

# **RNA-polymeraasi I:n transkription aloitus yksittäismolekyylitasolla**

Pro gradu-tutkielma

Turun yliopisto

Biokemian laitos

Biokemian tutkinto-ohjelma (FM)

Heinäkuu / 2020

Emil Aalto-Setälä

*Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -järjestelmällä*

TURUN YLIOPISTO

Biokemian laitos

EMIL AALTO-SETÄLÄ: RNA-polymeraasi I:n transkription aloitus  
yksittäismolekyyllitasolla

Pro gradu -tutkielma, 70 s, 3 liites.

Biokemia

Heinäkuu / 2020

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -järjestelmällä

-----

DNA:sta tuotetaan RNA:ta transkriptiossa ja tuotetusta RNA:sta tuotetaan muun muassa proteiineja. RNA:n tuottamisesta ovat vastuussa RNA-polymeraaseiksi kutsutut entsyymit. Eukaryooteilla on kolme RNA-polymeraasia, joista ribosomaalista RNA:ta tuottaa RNA-polymeraasi I. RNA-polymeraasi I:n yksityiskohtainen toimintamekanismi on vielä puutteellisesti tunnettua. Erityisesti RNA-polymeraasi I:n transkription aloituskompleksin muodostumisen dynamiikka on vielä pitkälti selvittämättä. RNA-polymeraasi I vaatii avustavia proteiineja, jotta se voi sitoutua promoottoriinsa ja aloittaa transkription. RRN3:n ja CF:n avulla RNA-polymeraasi I muodostaa minimaalisen transkription aloituskompleksin. TBP ja UAF täydentävät minimaalisen transkription aloituskompleksin ja lisäävät sen transkriptioaktiivisuutta kymmenkertaisesti. Vielä ei tiedetä, missä järjestyksessä nämä proteiinit sitoutuvat promoottorille ja onko sitoutuminen dynaamista vai vakautuuko syntynyt kompleksi promoottorin ympärille.

RNA-polymeraasin I:n aloituskompleksin sitoutumista DNA:han tutkittiin aluksi elektroforeettisella siirtymäkokeella sekä anisotropiakokeella. Näitä tuloksia pidettiin kuitenkin kvalitatiivisina ja erittäin paljon vaikeasti puhdistettavia proteiineja kuluttavana. Nämä kokeet eivät myöskään ole reaaliaikaisia, minkä takia dynaaminen informaatio oli hyvin rajallista. Näin ollen siirryttiin käyttämään yksittäisten molekyylien fluoresenssimikroskopiaan perustuvaa tutkimusmenetelmää, mikä mahdollisti DNA-proteiinikompleksin muodostumisen seuraamisen reaaliajassa. Tämä kuitenkin vaatii kokeissa käytettyjen proteiinien leimaamista ja huomattavan määrän data-analyysimenetelmien kehittämistä.

Tulosten mukaan RRN3 ei yksinään sitoudu DNA:han, CF ja RNA-polymeraasi I yksinään sitoutuvat promoottoriin dynaamisesti kahdella tavalla muodostaen joko lyhyt- tai pitkäikäisen kompleksin DNA:n kanssa. RNA-polymeraasi I:n, RRN3:n ja CF:n yhdessä muodostaman minimaalisen transkription aloituskompleksin sitoutuminen on myös olevan kaksivaiheista. Alustavien tulosten mukaan UAF ja TBP vakauttavat CF:n sitoutumista promoottoriin täydellisessä transkription aloituskompleksissa.

Luotujen menetelmien ja saatujen tulosten avulla on RNA-polymeraasi I:n aloituskompleksin sitoutumisjärjestys promoottorille mahdollista selvittää jatkotutkimuksissa.

Asiasanat: RNA-polymeraasi I, kinetiikka, yksittäismolekyyli fluoresenssi, transkriptio

Haluan kiittää ohjaajaani Anssi Malista työn mahdollistamisesta sekä kaikesta saamastani tuesta. Kiitos myös koko RNAP/RET-laboratorion väelle.

Haluan myös kiittää anniskeluliike Porttia sekä seurusteluravintola Kerttulin Kievaria niistä lukemattomista tunteista, joita vietin siellä tuskaillessani opinnoissani tai tämän työn parissa. Oispa kaljaa.

Sekä ystävät, jotka ovat minua tukeneet: kiitos!

Antti Aalto-Setälän muistolle.

# Sisältö

1. Johdanto .....	5
2. Kirjallisuuskatsaus .....	7
2.1 RNA-polymeraasien rakenne ja avustavat proteiinit .....	7
2.2 Transkriptiosykli .....	9
2.3 RNA-polymeraasi I:n transkription aloitus .....	19
2.3.1 Välttämätön aloituskompleksi .....	19
2.3.2 Täydellinen aloituskompleksi .....	25
2.4 RNA-polymeraasi I:n transkription pidennysvaihe .....	27
2.5 RNA-polymeraasi I:n transkription lopetusvaihe .....	28
2.6 Leivinihiivan RNA-polymeraasi I:n ja ihmisen RNA-polymeraasi I:n vertailu .....	30
2.7 Ihmisen RNA-polymeraasi I lääkekehityksen kohteena .....	32
2.8 Työn tarkoitus .....	33
3. Materiaalit ja menetelmät .....	35
3.1 Proteiinien leimapaikkojen suunnittelu .....	35
3.2 Proteiinien tuotto ja puhdistus E. coli:ssa .....	35
3.2.1 RRN3 ja sen variantit .....	35
3.2.2 CF ja sen variantit .....	36
3.2.3 TBP ja sen variantit .....	37
3.3 Proteiinien tuotto ja puhdistus S. cerevisiae:ssa .....	38
3.3.1 Pol I .....	38
3.3.2 UAF .....	39
3.4 Proteiinien leimaus .....	40
3.5 Sitoutumiskokeet .....	41
3.5.1 EMSA .....	41
3.5.2 Anisotropia .....	41
3.6 Yksittäisten molekyylien fluoresenssimikroskopia .....	42
4. Tulokset .....	44
4.1 RNA-polymeraasi I:n transkription aloituskompleksin proteiinien puhdistaminen .....	44
4.1.1 Proteiinien puhdistaminen .....	44
4.1.2 Proteiinien karakterisointi .....	45
4.2 Proteiinien käytön suunnittelu reaaliaikaisiin sitoutumiskokeisiin .....	48
4.3 Mutatoitujen proteiinien tuotto ja karakterisointi .....	51
4.4 Proteiinien leimaus .....	53
4.5 Mikroskooppiprotokollan luominen .....	54
4.6 Reaaliaikaiset sitoutumiskokeet .....	55
4.7 Mikroskooppidatan analysointivaiheen kehittäminen .....	56
4.8 Käytettyjen fluoroforien elinajat .....	58
4.9 Välttämättömän aloituskompleksin komponenttien sitoutuminen promoottoriin .....	59

5.	Tulosten tarkastelu .....	64
5.1	Puhdistetun Pol I –transkriptiokoneiston valmistaminen biofysikaalisia tutkimuksia varten	64
5.2	Superresoluutiomikroskopian haasteet ja niiden ylitse pääseminen .....	65
5.3	Pol I välttämättömän aloituskompleksin osatekijöiden sitoutuminen promoottoriin on dynaamista .....	67
5.4	Luotujen menetelmien avulla Pol I aloituskompleksin sitoutumisjärjestyksen selvittäminen on mahdollista .....	69
6.	Yhteenveto .....	70
7.	Lähteet.....	71
	Liite 1 – Plasmidikartat.....	86
	Liite 2 – Kirjoitetut Excel-makrot.....	87
	Makro 1: useiden tekstitiedostojen tuonti yhteen Excel-työkirjaan eri välilehdille.....	87
	Makro 2: välilehtien sisältämän numeerisen informaation koonti yhdelle välilehdelle, jossa ensimmäiseen sarakkeeseen luodaan kuvanumerot .....	88

## Lyhenteet

22-pep = 22 aminohappoa pitkä polypeptidi, jolla inhiboidaan Pol I transkriptiota

RP<sub>c</sub> = Suljettu kompleksi (engl. closed complex), jossa RNA-polymeraasi on sitoutuneena DNA:han, muttei vielä avannut DNA:ta

CF = (engl. Core Factor) Pol I:n välttämätön aloitustekijä, joka koostuu proteiinialayksiköistä RRN6, RRN7 ja RRN11.

DSIF = DRB molekyylin herkkyyttä kasvattava tekijä (engl. DRB sensitivity inducing factor), joka edistää Pol II:n pysähtymistä promoottorin läheisyyteen.

ELL = Pol II:n yleinen transkription pidennysvaiheen transkriptiotekijä

EMSA = elektroforeettisen siirtymän muutoskoe (engl. electrophoretic mobility shift assay), jossa tutkitaan leimatun komponentin liikkumista natiivilla (ei-denaturoivalla) geelillä.

hRRN3 = ihmisen RRN3 (engl. human RRN3)

Kryo-EM = kryoelektronimikroskopia (engl. cryo electron microscopy)

LB = bakteerien kasvatukseen käytetty elatusaine (engl. lysogey broth)

NELF = negatiivinen pidennystekijä (engl. negative elongation factor), joka edistää Pol II:n pysähtymistä promoottorin läheisyyteen

Nsi1 = ei-transloitavaa sekvenssiä hiljentävä proteiini (engl. non-translated sequence (NTS) silencing protein), Pol I:n transkription lopetustekijä

NTP = nukleosiditriposfaatti (engl. nucleoside triphosphate)

RP<sub>o</sub> = Avoin kompleksi (engl. open complex), jossa RNA-polymeraasi on sitoutunut ja avannut DNA:n

OD = Optinen tiheys (engl. optical density)

pAzF = p-Azido-L-fenyyialaniini (engl. p-Azido-L-phenylalanine), luonnon aminohappo

PEG = Polyetyleeniglykoli

Pol I = RNA-polymeraasi I

Pol II = RNA-polymeraasi II

Pol III = RNA-polymeraasi III

PPi = Pyrofosfaatti

P-TEFb = transkription pidennysvaihetta edistävä tekijä b (engl. positive transcription-elongation factor b), joka fosforyloi DSIF ja NELF proteiineja ja edistää Pol II:n transkription etenemisessä

PTRF = Pol I transkription vapauttava tekijä (engl. Pol I transcript release factor), nisäkkäiden Pol I:n transkription lopetustekijä

rDNA = ribosomaalinen DNA

rRNA = ribosomaalinen RNA

SL1 = valikoiva tekijä 1 (engl. selective factor 1), ihmisen RNA-polymeraasi I:n aloitusta avustava proteiini

TAF = TBP:hen liittyvä tekijä (engl. TBP associated factor)

TBP = TATA-alueeseen sitoutuva proteiini (engl. TATA binding protein), joka on osa RNA-polymeraasien I, II ja III initiaatiokomplekseja.

TFIIA = Transkriptiotekijä II A (engl. transcription initiation factor II A), Pol II:n aloitustekijä

TFIIB = Transkriptiotekijä II B, Pol II:n aloitustekijä

TFIID = Transkriptiotekijä II D, Pol II:n aloitustekijä

TFIIE = Transkriptiotekijä II E, Pol II:n aloitustekijä.

TFIIF = Transkriptiotekijä II F, Pol II:n aloitustekijä.

TFIIH = Transkriptiotekijä II H, Pol II:n aloitustekijä.

TFIIIA = Transkriptiotekijä III A, Pol III:n aloitustekijä

TFIIIB = Transkriptiotekijä III B, Pol III:n aloitustekijä

TFIIIC = Transkriptiotekijä III C, Pol III:n aloitustekijä

TIRF = totaalisen sisäisen heijastuman fluoresenssi (engl. Total internal reflection fluorescence); tutkimismenetelmä, jossa mikroskooppilasin pintaan sitoutuneet fluorofoorin sitoutuvat kompleksit näkyvät, mutta vapaat kompleksit eivät.

TTF-I = transkription lopetustekijä I (engl. transcription termination factor I), nisäkkäiden Pol I:n transkription lopetustekijä

tWH = tandemiivekäheliksi (engl. tandem winged helix), proteiinidomeeni, joka muistuttaa heliksi-kääntö-heliksi proteiinidomeenia.

UAF = ylävirran aktivoiva tekijä-proteiini (engl. Upstream activating factor), joka koostuu viidestä alayksiköstä ja aktivoi Pol I:n välttämätöntä aloituskompleksia.

UBF = ylävirtaan sitoutuva tekijä (engl. upstream binding factor), ihmisen Pol I:n aloitustekijä



## 1. Johdanto

Elävien organismien perimä koostuu DNA:sta. DNA:n ohjeen mukaan valmistetaan sen RNA:ta, joka siirtää DNA:n tiedon ribosomeille, jossa tästä RNA:sta luodaan proteiineja. RNA-polymeraasit ovat entsyymejä, jotka tuottavat DNA:sta RNA:ta ja välittävät tämän tiedon ribosomeille. Ribosomeille siirtyvää RNA:ta kutsutaan lähetti-RNA:ksi. Osaa RNA-polymeraasien luomasta RNA:sta ei kuitenkaan kuljeteta ribosomeille, vaan niillä on solun säätelyyn liittyviä tehtäviä. Tällaisia ovat esimerkiksi lyhyet häiritsevät RNA:t (engl. *short interfering RNA*) tai ribosomaalinen RNA (rRNA), joka toimii osana ribosomeja. RNA-polymeraasin toimintaa, eli RNA:n tuottamista DNA:n geneettisen koodin mukaan, kutsutaan transkriptioksi. Transkriptiossa RNA-polymeraasi pidentää RNA:ta yksi emäs kerrallaan RNA-polymeraasin sisälle kehittyvän DNA/RNA-hybridin avulla ja syntyvää yksinauhaista RNA:ta työnnetään ulos RNA-polymeraasissa sijaitsevassa RNA:n ulosmenokanavassa. Kun RNA-ketju on pidentynyt ja transkriptio lopetettu, lähetti-RNA siirtyy ribosomeille ja siitä tuotetaan proteiineja translaatioksi kutsutussa prosessissa.

RNA-polymeraaseja on jokaisessa elävässä organismissa, mutta niiden koostumus vaihtelee jonkin verran. Arkeilla ja bakteereilla on vain yksi RNA-polymeraasi, joka on vastuussa kaikesta RNA:n tuotosta. Eukaryooteilla on sen sijaan kolme eri tehtäviin erikoistunutta RNA-polymeraasia: RNA-polymeraasi I (Pol I) tuottaa suurimman osan rRNA:sta, RNA-polymeraasi II (Pol II) tuottaa lähetti-RNA:ta ja RNA-polymeraasi III (Pol III) tuottaa pääosin siirtäjä-RNA:ta. Eukaryoottien RNA-polymeraasit ovat monista alayksiköistä koostuvia proteiinikomplekseja, ja ne jakavat keskenään osan alayksiköistään. Kasveilla on muista eukaryooteista erottuen viisi RNA-polymeraasia, edellä mainittujen kolmen RNA-polymeraasin lisäksi RNA-polymeraasit IV ja V. Ne tuottavat lyhyitä häiritseviä RNA:ita. (Drouin ja Carter, 2010; Steitz, 1998; Werner ja Grohmann, 2011; Woychik, 2005; Woychik ja Reinberg, 2001) RNA-polymeraasit ovat kehittyneet divergenttisellä evoluutiolla, eli niillä on kaikilla yksi yhteinen esimuoto, joka todennäköisesti koostui nykyisten bakteerien RNA-polymeraasien tapaan viidestä alayksiköstä (Drouin ja Carter, 2010).

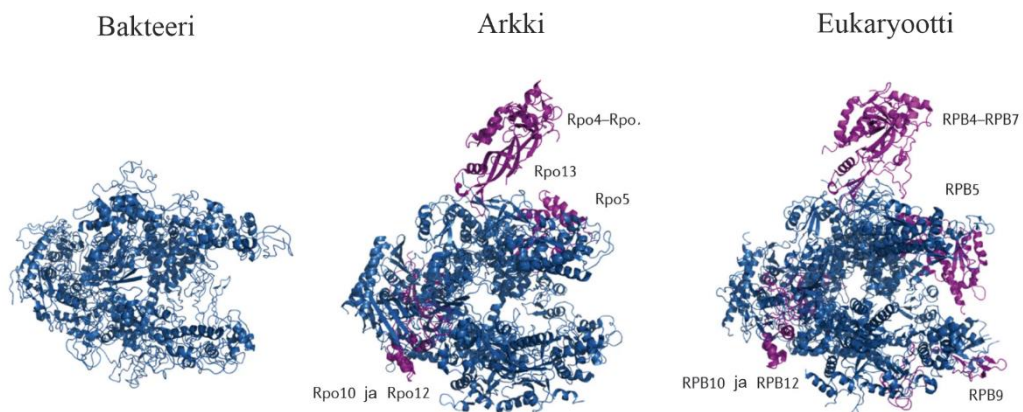
Kaikki eukaryoottien RNA-polymeraasit muistuttavat rakenteellisesti toisiaan ja niillä on samoja toimintoja eri eliölajeissa. Vaikka RNA-polymeraasien yleisrakenne on samankaltainen, niissä on myös toisistaan eroavia alayksiköitä ja ne käyttävät erilaisia

avustavia proteiineja. Nämä erot mahdollistavat RNA-polymeraasien toiminnallisen erilaistumisen. Eukaryoottien RNA-polymeraaseista kaksi on ollut laajimman tutkimuksen alla; ihmisen (*Homo sapiens*) sekä leivinihiivan (*Saccharomyces cerevisiae*). Molempien organismien eniten ja pisimpään tutkittu RNA-polymeraasi on Pol II. Ihmisen Pol II toiminta on monimuotoisempaa ja monimutkaisempaa kuin leivinihiivan, mutta myöskin ihmisen genomi on paljon suurempi ja monimuotoisempi verrattuna leivinihiivan genomiin. (Botstein ja muut, 1997; Cramer ja muut, 2008) Jotta leivinihiivan ja ihmisen RNA-polymeraasien toimintaa ja rakennetta voidaan ymmärtää paremmin, on syytä tarkastella ensin yksinkertaisempaa RNA-polymeraasimallia. Bakteerien RNA-polymeraasi on paljon tutkittu ja Pol II:n toiminta muistuttaa sen toimintaa, jolloin sen toiminnan tarkastelu yhdessä Pol II kanssa on mielekästä. Puhuttaessa Pol II:sta viitataan sillä leivinihiivan Pol II kompleksiin, ellei toisin mainita.

## 2. Kirjallisuuskatsaus

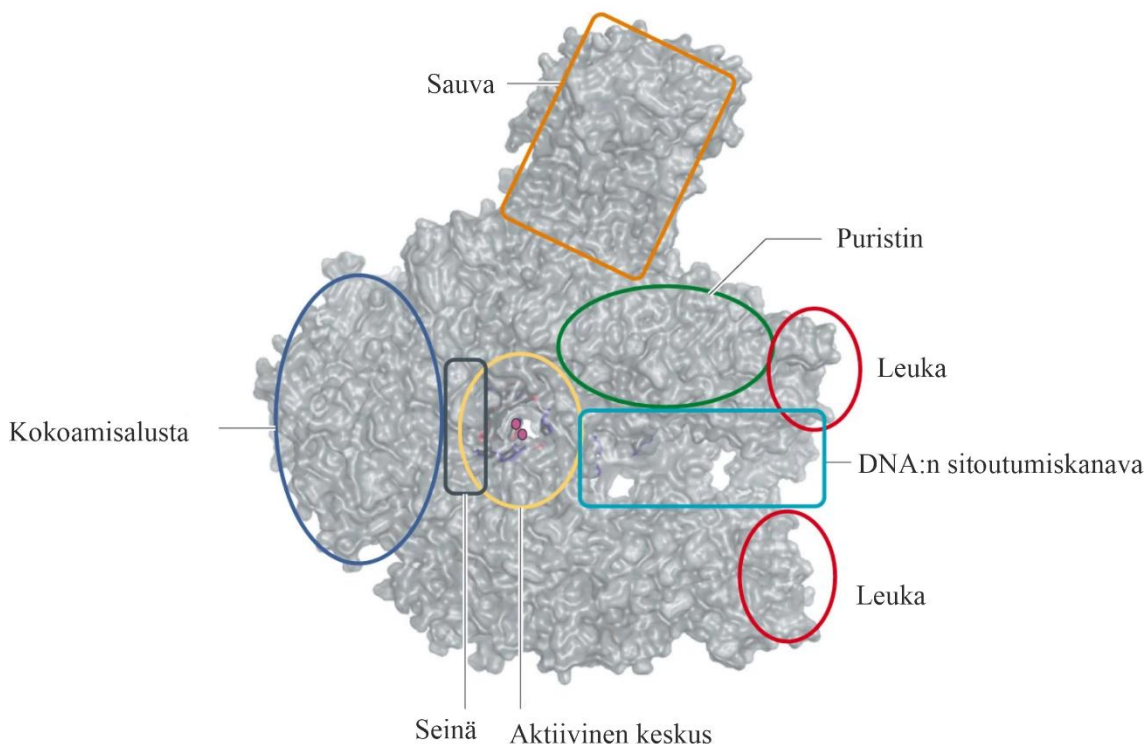
### 2.1 RNA-polymeraasien rakenne ja avustavat proteiinit

Monialayksikköisten RNA-polymeraasien rakenne on hyvin samankaltainen. Yksinkertaisin RNA-polymeraasi löytyy bakteereilta, joilla on vain viisialayksikköinen RNA-polymeraasi. Sen alayksiköt,  $\beta$ ,  $\beta'$ ,  $\alpha$  sekä  $\omega$  muodostavat toiminnallisen RNA-polymeraasin, jossa  $\alpha$ -alayksikköä on kaksi kappaletta. Eukaryooteilla, joilla on kolme RNA-polymeraasia, eniten bakteeria muistuttava RNA-polymeraasi on Pol II. Pol II koostuu 12 alayksiköstä ja toimii niin sanottuna yleispolymeraasina, joka tuottaa lähetti-RNA:ta. Vaikka Pol II sisältää seitsemän alayksikköä enemmän, on se silti hyvin saman näköinen kuin bakteerien RNA-polymeraasi (kuva 1). Oikeastaan kaikki RNA-polymeraasit muistuttavat toisiaan, niiden eroavaisuudet johtuvat erilaisista avustavista proteiineista, joita eukaryoottiset RNA-polymeraasit vaativat toimiakseen.



**Kuva 1.** RNA-polymeraasien vertailua eri organismien välillä. Vasemmalla on bakteerien RNA-polymeraasi, keskellä arkkien RNA-polymeraasi ja oikealla Eukaryoottinen RNA-polymeraasi II. RNA-polymeraasien yleisarkkitehtuuri on hyvin samanlainen ja suurimmat erot johtuvat monimutkaisempien organismien RNA-polymeraasien suuremmasta alayksikkömäärästä. Kuva muokattu lähteestä Werner ja Grohmann (2011).

RNA-polymeraasien yleisarkkitehtuuri muistuttaa ravun saksia ja koostuu seitsemästä domeenista (kuva 2). Nämä domeenit sijaitsevat eri alayksiköissä RNA-polymeraasista riippuen. Ulommainen domeeni on kaksijakoinen leuka (engl. *jaw*), joka sitoo alavirran DNA:ta. Leukojen välissä sijaitsee DNA:ta sitova halkio (engl. *DNA binding cleft*), joka nimensä mukaisesti sitoo DNA:ta. DNA kulkee halkion kautta aktiiviseen keskukseen (engl. *catalytic center*), jossa itse transkriptio tapahtuu. Katalyyttisen keskuksen ja leuan välissä, halkion yläpuolella, on puristin (engl. *clamp*). Sauva (engl. *stalk*), joka sijaitsee



**Kuva 2.** RNA-polymeraasien yleisrakenne. Kyseessä eukaryoottien RNA-polymeraasi II. Leuat, DNA:n sitoutumiskanava, koura, sauva sekä aktiivinen keskus eritelty kuvassa. Myös kokoamisalusta ja seinä eritelty kuvassa, muttei tekstissä. Kuva muokattu lähteestä Werner ja Grohmann (2011).

puristimen yläpuolella, ohjaa pidentyvää RNA:ta ulos RNA-polymeraasista. (Cramer ja muut, 2001; Vannini ja Cramer, 2012; Werner ja Grohmann, 2011)

Kaikki RNA-polymeraasit myös toimivat samankaltaisesti. Niiden toiminnassa on kolme vaihetta: transkription aloitus (engl. *transcription initiation*), transkription pidennysvaihe (engl. *transcription elongation*) ja viimeisenä transkription lopetusvaihe (engl. *transcription termination*). Transkription aloitusvaiheessa avustavat proteiinit sitoutuvat DNA:han tai itse polymeraasiin ja näin sitovat polymeraasin promoottori. Aloitusvaiheen avustavia proteiineja on bakteerien polymeraaseillakin useita, ja ne tunnetaan  $\sigma$ -tekijöinä. Pol I koneistossa toimii neljä avustavaa proteiinikompleksia: ylävirran aktivoiva proteiini (engl. *upstream activating factor*, UAF), TATA-alueeseen sitoutuva proteiini (engl. *TATA-binding protein*, TBP), ydintekijäproteiini (engl. *core factor*, CF) sekä RRN3 (Vannini ja Cramer, 2012). Pol II taas tarvitsee vähintään transkriptiotekijä II B -proteiinin (engl. *transcription initiation factor II B*, TFIIB), TFIIE:n, TFIIF:n, TFIIH:n sekä TBP:n ja TBP:hen sitoutuvat avustavat proteiinit- TFIIA ja TFIID (Roeder, 1996). Pol III tarvitsee transkription aloitukseen proteiinit TFIIA, TFIIB ja TFIIC (Abascal-

Palacios ja muut, 2018; Kassavetis ja muut, 1999; Kassavetis ja muut, 2001). Eroavaisuudet RNA-polymeraasien transkription aloituksen avustavien proteiinien määrissä on osin selitettävissä sillä, että Pol I ja Pol III sisältävät enemmän alayksiköitä rakenteessaan kuin Pol II: Pol I sisältää 14 alayksikköä ja Pol III 17 alayksikköä (taulukko 1). Transkription aloitukseen nämä kaikki avustavat proteiinit eivät ole välttämättömiä, sillä esimerkiksi Pol II pystyy toimimaan ilman TBP:hen sitoutuvia avustavia proteiineja (TFIIA ja TFIID) (Roeder, 1996). Pol I puolestaan pystyy toimimaan ilman UAF ja TBP-proteiineja, mutta Pol I aktiivisuus on silloin vain 2 – 10 % verrattuna täydelliseen kompleksiin (Keener ja muut, 1998). Tämä viittaisi siihen, että solu pystyy rajoittamaan RNA-polymeraasien toimintaa rajoittamalla aktivoivien proteiinien tuottoa.

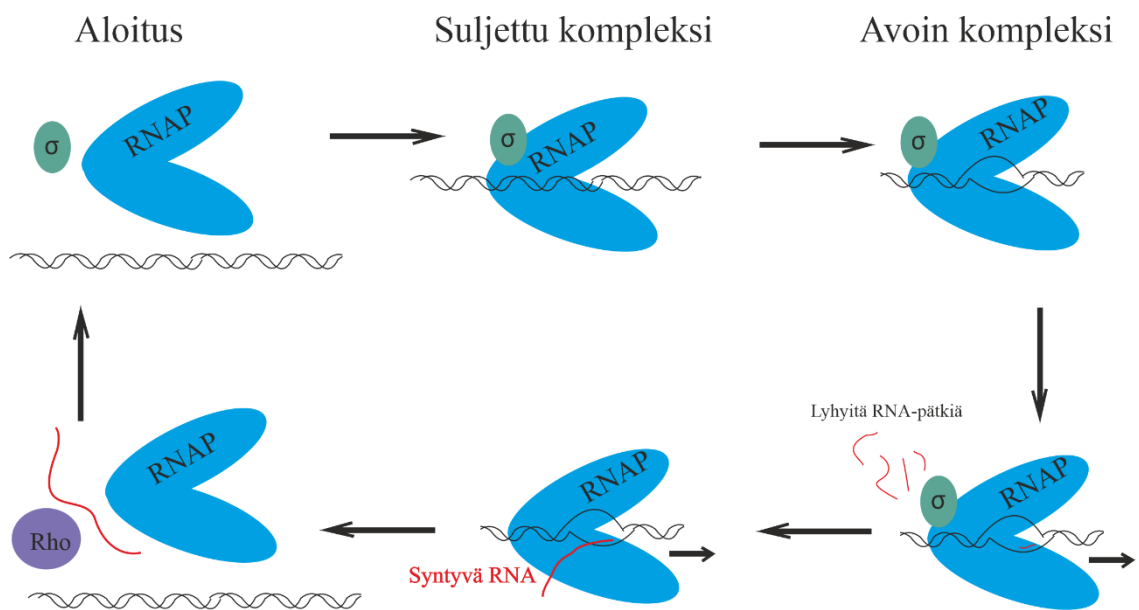
**Taulukko 1.** *Leiviniivan RNA-polymeraasien alayksiköt. Toisiaan vastaavat alayksiköt ovat vierekkäin. Jokainen RNA-polymeraasi koostuu kymmenen alayksikön ytimestä. Pol I ja Pol III sisältävät Pol II:n RPB7/RPB4 alakompleksin kaltaisen kompleksin. Pol I ja Pol III sisältävä myös TFIIIF kaltaisen alakompleksin, jota ei ole Pol II:ssa, mutta TFIIIF on Pol II:n transkription aloitusta avustava proteiini. Tämän lisäksi Pol III:lla on vielä oma, kolmesta alayksiköstä koostuva alakompleksi.*

<b><i>Sachharomyces Cerevisiae</i> RNA-polymeraasien alayksiköt</b>			
	<b>Pol I</b>	<b>Pol II</b>	<b>Pol III</b>
Kymmenen alayksikön ydin	A190	RPB1	C160
	A135	RPB2	C128
	AC40	RPB3	AC40
	AC19	RPB11	AC19
	RPB5	RPB5	RPB5
	RPB6	RPB6	RPB6
	RPB8	RPB8	RPB8
	RPB10	RPB10	RPB10
	RPB12	RPB12	RPB12
	A12	RPB9	C11
RPB7/RPB4 alakompleksi	A43	RPB7	C25
	A14	RPB4	C17
TFIIIF kaltainen alakompleksi	A49	-	C37
	A34.5	-	C53
Pol III alakompleksi	-	-	C82
	-	-	C34
	-	-	C31

## 2.2 Transkriptiosykli

Transkriptiosyklin mekanistinen kuvailu keskittyy bakteerien RNA-polymeraasin tutkimuksiin, mutta eukaryoottien RNA-polymeraaseissa transkription aloitusvaihe alkaa avustavien proteiinien sitoutumisesta promoottori-DNA:han sekä itse RNA-polymeraasiin. Tämän jälkeen avustavat proteiinit vuorovaikuttavat keskenään, ohjaten

RNA-polymeraasin promoottorille. Tätä kompleksia, jossa RNA-polymeraasi avustavine proteiineineen on sitoutunut promoottoriin, kutsutaan suljetuksi kompleksiksi (engl. *closed complex*, RP<sub>c</sub>) (kuva 3). Bakteerit muodostavat saman kompleksin, mutta ilman erillisiä DNA:han sitoutuvia avustavia proteiineja. Tämän jälkeen RNA-polymeraasi yhdessä avustavien proteiinien kanssa sulattaa DNA:ta, jolloin sen kaksi juostetta irtoavat toisistaan transkriptiokuplan alueelta. Toinen juosteista ohjataan RNA-polymeraasin sisään aktiiviseen keskukseen. Tätä juostetta kutsutaan mallijuosteeksi (engl. *template strand*). Mallijuosteen emäsjärjestyksen avulla RNA-polymeraasi tuottaa RNA:ta. Kun DNA:n mallijuoste on RNA-polymeraasin sisällä, toinen DNA-juosteista kulkee RNA-



### Transkription lopetus      Transkription pidennysvaihe      Keskeytyvä aloitus

**Kuva 3.** Transkriptiosyklin vaiheet bakteerien RNA-polymeraasilla havainnoituna. Transkriptio alkaa, kun RNA-polymeraasi (RNAP-kuvassa) ohjataan promoottorille, jolloin se muodostaa avustavien proteiinien, bakteerien tapauksessa  $\sigma$ -tekijän, kanssa suljetun kompleksin. RNA-polymeraasi sulattaa promoottorin, avaten sen ja muodostaen avoimen kompleksin. Avoimen kompleksin muodostuttua RNA-polymeraasi käy keskeytyvän aloituksen läpi niin moneen kertaan, kunnes se pääsee etenemään pidennysvaiheeseen. Pidennysvaiheessa RNA-polymeraasi tuottaa mallijuosteen mukaista RNA:ta ja etenee DNA:ta pitkin. Lopetusvaiheessa RNA-polymeraasi irtoaa DNA:sta ja täysimittainen RNA irtoaa RNA-polymeraasista. Bakteereilla eräs lopetusvaiheen proteiini on Rho, joka esiteltä kuvassa. Tämän jälkeen RNA-polymeraasi kykenee aloittamaan uuden transkriptiosyklin.

polymeraasin pinnalla. Tässä vaiheessa olevaa RNA-polymeraasin ja DNA:n muodostamaa kompleksia kutsutaan avoimeksi kompleksiksi (engl. *open complex*, RP<sub>o</sub>), joka on käytännössä valmis aloittamaan transkription (kuva 3). Avoimen kompleksin muodostumisen jälkeen siirrytään transkription aloitusvaiheeseen, jolloin RNA-polymeraasi on alustavasti transkriptoiva kompleksi (engl. *initially transcribing complex*,

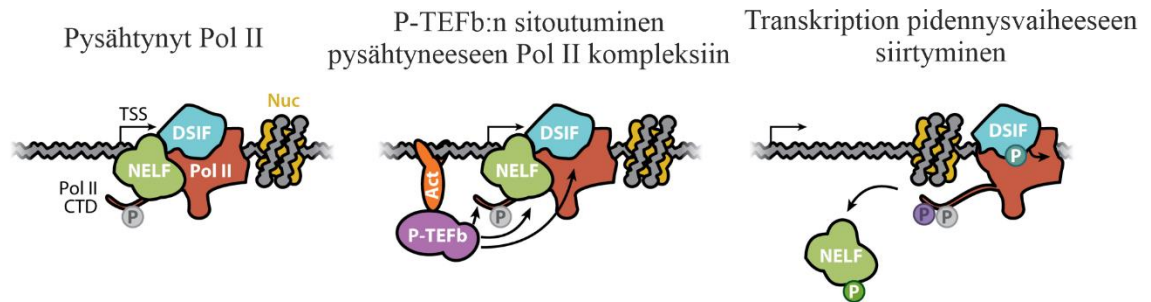
RP<sub>ITC</sub>). Tässä vaiheessa RNA-polymeraasilla on kaksi kilpailevaa reaktiota ennen siirtymistä pidennysvaiheeseen: keskeytyvä aloitus (engl. *abortive initiation*) ja promoottorilta karkaaminen (engl. *promoter escape*). Keskeytyvässä aloituksessa RNA-polymeraasi tuottaa lyhyitä, 2-12 nukleotidiä pitkiä RNA-pätkiä, jotka se sitten peruuttamattomasti vapauttaa ennen uuden RNA-molekyylin synteesin aloittamista. Promoottorilta karkaaminen vaatii RNA-polymeraasin irtautumisen promoottorin ydinalueesta. Kun RNA-polymeraasi onnistuu karkaamaan promoottorilta, RP<sub>ITC</sub> muuttuu transkription pidennyskompleksiksi, joka lopulta tuottaa täysimittaisen RNA:n. (Hsu, 2002) Transkription aloitusvaiheessa, ennen promoottorilta karkaamista, DNA:han sitoutunut RNA-polymeraasi pysyy paikallaan (Li ja McClure, 1998). RNA-polymeraasi vetää kuitenkin RNA:n syntetisoimisen yhteydessä alavirran DNA:ta sisäänsä mekanismeissa, jota kutsutaan rytistämiseksi (engl. *scrunching*). Rytistäminen purkautuu, kun RNA-polymeraasi vapauttaa RNA:n keskeytyvässä aloituksessa tai kun RNA-polymeraasi karkaa promoottorilta transkription pidennysvaiheeseen. (Kapanidis ja muut, 2006; Revyakin ja muut, 2006) Vaikka RNA-polymeraasi aloittaisi transkription tehokkaasti, siirtyen suoraan promoottorilta karkaamiseen, on transkription aloituksessa havaittavissa sama rytistäminen (Revyakin ja muut, 2006). Transkription pidennysvaiheessa rytistämistä ei ole havaittu (Abbondanzieri ja muut, 2005) siitä huolimatta, että RNA-polymeraasi saattaa pysähtyä kymmeniksi sekunneiksi (Artsimovitch ja Landick, 2000). Keskeytettyä aloitusta tapahtuu sekä *in vitro* että soluissa. Keskeytetyn aloituksen todennäköisyyteen vaikuttaa se, kuinka hyvin promoottori vakauttaa aloituskompleksia. Mikäli promoottorin ja RNA-polymeraasin vuorovaikutukset ovat vahvoja, tapahtuu keskeytettyä aloitusta enemmän. (Goldman ja muut, 2009) Promoottorin ja RNA-polymeraasin väliseen vuorovaikutuksen vahvuuteen vaikuttaa promoottorin sekvenssi: esimerkiksi bakteerien RNA-polymeraasilla vuorovaikutuksen vahvuuteen vaikuttaa kohdan -10 ja transkription aloituskohdan (+1) välinen erottaja-alue (engl. *discriminator region*)  $\sigma^{70}$  riippuvaisilla promoottoreilla (Ruff ja muut, 2015), mutta myös -10 ja -35 alueet sekä niiden välinen erottaja (engl. *spacer*) promoottorilla vaikuttavat vuorovaikutuksen vahvuuteen (Heyduk ja Heyduk, 2018). Vuorovaikutuksen voimakkuudesta riippumatta promoottorilta karkaamiseen on esitetty hypoteesi: rytistämisestä johtuva stressi, jolloin RNA-polymeraasi vetää alavirran DNA:ta sisäänsä kolmessa vaiheessa, nostaen stressiä polymeraasin sisällä, mikä epävakauttaa vuorovaikutukset promoottorin kanssa ja mahdollistaa RNA-polymeraasin karkaamisen promoottorilta. (Henderson ja muut, 2019).

