Supplementary Information

Regulatory landscapes and structural choreography of transcription initiation in spirochetes

Vilma K. Trapp[†], Tingting Wang[†], Tarek Hilal, Janne J. Makinen, Juho Kotikoski, Paavo J. Tavi, Ville Levola, Sari Paavilainen, Bernhard Loll, Markus C. Wahl*, Georgiy A. Belogurov*

[†] Shared first authors; these authors contributed equally to this work.

* Correspondence to: markus.wahl@fu-berlin.de; gebelo@utu.fi

Supplementary methods

S. africana cultivation

S. africana was routinely cultivated in an alkali medium (pH 9.2-9.5) modified from the DSMZ alkaliphilic spirochaete medium #700. The medium was buffered by a mixture of Na_2CO_3 (10 g/l) and $NaHCO_3$ (15 g/l), contained NaCl (50 g/l) to maintain high-salinity [1], maltose (5 g/l) as a major carbon and energy source and Na_2S (0.5 g/l) and cysteine hydrochloride (0.5 g/l) as reducing agents. The medium also contained NH_4Cl (0.5 g/l), K_2HPO_4 (0.5 g/l), yeast extract (0.5 g/l), 10 ml/l of vitamin solution, 1 ml/l of trace element solution and 1 mg/l of resazurin as redox indicator. For cultivation at pH 8.5, $NaHCO_3$ was increased to 25 g/l and Na_2CO_3 was omitted. For monitoring the survival at pH 7.5, HEPES-Na (HEPES 12 g/l, NaOH 1 g/l) was used as a buffering agent, Na_2CO_3 and $NaHCO_3$ were omitted, and NaCl was increased to 90 g/l.

A continuous culture was grown in completely filled 6.5 ml soda-lime glass tubes with polypropylene screw caps and thermoplastic elastomer seal (or up to 1 l screw caps bottles when desired) without shaking at 25°C, pH 9.2-9.5 and transferred every 3 days (OD_{600} 0.3-0.5; 25 µl to 6.5 ml). Alternatively, 3-day old cultures were supplemented with glycerol (16% final), stored at -80°C and revived when desired. On day 4 cultures reached saturating OD_{600} ~0.8, cells maintained spiral shape and motility, and the culture remained evenly turbid. On day 5 culture health deteriorated, a large fraction of cells sedimented to the bottom of the tube and morphed into spherical bodies. When grown at 30°C, cultures reached saturation on day 3 and the viability deteriorated on day 4. When grown at 25°C, pH 8.5, cultures reached saturation on day 7 and remained healthy on day 12 after which the viability rapidly deteriorated. When 3-day old cultures pre-grown at pH 9.2-9.5 were pelleted and transferred to pH 7.5 medium, cells remained spiral and motile on day 2 followed by the massive die out on days 3-4, but a small number of spiral and motile cells were observed up to at least day 7.

MODIFIED ALKALIPHILIC SPIROCHAETE MEDIUM

Main solution

10	g	Na_2CO_3
15	g	$NaHCO_3$
50	g	NaCl
0.5	g	K_2HPO_4
0.5	g	NH_4Cl
0.5	g	Cysteine hydrochloride
0.5	g	Na_2S x 9 H_2O
0.5	g	Yeast extract
5	g	Maltose
1	mg	Resazurin
10	ml	**Wolin's vitamin solution**
1	ml	**Modified Wolin's mineral solution II**
1000	ml	Distilled water

Filter sterilize, warm up to 60 C and cool down to room temperature. Expected pH 9.2-9.5

Modified Wolin's mineral solution II

1.5	g	Nitrilotriacetic acid
3	g	$MgSO_4$ x 7 H_2O
0.5	g	$MnSO_4$ x H_2O
1	g	NaCl
0.1	g	$FeSO_4$ x 7 H_2O
0.18	g	$CoSO_4$ x 7 H_2O
0.1	g	$CaCl_2$ x 2 H_2O
0.18	g	$ZnSO_4$ x 7 H_2O
0.01	g	$CuSO_4$ x 5 H_2O
0.02	g	$AlK(SO_4)_2$ x 12 H_2O
0.01	g	H_3BO_3
0.01	g	Na_2MoO_4 x 2 H_2O
0.025	g	$NiCl_2$ x 6 H_2O
0.3	mg	Na_2SeO_3 x 5 H_2O
1000	ml	Distilled water

First dissolve nitrilotriacetic acid and adjust pH to 6.5 with KOH, then add minerals. Adjust final to pH 7.0 with KOH.

Wolin's vitamin solution

2	mg	Biotin
2	mg	Folic acid
10	mg	Pyridoxine hydrochloride
5	mg	Thiamine HCl
5	mg	Riboflavin
5	mg	Nicotinic acid
5	mg	Calcium D-(+)-pantothenate
0.1	mg	Vitamin B12
5	mg	p-Aminobenzoic acid
5	mg	(DL)-alpha-Lipoic acid
1000	ml	Distilled water

Purification of *S. africana* RNAP from *S. africana* cells

Native *S. africana* RNAP was purified from *S. africana* cells by a combination of heparin, size-exclusion and anion exchange chromatography. *S. africana* cultures were grown for 3-4 days at 25-30°C until the early saturation at OD_{600} ~0.8 (6.5 ml inoculum per 1 l bottle of standard medium with pH 9.2-9.5). *S. africana* cells were harvested by centrifugation at 10,000 × g for 30 min at 4°C (~15 g from 6 l of culture, stored at -80°C), resuspended in 10% Buffer B supplemented with EDTA-free protease inhibitor cocktail (Pierce or Roche) and disrupted by a mild sonication. At least 30 ml of lysis buffer was used per 5 g of cells. Lysates were clarified by centrifugation at 40-60,000 × g for 0.5-1 h at 4°C and loaded onto a 5 ml HiTrap® Heparin HP column (Cytiva) equilibrated with 10% Buffer B. The column was washed with 15 ml of 10% Buffer B and proteins were eluted with a steep gradient of Buffer B (30 ml, 10-100%). All protein-containing fractions were pooled, concentrated to 2 ml using a Amicon Ultra-15 (15 ml) centrifugal filter unit with Ultracel-3 (3 kDa cutoff) membrane and loaded onto a 120 ml HiPrep™ Sephacryl™ S-400 HR 16/60 gel filtration column (Cytiva) equilibrated with 10% Buffer B. RNAP eluted from the column between 70 and 80 ml. Fractions containing RNAP were pooled, concentrated using ultrafiltration concentrators, dialyzed against the storage buffer and stored at -20°C (short term) or -80°C (long term). Alternatively, RNAP-containing fractions were loaded onto a 6 ml ResourceQ column (Cytiva) equilibrated with 10% Buffer B, washed with 15-30 ml of 10% buffer B, and RNAP was eluted with a 100 ml gradient of buffer B (10-100%) followed by concentration and dialysis.

Total RNA isolation from *S. africana*

Total cell RNA was isolated using TRIzol® Max™ Bacterial RNA Isolation Kit (Thermofisher). Cells from 6.5 ml culture were harvested by centrifugation at 4,000 × g for 10 min at 4°C and resuspended in 200 µl of Max Bacterial Enhancement Reagent preheated at 95°C. The sample was then dissolved in 1 ml TRIzol and extracted twice with 200 µl of cold chloroform. RNA was then precipitated from the aqueous phase by addition of 500 µl of cold isopropanol. The RNA pellet was washed with 1 ml of 75% ethanol and dissolved in 50 µl of RNase-free water. The sample was then treated with DNase (TURBO DNA-free™ Kit, Thermofisher). Briefly, the sample was supplemented with DNase buffer, DNase, diluted to 100 µl, and incubated for 20 min at 37°C. DNase and divalent cations were then removed by the proprietary DNase removal matrix supplied with the kit. RNA was then additionally cleaned up using Qiagen RNAeasy mini kit using Clean-up protocol. Briefly, the RNA sample was supplemented with 350 µl of high-salt proprietary buffer containing guanidine thiocyanate (Buffer RLT), 950 µl of 99% ethanol (up from the default 250 µl of ethanol to facilitate binding of small RNAs) and applied to the silica-based spin column. The column

was washed with the low-salt proprietary buffer (Buffer RPE) containing 75% ethanol, and RNA was eluted with 50 µl of RNase-free water.

Transcriptome sequencing and analysis

Total RNA was isolated from a 3-day-old *S. africana* culture grown at pH 9.2-9.5 until early saturation (OD_{600} ~0.8) at 30 °C. Total RNA (100 ng) was converted into a dual-indexed library using Illumina® Stranded Total RNA Prep, Ligation with Ribo-Zero Plus kit and sequenced with Illumina NovaSeq 6000 SP v1.5 (2 × 100 bp). The library preparation and RNA sequencing were performed by the Finnish Functional Genomics Centre Facility (FFGC). The RNA sample and library quality were monitored using an Agilent Bioanalyzer 2100 instrument and software (Agilent). The read quality was inspected using FastQC.

The automated transcriptome assembly was conducted using Rockhopper v2.03 [2] that aligned raw reads (FASTQ format) to the annotated *S. africana* DSM8902 genome [3], produced coverage traces (WIG format) and a table containing raw read counts and expression values (analogues of FPKM but normalized by the upper quartile of gene expression, excluding genes with zero expression) for annotated genome features (**Supplementary Data File**). Transcriptome coverage was explored using Integrative Genomics Viewer [4]. For figure preparation, strand-specific transcriptome coverage of selected regions of the genome was extracted from WIG files using a python script and plotted using Origin 2015 software (Origin Labs).

