Regulation of Cytoplasmic RNA Polymerase II

by

Annamaria Hlavata

March 2025

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:
Beatriz Vicoso, PhD, Chair
Carrie Bernecky, PhD
Florian Schur, PhD
Dea Slade, PhD

The thesis of Annamaria Hlavata, titled *Regulation of Cytoplasmic RNA Polymerase II* is approved by:

Supervisor: Carrie Bernecky, PhD, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Florian Schur, PhD, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Dea Slade, PhD, Max Perutz Labs, Vienna, Austria; Medical University of Vienna, Vienna, Austria

Signature: _____

Defense Chair: Beatriz Vicoso, PhD, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

ISBN: 978-3-99078-055-8

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I accept full responsibility for the content and factual accuracy of this work, including the data and their analysis and presentation, and the text and citation of other work.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Annamaria Hlavata
March 2025

Signed page is on file

Abstract

Gene expression is crucial for cell differentiation, development and survival of organisms. It consists of several steps, starting with transcription that is mediated by RNA polymerases. These are protein machineries transcribing and producing different types of RNAs. Although, the individual steps of transcription by RNA polymerase II (Pol II) as well as the structure of Pol II has been extensively studied, surprisingly, there is still little known about its regulation and assembly in cytoplasm. Among the proteins that are important in biogenesis of Pol II are RNA polymerase II associating proteins (RPAP) and small GPN-loop GTPases (GPN). Both of these protein groups were shown to take essential part in assembly of Pol II.

The aim of this project was to deepen our knowledge in regulation of Pol II in the cytoplasm as well as the proteins involved in this process. Techniques of structural biology, biochemistry and cell biology were employed to study and characterize cytoplasmic Pol II and its interacting partners.

This study shows for the first time the structure of cytoplasmic Pol II at high resolution. The structure also reveals proteins interacting with Pol II in cytoplasm, namely GDOWN1, RPAP2. Comparing the structure of cytoplasmic Pol II with transcribing Pol II revealed striking difference in clamp region that is not in closed state. Furthermore, GDOWN1 and RPAP2 make steric clashes with various transcription factors bound to Pol II during different stages of transcription. Even though GPN1 and GPN3 proteins were not resolved in the cytoplasmic Pol II structure, they are part of the complex and their interaction with Pol II was confirmed *in vitro*. RPAP2 stabilizes these proteins on Pol II and several experiments suggest that they interact with the clamp region. In addition, GDOWN1, RPAP2 and GPNs might keep clamp in open or partially open state. Based on these results I propose a novel model of regulation of Pol II in cytoplasm. GDOWN1, RPAP2, GPN1 and GPN3 bind to Pol II in cytoplasm and doing so they can prevent pre-mature binding of DNA or RNA and different transcription factors to Pol II in cytoplasm or before engaging in transcription nucleus.

This research contributes to the current knowledge of molecular mechanisms of Pol II regulation in cytoplasm.

Acknowledgement

This thesis, which has been a result of several years of dedication and hard work, would not have been possible without support and encouragement of many people. They played a crucial role for me to make it here.

Very special thanks belongs to my supervisor, Carrie Bernecky, for her leadership and for giving me the opportunity to be part of her research group where I was able to gain a lot of experience and grow my scientific potential. Her expertise and guidance have been invaluable during my PhD. She was always positive about the project and enthusiastic whenever I showed her some exciting results. Her motivation was a big driving force through my whole PhD journey and because of that, I was able to tackle the problems I faced with optimism.

I would also like to thank to the members of my PhD committee, Dea Slade and Florian Schur, for their constructive feedback and giving me directions on what to focus in the PhD project. I always felt comfortable during our meetings and there was always friendly and positive atmosphere.

Thanks to the lab members of Bernecky group for creating friendly atmosphere and place for scientific discussion.

I would also like to thank all the collaborators involved in this project. Namely, Clemens Plaschka and his lab for generating the cell lines used in this project and support in the mass spectrometry experiments. I also appreciate recent collaboration with Franz Herzog lab. They helped me with the cross-linking mass spectrometry experiments.

I would also like to acknowledge the ISTA Facilities: Lab Support Facility, Protein Services and Electron Microscopy Facility (EMF) and Scientific Computing. EMF for their support during data collections and troubleshooting, especially Valentin. Scientific Computing for solving quickly any issues related with cluster.

I would like to thank to ISTA for providing the resources and academic environment that allowed me to pursue my research.

Finally, I cannot forget to thank my family and closest friends. My parents, for always standing by me and supporting me and my decision to become a scientist. They are my rock and I can always rely on them and discuss all my troubles whenever I need. Julian, thank you for your patience with me. I think without your constant reassurance: “You can do it!” and “Everything is going to be fine!” I would have much harder time. Also, big thanks to all my friends.

About the author

Annamaria Hlavata, MSc., is a PhD candidate at Institute of Science and Technology Austria (ISTA) in the lab of Carrie Bernecky, PhD. Annamaria obtained her BSc and MSc titles at Comenius University in Bratislava, Slovakia, in the Department of Genetics. Afterwards, Annamaria joined ISTA and her PhD project focuses on understanding regulation of mammalian RNA polymerase II in cytoplasm.

During her PhD journey, Annamaria acquired experience in various techniques of cryo-electron microscopy (cryoEM), biochemistry and cell biology. During her PhD she also presented her findings to the Vienna cryoEM community. Annamaria is now in process of preparing a manuscript.

List of Figures

List of Tables

List of Abbreviations

AID	auxin-inducible degron
BS3	bis(sulfosuccinimidyl)suberate
CAS	cytoplasmic anchoring sequence
CDC73	Parafibromin
cryoEM	cryo-electron microscopy
CTD	C-terminal domain of RNA polymerase II
CTR9	RNA polymerase-associated protein CTR9 homolog
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSBU	disuccinimidyl dibutyric urea
DSIF	DRB sensitivity inducing factor
DSSO	disuccinimidyl sulfoxide
GA	glutaraldehyde
GDOWN1	DNA-directed RNA polymerase II subunit GRINL1A
GPN	GPN-loop GTPase
GTF	general transcription factors
HSP	heat-shock protein
IAA	auxin, indolic acid
INT	Integrator
Iwr1	interacting with RNA polymerase II protein 1
LEO1	RNA polymerase-associated protein LEO1
LMB	leptomycin B
MED	Mediator
MS	mass-spectrometry
NELF	negative elongation factor
NES	nuclear export signal
NLS	nuclear localization signal
NPC	nuclear pore complex
NTP	nucleoside triphosphate
PAF1	RNA polymerase II-associated factor 1 homolog
PIC	pre-initiation complex
Pol	RNA polymerase
P-TEFb	positive transcription elongation factor
RNA	ribonucleic acid
RPAP	RNA polymerase II associated protein
RPB	DNA-directed RNA polymerase II subunit
RTF1	RNA polymerase-associated protein RTF1 homolog
SPT4/5/6	transcription elongation factor SPT4/5/6
TAFs	TFIID-associated factors
TBP	TATA-binding protein
TFIIA	transcription factor II A
TFIIB	transcription factor II B
TFIID	transcription factor II D
TFIIF	transcription factor II F
TFIIH	transcription factor II H
TFIIS	transcription elongation factor II-S
TRiC	T-complex protein ring complex
WDR61	WD repeat-containing protein 61
XL-MS	cross-linking mass-spectrometry

Table of Contents

1. Introduction

1.1. Scientific background

Gene expression is essential for maintaining the integrity of cells and organisms. In higher organisms, it plays a crucial role in cell differentiation and metabolism (Levine *et al.*, 2014). Any disruption in this process can lead to disease (Herz *et al.*, 2014). Transcription is the first step of gene expression and is carried out by enzymes, called RNA polymerases. While mitochondrial RNA polymerase consist of one single protein, nuclear RNA polymerases are multisubunit complexes (Gaspari *et al.*, 2004). Since transcription is important to maintain cellular homeostasis, structure and function of RNA polymerases as well as general transcription factors (TFs) have been heavily studied. However, there is not much known about biogenesis of these complexes and their regulation during this process. Many similar protein machineries are assembled with help of different chaperones and/or specific assembly factors and are highly regulated (Makhnevych & Houry, 2012). Correct assembly of protein complexes is essential for the cells and any discrepancy can lead to diseases or death (Klinge & Woolford Jr., 2019). Therefore, it is important to better understand the underlying mechanism. Furthermore, altered patterns in transcription are prominent in many diseases, especially in cancer (Bradner *et al.*, 2018). Thus, studying and developing inhibitors of transcription is a very promising strategy in fighting cancer and other diseases. Understanding regulation of cytoplasmic RNA polymerases can help in developing new drugs (Mapp *et al.*, 2015; Bradner *et al.*, 2017). The proposed project focuses on the characterization of regulation of mammalian RNA polymerase II in cytoplasm and proteins involved in this process.

1.2. Transcription mediated by RNA polymerases

Eukaryotic transcription in nucleus is facilitated by three polymerases. RNA polymerase I (Pol I) is located in nucleolus, transcribes 28S, 5.8S and 18S rRNAs and consists of 13 subunits. RNA polymerase III (Pol III) synthesizes tRNAs, 5S rRNA and other small RNAs. It comprises 17 subunits. Protein-coding genes are transcribed by 12-subunit RNA polymerase II (Pol II) into mRNAs. Pol II also produces small nuclear RNAs, micro-RNAs and small interfering RNAs. The three above-mentioned eukaryotic RNA po-lymerases have a 10-subunit core and additional peripheral subunits. Five of the core subunits are shared between the three polymerases - Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 (Cramer *et al.*, 2008).

1.2.1. Structure of Pol II

Structures of yeast and mammalian Pol II at various stages of transcription have been resolved using X-crystallography and cryo-electron microscopy (cryoEM) techniques in past years. These results brought information about how the particular transcription steps work and opened the way to new drug development. The most prominent features of Pol II are clamp, cleft and stalk (Fig. 1). The cleft, formed by RPB1 and RPB2, is positively charged, enabling accommodation of single-stranded promoter DNA upon promoter melting in the active site of the enzyme (He *et al.*, 2016). There, nucleotide triphosphates (NTPs) are used for RNA synthesis and formation of an RNA:DNA hybrid that is later separated by the rudder and lid domains of RPB1, whereas RNA exits via the RNA exit channel and DNA along the protrusion (Fig. 1). The clamp opens and closes during different stages of transcription. Further separation and unwinding of DNA template is facilitated by the bridge helix, when DNA

passes near it (Bruckner & Cramer, 2008; Bernecky *et al*., 2016). Peripheral subunits of Pol II RPB7/Rpb7 and RPB4/Rpb4 form the stalk that is located close to RPB6 (Fig. 1). The stalk provides interaction sites with various initiation and elongation factors (Bushnell & Kornberg, 2003). One very important region of RPB1 is the C-terminal domain (CTD). It consists of several heptapeptide repeats with consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$. The CTD participates in transcription regulation and serves as a substrate for phosphorylation. Throughout the transcription, CTD is dynamically phosphorylated and dephosphorylated depending on the transcription stage. Phosphorylation state of the CTD residues is important for interaction with different proteins regulating transcription (Heidemann *et al*., 2013, Harlen & Churchman, 2017). Pol II recruited to promoter sites has unphosphorylated CTD (Lu *et al.*, 1991). After binding to promoter region of the gene Pol II is phosphorylated on Ser5 and Ser7 residues of CTD that causes promoter escape of Pol II (Wong *et al*., 2014). Due to its high flexibility, CTD is not resolved in Pol II structures (Fig. 1).

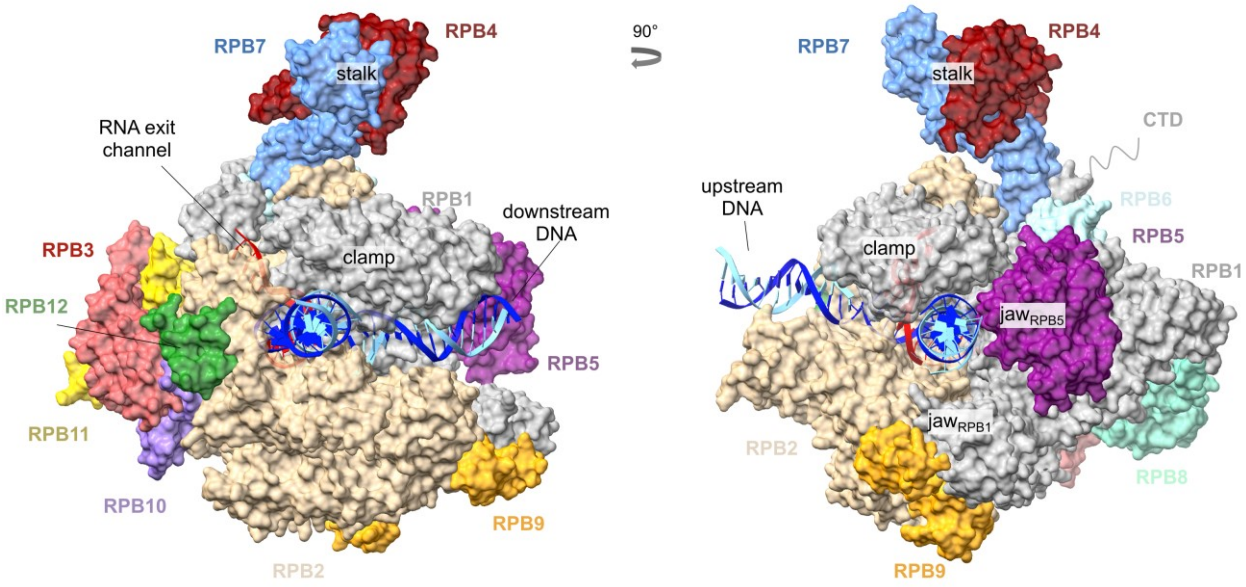


Figure 1: Schematic of mammalian Pol II structure. It shows the subunits of Pol II and DNA-RNA scaffold bound in the cleft of the Pol II. Pol II subunits, DNA and RNA are color-coded. Surface representation of the transcribing Pol II model. Structure adapted from Bernecky *et al.* 2016.

1.2.2. Regulation of Pol II during transcription

Pol II is highly regulated during transcription in nucleus. Initiation is the first step in transcription mediated by Pol II. First, a pre-initiation complex (PIC) needs to be formed. PIC consists of around 80 proteins, including Pol II, TFs and Mediator complex. A correct assembly of PIC is essential for gene expression. In the canonical model of PIC assembly, TFIID, containing TATA-binding protein (TBP) and 13 TFIID-associated factors (TAFs), recognizes the promoter DNA and places TBP on the promoter (Burley and Roeder, 1996; Hou and Kraus, 2021). Transcription factor TFIIA aids proper binding of TBP to DNA. Once TBP and TFIIA are bound to the DNA promoter, TFIIB is recruited. Presence of TFIIB allows Pol II-TFIIF complex, formed in the meantime, to bind to the promoter region (Kim *et al*., 1993; Imbalzano *et al*., 1994). At this stage, Pol II CTD domain remains unphosphorylated and is recognized by Mediator. Mediator is a multisubunit complex and has a role in an activator-dependent transcription and formation of PIC (Malik & Roeder, 2005). Afterwards, TFIIE and TFIIH

are recruited to Mediator and TFIIH phosphorylates CTD. DNA is structurally rearranged that leads to DNA promoter opening and forming a transcriptional bubble (Goodrich and Tjian, 1994; Holstege *et al*. 1996; Plaschka *et al*., 2016). Phosphorylation of CTD also causes dissociation of Pol II from Mediator and other TFs and promoter escape (Max *et al*., 2007).

In early elongation, promoter-proximal pausing of Pol II is crucial step for correct onset of productive elongation of transcription. Paused Pol II can enter either elongation and produce mRNAs or undergo termination and it is released from the DNA. Pol II transcribes around 25-50 nt until it is paused and stabilized by DSIF and NELF proteins. These proteins also block further transcription (Vos *et al*., 2018). DSIF protein consists of SPT5 and SPT4. SPT5 protein is especially important as it is involved in capping of newly synthesized mRNAs. Stalled Pol II is released by P-TEFb factor and PAF complex, consisting of PAF1, LEO1, CTR9, CDC73, WDR61 and RTF1 (Yu *et al*., 2015, Lu *et al*., 2016). This process is highly regulated by cell. After successful release of Pol II, elongation of nascent RNA resumes. If the paused Pol II fails to continue the transcription, it can be terminated prematurely. Integrator is a multisubunit complex that can bind paused Pol II and cleaves the short RNA transcript (Fianu *et al*., 2021). Furthermore, Integrator also removes Pol II from the DNA (Fianu *et al*., 2024).

During active elongation, Pol II can be stalled again when it encounters nucleosomes. Pol II overcomes this barrier with aid of various elongation factors, among them DSIF, SPT6, PAF complex and TFIIS (Farnung *et al*., 2022).

Last step of transcription is termination that involves many more factors, for example TTF2 (Liu *et al*., 1998). It was also reported that TTF2 is involved in mitotic repression of transcription elongation (Yiang *et al*., 2004).

1.3. RNA polymerase II biogenesis and regulation

As mentioned above, there has been a substantial research done to understand the mechanism and regulation of transcription mediated by Pol II. How is the Pol II regulated and assembled before engaging in transcription is not yet fully understood. Correct assembly of Pol II is essential for the function of cells. Multiple studies support the hypothesis that subcomplexes of RNA polymerases are formed in cytoplasm and then put together to create the whole complex ready for nuclear import. First, subassemblies of Pol II - Rpb1-Rpb8, Rpb2-Rpb3-Rpb10-Rpb11 and heterodimer Rpb4/7 - can be isolated separately supporting the idea of ordered Pol II assembly (Edwards *et al*., 1991; Kimura *et al*., 1997). Second, α-amanitin treatment in cells causes enrichment of partially assembled Pol II complexes and cytoplasmic accumulation of RPB3 subunit (Boulon *et al*., 2010). This drug leads to transcriptional arrest of Pol II. α-amanitin interacts with the Rpb1 subunit, leading to stalling of Pol II followed by Rpb1 degradation (Nguyen *et al*., 1996). Co-treatment with a Crm1 exportin inhibitor, leptomycin B (LMB), caused retention of RPB3 in the nucleus and strikingly, accumulation of RPB1 in cytoplasm. Newly synthesized α-amanitin resistant RPB1 could not enter the nucleus, possibly because a factor facilitating RPB1 import was trapped in the nucleus. Surprisingly, cells depleted of any Pol II subunit accumulated RPB1 in the cytoplasm leading to the conclusion that RPB1 needs to be part of full Pol II to be transported into the nucleus. It also suggests that Pol II enters the nucleus as fully assembled enzyme (Boulon *et al*., 2010).

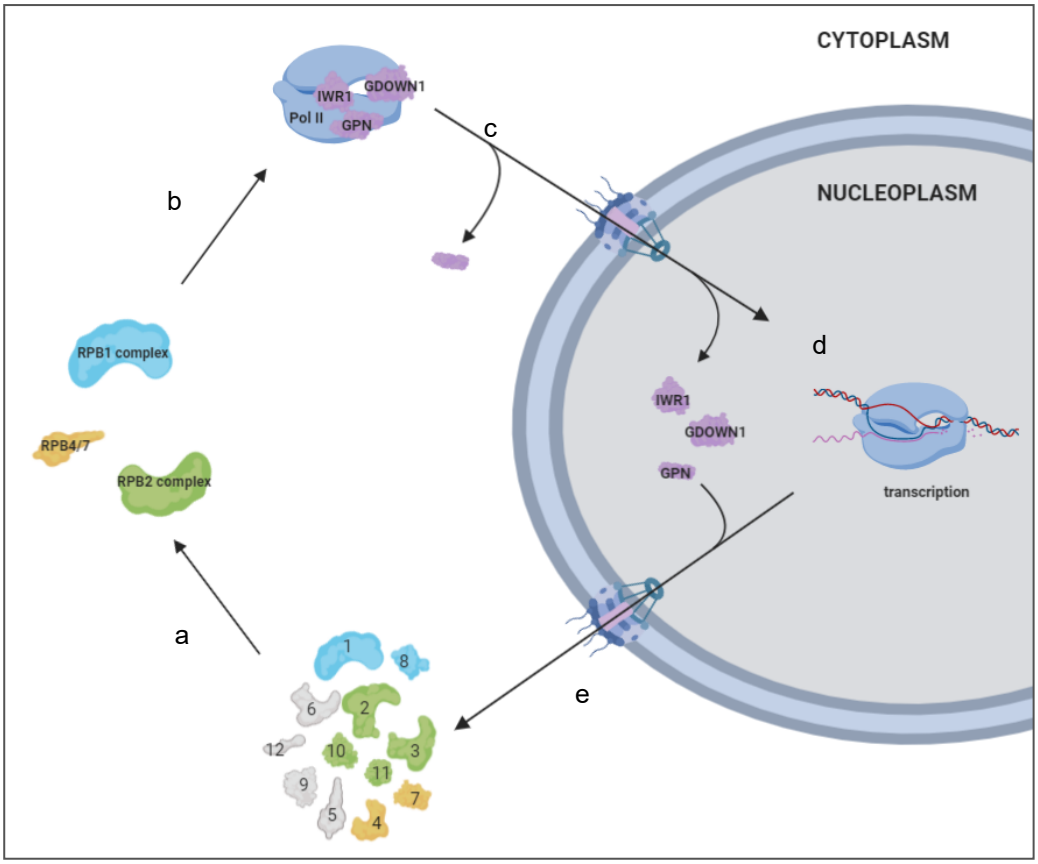


Figure 2: Proposed model of biogenesis of Pol II. (**a**) Free subunits of Pol II are first assembled into Rpb1 (blue), Rpb2 (green) and Rpb4/7 subcomplexes (orange) in the cytoplasm. (**b**) These intermediates form fully assembled Pol II with help of assembly factors (in violet). (**c**) Fully assembled Pol II is transported into the nucleus. This import is facilitated possible by Iwr1 and GPNs. Release of assembly factors from Pol II can happen before or after entering the nucleus. Some can compete with GTFs until Pol II is properly bound to promoter regions (GDOWN1). (**d**) Pol II is engaged in transcription after nuclear import and assembly factors dissociate. (**e**) Later, Pol II is degraded or transported back to the cytoplasm together with released assembly factors, where they can engage in another round of Pol II assembly; adapted from Wild & Cramer, 2012.

The main steps of the Pol II life cycle have been proposed as follows. First, Pol II is assembled in cytoplasm and then transported to nucleus, where gene transcription occurs (Fig. 2 a-e; Wild & Cramer, 2012). Last, Pol II is degraded or transported back to cytoplasm together with other assembly factors after transcription (Fig. 2 e).

Since RNA polymerases share some subunits, it is highly possible there are general and specific RNA polymerase assembly factors. Several studies have focused on characterization of proteins involved in the biogenesis of RNA polymerases.

1.3.1. Protein folding and chaperones

The first step of biogenesis of protein complexes like RNA polymerases is proper folding of newly-synthetized subunits. This process is highly complex also due the vast number of conformations a protein chain can take. The protein must overcome kinetic barriers in order to fold correctly. A failure to do so can result in misfolding and aggregation (Bartlett and Radford, 2009). Chaperones are important for proper folding of proteins after translation and prevent aggregates formation under stress conditions (Finka *et al.*, 2016). These often-multisubunit protein complexes recognize the hydrophobic residues of newly synthesized or misfolded proteins and bind to them in ATP-

dependent manner (Langer *et al*., 1992). There are three major classes of chaperones involved in *de novo* protein folding: HSP70 system, HSP90 system and chaperonins.

1.3.1.1. HSP70 class

HSP70 belongs to the key chaperones in protein folding and general proteostasis. HSP70 consists of an ATPase domain and a peptide binding domain (Mayer, 2010). In the ATP-bound state, HSP70 has a lower affinity for proteins. When ATP is hydrolyzed to ADP, it adopts a closed conformation and locks the peptide. This allows the proteins to fold properly. Once a nucleotide exchange factor exchanges ADP to ATP, HSP70 releases the folded protein. This type of folding is preferred by fast-folding proteins (Sharma *et al*., 2010).

1.3.1.2. Chaperonins

Another class of chaperones are chaperonins. They are relatively large protein complexes forming a double-ring structure that fully encloses the substrate proteins, typically around 60 kDa in size (Horwich and Fenton, 2009). There are two groups of chaperonins: HSP60 and TRiC (also known as CCT).

HSP60 protein in eukaryotes is present in organelles of endosymbiotic origin, such as chloroplasts and mitochondria. HSP60 has either single or double ring form, depending on of the presence on ATP and HSP10 co-chaperone. When HSP60 interacts with HSP10 in the presence of ATP it forms double rings. HSP10 serves as a lid to enclose the unfolded proteins in the HSP60 cavity. Once ATP is hydrolyzed, HSP10 is released and HSP60 is again in the single ring form (Ishida *et al*., 2018).

TRiC/CCT chaperonins are HSP10 independent and they function in cytosol and they are critical for correct actin and tubulin folding (Grantham *et al*., 2006, Munoz *et al*., 2011). They contain eight subunits in one ring. Their apical domains have protrusions that serve as a lid, thus they do not require HSP10. The lid-like structure are similar to HSP60 closing and opening depending on ATP/ADP bound state. Nevertheless, the process of encapsulation of TRiC complex is much slower than in HSP60-HSP10 cycle (Douglas *et al*., 2011).

1.3.1.3. HSP90 class

HSP90 serves as a central hub, regulating several key signaling pathways. It acts downstream of HSP70, playing a crucial role in the structural maturation and conformational regulation of various protein complexes (Taipale *et al*., 2010, McClellan *et al*., 2007). HSP90 is a homodimer that is assembled through the C-terminal domains. The N-terminal domains bind and hydrolyze ATP. N-terminal and C-terminal domains are connected via the middle domain. The middle domain plays a role in the substrate binding and also interacts with other co-chaperones (Ali *et al*., 2006). It is proposed that once the substrate and ATP is bound to HSP90, the N-terminal domains dimerize and close. When ATP is hydrolyzed N-terminal domains separate, releasing the folded substrate. This whole process is regulated by multiple co-chaperones, for example R2TP/Prefoldin-like complex (Shiau *et al*., 2006).

HSP90 and co-chaperone RT2P are involved in assembly of Pol II. Upon α-amanitin treatment, both HSP90 and RT2P complex associated with partially assembled Pol II in cytoplasm (Boulon *et al.*, 2010; Forget *et al.*, 2010). These protein-folding complexes were also shown to associate with Pol I subunits A135 and AC40. The latter is shared with Pol III, which suggests their general role in assembly of RNA polymerases (McClellan *et al.*, 2007).

1.3.2. Pol II assembly factors

1.3.2.1. GPN proteins

A family of GTPases containing a GPN (Gly-Pro-Asn) loop motif was described several years ago (Gras *et al.*, 2007). They associate with assembling Pol II and participate in Pol II biogenesis (Boulon *et al.*, 2010; Forget *et al.*, 2010). Human GPN1 and GPN3 are important in Pol II assembly and its nuclear transport (Figure 2 c). Depletion of GPN1 or GPN3 or inhibition of GPN1 GTPase activity resulted in accumulation of RPB1 in the cytoplasm. Both GPN1 and GPN3 associate with isolated RPB1, RPB4 and RPB7 subunits *in vitro*. Interestingly, only GPN1 has GTPase activity, while GPN3 is inactive (Carre & Shiekhattar, 2011). Activity of GPN2 was not yet tested. Importantly, GPN1 and GPN3 form a stable complex in cells and they are shuttled together between cytoplasm and nucleus. Stability of this complex is essential for the cells and explains why mutations in one of the GPNs can lead to defects in Pol II assembly (Méndez-Hernández *et al*., 2014). Interestingly, when GPN1 is mutated in the GTP binding pocket it can be still shuttled into nucleus, but it does not interact with Pol II or GPN3 there (Carre & Shiekhattar, 2011). Yeast homologs of GPNs also participate in Pol II assembly, suggesting their high conservation among different species (Staresincic *et al.*, 2011; Zeng *et al.*, 2018). Mutants of Gpn2 and Gpn3 show defects in Rpb1, Rpb2 and Rpb3 localization and these RNA polymerase subunits are accumulating in the cytoplasm. These mutants also appear to affect the elongation of transcription, as mutants are more sensitive to transcription inhibitor than wild type cells. In addition, it was observed that Gpn2 and Gpn3 are important in localization of Pol III (Minaker *et al.*, 2013). Interestingly, it was proposed that GPN1 function in assembly of Pol II may be coupled to microtubules. CCT complex, which is important in microtubule folding and polymerization, interacts with Pol II (Dekker *et al.*, 2008) and GPN1 (Forget *et al.*, 2010). Treatment with benomyl, a drug promoting depolarization of microtubules, caused cytoplasmic accumulation of Rpb1. Moreover, Gpn1 yeast mutants are more sensitive to this treatment, linking the biogenesis of Pol II and microtubule polymerization (Forget *et al.*, 2010).

Solved crystal structure of yeast Gpn1 GTPase core in the past showed two different conformations depending on the bound nucleotide (Niesser at al., 2016). Gpn1 has two extra domain compared to prokaryotic homolog that are named insertions. Gpn1-GDP core contains several helices and β-sheets. GTP binding pocket of Gpn1 consists of G1-G5 motifs that are important for the harboring of GTP/GDP and hydrolysis. Overall, Gpn1-GDP structure assumes a closed conformation (Fig. 3 a). When GTP nonhydrolyzable analog GMPPCP was bound to Gpn1, the overall structure showed visible rearrangements (Fig. 3 b). Both insertions undergo a conformational change and are straightened. One of the helices is also rearranged and gets to closer proximity to the GTP/GDP binding pocket (Fig. 3 b). GTP bound Gpn1 is in open state and it also contains a hydrophobic pocket that is common among the chaperone proteins. Gpn1 exhibits chaperone activity by preventing aggregate formation. Furthermore, peptide binding causes the opening of the pocket, release of the GDP and enabling the rebinding of GTP. The release of the peptide is controlled by GTP hydrolysis. Gpn1 interacts directly with some regions of yeast Pol II when binding assays with short Pol II-derived peptides were used. Gpn1 mostly interacted with Pol II residues that were hydrophobic, proposing role of Gpn1 as Pol II chaperone. Many of confirmed Pol II peptide interactions with Gpn1 were within the clamp region formed by Rpb1 and Rpb2 domains. Gpn1 had the same affinity to Pol II peptides regardless of the bound GTP or GDP (Niesser *et al*., 2016)

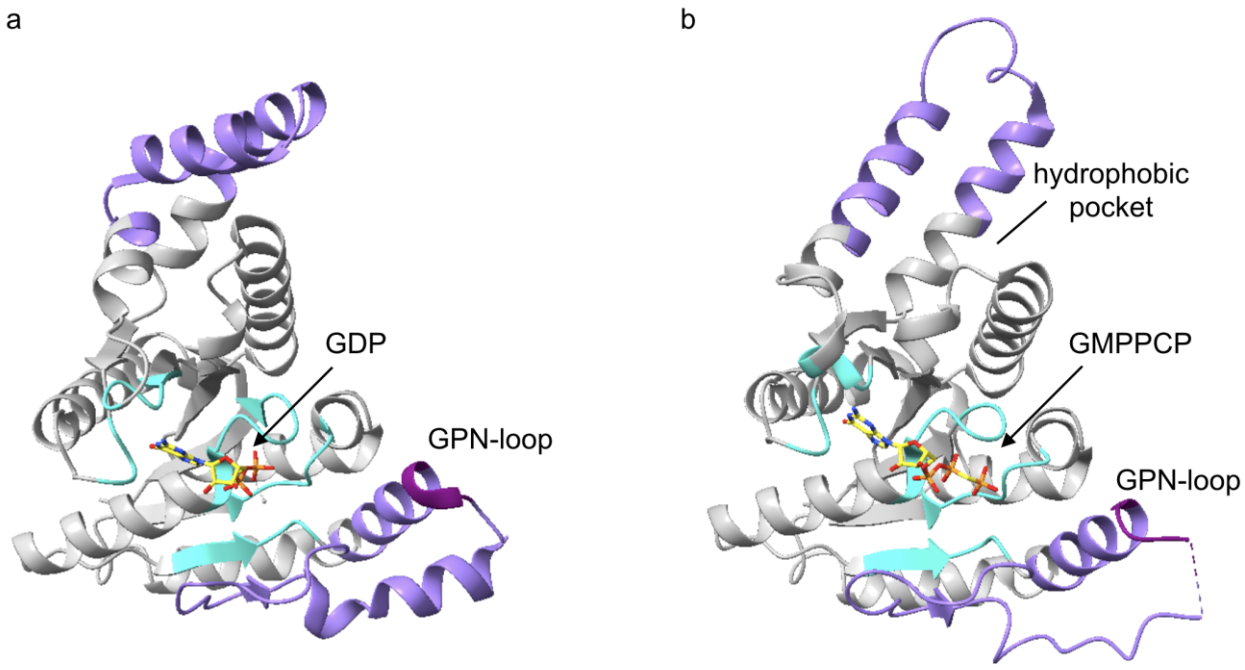


Figure 3: Structure of yeast Gpn1 GTPase core. (**a**) Crystal structure of yeast homolog of GPN1 GTPase core in closed conformation. Bound to GDP – yellow, G motifs – turquoise, insertions – medium purple, GPN-loop - purple. (**b**) Crystal structure of yeast homolog of GPN1 GTPase core in open conformation. Bound to GMPPCP – yellow, G motifs – turquoise, insertions – medium purple, GPN-loop - purple. (**a,b**) Adapted from Niesser *et al*., 2016 (PDB: 5HCI, 5HCN).

1.3.2.2. RPAP proteins

RNA polymerase II associated proteins 1/2/3 (RPAP1/2/3) are also among proteins that were identified to interact with Pol II and participate in Pol II assembly (Jeronimo *et al*., 2007, Boulon *et. al*, 2010, Forget *et al*., 2010).

RPAP2 is preferentially localized in cytoplasm, but shuttles between nucleus and cytoplasm. Similar to Pol II, the nuclear export of RPAP2 is dependent on CRM1/NES pathway (Forget *et al*., 2013). It was previously reported that RPAP2 is a phosphatase of Pol II CTD even though some studies on yeast homolog Rtr1 showed that the protein lacks phosphatase active site (Eglof *et al*., 2012, Xiang *et al*., 2012, Hsu *et al*. 2014). Recent studies did not observe any phosphatase activity *in vitro* when they used Pol II as a substrate (Wang *et al*., 2022). Depletion of RPAP2 leads to accumulation of the largest Pol II subunit in cytoplasm suggesting its involvement in Pol II assembly (Forget *et al*., 2013).

High resolution structure of Pol II-RPAP2 showed binding of RPAP2 to region of Pol II that also binds downstream DNA. Only the N-terminal region of RPAP2 was resolved in high-resolution because the C-terminal region has high flexibility. C-terminal region of RPAP2 is interacting with CTD of Pol II but it is not essential for binding to Pol II (Forget *et al*., 2013, Fianu *et al*., 2021, Wang *et al*., 2022). The RPAP2 modeled region contains an N-terminal domain (NTD, 41-182) and a TFIIF inhibitory (TFIIFi) region. NTD consists of five helices and a zinc finger (Fig. 4). Helices of NTD are interacting with the RPB5 jaw and the two β sheets forming zinc finger motif bind the RPB1 jaw. TFIIFi region binds RPB1 jaw and RPB2 (Fig. 4). Binding of RPAP2 in DNA-downstream region suggests that Pol II is inactive while RPAP2 is bound and cannot transcribe or bind DNA. Full-length RPAP2 cannot be displaced from Pol II

after adding TFs or DNA-RNA scaffold (Wang *et al*., 2022). Interestingly, RPAP2 containing only N-terminal region (1-215) dissociates from Pol II after adding transcription factors to a reaction *in vitro* (Fianu *et al*., 2021). RPAP2 does not distinguish between phosphorylated and unphosphorylated Pol II and interacts with both forms (Wang *et al*., 2022).

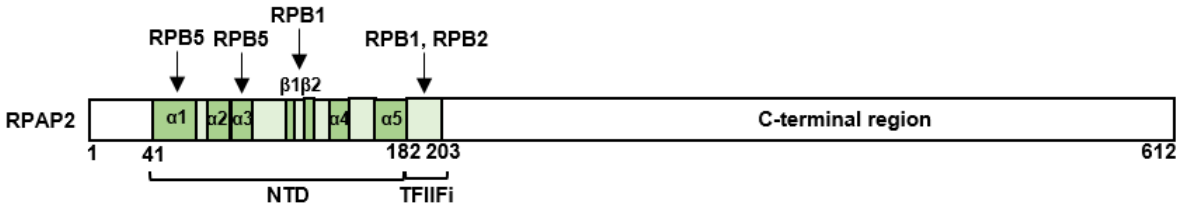


Figure 4: Domains of RPAP2 protein. Resolved regions of human RPAP2 in green. NTD – N-terminal domain, TFIIFi – TFIIF inhibitory region. Binding site of Pol II RPB subunits marked with arrow. Adapted from Wang *et al*., 2022.

It was proposed that RPAP2 mediates a checkpoint before transcription initiation. RPAP2 directly prevents recruitment of TFIIF to Pol II thus formation of PIC. Upon RPAP2 depletion from cells, levels of TFIIF and Pol II at promoters were increased but number of elongating Pol II decreased. While induction of wild-type (wt) RPAP2 rescued this phenotype, RPAP2 mutant lacking the Pol II binding region showed the same levels of TFIIF accumulation at promoters as in RPAP2 depleted cells (Wang *et al*., 2022).

It was reported that GPN1 and RPAP2 proteins interact with each other through the C-terminal domain of RPAP2. When GPN1 was silenced, RPAP2 was trapped in the nucleus leading to the conclusion that GPN1 is important for recycling of RPAP2 after fulfilling its function in assembly of Pol II (Forget *et al*., 2013). Importance of GPN1 in recycling further supports also presence of nuclear export signal (NES, Carre & Shiekhattar, 2011).

Although, the exact role of RPAP1 in assembly of Pol II has not been established yet, it was observed that RPAP1 links Pol II with protein folding chaperones. RPAP1 interacts with RPB3 and RPB11 of Pol II (Jeronimo *et al*., 2007). RPAP1 is a cytoplasmic protein, but treatment of mammalian cells with LMB, entraps RPAP1 in nucleus, similar to the case of RPAP2 mentioned above (Lynch *et al*., 2018).

RPAP3 is a part of RT2P complex and it serves as a linker between HSP90 and R2TP co-chaperone complex but also participates in recruitment of protein subunits that need to be assembled. RPAP3 provides a flexible bridge between HSP90 and the rest of the RT2P complex, which allows HSP90 to interact with various target proteins (Martino *et al*., 2018). As mentioned above, RPAP3 as part of RT2P complex interacts with the RPB1, RPB2 and RPB5 subunits of Pol II and subunits of Pol I, suggesting a more general role of RPAP3 in assembly of RNA polymerases (Boulon *et al*., 2010, Martino *et al*., 2018).

