UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE CIENCIAS

DEPARTAMENTO DE BIOLOGÍA MOLECULAR

THE CRYO-EM STRUCTURE OF RNA POLYMERASE I STALLED AT UV LIGHT-INDUCED DAMAGE UNRAVELS A NEW MOLECULAR MECHANISM TO IDENTIFY LESIONS ON RIBOSOMAL DNA

TESIS DOCTORAL

AUTOR

Marta Maria Sanz Murillo

Licenciada en Bioquímica

DIRECTOR

Carlos Fernández-Tornero

Centro de Investigaciones Biológicas (CIB) - CSIC

Madrid, 2019

AGRADECIMIENTOS

Han sido tanta gente la que ha estado conmigo en este proceso que al llegar al final de esta etapa y escribir estos agradecimientos se me hace difícil resumirlos en unas pocas páginas, pero vamos a ello…

A mi director de tesis, el Dr. Carlos Fernández-Tornero, por permitirme realizar este trabajo en su laboratorio, por su supervisión y por lo que me ha enseñado a lo largo de estos años. Entré al laboratorio siendo una estudiante de Licenciatura en Bioquímica que había cogido una pipeta en contadas ocasiones y ahora salgo con esta tesis "bajo el brazo". Gracias por tu paciencia para enseñarme, por tu dedicación y por tus consejos.

A mis compañeros de laboratorio. A Eva, Nuria, Rocío, Jaime, Nicholas, Miguel, Srdja, Carolina, Álvaro Gándara, Álvaro Ras, Alicia, Adrián y Margarita. Gracias por estar conmigo, por aconsejarme, por compartir los días buenos y los no tan buenos y por los ratitos de risas en el laboratorio.

Gracias al grupo de Biología Estructural de Proteínas; a Antonio, Elena, Mer, Fede y Javier, por sus consejos y su ayuda. En especial tengo que dar las gracias a mi amiga Irene, que empezamos en esto a la par y hemos compartido el camino juntas. Los días de experimentos y escritura en la biblioteca no hubieran sido iguales sin ti. Gracias por compartir conmigo los "after lab" y los domingos de Rastro.

Al Profesor Dong Wang por permitirme hacer una estancia en su laboratorio en la Universidad de San Diego, por su ayuda y por sus valiosas recomendaciones. Gracias al Dr. Jun Xu por enseñarme a realizar los experimentos, por guiarme en el trabajo diario, por su amistad y sus consejos (no los olvidaré).

Al ministerio de Ciencia, Innovación y Universidades por la ayuda BES-2014-070708 para la realización de esta tesis.

Al grupo de Complejos Macromoleculares en la respuesta a Daños en el ADN del CNIO por sus consejos para el procesamiento de datos; en especial a Andrés.

A Rafa, Begoña y Fernando del Servicio de Microscopía Electrónica del CIB.

A la Dra. Olga Calvo por acogerme en su laboratorio y su paciencia para enseñarme.

A mis amigos con los que compartí años de estudio (y fiestas) en la Universidad; Borja, Pablo, Celsa, Vicky, Pablito y Bea. Aunque cada uno esté en un sitio y no nos veamos tanto como antes, compartir esos años con vosotros no se me olvidarán nunca.

A mis amigos; Miriam, Miguel, Alberto, Isis, Mamen, María, Marcos, Danielito, Oscar, Miguel, Daniel, Carmen y Javi. Gracias por estar ahí y por sacarme una sonrisa siempre. Por hacerme ver que los problemas a veces no son para tanto.

A Isabel, Carlos, João y a toda la familia portuguesa.

A las personas que se fueron, pero en las que pienso cada día. A mis abuelos Pilar, Felipe, Lorenza y Román. A mi tía Mari Nieves.

A mi hermana Sara y a mi cuñado Daniel. Gracias por estar siempre, no tengo palabras suficientes para describir lo que significáis para mí, sois increíbles y os quiero muchísimo.

A Jaime, que lo encontré por casualidad y se convirtió en alguien imprescindible en mi vida. Gracias por estar a mi lado, animándome a no rendirme y empujándome a continuar. Gracias por ayudarme, por quererme, por todo…

A mis padres, Flora y Goyo, porque sin ellos como dice la canción de Amaral "no soy nada". Gracias por haber sido el mejor ejemplo que podía tener, por cuidarme, por ayudarme, por aguantarme los días malos y sentir mis alegrías como vuestras. Esta tesis no es para que sintáis orgullo de vuestra hija; yo estoy orgullosa de tener unos padres como vosotros. Os quiero con toda mi alma.

Gracias a todos los que me han ayudado a llegar hasta aquí y que forman parte de mi vida día a día. Sin todos vosotros no sería la persona que soy.

SUMMARY

In eukaryotic cells, three RNA polymerases transcribe the genome, each specialized in transcribing a specific set of genes. Pol II synthesizes mRNA, Pol III produces short untranslated RNAs and Pol I transcribes ribosomal DNA (rDNA). The latter produces the rRNA precursor, which after maturation constitutes the backbone of the ribosome. Pol I accounts for approximately 60% of the total transcriptional activity in growing cells and also carries out the supervision of rDNA integrity. Therefore, it is a key determinant for the control of the normal function of the cell. Environmental threats can generate DNA lesions that are cytotoxic for the cell and one of the most known is UV-light. The principal DNA damage produced by this external agent is *cis-syn* cyclobutane pyrimidine dimers (CPDs), a bulky DNA lesion that can introduce distortions in the DNA helix, thus obstructing fundamental processes such as transcription. The main goal of this Ph.D. Thesis is understanding the structural basis of Pol I stalled at UV light-induced DNA damage. The principal contribution is the cryo-EM structure at 3.6 Å resolution and the derived atomic model of Pol I in elongation complex containing a CPD lesion at the DNA TS. This structure shows that the CPD lesion induces an early translocation intermediate, along with several conformational rearrangements in Pol I structural elements inside the DNA binding cleft, which contribute to enzyme stalling. The structure revealed that the BH residue R1015 plays a relevant role for enzyme arresting, which was confirmed by mutational analysis using *E.coli* RNA polymerase as a model system. *In vitro* transcription assays comparing the Pol I and Pol II behavior in the presence of CPD reveal that, while Pol II can slowly bypass the lesion, Pol I stalls right before the lesion due to the balance between a slow nucleotide incorporation and a fast-intrinsic RNA cleavage activity. Altogether, our results reveal the molecular mechanism of Pol I stalling at CPD lesions, which is distinct from Pol II arrest. This PhD Thesis opens the avenue to unravel the molecular mechanisms underlying cell endurance to lesions on rDNA.

PRESENTACIÓN

En las células eucariotas, tres ARN polimerasas (Pols) transcriben el genoma, y cada una es especialista en transcribir un conjunto específico de genes. Pol II sintetiza ARNm, Pol III produce ARNs cortos no transcritos y Pol I transcribe el ADN ribosomal (ADNr). Éste último produce el ARNr precursor, el cual después de su procesamiento constituye el esqueleto del ribosoma. Pol I representa aproximadamente el 60% de la actividad transcripcional total en células en crecimiento y además lleva a cabo la supervisión de la integridad del rADN. Por lo tanto, es una clave determinante para el control del normal funcionamiento de la célula. Las amenazas medioambientales pueden generar lesiones en el ADN que son citotóxicas para las células, y una de las más conocidas es la luz ultravioleta (UV). El principal daño en el ADN producido por este agente externo es el *cis-syn* dímero de pirimidina ciclobutano (DPC), una lesión voluminosa en el ADN, la cual puede introducir distorsiones en la hélice de ADN y, por lo tanto, puede obstruir procesos fundamentales como la transcripción. El principal objetivo de esta tesis doctoral es entender las bases estructurales de Pol I bloqueado por un daño en el ADN inducido por la luz UV. La principal contribución es la estructura por crio microscopía electrónica a 3.6 Å de resolución y el modelo atómico derivado de Pol I en complejo de elongación conteniendo una lesión DPC en la hebra molde de ADN. Esta estructura muestra que la lesión de DPC induce un intermedio de translocación temprano que, sumado a varias reorganizaciones conformacionales de elementos estructurales de Pol I dentro de la hendidura donde se produce la unión del ADN, contribuyen a detener a la enzima. La estructura reveló que el residuo de arginina 1015 de la hélice puente es importante para el bloqueo de la enzima, lo que fue confirmado por análisis mutacionales usando la ARN polimerasa de *E.coli* como sistema modelo. Los ensayos de transcripción *in vitro* comparando el comportamiento de Pol I y Pol II en presencia de DPC revelan que mientras Pol II puede saltar lentamente la lesión, Pol I se bloquea delante de la lesión debido al balance entre una incorporación de nucleótidos lenta y una actividad intrínseca de corte del ARN rápida. En conjunto, nuestros resultados revelan el mecanismo molecular de Pol I bloqueado por una lesión DPC, el cual es diferente de Pol II detenido. Esta tesis doctoral abre una vía para descifrar los mecanismos moleculares que subyacen en la resistencia celular a las lesiones en el ADNr.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Abs_{260nm}	absorbance at 260 nanometers
Abs_{280nm}	absorbance at 280 nanometers
APS	ammonium persulfate
Un	amount of enzyme required to digest 1 µg of λ DNA (HindIII digest) in 1 hour at 37°C in a total reaction volume of 50 µl
U	amount of enzyme that will incorporate 10 nanomol of dNTP into acid insoluble material in 30 minutes at 74°C
°C	degree Celsius
g	gram
g	gravitational force
h	hour
kDa	kilodalton
keV	kiloelectronvolt
kV	kilovolt
L	litre
µg	microgram
µL/min	microliter per minute
µm	micrometer
µM	micromolar
mg	milligram
min	minute
mL	milliliter
mm	millimeter
mM	millimolar
min	minute
M	molar
nm	nanometer
nt	nucleotide
OD_{600nm}	optical density at 600 nanometers
% (v/v)	percentage of a liquid product volume per volume of solution using the same volume units
% (w/v)	percentage of weight of solute in grams per volume of solution in milliliters
PDB	Protein Data Bank
pmol	picomole
RO	run-off
rpm	revolutions per minute
RT	room temperature
s	second
TEMED	tetramethylethylenediamine
UV	ultraviolet
V	volt

INTRODUCTION

1. Transcription

The Central Dogma of Molecular Biology (Crick 1970) describes how genetic information is transferred from DNA to RNA and, eventually, to proteins in a one-way direction. The first sequential step, where DNA is used as template to obtain the different types of RNA (Perdew 2006), is known as transcription and occurs in all living organisms. RNA molecules can directly play a role inside the cell, for instance constituting the backbone of the ribosome, or can encode information that is used for protein synthesis. The latter process is known as translation and constitutes the second stage in gene expression.

Together with the ribosome, which translates the RNA chains, the central component of the gene expression is RNA polymerase (RNAP), responsible for DNA transcription. RNAP utilizes the DNA strand as template to synthesize RNA molecules, using ribonucleoside triphosphates. The RNA monomers are composed by a common backbone (the sugar ribose and three phosphate groups) and one of the four nitrogenous bases: adenine (A), guanine (G), cytosine (C) or uracil (U). RNAP extends the RNA chain in the 5' to 3' direction by the formation of phosphodiester bonds between the 3'-OH end of the RNA strand and the α-phosphate of the ribonucleotide, releasing a pyrophosphate ion.

While a single-subunit RNAP encoded by bacteriophages T7 or SP6 can directly recognize the promoter site without the presence of auxiliary regulatory factors (Cheetham and Steitz 2000, Steitz 2009), bacterial, archaeal and eukaryotic multi-subunit RNAPs are highly regulated with auxiliary proteins, known as transcription factors (TFs). TFs modulate the RNAPs activities during the three phases of the transcription cycle: initiation, elongation and termination. Besides TFs, other molecules such as RNA or single nucleotides can also modulate transcription. Bacterial, archaeal and eukaryotic RNAPs are highly conserved in evolution, showing a common structural framework, and operate by related molecular mechanisms (Werner et al 2007).

Specific DNA sequences where the transcription of a gene is initiated are known as promoters and are located near the transcription start sites (TSSs). Transcription initiation commences with the sequential assembly of TFs and RNAPs at promoter DNA to form the pre-initiation complex (PIC) (Murakami et al 2013). Once the macromolecular assembly is ready, the PIC transitions from the closed complex (CC), where the DNA is double-stranded, to the open complex (OC) in which the DNA is melted (Sainsbury et al

2015, He et al 2016, Plaschka et al 2016, Sadian et al 2017). The template strand (TS), which is defined as the DNA chain used by RNAP to attach complementary bases during transcription, reaches the active site. Then, RNAP starts to move along the TS to produce the first phosphodiester bonds. At this stage, the complex is termed initially transcribing complex (ITC) (Liu et al 2011). When the transcript length reaches about 25 nt, RNAP is released from the promoter and elongation starts. The melted DNA region, the synthesized RNA molecule and the enzyme form the elongation complex (EC). The termination step is also highly regulated. In this stage, the EC stops due to different signals, such as termination TFs or termination signal sequences. RNAP is disassembled from the DNA to re-initiate transcription, the DNA recovers the double helix structure and the RNA chain is released.

In prokaryotes, transcription of the entire genome is carried out by a single RNAP. This process takes place in the cytoplasm and is coupled to translation. Eukaryotic cells use three different RNAPs and each of them is specialized in the transcription of distinct gene types. The RNA chains are synthesized in the cell nucleus and their production is not coupled to translation. Most often, immature RNA requires processing in order to be exported to the cytoplasm. Yeast RNA polymerase I (Pol I) synthesizes the ribosomal RNA precursor (35S rRNA), later processed to mature rRNAs (25S, 18S and 5.8S), which form the ribosome scaffold. RNA polymerase II (Pol II) mainly transcribes mRNAs and RNA polymerase III (Pol III) produces short untranslated RNAs such as tRNAs and the 5S rRNA, which also forms part of the ribosome.

Archaeal transcription is a hybrid of eukaryotic and bacterial transcriptional characteristics. RNA synthesis takes place in the cytoplasm, is coupled to translation and all genes are transcribed by a single RNAP. However, the composition and structure of the RNAP, as well as the most important TFs, are similar to the eukaryotic kingdom.

2. RNA Polymerases

2.1. Bacterial RNA Polymerase

Bacterial RNA polymerase (bRNAP) carries out the synthesis of all cellular RNA and is composed by five subunits (β', β, two identical α subunits and ω) (Murakami et al 2002), which form the core. In spite of being catalytically active, the bRNAP core must bind sigma factor (σ) to constitute the bRNAP holoenzyme and initiate transcription. Single-

polypeptide σ is required to locate the promoter DNA and assists bRNAP to melt the DNA duplex and form the OC (Campbell et al 2002, Bae et al 2015). β' and β are the largest subunits and form a characteristic structure known as *crab-claw* (Zhang et al 1999), conforming a central cleft and the active site (Figure 1). The two α subunits interact with promoter DNA and are involved in heterodimer β'β assembly. Subunit ω is the smallest and it is involved in assembly and conformational maintenance of bRNAP. It is the only subunit that can be deleted without affecting the survival of the organism (Mathew and Chatterji 2006).

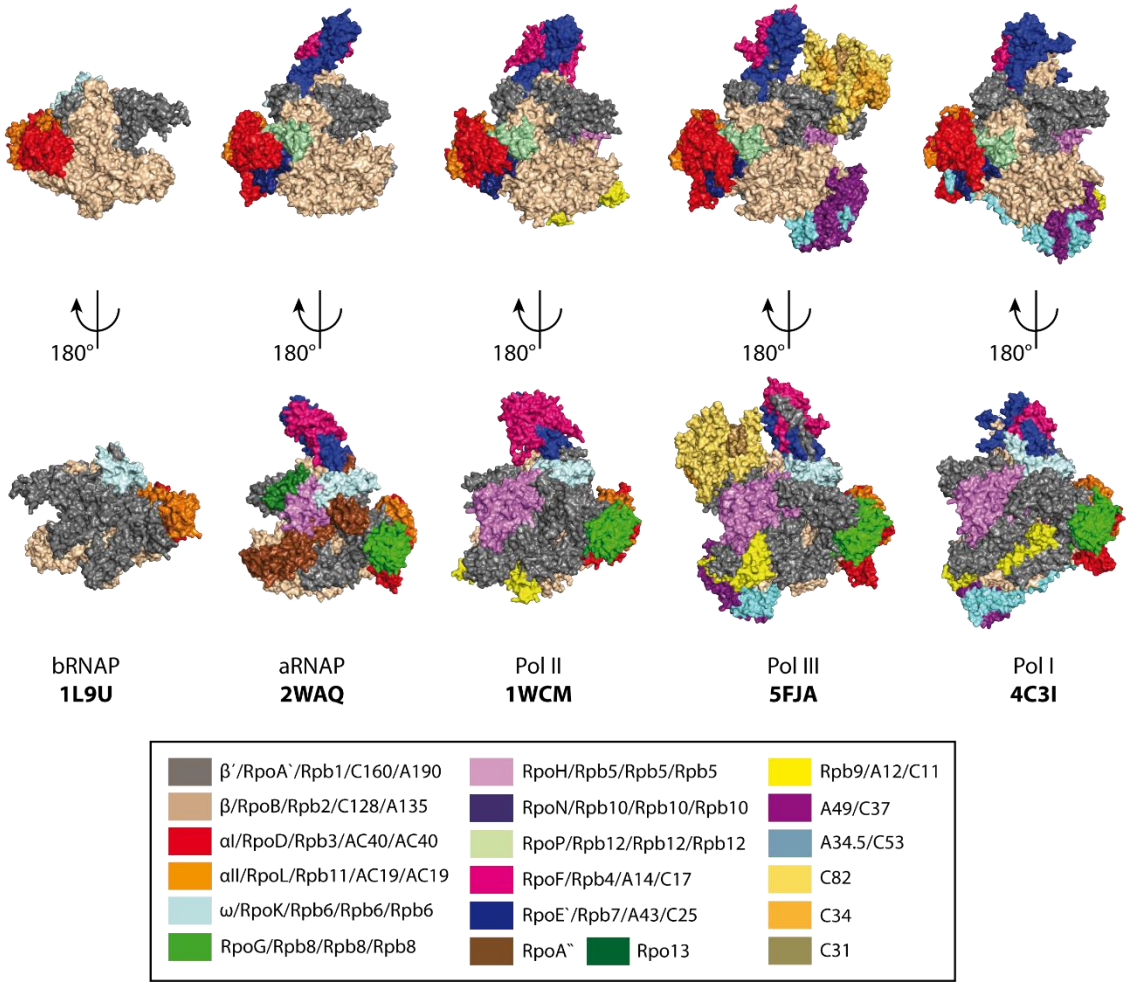

Figure 1. Structures of the multi-subunit RNAPs. Atomic models represented as surfaces. Conserved subunits are represented in the same color. PDB codes are indicated below.

2.2. Archaeal RNA Polymerase

Distinct X-ray crystal structures of archaeal RNA polymerase (aRNAP) revealed its architecture (Hirata et al 2008, Korkhin et al 2009). Like bRNAP, it transcribes all genes. The five subunits forming the aRNAP core are conserved in eukaryotes and bacteria.

Depending on the archaeal species, this enzyme can be composed from 11 to 13 subunits (Figure 1). Polypeptides RpoA' and RpoA'', encoded from two genes (Langer et al 1995), interact to form the RpoA'+RpoA'', orthologous to bacterial β' (Table 1). The second largest subunit, RpoB, is orthologous to bacterial β subunit and forms the cleft with RpoA+RpoA'', which contains the active site. RpoD and RpoL subunits are orthologs of the two α subunits in bacteria and are responsible for the stability and assembly of the largest heterodimer. RpoK subunit (ortholog of ω subunit in bacteria) helps to the interaction between RpoA'+RpoA'' and the stalk, a sub-complex formed by RpoE and RpoF. This heterodimer protrudes from aRNAP and is not present in bRNAP. Subunits RpoH, RpoG, RpoN and RpoP and their orthologs are only present in archaea and eukaryotic RNAPs.

2.3. Eukaryotic RNA Polymerases

Eukaryotic cells express three nuclear RNAPs. Although these enzymes share the catalytic mechanism, there are important differences between them, such as the number of subunits, the size and other special features.

2.3.1. RNA Polymerase II

RNA polymerase II (Pol II) comprises 12 subunits with a total mass of 500 kDa (Figure 1) and produces all mRNAs and several small nuclear RNAs. The two largest subunits are Rpb1 and Rpb2, which form the cleft and the active site and share homology with β' and β in bacteria and RpoA'+RpoA'' and RpoB in archaea. Rpb3 and Rpb11 form a heterodimer and function as the α/α or RpoD/RpoL subcomplexes. Pol II presents five subunits in common with the other two eukaryotic RNAPs (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12). These subunits have homologs in archaea (Minakhin et al 2001) and are specially important to maintain the architecture and assembly. Rpb9 does not have homologs in archaea and it interacts with Rpb1 and Rpb2. Rpb4 and Rpb7 form a stalk, such as RpoE and RpoF in archaea.

2.3.2. RNA Polymerase III

RNA polymerase III (Pol III), the largest among all RNAPs, is composed of 17 subunits with a total weight of approximately 700 kDa (Figure 1). It produces small non-coding RNAs, including all tRNAs and the 5S rRNA. Subunits C160 and C128, with homologs

in all multi-subunit RNAPs (Table 1), form the cleft and the active center of the enzyme. The heterodimer AC40 and AC19 is shared with Pol I and it is homologous of Rpb3/Rpb11 subcomplex in Pol II. The Pol III core is completed by subunit C11, involved in RNA cleavage. C11 N-terminal and C-terminal domains are homologous of Rpb9 and TFIIS from Pol II, respectively (Chédin et al 1998), and the latter is involved in RNA cleavage activity in Pol II (Wind and Reines 2000). The Pol III stalk is formed by subunits C17 and C25, which is involved in transcription initiation. The C53/C37 heterodimer is related to TFIIF, which has an elongation-stimulatory activity in Pol II (Knutson et al 2016). Finally, the C82/C34/C31 heterotrimer is positioned on C160 clamp domain and it is involved in transcription initiation (Wei and Chen 2018). C34 and C82 share homology with TFIIE, which acts in the transition from initiation to elongation in Pol II transcription (Compe et al 2019).

2.3.3. RNA Polymerase I

RNA polymerase I (Pol I) is a macromolecular enzyme (Figure 1) composed of 14 subunits that produces rRNA, which after maturation conforms the backbone of the ribosome. The two largest Pol I subunits, A190 and A135, constituting the DNA-binding cleft, contain regions homologous to Pol II subunits Rpb1 and Rpb2 respectively (Kuhn et al 2007). They also are homologous to the cleft-forming subunits from bacteria, archaea and Pol III (Table 1). The A12.2 N-terminal domain in Pol I is counterpart to Rpb9 in Pol II and the C11 N-terminal region in Pol III. The A12.2 C-terminal domain, which confers RNA cleavage activity to the Pol I active center, is homologous to Pol II TFIIS and the C11 C-terminal domain in Pol III. The A14/A43 heterodimer forms the stalk and it is homologous to the archaeal and eukaryotic stalk heterodimers. The enzyme is completed with the peripheral subcomplex A49/A34.5, which is structurally and functionally analog to the TFIIF heterodimer. This heterodimer is required for elongation-stimulatory activity (Kuhn et al 2007, Geiger et al 2010) and its deletion decreases Pol I activity (Huet et al 1975, Liljelund et al 1992). The A49/A34.5 dimerization module is anchored opposite the stalk similarly to TFIIF in Pol II (Vannini and Cramer 2012). The A49 C-terminal domain is homologous to TFIIE and is located close to the DNA-binding cleft in the EC.

Table 1. Homology in RNAP subunits.

Prokaryotes		Eukaryotes		
bRNAP	**aRNAP**	**Pol II**	**Pol III**	**Pol I**
β'	RpoA'+RpoA"	Rpb1	C160	A190
β	RpoB	Rpb2	C128	A135
α	RpoD	Rpb3	AC40	
α	RpoL	Rpb11	AC19	
ω	RpoK	Rpb6 (ABC23)		
	RpoH	Rpb5 (ABC27)		
	RpoG	Rpb8 (ABC14.5)		
	RpoN	Rpb10 (ABC10β)		
	RpoP	Rpb12 (ABC10α)		
	RpoF	Rpb4	C17	A14
	RpoE	Rpb7	C25	A43
		Rpb9	C11-Nt	A12.2-Nt
		TFIIS-Ct	C11-Ct	A12.2-Ct
		TFIIFα	C37	A49-Nt
		TFIIFβ	C53	A34.5
		TFIIEβ	C34	A49-Ct
		TFIIEα	C82	
			C31	

3. **RNA polymerases show common features**

There are several structural and molecular features shared between archaea, bacterial and eukaryotic RNAPs that suggest a common ancestry (Langer et al 1995, Hirata et al 2008). This is reflected in the sequence and structure of their subunits, the interactions with TFs and the molecular mechanisms (Werner and Grohmann 2011). First, their two largest subunits form a DNA-binding cleft and the active site of the enzyme is located within. In the active site, three conserved aspartate residues coordinate a Mg^{2+} ion, which is required for catalytic activity. Beyond the active site, the DNA path is blocked by a structural element named as "wall" (Cramer et al 2001). A fundamental conserved element, the

bridge helix (BH), lies inside the largest subunit and is involved in enzyme translocation along template DNA, likely through a fully folded/partially unfolded transition that allows displacement of the protein complex toward downstream DNA (Cheung and Cramer 2012). The five-subunit core is homologous in the three kingdoms, suggesting a common background. Furthermore, the enzymes show two channels, one for the substrate NTPs and the other for the RNA product (Khatter et al 2017).

Another important element is the subcomplex that forms the stalk, which does not exist in bRNAP but is a recognizable feature from archaea to eukaryotic RNAPs and modulates the three transcription cycle steps (Werner and Grohmann 2011). The stalk is dissociable in Pol II, but is strongly bound in aRNAP (Grohmann et al 2009), Pol I and Pol III (Fernández-Tornero et al 2013, Hoffmann et al 2015). The position of the stalk in relation with the core is highly conserved between aRNAP and the three eukaryotic RNAPs (Fernández-Tornero et al 2007).

The folding of conserved elements around the active site is essentially identical in nuclear RNAPs and aRNAP (Hirata and Murakami 2009), most differences between the enzymes being greater as the distance to the central part increases. Almost all the structural discrepancies between archaeal and eukaryotic RNAPs can be classified as simple addition of eukaryotic RNAP-specific polypeptides to aRNAP rather than changes in the main RNAP architecture. It seems likely that the structure of aRNAP has been maintained from the ancestor of archaea and eukaryotes, whereas nuclear RNAPs further accompanied the evolutionary lines of the eukaryotic kingdom.

4. RNA polymerases transcription elongation

For gene transcription to occur, a fundamental structure of nucleic acids must be maintained: the transcription bubble. This bubble is formed by a double DNA helix which displays a central region with approximately 15 impaired nts and an RNA chain with 8 or 9 residues hybridized with the TS (Figure 2A). Structural studies using yeast cells have identified two conserved elements adjacent to the active site with irreplaceable function in translocation: the aforementioned BH and the trigger loop (TL) (Cramer et al 2001, Gnatt et al 2001) (Figure 2B). The BH, highly conserved in sequence, divides the RNAPs central cleft into a main DNA loading channel and a secondary channel for NTP entry (Zhang et al 1999, Zhang et al 2015). The BH is straight in elongating RNAPs and oscillates between folded and bent states during the transcription cycle, facilitating the

entrance of nts into the active site. The TL is a flexible element that swings between two positions: near the active site or closer to the downstream DNA. This oscillating movement allows establishing both direct and indirect contacts with the RNAP active center and is involved in correct NTP selection at the nucleotide addition site. (Wang et al 2006).

During transcription elongation, the RNAPs move stepwise along a DNA TS and repeatedly perform the nucleotide addition cycle (NAC), a mechanism conserved in all three kingdoms of life (Brueckner et al 2009). The TS is positioned close to the active site and contains the *i*+1 position, which is defined as the nucleotide that is complementary to the incoming NTP substrate (Murakami et al 2002). The NAC starts when a NTP substrate extends the growing 3′-end of the RNA chain by DNA template-directed formation of an RNA phosphodiester bond and the release of a pyrophosphate ion (Brueckner et al 2009) (Figure 2C). At post-translocation state, *i*+1 position remains empty, the BH is folded and TL is open. NTP entry and closure of the active center leads to complete folding of the TL. NTP binding requires the presence of two catalytic metal ions (A and B). Metal ion A is persistently bound to the active site by three conserved aspartate residues, while metal ion B enters with NTP substrate. The catalysis reaction and the consequent liberation of a pyrophosphate ion destabilize the TL closed conformation and break the TL contacts with the active site, leading to TL return to the open state. At this stage, termed the pre-translocated state, the BH bends to allow the forward movement of the DNA/RNA hybrid to a new position. Downstream DNA is translocated until the next nt in the TS reaches the *i*+1 position (Brueckner et al 2009). The BH reversion to the straight state without movement of nucleic acids creates an empty site at position *i*+1 for entrance of the next nucleotide, completing the NAC (Gnatt et al 2001) (Figure 2C). Structural comparison of RNAPs at different translocation states revealed that the straight/bent states shown by the BH are important for the interaction between nucleic acids and protein during translocation of the enzyme and the orchestrated movements of the TL are essential for catalysis (Cheung et al 2012).

