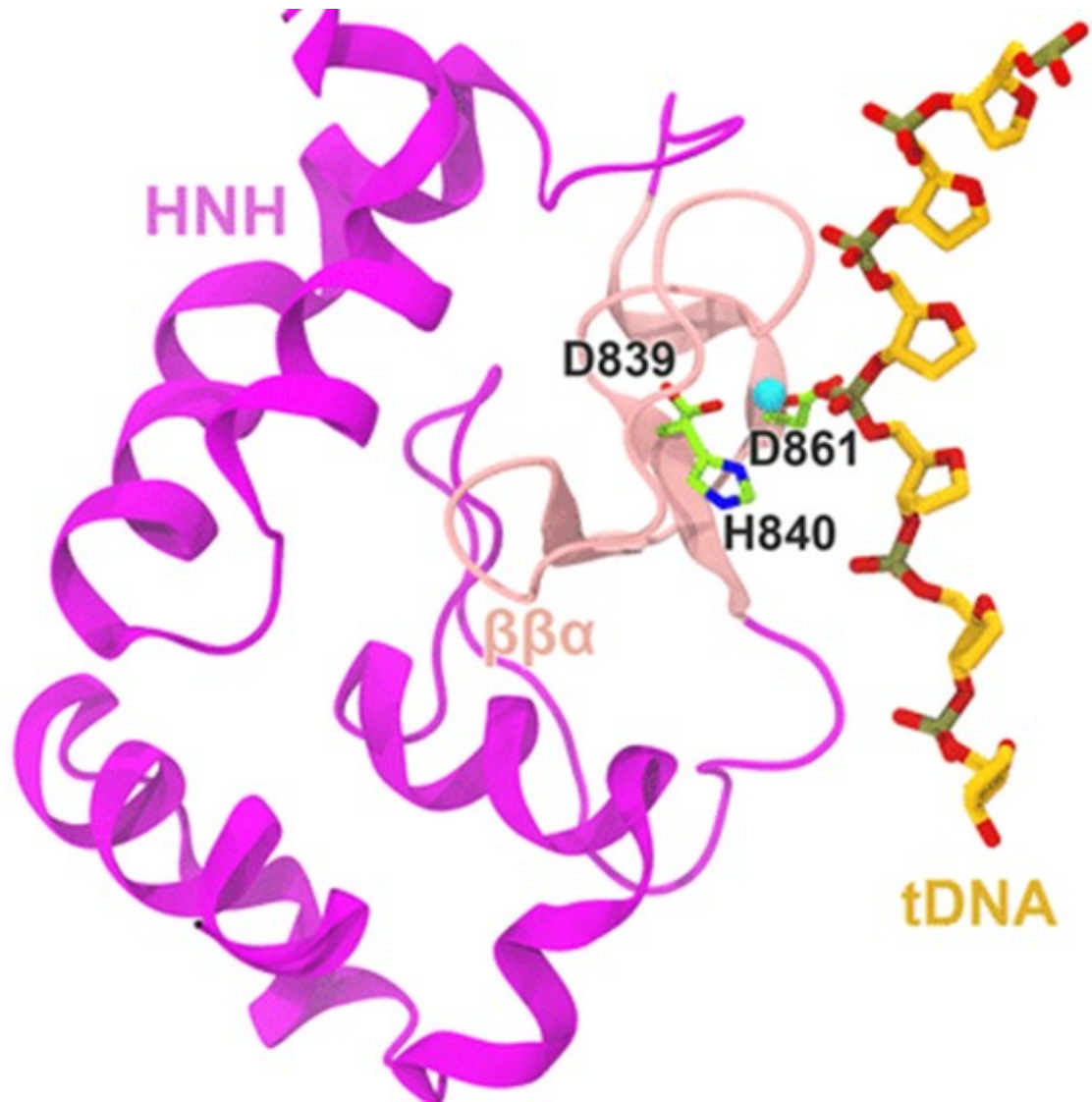
Kuva 9 SpCas9-proteiinin kaavio ja kokonainen apo-muoto – A: Jinek *ym.*, (2014) julkaisema SpCas9:n apo-muodon kristallirakenne [PDB (engl. Protein Data Bank) koodi 4CMP]. B: kaavio proteiinin geenialueesta lohkojen ja domeenien perusteella. Kuvassa Cas9:n domeenit ovat kuvattuna eri väreillä. Niin sanotut linkkerialueet ovat merkittyinä keltaisella. Kuva on tehty Jiangin ja Doudnan (2015) artikkelin perusteella.

3.3 Nukleaasilohko

Nukleaasilohkossa, eli NUC-lohkossa sijaitsee PI-domeeni, jolla on REC-domeenin tavoin ainutlaatuinen vain Cas9-proteiiniperheessä tunnettu rakenne (Nishimasu *ym.*, 2014). PI-domeeni on pitkä rakenne, joka koostuu seitsemästä α -kierteestä, yhdestä kolmiketjuisesta vastakkaissuuntaisesta β -levystä, yhdestä viisiketjuisesta vastakkaissuuntaisesta β -laskoksesta ja yhdestä kaksiketjuisesta vastakkaissuuntaisesta β -laskoksesta (Nishimasu *ym.*, 2014). Sidotun komplementaarisen DNA-juosteen asento ja RuvC-domeenin aktiivisen keskuksen asento viittaavat siihen, että asentonsa perusteella PI-domeeni tunnistaa ei-komplementaarisen DNA-juosteen PAM-sekvenssin (Nishimasu *ym.*, 2014). PI-domeenin toiminnalliseen tärkeyteen myös viittaa se, että domeenin deleetio estää Cas9:n DNA-katkaisuaktiivisuuden (Nishimasu *ym.*, 2014). PI-domeeni on Cas9:n apo-muodossa rakenteeton, viitaten edellä mainittuun apo-muodon inaktiivisuuteen (Jiang ja Doudna, 2017).

NUC-lohkossa kaksi domeenia ovat vastuussa kohde-DNA:n leikkaamisesta: HNH, joka leikkaa crRNA:han komplementaarisen DNA-sekvenssin ja RuvC, joka leikkaa ei-komplementaarisen sekvenssin; tämä johtaa molempien juosteiden niin sanottuun tylppäpäiseen (engl. blunt) DSB-katkokseen, jossa molemmat DNA-päät ovat saman pituiset (Gasiunas *ym.*, 2012; Jinek *ym.*, 2012). Jommankumman domeenin aktiivisen keskuksen mutaatio (HNH:n His840 tai RuvC:n Asp10 alaniiniksi) taas saa aikaan sen, että vain yksi DNA-juoste katkaistaan (engl. nicking) (Gasiunas *ym.*, 2012; Jinek *ym.*, 2012).

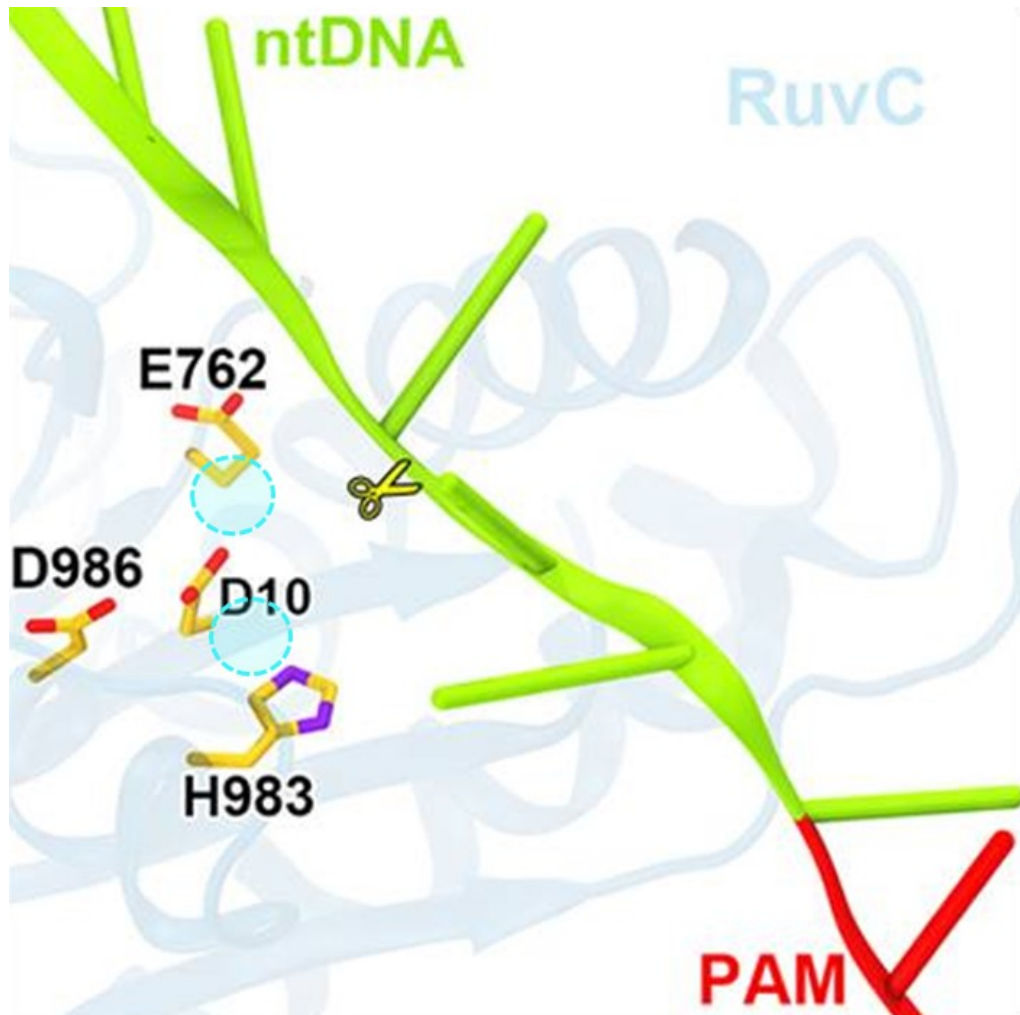
HNH-domeeni koostuu kaksijuosteisesta antiparalleelista β -levystä ja tämän sivussa on neljä α -kierrettä (Nishimasu *ym.*, 2014). Domeeni on HNH-endonukleaasien kanssa rakenteellisesti samankaltainen $\beta\beta\alpha$ -metalli-laskos (Nishimasu *ym.*, 2014). Tällainen laskos on esimerkiksi faagin T4-endonukleaasi VII:llä, jonka takia HNH-domeeni on 20 % identtinen kyseisen T4-endonukleaasin kanssa (PDB koodi 2QNC) (Biertümpfel, Yang ja Suck, 2007; Nishimasu *ym.*, 2014). HNH-domeenin aktiivisessa keskuksessa on seuraavat katalyyttiset aminohapot: His840, Asn854 ja Asp839, mutta myös Mg^{2+} -ioni (kuva 10) (Zuo ja Liu, 2017). Kuvassa 10 on esitettyinä HNH-domeenin $\beta\beta\alpha$ -metalli-laskos. Aktiivisen keskuksen aminohapot, sekä Tyr823 ja Arg864 vastaavat T4-endonukleaasin aktiivisen keskuksen aminohappoja (Zuo ja Liu, 2017). HNH-domeeni leikkaa kohde-DNA:n nukleiinihapposekvenssin (komplementaarisen sekvenssin) yhden metalli-ionin mekanismilla, kuten T4-endonukleaasi VII (Nishimasu *ym.*, 2014).



Kuva 10 SpCas9-HNH-domeenin aktiivinen keskus – Kuvassa on esitettyä SpCas9:n HNH-domeenin Zuon ja Liun (2017) tMD-simulaatiolla (engl. targeted molecular dynamics) saama aktiivinen keskus: His840 (H840), Asp839 (D861) ja Asp861 (D861), sekä Mg^{2+} -ioni (sininen pallo). Kuvassa on esitettyä HNH:n $\beta\beta\alpha$ -metalli-laskos (pinkki) sitoutuneena kohde-DNA:n komplementaariseen juosteeseen (kuvassa tDNA). Kuva on muokattu Zuon ja Liun (2017) artikkelista.

RuvC-domeeni taas koostuu kuuden sekalaisesti järjestyneen juosteen β -levystä, jonka sivussa on 7 α -kierrettä ja lisäksi kaksi kaksi-juosteista antiparalleelia β -levyä (Nishimasu *ym.*, 2014). Domeenilla on samankaltainen rakenne retroviruksien integraasisuperperheen proteiinien kanssa, joilla on RNAasi H-laskos, kuten esimerkiksi *E. colin* RuvC-proteiini (PDB koodi 1HJR, 14 % identtinen) (Ariyoshi *ym.*, 1994; Nishimasu *ym.*, 2014). RuvC:n aktiivisessa keskuksessa on seuraavat katalyyttiset aminohapot: Asp10, Glu762, His983 ja Asp986, sekä kaksi magnesium-ionia (kuva 11)

(Nishimasu *ym.*, 2014; Zuo ja Liu, 2016). RuvC-domeeni leikkaa kohdesekvenssin ei-komplementaarisen juosteen kahden metalli-ionin mekanismilla (kuva 11), kuten edellä mainitut retroviruksien nukleasit (Jinek *ym.*, 2014; Nishimasu *ym.*, 2014). Molemmissa HNH:n ja RuvC:n aktiivisissa keskuksissa on yhden ja kahden metalli-ionin mekanismille tyypilliset aminohapot: aspartaatti (Asp) ja histidiini (His) (Yang, 2008).



Kuva 11 SpCas9-RuvC-domeenin aktiivinen keskus – Kuvassa on esitettyä SpCas9:n RuvC-domeenin (PDB koodi 4CMQ) aktiivinen keskus: Asp10 (D10), Glu762 (E762), His983 (H983) ja Asp986 (D985). Sinisellä katkoviivalla on merkitty Mg²⁺-ionien sijainti suurin piirtein. Aktiivinen keskus sitoo ja katkaisee (kuvassa sakset) kohde-DNA:n ei-komplementaarisen juosteen (kuvassa ntDNA). Kuvassa on myös merkittynä PAM-sekvenssi. Kuva on muokattu Zuon ja Liun (2016) artikkelista.

4. CRISPR-Cas ja geneettinen muokkaus

4.1 CRISPR-systeemin löytö geenieditoinnin kannalta

Tyypin II CRISPR-Cas-systeemiä tutkimalla huomattiin, että systeemi soveltuu geenieditointiin. Ensin huomattiin, että systeemin Cas9 tekee DNA:n kaksoisjuosteeseen DSB:itä (Garneau *ym.*, 2010). Sitten huomattiin *Streptococcus thermophiluksen* tyypin II CRISPR-systeemin (nimellä CRISPR3/Cas) olevan modulaarinen, eli kyseinen systeemi voidaan siirtää bakteerikannasta toiseen toimivana (*E. coli*); systeemin Cas9-proteiinia voidaan siis ilmentää heterologisesti muissa bakteerilajeissa toimivana ja systeemin proteiineista vain Cas9:ää tarvitaan geenihäirintään (Sapranauskas *ym.*, 2011). Cas9-proteiinin identifioinnin jälkeen (*S. thermophiluksessa* ja *S. pyogenesissä*) tuli ilmi, että Cas9:n sekvenssispesifistä endonukleasiaktiivisuutta voitaisiin käyttää geenieditoinnissa (Gasiunas *ym.*, 2012; Jinek *ym.*, 2012).

Ensimmäisen kerran Tyypin II CRISPR-Cas-systeemin häirintämekanismia käytettiin geenieditoinnissa nisäkässoluissa; editoinnissa Cas9:sää käytettiin geenien leikkaamiseen ja homologista rekombinaatiota korjaamaan sekvenssikatkoksen (Cong *ym.*, 2013; Mali *ym.*, 2013). Tyypin II systeemin geenieditointiin soveltuvuuden ilmeneminen on johtanut nykyään molekylaarisessa solubiologiassa laajasti käytetyn CRISPR/Cas9-geenieditointisysteemin syntyyn (Hsu *ym.*, 2014). Nykyään on taltioitu noin 9000 CRISPR-tietuetta ja CRISPR-alue on löydetty 87 % arkeista ja 45 % bakteereista vuonna 2017 päivitetyn CRISPRdb-tietokannan (Grissa *ym.*, 2007) perusteella. Lukuihin voi kuitenkin vaikuttaa, että bakteereja on sekvensoitu noin 30 kertaa enemmän arkkeihin verrattuna. Lukuisia CRISPR-systeemejä on siis vielä karakterisoimatta, joten eri CRISPR-systeemeistä voidaan vielä löytää uusi geenieditointiin soveltuva häirintäsysteemi.

4.2 CRISPR/Cas bakteerien muokkauksessa: mahdollisuudet ja vaikeudet

CRISPR/Cas9-geenieditointisysteemiä on käytetty eniten lähinnä eukaryooteissa: Cas9:llä on tehty näiden genomeihin spesifisiä DSB-katkoksia, jotka on korjattu joko homologisella rekombinaatiolla tai ei-homologisella rekombinaatiolla (engl. non-homologous end joining, NHEJ) (Luo, Leenay ja Beisel, 2016). Bakteereissa vastaava geneettinen muokkaaminen on haasteellista, koska toisin kuin eukaryootit, bakteerit eivät

usein pysty korjaamaan CRISPR-systeemin tekemiä DSB:itä (Luo, Leenay ja Beisel, 2016). Tämän takia esimerkiksi homologinen rekombinaatio on todennäköisesti heikkoa ilman ulkoisia korjaussysteemejä (DNA-rekombinaasit) (Luo, Leenay ja Beisel, 2016). CRISPR-systeemien tekemien katkokkien korjaus olisi puolustuksen kannalta epäsuotuisaa, joten Luo, Leenay ja Beisel (2016) mukaan bakteerien CRISPR-Cas-systeemit todennäköisesti kehittyivät vastustamaan korjausmekanismeja puolustusjärjestelmän toiminnan hyväksi (Luo, Leenay ja Beisel, 2016). Esimerkiksi tyypin II systeemissä Cas9-crRNA-dupleksi pysyy kohde-DNA:han sitoutuneena katkaisun jälkeen (Gasiunas *ym.*, 2012) tai tyypin I systeemissä Cas3 toimii kohteensa hajottavana eksonukleaasina (Sinkunas *ym.*, 2011).