Eukaryoottien keskeytetty aloitus ja promoottorilta karkaaminen ovat monimutkaisempia tapahtumia, sillä aloituskompleksit sisältävät DNA:han itsenäisesti sitoutuvia, RNA-polymeraasin sitoutumista avustavia proteiineja. Pol II:n tapauksessa keskeytetty aloitus tapahtuu lyhyemmällä, 4-5 nukleotidiä pitkällä RNA-pätkillä, minkä jälkeen Pol II R<sub>P</sub>ITC on paljon vakaampi. Promoottorilta karkaaminen tapahtuu, kun Pol II on tuottanut yli 5 nukleotidiä pitkän RNA:n, eli R<sub>P</sub>ITC muuttuu RNA:n tuottokompleksiksi (engl. *elongation complex*). (Dvir, 2002) Pol II:n täytyy kuitenkin sitoutua promoottorilta karkamiseen (engl. *escape commitment*), ennen kuin se voi muodostaa EC:n. Sitoutuminen promoottorilta karkamiseen vaatii vähintään 4 nukleotidiä pitkän RNA:n ja Pol II:n on siirryttävä transkription aloituspaikasta kohti geenin alavirtaa. Tällöin suurin osa Pol II molekyyleistä karkaa promoottorilta transkription pidennysvaiheeseen. (Kugel ja Goodrich, 2002)

Transkription pidennysvaihe alkaa, kun RNA-polymeraasi on onnistunut karkamaan promoottorilta. Bakteerien RNA-polymeraasit sekä eukaryoottien Pol I ja Pol III alkavat tuottaa tehokkaasti RNA:ta, mutta Pol II ei vielä tähän kykene, vaan sen täytyy päästä vielä yhden esteen yli. Tätä estettä kutsutaan promoottorin läheisyyteen pysähtymiseksi (engl. *promoter proximal arrest* tai *promoter proximal pausing*). (Core ja Adelman, 2019) Promoottorin läheisyyteen pysähtymistä havaitaan useimmiten Pol II:lla, mutta esimerkiksi kolibakteerin (*Escherichia coli*) RNA-polymeraasi pysähtyy  $\sigma^{38}$  avulla promoottorin läheisyyteen (Petushkov ja muut, 2017). Kolibakteerilla on havaittu myös promoottorin läheisyyteen pysähtymistä  $\sigma^{70}$ -sisälvissä RNA-polymeraasikomplekseissa, jolloin  $\sigma^{70}$ -aloitustekijä on kiinni RNA-polymeraasissa sen aloitettua transkription ja RNA-polymeraasi pysähtyy kohtaan +17 lacUV5-promoottorilla (Brodolin ja muut, 2004). Pol II:lle promoottorin läheisyyteen pysähtyminen on tärkeä toiminnan säätelyvaihe, jossa solu päättää Pol II transkription pidennysvaiheeseen etenemisestä ja edelleen geenin ilmenemisen aktivoimisesta. Pysähtyminen toimii myös yhtenä tarkastuspisteenä transkription laadunvalvonnassa. Pysähtyminen voi myös vaikuttaa alavirran RNA:n prosessointiin, kuten RNA:n 5'-pään lakittamiseen (engl. *5'-end capping*) tai RNA:n silmukointiin (engl. *splicing*) (Akhtar ja muut, 2019). Pol II:n pysäyttämistä promoottorin läheisyyteen on vastuussa usea proteiini: DRB molekyylin herkkyttä kasvattava tekijä (engl. *DRB sensitivity inducing factor*, DSIF) ja negatiivinen pidennystekijä (engl. *negative elongation factor*, NELF) (kuva 4). DSIF on konservoitunut hiivoista ihmisiin, mutta NELF on konservoitunut vain



nisäkkäiden ja banaanikärpäsen välillä. (Saunders ja muut, 2006) DSIF ja NELF eivät itsessään aiheuta promoottorin läheisyyteen pysähtymistä, vaan siitä on vastuussa Pol II

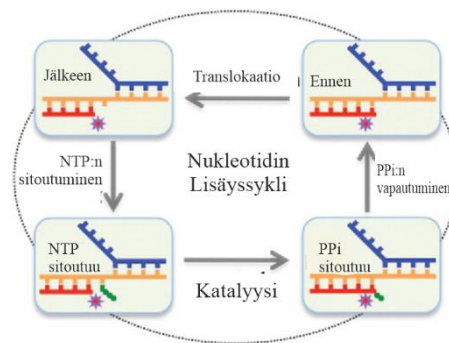


**Kuva 4.** Yleiskuva Pol II:n promoottorin läheisyyteen pysähtymisestä. Pol II pysähtyy promoottorin läheisyyteen DSIF ja NELF-proteiinien avustamana. Tämän jälkeen P-TEFb fosforyloi NELF-proteiinin, jolloin NELF irtoaa kompleksista ja Pol II pääsee jatkamaan transkriptiota. Kuva muokattu lähteestä Kwak ja Lis (2013).

aloitustekijä TFIID, ainakin *in vitro* promoottorilla, jossa ei ole nukleosomeja. DSIF ja NELF edesauttavat promoottorin läheisyyteen pysähtymistä TFIID:n läsnäollessa. (Fant ja muut, 2020) DSIF ja NELF välitteisessä promoottoriin läheisyyteen pysähtymisestä Pol II pääsee karkaamaan, kun TFIIH fosforyloi Pol II C-terminaalista päätä DSIF:n ja NELF:n sitouduttua. Tämän jälkeen transkription pidennysvaihetta edistävä tekijä b (engl. *positive transcription-elongation factor-b*, P-TEFb) fosforyloi DSIF:n sekä NELF:n. P-TEFb sitoutuu DSIF:n kanssa, jolloin NELF irtoaa kompleksista ja pidennysvaihe on valmis alkamaan. (Adelman ja Lis, 2012; Fant ja muut, 2020; Kwak ja Lis, 2013; Saunders ja muut, 2006) Edellä mainitun promoottorin ohjaaman pysähtymisen lisäksi on olemassa toisenlaista promoottorin läheisyyteen pysähtymistä, joka on nukleosomien aiheuttamaa. Tällöin Pol II pysähtyy lähimmän nukleosomin eteen, mutta Pol II:lla on lukuisia keinoja päästä sen ohi. (Kwak ja Lis, 2013) Promoottoriin läheisyyteen pysähtymistä tapahtuu transkription ensimmäisen 50 nukleotidin aikana. Promoottorin läheisyyteen pysähtymiseen johtaa jokin kolmesta mekanismista: promoottoriin sitoutuneet aloitustekijät ”vetävät” RNA-polymeraasia takaisin päin, syntyvä RNA ei ole tarpeeksi pitkä muodostaakseen liukumista estäviä rakenteita tai taempaa saapuvalla RNA:n tuottokompleksilla ei ole tarpeeksi tilaa ”puskea” peruuttanutta (engl. *backtracked*) RNA:n tuottokompleksia eteenpäin. (Nudler, 2012)

Transkription pidennysvaihe nimensä mukaisesti tuottaa pitenevää RNA:ta DNA-mallin mukaisesti. Transkription pidennysvaiheeseen päästessä on RNA jo pidentynyt usean emäksen verran. Osa tästä RNA:sta on yksittäisenä ketjuna RNA-polymeraasin

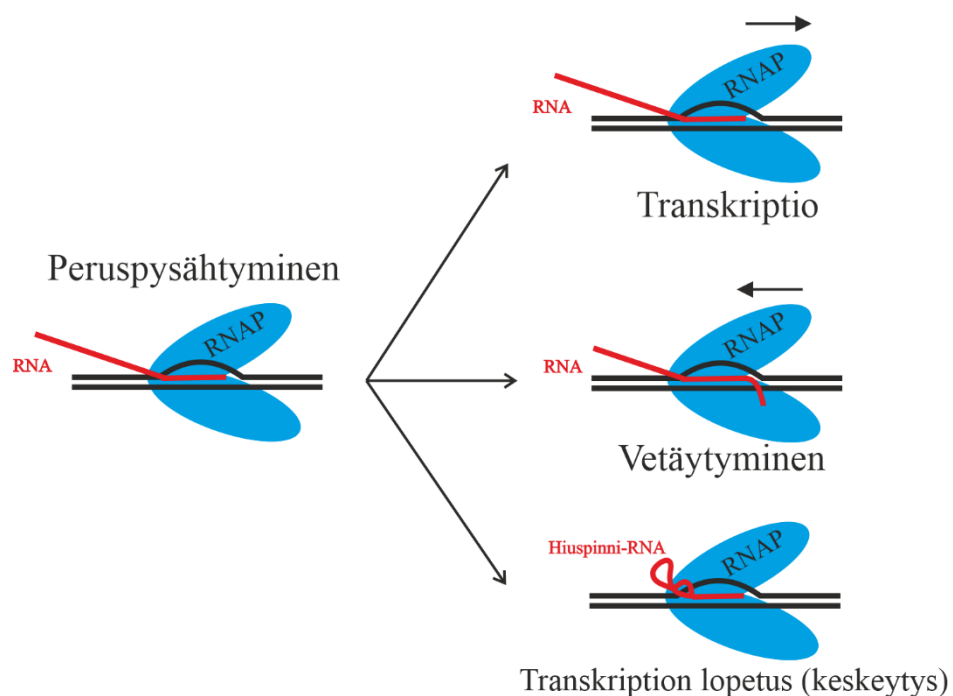
ulosmenokanavassa (7-11 emästä) tai polymeraasin ulkopuolella, osa vielä DNA/RNA-hybridinä RNA-polymeraasin sisällä aktiivisessa keskuksessa (9-10 emästä). RNA-polymeraasi pidentää RNA:ta sen 3'-päästä katalyysissä käyttäen apuna magnesiumioneja (Steitz, 1998). Laukaisinsilmukka (engl. *trigger loop*) on vastuussa uuden nukleotidin liittämistä kasvavan RNA:n 3'-päähän. Laukaisinsilmukka myös tarkistaa, onko liitetty nukleotidi mallijuosteeseen sopiva. (Wang ja muut, 2006) Tehokas nukleotidin liittämisykli koostuu neljästä vaiheesta (kuva 5): nukleosiditriphosfaatin (engl. *nucleoside triphosphate*, NTP) sitoutumisesta, katalyysistä, pyrofosfaatin (PP<sub>i</sub>) vapautumisesta ja translokaatiosta. RNA-polymeraasi valitsee mallijuostetta vastaavan



**Kuva 5.** Nukleotidin lisäyssi. Pidentävä RNA saa uuden nukleotidin neljässä vaiheessa. Ensin translokaation jälkeisessä tilassa olevaan RNA-polymeraasiin sitoutuu uusi NTP, jolloin tapahtuu katalyysi, nukleotidi liittyy pidentyvään RNA-ketjuun. Pyrofosfataasi irtoaa ja RNA-polymeraasi suorittaa translokaation. Kuva muokattu lähteestä Zuo ja Steitz (2017).

NTP:n, joka sitoutuu polymeraasiin. RNA-polymeraasin täytyy olla translokaation jälkeisessä -tilassa (engl. *post-translocational*) NTP:n sitomiseksi, jolloin sen aktiivinen keskus on avoin sitoutumiselle. (Zuo ja Steitz, 2017) Tämän jälkeen laukaisinsilmukka laskostuu ja yhdessä magnesiumin kanssa katalysoi NTP:n RNA-ketjun 3'-päähän (Wang ja muut, 2006). Tämän jälkeen NTP:stä irtoaa PP<sub>i</sub> ja translokaatio on valmis alkamaan (Malinen ja muut, 2012). Translokaatiossa DNA/RNA-hybridi siirtyy eteenpäin yhden nukleotidin verran, jolloin kehittyvä RNA (engl. *nascent RNA*) irtoaa DNA -templaattista DNA/RNA hybridin ylävirrassa. Ylävirrassa yksi DNA-emäspari liittyy yhteen, kun alavirrassa yksi pari eroaa toisistaan, jolloin transkriptiokupla pysyy saman mittaisena. Pidentävä RNA ohjataan erilliseen ulosmenokanavaan. DNA/RNA-hybridin erottamisesta on vastuussa kansisilmukka-domeeni (engl. *lid loop*), joka muodostaa steerisen esteen DNA/RNA-hybridille. Myös ylävirran transkriptiokuplan sulkeutuminen on tärkeää DNA/RNA-hybridin irtoamiselle. (Vassylyev ja muut, 2007) Translokaation jälkeen RNA-polymeraasi on valmis sitomaan seuraavan NTP:n.

Ideaalitilanteessa RNA-polymeraasi etenee pidennysvaiheen alusta loppuun vailla häiriötekijöitä, mutta yleensä RNA-polymeraasi pysähtyy kesken transkription. Pysähtymistä on joko väliaikaista, jolloin kyseessä on tauko (engl. *pause*) ja RNA-polymeraasi kykenee vielä jatkamaan tuon tauon jälkeen transkriptiota. Mikäli RNA-polymeraasi ei kykene jatkamaan transkriptiota, on kyseessä keskeytys (engl. *arrest*). (Artsimovitch ja Landick, 2000) RNA-polymeraasin pysähtyessä pidettyvän RNA:n 3'-pää ei enää ole aktiivisessa keskuksessa, jolloin pidennyskompleksi ei ole enää aktiivinen. Pysähtynyt RNA-polymeraasi pysyy kuitenkin paikallaan DNA:ssa. (Nudler, 2012) Pysähtyminen on yksi transkription pidennysvaiheen laadunvarmistusmekanismeista, joilla RNA-polymeraasi varmistaa syntyvän RNA:n emäsjärjestyksen oikeellisuuden. Peruspysähtyminen (engl. *elemental pause*) on lyhytaikainen pysähtyminen, johon RNA-polymeraasilla on aina mahdollisuus joutua (kuva 6). Pysähtyminen aiheuttaa myös rakenteellisia muutoksia RNA-polymeraasissa. (Artsimovitch ja Belogurov, 2018) RNA-polymeraasi voi päästä jatkamaan transkriptiota samasta kohdasta johon se pysähtyi, mutta joskus se joutuu vetäytyneeseen tilaan (engl. *backtracking*). Vetäytyessään RNA-polymeraasi siirtyy takaisinpäin yhden tai useamman emäksen verran. Mikäli RNA-



**Kuva 6.** Havainnekuva bakteerien RNA-polymeraasin pysähtymisestä. Pysähtynyt RNA-polymeraasi (RNAP) voi joko jatkaa transkriptiota peruspysähtymisen jälkeen tai se voi esimerkiksi siirtyä vetäytymiseen, jossa RNA-polymeraasi siirtyy taaksepäin päästäen RNA:ta ulos. RNA-polymeraasilla on muutamia tapoja selvittää vetäytyneestä tilasta, se voi esimerkiksi leikata ulos puskeensa RNA:n ja jatkaa tästä transkriptiota. Tämä on todennäköistä, mikäli RNA-polymeraasi on tehnyt virheen transkriptiossa. RNA-polymeraasi voi myös peruspysähtymisensä kautta joutua keskeytettyyn tilaan, jolloin transkriptio lopetetaan.

polymeraasi vetäytyy useamman emäksen, purkautuu DNA/RNA-hybridi osin ja jo syntynyttä RNA:ta pusketaan ulos nukleotidien sitoutumispaikasta. Mikäli RNA-ketjussa on väärä emäs, RNA-polymeraasi katkaisee sen endonukleolyysillä, jonka jälkeen RNA-polymeraasi jatkaa transkriptiota normaalisti. Mikäli korjattavaa ei ole, on RNA-polymeraasin päästävä siirtymään vetäytyneestä tilasta, jolloin RNA-polymeraasi voi käyttää ulos työntämäänsä RNA-ketjua ja tämän jälkeen jatkaa RNA:n pidentämistä. (Shaevitz ja muut, 2003)

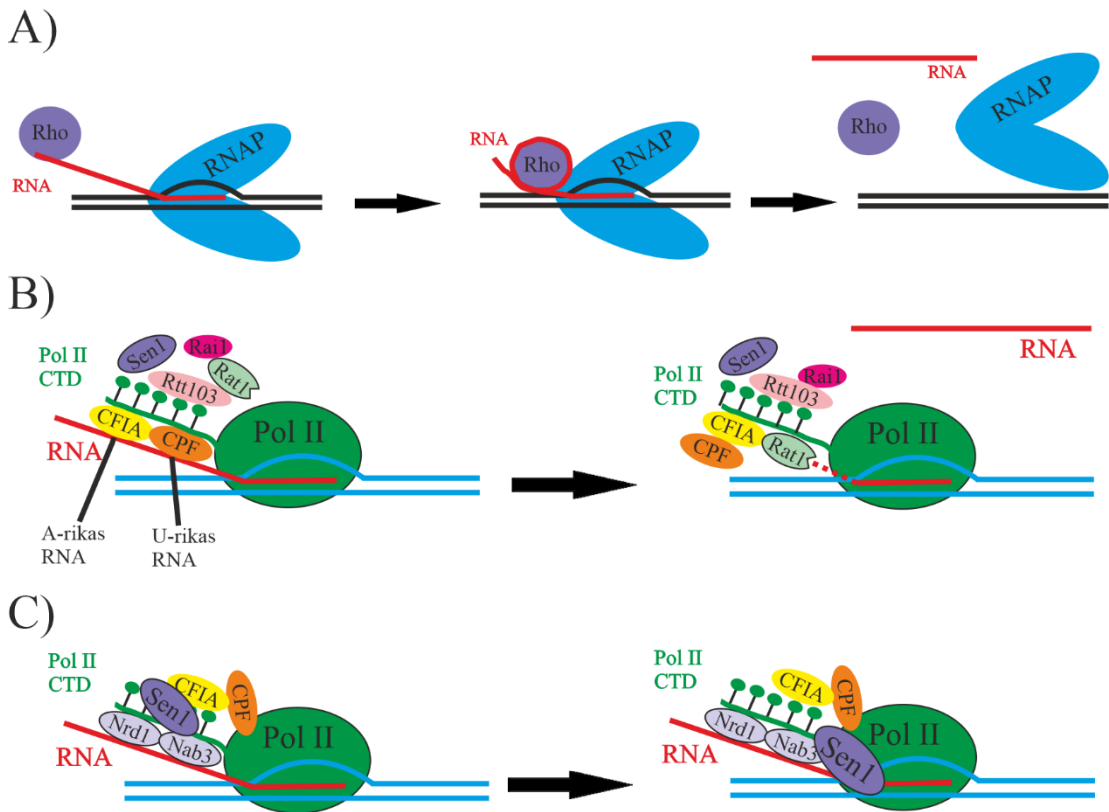
RNA-polymeraaseilla on useita pidennystekijöitä, jotka joko edesauttavat transkriptiota tai hidastavat sitä. Kolibakteerilla on ainakin kaksi tällaista tekijää; NusG ja NusA. NusG edesauttaa transkriptiota vähentämällä RNA-polymeraasin pysähtymisaikaa, kun taas NusA hidastaa transkriptiota lisäämällä RNA-polymeraasin pysähtymistiheyttä ja yksittäisen pysähtymisen kestoa. (Artsimovitch ja Landick, 2000) NusG:n ei havaittu vaikuttavan yksittäiseen RNA:n pidentämisreaktioon eli katalyyttiseen sykliin, joten sen transkriptiota edistävä vaikutus perustuu pelkästään RNA-polymeraasin pysähtelyn ja vetäytymisen vähentämiseen (Turtola ja Belogurov, 2016). NusG on myös ainoa kaikissa eliöissä konservoitunut transkriptiotekijä, leivinhiivalla NusG:tä vastaava proteiini on Spt5. Kolibakteerilla on monia muitakin transkription pidennystekijöitä, kuten Gre-proteiiniperheen proteiinit (Roberts ja muut, 2008). Pol II yleisinä pidennystekijöitä ovat P-TEFb, TFIIS, jotka estävät Pol II:n keskeytystä, sekä TFIIIF, elongiini ja ELL, jotka vähentävät polymeraasin taukoja (Reines ja muut, 1996).

Transkription lopetusvaihe voi alkaa monesta syystä, joista toivottavin on transkriptoitavan geenin päättymisen. Transkription lopetus on siis tärkeä säätelyvaihe transkriptiossa, jolloin väärin muodostuneet RNA-ketjut voidaan hävittää tai oikein muodostuneet RNA-ketjut siirretään translaatioon. Siirtyminen transkription lopetukseen voi tapahtua bakteereilla kahdella tavalla. Luontaisessa lopetuksessa (engl. *intrinsic termination*) ei vaadita RNA-polymeraasin, DNA:n ja RNA:n lisäksi mitään muuta. Toinen tapa on erillisistä lopetustekijöistä (engl. *termination factor*) riippuva tapa. Bakteerien RNA-polymeraasien tunnetuin lopetustekijä on Rho, joka hajottaa RNA:n pidennyskompleksin. (Ray-Soni ja muut, 2016) Luontaisessa lopetuksessa RNA-polymeraasi tunnistaa lopetussekvenssin, joka koostuu ylävirran G-C-alueesta ja sitä seuraavasta T-rikkaasta alueesta. Tästä alueesta RNA-polymeraasin luoma RNA-sekvenssi johtaa lopetushiuspinninrakenteen (engl. *terminator hairpin*) muodostumiseen. (Carafa ja muut, 1990) RNA-polymeraasi lopettaa transkription kolmessa vaiheessa.

Ensimmäiseksi RNA:n pidennyskompleksi pysähtyy mallijuosteen T-rikkaan alueen päähän. T-rikas sekvenssi johtaa heikkoon DNA/RNA-hybridin, jolloin RNA-polymeraasin kansi ja koura-domeenit löystyvät. RNA:lla on tilaa muodostaa ulosmenokanavaan lopetushiuspinnirakenne ja näin pakottaa RNA-polymeraasin konformaatiomuutokseen. Tällöin RNA-polymeraasi on avoimempi ja RNA pääsee irtoamaan DNA/RNA-hybridista. (Gusarov ja Nudler, 1999; Ray-Soni ja muut, 2016) Luontaisen lopetuksen lisäksi on Rho-riippuvaista lopetusta (kuva 7 A). Molemmat tavat ovat sekvenssistä riippuvaisia, ja noin puolet lopetuksista on Rho-riippuvaisia. Rho tunnistaa sekvenssin syntyvässä RNA:ssa. Rho sitoutuu RNA-polymeraasin transkriptoimaan Rho-alueeseen, joka sijaitsee ennen varsinaista lopetussekvenssiä, ja liikkuu sitä pitkin kohti RNA-polymeraasia. Rho liikkuu RNA:ta pitkin nopeammin kuin RNA-polymeraasi transkriptoi RNA:ta, joten se saavuttaa RNA-polymeraasin. Kun Rho saavuttaa RNA-polymeraasin, se avaa helikaasi-aktiivisuutensa avulla DNA/RNA hybridin. Tämän jälkeen Rho poistaa syntyneen RNA:n, jolloin transkriptiokupla luhistuu ja RNA-polymeraasi irtoaa DNA:sta. (Ray-Soni ja muut, 2016) Pol II systeemissä transkription lopetus tapahtuu joko poly(A)-riippuvaisesta lopetuksesta (engl. *Poly(A)-dependent termination*) tai Sen1 proteiinista riippuvaisesta lopetuksesta (kuva 7 B ja C). Poly(A)-riippuvaisessa lopetuksessa Pol II päättyy transkriptoimaan geenin loppua, jossa sijaitsee poly(A)-alue. Pol II transkriptoi poly(A)-alueen, jonka jälkeen se pysähtyy ja endonukleolyttisesti katkaisee syntyneen RNA:n. RNA:n katkaisusta on vastuussa erillinen proteiini, RNA-trafiikki proteiini (engl. *RNA trafficking protein*, Rat1). Rat1 on 5'-3' eksoribonukleaasi, joka vaatii transkription lopetukseen myös avustavia proteiineja (kuva 7 B). Rat1 syö tietään RNA:ta pitkin kohti Pol II:ta, jolloin Rat1:n ja Pol II:n välinen törmäys aiheuttaa transkription lopetuksen. RNA vapautuu, kun Rat1 alkaa katkaisemaan RNA:ta 5'-3' suunnassa. Toinen tapa transkription lopetukseen Sen1 proteiini aiheuttama transkription lopetus, jossa Sen1 avaa helikaasi-aktiivisuutensa

avulla DNA/RNA-hybridin Pol II sisällä. Molemmat Pol II transkription lopetustavat vaativat kuitenkin useita avustavia proteiineja. (Kuehner ja muut, 2011)

Bakteerien ja eukaryoottien RNA-polymeraasien toiminta transkription aloituksesta sen lopettamiseen on hyvin samankaltaista, mutta suurin ero näiden transkriptiosysteemien välillä on avustavissa proteiineissa. Myös leivinihiivan RNA-polymeraasien I, II ja III välillä suurin ero on avustavissa proteiineissa, sillä itse polymeraasit ovat hyvin samankaltaisia. Kun ymmärrämme transkription perusmekanismin, on myös jonkin yksittäisen RNA-polymeraasityypin toiminnan ymmärtäminen helpompaa. Toistaiseksi vähiten tutkittu kolmesta eukaryoottisesta polymeraasista on Pol I, jonka tutkimus on tosin viimeisen kymmenen vuoden aikana elpynyt. Koska kokeellinen pro gradu -tutkimukseni keskittyi Pol I:een, esittelen seuraavaksi tarkemmin sen toiminnan

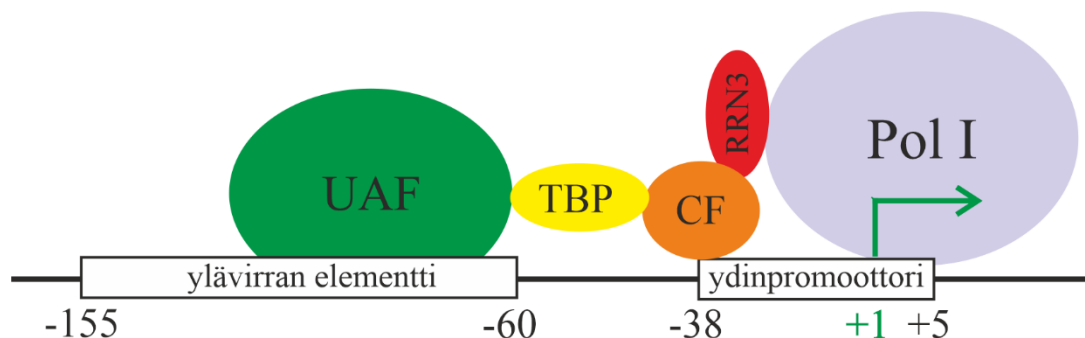


**Kuva 7.** Transkription lopetusvaihe bakteereilla ja leivinihiivalla. **A)** Rho-proteiinista johtuva bakteerien RNA-polymeraasin lopetusvaihe. Rho-tunnistaa RNA:ssa tietyn sekvenssin, johon se tarttuu ja jota pitkin se etenee kohti RNA-polymeraasia. Saavuttaessaan RNA-polymeraasin se on kietonut RNA:ta ympärilleen. Rho avaa helikaasi-aktiivisuutensa avulla RNA-polymeraasin sisällä olevan DNA/RNA-hybridin ja poistaa syntyneen RNA:n jolloin transkriptiokupla luhistuu ja RNA-polymeraasi irtaoo DNA:sta. **B)** Poly(A)-riippuvainen transkription lopetusvaihe leivinihiivalla. Lopetusvaiheen avustavat proteiinit fosforyloivat Pol II:n C-terminaalista päätä (Pol II CTD), jolloin avustavat proteiinit pääsevät sitoutumaan. Avustavat proteiinit tunnistavat A-rikkaan sekvenssin eli Poly(A) sekvenssin, jonka jälkeen Rat1 katkaisee RNA:n 5'-3' suunnassa, sillä se on eksoribonukleaasi. **C)** Sen1 riippuvainen transkription lopetusvaihe. Erilaisen lopetussekvenssin tunnistettuaan Sen1 ohjautuu transkriptiokuplan reunalle, jossa se helikaasi-aktiivisuutensa avulla avaa DNA/RNA-hybridin, mikä lopulta johtaa RNA:n vapautumiseen.

ymmärtämisen nykytilan, käyttäen esimerkkinä erityisesti eniten tutkittua leivinihiivan Pol I:tä.

### 2.3 RNA-polymeraasi I:n transkription aloitus

Leivinihiivassa Pol I tuottaa pelkästään ribosomaalista 35 S RNA:ta, jonka solu pilkkoo edelleen 18 S, 5,8 S ja 25 S rRNA:ksi. Tämä 35S rDNA sijaitsee kromosomissa XII noin 100–140 toistona (Petes, 1979). Nopeasti kasvavissa soluissa rRNA:n määrä on 80 % kaikesta RNA:sta ja Pol I:n aktiivisuus 60 % kaikesta transkriptiosta (Warner, 1999). rRNA:n suuren tuottotason selittää useiden geenikopioiden lisäksi se, että yksittäistä rDNA-geeniä transkriptoi yhtä aikaa monta peräkkäin etenevää Pol I:a (Hontz ja muut, 2008; Miller ja Beatty, 1969). Pol I -transkriptiokoneisto aloittaa transkription aina samalta promoottorilta. Tämän promoottorin välttämätön alue alkaa 155 emästä ylävirtaan transkription aloituskohdasta ja jatkuu 27 emästä alavirtaan, eli promoottorin sijainti on tällöin lyhennetyksi -155/+27, jossa transkription aloituskohta on +1 (Choe ja muut, 1992; Keener ja muut, 1998; Musters ja muut, 1989; Petes, 1979). Tämä promoottori on jaettavissa kahteen elementtiin: ylävirran elementtiin ja ydinelementtiin. Ylävirran elementti sijoittuu promoottorissa kohtaan -155/-60 ja ydinelementti sijoittuu kohtaan -38/+5 (Keener ja muut, 1998) (kuva 8). Koska Pol I tuottaa vain rRNA:ta, on sen promoottori helppo selvittää.



**Kuva 8.** Malli Pol I aloituskompleksin sitoutumiselle. UAF sitoutuu ylävirran elementtiin, joka sijaitsee välillä -155/-60. CF sitoutuu ydinpromoottoriin, joka sijaitsee välillä -38/+5. TBP:n arvellaan sitoutuvan CF:n ja UAF:n väliin, mutta varmaa tietoa TBP:n sitoutumisesta ja osallisuudesta ei vielä ole. UAF, TBP ja CF muodostavat pohjan, jolla Pol I - RRN3-kompleksi sitoutuu promoottoriin ja aloittaa RNA:n tuoton kohdasta +1.

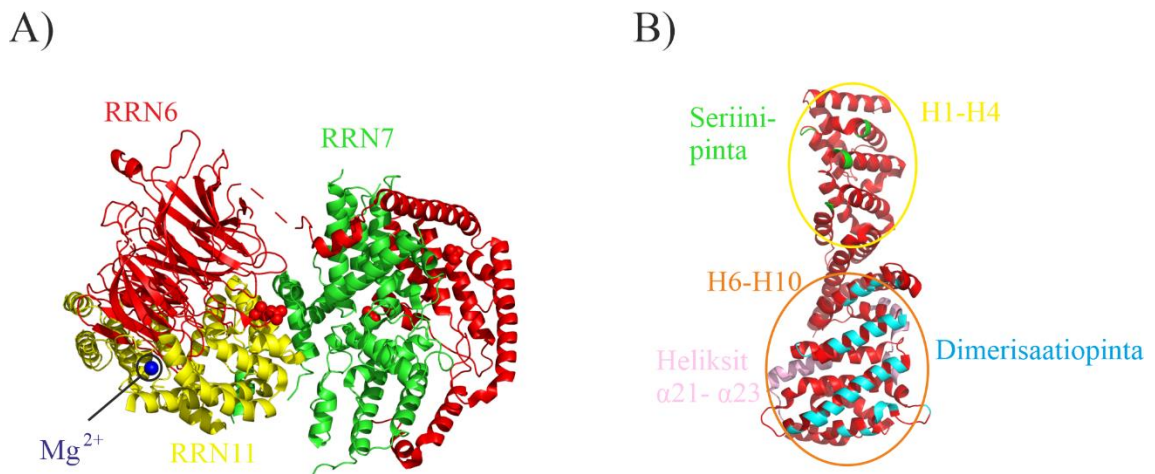
#### 2.3.1 Välttämätön aloituskompleksi

Transkription aloitus *in vitro* vaatii Pol I:n lisäksi vain kaksi proteiinia, RRN3:n ja CF:n. Tätä kutsutaan välttämättömäksi aloituskompleksiksi, joka toimii vain 2–10 % teholla verrattuna täydelliseen aloituskompleksiin. Aloitustekijöistä lähimmäksi transkription aloituskohtaa (+1) sitoutuu alayksiköistä RRN6, RRN7 ja RRN11 koostuva

proteiinikompleksi CF (Lalo ja muut, 1996) (kuva 8). CF sitoutuu promoottorin ydinpromoottorin yläosaan, mutta sen tarkasta sitoutumiskohdasta on ollut paljon erimielisyyksiä. Tämänhetkisen tiedon mukaan se sitoutuu välille -32/-13 (Kulkens ja muut, 1991; Sadian ja muut, 2017; Sadian ja muut, 2019). Koska Pol I promoottoreilla lajien välillä ei ole homologiaa, on myös mahdollista, ettei CF tunnista mitään tiettyä emäsjärjestystä, vaan DNA:n paikallisen rakenteen. Yhdeksi tällaiseksi rakenteeksi on ehdotettu DNA:n GC-rikasta pientä uurretta (engl. *minor groove*). DNA:n rakenteen tunnistamista tukee myös se, että hiivan CF sitoutuu emäsjärjestykseltään erilaiseen ihmisen rDNA-promoottoriin. CF:n on siis ehdotettu tunnistavan Pol I-promoottorin GC-rikas alue. (Jackobel ja muut, 2019) Kvantitatiiviset parametrit, jotka määrittävät CF:n sitoutumisen tiettyyn kohtaan rDNA-promoottoria ovat edelleen huonosti ymmärrettyjä. Muilla DNA:han sitoutuvilla aloitustekijöillä on CF:stä poiketen sekvenssispesifisyyttä (Chasman ja muut, 1993; Jackobel ja muut, 2019; Keys ja muut, 1996). CF on kuitenkin välttämätön aloitustekijä. (Peyroche ja muut, 2000). Tämä näkyy myös Pol I-RRN3-CF-rakenteissa (Engel ja muut, 2017; Sadian ja muut, 2019). Itsenäisenä proteiinina CF koostuu kahdesta moduulista (kuva 9 A). Moduuli 1 koostuu RRN6 C-terminaalista päästä sekä RRN11 proteiinista ja moduuli 2 puolestaan koostuu RRN6 N-terminaalista päästä sekä RRN7 proteiinista. Moduulien välissä on ohut sarana-alue (engl. *hinge region*), joka liittää moduulit toisiinsa. RRN6 käytännössä sulkee RRN7:n ja RRN11:ta sisäänsä CF:n vapaassa muodossa. (Engel ja muut, 2017) CF sitoutuu DNA:han alayksiköillä RRN11 ja RRN7, mikä johtuu todennäköisimmin CF:n rakenteen muuttumisesta sen sitoutuessa DNA:han (Pilsel ja Engel, 2020). Alayksikkö RRN7 muistuttaa Pol II aloituskompleksin transkriptiotekijää TFIIB (Blattner ja muut, 2011; Engel ja muut, 2017), mutta niiden samankaltaiset rakenteet omaksuvat erilaiset asennot komplekseissa (Pilsel ja Engel, 2020). CF sitoutuu TBP:hen todennäköisimmin RRN6



alayksikön kautta, mutta asia ei ole täysin selvä, sillä TBP:n on puolestaan osoitettu vuorovaikuttavan sekä RRN7:n että RRN11:n kanssa (Steffan ja muut, 1996).



