Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

B.Sc. Agata D. Misiaszek

born in: Wrocław, Poland

Oral examination: 15$^{\text{th}}$ of September 2022

Structural studies of the human RNA polymerase I and the role of human RNA polymerase I and III in diseases.

Referees:

Prof. Dr. Irmgard Sinning

Prof. Dr. Janosch Hennig

SUMMARY

In eukaryotes, the task of RNA synthesis is divided between three RNA polymerases (RNAPs). RNA polymerase I (Pol I) produces ribosomal RNA (rRNA), and Pol III specialises in making short, non-transcribed RNAs, which together constitute the majority of the RNA pool. Dysregulation of Pol I and Pol III function leads to complex diseases such as cancer or developmental disorders like Treacher Collins Syndrome or Hypomyelinating Leukodystrophy. Traditionally, RNAPs have been studied in *Saccharomyces cerevisiae*. Yet, to gain insights into the complex human disorders and allow targeted drug design, understanding the structure and function of human RNAPs is crucial. Here, I present insights into the structure and function of the human Pol I and explore the role of human Pol I and Pol III in the aetiology of developmental disorders.

Firstly, I employed the CRISPR-Cas9 genome editing technique to endogenously tag a subunit shared exclusively between Pol I and III (the RPAC1 subunit). Using introduced affinity tags, I purified native complexes to high homogeneity. Additionally, recombinant expression systems for producing human Pol I and its co-factors have been established.

Cryo-electron microscopy (cryo-EM) allowed me to determine the structure of human Pol I in various functional states such as an elongating complex, bound to initiation factor RRN3 and bound to an open DNA scaffold, at 2.7 to 3.3 Å resolution. In the elongating Pol I structure, I could observe a double-stranded RNA in the RNA exit tunnel, which may support Pol I processivity. Analysis of the stalk sub-complex revealed that in humans, it constitutes only one subunit allowing the stalk to bend, for example, upon binding the initiation factor RRN3. Different modes of the human Pol I inactivation could also be explored.

Finally, I could use the high-resolution structure of the human Pol I, and the recently solved structure of the human Pol III, to map disease-associated mutations. While some of those mutations are located in the subunits shared between Pol I and III, they might affect the two complexes differently. To further study the role of those mutations, I used the CRISPR-Cas9 system to introduce selected mutations into the endogenous subunits. Using inserted tags, I could study the effects of mutations on the localisation and assembly of the human Pol I and Pol III complexes. Further studies will help in understanding the molecular phenotypes underlying complex developmental disorders.

ZUSAMMENFASSUNG

Die RNA Synthese ist in Eukaryonten unter drei RNA Polymerasen (RNAPs) aufgeteilt. Die RNA Polymerase I (Pol I) produziert ribosomale RNA (rRNA), und Pol III ist auf die Herstellung kurzer, nicht transkribierter RNAs spezialisiert, die zusammen den Großteil der gesamten RNA ausmachen. Fehlregulationen der Pol I- und Pol III-Funktionalität führen zu komplexen Erkrankungen wie zum Beispiel Krebs oder den Entwicklungsstörungen Treacher-Collins-Syndrom oder hypomyelinisierende Leukodystrophie. Traditionell wurden RNAPs in *Saccharomyces cerevisiae* untersucht. Um jedoch Einblicke in die komplexen menschlichen Erkrankungen zu gewinnen und ein gezieltes Arzneimitteldesign zu ermöglichen, ist das Verständnis der Struktur und Funktion menschlicher RNAPs von entscheidender Bedeutung. Ich präsentiere hier Details in der Struktur und Funktion der menschlichen Pol I und untersuche die Rolle von Pol I und Pol III in der Ätiologie von Entwicklungsstörungen.

Als erstes habe ich die CRISPR/Cas9-Methode eingesetzt, um eine Untereinheit (RPAC1) endogen zu markieren, die ausschließlich in Pol I und III vorwander ist. Durch hinzugefügte Affinitäts-Tags reinigte ich native Komplexe in hoher Homogenität auf. Außerdem wurden rekombinante Expressionssysteme zur Herstellung von humaner Pol I und ihren Cofaktoren etabliert.

Kryoelektronenmikroskopie (Kryo-EM) erlaubte mir, die Struktur von humaner Pol I in verschiedenen funktionellen Konformationen mit Auflösungen von 2,7 bis 3,3 Å zu bestimmen, wie z.B. als Elongationskomplex, als Komplex mit Initiationsfaktor RRN3 oder als Komplex mit einer entwundenen DNA. In der Pol I-Elongationskomplex-Struktur konnte ich eine doppelsträngige RNA im RNA-Ausgangstunnel detektieren, welche die Pol I-Aktivität unterstützen könnte. Die Analyse des Stalk-Subkomplexes ergab, dass er im Menschen nur aus einer Untereinheit besteht, die es ermöglicht, sich flexibel bei z.B. Bindung des Initiationsfaktors RRN3 zu biegen. Zusätzlich konnten auch verschiedene Modi der menschlichen Pol I-Inaktivierung untersucht werden.

Schließlich konnte ich die hochauflösende Struktur der humanen Pol I und die kürzlich bestimmte Struktur der humanen Pol III nutzen, um krankheitsassoziierte Mutationen zu kartieren. Einige dieser Mutationen sind innerhalb der geteilten Untereinheiten zwischen Pol I und Pol III lokalisiert, sie könnten aber die beiden Komplexe unterschiedlich beeinflussen. Um die Auswirkungen dieser Mutationen weiter zu untersuchen, habe ich die CRISPR/Cas9-Methode verwendet, um ausgewählte Punktmutationen der endogenen Untereinheiten zu erzeugen. Anhand

hinzugefügter Protein-Tags konnte ich die Auswirkungen von Mutationen auf die Lokalisierung und die Zusammensetzung der Pol I- und Pol III-Komplexe untersuchen. Zukünftige Studien werden helfen, die molekularen Phänotypen dieser komplexen Entwicklungsstörungen zu charakterisieren.

PUBLICATIONS

Girbig M.*, **Misiaszek A.D.***, Vorländer M.K., Lafita A., Grötsch H., Baudin F., Bateman A., Müller C.W. (2021). Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states. *Nature Structural and Molecular Biology,* 28(2): 210-19.

* equal contribution

Misiaszek A.D., Girbig M., Grötsch H., Baudin F., Lafita A., Murciano B., Müller C.W. (2021). Cryo-EM structures of human RNA polymerase I. *Nature Structural and Molecular Biology*, 28(12): 997-1008.

Girbig, M., **Misiaszek, A.D.**, Müller, C.W. (2022) Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nature Reviews Molecular Cell Biology*, doi: 10.1038/s41580-022-00476-9.

In preparation

Perrier S., Maegawa G. H. B., **Misiaszek A.D.**, Tran L. T., Müller C.W., Thiffault I., Bernard G. Novel pathogenic variants in POLR3K cause POLR3-related leukodystrophy. Clinical Genetics – Case Report.

ACKNOWLEDGEMENTS

This thesis summarises the immense personal and professional growth I underwent throughout the last four years. My knowledge has expanded, and I have acquired many new skills which will undoubtedly help me in the future. I will fondly remember the time I spent at EMBL, mainly thanks to all the people that helped me along the way, whom I would like to thank here.

Firstly, I would like to thank my supervisor, Christoph Müller for the opportunity to develop projects collected in this thesis. Thank you for your support and help in shaping the ideas, while also giving me the freedom to explore and follow my curiosity.

I am grateful to every member of the Müller group for always being there, ready to answer all the questions and offer their help. The atmosphere in the group made me happy and excited about coming to the lab and I always felt supported in all my endeavours.

I want to thank Helga Grötsch for welcoming me into the group and patiently showing me the basics of everything. Further, thank you for your help with cell culture, CRISPR-Cas9 and for being there to help with anything I could need.

Working alongside Mathias Girbig has been incredibly rewarding. Thank you for showing me all the processing and model-building tips and tricks and for coming on the paper-writing journeys together. Without your input, none of my work would look the way it ultimately does.

I wish to show my appreciation to Florence Baudin for all her help with biochemistry and transcription assays. Thanks to Brice Murciano for producing human RRN3 and being there whenever I needed help with misbehaving Äktas.

A large part of the mutational studies would not exist without the involvement of Granita Lokaj. Thank you for your excitement, motivation and independence. All those hours spent in the cell culture have not gone unnoticed and I am truly grateful I had a chance to work with you. Learning how to share my ideas with you has been one of the most rewarding parts of this project.

Having Wolfram Seifert Davila and Anna Lippert do their PhDs alongside me meant that I always had great energy around. Thank you for all the scientific and personal discussions, and the opportunity to learn from each other. Interactions with all the post-docs, Inessa De, Herman Fung, Inyoung Hwang and Sid Narasimhan, have always been very valuable and gave me a fresh perspective on all aspects of my work.

Also, thanks to Jonas Wiedenhausen for taking over the group's support and your German superpowers, also reflected in the Summary of this thesis.

Throughout my project, in a truly EMBL interdisciplinary spirit, I tried many different techniques with the help of the EMBL Core Facilities. The Proteomics Core Facility has been great at giving out advice and analysing all needed samples for me. The Advanced Light Microscopy Facility, particularly Aliaksandr Halavatyi, Stefan Terjung and Marko Lampe, were always willing to discuss my experiments and help me set up all experiments. Flow Cytometry Core Facility, specifically Daniel Gimenes, sorted all the cells and helped me make the needed cell lines.

Further, I am incredibly grateful to Thomas Hoffmann and the EMBL IT support for maintaining the cluster, answering all my questions and solving issues I encountered. Thanks to Felix Weis and Wim Hagen I had access to perfectly maintained cryo-EM microscopes and could learn from the experts, which resulted in top-quality data. I always appreciated the interest and vast knowledge of Sebastian Eustermann, who helped me develop less standard processing strategies. I am also grateful for administrative support and all help from Alena Fröhlichová.

I was also lucky to get input from many knowledgeable people who shared their protocols and spent time answering many of my questions. In particular, François Dossin and Matteo Trovato helped me set up the CRISPR-Cas9 editing experiments.

I am also grateful to my Thesis Advisory Committee members, Janosch Henning, Jan Kosiński and Ed Hurt, for accompanying me on the journey through this thesis and giving me helpful suggestions. I also acknowledge the long-lasting support of Caroline Lynn Kamerlin.

The last four years have been filled with a lot of fun and great memories, thanks to amazing people from my batch and all friends I made in EMBL and outside, especially: Alberto, Andrea, Anna, Ashwin, Chris, Daniel, Elisa, Felix, Jesús, Kai, Karolina, Luca, Lucia, Maja, Marta, Matteo. I am also very grateful to friends supporting me from afar: Ada, Arek, Asia, Daria, Ewa, Indrė, Kamil, Lina, Lucie, Maciek, Mantas, Martyna, Róża. Thank you for all the chats, laughs and support!

I am also very fortunate to have the constant support from my family, who were always there to listen, give their advice and reassure me. Thank you to my parents for being there for me and staying curious about my work and plans. My time in Heidelberg would not have been the same without my amazing partner, Luigi Vigani. I am incredibly grateful to have you in my life, and thank you for your love and support.

CONTENTS

Chapter I: Introduction

Chapter II: Results and Discussion

Chapter I: Introduction

INTRODUCTION

At the centre of the function of any organism lies the central dogma of molecular biology, which describes the sequential flow of information from DNA to RNA and finally to proteins. Two sequential processes, transcription and translation, govern those transformations [1,2]. DNA-dependent RNA polymerases (RNAPs) are enzymes responsible for the process of transcription since they are responsible for RNA synthesis using a DNA template. Most RNAPs used throughout the tree of life are multisubunit protein complexes derived from one common ancestor [3]. While prokaryotes have only one RNAP, diversification and specialisation of the RNAPs occurred in eukaryotes. Each of the three (or, in the case of plants, five [4]) eukaryotic RNAPs has a different transcriptional profile [3]. RNA polymerase I (Pol I) produces only one type of transcript, the pre-rRNA; Pol II transcribes the most extensive diversity of targets that include both coding and non-coding regulatory sequences, while Pol III synthesises short, non-translated RNAs such as tRNAs. As such, Pol I and Pol III also indirectly play a role in the process of translation since produced by Pol I rRNA constitutes the catalytic portion of the ribosome, and the tRNA produced by Pol III is necessary to read out the consecutive RNA codons and to bring amino acids to the growing polypeptide chain.

In this introduction, I highlight the similarities and differences between the eukaryotic RNAPs, especially emphasising Pol I adaptations to its function. Moreover, I show how changes in Pol I and III function can play a role in diseases. Further, an introduction to the main techniques: cryo-electron microscopy (cryo-EM) and CRISPR-Cas9 gene editing, which are the basis of work collected in this thesis, is included.

1.1 Eukaryotic DNA-dependent RNA polymerases

Most molecular machines that produce RNA in living organisms are multisubunit RNAPs that use the DNA template for RNA production. Given the focus of this thesis, I only describe those canonical complexes. At the same time, it is essential to note that RNA-dependent RNA polymerases can also produce RNAs. They are common in viruses with RNA-based genomes, where they are necessary for both genome duplication and expression [5]. In some species of eukaryotes, RNA-dependent RNA polymerases are also present and can be involved in

RNA interference pathways [6]. Other than multisubunit RNAPs, also small, single subunit RNAPs exist. The most commonly known example of this RNAP type is T7 phage RNAP [7]. Single-subunit RNAPs are also found in mitochondria [8] and chloroplasts [9].

Additionally to the three RNAPs conserved across eukaryotes, plants have two other RNAPs. Pol IV and V are specialised versions of Pol II used in RNA-mediated transcriptional silencing [4]. Those complexes are, however, not covered in this thesis. While mainly eukaryotic RNAPs are discussed in this section, it is essential to note that they share a significant degree of structural similarity with prokaryotic and archaeal RNAPs [3].

1.1.1 The transcription cycle

The primary function of RNAPs is the production of RNA from a DNA template in the process known as transcription. The general process of transcription is conserved between all RNAPs from all domains of life. It can be generally divided into three phases: (i) initiation, (ii) elongation and (iii) termination (Fig. 1.1 a). In a recent review from the Müller group, which I co-authored [10], each stage was reviewed in detail, including similarities and differences between the three RNAPs. The characteristic features of each of those phases include:

(i) **Initiation** – RNAP is recruited to the promoter DNA. It is positioned/activated by general transcription factors (GTFs), which aid in its positioning. RNAP and its co-factors form a pre-initiation complex (PIC) (Fig. 1.1 b). The number and identity of GTFs vary between different RNAPs. Initially, RNAP binds to a double-stranded DNA forming the so-called closed complex (CC) (Fig. 1.1 c). Further, the melting of the double-stranded DNA occurs, and an open transcription bubble is formed, with the template and non-template strands separated from each other. This state is called the open complex (OC) (Fig. 1.1 c). The initial steps of the nucleotide addition cycle are described in detail in [11]. Firstly, the RNAP selects and positions the transcription start site (TSS) inside the active centre. Since TSS distance from the promoter (which determines the site of the opening of the transcription bubble) varies, the TSS is positioned via a scrunching mechanism. Transcription begins with the addition of two nucleoside triphosphates (NTPs) termed the initiating and extending NTP in the active site. They base pair to the

template strand in positions +1 and +2, and the first phosphodiester bond is formed. RNAPs can also initiate transcription using a pre-formed RNA primer in the process of primer-dependent initiation. Nucleotides are added to the growing nascent RNA chain step-wise, and an initially transcribing complex (ITC) is formed (Fig. 1.1 c). After adding each nucleotide to the 3' end of the growing RNA, the template strand is translocated by ~3.4 Å, while the RNAP remains bound to the promoter sequence. When the RNA chain extends to the length of 11-15 nucleotides, the accumulated free energy allows it to overcome the energy of RNAP binding to the promoter, and the promoter escape takes place.

(ii) **Elongation** – The elongation complex (EC) is formed after promoter escape. The nascent RNA chain is extended and RNAP translocates along the DNA. The template DNA-RNA hybrid is 8 – 9 base pairs (bp) long and is tightly bound in the active site. The 5' end of the RNA extends into the RNA exit tunnel. The growing RNA extending out of the RNA exit tunnel is co-transcriptionally folded [12] and processed [13]. The transcription rate is not uniform over the length of the transcribed sequence: RNAPs pause and even backtrack [14]. The speed of transcription is influenced by many factors, such as the DNA sequence or the binding of accessory factors. While *in vitro* all RNAPs exhibit the capability to elongate independently, *in vivo* especially Pol I and II require additional elongation factors to support their processivity.

(iii) **Termination** – At the end of the transcription cycle, the RNAP needs to release the nascent RNA and be released from the DNA template. Termination needs to occur at the correct site to produce a complete product. In Pol I and Pol II additional factors ensure fidelity of termination, while in Pol III a specific DNA sequence is sufficient to induce disassembly of the complex. Transcription termination is the most poorly understood stage of the transcription cycle, with little structural information present. Following termination, RNAPs are suggested to undergo recycling, allowing them to enter the new transcription cycle.

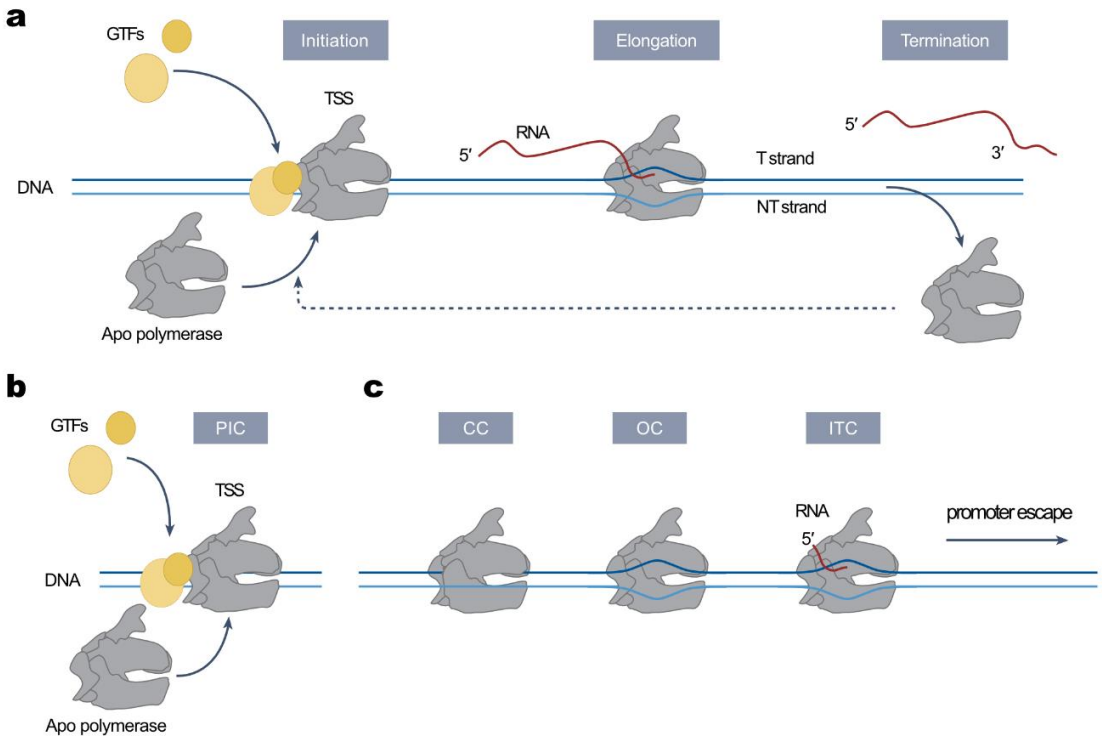

Fig. 1.1. General transcription cycle of the RNAP. a| Cartoon representation of the overall transcription cycle, including initiation, elongation and termination in which RNAPs produce the RNA. GTFs – general transcription factors, TSS – transcription start site, T strand – template strand, NT strand – non-template strand. **b|** GTFs bind to the promoter DNA and recruit RNAP. Together they form a Pre-Initiation Complex (PIC). **c|** Sequence of structural changes during initiation. RNAP bound to double-stranded DNA forms a closed complex (CC). After the opening of the DNA bubble, an Open Complex (OC) is formed. RNA synthesises, and an Initially Transcribing Complex (ITC) is formed, which then escapes from promoter DNA.

1.1.2 Key structural elements of eukaryotic RNA polymerases

RNAPs are highly conserved, multisubunit complexes structurally related across all living organisms [3]. Already early electron microscopy studies revealed the overall shape of the complex that resembles a crab claw with a groove running across the middle of the complex where the DNA could be bound [15,16] (Fig. 1.2 a, b). The core is made up mainly of the two largest subunits, which in eukaryotes are further complemented by smaller subunits, in many cases, shared between all RNAPs [17].

All RNAPs share the same structural elements that allow them to perform their function. Functional elements were first named and described in the atomic structures of Pol II [18–20], and then Pol I [21] and Pol III [22]. Most important, conserved structural elements include (Fig. 1.2):

- **DNA-binding cleft** - positively charged groove formed by two largest RNAP subunits running along the length of the RNAP where DNA is bound. The cleft can open and close (or widen and narrow) thanks to the movement of the clamp, which allows tighter binding of the nucleic acid scaffold. The active site is located at the base of the DNA-binding cleft [20–22].
- **Clamp** - also referred to as "hinged" domain, a mobile element that clamps the DNA. Composed of mainly the N-terminal region of the largest RNAP subunit and the C-terminal portion of the second largest subunit [20–22].
- **Stalk** – a domain formed by two subunits (or one subunit in the case of Pol I in most species, excluding some fungi [23,24]) located on top of the clamp and close to the RNA exit tunnel. It is an assembly platform for initiation factors [21,22].
- **Dock** – part of the largest RNAP subunit involved in binding initiation factors such as TFIIB in the Pol II system [19] or RRN3 in the Pol I system [23].
- **Jaws** – domains located at the end of the DNA-binding cleft on the opposing sites. They contact the downstream DNA and show mobility allowing them to clamp on the downstream DNA [20,21].
- **Wall** – part of the second-largest subunit found directly behind the active site, which re-directs and tilts the DNA-RNA hybrid [20].
- **RNA exit tunnel** – a tunnel located above the DNA-RNA hybrid in the clamp direction through which nascent RNA longer than 11 nt extends. It encapsulates the RNA for about six more residues. Specific loops termed "rudder", "lid" and "zipper" extending from the clamp position the RNA in the exit tunnel [18].
- **Funnel/pore 1** - a secondary tunnel found below the DNA helix that acts as the entry site for NTPs, as a site of exit for the backtracked RNA and as an entry site of elongation factors that rescue the paused and/or backtracked state [20].
- **Bridge helix** – helix that connects core and clamp domains across the cleft. It contacts the downstream DNA. Bridge helix changes its conformation throughout the transcription cycle between "folded" or "straight" state and "unfolded" or "bent" conformations allowing translocation to occur without complex disassembly [19,21].
- **Aspartate loop** – active site element, a conserved triad of aspartates from the largest RNAP subunit.

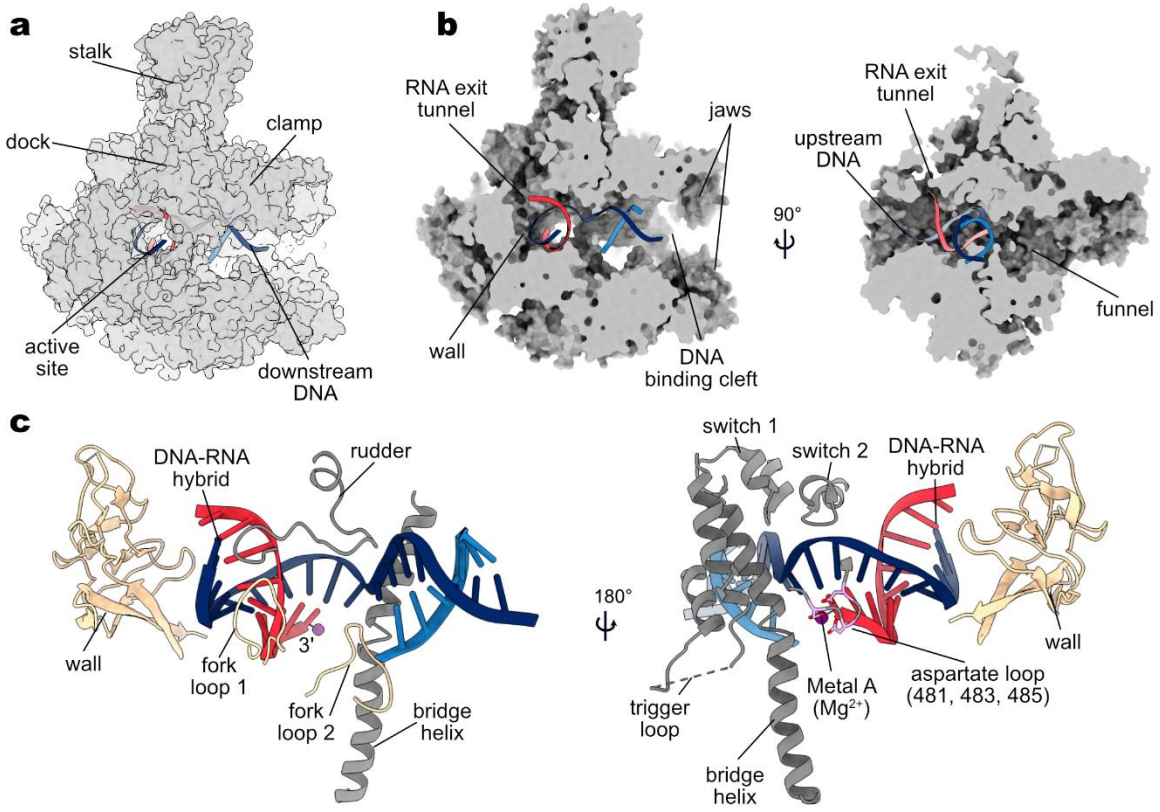

Fig. 1.2. Structural elements of eukaryotic RNAP. a| General features of the RNAP structure. **b|** Functional tunnels, jaws, and wall annotated on the RNAP structure. Pol II elongation structure is shown as a cross-section from the side (left) and front (right). **a,b|** Protein surface is shown in grey, with DNA (blue) and RNA (red) shown in cartoon representation. **c|** Active site elements shown in cartoon representation. All surrounding chains are removed for clarity. The largest RNAP subunit is coloured in grey, while the second-largest RNAP subunit is shown in wheat. Catalytic Mg^{2+} ion is represented as a pink sphere. All panels are made using the *S. cerevisiae* Pol II elongation structure PDB: 1Y1W [25].

- **Trigger loop** – loop coming from the largest RNAP subunit found close to the bridge helix. It changes its conformation from "open" or "disordered" to "closed" or "ordered" conformation throughout the transcription cycle [21] (further discussed in section 1.1.3).
- **Hybrid binding domain** – The region surrounding the active site, coming from the second largest subunit of RNAP [19].
- **Fork loops 1 and 2** – two loops coming from the second largest RNAP subunit. Fork loop 1 maintains the upstream portion of the transcription bubble, while fork loop 2 positions the downstream edge of the bubble and directs the NT-strand [18].

- **Rudder** – loop near the active centre that participates in separating the RNA from the DNA in the DNA-RNA hybrid region and maintains the upstream end of the hybrid [19].
- **Switches** – flexible regions that allow connection and movement between clamp and cleft. Switches 1, 2 and 3 are longer and more flexible than switches 4 and 5. They also contact the DNA-RNA hybrid, possibly coupling changes in the cleft width to the presence of the nucleic acids [18,19].

1.1.3 Active site and the mechanism of nucleotide addition

The key functionality of RNAP is the synthesis of RNA using a DNA template. The DNA binds in the central cleft of the RNAP at the active site. An open DNA bubble is separated into the template (T) and non-template (NT) strands, stabilised by the active site elements. The open bubble is 11-12 nt [26] in length with an RNA-DNA hybrid of 8-9 nt [10]. The DNA is opened, and the bubble and DNA-RNA hybrid are maintained by the rudder and fork loop 1. The downstream DNA is positioned by fork loop 2. The DNA strands re-anneal immediately after the separation of the T-strand from RNA. The RNAP clamp is closed during elongation, tightly binding the DNA scaffold [10].

Focusing only on the mechanism that involves the nucleic acids, three different stages can be separated along the nucleotide addition cycle: (i) NTP binding, (ii) catalysis and (iii) pyrophosphate ion (PPi) release. At the beginning of the cycle, the 3' end of the nascent RNA chain is bound in the active site with an empty nucleotide addition site (termed i+1). This state is referred to as a post-translocation state. Incoming NTP enters the active site through a secondary entry site called the funnel and base-pairs to the T-strand [27]. Catalysis occurs via two metal ion mechanisms. Metal ion A (Mg^{2+}) is permanently bound in the active site coordinating the 3' end of the nascent RNA. Metal ion B (Mg^{2+}) instead coordinates the incoming NTP. The deprotonated 3' hydroxyl group of the RNA acts as the nucleophile and attacks the NTP-α-phosphate of the NTP. This, in turn, results in a formation of a new phosphodiester bond and the release of PPi [26]. The addition of a nucleotide is energetically favourable, with a free energy change of about −5.6 kcal/mol [28,29]. When the new bond is formed, RNAP is in the pre-translocated state with the i+1 site occupied by the nucleotide.

Many structural elements of the RNAP are necessary to allow the transition through described steps of the nucleotide addition cycle. The catalytic ions are coordinated by three conserved aspartate and/or glutamate residues of both largest RNAP subunits [26] (in the case of Pol I, all catalytic residues are aspartates from the RPA1 subunit [23]). The bridge helix and trigger loop are two essential active site elements that undergo conformational changes throughout the nucleotide addition cycle. In the post-translocated state, the bridge helix is completely folded and the trigger loop is disordered (also called "open"), allowing NTP to enter the active site. Both bridge helix and trigger loop bind the NTP and correct base pairing between the inserted NTP and T-strand of DNA induces folding of the trigger loop into a hairpin ("folded" or "closed" state), which is associated with about 20 Å displacement of this structural element [30]. The trigger loop aids in discrimination between the complementary and non-complementary base pairing of the NTP and T-DNA and between NTPs and deoxynucleoside triphosphates (dNTPs). Folded trigger loop helps position the nucleic acids for the nucleophilic attack. After a new phosphodiester bond is formed, the trigger loop again unfolds. This cycling between "open" and "closed" conformations of the trigger loop occurs on a millisecond scale with the addition of each nucleotide [30]. Further, the PPi is released, which might be promoted by the presence of the following complementary NTP [27]. Yet, for the RNAP to be ready to add the next NTP, the active site needs to be re-set to the pre-translocation state with the empty i+1 site. RNAP is proposed to oscillate between pre- and post-translocation states in agreement with the Brownian ratchet model. Unfolding and bending the bridge helix, which contains an unstable α-helical region, helps push the RNAP to the pre-translocation state [26,31].

In addition to the oscillation between pre-and post-translocated states, the RNAP can also ratchet. In ratcheted state, RNAP has an open DNA-binding cleft, kinked bridge helix and can have an additional factor/subunit occupy the NTP entry funnel [32] (in the case of Pol II, elongation TFIIS factor plays this role, and in Pol I it is fulfilled by the C-terminal domain of the RPA12 subunit [23]). The bridge helix that spans across the cleft undergoes changes in its conformation that allow linking movements of the rigid components to the ratcheted motion [31]. This conformation is also associated with the backtracked state. In backtracked state, the 3' end of RNA dissociates from the T-DNA and exits the active site via a pore termed funnel, opposite the canonical RNA exit tunnel. To rescue the backtracked state, the RNA must be cleaved to allow the 3' RNA end to return to the active site [26].

1.1.4 Specialisation of eukaryotic RNA polymerases

In prokaryotes, only one RNAP is responsible for the transcription of the entire genome. Eukaryotic RNAPs share homology with this basic bacterial RNAP composed of only four subunits. Those subunits form a catalytic core also present in all other species. Archaeal RNAPs are of intermediate complexity, having 11-13 subunits [26]. Eukaryotic RNAPs are the most complex, standing at 12-17 subunits and up to 0.7 megadaltons (MDa) in size [33]. Given the high level of structural homology and use of an identical molecular mechanism to synthesise the RNA, it is highly likely that all currently known RNAPs evolved from one complex found in the last universal common ancestor (LUCA) [3]. The increased complexity of the RNAP structure is likely a response to the increasing complexity of the genome and its reduced accessibility due to packaging into chromatin [34].

The three eukaryotic RNAPs are found in different cellular compartments and transcribe non-overlapping subsets of genes (Fig. 1.3 a, b). Pol I is found solely in the nucleolus, where it produces only one type of transcript – the pre-rRNA. Pol II localises mainly in the nucleus but can also be found in the nucleolus. Pol II transcribes the most extensive repertoire of genes, including coding genes giving rise to mRNA and many non-coding genes. Pol III operates mainly in the nucleus, synthesising a range of short, non-transcribed RNAs. Pol III can also be found in the cytoplasm, where it plays a role in the innate immune system [10].

Differentiation of the eukaryotic RNAPs allowed them to specialise in the transcription of their specific target genes (Fig. 1.3 a). Included adaptations involved modification of existing subunits and the incorporation of novel subunits, often related to GTFs used by other RNAPs [10] (Fig. 1.1 c). Specialisation was needed to allow different levels of control over each subset of genes. This way, transcription outputs can better respond to various environmental stimuli, and both the abundance and the diversity of transcripts can be regulated. The majority of the innovation is found in the peripheral subunits and the active site is highly conserved, as the mechanism of the nucleotide addition is crucial to the cell's survival and, as such, allows little modification. Peripheral subunits on the other hand, when modified, allow specialised regulation of the whole complex, which drives RNAP diversification [34].

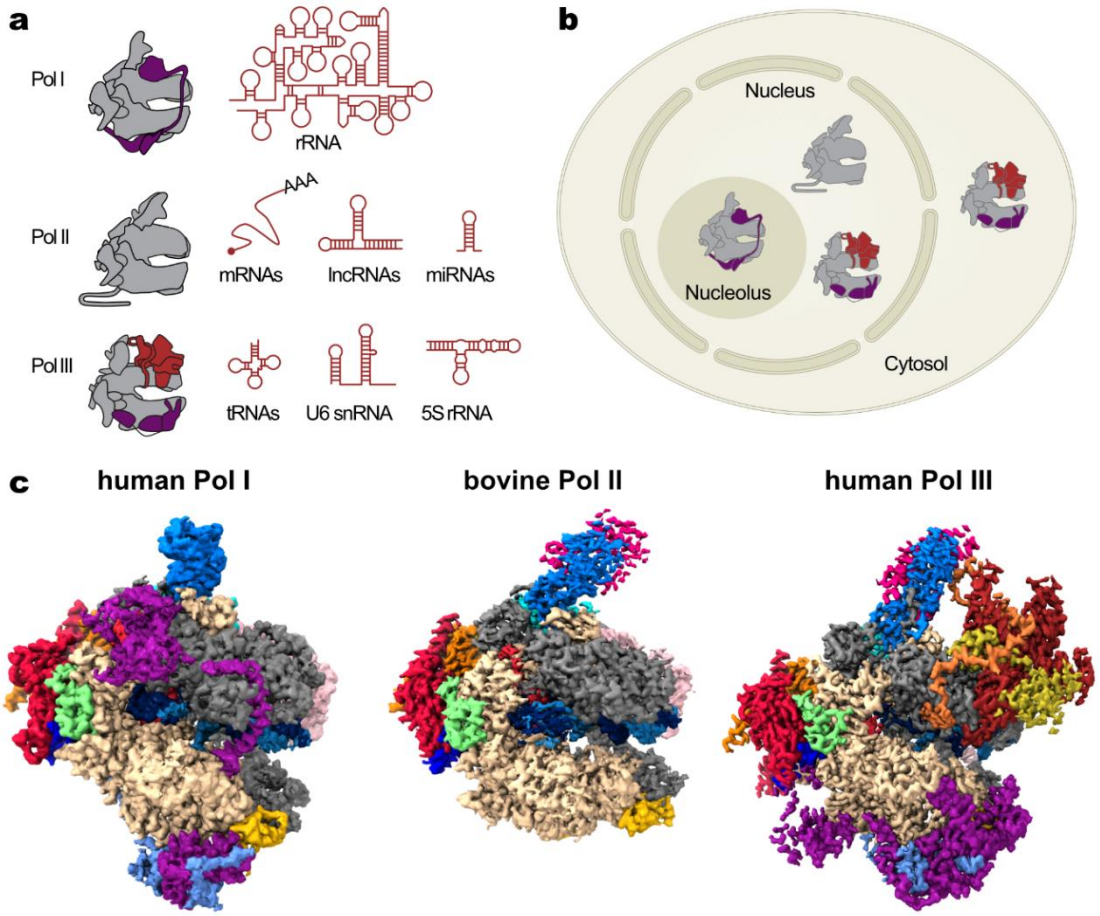

Fig. 1.3. The specialisation of the eukaryotic RNAPs. a| The three eukaryotic RNAPs in cartoon representation next to an example of their target transcript. **b|** Localisation of the three eukaryotic RNAPs into cell compartments (cytoplasm, nucleus and nucleolus as labelled). **c|** Gallery of mammalian RNAPs in elongation state. Cryo-EM densities of the RNAPs are shown coloured by subunits. Homologous subunits are shown in the same colour. PDBs 7OB9 [23], 5FLM [35] and 7AE1 [36] were used to visualise Pol I, II and III, respectively. Panels **a** and **b** are adapted from [10].

Pol III is the largest among the eukaryotic RNAPs and is speculated to have differentiated first [34]. It has stably integrated many GTFs to reach 17 subunits making up the basal complex. Thus, Pol III requires the least additional transcription factors: it can elongate and terminate without additional factors and uses only two or three initiation factors. This made Pol III particularly specialised for fast transcription of short and abundant genes. Pol I is an intermediate RNAP in terms of its complexity. In most species, it has 13 subunits (in yeast 14) and requires several factors at every stage in the transcription cycle. Those additional factors allow modulation of rRNA production in response to environmental stimuli, which is particularly important given the high energy expenditure associated with producing

such large transcripts. Transcription factors also help increase Pol I processivity and transcription rate, which are needed to transcribe the very long target gene of Pol I. Pol II is the smallest eukaryotic RNAP with 12 subunits. On the other hand, it has the most diverse repertoire of GTFs. Many factors are required for Pol II function at every stage of the transcription cycle. This allows fine-tuning the control over gene expression performed by Pol II in response to various internal and external stimuli [10]. In summary, diversification of the eukaryotic RNAPs allows them to perform their function most efficiently.

1.2 Pol I transcription

1.2.1 Pol I discovery

Eukaryotic transcription has been an important field of study for over 50 years, as captured in a recent review [37]. RNAP activity, understood as the incorporation of NTPs into the growing RNA chain, was first shown in 1959 using a rat liver extract [38]. Subsequently, bacterial RNAP was the main focus of studies, as described in detail by Hurwitz, 2005 [39]. Only in 1969, three distinct eukaryotic RNAPs were identified in the sea urchin embryos. They were separated using anion exchange chromatography. The three identified species eluted in distinct peaks and were numbered from I to III according to the sequence in which they eluted from the column [40]. Subsequent purification of the three RNAPs from *Xenopus laevis* and *Saccharomyces cerevisiae* allowed the separation of individual subunits and noting their distinct molecular weights [41,42]. Further, it was realised that the identified complexes had different sensitivities to certain drugs, such as α-amanitin, with Pol I being resistant to the drug, Pol III being mildly affected and Pol II showing high sensitivity to it [43,44]. This observation initiated efforts to study each of the RNAPs in greater detail.

Following studies identified basal transcription factors [45–48] and promoter elements [49,50] necessary for Pol I function. These results allowed the reconstitution of specific Pol I activity *in vitro* [51,52]. Finally, a low resolution (~30 Å) structure of yeast Pol I was obtained [51,53]. High-resolution crystal structures of yeast Pol I were solved in 2013 [54,55], followed by cryo-EM structures of elongating yeast Pol I [56,57]. In last years, multiple additional structures of yeast Pol I in various conformations became available, such as structures of Pol I bound to initiation factor RRN3 [58,59], bound to a nucleotide analogue that also lacks some Pol I-specific subunits [60] and also in complex with initiation factors forming a complete PIC [61–63]. Other than *S. cerevisiae* Pol I, structures of the *Schizosaccharomyces pombe* Pol I are available [64]. Only in the last year, human Pol I structures were solved by myself and the Yang and Xu groups [23,24].

1.2.2 Pol I specialisation

Pol I is a highly specialised enzyme that produces only one type of transcript – the pre-rRNA. Two main challenges are associated with rRNA production: the long length of the pre-RNA transcript and the high demand for the rRNA production needed to meet the ribosome biogenesis demand.

To allow more efficient transcription initiation and elongation, Pol I has stably integrated many general transcription factors found in the Pol II system into the basal complex (details discussed in the 1.2.3 and 1.2.5 sections) [10]. At the same time, additional factors such as DRB sensitivity-inducing factor (DSIF) (Spt4/5 in yeast) or PAF1 complex are still associated with Pol I, which further increase its transcription rate [65–68]. Compared to Pol II and Pol III, the Pol I machinery is of medium size and complexity. As such, its assembly efficiency and the flexibility of regulation land in the middle between Pol II and Pol III [10]. It has been speculated that to increase Pol I processivity and transcription rate further, rRNA folding and processing may play a role [23,69]. Additionally, during transcription, the Pol I clamp closes, more tightly engaging the DNA-RNA hybrid, which might help with the complex's processivity [23,56]. Taken together, it becomes apparent that Pol I is specialised in producing long RNA transcripts.

Between 80-90% of the total RNA pool in a eukaryotic cell is attributed to rRNA [70]. Many transcriptionally active rDNA repeats are present in the genome to mitigate the rRNA production pressure. The total numbers vary between species, ranging from 150-200 found in yeast and 200-600 repeats found in the human genome [71,72]. Not only are there many copies of the rDNA gene that can be transcribed simultaneously, but multiple Pol I can transcribe each copy. Using electron microscopy of chromatin spreads, convoys of actively transcribing Pol I that form the characteristic "Miller tree" have been described [57,73]. Having many active copies of the rDNA gene transcribed by multiple Pol I complexes allows for meeting the high demand for the pre-rRNA.

1.2.3 Pol I architecture

The general architecture of Pol I is conserved across the tree of life [21,23]. The Pol I subunit nomenclature is summarised in Fig. 1.4 and Table 1.1. The Pol I core resembles all other eukaryotic RNAPs and comprises the two, largest subunits (in humans, RPA1 and RPA2) complemented by five subunits shared with Pol II and Pol III and two subunits (RPAC1 and RPAC2 in humans) exclusively shared with Pol III. The two Pol I – Pol III specific subunits constitute the RNAP biogenesis platform [74].

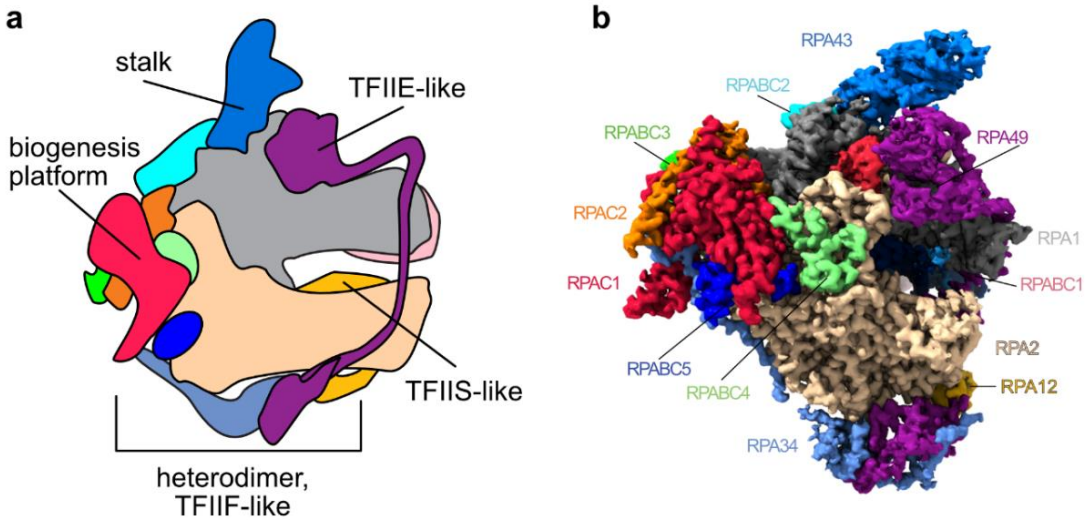

Fig. 1.4. The general architecture of Pol I. a| Pol I in cartoon representation with functional domains annotated. **b|** Surface representation of the human Pol I (PDB: 7OB9 [23]) with all subunits marked.

On top of the core, an additional domain termed the stalk is located (Fig. 1.4 a). In the case of Pol I, in most species, except for some fungi, including *S. cerevisiae* and *S. pombe*, the stalk is made up of only one subunit. At the same time, Pol II and Pol III have a stalk consisting of two subunits (further revisited in section 3.2.5) [23,24]. Additionally, Pol I has stably integrated several subunits homologous to Pol II general transcription factors (Fig. 1.4 a). The TFIIE-TFIIF-like chimaera is called the heterodimer and comprises the RPA34 and RPA49 subunits [75]. The dimerisation module, located opposite the stalk, is homologous to TFIIF. The C-terminus of the RPA49 subunit forms a tandem winged-helix domain (tWH) which binds close to the upstream DNA [61] and in another conformation next to the stalk and the RNA exit tunnel [23,56]. This domain is homologous to TFIIE and is thought to play a role in transcription initiation and elongation [75]. Pol I has also incorporated RNA cleaving activity as the C-terminal domain of the RPA12 subunit, akin to the Pol II general elongation factor TFIIS [76].

Table 1.1. Subunit nomenclature for Pol I in yeast and human and their Pol II and Pol III counterparts. Percent sequence identity between human and yeast subunits and their molecular weights are given.

Identity Human vs Yeast (%)	**Mass**		**RNAP domains**	**Pol I**			**Pol II**	**Pol III**
	Human	**Yeast**		**Human**		**Yeast**	**Human**	
				gene	**protein**	**protein**	protein	
40.77	195 kDa	190 kDa	**Homologous core subunits**	POLR1A	RPA1	A190	RPB1	RPC1
45.06	128 kDa	135 kDa		POLR1B	RPA2	A135	RPB2	RPC2
45.18	39 kDa	38 kDa		POLR1C	RPAC1	AC40	RPB3	RPAC1
44.63	15 kDa	16 kDa		POLR1D	RPAC2	AC19	RPB11	RPAC2
36.21	14 kDa	14 kDa		ZNRD1	RPA12 (N-term)	A12.2 (N-term)	RPB9	RPC10 (N-term)
					RPA12 (C-term)	A12.2 (C-term)	TFIIS (C-term)	RPC10 (C-term)
43.48	25 kDa	25 kDa	**Shared core subunits**	POLR2E	RPABC1	ABC27	RPB5	RPABC1
56.69	14 kDa	18 kDa		POLR2F	RPABC2	ABC23	RPB6	RPABC2
35.97	17 kDa	17 kDa		POLR2H	RPABC3	ABC14.5	RPB8	RPABC3
74.63	8 kDa	8 kDa		POLR2L	RPABC5	ABC10β	RPB10	RPABC5
37.93	7 kDa	8 kDa		POLR2K	RPABC4	ABC10α	RPB12	RPABC4
-	-	15 kDa	**Stalk subcomplex**	-	-	A14	RPB4	RPC9
21.45	37 kDa	36 kDa		TWISTNB	RPA43	A43	RPB7	RPC8
19.95	47 kDa	47 kDa	**Heterodimer**	POLR1E	RPA49 (N-term)	A49 (N-term)	TFIIFα	RPC5
					RPA49 (C-term)	A49 (C-term)	TFIIE	RPC5
25.00	55 kDa	27 kDa		CD3EAP	RPA34	A34.5	TFIIFβ	RPC4

1.2.4 Structure of the rDNA gene repeat

Each single rRNA gene in humans is 43 kb long, out of which 13 kb code for the 47S pre-rRNA. It is later spliced to give 18S, 5.8S and 28S rRNA [77] (Fig. 1.5 a). rDNA repeats form linear arrays with repeats arranged in head-to-tail orientation. Those repeats can be spread across many chromosomes, like in the case of humans, where they are found on the short arms of acrocentric chromosomes, which in humans are chromosomes 13, 14, 15, 21 and 22. Sometimes, like in the case of *S.* cerevisiae, all repeats can also be found in one locus [78]. All together, repeats localise to the nucleolus where rDNA transcription occurs [71].

At the beginning of the coding sequence, the core promoter region is located (Fig. 1.5 a). In humans, it is bound by the Selectivity Factor 1 (SL1) in complex with the TATA-box Binding Protein (TBP). Pol I binds directly to the ribosomal initiator element (rInr), which is functionally related to the Pol II initiator element (Inr) [78].

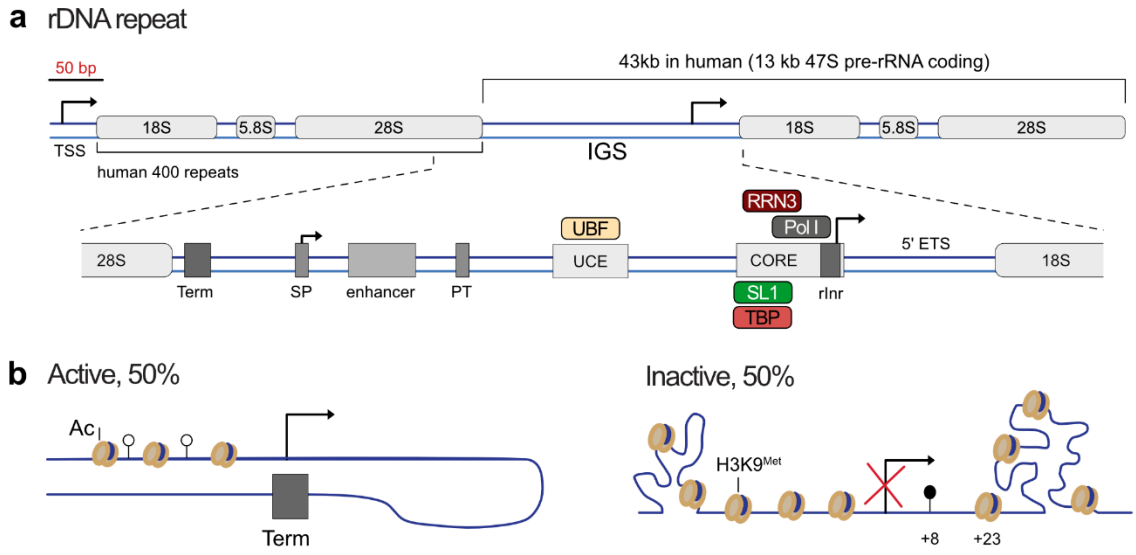

Fig. 1.5 Structure of an rDNA repeat. a| Structure of the rDNA repeat in cartoon representation. Sites of binding of Pol I transcription factors and Pol I are indicated by coloured squares. TSS - transcription start site, IGS – intergenic spacer, Term – terminator sequence, SP – spacer promoter, PT – proximal terminator, UCE – upstream control element, rInr – ribosomal Initiator, 5' ETS - 5′ external transcribed spacer. **b|** Epigenetic markers determine the rDNA's active or inactive status. Arrow indicates transcription start site. Numbers indicate positions with respect to the transcription start site. Empty circles indicate demethylated DNA, and filled circles indicate methylated DNA. Ac – acetylation. Panel **a** is adapted from [10].

The rInr is a part of the core promoter region. Further upstream from the transcription start site, another regulatory element is located. In humans, it is termed the upstream control element (UCE) and is bound by the Upstream Binding Factor (UBF), which *in vivo* stimulates transcription above the basal level [79]. Interestingly, UBF is dispensable *in vitro*, and no stimulatory effect on transcription has been observed with its addition [80]. The functionally homologous counterpart of human UBF in yeast is the Upstream Activating Factor (UAF) [81], which however, shares no structure or sequence identity. In yeast, UAF forms a complex with TBP. The recently solved structure of UAF suggests that it is needed to sequester TBP from Pol II and thus push the equilibrium towards the formation of the Pol I initiation complex [82]. Since in humans UBF only interacts with TBP, and instead SL1 forms a stable complex with TBP, it remains to be seen if UAF and UBF function similarly. It was suggested that both UBF and UAF can bend and possibly loop DNA, creating permissive transcription structures and allowing for the efficient binding of transcription factors [83]. DNA looping might play an especially crucial role in the case of UBF decorating the DNA throughout the rDNA repeat and not only at the regulatory sequences [84].

The rRNA coding regions are interspaced by the non-coding Intergenic Spacer (IGS), which has a regulatory role (Fig. 1.5 a). It contains various elements such as enhancers, space promoters (SP) and proximal terminators (PT), which can affect the rRNA expression [85]. SP may play a role in increasing the numbers of Pol I bound to the DNA [86]. It may also allow for the initiation of non-coding transcripts that can further associate with chromatin remodellers such as nucleolar remodelling complex (NoRC) and cause silencing of adjacent rDNA repeats [87]. IGS-derived transcripts produced by Pol I can also affect the epigenetic state of rDNA repeats and thus regulate its expression [88]. SP may also initiate transcription by Pol II, which then in mammalian cells has been shown to create R-loops, which in turn regulate Pol I transcription [89]. PT are additional terminator sequences scattered across the IGS. They function in the termination of IGS-derived transcripts but might also have a regulatory role in transcription enhancement via chromatin looping, which could stimulate initiation [90].

Given the highly repetitive nature of rDNA sequences, those loci are mutation and recombination hotspots. IGS is the integration site of many retrotransposons and contains many microsatellites, long repeats or butterfly repeats, among others [91]. All discussed features of the IGS and the constant transcription pressure lead to the genomic instability of the rDNA repeats. Thus, a high variation in copy number and repeat variability is observed between individuals of the same species [92]. The rDNA repeat's general organisation between species is highly conserved, but the individual elements are highly species-specific [93,94].

At any given time, only about 50% of rDNA genes are transcriptionally active [95] (Fig. 1.5 b). This distinction can vary even between different tissues in the same organism [94] and is maintained by many chromatin features, including epigenetic marks. As such, the active repeats show permissive chromatin and histone marks such as hypomethylation of DNA, histone acetylation, enrichment of UBF or looping of the DNA. Looping of the rDNA is supposed to arise by promoter-terminator interactions. It could be mediated by transcription termination factor TTF-I which in this model would facilitate efficient transcription and Pol I recycling [90,96]. The inactive rDNA repeats instead resemble heterochromatin. They are compacted by nucleosomes and carry repressive histone marks while the DNA is methylated [96].

1.2.5 Pol I transcription cycle

The Pol I transcription cycle follows the general pattern of all RNAPs transcription cycles described in sections 1.1.1. Here, I will thus focus on Pol I-specific adaptations throughout transcription initiation, elongation and termination.

(i) Initiation

To become a part of the PIC bound to the DNA, Pol I must first be activated by a Pol I-specific initiation factor called RRN3 [97,98]. In humans, Pol I exists in two distinct, identified forms: Pol Iα constituting 90% of the total pool and Pol Iβ making up 10% of the total Pol I. While both forms can sustain transcription *in vitro* on a sheared DNA template, only Pol Iβ, which is pre-bound by RRN3, can initiate specific transcription from the rDNA promoter [99]. In yeast, an association of RRN3 with Pol I prevents it from the formation of inactive dimers [100,101]. Pol I dimerisation has not been observed in humans [23], suggesting that the activation mechanism differs between species. The binding of RRN3 to Pol I is controlled by post-translational modifications allowing its regulation in response to stress, nutrient availability or stage of the cell cycle [102–107]. The structures of Pol I bound to RRN3 are available for both yeast and humans [23,59,61].

RRN3 bridges Pol I and the pre-bound factors at the promoter elements of the Pol I PIC, in yeast termed UAF and core factor (CF), while in humans, called UBF and SL1. It is unclear which element needs to bind first, but it is speculated that the DNA remodelling performed by UAF/UBF is vital for other factors to bind [81,82,84,108]. The post-translational modifications also play a role in the modulation of UBF binding to other components of the PIC, especially TBP [109]. PIC structures are so far only available from yeast [57,58,61–63]. Pol I binds the CF via the protrusion from the A135 subunit [62]. Additionally, the tWH domain of the A49 subunit contacts the upstream DNA in the vicinity of the CF [61]. The human PIC structure will likely differ from the yeast one, given that CF does not bind TBP while SL1 does [110].

In the context of the correctly assembled PIC, Pol I can bind the promoter DNA (Fig. 1.6 a). To allow binding of the template, a large extension from the largest Pol I subunit, called the DNA-mimicking loop or the expander, that lines the DNA binding cleft needs to be displaced [21,23,54]. Subsequently, the C-terminus of the RPA12 subunit needs to be displaced from the active site [62,63]. Concomitantly, the DNA binding cleft contracts, more tightly engaging the template [23,56,62,63].

In this transition, the tWH domain of the A49 subunit (RPA49 in humans) moves away from the upstream DNA and associates close to the RRN3, aiding in its displacement in the transition to the promoter escape [61] (Fig. 1.6 a). UBF may also play a role in the transition between initiation and elongation, aiding promoter escape [111]. After promoter escape, UBF and SL1 remain pre-bound to the promoter, ready for another round of transcription initiation [112].

(ii) Elongation

Pol I can sustain transcription by itself when supplied with an open DNA template and an RNA primer or even with a DNA template with an overhang [113]. In the elongation state, Pol I has a contracted DNA binding cleft, and a tWH domain of the RPA49 subunit positioned close to the RNA exit tunnel forming its extension [23,61]. Necessary for the Pol I processivity and possibly to maintain the open DNA bubble is the helix-turn-helix (HTH) motif found in the RPA49 linker. In the human Pol I structure, it is poised to interact with downstream DNA, and through mutagenesis studies, it was shown to be important for Pol I activity [114]. Another Pol I-specific subunit that has been implicated in the stimulation of the elongation rate and stabilisation of the complex is RPA12 [60,115–117]. While Pol I is processive by itself, the transcription rates measured *in vitro* (transcription rate of 53 ± 4 nt/s [118]) are lower compared to estimated rates for the *in vivo* elongation (transcription rate of 95 nt/s [119]). Similarly, it has been shown with NETSeq that the pause sites do not correlate between the *in vitro* and *in vivo* experiments [120]. It is thus clear that a range of additional factors contribute to the high speed and processivity of Pol I in the chromatin context [121].

Pol I has been shown to interact with some elongation factors implicated in Pol II transcription (Fig. 1.6 b). It associates with DSIF (Spt4/5 in yeast), which is supposed to induce pausing early in the elongation, but promote a higher transcription rate later in the transcription cycle, especially in GC-rich regions [65,66,122–124]. Another factor associated with Pol I is the Paf1 complex, which helps relieve transient Pol I pausing [68]. It is debated whether the heterodimer can dissociate from the Pol I core during elongation or if it can compete with other elongation factors [125]. TFIIH has been shown to decorate the rDNA throughout the promoter and the length of the coding region.

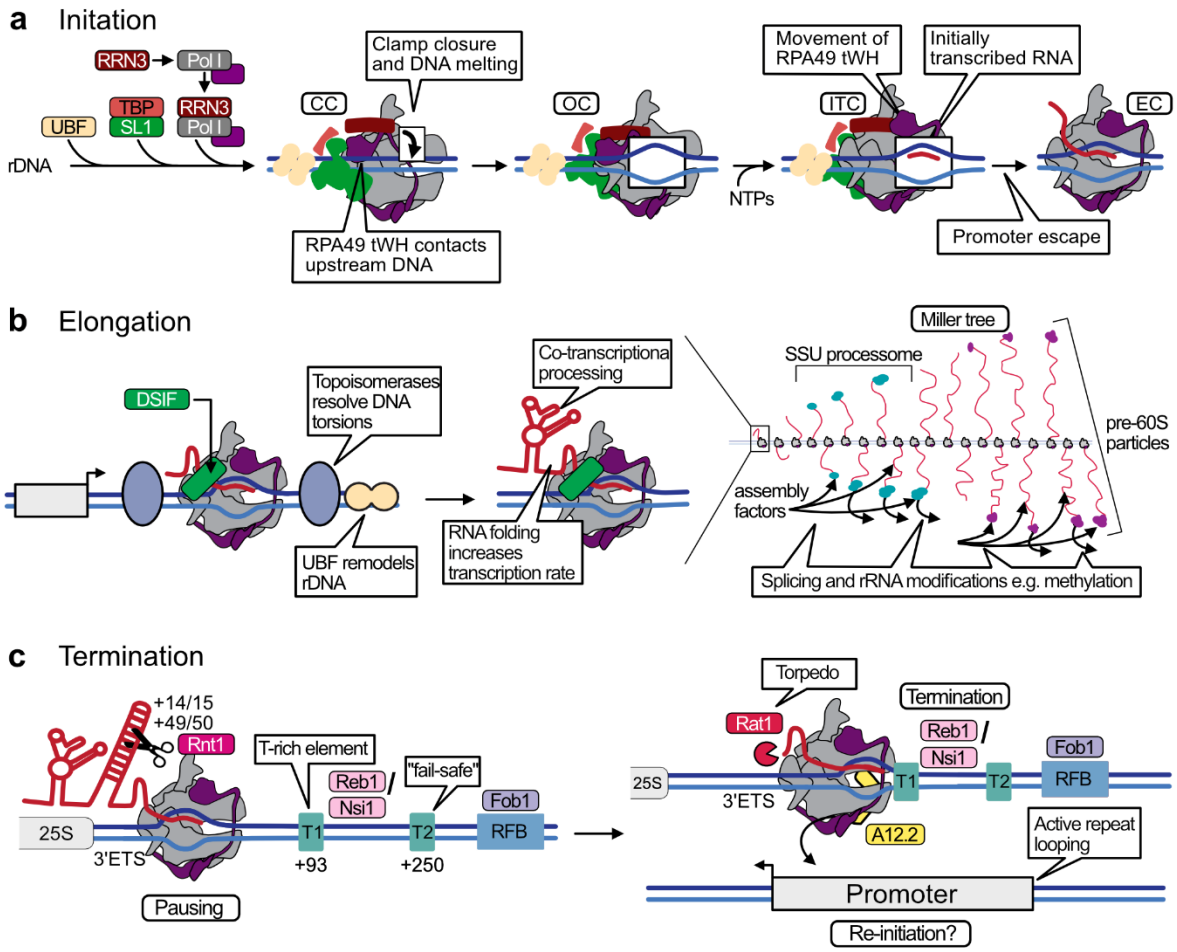

Fig. 1.6. Pol I transcription cycle. a| Pol I initiation. Structural rearrangement in progression from closed complex (CC), where clamp closing occurs, through the open complex (OC) to initially transcribing complex (ITC), where nucleotides (NTPs) are added to the growing RNA chain and the tWH domain of RPA49 is repositioned. After promoter escape, an elongation complex (EC) is formed. **b|** Pol I elongation. Pol I binds elongation factors such as DRB sensitivity-inducing factor (DSIF). Topoisomerases and UBF are needed to maintain the correct DNA structure. Zoomed-out view: multiple Pol I complexes form the Miller tree structure on the rDNA repeat when they simultaneously transcribe one gene. Additional factors process nascent RNA. **c|** Pol I termination in yeast. Rnt1 performs endonucleolytic cleavage (indicated by scissors) across a specifically designated position in a formed RNA stem-loop. All indicated positions are downstream from the 25S coding sequence. T1 – terminator 1, RFB – replication fork barrier. In all panels, Pol I is displayed in grey, with the heterodimer subunits in purple. All factors are shown as coloured squares. DNA is shown in blue and RNA in red. Figure adapted from [10].