CRISPR-Cas-systeemeistä tyypin II systeemiä ja tämän Cas9:sää on lähinnä tuotu ja käytetty kohdeorganismeissa kompaktisuutensa takia, mutta vaihtoehtoisesti bakteereissa voitaisiin mahdollisesti käyttää hyväksi esimerkiksi näiden omia CRISPR-systeemejä; bakteerien aktiiviset CRISPR-Cas-systeemit voitaisiin ottaa käyttöön tuomalla bakteeriin synteettinen CRISPR-alue (Luo, Leenay ja Beisel, 2016). Kyseistä strategiaa on jo käytetty esimerkiksi sekalaisesta bakteerikasvatuksesta *E. coli*n spesifiseen tuhoamiseen (antibioottien sijaan) omalla I-E-systeemillään: Gomaa *ym.* (2014) tutkimuksessa bakteerikannan tuhoaminen läheni 100 % bakteerin geeniä kohdentavan välkkeen sisältävällä plasmidilla (Gomaa *ym.*, 2014). Tämän lisäksi omaa CRISPR-systeemiä voi mahdollisesti käyttää isojen genomialueiden deletoinnissa, mutta ongelmana on menetelmän toksisuus isäntäsolulle (Vercoe *ym.*, 2013). Gomaa *ym.* (2014) ja Vercoe *ym.* (2013) mukaan oman genomien kohdennus vaikuttaisi nopeuttavan bakteerien evoluutiota esimerkiksi uudelleen järjestelemällä kantojen genomia ja muuttamalla patogeenisuutta, mutta kannat kuitenkin pyrkivät pääsemään eroon autoimmuunista kohdennusta ajavista CRISPR-välkkeistä (Vercoe *ym.*, 2013; Gomaa *ym.*, 2014).

Suurella osalla bakteereista on tyypin I tai tyypin III CRISPR-systeemi (Luo, Leenay ja Beisel, 2016). Isäntien omaa tyypin I systeemiä voitaisiin mahdollisesti käyttää esimerkiksi hypertermofiilissä bakteereissa, joissa perinteinen Cas9 ei toimi; näin ei myöskään tarvitsisi tuoda vieraita Cas-systeemien proteiineja (Luo, Leenay ja Beisel, 2016). Tyypin III systeemiä taas voitaisiin käyttää geenien hiljentämisessä: kuten edellä on mainittu, tyypin III systeemi hajottaa DNA:n lisäksi RNA:ta, joten systeemiä voitaisiin käyttää spesifisten geenien hiljentämisessä mRNA:n (lähetä-RNA:n) hajotuksella. Tämä on osoitettu *Sulfolobus solfataricus* tyypin III systeemillä käyttäen hyväksi edellä

mainittua systeemille ominaista oman genomin DNA-hajotuksen inaktivaatiota (Zebec *ym.*, 2014). Pitää kuitenkin ottaa huomioon, että bakteerien omien CRISPR-systeemien käyttöönotto geenieditoinnissa voi häiritä näiden luonnollista immuunipuolustusta (Luo, Leenay ja Beisel, 2016). Tämän lisäksi kaikkien kohdekantojen CRISPR-systeemit jouduttaisiin luokittelemaan systeemikohtaista PAM-sekvenssiä myöten, ja systeemien mahdollisesti olosuhteista riippuva toiminta jouduttaisiin varmistamaan (Luo, Leenay ja Beisel, 2016).

CRISPR-häirintämekanismien lisäksi adaptaation välkkeen integraatio- ja hankintamekanismia voitaisiin mahdollisesti käyttää bakteerien geenieditoinnissa sekvenssispesifisenä rekombinaasina, jolla pystyisi tehokkaasti liittämään synteettistä DNA:ta haluttuun geenialueeseen. Systemi normaalisti liittää kohteeseensa lyhyitä välkkeitä, mutta systeemiä voitaisiin mahdollisesti muokata liittämään eri kokoisia sekvenssejä kohteeseensa. (Luo, Leenay ja Beisel, 2016)

4.3 CRISPR/Cas9:n käyttö *Streptomyces*-bakteereissa

4.3.1 Perinteinen vastaan uusi genomienmuokkaus *Streptomyceeteissä*

Ennen CRISPR/Cas9-systeemiä, *Streptomyces*-kantojen geneettinen muokkaaminen on ollut aikaa vievää, työlästä ja muokkaus toimii matalalla tehokkuudella uuteen systeemiin nähden (Cobb *ym.*, 2015; Huang *ym.*, 2015). Yleisesti geneettistä muokkausta on tehty muun muassa PCR-fragmenteilla ohjatulla λ -red-rekombinaatiolla (Datsenko *ym.*, 2000). Menetelmän matalan tehokkuuden lisäksi kyseisessä rekombinaatiossa syntyy arpisekvenssiä, joka voi odottamattomasti rekombinoitua esimerkiksi toisen arpisekvenssialueen kanssa tai uusien PCR-fragmenttien kanssa, jos kohdekantaa erehtyy muokkaamaan useammin kuin kerran (Datsenko *ym.*, 2000). Esimerkiksi *Streptomyces coelicolorin* geosmiinisynteesiin liittyvien geenien muokkaaminen kosmidissa λ -red-rekombinaatiolla onnistui vain 10 – 30 %:ssa kloonista (Gust *ym.*, 2003). Homologinen rekombinaatio voi kuitenkin tapahtua suhteellisen tehokkaasti *Streptomyces*-mallikannoissa, kuten *S. lividansiksessa* ja *S. albuksessa* ilmankin CRISPR/Cas9:sää (40 – 60 %:n luokkaa) (Zhang *ym.*, 2017).

CRISPR/Cas9-systeemi taas toimii *Streptomyces*-bakteereissa jopa 100%:n tehokkuudella (Cobb *ym.*, 2015; Huang *ym.*, 2015). Tehokkuus on 100 %:n luokkaa lyhyillä inserteillä (100 bp), mutta selvästi matalampi (80 %:n luokkaa) pitkillä inserteillä

(1 kb) (Zhang *ym.*, 2017). CRISPR/Cas9-menetelmän suurempi rekombinaatiotehokkuus on tärkeä hankalissa matalan rekombinaatiotehokkuuden kannoissa, kuten *Streptomyces roseosporuksessa* jossa rekombinaatiota ei tapahdu lähes ollenkaan ilman CRISPR/Cas9:sää, mutta CRISPR/Cas9-menetelmällä yli 40 %:n tehokkuudella (Zhang *ym.*, 2017).

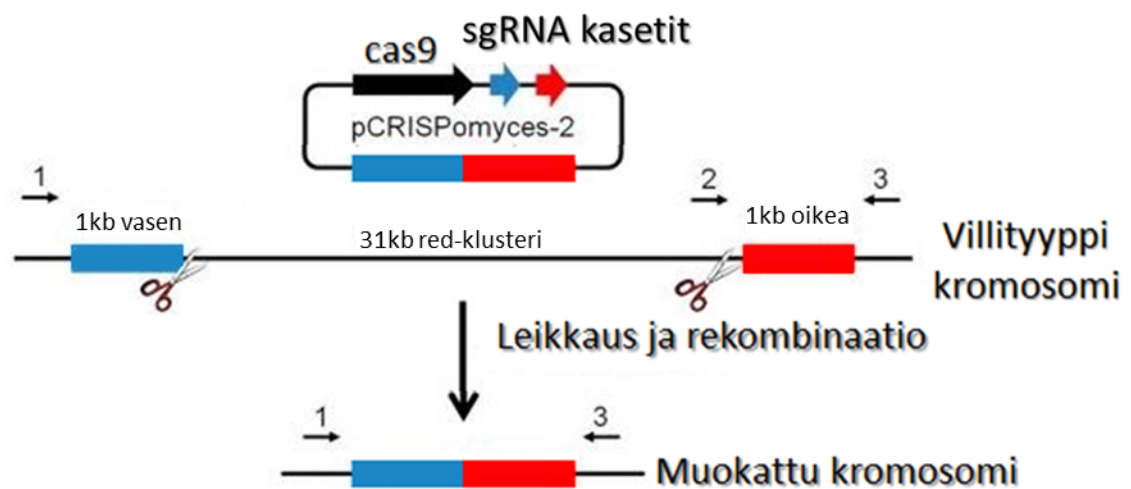
Korkean tehokkuuden lisäksi CRISPR/Cas9-systeemillä voidaan helposti muokata kohdekannan genomia ilman geenikirjastokosmideja, toisin kuin PCR-kohdennettu λ -red-rekombinaatio (Huang *ym.*, 2015). CRISPR/Cas9 on kuitenkin vielä nuori menetelmä *Streptomyces*-bakteerien muokkauksessa. Cobb *ym.*, (2015) osoittivat menetelmän soveltuvuuden *S. lividans*- ja *S. viridochromogenes*-kannassa, kun taas Huang *ym.* (2015) osoittivat menetelmän soveltuvuuden *S. coelicolor*- ja *S. pristinaespiralis*-kannassa.

4.3.2 *Streptomykeettien* muokkaus CRISPR/Cas9-plasmideilla

Cobb *ym.* (2015) muokkasivat ensimmäisinä *Streptomyces*-kantoja CRISPR/Cas9-systeemillä. He käyttivät yksinkertaista kaksiosaista pCRISPomyces-2-plasmidia, jossa on *Streptomyces*-kodonioptimoitua Cas9:sää ja sgRNA:ta koodaava alue. He muokkasivat useaa *S. lividans*-kannan geeniä, ja tämän lisäksi kahta *S. viridochromogenes*-kannan geeniä pCRISPomyces-2-plasmidin laajemman soveltuvuuden osoittamiseksi. Ensimmäisen systeemin lisäksi Huang *ym.* (2015) ja Tong *ym.* (2015) kehittivät oman CRISPR/Cas9-plasmidinsa.

Cobb *ym.* (2015) osoittivat CRISPR/Cas9-plasmidinsa korkean geenieditointitehokkuuden muokkaamalla undekyyliprodigiosiinigeeniklusterin *redN*-geeniä (*red*-klusteri) ja aktinorodiinigeeniklusterin *actVA-ORFS*-geeniä (*act*-klusteri) *S. coelicolorissa*. Geneettinen muokkaus onnistui korkealla 70 – 100 %:n tehokkuudella. He käyttivät geenieditointisysteemiään useamman geenin muokkaamiseen kerralla (engl. multiplex editing). Kuvassa 12 on *red*-geeniklusterin pCRISPomyces-2-plasmidilla muokkauksen periaate: he kohdensivat geeniklusterin DNA:n leikkauksen kahdesta kohtaa kahdella sgRNA:lla ja korjasivat tämän homologisella rekombinaatiolla. Genomin kanssa rekombinoituvan homologisen kasetin välistä puuttuu poistettava sekvenssi, joten tähän syntyy deleetio. Kyseisessä tutkimuksessa käytetty pCRISPomyces-2-plasmidi, sekä tämän käyttöön kuuluva CRISPR/Cas9-protokolla valittiin tutkielman antibioottitutkimukseen. (Cobb *ym.*, 2015)

Cobb *ym.* (2015) kokeilivat kaksiosaisen pCRISPomyces-2-plasmidin lisäksi kolmiosaista pCRISPomyces-1-plasmidia, jossa oli Cas9:n lisäksi valmiiksi toimivan sgRNA:n sijaan tracrRNA:n ja crRNA:n sisältävä CRISPR-aluekasetti. He huomasivat, että crRNA-tracrRNA-kohdentaminen toimi heikosti sgRNA:han verrattuna. Tämä johtui todennäköisesti siitä, että pre-crRNA-prosessointi ei toiminut kunnolla kohdekannassa. (Cobb *ym.*, 2015)



Kuva 12 Kaksiosaisen pCRISPR/Cas9-plasmidin toimintaperiaate – Kuvassa on esitetty miten Cobb *ym.* (2015) käyttivät pCRISPomyces-2-plasmidia red-geeniklusterin muokkaamiseen. He käyttivät kahta sgRNA-kasettia leikatakseen geeniklusterin kahdesta kohtaa, yhdistäen kaksi 1 kb:n aluetta homologisella rekombinaatiolla (genomiin homologinen sininen ja punainen kasetti). Kuva on muokattu Cobb *ym.* (2015) artikkelista.

Huang *ym.* (2015) käyttivät myös tutkimuksessaan kaksiosaista CRISPR/Cas9-plasmidia (pKCcas9dO), joka ilmentää sgRNA:ta ja *S. pyogenes* kannan *S. coelicolorille* kodonioptimoitua Cas9:sää (ScoCas9). Kyseinen genomieditointiplasmidi pohjautuu lämpöherkkään pKC1139-plasmidiin, mutta toisin kuin Cobb *ym.* (2015), he käyttivät Cas9:n ilmentämisessä indusoitavaa promoottoria. He osoittivat sgRNA:n kohdennusspesifisyyden ja tehokkuuden *S. coelicolorissa*, tehden kromosomaalisia katkoksia kyseisiin bakteereihin CRISPR/Cas9-plasmidilla. Huang *ym.* (2015) osoittivat geenieditointisynteesin tehokkuuden muokkaamalla *act-* ja *red-* antibioottibiosynteesigeeniklusterin geenejä kuten Cobb *ym.* (2015), mutta *S. coelicolorissa*. Tämän lisäksi he muokkasivat kannan CDA-geeniklusterin geenejä.

Deleetiot onnistuivat *act*-klusterissa 100 %:n tehokkuudella, mutta *red*-klusterissa vain 30 – 70 %:n tehokkuudella. (Huang *ym.*, 2015).

Tong *ym.* (2015) kehittivät myös oman CRISPR/Cas9-plasmidinsa, joka pohjautuu lämpöherkkään pGM1190-plasmidiin. Huang *ym.* (2015) tavoin he käyttivät indusoitavaa promoottoria, käyttivät isäntänä *S. coelicoloria*, sekä muokkasivat *act*-geeniklusterin geenejä. Tong *ym.* (2015) käyttivät joko homologista- tai NHEJ-rekombinaatiota. Homologisella rekombinaatiolla geenien deleetio oli 100 %:n luokkaa, mutta NHEJ-rekombinaation tehokkuus vain noin 3 – 50 %. He kuitenkin saivat parannettua NHEJ-menetelmän tehokkuutta heterologisesti ilmennetyn DNA-ligaasin avulla (noin 40 – 80 %). He myös osoittivat deleetiokantojen koko genomien sekvensoinnilla, ettei heidän CRISPR/Cas9-plasmidinsa ollut aiheuttanut tahattomia geneettisiä muutoksia genomeihin. (Tong *ym.*, 2015)

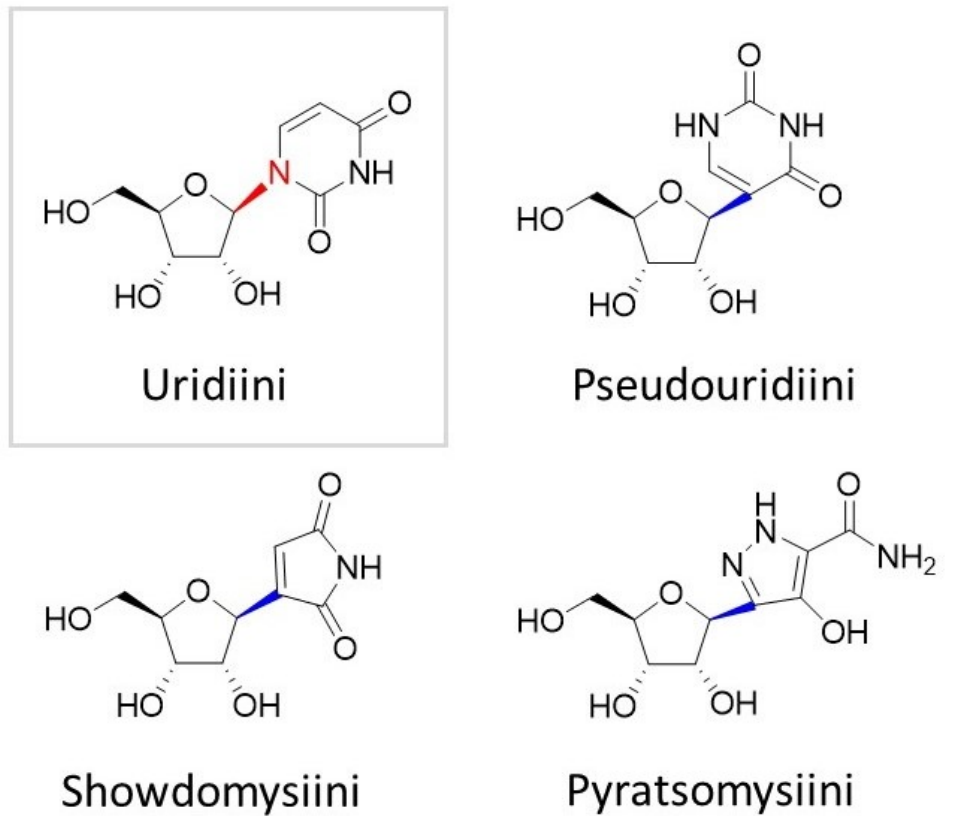
Geenien deleetion lisäksi Tong *ym.*, (2015) käyttivät niin sanottua CRISPR/dCas9-systeemiä. He käyttivät Cas9:n inaktiivista HNH- ja RuvC-domeenien aktiivisten keskusten kaksoismutanttia (dCas9; RuvC:n Asp10, sekä HNH:n His840 vaihdettu alaniiniksi) hiljentämään *S. coelicolorin act*-geenejä. Katalyyttisesti inaktiivinen dCas9 pystyy sitoutumaan kohde-DNA:han sgRNA:n avulla estäen sekvenssispesifisesti geenin transkription ja hiljentäen näin geenin toiminnan (CRISPRi). (Tong *ym.*, 2015)