As an alternative analysis method, raw reads (FASTQ format) were mapped to the *S. africana* genome using Bowtie2 v2.5.3 [5]. The resulting alignment in SAM format was sorted and indexed using SAMtools v1.19 [6]. Reads mapping to the genome features (Gene Transfer Format, GTF) were then counted using HTSeq v2.0.3 [7]. The HTSeq read counts were normalized using the transcripts per million (TPM) method (**Supplementary Data File**).

Analysis of *in vitro* transcription products using Oxford nanopore sequencing

To estimate transcription across the entire linearized transcription template (P-gre template was used as an example), *in vitro* transcription was performed as in the FLAP assays, but the reaction volume was scaled up to 0.5 ml and DFHBI-1T was omitted. Product RNAs were purified using the RNA easy kit (Qiagen) and polyadenylated with *E. coli* PAP (poly-A polymerase, NEB) to enable ligation of the sequencing adapters to the 3' ends of the RNAs. RNAs were purified using Agencourt RNAClean XP beads (Beckman Coulter), the sequencing library was prepared using the Direct RNA sequencing kit (Oxford nanopore) and sequenced using a MinION Mk1D instrument (Oxford nanopore). Reads were aligned to the linearized P-gre transcription template using MinKNOW, alignment files were split according to the strand using SAMtools[6], loaded into Ugene[8], strand-specific coverage was exported and ultimately plotted using Origin 2015 software.

In the transcription start site mapping experiment, seven *in vitro* transcription reactions were performed individually, each using a dedicated linearized plasmid template (six plasmids from the main set and the parent pOP004 plasmid featuring P-gre promoter with the altered initially transcribed region). One set of reactions was performed using *S. africana* holoenzyme and another set using *E. coli* holoenzyme. After quenching with EDTA, the product RNAs in each set were pooled together, purified using the RNA easy kit and polyadenylated with *E. coli* PAP. Polyadenylated RNAs were then bound to Oligo-d(T)25 magnetic beads (NEB) in the presence of 0.5 M LiCl and treated with RppH (RNA pyrophosphohydrolase, NEB) to convert 5' triphosphate moieties into single phosphate moieties. A 24 nt RNA adapter (CACACGCACACACAACCAGAGGAG) was then ligated to the 5' ends of the product RNAs by T4 RNA ligase I (NEB) in the presence of 12.5% (v/v) PEG. RNAs were released from beads by heating to 75 °C for 2 min. RNAs were purified using Agencourt RNAClean XP beads, the sequencing library was prepared using the Direct RNA sequencing kit (Oxford nanopore) and sequenced using a MinION Mk1D instrument (Oxford nanopore).

The reference RNA sequences were constructed for each of the seven promoters assuming that transcription initiates at a purine separated by 6 or 7 bp from the -10 promoter element or at a purine separated by 7 bp from the -10 element if the former is not present. Reads (FASTQ format) were aligned to the reference RNAs using MinKNOW (BAM format output). The default alignment settings in MinKNOW are relaxed enough to recover reads with small indels at the adapter-RNA 5' junction and the same reads were recovered if the reference sequences were constructed using transcription start sites ±2 nucleotides. Reads that overlap with the adapter sequence (first 24 nt) were extracted (BAM format output) and converted back to a FASTQ format list using SAMtools. Reads were then loaded into Ugene and searched with sequence patterns containing 6 nucleotides of the adapter and 7 nucleotides of RNA transcribed from the predicted transcription start site ±2 nucleotides. Additional patterns were designed after slippage was detected by manual inspection of read alignments. To perform the manual inspection, extracted reads were realigned to the reference RNAs using Muscle [9] to reveal sequence insertions that are obscured when viewing BAM format alignments.

***In vitro* crosslinking with DSSO followed by mass spectrometry analysis**

S. africana holoenzyme (100 µl, 15 µM) with and without CarD (15 µM) was dialyzed against high-salt phosphate buffered saline (PBS; pH 7.4, 300 mM NaCl) resulting in 150 µl of 10 µM solution. Proteins were crosslinked with 3 mM disuccinimidyl sulfoxide (DSSO) for 30 min at 25 °C, quenched with Tris-HCl pH 7.9 (30 mM final) and dialyzed against PBS (pH 7.6, 150 mM NaCl). Samples were treated with urea, proteins were reduced with 10 mM dithiothreitol (in 50 mM Tris-HCl, pH 8.0), alkylated with 40 mM iodoacetamide and

digested overnight with sequencing grade modified trypsin (Promega). Peptides were desalted using Sep-Pak tC18 96-well plate (Waters) and dried in a vacuum centrifuge. The dried peptide samples were dissolved with 0.1% formic acid and 1500 ng were subjected to the Liquid Chromatography-Electrospray Ionization-Tandem Mass Spectrometry (LC-ESI-MS/MS) analysis. The LC-ESI-MS/MS analyses were performed on a nanoflow HPLC system (Easy-nLC1000, Thermo Scientific) coupled to the Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with a nano-electrospray ionization source and FAIMS Pro interface (Thermo Scientific). Compensation voltages of -40 V, -60 V, and -75 V were used. Peptides were first loaded on a trapping column and subsequently separated inline on a 15 cm C18 column (75 µm x 15 cm, ReproSil-Pur 3 µm 120 Å C18-AQ, Dr. Maisch HPLC GmbH). The mobile phase consisted of water with 0.1% formic acid (solvent A) or acetonitrile/water (80:20 (v/v)) with 0.1% formic acid (solvent B). Peptides were eluted with the following gradient: from 5% to 21% of solvent B in 62 min, from 21% to 36% of solvent B in 48 min, from 36% to 100% of solvent B in 5 min, followed by 5 min at 100% of solvent B. MS data was acquired automatically by using Thermo Xcalibur 4.7 software (Thermo Fisher Scientific). Crosslinks were analyzed using the MS2-MS2-MS3 method where MS1 spectra were acquired in the Orbitrap mass analyzer with mass range of 375–1500 m/z at a resolution of 60,000. For the MS2-MS3-MS2 fragmentation method, sequential CID and ETD spectra were acquired for each precursor. MS3 scans were triggered by a targeted mass difference of 31.9721 detected in the MS2 scan. The MS3 scan was performed in the ion trap with CID fragmentation. Pairs of crosslinked peptides were identified by searching the raw data against the *S. africana* RNAP holoenzyme subunits using Proteome Discoverer 3.2 software (Thermo Scientific) and the XlinkX 3.2 algorithm (**Supplementary Data File**).

Supplementary Table 1. Bacterial species employed in analysis presented in Figure 1.

Kingdom	Superphylum	Phylum		Species
Pseudomonadati	Pseudomonadota and related phyla	Pseudomonadota	eco	*Escherichia coli*
			cvi	*Chromobacterium violaceum*
			rsp	*Cereibacter sphaeroides*
		Bdellovibrionota	bba	*Bdellovibrio bacteriovorus*
		Myxococcota	mxa	*Myxococcus xanthus*
		Thermodesulfobacteriota	gsu	*Geobacter sulfurreducens*
		Nitrospirota	nmv	*Nitrospira moscoviensis*
		Nitrospinota	tye	*Thermodesulfovibrio yellowstonii*
		Acidobacteriota	aca	*Acidobacterium capsulatum*
			nli	*Nitronauta litoralis*
	ACT	**A**quificota	aae	*Aquifex aeolicus*
			pmx	*Persephonella marina*
		Campylobacterota	tdn	*Sulfurimonas denitrificans*
			nap	*Nautilia sp. PV-1*
		Thermosulfidibacterota	ttk	*Thermosulfidibacter takaii*
	FCB+	**F**ibrobacterota	fsu	*Fibrobacter succinogenes*
		Gemmatimonadota	gau	*Gemmatimonas aurantiaca*
		Chlorobiota	cte	*Chlorobaculum tepidum*
		Bacteroidota	bth	*Bacteroides thetaiotaomicron*
		Rhodothermota	sru	*Salinibacter ruber*
		Balneolota	cprv	*Cyclonatronum proteinivorum Omega*
	PVC+	**P**lanctomycetota	rba	*Rhodopirellula baltica*
			hbs	*Humisphaera borealis*
		Kiritimatiellota	vbl	*Kiritimatiella glycovorans*
		Verrucomicrobiota	ote	*Opitutus terrae*
		Lentisphaerota	lpro	*Lentisphaera profundi*
		Chlamydiota	ctr	*Chlamydia trachomatis*
		Spirochaetota	sfc	*Spirochaeta africana*
			ock	*Oceanispirochaeta crateris*
			lbf	*Leptospira biflexa*
			taqu	*Thermospira aquatica*
Fusobacteriati		Fusobacteriota	fnu	*Fusobacterium nucleatum*
			haby	*Haliovirga abyssi*
Thermotogati	**DST**	**D**einococcota	dra	*Deinococcus radiodurans*
			ttj	*Thermus thermophilus*
		Synergistota	amo	*Acetomicrobium mobile*
			tai	*Thermanaerovibrio acidaminovorans*
		Thermotogota	tma	*Thermotoga maritima*
			mpg	*Mesotoga prima*
			tta	*Pseudothermotoga thermarum*
Bacillati		Cyanobacteriota	syn	*Synechocystis sp. PCC 6803*
			gvi	*Gloeobacter violaceus*
		Armatimonadota	ccz	*Chthonomonas calidirosea*
			fgi	*Fimbriimonas ginsengisoli*
		Chloroflexota	det	*Dehalococcoides mccartyi*
			tro	*Thermomicrobium roseum*
			cap	*Caldilinea aerophila*
		Actinomycetota	mtu	*Mycobacterium tuberculosis*
			sco	*Streptomyces coelicolor*
		Bacillota	bsu	*Bacillus subtilis*
			chy	*Carboxydothermus hydrogenoformans*
			dsy	*Desulfitobacterium hafniense*

Bold letters explain the names of superphyla

Supplementary Table 2. Cryo-EM data collection, refinement and validation statistics.