1.3.3. Nuclear transport of RNA polymerases

Once Pol II is assembled in cytoplasm, it needs to be transported into nucleus (Boulon *et al*., 2010). What is the mechanism of nuclear import of Pol II? Fully assembled RNA polymerases have molecular weights ranging from 500 kDa up to 700 kDa.

They are, therefore, too large to pass through the nuclear pore via diffusion, a mechanism that allows the passage of molecules around 90 kDa (Wang & Brattain, 2007). Analyses of the amino acid sequence of all Pol II subunits revealed that there is no nuclear localization signal (NLS) or nuclear export signal (NES). Thus, there must be a protein responsible for transport of assembled RNA polymerases into the nucleus. Iwr1 in *Saccharomyces cerevisiae* was described as a protein facilitating the nuclear import of Pol II and Pol III (Figure 2; Kosugi *et al.*, 2009; Czeko *et al.*, 2011; Minaker *et al.*, 2013). A bipartite NLS located on the N-terminus of Iwr1 directs the nuclear import of Pol II. A NES in Iwr1 is responsible for Iwr1 recycling. Deletion of *S. cerevisiae* Iwr1 leads to aberrant localization of the Rpb1 and Rpb3 subunits. Although loss of Iwr1 in yeast is not lethal, yeast cells get sick and display altered morphology. This phenotype was rescued by full length Iwr1 and by the putative human Iwr1 homolog SCL7A60S (hIWR1; Czeko *et al.*, 2011). The low-resolution structure of yeast Iwr1 bound to Pol II revealed an extra density in Pol II active centre cleft, which consists of Rpb1 and Rpb2. This suggests that Iwr1 can bind to fully assembled Pol II and function as a quality control checkpoint for the biogenesis of Pol II (Czeko *et al.*, 2011). The exact transport of Pol II into the nucleus remains to be yet discovered.

All the findings mentioned above indicate that assembly of Pol II can be another step of transcription regulation in cells.

1.4. GDOWN1 factor role implicated in regulation of transcription and cytoplasmic Pol II

As mentioned above, Pol II is composed of 12 subunits. A distinct form of Pol II has been described (Pol II (G)), which includes the additional protein GDOWN1. GDOWN1 is specific for higher metazoans (Hu *et al.*, 2006). GDOWN1 is an alternatively spliced product of GRINL1A complex transcription unit and is ubiquitously expressed (Roginski *et al.*, 2002; Fagerberg *et al.*, 2014). The protein is associated with Pol II stochiometricaly and binds tightly to Pol II. High-salt washes did not lead to dissociation of GDOWN1 from the Pol II. It has been shown that binding of GDOWN1 to Pol II depends on the GDOWN1 phosphorylation state. Mass-spectrometry analysis of GDOWN1 revealed Ser270 (S270) as a phosphorylation site. If GDOWN1 is phosphorylated, the binding to Pol II is weaker (Guo *et al.*, 2014). Pol II (G) has been estimated to comprise from 30 to 50 % of total Pol II purified from pig and bovine tissues (Hu *et al.*, 2006) and 5 % of purified Pol II from HeLa nuclear extracts (unpublished observations). Knock-down of GDOWN1 has a toxic effect in HeLa cells (Cheng *et al.*, 2012). Concordantly, attempts to obtain stable adult mutants of GDOWN1 in *Drosophila melanogaster* have failed (insertions in gene or using siRNA to knock down), indicating the importance of GDOWN1 in development (Jishage *et al.*, 2018).

1.4.1. Transcription regulation

GDOWN1 has been reported to directly influence transcription. GDOWN1 and Pol II were identified in protein complexes interacting with one of the subunits of Mediator (Sato *et al.*, 2004). Under *in vitro* conditions, Pol II (G) is active in a nonspecific elongation assay using a tailed DNA template, but fails to mediate activator-dependent transcription, unless Mediator is present (Hu *et al.*, 2006). Importantly, it was observed that GDOWN1 represses basal transcription, which suggests a potential general regulatory role for the protein. Studies also showed that GDOWN1 inhibits PIC formation and in this way blocks the initiation of transcription. GDOWN1 prevents binding of TFIIF and TFIIB to Pol II. Under the conditions investigated, only 10-fold excess of

TFIIF could partially dissociate GDOWN1 from Pol II *in vitro*, and 20-fold molar excess of TFIIF was necessary for activating transcription (Cheng *et al.*, 2012; Jishage *et al.*, 2012). A structure of Pol II (G) revealed low-resolution GDOWN1 density near subunits RPB1, RPB2, RPB3 and RPB10. GDOWN1 density overlaps binding sites of TFIIF and TFIIB on Pol II (Jishage *et al.*, 2018). Furthermore, GDOWN1 blocks binding and activity of termination factor TTF2 (Cheng *et al.*, 2012). In the genome, GDOWN1 is present in a majority of the peaks of paused Pol II upon flavopiridol treatment, specifically on actively transcribed genes. The distribution of Pol II(G) was concentrated at the 5' end of the genes (Cheng *et al.*, 2012). These results led to the interpretation of GDOWN1 as a novel transcription factor. The exact role of GDOWN1 in cells, as a biogenesis factor or a transcription factor, is yet to be discovered. It is possible that GDOWN1 can have two distinct functions in regulation of gene expression, during regulation of Pol II in cytoplasm and transcription itself.

1.4.2. Functions in cells

Recently, several studies shed more light about the function of GDOWN1 in cells. Interestingly, they did not observe GDOWN1 associating with paused Pol II or elongation complexes in cells. The ChIP-seq data showed that while Pol II was significantly enriched at the sites of pausing and productive elongation on genes, GDOWN1 signal was enriched only weakly compared to control, mostly at the paused Pol II sites. In addition, depletion of GDOWN1 did not have a significant effect on interphase transcription by Pol II (Ball *et al*., 2022).

Several recent studies also showed that GDOWN1 is predominantly localized in cytoplasm, but it is shuttled to nucleus (Jishage *et al*., 2018, Ball *et al*. 2022, Zhu *et al*., 2022). In the cytoplasm, GDOWN1 co-localized with several subunits of DSIF and NELF complexes as well as Pol II (Zhu *et al*., 2022). Interestingly, GDOWN1 contains two nuclear export signals (NES) between 191-201 and 332-340 amino acids and a cytoplasmic anchoring signal (CAS) at the very C-terminus of the protein. CAS is interacting with nuclear pore complex (NPC) in cytoplasm, anchoring it there (Zhu *et al*., 2022). While full GDOWN1 protein is localized in cytoplasm upon LMB treatment, GDOWN1 with mutated CAS motif is sensitive to LMB and accumulates in nucleus. Aberrant accumulation of GDOWN1 in nucleus causes reduced cell growth and negatively correlates with transcription levels (Zhu *et al*., 2022). It was also observed that GDOWN1 shuttling into nucleus is increased during stress condition, suggesting it has a role during stress adaptation (Zhu *et al*., 2022).

It was proposed that GDOWN1 has role during the mitosis as transcription repressor. GDOWN1 is localized mainly in nucleus during the early steps of mitosis before the nuclear envelope breakdown. It also is heavily phosphorylated during this time that weakens the interaction with Pol II (Ball *et al*., 2022).

1.4.3. Proposed role in Pol II assembly

There are several studies connecting GDOWN1 with assembling Pol II, Pol II assembly factors, or proteins responsible for protein complex formation (Forget *et al.*, 2010; Boulon *et al.*, 2010; Jeronimo *et al.*, 2007). One intriguing possibility is that the function of GDOWN1 may be particularly important in the regulation and assembly of Pol II under stress conditions. The involvement of GDOWN1 under conditions of cellular stress has been previously observed (Jishage *et al.*, 2012). The interaction partners of GDOWN1 in the cytoplasmic fraction have not yet been identified. Thus, identification of GDOWN1 cytoplasmic interacting partners would provide insights into the stage of Pol II regulation at which GDOWN1 acts.

2. Research objectives

RNA polymerases are essential for gene expression and their correct assembly and regulation in cytoplasm is an important step in biogenesis of RNA polymerases. Previous studies have identified proteins interacting with partially assembled RNA polymerases (Boulon *et al.*, 2010, Forget *et al.*, 2010). Several of them were confirmed as assembly factors for Pol II (Carre & Shiekhattar, 2011). However, the exact mechanism how Pol II is regulated in cytoplasm have not been yet understood. Therefore, characterizing the interactions between the Pol II and assembly factors as well as the structure of cytoplasmic Pol II is important to gain more insight in the crucial step of their regulation outside of nucleus.

The goal of this study is to provide molecular and structural insights into human cytoplasmic Pol II regulation. I focus on complexes of GDOWN1. GDOWN1 protein is abundant in cytoplasm and might serve as a cytoplasmic anchor of Pol II. The experiments proposed here can provide new insights of Pol II regulation in cytoplasm.

First aim of the project is to identify interacting partners of assembled Pol II in different cell compartments. These experiments can answer several questions. What are the differences between cytoplasmic and nuclear Pol II before engaging in transcription? Which proteins remain bound to Pol II after shuttling to nucleus? Are all assembly factors of Pol II associating with Pol II in cytoplasm or nucleus at the same time?

Second aim is to understand the molecular mechanisms of Pol II regulation in cytoplasm. Resolving the structure of cytoplasmic Pol II is crucial to characterize how the interacting proteins might be involved in control of Pol II in cytoplasm. The structure is supplemented with cross-linking mass-spectrometry data in case some parts of the complex are not resolved into high resolution.

Third aim of the study is to reconstitute the cytoplasmic Pol II *in vitro*. This approach reduces the complexity of the experiments and allows me to the work with pure proteins. This can improve the structure even further. In addition, I can apply various biochemistry assays to better characterize proteins involved in regulation of cytoplasmic Pol II.

Fourth aim is to address the functional roles of GDOWN1 in cells. Even though recent studies showed promising results, there is still little known about the function of GDOWN1 and Pol II regulation in cytoplasm in the context of living cells. Therefore, *in vivo* observations can provide important input, for example at which stages of Pol II regulation in cytoplasm this protein plays a crucial role.

The proposed study provides more insight into regulation of Pol II in cytoplasm. A structure of cytoplasmic Pol II together with the aforementioned aims is valuable to understand the regulatory pathways that lead towards the active Pol II. Studying cytoplasmic Pol II regulation can have a broader relevance, for example in cell growth regulation or cancer drug research.

3. Results

3.1. Interacting partners of Pol II in cytoplasm and nucleus

In order to better understand how human Pol II is regulated in cytoplasm and which proteins are associating with it in different compartments, a human cell line K562 endogenously expressing GFP-tagged GDOWN1 protein was produced. Successful expression of N-terminal GFP-tagged GDOWN1 and viability of generated K562 cell line were confirmed by collaborator (data not shown). Experimental design of cell line production is shown in Figure 5 a. The correct fractionation of cells was confirmed by Western blot (Fig. 5 b, c).

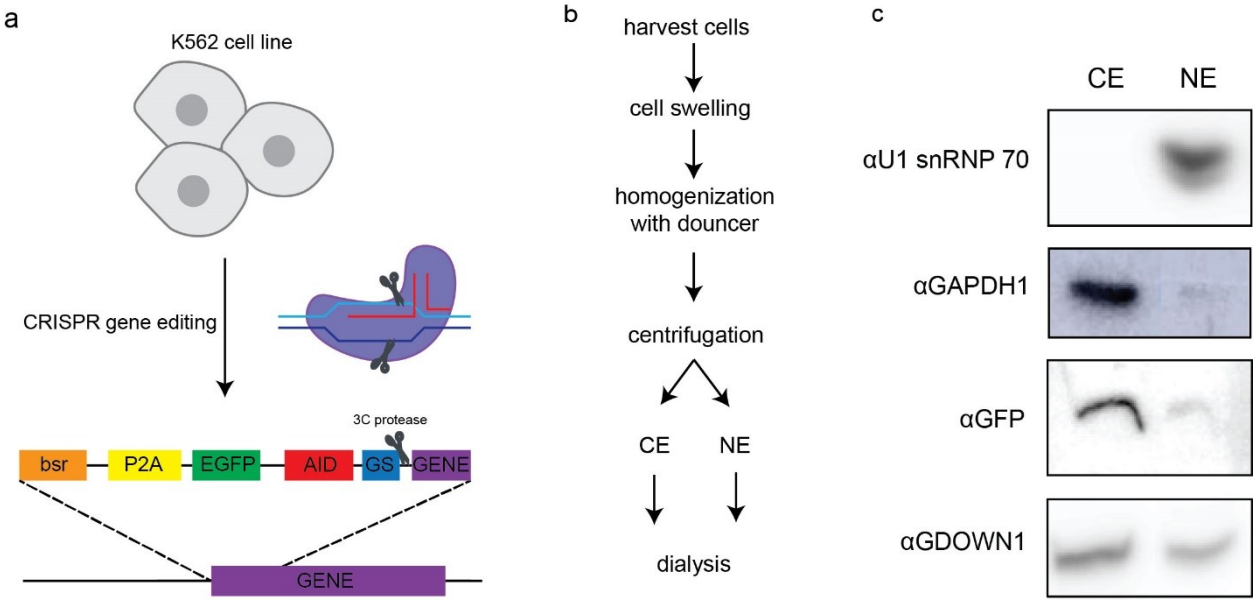


Figure 5: Production of K562 cell line expressing endogenously GFP-tagged GDOWN1 protein. (**a**) Schematic of K562 GFP-GDOWN1 cell line production. The donor construct consists of blasticidin resistance marker (bsr); an intervening porcine teschovirus-1 2A (P2A) which would release the resistance marker during translation; EGFP and auxin-inducible degron (AID) domain separated each by two Gly-Ser (GS) linker; human rhinovirus (3C) protease cleavage site and sequence of first amino acids GDOWN1 (gene). Cell generated by D. Riabov-Bassat (Plaschka lab). (**b**) Schematic of cell separation to cytoplasmic (CE) and nuclear extract (NE). Protocol adapted from Mayeda and Krainer, 1999. (**c**) Correct cell fractionation of K562 GFP-GDOWN1 cells confirmed by Western blot of cytoplasmic and nuclear fractions using antibodies against NE (U1 snRNP70) and cytoplasmic (GAPDH1) markers. Expression of GFP-GDOWN1 confirmed with antibodies against GFP and GDOWN1.

GFP-GDOWN1 pull-downs from cytoplasm and nucleus of K562 cells showed that GDOWN1 associates with all subunits of Pol II enzyme in both compartments (Fig. 6 a, b), meaning all subunits were significantly enriched and had the highest fold change among other identified proteins. Furthermore, I also detected proteins acting as assembly factors of Pol II significantly enriched in the pull-downs (Fig. 6 a, b). Specifically, they were GPN1, GPN2, GPN3, RPAP1, RPAP2 and RPAP3. As mentioned before many of them are essential for correct assembly of Pol II and nuclear shuttling (Carre & Shiekhattar, 2011, Forget *et al.*, 2013). These results suggest there is a stable and strong interaction with Pol II, GDOWN1 and Pol II assembly factors.

Among the proteins that were identified in cytoplasm as well as in nucleus as top hits was RECQL5 helicase (Fig. 6 a, b). RECQL5 is the only helicase from RECQ helicase family directly interacting with Pol II (Aygun *et al.*, 2008). This helicase is

involved in DNA damage repair. It has also been suggested that it inhibits Pol II transcription to prevent colliding with DNA repair machinery, based on low-resolution structure of Pol II and RECQL5, where RECQL5 binds to RPB1 jaw region of Pol II (Kassube *et al.*, 2013).

Other significantly enriched proteins from the pull-downs of GDOWN1 in nucleus were also subunits of Mediator (MED9, MED15, MED 16, MED 20, etc; Fig. 6 a, b). It was previously suggested that GDOWN1 might be a functional interactor of MED (Hu *et al*., 2006). GDOWN1 was also associating in nucleus with components of Integrator complex (INT, Fig. 6). I also identified presence of some general transcription factors (GTF2F1, GTF2F2) in cytoplasmic pull-downs (Fig. 6 a). Among other proteins that were enriched in both compartments was also HSP90, which together with RPAP3 was shown to associate with several Pol II subunits and is involved in Pol II assembly (Boulon *et al*., 2010).

The results from mass-spectrometry suggested a stoichiometric binding of GDOWN1, assembly factors and Pol II in cytoplasm as well as in nucleus. The normalised area shows, that many of the Pol II subunits were represented in similar ratio. Important to mention is that GPN1, GPN3 and RPAP2 were more enriched than other assembly factors (Fig. 6 a, b, Fig. 7 a). I was also interested how big fraction of Pol II is bound to GDOWN1 in both compartments. Western blot analysis of GDOWN1 pull-downs from CE and NE showed that GDOWN1 interacts with RPB1 of Pol II in both cell compartments, supporting the mass-spectrometry data (Fig. 6 a, b). Moreover, all RPB1 subunits in cytoplasm were bound to GDOWN1, while in nucleus there was GDOWN1-free fraction of RPB1 (Fig. 7 b). Even though I tested only presence of RPB1, I could speculate that GDOWN1 interacts with mostly fully assembled Pol II.

With these experiments, I successfully identified interacting partners of GDOWN1 in nucleus and cytoplasm. In both compartments, GDOWN1 associates mainly with Pol II and its assembly factors -GPN and RPAP proteins. This suggests that Pol II with bound assembly factors and GDOWN1 is shuttled from cytoplasm into nucleus and stays intact for some time before the Pol II engages in transcription and associating proteins dissociate.

Since I wanted to better characterize the regulation of Pol II in cytoplasm, I further focused on cytoplasmic Pol II complexes.

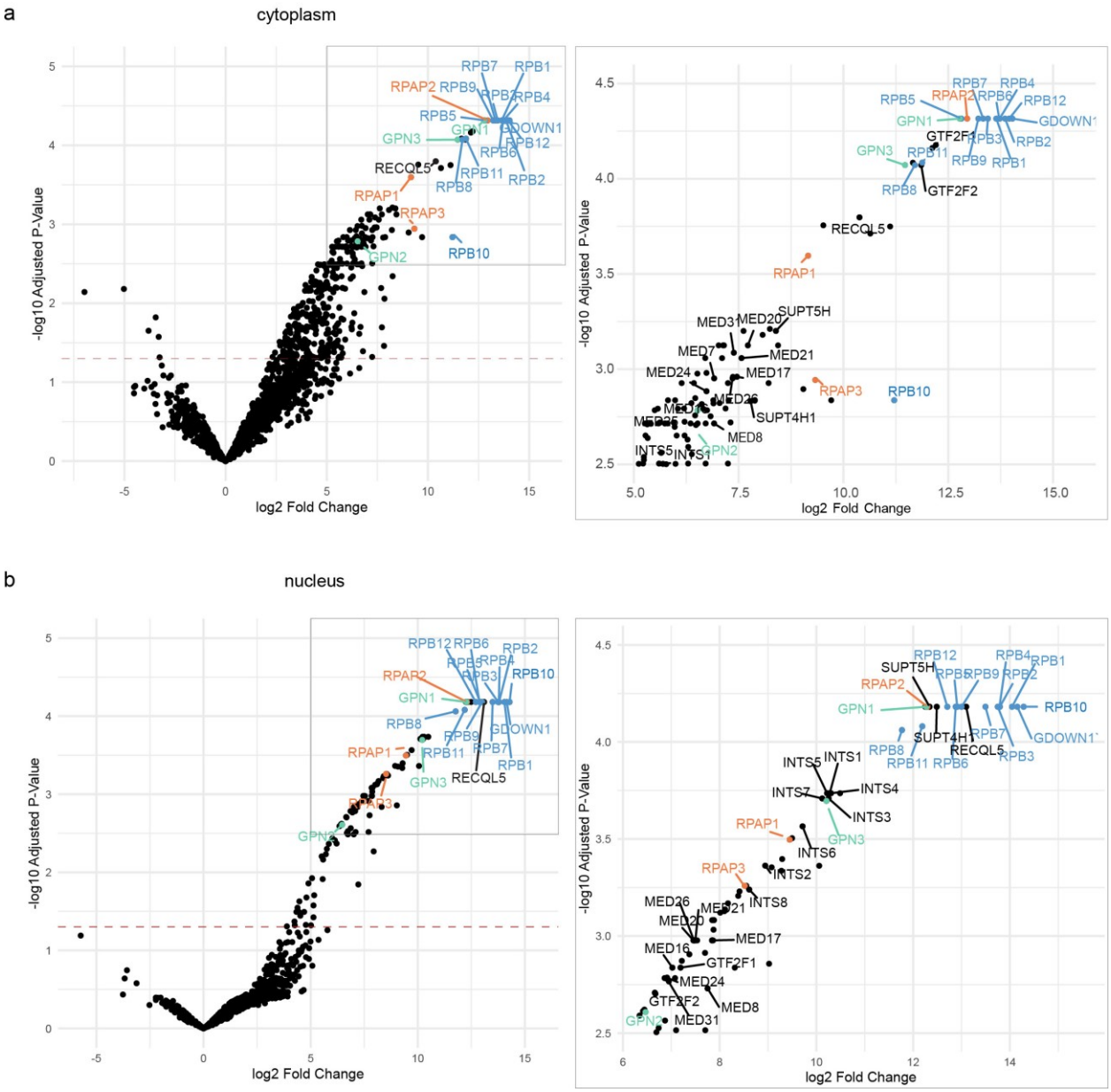


Figure 6: Mass-spectrometry analysis of GDOWN1 pull-downs from cytoplasm and nucleus of K562 cells endogenously expressing GFP-tagged GDOWN1. (**a**, **b**) Volcano plot of proteins identified in pull-downs from cytoplasm (**a**) and nucleus (**b**). Log2 fold change was plotted against -log10 adjusted p-values. Fold change characterizes enrichment of protein from pull-downs compared to controls. Right panels of **a** and **b** are zoom-ins of the squared areas in volcano plots. ($p<0.05$ - red dash). MS analysis performed at VBC Mass Spectrometry Facility.

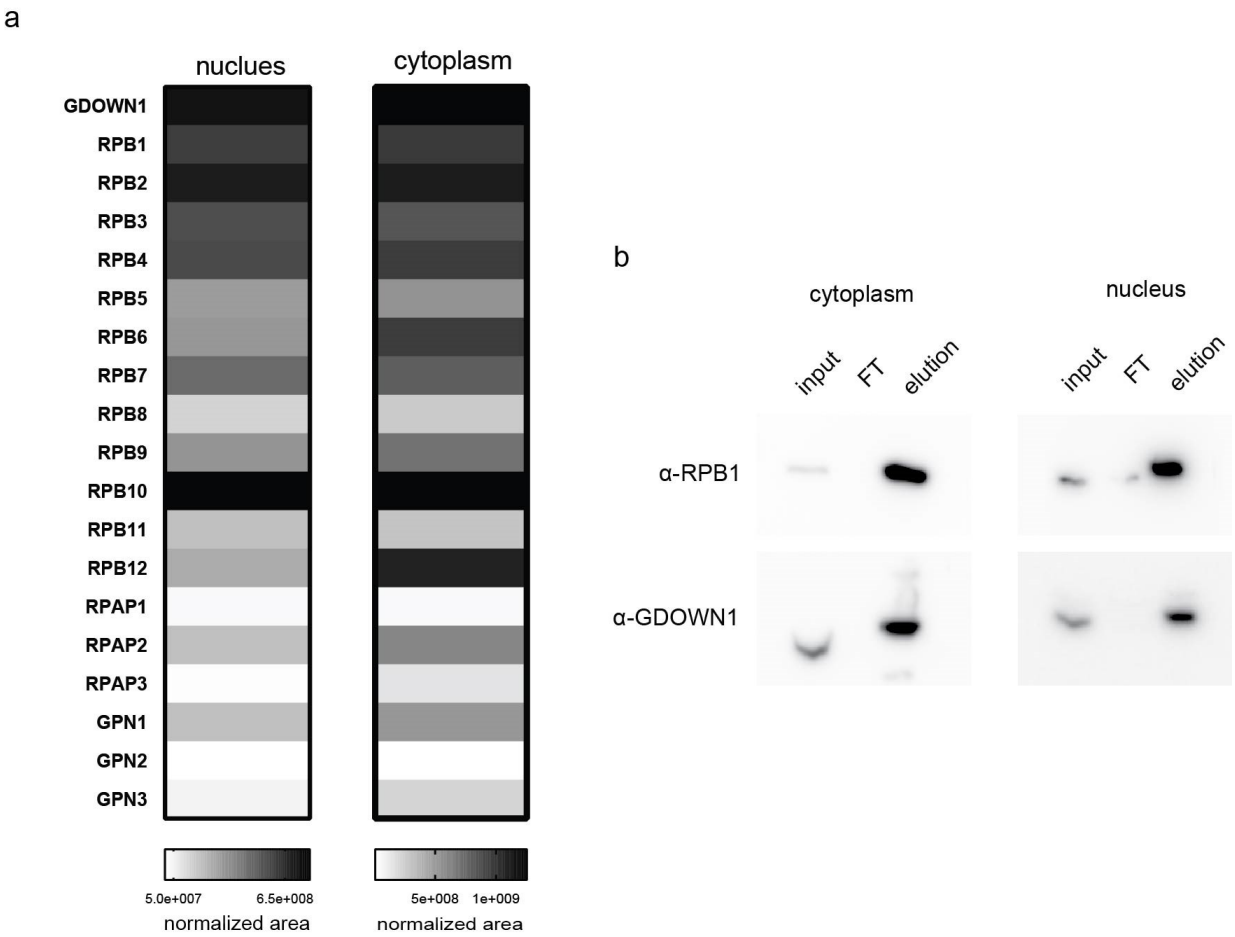


Figure 7: Interaction of GDOWN1 and Pol II in cytoplasm (CE) and nucleus (NE). (**a**) Heat map of normalized areas of idenetified proteins of interest from GDOWN1 pull-downs. (**b**) Western blot of GDOWN1 pull-down from CE and NE. Presence of proteins in input, flow-through (FT) and elution was confirmed by incubating with antibodies against GDOWN1 and CTD of RPB1.

3.2. Characterization and structure of cytoplasmic Pol II complexes

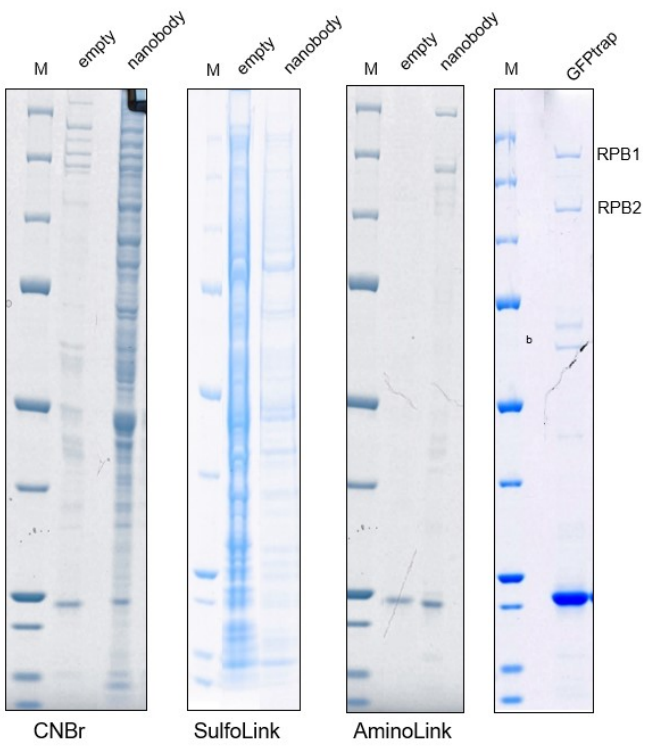
Next step of the project was to determine the structure of cytoplasmic Pol II complex to gain insight into molecular mechanisms of Pol II regulation in cytoplasm. To accomplish this, I had to optimize a homemade GFP-nanobody coupled resin capable of being produced in large volumes while ensuring a high protein yield, as the structural studies required substantial protein amounts.

I successfully purified GFP-nanobody and then tested three different resins that were commercially available. Cyanogen bromide-activated-sepharose (CNBr) resin is commonly used in our lab to prepare Pol II antibody column for Pol II purification. Advantage of this resin is that it has high coupling capacity and could have provided an ideal protein yield, but on the other hand the orientation of coupled proteins is random. Another tested resin was SulfoLink resin. For this purpose, the GFP-nanobody contained cystein linker on the C-terminal end. Coupling to SulfoLink resin is via reduced sulfurhydryl group of the nanobody to iodoacetyl group of the beads, forming a thioeter bond. This would ensure the nanobody is attached to the resin through the cystein linker in one orientation only, which ensures better binding with the protein of interest and should improve the background. The last resin that I tested was AminoLink Plus coupling resin. This resin contains aldehyde functional groups that react with

primary amines of proteins and form a stable bond in presence of sodium cyanoborohydride (Domen *et al.*, 1990).

CNBr and SulfoLink resins performed suboptimally. Both had high background due to unspecific binding and slight leaking of the GFP-nanobody (Fig. 8 a, b). Aminolink resin proved to be the best option after the optimization of all necessary steps and its performance was similar to commercially available GFPtrap resin (ChromoTek, Fig. 8 a). The sumary of the resin performance is in Figure 8 b. RPB1 and RPB2, the two largest subunits of Pol II, are clearly visible on the gel (Fig. 8 a). I decided to continue with AminoLink resin in further experiments. It is important to emphasize that this step of optimization was crutial for the further experiments to successfully purify cytoplasmic Pol II. Thus, I invested a substantial effort and time in these experiments.

a



b

Resin	Notes
SulfoLink	• One-site direction • High background • Leaking of nanobody • Low yield
CNBr	• High coupling capacity • Random orientation of nanobody • High background • Nanobody leaking • Low yield
AminoLink	• Low background • No leaking • Performance comparable to GFPtrap • Random orientation of nanobody

Figure 8: Optimization of home-made GFP-nanobody resin. (**a**) Pull-downs of GFP-GDOWN1 from K562 cells using different resins were eluted with 3C protease (CNBr resin, AminoLink Plus resin, GFPtrap resin) or Gly-HCl (SulfoLink), seperated with SDS-PAGE gel and Coomassie stained. Resins were either coupled with GFP-nanobody (nano) or they were empty. M - Precision Plus Protein™ Unstained Standards (BioRad Laboratories). (**b**) Summary table of tested resin and their performance.

The pull-downs of GDOWN1 from cytoplasm were subjected to sucrose gradient fractionation to identify presence of one or more complexes of Pol II, GDOWN1 and other interacting partners (Fig. 9 a). The fractions were analysed using SDS-PAGE gel separation, Western blot and MS analysis (Fig. 9 b-d). Based on prediction of gradient profile, Pol II bound to GDOWN1 and other assembly factors identified by MS analysis should be present in fractions 16-18. I confirmed the presence of Pol II and GDOWN1 in selected fractions by Western blot using antibodies against RPB1 and GDOWN1 (Fig. 9 c). GDOWN1 was present in first fractions as well as in later ones (Fig 9 b, c). This can be explained by either forming different complexes with

other proteins (or partially assembled Pol II) or the GDOWN1 complexes are dissociating and fall apart during sample handling and long ultracentrifugation (Fig. 9 a). On the other hand, RPB1 is present only in later fractions and co-eluted with GDOWN1 suggesting that these fractions contain the cytoplasmic Pol II (Fig. 9 c). Pooled fractions containing RPB1 and GDOWN1 based on the Western blot were separated on SDS-PAGE gel and the prominent bands were cut out and analysed by MS (Fig. 9 d). The pooled fractions contained GDOWN1, several Pol II subunits and RECQL5, GPN1, GPN2, GPN3 and RPAP2 (Fig. 9 d). This result encouraged me that it is possible to separate the cytoplasmic Pol II with all interacting proteins of interest and in sufficient quantity and I further continued with optimizing cryoEM sample preparation.

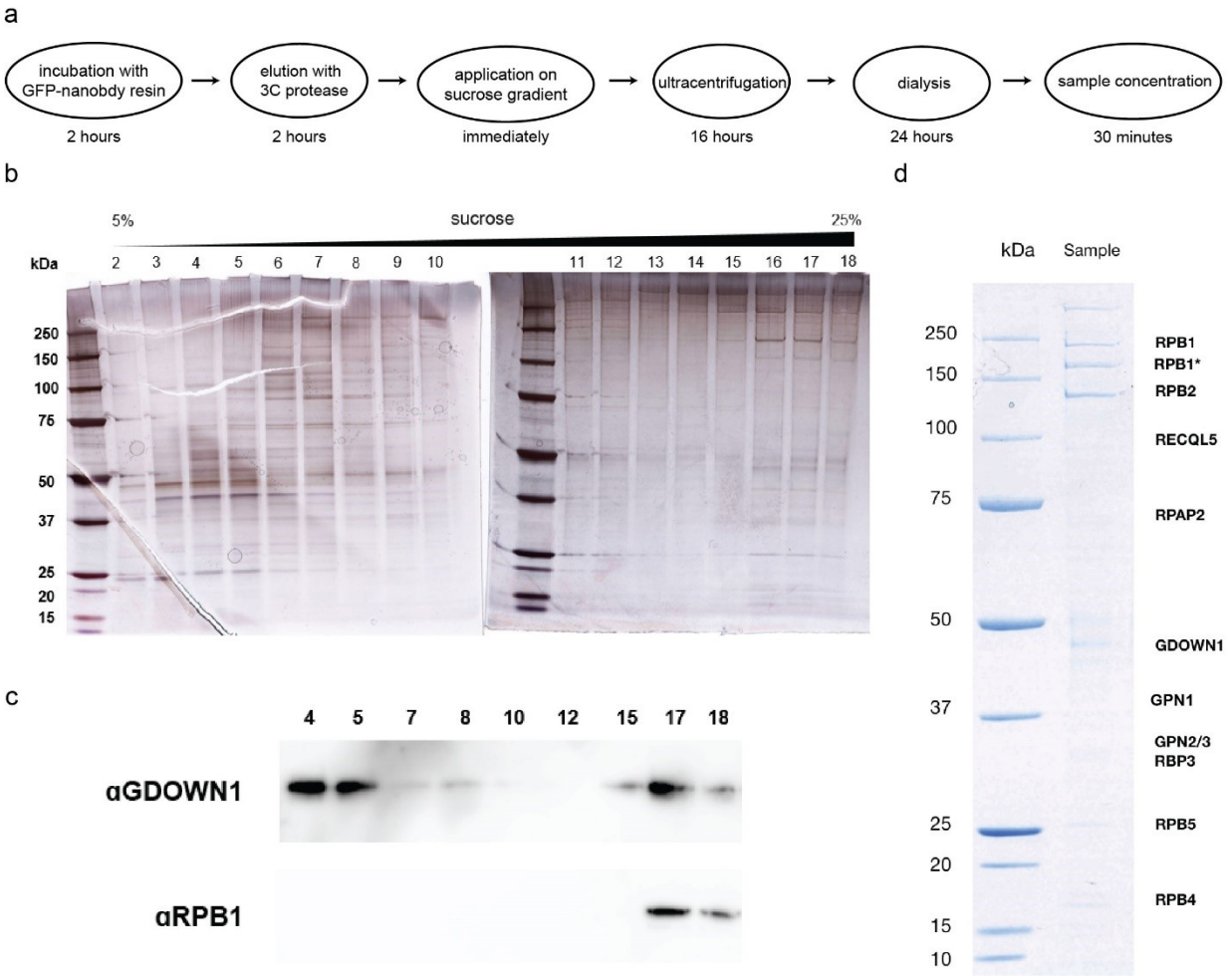


Figure 9: Separation of GDOWN1 complexes from cytoplasm using 5-25% sucrose gradient. (**a**) Schematic of experimental design for sucrose gradient separation of GDOWN1 pull-downs from cytoplasm. CE are first incubated with homemade GFP-nanobody resin for 2 hours. GDOWN1 complexes are eluted with 3C protease for 2 hours. Eluted proteins are immediately applied on sucrose gradient and centrifuged for 16 hours. Samples containing cytoplasmic Pol II complexes are dialysed against buffer with no sucrose for 24 hours and concentrated to smaller volume. (**b**) SDS-PAGE gel of sucrose gradient fractions, stained with silver. (**c**) Western blot of selected fractions from sucrose gradient using antibodies against GDOWN1 and RPB1. (**d**) SDS-PAGE gel of pooled and concentrated gradient fractions containing GDOWN1 and Pol II. Bands were cut out of gel and send for MS analysis. MS experiments performed by VBC MS facility (* RPB1 without CTD) kDa - Precision Plus Protein™ Unstained Standards (BioRad Laboratories).

In order to prepare the samples for structural studies, I first tested the stability of the complex with negative stain and mass photometry. Fractions of sucrose gradient

containing cytoplasmic Pol II were negative stained and imaged with TEM microscope. From initial screening of the grids, I saw that the complex is falling apart on the grids and the particles were heterogeneous in their size (Fig. 10 a). I made the same observation measuring the samples with mass-photometry. There was one prominent peak of 550 kDa; roughly matching the molecular weight of Pol II monomer (Fig. 10 c). For this reason, I decide to implement GraFix step that will cross-link the complexes while centrifuging in the sucrose gradient (Stark & Holger, 2010). This approach stabilized the complex and I observed mostly intact Pol II complex in negative stained grids (Fig. 10 b). In addition, mass photometry of cross-linked complexes showed presence of higher molecular weight peaks that could represent the cytoplasmic Pol II complexes (Fig. 10 d).