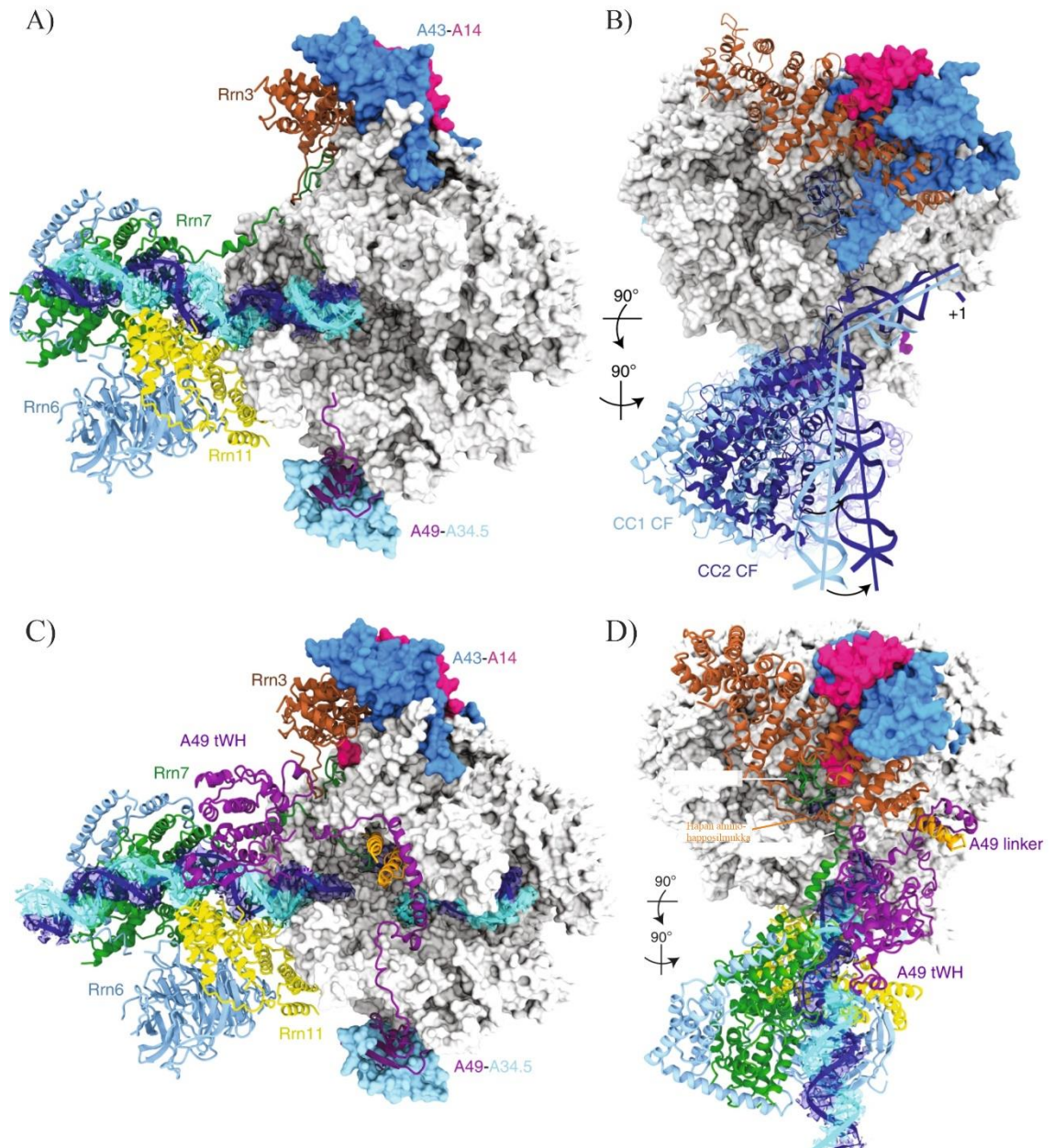
**Kuva 9.** CF:n ja RRN3:n rakennekuvat. **A)** CF koostuu kolmesta alayksiköstä: RRN6:sta, RRN7:stä ja RRN11:sta (PDB id: 5O7X). Kasautuakseen CF tarvitsee myös magnesiumioneja, joista tyypillisesti yksi erotetaan ratkaistuissa kide- ja kryo-EM-rakenteissa. Vasemmalla on moduuli 1, joka koostuu RRN6:n N-terminaaliosasta ja RRN11:sta, oikealla on moduuli 2, joka koostuu RRN6:n C-terminaaliosasta ja RRN7:stä. Moduuleja yhdistää sarana-alue. **B)** RRN3-rakenteen seriinipinta on merkitty vihreällä, dimerisaatiopinta vaaleansinisellä ja heliksit  $\alpha 21$ -  $\alpha 23$  vaaleanpunaisella (PDB id: 3T1J). HEAT-toistojaksot H1-H4 on ympyröity keltaisella ja H6-H10 oranssilla.

Aloitustekijöistä RRN3 sitoutuu suoraan Pol I proteiiniin muodostaen holoentsyymin. RRN3 sitoutuu Pol I proteiiniin myös ilman promoottorin läsnäoloa. (Yamamoto ja muut, 1996) RRN3:lla ja DNA:lla ei ole osoitettu olevan vuorovaikutusta keskenään. RRN3 sitoutuu Pol I alayksikköön A43 (Peyroche ja muut, 2000), tarkemmin sanottuna Pol I sauvan N-terminaaliseen päähän (Engel ja muut, 2013). RRN3 ylettyy myös alayksikön A190 ja AC40-AC19 heterodimeerin pintaan, vakauttaen Pol I monomeeristä, transkriptioon kykenevää muotoa (Engel ja muut, 2016; Pilsl ja muut, 2016). Ilman RRN3:a Pol I ei pysty tehokkaaseen transkriptioon (Milkereit ja Tschochner, 1998). Kuten Pol I, myös RRN3 esiintyy vapaassa muodossa inaktiivisena homodimeerinä (Blattner ja muut, 2011; Engel ja muut, 2013). Rakenteellisesti RRN3 koostuu yhteensä kymmenestä HEAT-toistojaksosta, H1-H10 (Andandre ja Bork, 1995). Näistä H6-H10 osallistuvat RRN3 dimerisoitumiseen kolmen yksittäisten  $\alpha$ -heliksin  $\alpha 21$ -  $\alpha 23$  avulla. H1-H4 sisältävät seriinipinnan (engl. *serine patch*), joka viittaa RRN3:n toiminnan muokkaamiseen fosforylaation avulla. RRN3:n rakenne on konservoitunut, joten ne ovat melko samanlaiset niin leivinhiivalla kuin ihmiselläkin (kuva 9). (Blattner ja muut, 2011) RRN3 siis muuttaa dimerisaatiomodulin rakennetta sitoutuessaan, mikä viittaa RRN3:n

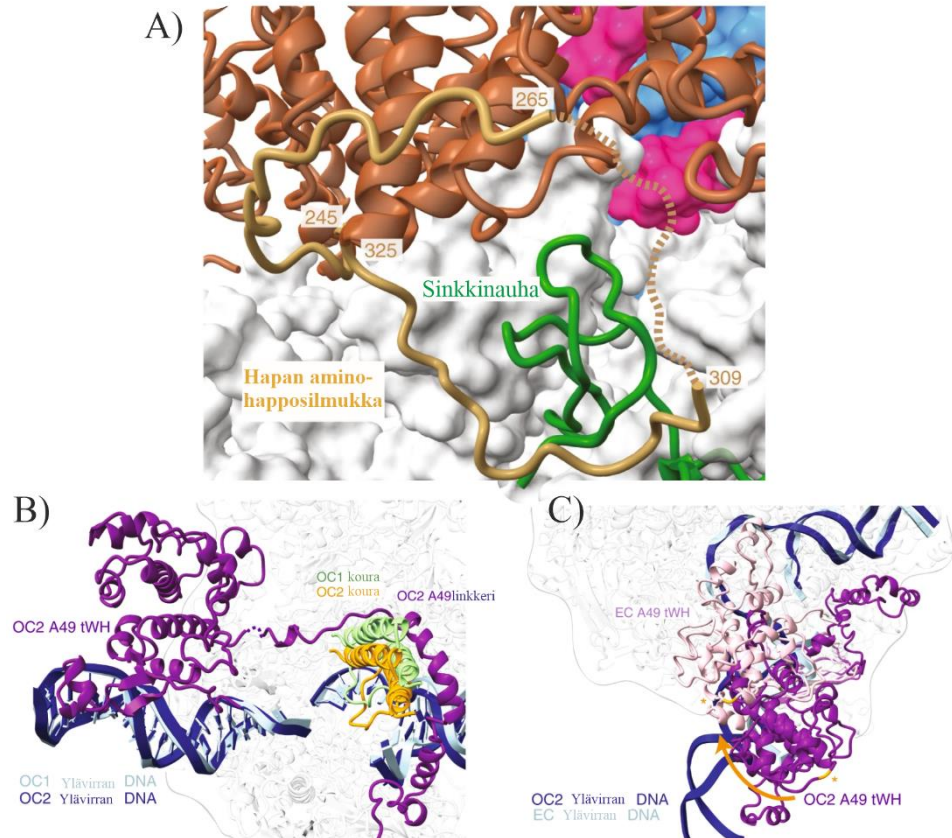
toimintaan Pol I monomeeriä stabiloivana proteiinina. (Engel ja muut, 2016; Pilsl ja muut, 2016)

Välttämätön aloituskompleksi sitoutuu promoottoriin todennäköisesti CF:n ohjaamana. CF:n moduuli 1 suuntautuu holoentsyymiin Pol I halkioon päin, moduuli 2 sen sijaan osoittaa pois päin holoentsyymistä. CF:n alayksikkö RRN7, joka sijaitsee moduulissa 2, on kuitenkin vuorovaikutuksessa RRN3:n kanssa (kuva 10). Tämä vuorovaikutus vaikuttaisi olevan ainoa sitoutumiskohta välttämättömän aloituskompleksin avustavien proteiinien välillä. (Engel ja muut, 2017) Pol I sitoutuu DNA:han halkionsa avulla. Pol I: halkio on DNA:han sitoutumattomassa muodossaan leveämpi, noin 42 Å, kuin muilla RNA-polymeraaseilla (Tafur ja muut, 2016). Pidentynyt muoto halkion läpäisevästä siltaheliksistä on vastuussa halkion auki pitämisestä, mutta laajaa halkiota vakauttaa myös erillinen levittäjä-domeeni (engl. *expander*). DNA:n sitoutuessa levittäjä kuitenkin siirtyy, jolloin aiemmin epäjärjestäytynyt siltaheliksien keskiosa laskostuu alfa-kierteeksi, minkä seurauksena siltaheliksien kokonaispituus lyhenee ja Pol I:n halkio supistuu. (Engel ja muut, 2013; Engel ja muut, 2017; Fernández-Tornero ja muut, 2013; Pilsl ja muut, 2016) Levittäjä vakauttaa myös Pol I alayksikön A12.2 sinkkinauhamotiivia (Fernández-Tornero ja muut, 2013; Kuhn ja muut, 2007). Pol I alayksikön A12.2 C-terminaalinen pää yltää Pol I:n aktiiviseen keskukseen asti ja sen poistuminen aktiivisesta keskuksesta on yksi edellytys Pol I:n transkription aloitukselle. A12.2 C-terminaalisen päään rooli on todennäköisesti keskeytetyssä aloituksessa, jossa se pilkkoo syntyvät lyhyet RNA:t. (Engel ja muut, 2017; Sadian ja muut, 2017) Välttämättömän aloituskompleksin sitouduttua DNA:han on kyseessä suljettu kompleksi ( $RP_c$ ), joka on havaittu Pol I:n tapauksessa kahdessa erilaisessa tilassa (kuva 10 A ja B). On mahdollista, että nämä tilat edustavat peräkkäisiä välivaiheita avoimen kompleksin muodostusmekanismissa. Molemmille  $RP_c$ -tiloille on yhteistä, että CF sitoutuu ylävirran DNA:han ja aiheuttaa noin 60° taivutuksen pois päin Pol I:n DNA:ta sitovasta halkiosta (kuva 10 B). Ero näiden kahden tilan välillä on ylävirran DNA:n kulman sijainti verrattuna holoentsyymiin. Aloituskompleksin esiintyessä  $RP_c$ :nä levittäjä ei ole vielä siirtynyt halkiosta.  $RP_c$ :n rakenteissa, jotka Sadian ja muut (2019) hiljattain selvittivät, näkyy myös toinen mielenkiintoinen seikka, josta ennen tätä ei ollut täyttä varmuutta: RRN3:n hapan aminohapposilmukka (engl. *acidic loop*), jota ei ennen ole rakenteissa selkeästi nähty, on kontaktissa RRN7:n sinkkinauhadomeeniin ja tätä kautta ankkuroi CF:n Pol I:een (kuva 11 A). Tämä vaikuttaisi olevan myös ainoa sitoutumiskohta RRN3:n ja CF:n välillä.

RRN3:n ja RRN7:n välinen kontakti vakauttaa CF:n Pol I pintaan, mutta alustavan transkription edetessä RRN7 rakenne törmää RRN3:n happamaan silmukkaan, jolloin sen rakenne häviää ja CF irtoaa kompleksista. (Sadian ja muut, 2019) Tämä malli on esitetty kryoelektronimikroskopiolla, mutta tätä mallia ei ole testattu toiminnallisesti.



**Kuva 10.** Pol I suljettu ja avoin kompleksi kahdesta eri kulmasta. **A)** Suljettu kompleksi, jossa Pol I on sitoutunut DNA:n ympärille, muttei vielä avannut sitä. RRN3 sijaitsee A43-A14 domeenin lähellä, kaukana DNA:sta. RRN3:n ja CF:n välinen kontakti ei ole nähtävissä tästä kuvakulmasta. Suljetun kompleksin rakennekuvista myös puuttuu A49 tWH. **B)** Suljettu kompleksi, mutta käännettynä. Tässä kuvassa myös suljetun kompleksin kaksi vaihetta, RPC 1 ja RPC 2 (CC1 ja CC2 kuvassa) ovat nähtävillä. RPC 1:n ja RPC 2:n ero on siinä, että ne taivuttavat DNA:ta hivenen eri verran, RPC 2 tilassa DNA on taivutetumpaa. **C)** Avoin kompleksi, jossa DNA on avattu. Avoimessa kompleksissa A49 tWH on nähtävissä, ja se sitoutuu RRN3:n ja CF:n väliin DNA:n ympärille. Suurta rakenteellista eroa suljettuun kompleksiin ei ole nähtävissä. **D)** Avoin kompleksi, mutta käännettynä. Tässä kuvassa näkyy juuri ja juuri RRN3:n ja RRN7:n välinen kontakti, mutta se on esitelty tarkemmin kuvassa 11. Kuva muokattu lähteestä Sadian ja muut (2019).



**Kuva 11.** Yksityiskohtia Pol I:n välttämättömästä aloituskompleksista. **A)** RRN3:n ja RRN7:n välinen kontaktipinta. RRN3:n hapan aminohapposilmukka ja RRN7:n sinkkinauha ovat kontaktissa toisiinsa, ollen toistaiseksi ainoa kontaktipinta RRN3:n ja CF:n välillä. **B)** A49 tWH:n sijainti avoimessa kompleksissa. A49 tWH sitoutuu ylävirran promoottoriin, mutta se vaatii puristin-domeenin supistumisen ennen sitoutumistaan DNA:han. **C)** A49 tWH:n asennot avoimessa kompleksissa ja transkriptoivassa pidennyskompleksissa. A49 tWH:n on liikuttava paljon, sillä se on pidennyskompleksin rakennekuvissa verrattuna ylävirran DNA:n tiellä. Kuva muokattu lähteestä Sadian ja muut (2019).

Välttämättömästä aloituskompleksista alavirran DNA:han sitoutuu kaksi domeenia, CF:n RRN7 ja RRN11 ja Pol I alayksikön A49 kaksoissiivekäs heliksi-domeeni (engl. *tandem winged helix*, tWH), joka sijaitsee alayksikön C-terminaaliosassa (Sadian ja muut, 2019) (kuva 11 B). Vaikka A49 tWH:n rooli onkin ollut epäselvä aikaisemmissa rakennekuvissa (Fernández-Tornero ja muut, 2013; Neyer ja muut, 2016; Pilsl ja muut, 2016), ei sen rooli DNA:ta sitovana domeenina ole yllätys, sillä A49 tWH muistuttaa hyvin paljon Pol II aloitustekijän TFIIE:n alayksikköä  $\beta$ . A49 on osa heterodimeeriä A49-A34.5, joka muodostuu kahdesta osasta, joista toinen on edellä mainittu A49 tWH ja toinen on A49-A34.5 dimerisaatiomoduuli. Tämä dimerisaatiomoduuli muistuttaa Pol II aloitustekijää TFIIF. (Fernández-Tornero ja muut, 2013; Geiger ja muut, 2010; Kuhn ja muut, 2007) Vaikka A49 tWH sitoutuu DNA:han, ei sillä ole DNA:n avaamiseen soveltuvia elementtejä kuten TFIIE:ssä, jolloin sillä ei todennäköisesti ole myöskään roolia DNA:n avaamisessa. DNA on kuitenkin avattava, jotta Pol I voi muuttua  $RP_c$ :tä

RP<sub>o</sub>:ksi. Pol I kompleksissa tätä roolia todennäköisesti hoitaa peräsin (engl. *rudder*), mutta DNA:n avaamisen tarkempi mekanismi on yhä selvittämättä. Kuitenkin Pol I:n ollessa RP<sub>o</sub>-tilassa, A49:n ja RRN7:n linkkerialueet ovat hyvin lähellä transkriptiokuplan reunaa, jolloin ne todennäköisesti vakauttavat ylävirran DNA:ta transkriptiokuplan avautumisen alussa (kuva 10 C ja D). (Sadian ja muut, 2019) RP<sub>o</sub>:n muodostumisen jälkeen Pol I puristin alkaa sulkeutua, jolloin halkio yhä pienenee, Pol I muuttuu alustavasti transkriptiivaksi kompleksiksi (RP<sub>ITC</sub>) ja aloittaa tällöin RNA:n tuoton (Tafur ja muut, 2016). RP<sub>ITC</sub>:n muodostuttua rakennekuvista puuttuu yleensä A49 tWH (Engel ja muut, 2017; Fernández-Tornero ja muut, 2013; Pilsl ja muut, 2016). Kun A49 tWH:n rakenteen selvittäminen onnistui RP<sub>ITC</sub>:ssä, RRN3:n katosi rakenteesta todennäköisesti siksi, että A49 siirtyy paikaltaan RNA:n pidentyessä tarpeeksi (kuva 11 C). Tämä johtaa törmäykseen RRN3:n kanssa, jolloin se irtoaa kompleksista. (Han ja muut, 2017). Tämä viittaa A49 tWH:n rooliin myös RRN3:n irrottajana, jolloin Pol I pääsisi karkaamaan promoottorilla.

### 2.3.2 Täydellinen aloituskompleksi

Täydellinen aloituskompleksi koostuu välttämättömän aloituskompleksin, eli promoottorin, Pol I:n, RRN3:n ja CF:n lisäksi UAF ja TBP-proteiineista (kuva 8). Täydellinen aloituskompleksi on vähintään 10-kertaa aktiivisempi kuin välttämätön aloituskompleksi. (Keener ja muut, 1998) Aloituskompleksin avustavista proteiineista UAF sitoutuu kauimmas transkription aloituspaikasta, ylävirran elementtiin (-155/-60) (Keys ja muut, 1996). UAF koostuu kuudesta alayksiköstä: RRN5, RRN9, RRN10, UAF30 sekä histoneista H3 ja H4 (Keener ja muut, 1997; Keys ja muut, 1996; Nogi ja muut, 1991; Siddiqi ja muut, 2001). UAF:n rooli Pol I välitteisen rDNA:n transkription aktivoimisen lisäksi on estää Pol II välitteinen rDNA:n transkriptio, minkä UAF tekee kromatiinirakenteeseen vaikuttamalla (Siddiqi ja muut, 2001). UAF muodostaa vakaan kompleksin promoottorin kanssa, muodostaen pohjan transkription aloitukselle ja näin auttaa CF:n ja Pol I:n sitoutumista (Keys ja muut, 1996). RRN5, RRN9 ja RRN10 ovat tärkeitä UAF:n toiminnalle, sillä mutaatiot näissä tai näiden alayksiköiden puuttuminen hidastaa merkittävästi hiivan kasvunopeutta (Keys ja muut, 1996; Nogi ja muut, 1991). UAF30 ei ole välttämätön alayksikkö UAF:n toiminnalle, joskin se vaaditaan täyteen transkriptioaktiivisuuteen (Siddiqi ja muut, 2001). *In vivo* UAF30:n puuttuminen hidastaa rDNA:n transkriptiota jopa 70 %, vaikka rDNA on yliladattuna polymeraaseilla. (Hontz ja muut, 2008; Siddiqi ja muut, 2001) Yksittäisissä deleetio-solukannoissa testattuna UAF

kasaantuu myös ilman yksittäisiä alayksiköitä. Poistettaessa RRN5 alayksikkö UAF kompleksista, histoneita H3 ja H4 havaitaan vähemmän. RRN5:llä on siten todennäköisesti rooli kyseisten histonien sitouttamisessa tai vakauttamisessa. (Smith ja muut, 2018) Histonien rooli UAF-proteiinikompleksissa on vielä osin selvittämättä, sillä histonien pääasiallinen tehtävä on toimia nukleosomeissa ja kromatiineissa, jossa ne pakkaavat DNA:n tiiviiksi. Histonit koostuvat kahdesta domeenista: rungosta ja hännästä (engl. *core* ja *tail*). (Zhou ja Bai, 2019) Histonin H3 häntä-domeenilla on UAF:n kasaantumista edistävä vaikutus, mutta myös pelkkä runko riittää UAF:n kasaamiseen. Toisaalta histonin H3 tai H4 puuttuessa UAF kykenee muodostamaan kompleksin, jolloin histonien toiminta UAF-kompleksissa on vielä selvittämättä. Tavallisessa UAF-kompleksissa on kuitenkin vain yksi kappale kutakin alayksikköä, pois lukien H3, jota on kaksi kappaletta. (Smith ja muut, 2018) Histonien puuttumista UAF-kompleksista *in vivo* on vaikea tutkia, sillä histonien poisto on todennäköisesti letaali mutaatio, toisin kuin UAF:n RRN- tai UAF30-alayksiköiden poisto. Vaikka UAF:n kaikki alayksiköt tunnetaan ja se on jopa onnistuttu tuottamaan rekombinanttiproteiininä (Smith ja muut, 2018), ei UAF:n kolmiulotteista rakennetta ole vielä selvitetty. Todennäköisesti UAF sitoo DNA:ta nukleosomien tapaan kiertämällä DNA:ta ympärilleen, sillä sen sitoma DNA-alue on melko pitkä. Tämä selittäisi myös histonien H3 ja H4 olemassaolon UAF-kompleksissa. (Keener ja muut, 1998)

TBP on osa jokaisen eukaryootti-RNA-polymeraasin toimintaa, myös Pol I:n. TBP on pieni 27 kDa yksialayksikköinen proteiini, joka sitoutuu DNA:han nimensä mukaisesti alueelle, jossa on TATA-emästoisto tai jokin sen variaatio (Chasman ja muut, 1993). TBP:n toiminta tunnetaan hyvin Pol II ja Pol III transkription aloitustekijänä, mikäli promoottorissa on TATA-alue, mutta TATA-alueettoman promoottorin tapauksessa TBP:n toiminta tunnetaan huonosti. Huonoiten TBP:n toiminta tunnetaan kuitenkin Pol I aloituskoneistossa. Pol I promoottorissa ei ole TATA-aluetta, johon TBP voi spesifisesti sitoutua (Choe ja muut, 1992; Keener ja muut, 1998). Kuitenkin TBP vaaditaan täydelliseen Pol I transkription aloituskoneistoon ja TBP stimuloi rRNA:n tuottoa soluissa (Aprikian ja muut, 2000). Näistä havainnoista huolimatta ei ole varmaa, miten TBP stimuloi soluissa Pol I aloituskompleksin muodostumista. On ehdotettu, että TBP muodostaisi sillan CF:n ja UAF:n välille. (Aprikian ja muut, 2001) TBP sitoutuu kuitenkin paremmin UAF:ään kuin CF:ään, mutta stimuloi Pol I transkriptiota myös ilman UAF:ää. TBP:n rooli aloituskompleksissa voi liittyä myös vahvasti CF:ään, sillä

TBP:n on osoitettu vakauttavan CF:n ja promoottorin välistä vuorovaikutusta. (Steffan ja muut, 1996) TBP on vuorovaikutuksessa UAF:n RRN9-alayksikön kanssa, todennäköisesti useastakin eri kohdasta (Steffan ja muut, 1996; Steffan ja muut, 1998). TBP:n ylituotto korvaa UAF:n poistamisen vaikutuksen Pol I:n aktiivisuuteen mutta ei CF:n poistamisen vaikutusta (Aprikian ja muut, 2000). Vaikka TBP:n on osoitettu sitoutuvan sekä UAF:ään että CF:ään, sitoutumisen varsinainen mekanismi ja rooli on vielä selvittämättä. Samoin on yhä selvittämättä, sitoutuuko TBP Pol I -promoottoriin vai toimiiko se vain siltana UAF:n ja CF:n välillä.

#### *2.4 RNA-polymeraasi I:n transkription pidennysvaihe*

Pol I:n siirtyminen aloituskompleksista RNA:ta pidentävään kompleksiin, eli RNA:n pidennysvaiheeseen, vaatii aloitustekijöiden irtoamisen Pol I:stä. Siirtymävaiheessa aloituksesta RNA:n pidentämiseen CF poistuu kompleksista todennäköisesti sen takia, että CF:n alayksikkö RRN7 on pidentyvän RNA:n tiellä. CF:n poistuessa myös RRN3:n hapan aminohapposilmukka, joka on RRN3:n ja RRN7:n vuorovaikutukseen vaadittava motiivi, siirtyy. RRN3 irtoaa siirtymävaiheessa, kun A49 tWH liikkuu aiheuttaen näiden kahden osan välille steerisen esteen. (Han ja muut, 2017; Sadian ja muut, 2019) Kuten aiemmin on mainittu, on pidennysvaiheen Pol I rakenne erilainen kuin aloitusvaiheen kompleksin. Transkriptoivassa Pol I:ssä ei tämänhetkisen käsityksen mukaan ole kiinni avustavia proteiineja (Neyer ja muut, 2016; Sadian ja muut, 2019), toisin kuin esimerkiksi Pol II:ssa (Shilatifard ja muut, 2003). Pol I:n karatessa promoottorilta todennäköisesti irtoavat aloitustekijät, lukuun ottamatta UAF:ää, joka pysyy promoottorilla ohjaten uusia Pol I aloituskomplekseja DNA:lle (Aprikian ja muut, 2001).

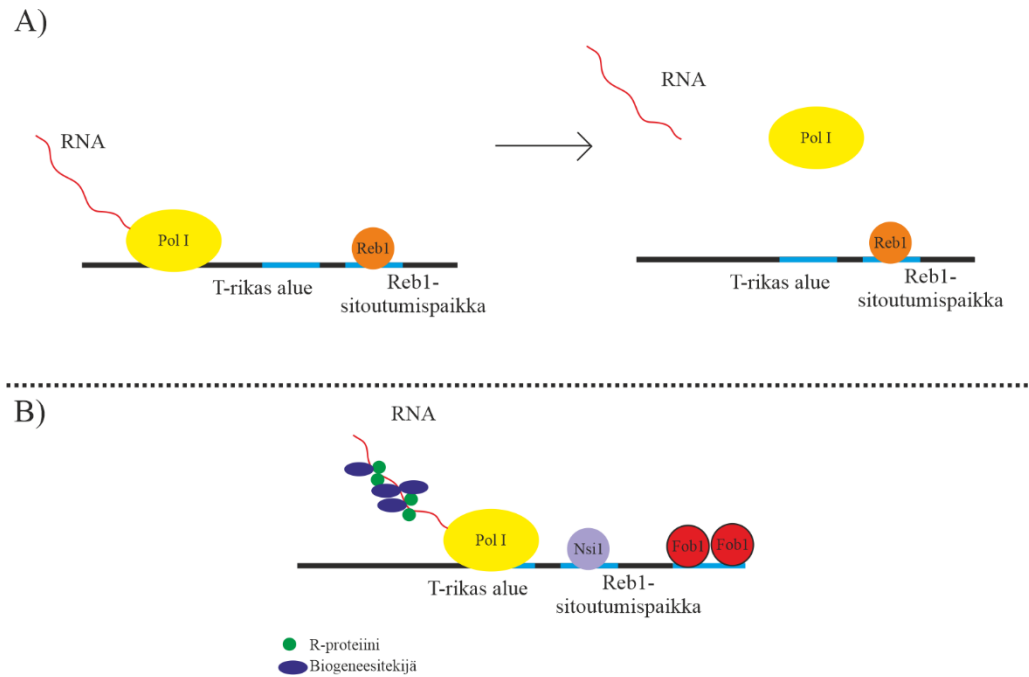
Pol I transkription pidennysvaihetta on tutkittu vähän verrattuna Pol II:een. Laukaisinsilmukka, joka on kaikkien monialayksikköisten RNA-polymeraasien välillä säilynyt rakenne, vaikuttaa suoraan RNA:n pidennysnopeuteen (Cramer ja muut, 2008). Laukaisinsilmukan toiminta käsiteltiin kappaleessa 2.2. Laukaisinsilmukoiden toiminnoissa on silti eroavaisuuksia Pol I ja Pol II välillä. Esimerkiksi mutaatio Rpb1-E1103G Pol II systeemissä kasvattaa transkriptionopeutta, kun taas vastaava mutaatio Pol I systeemissä (A190-E1124G) hidastaa transkriptionopeutta. Pol I laukaisinsilmukka ei siedä mutaatioita niin paljon kuin muut polymeraasit, sillä mutaatiot yleensä johtavat hitaampaan transkriptioon soluissa (Viktorovskaya ja muut, 2013). Villityypin Pol I:n maksimaalinen transkriptionopeus ( $k_{max}$ ) on 67 emästä sekunnissa ja vaadittava

emäspitoisuus puolessa transkriptionopeudessa  $K_{1/2} = 142 \mu\text{M}$ . A190-E1124G mutaation sisältämän Pol I transkriptionopeus on vain 19 emästä sekunnissa ja  $K_{1/2} = 64 \mu\text{M}$  (Scull ja muut, 2019). Toisessa tutkimuksessa RNA:n pidentämisnopeus villityypin Pol I:llä on 32 emästä/s ja Pol II:lla 18 emästä/s. Pol I myös pysähtyy harvemmin kesken transkription, kun taas pysähdystilan kesto on sama molemmilla polymeraaseilla. Vaikka Pol I joutuisi peruuttamaan useita emäksiä taaksepäin, kykenee se palauttamaan transkriptioaktiivisuutensa. Pol II kykenee myös peruuttamaan ja palauttamaan transkriptioaktiivisuutensa, mutta Pol II:lla suurin mahdollinen peruutusmatka, josta transkriptioaktiivisuus voi vielä palata on noin puolet lyhyempi kuin Pol I:llä. (Lisica ja muut, 2016) Tämä johtuu todennäköisesti siitä, että Pol I sisältää rakenteessaan RNA:ta pilkkovan alayksikön, A12.2:n, kun Pol II:ssa RNA:ta pilkkovana proteiinina toimii erillinen transkriptiotekijä TFIIS (Tafur ja muut, 2019). Myös palautumisaika peruutustilasta aktiiviseen transkriptioon oli Pol I:llä noin kolmanneksen lyhyempi kuin Pol II:lla. (Lisica ja muut, 2016) Pol I:n nopeampi transkriptionopeus johtuu todennäköisesti siitä, että se sisältää heterodimeerin A49-A34.5. Tämä heterodimeeri stimuloi transkriptiota ja vastaa osin Pol II transkriptiotekijää TFIIF. Tämän tarkempaa teoriaa ei toistaiseksi ole. (Geiger ja muut, 2010; Kuhn ja muut, 2007; Sadian ja muut, 2019)

### *2.5 RNA-polymeraasi I:n transkription lopetusvaihe*

Kaikkien RNA-polymeraasien transkriptio geenin lopussa päättyy kontrolloidusti. Varhaisin malli Pol I:n transkription lopetusvaiheelle on pysäytystekijä Reb1:n sitoutuminen geenin päässä olevaan pysäytyssekvenssiin (Lang ja Reeder, 1995). Pol I:n transkriptioiman geenin päässä on T-rikas alue, joka vaikuttaisi toimivan irrotuselementtinä. Reb1 sitoutuu tästä T-rikkaasta alueesta 15-emästä alavirtaan ja aiheuttaa transkription lopetuksen tällä alueella (kuva 12 A). Pol I alayksikkö A12 vaaditaan Reb1 välitteiseen transkription lopetukseen, joten kyseiset proteiinit todennäköisesti vuorovaikuttavat keskenään. (Prescott ja muut, 2004) Transkription lopetusvaihe on Pol I:llä ja bakteerien RNA-polymeraasilla hyvin samankaltainen, sillä molemmilla transkriptio päättyy T-rikkaaseen alueeseen (kuva 7 A ja kuva 12 A). Reb1 välitteinen transkription lopetus on yksinkertaisin mahdollinen malli, joka toimii *in vitro*. Soluissa transkription lopetusvaihe on todennäköisesti paljon monimutkaisempi, eikä





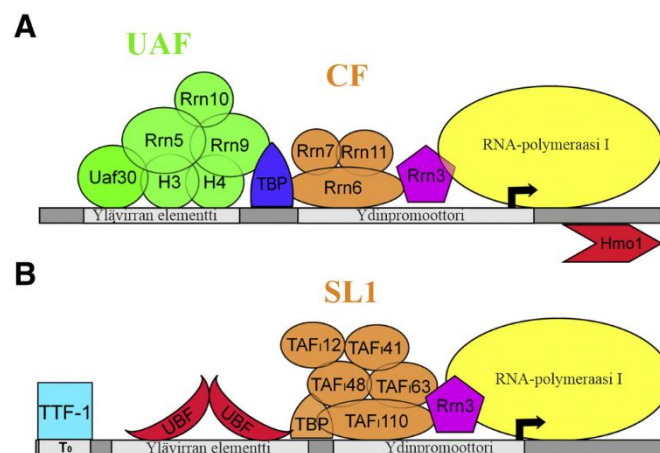
**Kuva 12.** Hiivan Pol I:n transkription lopetus *in vitro* ja *in vivo*. **A)** *In vitro* Reb1 sitoutuu Reb1-sitoutumispaikkaan, joka on 15-emästä alavirtaan T-rikaasta alueesta. Transkriptio loppuu, Pol I irtoaa DNA:sta ja RNA irtoaa Pol I:stä. **B)** *In vivo* lopetusvaihe on paljon monimutkaisempi, eikä sen toimintamekanismista ole täyttä varmuutta. Todennäköisesti ainakin Nsi1 edesauttaa Pol I:n transkription lopetusvaiheessa. Esitetty malli on yksi vaihtoehto transkription lopetukselle solussa.

