

Immunokomplekseja sitovien vasta- aineiden tunnistaminen sekvenssianalyysillä

Diplomityö
Turun yliopisto
Biokemian laitos
Biotekniikka
Syyskuu 2019
Samu Sontag

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

Tiivistelmä

Uuden polven sekvensointimenetelmät ovat kehittyneet viimevuosina nopeasti. Sekvensointilaitteiden kehittymisen myötä menetelmät ovat muuttuneet yhä luotettavammiksi, joustavimmiksi ja edullisimmiksi. Tämä on mahdollistanut uusien vasta-aineseulontamenetelmien kehittämisen, ja ennustaa uuden polven sekvensointimenetelmiin pohjautuvien tekniikoiden yleistymistä vasta-ainekehityksessä.

Turun yliopiston biotekniikan laitoksella kehitettyjen synteettisten vasta-ainekirjastojen avulla on mahdollista seuloa vasta-aineita, joita on vaikea kehittää perinteisen eläinten immunisaation avulla. Tällä hetkellä faaginäyttötekniikan ongelmana on parhaiden vasta-aineiden löytäminen tuhansien kloonikandidaattien joukosta. Uuden polven sekvensointimenetelmät tarjoavat mahdollisuuden valvoa vasta-aineiden rikastumista faaginäyttöseulonnan aikana. Lisäksi käyttämällä DNA-pohjaisia viivakooditunnisteita voidaan sekvensoida yhden ajon aikana jopa kymmeniä eri näytteitä. Yhdistämällä kirjastoja voidaan vasta-aine kehitykselle selvittää optimaaliset olosuhteet pelkän uuden polven sekvensoinnista saadun tiedon perusteella.

Tässä työssä hyödynnettiin uuden polven sekvensointitekniikkaa sekä bioinformatiikan työkaluja skatoli-, testosteroni-, mikrokystiini-LR- ja mikrokystiini-RR-immunokomplekseja sitovien vasta-aineiden tunnistamisessa, sekä niiden sitomisominaisuuksien määrittämisessä. Näytteinä käytettiin faaginäyttötekniikan ensimmäisen ja neljännen selektiokierroksen jälkeen eristettyjä kokonaisia scFv-M13-faageja sekä niistä eristettyä faagemidi-DNA:ta. Näytteet viivakoodattiin DNA-tunnistusalukkeilla ja sekvensoitiin Illumina Miseq sekvensointialustalla. Bioinformatiikkaa hyödyntäen sekvensseistä pystyttiin tunnistamaan CDRH3-alueet, joiden rikastumiskertoimia vertailemalla määritettiin sekvenssit, jotka ennustivat hyvää sitoutumista immunokomplekseja vastaan. Lisäksi sitoutumisominaisuuksien määrittämiseksi sekvenssien rakenteita vertailtiin toisiinsa klusterianalyysissä. Tutkimuksessa onnistuttiin löytämään uusia hyvin rikastuneita, mikrokystiini-LR:lle spesifisiä CDRH3-sekvenssejä, joita ei ole aiemmin pystytty Turun yliopistossa havaitsemaan perinteisiä menetelmiä käyttäen. Testosteroni- ja skatolikirjastoissa

sekvenssien monimuotoisuus oli huomattavasti suurimpi kuin pidemmälle rikastetuissa mikrokystiinikirjastoissa. Skatolin tai testosteroni muodostamille immunokomplekseille ei tässä työssä onnistuttu sekvensointidatan perusteella löytämään muita selkeästi enemmän rikastuneita ja spesifisempinä erottuvia CDRH3-sekvenssejä.

Abstract

Next generation sequencing methods have evolved rapidly over the recent years. Through the development of sequencing instruments, sequencing methods have become more robust, reliable, flexible and cost effective. This has made it possible to develop new screening methods for antibody research, which predicts technologies based on sequencing to become ever more popular in antibody engineering in future.

With the use of synthetic antibody libraries developed in the department of biotechnology in University of Turku it is possible to obtain antibodies that would be extremely difficult to achieve with conventional methods based on immunization of animals. At the moment the challenge is how to find the potential antibodies among the thousands of candidates. Next-generation sequencing methods offers a possibility to monitor the enrichment of antibodies simultaneously with the phage display selection process. Additionally, with the use of DNA-based barcoded sequencing primers, it is possible to analyze up to tens of libraries in a single next-generation sequencing run. By combining different libraries optimum conditions for antibody development can be revealed with the information gained from next generation sequencing only.

In this study next-generation sequencing was used in combination with bioinformatics tools in order to identify skatole-, testosterone-, microcystin-LR and microcystin-RR-immune complex binding antibodies. The samples used were whole scFv-M13-phages and phagemids extracted from phages that had been isolated after the first and fourth phage display selection rounds. Samples were barcoded with DNA-based identification primers and sequenced with Illumina Miseq sequencing platform. With the use of bioinformatics tools CDRH3-regions of the sequences were identified and analyzed. By determining and comparing abundances of different CDRH3s between the libraries, sequences predicting good binding properties were identified. Additionally, with cluster analysis the sequence structures were compared in order to determine the properties affecting binding. In the study new and efficiently enriched CDRH3-sequences were found that showed high specificity against microcystin-LR. These sequences had not been able to be found in University of Turku before with using conventional methods. In testosterone- and skatole libraries the diversity of sequences was significantly higher

compared to the sequences in the further enriched microcystin libraries. In this study based on sequencing data only no CDRH3-sequences enriched clearly more efficiently than others while having also high specificity could be found from skatole or testosterone libraries.

Sisällysluettelo

LYHENTEET	1
JOHDANTO	3
SKATOLI JA KARJUNHAJU	5
TESTOSTERONI JA HYPOGONADISMI	8
MIKROKYSTIINIT JA SINILEVÄ	11
KILPAILEVA JA EI-KILPAILEVA IMMUNOMÄÄRITYS	14
IMMUNOKOMPLEKSEJA SITOVIIIN VASTA-AINEISIIN PERUSTUVA EI-KILPAILEVA IMMUNOMÄÄRITYS.....	16
IMMUNOKOMPLEKSEJA SITOVIIEN VASTA-AINEIDEN KEHITYS.....	17
FAAGINÄYTTÖTEKNIikka	18
FAAGINÄYTTÖTEKNIikka JA UUDEN POLVEN SEKVENSOINTI	20
MATERIAALIT JA MENETELMÄT	22
KIRJASTOT	22
SEKVENSOINTIKIRJASTOJEN VALMISTELU	23
SEKVENSOINTI	25
NGS DATA-ANALYYSI.....	25
TULOKSET JA NIIDEN TARKASTELU	30
NGS-DATANKÄSITTELY	32
PÄÄTELMÄT	39
LÄHTEET	40

Lyhenteet

aa	aminohappo (engl. amino acid)
bp	emäspari (engl. base pair)
CDR	spesifisyyttä määrittävä alue (engl. complementary determining region)
CDRH3	raskasketjun kolmas spesifisyyttä määrittävä alue (engl. third complementary determining region of heavy chain)
Da	Dalton, atomimassayksikkö $1,66 \times 10^{-24}$ g eli 1/12 hiiliatomin massasta
dsDNA	kaksijuosteinen DNA (engl. double-stranded DNA)
EC50	konsentraatio, jolla saavutetaan 50 % maksimaalisesta antigeenin sitoutumisesta (engl. effective concentration 50 %)
ELISA	entsyymivälitteinen immunosorbenttimääritys (engl. enzyme-linked immunosorbent assay)
HPLC	korkean erotuskyvyn nestekromatografia (engl. high-performance liquid chromatography)
IC	immunokompleksi (engl. immune complex)
LD50	kuolettava annos 50 %:lle (engl. median lethal dose)
LH	luteinisoiva hormoni (engl. luteinizing hormone)
Mc0	kirjasto, jonka faaginäyttöselektiossa ei ollut käytetty mikrokystiiniä
Mc00	kirjasto, jonka faaginäyttöselektiossa ei ollut käytetty mikrokystiiniä eikä sitojavasta-ainetta
Mc0st	Ennen mikrokystiini-faaginäyttöselektiota eristetty faagistokki
Mc3rd	Kolmannen mikrokystiini-faaginäyttöselektion jälkeen eristetty faagistokki
McLR01	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 0,1 nM mikrokystiini-LR:ää sitoutuneena vasta-aineeseen.
McLR1	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 1 nM mikrokystiini-LR:ää sitoutuneena vasta-aineeseen.

McLR30	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 30 nM mikrokystiini-LR:ää sitoutuneena vasta-aineeseen.
McRR01	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 0,1 nM mikrokystiini-RR:ää sitoutuneena vasta-aineeseen.
McRR1	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 1 nM mikrokystiini-RR:ää sitoutuneena vasta-aineeseen.
McRR30	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 30 nM mikrokystiini-RR:ää sitoutuneena vasta-aineeseen.
scFv	yksiketjuinen variaabeli fragmentti (engl. single-chain variable fragment)
SHBG	sukupuolihormoneja sitovaan globuliiniin (engl. sex hormone-binding globulin)
SKA0	Kirjasto, jonka faaginäyttöselektiossa ei ollut käytetty skatolia.
SKA50	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 50 pM skatolia sitoutuneena vasta-aineeseen.
SKA500	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 500 pM skatolia sitoutuneena vasta-aineeseen.
SKA5000	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 5000 pM skatolia sitoutuneena vasta-aineeseen.
ssDNA	yksijuosteinen DNA (engl. single-stranded DNA)
TES0	Kirjasto, jonka faaginäyttöselektiossa ei ollut käytetty testosteronia.
TES50	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 50 pM testosteronia sitoutuneena vasta-aineeseen.
TES500	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 500 pM testosteronia sitoutuneena vasta-aineeseen.
TES5000	Kirjasto, jonka faaginäyttöselektiossa oli käytetty 5000 pM testosteronia sitoutuneena vasta-aineeseen.

Johdanto

Vaikka makromolekyylit ovat tärkeitä analyyttejä kliinisessä diagnostiikassa, valtaosa ympäristötieteiden, toksikologian, lääkeainepitoisuuksien seurannan ja bioturvallisuuden kannalta merkittävistä analyteistä on pienmolekyylejä (Carlomagno ja muut 2014). Immunokemiassa alle 5000 Daltonin (Da) painoiset pienmolekyylit, jotka reagoivat spesifisesti vasta-aineiden kanssa, mutta eivät voi aiheuttaa immuunivastetta ilman että, ne ovat sitoutuneena kantajaproteiiniin tai johonkin muuhun niitä suurempaan antigeeniseen molekyyliin on luokiteltu hapteneiksi. (Li ja muut 2018, Omi ja muut 2015). Hapteneilla voidaan viitata endogeenisiin yhdisteisiin, kuten hormoneihin, eksogeenisiin lääkeaineisiin tai luonnollisiin ja epäluonnollisiin ympäristökontaminantteihin kuten sienimyrkkyihin ja torjunta-aineisiin. Hapteenien mittaaminen niin ympäristö- ja elintarviketurvallisuuden valvonnassa kuin kliinisen diagnostiikan sovelluksissa edellyttää nopeita ja herkkiä määrittästeknologioita maksimaalisen tehokkuuden ja oikeanmukaisuuden varmistamiseksi. Hapteenien määrittämiseen liittyvät julkaisut ovat kasvavassa määrin käsitelleet korkean erotuskyvyn nestekromatografiaan (engl. high-performance liquid chromatography, HPLC), kaasukromatografiaan, nestekromatografia-tandem-massaspektrometriaan sekä pinta-plasmoniresonanssiin perustuvia menetelmiä. Vaikka näillä menetelmillä päästään korkeaan spesifisyyteen ja herkkyyteen, ne eivät sovellu nopeatempoiseen kenttäkäyttöön johtuen niiden vaatimasta monimutkaisesta ja usein aikaa vievästä näytteiden esikäsittelystä, kalliista laitteistosta sekä erityiskoulutetusta henkilökunnasta. Immunomäärittästekniikoilla, jotka perustuvat vasta-aineiden ja antigeenien sitoutumisominaisuuksiin, on muihin tekniikoihin nähden monia etuja kuten alhainen hinta, helppokäyttöisyys, suuri kapasiteetti sekä nopeus. (Li ja muut 2018.)

Hapteenien havaitsemiseen käytettäviä teknologioita rajoittaa hapteenien pieni koko. Pienen molekulaarisen kokonsa takia hapteenimolekyyleissä ei ole tilaa kahdelle vasta-aineelle sitoutua saman aikaisesti, mikä estää antigeenin kahden epitoopin tunnistamiseen perustuvien, ei-kilpailevien kerros-immunomäärittysten kehittämisen niille. Tähän päivään asti hapteenien havaitsemiseen on pääosin käytetty kilpailevia immunomäärittäksiä, joiden analyttinen herkkyys tai tarkkuus eivät yllä ei-kilpailevien

määritysten tasolle. (Omi ja muut 2015.) Tarve paremmille pienmolekyyliäärityksille on kiihdyttänyt immunokomplekseja (engl. immune complex, IC) sitovien vasta-aineiden kehitystä, joita hyödyntämällä on mahdollista kehittää ei-kilpailevaan formaattiin perustuvia immunomäärityksiä myös pienmolekyyleille (González-Techera ja muut 2007). Auto-antigeensyydestä ja epävakaudesta johtuen IC:ja sitovien vasta-aineiden tuottaminen on kuitenkin vaikeaa perinteisillä eläinten immunisaatioon perustuvilla menetelmillä (Pulli ja muut 2005).

Faaginäyttökniikka on tehokas työkalu sitojaproteiinien seulomiseen, jolla on nykyään vakiintunut asema vasta-ainekehityksessä. Menetelmässä kohdetta spesifisesti tunnistavat molekyylit selektoidaan kirjastoista rikastuttamisprosessin avulla. (Yang ja muut 2017). Vasta-ainekirjastoja ja faaginäyttökniikkaa hyödyntäen on mahdollista tuottaa IC:ja sitovia vasta-aineita in vitro ja samalla kiertää eläinten immunisaatioon liittyviä ongelmia (Arola ja muut 2016, Lövgren ja muut 2016). Yksittäisen faagikloonin monistumiseen isäntäsolussa vaikuttaa monet tekijät. Tekniikan rajoitteena on se, ettei rikastumisastetta pystytä kontrolloimaan selektiokierrosten aikana, jonka seurauksena osasta klooneja tulee kirjastoissa yliedustettuja. Riskinä on menettää hyviä sitojia, sillä joissain tilanteissa huonommin rikastuneet kloonit, jotka tulevat harvemmin valituiksi, saattavat myös pitää sisällään hyviä sitojia. (Lövgren ja muut 2016, Yang ja muut 2017).

Viime aikoina uuden polven sekvensointimenetelmiä (engl. next-generation sequencing, NGS) on alettu käyttää kokonaisten faaginäyttökniikan selektiokierrosten jälkeisten faagivasta-ainekirjastojen analysoimiseen. Sekvensointitietoa ja sen käsittelyyn kehitettyjä tietokoneohjelmia hyödyntämällä voidaan selvittää vasta-aineen spesifisyyden määrittävien alueiden (engl. complementary determining region, CDR) emäsjärjestys ja erottaa analyyttille spesifejä sekvenssejä ilman sattumanvaraiseen pesäkkeiden valintaan perustuvaa seulontaa. (Lövgren ja muut 2016.)

Tässä työssä edellä mainittuja menetelmiä hyödynnettiin IC:ja sitovien vasta-aineiden seulontaan. Kohdeanalyytteinä oli neljä eri hapteneia: skatoli, testosteroni, mikrokystiini-LR ja mikrokystiini-RR. Tavoitteena oli löytää vasta-aineita, jotka tunnistavat spesifisesti kyseisten hapteenien ja niitä sitovien vasta-aineiden muodostamia IC:ja. Seulonnassa analysointi rajattiin vasta-aineen spesifisyyden kannalta merkittävimmän rakenteen, raskasketjun kolmannen spesifisyyttä määrittävän alueen (engl. third complementary

determining region in heavy chain, CDRH3) sekvensseihin. CDRH3-sekvenssien perusteella pyrittiin määrittämään ainutlaatuiset kloonit ja klooniperheet, sekä selvittämään sitoutumisominaisuuksiin vaikuttavia sekvenssirakenteita. Työssä oli käytössä faaginäyttötekniikalla aiemmin eristettyjä vasta-ainefaageja, jotka oli selektoitu Turun yliopiston synteettisistä yksiketjuisten variaabeleiden fragmenttien (engl. single-chain variable fragment, scFv) vasta-ainekirjastoista edellä mainittuja IC:ja vastaan (Brockmann ja muut 2011, Huovinen ja muut 2013). Vasta-ainefaagit sekvensoitiin Illumina Miseq sekvensointiteknologiaa sekä DNA-viivakoodattujen tunnistusalukkeita hyödyntäen, joiden avulla eri kirjastoista peräisin olevat sekvenssit identifioitiin sekvensoinnin jälkeen. Sekvenssianalyysissä IC:lle spesifisiä sekvenssejä etsittiin vertailemalla eri analyttikirjastojen CDRH3 frekvenssejä keskenään, jonka jälkeen spesifisyyteen vaikuttavia aminohappojen (engl. amino acid, aa) sekvenssirakenteita kartoitettiin klusterianalyysillä.

