STRUCTURAL CHARACTERISATION OF *PYROCOCCUS FURIOSUS* RNA POLYMERASE APO ENZYME AND ELONGATION COMPLEXES

DISSERTATION

zur Erlangung des Doktorgrades der Naturwissenschaften (DR. RER. NAT.)

der Fakultät für Biologie und Vorklinische Medizin

der Universität Regensburg

vorgelegt von

Daniela Madalina Tarau

aus

Romania

Regensburg, September 2025

Das Promotionsgesuch wurde eingereicht am:

22.09.2025

Die Arbeit wurde angeleitet von: **Prof. Dr. Dina Grohmann**

Unterschrift:

Table of Contents

Acknowledgements

I would like to thank Prof. Dr. Dina Grohmann for the opportunity to undertake this PhD project. I also wish to thank the colleagues at the Microbiology and Archaea Centre department for the wonderful time and teamwork.

I thank the Structural Biochemistry department for providing equipment and computational support. Special thanks go to Prof. Dr. Reinhard Rachel for the electron microscopy training. I also appreciate the external collaborators: the Würzburg University EM facility, the MPI Mass Spectrometry facility in Göttingen, and the Bayreuth University NMR group. Lastly, I would like to thank all the bachelor's and master's students I met during my PhD journey.

1. Introduction

Transcription is a fundamental and tightly regulated process occurring in all living cells. It can be defined as the process by which RNA molecules are produced from a DNA template, catalysed by a DNA-dependent RNA polymerase (RNAP). Transcription is part of the Central Dogma of Biology model, introduced by Francis Crick[1], which describes the directional flow of genetic information from DNA to RNA and finally to protein. Through transcription, the RNA molecules produced can be classified into three main types: messenger RNA (mRNA), which contains the coding information in triplets for a specific protein; and ribosomal RNA (rRNA) and transfer RNA (tRNA), which are involved in producing components of the ribosome.

The transcription process is a complex and tightly regulated multistep mechanism divided into three distinct phases: initiation, elongation, and termination. During initiation, the events include identification of the promoter region, RNA polymerase tethering, and transcription bubble formation. During the elongation phase, the RNA molecule is synthesised as the RNA polymerase advances along the DNA template. Finally, transcription concludes with the release of the newly synthesised RNA molecule and recycling of RNA polymerase[2–6].

Although RNA production occurs in every living cell, there are several differences between prokaryotic and eukaryotic transcription, particularly in the necessary transcription factors for each process stage, as well as the composition and structure of RNAP. Nevertheless, the main set of core subunits is conserved sequence-wise and structure-wise[7].

Bacterial RNAP enzyme is overall conserved in all bacterial species. It is a multi-subunit complex formed by a homodimer of the α-subunit, two highly similar β- and β'-subunit, and a dissociable subunit sigma (σ)[5,8,9]. The latter plays a crucial role in promoter region recognition of a gene or gene cluster. Bacteria contain different σ-factors specialised in different categories of gene recognition. The RNAP lacking the σ-subunit is referred to as "core enzyme", whereas when the σ-subunit is associated with the core enzyme, it is called "holoenzyme". Transcription initiation in Bacteria starts with the RNAP holoenzyme association with the promoter region located upstream of the gene of interest. Afterwards, the synthesis of RNA starts using one of the DNA strands as a template strand (TS), and it proceeds in a 5' -> 3' direction. Meanwhile, the non-template strand (NTS) is kept away from the catalytic site. After transcribing approximately nine nucleotides, typically, the σ-factor dissociates from the core enzyme, and the elongation phase starts.

During the elongation phase, the RNA chain begins to form, proceeding forward along the DNA template strand. In this passage, the newly incorporated ribonucleotides remain bound to the TS-DNA, forming a DNA-RNA hybrid of around 9-10 base pairs. Transcription elongation does not proceed continuously at a constant rate, but it is subjected to many pausing events that can be caused by base misincorporation or DNA sequence composition. When a base misincorporation happens, the polymerase back-translocates, forming a temporarily (backtracked) or irreversibly (arrested) inactive complex, shifting away the RNA's 3'-end from the catalytic centre. The misincorporated base is then

removed by intrinsic or factor-dependent nucleolytic activity. Additionally, DNA sequence motifs can pause the transcription, an example of which is the Elemental Pause Sequence (EPS). On the contrary, there are a few factors that positively stimulate elongation. The main factor stimulating TEC progression is NusG (N-utilisation substance G). In *E. coli,* it is a ~21 kDa protein composed of two domains: the NGN domain and the Kyprides, Ouzounis, Woese (KOW) domain. NusG in Bacteria covers multiple functions. It increases processivity and reduces pausing events, but it also supports transcription termination by interacting with the transcription termination Rho or by interacting with the ribosomal protein S10 during transcription-termination coupling[10–13].

Finally, transcription termination can be achieved in three different ways in Bacteria. First transcription elongation and translation can occur simultaneously in some bacterial species forming the transcription-translation coupling complex or expressome[10,11]. The two processes were found to be linked rather by NusG, NusA or RfaH, or simply linked by the nascent RNA. Secondly, the transcription termination can intrinsically occur on genes with a defined sequence. The transcribed RNA molecule must contain a GC-rich region preceding a U-tract at its 3'-end. So long as the GC-rich region can form a "hairpin loop", the RNA release from the TEC is able to occur without additional factors involvement. If the sequence requirements are not fulfilled, Rho termination factor comes into play[14,15]. Rho facilitates transcription termination and RNA release in an ATP-dependent manner. Rho tethering and association is also encouraged by the cofactors NusA and NusG which interacts through its KOW domain[15].

Eukaryotic transcription is more complex than bacterial transcription. While bacterial RNAP can transcribe the whole genetic repertoire, in eukaryotes there are three polymerases devoted to transcribing the nuclear DNA into different classes of RNA. Promoter region also differs according to which class of RNA polymerase is involved in the process. In addition, transcription phases regulation relies on many more transcription factors.

Nuclear DNAs are transcribed by RNA polymerase I (Pol I), RNA polymerase II (Pol II) and RNA polymerase III (Pol III). Pol I transcribes the precursor rRNA of three out of the four rRNAs present in the ribosome (28S rRNA, 18S rRNA, and 5.8S rRNA)[16]. This enzyme is found in the nucleolus, where ribosome maturation occurs as the precursor rRNA is being synthesised. Pol II mainly synthesises mRNA, the class of RNAs devoted to protein coding. Pol II is also transcribing RNAs with regulatory functions, such as microRNAs, long non-coding RNAs, and small nuclear RNAs. Pol III generates a variety of short RNAs, including tRNAs precursors and the smallest rRNA, 5S. Mitochondrial and chloroplast RNAPs, on the other hand, resemble bacterial structures and mechanisms of action, which agrees with the hypothetical origin of these organelles being endosymbiotic bacteria. Plants also possess a Pol IV and Pol V devoted to the production of RNAs responsible for transcriptional silencing pathways[17].

The three nuclear RNA polymerases share similar characteristics concerning the motifs and domains involved in catalysis and the overall transcription process. However, subunit composition, structural elements and transcription factors differ among the three systems. All RNAPs share a horseshoe-shaped core with the catalytic site in its centre. Moreover, the three eukaryotic RNAPs present a stalk domain formed by two subunits in Pol II and Pol III and one subunit in Pol I[2,18,19]. This subunit is not conserved in Bacteria. Besides the main subunits, Pol I and III have additional subunits[18,19], homologous to Pol II general transcription factors (GTFs) as listed in Table 1[20].

The transcription process phases share similarities with the previously described bacterial system, with a few exceptions. Transcription initiation is different in each of the three systems. It relies on the auxilium of general transcription factors. In yeast, Pol I requires Rrn3[21,22], the Core Factor (CF) complex[23], the upstream activating factor (UAF)[24] and the TATA-binding protein (TBP). Pol II transcription initiation requires TFIIA, TFIIB, TFIID, TFIIE and TFIIH together with the mediator complex as a cofactor[20,25]. TBP is also recruited to the promoter region during transcription initiation of Pol II, and it is recruited either by TFIID or the SAGA complex[26,27]. In the RNA polymerase II system, the transition from initiation to elongation is regulated by phosphorylation of Pol II C-terminal domain (CTD) in metazoans by -TEFb (CDK9/Cyclin T), and complemented by CDK12 toward the 3' end of genes[28]. Soon after initiation, Pol II is paused near the promoter region by DSIF (Spt4/5 and NusG homolog) and NELF, which stabilise the paused complex[29–31]. P-TEFb triggers pause release by phosphorylating not only the CTD but also DSIF and NELF, thereby converting them from inhibitory to supportive roles in elongation[32]. Elongation factors and complexes such as FACT, TFIIS, ELL, Paf1C, and the Super Elongation Complex (SEC) further modulate Pol II processivity, coordinate co-transcriptional RNA processing, and help navigate chromatin structure during elongation[33,34]. The Pol II termination phase starts with the recognition of the polyadenylation signal (poly A) and cleavage by polyadenylation specificity factor (CPSF). CPSF73, the endonuclease within CPSF, cleaves the nascent transcript, allowing poly(A) polymerase to add a poly(A) tail[35,36]

Lastly, Pol III requires fewer GTFs that differ depending on the transcribed region[37–40]. During initiation, TFIIIA, TFIIIB and TFIIIC GTFs are mainly involved in the process[41]. Unlike other polymerases, Pol III can complete RNA chain elongation and termination autonomously, without relying on additional cofactors.

Archaeal RNA polymerase, on the other hand, is a multisubunit enzyme that shares notable structural and functional similarities with eukaryotic RNA polymerase II, reflecting their common evolutionary origin. Like Pol II, the archaeal enzyme consists of 12–14 subunits that closely resemble their eukaryotic counterparts in sequence, structure, and organisation. This conversation extends to the catalytic core, transcription bubble formation, and the fundamental mechanism of nucleotide addition. Archaeal RNA polymerase also depends on basal transcription factors, such as TBP and TFB, which are homologous to eukaryotic general transcription factors, highlighting the parallels in promoter recognition and initiation. Simultaneously, archaeal transcription is more streamlined than in eukaryotes, lacking the complex array of additional regulatory factors and a C-terminal domain in the largest subunit of RNA polymerase. These shared features make archaeal RNA polymerase an invaluable model for understanding the evolution and mechanistic principles of the eukaryotic Pol II transcription system. Details on the structure and function of archaeal RNAPs are described in the following paragraphs.

1.1 Archaeal RNA polymerases anatomy

Archaeal RNA polymerases (RNAPs) are one example of molecular complexes resembling the eukaryotic multisubunit RNAP II, rather than the bacterial RNA polymerase complex. Genetic and structural studies have demonstrated that the two main archaeal polymerases differ between the Euryarchaeota and Crenarchaeota phyla. The two represent the major phyla of the Archaea, which encompass most of the species known to date. The Euryarchaeota phylum comprises Thermococcales, Thermoplasmatales, Halobacteriales, Archaeoglobales, Methanococcales, Methanopyrales, Methanobacteriales, Methanocellales, Methanomicrobiales and Methanosarcinales orders. Crenarchaeota phylum, on the other hand, contains the Thermoproteales, Sulfolobales, and Desulfurococcales orders[42]. The subunit composition varies between 11 and 13, depending on their origin: Euryarchaea or Crenarchaea, respectively (Figure 1 and Table 1). RNAP subunits can be classified based on their function: assembly platform, catalytic, and auxiliary subunits, as illustrated in Figures 1 and 2.

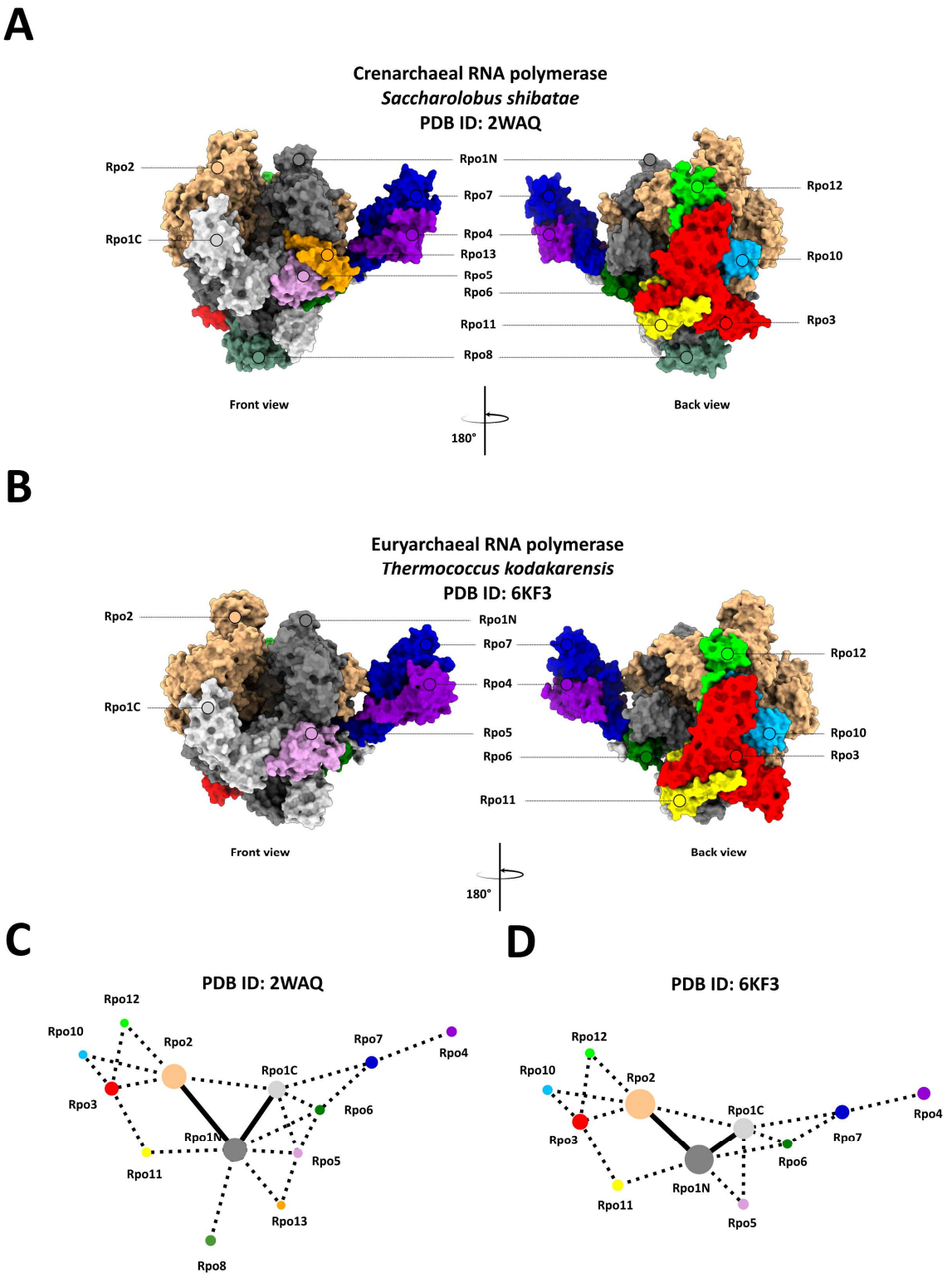


Figure 1. A. Subunit composition of crenarchaeal RNA polymerase from *Saccharoloubus shibatae* seen from the front and the back view (PDB: 2WAQ). The catalytic subunits, Rpo1N, Rpo1C and Rpo2 are colored in dim grey, light grey, and beige, respectively. The assembly platform subunits Rpo3 and Rpo11 are depicted in red and yellow, respectively. The stalk domain (Rpo4/Rpo7) is displayed in purple and indigo blue. Auxiliary subunits Rpo5, Rpo6, Rpo8, Rpo12, Rpo13, and Rpo10 are shown in lilla, dark green, forest green, light green, and orange, respectively. **B.** Subunit composition of euryarchaeal RNA polymerase from *Thermococcus kodakarensis* illustrated from the front and the back view (PDB: 6KF3). The catalytic subunits, Rpo1N, Rpo1C, and Rpo2 are colored in dim grey, light grey, and beige, respectively. The assembly platform subunits Rpo3 and Rpo11 are depicted in red and yellow, respectively. The stalk domain (Rpo4/Rpo7) is displayed in purple and indigo blue. Auxiliary subunits Rpo5, Rpo6, Rpo12, and Rpo10 are shown in lilla, dark green, and light green, respectively. **C./D.** Graphic representation of the interface analyses of *Saccharoloubus shibatae* and *Thermococcus kodakarensis* RNAPs, respectively. Analysis conducted in ChimeraX.

The catalytic subunits are mainly the largest, Rpo1/RpoA (~145 kDa in *Pyrococcus furiosus*) and Rpo2/RpoB (~127 kDa in *Pyrococcus furiosus*). These three subunits collectively create the binding channel, with its main subdomains conferring its "crab-claw" appearance (Figure1A/B). Rpo1 forms the upper jaw domain, which in eukaryotic counterparts interacts with the downstream DNA during transcription elongation[8,9]. Another critical domain in the Rpo1N subunit is represented by the clamp domain, which is the most flexible part of the polymerase. The clamp domain is mainly involved in open/closed complex formation. In the closed conformation, the clamp domain is found in the vicinity of the protrusion domain, and it is often referred to as the conformation in which the polymerase is active, rather than in its apo state or bound to nucleic acid residues[44–47]. The clamp domain is also the binding site of fundamental transcription factors such as TFEα/β and Spt4/5[44,48,49]. Archaeal polymerase also assumes an open clamp conformation, in which the distance from the protrusion domain increases significantly. The open clamp conformation is predominantly found in inactive complexes[47,50] or when associated with transcription factor TFEα during transcription initiation[44]. Clamp opening and closing are key dynamic feature that guarantees DNA accommodation, bubble formation, and overall transcription progression[3].

Rpo1N also comprises the catalytic centre, which consists of a permanently bound magnesium ion (Mg2+) coordinated by three aspartic acid residues, also known as the "Metal A" motif. In the *Pyrococcus furiosus* enzyme, the aspartic acid triad is represented by the residue's positions 458, 460, and 462. Together with the catalytic centre, a series of motifs contribute to the RNA synthesis reaction. Indeed, Rpo1N also contains the "bridge helix" collocated in the active centre. The bridge helix connects the upper part of the RNAP to the lower one. Its importance is emphasised by the fact that the primary amino acid sequence is highly conserved across Bacteria, Archaea, and Eukaryotes. The bridge helix is a slightly curved, long helix (Figure 2). Its primary role is to coordinate the forward translocation of the template DNA nucleotide, preserving the register of transcription[25]. To do so, the bridge helix assumes two different conformations during nucleotide incorporation: stretched[51] and unstretched. In correspondence with the catalytic centre, the bridge helix can transition to a "kinked" version[51–53], where a few amino acid residues become unstructured. The stretching and relaxation of the bridge helix is thought to be essential for moving the RNAP forward along the DNA template strand, facilitating RNA synthesis[53]. Mutagenesis studies conducted on recombinantly purified and assembled RNA polymerase from *Methanocaldococcus jannashii* elucidated the importance of the bridge helix motif upon transcription progression[53,54]. Residues in contact with incoming rNTP and DNA template strand mutations result in complete loss of function, while mutations of the other residues cause either decreased activity or an increase in transcription rate when serine 824 is converted to proline[53,54].

Additionally, Rpo1N features three loops that have structural relevance in ensuring that the enzyme can accurately and efficiently transcribe the DNA into RNA. The three loops are the zipper, lid, and rudder[55]. The zipper loop aids in DNA stabilisation during the transcription bubble formation. The lid loop is a key structural motif involved in overcoming the RNA association to the template strand and redirecting it towards the RNA exit channel. The rudder loop plays a crucial role in maintaining the transcription bubble in an open state due to its position near the catalytic centre in the RNA polymerase[25,47,56,57]. These functions have been extensively investigated in eukaryotic RNA polymerase II, but their structural similarities and sequence composition suggest a conserved function in Archaea[43]. The Rpo1C subunit also features a fundamental functional motif, the trigger loop. The latter is involved

in nucleotide incorporation and maintains the fidelity of the RNA synthesis. Structural studies have proven that the trigger loop oscillates between the pore (open state) and towards the catalytic centre in RNA polymerase II (closed state)[52] during nucleotide incorporation[58–60]. All multi-subunit DNA-directed RNA polymerases have this motif, and its importance is underlined by an altered transcription rate when the trigger loop is deleted[52,58,60]: point mutations in the trigger loop can result in a lethal phenotype[61] in *S. cerevisiae*, reduced processivity of the RNA polymerase[61,62], or a "superactive" form, thought to favour its closed state[53,62].

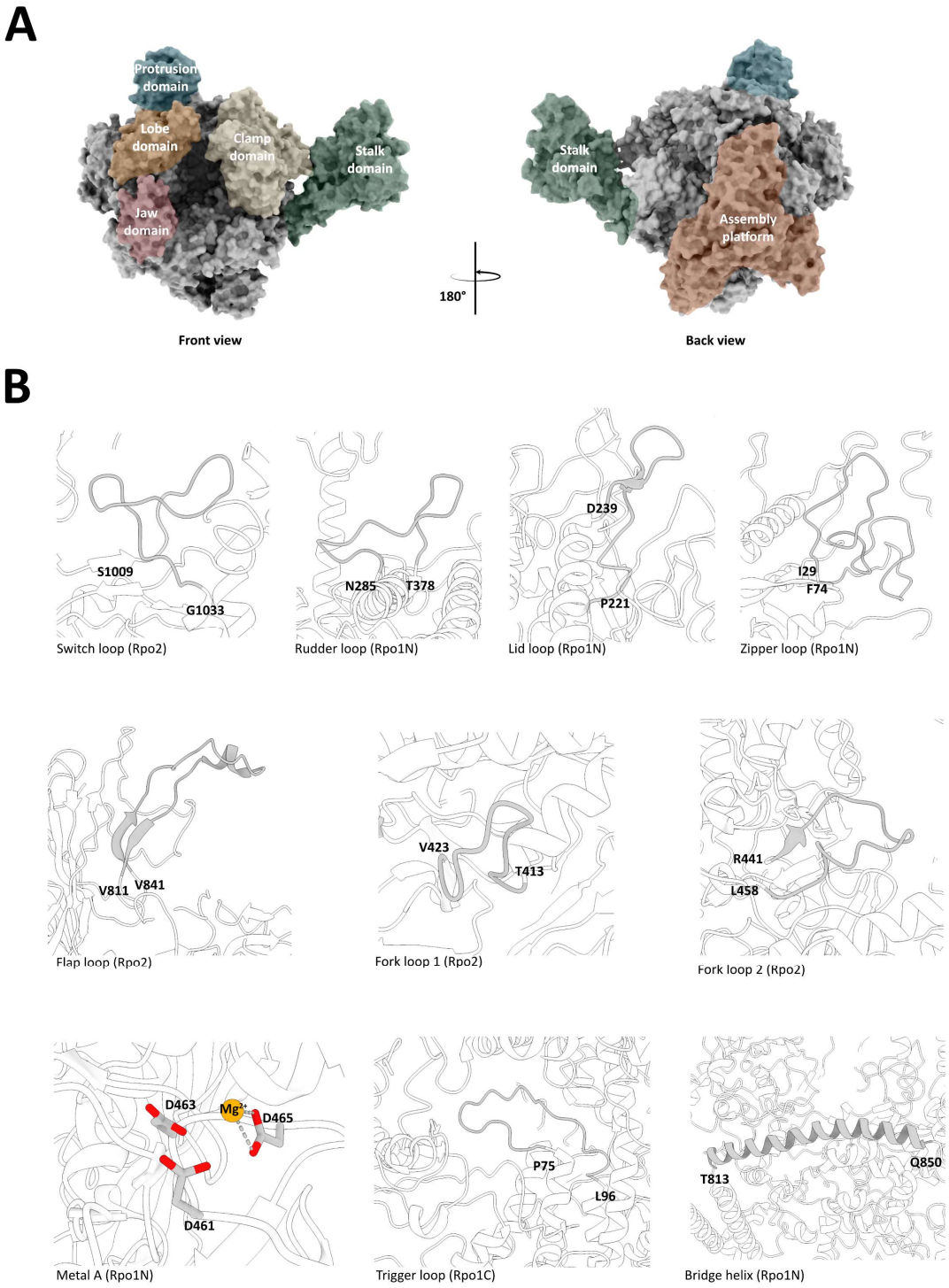


Figure 2. Schematic representation of archaeal RNA polymerase domains and motifs. A. *Thermococcus kodakarensis* RNA polymerase 3D structure (PDB ID: 6KF3) in surface representation from ChimeraX[63]. The protrusion, lobe, jaw, clamp, stalk domains, and assembly platform are highlighted in different colours. **B.** List of motifs involved in DNA bubble formation or catalysis. The first two rows represent the *tko* RNAP motifs involved in DNA melting and bubble maintenance throughout the transcription process. In the last row, the motifs involved in nucleotide incorporation are shown.

Rpo2 forms the protrusion and lobe domains, which are involved in DNA-binding channel closure, and it extends toward the clamp domain, interacting with Rpo1 and forming the RNA exit channel. A hallmark of archaeal RNAPs is that the two largest subunits can be found in a split form in some

organisms (Rpo1'/Rpo1N and Rpo1''/Rpo1C; Rpo2'/Rpo2N and Rpo2''/Rpo2C)[44,46,50,64]. Nonetheless, the splitting does not interfere with either protein folding or their function. Unpublished observations have shown that artificial fusion did not inhibit their function either [3]. Similar to Rpo1N, Rpo2 carries some functional loops, including fork loops and flap loops. The fork loops are situated in the catalytic centre, and their function in eukaryotic RNA polymerase II is to assist in separating the DNA strands, ensuring that the non-template strand is kept away from the active centre[43]. The flap loop is located close to the RNA exit channel and plays a crucial role in guiding the nascent RNA away from the DNA-RNA hybrid[43,65].

The assembly platform consists of Rpo3/RpoD-Rpo11/RpoL (~29.8 kDa and ~11.1 kDa, respectively, in *Pyrococcus furiosus*) heterodimer, which is situated at the back of the polymerase (Figure 1C/D). The Rpo3 subunit from crenarchaeal RNAPs contains an iron-sulfur cluster coordinated by four cysteine residues[47]. These four cysteine residues are conserved in the eukaryotic RPB3 subunit, but they coordinate a Zn^{2+} instead[25,30,43,65]. Nevertheless, the iron-sulfur cluster or Zn^{2+} binding domain is not conserved in known euryarchaeal Rpo3 subunit structures[30,47].

Rpo10/RpoN (~7.8 kDa in *Pyrococcus furiosus*) and Rpo12/RpoP (~5.8 kDa in *Pyrococcus furiosus*) are two small subunits containing a Zn^{2+} binding domain and fill the concave depression in Rpo2, functioning as adaptors between Rpo3 and the Rpo2 subunits.

The stalk domain is formed by subunits Rpo7-Rpo4/RpoE-RpoF (~21.7 kDa and ~14.1 kDa, respectively in *Pyrococcus furiosus*). This domain represents the main difference between bacterial RNAP versus archaeal and eukaryotic RNAP II, since it is not conserved in Bacteria. Rpo4/Rpo7 subunits are involved in transcription elongation by interacting with the newly synthesized RNA oligonucleotide[66]. Depletion of the stalk domain has revealed that the RNAP retains its catalytic activity, but processivity and precision are reduced[66].

Rpo5/RpoH (~9.2 kDa in *Pyrococcus furiosus*) is situated at the lower side of the DNA-binding channel, more precisely in correspondence with the downstream DNA double-strand accommodation site. Archaeal Rpo5 lacks one domain, which is present in eukaryotic polymerases and forms the lower jaw, creating a network of interactions with the DNA during transcription elongation[30,67].

Rpo6/RpoK (~6.2 kDa in *Pyrococcus furiosus*) is a small subunit that connects the stalk domain to the large subunit Rpo1. Rpo8 (~15.1 kDa in *Saccharolobus shibatae* and *Sulfolobus acidocaldarious*) and Rpo13 (~12.3 kDa in *Saccharolobus shibatae* and *Sulfolobus acidocaldarious*) are only found in crenarchaeal RNAPs (Figure 1A).

Rpo8/RpoG, equivalent to RPB8 in eukaryotes, is located between the assembly platform and the pore channel, whereas Rpo13 is the only archaeal-specific subunit, which interacts with Rpo5 (Figure 1C). A peculiarity of archaeal and bacterial RNA polymerase is that its function is not dependent on post-translational modifications, unlike eukaryotic RNA polymerase II[28]. Studies were conducted on the subunit association of *Pfu* RNAP[68] and *Methanocaldococcus jannashii* RNA polymerase[69]. Recombinant complexes were still able to bind to the transcription start site and synthesize RNA. The discovery has led to an understanding of the mechanism of action of transcription factors, such as Spt4/5[49]. Still, it has also allowed for the fluorescent labelling of the subunits and elucidated the dynamics of the

polymerase during the initiation phase employing single-molecule Förster Resonance Energy Transfer (smFRET) experiments[48,64,70].

Table 1. List of conserved RNA polymerases subunits and basal transcription factors.

Bacteria	**Archaea**	**Eukaryotes** RNAPII	RNAPIII	RNAPI	Plant RNAPIV‡	Plant RNAPV‡
RNAP subunits						
β-subunit	Rpo1 (RpoA)	RPB1	C160	A190	NRPD1	NRPE1
β-subunit	Rpo2 (RpoB)	RPB2	C128	A135	NRPD/E2	NRPD/E2
α-subunit	Rpo3 (RpoD)	RPB3	AC40	AC40	RPB3 [1]	RPB3 [1]
α-subunit	Rpo11 (RpoL)	RPB11	AC19	AC19	RPB11	RPB11
ω-subunit	Rpo6 (RpoK)	RPB6	RPB6	RPB6	RPB6 [1]	RPB6 [1]
	Rpo5 (RpoH)	RPB5	RPB5	RPB5	RPB5 [3]	NRPE5
	Rpo8 (RpoG) (-)	RPB8	RPB8	RPB8	RPB8 [1]	RPB8 [1]
	Rpo10 (RpoN)	RPB10	RPB10	RPB10	RPB10	RPB10
	Rpo12 (RpoP)	RPB12	RPB12	RPB12	RPB12	RPB12
	Rpo4 (RpoF)	RPB4	C17	A14	NRPD/E4	NRPD/E4
	Rpo7 (RpoE)	RPB7	C25	A43	NRPD7 [1]	NRPE7
		RPB9	C11	A12	NRPD9b	RPB9
	Rpo13 (-)					
Transcription factors						
		TFIIFα (RAP74)	C53 (C4)*	A49*		
		TFIIFβ (RAP30)	C37 (C5)*	A34.5*		
	TFEα	TFIIEα	C82*			
	TFEβ	TFIIEβ	C34*			
			C31*			
	TBP	TBP	TBP	TBP		
	TFB	TFIIB	BRF1			
		TFIIA				
		TFIIH				
	TFS	TFIIS	TFIIS			
	Spt4	SPT4	SPT4			
NusG	Spt5	SPT5	SPT5			
NusA	NusA					
Rho	aCPSF1/FttA					
σ-factors						

*Alternative names frequently found in the literature are provided in brackets. ‡Numbers within square brackets represent the count of orthologs for RNAPIV and RNAPV subunits. (-) Present in certain archaeal species but not universally observed.

1.2 Archaeal Transcription Cycle

Transcription is a fundamental process that is widely conserved across all domains of life. The key player in this process is RNA polymerase, a multisubunit complex that is able to synthesize RNA using DNA as a template. The polymerisation reaction requires one out of the four ribonucleosides 5'-triphosphate to synthesize an RNA chain. The newly synthesized RNA oligonucleotide is complementary to its template strand DNA (TS). Additionally, RNA polymerases do not require a primer to initiate *transcription de novo*. As the RNA is elongated in 5'-3' direction, the 3' end of the RNA acts as a nucleophile for the incoming nucleoside α-phosphate, attaching it to the growing RNA and releasing a diphosphate. The first incoming nucleoside 5'-triphosphate also carries a second Mg^{2+} ion (metal 2), which, together with the RNAP's catalytic Mg^{2+} (metal 1), is necessary for the catalytic reaction. RNA polymerase activity is regulated by a series of transcription factors (TF) that interact with

the polymerase and nucleic acids throughout the different transcription phases. The archaeal transcription cycle mirrors the eukaryotic one in a much simpler manner.

Also, the RNA polymerase resembles eukaryotic RNA polymerase II in terms of subunit composition and transcription factors involved in the process (Figures 1A/B and 3B). Despite all the homologies with the eukaryotic system, the genomic organisation shares similarities with the bacterial one: genes are organised in polycistronic operons, mRNAs are cap-less and rely on the Shine-Dalgarno sequence or leaderless mRNAs in the transcription process[71,72]. The archaeal transcription cycle also consists of the same three phases as bacterial and eukaryotic transcription: (i) initiation, (ii) elongation, and (iii) termination (Figure 3).

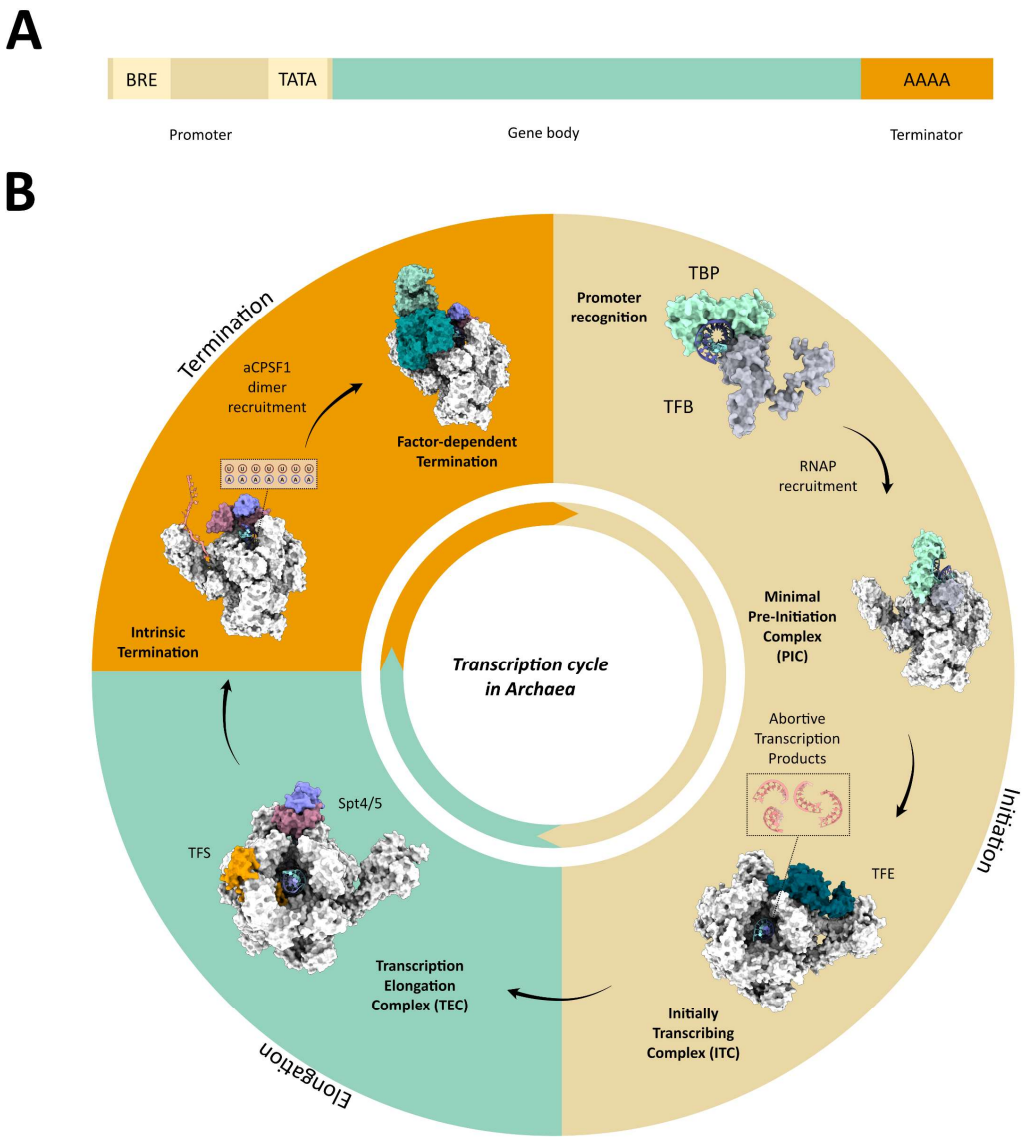


Figure 3. A. Schematic representation of an archaeal gene, starting with promoter region containing the BRE element and the TATA box (beige); the gene body (green); and the terminator A-rich tract (orange). **B.** Wheel representing the transcription cycle in Archaea. In beige are three complexes formed during the transcription initiation phase. Starting with promoter recognition by TBP and TFB, followed by RNA polymerase recruitment and formation of the minimal Pre-Initiation Complex (PIC), and finally ending with the association of TFE and the Initially Transcribing Complex (ITC). The elongation phase is depicted in green, and the Transcription

Elongation Complex (TEC) is formed. Elongation is facilitated by the cleavage factor TFS and processivity factor Spt4/5. The cycle ends with transcription termination (orange), which can occur through either intrinsic or factor-dependent mechanisms (aCPSF1 dimer, PDB ID: 9BCT).

1.3.1 Transcription initiation

Transcription initiation marks the beginning of the RNA synthesis process. During this phase, key events include the recognition of the promoter region, recruitment of RNAP, and DNA melting. Archaea utilize a simplified, eukaryotic-like mechanism for initiating transcription[3,46,54,68]. Indeed, archaeal initiation is the phase that involves two transcription factors, TBP (TATA-box Binding Protein) and TFB (Transcription Factor B), that recognise and bind to the TATA-box and the BRE element, respectively[6,73–75]. TBP is composed of a highly conserved DNA-binding core domain and a eukaryotic-specific N-terminal domain. The core domain consists of a bipartite symmetrical structure, which can recognize dinucleotide repeats (TATA repeats). The TATA-box is located approximately 25 nucleotides upstream of the transcription start site (TSS)[76]. The binding of TBP to the promoter region has been shown to induce DNA bending of 90° in eukaryotes[73,77,78].

In Archaea, TFB association with the BRE element is mainly required to stabilize DNA bending[74,79–81]. TFB, as well as eukaryotic TFIIB, is comprised of two domains connected by a flexible linker. The C-terminus domain binds to the TBP-DNA complex, forming contacts with the upstream and downstream regions of the BRE sequence element located upstream of the TATA box. The N-terminus domain contains a zinc ribbon formed by four cysteines or three cysteines and one histidine residue in most species. The NTD zinc ribbon seems to be fundamental for establishing contacts with the RNAP subunit RpoK. It is also important to mention that not all archaeal species display a conserved zinc ribbon, but in species such as *Sulfolobus acidocaldarius* or *Aeropyrum pernix,* two out of four cysteines are replaced by a serine or a threonine residue[74,82]. Subsequently, the RNAP associates with TBP and TFB on the transcription start site located 25 base pairs downstream of the TATA element[83], forming the minimal pre-initiation complex (PIC). TBP and TFB play a fundamental role in RNAP orientation and are sufficient to initiate the *de novo* synthesis of RNA[84].

Nevertheless, the transcription factor TFEα/β association with the minimal PIC connecting clamp domain and stalk domain also assists in the transcription initiation, regulating the clamp opening and closing[44,48,76]. Many euryarchaeota, including *Pyrococcus furiosus* and *Methanocaldococcus jannashii,* utilise only TFEα. Indeed, not all archaeal species have retained the gene for the *tfeB* subunit. For example, in *Haloferax vulcanii,* it has been shown that *tfeB* is not essential[85], while in other species, such as *Saccharolobus solfataricus,* it is an essential transcription general factor[86]. TFEα interacts with the RNAP by bridging the clamp domain to the stalk domain. On the clamp domain, TFEα occupies the same spot as the transcription elongation factor Spt5[44,48] in correspondence to the coiled-coil motif. Crosslinking experiments have demonstrated that archaeal TFEα also interacts with upstream DNA during transcription initiation[87]. TFEα contains a zinc ribbon domain with a Zn^{2+} ion coordinated by four cysteines in proximity to the stalk binding site, which are not conserved in all archaeal species[75]. Structural insights into TFEα binding to the euryarchaeal RNAP have elucidated that in the absence of nucleic acids, the RNAP is found in an open conformational state. In contrast, when double-stranded DNA is added to the RNAP-TFEα complex, the clamp domain adopts a closed conformation[44,48]. The

RNA polymerase assumes two different conformations during this final part of transcription initiation, which are also referred to as the Open Complex (OC) and the Closed Complex (CC). The oscillation between the two major conformational changes is crucial for transcription bubble formation, promoter escape, and RNAP progression.

1.3.2 Transcription elongation

Elongation represents the phase when RNA is being synthesised with high processivity. This process is helped mainly by two transcription factors, Spt4/5 and TFS. The polymerase begins synthesizing RNA with the formation of the initially transcribing complex (ITC). But to do so, promoter escape and factor swapping must occur first. TFEα binds to the same site as transcription elongation factor Spt4/5[48]. Therefore, TFEα/β must be replaced before the elongation phase starts. Spt5/DSIF in archaea and eukaryotes or NusG (N-utilisation substance G) in Bacteria is the only transcription factor conserved across all domains of life. This has drawn attention to the elongation first theory[3] speculating that the origin of transcription might have started from this phase.

Archaeal Spt5 is composed of an NGN domain, which interacts with the RNAP clamp core domain[49], and one Kyprides, Ouzounis, Woese (KOW) domain, resembling bacterial NusG[3,88]. On the other hand, eukaryotic SPT5/DSIF differs from archaeal and bacterial homologs by binding mechanism, and it can contain up to seven KOW domains[29]. Archaeal and eukaryotic Spt5/SPT5 act as a heterodimer in complex with Spt4, a compacted zinc-ribbon motif protein. Literature reports that the binding web involves a hydrophobic pocket of the NGN domain of Spt5 and NusG, where the coiled-coil motif of the RNAP clamp-core helices accommodates[6,29,30,64,89]. In addition, the NGN domain also interacts with the protrusion domain of the RNA polymerase, impeding the transition to an open-clamp conformation. The transcription elongation process in Archaea occurs in the absence of transcription factors, although the transcription rate is notably increased by the addition of Spt5 to the TEC *in vitro*[49,90,91]. As well as for the bacterial NusG, the archaeal KOW domain is involved in the recruitment of transcription termination factor aCPSF1, and it is hypothesised to be the linker between transcription and translation processes[92–95].

At the initial stage of the elongation phase, short RNA fragments (8-12 NTP) are produced[3,96]. These nascent transcripts, also referred to as abortive products, do not contribute to productive elongation. This phenomenon is hypothesised to result from persistent and stable contacts of the RNAP, promoter DNA, and transcription factors in the initiation complex[97–99]. Their interactions are thought to be sufficiently strong to slow down the transition from the initiation to the elongation phase, thereby causing repeated cycles of transcript synthesis before the RNAP successfully escapes the promoter region and enters the elongation phase. It is hypothesised that TFS (Transcription Factor S), known from eukaryotes to be involved in cleavage of these short RNA transcripts[100,101], might also be involved in RNA degradation of abortive products. Archaeal genomes encode different isoforms of TFS, which are involved in transcription proofreading, and some paralogs exhibit an antiviral effect[47,102]. *In vitro* transcription studies conducted on *Methanococccus* have elucidated the importance of TFS in pausing escape and cleavage of RNA 3'-end after misincorporation[103]. Recently, the 3D structure of a TFS4 paralog and RNAP complex from *Sulfolubus acidocaldarius* has revealed that it binds between the

RNAP lobe and upper-jaw domains, extending towards the pore channel. The RNAP is found in an open clamp conformation, which inhibits the ability to seal the DNA binding channel and therefore prevents transcription initiation and elongation phases of viral toxins[47,102].

In addition to the well-characterised transcription elongation factors TFS and Spt4/5, NusA also plays a significant role in modulating pausing events in bacteria. Specifically, bacterial NusA has been shown to both stimulate and prevent transcriptional pausing[104–106]. In contrast, the role of archaeal NusA homologs appears to diverge from that of the bacterial counterparts, despite partial sequence conservation. Structural and functional studies have demonstrated that the two archaeal NusA1 and NusA2 have a preferential affinity for single-stranded CU-rich RNA[107,108]. However, current evidence suggests that neither NusA1 nor NusA2 is likely to associate with the elongating RNA polymerase stably.

1.3.3 Transcription termination

Finally, transcription ends with the termination phase. During termination, the newly synthesised RNA oligonucleotide detaches from the template strand, and the RNA polymerase recycling occurs. In Archaea, termination can be induced by intrinsic sequence motifs and by external protein factors. Intrinsic termination, similarly to bacterial termination[109,110], relies on a specific sequence region present at the 3'-end of the transcript[111,112]. In Archaea, it has been proven that short uridine tracts are enough to properly terminate the transcription process according to *in vitro* studies conducted on *Methanothermobacter thermautotrophicus* RNAP[113]. This sequence length is not identical to all species of archaea; in particular, *in vivo* and *in vitro* studies have proven that transcription termination in euryarchaea occurs downstream of the uridine tracts without secondary structure formation[112,114]. Regarding crenarchaeal termination signals, studies confirm the presence of a uridine-rich region divided into two positions, one found proximal to the terminator and a secondary region positioned -20 to -12 nucleotides upstream[112,115].

Factor-dependent termination in archaea has been described to be mediated by the factor Eta, although a general function as a termination factor at all termination events has not been proven[116,117]. On the other hand, one transcription factor, aCPSF1 (FttA), has been proposed as an actual component of the transcription termination complex[92,93,118]. aCPSF1 has a eukaryotic homologue, CPSF73, and exhibits an endoribonucleolytic activity. It acts as a dimer, and each monomer is composed of a catalytic bimetal centre (two Zn^{2+}) coordinated by six histidine residues, one metallo-β-lactamase domain, one β-CASPase (β-CASP) domain, and two K homology domains (KH1 and KH2). Recent studies have demonstrated that the mechanism of action *in vitro* of aCPSF1 from *Thermococcus kodakarensis* is positively influenced by the presence of Spt4/5, suggesting that the termination phase resembles bacterial NusG-Rho-mediated termination[15]. A 3D model of the transcription termination complex *Thermococcus kodakarensis* has been recently solved[93,109]. In this study, the authors have used an inactive aCPSF1^{H225A} mutant converting one of the histidine residues to alanine, thereby coordinating the two catalytic zinc ions. This study demonstrated that aCPSF1 binds both to the RNAP, more precisely to the stalk domain, and to the KOW domain of Spt5.

Transcription elongation and termination in archaea is also hypothesized to be linked to the translation machinery, as has been shown in Bacteria[10–13,95]. This phenomenon is known as transcription-translation coupling, and, due to the absence of a nuclear membrane in prokaryotes, the two processes are thought to occur simultaneously; however, no studies to date have proven this. In some Bacteria, the coupling of the two machineries is bridged by NusG, NusA, or RfaH factors[11–13]. Nevertheless, the presence of these factors is not mandatory for the occurrence of transcription-translation coupling, according to studies conducted *in vitro*[12,12]. Recent studies in *Bacillus subtilis* have also shown that transcription and translation coupling is not conserved throughout all prokaryotes[119]. Since Spt5 is highly homologous to NusG, it is expected that archaeal Spt5 might also bridge the transcription-translation machineries by direct binding of the Spt5-NGN domain to the elongating RNAP and by binding of the Spt5-KOW domain to the ribosomal protein S10[95].

1.3 Aim of the study

The aim of this study is to determine the atomic structure of *Pyrococcus furiosus* RNA polymerase during different steps of the transcription cycle. A key question of this study was how the transcription elongation factor Spt4/5 influences the conformational state of the RNAP and the TEC. Cryo-Electron Microscopy (Cryo-EM) and Single Particle Analysis (SPA) were employed to determine the 3D structure of *Pfu* RNAP transcription complexes in its apo-state, Transcription Elongation Complex (TEC), and the structural influence of the transcription elongation factor Spt4/5 on the apo RNAP and the TEC. Connecting structural information with the functionality of the RNAP, biochemical experiments should provide insights into the structure-function relationship in the TEC. Additionally, an attempt was made to verify the interaction contact network between the Spt5-KOW domain and ribosomal protein S10 in the context of the transcription-translation complex. Several methods were employed to determine the binding of the two proteins, including MST thermophoresis, mass spectrometry analysis, and crosslinking experiments. With the recent advances in the electron microscopy, such as improvement of electromagnetic lenses, introduction of Field Emission Gun (FEG) and most importantly the introduction of electron direct detectors, alongside to X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and single-molecule FRET (smFRET), cryo-EM is now one of the structural biology tools that allow the collection of structural information with atomic resolution.

1.4 *Pyrococcus furiosus* as archaeal model organism

The term "Archaea" comes from Greek, and it means "ancient" or "primitive". In fact, archaea are believed to have existed on Earth a long time ago and represent the precursors of bacterial and eukaryotic cells. Archaea were torn apart from Bacteria for the first time in 1977 by Carl Woese, while conducting studies on ribosomal RNA[120]. In this study, the author has divided Bacteria into three categories: *eubacteria*, comprising the typical bacteria known to date; *archeobacteria,* including the methanogenic Archaea; and *urkaryotes*, as the endosymbionts of eukaryotic cells (mitochondria and chloroplasts). To date, Archaea are considered single-celled microorganisms that constitute one of the three primary domains of life, alongside Bacteria and Eukarya. Extremophilic archaea thrive in various challenging environments, including terrestrial hot springs of volcanic origin, hydrothermal vents on the ocean floor, and a diverse range of saline, acidic, alkaline, and anaerobic settings[121], including the human gut[122]. *Pyrococcus furiosus* (*Pfu*) was isolated from Vulcano Island, Sicily, Italy[123]. The hot water vents close to the island's beach are known for reaching a depth of approximately 3 meters (Porto di Levante). This accessibility facilitates sample collection, distinguishing it from the challenges associated with deep-sea settings. Alongside *Pfu,* numerous other well-studied hyperthermophiles, such as *Pyrococcus woesei*, *Thermococcus celer,* and *Pyrodictium occulum*, have also been isolated from this source.

Pyrococcus furiosus was first described in 1986[123] as a member of the genus *Pyrococcus*, known as the "fireball". The reason for choosing its species name (*furiosus*) is mainly due to its doubling time of approximately 37 minutes, together with its strong motility. *Pfu* grows optimally in anaerobic

conditions using sulphur compounds for its respiration, at 100°C, and prefers sugar (such as starch) for its metabolism. *Pfu* belongs to the Euryarcheota phylum or Methanobacteriota phylum and the order of *Thermococcales*. Over the years, *Pfu* has been extensively studied due to its peculiar metabolic pathways and has been used as a resource for robust enzymes. One of the most famous examples of industrial application of *Pfu* enzymes is the DNA polymerase I[124], which possesses a 3'-to-5' exonuclease activity. Its proofreading ability reduces the error rate in PCR reactions when compared to Taq-DNA polymerase.

2. Results

2.1 Spt4/5 purification

Spt4 and Spt5 are essential transcription elongation factors that tightly associate with each other to form a heterodimer. To determine their function and structure in the context of transcription elongation, a large-scale purification was necessary. Recombinant Spt4 and Spt5 overexpression was performed in chemically competent *E. coli* Rosetta (DE3) cells separately. Given the tight interaction between the two proteins, it was possible to purify the complex by exploiting the Spt4 6xHis-tag after simply mixing the two lysates obtained as described in chapter 3.2.5. Hence, the purification of the Spt4/5 heterodimer was carried out in a two-step purification: through a first step of affinity chromatography and a size exclusion chromatography step. The chromatogram profiles from the elution steps of the affinity and size exclusion chromatography runs showed high absorbance at 280 nm. The presence of proteins with matching molecular weights for Spt4 and Spt5, at 7.8 kDa and 16.8 kDa, respectively, was verified by SDS-PAGE, confirming that the two proteins co-purified (Figure 4).

The concentration of the final protein was determined using absorption spectroscopy and calculated to be 5.3 mg/mL. Considering the extinction coefficient of Spt4-6xHis-tag and Spt5 being 8,480 M^{-1} cm^{-1} and 7,450 M^{-1} cm^{-1}, respectively, calculated on the ProtParam (Expasy)[125] Using an online tool, the concentration of the heterodimer is determined to be 332.7 µM. The amount of protein obtained corresponds to 10.6 mg of highly concentrated and pure heterodimeric complex, obtained from the overexpression of each partner in 1 L. The purity of the complex also benefits from the heat-shock step in the purification protocol, which can be performed because both proteins are thermostable.

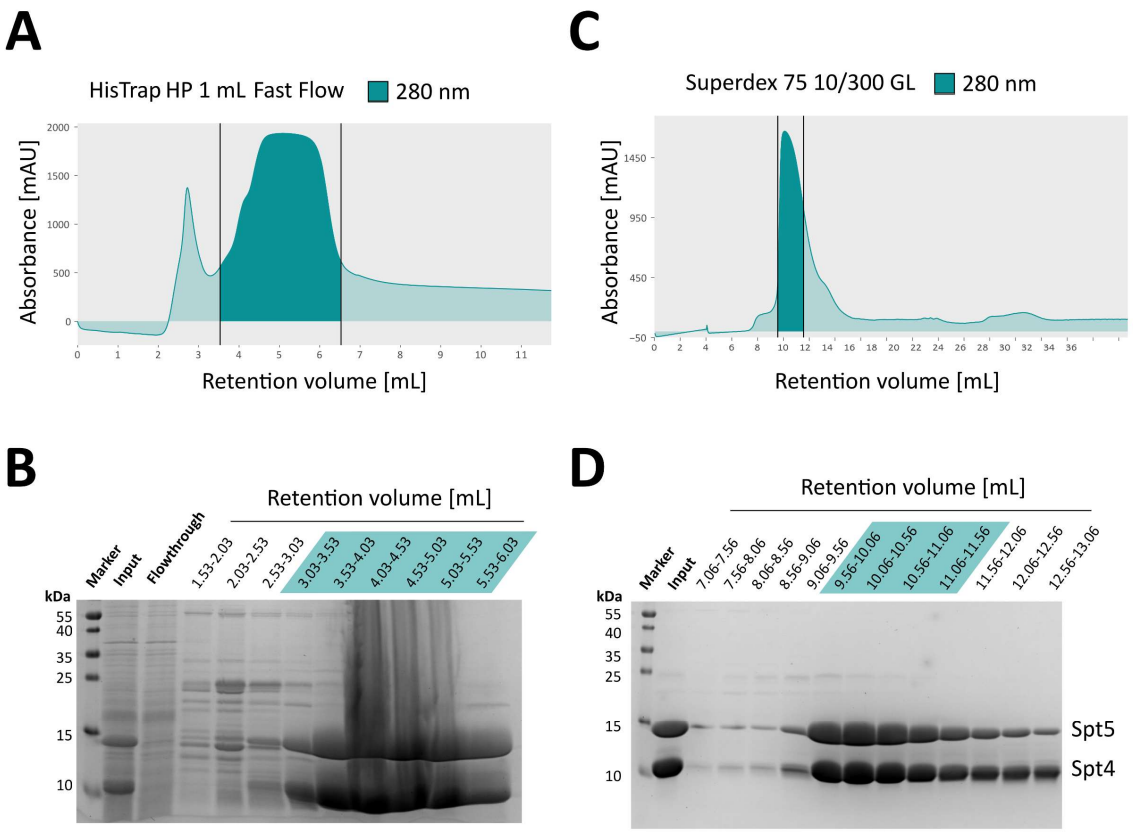


Figure 4. Purification of the transcription elongation factor *Pfu*Spt4/5. A. Affinity chromatography run (performed as gradient elution), absorbance profile with highlighted fractions corresponding to the Spt4/5 heterodimer peak. **B.** 16% SDS-PAGE loaded with input, flowthrough, and fractions corresponding to the peaks of the affinity chromatography elution. The highlighted fractions (turquoise) were collected. **C.** Chromatogram from a size exclusion chromatography run performed using the selected fractions from the affinity chromatography shown in panels **A.** and **B.** Highlighted is the elution peak containing Spt4/5. **D.** 16% SDS-PAGE loaded with input, flowthrough, and fractions corresponding to the peaks of the size exclusion chromatography elution. The highlighted fractions were collected and stored at -80°C until further use.