	RPo (P-gre)				RPc (P-gre)	CarD-RPo (P-gre)	CarD-RPo (P-rrna)
State	Monomer I	Monomer II	Monomer III	Monomer IV	Dimer	Monomer I	Monomer
EMDB ID	56405	56406	56407	56408	56409	56410	56411
PDB ID	9TXX	9TXY	9TXZ	9TY0	9TY1	9TY2	9TY3
Data collection and processing							
Microscope	FEI Titan Krios G3i						
Voltage [keV]	300						
Camera	Falcon 3EC						
Magnification	96,000						
Pixel size at detector [Å/pixel]	0.832						
Total electron exposure [e^-/Å^2]	40						
Exposure rate [e^-/pixel/s]	0.7						
Frames per exposure	29						
Defocus range [μm]	0.5 - 2.5						
Automation software	EPU (version 2.8.1)						
Micrographs used	6,990				3,505	4,204	5,406
Total extracted particles	1,925,988				361,110	566,552	2,260,801
Final particles	163,995	52,968	70,566	47,627	90,219	56,632	394,310
Point-group or helical symmetry parameters	C1	C1	C1	C1	C2	C1	C1
Global resolution [Å]; $FSC_{0.143}$ (unmasked / masked)[a]	3.7 / 3	4.2 / 3.3	4.0 / 3.2	4.2 / 3.3	3.9 / 3.2	3.8 / 3.2	3.3 / 2.8
Local resolution range [Å]	2.4 - 38	2.5 - 40	2.5 - 40	2.6 - 40	2.5 - 40	2.6 - 43	2.2 - 40
Map sharpening *B* factor [Å^2] / *B* factor range	-91	-75	-83	-73	-85	-76	-94
Map sharpening method	Local *B* factor						
Refinement package	PHENIX (version 1.21_5419); real.space.refine						
Model composition							
Non-hydrogen atoms	30,517	31,358	31,235	31,269	61,999	32,315	32,097
Protein residues	3,578	3,658	3,646	3,643	3,436	3,753	3,764
Nucleic acid residues	102	110	112	115	145	124	109
Mg^{2+} ions	1	1	1	1	2	1	1
Zn^{2+} ions	2	2	2	2	4	2	2

Model Refinement							
Model-map scores							
CC[b] (mask)	0.86	0.86	0.85	0.78	0.85	0.75	0.86
CC (volume)	0.85	0.86	0.84	0.78	0.84	0.73	0.85
Average *B* factors [Å^2]							
Overall	154	216	191	183	157	125	140
Protein	146	211	170	177	154	120	125
Nucleic acid	264	276	459	252	220	182	348
Mg^{2+} ions	124	142	128	155	111	125	110
Zn^{2+} ions	142	162	144	162	154	92	125
Rmsd[c] from ideal values							
Bond lengths [Å]	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Bond angles [°]	0.616	0.545	0.546	0.568	0.488	0.581	0.532
Validation[d]							
MolProbity score	2.07	2.15	2.15	2.22	2.09	2.02	1.69
CaBLAM outliers [%]	2.6	2.6	2.8	2.8	2.6	1.7	1.8
Clash score	11.6	11.3	11.6	12.4	10.8	9.9	8.7
Poor rotamers [%]	1.9	2.9	2.6	3.0	2.5	2.7	1.7
Cβ deviations	0	0	0	0	0	0	0
EMRinger score	2.05	2.10	2.01	2.02	2.46	2.48	2.97
Ramachandran plot							
Favored [%]	96.0	96.4	96.3	96.2	96.4	97.0	97.3
Allowed [%]	4.0	3.6	3.7	3.8	3.6	3.0	2.7
Outliers [%]	0	0	0	0	0	0	0
Ramach. *Z*-score (rmsd)							
Whole	0.57 (0.14)	0.69 (0.14)	0.78 (0.14)	0.62 (0.14)	0.81 (0.10)	0.95 (0.14)	1.26 (0.14)
Helix	1.85 (0.14)	1.91 (0.14)	1.85 (0.14)	1.83 (0.14)	1.86 (0.10)	1.72 (0.14)	2.05 (0.14)
Sheet	-0.30	-0.08 (0.25)	0.17 (0.25)	-0.12 (0.24)	0.11 (0.18)	0.41 (0.24)	0.61 (0.24)
Loop	(0.24)	-0.61 (0.15)	-0.49 (0.15)	-0.60 (0.15)	-0.46 (0.10)	-0.22 (0.15)	-0.13 (0.15)
	-0.65						
	(0.15)						

[a] FSC, Fourier shell correlation

[b] CC, correlation coefficient

[c] Rmsd, root-mean-square deviation

[d] Using MolProbity (PMID: 29067766)

Supplementary Table 3. *E. coli* protein expression vectors used in this study.

Name	Description	protease	Source/reference
pVS10	wild-type *E. coli* RNAP (α-β-β'_his_6-ω)		[10]
pVS14	ΔSi3 *E. coli* RNAP (α-β-β'_his_6-ω)		[11]
pIP008	βS351N *E. coli* RNAP (α-his_6_β-β'-ω)		this work
pJM019	wild-type *S. africana* RNAP (α-β-β'_his_6-ω)		this work
pIP007	βN490S *S. africana* RNAP (α-his_6_β-β'_his_6-ω)		this work
pET-σ70	wild-type *E. coli* σ70 (his6_thrombin_σ70)	thrombin	[12]
pGB217	wild-type *S. africana* σ70 (his6_SUMO_σ70)	ULP1	this work
pGB179	*S. africana* CarD (his6_TEV_CarD)	TEV	this work
pGB175	*S. africana* DksA (his6_TEV_DksA)	TEV	this work
pRF2	*E. coli* DksA (his10_TEV_DksA)	TEV	[13]

Sequences of the plasmids are included in the **Supplemetary Data File**.

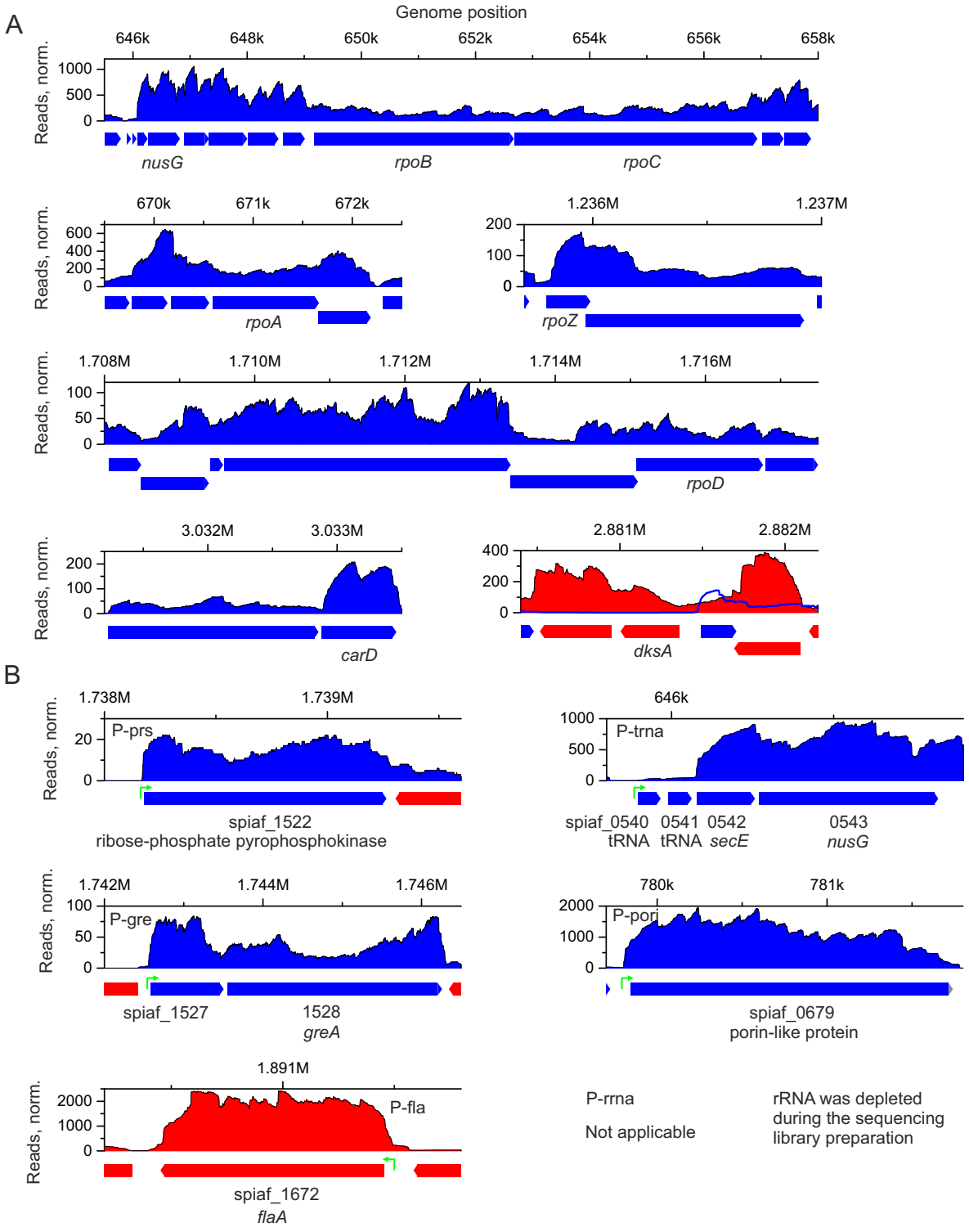
Supplementary Table 4. Plasmids encoding transcription templates.