OSA II – KOKEELLINEN OSA

5. Tutkimuksen tavoite

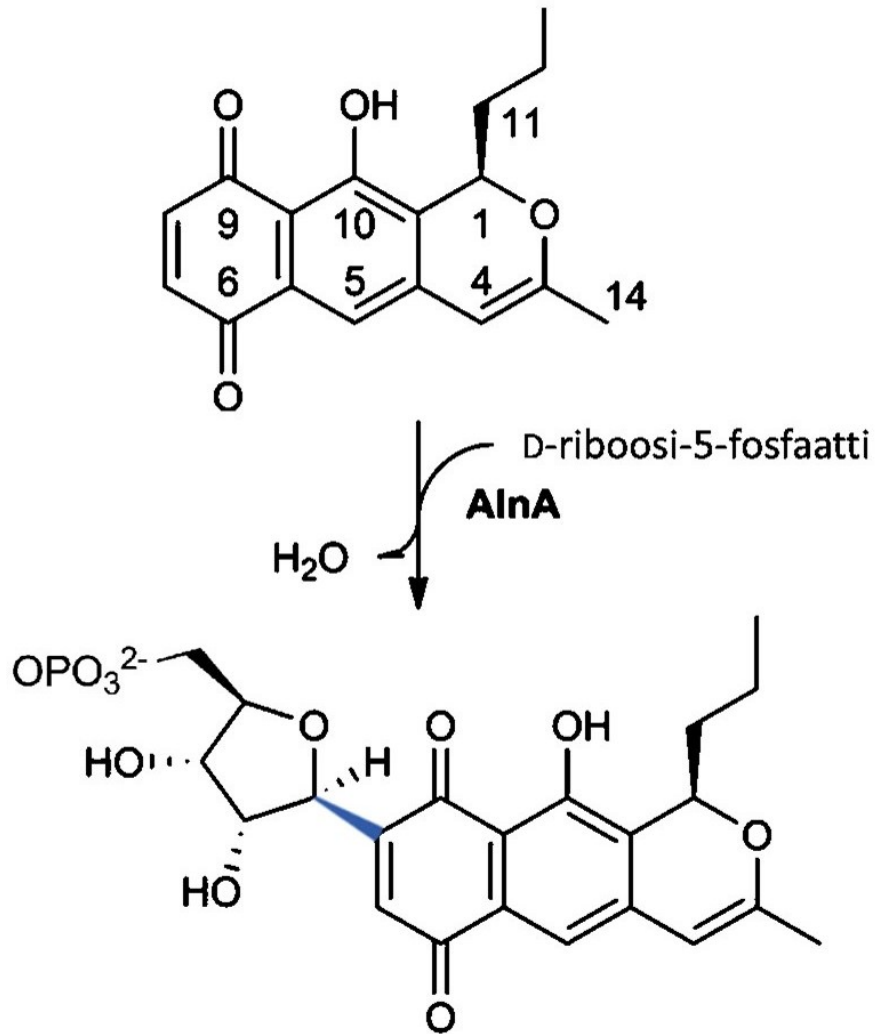
C-glykosidiset antibiootit ovat metabolialtaan vähän tutkittuja antibiootteja, joilla on kovalenttinen hiili-hiili-sidos sokeriosan ja molekyylin aglykoniosan välillä (Wellington ja Benner, 2006). C-glykosideja ovat muun muassa C-nukleosidit, joilla on nukleosidianalogeina N-nukleosidien hiili-typpi-sidoksen sijaan hiili-hiili-sidos riboosin ja emäksen välillä: kyseinen sidos ei hydrolysoitu (De Clercq, 2016). C-nukleosidiantibiootit ovat biologisesti aktiivisia, ja ovat näin tärkeitä lääketieteen kannalta. Kuvassa 13 on esitettyinä C-nukleosidit: pseudouridiini, showdomysiini ja pyratsomysiini, sekä N-nukleosidi uridiini. C-nukleosidit vaikuttavat muun muassa eliöiden transkriptiokoneistojen toimintaan (Chellat *ym.*, 2017). Showdomysiini on sytotoksinen antibiootti, mutta mahdollinen syöpälääke (Roy-Burman *ym.*, 1968; Rabinowitz *ym.*, 1979). Pyratsomysiini on taaskin antibiootti, jolla on antiviraalisia ominaisuuksia. Yhdisteellä on ribavirin-viruslääkkeeseen verrattuna yksi OH-ryhmä enemmän, joka sijaitsee 5'-hiilessä. Pyratsomysiini on kuitenkin todettu liian sytotoksinen kyseiseen tarkoitukseen. Sytotoksisuutensa takia pyratsomysiiniä on myös yritetty kehittää syöpälääkkeeksi, mutta yhdisteen on todettu yksinään tehoavan huonosti (Johnson *ym.*, 1981).

C-glykosidien tutkimuksessa on sokeriosan ja emäksen välisen kovalenttisen C-C-sidoksen synteessin selvittäminen tärkeää, koska kyseisen reaktion katalysoivia entsyymejä on tutkittu vähän. Työn tavoitteena oli löytää *Streptomyces*-kantojen vähän tutkittujen C-nukleosidimetaboliageeniklustereiden genejä ja osoittaa näiden funktio, jotta kyseisten antibioottien biosynteesistä saataisiin lisää tietoa.



Kuva 13 C-nukleosidit ja N-nukleosidit – Kuvassa C-nukleosidien C-C-sidos merkitty sinisellä (pseudouridiini, showdomysiini ja pyratsomysiini) ja N-nukleosidin C-N-sidos punaisella (uridiini).

Kohteena oli tutkia showdomysiiniklusterin ja pseudouridimysiiniklusterin geenejä. Tarkoituksena oli selvittää geenien mahdollinen suhde antibioottien metaboliaan deletoimalla seuraavat geenit klustereistaan: *Streptomyces showdoensiksen* showdomysiiniklusterin *sdmA*, sekä *S. albuksen* pseudourinomysiiniklusterin *stuC* ja *stuG*. Tarkoituksena oli käyttää CRISPR/Cas9-menetelmää ensin rakentaen tarvittavat geeninhäirintäkonstruktit, joilla geenideleetio oli tehtävä. Tutkimuksessa oli tarkoitus nähdä, eroaako geenidelektion yhteydessä kohteena olevan antibiootin tuotto villityyppikantaan verrattuna. Tarkoituksena oli myös tarkistaa komplementaatiolla ja geenityypityksellä, että deleetio on tapahtunut halutusti bakteerigenomiin. Komplementaatiossa onnistuneesti deletoituneen geenin toiminta palautetaan, todistaen mahdollisen hypoteesin.



Kuva 14 Alnumysiinin biosynteesin C-glykosidinen sidos – Kuvassa on AlnA-entsyymin reaktio, jossa muodostuu C-glykosidinen C-C-sidos (sinisellä). Kuva on muokattu Oja *ym.* (2012, 2013) artikkeleista.

Showdomysiinin ja pseudouridimysiinin tuottajaorganismit sekvensoitiin ja lupaavanolaiset geeniklusterit valittiin sekvenssianalyysin perusteella. Showdomysiinin kovalenttisen C-glykosidisen C-C-sidoksen valmistaman geenin olemassaolosta saatiin vihiä alnumysiini-A:n (kuva 14) kovalenttisen C-C-sidoksesta vastuussa olevasta *alnA*-C-glykosyntaasigeenistä, joka on homologinen showdomysiinin *sdmA*-geeniin (Oja *ym.*, 2012, 2013). Vihä saatiin pseudouridimysiinin geeniklusterin olemassaolosta sillä, että löydettiin pseudouridiinisyntaasia koodaava sekvenssi. Geeniklusterista valittiin häirintään *stuG*-geeni, mutta *stuC*:n häirintä oli myös suunnitelmassa.

6. Materiaalit ja menetelmät

6.1 Plasmidikonstruktioiden valmistaminen ja kloonaukset

6.1.1 Q5-PCR-reaktio *sdmA*-deleetiointisegmenttien valmistamiseksi

Geenieditointifragmentit (*sdmA*-5' ja -3') syntetisoitiin Q5-PCR:llä 20 µl kokonaistilavuuden PCR-reaktiolla [Q5 reaktio puskuri (New England Biolabs Inc., NEB), Q5 High GC enhancer (NEB), templaatti-DNA (100 ng; ATCC15227 genomisen DNA), 0,6 µM for- ja rev-alkukkeet, 3 % DMSO (dimetyylisulfoksidi; Thermo Fisher), dNTP Mix (200 mM jokaista dNTP:tä) ja Q5 High-fidelity DNA-polymeraasi (0,2 µl/reaktio)]. Alukkeet ovat listattuna taulukossa 1 (SdmAfor3 ja rev3, sekä SdmAfor5 ja rev5).

Ennen PCR:ää kromosomaalinen DNA laimennettiin ja denaturoitiin 98 °C:ssa 10 minuuttia. PCR ohjelma oli seuraavanlainen: 2 minuutin 98 °C:n denaturaatio ilman DNA-polymeraasia (engl. "hot start"). Polymeraasin lisäyksen jälkeen jatkettiin seuraavalla syklillä: 98 °C:n denaturaatiovaihe 10 sekuntia, 59,8 °C:n (*sdmA*-5') tai 62,6 °C:n (*sdmA*-3') alukkeidenliittymisvaihe 30 sekuntia ja 72 °C:n pidentymisvaihe 30 sekuntia. Kyseinen sykli toistettiin 30 kertaa. Reaktio lopetettiin 2 minuutin 72 °C:n pidentymisvaiheella (Agilent Technologies SureCycler 8800).

6.1.2 NEBuilder-reaktio *sdmA*-deleetiokonstruktion kokoamiseksi

NEBuilder-reaktiota varten geeniin spesifisen välikealueen (PS-sekvenssi; eroaa osittain alkuperäisestä määritelmästä) omaava pCRISPomyces-2 digestoitiin *Xba*I:llä (FastDigest, Thermo Fisher) 37 °C:ssa, tarkastettiin agarosielektroforeesilla (1 % geeli; w/v) ja eristettiin GeneJet Gel Extraction Kit:illä (Thermo Fisher). 1 kb:n geenieditointifragmentit liitettiin digestoituihin ja eristettyyn pCRISPomyces-2-plasmidiin NEBuilder® HiFi DNA Assembly Cloning Kit:illä (NEB) reaktioprotokollan mukaan (käytettiin vain puolet "master mix":ä per reaktio). Reaktioita inkuboitiin tunti 50 °C:ssa, jotta reaktion saanto olisi parempi.

6.1.3 Phusion PCR-reaktio *stuG*-insertin valmistamiseksi

Komplementaatiofragmentit syntetisoitiin Phusion-PCR:llä (Thermo Fisher) 20 µl kokonaistilavuuden reaktiolla [Phusion-GC-puskuri, 200 µM dNTP-seos, alukkeet 0,5 µM, <250 ng *Streptomyces albus* kromosomaalista DNA:ta, DMSO 3 % (v/v), 1 U Phusion-DNA-polymeraasi]. Alukkeet ovat taulukossa 1 (StuGfwdXba ja StuGrevHindbgI).

Kromosomaalinen DNA laimennettiin ja denaturoitiin 98 °C:ssa 10 minuuttia. PCR-ohjelma oli seuraavanlainen: Ensin 2 minuutin 98 °C denaturaatio ilman polymeraasia, jäähdytys huoneen lämpöön ja polymeraasin lisäys, jonka jälkeen ajettiin PCR-sykli. Seuraava sykli toistettiin 30 kertaa: 98 °C:n denaturaatiovaihe 10 sekuntia, alukkeidenliittymisvaihe (60,6, 61,6, 64,2 ja 66,0 °C) 20 sekuntia ja 72 °C:n pidentymisvaihe 25 sekuntia. Sykliä jälkeen lopullinen pidentymisvaihe 72 °C:ssa 10 minuuttia (Agilent Technologies SureCycler 8800).

Aluke	Suunta	Kuvaus	Sekvenssi
SdmAfor3	FOR	3'-puolen <i>sdmA</i> homologia-alue CRISPR/Cas:lle	5'-CCTTTTTACGGTTCCTGGCCTCTAG-ACCCACTGCATCGCCGACCCG-3'
SdmArev3	REV	3'-puolen <i>sdmA</i> homologia-alue CRISPR/Cas:lle	5'-TTCGACCCGACTCTGCTCGCACT-AGTCGACCACCGTGCTGGCG-3'
SdmAfor5	FOR	5'-puolen <i>sdmA</i> homologia-alue CRISPR/Cas:lle	5'-CGCCAGCACGGTGGTCGACTAGTG-CGAGCAGAGTCGGGTCGAA-3'
SdmArev5	REV	5'-puolen <i>sdmA</i> homologia-alue CRISPR/Cas:lle	5'-TTGCCGCCGGGCGTTTTTTATCTAG-AGCTTCGACTACTCCGGCTCGTACC-3'
StuG-fwdXba	FOR	<i>stuG</i> -geeni-insertti komplementaatioon	5'-GCGTCTAGAAGAGAAAGGGCGGA-CAAACC-3'
StuG-revHindbgI	REV	<i>stuG</i> -geeni-insertti komplementaatioon	5'-GCGAAGCTTAGATCTGGGCCTGCC-GCTCAGAC-3'

Taulukko 1 Aluketaulukko – Käytetyt *sdmA*- ja *stuG*-PCR-alukkeet.

6.1.4 Plasmidikloonauksen ja eristys *E. coli*sta

Plasmidikloonaukset tehtiin seuraavanlaisesti: Insertit (PCR-tuotteet) ja käytetyt plasmidit tarkastettiin agarosigeelielektroforeesilla (1 % geeli; w/v). FastDigest™-restriktioentsyymejä (Thermo Fisher), FastAP™:a (Thermosensitiivinen Alkalinen Fosfataasi, Thermo Fisher) ja T4-ligaasia (Thermo Fisher) käytettiin plasmidikonstruktioiden valmistamisessa. Kaupallisia GeneJET-reagenssisettejä (Thermo Fisher) käytettiin DNA-eristyksessä. Gel Extraction Kit:ä käytettiin DNA-bandien eristämiseksi geeliltä, kun taas Plasmid Miniprep Kit:ä *Escherichia coli*-plasmidien eristämiseen (sentrifuugi: Heraeus Pico 17; >12,000 x g.) *Streptomyces*-plasmidien eristyksessä käytettiin erilaista protokollaa. Työssä käytetyt plasmidit ovat taulukossa 2.

Plasmidi	Kuvaus	Viite tai lähde
pCRISP2-dsdmA		Tämä työ; Palmu <i>ym.</i> (2017)
pCRISP2-dstuC3		Tämä työ
pCRISP2-dstuG2		Julkaisematon
pCRISP2-sdmA-PS		Palmu <i>ym.</i> (2017)
pCRISP2-stuC-PS		Julkaisematon
pIJE486		Ylihonko <i>ym.</i> (1996)
pIJEstuG	<i>stuG XbaI-HindIII</i> insertti + pIJE486	Tämä työ
pMS-dstuC3	Synteettinen	Tämä työ; Invitrogen, Thermo Fisher
pS-GK	Lisätty kanamysiini resistenssi	Julkaisematon; muokattu Bai <i>ym.</i> , (2015) muokkaamasta pIJ8660-plasmidista
pS-GstuG	<i>stuG XbaI-BglIII</i> insertti + pS-GK	Tämä työ; kanamysiini resistenssi (K) korvattu <i>stuG</i> -geeni-insertillä
pUC19		Yanisch-Perron <i>ym.</i> (1985)
pUCstuG	<i>stuG XbaI-HindIII</i> insertti + pUC19	Tämä työ

Taulukko 2 Plasmiditaulukko – Työssä käytetyt plasmidit.

6.1.5 *E. coli*-tranformaatio

Kompetentit *E. coli*-solut (TOP10 tai ET12567/pUZ8002) sulatettiin jäällä, lisättiin ligaatioseos ja inkuboitii 30 minuuttia jäällä. Soluja inkuboitii 42 °C:ssa 45 sekuntia ja 5 minuuttia jäällä. Lisättiin 2xTY-mediumia (engl. Yeast extract tryptone; kokonaistilavuus 250 µl), ja soluja esikasvatettiin tunti 37 °C:ssa (ravistelu 200 – 300 rpm.) Bakteerit levitettiin antibiootti LA-maljoille (engl. Luria agar; apramysiini 25 µg/ml) ja kasvatettiin yön yli (noin 16 h) 37 °C:ssa. Valikoituneet kloonit poimittiin ja niiden plasmidit eristettiin. Plasmidin tarkistamiseen käytettiin Eurofins Genetics TubeSeq-sekvensointipalvelua (Eurofins MWG Operon Inc., Ebersberg, Saksa).

6.2 *Streptomyceettien* kloonaus

6.2.1 Konjugaatio *E. coli*n ja *Streptomyceettien* välillä

Streptomyces-itiöt valmistettiin seuraavanlaisesti: Itiöt raaputettiin MS-agarmaljoilta (agar, jossa 1:1 mannoosia ja soijajauhoa) ja suspensoitiin veteen. Noin minuutin voimakkaan sekoituksen jälkeen itiöt suodatettiin pumpulin läpi ja liuos sentrifugoitiin 10 minuutin ajan 4000 rpm:ssä, 4 °C (Hettich Universal 30F; 1 424A). Itiöt resuspensoitiin 20 % glyseroliin (v/v) ja säilöttiin -20 °C:ssa.

ET12567/pUZ8002-solut esikasvatettiin yön yli (noin 16 tuntia) LB-mediumissa (engl. Luria broth) antibioottiselektiolla [50 µg/ml apramysiini (apramysiini tiosulfaatti, sigma), 25 µg/ml kanamysiini (kanamysiini sulfaatti *Streptomyces kanamyceticuksesta*, Sigma-Aldrich) ja 19 µg/ml kloramfenikoli (sigma)] 37 °C:ssa (sekoitus 200 – 300 rpm).

Yön yli kasvatetut ET/U-solut inokuloitiin 1/100 laimentaen LB-mediumiin antibioottiselektiolla (kuten yllä). Soluja kasvatettiin 37 °C:ssa, kunnes saavutettiin 0,4-0,5 OD₆₀₀ (GeneQuant 1300 Spektrofotometri, GE Healthcare). Kasvatetut ET/U-solut pestiin kahdesti alkuperäistilavuudella LB:tä. Pesuissa solut sentrifugoitiin 10 minuuttia 4600 rpm:ssä, 4 °C (Hettich Universal 30F; 1424A). Pelletti resuspensoitiin pesujen jälkeen 1/10 tilavuuteen LB:tä.