2.2 *Pyrococcus furiosus* RNA polymerase purification

Pfu RNAP was endogenously purified using a genetically engineered strain (PMUR1), which David Pöllmann[126] previously produced. In this strain, a StrepII-6xHis tag is encoded at the C-terminal end of the Rpo3 genomic copy. To obtain enough protein material for structural studies, a large-scale cultivation of the PMUR1 *P. furiosus* mutant in a total volume of 100 L in the in-house fermentation facility was performed. The cultivation yielded 117 g of cell pellet after centrifugation, which was plunge frozen in liquid nitrogen and stored at -80°C until further use. The cell pellet was resuspended by bead-beating and broken through sonication as described in Chapter 3.2.5; the lysate was further cleared by ultracentrifugation to eliminate residual S-layer protein and precipitates produced by *P. furiosus'* metabolism. Finally, the lysate was sterile-filtered and loaded onto a nickel-based column to perform affinity chromatography exploiting the 6xHis-tag. After loading the approximately 350 ml of lysate, the column was washed with *Pfu* RNAP Lysis Buffer until the absorbance at 280 nm dropped near to 0 mAU. The elution was performed in one step, using 100% *Pfu* RNAP Elution Buffer 1

containing 250 mM imidazole with the aim of eluting most of the protein in a small volume. The corresponding chromatogram of the affinity purification elution step is shown in Figure 5A. The chromatogram exhibits a high peak spanning over 14 fractions, each 500 µL in volume. SDS-PAGE verified the presence of Pfu RNAP, and the fractions corresponding to the peak were pooled and loaded on a gel filtration column.

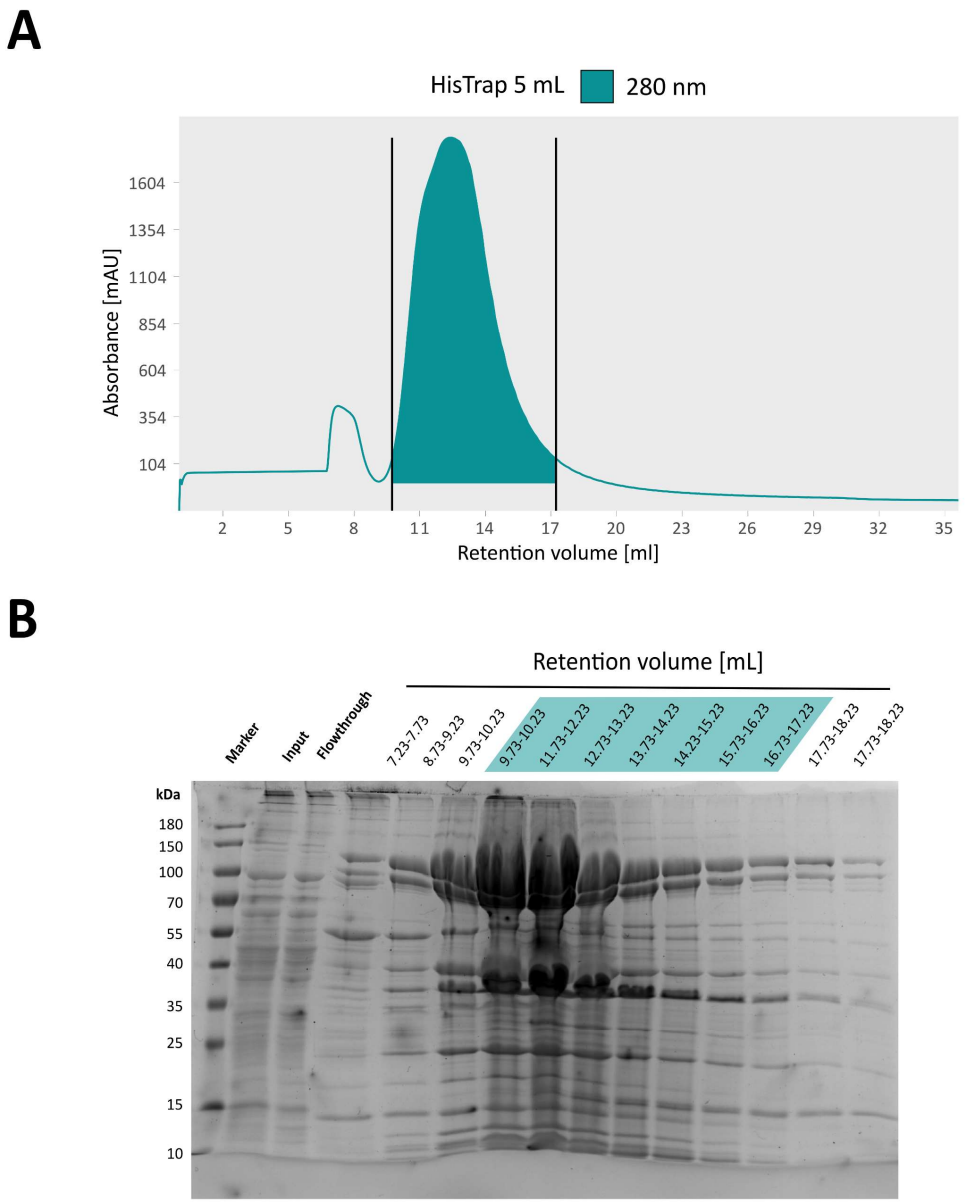


Figure 5. A. Chromatogram from the affinity chromatography run performed as step elution of the *P. furiosus* RNA polymerase. The peak containing the polymerase is highlighted in turquoise. **B.** 16% SDS-PAGE is showing the input, flowthrough, and fractions corresponding to the peak. The highlighted fractions were collected.

Since the amount of eluate from the affinity chromatography was very high, size exclusion was performed on a column with 360 mL bed volume. The reason for not concentrating the affinity chromatography eluate to a lower volume was that *Pfu* RNAP tends to lose its stalk domain (subunits Rpb4/7), as previously known (Dr. Robert Reichelt). The size exclusion chromatography was performed using isocratic elution in 1 mL fractions. SDS-PAGE was used to assess the purification success, and the polymerase was divided into two pools. In the first pool, the most concentrated fractions were

collected (Figure 6). This preparation procedure resulted in a final concentration of 1.685 mg protein /mL. This part of the eluate was aliquoted in 100 µL aliquots and stored at -80°C. This preparation was primarily used for structural studies, as it was the least subjected to laboratory processes, from the cell to the protein mixture, making it the most native fraction. Nevertheless, the rest of the size exclusion chromatography eluate was also collected and concentrated, yielding a preparation with a final concentration of 1.70 mg/mL for a total amount of 15,5 mg of *Pfu* RNAP from a 100L cell culture. The remaining eluate was also aliquoted, plunge-frozen in liquid nitrogen, and stored at -80°C. This second pool was mainly used for *in vitro* transcription assays. The presence of nucleic acids co-purified with the *Pfu* RNAP was not detectable during the purification steps due to limitations of the Äkta system. The two Äkta systems used for affinity and size exclusion chromatography were capable of monitoring only one wavelength, which was set to 280 nm. The *Pfu* RNAP subunits were mostly visible at the expected molecular weight (Table 2) on the 4%-20% gradient SDS-PAGE run loaded with an optimal amount, as shown in Figure 7A, whereas the purification SDS-PAGEs (Figure 5B and 6B) were overloaded.

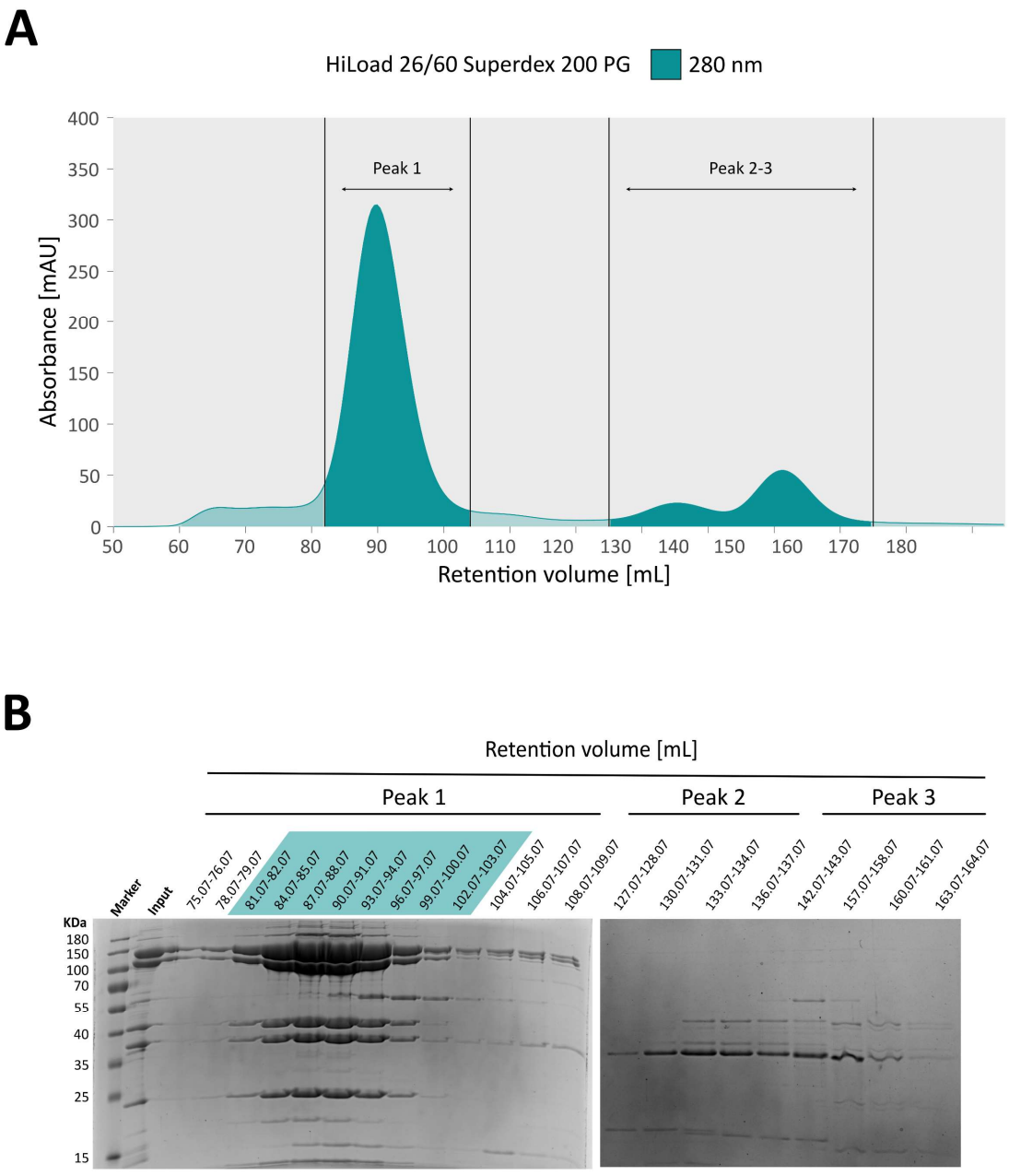


Figure 6. A. Chromatogram of the size exclusion chromatography step to purify the *P. furiosus* RNA polymerase showing the protein absorbance at 280 nm the three different elution peaks were analysed on a 16% SDS-PAGE. **B.** On the left gel, in Peak 1, are contained the fractions corresponding to the full RNA polymerase complex. Peaks 2 and 3 mainly contain free Rpo3 subunit and few other contaminants.

Table 2 List of *Pyrococcus furiosus* RNA polymerase subunits with their correspondent molecular weight and extinction coefficient.

Pfu RNAP Subunit	*Molecular Weight (kDa)*	*Extinction coefficient*
Rpo2	127.0	126,060
Rpo1N	103.5	94,200
Rpo1C	44.4	14,900
Rpo3-tags	31.6	38,390
Rpo7	21.7	22,920
Rpo4	14.1	12,950
Rpo11	11.1	8,480
Rpo5	9.2	7,450
Rpo10	7.7	7,450
Rpo6	6.2	1,490
Rpo12	5.8	4,470

Subsequently, the molecular weight of the *Pfu* RNAP was determined by mass photometry, a powerful and label-free technique that enables the precise measurement of molecular mass in solution (Chapter 3.2.11). Moreover, mass photometry facilitates the detection of potential contaminants, degradation products, or the presence of oligomeric species. Before measurement, the sample buffer was exchanged to *Pfu* RNAP cryo-EM Buffer, since its original buffer contained 20% glycerol. This passage was necessary because high concentrations of glycerol are incompatible with mass photometry. Figure 7C shows that, despite the number of counts, the main population of the sample primarily accumulates at a molecular mass of 398 kDa. The actual molecular weight of the 11 subunits *Pfu* RNAP is 382.3 kDa (Table 2), but this could be because the standard deviation for this measurement is 17.8 kDa. To assess the purity of the RNAP preparation, a more sensitive approach was employed.

According to laboratory knowledge, it is possible to purify transcription elongation complexes directly from *P. furiosus* cells following a mild purification protocol. Hence, transcription factors that are tightly associated with the *Pfu* RNAP in the three different phases of transcription could be present in the final step of the purification in minor amounts. The gradient 4%-20% SDS-PAGE of the *Pfu* RNAP in Figure 7A shows a band for almost all subunits, although the detection of co-purified proteins is incompatible with this technique. To determine whether transcription elongation factor Spt5 was present in the final RNAP preparation, a Western blot was performed using an antibody directed against Spt5. A small volume of the purified *Pfu* RNAP was loaded onto a 16% SDS-PAGE gel. Alongside the polymerase, recombinantly expressed and purified negative and positive controls were also loaded. For the positive controls, 100 µg each of wild-type Spt4/5, the Spt4/5$^{H12A/R64A/H65A/R67A}$ mutant used in the *in vitro* assays described in Chapter 2.7, and wild-type Spt5 were included. Additionally, 100 µg of recombinant Spt4 was added to the run. As shown in Figure 7B, Spt5 was successfully co-purified with the *Pfu* RNAP.

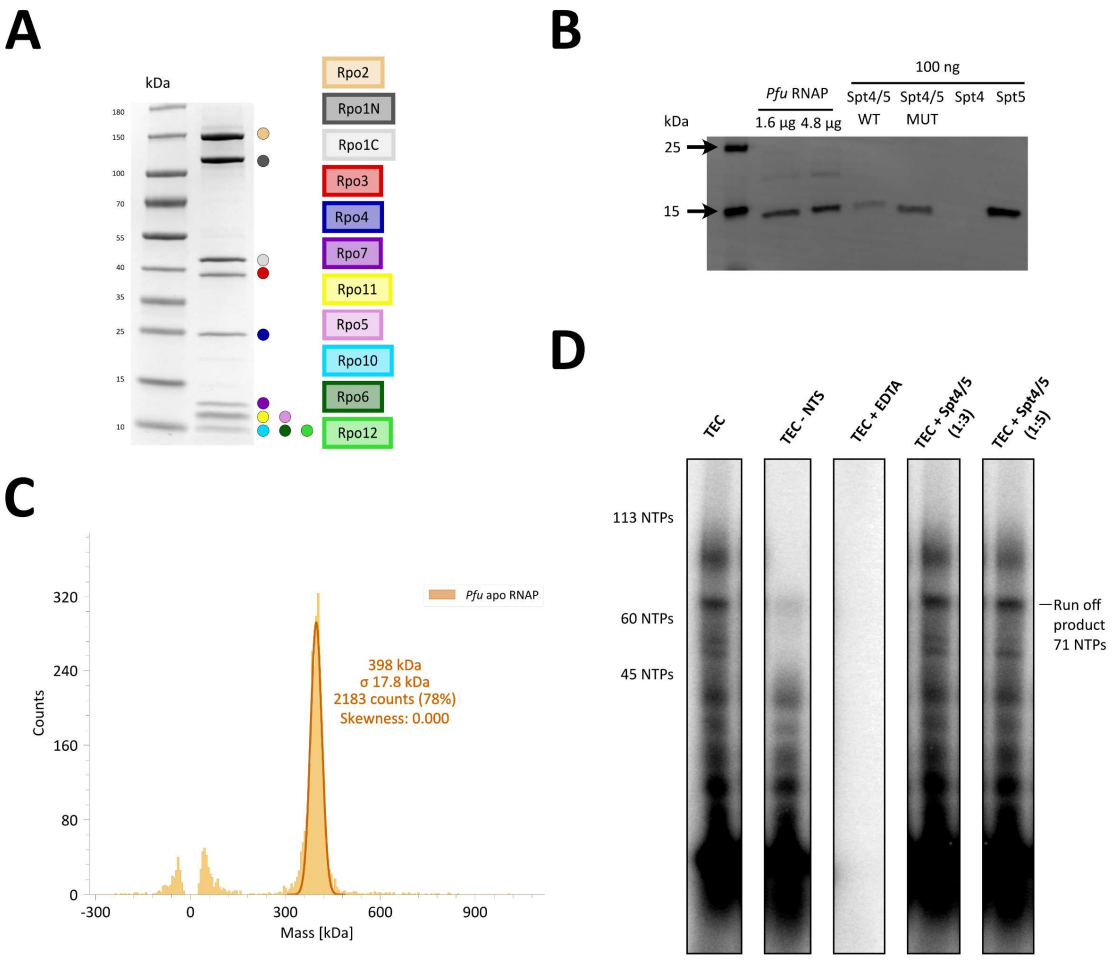


Figure 7. Homogeneity and activity tests of natively purified *Pfu* RNAP. A. SDS-PAGE analysis using a polyacrylamide gradient (gradient concentration 4%-20%) shows bands for all RNAP subunits, indicating successful purification**. B.** Western blot analysis of samples derived from affinity-purified *Pfu* RNAP from a cell lysate from a *Pfu* variant that expresses His-tagged subunit Rpo3. Recombinantly expressed and purified wild-type or mutant Spt4/5$^{H12A/R64A/H65A/R67A}$ (100 ng), as well as the individual recombinant proteins Spt4 and Spt5 (100 ng), were loaded as controls. For detection, an antibody directed against the elongation factor Spt4/5 was used. **C.** Mass photometry analysis of *Pfu* RNAP (ochre graph). The measurement displays a peak with the greatest number of counts slightly higher than the expected molecular weight of 382 kDa **D.** *Pfu* RNAP activity test with radioactivity detection. In the first lane, the full transcription elongation complex was assembled; in the second lane, the TEC was assembled without the non-template strand; in the third lane represents the negative control, in the last two lines, two different concentrations of Spt4/5 were added to the TEC in the transcription reaction.

Next, the capacity of *Pfu* RNAP to bind a synthetic elongation scaffold and its polymerisation activity were tested. This is a critical assessment to confirm the activity of the protein complex before proceeding with structural biology techniques. To this end, *in vitro* transcription assays were initially performed, building on previous knowledge from *in vitro* assembled *Mj* RNAP activity studies[49,66]. Figure 7D illustrates a successful transcription assay. Starting from a synthetic elongation scaffold, *Pfu* RNAP can elongate the partially annealed RNA until the end of the DNA template strand. The transcription progress is monitored by radioactivity detection, as a minimal amount of ^{32}P-UTP is added to the rNTP mixture in the transcription reaction. The negative control also does not show a signal, as EDTA chelates the magnesium ions strictly required for the polymerization reaction. The addition of

Spt4/5 in 1:3 and 1:5 molar ratios with respect to the polymerase concentration does not result in a notable increase in signal.

2.3 *Pyrococcus furiosus* RNA polymerase apo-form structural characterisation

Although only a limited number of archaeal RNA polymerases apo-structures are currently available in the Protein Data Bank[127], a complete atomic structure of *Pfu* RNAP has yet to be fully resolved. This gap in structural data has hindered a comprehensive understanding of how *Pfu* RNAP compares to the already characterized transcription machineries. The apo-form, representing the unbound, transcriptionally inactive state, serves as a crucial reference for understanding conformational changes associated with transcription initiation and elongation. To address this and determine whether *Pyrococcus furiosus* transcription machinery assumes the same conformations as the already characterized archaeal RNAPs, the atomic structure was determined. A step-by-step approach was employed, starting with negative staining analysis and then performing a high-resolution cryo-EM data collection.

The atomic structure of *Pyrococcus furiosus* RNA polymerase apo-form was determined following the sample preparation steps described by Passmore and Russo in 2016[128]. First, protein complex homogeneity was confirmed by mass photometry measurements (Figure 7C). Next, to determine the particle distribution over the concentration relationship, a negative staining dataset of 23 micrographs was acquired. The sample preparation consisted of simply diluting the stock *Pfu* RNAP to a final concentration of 50 µg of protein in 100 µl final volume. The dataset was acquired at 40,000x magnification and a pixel size of 2.7 Å. The microscope was operated using SerialEM[129] Software and the micrographs were further processed in RELION 3.1[130] (Figure 8A). 72,089 particles were initially picked using the Laplacian of Gaussian auto-picking tool, which is incorporated in RELION. After a round of 2D classification, 17 classes containing 45,662 particles were selected for 3D analysis. It is essential to note that 3D reconstructions from negative staining datasets only allow the determination of the overall architecture of the complex under analysis. In this case, the final 3D classification yielded two classes with slightly different conformations: an open clamp conformation (Class 1 - 16,068 particles) and a closed-clamp conformation (Class 2 - 16,558 particles) as illustrated in Figure 8B.

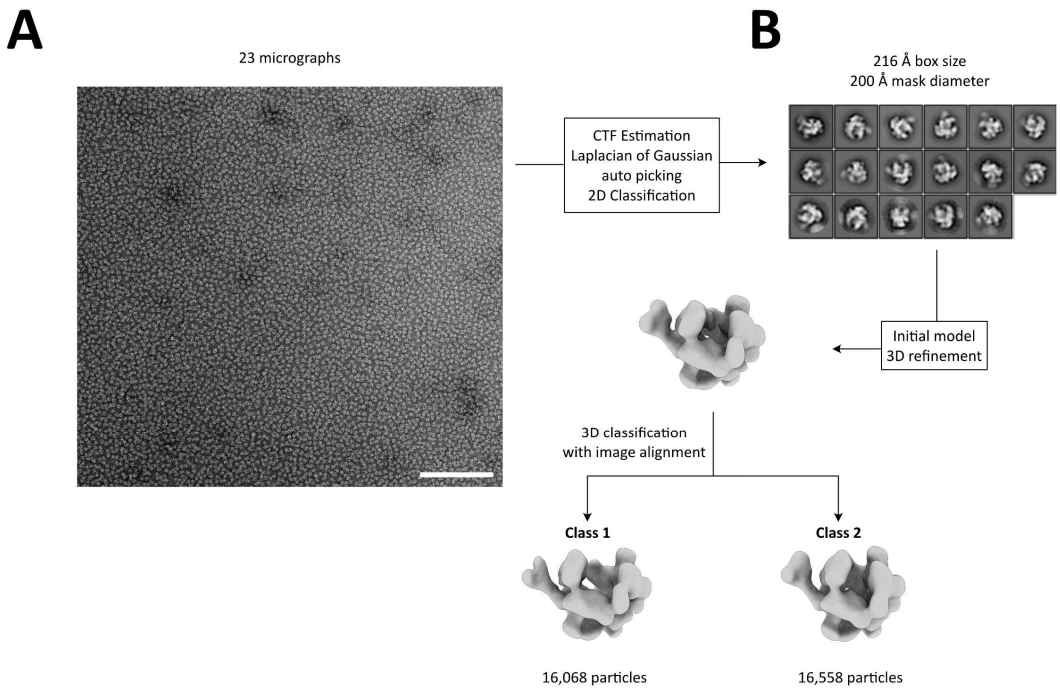


Figure 8. Negative staining analysis on *Pfu* RNAP apo enzyme. A. Negative staining analysis of the apo *Pfu* RNAP comprising one of the 23 micrographs collected (scale bar 500 µm). **B.** Shown is the Relion 3.1 processing scheme reuslting the 2D classes from which the two final 3D conformations were obtained. The first 3D class (left) is displaying a wider clamp opening compared to the second 3D class (right) which is found in a closed clamp conformation.

The negative staining dataset micrographs show particles at high density, making it a suitable concentration for the single particle cryo-EM technique. Specimen preparation for cryo-EM consisted of running an isocratic size exclusion chromatography of 168 µg of *Pfu* RNAP using the Superose 6 Increase 3.2/300 column. This passage was necessary to exchange the buffer to the *Pfu* RNAP Cryo Buffer to replace the glycerol-containing *Pfu* RNAP buffer.

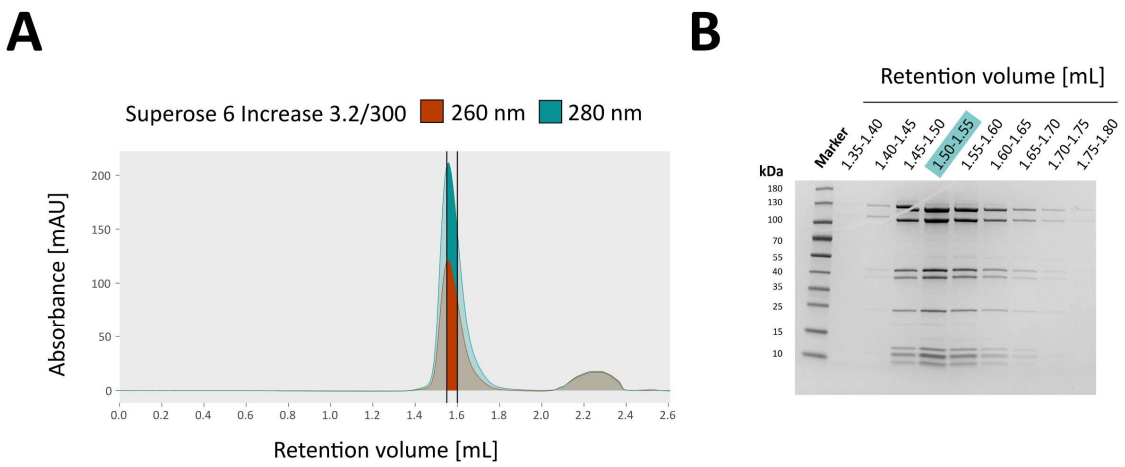


Figure 9. Sample preparation for cryo-EM data collection and analysis. A. Chromatogram of the *Pfu* RNAP apo enzyme sample preparation with highlighted fraction used for cryo-EM analysis. **B.** 4%-20% SDS-PAGE of the fractions corresponding to the peak fractions from the chromatogram in **A.** Highlighted are the fractions used for cryo-EM grid preparation.

The first dataset was collected at the University of Würzburg, in the Böttcher Group. The micrographs of this first dataset were overcrowded, and the sample was prepared again using only 100 µg of *Pfu* RNAP as input (Figure 9A). The mixture was run over the same column as before, and the success of the run was assessed by SDS-PAGE (Figure 9B). The purified *Pfu* RNAP was used for data acquisition at the in-house CRYOARM™ 200 (JEM-Z200FSC) Field Emission Cryo-Electron Microscope (Jeol). The dataset was processed in Relion 4.0 and CryoSPARC, yielding three density maps with nearly atomic resolution (Appendix Figure 36). The processing analysis has resulted in two main conformations in Relion 4.0: the expanded conformation and super-contracted conformation. Starting with 8,212 movies, a set of 1,209,249 particles was selected using Topaz and a few rounds of 2D classification. The selected particles were then used to generate a 3D initial model, which served as a reference to start the 3D classification. After four rounds of 3D classification, the two conformations were clearly distinguishable from each other. After a few other refinement steps, the final maps of the expanded conformation and super-contracted conformation had a nominal resolution of 3.2 Å from 161,531 particles and 3.4 Å from 107,970 particles, respectively. The CryoSPARC processing was performed by Dr. Michael Pilsl, and the procedure is described in *Tarau et al. 2024*[131]. These results have allowed the reconstruction of three conformations (Appendix Table 13) of the *Pfu* RNAP as illustrated in Figure 10.

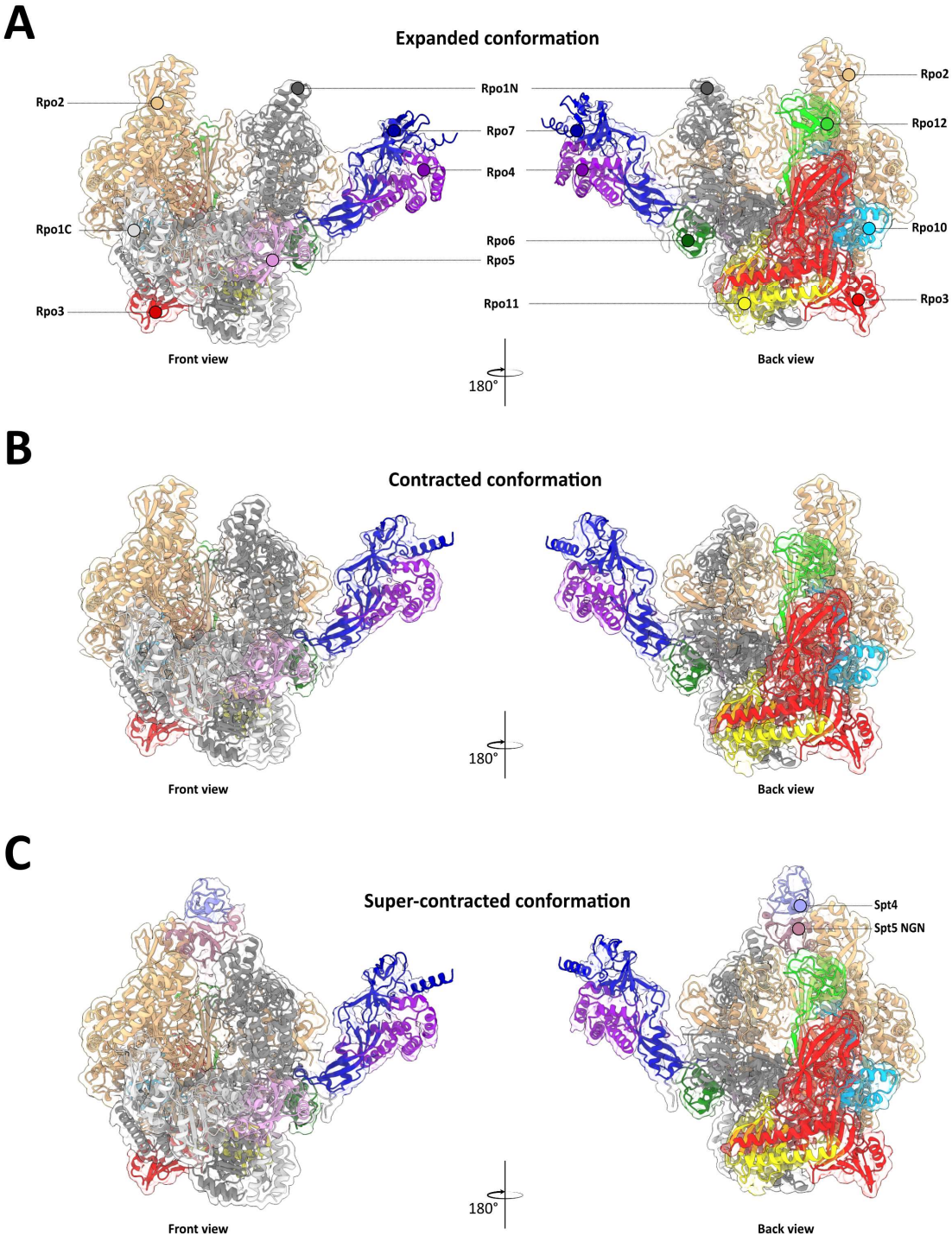


Figure 10. Structure of the euryarchaeal apo RNA polymerase and binary complex with Spt4/5 from *Pyrococcus furiosus*. A. Atomic model in ribbon representation fitted in the corresponding refined density map of *Pfu* RNAP in its expanded conformation in two different views with subunit denomination. **B.** Structural model in ribbon representation fitted in the corresponding refined density map or the *Pfu* RNAP in its contracted conformation in two different views. **C.** Structural model in ribbon representation fitted in its corresponding density map of *Pfu* RNAP and Spt4/5 binary complex in a super-contracted conformation, seen in two different orientations

The expanded conformation matches the crystal structure of apo RNA polymerase from *Themococcus kodakarensis*[50]. In this state, *Pfu* RNAP shows a wide-open DNA-binding channel, characterized by substantial movements of the clamp and stalk domains relative to protrusion, lobe, and jaw domains. The distance between the two edges of the channel, measured between the protrusion domain and the clamp coiled-coil motif, is 54.3 Å, as calculated using UCSF ChimeraX[63] distance analysis (Figure 11A). Two amino acid residues from Rpo2, forming the protrusion domain and Rpo1N from the coiled-coil motif in the clamp domain, were selected for distance measurements: Rpo2-R^{377} and Rpo1N-G^{259}, respectively. The overall map quality resulted in an optimal amino acid assignment (Figure 11B) for the most static parts of the polymerase. On the other hand, mobile elements, such as stalk and clamp domains, resulted in a lower local resolution, making amino acid assignment challenging (Appendix Figure 38).

To overcome this issue, an additional map was obtained with a focus on the stalk domain, allowing for the rigid fitting of the main chains into it (Figure 11C). This was possible by using a feature called "Multi-body Refinement"[132] in the RELION software, which can refine separately two or more parts of a specimen. By creating a mask around the stalk domain and one around the rest of the *Pfu* RNAP, a roughly 4 Å resolution map was obtained. Despite the rigid fitting, the structural model was refined in PHENIX over the full map from the last 3D Refinement job (Appendix Figure 36).

Another notable feature of the expanded conformation is the overly flexible Rpo2 loop spanning from position 1004 to 1025, which in eukaryotic RNA polymerase II is denominated as the switch three loop[57]. This functional loop is a crucial element in transcription elongation as it contacts the DNA template strand in the elongating state. The switch three loop stabilises in the closed-clamp conformations. Clamp domain opening enables the formation of a weak interface between clamp domain subunit Rpo2 and stalk domain subunit Rpo7 (Figure 11D/E), which is not observed in the contracted conformations 3D model structures. The clamp opening seems to be facilitated by the stalk domain moving in the opposite direction with respect to the clamp module.

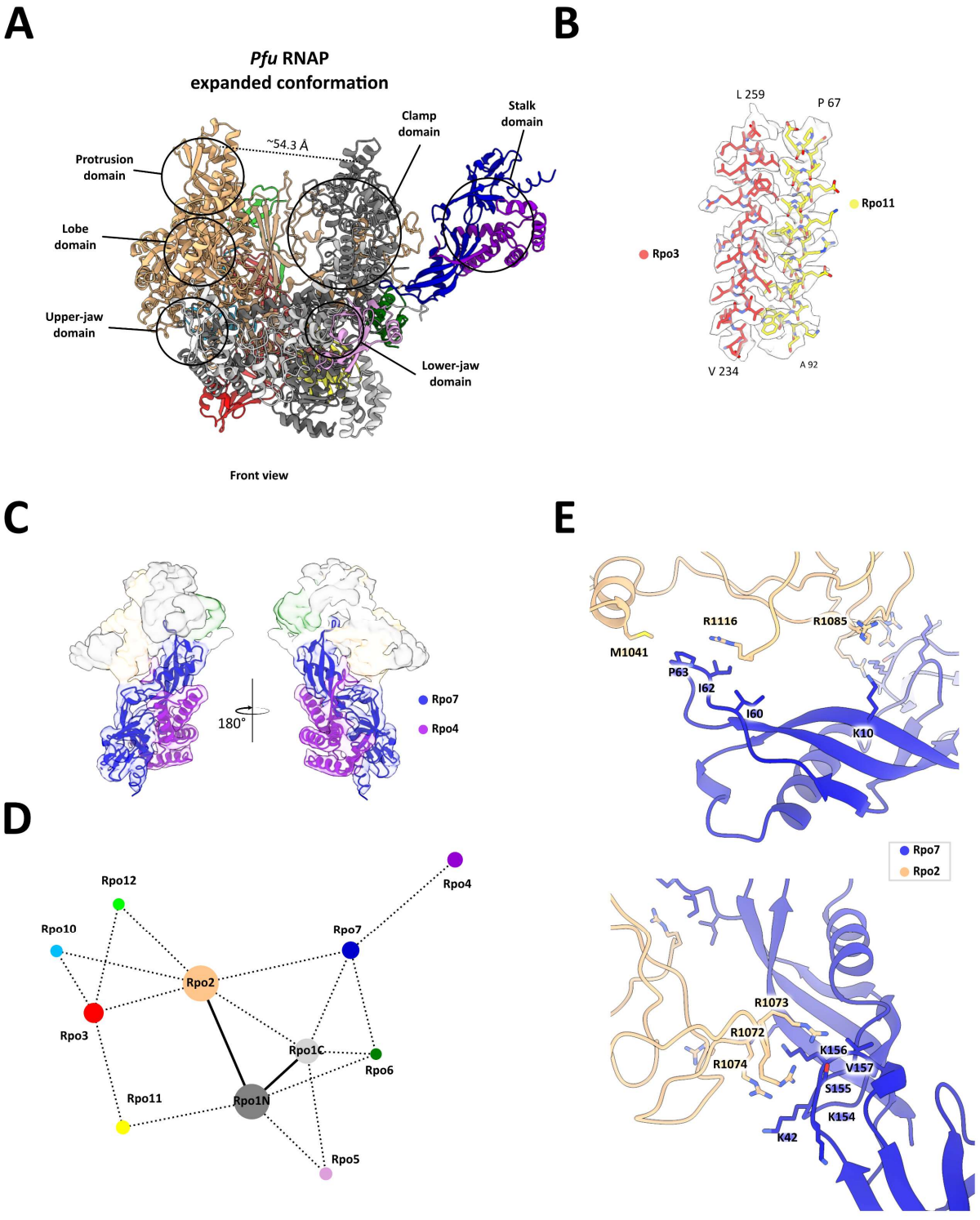


Figure 11. *Pfu* RNA in the expanded conformation establishes an interface between clamp domain and stalk domain. A. Structural model of the *Pfu* RNAP open clamp conformation with domain denomination. The distance of ~54.3 Å spanning between protrusion domain and clamp coiled-coil motif was calculated in ChimeraX using the α-carbon of the amino acid residues Rpo2-R^{377} and Rpo1N-G^{259}. **B.** Details of the atomic model corresponding to the *Pfu* RNAP expanded conformation. Alpha helices corresponding to Rpo3 (V^{234}-L^{259}) and Rpo11 (P^{67}-A^{92}) heterodimer with the respective sharpened density. **C.** Stalk domain (Rpo7/Rpo4) of *Pfu* RNAP fitted in the Multi-body Refinement map. **D.** Subunit interface analysis network performed in ChimeraX of the *Pfu* RNAP expanded conformation. **D.** Amino acid residues of the two regions involved in the interface network between clamp domain (Rpo2) and stalk domain (Rpo7) established in the expanded conformation.

After data processing in cryoSPARC[133], performed by Dr. Michael Pilsl, an additional map was reconstructed. This conformation represents the contracted apo RNA polymerase, which has also been observed for other archaeal RNA polymerase structures[44–47]. The analysis in cryoSPARC also shed light on the continuous conformational changes from the contracted to expanded conformation via 3D variability analysis[134].

The last conformation solved in this dataset was the so-called super-contracted conformation. Here, *Pfu* RNAP is associated with transcription elongation factor Spt4/5, but not with nucleic acids —a configuration not previously described, as Spt4/5 has been described to associate only with RNAP during the elongation phase of transcription. Spt4/5 sealed the DNA entry channel by a web of interactions involving the clamp, protrusion, and lobe domain (Figure 10C and 12A). Distance analysis confirmed that the clamp domain closes further compared to the other contracted conformations of the apo *Pfu* RNAP and TECs (Figure 12B). An indication of the absence of nucleic acids in the catalytic centre is also provided by the fact that Spt4/5 assumes a position too deep in the polymerase DNA entry channel, which would clash with a DNA double helix.

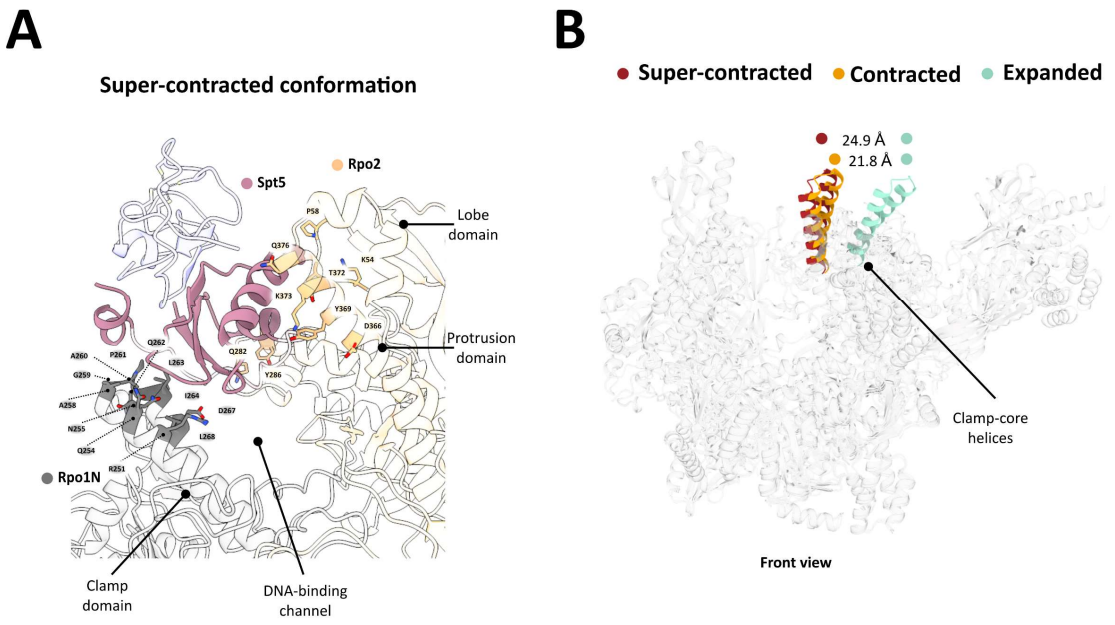


Figure 12. *Pfu* RNAP super-contracted conformation and comparison of the position of all clamp conformations determined for the apo states. A. Adjusted back view of the Spt4/5 and RNAP clamp, lobe, and protrusion domain. The schematic representation highlights the amino acid residues involved it the interaction network of Spt5 and the three domains, sealing off the DNA-binding channel. **B.** Direct comparison of the three apo enzyme conformations globally aligned in ChimeraX. Close-up of the clamp coiled coil helices that undergo a movement of 21.4 Å upon transition from the open to closed state and additional ~3 Å when Spt4/5 is bound to the complex. Contraction distance was measured between the Cα positions of Proline 261 within subunit Rpo1N in the three models.

2.4 *Pyrococcus furiosus* transcription elongation complex structural characterisation

The structural characterisation of the archaeal transcription elongation complex represents one of the primary objectives of this study. This work reports, for the first time in the archaeal domain, the atomic-resolution structure of this conformational state, representing a substantial advance in understanding the archaeal transcription mechanisms. Here, the atomic structure of *Pfu* transcription elongation complex was obtained by following standard strategies for structural determination of protein-nucleic acid complexes. The procedure followed a step-by-step approach, starting with negative staining analysis to determine the optimal ratio between concentration and particle distribution. Afterwards, a screening cryo-EM data collection was acquired to test different grid types. Finally, the RNAP-nucleic acids complex was optimised, and a high-resolution cryo-EM dataset yielded an atomic resolution map.

The Transcription Elongation Complex (TEC) was first assembled with the long scaffold (Paragraph 3.2.7). The mixture contained 50 µg of *Pfu* RNAP incubated with the synthetic scaffold, and size exclusion chromatography was performed to eliminate the excess of nucleic acids (Figure 13A). The elution profile in Figure 13A suggests that the RNAP-scaffold complex eluted in the first peak, as indicated by the nearly equal heights of the absorbance at wavelengths of 260 nm and 280 nm. Additionally, based on SDS-PAGE analysis (Figure 13B), the polymerase fractions also eluted in the first peak. In contrast, the second peak does not display any protein band, suggesting that it mainly contains nucleic acid residues. Hence, the most concentrated fraction was selected for further analysis through negative staining. A small dataset of 161 micrographs was acquired using SerialEM[129] at 40,000x magnification and with a pixel size of 2.7 Å. The micrographs were imported and processed in RELION 3.1[130,135]. First, a CTF estimation job using GCTF[136] function was performed, and particles were picked using the Laplacian-of-Gaussian auto picker from RELION. The particles were then extracted with a box size of 216 Å, and 2D classification was performed using a mask diameter of 200 Å. A subset of 192,171 particles was selected for initial 3D model generation. To investigate the sample heterogeneity, 3D classification was run. The final three classes exhibited similar features: an increased density in the DNA binding channel and a contracted conformation of the RNA polymerase. Besides the monomeric complex composed of the RNAP and elongation complex, three 2D classes corresponding to 8,654 particles also showed the presence of dimeric complexes (Figure 13C). A 3D initial model was performed on the dimeric conformation; however, map interpretation was hindered by preferential orientation (Figure 13C).

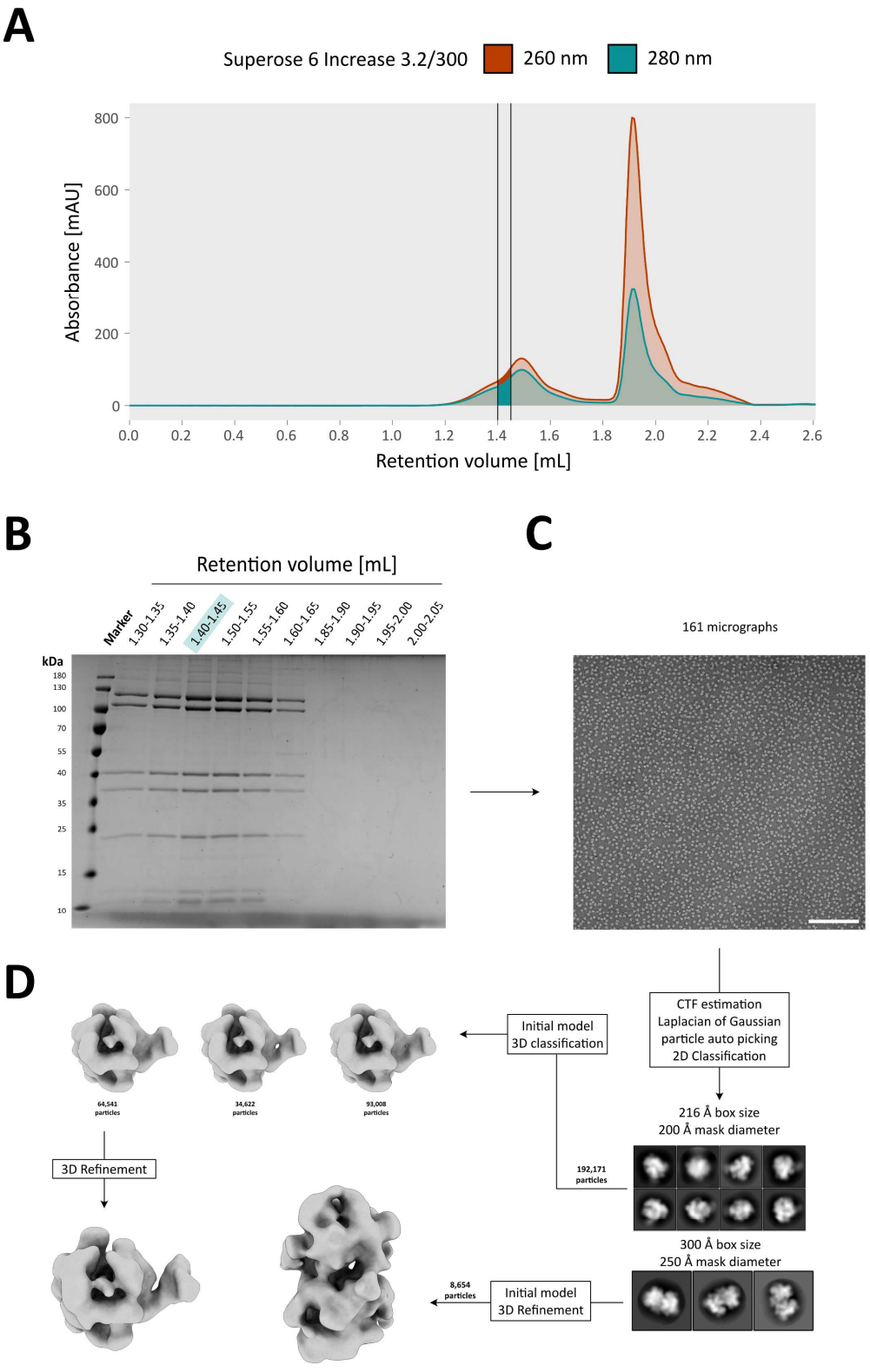


Figure 13. Negative staining workflow of the *Pfu* transcription elongation complex. A. Chromatogram of the size exclusion chromatography run of *Pfu* TEC shows a homogeneous elution with absorbance at 260 nm (orange) and at 280 nm (petrol). **B.** 4%-20% SDS-PAGE with the fractions corresponding to the two peaks of the chromatography run. Notably the first peak presents all *Pfu* RNAP subunits, whereas the second peak represents the excess of nucleic acids as there is no protein content detected by Coomassie staining. **C.** Representative negative staining micrograph of the highlighted fraction from **A.** and **B.** (scale bar 500 µm). The specimen distribution and concentration appear optimal. **D.** RELION processing tree of the negative staining dataset of 161 micrographs. Starting from right-to-left, particle CTF Estimation, Particle Picking and 2D classification were performed. In a first set, the monomeric form of the TEC was selected and processed resulting in a single conformation. Three 2D classes presenting dimeric features were also used for generating a 3D map.

Following the positive negative staining results, the TEC was assembled again for cryo-EM screening. The scaffold with long DNA and short RNA was assembled using 100 µg of *Pfu* RNAP. The mixture was then run on a Superose 6 Increase 3.2/300 GL column to remove excess nucleic acids and to replace the buffer with a glycerol-free buffer solution, which is more suitable for cryo-EM. The chromatogram (Figure 14A) exhibits a broad first peak, which corresponds to the polymerase-nucleic acids complex, as confirmed by SDS-PAGE examination (Figure 14B). The three most concentrated fractions, according to Nanodrop absorbance measurements of 0.51 mg/mL, were collected and mixed.

From the cryo-EM dataset, a 3.7 Å density map was obtained, formed by 177,459 particles. The dataset was recently processed in the newest version of RELION 5.0, where a new feature, DynaMight[137] was introduced in its pipeline. The entire processing scheme, along with a detailed description of sample preparation, is reported in Appendix Chapter 5.2. The DynaMight tool generates a series of 3D maps reflecting the flexible heterogeneity of the refined particles. The DynaMight tool allowed the investigation of the molecular motion of the *Pfu* RNAP at the level of the 3D density map (Figure 14C). This analysis has also revealed a high motion of the upstream DNA duplex, which could explain why the final map displays such a weak density for this part of the scaffold. The clamp and stalk domains are also motile, and they appear to interact with each other as in the expanded conformation of the apo form. Nevertheless, the clamp opening from the DynaMight analysis is not as drastic as in the expanded conformation of the apo enzyme (Figure 14C-petrol).

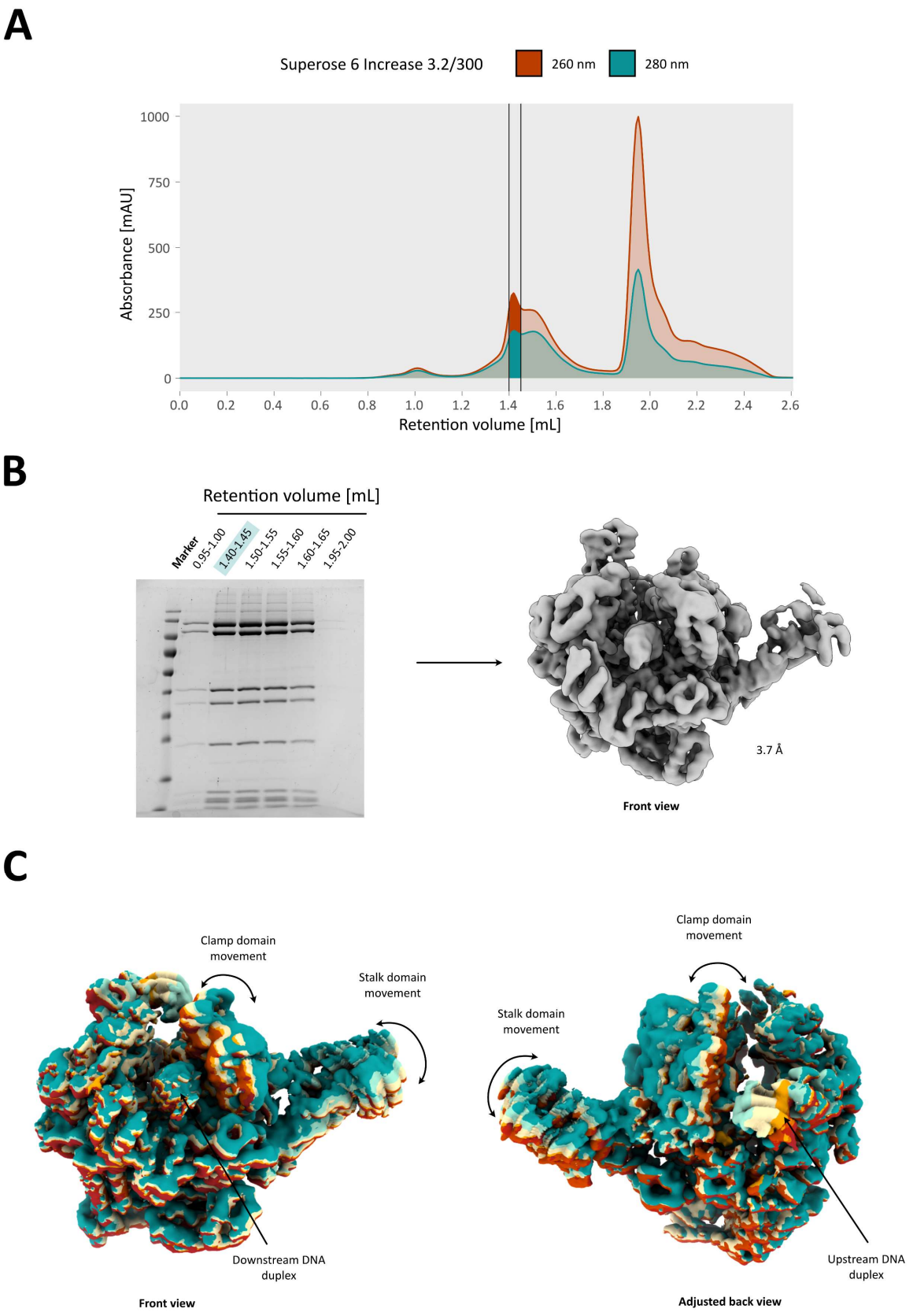


Figure 14. Sample preparation and cryo-EM data collection of the screening dataset. A. Chromatogram of the size exclusion chromatography run of *Pfu* TEC shows a double peak elution with higher absorbance at 260 nm (orange) than 280 nm (petrol) in the first peak corresponding to the polymerase. **B.** 4%-20% SDS-PAGE with the fractions corresponding to the two peaks of the chromatography run. Notably, the double peak presents all *Pfu* RNAP subunits, whereas the second peak represents the excess of nucleic acids, as there is no signal detectable after Coomassie staining. The marked fraction was used for the cryo-EM screening data

collection, and the final 3D density map is shown on the right. **C.** DynaMight analysis of the screening dataset. Movements of the clamp and stalk domains were notable, together with the upstream DNA being the most dynamic part of the map.