It does not influence the Pol I initiation but instead leaves the promoter bound to Pol I and stimulates elongation [126]. TFIIH could potentially interact with the tWH domain of the RPA49 subunit, especially in its conformation bound to the upstream DNA. Still, the exact mechanism and context of such interaction need to be investigated [127]. Pol I has also been shown to depend on chromatin remodellers

such as SWI/SNF, nucleolin and possibly Chd1p, Isw1p or Isw2p [128–130]. Chromatin structure is also influenced by UBF, which in mammals ensures an open chromatin state which allows transcription of active rDNA repeats [131]. Post-translational modifications can also affect the Pol I elongation rate. For example, dephosphorylation of the A43 subunit increases the Pol I elongation speed [132].

During transcription, Pol I forms convoys on the rRNA gene (Fig. 1.6 b, right panel). When visualised, they have a characteristic "Christmas tree" pattern and create so-called Miller trees where the DNA forms the spine of the structure and nascent RNA protrudes to the sides with the ribosome processomes located at the ends of the growing RNA chains [73,133]. While it is not clear whether the subsequent Pol I complexes can interact with each other, it becomes apparent that transcription speed needs to be correlated to avoid steric clashes [134]. Additionally, coordinated speed and spacing of the Pol I complexes might aid in resolving DNA torsions created during transcription [134]. Still, topoisomerases need to also be present throughout the rRNA gene, and when they are deleted, an accumulation of R-loops blocking transcription is observed [135]. Regulation of the speed of elongation is also crucial for the folding of rRNA and ribosome assembly. Disturbances in the elongation speed result in rRNA processing and ribosome biogenesis defects [136]. Not only does the speed of transcription influence RNA folding, but the formation of the RNA structures can influence the speed of transcription. Structured RNA in the RNA exit tunnel correlates with the faster transcription rate since it is energetically favourable and prevents backtracking [69]. While it is clear that there are multiple layers of Pol I elongation regulation, mechanistic insights into the majority of those processes are still lacking.

(i) Termination

Termination is a crucial step of the transcription cycle as it releases the produced RNA and recycles RNAP allowing it to enter a new transcription cycle and preventing transcription of aberrant regions, which could lead to increased genome instability [137,138]. Pol I termination is the least understood step of the transcription cycle, with still more information available for its mechanism in yeast compared to higher eukaryotes.

In yeast (Fig. 1.6 c), transcript release and Pol I detachment are decoupled. Nascent RNA is co-transcriptionally cleaved and processed during the ribosome biogenesis pathway. At the end of the 25S gene, a specific sequence folds into a stem loop that

is the endonucleolytically cleaved by Rnt1 [139–141]. Subsequently, Pol I pauses when it reaches the T1 terminator sequence, which is located 93 bp downstream from the 3' external transcribed spacer (ETS) of the rRNA coding sequence [142]. It is bound by Reb1, as determined in *in vitro* studies [143] or NTS1 silencing protein (Nsi1), as suggested by *in vivo* observations [144]. Pol I pausing facilitates RNA degradation performed by Rat1 exonuclease. It has been suggested that Rat1 functions in a "torpedo" mechanism and not only degrades the remaining RNA but also facilitates Pol I disassembly by physical clash [145]. Yet, neither details of the "torpedo" mechanism nor the steric clash dislocating RNAP have been recapitulated experimentally [146,147]. Essential for the transcript release and complex disassembly is also the A12.2 subunit which harbours the RNA cleavage activity and induces cleft opening when inserted into the active site. A12.2 subunit interacts with Reb1 and is necessary for termination [144,148]. When deleted, Pol I complexes are found distributed throughout the IGS since they fail to terminate correctly [149]. Usually, termination occurs over a T-rich region upstream from the T1 sequence [150]. In the case of the read-through, however, there is a second termination site around 250 bp downstream from the 3' ETS [142]. Another regulatory sequence that ensures Pol I termination is the replication fork barrier (RFB) bound by Fob1. While Fob1 may function as an additional termination factor, its primary function is to ensure no steric clashes between Pol I and the DNA replication machinery [137,145].

In humans, multiple short terminator sequences are present downstream from the rRNA coding sequence. They are bound by TTF-I, which induces DNA bending leading to transcription termination [151–154]. Pol I dissociation and transcript release are assisted by the RPA12 subunit of Pol I and transcript release factor (PTRF) [155]. In mammals, terminator sequences act as the RFB. The first terminator sequence operates only in one direction, but other sequences can act bi-directionally to prevent Pol I and DNA replication machinery clashes [156]. TTF-I allows physical looping of DNA and juxtaposition of initiation and termination sequences [90]. Yet direct evidence of Pol I reinitiation is still missing.

1.2.6 Pol I in health and disease

Pol I is an essential molecular machinery which plays a key role in cellular homeostasis. Ribosome function underlies all cellular processes, and as such its biogenesis needs to be tightly regulated. Additionally, rRNA synthesis is a highly energetically demanding process that must be controlled. Pol I activity is regulated in response to nutrient availability or stress [157–159]. It also underpins the correct progression of the cell cycle [160,161]. Imbalance in Pol I function is associated with complex disorders. In the case of Pol I upregulation, cancer transformation of cells occurs. Developmental disorders arise when Pol I is downregulated, or its function is impaired. The remainder of this section will discuss in more detail the role of Pol I in those abnormal states. Additionally, Pol I activity has also been implicated in longevity studies. In flies, it has been shown that mild downregulation of Pol I prolongs the organism's lifespan. Pol I was also implicated in organismal growth [162] and even in cell fate decisions of stem cells [163]. We are only beginning to understand the complex role of Pol I in the function of an organism.

(i) Pol I in cancer

Already in the 19th century, the size and number of nucleoli were used as a marker for cancer malignancy [164–166]. Since Pol I upregulation is required for cancer progression, size, the number of nucleoli, and the level of Pol I activity have been used as markers for cancer diagnosis [167–173]. In the direct line of reasoning, increased demand for ribosomes directly correlates with increased demand for proteins for cell growth and division [174]. However, a more extensive Pol I and ribosome biogenesis role in cancer progression has recently been identified. Pol I and the rRNA production are required for nucleolus formation, which in turn plays a role in many processes maintaining the cells' activity [175–178]. Those additional functions include nucleolar sensing of the ribosome biogenesis status and the capability to trigger a nucleolar stress response that, via the p53 pathway, results in cell cycle arrest or even apoptosis [164]. The nucleolus also plays a role in maintaining genome stability, DNA repair and telomere integrity, among others [166,175,179,180].

Pol I has thus emerged as a promising drug target for various cancer types [166]. Several drugs targeting components of the Pol I machinery are already in clinical use, while small molecule inhibitors targeting Pol I itself are under clinical trials. Taking into account only the compounds that directly target Pol I, the CX-5461 drug shows significant promise. It was first reported to selectively induce apoptosis

in B-cell lymphoma and was later used in phase I clinical trials to treat hematologic malignancies. CX-5461 is thought to target Pol I transcription initiation and prevent its association with SL1, which triggers the p53 pathway in cancer cells [164,181]. Another Pol I-specific inhibitor is BMH-12, which functions in a not fully defined manner. It is proposed to target Pol I transcription elongation and act as a DNA intercalator. Yet further studies are needed to fully define its mode of action and therapeutic potential [164,182,183]. Having a high-resolution structure of the human Pol I and in the future, its direct binding partners will allow more targeted drug design via, for example, fragment screening approaches. A complete understanding of the human Pol I transcription cycle is needed to choose specific targets for efficient inhibition.

(ii) Pol I and development

Given the central role of Pol I in ribosome biogenesis, it is no surprise that impairment of Pol I or its associated factors' function leads to various diseases, collectively termed ribosomopathies [184]. Aetiology of Pol I-associated diseases occurs during development, possibly due to the higher demand for ribosome production in this period. Pluripotent stem cells have been observed to have higher rates of rRNA production, and lowering of the Pol I activity leads to cell differentiation [185–188]. Many diseases arise from disturbances to proteins associated with Pol I. Each has specific symptoms, but they also share many pathologies such as symptoms of ageing, neuronal issues which usually arise during development and skeletal, facial or limb malformations [184]. Up to date, only a few disorders with disease-causing mutations mapping directly to the Pol I complex have been described. They all arise during development and show craniofacial or neuronal abnormalities.

- **Treacher Collins Syndrome (TCS)**

A disorder characterised by congenital craniofacial abnormalities. In most cases, it is caused by mutations in the treacle protein, which interacts with UBF and plays a role in ribosome biogenesis [189]. In other patients, mutations in RPA2, RPAC1 and RPAC2 have been identified as the underlying cause of TCS. Mutations associated with TCS are autosomal dominant (in the RPA2 subunit and the majority of mutations in the RPAC2 subunit) or recessive (in the RPAC1 subunit and rare cases such as L55V mutation in the RPAC2 subunit) [190–192]. Diminished functionality of Pol I (and/or Pol III) results in reduced ribosome numbers in neural crest cells

and the neuroepithelium at a crucial developmental stage. In such cases, apoptosis of some progenitor cells might be triggered via the p53 pathway leading to the observed deformations [189,191]. TCS patients show a broad spectrum of disorder severity with currently poor ability to predict disease outcomes [189,193]. Current TCS preventive therapies focus on p53 pathway inhibition. The caveat of this approach is induction of tumorigenesis as a side-effect of the treatment [194,195]. It thus becomes apparent that a more detailed understanding of TCS aetiology is needed to devise new treatment methods.

- **Acrofacial Dysostosis (AD), Cincinnati type**

A cranioskeletal malformation syndrome resulting from mutations in the RPA1 subunit. It is an autosomal dominant condition first described by Weaver *et al.* (2015) [196]. AD is characterised by craniofacial abnormalities and limb deformations (in two out of three patients). The craniofacial phenotype was recapitulated in the zebrafish model. It is speculated to arise from increased cell death caused by the activation of the p53 pathway, which diminishes the population of the migratory neural crest cells. The possible mechanism in which the limb abnormalities arise is not known.

- **Neurodegenerative Leukodystrophy**

A recently described condition associated with homozygous mutation S934L in the RPA1 subunit of Pol I. The disease manifests itself with various issues, including cerebral atrophy and intellectual disability. A reduced amount of the RPA1 subunit in the nucleolus was also reported. Due to the low number of patients with the reported mutation (only two brothers), further studies are needed to understand the causation of Pol I-related leukodystrophy [197].

Another disease in which causative mutations map to Pol I subunits shared with Pol III is Hypomyelinating Leukodystrophy. However, it is speculated that the observed phenotype is mainly caused by Pol III disturbance [23,198], which will be discussed in more detail in the 1.3.2 section.

1.3 Pol III transcription

1.3.1 Pol III specialisation

Pol III specialises in producing short, non-coding RNAs, including transfer RNAs (tRNA), the 5S rRNA and the U6 small nuclear RNA (snRNA). Transcripts produced by Pol III are short (only 70-340 nucleotides) but highly abundant, with each mammalian cell estimated to contain 3-7 x 10^7 tRNA molecules [70,199]. Pol III mainly localises to the nucleus, but a small pool of Pol III molecules are also found in the cytoplasm, where they function as a part of the innate immune system and sense viral RNA [200,201]. The crucial adaptation of Pol III is the capability to synthesise RNA with only a minimal repertoire of additional factors, which allows it to be exceptionally efficient [10]. Thus, Pol III has stably integrated many additional factors into the core complex, making it the largest among eukaryotic RNAPs (Fig. 1.8).

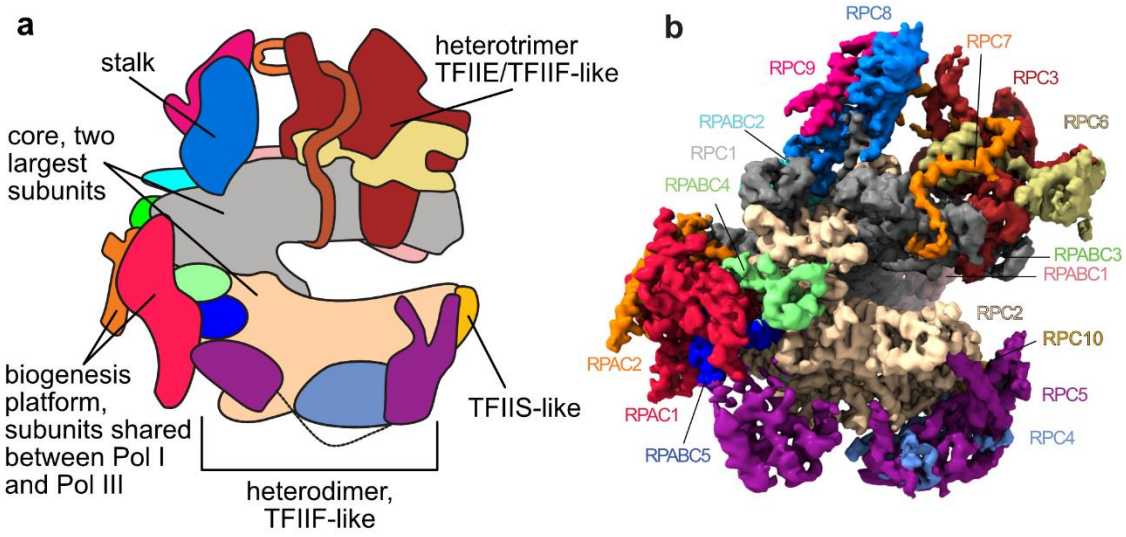

Fig. 1.8. The general architecture of human Pol III. a| Pol III in cartoon representation with functional domains annotated. **b|** Surface representation of the human Pol III (PDB: 7A6H [36]) with all subunits marked.

In addition to the subunits homologous between all core RNAPs, Pol III has two additional subcomplexes: the heterodimer, which fulfils the function of TFIIF found in the Pol II system and the heterotrimer, which is a TFIIE-TFIIF-like hybrid [36,202]. Similarly to Pol I, Pol III also incorporated a TFIIS-like subunit with an RNA cleaving activity [203]. Pol III requires only a few initiation factors throughout the transcription cycle, namely TFIIIA, TFIIIB, TFIIIC and SNAPc (in higher eukaryotes). They are used modularly depending on the transcribed gene type

associated with a specific promoter structure [10,204]. Pol III does not require any elongation or termination factors. It can efficiently terminate by itself using a stretch of thymidines on the non-template strand as the termination signal [205]. Thanks to long-standing efforts, there are now multiple Pol III structures available, including the apo and elongating structures from yeast [202] and humans [36,206–208], yeast PIC structures [209–211], Pol III bound by a repressor Maf1 [212] as well as recent structures that shed light on the Pol III termination mechanism [213,214].

1.3.2 Pol III in health and disease

Pol III produces transcripts that play a role in various crucial cellular homeostasis processes. Among other functions, it makes 5S rRNA and thus contributes to the ribosome biogenesis, plays a central role in translation regulation by the production of tRNAs, influences Pol II transcription via 7SK snRNA, and affects protein translocation since 7SL snRNA is a part of the signal recognition particle and plays a role in the processing of other RNAs via U6 RNA [215]. Given the involvement of Pol III in many processes in the cell, any perturbations to its function result in diseases which span many different phenotypes.

Similarly to Pol I, Pol III also contributes to tumorigenesis when upregulated. Pol III is under the repression of tumour suppressors such as Maf1 and can be activated by proto-oncogenes [216–218]. Additionally, in humans, there is a particular isoform of one of the heterotrimer subunits, RPC7α, specifically expressed during development and by cancerous cells, which may allow Pol III to escape from Maf1 regulation [36,219,220]. Pol III is thus a promising target for anticancer drugs, with triptolide being the first compound identified as inhibiting Pol III transcription, which has a potential for therapeutic use [221,222]. Investigations into the function of another identified Pol III inhibitor, ML-60218, still need to be performed [223]. Pol III inhibition might not only serve as a way to treat cancer but also combat ageing. In animal models, Pol III inhibition has been shown to increase longevity [224]. On the other hand, loss of the Pol III repressor, Maf1, might prevent obesity [225].

Mutations found directly in Pol III can lead to its decreased activity and further influence the function of the cell. It is, however, difficult to quantify those results given the range of transcripts that Pol III produces. Small and structured transcripts are also inherently more difficult to quantify which further increases the difficulty of such analyses [215]. Nevertheless, several disorders whose aetiology is attributed to

the disturbance of Pol III activity have been described (Fig. 1.9). While most of them give a congenital phenotype, there are also known Pol III mutations whose effect is only notable upon viral infection, such as with varicella-zoster virus. Patients then show increased susceptibility to the infection and heightened symptoms, presumably caused by dysfunction of the innate immunity system where Pol III also plays a role [226,227]. When considering developmental disorders, one commonality between different diseases is the effect on neuronal development. How a mutation in a ubiquitous enzyme such as Pol III can result in a tissue-specific phenotype is under investigation. One hypothesis states that the perturbed function of Pol III leads to reduced levels of transcripts which in turn regulate the expression of particular mRNAs crucial for the function and development of cells in the central nervous system [198,228]. While developmental disorders are inherently difficult to treat, specific approaches such as cell therapy, gene therapy, and gene editing techniques might offer promise. However, an understanding of the disease aetiology is needed to efficiently use them to apply the appropriate corrective measures [229].

- **Hypomyelinating Leukodystrophy (HLD)**

HLD is a neurodevelopmental disorder that affects the cerebral white matter with its degeneration leading to disability and often premature death [229]. The severity of the condition has a vast spectrum, from mild to extremely severe and with a life expectancy of patients that varies between a few months to late adulthood (>40 years) [230]. Intellectual disability, developmental delay, myopia and issues with motor function are among the most common symptoms [230]. HLD is usually diagnosed with a Magnetic Resonance Imaging (MRI) scan, which reveals hypomyelination. It is an autosomal recessive disease, and often the compounded effect of two different mutations is needed to give rise to the phenotype [231]. Many mutations giving rise to HLD are found in Pol III-specific subunits: RPC1, RPC2 and RPC10. Additionally, mutations found in RPAC1, the larger subunit shared exclusively between Pol I and Pol III, have also been linked with the disorder [198]. Recently, mutations in the RPC7α subunit, an isoform of one of the heterotrimer subunits only expressed during development or in cancer cells, have also been described [232]. Many described mutations are single-nucleotide substitutions that result in missense mutation, but also a variety of frameshifts and splice site variants leading to truncations have been reported [233–235]. The severity of the condition can sometimes be linked to the mutated subunit, with RPC2 mutations often giving milder phenotypes while RPAC1 and some RPC1 mutations result in the most severe

outcomes [198,230,233]. The pattern is however not complete and given that in the majority of cases, the combined effect of two different mutations has to be taken into account (the majority of patients have compound heterozygous mutations), it is often difficult to predict the phenotypic outcome given the genetic background. The cause of the disorder is attributed to lowered levels of mutated transcripts, which may undergo nonsense-mediated decay or lead to proteins impairing the assembly of the complex [234,236,237]. In both cases, a lower number of functional Pol III is present, leading to decreased binding to target genes and manifests itself in lower levels of 5S rRNA, 7SL RNA and specific tRNAs [236,238].

- **Wiedemann - Rautenstrauch Syndrom (WDRTS)**

A developmental disorder characterised by pre-natal growth retardation and progeroid appearance. Affected individuals often have a poor life expectancy prognosis. It arises from recessive mutations in the RPC1 subunit [239–241]. In the majority of cases, WDRTS arises from a mutation leading to a truncation of a protein, which is found *in trans* with another truncating mutation or with a point mutation [241].

Mutations in subunits shared between Pol I and Pol III have also been linked to TCS as described in the 1.2.6 section, however, it is not yet established if Pol III function is affected by them. In other reports with limited numbers of patients, Pol III mutations have been implicated in conditions such as primary ovarian insufficiency [242] or intellectual disability [243]. While further studies are needed better to understand Pol III's role in those cases, it becomes apparent that disturbance of Pol III's function can have varied and far-reaching consequences for human health.

1.4 Cryo-Electron Microscopy

Cryo-electron microscopy (cryo-EM) is one of the main techniques for structural determination. Its main advantage over X-ray crystallography is the circumvention of the need for protein crystals, allowing cryo-EM to be used for a broader range of molecular targets. The first electron microscopes were developed in the 1930s, but due to the high electron power and vacuum requirement, they were not originally used to image biological samples [244,245]. Initially, biological samples were only imaged using negative stain, where heavy metals stain was added to the sample, allowing reconstruction of the low-resolution envelope of the molecule of interest [246]. Several improvements needed to be made to enable the study of biological molecules. Firstly, the sample had to be preserved, which was achieved by vitrification [247]. It is a technique enabling fast freezing of aqueous solutions which does not allow the crystal lattice of ice to form and instead preserves the liquid-like properties of the aqueous solutions in a frozen state. Secondly, the electron damage to the specimen needed to be reduced, which was achieved by imaging under low electron dose and using the protective effect of the cryogenic conditions [248,249]. Acquired images needed to be averaged to increase the signal-to-noise ratio and allow obtaining 3D reconstructions from 2D projections, which was achieved with computational advancements [250,251]. Another crucial innovation was the development of direct electron detectors, which improved image quality and acquisition time [252]. Thanks to the joined effect of all those advancements, the so-called "resolution revolution" took place. It took cryo-EM from being the technique of obtaining low-resolution "blobs" to a method for obtaining atomic-resolution structures [253,254].

The typical cryo-EM workflow currently involves sample preparation and vitrification, data collection and processing. A general overview of those steps, focused on the standard pipeline for single-particle cryo-EM structure reconstruction, is presented as follows.

(i) Sample preparation and vitrification

The pipeline for sample preparation has been reviewed in detail [255]. The excellent quality of the sample is the first important consideration in obtaining a high-resolution structure. The sample should be purified, ideally to high homogeneity. It should not aggregate, and the protein particles need to be intact. Furthermore, the buffer composition in which the sample is suspended should be considered to

obtain the highest possible contrast. Thus additives such as glycerol, sucrose or high salt concentration should be avoided.

A small volume (typically 1-5 µl) of the sample in solution is then applied to a cryo-EM grid. The grid is typically 3 mm in diameter and is made out of a metal mesh, typically copper or gold. Over the metal grid, a support film is deposited, which can have various properties. The overlaid film forms small holes which will later be imaged to capture the images of the protein complex of interest. The grid's surface is made hydrophilic to facilitate the adsorption of proteins. This is usually achieved by glow discharging or plasma cleaning [256].

The sample is then applied to the grid, its excess is blotted, and the grid is rapidly frozen, usually using liquid ethane. When using automatic plunge-freezing equipment such as Vitrobot (ThermoFisher), parameters such as length of blotting, the force of blotting, the time between sample deposition and blotting, as well as temperature and humidity of the chamber can be controlled. Sample optimisation is usually done on a trial-and-error basis. The goal is to produce a grid in which the protein sample is evenly distributed across the grid, the ice is not too thick nor too thin, and there is little ice contamination. Having a well-distributed sample means that many projections can be acquired in a shorter time, giving more data to process. If the sample is too thick, the image's contrast will worsen. Additionally, particles might then be found above one another, hindering reconstruction. If the ice is too thin, particles might be excluded from the layer. Extensive damage from contact with the air-water interface can also occur, and particles may adopt preferential orientation, which would prevent obtaining the complete 3D reconstruction. Extensive ice contamination reduces the areas where electrons can penetrate and causes electrons to diffract, worsening the data quality. It can also interfere with the integrity of the imaged molecules by disturbing the specimen or its hydration shell.

(ii) Data collection

Following vitrification, the cryo-EM grid is transferred to the transmission electron microscope (TEM) under cryogenic conditions, where it is loaded into the column under a vacuum. Low temperature preserves the specimen's integrity while vacuum prevents the scattering of electrons by air molecules [255]. Typically, the TEM is operated with high-energy electrons (100-300 keV) to allow structure determination to high resolution. Electrons are emitted from the source, typically a field emission gun, nowadays usually a tungsten filament. It is heated to extract and accelerate

electrons. The 80-300 kV voltage is typically used for electron extraction, with higher voltage giving a brighter and more coherent beam. Electrons pass through a series of lenses that allow them to be focused on the sample. Since the lenses used are electromagnetic, they produce a series of aberrations, such as chromatic, spherical and axial astigmatism. Those can be partially counteracted by image processing or imaging with defocus [245]. Once the electrons reach the sample, they interfere with the biological molecules. Electrons which undergo inelastic scattering have the appropriate phase shift to contribute to the image formation. Since biological samples are composed mainly of atoms that only weakly scatter electrons, the produced projection has very low contrast. A detector then records the scattered electrons. Direct electron detectors are most commonly used and give high sensitivity and fast acquisition rates, resulting in the highest quality images [252]. Multiple frames at low electron doses are collected and summed up to give an image. The dose fractionation reduces the sample damage, and having a "movie" instead of one frame allows for better estimation and thus correction for the specimen movement [257]. To further improve the contrast, images are collected at a defined defocus (usually -0.25 μm to -3 μm). Data collection is currently automated, with SerialEM being the most commonly used software [258].

(iii) Data processing

Data processing begins with "on-the-fly" pre-processing of the data. Individual frames are aligned. Further, to restore high-resolution features, data need to be corrected for the optical distortions introduced by the microscope. Those together with the aberrations introduced by the defocus are described by the contrast transfer function (CTF) as a function of spatial frequency. The CTF can be estimated and modelled based on the power spectrum, which is calculated as a Fourier transform of the squared amplitudes and visualised as the typical diffraction pattern with alternating light and dark rings [259,260]. The CTF model allows more accurately estimating the local defocus values, which initially was done per micrograph but has recently been performed on a per-particle basis. This pre-processing can be done using, for example, Warp [261]. Warp is also one of the dedicated software that allows particle picking. In this process, the obtained micrograph is segmented, and only images assumed to be projections of the molecule of interest are extracted [262]. Those are then averaged and classified to finally be arranged in the 3D space and give rise to a 3D reconstruction. This pipeline is based on the assumption that the particles suspended in the thin ice layer assume random orientations, and as such, obtained

projections come from all angles. Given that each projection's angle is unknown, they are determined by fitting in the Fourier space. This can be achieved using, for example, the Bayesian approach to 3D reconstruction, which is harnessed by RELION (REgularized LIkelihood OptimizatioN) [263]. This strategy incorporates a starting model, which is used as a reference and the angles of all projections are assigned to give a model with the highest likelihood of being correct, given both experimental and supplied data. Once a 3D reconstruction is determined, the contributing projections can be further sorted and classified to account for the conformational heterogeneity in the sample [264]. This is often an iterative process that requires a lot of user input. Care should be taken to not bias the reconstruction by supplying a reference model that contains too much information. Thus, the used reference is often low-pass filtered to lower resolution. Once a consensus reconstruction is achieved, validation is needed to confirm that the obtained map does not suffer from overfitting and other biases. This is traditionally done by estimating a Fourier shell correlation (FSC) curve with a standard threshold of 0.143. In this process, two independent half-maps are used [265]. This procedure also allows the estimation of the map resolution [245]. The obtained reconstruction is then ready to be interpreted through model building.

Recent years saw a wave of improvements in the field of single-particle cryo-EM. The technique offers many insights into structures of macromolecules which were not previously accessible. Many improvements are yet to be made, especially in pipeline automation, more efficient sample screening and model validation [266]. Nevertheless, cryo-EM has become the technique of choice for many structure determination projects, especially for larger complexes which do not crystallise and where only small quantities of the starting material are available.

1.5 CRISPR-Cas9 gene editing

In a biological sense, CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR Associated System) is a form of adaptive immunity prokaryotes use to protect themselves from phage viral infections [267]. It relies on short DNA sequences that allow the detection and silencing of foreign nucleic acids in a sequence-specific manner [268]. The adaptive immune system functions by first integrating a short sequence of the invading pathogen's genome into the CRISPR array [269–271]. It is then transcribed together with the spacer element into precursor CRISPR RNA (pre-crRNA), which can be cleaved to form mature crRNA [272]. crRNA is complementary to, and thus can pair with, the genomic sequences of the invading pathogens. Thanks to the regulatory sequences coming from the spacer element, crRNA can bind Cas proteins and direct them to the foreign DNA. Cas proteins then mediate cleavage and, as a result, degradation of the invading DNA protecting the bacterium [268,273,274].

CRISPR-Cas9 discovery began with the short, repeated sequences arranged as direct repeats spotted in the bacterial DNA [275]. Subsequently, the architecture of the locus was characterised, and the term CRISPR was coined [276]. In 2010, the role of the crucial enzyme, Cas9, was characterised in *Streptococcus pyogenes*. It binds to the target sequence and cuts it [273]. The whole system's function and its potential in gene editing were described a couple of years later [268]. Adaptation of CRISPR genome editing in mammalian cells was developed soon afterwards [277].

The type II CRISPR-Cas system is the most widely used for genome editing approaches and thus is the focus here. CRISPR-Cas9 genome editing process is now well understood, with various structures available for editing stages [278]. In this system, the central enzyme is Cas9, a ribonuclease RNase III. It is the only element required for the RNA-directed cleavage of the double-stranded DNA [268]. Here, Cas9 protein binds an RNA scaffold of two RNAs: crRNA and trans-activating CRISPR RNA (tracrRNA). crRNA contains a sequence directly complementary to the target DNA and a small spacer sequence from the CRISPR array. tracrRNA binds to the repeat sequence of the crRNA and activates the Cas9 enzyme [268]. The two RNAs can also be artificially linked to form a single-guide RNA (sgRNA), simplifying the genetic modification system [268]. The selection of the target sequence depends on the 3–5 bp-long protospacer adjacent motif (PAM), which is specific to each Cas protein and, in the original system, allowed bacteria to distinguish between

self and foreign DNA. The PAM sequence is not inserted into the CRISPR array but is essential to select the cutting site [279]. After PAM recognition, target DNA is unwound while simultaneously pairing to the crRNA sequence. Subsequently, both DNA strands are cut using two separate active sites in the presence of Mg^{2+} ions. For the cutting to occur, DNA must first assume a "kinked" conformation, which is only possible in the absence of mismatches. This allows for more precise target selection and prevents off-target cutting [280]. After the cut, the desired change in the genome is produced by the action of the cell's DNA repair pathways. Different repair pathways give various genome modifications. The prevalent DNA repair pathway is the non-homologous end joining (NHEJ). For many CRISPR-Cas9 genome editing applications, this is not a desirable DNA repair pathway as it is error-prone and usually results in non-specific, short DNA deletions. Instead, the homology-directed repair (HDR) pathway results in precise repair from a supplied DNA template, thus allowing the insertion of any desired sequence. The issue with HDR is the low efficiency of the pathway. Therefore many strategies that boost the performance of the HDR pathway, such as supplying an excess of the repair template or tethering it to the nuclease, are under development [281,282].

The CRISPR-Cas9 genome editing system currently experiences a boom in potential applications. Various modifications to the Cas9 enzyme allow cleavage of only one DNA strand or precise tagging of the genomic locus without cutting. Using different Cas proteins allows induction of desired deletions and insertions and the accurate introduction of point mutations. The main challenges in using the system are associated with off-target effects. It occurs when Cas9 cuts the genome in a location other than desired, which might lead to unknown genomic changes. Efficient delivery of the CRISPR-Cas9 components to the target remains an issue in some cases. Reliance on the native DNA repair systems proves that genome editing is often inefficient and cumbersome. While CRISPR-Cas9 has already offered many advances in scientific exploration and is approaching medical use, many advancements still need to be made [278,281–283]. There is immense potential for developing yet more efficient, faster and safer methods to use CRISPR-Cas9 in genome editing.

1.6 Aims and scope of this thesis

Pol I is one of the key macromolecular machines needed for the function of every eukaryotic cell. Over the last decades, our understanding of Pol I function, especially in yeast, has grown tremendously. Yet, no structural and only limited biochemical information is available for the human Pol I. Given the significance of Pol I in the biomedical field, it is essential to expand the understanding of Pol I function, ideally directly in human cells. Further, the interplay between Pol I and Pol III in the aetiology of complex disorders is understudied, hindering the studies of the related diseases. Thus, this thesis focuses on three aims, which contribute to expanding knowledge about human Pol I.

To allow structural and biochemical studies of human Pol I, a purified sample of the complex is necessary. The obtained complex needs to be intact, pure and active to allow meaningful studies. To produce a purified sample of human Pol I and Pol III, I used the CRISPR-Cas9 system. It allowed me to endogenously tag and further purify complexes of interest. The high quality and purity of the sample in sufficient amounts allow further studies to be performed. To further expand the repertoire of human Pol I – related proteins for future studies, I started to produce Pol I initiation factors. When expanded, this work will give insights into the function of Pol I in the context of the PIC. Pol I purified from the endogenous source gives great insight and offers a chance to study the complex process in the native state. Yet to allow the ease of modification of this system, I also prepared the means to obtain recombinant human Pol I. The combination of endogenous Pol I purification with the recombinant production of the accessory factors and Pol I itself opens the avenue for various *in vitro* studies. Structural, biochemical or biophysical studies of the human Pol I system are now possible, and some of their examples are covered further in this work.

Secondly, I aimed to obtain the structure of the human Pol I. Purified, endogenous human Pol I is the perfect target for structural studies using cryo-EM. I solved the structures of the human Pol I in its elongating state, bound to the initiation factor RRN3 and in Open Complex conformation at a resolution of 2.7 to 3.3 Å. Obtained structures give insights into several aspects of human Pol I function. Looking at the conformation of the RNA, I could understand how Pol I can accommodate structured nascent RNA in its exit tunnel and can speculate how this can be beneficial for the transcription rate. Analysis of the stalk conformation uncovered that in humans and the majority of other species, with yeast being an exception, the Pol I stalk is made up of one subunit only, which allows it more flexibility in the binding initiation factors such as RRN3. Obtained structures also give insights into several other

features of Pol I, such as the mode of binding of the heterodimer to the core or different strategies for complex inactivation (via the DNA mimicking loop or via active site rearrangements). Overall, newly obtained structures expand the understanding of human Pol I function and can potentially be used in drug design.

Having obtained the structures of human Pol I and Pol III (performed by Dr. Mathias Girbig,) I could analyse the potential role of many disease-causing point mutations. Especially interesting are the mutations found in the two subunits shared between Pol I and Pol III. Analysis of the two structures showed that they might affect the two complexes differently, potentially contributing to different disorders. To better understand the effect of those disease-associated mutations on the function of Pol I and Pol III, I used CRISPR-Cas9 to introduce them directly into the endogenous subunit in cells. Tracking of the subunit's localisation and studies of Pol I and Pol III assembly give a starting point for understanding how complex developmental disorders arise on the molecular level.

Chapter II: Results and Discussion

PRODUCTION OF HUMAN POL I AND POL III

This section describes the approach I employed to obtain pure and enzymatically active human Pol I and Pol III required for structural studies using cryo-EM. When designing the purification strategy at the beginning of my project, I had to confront the issue of incomplete information regarding the composition of the human Pol I. While the yeast Pol I is composed of 14 subunits, homologues of only 13 of them had been identified in humans. The identity of the smaller stalk subunit (A14 in yeast) homologue in humans was unknown. Thus, the traditional strategy of protein overexpression could not be used. Instead, I decided to employ the CRISPR-Cas9 genome editing strategy to introduce an affinity tag on one of Pol I's identified subunits, which allowed purification of the complex.

The Müller group has a well-established purification strategy for the yeast Pol I and Pol III [21]. In the approach, the larger subunit shared exclusively between Pol I and Pol III – AC40 – is endogenously tagged on the C-terminus. The introduced His-tag allows the initial clearing of the complexes of interest from the lysate, which are further separated from contaminants using a Heparin column. Finally, the two polymerases can be separated by ion-exchange chromatography on a MonoQ column. I designed the human Pol I and Pol III purification protocol to model the established approach for yeast Pol I and Pol III purification. Thus, the tag was inserted on the C-terminus of the RPAC1 subunit, which is the homologue of the AC40 yeast subunit.

The endogenous tagging of Pol I and Pol III has many advantages. It circumvents issues of complex assembly and insertion of correct post-translational modifications allowing the study of complexes in the closest to native conditions. Purification of both Pol I and Pol III simultaneously is also cost-efficient. Yet, there are also disadvantages to the employed strategy. Endogenous tagging with CRISPR-Cas is time-consuming and yields a system that is difficult to modify. Since all Pol I and Pol III components are required for cell viability, only specific mutations or truncations can be introduced. Thus, as described at the end of this section, I sought to establish a recombinant expression system for Pol I. Similarly, Pol I co-factors can be efficiently purified from recombinant sources.

2.1 Endogenous tagging of the RPAC1 subunit

While designing the strategy to tag Pol I and Pol III endogenously, I focused on the final application of the cell line, which is to be used for protein purification. Thus, tagging efficiency and ease of use of the resulting cell line were the most critical considerations. Accordingly, I used standard components of the CRISPR-Cas system, including *Streptococcus pyogenes* Cas9 and single-guide RNA (sgRNA). Thus, all details presented in this section concern the *S. pyogenes* Cas9.

The strategy to endogenously tag the RPAC1 subunit (encoded by the POLR1C gene) I have developed is based on a protocol from the Ellenberg group [284] and the suggestions provided by Dr. Francois Dossin. Dr. Helga Grötsch supported basic cell culture work.

2.1.1 crRNA selection via T7 endonuclease assay

The first step in the CRISPR-Cas9 genome editing approach is the selection of crRNAs appropriate for the target locus. crRNA is 20 bp in length and complementary to the genome sequence preceding a consensus PAM sequence (here NGG). Cas9 introduces double-stranded (ds) DNA breaks usually 3 bp upstream from the PAM site [268]. The dsDNA break should be located as close as possible to the insertion site for the efficient integration of the desired insert. To design the crRNAs, I used the online tool CRISPOR [285] and selected three crRNAs closest to the stop codon found in exon 9 of the POLR1C gene (Fig. 2.1 a). Selected crRNA sequences were cloned into a plasmid containing Cas9-GFP and the sgRNA cloning backbone (Addgene PX458 plasmid [286]). The sgRNA consists of the crRNA specific to the target sequence to be cut by Cas9 and the RNA backbone called tracrRNA, which binds and activates the Cas9 nuclease [268,273,278].

To assess the cutting efficiency of the three selected crRNAs, I performed the T7 endonuclease I assay. The T7 endonuclease I recognises and cleaves mismatched and non-canonical DNA structures such as Holliday junctions [287–289]. Efficient DNA cleavage performed by Cas9 in the absence of a repair template stimulates non-homologous end joining (NHEJ) and other DNA repair pathways that lead to the accumulation of such DNA structures and mismatches.

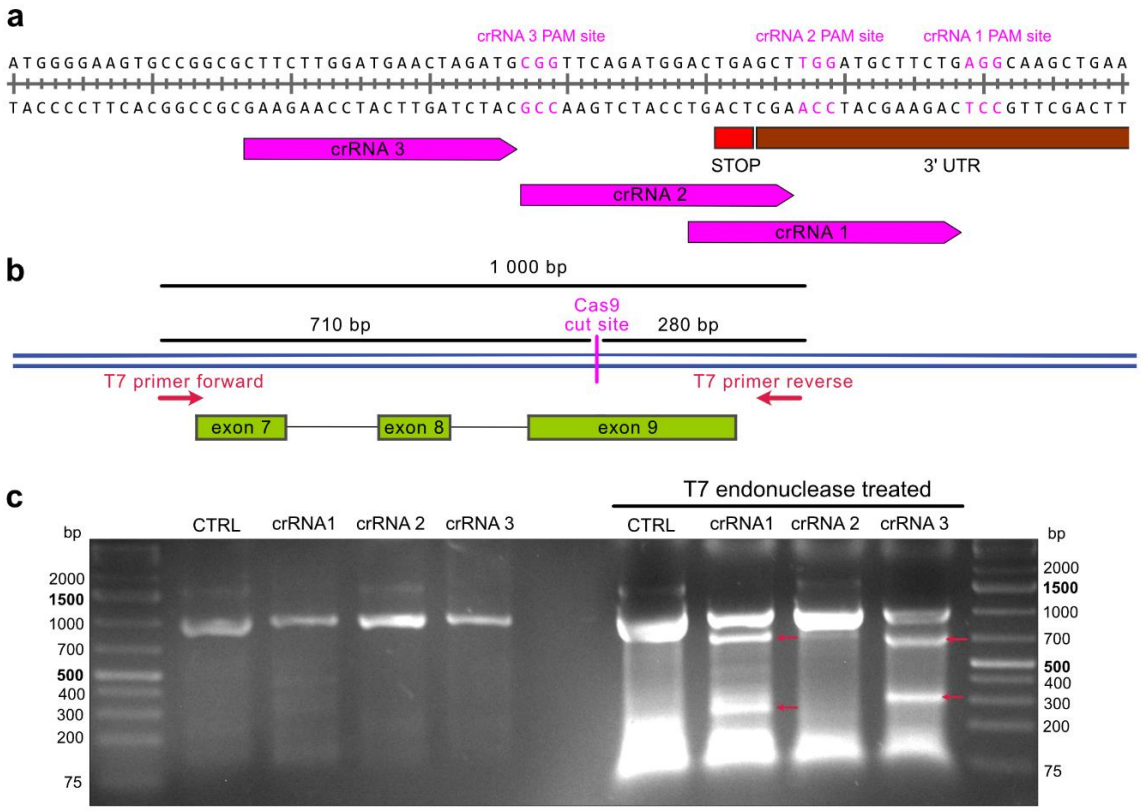

Fig. 2.1. T7 endonuclease I assay to select the most efficient crRNA. **a|** C-terminal sequence of the POLR1C gene coding for the RPAC1 subunit shared between Pol I and Pol III. Locations of the STOP codon (red) as well as the 3'untranslated region (UTR) (brown) are indicated. Designed crRNAs are marked with magenta arrows, and their corresponding PAM sequences are marked by magenta colouring on the DNA. **b|** Design of the primers (red arrows) for the T7 endonuclease I assay. The Cas9 cut site is indicated in magenta. Black lines show the total size of the amplicon (~1 000 bp) and the sizes of the DNA fragments created after T7 endonuclease I cleavage. POLR1C exons are indicated in green. **c|** The result of the T7 endonuclease I assay. Control (CTRL) was not transfected with Cas9-sgRNA plasmid. Lines annotated with crRNA have been transfected with a plasmid carrying Cas9 and the sgRNA scaffold with the indicated crRNA cloned into it. The left panel shows the undigested product of the genomic DNA amplification with the T7 primers. The right panel shows the same product after incubation with T7 endonuclease I (T7 endonuclease treated). Red arrows mark the expected products of the digestion.

To test whether the selected crRNAs efficiently target Cas9 to the C-terminus of the POLR1C gene, HEK293T cells were transfected with the prepared plasmids. Subsequently, I amplified a 1 kb fragment around the site where Cas9 would introduce dsDNA breaks (Fig. 2.1 b). Primers used for the fragment amplification were designed so that the cut site would fall away from the centre of the amplicon. Thus, if the T7 endonuclease I efficiently cuts the mismatched DNA, and two fragments of distinct sizes (~300 and 700 bp) would be detected.

crRNA1 and crRNA3 were more efficient at targeting Cas9 to the desired locus than crRNA2 (Fig. 2.1 c). I selected crRNA1 for further insertion of the RPAC1 tag since this guide spans the tag insertion site. Thus, correct insertion of the DNA fragment will destroy the crRNA1 binding site, preventing the DNA's re-cutting.

2.1.2 Introduction of the C-terminal tag on the RPAC1 subunit

In parallel, I designed the repair template to introduce the desired sequence into the POLR1C gene. The repair template consists of the sequence to be introduced and homology arms (HA) that are complementary to the two sides of the insertion site (Fig. 2.2 a). HA used in this case were 800 bp long since the designed insert is relatively long (over 1.5 kb). The template sequence will be introduced into the genome using cells' DNA repair pathways.

The insert, which I designed, contains several elements interspaced by flexible GS linkers (Fig. 2.2 a, lower panel). The mCherry fluorescent protein allows tracking of the subunit's location, confirming that the tag's introduction does not change the cellular localisation of the target complexes. TwinStrep-tag and 6x His-tag allow efficient, 2-step affinity purification yielding target complexes separated from contaminants. Genome editing with the CRISPR-Cas9 approach usually has low efficiency, with several aspects of the pipeline, such as transfection efficiency or HDR efficiency, being possible culprits [281,282]. By including a step of antibiotic selection, I could significantly boost the number of positive clones. This was especially important given that the used cell line, HEK293, has an unstable genome with polyploidy reported in the region of interest [290]. This means that more than two alleles of the gene must be tagged for homozygosity. Blasticidin S antibiotic and its resistance gene coding for blasticidin S deaminase (BSD) were selected, due to the ease of use of blasticidin S and the small gene size of BSD [291]. A P2A self-cleaving peptide included preceding the BSD allowed the separation of the additional protein from the tag attached to the RPAC1 subunit [291]. Reducing the size of the tag attached to the subunit could be necessary for the correct assembly and function of the whole complex.

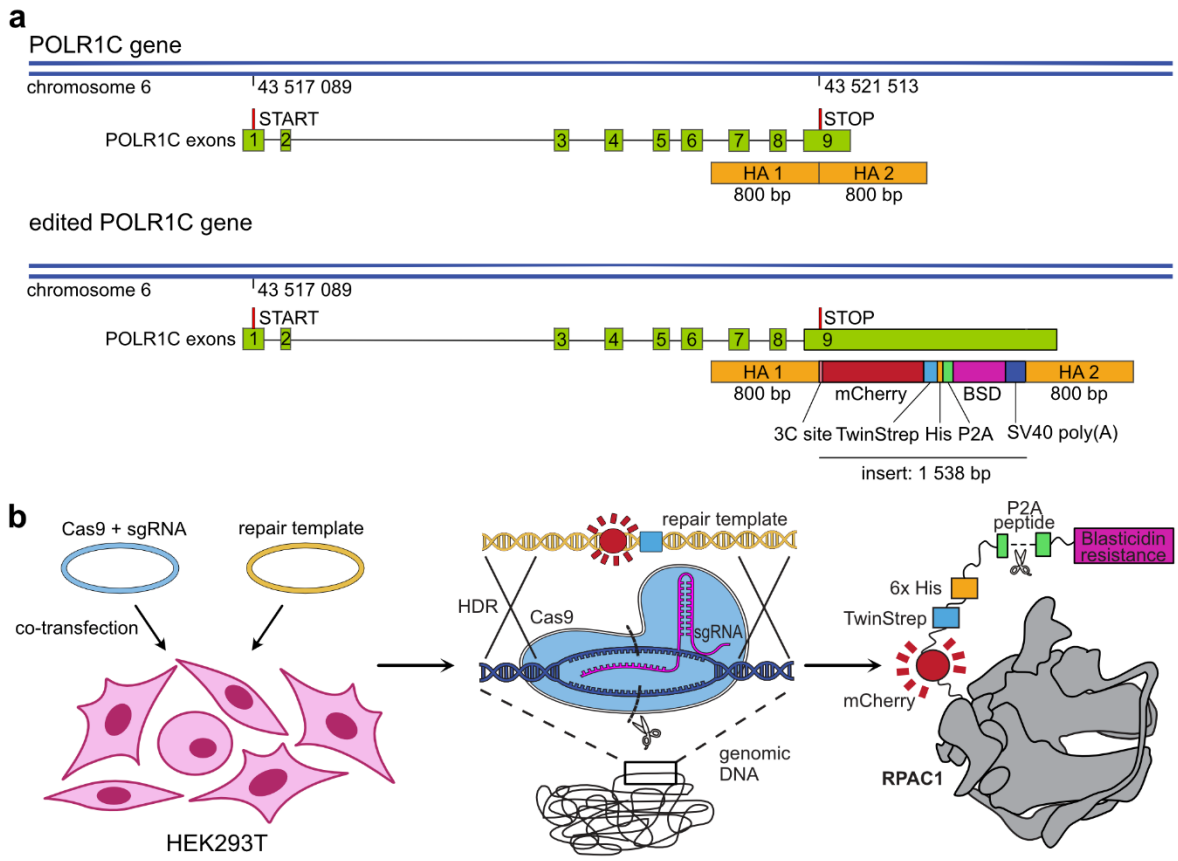

Fig. 2.2. Endogenous tagging of the RPAC1 subunit. a| Genomic context of the POLR1C gene. Exons coding for the RPAC1 subunit are shown as green rectangles; red lines mark start and stop codons; homology arms (HA) are included in the repair template and are shown as orange rectangles. Lower panel: the POLR1C gene after the successful introduction of the tag on the 3'end of the gene. Tag components are marked with coloured rectangles: HRV 3C protease cutting site (3C site), mCherry fluorescent protein (mCherry), Twin-Strep-tag (TwinStrep), 6x His-tag (His), self-cleaving P2A peptide (P2A), blasticidin S deaminase gene conferring resistance to blasticidin S (BSD) and Simian virus 40 poly(A) signal (SV40 poly(A)). All elements are drawn to scale. **b|** Cartoon scheme of the CRISPR-Cas genome editing. HEK293T cells (pink) are co-transfected with the plasmid carrying Cas9-sgRNA (blue) and repair template containing the insert (yellow). Once expressed, Cas9 (blue) is targeted to the desired genomic location by the sgRNA (pink) complementary to the genomic sequence crRNA. It then introduces dsDNA break (marked by scissors), which is then repaired by cells' DNA repair pathways such as homology-directed repair (HDR), which uses the supplied repair template (yellow) carrying the desired insert (represented by mCherry in red and TwinStrep in blue). Once the insert is successfully integrated, it is expressed giving rise to a protein complex (here for simplicity shown as Pol I only) with a tag on the engineered subunit.

In CRISPR-Cas9 genome editing, it has been observed that the DNA surrounding the site of the dsDNA break is sometimes recessed [292]. In the case of the 5' side of the dsDNA break, cells with introduced mismatches will likely be selected out. If any codon shifts are introduced, the BSD gene will likely not be correctly expressed.

Following the insertion site is the 3'UTR, which could be significantly shortened while still allowing protein expression. To prevent decreased levels of gene expression related to shortened 3'UTR, additional polyadenylation (poly(A)) sequence derived from the Simian virus 40 (SV40) is also included in the tag [293]. Additionally, located directly after the RPAC1 coding sequence and before the mCherry is an HRV 3C protease cleavage site (3C site) [294,295], which would allow cleaving of the entirety of the tag in case it is needed.

HEK293T cells were co-transfected with the plasmid carrying the repair template and a plasmid carrying Cas9 and sgRNA (Fig.2.2 b). Cas9 is then targeted to the 3'end of the POLR1C gene by sgRNA carrying sequence complementary to the locus where it introduces a dsDNA cut. DNA repair pathways such as Homology Directed Repair (HDR) need to be encouraged to promote the insertion of the fragment of interest. By providing an excess of the repair template that can anneal around the site of the dsDNA break introduction, the equilibrium can be skewed towards using this template for high fidelity repair of the break without the introduction of mismatches [282]. As a result, the introduced tag becomes part of the open reading frame and is transcribed and later translated as a part of the POLR1C gene. The protein expression levels and their post-translational modifications are not affected by the insertion of the tag.

2.1.3 Selection of monoclonal cell lines

To ensure the highest possible yield during protein purification, it is vital to ensure all alleles of the targeted gene carry the inserted tag. Additionally, CRISPR-Cas9 genome editing can introduce off-target DNA breaks, which may affect the viability or characteristics of cells. To ensure reproducible cell growth and high protein recovery, I aimed to derive a monoclonal cell line with a homozygously introduced insert.

After recovery, cells were transfected with both plasmids and subjected to blasticidin S selection. Subsequently, single-cell derived lines needed to be isolated. I attempted to separate cells using fluorescence-activated cell sorting (FACS), yet the viability of the single cells was extremely low (not shown). Thus, I turned to manual picking of colonies derived from single cells.

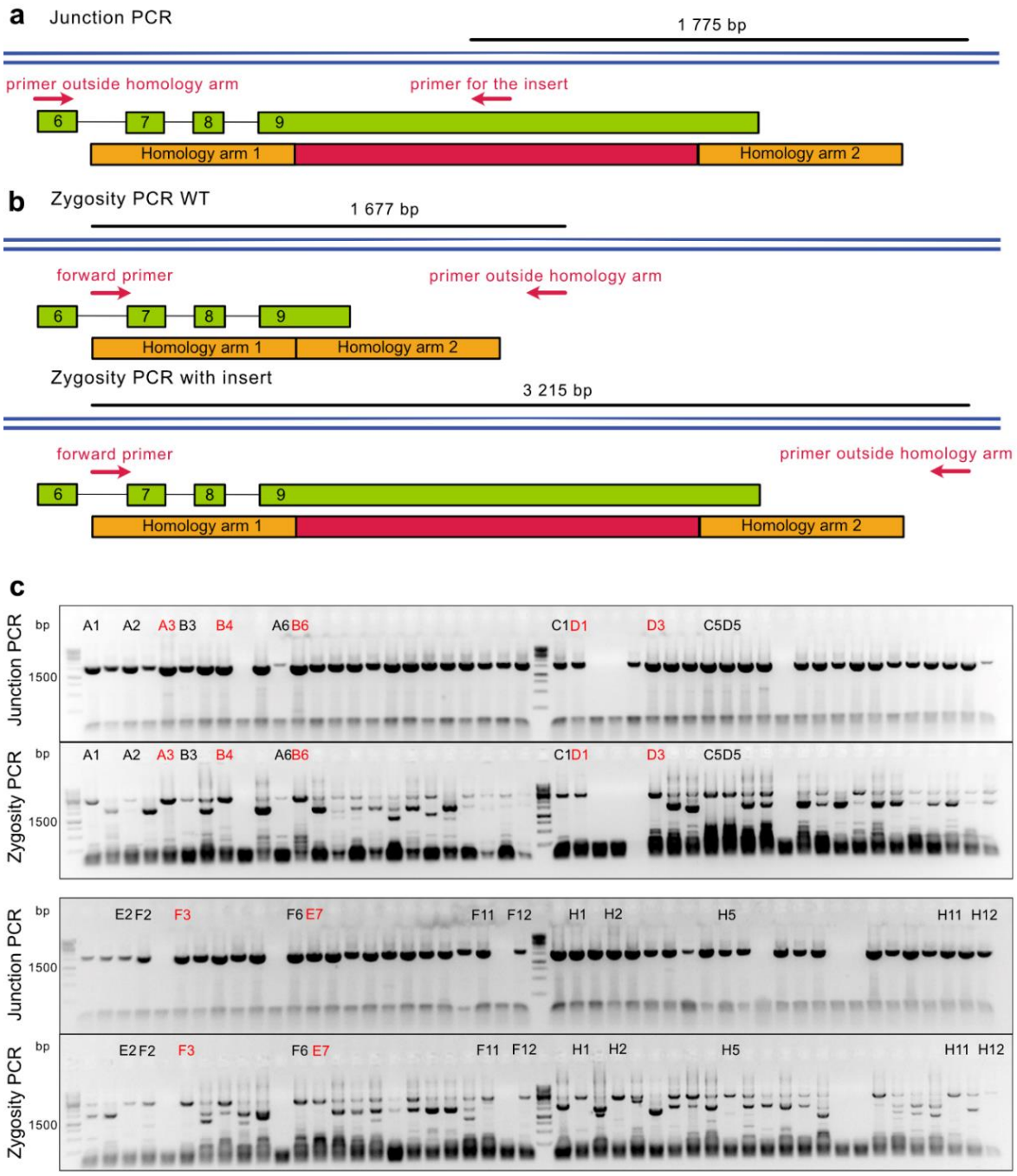

Fig. 2.3. Genotyping of the monoclonal cell lines. a| Junction PCR, which uses a primer pair (red arrows) with one primer annealing to the sequence within the insert and the second primer annealing outside of the homology arm. **b|** Zygosity PCR, where used primers (red arrows) anneal outside of the insert site on either side of the introduced sequence. In the case of the wild-type (WT) gene copy with no insert (upper panel), the PCR product is significantly smaller than in the case of the gene copy with an insert (lower panel). Green rectangles show locations of the POLR1C gene exons; orange boxes indicate locations of the homology arms; red rectangle indicates the insert location. The black bar shows the expected size of the PCR product. **c|** Genotyping PCR gels. Each clone from the 96-well plate is subjected to the junction PCR and zygosity PCR (aligned one above the other). Clones that show homozygous insertion of the tag are annotated with the well number. Those further expanded are marked in red.

Derived cell lines were then genotyped to determine which clones were going to be further expanded and validated. I performed the initial genotyping on cell lysates relying on two PCR reactions: junction and zygosity PCR (Fig. 2.3 a and b respectively). Junction PCR is performed to confirm the presence of the insert. One primer annealing to the sequence inside the tag (here to the TwinStrep tag sequence) and another primer annealing outside the tag and the homology arm were used. As a result, a band is only observed if the insert is present and if it is present at the correct genomic location. While unlikely, it is possible that the insert could be randomly integrated in frame with another gene, giving rise to blasticidin S resistance. Zygosity PCR allows determining if all alleles of the edited gene carry the insert. Both primers are designed to anneal outside of the insert site on either side of the insertion site. When used for DNA amplification from the genomic DNA, they will give rise to products of distinct sizes. When annealing to a WT copy of the gene, the product will be significantly smaller than when the gene carrying the insert serves as a template. Only the higher molecular weight band can be observed when all alleles are successfully edited. A lower molecular band is also present if not all gene copies are edited.

Having isolated 96 clonal lines derived from single cells, I subjected them to the outlined genotyping strategy (Fig. 2.3 c). Unsurprisingly, many clones showed a product of correct molecular weight in the junction PCR. Its absence can be attributed to incorrect cell picking or low viability of the picked cell resulting in no cells in the genotyped well, integration of the insert into a random genomic location or the survival of the clone by chance, without the BSD gene. Interestingly, many clones (labelled with the well number) showed homozygous insertion of the tag into all alleles. It can be speculated that the insertion of more copies of the tag might be beneficial for the cell's survival as it will then express more of the BSD gene product better inactivating blasticidin S. As expected, many clones show insertion of the tag on only some alleles, which can be detected by the presence of bands of both molecular weights. Different distribution of the band intensities between the higher and lower molecular weight bands can be attributed to a different number of alleles with and without the insert. Any bands of other than expected sizes in the zygosity PCR are a testament to how unpredictable DNA repair pathways may be [296]. It is possible that only a portion of the insert was incorporated or that part of the original gene was lost while the dsDNA break was repaired.

2.1.4 Validation of the final cell line

Several lines that showed the homozygous incorporation of the insert were selected and expanded so that I could further validate them and choose the line with the best growth profile.

The initial junction and zygosity PCRs were performed on the cell lysate. To remove the potential contaminants and better control DNA amounts, I extracted genomic DNA using the Mammalian GenElute kit (ThermoFisher) and repeated the genotyping PCRs (Fig. 2.4 a). Most clones showed products consistent with the initial screening, however, many more unspecific products could be detected. For successful use of the cell line, the insert must not only be present in the genome but also needs to be correctly expressed. To test for the expression of the tag, I performed Western Blot (WB) using an anti-mCherry antibody (Fig. 2.4 b, upper panel). GAPDH levels were checked using the anti-GAPDH antibody as a loading control (Fig. 2.4 b, lower panel). All selected clones showed expression of the tag measured by the mCherry expression, while no mCherry signal was detected in the WT cells. Further, I aimed to check the correct localisation of the tagged complexes inside cells to confirm that the introduced tag does not affect the function of RNAPs (Fig. 2.4 c). Since Pol III is found at lower than Pol I concentrations and its localisation is more dispersed, I expected only to be able to identify the signal coming from Pol I, which should localise to the nucleolus. Indeed, with the use of live-cell confocal imaging with the addition of the Höchst DNA stain, I could confirm that the signal localises to the nucleolus as expected. The nucleolus shows as a region that excludes Höchst staining due to lower DNA concentration in the area [297]. The mCherry signal localises precisely to those regions where DNA stain is excluded forming characteristic speckles [298]. Additionally, to all those validation measures, I monitored the growth rate of the selected clones to ensure that it is comparable to the WT cells.