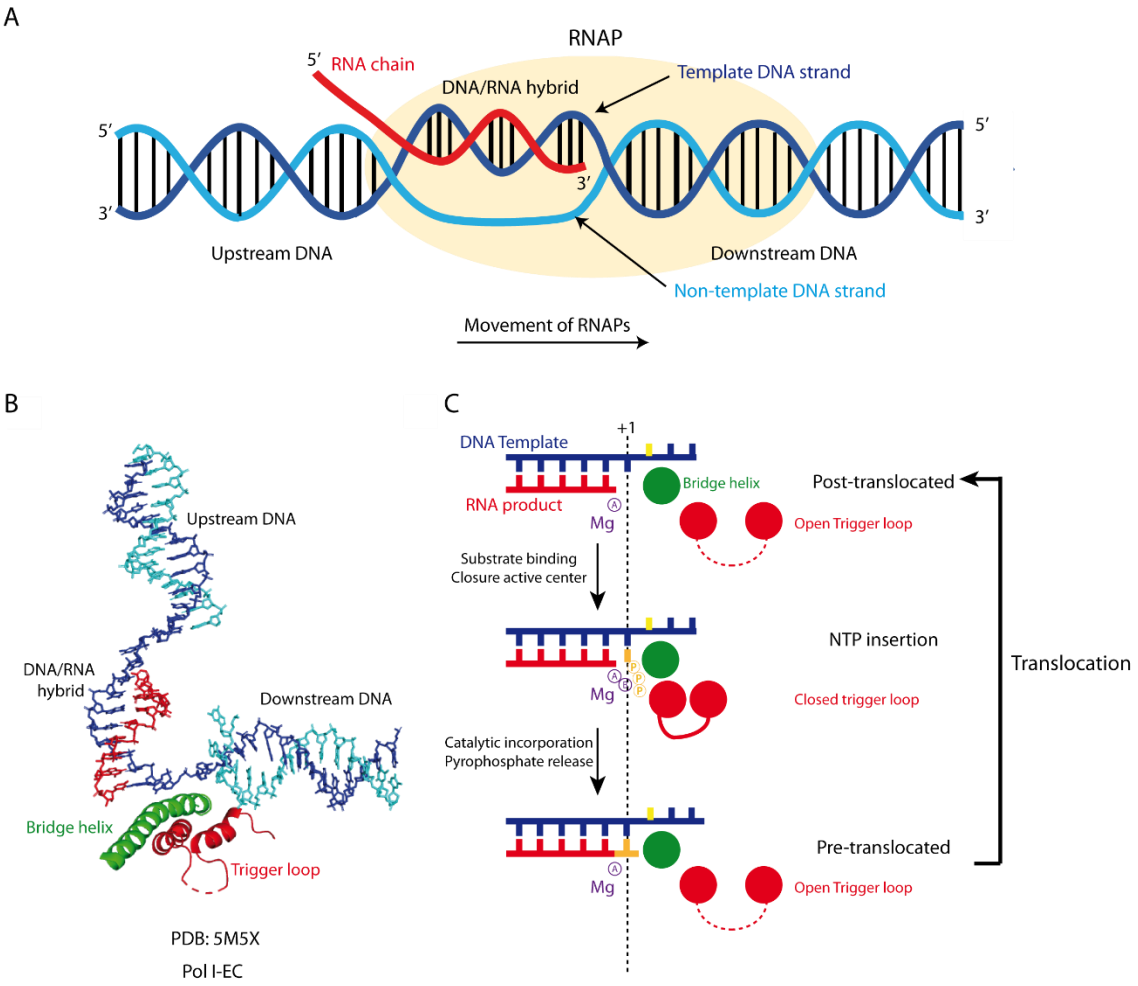

Figure 2. Elements associated to transcription elongation and different stages at the nucleotide addition site. (A) Schematic figure of the transcription bubble, showing the TS, NTS and RNA chain in blue, cyan and red, respectively. RNAP is colored in light yellow. (B) Atomic model of Pol I transcription bubble, BH and TL (extracted from Pol I-EC (Tafur et al 2016)). The BH and TL are colored in green and red, respectively. (C) Schematic representation of the NAC (adapted from Brueckner and Cramer 2008 with modifications). The vertical dashed lines indicate *i*+1 position.

Although structural studies have enhanced the understanding of the transcription process, only static snapshots of the starting and ending points of NAC were achieved (Figure 2C). Molecular dynamics studies revealed that NAC presents four metastable states and two of them are structurally unidentified intermediate states. These studies showed that two conserved amino acids at the BH of Pol II, T831 and Y836, directly interact with the DNA transition nucleotide *i*+1 and play an important role in the DNA/RNA hybrid translocation (Silva et al 2014).

Elongation stage is not a smooth and continuous process; RNAPs are prone to pausing when they sense certain DNA sequences, encounter a DNA lesion in the TS or

misincorporate a noncomplementary nucleotide. This temporary arrest situation represents a rate-limiting step during transcription elongation (Kireeva and Kashlev 2009). In these cases, RNAPs may delay the extension of the RNA chain (Weixlbaumer et al 2013). In prokaryotes, pausing regulates the expression of operons, such as *trp* or *his* (Landick and Yanofsky 1987, Chan and Landick 1989), maintain the transcription-translation coupling and plays an important role in termination (Landick 2006). The intermediate state of bRNAP reveals a relaxed, open-clamp conformation, similar to the nucleic acid free bRNAP core. Inside the cleft, the enzyme exhibits a kinked BH and a widened RNA exit channel. The kinked BH blocks the TS *i*+1 position and obstructs the NTP substrate entry, and the opened RNA exit channel allows the formation of a stable RNA hairpin (Weixlbaumer et al 2013). In eukaryotic cells, Pol II is paused in the promoter-proximal regions (Muse et al 2007, Core and Lis 2008). Among other functions, pausing can provide the time to load the elongation regulators, be involved in splice-site selection and aid adequate RNA folding (Toulokhonov et al 2007). Structural studies in Pol II revealed that the pause state corresponds to an inactive intermediate between pre-translocated and post-translocated stages, which results from a rearrangement of the EC that inhibits nt addition (Sydow et al 2009).

5. Pol I transcription initiation

Ribosome biosynthesis is a central cellular process extremely regulated and a huge metabolic effort that in eukaryotes requires the coordinated action of all three nuclear RNAPs (Laferté et al 2006, Schneider et al 2007). In fact, Pol I transcription, as a crucial step for ribosome biogenesis, is regulated in different manners, such as disruption of inactive Pol I homodimers into monomers (Torreira et al 2017) or adjustments in the number of active rRNA genes and chromatin structure (Birch and Zomerdijk 2008). In particular, Pol I initiation is specially modulated and recent studies have provided further structural and biochemical insights into this regulation (Engel et al 2017, Han et al 2017, Sadian et al 2017, Torreira et al 2017).

5.1. rDNA

Several decades ago, Pol I was observed transcribing the rDNA repeats by electron microscopy and forming specific structures known as Miller spreads, resembling Christmas trees (Miller and Beatty 1969). Intercalated matrix units (Mx) and matrix-free

segments (S) were visible, as well as Pol I molecules loaded on a single rDNA gene (black dots in the "tree trunks") (Figure 3A). The detailed examination of Mx revealed that each of them is composed by hundreds of fibrils connected by one end to the "tree trunk". These fibrils are the 35S pre-rRNA and they increase in length from the beginning to the end of the rDNA gene (Figure 3A).

In the yeast nucleolus, there are 150-200 copies of rRNA genes of about 9 kb each (Russell and Zomerdijk 2005), located on chromosome XII (Petes 1979). The repeat are organized as tandem head-to-tail repeats and constitute the nucleolar organizing regions (NORs) (Russell and Zomerdijk 2005). Each rDNA repeat contains the small 5S rRNA gene, the transcribed region coding for the 35S pre-rRNA and non-transcribed regions, such as Pol I promoter, which contains two main regions, the upstream activating sequence (UAS) and the core element (CE) (Figure 3B).

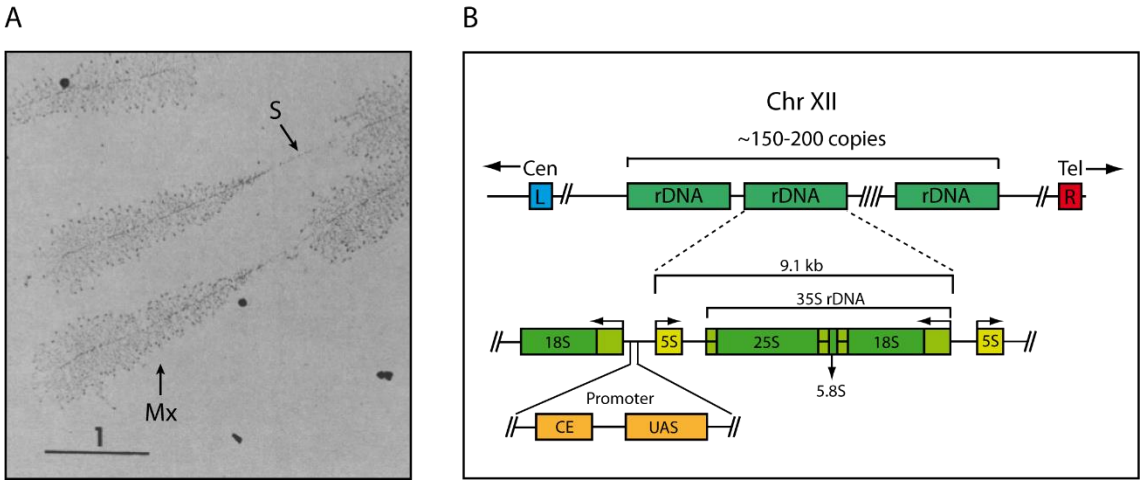

Figure 3. rDNA organization. (A) Miller spread of a *Triturus viridescens* oocyte (Miller and Beatty 1969). Scale bar - 1 µm. Mx, matrix units; S, matrix-free segments. (B) Schematic of yeast rDNA organization. Chr, chromosome; Cen, centromere; Tel, telomere; CE, core element; UAS, upstream activating sequence.

5.2. The assembly of the Pol I initiation complex

The first step in transcription is the assembly of a PIC (Knutson et al 2014, Han et al 2017, Sadian et al 2017). Yeast requires a unique set of general transcription factors (GTFs): the Upstream Activating Factor (UAF), the TATA-binding protein (TBP), the Core Factor (CF) and the regulatory factor Rrn3 (Schneider 2011). GTFs are found only in the promoter (Bier et al 2004) and do not participate in the elongation stage. They bind sequentially to specific DNA sequences in the promoter: the UAS and the CE.

First of all, UAF forms a stable complex with the UAS in the promoter, and seems not to require CF, Pol I or Rrn3 for the interaction (Goetze et al 2010). Although the rDNA promoter does not contain the TATA-box, the next step is the recruitment of TBP, likely playing a role, together with the UAF, in the attachment of the CF to the PIC (Steffan et al 1996). The third step is the interaction of the CF with the CE. This three-subunit complex (Lalo et al 1996, Lin et al 1996) contacts Pol I and Rrn3 and plays a central role in PIC assembly. *In vitro* assays revealed that the CF does not require UAF and TBP for promoter recognition, and it is essential to direct Pol I:Rrn3 complex to the TSS and basal transcription initiation (Keener et al 1998, Pilsl et al 2016, Engel et al 2017). Once UAF, TBP and CF form a stable complex, the last step is the recruitment of Pol I in complex with Rrn3. This regulatory factor is functionally conserved in eukaryotic cells (Schnapp et al 1993, Bodem et al 2000, Moorefield et al 2000) and interacts with the A43 stalk subunit of Pol I (Peyroche et al 2000), the A14 subunit, the A135 stalk-binding region, the A190 dock domain, the AC40/AC19 heterodimer and Rpb6 (Engel et al 2016, Pilsl et al 2016, Torreira et al 2017). Upon recruitment of Pol I:Rrn3 to the TSS, forming the CC, the double helix DNA is melted and forms the OC. The DNA opening permits the TS to reach the active site, and downstream DNA is positioned within the cleft. Pol I, similar to other eukaryotic RNAPs, forms the ITC. Once the nascent RNA is between 8 and 12 nt in length, Pol I clears the promoter (Kahl et al 2000) and converts the ITC to a stable EC. Rrn3 leaves the transcribing complex very early in the elongation stage and is not part of a stable Pol I-EC (Bier et al 2004). Occasionally, the step from ITC to EC is unsuccessful and the enzyme and a short RNA transcript are released.

6. Conformational states in Pol I

There are three major conformations of the Pol I enzyme (Fernández-Tornero 2018). X-ray crystal structures of yeast Pol I (Engel et al 2013, Fernández-Tornero et al 2013) were nearly identical and described Pol I in the inactive dimeric state. The principal interaction in the dimers involves A43-Ct (251-326) insertion in the cleft of the neighboring monomer. A43-Ct can be divided into three regions; a long α-helix (amino acids 273 to 293), a β-hairpin (298 to 309) and an acidic C-tail (311 to 326). *In vivo* mutational assays with the β-hairpin partially truncated and deleted C-tail (A43 Δ307-326) showed the disruption of Pol I dimers (Torreira et al 2017). In addition to homodimerization, the principal main inactivation characteristics are that the BH is partially unfolded in its

central region, a DNA-mimicking loop prevents downstream DNA accommodation near the active site and an open cleft of about 42 Å in width is unable to trap the DNA (Figure 4A). Furthermore, the A12.2-Ct involved in the RNA cleavage, is placed inside the pore (Kuhn et al 2007).

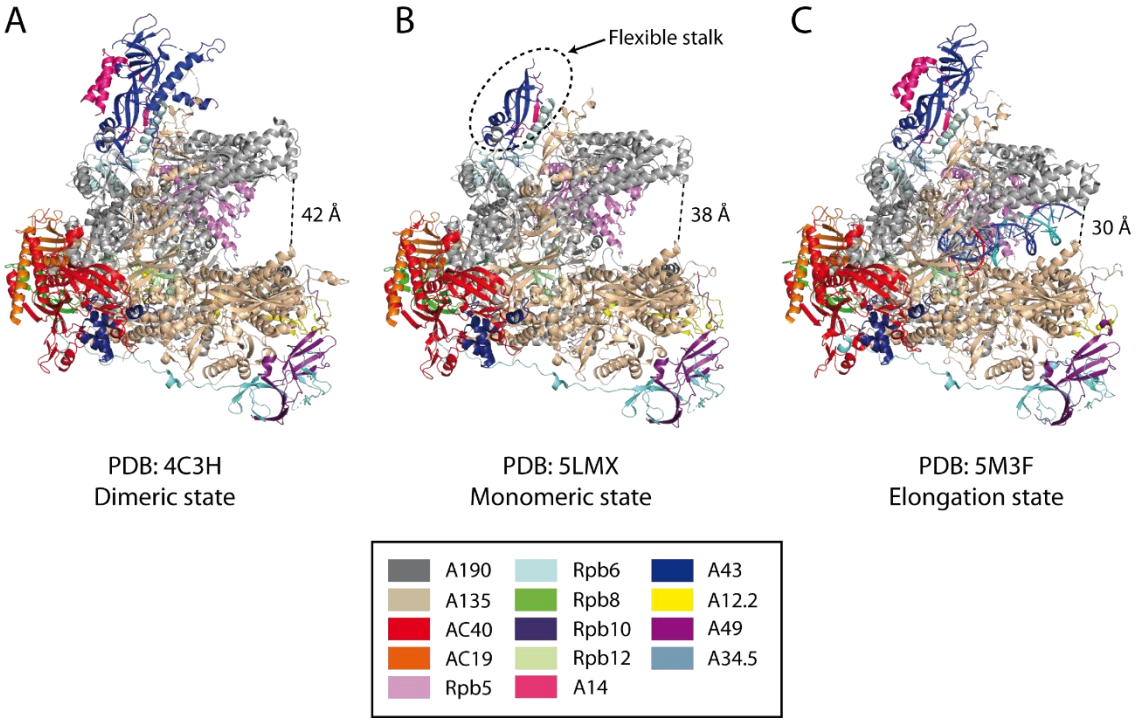

Figure 4. Three major conformational states of Pol I. The subunits forming the structures are colored according to the bottom legend. (A) Dimeric Pol I (PDB:4C3H). (B) Monomeric Pol I (PDB:5LMX). (C) Monomeric Pol I in elongation state with an artificial transcription bubble (PDB:5M3F).

Pol I activation involves the disruption of the dimers, reorganization of the DNA-mimicking loop and cleft closing. The free monomeric structure presents a semi-expanded cleft of about 38 Å in width (Figure 4B), a partially unfolded BH and disordered TL and DNA-mimicking loop. The A12.2-Ct domain has been found either disordered or partially ordered inside the pore. In addition, the stalk (heterodimer A14/A43) appears flexible in monomeric Pol I (Figure 4B). Pol I in the ITC and EC is represented by the third conformational state (Figure 4C). This conformation is defined by a full closure of the cleft (about 30 Å in width) to maintain the nucleic acids bound at the active site and a fully ordered BH (Neyer et al 2016, Tafur et al 2016, Sadian et al 2017). Other two important features are that the DNA-mimicking loop is disordered, as expected by the presence of nucleic acids in the cleft, and the A12.2-Ct domain is flexible, as the

elongating Pol I does not require its cleavage domain in place while polymerizing the RNA.

7. DNA lesion recognition and repair pathways

The maintenance of genomic DNA integrity is essential for the normal function of the cell, as well as for the cell survival (Pani and Nudler 2017). Consequently, chemical changes in the genetic material are connected to DNA mutagenesis, cancer and apoptosis. DNA damage can be produced by exogenous agents like genotoxic chemical products, radiation or pollution, or endogenous sources, such as reactive oxygen species (Sancar et al 2004, Hanawalt et al 2008).

The universal repair mechanism is the nucleotide excision repair (NER), which removes longer sections containing diverse bulky and helix-distorting lesions (Wirth et al 2016, Peyresaubes et al 2017). Defects in NER activity are associated with several rare autosomal recessive human disorders, including Cockayne syndrome (CS) and Xeroderma pigmentosum (XP). CS is characterized by neurological dysfunction, including retinal degradation and XP exhibits a high incidence of internal tumors. Sunlight hypersensitivity and high skin cancer in different severity levels are hallmarks of the two diseases (Hanawalt and Spivak 2008).

Multiple factors affect the NER capability, like DNA and histone modifications, condensed chromatin domains or DNA dynamics (Yang et al 2019). NER can be divided into two subpathways: global genome-nucleotide excision repair (GG-NER), which deals with damage anywhere in the genome (Hanawalt and Spivak 2008) and transcription coupled-nucleotide excision repair (TC-NER), which is activated when RNAPs are blocked by bulky DNA lesions in the template DNA of actively transcribed genes (Tornaletti and Hanawalt 1999) (Figure 5). TC and GG-NER are divided into four principal steps: damage recognition, verification, incision (pre-incision repair complex assembly and dual incision steps) and gap filling (DNA synthesis and ligation steps), and the mechanisms only differ in the first step.

In GG-NER of yeast, the lesion is sensed by the Rad4-Rad23 complex (Figure 5), which is orthologous to human XPC-RAD23B. Rad4 repair factor is localized to the DNA lesion and indirectly recognizes diverse chemically and structurally DNA double helix distortions *in vitro* (Kong et al 2016, Mu et al 2018). The lesion recognition by Rad4-

Rad23 complex is necessary to trigger GG-NER. The transcription factor TFIIH contains the two DNA helicases Rad3 and Rad25 that translocate on ssDNA at opposites directions and are essential for creating a bubble structure (Sung et al 1987, Guzder et al 1994, Sung et al 1996). The two endonucleases, Rad1–Rad10 complex and Rad2, incise the damaged DNA strand on the 5′- and 3′-side of the lesion, respectively and both show structure-specific activity (Prakash and Prakash 2000). The damage binding factors include Rad14 that binds with high specificity to UV damage DNA and RPA, a heterotrimeric protein complex essential to protect an exposed single stranded DNA (ssDNA) and DNA replication (Yates et al 2018). In TC-NER subpathway, DNA damage is initially recognized by RNAPs (Hanawalt and Spivak 2008), which arrest at the DNA lesion and trigger the repair (Lainé and Egly 2006, Lindsey-Boltz and Sancar 2007, Wang et al 2018). This process is dependent on the repair proteins Rad26 and Rad28 in yeast cells, homologs to human Cockayne syndrome proteins CSB and CSA, respectively. These transcription-repair factors, together with the TFIIH complex and Rad2, remove Pol II from the lesion and recruit the subsequent NER factors (Rad14, Rad1-Rad10 complex and RPA) (Mitchell et al 2003) (Figure 5). Both NER subpathways eliminate the DNA damage by dual incisions bracketing the lesion, removing a DNA fragment between 24 and 32 nucleotides in eukaryotes and 12 to 13 nucleotides in prokaryotes (Sancar and Reardon 2004). The new DNA fragment is produced by DNA polymerases δ, ε or κ (Ogi and Lehmann 2006) using the NTS as template. The NER cycle ends with the ligation of the newly synthesized DNA.

In yeast rDNA, TC-NER occurs only in active repeats (Conconi et al 2002). Occasionally, blocked Pol I is displaced by nucleosomes and GG-NER removes the DNA damage from the TS (Charton et al 2015). TC-NER in rRNA genes is independent of Rad26, an essential repair factor of Pol II transcribed genes (Verhage et al 1996). Yeast Rad4 repair protein is needed to remove DNA lesions in the intergenic space and in the inactive rRNA genes, whereas it is not essential for TC-NER in active rDNA.

In mammalian cells, the DNA repair proteins of TC-NER in Pol II genes have been found in Pol I transcription sites, such as CSB, TFIIH and other components of the NER machinery (Hannan et al 1999, Bradsher et al 2002, Iben et al 2002). However, a direct role for these factors in the TC-NER of rDNA has yet to be demonstrated. In contrast to the TC-NER observed in many genes of mammalian cells transcribed by Pol II (Hanawalt and Spivak 1999), no evidence for TC-NER was found in the rDNA of higher eukaryotes (Yang et al 2019).

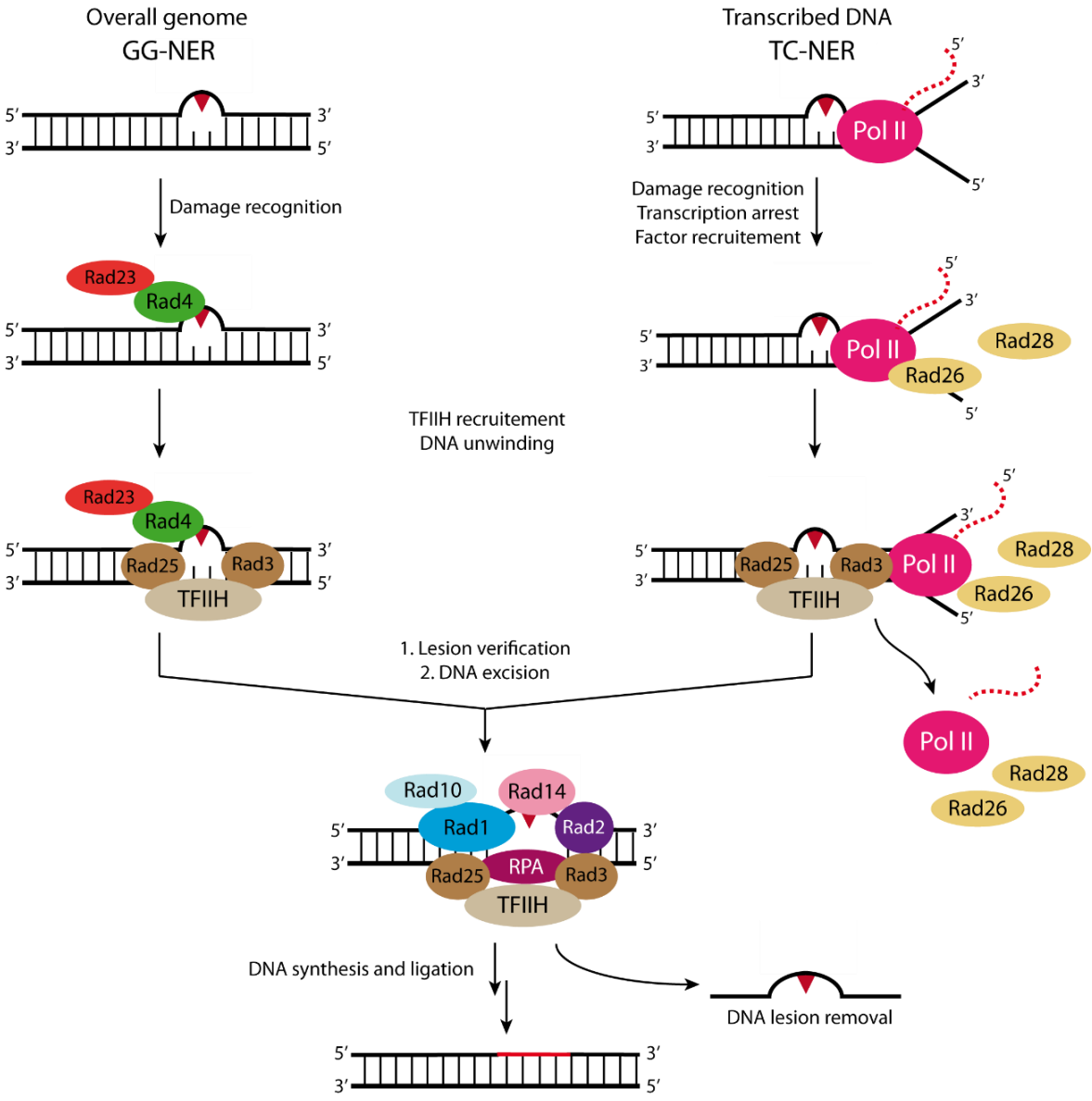

Figure 5. Steps of two NER subpathways (figure adapted from Mitchell et al 2003). Common steps of NER mechanisms are depicted in the middle of the figure. The lesion is represented as a maroon triangle.

8. The arrest of RNA polymerase II by UV light-induced lesions

One of the best-known environmental threats is UV light, which can generate bulky DNA lesions that are cytotoxic (Mitchell et al 2003). The main UV light-induced DNA lesions are *cis-syn* cyclobutane pyrimidine dimers (CPDs) and 6-4 photodimers (Ravanat et al 2001), that covalently crosslink two adjacent nucleotides in the same strand. These UV light-induced photoproducts produce an important distortion of the local DNA duplex structure that interferes with base pairing and obstructs fundamental processes such as transcription (Hoeijmakers 2001).

The first structural studies about CPD-induced blocked transcription were performed using Pol II, the best-studied eukaryotic RNA polymerase. There are several structures of Pol II in EC with a transcription bubble containing the bulky CPD lesion at different positions (Pol II-EC_{CPD}) (Figure 6). While the overall structures of those Pol II-EC_{CPD} are nearly identical to the damage-free Pol II-EC (Kettenberger et al 2004), important differences are observed in the vicinity of the active center.

The X-ray structure containing the CPD lesion at *i*+2/*i*+3 position exhibited a post-translocated state (Figure 6A). This complex was used to perform time-dependent RNA transcription assays, which showed that Pol II is blocked after nucleotide incorporation opposite both pyrimidines. The incorporation rate opposite the 3'-T was approximately 16 nt/h, while the addition rate opposite 5'-T was 2.4 nt/h. When the CPD damage has reached the active site and remains stably accommodated at *i*-1/*i*+1 position, also revealing a post-translocated state (Figure 6C).

Subsequent structural studies by X-ray crystallography obtained the Pol II-EC_{CPD} containing the DNA damage at position *i*+1/*i*+2. This structure showed that the pyrimidine dimer is located above the BH and cannot reach the canonical position *i*+1 site (Figure 6B) (Walmacq et al 2012). This translocation represents an intermediate between pre- and post-translocated states for undamaged TS (Silva et al 2014). Complementary RNA transcription assays using two different mismatched RNAs opposite the CPD lesion at *i*-1/*i*+1 position concluded that Pol II can bypass the pyrimidine dimer if it incorporates adenine opposite the 5'-T (Brueckner et al 2007). Therefore, the Pol II behavior in the presence of the CPD lesion results in the non-

template AMP incorporation according to the A-rule followed by two possibilities; an addition of another AMP, leading to lesion bypass or a misincorporation of UMP, leading to arrest. The incorporation of the first adenine together with slight conformational changes of BH and TL helps to stabilize the CPD lesion into the active site. From the branch-point, the second incorporation event drives Pol II between two pathways. Blocked Pol II presents the CPD lesion within the active site, which is not accessible to the repair factors (Walmacq et al 2012). Whereas blocked Pol II at the lesion recruits the repair machinery, this is not enough for lesion removal (Hanawalt and Spivak 2008).