Since the dimeric form of the polymerase persisted after cryo-EM dataset processing, the annealing parameters and scaffold design were investigated further. Initially, Electrophoretic Mobility Shift Assays (EMSAs) were carried out to determine whether dimer formation was temperature-dependent. The original scaffold was tested by using an RNA oligonucleotide labelled with a Cy5 dye at the 5' end. The conducted EMSA run on the original scaffold (Figure 15A) reveals two shifts when the *Pfu* RNAP is assembled onto the scaffold with respect to the nucleic acids alone in the first two lanes. In addition, the results of the EMSA (Figure 16) suggest that dimer formation is likely to be temperature independent. Afterwards, scaffold length was tested. A shorter DNA template strand and non-template strand (Chapter 3.2.8, Table 12) together with labelled RNA oligonucleotide were incubated with the polymerase at different temperatures. This considerably reduced the formation of polymerase dimers, as evidenced by the observation of mostly one strong shift (Figure 15B).

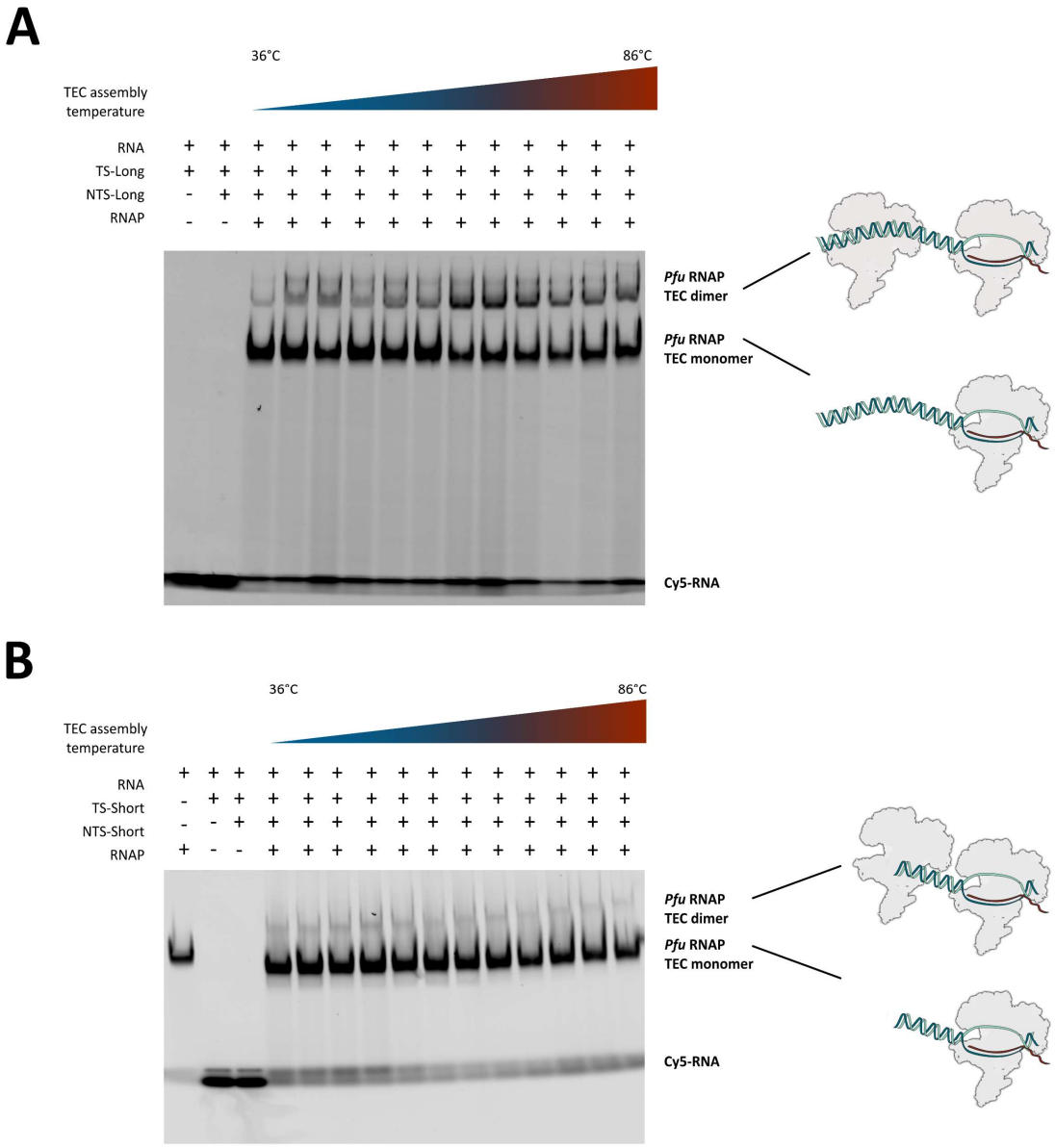


Figure 15. ***In vitro*** **analysis of** ***Pfu*** **RNAP association with different synthetic transcription scaffolds. A.** Electrophoretic Mobility Shift Assay (EMSA) of the RNA polymerase associated with the original scaffold used in *Mj* RNAP studies[49,66]. The loading scheme includes DNA-RNA (first lane) and full scaffold alone (second lane), and the TEC assembled at a gradient temperature ranging from 36°C to 86°C (lanes 3 to 12). This assay exhibits two shifts, one for a dimeric TEC and one for a monomeric TEC. **B.** Same assay as in **A.,** using shorter DNA oligonucleotides. In this case, an additional control of *Pfu* RNAP and RNA incubated at 80°C shows a shift (first lane). The *Pfu* TEC samples exhibit predominantly a defined shift corresponding to the monomeric form.

Another sample was prepared for cryo-EM analysis as described before. This time, the partially complementary RNA and short DNA template and non-template strands were associated with *Pfu* RNAP, instead. Notably, the chromatogram in Figure 16 shows a sharp and narrow first peak representing the most concentrated fraction of 0.95 mg/mL, which is almost double the value compared to the previous long scaffold. The fraction was collected and plunge frozen onto freshly glow-discharged gold grids. The dataset acquisition parameters, processing scheme and model building are discussed in detail in Appendix Chapter 5.2.

Pfu TEC atomic structure provides insights into the archaeal RNA synthesis process. Its overall architecture resembles the eukaryotic RNA polymerase II complex[30,67]. In this structure, all subunits were present and arranged in a contracted conformation (Figure 16C). The stalk domain, formed by subunits Rpo7/Rpo4, and upper jaw from the Rpo1C subunit, represents the most mobile features, as also suggested by the low local resolution (Appendix Figure 42).

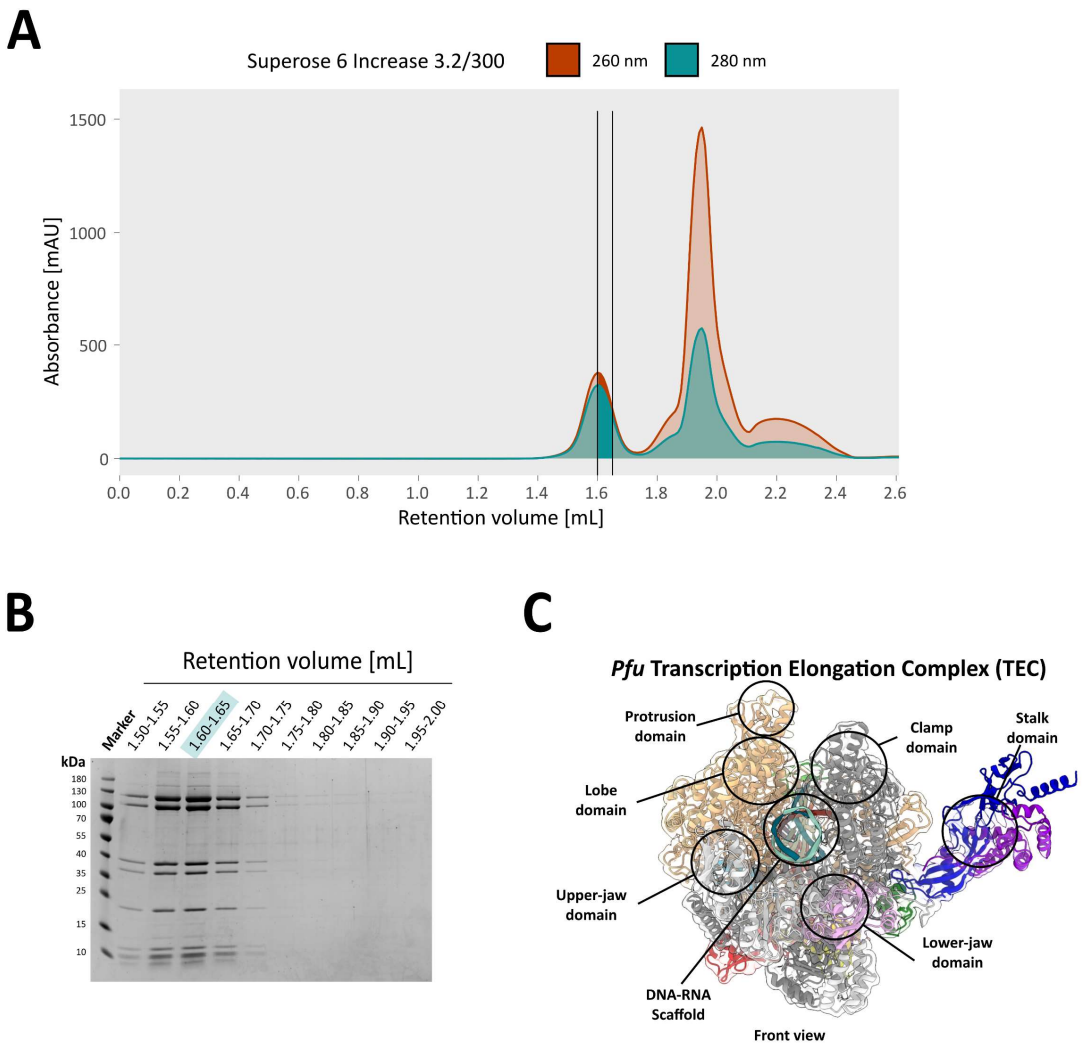


Figure 16. Sample preparation workflow for high-resolution cryo-EM data collection. A. Chromatogram of the size exclusion chromatography shows a single sharp peak corresponding to the *Pfu* TEC. The presence of the scaffold is also suggested by the higher absorbance at 260 nm than at 280 nm. **B.** 4%-20% gradient SDS-PAGE confirming the presence of the Pfu RNAP in the first peak from **A.** and the excess of scaffold in the second peak. The fraction used for further analysis is highlighted in petrol. **C.** Structure of *Pfu* TEC fitted into its refined density map with domain denomination.

The DNA-binding channel was occupied by an additional density, which allowed for the partial reconstruction of the path of the synthetic DNA-RNA elongation scaffold. Indeed, the reconstruction of the downstream DNA duplex and DNA-RNA hybrid was possible. Nevertheless, the assignment of the upstream DNA double helix and nascent RNA transcript sequence in 3D space was hindered by the low-resolution density, possibly due to high flexibility. The same is true for the NTS in correspondence with the transcription bubble. Out of the 35 nucleotides for TS only 20 were possible to be assigned from position 2 to 21 and 10 nucleotides for the NTS from position 25 to 34 considering 5' to 3' orientation numeration. As for the 19 rNTP-long RNA, only the 9-mer forming the DNA-RNA hybrid

from position 19 to 11 was modelled. The nucleotides that were built in the final 3D model are represented as full colour in Figure 17B, whereas the rest of the scaffold is coloured in grey.

The partial reconstruction of the synthetic scaffold has enabled analysis of protein-nucleic acid interactions. Most interactions are of an electrostatic nature between the *Pfu* RNAP DNA-binding channel's positive interface and nucleic acids' negative backbone, with few exceptions. The bridge helix amino acid residues interact with the sugar and base of the nucleotide in position -1 from the TS, as shown in Figure 17C. The bridge helix seems to hold the TS chain from shifting to the next nucleotide, and it physically separates positions -1 and -2 in the DNA TS and adopts a fully α-helix fold, also defined as "unstretched state". The RNA-RNAP interactions principally involve the first ribonucleotide of the catalytic centre (Figure 17A). The NTS interactions with the RNAP are limited to the upstream DNA duplex, and its path cannot be thoroughly followed due to the high flexibility of the DNA. Interestingly, the Q447 residue from the Rpo2 subunit positioned in the fork loop seems to be involved in transcription bubble formation and maintenance by specifically interacting with the base of the NTS nucleotide in position -1 in both TEC complexes investigated (Figure 17 and Figure 22). *Pfu* TEC is found in a post-translocated state as indicated by the interaction between the catalytic magnesium (Metal A) and the ribonucleotide in position +1. Unlike the eukaryotic counterpart, RNA polymerase II, the upper and lower jaw domains of *Pfu* RNAP, formed by subunits Rpo1C and Rpo5, respectively, do not retain any contact with the exiting DNA duplex in the high-resolution structure solved in this study.

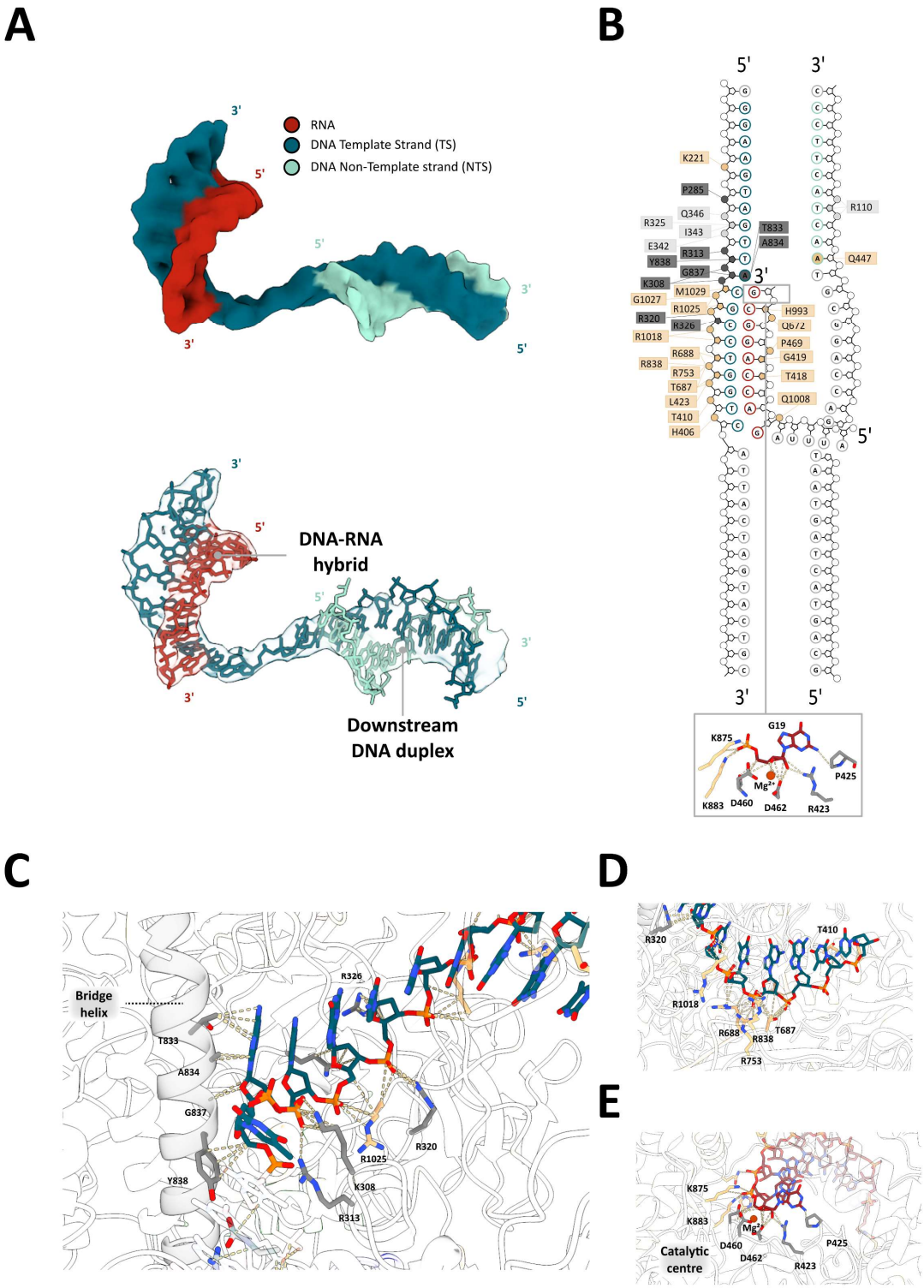


Figure 17. DNA-RNA scaffold density and interaction network of the nucleic acids with the *Pfu* RNAP. A. 3D refined density map (top) of the nucleic acid scaffold components of the TEC (coloured as follows: TS-petrol, NTS-light green, and RNA-red) and the partial structure model of scaffold fitted in the electron density. **B.** Schematic representation of *Pfu* RNAP and nucleic acids contacts in the TEC (calculated using ChimeraX). The colours of the amino acid boxes correspond to the subunit colour-coding which was used in Figure 9. Full colour indicates high-confidence assignments within the 3D maps, whereas grey-coloured areas represent parts of the scaffold that could not clearly be assigned from the reconstruction. **C.** Close-up of the DNA template strand and the interaction network established with the *Pfu* RNAP catalytic centre. **D.** Close-up of the DNA template strand and *Pfu* RNAP interaction network in correspondence with the DNA-RNA hybrid formation. **E.** *Pfu* RNAP catalytic centre and RNA oligonucleotide interaction pattern.

2.5 *Pyrococcus furiosus* transcription elongation complex with Spt4/5

To elucidate the influence of transcription elongation factor Spt4/5 on the TEC, TEC was formed in the presence of Spt4/5 and subjected first to negative staining. The *Pfu* RNAP was in first place when incubated with the synthetic elongation scaffold (TS-long, NTS-long, and T40-short; Table 12). Then, the Spt4/5 heterodimer was added to the TEC and incubated for an additional 15 minutes at 65°C. Since the DNA-RNA scaffold and Spt4/5 were added in excess with respect to the *Pfu* RNAP, a size exclusion chromatography run was performed to separate the TEC from free nucleic acids and Spt4/5 (Figure 18A). To confirm the presence of the TEC in the first peak of the chromatogram, an SDS-PAGE with a small amount of the fractions corresponding to the two peaks was performed (Figure 18A). The highlighted fraction was then blotted onto a negative staining grid as described in Chapter 3.2.14, and a small dataset of 116 micrographs was acquired and processed. The outcome of the dataset was three conformations, two of which contained an extra density on the clamp domain coiled-coil motif, which is where Spt4/5 was detected in structural and biochemical studies (Figure 18B).

Following successful negative staining screening, the sample was scaled up for acquisition of a screening cryo-EM dataset. As evident from the chromatogram in Figure 18A, there were two populations in the polymerase peak, similar to those in the TEC sample. Data processing has revealed the presence of dimeric complex formation (see previous chapter), which was visible in 2D classifications. In addition to this result, the final map exhibited a weak density for Spt4/5 (data not shown). Given the latter results, sample preparation was improved by using a shorter TS and NTS; the RNA oligonucleotide assembled was chosen to have a longer exiting sequence and to ensure Spt4/5 occupancy on the polymerase, BS_3 crosslinking was performed.

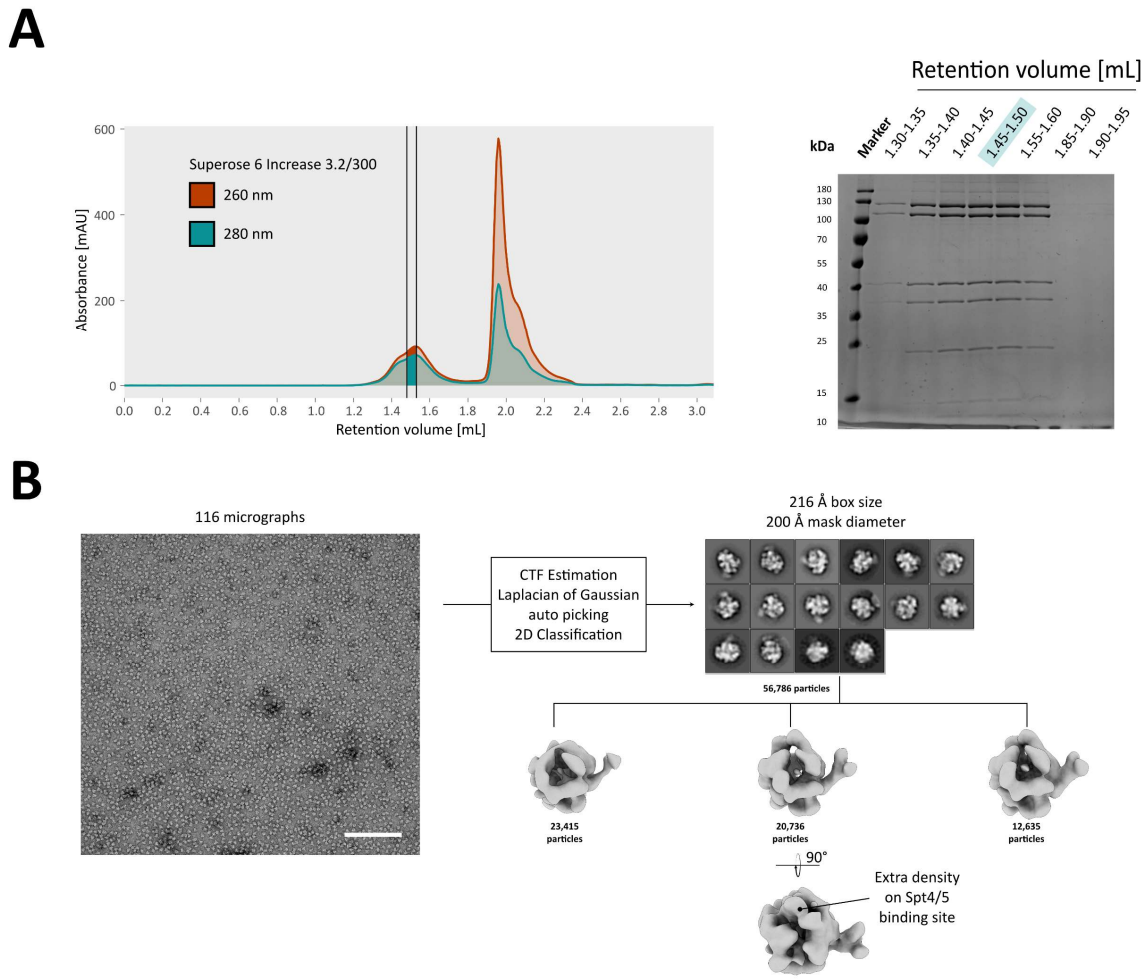


Figure 18. Screening workflow for the *Pfu* TEC with transcription factor Spt4/5. A. Sample preparation for the negative staining data collection and processing, including chromatogram from size exclusion chromatography and corresponding SDS-PAGE using a 4%-20% gradient gel to resolve the protein content of the fractions corresponding to the peak fractions. **B.** Schematic representation of the negative staining data collection and processing in RELION. On the left side, a micrograph is shown (scale bar: 500 µm), and the processing tree to the right side with 2D classification, three classes from which the two out of the three presenting an extra density on the Spt4/5 binding site.

The reason for using a longer RNA in this set of experiments was that in its eukaryotic counterpart, the KOW domains of SPT5 interact with the nascent RNA oligonucleotide, stabilising the interaction between Pol II and SPT4/5[29,30]. For sample implementation, different BS_3 concentrations (Paragraph 3.2.9) were tested to determine the necessary amount to crosslink *Pfu* RNAP and Spt4/5 in the elongation complex (see Table 3). Initially, an apo Pfu RNAP crosslinking titration was carried out with a range of BS3 concentrations (Figure 19A). Following this, the whole complex was assembled with an excess of recombinant Spt4/5, and another titration was carried out with BS_3. Unfortunately, the crosslinking was unsuccessful, likely due to the excessive amount of nucleic acids in the reaction (Figure 19B).

Table 3. The amounts of the different components of the nine reactions of *Pfu* RNA polymerase apo enzyme.

Stock concentration										*Unit*
Reaction number		1	2	3	4	5	6	7	8	9	
Final concentration [BS_3]		0	0.16	0.40	0.80	1.20	1.60	2.00	2.60	3.20	mM
BS_3	50 mM	-	-	-	-	0.24	0.32	0.40	0.52	0.64	µL
BS_3	5 mM	-	0.32	0.80	1.60	-	-	-	-	-	µL
Pfu RNAP	1.6 mg/mL	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	µL
Buffer		9.10	8.78	8.30	7.50	8.86	8.78	8.70	8.58	8.46	µL
Lys/Asp 25°C/20 min.	100 mM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	µL
NH_4HCO_3 25°C/20 min.	500 mM	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50	µL

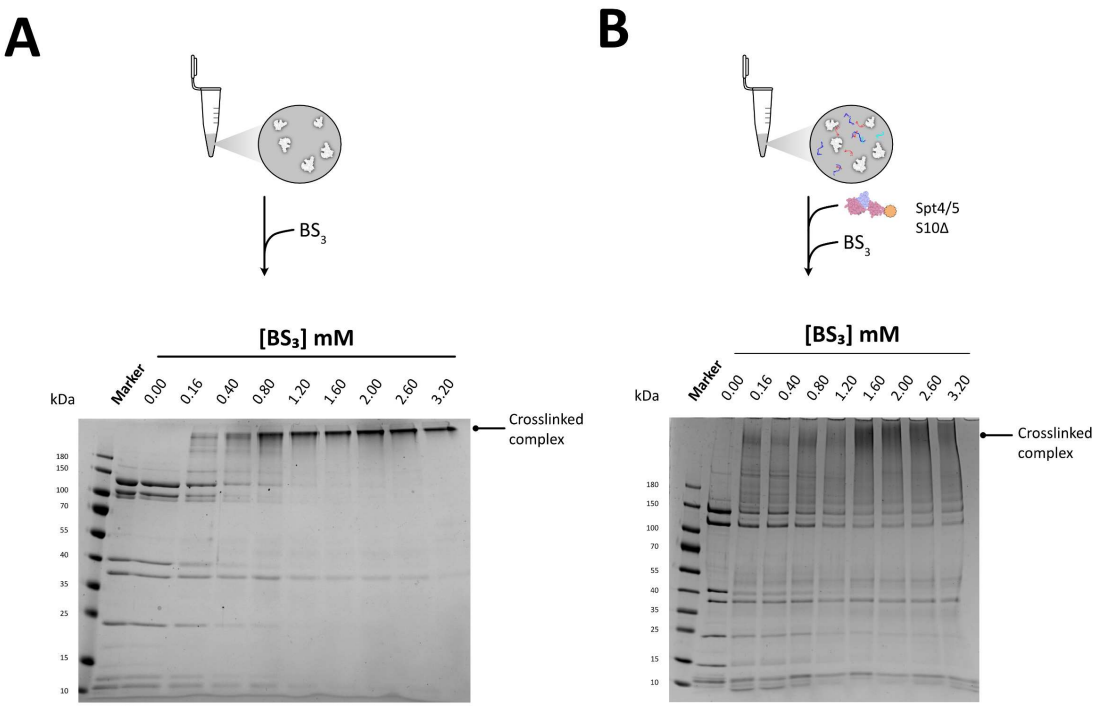


Figure 19. Optimal BS_3 concentration determination for *Pfu* RNAP and *Pfu* TEC through BS_3 titration. A. BS_3 titration of the *Pfu* apo RNAP 4%-20% gradient SDS-PAGE also containing one µg RNA polymerase complex as positive control in the second lane and the reactions in lanes three to twelve. The crosslinked complex forms completely at a final concentration of 1.20 mM BS3 (reaction 5). **B.** BS_3 titration of the *Pfu* TEC in complex with Spt4/5 and S10Δ 4%-20% gradient SDS-PAGE, also containing one µg RNA polymerase complex as positive control in the second lane and the reactions in lanes three to ten. The addition of TEC components inhibits the crosslinking reaction at the same concentration range used for *Pfu* RNAP alone.

To improve chemical crosslinking of the TEC with Spt4/5, the *Pfu* TEC was first assembled, and the excess nucleic acids were eliminated via size exclusion chromatography. At this point, Spt4/5 and S10Δ mutant were added to the reaction, and another titration was performed. S10Δ mutant was provided by collaborators from the University of Bayreuth, Dr. Stefan Knauer and MSc. Tim Engelgeh. S10Δ is part of the ribosomal protein, which is theoretically interacting with the KOW domain of Spt5 transcription elongation factor (see Chapter 2.7). Even though the crosslinking occurred this time, S10Δ most probably failed to associate with the complex, because an intense band at the expected height is present in the second peak of the second chromatogram, as shown in the SDS-PAGE Figure 20C/D. After the titration, 3 mM of BS_3 was used in the upscaled experiment (50 µg). The excess of crosslinker and quencher was eliminated by a second gel filtration run, as shown in Figure 20C. The most concentrated fraction of the second round (Figure 20D) was blotted onto a copper grid, and negative staining was performed. Data processing of the negative staining dataset was separated into three classes, one of which did not exhibit any extra density (Figure 20E). Given that two of the three low-resolution maps presented an extra density on the Spt4/5 binding site on the polymerase, this procedure was repeated for cryo-EM analysis.

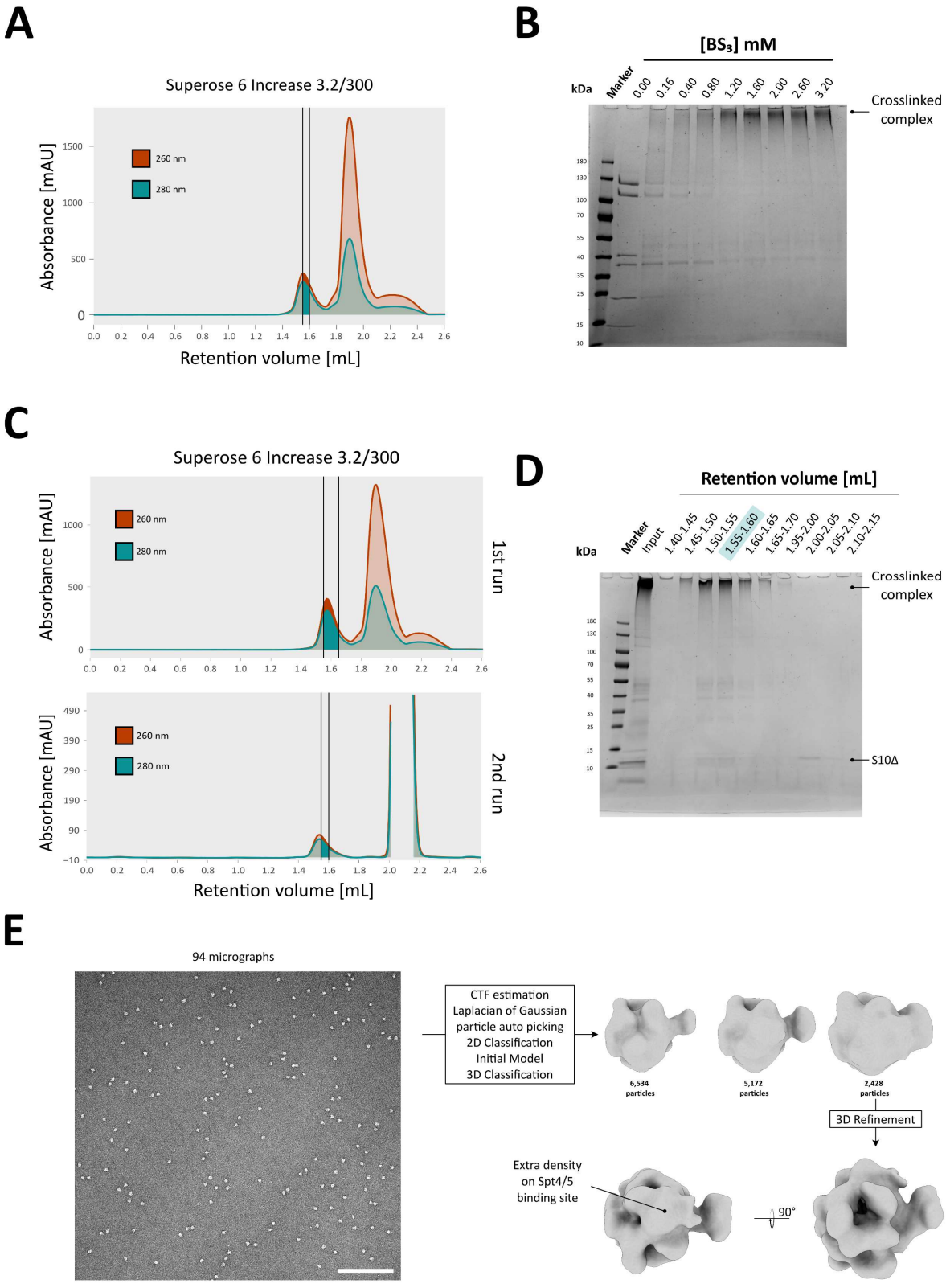


Figure 20. Chemical crosslinking approach for *Pyrococcus furiosus* transcription elongation complex. A. Chromatogram of the size exclusion chromatography run of the *Pfu* TEC with marked fractions used for chemical crosslinking titration. **B.** 4%-20% gradient SDS-PAGE of *Pfu* TEC with increasing concentrations of BS_3. **C.** Size exclusion chromatography of the *Pfu* TEC in the first run, with marked fractions that were used for chemical crosslinking. The second run represents the size exclusion chromatography run, intended to separate the crosslinked complex from excess of BS_3, quencher, and additional factors. **D.** 4%-20% gradient SDS-PAGE containing the fractions from the second run shown in **C.,** with the highlighted fraction used for negative staining. **E.** Negative staining processing scheme (left: a micrograph with a scalebar of 500 nm), and the processing steps from RELION software. The third class was presenting an extra density on the Spt4/5 binding site.

Given the promising results from the negative staining analysis, the same procedure was repeated. The chromatograms in Figure 21A demonstrate that the crosslinking procedure was reproducible, and the 4%-20% SDS-PAGE in Figure 21B also confirmed the success of the crosslinking reaction. The highlighted fraction was plunge-frozen, and a high-resolution cryo-em dataset was collected at the in-house facility (Appendix Chapter 5.3)

The heterogeneity of the latter dataset was highly unusual and noteworthy. A starting set of 6,076 micrographs was selected for performing particle autopicking. After a few rounds of 2D classification, 828,076 particles were selected for 3D initial model generation. After four rounds of 3D classification, a 3.3 Å nominal resolution map was obtained with a weak density for Spt4/5. Two more rounds of focused 3D classification were run, and a final map of 3.4 Å nominal resolution was obtained from 52,527 particles (Appendix Chapter 5.3, Figure 41). The final density map of the 3.4 Å reconstruction has allowed the atomic structure analysis of the complex. As in the *Pfu* TEC, the scaffold and all subunits of the polymerase were present. The overall conformation of the *Pfu* RNAP was contracted with a highly flexible stalk domain (Appendix Chapter 5.3, Figure 42). In this reconstruction (Figure 21C), only the Spt5 NGN domain could be unambiguously fitted into the electron density, and Spt4 was rigidly fitted in the corresponding position due to low local resolution. Unfortunately, the KOW domain remains too flexible to be reconstructed. A visible effect of Spt4/5 on the *Pfu* TEC is the stabilisation of the upstream DNA duplex, as evidenced by the improved density resolution compared to the TEC alone.

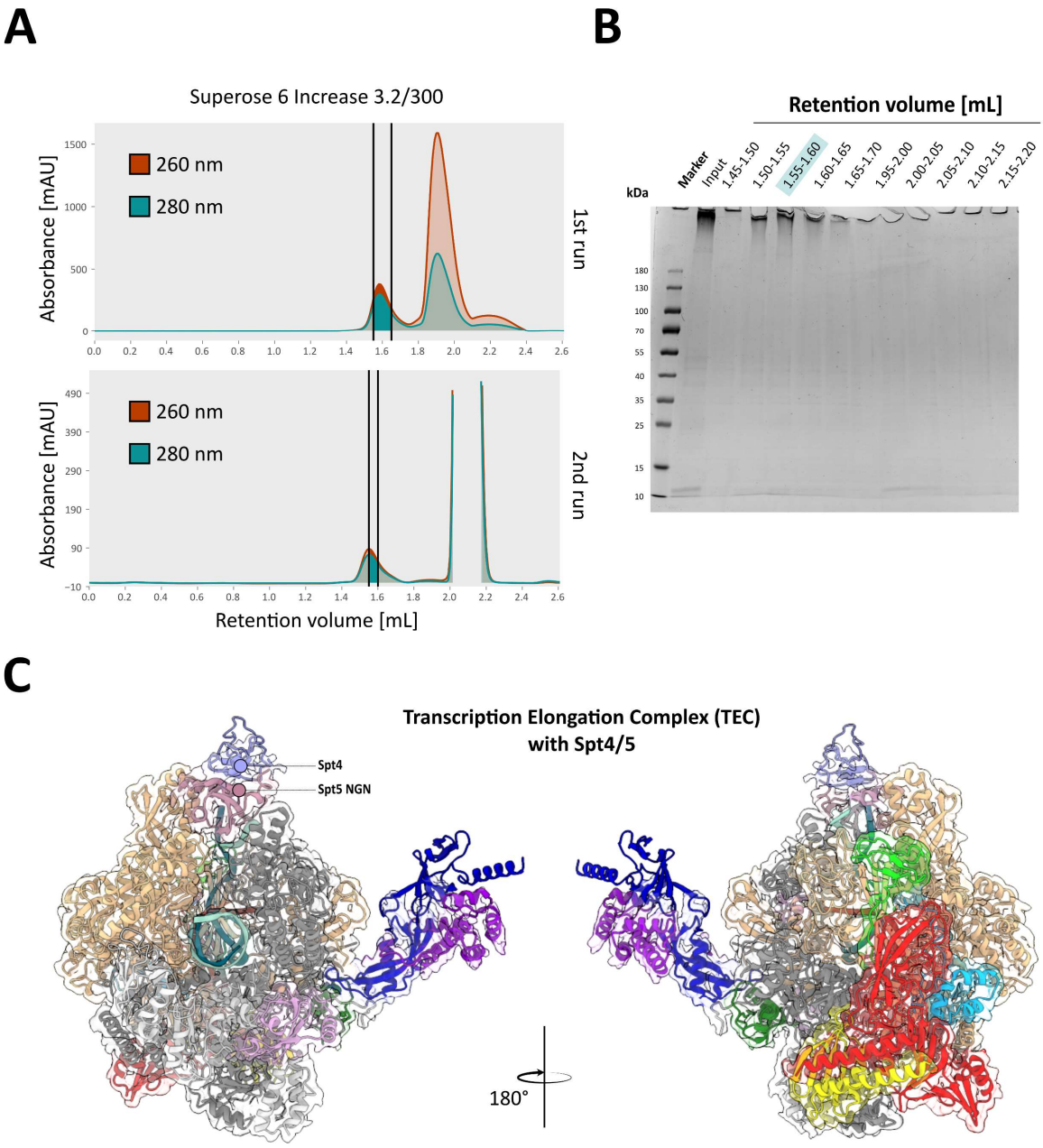


Figure 21. Sample preparation of the high-resolution cryo-EM data collection and final 3D model of *Pfu* transcription elongation complex with Spt4/5. A. Chromatograms of the size exclusion chromatography performed first to eliminate the excess of nucleic acids (upper panel) and after chemical crosslinking to eliminate the excess of proteins, BS_3, and quencher (second chromatogram, lower panel). The highlighted fractions of the first run were used for crosslinking and for cryo-EM sample preparation. **B.** 4%-20% gradient SDS-PAGE showing the elution fractions from the second size exclusion chromatography run in **A.** The fraction used for plunge-freezing are highlighted. **C.** 3D structure of the *Pfu* TEC in complex with Spt4/5 fitted into the corresponding refined map. The Pfu RNAP subunits are coloured with the same scheme as in Figure 10A.

This effect allowed the reconstruction of a longer portion of the transcription bubble and elucidated interaction patterns between *Pfu* RNAP domains and upstream DNA. Additionally, the use of a longer RNA overhang sequence has enabled the identification of a poorly resolved density in the RNA exit channel. Nevertheless, the low resolution of the density hindered the assignment of the whole RNA sequence; therefore, only the 9 nucleotides involved in DNA-RNA hybrid formation were built. Out of the 35 nucleotides of the TS, the first 26 were unambiguously modelled. The NTS strand nucleotides visible in the 3D map reconstruction were from positions 10 to 15 and 25 to 35, with a gap in between, corresponding to the transcription bubble portion of the scaffold. *Pfu* RNAP contact pattern with the

scaffold only shows minor differences when compared to the TEC alone structure (Figure 22B). It mainly involves positively charged amino acids and the negatively charged phosphate backbone of the nucleic acids. The bridge helix is found again in between position -1 and -2 of the DNA TS, and interestingly, lysine 308 from Rpo1N seems to be involved in transcription progression, given that it interacts with phosphates of both -1 and +1 on the template strand. The most notable interaction involving base and amino acids occurs between an adenosine base in position +10 and leucine 228 and isoleucine 232 (Rpo1) as illustrated in Figure 22. This interaction happens in correspondence to the transcription bubble formation on the upstream DNA duplex. The RNA oligonucleotide contacts with the enzyme are focused on position +1 (G19) in the catalytic centre (Figure 22B) and interactions in the RNA exit channel (Figure 22E).

In summary, the first atomic reconstructions of archaeal transcription elongation complexes were obtained. The *Pfu* TEC elongation scaffold was partially resolved, allowing the assignment of the downstream DNA duplex and the DNA–RNA hybrid within the catalytic centre. In contrast, density corresponding to the upstream DNA duplex was progressively lost with decreasing nominal resolution, indicating that although engaged with the polymerase, this region is highly flexible. Association of Spt4/5 with the TEC stabilised the upstream duplex, enabling reconstruction of an additional segment of the synthetic scaffold. This observation provides mechanistic insight into the function of Spt5 within the archaeal TEC. Moreover, inspection of the amino acid sequence proximal to the reconstructed upstream DNA identified four positively charged residues, establishing a structural basis for subsequent structure-to-function analyses, which are described in the following chapter.

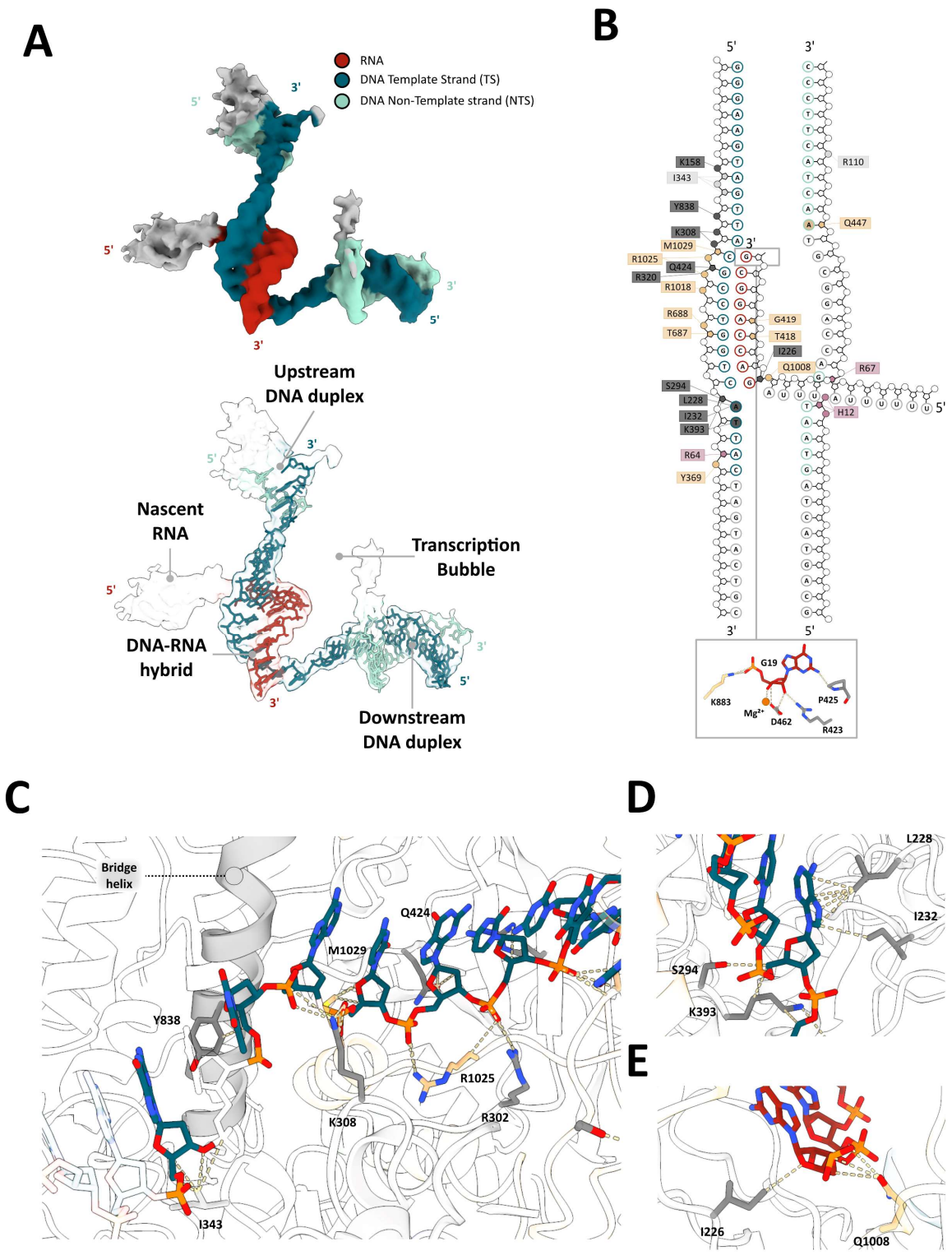


Figure 22. Analysis of the interaction pattern between *Pyrococcus furiosus* RNA polymerase and the DNA-RNA scaffold in the transcription elongation complex with Spt4/5. A. Density of the elongation complex scaffold (top) with Spt4/5coloured according to its components: template strand (petrol), non-template strand (light green) and RNA oligonucleotide (red). In the bottom panel, the structural model of the elongation scaffold in the corresponding refined map is shown, with element and conformation designations. **B.** Schematic representation of *Pfu* RNAP and nucleic acids contacts calculated with ChimeraX of the TEC in the presence of Spt4/5. Full colour indicates a high confidence assignment within the 3D density map, whereas faint colour indicates parts of the scaffold that could not be assigned in the reconstruction. **C.** Detail of the template strand pathway in *Pfu* RNAP with focus on the catalytic centre showing the interaction network between amino acid residues and nucleic acids. **D.** Close up of the DNA template strand and *Pyrococccus furiosus* RNA polymerase interaction network (dark grey amino acid residues corresponding to Rpo1N and beige amino acid residues corresponding to Rpo2) showing the correspondence of the downstream DNA duplex. **E.** Zoom into the RNA exit channel and RNA oligonucleotide interaction pattern.

2.6 Functional investigation of the Spt5 NGN domain interaction with the DNA

Considering the structural findings on the transcription elongation factor Spt4/5 and its effects on the *Pyrococcus furiosus* transcription elongation complex (*Pfu* TEC), a more in-depth analysis was conducted to explore whether the interaction between the DNA and Spt5 might stabilise the DNA. The atomic model of the *Pfu* TEC in complex with Spt4/5 provided critical insights into the spatial positioning of the Spt4/5 main chain. However, due to limitations in local resolution within the cryo-EM data, several side chains were only partially assigned with high confidence. Despite this constraint, it was possible to hypothesise that four basic amino acid residues—Histidine 12, Arginine 64, Histidine 66, and Arginine 67—likely play a key role in interacting with the incoming DNA strand, as illustrated in Figure 23A.

These four residues, positioned in close proximity to the phosphate backbone of the DNA, appear to be particularly relevant at the upstream edge of the transcription bubble, where the DNA duplex begins to unwind. This positioning indicated that these positively charged residues could promote interactions essential for stabilising the transcription bubble. To verify the functional relevance of these residues, an Spt5 mutant variant was generated, in which the four basic residues were replaced with alanine. This mutant was subsequently tested in an *in vitro* transcription assay using pre-formed *Pfu* TEC to evaluate its impact on transcription processivity (Figure 23B). The TEC was assembled under the same conditions and DNA-RNA scaffold as for the high-resolution structural analysis presented in Paragraphs 2.4 and 2.5 (TS-short, NTS-short and RNA-long, Chapter 3.2.8, Table 12).

The experimental results demonstrated that wild-type Spt4/5 significantly reduced transcriptional pausing events and enhanced the processivity of *Pfu* RNA polymerase (RNAP). In contrast, the Spt5 mutant, lacking the positive charge at these key residues, exhibited a diminished ability to stimulate transcription elongation (Figure 23C). Despite this loss of stimulatory effect, the overall stability of the complex and its ability to form an interaction with its Spt4 partner remained unaltered, indicating that the structural integrity of the mutant and the Spt4/5 complex was not compromised.

These structure-to-function analyses highlighted the importance of the two histidine and two arginine residues in supporting the transcriptional activity of Spt4/5. Specifically, the interaction between these positively charged residues and the phosphate backbone of the upstream region of the transcription bubble appears to be crucial for the stimulatory function of Spt4/5 in the *Pfu* transcription elongation phase. This highlights a likely mechanism by which Spt4/5 promotes the efficient progression of the transcription machinery, stabilising the transcription bubble and thereby reducing pausing events.

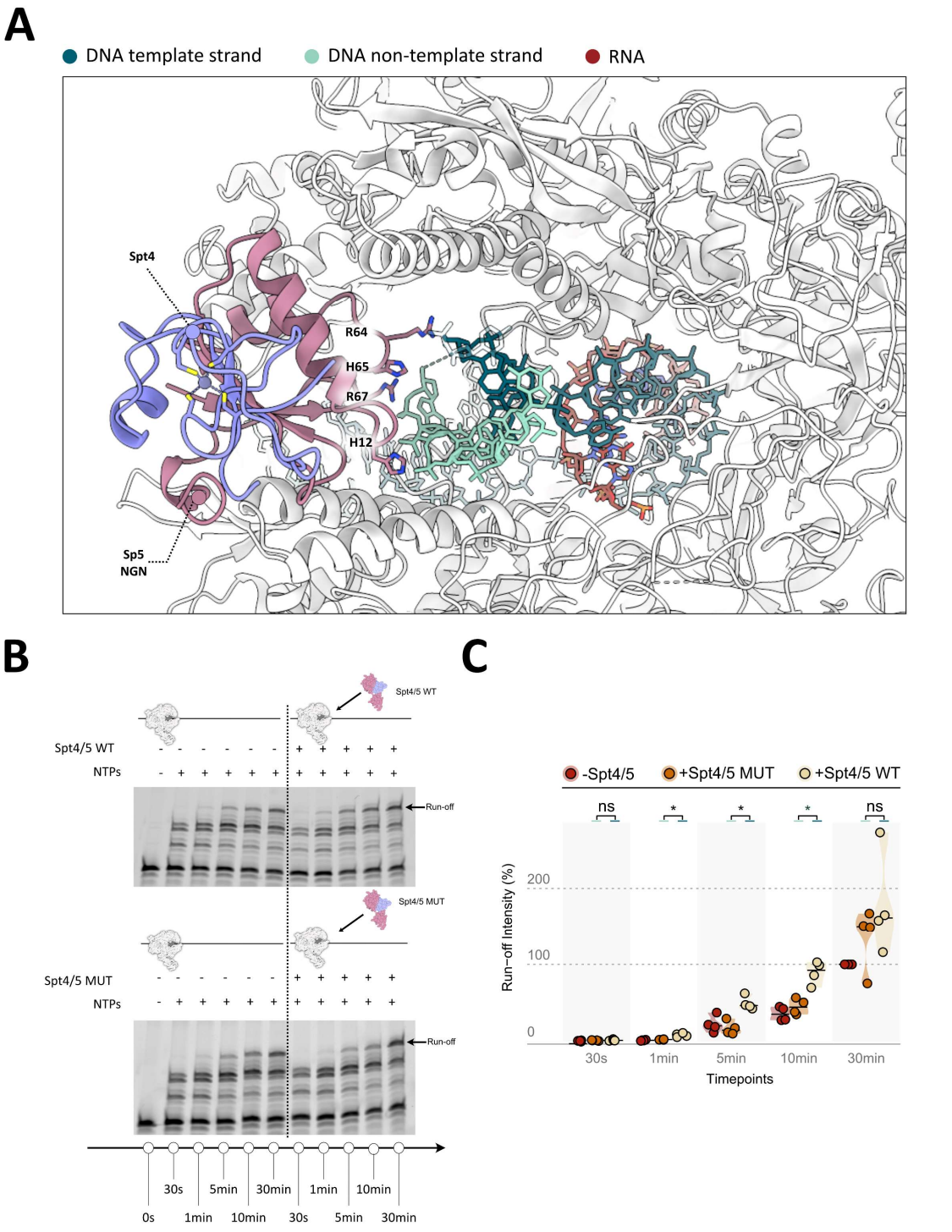


Figure 23. In vitro analysis of the effect of the interface of the Spt5 NGN domain and the DNA on transcription elongation. **A.** Close-up view of the interaction of Spt5 with the phosphate backbone of the non-template strand. Spt5 side chains (H12, R64, H65 and R67) that contact the DNA are indicated. Spt5 is coloured in dusky pink and Spt4 in purple. **B.** Transcription elongation assay using wt *Pfu* RNAP (0.6-0.7 µg) and Spt4/5 wt or the Spt4/5 mutant ($Spt5^{H12A/R64A/H65A/R67A}$, 1.7 µM). As a reference, transcription assays were performed in the absence of Spt4/5. Reactions were stopped at 30 s, 1 min, 5 min, 10 min, and 30 min. **C.** Diagram showing the quantified run-off transcript levels shown in panel **B.** Violin plot illustrating run-off intensity (%) over time (30 s, 1 min, 5 min, 10 min, 30 min) for a reaction without Spt4/5 (dark red), for a reaction with Spt4/5 MUT (orange), and a reaction with Spt4/5 WT (beige) samples. Individual data points are overlaid as jittered points for clarity. Medians displayed with crossbars. Statistical significance indicated by NS ($p > 0.05$) and * ($p < 0.05$) based on Mann-Whitney U tests comparing +Spt4/5 MUT and +Spt4/5 WT at each time point.

2.7 Spt5 and ribosomal protein S10 interaction studies

Archaea are deprived of cellular compartments, unlike Eukaryotes; therefore, it is most likely that the transcription and translation processes occur simultaneously[95]. Indeed, it has been proven for some bacterial species that the nascent RNA synthesised by the RNA polymerase is being promptly loaded onto the ribosome for translation. This phenomenon of a physically coupled RNAP and ribosome has been described in the literature as "expressome"[10] or transcription-translation complex (TTC)[11–13]. Since most of the structural studies have described an NusG- or NusA-coupled complex, the hypothesis was that Spt5 in *P. furiosus* might mediate the coupling of the RNAP and ribosome in archaea. To assess whether the *Pfu* RNAP and ribosome can interact directly, we first examined the interaction between the single proteins Spt5 and S10. The latter is a ribosomal protein situated in the head domain of the 30S particle, and in bacteria, it interacts with the KOW domain of NusG. Based on AlphaFold 2[138] predictions, suggest that the two proteins interact in a homologous manner in Archaea (Figure 24C/D).