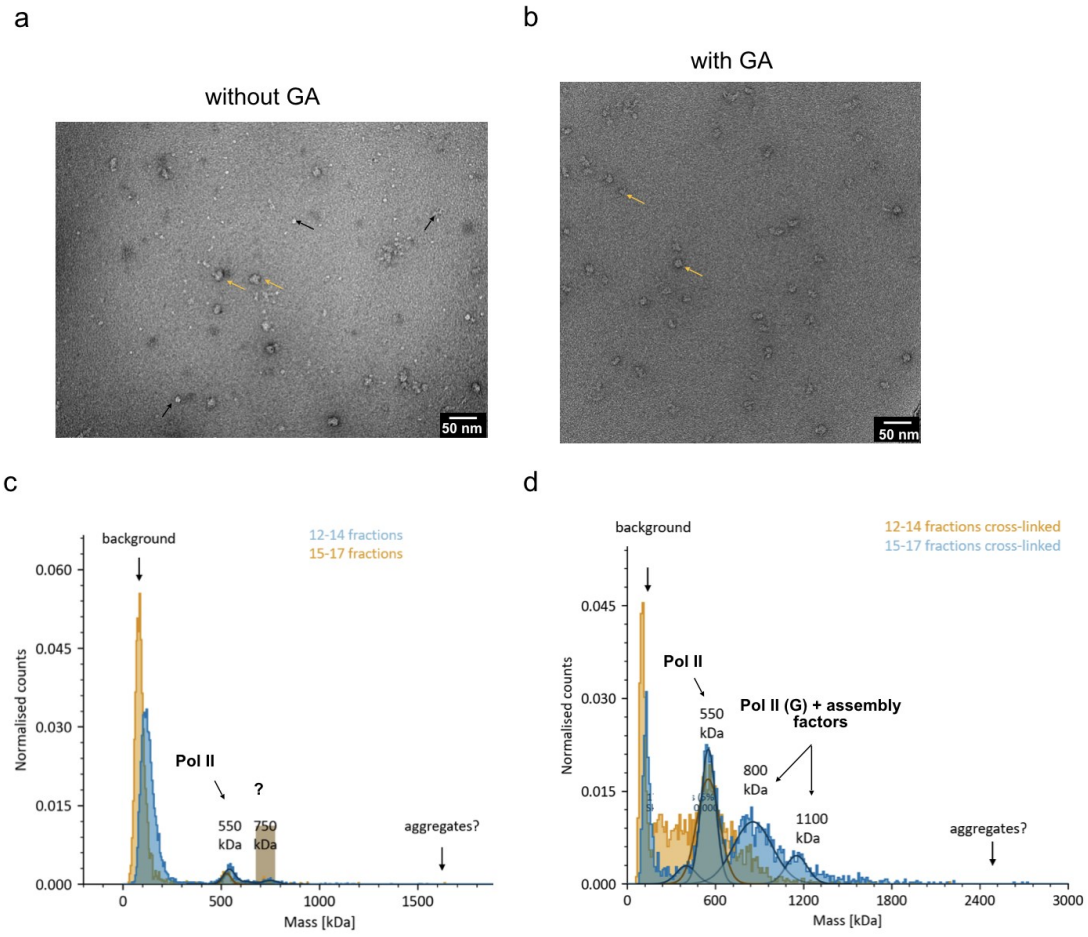


Figure 10: Optimization of grid preparation for electron microscopy using GraFix (Stark & Holger, 2010). (**a, b**) Negative staining of cytoplasmic Pol II separated on sucrose gradient with (**a**) or without glutaraldehyde (**b**). GA – glutarahldehyde, cytoplasmic Pol II particles are marked with yellow arrows, smaller particles of dissociating proteins marked with black arrow. 1,37Å pixel size, scale bar - 50 nm. (**c, d**) Mass photometry of non-cross-linked (**c**) or cross-linked (**d**) cytoplasmic complex separated by sucrose gradient. Data were collected with Refeyn TwoMP and analyzed by Refeyn Discovery^{MP} software. BSA was used to calibrate the instrument.

Since the particle density was low, I also included prolonged incubation of sample with carbon layer prior applying it on the grid (Fig. 11 a). This allowed me to increase incubation time of Pol II complexes with carbon. After screening the grids in TEM microscope, I could see that the particle density as well as their distribution improved and even shorter incubation time was sufficient to obtain an ideal grid quality for negative staining (Fig. 11 b, c, d).

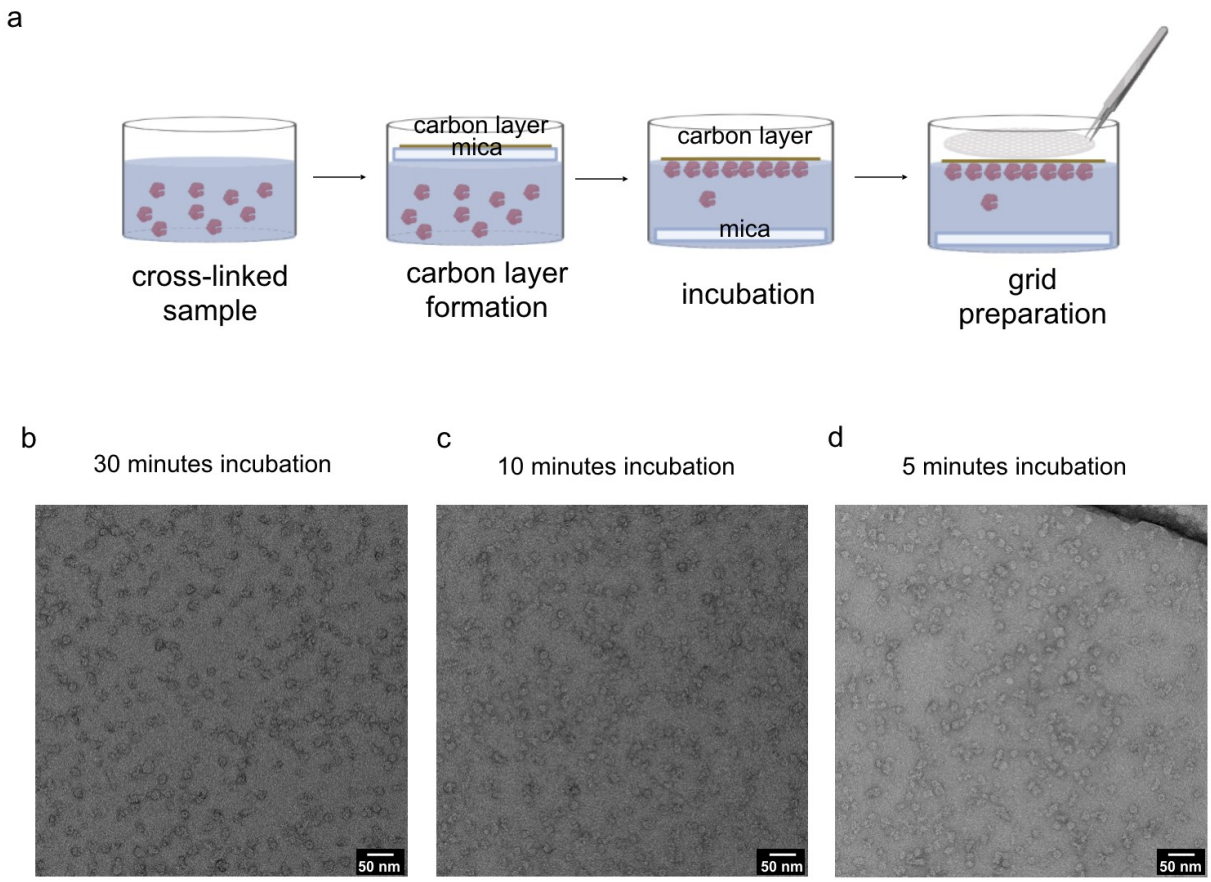


Figure 11: Optimization of grid preparation using carbon as a support layer. (**a**) Schematic of experimental work-flow. Sample in solution is transferred into wells and layered with carbon-coated mica. Mica is separated from carbon and sample is incubated with carbon layer. After the incubation, the grid is placed on top of the carbon and removed again. Carbon with the sample is attached to the grid. (**b-d**) Representative micrographs of negative stained cytoplasmic Pol II incubated with carbon for 30 minutes (**b**), 10 minutes (**c**) and 5 minutes (**d**). 1,37Å pixel size, scale bar - 50 nm.

I collected a small data set of the negative stained grid with the best particle density and distribution (Fig. 12 a). 2D classes of picked particles correspond with size and shape to Pol II (Fig. 12 b). The impurities in the sample represented around 5 % of all picked particles from the micrographs. Selected Pol II 2D classes were used for 3D refinement and resulted in low resolution 3D density. I fitted the density map with the model of transcribing mammalian Pol II (Bernecky *et al.*, 2016; Fig. 12 c). Although the resolution of the structure is low, I observed extra densities when compared to the reported Pol II model (Fig. 12 c). One of these extra densities is present near downstream DNA binding region of the Pol II. This location would be consistent with RPAP2 and RECQL5 densities observed in previous structures. These two proteins bind in this region possibly to prevent interaction with DNA and transcription factors (Kassube *et al.*, 2013; Fianu *et al.*, 2021, Wang *et al*., 2022). There is also an extra density close to stalk/clamp region of Pol II (Fig. 12 c). Based on the low resolution structure I could only speculate if there are GPNs bound to Pol II and a higher resolution structure was needed to better characterize the structure of cytoplasmic Pol II.

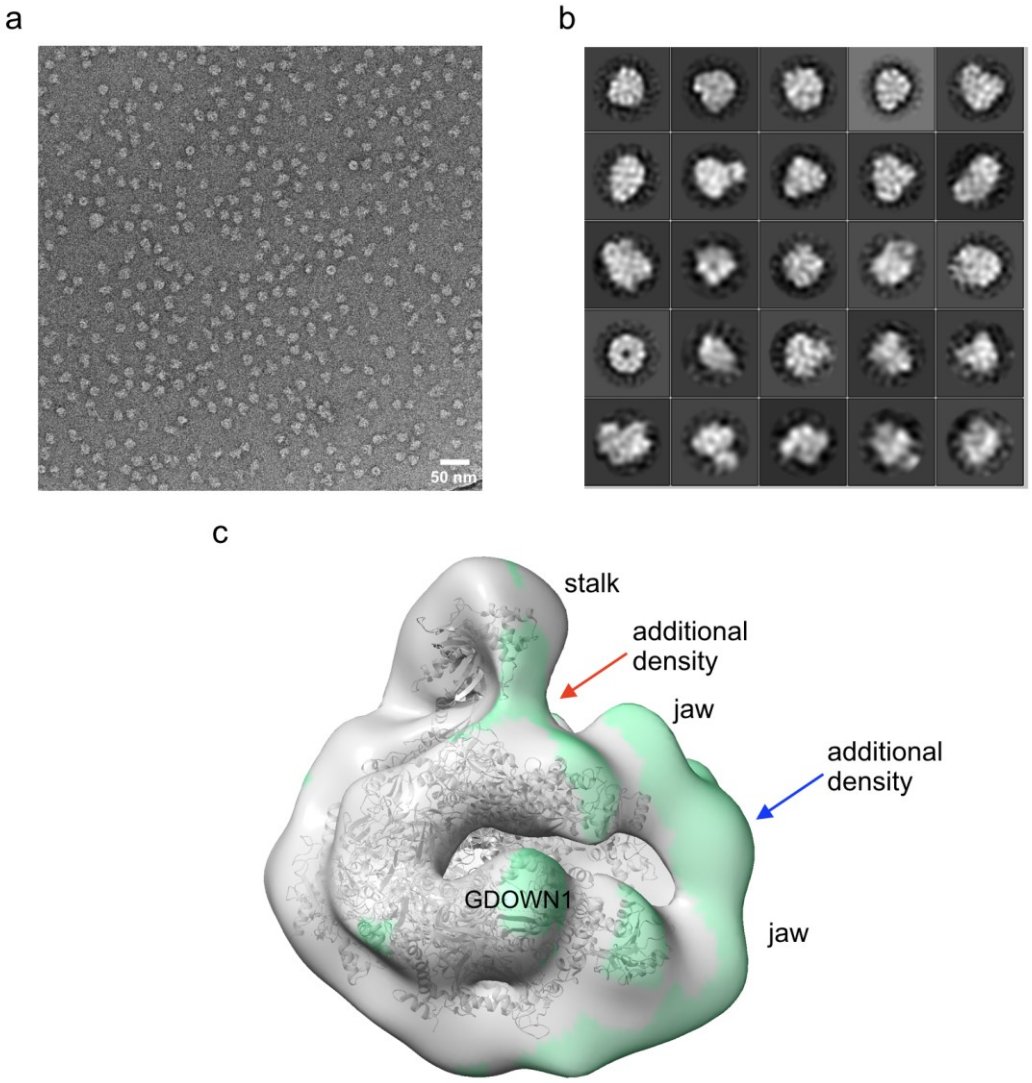


Figure 12: Low resolution density map of cytoplasmic Pol II. (**a**) Example of negative stained micrograph of cytoplasmic Pol II. 1,37Å pixel size, scale bar - 50 nm (**b**) 2D classification of particles picked from collected micrographs. (**c**) 3D refinement of cytoplasmic Pol II complex. Transcribing Pol II structure (grey, PDB:5FLM, Bernecky *et al*., 2016) is placed into density of cytoplasmic Pol II complex (green). Extra densities present near Pol II jaws (blue) and stalk region (red). 2D classification and 3D refinement done in RELION 3.1 (Zivanov *et al*., 2018).

Next steps were to prepare sample for cryoEM. Since the pre-incubation with carbon proved successful for the negative stain, I applied it also for the cryoEM grid preparation. I noticed that even though shorter incubation with carbon in case of negative stained grids worked very nicely, cryoEM grids did not have enough particles for a data collection (Fig. 13). I had had to prolong the incubation to 30 minutes and adjust the blotting conditions. I was successful in preparing grids with good quality of ice thickness, particles density and distribution. Those grids were further used for data collection.

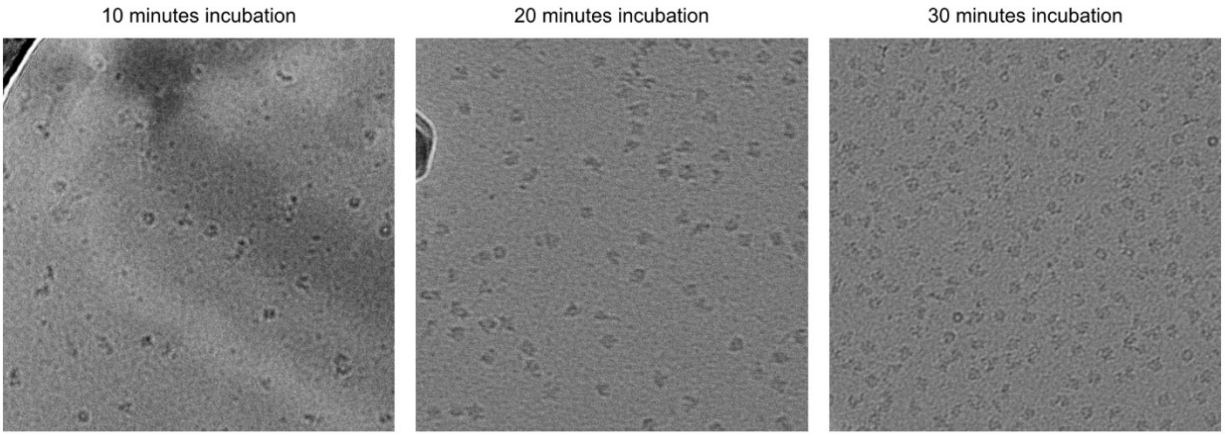


Figure 13: Optimisation of grid preparation for cryoEM. Cytoplasmic Pol II samples were pre-incubated with 2 nm carbon layer for 10, 20 or 30 minutes prior blotting and plunge freezing in liquid ethane. Grids were screened using Thermo Fisher Glacios transmission electron microscope (200 kV), 0.95Å pixel size.

Micrographs of the grids with ideal particle density and distribution were collected using Thermo Fisher Titan Krios G3i transmission electron microscope (300 kV, Fig. 14).

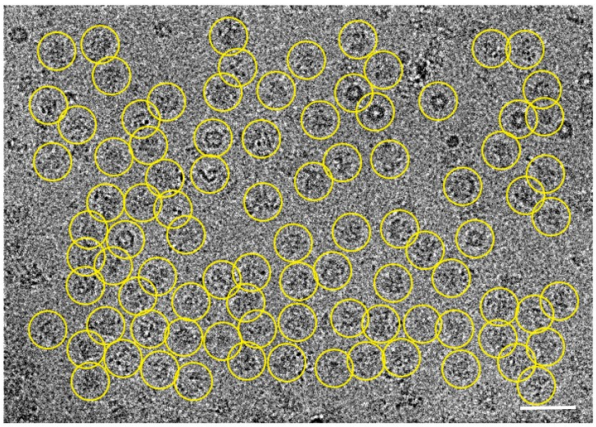


Figure 14: Representative micrograph from Thermo Fisher Titan Krios G3i transmission electron microscope (300 kV). Particles (yellow circle) picked with Warp (Tegunov, 2019). defocus -1.9 µm, 0.65Å pixel size, scale - 50 nm.

2D classification yielded multiple 2D classes that are comparable to 2D classes of transcribing Pol II (Fig. 15 a). Interestingly, they look slightly different to transcribing Pol II, mainly in the stalk/clamp region where there was a fuzzy region present, suggesting an extra density (Fig. 15 a). Since the sample used for grid preparation was heterogeneous even after gradient separation, I observed presence of other 2D classes that were not Pol II. Most prominent 2D classes were the proteasome and AAA+ ATPase that are present in the cytoplasmic pull-downs (Fig. 15 b). They are comparable to already published 2D classifications of 20S proteasome and p97 AAA+ ATPase (Fig. 15 b). These impurities represented only 5 % of the all particles.

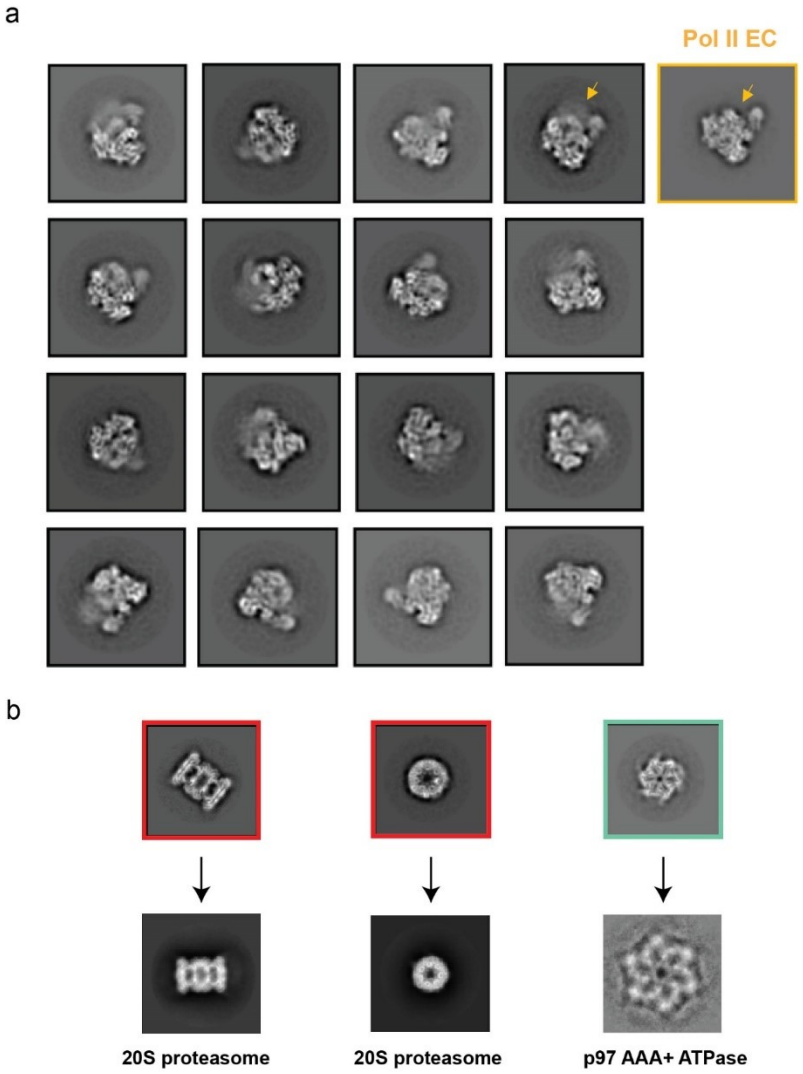


Figure 15: 2D classification of picked cytoplasmic Pol II particles. (**a**) 2D classes of cytoplasmic Pol II from collected dataset. For comparison Pol II EC 2D class is marked in yellow and differences between cytoplasmic and Pol II EC 2D class marked with arrow. Adapted from Bernecky *et al*., 2016 (**b**) 2D classes of contaminants present in our cytoplasmic Pol II sample. In red - 2D classes of proteasome compared to 2D class of proteasome 20S (Guan *et al*., 2020), in turquoise - 2D classes of AAA+ ATPase compared to p97 AAA+ ATPase (Banarjee *et al*., 2016).

After successful 2D classification and 3D refinement, a post-processed map was generated (Fig. 16 a). I managed to produce a structure with nominal resolution 2.72 Å. I built the structural model of cytoplasmic Pol II by combining already published structure of *in vitro* reconstituted human Pol II-RPAP2 complex and AlfphaFold3 (AF3) predicted Pol II (G) complex without clamp region (Wang *et al*., 2022, Abramson *et al*., 2024). Structural model of cytoplasmic Pol II fits the overall Pol II density extra densities corresponding to RPAP2 and GDOWN1 proteins (Fig. 16 a). The clamp region of Pol II was not resolved to high resolution in my structure and was partially visible only when I filtered the map to lower resolution (Fig. 16 b).

RPAP2 density is present in the DNA down-stream region of Pol II (Fig. 16 a). Similar to already published *in vitro* structures of Pol II-RPAP2 complex, only N-terminal part of RPAP2 bound to the core of the Pol II was resolved (Fig 16 a; Fianu *et al*., 2021, Wang *et al*., 2022). Altogether RPAP2 has 69 kDa molecular weight (612 amino acids), but only region between 1-215 amino acids was resolved (Fig. 16 a). RPAP2

has a long C-terminal domain that was shown to bind CTD domain of RPB1 on Pol II. This part of Pol II is very flexible and cannot be resolved in my cryoEM structure.

Another prominent extra density on the Pol II is GDOWN1 (Fig. 16 a). In previously published *in vitro* structure, only a small part of GDOWN1 helix was resolved (Jishage *et al.*, 2018). My structure revealed this helix is extending further to the space above the RPB2 towards the clamp region (Fig. 16 a). Upon closer look, I see a nice fit of the structural model of cytoplasmic Pol II with the GDOWN1 density (Fig. 19 a). I named the modelled helices of GDOWN1 as α1 (21-42), α2 (49-74), α3 (59-63) and α4 (306-335, Fig. 19 a). Helix α4 and α3 binds to RPB2 subunit (Fig. 16 a, Fig. 19 a). Helices $\alpha 1\text{-}2_{GDOWN1}$ were not resolved in density, but they are predicted to bind close to RPB10 and RPB3 subunits (Fig. 19 a).

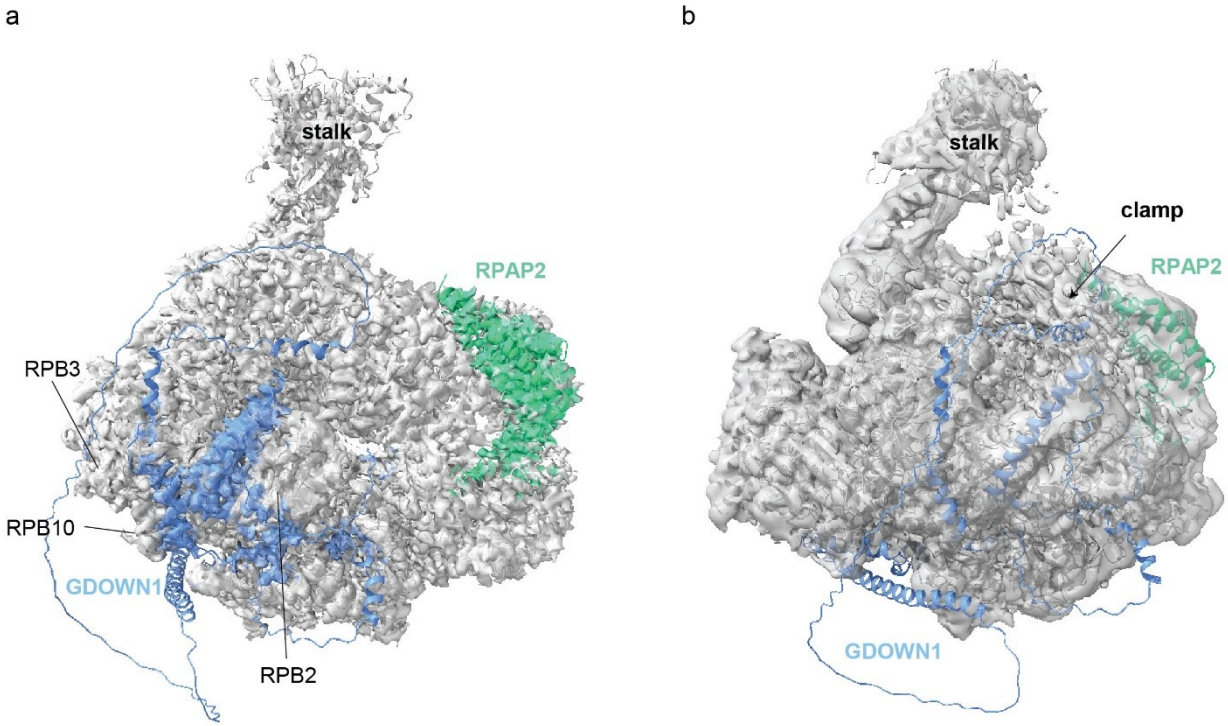


Figure 16: Overall structure of endogenous cytoplasmic Pol II. (**a**) A cryoEM map of cytoplasmic Pol II fitted with structural model of cytoplasmic Pol II. Structural model of cytoplasmic Pol II was generated combining published Pol II-RPAP2 model (PDB: 7F4G, Wang *et al*., 2022) and AlfphaFold3 (AF3) prediction of Pol II (G) complex without clamp region. Pol II in grey, GDOWN1 in light blue and RPAP2 in green. (**b**) CryoEM density map of cytoplasmic Pol II filtered to 5 Å resolution. Cytoplasmic Pol II structural model fitted into the density. Pol II in grey, GDOWN1 in blue, RPAP2 in green.

Angular distribution of particle views showed there is slight preferred orientation, but all views are represented (Fig. 17). This is due to usage of carbon support layer that causes preferred orientation of Pol II in general.

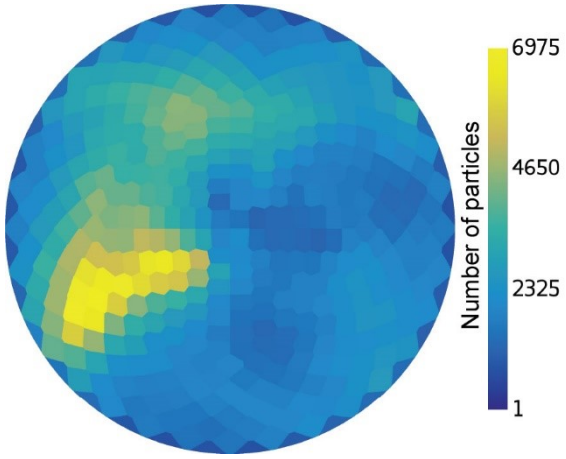


Figure 17: Angular distribution plot of particles for the cytoplasmic Pol II. Specific coloured based on the particle number representing them. Color key shows number of particles corresponding to specific color. Generated by Warp (Tegunov *et al*., 2018).

Local resolution maps generated by Occupy and cryoSPARC show that core of the cytoplasmic Pol II is resolved to higher resolution but not the stalk (Fig. 18 a, b; Punjani *et al*., 2017, Forsberg *et al*., 2023). RPAP2 density is resolved to lower resolution. Helices α3-4_{GDOWN1} show also lower resolution compared to core of Pol II (Fig. 18 a, b). Important to mention is that Occupy generated scale does not show the local resolution values, but estimates the relative contrast degradation within the map. If the resolution is high then the contrast between pixels is high as well. In poorly-resolved regions the contrast between the pixel values is low (Forsberg *et al*., 2023).

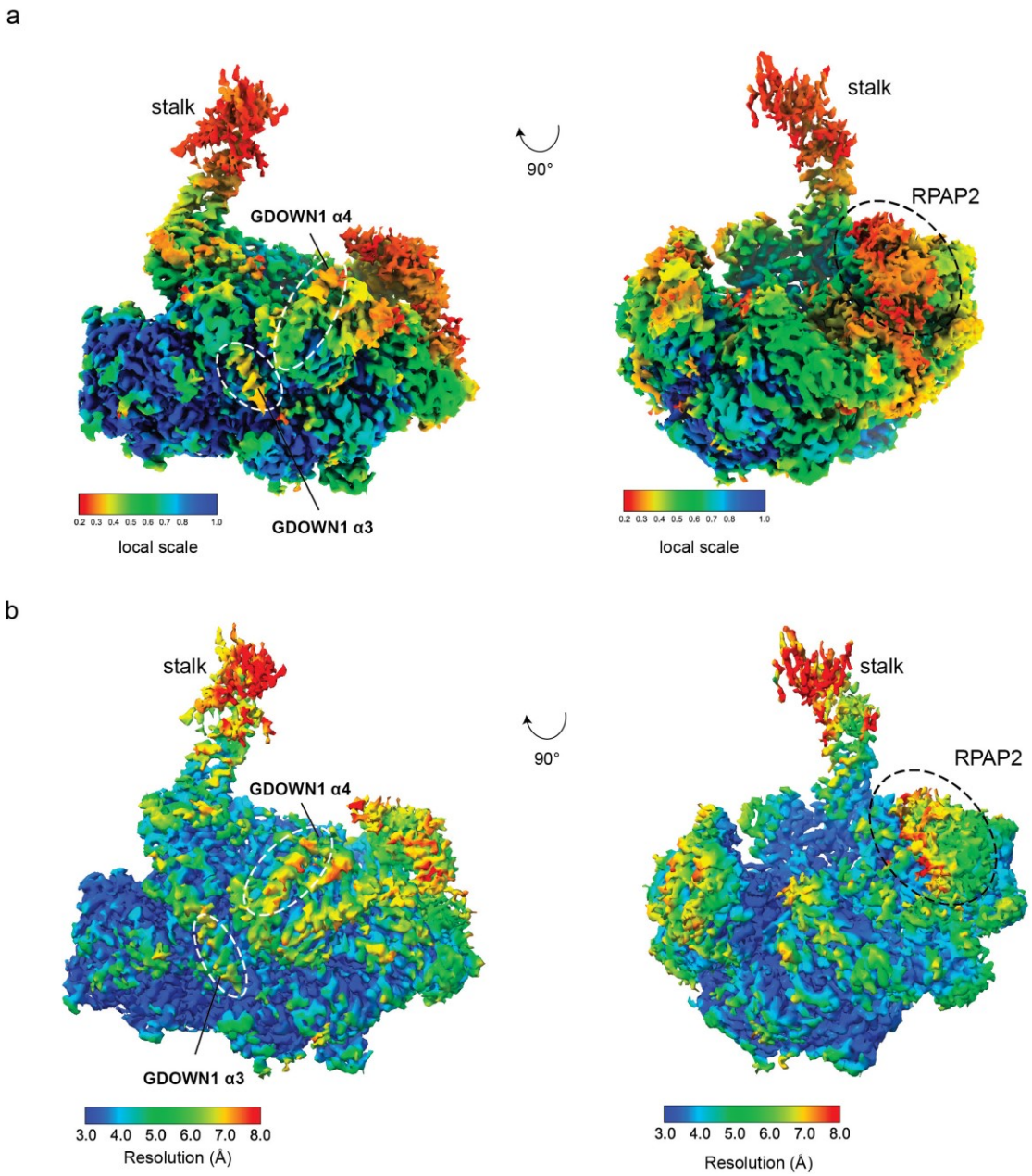


Figure 18: Local resolution of density map of cytoplasmic Pol II. (**a**) Cytoplasmic Pol II density map coloured according to local scale generated by Occupy. 1 - highest relative resolution, 0 - lowest relative resolution (Forsberg *et al*., 2023). (**b**) Local resolution map of cytoplasmic Pol II generated by cryoSPARC (Punjani *et al*., 2017). Stalk, GDOWN1 helices and RPAP2 marked.

For AF3 prediction of GDOWN1, I used protein sequences of GDOWN1 and Pol II subunits without clamp (subunits RPB1 361-1487, RPB2 1-1061, RPB3, RPB4, RPB5, RPB6 46-127, RPB7, RPB8, RPB9, RPB10, RPB11, RPB12). The prediction had ipTM and pTM scores 0.83 and 0.86 respectively. The individual pIDDT scores of GDOWN1 show that the helices α3 and α4 matching my density were predicted with high confidence (Fig. 19 b). Helix α1 and α2 have a bit lower confidence score, but it is still considered high (Fig. 19 b). The important regions of GDOWN1 like the two NES, CAS or S270 residue are mostly in the unstructured parts of the GDOWN1 prediction, or in parts with low confidence score. The only region that is present in my density corresponds to helix α4 is part of NES 332-340 (Fig. 19 b).

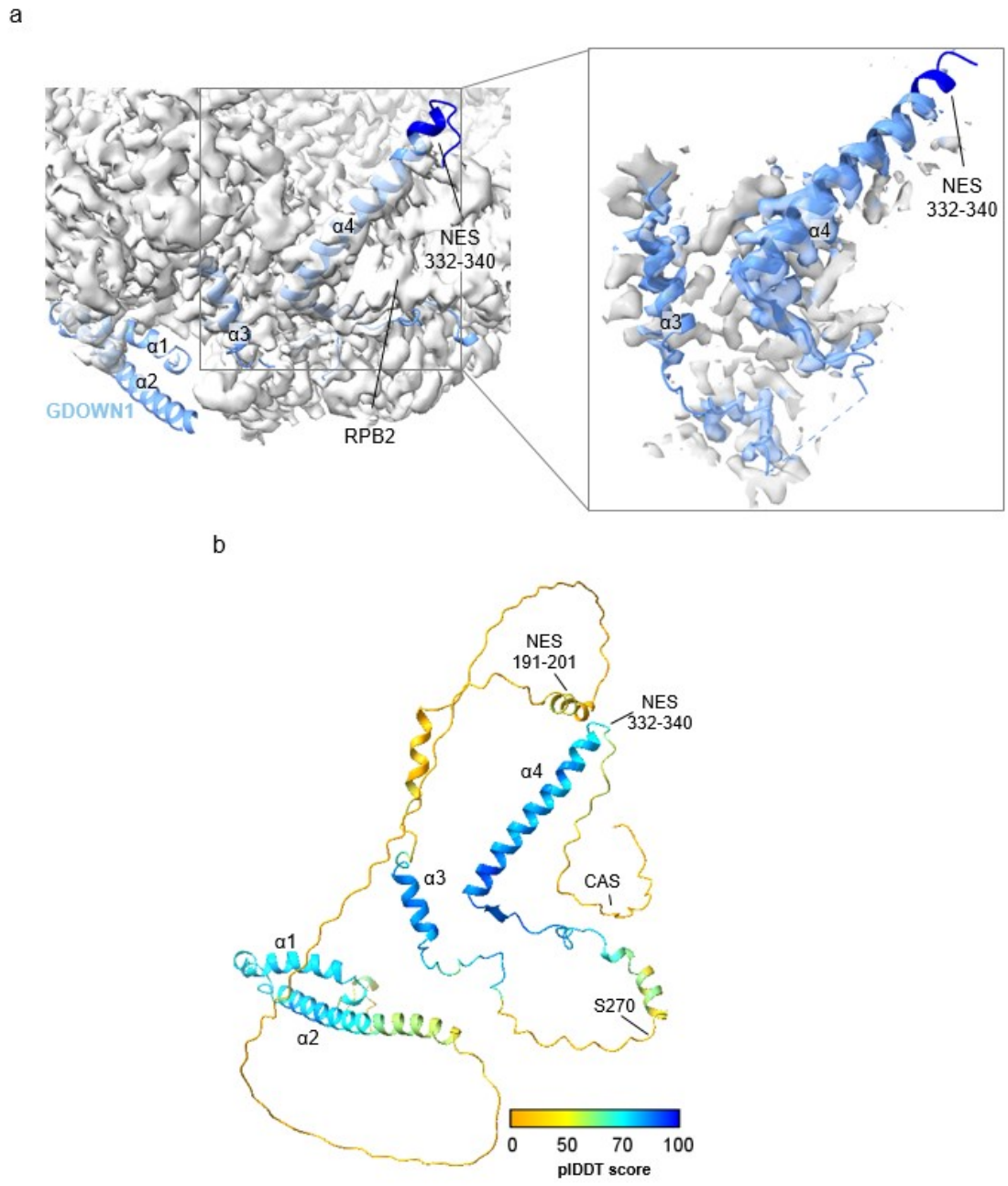


Figure 19: Characterization of GDOWN1 structural model. (**a**) Zoom in of cytoplasmic Pol II density map with fitted GDOWN1 structural model showing only residues with pIDDT score higher than 50. In grey cytoplasmic Pol II density, in light blue GDOWN1. (**b**) GDOWN1 structural model predicted by AF3 colored by the pIDDT scores generated. NES 191-201, NES 332-340, CAS and S270 are labelled. Visualized with ChimeraX (Pettersen, 2021).

In order to improve the resolution of RPAP2 and GDOWN1 a local 3D classification using masks over RPAP2 and GDOWN1 with RPB3 regions was performed. Particles from 3D classes with well-resolved the RPAP2 or GDOWN1 were further used for 3D refinement. In both cases, there was an improvement in resolving the features of RPAP2 and GDOWN1 (Fig. 20 a, b). RPAP2 density overlaps better with RPAP2 helices from the structural model (Fig. 20 a).

A big improvement is also in case of GDOWN1. In previous density map of cytoplasmic Pol II, I could not resolve the density for helix α1 and α2 that should be close to RPB3 and RPB10 subunits of Pol II (Fig. 16 a, 19 a). 3D classification of this

region revealed one class that had densities of helices α1-2$_{GDOWN1}$. Further 3D refinement resulted in resolved regions of those helices as well. The predicted GDOWN1 structure fits the densities for the helix α1 and α2 (Fig. 20 b).

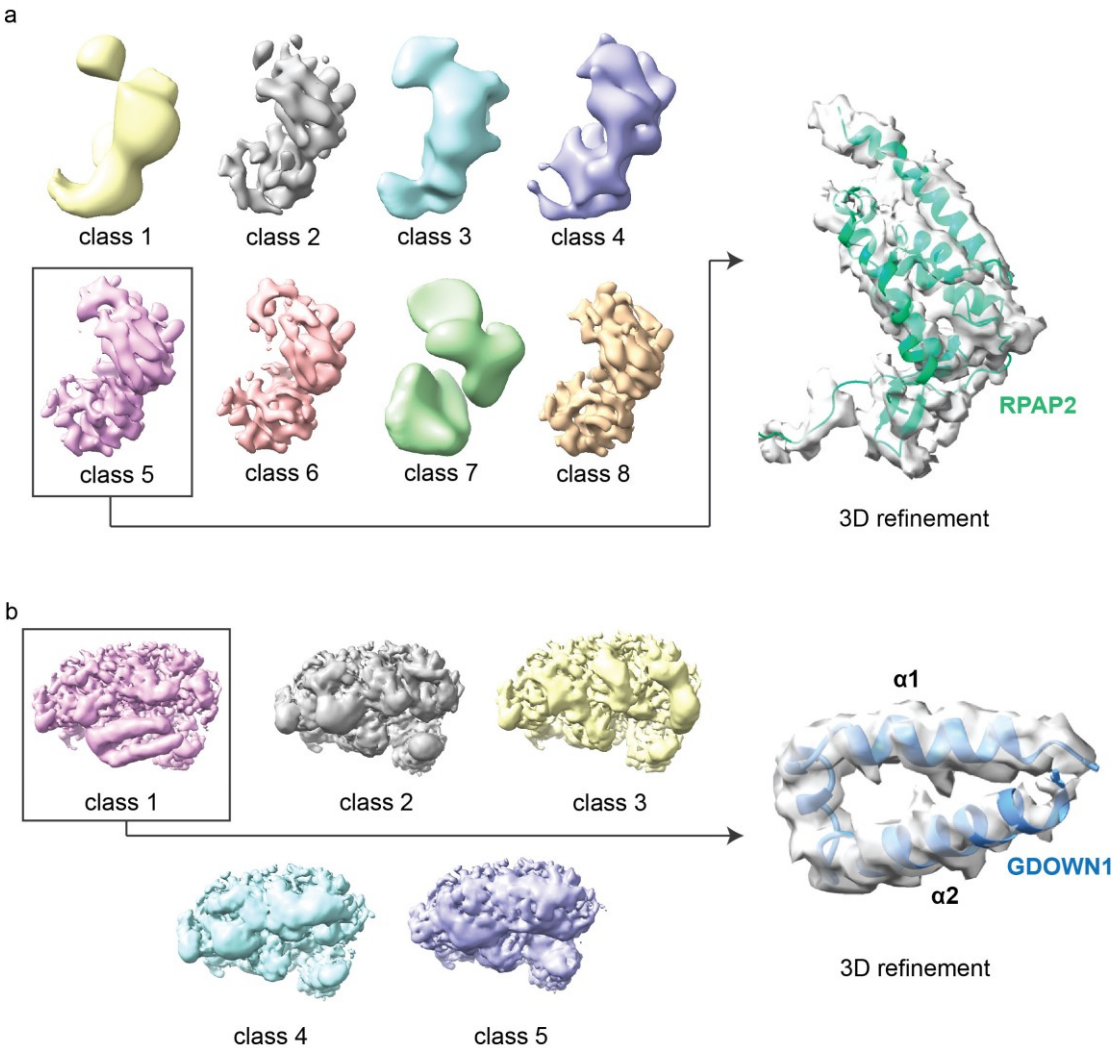


Figure 20: 3D classification of RPAP2 and GDOWN1 regions of cytoplasmic Pol II. (**a**) 3D classes with RPAP2 mask and the 3D refinement of the best resolved RPAP2 is showed. RPAP2 structural model (in green) fit into the RPAP2 density (in grey). (**b**) 3D classes with GDOWN1 and RPB3 region mask and the 3D refinement of the best resolved GDOWN1 helices α1 and α2 is showed. GDOWN1 structural model (in blue) fit into the GDOWN1 density (in grey).