Valmiit itiöt sulatettiin, lisättiin 2xTY-mediumi (kokonaistilavuus 500 µl), tehtiin 10 minuutin lämpöshokki 50 °C:ssa ja itiöiden annettiin jäähtyä huoneenlämpöiseksi. Pestyt ET/U-solut sekoitettiin (1:1; v/v) jäähtyneisiin itiöihin ja annettiin seistä huoneenlämmössä 15 minuuttia. Solut kerättiin ja levitettiin MS-agarmaljoille, joissa MgCl₂ (10 mM). Noin 16 – 20 tunnin 30 °C:ssa kasvatuksen jälkeen maljoille levitettiin nalidiksiinihappo (25 µg/ml; Sigma-Aldrich) ja plasmidiselektioantibiootti (apramysiini 50 µg/ml). Maljoja inkuboitiin 30 °C:ssa, kunnes mahdolliset konjugaatiokloonit olivat valmiita poimittaviksi.

6.2.2 Plasmidin pudotus *Streptomyces*-bakteerista

Plasmidi pudotettiin valituista *Streptomyces*-konjuganteista, ensin kasvattaen pumpulilla suodatettuja ja laimennettuja itiöitä 30 °C:ssa TSA-maljalla (engl. Tryptic soy agar). Itiöivät pesäkkeet poimittiin ja inokuloitiin TSB-nestekasvatukseen (engl. Tryptic soy broth). Plasmidipudotus tehtiin kasvattaen soluja 40 °C:ssa 72 h.

6.2.3 *Streptomyces*-protoplastien valmistus

Streptomyces inokuloitiin nestekaskavatuksista joko *S. lividans* TK24-kasvatuksessa YEME-mediumiin (Kieser *ym.*, 2000) tai *S. albus* DSM40763-kasvatuksessa TSB-mediumiin (mediumeihin lisättyinä 5 mM MgCl₂ ja 0,5 % glysiiniä). Soluja kasvatettiin 1 – 3 päivää ja ne sentrifugoitiin 10 minuuttia 3000 rpm (Megafuge 1.0.) Pelletit pestiin kahdesti 10,3 % sakkaroosiliuoksella [(w/v); P+-puskurissa (Kieser *ym.*, 2000)] ja sentrifugoiden kuten edellä. Solut käsiteltiin lysotsyymillä (0,45 µm steriilisuodatettu 2 mg/ml, P+ puskurissa) inkuboiden 30 minuuttia 30 °C:ssa.

Liuokset sekoitettiin ja niitä inkuboitiin vielä 18 minuuttia 30 °C:ssa. P+-puskuria lisättiin 1:1 (v/v) ja liuokset sekoitettiin. Solut suodatettiin pumpulin läpi ja sentrifugoitiin 7 minuuttia 3000 rpm 4 °C:ssa. Protoplastit resuspensoitiin P+-puskuriin, säilöttiin -20 °C:ssa yön yli ja tämän jälkeen säilöttiin -80 °C:ssa.

6.2.4 *Streptomyces*-protoplastitransformaatio

Protoplastit sulatettiin, sentrifugoitiin 7 minuuttia 3000 rpm (Heraeus Pico 17; roottori Ch.500005PP), supernatantti kaadettiin pois ja pelletti resuspensoitiin jääneeseen supernatanttiin. Plasmidi-DNA lisättiin soluihin ja liuos sekoitettiin. 25 % PEG-4000 (polyetyleeniglykoli; w/v) lisättiin (alkuperäistilavuus) välittömästi ja liuos sekoitettiin hyvin. Protoplastiliuos levitettiin varovasti R2YE-maljoille (Kieser *ym.*, 2000). 16 – 24 tunnin kasvatuksen jälkeen maljoille lisättiin antibiootti (tiosstreptoni 10 – 15 µg/ml.) *Streptomyces*-soluja kasvatettiin, kunnes transformantit voitiin poimia.

6.3 DNA:n eristys *Streptomyces*-soluista

6.3.1 Kromosomaalisen DNA:n eristys

S. albus DSM40763 (villityyppi ja *ΔstuG*) kasvatettiin TSB:ssä 30 °C:ssa 4 päivää. Solut sentrifugoitiin 4000 x g 15 minuuttia (Megafuge 1.0). Soluja lyysattiin 2,5 tuntia 37 °C:ssa [2,5 mg/ml lysotsyymi (Sigma, kananmunasta), 25 mM Tris-HCl, 25mM EDTA (etyleenidiamiinitetraetikkahappo) ja 15 % sakkaroosi (w/v).] Ensimmäisen inkubaation jälkeen lisättiin lysotsyymiä (1,5 mg) ja lyysausta jatkettiin 37 °C:ssa 1,5 tuntia. Lisättiin 0,7 mg proteinaasi-K:ta (Sigma, *Tritirachium albuksesta*) ja 1 % SDS (v/v; natriumdodekyylisulfaatti), lyyssaatti sekoitettiin kunnolla ja inkuboitiin 1,5 tuntia 50 °C:ssa. Liuokset jäähdytettiin jäällä. Lisättiin 0,3 M NaAc (natriumasetaatti), jonka jälkeen tehtiin reilu sekoitus ja 5 minuutin inkubaatio jäällä. Lisättiin 1:1 puskuroitua

fenolia (v/v; Sigma-aldrich, fenoli liuos), nestefaasit sekoitettiin kunnolla ja erotettiin 10 minuutin 3000 rpm sentrifuugauksella (Heraeus Multifuge X1R; TX-400). Vesifaasi otettiin talteen ja fenoliuutto toistettiin. Lisättiin 1:1 kloroformia (v/v), nestefaasit sekoitettiin ja erotettiin sentrifugoimalla kuten edellä. Vesifaasi otettiin taas talteen ja lisättiin 1:1 (v/v) 70 % isopropanolia. Kromosomaalisen DNA:n annettiin sakkautua noin 10 minuuttia. DNA poimittiin tikulla, pestiin kylmässä 70 % etanolissa ja ilmakeivattiin. Kuivattu DNA-pelletti resuspensoitiin TE-puskuriin (pH 8.0) ja säilöttiin 4 °C:ssa.

6.3.2 *Streptomyces*-plasmidin eristys

Streptomyces lividans TK24/pIJE486-kantaa kasvatettiin TSB:ssä 30 °C:ssa 6 päivää, solut fuugattiin alas 4000 rpm:ssä 10 minuuttia (Heraeus Megafuge 1.0, #2704 roottori #7570 F; 4 °C) ja pelletti säilöttiin -20 °C:ssa. Solut sulatettiin, resuspensoitiin ja lyysaattiin (2 mg/ml lysotsyymi; sama puskuri, kuten kromosomaalisen DNA:n eristyksessä) 37 °C:ssa 30 minuuttia (kokonaistilavuus 2 ml.) Lisättiin alkalinen SDS-liuos (0,1 M NaOH ja 0,67 % SDS; v/v) ja solulysaattia inkuboitiin 20 minuuttia 55 °C:ssa. Liuos sekoitettiin ja lisättiin 1/6-tilavuutta hapokasta fenolikloroformia (fenolia ja kloroformia 1:1; w/v), nestefaasit sekoitettiin ja erotettiin sentrifugoimalla 10 minuuttia 5000 rpm (Heraeus Multifuge X1R). Vesifaasi otettiin talteen ja lisättiin kloroformi (puolet vesifaasin tilavuudesta), nestefaasit sekoitettiin ja sentrifugoitiin kuten edellä. Vesifaasi otettiin taas talteen ja lisättiin 0,3 M NaAc (puskuroimaton), tehtiin sekoitus ja lisättiin 1:1 isopropanolia (v/v). DNA:n annettiin sakkautua 15 minuuttia huoneenlämmössä ja sentrifugoitiin 15 minuuttia 5000 rpm. Supernatantti otettiin pois, pelletti pestiin kylmällä 70 % etanolilla, sentrifuugaus tehtiin kuten edellä ja DNA-pelletti kuivattiin. Pelletti resuspensoitiin TE-puskuriin (pH 8) ja käsiteltiin RNAasi-A:lla (1 mg/ml) 37 °C:ssa tunnin. Plasmidi-DNA säilöttiin -20 °C:ssa.

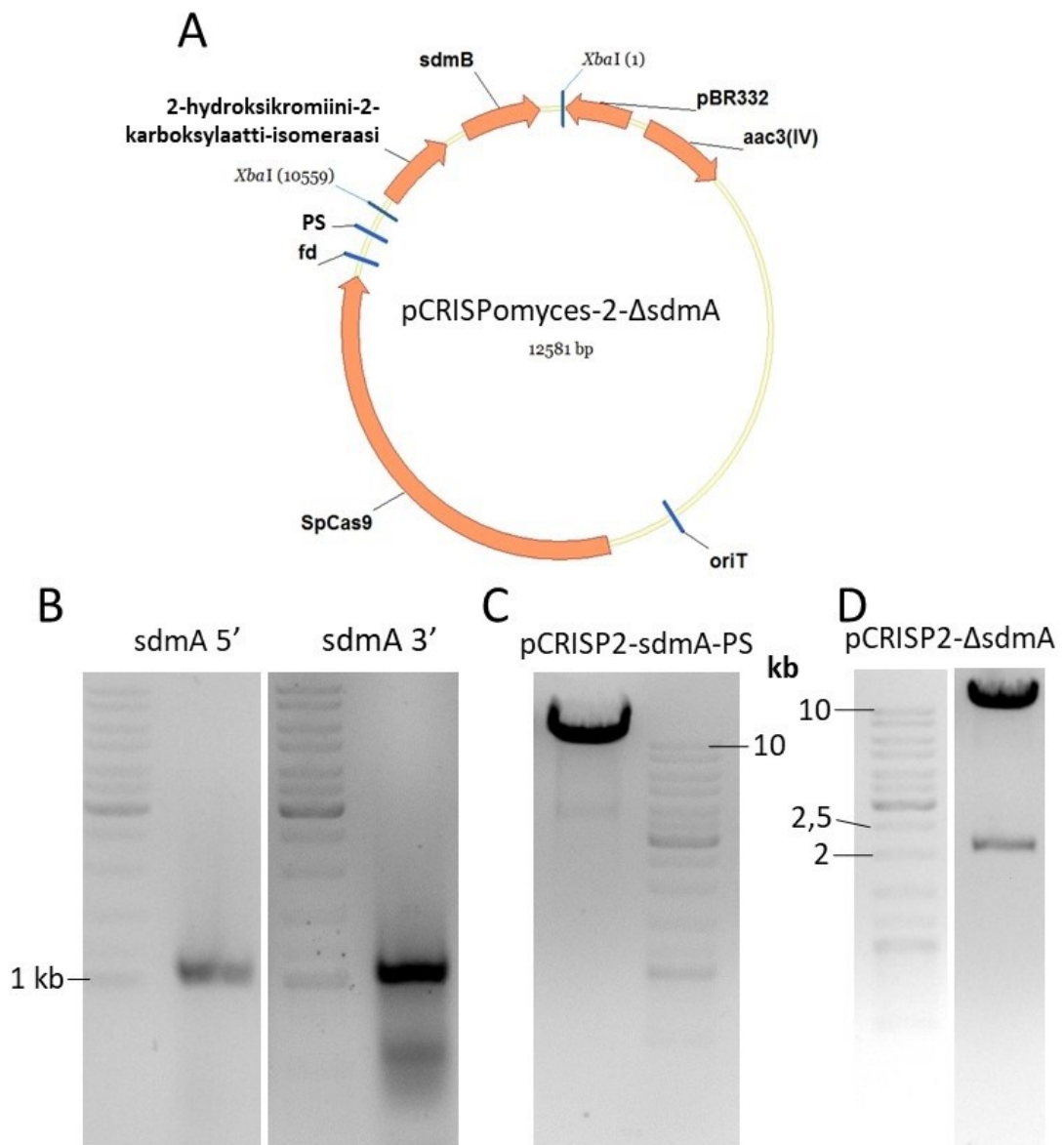
7. Tulokset

7.1 Showdomysiinin synteesireitin selvittäminen

7.1.1 Showdomysiinin *sdmA*-geenin inaktivointikonstrukt

Tarkoituksena oli saada selville, onko *sdm*-geeniklusterilla osaa showdomysiinin biosynteesissä. Tätä varten valmistettiin CRISPR/Cas9-geenihäirintäkonstruktit klusterin *sdmA*-geenin delemiseksi. Homologista rekombinaatiota varten deleetion sisältävät korjausfragmentit (PCR:llä syntesoidut 1 kb fragmentit; kuva 15, B) lisättiin onnistuneesti pCRISPomyces-2-plasmidiin, johon tutkimusryhmä oli lisännyt *sdmA*-protovälkkeen. Plasmidiin liitetty PS-fragmentti (luonnollisessa CRISPR-alueessa tätä kutsuttaisiin pelkästään välikkeeksi) koodaa *sdmA*-geenin keskelle homologista sgRNA:ta, joka taas kohdistaa plasmidin ilmentämän Cas9:n leikkaamaan *sdmA*:n. PS-plasmidin analyysi on kuvassa 15, C. Plasmidin Cas9:sää koodaava geeni on kodonoptimoitu *Streptomyces*-bakteereille.

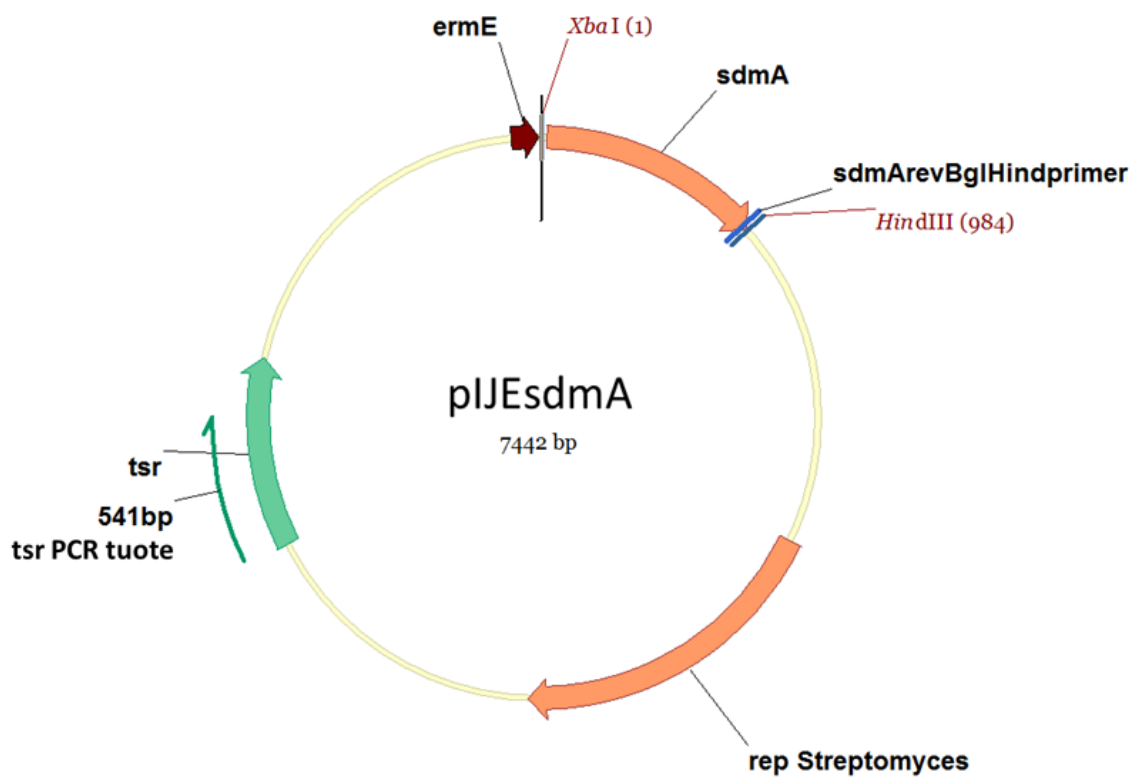
Valmis *sdmA*-geenin deleetiota varten koottu pCRISPomyces-2- Δ *sdmA*2-deleetioplasmidi varmistettiin oikeaksi (kuva 15, D; restriktioanalyysi ja sekvensointi). Kuvan 15, A plasmidikaavakuvasta voi nähdä, että *Xba*I-alueiden välisessä korjausfragmentissa (2 kb:n homologiakasetti) on *sdmA*:n ympäröivät fragmentit. Itse *sdmA*:n sekvenssi on poistettu välistä, näin tehden *sdmA*-geenideleetion *sdm*-geeniklusteriin homologisessa rekombinaatiossa (kuva 15, A; *sdmA* on *sdmB*:n ja 2-hydroksikromiini-2-karboksylaatti-isomeraasigeenin välissä).



Kuva 15 SdmA deletion käytetty plasmidi ja tämän kokoaminen – Kuvassa A on pCRISPomyces-2-ΔsdmA-plasmidikaavio, B on PS-plasmidiin liitettävät 1 kb PCR-tuotteet (5'- ja 3'-fragmentti; yhdessä 2 kb), C on pCRISP2-sdmA-PS-plasmidin restriktioanalyysi (*XbaI*) ja D on valmiin pCRISP2-ΔsdmA2-plasmidin restriktioanalyysi (*EcoRI* ja *NotI*). Plasmidin 1 kb tuotteet yhdistettynä 2 kb insertiksi, joka toimii plasmidissa homologisen rekombinaation kasettina. Insertti on kuvassa *XbaI*-kohtien välissä: *sdmB* ja isomeraasigeeni, joiden välistä *sdmA* on poistettu. PS on plasmidin sgRNA:ta koodaava välike.