Skatoli ja karjunhaju

Euroopan Unionissa teurastetaan vuosittain yli 250 miljoonaa sikaa lihan tuotannon yhteydessä. Noin kaksi kolmasosaa näistä sioista tuotetaan viidessä valtiossa, jotka ovat Saksa (23 %), Espanja (16,1 %), Ranska (10,0 %), Puola ja Tanska (noin. 8 %). Esimerkiksi Belgiassa teurastetaan vuosittain noin 11 miljoonaa sikaa, mikä on noin 4,5 % EU:n sikojen kokonaistuotantomäärästä. Ruokavalion, asumusten sekä geneettisen jalostuksen voimakas kehitys kaupallisen tehokkuuden maksimoimiseksi ovat osaltaan syitä siihen miksi moderni sika ja sen kasvatukseen liittyvät tuotantojärjestelmät ovat muuttuneet erittäin paljon alkuperäisestä sian kotikasvatuksesta. (Bekaert 2012.)

Eläinten hyvinvointi on avainkysymys kasvattajille suurimmassa osassa maailmaa. Erityisesti se on tärkeää niille kasvattajille, jotka haluavat viedä porsaanlihatuotteita ulkomaille ilman rajoitteita. Kuluttajien ja jälleenmyyjien kasvaneet vaatimukset eettisemmän sianlihan tuotantomenetelmille painostavat hallituksia laatimaan joissain tapauksissa jopa ylimäärin asetuksia ja säännöksiä EU:n karjan ja sian kasvattajille.

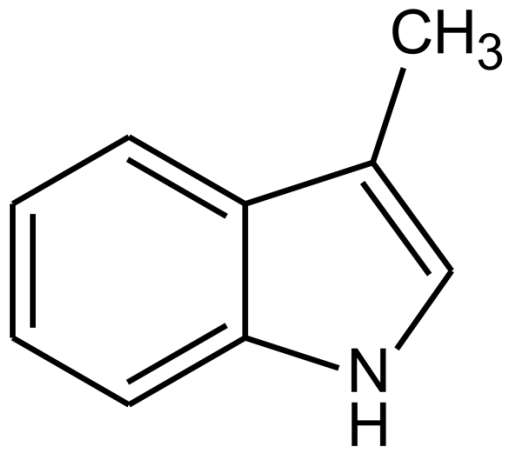
Esimerkiksi EU edellyttää, että sikoja tulee pitää ryhmissä, pois lukien emakot, joita saa pitää erillään alkaen yksi viikko ennen poikimista ja päättyen viimeistään neljä viikkoa poikimisen jälkeen. (Bekaert 2012.)

Uroseläinten kastraatiota on käytetty yleisesti lihantuotannossa jo satojen vuosien ajan. Suurin syy kastroimiseen on ollut urosten luonteen ja käyttäytymisen kontrollointi, minkä avulla urospuolisten eläinten käsittelyä on voitu helpottaa. Kastroimalla on myös saavutettu ruhon korkeampi rasvapitoisuus, mikä on aikoinaan ollut haluttu ominaisuus. Viime aikoina kuluttajien tottumukset ovat kuitenkin muuttuneet ja ihmiset ovat alkaneet vaatia vähärasvaisempaa lihaa. Tämän ja edullisemman tuotantohinnan takia suurin osa maista on luopunut naudan sekä lampaan kastroinnista. Sian kohdalla kirurgista kastrointia käytetään yhä poistamaan epämiellyttäväksi koettua karjun hajua, mitä vapautuu kastroimattoman urossian lihasta sitä kuumennettaessa. Lisäksi kastraatiolla voidaan kontrolloida urossikojen seksuaalista sekä aggressiivista käyttäytymistä. (Bekaert 2012, Bonneau 1998.) Tähän mennessä urossikojen kirurginen kastratio ilman anestesiaa on ollut yleisin menetelmä karjun hajun kontrolloimiseksi Euroopassa, mikä on tarkoittanut vuosittain noin 94 miljoonan porsaan kastraatiota (Fredriksen ja muut 2009). Viime vuosina kastraatiosta on kuitenkin tullut merkittävä kysymys eläinten hyvinvoinnin suhteen. Tutkimukset ovat osoittaneet ilman anestesiaa suoritettavan kirurgisen kastraation aiheuttavan kipua sioille, vaikka se tehtäisiin porsaiden ollessa erittäin nuoria (Marx ja muut 2003). Euroopassa suurin osa kastroimattomista urossioista kasvatetaan Iso-Britanniassa, Irlannissa, Espanjassa sekä Portugalissa, kun taas muissa EU-maissa suurin osa urosporsaista kastroidaan niiden ollessa nuoria. Kastroimattoman urossian lihaan suhtaudutaan monissa maissa kriittisesti ja esimerkiksi Tanskan yritys viedä kastroimattomia urospuolisia sikoja herätti vastustusta etenkin sen pääasiallisten asiakkaiden toimesta Saksassa ja Ranskassa, vaikka EU:n direktiivissä sallitaankin kastroimattomista urossioista peräisin olevan lihan kauppa ilman rajoituksia alle 80 kg painoisien ruhojen osalta ja edellytetään yli 80 kg painoisten ruhojen tutkiminen karjun hajun varalta. Direktiivin ongelmana on, ettei 80 kg rajalle ole olemassa tieteellistä näyttöä, ja se, ettei EU:ssa ole yleisesti hyväksyttyä objektiivista mittaumenetelmää karjun hajun arvioimiseksi teurastuslinjalla. (Bonneau 1998.)

Yksi syy olla kastroimatta urossikoja on niiden huomattavasti alhaisemmat tuotantokulut verrattuna kastroitujen sikojen kasvatukseen. Kun kastrointi jätetään pois vältytään

kirurgisen toimenpiteen kuluilta, mahdollisilta eläinten menetyksiltä ja komplikaatioilta sekä operaation porsaalle aiheuttamilta traumaailta. Kastroimattomat urossiat tarvitsevat vähemmän ravintoa kasvaakseen ja usein kasvavat nopeammin kuin kastroidut lajitoverinsa. Lisäksi niillä on parempi vastustuskyky sairauksien varalle, mikä entisestään parantaa niiden tehokkuutta tuotannossa. Kastroimattomuus voi myös parantaa ruhon ja lihan tiettyjen ominaisuuksien laatua. (Bonneau 1998.) Kastroimattomien sikojen puhtaan lihan osuus ruhosta on suurempi kuin kastroitujen sikojen, mikä tekee niiden kasvatuksesta edullisempaa lihantuotannossa (Walstra 1974). Lisäksi kastroimattomat siat vapauttavat vähemmän typpipäästöjä ympäristöönsä (Desmoulin et al., 1974). Kastration poisjättämisen merkittävin haittavaikutus on kuitenkin sen ilmeinen vaikutus lihan aistittavaan laatuun, joka ilmenee yleisesti tunnettuna karjun hajuna (Bonneau 1998).

Karjun haju johtuu pääasiallisesti kahden yhdisteen, skatolin (3-metyyli-indoli) ja urossian sukupuolihormoni androstenonin (5-alfa-androst-16-en-3-oni) kerääntymisestä urossikojen rasvakudokseen. Yhdisteet aiheuttavat virtsaa ja ulostetta muistuttavan epämiellyttävän aromin, joka vapautuu sianlihasta sitä kuumennettaessa. Skatoli (kuva 1) on pieni, noin 130 Da kokoinen yhdiste, jota muodostuu paksusuolen bakteerien tryptofaanimetabolian sivutuotteena sekä uros- että naarassioilla. (Leivo ja muut 2016, Vold 1970.) Urossioilla testikulaariset steroidit, erityisesti androstenoni, kuitenkin inhiboivat skatolin hajoamista maksassa, mikä johtaa suurempiin pitoisuuksiin kastroimattomilla urossioilla verrattuna naarassikoihin (Whittington ja muut 2004). Kirurginen kastraatio on pitkään ollut pääsääntöinen metodi karjun hajuun kontrolloimiseksi urossioilla. Viime vuosina porsaiden kirurginen kastraatio on alkanut herättää paljon eettisiä kysymyksiä, ja se on saanut runsaasti kritiikkiä julkisessa keskustelussa, minkä johdosta sitä on alettu välttelemään maailmanlaajuisesti. Vastatakseen näihin huoliin EU on antanut ohjeistuksen kirurgisten kastraatioiden lopettamisesta vuoteen 2018 mennessä. Tämä kehitys on lisännyt tarvetta nopeiden ja suhteellisen edullisten metodien kehittämiseen skatolin ja muiden karjun hajuun liittyvien yhdisteiden tunnistamiseksi. Tähän mennessä suurin osa skatolin tunnistamismenetelmistä on perustunut fysikaaliseen erotteluun, kromatografisiin menetelmiin, joiden käyttö vaatii usein kehittyneitä laitteistoa sekä koulutettua henkilökuntaa. Vain muutamia immunologisia menetelmiä on kehitetty, jotka nekään eivät vielä ole rutiinikäytössä. (Leivo ja muut 2016.)



Kuva 1. Skatolimolekyyli.

Testosteroni ja hypogonadismi

Kiveksen kaksi pääasiallista tehtävää on tuottaa siittiöitä sekä miessukupuolihormonia eli testosteronia (O'Donnell ja muut 2017). Testosteroni on steroideihin kuuluva, suhteellisen pieni, noin 288 Da kokoinen molekyyl (kuva 2), joka on miehen tärkein androgeeni (Kicman 2008). Se on hormoneista voimakkaimmin vastuussa miehen seksuaalisuudesta ja vastaa suurelta osin maskuliinisuuden liitettävistä ominaisuuksista kuten lihasmassasta, parrankasvusta, sukuvietistä ja siemennesteen tuotosta. Tämän lisäksi sillä on muita tärkeitä fysiologisia sekä psykologisia rooleja elimistössä. Testosteronin tuotanto alkaa miehillä vähentyä hiljalleen 30 ikävuoden jälkeen, minkä seurauksena veren testosteronikonsentraatio laskee miehillä tasaisesti ikääntymisen myötä. Testosteronin lasku saattaa miehillä aiheuttaa seksuaalisen halukkuuden vähenemistä, masennusta, väsymystä, energiatasojen laskua ja unettomuutta. (Almairan 2018.)

Testosteronin tuotanto aktivoituu miehen kehossa, kun aivokuori antaa hypotalamukselle signaalin stimuloida testosteronin tuotantoa. Tämän seurauksena hypotalamus vapauttaa gonadotropiinin vapauttajahormonia, joka edelleen stimuloi aivolisäkettä vapauttamaan gonadotropiineja: luteinisoivaa hormonia (engl. luteinizing hormone, LH) ja follikkelia

stimuloivaa hormonia. Vapauduttuaan verenkiertoon LH aktivoi kiveksen Leydigin solujen testosteronituotannon, jossa kolesterolista muokataan testosteronia. Kun testosteronitasot ovat nousseet sopivalle tasolle, aivolisäke hidastaa LH:n eritystä negatiivisen palautesäätelyn kautta vähentäen testosteronin tuotantoa. Testosteronin tuotanto on monimutkaisen prosessi, jonka yhteydessä monet potentiaaliset ongelmat voivat aiheuttaa testosteronitasojen laskemisen. Pienetkin muutokset kiveksissä, hypotalamuksessa tai aivolisäkkeessä voivat johtaa hypogonadismiin. (Kumar ja muut 2010.)

Hypogonadismilla tarkoitetaan sukupuolirauhasten, munasarjojen tai kivesten, heikentynyttä toimintaa. Sukupuolirauhasten tehtävä on tuottaa hormoneja kuten testosteronia, estradiolia, anti-Mullerian hormonia, progesteronia, inhibiini B:tä, aktiviiniä sekä naisilla munasoluja ja miehillä siittiöitä. Miesten hypogonadismi on erikseen karakterisoitu tilaksi, jossa kivekset eivät pysty tuottamaan tarpeeksi testosteronia. Tämä ilmenee miehillä testosteronin alentuneina plasmapitoisuuksina. Hypogonadismien testosteronipitoisuuden raja-arvona pidetään 300 ng/dl plasmassa yhdessä vähintään yhden kliinisen oireen kanssa. Tavallista alhaisempi testosteronipitoisuus miehillä voi aiheuttaa hedelmättömyyttä, sekundaaristen sukupuoliominaisuuksien vähänemistä, anemiaa, lihasten surkastumista, vyötärölihavuutta, seksuaalielämän häiriöitä, energian vähenemistä, masentuneisuutta, ärsyntyneisyyttä, keskittymisvaikeuksia, osteoporoosia, kolesterolitasojen muutoksia sekä kuumia aaltoja. Aikaisella diagnosoinnilla voidaan vähentää hypogonadismiin liittyviä riskejä. Pojilla voidaan esimerkiksi ennaltaehkäistä myöhästyvään puberteettiin liittyviä haittoja ja miehillä osteoporoosin ja muihin oireisiin liittyviä ongelmia. Hypogonadismien diagnostiikka perustuu pääasiassa oireisiin sekä veren testosteronipitoisuuden mittaamiseen. (Kumar ja muut 2010.)

Hypogonadismi voi olla joko hypothalamus-aivolisäkeperäistä, kivesperäistä tai niiden yhdistelmä. Myöhäisellä iällä alkavan miesten hypogonadismien ajatellaan olevan ikääntymisestä aiheutuva sairaus. Väestön ikääntymisen myötä miesten hypogonadismien suhteellisen osuuden ennustetaan kasvavan. Viime vuosina hypogonadismi on yleistynyt voimakkaasti ja tällä hetkellä 50, 60, 70 ja 80 vuotiailla miehillä noin 12 %, 19 %, 28 % ja 49 %: lla täyttyvät miesten hypogonadismien kriteerit. Maailman terveysjärjestö WHO:n arvion mukaan vuoteen 2020 mennessä yli 65 vuotiaiden määrä tulee

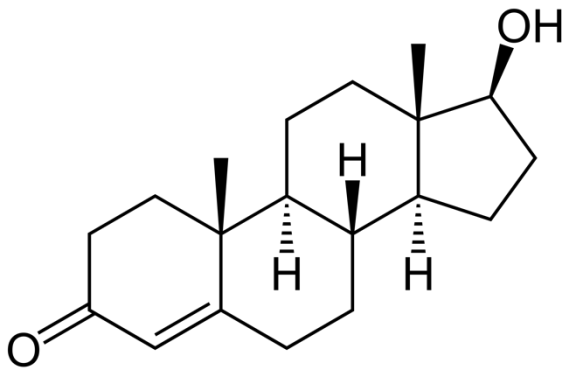
ensimmäistä kertaa ylittämään alle viisi vuotiaiden ihmisten määrän. Samojen arvojen mukaan yli 65 vuotiaiden määrä tulee kasvamaan nykyisestä 524 miljoonasta noin 1,5 miljardiin vuoteen 2050 mennessä. Arvioiden mukaan USA:ssa tulee vuoteen 2025 mennessä olemaan 6,5 miljoonaa miesten hypogonadismia sairastavaa miestä. Lisäksi ikääntymisen myötä kumuloituvat komorbiliteetit, sekä lihavuuden ja diabeteksen yleistyminen voivat osaltaan lisätä miesten hypogonadismin yleistymistä. Yleistyvien oireiden myötä tarve tehokkaille ja edullisille testosteronimäärityksille on kasvussa. (Amin ja muut 2016, Carnegie 2004, Kumar ja muut 2016.)

Terveillä miehillä suurin osa testosteronista kiertää plasmassa komplekseina sitoen eri proteiineja erilaisilla affiniteeteilla. Kokonaistestosteronista noin 50 % on heikosti sitoutuneena albumiiniin, noin 44 % on sitoutunut voimakkaasti sukupuolihormoneja sitovaan globuliiniin (engl. sex hormone-binding globulin, SHBG) ja noin 4 % on heikosti sitoutuneena kortikosteroideja sitovaan globuliiniin. Loput noin 2 % testosteronista kiertää plasman vesiliuoksessa vapaana molekyylinä. Kehon testosteronista biologisesti aktiivisen osuuden muodostaa kokonaistestosteroni miinus SHBG:tä sitovan testosteronin osuus. (Trost ja Mulhall 2016.)