I also wanted to supplement the cryoEM structure further; more specifically the unresolved parts and I decided to perform a cross-linking mass-spectrometry (XL-MS). In order to do that I first needed to optimize the cross-linking conditions. I tested multiple cross-linkers that were available. First, I tested BS3 cross-linker and optimized the ideal concentration on pure Pol II (Fig. 21 a). It showed promising results and the cross-linked cytoplasmic Pol II complexes were separated on SDS-PAGE gel, cut out and send for XL-MS analysis. Unfortunately, the XL-MS identified only very few cross-links (data not shown).

I decided to test also DSSO and DSBU cleavable cross-linkers. The advantage of cleavable cross-linkers is that they improve the identification of the cross-linked peptides. Since the XL-MS identifies not only cross-linked but also non-cross-linked peptides, use of cleavable cross-linkers decrease the complexity of the sample. They

are cleaved during the analysis creating specific ions that are later use to confidently identify the cross-linked peptides. This reduces occurrence of false positive cross-links identified (Kao *et al*., 2011). Similarly as with BS3, I first optimized cross-linking reaction on pure Pol II. I observed that DSBU at lower concentrations is inefficient to fully crosslink Pol II while at higher concentrations it causes protein complex aggregation (Fig. 21 b). DSSO cross-linker proved to be better for the optimization of the conditions to obtain cross-linked Pol II complexes (Fig. 21 c). I prepared the cytoplasmic Pol II sample with the concertation of DSSO that worked the best and submitted to XL-MS analysis.

There were very few cross-links identified in the cytoplasmic Pol II cross-linked with DSSO (Fig. 21 d). They were in agreement with published literature, but they did not supplemented my structure further. Interestingly, I identified a cross-link between GDOWN1 and DOCK7 (Fig. 21 d). DOCK7 belongs to DOCK family of guanine nucleotide exchange factors (GEFs) that are responsible for activating small GTPases (Watabe-Uchida *et al*., 2006, Gao *et al*. 2021). Initially, I assumed that this protein could be involved in regulation of GPNs in cytoplasm. However, there was no DOCK7 identified in the MS results of the GDOWN1 pull-downs (Figure 6). Upon further inspection of MS results from GDOWN1 pull-downs, the short peptide sequence that was assigned to DOCK7 was also found in PDCD11 protein. PDCD11 was identified as resin background. Thus, identification of DOCK7 in XL-MS of cytoplasmic Pol II is likely a false positive result. Even though identification of DOCK7 in the XL-MS was exciting, due to this observation I cannot assume it is indeed taking part in GPNs nucleotide exchange.

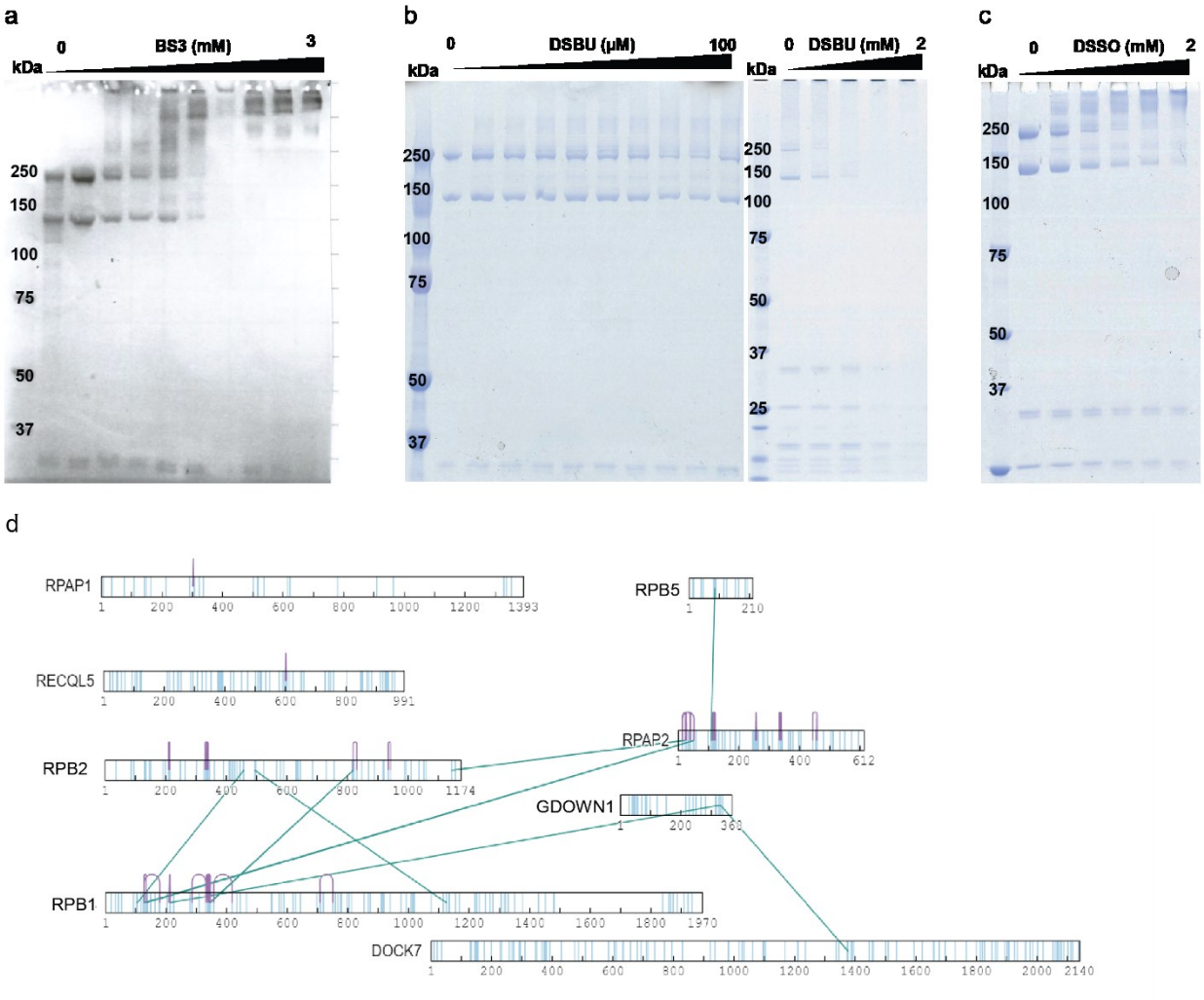


Figure 21: Cross-linking mass-spectrometry of cytoplasmic Pol II. (**a-c**) Optimization of cross-linking using different cross-linkers. Pure Pol II was used for optimization of ideal concentration of BS3 (**a**), DSBU (**b**) and DSSO (**c**). Complexes were separated on SDS-PAGE gel and Coomassie stained. kDa - Precision Plus Protein™ Unstained Standards (BioRad Laboratories). (**d**) Schematic of identified cross-linked peptides from cytoplasmic sample cross-linked with 2 mM DSSO. It shows inter- (green) and intra-molecular (violet) cross-linked peptides. Cross-links visualized using xiNET (Combe *et al*., 2015). XL-MS performed in Max Perutz Vienna MS facility.

I then started a collaboration with MS focused research group that had previous experience with challenging endogenous samples as mine. The cytoplasmic Pol II was directly cross-linked with DSSO on beads. This approach improved the results and there were more cross-links identified, mainly between the Pol II subunits (Fig. 22 a). Unfortunately, there were no cross-links identified between GPNs and other proteins, thus I did not obtain more insight where those proteins might bind the Pol II. Nevertheless, this experiment showed that the complex is in agreement with my structure. I can also notice that the Cα-Cα distances are within the limits of 25 Å, thus the identified cross-links correspond with the structure of cytoplasmic Pol II (Fig. 22 b, c).

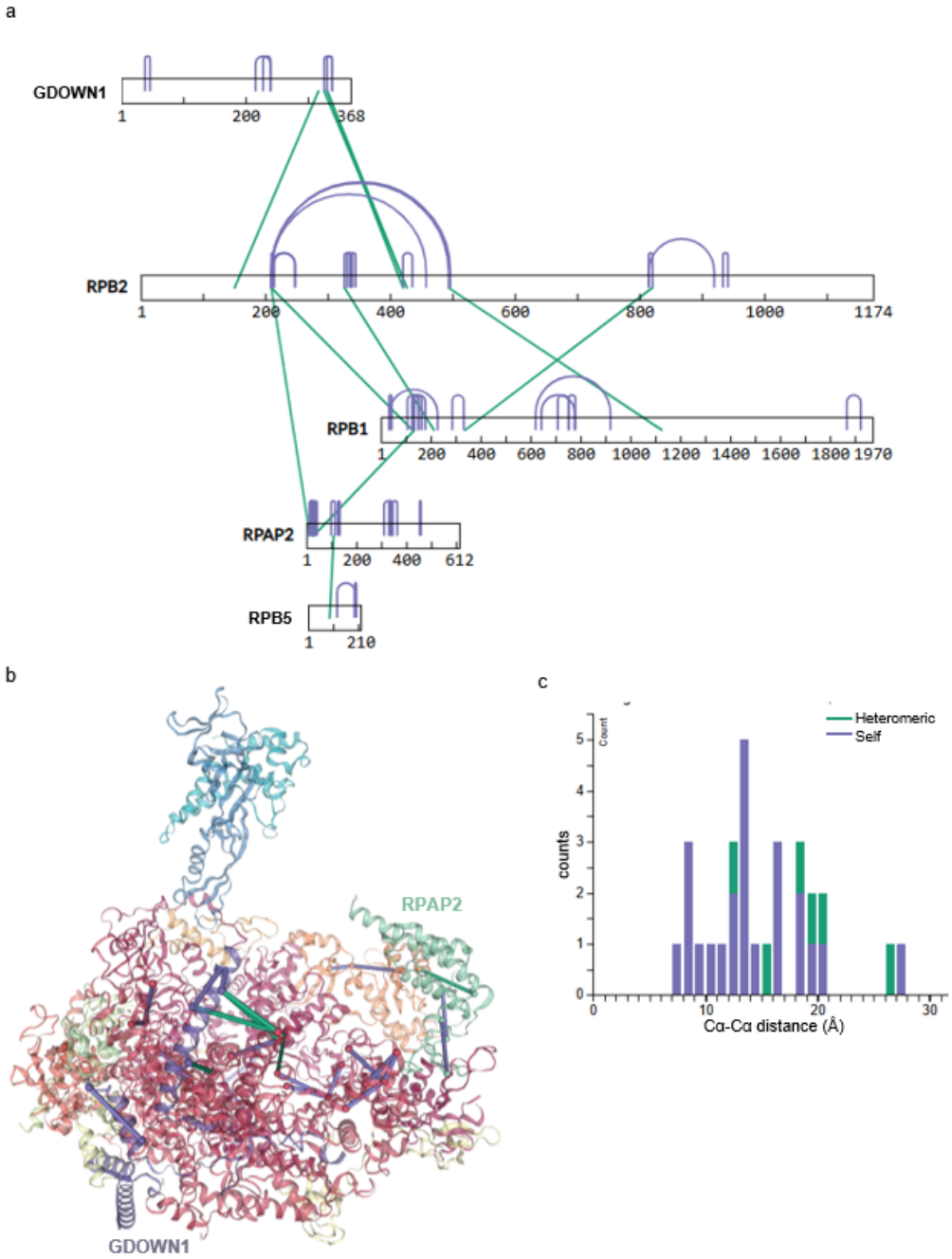


Figure 22: XL-MS analysis of cytoplasmic Pol II complexes. (**a**) Schematic of identified cross-links in the cytoplasmic Pol II. Violet lines - self cross-linked peptides, green lines - heteromeric cross-linked peptides. (**b**) Structure of cytoplasmic Pol II complex with mapped cross-links. (**c**) Plot of Cα-Cα distances between the identified cross-links using structure of cytoplasmic Pol II as reference. Violet - self cross-linked peptides, green - heteromeric cross-linked peptides. Visualized with xiView (Combe *et al*., 2024). XL-MS analysis done by B. Neuditschko (Herzog lab).

To sum up, I successfully solved the structure of cytoplasmic Pol II. The structure is in agreement with published structures of *in vitro* reconstituted complexes of Pol II-RPAP2 and Pol II (G). I managed resolve GDOWN1 bound to this complex showing improved densities α1-2$_{GDOWN1}$ close to RPB3 subunit and α4$_{GDOWN1}$ that hints its involvement in interaction with clamp region and having a role in regulation of cytoplasmic Pol II.

3.3. Characterization of GPN1 and GPN3 binding to Pol II and RPAP2

Since the cytoplasmic Pol II structure did not reveal extra density with high enough resolution corresponding to GPNs and XL-MS did not yield sufficient amount of interesting cross-links, I decided to better characterize the role of GPNs and RPAP2 in regulation of Pol II in cytoplasm. I first started with optimizing the expression and purification of GPN1 and GPN3. Initially, I chose *E. coli* cells as expression system. I was able to detect GPN1 and GPN3 upon induction of expression (Fig. 23 a). A small-scale test purification revealed that I could purify GPN1, but GPN3 was insoluble (Fig. 23 b). GPN2 ORF was also cloned into *E. coli* vector, but the expression was not tested as I changed the expression system to insect cells after the failed GPN3 test purification. GPN1, GPN2 and GPN3 were separately expressed in insect cells but with very low protein yield. I had to confirm the presence of expressed GPN1, GPN2 and GPN3 in test purification using Western blot (Fig. 23 c).

To improve protein yield, GPN1 and GPN3 were co-expressed in insect cells. I choose these two GPNs, because in most literature this heterodimer is mainly studied (Carre & Shiekhattar, 2011, Méndez-Hernández *et al*., 2014). In addition, from the MS of pull-downs it was clear that GPN1 and GPN3 are more enriched in CE than GPN2 (Figure 6 a, 7 a). Co-expression of GPN1-GPN3 massively increased the solubility of the complex and improved the yield of the protein purification (Fig. 23 d).

Additionally, I confirmed that GPN1-GPN3 complex has GTPase activity using malachite green phosphate assay to measure phosphate produced after GTP hydrolysis (Fig. 23 e). When GPN1-GPN3 was incubated with GTP there is 6x fold change in presence of phosphate compared to control buffer (Fig. 23 e). There was slight increase in presence of phosphate in GPN1-GPN3 only samples that can be explained by phosphate carry-over from protein purification (Fig. 23 e)

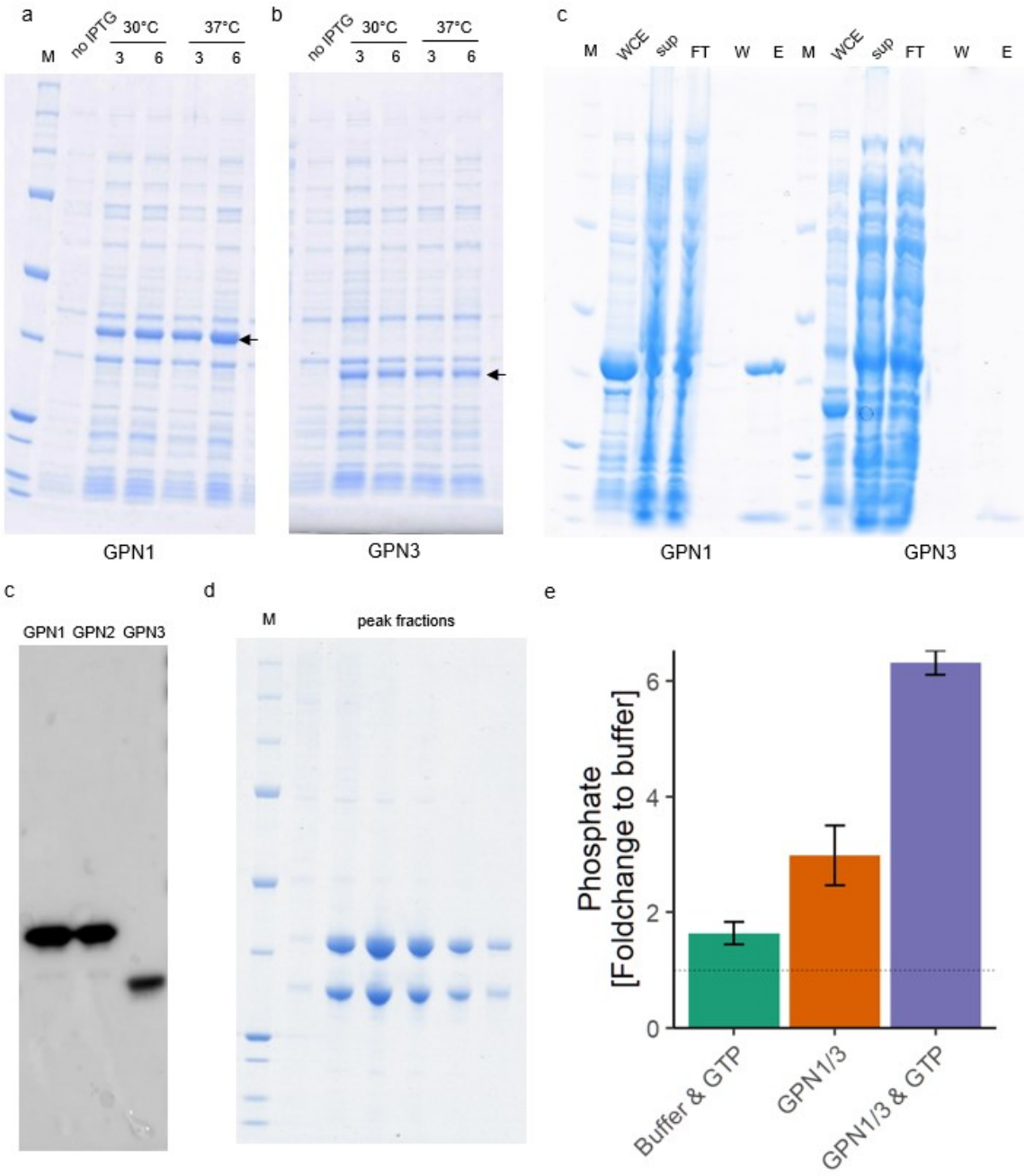


Figure 23: GPN protein purification and GTPase activity assay. (**a**) Test expression of GPN1 and GPN3 expressed in *E. coli* BL21(DE3)RIL cells. Cells were expressing the proteins for 3 or 6 hours at 30°C or 37°C. Samples were separated on SDS-PAGE gel and Coomassie stained. Proteins of interest marked with arrow. (**b**) Test purification of *E. coli* expressed GPN1 and GPN3. Samples were separated on SDS-PAGE gel and Coomassie stained. WCE - whole cell extract, sup - supernatant, FT - flow-through, W - wash, E - elution. M - Precision Plus Protein™ Unstained Standards (BioRad Laboratories). (**c**) Test purification of GPN1, GPN2 and GPN3 expressed in Hi5 insect cells. Samples were first separated by SDS-PAGE, the GPNs were detected with Western blot using antibody against His-tag. (**d**) Size exclusion of GPN1-GPN3 complex co-expressed in Hi5 insect cells. SDS-PAGE gel with peak fractions from size exclusion stained with Coomassie. M - Precision Plus Protein™ Unstained Standards (BioRad Laboratories). (**e**) GTPase activity of GPN1-GPN3 complex. Malachite Green Phosphate Assay kit (Sigma-Aldrich) was used to measure presence of phosphate after incubation of GPN1-GPN3 with GTP. Dash line is 1. Plot generated with help of J. Stopp.

To test the interaction of GPN1-GPN3 with Pol II, I formed complex of Pol II and GPN1-GPN3 and applied it on size exclusion column. I noticed that Pol II and GPN1-

GPN3 complex did not elute together in the same fractions (Fig. 24 a). This could be explained that either the complex did not form or the interaction between Pol II and GPN1-GPN3 is weak and conditions of SEC were too harsh. Therefore, I chose a different approach. I purified GPN1-GPN3 complex with intact His-tag on GPN1 and immobilized the complex on the magnetic Ni-beads. Then, I incubated Pol II with the beads and eluted bound proteins. This experiment confirmed the interaction of GPN1-GPN3 with Pol II and based on the intensity of bands, one can assume the complex forms in stoichiometric manner (Fig. 24 b). These observations showed that the binding to Pol II is present but it is weak.

I was also interested in which part of Pol II is the binding site for GPN1-GPN3. Previously published experiments suggested that CTD domain of RPB1 interacts with GPNs (Carre and Shiekhattar, 2011). I purified hCTD domain of Pol II. Since the protein tends to phase separate and degrade, I kept the His-MBP tag attached. Similarly as with GPN1-GPN3 pull-downs, I first immobilized hCTD on Ni-beads and then incubated with GPN1-GPN3 complex and eluted the bound proteins. In order to exclude unspecific binding of GPN1-GPN3 to still present MBP tag, additionally I purified MBP protein as well. I confirmed no unspecific binding of tested GPNs to MBP either (Fig. 24 c). This experiment showed no evidence of interaction between hCTD and GPN1-GPN3 suggesting a different binding site of GPNs on Pol II (Fig. 24 c).

From my structure of cytoplasmic Pol II, I saw there was a density in the area of clamp that could not be resolved into higher resolution (Fig. 16 b). I hypothesized that a clamp region can potentially be a binding site for GPN1-GPN3. One of the prominent changes in conformation of Pol II upon binding to DNA and engaging in transcription is closing of the clamp region. This region is flexible when no DNA/RNA is bound to Pol II and is hard to resolve. Thus, I tested binding of GPN1-GPN3 on Pol II elongation complex (Pol II EC). Pol II EC consisted of Pol II and DNA-RNA scaffold. From the pull-downs, it is visible that GPN1-GPN3 does not have affinity to Pol II EC (Fig. 24 b). This suggests that after clamp is closed, the binding site for GPN1-GPN3 is not accessible and GPNs cannot bind to Pol II anymore. To further, test the hypothesis I am also in process of cloning clamp regions of Pol II into insect cell expression vector. I plan to express and purify the Pol II clamp region and then test the interaction between GPN1-GPN3 and Pol II clamp using pull-downs or mass photometry.

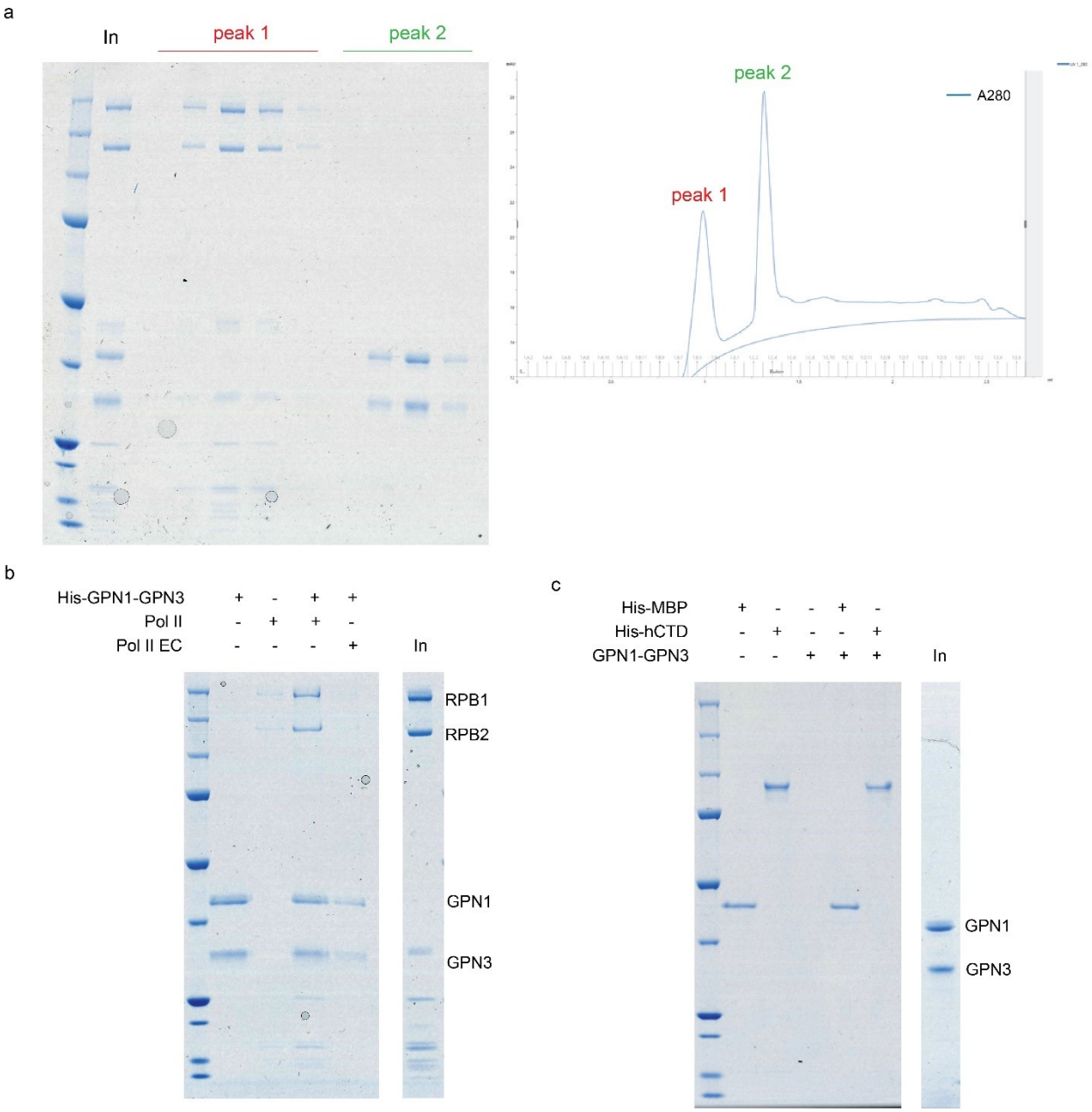


Figure 24: Interactions between GPN1-GPN3 and Pol II complexes. (**a**) Size exclusion of Pol II and GPN1-GPN3 complex showed that they do not elute together. blue line - A280 curve. (**b**) Pull-down of His-GPN1-GPN3 with Pol II and Pol II EC. (**c**) Pull-down of His-hCTD and GPN1-GPN3. MBP serves as an internal control since hCTD has still MBP-tag attached. SDS-PAGE gels were Coomassie stained. M - Precision Plus Protein™ Unstained Standards (BioRad Laboratories).

One of the potential proteins that could stabilize GPN1-GPN3 on Pol II was RPAP2. It was previously confirmed that RPAP2 interacts with GPN1, but no evidence was shown of its interaction to GPN1-GPN3 complex. To confirm binding of GPN1-GPN3 to RPAP2, I used mass photometry. GPN1-GPN3 complex and RPAP2 are measured as 75 kDa and 69 kDa peak respectively (Fig. 25). When mixed they form a complex that shows as a 144 kDa molecular weight peak (Fig. 25). I successfully confirmed the interaction between GPN1-GPN3 and RPAP2 with this experiment.

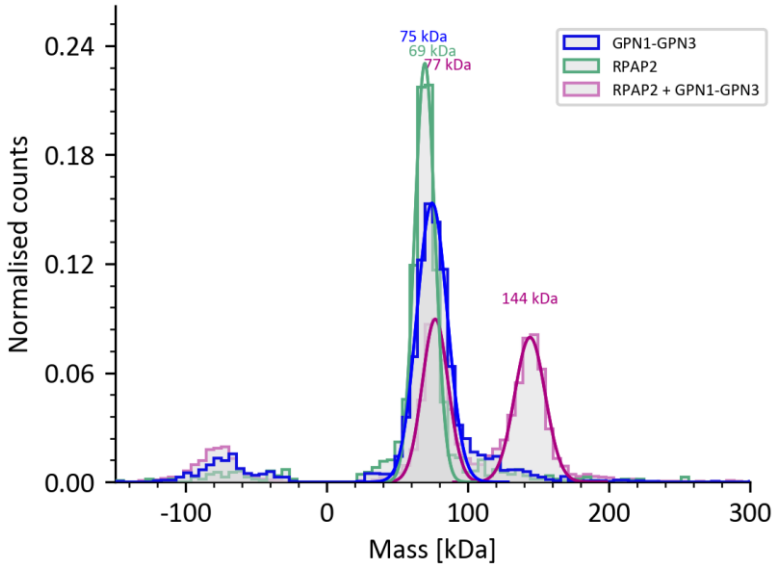


Figure 25: Interaction between RPAP2 and GPN1-GPN3. Mass photometry data collected with Refeyn TwoMP shows separate peaks of 69 kDa nad 75 kDa corresponding to RPAP2 and GPN1-GPN3 respectively. When proteins were combined there is another peak with 144 kDa, which corresponds to size of the whole complex. Data were analyzed with Refeyn Discover^{MP} software and instrument was calibrated with MfP1 standard prior the measurement.

I also predicted the structure of RPAP2-GPN1-GPN3 complex using AF3 (Abramson *et al.*, 2024). For this purpose, I removed the N-terminal region of RPAP2 between 1-320 amino acids. The prediction resulted in the ipTM score 0.67 and pTM score 0.69. This suggests that the predicted structure might be similar to the real structure. The structure shows that GPN1, GPN3 and RPAP2 might form a compact complex with multiple regions that can interact with each other (Fig. 26 a). The individual pIDDT scores show that the parts of GPN1, GPN3 and RPAP2 that are in close proximity with each other are predicted with very high confidence (Fig. 26 b). I also noticed that some regions are unstructured and in general with low pIDDT score (Fig. 26 b). Also PEA (predicted align error) values that serve to estimate the error in the relative position and orientation between two residues shows that some regions have quite low predicted error while there are regions with higher estimated error (Fig. 26 c).

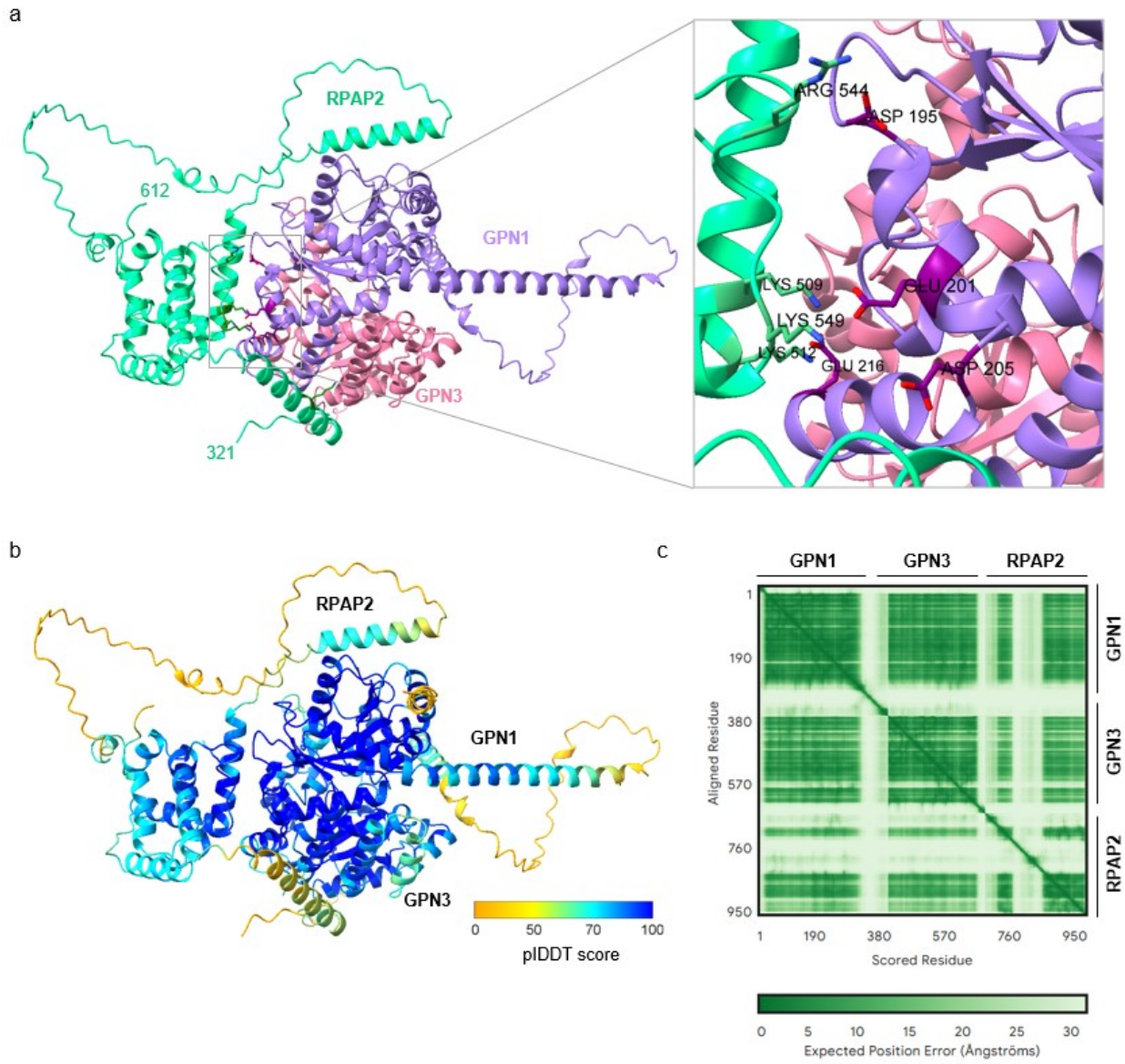


Figure 26: Predicted structure of RPAP2, GPN1 and GPN3 complex. (**a**) AlphaFold3 (AF3) prediction of RPAP2 and GPN1-GPN3 complex (Abramson *et al*., 2024). Squared region is zoom-in of potential interaction sites between GPN1 and RPAP2, regions are highlighted and amino acids labelled. GPN1 – violet, GPN3 – pink, RPAP2 - green (**b**) Predicted structure of RPAP2-GPN1-GPN3 coloured based on pIDDT scores. (**c**) Plot shows the PEA values of the predicted structure. Generated by AF3 (Abramson *et al*., 2024).

Furthermore, complex of RPAP2 and GPN1-GPN3 was analysed by XL-MS. The analysis identified multiple cross-linked peptides between the proteins. There were multiple residues RPAP2 in N-terminus, NTD and C-terminal region that cross-linked with GPN1 and GPN3 (Fig. 27 a). RPAP2 region 108-125 is interacting with GPN3 region 60-66 amino acids (Fig. 27 a). This is an interesting observation because based on previously published results it is C-terminal region but not NTD of RPAP2 is the binding site for GPN1 (Forget *et al*., 2013). It can be explained by the fact that RPAP2-GPN1-GPN3 complex is packed tightly and these regions get to close proximity. I also mapped the cross-links into the predicted structure of RPAP2-GPN1-GPN3 complex and checked the Cα-Cα distances. Many of the identified cross-links are above the maximum 25 Å distance of cross-linker, which suggests that these regions can have different structure or orientation (Fig. 27 b, c).