Taken together, I selected the D3 clone to be further used for Pol I and Pol III purification. All the validation tests ensured that this cell line has the desired tag incorporated into all alleles of the POLR1C gene and correctly expresses the tag, which does not influence the localisation of RNAPs inside cells.

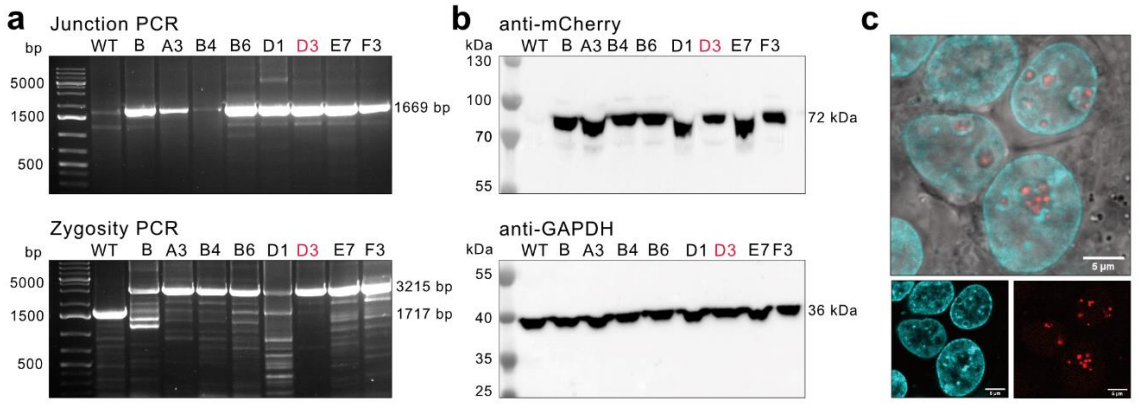

Fig. 2.4. Validation of selected monoclonal cell lines. a| Genotyping PCRs: junction PCR (upper panel) and zygosity PCR (lower panel) of wild-type (WT) and selected monoclonal cell lines. Expected band sizes are indicated. **b|** Western blots of WT cells and selected monoclonal cell lines using an anti-mCherry antibody (upper panel) and anti-GAPDH antibody as a loading control (lower panel). The expected molecular weight is annotated on the right. Clones are annotated with the number of the well in which they were initially located after single-cell clones isolation. The cell line chosen for further experiments (D3) is red. **c|** Confocal microscopy image of the D3 cell line. The upper panel shows a merge between the transmission image and fluorescent channels (shown as split channels below) of the DNA stained with Höchst (cyan) and Pol I and Pol III with endogenously introduced tag carrying mCherry (red). The cropped lower panel in **a** and upper panel in **b** were published in [36]. Panel **c** is adapted from [23].

2.1.5 Selection of the cell line used for the genetic engineering

The main concern associated with the strategy to purify the endogenous human Pol I and Pol III regarded the potential low yield of the obtained protein after purification. To circumvent the issue, I aimed to use the Expi293F cell line (ThermoFisher), which was developed to allow cell growth at high density (over 15 x 10^6 cells/ml according to the producer). While the genetic engineering proved successful in the Expi293F cells (Fig. 2.5), I could not isolate a monoclonal cell line due to issues with cell expansion from single cells after FACS. A polyploid cell line could also be used, given that the antibiotic selection increased the ratio of cells carrying the insert. Yet, high heterogeneity between cells prevented precise genotyping of the cell line. Thus, selecting a monoclonal cell line was essential to verify the tag's homozygous insertion to all copies of the POLR1C gene, allowing higher protein recovery during purification and ensuring the stability of the cell line for future use. With adherent HEK293T cells, I could isolate monoclonal cell lines. In a novel approach, I adapted them to grow in suspension in the Expi293 Expression Medium, yielding high cell concentrations (a concentration of 7 x 10^6 cells/ml was routinely used).

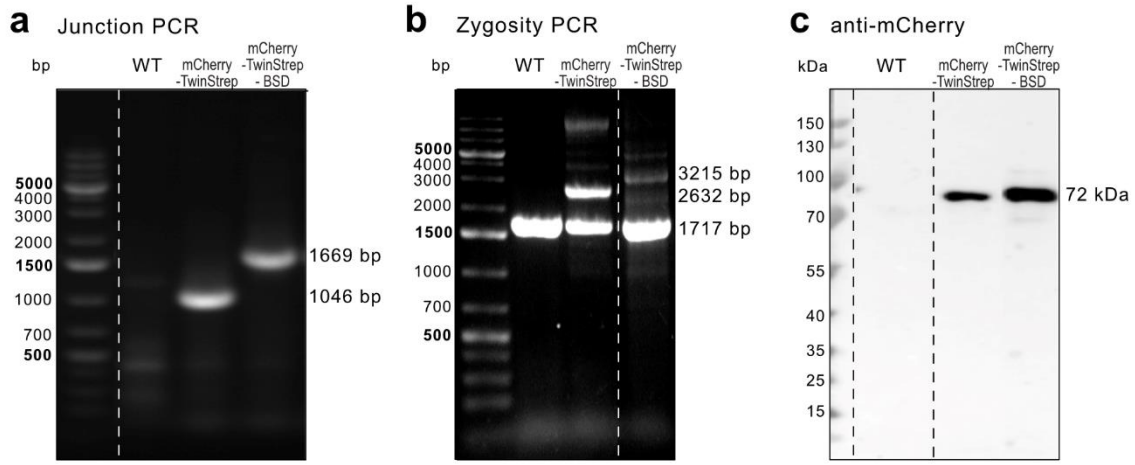

Fig. 2.5. Validation of CRISPR-Cas edited Expi293 cells. a| Junction PCR. **b|** Zygosity PCR. **c|** Western blot with anti-mCherry antibody. In each panel from left to right: WT cells, cells containing the insert with mCherry-TwinStrep-His-tags and cells with the full insert containing additionally the P2A peptide, BSD gene and SV40 poly(A) signal. The expected molecular weight is indicated on the right.

2.1.6 Construction of additional cell lines

In addition to the cell line carrying a mCherry-TwinStrep-His-tag on the C-terminus of the RPAC1 subunit, I also constructed three other cell lines using the approach outlined (Fig. 2.6). The modifications included:

- RPA2 subunit tagged with mCherry-TwinStrep-His-tag
- RPA2 subunit tagged with FLAG-Halo tag
- RPAC2 subunit tagged with mCherry-TwinStrep-His-tag

The cell lines with the RPA2 subunit tagged were created to allow different purification approaches in case the separation of Pol I and Pol III was not possible. Ultimately, they were not used since purification of both Pol I and Pol III simultaneously is more time and cost-effective. Additionally, the cell line with the Halo tag is compatible with super-resolution studies since an organic dye can be coupled to the Halo tag. This avenue of *in cellulo* Pol I studies remains to be explored.

The cell line with the tag on the RPAC2 subunit (Fig. 2.6 c), the second subunit specifically shared between Pol I and Pol III, was constructed to allow studies of the influences of disease-associated mutations on Pol I and Pol III. The use of this cell line is further explored in section 4.2. This cell line was constructed with the help of Granita Lokaj, a student under my supervision.

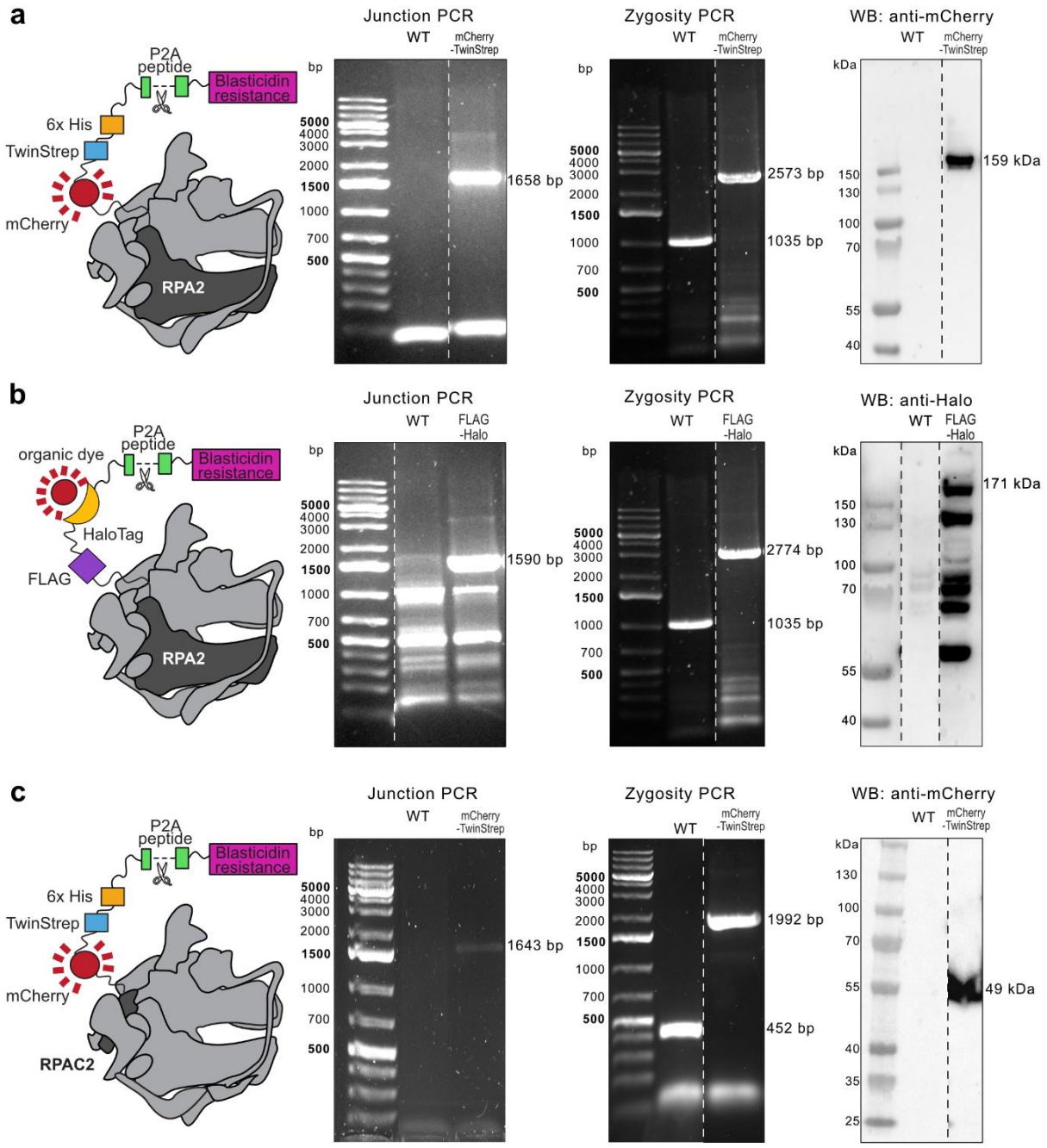

Fig. 2.6. Validation of additional cell lines that were constructed. **a|** RPA2 subunit tagged with mCherry-TwinStrep-His tag. **b|** RPA2 subunit tagged with FLAG-Halo tag. **c|** RPAC2 subunit tagged with mCherry-TwinStrep-His tag. For each cell line, with the modification shown in cartoon representation, validation involved (left to right): junction PCR, zygosity PCR and Western blot (WB). Sizes of the expected products are indicated on the right.

2.1.7 Conclusions

In section 2.1, I have outlined the methodology I established to insert endogenous tags on the subunits of interest. The approach has been used to create four cell lines, highlighting its versatility. As listed below, several adjustments have been incorporated into the method, making it particularly suitable for further use in protein purification.

(i) **Design features of the inserted tag.**
When selecting the elements incorporated into the tag, the versatility of its use was my main focus. Thus, the tag is relatively large, which for other approaches where minimal disturbance of the cells is desired might not be optimal. Insertion of the additional poly(A) signal ensures high gene expression.

(ii) **Ease of cell line selection.**
The antibiotic selection step boosts the ratio of positive clones and removes the need for additional screening steps. Hand-picking of homozygous colonies speeds up the tagging procedure.

(iii) **Ensuring homozygous tag insertion.**
Homozygous tag insertion is essential to increase the yield of the purified proteins. The use of high antibiotic concentration for clone selection increases the tag incorporation.

(iv) **Set of validation methods compatible with the downstream application of the cell line.**
Selected validation methods are quick, inexpensive and require only standard laboratory equipment. Additional steps such as whole-genome sequencing have not been included in the pipeline for now.

(v) **Selection of the cell line and adaptation to cell growth in suspension.**
HEK293T cells have been selected because of their ease of handling and the opportunity to adapt them to growth in suspension in a medium that supports the high-density growth of cells. This was crucial in obtaining sufficient starting material for the purification.

2.2 Purification of human Pol I and Pol III

After obtaining a stable cell line with endogenously tagged Pol I and Pol III, which could be grown in suspension, I purified the target complexes. For cryo-EM studies, a homogenous sample of high purity is required. Protein concentrations between 0.05–5 µM are commonly used for the sample preparation [299].

2.2.1 Small-scale purification of human Pol I and Pol III

The suspension cell line was grown to 7 x 10^6 cells/ml. At higher concentrations, the viability of the cells dropped below 95%, which could negatively affect protein yields. I used 300-600 ml of cell suspension for initial trials, corresponding to a pellet of 5-10 g. The purification strategy was based on the affinity pulldowns, first using Ni-NTA beads and further with Strep-Tactin beads. Initially, I tracked the purification progress using WB (Fig. 2.7 a) since protein yield was low and could only be detected on an SDS-PAGE gel with a silver stain showing many bands (Fig. 2.7 b). Still, those initial trials proved the concept before moving to a larger scale of purification.

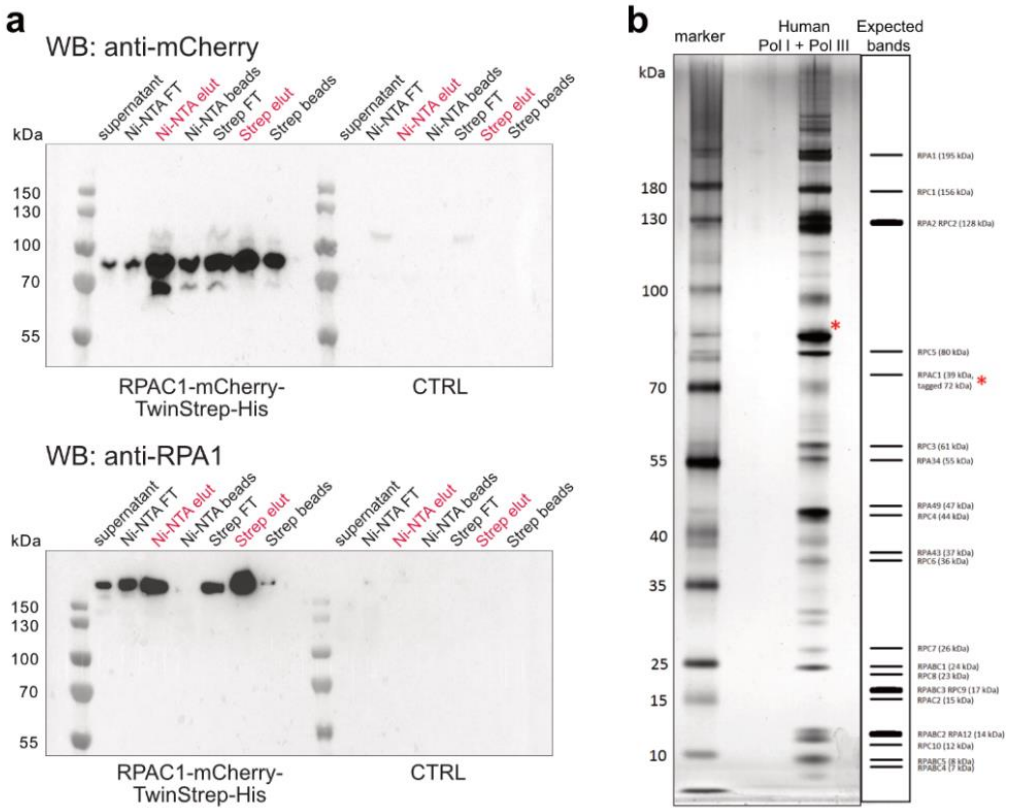

Fig. 2.7. Initial human Pol I and Pol III purification trials. a| Western blots (WB) of the fraction collected throughout the purification. Anti-mCherry (upper panel) and anti-RPA1 (lower panel) antibodies were used. WT cells were used as a control (CTRL). Flow-through (FT), elution (elut). Elution fractions are marked in red. **b|** Silver stained SDS-PAGE gel of final sample. Expected bands are shown on the right. A red asterisk marks the tagged subunit.

2.2.2 Human Pol III purification

In an optimised purification, I used 3 L of cell suspension, which produced a pellet of about 50 g. The general pipeline I have developed included: cell lysis and sonication, ultracentrifugation, lysate filtering, Ni-NTA affinity purification, Strep-Tactin affinity purification, an anion-exchange chromatography (MonoQ) and buffer exchange coupled to sample concentration. Essential modifications that I introduced which allowed for higher protein recovery and increased purity involved:

- addition of benzonase nuclease to the lysis buffer to efficiently remove the nucleic acid contamination,
- resuspension of the cell pellet in a large volume of the lysis buffer (using 300 mL of lysis buffer per 50 g of cell pellet) to reduce clogging of the Ni-NTA column,
- using the Ni-NTA column instead of beads,
- using Strep-Tactin beads instead of the column,
- use of a small anion-exchange column (below 1 ml of bed volume) run on the Äkta Pure Micro, which allowed eluting in fractions of up to 50 µl.

The purification protocol yielded a reproducible peak pattern obtained during anion-exchange chromatography (Fig. 2.8 a). Similarly, as in the case of the yeast Pol I and Pol III purification, human Pol I eluted earlier than human Pol III. Unfortunately, the small peak corresponding to Pol I only contained an incomplete complex with heterodimer subunits in sub-stoichiometric ratios. Additionally, what is not the case for yeast polymerase purification, an intermediate peak that contained both Pol I and Pol III was observed. Most of the complete Pol I eluted in this peak, together with Pol III, which lacks the heterotrimer subunits. Further, Pol III eluted in two distinct peaks. When samples from each peak were run on an SDS-PAGE gel (Fig. 2.8 b), a high yield of especially Pol III with all subunits in stoichiometric ratios could be appreciated.

After concentration, I could obtain up to 50 µl of human Pol III at concentrations reaching 4 mg/ml. The identity of all Pol III subunits could be confirmed with mass spectrometry (MS) (Fig. 2.8 c). A transcription assay performed by Dr. Florence Baudin confirmed that human Pol III is transcriptionally active (Fig. 2.8 d). Human Pol III can extend the radioactively labelled RNA primer bound to an open DNA bubble in the presence of NTPs.

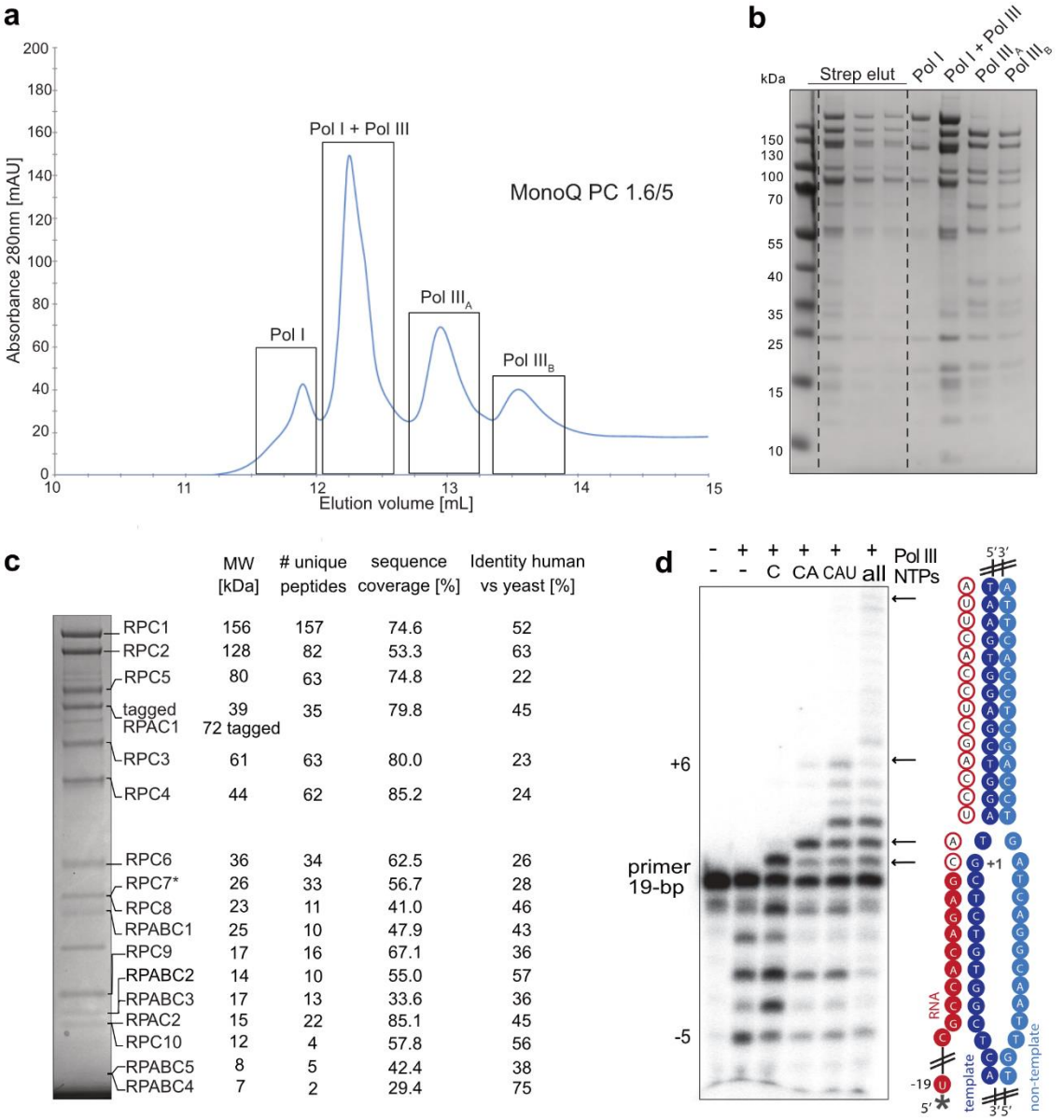

Fig. 2.8. Human Pol III purification. **a|** Anion-exchange chromatography profile using MonoQ PC 1.6/5 column. The four visible peaks with the absorbance at 280 nm are annotated with the identified complexes. **b|** Coomassie blue-stained SDS-PAGE gel of selected fractions from the purification. Strep-Tactin beads elution (Strep elut) and samples from the peaks separated on the ion-exchange column are shown. **c|** Concentrated sample of human Pol III run on an SDS-PAGE gel visualised with Coomassie blue stain. The identity of the bands was confirmed with MS, with the main statistics shown on the right. The percentage identity of each subunit between humans and *S. cerevisiae* is also shown. **d|** *In vitro* transcription assay using represented as a cartoon artificial DNA bubble (blue) with a 5' radioactively labelled (asterisk) RNA primer (red). The presence or absence of Pol III and NTPs is indicated on top. Expected products are marked with arrows. Panels **c** and **d** are taken from [36] in a modified form.

2.2.3 Human Pol I purification

The established strategy allowed me to purify intact human Pol III reproducibly and, to some extent, human Pol I. Yet, the protocol needed to be optimised to recover a complete human Pol I. In the end, I introduced three main changes:

- Replaced 150 mM $(NH_4)_2SO_4$ with 200 mM CH_3CO_2K in the purification steps from the wash of Strep-Tactin beads onwards. I opted to use $(NH_4)_2SO_4$ in the initial purification steps as it efficiently stripped away contaminants. CH_3CO_2K has lower ionic strength and thus might help stabilise the complete Pol I.
- Changed the pH of the buffers from 7.5 to 8.0 for all steps except for the last ion-exchange column. This choice of buffer pH directly mimics the established yeast Pol I and Pol III purification protocol.
- Used a MiniQ 4.6/50 PE column instead of MonoQ PC 1.6/5. This change was motivated by column clogging leading to overpressure issues while using the MonoQ PC 1.6/5 column. While the MiniQ 4.6/50 PC column is larger (0.8 ml compared to 0.1 ml in the case of MonoQ PC 1.6/5), it has a smaller particle size (3 µm instead of 10 µm), giving it a higher resolution.

All eluted proteins were buffer exchanged into $(NH_4)_2SO_4$, as they precipitated on the concentrator when CH_3CO_2K was used.

With the outlined changes in place, the purification profile obtained during anion-exchange chromatography changed (Fig. 2.9 a). Two peaks containing Pol I and only one broad peak containing Pol III were present. Only a small peak with both Pol I and Pol III, running close to the second Pol I peak, was observed (Fig. 2.9 b). Only the first peak containing Pol I (annotated as Pol I_A in Fig. 2.9) was used for further studies to ensure higher homogeneity of the sample and lack of contamination with Pol III subunits coming from the mixed Pol I-Pol III peak. While the peak absorbance at 280 nm observed for each species is significantly lower compared to the previous purification conditions, the recovered protein yield is comparable because of the larger elution volume. After concentration and buffer exchange, I obtained about 50 µl of Pol I at 2 mg/ml. Recovered Pol I contained all subunits as confirmed by MS analysis (Fig. 2.9 c). Pol I was also confirmed to be transcriptionally active through an *in vitro* primer extension assay performed by Dr. Florence Baudin (Fig. 2.9 d).

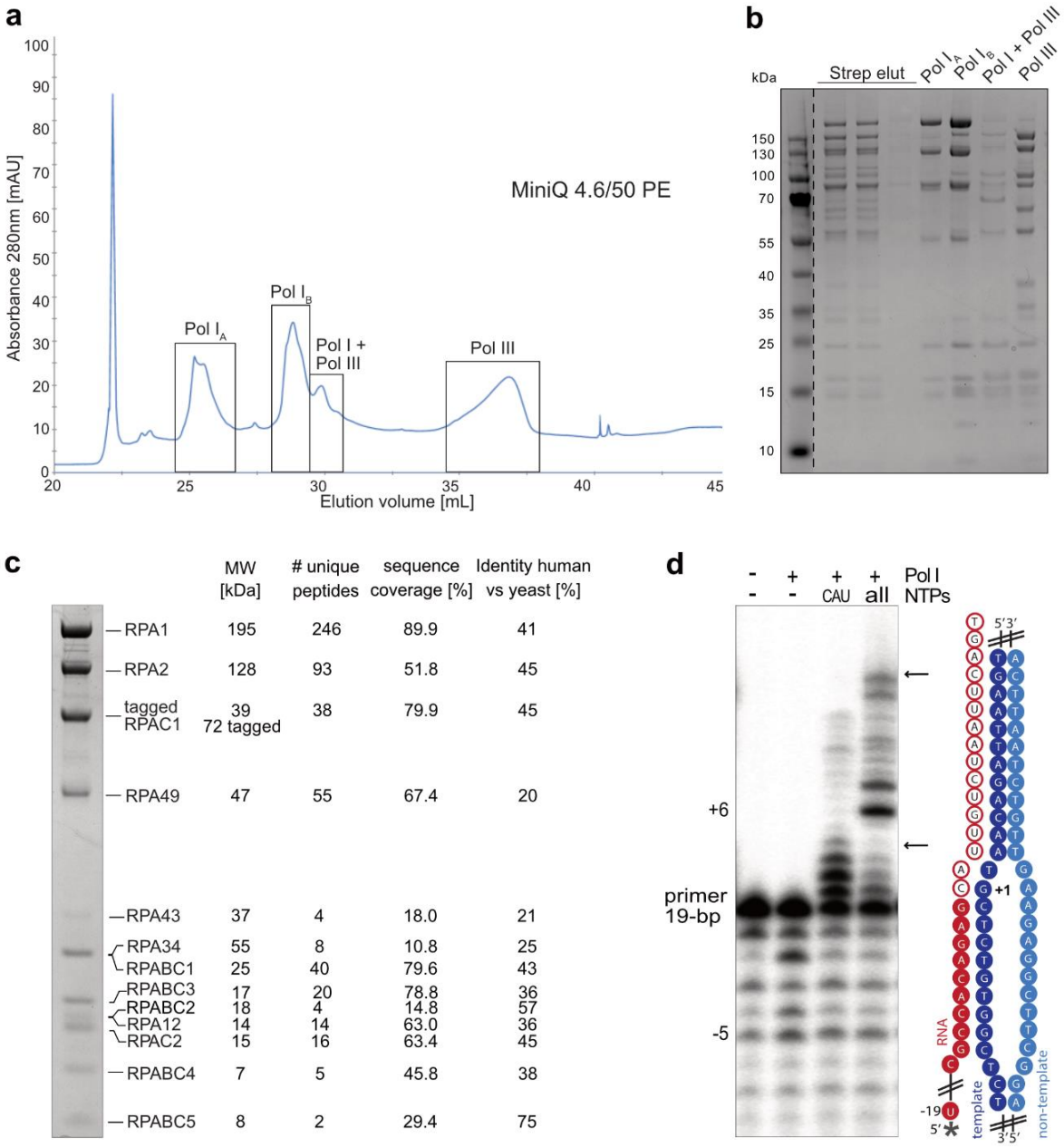

Fig. 2.9. Human Pol I purification. **a|** Anion-exchange chromatography profile using MiniQ 4.6/50 PE column. The four visible peaks with the absorbance at 280 nm are annotated with the identified complexes. **b|** Coomassie blue-stained SDS-PAGE gel of selected fractions from the purification. Strep-Tactin beads elution (Strep elut) and samples from the peaks separated on the ion-exchange column are shown. **c|** Concentrated sample of human Pol I run on an SDS-PAGE gel visualised with Coomassie blue stain. The identity of the bands was confirmed with MS, with the main statistics shown on the right. The percentage identity of each subunit between humans and *S. cerevisiae* is also shown. **d|** *In vitro* transcription assay using represented as a cartoon artificial DNA bubble (blue) with a 5' radioactively labelled (asterisk) RNA primer (red). The presence or absence of Pol I and NTPs are indicated on top. Expected products are marked with arrows. Panels **c** and **d** are taken from [23] in a modified form.

2.3 Recombinant components of human Pol I machinery

In future studies, determining the human Pol I PIC structure is highly interesting. Thus, I aimed to initiate the production of the Pol I co-factors required. Additionally, a recombinant Pol I expression system might prove necessary for further studies of Pol I mutations. In this section, I present the intermediate results of constructing the recombinant expression systems for Pol I and its co-factors.

2.3.1 Recombinant Pol I

Full-length genes for ZNRD1 (RPA12 subunit), CD3EAP (RPA34 subunit), TWISTNB (RPA43 subunit) and POLR1E (RPA49 subunit) were ordered from DNA SU (details in Table A1 in the Appendix). I amplified them (Fig. 2.10 a) and cloned them into the pcDNA3 backbone using the restriction-free cloning method (see section 6.1.1).

POLR1A gene (coding for RPA1 subunit) was not commercially available. I thus cloned it from the HeLa cDNA kindly provided by Maja Starostecka. Given the large gene size (5 160 bp), I decided to split it into two parts and clone them separately (Fig. 2.10 b). Using Sanger sequencing, I confirmed the presence of the whole POLR1A gene in the pcDNA3 backbone. Unfortunately, sequencing revealed the presence of two mutations: the CGA codon mutated into CAA resulting in R339Q mutation and the CGT codon mutated to AGT, giving rise to the R1506S mutation that arose during the cloning procedure. Additionally, in the POLR1A gene at position 310-318 in the coding sequence, a PmeI restriction site is located. Given that the cloning strategy is based on the biGBac method, which uses PmeI to release multi-gene gene cassettes, this sequence also needed to be changed [300]. Thus I introduced a silent TTA to TTG mutation. To introduce all those point mutations, I used a restriction-free cloning strategy [300] with two annealed primers carrying the desired mutation acting as a mega primer (section 6.1.2.). The introduction of all desired changes was confirmed with Sanger sequencing.

The only available clone of the POLR1B gene (coding for the RPA2 subunit) was missing half of exon 7 and exons 8, 9 and 10. Firstly, I cloned the available part of the gene into the pcDNA3 backbone (Fig. 2.10 c). Next, I amplified the sequence of the missing exons from the cDNA and inserted it into the remaining part of the gene. Lastly, I added a C-terminal TwinStrep-His-tag to the complete construct. Each cloning step was validated by Sanger sequencing.

Remaining genes which are shared with Pol III: POLR1C (RPAC1 subunit), POLR1D (RPAC2 subunit), POLR2E (RPABC1 subunit), POLR2F (RPABC2 subunit), POLR2H (RPABC3 subunit), POLR2K (RPABC4 subunit) and POLR2L (RPABC5 subunit) were provided by Dr. Mathias Gribig already cloned into the pcDNA3 vector.

Subsequently, using the biGBac protocol [300] (details in section 6.1.3.), I aimed to combine the single genes into the multigene construct. Initially, I tried to form 3 plasmids with 4 - 5 genes in Big1a, Big1b and Big1c plasmids. The strategy, however, did not work due to the low efficiency of integration of gene cassettes in the 1st step of the biGBac assembly protocol. Instead, I divided the genes into five separate vectors. The cloning was performed by Granita Lokaj, an internship student under my supervision. Genes were paired based on their location within the complex and their sizes to simplify cloning. Created plasmids contained (Fig. 2.10 d):

- Big1a: POLR1C, POLR1D, POLR2L
- Big1b: TWISTNB, CD3EAP, POLR1E
- Big1c: POLR2E, POLR2H, POLR2F
- Big1d: POLR1A, ZNRD1
- Big1e: POLR1B, POLR2K

A diagnostic digest with SwaI was performed to check for cassettes' integration. After the assembly of the target plasmid, SwaI restriction sites are present between each cassette, which contains the gene of interest, CMV promoter and bGH polyadenylation signal. After SwaI digestion, cassettes and the Big1 backbone (2 640 bp long) are released. Each plasmid showed expected band sizes compared to the simulated agarose gel pattern (Fig. 2.10 d, left). The absence of mutations and correct integration of all genes was confirmed with Sanger sequencing.

Finally, all gene cassettes from all five plasmids could be released with PmeI digest and integrated into one Big2abcde plasmid backbone. Individual gene cassettes are again released when the final plasmid is digested with SwaI. Compared to the simulated agarose gel (Fig. 2.10 e, left), one of the clones showed the expected pattern. Insertion of all genes was validated with whole plasmid sequencing.

The obtained plasmid is ready for the expression of recombinant human Pol I in mammalian cells. Dr. Mathias Girbig tested this strategy for producing human Pol III. Purification of recombinant Pol III yields comparable protein amounts to the purification of the endogenous complex. Access to the recombinant Pol I expression

system allows for a more straightforward introduction of mutations and truncations. The introduction of endogenous mutations is cumbersome and limited to only mutations that support cell viability. In the recombinant expression system, any modification can be introduced. Overexpressing all subunits from one plasmid helps maintain the stoichiometry of the complex.

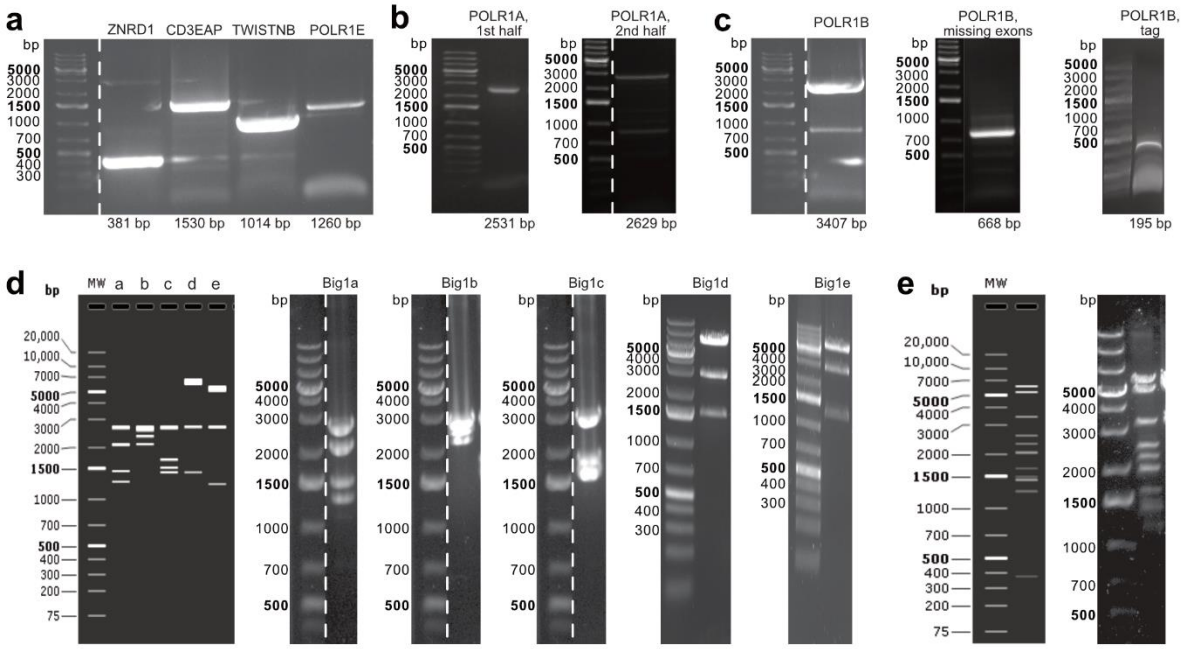

Fig. 2.10. Cloning recombinant human Pol I. a| Amplification of ZNRD1, CD3EAP, TWISTNB and POLR1E genes before cloning them into pcDNA3 backbone using the restriction-free cloning method. **b|** Amplification of POLR1A gene from cDNA, with 1st half of the sequence (left) and 2nd half of the sequence (right) done separately. **c|** Amplification of POLR1B gene, with commercially acquired clone missing middle exons (left), amplification of the missing sequence covering half of exon 7 and exons 8-10 (middle) and amplification of the added C-terminal tag (right). **d|** Diagnostic digest with SwaI restriction enzyme performed after the Pol I genes were assembled into five Big1 plasmids with simulated agarose gel (left) and individual Big1 plasmid digests (right). **e|** Diagnostic digest with SwaI restriction enzyme performed after combining all human Pol I genes into one Big2abcde plasmid. Individual gene cassettes are released giving a pattern consistent with the simulated agarose gel (left). The molecular weight (MW) ladder with indicated sizes of the bands is shown on the left of each gel. Expected product sizes are indicated below the gel or on the simulated agarose gel on the left of the panel. Granita Lokaj produced gels presented in panels **d** and **e**.

2.3.2 SL1

In humans, SL1 comprises five subunits (TAF1A, TAF1B, TAF1C, TAF1D and TAF12) and functions in a complex with TBP [301–305]. The yeast functional counterpart of SL1 is CF, which comprises only three subunits and does not interact with TBP [61–63]. The three yeast CF subunits have human homologues: Rrn6 is homologous to TAF1A, Rrn7 to TAF1B and Rrn11 to TAF1; however, with low sequence identity [300].

SL1 is highly post-translationally modified [306,307], and thus, in the absence of some post-translational modifications or assembly factors only found in human cells, it could be unstable. To circumvent these issues, I attempted to express SL1 in human cells. For cloning, I used synthetic genes already present in the Müller group.

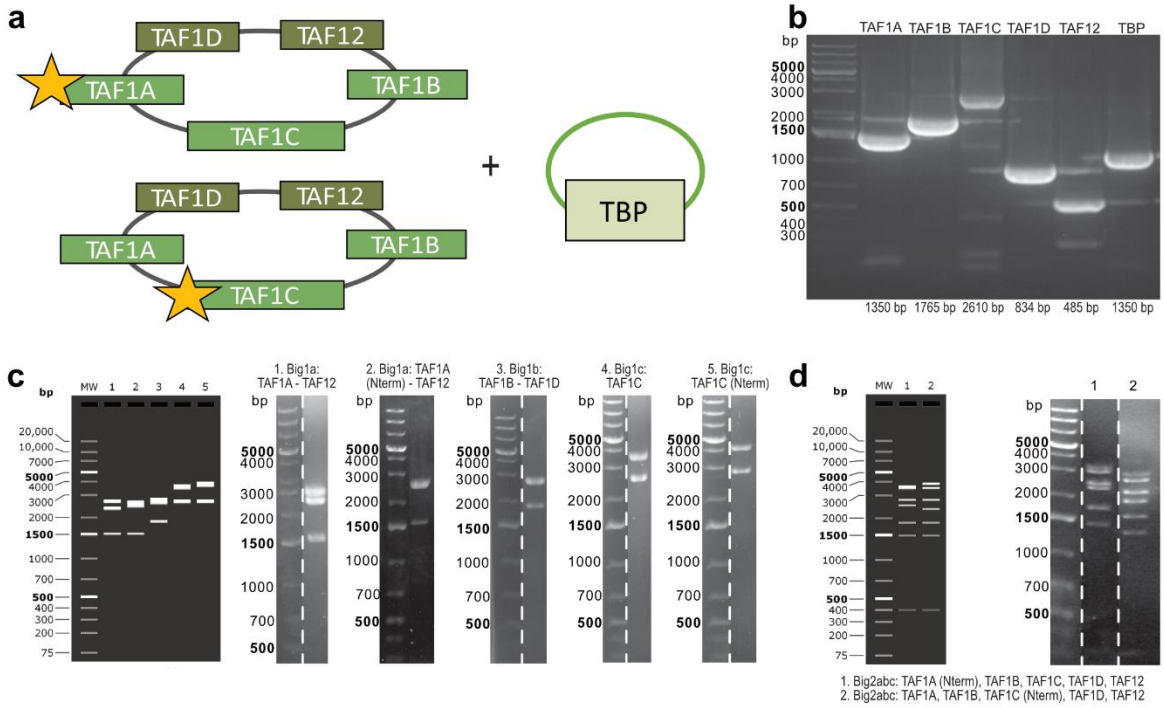

Fig. 2.11. Cloning recombinant SL1. a| Cartoon representation of the co-expression strategy for the SL1 complex. All SL1-specific subunits are cloned into a standard plasmid with one of the two chosen tags (star) and co-transfected with a second plasmid carrying TBP. **b|** Amplification of SL1 subunits and TBP. Expected product sizes are indicated below the gel. **c|** Diagnostic digest with SwaI restriction enzyme performed after the assembly of SL1 genes into Big1 plasmids with simulated agarose gel (left) and individual Big1 plasmid digests (right). **d|** Diagnostic digest with SwaI restriction enzyme performed after combining all SL1 genes into one Big2abc plasmid. Individual gene cassettes are released giving a pattern consistent with the simulated agarose gel (left). The molecular weight (MW) ladder with indicated sizes of the bands is shown on the left of each gel.

To produce yeast CF, in the Müller group, Rrn6 and Rrn7 have been tagged on the N-terminus [61,62]. However, in the Cramer group, the strategy was to insert a tag on the C-terminus of Rrn6 and the N-terminus of Rrn7 [63]. Published strategies to tag SL1 either for purification or *in vivo* studies introduced tags on the N-terminus of TAF1A, TAF1C [302] or TAF1D [303] or the C-terminus of TAF1A [308]. Taking this information into account, I decided to make two constructs in parallel: with TAF1A or TAF1C N-terminally tagged (Fig. 2.11 a). Those are two larger subunits within the complex, so the insertion of a tag in this location could help maintain the completeness of the assembly during purification. To allow a more versatile purification approach, I selected the TwinStrep-His-tag.

I cloned each of the five SL1 genes into the pcDNA3 vector for biGBac assembly (Fig. 2.11 b). I also cloned TBP into the pcDNA3 vector, which can be directly expressed in mammalian cells (Fig. 2.11 b). Given the high efficiency of transfection and expression of TBP, I decided to keep TBP on a separate plasmid and co-transfect it with the plasmid containing all five SL1 subunits (Fig. 2.11 a). Having the other genes on one plasmid could support the stoichiometric assembly of the complex.

I originally planned to directly clone all five genes into Big1a, which allows the assembly of up to 5 genes. However, this strategy proved unsuccessful, and thus I split the genes into three Big1 plasmids (Fig. 2.11 c). Gene combinations included:

- Big1a: TAF1A (with or without N-terminal tag) and TAF12
- Big1b: TAF1B and TAF1D
- Big1c: TAF1C (with or without N-terminal tag)

Subsequently, I combined cassettes from all three Big1 plasmids into the Big2abc vector. I obtained two versions of the final plasmid (Fig. 2.11 d): one with TAF1A carrying the N-terminal tag and one with TAF1C tagged on the N-terminus. Prepared plasmids are ready to be used to purify recombinant SL1, which I have not yet optimised due to time constraints.

2.3.3 RRN3

Human recombinant RRN3 was produced by Dr. Brice Murciano. Thus the purification strategy is only briefly mentioned here.

RRN3 with a His-tag on the N-terminus was expressed in the LOBSTR *Escherichia coli* strain. A three-step purification involved Ni-NTA affinity pulldown with cleavage of the tag and reverse Ni-NTA followed by anion exchange and size-exclusion chromatography (Fig. 2.12). The pure sample was visualised using coomassie blue-stained SDS-PAGE gel with a band at the expected size of 74 kDa. Prepared protein could directly be used for the cryo-EM studies after incubation with human Pol I.

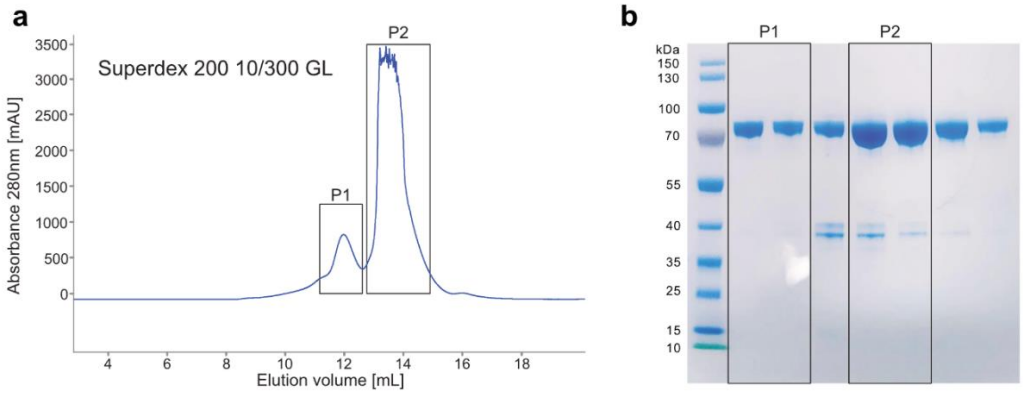

Fig. 2.12. Recombinant RRN3 purification. a| Size-exclusion chromatography profile using Superdex 200 10/300 GL column. There are two visible peaks in the absorbance at 280 nm. **b|** Coomassie blue-stained SDS-PAGE gel of selected fractions from the purification. Annotated peaks are marked. Purification performed by Dr. Brice Murciano.

2.3.4 Conclusions

The outlined methodology allowed me to obtain pure, intact and active human Pol I and Pol III complexes in quantities sufficient for cryo-EM studies. While using CRISPR-Cas9 to endogenously tag proteins for their purification has grown in popularity, when I was establishing described methods, only one study using a similar approach was available [309]. It is remarkable that the purification of two large, multisubunit complexes using one tag is possible and yields sufficient protein quantities for structural studies. Additionally, I outlined the initial steps in producing the recombinant components of the human Pol I machinery. In the follow-up studies, they will play an essential role in investigations of various stages in the Pol I transcription cycle.

STRUCTURES OF THE HUMAN POL I

3

Having acquired a pure, homogenous and active sample of native human Pol I, as outlined in section 2, I set out to determine its atomic structure. The obtained protein quantities were sufficient for efficient cryo-EM studies.

This section discusses the structural studies of human Pol I in various functional states. I investigated Pol I in the apo form, bound to the complete transcription bubble (DNA-RNA scaffold) mimicking the elongation complex (EC), bound to an open DNA bubble in Open Complex (OC) conformation as well as bound to the initiation factor RRN3. The obtained high-resolution structures allowed me to gain insight into peculiarities of the human Pol I, which are also discussed.

3.1 Cryo-EM structure of the apo Pol I

In my initial attempts, I aimed to determine the structure of apo human Pol I. Due to the potential ease of handling, I started the sample optimisation without any binding partners.

3.1.1 Sample preparation optimisation

Various conditions have been previously used in the structure determination of the yeast Pol I complexes using single-particle cryo-EM (Table 3.1).

Using conditions that previously gave high-resolution structures as a started point, I screened many buffer compositions and combinations of settings. All grids were prepared using the Vitrobot Mark IV plunge-freezer and screened using the FEI Talos Arctica. During the optimisation process, I tried adjusting:

- grid type (copper, copper with a continuous carbon layer and gold),
- blotting conditions (testing a range of blot force, blot time and wait time),
- protein concentrations (from 0.1 mg/ml to 2 mg/ml),
- buffer composition (the basic buffer contained 15mM HEPES pH 7.5, 150 mM $(NH_4)_2SO_4$, 2mM DTT; further 5mM $MgCl_2$ was added or not; also sample directly taken after elution from the ion-exchange column was used),
- additives (4mM CHAPSO).

Table 3.1. Grid preparation conditions for the previously published yeast Pol I structures from the Müller group. EC – elongation complex, OC – open complex, PIC – pre-initiation complex, Pol I* - Pol I without heterodimer, CC – closed complex. WT – wait time, BT – blot time, BF – blot force.

Structure	**Pol I EC1**	**OC, EC2, EC-tWH**	**Pol I PIC**	**Pol I+GpCpp, Pol I***	**Pol I PIC**
Reference	Tafur, *et al.*, 2016 [56]	Tafur, *et al.*, 2016 [56]	Sadian, *et al.*, 2017 [62]	Tafur, *et al.*, 2019 [60]	Sadian, *et al.*, 2019 [61]
PDB	5M5X	5M5W, 5M5Y, 5M64	5OA1	6HKO, 6HLS	6RQH, 6RUO, 6RQT
Grid type	holey carbon 1.2/1.3 Mo	holey carbon 2/1 copper	holey copper 2/1	holey carbon 2/1 copper	holey carbon 2/1 copper
Buffer composition	150mM $(NH_4)_2SO_4$ 15mM HEPES pH 7.5 10mM DTT	150mM $(NH_4)_2SO_4$ 15mM HEPES pH 7.5 10mM DTT	100mM CH_3CO_2K, 50mM HEPES pH 7.5 10mM DTT 5mM $Mg(CH_3COO)_2$	150mM $(NH_4)_2SO_4$, 15mM HEPES pH 7.5 10mM DTT 1mM $MgCl_2$	100mM CH_3CO_2K 50mM HEPES pH 8 10mM DTT
Blotting conditions	WT 15s, BT 8s	WT 15s, BT 8s	WT 15s, BT 8s	WT 30s, BT 3s, BF 3	WT 30s, BT 3s, BF 2

The best reconstruction I obtained was from holey copper 2/1 grids, on which Pol I sample at 2 mg/ml in 150 mM $(NH_4)_2SO_4$, 15 mM HEPES pH 7.5, 5 mM $MgCl_2$ and 10 mM DTT supplemented with 4 mM CHAPSO was deposited. The grid was incubated for 10 s and blotted for 4 s with blot force "4". While the particles appeared to be well distributed and homogenous, processing revealed issues with the intrinsic dynamics of the complex. I could resolve only the lower lobe, with the upper clamp and stalk showing fuzzy density (Fig. 3.1). Since the clamp's motion is likely continuous, I could not obtain discrete classes of fixed conformations. The Engel group also made a similar observation confirming the intrinsic flexibility of human Pol I [310].

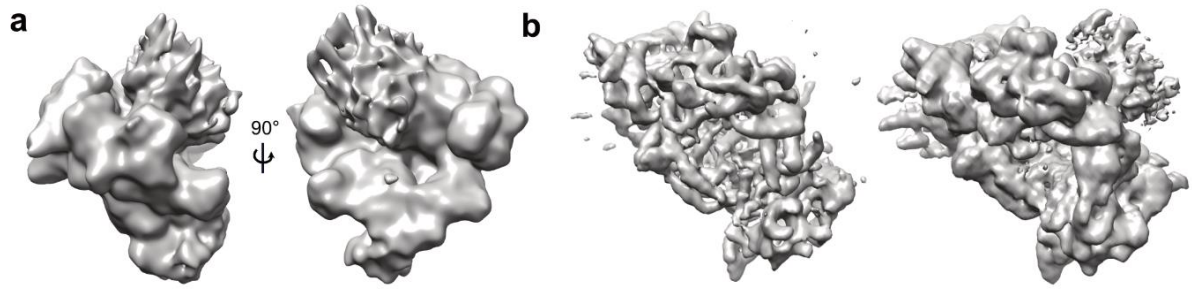

Fig. 3.1. Initial apo Pol I reconstructions. a| Initial reconstruction obtained using cryoSPARC from 86k particles. **b|** Reconstruction was obtained using RELION with 450k particles. Left: higher threshold where secondary structure elements are visible. Right: lower threshold shows blurry clamp density.

3.1.2 Optimised sample preparation and data processing

To surpass the flexibility issue, I aimed to lock the clamp by addition of 3 times molar excess of the DNA-RNA scaffold mimicking the elongation bubble. Using the same buffer composition and grid type, I adjusted the blotting conditions to no wait time, blot time of 2 s and blot force "2". A small dataset of 1580 micrographs collected on FEI Talos Arctica at 92000x magnification showed a promising initial reconstruction (Fig. 3.2 a, upper panel). Thus, I collected a larger dataset of 22 k movies on Titan Krios with a K3 direct detector. I used a magnification of 105000x, corresponding to a pixel size of 0.822 Å/pixel. Data collection details can be found in Table A4 in the Appendix. Using automatic particle picking in WARP with the standard reference, I obtained almost 1M particles. The employed processing pipeline relied on RELION (Fig. 3.2 a). While I tested multiple processing approaches, only the most successful one is outlined here. To increase the processing speed, particles were binned two times and initially split into five subsets. The reconstruction obtained from the Talos Arctica dataset was initially used as a reference. The strategy relied on a series of 3D classifications, with or without masking the less-well resolved regions.

Ultimately, I sorted out two states: with more density in the clamp region (Map A) or more density in the stalk region (Map B). No density that could be assigned to the DNA was present, which explained the observed flexibility of the clamp. The used salt concentration was likely too high for the DNA binding. Thus, I refer to the dataset as "apo Pol I".

The obtained structure reached the average resolution of 4.7 Å for Map A with enriched clamp and 3.7 Å for Map B with enriched density in the stalk region (Fig. 3.2 b, c). For both maps, the resolution in the core approached 3 Å, with the upper clamp having the lowest resolution. While in Map B, distinct structural features can be observed in the stalk, in Map A, only fuzzy, noisy density is present in the clamp region with no defining features.

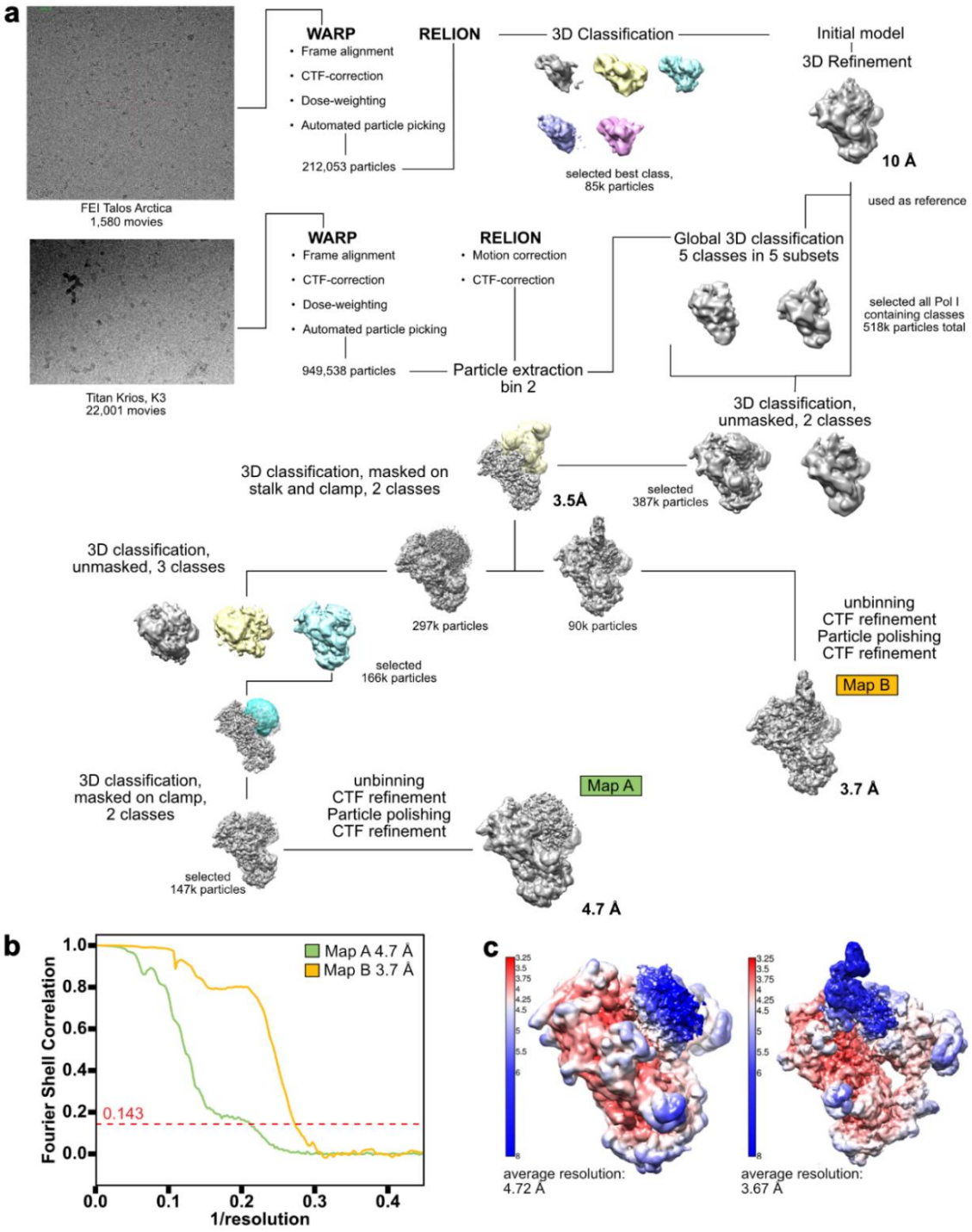

Fig. 3.2. Human apo Pol I processing strategy. a| Main steps of the processing pipeline for the apo Pol I dataset. The initial dataset was collected on FEI Talos Arctica, and a high-resolution dataset was collected on Titan Krios. Transparent surfaces show masks used for 3D classification. The first mask (yellow) was created by fitting the yeast Pol I structure and using it to create a mask over the clamp and stalk. The second mask (cyan) was created by map segmentation to enrich clamp features. Obtained two maps correspond to a density with more signal in the clamp region (Map A) or the stalk region (Map B). **b|** FSC curves for the final average resolution obtained by RELION post-processing (FSC = 0.143). **c|** Local resolution estimation of Maps A and B obtained with the script implementation in RELION.

3.1.3 Structure analysis of apo Pol I

Due to the overall low resolution of the structure, I was unable to build the atomic model of the apo Pol I. Fitting of both yeast Pol I and later the human Pol I EC structure (Fig. 3.3) revealed that the obtained density correlates well with the Pol I structure in the core. All distinct helices and active site elements could be perfectly fitted, including the bridge helix. The heterodimer was also visible in the structure, albeit at a lower resolution.

The most striking features I could observe gave testament to the upper clamp's flexibility. In the case of Map A, where density was enriched in the clamp region, even at the lower threshold, the density did not cover the jaw. Instead, there is a build-up of fuzzy density below the expected stalk location, indicating that perhaps the clamp can open at a wide-angle. In Map B, a defined density likely corresponding to the stalk can be observed. Yet, it is in a different position than the Pol I EC. The stalk is tilted further away from the clamp, covering the RNA exit tunnel.

While due to the poor quality of the EM density, I could not draw any concrete conclusions regarding the structure of the apo Pol I, this dataset allows appreciation of the flexibility of the complex. The clamp moves in a continuous motion that spans wide clamp opening angles. The stalk can also tilt away, occupying a previously unseen location above the RNA exit tunnel. The biological significance of the observed extensive range of motions remains to be investigated.

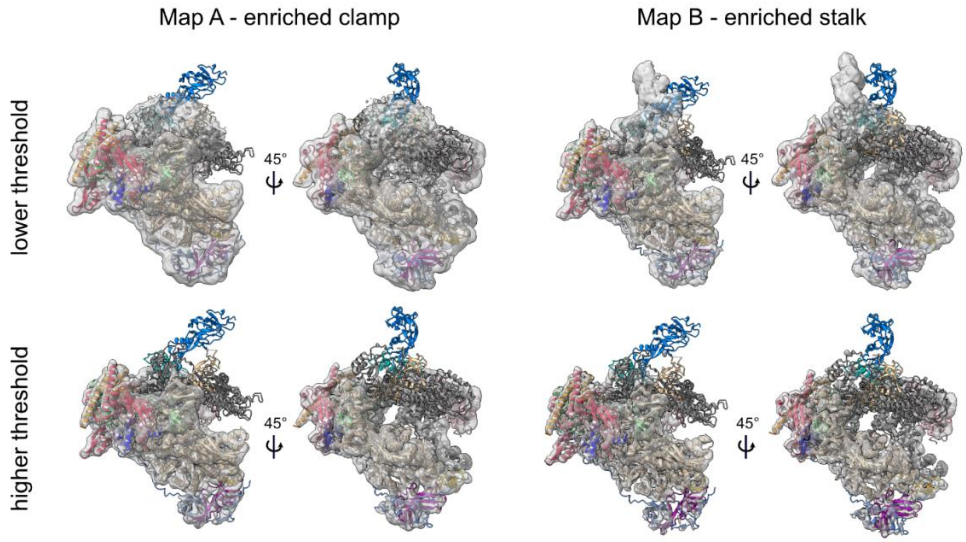

Fig. 3.3. Apo human Pol I fitted with the human Pol I EC structure. PDB 7OB9 [6] was used with nucleic acids not shown. Subunits are coloured as in Fig.1.4. The structure is fitted into Map A (left panel) and Map B (right panel) using the fit_inMap command in ChimeraX. The density is shown at lower and higher thresholds (upper and lower panels, respectively).