Testosteronipitoisuuden määrittäminen tarkasti on pysynyt haasteellisena tähän päivään asti. Tällä hetkellä käytössä olevat kokonaistestosteronin mittaamenetelmät perustuvat joko kilpailevaan immunomääritykseen tai massaspektrometriaan. Kilpailevan testosteroni-immunomäärityksen huonona puolena on sen heikko tarkkuus suurilla ja pienillä testosteronipitoisuuksilla, sekä aikaa ja erikoisosaamista vaativa testosteronin eristys näytteestä. Massaspektrometrillä päästään hyvään tarkkuuteen, mutta sen käyttöä rajoittaa instrumenttien, sekä käytön ja instrumenttien kallis hinta. Kumpikin menetelmästä vaativat lisäksi harjaantunutta henkilökuntaa testin suorittamiseen. Kilpailevaan immunomääritystekniikkaan perustuva kokonaistestosteronin määrittäminen on sen hinnan ja tehokkuuden ansiosta yleisin kliinisessä työssä käytetty hypogonadismin testi. (Trost ja Mulhall 2016.)

Vapaata testosteronia pidetään yleisesti parempana hypogonadismin indikaattorina kuin kokonaistestosteronia johtuen ikääntymisestä, ylipainosta, hormonitasojen vaihteluista tai erilaisten sairauksien aiheuttamista SHBG:n pitoisuuden vaihteluista. Lisäksi vapaan testosteronin ajatellaan olevan parempi indikaattori erityisesti tapauksissa, joissa

kokonaistestosteronipitoisuudet ovat erittäin alhaiset. Vapaan testosteronin pitoisuus voidaan mitata joko suoraan määrittelyllä, kuten tasapainodialyysillä tai laskinnallisesti kokonaistestosteronipitoisuuden avulla. Tällä hetkellä käytössä olevat immunomääritykset vapaalle testosteronille ovat kilpailevaan formaattiin perustuvia, joilla ei päästä tarvittavaan herkkyyteen pienillä testosteronipitoisuuksilla. (Trost ja Mulhall 2017.) Tässä työssä tutkittiin mahdollisuuksia kehittää testosteronin ja vasta-aineen muodostamaa immunokompleksia tunnistavia vasta-aineita, jotka potentiaalisesti voisivat mahdollistaa helppokäyttöisemmän ja herkemmän, ei-kilpailevan immunomäärityksen kehittämisen vapaan testosteronin pitoisuuksien mittaamiseen.



Kuva 2. Testosteronimolekyyli.

Mikrokystiinit ja sinilevä

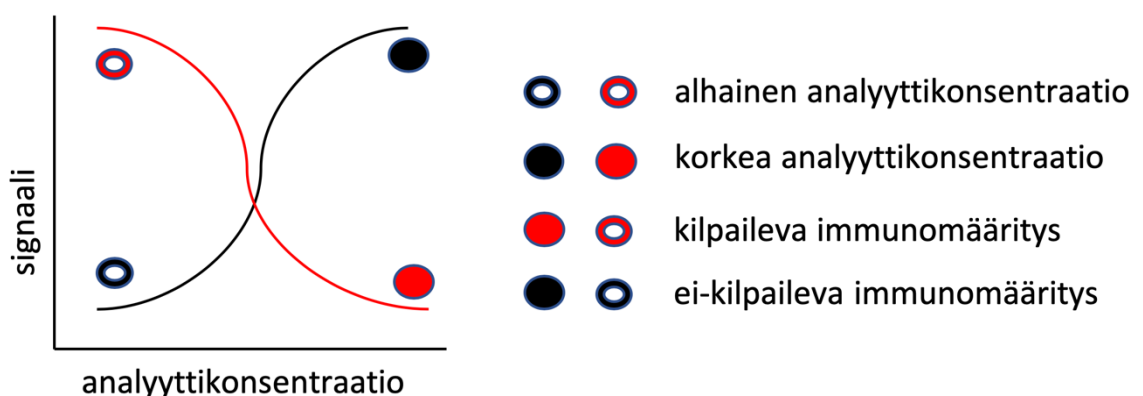
Syanobakteerit eli sinilevät ovat fotosynteesisiä organismeja, joita löytyy sekä makean veden ympäristöistä, että suolaisista vesistöistä (Schmidt ja muut 2014). Syanobakteerit ovat gramnegatiivisia prokaryootteja jotka pystyvät kasvamaan äärimmäisissäkin olosuhteissa kuten ultravioletivalossa, korkeassa suolapitoisuudessa, alhaisissa ravinnepitoisuuksissa sekä vähäisessä valossa (Kaushik ja Balasubramanian 2013). Monet syanobakteerilajit tuottavat sekundäärisiä metaboliitteja, jotka ovat myrkyllisiä kehittyneemmille eliöille. Myrkylliset metaboliitit voidaan jakaa hepatotoksiineihin,

dermatotoksiineihin, neurotoksiineihin ja sytotoksiineihin. Mikrokytiinit ovat ryhmä hepatotoksiineja, eli maksamyrkyihin kuuluvia syanobakteerien tuottamia sekundäärisiä metaboliitteja. Mikrokytiini-LR on niistä yleisin ja myös myrkyllisin. Myrkyllisyys johtuu muun muassa proteiinifosfataasien inhiboinnista. Mikrokytiini-LR häiritsee myös solurangan muodostumista sekä aiheuttaa maksakudoksen oksidatiivisen stressin lisääntymistä. Rotilla tehtyjen kokeiden mukaan oraalisesti nautittuna mikrokytiini-LR:n kuolettava annos 50 %:lle rotista (engl. median lethal dose, LD50) on noin 5 mg/kg ruumiinpainossa. Vertailun vuoksi syanidin vastaava luku on noin 3 mg/kg. Lisäksi pitkäaikainen altistuminen pienillekin mikrokytiini-LR määrille voi edistää kasvainten kehittymistä maksaan ja muihin elimiin. Mikrokytiini-RR on toinen, vähemmän myrkyllinen mikrokytiinin muoto, jonka LD50 on noin 50 mg/kg. (Schmidt ja muut 2014.)

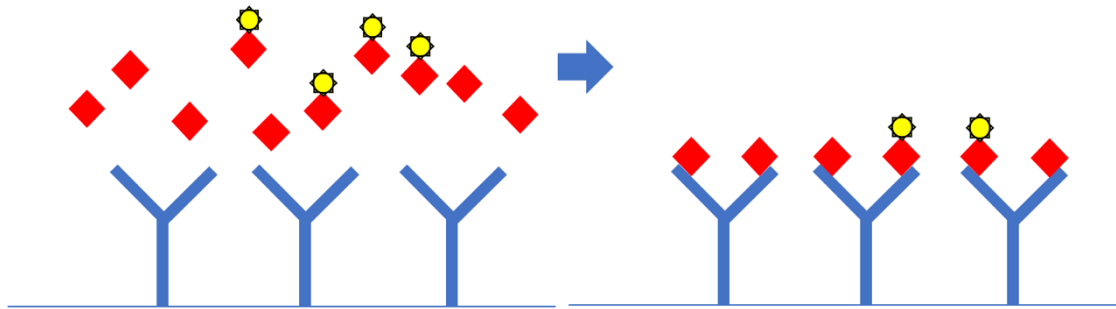
Sinileväkukinnot ovat vuosien mittaan yleistyneet, ja ihmisen aiheuttamat typpi- ja fosforipäästöt luonnonvesiin ovat edesauttaneet myrkyllisten leväkukintojen yleistymistä (Schmidt ja muut 2014). Tutkimusten mukaan makean veden kukinnoista noin 59 % sisältää myrkyllisiä metaboliitteja, joista hepatotoksiinit ovat neurotoksiineja yleisempiä. Sinilevien massakukinnot ovat yleinen ja maailmanlaajuinen ilmiö, joka uhkaa ihmisten hyvinvointia. (Rantala ja muut 2006.) Ihmiskehoon mikrokytiinejä päätyy suurimmaksi osaksi saastuneiden merenelävien, viljelyskasvien tai juomaveden nauttimisen yhteydessä. Tämä voi johtaa sairastumiseen tai pahimmassa tapauksessa jopa kuolemaan. (Rezaitabar ja muut 2017, Miller ja Russel 2017, Gaget ja muut 2017.) Esimerkiksi vuonna 1996 Brasiliassa 70 ihmistä sai surmansa altistuttuaan mikrokytiineille niin sanotussa ”Caruarun onnettomuudessa” (Liu ja Sun 2015). Kansainvälinen syöväntutkimuslaitos IARC on luokitellut mikrokytiini-LR:n mahdollisesti ihmiselle karsinogeeniseksi aineeksi (Lundqvist ja muut 2017). Maksan lisäksi mikrokytiini-LR:n on havaittu olevan haitallinen myös kivesten toiminnalle aiheuttaen muutoksia miesten hormonitasoissa, morfologisia muutoksia siittiöille sekä siittiöiden liikkumiskyvyn huononemista (Lone ja muut 2015). Maailman terveysjärjestö WHO on ilmoittanut mikrokytiini-LR:n päivittäiselle saannille ruuasta ihmisen eliniän aikana ylärajaksi 0,04 µg ruumiin painokiloa kohti päivässä ja juomaveden mikrokytiini-LR pitoisuuden turvallisuusrajaksi 1 µg/l (Schmidt ja muut 2014). Mikrokytiini-LR on tähän mennessä ainoa mikrokytiini, jolle on asetettu tämän kaltaiset ohjeistukset, mikä korostaa sen myrkyllistä profiilia ja lisää tarvetta kehittää herkempiä menetelmiä mikrokytiini-LR:n

Kilpaileva ja ei-kilpaileva immunomääritys

Immunomäärityksiä on olemassa kahta pääasiallista reaktiomuotoa, kilpaileva ja ei-kilpaileva (kuva 4). Joissain tapauksissa molekyyleissä on niiden pienen koon tai rakenteen johdosta tilaa vain yhdelle immunologiselle sitoutumiskohdalle. Näille analyyteille kilpaileva määrittäminen on ei-kilpailevaa sopivampi. Kilpailevassa määrittämisessä (kuva 5) kohdeanalyytti kilpailee samanlaisten, mutta leimattujen ja vakio konsentraatioissa pidettävien analyyttien kanssa sitoutumisesta immobilisoituun vasta-aineeseen. Kun sitoutumisreaktio saavuttaa tasapainotilan, sitoutumattomat analyytit pestään pois ja leimattujen analyyttien tuottama signaali mitataan, mikä on kääntäen verrannollinen kohdemolekyylin pitoisuuteen nähden. Kilpailevan määrittäksen herkkyyttä rajoittaa käytettyjen antigeenien ja vasta-aineiden välinen affiniteetti. Jos sitoutumisvakio on suurempi kuin 10^{12} mol/l voidaan saavuttaa 10^{-12} mol/l herkkyys. Valitettavasti vasta-aineiden sitoutumisvakio on yleensä kuitenkin alle 10^{10} mol/l, mikä takia kilpailevien määrittästen herkkyys jää yleensä suhteellisen alhaiseksi. Pienillä konsentraatioilla menetelmän virhemarginaalit kasvavat myös suuriksi, mikä johtaa huonoon tarkkuuteen ja toistettavuuteen sekä kapeaan lineaariseen mittausalueeseen. Tämä rajoittaa kilpailevien immunomääritysten käytettävyyttä esimerkiksi ympäristökontaminanttien havaitsemisessa. (Li ja muut 2018.)

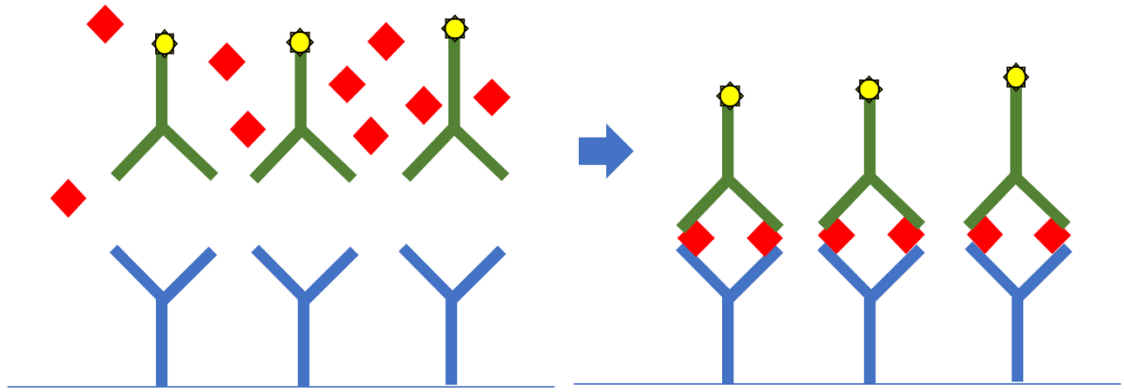


Kuva 4. Kilpailevan ja ei-kilpailevan immunomäärityksen ero. Kilpailevan immunomäärityksen signaali korreloi negatiivisesti analyttikonsentraatioon kun taas ei-kilpailevan immunomäärityksen signaali korreloi positiivisesti.



Kuva 5. Kilpaileva immunomääritys. Kohdeanalyytti ”kilpailee” sitoutumisesta immobilisoituun vasta-aineeseen muuten identtisen, mutta leimatun molekyylin kanssa, jonka pitoisuus tunnetaan. Sitoutumattomat molekyylit pestään pois, jonka jälkeen sitoutuneista leimoista voidaan lukea analyyttikonsentraatioon nähden kääntäen verrannollinen signaali.

Ei-kilpailevia immunomäärityksiä (kuva 6) käytetään yleensä havaitsemaan makromolekyyliä, joissa on useampia sitoutumiskohtia vasta-aineille. Tärkein ei-kilpaileva immunomääritysformaatti on niin sanottu kerros-ELISA, jossa kahta vasta-ainetta käytetään sitomaan samaa antigeeniä. Toinen vasta-aine on sidottu kiinteälle alustalle ja toimii sitojana kohdeantigeenille. Toinen vasta-aine aine on leimattu ja sitä käytetään tunnistamaan kohdeantigeeni. Leiman tuottama signaali on suoraan verrannollinen kohdeantigeenin pitoisuuteen näytteessä. Tehokkaan ei-kilpailevan immunomäärityksen kehittäminen hapteneille on riippuvainen käytössä olevista vasta-aineista. Vaikka vasta-aineita on tarjolla paljon eri spesifisyyksillä, on niiden käytössä omat rajoitteensa, ja joskus sovellusten vaatimia herkkyys, affiniteetti, stabiilius tai muita vaatimuksia ei voida täyttää. Näiden ongelmien kiertämiseksi on kehitetty uusia tunnistusmolekyyliä, kuten reseptoriproteiineja, rekombinanttivasta-aineita ja aptameerejä. Yksi uusi rekombinanttivasta-aineiden luoka on IC:ja tunnistavat vasta-aineet. Näitä hyödyntämällä on mahdollista kehittää ei-kilpailevia immunomäärityksiä myös hapteneille. (Li ja muut 2018.)



Kuva 6. Ei-kilpaileva kerros-ELISA-immunomääritys. Kohdeanalyytti sitoutuu immobilisoituun primääriseen vasta-aineeseen. Tämän jälkeen toinen, leimattu vasta-aine sitoutuu saman kohdeanalyytin toiseen epitooppiin muodostaen vasta-aine-analyytti-vasta-aine kerrosrakenteen. Sitoutumattomat vasta-aineet ja analyytit pestään pois, jonka jälkeen leimattujen vasta-aineiden avulla voidaan lukea suoraan kohdeanalyytin konsentraatioon verrannollinen signaali.