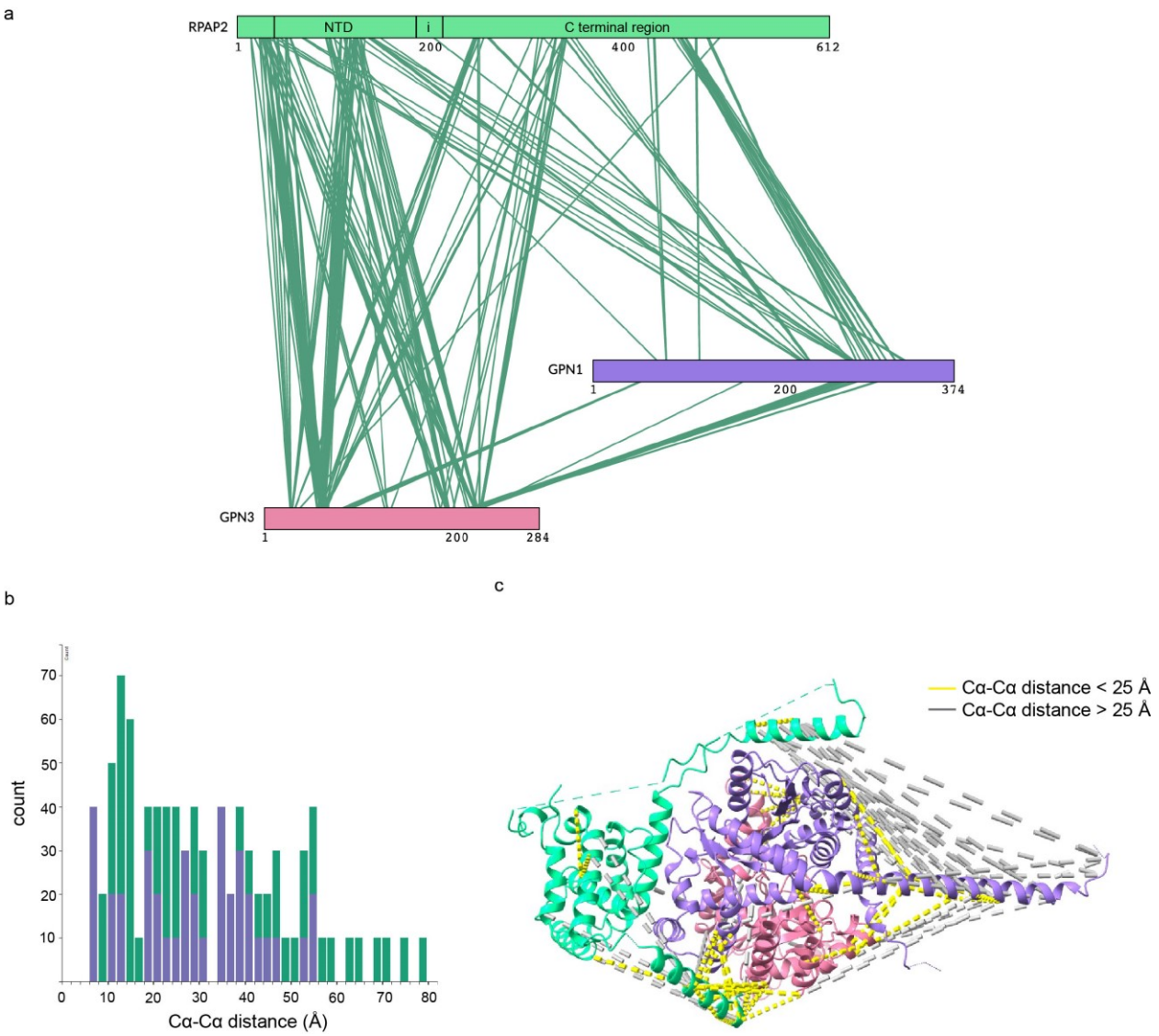


Figure 27: Cross-linking mass-spectrometry of RPAP2-GPN1-GPN3 complex. (**a**) Schematic of identified cross-links in the complex or RPAP2 (green), GPN1 (violet) and GPN3 (pink). Green lines - heteromeric cross-links. NTD - N-terminal region, i - TFIIF inhibitory domain. Generated by xiView. (**b**) Plot of Cα-Cα distances between the identified cross-links. Green lines - heteromeric cross-links, violet - self cross-links Plot generated with xiView. (**c**) AF3 predicted structure of RPAP2-GPN1-GPN3 complex with mapped cross-links. Only residues with pIDDT > 35 showed. yellow - Cα-Cα distance < 25Å, grey - Cα-Cα distance > 25Å. Cross-links mapped using XMAS plugin in ChimeraX (Pettersen *et al*., 2021). XL-MS analysis done by B. Neuditschko (Herzog lab).

Interestingly, there is a region between RPAP2 and GPN1 that contains amino acids of opposite charge and it can be a potential binding site of these two proteins (Fig. 26 a). This observation together with results from XL-MS of RPAP2-GPN1-GPN3 complex and available literature helped me to design several mutants of RPAP2 and GPN1 to better characterize binding site between these proteins (Fig. 28 b).

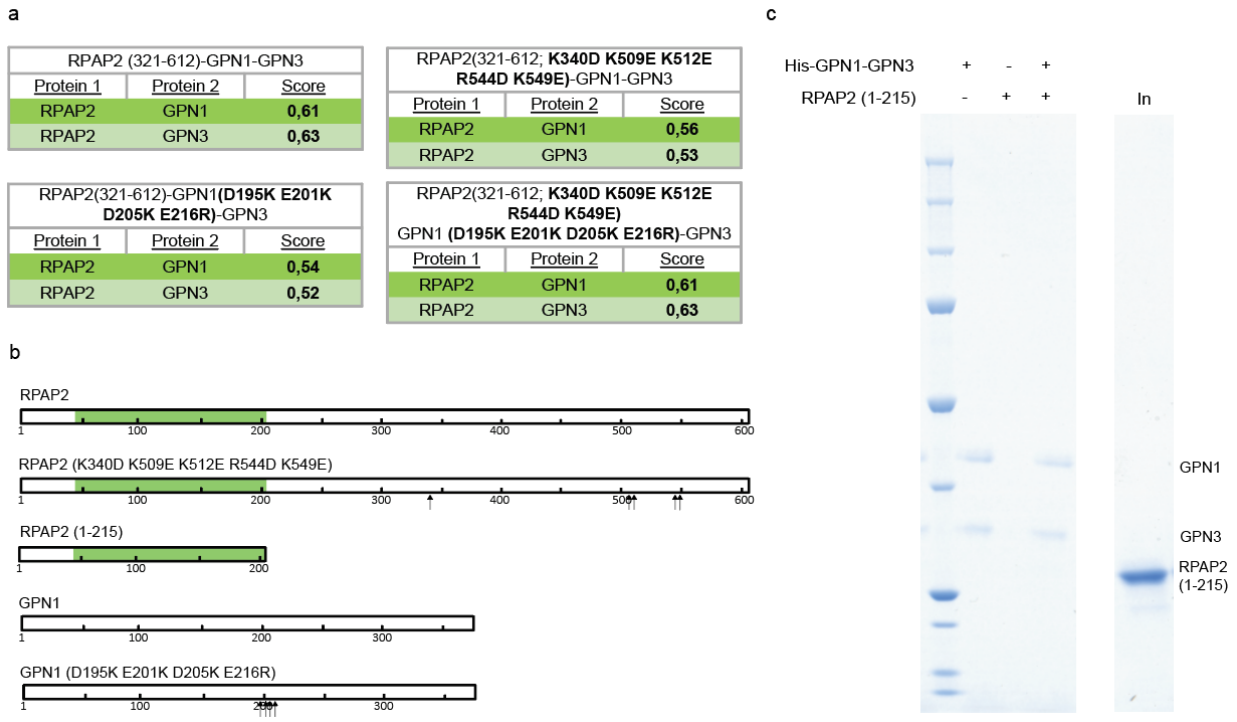


Figure 28: Interactions between mutants of RPAP2 and GPN1-GPN3. (**a**) iPTM scores of interactions between RPAP2, GPN1 and GPN3 wt and the mutated versions of the proteins. iTPM score generated by AF3 (Abramson *et al*., 2024). (**b**) Schematic of proteins and the positions of mutations in amino acid sequence. Green - region of RPAP2 that has been structurally solved. Arrows point at the mutation site. (**c**) Pull-down assay of His-GPN1-GPN3 and RPAP2(1-215). Samples were separated on SDS-PAGE gel and Coomassie stained. M - Precision Plus Protein™ Unstained Standards (BioRad Laboratories).

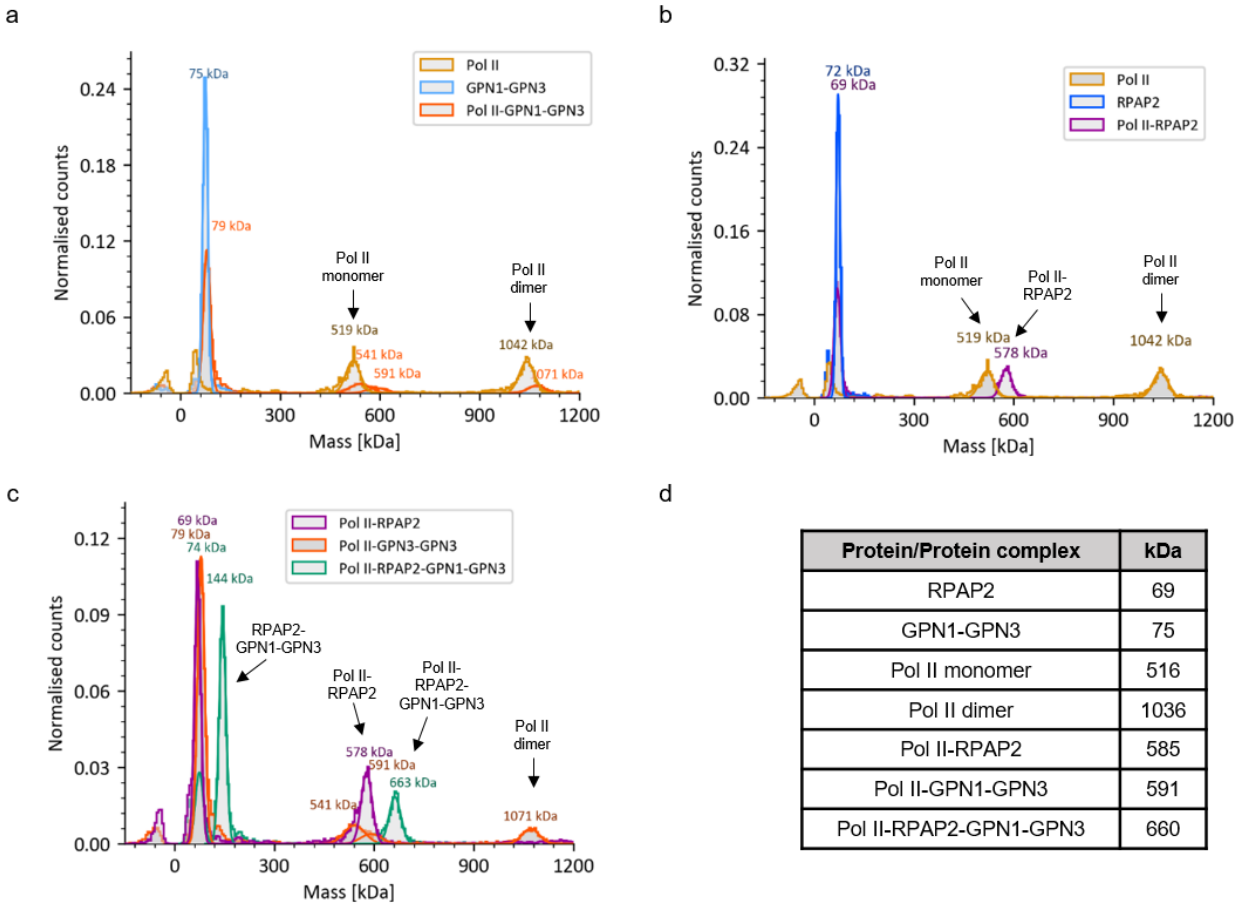
A previously published paper about structure of RPAP2 with Pol II showed that C-terminal region of RPAP2 is predominantly flexible and unstructured (Fianu *et al*., 2021; Wang *et al*., 2022). Since N-terminal part of RPAP2 is bound to Pol II it suggests that C-terminal part will be important for GPN binding. I expressed and purified truncated version of RPAP2 (1-215). Pull-down with immobilised GPN1-GPN3 revealed that RPAP2 truncated does not interact with GPN1-GPN3 complex (Fig. 28 c). Even though there were cross-links identified between N-terminal region of RPAP2 and GPN1 and GPN3, they are not important for the binding of these proteins (Fig. 27 a). Thus, the binding site for GPN1-GPN3 likely lies in C-terminal region of RPAP2.

Furthermore, I designed mutants based on AF3 predictions of RPAP2-GPN1-GPN3 complex. Specifically in RPAP2 I mutated following amino acids, K340D, K509E, K512E, R544D, K549E, in GPN1 D195K, E201K, D205K and E216R (Fig. 28 b). Amino acids were mutated to opposite charge. The confidence score were reduced when mutation were introduced in RPAP2 or GPN1 in AF3 prediction (Fig. 28 a). Furthermore, when I designed rescue phenotype mutants and the confidence score increased to similar values as wild-type proteins (Fig. 28 a). I am in the process of preparing these mutants to test this hypothesis. I will use mass photometry to test the interactions as this proved to be efficient and reliable method for this kind of experiment.

3.4. Structure of *in vitro* reconstituted Pol II (G)-RPAP2-GPN1-GPN3 complex

I also wanted to resolve the structure of *in vitro* Pol II complex containing GPN1, GPN3, RPAP2 and GDOWN1 and see if there are some differences compared to cytoplasmic Pol II structure. Working with pure proteins could also improve the resolution of some parts of the complex.

Thus, I tested the interaction of RPAP2, GPN1-GPN3, and GDOWN1 with Pol II. Generally, Pol II forms monomers and dimers in solution, which I confirmed by mass photometry. There are always two peaks present, around 516 kDa for monomer and 1032 kDa for dimers (Fig. 29 a, b, d). When I added GPN1-GPN3 to Pol II, I did not see a significant shift of the peaks, but there was very small peak corresponding to 591 kDa that would be the molecular weight of Pol II-GPN1-GPN3 complex (Fig. 29 a, d). It is interesting, as it suggests the complex is formed, but under these conditions it is not very stable, as it was the case with size exclusion (Fig. 24 a). When I added RPAP2 to Pol II I observed a dramatic shift toward monomeric Pol II. Dimer peak disappeared and instead there was a new peak of 578 kDa, which correspond roughly to Pol II-RPAP2 complex (Fig. 29 b, d). This observation suggests that RPAP2 keeps Pol II monomeric. Surprisingly, when I formed complex of Pol II with RPAP2 and GPN1-GPN3 I experienced the same shift in peaks toward the monomer. A 663 kDa peak was formed which is consistent with the expected size of the Pol II-RPAP2-GPN1-GPN3 complex (Fig. 29 c, d). From these observations, I can conclude that RPAP2 binds to Pol II and makes it monomeric. Additionally, RPAP2 helps to stabilize GPN1-GPN3 complex on Pol II.



d

Protein/Protein complex	kDa
RPAP2	69
GPN1-GPN3	75
Pol II monomer	516
Pol II dimer	1036
Pol II-RPAP2	585
Pol II-GPN1-GPN3	591
Pol II-RPAP2-GPN1-GPN3	660

Figure 29: Mass photometry of purified proteins and their complexes. (**a**) Pol II and GPN1-GPN3 interaction. (**b**) Pol II and RPAP2 interaction. (**c**) Pol II, RPAP2 and GPN1-GPN3 interaction. (**d**) Summary table of expected molecular weights (kDa) of proteins and protein complexes. Plots were generated using Refeyn Discover^{MP} software. Mass photometer was calibrated prior the usage with BSA.

Similarly, I tested the stability of the *in vitro* complexes by applying them on size exclusion column. I confirmed that once RPAP2 protein is bound to Pol II, GPN1-GPN3 is stabilized and the whole complex stays compact and elutes in same fractions (Fig. 30 a). Also, adding GDOWN1 to mixture resulted in formation of Pol II(G)-RPAP2-GPN1-GPN3 complex (Fig. 30 b). Thus, I further continued with preparing *in vitro* reconstituted complexes for cryoEM.

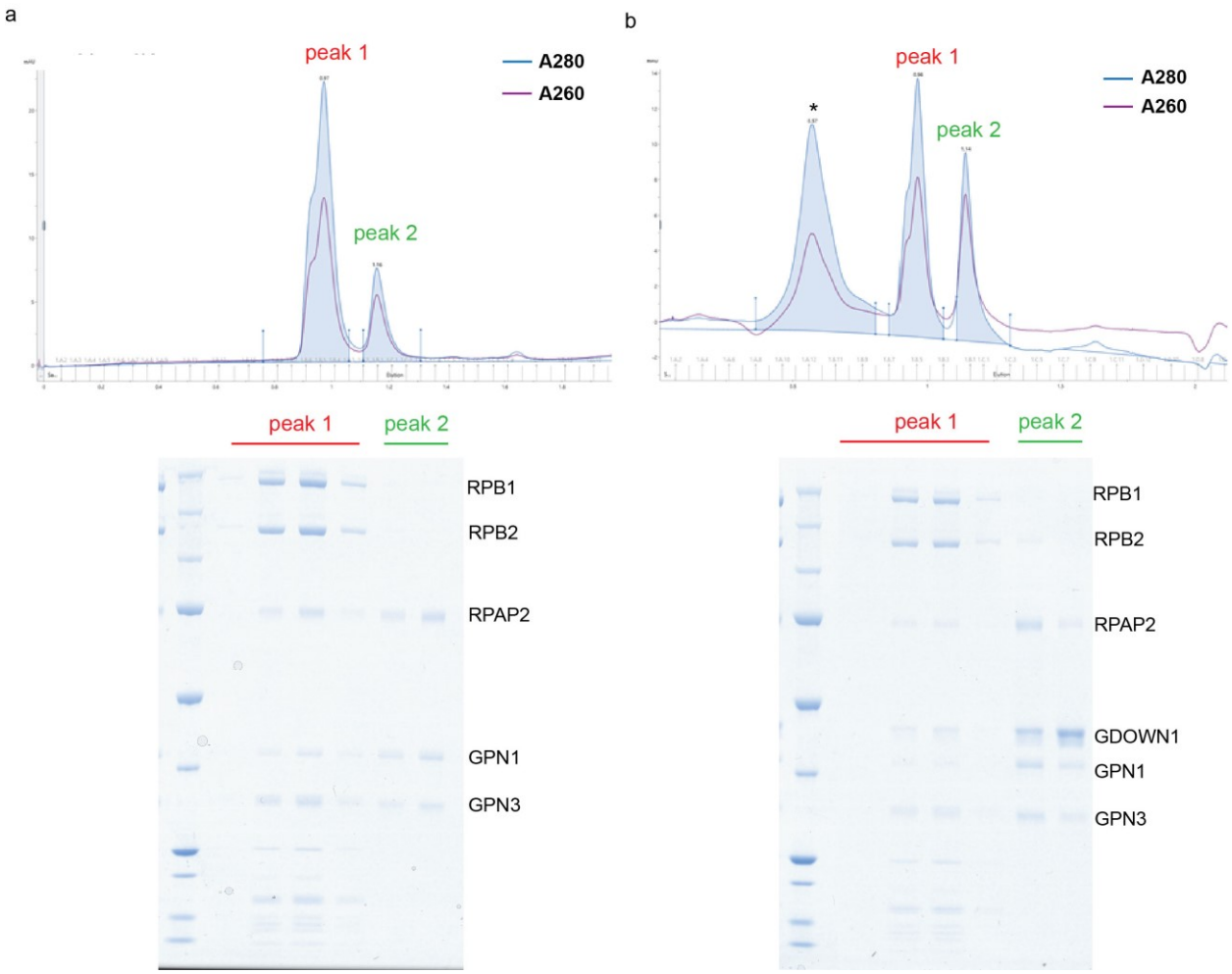
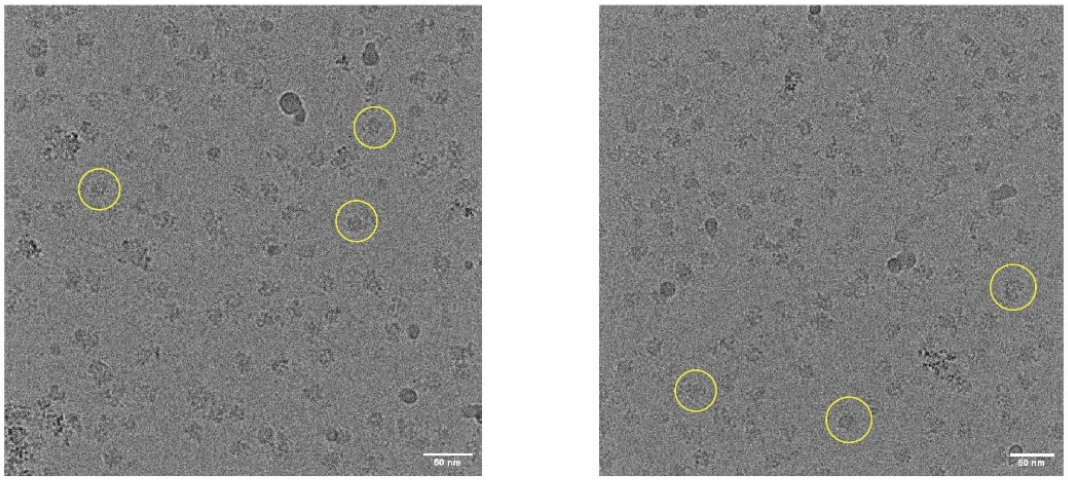


Figure 30: Size exclusion of Pol II-RPAP2-GPN1-GPN3 complex (**a**) and Pol II (G)-RPAP2-GPN1-GPN3 complex (**b**). Additional peak ***** in **b** correspond to washing step of the SEC column from the previous run. M - Precision Plus Protein^{TM} Unstained Standards (BioRad Laboratories).

Samples of *in vitro* reconstituted Pol II (G)-RPAP2-GPN1-GPN3 were cross-linked with 0.09% GA prior applying on cryoEM grids to stabilize the complex. I also screened the prepared grids on 200 kV Glacios TEM microscope and collected a small dataset of 3363 micrographs (Fig. 31 a). Particles from collected micrographs were picked in Warp and classified into 2D classes using RELION 5.1 (Fig. 31 b; Scheres, 2012).

a



b

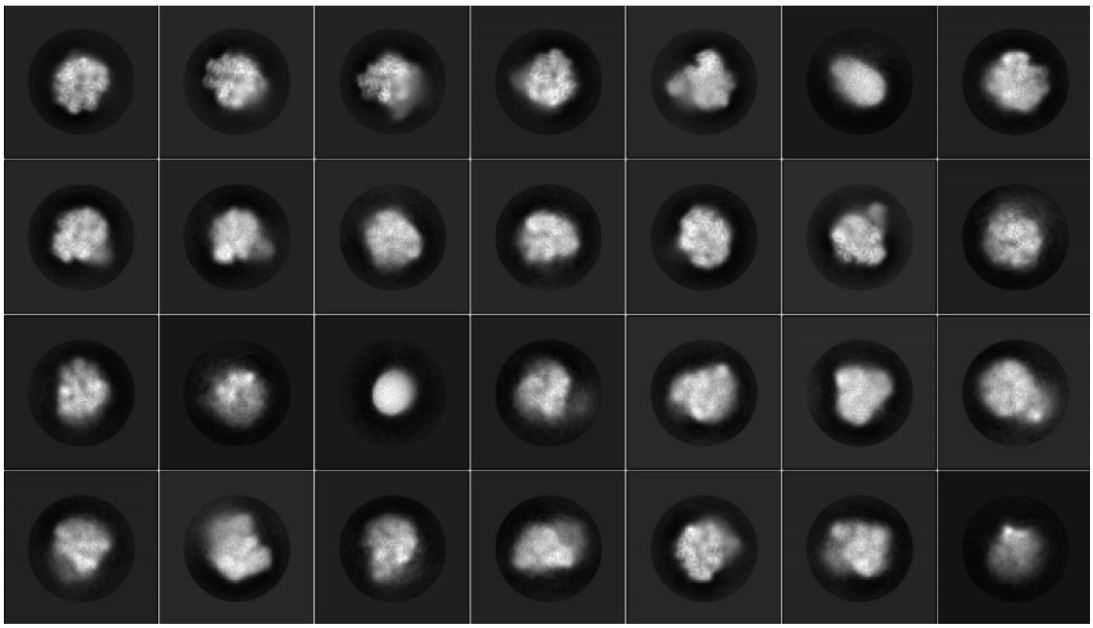


Figure 31: CryoEM collection of *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 complexes with 200 kV Glacios TEM microscope. (**a**) Representative micrographs. Defocus -2.6 µm and -2 µm respectively. 0.95Å pixel size, scale bar - 50 nm. (**b**) Representation of 2D classes from picked particles. Generetad with RELION 5.1 (Scheres, 2012).

This small data collection resulted in a 3D density map with nominal resolution 5.3 Å. I fitted the structural model of cytoplasmic Pol II in the density map and there are prominent densities that correspond to RPAP2 and GDOWN1, but no clamp density was identified (Fig. 32). Density of GDOWN1 in *in vitro* Pol II complex shows improvements in helix $\alpha4_{GDOWN1}$ that is prolonged towards the clamp region of Pol II (Fig. 32). Much improved density quality shows also helix $\alpha1_{GDOWN1}$ and $\alpha2_{GDOWN1}$ without additional local 3D classification as in case with cytoplasmic Pol II (Fig. 32).

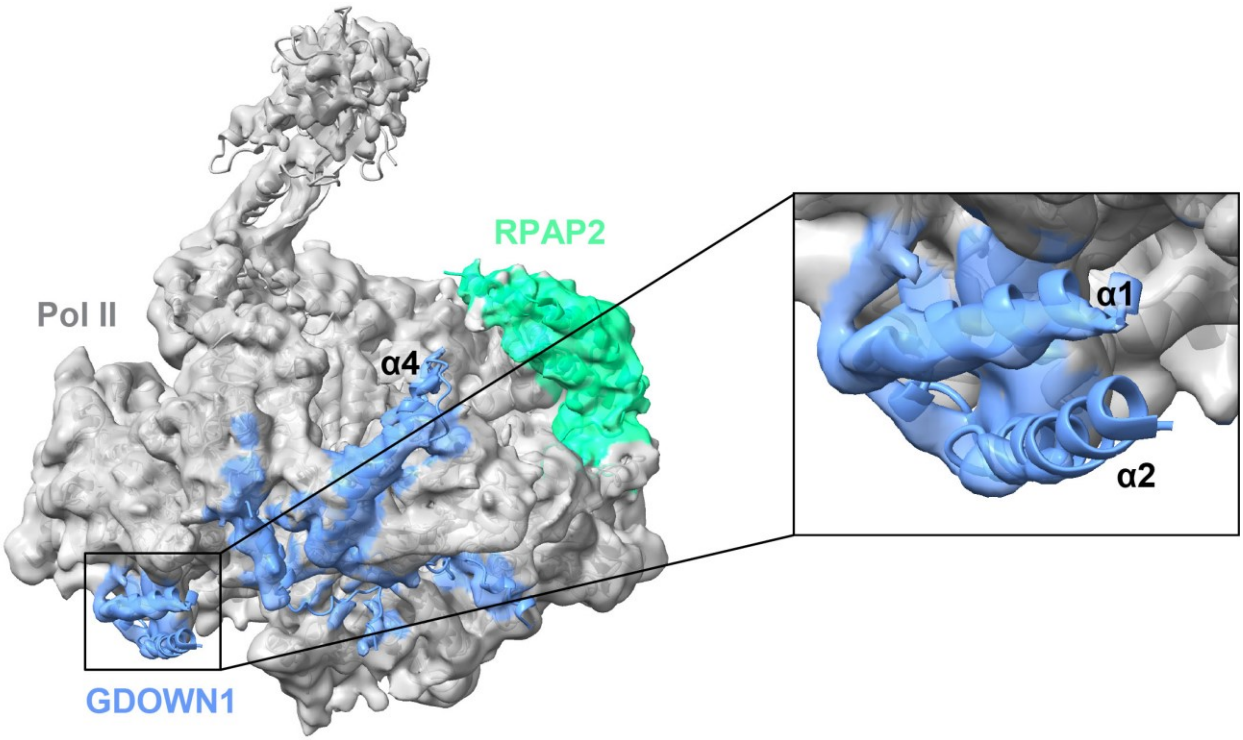


Figure 32: Structure of *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 complex. A cryoEM map of Pol II (G)-RPAP2-GPN1-GPN3 fitted with structural model of cytoplasmic Pol II. Pol II in grey, GDOWN1 in blue and RPAP2 in green. Helices of GDOWN1 labelled in zoom-in of the region.

I also collected around 20 000 micrographs of *in vitro* reconstituted Pol II (G)-RPAP2-GPN1-GPN3 complexes with Thermo Fisher Titan Krios G3i transmission electron microscope (300 kV) at 130000x nominal magnification (Fig. 33). I am currently processing and analysing this dataset.

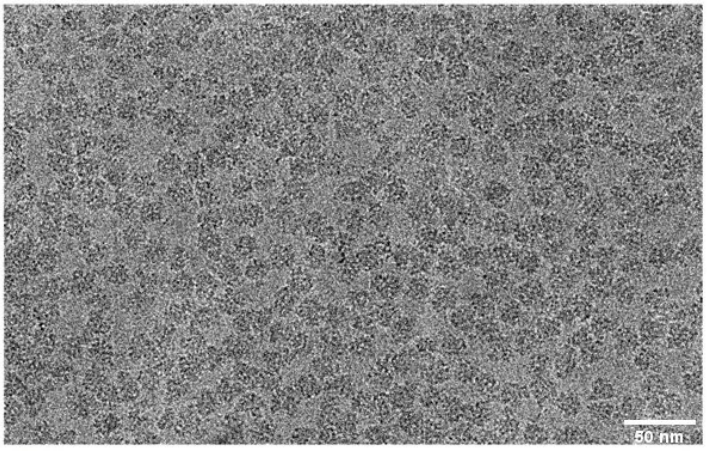


Figure 33: Representative micrographs of in vitro Pol II (G)-RPAP2-GPN1-GPN3 complex obtained by Thermo Fisher Titan Krios G3i transmission electron microscope (300 kV). 0.65Å pixel size, scale bar - 50 nm, defocus -1.5 µm.

Again, I was not able to identify density of GPNs in my structure, but I saw low-resolution density for the clamp region. This suggests that GPNs potentially bind Pol II in this region and the high flexibility of this part of Pol II makes it hard to resolve it as in case of the cytoplasmic Pol II structure. I decided to prepare cross-linked complexes of *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 and subject them to XL-MS. I was able to

identify several cross-links in the *in vitro* sample (Fig. 34 a, b). Many of them were between the subunits of Pol II and they were in agreement with the current knowledge of the Pol II structure (Bernecky *et al*., 2016, Fig. 34 a). No cross-links were identified for RPB12 with Pol II or other proteins in the sample. GDOWN1 interacts mainly with clamp region of RPB1, but also with RPB2, RPB3 and RPB10, which is in agreement with the published data (Jishage *et al*., 2018, Fig 34 a, b).

Cross-linked peptides were identified between the very N-terminus of RPAP2 and RPB1 clamp, while NTD domain of RPAP2 was interacting mostly with RPB2 and RPB5 subunits of Pol II, which corresponds to its location in the cytoplasmic Pol II structure (Fig. 34 a). RPAP2 protein was also interacting through its N-terminus and C-terminal region with GPN1 and GPN3 (Fig. 34 b).There were no cross-links identified in NTD region of RPAP2 and GPN1 or GPN3 (Fig. 34 b). Surprisingly, C-terminal part of GDOWN1 protein and C-terminal region of RPAP2 region was cross-linked suggesting an interaction between those proteins (Fig. 34 b).

Importantly, I identified cross-linked peptides between GPN1, GPN3 and RPB1 subunit. C-terminus of GPN1 cross-linked with clamp of RPB1 supporting the hypothesis that GPN1-GPN3 complex interacts with clamp of Pol II. There were few cross-links between CTD of Pol II and GPN1 and GPN3 (Fig. 34 b). Since this region of Pol II is flexible, it could be in close proximity to GPNs and that can explain identification of these cross-linked peptides.

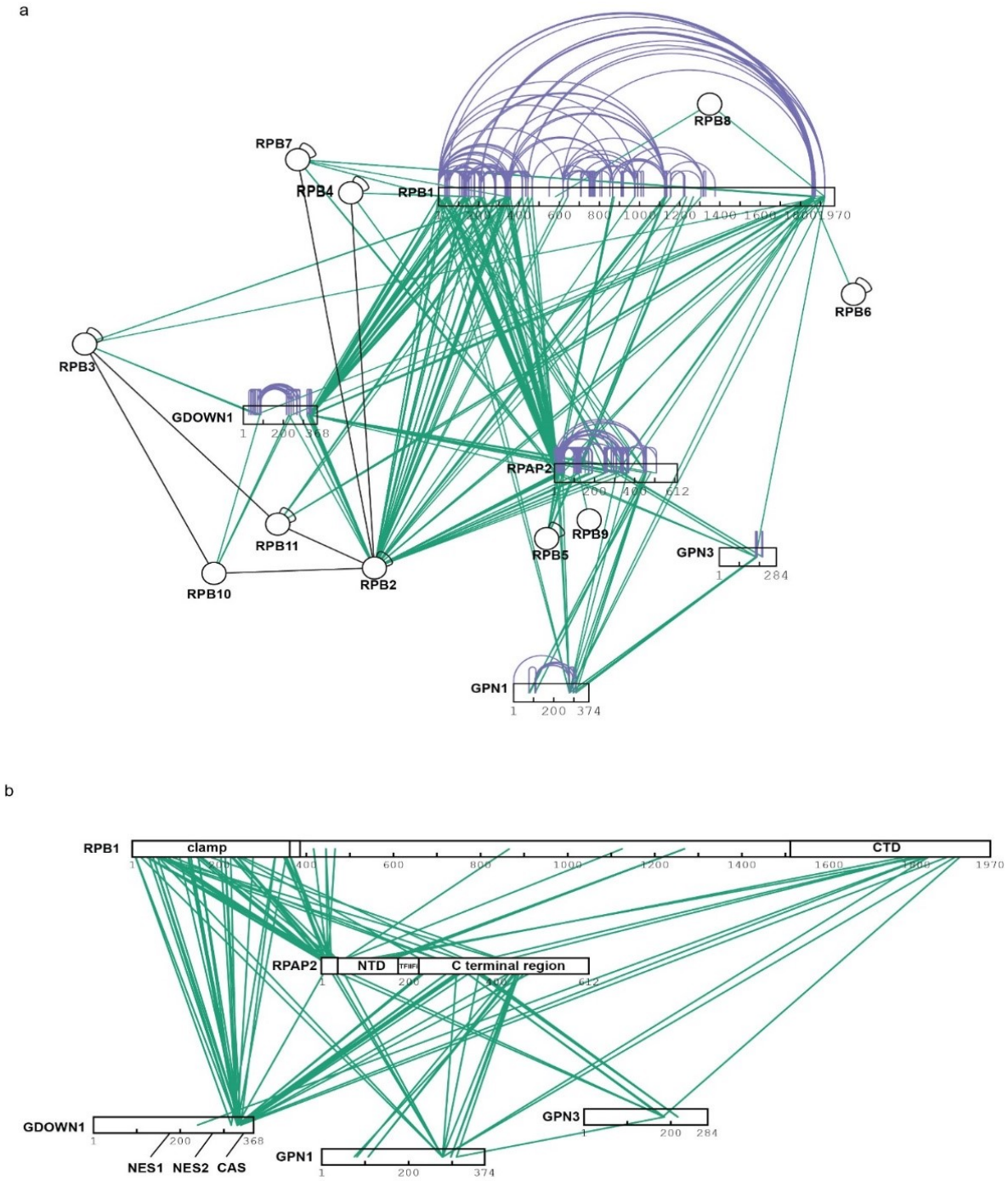


Figure 34: Cross-linking mass spectrometry analysis of in vitro reconstituted Pol II (G)-RPAP2-GPN1-GPN3 complex (**a**) Diagram of cross-links identified cross-linked peptides in *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 complex (**b**) Inter-molecular cross-link map of RPB1, GDOWN1, RPAP2, GPN1 and GPN3 only. Important regions of proteins are marked: RPB1 - clamp, CTD; RPAP2 - NTD, TFIIFi, C terminal region; GDOWN1 – NES1, NES2 and CAS sequences. Green lines - heteromeric cross-links, violet – self cross-links. Visualized with xiView. XL-MS analysis done by B. Neuditschko (Herzog lab).

In terms of Cα-Cα distances between the cross-links, most of them were in range of 325 Å when I use my model of the cytoplasmic Pol II as reference to calculate the distances (Fig. 35 a, b). The peptides that were cross-linked between stalk (RPB4) and RPB2 exceeding 25 Å are fitted better to Pol II dimer model (Fig. 35 c, d). This is because the *in vitro* sample submitted to XL-MS was not separated on size exclusion after the cross-linking and can contain small fraction of Pol II dimers.

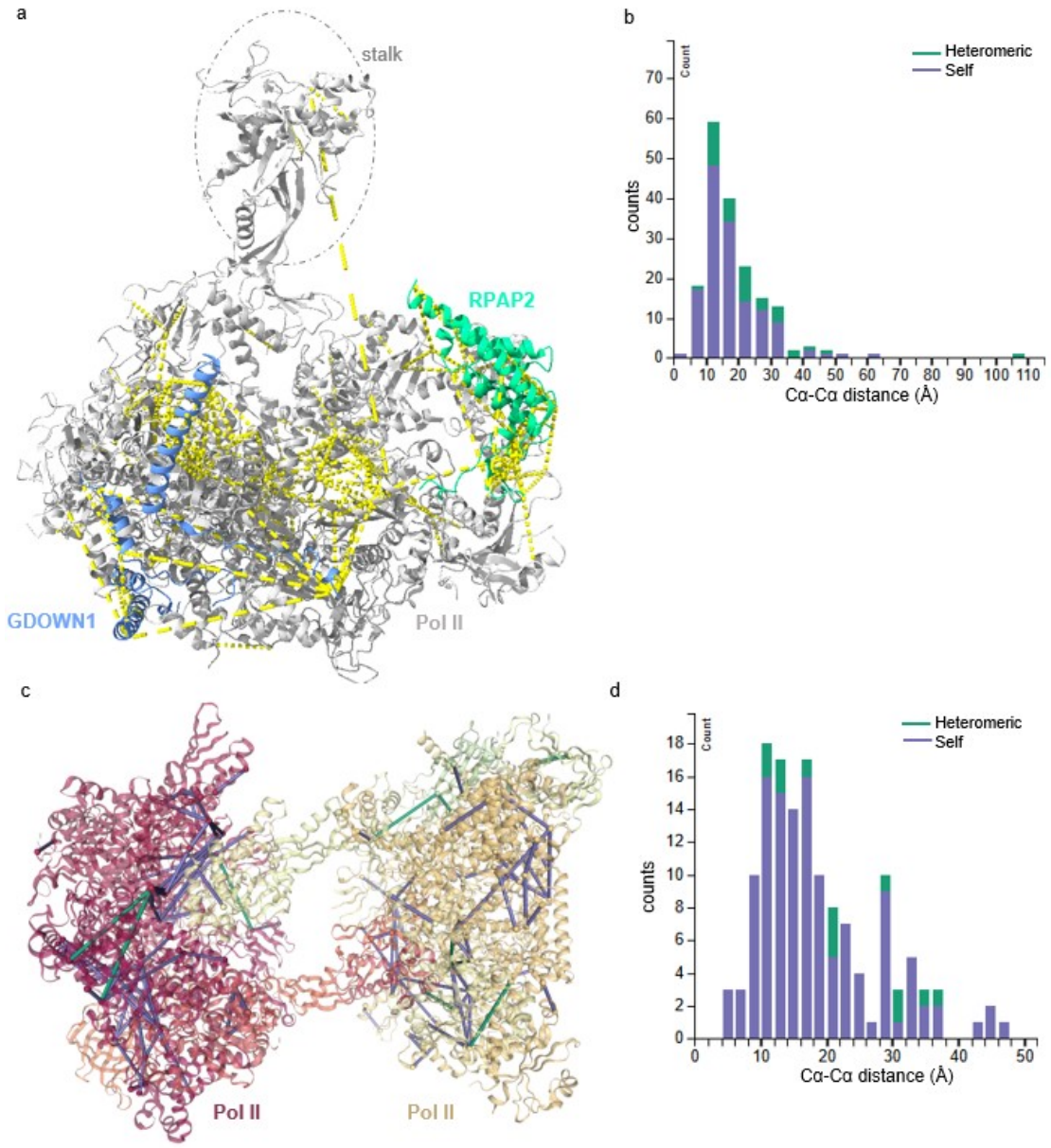


Figure 35: Molecular characterization of *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 by XL-MS. (**a**) Mapped cross-links from XL-MS of *in vitro* Pol II (G)-RPAP2-GPN1-GPN2 into model of cytoplasmic Pol II. Pol II in grey, GDOWN1 in blue, RPAP2 in green, cross-linked residues in yellow dashes. Visualized by XMAS plugin in ChimeraX (Lagerwaard *et al*., 2022). (**b**) Cα-Cα distance between identified cross-links using model of cytoplasmic Pol II using xiView (Combe *et al*., 2024) (**c**) Mapped cross-links from XL-MS of *in vitro* Pol II (G)-RPAP2-GPN1-GPN2 to model of Pol II dimer (PDB: 7OZN, Aibara *et al*., 2021). Violet lines - self cross-links, green lines - heteromeric cross-links. Visualized by xiView (Combe *et al*., 2024). (**d**) Cα-Cα distance between identified cross-links using model of Pol II dimer (PDB: 7OZN, Aibara *et al*., 2021) as reference using xiView (Combe *et al*., 2024).