7.1.2 Showdomysiinin *sdmA*-geenidelektion komplementaatio

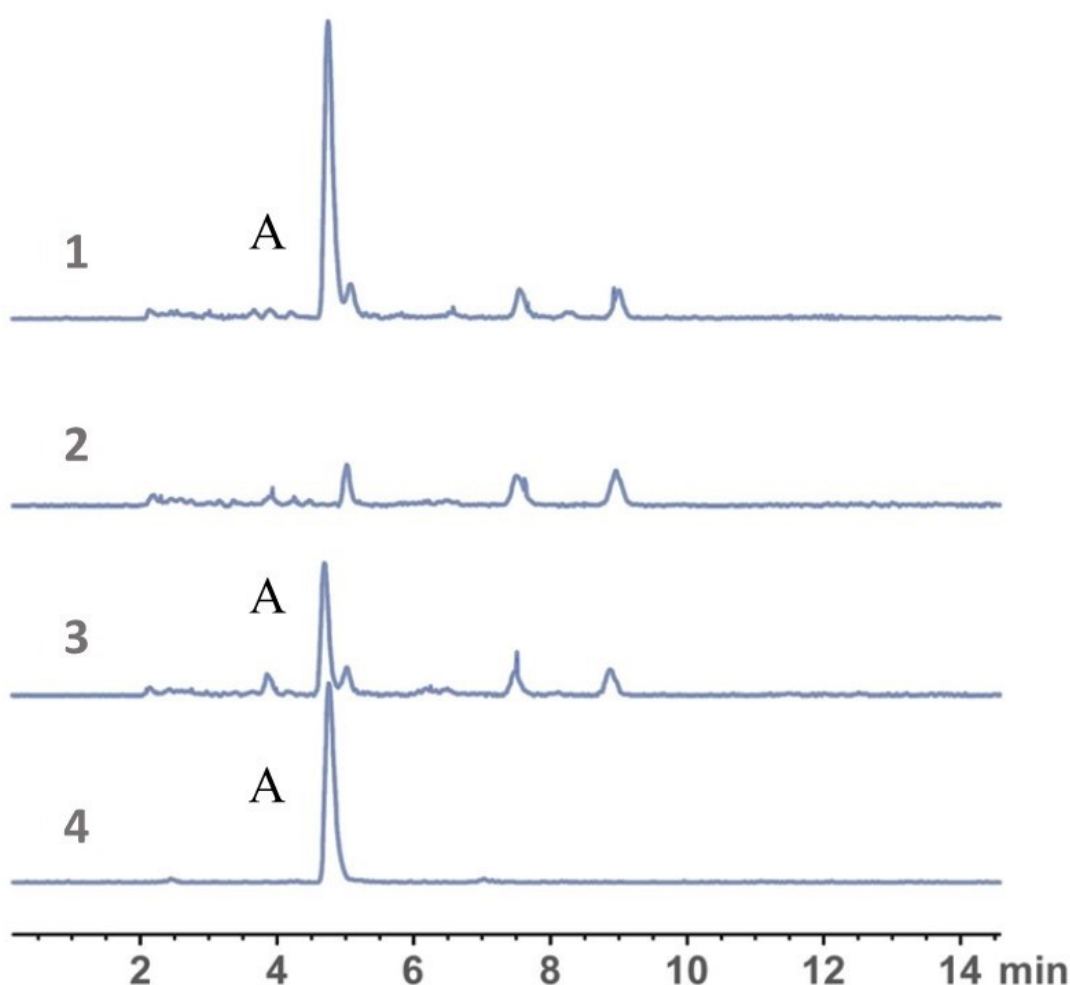
Tutkimusryhmä kokosi pIJE486sdmA-komplementaatioplasmidin (kuva 16), jolla *S. showdoensis-ΔsdmA* komplementoitiin deletion spesifisyyden varmistamiseksi. Plasmidissa on *Streptomyces*ssä toimiva *ermE*-promootori, jonka alle *sdmA*-geeni on liitetty. Plasmidin tarkoituksena on palauttaa *sdmA*-geenin toiminta onnistuneeseen delectiokantaan.



Kuva 16 SdmA-geenin komplementaatioon käytetty plasmidi – Kuvassa on pIJE486-sdmA:n plasmidikaavio. Plasmidin pohjana on pIJE486 (Ylihonko *ym.*, 1996), jossa on tehokas *ermE*-promootori (Bibb *ym.*, 1985) kloonattuna tämän polylinkkerialueeseen. Alun perin pIJE486 pohjautuu pIJ486-plasmidiin (Ward *ym.*, 1986), joka on *Streptomyceeteissä* korkean replikaatioluvun plasmidi.

7.1.3 Geenidelektion vaikutus showdomysiinin tuottoon

Tutkimusryhmä teki LC-MS tuottoanalyysin muokatuille *S. showdoensis* kannoille. *S. showdoensis* ATCC15227-villityypin, *S. showdoensis-ΔsdmA:n*, *S. showdoensis ΔsdmA/pIJE486-sdmA:n* ja kaupallisen yhdisteen nestekromatografia-, sekä massaspektrometridataa vertailemalla saatiin selville, että *sdmA:n* deleetio esti showdomysiinin tuoton (yhdistepiikki A). Tämän voi nähdä kuvassa 17, jossa villityypillä (1), komplementoidulla kannalla (3) ja kaupallisella yhdisteellä on yhdistepiikki A (4), mutta *sdmA*-deleetiokannalla ei ole piikkiä (2). Tutkimusryhmä on julkaissut kyseiset tulokset. (Palmu *ym.*, 2017)

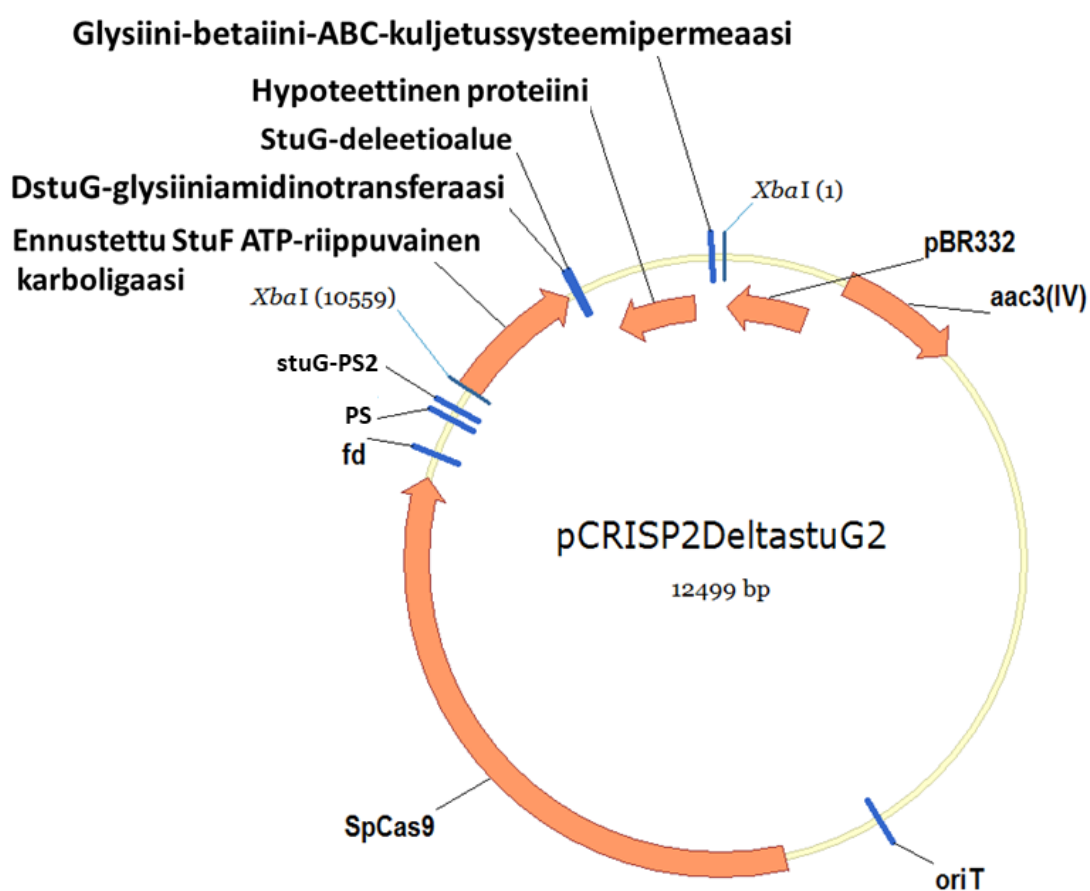


Kuva 17 Showdomysiinin tuotto – Kuvassa on yhdistepiikin A perusteella eriteltyt ionikromatogrammit. Kromatogrammi 1 on *S. showdoensis* ATCC15227-villityypikanta, 2 on *S. showdoensis-ΔsdmA*, 3 on *S. showdoensis ΔsdmA-pIJE486-sdmA* ja 4 on kaupallinen yhdiste. Yhdistepiikki A on positiivisen moodin ioni $[M + H]^+$, jonka m/z on 230,07 ja rakennekaava on $C_9H_{12}NO_6$. Kuva on muokattu Palmu *ym.* (2017) artikkelista.

7.2 Pseudouridimysiinin synteesireitin selvittäminen

7.2.1 Pseudouridimysiinin *stuG*-geenin inaktivoitikonstrukt

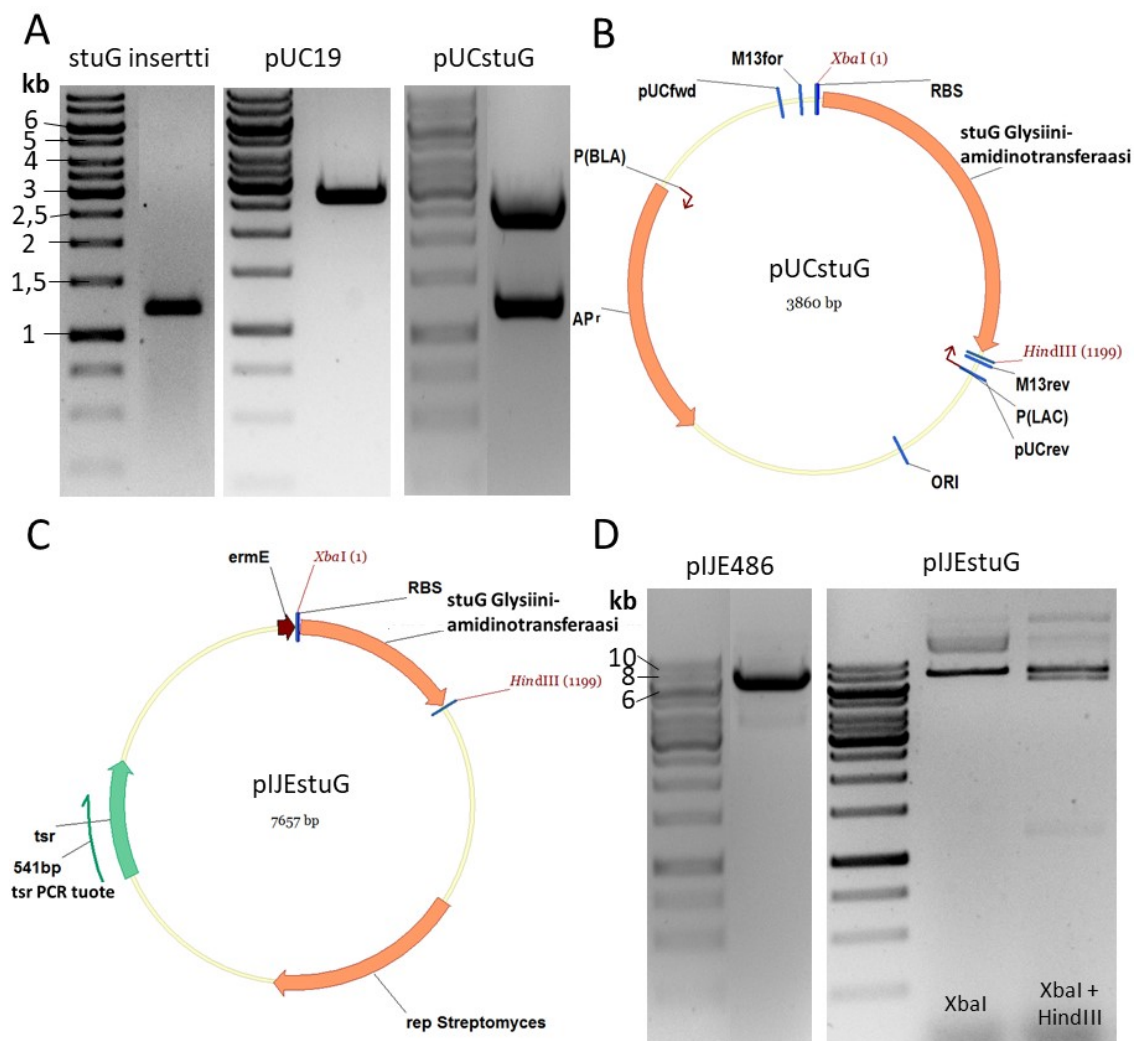
Tutkimusryhmä teki *stuG*-deleetiokonstruktiin (pCRISPomyces-2- Δ stuG2, kuva 18), jolla tehtiin mahdollinen *stuG*-geenin deleetio DSM40743 *Streptomyces albukseen*. Geemideleetiota varten pCRISPomyces-2 plasmidiin on liitettyä *stuG*:hen homologinen sgRNA-kasetti (kuvassa 18 *stuG*-PS2), *stuG*-deleetioalue ja tämän ympäröivät geenialueet, kuten *stuF*-geeni (liitetty *Xba*I:llä). Plasmidissa on *Streptomyces*-bakteereille kodonioptimoitu SpCas9, kuten myös *sdmA*-deleetioplasmidissa.

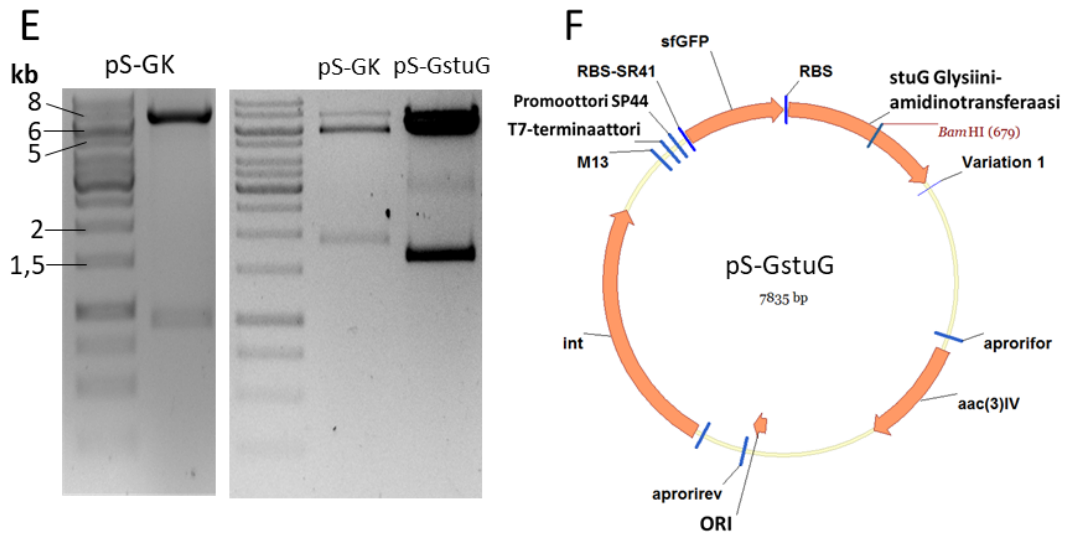


Kuva 18 Työssä *stuG*-geenin deleetioon käytetty plasmidi – pCRISPomyces-2- Δ stuG2:n plasmidikaavio. Plasmidissa on liitettyä *Xba*I:llä homologiakasetti, jossa on *stuG*:n deleetio ja tämän ympäröivät geenialueet (kuten *stuF*-geeni). Plasmidissa on lisäksi *Streptomyces*-bakteereille kodonioptimoitu SpCas9, sekä *stuG*:tä kohdentava PS (sgRNA-kasetti; *stuG*-PS2).

7.2.2 Pseudouridimysiinin *stuG*-geenidelektion komplementaatio

S. albus DSM40743:en *stuG*-geenidelektion spesifisyyden varmistamiseksi valmistettiin komplementaatioplasmidit. Tätä varten syntetisoitiin onnistuneesti 1,2 kb *stuG*-geeni-insertti, joka liitettiin pUC19-plasmidiin kloonausta varten. Tämä tarkastettiin restriktioanalyysillä ja sekvensoinnilla (kuva 19; A ja B). Tarkastettu insertti liitettiin onnistuneesti pIJE486-plasmidiin (Ylihonko *ym.*, 1996) *stuG*:n ilmentämistä varten, ja kyseinen pIJEstuG-komplementaatioplasmidi tarkastettiin restriktioanalyysillä (kuva 19; C ja D). Insertti myös liitettiin integraasin sisältämään pS-GK-plasmidiin [muokattu pIJ8660 (Bai *ym.*, 2015)] *stuG*:n liittämiseksi takaisin kohdekantaan. Valmis pS-GstuG-komplementaatioplasmidi analysoitiin restriktiolla (kuva 19; E ja F). Komplementaatio ei kuitenkaan onnistunut kyseisillä plasmideilla, eli mahdollisen delektion spesifisyyttä ei ole varmistettu komplementaatiolla.