Immunokomplekseja sitoviin vasta-aineisiin perustuva ei-kilpaileva immunomääritys

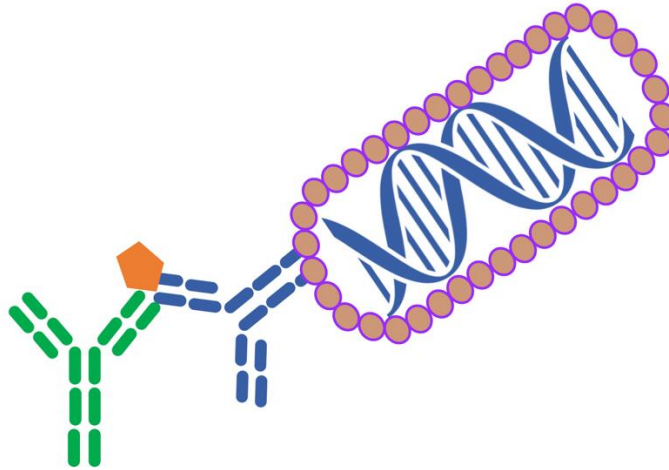
Ei-kilpailevat immunomääritykset perustuvat kahden vasta-aineen spesifiseen sitoutumiseen saman analyytin kahteen eri epitooppiin. Koska pienissä hapteneimolekyyleissä ei ole tarpeeksi epitooppeja kahdelle vasta-aineelle sitoutua samanaikaisesti, kilpailevat immunomääritykset ovat olleet pääasiainen valinta hapteneiden tunnistamisessa alhaisemmasta sensitiivisyydestä ja spesifisyydestä huolimatta. (Pulli ja muut 2005.)

Tähän mennessä ainoa tapa saavuttaa ei-kilpaileva immunomääritysformaatti hapteneille on ollut hyödyntää vasta-aineita, jotka tunnistavat primäärisen vasta-aineen ja pienmolekyylisanalyytin muodostaman IC:n (Arola ja muut 2016). IC:ja sitovat vasta-aineet hyödyntävät vasta-aineen ja antigeenin sitoutumiskohtaan muodostuvaa uutta epitooppia, johon IC:ja tunnistava vasta-aine spesifisesti sitoutuu. Tätä ominaisuutta hyödyntämällä voidaan kehittää IC:ja sitoviin vasta-aineisiin pohjautuva ei-kilpaileva immunomääritys myös hapteneille. (Li ja muut 2018.)

Immunokomplekseja sitovien vasta-aineiden kehitys

Ei-kilpailevan IC-määrityksen kehityksessä suurin haaste on tuottaa IC:ja spesifisesti tunnistavia vasta-aineita (Li ja muut 2018). Ensimmäiset IC:ja tunnistavat monoklonaaliset vasta-aineet kehitettiin käyttämällä perinteistä eläinten immunisaatioon perustuvaa menetelmää (Arola ja muut 2016). Näiden vasta-aineiden kehitys immunisaatiolla on kuitenkin työlästä ja tuottaa harvoin toivottua lopputulosta (Li ja muut 2016). Immunisaatiomenetelmän rajoitteena on vasta-aine-analyytti-kompleksien stabiilisuus elimistössä sekä usein niiden autoantigeenisuus (Arola ja muut 2016). Osa kehitystyön ongelmista johtuu myös pienistä variaatioista IC:eissa antigeenin ja vasta-aineen sitoessaan toisiaan. Lisäksi yli 85 % potentiaalisista sitoutumiskohdista hautautuu IC:n sisäosiin antigeenin ja vasta-aineen sitoutuessa toisiinsa, tehden näistä saavuttamattomia. Monimutkaisten IC-rakenteiden ja potentiaalisten sitoutumiskohtien ennustaminen on vaikeaa, mikä hankaloittaa niitä luotettavasti tunnistavien vasta-aineiden kehittämistä. (Li ja muut 2018.)

Immunisaation ongelmien välttämiseksi ei-kilpailevan IC-määrityksen kehityksessä voidaan hyödyntää faaginäyttökniikkaa, joka sallii vasta-aineiden seulonnan in vitro (Arola ja muut 2016, Li ja muut 2018). Faaginäyttökniikkaa ja universaaleja vasta-ainekirjastoja hyödyntämällä on mahdollista tuottaa sitojia lähes mille tahansa kohdemolekyylille nopeasti, sekä lisäksi molekyyleille, joita on vaikea saavuttaa perinteisellä immunisaatiomenetelmällä (Lövgren ja muut 2016). Tässä tekniikassa faaginäyttökirjastosta valikoidaan faaginäyttöselektiossa ne faagikloonit, jotka ilmentävät IC:ja spesifisesti sitovia vasta-aineita pinnallaan (kuva 7) (Li ja muut 2018).



Kuva 7. Immunokompleksia sitova vasta-ainefaagi.

Faaginäyttötekniikka

Faaginäyttötekniikka on George Smith:n vuonna 1985 kehittämä menetelmä, jossa polypeptidejä ilmennetään lysogeenisten säikeisten bakteriofaagien pinnalla (Smith 1985). Pelkistetysti kuvailen menetelmässä verrataan faagien pinnalla ilmennettävien proteiinien ja kohdeligandin välisiä proteiini-ligandi-vuorovaikutuksia perustuen niiden sitoutumisen affiniteettiin (Marintcheva 2018). Teknologia perustuu faagien fenotyypin ja genotyypin välillä olevaan suoraan fyysiseen yhteyteen, jossa faagin pinnalla ilmennettävää proteiinia koodaava geeni on pakattuna saman viruspartikkelin sisään yksijuosteisena DNA:na (engl. single-stranded DNA, ssDNA), jonka pinnalla ilmennettävä peptidi ekspressoidaan fuusioproteiinina faagin pintaproteiinin kanssa. (Kaplan ja Gershoni 2012.) Tämän genotyyppi-fenotyyppi-yhteyden ansiosta samassa bakteerissa tuotetut virukset ovat keskenään identtisiä. Faaginäyttötekniikalla voidaan luoda jopa 10^{10} variantin kokoisia kirjastoja, joita voidaan hyödyntää esimerkiksi vasta-aineiden kehityksessä (Bazan ja muut 2012).

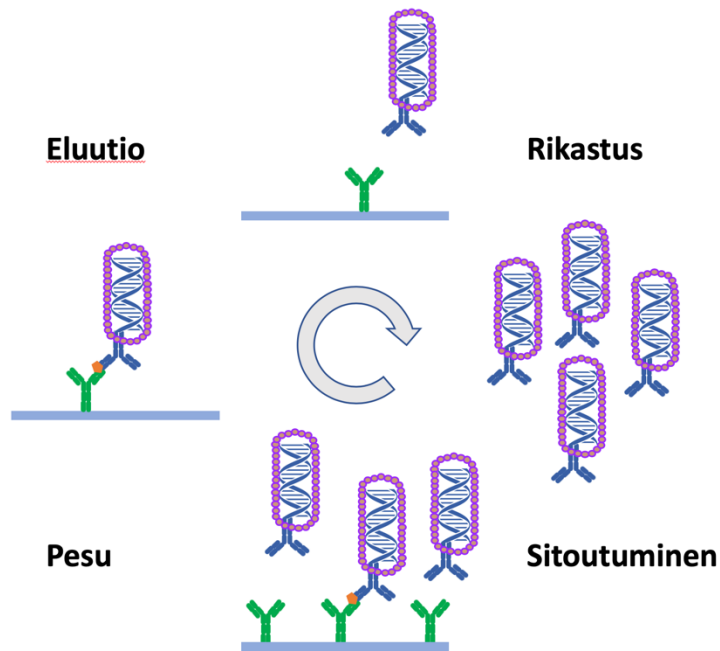
Faaginäyttötekniikka kehitettiin alun perin käyttäen niin sanottuja filamenttifaageja (Marintcheva 2018). Säikeiset faagit ovat ryhmä ei-lyyttisiä faageja, joilla on ssDNA:sta koostuva rengasmaisen genomi. M13 on ryhmän faageista yleisimmin käytetty. Se on

noin 7 nm leveä ja 900 nm pitkä, putkimainen bakteriofaagi, jonka genomi on noin 6400 emäsparin (engl. base pair, bp) pituinen. M13-faagin suurin etu suhteessa muihin faageihin on sen helppokäyttöisyys ja puhdistuksen vaivattomuus. M13-faagi sitoutuu bakterisolun F pilukseen, jonka seurauksena bakteeri infektoituu viruksen genomilla. Infektoituneet bakterisolut toimivat tuotantokoneistona uusien viruspartikkeleiden monistamisessa. M13-faagit eivät tuhoa bakteereita vaan ainoastaan hidastavat bakteerien jakautumista viruspartikkeleiden monistamisesta aiheutuvan stressin takia. (Rami ja muut 2017.)

Kirjastot koostuvat lukemattoman suuresta määrästä DNA-klooneja, joiden sisällä yksittäiset kiinnostuksen kohteena olevat sekvenssivariantit ovat kloonattuna samaan DNA-runkoon yhdessä monien replikaation, transkription, translaation ja kohdeproteiinin ilmentämisen säätelystä vastaavien sekvenssien kanssa. Faaginäyttökniikka alkaa kloonamalla tutkittavan proteiinin sekvenssi osaksi faagimidivektoria tai rekombinanttifaagia. Faagimidivektorit ovat niiden paremman vakauden ja muokattavuuden takia faagivektoreita suositumpi työkalu. Faagimidivektori sisältää sekä plasmidin että faagin replikaation aloituskohdat, sekvenssin DNA:n pakkaamiseen, valintamarkkerin sekä pintaproteiinin kloonamista, ekspressiota ja havaitsemista helpottavan proteiinikasetin. Faagimidit voidaan replikoida isäntäsolussa joko itsenäisesti tavallisena plasmidina, mikä sallii niiden muokkaamisen tai faagikoneiston avulla M13-auttajafaagin infektoidessa isäntäsolun, mikä mahdollistaa tutkittavaa proteiinia pinnallaan ilmentävien viruspartikkelien kokoamisen. Faaginäyttökirjastot luodaan tavallisilla kloonausmenetelmillä käyttäen faagimidin tai faagin genomien kaksijuosteisesta DNA (engl. double-stranded DNA, dsDNA) replikaatiomuotoa. Kirjastot transformoidaan kompetentteihin *E. Coli* bakteereihin, jotka alkavat tuottaa tutkittavaa proteiinia ilmentäviä rekombinanttifaageja.

Faaginäyttöselektiossa hyödynnetään affiniteettikromatografiaa muistuttavaa menetelmää, jossa rekombinanttifaageja seulotaan vasten kiinnostuksen kohteena olevaa analyyttiä ja identifioidaan sitä sitovat faagit. Selektiossa analyytti altistetaan faagien pinnalla ilmennetyille proteiineille, joista osa sitoutuu kohdemolekyylin. Sitoutumisen jälkeen sitoutumattomat faagit pestään pois ja sitoutuneet faagit eluoidaan. Eristetyt faagit rikastutetaan bakteereissa ja uusi, analyyttille kohdennettu kirjasto syötetään uuteen selektioon. Usean selektiokierroksen myötä jäljelle jää faageja, joiden ilmentämät

proteiinit omaavat hyvän affiniteetin analyyttiä kohtaan. Korkean affiniteetin faagit siirretään tuottumaan bakteereihin, jonka jälkeen pesäkkeistä valitaan yksittäisiä kloonveja, joiden ominaisuudet karakterisoidaan. Lopuksi analysoidaan eroteltujen faagien spesifisyys, affiniteetti sekä DNA-sekvenssit. (Marintcheva 2018.) Kuvassa 8 on esitetty havainnollistava kuva faaginäyttöselektiosta IC:ja sitoville vasta-aineille.



Kuva 8. Faaginäyttöselektio immunokompleksia vastaan. Vasta-aineita ilmentävät faagit ohjataan alustalle, joka on pinnoitettu immobilisoiduilla vasta-aine-antigeenikompleksilla, joihin osa vasta-ainefaageista spesifisesti sitoutuu. Sitoutumattomat vasta-ainefaagit pestään pois, jonka jälkeen sitoutuneet faagit eluoidaan. Eristetyt faagit syötetään samalle alustalle ja selektio toistetaan uudelleen. Useamman kierroksen myötä immunokomplekseja spesifisesti sitovat vasta-ainefaagit rikastuvat.

Faaginäyttötekniikka ja uuden polven sekvensointi

Faaginäyttötekniikka on vakiintunut tapa tuottaa spesifejä rekombinanttivasta-aineita erilaisten sovellusten tarpeisiin. Yhdessä universaalien vasta-ainekirjastojen kanssa sen avulla voidaan tuottaa tehokkaasti sitoja lähes mille tahansa analyyttille ilman tarvetta käyttää eläimiä. Menetelmällä voidaan kehittää vasta-aineita myös kohteille, joita on

vaikea saavuttaa eläinten immunisaatiolla. Faaginäyttötekniikan selektio perustuu tyypillisesti kahteen vaiheeseen, joista ensimmäisessä rikastutetaan analytyttiin sitoutuvia faageja. Toisessa vaiheessa bakteereissa rikastutetut ja yksittäisistä pesäkkeistä eristetyt vasta-ainekloonit analysoidaan immunomäärityksiin perustuvilla aktiivisuusanalyysillä. Ongelmana on, että yksittäisen faagivasta-aineen rikastumiseen bakteerissa vaikuttaa useat tekijät ja aina parhaimman affiniteetin omaavat proteiinit eivät syystä tai toisesta rikastu bakteereissa yhtä tehokkaasti kuin muut proteiinit. Tavallisessa faaginäyttötekniikan seulontavaiheessa vasta-aineet valitaan bakteeripesäkkeistä manuaalisesti, joista yleensä parhaiten rikastuneet tulevat valituiksi. Menetelmän riskinä on menettää hyviä sitoja, sillä on mahdollista, että huonommin rikastuneet pesäkkeet, jotka eivät tule valituiksi, sisältävät silti potentiaalisia vasta-aineita. Hyvien potentiaalisten sitojien menettämisen riski on erityisen suuri silloin kun kohdeanalytyttinä faaginäyttöselektiossa on monimutkainen rakenne. Riskin minimoimiseksi seulonnassa voidaan hyödyntää uuden sukupolven sekvensointimenetelmiä ja sekvensoida kaikki faaginäyttöselektion jälkeiset kloonit. (Lövgren ja muut 2016.)

NGS-sekvensointi on mullistanut genomiikan. Sen vaikutus vasta-aineseulontaan on ollut hitaampaa, mutta tulee todennäköisesti olemaan yhtä vaikuttava. Monella tapaa syväsekvensointi ja erilaiset näyttötekniikat lähestyvät täydellistä yhdistelmää. Kirjastoanalyysissä sekvensoitujen emästen kokonaismäärä ei ole yhtä tärkeää kuin lukujen pituus tai niiden lukumäärä. Tällä hetkellä yhdellä ainoalla Illumina Miseq sekvensointiajolla voidaan tuottaa noin 40 miljoonaa lukua. Synteettinen vasta-ainekirjasto voi pitää sisällään yli 10^9 erilaista varianttia. Yhden faaginäyttöselektiokierroksen jälkeen jäljelle jää enää kuitenkin vain noin 10^6 varianttia, mikä sallii useamman selektoitujen kirjastojen kokonaisvaltaisen analysoimisen yhdellä Miseq ajolla. Useamman selektiokierroksen myötä varianttien määrä pienenee entisestään samalla kun positiivisten kloonien osuus kasvaa huomattavasti. Tällä hetkellä saatavilla on monia erilaisia NGS-teknologioita, mutta Miseq:n tuotettujen lukujen suuri määrä ja suhteellisen edullinen hinta on tehnyt siitä suosituimman vaihtoehdon kirjastoanalyysissä. (Glenville ja muut 2015.)

Illumina-sekvensointitekniikka mahdollistaa noin 600 bp:n lukupituuden (V3 2 x 300 bp). Illumina-sekvensoinnin ja Turun yliopiston synteettisten scFv-vasta-ainekirjastojen avulla voidaan selvittää vasta-aineen kaikki CDR-sekvenssit. Sekvenssitiedon avulla

rikastetuista kirjastoista voidaan myös erotella yksittäisiä mielenkiintoisia vasta-aineklooneja ilman sattumanvaraiseen pesäkkeiden valintaan perustuvaa seulontaa. (Lövgren ja muut 2016.)

Materiaalit ja menetelmät

kirjastot

Tutkimuksessa käytettiin aiemmin Turun yliopistossa faaginäyttötekniikalla eristettyjä vasta-aineen ja antigeenin muodostamia IC:ja tunnistavia kokonaisia scFv-M13-vasta-aine-faageja sekä faageista eristettyä faagimidi-DNA:ta. Faaginäyttöselektiossa kohdeanalyytinä oli käytetty skatolia, testosteronia, mikrokystiini-LR:ää ja mikrokystiini-RR:ää. Analyytit olivat selektioissa sitoutuneena immobilisoituun vasta-aineeseen muodostaen IC:n. Jokaisella analyytillä oli tehty kolme rinnakkaista selektiota käyttäen eri analyyttikonsentraatioita sekä selektiot ilman analyyttiä.