3.2 Cryo-EM structure of the Pol I Elongation Complex (EC)

Based on the experience gained while working with the apo Pol I dataset, it became apparent that the addition of a nucleic acid scaffold is necessary to attempt structure determination of the human Pol I. Previous work, focused on yeast Pol I, showed that the clamp contracts upon the addition of the nucleic acid scaffold, especially when it resembles the elongation state [56]. Allowing Pol I to engage its substrate could lock the clamp conformation and allow its averaging. In this section, I show that binding the DNA-RNA bubble resembling the elongation scaffold allows for the structure determination of the human Pol I. Further, I attempt to interpret the observed features and derive their possible biological significance.

3.2.1 Sample preparation and structure determination

In the previous attempts, while the nucleic acid scaffold was present in the cryo-EM sample, it did not bind to Pol I. I suspected this might be due to the high (150 mM $(NH_4)_2SO_4$) salt concentration. Because of the low sample availability, a traditional buffer screen could not be performed. Instead, I selected some potential buffer compositions to use in buffer exchanging of the sample. During purification, CH_3CO_2K is used to maintain the stability of Pol I. Thus, I initially attempted to use CH_3CO_2K for grid preparation. Unfortunately, it led to sample aggregation on the concentrator and further caused issues with sample spreading on the grid. Further, I attempted to lower the salt concentration, while still using $(NH_4)_2SO_4$. The lowest salt concentration at which I could still concentrate the sample without significant aggregation and which would spread on the grid was 70 mM $(NH_4)_2SO_4$. I annealed the nuclei acids by short heating and cooling cycles (details are outlined in section 6.4.1) and added it in 1.5x molar excess to the protein sample at 0.7 mg/mL. The final mixture was spotted on mesh 200, Cu R2/1 grids (Quantifoil) and plunge-frozen using Vitrobot Mark IV with wait time 0 s, blot time 0 s and blot force “3”.

Over 10,000 movies were collected on the Titan Krios TEM at 105,000x magnification. The successful processing pipeline is outlined in Fig. 3.4. WARP was used for particle picking. Initially, I obtained over 2 M particles, which were first cleaned using cryoSPARC 2D classification to allow the building of the initial reconstruction. Further, all particles picked by WRAP were imported into RELION with 4 times binning.

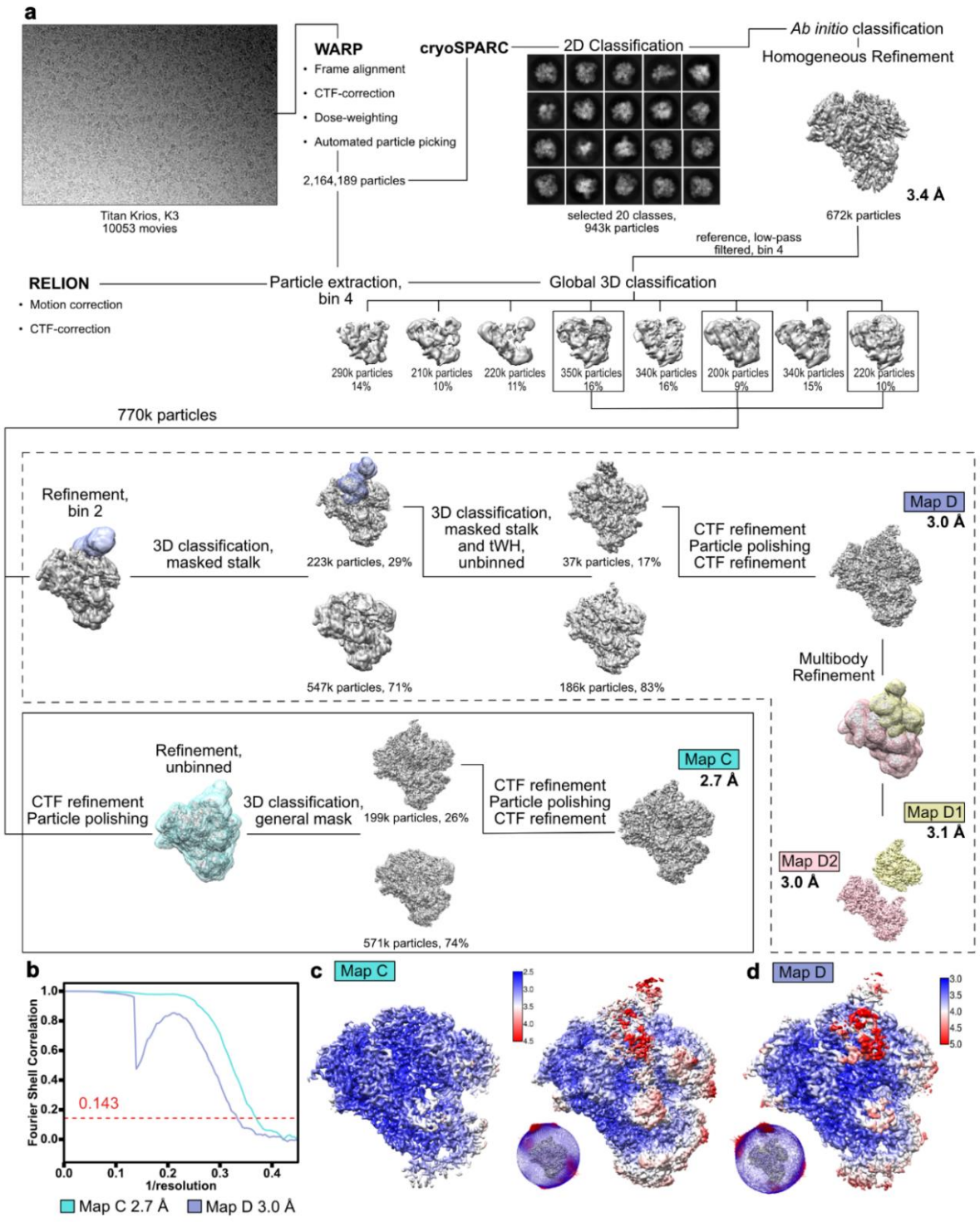

Fig. 3.4. Pol I EC data processing. a| Processing pipeline for the Pol I EC dataset. An example micrograph is shown and subsequent steps are outlined. Transparent, coloured surfaces show used masks. Presented are unsharpened maps with individually adjusted threshold levels (grey). **b|** FSC curve for Maps C and D as obtained from RELION's PostProcessing (FSC = 0.143). **c|** Local resolution estimation as obtained using RELION's implementation. Map C is shown at a higher (left) and lower (right) threshold to allow both the high-resolution core and lower-resolution peripheral elements to be visible. Angular distribution plots are shown at the bottom. Adapted from [23].

Binning not only sped up the processing but also was crucial in the initial classifications to allow focusing on more general features first. Again, the strategy I employed is based on a series of 3D classifications, first without masking and further with masks applied to the poorly resolved regions. Throughout the classification, the particles were successively unbinned, which allowed the focus to be shifted from the general shape of the particles to the high-resolution features. To obtain the highest resolution Map C at a nominal resolution of 2.7 Å, an overall mask covering the whole expected shape of the complex was used. In other cases, precise masks in regions requiring improvement (such as the stalk and further the tWH domain of the RPA49 subunit) were applied to recover more signal. This gave rise to Map D at 3.0 Å resolution. Finally, to still account for the movement of the clamp, Multibody Refinement was used. It allowed me to better define specific features in the upper clamp and stalk regions.

All obtained maps were used for model building. The initial model was prepared based on the yeast Pol I, with homology models for the human subunits used. Dr. Mathias Girbig supported the model building, especially in the stalk, tWH domain of the RPA49 subunit and RNA regions. Especially the core of the complex was very well resolved, allowing me to build all active site elements confidently.

3.2.2 General architecture of human Pol I

The cryo-EM map allowed me to assign the identity of 13 known subunits forming human Pol I (Fig. 3.5). The general shape of the complex is conserved with the yeast Pol I, despite the relatively low sequence identity of all subunits ranging between 20 and 45% for Pol I-specific subunits. The core has a horse-shoe shape with attached on one side the stalk made up of RPA43 subunit and on the other side the heterodimer formed by the RPA34 and RPA49 subunits. The stalk is the site of attachment of initiation factors, including RRN3 [59,61–63]. The TFIIE-TFIIF-like heterodimer plays a role in the initiation and potentially elongation stages of the transcription cycle [75]. The C-terminal tWH domain of the RPA49 subunit is located close to the RNA exit tunnel and the stalk. Thus, the linker of the RPA49 subunit spans across the clamp, close to the downstream DNA [75,125,311]. I could only build the N-terminal domain of the TFIIS-like RPA12 subunit.

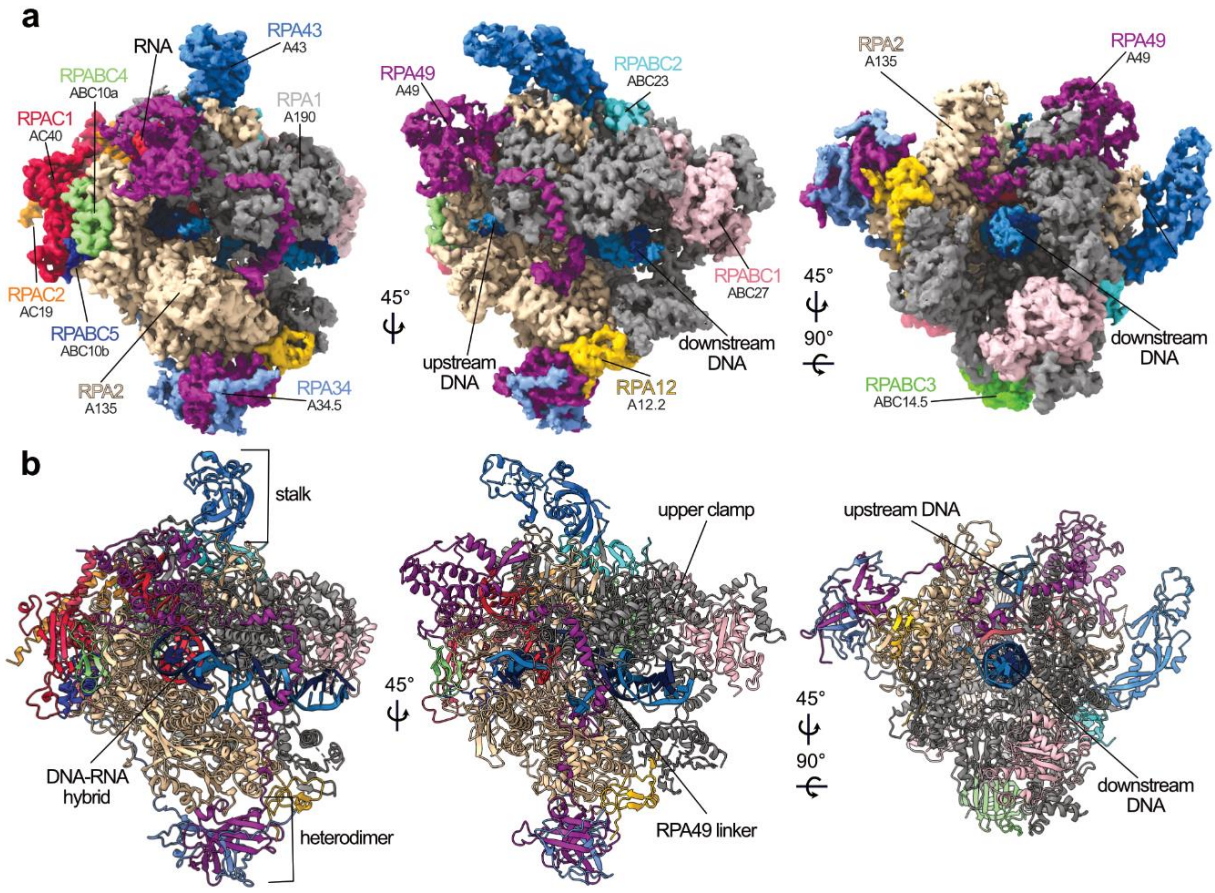

Fig. 3.5. The general architecture of the human Pol I EC. a| Cryo-EM density coloured by the assigned subunits. Yeast subunit terminology is given below. Composite Map E was used for the representation. **b|** Atomic model built into the cryo-EM density. Structural domains are annotated. Adapted from [23].

3.2.3 Nucleic acids interactions

In assembling the nucleic acid scaffold, a 43 bp DNA duplex containing a mismatch over 12 nucleotides was annealed with a 19-nucleotide RNA primer. The density corresponding to the nucleic acids is of very high quality. The active site is the best-resolved part of the structure, with a resolution of 2.6 Å. Thus, I could build a large part of the scaffold and confidently assign residues interacting with the nucleic acids. A majority of the downstream DNA, the whole length of the DNA-RNA hybrid, a significant portion of the RNA protruding from the exit tunnel and part of the upstream DNA were built. Only a portion of the upstream DNA and non-template DNA strand could not be assigned (Fig. 3.6 a).

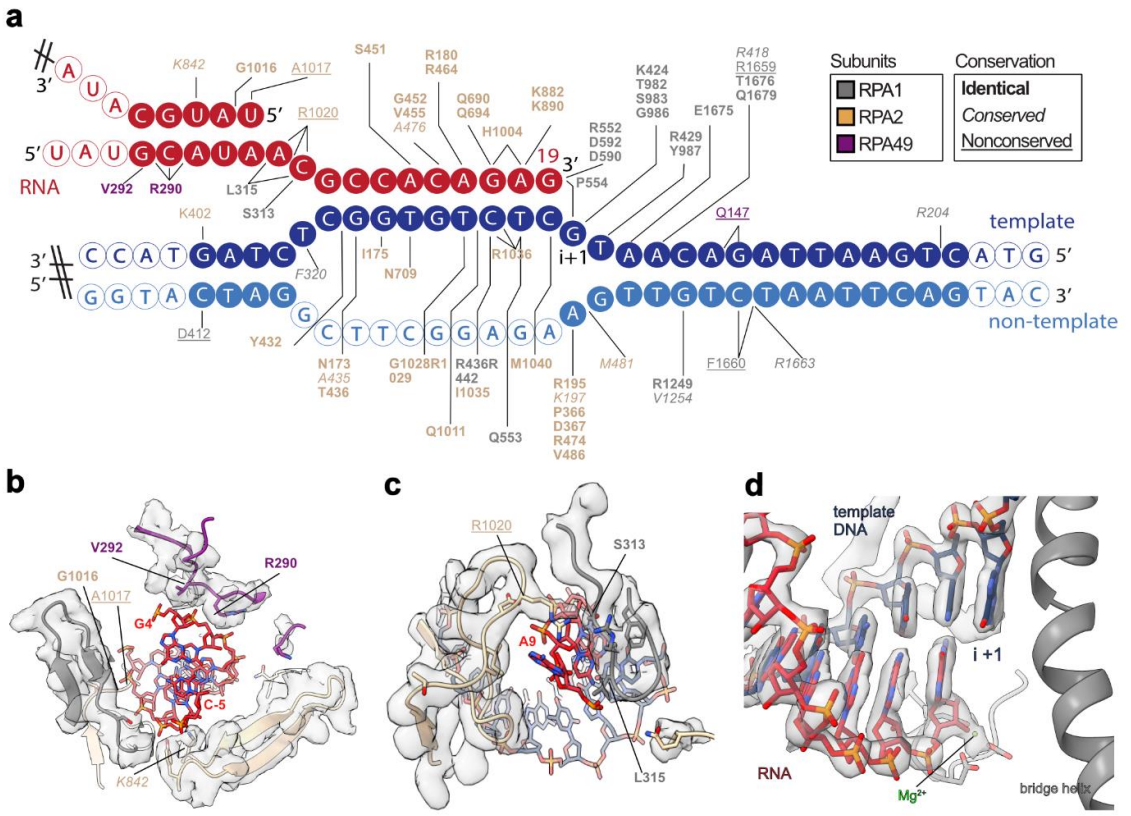

Fig. 3.6. Interactions between the nucleic acids and Pol I. a| Schematic representation of the nucleic acid scaffold with all interacting residues within 5 Å indicated. The filled circles indicate residues built in the structure, and empty circles indicate non-modelled parts of the scaffold. Amino acids are coloured according to their subunit and their conservation to the yeast Pol I is marked. **b|** View at the top of the RNA exit tunnel. **c|** Cross-section view from the top into the RNA separating from the DNA strand. **d|** Active site. The magnesium ion is indicated in green, and the position of the i+1 nucleotide is annotated. **b-d|** Transparent grey surface shows the corresponding cryo-EM density. Adapted from [23].

The two largest subunits, RPA1 and RPA2, make the majority of the nucleic acid contacts. Most residues contacting the nucleic acids are identical between yeast and human. Given that those residues are crucial for Pol I's function, their conservation is not surprising [312–314]. Additionally, the N-terminal tWH domain of the RPA49 subunit is close to the RNA exit tunnel and contacts the RNA. At the same time, the RPA49 linker spans the length of the clamp and contacts the downstream DNA.

At the top of the RNA exit tunnel the passage is wide, easily accommodating two strands of RNA (Fig. 3.6 b). Further, the RNA exit tunnel becomes narrower, with a loop from the RPA2 subunit positioning the RNA separating from the DNA-RNA hybrid (Fig. 3.6 c). The RNA here is contacted by residues R1020 and L315 of RPA2 and RPA1 respectively. The DNA-RNA hybrid is 9 bp long, with the tenth base poised to form weak interactions (it is, however, mismatched in the used template).

The obtained structure is in the post-translocated state (Fig. 3.6 d), with an empty site at the i+1 position awaiting the incoming nucleotide. I could confidently build the magnesium ion coordinating the 3' end of the RNA in the active site. All three catalytic aspartates are also well resolved, including the conformation of the side chains. The second magnesium ion could not be assigned, which is a common observation in all RNAP structures [19,33].

Structures of two more states along the transcription cycle are available for the human Pol I from the Yang and Xu groups [24]. The reported structure in the post-translocated state is identical to the one I obtained, confirming the findings. Additionally, the pre-translocation and backtracked states are also available, expanding our understanding of human Pol I function.

3.2.4 Double-stranded RNA in the RNA exit tunnel

When building the RNA in the exit tunnel, I noticed that the density corresponding to the RNA indicated the presence of more than one chain (Fig. 3.7 a). Looking closely at the RNA sequence, I noticed that the primer is self-complementary at its 5' end. Together with Dr. Mathias Girbig, who helped me assign the density in this region, we debated placing either an RNA hairpin or a double-stranded RNA made up of two separate primers. We decided to build a double-stranded RNA (A-form) (Fig. 3.7 b) based on a perfect RNA helix, as this conformation fitted the density better. The remaining parts of the primers are not visible, which can be attributed to the chain's flexibility outside the exit tunnel. A similar RNA conformation has only been captured in the case of the bacterial RNAP in the paused state [315].

Such a large portion of the RNA visible in the RNAP structure is unique. Even considering only the RNA strand attached to the DNA, I assigned 16 out of 19 nucleotides. In the yeast Pol I structures, only 7 to 13 nucleotides could be traced [60], and in human Pol III only 5 [316] to 6 [36] nucleotides were assigned. Interestingly, I could observe the double-stranded RNA in the exit tunnel, even though no such structures were observed in the human Pol III structure, which was prepared using the same RNA primer. Thus, it can be speculated that Pol I has a unique propensity for stabilising RNA structures in its RNA exit tunnel.

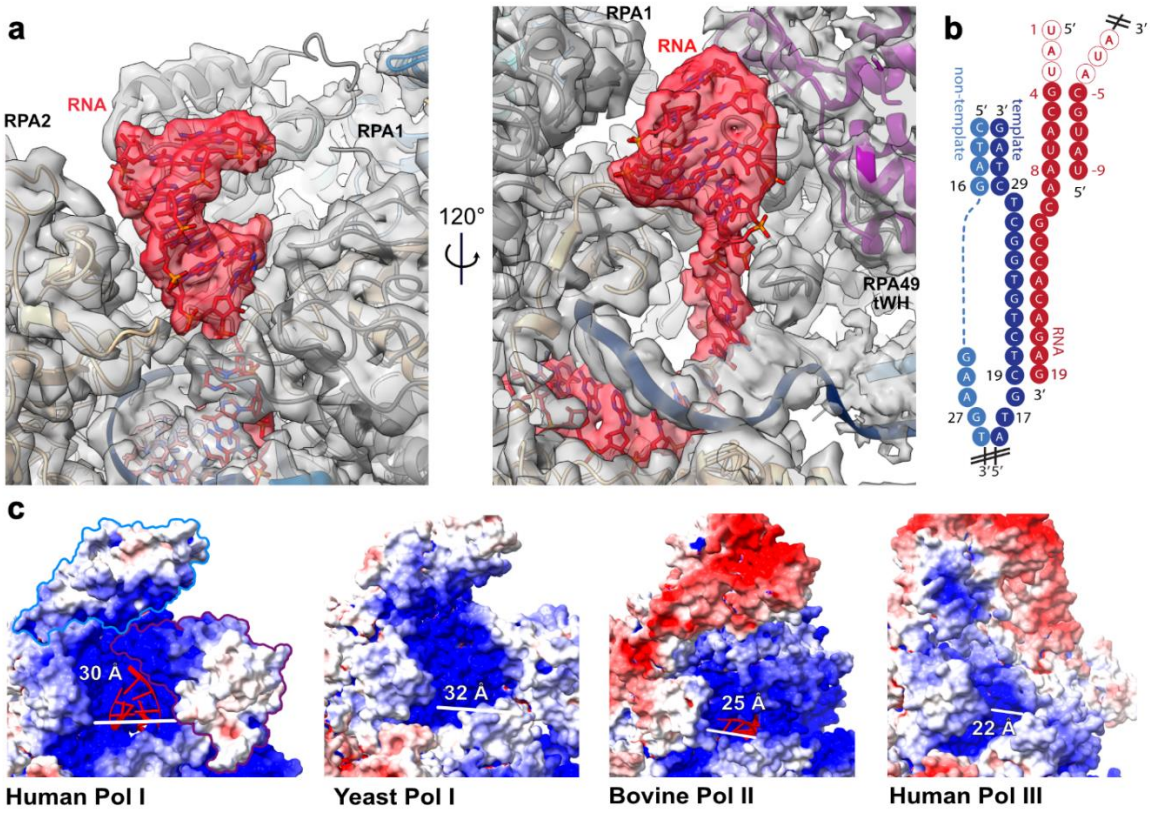

Fig. 3.7. RNA in the RNA exit tunnel. a| Close-up view of the RNA in the RNA exit tunnel. The cryo-EM density corresponding to the RNA is shown as a red, transparent surface. The cryo-EM density corresponding to the surrounding proteins is shown in grey. **b|** Schematic representation of the built nucleic acids. The RNA is double-stranded, with 2 separate primers annealing together due to self-complementarity. **c|** Electrostatic charge distribution over the surface in the vicinity of the RNA exit tunnel. Human Pol I is compared to yeast Pol I (PDB: 5M5X, [56]), bovine Pol II (PDB: 5FLM; [35]) (right top panel) and human Pol III (PDB:7AE1; [36]) (all left to right). The APBS software was used to calculate the charge distribution [317]. RNA in the exit tunnel is shown as a red cartoon. Outlines of the RPA43 stalk subunit (blue) and tWH of the RPA49 subunit (purple) are shown on the left panel. The tunnel width was measured using ChimeraX from backbone to backbone using: for human Pol I residues RPA49 K362 to RPA1 S508, for yeast Pol I residues A49 N354 to A190 S544, for Pol II residues RPB1 K434 to RPB2 Q838 and for Pol III residues RPC1 Y434 to RPC2 A798. Adapted from [23].

To further investigate whether Pol I has any special features that might stabilise the structured RNA, I analysed the RNA exit tunnels in different RNAPs. Pol I has a wider and more positively charged exit tunnel compared to Pol II and III (Fig. 3.7 c). The tWH domain of RPA49, unique to Pol I, extends and shields the RNA exit tunnel. The broad space can accommodate structured RNA and the charge may help it fold.

rRNA is an extensively large and highly structured transcript that must fold in specific conformations to function. While later in the assembly pathway, many

additional factors chaperone rRNA folding [13], it is possible that initial folding steps are facilitated by the Pol I itself. The formation of the RNA structures might positively affect the transcription rate and processivity. Stable RNA structures might prevent RNA's re-entry into the active site and thus backtracking, positively influencing the processivity. Permissive Brownian motion experienced by the folded nucleotide chain, as opposed to the extended, unfolded one may also stimulate the transcription rate. It has been shown that folded RNA structures correlate with faster transcription rates in the case of all RNAPs [69]. This could be especially important in the case of Pol I, which needs to transcribe particularly long genes (13 kb in humans or 6.9 kb in *S. cerevisiae*) [85].

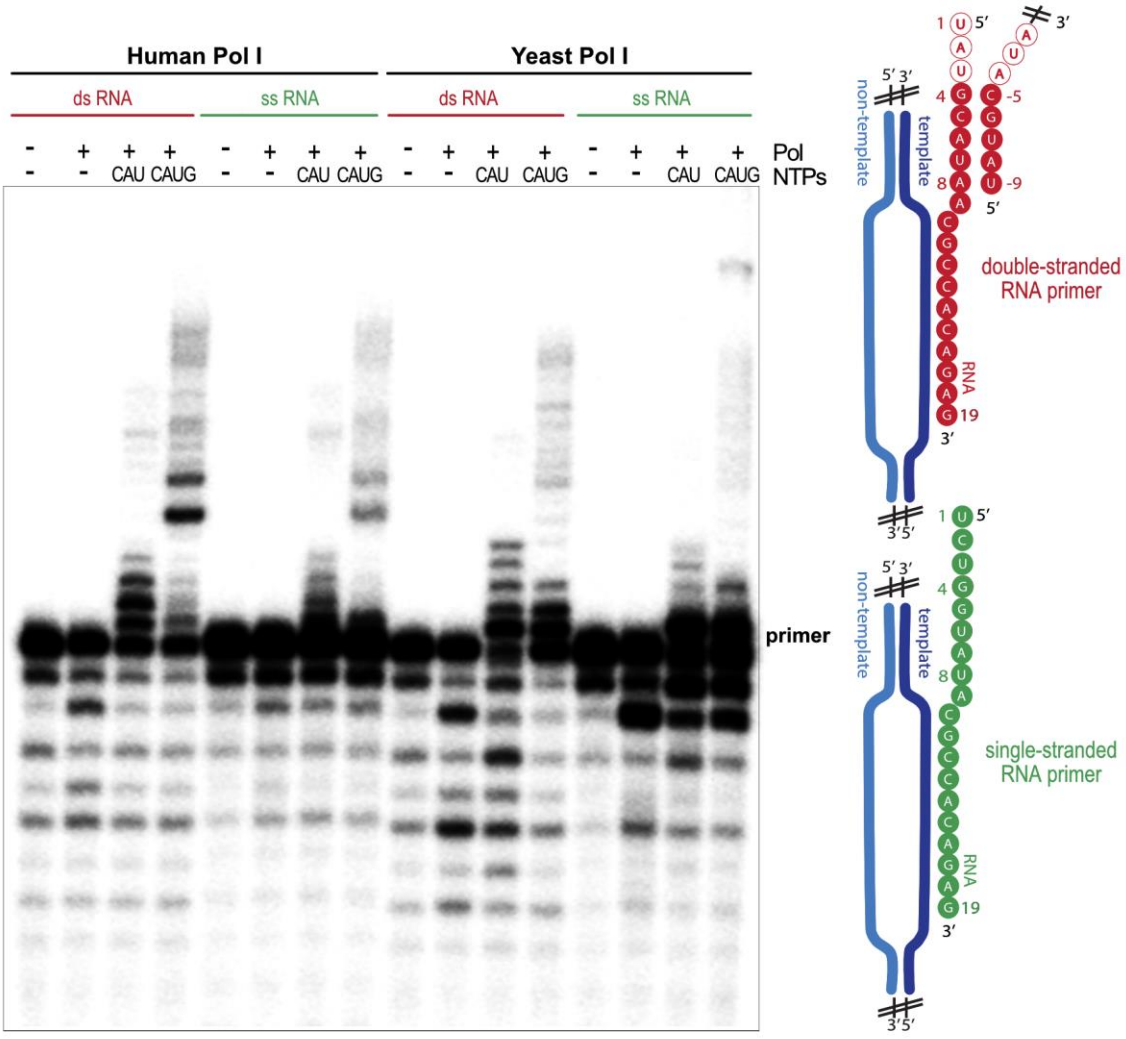

Fig. 3.8. Differences in primer extension when single- or double-stranded RNA is used. Human and yeast Pol I (as indicated) were incubated with scaffolds shown on the right in cartoon representation, which contained either a double-stranded RNA (as used in structure determination) (red) or single-stranded RNA, which is not self-complementary (green). The presence or absence of Pol I and NTPs is indicated at the top. Adapted from [23].

In an *in vitro* primer extension assay performed by Dr. Florence Baudin a difference in transcription rate between single-stranded and double-stranded RNA could be observed (Fig. 3.8). The yeast and human Pol I were supplied with the same artificial DNA-RNA transcription bubble that only differs by the RNA 5' end sequence so that one primer can self-anneal and form double-stranded RNA, while the other remains single-stranded. The double-stranded primer was extended more efficiently in the same length of time. On the other hand, the single-stranded primer was cleaved more often, especially when only one (in the case of human Pol I) or two (in the case of yeast Pol I) nucleotides were removed. Those results support previously reported observations that structured RNA in the RNA exit tunnel supports a faster transcription rate and prevents backtracking.

3.2.5 Pol I stalk is composed of one subunit only

All RNAPs share a domain termed the stalk that serves as a binding platform for some initiation factors. The stalk is located close to the RNA exit tunnel. In Pol II, Pol III and yeast Pol I, the stalk consists of two subunits (Fig. 3.9). The larger yeast Pol I subunit A43 is homologous to Rpb7 in Pol II and C25 in Pol III; the smaller yeast Pol I subunit A14 is homologous to Rpb4 in Pol II and C17 in Pol III [21,202]. The larger stalk subunit is reasonably well conserved between species. It consists of two main domains: the Rpb7-like domain, attaching the subunit to the core, and the C-terminal OB domain (Fig. 3.9 a, b). Non-structured extensions flank structured domains in many species (except, for example, *S. pombe*). In the case of the smaller stalk subunit, no homologue of the yeast A14 subunit could be identified in humans using structure or sequence homology searches. When investigating the cryo-EM density I obtained, it became apparent that the human Pol I stalk is made up of one subunit only. While the cryo-EM density in this region is of worse quality than in the active site, no additional density that could not be assigned to the RPA43 subunit could still be observed (Fig. 3.9 d). On the contrary, in yeast Pol I, clear density corresponding to the α-helices of the A14 subunit is visible (Fig. 3.9 c).

In yeast, the A14 subunit is positioned at the hinge between the Rpb7-like domain and the OB domain (Fig. 3.9 e). In humans, this position is instead taken up by an N-terminal helix of RPA43 (Fig. 3.9 f). A structurally homologous helix is also present in the yeast A43 subunit. Yet, it is tightly bound to the core by A135 and ABC23 extensions, which are absent in the human homologues of those subunits (Fig. 3.9 e, f).

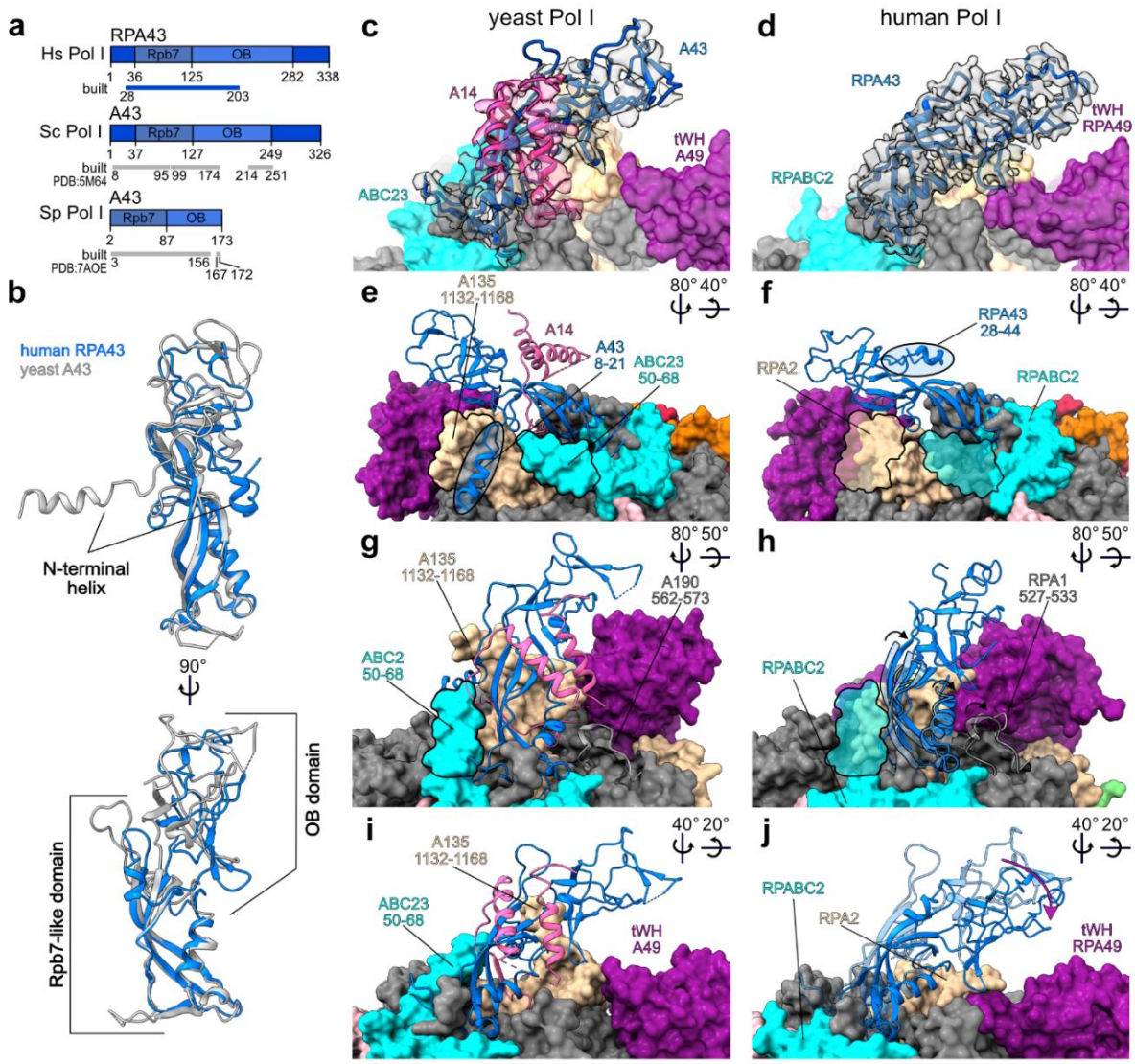

Fig. 3.9 Human stalk is made up of only the RPA43 subunit. a| Domain composition of the larger Pol I stalk subunit in different species. Human RPA43 subunit (PDB: 7OB9 [23]), *S. cerevisiae* A43 (PDB: 5M64 [56]) and *S. pombe* A43 (PDB: 7AOC [64]) from top to bottom. The bar below indicates the areas that are built in each structure. **b|** Structure overlay of human RPA43 and *S. cerevisiae* A43 subunit. Structural domains are shown. **c,e,g,i|** *S. cerevisiae* stalk. **d,f,h,j|** Human stalk. **c-j|** Core subunits are shown in surface representation, coloured as in Fig. 3.5. Stalk subunits are shown in cartoon representation. Discussed extensions and structural features are labelled. Arrows in the upper right corners of all panels show view rotations. **f,h|** Transparent outlines correspond to yeast extensions absent in human structures as labelled in **e,g** respectively. **h,j|** Transparent cartoons show yeast A43 features from **g,i** respectively to indicate the change in position to the RPA43 subunit as indicated by arrows. Adapted from [23].

The absence of additional extensions that tightly bind the stalk to the core means that the human stalk can be much more flexibly positioned. The whole stalk leans towards the tWH domain of the RPA49 subunit, which is further allowed by the RPA1 527-533 loop. In yeast, this loop is highly structured, supporting the stalk positioning from the opposite to the extensions site (Fig. 3.9 g). In humans, the loop is flexible and bends together with the stalk (Fig. 3.9 h). As a result, the whole human stalk leans forward towards the tWH domain of the RPA49 subunit and above the RNA exit tunnel (Fig. 3.9 i,j). The large flexibility of the human Pol I stalk is consistent with the observations I have made using the apo Pol I structure (section 3.1.2).

Further, I was curious to find out the prevalent subunit composition for Pol I and whether yeast or human structure is the outlier. Dr. Aleix Lafita performed a phylogenomic analysis that looked at the presence or absence of homologues of Pol I subunits A43 and A14 and Pol II-specific RPB4 subunit across the tree of life. While the RPB4 subunit, the Pol II ortholog of the A14 subunit, is conserved throughout the tree of life, homologs of A14 are only found in some species from a division of Fungi, Ascomycota (Fig. 3.10 a). This finding suggests that yeast Pol I is the outlier and in most species Pol I is composed of 13 subunits.

To further investigate if the emerging pattern correlates with the presence or absence of extensions that bind the stalk in yeast, Dr. Aleix Lafita further extended the phylogenomic analysis to include those subunits. It became apparent that the insertion within the A135 subunit is also specific to Fungi, albeit its length differs (Fig. 3.10 a, b). The insertion in the N-terminus of the ABC23 subunit is less conserved and appears in more species (Fig. 3.10 a). Further, using sequence alignments, I looked at the agreement with the consensus sequence and ungapped occupancy around the insert of interest (Fig. 3.10 c, d). Only a few percent of sequences agree with the consensus sequence (Fig. 3.10 c) and many gaps in residue occupancy are detected (Fig. 3.10 d). Thus, further investigation is needed into the evolutionary relationship between the ABC23 subunit extension and stalk subunits.

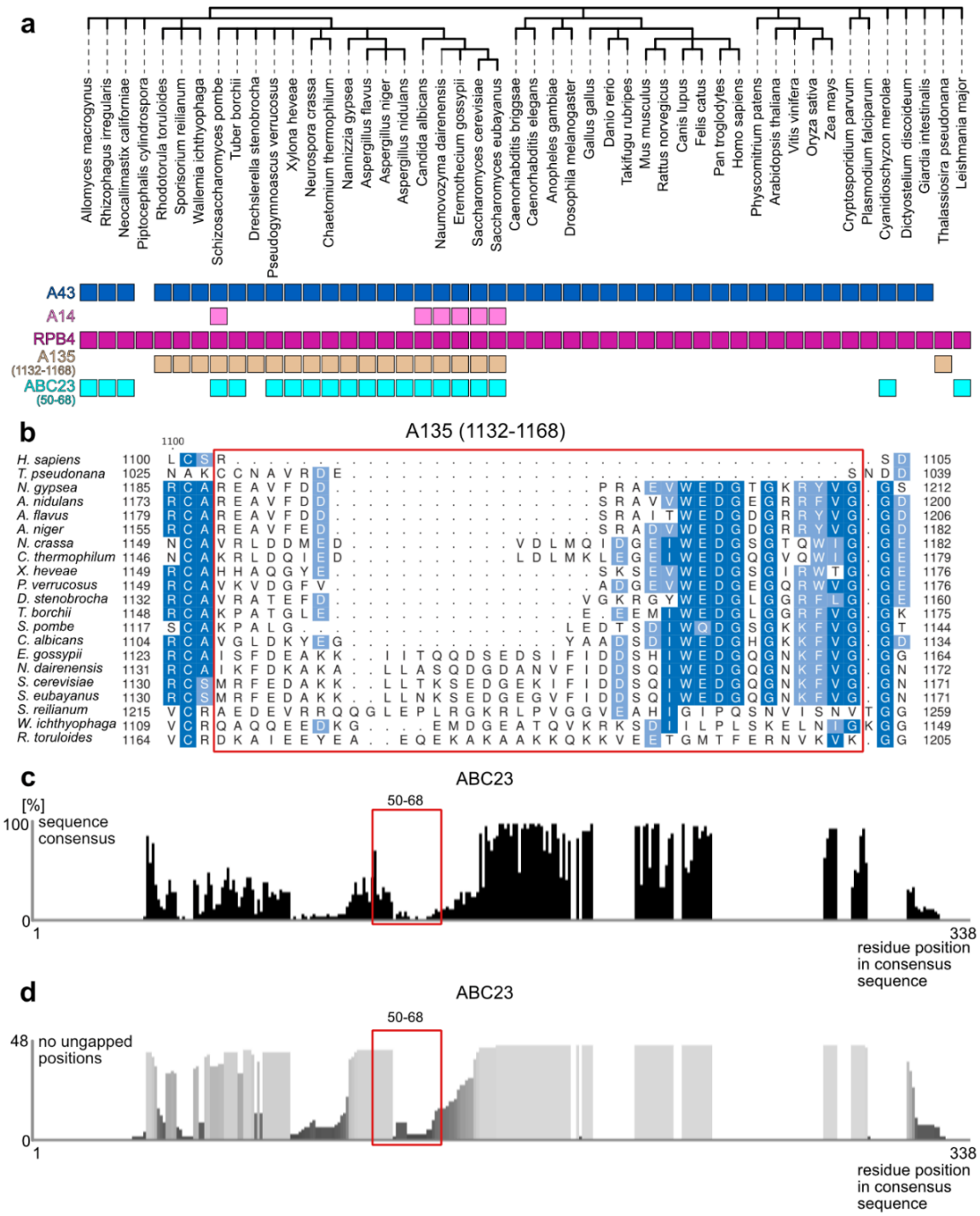

Fig. 3.10 Phylogenomic analysis. **a|** Tree of life of selected species with presence (indicated by a filled square) of homologues of the Pol I A43 and A14 subunits, Pol II RPB4 subunit and discussed extensions in A135 and ABC23 subunits. **b|** Sequence alignment of the A135 extension. Sequences in species where the presence of the extension was identified and the sequence of human RPA2 on top are shown. **c|** The ABC23 subunit sequence with the sequence consensus score plotted (percentage of sequences which contain the consensus residue). **d|** The ABC23 subunit occupancy plot. The number of ungapped residues (a total of 48 sequences were investigated) in each position over the consensus alignment. **b-d|** Red box indicates the location of the extension of interest. Taken from [23].

3.2.6 Structure of the RPA34-RPA49 heterodimer

The Pol I-specific subcomplex consisting of RPA34 and RPA49 subunits is termed the heterodimer. It consists of the dimerization domain opposite the stalk and long extensions attaching it to the core [21,75,125]. The heterodimer is functionally related to the Pol II general transcription factors: TFIIF (considering the dimerisation domain) and TFIIE (when accounting for the tWH domain of the RPA49 subunit) [61,62,75,125,318]. While the sequence conservation between yeast and humans is particularly low (20% for RPA49 and 25% for RPA34), their overall architecture and function remain unchanged.

The RPA49 subunit is anchored via the dimerisation domain to the RPA34 subunit. It then extends on the surface of the core with a long linker that positions the C-terminal tWH domain next to the RNA exit tunnel and stalk (Fig. 3.11 b). After thorough 3D classification using appropriate masks, I recovered enough cryo-EM density to build almost the entirety of the subunit (Fig. 3.11 a), which is unique compared to the other available structures of the human Pol I [24]. Previously, the C-terminal tWH domain of the RPA49 subunit was resolved only in some of the yeast Pol I structures [56]. While the general position of the tWH domain is consistent between yeast and human structures (Fig. 3.10 c), the domain extends closer to the upstream DNA in yeast. The position of the tWH domain changes depending on the state in the transcription cycle [319]. Thus, a conformation of the human RPA49 subunit even more similar to yeast might still be observed.

The RPA49 linker is tightly bound to the clamp. It wraps around the clamp coil-coils (Fig. 3.11 d) and assists in clamp closure upon DNA binding. Partially disordered loop connecting the two coil-coils changes the conformation when the RPA49 linker binds. In the structure of human Pol I bound to RRN3, the loop adopts a conformation that could clash with the RPA49 linker if it was bound.

Both the tWH domain and a motif within the RPA49 linker termed the helix-turn-helix motif (HTH) have DNA binding capability [61,114]. The HTH motif is located close to the downstream DNA (Fig. 3.11 e) and is functionally conserved across different species, regardless of only partial sequence conservation (Fig. 3.11 f). *In vitro* mutational studies show that both helices within the HTH motif can bind DNA [114]. In the current human Pol I structure, only helix 2 is poised for the DNA interaction, while helix 1 interacts with the RPA2 lobe (Fig. 3.11 e right panel).

Yet, since the conformation of the tWH domain has been observed to change, it is also possible that the conformation of the HTH motif can be altered. While the HTH motif's sequence and position might change, its presence is required for cell proliferation [114].

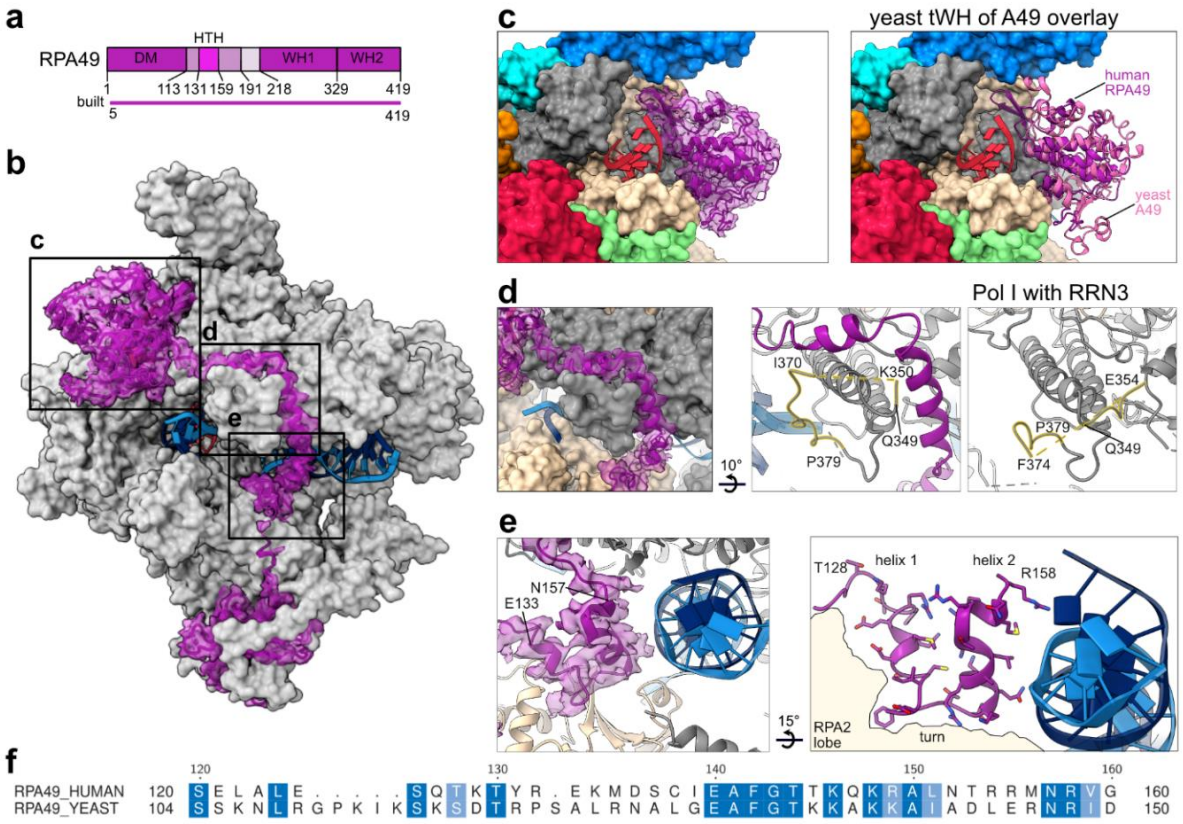

Fig. 3.11. The RPA49 subunit: linker spanning the clamp and the tWH domain. **a|** Schematic representation of subunit's domains. The purple bar below indicates regions included in the structure. DM – dimerisation domain, HTH – helix-turn-helix motif, WH – winged-helix domain. **b|** Structure overview with the RPA49 subunit marked in purple. The nucleic acids scaffold is shown in cartoon representation. **c|** Position of the tWH domain (purple) shown as cartoon fitted into the transparent cryo-EM density. Right panel: yeast tWH domain (PDB: 5M64 [56]) (pink) overlaid on the human structure. **d|** The linker wraps around the clamp coil-coils. Middle and right panels: coil-coils are connected by a linker (golden highlight) that changes conformation in the absence of the RPA49 linker (from Pol I-RRN3 structure). **e|** Helix-turn-helix (HTH) motif positioned close to the downstream DNA (blue). RPA49 subunit shown in cartoon representation and fitted into the cryo-EM density. Right panel: wheat outline shows the position of the RPA2 subunit. **f|** Sequence alignment of the HTH motif between humans and yeast. Identical residues are coloured in dark blue, similar residues in light blue and dots indicate gaps in the sequence. Taken from [23].

The second subunit making up the heterodimer is the RPA34 subunit. It associates with the RPA49 subunit via the dimerisation domain. The C-terminus is made up of a large, unstructured extension twice as large in humans compared to yeast (Fig. 3.12 a). Only the 40 first residues of these extensions running along the core made up by the RPA2 subunit, and further anchoring close to the RPAC1 and RPAC2 subunits could be assigned (Fig. 3.12 c) [21,54]. While the extension is similarly positioned in both species, its sequence is highly divergent (Fig. 3.12 b). The human RPA34 extension is proline-rich. It forms a rigid chain that perfectly fits with the residues coming from the core. The poly-proline stretch is, on one side, highly hydrophobic, and on the other side, has many sites accessible for hydrogen bonding (Fig. 3.12 c). Together, this creates a favourable area for protein-protein interactions [320]. In yeast, the interaction is based on charged residues (Fig. 3.12 d), which can form hydrogen bonds with the core. Regardless of the sequence details, both extensions tightly anchor the heterodimer to the core.

The role of the C-terminal disordered extension of the RPA34 subunit is not well understood. It is possible that the C-terminal extension would become ordered upon interaction with specific transcription factors. It was suggested to interact with UBF [79] and possibly SL1 based on *in vivo* studies [321]. In the disordered state, its role may be modulated through post-translational modifications. The C-terminal extension was shown to be a target of acetylation [322], which allows it to diffuse out of the nucleolus upon starvation [321]. The extension might also have a role in Pol I assembly as it contains a nucleolar localisation signal [323]. Further studies, using various methods, are needed to uncover its role fully.

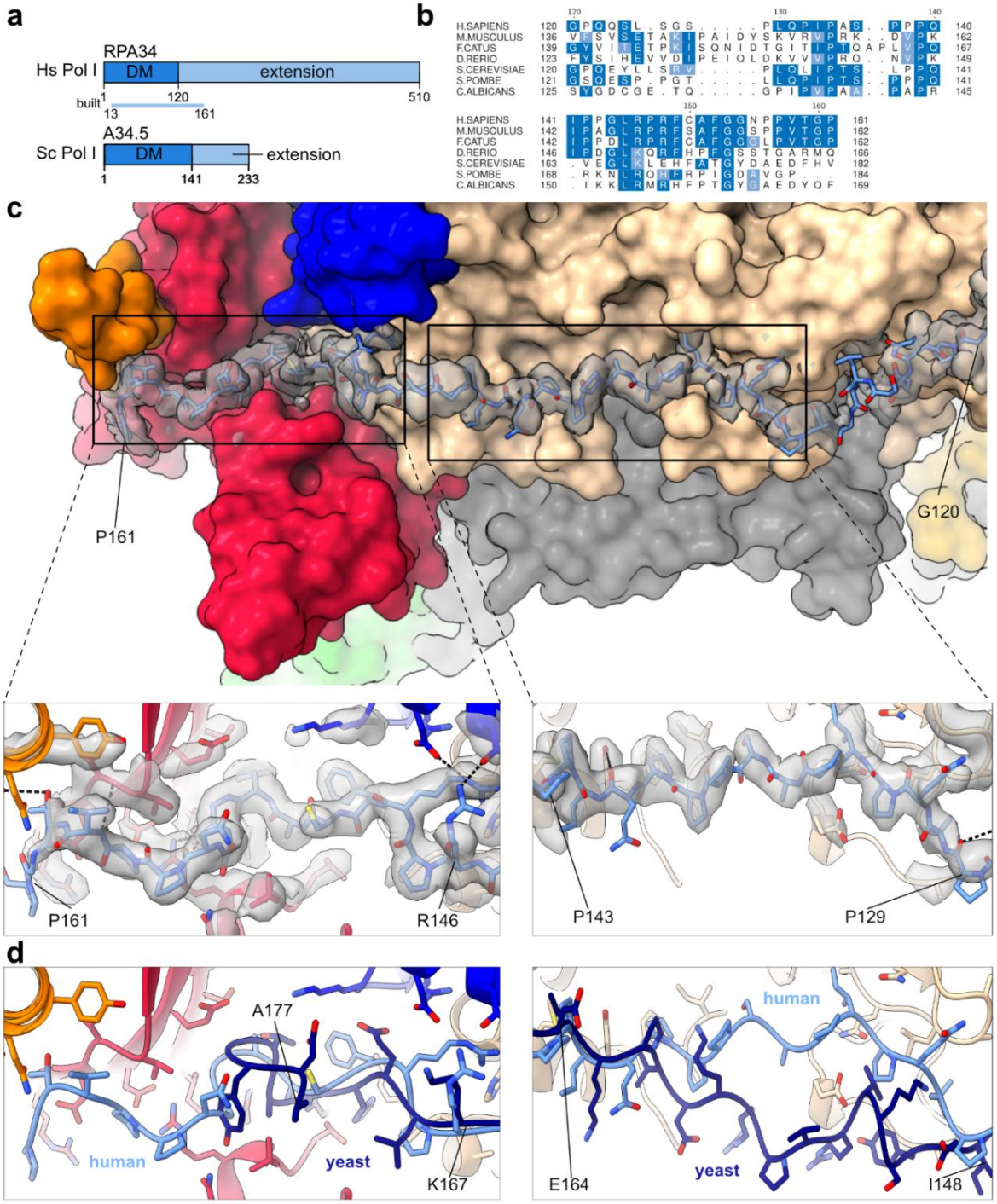

Fig. 3.12. The RPA34 extension. a| Schematic representation of the RPA34 subunit with domains annotated. The region modelled in the structure is denoted with a blue bar. DM – dimerisation domain. **b|** Sequence alignment of the modelled portion of the extension. Identical residues are coloured in dark blue, similar residues in light blue and dots indicate gaps in the sequence. **c|** Path of the modelled RPA34 extension (stick representation with cryo-EM density shown in transparent grey) on the Pol I core (surface representation). Black dotted lines indicate hydrogen bonds, and solid grey lines show contacts. **d|** Comparison between human (light blue) and yeast PBD: 5M64 [56] (dark blue) extension. **c, d|** In the zoom-in, residues within 5Å from the extension are shown in stick representation. Taken from [23].

3.3 Structure determination of human Pol I-RRN3 and Pol I OC

To gather more information about human Pol I in various functional states, I decided to study Pol I in complex with the initiation factor RRN3. The addition of a co-factor binding next to the stalk was interesting, as it allowed me to gather more information about this area of the Pol I. It is debated whether Pol I binds RRN3 first before engaging with the open DNA or if Pol I first binds the DNA and then associates with RRN3. Regardless, RRN3 should still be stably associated with Pol I bound to the open DNA template. Thus, I decided to include in the sample an open DNA scaffold devoid of RNA that would mimic the initial steps in Pol I transcription. Ultimately, I could sort out two states: one where Pol I was bound to RRN3 and another where it was engaging with the open DNA forming OC, which are discussed in detail in this section.

3.3.1 Sample preparation and data processing

The sample preparation was highly similar as for the Pol I EC sample. Pol I at a concentration of 0.85 mg/ml was pre-incubated with RRN3 expressed and purified by Dr. Brice Murciano (as outlined in section 2.3.3) mixed in equimolar ratios. After 20 min, 1.5x molar excess of the open DNA scaffold was added and further incubated for 30 min. The sample was plunge-frozen in liquid ethane using the same parameters as for the Pol I EC sample.

Firstly, I screened obtained grids on the FEI Talos Arctica microscope and collected a small dataset that allowed me to obtain a first Pol I-RRN3 reconstruction that was further used as a reference (Fig. 3.13 a, upper panel). Then, I collected a larger dataset of over 14 k movies using the Titan Krios microscope. The general processing strategy I used in the case of this dataset is very similar to the Pol I EC dataset. The main difference was to divide the dataset into five subsets to speed up the initial processing further. Selected particles were pooled together and classified using appropriate masks (Fig. 3.13 a). Overall, the achieved resolution for the Pol I-RRN3 and Pol I OC datasets is lower than the Pol I EC (Fig. 3.13 b, c). For Pol I-RRN3 structure, the resolution reaches 3.2 Å and can be improved up to 3.1 Å, while an overall resolution of 3.3 Å is achieved for the Pol I OC structure.

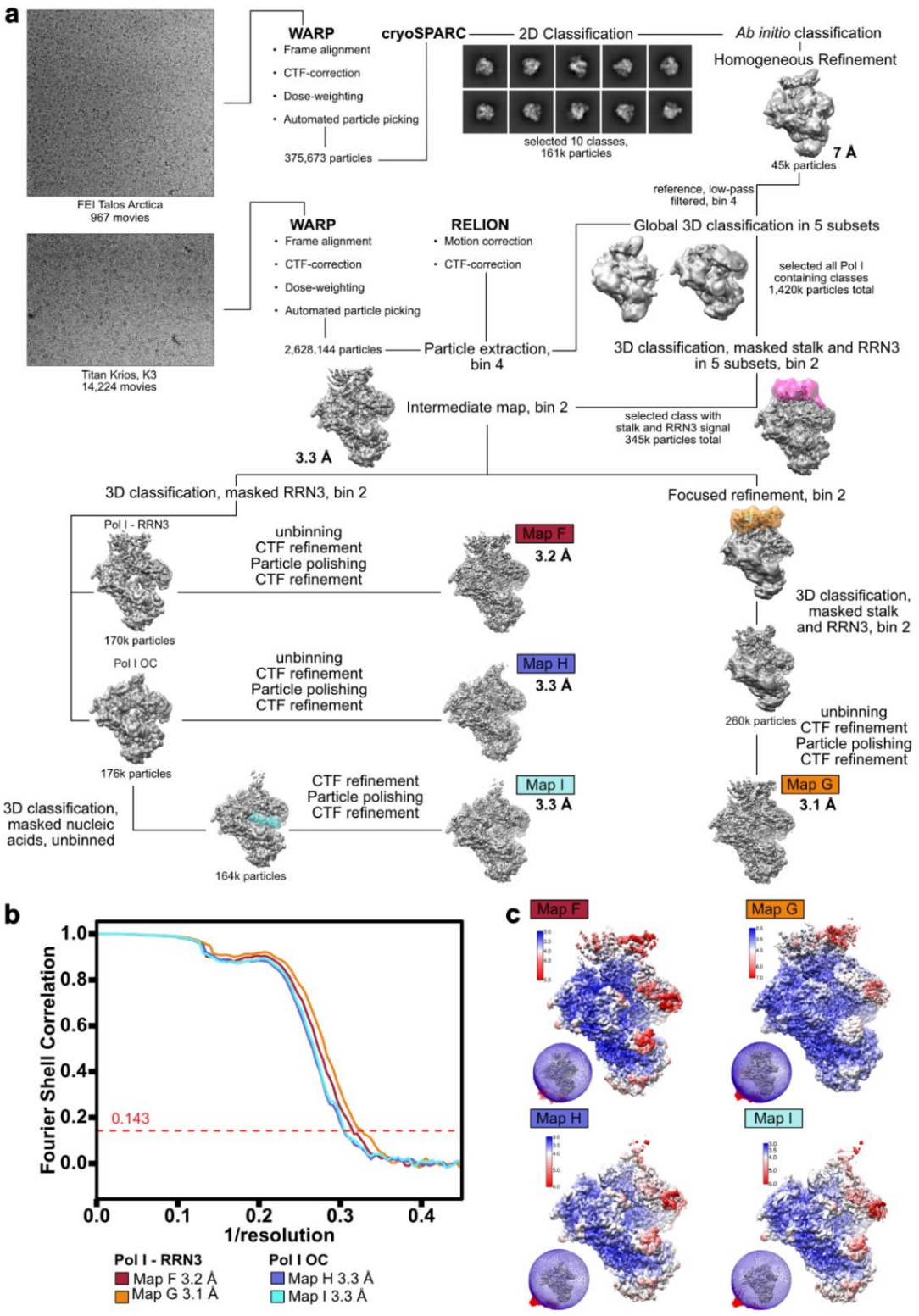

Fig. 3.13. Pol I-RRN3 and Pol I OC data processing. a| Processing pipeline for the Pol I-RRN3 and Pol I OC dataset. An example micrograph is shown and subsequent steps are outlined. Transparent, coloured surfaces show used masks. Presented are unsharpened maps with individually adjusted threshold levels (grey). **b|** FSC curve for Maps F - I as obtained from RELION's PostProcessing (FSC = 0.143). **c|** Local resolution estimation as obtained using RELION's implementation. Angular distribution plots are shown at the bottom. Taken from [23].