Reb1:n toiminnasta Pol I transkription lopetusvaiheessa ole vielä täyttä varmuutta, sillä se toimii myös useissa Pol II geneissä transkription lopetustekijänä. (Németh ja muut, 2013) Hiljattain on löydetty lisää Pol I lopetustekijöitä, kuten ei-transloitavaa sekvenssiä hiljentävä proteiini (engl. *non-translated sequence (NTS) silencing protein*, Nsi1). Nsi1 on Reb1-homologi, joka vuorovaikuttaa rDNA:n tuottoa hiljentävien proteiinien kanssa. Se sitoutuu myös DNA:han samanlaisella Myb-rakenteella kuin Reb1. rDNA:ta sisältävässä kromosomi XII:ssä on noin 150 35S-DNA-toistoja, joiden välissä on kaksi ei-transloitavaa sekvenssiä: NTS1 ja NTS2: Nsi1 sitoutuu näistä NTS1-alueelle ja on tällöin NTS1-spesifinen hiljentäjä. Se, millä mekanismeilla Nsi1 aiheuttaa transkription lopetuksen, on vielä osin tuntematon. On kuitenkin osoitettu, että Nsi1 vaatii Fob1-proteiinin, jotta se voi tehokkaasti sitoutua DNA:han soluissa. (Ha ja muut, 2012; Reiter ja muut, 2012) Fob1 on replikaationestoproteiini, joka ohjaa rDNA:n transkription vain yhteen suuntaan (Defossez ja muut, 1999). Nsi1 kykenee myös yksinään transkription lopetukseen *in vitro* (Merk1 ja muut, 2014). Soluissa Nsi1 ja Reb1 eivät yksinään aiheuta transkription lopetusta Pol I koneistolla, mutta molempia todennäköisesti tarvitaan. Reb1:n ja Nsi1:n lisäksi todennäköisesti tarvitaan muitakin avustavia proteiineja, kuten mainittu Fob1. Tunnetuin mekanismi Pol I:n transkription lopetukseen on kuitenkin Reb1-välitteinen transkription lopetus. Mahdollisesti myös syntyvää RNA:ta transkription

kanssa samanaikaisesti muokkaavat proteiinit vaikuttavat Pol I:n transkription lopetukseen (kuva 12 B). (Németh ja muut, 2013)

### 2.6 Leivinihiivan RNA-polymeraasi I:n ja ihmisen RNA-polymeraasi I:n vertailu

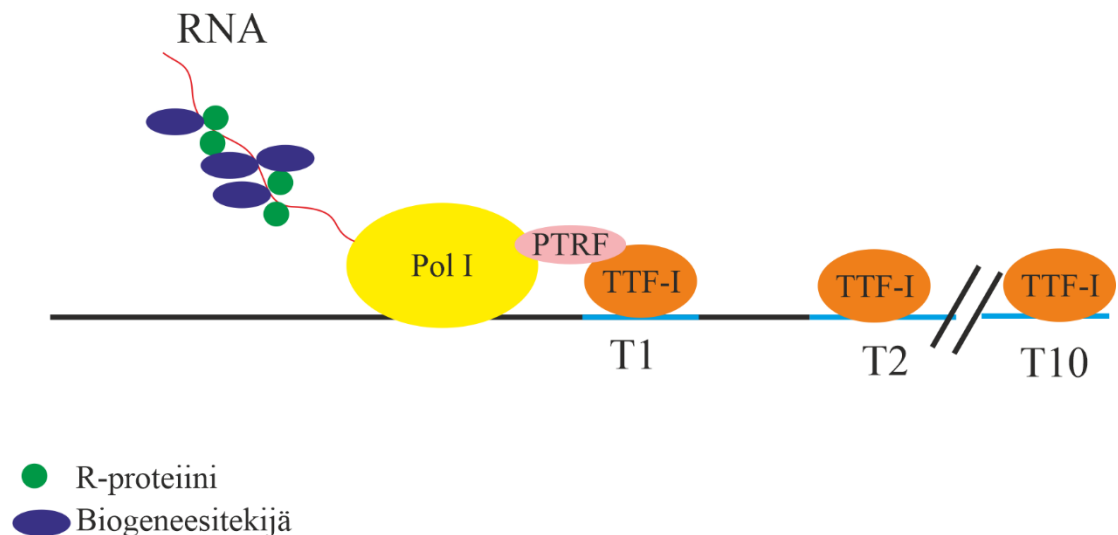
Ihminen ja leivinihiiva ovat molemmat eukaryootteja, mutta ihmisen genomi on suurempi ja monimutkaisempi. Siitä huolimatta myös ihmisellä on samanlaiset RNA-polymeraasit kuin leivinihiivalla ja niiden toimintatapa on samankaltainen. Ihmisen ja leivinihiivan Pol I koostuvat hyvin samanlaisista alayksiköistä ja ihmisen Pol I alayksiköt on nimetty pääosin leivinihiivan Pol I:n alayksiköiden mukaan. Leivinihiivan Pol I alayksiköistä vain A14:ta vastaava proteiini puuttuu ihmisen Pol I:stä. Ihmisen Pol I transkriptoimaa esi-RNA:ta kutsutaan 47S-RNA:ksi ja leivinihiivan 35S-RNA:ksi. (Russell ja Zomerdijk, 2006) Aloitustekijöistä vain RRN3 on säilynyt leivinihiivasta ihmiseen (kuva 13). Ihmisen RRN3:n (engl. *human RRN3*, hRRN3) toiminta Pol I:n transkription aloituksessa vaatii leivinihiivasta poiketen useita fosforylaatioita, joilla säädelään Pol I:n toimintaa (Cavanaugh ja muut, 2002). hRRN3:n lisäksi leivinihiivan ja ihmisen aloitustekijöillä ei ole homologiaa. Valikoiva tekijä 1 (engl. *selective factor 1*, SL1) vastaa ihmisen Pol I:n aloitustekijänä hiivan CF:a ja CF voi ainakin jossain määrin sitoutua SL1:n sitoutumispaikkaan ihmisen Pol I promoottorilla (Jackobel ja muut, 2019; Schneider, 2012) (kuva 13). SL1 koostuu TBP:stä sekä useista TBP:hen liittyvistä tekijöistä (engl. *TBP associated factors*, TAF) (Comai ja muut, 1994) (kuva 13 B). SL1 sitoutuu promoottorin ydinelementtiin, jonka avulla se ohjaa hRRN3:n yhdessä Pol I:n kanssa promoottorille (Miller ja muut, 2001). SL1-hRRN3-välitteinen Pol I:n sitoutuminen



**Kuva 13.** Pol I aloituskompleksien vertailua. **A)** Leivinihiivan Pol I, joka koostuu viidestä proteiinikompleksista: UAF:stä, TBP:stä, CF:stä, RRN3:sta ja itse Pol I:stä. **B)** Ihmisen Pol I aloituskompleksi. Se koostuu ainakin TTF-I:stä, UBF:stä, SL1:stä, hRRN3:sta ja Pol I:stä. Molemmat aloituskompleksit ovat toistensa kaltaisia, mutta CF ja SL1 eivät ole homologeja, eikä UAF ja UBF. Kuva muokattu lähteestä Schneider (2012).

promootorille vastaa hyvin paljon CF-RRN3-välitteistä leivinihiivan Pol I:n sitoutumista promootorille, vaikka proteiinikomplekseina CF ja SL1 ovat hyvin erilaisia eikä niiden kesken ole rakennehomologiaa. Leivinihiivan UAF:n sukulaisproteiinia ei ole löytynyt ihmisestä. UAF:n sijaan ihmisen ja muiden nisäkkäiden Pol I -koneistot käyttävät rakenteellisesti yksinkertaisempaa promootorin ylävirtaan sitoutuvaa tekijää UBF (engl. *upstream binding factor*). UBF stimuloi ihmisen Pol I:n transkriptiota, mutta se ei ole transkriptiolle välttämätön (Steffan ja muut, 1996). UBF sitoutuu ylävirran kontrollielementtiin, joka sijaitsee välillä -156/-107 (Learned ja muut, 1986). Homodimeerinen UBF kykenee taivuttamaan DNA:ta ympärilleen: UBF:n sitoutuessa DNA:han DNA kiertyy yhden kierroksen verran UBF:n ympärille noin 140 emäksen matkalta. Alayksiköiden sisällä on paljon sisäistä toistoa, jotka todennäköisesti osallistuvat DNA:n kiertämiseen UBF:n ympäri. (Russell ja Zomerdijk, 2006) Leivinihiivan ja ihmisen Pol I aloituskompleksien toiminta vaikuttaa hyvin samankaltaisilta, vaikka aloitustekijöistä vain yksi on yhteinen molemmilla. Loput aloitustekijöistä muistuttavat toisiaan enemmän toiminnaltaan kuin rakenteeltaan. Keskeisimpien aloitustekijöiden SL1:n, hRRN3:n ja UBF:n lisäksi ihmisen Pol I aloituskompleksin toiminnan säätelyyn osallistuu myös muita transkriptiotekijöitä. (Drygin ja muut, 2010; Schneider, 2012)

Korkeammilla eukaryooteilla on konservoitunut Reb1:n kaltainen proteiini, transkription lopetustekijä I (engl. *transcription termination factor I*, TTF-I) (kuva 14). TTF-I on



**Kuva 14.** Nisäkkäiden Pol I transkription lopetusvaihe *Mus Musculus*:n lopetusvaiheella näytettynä. *Mus musculus*:en Pol I:n lopetusvaihetta on tutkittu nisäkkäistä eniten, mutta sen toimintamekanismi on samanlainen muun muassa ihmisellä. TTF-I sitoutuu DNA:han alavirtaan Pol I:n lopetusalueesta, ja PTRF sitoutuu Pol I:n ja TTF-I:n väliin. PTRF ei ole välttämätön Pol I:n transkription lopetukselle, mutta se edesauttaa transkription lopetusta

löydetty ainakin ihmisestä, rotasta, sammakosta ja hiirestä, joista viimeisimmässä sen toimintaa on tutkittu eniten. (Németh ja muut, 2013) TTF-I:n on osoitettu toimivan kromatiinirakenteen uudelleenjärjestämisessä, jolloin sen rooli ei ole pelkästään transkription pysäyttäjä geenin lopussa vaan myös geenin transkription määrää lisäävä (kuva 13 B) (Längst ja muut, 1997; Németh ja muut, 2008). Yhdessä TTF-I:n kanssa transkriptiota pysäyttämässä on myös Pol I transkriptin vapauttava tekijä (engl. *Pol I transcript release factor*, PTRF), joka toimii *in vitro* yhteistyössä myös hiivan Reb1 proteiinin kanssa (kuva 14). PTRF sitoutuu sekä TTF-I:een että Pol I:een, pysäyttäen Pol I:n ja RNA:n tuoton. (Jansa ja Grummt, 1999) PTRF on konservoitunut selkärangkaisissa, mutta se puuttuu leivinhiiivasta. Hiiren (*Mus musculus*) Pol I transkription lopetuksessa primäärisen pysähdyskohdan (T1) lisäksi useita TTF-I proteiineja on sitoutunut DNA:n alavirtaan (kohdat T2-T10) muodostaen transkription pysäyttämisen vikaturvan. PTRF toimii siltana Pol I ja TTF-I välissä, jolloin Pol I pysähtyy ja transkriptio loppuu. (Németh ja muut, 2013) Ihmisellä on samat transkription lopetustekijät, mutta sen mekanismia ei ole tutkittu tarkasti.

### 2.7 Ihmisen RNA-polymeraasi I lääkekehityksen kohteena

Pol I:n tuottaman rRNA:n tehokas transkriptointi on yksi solujen nopean kasvun vaatimus. Näin on myös hallitsemattomasti jakautuvilla syöpäsoluilla. Pol I:n toiminnan estäminen johtaa solusyklin pysähtymiseen ja lopulta apoptoosiin eli ohjattuun solukuolemaan. Apoptoosista on vastuussa kasvaimia hiljentävä proteiini p53, jota ylituotetaan Pol I:n pysähdyttyä. Käytännössä Pol I:n pysähdyksestä syntyvä nukleolaarinen stressi aloittaa p53:n, jonka määrä solussa suurenee Pol I:n toiminnan pysähdyttyä. (Drygin ja muut, 2011) Pol I:n toiminnan estäminen aiheuttaa syöpäsolujen spesifistä kasvun estämistä ja kuolemaa myös ilman p53:a, mikä on mahdollisten terapeuttisten sovellusten kannalta tärkeää, sillä yli 50 %:ssa kasvaimia mutatoituneen p53:n toiminta on estynyt (Peltonen ja muut, 2014; Perri ja muut, 2016). Näin ollen Pol I:n toiminnan tunteminen luo mahdollisuuden kehittää uusia syöpälääkkeitä. Syöpälääkkeille on kova tarve. Vuonna 2018 uusia syöpädiagnooseja todettiin 18 miljoonaa, ja syöpäkuolemia 9,6 miljoonaa (Bray ja muut, 2018). Pol I transkriptioon vaikuttavia lääkeaineita on kehitetty useita, joskin mikään käytössä olevista ei vaikuta suoraan Pol I:een (Ferreira ja muut, 2020). Ensimmäinen kliinisiin kokeisiin päässyt Pol I inhibiittori, jonka toimintaa ei tosin täysin tunneta, on CX-3543. CX-3543, toiselta nimeltään quarflokxiini (engl. *quarfloxin*) on pienimolekyylinen, joka estää nukleoliinin

vuorovaikutuksen rDNA:n kanssa, jolloin se inhiboi Pol I:n transkriptioaktiivisuutta. Tämä johtaa ohjattuun solukuolemaan. (Drygin ja muut, 2009) Sen biologinen hyötyosuus (engl. *bioavailability*) oli kuitenkin niin huono, ettei kliinisiä kokeita jatkettu (Ferreira ja muut, 2020). Hyvin pian CX-3543:n jälkeen kehitettiin Pol I spesifinen inhibiittori, CX-5461, joka on CX-3543 analogi. CX-5461 estää Pol I transkriptiota häiriten SL1:n ja promoottorin välistä vuorovaikutusta. (Drygin ja muut, 2011) SL1 on vastuussa hRRN3-Pol I:n ohjaamisesta rDNA:lle (Miller ja muut, 2001). SL1:llä voi olla useitakin toimintamekanismeja, sillä CX-5461:n on raportoitu vaikuttavan myös guaniini-emästoistoja sisältäviin DNA:han muodostuviin G-quadrupleksi-rakenteisiin. Nämä rakenteet estävät monia DNA:lla tapahtuvia prosesseja, mukaan lukien ihmisen rDNA:n transkriptiota. (Xu ja muut, 2017). CX-5461:n on osoitettu olevan myös topoisomeraasi-2 myrkky (Bruno ja muut, 2020). Joka tapauksessa CX-5461:n on osoitettu tappavan syöpäsoluja, ja se on nyt kliinisissä kokeissa (Drygin ja muut, 2011). Ensimmäisen vaiheen kliininen koe on suoritettu potilaille, joilla oli pahanlaatuisia hematologisia kasvaimia. CX-5461:n osoitettiin inhiboivan rDNA:n transkriptiota potilaiden kasvaimissa. Transkription väheneminen vaihteli 4 – 68,5 %:n välillä, mutta annostellun CX-5461:n määrällä ei näyttänyt olevan tähän vaikutusta välillä 25 – 250 mg / ml. (Khot ja muut, 2019) Pienmolekyylien lisäksi on kehitetty uusi menetelmä estää Pol I transkriptio 22-amiohappoa pitkän peptidin (22-pep) avulla. Tämä strategia perustuu Pol I – hRRN3 vuorovaikutuksen estämiseen, jossa 22-pep muistuttaa sitä osaa Pol I:n alayksiköstä A43, johon hRRN3 sitoutuu. Peptidin läsnäollessa hRRN3 ei sitoudu Pol I:een eikä transkription aloituskompleksi muodostu promoottorille, mikä estää rDNA:n transkription. (Rothblum ja muut, 2014) 22-pep:n kliiniset kokeet ovat suunnitteilla (Ferreira ja muut, 2020).

### *2.8 Työn tarkoitus*

Pol I:n transkription aloitusta on tutkittu sekä biokemiallisesti että rakenteellisesti, ja näiden havaintojen avulla on kyetty luomaan malli transkription aloitusmekanismille. Nykyinen malli on kuitenkin staattinen ja siitä puuttuvat erityisesti transkription aloituskompleksin kasaantumista kuvaavat kvantitatiiviset parametrit, esimerkiksi eri transkriptiotekijöiden sitoutumis- ja irtoamisnopeudet. On myös mahdollista, että mallissa on perustavanlaatuisia puutteita, jotka voivat paljastua vain tekemällä tarkempia, reaaliaikaisia mittauksia transkriptiokompleksin muodostumisesta ja alkuvaiheen transkriptiosta. Nämä kuitenkin mahdollistavat muun muassa biofysikaaliset

tutkimukset. Pro gradu –tutkielmani yleisenä tavoitteena oli kehittää biofysikaalisten mittausmenetelmien kokonaisuus Pol I:n transkription aloitusvaiheen tutkimiseksi, erityisesti hyödyntäen yksittäisten molekyylien havainnointia fluoresenssimikroskopian avulla. Ensimmäisenä osatavoitteenani on luoda toimivat Pol I aloituskompleksin proteiinien tuotto- ja puhdistusprotokollat. Toisena osatavoitteena on näiden puhdistettujen proteiinien toimivuuden osoittaminen biokemiallisin mittauksin. Kolmantena osatavoitteena on proteiinien hallittu leimaaminen fluoresenssileimoilla. Neljäntenä osatavoitteena on kehittää yksittäismolekyyli fluoresenssimikroskoopin käyttö- ja data-analyysiprotokollat. Viidentenä tavoitteena on selvittää ainakin osa transkription aloituskompleksin kasaantumisen kvantitatiivisista parametreista.

### 3. Materiaalit ja menetelmät

#### 3.1 Proteiinien leimapaikkojen suunnittelu

Käytettyjen rekombinanttiproteiinien kohdennetut leimapaikat suunniteltiin käyttäen Pymol -ohjelmistoa (DeLano, 2002) proteiinien visualisointiin sekä FPS-ohjelmaa (Kalinin ja muut, 2012) fluoresoivien leimamolekyylien liikeratojen ja mahdollisten steeristen esteiden visualisoimiseksi. Proteiinien rakennekuvat saatiin proteiinidatapankista (<https://www.rcsb.org/>), ja niiden proteiinidatapankkitunnisteet (PDB id) ovat 5n5z, 5n60, 5n61, 4c3h, 5oa1 ja 5m64. Toisena leimausstrategiana käytettiin SNAP-proteiiniin kovalenttisesti kiinnittyviä fluoresoivia leimoja, ja tätä SNAP-proteiinia päätettiin tuottaa CF:n alayksikkö RRN11:sta C-terminaalisisessä päässä. TBP:n leimaamiseen käytettiin aiemmin käytettyjä leimapaikkoja ja strategioita (Banik ja muut, 2001; Treutlein ja muut, 2012).

#### 3.2 Proteiinien tuotto ja puhdistus *E. coli*:ssa

##### 3.2.1 *RRN3* ja sen variantit

*S. cerevisiae*:n *RRN3* (YKL125W) geeni PCR-kloonattiin pET15-tuottoplasmiin (Novogen), johon oli lisätty C-terminaalinen TEV-10xHis -motiivi. Saadun *RRN3*-tuottoplasmin nimi oli pAM037. Proteiinia tuotettiin joko *Escherichia coli* T7-express LysY/lq -kannassa (New England Biolabs) tai B95.ΔA -kannassa (Mukai ja muut, 2015). Solut kasvatettiin LB-elatusaineessa (10 g/l tryptoni, 5 g/l hiivauute, 5 g/l NaCl) ravistelussa 37°C:ssa kunnes OD<sub>600</sub> ≈ 0,8, minkä jälkeen proteiinien ylituotto aloitettiin lisäämällä 0,2 mM IPTG ja soluja kasvatettiin huoneen lämmössä yön yli. Solut kerättiin sentrifugioimalla (6750 x g, 15 min, 4°C) ja pakastettiin -80°C. Kerätyt solut suspensoitiin puskuri A:han [50 mM HEPES-KOH (pH 7,8), 10 % glyseroli, 500 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM β-merkaptotetanoli], johon oli lisätty 0,2 % Tween-20, 20 mM imidatsolia, 0,12 mg/ml lysosyyymiä, 0,1 mM EDTA ja yksi proteaasi-inhibiittoritabletti (Roche, EDTA-vapaa). Solut hajotettiin sonikoimalla ja solujäte sentrifugioitiin 53000 x g, 1 h, 4°C. Supernatantti kaadettiin nikkelimatriisilla (matriisin tilavuus noin 1 ml, Ni Sepharose 6 Fast Flow, GE Healthcare) pakattuun PD-10 kolonniin (Bio-Rad), annettiin valua kolonnin läpi painovoimaisesti, pestiin 50 mM imidatsolia sisältävällä puskuri A:lla ja eluoiitiin 400 mM imidatsolia sisältävällä puskuri A:lla. Eluantti laimennettiin 1:3 puskuri B:llä [20 mM HEPES-KOH (pH 7,8), 5 % glyseroli, 5 mM MgCl<sub>2</sub>, 0,1 mM EDTA, 1 mM β-merkaptotetanoli] ja ladattiin kationinvaihtokolonniin (Resource Q 6 ml, GE Healthcare) 0,22 μM ruiskusuodattimen läpi sylinteripumpulla (virtausnopeus 1,5 ml

/ min). Resource Q-kolonne kiinnitettiin Äkta prime nestekromatografiin (GE Healthcare) ja kiinnittyneet proteiinit eluoiitiin 0,1 – 1,5 M / 100 ml NaCl-gradientilla virtausnopeudella 1 ml / min. RRN3 irtosi kolonnista 450 mM NaCl - pitoisuudessa. RRN3 konsentroidiin konsentraattorilla (Millipore, 3K MWCO) ja dialysoitiin puskuri C:ssä [20 mM HEPES pH (7,8), 50 % glyseroli, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 0,1 mM EDTA, 0,1 mM DTT] ja säilöttiin -20°C.

Kohdennettua leimausta varten luotiin RRN3-mutantit, jotka sisältävät epäluonnollisen aminohapon p-azido-L-fenyyialaniinin (pAzF). RRN3 N201pAzF variantin geeni rakennettiin käyttämällä megaprimer-strategiaa (Tyagi ja muut, 2004), missä AAT-kodoni (N201) korvattiin TAG-kodonilla. RRN3 N66pAzF tuotettiin synteettisellä fragmentilla (Eurofins Genomics). Molemmat mutaatiot rakennettiin samaan tuottoplasmiin kuin villityypin RRN3. Nämä RRN3 variantit tuotettiin B95.ΔA soluissa, joihin oli transformoitu pEVOL-pAzF plasmidi (Addgene # 31186; tuottaa tRNA/tRNA-syntetaasiparin, joka tunnistaa TAG-kodonin) (Chin ja muut, 2002). Kun OD<sub>600</sub> ≈ 0,8, aloitettiin mutantti RRN3:n tuotto lisäämällä LB-elatusaineeseen 1 mM pAzF luonnoton aminohappo, 0,2 % arabinoosia ja 1 mM IPTG ja soluja kasvatettiin pimeässä ja huoneen lämmössä yli yön. Solut kerättiin kuten edellä. pAzF -aminohapon sisältämä RRN3 putsattiin kuten natiivi-RRN3, mutta ilman pelkistäviä reagensseja (β-merkaptotetanolia ja DTT:tä) ja suojattuna valolta.

### 3.2.2 CF ja sen variantit

*S. cerevisiae*:n CF-proteiinin alayksiköt RRN7 (YJL025W) ja RRN11 (YML043C) PCR-kloonattiin muokattuun pET36b -tuottoplasmiin (Novogen), RRN7 kloonattiin restriktioentsyymeillä *NcoI* ja *SacI*, RRN11 kloonattiin restriktioentsyymeillä *NdeI* ja *NotI*. CF:n alayksikkö RRN6 (YBL014C) PCR-kloonattiin pET15b -tuottoplasmiin (Novogen) restriktioentsyymeillä *NcoI* ja *SacI*. Plasmidit transformoitiin *E. coli* T7-express LysY/lq -kantaan ja molemmat plasmidit sisältävät solut valikoitiin yhtäaikaaisesti ampicilliini (100 µg/ml) ja kanamysiini (30 µg/ml) selektiossa LB-maljoilla. CF tuotettiin ja putsattiin kuten RRN3, mutta käyttäen kationinvaihtokolonnin sijaan hepariinikolonnia (HiTrap Heparin HP 5 ml, GE Healthcare). CF eluotui 700 mM NaCl pitoisuudessa. CF dialysoitiin puskuri C:hen, jossa oli 1 M NaCl saostumisen estämiseksi.

Luonnottoman aminohapon sisältävän CF (RRN6 Y768pAzF) geeni rakennettiin käyttämällä synteettistä fragmenttia (Eurofins Genomics), jolla korvattiin *AdeI* ja *SacI*



restriktioentsyymien välinen alue villityypin RRN6-geenistä. Tuotto suoritettiin kuten RRN3:n luonnottoman aminohappojen sisältämien proteiinien kanssa, mutta käyttäen 0,5 mM pAzF. Putsaus suoritettiin kuten natiivin CF:n putsaus, mutta suojattuna valolta.

SNAP-fuusioproteiinin sisältävä CF (RRN11-SNAP) tuotettiin kloonamalla pSNAPf -plasmidista (New England Biolabs) SNAP-alue, joka yhdistettiin RRN11 C-terminaaliseen päästä linearisoituun plasmidiin NEBuilder HiFi DNA Assembly -kittiä käyttäen (New England Biolabs). SNAP-tagin sisältämä CF tuotettiin ja puhdistettiin kuten villityypin rekombinantti CF.

### 3.2.3 TBP ja sen variantit

*S.cerevisiae*:n natiivi TBP (SPT15/ YER148W) kloonattiin PCR:llä pET28b -tuottoplasmiin käyttäen restriktioentsyymeinä *NcoI* ja *XhoI* luoden plasmidin pAM027. TBP tuotettiin T7 express lysY/lq -soluissa kuten edellä. TBP putsattiin kuten RRN3, käyttäen puskuri A:n sijaan puskuri D:tä [50 mM Tris-HCl (pH 6,9), 5 % glyseroli, 500 mM NaCl, 1 mM  $\beta$ -merkaptotetanoli], puskuri B:n sijaan puskuri C:tä [50 mM Tris-HCl (pH 6,9), 5 % glyseroli, 0,1 mM EDTA, 1 mM  $\beta$ -merkaptotetanoli], nikkelimatriisin jälkeen käytettiin hepariinikolonnia (HiTrap Heparin HP 5 ml, GE Healthcare) ja hepariinikolonnin jälkeen anioninvaihtokolonnia (Resource S 6 ml, GE Healthcare). TBP eluoitui hepariinikolonnista 700 mM NaCl pitoisuudessa ja anioninvaihtokolonnista 350 mM NaCl pitoisuudessa. TBP konsentroidtiin kuten edellä ja dialysoitiin puskuri F:ssä [20 mM Tris-HCl (pH 6,9), 50 % glyseroli, 150 mM NaCl, 0,1 mM EDTA, 0,1 mM TCEP] ja säilöttiin -20°C.

TBP variantit (C78A S128C C164A; S61C C78A C164A; N-terminaalinen pAzF; C-terminaalinen pAzF) luotiin kaikki käyttäen synteettisiä fragmentteja (Invitrogen ja Eurofins Genomics) sekä NEBuilder HiFi DNA-assembly -kittiä (New England Biolabs). Villityypin TBP tuottoplasmi pAM027 katkaistiin restriktioentsyymeillä *NdeI* ja *HindIII*, ja tämä pala korvattiin suunnitellut mutaatiot sisältävällä kodonioptimoidulla synteettisellä geenipalalla. Yksittäisten kysteiinitähteiden sisältämät TBP-proteiinit tuotettiin ja putsattiin kuten villityypin TBP. pAzF:n sisältämät TBP-variantit tuotettiin kuten pAzF:n sisältämä CF ja putsattiin kuten villityypin TBP ilman pelkistäviä reagensseja ja valolta suojattuna.

### 3.3 Proteiinien tuotto ja puhdistus *S. cerevisiae*:ssa

#### 3.3.1 Pol I

Natiivi Pol I puhdistettiin joko valmiista *S. cerevisiae* hiivakannasta SC1613 (MAT $\alpha$  ura3-52 leu2-3,112 YPR110c::TAP-KIURA3, Euroscarf) tai itse luodusta hiivakannasta AMy008 (MAT $\alpha$  ura3-52 trp1 leu2- $\Delta$ 1 his3- $\Delta$ 200 pep4::HIS3 prb1- $\Delta$ 1.6R can1 RPA190-TEV-proteinA::kanMX6), jossa RPA190 -alayksikön C-terminaaliseen päähän on liitetty TEV-proteiini-A -puhdistuskahva. TEV-proteiini-A -puhdistuskahva eristettiin TAP-puhdistuskahvasta (Puig ja muut, 2001). AMy008 hiivakanta luotiin käyttäen homologista rekombinaatiota, vakiintuneita transformaatiomenetelmiä sekä antibiottimarkkerigeeniä kanMX6 (Gietz ja Schiestl, 2007; Goldstein ja McCusker, 1999; Wach ja muut, 1994). Solut fermentoitiin mukautetussa YPD-elatusaineessa (10 g/l peptoni, 5 g/l hiivauute, 2 % glukoosi), esikasvatukset tehtiin joko selektiivisessä minimielatusaineessa (1,7 g/l hiivatyypipohja ilman aminohappoja ja ammoniumsulfaattia (yeast nitrogen base without amino acids without ammonium sulfate, Alfa Aesar) 5 g/l ammoniumsulfaatti, 2 g/l hiiva-aminohappolisää (yeast synthetic drop-out medium supplement without tryptophan, histidine, leucine, uracil, Sigma Aldrich), 20 mg/l L-tryptofaani, 20 mg/l L-histidiini, 30 mg/l L-leusiini; SC-URA) tai selektiivisessä YPD-elatusaineessa (20 g/l peptoni, 10 g/l hiivauute, 2% glukoosi, 200  $\mu$ g/ml G418 (Sigma-Aldrich)). Hiivakantoja fermentoitiin (NLF22, Bioengineering) ja solut kerättiin logaritmisessa kasvuvaiheessa, jolloin Pol I on aktiivisin. Fermentointitilavuus oli yhteensä 60 litraa, josta soluja saatiin noin 150 g ja ne säilöttiin -80°C. Puhdistus aloitettiin vasaroimalla jäänyt hiivakimpale pienemmiksi paloiksi, jotka sulatettiin 150 ml:ssä puskuri G:tä (250 mM Tris-HCl pH 7,9, 20 % glyseroli, 250 mM ammoniumsulfaatti, 1 mM EDTA, 10 mM MgCl<sub>2</sub>, 10  $\mu$ M ZnCl<sub>2</sub>, 3 proteaasi-inhibiittoritablettia (Roche), 1 mM  $\beta$ -merkaptotetanolin). Solut hajotettiin helmivatkaimessa (engl. *bead beater*), jota viilennettiin jäällä. Toimintakierto oli 30 s päällä, 90 s tauolla, 3x10 min ajan ja väleissä 5 min tauot estämään vatkaimen ja solujen liiallista lämpenemistä. Solulysaatti erotettiin helmistä suodattamalla suodatinkankaan läpi (Miracloth, huokoskoko 22-25  $\mu$ m Millipore) Supernatantti kerättiin talteen ja sentrifugioitiin 30 min ajan 10 000 rpm (Beckman-Coulter JA-10 roottori), 4 °C. Supernatantti kerättiin ja sitoutettiin hepariinimatriisiin (10 ml Heparin Sepharose 6 Fast Flow, GE Healthcare) kevyesti pyörittämällä 1 h ajan 4°C lämmössä. Hepariinimatriisi sentrifugioitiin 1 min 500 g 4°C ja supernatantti kaadettiin pois. Hepariinimatriisi pestiin puskuri H:lla [50 mM Tris-HCl (pH 7,9), 20 % glyseroli, 1 mM EDTA, 10 mM MgCl<sub>2</sub>,

10  $\mu\text{M}$   $\text{ZnCl}_2$ , 250 mM ammonium sulfaatti, 1 mM  $\beta$ -merkaptetoetanol] kahdesti. Pol I (ja Pol III SC1613 kannasta puhdistettuna) eluoiitiin puskuri H:llä, jossa oli 1 M ammoniumsulfaatti. Eluoiu Pol I laimennettiin suhteessa 1:2 puskuri I:llä [50 mM Tris-HCl (pH 7,9), 20 % glyseroli, 1 mM EDTA, 10 mM  $\text{MgCl}_2$ , 10  $\mu\text{M}$   $\text{ZnCl}_2$ , 0,1 % Tween-20, 1 mM  $\beta$ -merkaptetoetanol] ja sidottiin IgG-matriisiin (1,5 ml IgG Sepharose 6 Fast Flow, GE Healthcare) kevyesti pyörittämällä 5 h 4°C lämmössä. Matriisi kaadettiin tyhjään Poly-Prep-kolonnein (BioRad) ja supernatantin annettiin valua kolonnin läpi painovoimaisesti. Matriisi pestiin 10 ml puskuri J:tä (50 mM Tris-HCl, 20 % glyseroli, 1 mM EDTA, 10 mM  $\text{MgCl}_2$ , 10  $\mu\text{M}$   $\text{ZnCl}_2$ , 0,05 % Tween-20, 250 mM ammoniumsulfaatti, 1 mM  $\beta$ -merkaptetoetanol) ja annettiin valua tyhjäksi. Kolonnin pohja suljettiin ja kolonnein lisättiin puskuri J:tä, jossa oli 200  $\mu\text{g}$  / ml TEV-proteaasia ja ravisteltiin kevyesti yön yli 4°C lämmössä. Resiini valutettiin painovoimaisesti kuivaksi ja kerättiin talteen, jonka jälkeen sitä pestiin vielä puskuri J:llä. Eluoiu proteiini laimennettiin puskuri K:lla (50 mM Tris-HCl pH 7,9, 0,5 mM EDTA, 1 mM  $\text{MgCl}_2$ , 10  $\mu\text{M}$   $\text{ZnCl}_2$ , 1 mM  $\beta$ -merkaptetoetanol) suhteessa 1:1 ja ladattiin kationinvaihtokolonnein (Resource Q 6 ml, GE Healthcare) 0,22  $\mu\text{m}$  ruiskusuodattimen läpi sylineripumpulla (virtausnopeus 1,5 ml / min). Resource Q -kolonnein kiinnitettiin Äkta prime -nestekromatografiin ja kiinnittyneet proteiinit eluoiitiin 125 – 1000 mM ammoniumsulfaattigradientilla virtausnopeudella 1 ml / min. Pol I eluoiu 250 mM ammoniumsulfaattipitoisuudessa. Kationinvaihdon jälkeen proteiini konsentroidiin, jaettiin eriin ja pikajäädettiin nestetyössä, minkä jälkeen proteiinit säilöttiin -80°C.

### 3.3.2 UAF

*S. cerevisiae*:n natiivin UAF-proteiinikompleksin tuottotaso on pieni (Ghaemmaghami ja muut, 2003), joten muodostettiin strategia UAF:n alayksiköiden ylituottamiseksi. Koska histoneita on solussa paljon enemmän kuin muita UAF:n komponentteja, kloonattiin UAF:n alayksiköistä vain RRN5, RRN9, RRN10 ja UAF30, kahteen hiivan tuottoplasmiin. RRN9 ja RRN10 PCR-kloonattiin BYP7586 hiivatuottoplasmiin (Ishii ja muut, 2014) ja UAF30 sekä synteettinen fragmentti RRN5-TEV-Proteiini-A (InVitrogen) BYP7588 hiivatuottoplasmiin. Proteiini-A on TAP-puhdistuskahvan ulommainen osa. Tuottoplasmit transformoitiin BY5625 ( $\text{MAT}\alpha$  ura3-52 trp1 leu2- $\Delta$ 1 his3- $\Delta$ 200 pep4::HIS3 prb1- $\Delta$ 1.6R can1) hiivakantaan yhdessä RRN5 $\Delta$ LEU2 poistofragmentin kanssa. Muodostuneessa AMy010 pitäisi histoneita lukuun ottamatta ylituottaa kaikki UAF:n alayksiköt ja puhdistuskahvallisen RRN5-proteiinin villityypin

version geeni on poistettu. UAF tuotettiin esikasvattamalla AMy010-hiivoja selektiivisessä minimielatusaineessa (SC-LEU/TRP/URA; kuten Pol I SC-URA, mutta ilman tryptofaania ja leusiinia), jonka jälkeen sitä tuotettiin fermentoimalla kuten Pol I. UAF puhdistettiin kuten Pol I, mutta IgG sitouttamisvaiheessa puskurin suola vaihdettiin ammoniumsulfaatista KCl:iin, ja IgG:n jälkeinen puhdistusvaihe suoritettiin SP-kationinvaihtomatriisilla (SP Sepharose 6 Fast Flow, GE Healthcare).

### 3.4 Proteiinien leimaus

Luonnottoman aminohapon sisältämät proteiinit leimattiin DyLight 650 fosfiinifluoroforilla (ThermoFisher Scientific, kataloginnumero 88911) valmistajan ohjeen mukaan Staudingerin ligaatiolla. Ylimääräinen väri poistettiin Zeba Desalting Spin -kolonneilla (7K MWCO, 0,5 ml, ThermoFisher Scientific). SNAP-proteiinit leimattiin TMR-star -fluoroforilla (New England Biolabs, kataloginnumero S9105) valmistajan ohjeen mukaan, mutta puskurina käytettiin PBS:n sijaan leimauspuskuria [20 mM HEPES-KOH (pH 7,8), 1 M NaCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT], sillä CF-SNAP leimaus ei onnistunut ilman Mg<sup>2+</sup> läsnäoloa ja CF sakkaantui alhaisemmissa suolapitoisuuksissa. Ylimääräinen väri poistettiin kuten edellä. Proteiinit leimattiin kysteiinitähteistä DyLight 550-maleimidifluoroforilla (ThermoFisher Scientific, kataloginnumero 62290) valmistajan ohjeen mukaan, ja ylimääräinen väri poistettiin kuten edellä. Leimattujen proteiinien konsentraatio laskettiin seuraavalla kaavalla 1:

$$c = (A_{280} - (A_{\max} \cdot KK)) / \epsilon_{\text{proteiini}} \quad (\text{Kaava 1})$$

Kaavassa  $A_{280}$  on absorbanssin arvo aallonpituudella 280nm,  $A_{\max}$  absorbanssin arvo leiman maksimiemissioaallonpituudella, KK on korjauskerroin, jolla korjataan leiman aiheuttama absorbanssi 280nm aallonpituudelle (jokaiselle leimalle ominainen) ja  $\epsilon_{\text{proteiini}}$  on kyseisen leimatun proteiinin ekstinktiokvatio. Leimaustehokkuus laskettiin seuraavalla kaavalla 2:

$$\text{Tehokkuus} = A_{\max} / (\epsilon_{\text{ff}} \cdot c_{\text{proteiini}}) \quad (\text{Kaava 2})$$

Kaavassa  $c_{\text{proteiini}}$  on leimatun proteiinin konsentraatio  $\epsilon_{\text{ff}}$  on fluoroforin ekstinktiokvatio.