Skatoli- ja testosteroni-IC:a vastaan valikoidut faagit oli eristetty faaginäyttötekniikan ensimmäisen selektiokierroksen jälkeen synteettisestä scFv-P-vasta-ainekirjastosta (Huovinen ja muut 2013). Skatoli-IC-faaginäyttöselektiossa oli käytetty immobilisoitua scFv (ESKA.A6) sitojavasta-ainetta (Turun yliopisto, Suomi) ja 0 pM, 50 pM, 500 pM ja 5 nM skatolia. Skatoli-IC-selektiosta eristetyistä faagivasta-ainekirjastoista käytetään tässä työssä lyhennyksiä SKA0, SKA50, SKA500 ja SKA5000 selektioissa käytettyjen skatolipitoisuuksien mukaisesti. Testosteroni-IC-faaginäyttöselektiossa sitojana oli käytetty immobilisoitua Fab SPTN-5 vasta-ainetta (Medix Biochemican, Suomi) ja 0 nM, 50 pM, 500 pM ja 5 nM testosteronia. Testosteroni-IC-selektioista eristetyistä faagivasta-ainekirjastoista käytetään lyhennyksiä TES0, TES50, TES500 ja TES5000 selektiossa käytettyjen testosteronipitoisuuksien mukaisesti. Sekä SKA- että TES-sekvensointikirjastojen valmisteluun käytettiin kokonaisia faageja.

Mikrokystiini-LR-IC:a ja mikrokystiini-RR-IC:a vastaan valikoidut faagit oli eristetty faaginäyttötekniikan neljännen selektiokierroksen jälkeen synteettisistä vasta-ainekirjastoista scFv-M ja scFv-P (Brockmann ja muut 2011, Huovinen ja muut 2013). Faageista oli tämän jälkeen eristetty niiden scFv-pEB32x-plasmidit käyttäen GeneJet Plasmid Miniprep Kit:iä (Thermo Fisher Scientific, USA). Sekä mikrokystiini-LR:n että mikrokystiini-RR:n selektioissa oli käytetty 0,1 nM, 1 nM ja 30 nM analyyyttiä sekä sitojana immobilisoitua Fab 226A2-vasta-ainetta (Turun yliopisto, Suomi). Mikrokystiini-LR-IC-selektioista eristetyistä kirjastoista käytetään tässä työssä lyhennyksiä McLR01, McLR1, McLR30 selektioissa käytettyjen mikrokystiini-LR-pitoisuuksien mukaan. Mikrokystiini-RR-IC-selektioista eristetyistä kirjastoista käytetään lyhennyksiä McRR01, McRR1 ja McRR30 selektioissa käytettyjen mikrokystiini-RR-pitoisuuksien mukaisesti. Lisäksi mukana oli näyte, jonka selektiossa ei ollut käytetty yhtään mikrokystiiniä (Mc0), näyte jonka selektiossa ei ollut käytetty mikrokystiiniä eikä sitojavasta-ainetta (Mc00), faagistokkinäyte kolmannen selektiokierroksen jälkeen (Mc3rd) sekä faagistokkinäyte ennen ensimmäistä selektiota (Mc0st).

Sekvensointikirjastojen valmistelu

Mikrokystiininäytteiden scFv-pEB32x-plasmidit linearisoitiin HindIII restriktioentsyymillä käyttäen Recommended Protocol for Digestion -protokollaa (Thermo Fisher Scientific, USA), 10 ul reaktiotilavuutta sekä 2 µg DNA:ta per näyte.

Sekvensointikirjastot valmistettiin liittämällä jokaiseen näytteeseen ainutlaatuinen pari viivakoodattuja sekvensointialukkeita (Sigma-Aldrich, USA), joiden avulla eri näytteistä peräisin olevat sekvenssit pystyttiin tunnistamaan sekvensoinnin jälkeen. Alukkeiden liittäminen näytteisiin tehtiin PCR:ssä, missä samalla rikastutettiin faagien scFv-pEB32x plasmidien CDRL3 – CDRH3 alueet. Sekvensointialukkeet ja niiden rakenne, adapteri-indeksi-hybridisoituva alue, on lueteltu taulukossa 1.

Taulukko 1. Viivakoodatut sekvensointialukkeet. Alukkeen rakenne: Adapteri – viivakoodi – hybridisoituva alue.

Alukke	Sekvenssi 5' - 3'
REV_B	<u>CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCC</u> <u>GATCT-GTTACGTGGTTG-ACTTCGCGGTGTACTATT</u>
REV_D	<u>CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCC</u> <u>GATCT-ACTTTAAGGGTGT-ACTTCGCGGTGACTATT</u>
REV_E	<u>CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCC</u> <u>GATCT-TGTTGCGTTTCTGT-ACTTCGCGGTGACTATT</u>
REV_F	<u>CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCC</u> <u>GATCT-ATGTCCGACCAAGT-ACTTCGCGGTGACTATT</u>
REV_G	<u>CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCC</u> <u>GATCT-ACAGCCACCCATCGA-ACTTCGCGGTGACTATT</u>
REV_H	<u>CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCC</u> <u>GATCT-TGTCTCGCAAGCCGA-ACTTCGCGGTGACTATT</u>
FWD_1	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-CCTAAACTACGG-GACTAGTGTACCCTGACC</u>
FWD_2	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-TGCAGATCCAAC-GACTAGTGTACCCTGACC</u>
FWD_3	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-CCATCACATAGG-GACTAGTGTACCCTGACC</u>
FWD_5	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-ACTTTAAGGGTGA-GACTAGTGTACCCTGACC</u>
FWD_6	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-GAGCAACATCCTA-GACTAGTGTACCCTGACC</u>
FWD_7	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-TGTTGCGTTTCTTC-GACTAGTGTACCCTGACC</u>
FWD_8	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-ATGTCCGACCAATC-GACTAGTGTACCCTGACC</u>
FWD_9	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-AGGTACGCAATTC-GACTAGTGTACCCTGACC</u>
FWD_10	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-ACAGCCACCCATCTA-GACTAGTGTACCCTGACC</u>
FWD_11	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-TGTCTCGCAAGCCTA-GACTAGTGTACCCTGACC</u>
FWD_12	<u>AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC</u> <u>GATCT-GAGGAGTAAAGCCTA-GACTAGTGTACCCTGACC</u>

SKA0, SKA50, SKA500, SKA5000, TES0, TES50, TES500, TES5000 ja Mc0st kirjastot monistettiin Q5 High-Fidelity DNA Polymeraasilla (New England Biolabs, USA) valmistajan ohjeiden mukaisesti, sillä erotuksella että alukkeita käytettiin 0,15 µM ja mastermiksina käytettiin NEBNext Ultra II Q5 Master Mix:iä (New England Biolabs, USA). PCR-olosuhteet olivat: 98 °C 3 min., 6 x (98 °C 10 s → 60 °C 15 s → 72 °C 20 s), 30 x (98 °C 10 s → 72 °C 20 s) ja 72 °C 3 min. Mc0, Mc00, McLR01, McLR1, McLR30, McRR01, McRR30 sekä Mc3rd kirjastojen PCR:ssä käytettiin Q5 High-Fidelity DNA Polymeraasia valmistajan ohjeiden mukaisesti sillä erotuksella, että

alukkeita käytettiin 0,15 μ M. PCR-olosuhteet olivat: 95 °C 3 min., 15 x (98 °C 20 s → 62 °C 20 s → 72 °C 30 s). Kaikki PCR-ajot suoritettiin C1000 Touch Thermal Cycler:illa (BioRad, USA).

PCR-tuotteet ajettiin 1 % agarosigeelielektroforeesissa 65 voltin jännitteellä yhden tunnin ajan. Geeliväriä käytettiin Invitrogen SYBR Safe DNA Gel Stain 10000 x in DMSO (Invitrogen, USA), DNA-mitta-asteikkona GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, USA) ja latausleimana Thermo Scientific 6 x DNA Loading Dye (Thermo Fisher Scientific, USA). Ajon jälkeen geeliltä eroteltiin noin 600 bp:n fragmentit, ja puhdistettiin GeneJET Gel Extraction Kit:illä (Thermo Fisher Scientific, USA) valmistajan ohjeiden mukaisesti ja eluoiitiin 30 μ l tilavuuteen.

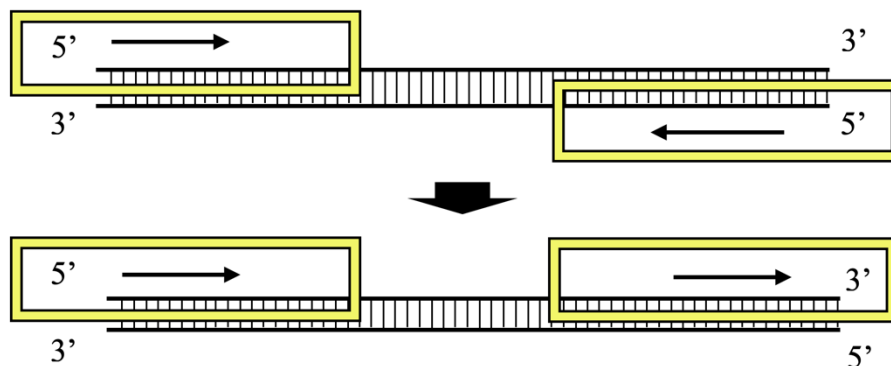
Sekvensointi

Puhdistettujen PCR tuotteiden molaarinen konsentraatio määritettiin Qubit 2.0 fluorometrillä (Invitrogen/LifeTechnologies/ThermoFisher Scientific, USA) käyttäen iQuant™ High Sensitivity dsDNA Quantitation Kit:iä (GeneCopoeia, USA), sekä Agilent High Sensitivity D1000 ScreenTape System:iä (Agilent, USA). Indeksoidut kirjastot yhdistettiin ja 6 pM näyte, jossa oli 10 % phiX viraalista DNA:ta sekvensoitiin Illumina Miseq instrumentilla (Illumina, USA). Sekvensointiin käytettiin V2 2 x 250 bp Illumina Miseq Reagent Kit:iä (Illumina, USA).

NGS data-analyysi

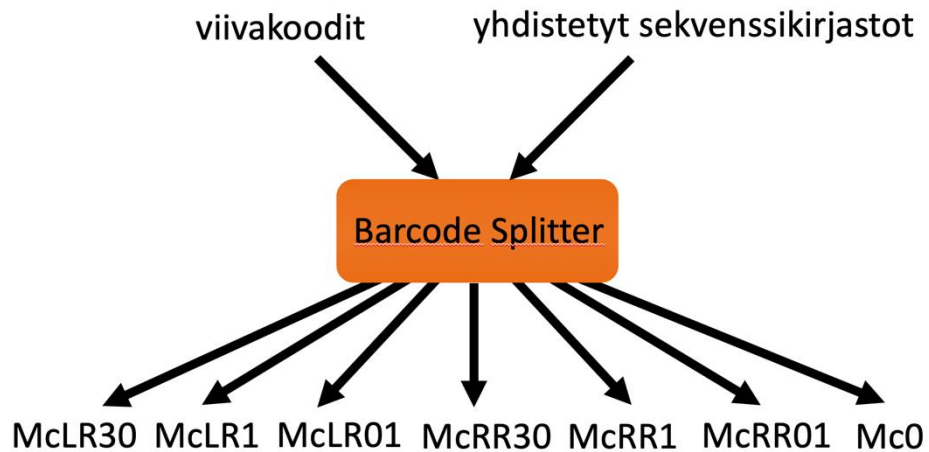
Illumina Miseq:stä saatiin ulos kaksi fastq-tiedostoa, joista toinen piti sisällään kaikkien näytteiden kaikki etuperinluvut ja toinen kaikki takaperinluvut. Datan esikäsittelyvaiheessa NGS-ajon laatu tarkistettiin FastQC-ohjelmalla (Andrews 2010).

Tämän jälkeen etu- ja takaperinluvut yhdistettiin yhteen tiedostoon. Koska sekvensoinnissa takaperinluku sekvensoitiin komplementaarista juostetta käyttäen vastakkaiseen suuntaan, piti se ennen yhdistämistä kääntää 3' – 5' suuntaisesta 5' – 3' suuntaiseksi ja sen jälkeen komplementaariseen muotoon (kuva 9). Lukuja ei voitu liittää toisiinsa päällekkäisten sekvenssien perusteella, koska 250 emäksen lukupituudet olivat tähän liian lyhyet, mutta koska etu- ja takaperinluvut olivat tiedostoissa samassa järjestyksessä, luvut voitiin yhdistää toisiinsa rivi kerrallaan. Lukujen yhdistäminen tehtiin käyttämällä ScFvMiner combiner -skriptiä (Lövgren ja muut 2016).



Kuva 9. Takaperin-luvun kääntäminen 3' – 5' suuntaisesta 5' – 3' suuntaiseksi ja komplementaariseksi.

Seuraavassa vaiheessa eri näytekirjastoista peräisin olleet sekvenssit eroteltiin omiin tiedostoihin viivakoodattujen sekvensointialukkeiden perusteella. Tähän käytettiin FASTX-Toolkit:n Barcode Splitter -skriptiä (Hannon 2009). Sekvenssien erottelu tehtiin kahdessa vaiheessa. Ohjelma etsi ensin kaikista sekvensseistä etuperinviivakoodit ja erotteli saman viivakoodin omaavat sekvenssit omiin tiedostoihin. Tämän jälkeen jokaiselle luodulle tiedostolle tehtiin haku takaperinviivakoodeilla. Sekvenssit eri kirjastojen mukaisilla viivakoodiyhdistelmillä sijoitettiin omiin tiedostoihin. Sekvenssien erottelu on havainnollistettu kuvassa 10.



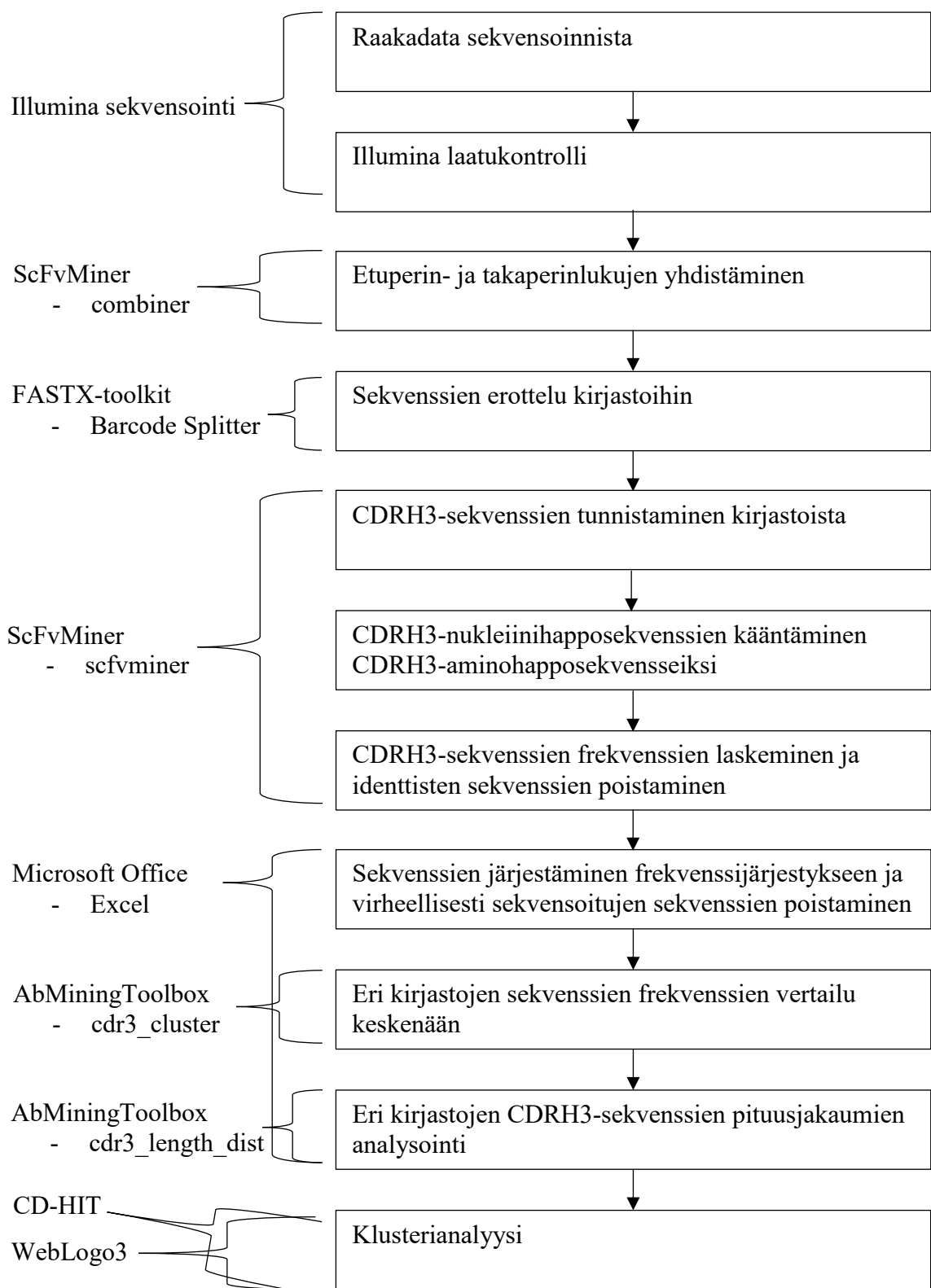
Kuva 10. Mikrokystiini-IC:ja tunnistavien vasta-aineiden sekvenssien erottelu omaiin tiedostoihin viivakoodattujen alukkeiden ja FASTX-toolkit Barcode Splitter:n avulla.