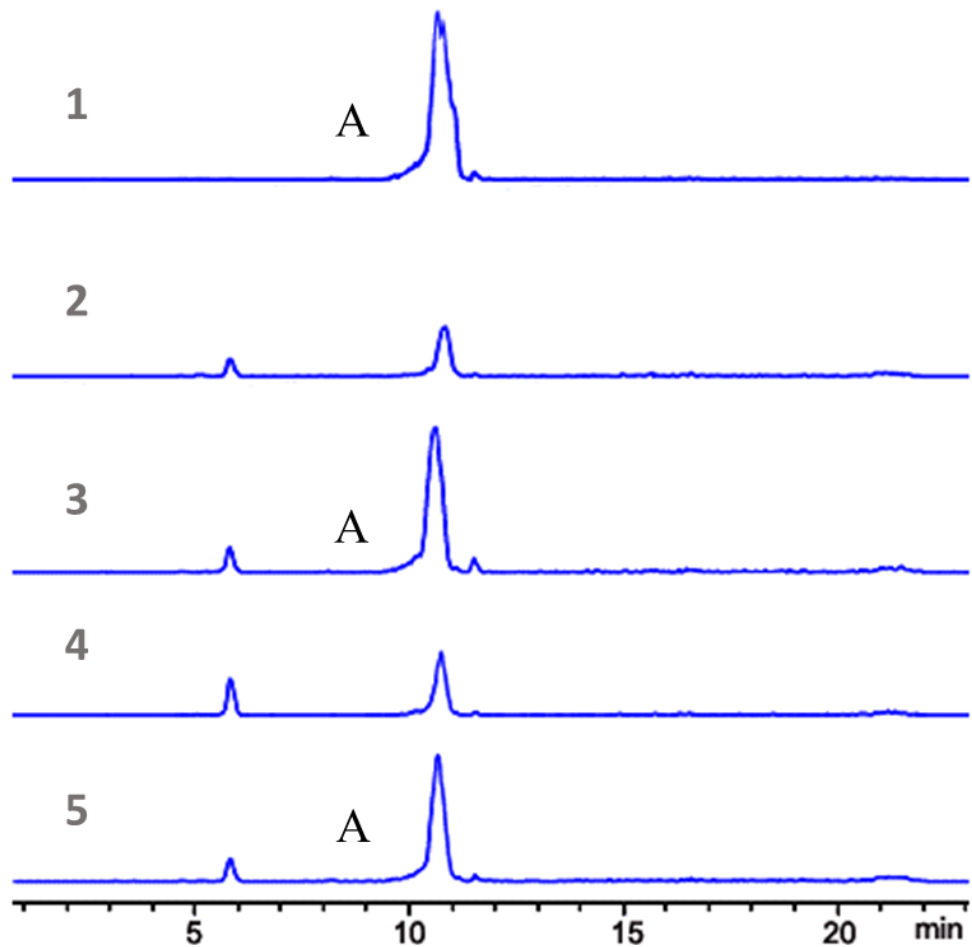




Kuva 19 Plasmidit *stuG*:n komplementaatioon – Kuvassa A on 1,2 kb *stuG*-PCR-tuote, pUC19:n ja pUCstuG:n restriktioanalyysi (*Xba*I ja *Hind*III) ja B on pUCstuG-plasmidikaavio. Kuva C on pIJEstuG-plasmidikaavio, D on pIJE486:n (*Xba*I ja *Hind*III) ja pIJEstuG:n restriktioanalyysi (*Xba*I ja *Hind*III). Kuva E on pS-GK:n ja pS-GstUG:n restriktioanalyysi (*Bam*HI ja *Xba*I) ja F on pS-GstUG-plasmidikaavio: kyseisessä pS-GK-plasmidista [pIJ8660 (Bai *ym.*, 2015)] muokatussa plasmidissa on faagin integraasigeeni (*int*), jolla *stuG*-geeni-insertti (transferaasi) liitetään bakteeriin. Plasmidin sfGFP (engl. green fluorescent protein) on taas fluoresoiva proteiini.

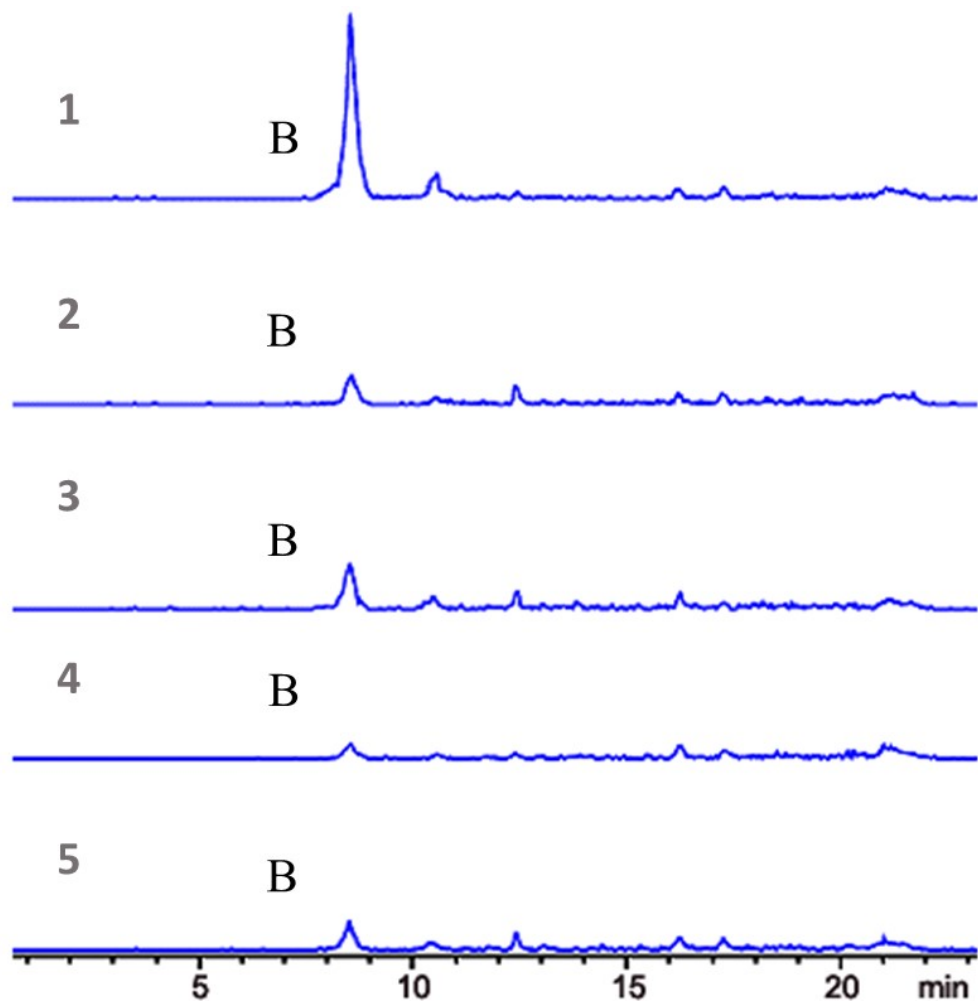
7.2.3 Geenidelektion mahdollinen vaikutus *stu*-yhdisteiden tuottoon

Streptomyces albus DSM40763-villityypillä ja *S. albus* Δ *stuG*-klooneilla tehtiin tuottokasvatus, ja saadut tuotteet eristettiin ja analysoitiin LC-MS:llä. Villityypistä havaittiin yhdiste, joka on oletettavasti pseudouridimysiini (PUM), jonka massa on 486 (kuva 20, 1). Havaitun piikin A retentioaika on 10,5 min (kuva 20). Valitettavasti kromatogrammin piikissä on olkapää, tai kaksi piikkiä on päällekkäin. *S. albus* Δ *stuG*-kloonien (kuva 20; 2, 3, 4 ja 5) yhdistepiikit ovat selkeitä, mutta yhdisteen tuotto kloonien välillä vaihtelee paljon. Kuvassa 20 on esitettyinä kloonien kromatogrammit: 1 on *S. albus* DSM40763-villityyppi, 2 on *S. albus* Δ *stuG* K1-klooni, 3 on K3-klooni, 4 on K4-klooni ja 5 on K7-klooni. Yhdisteen tuotto on selvästi laskenut K1:ssä ja K4:ssä.



Kuva 20 Pseudouridimysiinin tuotto – Kuvassa *S. albus* DSM40743-kantojen metaboliittiprofiileista eriteltyt ionikromatogrammit yhdistepiikin A kohdalta. Piikki A on positiivisen moodin ioni $[M + H]^+$, jonka m/z on 487,7 ja rakennekaava on $C_{17}H_{27}N_8O_9$. Kuvan ionikromatogrammi 1 on *S. albus* DSM40763-villityyppi, 2 on *S. albus* $\Delta stuG$ K1-klooni, 3 on K3-klooni, 4 on K4-klooni ja 5 on K7-klooni.

PUM:in lisäksi havaittiin 470 massan yhdiste B, jonka retentioaika on 8,5 min (kuva 21). Tämä on oletettavasti deoksipseudouridimysiini (dPUM). *S. albus* DSM40763-villityypistä havaittiin vahva piikki (kuva 21, 1), kun taas *S. albus* $\Delta stuG$ -klooneissa havaittiin todella pieni piikki. Piikin korkeuden perusteella dPUM:n tuotto on mahdollisesti pudonnut kaikissa kannoissa (kuva 21, 2 – 5).



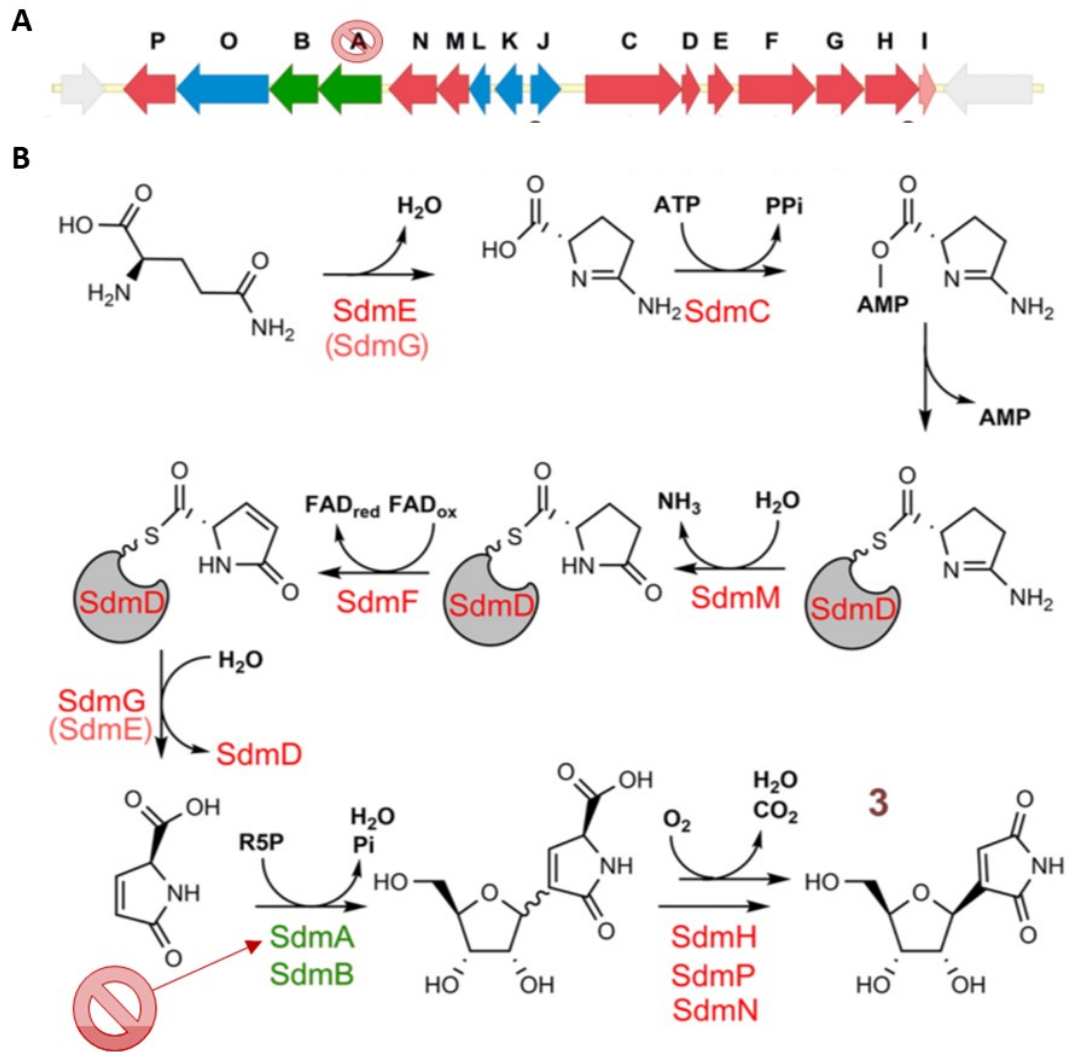
Kuva 21 Deoksipseudouridimysiinin tuotto – Kuvassa on *S. albus* DSM40743-kantojen metaboliittiprofiileista eriteltyt ionikromatogrammit yhdistepiikin B kohdalta. Piikki B on positiivisen moodin ioni $[M + H]^+$, jonka m/z on 471,7 ja molekyylikaava on $C_{17}H_{27}N_8O_8$. Kuvassa kromatogrammi 1 on *S. albus* DSM40763-villityyppi ja 2 – 5 ovat *S. albus* $\Delta stuG$ -kloonien (K1, K3, K4 ja K7 vastaavassa järjestyksessä) kromatogrammit.

8. Tulosten tarkastelu

8.1 Showdomysiini-metaboliareitin *sdmA*:n deleetion tulosten tarkastelu

Tutkimusryhmä sai selville, että *sdmA*:n deleetio (kuva 22, estomerkki) oli tärkeä syystä, että ilmentyvä entsyymi on vastuussa showdomysiinin C-nukleosidisen C-C-sidoksen synteesistä ribosin ja emäksen välillä. Tämä selittää, miksi *sdmA*-geenin deleetio estää tehokkaasti showdomysiinin tuottumisen. Tämän vahvistamiseksi tutkimusryhmä osoitti komplementaatiolla, että *sdmA*:n deleetio oli spesifinen, eikä deleetiolla huomattu haitallisia vaikutuksia muihin showdomysiinigeeniklusterin geeneihin. He myös varmistivat yhdisteen rakenteen NMR:llä (ydinmagneettinen resonanssi), että kannan tuottava yhdiste oli varmasti showdomysiini. Tämän lisäksi showdomysiinin koko metaboliareitti ja geeniklusteri saatiin selville osittain sekvenssianalyysillä ja kokeellisesti (kuva 22). SdmA on C-glykosyntaasi, joka on evolutiivisesti läheistä sukua pseudouridiinia hajottaville entsyymeille. (Palmu *ym.*, 2017)

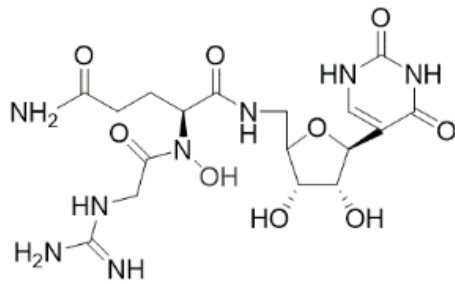
Showdomysiinin välituotteen C-glykosylaatioissa muodostuvalla C-C-sidoksella on spesifinen stereokemia aiemmin mainittuun alnumysiinin C-C-sidokseen verrattuna, jonka metaboliassa syntyy kahden stereokemialtaan erilaisen alnumysiinivälituotteen (alnumysiini C) raseeminen seos (Oja *ym.*, 2012; Palmu *ym.*, 2017). Showdomysiinin C-glykosylaatio tapahtuu muuten samankaltaisesti, kuten alnumysiinin biosynteesissä (Oja *ym.*, 2013; Palmu *ym.*, 2017). Esimerkiksi AlnB poistaa alnumysiinin välituotteesta ylimääräisen fosfaattiryhmän, jonka SdmB tekee todennäköisesti showdomysiinin synteesissä (Oja *ym.*, 2013; Palmu *ym.*, 2017). Tutkimusryhmän mukaan stereospesifisyyden selityksenä on, että showdomysiiniklusterin homologiset SdmH- ja SdmP-proteiinit (kuva 22) kuuluvat glutamaattirasemaasien ja aryyylimalonaattidekarboksylaasien proteiiniperheeseen (Palmu *ym.*, 2017). Kyseiset entsyymit ovat siis annotaation perusteella vastuussa showdomysiinin stereospesifisyydestä (Palmu *ym.*, 2017).



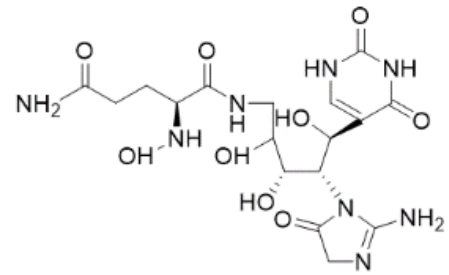
Kuva 22 Showdomysiinin hypoteettinen metaboliareitti – A: Showdomysiinin geeniklusteri, johon *sdmA*-geenin deleetio on merkitty punaisella estomerkillä. B: Showdomysiinin metaboliareitti, jossa deleetio on myös estomerkillä. Reitissä FAD on flaviiniadeniinidinukleotidi, Pi on fosfaatti, PPi on pyrofosfaatti ja R5P on D -riboosi-5-fosfaatti. Metaboliareitissä valmis showdomysiini on osoitettu numerolla 3. Kuva on muokattu Palmu *ym.* (2017) artikkelista.

8.2 Pseudouridimysiininmetaboliareitin *stuG*:n deleetion tulosten tarkastelu

Tutkimusta varten Pesic *ym.* (2014) artikkelin perusteella valittiin oletetusti strepturidiinia (STU) tuottava *S. albus* DSM40786-kanta. Kanta osoittautui tuottavan PUM:a STU:n sijaan. STU on siis itsessään väärin tulkitun NMR-kokeen perusteella määritely yhdiste (kuva 23), jota ei todennäköisesti ole olemassa. Tämän takia tutkimuksemme PUM-geeninimet ovat nimetty STU mukaan. Kuvan 24 metaboliareittiin on merkitty tutkimusryhmän geeninimet (STU) ja punaisella Sosio *ym.* (2018) mukaan nimettynä geenit.



Tarkka massa: 486,18
Molekyylipaino: 486,44
Pseudouridimysiini, PUM



Tarkka massa: 486,18
Molekyylipaino: 486,44
Strepturidiini, STU

Kuva 23 Pseudouridimysiini vastaan strepturidiini – Molemmilla yhdisteillä on sama massa ja molekyylikaava, koska strepturidiinin rakennekaava on tulosta väärin tulkitun NMR-kokeen perusteella.