### 3.5 Sitoutumiskokeet

#### 3.5.1 EMSA

Sitoutumiskokeita varten tuotettiin DY782-leimattuja (Eurofins) promoottoreita PCR-reaktiolla. Toinen PCR-alue sisälsi DY782-leiman ja toinen oli leimaamaton. Promoottorien pituus oli keskimäärin 550 emäsparia. Sitoutumiskokeissa TBP-, CF-, tai Pol I -proteiini sitoutettiin leimattuun DNA:han 20 min ajan huoneen lämmössä ennen geelille ajamista. Sitoutumisen jälkeisessä kilpailussa alkuperäisen promoottoriin sitouttamisen jälkeen lisättiin kilpaileva yhdiste (DNA tai hepariini) ja inkuboitiin vielä 20 min huoneen lämmössä ennen geelille lataamista. EMSA-näytteet ajettiin natiivi-PAGE geelillä [3 tai 4,5 % akryyliamidi/bis-akryyliamidi (19:1), 0,5x TBE (Invitrogen), 1 % glyseroli, 0,075 % APS, 0,086 % TEMED (ThermoFisher Scientific)] Kaikissa kokeissa käytettiin reaktiopuskuri A:ta [100mM HEPES-KOH (pH 7,8), 100 mM kaliumglutamaatti, 25µM ZnCl<sub>2</sub>, 5 % glyseroli, 2 mM magnesiumasetaatti, 0,1 mM EDTA] sekä latausvärinä bromofenolin sinistä. Geelit esiajettiin 10 mA 2 h, ja näytteitä ajettiin 10 mA kunnes sininen bromofenoli rintama oli geelin alaosassa. Geelit skannattiin Odyssey infrapunaskannerilla (LiCOR) ja analysoitiin Fiji -ohjelmalla (Schindelin ja muut, 2012).

#### 3.5.2 Anisotropia

Anisotropiamittauksia varten vaihteleva pitoisuus CF-proteiinia sitoutettiin 5 nM Pol I promoottoriin, jossa oli Atto425 leima (Eurofins). Promoottori tuotettiin PCR:llä, jossa toinen alue sisältää fluoroforin ja toinen ei. Sitoutuminen tapahtui reaktiopuskurissa 30 min ajan 30°C minkä jälkeen anisotropiaa mitattiin LS1-fluorometrillä (Perkin Elmer) 2 min ajan. Anisotropia laskettiin kaavalla 3:

$$A = (I_V - I_H) / (I_V + 2 \cdot I_H) \quad (\text{Kaava 3})$$

Kaavassa  $I_V$  on fluoresenssin intensiteettiarvo, kun valomonistinputki-detektorin edessä on pystysuoraan polarisoidun valon läpäisevä suodatin ja  $I_H$  on fluoresenssin intensiteettiarvo, kun detektorin edessä on vaakasuoraan polarisoidun valon läpäisevä suodatin. Viritysvalo polarisoitiin koko ajan pystysuoraan tasoon.

CF:n sitoutuminen Pol I promoottoriin laskettiin kaavalla 4:

$$A_{\text{havaittu}} = A_1 + (A_2 \cdot x^n / K_2) + ((A_0 \cdot K_1) / [CF]) / (1 + K_1 / [CF] + [CF]^n / K_2) \quad (\text{Kaava 4})$$

Kaava 2 perustuu oletukseen, että käytetyssä promoottori-DNA:ssa on yksi spesifinen ja useita (parametrin  $n$  mukainen määrä) identtisiä epäspesifisiä sitoutumispaikkoja CF:lle.  $A_0$  on vapaan promoottorin anisotropia,  $A_1$  on promoottorin anisotropia silloin, kun siihen on sitoutunut spesifiseen sitoutumispaikkaan yksi CF-molekyyli,  $A_2$  on promoottorin anisotropia silloin, kun siihen on sitoutunut suurin mahdollinen määrä CF-molekyyliä.  $K_1$  on spesifisen sitoutumispaikan dissosiaatiovakio.  $K_2$  on verrannollinen epäspesifisten sitoutumispaikkojen dissosiaatiovakio. Termi  $[CF]^n / K_2$  tuo sitoutumiskäyrän muotoon positiivista ko-operatiivisuutta. Käytettävissä oleva datan rajoissa ei ole mahdollista erottaa, johtuuko havaittu ko-operatiivisuus CF-molekyylien välisestä sitoutumisen edistämisestä vai promoottorin anisotropian epälineaarisesta muuttumisesta sitoutuneiden CF-molekyylien lukumäärän muuttujana.

### *3.6 Yksittäisten molekyylien fluoresenssimikroskopia*

Mikroskooppilasit (Thorlabs CG15XH1, paksuus  $170 \pm 5 \mu\text{m}$ ) pestiin 10 min 68 % typpihapossa, jonka jälkeen ne huuhdeltiin runsaalla vedellä ja pestiin vielä metanolilla ja kolmesti 100 % etanolilla. Puhtaat mikroskooppilasit silanoitiin kuten aiemmin on raportoitu (Chandradoss ja muut, 2014) käyttäen metanolin ja etikkahapon sijaan asetonia (3 % APTES asetonissa). Lasit pestiin asetonilla ja vedellä, jonka jälkeen ne kuivattiin typpikaasussa ja päällystettiin PEG-liuoksilla (Laysan Biotech) (Dulin ja muut, 2018). PEGyloidut lasit päällystettiin 0,5 mg/ml neutravidinilla, jonka jälkeen neutravidiniin linkitettiin 10 pM biotinyloitua Atto488-leimattua (Eurofins) tai Cy5-leimattua (Sigma Aldrich) promoottoria reaktiopuskuri B:ssä [100mM HEPES-KOH (pH 7,8), 100 mM kaliumglutamaatti, 25 $\mu\text{M}$  ZnCl<sub>2</sub>, 5 % glyseroli, 2 mM magnesiumasetaatti, 0,1 mM EDTA, 1 mM kysteamiinihydrokloridi, 0,2 mg/ml BSA].

Mikroskooppikokeet suoritettiin DeltaVision OMX v4 (GE Healthcare Life Sciences) -mikroskoopilla käyttäen Olympus APO N 60x Oil TIRF -objektiivia (immersioöljyn heijastusindeksi 1,516) ja etuvalaistua sCMOS-kameraa (resoluutio 2560 x 2160 pikseliä). Mikroskooppi tarkennettiin mikroskooppilasin pintaan, minkä jälkeen reaktiopuskuri B korvattiin kuvauspuskurilla (100mM HEPES-KOH, 100 mM kaliumglutamaatti, 25 $\mu\text{M}$  ZnCl<sub>2</sub>, 5 % glyseroli, 2 mM magnesiumasetaatti, 0,1 mM EDTA, 1 mM kysteamiinihydrokloridi, 0,2 mg/ml BSA, 0,5 mM TROLOX, 1 % glukoosi, 1 mM COT, 0,4  $\mu\text{g/ml}$  katalaasi, 1 mg/ml glukoosioksidaasi), jossa 0,5-5 nM kuvattavaa leimattua proteiinia.

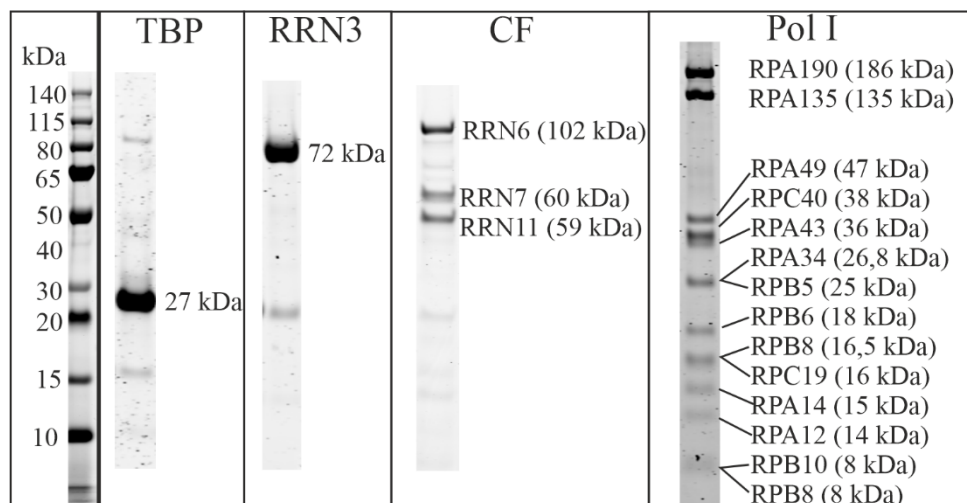
Mikroskoopilla nauhoitetut filmit käsiteltiin Fiji-ohjelmalla (Schindelin ja muut, 2012) muotoon, joka oli yhteensopiva iSMS ohjelman kanssa (Preus ja muut, 2015). iSMS:n avulla etsittiin ne promoottorimolekyylit, joihin tapahtui filmin aikana proteiinien sitoutumista. Molekyyliin signaali-intensiteetit laskettiin käyttäen aukkokuvamaskia (engl. *aperture photomask*). Kriteerit täyttävät molekyylit valittiin manuaalisesti iSMS-ohjelmassa. Hyväksymiskriteerit olivat: (a) fluoroforin signaali pystyttiin luotettavasti erottamaan naapurimolekyyliin signaalista, (b) fluoroforin signaalitaso oli mittaolosuhderiippuvaisen alarajan ja ylärajan välillä ja (c) fluoroforin signaalitaso pysyi vakaana. iSMS:stä saadut tiedostot muokattiin käyttökelpoiseen muotoon kahdella excel-makrolla, jotka kirjoitin (liite 2) Sitoutumistapahtumat, eli sitoutumisen alkaminen, loppuminen ja kesto, erotettiin signaaliradasta (engl. *trajectory*) yhteistyökumppanimme Subhas Chandra Beran (Interdisziplinäres Zentrum für Klinische Forschung, David Dulin laboratorio, Erlangen, Saksa) tekemällä Python-perusteisella ohjelmalla (BPF), joka hyödyntää signaaliprosessointikoodia (Truong ja muut, 2020). Sitoutumistapahtumat haettiin alhaalta ylöspäin segmentoinnilla (engl. *bottom-up segmentation*) käyttäen maksufunktiona (engl. *cost function*) Gaussianista prosessinmuutosta (engl. *Gaussian process change*). Sakkoarvoksi (engl. *penalty value*) asetettiin 150 ja hypyksi (engl. *jump*) 1. Saadut sitoutumistapahtumien kestoajat analysoitiin kineettisesti sekä Origin-ohjelmistolla (OriginLab Corporation) että yhteistyökumppanimme Subhas Chandra Beran (Interdisziplinäres Zentrum für Klinische Forschung, David Dulin laboratorio, Erlangen, Saksa) tekemällä Python-perusteisella ohjelmalla, joka hyödyntää signaaliprosessointikoodia (Truong ja muut, 2020).

## 4. Tulokset

### 4.1 RNA-polymeraasi I:n transkription aloituskompleksin proteiinien puhdistaminen

#### 4.1.1 Proteiinien puhdistaminen

Pol I aloituskompleksin tutkiminen *in vitro* vaatii, että avustavat proteiinit ja itse Pol I on puhdistettava. Yksialayksikköisille RRN3:lle ja TBP:lle sekä kolmealayksikköiselle CF:lle tuotto- ja puhdistusstrategiaksi valittiin kolibakteerissa (*E.coli*) tuottaminen rekombinanttiproteiineina. Tällöin proteiineja voidaan ylituottaa, jolloin niitä saadaan mahdollisimman paljon mahdollisimman pienellä vaivalla. Yksialayksikköiset RRN3 ja TBP tuotettiin yhden plasmidin avustuksella, mutta kolmealayksikköinen CF tuotettiin kahden plasmidin avustuksella. Kaikkien proteiinien puhdistusstrategia alkoi samoilla vaiheilla, solujen hajottamisella ja proteiinin kiinnittämällä nikkelipylvääseen. Tämä onnistuu siksi, että jokainen proteiini sisältää histidiinihännän, joka sitoutuu nikkeliin. Tämän jälkeen RRN3 puhdistettiin yksivaiheisella kationinvaihtopuhdistuksella, CF yksivaiheisella hepariinipylväspuhdistuksella ja TBP kaksivaiheisesti ensin hepariinipylväällä ja sitten anioninvaihtopylväällä. Kaikkien proteiinien puhdistusta arvioitiin Coomassie-värjättyillä SDS-PAGE geeleillä. RRN3, TBP ja CF onnistuttiin puhdistamaan hyvin (kuva 15), mutta RRN3 ja CF tapauksessa proteiinimäärät olivat paljon pienemmät. RRN3 saatiin 540 µg (7,5 nmol), CF 330 µg (1,5 nmol) ja TBP 10,3 mg (380 nmol). Koska RRN3:n ja CF:n tuottoskaalan kasvattaminen on kohtuullisen



**Kuva 15.** Coomassie-värjätty SDS-PAGE-kuvat Pol I aloituskompleksin puhdistuksista. TBP-kuva on hepariinipuhdistusvaiheesta, ja TBP puhdistettiin vielä kerran anioninvaihtokromatografialla, jolloin siitä saatiin vielä puhtaampaa. TBP on melko puhdasta jo hepariinipuhdistuksen jälkeen. CF kuva on myös hepariinipuhdistuksesta, ja kaikki kolme alayksikköä on nähtävissä. Jonkin verran epäpuhtauksia löytyy. RRN3 puhdistui melko puhtaaksi kationinvaihtokromatografialla, pieniä epäpuhtauksia on nähtävissä. Pol I puhdistui, vaikka kaikkia alayksiköitä ei ole nähtävissä selvästi tässä geelikuvassa. Kuvat eivät ole keskenään vertailukelpoisia, sillä niiden havainnointiin on käytetty skannausvaiheessa eri laserin intensiteettejä sekä kuvankäsittelyvaiheessa kirkkautta ja kontrastia on muutettu eri tavoin proteiinien havainnoimiseksi



yksinkertaista, niitäkin voidaan käyttää suhteellisen paljon proteiinia vaativissa biokemiallisissa kokeissa, esimerkiksi EMSA- ja anisotropiakokeissa.

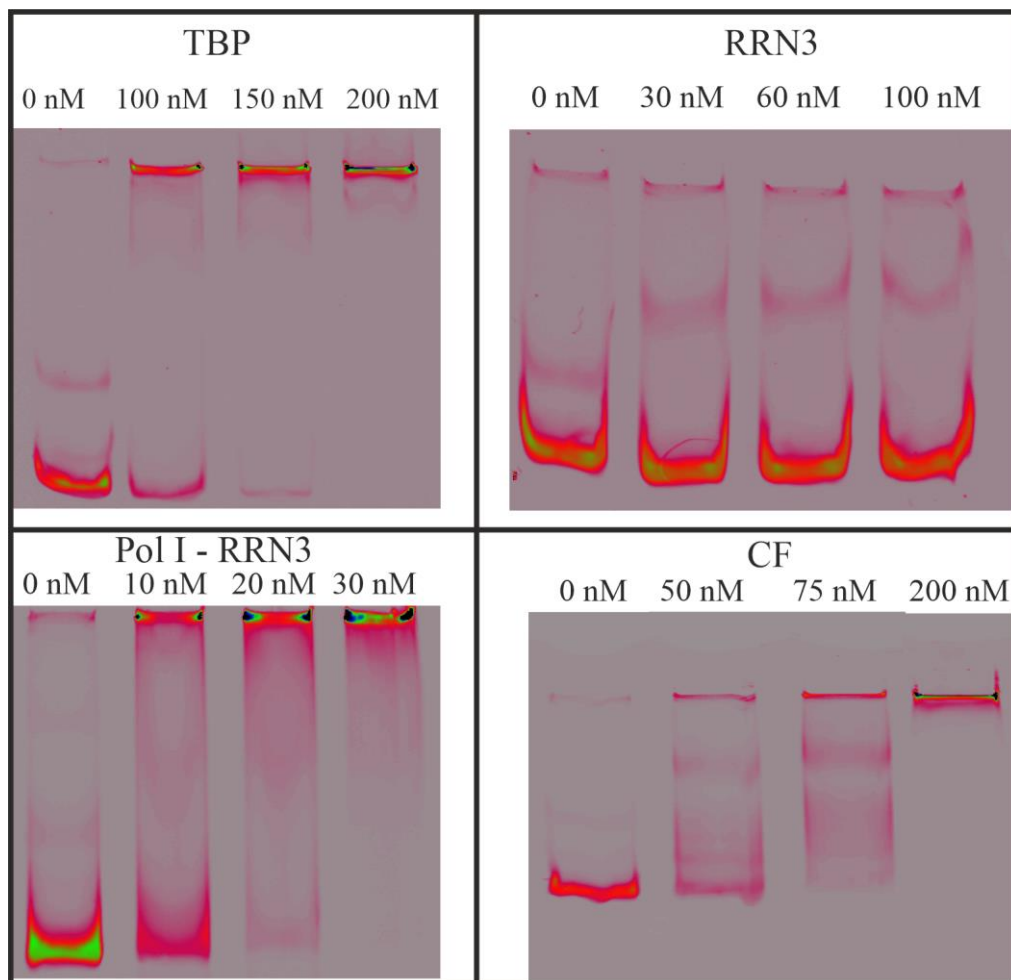
Koska Pol I sisältää 14 alayksikköä ja UAF, näiden proteiinien tuotto ja puhdistus tehtiin leivinhiiivassa. UAF:n puhdistukseen käytettiin itse luotuja hiivakantoja AMy007 ja AMy010 (BY5625 johdannaisia), joissa molemmissa UAF:n alayksiköt histoneita lukuun ottamatta ylituotetaan tuottoplasmideista. AMy010 eroaa AMy007:tä vain sillä, että AMy010-kannassa hiivan oma RRN5-geeni on poistettu. Plasmidin koodaamassa RRN5-proteiinissa oli proteiini A-puhdistuskahva, jolloin RRN5:n poistaminen hiivan genomista johtaa vain tämän puhdistuskahvan sisältävän UAF-proteiinin tuottoon. Jostain syystä en onnistunut puhdistamaan UAF:ää; kummastakaan ylituottokannasta ei ollut havaittavissa UAF:ää useiden tehtyjen puhdistuksien jälkeen. Proteiinin tuottotaso saattoi olla hyvin alhainen tai ylituotettu proteiini ei ollut liukoinen. Koska proteiini A:lle tai UAF:n alayksiköille ei ole kehitetty spesifiä kaupallista vasta-ainetta, UAF:n tuottamista ei voitu varmistaa Western-analyysillä. UAF hiivakantoja tuotettiin ja proteiinia puhdistettiin yhteensä kuusi kertaa, mutta mikään puhdistus ei tuottanut tulosta. Koska UAF:n tuottaminen ja puhdistaminen hiivasoluista ei toiminut, RET-laboratorion Olena Parilova sovelsi rekombinanttiproteiinin tuottomenetelmää kolibakteerissa (Smith ja muut, 2018) ja käytin lopulta tätä puhdasta proteiinia tässä työssä.

Pol I puhdistusstrategiana käytettiin TAP-puhdistusta, jossa Pol I / Pol III alayksikössä RPC40 on TAP-puhdistuskahva (Puig ja muut, 2001). Pol I puhdistukseen kehitettiin myös oma hiivakanta AMy008, jossa TAP-puhdistuskahvan proteiini A-osa on liitetty RPA190-alayksikön C-terminaaliseen päähän, mutta tätä hiivakantaa ei ehditty käyttää Pol I puhdistukseen. Pol I osoittautui haastavaksi puhdistaa, mutta siinäkin onnistuttiin (kuva 15). Yhteen puhdistuskertaan yhdistettiin 150 g fermentoitua hiivaa, joka oli kerätty eksponentiaalisessa kasvuvaiheessa. Tämä tarkoitti kolmea 20 litran fermentointia. Puhdistetun Pol I:n pitoisuus oli kuitenkin niin pieni (70 µg; 120 pmol), että sen käyttö kvalitatiivissa biokemiallisissa kokeissa päätettiin jättää hyvin vähälle.

#### 4.1.2 Proteiinien karakterisointi

Elektroforeettisella liikkuvuussiirtymäkokeella (engl. *electrophoretic mobility shift assay*, EMSA) pyrittiin osoittamaan puhdistettujen proteiinien toimivuus, tässä tapauksessa niiden sitoutuminen DNA:han. EMSA:ssa ei-denaturoivalla geelillä etenevä vapaa leimattu DNA etenee nopeammin kuin DNA, johon on sitoutunut proteiini. Tällöin

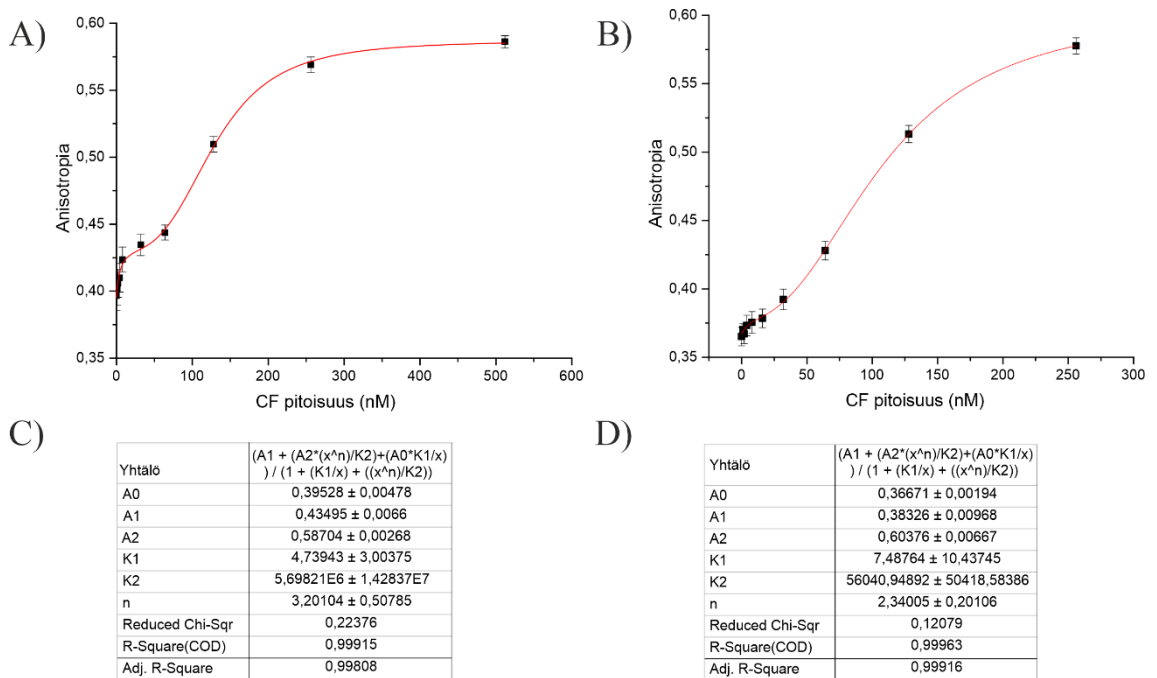
geelillä nähdään siirtymä. EMSA valittiin sen yksinkertaisuuden ja suhteellisen pienen proteiininkulutuksen vuoksi. EMSA:a hyödyntäen tarkistettiin kvalitatiivisesti, sitoutuuko Pol I, RRN3 ja CF itsenäisinä proteiineina Pol I promoottoriin. Tämän lisäksi tutkittiin, sitoutuuko Pol I-RRN3 kompleksi Pol I -promoottoriin. Myös TBP perustoiminta karakterisoiitiin käyttämällä Pol I -promoottoria, vaikka TBP:n rooli Pol I aloituskompleksissa on vielä epäselvä. CF sitoo kaiken promoottorin, kun sitä on 30–40 kertainen ylimäärä promoottoriin (5 nM) verrattuna (kuva 16). RRN3 ei sitoudu promoottoriin. Pol I sitoo kaiken promoottorin, kun sitä on 12-kertainen ylimäärä (data ei näkyvissä). TBP sitoo kaiken Pol I promoottorin, kun sitä on 40-kertainen ylimäärä. Sitoutumisen jälkeisessä kilpailussa CF ja TBP irtosivat Pol I promoottorilta, mikä viittaisi siihen, ettei niillä ole spesifisyyttä Pol I promoottoriin (data ei näkyvissä). Valitettavasti missään tapauksessa ei nähty geelillä tapahtuvaa siirtymää, vaan DNA:n sitonut proteiini jäi jumiin geelikaivoon. Tämä voi viitata siihen, että yhteen promoottori-



**Kuva 16.** Kokoelma EMSA-kuvia. TBP, CF ja Pol I - RRN3 sitoutuvat Pol I promoottoriin, mutta RRN3 yksinään ei. Näiden tulosten avulla pystyttiin jatkamaan seuraaviin kokeisiin. Pol I yksinään sitoutuu myös promoottoriin (data ei näkyvissä).

DNA-molekyyliin sitoutuu enemmän kuin yksi proteiinimolekyyli ja että muodostunut kompleksi on liian iso kulkeutuakseen geeliin.

Koska CF sitoutui Pol I promoottoriin ja sen tuottotaso oli kohtuullinen, suoritettiin CF:lle vielä muutama kvalitatiivinen koe, joissa tarkasteltiin CF:n yleistä taipumusta sitoutua DNA:han. CF:n sitoutumiskykyä Pol II -promoottorin sisältävään DNA:han ja satunnaiseen ei-promoottori DNA:han tutkittiin sekä EMSA:lla että anisotropia-kokeilla (kuva 17). Anisotropiassa mitataan fluoresenssileimatun molekyylin, tässä tapauksessa DNA:n, pyörimisnopeutta hyödyntäen polarisoitua viritys- ja emissiovaloa (kaava 3, materiaalit ja menetelmät -osa). Jos leimattu molekyyli on pieni, se pyörii nopeasti ja mitattu anisotropia on pieni. Jos leimattuun molekyyliin sitoutuu iso proteiinimolekyyli, muodostuneen kompleksin pyörimisnopeus hidastuu ja havaittu anisotropia nousee. EMSA-kokeiden perusteella CF ei erottele DNA:ta, johon se sitoutuu (dataa ei näytetty). Tällöin myös kahden eri epäspesifisen DNA:n eli Pol II promoottorin ja satunnaisen DNA:n käyttäminen turhaa, jolloin päädyttiin käyttämään satunnaista DNA:ta. Anisotropiatulosten mukaan CF ei erottele DNA:ta, mutta pienessä pitoisuudessa (<50 nM, kun DNA:n pitoisuus 5 nM) CF sitoutuu sekä promoottoriin että satunnaiseen



**Kuva 17.** Anisotropiakokeiden tulokset. **A)** Promoottorin anisotropia kasvavassa CF:n pitoisuudessa. Pienellä CF:n pitoisuudella (<50 nM) CF vaikuttaisi sitoutuvan spesifisti. **B)** Satunnaisen DNA:n anisotropia kasvavassa CF:n pitoisuudessa. Kuten promoottorilla, myös satunnaisella DNA:lla CF vaikuttaisi sitoutuvan tiukasti pienillä CF:n pitoisuuksilla (<50 nM). **C)** Käyrän sovituksen parametrit promoottorin anisotropialle kasvavassa CF:n pitoisuudessa. Vähennety Chi-neliön arvo (Reduced Chi-Sqr) kertoo, että käyrä on alisovitettu, eli dataa vaaditaan enemmän käyrän parempaan sovitukseen. Täydellinen sovitus on vähennety Chi-neliön arvolla 1. **D)** Käyrän sovituksen parametrit satunnaiselle DNA:lle kasvavassa CF:n pitoisuudessa.

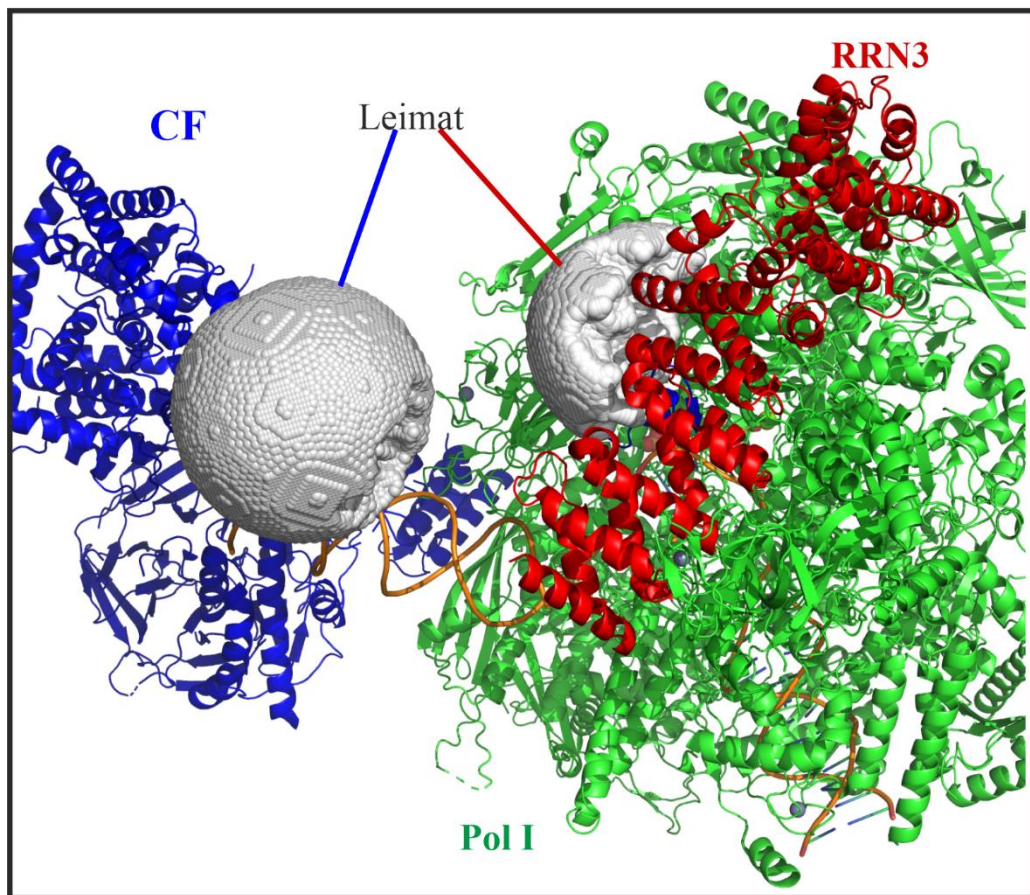
DNA:han tiukemmin. Suuremmilla pitoisuuksilla useita CF-molekyylejä sitoutuu epäspesifisesti sekä promoottori-DNA:han että satunnaiseen DNA:han (kuva 17). Vaadittavat määrät proteiinia ja leimattua DNA:ta ovat hyvin suuria anisotropiakokeiden suorittamiseen tarvittavalla toistomäärällä, joten tästä strategiasta päätettiin luopua. Kuvan 17 dataa ja tehtyjä päätelmiä on siten pidettävä kvalitatiivisina ja alustavina.

#### *4.2 Proteiinien käytön suunnittelu reaaliaikaisiin sitoutumiskokeisiin*

EMSA:n ja anisotropian sekä muiden harkittujen menetelmien, kuten stopped flow, heikkous on suuret näytemäärät. Hiivassa tuotettua Pol I:tä täytyisi tuottaa massiivisesti, jotta sen saisi riittämään kvantitatiivisiin EMSA, anisotropia tai stopped flow -mittauksiin. Toisaalta EMSA- (kuva 16) ja anisotropiatulokset (kuva 17) viittaisivat siihen, että ainakin yksittäisten Pol I -koneiston proteiinien kohdalla ero spesifisessä ja epäspesifisessä DNA:han sitoutumisessa ei ole kovin suuri. Tutkittava sitoutumisreaktio on siten aina hyvin todennäköisesti heterogeeninen koostuen rinnakkain tapahtuvasta spesifisestä ja epäspesifisestä DNA:han sitoutumisesta. Jotta olisi mahdollista erottaa aloituskompleksin proteiinien sitoutumisjärjestys promoottoriin, on sitoutumista tutkittava reaaliajassa ja erityyppiset sitoutumistyytit on pystyttävä erottamaan toisistaan. Tästä syystä päätin siirtyä käyttämään yksittäismolekyylifluoresenssi-tekniikkaa. Se minimoi näytteiden kulutuksen ja mahdollistaa erilaisten reaktiotyyppien havaitsemisen samassa näytteessä. Yksittäisten molekyylien kuvaaminen reaaliajassa vaatii molekyylien leimaamisen fluoroforeilla. Niinpä tutkin mahdollisia leimausstrategioita, aluksi keskittyen mahdollisuuteen hyödyntää FRETiä. Fluoroforien etäisyys toisistaan tehokkaan FRET reaktion saavuttamiseen voidaan määrittää (Vogel ja muut, 2014), mutta yleensä 50 % FRET-tehokkuus saavutetaan, kun fluoroforit ovat toisistaan 42-68 Å (4,2 – 6,8 nm) etäisyydellä (Lam ja muut, 2012). Proteiinien leimapaikat on siis suunniteltava siten, että ne ovat noin 50 Å toisistaan, kun Pol I aloituskompleksi on kasautunut. Leimapaikkojen suunnittelu näin tarkasti vaatii proteiinien rakennekuvien selvittämistä, jotta voidaan arvioida leimojen sijaintia ja näiden niiden välistä mahdollista vuorovaikutusta. Pol I:n, CF:n, RRN3:n sekä näiden kaikkien muodostaman välttämättömän aloituskompleksin rakenteet on selvitetty (Engel ja muut, 2013; Engel ja muut, 2017; Fernández-Tornero ja muut, 2013; Sadian ja muut, 2019). Näitä rakennekuvia tarkasteltiin Pymol-ohjelmalla, ja päädyttiin proteiinien kohdennettuun leimaukseen tietyistä aminohapoista. Proteiineiksi valittiin CF ja RRN3, sillä niiden

tuottaminen kolibakteerissa on mahdollista. Mutatoidessa proteiineja niiden tuottotaso yleensä heikkenee eikä Pol I:n heikkoa tuottotasoa haluttu enää pienentää.

CF:n ja RRN3:n rakenteista etsittiin alueita, joissa ei ollut laskostunutta rakennetta, kuten  $\beta$ -levyjä tai  $\alpha$ -heliksiä. Kun sopivat alueet löytyivät, tarkastettiin niiden mahdolliset kontaktit muihin rakenteen aminohappoihin. Kun sopivat alueet löytyivät, simuloitiin leimojen paikkoja hyödyntäen yhdessä Pymol-ohjelmiston kanssa FPS-ohjelmistoa. FPS-ohjelmisto simuloi leima-aineen liikettä, näyttäen kaikki mahdolliset konformaatiot jossa leima voi olla minä tahansa ajan hetkenä. Leimojen simulaatiot toteutettiin välttämättömän aloituskompleksin kanssa, jotta mahdolliset konformaatiomuutokset sitoutumisesta johtuen olisi näkyvissä. Näiden tulosten perusteella valittiin kaksi



**Kuva 18.** Kohdennettujen leimapaikkojen suunnittelun visualisointi. Visualisointiin käytettiin Pymol ohjelmaa ja leimojen liikkeen simuloimiseksi FPS-ohjelmaa (Kalinin ja muut, 2007). Pol I on vihreällä, CF sinisellä ja RRN3 punaisella. Valkoiset pallot edustavat leimojen mahdollista liikettä minä tahansa ajan hetkenä. Leimapaikoiksi valittiin RRN3:n N201 ja RRN6:n Y768, ja ne ovat noin 50 Å päässä toisistaan. Suunnitteluun ja visualisointiin käytettiin pohjarakenteena rakennetta, jonka Engel ja muut (2007) julkaisivat (PDB id 5N61).

leimapaikkaa, joilla on mahdollisuus FRET-reaktion niiden sitoutuessa aloituskompleksiksi: CF:n alayksikön RRN6 Y768 sekä RRN3:n N201 (kuva 18).