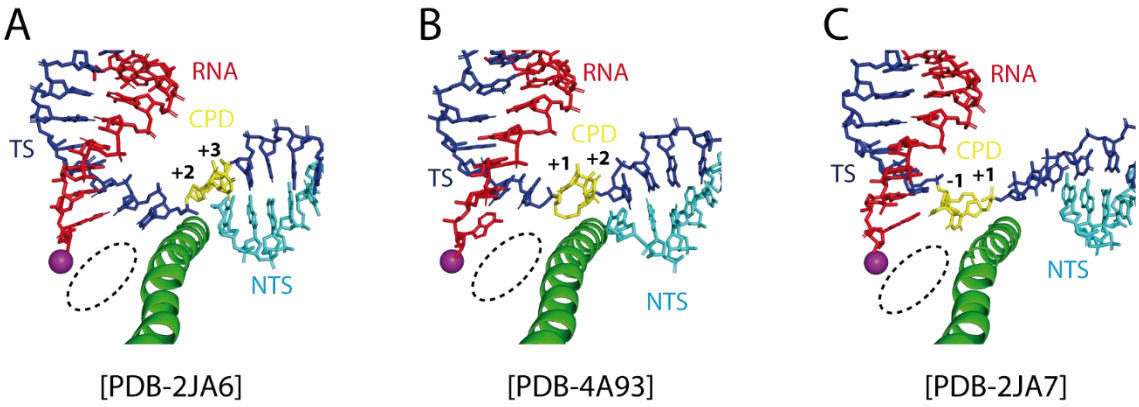

Figure 6. Structures of stalled Pol II-EC$_{CPD}$ showing the CPD lesion at different locations. RNA, TS, NTS, and the BH motif of RNA Pol II (Rpb1 810-847) are colored in red, blue, cyan and green, respectively. The CPD lesion is highlighted in yellow. The magnesium ion at the active site is colored in purple. The Pol II active site is indicated by a dashed circle.

9. The Nobel Prize in Chemistry for developing cryo-EM

Understanding how macromolecular complexes carry out their roles in cells is a central issue in molecular biology. To this aim, one fundamental study field is structural biology, which deduces the 3D arrangement of the atoms (Bai et al 2015). The most used techniques are nuclear magnetic resonance (NMR), X-ray crystallography (X-ray) and cryo-EM. Among others, one of the limiting factors to perform structural studies is the sample amount. Cryo-EM needs much less protein (about 0.1 mg) than X-ray crystallography. Besides, high-quality diffracting crystals are not required. Due to these reasons and the advances in the technique, the number of novel protein complex structures obtained by cryo-EM has been increasing (Fernandez-Leiro and Scheres 2016). Fundamental issues were improved in recent years: sample preparation, microscopes,

detectors and the image processing software. It is now possible to reach atomic resolution in 3D density maps of biological assemblies.

First observations of the specimens using EM were done by negative-staining, where the water is replaced by a dried solution of heavy-metal salt, although this method resolution is limited by the grain size (Brenner and Horne 1959). One of the most fundamental cryo-EM improvements consists on the preservation of the sample in a close-to-native state embedded in amorphous ice (not crystalline) after flash freezing in liquid ethane at liquid nitrogen temperatures. This was fundamental to gain a better understanding of biological function (Bai et al 2015).

In addition, the accelerating voltage has been increased from 120 kV to 300 kV and thermionic emission electron guns were replaced by field emission guns. The increase of the electron beam temporal stability was enhanced decreasing the temperature of the electron source cathodes; and the vacuum around frozen specimens was improved, reducing ice contamination and allowing data collection for days rather than hours, so much more images can be recorded, specially now with automatic image acquisition.

Likewise key to the development of cryo-EM were the new generation electron detectors, able to obtain information with unprecedent speed and sensitivity (Kühlbrandt 2014). First EM images were recorded on photographic film, and more recently, with charge-coupled device (CCD) cameras. These cameras work well at low energies, however their detective quantum efficiency (DQE) decrease at higher energies. This introduces noise to the received signal and eventually affects the 3D density map resolution. The new commercial cameras have a DQE much higher than film and CCD cameras. Therefore, they reach a high signal to noise ratio and they can detect individual events of 300 kV electrons (McMullan et al 2014). Direct electron detectors can continuously integrate electrons without any additional conversion step, storing and processing individual frames, which form a "movie". In fact, dealing with "movies" prompted other development, the alignment of individual movies frames, which improves the signal to noise ratio of individual particles and partially accounts for their beam induced motion (Brilot et al 2012, Campbell et al 2012).

Improvements in computer programs for image processing were performed as well to address the difficulties with heterogeneous samples. Maximum likelihood algorithms applied to microscopy (Sheres 2010, Sigworth et al 2010, Sheres 2012), were developed

from the method used previously by X-ray crystallography community (Henderson 2015). The RELION program (Sheres 2012), which is widely used, incorporates maximum likelihood algorithms.

Professors Richard Henderson, Jacques Dubochet and Joaquim Frank were awarded the 2017 Nobel Prize in Chemistry for their effort on developing cryo-EM for the high-resolution structure determination of biomolecules in solution. Richard Henderson succeeded in using an electron microscope to generate a 3D image of a protein at atomic resolution. Joachim Frank developed an image processing method in which the electron microscope's blurred 2D images are analyzed and merged to reveal a sharp 3D structure. Jacques Dubochet succeeded in vitrifying water, allowing the biomolecules to retain their natural shape even under the microscope vacuum. The enormous advance in this field over the last time was characterized by the term "Resolution Revolution" (Kühlbrandt 2014).

All the advances in cryo-EM were essential for the results described in this Ph.D. thesis. The obtained cryo-EM maps permitted to build the first published atomic model of Pol I stalled at the CPD lesion, revealing important mechanistic insights. The structures, complemented with biochemical studies, shed light on how Pol I is blocked upon encountering a bulky distortion in the template DNA.

OBJECTIVES

The main objective of this work was the 3D structural determination by cryo-EM of Pol I in elongation complex containing a CPD lesion in the position *i*+1/*i*+2 of the template DNA strand. To this aim, several objectives were planned.

- To assemble *in vitro* the transcription bubble efficiently and the Pol I-EC_{CPD} complex under native conditions.

- To prepare cryo-EM grids for data collection at a Titan Krios equipped with direct-electron detector to elucidate the 3D structure of Pol I-EC_{CPD} by cryo-EM at a quasi-atomic resolution.

- To obtain a refined atomic model for Pol I-EC_{CPD}.

- To compare Pol I-EC_{CPD} with Pol II-EC_{CPD} and both lesion-free Pol I-EC and Pol II-EC.

- To compare the behavior of Pol I and Pol II in the presence of a CPD lesion using *in vitro* transcriptional assays.

- To validate structural results by performing mutational analysis and *in vitro* transcription assays to confirm the role of certain amino acids.

MATERIALS AND METHODS

1. Materials

1.1. Media and additives

All media were prepared by dissolving the components in water and autoclaving at 121°C for 20 min.

Table 2. Cell culture media.

Name	Composition
Lysogeny broth (LB)	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl
LB plates	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 1.5% (w/v) agar
NYZ^+	1% (w/v) NZ amine, 0.5% (w/v) yeast extract, 0.5% NaCl, 0.4% (w/v) glucose*
YPD	1% (w/v) yeast extract, 2% (w/v) bactopeptone, 2% (w/v) glucose*
YPDA	1% (w/v) yeast extract, 2% (w/v) bactopeptone, 2% (w/v) glucose*, 0.003% (w/v) adenine hemisulfate
YPD plates	1% (w/v) yeast extract, 2% (w/v) bactopeptone, 2% (w/v) glucose*, 2% (w/v) agar

* A 40% (w/v) glucose stock solution was prepared separately, autoclaved at 110 °C and added under sterile conditions.

Table 3. Cell culture additives.

Stock	Work concentration
Ampicillin 1000x	100 µg/mL
Chloramphenicol 1000x	25 µg/mL
$MgSO_4$ 80x	12.5 mM
$MgCl_2$ 80x	12.5 mM

1.2. Native polyacrylamide gel electrophoresis

Native gels for protein samples were run at 120 V and 4°C in running buffer 1 (25 mM Tris, 192 mM glycine, 2 mM DTT).

Table 4. Recipe for native polyacrylamide gels.

Gels with 0.75 mm spacer plates (BIORAD)

	Percentage	
Stock	**4.5 %**	**7.0 %**
Sterile deionised H_2O	6.3 mL	5.7 mL
1.5 M Tris pH 8.8	2.5 mL	2.5 mL
40% Acrylamide/Bis Solution 19:1 (BIORAD)	1.1 mL	1.7 mL
APS 10% (w/v) (prepared from APS powder, BIORAD)	100 µL	100 µL
TEMED (BIORAD)	10 µL	10 µL

Nucleic acid polyacrylamide gels were run at 100 V and 4°C in running buffer TAE 1x (40 mM Tris, 20 mM acetic acid and 1 mM EDTA).

Table 5. Recipe for nucleic acid acrylamide gels.

Stock	**20%**
Sterile deionised H_2O	3.03 mL
30% acrylamide (BIORAD)	6.66 mL
TAE 50x	200 µL
APS 10% (w/v) (prepared from APS powder, BIORAD)	100 µL
TEMED (BIORAD)	20 µL

1.3. Denaturing urea polyacrylamide gel electrophoresis

Urea gels for *in vitro* transcription assays were run at 25 V and RT in running buffer TBE 1x (89 mM Tris base, 89 mM boric acid and 2 mM EDTA).

Table 6. Recipe for urea polyacrylamide gels.

Stock	12%
Urea (Merck)	56.7 g
TBE 10x	13.5 mL
Sterile deionized H_2O	96 mL
40% Bis-acrylamide 19:1 (BIORAD)	40.5 mL
APS 10% (w/v) (prepared from APS powder, BIORAD)	1 mL
TEMED (BIORAD)	100 µL

2. Assembly of the DNA-RNA scaffold containing the CPD lesion

In order to prepare stock solutions, non-template DNA strand (NTS - 5'GCAGCCTAGTTGATCTCATAGCCCATTCCTACTCAGGAGAAGGAGCAGAGCG-3'), CPD-containing template DNA strand (TS - 5'CGCTCTGCTCCTTCTCCTTTCCTCTCGATGGCTATGAGATCAACTAGGCTGC 3') where TT represents the thymine dimer and 10-mer RNA (5'- AUCGAGAGGA-3') HPLC purified (Trilink Biotechnology) were separately resuspended in sterile H_2O. Gloves were used whenever handling the nucleic acids to avoid DNase or RNase contamination. The DNA strands were incubated in equimolar amounts in Buffer A (10 mM HEPES pH 8.0, 150 mM NaCl) (Table 7) 0.22 µm filter-sterilized, heated 5 min to 95°C and slow-cooled to 4°C. Immediately, double equimolar amounts of the RNA strand were added (Table 7), the mixture heated 5 min to 45°C and gradually cooled down to 4°C. The optimal formation of the scaffold (Table 7) and controls (Table 8) was analyzed by 20% native acrylamide gels detected with ethidium bromide. All nucleic acids were quantified using Abs_{260nm} in a spectrophotometer (Nanodrop, Thermo Fisher).

Table 7. Reaction mixture for transcription bubble scale-up.

	TS-CPD/NTS/RNA scale-up
NTS (80 pmol/µL)	2.5 µL (200 pmol)
TS-CPD (65 pmol/µL)	3.1 µL (200 pmol)
Buffer A (10x)	2.0 µL
Sterile deionised H_2O	7.7 µL
RNA (85 pmol/µL)	4.7 µL (400 pmol)
Total volume	20 µL

Table 8. Reaction mixtures for transcription bubble controls.

	NTS	**TS**	**TS/NTS**	**RNA**
NTS (8 pmol/µL)	1.3 µL (10 pmol)	-------	1.3 µL (10 pmol)	------
TS-CPD strand (6.5 pmol/µL)	------	1.5 µL (10 pmol)	1.5 µL (10 pmol)	------
Buffer A (5x)	1.0 µL	1.0 µL	1.0 µL	1.0 µL
Sterile deionised H_2O	2.7 µL	2.5 µL	1.2 µL	2.8 µL
RNA strand (8.5 pmol/µL)	------	------	------	1.2 µL (10 pmol)
Total volume	5 µL	5 µL	5 µL	5 µL

3. A190 gene cloning into pRS315 vector

S. cerevisiae rpa190 inserted in pFL38 vector (courtesy by Olivier Gadal's laboratory) and pRS315 plasmid (Addgene) were digested with SalI/SacI restriction enzymes (New England Biolabs (NEB)) (Table 9) at 37°C overnight. 5 µL of both double digestions were run in an agarose gel to check digestion occurred. Next, the entire digestion volumes were loaded in a 1% (w/v) agarose gel and purified using QIAquick Gel Extraction Kit (Qiagen) following the commercial protocol with minor modifications. Both purified DNA fragments were eluted separately in 30 µL Elution buffer (10 mM Tris-HCl pH 8.5 at 25°C) and quantified by Abs_{260nm} in a NanoDrop spectrophotometer (Thermo Fisher).

Table 9. Reaction mixtures of *rpa190* and pRS315 plasmid digestions.

	Insert	**Vector**
Purified pFL38-A190 plasmid (173 ng/µL)	5.8 µL (1 µg)	--------
Purified pRS315 plasmid (257 ng/µL)	--------	3.9 µL (1 µg)
Buffer CutSmart 10x (NEB)	5.0 µL	5.0 µL
SalI – HF® 20 Un/µL (NEB)	1.0 µL	1.0 µL
SacI 20 Un/µL (NEB)	1.0 µL	1.0 µL
Sterile deionised H_2O	37.2 µL	39.1 µL
Total volume	50 µL	50 µL

The digested insert and vector were ligated overnight at 22°C using T4 DNA Ligase (NEB) at 1:1 and 3:1 insert:vector molar ratio. A control reaction without the A190 gene fragment was set following standard molecular biology protocols. The whole ligation was transformed in homemade chemically competent DH5α cells, which were plated in LB plates supplemented with ampicillin. Three colonies from the ligation reactions were digested with restriction enzymes to find positive digestion profiles and later sequenced (Secugen S. L.) to confirm A190 gene insertion into the plasmid.

4. Mutant yeast strain construction

The previously obtained pRS315-A190 plasmid was used to transform a yeast strain containing a tandem affinity purification tag (TAP-tag) (Rigaut et al 1999) at the AC40 subunit (AC40-TAP). This transformation was performed by the lithium acetate method (adapted from Giezt et al 1995). Colonies were selected on YPD plates containing appropriate selection markers and analyzed by colony PCR.

5. Pol I endogenous expression and purification

5.1. Pol I endogenous expression

A *S. cerevisiae* strain with the TAP-tag fused to the C-ter of A190 subunit was further modified to truncate the C-ter of A43 subunit (Δ307-326), thus avoiding the assembly of inactive dimeric Pol I enzymes (Torreira et al 2017). This strain was used in the cryo-EM reconstruction of Pol I-EC_{CPD}. Pol I *in vitro* transcription assays were performed using a genetically modified strain expressing the TAP-tag fused to the C-ter of AC40 subunit, and a strain further lacking the last 47 C-ter residues of A12.2 subunit. For all strains, and starting from glycerol stocks, fresh YPD plates were seeded and left 2 to 3 days at 30°C. One big and isolated colony of each strain was transferred to 5 mL YPDA supplemented with ampicillin to avoid bacterial contamination and was grown 24 h at 30°C and 220 rpm. The 5 mL pre-inoculum was used to start 500 mL YPDA cultures and the cells were grown 24 h at 30°C and 220 rpm. For A190-TAP strain, a 75 L fermenter was inoculated, while 12 L of final culture were inoculated for AC40-TAP strains, until cells reached OD_{600nm} 5.5 – 6.0. Then, cells were centrifuged 30 min at 8000 *g*, 4°C, and the pellet shipped in dry ice and stored at -80°C.

5.2. Pol I purifications

All chromatographic separations were performed at 4°C and the 0.22 µm filtered buffers and protein fractions kept on ice throughout the purification. Approximately 500 g of A190-TAP or 80 g of AC40-TAP cellular pellets were resuspended with equivalent volumes of Buffer L (250 mM Tris-HCl pH 7.4, 20% (v/v) glycerol, 250 mM ammonium sulfate, 1 mM EDTA pH 8.0, 10 mM $MgCl_2$, 10 µM $ZnCl_2$, 10 mM betamercaptoethanol) supplemented with protease-inhibitors (1 mM phenylmethanesulfonyl fluoride, 2 µg/ml leupeptine, 4 mM benzamidine, 1.4 µg/ml pepstatine A), phosphatase inhibitors (50 mM sodium fluoride, 2 mM sodium pyrophosphate, 5 mM betaglycerophosphate) and DNAse (DNase I recombinant, RNase-free, Roche). Cells were lysed at 4°C with glass beads using a BeadBeater (Biospec). The lysates were centrifuged 1 h at 25000 *g* at 4°C and the supernatants carefully transferred into a cooled flask. Approximately 4 mL and 0.5 mL of IgG Sepharose 6 Fast Flow resin (GE Healthcare) respectively were equilibrated in Buffer IgG (50 mM Tris-HCl pH 7.4, 5% (v/v) glycerol, 200 mM NaCl, 1 mM $MgCl_2$, 10 µM $ZnCl_2$ and 5 mM DTT) and incubated with the supernatants overnight at 4°C in a Stuart® roller mixer (model SB3) set to 8 rpm.

The whole soluble protein incubations with IgG resin were centrifuged at 4°C, 600 *g* and 5 min. The supernatants were carefully removed with a pipette and the IgG resins were washed five times (resuspending and centrifuging each time) using a total minimum of 10 column volumes (CV). After, the IgG resin bound A190-TAP Pol I was resuspended in 10 mL Buffer IgG while IgG resin bound AC40-TAP Pol I was resuspended in 2 mL Buffer IgG. 800 µL TEV protease (0.5 mg/ml) were added to A190-TAP Pol I IgG resin and 6 uL TEV protease (10 mg/mL) were added to AC40-TAP Pol I IgG resin and the mixtures were left incubating overnight at 4°C and 8 rpm in a roller mixer (Stuart®).

Pol I released into solution was recovered by separation of the supernatant and the resin by centrifugation at 4°C, 600 *g* and 5 min. The beads were washed several times with Buffer IgG to maximize Pol I recovery. The whole supernatants were applied into cooled Poly-Prep® Chromatography Columns (Bio-Rad) to completely remove the IgG beads. The flowthrough was centrifuged 30 min at 15000 *g* and 4°C and carefully transferred to a clean tube. A190-TAP Pol I sample was loaded in a Mono Q 5/50 GL (GE Healthcare) and AC40-TAP Pol I was loaded in a Mono Q PC 1.6/5 (GE Healthcare), both anion-exchange columns equilibrated in Buffer MQ A (20 mM Tris pH 7.4, 200 mM NaCl, 1 mM $MgCl_2$, 10 µM $ZnCl_2$, 5 mM DTT). A190-TAP Pol I was eluted running a gradient from 0% to 50% Buffer MQ B (20 mM Tris pH 7.4, 1 M NaCl, 1 mM $MgCl_2$, 10 µM $ZnCl_2$, 5 mM DTT) at 0.5 mL/min in 30 CV and collected in 500 µL fractions. AC40-TAP Pol I was eluted using the same buffers and gradient conditions at 50 µL/min in 3 CV and collected in 50 µL fractions. Fractions representing the chromatogram peak profiles were analyzed in 15% SDS-PAGE gels coomassie-stained to assess protein purity and help in the decision of the pools to form. 4.5% native gels of selected fractions, run at 120 V, 4°C, for 3 h, loading 2 µg of protein and likewise coomassie-stained, were also useful in the decision of which fractions to join. The pools were formed, concentrated in a 30 kDa cut-off Amicon Ultra-0.5 mL centrifugal filter (Merck Millipore) and quantified before and after concentration to monitor protein recovery. Aliquots were frozen in liquid N_2 and stored at -80 °C. The protein was stored at the salt concentration correspondent to anion-exchange chromatography elution. The purification protocol of A190-TAP Pol I retrieved 14-subunit Pol I lacking the last 49 residues of A43 (Pol I A43ΔCt) and Pol I ΔA49/A34.5 lacking the last 49 residues of A43 (Pol I A43ΔCt ΔA49/A34.5). Because AC40 subunit is shared by Pol I and Pol III, the purification protocol of AC40-TAP Pol I

yielded intact 14-subunit Pol I, 12-subunit Pol I ΔA49/A34.5 (Pol I ΔA49/A34.5), and intact 17-subunit Pol III.

6. *In vitro* transcription elongation assays

In vitro transcription elongation assay with full DNA/RNA scaffold was performed based on previous reported methods with slight modifications (Awrey et al 1998) . The DNA and RNA oligo sequences used in this assay were: CPD-containing template DNA was the same strand as structural studies; non-template DNA *in vitro* transcription assays (5'-GCAGCCTAGTTGATCTCATATTTCATTCCTACTCAGGAGAAGGAGCAGAGG - 3'), RNA oligos (8mer for transcription extension: 5'-AUCGAGAG-3'; 11mer for cleavage assay: 5'-AUCGAGAGGAA-3'). For transcription control assays with lesion-free scaffold, the nucleic acid sequences were: lesion-free template DNA (5'-CGCTCTGCTCCTTCTCCTTTCCTCTCGATGGCTATGAGATCAACTAG - 3'), non-template DNA (5'-CTAGTTGATCTCATATTTCATTCCTACTCAGGAGAAGGAGCAGAGCG - 3') and 8mer RNA oligo labeled. In all cases, HPLC-purified oligos were obtained from TriLink. An aliquot of 8-mer or 11-mer 5'-^{32}P labeled RNA was annealed with a 1.5-fold molar amount of template DNA strand and 2.0-fold amount of non-template *in vitro* transcription assays DNA strand to form the DNA/RNA scaffold in a Buffer A (20 mM Tris-HCl pH 7.5, 150 mM KCl and 5 mM DTT). An aliquot of DNA/RNA scaffold was incubated with a five molar excess amount of purified 12- subunit Pol II, 14-subunit Pol I or *E.coli* RNAP on ice for 10 min, followed by incubation at room temperature (23°C) for 10 min. The transcription was chased by adding an equal volume of solution C (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT and 2 mM NTP). Final reaction concentrations after mixing were 30 nM scaffold, 150 nM Pol I or Pol II and 1 mM NTP. Reactions were quenched at various times (0, 0.3, 1, 3, 10, 30 and 90 min) by addition of one volume Buffer STOP (0.5 M EDTA pH 8.0). For the transcript cleavage assay, the scaffold was incubated with the RNA polymerase for 10 min on ice. The cleavage reaction started by adding 5 mM $MgCl_2$, and the temperature switched to room temperature (23°C) by water bath. In the experiment with TFIIS-stimulated cleavage assay, TFIIS was added to the reaction system at time 0 (final concentration of TFIIS: 100 nM). Cleavage reactions were quenched at various time points (0, 1, 3, 10 and 30 min). For the chase experiment, 1 mM NTP was added to the system with additional 60

s incubation before quenching the reaction. The transcript was analyzed by 12% (w/v) denaturing urea/PAGE. The gel was visualized by phosphorimaging and analyzed using Image Laboratory software (BioRad).

7. *In vitro* assembly of Pol I-EC$_{CPD}$ complex

2 µg of 14-subunit Pol I A43ΔCt were incubated with the artificial transcription bubble in a 1:1 molar ratio 1 h at 20°C in 0.22 µm filter-sterilized Buffer I (10 mM HEPES pH 8.0, 150 mM NaCl, 5 mM DTT) (Table 10). Pol I-EC$_{CPD}$ mixture, 14-subunit Pol I A43ΔCt and artificial transcription bubble controls, were loaded in 4.5% and 7% native gels. The interaction was assessed with ethidium bromide to detect the nucleic acids and, subsequently, coomassie-stained to detect the protein.

Table 10. Reaction mixtures for Pol I-EC$_{CPD}$ and controls.

	DNA/RNA scaffold	**Pol I A43ΔCt**	**PolI-EC$_{CPD}$**
Pol I A43ΔCt 6µg/µL (10.2 pmol/µL)	----------	0.34 µL (3.4 pmol)	0.6 µL (6.8 pmol)
DNA/RNA scaffold (2 pmol/µL)	2.0 µL (4 pmol)	----------	3.4 µL (6.8 pmol)
Buffer I (5x)	1.2 µL	1.2 µL	1.2 µL
$MgCl_2$ (5x)	1.2 µL	1.2 µL	1.2 µL
Sterile H_2O	1.6 µL	3.3 µL	----------
50% GlyOH + Buffer I	6 µL	6 µL	6 µL
Final volume	12 µL	12 µL	12.4 µL

8. Cryo-EM sample preparation, data collection and image processing

8.1. Sample preparation and data collection

For cryo-EM analysis, 3.5 µL of Pol I-EC$_{CPD}$ at 0.22 mg/ml were applied to glow-discharged copper 400 mesh R 2/1 grids (Quantifoil) and incubated in the chamber of a FEI Vitrobot Mark III at 4°C and 95% humidity for 15 s. The grids were blotted for 3 sec

at an offset of –3 mm on both sides and vitrified by plunging into liquid ethane cooled down to liquid nitrogen temperature.

Data were collected on a FEI Titan Krios electron microscope operated at 300 kV, using a K2 summit direct electron detector (Gatan) and equipped with the FEI automated single particle acquisition software (EPU). Images were acquired at defocus values varying between 1.5 and 4.0 µm and nominal magnification of 47,170x, yielding a pixel size of 1.06 Å. The camera was operated in dose-fractionation counting mode collecting 32 frames per movie, with a dose rate of 5.9 e^- per pixel per sec for 8 sec total exposure.

8.2. Cryo-EM data processing

2056 movies of the vitrified Pol I-EC_{CPD} complex were averaged using optical flow correction implemented in Scipion (Abrishami et al 2015) and their CTF parameters were estimated using CTFFIND4 (Rohou et al 2015). In a first step, 1864 particles were picked manually and reference-free 2D classes were generated, five of which were used for template-based auto-picking after low-pass filtering to 20 Å. Approximately 460,000 particles were automatically selected and extracted with a 288 pixel box using Relion (Kimanius et al 2016), also employed for subsequent processing. Four rounds of reference-free 2D class averaging yielded a stack of 291,639 good-quality particles that were used to generate an initial 3D model employing a reference generated from the PDB entry 4C3I filtered to 60 Å. Four independent runs of 3D classification were performed using masks for the nucleic acid scaffold inside the cleft, A49-Nt/A34.5, the stalk, and upstream DNA plus the A49-Ct. In each case, the best classes were selected and subsequently refined, followed by correction of local defocus of the individual particles using GCTF (Zhang 2016) and local motion and radiation damage using particle polishing as implemented in Relion. Final post-processing was performed using automatic masking and B-factor sharpening. FSC and local resolution estimations were performed using the routines implemented in Relion (Kimanius et al 2016). Subsequent structural analysis was performed with the model derived from the initial map.

9. Structure modelling

The available structures of undamaged Pol I EC (PDBs: 5M3F, 5M5X) were fitted into the cryo-EM maps using UCSF Chimera (Pettersen et al 2004) and used as starting point for model building, which was performed in Coot (Emsley et al 2004). The structures

were refined using real space refinement as implemented in Phenix (Adams et al 2010). The same procedure was used for the cryo-EM map of Pol I-EC_{CPD} + upstream DNA with fully ordered A49, using as starting point the models 5W66 and 5M64. Refinement statistics of both structures are summarized in Table 11.

Table 11. Cryo-EM data collection and refinement statistics.

	Pol I-EC_{CPD}	**Pol I-EC_{CPD} + upstream DNA with A49 fully ordered**
Pixel size (Å/pixel)	1.06	
Number of grids	1	
Days of data collection	2	
All micrographs	2,056	
Selected micrographs	1,878	
Particles after 2D-class	291,639	
Final number of particles	254,079	60,297
Resolution (Å)	3.6	4.6
AccuracyRotations (°)	1.46	1.78
AccuracyTranslation (pixel)	0.80	1.05
Map sharpening B-factor ($Å^2$)	-94	-133
Ramachandran plot		
Outliers (%)	0%	0%
Allowed (%)	7.5%	8.5%
Favoured (%)	92.5%	91.5%
Map CC (around atoms)	0.711	0.711
RMSD bond lengths (Å)	0.06	0.05
RMSD bond angles (°)	0.87	0.82
All-atom clashscore	5.46%	6.70%
Rotamer outliers (%)	0.05%	0.05%
C-beta deviations	0	0
EMDB code	EMD-0146	EMD-0147
PDB code	6H67	6H68

10. Data availability

Pol I-EC_{CPD} and Pol I-EC_{CPD} + upstream DNA with A49 fully ordered cryo-EM maps (3.6 Å and 4.6 Å, respectively) were deposited in the Electron Microscopy Database under accession numbers EMD-0146 and EMD-0147. The derived atomic models were deposited in the Protein Data Bank under accession codes 6H67 and 6H68, respectively.