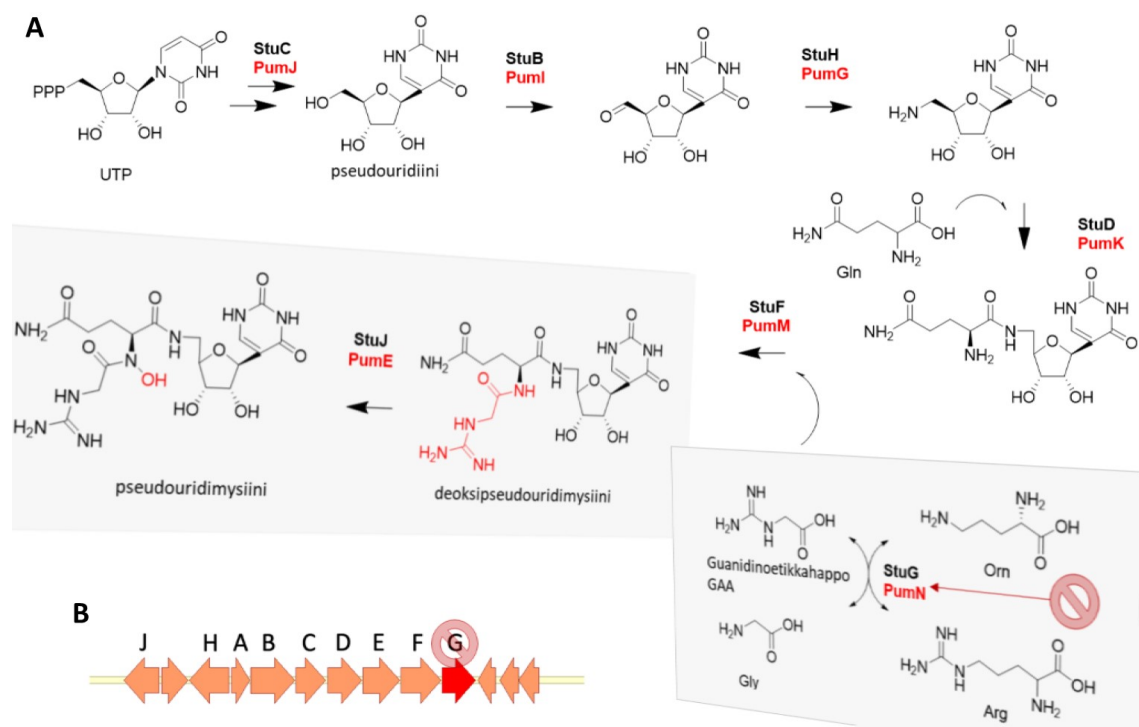
Sosio *ym.* (2018) julkaisivat artikkelin, jossa esitetään PUM:n metaboliareitti ja reitin geenit (kuva 24). Näiden geenien samankaltaisuus työssä löydettyihin *stu*-geeneihin vahvistaa käsityksen, että *S. albuksessa* tuottuva yhdiste on PUM, eikä STU. He tuottivat *Streptomyces* ID38640-villityyppikannassa PUM:a (myös tuottuu ID38673:ssa), ja estivät kannasta PUM:n tuottumisen *pumN*-deleetiolla ($\Delta stuG$). He myös osoittivat, että *Streptomyces* ID38673- $\Delta pumN$ -kanta tuotti PUM:a, kun kasvatukseen lisättiin guanidinoetikkahappoa (GAA). PUM:n saanto oli myös parempi GAA:n lisäyksen yhteydessä. Synteesiä ei kuitenkaan tapahtunut GAA:n kaltaisilla substraateilla (3-guanidinopropaanihappo, 4-guanidinobutaanihappo ja kreatiini), viitaten entsyymin spesifisyyteen tai huonoon substraatin sisäänottoon. (Sosio *ym.*, 2018)

Kuvassa 24 *stuG*-deleetio on esitetty estomerkillä. Hypoteettinen metaboliareitin reaktio täsmää osittain saatujen tulosten kanssa. Kuvassa 21 deoksipseudouridimysiinin tuotto laskee kaikissa *S. albus* $\Delta stuG$ -klooneissa. Tämä on selitettävissä sillä, että *stuG*-geeni on vastuussa GAA:n tuotosta. Ilmentyvä entsyymi katalysoi reaktiota, jossa glysiinistä ja arginiinista valmistetaan GAA:ta ja ornitiinia. GAA:ta taas tarvitaan dPUM:n tuottoon (StuF, kuva 24). GAA:n määrän pudotessa dPUM:n tuotto myös putoaa.

Tämän lisäksi pseudouridimysiinin tuoton pitäisi laskea, mutta näin ei ole. Kahdesta *S. albus* $\Delta stuG$ -kloonista PUM:n tuotto näyttäisi olevan laskenut, mutta kahdessa on vielä korkea piikki (kuva 20, piikki A). Tämän selittää *S. albus* $\Delta stuG$ K1:n genotyypitys, jonka perusteella haluttuun geeniklusteriin ei saatu aiheutettua geenideleetiota. Deleetiota ei siis todennäköisesti tapahtunut mihinkään valituista kannoista. Tämä johtuu todennäköisesti

epäonnistuneesta selektiosta, koska *S. albus* antibioottiresistenssistä oli puutteellisesti tietoa. Tähän johti myös *S. albus*-selektiomaljojen epäselvyys. Vaihtoehtoisesti käytetyllä bakteerikannalla on huono DSB:iden korjauskyky, jolloin Cas9:n katkaisema genomi ei korjaannu (Luo, Leenay ja Beisel, 2016). Koska Cas9 on heterologinen endonukleaasi, proteiinin toksisuus voi mahdollisesti myös vähentää konjugaatiotehokkuutta (Cobb, Wang ja Zhao, 2015). Saadut yhdisteiden tuoton erot voidaan mahdollisesti selittää satunnaisella tuoton vaihtelulla, koska tuottokasvatukset tehtiin Erlenmeyerpulloissa tarkasti säädellyn fermentorin sijaan.

S. albus DSM40743-villityypin analyysissä PUM:n kohdalla on lisäksi osittain päällekkäinen olkapääpiikki (kuva 20, 1). *S. albus* Δ *stuG*-kloonien (kuva 20, 2 ja 4) PUM:in yhdistepiikin kohdalla olevalla matalalla piikillä vaikuttaisi olevan sama retentioaika, kuten villityypin olkapääpiikillä. Kyseisellä piikillä on PUM:iin verrattuna sekunteja hitaampi retentio. Tämä toinen piikki on mahdollisesti jokin toinen yhdiste.



Kuva 24 Pseudouridimysiinin hypoteettinen metaboliareitti – A: Kuvassa on *pum*-geeniklusteri, jonka geenit ovat sijoitettuna metaboliareittiin *stu*-geeninimillä. Punaisella on merkitty löydetyt geenit Sosio *ym.* (2018) perusteella. Metaboliareitti on varmistettu tämän artikkelin perusteella. B: *pum*-geeniklusteri, jossa *stuG*-geenin deleetio on merkitty punaisella estomerkillä. Lisäksi *pum*-metaboliareitissä deleetio on merkitty myös estomerkillä.

9. Päätelmät ja yhteenveto

9.1 Showdomysiininimetaboliareitin tieto syöpälääketutkimuksessa
C-nukleosidiantibiootti showdomysiini on mahdollinen syöpälääke ja on alkyloivan maleimidirenkaan takia sytotoksinen (Roy-Burman *ym.*, 1968). Maleimidiryhmä alkyloi aminohappojen sulfhydryyliryhmiä fysiologisessa pH:ssa (Roy-Burman *ym.*, 1968). Showdomysiini näin inhiboi vapaita sulfhydryyliryhmiä sisältäviä entsyymejä ja spesifisesti pyrimidiinimetaboliaan osallistuvia entsyymejä (Roy-Burman *ym.*, 1968). Showdomysiini on esimerkiksi kaksi kertaa toksisempi lymfosyyttisiä leukemiasoluja (L1210) vastaan normaaleihin luuytimen soluihin verrattuna (Rabinowitz *ym.*, 1979). Maleimidi (N-etyylimaleimidi) on itsessään sytotoksinen ja tehoaa leukemiasoluihin, mutta vähemmän syöpäspesifinen showdomysiiniin verrattuna. Showdomysiini ja maleimidi inhiboivat myös *E. coli*n deoksiriboosisynteesiä (Uehara *ym.*, 1980).

Tutkimusryhmän julkaistusta tutkimuksesta (Palmu *ym.*, 2017) saadulla showdomysiininimetaboliareitin tiedolla (kuva 22) on todennäköisesti mahdollista muokata yhdisteen metaboliareittiä ja näin itse yhdistettä ilman synteettistä kemialla. Yhdistettä muokkaamalla voitaisiin mahdollisesti vähentää showdomysiinin haittavaikutuksia normaaleihin soluihin, mutta tämä vaikuttaa hankalalta maleimidiryhmän epäspesifisen aktiivisuuden takia. Showdomysiinin sokerin ja emäksen välisen hiili-hiili-sidoksen synteesiä koodaava *sdmA*-geeni on tärkeä löytö, koska kyseisten yhdisteiden sokerin ja emäksen välinen sidos ei hydrolysoitu. Tämän takia C-nukleosidiantibiootit ja näiden tutkimus ovat tärkeitä esimerkiksi lääketieteen näkökulmasta. Tietoa voitaisiin siis soveltaa mahdollisesti uusien syöpälääkkeiden kehittämiseen. Yksinkertaisesti tietoa voidaan myös käyttää hyväksi esimerkiksi yhdisteen saannon parantamisessa tehokkaammilla promoottoreilla.

9.2 Pseudouridimysiininimetaboliareitin tieto antibioottitutkimuksessa

PUM inhiboi spesifisesti bakteereiden RNAP:ita *in vitro* ja inhiboi Grampositiivisia ja -negatiivisia bakteereja nestekasvatuksessa. Tämän lisäksi PUM:n on tutkittu pysäyttävän bakteeri-infektion hiiressä (*Streptococcus pyogenes*; peritoniittimalli). PUM tehoaa antibioottisensitiivisiin ja resistentteihin bakteerikantoihin (resistenssi: aminoglykosidi, fluorokinoloni, glykopeptidi, kloramfenikoli, linkosamidi, lipopeptidi, makrolidi,

mupirosin, oksatsolidinoni, rifamysiini, tetrasykliini, trimetoprim ja β -laktaami), sekä moniantibioottiresistentteihin bakteerikantoihin. PUM on tulevaisuuden kannalta tärkeä antibiootti; bakteerit eivät voi kehittää resistenssiä PUM:aa vastaan normaalia transkriptiota haittaamatta, koska tämä sitoutuu RNAP:n NTP-sitoutumiskohtaan. PUM matkii uridiini-5'-trifosfaatin (UTP) rakennetta, ja näin kilpailee RNAP:hen sitoutumisesta UTP:n kanssa. UTP:n sitoutumisen estyessä RNA-synteesi lakkaa. (Maffioli *ym.*, 2017)

PUM on moniosainen, mutta metaboliareittinsä perusteella yksinkertainen antibiootti. Metaboliareitin tuntemuksen avulla voitaisiin mahdollisesti muokata PUM:n rakennetta metabolian tasolla. Tuntemalla metaboliareitin geenit, voidaan myös esimerkiksi parantaa yhdisteen saantoa uudella promootorilla tai metaboliaklusteri voidaan itsessään siirtää toiseen bakteerikantaan tai muuten koota uudelleen. Tietoa metaboliareitistä voitaisiin siis käyttää antibioottien tehokkaampaan tuotantoon, joten PUM:ia voitaisiin tulevaisuudessa tuottaa tehokkaammin antibioottihoitoa varten.

Tulevaisuudessa voitaisiin mahdollisesti lisätä PUM:n metaboliareittiin (kuva 24) geenejä, joilla voitaisiin koota PUM:n eri iteraatioita käyttäen eri metaboliatuotteita yhdisteen substraatteina. Haluttu hybridiantibiootti voitaisiin ilmentää muokatuilla tai jopa kimeerisillä proteiineilla suoraan bakteerissa. Kyseistä kombinatorista biosynteesiä on kokeiltu jo usealla antibiootilla (Floss, 2006). Teoriassa näin voidaan tuottaa uudenlaisia C-glykosidiantibiootteja suhteellisen helposti tai parantaa antibiootin tehokkuutta ilman synteettistä kemiaa. PUM:n metaboliaklusterin tuntemuksen avulla voitaisiin siis kehittää esimerkiksi uusia bioaktiivisia C-glykosidiyhdisteitä kombinatorisesti, ja näin kehittää uusia antibiootteja multiantibioottiresistenttejä sairaalabakteereja vastaan.

Koska toinen tutkimusryhmä selvitti jo *stuG*:n roolin, tutkimusta kannattaisi jatkaa esimerkiksi tutkimalla ilmentyvää StuC-proteiinia. Proteiinitutkimuksen lisäksi C-nukleosidien tutkimusta kannattaisi jatkaa etsimällä uusia antibiootteja homologian perusteella.

9.3 Bakteerien genominen muokkaus CRISPR-Cas:lla

Tyyppin II-A CRISPR-Cas-systeemiin pohjautuva CRISPR/Cas9-geenieditointisysteemi on osoittautunut toimivan tehokkaasti antibioottien metaboliaan osallistuvien geenien inaktivoinnissa, mutta vain eräissä *Streptomyces*-bakteereissa. Aiemmissä tutkimuksissa CRISPR/Cas9-plasmideja on lähinnä käytetty helposti muokattavissa *Streptomyces*-mallikannoissa, joka ei takaa systeemin toimivuutta muissa kannoissa. Tutkielman tutkimuksessa käytetyn pCRISPomyces-2-plasmidin on osoitettu toimivan *S. lividanssin* muokkauksessa, mutta esimerkiksi *S. coelicoloria* ei olla tähän asti pystytty muokkaamaan kyseisellä plasmidilla (Alberti ja Corre, 2019). CRISPR/Cas9-systeemin soveltuvuus siis vaihtelee bakteereissa kantakohtaisesti.

Albertin ja Corren, (2019) mukaan systeemiä käytettäessä ongelmia ovat lähinnä edellä mainittu Cas9:n toksisuus, sekä geneettisessä muokkauksessa tahattomat geneettiset muutokset. Cas9:n toksisuus johtuu siitä, että eräillä bakteerikannoilla DSB-rekombinaatio ei toimi. Toisaalta tarvittavaa isäntäsolun DSB-korjauskykyä voitaisiin parantaa heterologisesti ilmentämällä DNA-rekombinaasia, näin vähentäen Cas9:n toksisuutta geenieditoinnissa. Toksisuuden lisäksi Cas9:n ohjaamiseen käytetyn sgRNA:n sekundääriset rakenteet voivat myös mahdollisesti vähentää geenieditointitehokkuutta (Alberti ja Corre, 2019).

Tutkimustulosten perusteella pCRISPomyces-2-plasmidi sopii *S. showdoensiksen* muokkaukseen, mutta ei todennäköisesti sovellu käytetyn *S. albus*-kannan muokkaukseen. Tämä voi johtua matalasta rekombinaatiotehokkuudesta *S. albus*essa, mutta myös resistenssistä selektioon käytettyä apramysiinantibioottia vastaan. Hankalia kantoja varten voitaisiin mahdollisesti kehittää geenieditointimenetelmiä, joissa voitaisiin käyttää muiden CRISPR-Cas-systeemien häirintämekanismia tai näiden adaptaatiossa toimivia välikeintegraatiomekanismeja. Vaihtoehtoisesti genejä voitaisiin myös tutkia CRISPRi:n avulla hiljentämällä, eli ilman genomien DSB-katkoksia.

Lähteet

- Alberti, F. ja Corre, C. (2019) 'Editing streptomycete genomes in the CRISPR/Cas9 age.', *Natural product reports*. doi: 10.1039/c8np00081f.
- Anders, C. *ym.* (2014) 'Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.', *Nature*. Europe PMC Funders, 513(7519), pp. 569–73. doi: 10.1038/nature13579.
- Ariyoshi, M. *ym.* (1994) 'Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from *E. coli*.' , *Cell*, 78(6), pp. 1063–72. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7923356> (Accessed: 26 February 2019).
- Arslan, Z. *ym.* (2014) 'Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system.', *Nucleic acids research*. Oxford University Press, 42(12), pp. 7884–93. doi: 10.1093/nar/gku510.
- Bai, C. *ym.* (2015) 'Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*.' , *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(39), pp. 12181–6. doi: 10.1073/pnas.1511027112.
- Barrangou, R. *ym.* (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes.' , *Science (New York, N.Y.)*. American Association for the Advancement of Science, 315(5819), pp. 1709–12. doi: 10.1126/science.1138140.
- van Belkum, A. *ym.* (1998) 'Short-sequence DNA repeats in prokaryotic genomes.' , *Microbiology and molecular biology reviews : MMBR*, 62(2), pp. 275–93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9618442> (Accessed: 19 October 2018).
- Beloglazova, N. *ym.* (2008) 'A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats.' , *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 283(29), pp. 20361–71. doi: 10.1074/jbc.M803225200.
- Bibb, M. J., Janssen, G. R. ja Ward, J. M. (1985) 'Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*' , *Gene*. Elsevier, 38(1–3), pp. 215–226. doi: 10.1016/0378-1119(85)90220-3.
- Biertümpfel, C., Yang, W. ja Suck, D. (2007) 'Crystal structure of T4 endonuclease VII resolving a Holliday junction' , *Nature*. Nature Publishing Group, 449(7162), pp. 616–620. doi: 10.1038/nature06152.
- Bolotin, A. *ym.* (2005) 'Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin' , *Microbiology*. Microbiology Society, 151(8), pp. 2551–2561. doi: 10.1099/mic.0.28048-0.
- Brouns, S. J. J. *ym.* (2008) 'Small CRISPR RNAs guide antiviral defense in prokaryotes.' , *Science (New York, N.Y.)*. American Association for the Advancement of Science, 321(5891), pp. 960–4. doi: 10.1126/science.1159689.
- Charpentier, E. *ym.* (2015) 'Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity.' , *FEMS microbiology reviews*, 39(3), pp. 428–41. doi: 10.1093/femsre/fuv023.