Eroteltujen kirjastojen sekvensseistä seulottiin CDRH3 alueet käyttäen ScFvMiner – scfvminer -skriptiä (Lövgren ja muut 2016). Yksi-lukukehyksinen scFv-sekvenssirakenne variaabelien alueiden välissä oli kiinteä (Brockmann ja muut 2011; Huovinen ja muut 2013), mitä hyödyntäen ohjelma etsi CDRH3 sekvenssit tunnettujen aloitus- ja lopetussekvenssien (ctgccgtctactactgt-CDRH3-tggggtcagggtacact) välistä. Ohjelma lisäksi laski jokaisen CDRH3 sekvenssin esiintymän jokaisessa kirjastossa ja käänsi nukleinihapposekvenssin aa-sekvenssiksi. Tämän jälkeen sekvenssien määrät normalisoitiin ja virheelliset sekvenssit, eli ”X:si” transloidut sekvenssit poistettiin.

Eri kirjastojen CDRH3 frekvenssien vertailussa hyödynnettiin AbMining-ToolBox -ohjelmistopakettin cdr3_cluster -skriptiä (D’Angelo ja muut 2014). Skripti vertaili valittujen kirjastojen ainutlaatuisten CDRH3 -sekvenssien lukumääriä jokaisessa kirjastossa ja loi uuden tiedoston, jossa ainutlaatuisten CDRH3:n lukumäärät eri kirjastoissa oli esitetty rinnakkain. Kirjastoista haettiin sekvenssejä, joilla oli suuri rikastuskerroin, eli suuri frekvenssin suhde nollanäytteen saman sekvenssin frekvenssiin verrattuna, sekä hyvä spesifisyys. Mikrokystiini kirjastoista etsittiin CDRH3-sekvenssejä, joiden esiintyvyys oli suuri mikrokystiini-LR-kirjastoissa, mutta pieni nollakirjastoissa sekä mikrokystiini-RR-kirjastoissa. Skatoli- ja testosteronikirjastoista etsittiin CDRH3 sekvenssejä, joiden esiintyvyys oli suuri skatoli- ja testosteronikirjastoissa, mutta pieni nollakirjastoissa.

Mikrokystiini-LR-kirjastojen uniikkien CDRH3-sekvenssien pituusjakaumia analysoimalla selvitettiin kirjastojen eripituisten CDRH3:n diversiteetti ja sen muutokset faaginäyttöselektiossa käytetyn antigeenikonsentraation suhteen. Tähän käytettiin AbMining-ToolBox -ohjelmistopakettin cdr3_length_dist -skriptiä (D'Angelo ja muut 2014).

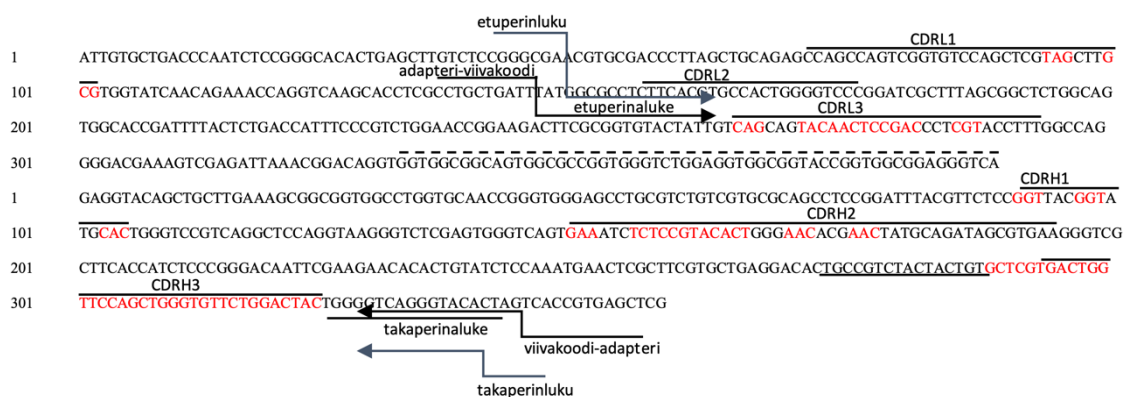
Kirjastoista poistettiin vain kerran esiintyvät CDRH3-sekvenssit, minkä jälkeen jokaisen kirjaston eripituiset CDRH3-sekvenssit jaettiin omiin tiedostoihin niiden frekvenssin mukaiseen suuruusjärjestykseen ja klusteroitiin CD-HIT (<http://weizhongli-lab.org/cd-hit/>) internet serverillä (Li ja muut 2001, Li ja muut 2002). Klusterit visualisoitiin WebLogo 3 (<http://weblogo.threeplusone.com/>) internet serverillä (Crooks ja muut 2004, Schneider ja muut 1990). NGS-datankäsittelyn työnkulku on esitetty kuvassa 11.



Kuva 11. NGS-sekvensointidatan analysoinnin työnkulku. Oikealla on kuvattu analyysin eri vaiheet. Vasemmalla on esitetty mitä ohjelmaa missäkin analyysin vaiheessa on käytetty.

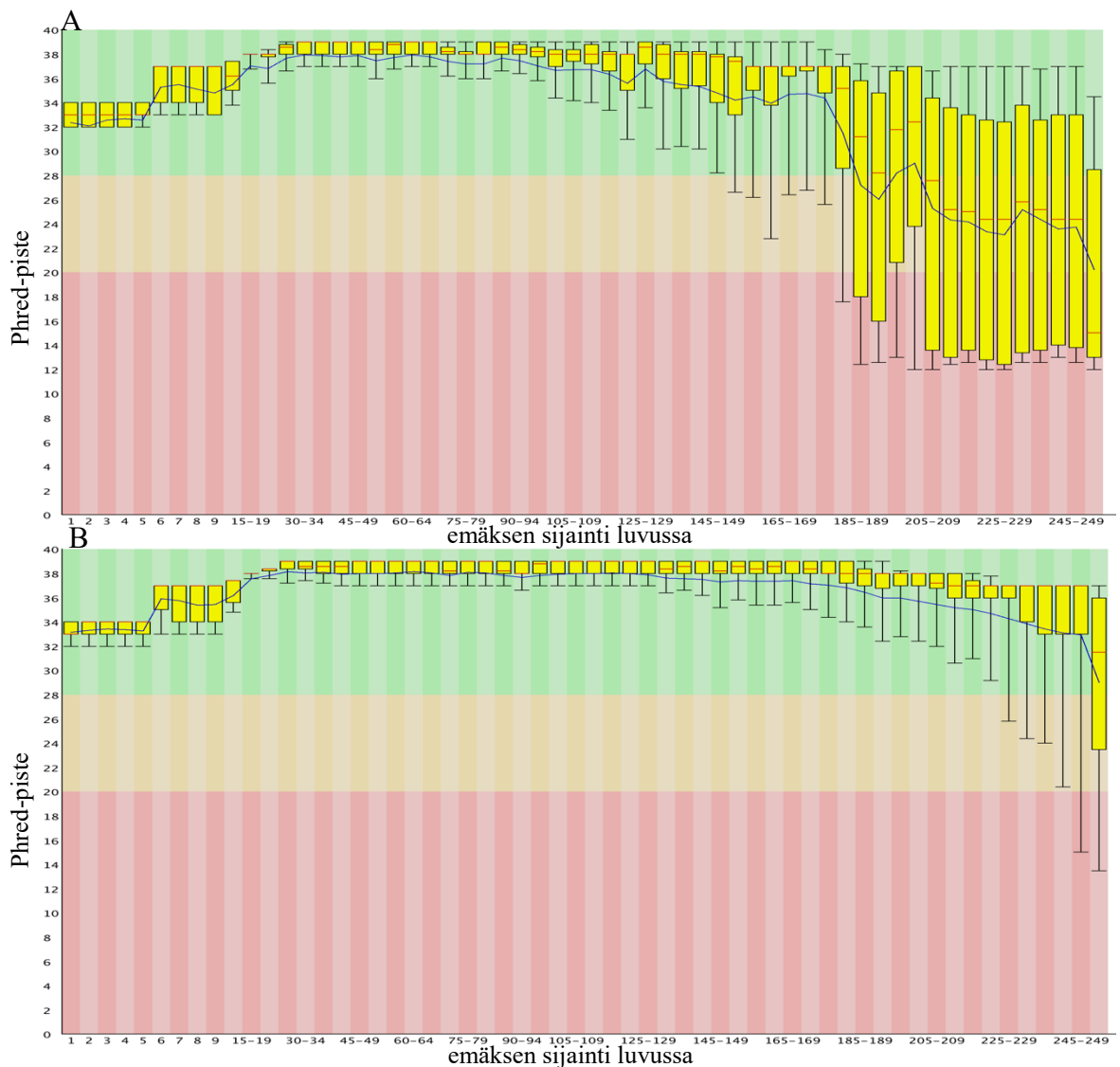
Tulokset ja niiden tarkastelu

Tässä työssä hyödynnetty sekvensointimenetelmä oli suunniteltu käytettäväksi erityisesti synteettisten scFv-faagivasta-ainekirjastojen scFv-M:n ja scFv-P:n kanssa (Brockmann ja muut 2011, Huovinen ja muut 2013). ScFv-fragmentti koostuu kevyt- ja raskasketjujen variaabeleista osista, jotka ovat noin 330 bp:n mittaisia ja siten liian pitkiä sekvensoitavaksi kokonaan Illumina Miseq V2 2 x 250 bp:n sekvensoinnilla. Käyttämällä hyväksi scFv-M ja scFv-P kirjastojen kahta erityisominaisuutta, oli kuitenkin mahdollista sekvensoida antigeenispesifisyyden kannalta keskeisimmät alueet. Ensinnäkin näissä kirjastoissa sekvenssien monimuotoisuus on rajattu yksinomaan viiden CDR:n alueelle ja sekvenssit CDR:ien välissä ovat vakioita. Monimuotoiset alueet ovat kevytketjussa CDRL1:ssä ja CDRL3:ssä, sekä raskasketjussa CDRH1:ssä, CDRH2:ssä ja CDRH3:ssä. Etäisyys CDRL1:n ensimmäisestä variaabelista emäksestä CDRL3:n viimeiseen variaabeliin emäkseen on 195 bp:a ja vastaava etäisyys CDRH1:stä CDRH3:een on 213-252 bp:a riippuen CDRH3 sekvenssin pituudesta. Toisekseen scFv-M- ja scFv-P-kirjastot perustuvat yhden lukukehyksen rakenteeseen. Tämä mahdollistaa yksittäisen, mahdollisimman lähelle ulommaisista CDR-silmukoita sitoutuvan PCR-alueparin käytön, jonka avulla voidaan tuottaa sekvensointiviivakoodit ja adapterit sisältävä amplikonikirjasto.



Kuva 12. Sekvensointikirjastojen valmistamiseen sekä sekvensointiin käytettyjen alukkeiden sijainnit scFv kirjastoissa. Monimuotoiset kodonit on merkitty punaisella, CDR-alueiden sijainnit on merkitty tekstin yläpuolella olevalla viivalla, kevyt- ja raskasketjun yhdistävä alue on merkitty yläpuolella olevalla katkoviivalla ja CDR-alueiden etsimiseen käytetyt sekvenssit on alleviivattu.

Tässä työssä sekvensoitiin scFv:n CDRL3- ja CDRH3-silmukoiden välinen alue, mikä vastaa suurimmasta osasta vasta-aineen sitoutumisen kannalta merkittävistä sekvenssivariaatioista (kuva 12). Sekvensointi aloitettiin raskasketjusta, jossa alue sitoutui CDRH3-alueen yläjuoksun puolelle ja sen jälkeen kevyetketju luettiin päinvastaiseen suuntaan alukkeeseen sitoutuessa CDRL3 alueen alajuoksun puolelle. NGS-ajosta saatujen sekvenssien kokonaismäärä oli sekä etu- että takaperinluvussa 11122751. Etuperinluvun keskimääräinen laatu (Phred) oli 37 ja takaperinluvun 33. Kummassakin esiintyi runsaasti identtisiä sekvenssejä, mikä ei ollut yllättävää käytettyjen kirjastojen osittain korkean rikastumisasteen takia. Etu- ja takaperinluvun laatu on esitetty kuvassa 13.



Kuva 13. Sekvensoinnin laatu. Yksittäisten emästen laatu etuperinluvussa (A) ja takaperinluvussa (B) niiden sijainnin mukaan.

NGS-datankäsittely

Illumina Miseq sekvensoinnista ulos saatu raakadata analysoitiin aluksi scFvMiner-ohjelmistopakettilla, jolla yhdistettiin etuperin- ja takaperinluvut. Sekvenssien yhdistämisessä ei voitu hyödyntää päällekkäin meneviä sekvenssejä lukujen lyhyiden takia, mutta koska sekvenssit olivat tiedostoissa samassa järjestyksessä, ohjelma yhdisti sekvenssit toisiinsa rivi kerrallaan. Lopputuloksena saatiin yksi fastq tiedosto, joka sisälsi kaikki sekvenssit. Seuraavassa vaiheessa sekvenssit eroteltiin FASTX-barcode-splitter -ohjelmaa käyttäen viivakoodien perusteella eri kirjastoihin ja sijoitettiin kirjastoa vastaavaan fastq tiedostoon. Tässä työssä analysoitiin vain sekvenssien CDRH3 alueet, sillä se vastaa suurimmasta osasta vasta-aineen sitoutumiseen vaikuttavasta monimuotoisuudesta. Merkkijonohakualgoritmin avulla sekvensseistä haettiin CDRH3-alueet, jotka käännettiin aa-sekvensseiksi. Samalla CDRH3-sekvenssien frekvenssit laskettiin jokaisesta kirjastosta. Faaginäyttötekniikassa jokaisen selektiokierroksen aikana kloonit, jotka eivät sitoudu kohdemolekyyliin tarpeeksi suurella affiniteetilla karsiutuvat, jonka jälkeen jäljelle jääneet kloonit rikastutetaan. Mitä useampi selektiokierros suoritetaan, sitä kapeammaksi rikastettujen faagien sekvenssien kirjo pienenee. Frekvenssien tarkastelu osoitti skatoli- ja testosteronikirjastojen, jotka oli eristetty faaginäyttöselektion ensimmäisen kierroksen jälkeen, olevan selvästi monimuotoisempia kuin pidemmälle rikastetut mikrokystiinikirjastot, jotka oli eristetty neljännen faaginäyttöselektiokierroksen jälkeen. Vähemmän rikastetuissa skatoli- ja testosteronikirjastoissa ainutlaatuisten sekvenssien määrä oli huomattavasti suurempi kuin mikrokystiinikirjastoissa. Tämän seurauksena skatoli- ja testosteronikirjastoista oli sekvensointidatan perusteella vaikea erottaa muita selkeästi voimakkaammin rikastuneita klooneja. Kirjastojen valmisteluvaiheessa TES5000 kirjaston monistaminen PCR:ssä ei onnistunut, joten sen osalta ei saatu tuloksia. McLR kirjastoissa sekvensseistä jopa 27 % ja McRR kirjastoissa 31 % olivat identtisiä. SKA-kirjastoissa 1 % sekvensseistä oli identtisiä, vastaava luku TES-kirjastoissa oli 2 %. Yhteenveto sekvensoinnista on esitetty taulukoissa 2, 3, 4, 5 ja 6.

Taulukko 2. Skatoli-kirjastojen NGS-sekvensointitulosten yhteenveto.