3.5. hIWR1 does not interact with Pol II

One of the steps in Pol II regulation in cytoplasm is its shuttling to the nucleus. Part of this project was focused on identification and characterization of proteins that might be responsible for this process.

Yeast Iwr1 is crucial for nuclear shuttling of Pol II in yeast and its deletion led to aberrant phenotypes, accumulation of Pol II in cytoplasm and in general low vitality of the cells (Czeko *et al*., 2011). There is a human homolog of yeast Iwr1 (hIWR1) that

has not been yet characterized. I wanted to test if the function of the human homolog will be the same as in case of the yeast protein. I successfully purified hIWR1, partially adapting the protocol used for purification of yeast Iwr1 (Czeko *et al.*, 2011). Initially, I tested interaction of hIWR1 and Pol II by mixing them together in 1:5 molar ratio and applying the mixture on SEC column. hIWR1 was not co-eluting together with the Pol II which suggested either no or weak binding between the proteins (Fig. 36 a, b). I also tested interaction of hIWR1 and Pol II (G), but again I did not confirm their interaction (Fig. 36 c, d).

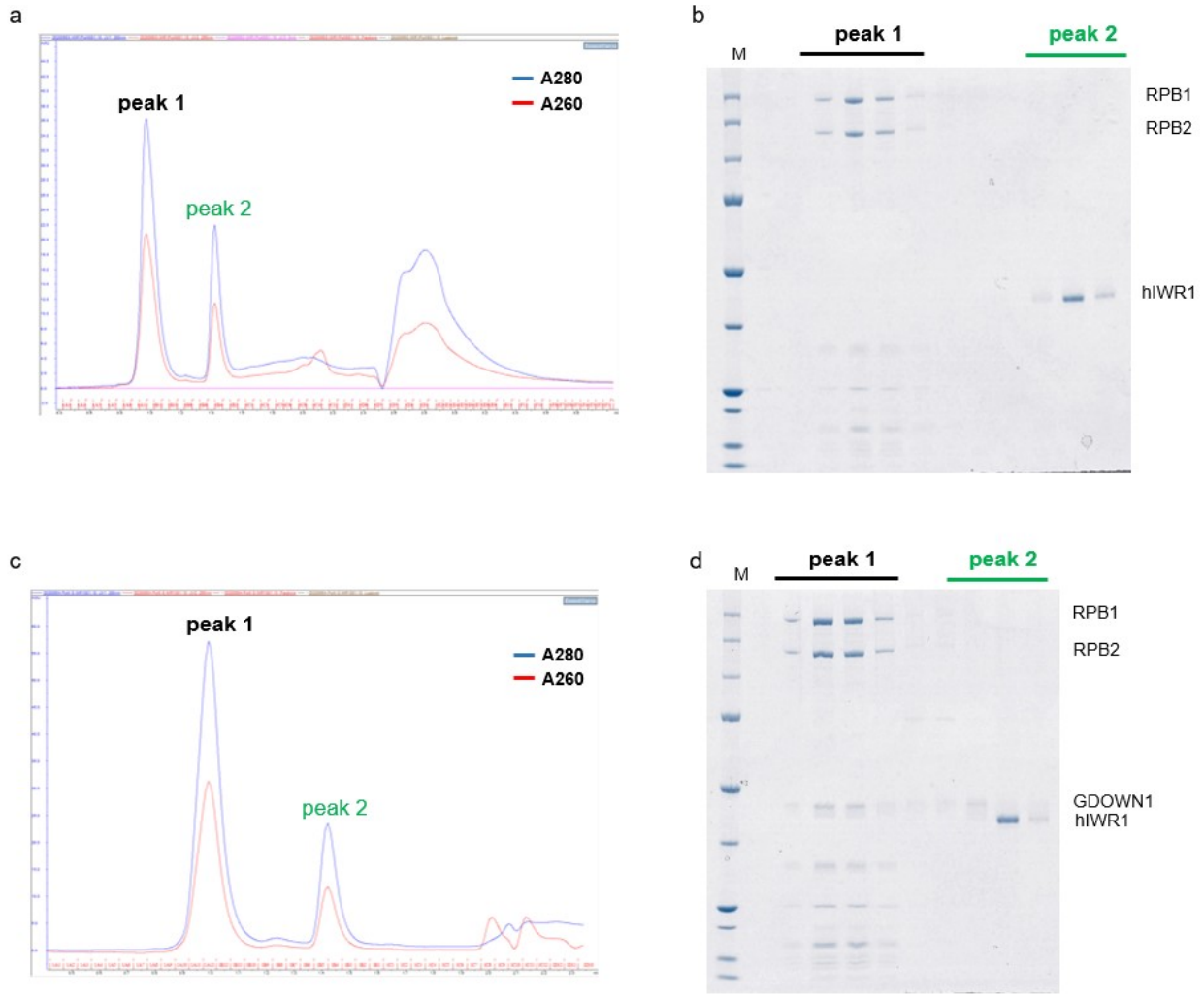


Figure 36: Complex formation of Pol II or Pol II (G) and hIWR1. (**a**, **b**) Size-exclusion chromatograms of Pol II + hIWR1 (**a**) and Pol II (G) + hIWR1 (**b**). (**c**, **d**) Peak fractions from SEC of Pol II + hIWR1 (**c**) and Pol II (G) + hIWR1 (**d**) analysed on SDS-PAGE gels. Coomassie stained. red line - A260, blue line - A280, M - Precision Plus Protein™ Unstained Standards (BioRad Laboratories).

I also tested interaction of hIWR1 with cytoplasmic Pol II complexes. I immobilized Pol II via GFP-GDOWN1 on GFP-nanobody resin and incubated the complex with His-hIWR1. The presence of Pol II, hIWR1 and GDOWN1 was identified using Western blot. I had to use His-tagged hIWR1 because the available antibodies against hIWR1 did not perform well enough for this kind of assay. I did not confirm the interaction of hIWR1 with cytoplasmic Pol II. All hIWR1 stayed unbound and did not elute together with GFP-GDOWN1 (Fig. 37). Since this method is less harsh than size exclusion, it suggests that there is no interaction between hIWR1 and Pol II. Important

to say is that hIWR1 was not identified among the significantly enriched proteins in the pull-downs of GDOWN1 from CE and NE (Fig. 6 a, b).

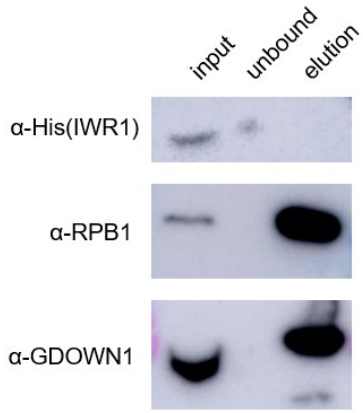


Figure 37: Western blot of cytoplasmic Pol II interaction with hIWR1. Samples from input, unbound proteins and elution were analysed using antibodies against CTD-domain of RPB1, GDOWN1 and His-tag (hIWR1).

There was an attempt to generate a K562 cell line endogenously expressing GFP-tagged IWR1. This strategy did not result in obtaining viable gene edited cell line, suggesting that hIWR1 is essential for the cells and different cloning strategy or method has to be applied. Thus, more experiments are needed in order to better understand role of hIWR1.

To sum up, I did not find evidence with above-mentioned applied methods that hIWR1 interacts with Pol II or is involved in its assembly.

3.6. Role of GDOWN1 *in vivo*

To further understand the role of GDOWN1 *in vivo*, generating of GDOWN1 knockdown cell line with auxin-inducible degron system (AID) was needed (Nishimura *et al*., 2009). This approach ensures a rapid depletion of GDOWN1 without potential negative effects of its deletion. K562 cell line with the TIR1 gene integrated into the GFP-GDOWN1 CRISPR cell lines was generated. First, the obtained clones were tested for the auxin (IAA) sensitivity. Some of the clones showed sensitivity to auxin and only those were used in later experiments (Fig. 38 a). A competition growth assay indicates decline in cell viability. Clones that are not sensitive to IAA maintain same viability, while IAA sensitive clones have reduced viability after longer incubation with IAA (Fig. 38 b). This speaks of importance of GDOWN1 in cells. I also optimized the IAA treatment times to see when GDOWN1 is completely depleted (Fig. 38 c). After 30 minutes of IAA treatment GDOWN1 levels are reduced and after 60 minutes no GDOWN1 was detected (Fig. 38 c).

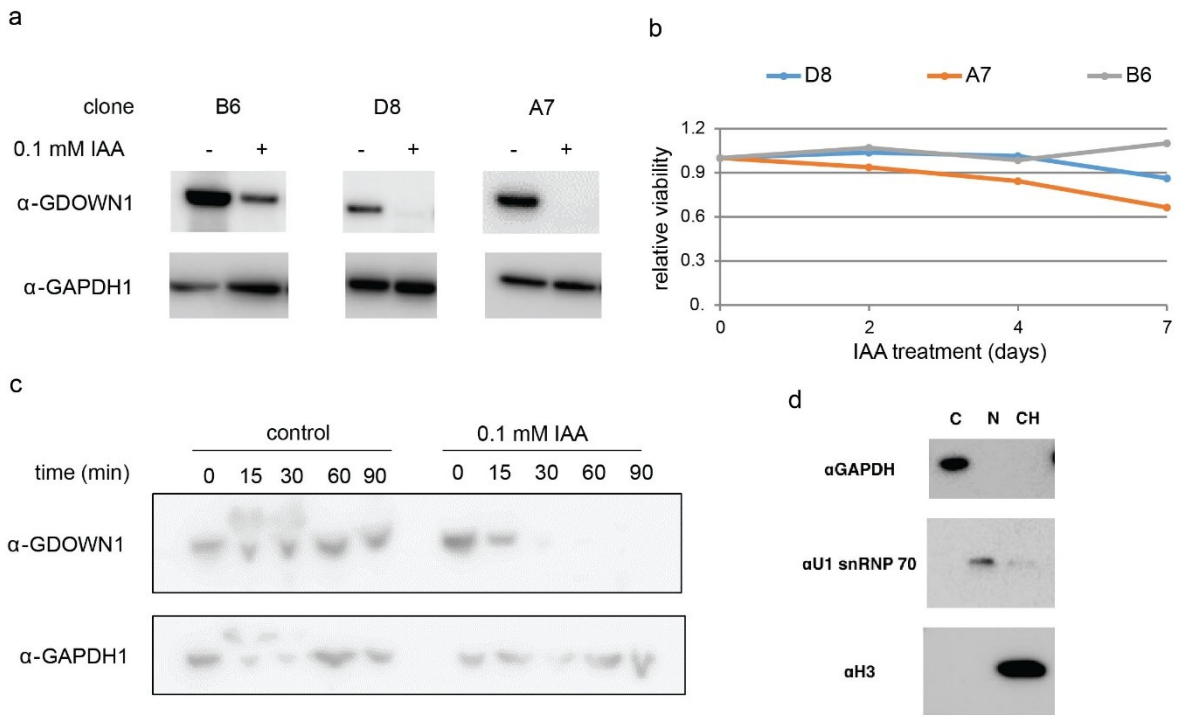


Figure 38: Cell generation of AID system. (**a**) Testing clones of K562 endogenously expressing GFP-GDOWN1 and TIR for auxin (IAA) sensitivity. Whole cells extracts were separated on SDS-PAGE gel, then Western blot was performed using antibodies against GDOWN1 and GAPDH1 (the latter as internal control). Experiments done by D. Riabov-Bassat (Plaschka lab). (**b**) Growth competition assay. Plot shows how depletion of GDOWN1 is affecting the cell viability of different clones compared to wt cells. Experiments done by D. Riabov-Bassat (Plaschka lab). (**c**) Optimization of GDOWN1 depletion in K562 cells. Cells were treated for specific times with 0.1 mM IAA. Presence of GDOWN1 in whole cells extract was observed by Western blot using antibody against GDOWN1 and GAPDH1 as internal control. (**d**) Correct cell fractionation of K562 GFP-GDOWN1 cells confirmed by Western blot of cytoplasmic (C), nuclear (N) and chromatin (CH) fractions using antibodies against GAPDH1 (C), U1 snRNP70 (N) and histone H3 (CH) markers.

I performed GDOWN1 depletion for 3.5 hours and 48 hours and prepared cytoplasmic and nuclear extracts. Correct cell separation was confirmed by Western blot (Fig. 38 d). Cytoplasmic extracts from the treated and control cells were subjected to MS analysis in order to identify if there is some change in abundancy of protein between the compartments after the GDOWN1 depletion. This approach did not yield clear results. In general, after 3.5 hours there were no major significant changes in abundancy of proteins of interest in cytoplasm between depleted and control cells (Tab. 1). In case of 48 hours depletion, I observed some significant changes, but since the depletion was too long, I cannot rule out potential secondary effect of the GDOWN1 depletion (Tab. 1).

Table 1: Summary of proteins identified in cytoplasm from K562 cells depleted of GDOWN1 for 3.5 hrs and 48 hrs. Arrows show increase or decrease of proteins of interest in cytoplasm when IAA treated and non-treated cells were compared. MS analysis done by VBC MS facility.

3.5 hours depletion				48 hours depletion			
protein	**fold change**	**p-value**	**change**	**Protein**	**fold change**	**p-value**	**change**
GDOWN1	-3.06	2.34E-04	↓	GDOWN1	-3.31	4.49E-05	↓
RPB2	-0.13	2.91E-01	↓	RPB2	-0.61	7.76E-04	↓
RPB8	0.2	4.73E-01	↑	RPB4	-0.54	5.38E-03	↓
RPB3	0.12	4.78E-01	↑	RPB8	-0.45	9.47E-03	↓
RPB7	0.46	1.62E-01	↑	RPB3	-0.4	3.44E-03	↓
RPB5	-0.46	4.14E-02	↓	RPB11	-0.4	7.46E-03	↓
RPB1	0.04	7.33E-01	-	RPB7	-0.34	8.71E-02	↓
RECQL5	0.02	8.63E-01	-	RPB9	-0.26	9.81E-02	↓
PCIF1	-0.95	3.18E-01	↓	RPB5	-0.08	5.44E-01	↓
GPN1	0.20	1.27E-01	↑	RPB1	-0.01	9.15E-01	-
GPN3	0.27	5.34E-01	↑	RECQL5	-0.24	4.28E-02	↓
RPAP1	-0.16	2.69E-01	↓	GPN1	-0.22	8.51E-03	↓
RPAP2	0.5	8.76E-02	↑	GPN3	-0.96	7.39E-04	↓
RPAP3	-0.06	3.84E-01	-	RPAP1	0.04	6.96E-01	-
				RPAP2	-0.14	1.62E-01	↓
				RPAP3	-0.06	3.84E-01	-

This prompted me to change the strategy and as other published papers about GDOWN1 did, include a stress condition in my depletion experiments. I decide to UV expose the cells after GDOWN1 depletion. UV-damage of DNA leads to stalling of Pol II followed by its ubiquitination and degradation (Somesh *et al.*, 2005; Verma *et al.*, 2011; Wilson *et al.*, 2012). Therefore, turnover of Pol II is faster and would require a higher Pol II assembly rate in cytoplasm. In this case, if GDOWN1 is involved in regulation of Pol II in cytoplasm or its assembly I could observe some changes in protein abundance in cytoplasm or nucleus after applied stress to the cells. Initially, I had to confirm the presence of DNA damage. For this purpose, I established a dot-blot technique in the lab. I was able to detect presence of pyrimidine dimers (CPDs) in genomic DNA (gDNA) isolated from exposed cells and confirmed that the DNA damage is present after two different doses of UV_{260} exposure even after 4 hours of recovery (Fig. 39 a). Due to time reasons, this part of the project is not fully done and there is a plan to resume these experiments in future. In short, I plan to deplete GDOWN1 from cells by incubating with IAA. Afterwards, I would expose the cells to UV that would lead to DNA damage. Cells would be allowed to recover in presence of IAA. In case GDOWN1 is involved in Pol II regulation during stress, its absence could potentially cause changes in abundance of different proteins in cell compartments or their re-localization. I would apply methods of mass-spectrometry and immunohistochemistry to observe these changes (Fig. 39 b).

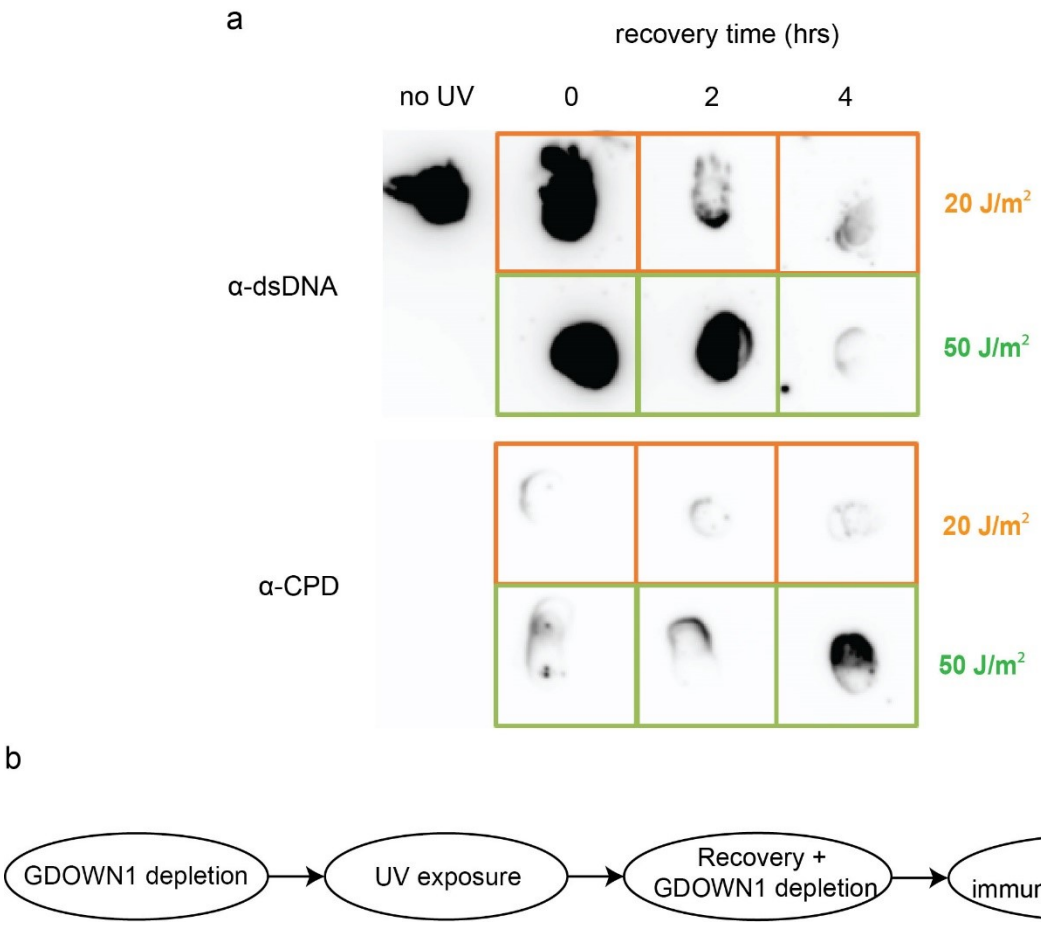


Figure 39: Confirmation of DNA damage of genomic DNA (gDNA) of UV-exposed K562 cells. (**a**) Dot-blot of gDNA isolated from K562 cells that were exposed to UV_{260} total dose of 20 J/m^2 or 50 J/m^2 and recovered for 0, 2 or 4 hours. Pyrimidines dimers were detected with anti-CPD antibody. Antibody against dsDNA served as internal control. (**b**) Schematic of proposed experimental design combing GDOWN1 depletion followed by UV exposure to better understand role of GDOWN1 in cells.

4. Discussion

Transcription and Pol II are essential for cells. The current literature is mainly focusing to better understand how the Pol II is engaged and regulated during the transcription of genes and how it unwinds from DNA. This process is taking place in nucleus, but before that, Pol II needs to be assembled in cytoplasm and then shuttled into nucleus. I speculate that during the time Pol II is in cytoplasm it needs to be regulated in order to be correctly assembled and not engage with other proteins like transcription factors or free mRNA. Even though assembly factors of Pol II have been partially described, there is still information missing about their mechanism and how they contribute to Pol II regulation in cytoplasm. The common phenotype for deleted or mutated assembly factors of Pol II is that RPB1 accumulates in cytoplasm and in most of the cases the cells are die (Carre & Shiekhattar, 2011). Some of these proteins might also act to prevent premature binding to free mRNA in cytoplasm as well as DNA in nucleus. Presence of GDOWN1 in these complexes also points to its additional role as regulator of cytoplasmic Pol II. Thus, I decided to better understand how assembled Pol II is regulated in cytoplasm and potentially during the nuclear shuttling.

In this project, I successfully solved the structure of mammalian cytoplasmic Pol I to high resolution. The structure shows that Pol II is bound to RPAP2 and GDOWN1 and has an open clamp conformation. Part of the cytoplasmic complex were also GPN proteins, but they were not resolved. Further experiments confirmed binding of GPN1-GPN3 complex to Pol II and their interaction with RPAP2. RPAP2 is stabilizing GPN1-GPN3 on Pol II. Based on the experiments with *in vitro* reconstituted complex of Pol II (G)-RPAP2-GPN1-GPN3, I hypothesize that a binding site for GPNs can lie in clamp of Pol II and it is inaccessible after clamp is closed. I propose that RPAP2, GDOWN1 and GPN proteins have function in regulating Pol II in cytoplasm by preventing aberrant interaction of transcription factors and free RNA with Pol II.

4.1. Pol II associates with assembly factors in cytoplasm and nucleus

First step in understating regulation of cytoplasmic Pol II was to identify proteins that are interacting with Pol II in this compartment. For this purpose, I generated K562 cell line endogenously expressing GDOWN1 tagged with GFP. As mentioned before in the text, GDOWN1 has very high affinity to Pol II and is abundant in cytoplasm (Hu *et al*., 2006, Zhu *et al*., 2022). Additionally, it was confirmed that GDOWN1 interacts with partially assembled Pol II (Forget *et al.*, 2010; Boulon *et al.*, 2010; Jeronimo *et al.*, 2007). Mass-spectrometry revealed that GDOWN1 associates with all Pol II subunits in cytoplasm as well as in nucleus. More importantly, I identified that assembly factors of Pol II were significantly enriched in both compartments. This suggests that Pol II bound to GDOWN1 is in majority fully assembled in cytoplasm and together with the assembly factors shuttled to nucleus. This is also supported by the observation that the largest Pol II subunit RPB1 was fully pulled-down from cytoplasm but not nucleus (Fig. 7 b). I hypothesize that GDOWN1 together with other assembly factors binds all Pol II in cytoplasm and the whole complex is transported to nucleus. In the nucleus, some of the Pol II can be engaged in transcription during which GDOWN1 is not interacting with Pol II anymore. GDOWN1 and assembly factors bound to Pol II in nucleus might also serve as regulators to prevent premature transcription initiation.

I identified also other proteins in pull-downs from nucleus that were enriched, but less compared to the assembly factors and Pol II subunits. There were several subunits of Mediator and DSIF subunits associating with GDOWN1 in nucleus. Since

there were data published that GDOWN1 is related to Mediator complex regulation during PIC assembly I hypothesized that a fraction of GDOWN1 interacts with Mediator subunits and that is the reason why it was enriched in this compartment (Hu *et al*., 2006, Zhu *et al*., 2022).

I also identified phosphorylation of GDOWN1 in position S270 that was present only in nucleus. Phosphorylation at position S270 weakens the affinity of GDOWN1 to Pol II (Guo *et al*., 2014). This is partially supported by the observation that in nucleus I identified GDOWN1-free Pol II (Figure 7 b). It represents the Pol II molecules, that are already engaged in transcription and GDOWN1 has already dissociated.

Another protein complex identified in the pull-downs, mainly in nucleus, was Integrator (INT). Since GDOWN1 was enriched in previous experiments in promotor paused regions and at the initiation site of Pol II, it is possible that when INT binds to paused Pol II, GDOWN1 can still weakly interact with it.

Interestingly, HSP90 was also enriched in cytoplasmic GDOWN1 pull-downs. HSP90 is involved in Pol II assembly through RPAP3 protein. RPAP3, as a part of RT2P complex, recruits Pol II subunits and serves as a linker between HSP90 and Pol II (Boulon *et al.*, 2010, Martino *et al.*, 2018). This observation suggests that small part of cytoplasmic Pol II might have been captured in partially assembled state.

I propose that Pol II in cytoplasm is bound to various assembly factors and GDOWN1 that stay on Pol II even after the nuclear import. Only when Pol II is recruited to promoter region of DNA these proteins dissociate from the Pol II.

4.2. Structural analysis of cytoplasmic Pol II complexes offered new insights in Pol II regulation

After successfully identifying interacting partners of cytoplasmic Pol II, I decided to structurally characterize the complex using cryoEM. It would offer more insight into how those interacting partners are regulating Pol II in cytoplasm. I was able to solve structure of endogenous cytoplasmic Pol II at nominal resolution of 2.7 Å. To my knowledge, this is the first time such structure was resolved. Cytoplasmic Pol II showed resolved core and partially resolved stalk of the Pol II and extra densities corresponding to GDOWN1 and RPAP2 proteins.

Structure of RPAP2 on cytoplasmic Pol II is consistent with previously published data. RPAP2 is bound to Pol II, mainly between jaws of RPB1 and RPB5 subunits, in downstream DNA region of Pol II. Similar to *in vitro* structure of Pol II-RPAP2, in cytoplasmic Pol II only N-terminal region of RPAP2 was resolved while C-terminal part of RPAP2 could not be resolved due to its high flexibility (Figure 16 a, Fianu *et al*., 2021, Wang *et al*., 2022).

However, structure of GDOWN1 in my cytoplasmic Pol II shows improvement compared to published data (Figure 16 a, Figure 19 a, Figure 20 b, Jishage *et al*., 2018). Previously published structure did not have assigned amino acid sequences for helices αII and αIII (Figure 40). My structure of GDOWN1 on cytoplasmic Pol II shows that α1, α2 and α3 helices contain amino acids between 21-43, 49-74 and 230-240, respectively. I was also able to model helix α4 of GDOWN1 further with assigned amino acids (Figure 40). All four helices were present in the density corresponding to GDOWN1 in structure of cytoplasmic Pol II (Figure 16 a, Figure 19 a, Figure 20 b).

Additionally, I did not observe any extra density corresponding to GDOWN1 binding to RPB1 dock region as previously described (Jishage *et al*., 2018).

GPN proteins were part of the cytoplasmic Pol II, but they were not resolved. A potential binding site for these proteins can be the clamp of Pol II that is present in the cryoEM density when it was filtered to lower resolution (Figure 16). Since there is not nucleic acid bound to Pol II in my structure, the clamp is in open state that causes high flexibility of this region and makes it challenging to resolve it. Thus, other methods were applied and are planned to confirm the binding site of GPNs in the clamp of Pol II.

While most of the studies put GDOWN1 and RPAP2 in context of regulating Pol II in nucleus before transcription initiation, here I show that Pol II is in complex with these proteins already in cytoplasm and they can act as regulators of Pol II in cytoplasm.

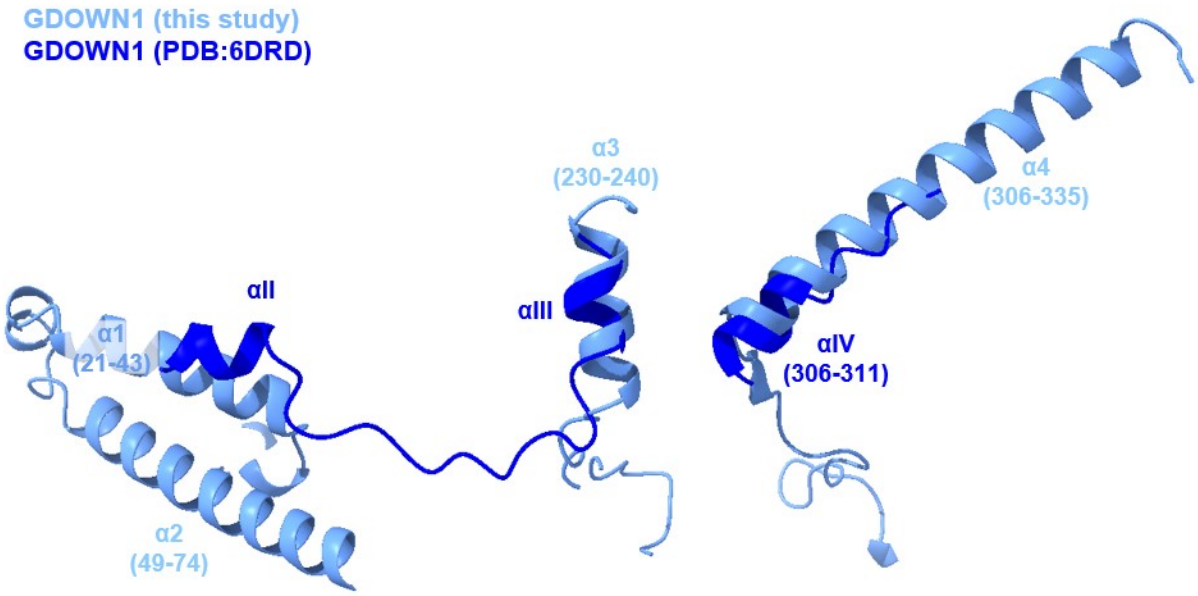


Figure 40: Superposition of cytoplasmic GDOWN1 structural model and published GDOWN1 structure (PDB: 6DRD, Jishage *et al*., 2018). Cytoplasmic GDOWN1 - light blue (only residues with pIDDT > 50 shown), GDOWN1 from Pol II (G) structure - blue.

4.2.1. Structure of cytoplasmic Pol II compared to pre-initiation complex

To better understand the regulatory effects of RPAP2 and GDOWN1 on Pol II in cytoplasm, I compared structure of cytoplasmic Pol II with available structures of Pol II during different stages of transcription.

First step of transcription is formation of PIC. Comparison of cytoplasmic Pol II and PIC structure revealed several steric clashes. In case of RPAP2, it was the same steric clash that has been previously reported (Wang *et al*., 2022). RPAP2 binds in the region of Pol II that interacts with TFIIF factor when PIC is formed (Figure 41). This can suggest that RPAP2 can also protect cytoplasmic Pol II from interacting prematurely with TFIIF in cytoplasm or in nucleus.

Position of GDOWN1 in cytoplasmic Pol II also revealed steric clashes with PIC. Helix $\alpha1_{GDOWN1}$ clashes with TFIIB present in PIC. Moreover, there is also small

part of predicted GDOWN1 structure that could clash with TFIIF factor in PIC (Figure 41).

The above-mentioned observations suggest that Pol II in cytoplasm and before engaging in transcription is unavailable to different initiation factors and RPAP2 and GDOWN1 can block these interactions. Additionally, I can speculate that GDOWN1 and RPAP2 has a functional role in regulation of cytoplasmic Pol II by helping to keep the clamp open or taking an advantage that the clamp of Pol II is open when no nucleic acids are bound.

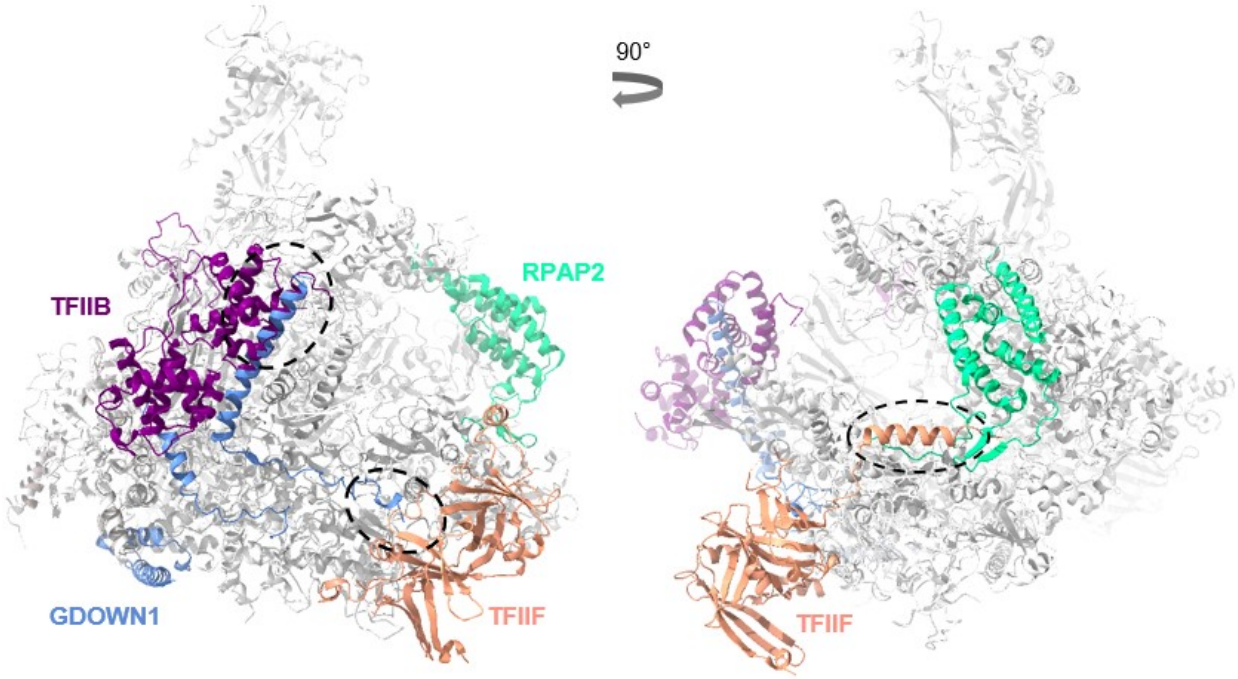


Figure 41: Comparison of cytoplasmic Pol II structure and PIC. Superposition of cytoplasmic Pol II structure into PIC (PDB: 7EG7, Chen *et al*., 2021). Some regions of PIC are omitted for clear view of the clashes. Pol II of PIC is in grey, highlighted are GDOWN1 (cornflower blue), RPAP2 (green), TFIIF (orange), TFIIB (violet). Clashes are indicated in dash circles.

4.2.2. Comparison of cytoplasmic Pol II and elongation complexes

After successful initiation of transcription, Pol II enters productive elongation. I observed a steric clash of RPAP2 is with downstream DNA bound to transcribing Pol II, which is again in agreement with published Pol II-RPAP2 structure (Figure 42, Wang *et al*., 2022). Another clash that is consistent with published data is with the clamp region of RPB1 that is in closed confirmation in the structure of transcribing Pol II (Figure 42, Wang *et al*., 2022). GDOWN1 in cytoplasmic Pol II also revealed steric clash with transcribing Pol II structure. Helix α1 would be in way of upstream DNA bound to Pol II (Figure 42). Interestingly, it was observed that Pol II (G) is transcriptionally active *in vitro* when tailed DNA template was used (Hu *et al*., 2006).

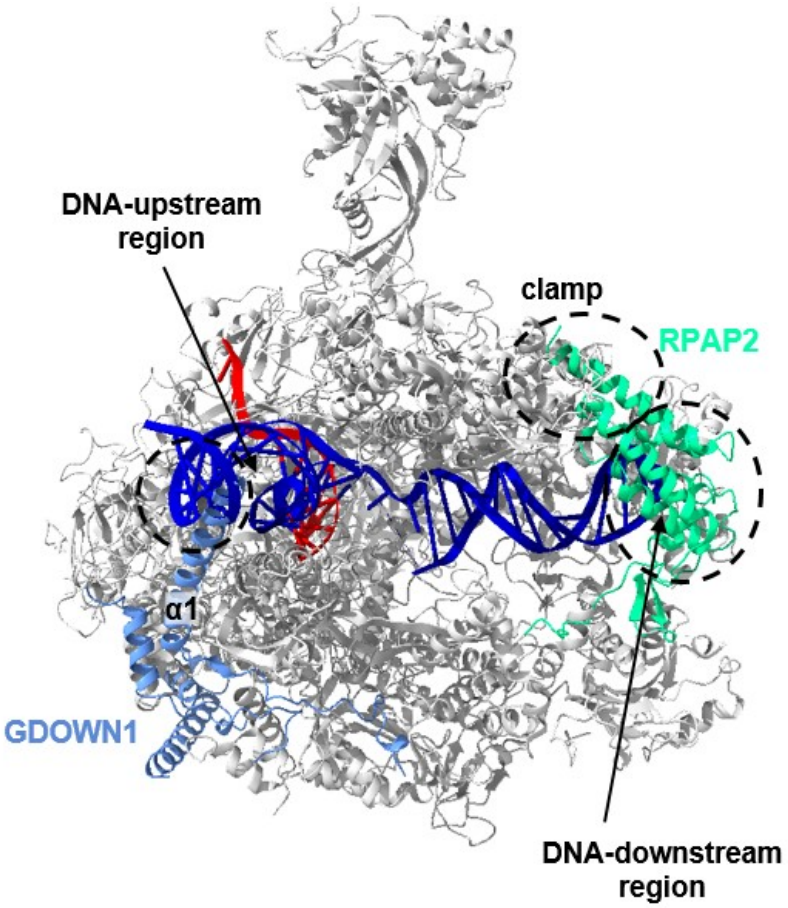


Figure 42: Structure of cytoplasmic Pol II with transcribing Pol II superimposed (PDB: 5FLM, Bernecky *et al*., 2016). Transcribing Pol II is in grey, highlighted are GDOWN1 (cornflower blue), DNA (blue), RNA (red), RPAP2 (green). Clashes are indicated in dash circles.

Furthermore, comparing the structure of cytoplasmic Pol II and paused Pol II in front of nucleosomes shows multiple clashes of GDOWN1 with components of PAF complex (Farnung *et al*., 2022, Figure 43). First clash is between helix $\alpha 4_{GDOWN1}$ with LEO1. Second clash is between helix $\alpha 3_{GDOWN1}$ and RTF1. There are also clashes between helix $\alpha 1_{GDOWN1}$ and PAF1 and RTF1 (Figure 43). There is again steric clash between RPAP2 and down-stream DNA of paused Pol II (Figure 43).