Name	Description	downstream gene in the genome	Source/reference
pOP004	P-gre with altered ITR	n/a	[14]
pVN003	P-rrna promoter	spiaf_2142, spiaf_2147, spiaf_2596	this work
pVN001	P-gre promoter	spiaf_1527	this work
pJK003	P-prs promoter	spiaf_1522	this work
pJK006	P-trna promoter	spiaf_0540	this work
pVL017	P-fla promoter	spiaf_1672	this work
pVL019	P-pori promoter	spiaf_0679	this work
pVL026	promoter-less	n/a	this work
pJK004	*E. coli* rrnB P1	*E. coli* rRNA genes (16S, 23S, 5S)	this work

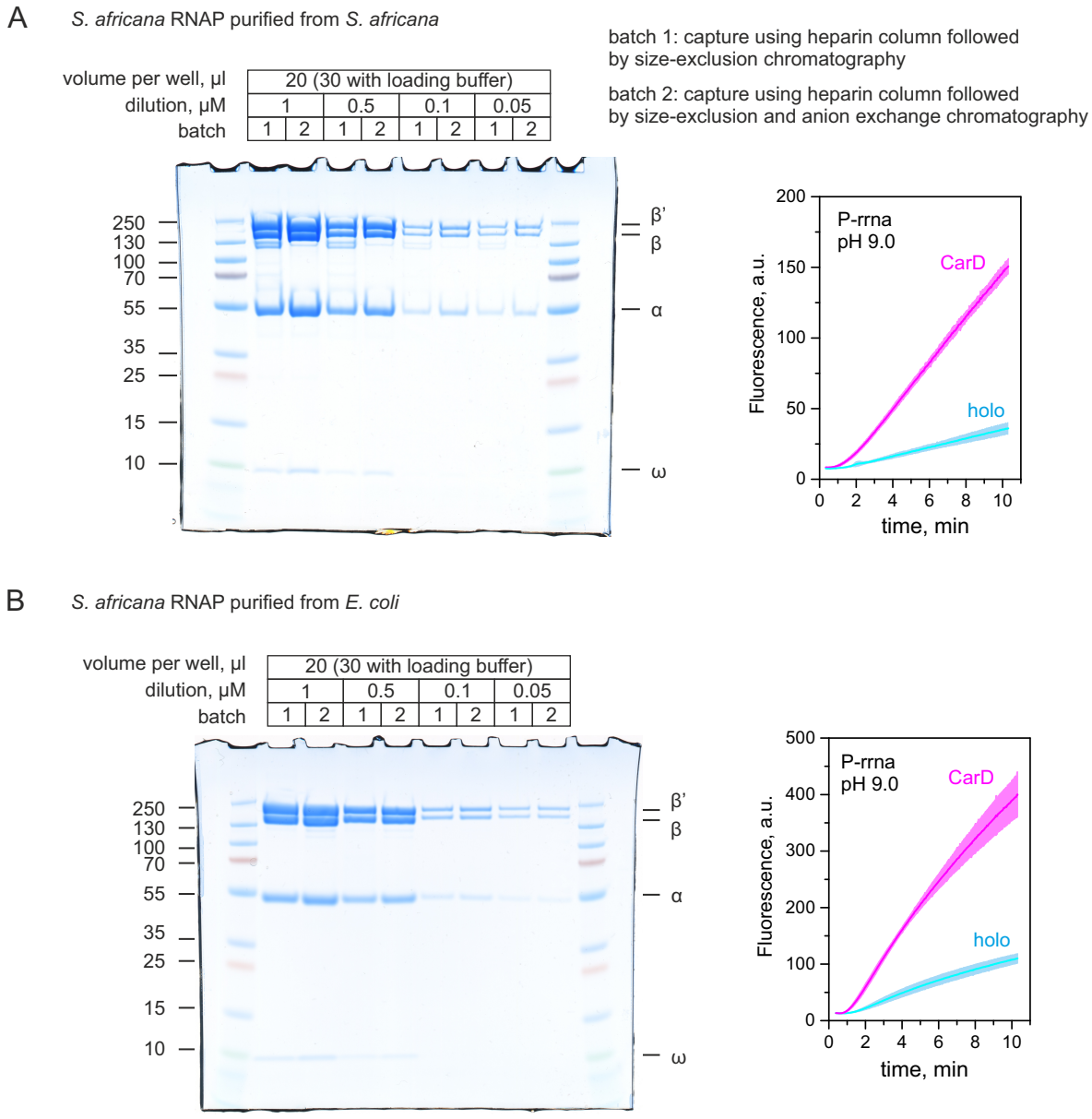
Sequences of the plasmids are included in a **Supplemetary Data File**.

Supplementary references

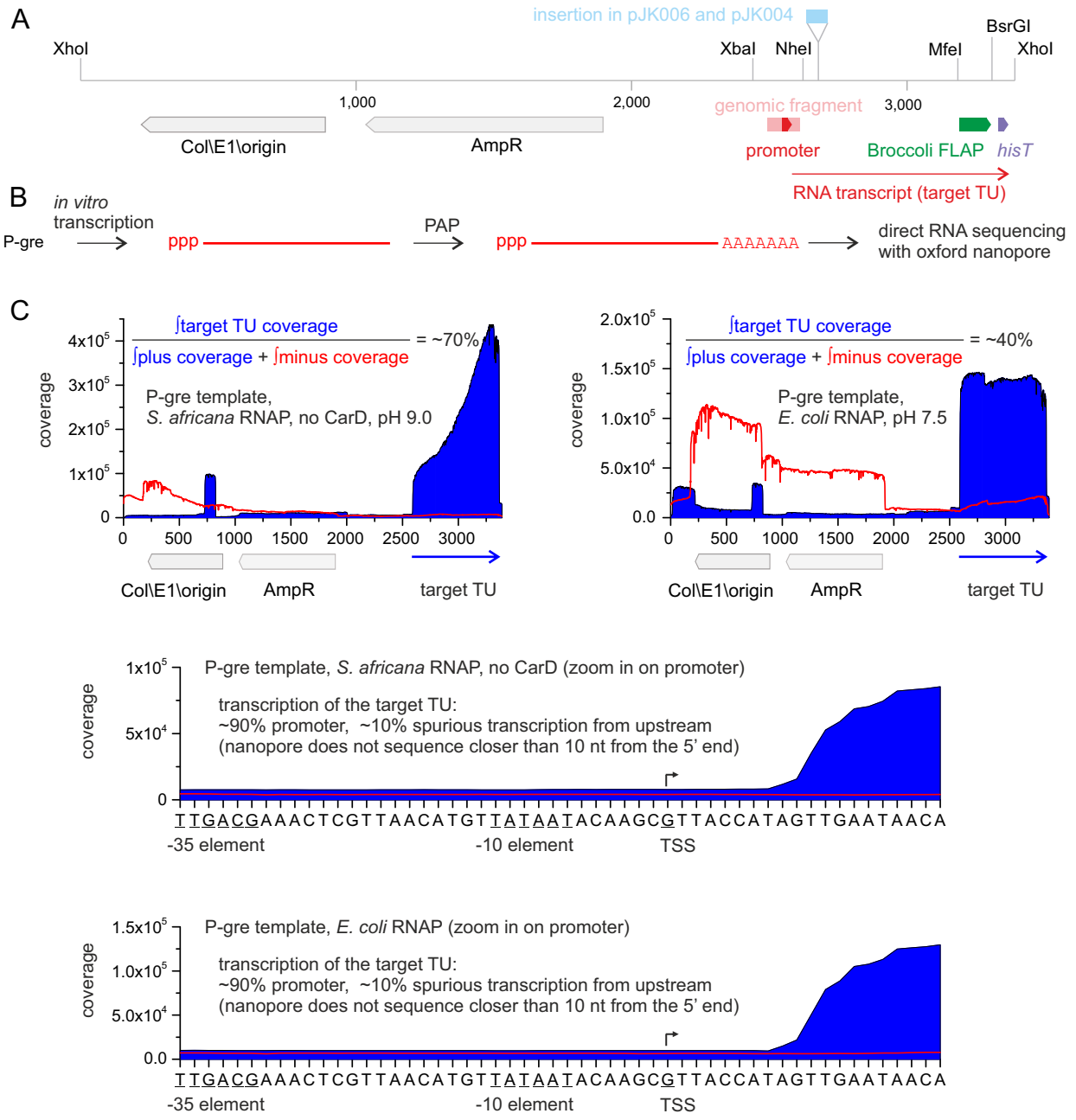
1. Zhilina, T. N. et al. Spirochaeta alkalica sp. nov., Spirochaeta africana sp. nov., and Spirochaeta asiatica sp. nov., alkaliphilic anaerobes from the Continental Soda Lakes in Central Asia and the East African Rift. Int. J. Syst. Bacteriol. **46**, 305–312 (1996).
2. McClure, R. et al. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res. **41**, e140 (2013).
3. Liolos, K. et al. Complete genome sequence of the halophilic bacterium Spirochaeta africana type strain (Z-7692(T)) from the alkaline Lake Magadi in the East African Rift. Stand. Genomic Sci. **8**, 165–176 (2013).
4. Robinson, J. T. et al. Integrative genomics viewer. Nat. Biotechnol. **29**, 24–26 (2011).
5. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**, 357–359 (2012).
6. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**, 2078–2079 (2009).
7. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics **38**, 2943–2945 (2022).
8. Okonechnikov, K., Golosova, O., Fursov, M. & UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics **28**, 1166–1167 (2012).
9. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**, 1792–1797 (2004).
10. Svetlov, V. & Artsimovitch, I. Purification of bacterial RNA polymerase: tools and protocols. Methods Mol. Biol. **1276**, 13–29 (2015).
11. Artsimovitch, I., Svetlov, V., Murakami, K. S. & Landick, R. Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. J. Biol. Chem. **278**, 12344–12355 (2003).
12. Marr, M. T. & Roberts, J. W. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science **276**, 1258–1260 (1997).
13. Furman, R. et al. pH dependence of the stress regulator DksA. PLoS ONE **10**, e0120746 (2015).
14. Huang, Y.-H. et al. Fluorogenic RNA aptamers to probe transcription initiation and co-transcriptional RNA folding by multi-subunit RNA polymerases. Meth. Enzymol. **675**, 207–233 (2022).