Valittujen aminohappojen jälkeen piti vielä valita leimausstrategia, joista ensin tarkasteltiin maleimidi-tekniikkaa, jossa maleimidiin liitetty fluorofori kiinnittyy kovalenttisesti kysteiniitähteeseen. Tämä tekniikka vaatisi kaikkien proteiinin muiden kysteiniin poistamista mutaatioilla. RRN3 sisältää 10 kysteiniä ja CF 21, joista osa sijaitsee laskostuneissa rakenteissa. Näin ollen maleimidi-tekniikan käyttö RRN3:n ja CF:n kohdalla ei ole käytännöllistä. Sen sijaan TBP sisältää vain kaksi kysteiniä, joten maleimidi-tekniikkaa käytettiin TBP:n leimaamisen kanssa, kuten aiemminkin on tehty (Banik ja muut, 2001; Treutlein ja muut, 2012). RRN3:n ja CF:n kohdalla päädyin lopulta käyttämään laajennettua geneettistä koodia leimausstrategiana. Laajennetun geneettisen koodin käyttö perustuu pysäytyskodonien (TAG, TAA ja TGA) uudelleen ohjelmoimiseen niin, että niiden avulla ribosomilla syntyvään proteiiniketjuun voidaan liittää luonnottoman aminohappo (engl. *unnatural amino acid*). Laajennetun geneettisen koodin hyödyntämiseen on monia strategioita (Kim ja muut, 2013). Valitsin proteiiniin liitettäväksi luonnottomaksi aminohapoksi 4-azido-L-fenyylialaniinin (pAzF), joka reagoi Staudingerin ligaatiolla proteiiniystävällisissä olosuhteissa fosfiiniryhmän sisältävien fluoroforien kanssa (Chakraborty ja muut, 2015). Tässä strategiassa korvataan halutun leimapaikan kodoni TAG pysäytyskodonilla. Tuotettavan proteiinin DNA-sekvenssin on siis päätyttävä joko TAA tai TGA pysäytyskodoneihin, jotta proteiinisynteesi loppuu oikeaan kohtaan. Tämän lisäksi tuotto-organismiin, tässä tapauksessa kolibakteeriin, on lisättävä TAG-kodonin tunnistava tRNA/tRNA-syntetaasipari, joka liittää kasvavaan proteiiniketjuun luonnottoman aminohapon. Tällainen plasmidi on kaupallisesti saatavana (Addgene #31186; (Chin ja muut, 2002)), jolloin sen on oltava yhdessä proteiinia tuottavan plasmidin kanssa saman solun sisällä. Näiden lisäksi on solujen elatusaineeseen lisättävä kohdeproteiinin tuottovaiheessa pAzF:ää. Tehokkaan proteiinituoton esteeksi voi vielä muodostua kolibakteerin oma TAG-kodonin tunnistava vapautustekijä 1-proteiini (engl. *release factor 1*, RF-1), joka lopettaa proteiinisynteesin. Vuonna 2013 Lajoie ja muut kehittivät kolibakteerikannan C-321, josta kaikki genomien 321 TAG-kodonia ja RF-1:n geeni oli poistettu. C-321 on siitä lähtien ollut suurella käytössä luonnottomien aminohappojen hyödyntämisessä, sillä luonnottomat aminohapot liitetään tarkasti proteiiniketjun TAG:tä vastaavaan sijaintiin. Valitettavasti C-321 kannan mutaatiot myös heikensivät sen kasvunopeutta 60 %:a (Lajoie ja muut, 2013). Mukai ja muut onnistuivat vuonna 2015 tuottamaan kolibakteerikannan B-95.ΔA:n, joka ei tunnista TAG-kodoneja mutta jonka kasvunopeus ei heikentynyt (Mukai ja muut, 2015). Valitsin tämän kannan pAzF:n liittämiseksi

proteiineihin. Tuotin myös villityypin RRN3:a ja CF:ää B95.ΔA soluissa mahdollisten tuottotaserojen havaitsemiseksi.

#### *4.3 Mutatoitujen proteiinien tuotto ja karakterisointi*

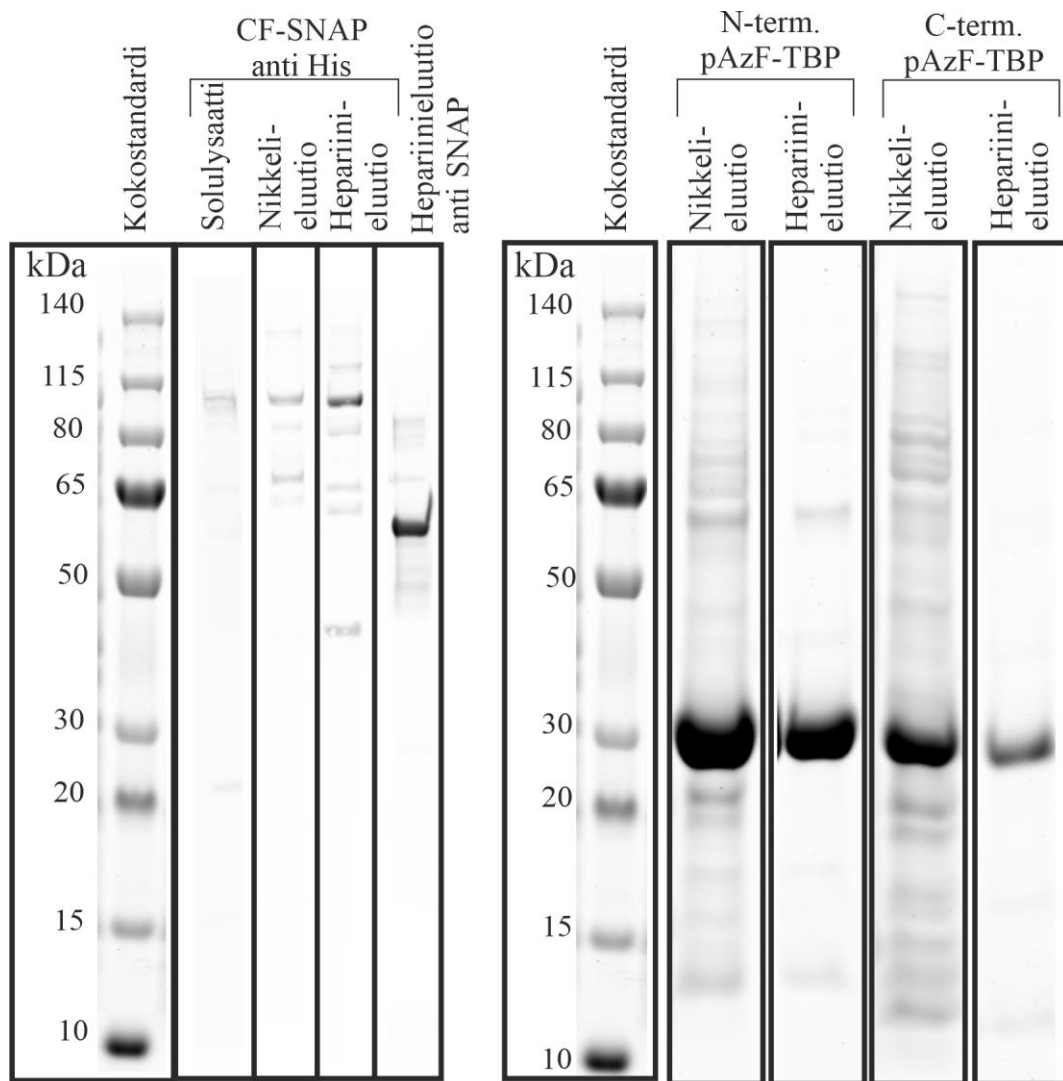
Yksittäisen kysteeinitähteen sisältämät TBP-variantit tuotettiin kuten villityypin TBP, mutta saanto oli paljon heikompi. S128C -yksittäiskysteinimuotoa saatiin 500 µg (17 nmol) ja S61C-yksittäiskysteinimuotoa 150 µg (5,1 nmol). pAzF:n sisältämiä RRN3:a ja CF:ää tuotettiin vain ¼ villityypin proteiinin tuottokasvatusten tilavuudesta materiaalien riittävyyden takaamiseksi. pAzF:n sisältämää CF tuottui melko huonosti ja sitä saatiin vain 38 µg (170 pmol). pAzF:n sisältämää RRN3a saatiin enemmän, 470 µg (6,5 nmol). pAzF:n sisältämien proteiinien toimintaa ei vielä tässä vaiheessa testattu, sillä niiden leimausreaktiot päätettiin suorittaa ensin.

Koska pAzF:n sisältämän CF:n tuottotaso oli todella heikko, päätettiin sen leimausstrategiaa vaihtaa SNAP-proteiinileimaukseen. Tässä strategiassa proteiinin päähän, yleensä sen C-terminaaliseen päähän, liitetään muutaman linkkeriaminohapon avulla SNAP-proteiini, jonka koko on 19,4 kDa. SNAP on entsyymi, joka liittää luonnollista substraattia muistuttavan fluoroforin kovalenttisesti aktiivisen keskuksen aminohappoonsa. Tämän strategian hyöty on se, että ylimääräistä plasmidia tai pAzF:ää ei tarvita. Haittana on se, että leimaus ei ole enää samalla tavalla tiettyyn pistemäiseen paikkaan kohdennettua. Yksittäismolekyylimikroskopiassa voidaan käyttää myös yhteislokalisaatiota (engl. *co-localization*), jossa pyritään esimerkiksi havaitsemaan leimatut proteiinit leimatulla DNA:lla vailla leimojen keskinäistä vuorovaikutusta. Yhteislokalisaatiolla ei kuitenkaan voida havaita proteiinien asemaa toisiinsa nähden. Päätin keskittyä tähän strategiaan FRET-kokeiden sijaan.

CF-SNAP fuusioproteiini tuotettiin kuten villityypin CF. CF-SNAP tuottui hivenen paremmin kuin villityypin CF ja sitä saatiin 500 µg (2,3 nmol). CF-SNAP:in puhdistumisen onnistuminen pystyttiin myös varmistamaan western-analyseilla. Coomassie-värjätty geeli osoitti kaikkien alayksiköiden läsnäolon (data ei näkyvissä). Western-analyysi osoitti erikseen histidiinihännällisen RRN6-alayksikön ja SNAP:n sisältämän RRN11-alayksikön. (kuva 19).

Tämän lisäksi halusin selvittää, huonontaako pAzF:n lisääminen hyvin tuottuvan proteiinin tuottoa. Proteiiniksi valittiin parhaiten tuottuva TBP, jonka N- ja C-

terminaaliin päihin TAG-kodoni, jolloin itse proteiinista ei korvattu mitään. TBP tuotettiin pAzF:n läsnäollessa ja puhdistettiin kuten villityypin TBP. TBP, jonka N-terminaalisessa päässä oli pAzF, tuottui todella hyvin ja sitä saatiin 6,9 mg (250 nmol) (kuva 19). C-terminaalisesta pAzF:n sisältämä TBP tuottui huonommin ja sitä saatiin 950 µg (33 nmol). Täten pAzF:n käyttö proteiinien tuotossa ei välttämättä heikennä proteiinien tuottotasoa. Luonnottoman aminohapon sijainti keskellä proteiinia saattaa vaikuttaa esimerkiksi sen laskostumiseen ja siten tuottotasoihin.



**Kuva 19.** CF-SNAP western-analyysi ja UAA-TBP:n coomassie-analyysi. CF-SNAP solulysaatissa on nähtävissä histidiinihännällinen RRN6 ja tämä havaittu viiva vahvistuu puhdistusvaiheiden edetessä. Viimeisestä vaiheesta ajettuna SNAP-proteiinin tunnistavassa western-analyysissä havaittiin myös SNAP-proteiinin sisältämä RRN11. pAzF:n sisältämien TBP-proteiinien puhdistus onnistui hyvin, havainnollistettuna ensimmäinen ja toinen puhdistusvaihe. C-terminaalisesta pAzF:n sisältämää TBP:tä hävisi jonkin verran nikkelieluution ja hepariinieluution välissä, mutta N-terminaalisesta pAzF:n sisältämää TBP:tä ei juurikaan hävinnyt.

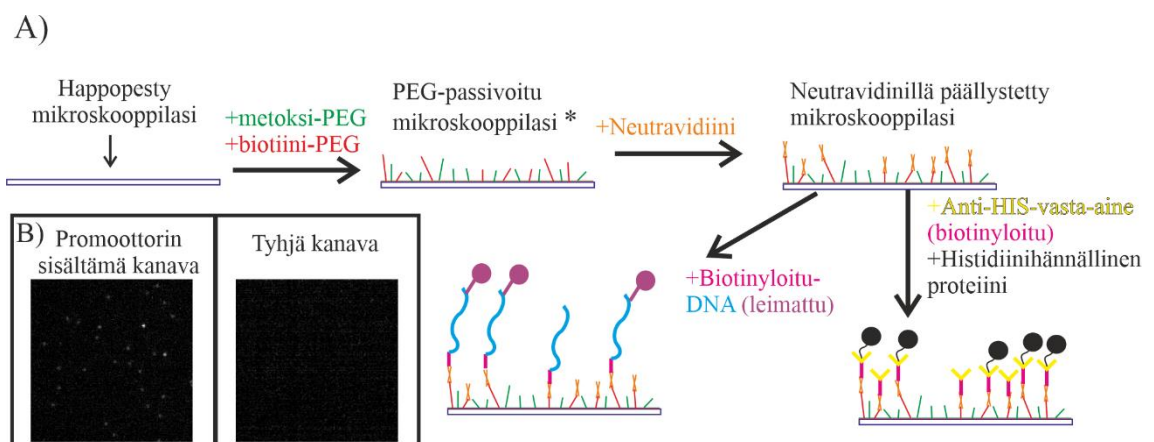


#### 4.4 Proteiinien leimaus

Yhteensä kolmea eri menetelmää käytettiin eri proteiinien leimaamiseen: maleimidi-tekniikkaa, SNAP-tekniikkaa sekä pAzF:n fosfiinileimaus. Näistä maleimidi- ja fosfiinileimaustekniikka ovat kohdennettuja leimauksia, kun SNAP taas on proteiinileima. Leimoiksi valittiin uuden sukupolven DyLight -fluoresoivat leimat, joista käyttöön valittiin DyLight550-maleimidi ja DyLight650-fosfiini. Maleimidileima kiinnittyy kysteiinitähteen tioliryhmään, jolloin leima on kohdennetusti kiinni tietyssä kohtaa proteiinia. Fosfiinileima reagoi pAzF:n atsidiryhmän kanssa Staudingerin ligaatioksi kutsutussa reaktiossa, jolloin myös se on kohdennetusti kiinni tietyssä kohtaa proteiinia. Maleimidireaktio on nopea, mutta Staudingerin ligaatio on hidas. Atsidit ovat herkkiä muutoksille, sillä atsidiryhmä on melko epävakaa. TBP:n maleimidileimauksessa saavutettiin 57 % leimaustehokkuus, RRN3:n leimaustehokkuus oli 21 %. CF leimautui Staudingerin ligaatiolla niin heikosti, että leimaustehokkuutta ei pystytty luotettavasti määrittämään. Maleimidillä on mahdollisuus optimoida leimaustehokkuutta, mutta se jätettiin ajan puutteen vuoksi pois. Atsidin fosfiinileimauksen optimointi ei myöskään ehtinyt tähän tutkielmaan, sillä tämänkin leimaustehokkuuden saavuttaminen vei jo paljon aikaa. Atsidin sisältämien proteiinien leimauksessa on todennäköisesti useampia virhelähteitä. Ensinäkin, proteiineja ei aluksi puhdistettu suojattuna kaikelta valolta koko puhdistusprosessin ajan, mikä ainakin nikkelipylväessä saattaa pelkistää ja inaktivoida pAzF-aminohapon atsidin. Toiseksi suuri osa fosfiineista oli todennäköisesti ehtinyt hapettumaan ennen leimausreaktiossa onnistumista. Tästä syystä CF päätettiin leimata SNAP-välitteisesti. Käytetty SNAP-proteiiniin sitoutuva leima oli TMR-star. Tämän avulla suoritetun leimauksen tehokkuus pienissä pitoisuuksissa osoittautui vaikeaksi. Käytetyn TMR-starin ekstinktiokvakioksi määriteltiin arvo  $57\,000\text{ M}^{-1}\text{ cm}^{-1}$ , jolloin pienillä CF- sekä TMR-star-pitoisuuksilla ei TMR-staria ole havaittavissa spektrofotometrillä. DyLight leimat havaitaan pienemmillä pitoisuuksilla, sillä niiden ekstinktiokvakiot ovat suurempia. DyLight550:n ekstinktiokvatio on  $150\,000\text{ M}^{-1}\text{ cm}^{-1}$  ja DyLight650:n ekstinktiokvatio on  $250\,000\text{ M}^{-1}\text{ cm}^{-1}$ . Kun CF-SNAP-proteiinia oli leimausreaktiossa  $28\ \mu\text{g}$  ( $130\ \text{pmol}$ ) ja TMR-star fluoroforia  $170\ \text{pg}$  ( $250\ \text{pmol}$ ), saavutettiin 78 % leimaustehokkuus. Näin kaikki tarvittavat proteiinit saatiin lopulta leimattua mikroskopiaan riittävällä tehokkuudella.

#### 4.5 Mikroskooppiprotokollan luominen

Turun Yliopistolla on kaksi TIRF-superresoluutiomikroskooppia, joista DeltaVision OMX (GE-Healthcare) valittiin sen neljän eri kanavan samanaikaisen kuvaamismahdollisuuden takia. TIRF-mikroskopia valittiin siksi, että mikroskooppilasille voidaan laittaa sisältämiä proteiineja eivätkä ne häiritse varsinaisia mikroskooppilasin pinnassa havaittuja sitoutumisia, koska viritysväloa ei ulotu kuin muutamia satoja nanometrejä lasin pinnasta näyteliuokseen. Mikroskooppilaseille on suoritettava muutamia käsittelyvaiheita, ennen kuin niitä voidaan käyttää kuvaamiseen. Ensin mikroskooppilasit on pestävä, silanoitava ja passivoitava spesifin linkkerin sisältämällä polyetyleeniglykolisekoituksella (engl. *polyethyleneglycol*, PEG) (kuva 20 A). Tämän jälkeen PEGyloidut mikroskooppilasit voidaan päällystää halutulla proteiinilla tai DNA:lla. Superresoluutiomikroskopiaa varten on mikroskooppilasin oltava hyvin puhtaita. Yleisimmin käytettyjä puhdistusprotokollia on mikroskooppilasin polttaminen korkeassa lämmössä (400°C) tai niiden puhdistaminen plasmalla, kuten Dulin ja muut (2018) tekivät. Koska kumpaakaan ei ollut helposti saatavilla, päädyttiin mikroskooppilasin happopesuun. Näin kaikesta ylimääräisestä materiaalista päästään eroon, eikä mikroskooppilaseilla pitäisi olla havaittavissa taustafluoresenssia mikroskoopilla. Seuraavaksi mikroskooppilasin pinnat silanoidaan, eli päällystetään negatiivisesti varautuneilla amiineilla. Tämän jälkeen mikroskooppilaseihin kiinnitetään PEG. Käytettävä PEG on sekoitus kahta tyyppiä: metoksi-PEG-sukinimidyyli-valeraatti joka passivoi mikroskooppilasin pinnan sekä



**Kuva 20.** Mikroskooppilasin valmistelu ja havainnekuva promootorikanavasta sekä tyhjästä kanavasta. **A)** Mikroskooppilasin valmistelun havainnekuva. Happopesty mikroskooppilasi on silanoitava ennen PEG-passivointia. Neutravidiinillä päällystettyyn mikroskooppilasiin liitettyyn joko biotinyloitu ja leimattu DNA tai histidiinihännän tunnistava vasta-aine (Anti-HIS-vasta-aine) sekä histidiinihännällinen proteiini, joka sisälsi usein leiman. **B)** Havainnekuva leimatulla promootorilla kuvatusta mikroskooppilasin pinnasta vasemmalla ja kuva kanavasta, jossa ei ollut promootoria. \*PEG-passivoitujen mikroskooppilasin valmistusvaiheessa säilöä 4°C:ssä noin kahden viikon ajan ennen käyttöä.

biotini-PEG-sukkinimidyylikarbonaatti, joka passivoi mikroskooppilasin pinnan ja sisältää pinnasta ulospäin olevan biotiiniryhmän, johon voidaan liittää biotiinin kanssa reagoivia ryhmiä (kuva 20 A). Kun mikroskooppilasit ovat päällystetty, voidaan niitä säilyttää noin kahden viikon ajan 4°C:ssa. Viimeisessä vaiheessa passivoidut, biotiinin sisältävät mikroskooppilasit päällystetään halutulla proteiinilla tai DNA:lla. Tämä vaatii välikappaleeseen neutravidinin, joka kiinnittyy pinnan biotiiniin antaen mahdollisuuden liittää itseensä vielä biotiinin sisältämä ryhmä. Tällöin syntyy (PEG)-biotini-neutravidin-biotini-(DNA) kerros. Proteiinit eivät sisältäneet DNA:ta, jolloin suoraan neutravidiniin pystytään kiinnittämään vain biotinyloitu DNA. Mikäli tietty proteiini haluttiin liittää mikroskooppilasin pintaan, lisättiin ensin neutravidiniin konjugoituvana substraattina biotinyloitu vasta-aine, joka sitoutui kohdeproteiinin histidiinihäntään. Kun mikroskooppilasi valmistelun protokolla oli valmis, pystyy sitoutumiskokeita tekemään mikroskoopilla.

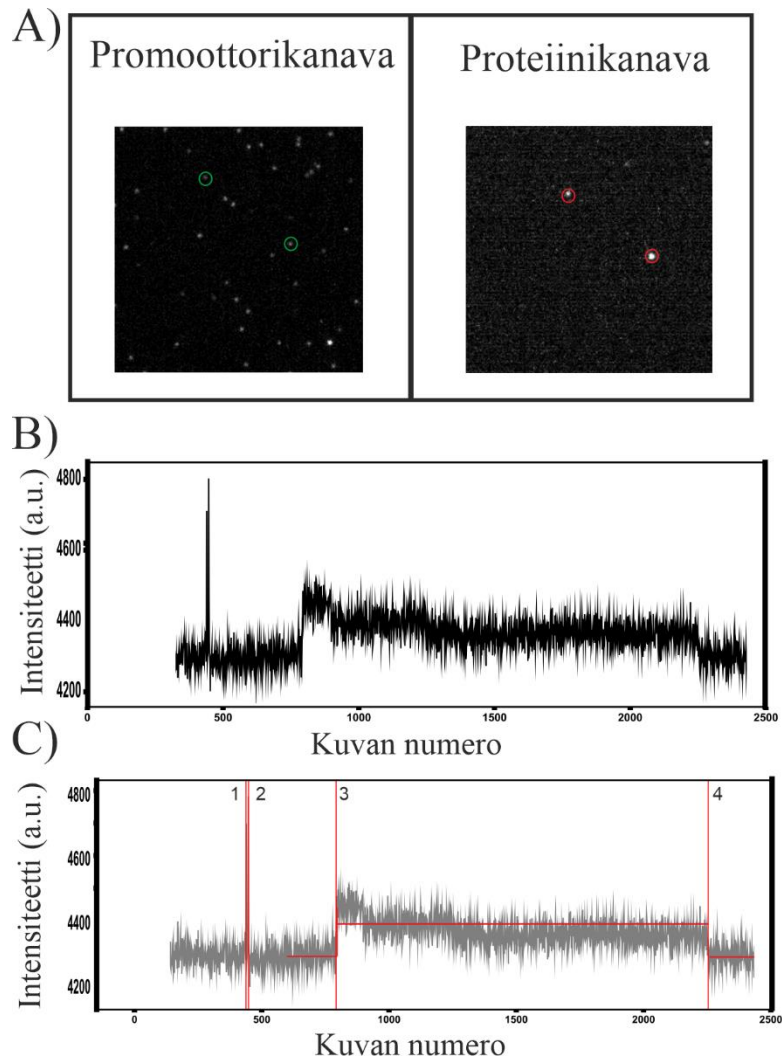
#### *4.6 Reaaliaikaiset sitoutumiskokeet*

Mikroskooppikokeen alkaessa täytyy ensimmäiseksi etsiä mikroskooppilasin pinta. DeltaVision OMX:llä on tähän oma automatisointinsa, mutta se ei löytänyt luotettavasti tyhjää mikroskooppilasin pintaa. Mikroskooppilasin pinnan löytymiseksi manuaalisesti oli lasin pinnalla oltava joko vasta-aineeseen kiinnitetty leiman sisältämä proteiini tai neutravidiniin kiinnitetty, biotinyloitu ja leimattu kohde-DNA (kuva 20). Reaaliaikaisissa sitoutumiskokeissa kiinnitettynä oli leimattu DNA, fluoroforin elinaikakokeissa leimattu proteiini. Kun leimattu DNA löytyi, vaihdettiin puskuri ja joukkoon lisättiin leiman sisältämä proteiini joko yksin (CF ja RRN3) tai yhdessä muiden proteiinien kanssa; leimatun proteiinin pitoisuus oli 0,5 – 5 nM. DeltaVision OMX:n kameroiden rajoitusten takia kahta kanavaa ei kyetty kuvaamaan samanaikaisesti, joten ensin kuvattiin promoottorikanavaa 400 – 600 ms valotusajalla ja 100 % laserin intensiteetillä promoottorien paikantamiseksi. Promoottorikanavan pitkä valotusaika pituus johtuu käytettyjen Atto488- ja Cy5-leimojen himmeystestä. Promoottorien paikallistamisen jälkeen mikroskooppi tarkennettiin proteiinikanavalle ja proteiini-DNA kompleksien muodostumista seurattiin 3001:n kuvan ajan. TMR-star:illa leimattua CF:ää kuvattiin 561 nm:n kanavalla, 100 % laserin intensiteetillä. Ennen lopullisen valotusajan valitsemista, tutkittiin proteiinikanavan signaalin ja taustan suhdetta valotusajoilla 20 – 150 ms. Tarkoituksena oli löytää lyhin mahdollinen valotusaika, jolla signaali on erotettavissa. Lopulta päädyttiin 50 ms:n valotusaikaan. TMR-star:in ollessa leimana, DNA-kanavana

oli joko 488 nm tai 642 nm. DyLight650:llä leimattu RRN3 kuvattiin 642 nm:n kanavalla, 5–50 ms valotusajalla. RRN3-DyLight650:n kanssa DNA-kanavana oli 488 nm. Kun filmi oli otettu, siirryttiin mikroskooppilasilla toiseen kohtaan, jossa kuvattiin uusi filmi. Yhtä mikroskooppilasia käytettiin alle 20 minuuttia, sillä kuvauspuskurissa oleva hapensitojaentsyymi alkaa laskemaan pH:ta tämän ajan ulkopuolella, eikä data tällöin ole luotettavaa.

#### *4.7 Mikroskooppidatan analysointivaiheen kehittäminen*

Kun mikroskooppikokeet oli suoritettu, linjattiin eri kanavilla kaikki tallennetut kuvasarjat käyttämällä DeltaVision OMX:n mukana tullutta SoftWorX -ohjelmaa. Kuvasarjojen linjaamisen jälkeen piti selvittää (1) miten promoottorikanavan kuva saadaan monistettua proteiinikanavan mittaiseksi, (2) miten promoottorikanavan ja proteiinikanavan kuvasarjat saadaan yhdistettyä toisiinsa ja mitä vaatimuksia kohdan 3 ohjelmilla on tähän, (3) miten ja mitä ohjelmaa käyttää molekyylien havainnointiin ja näiden intensiteettien seuraamiseen, (4) miten intensiteettien seuraamista ja mahdollista muuttumista voidaan käsitellä ja (5) miten näiden intensiteettien muutokset saadaan analysoitua. Kohdat 1 ja 2 päädyttiin tekemään Fiji-ohjelmistolla (Schindelin ja muut, 2012). Ennen kohdan 2 ratkaisua oli päätettävä, millä ohjelmalla molekyylien havainnointi suoritetaan. Koska kyseessä on yhteissijaintikokeet ja käytössä vain kaksi kanavaa, päädyttiin iSMS-ohjelmistoon (Preus ja muut, 2015). iSMS on suunniteltu yksittäismolekyyl-FRET-datan analysointiin, joten tulosten analyysiä ei voitu suorittaa tällä ohjelmistolla. Sillä voidaan kuitenkin seurata yksittäisiä molekyyliä ja näiden antamaa signaalia (kuva 21 A ja B). Nämä signaalit saadaan numeerisina arvoina ulos, jolloin tämän datana analysointi voidaan suorittaa muilla ohjelmilla. iSMS vaatii kuvasarjojen olemista vierekkäin, joten kuvat käsiteltiin kohdassa 2 sen mukaan. Tämän jälkeen intensiteettien muutosten aika, eli aika jonka proteiini viettää promoottorilla, on määritettävä jollain muulla ohjelmistolla. Tähän käytettiin Python-koodikielipohjaista signaaliprosessointikoodia (Truong ja muut, 2020). Signaaliprosessointikoodi etsi fluoresenssisignaalin muutospisteet (engl. *change point finder*, CPF) (kuva 21 C). Kun CPF havaitsi signaalin intensiteetin nousun ja laskun, se myös laski näiden tapahtuman välisen ajan, joka on kyseisen muodostuneen DNA–proteiinikompleksin elinaika. Kokosin kaikki havaitut yksittäisten kompleksien elinajat yhteen ja naalysoin saadun jakauman käyttäen joko Origin ohjelmistoa tai signaaliprosessointikoodiin (Truong ja muut, 2020) pohjautuvaa sovitusmenetelmää (engl. *exponential fit program*, EFP). Origin



**Kuva 21.** Data-analyysin vaiheet. *A)* Promoottorikanavasta etsitään hyvin erottuvat promoottorit (ympyröity vihreällä), joiden signaaliradan muutosta tutkitaan proteiinikanavalla. Kun signaaliradassa havaitaan muutos (ympyröity punaisella) eli proteiini sitoutuu promoottoriin, voidaan tämän sitoutumisen kesto määrittää. *B)* Signaaliradan havainnekuva. Signaaliradan intensiteetin muutokset voivat olla äkkinäisiä tai pidempiaikaisia. *C)* CPF-ohjelman havaitsemat signaaliradan muutokset.

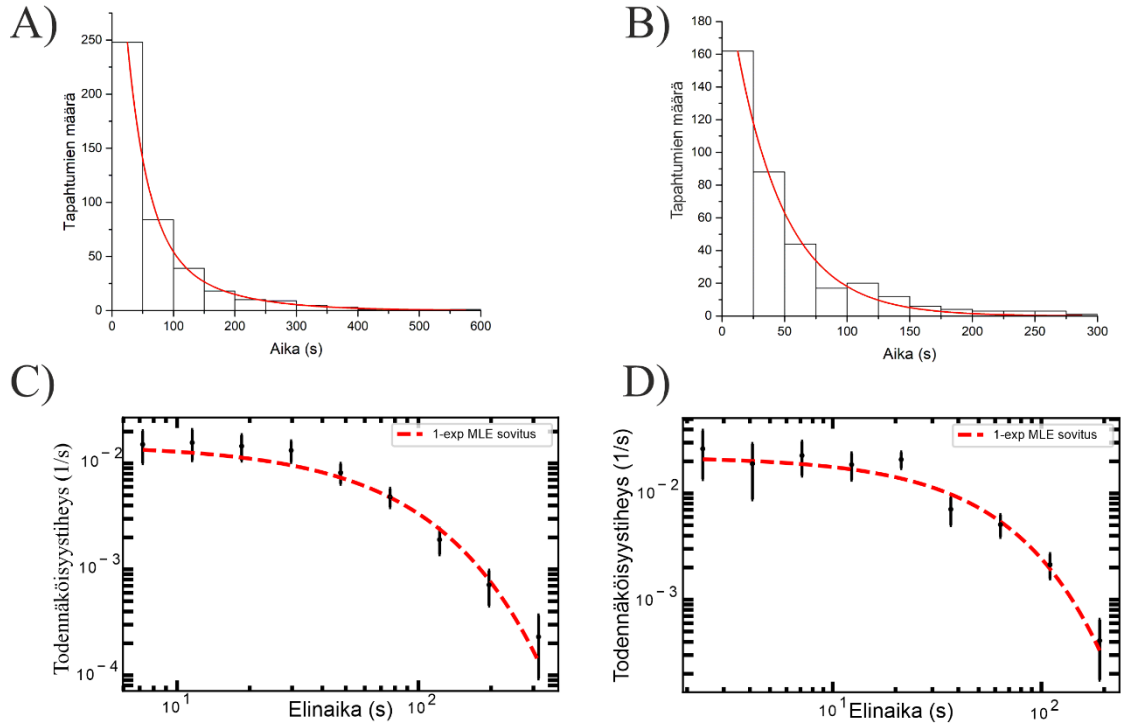
luokittelee kompleksien elinajat ryhmiin, esimerkiksi 1-50 ms, 51-100 ms ja niin edelleen, ja sovittaa saatuun histogrammikuvaajaan prosessia kuvaavan yhtälön, yksieksponttisen tai kaksieksponttisen yhtälön avulla (kuva 22 A ja B) sovittaa näiden pylväiden antaman datan joko yksieksponttisen yhtälön tai kaksieksponttisen yhtälön avulla. Sen sijaan EFP ei luokitele yhtälöön sovittamisvaiheessa DNA–proteiinikompleksien elinaikoja ryhmiin, vaan sovittaa yhtälön hyödyntäen kaikkia mittaustuloksia. Yhtälöön sovituksen visualisoimiseksi EFP kyllä tekee myös Origin-tyylisten histogrammikuvaajan, mutta tämä histogrammi ei suoraan vaikuta yhtälön sovitukseen (kuva 22 C ja D). EFP:n histogrammikuvaassa jokaisen dataluokan histogrammipylvään luotettavuusrajaa myös arvioidaan bootstrap-menetelmällä (engl. *bootstrapping*) jokaiselle pylväälle, jolloin EFP soveltuu paremmin hajanaisemman tai

vähemmän aikapisteitä sisältävän datan analyysille. Molemmissa ohjelmissa käytettiin ratkaistujen kineettisten parametrien virheen arvioimiseksi 95 %:n luottamusväliä. Mikäli dataa oli käytössä vähän ( $N < 400$  tapahtumaa), antaa EFP luotettavamman tuloksen kuin Origin.

#### *4.8 Käytettyjen fluoroforien elinajat*

Ennen komponenttien sitoutumisen tutkimista on syytä selvittää käytettävien fluoroforien elinaika (kuva 22). Fluoroforin sisältämät proteiinit kiinnitettiin mikroskooppilasiin histidiinihännän sisältämän vasta-aineen avulla (kuva 20 A). Tämän jälkeen fluoroforeja kuvattiin 3001:n kuvan ajan, jolloin seurattiin yksittäisten fluoroforien sammumista. Mikäli proteiinien sitoutumisaika on merkittävästi lyhyempi kuin fluoroforin elinaika samoilla mikroskoopin asetuksilla, voidaan mitattua proteiinin sitoutumisaikaa DNA:lla pitää luotettavana. Kineettisellä analyysillä ratkaistu fluoroforin keskimääräinen elinaika on se aika, minkä kuluttua puolet fluorofori-molekyyleistä on sammunut. Määritin elinajat DyLight650- sekä TMR-star -fluoroforeille, jotka ovat proteiineissa kiinni (RRN3-DyLight650 ja CF-TMR-star). TMR-star elinaika oli Originanalyysissä 50,7 s, 95 % luottamusväli 46,0 – 55,4 s ( $N = 419$ ) (kuva 22 A), DyLight650 elinaika oli 39,9 s, 95 % luottamusväli 35,9 – 43,9 s ( $N = 363$ ) (kuva 22 B). CF-TMR-starin sitoutumista voidaan luotettavasti mitata, mikäli sitoutumisen aika on lyhyempi kuin 20 s, RRN3-DyLight650 sitoutuminen on luotettavaa, kun sitoutuminen on lyhyempi kuin 15 s. Jos sitoutumisajat ovat yhtä pitkiä kuin fluoroforien elinajat, puolet fluoroforeista on sammunut eivätkä välttämättä irronneet promoottorilta. EFP analyysin mukaan TMR-star:in elinaika oli 66,8 s, 95 % luottamusväli 60,5 – 73,6 s ( $N = 416$ ) ja DyLight650 elinaika oli 45,3 s, 95 % luottamusvälillä 40,6 – 50,1 s ( $N = 359$ ). Analyysien antamien tulosten erot johtuvat muutamastakin asiasta, mutta yksi suurimmista syistä on ohjelmien datan käsittely. Origin ryhmittelee dataa, jolloin osa datapisteistä menee hukkaan. EFP ei ryhmittele dataa, jolloin se käyttää sovitukseen kaikkia saatuja datapisteitä. Suurella määrällä dataa Origin on luotettavampi, mutta pienemmällä määrällä dataa EFP on luotettavampi. Yhdenvertaisuuden vuoksi tulokset ilmoitetaan pääosin Origin-analyysinä. Tapauksissa, joissa dataa ei ole tarpeeksi luotettavan Origin-analyysin

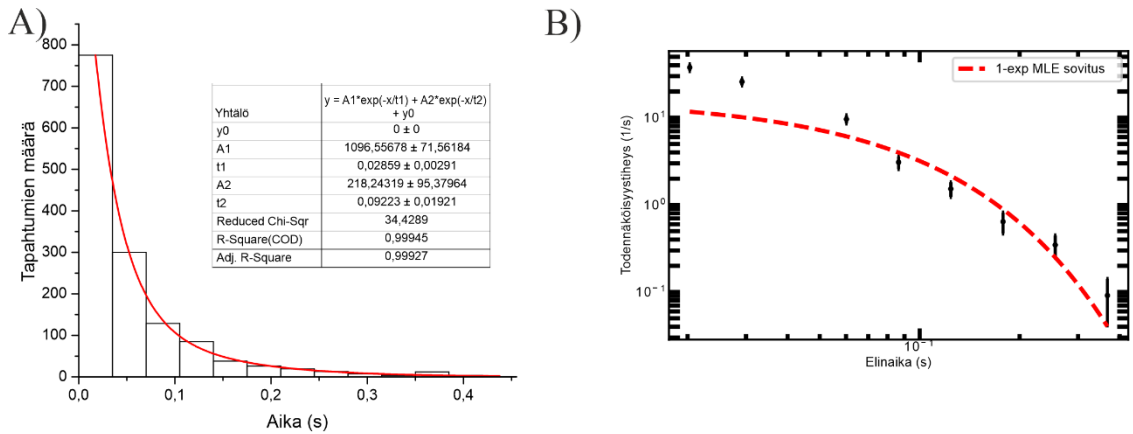
suorittamiseksi, käytetään pääasiallisena analyysikeinona EFP-sovitusta, mutta molemmat sovitukset esitetään.