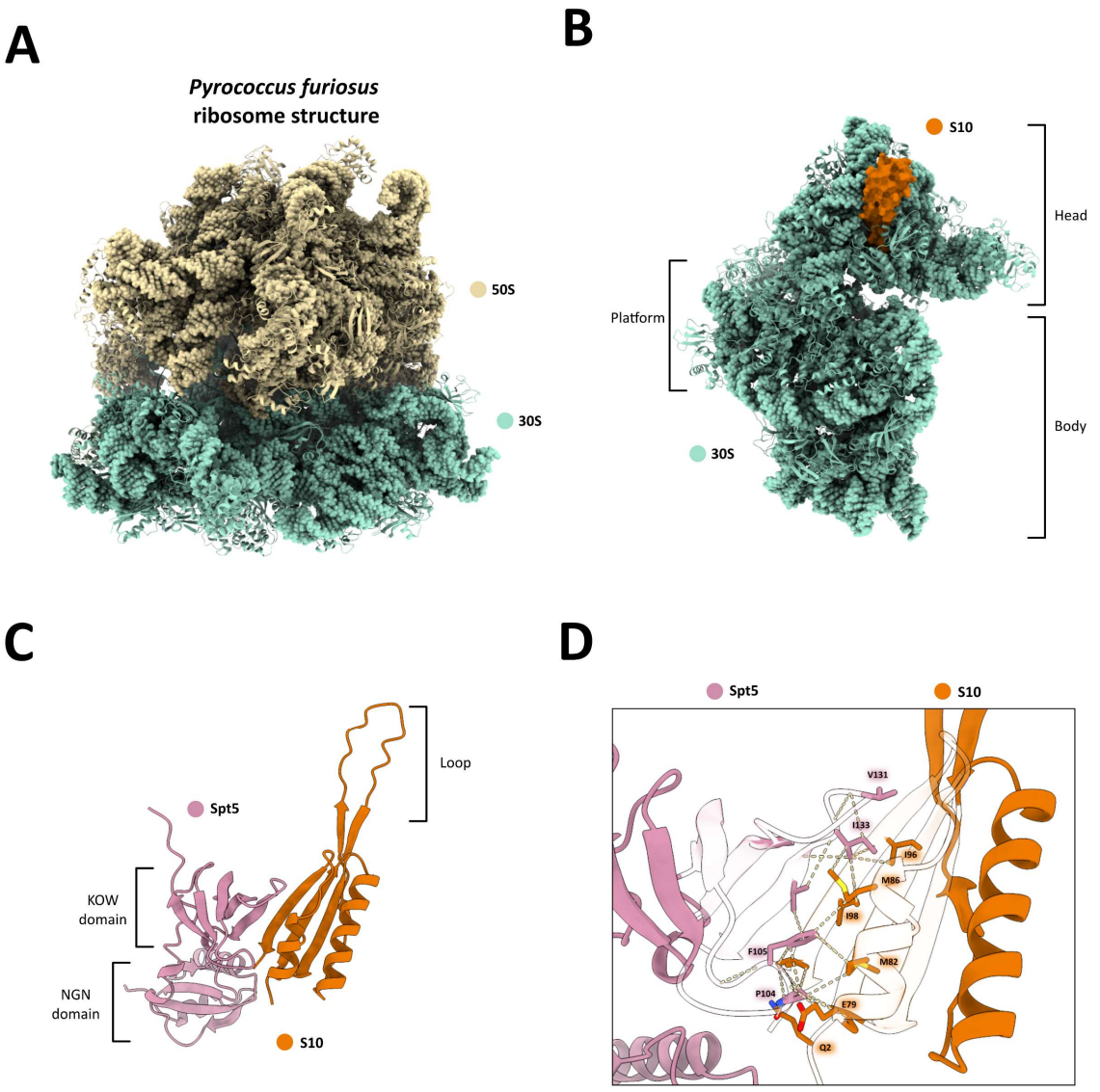


Figure 24. Structural organisation of the ribosome from *Pyrococcus furiosus* and hypothesized interaction pattern between transcription factor Spt5 and ribosomal protein S10. A. Cryo-EM structural model of *Pyrococcus furiosus* ribosome (PDB ID: 4V6U). The large subunit, 50S, is coloured in beige and the small subunit, 30S, is coloured in green. **B.** Structural model of *Pyrococcus furiosus*

small subunit 30S (PDB ID: 5JB3) showing prominent domains and ribosomal protein S10 coloured in orange. **C.** AlphaFold 2 multimer prediction of transcription factor Spt5 (dusty pink) and ribosomal protein S10 (orange). Protein domains are indicated. **C.** Contacts calculation performed in ChimeraX for the Spt5-S10 interaction based on the AlphaFold predicted structural model. Amino acids that might contribute to the interaction are highlighted.

Initially, a cysteine was inserted at the N-terminus of the recombinant protein sequence of S10 and labelled with maleimide dye Alexa647 as described in paragraph 3.2.6. The recombinant protein was expressed and purified as a full-length protein.

Cooperation partners from the University of Bayreuth (MsC. Tim Engelgeh, Dr. Stefan Knauer) have conducted NMR studies on the Spt5 KOW domain and the S10Δ mutant, discovering four amino acid residues on the Spt5 KOW domain involved in the interaction with the ribosomal protein. S10Δ mutant consists of a version of the protein lacking the loop motif (Figure 24C), which confers propensity to dimerization. On the other hand, the four amino acid residues valine 131, isoleucine 133, phenylalanine 105, and proline 104 from Spt5 were individually converted to alanine for further analysis. The point mutations were performed with mutagenesis primers as described in Chapter 3.2.7 and purified as described in Chapter 3.2.5. The generation of single mutants of the four amino acid residues was performed to achieve a better understanding of their interaction pattern with S10 through microscale thermophoresis (conducted by MSc Martin Brehm and in collaboration with Prof. Dr. Gernot Längst).

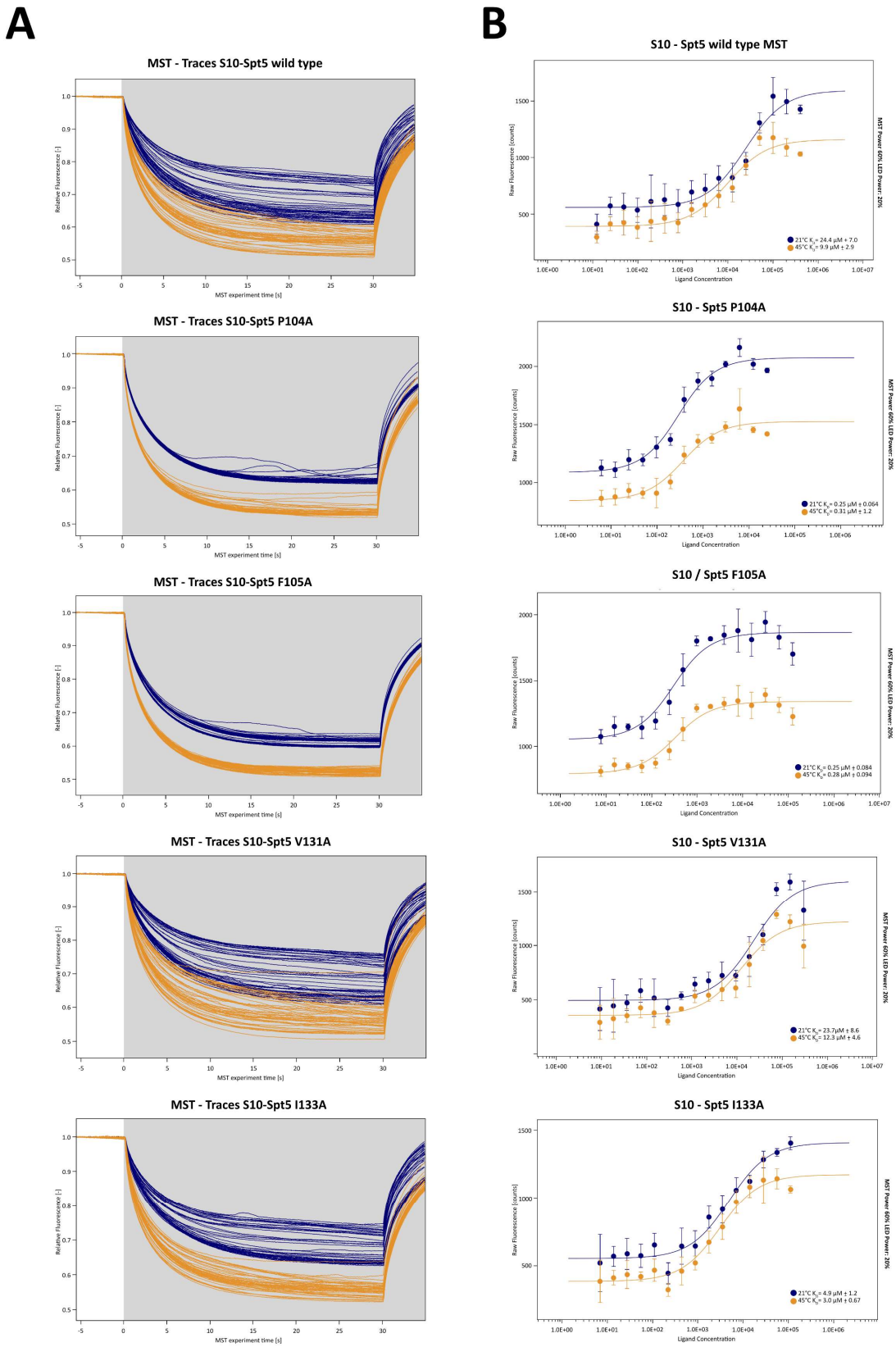


Figure 25. MST data analysis of the binding behaviour between fluorescently labelled S10 and Spt5 variants. A. MST traces curves of ribosomal protein S10 in combination with Spt5 wild type, Spt5 P104A mutant, Spt5 F105A mutant, Spt5 V131A mutant, and Spt5 I133A mutant. The curves represent are coloured according to the measurement temperatures used: 21°C in blue and 45°C in ochre. **B.** Correspondent fluorescent signal of the MST data analysis obtained from three replicates from each sample at the two different temperatures. Final K_D values and K_D confidence are indicated at the bottom right of each graph.

In this experimental setup, full-length S10 was used, which is known to dimerise in solution via the loop motif (Figure 24C). This motif binds to other ribosomal proteins and rRNA residues, remaining protected from the solution when associated with the small ribosomal subunit 30S; therefore, the dimerization does not occur (Figure 24B). For this reason, the reproducibility of MST measurements and data analysis is poor. Indeed, all Spt5 mutants were able to bind fluorescently labelled S10 (Figure 25). Additionally, Spt5 P104, Spt5F105, and Spt5 I133A show a better affinity to S10 than Spt5 wild type. Another drawback of this experiment was that full-length S10 degrades faster. Even though *P. furiosus* optimal growth temperature is 100°C, and recombinantly expressed proteins are still thermostable, the highest temperature to which measurements could be performed was 45°C due to machine limitations.

As the MST measurements remained inconclusive, an alternative approach for analysing the transcription-translation coupling was employed. This time, fresh cultures of different *Pyrococcus furiosus* mutants were grown as described in paragraph 3.2.2. The mutants chosen were MURPf 1, MURPf 4, MURPf 63, MURPf 75, MURPf 50, and MURPf 51. MURPf 1 and MURPf 4 can express the Rpo3 subunit fused to two tags, a STREP tag and a 6xHIS tag. The difference between these two mutants is that MURPf 1 presents the sequence of the tags in its genome, whereas MURPf 4 on a plasmid inserted with the shuttle vector methodology[139]. MURPf 63 corresponds to a *Pfu* mutant strain in which a plasmid containing the subunit Rpo1C fused to a STREP and 6xHIS-tag was inserted. MURPf 50 and MURPf 51 are mutants that contain the same plasmid as MURPf 63, with the difference that they also exhibit two mutations of threonine 82 converted to leucine and arginine, respectively. According to unpublished data conducted by Prof. Winfried Hausner, the T82L mutation leads to a hyperactive *Pfu* RNAP phenotype, whereas the T82R mutation impedes the *Pfu* RNAP from progressing into the elongation phase. Finally, MURPf 75 was used as a negative control in which only the HIS-tag is expressed in *Pfu*. The *Pfu* RNAP was pulled down from each mutant culture by exploiting the 6xHIS tag and using magnetic beads. Four replicates for each mutant were produced to obtain a statistically significant result. The final protein concentration of each replicate was determined using Qubit™ Protein and Protein Broad Range (BR) Assay Kits, and the results are shown in Table 4.

Table 4. Composition of the mass spectrometry sample preparation including *Pyrococcus furiosus* mutant, type of mutation and final concentration of the pulldown.

Replicate	Mutant	Name	Concentration
1	MURPf 1	Rpo3-STREP-HIS HmgCoAR genome	376 ng/µL
2			516 ng/µL
3			362 ng/µL
4			414 ng/µL
1	MURPf 4	Rpo3-STREP-HIS HmgCoAR plasmid	296 ng/µL
2			378 ng/µL
3			292 ng/µl
4			282 ng/µL
1	MURPf 63	Rpo1C-STREP-HIS HmgCoAR plasmid	346 ng/µL
2			386 ng/µL
3			360 ng/µL
4			444 ng/µL
1	MURPf 50	Rpo1C-T82L-STREP-HIS HmgCoAR plasmid	464 ng/µL
2			286 ng/µL
3			362 ng/µL
4			416 ng/µL
1	MURPf 51	Rpo1C-T82R-STREP-HIS HmgCoAR plasmid	346 ng/µL
2			288 ng/µL
3			302 ng/µL
4			316 ng/µL
1	MURPf 75	HISonLinker HmgCoAR plasmid	240 ng/µL
2			302 ng/µL
3			298 ng/µL
4			256 ng/µL

The samples were then analysed via mass spectrometry at the Max Planck Institute Göttingen (Dr. Henning Urlaub, Dr. Sabine König, and Olexandr Dybkov), and the resulting dataset was analysed by Dr. Felix Grünberger as described in *Tarau et al. 2024*[131].

By small-scale *Pfu* RNAP purification, the ribosomal proteins, even if present, remain in an insignificant amount, as shown in Figure 25. The ribosomal protein L3 from the large subunit of the ribosome (50S) appears enriched in two preparations. Nevertheless, L3, being the largest protein of the *Pfu* ribosome, can have a significant amount due to the mass spectrometry preference for larger proteins. The putative ribosomal protein S10 from the small subunit is present in insignificant numbers. On the other hand, due to its small size, detecting S10 with mass spectrometry might be challenging.

The volcano plots display all subunits of the *Pfu* RNAP and enriched proteins in all samples. These analyses also confirmed a tight association with archaeal transcription factors. Indeed, transcription initiation factor TFE and transcription elongation factor Spt5 are enriched in all samples. Even if Spt4 is absent in the data analysis plots, it may be due to the technique's limitation in detecting small proteins. Although Spt4 is well known to associate with transcription factor Spt5 and form a stable heterodimer.

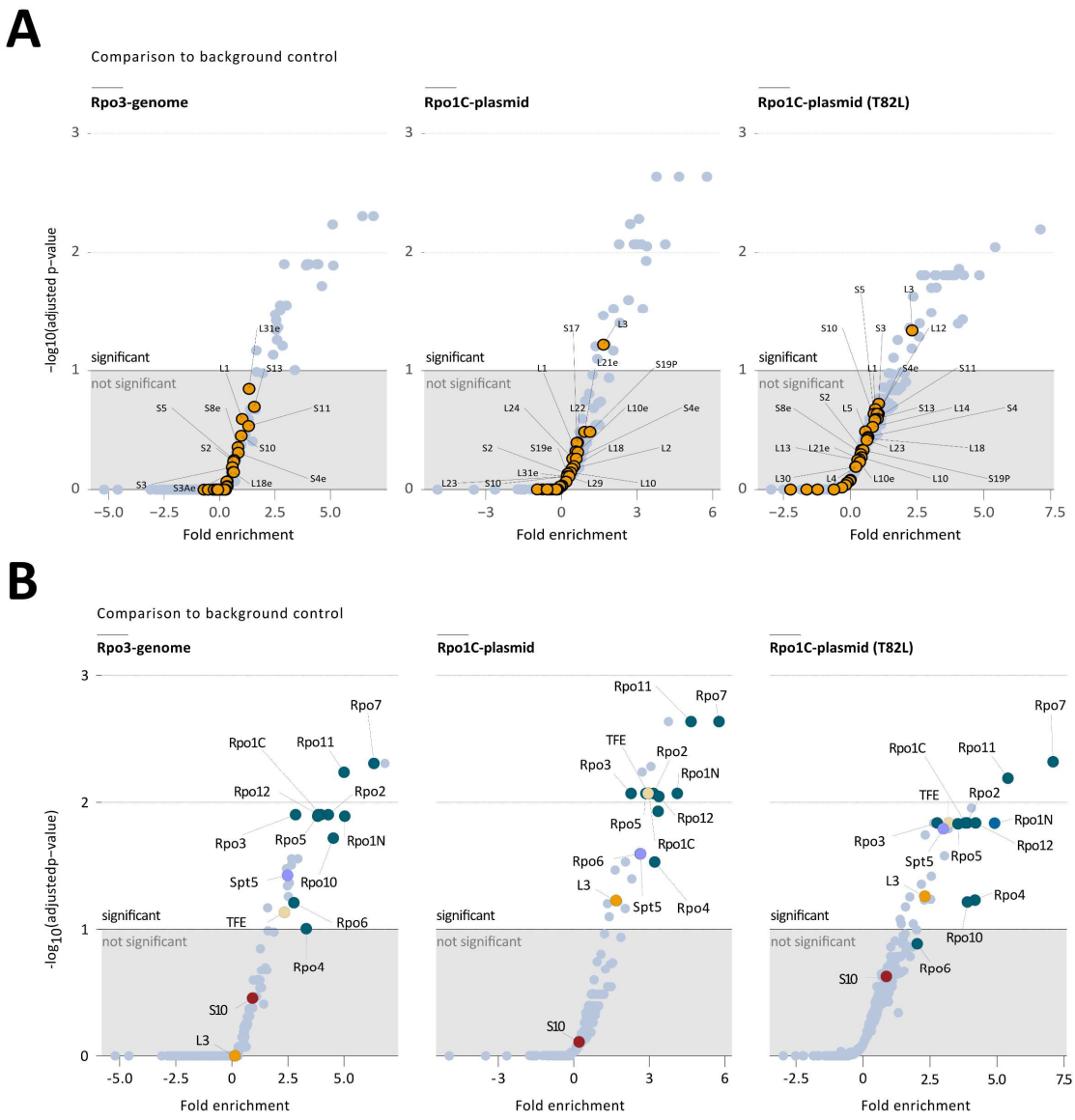


Figure 26. Mass spectrometry analysis of *Pfu* RNAP protein interactome. A. Volcano plot of mass spectrometry analysis conducted for the *P. furiosus* strains that encode a Strep-HIS-tagged variant of the subunit Rpo3, Rpo1C wild type, and Rpo1C T82L mutant from the shuttle vector (indicated as "-plasmid") with respect to the negative control ("His-tag only" mutant). Fold enrichment over background was calculated using a one-sided Wald test with multiple testing correction using Benjamin-Hochberg. The plots show only the ribosomal protein enrichments, with only one of them, L3, being significantly enriched in the case of the Rpo1C-plasmid tagged samples. **B.** Vulcano plots of the same samples analysed in the same way as described in **A.,** showing the RNA polymerase subunits in petrol and the enriched transcription factors together with the ribosomal protein S10, whose enrichment is not significant.

3. Materials and Methods

3.1 Materials

3.1.1 Chemicals

Table 5. Chemicals used and the respective supplier.

Chemical name	*Supplier*
Alexa Fluor™ 488 C5 Maleimide	Thermo Fisher Scientific
Alexa Fluor™ 647 C2 Maleimide	Thermo Fisher Scientific
BS_3 (bis(sulfosuccinimidyl)suberate)	Fisher Scientific
Agar-Agar	Roth
Agarose	Roth
Ampicillin	Serva
Ammonium Persulfate	Roth
Bacto-Tryptone	Fisher Scientific
Bacto-Yeast Extract	Fisher Scientific
C-40 CBB-G250	MP Biochemicals
CaCl x 2 H_2O (calcium chloride dihydrate)	Roth
$CoSO_4$ x 7 H_2O (cobaltous sulfate heptahydrate)	Fluka
$CuSO_4$ x 5 H_2O (copper sulfate pentahydrate)	Merck
Dithiothreitol (DTT)	Roth
$FeSO_4$ x 7 H_2O (ferrous sulfate heptahydrate)	Aldrich
Glycerol (87%)	Fisher Scientific
Glycine	Sigma
H_3BO_3 (boric acid)	Fisher Scientific
HEPES(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	Roth
Imidazole	Roth
Isopropyl-ß-D-thiogalactopyranosid (IPTG)	Fermentas
$KAl(SO_4)_2$ x 12 H_2O (chromium(III) potassium sulfate dodecahydrate)	Merck
KCl (potassium chloride)	Merck
KH_2PO_4 (monobasic potassium phosphate)	Merck
KI (potassium iodide)	Merck
KOH (potassium hydroxide)	Merck
$MgCl_2$ (magnesium chloride)	Roth
$MgCl_2$ x 6 H_2O (magnesium chloride hexahydrate)	Merck
$MgSO_4$ (magnesium sulfate)	Merck
$MgSO_4$ x 7 H_2O (magnesium sulfate heptahydrate)	Merck
$MnSO_4$ x H_2O (manganese sulfate monohydrate)	Merck

Na_2S (sodium sulfade)	Sigma
Na_2SeO_4 (sodium selenade)	Sigma
NaBr (sodium bromide)	Merck
NaCl (sodium chloride)	Roth
NaMoO x 2 H_2O (sodium molybdate dihydrate)	Roth
NaOH (sodium hydroxide)	Sigma
NH_4Cl (ammonium chloride)	Acros
NLS (N-lauronylsarcosine)	Sigma
Peptone	Bacton, Dickenson and company
Resazurin	Serva
ROTIPHORESE®NF-Acrylamide/Bis-solution 30 (29:1)	Roth
ROTIPHORESE®NF-Acrylamide/Bis-solution 40 (29:1)	Roth
SDS (sodium dodecyl sulfate)	Roth
$SrCl_2$ (strontium chloride)	Merck
Starch	Merck
SYBR-safe	Invitrogen
Tetramethylethylenediamine	Roth
Tris-(hydroxymethyl)-amino methane	Roth
Tween 20	Sigma
Urea	Roth
Yeast Extract	Fisher Scientific
$ZnSO_4$ x 7 H_2O (zinc sulfate heptahydrate)	Honeywell

3.1.2 Consumables

Immobilon®-E PVDF Membrane (Merck)

Dynabeads™ His-Tag Isolation and Pulldown (Thermo Fisher Scientific)

MagneHis™ Protein Purification System (Promega)

Ni-NTA Magnetic Agarose Beads (Qiagen)

G2400C Copper grids (Plano GmbH)

UltrAuFoil R 1.2/1.3 Gold 300 mesh (Quantifoil)

cOmplete Tablets EDTA-free, *EASY*pack (Roche)

3.1.3 Enzymes

Phusion® High-Fidelity DNA Polymerase (New England Biolabs)

3.1.4 Kits

Pierce™ Silver Stain Kit (Thermo Fisher Scientific)

Plasmid Miniprep Kit I, peqGOLD (VWR)

3.1.5 Instruments

Autoclave Sanoclav LaM-MCS (Maschinenbau Wolf GmbH)

Azura Bio purification – Two Step Purification System (Knauer)

- Azura® Assistant ASM 2.1L
- Azura® Assistant ASM 2.1L
- Azura® Assistant ASM 2.1L (Sample Injection Assistant)
- Azura® Pump P 6.1L
- Azura® UV Detector UVD CM 2.1S
- Foxy Fraction collector

Azura Bio purification – Two Step Purification System (Knauer)

- Azura® Assistant MWD 2.1L
- Azura® Assistant ASM 2.2L (Sample Injection Assistant)
- Azura® Assistant ASM 2.2L
- Azura® UV Detector UVD CM 2.1S
- Foxy Fraction collector

FastPrep-24™ 5G bead beating grinder and lysis system (MP Biomedicals)

Cressington Turbo Carbon Coater 208carbon (Cressington)

C1000 Touch™ Thermal Cycler (Bio-Rad)

Centrifuge 17S/RS (Heraeus Sepatech)

Ultracentrifuge Avanti™ J-25 (Beckman Coulter GmbH)

Optima MAX Ultracentrifuge (Beckman Coulter GmbH)

Avanti® J-25 High Speed Centrifuge (Beckman Coulter GmbH)

Optima LE-80K Ultracentrifuge (Beckman Coulter GmbH)

Centrifuge 5430 R (Eppendorf AG)

Centrifuge 5810 R (Eppendorf AG)

JA-10 Fixed-Angle Aluminium Rotor (Beckman Coulter GmbH)

JA-25.50 Fixed-Angle Aluminium Rotor (Beckman Coulter GmbH)

Type 70 Ti Fixed-Angle Titanium Rotor (Beckman Coulter GmbH)

Elmasonic S 30 H (Elma Schmidbauer Gmbh)

Imaging System ChemiDoc™ MP (Bio-Rad Laboratories Inc.)

Mini-PROTEAN Tetra Cell (Bio-Rad Laboratories Inc.)

Photometer Ultraspec® 10 (Biochrom GmbH)

Refeyn TwoMP mass photometer (Refeyn Ltd.)

NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific)

Ultrasound Homogenizers Sonopuls HD 2070 (Bandelin electronic GmbH & Co. KG)

Äkta Purifier system (GE Healthcare)

Incubation Shaker Multitron (Infors AG)

EasiGlow (TedPella)

Vitrobot Mark IV (Thermo Fischer Scientific)

JEM-2100F Field Emission Electron Microscope (JEOL Germany GmbH, Freising) with 200 kV FEG and F416 CMOS camera (TVIPS GmbH, Gauting, Germany)

CRYO ARM™ 200 (JEM-Z200FSC) Field Emission Cryo-Electron Microscope (JEOL, Japan) with Gatan K2 summit camera (Gatan)

Typhoon™ FLA 7000 (GE Healthcare)

Bio-Rad Gel Dryer Model 543 (Bio-Rad Laboratories)

VWR Doppio Thermocycler (VWR International)

C100 Touch (Bio-Rad Laboratories)

3.1.6 Chromatography columns

Table 6. Chromatographic columns and their respective suppliers

COLUMN TYPE	BRAND
HITRAP HEPARIN HP	Cytiva
HISTRAP HP 1 ML FAST FLOW	Cytiva
HISTRAP HP 5 ML FAST FLOW	Cytiva
SUPERDEX® 75 10/300 GL	GE Healthcare
SUPERDEX® 200 10/300 GL	GE Healthcare
SUPEROSE™ 6 INCREASE 3.2/300	GE Healthcare
HILOAD™ 26/60 SUPERDEX™ 200 PREP GRADE	GE Healthcare
HILOAD™ 16/60 SUPERDEX™ 75 PREP GRADE	GE Healthcare

3.1.7 Software and online tools

- RELION 3.1, 4.0 beta and 5.0
- SerialEM
- R studio
- ChimeraX 1.6
    - Isolde
    - AlphaFold 2
- UCSF Chimera 1.15
- Coot 0.9.8.1
- PHENIX 1.20.1 – 4487
- Unipro UGENE
- LocScale 2.0
- ProtParam (Expasy)
- Swiss-model (Expasy)
- 3DFSC
- PSIPRED (UCL Department of Computer Science: Bioinformatics Group)
- AcuireMP/DiscoverMP (Refrein)
- ImageLab (Bio-Labs Laboratories)
- SnapGene
- MO.Analysis (Version 2.2.6)

3.1.8 Media

Luria-Bertani Broth (LB medium), 1 L

- 10.0 g Bacto-Tryptone
- 5.0 g Yeast Extract

- 10.0 g NaCl
- add sterile filtered water to a final volume of 1 L

Luria-Bertani Agar (LB Agar) 1 L

- 10.0 g Bacto-Tryptone
- 5.0 g Yeast Extract
- 10.0 g NaCl
- 1.5% (w/v) Agar-Agar
- add sterile filtered water to a final volume of 1 L

½ SME – Anaerobic Media for *Pyrococcus furiosus* 1 L[123]

- 500.0 mL Synthetic Sea Water mixture

  Recipe for 1 L

  - 27.7 g NaCl (473.99 mM)
  - 7.0 g $MgSO_4$ x 7 H_2O (28.4 mM)
  - 5.5 g $MgCl_2$ x 6 H_2O (27.1 mM)
  - 0.75 g $CaCl_2$ x 2 H_2O (5.1 mM)
  - 0.65 g KCl (8.7 mM)
  - 0.1 g NaBr (0.97 mM)
  - 0.03 g H_3BO_3 (0.49 mM)
  - 0.015 g $SrCl_2$ x 6 H_2O (0.056 mM)
  - 0.1 mL KI – Lsg 0.1 % (v/v) (0.60 µM)
  - ddH_2O until final volume
- 0.5 g KH_2PO_4
- 0.5 g NH_4Cl
- 1.0 g Yeast Extract
- 1.0 g Peptone
- 1 mL Wolfe's Minerals 10x pH 1.0

  Recipe for 1 L

  - 30.0 g $MgSO_4$ x 7 H_2O (121.7 mM)
  - 5.0 g $MnSO_4$ x H_2O (29.6 mM)
  - 10.0 g NaCl (171.1 mM)
  - 1.0 g $FeSO_4$ x 7 H_2O (3.6 mM)
  - 1.8 g $CoSO_4$ x 7 H_2O (6.4 mM)
  - 1.0 g $CaCl_2$ x 2 H_2O (6.8 mM)
  - 1.8 g $ZnSO_4$ x 7 H_2O (6.3 mM)
  - 0.1 g $CuSO_4$ x 5 H_2O (0.4 mM)
  - 0.18 g $KAl(SO_4)_2$ x 12 H_2O (0.4 mM)
  - 0.1 g H_3BO_3 (1.6 mM)
  - 0.1 g Na_2MoO x 2 H_2O (0.4 mM)
  - 2.8 g $(NH_4)_2Ni(SO_4)_2$ x 6 H_2O (7.0 mM)

    - 0.1 g Na_2WO_4 x 2 H_2O (0.3 mM)
    - 0.1 g Na_2SeO_4 (0.5 mM)
    - ddH_2O to reach the final volume
    - Reach pH 1.0 with MH_2SO_4
  - 0.5 mL Resazurin (0.2% w/v)
  - 0.5 g Na_2S
  - 1.7 g Starch
  - NaOH until pH 6.5 is reached

3.1.9 Buffers

Elution Buffer 1

- 40.0 mM HEPES/KOH pH 7.4
- 50.0 mM NaCl
- 500.0 mM Imidazole
- 15.0% (v/v) Glycerol
- add sterile filtered water to a final volume of 1 L

EMSA Buffer 5x

- 200.0 mM HEPES/KOH pH 7.4
- 300.0 mM NaCl
- 2.5 mM $MgCl_2$ x 6 H_2O
- 0.5 mM EDTA
- 0.1 mg/mL BSA
- add sterile filtered water to a final volume of 1 L

Low Salt Buffer

- 40.0 mM HEPES/KOH pH 7.4
- 50.0 mM NaCl
- 15.0% (v/v) Glycerol
- add sterile filtered water to a final volume of 1 L

Low Salt Wash buffer

- 40.0 mM HEPES/KOH pH 7.4
- 50.0 mM NaCl
- 15.0% (v/v) Glycerol
- 2.5 mM DTT
- add sterile filtered water to a final volume of 1 L

High Salt Buffer

- 40.0 mM HEPES/KOH pH 7.4

- 1.0 M NaCl
- 15.0% (v/ v) Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Lysis Buffer

- 100.0 mM Tris/HCl pH 8
- 1.0 M NaCl
- 20.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Elution Buffer 1

- 100.0 mM Tris/HCl pH 8
- 300.0 mM NaCl
- 250.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Gel Filtration Buffer

- 100.0 mM HEPES/KOH pH 8
- 150.0 mM NaCl
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Cryo-EM Buffer

- 100.0 mM HEPES/KOH pH 8
- 150.0 mM NaCl
- 2.5 mM $MgCl_2$
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Homogenizing Buffer 300

- 100.0 mM HEPES/KOH pH 8
- 300.0 mM NaCl
- 20.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Homogenizing Buffer 600

- 100.0 mM HEPES/KOH pH 8

- 600.0 mM NaCl
- 20.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Homogenizing Buffer 1000

- 100.0 mM HEPES/KOH pH 8
- 1.0 M NaCl
- 20.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Elution buffer 2.1 300

- 100.0 mM HEPES/KOH pH 8
- 300.0 mM NaCl
- 250.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Elution buffer 2.1 600

- 100.0 mM HEPES/KOH pH 8
- 600.0 mM NaCl
- 250.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Elution buffer 2.1 1000

- 100.0 mM HEPES/KOH pH 8
- 1.0 M NaCl
- 250.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Elution buffer 2.2 300

- 100.0 mM HEPES/KOH pH 8
- 300.0 mM NaCl
- 500.0 mM Imidazole
- 2.5 mM $MgCl_2$

- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Elution buffer 2.2 600

- 100.0 mM HEPES/KOH pH 8
- 600.0 mM NaCl
- 500.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Pfu RNAP Elution buffer 2.2 1000

- 100.0 mM HEPES/KOH pH 8
- 1.0 M NaCl
- 500.0 mM Imidazole
- 2.5 mM $MgCl_2$
- 20.0% Glycerol
- add sterile filtered water to a final volume of 1 L

Phosphate-buffered saline (PBS)

- 137.0 mM NaCl
- 2.5 mM KCl
- 10.0 mM Na_2HPO_4
- 1.8 mM KH_2PO_4
- add sterile filtered water to a final volume of 1 L

Native Laemmli running buffer (10x)

- 240.0 mM Tris/HCl pH 8.3
- 1.9 M Glycine
- add sterile filtered water to a final volume of 1 L

Laemmli running buffer (10x)

- 240.0 mM Tris/HCl pH 8.3
- 1.9 M Glycine
- 35.0 mM SDS
- add sterile filtered water to a final volume of 1 L

Tris-acetate-EDTA (TAE) Buffer (50x)

- 2.0 M Tris
- 1.0 M Acetic acid
- 50.0 mM EDTA disodium salt
- add sterile filtered water to a final volume of 1 L

TBE Buffer (10x)

- 891.5 mM Tris
- 889.5 mM Boric acid
- 25.5 mM EDTA
- add sterile filtered water to a final volume of 1 L

WB-TBS (10x)

- 1,500.0 mM NaCl
- 100 mM Tris/HCl (from 1.0 M, pH 8.0 stock)
- add sterile filtered water to a final volume of 1 L

WB-TBST

- 10.0% v/v TBS (10x)
- 0.1% v/v Tween 20
- add sterile filtered water to a final volume of 1 L

WB-Blocking Solution

- 1x TBST
- 5.0% Milk Powder
- add sterile filtered water to a final volume of 50 mL

WB-Transfer Buffer

- 25.0 mM Tris
- 192.0 mM Glycine
- 20.0% v/v Methanol
- add sterile filtered water to a final volume of 1 L

3.2 Methods

3.2.1 Capillary Resuspension and Anaerobic Cultivation of *Pyrococcus furiosus* mutants

The strain collection at the German Archaea Centre at the University of Regensburg keeps the isolated microorganisms in glass capillaries for long-term conservation. These capillaries contain a few microlitres of highly concentrated cells, which are sealed by glass melting under powerful flame heat. Capillaries are stored in a tank that is constantly refilled with liquid nitrogen for the process known as cryopreservation. The cells stored in the capillaries can be resuspended in fresh medium by following these steps. First, the workflow must occur under sterile conditions to prevent contamination. Moreover, the process has to be performed as fast as possible to minimize contact with oxygen, which is often harmful to anaerobic species even at low concentrations. The capillary is cut at both ends, and the contents are collected using a 1 mL syringe and then transferred to the first serum bottle filled with an appropriate medium. This solution is considered the undiluted resuspension. The latter is then used for further dilutions: 1:20 and 1:40. For *Pyrococcus furiosus* mutants (see Chapter 3.2.1), this passage must be performed in ½ SME – Anaerobic Media for *Pyrococcus furiosus* supplemented with 10 µM simvastatin to limit the growth to the mutant only. Due to simvastatin degradation, *Pyrococcus furiosus* cells cannot be grown at the optimal temperature of 100°C, but at 85°C, meaning that the doubling time is longer.

3.2.2 *Pyrococcus furiosus* pMUR1 mutant generation

The pMUR1 strain was generated by David Pöllmann in 2010. To enable a simplified purification of the cellular native RNAP, a StrepII-His6-tag was introduced at the C-terminal end of subunit Rpo3 using an established genetic system for *Pfu* DSMZ 3638[140]. To this end, a double crossover recombination strategy with plasmid pMUR1 was employed to introduce the StrepII-His6-tag at the correct position of the *rpo3* gene into the genome of *P. furiosus*. This plasmid contains *rpo3* (PF1647) together with 250 bp of the upstream region. The corresponding sequence was amplified from wild-type DNA with the primer pair 1648/1647-BH1-F and PF1647_Fus-StrepII_Ri. The StrepII-His6-tag at the C-terminal end was provided by a double-stranded oligonucleotide (primer Fus_Strep_His_F and Fus_Strep_His_R). For the selection of the *Pyrococcus* transformants with the antibiotic simvastatin, the coding sequence of the hydroxymethylglutaryl CoA reductase from *Thermococcus kodakarensis* was used. The corresponding sequence was amplified from plasmid pMUR27 using the primer pair PF1647upTk0914F and TK0914-R54. Furthermore, the plasmid includes a 950 bp sequence downstream of *rpo3* to enable double crossover recombination. This fragment was amplified from genomic DNA using the primer pair Tk0914PF1647dwF and Fus004_BamHI-R. All DNA fragments were combined step-by-step by single

overlap extension PCR. The fused fragment was hydrolyzed with BamHI and ligated into a pUC19 vector, which was also hydrolyzed with BamHI and dephosphorylated with calf intestinal phosphatase.

The resulting plasmid pMUR1 was sequenced, and one µg of the linearized version was used for transformation of Pfu as described previously[139,140]. Transformants were enriched with 10 µM simvastatin in ½ SME-starch liquid medium at 85°C for 48 h, and pure cultures were isolated after plating the cells on solidified medium containing 10 µM simvastatin. The integration of the DNA fragment into the genome via double crossover was verified by sequencing of the corresponding PCR products.

3.2.3 Gel Electrophoresis

Agarose gel electrophoresis

Agarose gel electrophoresis is a common molecular biology technique used to separate, analyse, and recover DNA fragments based on their size. In this method, 1% or 1.5% agarose is dissolved in TAE buffer, and SYBR Safe 10,000x (Thermofisher Scientific) is added to the mixture for nucleic acid detection. The agarose mixture is then poured into a gel casting support. When an electric field is applied, the negatively charged DNA migrates toward the positive electrode. The gel's pore size influences how far DNA fragments of different lengths can travel: smaller fragments move more quickly through the matrix, resulting in separation by size. The DNA fragments can be visualized using UV or blue light. This method is widely used for checking PCR products, assessing restriction enzyme digests, purifying DNA for cloning, and many other downstream applications[141,142].

Denaturing gel electrophoresis- SDS-PAGE

SDS-PAGE, or sodium dodecyl sulphate – polyacrylamide gel electrophoresis, is a widely used method for protein separation that operates by denaturing proteins and distributing a uniform negative charge-to-mass ratio[143–145]. In this technique, proteins are treated with SDS – a strong anionic detergent – that binds at an approximately constant ratio and disrupts non-covalent interactions, linearizing polypeptides while masking intrinsic charges[146]. Additional reducing agents, such as DTT and β-mercaptoethanol, are included to eliminate disulfide bridges. The denaturation is then performed by incubating for 5-10 minutes at 95°C. The protein mixture is loaded onto a polyacrylamide gel, which separates proteins based on their molecular weight (smaller proteins run faster than larger proteins). The run is performed by applying a constant voltage (180V-200V) to the running system (Mini-PROTEAN® 3 Cell, Bio-Rad). Many systems employ a two-layer design, consisting of a stacking gel (characterized by a low acrylamide concentration and low pH) that concentrates proteins into narrow bands, followed by a resolving gel (with a higher acrylamide concentration and higher pH) that separates them by size. Protein bands can be detected using Coomassie brilliant blue staining, Colloidal staining, or silver staining.

Most of the SDS-PAGEs performed for this work were 16% polyacrylamide or precast 4%-20% polyacrylamide gels. The samples were mixed with 6x Loading Dye (Table 7). The detection method employed was principally colloidal Coomassie staining and destaining.

Table 7. Recipes of gels and buffers used to perform SDS-PAGE

6x Loading dye	
4x Tris/HCl/SDS	70.0% (w/v)
Glycerol	30.0% (v/v)
SDS	10.0% (w/v)
Bromophenol blue	0.012% (w/v)
DTT	600.0 mM
10x SDS running Buffer (1L)	
Tris	250.0 mM
Glycine	1.920 M
SDS	1.0% (w/v)
H_2O	Until 1L
Colloidal Coomassie staining solution (1L)	
Aluminium sulfate x 18 H_2O	5.0% (w/v)
Ethanol (99%)	10.0% (v/v)
Coomassie Brilliant Blue (CBB-G250)	0.02% (w/v)
Ortho-Phosphoric acid (85%)	2% (v/v)
H_2O	Until 1L
Colloidal Destaining solution (1L)	
Ethanol (99%)	10% (v/v)
Ortho-Phosphoric acid (85%)	2% (v/v)
H_2O	Until 1L
4% Stacking gel	
ROTIPHORESE®NF-Acrylamide/Bis-solution 30 (29:1)	1.0 mL
Tris/HCl, pH 6.8 (0.5M)	1.9 mL
H_2O	4.7 mL
SDS (10%)	100.0 µL
APS (10%)	100.0 µL
TEMED	10 µL
16% Separating gel	
ROTIPHORESE®NF-Acrylamide/Bis-solution 30 (29:1)	8.0 mL
Tris/HCl, pH 8.8 (1.5M)	3.75 mL
H_2O	3.15 mL
SDS (10%)	100 µL
APS (10%)	100 µL
TEMED	10 µL

Native gel electrophoresis and ElectroMobility Shift Assays (EMSAs)

Native gel electrophoresis, also known as non-denaturing PAGE, separates proteins in their native, folded state based on their isoelectric point. Unlike SDS-PAGE, this technique omits denaturing agents such as detergents and reducing agents, so protein migration depends on their intrinsic properties. EMSA, or electromobility shift assays, are an application of native gel electrophoresis designed for nucleic acid-protein interaction studies. The principle exploits the natural negative charge of nucleic acids to migrate in an electric field, and when a protein binds to it, the migration slows down, producing a shifted band[147].

Denaturing urea polyacrylamide gel electrophoresis (Urea-PAGE)

Urea-PAGE (or denaturing polyacrylamide gel electrophoresis) is a high-resolution method widely used in molecular biology for separation of single-stranded DNA or RNA[148,149]. It employs a polyacrylamide mixture containing 6-8 M urea, aiming to disrupt the hydrogen bonds and secondary structures of nucleic acids, thereby efficiently denaturing samples so that they migrate according to their weight rather than their conformation. Typically, the electrophoretic run is performed at 45-55°C temperatures to further promote the maintenance of linear, denatured molecules[150]. The technique is suitable for many applications, from enzymatic reactions to oligonucleotide purifications or ensuring transcription integrity.

For this study, urea-PAGE was necessary for monitoring transcription progression of *Pfu* RNAP on an elongation transcription scaffold. Through urea-PAGE, it is possible to distinguish the transcription product with single-base resolution, making it useful for characterising the polymerase kinetics of RNA synthesis. More details about the experimental setup are explained in Chapter 3.2.17 *In vitro* Transcription Assays.

3.2.4 Protein Purification from *Pyrococcus furiosus*

The cell pellet obtained from cultivation of *Pyrococcus furiosus* mutant described above (Paragraph 3.2.1), was resuspended in 2.5 ml per g of cells lysis buffer (100 mM Tris/HCl pH 8.0, 1 M NaCl, 20 mM imidazole, 20% (v/v) glycerol, 2.5 mM $MgCl_2$) and 1 EDTA free protease inhibitor cocktail tablet (Roche) per 50 ml of lysis buffer. 0.5 g of glass beads (0.106 µm diameter, Sigma-Aldrich) were added to the mixture, and the pellet was resuspended by bead beating using the FastPrep-24™ 5G bead beating grinder and lysis system from MP at a speed of 6.5 m/sec for 40 sec. Afterwards, cells were lysed by sonication using Bandelin Sonopuls HD 2070 at 60% amplitude for 6 min. Sonication was repeated five to six times. Cell lysis was confirmed by light microscopy. After lysis, the crude extract was centrifuged at 15,000 g for 10 min (Eppendorf 5810 centrifuge) to pellet cell debris and glass beads. The supernatant was transferred and subjected to ultracentrifugation at 100,000 g using a type 70 Ti Rotor in the Optima Le-80K Ultracentrifuge (Beckmann).

The supernatant was carefully collected and filtered using 0.45 µm syringe filters (Cytiva). The polymerase was purified using affinity chromatography exploiting the 6xHis-tag. To this end, the filtrate was applied to a HisTrap™ 5 ml Fast Flow column (Cytiva) equilibrated in lysis buffer using an Äkta Purifier fast protein liquid chromatography (FPLC) system (GE Healthcare). The column was washed with lysis buffer (100 mM Tris/HCl, pH 8.0, 1 M NaCl, 20 mM imidazole, 20% (v/v) glycerol, 2.5

mM $MgCl_2$) until the absorbance at 280 nm dropped close to 0 mAU. Protein bound to the matrix was eluted through a step elution with 30 mL 100% elution buffer (100 mM Tris/HCl, pH 8.0, 300 mM NaCl, 250 mM imidazole, 20% (v/v) glycerol, 2.5 mM $MgCl_2$). Subsequently, the RNA polymerase was further purified by performing a size exclusion chromatography step (column: HiLoad 26/600 Superdex 200 PG from Sigma-Aldrich). The following running buffer was used for the size exclusion chromatography: 100 mM HEPES/NaOH, pH 8.0, 150 mM NaCl, 20% (v/v) glycerol, 2.5 mM $MgCl_2$. Elution fractions were collected and analyzed using a denaturing SDS polyacrylamide gel (PAGE) (see Chapter 3.2.3). The fractions containing the RNA polymerase were aliquoted, flash-frozen, and stored at -80°C until further use. The concentration of 1.638 mg/mL was determined by absorption spectroscopy at 280 nm wavelength using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific).

3.2.5 Recombinant Protein Purification of Thermostable Proteins

To express and purify recombinant proteins derived from the hyperthermophilic archaeon *P. furiosus*, the cloning procedure of the genes of interest resembles most of the steps of common procedure with one additional step. First, the gene of interest is introduced in an expression plasmid using the Gibson Assembly[151] strategy. In our case, the two genes for *spt4* (GeneID 41712045) and *spt5* (GeneID 41713813) from *Pyrococcus furiosus* were amplified via PCR and cloned into pET14b and pET17b plasmids through Gibson Assembly. Dr. Florian B. Heiss did this step as his bachelor's thesis[152]. The two plasmids both harbor an ampicillin resistance gene, and in the case of Spt4, the recombinant protein is expressed as a fusion with a 6xHis tag at the C-terminus.

Escherichia coli Rosetta (DE3) chemically competent cells (Sigma-Aldrich) are transformed through heat-shock protocol, which consists of 1) 30 min incubation of the plasmid (200 ng) with 50-100 µL of competent cells on ice to allow the plasmid to coat the membrane surface of the cell, 2) incubation at 42°C for 45 seconds to increase membrane permeability and allow plasmid uptake, 3) five minutes incubation on ice to restore cell membrane conditions, 4) addition of 500 µL LB medium and incubation at 37°C shaking for approximately 1 hour to allow cells to double after the transformation step. 5) spreading the cell suspension onto an LB agar plate containing the antibiotic of interest (100 µg/mL) and incubating at 37°C until colonies were grown to a sufficient colony size.

After colonies have formed, one colony was picked and resuspended in 200 µL LB medium and spread on a fresh LB^{AMP} agar plate. The plate was incubated overnight at 37°C. This will serve as a pre-culture for the higher volume overexpression.

The next day, 5 mL of LB medium is used to resuspend the bacteria from the pre-culture plate and transferred to 500 mL LB medium supplemented with ampicillin (100 µg/mL). The resulting culture is then incubated at 37°C, 200 rpm until it reaches an adequate cell density. The latter is measured using spectroscopic approaches, specifically by measuring the optical density at 600 nm (OD_{600}). The OD_{600} before starting incubation was typically around ~0.05. Induction of gene expression is performed at an OD_{600} of ~0.6. Protein overexpression is induced by isopropyl-β-d-thiogalactopyranoside (IPTG), which induces the expression of the T7 polymerase present in the chromosome of Rosetta (DE3) chemically

competent cells. The T7 promoter is then able to recognise the promoter and terminator of the inserted plasmid, and it can drive a high-level transcription of the target gene[153]. The resulting mRNA is translated by the host ribosomes, resulting in the efficient accumulation of the recombinant protein. Therefore, when the desired OD_{600} is reached, IPTG (at a final concentration of 0.5 mM) is added to the bacterial culture. At this point, the cells typically begin to express the recombinant protein in large amounts. After expression induction, the temperature is lowered to 18°C, and the overexpression can be carried out overnight, or the overexpression can be carried out at 37°C for three to four hours. The cells are then pelleted by centrifugation at 5,000 g/4°C for 15 minutes using a JA-10 fixed-aluminium rotor in the Avanti J-25 refrigerated floor centrifuge (Beckman). Afterwards, the supernatant is discarded, and the pellets are resuspended in PBS buffer and collected in a 50 ml falcon tube. The falcon tube is then placed in the centrifuge 5810 R (Eppendorf), and the cells are pelleted down at 5,000 g/4°C for 15 minutes. The supernatant is discarded again, and the pellet is stored at -80 °C for further use.

To verify that the overexpression was successful, a 1 mL aliquot of the culture at an optical density (OD600) of 0.6 before IPTG addition and a 1 mL aliquot of the culture before cell harvesting are collected in a 1.5 mL Eppendorf tube. The small culture aliquots are pelleted in a bench centrifuge 5430 R (Eppendorf) for 1 minute at 10,000 g, at room temperature. The pellets are resuspended in 1x protein loading buffer, using 100 µL for resuspension before induction and 300 µL for resuspension before harvesting. The suspensions are then incubated at 95°C in a thermoblock for 10 minutes. Afterwards, 5 µl, 10 µl, and 15 µl of each suspension are loaded on an SDS-PAGE and run at 180 V constant voltage. The gel is then stained in colloidal Coomassie blue solution and destained in colloidal destaining solution. The success of overexpression can be appreciated by comparing the pellets before IPTG addition and before harvesting, which should exhibit an additional band at the expected molecular weight of the protein of interest.

The cell pellet corresponding to the whole overexpression is resuspended in Low Salt Buffer and sonicated five times for 30 seconds using a Bandelin Sonopuls HD 2070 at 50% amplitude and cycle 5. Cell debris is eliminated by centrifugation (JA-25.50 fixed rotor; 1 hour; 20,000 g; 4°C). Spt4 and Spt5 maintain their thermostability even when recombinantly expressed; therefore, the lysate is incubated for 20 minutes at 80°C in a water bath to denature the proteins of the expression host *E. coli*. The precipitates formed by the heat-shock are eliminated via centrifugation (JA-25.50 fixed rotor, 20,000 g, 15 minutes, 4°C). The supernatant is then sterile filtered using a 0.2 µm syringe filter. The protein is then purified in a first step by affinity chromatography and eluted by an elution gradient. Based on the complex, it can be used as an imidazole-containing elution buffer for proteins with a His-tag or high salt concentration for proteins with affinity for nucleic acids. Spt4, Spt4/5, and S10 were purified via their 6xHis-tag using HisTrap 1 mL Fast Flow column (Cytiva), while Spt5 was purified using HiTrap Heparin HP 1 mL.

3.2.6 Fluorescent Labelling of Proteins

There are many techniques to fluorescently label proteins[154–156]. For our investigations, we performed maleimide fluorescence labelling[157]. For this technique, a fluorophore is coupled to a maleimide, which

is a highly reactive chemical group that contains an alkene (C=C) structure. This double bond is essential for the maleimide's reaction with thiol groups (-SH) that are found in cysteine residues of proteins. Cysteine's nucleophilic thiol group undergoes Michael addition with the maleimide double bond. The protein is then labelled with the maleimide fluorophore as a result of the stable covalent link that is formed between the maleimide and thiol groups. Here, this procedure was used to label the ribosomal protein S10 and the S10Δ version fluorescently. The S10 protein from *Pyrococcus furiosus* does not contain any natural cysteine residues. To genetically engineer a cysteine residue into this protein, a DNA triplet coding for a cysteine residue was inserted into the expression plasmid. The cysteine insertion was designed to be expressed at the C-terminus of the protein. The modified protein was expressed and purified using the same procedures described in Chapter 3.2.5 except for the affinity chromatography step. At this point, once the protein is immobilised on the column, an additional wash step with Wash Low Salt Buffer, which contains DTT, is performed. This step is fundamental for reducing the cysteine in the protein. Nevertheless, DTT has to be washed away with regular Low Salt Buffer, because it also contains a thiol group. If present in the eluted protein solution, the thiol group of DTT would competitively react with the maleimide group, thereby lowering the protein labelling efficiency. The two washing steps were performed with 10 column volumes of the respective buffers. After the protein was eluted from the column, its concentration of the protein was determined using the Lambert-Beer Law. Based on the protein concentration, the labeling was carried out using a 1:10 molar ratio of protein with respect to the dye, usually in a 100 µL final volume. The dye used was Alexa Fluor 647, and the incubation time varied based on the temperature, with a duration of 4 hours at room temperature and overnight at 4 °C.

The excess of dye was eliminated via size exclusion chromatography using Superdex® 75 10/300 GL column through isocratic elution in Low Salt Buffer. This chromatography step dilutes the protein content considerably. The protein was then concentrated using Centrisart® 5,000 D concentrators (Vivaproducts). These concentrators are well-suited to prevent the non-specific attachment of the fluorophores to the plastic material or filter material of the concentrator.

3.2.7 Protein Mutagenesis

Residues of interest were mutated via mutagenic primer PCR. Spt5 amino acids were mutated to alanine for the *in vitro* transcription assays from Chapter 2.6 and binding assays explained in chapter 2.7. A cysteine residue was also inserted in the sequence of S10 and S10Δ mutant to perform maleimide fluorescent labelling for binding assays reported in chapter 2.7. The forward mutagenesis primer was phosphorylated at the 5' end and contained the mutated sequence at the respective positions (Table 9 and Table 10). The reverse primer was designed to cover the complementary sequence just upstream of the forward primer with a PCR mixture listed in Table 8. After performing a gradient PCR to amplify the whole plasmid with the conditions shown in Table 9, the successful amplification was verified by 1% agarose gel electrophoresis. 5 µL from the PCR product was circularized using 1 µL of 1x T4 DNA ligase and 2 µL ligation buffer 10x (NEB) in a final volume of 20 µL. To stop the ligation reaction, the sample was incubated at 60°C for 1 h to denature the T4 DNA ligase. 5 µL from the ligation mixture was used to transform 100 µL *E. coli* DH5α competent cells. Transformants were selected on 100 µg/mL ampicillin-containing LB agar plates (incubation of plates overnight at 37°C). A colony was collected and resuspended in 5 mL of liquid LB medium containing

100 µg/mL ampicillin, and then cultivated overnight at 37°C. Plasmid isolation was performed using the peqGOLD Plasmid Miniprep Kit I (VWR Life Science). Successful mutagenesis was verified by Sanger sequencing.