Supplementary Fig. 1 | Strand-specific transcriptome coverage corresponding to the components of *S. africana* transcription system (A) and promoters investigated in this study (B). Annotated genes are plotted below the transcriptome graphs. KEGG (www.kegg.jp) gene numbers (spiaf_xxxx) and selected gene names are indicated below the annotations. Coloring: positive strand coverage and annotations blue, negative strand coverage and annotations red.

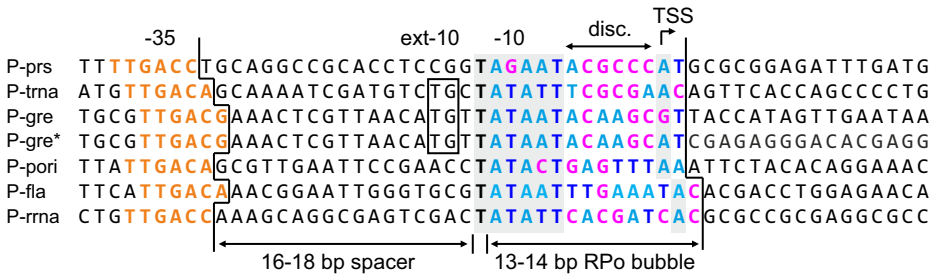


Supplementary Fig. 2 | Transcriptional activity of *S. africana* RNAP purified from *S. africana* (A) and *E. coli* (B) at P-rrna promoter. SDS-PAGE analysis was performed using the MES buffer system (50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3). Transcription activity was monitored using FLAP assay (1 µM RNAP, 4 µM σ, 50 nM DNA template and where indicated 10 µM CarD). Coloring: holoenzyme (cyan), in the presence of CarD (magenta). Promoter sequences and schematic of the FLAP assay are presented in the main text Figure 2.

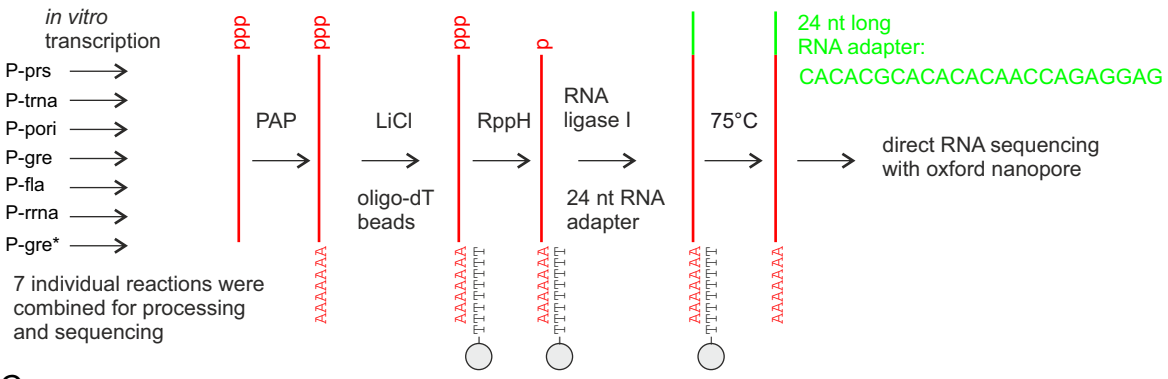


Supplementary Fig. 3 | Analysis of RNA products transcribed by *S. africana* and *E. coli* holoenzymes from a linearized plasmid using direct RNA sequencing. (A) Schematics of transcription templates used in this study. pVN001, pVN003, pJK003, pVL017, pVL019 and pOP005 plasmids differed only in the *Xba*I-*Nhe*I module containing plasmid-specific fragment of *S. africana* genome (pink) including the promoter (red). pJK006 additionally contained an insertion in the *Nhe*I-*Mfe*I module. *Mfe*I-*BsrG*I module encoded Broccoli FLAP embedded in tRNA scaffold, *BsrG*I-*Xho*I module encoded intrinsic transcription terminator. **(B)** Schematics of the experiment. *In vitro* transcription reactions (0.5 ml) were performed as in FLAP assays. Product RNA were purified using RNA easy kit (Qiagen), polyadenylated with *E. coli* poly-A polymerase (PAP). The sequencing library was prepared using Direct RNA sequencing kit (Oxford nanopore), and sequenced using MinION Mk1D instrument. **(C)** Reads were aligned to the P-gre transcription template using MinKNOW software, alignment files were split according to the strand using SAMtools, loaded into Ugene, per base coverage was exported and plotted using Origin software.

A



B



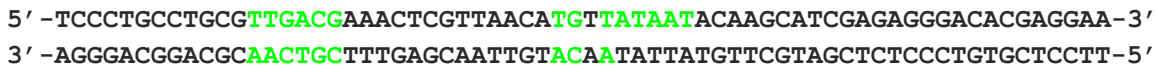
C

Search pattern	Read counts *Sfc*	*Eco*	
adapter P-prs			
GAGGAG GCGCGGA	0	7	TSS+2 or degradation
GAGGAG TGCGCGG	1	2	TSS+1 or degradation
GAGGAG ATGCGCG	**237**	**1525**	**TSS predicted**
GAGGAGCATGCGCG	2	0	TSS-1
adapter P-trna			
GAGGAG AGTTCAC	21	33	TSS+2 or degradation
GAGGAG CAGTTCA	0	7	TSS+1 or degradation
GAGGAG ACAGTTC	**609**	**1441**	**TSS predicted**
GAGGAG AACAGTTC	5	7	TSS-1
GAGGAG GAACAGTTC	1	1	TSS-2
GAGGAG CACAGTTC	2	6	slippage+degradation
GAGGAG ACACAGTTC	23	107	slippage
GAGGAGACACACAGTTC	1	1	slippage
adapter P-gre			
GAGGAG TTACCAT	22	1	TSS+1 or degradation
GAGGAG GTTACCA	**1365**	**375**	**TSS predicted**
GAGGAGCGTTACCA	4	1	TSS-1
adapter P-rrna			
GAGGAG CGCGCCG	2	0	TSS+1 or degradation
GAGGAG ACGCGCC	**172**	**228**	**TSS predicted**
GAGGAG CACGCGCC	61	131	TSS-1
GAGGAGTCACGCGCC	0	2	TSS-2

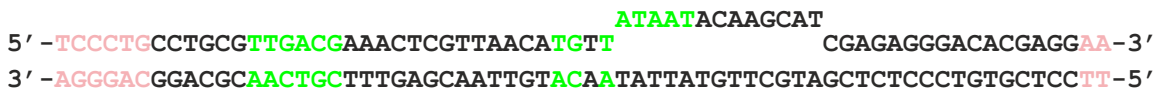
Search pattern	Read counts *Sfc*	*Eco*	
adapter P-pori			
GAGGAG ATTCTAC	16	1	TSS+2 or degradation
GAGGAG AATTCTA	31	12	TSS+1 or degradation
GAGGAG AAATTCT	**375**	**74**	**TSS predicted**
GAGGAG TAAATTCT	55	0	TSS-1
GAGGAG AAAATTCT	212	129	slippage
GAGGAG AAAAATTCT	25	29	slippage
GAGGAG AAAAAATTCT	6	3	slippage
GAGGAGAAAAAAATTCT	3	3	slippage
adapter P-fla			
GAGGAG ACGACCT	31	24	TSS+2 or degradation
GAGGAG CACGACC	17	7	TSS+1 or degradation
GAGGAG ACACGAC	**1868**	**755**	**TSS predicted**
GAGGAG TACACGAC	18	4	TSS-1
GAGGAG ATACACGAC	18	3	TSS-2
GAGGAG CACACGAC	5	2	slippage+degradation
GAGGAG ACACACGAC	129	86	slippage
GAGGAG ACACACACGAC	7	14	slippage
GAGGAGACACACACACGAC	0	1	slippage
adapter P-gre*			
GAGGAG AGAGGGA	1	4	TSS+4 or degradation
GAGGAG GAGAGGG	15	9	TSS+3 or degradation
GAGGAG CGAGAGG	15	6	TSS+2 or degradation
GAGGAG TCGAGAG	7	9	TSS+1 or degradation
GAGGAG ATCGAGA	**1763**	**2820**	**TSS predicted**
GAGGAGCATCGAGA	6	0	TSS-1

Supplementary Fig. 4 | Transcription start site mapping using 5' adapter ligation followed by direct RNA sequencing. (A) Promoter sequences used in this study aligned by -10 elements. Color coding as in main text **Fig. 2**. **(B)** The schematics of *in vitro* RNA processing before sequencing. **(C)** Counts of reads that map to the adapter-RNA 5' end junction and contain indicated sequence patterns. RNAs were transcribed individually from 7 templates and combined for processing and sequencing. Separate sequencing runs were performed for *S. africana* and *E. coli* holoenzymes. Search patterns contain 6 nucleotides of the adapter and 7 nucleotides of RNA transcribed from the predicted TSS ±2 nucleotides. Additional patterns were designed after slippage was detected by manual inspection of read alignments. Patterns and read counts corresponding to the predicted TSS are indicated in bold text. RNA sequencing produced U-containing reads, but MinKNOW aligner converted U to T. Sequence patterns were then designed to feature T in place of U.