RESULTS

1. Expression and purification of Pol I

In order to obtain the protein complex with high purity and homogeneity and due to the difficulty to assemble the enzyme from the individual components or through co-expression strategies, Pol I was obtained from yeast endogenous expression. The strain contained a deletion from residues 307 to 326 at the A43 subunit (A43ΔCt). Recent studies revealed that this deletion was enough to disrupt Pol I dimers formation *in vivo* (Torreira et al 2017); therefore, the modified strain allowed us to obtain the enzyme in the monomeric state. The tandem affinity purification (TAP) tag method to purify proteins complexes preserving their structural integrity was used (Rigaut et al 1999). As expected, the yield was significantly lower compared to overexpression protocols. The most successful purification using A190-TAP strain yielded 35 µg of Pol I per L of fermenter culture. In this purification, Pol I was isolated using two chromatography steps. The purification started with a high-affinity step based on the interaction between the protein A in TAP-tag and the IgG resin, allowing the removal of most contaminants at a very early stage. After extensive washing of the IgG beads and elution by cleavage with the TEV protease, a fraction containing the enzyme was obtained. An additional purification step was an anion-exchange chromatography. Using a NaCl gradient, this step separated Pol I from nucleic acids and completely removed the TEV protease and minor contaminants. The purification allowed the identification of different Pol I variants with different subunit composition (Table 12). Importantly, an enriched 14-subunit Pol I population was systematically obtained, constituting the most significant isolated species, which was used in the cryo-EM studies.

Table 12. Pol I variants detected in the A190-TAP purification.

Subpopulation	Nomenclature	Description
Pol I A43ΔCt	Pol I	14-subunit Pol I lacking last 20 C-ter residues in A43 (A43Δ307-326)
Pol I A43ΔCt ΔA49/A34.5	Pol I*	12-subunit Pol I lacking last 20 C-ter residues in A43 (A43Δ307-326) and A49/A34.5 heterodimer

The A190-TAP Pol I anion-exchange chromatogram revealed one asymmetric peak showing a “shoulder” at higher salt concentration (Figure 7A). As expected, analysis of the eluted fractions by SDS-PAGE demonstrated that A43 migrated at 37 kDa and not around 43 kDa (Figure 7B), proving that the enzyme lacked full length A43. Also, SDS-

PAGE analysis of the latest anion-exchange fractions showed 12-subunit Pol I (Pol I*) lacking A49 and A34.5 subunits (Figure 7B).

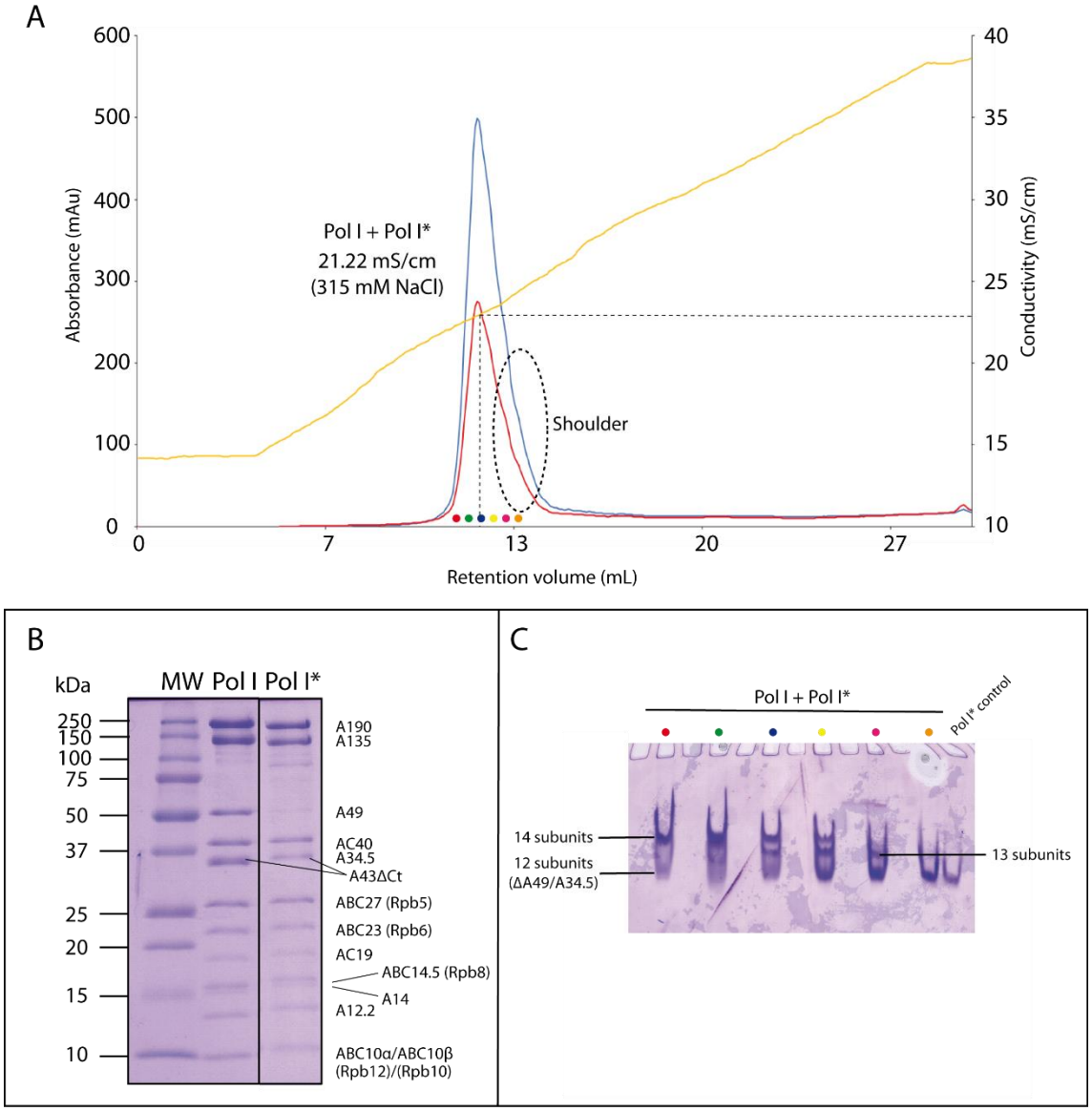

Figure 7. A190-TAP A43ΔCt Pol I purification. (A) Anion-exchange chromatography of Pol I. Abs_{280nm} and Abs_{260nm} curves are in blue and red, respectively, and conductivity in orange. Red, green, blue, yellow, pink and orange spheres represent fractions analyzed in (C). (B) Coomassie-stained 15% SDS-PAGE showing Pol I and Pol I*. (C) 4% native gel of Pol I + Pol I* peak in (A).

When loading individual fractions of the peak in a native gel, three bands were observed (Figure 7C). The intensity of the slow-migrating band is higher at low salt concentration. The relative intensities change as the peak elutes and eventually, the last fraction contains only the fast-migrating band. The upper band corresponded to Pol I and the band below to Pol I*. Thus, the anion exchange chromatography did not completely resolve 14-subunit Pol I and Pol I*. Additionally, in the native gels, an extra band between Pol I and

Pol I* was visible (Figure 7C). This extra band in the native gel corresponds to Pol I A43ΔCt ΔA49. In fact, the existence of Pol I variants lacking only A49 or A34.5 might occur, as both subunits present extended arms fixing them to A135 in Pol I (Fernández-Tornero et al 2013).

2. *In vitro* assembly of the DNA/RNA scaffold and Pol I-EC$_{CPD}$

To understand the structural organization of Pol I in the presence of a CPD lesion, in collaboration with Professor Dong Wang's laboratory at University of California San Diego (UCSD), a DNA/RNA scaffold was designed in order to mimic nucleic acids during elongation. This DNA/RNA scaffold contained a CPD lesion at positions *i*+1/*i*+2 of the TS. It consisted of a 52-mer DNA duplex with a 12-nucleotide mismatched bubble region and a 10-mer RNA strand complementary to the DNA TS. Analysis of the DNA/RNA scaffold in a polyacrylamide native gel reveals different electrophoretic mobility for single-stranded chains, TS/NTS double helix strands and the full DNA/RNA scaffold. Thus, the DNA/RNA scaffold was efficiently formed (Figure 8A). Moreover, there is a small signal for free TS/NTS double helix.

In order to check Pol I-EC$_{CPD}$ formation, the enzyme was incubated 1 hour at 20°C with different molar ratios, and the incubations were analyzed by 4.5% and 7% native gels (Figure 8B). As controls, Pol I and the DNA/RNA scaffold were run separately. Analysis of the coomassie-staining reveals that there is a migration shift between Pol I and Pol I-EC$_{CPD}$, suggesting protein complex and nucleic acid scaffold interaction (Figure 8B, upper left panel). In addition, ethidium bromide detection of the DNA/RNA scaffold in complex with Pol I generated a signal overlapping with the coomassie-stained band (Figure 8B, all panels), confirming the interaction. Also, the DNA/RNA scaffold control migrates faster than the nucleic acids in the elongating complex (Figure 8B, lower right panel). Finally, it can be observed a low percentage of free DNA/RNA scaffold in the 1:1.2 molar ratio incubation, suggesting that Pol I was completely occupied by the DNA/RNA scaffold and a small amount of nucleic acids is free.

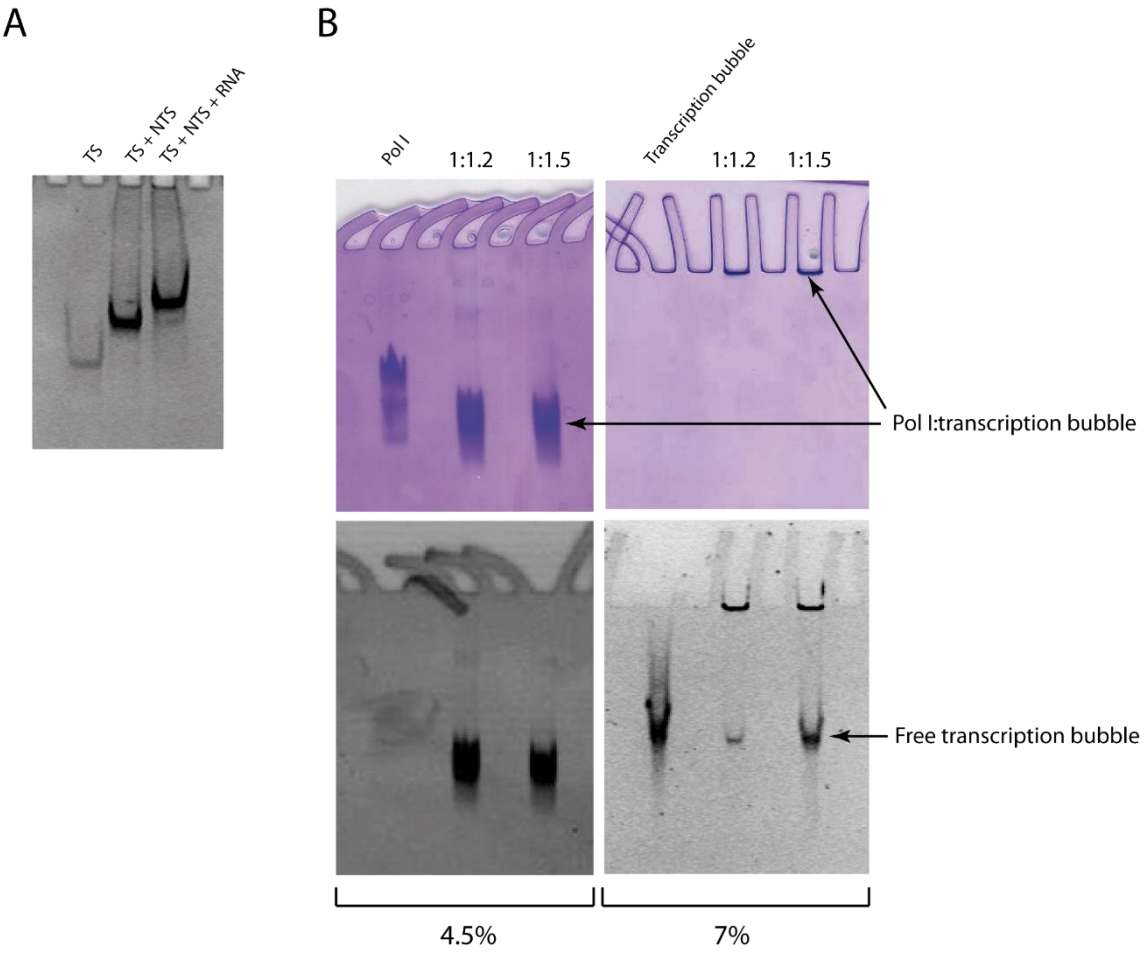

Figure 8. Electrophoretic mobility shift assay testing DNA/RNA scaffold and Pol I-EC$_{CPD}$ assemblies. (A) 20% acrylamide native gel of artificial transcription bubble assembly detected with ethidium bromide. (B) 4.5% (left side) and 7% (right side) acrylamide native gels of Pol I incubation with artificial transcription bubble and controls. Upper panels: coomassie-staining. Lower panels: ethidium bromide detection.

3. Cryo-EM structure of Pol I-EC$_{CPD}$

With the objective of obtaining the cryo-EM structure of Pol I-EC$_{CPD}$ at high resolution, we collected images in a Titan Krios equipped with a direct electron detector. The 2D micrographs revealed adequate protein concentration in the grid and a proper ice thickness (Figure 9A). Two sequential pre-processing steps were done; first, 2,056 movies were aligned using optical flow correction, implemented in Scipion (Abrishami et al 2015), then the CTF parameters of each micrograph were estimated using Ctffind4 (Rohou and Grigorieff 2015). A manual micrograph selection was done by inspection of Thon rings, allowing to discard 178 micrographs based on three factors: defocus not in the range -1.0 to -4.0 µm, drift or the astigmatism. Reference-free 2D averages showed a significant level of detail (Figure 9B), revealing secondary structure elements. To avoid

the bias in the first 3D refinement, we used as reference the 14-subunit crystal structure of Pol I in the inactive state (PDB: 4C3I) (Fernández-Tornero et al 2013) filtered at 60 Å, which shows the open cleft and does not contain nucleic acid scaffold inside. To obtain the structural information of the different regions of interest, four independent runs of focused 3D classification were performed (Figure 9C), using as reference the previous 3D-refine cryo-EM map: i) in the DNA/RNA scaffold within the cleft, ii) in the A49-Nt/A34.5 heterodimer, iii) in the stalk, and iv) in the upstream DNA plus A49-Ct.

The first 3D classification was focused on the artificial transcription bubble. The particles were divided in four classes and three of them contained density for the area of interest. The particles from these classes were joined together and 3D refinement was performed. The cryo-EM map yielded a resolution of 4.2 Å, allowing recognition of α-helices and β-strands. After per-particle CTF determination (Zhang et al 2016), per-particle motion correction (particle polishing in RELION) and post-processing map sharpening, the procedure yielded a map of Pol I-EC_{CPD} identical to the 4.2 Å reconstruction but attaining a resolution of 3.6 Å (Figure 9C). Besides, local resolution at regions close to the active site showed detail up to 3.4 Å. Despite having a slight DNA/RNA scaffold excess in the sample, around 12.4% of the total particles after 2D-classification did not exhibit DNA/RNA scaffold or showed poor density map for it.

Mobile regions were modeled using maps derived from focused 3D classifications. The A49-Nt/A34.5 heterodimer module in certain conditions can dissociate from the enzyme (Huet et al 1976) and the 3D focused classification allowed to sort between Pol I and Pol I*. Only 43.4% of the total particles after 2D-classification built the cryo-EM map, yielding a resolution of 4.2 Å. 3D focused classification based on the stalk contained 21% of the particles and the final cryo-EM map reached 4.4 Å of resolution. The last focused refinement was done using as mask the A49-Ct domain plus the upstream DNA. Particles composing this map were 20.7% of the outputted from 2D classification and the global resolution was 4.6 Å. Besides the A49-Ct and the upstream DNA, this map showed density that corresponds to the A49 linker, the stretch of residues that links A49-Ct and A49-Nt. The local resolution maps revealed that the highest resolution was in the enzyme core (blue color) while the more flexible areas extended toward the periphery (red color). FSC curves were calculated for each cryo-EM map (Figure 9D).

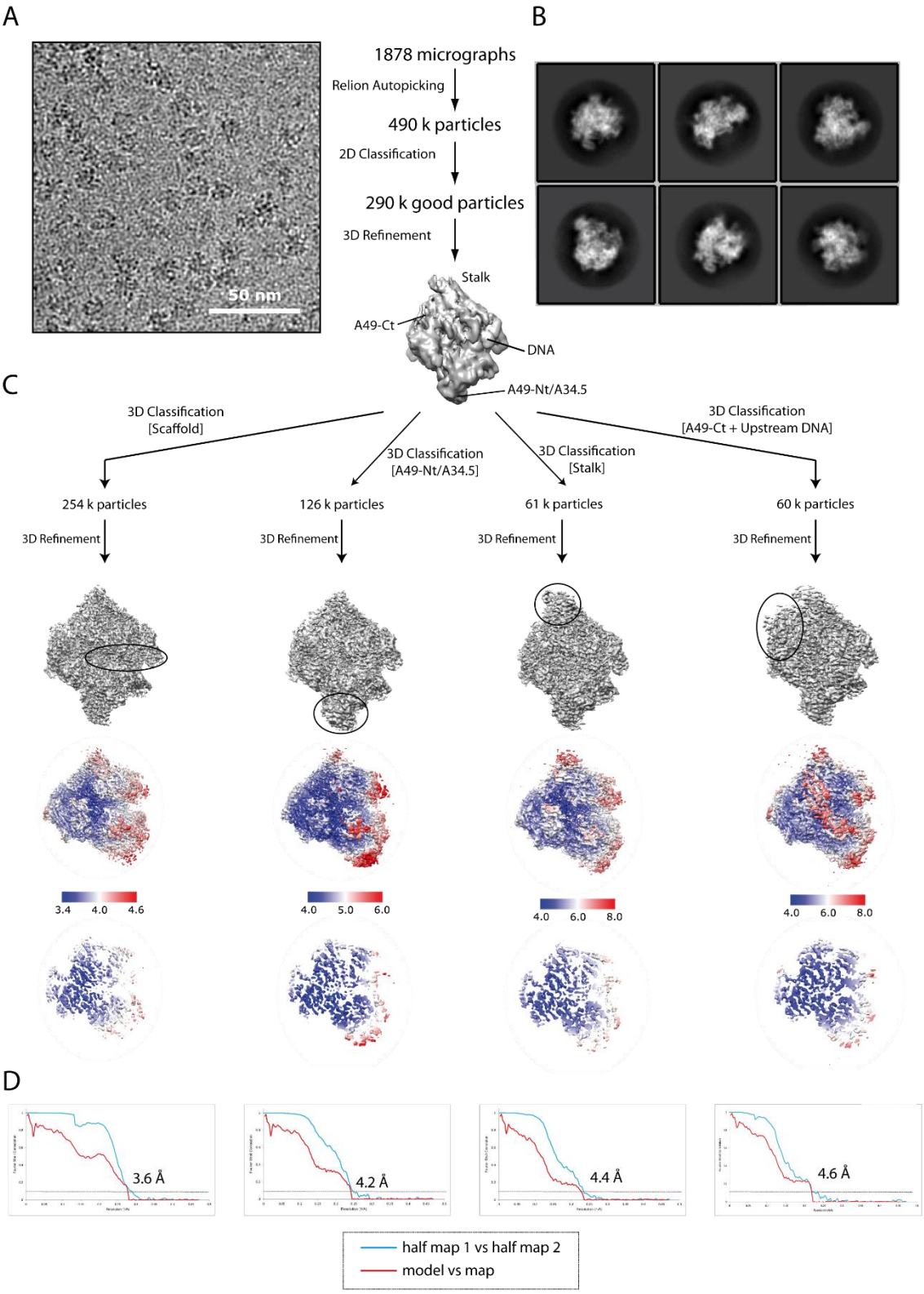

Figure 9. Cryo-EM image processing. (A) Typical field of the aligned movies. (B) Initial reference-free 2D averages showing a significant level of detail. (C) Data processing strategy showing the first 3D-refine with good-quality particles and the final maps resulting from the four focused classifications. Below, local resolution estimation according to the legend. For each map, an overall view (above legend) and a central slice (below legend) are presented. (D) FSC curves between half maps according to gold standard (blue) and between the cryo-EM map and derived atomic model (red).

The first three focused maps allowed us to reconstruct the atomic model of Pol I-EC_{CPD} at highest resolution (Figure 10A). The principal flexible regions lacking density in Pol I-EC_{CPD} structure were: A12.2-Ct, NTS at the mismatch, A49-Ct, upstream DNA and the A190 DNA-mimicking loop. Pol I core, the DNA/RNA hybrid and DNA downstream were refined against the 3.6 Å cryo-EM map. Although the second and third 3D-focused maps overall resolutions were lower (4.2 Å and 4.4 Å, respectively), these maps showed stronger density for A49-Nt/A34.5 heterodimer and the stalk, allowing its modelling. Structural details of significant regions in Pol I-EC_{CPD} corroborate the quality of the cryo-EM reconstruction and the derived atomic model (Figure 10B). Two important built regions inside the highest resolution map are shown: the CPD lesion (Figure 10C, upper right panel) and the active site including the magnesium ion (Figure 10C, lower right panel).

Even though the last map presented the lowest resolution (4.6 Å), its significance relies in the fact that it contains extra density with respect to the other maps. This map showed strong density for A49-Ct, A49 linker and the upstream DNA (Figure 11A), allowing to build these structural elements. However, the external area of the upstream DNA and the mismatch of NTS were not present in any of the maps (Figure 11B), likely due to their high flexibility. The 4.6 Å map also showed density for the DNA/RNA hybrid, downstream DNA, the A49-Nt/A34.5 module and the stalk.

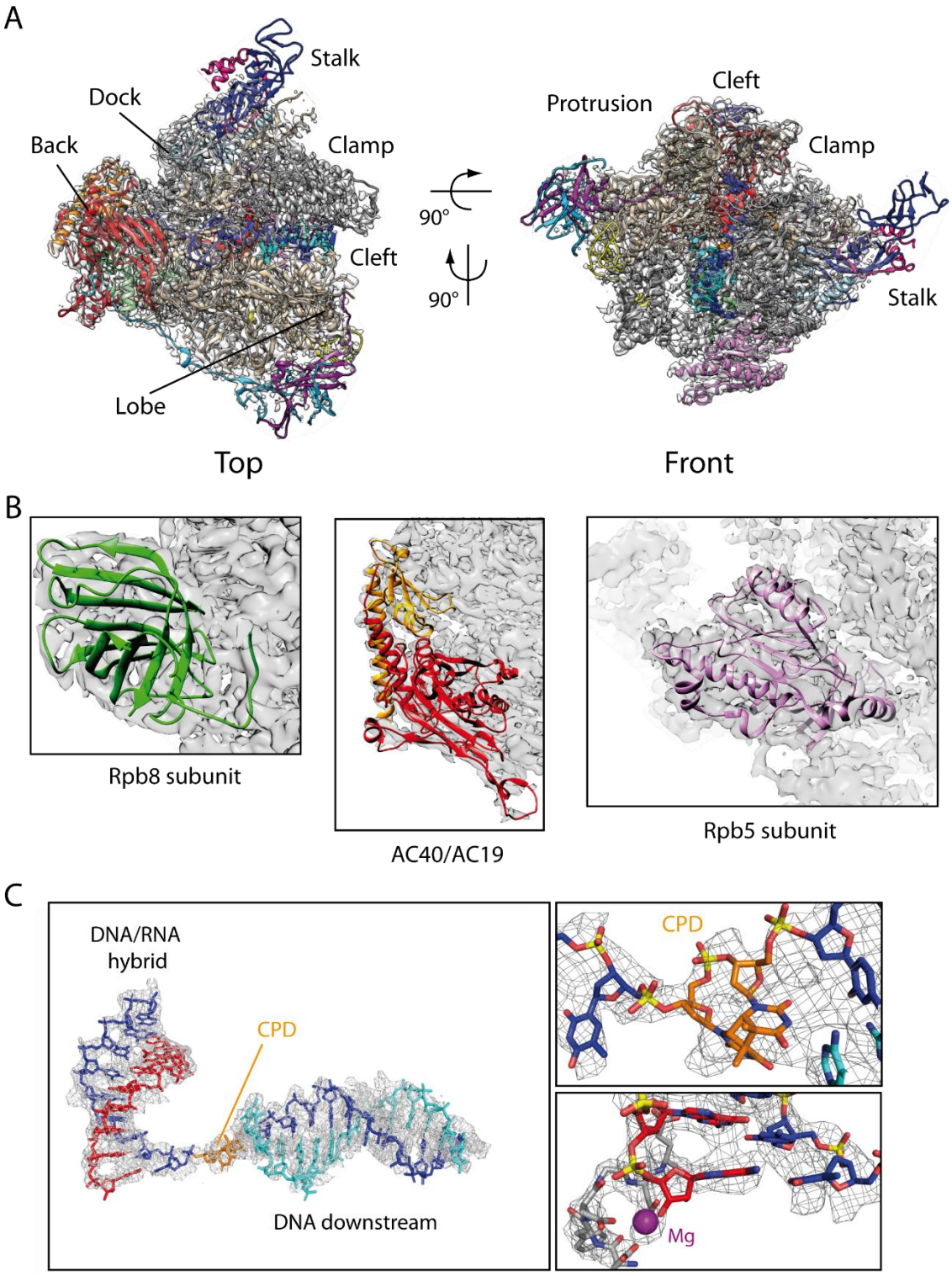

Figure 10. Cryo-EM structure of CPD-stalled Pol I. Extracted from Sanz-Murillo et al (2018). (A) Cryo-EM reconstruction of Pol I-EC_{CPD} at 3.6 Å resolution superposed with the derived atomic model, indicating the different subunits in the enzyme. (B) Close-up views of regions in the 3D map of Pol I-EC_{CPD} and the selected domains of the atomic model colored as in Figure 1. (C) Cryo-EM map and derived model of the scaffold in Pol I-EC_{CPD} at a resolution of 3.6 Å (left panel). Zoomed-in views around the CPD lesion (upper right panel) and the active site with the magnesium ion (lower right panel).

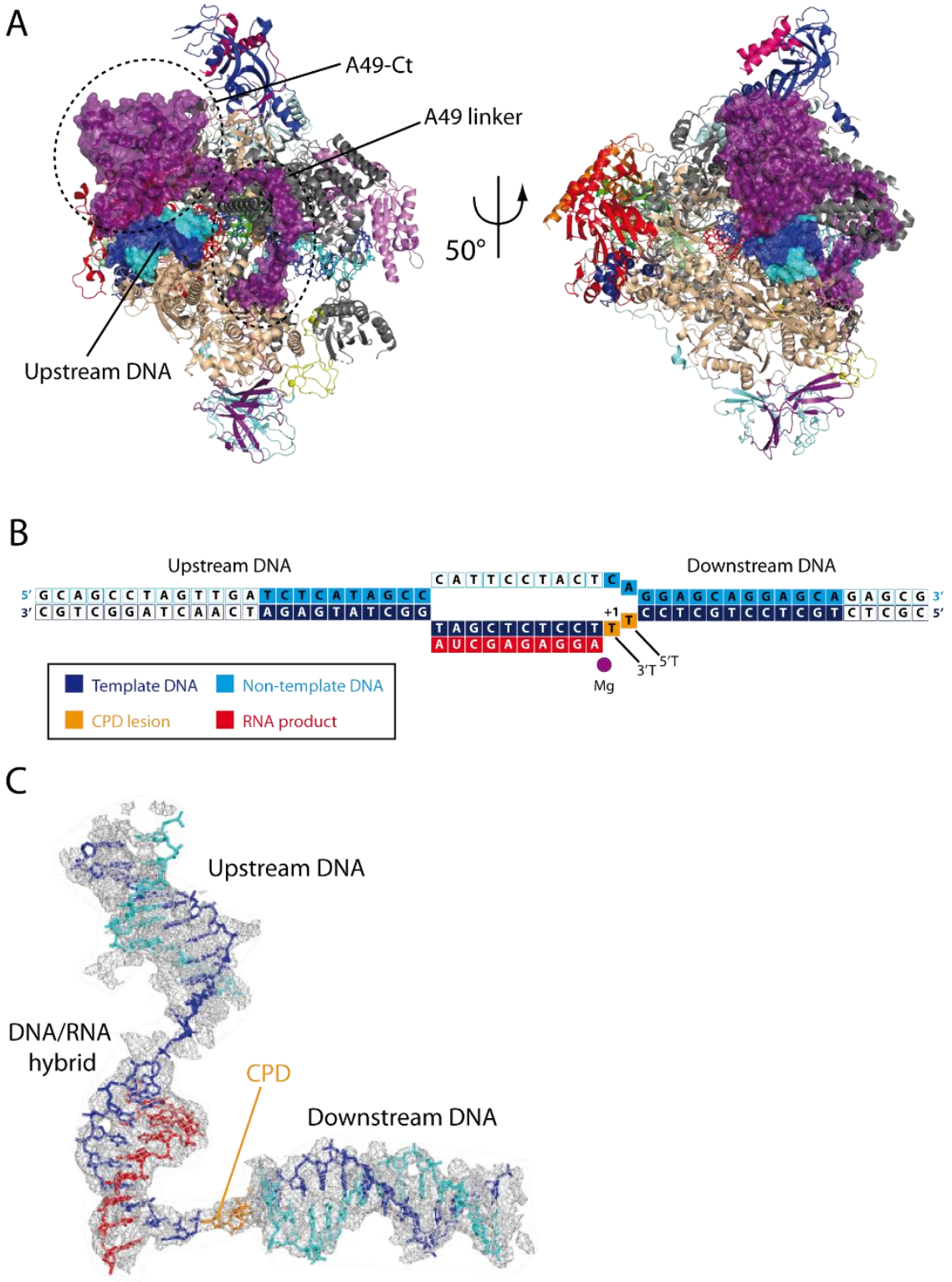

Figure 11. Cryo-EM structure of Pol I-EC$_{CPD}$ with A49 subunit fully-ordered. (A) Derived atomic model form the map showing the A49-Ct, A49 linker and upstream DNA represented as surface. (B) Schematic diagram of the artificial transcription bubble. Filled squares display the built nucleotides in the atomic model. (C) Cryo-EM map of the DNA/RNA scaffold at a 4.6 Å and derived atomic model.