Additionally, issues with anisotropy of the reconstruction coming from the inherent dynamics within the sample appeared, especially in the case of the Pol I-RRN3 structure (Fig. 3.14 a). I performed focused refinement to counteract those issues, which improved the density around RRN3. The improvement of the map could be measured using a 3D FSC implementation [324]. When the sample is entirely isotropic, all FSC curves from different angles are the same, giving rise to a perfectly spherical reconstruction of the FSC. If any issues occur, the value of the sphericity drops (Fig. 3.14). Focused refinement improved the overall resolution and the resolution in the best and worst orientations, partially counteracting anisotropy issues.

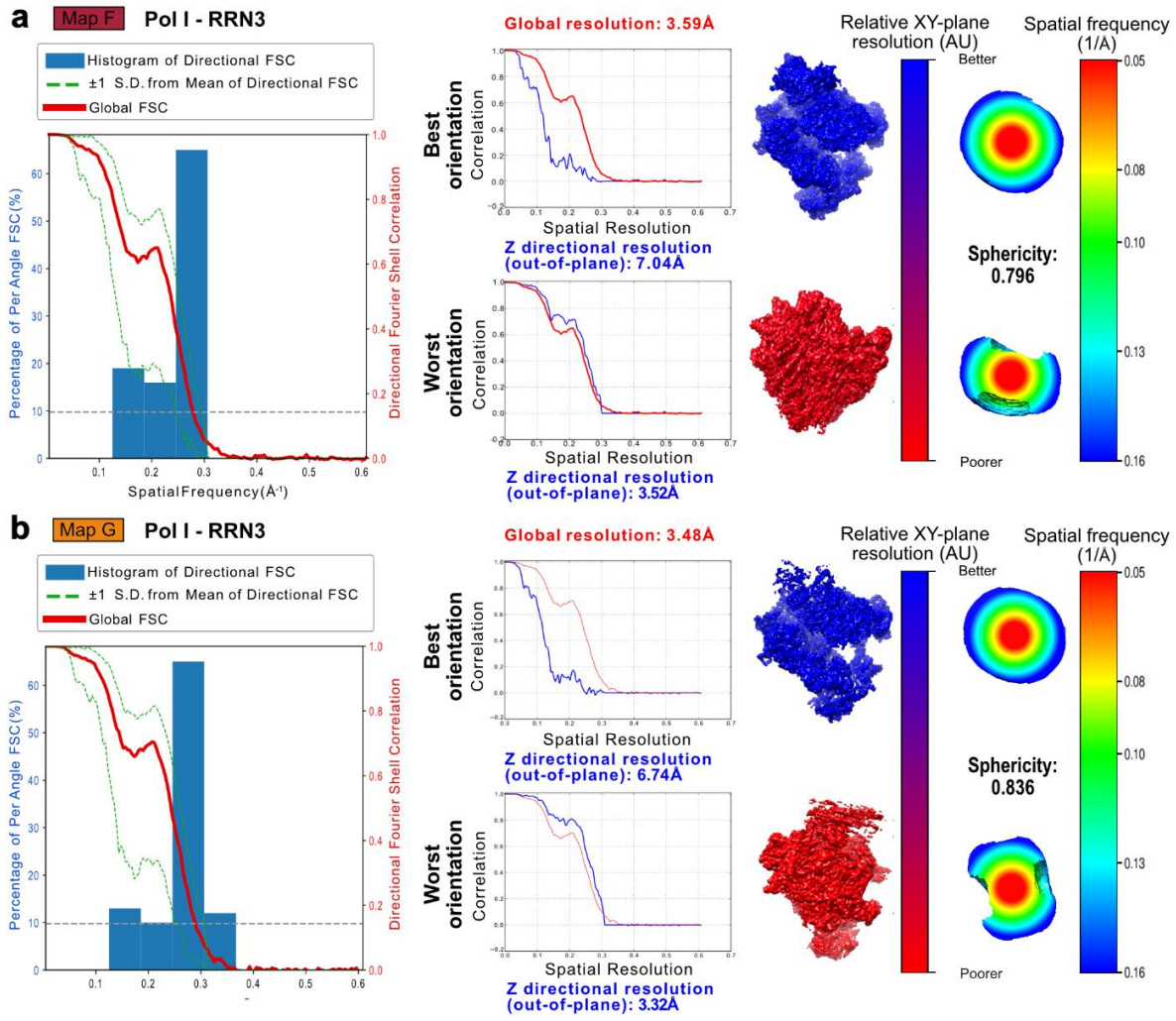

Fig. 3.14. Improvement of Pol I – RRN3 map isotropy with focused refinement measured using 3D FSC. a| Map F: Pol I-RRN3 **b|** Map G: Pol I-RRN3 after focused refinement. **a,b|** Left to right: Distribution of the directional FSCs plotted as a histogram with a global FSC curve. FSC in the best (top) and worst (bottom) XY orientation and the sphericity measure in those slices. Adapted from [23].

3.3.2 Differences between Pol I EC, Pol I OC and Pol I-RRN3

The three obtained Pol I structures allowed me to observe Pol I changes throughout the transcription cycle. The obvious differences include different nucleic acid scaffolds present in the structures, and binding of the initiation factor in the Pol I-RRN3 structure (Fig. 3.15). Further changes are associated with different stages in the transcription cycle. They include ordering the C-terminal domain of the RPA12 subunit and its insertion into the active site in the Pol I OC and Pol I-RRN3 structures (Fig. 3.15). On the contrary, the tWH domain of the RPA49 subunit is only ordered in the Pol I EC structure. Additionally, in the Pol I-RRN3 structure, a density corresponding to the DNA mimicking loop can be observed (discussed in section 3.4.3).

Depending on the stage in the transcription cycle, the DNA cleft's width changes (Fig. 3.16 a). The clamp is closed in the Pol I EC, tightly binding the nucleic acid scaffold. In the Pol I OC and Pol I-RRN3 structures, it widens by ~10 Å. This opening can also be associated with the insertion of the C-terminal domain of the RPA12 subunit and the displacement of the tWH domain of the RPA49 subunit – the linker does not cross the clamp helping in its closing [61,211,325]. In yeast, the transition between the Pol I OC and Pol I EC state is associated with narrowing the cleft by 7 Å [56], indicating that in human Pol I the clamp is more flexible.

Several changes to the active site elements are concomitant with the different stages in the Pol I transcription cycle (Fig. 3.16 b). Overall, most active site elements are ordered and positioned in the active conformation in the Pol I EC structure and disordered or folded away in Pol I OC and Pol I-RRN3 structures. The difference is particularly striking in the bridge helix and trigger loop case. The bridge helix spans across the cleft, so it must become disordered with the widening of the cleft to give more space. The trigger loop is displaced by the insertion of the RPA12 C-terminal domain, as further discussed in section 3.5.1. In the case of the lid loop and the hybrid binding region, they are fully ordered in the Pol I EC and partially ordered in the Pol I OC. It is possible to speculate that binding of nucleic acids, even in an inactive conformation as in the Pol I OC structure, helps in the positioning of certain active site elements that are otherwise disordered when the DNA-mimicking loop is inserted, as seen in the case of the Pol I-RRN3 structure. Having structures of various conformations within the transcription cycle allows for studying the transitions necessary for the correct function of the complex.

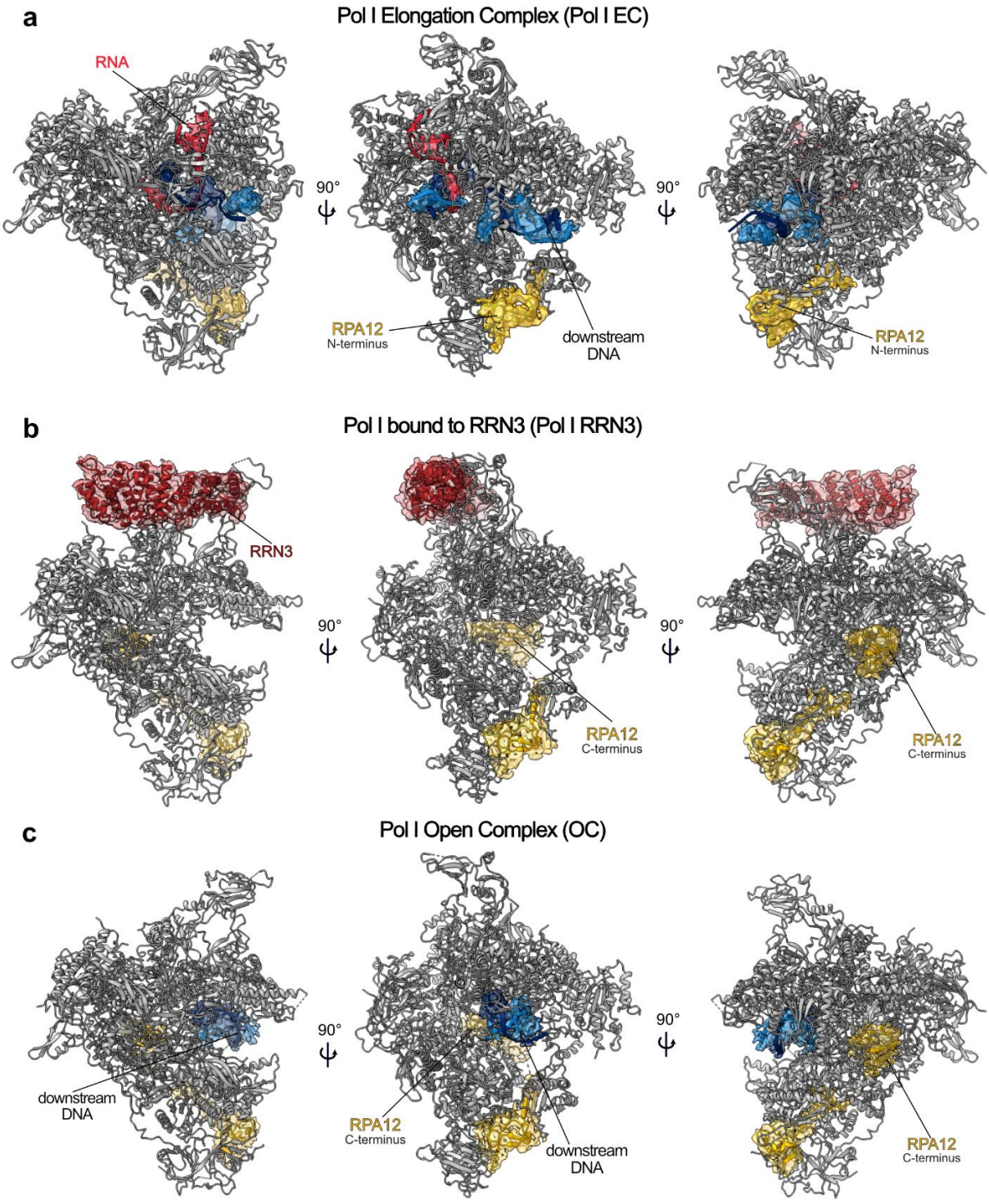

Fig. 3.15. Overview of differences between obtained human Pol I structures. a| Pol I EC. **b|** Pol I-RRN3. **c|** Pol I OC. All structures are shown in cartoon representation (grey) with nucleic acids coloured (DNA – blue, RNA – red). RPA12 subunit is shown in yellow and RRN3 in maroon. Transparent surfaces correspond to the cryo-EM density. Panels **b** and **c** are adapted from [23].

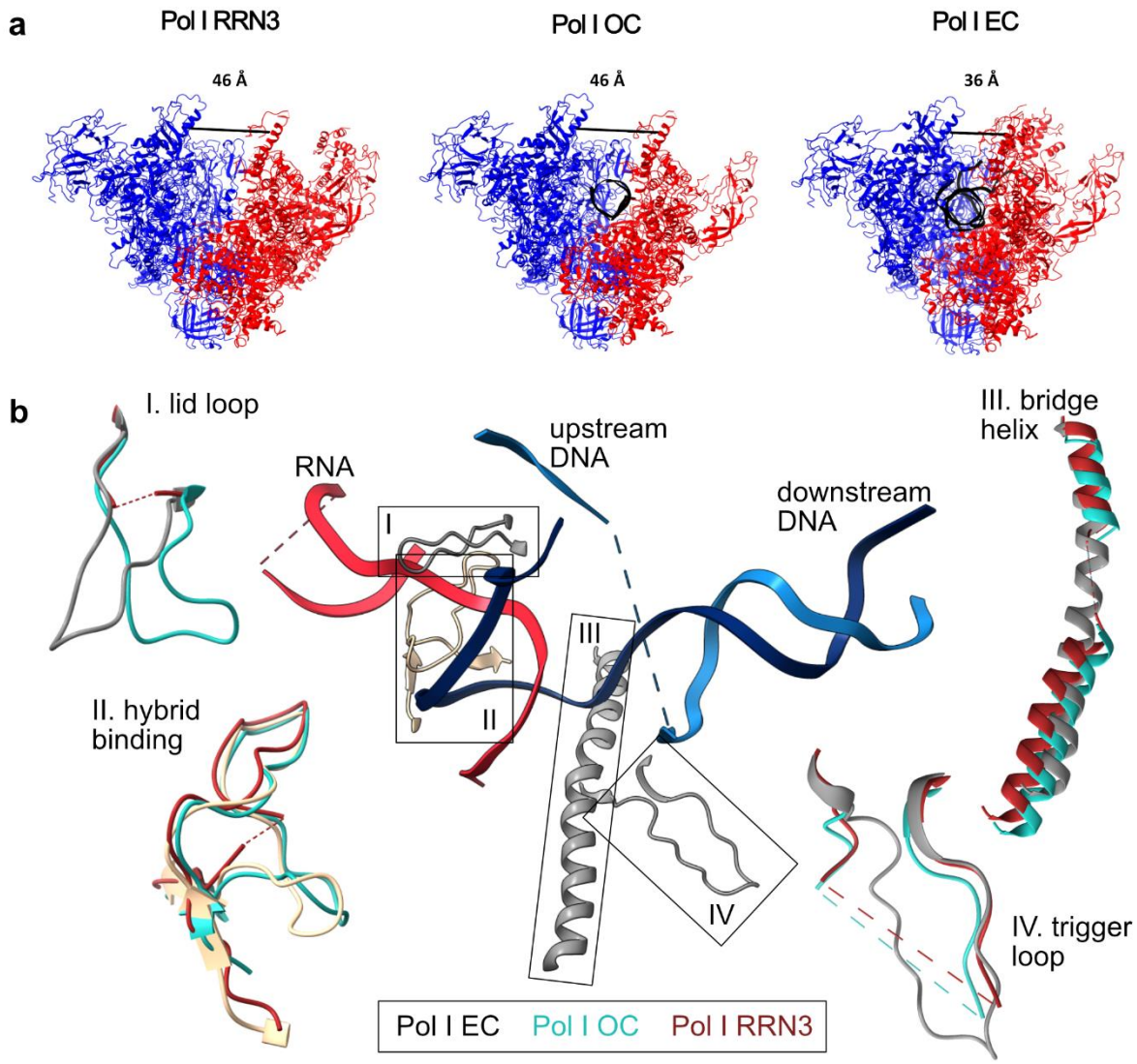

Fig. 3.16. Differences between Pol I structures. a| Width of the clamp opening is shown as a black line, measured from L321 of the RPA1 subunit (clamp core) to I396 of the RPA2 subunit (protrusion). The two mobile domains are shown in red and blue; nucleic acids are marked in black. **b|** Comparison of the active site elements. The Pol I EC structure is shown in the overview with all three structures overlaid in the zoom-in of the active site elements. Pol I EC is coloured according to Fig. 3.5, Pol I OC is shown in cyan and Pol I-RRN3 is in maroon. Panel **a** is taken from [23].

3.4 Cryo-EM structure of Pol I bound to the initiation factor RRN3

RRN3 is a Pol I-specific initiation factor that primes it for transcription initiation [99]. Several structures of yeast Pol I bound to RRN3 either alone [59] or in the context of a full PIC [61–63] are available. In yeast, RRN3 is thought to prevent dimerisation of Pol I and thus bring it into the active state [59,64]. No dimerisation of human Pol I has been observed, but the activating role of RRN3 in the transcription cycle is likely conserved between species [326,327].

3.4.1 RRN3

Human RRN3 binds Pol I in the expected position contacting the stalk and the RPA1 dock. The overall position matches when superimposed with the available yeast Pol I-RRN3 structures (Fig. 3.17 a). In yeast, the positioning of RRN3 has been strongly influenced by the tWH domain of the RPA49 subunit. RRN3 can slide on the Pol I dock either forward, when the tWH domain is lower, engaging with the upstream DNA, or backwards when the tWH domain is positioned closer to the stalk, potentially aiming to help dissociate RRN3 [61,125] (Fig. 3.17 a). In the case of the human Pol I-RRN3 structure, the binding of the tWH domain of the RPA49 subunit is not observed. Thus, RRN3 assumes a "neutral" middle position on the dock. In the yeast structures, a larger portion of the N-terminus is ordered compared to the human Pol I structure, potentially owing to the interaction with the tWH domain of the A49 subunit. It has been shown that the RPA49 subunit can interact with RRN3 also in humans [328], so it is possible that at a different stage in the transcription cycle larger part of RRN3 would become ordered upon such interaction.

Post-translational modifications modulate the interactions between Pol I and RRN3. The phosphorylation status can affect both binding and transcription activation of Pol I by RRN3, with eight phosphorylation sites mapped so far [102,103,106,107,329] (Fig. 3.17 b). Many RRN3 phosphorylation sites lie in disordered regions, which might become ordered at different stages of the transcription cycle or upon interaction with initiation factors. Interestingly, many residues must be phosphorylated to allow RRN3 to function. Given that the RRN3 I used to obtain this structure was recombinantly made in *E. coli* by Dr. Brice Murciano, it is unlikely to have the exact required phosphorylation pattern [132].

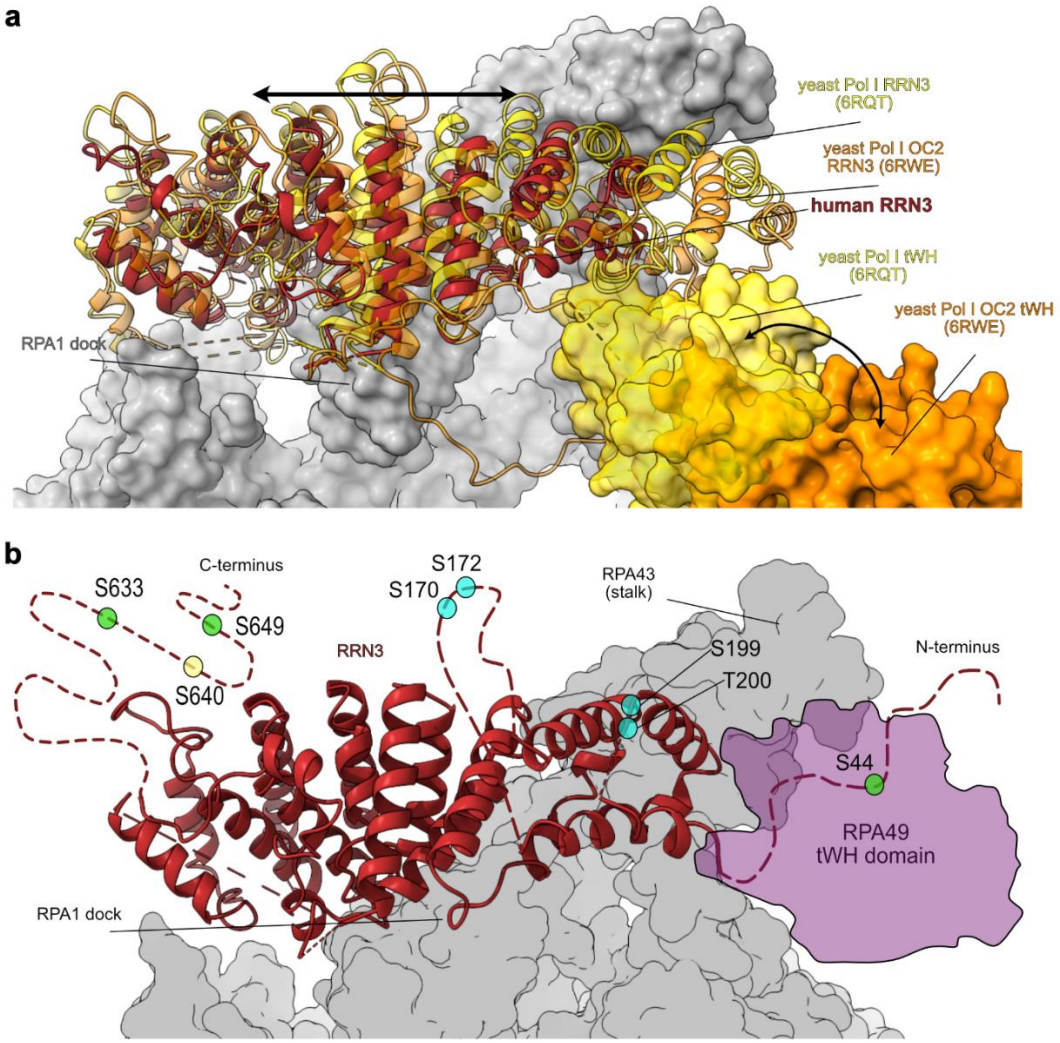

Fig. 3.17. RRN3 binding to Pol I. a| Superimposition of the human Pol I-RRN3 structure PDB: 7OBA (maroon) with RRN3 and tWH domain of A49 subunit from yeast structures (transparent, yellow and orange). **b|** Post-translational modifications of human RRN3 (cartoon representation, maroon). Pol I core is shown in grey surface representation. Unstructured regions are shown as dotted lines, not to scale. Phosphorylation sites are marked with transparent circles: activating sites in green [102,106], inactivating sites in cyan [102,103,107] and a site of unknown role in yellow [329]. The putative position of the tWH domain of RPA49 is shown as a purple transparent overlay. Panel b is taken from [23].

Cavanaugh *et al.*, 2002 [104] showed that dephosphorylated RRN3 has a reduced binding capability to the isolated RPA34 subunit. Here, however, binding is observed in the context of the complete Pol I complex, where RRN3 also binds to the RPA1 clamp core, suggesting that the interaction between Pol I and RRN3 is complex and relies on an interplay of multiple factors, phosphorylation pattern being only one of them. Those activating phosphorylation sites (such as S44, S633 and S649) may also be required to stimulate Pol I activity and not necessary for binding [102,106]. They could be involved in the SL1 binding instead. Other phosphorylation sites are only

important at specific steps in the transcription cycle; for example, in the case of residues S170 and S172, their phosphorylation reduces the affinity of RRN3 for Pol I, which could be significant in RRN3 release from the complex in transition from initiation to elongation [107]. Phosphorylation sites in the disordered regions are targets of many signalling cascades like the mitogen-activated protein kinase (MAPK) cascade [106]. They are vital to the regulation of Pol I activity in response to proliferation signals [106], nutrient availability [102] or stress [103]. Discerning the mode of action of those modifications throughout the transcription cycle will be highly important to understand the modulation of Pol I activity.

3.4.2 RRN3 binding induces tilting of the Pol I stalk

While RRN3 binds the dock of the RPA1 subunit, its main interaction site is the stalk. The binding of RRN3 induces a conformational change in which the stalk swings away by approximately 5 Å (Fig. 3.18 a). The lower, Rpb-7-like domain of the RPA43 subunit remains tightly anchored to the core and only the upper OB domain swings away. This movement is allowed by the higher flexibility of the human stalk compared to the yeast one. In yeast, the A14 subunit is positioned between the two domains of A43, stabilising their relative position (Fig. 3.18 b). In humans, where the homologue of the A14 subunit is absent, a hinge is introduced between the two domains of the RPA43 subunit. This allows the independent movement of the domains.

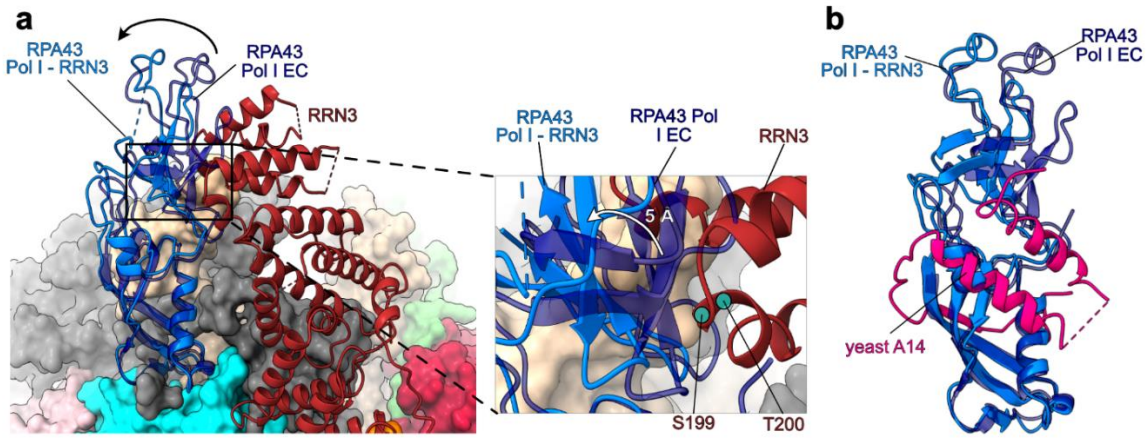

Fig. 3.18. RRN3 binding induces stalk tilting. **a|** Superimposition of the human Pol I-RRN3 structure with the RPA43 subunit from Pol I EC (dark blue, transparent). In the zoom-in, sites of inactivating phosphorylation are marked with cyan transparent circles. Arrows show the direction of stalk tilting from Pol I EC structure to Pol I-RRN3 structure. **b|** RPA43 subunits from Pol I EC (light blue) and Pol I-RRN3 structure overlaid with the smaller stalk subunit A14 from yeast PBD: 5M64 (pink). Panel **a** is taken from [23].

The closest point of contact between RRN3 and RPA43 is concomitant with the target site of two inactivating phosphorylations (Fig. 3.18 a). The two adjacent residues, S199 and T200 [107] might act as a phospho-switch. The introduction of a post-translational modification at this site could lead to a steric clash or change in the charge distribution, ultimately leading to a loss of favourable contacts between RRN3 and RPA43. Stalk flexibility might allow for conformational selection and more specific binding between RRN3 and Pol I.

3.4.3 DNA-mimicking loop

The Pol I-RRN3 structure is the only one I have obtained with no nucleic acid scaffold bound in the active site. Instead, a weak cryo-EM density is lining the DNA-binding cleft and occluding the active site (Fig. 3.19 b). While, due to the low quality of the density, I was not able to unambiguously assign the residue register, this density likely corresponds to the DNA-mimicking loop, also called the expander. The RPA1 subunit has a large insertion within its jaw (Fig. 3.19 a), which is fully disordered in the Pol I EC structure. A similar insertion is present in the yeast homologue of the RPA1 subunit, the A190 subunit. In the yeast Pol I crystal structure [21], the DNA-mimicking loop was shown to be ordered, and its position overlaps with the density visible in the cleft of the Pol I-RRN3 structure (Fig. 3.19 c, right panel). When the observed density is superimposed onto the Pol I EC structure, it directly overlaps with the bound nucleic acids. This indicates that the DNA-mimicking loop must be removed for the nucleic acids to bind.

The density corresponding to the first 10 residues of the DNA-mimicking loop was of sufficient quality, allowing me to build this residue stretch. When superimposing the cryo-EM density corresponding to this loop onto the Pol I EC structure, I noticed that it directly overlapped with the HTH motif from the linker of the RPA49 subunit. It can be thus speculated that the ordering of the DNA-mimicking loop inside the DNA binding cleft prevents the association of the RPA49 linker preventing it from closing the clamp and thus maintaining the open state of the complex [56,325].

The mechanism of Pol I inactivation using the DNA-mimicking loop appears to be conserved between yeast and humans. Yet, there is little sequence identity between the proteins from the two species in this region (Fig. 3.19 e). The insertion harbours a high number of charged residues, such as glutamic and aspartic acids, which contribute to the negative charge of the loop, mimicking the charge distribution of the DNA.

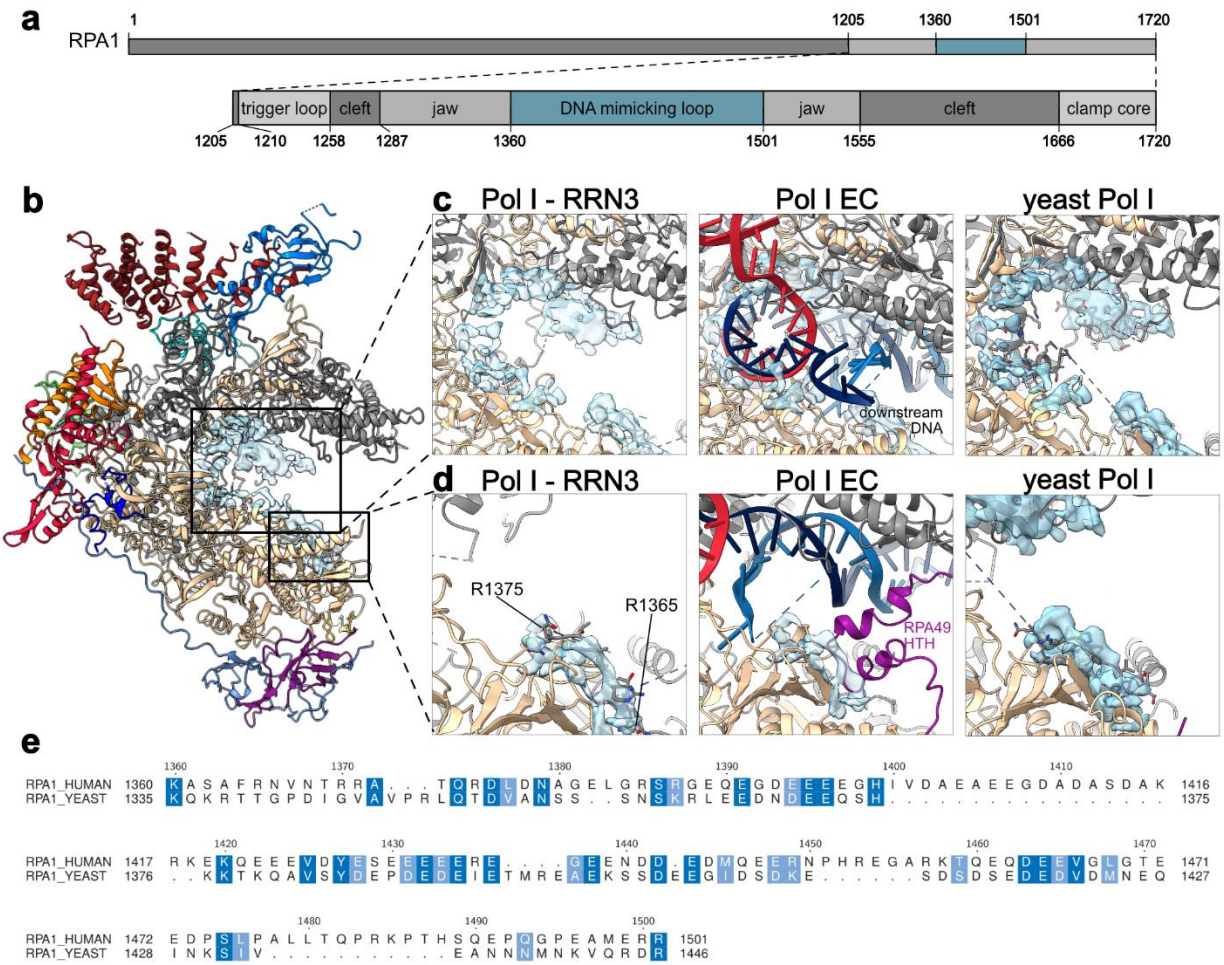

Fig. 3.19. DNA-mimicking loop in the Pol I-RRN3 structure. a| Domain representation of the RPA1 subunit with a zoom-in on the region of the DNA-mimicking loop (blue). **b|** Overview of the Pol I-RRN3 structure with the density likely corresponding to the DNA-mimicking loop shown as a blue transparent surface. **c|** Zoom-in on the DNA binding cleft and **d|** on the first 10 residues of the DNA mimicking loop in Pol I-RRN3, Pol I EC and yeast Pol I (PDB: 4C3J [21]) structures (left to right) with density lining the cleft in Pol I-RRN3 structure overlaid as a blue transparent surface. **e|** Sequence alignment of the human and yeast DNA-mimicking loop. Identical residues are coloured in dark blue, similar residues in light blue and dots indicate gaps in the sequence. Taken from [23].

The DNA binding cleft of Pol I is highly positively charged, and thus the DNA-mimicking loop likely forms many transient electrostatic interactions with the residues that, in the functional state, will bind the DNA. The lack of specific interactions might explain the poor quality of the cryo-EM density: the DNA-mimicking loop does not adopt any particular conformation inside the DNA-binding cleft, and thus it is not possible to average out a specific conformation for this region. This mode of function could be necessary for the element's functionality, which must be easily displaced to allow DNA binding.

3.5 Cryo-EM structure of the Pol I Open Complex

The Pol I OC structure captures an inactive state of the human Pol I. While an open DNA scaffold is bound in the cleft, it is not engaged with the active site elements, and the cleft is widely open. The Pol I OC structure represents a state in which the RNA synthesis still has to occur or where the RNA was cleaved using the RPA12 subunit catalytic activity. A similar structure has been observed for yeast Pol I [56].

3.5.1 Pol I in an inactive state

Several features are associated with the inactivation of Pol I, namely the opening of the cleft, the disordering of the active site elements and the insertion of the C-terminal domain of the RPA12 subunit into the active site [56]. Tight binding of the nucleic acid scaffold associated with closed cleft conformation supports complex' processivity. In the Pol I OC structure, the DNA scaffold is bound away from the active site in a non-productive conformation (Fig. 3.20 a, b). The density corresponding to the DNA, especially in the unwound and upstream portion, is weak, as expected in the case of flexible binding. Since I could not unambiguously assign the register of nucleotides, a large part of the nucleic acid scaffold is excluded from the structure.

Active site elements position the nucleic acid scaffold in a productive conformation where the catalytic addition of an incoming NTP can occur. Contrary to the Pol I EC structure, no density for the catalytic magnesium ion could be assigned in the Pol I OC structure. The active site residues (especially D590 in the RPA1 subunit) have weaker density, which might be caused by their flipping out in a mechanism previously suggested to occur in yeast Pol I [54]. Unfolding of the bridge helix might be associated with clamp opening, as in this conformation, it needs to stretch over a longer distance (Fig. 3.20 d). Yet, the bridge helix cannot efficiently position the nucleic acid scaffold in this conformation. The trigger loop might be disordered due to the insertion of the C-terminal domain of the RPA12 subunit, which occupies the same position as the trigger loop in the Pol I EC structure (Fig. 3.20 e, j).

The C-terminal domain of the RPA12 subunit is functionally related to TFIIS. It has RNA cleavage activity [311,330] – thus, its insertion into the active site is associated with a backtracked or termination state of Pol I. The C-terminus of the RPA12 subunit is observed in a fully inserted conformation with an extended catalytic loop harbouring catalytic aspartates (Fig. 3.20 b). In the previously observed structures

of the yeast Pol I OC, the catalytic loop of the A12.2 subunit was retracted and disordered [56]. Insertion of the TFIIS-like domain of the RPA12 subunit affects the trigger loop's conformation and induces the gating tyrosine flipping (Fig. 3.20 c, h). In humans, Y687 from the RPA2 subunit - the gating tyrosine, occludes the backtracking tunnel and prevents RNA backtracking. Changes to its conformation were first observed by Cheung and Cramer, 2011 [331] in the reactivation intermediate of the backtracked Pol II. The gating tyrosine was flipped in this structure to avoid contact with the backtracked RNA. In the human Pol I OC structure, no RNA is present, and thus the tyrosine flipping must be induced by a different mechanism, potentially the RPA12 subunit C-terminus insertion.

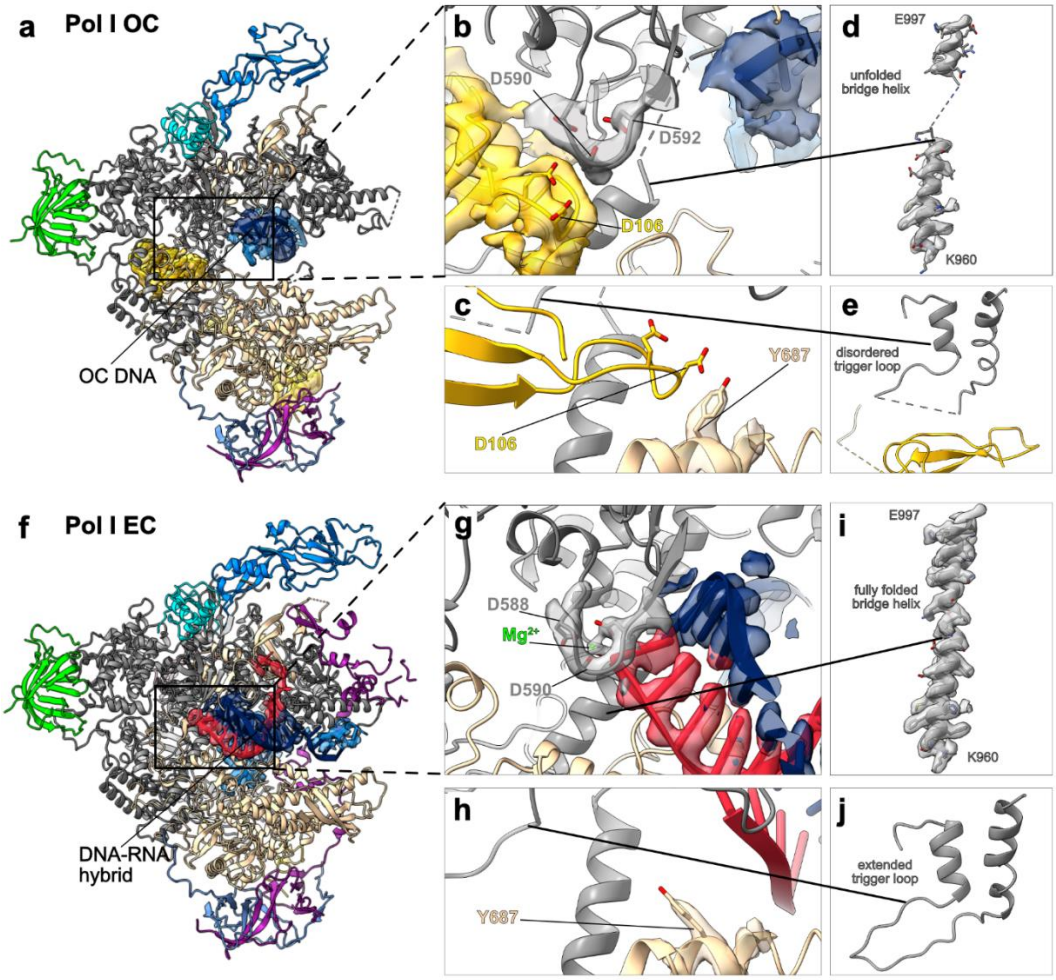

Fig. 3.20. Pol I OC represents an inactivated state. a| Overview of the Pol I OC structure in cartoon representation. **b|** Zoom-in on the active site of the Pol I OC. **c|** Conformation of the gating tyrosine (Y687) with the insertion of the C-terminal domain of the RPA12 subunit (yellow). **d,i|** Conformation of the bridge helix. **e,j|** Conformation of the trigger loop. **f|** Overview of the Pol I EC structure in cartoon representation. **g|** Zoom-in on the Pol I EC active site. Catalytic magnesium ion is shown in green. **h|** Conformation of the gating tyrosine. In all panels, transparent surfaces show cryo-EM density for the annotated elements. Figure taken from [23].

3.1 Conclusions and outlook

In this chapter, I have presented the structural analysis of human Pol I in different functional states. Firstly, I obtained lower resolution (3.7 Å - 4.7 Å resolution) structures of apo Pol I, which revealed the intrinsic flexibility of the complex that hindered obtaining the complete Pol I structure. Further, I obtained high-resolution structures of Pol I EC (at 2.7 Å resolution), Pol I – RRN3 (at 3.1 Å resolution) and Pol I OC (at 3.3 Å resolution). While the general architecture of the Pol I is conserved throughout the tree of life, several features that offer novel insights into the function of the human Pol I could be observed.

(i) Using the self-complementary RNA primer, I showed that Pol I can accommodate structured RNA in its RNA exit tunnel, which may allow it to transcribe at a higher rate.

(ii) I confirmed the subunit composition of the human Pol I. I showed that the stalk is composed of one subunit only, while previous studies of the yeast Pol I focused on species that are outliers within the tree of life in terms of the subunit composition of this domain.

(iii) The general strategy for associating the TFIIE-TFIIF-like heterodimer with the Pol I core is conserved between species, regardless of the low sequence identity.

(iv) Binding of the initiation factor RRN3 can be modulated by post-translational modifications and can induce conformational changes within Pol I by bending the stalk, which in humans is highly flexible.

(v) Strategies to inactivate Pol I such as the opening of the cleft, insertion of the DNA-mimicking loop, insertion of the C-terminal domain of the RPA12 subunit or disordering of the active site elements are also conserved between species.

While the novel human Pol I structures increase our understanding of Pol I transcription, each of the drawn conclusions opens new questions and requires further follow-up studies.

(i) In determining the Pol I EC structure, I used a pre-formed double-stranded RNA possibly only stabilised by the Pol I. Further studies of the co-transcriptional formation of secondary RNA structures within the exit tunnel and specific adaptations of the human Pol I to facilitate this process

are needed to expand our understanding of this observation and its significance.

(ii) While I could show that RRN3 can induce stalk bending, possibly allowing for the conformational selection of the association of Pol I with its transcription factors, possible further benefits of having a smaller and more flexible stalk need to be investigated.

(iii) The currently obtained structures shed light on the association of the heterodimer with the Pol I core. Yet, in the case of the RPA34 C-terminal extension, I could only resolve ~10% of it. The role of the remaining portion of the extension is not well understood. It will be interesting to see if it can become ordered throughout the transcription cycle, for example, upon binding accessory factors. Similarly, only one position of the C-terminal tWH domain of the RPA49 subunit is currently available for human Pol I. It remains to be determined if it can change its conformation throughout the transcription cycle, as it is the case for yeast Pol I.

(iv) In the Pol I – RRN3 structure I obtained, a large portion of RRN3, which harbours many phosphorylation sites, remains disordered. Further structural studies, for example, with addition of SL1, may shed light on the structure of those regions and the role of phosphorylation for Pol I activation.

(v) While the general strategies to inactivate Pol I are conserved between species, dimerisation which is a crucial strategy to inactivate yeast Pol I, has not been observed for human Pol I. Further confirmation of this observation and its implications need to be performed. Similarly, careful consideration of the RPA12 RNA cleavage activity is required, as some differences between yeast and human Pol I regarding backtracking activity have been observed.

(vi) The structure of Pol I bound to RRN3 lies ground for further studies of the complete human Pol I PIC.

POL I AND POL III IN DISEASE

Both Pol I and Pol III play a prominent role in the homeostasis of every cell. As described in sections 1.2.6 and 1.3.2, Pol I and Pol III upregulation is necessary for the transformation of cancer cells and impairment of Pol I and Pol III function is linked with many diseases, most prominently developmental ones. It is interesting to note that development is possible despite the impairment of the function of such essential enzymes. Even more curiously, only specific tissues are affected by Pol I and Pol III mutations. While a growing number of patients are diagnosed with Pol I- and Pol III-related conditions, understanding the role of each mutation in the aetiology of the diseases is still lacking. Conditions linked to Pol I and Pol III have a broad spectrum of symptoms and disease severity. Thus, extending the understanding of those disease processes is necessary to support the appropriate diagnosis and care of patients suffering from those complex disorders.

4.1 Mapping of disease-causing mutations

The aim of mapping and further studying the disease-associated mutations of human Pol I and Pol III was to gain insights into the conditions that affect people. Previous studies used mutagenesis also to gain insight into the function of Pol I and III in general, making important contributions such as discovering a hyperactive Pol I mutant [325] or furthering the understanding of the role of the heterotrimer during Pol III transcription initiation [332]. Those studies, however, mainly focused on yeast RNAPs making the extrapolation to the human complexes not direct. Additionally, whether those mutations could be found in humans and still support the development and function of a complex organism was not established. Thus, for the presented analysis, I only selected mutations that were found in human patients. With the overall improvements in sequencing technologies, genome sequencing for patient care is more commonly used, offering new avenues for diagnostics or treatment in some cases [333]. With efforts from many research groups and clinicians [190–192,196,198], many mutations found in Pol I and Pol III, which are thought to underlie complex disorders, have now been described. Many types of mutations have been found, including frameshift mutations (including insertions, deletions and duplications) [191,196], splice site variants [198,232] and missense mutations. In many cases, those mutations lead to large truncations or exon skipping resulting in highly altered proteins. However, many described mutations result only in a point substitution.

High-resolution structures of human Pol I and Pol III represent a unique tool allowing their analysis. Thus, in this section, I will only focus on analysing Pol I and Pol III point mutations derived from the patient studies.

A complete list of the considered mutations and references to the studies describing them are found in Tables A6 and A7 in the Appendix. To allow a more comprehensive analysis and comparison of different mutations, I derived two values that describe the overall position of each residue within the structure: the solvent-accessible surface area (SASA) of each residue and the number of other residues it makes contact with (residue degree). These values, together with a visual inspection of the location of each residue, allowed me to classify each affected residue into one of four categories:

Type I – maintaining intrasubunit contacts – characterised by high degree value and low SASA, a residue is found in the hydrophobic core of the subunit;

Type II – found in a functional element;

Type III – maintaining intersubunit contacts – residue makes contacts with other subunits characterised by high degree and low SASA score;

Type IV – located on a surface – above-average SASA value and low degree score.

4.1.1 Disease-associated mutations of Pol I

Having a human Pol I structure at high resolution (in the core, resolution approaches 2.5 Å), I could confidently map all disease-associated mutations. I have considered 13 mutations linked to Treacher Collins Syndrome (TCS) or Acrofacial Dysostosis, Cincinnati type (AD), which are thought to affect Pol I function. Affected residues are found in the structure's best-resolved parts, allowing confident analysis. Visual inspection confirmed that the cryo-EM density for all residues in question is of good quality, also for the side chains. Mutations linked to HLD are excluded from this analysis as they are more likely related to Pol III's disturbance. Their potential impact on Pol I is explored in section 4.2.

Mutations in the Pol I-specific subunits cluster around the active site (Fig. 4.1). They are found in the elements interacting with the nucleic acids (like R1003 from the RPA2 subunit found in the hybrid binding domain in Fig. 4.1 c) or with functional elements of the active site (like E593 from the RPA1 subunit found directly adjacent to the catalytic aspartate loop presented in Fig. 4.1 b or S682 from the RPA2 subunit

interacting with the bridge helix shown in Fig. 4.1 d). All those mutations may negatively impact the nucleotide addition cycle by affecting the structural rearrangements of the active site. Additionally, one mutation, V1299 in the RPA1 subunit (Fig. 4.1 e), is found in the region interacting with the RPA12 subunit. It might be necessary for the correct insertion of the C-terminal domain of the RPA12 subunit into the active site to allow RNA cleavage. Despite being linked to two different disorders, all mutations found in the core of Pol I are dominant [192,196].

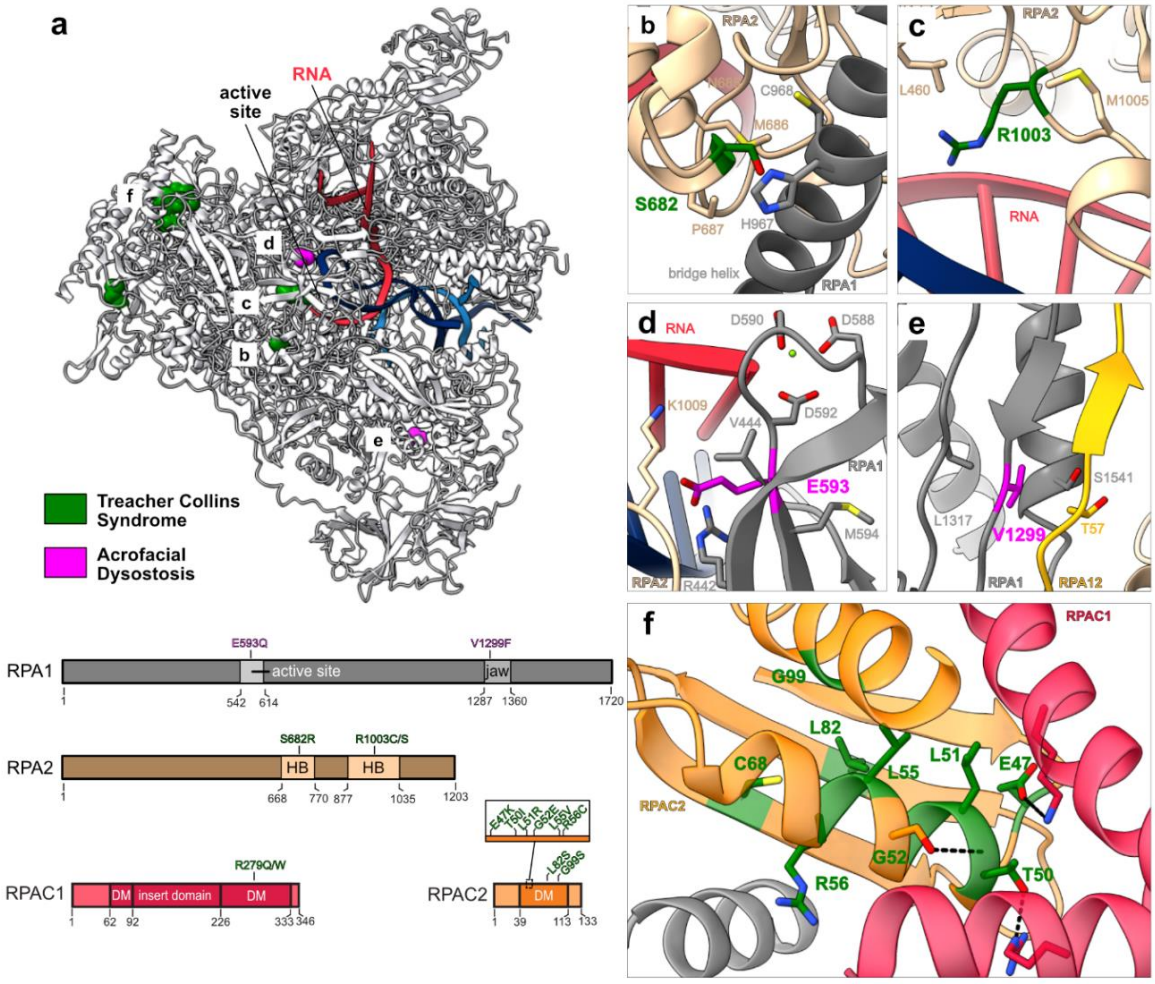

Fig. 4.1. Disease-associated mutations of Pol I. a| Disease-associated mutations of human Pol I are marked as coloured spheres. Affected subunits are shown in a cartoon domain representation with positions of the point mutations indicated. HB – hybrid binding domain. **b-f|** Close-up views of the residues affected by mutation in diseases. Subunits, as well as functional elements, are annotated. Black dotted lines show hydrogen bonds. Residues within 5 Å from the residue in question are shown in stick representation. Figure adapted from [23].

The other mutations associated with TCS are found in the two subunits shared exclusively between Pol I and Pol III, the RPAC1 and RPAC2 subunits. The majority of those mutations are dominant [191], except the R279Q/W mutation found in the RPAC1 subunit [191] and the L55V mutation found in the RPAC2 subunit [12,190]. The R279Q/W mutation has been described in patients showing compound heterozygous mutations in trans with a POLR1C splice variant leading to exon skipping or nonsense mutations resulting in truncated proteins [191]. Those mutations are familial, with carriers showing no symptoms, indicating that both alleles of the POLR1C gene need to be affected and that the point mutation contributes to the phenotype. Given that other TCS-related mutations are found in Pol I-specific subunits or in treacle protein which associates with UBF and plays a role in ribosome biogenesis [189], it can be speculated that TCS arises from Pol I perturbation. Further discussion covering the potential relationship of TCS with Pol I and Pol III perturbance is covered in section 4.2.

4.1.2 Disease-associated mutations of Pol III

The high-resolution structure of the human Pol III was solved in the Müller group by Dr. Mathias Girbig. By mapping 106 mutations on to the structure, I could gain insights into the potential roles of different mutations in the development of various conditions. Two neurodevelopmental conditions, Hypomyelinating Leukodystrophy (HLD) and Wiedemann-Rautenstrauch Syndrome (WDRTS), were the main focus of the analysis. Additionally, some mutations found in Pol III are also linked to the Varicella Zoster Virus susceptibility (VZV). Those mutations are mostly found on the periphery of the complex (Fig. 4.2), and affected residues are often solvent-exposed, potentially reducing the severity of the mutation. It can be speculated that those residues play a vital role in the activity of Pol III in the cytoplasm, where it detects the viral DNA and triggers innate immunity pathways [226,227].

Mutations associated with neurodevelopmental conditions are spread throughout the core of Pol III. Both HLD and WDRTS are recessive disorders. In most cases, the effects of two different mutations are needed to give a phenotype (only 12 out of 76 point mutations found in the RPC1 and RPC2 subunits have been observed in the homozygous state). Most affected residues are found in the two largest Pol III subunits, evenly distributed across the sequence (Fig. 4.2 a). Some of the mutation hotspots are located along the DNA binding cleft, potentially affecting the Pol III affinity for the target DNA (Fig. 4.2 b).

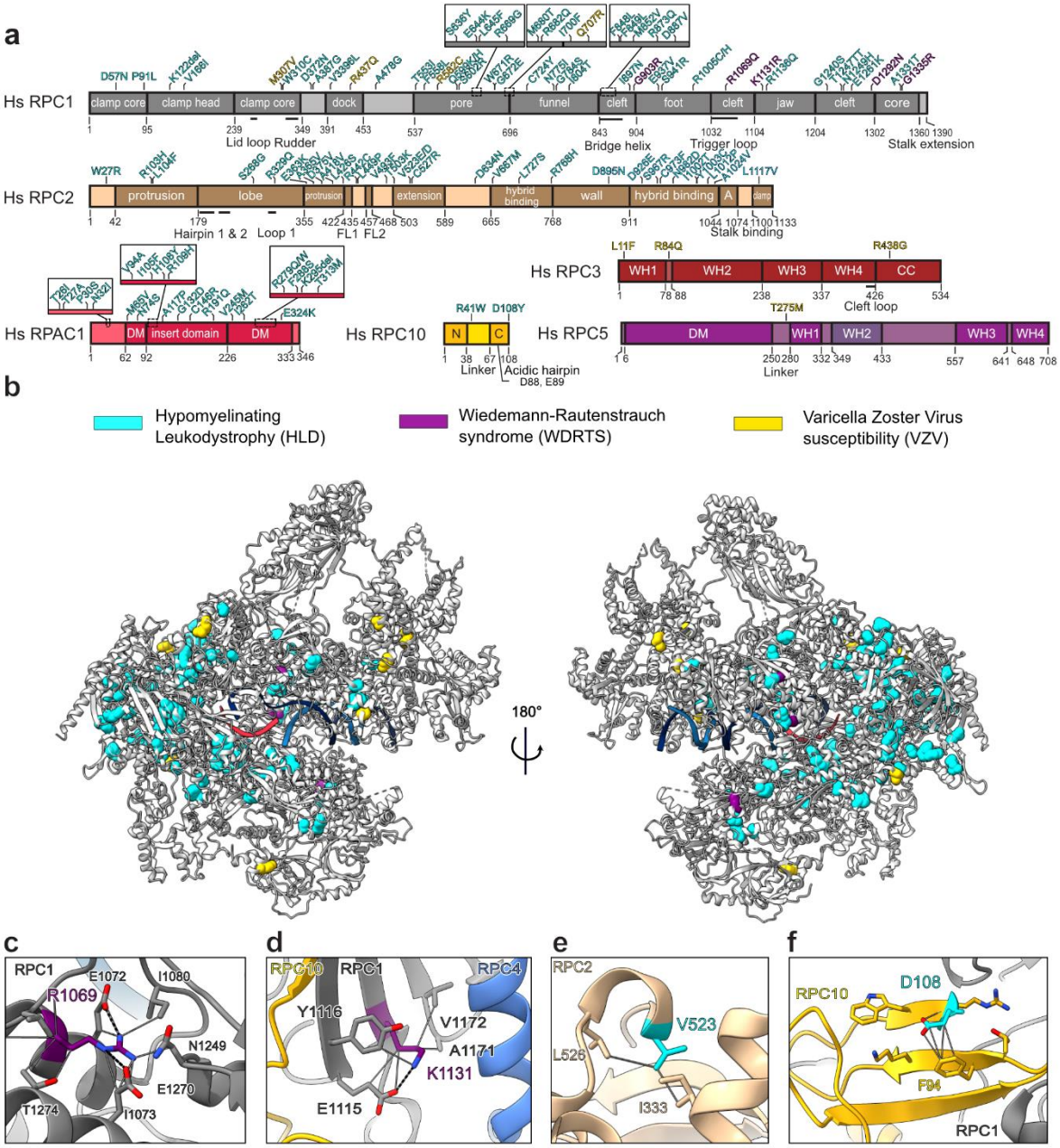

Fig. 4.2. Disease-associated mutations of Pol III. a| Cartoon domain representation of all affected subunits with the point mutations marked. Functional domains are annotated. A - anchor, DM – dimerisation domain, WH – winged-helix domain, N – N-terminal domain, C – C-terminal domain. **b|** Residues affected by disease-associated point mutations mapped on the Pol III structure (PDB: 7AE1) as coloured spheres according to the disease type indicated on top. **c-f|** Zoom-in view on selected residues affected by mutations (in a colour corresponding to the associated disease). Contacts are shown as solid grey lines and hydrogen bonds as dotted black lines. Residues making contact with the residue of interest are annotated. **e|** Panel made using PDB: 7D58 [316]. Figure adapted from [36].

While most point mutations are associated with HLD, a few are linked to WDRTS, which is sometimes regarded as an extreme form of HLD. Many RPC1 truncations have been described in patients with HLD or WDRTS [239,241]. However, in HLD, those are always *in trans* with a point mutation, while WDRTS protein truncations are often *in trans* with another nonsense variant [239,240]. WDRTS is hypothesised to arise from a specific combination of compound heterozygous mutations, where one would often be milder and the other more deleterious [241]. Out of the point mutations linked to WDRTS, three (G903R, D1292N and G1335R) are found in a core of the subunit (classified as type I) with a low SASA and high count of contacts, while the R1069Q mutation (Fig. 4.2 c) is found in the trigger loop. Those mutations likely have deleterious effects on Pol III and are coupled with mild intronic variants [241]. The remaining mutation associated with WDRTS, K1131R (Fig. 4.2 d), has been classified as type III with a significantly higher SASA and fewer contacts. This mutation might have a milder effect, coupled with a deletion of exons 12-15, yielding a non-functional RPC1 [241].

In the case of mutations associated with HLD, a compound effect of two different point mutations makes it challenging to discern the individual contributions of each substitution. Some correlations have been made between the severity of the condition and the affected subunit, with mutations found in the RPC2 subunit generally giving a milder phenotype [230]. One of the most common and also mildest (especially when homozygous) mutations [230], V523E in the RPC2 subunit (Fig. 4.2 e), forms fewer contacts and is more solvent-exposed. Some rare mutations map to other Pol III subunits, such as RPC10. Up to date, two point mutations have been described: the R41W mutation found in the linker region [238] and the novel D108Y mutation affecting the C-terminal residue (Fig. 4.2 f) recently described by our collaborators (Bernard group, submitted). I have analysed the potential impact of the D108Y mutation on Pol III's function. The RPC10 subunit undergoes large conformational rearrangements allowing the C-terminal domain with the catalytic, RNA-cleaving function to insert into the active site. The RPC10 linker mediates this crucial change of conformation, and the R41W substitution might impede its correct alignment. In the conformation with the C-terminal domain of the RPC10 subunit inserted into the funnel, the local arrangement of the D108 residue could be analysed [316]. A bulky residue in this location may lead to steric clashes or impede the C-terminus insertion as the tunnel is relatively narrow. Impairment of the RNA cleavage activity in Pol III results in unusual disease symptoms. Besides the typical features leading to the HLD diagnosis, patients exhibit gastrological symptoms not usually observed. In a zebrafish model, impairment of RPC2 subunit function leads to

underdevelopment of the digestive system, which can be rescued by overexpression of the RPC10 subunit [334]. These observations hint at a unique role of the RPC10 subunit during gut development that will require further studies.

Another subunit affected by HLD-associated mutations is the RPAC1 subunit. Since it is shared with Pol I, those mutations might also affect Pol I, as further discussed in section 4.2. Understanding the genotype-phenotype relationship in the case of Pol III mutations is very complicated due to the number of Pol III subunits affected by mutations and the number of mutation combinations described. A better understanding of the disease processes and access to high-resolution structures of human Pol III should help develop more accurate predictions of disease outcomes in the future.