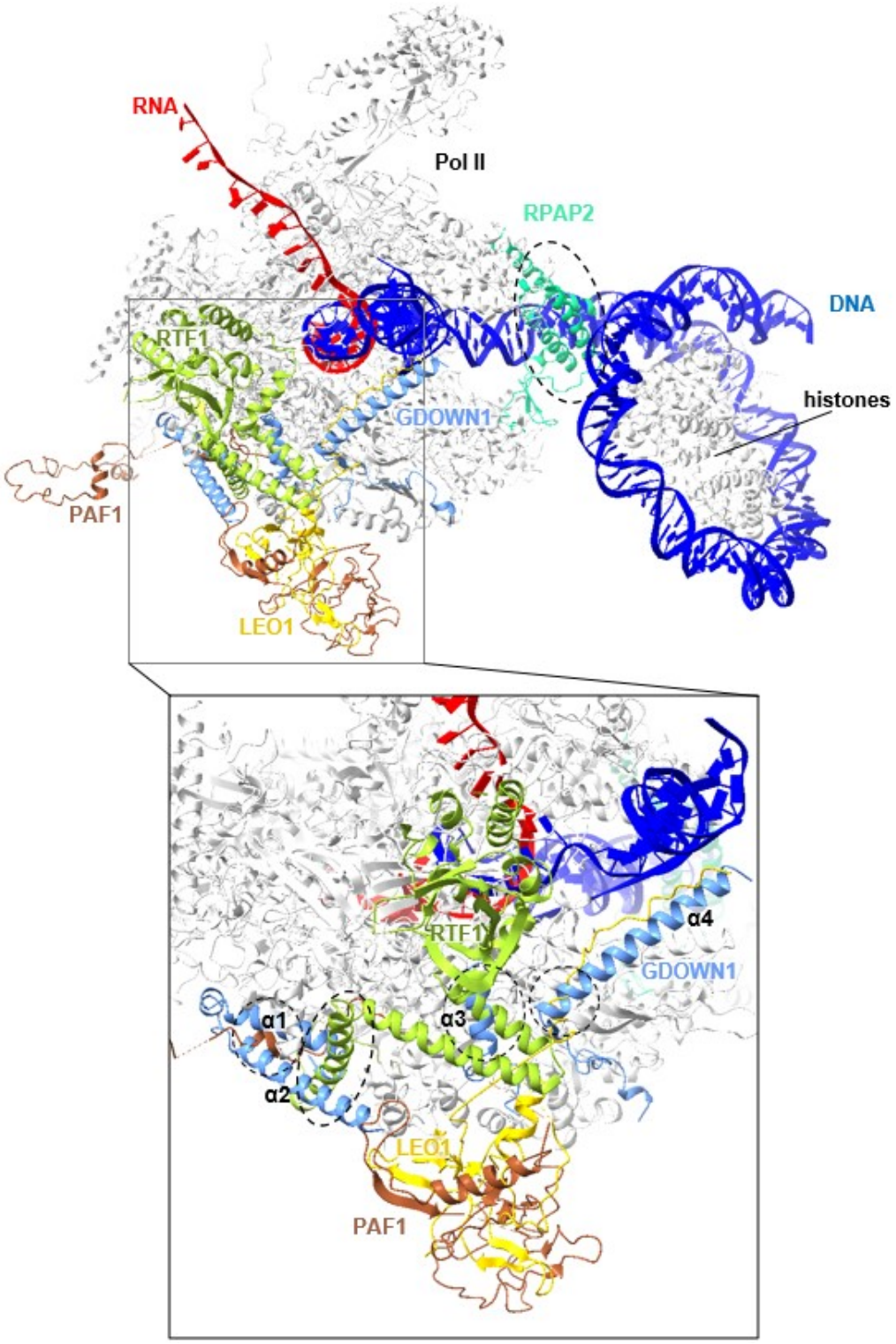


Figure 43: Structure of cytoplasmic Pol II with elongating Pol II superimposed (PDB: 8A3Y, Farnung *et al.*, 2022). Elongating Pol II and histone is in grey, highlighted are GDOWN1 (cornflower blue), DNA (blue), RNA (red), RPAP2 (green), PAF1 (brown), LEO1 (yellow), RTF1 (olive green). Clashes are indicated in dash circles.

4.2.3. Cytoplasmic Pol II and termination of transcription

Previously published data showed that GDOWN1 also prevents binding transcription termination factor TTF2 to Pol II *in vitro* (Guo *et al.*, 2012). Since structure of TTF2 with Pol II has not been yet resolved, AF3-predicted structure of Pol II-TTF2 was

used to compare it with structure of cytoplasmic Pol II (Figure 44). There can be several clashes between all 4 α-helices of GDOWN1 and TTF2 bound to Pol II (Fig. 44 a). These regions of TTF2 are also predicted with high confidence scores (Fig. 44 b). Most prominent clashes are between helix $\alpha4_{GDOWN1}$ and β-sheet region of TTF2 (Fig. 44 a). Another clash can be observed between helices $\alpha1\text{-}2_{GDOWN1}$ and TTF2 region 412-460 (Figure 44 a, c). Interestingly, N-terminal region and C-terminal region of GDOWN1 containing α1-2 and α3-4 respectively are important for blocking TTF2 binding to Pol II (Guo *et al.*, 2012). One can also notice that $\alpha1\text{-}2_{GDOWN1}$ and TTF2 region 412-460 are structurally very similar and they both contain a highly conserved LPDKG sequence (Fig. 44 c). It was showed previously that when these amino acids were deleted in GDOWN1 protein, it lost its ability to block TTF2 binding to Pol II (Guo *et al.*, 2012). These above mentioned observations supported by the published literature show evidence that the TTF2 structure prediction is correct.

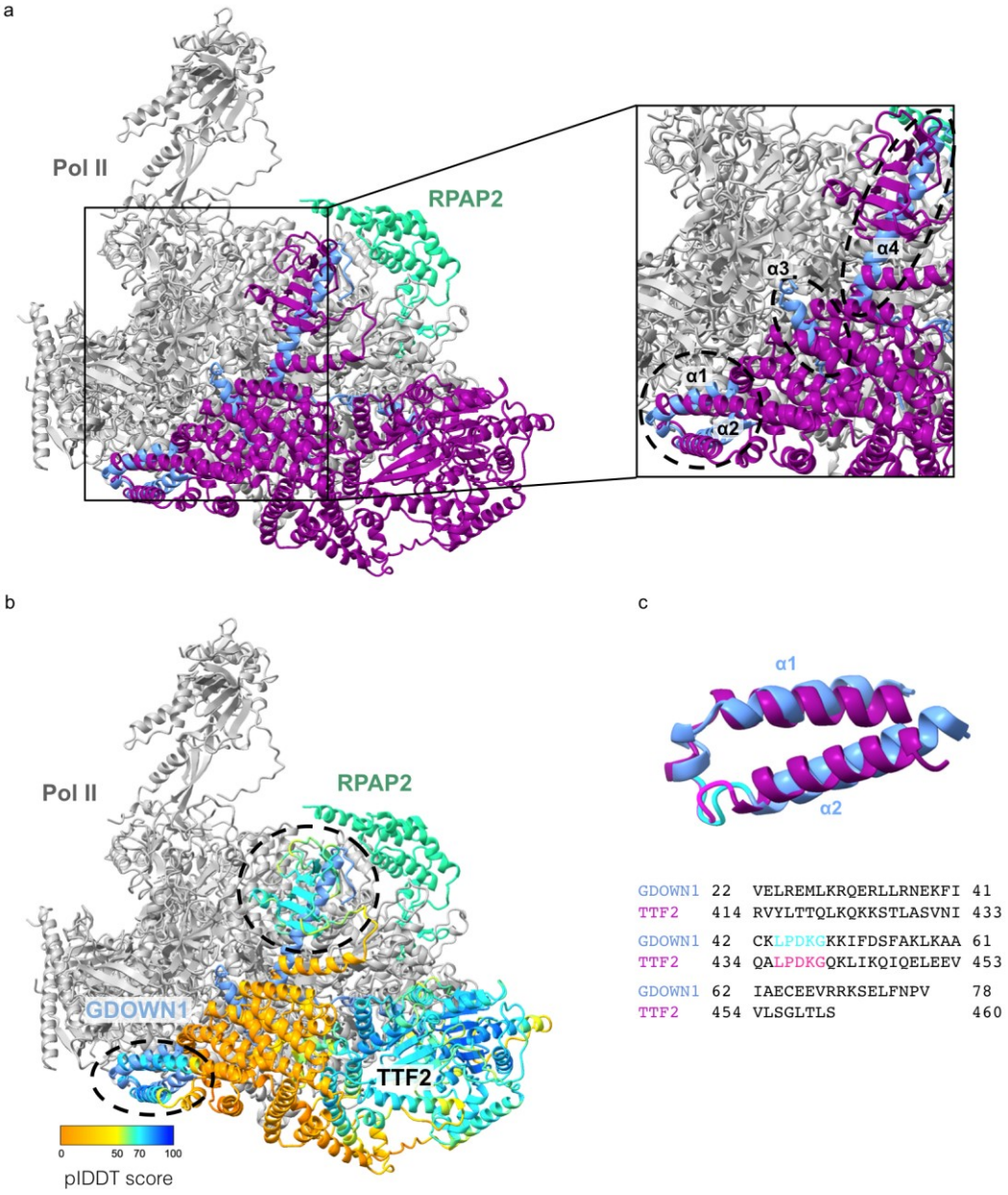


Figure 44: Comparison of structure of cytoplasmic Pol II with Pol II-TTF2 superimposed. (**a**) Pol II-TTF2 complex was predicted with AF3 (Abramson *et al.*, 2024). Pol II is in grey, highlighted are GDOWN1

(cornflower blue), RPAP2 (green) and TTF2 (violet). α1-4$_{GDOWN1}$ helices are labelled. Clashes indicated in dash circles. (**b**) TTF2 structural model predicted by AF3 coloured by the pIDDT scores. Pol II in grey, GDOWN1 in cornflower blue, RPAP2 in green. Clashes of TTF2 regions with high pIDDT scores and GDOWN1 marked with dashed circles. (**c**) Structure of GDOWN1$_{α1-2}$ (22-78, cornflower blue) and TTF2 (414-460, violet) superimposed and amino acid sequence alignment of the regions. Conserved LPDKG sequence in GDOWN1 and TTF2 highlighted in cyan and magenta respectively. Visualized with ChimeraX (Pettersen, 2021).

These observations of cytoplasmic Pol II compared to structures of Pol II in different stages of transcription suggest that proteins bound to cytoplasmic Pol II might block premature binding of DNA or RNA and different transcription factors to Pol II in cytoplasm and before engaging in transcription.

4.3. Pol II clamp as a possible binding site of GPN1-GPN3

One of the most striking difference between cytoplasmic Pol II and other published Pol II structures that are already involved in transcription is the clamp. Structures of Pol II engaged in transcription have clamp in a close state (Bernecky *et al*., 2016, Farnung *et al*., 2022). Pol II that is not bound to transcription factors or DNA has clamp in an open state (Cramer *et al*., 2001). Since in my experiments there is no DNA-RNA scaffold or transcription factors but there are other proteins bound, one can assume that clamp is not in a closed state. This causes high flexibility of this region that makes it hard to resolve it into higher resolution. When the structure of cytoplasmic Pol II was filtered to lower resolution there was an extra density localized in clamp region of Pol II (Figure 16 b).

Since no extra density of GPNs was resolved on cytoplasmic Pol II, it prompted me to supplement the structure with XL-MS. Due to high complexity of the sample this analysis proved to be rather challenging. After several attempts and different tested strategies, XL-MS of cytoplasmic Pol II did not reveal any interactions between Pol II and GPN1 and GPN3 that would supplement the structure that I obtained.

I further used *in vitro* reconstituted complexes of Pol II with GDOWN1, RPAP2, GPN1 and GPN3 for characterization of the molecular mechanisms of Pol II regulation in cytoplasm.

In previous studies focused on GPNs and Pol II interactions, they showed that GPN1 is interacting with CTD of Pol II. Contrary to their observations, I did not see binding of GPN1-GPN3 complex to CTD of Pol II (Carre & Shiekhattar, 2011). One of the explanations can be that the experimental design of the published studies is different from mine. I was working with pure proteins, while in previous publications they included only one affinity step to purify GPNs or hCTD. There could have been other proteins in their pull-downs that could contribute to the observed interaction with CTD, for example RPAP2. RPAP2 binds CTD and has stable and high affinity to GPN1-GPN3 complex based on my experiments and published data (Figure 25; Egloff *et al*., 2012, Forget *et al*., 2013).

Nevertheless, I was able to observe the binding of GPN1-GPN3 complex to Pol II, but this interaction seemed to be of low affinity and GPN1-GPN3 tend to dissociate from the complex when harsher separation methods are applied (Figure 24 a, b). This was overcame by adding RPAP2 to the Pol II-GPN1-GPN3 complex suggesting that RPAP2 is stabilizing GPN1-GPN3 on Pol II complex (Figure 29, Figure 30).

I also reconstituted *in vitro* complex of Pol II, GDOWN1, RPAP2 and GPN1-GPN3 for structural studies. This approach reduced the complexity of the endogenous cytoplasmic Pol II. It also allowed me to work with pure proteins in higher concentration and I did not have need carbon support layer for grid preparation. I am in process of solving the cryoEM structure of *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 complex. Preliminary results from a small dataset collection of micrographs show that the structure of *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 complex is almost identical to cytoplasmic Pol II structure (Figure 16, Figure 32). *In vitro* Pol II complex has improved GDOWN1 density for helices α1-4 that match the predicted structure of GDOWN1 (Figure 32). Possible explanations for this improvement are that first, I was not limited in protein amount as much as in case of cytoplasmic Pol II. It allowed me to apply proteins directly on grids without using a support layer. Second, *in vitro* reconstituted complex of Pol II, GDOWN1, RPAP2 and GPN1-GPN3 is homogenous and more stable, as the preparation time is much shorter than in case of the cytoplasmic Pol II samples. Third, direct application of sample on grids without using carbon layer could improve particle orientation and contrast. Similar to the cytoplasmic Pol II structure I did not resolve density corresponding to clamp of Pol II or GPN1-GPN3 (Figure 32).

XL-MS of the *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 complex revealed important cross-links. Several cross-linked peptides between GPN1 and clamp region of RPB1 subunit were identified (Figure 34 b). The cross-links were mapped into the structure of cytoplasmic Pol II. Most of the identified cross-links were within the range of maximal theoretical Cα-Cα distance of BS3, 25 Å (Figure 35 a, b, Gong *et al*., 2020). There are few cross-links slightly above 25 Å and one outlier around 100 Å (Figure 35 a, b). Since the complex was not separated on size-exclusion column prior the XL-MS, there could be present Pol II dimers. Indeed the outliers between the stalk and RPB2 subunit have Cα-Cα distance less than 25 Å when mapped to Pol II dimer structure (Figure 35 b, c).

Unresolved clamp of Pol II and identification of cross-linked peptides between GPN1 and Pol II clamp region suggests that GPN1-GPN3 complex can interact with the open clamp of Pol II. This is also supported by the observation that GPN1-GPN3 does not interact with Pol II EC (Figure 24 c). I argue that once the Pol II clamp closes around the DNA-RNA scaffold, GPN1-GPN3 cannot interact with the clamp residues because they are not available anymore in this conformation. As mentioned above, cytoplasmic Pol II structure has unresolved clamp region. I speculated that this could be a potential binding site for GPNs as I did not observe any extra density in my structures corresponding to them.

Based on the structural and biochemical experiment in this project I hypothesize that GDOWN1, RPAP2 and GPN1-GPN3 serve not only as assembly factors of Pol II in cytoplasm but also play a key role in its regulation. They can prevent pre-mature binding of transcription factors to Pol II in cytoplasm and nucleus. Proteins GDOWN1, RPAP2 and GPN1-GPN3 might work together in cytoplasm to keep Pol II clamp open. They also prevent binding various transcription factors to Pol II. In addition, keeping Pol II clamp in open state may prevent stable interaction with free mRNAs in cytoplasm. mRNA transported from nucleus to cytoplasm freely diffuses in this compartment unless it is located by ribosomes and immediately translated or it is bound to specific anchor proteins that will locate it in specific region of cytoplasm (Shahbabian *et al*., 2011). This feature is especially important during cell division and developmental stages. mRNA in cytoplasm has poly(A) tail and 5'-cap that prevents its degradation

by different pathways (Siwaszek *et al*., 2014). However, there is not much know if Pol II binds free mRNA in cytoplasm or if there is a mechanism to prevent it. More experiments need to be done to test this hypothesis. For example competitive binding assays of GPN1-GPN3 complex and RPAP2 with Pol II and adding DNA-RNA scaffold or other RNA. Another proposed experiment can be testing interaction of GPN1-GPN3 and clamp region of Pol II.

4.4. Prediction of RPAP2-GPN1-GPN3 complex

Previously, it was reported that GPN1 interacts mainly with C-terminal part of RPAP2. I decided to take a closer look at these interactions and include GPN3. The interaction between RPAP2 and GPN1-GPN3 was confirmed (Figure 25). I also confirmed GTPase activity of the GPN1-GPN3 complex (Figure 23 e). Previously published study showed that only GPN1 and not GPN3 has GTPase activity, even though critical residues for nucleotide binding are conserved between these two GPNs (Carre & Shiekhattar, 2011, Niesser *et al*., 2016). Thus, I can assume that GPN1 contributes to GTPase activity in the complex of GPN1-GPN3.

Similar to published data I did not see interaction between N-terminal region of RPAP2 and GPN1-GPN3 complex (Figure 28 c). This observation was previously made only with GPN1 (Forget *et al*., 2013). However, XL-MS of RPAP2-GPN1-GPN3 complex revealed not only cross-links between C-terminal region of RPAP2 and GPN1 and GPN3, but also between RPAP2 NTD region (Figure 27 a). Since I observed abolished binding of GPN1-GPN3 complex to RPAP2 (1-215), I conclude that GPN1-GPN3 complex does not have strong affinity to RPAP2 NTD region. Identification of cross-links in this region can be explained by close packing of the proteins when they are not bound to Pol II. As mentioned before GPN1 is important for RPAP2 nuclear exit (Forget *et al*., 2013). It suggests that RPAP2-GPN1-GPN3 complex is recycled to cytoplasm together after dissociating from Pol II in a packed form.

Other argument that interaction between GPN1-GPN3 complex and RPAP2 N-terminal region is rather weak are the results of *in vitro* Pol II (G)-RPAP2-GPN1-GPN3 complex XL-MS. There was only one cross-link identified between NTD of RPAP2 and GPN1 and none for GPN3 (Figure 34 b). This region of RPAP2 is bound to DNA-downstream region of Pol II. Presence of this one cross-link can be also explained that the *in vitro* complex preparation for XL-MS did not include additional sizing step and there could be some unbound RPAP2-GPN1-GPN3 complexes that contributed to identification of this cross-link.

AF3 was used to predict the structure of RPAP2-GPN1-GPN3 (Figure 26). The identified cross-links in the complex were mapped to predicted structure (Figure 27 b, c). It shows that some Cα-Cα distances are within the 25 Å that is the maximal distance for DMTMM (Leitner *et al*., 2014). This observation suggests that the predicted parts with Cα-Cα distance higher than 25 Å can be positioned differently in the real structure. To improve the structure prediction of RPAP2-GPN1-GPN3 complex, one can also employ AlphaLink, which integrates the results from XL-MS into AlphaFold2. This combined approach can yield better predictions of protein structure (Stahl *et al.*, 2023). Interestingly, the predicted model favors closed confirmation of GPN1 (Fig. 44 a). Superposition of predicted RPAP2-GPN1-GPN3 structure with open conformation of yeast Gpn1 shows steric clash between one of the RPAP2 helices and insertion region of Gpn1 (Fig. 44 b). It is possible that GPN1 in complex with RPAP2 is in closed conformation.

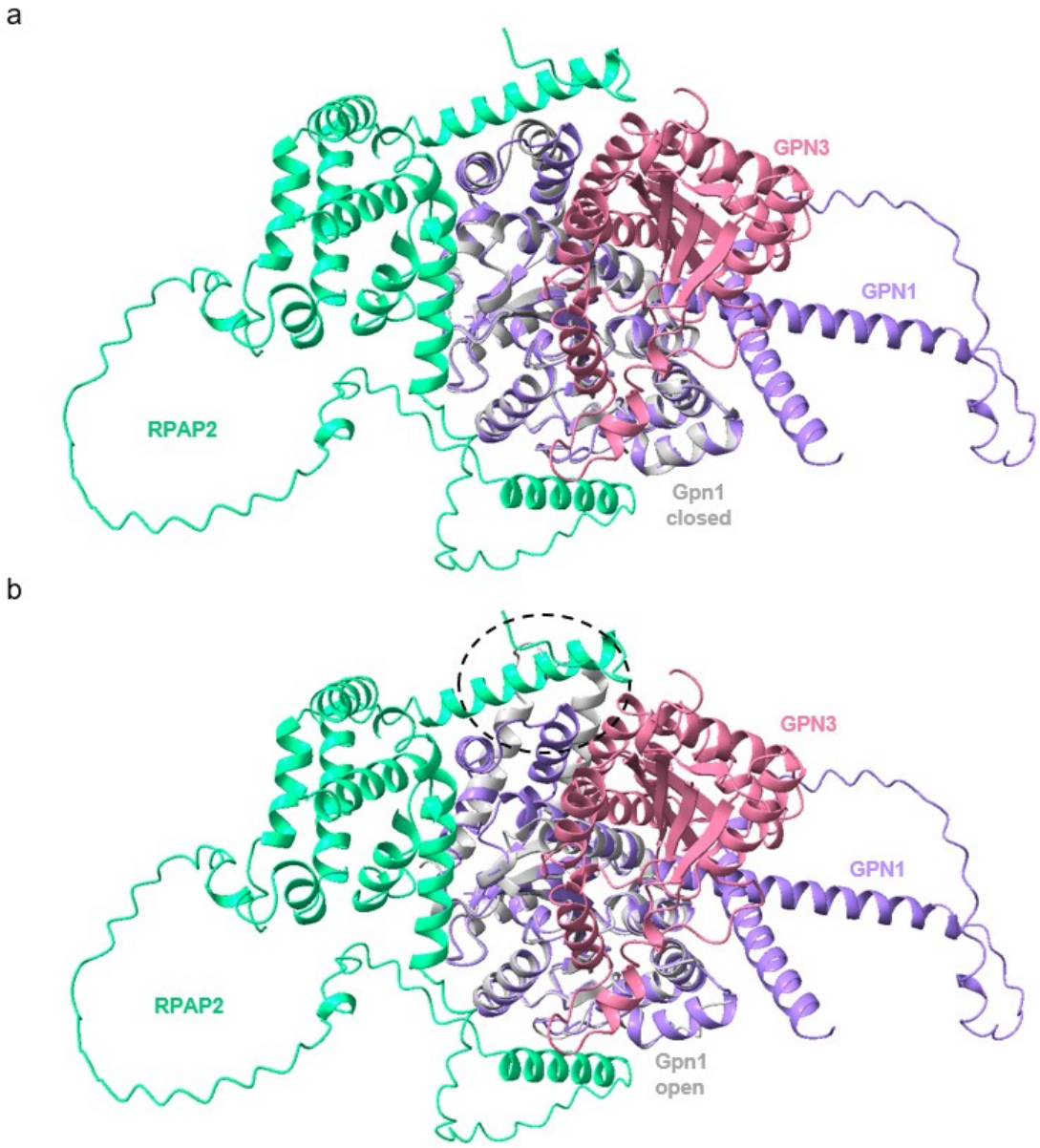


Figure 45: Overlay of RPAP2-GPN1-GPN3 AF3 prediction and yeast Gpn1 crystal structure. (**a**) Gpn1 in closed state in grey (PDB: 5HCI, Niesser at al., 2016), RPAP2 in green, GPN1 in violet, GPN3 in pink. (**b**) Gpn1 in open state in grey (PDB: 5HCN, Niesser *et al*., 2016), RPAP2 in green, GPN1 in violet, GPN3 in pink.

Identification of cross-links and prediction of RPAP2-GPN1-GPN3 complex made it possible to the design mutants of RPAP2 and GPN1-GPN3 in the predicted binding regions. Charged residues of RPAP2 K340D, K509E, K512E, R544D, K549E and GPN1 D195K, E201K, D205K and E216R were mutated to residues with opposite charge. iPTM score of the mutant complexes showed decrease, while rescue mutant complex reached the same iPTM score as wt (Figure 26 a, Figure 28 a, b). I plan to test these mutants to characterize the binding site of RPAP2-GPN1-GPN3 in more detail.

4.5. Role of hIWR1 as Pol II nuclear importer not confirmed

Nuclear importer proteins often facilitate nuclear transport of multisubunit protein complexes (Wang & Brattain, 2007). These proteins usually contain NLS. Pol II subunits do not have one. In yeast, Iwr1 protein facilitates Pol II import from cytoplasm

to nucleus. Iwr1 protein contains NLS and binds to Pol II. There is a human homolog of Iwr1 (hIWR1) which function was not yet characterized. As shown in the results, I tried to test the interaction of purified hIWR1 and Pol II with various methods, but I did not see any binding (Figure 36, Figure 37). In addition, I did not identify hIWR1 in the GDOWN1 pull-downs in cytoplasm or in nucleus. It can also mean that this interaction is transient and depends on some other post-translational modification of either hIWR1 or Pol II assembly factors, or there is different nuclear importer protein in mammals for Pol II nuclear transport.

Nevertheless, hIWR1 protein can have other essential role in cells, as the attempts to created GFP-tagged hIWR1 human cell line failed due to nonviable cells.

4.6. Function of GDOWN1 as cytoplasmic Pol II regulator in cells remains unknown

Even though a substantial effort was invested into unraveling role of GDOWN1 in cells, I obtained mostly inconclusive results. Some of my observations were consistent with published literature. Specifically, GDOWN1 is important for cells and its depletion leads to reduced viability (Figure 38 b). Short GDOWN1 depletion did not show drastic difference in abundance of proteins between cytoplasm and nucleus (Table 1). On the other hand, longer depletion of GDOWN1 revealed some significant differences in the abundance of proteins of interest but I cannot exclude a secondary effect of longer GDOWN1 depletion with this experimental design (Table 1).

Since there are couple of published articles showing role of GDOWN1 in stress response, I included combination of GDOWN1 depletion and UV exposure into the experimental design. I reasoned that UV exposure would lead to damage of DNA. This causes stalling of Pol II which is further ubiquitinated and degraded (Somesh *et al.*, 2005; Verma *et al.*, 2011; Wilson *et al.*, 2012). I hypothesized that as a result, turnover of Pol II is faster and would require a higher Pol II assembly rate and regulation of Pol II in cytoplasm. Experiments to optimize the UV exposure and the treatment and recovery time were initiated but could not be finished in time.

In future, there is a potential that these experiments will yield interesting results.

4.7. Summary and future directions

In summary, this project contributed to better understanding the molecular mechanism of Pol II regulation in cytoplasm. The structure of cytoplasmic Pol II was resolved to high resolution. The key differences between Pol II in cytoplasm and Pol II engaged in various steps of transcription were characterized. Based on the structural and biochemical data, I propose a model of Pol II regulation in cytoplasm. GDOWN1, RPAP2 and GPN1-GPN3 are interacting with Pol II in cytoplasm. These proteins prevent binding of transcription factors and free mRNA to Pol II in cytoplasm. GPNs are stabilized on Pol II by RPAP2 and might mainly bind to clamp region of Pol II further inhibiting mRNA binding in cytoplasm and potentially DNA in nucleus. In addition, GDOWN1, RPAP2 and GPNs might prevent closing of clamp and keeping it in open state. Further experiments and more extensive processing of cryoEM data of *in vitro* reconstituted Pol II (G)-RPAP2-GPN1-GPN3 are needed to confirm the proposed model of cytoplasmic Pol II regulation.

Unfortunately, I was not able to resolve density of GPNs in both cytoplasmic Pol II or *in vitro* reconstituted Pol II (G)-RPAP2-GPN1-GPN3 structure. Nevertheless,

all results that I showed so far suggest that the binding site for GPNs is in clamp region. For this purpose, I plan to test interaction of GPN1-GPN3 complex with purified recombinant clamp region. The clamp parts of Pol II have been cloned into insect cells expression vector. I plan to co-express and purify it. I propose testing the interaction of GPN1-GPN3 complex with clamp using pull-downs with immobilized GPN1-GPN3 or mass photometry.

Regulation of Pol II in cytoplasm might also include preventing free RNA binding to the protein complex. To further characterize how is the affinity of the cytoplasmic Pol II to free RNA affected, I suggest following experiments. I plan to prepare various labelled RNA molecules of different lengths and test their interaction with Pol II (G)-RPAP2-GPN1-GPN3 complex using gel mobility shift assay. I hypothesize that GDOWN1, RPAP2 and GPN1-GPN3 can reduce the affinity of Pol II to RNAs. These proposed experiments can expand the knowledge about the role of these proteins in regulation of Pol II in cytoplasm.

Furthermore, role of GDOWN1 in cytoplasm remains unknown and I would like to contribute to this part of the research. As showed above, I initiated pilot experiments with combined depletion of GDOWN1 and UV-exposure. I propose experiments where I first deplete GDOWN1 from cells and then UV-expose them followed by recovery for specific time. Then I would test abundancy of proteins in different compartments by Western blot, immunohistochemistry or mass-spectrometry.

This research has a potential to improve our understanding of Pol II regulation before it is shuttled into nucleus and engaged in transcription. As mentioned above, cancer cells have dysregulated transcription and often increased transcription levels (Bradner *et al*., 2018). It is important to know all the regulatory steps of Pol II even before it starts transcription, so better drug design strategies can be applied. Structural and biochemical characterization of cytoplasmic Pol II complex can help identify promising sites that can be targeted by inhibitors disrupting the regulatory path of Pol II in cytoplasm. This can lead to reduced transcription rate in cancer cells.

5. Methods

Vectors and sequences:

The human homolog of Iwr1 SLC7A6OS (hIWR1; UniProt Q96CW6), GPN1 (UniProt Q9HCN4), GPN2 (UniProt Q9H9Y4), GPN3 (UniProt Q9UHW5) and RPAP2 (Q8IXW5) open reading frames (ORFs) were first amplified from cDNA and separately cloned into pOPINB expression vector using In-Fusion seamless cloning. Shortly described, each amplified ORF has around 15-bp extension that are complementary to the vector end. The pOPINB vector was cleaved with KpnI and HindIII restriction enzymes (New England Biolabs). Both insert and linearized vector were separated on agarose gel, isolated and then incubated together with In-Fusion® Snap Assembly Master Mix (Takara) for 15 minutes at 50°C. This ensured that the 15bp overlaps at the ends of vector and insert were fused. Chemically competent *E. coli* DH5α cells were transformed with the mixture, plasmids isolated and verified by sequencing for correct insertion of ORFs into the vectors. hIWR1, GPN1, GPN2 and GPN3 are tagged on N-terminus with 6x-histidine(His) tag. RPAP2 contains N-terminal 6x-His-MBP tag. All constructs also contain a 3C protease cleavage site after the tags.

GFP-nanobody plasmid was obtained from Addgene (#49172). It contained C-terminal 6xHis tag. Additionally, two lysines (KK) were added on C-terminus of the GFP-nanobody before 6xHis tag using site-directed mutagenesis (SDM) using two single primer PCRs in parallel (Edelheit *et al*., 2018). Briefly, two PCRs are run in parallel with each one of the complementary primers (forward and reverse) containing the mutations of interest. The reactions are then combined, denatures, annealed and digested by DpnI (New England Biolabs) to remove the parental plasmid. Reannealed plasmids were transformed into competent DH5α cells. Correct insertion of two lysines was confirmed by sequencing.

Furthermore, ORFs of GPN1, GPN2, GPN3 and RPAP2 were cloned into 438A (GPN1, GPN2, GPN3) and 438C (RPAP2) insect cells expression vectors, using In-Fusion seamless cloning as described above. Vectors and inserts were amplified by PCR. Afterwards, GPN1 and GPN3 ORFs were cloned into the same 438A vector using ligation-independent cloning (LIC; Gradia *et al*., 2017). Shortly, destination plasmid (438A-HisGPN1) was cleaved by SwaI (New England Biolabs). GPN3 was cleaved out of 438A-GPN3 plasmid using PmeI enzyme (New England Biolabs). Destination vector with 6xHis-GPN1 and GPN3 insert were treated with T4 DNA polymerase and single nucleotide (dGTP or dCTP) mixed together and transformed into DH5α cells. Isolated plasmid were confirmed by sequencing.

Truncated version of RPAP2 contained 1-215 amino acids (RPAP2trun). Mutant was prepared by round-the-horn mutagenesis approach. Shortly, primers were designed to have opposite direction and border with the sequence to be deleted. Plasmid was amplified by PCR, ends phosphorylated and ligated, then transformed into DH5α cells. Positive clones were confirmed by sequencing.

A plasmid with the human Pol II CTD (hCTD, 1593-1970 residues of RPB1) was a kind gift from P. Cramer (Max Planck Institute for Biophysical Chemistry, Göttingen). A plasmid encoding MBP protein was constructed by deleting hCTD using round-the-horn mutagenesis methodology as described above.

For specific primer sequences see table list of oligos attached.

Protein expression and purification:

Protease inhibitors (PIs) included 1 mM PMSF, 1 mM benzamidine, 60 µM leupeptin, and 200 µM pepstatin. All steps were performed at 4°C unless otherwise specified.

Denatured proteins mixture was prepared as followed. *E. coli* BL21(DE3)RIL culture was grown over-night in LB medium (30 ug/mL chloramphenicol) at 37°C. Cells were collected, centrifuged and resuspended in respective lysis buffer based on the protein purification. Cell mixture was sonicated on ice and the incubated at 65 °C for 15 minutes to denature the proteins. Denatured protein were centrifuged in table-top centrifuge at max speed, 30 minutes. Proteins were divided into 20mg aliquots, flash frozen in liquid nitrogen (LN) and stored at -80°C.

GFP-nanobody with KK linker was expressed in *E. coli* BL21(DE3)RIL cells in LB medium supplemented with 30µg/mL ampicillin and 100µg/mL chloramphenicol. The culture was grown to OD_{600} 0.6-0.8 at 37°C and then protein expression was induced by 0.5 mM IPTG for 16 hrs at 18°C . Cells were collected and resuspended in NaPi buffer (50 mM phosphate buffer pH 7.4 @ RT, 200 mM NaCl) with addition of 1x PIs and 10 mM imidazole pH 8.0 @ RT, then sonicated on ice. Lysate was centrifuged (50 000g, 4°C, 50 min) and supernatant was applied on HisTrap column HP (1 mL) equilibrated in NaPi buffer with 10mM imidazole. Column was washed with 5CV of NaPi buffer with 30mM imidazole. Column was then incubated for 10 minutes with ATP Wash buffer (NaPi buffer, 30 mM imidazole pH 8.0 @ RT, 5mM ATP, 2mg/mL denatured proteins) at RT to remove excess of chaperones and then put back to 4°C. GPN-nanobody was eluted over gradient of imidazole pH 8.0 @ RT in NaPi buffer (30 mM to 500 mM) over 10CV. Peak fractions were checked on SDS-PAGE gel, pooled and concentrated to 500uL using a 10,000 MWCO Amicon Ultra Centrifugal Filter (Merck Millipore). GFP-nanobody was applied on size exclusion columns Superdex Increase 75 10/300 GL (Cytiva) equilibrated in Coupling buffer (50 mM sodium phosphate pH 7.2 @ RT, 150 mM NaCl). Peak fractions were assessed by SDS-PAGE and fractions containing GFP-nanobody were pooled and concentrated using 10,000 MWCO Amicon Ultra Centifugal Filter (Merck Millipore). Size exclusion buffer for GFP-nanobody with cysteine linker contained 50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5mM EDTA pH 8.5.

hIWR1 protein was expressed in *E. coli* BL21(DE3)RIL cells in LB medium supplemented with 30µg/mL ampicillin and 100µg/mL chloramphenicol. The culture was grown to OD_{600} 0.6-0.8 at 37°C and then protein expression was induced by 0.5 mM IPTG for 16 hrs at 18°C . Cells were collected and resuspended in Buffer A (20 mM Tris-HCl pH 8.0 at RT, 300 mM NaCl, 10% glycerol, 10 mM imidazole pH 8.0 at RT, 2 mM DTT) with 10 mM imidazole pH 8.0 at RT and 1x PIs. Cells were sonicated on ice. Lysates were centrifuged for 50 minutes, 50 000g and supernatant was applied on HisTrap column HP 1 mL (Cytiva) equilibrated in Buffer A 10mM imidazole. Column was washed with 5CV of Buffer A with 30 mM imidazole. Bound protein was eluted over linear gradient of 10 CV Buffer A with 30 mM to 500 mM imidazole. Peak fractions were checked on SDS-PAGE gel, pooled and dialyzed in Buffer A with 3C protease over-night. Cleaved protein was applied on HisTrap column equilibrated in Buffer A. Flow-through was collected and concentrated using 10,000 MWCO Amicon Ultra Centrifuge Filter (Merck Millipore) to 500uL. hIWR1 was applied on size exclusion column Superdex Increase 200 10/300 GL (Cytiva) equilibrated in Buffer A. Peak fractions

were assessed by SDS-PAGE and hIWR1 containing fractions were pooled and concentrated using 30,000 MWCO Amicon Ultra Centrifugal Filter (Merck Millipore). Protocol modified from Czeko *et al*., 2011.

GPN1-GPN3 proteins were co-expressed in insect cells. The expression was performed by Protein Facility at ISTA. Pellets were resuspended in Lysis buffer (20 mM HEPES pH 7.9 at RT, 200 mM NaCl, 1 mM DTT, 30 mM imidazole pH 8.0 @ RT, 1xPIs). Cells were sonicated on ice and cleared using ultra-centrifuge. Cleared lysate was applied on HisTrap HP 1 mL column (Cytiva) equilibrated in Lysis buffer and washed with Lysis buffer supplemented with 50 mM imidazole pH 8.0 at RT. GPN1-GPN3 complex was eluted with linear gradient from 50 mM to 500 mM imidazole. Fractions containing complexes were pooled, cleaved with 3C protease over night while dialyzing against Lysis buffer without PIs. Next day, cleaved proteins were applied on HisTrap column and flow-through collected. Proteins were concentrated to 500 ul using 10,000 MWCO Amicon Centrifugal Filter (Merck Millipore) and separated on Superdex Increase 75 10/300 GL (Cytiva) equilibrated in Size exclusion buffer (20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT). Peak fractions were analyzed by SDS-PAGE gel, pooled, aliquoted and frozen in LN. Proteins were stored at -80 °C until further use. In case of His-GPN1-GPN3 complexes, cleavage and reverse Ni-NTA step was omitted. GTPase activity of GPN1-GPN3 complex was verified by Malachite Green Phosphate Assay Kit (Sigma-Aldrich).