Table 8. Mutagenesis PCR reaction.

Substance	Stock concentration	Volume (µL)
Phusion HF Buffer	5x	10.0
dNTP mix	10 mM	1.0
Forward Primer	10 µM	2.5
Reverse Primer	10 µM	2.5
Phusion DNA Polymerase	2 U/µL	0.5
Template	50 to 100 ng/µL	1.0
DEPC H_2O		32.5
Total volume		50.0

Table 9. Mutagenesis PCR program conditions.

Cycles	Temperature	Time	Volume
1x	98°C	30 sec	25 µL
31x	98°C	15 sec	
	Gradient 62.3-50.3	30 sec	
	72°C	4 min	
1x	72°C	5 min	

Table 10. Spt5 mutagenesis primers.

Oligo Name	DNA sequence (5'-3')	Modification
Spt5_F105A_F_DT	GCT AAG GGA CAA AAG GCA AAA GTT G	Modification at 5' Phosphorylation
Spt5_F105A_R_DT	TGG ACC GGC AAT AAC TTC AAC	
Spt5_I133A_F_DT	GCT CCC GTT ACC ATA AAG GGA GAT TAC	Modification at 5' Phosphorylation
Spt5_I133A_R_DT	GGG AAC TAT GGC ATC AAT GAA C	
Spt5_P104A_F_DT	GCT TTT AAG GGA CAA AAG GCA AAA G	Modification at 5' Phosphorylation
Spt5_P104A_R_DT	ACC GGC AAT AAC TTC AAC GAG G	
Spt5_V131A_F_DT	GCT CCC ATT CCC GTT ACC ATA AAG	Modification at 5' Phosphorylation
Spt5_V131A_R_DT	TAT GGC ATC AAT GAA CTG AAC AAC AAC	
Spt5I133A_R _2x	GGG AGC TAT GGC ATC AAT GAA C	
Spt5F105A_R_2x	AGC ACC GGC AAT AAC TTC AAC	

Spt5I133A_R _2x2	GGG AGC TAT GGC ATC AAT G	
Spt5_V131A_F_2x	GCT CCC GCT CCC GTT ACC ATA AAG	Modification at 5' Phosphorylation
Spt5_V131A_R_2x	TAT GGC ATC AAT GAA CTG AAC AAC AAC	
Spt5_R67A	GCT GGT GTT CTT CCA GGA GAG GTT C	Modification at 5' Phosphorylation
Spt5_R67A_R	AGC GTG TCT AAT TCC CCT TAT TGC C	
Spt5_R64A_H65A_F	GCT GCT GCT GCT GGT GTT CTT C	Modification at 5' Phosphorylation
Spt5_R64A_H65A_R	AAT TCC CCT TAT TGC CTC ATC CAC AAC GCC	
Spt5_H12A_F	GCT GGA CAG GAA GAG ACA ACT G	Modification at 5' Phosphorylation
Spt5_H12_R	AGT CAC TCT CAC AGC AAA GAT TTT CC	

Table 11. S10/S10Δloop mutagenesis primers.

Oligo Name	**DNA sequence (5'-3')**	**Modification**
S10DeltaLoopCysMut_Forward	TGT GGA CAA AAG GCA CGG ATC AAA ATT GC	Modification at 5' Phosphorylation
S10DeltaLoopCysMut_Reverse	CAT GGC GCC CTG AAA ATA AAG ATT CTC	
Cys_mut_2ndtry_Forward	CAG GGC GCC ATG TGT GGA CAA AAG GCA CGG	
Cys_mut_2ndtry_Reverse	CCG TGC CTT TTG TCC ACA CAT GGC GCC CTG	
S10Delta_Cysmut-F	TGT GGA CAA AAG GCA CGG ATC	Modification at 5' Phosphorylation
S10Delta_Cysmut_R	CAT GGC GCC CTG AAA ATA AAG	

3.2.8 Transcription Elongation Complex assembly

To mimic the elongating state of *Pfu* RNAP, a synthetic transcription elongation scaffold was used to assemble a transcription elongation complex (TEC). For this purpose, we exploited the protocols previously described by Hirtreiter et al.[49] . The components of a synthetic elongation scaffold comprise a DNA Template Strand (TS), an RNA oligonucleotide partially annealing to the TS, and a DNA Non-Template Strand (NTS). The Transcription Elongation Complex assembly consists of a few steps: first, the TS is annealed to the RNA oligonucleotide in a 1:9 molar excess. The annealing occurs through three minutes of incubation at 95°C and slowly cooling down in a thermocycler. Secondly, the *Pfu* RNAP

is added to the pre-annealed RNA-TS in a 1:3 molar excess, respectively, by incubation at 80°C for 15 minutes. Lastly, the NTS is added to the mixture in a 1:3 molar excess of NTS with respect to the *Pfu* RNAP. To anneal the NTS to the preformed complex, the mixture was finally incubated at 80°C for 10 minutes. For our studies, we have used two different scaffolds. One consisting of long DNA oligonucleotides (TS-long and NTS-long) and a short RNA (T40-short). The second one was formed by two short DNA oligonucleotides (TS-short and NTS-short) and a long RNA oligonucleotide (T40-long) (Table 12). The characteristics of the T40-short RNA oligonucleotide are that it anneals to the TS from position 13 to 21 (nine nucleotides complementary) and five nucleotides non-complementary.

Table 12. Oligonucleotides used for synthetic transcription bubble formation.

NAME	SEQUENCE
TS-LONG	5'-CCACCCTTACCTCCACCATATGGGAGATCCATTAGAGTAGTTAAGATGAAGTAGTTACGCCTGGTCATTACTAGTACTGCCGG-3'
NTS-LONG	5'-CCGGCAGTACTAGTAATGACCAGGCGTAACTACTTCATCTTAACTACTCTAATGGATCTCCCATATGGTGGAGGTAAGGGTGG – 3'
T40-SHORT	5'-AUUUAGACCAGGCG-3'
TS-SHORT	5'- GGG AAGTAGTTACGCCTGGTCATTACTAGTACTGC -3'
NTS-SHORT	5'- GCAGTACTAGTAATGACCAGGCGTAACTACTTCCC -3'
T40-LONG	5'-UUUUUUUUAUUUAGACCAGGCG-3'

3.2.9 Protein chemical crosslinking

To chemically crosslink proteins, the protocol from Pilsl et al.[158] was adapted to *Pfu* RNAP. To this end, glutaraldehyde and BS3 (Bis(sulfosuccinimidyl)suberat) were used. Glutaraldehyde crosslinking was not suitable for *Pfu* RNAP due to the formation of over-crosslinking artifacts; therefore, the BS3 crosslinking was optimized instead. The purpose of the crosslinking reaction was to increase Spt4/5 occupancy on the transcription elongation complex. BS_3 is formed by an eight-carbon spacer chain and two amine-reactive N-hydroxysulfosuccinimide (NHS) esters at its ends. NHS esters form a stable amine bond with primary amines, which are generally found on lysine residues. This reaction happens only when lysine residues (K) or N-termini of each polypeptide are in close proximity, in the range of the carbon arm length (11.4 Å). This feature makes BS_3 a mild crosslinker with a low probability of inducing over-crosslinked precipitate formation.

3.2.10 Sample preparation for mass spectrometry analysis

P. furiosus pMUR1, pMUR4, pMUR50, pMUR51, pMUR63 and pMUR75 strains were resuspended from a fresh capillary from the Regensburg Bacterial Bank and cultivated in 40 ml ½ SME medium supplemented with 0.1% (w/v) yeast extract, 0.1% (w/v) peptone, 0.4% (w/v) sodium pyruvate and 0.05% (w/v) Na_2S and using a N_2 gas phase (80 kPa pressure). The medium was furthermore supplemented with 10 µM simvastatin to restrict the growth of the *P. furiosus* mutants that carry the genetically modified cells. All mutants except for the pMUR1 strain were genetically engineered with the strategy described in Waege et al., 2010[139]. The re-cultivated cultures were grown at 85°C for approximately 72 h, shaking. Five cultures of each mutant were inoculated in the same medium in the absence of simvastatin and grown at 95°C for 12 h. The cultures were pelleted at 4,000 g for 30 min at 4°C.

The mutant pMUR1 carries a genetically modified *rpo3* gene. In the modified version, two affinity tag sequences, 6xHis-tag and a STREP II tag, were fused to the *rpo3* gene. These insertions confer the expression of all endogenous Rpo3 proteins in fusion with the two affinity tags located at its N-terminus. The mutagenesis procedure was described in paragraph 3.2.1.

The shuttle vector of pMUR4 contains the same mutation as pMUR1, but instead of a genomic insertion, the *rpo3* gene is genetically fused with two tags (a Strep-tag and a 6xHis-tag), which are expressed from a plasmid. Therefore, cellular Rpo3 proteins are a mixture of wild-type genomic Rpo3 and Rpo3 fused with the two tags. The „His-tag only" mutant was used as a negative control since only the His-tag, but not the *rpo3* gene, is encoded on the expression plasmid. Samples pMUR63, pMUR50 and pMUR51 are mutants transformed with a plasmid able to express subunit Rpo1C fused to a STREP II-tag and 6xHis-tag. The difference between the three mutants is that pMUR63 expresses a wild-type tagged protein, while the other two express two variants of subunit Rpo1C, e.g. proteins that carry the mutations T82L and T82R, respectively.

The re-cultivated cultures were grown at 85°C for approximately 72 h. Five cultures of each mutant were inoculated in the same medium in the absence of simvastatin and grown at 95°C for 12 h. The cultures were pelleted at 4,000 g for 30 min at 4°C. The cells were plunge-frozen in liquid nitrogen and kept at -80°C until further use.

At first, a few purification tests were run on cell pellets from pMUR1 pellets to verify the buffer conditions and the efficiency of purification approaches. For the buffer screening, three different salt concentrations were tested: 300 mM, 600 mM, and 1 M NaCl. The quantity of salt was tested at all stages of the purification process, namely homogenization, washing, and elution.

600 mM NaCl were used in the buffer, given the results of the first screening. Next, magnetic beads from three different companies were used: Dynabeads His-Tag Isolation and Pulldown (Thermo Fisher Scientific), MagneHis™ Protein Purification System (Promega) and Ni-NTA Magnetic Agarose Beads (Qiagen). Since the magnetic beads from Promega for His-tagged protein purifications performed notably better than the other two, they were used for the next steps.

After finding the right conditions, the experiment was carried out using one of the five pellets from each mutant to determine the reproducibility of the improved protocol to the actual samples before mass spectrometry analysis. Due to the reduced starting material, cell pellets were resuspended in 700 µl *Pfu* RNAP Homogenising Buffer 600 and resuspended by vortexing. The cells were lysed by sonication using Bandelin Sonopuls HD 2070 at 25% amplitude for three minutes. The sonication was repeated six times and performed on ice. Cell debris was removed by ultracentrifugation at 20,000 g for 1 hour at 4°C using a Beckman Optima MAX Ultracentrifuge and a TLA-55 rotor. The supernatant was carefully collected and loaded onto 30 µl MagneHis™ beads. The samples were eluted twice by incubating the beads with 50 µl *Pfu* RNAP Elution Buffer 2.2 600 for 15 min. The success of the procedure was assessed by SDS-PAGE and silver staining using the Pierce™ Silver Stain Kit according to the manufacturer's instructions.

The same procedure was repeated for the remaining samples, and the eluates were precipitated in 250 µL 99.5% acetone (pre-cooled to -20°C) by incubating the samples at -20°C for one hour. The precipitated protein mixture was centrifuged at 18,000 g for one hour at 4°C. The supernatant was

carefully removed, and the pellet was resuspended in 5 µl of 6x protein loading buffer solution. The samples were then sent away for mass spectrometry measurement and analysis.

3.2.11 Mass Photometry

Mass photometry is an easy-to-use technique for determining the mass of biomolecules in solution. The principle behind it is that the sample of interest is applied to a support, such as a glass slide. As the biomolecules collide with the support, they generate a small yet detectable scattering signal. The latter is directly proportional to the sample mass, meaning that the bigger the mass of the biomolecule of interest, the stronger the scattering signal will be. The measurement is a result of the interference between the light reflected by the support (glass slide) and the scattered light by the molecules. The measurement is referred to as mass photometry contrast or interferometric contrast, which is directly correlated to the molecular mass of the molecule[159]. Mass photometry can detect molecular masses ranging from 30 kDa to 5 MDa. It is suitable for either protein solutions and nucleic acids solutions and lipids or a mixture of all. The application of this powerful technique does not only apply to molecular masses of single isolated biomolecules but also allows us to study complex formation or oligomerisation. Another advantage is that the measurement is carried out in solution, allowing for quick screening of buffer conditions. Additionally, it is a label-free technique that does not require labelling with bulky dyes, which can interfere with correct molecule folding, assembly, or reactivity.

For mass photometry measurements, the Refeyn Two MP (Refeyn) is used. The machine is mostly automated except for the calibration and sample application. Two specialised software packages, AcquireMP for acquisition and DiscoverMP for data processing, are necessary for the machine to function and for data analysis. Other consumables include silicon gaskets with six spots distributed in two rows 3x3, coverslips (Refeyn), immersion oil (Zeiss Immersol 518 F) for the objective, analytical isopropanol and lens cleaning tissues for cleaning the oil from the objective.

3.2.12 Microscale Thermophoresis

MicroScale Thermophoresis (MST) was employed in this thesis to investigate the interaction between the transcription elongation factor Spt5 and the ribosomal small subunit protein S10. MST is a biophysical technique that detects molecular interactions by measuring changes in the motion of molecules within a microscopic temperature gradient. The gradient is generated by an infrared laser focused into a thin glass capillary containing the binding partners[160,161]. Upon heating, molecules redistribute according to size, charge, hydration shell, and binding state.

In this study, ribosomal protein S10 was fluorescently labelled with Alexa Fluor 647 (see Section 3.2.6), enabling detection of its thermophoretic behaviour. A titration series of unlabelled Spt5 was prepared and mixed with a constant amount of labelled S10 across 15 capillaries. Fluorescence changes in response to the local temperature gradient were recorded and compared to a reference measurement. If binding occurs, the interaction alters the thermophoretic mobility of the labelled protein, producing

a concentration-dependent signal. These data can be fitted to generate a binding curve and derive quantitative affinity parameters (Figure 26).

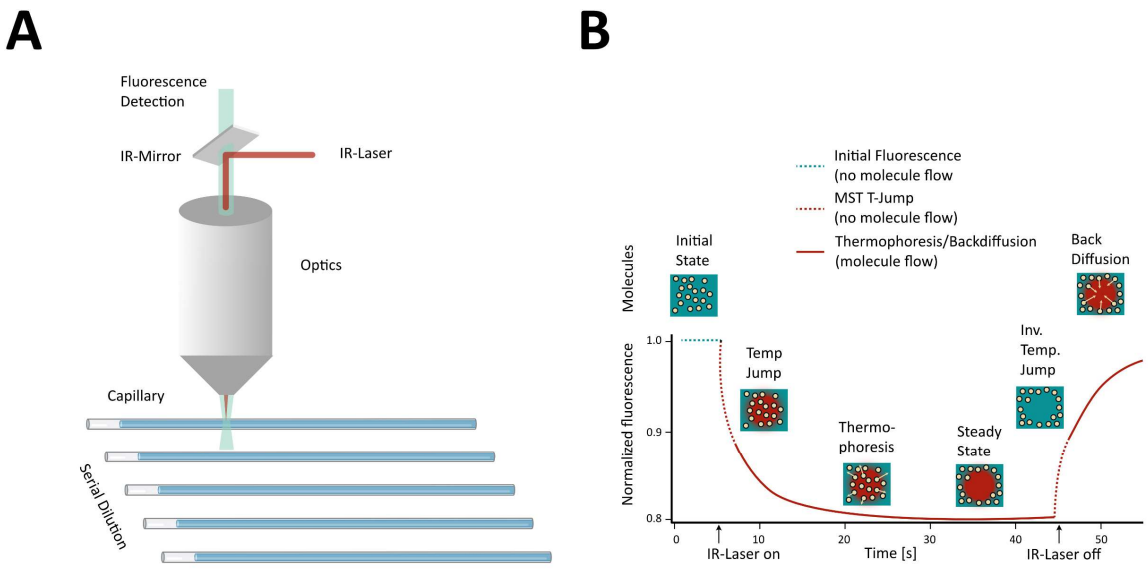


Figure 27. Schematic representation of experimental setup and MST signal. A. Microscale Thermophoresis (MST) is performed in a capillary containing a four µL sample. Fluorescence inside the capillary is both excited and detected using the same optical pathway. Localised heating of the sample is induced by an infrared (IR) laser in the fluorescence detection zone. The resulting temperature jump (T-Jump) and thermophoretic movement of molecules are detected as fluorescence changes occurring on different time scales. **B.** The panel on the right illustrates a representative MST signal. At the beginning, molecules are evenly distributed, producing a stable "initial fluorescence." Once the IR laser is activated, a rapid T-Jump occurs, followed by thermophoretic migration of molecules. A decrease in fluorescence is recorded over ~30 s. After switching off the IR laser, an inverse T-Jump occurs, and molecules redistribute through pure mass diffusion ("backdiffusion"), providing insights into molecular size**.** The figure was taken from *M. Jerabek-Willemsen et al., 2011*[161]

3.2.13 Western Blotting

Western Blotting is an essential technique for the detection of a specific protein in a mixture of proteins. It exploits the high specificity of a monoclonal antibody against the protein of interest and a labelled secondary antibody for detection, to give a sensitive result. The technique consists of three main steps: (1) protein separation, (2) protein transfer and (3) protein detection.

Initially, the protein mixture is denatured in loading dye and run onto an SDS-polyacrylamide gel. In this study, a 16% polyacrylamide concentration was mainly used due to the reduced size of the proteins of interest. The gel is run for approximately 40 minutes at 180 V.

In the next step, the proteins from the gel are transferred to a solid carrier. Here, Immobilon®-PSQ PVDF membrane (Merck) was used. To achieve this passage, the so-called “sandwich” has to be assembled. The components include two sponges, four Whatman filter papers, the membrane and the gel, which are equilibrated in WB-transfer buffer for 15 minutes before the assembly. The sandwich is then assembled as shown in Figure 27 and placed in a chamber. The transfer was carried out overnight at 4°C at 30V/90mA or at room temperature for one hour 90V/350 mA with constant voltage.

After the transfer, the sandwich is disassembled, and the membrane is collected and blocked by incubation in WB-Blocking solution for at least one hour at room temperature. The membrane is then washed in WB-TBST and incubated in blocking solution supplemented with a rabbit primary antibody in a 1:1000 volume ratio for four hours or overnight. The membrane is then washed with WB-TBST and incubated again in WB-Blocking solution and secondary antibody. Here, a chicken anti-rabbit IgG (H+L) secondary antibody, Alexa Fluor™ 647 (Invitrogen) in a 1:10000 volume ratio was used. The incubation with the secondary antibody was carried out for 1 hour, shaking at room temperature. Afterwards, the membrane is washed in WB-TBST Buffer three times, and the fluorescence of the secondary antibody is detected using a ChemiDoc™MP Gel Imaging System (BIORAD).

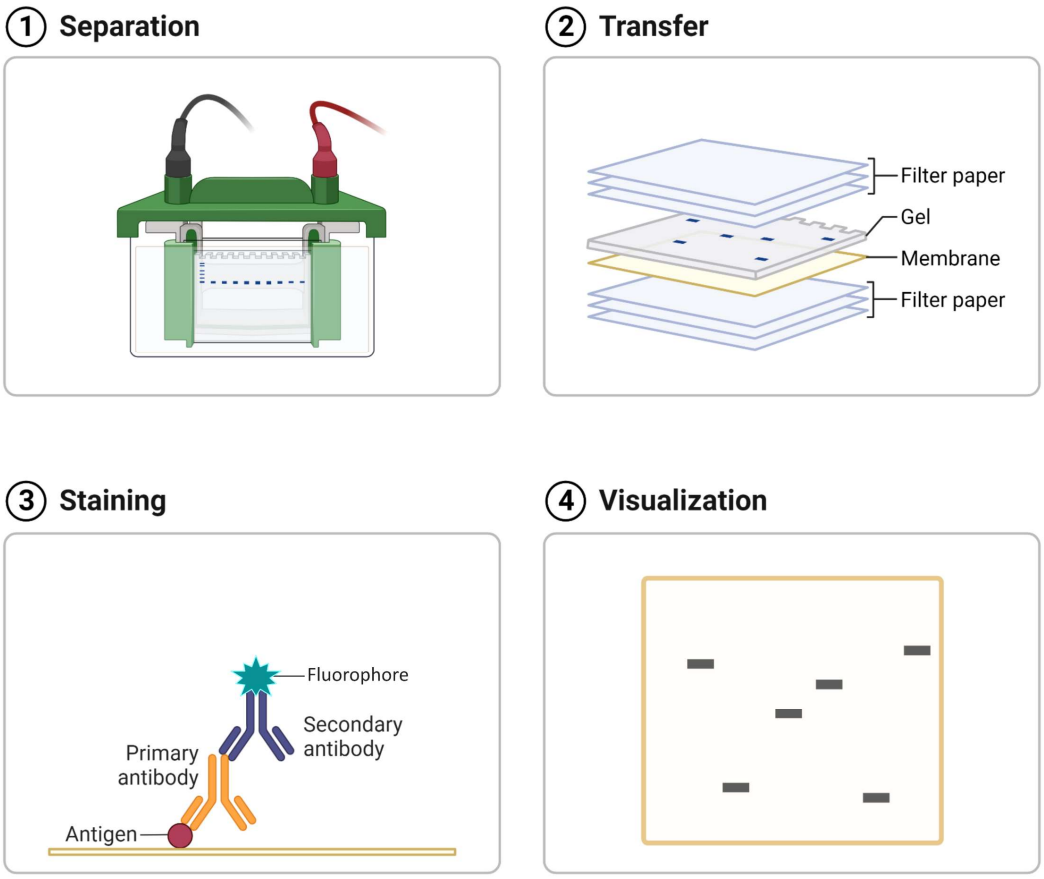


Figure 28. Schematic depiction of the four different steps in a Western Blotting experiment. The first step consists in protein separation through gel electrophoresis. Secondly, the so-called western blot sandwich is assembled, and the protein is transferred onto the membrane. Subsequently, the membrane is incubated with primary antibody targeting the protein of interest and with the labelled secondary antibody targeting the primary antibody. Finally, visualisation of signal is determined by detecting the label on the secondary antibody. Image taken from BioRender.

3.2.14 Negative Staining

3.2.14.1 *Grid preparation*

For negative staining, empty copper grids coated with continuous carbon film were used. The first step in preparing the grids is to spray a thin layer (10-15 nm) of carbon onto mica sheets. Mica is a class of silicates[162] with a perfect basal cleavage, conferring a thin and flat support for carbon sedimentation. The advantage of using mica is that the carbon will take the shape of the flat mica surface which is ideal for EM-grid support. Additionally, the space between two sheets of mica will be sterile. For these reasons, a mica sheet was freshly divided into two pieces using a razor blade just before introducing it in the carbon evaporator machine. The freshly divided mica sheets are then placed in the 208 carbon High Vacuum Carbon Coater (Cressington Scientific Instruments) together with two coal rods. To evaporate carbon, the small coal rod is first heated up by applying current with the shield on, eliminating contaminants or other gases from the coal rod. This passage is repeated at least three times. Afterwards, the carbon evaporation can start. The carbon is evaporated by applying high voltage

for 9.9 seconds through the coal rod. Usually, one round of coating leads to a 5-6 nm thick carbon layer; therefore, this step has to be repeated once or twice. Subsequently, the carbon-coated mica sheets are stored at room temperature and can be used at least two months after generation.

G2400C (Plano GmbH) Copper grids are used for negative staining analysis. These copper grids have 400 empty squares. 40 to 50 grids are glow discharged and placed on a piece of filter paper placed under water in a chamber**.** After all the grids are underwater, the carbon-coated mica is then submerged from the slide of the chamber. Finally, the water level is decreased, and the carbon layer is deposited on the grids. To improve the carbon coating, the freshly prepared grids are incubated at 60°C for approximately two hours. Afterwards, the grids are ready to use or can be stored at room temperature.

3.2.14.2 *Sample preparation*

The apo *Pfu* RNAP and transcription elongation complexes were first assembled as described in Chapter 3.2.8 and run on a Superose 6 Increase 3.2/300 with an isocratic elution in *Pfu* RNAP in cryo-EM Buffer. The input of polymerase was half the amount (50 µg) with respect to the cryo-EM because the continuous carbon increases the adherence of the specimen onto the grid, consequently the amount of input sample had to be reduced to avoid overcrowding. 3-5 µL of the sample are applied onto a freshly glow-discharged grid and incubated for 1 minute. The excess of liquid is then blotted with a Whatman paper filter, and the grid is washed in double-distilled water. The excess of water is again removed with filter paper, and the 1%-2% uranyl formate (5 µL) solution is applied to the grid, incubated for 25 seconds, and blotted with Whatman filter paper. This procedure is repeated three times. The grids are then air-dried and stored at room temperature until data acquisition.

3.2.14.3 *Data collection and processing*

Micrographs are acquired at the inhouse 200 kV FEG transmission electron microscope (Jeol), using an F416 CMOS camera. The software used for data acquisition is SerialEM[129] and it was programmed for automatic acquisition. Dataset sizes ranged between 40 to 168 micrographs per sample (see Results sections 2.3, 2.4 and 2.5). The magnification used for the data acquisitions was 80-fold for creating a montage to have an overview of the grid conditions concerning carbon film integrity and bulky contamination. Next, mid-maps are acquired at 1,500-fold magnification which can fit one of the 400 squares of the grid used to have an overview of staining intensity and the choice of the acquisition points. The final dataset was acquired at 40,000x magnification with a pixel size of 2.7 nm. At this magnification the protein sample appears bright against the stained dark grey background.

Processing of negative staining datasets consists in importing the micrographs to RELION[130,135], and running: CTF estimation, Particle picking, 2D classification, Initial Model, and 3D classification. The algorithms are described in detail in the next chapter.

3.2.15 Cryo-Electron Microscopy (Cryo-EM)

3.2.15.1 Sample preparation

The first step in cryo electron microscopy experiments is to examine protein homogeneity and composition within the sample using biochemical approaches. The goal is to avoid contaminating proteins or compounds and precipitated adducts that might interfere with complex stability as this will later slow down computational evaluation. Such biochemical approaches comprise evaluation of purity through SDS-PAGE and mass spectrometry. The latter will identify the nature of the contaminant. Concerning aggregate formation, this can be easily evaluated through size exclusion chromatography, using a gel filtration column. Complex formation can be confirmed for example by native-PAGE analysis or size exclusion chromatography. Another approach recently developed for investigating protein purity, aggregation and complex formation is mass photometry (see Chapter 3.2.11). In this study, RNAP complexes were prepared using an excess of binding partner (Chapter 3.2.8). The complex was, then, separated from the excess of the binding partner by size exclusion chromatography using Superose 6 Increase 3.2/300 GL column in a micro Äkta purification system.

3.2.15.2 Vitrification

Vitrification is the process through which water transforms from liquid to an amorphous state without inducing ice crystals nucleation[163]. This process depends on temperature and pressure applied to an aqueous solution. To ensure reproducibility and ease of handling of the procedure, three automated system machines were developed: Vitrobot (Thermo Fisher Scientific), EM GP (Leica) and Cryoplunge 3 (Gatan). For this study, the Vitrobot Mark IV from Thermo Fisher Scientific was used.

The first step in the procedure is choosing an adequate grid type. Cryo-EM grids are characterised by a metallic carrier which can be made up by copper, gold or nickel and a holey mesh. Concerning holey meshes, there are mainly kinds in use: Quantifoil and Lacey. The first one has a distinct regular pattern, presenting same size of equidistant holes all over the mesh, whereas lacey films are characterised by irregular-sized holey meshes. The cheapest and most commonly used combination in cryo-EM and single particle analysis are Quantifoil carbon films combined with copper carrier. To increase adherence to the holey, one additional ultra-thin layer of carbon or graphene oxide can be applied on top of it.

Before applying the sample on the grid, the latter needs to be glow discharged. This step is important for cleaning the surface of the grid from contaminants, but also for temporarily rendering the surface of the grid hydrophilic. This step is achieved by exposing the surface containing the film to plasma. The exposure of the grids to plasma depends on the nature of the film.

Sample application occurs in an isolated chamber of the vitrifying machine which has to be set to low temperature and high humidity. Since humidity and temperature are inversely proportional, the perfect conditions for the chamber are 4°C and 100% humidity. This passage is important for avoiding evaporation once the excess of sample liquid has been blotted away. The goal is to obtain a very thin layer of proteinaceous solution and evaporation of such thin layer occurs at fast rates unless the humidity is set at high percentage.

In the blotting step, the excess of liquid is drained away. There are different methods of blotting, each aiming to achieve different ice distributions. For example, the Vitrobot is equipped with two curved blotters, which apply pressure on the grid from behind and front of the sample creating a gradient ice thickness. On the other hand, EM GP system is equipped only with one blotter which can either blot from the front or from the back of the grid (back-blotting) resulting in a uniform ice thickness layer formation on top of the grid. This technique is useful for filamentous samples because it increases the concentration on the front part by repeating sample application and back-blotting for few times. There are few parameters to take in consideration that will interfere with the ice distribution on the grid, namely: waiting time, blotting force, blotting time, and blotting paper type. The blotting time is the incubation of the sample to the grid just after application. This step is not necessary, but it can be useful to allow the sample to migrate towards the glow discharged film. Nevertheless, it has also been shown to increase the formation of artifacts due to air-water interface effects[164]. Blotting force is the amount of pressure being exerted on the grid and blotting time is the amount of time (usually seconds) that the blotters exert this force on the grid. Filter paper types also influence the distribution of the ice thickness, depending on their porosity, they might have a different capillarity influencing the rate at which the excess of sample is being taken away.

Finally, the grid containing a thin layers of sample solution is being plunged in a cryogenic solution. The most often used cryogenic solution is liquid ethane maintained at melting temperature (approximately -182.8°C) by liquid nitrogen (approximately -196°C). This is the passage where the sample is being vitrified and from this point on it must be handled at temperatures below -137°C to prevent devitrification of the ice layer.

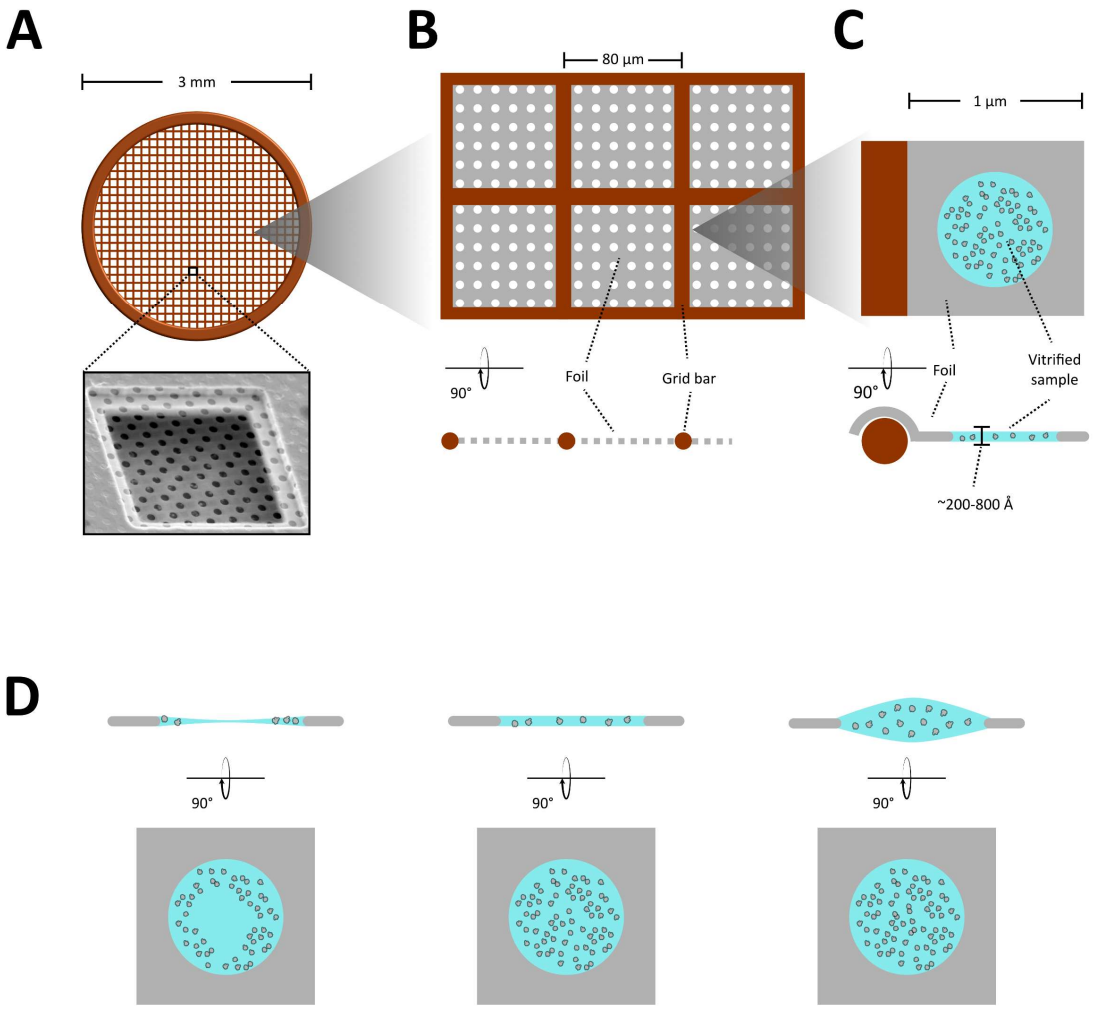


Figure 29. Schematic representation of a TEM Quantifoil grid and vitrification examples. A. Drawing of a copper holey Quantifoil grid with a scanning EM micrograph showing the equidistant hole pattern of its carbon support (picture taken from Fisher Scientific website https://www.fishersci.com/shop/products/quantifoil-r-2-2-400-au-1/501902569). **B.** Drawing of six squares zoom in seen from the front and side view. **C.** Drawings of a single hole zoom in containing the ideal distribution of particles and vitrified water thickness from the front view and from the side view. **D.** Three examples of sample vitrification showing one optimal situation (middle) and two effects of rather too thin vitreous water layer (left) or too thick vitreous water layer (right).

3.2.15.3 Electron microscope and data acquisition

Electron microscope anatomy

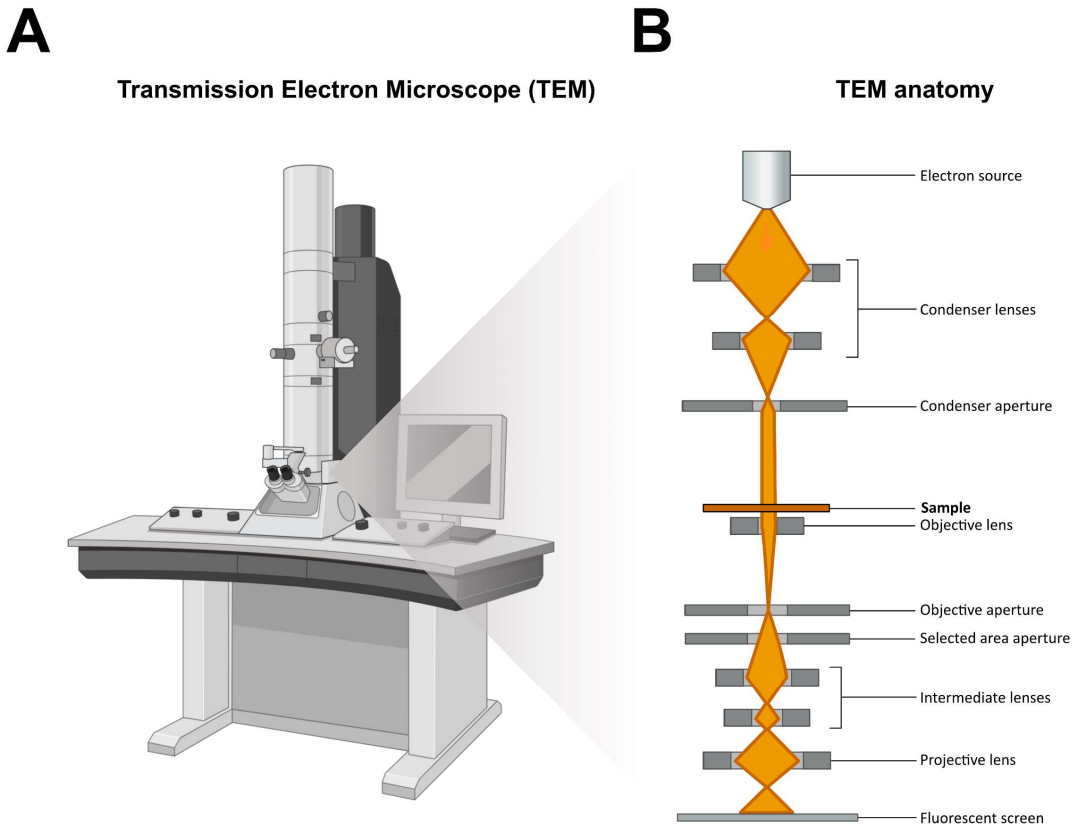


Figure 30. Transmission electron microscope anatomy and electron beam path. A. Image of a transmission electron microscope (TEM). **B.** Scheme of the TEM components and the electron beam path. Figures taken and modified from BioRender.

Electron gun

Electron microscopes employ various types of electron guns to generate and focus electron beams for imaging specimens at extremely high resolution. The most common types include thermionic emission guns, field emission guns, and cold field emission guns. Thermionic emission guns work by heating a filament to emit electrons, offering a relatively simple and cost-effective solution[165]. However, they produce relatively low current densities and require frequent filament replacement. In addition to thermionic emission guns, field emission guns, and cold field emission guns[166], another type of electron gun used in electron microscopes is the LaB_6/CeB_6 (Lanthanum hexaboride or Cerium hexaboride) electron source. LaB6[167] and CeB6[168] electron guns offer high brightness and stable emission, making them suitable for high-resolution imaging applications. They provide excellent spatial coherence and beam stability, contributing to sharp and detailed imaging. However, LaB6 sources are sensitive to contamination and degradation over time, requiring careful maintenance and occasional replacement. Despite this drawback, their superior performance in terms of brightness and stability makes them a

preferred choice for many advanced electron microscopy applications. Field emission guns[169] utilize high electric fields to extract electrons from a (sharpened) Tungsten tip, providing exceptionally high brightness and resolution. Key importance: small spot size; small delta E/E. They are complex and expensive to manufacture and operate, and they require stringent vacuum conditions. Cold field emission guns combine the benefits of field emission guns with lower power consumption and longer lifespan; in addition, as being operated at room temperature (rather than at 1700 K, as Schottky FEG's), they have the lowest = best delta E/E values[166,170].

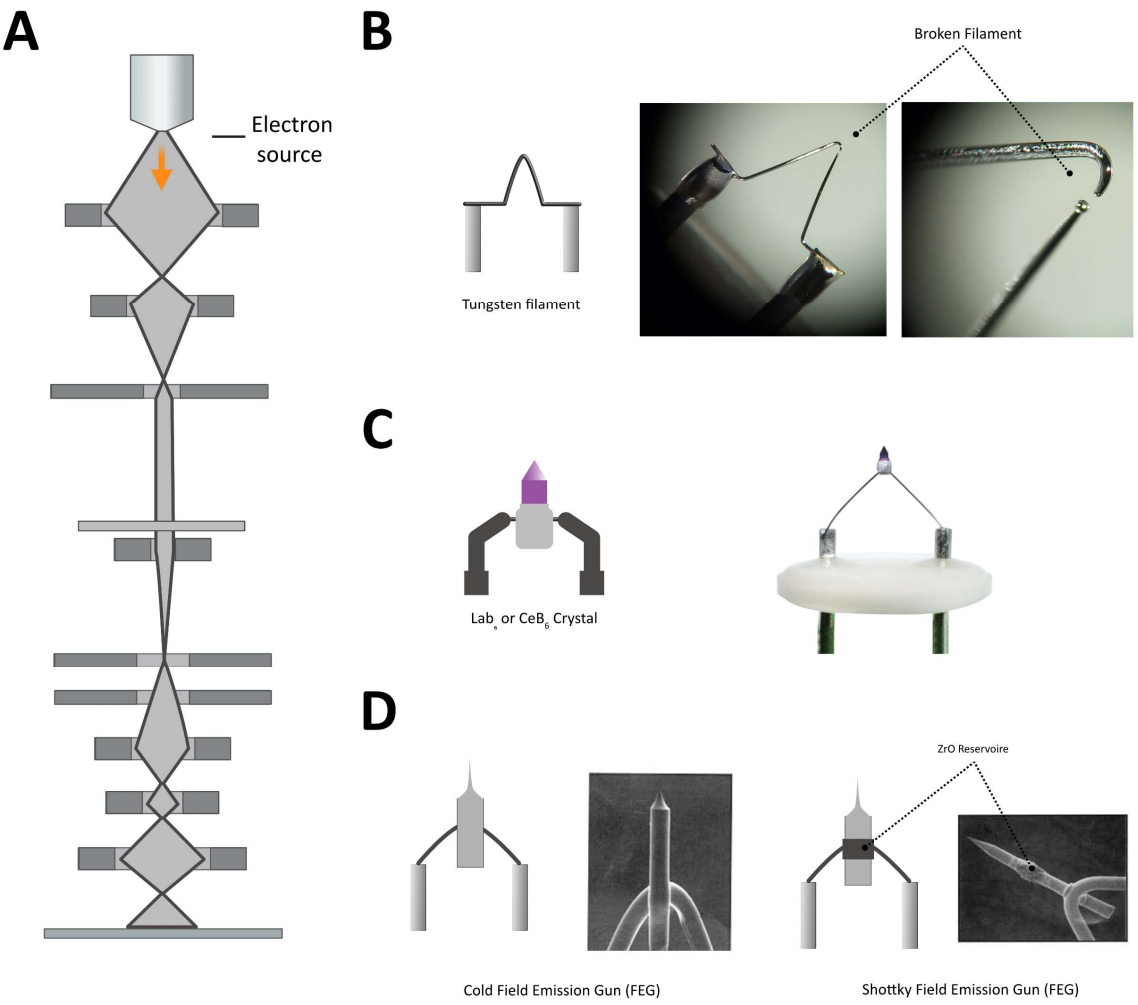


Figure 31. Transmission electron microscope electron sources. A. Depiction showing the electron gun position in the microscope. **B.** Example of a tungsten filament electron source and two pictures of a broken filament. **C.** LaB_6 and CeB_6 crystal electron guns drawing, and picture taken from EMSIS ASIA online shop https://emsisasia.com/?s=lab6. **D.** Examples of Field Emission Gun (FEG) and Shottky Field Emission Gun. The main difference between the two is that Shottky FEGs have a ZrO reservoir meant to improve its performance at high temperatures of operation. SEM pictures take from https://joachimfranklab.org/

Stage

The stage is the part of the microscope where the specimen is inserted. In the case of Transmission Electron Microscopes, the sample is introduced in the stage by grid holders. Older generations of microscopes are equipped with traditional side-entry holders which allow the introduction of one grid at the time. On the other hand, high resolution cryo-microscopes are equipped with an auto-loading

system which can store up to 12 samples. The cryo-grids are sealed to cartridges by means of C-clips and the cartridges are assembled in a cassette. Once the cassette is ready, it can be inserted in the microscope.

Electron detectors

Electron detection has significantly improved when compared to the first prototypes of transmission electron microscopes. The first kind of electron micrographs were printed onto films[171]. The digitalisation of data acquisition occurred with the introduction of charge-coupled device (CCD) cameras. CCDs are able to image through optics fibres or lens on top of which the energy of the primary electron in a phosphor or scintillator is deposited. This indirect detection's major drawback is that resolution is reduced while light is subjected to multiple scattering events while travelling through the phosphor or optical interfaces[172,173].

The new state-of-art cryo-electron microscopes used for high resolution data acquisition are equipped with direct electron detectors. Direct detectors are able to detect incident electrons and transform them in signal without photon conversion. Direct detectors allow a higher sensitivity and better signal-to-noise ratio compared to the film or CCD cameras detection. Additionally, the speed of acquisition of direct detectors is so high that instead of single micrographs, a multitude of frames per second can be acquired in a movie format. Different acquisition modes were also designed for better signal-to-noise ratio on the direct detectors. The three modalities of acquisition are integrative, counting, and super-resolution mode. Integration mode involves integrating the signal generated by incident electrons over the exposure time. Instead of counting individual events, it rather accumulates the total signal after a certain exposure period. In counting mode, the direct detector operates by counting each electron event as they hit the detector's surface. Super-resolution mode, on the other hand, leverages sub-pixel localization of electron impacts to achieve finer spatial resolution than the physical pixel size of the detector, effectively increasing resolution.

3.2.15.4 Data processing

Due to bad signal-to-noise ratio in cryo-EM that investigates protein mixtures, an articulated image processing is fundamental for reconstructing a detailed coulombic density map. Electrons passing through protein atoms (Carbon, Hydrogen, Nitrogen, Oxygen, Sulphur) mostly do not scatter, therefore the signal is limited. For this reason, the defocus values cover a range in order to increase the contrast of proteinaceous solutions and their visibility. Nevertheless, changes in defocus values also result in loss of high frequency information corresponding to details of amino acids side chain positions, for example. There are two main software packages in which analyses of cryo-EM data are performed: CryoSPARC[133] and RELION[130,135]. The two software are compiled to function on Linux-based operating systems, and they include a multitude of functions and algorithms. The process begins with motion correction using tools like MotionCor2 or RELION's internal implementation, which aligns frames in raw micrograph datasets to compensate for beam-induced motion. Next, CTF estimation is performed to account for the contrast transfer function of the microscope, a critical step for accurate downstream processing using two main algorithms gCTF[136] or CTFFIND[174]. The two algorithms are meant for

calculation of movie-averaged power spectra. The main differences consist in executability (gCTF runs on GPU and CTFFIND runs on CPU) and licenses (CTFFIND is open source). Particles are then picked, either manually, semi-automatically, or through machine-learning-based methods integrated into RELION, followed by extraction to crop individual particles from the micrographs. These particles undergo 2D classification, which groups them based on similar orientations and features, helping to filter out noise, contaminants, or particles of poor-quality.

The remaining high-quality particles are used for 3D classification, where RELION allows for the sorting of particles into different structural conformations or states. This step is crucial for handling flexible or heterogeneous samples. The selected particles are further refined through 3D auto-refinement, which aligns particles iteratively and enhances the resolution of the reconstruction. RELION also supports local refinement to improve specific regions of interest within a structure. After refinement, post-processing is applied to sharpen the density map, often using B-factor corrections, and to mask the volume for accurate resolution estimation by Fourier shell correlation (FSC). Additionally, a mask in RELION which creates an artificial envelope around our 3D density of interest can be applied to post-processing or 3D auto-refine, to restrict the analysis in the defined box to what is inside the defined mask. 3D auto-refine using a mask is performed usually as the very last step of the processing if at all due to a high probability to introduce a bias at this point. However, for post-processing it will be mandatory to insert a mask. Since this process is performed as one of the last steps the resolution of the refined map will be high, therefore the mask will have to be lowpass filtered to 15-20 Å.

To further enhance the quality of cryo-EM data and enable higher-resolution reconstructions, it is strongly advised to perform Bayesian polishing and CTF refinement. Bayesian Polishing refines the motion correction process by estimating beam-induced motion trajectories for individual particles, going beyond the global corrections applied during initial processing. Using Bayesian inference, it optimizes frame weights to account for the effects of radiation damage and noise, thereby improving the signal-to-noise ratio and enhancing particle image quality. Complementing this, CTF refinement improves the accuracy of the optical parameters describing the microscope's influence on image contrast. It estimates defocus values for individual particles instead of assuming uniform defocus across micrographs, accounts for beam-tilt misalignments, corrects higher-order aberrations, and adjusts for anisotropic magnification variations. Together, these steps refine particle alignments and correct optical artifacts, pushing the resolution limits of cryo-EM reconstructions and ensuring that even the finest structural details are accurately resolved.

3.2.16 Model Building and Validation

At this stage of the process, the atomic model can be generated by placing the amino acid residues in 3D space following the map densities. Model building can be performed on maps of a resolution better than 4 Å. In this study, this process was started with the map corresponding to the *Pfu* transcription elongation complex sample. The final resolution of this sample was 3 Å resolution. Initially, homology models of all subunits were generated using the online tool SWISS-MODEL (Expasy). This tool relies on existing structures and generates a template-based homology model. In our case, the template chosen was the RNA polymerase in complex with TFEα and promoter DNA from *Thermococcus kodakarensis*[44]

(PDB: 6KF9). The generated subunits were aligned in 3D space using the MatchMaker option with default parameters in UCSF Chimera to the *tko* complex 6KF9 and then grouped to a single pdb file using the Copy/Combine function in the same software.

The generated model was then fitted in the refined final map and transferred to winCoot[175] for model building. In winCoot, the coordinates were first rigidly fitted to the refined and sharpened density maps. In the following step, the amino acid residues side chains were assigned to the corresponding densities with the Real Space Refine Zone function, using Planar Peptide Restraints, Trans Peptide Restraints, Main Chain Restraints, and a refinement weight of 60.00. The refinement map was switched from time to time based on the local resolution. After all subunits were assigned to the densities, the DNA and RNA were also built following a bacterial TEC scaffold[60] which was then mutated to the sequence of the short TS, NTS and RNA.

Subsequently, the Geometry Analysis plot in the "Validate" option was checked and geometry outliers were fixed. The same was fulfilled for the Ramachandran outliers, but this time turning on Ramachandran Restraints. The Ramachandran outliers were fixed only for the amino acid residues side chains where the density was poor, whereas for the regions of the map displaying a clear side chain density, it was left as outlier. To further fix severe clashes or low probability side chain conformations, the coordinates and maps were transferred to ISOLDE[176], a model building tool exploiting molecular dynamics algorithms. This tool is available as an extension of UCSF ChimeraX program[63]. This was followed by five cycles Refinement in the PHENIX software[177]. The success was evaluated by MolProbity[178] results list, an option incorporated in the PHENIX software. In the first runs, the results were presenting several outliers which were fixed in the input coordinate file and refined again until the outliers were reduced to minimum amounts.

The remaining four atomic structures were built according to the *Pfu* TEC 3D model. The TEC coordinates were fitted with or without the nucleic acid scaffold to the other maps. The *Pfu* Spt4/5 3D model was first rigidly fit from PDB: 3QQC[179] which belongs to the *Pfu* RNAP clamp domain fragment in complex with Spt5 NGN domain and Spt4. The super-contracted conformation, contracted conformation, and Transcription Elongation Complex associated to Spt4/5 were fitted in winCoot and ISOLDE. While on the other hand, the expanded *Pfu* RNAP conformation homology model was generated again with SWISS-MODEL using as a template PDB: 4QIW which is a *tko* RNAP structure with an expanded conformation. In addition to SWISS-MODEL also AlphaFold 2[138] was run from Google Colab for better understanding of the clamp domain architecture.

The five final atomic models' statistical analysis is listed in Tables 8, 9 and 10 in the Appendix chapters 5.1, 5.2, and 5.3 respectively.

3.2.17 *In vitro* Transcription Assays

TEC was assembly was done similarly as described in Paragraph 3.2.8 in the Methods concerning the ratio between the nucleic acids and RNA polymerase and incubation temperature and time. Monitoring of RNA synthesis was possible by incorporation of a small amount of radiolabelled UTP (α-P32-UTP, 1 µC) to the transcription mixture. The activity tests were carried out using TS-Long and NTS-

Long DNA oligonucleotides, and the T40-Short RNA oligonucleotide (Table 12 from methods Chapter 3.2.8). Initially, the RNA was annealed to the DNA template strand in a 9:1 molar ratio (final concentration of 40 mM DNA-RNA hybrid) by incubation at 95°C for 3 minutes and cooled down slowly. The *Pfu* RNAP was then added to the pre-annealed TS-RNA together with EMSA Buffer 5x and DEPC water. Subsequently, the rest of the reaction components were added to the mixture, components comprising NTS, NTPs and radiolabelled UTP. In some samples, Spt4/5 (1:3 molar ratio, ad 1:5 molar ratio excess with respect to the *Pfu* RNAP) was also included to investigate whether is influences the transcription processivity. As a negative control, to assess that the final radioactive signal derives only from the transcription reaction, EDTA is always added to one sample. EDTA's mechanism of action is consisting in sequestering a variety of divalent cations such as Mg^{2+} which is fundamental for the polymerisation reaction. To decrease the unspecific binding of *Pfu* RNAP to nucleic acids, 0.5% N-Lauroylsarcosine was also mixed to the samples.

The transcription reaction was performed at 80°C for ten minutes in 30 µL reaction volume. The protein content was then denatured by adding 2x BPB and formamide solution and incubating at 95°C for ten minutes. 8 µl of each denatured samples were loaded onto a pre-heated 25% polyacrylamide denaturing midi-gel (7 M Urea, in 1x TBE Buffer). The pre-heating was achieved by running the empty gel at 100 V, 15-30 mA for 30 minutes to one hour. The samples were run for approximately one hour at 16 mA. After the run, the gel is carefully placed on a piece of whatman paper and wrapped in plastic foil. The gel is then dried for 40 minutes under vacuum and placed in the imaging box with the gel facing an exposure screen. The next day the exposure screen signal is measured at the GE Typhoon FLA 7000 with the Phosphoimaging program (50 µm).

For the second approach, instead of incorporating radioactive UTP fluorescence was used as a sensitive readout signal. This time, the TEC was assembled using a Cy5 labelled RNA (label linked to the 5' end of the RNA) that was part of the elongation scaffold. The overall TEC assembly protocol is the same as described in Chapter 3.2.8 with the difference that the RNA was incubated with the template strand in a molar ratio of 3:1 respectively. The premixed RNA-TS was annealed to 50 µg of *Pfu* RNAP (final concentration: 4.3 µM) with a molar ration of 3:1. Finally, the NTS was added to the mixture also this time with a molar ratio of 3:1 with respect to the *Pfu* RNAP. To remove the excess of elongation scaffold and dilute endogenous amounts of Spt4/5, a Nick column (GE Healthcare) was used, and fractions were collected manually. The presence of the TEC in the fractions was assessed by running a 4-15% gradient SDS-PAGE using Mini-PROTEAN® TGX™ Precast Protein gels (BIORAD). The concentration of the RNAP in these fractions was determined by absorption spectroscopy at a wavelength of 280 nm using NanoDrop™ One/OneC Microvolume UV-Vis spectrophotometer (Thermo Fisher Scientific) and the most concentrated fraction (0.6-0.7 mg/mL) was used for transcription assays. The transcription reactions were performed as a time course in the absence of Spt4/5, with Spt4/5 wild-type and the Spt4/5 mutants. The first reaction mixture contained 1 µl of TEC, 3 µl 5x EMSA buffer (200 mM HEPES/KOH pH 7.4, 300 mM NaCl, 2.5 mM $MgCl_2$ x 6 H_2O, 0.5 mM EDTA, 0.1 mg/ml BSA), and DEPC water to arrive to a final volume of 13 µl. 1.5 µl from the nucleotide mix (ATP/CTP 1 mM, UTP 0.02 mM stock solution) was added just before incubation at 80°C. The reaction was stopped through ion chelation by the addition of 0.5 µl EDTA (0.5 M stock solution) after 30 s, 1 min, 5 min, 10 min, and 30 min. The samples were denatured by the addition of 15 µl of 100% formamide buffer containing 0.05% bromophenol blue and incubation at 95°C for 10 min. The other two reactions were prepared

the same way with the difference that 1 µl of Spt4/5 wild-type or Spt4/5 mutant was added, corresponding to a final concentration of 1.7 µM. For each reaction, a control was also prepared which consisted of the same reaction mixture, but the nucleotides were omitted from these samples. The elongation product was evaluated by running a denaturing PAGE (with 8M Urea and 20% v/v acrylamide) at 50°C constant temperature and 25 W. The gel plates were 16 x 20 cm and 0.4 mm spacers were used. The gel was loaded with 8 µl of the denatured sample. Detection of the Cy5 fluorophore was carried out using the ChemiDoc™MP Gel Imaging System (BIORAD), and signal quantification was performed using the software ImageLab (BIORAD). To evaluate the effect of Spt4/5on elongation, the run-off product from the reaction without Spt4/5 from the 30 min sample was equalled to 100% band intensity in each gel.