4. The CPD lesion occupies an intermediate position above the Pol I bridge helix

With the objective of determining the translocation cycle stage of the Pol I-EC_{CPD}, alignment of different atomic models of Pol II or bRNAP with Pol I-EC_{CPD} structure were carried out, based on BH sequence (Figure 12). Five different structures were used, three of them representing a distinct translocation state inside the transcription cycle (PDBs 1I6H pre-translocated state, 1Y1W post-translocated state and 2E2H NTP insertion state), PDB 4A3D showing an initial RNA addition intermediate (Cheung et al 2011) and PDB 6FLP exhibiting a paused bRNAP (Guo et al 2018).

The CPD thymines in Pol I-EC_{CPD} accommodate at a position that is similar to that occupied by nucleotides *i*+1/*i*+2 in the Pol II structure of the pre-translocated state (Figure 12A), but in Pol I they are slightly more advanced toward the active site. Besides, the base pair immediately upstream of the lesion occupies an intermediate location between canonical positions *i* and *i*−1 in the Pol II pre-translocated state, suggesting that while the DNA/RNA hybrid moves forward to the post-translocation state, the downstream DNA remains blocked. Comparison with Pol II post-translocated state showed that in Pol-EC_{CPD}, the NTP entry site is partially occluded due to the position of the 3′ end RNA (Figure 12B). Superposition with Pol II in the pre-insertion stage showed that the incoming NTP lies at 1.9 Å from the base at the 3′end of RNA molecule in Pol I-EC_{CPD} (Figure 12C). This distance is too short to allow nucleotide entrance.

Furthermore, a similar configuration of the DNA/RNA hybrid has been observed for Pol II initiation complexes with four- to six-nucleotide RNAs (Cheung et al 2011) (Figure 12D). Paused bacterial RNAP (Guo et al 2018, Kang et al 2018) (Figure 12E) also exhibits an intermediate translocation state similar to Pol I-EC_{CPD}. The RNA strand adopts a post-translocated state, still maintaining TS pairs. Besides, the TS presents a halfway conformational state, leading the *i*+1 TS base pair located between the thymine dimer (Guo et al 2018). Overall, the results show that Pol I-EC_{CPD} complex presents an intermediate translocation state that is incompatible with nucleotide addition opposite the CPD lesion.

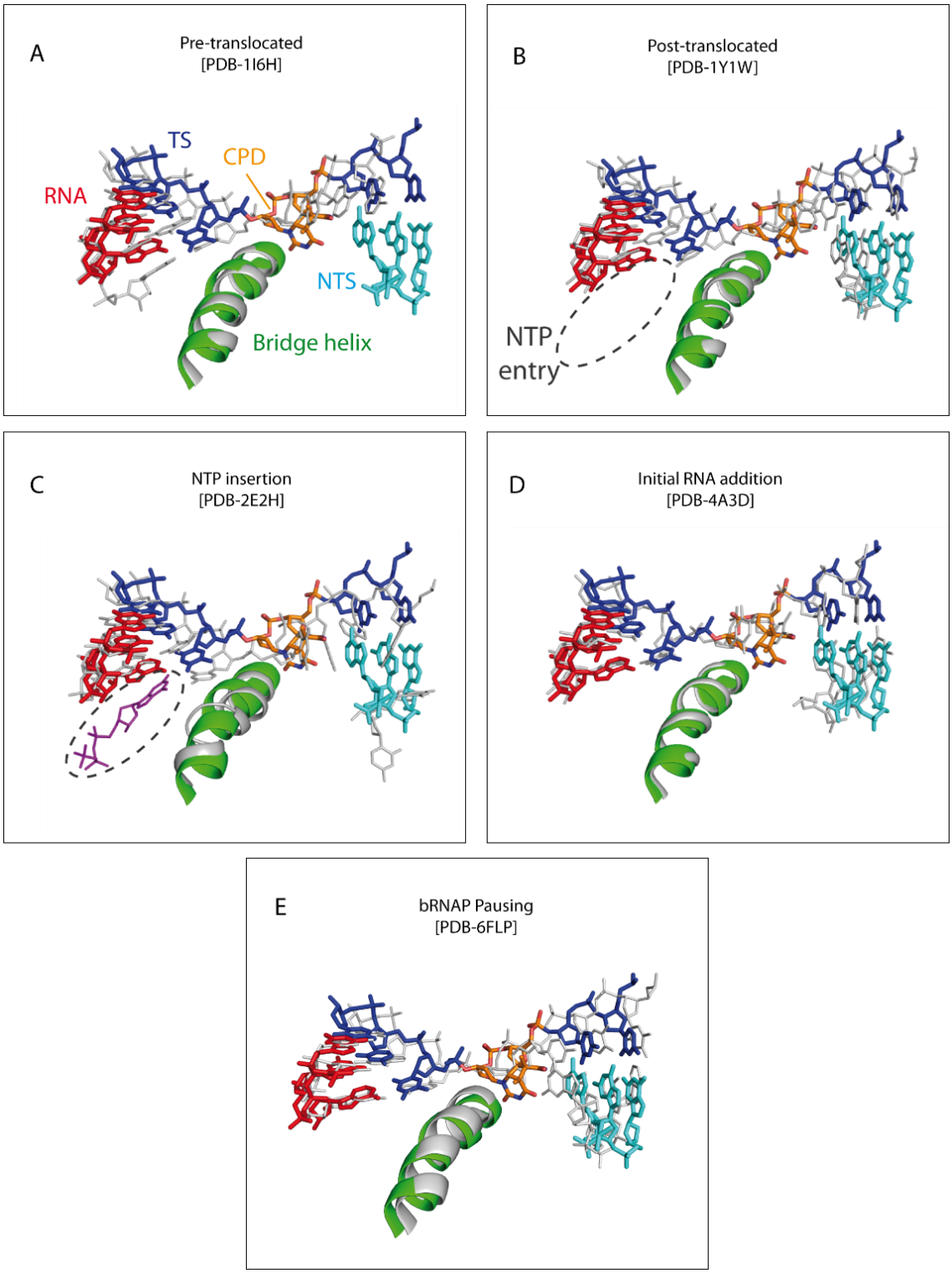

Figure 12. Translocation state of CPD-stalled Pol I. Extracted from (Sanz-Murillo et al 2018) Superposition of Pol I-EC_{CPD} with Pol II in the pre-translocated (A), post-translocated (B), and NTP insertion (C) stages of the nucleotide addition cycle, Pol II in an initial RNA addition intermediate (D) and paused bRNAP (E).

5. CPD-Mediated stalling is different in Pol I and Pol II

To obtain further information about CPD-mediated Pol I stalling, we compared our structure with that of CPD-stalled Pol II (Walmacq et al 2012) and Pol II inhibited by α-amanitin (Brueckner and Cramer 2008), both structures showing an intermediate translocation state. To do this, atomic structure models were aligned using the BH as reference point (Figure 13).

In the Pol II-EC_{CPD}, the CPD thymines at positions *i*+1/*i*+2 are disengaged from downstream DNA but the 3'T is not able to reach the canonical position of the active site (Figure 13A, upper panel). This situation leaves an enlarged NTP binding site that allows for non-templated addition (Walmacq et al 2012). In contrast, the NTP entry site in CPD-stalled Pol I is reduced by the nucleotide at the 3' end of RNA (Figure 13A, upper panel). Besides, the structure revealed that the CPD lesion in Pol I-EC_{CPD} and the immediately upstream base pair at the *i* position lie about 7 Å backward, in comparison with the equivalent Pol II structure. This suggests that Pol I shows an early blockage of translocation upon encounter of the CPD lesion in comparison to Pol II. The kinked BH and the wedged TL of Pol I-EC_{CPD} are absent in the Pol II-EC_{CPD} (Figure 13A, below panel).

We further compared the Pol I-EC_{CPD} structure with that of Pol II inhibited by α-amanitin (Brueckner and Cramer 2008, Kaplan et al 2008). Comparing both structures, the scaffold in α-amanitin-arrested Pol II showed a configuration where downstream nucleotides at positions *i*+1/*i*+2 are located similarly as in CPD-stalled Pol I (Figure 13B, upper panel). However, the *i*+1 base in α-amanitin-arrested Pol II is crossed over the bridge helix and reaches the canonical *i*+1 template position, while the 3'T in the CPD lesion in CPD-stalled Pol I remains on the downstream side of the BH (Figure 13B, upper panel), which is likely due to the covalent bonds between the thymines in the CPD lesion, forming a cyclobutane. Moreover, in the case of Pol I-EC_{CPD}, the RNA base pair at the *i* position is tilted toward the NTP site and its DNA template counterpart is retarded and tilted with respect to that observed in the α-amanitin-inhibited Pol II. With respect to the BH, this element, in both cases, is kinked and the BH residues that establish contacts with template DNA in Pol I-EC_{CPD} are either not conserved or differently oriented in Pol II inhibited by α-amanitin (Figure 13B, below panel).

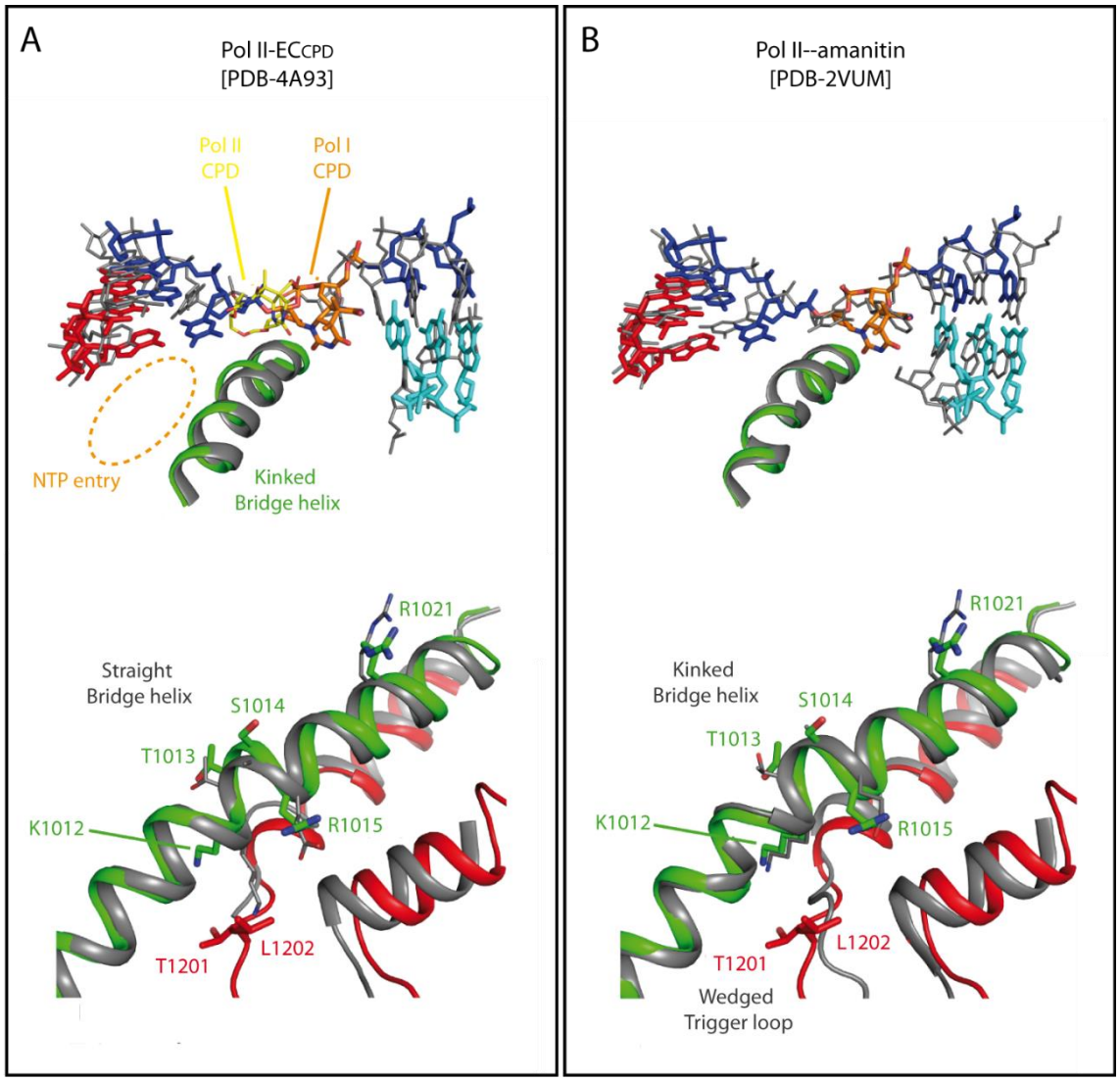

Figure 13. Comparison of CPD-stalled Pol I and Pol II. Extracted from (Sanz-Murillo et al 2018). (A, upper panel) Side view of the superposition with Pol II-EC$_{CPD}$, with Pol II shown in gray and the CPD lesion shown in yellow. (A, below panel) Close-up view where only the bridge helix and trigger loop are shown. (B, upper panel) Side view of the superposition with α-amanitin-inhibited Pol II shown in gray. (B, below panel) Close-up view, where only the bridge helix and trigger loop are shown.

These results suggest that the mechanism of Pol I stalling at CPD lesions is caused by blockage of translocation rather than by nucleotide misincorporation opposite the lesion, as described for Pol II (Brueckner et al 2007). The comparison with Pol II-EC$_{CPD}$ and α-amanitin-inhibited Pol II structure confirmed that Pol I-EC$_{CPD}$ is stalled at an early stage of translocation.

6. Specific contacts with Pol I stabilize the CPD lesion

To obtain further insights into CPD-mediated Pol I stalling, Pol I-EC_{CPD} was compared with the known atomic models for lesion free Pol I-EC. The configuration of Pol I-EC_{CPD} structure is overall similar to that found in these atomic structures (Neyer et al 2016, Tafur et al 2016). However, the stalled Pol I represented an intermediate of translocation, whereas the transcribing Pol I structures corresponded to the canonical post-translocated state. In addition, the DNA upstream of Pol I-EC_{CPD} with fully-ordered A49 showed a tilt in comparison with the cryo-EM structure of lesion-free Pol I-EC (Figure 14A).

Inside the cleft, several Pol I structural elements undergo conformational changes that are likely relevant for enzyme stalling. The cleft in the Pol I-EC_{CPD} is closed, compatible with the recently defined conformation III of the enzyme (Fernández-Tornero 2018). In lesion-free Pol I-EC, the BH, comprising residues 992-1028 in subunit A190, presents a fully regular helical configuration, while the BH in CPD-stalled Pol I-EC is kinked at its central region (residues 1009-1015 in A190). This kink showed a maximum distance of 1.2 Å toward the TS respect to the fully regular helix (Figure 14B and 14C).

Within the kinked region, the R1015 in A190 subunit is conserved in Pol I from yeast to humans, as shown in the multi-alignment, but does not exist in other types of RNAPs, including bacterial RNAP, archaeal RNAP and eukaryotic Pol II and Pol III. Besides, these alignments revealed that S1014 is not conserved in eukaryotic Pol II, Pol III and bacterial RNAP. However, a few organisms present an amino acid serine in Pol II and archaeal RNAP (Figure 14D) at this position.

Strikingly, the side chain of R1015 in Pol I-EC_{CPD} complex lies at 5.3 Å of distance from the 3'T in the CPD lesion, which is compatible with a cation-π interaction, while in lesion-free EC this distance is 8.2 Å, which is too far for such interaction (Figure 14E). Moreover, S1014 lies at hydrogen bond distance from the TS backbone next to the CPD lesion, unlike Pol I-EC, where this residue is toward the opposite side of the TS. In Pol I-EC_{CPD}, R1021 lies at hydrogen bond distance from the backbone phosphate within the thymine dimer (Figure 14C), maintaining a similar position in Pol I-EC. The BH kink associates with a wedged conformation of the trigger loop, such that residue T1201 lies next to BH residue K1012. Additional rearrangements inside the Pol I cleft mainly affect switch loop 2 in A190, which is involved in clamp swinging in Pol II (Cramer et al 2001). In Pol I-EC_{CPD}, this loop approaches downstream DNA and contacts the backbone

phosphate within the CPD lesion through R468. Moreover, two basic residues, K462 and K463, alter their configuration to approach the template strand, with K463 lying at hydrogen bond distance (Figure 14F). These results suggest that the structural rearrangement inside the cleft generates a network of interactions around the CPD lesion that likely contributes to enzyme stalling.

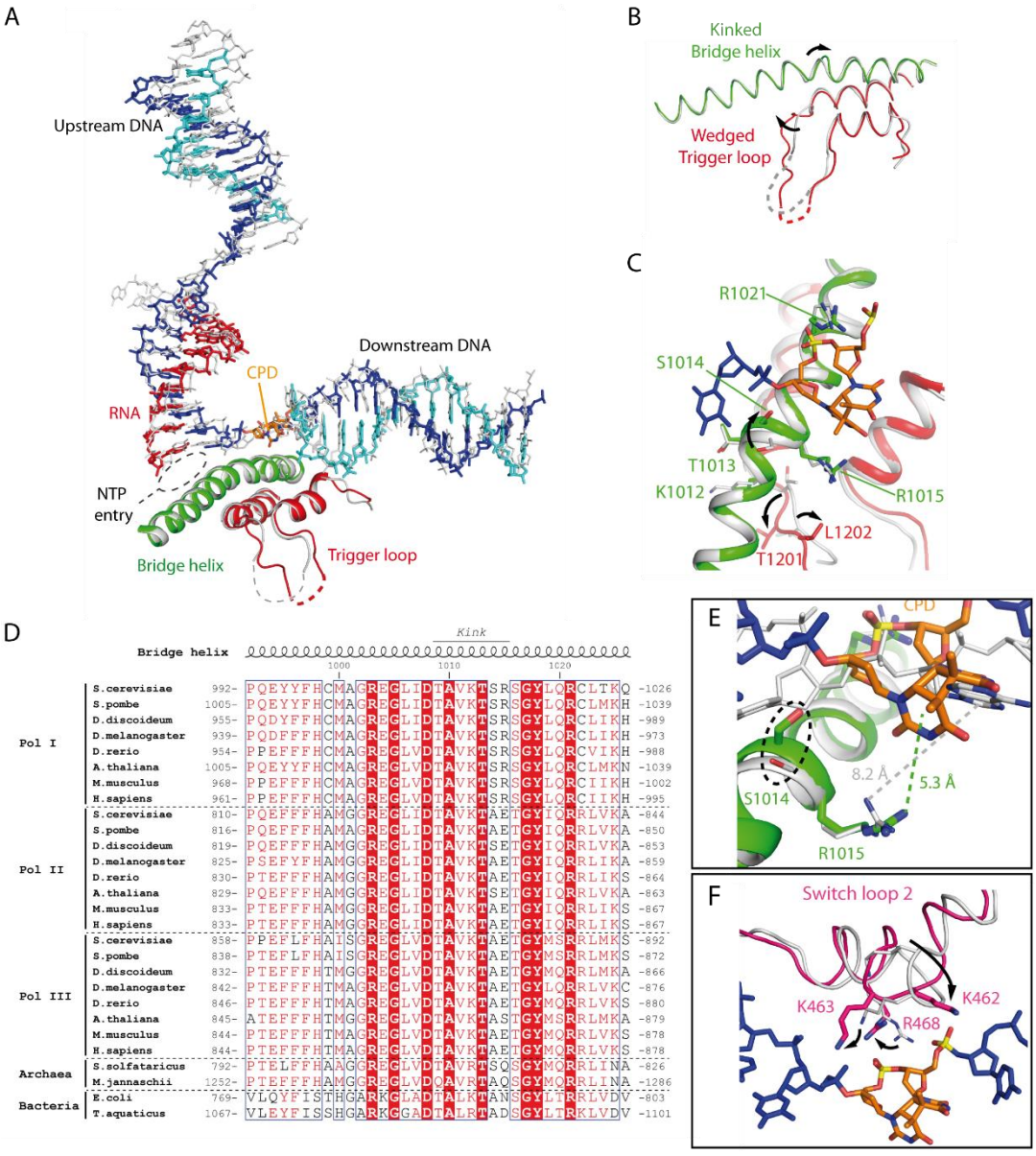

Figure 14. Comparison of CPD-stalled Pol I with undamaged Pol I-EC. (Extracted from Sanz-Murillo et al 2018). (A) Side view of the superposition with undamaged Pol I-EC shown in gray. (B) Front view where only the BH and the TL are shown. (C) Close-up view around the BH and TL. (D) Sequence alignment of RNA polymerase BH. Fully and partially-conserved residues are boxed in red and white, respectively. Top numbering is for S. cerevisiae Pol I. (E) Close-up view around the CPD lesion, with distances between the side chain of R1015 and the closest base in downstream TS. (F). Close-up view around switch loop 2.

7. Pol I blocks at CPD lesions

For the purpose of investigating the behavior of Pol I in the presence of UV light-induced DNA damage, in collaboration with Professor Dong Wang's laboratory, *in vitro* transcription elongation assays were carried out. The transcriptional activity of Pol I was tested using a nucleic acid scaffold containing a site-specific CPD lesion on the template strand (Figure 15A). The artificial scaffold includes an 8-mer RNA molecule whose 3' end base-pairs with the TS two nucleotides before the CPD lesion (Figure 15B).

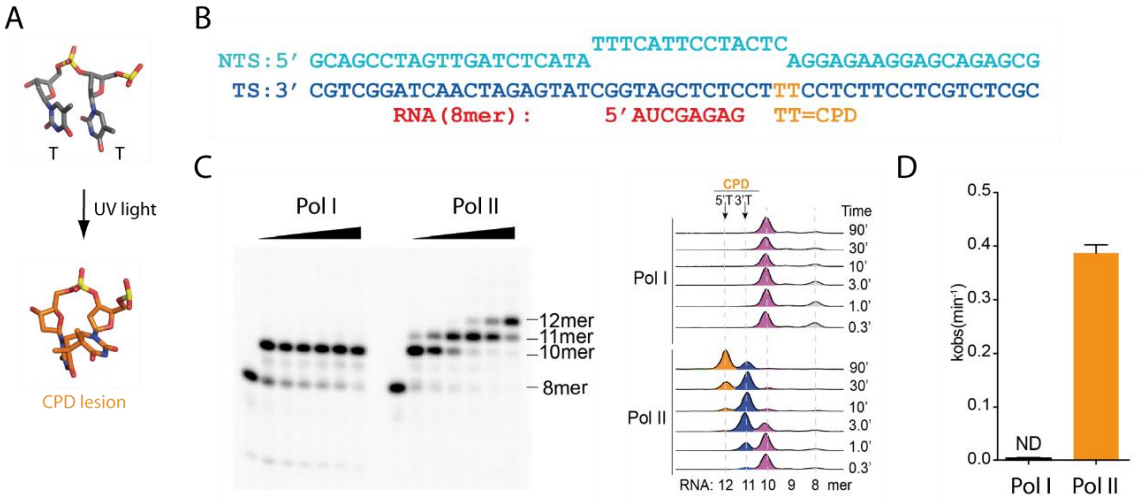

Figure 15. CPD lesion effect on Pol I and Pol II elongation. Extracted from (Sanz-Murillo et al 2018). (A) Structure of the CPD lesion. (B) Scheme of DNA/RNA scaffold. Template DNA, non-template DNA and RNA are shown in blue, cyan and red respectively. Orange corresponds to CPD lesion. (C) *In vitro* transcription assays of Pol I and Pol II on a scaffold containing a site-specific CPD lesion in the presence of 1 mM NTPs. Different RNA lengths are highlighted in different colors, with time points indicated on the graph in min. (D) Pol I vs Pol II k_{obs} graphic (ND, not detectable).

Transcription assays of Pol I versus Pol II showed that upon NTP addition two nucleotides were incorporated to the RNA molecule by Pol I and it stalled when the CPD lesion reached the active site. Approximately 90% of the RNA molecules at time point 20 s (0.3 min) are 10-mer, while 10% remained 8-mer and a slight band appeared as 9-mer RNA, this latter remained even at long time (90 min incubation), which indicates the persistent presence of an intermediate (Figure 15C). Transcription extension beyond the 10-mer RNA is not observed even after 90 min of incubation. At short time reaction, most Pol II stalled at 10-mer RNA, with a slightly presence of 11-mer, but slowly extends to 11-mer and 12-mer after nucleotide incorporation opposite the CPD lesion (Figure 15C). Both gels were represented as intensity signal graphics, clearly showing that Pol I was blocked

at 10-mer RNA, while Pol II is blocked after incorporation of two nucleotides opposite the thymine dimer. Nucleotide addition opposite the upstream thymine (3'-T) is efficiently achieved (within 3 min of incubation), while addition opposite the 5'-T is much slower (90 min incubation). Besides, the incorporation rate constant (k_{obs}) for both enzymes were measured, and Pol II k_{obs} is significantly higher than that for Pol I (Figure 15D), supporting the result.

Because Pol I has intrinsic 3'-RNA cleavage activity due to the presence of the A12.2 subunit (Kuhn et al 2007), the final Pol I product transcription may be an equilibrium between polymerization and cleavage activities. To dissect the polymerization and intrinsic cleavage activities in Pol I, we performed transcription assays using as controls a purified mutant harboring a truncation in A12.2-Ct (Pol I ΔA12-Ct) that abolishes its strong cleavage activity. Because this strain was tagged at AC40 subunit, we first performed the AC40-TAP Pol I purification and later checked that the presence of the TAP-tag at different sites (A190 or AC40) did not influence the outcome of the *in vitro* transcription assays.

7.1. AC40-TAP Pol I isolation for mutational studies

Pol I purification from AC40-TAP strain was performed using the previous protocol. The isolation yielded 14 µg of total Pol I per liter of culture. Due to the fact that the tagged subunit is shared by Pol I and Pol III, both enzymes were isolated. Unlike the A190-TAP purification, the anion-exchange chromatogram showed two asymmetric peaks (Figure 16A). The first corresponded to Pol I elution, which showed a "shoulder" in the ascending part of the curve. SDS-PAGE analysis of eluted fractions demonstrated the "shoulder" eluted 14-subunit Pol I enzyme (Figure 16B, lane 1), while the peak with the highest absorbance maximum corresponded to 12-subunit Pol I A49/A34.5 (Figure 16B, lane 2). The peak with the second highest absorbance maximum corresponds to Pol III elution at 391 mM NaCl (Figure 16A). SDS-PAGE analysis of Pol III peak showed a high purity and homogeneity of the enzyme (Figure 16C). All fractions collected from the second peak contained 17 subunits at the expected migration region, which suggested no degradation.