- Chellat, M. F. ja Riedl, R. (2017) 'Pseudouridimycin: The First Nucleoside Analogue That Selectively Inhibits Bacterial RNA Polymerase', *Angewandte Chemie International Edition*. Wiley-Blackwell, 56(43), pp. 13184–13186. doi: 10.1002/anie.201708133.
- Chylinski, K. *ym.* (2014) 'Classification and evolution of type II CRISPR-Cas systems.', *Nucleic acids research*. Oxford University Press, 42(10), pp. 6091–105. doi: 10.1093/nar/gku241.
- De Clercq, E. (2016) 'C-Nucleosides To Be Revisited', *Journal of Medicinal Chemistry*. American Chemical Society, 59(6), pp. 2301–2311. doi: 10.1021/acs.jmedchem.5b01157.
- Cobb, R. E., Wang, Y. ja Zhao, H. (2015) 'High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system.', *ACS synthetic biology*. American Chemical Society, 4(6), pp. 723–8. doi: 10.1021/sb500351f.
- Cong, L. *ym.* (2013) 'Multiplex genome engineering using CRISPR/Cas systems.', *Science (New York, N.Y.)*. NIH Public Access, 339(6121), pp. 819–23. doi: 10.1126/science.1231143.
- Datsenko, K. A. *ym.* (2012) 'Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system', *Nature Communications*. Nature Publishing Group, 3(1), p. 945. doi: 10.1038/ncomms1937.
- Datsenko, K. A. ja Wanner, B. L. (2000) 'One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 97(12), pp. 6640–5. doi: 10.1073/pnas.120163297.
- Deltcheva, E. *ym.* (2011) 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III', *Nature*, 471(7340), pp. 602–607. doi: 10.1038/nature09886.
- Deveau, H. *ym.* (2008) 'Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*.', *Journal of bacteriology*. American Society for Microbiology (ASM), 190(4), pp. 1390–400. doi: 10.1128/JB.01412-07.
- Erdmann, S. ja Garrett, R. A. (2012) 'Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms.', *Molecular microbiology*. Wiley-Blackwell, 85(6), pp. 1044–56. doi: 10.1111/j.1365-2958.2012.08171.x.
- Fineran, P. C. *ym.* (2014) 'Degenerate target sites mediate rapid primed CRISPR adaptation.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(16), pp. E1629-38. doi: 10.1073/pnas.1400071111.
- Fineran, P. C. ja Charpentier, E. (2012) 'Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information', *Virology*. Academic Press, 434(2), pp. 202–209. doi: 10.1016/J.VIROL.2012.10.003.
- Floss, H. G. (2006) 'Combinatorial biosynthesis--potential and problems.', *Journal of biotechnology*. NIH Public Access, 124(1), pp. 242–57. doi:

10.1016/j.jbiotec.2005.12.001.

Garneau, J. E. *ym.* (2010) 'The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA', *Nature*. Nature Publishing Group, 468(7320), pp. 67–71. doi: 10.1038/nature09523.

Gasiunas, G. *ym.* (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 109(39), pp. E2579–86. doi: 10.1073/pnas.1208507109.

Gomaa, A. A. *ym.* (2014) 'Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems', *mBio*. Edited by S. Y. Lee, 5(1), pp. e00928–13. doi: 10.1128/mBio.00928-13.

Grissa, I., Vergnaud, G. ja Pourcel, C. (2007) 'The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats', *BMC Bioinformatics*, 8(1), p. 172. doi: 10.1186/1471-2105-8-172.

Gust, B. *ym.* (2003) 'PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 100(4), pp. 1541–6. doi: 10.1073/pnas.0337542100.

Hale, C. R. *ym.* (2012) 'Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs.', *Molecular cell*, 45(3), pp. 292–302. doi: 10.1016/j.molcel.2011.10.023.

Heler, R. *ym.* (2015) 'Cas9 specifies functional viral targets during CRISPR-Cas adaptation.', *Nature*. NIH Public Access, 519(7542), pp. 199–202. doi: 10.1038/nature14245.

Hille, F. ja Charpentier, E. (2016) 'CRISPR-Cas: biology, mechanisms and relevance.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. The Royal Society, 371(1707). doi: 10.1098/rstb.2015.0496.

Hochstrasser, M. L. *ym.* (2014) 'CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(18), pp. 6618–23. doi: 10.1073/pnas.1405079111.

Holm, L. ja Sander, C. (1993) 'Protein Structure Comparison by Alignment of Distance Matrices', *Journal of Molecular Biology*, 233(1), pp. 123–138. doi: 10.1006/jmbi.1993.1489.

Hsu, P. D., Lander, E. S. ja Zhang, F. (2014) 'Development and applications of CRISPR-Cas9 for genome engineering.', *Cell*. NIH Public Access, 157(6), pp. 1262–78. doi: 10.1016/j.cell.2014.05.010.

Huang, H. *ym.* (2015) 'One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*', *Acta Biochimica et Biophysica Sinica*. Oxford University Press, 47(4), pp. 231–243. doi: 10.1093/abbs/gmv007.

Ishino, Y. *ym.* (1987) 'Nucleotide sequence of the *iap* gene, responsible for alkaline

phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product.', *Journal of bacteriology*. American Society for Microbiology (ASM), 169(12), pp. 5429–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3316184> (Accessed: 20 October 2018).

Jansen, R. *ym.* (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes.', *Molecular microbiology*, 43(6), pp. 1565–75. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11952905> (Accessed: 19 October 2018).

Jiang, F. ja Doudna, J. A. (2015) 'The structural biology of CRISPR-Cas systems.', *Current opinion in structural biology*. Howard Hughes Medical Institute, 30, pp. 100–111. doi: 10.1016/j.sbi.2015.02.002.

Jiang, F. ja Doudna, J. A. (2017) 'CRISPR–Cas9 Structures and Mechanisms', *Annual Review of Biophysics*, 46(1), pp. 505–529. doi: 10.1146/annurev-biophys-062215-010822.

Jinek, M. *ym.* (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.

Jinek, M. *ym.* (2014) 'Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.', *Science (New York, N.Y.)*. Howard Hughes Medical Institute, 343(6176), p. 1247997. doi: 10.1126/science.1247997.

Johnson, R. K. *ym.* (1981) 'Influence of pyrazofurin on the toxicity and antitumor activity of fluorinated pyrimidines in vivo', *Advances in Enzyme Regulation*. Pergamon, 19, pp. 309–333. doi: 10.1016/0065-2571(81)90022-4.

Jore, M. M. *ym.* (2011) 'Structural basis for CRISPR RNA-guided DNA recognition by Cascade.', *Nature structural & molecular biology*, 18(5), pp. 529–36. doi: 10.1038/nsmb.2019.

Kazlauskienė, M. *ym.* (2016) 'Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition', *Molecular Cell*. Cell Press, 62(2), pp. 295–306. doi: 10.1016/J.MOLCEL.2016.03.024.

Kieser, T. *ym.* (2000) *Practical Streptomyces genetics*. John Innes Foundation.

Kim, T.-Y. *ym.* (2013) 'Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity', *Biochemical and Biophysical Research Communications*, 441(4), pp. 720–725. doi: 10.1016/j.bbrc.2013.10.122.

Koonin, E. V. ja Krupovic, M. (2015) 'Evolution of adaptive immunity from transposable elements combined with innate immune systems', *Nature Reviews Genetics*, 16(3), pp. 184–192. doi: 10.1038/nrg3859.

Leon, L. M., Mendoza, S. D. ja Bondy-Denomy, J. (2018) 'How bacteria control the CRISPR-Cas arsenal', *Current Opinion in Microbiology*. Elsevier Current Trends, 42, pp. 87–95. doi: 10.1016/J.MIB.2017.11.005.

Levy, A. *ym.* (2015) 'CRISPR adaptation biases explain preference for acquisition of foreign DNA.', *Nature*. Europe PMC Funders, 520(7548), pp. 505–510. doi: 10.1038/nature14302.

- Lillestøl, R. K. *ym.* (2009) ‘CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties’, *Molecular Microbiology*, 72(1), pp. 259–272. doi: 10.1111/j.1365-2958.2009.06641.x.
- Luo, M. L., Leenay, R. T. ja Beisel, C. L. (2016) ‘Current and future prospects for CRISPR-based tools in bacteria.’, *Biotechnology and bioengineering*. NIH Public Access, 113(5), pp. 930–43. doi: 10.1002/bit.25851.
- Maffioli, S. I. *ym.* (2017) ‘Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase.’, *Cell*. NIH Public Access, 169(7), p. 1240–1248.e23. doi: 10.1016/j.cell.2017.05.042.
- Makarova, K. S. *ym.* (2015) ‘An updated evolutionary classification of CRISPR-Cas systems.’, *Nature reviews. Microbiology*. NIH Public Access, 13(11), pp. 722–36. doi: 10.1038/nrmicro3569.
- Mali, P. *ym.* (2013) ‘RNA-guided human genome engineering via Cas9.’, *Science (New York, N.Y.)*. NIH Public Access, 339(6121), pp. 823–6. doi: 10.1126/science.1232033.
- Marraffini, L. A. ja Sontheimer, E. J. (2010) ‘Self versus non-self discrimination during CRISPR RNA-directed immunity.’, *Nature*. NIH Public Access, 463(7280), pp. 568–71. doi: 10.1038/nature08703.
- Mojica, F. J. M. *ym.* (2000) ‘Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria’, *Molecular Microbiology*. Wiley/Blackwell (10.1111), 36(1), pp. 244–246. doi: 10.1046/j.1365-2958.2000.01838.x.
- Mojica, F. J. M. *ym.* (2005) ‘Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements’, *Journal of Molecular Evolution*. Springer-Verlag, 60(2), pp. 174–182. doi: 10.1007/s00239-004-0046-3.
- Mojica, F. J. M. *ym.* (2009) ‘Short motif sequences determine the targets of the prokaryotic CRISPR defence system’, *Microbiology*. Microbiology Society, 155(3), pp. 733–740. doi: 10.1099/mic.0.023960-0.
- Nam, K. H. *ym.* (2012) ‘Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein.’, *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 287(43), pp. 35943–52. doi: 10.1074/jbc.M112.382598.
- Nishimasu, H. *ym.* (2014) ‘Crystal structure of Cas9 in complex with guide RNA and target DNA.’, *Cell*. NIH Public Access, 156(5), pp. 935–49. doi: 10.1016/j.cell.2014.02.001.
- Nuñez, J. K. *ym.* (2014) ‘Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity.’, *Nature structural & molecular biology*. Howard Hughes Medical Institute, 21(6), pp. 528–34. doi: 10.1038/nsmb.2820.
- Nuñez, J. K. *ym.* (2015) ‘Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity.’, *Nature*. Howard Hughes Medical Institute, 519(7542), pp. 193–8. doi: 10.1038/nature14237.
- Oja, T. *ym.* (2012) ‘Biosynthetic pathway toward carbohydrate-like moieties of

alnumycins contains unusual steps for C-C bond formation and cleavage.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 109(16), pp. 6024–9. doi: 10.1073/pnas.1201530109.

Oja, T. *ym.* (2013) ‘Structural basis for C-ribosylation in the alnumycin A biosynthetic pathway.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(4), pp. 1291–6. doi: 10.1073/pnas.1207407110.

Osawa, T. *ym.* (2015) ‘Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog’, *Molecular Cell*. Cell Press, 58(3), pp. 418–430. doi: 10.1016/J.MOLCEL.2015.03.018.

Palmu, K. *ym.* (2017) ‘Discovery of the Showdomycin Gene Cluster from *Streptomyces showdoensis* ATCC 15227 Yields Insight into the Biosynthetic Logic of C-Nucleoside Antibiotics.’, *ACS chemical biology*, 12(6), pp. 1472–1477. doi: 10.1021/acscchembio.7b00078.

Pesic, A. *ym.* (2014) ‘Isolation and structure elucidation of the nucleoside antibiotic strepturidin from *Streptomyces albus* DSM 40763’, *The Journal of Antibiotics*, 67(6), pp. 471–477. doi: 10.1038/ja.2014.16.

Pourcel, C., Salvignol, G. ja Vergnaud, G. (2005) ‘CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies’, *Microbiology*. Microbiology Society, 151(3), pp. 653–663. doi: 10.1099/mic.0.27437-0.

Pul, U. *ym.* (2010) ‘Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS.’, *Molecular microbiology*, 75(6), pp. 1495–512. doi: 10.1111/j.1365-2958.2010.07073.x.

Rabinowitz, M., Uehara, Y. ja Vistica, D. T. (1979) ‘Differential competition with cytotoxic agents: an approach to selectivity in cancer chemotherapy.’, *Science (New York, N.Y.)*, 206(4422), pp. 1085–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/493993> (Accessed: 20 June 2018).

Rath, D. *ym.* (2015) ‘The CRISPR-Cas immune system: Biology, mechanisms and applications’, *Biochimie*. Elsevier, 117, pp. 119–128. doi: 10.1016/J.BIOCHI.2015.03.025.

Redding, S. *ym.* (2015) ‘Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System’, *Cell*. Cell Press, 163(4), pp. 854–865. doi: 10.1016/J.CELL.2015.10.003.

Roy-Burman, S., Roy-Burman, P. ja Visser, D. W. (1968) ‘Showdomycin, a new nucleoside antibiotic.’, *Cancer research*. American Association for Cancer Research, 28(8), pp. 1605–10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/5691758> (Accessed: 13 June 2018).

Samai, P. *ym.* (2015) ‘Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity’, *Cell*. Cell Press, 161(5), pp. 1164–1174. doi: 10.1016/J.CELL.2015.04.027.

- Sapranaukas, R. *ym.* (2011) ‘The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*.’, *Nucleic acids research*. Oxford University Press, 39(21), pp. 9275–82. doi: 10.1093/nar/gkr606.
- Semenova, E. *ym.* (2011) ‘Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence.’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), pp. 10098–103. doi: 10.1073/pnas.1104144108.
- Sinkunas, T. *ym.* (2011) ‘Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system’, *The EMBO Journal*, 30(7), pp. 1335–1342. doi: 10.1038/emboj.2011.41.
- Sosio, M. *ym.* (2018) ‘Analysis of the Pseudouridimycin Biosynthetic Pathway Provides Insights into the Formation of C-nucleoside Antibiotics’, *Cell Chemical Biology*, 25(5), p. 540–549.e4. doi: 10.1016/j.chembiol.2018.02.008.
- Sternberg, S. H. *ym.* (2014) ‘DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.’, *Nature*. NIH Public Access, 507(7490), pp. 62–7. doi: 10.1038/nature13011.
- Swarts, D. C. *ym.* (2012) ‘CRISPR interference directs strand specific spacer acquisition.’, *PLoS one*. Public Library of Science, 7(4), p. e35888. doi: 10.1371/journal.pone.0035888.
- Tong, Y. *ym.* (2015) ‘CRISPR-Cas9 Based Engineering of Actinomycetal Genomes’, *ACS Synthetic Biology*. American Chemical Society, 4(9), pp. 1020–1029. doi: 10.1021/acssynbio.5b00038.
- Treangen, T. J. *ym.* (2009) ‘Genesis, effects and fates of repeats in prokaryotic genomes’, *FEMS Microbiology Reviews*. Oxford University Press, 33(3), pp. 539–571. doi: 10.1111/j.1574-6976.2009.00169.x.
- Uehara, Y., Fisher, J. M. ja Rabinovitz, M. (1980) ‘Showdomycin and its reactive moiety, maleimide. A comparison in selective toxicity and mechanism of action in vitro.’, *Biochemical pharmacology*, 29(16), pp. 2199–204. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7426025> (Accessed: 26 June 2018).
- Vercoe, R. B. *ym.* (2013) ‘Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands’, *PLoS Genetics*. Edited by D. Hughes, 9(4), p. e1003454. doi: 10.1371/journal.pgen.1003454.
- Ward, J. M. *ym.* (1986) ‘Construction and characterisation of a series of multi-copy promoter-probe plasmid vectors for *Streptomyces* using the aminoglycoside phosphotransferase gene from Tn5 as indicator.’, *Molecular & general genetics : MGG*, 203(3), pp. 468–78. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3018431> (Accessed: 11 December 2018).
- Wei, Y. *ym.* (2015) ‘Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*’, *Nucleic Acids Research*. Oxford University Press, 43(3), pp. 1749–1758. doi: 10.1093/nar/gku1407.
- Wei, Y., Terns, R. M. ja Terns, M. P. (2015) ‘Cas9 function and host genome sampling

in Type II-A CRISPR-Cas adaptation.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 29(4), pp. 356–61. doi: 10.1101/gad.257550.114.

Wellington, K. W. ja Benner, S. A. (2006) ‘A review: synthesis of aryl C-glycosides via the heck coupling reaction.’, *Nucleosides, nucleotides & nucleic acids*, 25(12), pp. 1309–33. doi: 10.1080/15257770600917013.

Wiedenheft, B. *ym.* (2011) ‘RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(25), pp. 10092–7. doi: 10.1073/pnas.1102716108.

Xue, C., Whitis, N. R. ja Sashital, D. G. (2016) ‘Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity’, *Molecular Cell*. Cell Press, 64(4), pp. 826–834. doi: 10.1016/J.MOLCEL.2016.09.033.

Yang, W. (2008) ‘An equivalent metal ion in one- and two-metal-ion catalysis’, *Nature Structural & Molecular Biology*, 15(11), pp. 1228–1231. doi: 10.1038/nsmb.1502.

Yanisch-Perron, C., Vieira, J. ja Messing, J. (1985) ‘Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors.’, *Gene*, 33(1), pp. 103–19. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2985470> (Accessed: 10 March 2018).

Ylihonko, K. *ym.* (1996) ‘A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and complementation of early-block mutations in the anthracycline pathway.’, *Molecular & general genetics : MGG*, 251(2), pp. 113–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8668120> (Accessed: 10 March 2018).

Yosef, I., Goren, M. G. ja Qimron, U. (2012) ‘Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*.’, *Nucleic Acids Research*. Oxford University Press, 40(12), pp. 5569–5576. doi: 10.1093/nar/gks216.

Zebec, Z. *ym.* (2014) ‘CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*’, *Nucleic Acids Research*, 42(8), pp. 5280–5288. doi: 10.1093/nar/gku161.

Zhang, M. M. *ym.* (2017) ‘CRISPR–Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters’, *Nature Chemical Biology*. Nature Publishing Group, 13(6), pp. 607–609. doi: 10.1038/nchembio.2341.

Zuo, Z. ja Liu, J. (2016) ‘Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations’, *Scientific Reports*. Nature Publishing Group, 6(1), p. 37584. doi: 10.1038/srep37584.

Zuo, Z. ja Liu, J. (2017) ‘Structure and Dynamics of Cas9 HNH Domain Catalytic State’, *Scientific Reports*, 7(1), p. 17271. doi: 10.1038/s41598-017-17578-6.