4.2. Mutations shared between Pol I and Pol III

Pol I and Pol III exclusively share two subunits: the RPAC1 and RPAC2 subunits, which make up the primary RNAP biogenesis platform [74]. Many mutations that are linked to TCS and HLD map to those subunits. Given that other TCS-related mutations map to Pol I-specific subunits and other HLD-mutations map to Pol III-specific subunits, it seems that despite the mutations being found in both complexes, they might only affect one of them. Understanding how a mutation may affect one or the other complex exclusively will aid in understanding molecular mechanisms governing each disease, potentially aiding diagnostics or treatment.

4.2.1 Possible differences in mutations' effects on Pol I and Pol III

In total, 30 disease-associated mutations have been found in the RPAC1 and RPAC2 subunits (Fig. 4.3 a-c). Most residues mutated in TCS and HLD are classified as type III mutations since they are found on the interfaces between subunits or as type I since they are responsible for the stability of the hydrophobic core. Thus, substitutions of those residues may disrupt the biogenesis platform assembly and, ultimately, the whole complex. During assembly, additional factors bind the RPAC1-RPAC2 dimer and help its folding. It is currently unknown how the subsequent assembly processes discriminate between Pol I- and Pol III-specific subunits [74]. A residue found in the RPAC1 or RPAC2 subunit may also be used to connect with the Pol I- or Pol III-specific subunit. Those substitutions might affect interactions in only one of the RNAPs.

When analysing and comparing the structures of the human Pol I and Pol III, I noticed that several domains of the RPAC1-RPAC2 dimer look different in the two complexes, despite being the product of the same genes. Since solved structures come from a protein sample purified and prepared similarly (see section 2.2), the observed differences likely arise from the inherent differences between the RNAPs. Two regions showed the most striking differences: the region around R279 residue in the RPAC1 subunit (Fig. 4.3 d) and the N-terminus of the RPAC1 subunit (Fig. 4.3 e).

Mutations of residue R279 in the RPAC1 subunit (Fig. 4.3 d) have been linked to TCS and HLD. In Pol I, the residue makes a lot of contacts and is shielded from solvent by the C-terminal extension of the RPA34 subunit, which docks in between the RPAC1 and RPAC2 subunits. Instead, in Pol III, the residue is solvent-exposed and makes fewer contacts with the neighbouring residues. It can be speculated that a mutation at this location might thus affect Pol I more than Pol III.

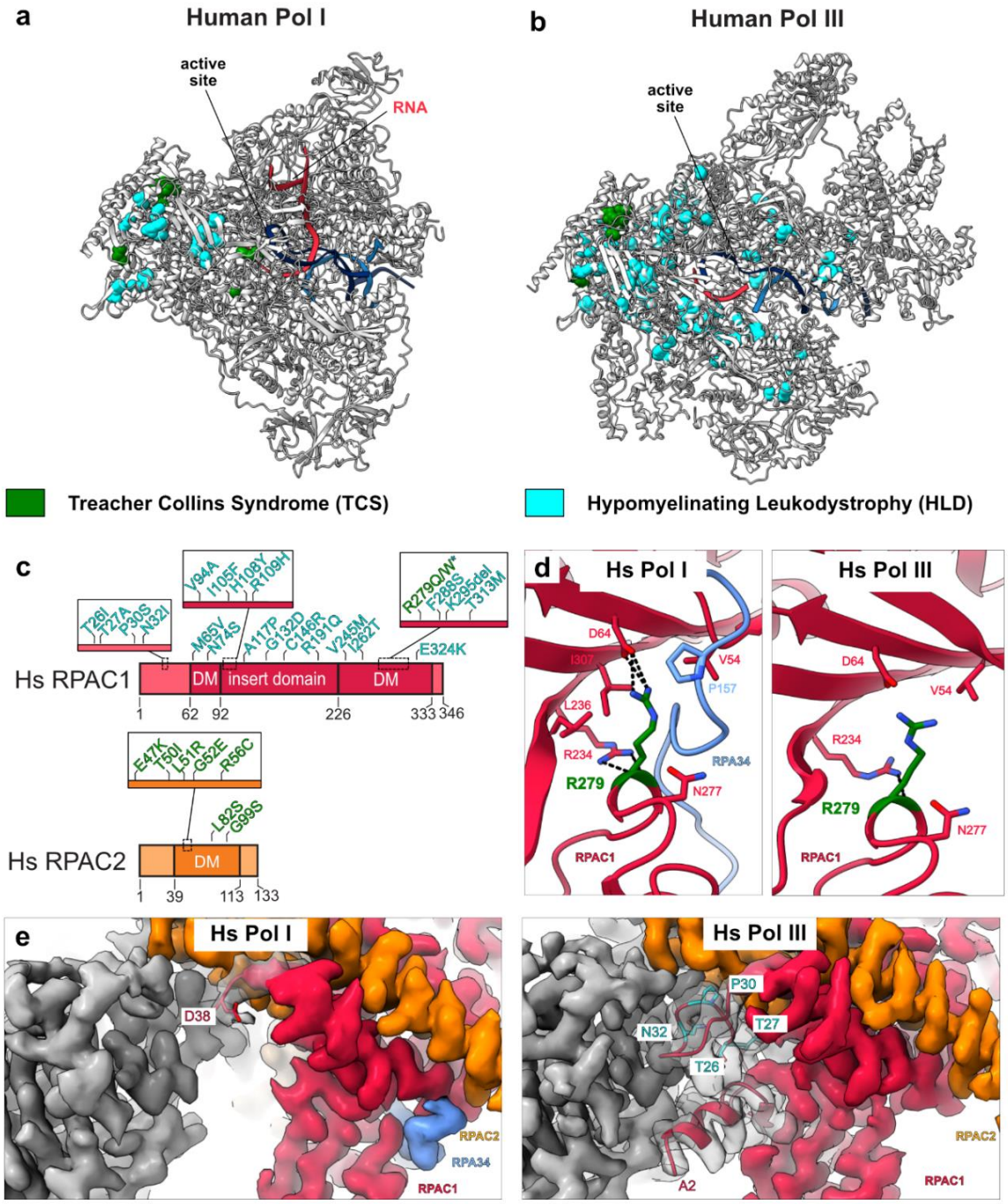

Fig. 4.3. Differences between Pol I and Pol III in regions affected by mutations. **a|** Human Pol I and **b|** Pol III in cartoon representation with Treacher Collins Syndrome (TCS) and Hypomyelinating Leukodystrophy (HLD) mutations marked as coloured spheres. Protein is shown in grey; nucleic acids are coloured with DNA in blue and RNA in red. **c|** Cartoon diagram representing shared exclusively between Pol I and Pol III subunits with disease-associated mutations mapped. The R279Q/W mutation marked with a star has been associated with TCS and HLD. **d|** Close-up view of the R279 residue from RPAC1 in Pol I (left) and Pol III (right). **e|** Close-up view of the RPAC1 subunit N-terminus in Pol I (left) and Pol III (right). Cryo-EM density is shown in transparent grey. HLD-associated mutations are marked in cyan. Panels **d** and **e** are adapted from [23].

The N-terminus of the RPAC1 subunit (Fig. 4.3. e) is fully disordered in the structure of the human Pol I. I could only build the structure starting at residue D38. On the contrary, all residues could be assigned in the Pol III structure. Four residues that have been reported to be mutated in HLD are found in this region. When mutated, they will more likely affect the interactions within Pol III, as they participate in more interactions within this complex.

Described observations are consistent with the findings reported by Thiffault *et al.*, 2015 [236]. The authors focused on two HLD-related mutations of the RPAC1 subunit, N32I (found in the N-terminal region) and N74S (found in the dimerisation domain). They could show, using overexpression of a single subunit with or without mutation, that substitution of N32 or N74 residues results in the impairment of the Pol III assembly while not affecting Pol I. They also investigated the effects of the R279Q mutation in the RPAC1 subunit and concluded that the mutation has an impact only on the localisation of Pol I. To follow up on those findings, I wanted to determine if the described pattern could be extrapolated to other mutations found in the RPAC1 and RPAC2 subunits. Additionally, to study the effects of mutations in the closest to native conditions, I decided to introduce the point mutations into the endogenous RPAC1 and RPAC2 subunits using the CRISPR-Cas9 method.

4.2.2 Construction of cell lines carrying mutations of interest

Given the time and resource constraints, I had to select a small number of mutations mapping to the RPAC1 and RPAC2 subunits, which would cover a wide range of described conditions. Based on the comparison of the human Pol I and Pol III structures, I selected three mutations in each RPAC1 and RPAC2. For the HLD-associated mutations, I picked mutations with different modes of inheritance, with one mutation that is found only as a compound heterozygote, one that has been described as homozygous and one that has been observed in both combinations. I also decided to investigate the R279Q mutation in the RPAC1 subunit, as it is the only mutation linked to TCS and HLD.

I also selected three mutations mapping to the RPAC2 subunit associated with TCS. Due to time limitations, I have not yet inserted them into the endogenous subunit, but plan on performing the experiments in the near future. The complete list of selected mutations can be found in Table 4.1.

Table 4.1. Selected disease-causing point mutations, their inheritance type, and their potential influence on human Pol I and III.

Mutation	Subunit	Disease	Inheritance	Predicted outcome
T26I	RPAC1	HLD	Recessive, found as a compound heterozygous and biallelic variant [233]	It may affect only Pol III, located in the disordered Pol I region.
I105F	RPAC1	HLD	Recessive, compound heterozygous [233]	It affects RPA2/RPC2 interaction more in Pol III.
R279Q	RPAC1	HLD/TCS	Recessive, found as a compound heterozygous (in TCS) and biallelic variant (in HLD) [233]	A greater effect in Pol I, residue buried by the RPA34 subunit.
F288S	RPAC1	HLD	Recessive, found as a biallelic variant [335]	Interacts with RPAC1 N-terminus, which is only ordered in Pol III.
L51R	RPAC2	TCS	Familial, heterozygous – likely dominant [191]	The buried surface introduces charge repulsion / steric clashes.
R56C	RPAC2	TCS	Familial, heterozygous – likely dominant [191]	Disruption of interaction with other residues or folding issues.
G99S	RPAC2	TCS	*De novo*, heterozygous – likely dominant [191]	Residue buried inside the core may affect the stability or lead to folding issues.

To test the effect of the mutation on subunit localisation and complex assembly, I aimed to introduce the mutations into the already tagged subunit (Fig. 4.4 left panels and section 2.1). The methodology used to introduce the endogenous point mutations has been developed in the Noh group [336]. In short, it relies on the delivery of the Cas9 protein with the crRNA-tracrRNA scaffold and repair template carrying the mutation of interest into the cell line of choice via nucleofection. The used Cas9 protein has a monomeric streptavidin tag, and the repair template is biotinylated, tethering the two components to each other and increasing the local concentration of the repair template, promoting the HDR DNA repair pathway. Further, the tracrRNA used is labelled with a fluorescent ATTO550 tag allowing the selection of cells that have been successfully transfected via FACS (Fig. 4.4). Following the expansion of the selected clones, cell lines can be genotyped using a mismatch quantitative PCR (qPCR). Two pairs of primers are designed, one complementary to the WT sequence and one complementary to the mutated sequence. A primer perfectly complementary to the genomic sequence will anneal to the target sequence faster and give products earlier than a primer containing a mismatch. The cycle threshold numbers (Ct) with each primer set are compared to

determine which allele combination is found in the cell line. In a clone with homozygous mutation insertion, the Ct value with the primer complementary to the mutated sequence will be lower than with the primer complementary to the WT sequence. Suppose both primer sets give common, low Ct values (comparable to the value obtained with the WT primer set in the negative control). In that case, the edited clone is likely heterozygous for the mutation, as it has sequences perfectly complementary to both primers. After pre-screening with the mismatch qPCR, candidate cell lines are validated using Sanger sequencing to confirm the insertion of the mutation.

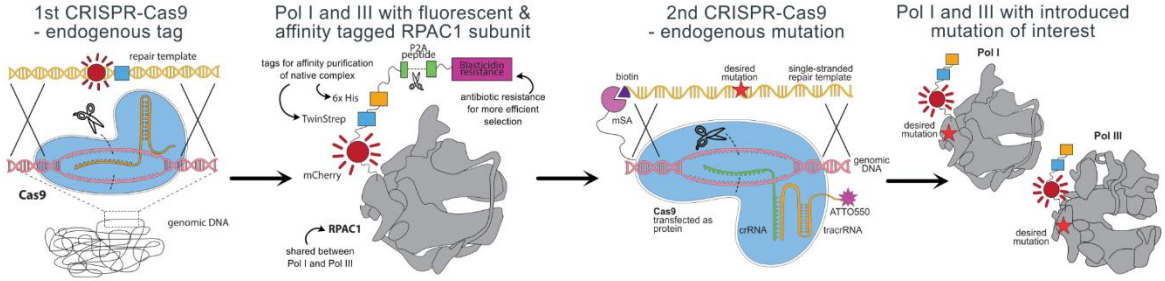

Fig. 4.4 Cartoon representation of the pipeline used to introduce endogenous point mutations. Cas9 is shown in blue, repair template in yellow, and RNAPs in grey. The components of the inserted tag are annotated. Introduced point mutation is marked with a red star.

Using the described methodology, with two subsequent CRISPR-Cas9 genome editing steps, stable cell lines where native Pol I and Pol III contain a mutated subunit tagged with fluorescent and affinity tags are created. With help from Granita Lokaj during the introduction of the point mutations, I could obtain four cell lines with the desired mutations in the RPAC1 subunit. Mutations are inserted homozygously to eliminate the background coming from the WT protein. Three cell lines were selected for each mutation to act as biological replicates. The biological replicates account for potential off-target effects from the CRISPR-Cas9 genome editing strategy and clonal variability resulting from HEK293 cells' genomic instability. Due to the time constraints, cell lines with mutations in the RPAC2 subunit are yet to be made. However, as described in section 2.1.6, a cell line with RPAC2 subunit tagged with fluorescent and affinity tags has already been prepared and validated, leaving only the mutations to be inserted.

4.2.3 Changes in the subunit localisation with mutations

Having constructed cell lines that carry a disease-associated mutation in a subunit tagged with a fluorescent signal, I could investigate the changes to the localisation of the subunit. Under normal conditions, Pol I localises to the nucleolus, which gives the brightest signal due to the high density of the complex. Pol III is distributed across the nucleus and in smaller amounts in the cytoplasm [10]. Using a combination of a DNA dye and transmission signal, I could estimate the boundaries of all those compartments. The nucleolus excludes the DNA stain due to the lower DNA concentration [297]. One of the challenges associated with imaging HEK293 cells is their loose adherence to the surface. Thus, I opted to use Höchst dye which is compatible with live-cell imaging and does not require extensive washing or cell permeabilisation. Additionally, HEK293 cells are thicker than the more strongly adherent cells, increasing the bleed-through issues from other planes. This problem is compounded by HEK293 cells having multiple nucleoli. The high brightness of their signal can skew the intensity in the neighbouring planes. To mitigate those issues, I decided to use confocal microscopy, which helps reduce the out-of-focus fluorescence [337]. The imaging conditions were set with advice from the EMBL Advanced Light Microscopy Facility, and the acquisition of images was performed together with Granita Lokaj.

For each of the three biological replicates of each cell line carrying one of the four mutations, I acquired ten images with a field of view covering 10 – 20 cells (depending on the cell density). Images were acquired under the same settings over two consecutive days. Control cells were also imaged: a WT control with no tag inserted (negative control) as well as a cell line with the tag on the RPAC1 subunit (in section 2.1 annotated as D3) but no mutations (positive control). Images were adjusted using the same transformation applied to all images to allow direct comparison. Representative images are presented in Fig. 4.5.

As expected, no mCherry signal was detected in the WT cells (Fig. 4.5 a). The D3 cell line with a tag inserted on the RPAC1 subunit shows the expected signal distribution with the brightest spots localising in the nucleolus marked by the absence of the signal given by the Höchst dye and with a diffuse signal covering the nucleus (Fig. 4.5 b). Under visual inspection, the introduction of the mutation changes the localisation of the tagged subunit.

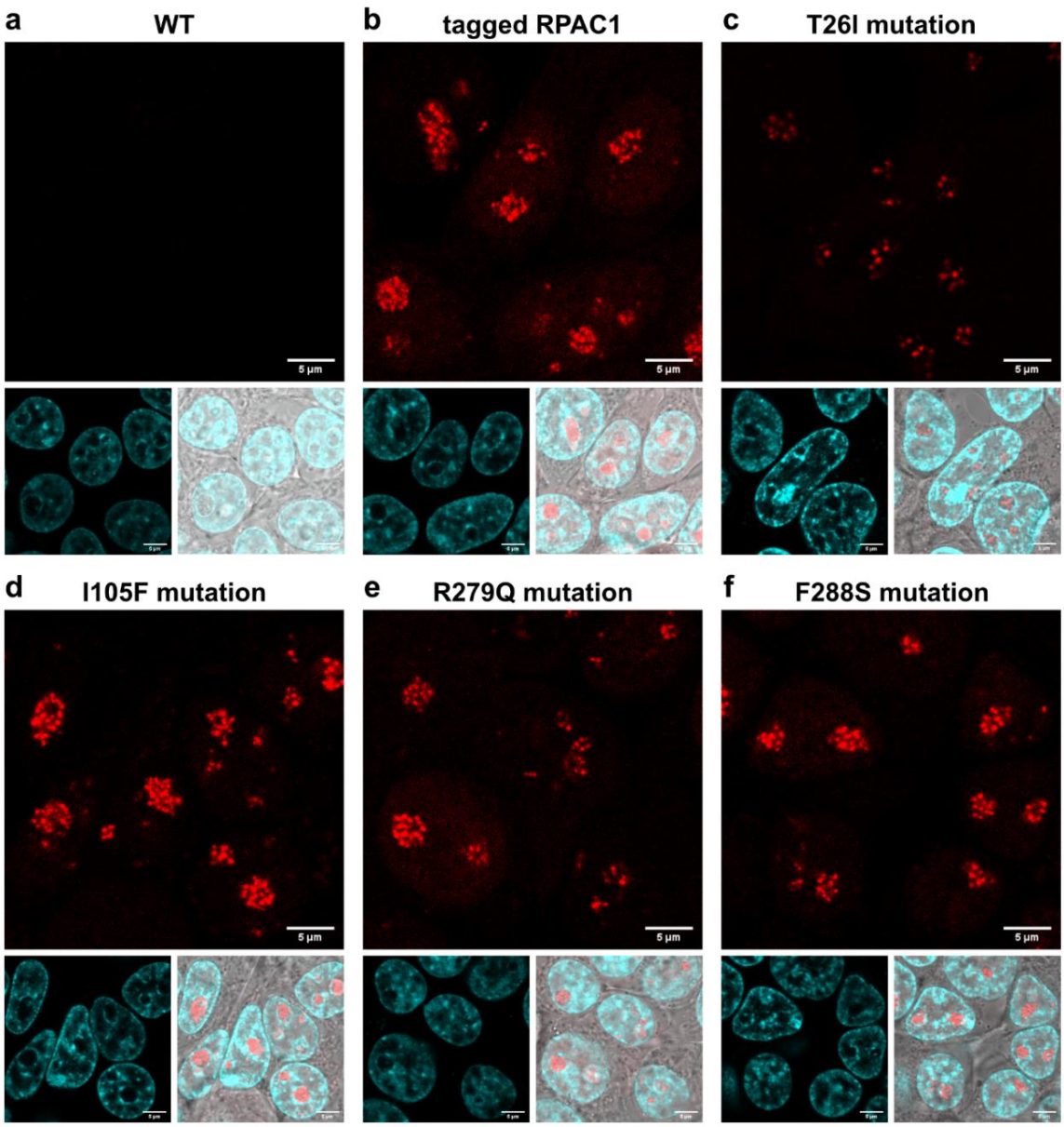

Fig. 4.5. Confocal microscopy images of cell lines carrying desired point mutations. Top panel: mCherry signal corresponding to endogenously tagged RPAC1 subunit. Bottom left panel: Höchst staining was used to mark the DNA. Bottom right panel: Overlay of mCherry and Höchst signal over transmission image (grey). Scale bar = 5 µm. **a|** Wild-type (WT) HEK293T cells. **b|** HEK293T cells with an endogenous tag on RPAC1 subunit (D3 cell line). **c-f|** D3 cell line with introduced **c|** T26I mutation, **d|** I105F mutation, **e|** R279Q mutation and **f|** F288S mutation in the RPAC1 subunit.

The most striking change in distribution is observed for the T26I mutation (Fig. 4.5 c). The diffuse signal in the nucleus is absent in this line, and the signal in the nucleolus is dimmer. There is also a smaller number of spots in each nucleolus. This cell line was also showing the most distinct abnormal growth phenotype. Cells grew about 30-50% slower than the other lines (to obtain confluency after two days, they had to be split in a 1:5 ratio compared to the other lines that were split in a 1:10 ratio). After transfection, a lower number of cells were viable, which translated

to the recovery of a smaller number of clones following FACS. The T26I mutation is located in the N-terminus of the RPAC1 subunit, and it was predicted to only affect Pol III. Yet, given the noted changes in the localisation and growth issues, the mutation appears to be more deleterious than predicted and affects both Pol I and Pol III. No increase of the signal in the cytoplasm is observed, indicating that mutation of the T26 residue of the RPAC1 subunit is not affecting the import of the complex. Instead, it might reduce the stability of the subunit leading to its degradation and an overall decrease in the observed signal.

Two other mutations associated with HLD, I105F (Fig. 4.5 d) and F288S (Fig. 4.5 f) mutations in the RPAC1 subunit, show similar changes in the signal localisation. The signal intensity observed in the nucleolus is similar to the positive control. A slight reduction of the signal in the nucleus can be noted. The differences, however, are not strongly pronounced and careful analysis of the differences in intensity is needed to confirm the effect. Nevertheless, a signal reduction in the nucleus indicates adverse effects on the Pol III function, consistent with the predicted mode of action of the mutations. Both I105F and F288S mutations are expected to affect the interfaces which differ between Pol I and Pol III (Table 4.1). Affected residues are on the interaction surface with the RNAP-specific subunit (in the case of I105 residue) or interact with the N-terminus of the RPAC1 subunit (in the case of F288 residue), which can be folded or not. Changes in those areas might affect assembly and thus localisation of RNAPs.

Based on the differences in structures, the R279Q mutation (Fig. 4.5 e), associated with both HLD and TCS, was supposed to affect Pol I more than Pol III. In some cells, the signal in the nucleolus looks less bright compared to the positive control. Similar observations have been made using the overexpressed RPAC1 subunit with the R279Q mutation [236]; however, in this study, the exclusion of the subunit from the nucleolus was more pronounced. Yet, in the approach by Thiffault *et al.,* 2015, the native WT RPAC1 subunit is still present in cells, and it is possible that the mutated subunit was preferentially not incorporated into the functional Pol I complexes. Quantifying the signal will determine whether observed differences are indeed significant and match the observations previously made.

4.2.4 Effects of mutations on the Pol I and Pol III assembly

Looking only at the localisation of the tagged subunit, it is impossible to state whether it constitutes a part of a functional RNAP or if it just localises to the correct cellular compartment. RNAPs are stable protein complexes that can be purified using a tag on only one subunit, as shown in section 2.1. Therefore, a partial affinity purification coupled with mass spectrometry (MS) can be used to assess the impact of a mutation on the assembly of the complex. I took advantage of the inserted affinity tags on the mutated subunit, which allow for binding of the RPAC1 subunit to Strep beads using the TwinStrep-tag. To reduce the background coming from the non-specific binding of proteins to beads, the 3C protease was used to release RPAC1 from the beads. Following one-step affinity purification, samples were submitted for MS analysis to the EMBL Proteomics Core Facility with the positive control (D3 cell line) and negative control (WT cell line). A quantitative MS analysis was performed. Thanks to Tandem Mass Tags (TMT), samples can be multiplexed, allowing directly comparing the protein quantities between the samples [236]. Granita Lokaj performed affinity purification with support from Jonas Wiedenhausen. The EMBL Proteomics Core Facility performed the MS experiment and data analysis resulting in the presented plots.

Excluding contaminants, 150 proteins were identified by MS in the sample. All subunits constituting Pol I and Pol III, except for the Pol I-specific RPA12 subunit, have been identified. Of those, 58 proteins have been identified in all samples, including all Pol I and Pol III subunits except for RPABC2, RPABC4 and RPC10. Subunits that were not detected are small, which gives rise to fewer peptides and more problematic detection in the MS experiment. Only the proteins identified in all samples have been considered in the analysis. The three biological replicates submitted for each mutation have been treated as a separate samples. Heterogeneity between the biological replicates has been detected, which might arise from actual differences between the lines or technical differences introduced during the affinity purification. Due to time constraints, I have not yet investigated the source of this variability, but I intend to perform such an analysis soon.

Amounts of proteins obtained through the affinity purification with the tagged RPAC1 subunit used as bait have been quantified and compared between the samples. Ratios of protein quantities compared to the positive control have been derived and presented as a heatmap (Fig. 4.6 a). In the WT, acting as the negative control, each protein of interest is less abundant than in the positive control. This indicates that using the RPAC1 subunit as bait enriches the sample in proteins belonging to Pol I and Pol III complexes. In protein samples derived from the mutant

samples, many subunits of interest are less abundant. This indicates that the introduction of a mutation may indeed disturb the assembly of the complexes. Interestingly, some subunits have been detected as higher than the control levels. Those belong mainly to the peripheral domains of the complex or are located away from the biogenesis platform. Nevertheless, a higher ratio of some subunits might also hint at assembly issues with non-canonical assemblies being formed.

Differences in the ratio of the protein amount between the samples coming from cell lines with introduced mutations and between the positive control can also be used to identify protein hits or candidates that are either significantly reduced or increased in the sample of interest. Those can be plotted over the abundance of each protein in the sample of interest, expressed as the TOP3 value, which is derived from the mean of the three most abundant peptides for the protein in the sample [338].

For each mutation, the biological replicate which showed the most changes in the protein abundances compared to the positive control was analysed (Fig. 4.6 b). The two mutations noted as homozygous, T26I and F288S, showed the most identified hits. In the sample coming from the T26I cell line, subunits belonging to both Pol I and Pol III have been identified as less abundant. This is consistent with the observed changes in the localisation of the RPAC1 subunit in this cell line, where its localisation to both nucleus and nucleolus was reduced. In the sample from the cell line with F288S mutation inserted, subunits belonging to the core of Pol III and some subunits making up Pol I are found to be less abundant. This might suggest that the mutation generally compromises Pol III's assembly but also, to a lesser extent, influences Pol I assembly.

Less significant changes in the protein abundances have been detected between the positive control and the samples coming from the cell lines with I105F or R279Q mutations. In the case of the I105F cell line, more of the Pol III-specific subunits are candidates for being less abundant compared to the positive control. This could suggest that indeed mainly the assembly of Pol III is affected. In the sample coming from the R279Q cell line no significant differences in protein quantities are detected, suggesting that the assembly of neither Pol I nor Pol III is affected, which is consistent with previous findings by Thiffault *et al.*, 2015 [236].

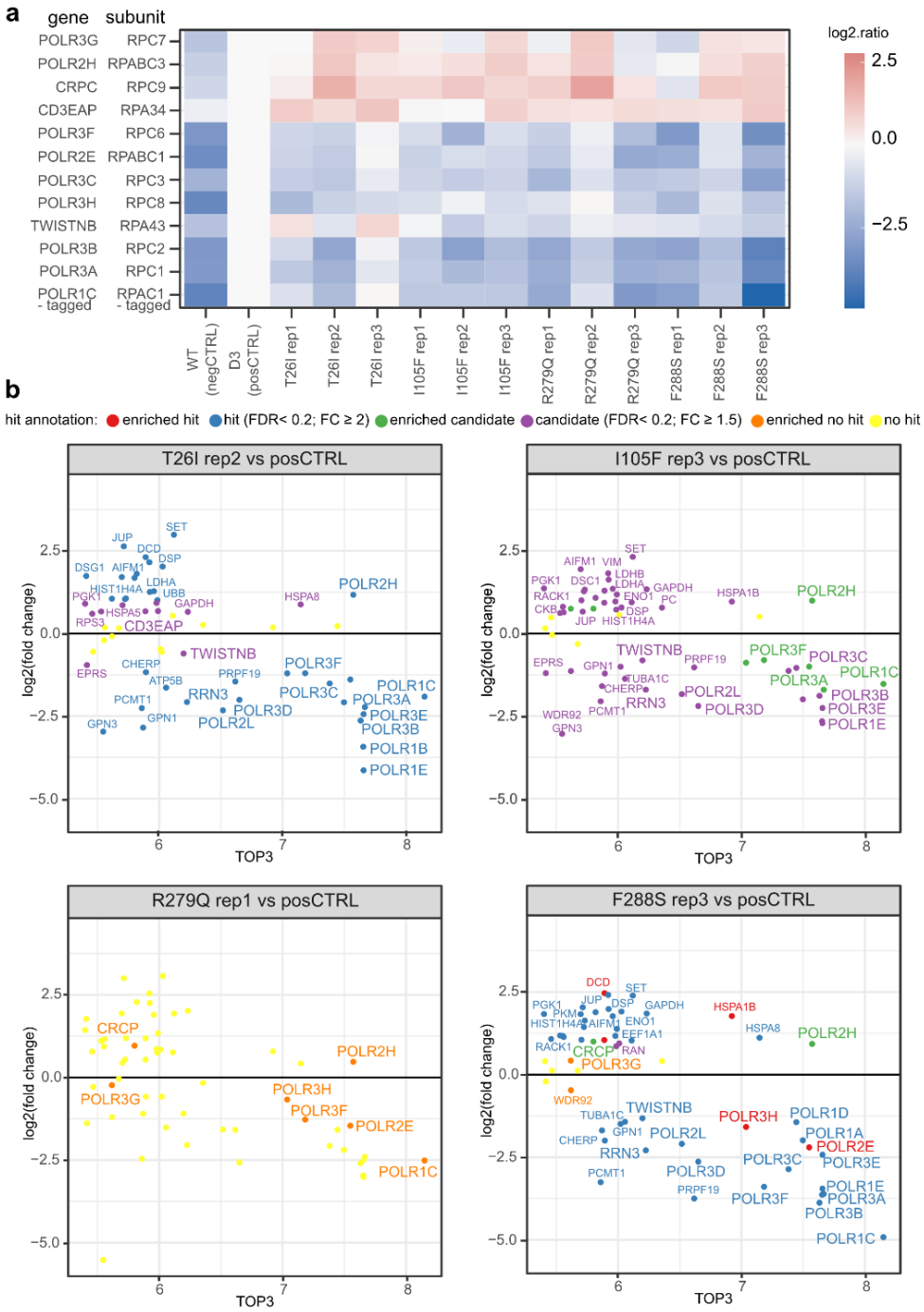

Fig. 4.6. Quantitative mass spectrometry results following a pulldown using endogenously tagged RPAC1 as bait. a| Heat map showing the log2 fold changes (log2.ratio) in the levels of identified proteins (labelled with gene and protein name) in cell lines carrying a mutation in the RPAC1 subunit. Positive control (D3 posCTRL) only has a tag on the RPAC1 subunit, and negative control (WT negCTRL) with no tag are included on the left. For each mutation, three biological replicates are shown. The log2.ratio is calculated with respect to posCTRL. Only components of Pol I and Pol III are shown. **b|** TOP3 plots showing the mean of the three most abundant peptides identified in a sample over fold change (expressed on log2 scale) in protein abundance between cell line carrying a mutation and positive control. One selected biological replicate with the most significant differences for each mutation is presented. The colour code is included on the top of the panel. FDR - false discovery rate; FC – fold change.

4.2.5 Conclusions and further studies

Only preliminary results are described in sections 4.2.3 and 4.2.4. To understand the effect of mutations on the Pol I and Pol III function, many further studies are needed, some of which are outlined in this section. Additionally, I plan to also construct cell lines with mutations in the RPAC2 subunit and analyse them in the same manner as the cell lines with mutations in the RPAC1 subunit.

- **Cell lines validation**

The correct insertion of the mutation at the desired location has been confirmed using Sanger sequencing. Yet, it must be assured that observed effects are a direct result of the introduced mutation and not an effect of an off-target mutation introduced by CRISPR-Cas9. Thus, whole-genome sequencing of the constructed cell lines should be performed, followed by an analysis of the potential effects of all *de novo* mutations found. Having three biological replicates for each mutation partially makes up for any variation introduced during the genome editing process, yet further confirmation of the genetic make-up of cell lines will strengthen the observations.

Further, it must be confirmed that observed differences in subunit localisation or assembly of the RNAPs result from protein malfunction, not protein expression changes. A qPCR experiment should be performed to compare the mRNA levels in all cell lines and ensure they are comparable.

- **Confocal microscopy**

Acquired microscopy images give a good indication that there are differences present in the localisation of the tagged subunit with the introduction of the mutation. The differences in the signal intensity need to be however quantified, and their statistical significance needs to be established. One challenge associated with quantifying the signal is the automation of cell segmentation. Automatic detection of cellular compartments would aid in a higher throughput of the analysis, allowing measurements across more cells. One solution to the segmentation problem could be to add distinct fluorescent markers to the plasma membrane and the nucleolus. Several kits compatible with live-cell imaging are available for membrane staining, such as CellBrite, MemBrite (biotium) or CellLight (Thermo Fisher). To segment the nucleoli, either exclusion of the Höchst dye can be measured, or a specific stain such as CytoPainter (Abcam) can be used.

To capture and compare the signal coming from all nucleoli in a cell, a 3D reconstruction is needed. HEK293 cells are round, and the nucleoli are found at different distances from the coverslip. Acquiring images at other planes and

reconstructing a 3D image of each cell will allow more precise quantification of the signal and account for the potential bleed-through from different planes.

- **Mass spectrometry**

To test for the changes in the assembly of Pol I and Pol III, a one-step affinity purification followed by MS is used. In addition to the biological replicates used, technical replicates are needed to account for differences arising from the sample preparation. Additionally, to reduce the impact of the affinity purification on the result, the second tag present on the bait subunit (a His-tag) can be used. Different beads used for sample preparation will reduce the effect of non-specific protein binding.

- **Cell proliferation**

Assays which measure cell proliferation to assess the impact of the mutation on cell growth should be performed. From general observations, mutations lead to slower cell growth. This phenotype, however, needs to be quantified, and the effect on the cell cycle should be investigated. For example, an MTT assay can quantify cell viability, and FACS-based methods can help track the cell-cycle progression.

- **Effects of RNAP activity**

Introduced mutations may not only affect the localisation and assembly of the complex but also its activity. Transcriptional capabilities of the complexes carrying the mutation can be measured *in vitro*. The complex would need to be purified to perform such studies, either directly from the cell line carrying the mutation or using the recombinant Pol I and Pol III expression systems.

Given that the introduced mutations are located away from the active site, it is possible that they will not directly affect the processivity of the enzyme but instead might disrupt interactions with transcription factors. Chromatin Immunoprecipitation Sequencing (ChIP-Seq) can be performed to measure the occupancy of Pol I and Pol III on their target genes. Changes in occupancy and transcriptional output can be measured by testing for the levels of pre-rRNA, tRNAs and 5S rRNA. Those experiments would have to be performed in collaboration, for example, with the Zaugg group.

Chapter III: Outlook and Future perspectives

OUTLOOK AND FUTURE PERSPECTIVES

5.1 Use of CRISPR-Cas9 for structural studies

New applications constantly emerge with the growing number of developments in CRISPR-Cas9 technology. One of them has been presented in this thesis, where I used CRISPR-Cas9 to introduce an endogenous tag on one of the complex's subunits of interest, allowing me to purify the intact protein assembly. The employed strategy has several advantages, including studying complexes closer to native conditions. Subunits in the purified complex are in their stoichiometric ratios, and if a mild purification strategy is used, interactions with associated factors can also be conserved. The purified complex will contain post-translational modifications required for its function in cells. It is possible to envision that this strategy might allow testing how the protein complex changes with perturbances. After all, one could purify the complex of interest after treating cells with a drug, applying any environmental stimuli or selecting a particular stage in the cell cycle.

Further, CRISPR-Cas9 technology has a growing role in *in situ* structural studies, such as in Correlative Light and Electron Microscopy (CLEM) approaches used in the cryo-Electron Tomography (cryo-ET) pipeline for localisation of the target proteins. Fluorescent tags allow guiding the data acquisition and annotation of cryo-ET data when the two signals are overlaid. There is a growing need to access structural information directly from cells, and CRISPR-Cas9 has a significant role in allowing those developments.

5.2 Structure of human Pol I

The high-resolution structure of human Pol I allows us to gain many insights into the function of this essential complex. However, it also opens many new questions, as outlined in section 3.6. To fully understand Pol I function throughout the transcription cycle, more structures need to be solved, ideally with additional factors. With the cloning of the Pol I initiation factors, I have laid the ground for studying the human Pol I PIC. While an isolated structure of the Pol I PIC would already answer many questions in the field, looking at Pol I in the context of chromatin would be even more fascinating. Cryo-ET studies by Dr. K. H. Herman Fung revealed that the nucleolus is a very densly packed compartment that cannot thus be directly imaged and segmented. Instead, Dr. Siddarth Narasimhan is attempting to visualise Pol I transcription units in an *ex vivo* preparation based on the Miller tree preparation approach. In both cases, the Pol I structure itself will help annotate larger assemblies.

There is also potential biomedical use of the human Pol I structures I obtained. Pol I is an important cancer drug target. High-resolution structures can be used for drug design, for example, via fragment screening. Additionally, there are small-molecule compounds (discussed in section 1.2.6) in development, such as CX-5461 or BMH-12, whose mode of action is not yet understood. Studies using human Pol I might help shed light on their activity. Finally, purified human Pol I could be used in high-throughput drug screens.

5.3 Role of human Pol I and Pol III in developmental disorders

Mutations in Pol I and Pol III are the underlying cause of some complex developmental disorders. Their exact role in the aetiology of the diseases is not yet understood. It is particularly puzzling how some mutations found in the subunits shared between Pol I and Pol III can give rise to two different disorders: Treacher Collins Syndrome (TCS) and Hypomyelinating Leukodystrophy (HLD). By identifying other proteins affected by mutations related to those disorders, it can be speculated that they arise from the perturbations in Pol I and Pol III, respectively, even though both RNAPs would carry a particular mutation. This theory is supported by the differences in the structures of the human Pol I and Pol III and some studies based on the overexpression of the subunit with a mutation [236]. Yet further studies are needed to confirm the impact of the mutations on the Pol I and Pol III function.

First, I aimed to introduce the mutations of interest into the endogenous subunit in HEK293 cells. Studies of such cell lines will allow approximating the effect of a mutation on a cellular level. Yet, HEK293 cells are not a perfect model, being an immortalised cell line. Most observed disease symptoms are related to neuronal or craniofacial development issues. HEK293 cells are derived from embryonic cells transformed with an adenovirus. It has also been shown that they exhibit some features typical for neuronal tissue [339]. As such, among the established cell lines, HEK293 cells are a good approximation for the tissue in which malfunctions are observed. Yet, an even better approximation would come from studies of stem cells differentiated into neural progenitor cells and neurons. Using such cultures, specific stages along embryonic development affected by the disease could be studied. Another opportunity to explore the role of mutations affecting Pol I and Pol III would come from mouse models. The first mouse model of HLD is based on an introduction of W671R and G672E mutations in the RPC1 subunit, which recapitulates some of the clinical features of the disease [340]. The mouse model for TCS is still missing, as are any models aiming to study the mutations found in the subunits shared between Pol I and Pol III.

Given that most mutations affecting Pol I and Pol III are developmental, treatment opportunities are vastly limited. Yet the broad spectrum of symptoms associated with the disorders raises even more problems in patients' care. With current knowledge, it is impossible to predict the severity or, in some cases, the type of symptoms the patient will experience. With children commonly diagnosed very young, lack of access to such information makes planning medical care extremely

difficult. Additionally to all described issues, more novel mutations are continuously being described in new patients. The development of a tool to even approximately estimate the severity of the disease is thus of burning importance. Either an *in silico* model or a cell-based assay which could estimate the issues a patient might be facing would significantly ease the burden of care. Yet to develop such a model, considerably more information regarding the role of the mutations in the development of both TCS and HLD is needed.

Chapter IV: Materials and Methods

MATERIALS AND METHODS

6.1 Molecular Biology

6.1.1 Restriction-free cloning

Protocol for the restriction-free cloning method is based on the method developed by Van Den Ent et al. [341], which was further adapted for use in the laboratory by Dr. Helga Grötsch. The method allows for the insertion of any target sequence into any plasmid. It relies on two subsequent polymerase chain reactions (PCR).

A set of two primers was used to perform restriction-free cloning. Each primer has two parts: one annealing to the insert and one annealing to the vector. The melting temperature (T_m) of the section annealing to the insert equalled 57°C, while the section annealing to the target vector equalled 64°C.

In the first PCR reaction, plasmid or cDNA with the desired insert is used as a template with PCR mix shown in Table 6.1 and PCR conditions in Table 6.2.

Table 6.1 PCR mix components for the first PCR reaction in restriction-free cloning.

Component	Amount
5x High-Fidelity PCR Green buffer (ThermoFisher)	10 µl
dNTPs (10 mM)	2 µl
Template DNA	20 ng
Primers 1 + 2 (20 pmol/µl)	1.25 µl each
Phusion polymerase	0.5 µl
Water	Up to 50 µl

Table 6.2 PCR settings for the first PCR reaction in restriction-free cloning. 35 rounds of amplification were used.

Stage	Temperature	Duration
Initial Denaturation	98°C	30 s
Denaturation	98°C	30 s
Annealing	55°C	30 s
Extension	72°C	20-30 s/kb
Final Extension	72°C	double Extension

PCR products were run on 1% agarose gel cast in 1x TBE with 5 µl of SYBR Safe DNA stain (Life Technologies) added. The gel was run at 100 V for 30-45 min, depending on the size of the product. Visible DNA fragments were cut out with a clean blade and gel purified using MinElute Gel Extraction Kit (Qiagen) according to the manufacturers' instructions.

The purified DNA fragment was used as a mega primer for the second PCR reaction, with PCR mix components as specified in Table 6.3 and PCR conditions in Table 6.4.

Table 6.3 PCR mix components for the second PCR reaction in restriction-free cloning.

Component	Amount
5x High-Fidelity PCR Green buffer (ThermoFisher)	10 µl
dNTPs (10 mM)	2 µl
Vector DNA	20 ng
Mega primer	100 ng
Phusion polymerase	0.5 µl
Water	Up to 50 µl

Table 6.4 PCR settings for the second PCR reaction in restriction-free cloning. 35 rounds of amplification were used.

Stage	Temperature	Duration
Initial Denaturation	98°C	30 s
Denaturation	98°C	30 s
Annealing	55°C	30 s
Extension	72°C	1 min/kb, >7 kb: 1.5 min/kb
Final Extension	72°C	double Extension

PCR product was digested with DpnI Fast Digest enzyme (NEB) for 2 h to remove the remainder of the starting vector. It was then transformed into XL1-Blue chemocompetent cells. The resulting colonies were isolated and grown overnight in 5 ml of LB medium with appropriate antibiotics. Plasmid DNA was extracted with a Miniprep kit (ThermoFisher) according to the manufacturers' instructions. Integration of the insert was checked using Sanger sequencing (Eurofins Genomics).

6.1.2 Mutagenesis via restriction-free cloning method

To use restriction-free cloning for the introduction of point mutations, only the second PCR reaction is performed. A set of two complementary to each other primers was designed with $T_m = 64°C$ on each side of the desired mutation. All settings described in section 6.1.1 were used, using 100 ng of each mutagenesis primer instead of the mega primer.

6.1.3 biGBac cloning

Protocol for the biGBac cloning method is based on the method developed by Weissmann *et al.*, [300], which was further adapted for use in the laboratory by Dr. Mathias Girbig. The method allows the integration of many genes into one expression vector. All required primers and vectors were already present in the laboratory. Constructs I have cloned are designed for expression in mammalian cells.

As a first step, all genes to be assembled were cloned into a plasmid (here pcDNA3) containing the desired promoter (here CMV) and the polyadenylation signal (here bGH polyadenylation signal) using restriction-free cloning described in section 6.1.2. Given the subsequent steps, all SwaI and PmeI restriction enzyme target sites were mutated.

Up to 5 genes can be combined into a pBig1 plasmid (in this thesis, only up to 3 genes were successfully integrated into one plasmid using this procedure). Each gene to be inserted into the pBig1 plasmid was amplified with two biGBac primers with appropriate overhangs matching repeat sequences in the plasmid. Care must be taken that regardless of the total number of genes combined, the first primer is complementary to the alpha site and the last primer to the omega site on the target pBig1 plasmid. Inserts were amplified using the PCR mix specified in Table 6.5 and the PCR settings listed in Table 6.6.

Table 6.5 PCR mix components for the insert amplification for cloning into pBig1.

Component	Amount
5x High-Fidelity PCR Green buffer (ThermoFisher)	10 µl
dNTPs (10 mM)	1 µl
Template plasmid	5 ng
Primers 1 + 2 (20 pmol/µl)	1.25 µl each
Phusion polymerase	0.5 µl
Water	Up to 50 µl

Table 6.6 PCR settings for the insert amplification for cloning into pBig1. 35 rounds of amplification were used.

Stage	Temperature	Duration
Initial Denaturation	98°C	30 s
Denaturation	98°C	30 s
Annealing	69°C	30 s
Extension	72°C	15-30 s/kb
Final Extension	72°C	10 min

1 µl of DpnI Fast Digest enzyme (NEB) was added and the sample was incubated for 1 h at 37°C before inactivation for 20 min at 80°C. The resulting product was run on 1% agarose gel and gel purified as described in section 6.1.1 for the restriction-free cloning method.

Target pBig1 plasmid was linearised by SwaI restriction enzyme digestion. 1000 ng of the pBig plasmid were mixed with 1 µl of SwaI (NEB) enzyme and 1 µl of 10x NEB buffer O in a total volume of 10 µl and incubated overnight at 30°C. SwaI was inactivated by 20 min incubation at 65°C. Linearised plasmid was run on a 1% agarose gel with an uncut plasmid used as a control (linearised plasmid runs slightly

higher on the agarose gel than the uncut vector). It was then gel purified using MinElute Gel Extraction Kit (Qiagen) according to the manufacturers' instructions.

Amplified inserts were inserted into the linearised pBig1 plasmid using Gibson assembly. 100 ng of linearised pBig1 plasmid was mixed with 5x molar excess of each insert. In the case of POLR1B and POLR2K genes cloned into the pBig1e backbone, a 2:1 ratio of vector to insert was used for the POLR1B gene and a 1:3 ratio was used for the POLR2K gene. 5 µl of the DNA mix was added to 15 µl of the Gibson reaction mix [342] on ice and incubated at 50°C for 1 h. The mixture was transformed into NEB10Beta chemical competent cells and plated on gentamycin-containing plates for selection. After overnight incubation, colonies were picked, grown in a 5 ml starting culture and plasmid DNA was extracted using a Miniprep kit (ThermoFisher) according to the manufacturers' instructions. A diagnostic digest was performed to confirm the integration of the inserts into the pBig1 backbone. 1 µl of plasmid DNA was mixed with 0.5 µl of SwaI restriction enzyme, 1 µl of 10x NEB buffer O with 7.5 µl of water added. Digestion was performed overnight at 30°C. Products were run on the 1% agarose gel and obtained pattern was compared with the expected one from *in silico* digestion simulation. Plasmids with correct band patterns were sequenced using Sanger sequencing (Eurofins Genomics).

To combine more genes into a larger construct, pBig1 plasmids can be used to release the multigene cassettes cloned into them and they can be further combined into pBig2 plasmids. The appropriate pBig2 plasmid with the cloning sites corresponding to the number of pBig1 plasmids being combined was linearised. 1 µg of stock pBig2 plasmid was mixed with 1 µl of PmeI restriction enzyme (NEB) with 1 µl of 10x Cut Smart (NEB) in a total volume of 10 µl and incubated overnight at 37°C. The product was run on a gel and purified as in the case of the pBig1 plasmids linearised with SwaI.

To assemble inserts into a multigene construct, linearised pBig2 plasmid is mixed with pBig1 plasmids. For assembly of SL1 constructs, a 1:5 vector to insert ratio was used, while for the human Pol I, a 1:2.5 ratio worked best. 33 ng of purified, linearised pBig2 were used and amounts of pBig1 plasmids were adjusted accordingly. DNA mix was diluted in water to a total of 12.5 µl, to which 4 µl of 5x isothermal reaction buffer (IRB) was added. 1 µl of PmeI restriction enzyme was added, and the reaction was allowed to proceed for 2 h at 37°C to release the multigene cassettes from the pBig1 plasmids. Subsequently, 2.5 µl of Gibson Enzyme mix (8 µl of Taq DNA ligase,

1 µl of T5 exonuclease 1:30 pre-diluted in 1x IRB and 1 µl of Phusion DNA polymerase) was added to the reaction on ice before the incubation for 1 h at 50°C. The whole mix was transformed into NEB10Beta chemical competent cells according to the manufacturers' instructions. Cells were plated on agar plates containing chloramphenicol for selection. After overnight incubation at 37°C, colonies were harvested, grown and plasmid DNA was extracted to be subjected to the diagnostic digest with SwaI as in the case of the pBig1 constructs validation. Plasmids with the correct band pattern in the diagnostic digest were sent for whole-plasmid sequencing (MGH CCIB DNA Core).

6.1.4 Western Blotting

Semi-dry western blotting (WB) was performed using established laboratory protocol. When whole-cell lysate was used as the input, around 2 x 10^6 cells were collected. They were spun down at 500 rpm for 5 min in a tabletop centrifuge, the supernatant was removed, and cells were once washed in PBS. Subsequently, the pellet was resuspended in 100 µl of cold WB lysis buffer (40 mM Tris pH 7.4, 1 mM EDTA, 10% glycerol, 10 mM DTT, 1% Triton-X, 0.08% SDS, cOmplete protease inhibitor cocktail (ThermoFisher)) and applied to the QiaShredder columns (Qiagen). Samples were spun down at 14,000 rpm for 2 min. 80 µl of the resulting lysate or sample from the purification was mixed with 20 µl of 5x SDS loading dye and boiled. Commercial SDS-PAGE gel (ThermoFisher) was then run in MES buffer for 35 min at 200 V.

The gel was then transferred to the semi-dry blotting transfer machine. The layers assembled included: thick Whatman paper, PVDF membrane, SDS-PAGE gel with sample and thick Whatman paper. The Whatman paper and membrane were wetted with the WB transfer buffer (38.6 mM glycine, 47.9 mM Tris, 20% MetOH) before assembling the layers. The blot was run at 15 V for 75 min. After transfer, the membrane was stained with Ponceau S stain to note whether the transfer was successful and protein concentrations comparable. The membrane was then washed with PBS until the stain was removed, and the membrane was then washed with water.

All subsequent steps were done in WB wash buffer: 4% skim milk, 0.1% Tween in PBS. The membrane was washed with WB wash buffer for 30 min, and then the primary antibody was added and incubated for 1 h at room temperature or overnight at 4°C. Anti-mCherry antibody (Abcam, ab167453) was used in 1:2000 dilution,

anti-RPA1 (Santa Cruz, sc-48385) in 1:2000, anti-GAPDH (Cusabio, CSB-MA000195) in 1:5000 and anti-Halo (Promega, G928A) in 1:2000. The membrane was then washed for 30 min with WB wash buffer exchanged three times. A secondary antibody was then applied and incubated for 1 h. All primary antibodies come from rabbit except for the anti-RPA1, a mouse antibody. Anti-rabbit IgG antibody (ThermoFisher, G-21234) was used in 1:20000 dilution while the anti-mouse IgG antibody (ThermoFisher, G-31430) was used in 1:5000 dilution. The membrane was then washed for 1 h, with the WB wash buffer exchanged 4 times before being transferred into PBS. Chemiluminescence was detected using Pierce ECL Western Blotting Substrate (ThermoFisher) used according to the manufacturers' instruction and recorded with BioRad ChemiDoc Imaging System.

6.1.5 Mammalian cell culture

Adherent HEK293T cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) with 4.5 g/L glucose supplemented with 10% Fetal Bovine Serum (FBS) and 1% L-glutamine in 37°C incubator at 5% CO_2. They were usually grown to confluency and split every 2-3 days at 1:20 into fresh medium. Cells were grown without any antibiotics except for selection with blasticidin S and supplementation with Penicillin-Streptomycin (Pen-Strep) during the introduction of point mutations.

Suspension HEK293T cells were grown in Expi293F medium (ThermoFisher) up to a density of 7 x 10^6 cells/ml. They were seeded at a density between 0.3-0.5 x 10^6 cells/ml. Cells were incubated at 37°C, 80% humidity 8% CO_2 and shaking at 125 rpm.

All cell lines were regularly tested for mycoplasma contamination.

6.2 CRISPR-Cas9

I developed the CRISPR-Cas9 tagging protocol based on the one published by Koch et al. [284] and advice provided by Dr. Francois Dossin. Dr. Helga Grötsch supported cell culture work. The protocol for introducing mutations was developed by Matteo Trovato [336,343]. The introduction of point mutations was partially done by Granita Lokaj under my supervision.

6.2.1 crRNA and donor template for tag insertion

Once the site where either the insert or the point mutation is to be introduced, the sequence of the genomic locus was found, here using hek293genome.org [290] since only HEK293 cells were used. Then, using an online tool CRISPOR [285] suitable crRNAs were found. Ideally, they should be no further than 10-20 bp from the desired cut size. Also, automatically provided statistics such as off-target rate were used to pick the three best crRNAs. The sequence of the crRNA omitting the PAM was selected and overhangs which simulate BbsI restriction enzyme products were added in the correct orientation, assuming that the PAM site should be on the 3' end. The general pattern used included (with N standing for any base within the crRNA sequence):

Sense crRNA:	caccgNNNNNNNNNNNNNNNNNNNN
Antisense crRNA:	aaacNNNNNNNNNNNNNNNNNNNc

crRNAs were ordered as standard primers from Sigma-Aldrich. They were then cloned into the pSpCas9(BB)-2A_GFP (Addgene PX458) vector, containing a sgRNA scaffold sequence and Cas9 expression cassette. 2.5 µg of target vector was digested with 3 µl of BpiI Fast Digest enzyme (Thermo Fisher) with 6 µl of 5X Green Fast Digest buffer in a total volume of 60 µl for 1 h at 37°C. Subsequently, primers were annealed by mixing 2 µg of each primer in the total volume of 50 µl of annealing buffer (10mM Tris, pH 7.5-8.0, 50 mM NaCl) and heating to 95°C for 4 min. Primers were then cooled to room temperature on the bench for 45 min. 5 µl of the annealed primers were then diluted with nuclease-free water to 50 µl total. The ligation reaction was then set up with 7.5 µl of diluted, annealed oligos, 0.5 µl of linearised PX458 vector, 1 µl of 10x ligase buffer (Thermo Fisher), 0.5 µl of T4 ligase (Thermo Fisher) in the total volume of 10 µl. Ligation was performed overnight at room temperature. T4 ligase was then inactivated by 10 min incubation at 65°C. The mixture was then cooled on ice and transformed into XL1-Blue cells. The ready

plasmid was purified using a Miniprep kit (ThermoFisher) according to the manufacturers' instructions. Integration of the insert was checked using Sanger sequencing (Eurofins Genomics).

The donor plasmid was designed to have 700 bp of homology arms flanking from each side of the desired insert. The tag was inserted directly in front of the STOP codon. The donor plasmid was synthesised using GeneScript gene synthesis service into the pUC57-Mini plasmid into the EcoRV site, keeping the unique EcoRV sites flanking the donor sequence.

Both PX458 plasmid with sgRNA and Cas9 and the donor plasmid, were Mira prepped before transfection to obtain higher DNA yields [344].

6.2.2 T7 endonuclease I assay

To test which crRNA allows the most efficient cutting, I performed the T7 endonuclease I assay. Cells were transfected using the standard protocol outlined in section 6.2.3 with 10.5 µg of plasmid DNA containing sgRNA and Cas9.

72 h post-transfection, 5 x 10^6 cells were harvested, and genomic DNA was extracted using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma). A primer pair was designed to amplify a fragment around the region of interest at a total size of 1 000 bp, with the cut site located not directly in the middle of the amplified fragment, instead dividing it into 300 and 700 bp fragments. The PCR reaction was then performed with the PCR mix shown in Table 6.7 and the PCR conditions in Table 6.8.

Table 6.7 PCR mix components for the T7 endonuclease I assay.

Component	Amount
5x HotStar PCR buffer (ThermoFisher)	4 µl
dNTPs (10 mM)	2 µl
Genomic DNA	200 ng
Primers 1 + 2 (10 µM)	2 µl each
Phusion polymerase	1 µl
Water	Up to 20 µl

Table 6.8 PCR settings for the T7 endonuclease I assay. 35 rounds of amplification were used.

Stage	Temperature	Time
Initial Denaturation	95°C	5 min
Denaturation	95°C	30 s
Annealing	70°C	45 s
Extension	72°C	1 min
Final Extension	72°C	10 min

5 µl of the PCR product was kept for further comparison. 15 µl of the product was heated to 95°C for 5 min and then cooled down on the bench to denature and reanneal

the DNA. It was then supplemented with 1.8 µl of 10x buffer 2 (NEB) and 1.2 µl of T7 endonuclease I (NEB). The reaction mixture was incubated at 37°C for 30 min. Products were run on a 2% agarose gel alongside the previously set aside uncut PCR product. The best performing crRNA was then selected based on the presence of characteristic 300 and 700 bp bands (detailed in section 2.1.1). If the chosen crRNA is fully present in the donor sequence (it does not span the stop codon directly), the PAM sequence in the donor plasmid needs to be mutated to avoid cutting the donor plasmid. This was done in the case of POLR1B tagging.

6.2.3 Transfection with Cas9-crRNA and donor plasmids

Cells were split 1:5 from confluent 24 hours before transfection to achieve c.a. 85% confluency on the day of transfection. In total, 10.5 µg of DNA (5.25 µg of each donor plasmid and plasmid with sgRNA-Cas9) was mixed with 500 µl of pre-warmed Opti-MEM (Gibco) (with no DNA added for the mock transfection). 21 µl of polyethylenimine (PEI) was also mixed into 500 µl of Opti-MEM. DNA and PEI-containing aliquots were joined by pipetting and left to stand at room temperature for 15 min. From the plate to be transfected medium was aspirated, cells were washed once with PBS, and then the DNA-PEI mixture in Opti-MEM was added dropwise to the plate. Cells were covered with 10 ml of complete medium and incubated overnight. The medium was exchanged the next day and cells were allowed to recover.

6.2.4 Clone selection by antibiotic selection and hand-picking

After cells recovered from transfection (usually 3-5 days) antibiotic selection can be started. 10 µg/ml of blasticidin S (Thermo Fisher) was used and maintained until the mock-transfected cells were no longer viable. Cells that were transfected and are still viable were then seeded on a 15 cm dish at low density (usually two concentrations, one at 1-3 x 10^4 cells/ml and one at 1-3 x 10^3 cells/ml were used) in 13 mL total. They were grown for 2 days to allow them to form colonies derived from single cells. The medium was then aspirated, and cells were overlaid with PBS. Monoclonal colonies were then picked by hand using a pipette into a round-bottom 96-well plate filled with 30 µl of trypsin. Trypsin was then neutralised with 170 µl of complete medium. Each plate was then split into two identical replicates to expand one and use the other for genotyping. Each plate was topped with 100 µl of medium up to 200 µl total in each well.

6.2.5 Genotyping

When clones reach at least 50% confluency (3-5 days after picking monoclonal clusters), they can be used for genotyping. The medium was aspirated and the plate was placed briefly in a freezer to help with lysis. Subsequently 50 µl of genotyping buffer (100 mM Tris pH 7.5, 200 mM NaCl, 0.5% Tween, and freshly supplemented 1:100 with proteinase K (20 mg/mL) (Roche)) was added to each well. The plate was sealed and incubated overnight at 56°C. 100 µl of water was added to each well, and the plate was incubated for 15 min at 95°C to denature proteinase K. PCR reaction was run using a PCR master mix composition as described in Table 6.9 and with settings collected in Table 6.10. DNA was added, keeping the well pattern. Both junction and zygosity PCR were performed with primer pairs designed as indicated in section 2.1.3. PCR products were run on 1% agarose gel fitting 96 samples for 20 min at 120 V.

Table 6.9 PCR master mix components for 1 plate of genotyping PCR. Topped up with 3 µl of DNA from each well.

Component	Amount
5x PCR green buffer (ThermoFisher)	408 µl
dNTPs (10 mM)	81 µl
$MgCl_2$	326 µl
Primer mix (20 µl of each primer (100 ng/µl) + 160 µl water)	163 µl
Phusion polymerase	15 µl
Water	744 µl

Table 6.8 PCR settings for the genotyping PCR. 38 rounds of amplification were used.

Stage	Temperature	Duration
Initial Denaturation	95°C	3 min
Denaturation	95°C	30 s
Annealing	66°C	30 s
Extension	68°C	3.5 min
Final Extension	68°C	10 min

6.2.6 Clone validation

Based on genotyping PCR, clones that contained a homozygous tag insertion were expanded. They were harvested and their genomic DNA was extracted using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma). Junction and zygosity PCR were run using PCR mix components as specified in Table 6.11 and PCR settings from Table 6.12. PCR products were run on 1% agarose gel and the band pattern was compared with the expected one for the homozygous or heterozygous insertion of the tag.

Table 6.11 PCR components for junction and zygosity PCR of selected clones.

Component	Amount
2x Q5 Master Mix (NEB)	7.5 µl
Forward primer (10 µM)	1.5 µl
Reverse primer (10 µM)	1.5 µl
DNA	150 ng
Water	Up to 15 µl

Table 6.12 PCR settings for junction and zygosity PCR of selected clones. 35 rounds of amplification were used.