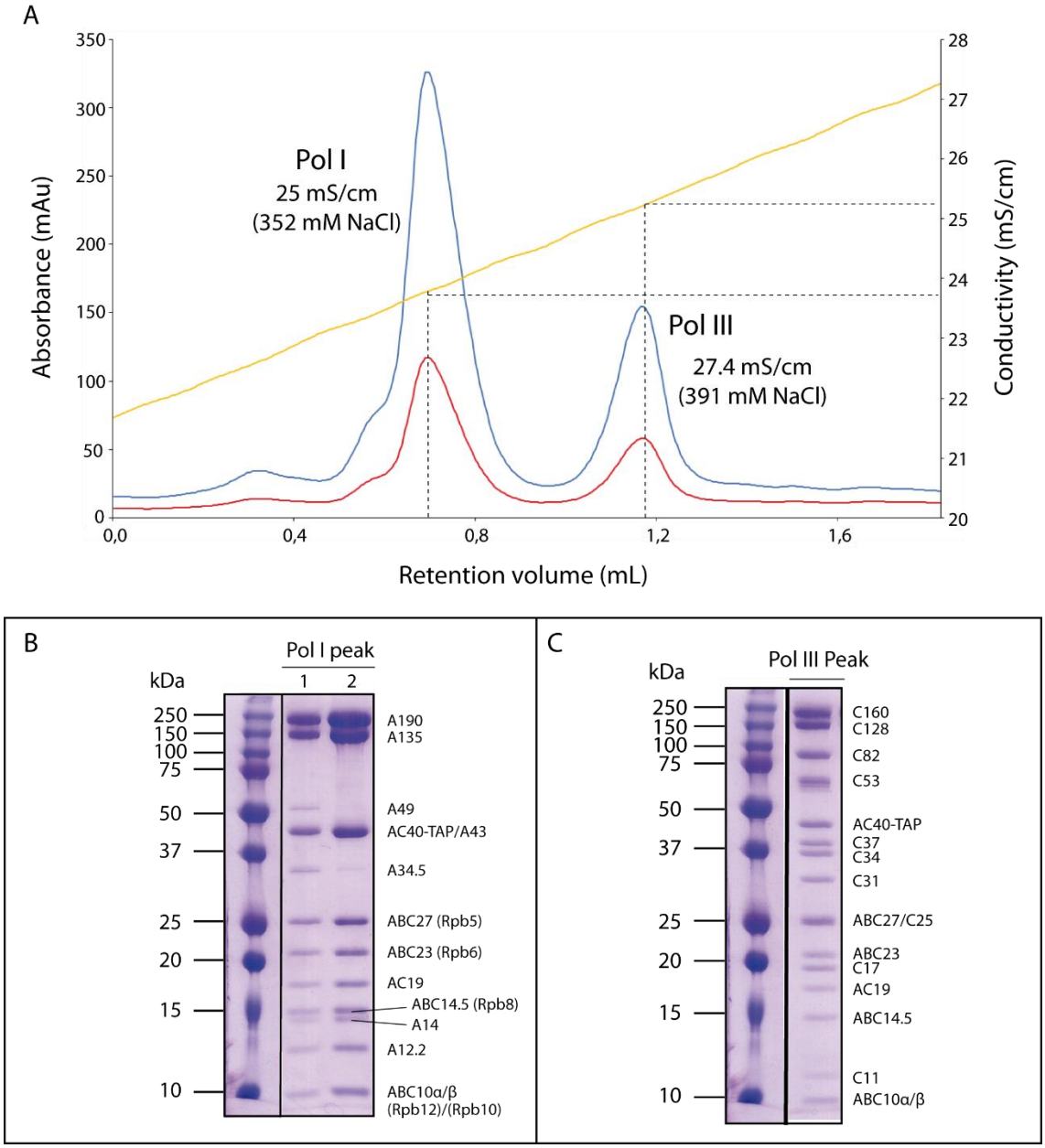

Figure 16. AC40-TAP Pol I and Pol III purification. (A) Anion-exchange chromatography of Pol I and Pol III. Abs_{280nm} and Abs_{260nm} curves are in blue and red, respectively, and conductivity in orange. (B) Coomassie-stained 15% SDS-PAGE gels of Pol I (left panel) and Pol III (right panel) peak.

Table 17. Pol I and Pol III variants detected in the AC40-TAP purification.

Nomenclature	Description
Pol I	14-subunit Pol I, complete
Pol I ΔA49/34.5	12-subunit Pol I lacking A49/34.5 heterodimer
Pol III	17-subunit Pol III, complete

Pol I enzyme used for *in vitro* transcriptional assays was tagged at AC40 subunit. To confirm that the change of tandem affinity position did not affect the biochemical results, *in vitro* transcription assays were performed using a lesion-free complete scaffold (Figure 17A). Both Pol I enzymes showed equal result in the urea gel (Figure 17B). Therefore, the position of the affinity tag does not affect the transcriptional assays.

Figure 17. Comparison of Pol I purified from different yeast strains. (A) Scheme of lesion-free DNA/RNA scaffold. Cyan, blue and red colors correspond to non-template DNA, template DNA and RNA, respectively. (B) *In vitro* transcription assays of AC40-TAP and A190-TAP Pol I on a lesion-free scaffold in the presence of 20 µM NTPs. Time points were 0 s, 20 s, 1, 3 and 10 min.

7.2. The Pol I RNA-cleavage activity is involved in CPD-induced stalling

To test if the RNA cleavage activity is involved in Pol I stalling, we first performed *in vitro* transcription assays using the previous transcription bubble containing the CPD lesion (Figure 15B). In contrast to AC40-TAP Pol I-WT, Pol I ΔA12.2-Ct mutant incorporated nucleotides opposite CPD lesions to generate 11-mer and 12-mer like Pol II, while Pol II+TFIIS led to a greater reduction of 11-mer and 12-mer products (Figure 18). The results also show that the *in vitro* RNA cleavage activity of AC40-TAP Pol I-WT is faster than Pol II+TFIIS (Figure 18).

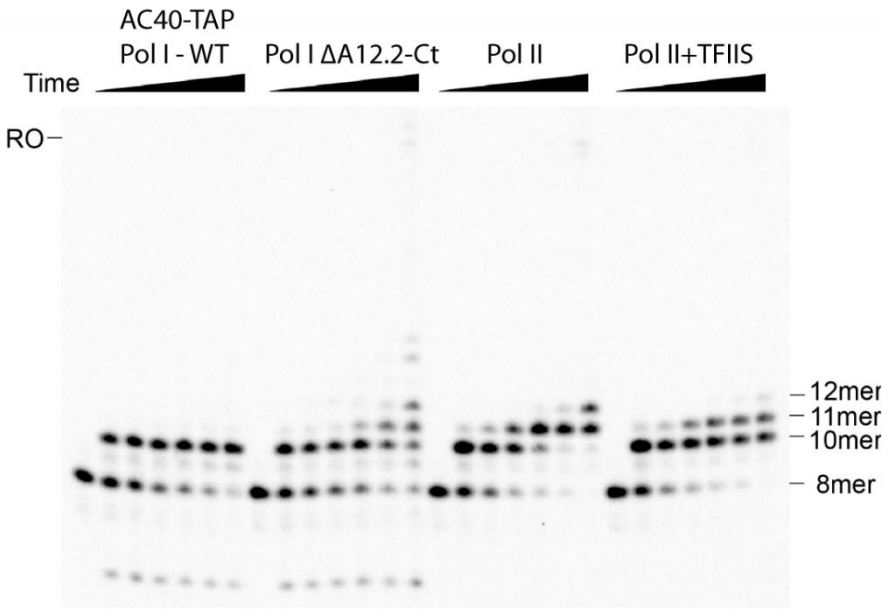

Figure 18. RNA cleavage activity effect in Pol I and Pol II to bypass the CPD lesion. Comparison of Pol I and Pol II processing of the CPD lesion in the presence (AC40-TAP Pol I-WT and Pol II+TFIIS) and absence (Pol I ΔA12.2-Ct and Pol II) of strong RNA cleavage activity, using the same conditions. RO indicates an expected position of the run-off product.

In vitro cleavage assays were performed to analyze the RNA cleavage activity effect when the RNA product mimics the insertion opposite of the 3′-T in the CPD lesion. AC40-TAP Pol I-WT, Pol I ΔA12.2-Ct, Pol II and Pol II+TFIIS were assembled with a scaffold containing an RNA 11-mer (Figure 19A), mimicking the insertion product opposite the 3'T in the CPD lesion. The reactions were performed by adding 5 mM $MgCl_2$ and the results showed that while the 11-mer is readily cleaved to short transcripts (9-mer and 8-mer at later points) in AC40-TAP Pol I-WT and Pol II+TFIIS systems, it remains stable in Pol I ΔA12.2-Ct complex, even in the presence of NTPs. Pol II did not show nucleotide incorporation at early times, but in the presence of 1 mM NTPs and after 30 min of incubation, the RNA 11-mer could be slowly extended showing a low intensity band at 12-mer, suggesting that a small percentage of Pol II is able to add a nucleotide opposite the 5'-T (Figure 19B), contrary to Pol I ΔA12.2-Ct. This revealed that Pol II bypasses CPD lesions more efficiently than Pol I ΔA12.2-Ct. In systems with strong intrinsic cleavage activity (Pol I-WT and Pol II+TFIIS), the appearance of the 10-mer product is found to concur with the disappearance of 11-mer in the presence of NTP (Figure 19B).

To test if the RNA product after cleavage remains active, *in vitro* transcription assays were performed by addition of 1 mM NTPs and RNA cleavage assay after 60 s incubation. The results revealed that the shortened transcripts produced by Pol I-WT and Pol

II+TFIIS in the presence of magnesium can be chased by adding nucleotides to the 10-mer in both systems. As expected, Pol I ΔA12.2-Ct and Pol II cannot cleave the 11-mer, and addition of 1 mM NTPs did not stimulate RNA extension in neither RNA polymerase (Figure 19C).

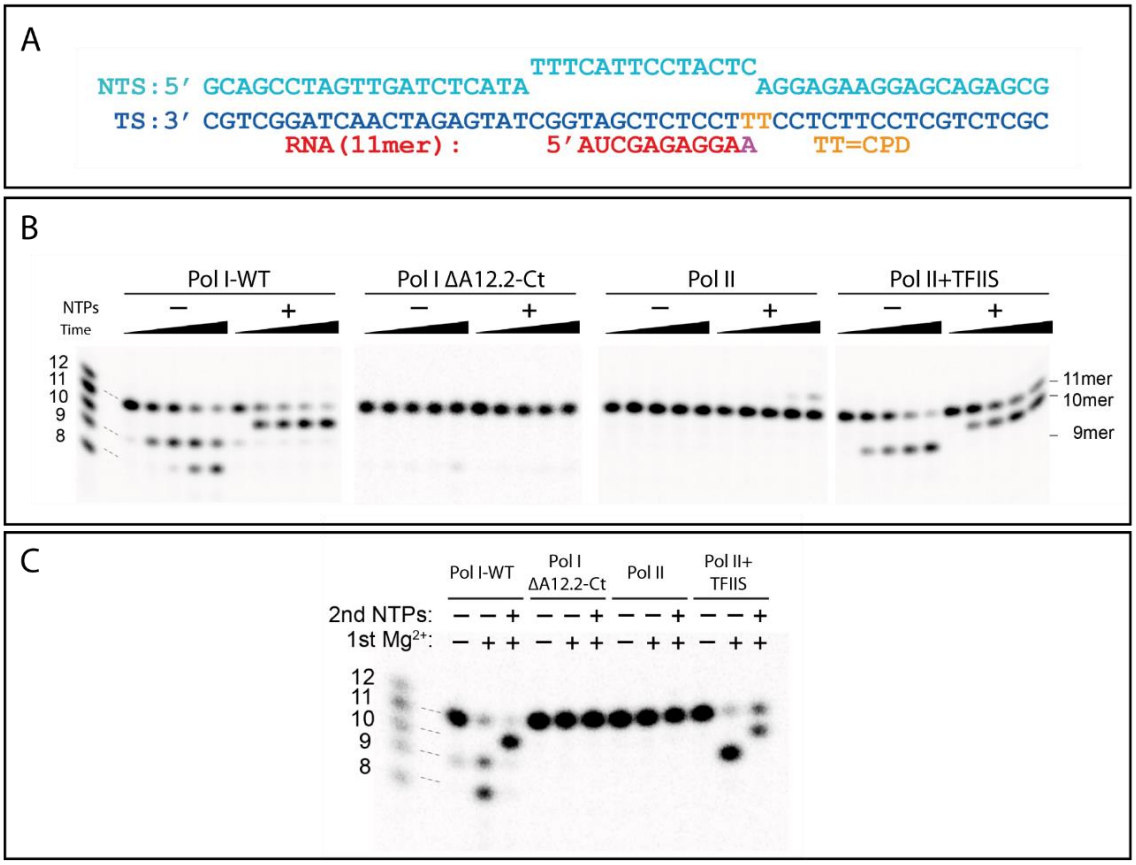

Figure 19. CPD lesion effect on cleavage assays. Extracted from (Sanz-Murillo et al 2018). (A) Schematic transcription bubble used for in vitro transcription assays. TS and NTS are shown in blue and cyan, respectively. Thymines of the CPD lesion are shown in orange. RNA is shown in red, and the adenine nucleotide opposite to 3′-T is shown in purple. (B) *In vitro* transcription assays using a scaffold with an 11-mer RNA that mimics transcript extension opposite the 3'-T in the CPD lesion. The cleavage reaction was started by adding 5 mM $MgCl_2$ at room temperature and stopped at 1, 3, 10 and 30 min. (C) The cleaved RNA product remains active and can be chased by adding 1 mM NTPs.

The results showed distinct behaviors between Pol I and Pol II upon CPD lesion encounter. While Pol II can insert additional nucleotides opposite the damage (11-mer and 12-mer), Pol I blocked right before the CPD lesion (10-mer). This difference is due to the combination of slower incorporation opposite the CPD lesion and faster intrinsic cleavage activity in Pol I.

8. Specific bridge helix residues play an important role in transcription processing of CPD lesions

Our structural results identified two residues of the BH that are fundamental for CPD lesion blockage. The alignment (Figures 14D and 20A) showed that the R1015 residue in A190 subunit is unique in Pol I. The main difference is that the side chain has a positive charge, while both Pol II and Pol III have a negatively charged glutamate and *E. coli* shows a polar amino acid residue. S1014 is also unique in Pol I, while Pol II, Pol III and *E. coli* RNAP have an alanine residue. To investigate the role of these amino acids in CPD lesion recognition, in collaboration with Dr. Georgiy A. Belogurov's laboratory, we used *E. coli* RNAP as a model system, due to the simplicity of the system.

To test the effect of N792 substitution, we used a full transcription bubble containing a CPD lesion on the template strand with an 8-mer RNA molecule (Figure 20B). bRNAP wild type (bRNAP-WT) as a control and mutants were incubated with the nucleic acids and the elongation triggered by adding 1 mM NTPs (Figure 20C). The urea gel showed that all three RNAPs incorporated a nucleotide opposite 3'-T, but not at the same rate. The N792D mutant showed an 11-mer RNA band at very early time, while N792R mutant presented an 11-mer RNA at 10 min time point. Compared all three rate constants (k_{obs}), the value for Pol II-like RNAP was 4.1 times larger than RNAP-WT. This latter k_{obs} value was 2.6 times larger than Pol I-like RNAP (Figure 20D). These results confirmed that Pol II-like bRNAP incorporate nucleotides opposite the CPD lesion faster than Pol I-like, meaning that the residue in that specific BH position plays an important role in transcriptional processivity.

To know if S1014 residue has effect in transcription processing over a CPD lesion in Pol I (equivalent to A791 in bRNAP), two mutant strains were used: one single mutant, A791S, and a double mutant, A791S/N792R, using as control bRNAP-WT and the previous studied mutant N792R (Pol II-like). *In vitro* transcription assays showed that all four strains could introduce a nucleotide opposite the 3'-T of the CPD lesion (11-mer) (Figure 20E), but the k_{obs} showed different values (Figure 20F). The bRNAP-N792R and bRNAP-A791S/N792R k_{obs} are smaller than bRNAP-A791S and bRNAP-WT values. The difference between bRNAP-WT and bRNAP-A791S k_{obs} values is not significant. The substitution of N792 (equivalent to R1015 in Pol I and E833 in Pol III) seems to play the principal role for the CPD processing.

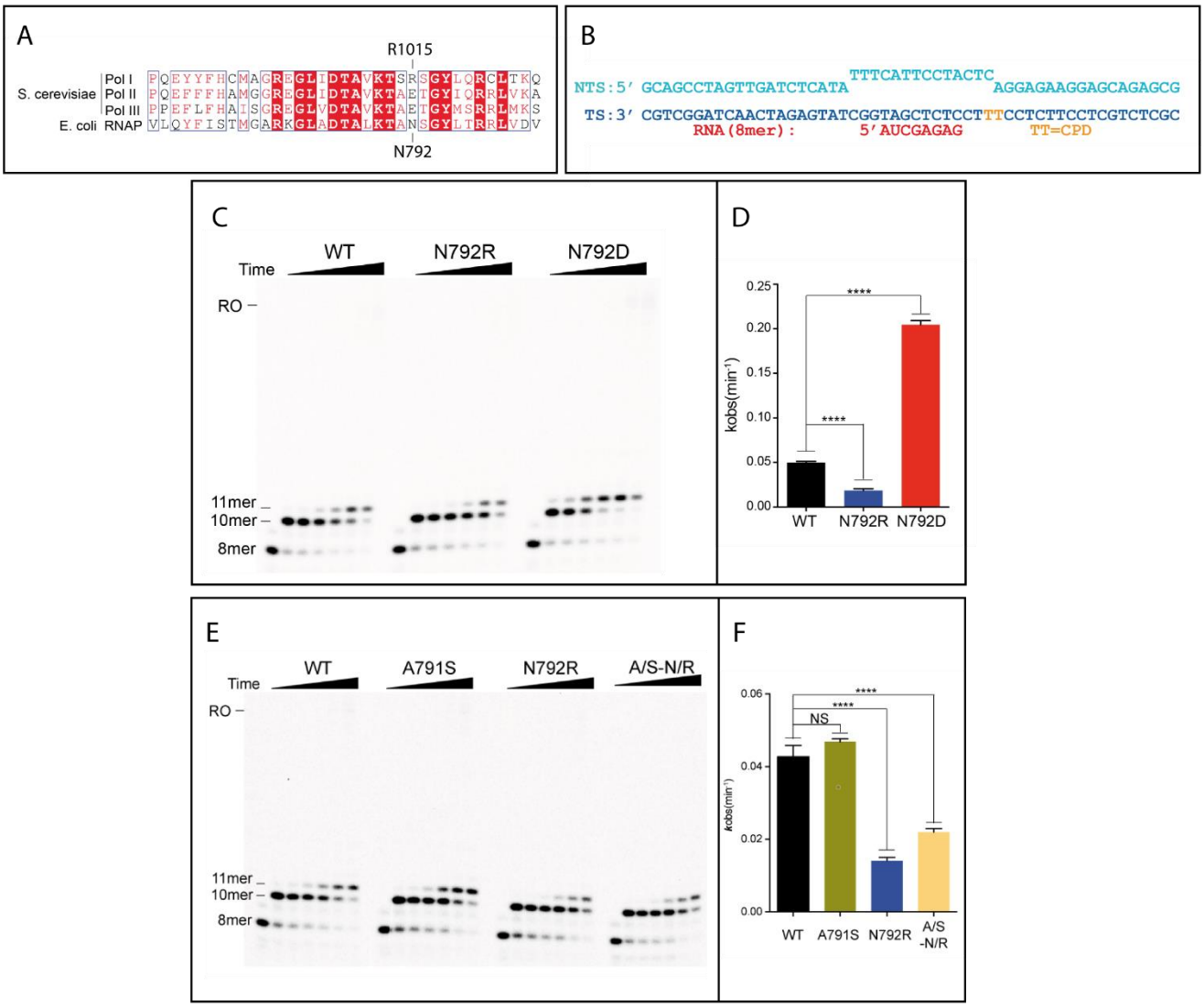

Figure 20. Mechanism of CPD lesion-induced Pol I stalling. Extracted from (Sanz-Murillo et al 2018). (A) Pol I-specific residues revealed by sequence alignment of the BH region for the three RNAPs in yeast and *E.coli* RNAP. Fully and partially conserved residues are boxed in red and white, respectively. A190-R1015E in Pol I and β'-N792 in bRNAP are labeled on the top and bottom, respectively. (B) DNA/RNA scaffold used for *in vitro* transcription assays. (C) Comparison between N792 mutants in the *E. coli* RNAP system using *in vitro* transcription assays. The NTP concentration is 1 mM, with time points as in Fig. 15C. (D) Quantification of the 10-mer RNA extension in (C). (E) *In vitro* transcription assay of WT and mutants of *E. coli* RNAP. The reaction was chased by adding 1 mM of NTPs and stopped at identical time points as in Fig. 15C. (F) Quantification of the 10-mer RNA extension in (E). All data are mean and SD (n=3). NS, not significant.

DISCUSSION

In this work, we provide biochemical and structural evidence for the mechanism of Pol I stalling by CPD lesions, which are commonly formed on DNA as a result of UV light irradiation. Pol I was endogenously expressed and purified from yeast with high quality, and the isolated enzyme was composed by two populations: Pol I and Pol I*. Image processing yielded a 3.6 Å resolution cryo-EM map of Pol I stalled at a CPD lesion. The comparison between undamaged Pol I-EC and Pol I-EC_{CPD} atomic models showed that the main structural changes occur at the Pol I-EC_{CPD} cleft, where a cation-π interaction is established between R1015 and the CPD lesion. Comparison with Pol II-EC at different translocation states defined that our structure displays an early intermediate translocation stage. Biochemical assays revealed that Pol I is persistently blocked when the CPD lesion reaches the active site, contrary to Pol II, which is able to incorporate nucleotides opposite of both thymines upon long incubation. A balance between a slow nucleotide incorporation and fast cleavage explains Pol I stalling. These combined results revealed that Pol I and Pol II present different mechanisms to deal with the CPD damage.

1. The optimization of Pol I isolation and Pol I-EC_{CPD} assembly

Previous studies identified the A43 C-terminal tail as an important element to form Pol I homodimers (Engel et al 2013, Fernández-Tornero et al 2013). Its dimerization role was confirmed *in vivo* by fluorescence microscopy studies (Torreira et al 2017) and phenotypic assays demonstrated that the truncated strain does not have its growth affected. This region is present but only partially conserved in most fungi and vertebrates (Beckouët et al 2011), and it is involved in the regulation of the activation of rRNA synthesis. As a monomer and in elongation, the A43 C-terminal tail is fully disordered (Engel et al 2016, Neyer et al 2016, Pilsl et al 2016, Tafur et al 2016, Torreira et al 2017), suggesting that this domain does not play a role modulating the elongation stage. Based on these facts, it was decided to use the A190-TAP Pol I lacking the A43 C-terminal tail in the structural studies, allowing the obtention of an active monomeric enzyme.

To purify the Pol I multi-subunit enzyme, a protocol previously established (Moreno-Morcillo et al 2014) was modified to increment the yield and reduce the time. The former protocol included three steps: low-affinity heparin, high-affinity IgG resin and polishing using anion-exchange. In the improved protocol, the first chromatography was removed.

Using IgG resin as the only affinity purification step, quite pure sample was obtained. Like in the former protocol, the anionic exchange permitted the removal of TEV protease and nucleic acids still present in the sample, further increasing protein purity.

Using this improved protocol, Pol I was reproducibly isolated with a minor presence of Pol I*. The continuous presence of the enzyme lacking A49/A34.5 could be due to both macromolecular complexes establishing an equilibrium *in vivo*. Lack of the heterodimer, or at least one of the subunits, alters the Pol I template specificity and its capacity to transcribe (Huet et al 1975). A recent report suggests that the A49/A34.5 dissociation *in vivo* is specifically involved in the regulation of Pol I transcription initiation and early elongation (Tafur et al 2019). *In vitro* assays showed that the presence of the heterodimer is strongly involved in promoter-dependent transcription, and the addition of the A49-Ct domain to Pol I* is enough to reestablish this transcriptional activity (Pilsl et al 2016). However, not only the absence of these subunits is involved in rDNA processing. Recent biochemical studies found that specific mutations at the interface formed by the two largest subunits and A12.2 could restore the normal rDNA transcription and increases Pol I density on rDNA when A49 is absent (Darrière et al 2019). Specifically, it was proposed that the A135-F301S or A12.2-S6L mutations stimulate DNA capture and facilitate the cleft closure in the absence of A49. The combination of Pol I mutated with the presence of A49 results in a super-active Pol I (Darrière et al 2019). Whether the A49/A34.5 heterodimer takes part in all stages of transcription, or its field of activity is delimited to initiation stage and early elongation, can be an interesting field of study.

To generate the Pol I-EC_{CPD} complex, our approach was to use electromobility shift assays to test nucleic acid-nucleic acid and protein-nucleic acid interactions. First, the binding of the two DNA strands and the RNA chain to form the transcription bubble was assessed. The ternary nucleic acid complex was efficiently assembled using equivalent molar ratios of the three nucleic acid strands. After that, Pol I was incubated with increasing amounts of the transcription bubble to choose the best molar ratio. The different electromobility between Pol I and Pol I-EC_{CPD} demonstrated the interaction was occurring and the best molar ratio was 1:1.2 (Pol I: transcription bubble). The minor excess of nucleic acids was intended to achieve full Pol I occupancy.

2. **Complete atomic model of CPD-stalled Pol I**

Preserving macromolecular complexes in a near native state allows better understanding of their biological function. Therefore, cryo-EM was selected to perform the structural determination of Pol I-EC_{CPD} and Pol I-EC_{CPD} with fully-ordered A49. The resulting cryo-EM density maps were obtained from the same dataset. Pol I itself has demonstrated to be a source of heterogeneity, as even the samples most enriched in Pol I contained Pol I*. Data processing strategies, such as focused 3D classification, are widely used to classify related, but not identical protein complexes (Scheres et al 2007, van Heel et al 2009). Thus, a heterogeneous sample represents an opportunity rather than a disadvantage to analyze different components and/or conformations (Nogales and Scheres 2015). For other structural techniques such as X-ray crystallography, the heterogeneity would have probably been an obstacle, while it could be resolved by 2D and 3D classification implemented in single-particle cryo-EM.

Four regions were selected for 3D focused classification. The selection was based on the interest of the specific area, flexibility or dissociation of the region. The first 3D focused classification was based on the nucleic acid scaffold. This classification revealed that a small percentage of Pol I particles showed poor density for the DNA/RNA scaffold. This indicates that not all the Pol I molecules contained the bubble, although an excess of nucleic acids was used. Perhaps incubation conditions could be further optimized. This cryo-EM map attained the highest resolution and allowed us to build the atomic model of the Pol I core, and, more importantly, the Pol I cleft and the nucleic acids. In the map, the density for A49/A34.5 heterodimer was not observed. This may be because the cryo-EM map is constituted by a mix of Pol I-EC_{CPD} and Pol I*-EC_{CPD} particles.

The A49-Nt/A34.5 anchors the N-terminal domain of A12.2 subunit and contains a dimerization module conserved in Pol II initiation factor TFIIF and Pol III C37/C53 (Fernández-Tornero et al 2010, Geiger et al 2010, Vannini et al 2012). The 3D focused classification based on the A49-Nt/A34.5 heterodimer accounted for 50% of the input particles, even when a 14-subunit Pol I enriched fraction was used to prepare the sample (in the native gels, approximately 80% of the sample was 14-subunit Pol I, while 20% corresponded to Pol I*). The decrease in 14-subunit Pol I percentage could be explained taking into account that the heterodimer is easily dissociated (Huet et al 1975). Dissociation of A49/A34.5 has been related to the binding of elongation factors to Pol I

such as Spt4/5 and Paf1C (Tafur et al 2019). Both factors are also involved in the transition from initiation to elongation in Pol II, where the initiation factors TFIIE and TFIIF block Spt4/5 and Paf1C binding (Xu Y et al 2017). An additional hypothesis is that the heterodimer might have been partially disengaged during sample vitrification.

Even after a 3D classification focused on A49-Nt/A34.5 heterodimer, the cryo-EM map did not show density for the A49-Ct domain. This latter region is involved in processivity both *in vivo* and *in vitro* (Beckouet et al 2008), DNA-binding, initiation and elongation (Geiger et al 2010). The 3D-focused classification based on A49-Ct domain yielded a complete map, showing the A49-Nt/A34.5 heterodimer, the stalk and the transcription bubble, including the upstream DNA. However, the number of particles contained in this map was the lowest, approximately 20% of the 3D classification input, and, in consequence, the resolution was 4.6 Å.

Another region with poor density in the cryo-EM map at 3.6 Å is the stalk. This area is known to be flexible in monomeric Pol I, unlike the enzyme bound to Rrn3 and Pol I in the dimeric state (Engel et al 2013, Fernández-Tornero et al 2013, Engel et al 2016, Pilsl et al 2016, Torreira et al 2017). The stalk is involved in transcription initiation, mainly through the interaction with Rrn3. To obtain information on this area, a 3D focused classification was carried out, which presented 21% of the total 2D selected particles. This map defined the A14/A43 dimer, which is in a similar conformation as Pol I-EC stalk. A14/A43 subcomplex is anchored to Pol I and it co-operates with A49-Nt/A34.5 heterodimer through interactions of A49-Ct and the hydrophilic domains of A43 (Beckoüet et al 2011).

The complete elongation complex at 4.6 Å was modelled, excluding the A43 C-tail, absent from our sample, as well as the A12.2-Ct domain and the DNA-mimicking loop, which did not show cryo-EM density, likely due to flexibility. The C-terminus of A12.2, responsible for intrinsic Pol I RNA cleavage activity, is absent in the four maps. The same occurs in two monomeric Pol I structures and two Pol I:Rrn3 reconstructions (Pilsl et al 2016, Torreira et al 2017). Binding of the nucleic acids inside the cleft maintains the closed conformation and expulses the A12.2-Ct from the pore (Neyer et al 2016, Tafur et al 2016). In the reconstruction of Pol I-OC, density for the A12.2-Ct is present, as well as in one monomeric Pol I and one Pol I:Rrn3 map (Engel et al 2016, Neyer et al 2016, Tafur te al 2016); nonetheless, the catalytic hairpin in A12.2-Ct appears to be mobile. Most

likely, A12.2-Ct might displace (or at least its catalytic hairpin) and is excluded from the active site when the cleft is closed, allowing Pol I to establish extensive contacts with the RNA chain (Tafur et al 2016). Thus, the fact that A12.2-Ct remains flexible in the four maps of this work agrees with published results and it makes sense that a domain specialized in RNA cleavage is only stabilized under specific conditions at concrete time windows. Furthermore, recent work has revealed that A12.2-Ct domain can alternate between TFIIS and Rpb9 positions in Pol II, depending on the presence of the A49/A34.5 heterodimer (Tafur et al 2019). In the 12-subunit Pol I structure, the A12.2-Ct occupies the space left empty by A34.5, interacting with the A135 subunit (Tafur et al 2019).