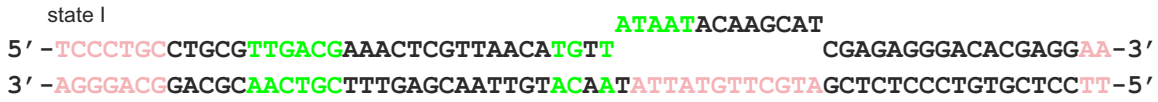
P-gre* promoter DNA (P-gre* = P-gre with an altered initially transcribed sequence):

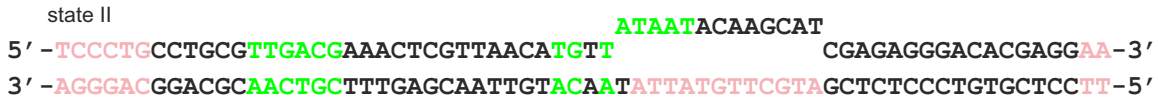


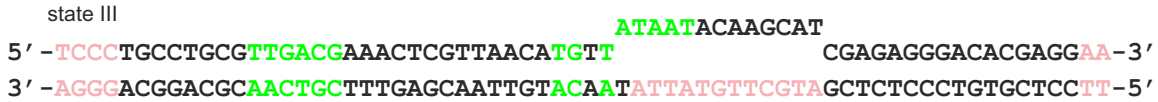
Transcription bubble in P-gre TIC-CarD (unresolved nucleotides colored pink):

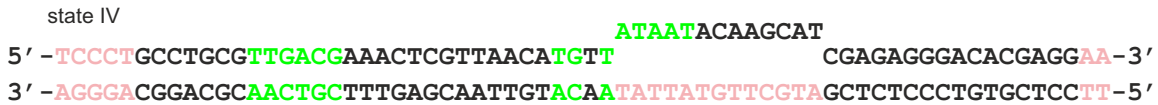


Transcription bubble in P-gre TIC (unresolved nucleotides colored pink):

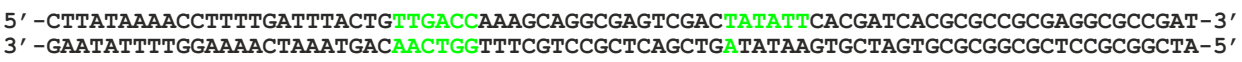




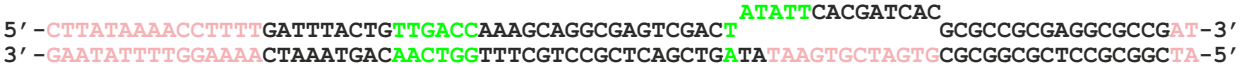




P-rrna promoter DNA:



Transcription bubble in P-rrna TIC-CarD (unresolved nucleotides colored pink):

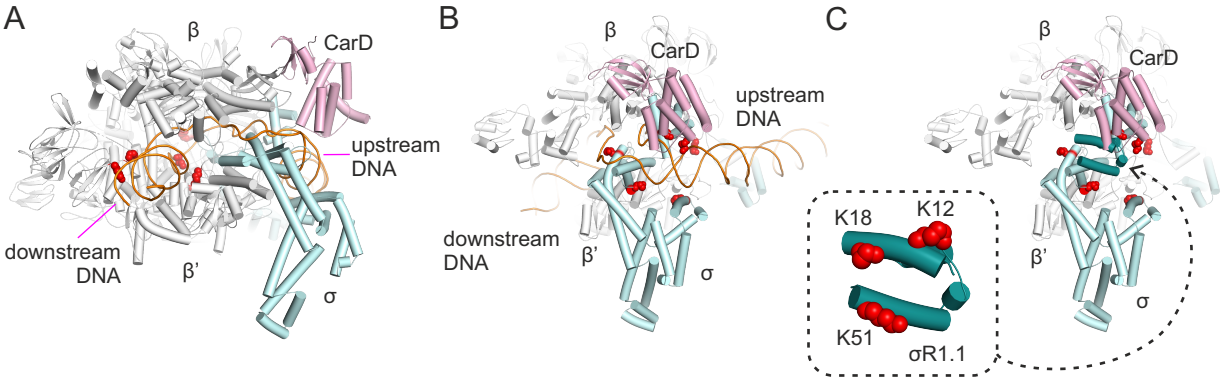


Supplementary Fig. 5 | Promoter scaffolds used to assemble *S. africana* holoenzyme - DNA complexes for cryoEM analysis. Promoter elements are shaded in green. Nucleotides that were not represented by the density and omitted from structural models are shaded light pink.

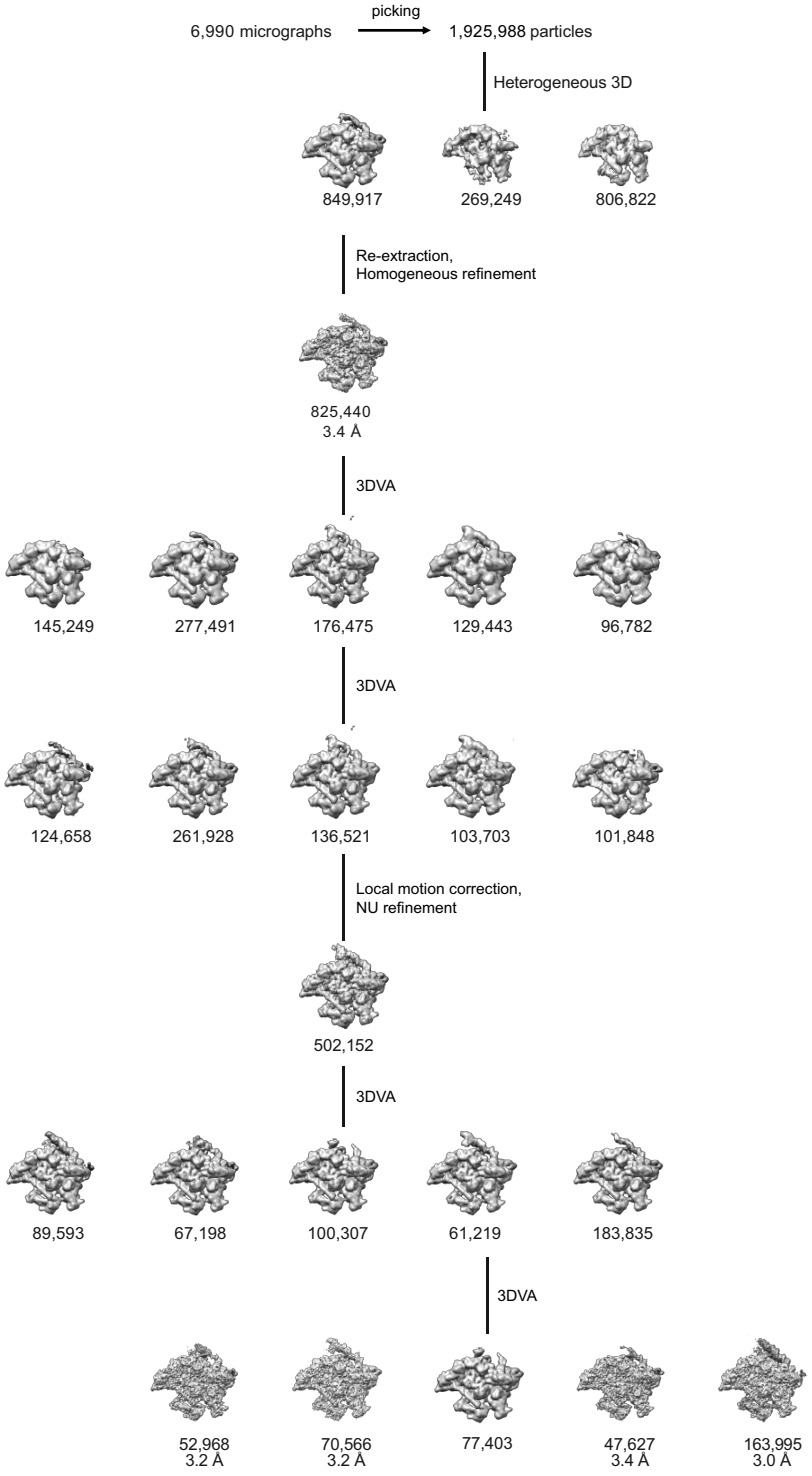
			Max. XlinkX Score				
			holo		holo-CarD		
σR1.1 Lys	-bridge-	other Lys	rep1	rep2	rep1	rep2	Location
K12	-DSSO-	β' K323	265	-	120	-	downstream DNA channel (Switch2)
K12	-DSSO-	β' K334	86	-	110	-	downstream DNA channel (Switch2)
K12	-DSSO-	σ K439	111	-	-	-	near upstream fork junction (σR3)
K12	-DSSO-	CarD K128	-	-	339	60	near upstream fork junction (CarD)
K18	-DSSO-	β'K323	76	84	65	-	downstream DNA channel (Switch2)
K18	-DSSO-	σ K51	182	-	171	-	σ1.1-σ1.1
K51	-DSSO-	β' K300	-	-	300	-	downstream DNA channel (Rudder)
K51	-DSSO-	β' K323	77	219	62	-	downstream DNA channel (Switch2)
K51	-DSSO-	β' K1150	144	184	-	-	downstream DNA channel (Jaw)
K51	-DSSO-	β' K1169	61	-	-	-	downstream DNA channel (Jaw)
K51	-DSSO-	β K800	110	-	-	-	near upstream fork junction (Flap)
K51	-DSSO-	β K1044	82	129	-	110	main channel (active site)
K51	-DSSO-	σ K18	182	-	171	-	σ1.1-σ1.1
K51	-DSSO-	σ K98	141	190	-	196	σ1.1-σ1.1
K51	-DSSO-	σ K173	138	211	-	118	near upstream fork junction (σNCR)
K51	-DSSO-	σ K369	78	-	-	-	near -10 element (σR2)
K51	-DSSO-	σ K395	191	183	236	144	near upstream fork junction (σR2)
K51	-DSSO-	σ K439	193	186	-	-	near upstream fork junction (σR3)
K51	-DSSO-	σ K472	-	-	-	238	near upstream fork junction (σR3)
K51	-DSSO-	CarD K128	-	-	65	-	near upstream fork junction (CarD)

holo = holoenzyme holo-CarD = holoenzyme with CarD DSSO = disuccinimidyl sulfoxide