Stage	Temperature	Time
Initial Denaturation	98°C	30 s
Denaturation	95°C	30 s
Annealing	65°C	30 s
Extension	68°C	3.5 min
Final Extension	68°C	7 min

The presence of the correctly expressed tag was also validated with WB using a primary antibody against the tag component was performed as described in section 6.1.4. Correct localisation of the tagged protein was confirmed with confocal microscopy performed as described in section 6.7.2.

6.2.7 Adaptation to growth in suspension

The introduction of the tag was performed in adherent HEK293T cells. Yet, for efficient purification, higher cell mass was needed; thus, the cells had to be adapted to grow in suspension. 5 ml of HEK293T cells grown in adherent culture in DMEM medium were added to 15 ml of Expi293 medium (Thermo Fisher) and placed on a shaker. After 5 days they were mixed 1:1 in 30 ml total with fresh Expi293 medium and resuspended by pipetting. Subsequently, they are passaged as normal suspension cells every 2 days with more pipetting to resuspend any cell clumps. After 3 – 4 weeks, the viability of the cells reached normal levels of Expi293 cells in suspension at 97-100% alive.

6.2.8 crRNA and donor design for the introduction of point mutations

crRNA was designed the same as for the introduction of inserts, but only one, the best crRNA, was chosen especially taking care for it to be close to the mutation site. A donor template with 35 bp on each side of the point mutation was also designed. They were both ordered from IDT, with crRNA being a part of a two-part RNA system together with a separate tracrRNA tagged with a fluorescent ATTO 550 label. The donor template was designed to have a 5' biotin label, which could bind to the monomeric streptavidin tag present on the Cas9 protein produced by the EMBL Protein Expression and Purification Core Facility.

6.2.9 Introduction of point mutations with Neon electroporation

The day before transfection, cells were split 1:3 from confluent plates. On the day of the transfection, the sgRNA complex was prepared by resuspending crRNA and tracrRNA in 10 µl of Nuclease Free Duplex buffer (IDT) and mixing 3 µl of each of them together, followed by heating to 95°C for 5 min and cooling on the bench for 20 min. For each transfection reaction, 20 pmol of Cas9 and 30 pmol of annealed sgRNA were mixed with 9 µl of Buffer R (IDT) and incubated for 20 min at room temperature. Then single-stranded DNA oligo (ssODN), acting as the repair template, was resuspended to a final concentration of 100 µM in resuspension buffer, and 30 pmol were added to the Cas9-sgRNA complex to be incubated for 20 min at room temperature. For each mutation, 6 samples were prepared: 4 samples to be transfected and 2 control samples.

Cells were harvested and pulled together. For each reaction, 1 x 10^6 cells were taken. They were spun down and washed with PBS and resuspended in the Cas9-sgRNA-ssODN mixture with the addition of the AltR-Electroporation Enhancer (IDT). Electroporation was done using Neon device set up according to manufacturers' instructions with settings of 1150 V, 20 ms, 2 pulses, which are the settings recommended for HEK293 cells [343]. After electroporation, cells were set at 37°C for 15 min to recover. They were then washed once with DMEM and resuspended in a pre-conditioned medium (1/3 conditioned medium and 2/3 fresh medium) with the addition of the HDR enhancer (IDT).

6.2.10 Clone selection by FACS

Cells were FACS sorted 48 h post-transfection at the EMBL Flow Cytometry Core Facility. Cells from all four transfected reactions were combined (same for the two controls), spun down by centrifugation at 500 rpm for 5 min and resuspended in 500 µl of the sorting medium (2% BSA and 2.5 mM EDTA in PBS) before straining with 40 µm cell strainer (Falcon BD). Cell viability dye was added to the sample, and cells were sorted using ATTO 550 fluorescence to enrich for the cells that were successfully transfected. Cells were grown for 2-3 weeks, and then surviving clones were used for further analysis.

6.2.11 Mismatch real-time quantitative PCR (RT-qPCR)

Mismatch RT-qPCR, which will allow detection of the insertion of point mutation, relies on primers complementary to the mutant or WT sequence. Primers were designed to have a T_m = 60-63°C, and a primer pair gave an amplification product of about 120 bp. The 3' end of one of the primers in the pair is annealing to the mutation site and is complementary to either the WT sequence or mutated one. Two primer pairs for each mutation (with either forward or reverse one annealing to the mutation site) were designed, and the better primer pair was chosen based on the primer melting curve.

The medium was aspirated from wells containing cells. 70 µl of direct lysis solution (Viagen Biotech) diluted 1:3 with water and supplemented with proteinase K to a final concentration of 0.5 mg/ml were added to each well, and then sample was transferred to a 96-well PCR plate. Samples were incubated for 2 h at 55°C followed by 45 min at 85°C and stored at 4°C. For each reaction, 1 µl of the primer mix (each at 10 µM), 10 µl of SYBRgreen Master Mix (Thermo Fisher) and 0.5 µl of cell lysate in a total of 20 µl were used. Non-transfected and bulk sorted positive cells were used as a negative and positive control, respectively. The RT-qPCR run was performed with the settings outlined in Tables 6.13 and 6.14.

Table 6.12 RT-qPCR settings used for mismatch PCR. 40 cycles were used.

Stage	Temperature	Time
Enzyme activation	98°C	30 s
Denaturation	95°C	30 s
Annealing/ Extension	65°C	30 s

Table 6.13 Dissociation performed at the end of the cycle to obtain primer melting curve.

Stage	Ramp rate	Temp.	Time
1	1.6°C/s	95°C	15 s
2	1.6°C/s	60°C	60 s
Dissociation	0.075°C/s	95°C	15 s

The cycle threshold (Ct) values were detected, and for each sample, the Ct value obtained with WT complementary and mutation complementary primer were compared. Ct_mutation was subtracted from the Ct_WT, and a positive value indicated that the point mutation is likely to present (see section 4.2.2). For each mutation, 3-5 clones were selected and expanded. Cells were then harvested, and their genomic DNA extracted. A region of about 1 kb around the mutation site was amplified, and the insertion of the mutation was validated using Sanger sequencing (Eurofins Genomics).

6.3 Protein expression and purification

6.3.1 Endogenous human Pol I and III purification

Components of all buffers are listed in Table 6.14. Purification for Pol I and Pol III follows the same steps, only with differences in buffer composition.

About 3 L of cells with tag inserted on the C-terminus of the RPAC1 subunit at 7 x 10^6 cells/ml were pelleted, giving in total about 50 g, washed with PBS and flash frozen. The cell pellet was resuspended in 300-350 ml of Lysis Buffer by rolling at 100 rpm and later stirring at 4°C. Cells were ruptured by sonication using a Branson Sonifier 250 with 7 cycles of 45 s turned on and 15 s turned off with output power set to 0.5 and duty cycle set to 0.5. The lysate was ultracentrifuged for 1 h at 45 000 rpm at 4°C. The supernatant was then collected and filtered using Whatman 1.2 µm filters. The sample was loaded on an equilibrated 5 ml HP HisTrap column (Merck) on an Äkta system. The column was washed with 20 column volumes of the Ni-NTA Buffer A and the protein of interest was eluted with a step elution using Ni-NTA Elution Buffer B. The first 8-10 ml usually contained the protein of interest. Ni-NTA elution was mixed with 3 ml of equilibrated Strep-Tactin beads slurry (IBA Lifesciences). The sample was incubated at 4°C while rotating at 100 rpm for 1 h. The sample mixed with Strep-Tactin beads was applied to a gravity column, washed with 50 ml of Strep-Tactin Wash Buffer and eluted in three fractions of 2 ml with Strep-Tactin Elution Buffer. Collected fractions were pulled and applied to an equilibrated MiniQ 4.6 50 PE column (Cytiva) running on an Äkta Micro system. The column was washed with 10 column volumes of MonoQ Buffer A and the proteins were eluted with a gradient over 40 column volumes. For Pol I purification, first a step of 25% MonoQ Buffer B over 3 column volumes was introduced, followed by a gradient over 40 column volumes. Proteins eluted in several peaks as outlined in sections 2.2.2 and 2.2.3. Fractions coming from desired peaks were pulled and concentrated using a 100 K 0.5 ml Amicon concentrator (Merck). The sample was then buffer exchanged on the concentrator by bringing the sample volume below 50 µl and resuspending in 0.5 ml of Buffer Exchange three times. The obtained sample was concentrated and used fresh for cryo-EM sample preparation or flash-frozen in liquid nitrogen for later biochemical assays.

Table 6.14 Buffer composition used for human Pol I and Pol III purification.

Pol I		Pol III	
Component	**Concentration**	**Component**	**Concentration**
Lysis Buffer			
HEPES (pH 8)	25 mM	HEPES (pH 8)	25 mM
$(NH_4)_2SO_4$	150 mM	$(NH_4)_2SO_4$	150 mM
$MgCl_2$	5 mM	$MgCl_2$	5 mM
$ZnCl_2$	10 uM	$ZnCl_2$	10 uM
glycerol	5%	glycerol	5%
Triton-X100	0.50%	Triton-X100	0.50%
imidazole	20 mM	imidazole	20 mM
2-Mercaptoethanol	2 mM	2-Mercaptoethanol	2 mM
protease inhibitors	1 tab / 50 ml	protease inhibitors	1 tab / 50 ml
benzonaze	0.05 µl/ml	benzonaze	0.05 µl/ml
Ni-NTA Buffer A			
HEPES (pH 8)	25 mM	HEPES (pH 8)	25 mM
$(NH_4)_2SO_4$	150 mM	$(NH_4)_2SO_4$	150 mM
$MgCl_2$	5 mM	$MgCl_2$	5 mM
glycerol	5%	glycerol	5%
imidazole	20 mM	imidazole	20 mM
2-Mercaptoethanol	2 mM	2-Mercaptoethanol	2 mM
Ni-NTA Elution Buffer B			
HEPES (pH 8)	25 mM	HEPES (pH 8)	25 mM
$(NH_4)_2SO_4$	150 mM	$(NH_4)_2SO_4$	150 mM
$MgCl_2$	5 mM	$MgCl_2$	5 mM
glycerol	5%	glycerol	5%
imidazole	300 mM	imidazole	300 mM
2-Mercaptoethanol	2 mM	2-Mercaptoethanol	2 mM
Strep-Tactin Wash Buffer			
HEPES (pH 8)	25 mM	HEPES (pH 7.4)	25 mM
CH_3CO_2K	200 mM	$(NH_4)_2SO_4$	150 mM
$MgCl_2$	5 mM	$MgCl_2$	5 mM
glycerol	5%	glycerol	5%
2-Mercaptoethanol	2 mM	2-Mercaptoethanol	2 mM
Strep-Tactin Elution Buffer			
biotin	20 mM	biotin	20 mM
HEPES (pH 8)	25 mM	HEPES (pH 7.4)	25 mM
CH_3CO_2K	200 mM	$(NH_4)_2SO_4$	100 mM
$MgCl_2$	5 mM	$MgCl_2$	5 mM
glycerol	5%	glycerol	5%
2-Mercaptoethanol	2 mM	2-Mercaptoethanol	2 mM
MonoQ Buffer A			
HEPES (pH 5)	HEPES (pH 5)	HEPES (pH 5)	HEPES (pH 5)
CH_3CO_2K	CH_3CO_2K	CH_3CO_2K	CH_3CO_2K
$MgCl_2$	$MgCl_2$	$MgCl_2$	$MgCl_2$
glycerol	glycerol	glycerol	glycerol
DTT	DTT	DTT	DTT

Table 6.14 (continued)

Pol I		Pol III	
Component	**Concentration**	**Component**	**Concentration**
MonoQ Buffer B			
HEPES (pH 7.4)	25 mM	HEPES (pH 7.4)	25 mM
CH_3CO_2K	2 M	$(NH_4)_2SO_4$	1 M
MgCl2	5 mM	$MgCl_2$	5 mM
glycerol	5%	glycerol	5%
DTT	5mM	DTT	10mM
Buffer Exchange			
HEPES pH7.5	15mM	HEPES pH7.5	15mM
$(NH_4)_2SO_4$	80mM	$(NH_4)_2SO_4$	150mM
$MgCl_2$	5mM	$MgCl_2$	5mM
DTT	10mM	DTT	10mM

6.3.2 RRN3 expression and purification

Dr. Brice Murciano performed RRN3 expression and purification. Human RRN3 was cloned with an N-terminal His tag into pETM11 plasmid (EMBL Protein Expression and Purification Core Facility). It was expressed in the *E. coli* LOBSTR strain (Kerafast) with an overnight growth in TB medium at 18°C. Expression was induced at $OD_{260nm} = 0.8–1.0$ using 0.05 mM of IPTG and the pellet was collected. All buffers used in the purification are outlined in Table 6.15. Cells were lysed in the lysis buffer and disrupted using Microfluidizer Processor M-110L (Microfluidics). The lysate was then centrifuged and supernatant applied to the equilibrated Ni-NTA beads (Qiagen). Sample was incubated for 1 h at 4°C while rolling at 100 rpm. The sample was applied to a gravity column. Beads were washed with Ni-NTA Buffer A and then protein of interest was eluted with Ni-NTA Elution Buffer B. TEV protease was added to the sample to cleave off the His – tag and the sample was dialysed overnight at 4°C into Buffer A. Reverse Ni-NTA was then performed by incubation of the sample for 30 min with Ni-NTA beads and the applying the sample to the gravity column to collect the flow-through. The sample was then applied to a MonoQ column (GE Healthcare), from which after the wash, RRN3 was eluted with a gradient of Buffer B run over 10 column volumes. Peak fractions were collected and concentrated using a 5-kDa concentrator (Corning). The concentrated sample was injected onto the equilibrated in Gel filtration buffer Superdex 200 increase 10/300 GL column (GE Lifesciences). Peak fractions were collected, concentrated and flash-frozen in liquid nitrogen.

Table 6.14 Buffer composition used for recombinant human RRN3 purification.

Component	Concentration
Lysis Buffer	
Tris (pH 7.5)	50 mM
NaCl	200 mM
glycerol	10%
imidazole	10 mM
2-Mercaptoethanol	2 mM
protease inhibitors	1 tab / 50 ml
Dnase 1	
lysozyme	
Ni-NTA Buffer A	
Tris (pH 7.5)	50 mM
NaCl	1 M
glycerol	10%
imidazole	40 mM
2-Mercaptoethanol	2 mM
Ni-NTA Elution Buffer B	
Tris (pH 7.5)	50 mM
NaCl	200 mM
glycerol	10%
imidazole	150 mM
DTT	2 mM
Buffer A	
Tris (pH 7.5)	20 mM
NaCl	200 mM
DTT	2 mM
Buffer B	
Tris (pH 7.5)	20 mM
NaCl	1M
DTT	2 mM
Gel filtration	
Tris (pH 7.5)	25 mM
NaCl	150 mM
DTT	2 mM

6.4 Biochemistry

6.4.1 DNA-RNA scaffold annealing

To assemble a DNA-RNA artificial transcription bubble, template and non-template DNA strands at 100 µM were mixed together and heated to 95°C for 2 min. They were then cooled on ice for 5 min. 2x hybridization buffer (40 mM HEPES pH 7.5, 24 mM MgCl2, 200 mM NaCl, 20 mM DTT) was then added to the sample. RNA at 100 µM was heated to 55°C for 1 min and added to the DNA diluted in hybridisation buffer on ice. The sample was then briefly incubated on ice before bringing it to room temperature by incubation on the bench. Sequences of all primers used in the assembly of nucleic acid scaffolds used in are listed in Table 6.15.

Table 6.15 Nucleic acid templates used in this thesis.

Primer	Sequence (5' – 3')
DNA, template strand	GTACTGAATTAGACAATGCTCTGTGGCTCTAGTACCATGAGCG
DNA, non-template strand	CGCTCATGGTACTAGGCTTCGGAGAAGTTGTCTAATTCAGTAC
RNA, 5' self-complementary	UAUGCAUAACGCCACAGAG
RNA, not self-complementary	UCUGGUAUACGCCACAGAG

6.4.2 *In vitro* primer extension assay

In vitro primer extension assays were performed by Dr. Florence Baudin. RNA primer was radioactively labelled with [γ-32P]ATP (10 mCi/ml, Hartmann Analytic) at 5' end using T4 PNK (NEB). RNA primer was then purified using PAGE. DNA-RNA scaffold was annealed as outlined in section 6.4.1, obtaining a scaffold at 16.67 µM which was then diluted in transcription buffer (20 mM HEPES pH 7.5, 60 mM $(NH_4)_2SO_4$, 10 mM $MgSO_4$, 10% glycerol, 10 mM DTT) to obtain scaffold concentration of 2 µM, of which 1 µl was typically used per reaction. It was mixed with 2.5 µM of Pol I and incubated for 10 min at room temperature. 5 µl of the protein-nucleic acid mix was supplemented with reaction start buffer, which consisted of transcription buffer with the addition of 1.3 mM of each nucleotide used in the reaction. Samples were incubated for 30 min at room temperature and then the reaction was stopped with the addition of formamide buffer. The sample was then incubated at 95°C for 3 min and loaded on a 17% denaturing gel (acrylamide/bis 19:1, 8 M urea, TBE 1%). Radioactivity was captured using a phosphor-imaging screen (Fujifilm) and recorded with Typhon FLA9500.

6.5 Structure determination via cryo-EM

6.5.1 Cryo-EM sample preparation

All cryo-EM grids were made using freshly purified Pol I. Details of the sample preparation conditions of the grids that resulted in the datasets described are collected in Table 6.16. All grids were plasma cleaned using a NanoClean plasma cleaner (Fischione Instruments, Model 1070) for 30 s using a 75%–25% argon–oxygen mixture.

Table 6.16 Grid preparation conditions for the apo Pol I, Pol I EC and Pol I-RRN3 / Pol I OC datasets. BF – blot force, BT – blotting time, WT – wait time.

	Apo Pol I	**Pol I EC**	**Pol I – RRN3 / OC**
Pol I concentration	3.56 mg/ml	0.7 mg/ml	0.85 mg/ml
Nucleic acid scaffold	3x molar excess DNA-RNA scaffold	1.5x molar excess DNA-RNA scaffold	1.5x molar excess DNA scaffold
Co-factors / additives	80 mM CHAPSO	-	Equimolar RRN3
Buffer composition	15 mM HEPES, pH 7.5, 150 mM $(NH_4)_2SO_4$, 10 mM DTT	15 mM HEPES, pH 7.5, 80 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 10 mM DTT	15 mM HEPES, pH 7.5, 80 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 10 mM DTT
Grid type	mesh 200, Cu R2/1 grids (Quantifoil)	mesh 200, Cu R2/1 grids (Quantifoil)	mesh 200, Cu R2/1 grids (Quantifoil)
Vitrobot settings	BF “2”, BT 2 s, WT 0 s humidity = 90%, 4°C	BF “3”, BT 0 s, WT 0 s humidity = 100%, 15°C	BF “3”, BT 0 s, WT 0 s humidity = 100%, 15°C

6.5.2 Cryo-EM data collection

Data collection settings for all datasets described in this thesis are summarised in Table 6.17. All datasets were collected via automatic acquisition using SerialEM [258]. Initial pre-processing and on-the-fly data quality control, especially for high-resolution datasets collected on Titan Krios TEM, were performed with Warp [261].

Table 6.17 Data collection statistics for datasets described in this thesis

	Apo Pol I	Pol I – RRN3/OC	Apo Pol I	Pol I EC	Pol I – RRN3/OC
Magnification	92,000	92,000	105,000	105,000	105,000
Voltage (kV)	200	200	300	300	300
Microscope	FEI Talos Arctica	FEI Talos Arctica	Titan Krios TEM	Titan Krios TEM	Titan Krios TEM
Detector	Falcon III	Falcon III	K3 direct	K3 direct	K3 direct
Electron exposure ($e^-/Å^2$)	38.1	43.4	50.9	67.3	41.3
Defocus range (μm)	0.5-2.5	0.5-2.5	0.6-1.4	1.0-2.5	0.75-2.25
Pixel size (Å)	1.566	1.566	0.822	0.822	0.822
Total no. movies recorded	1,580	967	22,001	10,053	14,224
No. of frames	12	12	38	38	40

6.5.3 Cryo-EM data processing

Details concerning the data processing pipeline are found in section 3 in Figures 3.2, 3.4 and 3.13 for apo Pol I, Pol I EC and Pol I-RRN3/Pol I OC datasets, respectively.

For the datasets acquired on FEI Talos Arctica, data was pre-processed (frame alignment, CTF estimation, dose weighting) using Warp 1.0.7W [261]. Particles were automatically picked using the standard BoxNet2Mask_20180918 network, assuming particles with a diameter of 220 Å. Particles were extracted in 200 px box and imported into cryoSPARC [345]. The 2D classification was performed, best classes were selected and used for *ab initio* classification with one class. The resulting model was refined using Homogenous Refinement (Legacy) and used as a reference after low-pass filtering to 20 Å using EMAN2 [346]. Particles picked by Warp were also imported into RELION 3.1 [263,347] and further processed using a combination of 3D classification and masked 3D classification to assess the heterogeneity of the datasets.

For all datasets acquired on Titan Krios TEM, initial data processing and particle picking were also performed with Warp [261], but particles were extracted with a 280 px box size. In the case of apo Pol I and Pol I-RRN3/Pol I OC datasets, particles were directly imported into RELION 3.1 [263,347] and the corresponding reference obtained from the FEI Talos Arctica dataset was used. In the case of the Pol I EC dataset, particles were pre-processed using cryoSPARC [345] as described for the FEI Talos Arctica datasets to obtain the reference that could be used in RELION 3.1 [263,347] for 3D classification. After importing into RELION, micrographs from all datasets were motion-corrected using RELION's own algorithm MotionCor2 [348] and CTF-corrected using Gctf [349]. To increase the processing speed, particles were binned (2 times for apo Pol I and 4 times for the other two datasets), and in the case of apo Pol I and Pol I-RRN3/Pol I OC datasets were split into 5 subsets for the initial global 3D classification. All classifications performed used T-parameter = 20, which I determined to allow the best separation of states for those datasets. After initial global classification, particles were further classified with different soft masks applied to regions of interest. Masks were made using UCSF Chimera [350] by fitting a structure into the region of interest and using the "molmap" command. The mask was then extended, and a soft edge was added in RELION. After a homogenous population of particles was sorted, particles were unbinned and subjected to CTF refinement, particle polishing and the second round of CTF refinement [347]. The reported resolution of all maps was obtained using the gold-standard Fourier shell correlation (FSC) with the 0.143 cut-off criterion implemented as a post-processing tool in RELION [351]. Local resolution was estimated with RELION's tool.

An additional processing step in the case of the Pol I EC dataset was the use of multibody refinement in RELION [352]. It allowed me to reconstruct two separate bodies, one covering the core of the complex and one encompassing the upper clamp, tWH domain of the RPA49 subunit and the stalk. Especially in the second body, more high-resolution features were visible. Given that I obtained a total of 4 maps corresponding to the Pol I EC structure (Maps C, D, D1 and D2), Phenix 1.18 "combine_focused_maps" tool [353] was used to obtain a composite map which included the highest-resolution features coming from all 4 maps.

The Pol I-RRN3 reconstruction suffered from visible anisotropy, especially in the RRN3 region. To assess the effects of directional resolution anisotropy, the 3DFSC program was used [354]. To correct some of the effects, focused refinement was

performed [352], which did not significantly increase the resolution of the map but helped alleviate some of the visual effects which came from the specimen movement.

6.5.4 Model building and refinement

Maps used for model building were sharpened using several tools such as Autosharpen from Phenix [353] or LocalDeblur [355] and LocScale [356] implemented within the CCP-EM [357] or Scipion [358]. Human Pol I homology model was built using PDB: 4C3I [21] as reference. Human subunits were then aligned either using the human Pol III structure (PDB: 7AE1 [36]) for the shared subunits or using homology models obtained from SWISS-MODEL Repository [359] or Phyre2 [360]. Structures of individual subunits were initially aligned and fitted into Map C using UCSF Chimera [350]. Further, they were manually rebuilt using Coot 0.9 [361]. Originally, homology models were fitter using "rigid body fit", but later ProSMART [362] restraints were generated and used to improve the initial fit to the density. During manual re-building of the models, secondary-structure predictions generated by PSIPRED [363] were used to model regions not covered by the homology models. Model building was supported by Dr. Mathias Girbig, who helped manually verify the final models I created.

Throughout the building process, structures were validated using a real-space refinement tool from Phenix 1.13 [364,365]. Refinement statistics found in Appendix (Tables A4-A6) were calculated using the MolProbity comprehensive validation tool [366]. UCSF ChimeraX [367] was used for the preparation of figures.

6.6 Analysis of disease-associated mutations

6.6.1 *In silico* analysis of mutations

For the analysis of the disease-associated mutations of Pol I and Pol III, EC structures have been used as they have the highest resolution (PDB: 7OB9 [23] and PDB: 7AE1 [36], respectively). The residues in question were first visually inspected in ChimeraX to ensure the high quality of the cryo-EM map. Residues where no cryo-EM density was present (for example, the N-terminus of the RPAC1 subunit for Pol I), have been excluded from further analysis. Only protein parts of the complex have been included in the analysis; thus, the nucleic acids have been removed from the PDB files. Subsequently, files were loaded into the Protein Contacts Atlas [368]. Since both RNAPs are too large to be loaded into the server, I split the PDB file into subdomains and loaded them separately, additionally recovering interaction surfaces between the subcomplexes by creating chimeric PDB files. Complex interaction maps from Cytoscape [369] were then downloaded for each complex. To calculate the degree of each residue, igraph [370] was used when run using the R studio. Subsequently, using the whole protein part of each model, solvent accessible surface area (SASA) was calculated using FreeSASA [371]. Values for all residues reported to be affected by the mutations of interest were collected in Tables A6 and A7 for Pol I and Pol III, respectively. Each mutation was assigned to a "type" based on the location within the complex, relation to structural elements, number of contacts and SASA.

6.6.2 Confocal microscopy

Confocal microscopy was performed together with Granita Lokaj, an internship student under my supervision.

Cells were seeded 48 h before imaging on 35 mm, 0.16-0.19 mm glass-bottom dishes (MatTek Life Sciences). 200 µl of cells collected from a confluent plate were seeded in 2 ml of the medium total. On the day of imaging, cells were overlaid with a medium containing 0.1 µg/ml final concentration of Höchst live DNA dye. Cells were imaged using Olympus FV3000 confocal microscope using a heating chamber set to 37°C with 5% CO_2. UPLSAPO x60 silicone objective was used. Höchst dye was excited using 405 nm excitation laser with 0.2% of the laser power and 500 V and fluorescence was detected within 413-513 nm emission range, while mCherry was excited with 561 nm laser with 1.2% of the laser power and 600 V and its emitted fluorescence was detected within 580-680 nm range. Transmitted light was recorded using a 405 nm excitation laser with 0.2% laser power and 350 V. The alignment of Köhler

illumination was performed before data collection. Imaging settings were adjusted using the D3 cell line and only half of the dynamic range for mCherry to avoid oversaturation on other images. Images were collected at the widest point of the nucleus. Images were processed using Fiji [372], with threshold adjustment performed identically across all images. The threshold was adjusted using the D3 images using the "Auto adjust" function, and then all other images were adjusted to the same values.

6.6.3 Pull-downs for mass spectrometry

Granita Lokaj performed pull-downs with help from Jonas Wiedenhausen. They were then submitted to the EMBL Proteomics Core Facility for the mass spectrometry analysis with protein quantification using TMT stable-isotope labelling.

For each sample, one 10 cm dish was used at 70-80% confluency. Cells were placed on ice and washed three times with PBS, then completely removed by aspiration. They were then resuspended in Lysis buffer (25 mM HEPES, pH 7.5, 150 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 5% glycerol, 0.5% Triton-X100, 20 mM imidazole, 2 mM 2-Mercaptoethanol supplemented with protease inhibitors and benzonase) and cells were scraped off the plate. Protein concentration was measured using Nanodrop and adjusted to ensure similar protein amounts. The sample was then centrifuged at 50,000 G for 30min at 4°C. Per each sample, 125 µl of Strep-Tactin beads (IBA Lifesciences) was applied to a 0.5 mL spin column (MoBiTec GmbH) and washed 2 times with water and then lysis buffer. Supernatant from each sample was applied to the equilibrated beads and incubated at 4°C while rotating at 100 rpm. The sample was washed 2 times with 500 µl of elution buffer (25 mM HEPES, pH 7.5, 150 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 5% glycerol, 20 mM imidazole, 2 mM 2-Mercaptoethanol). After every wash, columns were centrifuged for 2 min at 200 G. Subsequently, 150 µl of elution buffer were added to each column and supplemented with 10 µl of 3C protease (EMBL Protein Expression and Purification Facility). Proteins were cleaved overnight at 4°C while rotating at 100 rpm. They were then spun down for 2 min at 200 G and supernatant was collected. SDS-PAGE gel followed by silver staining was performed to compare the amounts of protein in each sample.

Chapter V: Appendix

APPENDIX 7

Table A1. Primers used for insertion of C-terminal tags on specified genes.

Primer use	Sequence (5‘ – 3‘)
	POLR1C
crRNA sense oligo	caccgACTGAGCTTGGATGCTTCTG
crRNA antisense oligo	aaacCAGAAGCATCCAAGCTCAGTc
T7 endonuclease I assay forward	GGGTTTCCTATAGGCACGTTCCCTCTTCCTCTCCAGGC
T7 endonuclease I assay reverse	GCACTCCAGCCTGGGCAACACAGTGAGACTGTCTC
Zygosity PCR forward	GGGTTTCCTATAGGCACGTTCCCTCTTCCTCTCCAGGC
Zygosity PCR reverse	CCCGGGACTAAAACAGGAAGAAATGTGACTGATGATTCTGGTCTTTCAAGC
Junction PCR forward	CGCTCAGCTGCGGCCTGGCCAAGAAATTG
Junction PCR reverse	CCTCCGCCAGACCCTCCCCCTTTCTCGAACTGG
	POLR1B
crRNA sense oligo	caccgTGTTGACCTTTTGGATTAAG
crRNA antisense oligo	aaacCTTAATCCAAAAGGTCAACAc
T7 endonuclease I assay forward	GCCTTTTACTGAGAGTGGGATGGTCCCAGACATTCTGTTCAATCCCC
T7 endonuclease I assay reverse	GCAGGGTGTAGGGAACCTGGGGTTGGAAAGG
Zygosity PCR forward	GTGTTGTGGATAACATCAAAGTGTGCAGTAATGACACTGGGAGTGG
Zygosity PCR reverse	CCAGAGAGATTCAGAAACCTTGCATACTTGGTTTTCAGGTCTTGGTTGAG
Junction PCR forward (mCherry-TwinStrep-His)	GGCGGAGGAAGTGGAGGCTCTGCTTGGTCTCATCCC
Junction PCR reverse (mCherry-TwinStrep-His)	GCACAGCACCATGCCTGGCTTCCAAAATGGATGAATC
Junction PCR forward (FLAG-Halo tagged line)	GTGTTGTGGATAACATCAAAGTGTGCAGTAATGACACTGGGAGTGG
Junction PCR reverse (FLAG-Halo tagged line)	CCTCGATAAAAACGTTCTGATCGATGATCAGCTTGCGGCCG
	POLR1D
crRNA sense oligo	caccgTCTCCTTGTATACTGCATAA
crRNA antisense oligo	aaacTTATGCAGTATACAAGGAGAc
T7 endonuclease I assay forward	CGGGATTTCTATGCCTTTCAACCCTCTCTTCATCCTCTAGCAGG
T7 endonuclease I assay reverse	CTTGTCTAGGCCTCTGATATTCTCAGTTTCCTCATTTTCAATGCTGGC
Zygosity PCR forward	CTTGTCTAGGCCTCTGATATTCTCAGTTTCCTCATTTTCAATGCTGGC
Zygosity PCR reverse	GGTACCCTTCCAGCTGTTGAGCCATTTCAGAGAGGC
Junction PCR forward	GGAAGCCGTAGTTGGGCCTGCACATTGATCATCTGTATCC
Junction PCR reverse	CCTCCGCCAGACCCTCCCCCTTTCTCGAACTGG

Table A2. cDNA clones ordered for cloning of the recombinant human Pol I.

Gene	Protein	UniProt code	Provider	Reference
POLR1B	RPA2	Q9H9Y6	BioCat	BC110833
TWISTNB	RPA43	Q3B726	DNASU	HsCD00398686
CD3EAP	RPA34	O15446	DNASU	HsCD00079619
ZNRD1	RPA12	Q9P1U0	DNASU	HsCD00079532
POLR1E	RPA49	Q9GZS1	DNASU	HsCD00861617

Table A3. Cryo-EM data collection statistics for apo Pol I.

	Map A apo Pol I with enriched clamp	Map B Apo Pol I with enriched stalk
Data collection and processing		
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e^{-}/Å^2)	50.9	50.9
Defocus range (μm)	0.6-1.4	0.6-1.4
Pixel size (Å)	0.822	0.822
Symmetry imposed	C1	C1
Total no. movies recorded	22,001	22,001
No. of frames	38	38
Initial particle images (no.)	949,538	949,538
Final particle images (no.)	146,558	90,609
Map resolution (Å)	4.72	3.67
FSC threshold	0.143	0.143
Map resolution range (Å)	3.38-15.36	3.43-9.27

Table A4. Cryo-EM data collection, refinement and validation statistics for Pol I EC. Table in modified form taken from [23].

	Map C (EMDB-12795) EC Pol I (PDB 7OB9)	Map D (EMDB-12795) Stalk tWH Focused Classification	Map D1 (EMDB-12795) Multibody Refinement Upper Clamp	Map D2 (EMDB-12795) Multibody Refinement Core	Map E (EMDB-12795) Composite
Data collection and processing					
Magnification	105,000	105,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300	300
Electron exposure (e^{-}/Å^2)	50.9	50.9	50.9	50.9	50.9
Defocus range (μm)	1.0-2.5	1.0-2.5	1.0-2.5	1.0-2.5	1.0-2.5
Pixel size (Å)	0.822	0.822	0.822	0.822	0.822
Symmetry imposed	C1	C1	C1	C1	C1
Total no. movies recorded	10,053	10,053	10,053	10,053	10,053
No. of frames	38	38	38	38	38
Initial particle images (no.)	2,171,736	2,171,736	2,171,736	2,171,736	2,171,736
Final particle images (no.)	198,822	37,180	37,180	37,180	-
Map resolution (Å)	2.7	3.0	3.1	3.0	-
FSC threshold	0.143	0.143	0.143	0.143	
Map resolution range (Å)	2.55-4.50	3.03-5.62	3.13-5.81	3.04-5.10	-
Refinement					
Initial model used (PDB code)	4C3I, 7AE1, 6LHR, 5M5X, 6RQT				
Model resolution (Å)	3.30				
Map sharpening *B* factor (Å^2)	-67				
Model composition					
Non-hydrogen atoms	36471				
Protein residues	4412				
Nucleotide residues	69				
Ligands	6x Zn, 1x Mg				
***B* factors (Å^2)**					
Protein	109.26				
Ligand	115.09				
R.m.s. deviations					
Bond lengths (Å)	0.008				
Bond angles (°)	1.26				
Validation					
MolProbity score	1.60				
Clashscore	4.64				
Poor rotamers (%)	0.31				
Ramachandran plot					
Favored (%)	94.70				
Allowed (%)	5.25				
Disallowed (%)	0.05				

Table A5. Cryo-EM data collection, refinement and validation statistics for Pol I – RRN3 and Pol I OC. Table in modified form taken from [23].

	Map F (EMDB-12796) Pol I – RRN3	Map G (EMDB-12796) Pol I – RRN3 Focused Refinement (PDB 7OBA)	Map H (EMDB-12797) Pol I OC (PDB 7OBB)	Map I (EMDB-12797) Pol I OC Nucleic acid Focused Classification
Data collection and processing				
Magnification	105,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300
Electron exposure (e^{-}/Å^2)	41.26	41.26	41.26	41.26
Defocus range (μm)	0.75-2.25	0.75-2.25	0.75-2.25	0.75-2.25
Pixel size (Å)	0.822	0.822	0.822	0.822
Symmetry imposed	C1	C1	C1	C1
Total no. movies recorded	14,224	14,224	14,224	14,224
No. of frames	40	40	40	40
Initial particle images (no.)	2,628,144	2,628,144	2,628,144	2,628,144
Final particle images (no.)	169,513	260.363	175,912	164,436
Map resolution (Å)	3.2	3.1	3.3	3.3
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.0-7.0	2.9-7.3	3.0-5.9	3.0-6.7
Refinement				
Initial model used (PDB code)		4C3I, 7AE1	4C3I, 7AE1, 5M5W, 6RUO	
Model resolution (Å)		3.1	3.6	
Map sharpening *B* factor (Å^2)		-103	-91	
Model composition				
Non-hydrogen atoms		35684	33051	
Protein residues		4399	4113	
Nucleotide residues		-	17	
Ligands		7x Zn	7x Zn	
***B* factors (Å^2)**				
Protein		65.99	75.19	
Ligand		88.12	119.54	
R.m.s. deviations				
Bond lengths (Å)		0.0111	0.0097	
Bond angles (°)		141	1.37	
Validation				
MolProbity score		1.77	1.76	
Clashscore		5.33	5.21	
Poor rotamers (%)		0.41	0.50	
Ramachandran plot				
Favored (%)		91.96	92.24	
Allowed (%)		7.99	7.74	
Disallowed (%)		0.05	0.02	

Table A6. Disease-causing point mutations in human Pol I. Mutations in bold have been reported as homozygous. SASA – Solvent Accessible Surface Area. Average for whole complex: SASA = 39.3 Å^2; degree = 15.

Related disease	Subunit	Mutation	SASA (Å^2)	Residue degree	Type
Acrofacial Dysostosis (AD)	RPA1	593 E>Q [196]	31.6	14	II
		1299 V>F [196]	6.1	17	II and III
Treacher Collins Syndrome (TCS)	RPA2	682 S>R [192]	1.1	19	II and III
		1003 R>C/S [192]	80.2	14	II
	RPAC1	279 R>Q/W [191,233,236]	15.0	18	I and III
	RPAC2	47 E>K [191]	9.5	18	III
		50 T>I [191]	2.3	20	III
		51 L>R [191]	0.0	19	III
		52 G>E [191]	0.0	14	I and III
		55 L>V [190]	0.0	21	I and III
		56 R>C [191]	25.5	21	I and III
		82 L>S [191]	0.0	17	I
		99 G>S [191]	1.4	15	I and III
Hypomyelinating Leukodystrophy (HLD)	RPAC1	**26 T>I** [233,236]	-	-	-
		27 T>A [233]	-	-	-
		30 P >S [233]	-	-	-
		32 N>I [233,236]	-	-	-
		65 M>V [233,236]	0.2	17	III
		74 N>S [233,236]	2.8	25	III
		94 V>A [233,236]	0.3	20	III
		105 I>F [233]	20.9	16	III
		108 H>Y [233]	19.6	21	III
		109 R>H [233,236]	14.7	22	III
		117 A>P [233]	2.3	16	I
		132 G>D [233,236]	5.0	8	I
		146 C>R [233,236]	0.1	21	I
		191 R>Q [233,236]	96.3	17	IV
		245 V>M [233,236]	7.6	16	I
		262 I>T [233,236]	14.7	19	I
		279 R>Q [233]	15.0	18	I and III
		288 F>S [335]	67.4	13	I and IV
		295 K deletion [233,236]	143.7	10	IV
		313 T>M [233]	7.6	15	I
		324 E>K [233,236]	34.9	17	I

Table A7. Disease-causing point mutations in human Pol III. In bold mutations have been reported as homozygous. SASA – Solvent Accessible Surface Area. Average for whole complex: SASA = 42.4 Å^2; degree = 15. Adapted from [36].

Related disease	Subunit	Mutation	SASA (Å^2)	Residue degree	Type
Hypomyelinating Leukodystrophy (HLD)	RPC1	57 D>N [373]	7.3	19	I
		91 P>L [235]	9.5	14	I
		122 K deletion [230]	40.4	20	I
		166 V>I [235]	7.7	16	II
		310 W>C [374]	12.1	24	I
		372 D>N [237]	0.9	20	III
		387 A>G [235]	0.0	21	I
		396 V>L [230]	0.6	18	I
		478 A>G [375]	0.1	15	I
		553 T>I [230]	0.0	20	I
		558 F>L [237]	27.1	16	III
		599 Q>K/**H** [230]	26.5	19	I
		602 S>R [235,376]	13.8	18	I
		636 S>Y [237]	0.4	17	I
		644 E>K [230]	34.1	18	II
		645 L>F [230]	6.1	17	I
		669 R>G [377]	55.0	17	I and IV
		671 W>R [235]	79.0	15	I and IV
		672 G>E [230,237]	28.4	11	I
		680 M>T [230]	1.9	22	I
		682 R>Q [375]	16.6	21	I
		700 I>F [373]	1.1	18	II
		724 C>Y [237]	0.3	17	II
		775 N>I [237]	11.5	20	I
		784 G>S [378]	15.1	10	I
		804 I>T [374]	19.7	12	II
		848 F>L [230]	0.0	21	II
		849 F>L [230]	1.2	20	II
		852 M>V [235,237,379]	9.9	22	II
		873 R>Q [274]	18.2	19	II
		887 D>V [230]	33.2	14	III
		897 I>N [234]	0.4	17	III
		937 E>V [230]	49.6	16	IV
		941 S>R [230]	9.1	18	I
		1005 R>C/H [234,235,237,380]	38.8	16	I
		1136 R>Q [230]	1.9	20	III
		1240 G>S [230]	6.4	9	I
		1247 T>TT [237]	44.3	12	I and IV
		1249 N>H [379]	31.3	14	I
		1261 E>K [235]	6.2	17	III

Hypomyelinating Leukodystrophy (HLD)	RPC1	1331 A>T [380]	0.0	17	I
	RPC2	**27 W>R** [230]	51.4	16	I and IV
		103 R>H [230]	12.6	23	I
		104 L>F [235]	2.6	20	II
		268 S>G [235]	1.5	15	I
		329 R>Q [230]	26.5	20	III
		363 E>K [381]	27.4	16	I
		365 A>V [381]	0.8	10	I
		375 D>V [381]	6.0	20	I
		418 A>V [230]	7.0	17	I
		426 L>S [381]	5.1	19	II
		442 R>C [235]	29.8	15	II
		449 L>P [230]	10.4	17	I
		493 V>F [230]	25.0	15	I
		503 T>K [382]	1.5	17	II
		523 V>E/D [230,235,382]	10.9	12	I
		527 C>R [235]	0.7	17	III
		634 D>N [375]	2.7	16	I
		667 V>M [230]	3.8	11	I
		727 L>S [230]	38.6	17	I
		768 R>H [234]	5.9	20	III
		895 D>N [230]	4.4	19	I
		926 D>E [234]	0.3	21	III
		967 S>R [230]	23.2	10	I
		973 C>F [230]	8.1	20	I
		982 N>D [230]	14.7	15	III
		1002 I>T [230]	0.0	21	I and III
		1003 Y>C [230]	0.0	20	I and III
		1012 L>P [230]	0.2	17	II
		1024 A>V [230]	22.1	13	III
		1117 L>V [230]	0.4	21	I and III
	RPAC1	**26 T>I** [233,236]	9.1	14	I and III
		27 T>A [233]	15.8	16	I
		30 P >S [233]	6.4	13	III
		32 N>I [233,236]	3.1	18	III
		65 M>V [233,236]	3.9	18	III
		74 N>S [233,236]	3.6	22	III
		94 V>A [233,236]	2.3	16	III
		105 I>F [233]	3.2	17	III
		108 H>Y [233]	3.6	21	III
		109 R>H [233,236]	1.3	23	III
		117 A>P [233]	0.3	17	I
		132 G>D [233,236]	2.6	8	I
		146 C>R [233,236]	0.4	17	I

Hypomyelinating Leukodystrophy (HLD)	RPAC1	191 R>Q [233,236]	101.9	14	IV
		245 V>M [233,236]	10.0	17	I
		262 I>T [233,236]	17.8	16	I
		279 R>Q [233]	65.3	14	I and IV
		288 F>S [335]	37.2	14	I
		295 K deletion [233,236]	144.4	8	IV
		313 T>M [233]	7.5	13	I
		324 E>K [233,236]	36.5	18	I
	RPC10	**41 R>W** [238]	91.4	12	III and IV
Wiedemann-Rautenstrauch syndrome (WDRTS)	RPC1	903 G>R [241]	1.1	8	I
		1069 R>Q [241]	12.6	24	II
		1131 K>R [241]	17.7	17	III
		1292 D>N [241]	4.7	22	I
		1335 G>R [241]	2.1	11	I
Treacher Collins Syndrome (TCS)	RPAC1	279 R>Q/W [191,233,236]	65.3	14	I and IV
	RPAC2	47 E>K [191]	3.4	15	III
		50 T>I [191]	2.0	16	III
		51 L>R [191]	0.0	16	III
		52 G>E [191]	0.1	13	I
		55 L>V [190]	0.0	22	I and III
		56 R>C [191]	19.9	22	I and III
		82 L>S [191]	0.1	20	I
		99 G>S [191]	1.2	14	I
Varicella Zoster Virus Susceptibility (VZV)	RPC1	307 M>V [227]	18.5	18	III
		437 R>Q [227]	68.2	19	IV
		582 R>C [226]	89.2	12	IV
		707 Q>R [227]	126.8	9	IV
	RPC3	11 L>F [227]	10.5	20	I
		84 R>Q [227]	20.7	18	I
		438 R>G [227]	19.6	18	I and III
	RPC5	275 T>M [226]	73.1	11	IV

Table A8. Primers used for insertion of point mutations into POLR1C gene.

Primer use	Sequence (5‘ – 3‘)
POLR1C T26I mutation	
crRNA	UACCGGGAAAGUCAGUAGUAGUUUUAGAGCUAUGCU
Donor template	CGTGGACAAATTCGTCCCTTTGCTCTAGGTCCATATTACTGACTTT CCCGGTAACTATTCCGGTTATGATG
Mismatch qPCR mutation primer forward	CGGAATAGTTACCGGGAAAGTCAGTAA
Mismatch qPCR mutation primer reverse	GCGGAACAGGGATGGGTC
Mismatch qPCR WT primer forward	CGGAATAGTTACCGGGAAAGTCAGTAG
Mismatch qPCR WT primer reverse	GCGGAACAGGGATGGGTC
Sanger sequencing forward	GTTGCCATGTCCCTTGAGGG
Sanger sequencing reverse	TCACTGACTTGAAACAAACCCTTC
POLR1C I105F mutation	
crRNA	UGAGAUUCUUGCUCACCGUCGUUUUAGAGCUAUGCU
Donor template	TGGTGTACAATAATACATCCATTGTTCAGGATGAGTTTCTTGCTCA CCGTCTGGGGCTCATTCCCATTCATGCTGATCCCCGT
Mismatch qPCR mutation primer forward	CCCAGACGGTGAGCAAGAAA
Mismatch qPCR mutation primer reverse	GGTCTGTTGGGTTTTTGCAGATAAGT
Mismatch qPCR WT primer forward	CCCAGACGGTGAGCAAGAAT
Mismatch qPCR WT primer reverse	GGTCTGTTGGGTTTTTGCAGATAAGT
Sanger sequencing forward	GAGGGAATTATCTAAGGGGGAG
Sanger sequencing reverse	CATATGCCTGGTATACACTGGAG
POLR1C R279Q	
crRNA	UGGCCAGAGUUGCCAACCCCGUUUUAGAGCUAUGCU
Donor template	ATTAGGTAAAAAGGTGGCCAGAGTTGCCAACCCCCAGCTGGATAC CTTCAGCAGAGAAATCTTCCGGAATG
Mismatch qPCR mutation primer forward	CAGAGTTGCCAACCCCCA
Mismatch qPCR mutation primer reverse	CATACTCACAGATATAATGATCTCGAACCC
Mismatch qPCR WT primer forward	CAGAGTTGCCAACCCCCG
Mismatch qPCR WT primer reverse	CATACTCACAGATATAATGATCTCGAACCC
Sanger sequencing forward	GTCAAGGGCATTGGTGAGAAC
Sanger sequencing reverse	GTGCACCTTGAAGCGCATG
POLR1C F288S mutation	
crRNA	CCUUCAGCAGAGAAAUCUUCGUUUUAGAGCUAUGCU
Donor template	CCAACCCCCGGCTGGATACCTTCAGCAGAGAAATCTCCCGGAATG AGAAGCTAAAGAAGGTTGTGAGGCTTG
Mismatch qPCR mutation primer forward	CCTTCTTTAGCTTCTCATTCCGGA
Mismatch qPCR mutation primer reverse	CATTAGGTAAAAAGGTGGCCAGAG
Mismatch qPCR WT primer forward	CTGGATACCTTCAGCAGAGAAATCTC
Mismatch qPCR WT primer reverse	ACTCATCTCACCACTTCATACTCAC
Sanger sequencing forward	GCTGACTAGGGAACTCAGCTG
Sanger sequencing reverse	CACCCTTGGTCACCTTCAG

REFERENCES

1. Crick, F.H. (1970) Central dogma of molecular biology. *Nature*. **227** (5258): 561–563.

2. Crick, F.H. (1958) On protein synthesis. *Symposia of the Society for Experimental Biology*. **12**: 138–163.

3. Werner, F. and Grohmann, D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews. Microbiology*. **9** (2): 85–98.

4. Ream, T.S., Haag, J.R., and Pikaard, C.S. (2014) Plant Multisubunit RNA Polymerases IV and V. in: M. Murakami, K., Trakselis (Ed.), Nucleic Acids and Molecular Biology, vol. 30., Springer, Berlin, Heidelberg pp. 289–308.

5. Venkataraman, S., Prasad, B.V.L.S., and Selvarajan, R. (2018) RNA dependent RNA polymerases: Insights from structure, function and evolution. *Viruses*. **10** (2): 76.

6. Pinzón, N., Bertrand, S., Subirana, L., Busseau, I., Escrivá, H., et al. (2019) Functional lability of RNA-dependent RNA polymerases in animals. *PLOS Genetics*. **15** (2): e1007915.

7. Sousa, R. and Mukherjee, S. (2003) T7 RNA polymerase. *Progress in Nucleic Acid Research and Molecular Biology*. **73**: 1–41.

8. Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., et al. (2011) Structure of human mitochondrial RNA polymerase. *Nature*. **478** (7368): 269–273.

9. Börner, T., Aleynikova, A.Y., Zubo, Y.O., and Kusnetsov, V. V. (2015) Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. **1847** (9): 761–769.

10. Girbig, M., Misiaszek, A.D., and Müller, C.W. (2022) Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nature Reviews Molecular Cell Biology*. doi: 10.1038/s41580-022-00476-9.

11. Winkelman, J.T., Nickels, B.E., and Ebright, R.H. (2021) Chapter 1: The transition from transcription initiation to transcription elongation: Start-site selection, initial transcription, and promoter escape. in: RNA Polymerases as Molecular Motors: On the Road, 2nd ed., Royal Society of Chemistry, pp. 1–24.

12. Schärfen, L. and Neugebauer, K.M. (2021) Transcription regulation through nascent RNA folding. *Journal of Molecular Biology*. **433** (14): 166975.

13. Jarrous, N., Mani, D., and Ramanathan, A. (2021) Coordination of transcription and processing of tRNA. *The FEBS Journal*. doi: 10.1111/febs.15904.

14. Noe Gonzalez, M., Blears, D., and Svejstrup, J.Q. (2020) Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nature Reviews Molecular Cell Biology*. **22** (1): 3–21.

15. Darst, S.A., Edwards, A.M., Kubalek, E.W., and Kornberg, R.D. (1991) Three-dimensional structure of yeast RNA polymerase II at 16 Å resolution. *Cell*. **66** (1): 121–128.

16. Darst, S.A., Kubalek, E.W., and Kornberg, R.D. (1989) Three-dimensional structure of Escherichia coli RNA polymerase holoenzyme determined by electron crystallography. *Nature*. **340** (6236): 730–732.

17. Borukhov, S. and Nudler, E. (2003) RNA polymerase holoenzyme: Structure, function and biological implications. *Current Opinion in Microbiology*. **6** (2): 93–100.

18. Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001) Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science*. **292** (5523): 1876–1882.

19. Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 ångstrom resolution. *Science*. **292** (5523): 1863–1876.

20. Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*. **288** (5466): 640–649.

21. Fernández-Tornero, C., Moreno-Morcillo, M., Rashid, U.J., Taylor, N.M.I., Ruiz, F.M., et al. (2013) Crystal structure of the 14-subunit RNA polymerase I. *Nature*. **502** (7473): 644–649.

22. Hoffmann, N.A., Jakobi, A.J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., et al. (2015) Molecular structures of unbound and transcribing RNA polymerase III. *Nature*. **528** (7581): 231–236.

23. Misiaszek, A.D., Girbig, M., Grötsch, H., Baudin, F., Murciano, B., et al. (2021) Cryo-EM structures of human RNA polymerase I. *Nature Structural & Molecular Biology*. **28** (12): 997–1008.

24. Zhao, D., Liu, W., Chen, K., Wu, Z., Yang, H., et al. (2021) Structure of the human RNA polymerase I elongation complex. *Cell Discovery*. **7** (1): 97.

25. Kettenberger, H., Armache, K.J., and Cramer, P. (2004) Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. *Molecular Cell*. **16** (6): 955–965.

26. Griesenbeck, J., Tschochner, H., and Grohmann, D. (2017) Structure and function of RNA polymerases and the transcription machineries. *Sub-Cellular Biochemistry*. **83**: 225–270.

27. Zuo, Y. and Steitz, T.A. (2016) A structure-based kinetic model of transcription. *Transcription*. **8** (1): 1–8.

28. Dickson, K.S., Burns, C.M., and Richardson, J.P. (2000) Determination of the free-energy change for repair of a DNA phosphodiester bond. *Journal of Biological Chemistry*. **275** (21): 15828–15831.

29. Frey, P.A. and Arabshahi, A. (1995) Standard free energy change for the hydrolysis of the alpha, beta-phosphoanhydride bridge in ATP. *Biochemistry*. **34** (36): 11307–11310.

30. Mazumder, A., Lin, M., Kapanidis, A.N., and Ebright, R.H. (2020) Closing and opening of the RNA polymerase trigger loop. *Proceedings of the National Academy of Sciences of the United States of America*. **117** (27): 15642–15649.

31. Weinzierl, R.O.J. (2010) The nucleotide addition cycle of RNA polymerase is controlled by two molecular hinges in the bridge helix domain. *BMC Biology*. **8** (1): 1–15.

32. Tagami, S., Sekine, S.I., Kumarevel, T., Hino, N., Murayama, Y., et al. (2010) Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature*. **468** (7326): 978–982.

33. Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., et al. (2008) Structure of eukaryotic RNA polymerases. *Annual Review of Biophysics*. **37**: 337–352.

34. Barba-Aliaga, M., Alepuz, P., and Pérez-Ortín, J.E. (2021) Eukaryotic RNA polymerases: The many ways to transcribe a gene. *Frontiers in Molecular Biosciences*. **8**: 663209.

35. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016) Structure of transcribing mammalian RNA polymerase II. *Nature*. **529** (7587): 551–554.

36. Girbig, M., Misiaszek, A.D., Vorländer, M.K., Lafita, A., Grötsch, H., et al. (2021) Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states. *Nature Structural & Molecular Biology*. **28** (2): 210–219.

37. Roeder, R.G. (2019) 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nature Structural & Molecular Biology*. **26** (9): 783–791.

38. Weiss, S.B. and Gladstone, L. (1959) A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *Journal of the American Chemical Society*. **81** (15): 4118–4119.

39. Hurwitz, J. (2005) The discovery of RNA polymerase. *Journal of Biological Chemistry*. **280** (52): 42477–42485.

40. Roeder, R.G. and Rutter, W.J. (1969) Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms. *Nature*. **224** (5216): 234–237.

41. Hager, G.L., Holland, M.J., and Rutter, W.J. (1977) Isolation of ribonucleic acid polymerases I, II, and III from Saccharomyces cerevisiae. *Biochemistry*. **16** (1): 1–8.

42. Sklar, V.E.F., Schwartz, L.B., and Roeder, R.G. (1975) Distinct molecular structures of nuclear class I, II, and III DNA dependent RNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America*. **72** (1): 348–352.

43. Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G., and Rutter, W.J. (1970) Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science*. **170** (3956): 447–449.

44. Weinmann, R. and Roeder, R.G. (1974) Role of DNA-Dependent RNA Polymerase III in the Transcription of the tRNA and 5S RNA Genes. *Proceedings of the National Academy of Sciences of the United States of America*. **71** (5): 1790–1794.

45. Lalo, D., Steffan, J.S., Dodd, J.A., and Nomura, M. (1996) RRN11 encodes the third subunit of the complex containing Rrn6p and Rrn7p that is essential for the initiation of rDNA transcription by yeast RNA polymerase I. *Journal of Biological Chemistry*. **271** (35): 21062–21067.

46. Steffan, J.S., Keys, D.A., Vu, L., and Nomura, M. (1998) Interaction of TATA-binding protein with Upstream Activation Factor is required for activated transcription of ribosomal DNA by RNA polymerase I in S. cerevisiae in vivo. *Molecular and Cellular Biology*. **18** (7): 3752–3761.

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

48. Learned, R.M., Cordes, S., and Tjian, R. (1985) Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I. *Molecular and Cellular Biology*. **5** (6): 1358–1369.

49. Choe, S.Y., Schultz, M.C., and Reeder, R.H. (1992) In vitro definition of the yeast RNA polymerase I promoter. *Nucleic Acids Research*. **20** (2): 279–285.

50. Grummt, I. (1982) Nucleotide sequence requirements for specific initiation of transcription by RNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*. **79** (22 I): 6908–6911.

51. Schultz, M.C., Choe, S.Y., and Reeder, R.H. (1991) Specific initiation by RNA polymerase I in a whole-cell extract from yeast. *Proceedings of the National Academy of Sciences of the United States of America*. **88** (3): 1004–1008.

52. Wilkinson, J.A.K. and Sollner Webb, B. (1982) Transcription of Xenopus ribosomal RNA genes by RNA polymerase I in vitro. *Journal of Biological Chemistry*. **257** (23): 14375–14383.

53. Bischler, N., Brino, L., Carles, C., Riva, M., Tschochner, H., et al. (2002) Localization of the yeast RNA polymerase I-specific subunits. *The EMBO Journal*. **21** (15): 4136–4144.

54. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D., and Cramer, P. (2013) RNA polymerase I structure and transcription regulation. *Nature*. **502** (7473): 650–655.

55. Fernández-Tornero, C., Moreno-Morcillo, M., Rashid, U.J., Taylor, N.M.I., Ruiz, F.M., et al. (2013) Crystal structure of the 14-subunit RNA polymerase I. *Nature*. **502** (7473): 644–649.

56. Tafur, L., Sadian, Y., Hoffmann, N.A., Jakobi, A.J., Wetzel, R., et al. (2016) Molecular Structures of Transcribing RNA Polymerase I. *Molecular Cell*. **64** (6): 1135–1143.

57. Neyer, S., Kunz, M., Geiss, C., Hantsche, M., Hodirnau, V.V., et al. (2016) Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature*. **540** (7634): 607–610.

58. Pilsl, M., Crucifix, C., Papai, G., Krupp, F., Steinbauer, R., et al. (2016) Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nature Communications*. **7**: 12126.

59. Engel, C., Plitzko, J., and Cramer, P. (2016) RNA polymerase I–Rrn3 complex at 4.8 Å resolution. *Nature Communications*. **7**: 12129.

60. Tafur, L., Sadian, Y., Hanske, J., Wetzel, R., Weis, F., et al. (2019) The cryo-EM structure of a 12-subunit variant of RNA polymerase I reveals dissociation of the A49-A34.5 heterodimer and rearrangement of subunit A12.2. *eLife*. **8**: e43204.

61. Sadian, Y., Baudin, F., Tafur, L., Murciano, B., Wetzel, R., et al. (2019) Molecular insight into RNA polymerase I promoter recognition and promoter melting. *Nature Communications*. **10**: 5543.

62. Sadian, Y., Tafur, L., Kosinski, J., Jakobi, A.J., Wetzel, R., et al. (2017) Structural insights into transcription initiation by yeast RNA polymerase I. *The EMBO Journal*. **36** (18): 2698–2709.

63. Engel, C., Gubbey, T., Neyer, S., Sainsbury, S., Oberthuer, C., et al. (2017) Structural basis of RNA polymerase I transcription initiation. *Cell*. **169** (1): 120-131.e22.

64. Heiss, F.B., Daiß, J.L., Becker, P., and Engel, C. (2021) Conserved strategies of RNA polymerase I hibernation and activation. *Nature Communications*. **12**: 758.

65. Schneider, D.A., French, S.L., Osheim, Y.N., Bailey, A.O., Vu, L., et al. (2006) RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. *Proceedings of the National Academy of Sciences of the United States of America*. **103** (34): 12707–12712.

66. Viktorovskaya, O. V., Appling, F.D., and Schneider, D.A. (2011) Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly. *Journal of Biological Chemistry*. **286** (21): 18825–18833.

67. Huffines, A.K., Edwards, Y.J.K., and Schneider, D.A. (2021) Spt4 promotes Pol I processivity and transcription elongation. *Genes*. **12** (3): 413.

68. Zhang, Y., Sikes, M.L., Beyer, A.L., and Schneider, D.A. (2009) The Paf1 complex is required for efficient transcription elongation by RNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*. **106** (7): 2153–2158.

69. Turowski, T.W., Petfalski, E., Goddard, B.D., French, S.L., Helwak, A., et al. (2020) Nascent transcript folding plays a major role in determining RNA polymerase elongation rates. *Molecular Cell*. **79** (3): 488–503.