4. Discussion and Conclusion

This study aimed to investigate transcription elongation in Archaea, with a particular focus on the structural aspects of the process and the recruitment and function of Spt4/5. The results presented here reveal the atomic structure of the *Pyrococcus furiosus* transcription machinery, providing new insights into how archaeal transcription is regulated and how it compares to the bacterial and eukaryotic systems. In the following discussion, these findings are considered in relation to the previous studies and their relevance to a broader understanding of transcription evolution.

4.1 Clamp motions and cleft contraction

When unbound to its substrate, the RNA polymerases can assume different conformations. Available structural data on RNA polymerases often show only one rigid conformation, especially in the case of X-ray crystallography[45,50,51,57,180]. Nevertheless, such enzyme function is characterised by drastic movements that are not easy to capture. The atomic structure of apo *Pfu* RNAP determined by cryo-em has allowed the reconstruction of three models representing the most challenging sample out of the three datasets collected. In its apo state, *Pfu* RNAP was captured in both contracted and expanded conformations. Additionally, flexibility analysis conducted in CryoSPARC[133] has revealed the existence of intermediate conformations. This demonstrated that *Pfu* RNAP continuously oscillates between the contracted and expanded conformations. Compared to crenarchaeal RNA polymerase, atomic models of *S. shibatae*[45] and *S. acidocaldarius*[47] have been solved in a contracted conformation with the two different techniques, X-ray crystallography and cryo electron microscopy, respectively. In contrast, the other euryarchaeal RNAP from *T. kodakarensis* was first solved as a crystal structure in an expanded conformation[180] and later in a contracted apo form through cryo-em[44].

The clamp domain is a crucial feature of all multi-subunit RNA polymerases, as evidenced by numerous single molecule FRET (smFRET) studies, structural evidence, and functional analyses[48,48,49,64,181,182]. Despite the information available, a comprehensive understanding of archaeal clamp domain dynamics remains unclear. In the case of crenarchaeal RNA polymerases, structural studies have predominantly revealed a contracted conformation[46,47]. This suggests a potentially less flexible mechanism compared to their euryarchaeal counterparts, whose RNAP structures are known to exhibit significant dynamics, oscillating between expanded and contracted states, considering structural data[44,50]. This inherent flexibility may play a crucial role in their functional adaptability during transcription.

This study provides new insights into the dynamics of the *Pfu* RNAP clamp domain by demonstrating that it exists in both open and closed conformations at high resolution. These findings fill the gaps in understanding the clamp dynamics, confirming previous investigations of smFRET conducted using a

recombinant *in vitro* reassembled *M. jannaschii* euryarchaeal RNAP. Furthermore, three-dimensional variability analysis from cryoSPARC has revealed the presence of intermediate conformations, suggesting that euryarchaeal RNA polymerase clamp domains fluctuate between open and closed states when not engaged with nucleic acids.

Comparatively, bacterial RNA polymerase exhibits an expanded conformation in its apo state, as demonstrated by smFRET studies, crystallographic, and cryo-EM structures[183,184]. This expanded state contrasts with observations in eukaryotic RNA polymerase II, which has only been captured in an expanded conformation in early crystal structures that lacked the stalk domain[57]. Additionally, studies on the dimeric structure of RNAP II from *Sus scrofa domesticus* indicate that the clamp region exhibits significant mobility, further complicating our understanding of its structural dynamics[185]. Structures of the transcribing bovine RNAP II showed that even in the TEC, a small subfraction of the RNAPs (10%) exhibited a mobile clamp in the absence of DSIF[30], the eukaryotic Spt4/5 homolog. Interestingly, RNAP sample preparation for crystallization was also performed without the stalk domain in the case of the archaeal RNAP from *Thermococcus kodakarensis*. The stalk was added once the crystals were formed[50]. In the resulting crystal structure of *Tko* RNAP, the RNAP adopts an expanded conformation that is highly similar to that of *Pfu* RNAP. In conclusion, eukaryotic apo RNA polymerase II appears unstable and highly flexible. Since archaeal RNA polymerase's closest homolog is eukaryotic RNAP II, the heterogeneity determined in *Pfu* apo RNA polymerase can explain the complexity of structural determination of apo RNAP II. In this study, it was also possible to observe the involvement of the stalk domain in clamp motions by establishing a network of interactions with the Rpo2 subunit in its expanded conformation. Hypothetically, the stalk domain function could also include the regulation of clamp mobility, without which the RNAP can be mostly found in an open clamp conformation[50,57].

In conclusion, although significant progress has been made in understanding the structural conformations of RNA polymerases, the unique dynamics of the clamp domain across different archaeal groups require further research. Exploring these differences could yield deeper insights into the evolutionary adaptations of RNA polymerases and their roles in transcription regulation.

4.2 Transcription elongation factor Spt4/5 binds to the RNA polymerase in absence of nucleic acids

In this study, the atomic structure of an archaeal RNA polymerase (RNAP) in its apo state, and in complex with the transcription elongation factor Spt4/5, is examined (Figure 34D). This structural insight provides a valuable comparison to the well-characterized interaction between bacterial RNA polymerase and the transcription factor NusG, which has been extensively documented in previous studies[89]. Unlike in bacteria, where NusG directly interacts with DNA-free RNA polymerase, eukaryotic transcription systems show a distinct behaviour. The eukaryotic counterpart of archaeal Spt4/5, the SPT4/5 complex or the DSIF factor, has not been observed to bind to RNA polymerase II (RNAP II) in the absence of newly synthesised RNA.

A possible explanation for this difference lies in the requirement for a nascent RNA strand in eukaryotes, which appears to be critical for the interaction between SPT4/5 and RNAP II. The newly

synthesized RNA interacts primarily with the KOW4 and KOW5 domains of SPT5, forming the major part of the interaction network around the RNA exit channel[29,30]. Only one KOW domain is present in prokaryotic systems, Spt5 and NusG, suggesting a divergence in the mechanism of action between prokaryotic and eukaryotic transcription elongation factors.

In eukaryotes, the KOW domains 4 and 5 of SPT5 are essential for stable binding to RNAP II, whereas in prokaryotes, the NusG-like Spt5 and NusG itself use their NGN domains for association with RNAP, and the single KOW domain is involved in ribosome and Rho/aCPSF1 termination factor recruitment[10,11,14,15,94,95]. This highlights a key evolutionary divergence in the function of these elongation factors.

Focusing on the archaeal transcription machinery, the residues A4 and Y42 of *Methanocaldococcus jannaschii* Spt5 are pivotal for interaction with RNAP[64]. These residues are conserved in *Pyrococcus furiosus*, and structural studies reveal a similar mode of interaction, where these residues fit into a hydrophobic pocket on the clamp domain of RNAP. Additionally, a hydrogen bond between Spt5 E78 and Rpo1N Q262 further stabilizes this interaction, a feature that has also been observed in crystal structures of this complex[179]. In conclusion, while the KOW4 and KOW5 domains of eukaryotic SPT5 play a primary role in mediating interactions within the RNAP II transcription elongation complex, prokaryotic Spt5 and NusG rely on the NGN domain as the primary player for their association with RNAP.

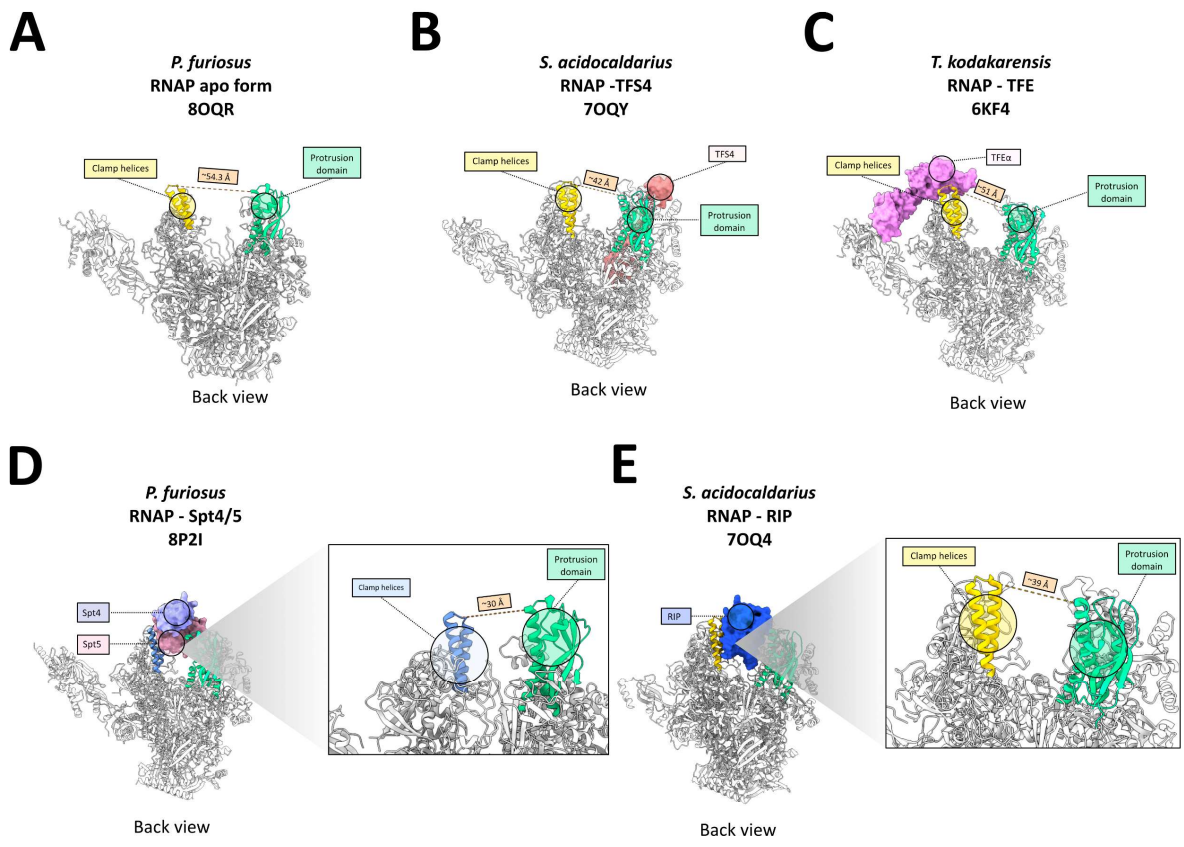


Figure 32. Conformation of the clamp domain across archaeal RNAPs and transcription complexes. A. In the apo state of Pfu RNAP, the clamp adopts an open clamp state with ~54.3 Å between Rpo1N-G259 in the coiled-coil motif on the clamp domain (yellow) and

Rpo2-R377 in the protrusion domain (green). B. *Sulfolobus acidocaldarius* RNAP in complex with TFS42 is characterized by a closed clamp domain with ~42 Å between residues Rpo1N-G257 and Rpo2-R373. C. *Thermococcus kodakarensis* TFEα shifts the clamp domain to an expanded state with a width between the coiled-coil motif (Rpo1N-A257) and protrusion (Rpo2-R380) of ~51Å. D. Spt4/5 binding to the Pfu RNAP in the absence of nucleic acids leads to a closed clamp conformation (blue). The distance between clamp (Rpo1N-G259) and protrusion domain (Rpo2 R377) is ~30 Å. E. The protein RIP2 is of viral origin and binds the RNAP from *Sulfolobus acidocaldarius*. Like Spt4/5, RIP binds inside the DNA-binding channel. However, the distance between clamp (Rpo1N G257) and protrusion (Rpo2-R373) differs by ~9 Å, with a more open clamp in the case of the RNAP-RIP complex.

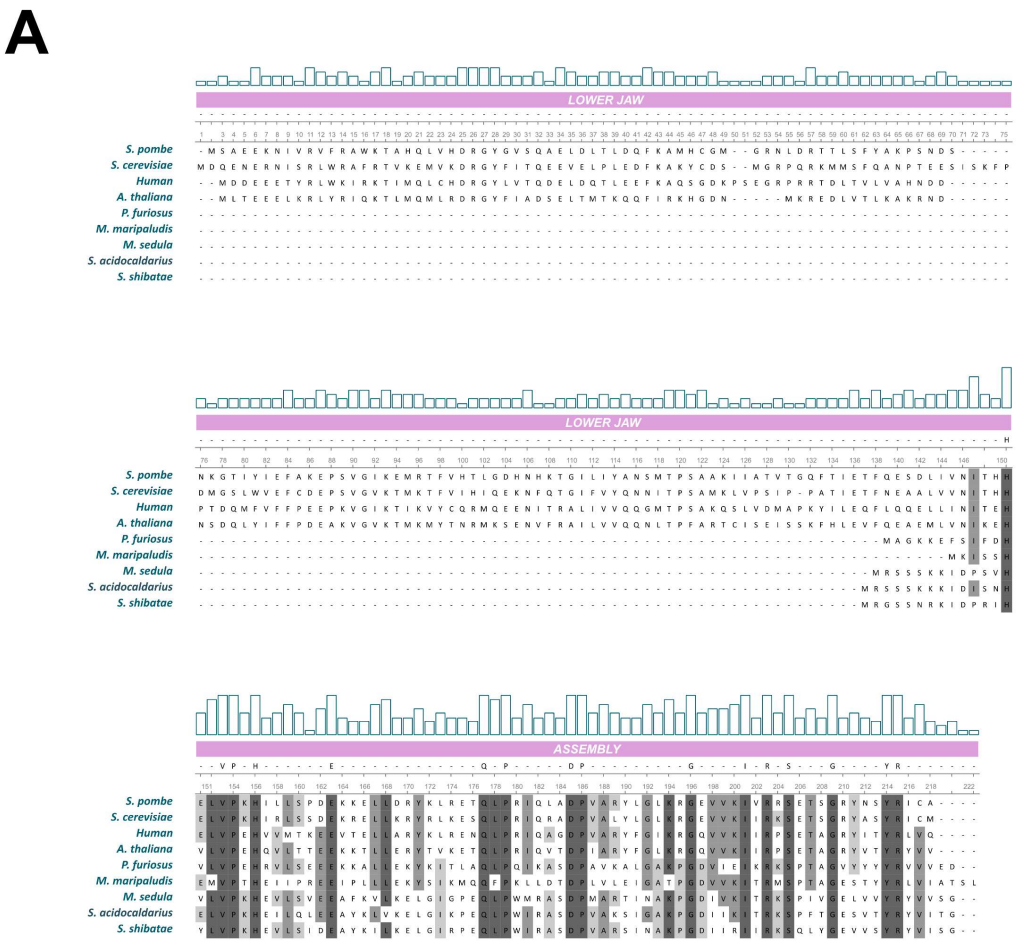
Likely, and in line with smFRET analyses *in vitro*, archaeal RNAP adopts various states in a dynamic equilibrium within the contracted and expanded boundaries described here. Spt4/5 apparently selects for a specific conformational state and can only bind contracted RNAPs (Figure 33D), resulting in super-contraction and removal from the conformational flexible pool of apo RNAPs. However, it cannot be ruled out that Spt4/5 binding induces the closure of the clamp and contraction of the RNAP instead. DNA loading into the binding channel requires an open state of the clamp domain, a transition supported by transcription factor TFE during the initiation phase, which was captured in structural analysis conducted on *Themococcus kodakarensis* RNAP (Figure 33C)[44,85]. When bound to TFEα *Tko* RNAP, the distance between the protrusion domain and the clamp domain is about 50 Å, which is approximately 4 Å less than the *Pfu* RNAP expanded conformation (Figure 33A). Hence, the RNA polymerase associated with Spt4/5 may represent a dormant form available in the cell that can readily be activated upon need. A similar binding behavior is observed in *Sulfolobus acidocaldarius* RNA polymerase inhibited by the RIP/ORF145 factor (Figure 33E). The latter binds between the clamp and protrusion domain, forcing the RNAP into a rigid contracted conformation and obstructing its DNA binding channel[47]. This mechanism of action is still unknown, and it contradicts the benefit of the virus reproduction since its genome does not encode for its own viral RNAP[186]. On the contrary, TFS4 locks the *Saci* RNAP in an open clamp state with a width between the protrusion and clamp domains of 42 Å, which inhibits the RNAP function (Figure 33B). In *P. furiosus*, a possible scenario how to activate *Pfu* RNAP for promoter-directed transcription initiation is the displacement of Spt4/5 by TFE from the clamp domain. A reverse scenario, such as the displacement of Spt4/5 by TFE to prepare RNAP for transcription initiation, could potentially activate this RNAP pool in conditions requiring increased transcriptional output or a shift in the transcriptional program. Supporting this notion, as previously demonstrated, TFE exhibits a higher affinity for RNAP within the context of the PIC compared to Spt4/5[48]. In conclusion, Spt4/5 is hypothesized to fixate the cleft in a contracted state, thereby inducing a reversible shutdown of *Pfu* RNAP activity.

4.3 Transcription elongation complex structure resembles eukaryotic RNAPII

The transcription elongation phase is characterised by the stage in which the RNA molecule is synthesized. In this thesis, it was demonstrated that *Pfu* RNAP associates to a synthetic DNA-RNA elongation scaffold and adopts a highly similar conformation to eukaryotic transcription elongation complexes[30,43,67,187,188] with a few differences. Firstly, the downstream DNA double helix is engaged in a web of interactions with the jaw domains[30]. The lower jaw domain of the eukaryotic RNA polymerase II interacts with the downstream DNA. Especially, the N-terminus of subunit RPB5, which is not

conserved throughout archaeal RNAPs, forms interactions with the DNA double strand. The sequence alignment in Figure 32 shows the difference in sequence and structural differences between archaeal Rpo5 and eukaryotic RBB5. Because archaeal Rpo5 is a smaller protein than RBP5, it is reasonable that contacts with the downstream DNA double helix are not occurring. Indeed, the missing domain is the one that, in eukaryotic RNAP II, forms the lower jaw domain interacting with the DNA[67] (Figure 32B).

The upper jaw domain also interacts with the downstream DNA-duplex in RNAP II TEC[30,67,187,188]. In the structures of elongating *Pfu* RNAP, this interaction is less evident or unable to be observed. In RNAP II the upper jaw is part of the largest subunit RPB1, while in *Pyrococcus furiosus* RNAP it is in the second part of the split largest subunit, Rpo1C. The structural flexibility of the upper jaw in *Pfu* RNAP appears to be quite pronounced, which adds a layer of complexity to its modelling. The assignment of the amino acid chain in this region was performed using rigid constraints, reflecting the challenge in modelling flexible domains. Despite this rigidity in the modelled structure, it cannot be fully ruled out that interactions between the upper jaw domain and downstream DNA may indeed occur during the transcription elongation process. This interaction could be dynamic and transient, potentially contributing to the elongation complex's function, even though it is not readily captured in the current 3D structural models.



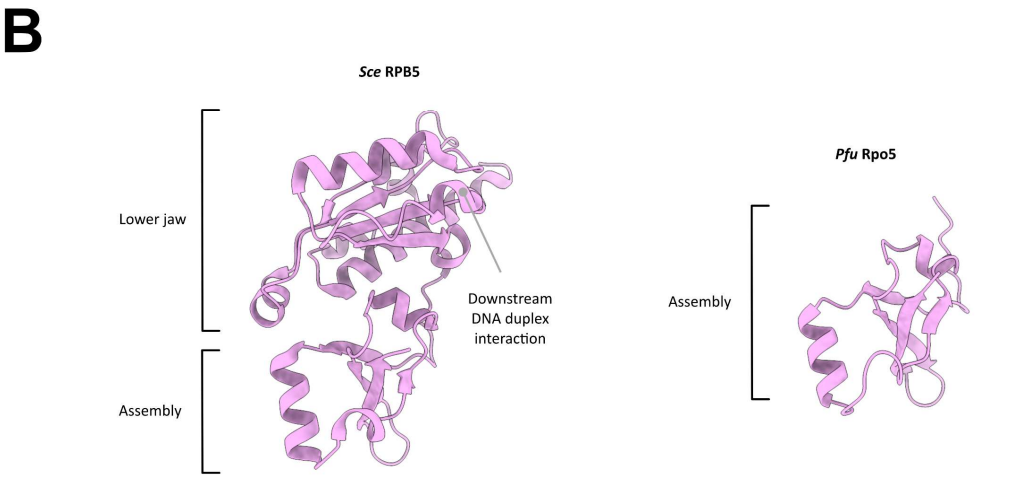


Figure 33. Comparison of RNA polymerase subunit Rpo5 /RPB5 from Archaea and Eukaryotes. A. Sequence alignment of eukaryotic RPB5 from *S. pombe*, *S. cerevisiae*, *H. sapiens* and *A. thaliana*, as well as archaeal Rpo5 from *P. furiosus*, *M. maripaludis*, *M. sedula*, *S. acidocaldarius* and *S. shibatae*. All eukaryotic RPB5 sequences show an additional domain of roughly 135 residues at their N-terminal end. B. Structural model of eukaryotic RPB5 from *S. cerevisiae* (PDB ID: 5C4X) on the left side with domain denomination. To the right, *P. furiosus* structural model of Rpo5 with domain denomination.

In multi-subunit RNA polymerase enzymes, the transcription elongation process has been described as involving numerous structural motifs within the enzyme. These motifs include fork loops, rudder

loop, lid loop, and zipper loop and their main task is to maintain the DNA strands separated during the entire process. The nucleotide incorporation is tightly regulated to maintain high fidelity. A flexible motif, the trigger loop, in the RNA polymerase active centre has been shown to interact with the incoming ribonucleotide and discriminate between the complementary and non-complementary NTPs[58]. In the *Pfu* TECs structures, no direct contact between the DNA template strand, the non-template strand, the RNA, and these motifs is visible. The trigger loop flexibility hindered full modelling of this region, and only the amino acidic main chain could be built, but side chain densities were assigned according to bond geometry rather than density tracking.

4.4 The Spt5-interface with upstream DNA in TECs is functionally important in *P. furiosus*

Archaeal Spt5 in complex with Spt4 and bacterial NusG function has been widely investigated. Throughout transcription progression, Spt5 and NusG are known to improve RNAP processivity and reduce pausing[49,89]. The NGN domain of both prokaryotic factors is found in all TEC structures bound between the clamp domain and protrusion domain, which secures the DNA binding channel and prevents the clamp domain from moving freely, thereby potentially preventing backtracking or termination events. The mechanism of action of eukaryotic SPT4/5 or DSIF differs from the prokaryotic counterpart, as the primary contact between elongation factor and RNAP II occurs at the level of the stalk domain and RNA exit channel[30]. Specifically, the linker domain between the KOW4 and KOW5 domains of SPT5/DSIF anchors to the newly synthesized RNA oligonucleotide and has been demonstrated to be crucial for SPT5 association with the TEC complex and stimulatory effects[31,189] (Figure 35C). Even though direct contact of SPT4/5 has not been observed in other eukaryotic polymerases, SPT5 seems to influence transcription rates in RNA polymerase I TEC *in vitro*[190].

Archaeal Spt5 only has one KOW domain[50], rendering it unlikely that this domain contributes to the stimulatory effect on transcription elongation[49]. In this study, the 3D structures confirm that the overall architecture and organization of archaeal TEC are conserved through the different domains of life. Nevertheless, DNA contacts may vary among species or transcription systems. In this study, it was impossible to reconstruct the KOW domain from *Pfu* Spt5. This may be due to the choice of scaffold, and the effect might vary in other situations where the emerging nascent RNA or the upstream DNA is longer.

Bacterial NusG has been reported to exhibit a dual effect on transcription progression. While *Escherichia coli* NusG stimulates RNAP processivity and reduces pausing events[89,191], in *Bacillus subtilis* it promotes pausing, especially in correspondence with the TTnTTT sequence on the template DNA strand[192]. In this study, the authors investigated the effect of *Bacillus subtillis* NusG on Mycobacterium tuberculosis TEC. They characterized a web of interactions occurring between NusG and the non-template DNA strand. The amino acids that were found to be involved in the pausing are W76, Y77, R80, N81, T86, K97, and Y11 (Figure 35A). This effect on pausing has also been observed in the pathogenic bacterium *Mycobacterium tuberculosis*[191]. In this study, four positively charged residues were identified that potentially form contacts with the upstream DNA duplex. These four amino acid

residues are not conserved throughout Archaea species. Mutation of H12, R64, H65, and R67 to alanine has resulted in a loss of processivity effect, indicating that they play a fundamental role in transcription bubble maintenance and upstream DNA duplex stabilization (Figure 35B).

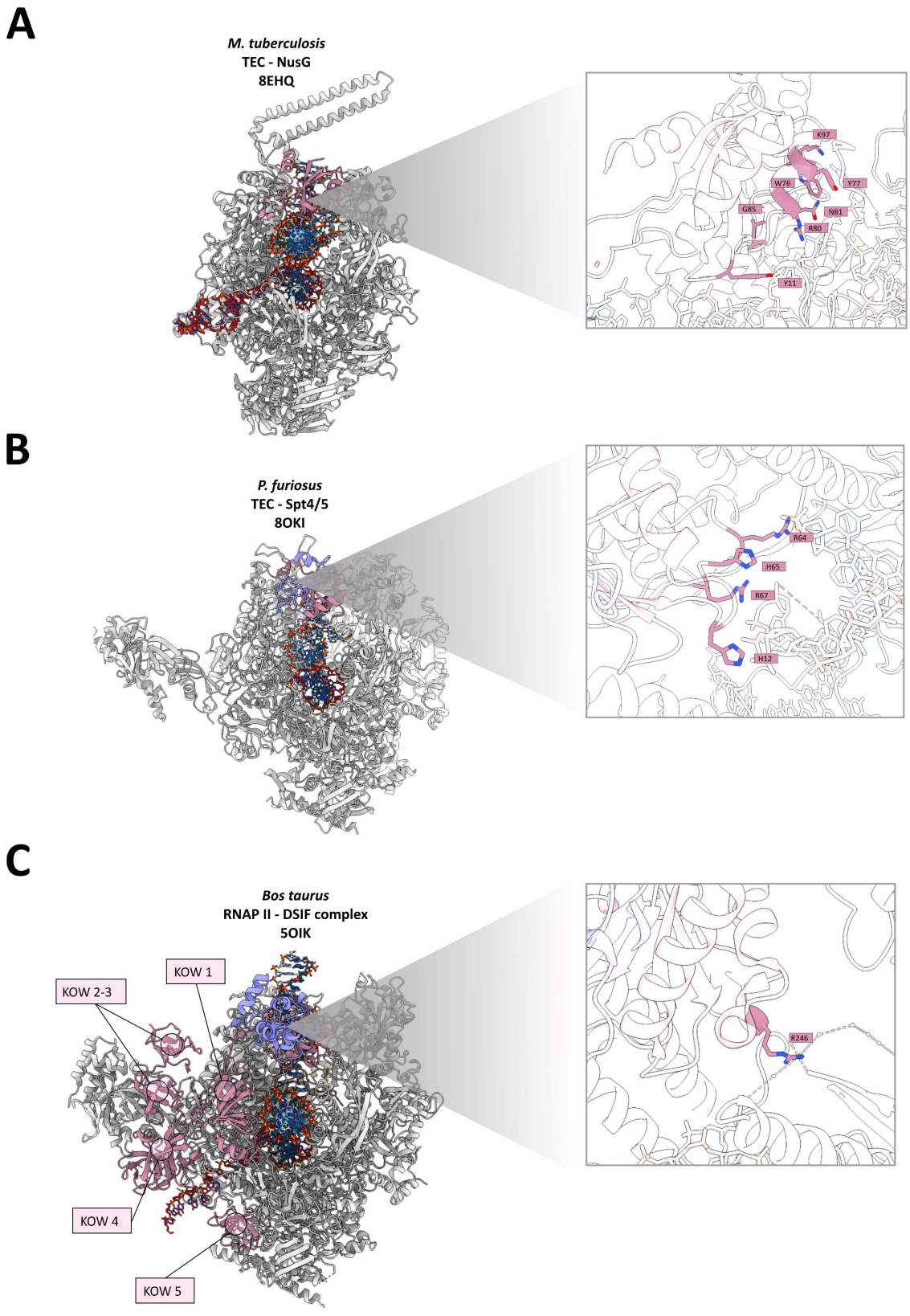


Figure 34. Transcription elongation complexes with NusG and Spt4/5 from the three domains of life. A. *Mycobacterium tuberculosis* transcription elongation complex structure (PDB: 8EHQ) with a close-up on Vishwakarma et al., an interaction interface involving W76, Y77, R80, N81, T86, K97, Y11, and the non-template strand was described in Bacteria[192]. **B.** Structure of human RNAP II TEC in complex with Spt4/5 homolog, DSIF. In eukaryotes, it has been described that Spt5-R246 is involved in transcriptional pausing[193]. **C.** In this study, the effect of R64, H65, R67, and H12 on transcription elongation in Archaea has been investigated.

TFE has also been shown to play a crucial role in DNA stabilization in the Initially Transcribing Complex (ITC) formation[44,48,90]. The stabilizing effect of transcription factors towards nucleic acid residues is far more important in hyperthermophilic and thermophilic Archaea given the elevated temperatures in which they thrive. Moreover, Spt5 is known to accompany the RNA polymerase throughout elongation and termination processes. The KOW domain of Spt5 has recently been demonstrated to bind the transcription termination factor aCPSF1 by cryo-EM[94] analysis and *in vitro* analyses[92,118]. The transcription termination rate of *Thermococcus kodakarensis* is indeed greater when Spt4/5 was added to the reaction. suggest a similar behavior of *Pfu* Spt4/5 involvement in both transcription termination and expressome formation.

5. Appendix

5.1 *P. furiosus* RNA polymerase apo enzyme data collection, processing, and model validation

100 µg of RNA polymerase was assembled on the synthetic DNA-RNA elongation scaffold as described in Chapter 3.2.8 and purified through size exclusion chromatography. The most concentrated fraction from the microÄkta run (0.622 mg/ml) was collected for further use. The sample was plunge-frozen with Vitrobot Mark IV (Thermofisher Scientific) onto gold grids (UltrAufoil R1.2/1.3; 300 mesh) which was freshly glow discharged with EasiGlow (TedPella) (1x 0.4 mbar; 15 mA; 30 sec).

The grids were then screened and imaged at the inhouse cryo electron microscope, CRYO ARM™ 200 (Jeol) equipped with Gatan K2 summit direct detector. A total of 8,212 movies in electron counting mode were collected at 50,000x magnification and a pixel size of 0.968 Å/pixel. The in-column energy filter was operated with a slit width of 20 eV. A total exposure of 40 e-/A2 was fractionated over 40 frames (1 e^-/pixel/s) at a defocus range from -0.6 µm to -2 µm using SerialEM software[129].

The 8,212 were imported and processed using RELION 4.0[130,135] software. The movie frames were aligned and dose weighted in RLION's implementation MotionCorr2[194] and Contrast Transfer Function (CTF) Estimation using GCTF[136]. Next, a subset of 6,482 micrographs were selected for further evaluation based on defocus values, astigmatism, or ice crystals contamination. Processing proceeded with particle picking, in a first step, Laplacian-of-Gaussian auto picking algorithm was used. The particles were then classified in 2D space and the best classes, representing different views of the specimen, were used for TOPAZ[195] training. The picked particles were then extracted with a box size of 260 Å and a binning factor of two. Afterwards, the particles were subjected to two rounds of 2D classification with a mask diameter of 180 Å in the final job. The classes presenting high resolution details were selected for 3D model generation. The Initial Model tool was run with four distinct classes with a starting number of 1,209,249 particles. One class of the initial 3D model was then used as a reference for several rounds of 3D classification. In the first round, two classes corresponding to 321,703 particles were discarded due to unknown density, probably due to erroneous picking and

overrepresentation of one view, yielding to misalignment to particles in 3D space. In a second 3D classification round, already two different conformations were distinguishable: expanded conformation and super-contracted conformation. At this point the set of particles was divided into two sets and analysed separately. 256,556 particles were discarded due to preferential view bias and unidentifiable densities around the main specimen's 3D density.

Two classes representing the super-contracted conformation containing 361,929 particles were classified once more, and 18,323 misaligned particles were discarded. The remaining particles were refined and unbinned, resulting in a 3.4 Å resolution density map. In this conformation, Spt4/5 density was present on the clamp domain; therefore, the extra density was used for a local mask generation. The mask was then imposed on the 3D map, and 3D classification without image alignment was run to identify the polymerase particles containing the Spt4/5 extra density. Finally, 107,970 particles were selected and refined without an additional mask. Furthermore, three rounds of CTF refinement and 3D refinement were performed to correct for beam tilt and astigmatism, and a Bayesian polishing tool was performed to improve particle quality. A final map of 3.4 Å resolution was obtained.

On the other hand, the expanded conformation initial 260,796 particles were further classified, and one class was discarded due to particle misalignment. The remaining 196,232 particles were 3D refined, CTF corrected, Bayesian polishing was performed. The consequent 3D refinement yielded to a 3.5 Å resolution density map. To further eliminate damaged and misaligned particles, a mask covering Clamp and stalk domain was generated and 3D classification without image alignment was run. Finally, three classes corresponding to 161,531 particles were refined and post-processed, giving as output a 3.2 Å resolution density map which was then used for the atomic model generation. The entire process is summarised in Figure 36.

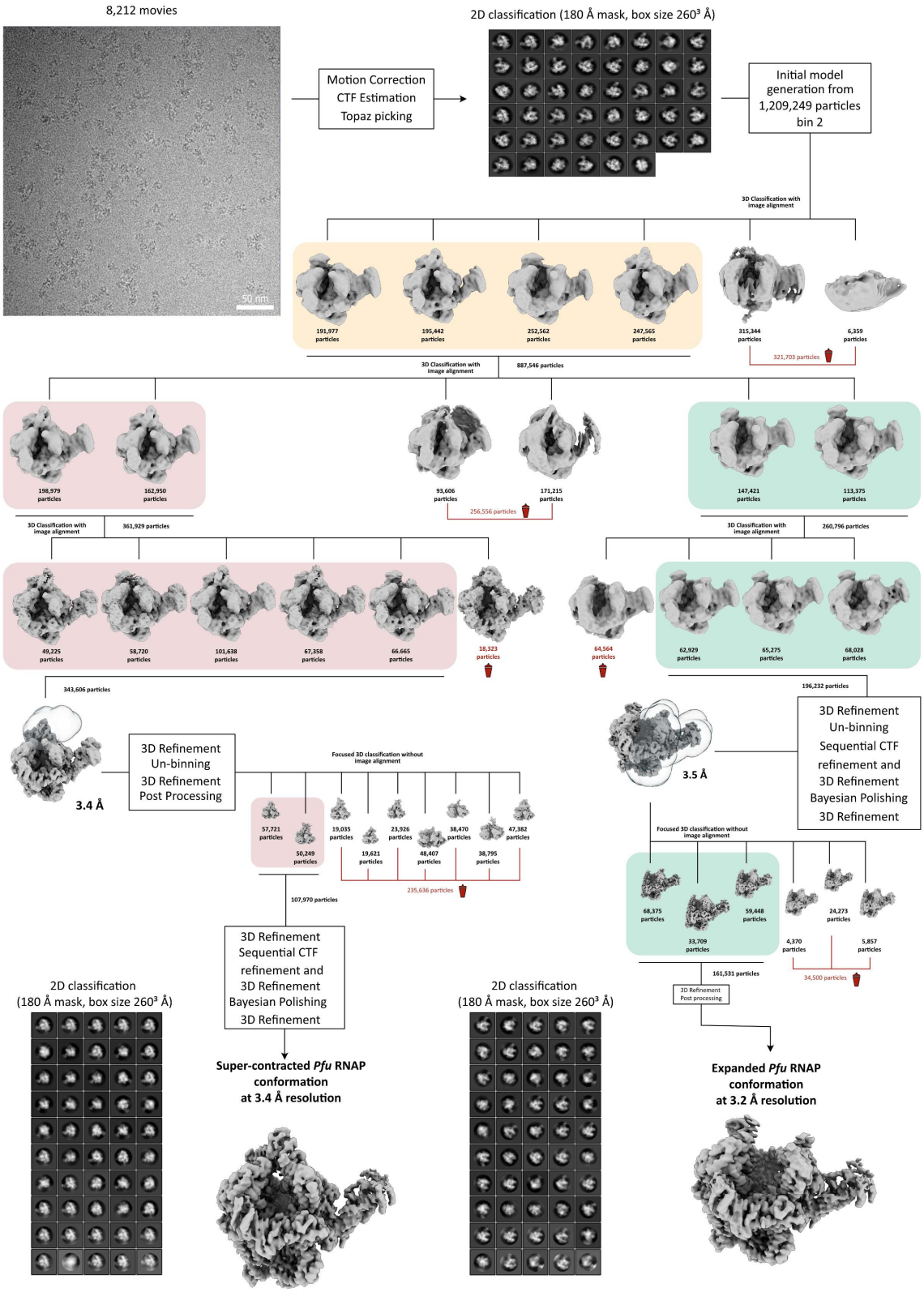


Figure 35. Cryo-EM and single particle analysis processing workflow for the apo *Pfu* RNAP sample, including representative micrograph, 2D averages, 3D classifications, final refined map in RELION 4.0.

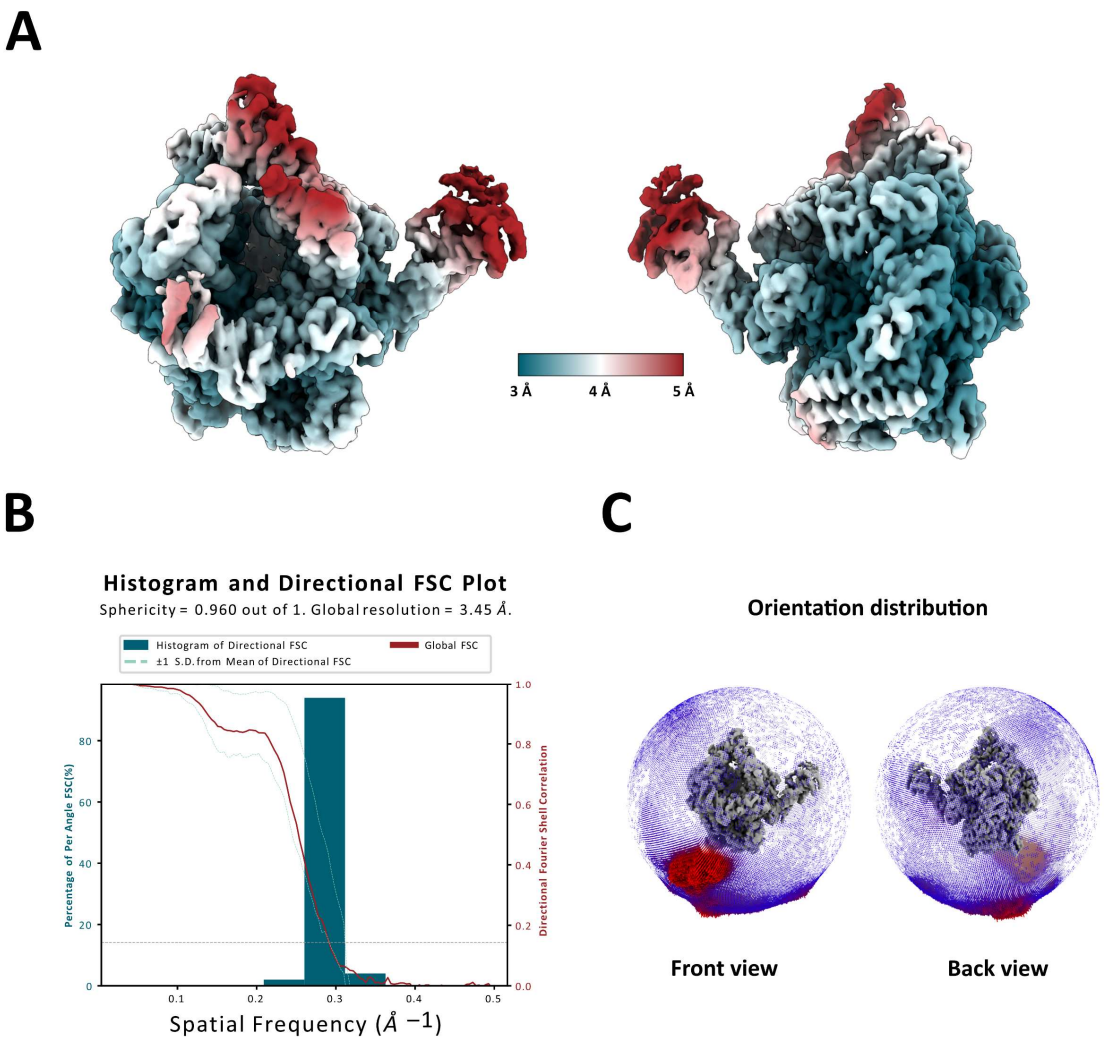


Figure 36. Map validation of the apo *Pfu* RNAP in its super-contracted conformation. A. The final refined map coloured according to local resolution. Ranging from 5 Å in red, to 4 Å in white, and 3Å in petrol. **B.** Directional FSC curves calculated with 3D FSC online tool with the gold standard cut-off at 0.143. **C.** Orientational distribution displayed in front view and back view.

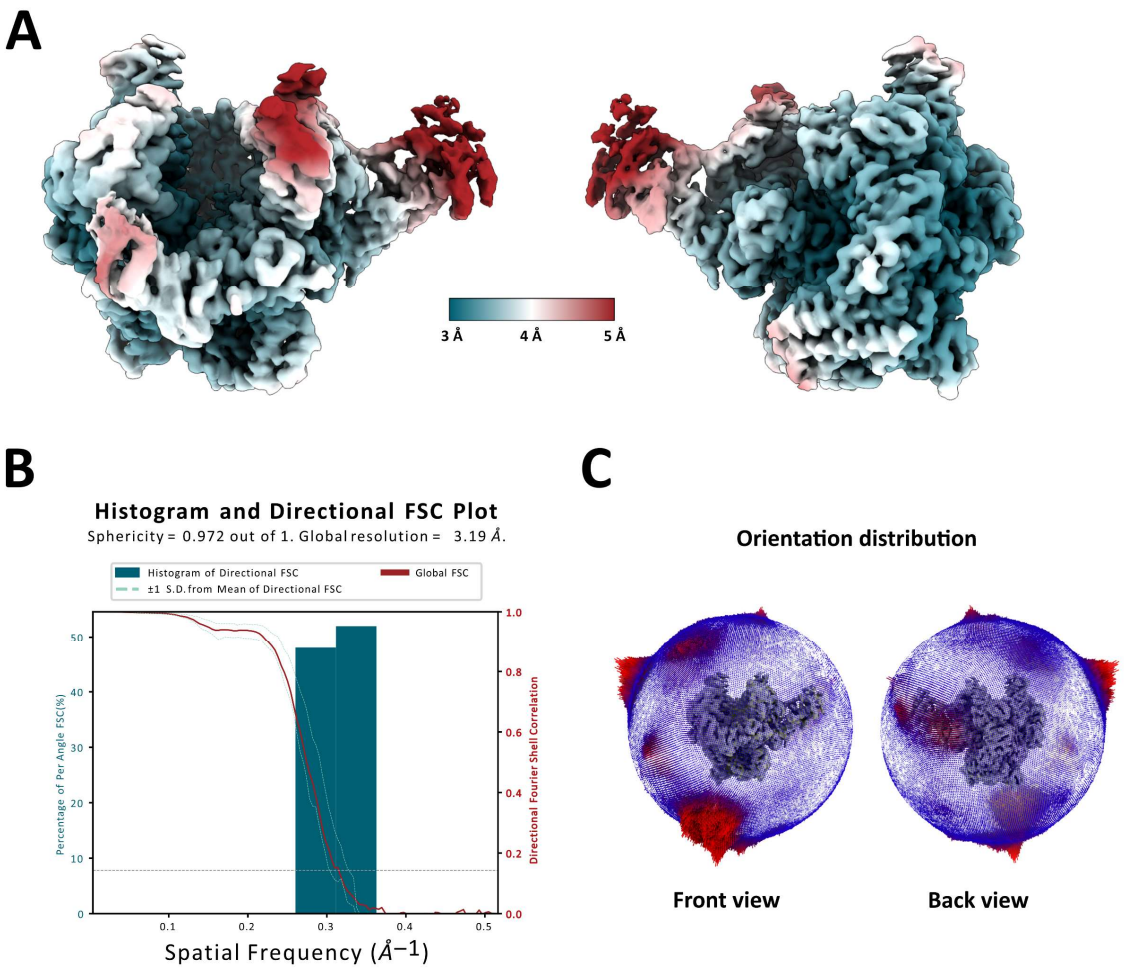


Figure 37. Map validation of the apo *Pfu* RNAP in its expanded conformation. A. The final refined map coloured according to local resolution. Ranging from 5 Å in red, to 4 Å in white, and 3Å in petrol. **B.** Directional FSC curves calculated with 3D FSC online tool with the gold standard cut-off at 0.143. **C.** Orientational distribution displayed in front view and back view.

The PDB model of the super-contracted *Pfu* RNAP conformation was generated using the TEC bound to the Spt4/5 structural model without the scaffold. The model was then adapted to the corresponding map in WinCoot software[175]. The clashes were reduced in the ChimeraX[63] implemented tool, ISOLDE[176]. Finally, the model was refined against the unsharpened map in PHENIX[177] and the statistics are shown in Table 13. On the other hand, the expanded conformation was built by using the overall structural model of the TEC with a few exceptions. Subunits composing the clamp domain (Rpo1N and Rpo2) were predicted again with AlphaFold 2[138] and built into the model using Coot[175] and Isolde[176]. The stalk domain was rigidly fitted in the multi-body refinement map, but the whole structural model was refined against its full unsharpened map in PHENIX[177]. The statistics from MolProbity are shown in Table 13.

Table 13. Cryo-EM data collection and atomic model refinement statistics of the *Pfu* RNAP apo states.

	Pfu Apo RNAP Expanded conformation	*Pfu* RNAP with Spt4/5	*Pfu* Apo RNAP Contracted conformation
EMDB	EMD-17130	EMD-17366	EMD-19033
PDB	8ORQ	8P2I	8RBO
Data collection and processing			
Magnification	50k	50k	50k
Voltage (kv)	200	200	200
Electron exposure (e^{-}/Å2)	40	40	40
Defocus range (μm)	-1 to -2.2	-1 to -2.2	-1 to -2.2
Pixel size (Å)	0.968	0.968	0.968
Symmetry imposed	C1	C1	C1
Initial coordinates (no.)	1,209,249	1,209,249	1,183,000
Final particle (no.)	161,531	107,970	111,194
Map resolution (Å)	3.2	3.4	3.0
FSC threshold	0.143	0.143	0.143
Refinement			
Homology model used (PDB code)	6kf9	6kf9	6kf9
Model composition			
Non-hydrogen atoms	26195	27486	26375
Protein residues	3266	3430	3290
Nucleotides	-	-	-
Ligands	5(Zn), 1(Mg)	6(Zn), 1(Mg)	5(Zn), 1(Mg)
APD (B factors)			
Protein (max)	98.61	245.70	98.61
Nucleotides (max)	-	-	
Ligand	30	30	30
Bonds (RMSD)			
Lengths (Å)	0.005 (0)	0.004 (0)	0.003 (0)
Angles (°)	0.545 (0)	0.626 (0)	0.557 (11)
Validation			
MolProbity score	1.69	1.57	1.53
Clashscore	8	8.44	9.3
Poor rotamers (%)	1.7%	2%	3%
Ramachandran plot			
Favored (%)	96.54	97.41	97.83
Allowed (%)	3.46	2.59	2.17
Disallowed (%)	0	0	0

5.2 Transcription Elongation complex data collections, processing, and model validation

For the attempt of plunge-freezing, different types of cryo-grids were used: Quantifoil R2/1, 300 mesh coated with an additional 2 nm thick continuous carbon layer; Quantifoil R2/2, 300 mesh coated with graphene oxide; Quantifoil R1.2/1.3, 300 mesh, coated with graphene oxide and without any additional

layer. The grids with graphene oxide were not glow-discharged, whereas the rest were glow-discharged twice using the following parameters: 0.4 mbar, 15 mA, 100 sec. Three microlitres of the sample was applied onto the grids and incubated for 60 seconds in the Vitrobot chamber at 4°C and 100% humidity. The excess of liquid was then blotted out with a force of 12, for 5 seconds and then transferred into liquid ethane. The grids were sent to University of Würzburg, Böttcher Laboratory for screening purposes. The best grid, Quantifoil R2/2, 300 mesh with graphene oxide, was chosen for an overnight data acquisition. 2,132 movies of 40 frames each were collected at 75,000x magnification in linear mode with a pixel size of 1.0635 Å.

The dataset was reprocessed recently in the newest version of Relion 5.0[137]. The movies were first motion corrected using RELION's implemented tool MOTIONCOR2[194]. Next, the frames were aligned, and CTF was estimated with CTFFIND[174] tool from RELION. Particle picking was initially performed manually to produce 2D classes suitable for reference-based particle picking. After two rounds of 2D classification, 487,165 particles were selected for 3D model generation using the unbiased initial model tool. Classes containing misaligned particles and preferential views were discarded throughout three rounds of 3D classification as shown in Figure 36. Finally, the best classes of particles were further processed by performing two cycles of CTF refinement, 3D refinement, and post-processing. A final Bayesian polishing job was carried out on the final map particles, resulting in a nominal resolution of 3.7 Å (Figure 36). The map local resolution, though, ranges from 3 Å to 6 Å in the regions of the polymerase known to be particularly flexible. The overall Fourier completeness is 91.5% and the angular distribution displays few preferential views (Figure 40B).

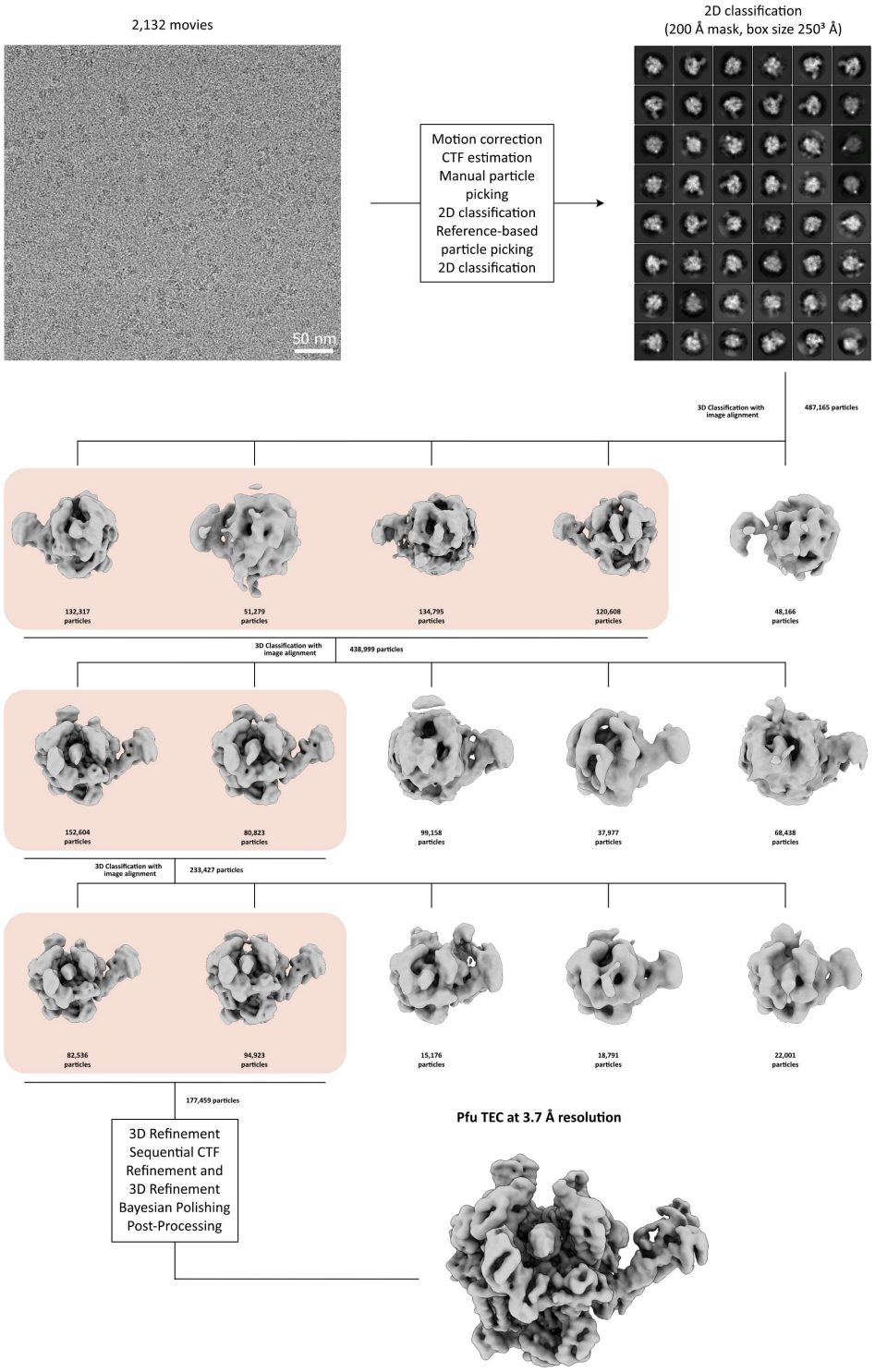


Figure 38. Cryo-EM and single particle analysis processing workflow of the TEC sample including representative micrograph, 2D averages, 3D classifications, final refined map in RELION 5.0.