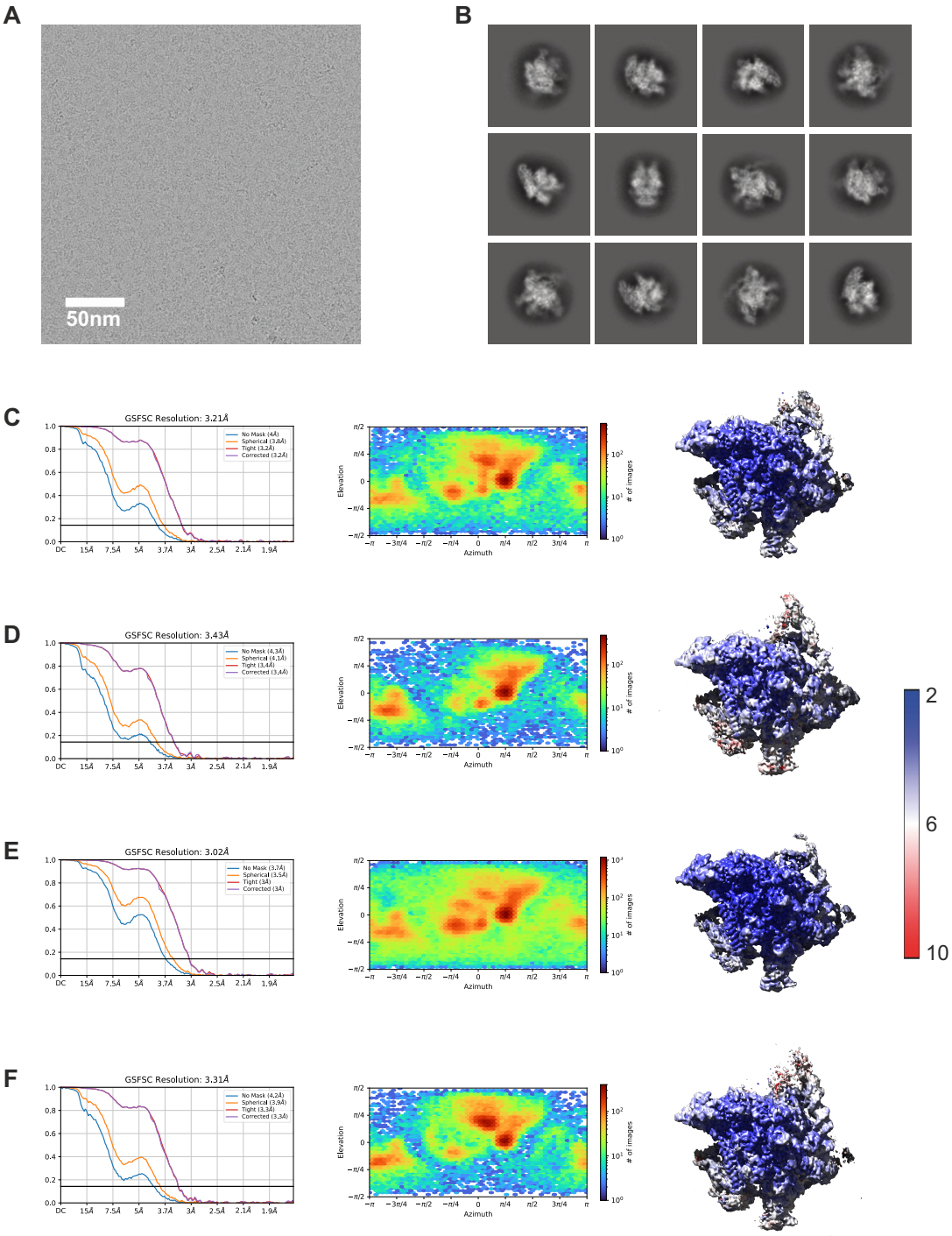
Max. XlinkX Score represents the likelihood that a detected cross-linked peptide is a true positive identification. Crosslinks with the score >60 are considered reliable.



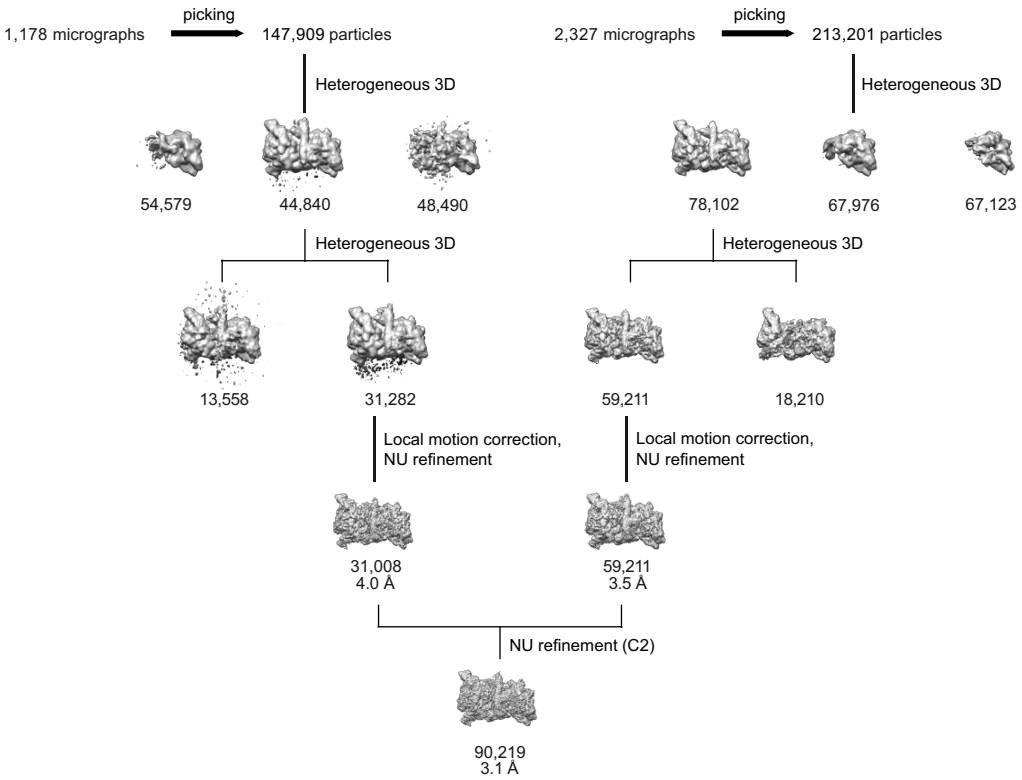
Supplementary Fig. 6 | Probing the location of σR1.1 in *S. africana* holoenzyme by *in vitro* crosslinking with disuccinimidyl sulfoxide (DSSO) followed by mass spectrometry analysis. Crosslinks involving Lys residues within the globular domain of σR1.1 (top table) were manually selected from the pool of all identified crosslinks (**Supplementary Data File**). Lys residues involved in crosslinks with σR1.1 lysines were then visualized (red spheres) on the structure of the *S. africana* RPo at the P-gre promoter with CarD. **(A)** σR1.1-crosslinkable Lys residues in Switch 2 region, the Rudder Loop and the Jaw domain suggest that σR1.1 is located in the downstream DNA channel. **(B)** σR1.1-crosslinkable Lys residues in σR2, σR3, σNCR and CarD suggest that σR1.1 is located in place of the upstream fork junction. **(C)** Same as in (B) but the promoter DNA was removed from the RPo structure and σR1.1 from the dimeric holoenzyme structure was modelled in using β subunit as a reference for structural alignment. The inset shows the zoomed in σR1.1 and σR1.1 Lys residues (red spheres) that form crosslinks discussed above.