Kirjasto	SKA0	%	SKA50	%	SKA500	%	SKA5000	%
Sekvenssien kokonaismäärä	379216		309950		288572		254094	
Ainutlaatuisten sekvenssien määrä	230799	61	190557	61	178142	62	164922	65
Singletonit ^a	189368	82	156541	82	146606	82	138598	84
Toistuvat sekvenssit ^b	41431	18	34016	18	31536	18	26324	16
Suurin frekvenssi	2823	1	2341	1	2129	1	1789	1
Rikkinäiset sekvenssit ^c	206	0	142	0	126	0	133	0
Vialliset sekvenssit ^d	2256	1	1837	1	1633	1	1661	1

Taulukko 3. Testosteroni-kirjastojen NGS-sekvensointitulosten yhteenveto.

Kirjasto	TES0	%	TES50	%	TES500	%
Sekvenssien kokonaismäärä	594559		482534		354051	
Ainutlaatuisten sekvenssien määrä	63793	11	54938	11	46678	13
Singletonit ^a	35141	55	31580	57	27575	59
Toistuvat sekvenssit ^b	28652	45	23358	43	19103	41
Suurin frekvenssi	11862	2	9688	2	7095	2
Rikkinäiset sekvenssit ^c	104	0	93	0	91	0
Vialliset sekvenssit ^d	1111	0	644	0	713	0

Taulukko 4. Mikrokystiini-LR-kirjastojen NGS-sekvensointitulosten yhteenveto.

Kirjasto	McLR01	%	McLR1	%	McLR30	%
Sekvenssien kokonaismäärä	452632		416913		404300	
Ainutlaatuisten sekvenssien määrä	2687	1	2612	1	2536	1
Singletonit ^a	1486	55	1456	56	1368	54
Toistuvat sekvenssit ^b	1201	45	1156	44	1168	46
Suurin frekvenssi	124189	27	99090	24	74453	18
Rikkinäiset sekvenssit ^c	84	0	85	0	71	0
Vialliset sekvenssit ^d	411	0	313	0	397	0

Taulukko 5. Mikrokystiini-RR- kirjastojen NGS-sekvensointitulosten yhteenveto.

Kirjasto	McRR01	%	McRR1	%	McRR30	%
Sekvenssien kokonaismäärä	324575		345544		350251	
Ainutlaatuisten sekvenssien määrä	2314	1	2346	1	2313	1
Singletonit ^a	1308	57	1320	56	1259	54
Toistuvat sekvenssit ^b	1006	43	1026	44	1054	46
Suurin frekvenssi	100452	31	104868	30	90529	26
Rikkinäiset sekvenssit ^c	44	0	55	0	67	0
Vialliset sekvenssit ^d	385	0	373	0	392	0

Taulukko 6. Muiden kirjastojen NGS-sekvensointi tulosten yhteenveto

Kirjasto	Mc0	%	Mc00	%	Mc3rd	%	Mc0st	%
Sekvenssien kokonaismäärä	286625		388108		295287		378533	
Ainutlaatuisten sekvenssien määrä	1959	1	1468	0	3470	1	4020	1
Singletonit ^a	1061	54	832	57	1947	56	2258	56
Toistuvat sekvenssit ^b	898	46	636	43	1523	44	1762	44
Suurin frekvenssi	95060	33	199671	51	100566	34	126382	33
Rikkinäiset sekvenssit ^c	43	0	53	0	50	0	85	0
Vialliset sekvenssit ^d	330	0	268	0	533	0	745	0

^a: Singletonit: vain kerran esiintyvien ainutlaatuisten sekvenssien määrä

^b: Toistuvat sekvenssit: useammin kuin kerran esiintyvien ainutlaatuisten sekvenssien määrä

^c: Rikkinäiset sekvenssit: sekvenssien määrä, joiden CDRH3 tunnistuksessa vain etuliite löytynyt

^d: Vialliset sekvenssit: virheellisesti transloitujen sekvenssien määrä

Ensimmäinen ja yksinkertainen vaihe hyvien vasta-ainekandidaattien tunnistamiseksi oli eri kirjastojen sekvenssien frekvenssien vertailu toisiinsa ja niistä spesifisimmin kohdemolekyylillä sitoneiden ja parhaiten rikastuneiden CDRH3-sekvenssien seulominen. Näytekirjastoista haettiin CDRH3-sekvenssejä, joiden frekvenssin suhde saman sekvenssin frekvenssiin nollanäytekirjastossa, eli rikastumiskerroin, oli mahdollisimman suuri. Lisäksi mikrokystiini kirjastoista etsittiin sekvenssejä, joilla oli korkea spesifisyys mikrokystiini-LR-IC:a kohtaan eli sekvenssejä, joiden frekvenssi oli suuri McLR-kirjastoissa, mutta pieni McRR-kirjastoissa. Skatolikirjastoissa CDRH3-sekvenssien rikastumiskerroin (SKA5000/SKA0) oli suurimmillaankin alle 4 ja kaikkien sekvenssien, joiden rikastuskerroin oli yli 2 frekvenssi jäi alle 50:een. Alhaisten frekvenssien ja rikastuskertoimien takia skatolikirjastoja ei analysoitu enää tätä vaihetta pidemmälle. Skatolikirjastojen suurimman rikastuskertoimen omaavat sekvenssit on lueteltu taulukossa 7. Testosteronikirjastoissa TES500-kirjaston CDRH3-sekvenssien suurin rikastumiskerroin oli yli 10, mutta frekvenssi oli suhteellisen alhainen. Testosteronikirjastojen suurimman rikastuskertoimen omaavat sekvenssit on lueteltu taulukossa 8. Alhaisten frekvenssien sekä TES5000 tulosten puuttumisen takia testosteronikirjastoja ei myöskään analysoitu pidemmälle.

Taulukko 7. Skatolikirjastojen suurimpien rikastuskertoimien CDRH3-sekvenssit ja niiden frekvenssit. SKA5000/SKA0 < 3 poistettu.

sekvenssi	SKA0 frekv.	SKA5000 frekv.	SKA5000/SKA0	pituus
TSRNY	0	42		5
ARPWPRFGY	8	30	3,7	9
ARVPDWVDY	10	36	3,6	9
AGAQFAY	10	34	3,4	7
AKRTWSDRVFSY	14	45	3,2	12

Taulukko 8. Testosteronikirjastojen suurimpien rikastuskertoimien CDRH3-sekvenssit ja niiden frekvenssit. TES500/TES0 < 2 sekvenssit poistettu.

sekvenssi	TES0 frekv.	TES500 frekv.	TES500/TES0	pituus
AREGWYSVFDY	24	257	10,7	11
AKGTLFYS	63	262	4,2	8
ARSWHFDD	48	124	2,6	8
ARWWDDV	658	1458	2,2	7
ARRWYYGYAFDY	89	191	2,2	12

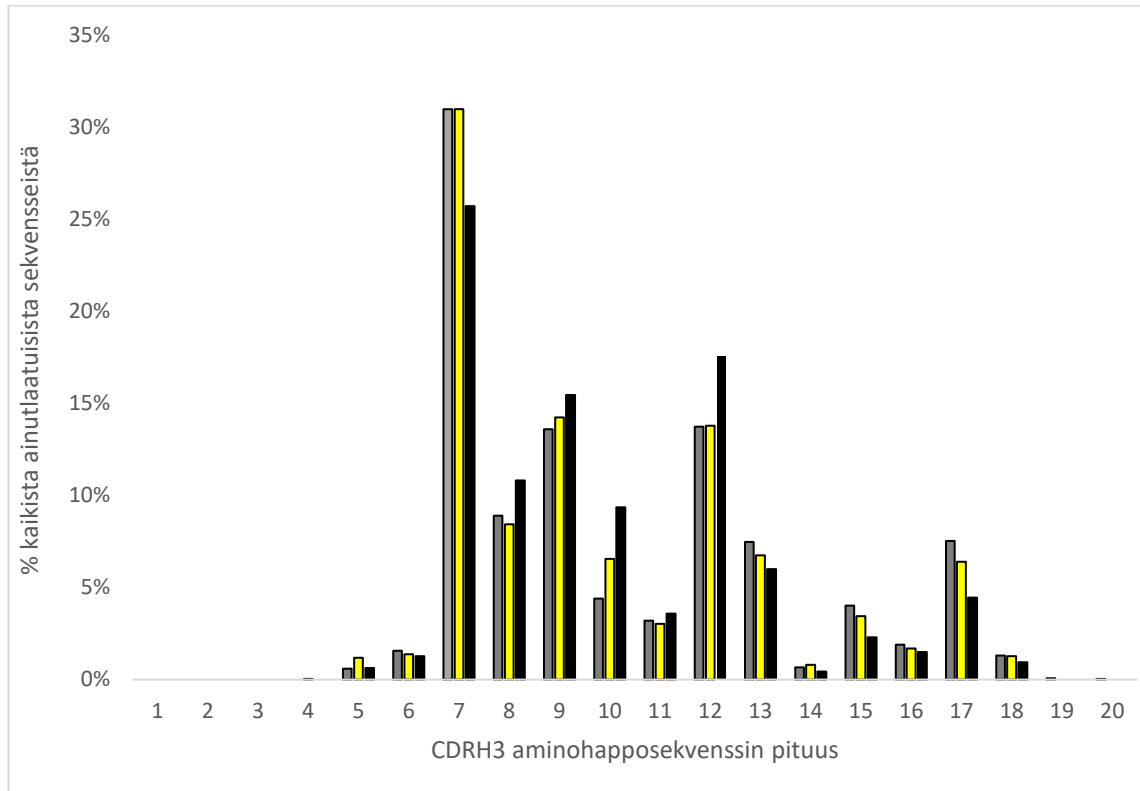
Mikrokystiinikirjastossa McLR30 CDRH3-sekvenssin rikastuskertoimen oli parhaimmillaan yli 55, mikä viittaa siihen, että kyseinen vasta-aine sitoo hyvin immunokompleksia verrattuna vapaaseen vasta-aineseen. McLR30 ja McRR30 kirjastojen CDRH3-sekvenssien rikastumiskertoimen oli parhaimmillaan yli 9, mikä viittaa kyseisen vasta-aineen tunnistavan mikrokystiini-LR:n muodostaman IC:n huomattavasti suuremmalla todennäköisyydellä kuin mikrokystiini-RR:n muodostaman IC:n (taulukko 9). Taulukon sekvensseistä RVDGRYDY oli sisäinen kontrolli, jonka EC50-arvo (engl. effective concentration 50 %, EC50), eli konsentraatio, jolla saavutetaan 50 % maksimaalisesta antigeenin sitoutumisesta, on noin 5 nM ja se tunnistaa mikrokystiini-LR:ää noin nelinkertaisella affiniteetilla verrattuna mikrokystiini-RR:ään. Suurimman rikastuskertoimen omaavan ARARWYPSYFDY-sekvenssin spesifisyys mikrokystiini-LR:ää kohtaan oli tulosten perusteella lähes identtinen, mutta rikastumiskertoimen oli noin 1,5-kertainen, jonka perusteella sen voidaan olettaa omaavan potentiaalisesti paremmat sitomisominaisuudet. Kolme suurimman rikastuskertoimen ja spesifisyyden omaavaa CDRH3-sekvenssiä, olivat kaikki 12 aa:n pituisia.

Taulukko 9. Mikrokystiini-LR kirjastojen suurimpien rikastuskertoimien CDRH3-sekvenssit, niiden frekvenssit, pituus sekä McLR:n ja McRR:n frekvenssien suhde. Alle 100 frekvenssin, $McLR/Mc0 < 2$ ja $McLR/McRR < 2$ sekvenssit poistettu.

sekvenssi	McLR30 frekv.	McRR30 frekv.	Mc0 frekv.	McLR30/ Mc0	McLR30/ McRR30	pituus (aa)
ARARWYPSYFDY	15723	1683	285	55,2	9,3	12
ARDSLYFAFDY	6359	785	257	24,8	8,1	12
AREGDLWFAFDY	582	97	16	37,5	6,0	12
ARDYEDASFDY	258	46	20	13,1	5,6	11
ARAGVWFFDY	6964	2505	568	12,3	2,8	10
APQGVYMDY	14810	4533	386	38,3	3,3	9
RVDGRYDY	4846	517	230	21,1	9,4	8

McLR-kirjastojen ainutlaatuisten CDRH3-sekvenssien pituusjakaumia analysoimalla selvitettiin miten faaginäyttöselektioissa käytetyt eri analyttikonsentraatiot vaikuttivat kirjastojen monimuotoisuuteen. McLR-kirjastojen ainutlaatuisten CDRH3-sekvenssien pituusjakauma oli suhteellisen samankaltainen eri kirjastojen välillä. Faaginäyttöselektion antigeenikonsentraation kasvaessa monimuotoisuus kasvoi eniten 10 ja 12 aa:n mittaisissa CDRH3-sekvensseissä, mikä viittasi suurempaan todennäköisyyteen löytää hyviä sitoja tämän mittaisista sekvensseistä. Kahdentoista aa:n

mittaisten CDRH3-sekvenssien monimuotoisuus oli myös toiseksi suurin 7 aa:n mittaisten sekvenssien jälkeen, joiden monimuotoisuus sen sijaan laski antigeenikonsentraation kasvaessa (Kuva 14).

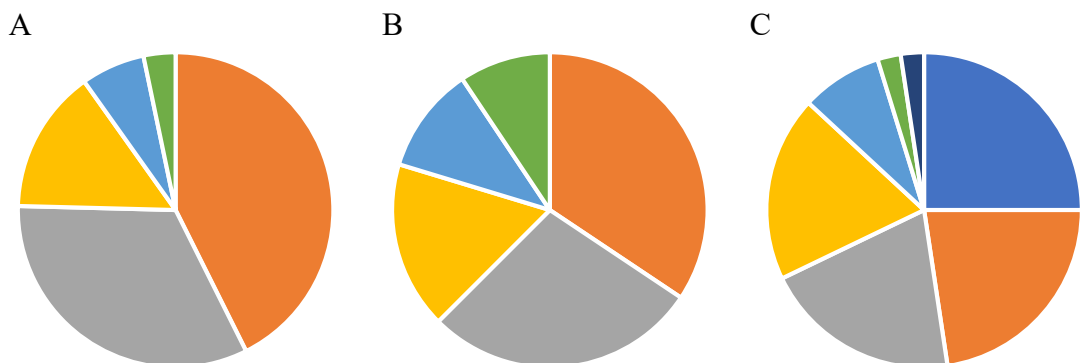


Kuva 14. Mikrokystiini-LR kirjastojen ainutlaatuisten CDRH3-aminohapposekvenssien pituusjakauma. McLR01 (■), McLR1 (■), McLR30 (■).

Kun eri kirjastojen CDRH3-sekvenssien frekvenssejä vertailtiin suoraan toisiinsa, vain suurimmat frekvenssit otettiin huomioon. Jotta voitiin analysoida laajempi kirjo potentiaalisia vasta-ainekandidaatteja, ja ottaa huomioon myös harvinaisempia sekvenssejä sekä analysoida faaginäyttöselektiossa käytetyn analyttikonsentraation vaikutusta sekvenssien koostumukseen, suoritettiin CDRH3-sekvensseille klusterianalyysi sitoutumiseen vaikuttavien CDRH3-sekvenssin rakenteiden tunnistamiseksi. Tähän käytettiin CD-HIT-ohjelmaa, jonka avulla klusterointi voitiin suorittaa nopeasti ja helposti. (Li ja muut 2001, Li ja muut 2002.) CD-HIT:n suoritustehon parantamiseksi ja ohjelman mahdollisesti aiheuttamien virheiden välttämiseksi CDRH3-sekvenssit järjestettiin niiden pituuden mukaan omiin tiedostoihin. Tiedostoista poistettiin myös rikastumattomat sekvenssit, eli vain kertaalleen esiintyneet sekvenssit. Lisäksi ennen tiedostojen syöttämistä ohjelmaan sekvenssit järjestettiin tiedostojen sisällä

frekvenssien mukaiseen suuruusjärjestykseen suurimmasta pienimpään riveittäin. Koska CD-HIT:ä ei ole suunniteltu erityisen lyhyiden peptidien analysoimiseen, lisättiin kaikkiin alle 10 aa:n mittaisiin sekvensseihin keinotekoisia aa:ja. Klusteroinnissa ohjelma sijoitti samaan klusteriin kaikki saman tiedoston sekvenssit, joiden ero ”pääsekvenssiin” eli klusterin suurimman frekvenssin omaavaan sekvenssiin oli yhden aa:n verran. Klusteroinnin jälkeen klusterit visualisoitiin WebLogo3:lla tulosten tulkitsemisen helpottamiseksi (Crooks ja muut 2004, Schneider ja muut 1990).