**Kuva 22.** Käytettyjen fluoroforien elinajan määrittäminen. **A)** Origin-analyysi TMR-star:in elinajasta. Kuvaajan lopussa käyrän sovitus ei ihan sovi käytettyihin lokeroihin. **B)** Origin-analyysi DyLight650:n elinajasta. **C)** EFP-analyysi TMR-star:in elinajasta. Toisin kuin Origin, EFP käyttää kaikkia saatuja arvoja käyrän sovittamiseen. **D)** EFP-analyysi DyLight650:n elinajasta. Elinajat on kaikissa määritetty yksittäisellä eksponentilla, eli fluoroforit sammuvat yhdessä vaiheessa.

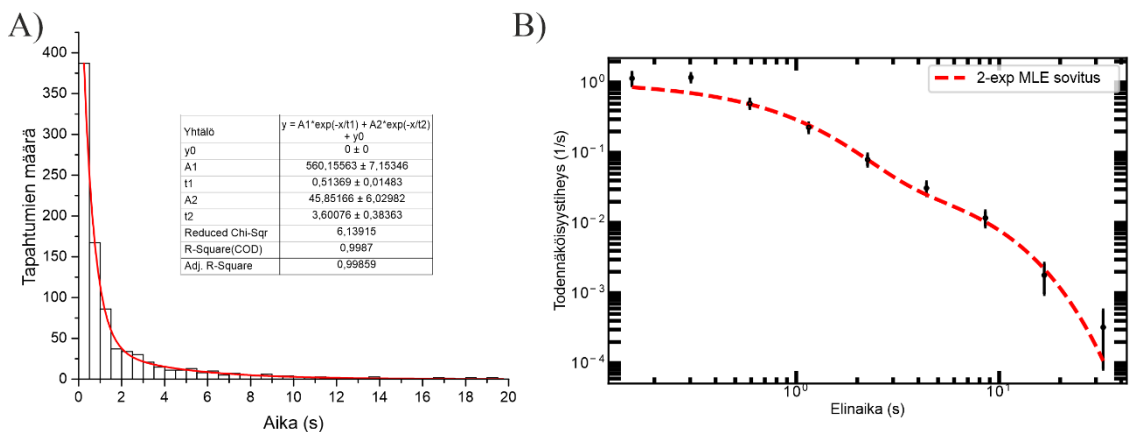
#### 4.9 Välttämättömän aloituskompleksin komponenttien sitoutuminen promoottoriin

RRN3 elinaika promoottorilla oli Origin-analyysissä kaksivaiheinen (kuva 23 A), joissa lyhyen sitoutumisen elinaika  $t_1 = 28,6$  ms (95 % luottamusväli 22,0 – 35,2 ms) ja toisen vakaamman sitoutumisen elinaika  $t_2 = 92,2$  ms (95 % luottamusväli 48,8 – 136 ms;  $N = 1414$ ). EFP-analyysissä RRN3:n elinaikaa ei pystytty kaksivaiheisesti sovittamaan, joten yksivaiheisen sovituksen mukaan RRN3:n elinaika promoottorilla oli 61,2 ms (95 % luottamusväli 58,2 – 64,7 ms) (kuva 23 B). Todennäköisesti RRN3:n elinaika promoottorilla oli vielä lyhyempi, sillä suurin osa sitoutumisista oli alle 5 kuvan mittaisia. Kuvaus suoritettiin nopeimmilla mahdollisilla asetuksilla. Kyseessä on todennäköisesti satunnainen havainto eikä DNA:han sitoutuminen, sillä RRN3:n määrä oli suhteellisen suuri (5 nM) verrattuna promoottoriin (10 pM). Tätä tukee myös se, että Origin-analyysin mukainen  $t_1$  on pienempi kuin sen ensimmäinen ryhmittymä, jolloin luotettavaa sitoutumisaikaa ei pystytä määrittämään (kuva 23 A).



**Kuva 23.** RRN3:n sitoutuminen promoottoriin. **A)** Origin-analyysi RRN3:n sitoutumisesta promoottoriin. Originanalyysin mukaan tapahtuma on kaksivaiheinen. **B)** EFP-analyysi RRN3:n sitoutumisesta. EFP:n käyrän sovitus ei onnistunut luotettavasti kahdella eksponentilla, joten yksiekspONENTTINEN käyrän sovitus oli ainoa mahdollisuus. Käyrä ei sovi kuvaajaan yhtä hyvin kuin Origin-analyysissä.

CF sitoutuminen promoottoriin on kaksivaiheista (N = 884) (kuva 24). Ensimmäisen eli lyhyen sitoutumisen elinaika Origin-analyysillä  $t_1 = 514$  ms (95 % luottamusväli 484 – 544 ms) ja toisen eli vakaamman sitoutumisen elinaika  $t = 3,60$  s (95 % luottamusväli 2,82 – 4,38 s) (kuva 24 A). EFP-analyysillä vastaavat elinajat olivat  $t_1 = 730$  ms (95 % luottamusväli 640 – 840 ms) ja  $t_2 = 5,27$  s (95 % luottamusväli 4,34 – 6,59 s) (kuva 24 B). Origin-analyysillä noin 8 %:a ja EFP-analyysillä 27,5 %:a sitoutumistapahtumista kuului vakaampaan kategoriaan. CF:n sitoutumista satunnaiseen DNA:han ei havaittu tarpeeksi. Analysoiduista 1063 promoottorista saatiin 55 signaalirataa, joita ei pystytty luotettavasti käyttämään. CF ei siten näyttäisi sitoutuvan ainakaan käytetyssä pienessä pitoisuudessa (500 pM) DNA:han, jossa ei ole Pol I -promoottoria.

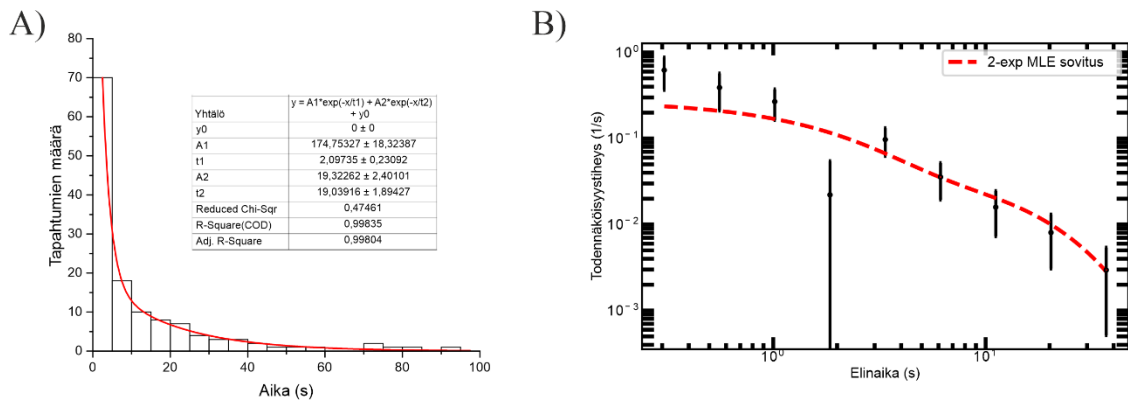


**Kuva 24.** CF:n sitoutuminen promoottoriin. **A)** Origin-analyysi CF:n sitoutumisesta promoottoriin. CF sitoutuu promoottoriin kaksivaiheisesti. **B)** EFP-analyysi CF:n sitoutumisesta promoottoriin. Myös EFP-analyysissä CF:n sitoutuminen oli kaksivaiheista, vaikka alku- ja loppupään sovitus eivät täysin sovi.

Satunnaisesti kysteineistään leimatun Pol I:n sitoutuminen promoottoriin vaikutti myös kaksivaiheiselta (kuva 25). Origin-analyysillä ensimmäisen lyhyen sitoutumisen elinaika



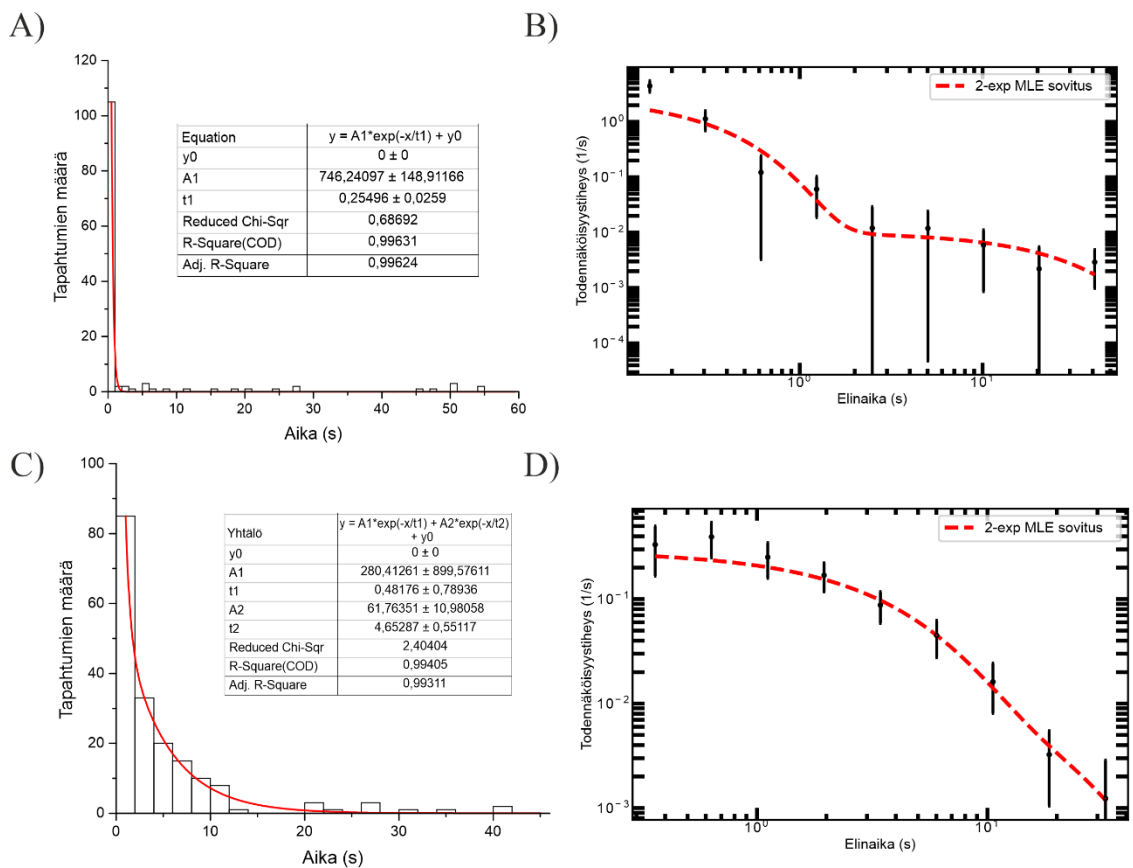
$t_1 = 2,10$  s (95 % luottamusväli 1,61 – 2,59 s) ja toisen vakaamman sitoutumisen  $t_2 = 19,0$  (95 % luottamusväli 15,0 – 23,1 s) (kuva 25 A). EFP-analyysillä ensimmäisen lyhyen sitoutumisen elinaika  $t_1 = 1,68$  s (95 % luottamusväli 0,93 – 2,73 s) ja toisen vakaamman sitoutumisen elinaika  $t_2 = 13,2$  s (95 % luottamusväli 10,5 – 16,0 s) (kuva 25 B). Todennäköisyys sitoutua vakaammin oli Origin-analyysillä noin 11 %:a ja EFP-analyysillä noin 61 %:a. Suuri hajonta johtuu pienestä otannasta (N = 126). Tarkoituksena oli katsoa, sitoutuuko Pol I itsestään DNA:han. Pol I:llä vaikuttaisi olevan monenlaisia, ainakin kahdenlaista, epäspesifejä sitoutumismuotoja. Koska kysteiniiniin liitetyn DyLight550:n elinaikaa ei ehditty määrittämään, tuloksiin on suhtauduttava alustavina. Varsinkin vakaamman sitoutumisen arvioituun keston voi vaikuttaa lyhentävästi fluoroforin sammuminen. Lisäksi osa Pol I -molekyyleistä saattoi sisältää useamman kuin yhden leiman, mikä myös sotkee analyysialgoritmeja.



**Kuva 25.** Satunnaisleimatun Pol I:n sitoutuminen DNA:han. **A)** Origin-analyysi Pol I:n sitoutumisesta promoottoriin. Sitoutuminen vaikuttaa olevan kaksivaiheista. **B)** EFP-analyysi Pol I:n sitoutumisesta promoottoriin. Myös EFP:llä tehdyssä analyysissä parhaiten sopiva yhtälö oli kaksivaiheinen, mutta EFP:llä käyrän sovitus ei sovi kuvaan.

Välttämättömän aloituskompleksin eli CF:n, RRN3:n ja Pol I:n muodostaman kompleksin sitoutumista promoottoriin tutkittiin yksivärikokeilla, jolloin mitattavana tekijänä oli joko leimattu CF tai leimattu RRN3 (kuva 26). Leimattujen proteiinien pitoisuus oli kymmenesosa muiden proteiinien pitoisuudesta. Leimatun RRN3:n sisältämän välttämättömän aloituskompleksin sitoutuminen promoottoriin riippui käytetystä analyysimenetelmästä (N = 130) (kuva 26 A ja B). Origin-analyysillä sitoutuminen oli yksivaiheista, ja sen kesto oli 255 ms (95 % luottamusväli 203 – 307 ms) (kuva 26 A). EFP-analyysillä sitoutuminen oli kaksivaiheista ja ensimmäisen lyhyen sitoutumisen elinaika  $t_1 = 270$  ms (95 % luottamusväli 230 – 320 ms) ja toisen sitoutumisen elinaika  $t_2 = 23,2$  s (95 % luottamusväli 15,7 – 31,5 s) (kuva 24 B). Vakaamman sitoutumisen todennäköisyys on 23,5 %:a Koska molemmat sitoutumiset

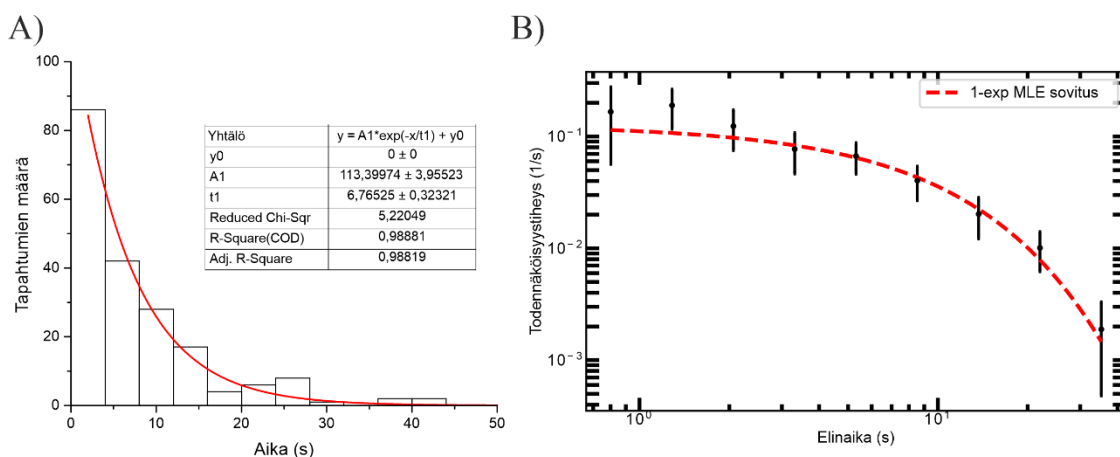
ovat nopeampia kuin DyLight650:n elinaika (EFP-analyysillä  $< 45,3$  s) ja koska sitoutumisten kesto eroaa selvästi pelkän RRN3:n sitoutumisdatasta (EFP-analyysillä yksivaiheinen sitoutuminen, jonka elinaika oli 61,2 ms), voidaan tulosta pitää varsin luotettavana. Leimattun CF:n sisältämä välttämätön aloituskompleksi sitoutuu kahdessa vaiheessa promoottoriin (kuva 27 C ja D). Origin-analyysillä ei saatu luotettavia arvoja, sillä ensimmäisen vaiheen amplitudin arvon  $a_1 = 280,95$  % luottamusväli oli  $-1600 - 2200$  ja ensimmäisen vaiheen sitoutumisen elinajan  $t = 480$  ms 95 % luottamusväli oli  $-1,17 - 2,13$  s (kuva 27 C). EFP-analyysillä sitoutuminen oli kaksivaiheista, joista ensimmäisen nopean sitoutumisen elinaika  $t_1 = 2,97$  s (95 % luottamusväli  $2,04 - 3,65$  s) ja toisen pidemmän sitoutumisen elinaika  $t_2 = 13,2$  s (95 % luottamusväli  $6,58 - 17,8$  s). Tässä tapauksessa molemmat sitoutumiset ovat pidempiä kuin pelkän CF:n sitoutumisajat (730



**Kuva 26.** Pol I välttämättömän aloituskompleksin sitoutuminen promoottoriin. Kaikissa reaktioissa on mukana Pol I, RRN3 ja CF, mutta leimattu proteiini vaihtelee. **A)** Sitoutuminen promoottoriin, jossa leimattuna proteiinina oli RRN3. Origin-analyysi. Sovitus on yksivaiheinen, todennäköisesti lyhyiden sitoutumisten takia. **B)** Sitoutuminen promoottoriin, jossa leimattuna proteiinina oli RRN3. EFP-analyysi. EFP-analyysin sovitus oli kaksivaiheinen, vaikka esitettyjen ryhmittymäpisteiden hajonta onkin melko suurta. **C)** Sitoutuminen promoottoriin, jossa leimattuna proteiinina oli CF. Origin-analyysi. Sitoutuminen on kaksivaiheista, mutta pidempiä sitoutumisia ei juuri oteta huomioon. Sovitus ei ole luotettava, sillä virheet ovat liian suuret 95 % luottamusvälillä. **D)** Sitoutuminen promoottoriin, jossa leimattuna proteiinina oli CF. EFP-analyysi. Sitoutuminen on kaksivaiheista ja pidemmät sitoutumiset otetaan paremmin huomioon kuin Origin-analyysissä.

ms ja 5,27 s), joten niitä voidaan pitää luotettavina, vaikka otanta oli melko pieni (N = 183).

Edellä mainittujen lisäksi ehdittiin vielä kokeilemaan UAF:n ja TBP:n vaikutus CF:n sitoutumiseen (kuva 28). Valitettavasti näitä ei ehditty tekemään erikseen, vaan UAF:n ja TBP:n vaikutus mitattiin yhdessä. Yhdessä UAF:n ja TBP:n kanssa, CF:n sitoutuminen vaikutti tapahtuvan vain yhdellä tavalla, jonka elinaika Origin-analyysillä oli 6,77 s (95 % luottamusväli 6,09 – 7,44 s) ja EFP-analyysillä 7,91 s (95 % luottamusväli 7,55 – 8,29 s; N = 126). Tämä elinaika on pidempi kuin CF:n oma pidempi sitoutumisaika promoottoriin (Origin-analyysillä 3,60 s ja EFP-analyysillä 5,27 s) ja toisin kuin CF:n sitoutuminen yksinään, tämä reaktio näyttäisi tapahtuvan yhdessä vaiheessa.



**Kuva 27.** UAF:n, TBP:n ja leimatun CF:n sisältämän kompleksin sitoutuminen promoottoriin. **A)** Origin-analyysi kompleksin sitoutumisesta. Sovitus on yksiekponenttinen ja reaktio tällöin yksivaiheinen. Käyrän sovitus ei ole täydellinen pienen otannan takia. **B)** EFP-analyysi kompleksin sitoutumisesta. EFP-sovitus on yksiekponenttinen kuten Origin-sovitus, vaikka alkupään ryhmittymät eivät käyrään sovi hyvin.

## 5. Tulosten tarkastelu

### *5.1 Puhdistetun Pol I –transkriptiokoneiston valmistaminen biofysikaalisia tutkimuksia varten*

Tutkiakseni Pol I:n aloituskompleksia, oli minun luotava protokollat proteiinien tuotolle ja puhdistukselle. Kaikki proteiinit saatiin puhdistettua UAF:ää lukuun ottamatta, mutta sitä saatiin käyttöön Olena Parilovan puhdistamana. Proteiinien saannot olivat melko hyviä Pol I:tä lukuun ottamatta. Pol I:n puhdistaminen vei paljon aikaa ja oli todella työlästä ottaen huomioon saadun proteiinin. Hiivaa piti ensin tuottaa yhteensä 60 litraa, jotta sitä saatiin solumassana 150 g. Se on riittävä määrä yhteen puhdistukseen. Tästä määrästä saatiin vain 70 µg proteiinia, joka ei mitenkään riitä esimerkiksi kvantitatiivisiin EMSA- tai anisotropia kokeisiin. Käytin Pol I:tä kuitenkin muutamiin EMSA-kokeisiin, joissa se sitoutui promoottoriin yhdessä RRN3:n kanssa melko pienissä pitoisuuksissa (kuva 16). Sen sijaan RRN3:n, CF:n ja TBP:n tuottotasot olivat melko hyviä, eikä niiden tuottaminen ja puhdistaminen ollut erityisen työlästä, ainakaan verrattuna Pol I puhdistukseen. Yksittäismolekyylifluoresenssikokeiden tekeminen oli viisasta, sillä se ei vaadi suuria määriä proteiineja. Yksittäismolekyylifluoresenssikokeilla havaitaan tarkemmin myös spesifiset ja epäspesifiset sitoutumiset, sillä Pol I –koneiston proteiinit sitoutuvat todennäköisesti heterogeenisesti, eli proteiineilla tapahtuu spesifistä ja epäspesifistä sitoutumista DNA:han rinnakkain. Hyötynä yksittäisten molekyylien tutkimisessa on myös sen reaaliaikaisuus.

RRN3:n ja CF:n leimapaikkojen suunnittelu rakennekuvien avulla oli aluksi haastavaa, mutta hyvät rakennekuvat tarjosivat mahdollisuuden suunnitella tarkat leimapaikat. pAzF:n sisältämän RRN3:n tuottotasot olivat hyvät, sillä N201pAzF mutanttia saatiin 470 µg ja villityypin proteiinia 540 µg. RRN3 onnistuttiin myös leimaamaan, vaikka leimaustehokkuus oli huono, vain 21 %. CF, jossa oli Y768:n tilalla pAzF ei onnistuttu tuottamaan tarpeeksi sen leimaamiseksi, joten aika- ja materiaalirajoitteiden myötä siirryttiin CF:n leimaamiseen SNAP-proteiinilla, joka onnistui 78 % tehokkuudella ja itse proteiinin tuotto onnistui hivenen paremmin kuin villityypin CF:n tuotto. SNAP-proteiinin sisältämän CF:n saanto oli 500 µg, kun villityypin CF:n tuotto oli 330 µg.

TBP tuottui proteiineista parhaiten, mutta sen käyttö jäi vähälle tässä pro gradu tutkielmassa. Tämä oli tietoinen valinta, sillä välttämättömän aloituskompleksin

tutkiminen oli etusijalla. Luoduilla TBP-mutanteilla pystytään kuitenkin leimaamaan TBP, käyttäen joko maleimidi-strategiaa tai Staudingerin ligaatiota.

### *5.2 Superresoluutiomikroskopian haasteet ja niiden ylitse pääseminen*

Reaaliaikaisten sitoutumisten tutkiminen yksittäismolekyylitasolla vaatii hyvin optimoidut näytteiden valmistus-, kuvantamis- ja data-analyysiprotokollat. Promoottorien tai mikroskooppilasin pintaan kiinnitettyjen fluoroforien on oltava riittävän harvassa, jotta niihin tapahtuvia sitoutumisia voidaan seurata luotettavasti. Promoottorien pitoisuudessa päädyttiin 10 pM pitoisuuteen, sillä korkeammilla pitoisuuksilla ei promoottoreja juuri erottanut toisistaan. Tämä johtaa myös mahdollisesti siihen, että joissain kohtaa mikroskooppilasia promoottorit ovat hyvin harvassa, mutta saatu data on käyttökelpoista ja laadukasta. Aikaa ei kuitenkaan ole hukattavaksi, sillä hapensitojaentsyymien lisätyä on yhdellä mikroskooppilasilla vain 20 minuuttia käyttöaikaa. Tämän jälkeen kuvauspuskurin pH laskee, eikä tulokset ole enää luotettavia. Kuvauspuskurista ei voi luopua, sillä ilman sitä fluoroforit sammuvat ennenaikaisesti. On liki sattumaa, mikäli mikroskooppilasi on täysin puhdas ja siihen kiinnitetyt fluoroforit ovat tarpeeksi tiheässä mutta tarpeeksi kaukana toisistaan. Seuraamalla ohjeita ja olemalla tarpeeksi huolellinen onnistuivat välttämään kontaminaatiot ja saamaan käyttökelpoista, laadukasta dataa.

Tämän jälkeen oli määritettävä parhaat mahdolliset laserien parametrit kaikille fluoroforeille, jotta tulokset ovat luotettavia. Tämä oli käytännössä kokeilun ja erheen avulla tehtävä, sillä ennen pääsyä analyysiohjelmalle ei voi tietää, onko signaalin ja taustan ero tarpeeksi hyvä signaalin erottamiseksi. Silmämääräisesti sitä voi toki arvioida, mutta liian pitkä valotusaika johtaa fluoroforien ennenaikaiseen sammumiseen. Tällöin pidempiä sitoutumisia ei välttämättä havaita lainkaan. Sekä TMR-starin että DyLight650:n tapauksessa päädyin 50 ms valotusaikaan, joka on nopein mahdollinen valotusaika sitoutumisten havaitsemiseen. Yksin RRN3:n tapauksessa käytin paljon lyhyempää valotusaikaa. Tämä johtui siitä, ettei useamman kuvan ajan pysyviä sitoutumisia havaittu pidemmällä valotusajoilla ja lyhyemmällä valotusajoilla ei havaittu mitään.

Suurimmaksi haasteeksi muodostui tämän massiivisen datan analysointi. Koska samanlaista analyysiä ei ole tehty työryhmässämme saatavilla olevilla ohjelmistoilla, täytyi siihen löytää oikeat työkalut, jotka lopulta löysinkin. Valitettavasti vapaasti

saatavista avoimen lähdekoodin (engl. *open source*) ohjelmista joita kokeilin ei täysin vastannut siihen mitä halusin tehdä. Kaikissa oli myös omat rajoitteensa, sillä suuri osa yksittäismolekyylifluoresenssiohjelmistoista on kirjoitettu MATLAB-ohjelmalla (MathWorks) ja niiden yksittäisiksi ohjelmiksi kasatut versiot ovat usein kömpelöitä, eivät kykene käyttämään tietokoneen tarjoamaa kapasiteettia tai kaatuvat ilman mitään erikoista syytä. Nämä toki ovat kaikilla avoimen lähdekoodin ohjelmissa ongelmana, sillä kukaan ei alituisen valvo näiden toimintaa, vaan ne perustuvat yhteisön panokseen. Toisin kuin muissa vapaan lähdekoodin ohjelmissa, MATLAB-pohjaisissa yksittäisiksi kasatuissa ohjelmissa ei näe mahdollista virheen syytä. Kaikesta tästä huolimatta löysin tarkoituksiini parhaiten sopivan ohjelman, iSMS:n. iSMS tarjosi kaiken sen mitä tarvittiinkin eli yksittäisten promoottorien kohdalla kyettiin seuraamaan proteiinin sitoutumista ja valikoimaan hyvät signaalit. Hyvien signaalien keräämisen jälkeen tulee eteen vaikein vaihe: varsinaisten analyysien tekeminen. Keräämiäni signaalien analysointiin ei soveltunut mikään löytämäni avoimen lähdekoodin ohjelmista. Onneksi yhteistyökumppanimme David Dulinin laboratoriossa oli kehitetty tarpeisiini sopiva ohjelma, jonka kirjoitti Python-koodikielellä Subhas Chandra Bera. Sen käyttäminen vaati kuitenkin jokaisen signaalin muokkaamisen ohjelmalle sopivaan muotoon. Tavoitteena oli saada monta signaalia vierekkäin samasta mittauksesta, jotta käsittely olisi nopeampaa. Tätä varten opettelin ja kirjoitin kaksi Excel-makroa, sillä tämän tekeminen käsin oli ja on liian aikaa vievää. Näillä makroilla (liite 2) tämän käsittelyvaiheen kesto lyheni noin 6-8 tunnista alle tuntiin. Tämän jälkeen pystyttiin käyttämään tätä signaaliratojen erottamiseen kirjoitettua ohjelmaa, joka siis tunnisti signaaliradan muutoksen alun ja lopun ja keston. Tätä kestoa kyettiin käyttämään itsessään olemassa olevalla Origin-ohjelmistolla, mutta haasteeksi osoittautui vähäisten, alle 400 sitoutumistapahtumaa sisältävien datojen käsittely. Yhteistyökumppanimme David Dulinin laboratoriossa oli kehitetty tähän sopiva ohjelma, tekijänä Subhas Chandra Bera. EFP ei ryhmitä dataa, vaan käyttää jokaista datapistettä luodessaan sovitusta. EFP luo samanlaisen histogrammikuvaajan kuin Origin, mutta tuo kuvaaja ei vaikuta datan analyysiin. EFP kuitenkin arvioi jokaisen visualisoimansa dataluokan luotettavuusrajan bootstrap-menetelmää. Näistä syistä päätin käyttää EFP ohjelmaa sitoutumisaikojen arviointiin Origin-ohjelman kanssa silloin, kun sitoutumistapahtumia oli vähän.

### *5.3 Pol I välttämättömän aloituskompleksin osatekijöiden sitoutuminen promoottoriin on dynaamista*

Saadut sitoutumisten arvot ovat niitä aikoja, joita yksittäiset proteiinit viettävät yksittäisillä promoottoreilla. Koska mikään proteiinikomplekseista ei ylittänyt käytettyjen leimojen elinaikoja, on turvallista sanoa sitoutumisten olevan dynaamisia. Mikäli proteiinikompleksien sitoutumisajat olisivat olleet pidempiä tai yhtä pitkiä kuin leimojen elinajat, ei sitoutumisen pysyvyyttä pystyittäisi luotettavasti arvioimaan. Suurimmalla osalla seuratuista signaaleista tapahtui vain yksi sitoutumistapahtuma koko filmin aikana, minkä takia sitoutumisvaiheen nopeutta ei kyetty selvittämään. Jatkossa sitoutumisvaiheen kinetiikkaa voidaan pyrkiä selvittämään käyttämällä suurempaa proteiinipitoisuutta, mikä lisää peräkkäisten sitoutumisten todennäköisyyttä samaan promoottorimolekyylisiin. Tämä saattaa myös lisätä epäspesifisten sitoutumisten määrää. Sitoutumisten kestot pystyin jo tässä työssä selvittämään.

Tulosteni mukaan RRN3 yksistään ei sitoudu DNA:han (kuva 23). Tämä tukee myös rakennekuvista huomattavaa RRN3:n kaukaista asemaa promoottoriin nähden (kuva 10). CF:llä vaikuttaisi olevan lyhyt ja pitkä sitoutumistapa promoottoriin (kuva 24), mutta pienellä CF:n pitoisuudella ei muuhun DNA:han. Tämä osin tukee hiljattain havaittua CF:n sitoutumista Pol I promoottorin muotoon (Jackobiel ja muut, 2019). Pol I:llä itselläänkin vaikuttaa sitoutuvan itsestään promoottoriin ja se sitoutuu kahdessa vaiheessa, kuten CF (kuva 25). Tämä sitoutuminen saattaa olla yleistä DNA:han sitoutumista, sillä tämä tulos on mitattu vain promoottorilla. Pol I sitoutuu promoottoriin pidemmäksi aikaa kuin CF; CF:n sitoutumisajat ovat 514 ms ja 3,60 s (Originin antamat arvot), kun taas Pol I:n sitoutumisajat ovat 2,10 s ja 19,0 s. Tämä tulos voi kuitenkin olla harhaan johtava, sillä Pol I oli satunnaisesti leimattu kysteiniitähteistään, jolloin tämä voi vaikuttaa Pol I:n toimintaan ja tulosten analysointiin. TBP:n tai UAF:n sitoutumista yksin Pol I promoottoriin ei tämän työn puitteissa ehditty tehdä.

Välttämätön aloituskompleksi, eli Pol I:stä, RRN3:sta ja CF:stä koostuva proteiinikompleksi, sitoutuu erilaisilla ajoilla riippuen leiman sisältämästä proteiinista (kuva 26). Valitettavasti kaksivärikokeita ei ehditty tekemään, ja varsinkin nopeimpien sitoutumisaikojen osalta se saattaisi olla mahdotonta tehdä tarkasti DeltaVision OMX:llä. Kun Pol I:n välttämättömässä aloituskompleksissa oli leimaproteiinina RRN3, sen sitoutumisajat ovat 270 ms ja 28,8 s. Tässä analyysissä on muista poiketen käytetty EFP-analyysiä, sillä Origin-analyysin datapisteiden ryhmittäminen vääristää tulosta (kuva 26

A ja B). Tämä voi myös olla pelkän Pol I – RRN3-kompleksin sitoutumista eikä välttämättömän aloituskompleksin sitoutumista, sillä Pol I ja RRN3 muodostavat yhdessä holoentsyymin (Pils1 ja muut, 2016). RRN3:n määritetty sitoutumisaika on 61,6 ms (EFP:n antama arvo), mutta todennäköisesti kyseessä on vain satunnaisesti pintaan sitoutuvia molekyyliä. Tässä tapauksessa Pol I aloituskompleksin sitoutumisajat ovat luotettavia, mutta on mahdollista, että nopeat (alle viisi kuvaa kestävät sitoutumiset) ovat ylliedustettuina datassa (kuva 26 A). On myös mahdollista, että RRN3 hakeutuu nopeasti DNA:n läheisyyteen, mutta ei jää sinne muiden proteiinien puuttuessa. Määritettäessä välttämättömän aloituskompleksin sitoutumisaika leimatun CF:n avulla, olivat sitoutumisajat EFP-analyysillä 3,19 s ja 22,4 s (kuva 26 D). Tässä ensimmäinen sitoutuminen on paljon pidempi kuin RRN3:n osoittamassa välttämättömän aloituskompleksin sitoutumisessa, mutta se on myös pidempi kuin CF:n itsensä sitoutuminen promoottoriin. Mahdollisesti Pol I tai RRN3 tai molemmat vakauttavat CF:n sitoutumista promoottoriin. Toinen sitoutuminen on leimatun CF:n tapauksessa hivenen pidempi kuin leimatun RRN3:n tapauksessa. Pidempi sitoutuminen voi teoriassa olla Pol I:n avoimen kompleksin muodostaminen tai promoottorin sitoutuminen, kun taas lyhyempi sitoutuminen voi olla Pol I:n suljetun kompleksin muodostaminen tai promoottorin ulkopuolelle sitoutumista tapauksessa, jossa jokin proteiini puuttuu. Origin-analyysi antoi hyvin poikkeavat arvot verrattuna EFP-analyysiin, sillä Origin-analyysin mukaan välttämättömän aloituskompleksin sitoutumisaikojen kestot olivat 482 ms ja 4,65 s (kuva 26 C). Näistä ensimmäinen on nopeampi kuin yksittäisen CF:n sitoutuminen promoottoriin (514 ms), mutta ero on todella pieni. Toinen vaihe on välttämättömällä aloituskompleksilla pidempi kuin yksittäisen CF:n sitoutuminen (3,60 s), mutta ero on vain yhden sekunnin. Tämä kaikki osoittaa sen, että kaksiväritutkimukset ovat tarpeellisia varsinaisen sitoutumisajan määrittämiseksi, mahdollisesti myös kolmiväritutkimukset. Tämä vaatisi myös Pol I:n leimamisen, joko kohdennetusti tai avustavan proteiinin, kuten SNAP-proteiinin, avulla. Tästäkin huolimatta, välttämätön aloituskompleksi tukee RRN3:n ja CF:n sitoutumista promoottorille.

Täydellistä aloituskompleksia ei päästy tutkimaan, sillä ongelmat UAF:n tuotossa kestivät pitkään. Kuitenkin Olena Parilova onnistui tuottamaan UAF:ää kolibakteerissa työni loppupuolella, joten ehdin tutkia, miten UAF ja TBP yhdessä vaikuttavat CF:n sitoutumiseen promoottoriin (kuva 27). UAF:n ja TBP:n läsnäollessa CF sitoutui promoottoriin yksivaiheisesti ja tämän sitoutumisen kesto oli 6,77 s (Origin-analyysillä).



Yksinään CF:n pidempi sitoutumisaika on 3,60 s, joten UAF ja TBP tai toinen näistä vakauttaa CF:n sitoutumista promoottoriin. Vaikuttaako UAF tai TBP yksistään CF:n sitoutumiseen vai onko kyseessä näiden yhteistyö? Tämä on yksi kysymyksistä, jota ei tässä työssä ehditty selvittää.