70. Palazzo, A.F. and Lee, E.S. (2015) Non-coding RNA: What is functional and what is junk? *Frontiers in Genetics*. **6**: 2.

71. Birch, J.L. and Zomerdijk, J.C.B.M. (2008) Structure and function of ribosomal RNA gene chromatin. *Biochemical Society Transactions*. **36** (4): 619–624.

72. Németh, A. and Längst, G. (2011) Genome organization in and around the nucleolus. *Trends in Genetics*. **27** (4): 149–156.

73. Miller, O.L. and Beatty, B.R. (1969) Visualization of nucleolar genes. *Science*. **164** (3882): 955–957.

74. Turowski, T.W. and Boguta, M. (2021) Specific features of RNA polymerases I and III: Structure and assembly. *Frontiers in Molecular Biosciences*. **8**: 349.

75. Geiger, S.R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., et al. (2010) RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Molecular Cell*. **39** (4): 583–594.

76. Chédin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. (1998) The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes & Development*. **12** (24): 3857–3871.

77. McStay, B. and Grummt, I. (2008) The epigenetics of rRNA genes: From molecular to chromosome biology. *Annual Review of Cell and Developmental Biology*. **24**: 131–157.

78. Long, E.O. and Dawid, I.B. (1980) Repeated genes in eukaryotes. *Annual Review of Biochemistry*. **49**: 727–764.

79. Panov, K.I., Friedrich, J.K., Russell, J., and Zomerdijk, J.C.B.M. (2006) UBF activates RNA polymerase I transcription by stimulating promoter escape. *The EMBO Journal*. **25** (14): 3310–3322.

80. Smith, S.D., O'Mahony, D.J., Kinsella, B.T., and Rothblum, L.I. (1993) Transcription from the rat 45S ribosomal DNA promoter does not require the factor UBF. *Gene Expression*. **3** (3): 229–236.

81. Knutson, B.A., Smith, M.L., Belkevich, A.E., and Fakhouri, A.M. (2020) Molecular Topology of RNA Polymerase I Upstream Activation Factor.

Molecular and Cellular Biology. **40** (13): e00056-20.

82. Baudin, F., Murciano, B., Fung, H.K.H., Fromm, S.A., Mattei, S., et al. (2022) Mechanism of RNA polymerase I selection by transcription factor UAF. *Science Advances*. **8** (16): 5725.

83. Grummt, I. (2003) Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes & Development*. **17** (14): 1691–1702.

84. O'Sullivan, A.C., Sullivan, G.J., and McStay, B. (2002) UBF binding in vivo is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat. *Molecular and Cellular Biology*. **22** (2): 657–658.

85. Goodfellow, S.J. and Zomerdijk, J.C.B.M. (2013) Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. *Sub-Cellular Biochemistry*. **61**: 211–236.

86. Grimaldi, G. and Di Nocera, P.P. (1988) Multiple repeated units in Drosophila melanogaster ribosomal DNA spacer stimulate rRNA precursor transcription. *Proceedings of the National Academy of Sciences of the United States of America*. **85** (15): 5502–5506.

87. Santoro, R., Schmitz, K.M., Sandoval, J., and Grummt, I. (2010) Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO Reports*. **11** (1): 52–58.

88. Mayer, C., Schmitz, K.M., Li, J., Grummt, I., and Santoro, R. (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. *Molecular Cell*. **22** (3): 351–361.

89. Abraham, K.J., Khosraviani, N., Chan, J.N.Y., Gorthi, A., Samman, A., et al. (2020) Nucleolar RNA polymerase II drives ribosome biogenesis. *Nature*. **585** (7824): 298–302.

90. Németh, A., Guibert, S., Tiwari, V.K., Ohlsson, R., and Längst, G. (2008) Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. *The EMBO Journal*. **27** (8): 1255–1265.

91. Gonzalez, I.L. and Sylvester, J.E. (1995) Complete Sequence of the 43-kb Human Ribosomal DNA Repeat: Analysis of the Intergenic Spacer. *Genomics*. **27** (2): 320–328.

92. Sylvester, J.E., Whiteman, D.A., Podolsky, R., Pozsgay, J.M., Respess, J., et al. (1986) The human ribosomal RNA genes: structure and organization of the complete repeating unit. *Human Genetics*. **73** (3): 193–198.

93. Heix, J. and Grummt, I. (1995) Species specificity of transcription by RNA polymerase I. *Current Opinion in Genetics & Development*. **5** (5): 652–656.

94. Parks, M.M., Kurylo, C.M., Dass, R.A., Bojmar, L., Lyden, D., et al. (2018) Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Science Advances*. **4** (2): 0665.

95. Gagnon-Kugler, T., Langlois, F., Stefanovsky, V., Lessard, F., and Moss, T. (2009) Loss of human ribosomal gene CpG methylation enhances cryptic RNA polymerase II transcription and disrupts ribosomal RNA processing. *Molecular Cell*. **35** (4): 414–425.

96. Zentner, G.E., Saiakhova, A., Manaenkov, P., Adams, M.D., and Scacheri, P.C. (2011) Integrative genomic analysis of human ribosomal DNA. *Nucleic Acids Research*. **39** (12): 4949–4960.

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

98. Moorefield, B., Greene, E.A., and Reeder, R.H. (2000) RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human. *Proceedings of the National Academy of Sciences of the United States of America*. **97** (9): 4724–4729.

99. Miller, G., Panov, K.I., Friedrich, J.K., Trinkle-Mulcahy, L., Lamond, A.I., et al. (2001) HRRN3 is essential in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters. *The EMBO Journal*. **20** (6): 1373–1382.

100. Milkereit, P., Schultz, P., and Tschochner, H. (1997) Resolution of RNA polymerase I into dimers and monomers and their function in transcription. *Biological Chemistry*. **378** (12): 1433–1443.

101. Milkereit, P. and Tschochner, H. (1998) A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *The EMBO Journal*. **17** (13): 3692–3703.

102. Mayer, C., Zhao, J., Yuan, X., and Grummt, I. (2004) MTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. *Genes & Development*. **18** (4): 423–434.

103. Mayer, C., Bierhoff, H., and Grummt, I. (2005) The nucleolus as a stress sensor: JNK2 inactivates the transcription factor TIF-IA and down-regulates rRNA synthesis. *Genes & Development*. **19** (8): 933–941.

104. Cavanaugh, A.H., Hirschler-Laszkiewicz, I., Hu, Q., Dundr, M., Smink, T., et al. (2002) Rrn3 phosphorylation is a regulatory checkpoint for ribosome biogenesis. *Journal of Biological Chemistry*. **277** (30): 27423–27432.

105. Chen, S., Seiler, J., Santiago-Reichelt, M., Felbel, K., Grummt, I., et al. (2013) Repression of RNA polymerase I upon stress is caused by inhibition of RNA-dependent deacetylation of PAF53 by SIRT7. *Molecular Cell*. **52** (3): 303–313.

106. Zhao, J., Yuan, X., Frödin, M., and Grummt, I. (2003) ERK-dependent phosphorylation of the transcription initiation factor TIF-IA is required for RNA polymerase I transcription and cell growth. *Molecular Cell*. **11** (2): 405–413.

107. Bierhoff, H., Dundr, M., Michels, A.A., and Grummt, I. (2008) Phosphorylation by Casein Kinase 2 facilitates rRNA gene transcription by promoting dissociation of TIF-IA from elongating RNA polymerase I. *Molecular and Cellular Biology*. **28** (16): 4988–4998.

108. Bell, S.P., Learned, R.M., Jantzen, H.M., and Tjian, R. (1988) Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science*. **241** (4870): 1192–1198.

109. Kihm, A.J., Hershey, J.C., Haystead, T.A.J., Madsen, C.S., and Owens, G.K. (1998) Phosphorylation of the rRNA transcription factor upstream binding factor promotes its association with TATA binding protein. *Proceedings of the National Academy of Sciences of the United States of America*. **95** (25): 14816–14820.

110. Comai, L., Zomerdijk, J.C.B.M., Beckmann, H., Zhou, S., Admon, A., et al. (1994) Reconstitution of transcription factor SL1: exclusive binding of TBP by SL1 or TFIID subunits. *Science*. **266** (5193): 1966–1972.

111. Panov, K.I., Friedrich, J.K., Russell, J., and Zomerdijk, J.C.B.M. (2006) UBF activates RNA polymerase I transcription by stimulating promoter escape. *The EMBO Journal*. **25** (14): 3310.

112. Panov, K.I., Friedrich, J.K., and Zomerdijk, J.C.B.M. (2001) A step subsequent to preinitiation complex assembly at the ribosomal RNA gene promoter is rate limiting for human RNA polymerase I-dependent transcription. *Molecular and Cellular Biology*. **21** (8): 2641–2649.

113. Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., et al. (2007) Functional architecture of RNA polymerase I. *Cell*. **131** (7): 1260–1272.

114. McNamar, R., Abu-Adas, Z., Rothblum, K., Knutson, B.A., and Rothblum, L.I. (2019) Conditional depletion of the RNA polymerase I subunit PAF53 reveals that it is essential for mitosis and enables identification of functional domains. *Journal of Biological Chemistry*. **294** (52): 19907.

115. Clarke, A.M., Huffines, A.K., Edwards, Y.J.K., Petit, C.M., and Schneider, D.A. (2021) Defining the influence of the A12.2 subunit on transcription elongation and termination by RNA polymerase I in vivo. *Genes*. **12** (12): 1939.

116. Appling, F.D., Schneider, D.A., and Lucius, A.L. (2017) Multisubunit RNA polymerase cleavage factors modulate the kinetics and energetics of nucleotide incorporation: An RNA polymerase I case study. *Biochemistry*. **56** (42): 5654–5662.

117. Scull, C.E., Lucius, A.L., and Schneider, D.A. (2021) The N-terminal domain of the A12.2 subunit stimulates RNA polymerase I transcription elongation. *Biophysical Journal*. **120** (10): 1883–1893.

118. Jacobs, R.Q., Ingram, Z.M., Lucius, A.L., and Schneider, D.A. (2021) Defining the divergent enzymatic properties of RNA polymerases I and II. *Journal of Biological Chemistry*. **296**: 100051.

119. Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., et al. (2002) A kinetic framework for a mammalian RNA polymerase in vivo. *Science*. **298** (5598): 1623–1626.

120. Clarke, A.M., Engel, K.L., Giles, K.E., Petit, C.M., and Schneider, D.A. (2018) NETSeq reveals heterogeneous nucleotide incorporation by RNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*. **115** (50): E11633–E11641.

121. Ucuncuoglu, S., Engel, K.L., Purohit, P.K., Dunlap, D.D., Schneider, D.A., et al. (2016) Direct characterization of transcription elongation by RNA polymerase I. *PLOS ONE*. **11** (7): e0159527.

122. Huffines, A.K., Edwards, Y.J.K., and Schneider, D.A. (2021) Spt4 promotes Pol I processivity and transcription elongation. *Genes*. **12** (3): 413.

123. Anderson, S.J., Sikes, M.L., Zhang, Y., French, S.L., Salgia, S., et al. (2011) The transcription elongation factor Spt5 influences transcription by RNA polymerase I positively and negatively. *Journal of Biological Chemistry*. **286** (21): 18816–18824.

124. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., et al. (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development*. **12** (3): 343–356.

125. Knutson, B.A., McNamar, R., and Rothblum, L.I. (2020) Dynamics of the RNA polymerase I TFIIF/TFIIE-like subcomplex: A mini-review. *Biochemical Society Transactions*. **48** (5): 1917–1927.

126. Assfalg, R., Lebedev, A., Gonzalez, O.G., Schelling, A., Koch, S., et al. (2012) TFIIH is an elongation factor of RNA polymerase I. *Nucleic Acids Research*. **40** (2): 650–659.

127. Nogales, E. and Greber, B.J. (2019) High-resolution cryo-EM structures of TFIIH and their functional implications. *Current Opinion in Structural Biology*. **59**: 188.

128. Jones, H.S., Kawauchi, J., Braglia, P., Alen, C.M., Kent, N.A., et al. (2007) RNA polymerase I in yeast transcribes dynamic nucleosomal rDNA. *Nature Structural & Molecular Biology*. **14** (2): 123–130.

129. Zhang, Y., Anderson, S.J., French, S.L., Sikes, M.L., Viktorovskaya, O. V., et al. (2013) The SWI/SNF chromatin remodeling complex influences transcription by RNA polymerase I in Saccharomyces cerevisiae. *PLOS ONE*. **8** (2): e56793.

130. Rickards, B., Flint, S.J., Cole, M.D., and LeRoy, G. (2007) Nucleolin is required for RNA polymerase I transcription in vivo. *Molecular and Cellular Biology*. **27** (3): 439–444.

131. Stefanovsky, V., Langlois, F., Gagnon-Kugler, T., Rothblum, L.I., and Moss, T. (2006) Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling. *Molecular Cell*. **21** (5): 629–639.

132. Fath, S., Kobor, M.S., Philippi, A., Greenblatt, J., and Tschochner, H. (2004) Dephosphorylation of RNA polymerase I by Fcp1p is required for efficient rRNA synthesis. *Journal of Biological Chemistry*. **279** (24): 25251–25259.

133. Osheim, Y.N., French, S.L., Keck, K.M., Champion, E.A., Spasov, K., et al. (2004) Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. *Molecular Cell*. **16** (6): 943–954.

134. Albert, B., Perez-Fernandez, J., Léger-Silvestre, I., and Gadal, O. (2012) Regulation of ribosomal RNA production by RNA polymerase I: Does elongation come first? *Genetics Research International*. **2012**: 276948.

135. El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010) Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes & Development*. **24** (14): 1546–1558.

136. Schneider, D.A., Michel, A., Sikes, M.L., Vu, L., Dodd, J.A., et al. (2007) Transcription elongation by RNA polymerase I is linked to efficient rRNA processing and ribosome assembly. *Molecular Cell*. **26** (2): 217–229.

137. Takeuchi, Y., Horiuchi, T., and Kobayashi, T. (2003) Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes & Development*. **17** (12): 1497–1506.

138. Mekhail, K., Seebacher, J., Gygi, S.P., and Moazed, D. (2008) Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature*. **456** (7222): 667–670.

139. Wu, H., Henras, A., Chanfreau, G., and Feigon, J. (2004) Structural basis for of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proceedings of the National Academy of Sciences of the United States of America*. **101** (22): 8307–8312.

140. Kufel, J., Dichtl, B., and Tollervey, D. (1999) Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA*. **5** (7): 909–917.

141. Henras, A.K., Sam, M., Hiley, S.L., Haihong, W.U., Hughes, T.R., et al. (2005) Biochemical and genomic analysis of substrate recognition by the double-stranded RNA binding domain of yeast RNase III. *RNA*. **11** (8): 1225–1237.

142. Reeder, R.H., Guevara, P., and Roan, J.G. (1999) Saccharomyces cerevisiae RNA polymerase I terminates transcription at the Reb1 terminator in vivo. *Molecular and Cellular Biology*. **19** (11): 7369–7376.

143. Lang, W.H. and Reeder, R.H. (1993) The REB1 site is an essential component of a terminator for RNA polymerase I in Saccharomyces cerevisiae. *Molecular and Cellular Biology*. **13** (1): 649–658.

144. Reiter, A., Hamperl, S., Seitz, H., Merkl, P., Perez-Fernandez, J., et al. (2012) The Reb1-homologue Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. *The EMBO Journal*. **31** (16): 3480–3493.

145. El Hage, A., Koper, M., Kufel, J., and Tollervey, D. (2008) Efficient termination of transcription by RNA polymerase I requires the 5′ exonuclease Rat1 in yeast. *Genes & Development*. **22** (8): 1069–1081.

146. Tollervey, D. (2004) Termination by torpedo. *Nature*. **432** (7016): 456–457.

147. Dengl, S. and Cramer, P. (2009) Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *Journal of Biological Chemistry*. **284** (32): 21270–21279.

148. Jaiswal, R., Choudhury, M., Zaman, S., Singh, S., Santosh, V., et al. (2016) Functional architecture of the Reb1-Ter complex of Schizosaccharomyces pombe. *Proceedings of the National Academy of Sciences of the United States of America*. **113** (16): E2267–E2276.

149. Prescott, E.M., Osheim, Y.N., Jones, H.S., Alen, C.M., Roan, J.G., et al. (2004) Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proceedings of the National Academy of Sciences of the United States of America*. **101** (16): 6068–6073.

150. Lang, W.H. and Reeder, R.H. (1995) Transcription termination of RNA polymerase I due to a T-rich element interacting with Reb1p. *Proceedings of the National Academy of Sciences of the United States of America*. **92** (21): 9781–9785.

151. Grummt, I., Rosenbauer, H., Niedermeyer, I., Maier, U., and Öhrlein, A. (1986) A repeated 18 bp sequence motif in the mouse rDNA spacer mediates binding of a nuclear factor and transcription termination. *Cell*. **45** (6): 837–846.

152. La Volpe, A., Simeone, A., D'Esposito, M., Scotto, L., Fidanza, V., et al. (1985) Molecular analysis of the heterogeneity region of the human ribosomal spacer. *Journal of Molecular Biology*. **183** (2): 213–223.

153. Evers, R., Smid, A., Rudloff, U., Lottspeich, F., and Grummt, I. (1995) Different domains of the murine RNA polymerase I-specific termination factor mTTF-I serve distinct functions in transcription termination. *The EMBO Journal*. **14** (6): 1248–1256.

154. Smid, A., Finsterer, M., and Grummt, I. (1992) Limited proteolysis unmasks specific DNA-binding of the murine RNA polymerase I-specific transcription termination factor TTFI. *Journal of Molecular Biology*. **227** (3): 635–647.

155. Jansa, P., Mason, S.W., Hoffmann-Rohrer, U., and Grummt, I. (1998) Cloning and functional characterization of PTRF, a novel protein which induces dissociation of paused ternary transcription complexes. *The EMBO Journal*. **17** (10): 2855–2864.

156. Akamatsu, Y. and Kobayashi, T. (2015) The human RNA polymerase I transcription terminator complex acts as a replication fork barrier that coordinates the progress of replication with rRNA transcription activity. *Molecular and Cellular Biology*. **35** (10): 1871–1881.

157. Salifou, K., Ray, S., Verrier, L., Aguirrebengoa, M., Trouche, D., et al. (2016) The histone demethylase JMJD2A/KDM4A links ribosomal RNA transcription to nutrients and growth factors availability. *Nature Communications*. **7**: 10174.

158. Woods, S.J., Hannan, K.M., Pearson, R.B., and Hannan, R.D. (2015) The nucleolus as a fundamental regulator of the p53 response and a new target for cancer therapy. *Biochimica et Biophysica Acta*. **1849** (7): 821–829.

159. Drygin, D., O'Brien, S.E., Hannan, R.D., McArthur, G.A., and Von Hoff, D.D. (2014) Targeting the nucleolus for cancer-specific activation of p53. *Drug Discovery Today*. **19** (3): 259–265.

160. Gómez-Herreros, F., Rodríguez-Galán, O., Morillo-Huesca, M., Maya, D., Arista-Romero, M., et al. (2013) Balanced production of ribosome components is required for proper G1/S transition in Saccharomyces cerevisiae. *Journal of Biological Chemistry*. **288** (44): 31689–31700.

161. Klein, J. and Grummt, I. (1999) Cell cycle-dependent regulation of RNA polymerase I transcription: the nucleolar transcription factor UBF is inactive in mitosis and early G1. *Proceedings of the National Academy of Sciences of the United States of America*. **96** (11): 6096–6101.

162. Ghosh, A., Rideout, E.J., and Grewal, S.S. (2014) TIF-IA-dependent regulation of ribosome synthesis in Drosophila muscle is required to maintain dystemic insulin dignaling and larval growth. *PLOS Genetics*. **10** (10): e1004750.

163. Zhang, Q., Shalaby, N.A., and Buszczak, M. (2014) Changes in rRNA transcription influence proliferation and cell fate within a stem cell lineage. *Science*. **343** (6168): 298–301.

164. Ferreira, R., Schneekloth, J.S., Panov, K.I., Hannan, K.M., and Hannan, R.D. (2020) Targeting the RNA polymerase I transcription for cancer therapy comes of age. *Cells*. **9** (2): 266.

165. Derenzini, M., Montanaro, L., and Treré, D. (2009) What the nucleolus says to a tumour pathologist. *Histopathology*. **54** (6): 753–762.

166. Bywater, M.J., Poortinga, G., Sanij, E., Hein, N., Peck, A., et al. (2012) Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53. *Cancer Cell*. **22** (1): 51–65.

167. Uemura, M., Zheng, Q., Koh, C.M., Nelson, W.G., Yegnasubramanian, S., et al. (2011) Overexpression of ribosomal RNA in prostate cancer is common but not linked to rDNA promoter hypomethylation. *Oncogene*. **31** (10): 1254–1263.

168. Williamson, D., Lu, Y.J., Fang, C., Pritchard-Jones, K., and Shipley, J. (2006) Nascent pre-rRNA overexpression correlates with an adverse prognosis in alveolar rhabdomyosarcoma. *Genes, Chromosomes & Cancer*. **45** (9): 839–845.

169. Batsotti, P., Ascoli, V., Nardi, F., and Marinozri, V. (1990) Silver staining of interphase nucleolar organizer regions in cytologic smears previously stained by the papanicolaou and may-Grünwald-Giemsa Techniques. *Diagnostic Cytopathology*. **6** (4): 289–296.

170. Quin, J.E., Devlin, J.R., Cameron, D., Hannan, K.M., Pearson, R.B., et al. (2014) Targeting the nucleolus for cancer intervention. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. **1842** (6): 802–816.

171. Carotenuto, P., Pecoraro, A., Palma, G., Russo, G., and Russo, A. (2019) Therapeutic approaches targeting nucleolus in cancer. *Cells*. **8** (9): 1090.

172. Donizy, P., Biecek, P., Halon, A., Maciejczyk, A., and Matkowski, R. (2017) Nucleoli cytomorphology in cutaneous melanoma cells - A new prognostic approach to an old concept. *Diagnostic Pathology*. **12**: 88.

173. Derenzini, M., Betts, C.M., Ceccarelli, C., and Eusebi, V. (1986) Ultrastructural organization of nucleoli in benign naevi and malignant melanomas. *Virchows Archiv. B, Cell Pathology*. **52** (4): 343–352.

174. Van Riggelen, J., Yetil, A., and Felsher, D.W. (2010) MYC as a regulator of ribosome biogenesis and protein synthesis. *Nature Reviews. Cancer*. **10** (4): 301–309.

175. Iarovaia, O. V., Minina, E.P., Sheval, E. V., Onichtchouk, D., Dokudovskaya, S., et al. (2019) Nucleolus: A Central Hub for Nuclear Functions. *Trends in Cell Biology*. **29** (8): 647–659.

176. Dousset, T., Wang, C., Verheggen, C., Chen, D., Hernandez-Verdun, D., et al. (2000) Initiation of nucleolar assembly is independent of RNA polymerase I transcription. *Molecular Biology of the Cell*. **11** (8): 2705–2717.

177. Jiménez-García, L.F., Segura-Valdez, M.D.L., Ochs, R.L., Rothblum, L.I., Hannan, R., et al. (1994) Nucleologenesis: U3 snRNA-containing prenucleolar bodies move to sites of active pre-rRNA transcription after mitosis. *Molecular Biology of the Cell*. **5** (9): 955–966.

178. Ochs, R.L., Lischwe, M.A., Shen, E., Carroll, R.E., and Busch, H. (1985) Nucleologenesis: Composition and fate of prenucleolar bodies. *Chromosoma*. **92** (5): 330–336.

179. Tsekrekou, M., Stratigi, K., and Chatzinikolaou, G. (2017) The nucleolus: In genome maintenance and repair. *International Journal of Molecular Sciences*. **18** (7): 1411.

180. Lindström, M.S., Jurada, D., Bursac, S., Orsolic, I., Bartek, J., et al. (2018) Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. *Oncogene*. **37** (18): 2351–2366.

181. Bywater, M.J., Poortinga, G., Sanij, E., Hein, N., Peck, A., et al. (2012) Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53. *Cancer Cell*. **22** (1): 51–65.

182. Jacobs, R.Q., Huffines, A.K., Laiho, M., and Schneider, D.A. (2022) The small molecule BMH-21 directly inhibits transcription elongation and DNA occupancy of RNA polymerase I in vivo and in vitro. *Journal of Biological Chemistry*. **298** (1): 101450.

183. Peltonen, K., Colis, L., Liu, H., Jäämaa, S., Moore, H.M., et al. (2010) Identification of novel p53 pathway activating small-molecule compounds reveals unexpected similarities with known therapeutic agents. *PLOS ONE*. **5** (9): e12996.

184. Hannan, K.M., Sanij, E., Rothblum, L.I., Hannan, R.D., and Pearson, R.B. (2013) Dysregulation of RNA polymerase I transcription during disease. *Biochimica et Biophysica Acta*. **1829** (3–4): 342–360.

185. Sharifi, S. and Bierhoff, H. (2018) Regulation of RNA polymerase I transcription in development, disease, and aging. *Annual Review of Biochemistry*. **87**: 51–73.

186. Hayashi, Y., Kuroda, T., Kishimoto, H., Wang, C., Iwama, A., et al. (2014) Downregulation of rRNA transcription triggers cell differentiation. *PLOS ONE*. **9** (5): e98586.

187. Poortinga, G., Wall, M., Sanij, E., Siwicki, K., Ellul, J., et al. (2011) C-MYC coordinately regulates ribosomal gene chromatin remodeling and Pol I availability during granulocyte differentiation. *Nucleic Acids Research*. **39** (8): 3267–3281.

188. Hein, N., Cameron, D.P., Hannan, K.M., Nguyen, N.Y.N., Fong, C.Y., et al. (2017) Inhibition of Pol I transcription treats murine and human AML by targeting the leukemia-initiating cell population. *Blood*. **129** (21): 2882–2895.

189. Trainor, P.A., Dixon, J., and Dixon, M.J. (2008) Treacher Collins syndrome: Etiology, pathogenesis and prevention. *European Journal of Human Genetics*. **17** (3): 275–283.

190. Schaefer, E., Collet, C., Genevieve, D., Vincent, M., Lohmann, D.R., et al. (2014) Autosomal recessive POLR1D mutation with decrease of TCOF1 mRNA is responsible for Treacher Collins syndrome. *Genetics in Medicine*. **16** (9): 720–724.

191. Dauwerse, J.G., Dixon, J., Seland, S., Ruivenkamp, C.A.L., Van Haeringen, A., et al. (2010) Mutations in genes encoding subunits of RNA polymerases I and III cause Treacher Collins syndrome. *Nature Genetics*. **43** (1): 20–22.

192. Sanchez, E., Laplace-Builhé, B., Mau-Them, F.T., Richard, E., Goldenberg, A., et al. (2020) POLR1B and neural crest cell anomalies in Treacher Collins syndrome type 4. *Genetics in Medicine*. **22** (3): 547–556.

193. Watt, K.E., Macintosh, J., Bernard, G., and Trainor, P.A. (2022) RNA Polymerases I and III in development and disease. *Seminars in Cell & Developmental Biology*. doi: 10.1016/j.semcdb.2022.03.027.

194. Dixon, J., Jones, N.C., Sandell, L.L., Jayasinghe, S.M., Crane, J., et al. (2006) Tcof1/Treacle is required for neural crest cell formation and proliferation deficiencies that cause craniofacial abnormalities. *Proceedings of the National Academy of Sciences of the United States of America*. **103** (36): 13403–13408.

195. Jones, N.C., Lynn, M.L., Gaudenz, K., Sakai, D., Aoto, K., et al. (2008) Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function. *Nature Medicine*. **14** (2): 125–133.

196. Weaver, K.N., Watt, K.E.N., Hufnagel, R.B., Navajas Acedo, J., Linscott, L.L., et al. (2015) Acrofacial Dysostosis, Cincinnati Type, a mandibulofacial dysostosis syndrome with limb anomalies, is caused by POLR1A dysfunction. *American Journal of Human Genetics*. **96** (5): 765–774.

197. Kara, B., Köroğlu, Ç., Peltonen, K., Steinberg, R.C., Maraş Genç, H., et al. (2017) Severe neurodegenerative disease in brothers with homozygous mutation in POLR1A. *European Journal of Human Genetics*. **25** (3): 315–323.

198. Lata, E., Choquet, K., Sagliocco, F., Brais, B., Bernard, G., et al. (2021) RNA Polymerase III subunit mutations in genetic diseases. *Frontiers in Molecular Biosciences*. **8**: 752.

199. Lafontaine, D.L.J., Riback, J.A., Bascetin, R., and Brangwynne, C.P. (2020) The nucleolus as a multiphase liquid condensate. *Nature Reviews Molecular Cell Biology*. **22** (3): 165–182.

200. Ablasser, A., Bauernfeind, F., Hartmann, G., Latz, E., Fitzgerald, K.A., et al. (2009) RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nature Immunology*. **10** (10): 1065–1072.

201. Chiu, Y.H., MacMillan, J.B., and Chen, Z.J. (2009) RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell*. **138** (3): 576–591.

202. Hoffmann, N.A., Jakobi, A.J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., et al. (2015) Molecular structures of unbound and transcribing RNA polymerase III. *Nature*. **528** (7581): 231–236.

203. Chédin, S., Riva, M., Schultz, P., Sentenac, A., and Caries, C. (1998) The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes & Development*. **12** (24): 3857–3871.

204. Dergai, O. and Hernandez, N. (2019) How to recruit the correct RNA polymerase? Lessons from snRNA genes. *Trends in Genetics*. **35** (6): 457–469.

205. Arimbasseri, A.G., Rijal, K., and Maraia, R.J. (2013) Transcription termination by the eukaryotic RNA polymerase III. *Biochimica et Biophysica Acta*. **1829** (3–4): 318–330.

206. Wang, Q., Li, S., Wan, F., Xu, Y., Wu, Z., et al. (2021) Structural insights into transcriptional regulation of human RNA polymerase III. *Nature Structural & Molecular Biology*. **28** (2): 220–227.

207. Ramsay, E.P., Abascal-Palacios, G., Daiß, J.L., King, H., Gouge, J., et al. (2020) Structure of human RNA polymerase III. *Nature Communications*. **11**: 6409.

208. Li, L., Yu, Z., Zhao, D., Ren, Y., Hou, H., et al. (2021) Structure of human RNA polymerase III elongation complex. *Cell Research*. **31** (7): 791–800.

209. Vorländer, M.K., Khatter, H., Wetzel, R., Hagen, W.J.H., and Müller, C.W. (2018) Molecular mechanism of promoter opening by RNA polymerase III. *Nature*. **553** (7688): 295–300.

210. Abascal-Palacios, G., Ramsay, E.P., Beuron, F., Morris, E., and Vannini, A. (2018) Structural basis of RNA polymerase III transcription initiation. *Nature*. **553** (7688): 301–306.

211. Han, Y., Yan, C., Fishbain, S., Ivanov, I., and He, Y. (2018) Structural visualization of RNA polymerase III transcription machineries. *Cell Discovery*. **4** (1): 40.

212. Vorländer, M.K., Baudin, F., Moir, R.D., Wetzel, R., Hagen, W.J.H., et al. (2020) Structural basis for RNA polymerase III transcription repression by Maf1. *Nature Structural & Molecular Biology*. **27** (3): 229–232.

213. Hou, H., Li, Y., Wang, M., Liu, A., Yu, Z., et al. (2021) Structural insights into RNA polymerase III-mediated transcription termination through trapping poly-deoxythymidine. *Nature Communications*. **12**: 6135.

214. Girbig, M., Xie, J., Grötsch, H., Libri, D., Porrua, O., et al. (2022) Architecture of the yeast Pol III pre-termination complex and pausing mechanism on poly-dT termination signals. *bioRxiv*. doi: 0.1101/2022.02.28.482286.

215. Yeganeh, M. and Hernandez, N. (2020) RNA polymerase III transcription as a disease factor. *Genes & Development*. **34** (13–14): 865–882.

216. Lei, J., Chen, S., and Zhong, S. (2017) Abnormal expression of TFIIIB subunits and RNA Pol III genes is associated with hepatocellular carcinoma. *Liver research*. **1** (2): 112–120.

217. Khattar, E., Kumar, P., Liu, C.Y., Can Akincilar, S., Raju, A., et al. (2016) Telomerase reverse transcriptase promotes cancer cell proliferation by augmenting tRNA expression. *The Journal of Clinical Investigation*. **126** (10): 4045–4060.

218. Johnson, S.A.S., Dubeau, L., and Johnson, D.L. (2008) Enhanced RNA polymerase III-dependent transcription is required for oncogenic transformation. *Journal of Biological Chemistry*. **283** (28): 19184–19191.

219. Haurie, V., Durrieu-Gaillard, S., Dumay-Odelot, H., Da Silva, D., Rey, C., et al. (2010) Two isoforms of human RNA polymerase III with specific functions in cell growth and transformation. *Proceedings of the National Academy of Sciences of the United States of America*. **107** (9): 4176–4181.

220. Van Bortle, K., Marciano, D.P., Liu, Q., Chou, T., Lipchik, A.M., et al. (2022) A cancer-associated RNA polymerase III identity drives robust transcription and expression of snaR-A noncoding RNA. *Nature Communications*. **13**: 3007.

221. Liang, X., Xie, R., Su, J., Ye, B., Wei, S., et al. (2019) Inhibition of RNA polymerase III transcription by Triptolide attenuates colorectal tumorigenesis. *Journal of Experimental and Clinical Cancer Research*. **38**: 217.

222. Banerjee, S. and Saluja, A. (2015) Minnelide, a novel drug for pancreatic and liver cancer. *Pancreatology*. **15** (4): S39–S43.

223. Wu, L., Pan, J., Thoroddsen, V., Wysong, D.R., Blackman, R.K., et al. (2003) Novel small-molecule inhibitors of RNA polymerase III. *Eukaryotic Cell*. **2** (2): 256–264.

224. Filer, D., Thompson, M.A., Takhaveev, V., Dobson, A.J., Kotronaki, I., et al. (2017) RNA polymerase III limits longevity downstream of TORC1. *Nature*. **552** (7684): 263–267.

225. Bonhoure, N., Byrnes, A., Moir, R.D., Hodroj, W., Preitner, F., et al. (2015) Loss of the RNA polymerase III repressor MAF1 confers obesity resistance. *Genes & Development*. **29** (9): 934–947.

226. Carter-Timofte, M.E., Hansen, A.F., Christiansen, M., Paludan, S.R., and Mogensen, T.H. (2019) Mutations in RNA polymerase III genes and defective DNA sensing in adults with varicella-zoster virus CNS infection. *Genes and Immunity*. **20** (3): 214–223.

227. Ogunjimi, B., Zhang, S.Y., Sorensen, K.B., Skipper, K.A., Carter-Timofte, M., et al. (2017) Inborn errors in RNA polymerase III underlie severe varicella zoster virus infections. *The Journal of Clinical Investigation*. **127** (9): 3543–3556.

228. Coulombe, B., Derksen, A., Piana, R. La, Brais, B., Gauthier, M.-S., et al. (2021) POLR3-related leukodystrophy: How do mutations affecting RNA polymerase III subunits cause hypomyelination? *Faculty Reviews*. **10**: 12.

229. Perrier, S., Michell-Robinson, M.A., and Bernard, G. (2021) POLR3-related leukodystrophy: Exploring potential therapeutic approaches. *Frontiers in Cellular Neuroscience*. **14**: 487.

230. Wolf, N.I., Vanderver, A., Van Spaendonk, R.M.L., Schiffmann, R., Brais, B., et al. (2014) Clinical spectrum of 4H leukodystrophy caused by POLR3A and POLR3B mutations. *Neurology*. **83** (21): 1898–1905.

231. Gauquelin, L., Cayami, F.K., Sztriha, L., Yoon, G., Tran, L.T., et al. (2019) Clinical spectrum of POLR3-related leukodystrophy caused by biallelic POLR1C pathogenic variants. *Neurology Genetics*. **5** (6): e369.

232. Terhal, P.A., Vlaar, J.M., Middelkamp, S., Nievelstein, R.A.J., Nikkels, P.G.J., et al. (2020) Biallelic variants in POLR3GL cause endosteal hyperostosis and oligodontia. *European Journal of Human Genetics*. **28** (1): 31–39.

233. Gauquelin, L., Cayami, F.K., Sztriha, L., Yoon, G., Tran, L.T., et al. (2019) Clinical spectrum of POLR3-related leukodystrophy caused by biallelic POLR1C pathogenic variants. *Neurology Genetics*. **5** (6): e369.

234. Saitsu, H., Osaka, H., Sasaki, M., Takanashi, J.I., Hamada, K., et al. (2011) Mutations in POLR3A and POLR3B encoding RNA polymerase III subunits cause an autosomal-recessive hypomyelinating leukoencephalopathy. *American Journal of Human Genetics*. **89** (5): 644–651.

235. Daoud, H., Tétreault, M., Gibson, W., Guerrero, K., Cohen, A., et al. (2013) Mutations in POLR3A and POLR3B are a major cause of hypomyelinating leukodystrophies with or without dental abnormalities and/or hypogonadotropic hypogonadism. *Journal of Medical Genetics*. **50** (3): 194–197.

236. Thiffault, I., Wolf, N.I., Forget, D., Guerrero, K., Tran, L.T., et al. (2015) Recessive mutations in POLR1C cause a leukodystrophy by impairing biogenesis of RNA polymerase III. *Nature Communications*. **6** 7623.

237. Bernard, G., Chouery, E., Putorti, M.L., Tétreault, M., Takanohashi, A., et al. (2011) Mutations of POLR3A encoding a catalytic subunit of RNA polymerase Pol III cause a recessive hypomyelinating leukodystrophy. *American Journal of Human Genetics*. **89** (3): 415–423.

238. Dorboz, I., Dumay-Odelot, H., Boussaid, K., Bouyacoub, Y., Barreau, P., et al. (2018) Mutation in POLR3K causes hypomyelinating leukodystrophy and abnormal ribosomal RNA regulation. *Neurology Genetics*. **4** (6): e289.

239. Wambach, J.A., Wegner, D.J., Patni, N., Kircher, M., Willing, M.C., et al. (2018) Bi-allelic POLR3A loss-of-function variants cause autosomal-recessive Wiedemann-Rautenstrauch syndrome. *American Journal of Human Genetics*. **103** (6): 968–975.

240. Jay, A.M., Conway, R.L., Thiffault, I., Saunders, C., Farrow, E., et al. (2016) Neonatal progeriod syndrome associated with biallelic truncating variants in POLR3A. *American Journal of Human Genetics*. **170** (12): 3343–3346.

241. Paolacci, S., Li, Y., Agolini, E., Bellacchio, E., Arboleda-Bustos, C.E., et al. (2018) Specific combinations of biallelic POLR3A variants cause Wiedemann-Rautenstrauch syndrome. *Journal of Medical Genetics*. **55** (12): 837–846.

242. Franca, M.M., Han, X., Funari, M.F.A., Lerario, A.M., Nishi, M.Y., et al. (2019) Exome sequencing reveals the POLR3H gene as a novel cause of primary ovarian insufficiency. *The Journal of Clinical Endocrinology and Metabolism*. **104** (7): 2827–2841.

243. Saghi, M., InanlooRahatloo, K., Alavi, A., Kahrizi, K., and Najmabadi, H. (2022) Intellectual disability associated with craniofacial dysmorphism due to POLR3B mutation and defect in spliceosomal machinery. *BMC Medical Genomics*. **15** (1): 1–16.

244. Ruska, E. (1987) The development of the electron microscope and of electron microscopy. *Bioscience Reports*. **7** (8): 607–629.

245. Raimondi, V. and Grinzato, A. (2022) A basic introduction to single particles cryo-electron microscopy. *AIMS Biophysics*. **9** (1): 5–20.

246. Brenner, S. and Horne, R.W. (1959) A negative staining method for high resolution electron microscopy of viruses. *Biochimica et Biophysica Acta*. **34** (C): 103–110.

247. Dubochet, J. and McDowall, A.W. (1981) Vitrification of pure water for electron microscopy. *Journal of Microscopy*. **124** (3): 3–4.

248. Knapek, E. and Dubochet, J. (1980) Beam damage to organic material is considerably reduced in cryo-electron microscopy. *Journal of Molecular Biology*. **141** (2): 147–161.

249. Adrian, M., Dubochet, J., Lepault, J., and McDowall, A.W. (1984) Cryo-electron microscopy of viruses. *Nature*. **308** (5954): 32–36.

250. Frank, J., Shimkin, B., and Dowse, H. (2018) SPIDER - A modular software system for electron image processing. in: Single-Particle Cryo-Electron Microscopy, pp. 94–108.

251. van Heel, M. and Frank, J. (1981) Use of multivariates statistics in analysing the images of biological macromolecules. *Ultramicroscopy*. **6** (1): 187–194.

252. McMullan, G., Chen, S., Henderson, R., and Faruqi, A.R. (2009) Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy*. **109** (9): 1126–1143.

253. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., et al. (2020) Single-particle cryo-EM at atomic resolution. *Nature*. **587** (7832): 152–156.

254. Kühlbrandt, W. (2014) The resolution revolution. *Science*. **343** (6178): 1443–1444.

255. Thompson, R.F., Walker, M., Siebert, C.A., Muench, S.P., and Ranson, N.A. (2016) An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods*. **100**: 3–15.

256. Aebi, U. and Pollard, T.D. (1987) A glow discharge unit to render electron microscope grids and other surfaces hydrophilic. *Journal of Electron Microscopy Technique*. **7** (1): 29–33.

257. Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., et al. (2012) Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology*. **177** (3): 630.

258. Mastronarde, D.N. (2005) Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*. **152** (1): 36–51.

259. Wade, R.H. (1992) A brief look at imaging and contrast transfer. *Ultramicroscopy*. **46** (1–4): 145–156.

260. Huang, Z., Baldwin, P.R., Mullapudi, S., and Penczek, P.A. (2003) Automated determination of parameters describing power spectra of micrograph images in electron microscopy. *Journal of Structural Biology*. **144** (1–2): 79–94.

261. Tegunov, D. and Cramer, P. (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods*. **16** (11): 1146–1152.

262. Van Heel, M. (1982) Detection of objects in quantum-noise-limited images. *Ultramicroscopy*. **7** (4): 331–341.

263. Scheres, S.H.W. (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. **180** (3): 519–530.

264. Scheres, S.H.W. (2016) Processing of structurally heterogeneous cryo-EM data in RELION. *Methods in Enzymology*. **579**: 125–157.

265. Sorzano, C.O.S., Jiménez-Moreno, A., Maluenda, D., Martínez, M., Ramírez-Aportela, E., et al. (2022) On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy. *Acta Crystallographica Section D Structural Biology*. **78** (4): 410–423.

266. Danev, R., Yanagisawa, H., and Kikkawa, M. (2019) Cryo-electron microscopy methodology: Current aspects and future directions. *Trends in Biochemical Sciences*. **44** (10): 837–848.

267. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315** (5819): 1709–1712.

268. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. **337** (6096): 816–821.

269. Terns, M.P. and Terns, R.M. (2011) CRISPR-based adaptive immune systems. *Current Opinion in Microbiology*. **14** (3): 321–327.

270. Bhaya, D., Davison, M., and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation. *Annual Review of Genetics*. **45**: 273–297.

271. Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. **482** (7385): 331–338.

272. Behler, J. and Hess, W.R. (2020) Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods*. **172**: 12–26.

273. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. **471** (7340): 602–607.

274. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*. **329** (5997): 1355–1358.

275. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakatura, A. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*. **169** (12): 5429–5433.

276. Jansen, R., Van Embden, J.D.A., Gaastra, W., and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*. **43** (6): 1565–1575.

277. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*. **339** (6121): 819–823.

278. Wang, J.Y., Pausch, P., and Doudna, J.A. (2022) Structural biology of CRISPR–Cas immunity and genome editing enzymes. *Nature Reviews Microbiology*. doi: 10.1038/s41579-022-00739-4.

279. Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., et al. (2018) PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biology*. **16** (4): 504–517.

280. Bravo, J.P.K., Liu, M. Sen, Hibshman, G.N., Dangerfield, T.L., Jung, K., et al. (2022) Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature*. **603** (7900): 343–347.

281. Liu, G., Lin, Q., Jin, S., and Gao, C. (2022) The CRISPR-Cas toolbox and gene editing technologies. *Molecular Cell*. **82** (2): 333–347.

282. Nambiar, T.S., Baudrier, L., Billon, P., and Ciccia, A. (2022) CRISPR-based genome editing through the lens of DNA repair. *Molecular Cell*. **82** (2): 348–388.

283. Knott, G.J. and Doudna, J.A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science*. **361** (6405): 866–869.

284. Koch, B., Nijmeijer, B., Kueblbeck, M., Cai, Y., Walther, N., et al. (2018) Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing. *Nature Protocols*. **13** (6): 1465–1487.

285. Concordet, J.P. and Haeussler, M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*. **46** (W1): W242–W245.

286. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., et al. (2013) Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*. **8** (11): 2281–2308.

287. Hadden, J.M., Convery, M.A., Déclais, A.C., Lilley, D.M.J., and Phillips, S.E.V. (2001) Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I. *Nature Structural & Molecular Biology*. **8** (1): 62–67.

288. Sentmanat, M.F., Peters, S.T., Florian, C.P., Connelly, J.P., and Pruett-Miller, S.M. (2018) A survey of validation strategies for CRISPR-Cas9 editing. *Scientific Reports*. **8** (1): 888.

289. Babon, J.J., McKenzie, M., and Cotton, R.G.H. (2003) The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. *Molecular Biotechnology 2003 23:1*. **23** (1): 73–81.

290. Lin, Y.C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., et al. (2014) Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*. **5**: 4767.

291. IzuMi, M., Miyazawa, H., Kamakura, T., Yamaguchi, I., Endo, T., et al. (1991) Blasticidin S-resistance gene (bsr): A novel selectable marker for mammalian cells. *Experimental Cell Research*. **197** (2): 229–233.

292. Liang, X., Potter, J., Kumar, S., Ravinder, N., and Chesnut, J.D. (2017) Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of Biotechnology*. **241**: 136–146.

293. Zhao, J., Hyman, L., and Moore, C. (1999) Formation of mRNA 3′ ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews*. **63** (2): 405–445.

294. Cordingley, M.G., Callahan, P.L., Sardana, V. V., Garsky, V.M., and Colonno, R.J. (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *Journal of Biological Chemistry*. **265** (16): 9062–9065.

295. Cordingley, M.G., Register, R.B., Callahan, P.L., Garsky, V.M., and Colonno, R.J. (1989) Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in Escherichia coli. *Journal of Virology*. **63** (12): 5037–5045.

296. Yeh, C.D., Richardson, C.D., and Corn, J.E. (2019) Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology*. **21** (12): 1468–1478.

297. Pontvianne, F., Carpentier, M.C., Durut, N., Pavlištová, V., Jaške, K., et al. (2016) Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D Organization of the A. thaliana genome. *Cell Reports*. **16** (6): 1574–1587.

298. Maiser, A., Dillinger, S., Längst, G., Schermelleh, L., Leonhardt, H., et al. (2020) Super-resolution in situ analysis of active ribosomal DNA chromatin organization in the nucleolus. *Scientific Reports*. **10** (1): 7462.

299. Passmore, L.A. and Russo, C.J. (2016) Specimen preparation for high-resolution cryo-EM. *Methods in Enzymology*. **579**: 51–86.

300. Weissmann, F., Petzold, G., VanderLinden, R., Huis In't Veld, P.J., Brown, N.G., et al. (2016) BiGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*. **113** (19): E2564–E2569.

301. Comai, L., Tanese, N., and Tjian, R. (1992) The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell*. **68** (5): 965–976.

302. Comai, L., Zomerdijk, J.C.B.M., Beckmann, H., Zhou, S., Admon, A., et al. (1994) Reconstitution of transcription factor SL1: exclusive binding of TBP by SL1 or TFIID subunits. *Science*. **266** (5193): 1966–1972.

303. Gorski, J.J., Pathak, S., Panov, K., Kasciukovic, T., Panova, T., et al. (2007) A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. *The EMBO Journal*. **26** (6): 1560–1568.

304. Denissov, S., Van Driel, M., Voit, R., Hekkelman, M., Hulsen, T., et al. (2007) Identification of novel functional TBP-binding sites and general factor repertoires. *The EMBO Journal*. **26** (4): 944–954.

305. Mcconnell, J., Gazdar, A., Nisen, P., Leitch, M., Word, R.A., et al. (1994) Assembly of transcriptionally active RNA polymerase I initiation factor SL1 from recombinant subunits. *Science*. **266** (5193): 2015–2018.

306. Heix, J., Vente, A., Voit, R., Budde, A., Michaelidis, T.M., et al. (1998) Mitotic silencing of human rRNA synthesis: Inactivation of the promoter selectivity factor SL1 by cdc2/cyclin B-mediated phosphorylation. *The EMBO Journal*. **17** (24): 7373–7381.

307. Muth, V., Nadaud, S., Grummt, I., and Voit, R. (2001) Acetylation of TAFI68, a subunit of TIF-IB/SL1, activates RNA polymerase I transcription. *The EMBO Journal*. **20** (6): 1353–1362.

308. Ide, S., Imai, R., Ochi, H., and Maeshima, K. (2020) Transcriptional suppression of ribosomal DNA with phase separation. *Science Advances*. **6** (42): 5953–5967.

309. Ho, C.M., Beck, J.R., Lai, M., Cui, Y., Goldberg, D.E., et al. (2018) Malaria parasite translocon structure and mechanism of effector export. *Nature*. **561** (7721): 70–75.

310. Daiß, J.L., Pilsl, M., Straub, K., Bleckmann, A., Höcherl, M., et al. (2021) The human RNA polymerase I structure reveals an HMG-like transcription factor docking domain specific to metazoans. *bioRxiv*. doi: 10.1101/2021.12.22.473891.

311. Vannini, A. and Cramer, P. (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular Cell*. **45** (4): 439–446.

312. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016) Structure of transcribing mammalian RNA polymerase II. *Nature*. **529** (7587): 551–554.

313. Mayorov, A., Dal Peraro, M., Abriata, L.A., and Wilke, C. (2019) Active site-induced evolutionary constraints follow fold polarity principles in soluble globular enzymes. *Molecular Biology and Evolution*. **36** (8): 1728–1733.

314. Echave, J., Spielman, S.J., and Wilke, C.O. (2016) Causes of evolutionary rate variation among protein sites. *Nature Reviews Genetics*. **17** (2): 109–121.

315. Kang, J.Y., Mishanina, T. V., Bellecourt, M.J., Mooney, R.A., Darst, S.A., et al. (2018) RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. *Molecular Cell*. **69** (5): 802–815.

316. Wang, Q., Li, S., Wan, F., Xu, Y., Wu, Z., et al. (2021) Structural insights into transcriptional regulation of human RNA polymerase III. *Nature Structural & Molecular Biology*. **28** (2): 220–227.

317. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., et al. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Science*. **27** (1): 112–128.

318. He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., et al. (2016) Near-atomic resolution visualization of human transcription promoter opening. *Nature*. **533** (7603): 359–365.

319. Sadian, Y., Baudin, F., Tafur, L., Murciano, B., Wetzel, R., et al. (2019) Molecular insight into RNA polymerase I promoter recognition and promoter melting. *Nature Communications*. **10** (1): 1–13.

320. Kay, B.K., Williamson, M.P., and Sudol, M. (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal*. **14** (2): 231–241.

321. Yamamoto, K., Yamamoto, M., Hanada, K., Nogi, Y., Matsuyama, T., et al. (2004) Multiple protein-protein interactions by RNA polymerase I-associated factor PAF49 and role of PAF49 in rRNA transcription. *Molecular and Cellular Biology*. **24** (14): 6338–6349.

322. Penrod, Y., Rothblum, K., Cavanaugh, A., and Rothblum, L.I. (2015) Regulation of the association of the PAF53/PAF49 heterodimer with RNA polymerase I. *Gene*. **556** (1): 61–67.

323. Albert, B., Léger-Silvestre, I., Normand, C., Ostermaier, M.K., Pérez-Fernández, et al. (2011) RNA polymerase I-specific subunits promote polymerase clustering to enhance the rRNA gene transcription cycle. *Journal of Cell Biology*. **192** (2): 277–293.

324. Aiyer, S., Zhang, C., Baldwin, P.R., and Lyumkis, D. (2021) Evaluating local and directional resolution of cryo-EM density maps. *Methods in Molecular Biology*. **2215**: 161.

325. Darrière, T., Pilsl, M., Sarthou, M.-K., Chauvier, A., Genty, T., et al. (2019) Genetic analyses led to the discovery of a super-active mutant of the RNA polymerase I. *PLOS Genetics*. **15** (5): e1008157.

326. Blattner, C., Jennebach, S., Herzog, F., Mayer, A., Cheung, A.C.M., et al. (2011) Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes & Development*. **25** (19): 2093–2105.

327. Moorefield, B., Greene, E.A., and Reeder, R.H. (2000) RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human. *Proceedings of the National Academy of Sciences of the United States of America*. **97** (9): 4724–4729.

328. Yuan, X., Zhao, J., Zentgraf, H., Hoffmann-Rohrer, U., and Grummt, I. (2002) Multiple interactions between RNA polymerase I, TIF-IA and TAF I subunits regulate preinitiation complex assembly at the ribosomal gene promoter. *EMBO Reports*. **3** (11): 1082–1087.

329. Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., et al. (2008) A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*. **105** (31): 10762–10767.

330. Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D., and Cramer, P. (2011) Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *Journal of Biological Chemistry*. **286** (21): 18701–18707.

331. Cheung, A.C.M. and Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*. **471** (7337): 249–253.

332. Lefèvre, S., Dumay-Odelot, H., El-Ayoubi, L., Budd, A., Legrand, P., et al. (2011) Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. *Nature Structural & Molecular Biology*. **18** (3): 352–358.

333. Smith, M. (2017) DNA sequence analysis in clinical medicine, proceeding cautiously. *Frontiers in Molecular Biosciences*. **4**: 24.

334. Yee, N.S., Gong, W., Huang, Y., Lorent, K., Dolan, A.C., et al. (2007) Mutation of RNA Pol III subunit RPC2/POLR3B leads to deficiency of subunit RPC11 and disrupts zebrafish digestive development. *PLOS Biology*. **5** (11): e312.

335. Kraoua, I., Karkar, A., Drissi, C., Benrhouma, H., Klaa, H., et al. (2019) Novel POLR1C mutation in RNA polymerase III-related leukodystrophy with severe myoclonus and dystonia. *Molecular Genetics & Genomic Medicine*. **7** (9): e914.

336. Gehre, M., Buccitelli, C., Diaz, N., Korbel, J., and Noh, K.-M. (2019) Efficient strategies to detect genome editing and integrity in CRISPR-Cas9 engineered ESCs. *bioRxiv*. doi: 10.1101/635151.

337. Murray, J.M., Appleton, P.L., Swedlow, J.R., and Waters, J.C. (2007) Evaluating performance in three-dimensional fluorescence microscopy. *Journal of Microscopy*. **228** (3): 390–405.

338. Fabre, B., Lambour, T., Bouyssié, D., Menneteau, T., Monsarrat, B., et al. (2014) Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. *EuPA Open Proteomics*. **4**: 82–86.

339. Shaw, G., Morse, S., Ararat, M., and Graham, F.L. (2002) Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK293 cells. *The FASEB Journal*. **16** (8): 869–871.

340. Merheb, E., Cui, M.-H., Dubois, J.C., Branch, C.A., Gulinello, M., et al. (2021) Defective oligodendrocyte development and function in an RNA polymerase III mutant leukodystrophic mouse. *bioRxiv*. doi: 10.1101/2020.12.09.418657.

341. Van Den Ent, F. and Löwe, J. (2006) RF cloning: A restriction-free method for inserting target genes into plasmids. *Journal of Biochemical and Biophysical Methods*. **67** (1): 67–74.

342. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. **6** (5): 343–345.

343. Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., et al. (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology*. **208**: 44–53.

344. Pronobis, M.I., Deuitch, N., and Peifer, M. (2016) The Miraprep: A protocol that uses a miniprep kit and provides maxiprep yields. *PLOS ONE*. **11** (8): e0160509.

345. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017) CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*. **14** (3): 290–296.

346. Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*. **157** (1): 38–46.

347. Zivanov, J., Nakane, T., and Scheres, S.H.W. (2020) Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *Acta Crystallographica Section D Structural Biology*. **7** (2): 253–267.

348. Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., et al. (2017) MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*. **14** (4): 331–332.

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

350. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., et al. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. **25** (13): 1605–1612.

351. Rosenthal, P.B. and Henderson, R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*. **333** (4): 721–745.

352. Nakane, T., Kimanius, D., Lindahl, E., and Scheres, S.H.W. (2018) Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife*. **7**: e36861.

353. Liebschner, D., Afonine, P. V., Baker, M.L., Bunkoczi, G., Chen, V.B., et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D Structural Biology*. **75** (10): 861–877.

354. Zi Tan, Y., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., et al. (2017) Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature Methods*. **14** (8): 793–796.

355. Ramírez-Aportela, E., Vilas, J.L., Glukhova, A., Melero, R., Conesa, P., et al. (2020) Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics*. **36** (3): 765–772.

356. Jakobi, A.J., Wilmanns, M., and Sachse, C. (2017) Model-based local density sharpening of cryo-EM maps. *eLife*. **6**: e27131.

357. Burnley, T., Palmer, C.M., and Winn, M. (2017) Recent developments in the CCP-EM software suite. *Acta Crystallographica Section D Structural Biology*. **73** (6): 469–477.

358. de la Rosa-Trevín, J.M., Quintana, A., del Cano, L., Zaldívar, A., Foche, I., et al. (2016) Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of Structural Biology*. **195** (1): 93–99.

359. Bienert, S., Waterhouse, A., De Beer, T.A.P., Tauriello, G., Studer, G., et al. (2017) The SWISS-MODEL repository - New features and functionality. *Nucleic Acids Research*. **45** (D1): D313–D319.

360. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. **10** (6): 845–858.

361. Casañal, A., Lohkamp, B., and Emsley, P. (2020) Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Science*. **29** (4): 1055–1064.

362. Nicholls, R.A., Fischer, M., Mcnicholas, S., and Murshudov, G.N. (2014) Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallographica Section D Structural Biology*. **70** (9): 2487–2499.

363. Buchan, D.W.A. and Jones, D.T. (2019) The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Research*. **47** (W1): W402–W407.

364. Afonine, P. V., Poon, B.K., Read, R.J., Sobolev, O. V., Terwilliger, T.C., et al. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D Structural Biology*. **74** (6): 531–544.

365. Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Structural Biology*. **66** (2): 213–221.

366. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*. **35** (2): W375–W383.

367. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., et al. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*. **27** (1): 14–25.

368. Kayikci, M., Venkatakrishnan, A.J., Scott-Brown, J., Ravarani, C.N.J., Flock, T., et al. (2018) Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. *Nature Structural & Molecular Biology*. **25** (2): 185–194.

369. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*. **13** (11): 2498–2504.

370. Csardi, Gabor; Nepusz, T. (2006) The Igraph software package for complex network research. *InterJournal, Complex Systems*. **1695** (5): 1–9.

371. Mitternacht, S. (2016) FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research*. **5**: 189.

372. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nature Methods*. **9** (7): 676–682.

373. Timmons, M., Tsokos, M., Asab, M.A., Seminara, S.B., Zirzow, G.C., et al. (2006) Peripheral and central hypomyelination with hypogonadotropic hypogonadism and hypodontia. *Neurology*. **67** (11): 2066–2069.

374. Shimojima, K., Shimada, S., Tamasaki, A., Akaboshi, S., Komoike, Y., et al. (2014) Novel compound heterozygous mutations of POLR3A revealed by whole-exome sequencing in a patient with hypomyelination. *Brain and Development*. **36** (4): 315–321.

375. Wolf, N.I., Harting, I., Boltshauser, E., Wiegand, G., Koch, M.J., et al. (2005) Leukoencephalopathy with ataxia, hypodontia, and hypomyelination. *Neurology*. **64** (8): 1461–1464.

376. Arimbasseri, A.G. and Maraia, R.J. (2016) RNA polymerase III advances: Structural and tRNA functional views. *Trends in Biochemical Sciences*. **41** (6): 546–559.

377. Wolf, N.I., Harting, I., Innes, A.M., Patzer, S., Zeitler, P., et al. (2007) Ataxia, delayed dentition and hypomyelination: A novel leukoencephalopathy. *Neuropediatrics*. **38** (2): 64–70.

378. Arimbasseri, A.G. and Maraia, R.J. (2015) Mechanism of transcription termination by RNA polymerase III utilizes a non-template strand sequence-specific signal element. *Molecular Cell*. **58** (6): 1124–1132.

379. Terao, Y., Saitsu, H., Segawa, M., Kondo, Y., Sakamoto, K., et al. (2012) Diffuse central hypomyelination presenting as 4H syndrome caused by compound heterozygous mutations in POLR3A encoding the catalytic subunit of polymerase III. *Journal of the Neurological Sciences*. **320** (1–2): 102–105.

380. Potic, A., Brais, B., Choquet, K., Schiffmann, R., and Bernard, G. (2012) 4H syndrome with late-onset growth hormone deficiency caused by POLR3A mutations. *Archives of Neurology*. **69** (7): 920–923.

381. Djordjevic, D., Pinard, M., Gauthier, M.S., Smith-Hicks, C., Hoffman, T.L., et al. (2021) De novo variants in POLR3B cause ataxia, spasticity, and demyelinating neuropathy. *The American Journal of Human Genetics*. **108** (1): 186–193.

382. Tétreault, M., Choquet, K., Orcesi, S., Tonduti, D., Balottin, U., et al. (2011) Recessive mutations in POLR3B, encoding the second largest subunit of Pol III, cause a rare hypomyelinating leukodystrophy. *American Journal of Human Genetics*. **89** (5): 652–655.