The three maps focused on the DNA/RNA scaffold, A49-Nt/A34.5 and the stalk were used to build the first atomic model, in turn used as reference to construct the second model, which contains the A49-Ct and the DNA upstream, only using the cryo-EM map focused on both elements. Possibly, the absence of the A49-Nt/A34.5, the stalk, the A49-Ct and the upstream DNA areas in the high-resolution cryo-EM map might be resolved using crosslinking. Two previous reports published damage-free Pol I-ECs (Neyer et al 2016, Tafur et al 2016), one of them using glutaraldehyde crosslinking to fix the elongation complex (Neyer et al 2016). However, regardless of the chemical fixation, both Pol I-EC_S, when compared to Pol I-EC_{CPD} and Pol I-EC_{CPD} with fully-ordered A49, are essentially identical, only showing minor differences in the cleft, which are critical to understand CPD-induced Pol I stalling.

3. Pol I-EC_{CPD} adopts an early intermediate translocation state

The resolution achieved with the cryo-EM maps allowed model building of the enzyme and DNA/RNA scaffold, revealing the accurate position of the CPD lesion. To understand the translocation stage of Pol I-EC_{CPD}, comparative analysis of Pol II at different stages of elongation was performed (Gnatt et al 2001, Kettenberger et al 2004). These studies confirmed that, once the CPD lesion has reached *i*+1/*i*+2 position, Pol I is blocked at an early intermediate configuration in the translocation cycle, which is characterized by unexpected structural features. The captured intermediate state of Pol I-EC_{CPD} is attributed to the fact that intrastrand crosslinks at the CPD lesion twist the DNA strand and restrain the flexibility of the phosphodiester backbone of the TS, which impedes the regular translocation cycle. The CPD lesion accommodates above the BH, in a similar

position for *i*+1/*i*+2 to the pre-translocated state, and the 5'-T remains partially base-paired with its NTS counterpart, unlike Pol II at pre-translocated stage. The position of the *i* base pair in the Pol I-EC_{CPD} upstream is tilted toward the BH and occupies a non-canonical position, unlike in pre- and post-translocated states. The RNA 3'-end, which pairs the DNA base at the position *i*, occupies an intermediate position and partially impedes the NTP entry. This is confirmed by superposition of Pol I-EC_{CPD} with Pol II at the pre-insertion state (Wang et al 2006), as the incoming NTP remains 1.9 Å closer to the 3'end RNA chain in Pol I-EC_{CPD} than in the RNA canonical stage.

Interestingly, a similar configuration of DNA/RNA hybrid and downstream DNA was observed for Pol II ITCs with RNAs of four, five, six and seven nucleotides, which mimic the pathway from the Pol II-OC to Pol II-EC. Unlike Pol I-EC_{CPD}, these structures present a tilted DNA/RNA hybrid, where the RNA product always occupies a post-translocated state, leaving the NTP entry site free. Binding of the NTP substrate to the tilted complexes induces the standard conformation of the hybrid and stabilizes the fully post-translocated state (Cheung et al 2011).

As Pol I-EC_{CPD}, Pol II-EC inhibited by α-amanitin presents an intermediate translocation stage (Brueckner and Cramer 2008). Both structures display a configuration between the pre- and post-translocation state at the downstream DNA TS. Furthermore, the BH in both cases presents a shifted conformation towards to upstream DNA, which partially occupies the active site and the TL exhibits a wedged conformation, stabilizing the kinked BH. However, these structures also show two different features inside the enzymes. The inhibited Pol II structure shows a DNA/RNA hybrid at pre-translocated stage while Pol I presents an early intermediate state. The *i*+1 base at the blocked Pol II lies also above the BH; however, it occupies a different position, more advanced regarding the 3'-T of the CPD lesion. Despite having in both cases a blocked stage, these configurations are due to different reasons. Pol I-EC_{CPD} is stalled by restrictions on the movement of the CPD lesion, which cannot enter into the active site. Meanwhile, the Pol II-EC is blocked by a network of interactions between the α-amanitin and the surrounding BH and TL amino acids.

Several crystallographic structures of Pol II-EC_{CPD} containing the CPD lesion at different positions were described (Brueckner et al 2007, Walmacq et al 2012) and their configurations are slightly different in comparison with the Pol I-EC_{CPD}. In the case of

blocked Pol I, the CPD lesion remains slightly behind towards the downstream DNA. CPD-stalled Pol II at position *i*+1/*i*+2 revealed that the lesion is disengaged from downstream DNA, but the 3'-T is not able to reach the canonical *i*+1 position, leaving the regular site for the templating DNA unoccupied and an enlarged NTP entry site. *In vitro* transcription assays show that Pol II preferentially incorporates an adenine opposite the 3'-T. The space to accommodate the incoming NTP and the lack of DNA template in the correct position provides the conditions for non-templated addition, allowing the entry of a purine residue (Walmacq et al 2012). Testing the extension with a matched 5'-T adenine base pair, the RNA was elongated to the run-off transcript; however, the incorporation of an uridine nucleotide at this position maintains Pol II stalled (Brueckner et al 2007), consistent with the A-rule (Strauss 1991, Taylor 2002). The post-translocated state after the first adenine incorporation seems to be the mechanistic key between error-induced stalling or error-free bypass (Walmacq et al 2012).

Recent molecular dynamics studies identified two translocation intermediates between the pre- and post-translocated states in Pol II (Silva et al 2014). In the first intermediate, the DNA/RNA hybrid has been translocated and the base at *i*+1 is positioned above the bridge helix. In the second intermediate, the template base at *i*+1 has passed the bridge helix but not yet occupied the canonical templating position. The Pol I-EC_{CPD} presents an early intermediate that is only slightly advanced from the pre-translocated state, in which the base pair at the *i* position presents a unique midway configuration between the pre- and post-translocation states and the CPD lesion stalls earlier than the first intermediate identified by molecular dynamics. Altogether, the comparison with the Pol II-EC_{CPD} structures containing the CPD lesions at different positions, Pol II inhibited by α-amanitin, Pol II ITCs and taking into account the four metastable stages described in Pol II-EC translocation cycle by dynamic simulation studies (Silva et al 2014) allowed to conclude that Pol I-EC_{CPD} atomic model showed a novel early intermediate translocation state.

4. Structural rearrangements inside the DNA-binding cleft of Pol I-EC$_{CPD}$ explain the blocking

Overall, Pol I shows a similar configuration in Pol I-EC$_{CPD}$ and lesion-free Pol I-ECs (Neyer et al 2016, Tafur et al 2016). Nevertheless, the most significant structural changes occur inside the cleft and explain Pol I blockage by the CPD lesion. In Pol I-EC$_{CPD}$, the enzyme BH is kinked in its central region and the TL is wedged, whereas in the two available post-translocated Pol I-ECs the BH is fully folded and the TL remains flexible (Neyer et al 2016, Tafur et al 2016).

Specific rearrangements also occur in the amino acids that surround the CPD damage. The Pol I-specific R1015 residue establishes a cation-π interaction with the 3'-T of the CPD lesion. This contact is not possible in Pol I-EC structures because they present a post-translocated stage, and this arginine residue lies at a distance of 8.2 Å from the *i*+2 DNA TS position (Neyer et al 2016, Tafur et al 2016). S1014 at CPD-stalled Pol I lies at hydrogen bond distance with the phosphate DNA backbone next to the lesion, while in damage-free Pol I-ECs the residue is oriented towards the opposite side. The movement of the DNA TS could also be important for the S1014 contact, as the 3'-T phosphate approaches S1014 due to the sterically restricted movement of the CPD and the equivalent DNA backbone position in Pol-ECs is oriented toward the opposite side of the BH.

Multi-alignment of protein sequences of several prokaryotic and eukaryotic organisms exhibits interesting features. The central regions of the BHs show fully and partially-conserved residues; nevertheless, there are three consecutive highly variable amino acids that correspond to the positions S1014, R1015 and S1016 in Pol I A190 subunit. The area occupied by these three residues plays an important role in the translocation cycle, as around it the main variations in terms of position occur during the continuous bending motion of the BH. The positively-charged R1015 is fully conserved in Pol I, whereas Pol II and Pol III present a negatively-charged glutamate in this position. In archaea, the residue located in an equivalent position is a glutamine, while in bRNAP different residues are present: a polar asparagine in *E. coli* or a negatively-charged aspartate in *T. aquaticus*. S1014 is also strictly conserved in Pol I, while Pol II and Pol III display a hydrophobic alanine in almost all species with the exception of *D.discoideum* and *A.thaliana*, organisms that show a polar serine. S1016 is also completely conserved in Pol I, while Pol II and Pol III present a polar threonine. Overall, these three amino acids

are fully conserved between Pol I BHs, whereas Pol II and Pol III BHs amino acids are conserved between them.

In yeast, R1021 is conserved in the three RNAPs, as well as from bacteria to humans (Cramer et al 2001). The equivalent position in Pol II-EC_{CPD} containing the CPD lesion at *i*+1/*i*+2 shows its side chain turned towards the largest subunit and not at hydrogen bond distance of the 5'-T phosphate, unlike Pol I-EC_{CPD}, where the residue lies closer to the 3'-T phosphate. In Pol I-EC, this residue remains at the same position as Pol I-EC_{CPD}. Nevertheless, the distance is larger because the TS adopts its canonical position. Another strictly conserved residue in the BH is T1013. Molecular dynamic simulations and single-mutagenesis assays performed in Pol II revealed that T831, equivalent to T1013 in Pol I, acts as a checkpoint to examine the stabilities of the base pair in the active site. This residue detects the 3'-terminal base paring of DNA/RNA hybrid, promoting the RNA 3'-end nucleotide into a frayed state if the base pair is mismatched (Da et al 2016).

Inside the cleft, switch loop 2 plays a role in the transition to a transcribing complex, reordering when the nucleic acids enter the cleft and contacting the DNA/RNA hybrid in the active center (Cramer et al 2001, Gnatt et al 2001). The insertion of a glycine residue between the two conserved lysine residues in Pol II and Pol III suggests an extra mobility that Pol I switch loop 2 lacks. In Pol I-EC_{CPD}, switch loop 2 alters its configuration to approach the TS. K463 lies at hydrogen bond distance of the nucleotide upstream of the CPD lesion, while K462 remains at a distance which does not allow interaction. On the contrary, in Pol I-ECs this loop presents alternative arrangements, but never interacting with the DNA (Neyer et al 2016, Tafur et al 2016). R468, also belonging to switch loop 2, is conserved in the three RNAPs. This residue lies at hydrogen bond distance with the 3'T phosphate in the Pol I-EC_{CPD}, while in Pol II-EC_{CPD} R337 (equivalent to R468) does not interact with the CPD. Instead, R337 contacts the phosphate in the 5'-T, due to the different position of the CPD lesion. Undamaged Pol I-EC presents the R468 oriented slightly towards the enzyme, hampering the interaction with the canonical DNA TS. In summary, Pol I-EC_{CPD} blocked structure is secured by a unique set of specific interactions between the DNA TS and conserved Pol I regions surrounding the CPD lesion.

5. Specific mutations modify the transcriptional processing of CPD lesions

One of the principal defining features of RNAPs is the BH, which is implicated in translocation and fidelity of NTP incorporation (Kireeva et al 2012). The central role of Pol I R1015 in CPD lesion recognition was studied by single mutational analysis and *in vitro* transcription assays. The mutations were introduced in bRNAP, given that, in comparison to yeast, bacteria are simple organisms with straightforward genetic manipulation. Besides, this organism is used widely as a model system, and the presence of a polar residue in the required position allowed to distinguish the effect of the charge. Pol I-like bRNAP mutant (N792R) exhibited a k_{obs} 2.6 times less respect to bRNAP-WT, while the Pol II-like mutant (N792D) increased the transcriptional rate 4.1 times. These results showed that the residue located in this position plays an important role in the processing of the CPD lesion, allowing faster incorporation of the nucleotide opposite the 3'-T in Pol II and slower incorporation in Pol I. This correlates with the fact that the 3'-T in the CPD lesion and R1015 in Pol I-EC_{CPD} and contrary to Pol I-EC, establish a cation-π interaction. This type of interaction, among others, plays an important role in protein interactions (Dougherty 1996) and this kind of electrostatic attraction is achieved when an cationic side chain (K or R) is near an aromatic side chain (F, Y or W) (Gallivan and Dougherty 1999). In Pol I-EC_{CPD}, the cation-π interaction is produced using as aromatic electron donator the nitrogenous base of the 3'-T. Although this interaction explains the slow bypass of the DNA damage in Pol I, it did not fully explain the *in vitro* transcription results where Pol I is persistently blocked.

bRNAP mutations were also carried out to assess the role of Pol I S1014 in CPD lesion transcription. *In vitro* transcription assays revealed that A791S mutant has a distinct behavior in comparison to the N792R. In fact, this mutation increases 1.1 times the k_{obs}. This effect could suggest that the serine residue contributes in a minimal way to decelerate the CPD bypass. The double mutant A791S/N792R reduced the k_{obs} by half respect to bRNAP-WT, but this catalytic rate constant is still 1.3 times higher than the N792R single mutant constant. These assays confirmed the principal role of Pol I R1015 in CPD processing.

6. The Pol I elongation rate depends on the balance between nucleotide addition and intrinsic cleavage

In fast growing cells with a high demand of protein, Pol I needs to perform a fast elongation and has a strong backtracking and cleavage activity to prevent transcriptional arrest (Darzacq et al 2007, Lisica et al 2016). Pol I contains the A12.2-Ct domain, which is responsible for the intrinsic RNA cleavage, while Pol II needs to recruit an independent transcription factor, TFIIS (Kuhn et al 2007). *In vitro* transcription assays show that, unlike Pol II, Pol I is not able to bypass the CPD lesion even after long incubation at a high NTP concentration. Nevertheless, Pol I with A12.2-Ct truncated can bypass the CPD lesion, incorporating RNA nucleotides opposite both thymines, but is less prone than Pol II (without TFIIS) to incorporate nucleotides. Pol II incubated with TFIIS, which is homologous to A12.2-Ct, did not incorporate nucleotides opposite the 5'-T. Thus, *in vitro* RNA incorporation assays unveiled that the RNA cleavage activity plays an important role in CPD bypass.

The combination of both nucleotide incorporation and RNA cleavage activities in Pol I were dissected by *in vitro* cleavage assays, using an 11-mer RNA that mimics transcript extension opposite the 3'-T in the CPD lesion. Pol I-WT and Pol II+TFIIS were able to cleave the RNA transcript readily and their cleavage rates are similar. However, in the absence of RNA cleavage activity, truncated Pol I and Pol II do not cleave the RNA. After cleavage, Pol I and Pol II+TFIIS are able to incorporate nucleotides to the RNA product; nonetheless, Pol I does not reach again the 11-mer RNA length. Altogether, in spite of both displaying RNA cleavage activity, Pol I pyrophosphorolysis is stronger, as Pol I profits of A12.2-Ct being continually bound to the enzyme. Pol II has a weak intrinsic RNA cleavage with subunit Rpb9 and requires the TFIIS to improve its capacity (Lisica et al 2016). The distinct behavior of Pol I in relation to Pol II is due to the combination of slower nucleotide incorporation opposite the CPD lesion and much faster intrinsic RNA cleavage activity. These distinct mechanisms can be explained due to Pol II transcription system not requiring the constant presence of a cleavage factor. This enzyme can be more error-prone without provoking a lethal consequence in the cell. Nevertheless, Pol I system is only designed to transcribe ribosomal RNA genes in order to originate the backbone of the ribosome, therefore it must execute rapidly without introducing errors.

7. Distinct mechanistic models for Pol I and Pol II processing of CPD lesions

Knowing the specific features that can explain the Pol I blocking, the behavior of Pol I vs Pol II stalling at a CPD lesion was analyzed using *in vitro* transcription assays. In a previous work, a time course experiment using Pol II and a DNA/RNA scaffold with a CPD showed that RNA incorporation opposite 3'-T is slower than nucleotide addition opposite to a canonical DNA TS, and the second RNA incorporation opposite 5'-T is even slower. The combination of distortions produced by the pyrimidine dimer and misincorporation opposite the 5'-T result in Pol II blocking (Brueckner et al 2007). Unlike Pol II, Pol I can only extend the RNA molecule until the 3'-T reaches the active site.

The combination of the structural and biochemical data present in this Ph.D. thesis provides insights into the mechanism of Pol I stalling by CPD lesions. Pol I and Pol II showed mechanisms that significantly differ (Figure 21). Slow thymine dimer recognition, responsibility of R1015 in A190, in addition to a fast-intrinsic RNA cleavage carried out by A12.2-Ct, prevents the incorporation of an adenine, unlike in Pol II.

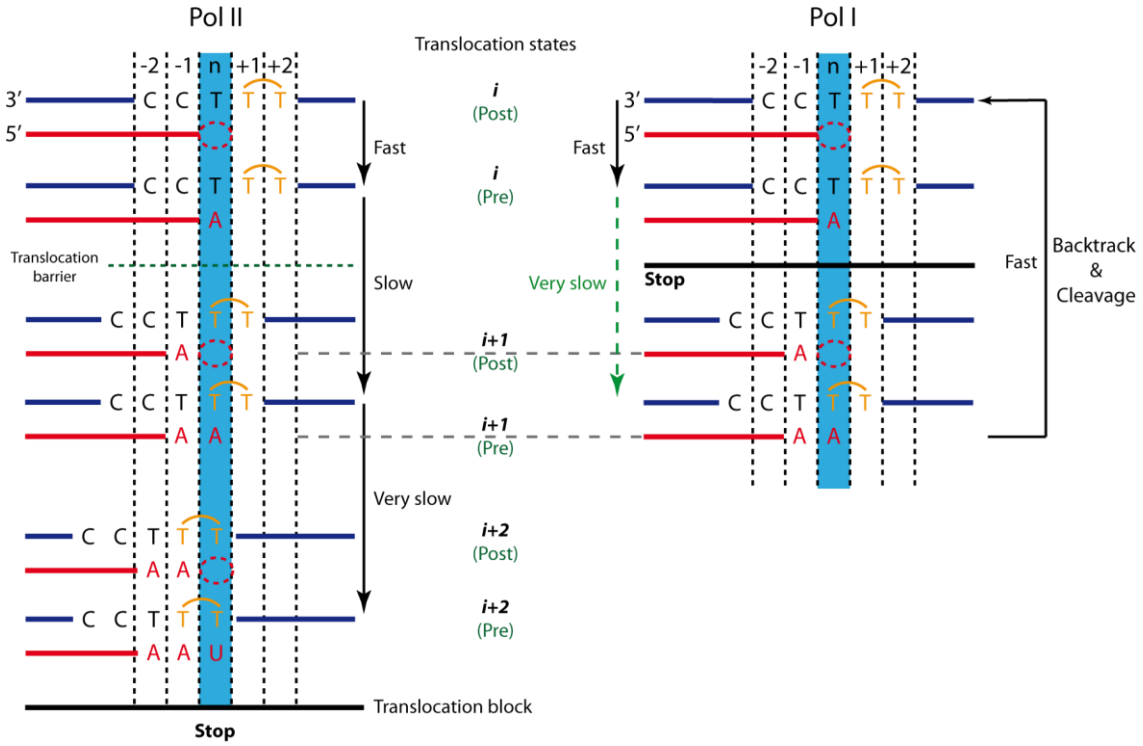

Figure 21. Mechanisms to deal with CPD lesions in Pol I and Pol II. (Extracted from Sanz-Murillo et al 2018). Scheme of distinct mechanisms of CPD lesion recognition by transcribing Pol II (Left) and Pol I (Right). The translocation barrier and the translocation block are indicated with a dashed line and a solid horizontal line, respectively.

Pol I cleaves this first added nucleotide and maintains an addition-cleavage equilibrium, resulting in an impediment to bypass the CPD. Due to the absence of the arginine residue, Pol II can incorporate an adenine opposite the 3'-T faster than Pol I lacking A12.2-Ct. The weak intrinsic RNA cleavage function of Rpb9 allows the untemplated addition of uracil opposite the 5'-T, leaving the enzyme blocked (Brueckner 2007, Walmacq 2012).

Over the last decade, structural studies of Pol II with several DNA lesions, such as a DNA alkylation (Xu et al 2017), an oxidative DNA lesion 8,5′-cyclo-2'-deoxyadenosine (CydA) (Walmacq et al 2015) or an abasic lesion (Wang et al 2018), were carried out to understand how this enzyme recognizes and processes the DNA damage (Wang et al 2018). The abasic lesion exhibits the same behavior as the CPD lesion under physiological conditions, following the A-rule (Wang et al 2018). Pol II-EC-CdyA preferentially performs the addition of an adenine opposite lesion, but it also can incorporate an adenine residue opposite the TS next to the lesion. This can generate mutant RNA transcripts, except at DNA sequences containing a thymine 5′ of the lesion (Walmacq et al 2015). Alkylation lesions revealed different transcriptional responses of Pol II, depending on the position of the ethyl group of thymine (Xu et al 2017). All these studies were performed using Pol II. Nonetheless, the mechanistic and biochemical insights of Pol I whit this kind of lesions is still not understood.

8. TC-NER in Pol II and Pol I

TC-NER in Pol II has been extensively studied, due to the fundamental role of this enzyme for the cell. The knowledge of this mechanism has benefited from the cooperation of several groups specialized in DNA repair and transcription. Previous reports have described how the repair factors act and are recruited when Pol II is arrested by bulky lesions. CSB has been considered as the master coordinator of TC-NER in mammalian cells, being among the first proteins to be recruited to the arrested Pol II (Wang et al 2018), avoiding the TFIIS-induced backtracking. The Pol II-CSB interaction activates the recruitment of TFIIH to catalyze the transcription bubble opening. From that point, the reparation steps are triggered sequentially. Rad26, the CSB ortholog in *S. cerevisiae*, was recently structural and biochemically characterized. Those observations revealed that the interaction between Rad26 and Pol II can help to distinguish different Pol II arrest scenarios, and the persistent arrest at a voluminous lesion, such as a CPD lesion, triggers the initiation of TC-NER (Xu et al 2017). The Pol II-Rad26 complex structure was solved

using single-particle cryo-EM (Xu et al 2017), showing that the repair factor is bound to upstream DNA and the interface between them is conserved from yeast to humans. Further biochemical and structural insights will contribute to understand better the molecular mechanism of TC-NER. Likely, this knowledge will help develop new chemotherapeutic drugs targeting TC-NER sub-pathway (Wang et al 2018).

The TC-NER mechanism in Pol I has not been described as in deeply as in Pol II. A previous report demonstrated that strand-specific repair occurs in transcriptionally active rRNA genes in yeast and it is not found in the inactive rDNA copies (Conconi et al 2002). Therefore, this suggests that TC-NER in Pol I exists as a repair mechanism in yeast (Charton et al 2015) and this process was demonstrated to be independent of Rad26, (Verhage et al 1996). However, a recent report has shown by analytical methods that TC-NER in rDNA is not involved in repairing damages caused by UV or by cisplatin in human and mouse cell lines (Yang et al 2019); nevertheless, the molecular basis of TC-NER mediated by Pol I in yeast remains unclear.

Over the last two decades, the special nucleolar environment of the rDNA and the involvement of Pol I in TC-NER has been identified as an interesting field to study. The structural and biochemical results present in this Ph.D. thesis pave the way to unravel the molecular mechanism underlying persistent enzyme stalling at CPD lesions on rDNA and provide important mechanistic insight into Pol I lesion recognition. In particular, the structure of Pol I-EC_{CPD} exhibits molecular interactions between Pol I and the bulky lesion and reveal novel mechanistic insight into the role of R1015 recognition. Future researches on how Pol I is involved in TC-NER and the sequential steps of the repair factors to resolve the damage would deep in our understanding of this essential cell process and its relation with the repair pathways. Importantly, several mechanistic questions deserve future investigation: which specific signals trigger TC-NER in rDNA? How are the repair factors recruited to the Pol I-EC_{CPD} for lesion verification? How are the elongation factors displaced when Pol I is blocked? At which stage is Pol I moved away from the CPD lesion? What is the fate of Pol I during lesion recognition? When and how are the repair proteins released from DNA lesions? What is the function of each of them? Finding answers to these questions would contribute to better understand the molecular mechanism of CPD lesion repair in rDNA. The work presented here represents the first mechanistic insight to resolve this puzzle.

CONCLUSIONS/CONCLUSIONES

CONCLUSIONS

1. Pol I persistently stalls when CPD reaches the active site. The cryo-EM structure of Pol I-EC_{CPD} presents an intermediate configuration between the pre- and post-translocated states with the CPD lesion lying above the bridge helix, which is slightly kinked at its central region.
2. Several conformational rearrangements create a network of interactions that stabilize enzyme stalling. R1015 at the central region of the BH establishes a cation-π interaction with the 3'-thymine of the CPD lesion and two basic residues in switch loop 2 alter their configuration to approach the TS.
3. Single mutational analysis using bRNAP as a model system and *in vitro* transcription assays reveal that R1015 plays an important role in the processing of CPD lesions.
4. Pol I stalls at CPD lesions due to the combination of slow nucleotide incorporation opposite the CPD lesion and fast intrinsic RNA cleavage activity of its A12.2-Ct domain.
5. The mechanism of Pol I stalling by CPD lesions significantly differs from that observed in Pol II.

CONCLUSIONES

1. La Pol I se bloquea de manera persistente cuando la lesión CPD alcanza el sitio activo. La estructura de crio-ME de Pol I-EC_{CPD} presenta una configuración intermedia entre los estados pre- y post-translocados con la lesión de CPD posicionada sobre la hélice puente, la cual está ligeramente doblada en su región central.
2. Diferentes reordenamientos en el sitio de unión de ADN en la enzima crean una red de interacciones que estabilizan el bloqueo. El residuo R1015 en la región central de la hélice puente establece una interacción catión-π con la 3'-T de la lesión CPD y dos residuos con carga positiva en el lazo conmutador 2 alteran su configuración para aproximarse a la hebra molde.
3. El análisis de mutaciones puntuales usando ARNPb como sistema modelo y los ensayos de transcripción *in vitro* revelan que el residuo R1015 juega un papel importante en el procesamiento de las lesiones CPD.

4. Pol I se bloquea en las lesiones CPD debido a la combinación de una lenta incorporación de nucleótidos de ARN frente a la lesión y una actividad de escisión del ARN rápida, debido a la presencia del dominio A12.2-Ct.
5. El mecanismo de bloqueo de Pol I por las lesiones DPC difiere significativamente del observado en Pol II.

BIBLIOGRAPHY

Abrishami, V., J. Vargas, X. Li, Y. Cheng, R. Marabini, C. O. S. Sorzano, J. M. Carazo (2015). "Alignment of direct detection device micrographs using a robust Optical Flow approach." J Struct Biol **189**(3): 163-176.

Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

Awrey, D. E., N. Shimasaki, C. Koth, R. Weilbaecher, V. Olmsted, S. Kazanis, X. Shan, J. Arellano, C. H. Arrowsmith, C. M. Kane and A. M. Edwards (1998). "Yeast Transcript Elongation Factor (TFIIS), Structure and Function: II: RNA polymerase binding, transcript cleavage, and read-through." Journal of Biological Chemistry **273**(35): 22595-22605.

Bae, B., A. Feklistov, A. Lass-Napiorkowska, R. Landick and S. A. Darst (2015). "Structure of a bacterial RNA polymerase holoenzyme open promoter complex." eLife **4**: e08504.

Bai, X. C., G. McMullan and S. H. W. Scheres (2015). "How cryo-EM is revolutionizing structural biology." Trends in Biochemical Sciences **40**(1): 49-57.

Beckouet, F., S. Labarre-Mariotte, B. Albert, Y. Imazawa, M. Werner, O. Gadal, Y. Nogi and P. Thuriaux (2008). "Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription." Molecular and cellular biology **28**(5): 1596-1605.

Beckouët, F., S. Mariotte-Labarre, G. Peyroche, Y. Nogi and P. Thuriaux (2011). "Rpa43 and its partners in the yeast RNA polymerase I transcription complex." FEBS Letters **585**(21): 3355-3359.

Bier, M., S. Fath and H. Tschochner (2004). "The composition of the RNA polymerase I transcription machinery switches from initiation to elongation mode." FEBS Letters **564**(1-2): 41-46.

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

Bodem, J., G. Dobreva, U. Hoffmann-Rohrer, S. Iben, H. Zentgraf, H. Delius, M. Vingron and I. Grummt (2000). "TIF-IA, the factor mediating growth-dependent control of ribosomal RNA synthesis, is the mammalian homolog of yeast Rrn3p." EMBO reports **1**(2): 171.

Bradsher J., J. A., L. Proietti de Santis, S. Iben, J. L. Vonesch, I. Grummt and J. M. Egly (2002). "CSB is a component of RNA polymerase I transcription." Mol Cell **10**(4): 819-829.