Eri kirjastojen pituusjakaumien analysoinnin perusteella oletettiin, että suurin todennäköisyys löytää potentiaalisia sekvenssejä olisi McLR-kirjastojen 12 aa:n mittaisista CDRH3-sekvensseistä. Kun McLR01-, McLR1- ja McLR30-kirjastojen 12 aa:n mittaisten sekvenssien klustereita vertailtiin toisiinsa, havaittiin että kirjastoissa McLR01 ja McLR1 oli 5 klusteria ja McLR30 oli 7 klusteria. McLR01-kirjastossa yksittäisten klusterien osuus koko kirjastosta oli myös suurempi kuin McLR1-kirjastossa, ja vastaavasti McLR1 kirjastossa yksittäisten klusterien osuus koko kirjastosta oli suurempi kuin McLR30-kirjastossa. Tämä tukee CDRH3-sekvenssien pituusjakaumien analysoinnin tuloksia ja viittaa siihen, että pienemmillä antigeenikonsentraatioilla faaginäyttöselektiossa harvemmat sekvenssit alkavat dominoida ja sekvenssien monimuotoisuus pienenee. McLR-kirjastojen 12 aa:n CDRH3-sekvenssien klustereiden vertailu on esitetty kuvassa 15.



Kuva 15. Kahdentoista aa:n mittaisten CDRH3-sekvenssien klusterit McLR01-kirjastossa (A) McLR1-kirjastossa (B) ja McLR30-kirjastossa (C). Sektorin koko esittää klusterien suhteellista kokoa kirjastossa. Eri kirjastojen saman värisillä klustereilla on sama pääsekvenssi: ARARWYPSYFDY (■), AREYYWTVVFDY (■), ARERYYPLVFDY (■), ARDSDLYFAFDY (■), AREYEWGVVFDY (■), AREGDLWFAFDY (■) ja ARARLYPSYFDY (■).

Koska kolme McLR spesifisintä sekvenssiä löytyivät McLR30-kirjaston 12 aa:n mittaisista CDRH3-sekvensseistä, analysoitiin ne kolme klusteria erikseen, joista nämä sekvenssit löytyivät. Nämä kolme sekvenssiä olivat myös kaikki oman klusterinsa pääsekvenssejä. CD-HIT:llä ja Weblogo3:lla analysoitiin näiden kolmen klusterin aarakennetta ja selvitettiin mitkä pääsekvenssin aa:sta olivat konservoituneita ja mitkä vaihtelevia. Kolme McLR spesifisimmän CDRH3-sekvenssin omaavaa klusteria on esitetty taulukossa 10. Kaikissa kolmessa klusterissa pääsekvenssin osuus sekvensseistä oli suurempi kuin kaikkien muiden sekvenssien määrä yhteensä.

Taulukko 10. Mikrokystiini-LR spesifisimpien CDRH3-sekvenssien klusterit McLR30-kirjastossa.

klusteri	pääsekvenssi	pituus	frekvenssi	jäseniä
	ARARWYPSYFDY	12	15723	21
	ARDSDLYFAFDY	12	6359	16
	AREGDLWFAFDY	12	582	2

Päätelmät

Tässä työssä onnistuttiin NGS-sekvensointia sekä sekvenssianalyysiä hyödyntämällä löytämään McLR spesifejä CDRH3-sekvenssejä, joita ei ole aikaisemmin Turun yliopistossa pystytty löytämään perinteisiä menetelmiä käyttäen. ”Parhaan” löydetyn sekvenssin spesifisyys McLR:ää kohtaan oli myös parempi kuin minkään aiemmin Turun yliopistossa löydetyn sekvenssin. Tämän työn aikana ehdittiin analysoida vain sekvenssien CDRH3-alueet. Jatkossa potentiaalisista sekvensseistä tulisi analysoida myös muut CDR-alueet ja tämän jälkeen testata vasta-aineiden toimivuus todellisuudessa. Löydettyjen CDRH3-alueiden todellisen vaikutuksen määrittämiseksi spesifisyyteen tulisi scFv-geenit kloonata vektoreihin, ekspressoida bakteereissa ja tämän jälkeen testata tuottuneiden vasta-aineiden spesifisyys esimerkiksi ELISA-määrittäyksillä. Klusterianalyysillä selvitettyjen sekvenssirakenteiden perusteella olisi mahdollista luoda uusi, kohdennettu kirjasto, jossa CDRH3-alueissa monimuotoisuudet olisi kohdennettu konservoituneiden alueiden väliin. Tämä mahdollistaisi potentiaalisesti entistä spesifisempien, vielä tuntemattomien, sekvenssien löytämisen.

Työssä saadut tulokset kannustavat jatkamaan NGS-sekvensoinnin sekä sekvenssianalyysin menetelmien hyödyntämistä ja kehittämistä vasta-ainekehityksessä jatkossa. NGS-sekvensoinnilla on valtava potentiaali, jonka avulla vasta-ainekehitystä on mahdollista nopeuttaa dramaattisesti ja löytää sekvenssejä, joita ei muuten ole mahdollista havaita. NGS-sekvensoinnin käytön suurimpana pullonkaulana vasta-aineseulonnessa lienee optimaalisten työkalujen puute. Suurin osa ohjelmistoista on suunniteltu käytettäväksi genomien tutkimiseen, eivätkä ne aina sovellu suoraan sekvenssianalyysin suorittamiseen vasta-aineiden löytämiseksi. Tässä työssä käytettyjä sekvenssianalyysin työkaluja tulisi jatkossa kehittää helppokäyttöisimmiksi ja mahdollisesti yhdistää yhden ohjelmiston alle. Tämä mahdollistaisi NGS-datan nopeamman käsittelyn ja helpottaisi NGS-teknologioiden käyttöön ottoa vasta-ainekehityksessä myös jatkossa.

Lähteet

- Akter, S., Vehniäinen, M., Meriluoto, J., Spoof, L., & Lamminmäki, U. (2017) Non-competitive ELISA with broad specificity for microcystins and nodularins. *Advances in Oceanography and Limnology*. 8(1).
- Andrews, S. (2010) FastQC: a quality control tool for high throughput sequencing data. < <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> >. (Luettu 8.1.2019).
- Almaiman, A. (2018) Effect of testosterone boosters on body functions: Case report. *Int J Health Sci (Qassim)* 12(2):86-90.
- Arola, H., Tullila, A., Kiljunen, H., Campbell, K., Siitari, H. & Nevalainen, T. (2016) Specific noncompetitive immunoassay for HT-2 mycotoxin detection. *Anal Chem* 88:2446-2452.
- Bazan, J., Całkosiński, I. & Gamian, A. (2012) Phage display—A powerful technique for immunotherapy. *Hum Vaccin Immunother.* 8(12):1817-1828
- Bekaert, K. (2012) Chemical and sensory detection of boar taint. Gentin yliopisto. Eläinlääketieteen tiedekunta. Väitöskirja.
- Bonneau, M. (1998) Use of entire males for pig meat in the European Union. *Meat Sci* 49(1):257-272
- Brockmann, E., Akter, S., Savukoski, T., Huovinen, T., Lehmusvuori, A., Leivo, J., Saavalainen, O., Azhayev, A., Lövgren, T., Hellman, J. & Lamminmäki, U. (2011) Synthetic single-framework antibody library integrated with rapid affinity maturation by VL shuffling. *Protein Eng Des Sel* 24(9):691-700.
- Carlomagno, M., Lassabe, G., Rossotti, M., Gonzales-Techera, A., Vanrell, I., & Gonzalez-Sapienza, G. (2014) Recombinant streptavidin nanopeptamer anti-immunocomplex assay for noncompetitive detection of small analytes. *Anal Chem* 86(20):10467-10473.
- Carnegie, C. (2004) Diagnosis of hypogonadism: Clinical assessments and laboratory tests. *Rev Urol* 6(Suppl 6):3-8.
- Crooks, G., Hon, G., Chanjonia J. & Brenner S. (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188-1190
- D'Angelo, S., Glanville, J., Ferrera, F., Naranjo, L., Gleasner, C., Shen, X., Bradbury, A. & Kiss, C (2014) The antibody mining toolbox. An open source tool for the rapid analysis of antibody repertoires. *Mabs* 6(1):160–172. < <https://sourceforge.net/projects/abmining/> >. (Luettu 9.1.2019).

Fredriksen, B., Font I Furnols, M., Lundstrom, K., Migdal, W., Prunier, A., Tuytens F. & Bonneau, M. (2009) Practice on castration of piglets in Europe. *Animal* **3(11)**:1480-1487.

Gaget, V., Humpage, A., Huang, Q., Monis, P. & Brookes, J. (2017) Benthic cyanobacteria: A source of cylindrospermopsin and microcystin in Australian drinking water reservoirs. *Water Res* **124**: 454-464.

Glenville, J., Angelo, S., Khan, T., Reddy, S., Naranjo, L., Ferrera, F. & Brandbury, A. (2015) Deep sequencing in library selection projects: what insight does it bring? *Curr Opin Struc Biol* **33**:146-160.

González-Techera, A., Vanrell, L., Last, J., Hammock, B. & González-Sapienza G. (2007) Phage anti-immune complex assay (PHAIA): a general strategy for noncompetitive immunodetection of small molecules. *Anal Chem* **79(20)**:7799-7806.

Hannon, G. (2009) FASTX-toolkit: FASTQ/A short-reads pre-processing tools. <http://hannonlab.cshl.edu/fastx_toolkit/index.html>. (Luettu 3.2.2019).

Herati, A., Cengiz, C. & Lamb, D. (2016) Assays of serum testosterone. *Urol Clin North Am* **43(2)**:177-184.

Huovinen, T., Syrjänpää, M., Sanmark, H., Brockmann, E., Azhayev, A., Wang, O., Vehniäinen, M. & Lamminmäki, U. (2013) Two ScFv antibody libraries derived from identical VL-VH framework with different binding site designs display distinct binding profiles. *Protein Eng Des Sel* **26(10)**:683-93.

Kaplan, G. & Gershoni, J. (2012) A general insert label for peptide display on chimeric filamentous bacteriophages. *Anal Biochem* **420**:68-72.

Kaushik, R. & Balasubramanian, R. (2012) Methods and approaches used for detection of cyanotoxins in environmental samples: A review. *Crit Rev Environ Sci Technol*

Kicman, A. (2008) Pharmacology of anabolic steroids. *Br J Pharmacol* **154(3)**: 502-521

Kumar, P., Kumar, N., Thakur, D. & Patidar, A. (2010) Male hypogonadism: Symptoms and treatment. *J Adv Pharm Technol Res* **1(3)**:297-301.

Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J. & Laustsen, A. (2018) Basics of antibody phage display technology. *Toxins (Basel)* **10(6)**:236

Leivo, J., Mäkelä, J., Rosenberg, J. & Urpo, L. (2016) Development of recombinant antibody-based enzyme-linked immunosorbent assay (ELISA) for the detection of skatole. *Anal Biochem* **492**:27-29.

Lövgren, J., Pursinheimo, J., Pyykkö, M., Salmi, J. & Lamminmäki, U. (2016) Next generation sequencing of all variable loops of synthetic single framework scFv—Application in anti-HDL antibody selections. *N Biotechnol* **33**:790-796. <<https://sourceforge.net/projects/scfvminer/>>. (Luettu 9.1.2019).

- Li, Y., Zhang, G., Mao, X., Yang, S., Ruyck, K. & Wu, Y. (2018) High sensitivity immunoassays for small molecule compounds detection - Novel noncompetitive immunoassay designs. *Trends Anal Chem* **103**:198-208
- Li, W., Jaroszewski, L. & Godzik, A. (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17(3)**:282-283.
- Li, W., Jaroszewski, L. & Godzik, A. (2002) Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* **18(1)**:77-82.
- Liu, J. & Sun, Y. (2015) The role of PP2A-associated proteins and signal pathways in microcystin-LR toxicity. *Toxicol Lett* **236(1)**:1-7.
- Lone, Y., Kumar, R. & Bhide, M. (2015) An overview of the toxic effect of potential human carcinogen Microcystin-LR on testis. *Toxicol Rep* **2**:289-296
- Lövgren, J., Pursiheimo, J., Pyykkö, M., Salmi, J. & Urpo, L. (2016) Next generation sequencing of all variable loops of synthetic single framework scFv—Application in anti-HDL antibody selections. *New Biotech* **33**:790-796.
- Marintcheva, B. (2018) Harnessing the power of viruses, s. 133-160 Academic Press, Cambridge, Massachusetts.
- Matochko, W., Chu, K., Jin, B., Lee, S., Whitesides, G. & Derda, R. (2012) Deep sequencing analysis of phage libraries using Illumina platform. *Methods* **58(1)**:47-55
- Marx, G., Horn, T., Thielebein, J., Knubel, B. & von Borell, E. (2003) Analysis of pain related vocalization in young pigs. *J Sound Vib* **266(3)**:687-698.
- Miller, A. & Russel, C. (2017) Food crops irrigated with cyanobacteria-contaminated water: an emerging public health issue in Canada. *Environ Health Rev* **60(3)**: 58-63
- O'Donnell, L., Stanton, P. & de Kretser (2017) Endocrinology of the Male Reproductive System and Spermatogenesis, s. 1 MDText.com, Inc, South Dartmouth, Massachusetts.
- Omi, K., Ando, T., Sakyu, T., Shirakawa, T., Uchida, Y. Oka, A., Ise, N., Aoyagi, K. & Goishi, K. (2015) Noncompetitive immunoassay detection system for haptens on the basis of antimetatype antibodies. *Clin Chem* **61(4)**:627-635
- Pulli, T., Höyhty, M., Söderlund, H. & Takkinen, K. (2005) One-Step homogeneous immunoassay for small analytes. *Anal Chem* **77**:2637-2642.
- Rami, A., Behdani, M., Yardehnavi, N., Habibi-Anbouhi, M. & Kazemi-Lomedasht, F. (2017) An overview on application of phage display technique in immunological studies. *Asian Pac J Trop Med* **7(7)**:599-602
- Rantala, A., Rajaniemi-Wacklin, P., Lyra, C., Lepistö, L., Rintala, J., Manckiewicz-Boczek, J. & Sivonen, K. (2006) Detection of microcystin-producing cyanobacteria in

Finnish lakes with genus-specific microcystin synthetase gene *e* (*mcyE*) PCR and associations with environmental factors. *Appl Environ Microbiol* **72(9)**:6101-6110

Rezaitabar, S., Sari, A., Bahramifar, N. & Ramezani, Z. (2017) Transfer, tissue distribution and bioaccumulation of microcystin LR in the phytoplanktivorous and carnivorous fish in Anzali wetland, with potential health risks to humans. *Sci Total Environ* **575**:1130-1138

Schmidt, J., Wilhelm, S. & Boyer, G. The fate of microcystins in the environment and challenges for monitoring. *Toxins (Basel)* **6(12)**:3354-3387.

Schneider, T. & Stephens, R. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18(20)**:6097-6100.

Smith, G. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**:1315-1317.

Trost, L. & Mulhall, J. (2016) Challenges in testosterone measurement, data interpretation, and methodological appraisal of interventional trials. *J Sex Med* **13(7)**: 1029-1046.

Turner, A., Dhanji-Rapkova, M., O'Neil, A., Coates, L., Lewis, A. & Lewis, K. (2017) Analysis of microcystins in cyanobacterial blooms from freshwater bodies in England. *Toxins* **10(1)**:39.

Vold, E. (1970) Fleischproduktionseigenschaften bei Ebern und Kastraten. IV. Organoleptische und gaschromatografische Untersuchungen wasserdampf-flüchtiger stoffe des rüchenspeckes von ebern. *Meldinger Nordlandbrukshoegskole*, **49**:1-25.

Walstra, P. (1974). Fattening of young boars: quantification of positive and negative aspects. *Livestock Production Science* **1**:187-196.

Whittington, F.M., Nute, G.R., Hughes, S.I., McGivan, J.D., Lean, I.J., Wood, J.D. & Doran, E. (2004). Relationships between skatole and androstenone accumulation, and cytochrome P4502E1 expression in Meishan X Large White pigs. *Meat Science* **67**:569-576.

Yang, W., Yoon, A., Lee, S., Kim, S., Han, J. & Chung, J. (2017) Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. *Exp Mol Med* **49(3)**:e308.

Zhang, Y., Chen, M., Li, H., Yan, F., Pang, P., Wang, H., Wu, Z. & Yang, W. (2017) A molybdenum disulfide/gold nanorod composite-based electrochemical immunosensor for sensitive and quantitative detection of microcystin-LR in environmental samples. *Sens. Actuator B-Chem* **244**:606-615.