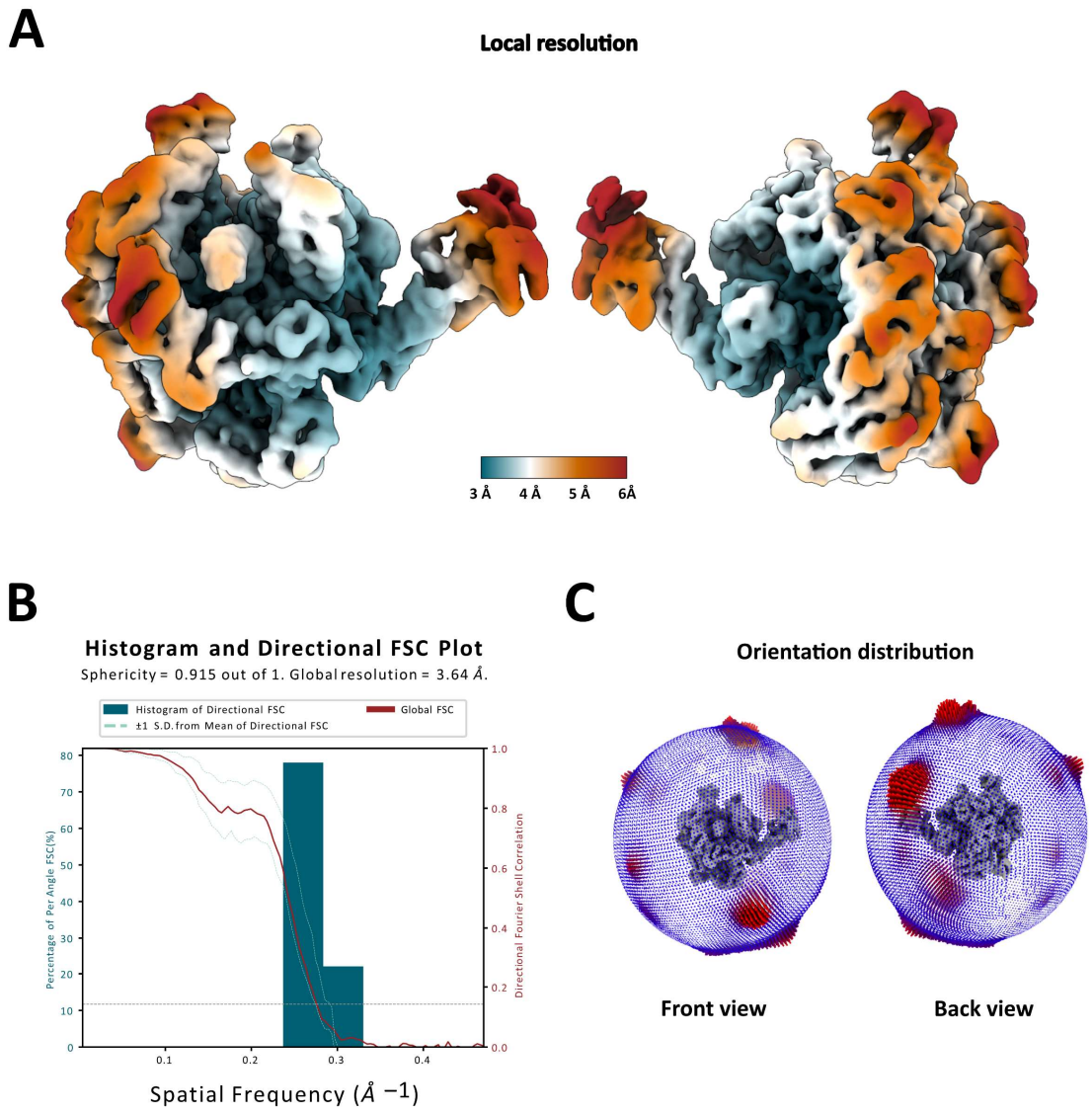


Figure 39. Map validation of the apo *Pfu* RNAP transcription elongation complex. A. The final refined map coloured according to local resolution. Ranging from 6 Å in red, to 3 Å in petrol with intermediate values at 4 Å in white and 6 Å in orange. **B.** Directional FSC curves calculated with 3D FSC online tool with the gold standard cut-off at 0.143. **C.** Orientational distribution displayed in front view and back view.

The second TEC dataset described in Chapter 2.4 corresponding to the short scaffold TEC, was plunge frozen with Vitrobot Mark VI (Thermo Fisher Scientific) onto gold grids (UltrAufoil R1.2/1.3; 300 mesh) which was freshly glow discharged with EasiGlow (TedPella) (2x0.4 mbar; 15 mA; 100 sec). The dataset was acquired at Rudolf Virchow-Zentrum - Center for Integrative and Translational Bioimaging, Böttcher Lab at the Titan-Krios G3 with an X-FEG source, 300 kV, Selectris Energy filter. A total of 2,952 movies were acquired in linear mode with a Falcon III direct electron detector. The movies were first motion corrected using RELION's MOTIONCORR2[194] executable tool. The movies were then aligned and CTF was estimated. After 2D classification filtering, a starting number of 1,074,946 particles binned by a factor of four, were used for 3D reconstructions analysis. 3D classification was executed to determine structure heterogeneity and eliminating classes corresponding to misaligned particles or exhibiting preferential views. Finally, after three rounds of 3D classification jobs, 317,487 particles

were un-binned, refined, and additional CTF refinement steps and Bayesian polishing was performed. The final reconstruction resolution yielded to 3.1 Å.

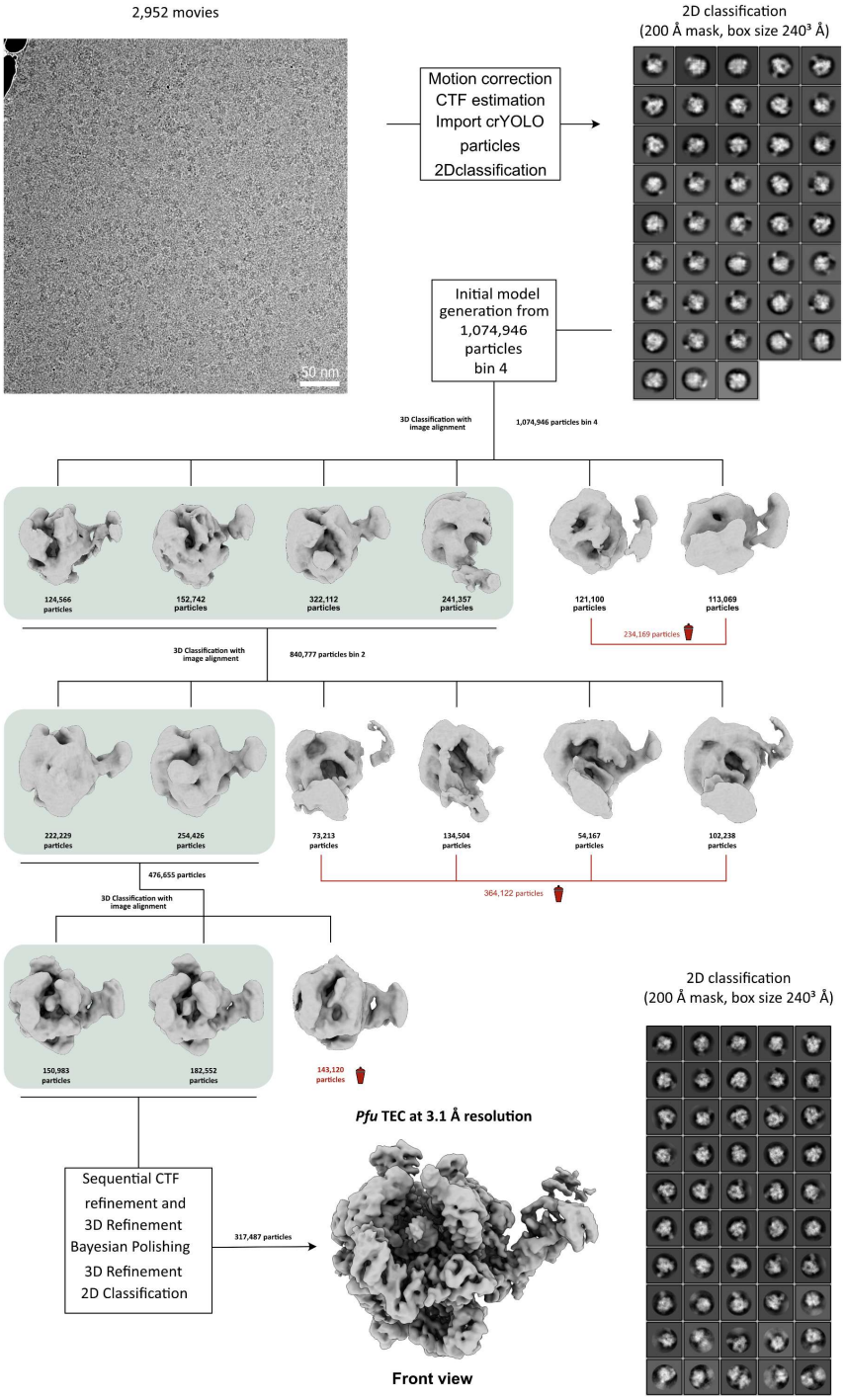


Figure 40. Cryo-EM and single particle analysis processing workflow of the TEC sample including representative micrograph, 2D averages, 3D classifications, final refined map in RELION 3.1.

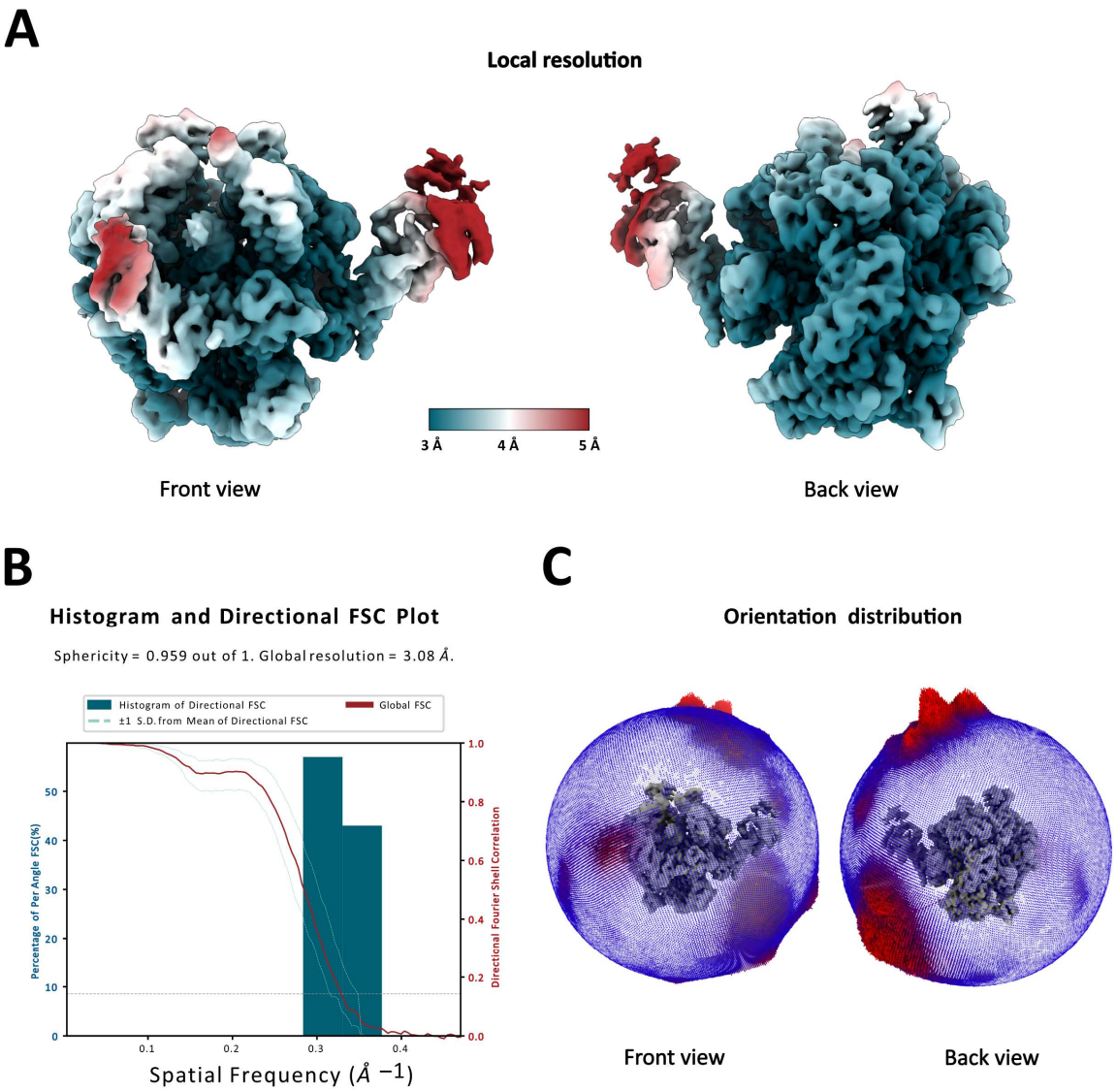


Figure 41. Map validation of the apo *Pfu* RNAP transcription elongation complex. A. The final refined map coloured according to local resolution. Ranging from 5 Å in red, to 4 Å in white and to 3Å in petrol. **B.** Directional FSC curves calculated with 3D FSC online tool with the gold standard cut-off at 0.143. **C.** Orientational distribution displayed in front view and back view.

To investigate the three-dimensional architecture of archaeal TEC, an atomic model was built in the density map in WinCoot[175] and ISOLDE[176] tool from ChimeraX[63] and refined in PHENIX[177]. The MolProbity statistics are listed in Table 9.

Table 14. Cryo-EM data collection and atomic model refinement statistics of the *Pfu* TEC sample.

	Pfu Elongation complex
EMDB	EMD-16809
PDB	8CRO
Data collection and processing	
Magnification	75k
Voltage (kv)	300
Electron exposure (e^{-}/Å2)	88.73
Defocus range (μm)	-1.2 to -2.0
Pixel size (Å)	1.0635
Symmetry imposed	C1
Initial coordinates (no.)	1,229,339
Final particle (no.)	333,535
Map resolution (Å)	3.0
FSC threshold	0.143
Refinement	
Homology model used (PDB code)	6kf9
Model composition	
Non-hydrogen atoms	27169
Protein residues	3289
Nucleotides	39
Ligands	5(Zn), 1(Mg)
APD (B factors)	
Protein (max)	214
Nucleotides (max)	214
Ligand	30
Bonds (RMSD)	
Lengths (Å)	0.003 (0)
Angles (°)	0.548 (0)
Validation	
MolProbity score	1.50
Clashscore	10
Poor rotamers (%)	2.1%
Ramachandran plot	
Favored (%)	98.28
Allowed (%)	1.72
Disallowed (%)	0

5.3 Transcription Elongation Complex associated with Spt4/5, data collection, processing, and model validation

The sample preparation of the TEC in complex with the transcription elongation factor Spt4/5 was described in the results section Paragraph 2.5. The crosslinked sample was plunge frozen with Vitrobot Mark VI (Thermo Fisher Scientific) onto gold grids (UltrAufoil R1.2/1.3; 300 mesh), which were freshly glow discharged with EasiGlow (TedPella) (2x0.4 mbar; 15 mA; 100 sec). To assess the specimen preparation success the grids were screened at the CRYO ARM^{TM} 200 (Jeol) equipped with Gatan K2 summit direct detector. After a first glimpse at the sample distribution and concentration, a data collection was set up in SerialEM[129] acquiring a total of 6,076 raw movies in super-resolution mode at

50,000x magnification and a pixel size of 0.968 Å. The in-column energy filter was operated with a slit width of 20 eV. A total exposure of 40 e^-/A^2 was fractionated over 40 frames (1 e^-/pixel/s) at a defocus range from -0.6 µm to -2µm with 0.2 µm steps.

The 6,076 raw movies were imported in RELION 4.0 beta. The movie frames were aligned and dose weighted in RELION's implementation MOTIONCORR2[194] and Contrast Transfer Function (CTF) estimation using GCTF[136]. A subset of 4,802 movies were selected excluding astigmatic, ice crystals contaminated or drifted micrographs. A first round of particle picking was performed using Laplacian-to-Gaussian tool in RELION 4.0 beta and yielding to 526,973particles picked which were then extracted with a box size of 260^3 Å with a binning factor of 4x. The extracted particles were classified in 2D in 50 classes initially afterwards the most representative classes corresponding to 345,386 particles were used as template tor Reference-based particle picking. The particles were again extracted with a box size of 260^3 Å and three rounds of 2D classification and subset selection were performed ending up with 828,076 particles. Afterwords, an initial 3D model was generated, and two rounds of 3D classification were run to eliminate misaligned particles of artifacts. Two additional 3D classification jobs were performed with the bin factor of 2x until particle extraction with the original pixel size. The final 3D map showed a weak density for Spt4/5, therefore a focused map including Spt4/5 and the stalk domain was generated and 3D classification without image alignment was carried out in the attempt to eliminate particles corresponding to *Pfu* TEC alone. A class formed by 140,120 particles was indeed empty and the particles from this class were excluded from further analysis. The remaining 182,018 particles were 3D refined and another focused mask incorporating stalk and whole clamp domain was generated for another round of 3D classification without image alignment. The result of this last passage consisted in the generation of three maps one of which was lacking a density for the NTS (class 2 – 49,231 particles). Class 1 and 3 were further refined together and separately with the best map representing class 3 individually refined with a final nominal resolution 3.4 Å and composed by 52,527 particles. The final density map was used for the atomic model fitting.

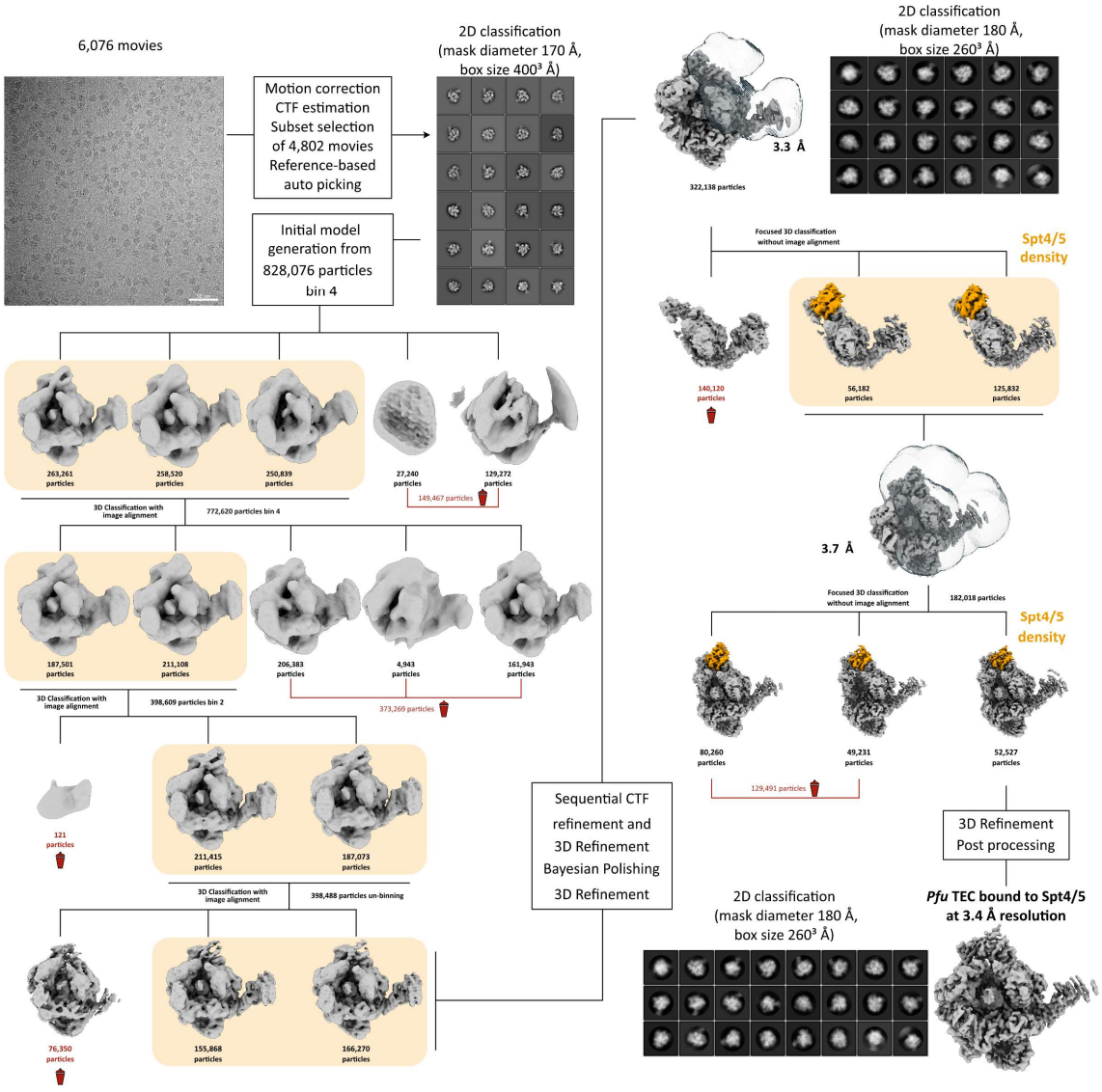


Figure 42. Cryo-EM and single particle analysis processing workflow of the TEC in the presence of Spt4/5 (crosslinked sample) including representative micrograph, 2D averages, 3D classifications, final refined map in RELION 4.0 beta.

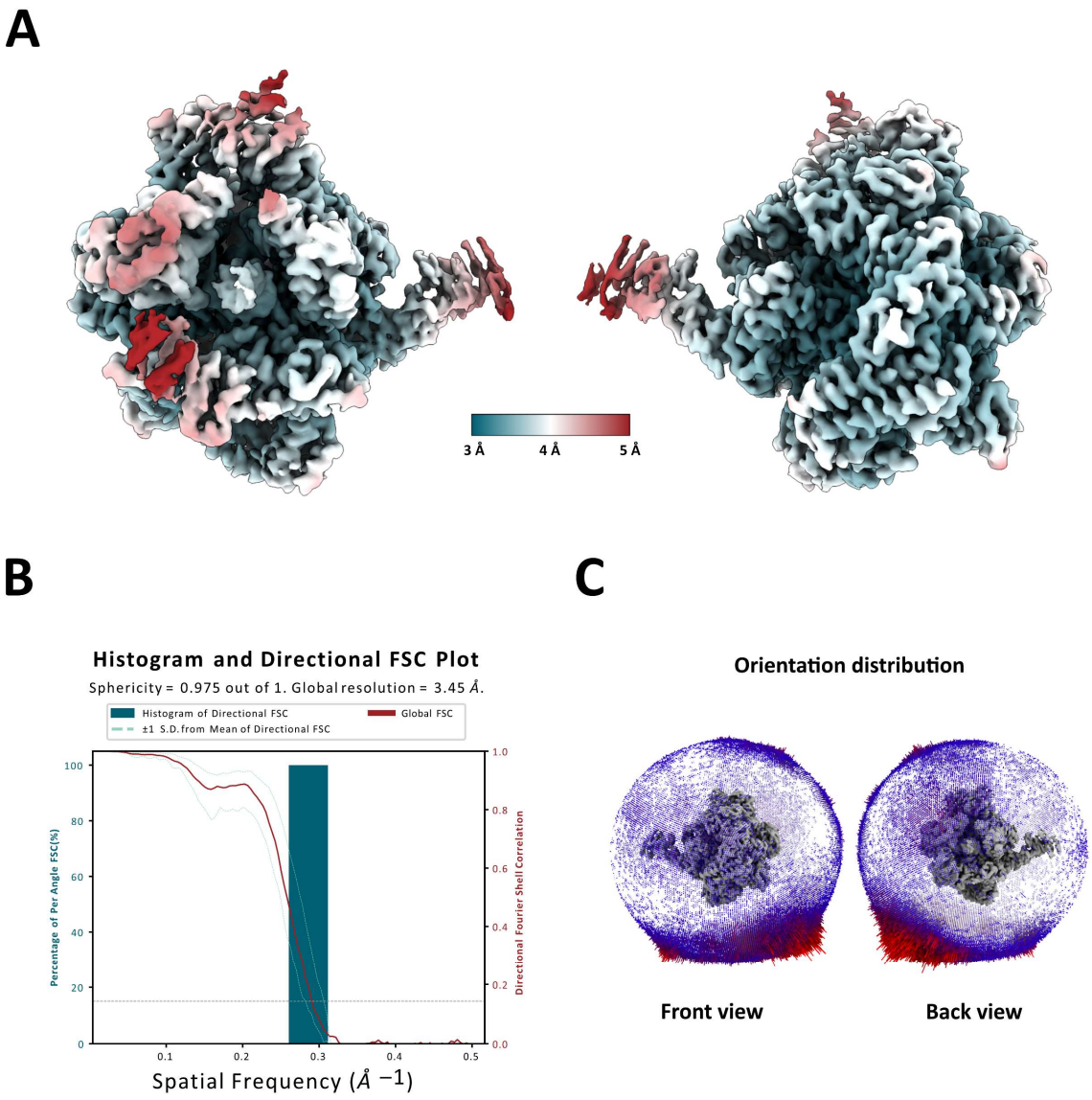


Figure 43. Map validation of the apo *Pfu* RNAP transcription elongation complex with Spt4/5. The final refined map is coloured according to local resolution. Ranging from 5 Å in red, to 4 Å in white and to 3 Å in petrol. B. Directional FSC curves calculated with the 3D FSC online tool with the gold standard cut-off at 0.143. C. Orientational distribution displayed in front view and back view.

To generate the three-dimensional model of archaeal TEC with Spt4/5, the TEC atomic model was used. Together, the Spt4 and Spt5 NGN domain were taken from their crystal structure (PDB ID: 3QQC) and fitted in the density map in WinCoot[175] and ISOLDE[176] tool from ChimeraX[63] and refined in PHENIX[177]. The MolProbity statistics are displayed in Table 10.

Table 15. Cryo-EM data collection and atomic model refinement statistics of the *Pfu* TEC in complex with Spt4/5.

	Pfu Elongation complex with Spt4/5
EMDB	EMD-16929
PDB	8OKI
Data collection and processing	
Magnification	50k
Voltage (kv)	200
Electron exposure (e^{-}/Å2)	40
Defocus range (μm)	-1 to -2.2
Pixel size (Å)	0.968
Symmetry imposed	C1
Initial coordinates (no.)	1,878,102
Final particle (no.)	322,138
Map resolution (Å)	3.4
FSC threshold	0.143
Refinement	
Homology model used (PDB code)	6kf9
Model composition	
Non-hydrogen atoms	28561
Protein residues	3260
Nucleotides	52
Ligands	6(Zn), 1(Mg)
APD (B factors)	
Protein (max)	245.60
Nucleotides (max)	214.13
Ligand	30
Bonds (RMSD)	
Lengths (Å)	0.002 (0)
Angles (°)	0.545 (5)
Validation	
MolProbity score	1.28
Clashscore	5.62
Poor rotamers (%)	2%
Ramachandran plot	
Favored (%)	98.09
Allowed (%)	1.88
Disallowed (%)	0

5.3 Collaborations or additional projects

5.3.1 *Pyrococcus furiosus* translation initiation factors (IFs) purification and complex formation

Translation, like transcription, can be divided into three main steps: initiation, elongation, and termination. In this study, translation initiation factors were investigated. As already described[196,197], archaeal transcription initiation factors (aIFs) are devoted to mRNA loading onto the small ribosomal subunit, and their resemblance to eukaryotic translation initiation factors (eIFs) is undeniable (Table 11). Translation initiation can also be divided into two stages. First, the ternary complex assembly occurs, consisting of e/aIF2 association to the methionated tRNA (Met-tRNAi^{Met}) in presence of GTP, but without GTP hydrolysis, in a combination called the ternary complex (TC). Following the initial stage, the messenger RNA (mRNA) positioning on the P-site of the small ribosomal subunit 30S (SSU) is correctly achieved by the association of the e/aIF2: Met-tRNAi^{Met} with the assistance of e/aIF1, e/aIF1A, and terminates with e/aIF5B[198]. Finally, translation initiation concludes with the recruitment of the large ribosomal subunit (50S) and the release of translation initiation factors.

Archaeal mRNA transcripts can present a Shine Dalgarno sequence or a 5' untranslated region (UTR) meant to interact with ribosomal RNA, but most transcripts are leaderless or devoid of any UTR and SD sequence. The mRNA loading occurring through the SD sequence or UTR limits the amount of "scanning" performed by the ternary complex in order to find the translation start codon AUG and stabilise the Met-tRNA$_i^{Met}$ onto the right position, therefore defining the proper reading frame[196,197].

Table 16. List of bacterial, archaeal, and eukaryotic translation initiation factors.

Bacteria	Archaea	Eukarya
IF3	aIF1	eIF1
IF1	aIF1A	eIF1A
-	aIF2(α, β, γ)	eIF2(α, β, γ)
IF2	aIF5B	eIF5B
-	-	eIF5
-	aIF2B (α, β, γ)	eIF2B (α, β, γ)
-	-	eIF3 (6-13 subunits)
-	aIF4A	eIF4F (4A, 4G, 4E)
-	-	eIF4B

Most structural studies have been performed on *Pyrococcus abyssi* translation initiation complexes[196,198], a highly homologous organism to *Pyrococcu furiosus*. MSc. Jasmin Heimerl has improved purification strategies already established by MSc. Martin Brehm[199] and tested the expected association of the different factors. Subsequently, mRNA binding was also investigated to determine if aIF proteins can bind to nucleic acid residues in the absence of methionated tRNA or SD sequence. Initially, all five proteins were predicted using AlphaFold 2 to gain a deeper understanding of the spatial rearrangements, as no structure is available for *Pfu* translation initiation factors (Figure 45C/D). As

shown in panel D, according to the prediction, the three subunits of aIF2 assemble as *P. abyssi* heterotrimer.

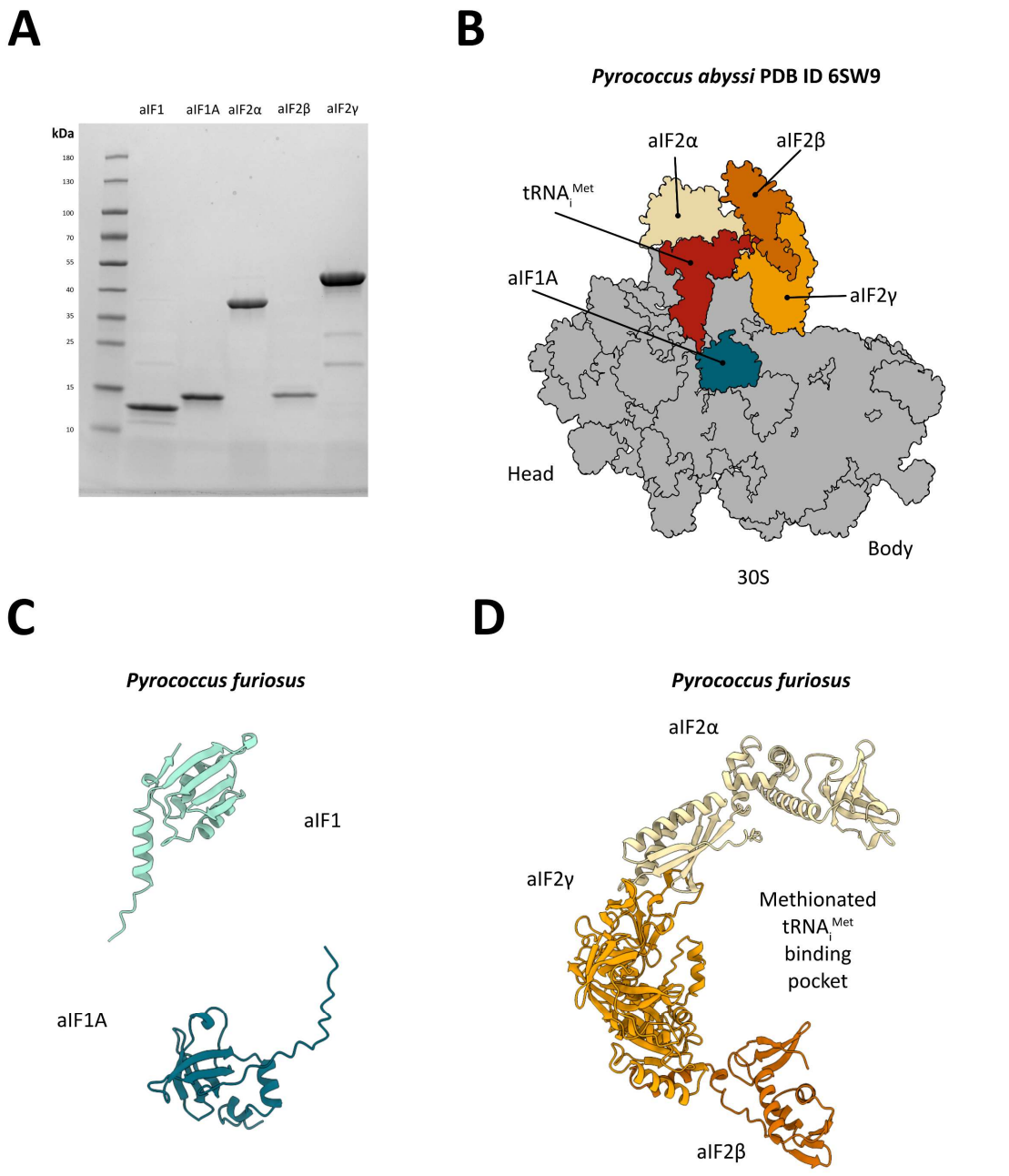


Figure 44. *Pyrococcus furiosus* translation initiation factors protein prediction. A. 4%-20% SDS-PAGE loaded with purified *Pyrococcus furiosus* translation initiation factors. **B.** Schematic representation of the *Pyrococcus abyssi* translation initiation complex PDB ID 6SW9[200] coloured according to the different components: 30S in grey, aIF1A in petrol, aIF2α in beige, aIF2β in dark orange, aIF2γ in orange, and the methionated tRNA$_i^{Met}$ in red. **C.** AlphaFold 2 prediction of *Pyrococcus furiosus* aIF1 in light green and aIF1A in petrol. **D.** AlphaFold 2 multimer prediction of *Pyrococcus furiosus* aIF2 subunits: aIF2α in beige, aIF2β in dark orange, and aIf2γ in orange.

The recombinant protein purification protocol was carried out for each protein as described in Chapter 3.2.5. Plasmids corresponding to aIF1A (pMUR 509), aIF2α (pMUR510), aIF2β (pMUR511), and aIF2γ (pMUR512) were successfully transformed and overexpressed in Rosetta (DE3) competent cells, whereas aIF1 (pMUR508) was successfully transformed and overexpressed only in *E. coli* strain C43 (DE3) chemically competent cells. The purification was performed through affinity chromatography by exploiting the 6xHIS-tags on each protein, and the most concentrated fractions were further purified

by size exclusion chromatography using a gel filtration column (Superdex® 75 10/300 GL). All proteins were loaded onto a 4%-20% gradient SDS-PAGE, as shown in Figure 42A. The final protein concentration and additional information are displayed in Table 17.

Table 17. *Pyrococcus furiosus* translation initiation factors with their corresponding molecular weight, overexpression volume, concentration, and extinction coefficient.

Archaeal Initiation Factor	Molecular weight (kDa)	Overexpression volume (mL)	Concentration (mg/mL)	Extinction coefficient (M^{-1} cm^{-1})
aIF1	13.38	500	0.369	2,980
aIF1A	15.04	500	0.720	25,105
aIF2α	33.80	500	1.80	43,890
aIF2β	18.19	500	0.167	10,680
aIF2γ	47.00	500	0.500	28,795

To biochemically verify if the purified proteins behave in a soluble manner, a multitude of mass photometry measurements were performed. It was first tested on a combination of two proteins after incubation at room temperature and at 65°C for 20 minutes. The proteins were not measured on their own since most of them have a mass inferior to 30 kDa which is the lowest threshold to achieve reliable results. The experiment was carried out using equimolar amounts of the components. The next table shows the different combinations, experimental setup and expected mass.

Table 18. Experimental setup for mass photometry analysis.

Reaction setup	Incubation temperature	Incubation time	Expected mass (kDa)
aIF2α/β/γ	Room temperature	20 minutes	98.9
	65°C	20 minutes	
aIF2α + aIF2β	Room temperature	20 minutes	51.9
	65°C	20 minutes	
aIF2α + aIF2γ	Room temperature	20 minutes	80.8
	65°C	20 minutes	
aIF2β + aIF2γ	Room temperature	20 minutes	65.1
	65°C	20 minutes	
aIF2α/β/γ + RNA (SD-gdh_oligo)	Room temperature	20 minutes	107.2
	65°C	20 minutes	
aIF2α/β/γ + RNA (No-SD-gdh_oligo)	Room temperature	20 minutes	107.2
	65°C	20 minutes	
aIF2α/β/γ + aIF1 + aIF1A + RNA (SD-gdh_oligo)	Room temperature	20 minutes	135.4
	65°C	20 minutes	
aIF2α/β/γ + aIF1 + aIF1A + RNA (No-SD-gdh_oligo)	Room temperature	20 minutes	135.7
	65°C	20 minutes	

Since the knowledge from the literature confirm that the three subunits of the aIF2 heterotrimer associate independently from the presence or absence of nucleic acid residues[196,198], there was first tested *Pfu* aIF2 assembly. The AlphaFold 2 multimer prediction tool has also hinted a highly probable association as illustrated in Figure 42D. In a first stage of the project aIF2 subunits namely aIF2α, aIF2β,

and aIF2γ were incubated in equimolar amounts and at two different temperatures, and their mass was measured (Figure 43). This passage is necessary for understanding if the three separately overexpressed and purified proteins would behave as the native complex. Although complex formation can be observed in both measurements at room temperature and 65°C, with a mass ranging from 80 kDa to 100 kDa, the replication of measurements led to heterogeneous complex formation. As we can observe in Figure 43 peaks around 60 kDa together with the proper mass peaks corresponding to the aIF2γ-aIF2β heterodimer mass. Whereas the mass detected at 65°C with the mass of ~83 kDa could represent the aIF2γ-aIF2α heterodimer formation.

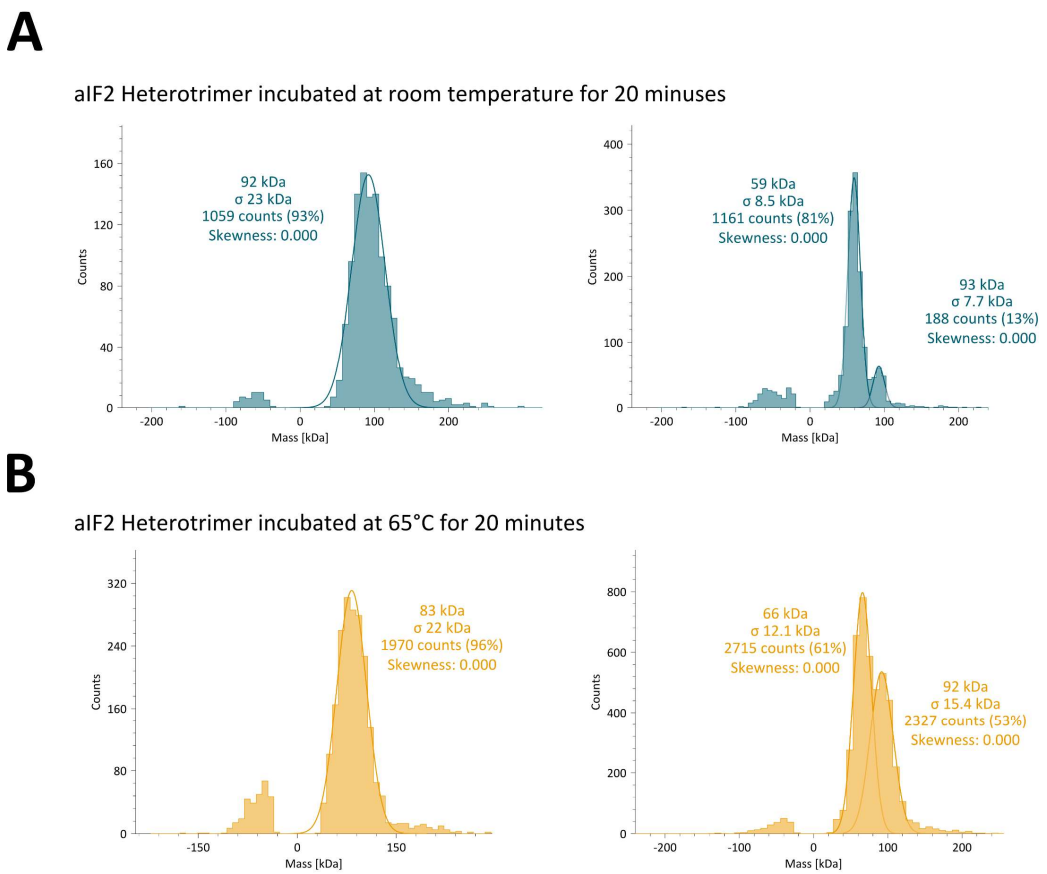


Figure 45. aIF2 heterotrimer formation. A. Mass photometry data analysis of the aIF2 heterotrimer incubated at room temperature. The left graph shows only one mass population at approximately 92 kDa. The right graph displays two distinct mass populations at 59 kDa with 81% of the molecules and one at 93 kDa with 13% of the molecules detected. **B.** Mass photometry data analysis of the aIF2 heterotrimer after incubation at 65°C. The left graph represents one homogeneous population of molecules at a mass of ~83 kDa. The right graph again shows two distinct mass populations at 66 kDa with 61% of the molecules and 92 kDa with 53% of the molecules detected.

To further elucidate the aIF2 association network, different combinations of the aIF2 subunits involving only two members were incubated and measured. First, aIF2α and aIF2β were incubated and measured at the mass photometer to confirm, in this case, that the association between these two subunits does not occur. Indeed, as displayed in Figure 44A the mass between aIF2α alone and in combination with aIF2β does not significantly differ. On the contrary both combinations, aIF2γ-aIF2β and aIF2α-aIF2γ, have shown a peak at a greater molecular mass with respect to a single subunit measurement (Figure 44B/C). Moreover, mass photometry measurements of subunit aIF2β have confirmed its instability in solution by displaying a peak at a molecular mass of twice greater than the protein's monomeric form, hinting oligomers formation (Figure 44D).

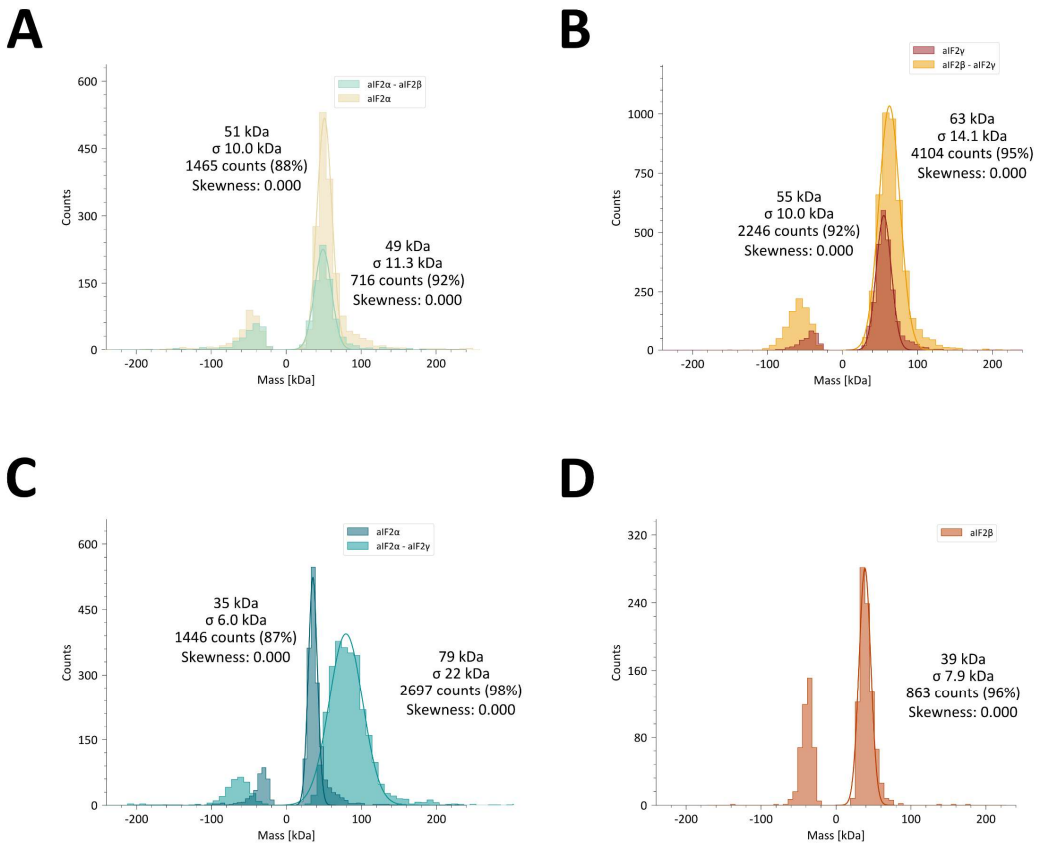


Figure 46. Mass photometry analysis of IF2 heterotrimer combinations. A. aIF2α and aIF2β combination mass photometry analysis (light green) with aIF2α alone as reference measurement (beige). **B.** aIF2γ and aIF2β combination mass photometry analysis (yellow) together with a measurement of aIF2γ alone as reference (red). **C.** aIF2α and aIF2γ combination mass photometry measurement (cyan) together with aIF2α as a reference measurement analysis (petrol). **D.** aIF2β mass photometry analysis showing dimer formation (orange).

The next step was to investigate the translation initiation factors affinity to nucleic acid oligonucleotides. To this end, two RNA oligonucleotides, gdh-SD-RNA, and gdh-No-SD-RNA, were incubated first with the aIF2 heterotrimer and then in combination including the other two members, aIF1 and aIF1A. The initiation factor aIF2 interaction to RNA was first assessed through mixing equimolar amounts of proteins and RNA. The two oligoribonucleotides were designed in a manner to have the same length, including the gene of glutamate dehydrogenase (gdh) with and without Shine-Dalgarno sequence. Furthermore, it was also tested the affinity of the complex to RNA oligonucleotides at two different temperatures to investigate if this parameter influences the outcome of the experiment. We can observe indeed that sequence alteration does not interfere with the binding as illustrated in Figure 45, but temperature seems to play an important role for complex formation. When the temperature is raised to 65°C, a population with a mass of 86-87 kDa appears for both RNA oligonucleotides (Figure 45A/C) or it is more prominent than at room temperature (Figure 45D). As for the combination of all translation initiation factors with the RNA oligonucleotides, after incubation for 20 minutes at room temperature and 65°C again, we could observe a similar behaviour as aIF2-RNA samples. Indeed, measurements of these samples have mostly revealed one prominent peak at a mass of 60-70 kDa and another peak corresponding to a mass of 90-100 kDa (Figure 45).

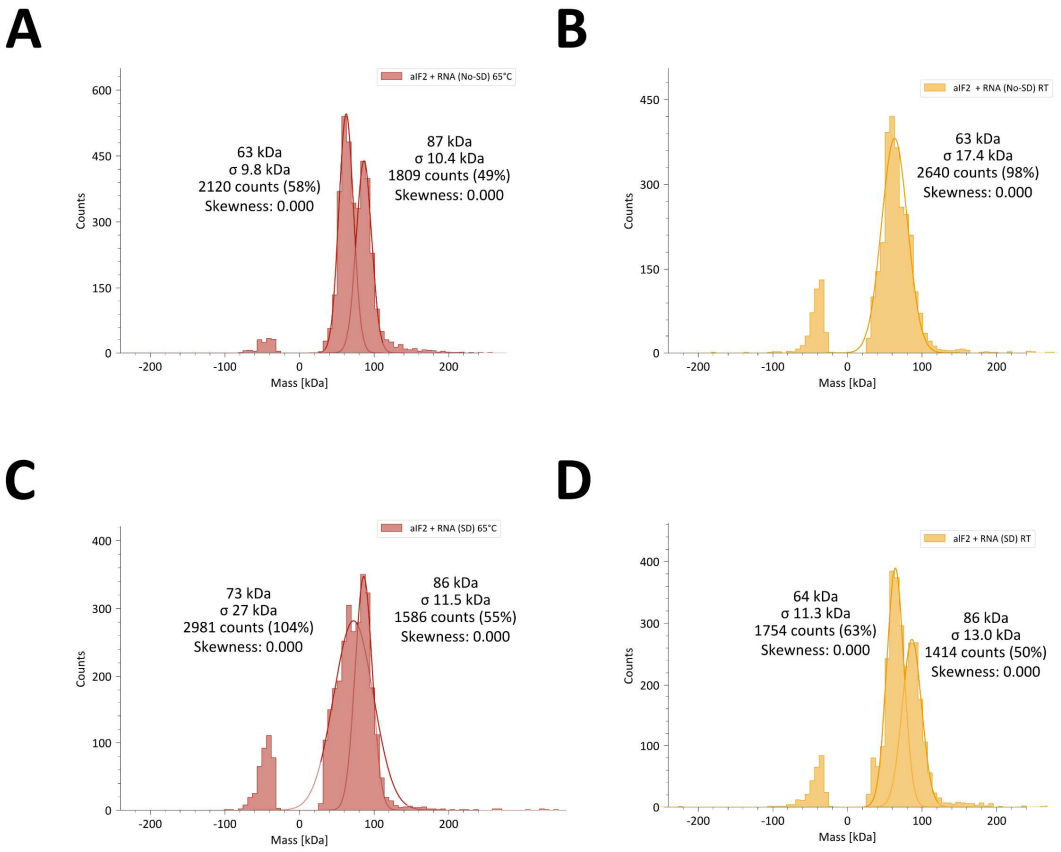


Figure 47. Mass photometry data analysis of aIF2 in complex with RNA. A. Mass photometry analysis of aIF2 heterotrimer incubated with *gdh* RNA without Shine Dalgarno sequence at 65°C. The curves represent two masses detected one corresponding to 63 kDa and another one at 87kDa (red). **B.** Same sample from **A.** was incubated again at room temperature. Data analysis shows a single peak at 63 kDa suggesting that possibly no association occurred. **C.** Mass photometry analysis of aIF2 heterotrimer incubated with *gdh* RNA with Shine Dalgarno sequence at 65°C. The curves represent two masses detected one corresponding to 73 kDa and another one at 86kDa (red). **D.** Same sample from **C.** was incubated again at room temperature. Data analysis shows a single peak at 64 kDa and at 86 kDa.

Measurements conducted on all aIFs equimolar concentrations and the same oligoribonucleotides shown above have given a similar outcome as aIF2 alone (Figure 46). Also, these results have confirmed that sequence composition does not have an impact on complex formation, but rather the temperature seems to play, again, a crucial role for the association. The overlapped graphs in Figure 46 show an increase in the mass of around 90-93 kDa when incubated at 65°C, with respect to the room temperature incubated samples where the main molecule population span at a mass of 64-65 kDa.

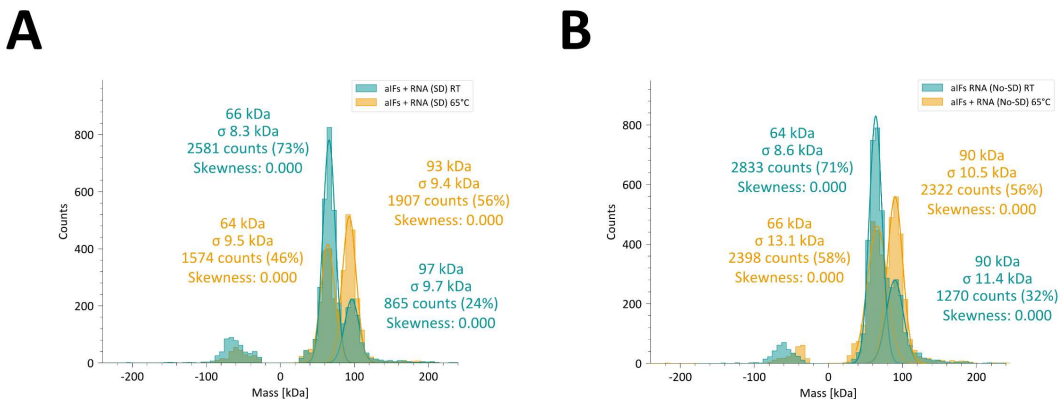


Figure 48. aIF1, aIF1A, aIF2 and RNA interaction analysis through mass photometry. A. Mass photometry analysis of aIF1, aIF1A, aIF2 and *gdh* RNA oligonucleotide with the Shine Dalgarno sequence measured at room temperature (cyan) and at 65°C (yellow) **B.** Mass photometry analysis of aIF1, aIF1A, aIF2 and *gdh* RNA oligonucleotide without the Shine Dalgarno sequence measured at room temperature (cyan) and at 65°C (yellow).

To investigate the efficiency and association rate of *Pyrococcus furiosus* aIFs and RNA, electromobility shift assays were performed. These experiments have not given a reproducible result given that only two out of eight EMSAs have shown a shift. In Figure 47 are reported the native gels that show a shift. For this experiment, 20 nM of *Pyrococcus furiosis* fluorescently labelled *gdh* RNA oligoribonucleotide was incubated with a range of aIFs concentrations starting from 0 nM up to 480 nM. Once again, the reactions were incubated rather at room temperature and 65°C. As illustrated in Figure 47A the native gels show a faint RNA shift when we add 24 or 50 times more aIFs than RNA. The native gels were also stained to confirm that the shift is due to the proteins associating to the RNA and in fact, the Coomassie staining displays a much more intense signal in correspondence to the fluorescence shift. This effect might be due to the protein intrinsic charge rather than to association to the RNA.

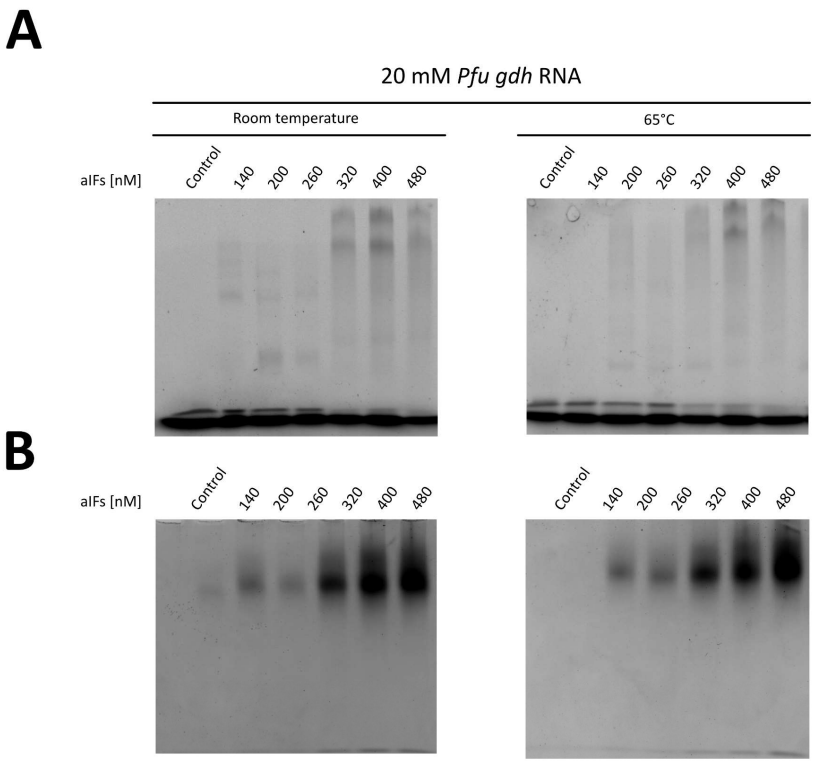


Figure 49. Electromobility shift assays of *Pyrococcus furiosus* translation initiation factors in complex with fluorescently labelled *gdh* RNA.A. Fluorescence detection at 488 nm wavelength of the 6% native gels loaded with the five aIFs in complex with *Pfu gdh* RNA (20nM). The samples loaded represent the control (RNA only), and molarity ratios of 1:7, 1:10, 1:13, 1:16, 1:20 and 1:24 of protein with respect to RNA concentration which was constant in all reactions. **B.** Coomassie staining detection of the native gels.

In the second set of EMSAs only aIF2 heterotrimer interaction with the *gdh* RNA oligonucleotide was verified. The goal of this experiment was to determine whether the shifted band pattern would vary. For this experiment, a constant concentration of 200 nM *Pfu gdh* RNA oligonucleotide was used and the range of protein was also scaled up. First, the aIF2 complex was assembled by mixing equimolar amounts of the three components: aIF2α, aIF2β, and aIF2γ. The aIF2 complex was titrated with a molar range of 200-1200 nM and the reactions were run onto a 6% native gel. The reactions were run at room temperature as well as at 65°C. The fluorescence detection at 488 nm wavelength has revealed two prominent shifts at 1 μM concentration of aIF2 against 200 nM of the RNA with most of the RNA present in a free version (Figure 48). Additionally, the temperature did not drastically influence the binding affinity.