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

Brilot, A. F., J. Z. Chen, A. Cheng, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher, R. Henderson and N. Grigorieff (2012). "Beam-induced motion of vitrified specimen on holey carbon film." Journal of structural biology **177**(3): 630-637.

Brueckner, F. and P. Cramer (2008). "Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation." Nature Structural &Amp; Molecular Biology **15**: 811.

Brueckner, F., J. Ortiz and P. Cramer (2009). "A movie of the RNA polymerase nucleotide addition cycle." Current Opinion in Structural Biology **19**(3): 294-299.

Brueckner, F., U. Hennecke, T. Carell and P. Cramer (2007). "CPD damage recognition by transcribing RNA polymerase II." Science **315**(5813): 859-862.

Campbell, E. A., O. Muzzin, M. Chlenov, J. L. Sun, C. A. Olson, O. Weinman, M. L. Trester-Zedlitz and S. A. Darst (2002). "Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit." Molecular Cell **9**(3): 527-539.

Campbell, M. G., A. Cheng, A. F. Brilot, A. Moeller, D. Lyumkis, D. Veesler, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher and N. Grigorieff (2012). "Movies of ice-embedded particles enhance resolution in electron cryo-microscopy." Structure (London, England : 1993) **20**(11): 1823-1828.

Chan, C. L. and R. Landick (1989). "The Salmonella typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins." Journal of Biological Chemistry **264**(34): 20796-20804.

Charton, R., L. Guintini, F. Peyresaubes and A. Conconi (2015). "Repair of UV induced DNA lesions in ribosomal gene chromatin and the role of “Odd” RNA polymerases (I and III)." DNA Repair **36**: 49-58.

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

Cheetham, G. M. T., and T. A. Steitz (2000). "Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases." Current Opinion in Structural Biology **10**(1): 117-123.

Cheung, A. C. M., and P. Cramer (2012). "A movie of RNA polymerase II transcription." Cell **149**: 1431-1437.

Cheung, A. C. M., S. Sainsbury and P. Cramer (2011). "Structural basis of initial RNA polymerase II transcription." The EMBO Journal **30**: 4755-4763.

Compe, E., C. M. Genes, C. Braun, F. Coin and J.-M. Egly (2019). "TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription." Nature Communications **10**(1): 2084.

Conconi, A., V. A. Bespalov and M. J. Smerdon (2002). "Transcription-coupled repair in RNA polymerase I-transcribed genes of yeast." Proceedings of the National Academy of Sciences **99**(2): 649.

Core, L. J. and J. T. Lis (2008). "Transcription Regulation Through Promoter-Proximal Pausing of RNA Polymerase II." Science **319**(5871): 1791.

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

Crick, F. (1970). "Central Dogma of Molecular Biology." Nature **227**(5258): 561-563.

Da, L.-T., F. Pardo-Avila, L. Xu, D.-A. Silva, L. Zhang, X. Gao, D. Wang and X. Huang (2016). "Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue." Nature Communications **7**: 11244.

Darrière, T., M. Pilsl, M.-K. Sarthou, A. Chauvier, T. Genty, S. Audibert, C. Dez, I. Léger-Silvestre, C. Normand, A. K. Henras, M. Kwapisz, O. Calvo, C. Fernández-Tornero, H. Tschochner and O. Gadal (2019). "Genetic analyses led to the discovery of a super-active mutant of the RNA polymerase I." PLOS Genetics **15**(5): e1008157.

Darzacq, X., Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair and R. H. Singer (2007). "In vivo dynamics of RNA polymerase II transcription." Nature Structural &Amp; Molecular Biology **14**: 796.

Dougherty, D. A. (1996). "Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp." Science **271**(5246): 163.

Emsley, P., and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Engel, C., T. Gubbey, S. Neyer, S. Sainsbury, C. Oberthuer, C. Baejen, C. Bernecky and P. Cramer (2017). "Structural Basis of RNA Polymerase I Transcription Initiation." Cell **169**(1): 120-131.e122.

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

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

Fernandez-Leiro, R. and S. H. W. Scheres (2016). "Unravelling biological macromolecules with cryo-electron microscopy." Nature **537**: 339.

Fernández-Tornero, C. (2018). "RNA polymerase I activation and hibernation: unique mechanisms for unique genes." Transcription **9**(4): 248-254.

Fernández-Tornero, C., B. Böttcher, U. J. Rashid, U. Steuerwald, B. Flörchinger, D. P. Devos, D. Lindner and C. W. Müller (2010). "Conformational flexibility of RNA polymerase III during transcriptional elongation." The EMBO journal **29**(22): 3762-3772.

Fernández-Tornero, C., B. Böttcher, M. Riva, C. Carles, U. Steuerwald, Rob W. H. Ruigrok, A. Sentenac, C. W. Müller and G. Schoehn (2007). "Insights into Transcription Initiation and Termination from the Electron Microscopy Structure of Yeast RNA Polymerase III." Molecular Cell **25**(6): 813-823.

Fernández-Tornero, C., M. Moreno-Morcillo, U. J. Rashid, N. M. Taylor, F. M. Ruiz, T. Gruene, P. Legrand, U. Steuerwald and C. W. Müller (2013). "Crystal structure of the 14-subunit RNA polymerase I." Nature **502**(7473): 644-649.

Gallivan, J. P. and D. A. Dougherty (1999). "Cation-pi interactions in structural biology." Proceedings of the National Academy of Sciences of the United States of America **96**(17): 9459-9464.

Geiger, S. R., K. Lorenzen, A. Schreieck, P. Hanecker, D. Kostrewa, A. J. R. Heck and P. Cramer (2010). "RNA Polymerase I Contains a TFIIF-Related DNA-Binding Subcomplex." Molecular Cell **39**(4): 583-594.

Gietz, R. D., R. H. Schiestl, A. R. Willems and R. A. Woods (1995). "Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure." Yeast **11**(4): 355-360.

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

Goetze, H., M. Wittner, S. Hamperl, M. Hondele, K. Merz, U. Stoeckl and J. Griesenbeck (2010). "Alternative Chromatin Structures of the 35S rRNA Genes in Saccharomyces cerevisiae Provide a Molecular Basis for the Selective Recruitment of RNA Polymerases I and II." Molecular and Cellular Biology **30**(8): 2028.

Grohmann, D., A. Hirtreiter and F. Werner (2009). "RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly in vitro." Biochemical Journal **421**(3): 339.

Guo, X., A. G. Myasnikov, J. Chen, C. Crucifix, G. Papai, M. Takacs, P. Schultz and A. Weixlbaumer (2018). "Structural basis for NusA stabilized transcriptional pausing." Molecular Cell **69**(5): 816-827.e814.

Guzder, S. N., P. Sung, V. Bailly, L. Prakash and S. Prakash (1994). "RAD25 is a DMA helicase required for DNA repair and RNA polymerase II transcription." Nature **369**(6481): 578-581.

Han, Y., C. Yan, T. H. D. Nguyen, A. J. Jackobel, I. Ivanov, B. A. Knutson and Y. He (2017). "Structural mechanism of ATP-independent transcription initiation by RNA polymerase I." eLife **6**: e27414.

Hanawalt, P. C. and G. Spivak (1999). Transcription-Coupled DNA Repair. Advances in DNA Damage and Repair: Oxygen Radical Effects, Cellular Protection, and Biological Consequences. M. Dizdaroglu and A. E. Karakaya. Boston, MA, Springer US**:** 169-179.

Hanawalt, P. C. and G. Spivak (2008). "Transcription-coupled DNA repair: two decades of progress and surprises." Nature Reviews Molecular Cell Biology **9**: 958.

Hannan, R. D., A. Cavanaugh, W. M. Hempel, T. Moss and L. Rothblum (1999). "Identification of a mammalian RNA polymerase I holoenzyme containing components of the DNA repair/replication system." Nucleic acids research **27**(18): 3720-3727.

He, Y., C. Yan, J. Fang, C. Inouye, R. Tjian, I. Ivanov and E. Nogales (2016). "Near-atomic resolution visualization of human transcription promoter opening." Nature **533**: 359.

Heel, M., R. Portugal and M. Schatz (2009). Multivariate Statistical Analysis in Single Particle (Cryo) Electron Microscopy.

Henderson, R. (2015). "Overview and future of single particle electron cryomicroscopy." Arch Biochem Biophys **581**: 19-24.

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

Hirata, A. and K. S. Murakami (2009). "Archaeal RNA polymerase." Current opinion in structural biology **19**(6): 724-731.

Hoeijmakers, J. H. (2001). "Genome manteniance mechanisms for preventing cancer." Nature **411**: 366-374.

Hoffmann, N. A., A. J. Jakobi, M. Moreno-Morcillo, S. Glatt, J. Kosinski, W. J. H. Hagen, C. Sachse and C. W. Müller (2015). "Molecular structures of unbound and transcribing RNA polymerase III." Nature **528**: 231.

Huet, J., J. M. Buhler, A. Sentenac and P. Fromageot (1975). "Dissociation of two polypeptide chains from yeast RNA polymerase A." Proceedings of the National Academy of Sciences of the United States of America **72**(8): 3034-3038.

Huet, J., F. Wyers, J. M. Buhler, A. Sentenac and P. Fromageot (1976). "Association of RNase H activity with yeast RNA polymerase A." Nature **261**: 431-433.

Iben, S., H. Tschochner, M. Bier, D. Hoogstraten, P. Hozák, J. M. Egly and I. Grummt (2002). "TFIIH plays an essential role in RNA polymerase I transcription." Cell **109**: 297-306.

Kahl, B. F., H. Li and M. R. Paule (2000). "DNA melting and promoter clearance by eukaryotic RNA polymerase I11Edited by R. Ebright." Journal of Molecular Biology **299**(1): 75-89.

Kang, J. Y., T. V. Mishanina, M. J. Bellecourt, R. A. Mooney, S. A. Darst and R. Landick (2018). "RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing." Molecular Cell **69**(5): 802-815.e801.

Kaplan, C. D., K. M. Larsson and R. D. Kornberg (2008). "The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin." Molecular Cell **30**(5): 547-556.

Keener, J., C. A. Josaitis, J. A. Dodd and M. Nomura (1998). "Reconstitution of Yeast RNA Polymerase I Transcription in Vitro from Purified Components." Journal of Biological Chemistry **273**(50): 33795-33802.

Kettenberger, H., K. J. Armache and P. Cramer (2004). "Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS." Molecular Cell **16**(6): 955-965.

Khatter, H., M. K. Vorländer and C. W. Müller (2017). "RNA polymerase I and III: similar yet unique." Current Opinion in Structural Biology **47**: 88-94.

Kimanius, D., B. O. Forsberg, S. H. Scheres and E. Lindahl (2016). "Accelerated cryo-EM structure determination woth parallelisation using GPUs in RELION-2." Elife **5**.

Kireeva, M. L. and M. Kashlev (2009). "Mechanism of sequence-specific pausing of bacterial RNA polymerase." Proceedings of the National Academy of Sciences **106**(22): 8900.

Kireeva, M. L., K. Opron, S. A. Seibold, C. Domecq, R. I. Cukier, B. Coulombe, M. Kashlev and Z. F. Burton (2012). "Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase." BMC biophysics **5**: 11-11.

Knutson, B. A., J. Luo, J. Ranish and S. Hahn (2014). "Architecture of the Saccharomyces cerevisiae RNA polymerase I Core Factor complex." Nature Structural &Amp; Molecular Biology **21**: 810.

Knutson, B. A., M. L. Smith, N. Walker-Kopp and X. Xu (2016). "Super elongation complex contains a TFIIF-related subcomplex." Transcription **7**: 133-140.

Kong, M., L. Liu, X. Chen, Katherine I. Driscoll, P. Mao, S. Böhm, Neil M. Kad, Simon C. Watkins, Kara A. Bernstein, John J. Wyrick, J.-H. Min and B. Van Houten (2016). "Single-Molecule Imaging Reveals that Rad4 Employs a Dynamic DNA Damage Recognition Process." Molecular Cell **64**(2): 376-387.

Korkhin, Y., U. M. Unligil, O. Littlefield, P. J. Nelson, D. I. Stuart, P. B. Sigler, S. D. Bell and N. G. A. Abrescia (2009). "Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure." PLoS biology **7**(5): e1000102-e1000102.

Kühlbrandt, W. (2014). "The Resolution Revolution." Science **343**(6178): 1443.

Kuhn, C. D., S. R. Geiger, S. Baumli, M. Gartmann, J. Gerber, S. Jennebach, T. Mielke, H. Tschochner, R. Beckmann and P. Cramer (2007). "Functional architecture of RNA polymerase I." Cell **11**(3): 319-324.

Laferté, A., E. Favry, A. Sentenac, M. Riva, C. Carles and S. Chédin (2006). "The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components." Genes & development **20**(15): 2030-2040.

Lainé, J. P. and J. M. Egly (2006). "Initiation of DNA repair mediated by a stalled RNA polymerase IIO." The EMBO Journal **25**(2): 387.

Lalo, D., J. S. Steffan, J. A. Dodd and M. Nomura (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.

Landick, R. (2006). "The regulatory roles and mechanism of transcriptional pausing." Biochemical Society Transactions **34**(6): 1062.

Landick, R. and C. Yanofsky (1987). "Isolation and structural analysis of the Escherichia colitrp leader paused transcription complex." Journal of Molecular Biology **196**(2): 363-377.

Langer, D., J. Hain, P. Thuriaux and W. Zillig (1995). "Transcription in archaea: similarity to that in eucarya." Proceedings of the National Academy of Sciences **92**(13): 5768.

Liljelund, P., S. Mariotte, J. M. Buhler and A. Sentenac (1992). "Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences **89**(19): 9302.

Lin, C. W., B. Moorefield, J. Payne, P. Aprikian, K. Mitomo and R. H. Reeder (1996). "A novel 66-kilodalton protein complexes with Rrn6, Rrn7, and TATA-binding protein to promote polymerase I transcription initiation in Saccharomyces cerevisiae." Molecular and Cellular Biology **16**(11): 6436.

Lindsey-Boltz, L. A. and A. Sancar (2007). "RNA polymerase: The most specific damage recognition protein in cellular responses to DNA damage?" Proceedings of the National Academy of Sciences **104**(33): 13213.

Lisica, A., C. Engel, M. Jahnel, É. Roldán, E. A. Galburt, P. Cramer and S. W. Grill (2016). "Mechanisms of backtrack recovery by RNA polymerases I and II." Proceedings of the National Academy of Sciences **113**(11): 2946.

Liu, X., D. A. Bushnell, D. A. Silva, X. Huang and R. D. Kornberg (2011). "Initiation Complex Structure and Promoter Proofreading." Science **333**(6042): 633.

Mathew, R. and D. Chatterji (2006). "The evolving story of the omega subunit of bacterial RNA polymerase." Trends in Microbiology **14**(10): 450-455.

McMullan, G., A. R. Faruqi, D. Clare and R. Henderson (2014). "Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy." Ultramicroscopy **147**: 156-163.

Miller, O. L. and B. R. Beatty (1969). "Visualization of Nucleolar Genes." Science **164**(3882): 955.

Minakhin, L., S. Bhagat, A. Brunning, E. A. Campbell, S. A. Darst, R. H. Ebright and K. Severinov (2001). "Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly." Proceedings of the National Academy of Sciences **98**(3): 892.

Mitchell, J. R., J. H. Hoeijmakers and L. J. Niedernhofer (2003). "Divide and conquer: Nucleotide excision repair battles cancer and ageing." Curr Opin Cell Biol(2): 232-240.

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

Moreno-Morcillo, M., N. M. Taylor, T. Gruene, P. Legrand, U. J. Rashid, F. M. Ruiz, U. Steuerwald, C. W. Müller and C. Fernández-Tornero (2014). "Solving the RNA polymerase I structural puzzle." Acta Crystallogr D Biol Crystallogr **70**(10): 2570-2582.

Mu, H., N. E. Geacintov, S. Broyde, J. E. Yeo and O. D. Schärer (2018). "Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair." DNA Repair **71**: 33-42.

Murakami, K., H. Elmlund, N. Kalisman, D. A. Bushnell, C. M. Adams, M. Azubel, D. Elmlund, Y. Levi-Kalisman, X. Liu, B. J. Gibbons, M. Levitt and R. D. Kornberg (2013). "Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex." Science **342**(6159): 1238724.

Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin and S. A. Darst (2002). "Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex." Science **296**(5571): 1285.

Murakami, K. S., S. Masuda and S. A. Darst (2002). "Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution." Science **296**(5571): 1280.

Muse, G. W., D. A. Gilchrist, S. Nechaev, R. Shah, J. S. Parker, S. F. Grissom, J. Zeitlinger and K. Adelman (2007). "RNA polymerase is poised for activation across the genome." Nature genetics **39**(12): 1507-1511.

Neyer, S., M. Kunz, C. Geiss, M. Hantsche, V. V. Hodirnau, A. Seybert, C. Engel, M. P. Scheffer, P. Cramer and A. S. Frangakis (2016). "Structure of RNA polymerase I transcribing ribosomal DNA genes." Nature **540**: 607-610.

Nogales, E. and S. H. Scheres (2015). "Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity." Molecular Cell **58**(4): 677-689.

Ogi, T. and A. R. Lehmann (2006). "The Y-family DNA polymerase κ (pol κ) functions in mammalian nucleotide-excision repair." Nature Cell Biology **8**(6): 640-642.

Pani, B. and E. Nudler (2017). "Mechanistic insights into transcription coupled DNA repair." DNA Repair **56**: 42-50.

Perdew, G. H., J. P. VandenHeuvel, and J. M. Petters (2006). Regulation of Gene Expression: molecular mechanism.

Petes, T. D. (1979). "Yeast ribosomal DNA genes are located on chromosome XII." Proceedings of the National Academy of Sciences of the United States of America **76**(1): 410-414.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Peyresaubes, F., C. Zeledon, L. Guintini, R. Charton, A. Muguet and A. Conconi (2017). "RNA Polymerase-I dependent transcription-coupled nucleotide excision repair of UV-Induced DNA lesions at transcription termination sites, in Saccharomyces cerevisiae." Photochem Photobiol **93**: 363-374.

Peyroche, G., P. Milkereit, N. Bischler, H. Tschochner, P. Schultz, A. Sentenac, C. Carles and M. Riva (2000). "The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3." The EMBO Journal **19**(20): 5473.

Pilsl, M., C. Crucifix, G. Papai, F. Krupp, R. Steinbauer, J. Griesenbeck, P. Milkereit, H. Tschochner and P. Schultz (2016). "Structure of the initiation-competent RNA polymerase I and its implication for transcription." Nature Communications **7**: 12126.

Plaschka, C., M. Hantsche, C. Dienemann, C. Burzinski, J. Plitzko and P. Cramer (2016). "Transcription initiation complex structures elucidate DNA opening." Nature **533**: 353.

Prakash, S. and L. Prakash (2000). "Nucleotide excision repair in yeast." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **451**(1): 13-24.

Ravanat, J. L., T. Douki and J. Cadet (2001). "Direct and indirect effects of UV radiation on DNA and its components." J Photochem Photobiol B **63**: 88-102.

Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann and B. Séraphin (1999). "A generic protein purification method for protein complex characterization and proteome exploration." Nat Biotechnol **17**(10): 1030-1032.

Rohou, A., and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." J Struct Biol **192**(2): 216-221.

Russell, J. and J. C. Zomerdijk (2005). "RNA-polymerase-I-directed rDNA transcription, life and works." Trends in Biochemical Sciences **30**(2): 87-96.

Sadian, Y., L. Tafur, J. Kosinski, A. J. Jakobi, R. Wetzel, K. Buczak, W. J. H. Hagen, M. Beck, C. Sachse and C. W. Müller (2017). "Structural insights into transcription initiation by yeast RNA polymerase I." The EMBO Journal **36**(18): 2698.

Sainsbury, S., C. Bernecky and P. Cramer (2015). "Structural basis of transcription initiation by RNA polymerase II." Nature Reviews Molecular Cell Biology **16**: 129.

Sancar, A. and J. T. Reardon (2004). Nucleotide Excision Repair in E. Coli and Man. Advances in Protein Chemistry, Academic Press. **69:** 43-71.

Scheres, S. H. W., H. Gao, M. Valle, G. T. Herman, P. P. B. Eggermont, J. Frank and J.-M. Carazo (2007). "Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization." Nature Methods **4**: 27.

Schnapp, A., G. Schnapp, B. Erny and I. Grummt (1993). "Function of the growth-regulated transcription initiation factor TIF-IA in initiation complex formation at the murine ribosomal gene promoter." Molecular and Cellular Biology **13**(11): 6723.

Schneider, D. A. (2011). "RNA polymerase I activity is regulated at multiple steps in the transcription cycle: recent insights into factors that influence transcription elongation." Gene **493**(2): 176-184.

Schneider, D. A., A. Michel, M. L. Sikes, L. Vu, J. A. Dodd, S. Salgia, Y. N. Osheim, A. L. Beyer and M. Nomura (2007). "Transcription Elongation by RNA Polymerase I Is Linked to Efficient rRNA Processing and Ribosome Assembly." Molecular Cell **26**(2): 217-229.

Sheres, S. H. (2010). "Classification of structural heterogeneity by maximum-likelihood methods." Methods Enzymol **482**: 295-320.

Sheres, S. H. (2012). "RELION: implementation of a Bayesian approach to cryo-EM structure determination." J Struct Biol **180**(3): 519-530.

Sigworth, F. J., P. C. Doerschuk, J. M. Carazo and S. H. W. Scheres (2010). "An introduction to maximum-likelihood methods in cryo-EM." Methods in enzymology **482**: 263-294.

Silva, D. A., D. R. Weiss, F. Pardo Avila, L.-T. Da, M. Levitt, D. Wang and X. Huang (2014). "Millisecond dynamics of RNA polymerase II translocation at atomic resolution." Proceedings of the National Academy of Sciences **111**: 7665-7670.

Steffan, J. S., D. A. Keys, J. A. Dodd and M. Nomura (1996). "The role of TBP in rDNA transcription by RNA polymerase I in Saccharomyces cerevisiae: TBP is required for upstream activation factor-dependent recruitment of core factor." Genes & Development **10**(20): 2551-2563.

Steitz, T. A. (2009). "The structural changes of T7 RNA polymerase from transcription initiation to elongation." Current opinion in structural biology **19**(6): 683-690.

Strauss, B. S. (1991). "The 'A rule' of mutagen specificity: A consequence of DNA polymerase bypass of non-instructional lesions?" BioEssays **13**(2): 79-84.

Sung, P., S. N. Guzder, L. Prakash and S. Prakash (1996). "Reconstitution of TFIIH and Requirement of Its DNA Helicase Subunits, Rad3 and Rad25, in the Incision Step of Nucleotide Excision Repair." Journal of Biological Chemistry **271**(18): 10821-10826.

Sung, P., L. Prakash, S. W. Matson and S. Prakash (1987). "RAD3 protein of Saccharomyces cerevisiae is a DNA helicase." Proceedings of the National Academy of Sciences **84**(24): 8951.

Sydow, J. F., F. Brueckner, A. C. M. Cheung, G. E. Damsma, S. Dengl, E. Lehmann, D. Vassylyev and P. Cramer (2009). "Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA." Molecular Cell **34**(6): 710-721.

Tafur, L., Y. Sadian, J. Hanske, R. Wetzel, F. Weis and C. W. Müller (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.

Tafur, L., Y. Sadian, N. A. Hoffmann, A. J. Jakobi, R. Wetzel, W. J. Hagen, C. Sachse and C. W. Müller (2016). "Molecular structures of transcribing RNA polymerase I." Mol Cell **64**: 1135-1143.

Taylor, J.-S. (2002). "New structural and mechanistic insight into the A-rule and the instructional and non-instructional behavior of DNA photoproducts and other lesions." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **510**(1): 55-70.

Tornaletti, S., and P. C. Hanawalt (1999). "Effect of DNA lesions on transcription elongation." Biochimie **81**: 139-146.

Torreira, E., J. A. Louro, I. Pazos, N. González-Polo, D. Gil-Carton, A. G. Duran, S. Tosi, O. Gallego, O. Calvo and C. Fernández-Tornero (2017). "The dynamic assembly of distinct RNA polymerase I complexes modulates rDNA transcription." Elife **6**.

Toulokhonov, I., J. Zhang, M. Palangat and R. Landick (2007). "A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement during Transcriptional Pausing." Molecular Cell **27**(3): 406-419.

Vannini, A. and P. Cramer (2012). "Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries." Molecular Cell **45**(4): 439-446.

Verhage, R. A., P. Van de Putte and J. Brouwer (1996). "Repair of rDNA in Saccharomyces cerevisiae: RAD4-independent strand-specific nucleotide excision repair of RNA polymerase I transcribed genes." Nucleic acids research **24**(6): 1020-1025.

Walmacq, C., A.C. Cheung, M. L. Kireeva, L. Lubkowska, C. Ye, D. Gotte, J. N. Strathern, T. Carell, P. Cramer and M. Kashlev (2012). "Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage." Cell **46**: 18-29.

Walmacq, C., L. Wang, J. Chong, K. Scibelli, L. Lubkowska, A. Gnatt, P. J. Brooks, D. Wang and M. Kashlev (2015). "Mechanism of RNA polymerase II bypass of oxidative cyclopurine DNA lesions." Proceedings of the National Academy of Sciences **112**(5): E410.

Wang, D., D. A. Bushnell, K. D. Westover, C. D. Kaplan and R. D. Kornberg (2006). "Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis." Cell **127**: 941-954.

Wang, W., C. Walmacq, J. Chong, M. Kashlev and D. Wang (2018). "Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II." Proceedings of the National Academy of Sciences **115**(11): E2538.

Wang, W., J. Xu, J. Chong and D. Wang (2018). "Structural basis of DNA lesion recognition for eukaryotic transcription-coupled nucleotide excision repair." DNA Repair **71**: 43-55.

Wei, Y.-y. and H.-T. Chen (2018). "Functions of the TFIIE-Related Tandem Winged-Helix Domain of Rpc34 in RNA Polymerase III Initiation and Elongation." Molecular and Cellular Biology **38**(4): e00105-00117.

Weixlbaumer, A., K. Leon, R. Landick and Seth A. Darst (2013). "Structural Basis of Transcriptional Pausing in Bacteria." Cell **152**(3): 431-441.

Werner, F. (2007). "Structure and function of archaeal RNA polymerases." Molecular Microbiology **65**(6): 1395-1404.

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

Wind, M. and D. Reines (2000). "Transcription elongation factor SII." BioEssays : news and reviews in molecular, cellular and developmental biology **22**(4): 327-336.

Wirth, N., J. Gross, H. M. Roth, C. N. Buechner, C. Kisker and I. Tessmer (2016). "Conservation and Divergence in Nucleotide Excision Repair Lesion Recognition." The Journal of biological chemistry **291**(36): 18932-18946.

Xu, J., I. Lahiri, W. Wang, A. Wier, M. A. Cianfrocco, J. Chong, A. A. Hare, P. B. Dervan, F. DiMaio, A. E. Leschziner and D. Wang (2017). "Structural basis for the initiation of eukaryotic transcription-coupled DNA repair." Nature **551**: 653.

Xu, L., W. Wang, J. Wu, J. H. Shin, P. Wang, I. C. Unarta, J. Chong, Y. Wang and D. Wang (2017). "Mechanism of DNA alkylation-induced transcriptional stalling, lesion bypass, and mutagenesis." Proceedings of the National Academy of Sciences **114**(34): E7082.

Xu, Y., C. Bernecky, C.-T. Lee, K. C. Maier, B. Schwalb, D. Tegunov, J. M. Plitzko, H. Urlaub and P. Cramer (2017). "Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex." Nature Communications **8**: 15741.

Yang, Y., J. Hu, C. P. Selby, W. Li, A. Yimit, Y. Jiang and A. Sancar (2019). "Single-nucleotide resolution analysis of nucleotide excision repair of ribosomal DNA in humans and mice." Journal of Biological Chemistry **294**(1): 210-217.

Yates, L. A., R. J. Aramayo, N. Pokhrel, C. C. Caldwell, J. A. Kaplan, R. L. Perera, M. Spies, E. Antony and X. Zhang (2018). "A structural and dynamic model for the assembly of Replication Protein A on single-stranded DNA." Nature Communications **9**(1): 5447.

Zhang, G., E. A. Campbell, L. Minakhin, C. Richter, K. Severinov and S. A. Darst (1999). "Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution." Cell **98**(6): 811-824.

Zhang, K. (2016). "Gctf:Real-time CTF determination and correction." J Struct Biol **193**: 1-12.

Zhang, N., J. Schäfer, A. Sharma, L. Rayner, X. Zhang, R. Tuma, P. Stockley and M. Buck (2015). "Mutations in RNA Polymerase Bridge Helix and Switch Regions Affect Active-Site Networks and Transcript-Assisted Hydrolysis." Journal of Molecular Biology **427**(22): 3516-3526.

CONTRIBUTION TO SCIENTIFIC ARTICLES

Sanz-Murillo, M., J. Xu, G. A. Belogurov, O. Calvo, D. Gil-Carton, M. Moreno-Morcillo, D. Wang and C. Fernández-Tornero (2018). "Structural basis of RNA polymerase I stalling at UV light-induced DNA damage." Proceedings of the National Academy of Sciences **115**: 8972-8977.