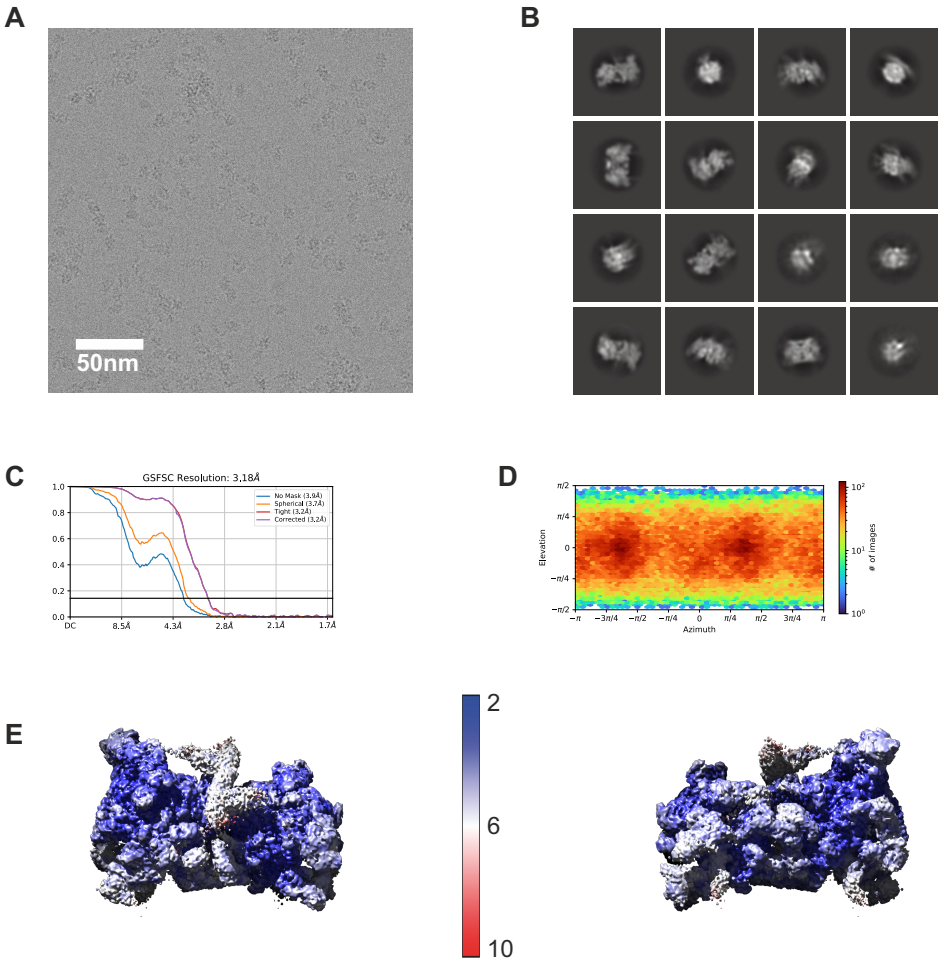
Supplementary Fig. 7 | Hierarchical clustering analysis of particle images of monomeric *S. africana* holoenzyme-DNA complex assembled at P-gre promoter.



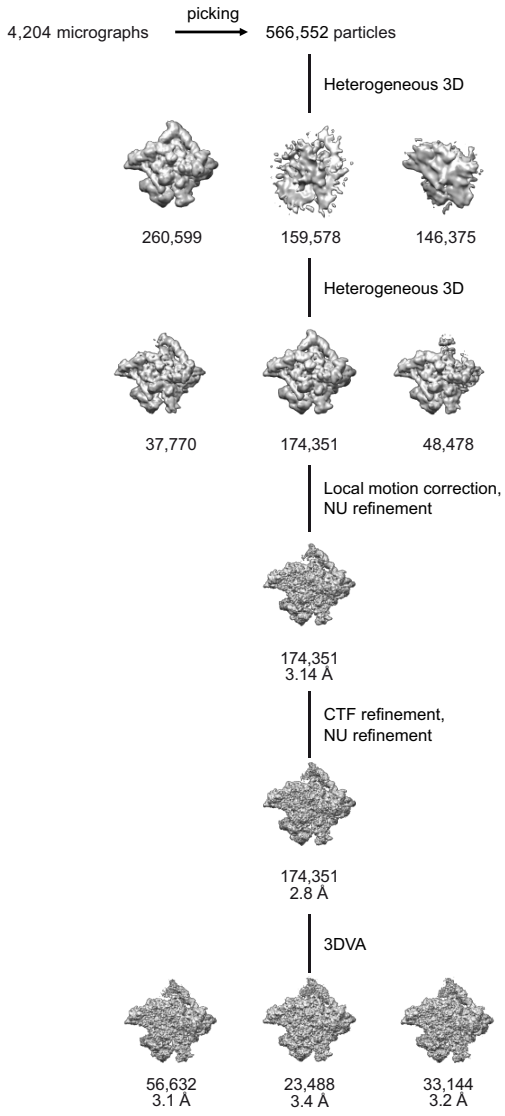
Supplementary Fig. 8 | CryoEM data processing, monomeric *S. africana* holoenzyme-DNA complex assembled at P-gre promoter. (A) Representative cryo-electron micrograph of the particles. Scale bar, 50 nm. **(B)** 2D class averages of the particle images. **(C-F)** Left, Fourier shell correlation (FSC) plots for half-maps of four states of the monomeric complex with 0.143 FSC criteria indicated; nominal resolutions 3.2 Å, 3.4 Å, 3.0 Å and 3.3 Å. Middle, the distributions of particle orientations show preferential orientation bias. Right, local resolution plots.



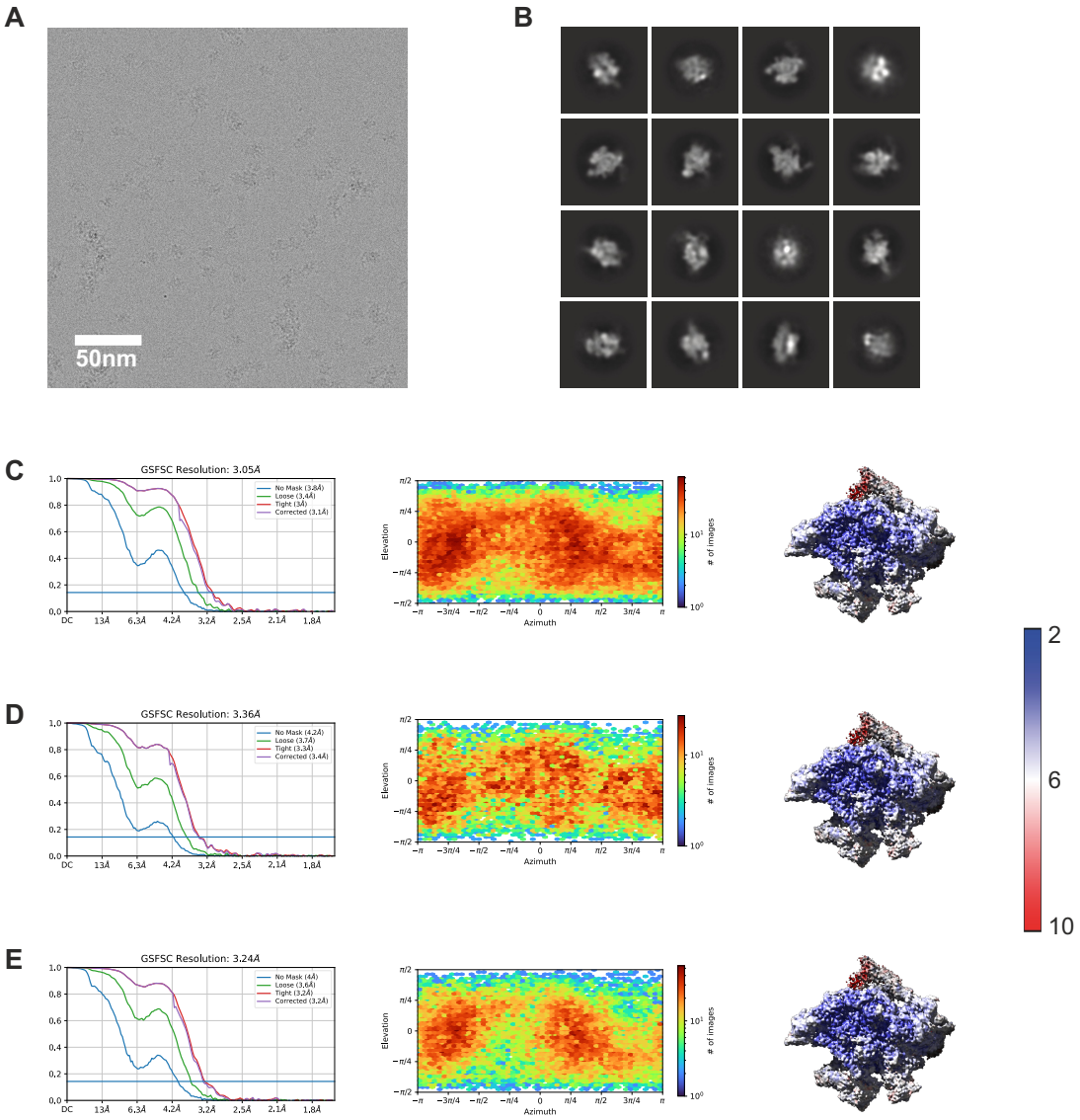
Supplementary Fig. 9 | Hierarchical clustering analysis of particle images of dimeric *S. africana* holoenzyme-DNA complex assembled at P-gre promoter.



Supplementary Fig. 10 | CryoEM data processing, dimeric *S. africana* holoenzyme-DNA complex assembled at P-gre promoter DNA. (A) Representative cryo-electron micrograph of the particles. Scale bar, 50 nm. **(B)** 2D class averages of the particle images. **(C)** Fourier shell correlation (FSC) plot for half-maps of the dimeric complex with 0.143 FSC criteria indicated; nominal resolution 3.2 Å. **(D)** The distribution of particle orientations shows preferential orientation bias. **(E)** Local resolution plots.



Supplementary Fig. 11 | Hierarchical clustering analysis of particle images of CarD-modified *S. africana* holoenzyme-DNA complex assembled at P-gre promoter.



Supplementary Fig. 12 | CryoEM data processing, CarD-modified *S. africana* holoenzyme-DNA complex assembled at P-gre promoter. (A) Representative cryo-electron micrograph of the particles. Scale bar, 50 nm. **(B)** 2D class averages of the particle images. **(C-E)** Left, Fourier shell correlation (FSC) plots for half-maps of three states of the complex with 0.143 FSC criteria indicated; nominal resolutions 3.0 Å, 3.3 Å and 3.2 Å. Middle, the distributions of particle orientations show preferential orientation bias. Right, local resolution plots.