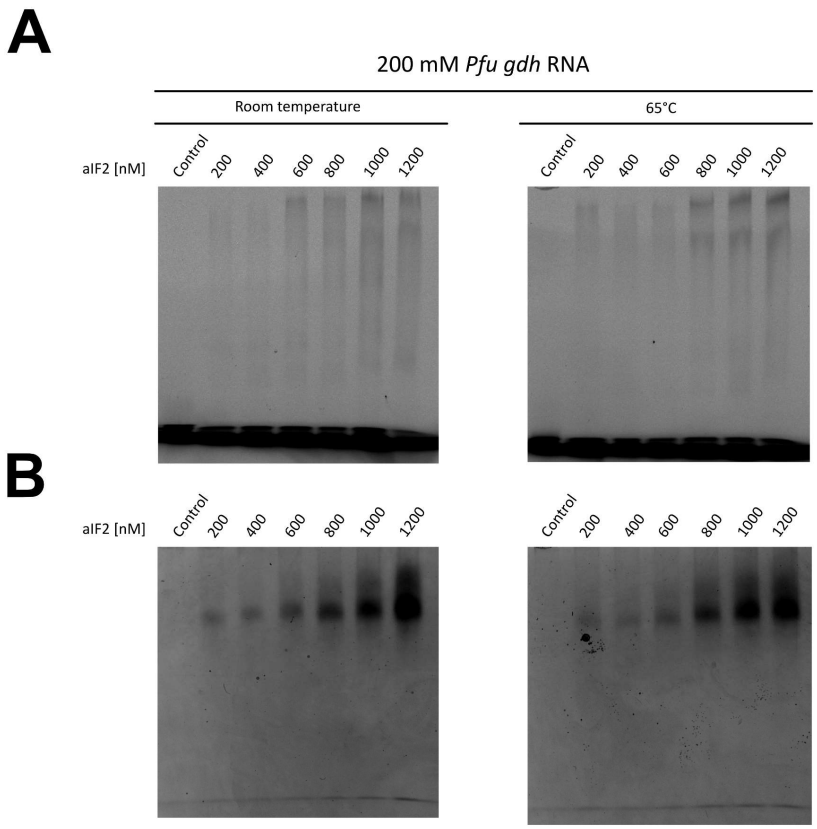


Figure 50. Electromobility shift assays of *Pyrococcus furiosus* aIF2 heterotrimer in complex with fluorescently labelled *gdh* RNA.A. Fluorescence detection at 488 nm wavelength of the 6% native gels loaded with aIF2 in complex with *Pfu gdh* RNA (200nM). The samples loaded represent the control (RNA only), and molarity ratios of 1:1, 1:2, 1:3, 1:4, 1:5, and 1:6 of protein with respect to RNA concentration which was constant in all reactions. **B.** Coomassie staining detection of the native gels.

In synthesis, the aim of the experiment was to implement the purification strategies with the scope to obtain a higher protein yield. The cloning and first overexpression attempt were performed by Martin Brehm in 2019. In our experimental setup, we aimed to upscale the overexpression volume and employ Äkta system (GE Healthcare) for the purification strategy. The first challenge represented the choice of overexpression *E. coli* strain. Due to striking similarities shared between aIF1 (also known as SUI1) and the translation inhibitor YciH[201] share a highly similar structure and sequence homology, causing cell death when recombinantly overexpressed. To overcome this issue, the expression vector was inserted in the *E. Coli* strains C41 (DE3) and C43 (DE3) for the overexpression step. These two strains are genetically engineered to overexpress toxic proteins. In the case of Martin Brehm's experimental design, the C41 (DE3) strain was able to grow after transformation, in Jasmin Haimerl's, instead, only C43 (DE3) strain was able to grow after the transformation with aIF1 gene plasmid. Nevertheless, transformed strain of *E. coli* has grown notably slower than Rosetta (DE3) standard strain.

Another challenge that has been faced was to overcome the propensity of aIF2β agglomeration in inclusion bodies. Upon different attempts, the standard purification, with lysate heat shock, affinity and size exclusion chromatography, a sufficient amount of aIF2β was successfully purified without the need of inclusion bodies protein renaturation. Nevertheless, aIF2β oligomerisation was also confirmed in mass photometry analysis, giving a mass peak corresponding to dimeric or oligomeric nature.

The overall procedure has revealed to be successful giving similar or higher yields of proteins. The lowest protein concentrations achieved represented the aIF1 (168 µg) and aIF2β-subunit (368 µg) for the reasons explained above. Whereas the highest protein amount purified represent aIF2γ (500 µg), aIF1A (720 µg) and finally aIF2α (1800 µg) from an overexpression volume of 500 ml.

Due to the successful purification of all translation initiation factors, biochemical analysis and characterisation was possible to perform. In this second part of the experiment, we aimed to characterise the efficiency of aIF2 heterotrimer association and eventual oligomerisation through mass photometry measurements. Furthermore, investigation of the binding activity of the translation initiation factors towards mRNA was also carried out by means of mass photometry and EMSAs. The masses of each individual protein were confirmed through mass photometry except for aIF2β-subunit. aIF2 heterotrimer formation has also appeared to be properly associating according to mass photometry measurements. This interaction seems to be favoured at higher temperatures when compared to room temperature incubation. In this study it was also confirmed that aIF2 subunits assembly occurs as predicted by AlphaFold 2 multimer tool, with no apparent interaction occurring between subunits β and γ.

Concerning interaction with messenger RNA, our results lack of reproducibility gives us the hint that in our experimental setup archaeal Ifs are not able to interact with RNA. This result might hint that the strongest interactions between aIFs and mRNA occurs indirectly through tRNA.

5.3.2 *Pyrobaculum aerophilum* S-layer purification for Atomic Force Microscopy (AFM)

All archaeal species characterized to date present a Surface layer (S-layer). The latter is formed by a transmembrane protein complex that oligomerises in a symmetrical fashion. The S-layer proteins are building blocks of 2, 3, 4, and 6 subunits of one type of (glyco-) protein, the highest symmetry according to the symmetry axis is p6, p4, p3, p2, and p1. This peculiar structure has been reported not only on archaeal cells, but also in pathogenic and non-pathogenic Bacteria[202–204]. Due to their crystalline nature, they form well-defined pores with a width in the nanometre range, making them industrially useful for small molecule filtration[205,206].

The S-layer is anchored outside of the cytoplasmic membrane in most Archaea. The surface layer protein contains a transmembrane domain, usually formed by an alpha-helix motif, and an extracellular domain involved in crystalline layer formation. The surface layer protein is the most abundant protein in the cells presenting an S-layer making it an extremely important component of the cell, given the energy spent to synthesize it. The S-layer function has been hypothesised to confer and maintain cell shape, which especially in extremophiles, it is crucial given the high pressure to which they are subjected in their native environments. Another peculiarity of S-layer is that the crystalline lattice is able to self-assemble.

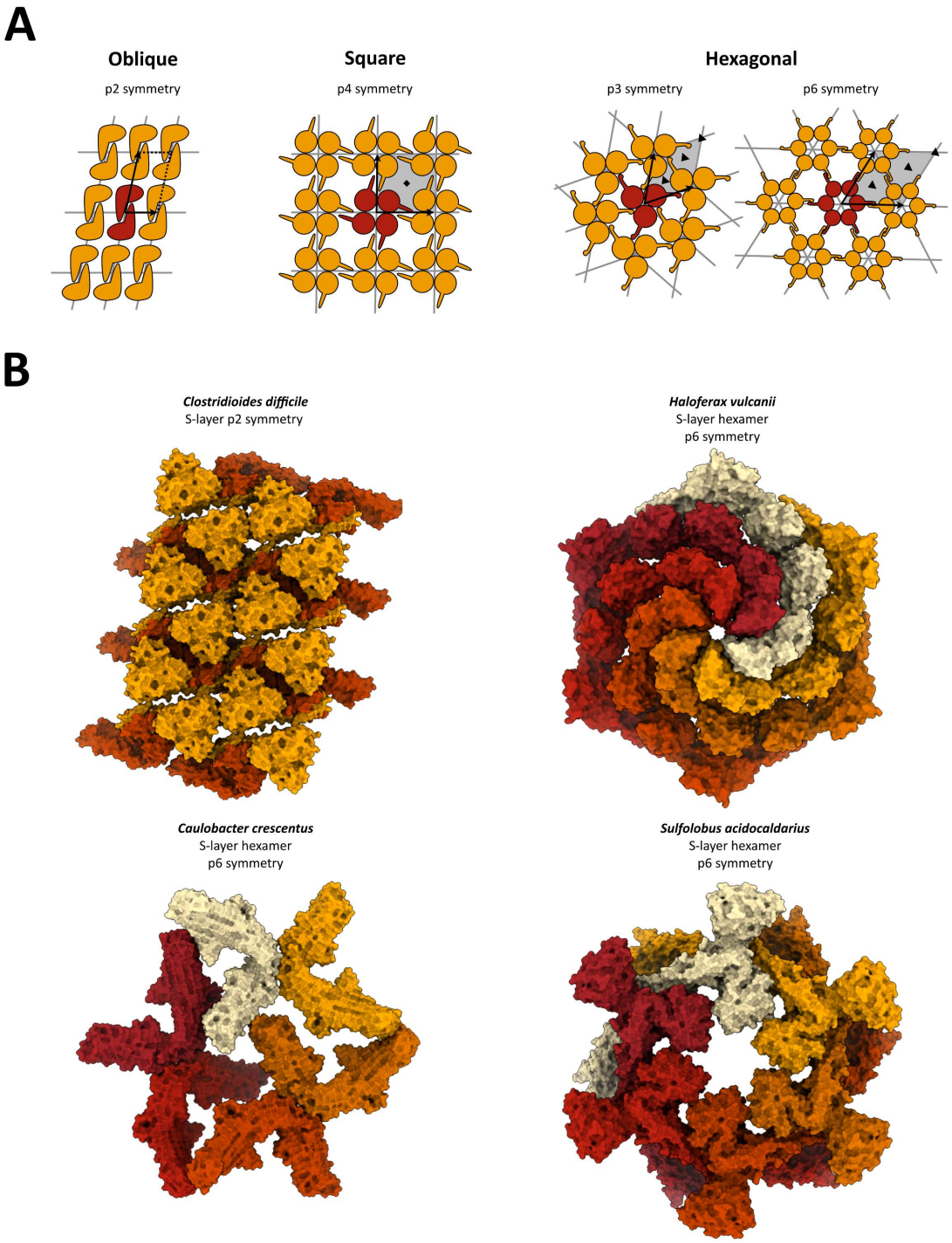


Figure 51. Surface layer symmetry schematic representation and examples. A. Schematic representation of oblique (p2 symmetry), square (p4 symmetry), and hexagonal (p3 and p6 symmetry) S-layer patterns. Scheme taken and modified from Uwe B. Sleytr et al[207] **B.** Solved structures of *Chlostridium difficile* S-layer with a p2 symmetry (PDB ID: 7QGQ), *Haloferax vulcanii* S-layer hexamer with a p6 symmetry (PDB ID: 7PTR), *Caulobacter crescentus* with a p6 symmetry (PDB ID: 5N8P), and *Sulfolobus acidocaldarius* S-layer with a p6 symmetry (PDB ID: 7QP0).

This study aimed to purify native S-layer latices from *Pyrobaculum aerophilum* isolate IM2, a hyperthermophilic archaeon which was first described in 1993[208]. *P. aerophilum* is a rod-shaped organism able to grow at temperatures ranging from 75°C to 105°C. The great advantage of using the S-layer from such organism comes from its robustness and resistance to harsh laboratory techniques

for protein denaturation. *P. aerophilum* S-layer protein assembles in a lattice with a p6 symmetry which makes it easy to identify through electron microscopy or atomic force microscopy (AFM).

To set up the experimental procedure, the protocol established by Prof. Dr. Reinhard Rachel was implemented by Katharina Vogl to achieve a better yield for AFM measurements. *P. aerophilum* IM2 cells were cultivated as described in Völkl et al[208]. One gram of *P. aerophilum* cell mass was resuspended by vortexing in 7 ml buffer 1 (20 mM MES/NaOH, pH 6.1, 1 mM $MgCl_2$). The cells were disrupted by sonication using a Bandelin Sonopuls HD 2070 for 3 minutes, cycle 5 at a power of 60%. The sonication was repeated 3 times with 1-minute breaks in between. Precipitates originating from the cultivation medium and cell debris were eliminated by short centrifugation at 2,254 g for 7 minutes at 4°C. The supernatant was collected, and 1.4 ml SDS-buffer (20 mM MES/NaOH, pH 6.1, 1 mM $MgCl_2$, 2% w/v SDS) was added. The mixture was incubated at 80°C for 1 hour to detach the S-layer from the cytoplasmic membrane. After centrifugation at 17,761 g for 2 minutes at 4°C, 3 phases are forming: supernatant, a white pellet, and a black pellet. The S-layer is found in the white pellet, which is carefully collected and resuspended in 200-400 µl of SDS buffer. Samples resulting from this S-layer preparation were used for initial TEM and AFM imaging attempts.

To further reduce the concentration of SDS from the S-layer preparation, an optimised purification strategy was developed. To this end, the white pellet was not resuspended in SDS buffer but in buffer 1 (20 mM MES/NaOH pH 6.1, 1 mM $MgCl_2$) instead. In addition, the final S-layer mixture was centrifuged 3 times at 15,871 g for 5-10 minutes at room temperature, and each time the supernatant was replaced with 200-300 µl fresh buffer 1 to reduce the final SDS concentration.

Negative staining was then performed on the white pellet of the S-layer preparation to confirm its proper purification and the amount of copurified contaminants. The negative staining procedure followed corresponds to the same used for single particle analysis purposes, and it is described in the methods Chapter 3.2.14.

Ultimately, the sample was exposed to AFM scanning sessions. Atomic force microscopy, or AFM, is a powerful tool allowing atomic imaging. It provides an incredibly detailed three-dimensional information of the sample surface. In our studies, it was employed frequency modulation AFM (FM-AFM) with a qPlus sensor introduced by *Giessibl in 1996*, and published in 1998[209]. The goal of the project was to design an ambient FM-AFM setup that allows for imaging of biological samples in both dry and liquid conditions. More details about the experimental setup are explained in *Seeholzer et al 2023*[210].

The first S-layer preparation yielded an optimal amount of S-layer pellet. The S-layer mixture was imaged at the transmission electron microscope and showed an elevated distribution of S-layer. Besides the high concentration, the crystalline lattice was also in optimal shape, with few examples of S-layer still maintaining the intact rod shape (Figure 50A/B). For this initial preparation, the S-layer white pellet was collected and resuspended in a buffer solution containing sodium dodecyl sulphate (SDS), which was undetectable in TEM micrographs. Nevertheless, when the sample was applied onto a freshly cut mica sheet and air-dried, a layer of SDS had occluded the sample from detection. Indeed, in AFM measurements, the lattice did not correspond to the S-layer lattice (Figure 50C). The lattice measured had a periodicity of 1.3 nm, which did not correspond to *P. aerophilum* S-layer nor the mica sheet. The small periodicity size hinted towards the hypothesis that the SDS present in the buffer

solution might have built sheets. The SDS adsorption on top of the S-layer, therefore, might have hindered their detection during the scanning process.

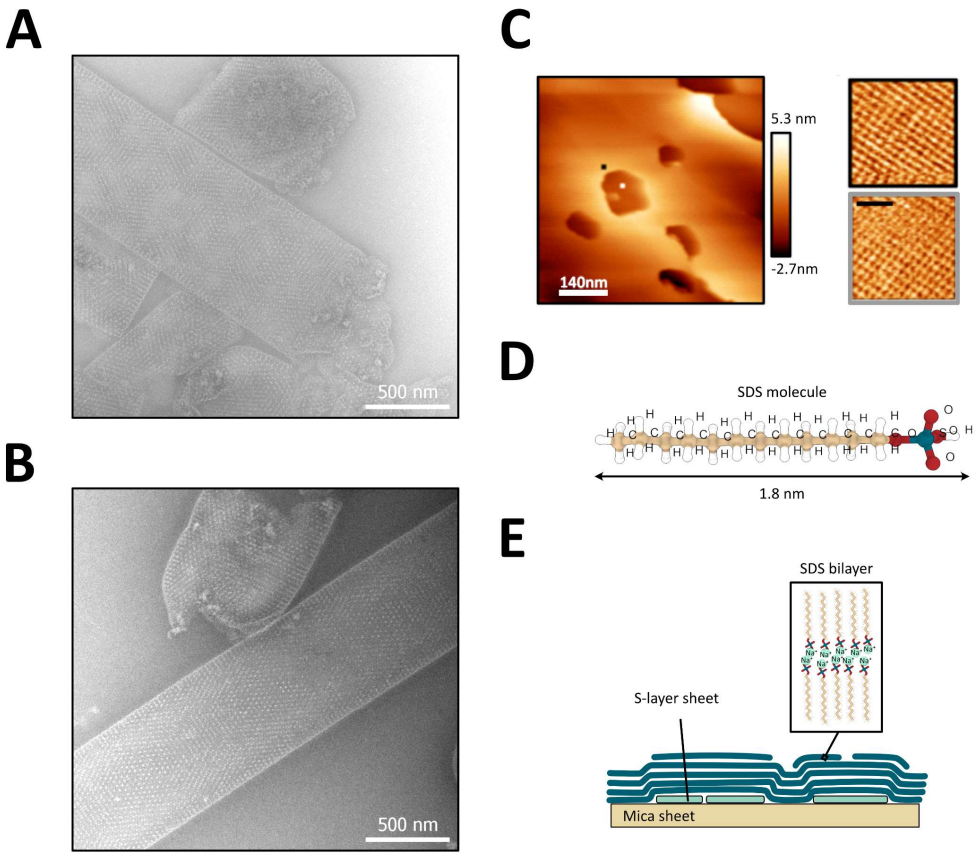


Figure 52. First attempt of *Pyrobaculum aerophilum* S-layer purification and imaging. A. and **B.** Representative negative staining micrographs of the purified *P. aerophilum* S-layer pellet with a scale bar of 500 nm. **C.** FM-AFM scans showing flat traces with several holes (left) on the right side are displayed the (black- and white-line squares) a lattice with a periodicity of 1.3 nm, which we propose to be SDS bilayers (picture taken from *Seeholzer et al 2023*[210]). **D.** 3D structural representation of the SDS molecule (PubChem CID: 3423265). **E.** Hypothetical distribution of the S-layer sheet and SDS bilayer.

To validate our hypothesis, we modified the standard S-layer preparation protocol by incorporating few additional steps following protein extraction. Given that the S-layer sheets are large, heavy macromolecules, they exhibit a tendency to sediment during centrifugation. We took advantage of this property by pelleting the S-layer from *P. aerophilum*, carefully discarding the supernatant, and then resuspending the pellet in an SDS-free buffer to reduce SDS final concentration in the mixture. Subsequently, the protein lattices were transferred onto two types of surfaces for analysis: one was blotted onto a freshly glow-discharged electron microscopy grid for negative staining, while the other was applied to freshly cleaved mica sheets for FM-AFT scanning. To assess the success of the extraction process, we performed negative staining and captured transmission electron micrographs (TEM). The resulting images revealed a homogeneous distribution of S-layer lattice fragments across the grid surface (Figure 51). Furthermore, some areas of the grid provided clear micrographs featuring single-layered lattices, allowing us to observe the characteristic p6 symmetry in the power spectrum of the images (Figure 51B/D). These observations confirm the successful extraction and preparation of the S-

layer protein for further structural analysis.

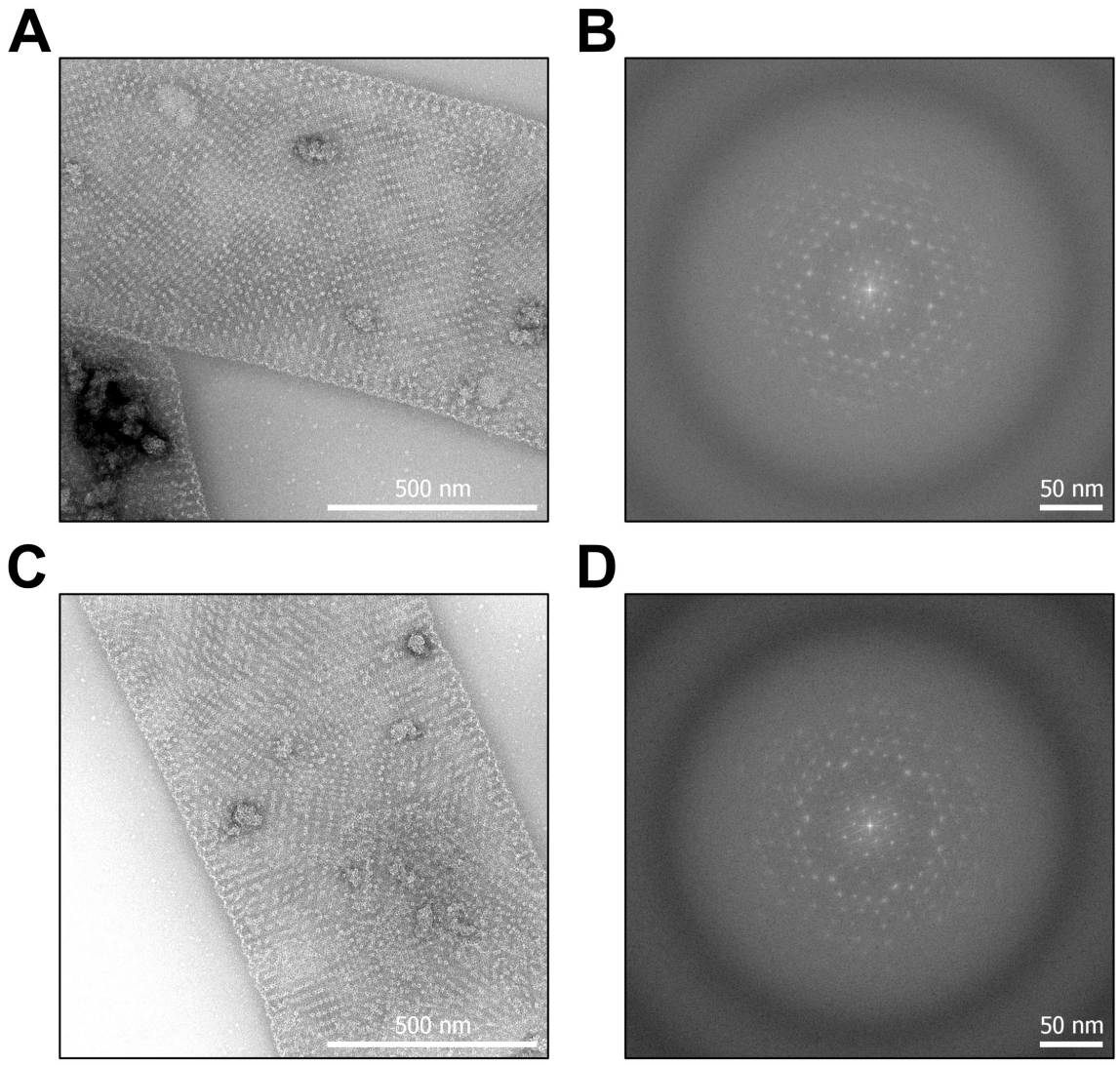


Figure 53. ***Pyrobaculum aerophilum*** **S-layer lattices purified with the implemented protocol negative staining micrographs. A.** and **C.** Negative staining micrographs displaying intact S-layer portions with a scale bar or 500 nm at the bottom right side. **B.** and **C.** Power spectrum of the micrographs in **A.** and **B.** showing the p6 symmetry diffraction pattern with 50 nm scale bar at the bottom right side.

On the other hand, dilution of SDS resulted helpful for the FM-AFM scanning. Even though the S-layer protein lattice was still partially likely covered with SDS, the structure is still visible. The S-layer derived from the last preparation was scanned from an air-dried mica and rehydrated and scanned under ultra-filtered water. Despite the two different measuring techniques, the S-layer lattice remained intact hinting that conformational changes due to rehydration are unlikely to occur. In the FM-AFM measurements many p6 symmetry elements were visible with a length of approximately 200 nm. The lattice constants were also calculated by measuring the centre-to-centre spacing between hexamers. An average spacing of (25.3 ± 1.7) nm and (26.0 ± 1.3) nm was found for the S-layer protein structure in air and liquid, respectively which is in agreement with the negative staining TEM micrographs.

In conclusion, the new qPlus-based FM-AFM setup for surface imaging in ambient and aqueous conditions has successfully imaged biological samples. *P. aerophilum* surface layer protein lattices extracted from the cell membrane have constituted a suitable biological sample for long scanning FM-

AFM period concerning robustness and low propensity towards degradation. Nevertheless, the best features of S-layer lattices also represented a limitation due to the lack of conformational changes in the two conditions, making it unsuitable for protein investigation. Furthermore, we demonstrate that FM-AFM can yield structural information on biological systems under biologically relevant conditions, achieving spatial resolution sufficient to decipher the structural organization of highly symmetric protein lattices like the S-layer. This might add FM-AFM to the advanced techniques providing high resolution details of the sub-10 nm range for matter or biological samples.

6. Bibliography

1. Crick, F. H. C. The Genetic Code. *Sci Am* **207**, 66–77 (1962).
2. Osman, S. & Cramer, P. Structural Biology of RNA Polymerase II Transcription: 20 Years On. *Annu. Rev. Cell Dev. Biol.* **36**, 1–34 (2020).
3. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**, 85–98 (2011).
4. Griesenbeck, J., Tschochner, H. & Grohmann, D. Structure and Function of RNA Polymerases and the Transcription Machineries. in *Macromolecular Protein Complexes* (eds. Harris, J. R. & Marles-Wright, J.) vol. 83 225–270 (Springer International Publishing, Cham, 2017).
5. Chen, J., Boyaci, H. & Campbell, E. A. Diverse and unified mechanisms of transcription initiation in bacteria. *Nat Rev Microbiol* **19**, 95–109 (2021).
6. Gehring, A. M., Walker, J. E. & Santangelo, T. J. Transcription Regulation in Archaea. *J Bacteriol* **198**, 1906–1917 (2016).
7. Decker, K. B. & Hinton, D. M. Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annu. Rev. Microbiol.* **67**, 113–139 (2013).
8. Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).
9. Campbell, E. A. *et al.* Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. *Molecular Cell* **9**, 527–539 (2002).
10. Kohler, R., Mooney, R. A., Mills, D. J., Landick, R. & Cramer, P. Architecture of a transcribing-translating expressome. *Science* **356**, 194–197 (2017).

11. Blaha, G. M. & Wade, J. T. Transcription-Translation Coupling in Bacteria. *Annu. Rev. Genet.* **56**, 187–205 (2022).

12. Wang, C. *et al.* Structural basis of transcription-translation coupling. *Science* **369**, 1359–1365 (2020).

13. Webster, M. W. *et al.* Structural basis of transcription-translation coupling and collision in bacteria. *Science* **369**, 1355–1359 (2020).

14. Molodtsov, V., Wang, C., Firlar, E., Kaelber, J. T. & Ebright, R. H. Structural basis of Rho-dependent transcription termination. *Nature* **614**, 367–374 (2023).

15. Hao, Z., Svetlov, V. & Nudler, E. Rho-dependent transcription termination: a revisionist view. *Transcription* **12**, 171–181 (2021).

16. Pitts, S. & Laiho, M. Regulation of RNA Polymerase I Stability and Function. *Cancers* **14**, 5776 (2022).

17. Zhou, M. & Law, J. A. RNA Pol IV and V in gene silencing: Rebel polymerases evolving away from Pol II's rules. *Current Opinion in Plant Biology* **27**, 154–164 (2015).

18. Fernández-Tornero, C. *et al.* Crystal structure of the 14-subunit RNA polymerase I. *Nature* **502**, 644–649 (2013).

19. Hoffmann, N. A. *et al.* Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**, 231–236 (2015).

20. Girbig, M., Misiaszek, A. D. & Müller, C. W. Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nat Rev Mol Cell Biol* **23**, 603–622 (2022).

21. Moorefield, B., Greene, E. A. & Reeder, R. H. RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4724–4729 (2000).

22. Blattner, C. *et al.* Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev.* **25**, 2093–2105 (2011).

23. Sadian, Y. *et al.* Structural insights into transcription initiation by yeast RNA polymerase I. *The EMBO Journal* **36**, 2698–2709 (2017).

24. Baudin, F. *et al.* Mechanism of RNA polymerase I selection by transcription factor UAF. *Sci. Adv.* **8**, eabn5725 (2022).

25.Liu, X., Bushnell, D. A. & Kornberg, R. D. RNA polymerase II transcription: Structure and mechanism. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**, 2–8 (2013).

26.Hahn, S. & Young, E. T. Transcriptional Regulation in *Saccharomyces cerevisiae* : Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. *Genetics* **189**, 705–736 (2011).

27.Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129–143 (2015).

28.Bowman, E. A. & Kelly, W. G. RNA Polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. *Nucleus* **5**, 224–236 (2014).

29.Decker, T.-M. Mechanisms of Transcription Elongation Factor DSIF (Spt4–Spt5). *Journal of Molecular Biology* **433**, 166657 (2021).

30.Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nat Struct Mol Biol* **24**, 809–815 (2017).

31.Qiu, Y. & Gilmour, D. S. Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing. *Journal of Biological Chemistry* **292**, 5555–5570 (2017).

32.Peterlin, B. M. & Price, D. H. Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell* **23**, 297–305 (2006).

33.Selth, L. A., Sigurdsson, S. & Svejstrup, J. Q. Transcript Elongation by RNA Polymerase II. *Annu. Rev. Biochem.* **79**, 271–293 (2010).

34.Wada, T. *et al.* FACT Relieves DSIF/NELF-Mediated Inhibition of Transcriptional Elongation and Reveals Functional Differences between P-TEFb and TFIIH. *Molecular Cell* **5**, 1067–1072 (2000).

35.Eaton, J. D. & West, S. An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. *Transcription* **9**, 321–326 (2018).

36.Eaton, J. D. *et al.* Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev* **32**, 127–139 (2018).

37.Sakonju, S., Bogenhagen, D. F. & Brown, D. D. A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5′ border of the region. *Cell* **19**, 13–25 (1980).

38. Bogenhagen, D. F., Sakonju, S. & Brown, D. D. A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3′ border of the region. *Cell* **19**, 27–35 (1980).

39. Das, G., Henning, D., Wright, D. & Reddy, R. Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. *The EMBO Journal* **7**, 503–512 (1988).

40. Kunkel, G. R. & Pederson, T. Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used. *Genes Dev.* **2**, 196–204 (1988).

41. Dergai, O. & Hernandez, N. How to Recruit the Correct RNA Polymerase? Lessons from snRNA Genes. *Trends in Genetics* **35**, 457–469 (2019).

42. Brochier-Armanet, C., Forterre, P. & Gribaldo, S. Phylogeny and evolution of the Archaea: one hundred genomes later. *Current Opinion in Microbiology* **14**, 274–281 (2011).

43. Kettenberger, H., Armache, K.-J. & Cramer, P. Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. *Molecular Cell* **16**, 955–965 (2004).

44. Jun, S.-H. *et al.* Direct binding of TFEα opens DNA binding cleft of RNA polymerase. *Nat Commun* **11**, 6123 (2020).

45. Wojtas, M. N., Mogni, M., Millet, O., Bell, S. D. & Abrescia, N. G. A. Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Research* **40**, 9941–9952 (2012).

46. Korkhin, Y. *et al.* Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS Biol* **7**, e1000102 (2009).

47. Pilotto, S. *et al.* Structural basis of RNA polymerase inhibition by viral and host factors. *Nat Commun* **12**, 5523 (2021).

48. Grohmann, D. *et al.* The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Molecular Cell* **43**, 263–274 (2011).

49. Hirtreiter, A. *et al.* Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Research* **38**, 4040–4051 (2010).

50. Hirata, A., Klein, B. J. & Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**, 851–854 (2008).

51. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719 (2002).

52. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell* **127**, 941–954 (2006).

53. Tan, L., Wiesler, S., Trzaska, D., Carney, H. C. & Weinzierl, R. O. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J Biol* **7**, 40 (2008).

54. Weinzierl, R. O. J. The RNA Polymerase Factory and Archaeal Transcription. *Chem. Rev.* **113**, 8350–8376 (2013).

55. Naji, S., Bertero, M. G., Spitalny, P., Cramer, P. & Thomm, M. Structure–function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Research* **36**, 676–687 (2008).

56. Cramer, P. *et al.* Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **37**, 337–352 (2008).

57. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science* **292**, 1863–1876 (2001).

58. Yuzenkova, Y. *et al.* Stepwise mechanism for transcription fidelity. *BMC Biol* **8**, 54 (2010).

59. Zhang, J., Palangat, M. & Landick, R. Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol* **17**, 99–104 (2010).

60. Vassylyev, D. G. *et al.* Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**, 163–168 (2007).

61. Kaplan, C. D., Larsson, K.-M. & Kornberg, R. D. The RNA Polymerase II Trigger Loop Functions in Substrate Selection and Is Directly Targeted by α-Amanitin. *Molecular Cell* **30**, 547–556 (2008).

62. Kaplan, C. D., Jin, H., Zhang, I. L. & Belyanin, A. Dissection of Pol II Trigger Loop Function and Pol II Activity–Dependent Control of Start Site Selection In Vivo. *PLoS Genet* **8**, e1002627 (2012).

63. Pettersen, E. F. *et al.* UCSF ChimeraX : Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82 (2021).

64. Nagy, J. *et al.* Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat Commun* **6**, 6161 (2015).

65. Bushnell, D. A. & Kornberg, R. D. Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: Implications for the initiation of transcription. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6969–6973 (2003).

66. Hirtreiter, A., Grohmann, D. & Werner, F. Molecular mechanisms of RNA polymerase—the F/E (RPB4/7) complex is required for high processivity in vitro. *Nucleic Acids Research* **38**, 585–596 (2010).

67. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016).

68. Goede, B., Naji, S., Von Kampen, O., Ilg, K. & Thomm, M. Protein-Protein Interactions in the Archaeal Transcriptional Machinery. *Journal of Biological Chemistry* **281**, 30581–30592 (2006).

69. Werner, F. & Weinzierl, R. O. J. A Recombinant RNA Polymerase II-like Enzyme Capable of Promoter-Specific Transcription. *Molecular Cell* **10**, 635–646 (2002).

70. Blombach, F., Smollett, K. L., Grohmann, D. & Werner, F. Molecular Mechanisms of Transcription Initiation—Structure, Function, and Evolution of TFE/TFIIE-Like Factors and Open Complex Formation. *Journal of Molecular Biology* **428**, 2592–2606 (2016).

71. Armache, J.-P. *et al.* Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Research* **41**, 1284–1293 (2013).

72. Gelsinger, D. R. *et al.* Ribosome profiling in archaea reveals leaderless translation, novel translational initiation sites, and ribosome pausing at single codon resolution. *Nucleic Acids Research* **48**, 5201–5216 (2020).

73. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).

74. Gietl, A. *et al.* Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Research* **42**, 6219–6231 (2014).

75. Hickey, A. J., De Macario, E. C. & Macario, A. J. L. Transcription in the Archaea: Basal Factors, Regulation, and Stress Gene Expression. *Critical Reviews in Biochemistry and Molecular Biology* **37**, 199–258 (2002).

76. Soppa, J. Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* **31**, 1295–1305 (1999).

77. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127–134 (1996).

78. Rengachari, S. *et al.* Structural basis of SNAPc-dependent snRNA transcription initiation by RNA polymerase II. *Nat Struct Mol Biol* **29**, 1159–1169 (2022).

79. Qureshi, S. A. & Jackson, S. P. Sequence-Specific DNA Binding by the S. shibatae TFIIB Homolog, TFB, and Its Effect on Promoter Strength. *Molecular Cell* **1**, 389–400 (1998).

80. Littlefield, O., Korkhin, Y. & Sigler, P. B. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13668–13673 (1999).

81. Lee, S. & Hahn, S. Model for binding of transcription factor TFIIB to the TBP-DNA complex. *Nature* **376**, 609–612 (1995).

82. Kawarabayasi, Y. Complete Genome Sequence of an Aerobic Hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1 (Supplement). *DNA Research* **6**, 145–152 (1999).

83. Hausner, W., Frey, G. & Thomm, M. Control regions of an archaeal gene. *Journal of Molecular Biology* **222**, 495–508 (1991).

84. Bell, S. D., Kosa, P. L., Sigler, P. B. & Jackson, S. P. Orientation of the transcription preinitiation complex in Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13662–13667 (1999).

85. Blombach, F. *et al.* Structural and functional adaptation of Haloferax volcanii TFEα/β. *Nucleic Acids Research* **46**, 2308–2320 (2018).

86. Blombach, F. *et al.* Archaeal TFEα/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *eLife* **4**, e08378 (2015).

87. Grünberg, S., Bartlett, M. S., Naji, S. & Thomm, M. Transcription Factor E Is a Part of Transcription Elongation Complexes. *Journal of Biological Chemistry* **282**, 35482–35490 (2007).

88. Klein, B. J. *et al.* RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 546–550 (2011).

89. Liu, B. & Steitz, T. A. Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res* **45**, 968–974 (2017).

90. Schulz, S. *et al.* TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl. Acad. Sci. U.S.A.* **113**, (2016).

91. Sanders, T. J. *et al.* TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin: Transcription through archaeal chromatin. *Mol Microbiol* **111**, 784–797 (2019).

92. Yue, L. *et al.* The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3′-end cleavage mode. *Nucleic Acids Research* **48**, 9589–9605 (2020).

93. Wang, C. *et al. Structural Basis of Archaeal FttA-Dependent Transcription Termination*. http://biorxiv.org/lookup/doi/10.1101/2023.08.09.552649 (2023) doi:10.1101/2023.08.09.552649.

94. You, L. *et al.* Structural basis of archaeal FttA-dependent transcription termination. *Nature* (2024) doi:10.1038/s41586-024-07979-9.

95. Weixlbaumer, A., Grünberger, F., Werner, F. & Grohmann, D. Coupling of Transcription and Translation in Archaea: Cues From the Bacterial World. *Front. Microbiol.* **12**, 661827 (2021).

96. Kapanidis, A. N. *et al.* Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science* **314**, 1144–1147 (2006).

97. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex. *Science* **296**, 1285–1290 (2002).

98. Kostrewa, D. *et al.* RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).

99. Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA Polymerase II–TFIIB Complex and the Transcription Initiation Mechanism. *Science* **327**, 206–209 (2010).

100. Kalogeraki, V. S., Tornaletti, S., Cooper, P. K. & Hanawalt, P. C. Comparative TFIIS-mediated transcript cleavage by mammalian RNA polymerase II arrested at a lesion in different transcription systems. *DNA Repair* **4**, 1075–1087 (2005).

101. Izban, M. G. & Luse, D. S. The increment of SII-facilitated transcript cleavage varies dramatically between elongation competent and incompetent RNA polymerase II ternary complexes. *J Biol Chem* **268**, 12874–12885 (1993).

102. Fouqueau, T. *et al.* The transcript cleavage factor paralogue TFS4 is a potent RNA polymerase inhibitor. *Nat Commun* **8**, 1914 (2017).

103. Lange, U. & Hausner, W. Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Molecular Microbiology* **52**, 1133–1143 (2004).

104. Jayasinghe, O. T., Mandell, Z. F., Yakhnin, A. V., Kashlev, M. & Babitzke, P. Transcriptome-Wide Effects of NusA on RNA Polymerase Pausing in Bacillus subtilis. *J Bacteriol* **204**, e00534-21 (2022).

105. Yakhnin, A. V. & Babitzke, P. NusA-stimulated RNA polymerase pausing and termination participates in the *Bacillus subtilis trp* operon attenuation mechanism *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11067–11072 (2002).

106. Mondal, S., Yakhnin, A. V. & Babitzke, P. Modular Organization of the NusA- and NusG-Stimulated RNA Polymerase Pause Signal That Participates in the Bacillus subtilis trp Operon Attenuation Mechanism. *J Bacteriol* **199**, (2017).

107. Shibata, R. *et al.* Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. *Biochemical and Biophysical Research Communications* **355**, 122–128 (2007).

108. Phung, D. K. *et al.* Archaeal NusA2 is the ancestor of ribosomal protein eS7 in eukaryotes. *Structure* **33**, 149-159.e6 (2025).

109. You, L. *et al.* Structural basis for intrinsic transcription termination. *Nature* **613**, 783–789 (2023).

110. Yarnell, W. S. & Roberts, J. W. Mechanism of Intrinsic Transcription Termination and Antitermination. *Science* **284**, 611–615 (1999).

111. Grünberger, F. *et al.* Uncovering the temporal dynamics and regulatory networks of thermal stress response in a hyperthermophile using transcriptomics and proteomics. *mBio* **14**, e02174-23 (2023).

112. Dar, D., Prasse, D., Schmitz, R. A. & Sorek, R. Widespread formation of alternative 3′ UTR isoforms via transcription termination in archaea. *Nat Microbiol* **1**, 16143 (2016).

113. Santangelo, T. J. & Reeve, J. N. Archaeal RNA Polymerase is Sensitive to Intrinsic Termination Directed by Transcribed and Remote Sequences. *Journal of Molecular Biology* **355**, 196–210 (2006).

114. Santangelo, T. J., Cubonová, L., Skinner, K. M. & Reeve, J. N. Archaeal Intrinsic Transcription Termination In Vivo. *J Bacteriol* **191**, 7102–7108 (2009).

115. Reiter, W.-D., Palm, P. & Zillig, W. Transcription termination in the archaebacterium *Sulfolobus* : signal structures and linkage to transcription initiation. *Nucl Acids Res* **16**, 2445–2460 (1988).

116. Walker, J. E., Luyties, O. & Santangelo, T. J. Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci. U.S.A.* **114**, (2017).

117. Wenck, B. R. & Santangelo, T. J. Archaeal transcription. *Transcription* **11**, 199–210 (2020).

118. Sanders, T. J. *et al.* FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat Microbiol* **5**, 545–553 (2020).

119. Johnson, G. E., Lalanne, J.-B., Peters, M. L. & Li, G.-W. Functionally uncoupled transcription–translation in Bacillus subtilis. *Nature* **585**, 124–128 (2020).

120. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088–5090 (1977).

121. Ricardo Chiavicchioli. *Archaea: Molecular and Cellular Biology*.

122. Duller, S. *et al.* Targeted isolation of Methanobrevibacter strains from fecal samples expands the cultivated human archaeome. *Nat Commun* **15**, 7593 (2024).

123. Fiala, G. & Stetter, K. O. Pyrococcus furiosus sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* **145**, 56–61 (1986).

124. Lundberg, K. S. *et al.* High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* **108**, 1–6 (1991).

125. *The Proteomics Protocols Handbook*. (Humana Press, Totowa, NJ, 2005). doi:10.1385/1592598900.

126. David Pöllmann. Anwendung des genetischen Pyrococcus Systems zur Gewinnung von Mutanten der RNA-Polymerase aus Pyrococcus furiosus. (Universität Regensburg, 2010).

127. Burley, S. K. *et al.* Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Research* **53**, D564–D574 (2025).

128. Passmore, L. A. & Russo, C. J. Specimen Preparation for High-Resolution Cryo-EM. in *Methods in Enzymology* vol. 579 51–86 (Elsevier, 2016).

129. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat Methods* **16**, 471–477 (2019).

130. Scheres, S. H. W. Amyloid structure determination in *RELION* -3.1. *Acta Crystallogr D Struct Biol* **76**, 94–101 (2020).

131. Tarău, D. *et al.* Structural basis of archaeal RNA polymerase transcription elongation and Spt4/5 recruitment. *Nucleic Acids Research* **52**, 6017–6035 (2024).

132. Nakane, T. & Scheres, S. H. W. Multi-body Refinement of Cryo-EM Images in RELION. in *cryoEM* (eds. Gonen, T. & Nannenga, B. L.) vol. 2215 145–160 (Springer US, New York, NY, 2021).

133. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

134. Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* **213**, 107702 (2021).

135. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**, 519–530 (2012).

136. Zhang, K. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* **193**, 1–12 (2016).

137. Schwab, J., Kimanius, D., Burt, A., Dendooven, T. & Scheres, S. H. W. DynaMight: estimating molecular motions with improved reconstruction from cryo-EM images. *Nat Methods* (2024) doi:10.1038/s41592-024-02377-5.

138. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

139. Waege, I., Schmid, G., Thumann, S., Thomm, M. & Hausner, W. Shuttle Vector-Based Transformation System for *Pyrococcus furiosus*. *Appl Environ Microbiol* **76**, 3308–3313 (2010).

140. Sommer, B. *et al.* Activation of a Chimeric Rpb5/RpoH Subunit Using Library Selection. *PLoS ONE* **9**, e87485 (2014).

141. Lee, P. Y., Costumbrado, J., Hsu, C.-Y. & Kim, Y. H. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *JoVE* 3923 (2012) doi:10.3791/3923.

142. Green, M. R. & Sambrook, J. Analysis of DNA by Agarose Gel Electrophoresis. *Cold Spring Harb Protoc* **2019**, pdb.top100388 (2019).

143. Shapiro, A. L., Viñuela, E. & V. Maizel, J. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochemical and Biophysical Research Communications* **28**, 815–820 (1967).

144. Shapiro, A. L. & Maizel, J. V. Molecular weight estimation of polypeptides by SDS-polyacrylamide gel electrophoresis: Further data concerning resolving power and general considerations. *Analytical Biochemistry* **29**, 505–514 (1969).

145. Dunker, A. K. & Rueckert, R. R. Observations on molecular weight determinations on polyacrylamide gel. *J Biol Chem* **244**, 5074–5080 (1969).

146. Matsumoto, H., Haniu, H. & Komori, N. Determination of Protein Molecular Weights on SDS-PAGE. in *Electrophoretic Separation of Proteins* (eds. Kurien, B. T. & Scofield, R. H.) vol. 1855 101–105 (Springer New York, New York, NY, 2019).

147. Stowell, J. A. W., Tang, T. T. L., Seidel, M. & Passmore, L. A. Gel-Based Analysis of Protein-Nucleic Acid Interactions. *Methods Mol Biol* **2263**, 321–339 (2021).

148. Albright, L. M. & Slatko, B. E. Denaturing Polyacrylamide Gel Electrophoresis. *CP Nucleic Acid Chemistry* **00**, (2000).

149. Maniatis, T., Jeffrey, A. & Van deSande, H. Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. *Biochemistry* **14**, 3787–3794 (1975).

150. Flett, F. & Interthal, H. Separation of DNA Oligonucleotides Using Denaturing Urea PAGE. in *DNA Electrophoresis* (ed. Makovets, S.) vol. 1054 173–185 (Humana Press, Totowa, NJ, 2013).

151. Gibson, D. G. Enzymatic Assembly of Overlapping DNA Fragments. in *Methods in Enzymology* vol. 498 349–361 (Elsevier, 2011).

152. Florian Benedict Johannes Maria Heiß. Heterologe Expression der Elongationsfaktoren Spt4 und Spt5 aus Pyrococcus furiosus. (Universität Regensburg, 2014).

153. Kielkopf, C. L., Bauer, W. & Urbatsch, I. L. Expression of Cloned Genes in *E. coli* Using IPTG-Inducible Promoters. *Cold Spring Harb Protoc* **2021**, pdb.prot102137 (2021).

154. Elia, N. Using unnatural amino acids to selectively label proteins for cellular imaging: a cell biologist viewpoint. *The FEBS Journal* **288**, 1107–1117 (2021).

155. Lai, Y.-T. *et al.* Rapid labeling of intracellular His-tagged proteins in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2948–2953 (2015).

156. Kim, Y. *et al.* Efficient Site-Specific Labeling of Proteins via Cysteines. *Bioconjugate Chem.* **19**, 786–791 (2008).

157. Nanda, J. S. & Lorsch, J. R. Labeling of a Protein with Fluorophores Using Maleimide Derivitization. in *Methods in Enzymology* vol. 536 79–86 (Elsevier, 2014).

158. Pilsl, M. *et al.* Preparation of RNA Polymerase Complexes for Their Analysis by Single-Particle Cryo-Electron Microscopy. in *Ribosome Biogenesis* (ed. Entian, K.-D.) vol. 2533 81–96 (Springer US, New York, NY, 2022).

159. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).

160. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* **1077**, 101–113 (2014).

161. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *ASSAY and Drug Development Technologies* **9**, 342–353 (2011).

162. Pauling, L. The Structure of the Micas and Related Minerals. *Proc. Natl. Acad. Sci. U.S.A.* **16**, 123–129 (1930).

163. Kanno, H., Speedy, R. J. & Angell, C. A. Supercooling of Water to -92°C Under Pressure. *Science* **189**, 880–881 (1975).

164. Glaeser, R. M. Proteins, interfaces, and cryo-EM grids. *Current Opinion in Colloid & Interface Science* **34**, 1–8 (2018).

165. Iqbal, M., Rafiq, M., Bhatti, S. A. & Aleem, F.-. The electron beam gun with thermionic hairpin-like cathode for welding and surface modifications. *Vacuum* **77**, 19–26 (2004).

166. Konings, S. & Bischoff, M. A Cold Field Emission Gun Optimized for Cryo-EM Applications. *Microsc Microanal* **26**, 566–567 (2020).

167. Zhukova, E. S. *et al.* Out-of-equilibrium electrons lead to record thermionic emission in LaB6 with the Jahn-Teller instability of boron cage. Preprint at https://doi.org/10.48550/ARXIV.1904.07132 (2019).

168. Peng, D. & Nakashima, P. N. H. Identification of the impurity phase in high-purity CeB_6 by convergent-beam electron diffraction. *Acta Crystallogr A Found Adv* **75**, 489–500 (2019).

169. Crewe, A. V., Eggenberger, D. N., Wall, J. & Welter, L. M. Electron Gun Using a Field Emission Source. *Review of Scientific Instruments* **39**, 576–583 (1968).

170. Olshin, P. K., Drabbels, M. & Lorenz, U. J. Characterization of a time-resolved electron microscope with a Schottky field emission gun. *Structural Dynamics* **7**, 054304 (2020).

171. Henderson, R. *et al.* Digitisation of electron microscope films: Six useful tests applied to three film scanners. *Ultramicroscopy* **107**, 73–80 (2007).

172. Faruqi, A. R. & McMullan, G. Electronic detectors for electron microscopy. *Quart. Rev. Biophys.* **44**, 357–390 (2011).

173. Faruqi, A. R. & Subramaniam, S. CCD detectors in high-resolution biological electron microscopy. *Quart. Rev. Biophys.* **33**, 1–27 (2000).

174. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216–221 (2015).

175. Emsley, P. & Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).

176. Croll, T. I. *ISOLDE* : a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).

177. Adams, P. D. *et al. PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).

178. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **27**, 293–315 (2018).

179. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity: RNA polymerase-Spt4/5 complex architecture. *The EMBO Journal* **30**, 1302–1310 (2011).

180. Jun, S.-H. *et al.* The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat Commun* **5**, 5132 (2014).

181. Chakraborty, A. *et al.* Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science* **337**, 591–595 (2012).

182. Duchi, D., Mazumder, A., Malinen, A. M., Ebright, R. H. & Kapanidis, A. N. The RNA polymerase clamp interconverts dynamically among three states and is stabilized in a partly closed state by ppGpp. *Nucleic Acids Research* **46**, 7284–7295 (2018).

183. Lin, W. *et al.* Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). *Molecular Cell* **70**, 60-71.e15 (2018).

184. Finn, R. D. Escherichia coli RNA polymerase core and holoenzyme structures. *The EMBO Journal* **19**, 6833–6844 (2000).

185. Aibara, S., Dienemann, C. & Cramer, P. Structure of an inactive RNA polymerase II dimer. *Nucleic Acids Research* **49**, 10747–10755 (2021).

186. Sheppard, C. *et al.* Repression of RNA polymerase by the archaeo-viral regulator ORF145/RIP. *Nat Commun* **7**, 13595 (2016).

187. Barnes, C. O. *et al.* Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble. *Molecular Cell* **59**, 258–269 (2015).

188. Wang, D. *et al.* Structural Basis of Transcription: Backtracked RNA Polymerase II at 3.4 Angstrom Resolution. *Science* **324**, 1203–1206 (2009).

189. Ehara, H. *et al.* Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921–924 (2017).

190. Anderson, S. J. *et al.* The Transcription Elongation Factor Spt5 Influences Transcription by RNA Polymerase I Positively and Negatively. *Journal of Biological Chemistry* **286**, 18816–18824 (2011).

191. Delbeau, M. *et al.* Structural and functional basis of the universal transcription factor NusG pro-pausing activity in Mycobacterium tuberculosis. *Molecular Cell* **83**, 1474-1488.e8 (2023).

192. Vishwakarma, R. K., Qayyum, M. Z., Babitzke, P. & Murakami, K. S. Allosteric mechanism of transcription inhibition by NusG-dependent pausing of RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218516120 (2023).

193. Dollinger, R. *et al.* Assessment of the roles of Spt5-nucleic acid contacts in promoter proximal pausing of RNA polymerase II. *Journal of Biological Chemistry* **299**, 105106 (2023).

194. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).

195. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153–1160 (2019).

196. Coureux, P.-D., Lazennec-Schurdevin, C., Bourcier, S., Mechulam, Y. & Schmitt, E. Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. *Commun Biol* **3**, 58 (2020).

197. Schmitt, E., Coureux, P.-D., Monestier, A., Dubiez, E. & Mechulam, Y. Start Codon Recognition in Eukaryotic and Archaeal Translation Initiation: A Common Structural Core. *IJMS* **20**, 939 (2019).

198. Kazan, R. *et al.* Role of aIF5B in archaeal translation initiation. *Nucleic Acids Research* **50**, 6532–6548 (2022).

199. Martin Brehm. Klonierung, Expression und Reinigung archaeeller Translationsinitiationsfaktoren. (Universität Regensburg, 2019).

200. Rego, N. & Koes, D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* **31**, 1322–1324 (2015).

201. Gogoi, P. & Kanaujia, S. P. Archaeal and eukaryal translation initiation factor 1 differ in their RNA interacting loops. *FEBS Letters* **592**, 1602–1610 (2018).

202. Messner, P., Steiner, K., Zarschler, K. & Schäffer, C. S-layer nanoglycobiology of bacteria. *Carbohydrate Research* **343**, 1934–1951 (2008).

203. Assandri, M. H., Malamud, M., Trejo, F. M. & Serradell, M. D. L. A. S-layer proteins as immune players: Tales from pathogenic and non-pathogenic bacteria. *Current Research in Microbial Sciences* **4**, 100187 (2023).

204. Bradshaw, W. J., Roberts, A. K., Shone, C. C. & Acharya, K. R. The structure of the S-layer of Clostridium difficile. *J. Cell Commun. Signal.* **12**, 319–331 (2018).

205. Raff, J. *et al.* S-Layer-Based Nanocomposites for Industrial Applications. in *Protein-based Engineered Nanostructures* (eds. Cortajarena, A. L. & Grove, T. Z.) vol. 940 245–279 (Springer International Publishing, Cham, 2016).

206. Schuster, B. & Sleytr, U. B. S-Layer Ultrafiltration Membranes. *Membranes* **11**, 275 (2021).

207. Sleytr, U. B., Schuster, B., Egelseer, E.-M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol Rev* **38**, 823–864 (2014).

208. Völkl, P. *et al.* Pyrobaculum aerophilum sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Appl Environ Microbiol* **59**, 2918–2926 (1993).

209. Giessibl, F. J. High-speed force sensor for force microscopy and profilometry utilizing a quartz tuning fork. *Applied Physics Letters* **73**, 3956–3958 (1998).

210. Seeholzer, T. *et al.* A Next-Generation qPlus-Sensor-Based AFM Setup: Resolving Archaeal S-Layer Protein Structures in Air and Liquid. *J. Phys. Chem. B* **127**, 6949–6957 (2023).

.