*5.4 Luotujen menetelmien avulla Pol I aloituskompleksin sitoutumisjärjestyksen selvittäminen on mahdollista*

Tässä pro gradu-työssä luotiin pohja Pol I aloituskompleksin sitoutumisen tutkimiseen. Kaikki lähti ideasta selvittää Pol I:n aloituskompleksin proteiinien sitoutumisjärjestys, sillä sitä ei ole aiemmin tehty. Ennen tätä on selvitettävä olemassa olevilla proteiineilla ja leimoilla, miten TBP sitoutuu promoottoriin, vakauttaako UAF TBP:n sitoutumista promoottoriin ja onko Pol I täydellisen aloituskompleksin sitoutuminen promoottoriin dynaamista vai vakautuuko aloituskompleksi paikalleen, kunnes se aloittaa RNA-synteesin NTP:n läsnäollessa. Näiden lisäksi on vielä määritettävä kontrollit jokaiselle proteiinille RRN3:a ja CF:ää lukuun ottamatta. RRN3 ei itsestään sitoudu DNA:han ja CF ei pienessä pitoisuudessa sitoudu satunnaiseen DNA:han, mutta sitoutuu promoottoriin. Tämän työn pohjalta on mahdollista selvittää Pol I aloituskompleksin sitoutumisjärjestys promoottoriin ja tutkia aloituskompleksin proteiinien välisiä etäisyyksiä ja rakenteellista dynamiikkaa hyödyntäen FRET-tekniikkaa.

## 6. Yhteenveto

Tässä työssä kehitettiin pohja Pol I:n aloituskompleksin tutkimukselle. Se aloitettiin proteiinien tuotosta ja puhdistuksesta ja näiden karakterisoinnista. Tämän jälkeen suunniteltiin proteiinien kohdennettu leimaus, jotta niitä voidaan tutkia yksittäismolekyylifluoresenssilla reaaliaikaisesti. Kaikissa kohdennetuissa leimauksissa ei onnistuttu, jolloin siirryttiin FRET:in tutkimisen sijaan yhteislokalisaation tutkimiseen ja osa proteiinien leimauksista muutettiin siihen sopivaksi. Näiden leimausstrategioiden ansiosta pääsin tutkimaan proteiinien sitoutumista promoottoriin reaaliaikaisesti superresoluutiomikroskopiolla. Suureksi esteeksi muodostui tämän datan käsittely ja analyysi, joiden selvittämiseen meni kauan. Lopulta kaikki tämä selvitettiin ja näin luotiin pohja Pol I:n aloituksen tutkimiselle. Tulevaisuudessa täytyy selvittää, miten täydellisen aloituskompleksin proteiinit sitoutuvat promoottorille ja mikä vaikutus UAF:llä ja TBP:llä on reaaliaikaiseen sitoutumiseen. Tämän jälkeen pystytään selvittämään Pol I:n aloituskompleksin sitoutumisjärjestys promoottoriin.

## 7. Lähteet

- Abascal-Palacios, G., Ramsay, E. P., Beuron, F., Morris, E. & Vannini, A. (2018). Structural basis of RNA polymerase III transcription initiation. *Nature*, **553**, 301–306
- Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature*, **438**, 460–465
- Adelman, K. & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.*, **13**, 720–731
- Akhtar, J., Kreim, N., Marini, F., Mohana, G., Brüne, D., Binder, H. & Roignant, J. Y. (2019). Promoter-proximal pausing mediated by the exon junction complex regulates splicing. *Nat. Commun.*, **10**:
- Andandre, M. A. & Bork, P. (1995). HEAT repeats in the Huntington's disease protein. *Nat. Genet.*, **11**, 115–116
- Aprikian, P., Moorefield, B. & Reeder, R. H. (2000). TATA binding protein can stimulate core-directed transcription by yeast RNA polymerase I. *Mol Cell Biol*, **20**, 5269–5275
- Aprikian, P., Moorefield, B. & Reeder, R. H. (2001). New model for the yeast RNA polymerase I transcription cycle. *Mol. Cell. Biol.*, **21**, 4847
- Artsimovitch, I. & Belogurov, G. A. (2018). Uneven braking spins RNA polymerase into a pause. *Mol. Cell*, **69**, 723–725
- Artsimovitch, I. & Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 7090–7095
- Banik, U., Beechem, J. M., Klebanow, E., Schroeder, S. & Weil, P. A. (2001). Fluorescence-based analyses of the effects of full-length recombinant TAF130p on the interaction of TATA box-binding protein with TATA box DNA. *J. Biol. Chem.*, **276**, 49100–49109
- Blattner, C., Jennebach, S., Herzog, F., Mayer, A., Cheung, A. C. M., Witte, G., Lorenzen, K., Hopfner, K. P., Heck, A. J. R., Aebersold, R. & Cramer, P. (2011). Molecular basis

of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev.*, **25**, 2093–2105

Botstein, D., Chervitz, S. A. & Cherry, J. M. (1997). Yeast as a model organism. *Science*, **277**, 1259–1260

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.*, **68**, 394–424

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D. & Heumann, H. (2004). The  $\sigma$ 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. *Nat. Struct. Mol. Biol.*, **11**, 551–557

Bruno, P. M., Lu, M., Dennis, K. A., Inam, H., Moore, C. J., Sheehe, J., Elledge, S. J., Hemann, M. T. & Pritchard, J. R. (2020). The primary mechanism of cytotoxicity of the chemotherapeutic agent CX-5461 is topoisomerase II poisoning. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 4053–4060

Carafa, Y. d'Aubenton, Brody, E. & Thermes, C. (1990). Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.*, **216**, 835–858

Cavanaugh, A. H., Hirschler-Laszkiewicz, I., Hu, Q., Dundr, M., Smink, T., Misteli, T. & Rothblum, L. I. (2002). Rrn3 phosphorylation is a regulatory checkpoint for ribosome biogenesis. *J. Biol. Chem.*, **277**, 27423–27432

Chakraborty, A., Mazumder, A., Lin, M., Hasaymayer, A., Xu, Q., Wang, D., Ebright, Y. W. & Ebright, R. H. (2015). Site-specific incorporation of probes into RNA polymerase by unnatural-amino-acid mutagenesis and Staudinger-Bertozzi ligation. *Bacterial Transcriptional Control. Methods in Molecular Biology*. Humana Press, New York.

Chandradoss, S. D., Haagsma, A. C., Lee, Y. K., Hwang, J. H., Nam, J. M. & Joo, C. (2014). Surface passivation for single-molecule protein studies. *J. Vis. Exp.*, **86**, 1–8

- Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. (1993). Crystal structure of yeast TATA-binding protein: a model for interaction with DNA. *Proc. Natl. Acad. Sci.*, **90**, 8174–8178
- Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. (2002). Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.*, **124**, 9026–9027
- Choe, S. Y., Schultz, M. C. & Reeder, R. H. (1992). In vitro definition of the yeast RNA polymerase I promoter. *Nucleic Acids Res.*, **20**, 279–285
- Comai, L., Zomerdijk, J. C. B. M., Beckmann, H., Zhou, S., Admon, A. & Tjian, R. (1994). Reconstitution of transcription factor SL1: Exclusive binding of TBP by SL1 or TFIID subunits. *Science*, **266**, 1966–1972
- Core, L. & Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.*, **33**, 960–982
- Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J. F. & Vannini, A. (2008). Structure of eukaryotic RNA Polymerases. *Annu. Rev. Biophys.*, **37**, 337–352
- Cramer, P., Bushnell, D. A. & Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science*, **292**, 1863–1876
- Defossez, P. A., Prusty, R., Kaeberlein, M., Lin, S. J., Ferrigno, P., Silver, P. A., Keil, R. L. & Guarente, L. (1999). Elimination of replication block protein Fob1 extends the life span of yeast mother cells. *Mol. Cell*, **3**, 447–455
- DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.*, **40**, 82–92
- Drouin, G. & Carter, R. (2010). Evolution of eukaryotic RNA Polymerases. *Encycl. Life Sci.*, September
- Drygin, D., Lin, A., Bliesath, J., Ho, C. B., O'Brien, S. E., Proffitt, C., Omori, M., Haddach, M., Schwaebe, M. K., Siddiqui-Jain, A., Streiner, N., Quin, J. E., Sanij, E.,

- Bywater, M. J., Hannan, R. D., Ryckman, D., Jaeres, K. & Rice, W. G. (2011). Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res.*, **71**, 1418–1430
- Drygin, D., Rice, W. G. & Grummt, I. (2010). The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu. Rev. Pharmacol. Toxicol.*, **50**, 131–156
- Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C. B., Proffitt, C., Trent, K., Whitten, J. P., Lim, J. K. C., Von Hoff, D., Jaeres, K. & Rice, W. G. (2009). Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. *Cancer Res.*, **69**, 7653–7661
- Dulin, D., Bauer, D. L. V., Malinen, A. M., Bakermans, J. J. W., Kaller, M., Morichaud, Z., Petushkov, I., Depken, M., Brodolin, K., Kulbachinskiy, A. & Kapanidis, A. N. (2018). Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nat. Commun.*, **9**, 1478
- Dvir, A. (2002). Promoter escape by RNA polymerase II. *Biochem. Biophys. Acta*, **1577**, 208–223
- Engel, C., Gubbey, T., Neyer, S., Sainsbury, S., Oberthuer, C., Baejen, C., Bernecky, C. & Cramer, P. (2017). Structural basis of RNA polymerase I transcription initiation. *Cell*, **169**, 120-131.e22
- Engel, C., Plitzko, J. & Cramer, P. (2016). RNA polymerase I-Rrn3 complex at 4.8 Å resolution. *Nat. Commun.*, **7**, 3–7
- Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D. & Cramer, P. (2013). RNA polymerase I structure ja transcription regulation. *Nature*, **502**, 650–655
- Fant, C. B., Levandowski, C. B., Gupta, K., Maas, Z. L., Moir, J., Rubin, J. D., Sawyer, A., Esbin, M. N., Rimel, J. K., Luyties, O., Marr, M. T., Berger, I., Dowell, R. D. & Taatjes, D. J. (2020). TFIID enables RNA polymerase II promoter-proximal pausing. *Mol. Cell*, 1–9

- Fernández-Tornero, C., Moreno-Morcillo, M., Rashid, U. J., Taylor, N. M. I., Ruiz, F. M., Gruene, T., Legrand, P., Steuerwald, U. & Müller, C. W. (2013). Crystal structure of the 14-subunit RNA polymerase I. *Nature*, **502**, 644–649
- Ferreira, R., Schneekloth, J. S., Panov, K. I., Hannan, K. M. & Hannan, R. D. (2020). Targeting the RNA polymerase I transcription for cancer therapy comes of age. *Cells*, **9**, 266
- Geiger, S. R., Lorenzen, K., Schrieck, A., Hanecker, P., Kostrewa, D., Heck, A. J. R. & Cramer, P. (2010). RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol. Cell*, **39**, 583–594
- Ghaemmaghami, S., Huh, W., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O’Shea, E. K. & Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, **425**, 737–741
- Gietz, R. D. & Schiestl, R. H. (2007). Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 1–4
- Goldman, S. R., Ebright, R. H. & Nickels, B. E. (2009). Direct detection of abortive RNA transcripts in vivo. *Science*, **324**, 927–928
- Goldstein, A. L. & McCusker, J. H. (1999). Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast*, **15**, 1541–1553
- Gusarov, I. & Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Mol. Cell*, **3**, 495–504
- Ha, C. W., Sung, M. K. & Huh, W. K. (2012). Nsi1 plays a significant role in the silencing of ribosomal DNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **40**, 4892–4903
- Han, Y., Yan, C., Nguyen, T. H. D., Jackobel, A. J., Ivanov, I., Knutson, B. A. & He, Y. (2017). Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. *Elife*, **6**, 1–24

- Henderson, K. L., Evensen, C. E., Molzahn, C. M., Felth, L. C., Dyke, S., Liao, G., Shkel, I. A. & Record, M. T. (2019). RNA polymerase: step-by-step kinetics and mechanism of transcription initiation. *Biochemistry*, **58**, 2339–2352
- Heyduk, E. & Heyduk, T. (2018). DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Res.*, **46**, 4469–4486
- Hontz, R. D., French, S. L., Oakes, M. L., Tongaonkar, P., Nomura, M., Beyer, A. L. & Smith, J. S. (2008). Transcription of multiple yeast ribosomal DNA genes requires targeting of UAF to the promoter by Uaf30. *Mol. Cell. Biol.*, **28**, 6709–6719
- Hsu, L. M. (2002). Promoter clearance and escape in prokaryotes. *Biochem. Biophys. Acta*, **1577**, 191–207
- Ishii, J., Kondo, T., Makino, H., Ogura, A., Matsuda, F. & Kondo, A. (2014). Three gene expression vector sets for concurrently expressing multiple genes in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **14**, 399–411
- Jackobel, A. J., Zeberl, B. J., Glover, D. M., Fakhouri, A. M. & Knutson, B. A. (2019). DNA binding preferences of *S. cerevisiae* RNA polymerase I Core Factor reveal a preference for the GC-minor groove and a conserved binding mechanism. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1862**, 194408
- Jansa, P. & Grummt, I. (1999). Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. *Mol. Gen. Genet.*, **262**, 508–514
- Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H. & Seidel, C. A. M. (2012). A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods*, **9**, 1218–1225
- Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S. & Ebright, R. H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, **314**, 1144–1147
- Kassavetis, G. A., Letts, G. A. & Geiduschek, E. P. (1999). A minimal RNA polymerase III transcription system. *EMBO J.*, **18**, 5042–5051



- Kassavetis, G. A., Letts, G. A. & Geiduschek, E. P. (2001). The RNA polymerase III transcription initiation factor TFIIIB participates in two steps of promoter opening. *EMBO J.*, **20**, 2823–2834
- Keener, J., Dodd, J. A., Lalo, D. & Nomura, M. (1997). Histones H3 ja H4 are components of upstream activation factor required for the high-level transcription of yeast rDNA by RNA polymerase I. *Proc. Natl. Acad. Sci.*, **94**, 13458–13462
- Keener, J., Josaitis, C. A., Dodd, J. A. & Nomura, M. (1998). Reconstitution of yeast RNA polymerase I transcription in vitro from purified components. *J. Biol. Chem.*, **273**, 33795–33802
- Keys, D. A., Lee, B. S., Dodd, J. A., Nguyen, T. T., Vu, L., Fantino, E., Burson, L. M., Nogi, Y. & Nomura, M. (1996). Multiprotein transcription factor UAF interacts with the upstream element of the yeast RNA polymerase I promoter and forms a stable preinitiation complex. *Genes Dev.*, **10**, 887–903
- Khot, A., Brajanovski, N., Cameron, D. P., Hein, N., Maclachlan, K. H., Sanij, E., Lim, J., Soong, J., Link, E., Blombery, P., Thompson, E. R., Fellowes, A., Sheppard, K. E., McArthur, G. A., Pearson, R. B., Hannan, R. D., Poortinga, G. & Harrison, S. J. (2019). First-in-human RNA polymerase I transcription inhibitor CX-5461 in patients with advanced hematologic cancers: Results of a phase I dose-escalation study. In *Cancer Discov.*, **8**, 1036-1049
- Kim, C. H., Axup, J. Y. & Schultz, P. G. (2013). Protein conjugation with genetically encoded unnatural amino acids. *Curr. Opin. Chem. Biol.*, **17**, 412–419
- Klein, C. & Struhl, K. (1994). Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. *Science*, **266**, 280–282
- Kuehner, J. N., Pearson, E. L. & Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.*, **12**, 283–294
- Kugel, J. F. & Goodrich, J. A. (2002). Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. *Mol. Cell. Biol.*, **22**, 762–773

- Kuhn, C. D., Geiger, S. R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R. & Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell*, **131**, 1260–1272
- Kulkens, T., Riggs, D. L., Heck, J. D., Planta, R. J. & Nomura, M. (1991). The yeast RNA polymerase I promoter: Ribosomal DNA sequences involved in transcription initiation and complex formation in in vitro. *Nucleic Acids Res.*, **19**, 5363–5370
- Kwak, H. & Lis, J. T. (2013). Control of transcriptional elongation. *Annu. Rev. Genet.*, **47**, 483–508
- Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H., Haimovich, A. D., Kuznetsov, G., Mercer, J. a, Wang, H. H., Carr, P. a, Mosberg, J. a, Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M. & Isaacs, F. J. (2013). Genomically recoded organisms expand biological functions. *Science*, **342**, 357–360
- Lalo, D., Steffan, J. S., Dodd, J. A. & Nomura, M. (1996). RRN11 encodes the third subunit of the complex containing Rrn6p and Rrn7p that is essential for the initiation of rDNA transcription by yeast RNA polymerase I. *J. Biol. Chem.*, **271**, 21062–21067
- Lam, A. J., St-Pierre, F., Gong, Y., Marshall, J. D., Cranfill, P. J., Baird, M. A., McKeown, M. R., Wiedenmann, J., Davidson, M. W., Schnitzer, M. J., Tsien, R. Y. & Lin, M. Z. (2012). Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat. Methods*, **9**, 1005–1012
- Lang, W. H. & Reeder, R. H. (1995). Transcription termination of RNA polymerase I due to a T-rich element interacting with Reb1p. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 9781–9785
- Learned, R. M., Learned, T. K., Haltiner, M. M. & Tijan, R. T. Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell*, **45**, 847-857
- Li, X. Y. & McClure, W. R. (1998). Characterization of the closed complex intermediate formed during transcription initiation by Escherichia coli RNA polymerase. *J. Biol. Chem.*, **273**, 23549–23557

- Lisica, A., Engel, C., Jahnel, M., Roldán, É., Galburt, E. A., Cramer, P. & Grill, S. W. (2016). Mechanisms of backtrack recovery by RNA polymerases I and II. *Proc. Natl. Acad. Sci.*, **113**, 2946–2951
- Längst, G., Blank, T. A., Becker, P. B. & Grummt, I. (1997). RNA polymerase I transcription on nucleosomal templates: The transcription termination factor TTF-I induces chromatin remodeling and relieves transcriptional repression. *EMBO J.*, **16**, 760–768
- Malinen, A. M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M. S. & Belogurov, G. A. (2012). Active site opening and closure control translocation of Multisubunit RNA polymerase. *Nucleic Acids Res.*, **40**, 7442–7451
- Merkel, P., Perez-Fernandez, J., Pilsl, M., Reiter, A., Williams, L., Gerber, J., Bohm, M., Deutzmann, R., Griesenbeck, J., Milkereit, P. & Tschochner, H. (2014). Binding of the termination factor Nsi1 to its cognate DNA site is sufficient to terminate RNA polymerase I transcription in vitro and to induce termination in vivo. *Mol. Cell. Biol.*, **34**, 3817–3827
- Milkereit, P. & Tschochner, H. (1998). A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *EMBO J.*, **17**, 3692–3703
- Miller, G., Panov, K. I., Friedrich, J. K., Trinkle-Mulcahy, L., Lamond, A. I. & Zomerdijs, J. C. B. M. (2001). hRRN3 is essential in the SL1-mediated recruitment of RNA polymerase I to rRNA gene promoters. *EMBO J.*, **20**, 1373–1382
- Miller, O. L. & Beatty, B. R. (1969). Visualization of nucleolar genes. *Science*, **164**, 955–957
- Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., Yokoyama, S. & Sakamoto, K. (2015). Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. *Sci. Rep.*, **5**:
- Musters, W., Knol, J., Mass, P., Dekker, A. F., van Heerikhuizen, H. & Planta, R. J. (1989). Linker scanning of the yeast RNA polymerase I promoter. *Nucleic Acids Res.*, **17**, 9661–9678

- Németh, A., Guibert, S., Tiwari, V. K., Ohlsson, R. & Längst, G. (2008). Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. *EMBO J.*, **27**, 1255–1265
- Németh, A., Perez-Fernandez, J., Merkl, P., Hamperl, S., Gerber, J., Griesenbeck, J. & Tschochner, H. (2013). RNA polymerase I termination: Where is the end? *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1829**, 306–317
- Neyer, S., Kunz, M., Geiss, C., Hantsche, M., Hodirnau, V. V., Seybert, A., Engel, C., Scheffer, M. P., Cramer, P. & Frangakis, A. S. (2016). Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature*, **540**, 607–610
- Nogi, Y., Loan, V. U. & Nomura, M. (1991). An approach for isolation of mutants defective in 35S ribosomal RNA synthesis in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.*, **88**, 7026–7030
- Nudler, E. (2012). RNA polymerase backtracking in gene regulation ja genome instability. *Cell*, **149**, 1438–1445
- Peltonen, K., Colis, L., Liu, H., Jäämaa, S., Zhang, Z., Hällström, T. A., Moore, H. M., Sirajuddin, P. & Laiho, M. (2014). Small molecule BMH-compounds that inhibit RNA polymerase I and cause nucleolar stress. *Mol. Cancer Ther.*, **13**, 2537–2546
- Perri, F., Pisconti, S., Scrapati, G. D. V. (2016). P53 mutations and cancer: a tight linkage. *Ann. Transl. Med.* **4**: 522
- Petes, T. D. (1979). Yeast ribosomal DNA genes are located on chromosome XII. *Proc. Natl. Acad. Sci. U. S. A.*, **76**, 410–414
- Petushkov, I., Esyunina, D. & Kulbachinskiy, A. (2017).  $\sigma^{38}$ -dependent promoter-proximal pausing by bacterial RNA polymerase. *Nucleic Acids Res.*, **45**, 3006–3016
- Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C. & Riva, M. (2000). The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J.*, **19**, 5473–5482

- Pils, M., Crucifix, C., Papai, G., Krupp, F., Steinbauer, R., Griesenbeck, J., Milkereit, P., Tschochner, H. & Schultz, P. (2016). Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nat. Commun.*, **7**, 1–12
- Pils, M. & Engel, C. (2020). Structural basis of RNA polymerase I pre-initiation complex formation and promoter melting. *Nat. Commun.*, **11**, 1–10
- Prescott, E. M., Osheim, Y. N., Jones, H. S., Alen, C. M., Roan, J. G., Reeder, R. H., Beyer, A. L. & Proudfoot, N. J. (2004). Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 6068–6073
- Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. (2015). iSMS: Single-molecule FRET microscopy software. *Nat. Methods*, **12**, 593–594
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. & Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **229**, 218–229
- Ray-Soni, A., Bellecourt, M. J. & Landick, R. (2016). Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.*, **85**, 319–347
- Reines, D., Conaway, J. W. & Conaway, R. C. (1996). The RNA polymerase II general elongation factors. *Trends Biochem. Sci.*, **21**, 351–355
- Reiter, A., Hamperl, S., Seitz, H., Merkl, P., Perez-Fernandez, J., Williams, L., Gerber, J., Németh, A., Léger, I., Gadal, O., Milkereit, P., Griesenbeck, J. & Tschochner, H. (2012). The Reb1-homologue Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. *EMBO J.*, **31**, 3480–3493
- Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, **314**, 1139–1143
- Roberts, J. W., Shankar, S. & Filter, J. J. (2008). RNA polymerase elongation factors. *Annu. Rev. Microbiol.*, **62**, 211–233
- Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.*, **21**, 327–335

- Rothblum, K., Hu, Q., Penrod, Y. & Rothblum, L. I. (2014). Selective inhibition of rDNA transcription by a small-molecule peptide that targets the interface between RNA polymerase I and Rrn3. *Mol. Cancer Res.*, **12**, 1586–1596
- Ruff, E. F., Thomas Record, M. & Artsimovitch, I. (2015). Initial events in bacterial transcription initiation. *Biomolecules*, **5**, 1035–1062
- Russell, J. & Zomerdijk, J. C. B. M. (2006). The RNA polymerase I transcription machinery. *Biochem. Soc. Symp.*, **73**, 203–216
- Sadian, Y., Baudin, F., Tafur, L., Murciano, B., Wetzel, R., Weis, F. & Müller, C. W. (2019). Molecular insight into RNA polymerase I promoter recognition and promoter melting. *Nat. Commun.*, **10**, 5543
- Sadian, Y., Tafur, L., Kosinski, J., Jakobi, A. J., Wetzel, R., Buczak, K., Hagen, W. J., Beck, M., Sachse, C. & Müller, C. W. (2017). Structural insights into transcription initiation by yeast RNA polymerase I. *EMBO J.*, **36**, e201796958
- Saunders, A., Core, L. J. & Lis, J. T. (2006). Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.*, **7**, 557–567
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nat. Methods*, **9**, 676–682
- Schneider, D. A. (2012). RNA polymerase I activity is regulated at multiple steps in the transcription cycle: Recent insights into factors that influence transcription elongation. *Gene*, **493**, 176–184
- Scull, C. E., Ingram, Z. M., Lucius, A. L. & Schneider, D. A. (2019). A novel assay for RNA polymerase I transcription elongation sheds light on the evolutionary divergence of eukaryotic RNA polymerases. *Biochemistry*, **58**, 2116–2124
- Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, **426**, 684–687

- Shilatifard, A., Conaway, R. C. & Conaway, J. W. (2003). The RNA polymerase II elongation complex. *Annu. Rev. Biochem.*, **72**, 693–715
- Siddiqi, I. N., Dodd, J. A., Vu, L., Eliason, K., Oakes, M. L., Keener, J., Moore, R., Young, M. K. & Nomura, M. (2001). Transcription of chromosomal rRNA genes by both RNA polymerase I and II in yeast *uaf30* mutants lacking the 30 kDa subunit of transcription factor UAF. *EMBO J.*, **20**, 4512–4521
- Smith, M. L., Cui, W., Jackobel, A. J., Walker-Kopp, N. & Knutson, B. A. (2018). Reconstitution of RNA polymerase I upstream activating factor and the roles of histones H3 and H4 in complex assembly. *J. Mol. Biol.*, **430**, 641–654
- Steffan, J. S., Keys, D. A., Dodd, J. A. & Nomura, M. (1996). The role of TBP in rDNA transcription by RNA polymerase I in *Saccharomyces cerevisiae*: TBP is required for upstream activation factor- dependent recruitment of core factor. *Genes Dev.*, **10**, 2551–2563
- Steffan, J. S., Keys, D. A., Vu, L. & Nomura, M. (1998). Interaction of TATA-binding protein with upstream activation factor is required for activated transcription of ribosomal DNA by RNA polymerase I in *Saccharomyces cerevisiae* In Vivo. *Mol. Cell. Biol.*, **18**, 3752–3761
- Steitz, T. A. (1998). A mechanism for all polymerases. *Nature*, **391**, 231–232
- Tafur, L., Sadian, Y., Hanske, J., Wetzel, R., Weis, F. & Müller, C. W. (2019). The cryo-EM structure of a 12-subunit variant of RNA polymerase I reveals dissociation of the A49-A34.5 heterodimer and rearrangement of subunit A12.2. *Elife*, **8**, 1–21
- Tafur, L., Sadian, Y., Hoffmann, N. A., Jakobi, A. J., Wetzel, R., Hagen, W. J. H., Sachse, C. & Müller, C. W. (2016). Molecular structures of transcribing RNA polymerase I. *Mol. Cell*, **64**, 1135–1143
- Treutlein, B., Muschielok, A., Jarecka, J., Jawhari, A., Buchen, C., Kostrewa, D., Hög, F., Cramer, P. & Michaelis, J. (2012). Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Mol. Cell*, **46**, 136–146

- Truong, C., Oudre, L. & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, **167**
- Turtola, M. & Belogurov, G. A. (2016). NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife*, **5**, 1–27
- Tyagi, R., Lai, R. & Duggleby, R. G. (2004). A new approach to “megaprimer” polymerase chain reaction mutagenesis without an intermediate gel purification step. *BMC Biotechnol.*, **4**
- Vannini, A. & Cramer, P. (2012). Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Mol. Cell*, **45**, 439–446
- Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, **448**, 157–162
- Viktorovskaya, O. V, Engel, K. L., French, S. L., Cui, P., Vandeventer, P. J., Pavlovic, E. M., Beyer, A. L., Kaplan, C. D. & Schneider, D. A. (2013). Divergent contributions of conserved active site residues to transcription by Eukaryotic RNA polymerases I and II. *Cell Rep.*, **4**, 974–984
- Vogel, S. S., Van der Meer, B. W. & Blank, P. S. (2014). Estimating the distance separating fluorescent protein FRET pairs. *Methods*, **66**, 131–138
- Wach, A., Brachat, A., Pöhlmann, R. & Philippsen, P. (1994). New heterologous Modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast*, **10**, 1793–1808
- Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. (2006). Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. *Cell*, **127**, 941–954
- Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.*, **24**, 437–440
- Werner, F. & Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.*, **9**, 85–98



Woychik, N. A. (2005). RNA polymerases ja the eukaryotic transcription machinery. *Encycl. Life Sci.*, 1–7

Woychik, N. A. & Reinberg, D. (2001). RNA polymerases: subunits and functional domains. *Encycl. Life Sci.*, 1–8

Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O’Neil, N. J., Santos, N. Dos, Silvester, J., Wei, V., Garcia, J., Kabeer, F., Lai, D., Soriano, P., Banáth, J., Chiu, D. S., Yap, D., Le, D. D., Ye, F. B., Zhang, A., Thu, K., Soong, J., Lin, S., Tsai, A. H. C., Osako, T., Algara, T., Saunders, D. N., Wong, J., Xian, J., Bally, M. B., Brenton, J. D., Brown, G. W., Shah, S. P., Cescon, D., Mak, T. W., Caldas, M., Stirling, P. C., Hieter, P., Balasubramanian, S., Aparicio, S. (2017). CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.*, **8**:

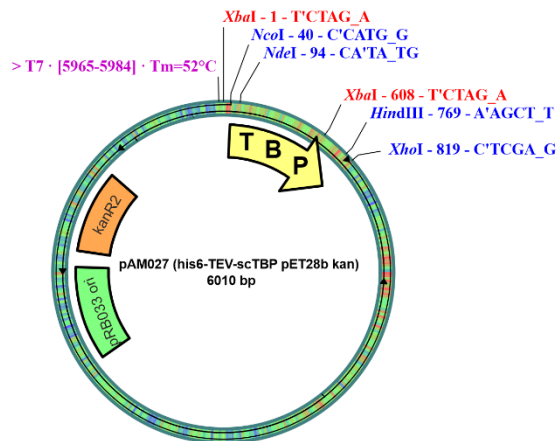
Yamamoto, R. T., Nogi, Y., Dodd, J. A. & Nomura, M. (1996). RRN3 gene of *Saccharomyces cerevisiae* encodes an essential RNA polymerase I transcription factor which interacts with the polymerase independently of DNA template. *EMBO J.*, **15**, 3964–3973

Zhou, B. R. & Bai, Y. (2019). Chromatin structures condensed by linker histones. *Essays Biochem.*, **63**, 75–87

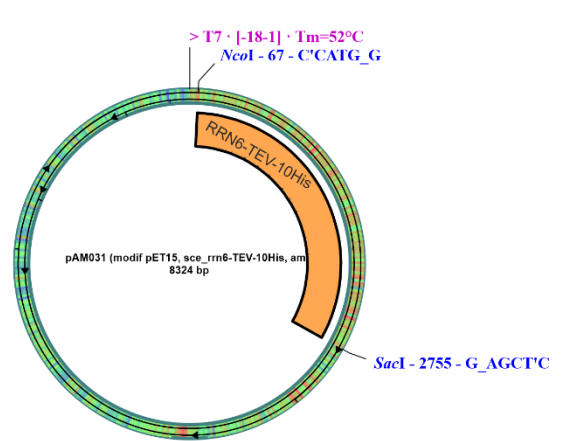
Zuo, Y. & Steitz, T. A. (2017). A structure-based kinetic model of transcription. *Transcription*, **8**, 1–8

## Liite 1 – Plasmidikartat

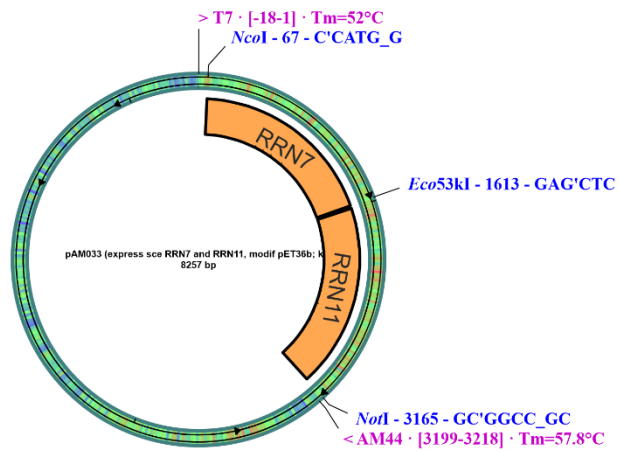
pAM027 (TBP-plasmidi)



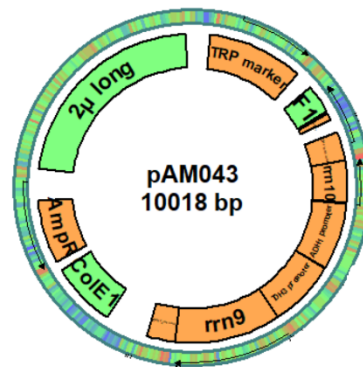
pAM031 (RRN6-plasmidi)



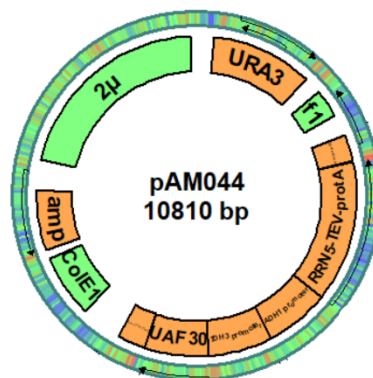
pAM033 (RRN7/RRN11-plasmidi)



pAM043 (UAF plasmidi 1)



pAM044 (UAF plasmidi 2)



## Liite 2 – Kirjoitetut Excel-makrot

*Makro 1: useiden tekstitiedostojen tuonti yhteen Excel-työkirjaan eri välilehdille.*

```
Sub CombineTextFiles()  
    Dim FilesToOpen  
    Dim x As Integer  
    Dim wkbAll As Workbook  
    Dim wkbTemp As Workbook  
    Dim sDelimiter As String  
  
    On Error GoTo ErrHjaler  
    Application.ScreenUpdating = False  
  
    sDelimiter = "|"   
  
    FilesToOpen = Application.GetOpenFilename _  
        (FileFilter:="Text files (*.txt), *.txt", _  
        MultiSelect:=True, Title:="ASCII Files to Open")  
  
    If TypeName(FilesToOpen) = "Boolean" Then  
        MsgBox "No Files were selected"  
        GoTo ExitHjaler  
    End If  
  
    x = 1  
    Set wkbTemp = Workbooks.Open(FileName:=FilesToOpen(x))  
    wkbTemp.Sheets(1).Copy  
    Set wkbAll = ActiveWorkbook  
    wkbTemp.Close (False)  
    wkbAll.Worksheets(x).Columns("A:A").TextToColumns _  
        Destination:=Range("A1"), DataType:=xlDelimited, _  
        TextQualifier:=xlDoubleQuote, _  
        ConsecutiveDelimiter:=False, _  
        Tab:=True, Semicolon:=False, _  
        Comma:=False, Space:=False, _  
        Other:=True, OtherChar:="|"   
    x = x + 1  
  
    While x <= UBound(FilesToOpen)  
        Set wkbTemp = Workbooks.Open(FileName:=FilesToOpen(x))  
        With wkbAll  
            wkbTemp.Sheets(1).Move After:=.Sheets(.Sheets.Count)  
            .Worksheets(x).Columns("A:A").TextToColumns _  
                Destination:=Range("A1"), DataType:=xlDelimited, _  
                TextQualifier:=xlDoubleQuote, _  
                ConsecutiveDelimiter:=False, _  
                Tab:=True, Semicolon:=False, _  
                Comma:=False, Space:=False, _  
                Other:=True, OtherChar:=sDelimiter  
        End With  
        x = x + 1  
    Wend  
  
ExitHjaler:  
    Application.ScreenUpdating = True  
    Set wkbAll = Nothing  
    Set wkbTemp = Nothing  
    Exit Sub
```

```

ErrHjaler:
    MsgBox Err.Description
    Resume ExitHjaler
End Sub

```

*Makro 2: välilehtien sisältämän numeerisen informaation koonti yhdelle välilehdelle, jossa ensimmäiseen sarakkeeseen luodaan kuvanumerot*

```

Sub CopyA6A3005toSummary()

    Dim sh As Worksheet
    Dim DestSh As Worksheet
    Dim CopyRng As Range
    Dim x As Integer

    With Application
        .ScreenUpdating = False
        .EnableEvents = False
    End With

    x = 2

    Set DestSh = ActiveWorkbook.Worksheets.Add
    DestSh.Name = "Summary"

    'If different cells need to be selected, change the range
    'A6:A3006 to whatever you need

    For Each sh In ActiveWorkbook.Worksheets
        If sh.Name <> "Summary" Then
            Set CopyRng = sh.Range("A6:A3006")
            CopyRng.Copy
            With DestSh.Cells(1, x)
                .PasteSpecial xlPasteValues
                Application.CutCopyMode = False
            End With
            x = x + 1
        End If
    Next

    Dim y As Integer
    y = 2
    Worksheets("Summary").Cells(1, 1).Value = 1
    While y <= 3001 'Enter here the amount of frames
        Worksheets("Summary").Cells(y, 1).Value = y
        y = y + 1
    Wend

ExitTheSub:
    Application.Goto DestSh.Cells(1)
    With Application
        .ScreenUpdating = True
        .EnableEvents = True
    End With
End Sub

```