RPAP2 full-length protein purification protocol is adapted from Fianu *et al*., 2021. Shortly, pellets of Hi5 cells expressing full-length RPAP2 resuspended in lysis buffer (25mM Tris-HCl pH 8.0 at RT, 500mM NaCl, 10% glycerol (v/v), 10µM $ZnCl_2$, 1mM DTT, 1xPIs) and lysed by sonication. The lysate was centrifuged and supernatant applied on 5 mL HisTrap HP column (Cytiva) equilibrated in lysis buffer. The column was washed with 10 CV of Wash buffer (25 mM Tris-HCl pH 8.0 at RT, 500 mM NaCl, 30 mM imidazole pH 8.0 at RT, 10% glycerol (v/v), 10 µM $ZnCl_2$, 1 mM DTT). An MBPtrap 5 mL column (Cytiva) equilibrated in Wash buffer was connected in tandem to the HisTrap column. Bound proteins were eluted from HisTrap column using linear gradient of Wash buffer from 30 mM to 500 mM imidazole pH 8.0 at RT. The HisTrap column was then detached and MBPTrap column was washed with 10 CV of Wash buffer and then the proteins were eluted with Maltose elution buffer (25 mM Tris-HCl pH 8.0 at RT, 500 mM NaCl, 30 mM imidazole pH 8.0 at RT, 10% glycerol (v/v), 10 µM $ZnCl_2$, 100 mM maltose, 1 mM DTT). Peak fractions were assessed by SDS-PAGE gel stained by Coomassie. RPAP2 containing fractions were pooled and cleaved with 3C protease over night at 4°C. Cleaved protein was applied on HisTrap column equilibrated in Wash buffer. Flow-through was collected and concentrated using 30,000 MWCO Amicon Ultra Centrifugal Filter (Merck Millipore). Concentrated sample was applied on a Superdex 200 10/300 increase column (Cytiva) equilibrated in size exclusion buffer (20 mM HEPES pH 7.9 at RT, 200 mM NaCl, 10% (v/v) glycerol, 10 µM $ZnCl_2$, 1 mM DTT). Peak fractions were analyzed by SDS–PAGE and Coomassie staining. Fractions containing RPAP2 were pooled and concentrated using a 30,000 MWCO Amicon Ultra Centrifugal Filter (Merck Millipore), aliquoted, frozen in liquid nitrogen, and stored at −80 °C (Fianu *et al*., 2021).

RPAP2 truncated was expressed in *E. coli* BL21(DE3)RIL cells. Cells were grown in LB with 30 µg/mL, kanamycin, 100µg/mL chloramphenicol to OD_{600} 0.6-0.8 and protein expressed adding IPTG (final concertation 0.5 mM) for 16 hours at 18 °C.

RPAP2 truncated protein was purified as RPAP2 full-length only omitting MBPtrap step. Protocol adapted from Fianu *et al*., 2021.

GDOWN1 protein was expressed in *E. coli* BL21(DE3)RIL in LB medium supplemented with 30µg/mL kanamycin and 100µg/mL chloramphenicol. The culture was grown to OD_{600} 0.6-0.8 at 37°C and then protein expression was induced by 1 mM IPTG for 6 hrs at 25°C . Cells were collected and resuspended in lysis buffer (Buffer A: 50 mM HEPES pH 7.5 @ RT, 10% glycerol, 300 mM NaCl, 1mM $CaCl_2$; 10mM imidazole pH 8.0 @ RT, 1x PIs) and sonicated on ice. Cell lysates were cleared by centrifugation for 45 minutes, 50 000g, 4°C and applied on Ni-NTA agarose (Qiagen) equilibrated in lysis buffer and then washed over 10 CV with Wash buffer (Buffer A, 15mM imidazole pH 8.0 at RT, 1xPIs). Excess of chaperones was removed by washing and incubating the resin with ATP wash buffer (Wash buffer, 5 mM ATP, 2 mg/mL *E. coli* denatured proteins). GDOWN1 protein was eluted with 10 CV of Elution buffer (Buffer A, 250mM imidazole pH 8.0 at RT). Peak fractions were checked on SDS-PAGE gel and those containing protein of interest pooled together. Buffer exchange was performed using PD-10 desalting columns into Lysis buffer with 1mM DTT and GDOWN1 was cleaved overnight at 4°C using 3C protease. Next day, GDOWN1 protein was applied on Ni-NTA resin equilibrated in Lysis buffer with 1mM DTT. Cleaved GDOWN1 was collected, diluted to 50 mM NaCl concentration and subjected to cation exchange using MonoS 5/50 GL column (Cytiva). Protein was eluted with linear gradient from 50mM to 1 M NaCl buffer. Peak fraction containing GDOWN1 protein were pooled diluted to final 50 mM NaCl concertation and applied on MonoQ 5/50 GL column (Cytiva). Protein was eluted with linear gradient from 50mM NaCl to 1 M NaCl. Peak fractions were assessed by SDS-PAGE and pooled. Protein was stored at -80°C.

Porcine Pol II was purified following, protocol adapted from Bernecky *et al*., 2016. Pig thymus was homogenized for 3 min in Buffer A (50 mM Tris-HCl pH 7.9 at 4 °C, 1 mM EDTA, 10 µM $ZnCl_2$, 10% glycerol, 1 mM DTT, protease inhibitors) using a 2 l blender (Waring). The homogenized material was centrifuged and the supernatant filtered through two layers of Miracloth. A 5% solution of polyethyleneimine, pH 7.9 at 25 °C, was added to a final concentration of 0.02%, and the material was stirred for 10 min then centrifuged. The resulting pellets were washed with buffer A before resuspension in Buffer A (0.15 M ammonium sulfate). After centrifugation, the conductivity of the supernatant was adjusted to that of Buffer A (0.2 M ammonium sulfate), and the resulting material was loaded on a 225-ml MacroPrepQ column equilibrated in buffer A (0.2 M ammonium sulfate). The column was washed with two column volumes of Buffer A (0.2 M ammonium sulfate), followed by Pol II elution with Buffer A (0.4 M ammonium sulfate). The eluate was precipitated by adding finely ground ammonium sulfate added to 50% saturation, and pellets were collected by centrifugation. The pellets were resuspended in Buffer A, and the conductivity was adjusted to that of Buffer A (0.15 M ammonium sulfate). The material was clarified by centrifugation, and further purified using a 5-ml gravity flow column of 8WG16 (αRPB1 CTD) antibody-coupled sepharose equilibrated in buffer A (0.15 M ammonium sulfate). After application of the input material, the antibody column was washed with five column volumes of Buffer A (0.5 M ammonium sulfate) and equilibrated to RT for 15 min. Pol II was eluted using Buffer A (0.5 M ammonium sulfate, 50% (v/v) glycerol), and Pol II-containing fractions were immediately mixed with Buffer A (2 mM DTT, lacking glycerol and protease inhibitors). The diluted material was centrifuged and subjected to anion exchange chromatography using a UNO-Q column equilibrated in Buffer A (0.1 M ammonium sulfate, 2 mM DTT, lacking protease inhibitors). Pol II was eluted using a linear gradient from

0.1 M to 0.5 M ammonium sulfate in Buffer A (2 mM DTT, lacking protease inhibitors). For the purification of 12-subunit bovine Pol II, the GDOWN1-free Pol II fraction was applied to a Sephacryl S-300 HiLoad sizing column equilibrated in Buffer B (150 mM NaCl, 5 mM HEPES pH 7.25 at 25 °C, 10 µM $ZnCl_2$, 10 mM DTT). For the purification of bovine Pol II containing GDOWN1, the GDOWN1-free Pol II fraction was incubated with a 3-times molar excess of human Gdown1 for 1 h at 4 °C before application to the Sephacryl S-300 HiLoad sizing column. Pol-II-containing fractions were concentrated using a 100,000 MWCO Amicon concentrator to a final concentration of 2-4 mg/mL. For PolII(G) sample, the PolII was incubated with human GDOWN1 in molar ratio 1:3 for 30 minutes, prior loading to Sepharycyl S-300 HiLoad column. Then concentrated as described above. Protocol adapted from Bernecky *et al*. 2016.

Human CTD domain and MBP protein were expressed separately in *E. coli* BL21(DE3)RIL in OD600 0.6-0.8 by adding 0.5mM IPTG at 37°C for 3-4 hours and then harvested and resuspended. Purification is based on the paper from Boehning *et al.*, 2018. *E. coli* pellets were sonicated and cleared by centrifugation. Lysate was loaded onto a 5mL HisTrap HP column (Cytiva) equilibrated in LB300 buffer (20mM HEPES pH 7.4 at RT, 300mM NaCl, 10% glycerol, 30mM imidazole pH 8.0 at RT, 1mM DTT, 1xPIs). The column was washed 20CV of HSB1000 (20mM HEPES pH 7.4, 1M NaCl, 10% glycerol, 30mM imidazole pH 8.0 at RT, 1mM DTT, 1xPIs), then again washed with 10CV of LB300 buffer. hCTD was eluted with 10CV of Elution buffer A (20mM HEPES pH 7.4 at RT, 300mM NaCl, 10% glycerol, 500mM imidazole pH 8.0, 1mM DTT, 1xPIs). Peak fractions containing hCTD protein were pooled and applied on amylose resin (New England Biolabs) that was washed with 20CV of HSB1000 buffer. Proteins were eluted with elution buffer B (20mM HEPES pH 7.4 at RT, 300mM NaCl, 10% glycerol, 30mM imidazole pH 8.0, 15mM maltose, 1mM DTT, 1xPIs). The fractions containing hCTD were then subjected to size-exclusion chromatography using Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in SEC300 buffer (20mM HEPES pH 7.4 at RT, 300mM NaCl, 10% glycerol, 1mM DTT). Peak fractions were checked on SDS-PAGE gel stained by Coomassie and hCTD was pooled and concentrated using 30,000 MWCO Amicon Ultra Centrifugal Filter (Merck Millipore). MBP protein was purified the same, omitting the amylose resin step.

For test purification, GPN1, GPN2 and GPN3 proteins were expressed in *E. coli* BL21(DE3)RIL at 30 or 37°C for 3 or 6 hours or in Hi5 insect cells in smaller volume. Cell pellets were mixed with lysis buffer (20 mM HEPES pH 7.9 at RT, 200 mM NaCl, 1 mM DTT, 1xPIs), sonicated on ice and applied to MagneHis beads (Promega). They were incubated for 30 minutes at 4°C then washed three-times with Wash buffer (20 mM HEPES pH 7.9 at RT, 200 mM NaCl, 1 mM DTT, 30 mM imidazole pH 8.0 @ RT, 1xPIs) and eluted with 500 mM imidazole buffer. Samples were separated on SDS-PAGE gel and Coomassie stained.

Cell lines production

Generation of a knock-in cell line expressing endogenous GFP–3C-Gdown1. K562 cells (DSMZ) were edited to express a eGFP–Gdown1 fusion protein using a modified CRISPR–Cas9 knock-in protocol. In brief, the gRNA was designed using the Benchling CRISPR gRNA design tool (Benchling; 5′-AGAATGTGCTCGCTGCCCCG-3′) and cloned into the plasmid pLCG (hU6-sgRNA-EFSSpCas9-P2A-mCherry), a gift from J. Zuber, IMP, Vienna. The 500 bp sequences flanking the Gdown1 start codon were obtained by PCR on gDNA obtained from K562 cells and subcloned into a pLPG vector, a gift from J. Zuber, digested with MluI using Gibson Assembly (NEB), which

produced the final vector pLPG-GFP-AID (5′ BlastR-P2A-eGFP AID-3C). K562 cells were grown in RPMI medium supplemented with 10% FBS (Sigma), 2% l glutamine (Gibco), 1% sodium pyruvate (Sigma) and 1% penicillin–streptomycin (Sigma) and transfected with the HDR donor and the Cas9 plasmids using a Neon electroporation device (Invitrogen) according to the user manual (for suspension cells). After 14 days of transfection and after several passages and blasticidin selection (10 µg ml, Invitrogen), cells were subjected to FACS using a BD FACSAria III (BD Biosciences). Cells expressing the eGFP tag were sorted into 96-well plates. After allowing single cells to regrow for approximately 2 weeks, clones with homogeneous GFP fluorescence were genotyped (primers: GCCAGAGTCTCCCAAATCCT and ACAAGTGTCTCGTGGGAAGT). For homozygously edited cells, expression of GFP–Gdown1 was analysed by western blotting with anti-POLR2M (HPA068141, Sigma Aldrich) and anti-GFP (A11122, Invitrogen) antibodies. Cell lines were produced by D. Riabov-Bassat from Plaschka lab (protocol adapted from Belen *et al*., 2023).

Preparation of cell extracts:

All steps were performed at 4°C unless otherwise stated. Cytoplasmic and nuclear extracts (NE, CE), 12 L of human K562 cells endogenously expressing GFP-AID-3C-GDOWN1 were grown in RPMI (supplemented with FBS, GlutaMax, sodium pyruvate and penicillin/streptomycin) to a density of 1×10^6 cells mL^{-1} at 37°C, stirred at 60 rpm and 5% CO_2. Cells were harvested and washed with cold 1x PBS then resuspended in Buffer A (10 mM HEPES pH 7.9 at RT, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 1x PIs) and let swell on ice for 10 minutes. Cells were homogenized with Douncer using pestle B. Homogenized material was centrifuged and CE = supernatant and NE = pellet separated. Nuclear fraction was resuspended in Low salt buffer (20 mM HEPES pH 7.9 at RT, 25% glycerol, 1.5 mM $MgCl_2$, 20 mM KCl, 0.5 mM DTT, 1x PIs) and High salt buffer (20 mM HEPES pH 7.9 at RT, 25% glycerol, 1.5 mM $MgCl_2$, 1.2 M KCl, 0.5 mM DTT, 1x PIs) in 1:1 ration and allowed to extract for 30 minutes gently stirring. Spin NE and dialyze the supernatant against Buffer E (20 mM HEPES pH 7.9 at RT, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT). Cytoplasmic fraction was diluted with 0.11 volume of Cytoplasmic buffer (0.3 M HEPES pH 7.9 at RT, 1.4 M KCl, 0.03 M $MgCl_2$) while stirring then dialyzed against Buffer E. Both CE and NE were aliquoted, frozen in liquid nitrogen and stored at -80°C. Correct separation of cell extracts was confirmed by Western blot analysis. Protocol adapted from Mayeda and Krainer, 1999.

Extracts for experiments with cells expressing GFP-AID-3C-GDOWN1 and TIR1 protein for AID degron system experiments were grown in RPMI (supplemented with FBS, GlutaMax and sodium pyruvate), at 37°C in 5% CO_2. Cell were harvested and washed in ice cold 1x PBS. Pellets were resuspended in Cell lysis buffer (10 mM Tris-HCl pH 7.4 at RT, 150 mM NaCl, 0.15 % IGEPAL) and let sit on ice for 10 minutes. Cell lysate was layered on top of Sucrose buffer (10 mM Tris-HCl pH 7.4 at RT, 150 mM NaCl, 24% sucrose) in tube. Lysates were centrifuged for 10 minutes, 3500 g, at 4°C. The top layer was carefully pipetted out and saved as cytoplasmic extract. The rest of sucrose buffer was discarded. The pellets were washed with ice cold 1x PBS + 0.5 mM EDTA. The nuclei were resuspended in Glycerol buffer (20 mM Tris-HCl pH 7.4 at RT, 75 mM NaCl, 0.5 mM EDTA, 50 % glycerol) and immediately mixed with same volume of Nuclear lysis buffer (10 mM Tris-HCl pH 7.4 at RT, 1 M urea, 0.3 M NaCl, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 1 % IGEPAL). Mixture was incubated on ice for 2 minutes and spin for 2 minutes, 13500 g, at 4°C. Nuclear fraction was saved and

pellets were incubated with 1x PBS + 2 mM $MgCl_2$ + bensonase (1:1000) for 1 hour at RT to release chromatin bound proteins. Protocol adapted from Conrad, 2017.

GFP-nanobady resin production:

All steps were performed at 4˚C unless stated otherwise. AminoLink Plus resin (Thermo Scientific) was equilibrated in Coupling buffer (50 mM sodium phosphate pH 7.2 at RT, 150 mM NaCl) and incubated with GFP-nanobody (0.3mg/100µL of resin) and $NaCNBH_3$ (final concentration 50 mM) for 4 hours and rotating. Resin was washed three times in Coupling buffer followed by blocking in 1M Tris-HCl pH7.4 @ 4°C with 50 mM $NaCNBH_3$ for 30 minutes, at 4°C and rotating. GFP-nanobody coupled resin was washed six times with 1 M NaCl and then three times 1x PBS, then stored at 4°C supplemented with 0.02% sodium azide.

SulfoLink resin (Thermo Scientific) was equilibrated in coupling buffer containing 50 mM Tris-HCl pH 8.5 and 5 mM EDTA pH 8.0. TCEP (in final concentration 25 mM) was added to GFP-nanobody with cysteine linker (0.3mg/100µL of resin) and the mixture was incubated with equilibrated SulfoLink resin for 4 hours at 4 °C. Unbound proteins were washed away three times with coupling buffer followed by blocking in 50 mM L-cystein-HCl for 30 minutes, rotating. GFP-nanobody coupled resin was washed six times with 1 M NaCl and then stored in 1x PBS at 4°C supplemented with 0.02% sodium azide.

Cyanogen bromide-activated-sepharose resin (Sigma Aldrich) was first swelled in 1 mM HCl for 30 minutes at RT and washed 10-times in MilliQ water then equilibrated in coupling buffer (0.1 M NaHCO3 pH 8.5 at RT, 0.5 M NaCl). GFP-nanobody was incubated with resin (0.3 mg/mL per 100 µL resin) for 4 hours, rotating. Unreacted GFP-nanobody was washed away 3-times with coupling buffer and resin blocked with 0.2 M glycine for 16 hours. The blocking solution was removed and resin washed extensively with coupling buffer, then low pH buffer (0.1 M acetate pH 4.0 at RT, 0.5 M NaCl) and again in coupling buffer. Resin was stored in 1x PBS, at 4° C supplemented with 0.02% sodium azide.

Pull-down of GFP-tagged GDOWN1 from cell extracts:

GFP-nanobody resin (home-made or GFPtrap) was equilibrated in Wash buffer (20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 0.05% Triton X-100, 2 mM MgCl2, 10% glycerol, 2 mM DTT). CE or NE was applied to it diluted 1:1 in Dilution buffer (20 mM HEPES pH 7.9 at RT, 7.5 mM KCl, 3 mM MgCl2, 1 mM DTT). Samples were incubated for 2 hours at 4°C. GFPtrap with bound proteins was washed five time in Wash buffer, then once in Last wash buffer (20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 2 mM MgCl2, 2 mM DTT). Proteins were eluted from resin with 3C protease for 2 hours at 4°C. Eluted proteins that were further analyzed by MS were flesh frozen in LN and stored at -80°C. Cytoplasmic Pol II complexes used for XL-MS or cryoEM were further processed (described below).

Cytoplasmic Pol II complexes preparation:

Sucrose gradients for GraFix were prepared as followed. Light sucrose solution (5% sucrose, 20 mM HEPES pH 7.9, 100 mM NaCl, 2 mM MgCl2, 2 mM DTT) and heavy sucrose solution (30% sucrose, 20 mM HEPES, 100 mM Nacl, 0.01% glutaraldehyde) were prepared separately. Five layers of sucrose gradient from 30% to 5% were layered on top of each other. Each layer was frozen in dry ice for 5 minutes before layering another one. Frozen gradients were thawed over night at 4°C. Cytoplasmic Pol II complexes eluted from homemade GFP-nanobody resin (as described

previously) were applied on the top of the sucrose gradient. Gradient with samples was centrifuged for 16 hours, 32 000 rpm, at 4°C. Gradient was fractionated and cytoplasmic Pol II fractions combined, quenched with 50 mM L-lysine pH 7.8 at RT and dialyzed for 24 hours against Dialysis buffer (20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 2 mM MgCl2, 2 mM DTT). Dialyzed samples were concentrated using 100,000 MWCO Amicon Centrifugal Filter (Merck Millipore).

In vitro Pol II reconstituted complexes preparation:

Pol II or Pol II (G) was incubated with GPN1-GPN3 and RPAP2 in molar ratio 1:2:2 for 30 minutes at 4°C. Samples were subjected to size exclusion on Superdex 200 Increase 3.2/300 (Cytiva) equilibrated in SEC buffer (20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 2 mM $MgCl_2$, 10 µM $ZnCl_2$, 2 mM DTT).

Elongation complex of Pol II (Pol II EC) was formed by mixing Pol II with nucleic acid scaffold in 1.5x molar excess. First, template DNA (sequence 5′-AA-GCTCAAGTACTTAAGCCTGGTCATTACTAGTACTGCC-3′) was pre-annealed with RNA (5′-UAUAUGCAUAAAGACCAGGC-3′), followed by incubating with Pol II 10 minutes at 4°C, then 10 minutes at 25°C. Afterwards, non-template DNA (sequence 5′-GGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT-3) was added to mixture and incubated for 25 minutes at 25 °C.

Testing interactions of proteins of interest using pull-downs:

His-GPN1-GPN3 complex was first incubated with MagneHis beads (Promega) for 30 minutes at 4°C. Unbound proteins were washed away with Low salt wash buffer (20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 2 mM $MgCL_2$, 1 mM DTT) and Pol II, Pol II EC or RPAP2trun were added in 1:2 molar ratio and incubated with beads for 30 minutes at 4°C. Beads were washed with High salt buffer (20 mM HEPES pH 7.9 at RT, 300 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT; for RPAP2trun experiment 30 mM imidazole pH 8.0 at RT was added) and the bound proteins eluted with 500 mM imidazole pH 8.0 at RT. The proteins were analyzed by SDS-PAGE gel stained with Coomassie. The same protocol applies for hCTD/MBP proteins and their interaction with GPN1-GPN3.

Cytoplasmic Pol II complexes were pulled-down using GFP-nanobody resin as described above. Before elution, immobilized complexes were incubated with 5 µg His-hIWR1 protein for 2 hrs minutes at 4°C and washed with buffer containing 20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 0.05% Triton X-100, 2 mM MgCl2, 10% glycerol, 2 mM DTT. Proteins were eluted with Gly-HCl pH 2.7 at RT. Samples separated on SDS-PAGE and blotted on PVDF transfer membrane (Thermo Fisher). Membrane was blocked in 1.5 % milk (Roth) and incubated with primary antibodies and afterwards with secondary antibody conjugated with HRP. Membranes were developed using SuperSignal West Pico Chemiluminescnet substrate (Thermo Fisher).

Mass photometry experiments:

Complex of GPN1-GPN3 and RPAP2 was formed by mixing the proteins in 1:2 molar ratio and incubated for 30 minutes on ice and diluted to final concertation 100 nM using Dilution buffer (20 mM HEPES pH 7.9 at RT, 100 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT). Similarly, Pol II complex with GPN1-GPN3, RPAP2 and GDOWN1 was prepared. Samples and respective controls were measured using Refeyn Two ^{MP}

mass photometer (Refeyn Ltd.). The movies were collected for 60 seconds. The movies were processed with Discover^{MP} software. The instrument was calibrated using BSA or MfP1 standard before each measurement.

Electron microscopy sample preparation and data collection:

For initial negative stain, EM imaging and screening fractions containing cytoplasmic Pol II complexes were adsorbed to 4.3-5 nm carbon, washed with water and stained with 2 % solution of uranyl formate. Samples for cryoEM were prepared as follows. Cytoplasmic Pol II complexes were pipetted in wells, layered with 2nm carbon and incubated for 30 minutes at 4°C. Using Quantifoil R1.2/1.3 grid carbon layer was fished out and the grid was plunged frozen in liquid ethane using Vitrobot Mark IV (humidity 100%, 4°C). Grids were stored in LN.

In vitro reconstituted Pol II (G)-RPAP2-GPN1-GPN3 was incubated with 0.09% glutaraldehyde for 1 hr at 4°C and quenched with L-lysine (final concentration 60 mM) prior applying on size exclusion column as described above. Peak fractions were pooled and diluted to final concentration 0.25 mg/mL. 3.5 µL of the complex were applied Quantifoil R1.2/1.3 holey carbon grids that had been glow discharged for 1 min (25 mA current, 7.0 x 10^{-1} mbar vacuum). Using a Vitrobot Mark IV set to 100% humidity and 4 °C, grids were blotted for 7 seconds and immediately plunge frozen in liquid ethane and stored in LN.

Grids were screened for particle density and particle orientation using a Thermo Fisher Glacios transmission electron microscope (200 kV) equipped with a Falcon III direct electron detector (cytoplasmic Pol II) or Falcon 4i direct electron detector (PolII(G)-RPAP2-GPN1-GPN3). A small dataset of *in vitro* reconstituted Pol II (G)-RPAP2-GPN1-GPN3 was collected on Glacios TEM at a nominal magnification of ×150000, pixel size 0.95Å. 3363 micrographs were collected using a defocus range of -1.4 to -2.6 µm with an exposure time 5.86 s. The total electron exposure was 69 e-/Å^2 distributed over 50 frames.

Grids were transferred to a Thermo Fisher Titan Krios G3i transmission electron microscope (300 kV) equipped with a Gatan K3 BioQuantum direct electron detector (energy slit width 10 eV) for data collection. Data sets were recorded using EPU 2.11. For cytoplasmic Pol II dataset, micrographs were collected at a nominal magnification of ×130000, pixel size 0.65Å. A total of 43,150 micrographs were collected using a defocus range of -1.27 to -2.75 µm with an electron exposure rate of 13.36 e-/Å^2/s and an exposure time 3.8 s. The total electron exposure was 75 e-/Å^2 distributed over 50 frames. For Pol II(G)-RPAP2-GPN1-GPN3 dataset, micrographs were collected at nominal magnification of x 130000, pixel size 0.65Å. A total of 19,569 micrographs were collected using a defocus range of -0.6 to -2.0 µm with an electron exposure rate of 18 e-/Å^2/s and an exposure time 1.6 s. The total electron exposure was 61 e-/Å^2 distributed over 50 frames.

Negative stain EM processing:

Micrographs collected with FEI Tecnai 12 transmission electron microscope operated at 120 keV at a nominal magnification 42000x were manually checked and particles picked in EMAN2 (Tang *et al*., 2007). 16452 picked particles were further 2D classified and 3D refined using RELION 3.1 (Scheres, 2012).

CryoEM data processing:

Micrographs of cytoplasmic Pol II and *in vitro* complex were first processed with Warp (Tegunov, 2019) and visually screened to remove micrographs with poorly visible Thon rings or too thick ice.

Micrographs of cytoplasmic Pol II were motion corrected, dose weighted and CTF estimated using Warp (Tegunov, 2019). Particle coordinates were selected by Warp too. Dataset was further processed in RELION 3.1 (Zivanov 2018). Several rounds of 2D classification were applied to clean the dataset. Several rounds of 3D refinement, defocus and astigmatism refinement were applied. Particles were polished followed by beam tilt, defocus and astigmatism refinement.

Micrographs of *in vitro* Pol II(G)-RPAP2-GPN1-GPN3 collected by Glacios TEM microscope were further processed in RELION 5.1, where they were motion corrected, dose weighted followed by CTF estimation using CTFFIND4 (Rohou, 2015). Particle coordinates were selected by Warp and further used in RELION 5. Particles were 2D classified and good-looking classes were further 3D refined (Scheres, 2012; Tegunov, 2019).

Modeling and refinement:

A model of human Pol II-RPAP2 combined with GDOWN1 model generated with AlphaFold3 (AF3) was rigid body fit into the cytoplasmic Pol II and Pol II (G)-RPAP2-GPN1-GPN3 density map using UCSF ChimeraX. Densities and models were visualized using UCSF ChimeraX (Pettersen, 2021, Abramson *et al*., 2024).

Mass-spectrometry experiments

The nano HPLC system (UltiMate 3000 RSLC nano system, Thermo Fisher Scientific) was coupled to an Q *Exactive* HF-X mass spectrometer equipped with a Proxeon nanospray source. Peptides were loaded onto a trap column (PepMap Acclaim C18, 5 mm × 300 µm ID, 5 µm particles, 100 Å pore size, Thermo Fisher Scientific) at a flow rate of 25 µl/min using 0.1% TFA as mobile phase. After loading, the trap column was switched in line with the analytical column (PepMap Acclaim C18, 500 mm × 75 µm ID, 2 µm, 100 Å, Thermo Fisher Scientific). Peptides were eluted using a flow rate of 230 nl/min, starting with the mobile phases 98% A (0.1% formic acid in water) and 2% B (80% acetonitrile, 0.1% formic acid) and linearly increasing to 35% B over the next 180 min. This was followed by a steep gradient to 95%B in 5 min, stayed there for 5 min and ramped down in 2 min to the starting conditions of 98% A and 2% B for equilibration at 30°C.

MS Analysis

The Q Exactive HF-X mass spectrometer was operated in data-dependent mode, performing a full scan (m/z range 380-1500, resolution 60,000, minimum intensity set to 4.8E4), followed each by MS/MS scans of the most abundant ions. MS/MS spectra were acquired using an isolation width of 1.0 m/z, AGC target value of 1E6, maximum injection time 60 ms, HCD with a collision energy of 28. Precursor ions selected for fragmentation (include charge state 2-6) were excluded for 60 s. The exclude isotopes feature was enabled.

Data Processing protocol

For peptide identification, the RAW-files were loaded into Proteome Discoverer (version 2.3.0.523, Thermo Scientific). All MS/MS spectra were searched using MSAmanda v2.0.0.12368 (Dorfer V. *et al*., J. Proteome Res. 2014 Aug 1;13(8):3679-84).

For the protein-identification-set, the peptide mass tolerance was set to ±5 ppm and fragment mass tolerance to ±8 ppm, the maximum number of missed cleavages was set to 2, using tryptic enzymatic specificity with proline restriction. Peptide and protein identification was performed in two steps. For an initial search the RAW-files were searched against the human_uniprot_reference_2020-12-05.fasta (20,541 sequences and 11395748 residues), supplemented with common contaminants using Beta-methylthiolation on cysteine as a fixed modification. The result was filtered to 1 % FDR on protein level using the Percolator algorithm (Käll L. *et al*., Nat. Methods. 2007 Nov; 4(11):923-5) integrated in Proteome Discoverer. A sub-database of proteins identified in this search was generated for further processing. For the second search, the RAW-files were searched against the created sub-database using the same settings as above and considering the following additional variable modifications: oxidation on methionine, glutamine to pyro-glutamate conversion at peptide N-terminal glutamine and acetylation on protein N-terminus. The localization of the post-translational modification sites within the peptides was performed with the tool ptmRS, based on the tool phosphoRS (Taus T. *et al*., J. Proteome Res. 2011, 10, 5354-62). Identifications were filtered again to 1 % FDR on protein and PSM level, additionally an Amanda score cut-off of at least 150 was applied. Proteins were filtered to be identified by a minimum of 2 PSMs in at least 1 sample. Peptides were subjected to label-free quantification using IMP-apQuant (Doblmann *et al*., 2019). Proteins were quantified by summing unique and razor peptides and applying intensity-based absolute quantification (iBAQ, Schwanhäusser *et al*., 2011). Protein-abundances-normalization was done using sum normalization, based on the MaxLFQ algorithm (Cox J. *et al*., 2014). Identified proteins were filtered to contain at least 2 peptides. Statistical significance of differentially expressed proteins was determined using limma (Smyth, G. K., 2004). The MS analysis was performed by VBC MS facility. Protocol adapted from Belen *et al*., 2023)

Cross-linking mass-spectrometry experiments:

BS3, DSBU and DSOO cross-linking of Pol II and cytoplasmic Pol II

Optimization of cross-linking conditions for cytoplasmic Pol II was first tested on pure Pol II. Pol II was incubated with BS3 (Thermo Fisher), DSSO (Thermo Fisher) or DSBU (Thermo Fisher) at 0 mM to 3 mM cross-linker concentration for 30 minutes at 4°C. Cross-linked complexes were separated on SDS-PAGE gels to check the effectiveness of cross-linking.

Complex of Pol II, GDOWN1, RPAP2 and GPN1-GPN3 was mixed in molar ratio 1:3:2:2 and incubated for 30 minutes at 4°C. The complex was incubated for 30 minutes with 3 mM BS3 (Thermo Fisher) and quenched with final concentration 60 mM Ammoniumbicarbonate (Sigma-Aldrich).

DMTMM crosslinking of RPAP2-GPN1-GPN3 complexes

Complex of GPN1-GPN3 and RPAP2 was formed mixing the proteins in 1:2 molar ratio and incubated for 30 minutes on ice. 22 µg of the RPAP2-GPN1-GPN3 complex at 1 µg/µL were incubated with 44 mM DMTMM (4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholiniumchlorid, Sigma-Aldrich) for 12 min at room temperature. Sample was loaded on pre-conditioned Zeba Spin columns and centrifuged for

1.5 min at 2500 g. 3 µL 1 M Ammoniumbicarbonate (Sigma-Aldrich) was added to the flow through.

DSSO crosslinking of cytoplasmic Pol II

Cytoplasmic Pol II samples were enriched using ProteoMiner protein enrichment kit (Bio-Rad Laboratories, Vienna, Austria) Therefore 50 µL ProteoMiner beads were washed twice and pre-conditioned with Last wash buffer. 300 µL sample was added to the beads and incubated for 2 h at room temperature. After the incubation supernatant was removed and beads were washed with buffer (pH 7.8). For crosslinking 50 µL of Disuccinimidyl Sulfoxide (DSSO, Cayman Chemical, Ann Arbor, MI, USA) was added at a final concentration of 1 and 1.5 mM and incubated for 6 minutes. Reaction was quenched using 1 M Ammoniumbicarbonate and submitted to proteolytic digestion.

Proteolytic digestion

Samples cross-linked with BS3, DSSO or DMTMM were prepared analogous from here on: 8M Urea was added to denature the protein. Proteins were reduced and alkylated using TCEP and IAA respectively. Samples were digested using Lys-C (FUJIFILM Wako Chemicals U.S.A. Corporation) for 1h and after dilution to 1 M Urea with Trypsin (Promega) overnight. Afterwards samples were acidified with TFA, peptides were cleaned using Sep-Pak tC18 1 cc Vac Cartridges (Waters) and dried in a vacuum centrifugation. Peptides were dissolved in 20 µL SEC buffer and injected onto a Superose 30 increase (Cytiva) to enrich fractions containing crosslinked peptides (ref: 10.1074/mcp.M111.014126–1). Three fractions (-1,0,+1) were dried and dissolved in mobile phase A (98% H_2O, 2% ACN, 0.1% FA).

MS measurements

Samples containing cross-linked peptides were injected onto an Ultimate 3000 RSLCnano System coupled to an Orbitrap Eclipse Tribrid Mass Spectrometer (both Thermo Fisher Scientific, Vienna, Austria). Peptides were separated on a PepMap RSLC EASY-Spray column (C18, 2 µm, 100 Å, 75 µm x 15 cm, Thermo Fisher Scientific). Separation occurred at 300 nL·min-1 with a flow gradient from 2-35% mobile phase B (2% H_2O, 98% ACN, 0.1% FA) within 60 min resulting in a total method time of 94 min. Orbitrap Eclipse was equipped with a FAIMS Pro alternating between three different CVs (-50,-60,-75) operating in positive mode. MS1 scans were recorded at a resolution of 60,000 @200 m/z in a range from 400-1600 m/z. Precursors with a charge of 3-8 were isolated within a 1.6 m/z window and fragmented with HCD with a stepped collision energy of 21,27 and 33 % NCE. MS2 spectra were recorded at a resolution of 30,000 @200 m/z.

Bioinformatics

Raw files were converted to mgf using MSConvert (ref:10.1038/nbt.2377) and searched using xiSEARCH (v1.8.5, ref: 10.15252/msb.20198994). Database consisted of the expected polymerase subunits and 40 additional proteins representing the most abundant proteins identified in a corresponding proteome profiling. To verify identified crosslinks xiFDR (v2.3) was employed at a maximum Residue Pair FDR of 1%. Due to the low protein and cross-link identifications xiFDR warned that the FDR

calculation was likely to be unreliable and identifications were also manually verified using xiView (Combe *et al*., 2024). XL-MS experiments were done by B. Neuditschko (Herzog lab).

GDOWN1 depletion and UV experiments:

Human K562 cells endogenously expressing GFP-AID-3C-GDOWN1 were grown in RPMI (supplemented with FBS, GlutaMax, sodium pyruvate) to confluence 70 % at 37°C in suspension flasks and 5% CO_2. Afterwards 0.1 mM 3-indolacetic acid (IAA, Sigma) was added to cells and they were incubated for 15,30, 60 or 90 minutes at 37°C, in suspension flasks and 5% CO_2. Depletion of GDOWN1 was verified by Western blot.

For UV exposure experiments K562 cells endogenously expressing GFP-AID-3C-GDOWN1 were grown in RPMI (supplemented with FBS, GlutaMax, sodium pyruvate) to confluence 70 % at 37°C in suspension flasks and 5% CO_2. Cells were washed in 1x PBS and UV exposed to 20, 50 or 100 J/m^2 UVP 254 nm. Exposed cells were recovered for specific times in RPMI media at 37°C in suspension flasks and 5% CO_2. Genomic DNA (gDNA) was isolated from cells using Monarach Genomic DNA Purification kit (New England Biolabs). Nitrocellulose membrane (Thermo Fisher) was equilibrated prior the sample application in 20x SSC buffer (3 M NaCl, 0.3 M sodium citrate). Control DNA and gDNA were first mixed with 0.5 volume 20x SSC buffer and incubated for 10 minutes at 95°C then put on ice for 10 minutes and 0.5 volume of 20x SSC was added to samples. 15 µL of samples was applied on the membrane and let dry. Membrane was then incubated for 10 minutes in Denaturation buffer (1.5 M NaCl, 0.5 M NaOH) followed by 10 minutes incubation in Neutralization buffer (1 M NaCl, 0.5 M Tris-HCl pH 7.0 @ RT). Afterwards, the membrane was baked at 80°C for 2 hours and then blocked in 1.5 % milk (Roth). Membrane was incubated with primary antibodies anti-CPD (Santa Cruz Biotechnology) and anti-dsDNA (Santa Cruz Biotechnology), then secondary antibodies. After incubation, membranes were developed using SuperSignal West Pico Chemiluminescnet substrate (Thermo Fisher).

Growth competition assay

Positive clones with Tir1 background were mixed 1:1 with WT control cells, plated in triplicates at 0.4 mil. cells/well and left to grow for 72h. They were screened by FACS for GFP (GDOWN1) and BFP (Tir1) levels and split into 2 plates. 100ug Auxin was added to the experiment plate. The cell growth competition was monitored every 2 days by FACS analysis. Experiments performed by D. Riabov-Bassat from Plaschka lab.

6. Materials

List of oligos

oligos	sequence
IWR1_pOPINB_fw	ctgtttcagggcccgATGGAGGCCGCCAGGACC
IWR1_pOPINB_re	atggtctagaaagctTCAGTCTGAATCCAGGTCGTGG
hGPN1_pOPINB_fw	ctgtttcagggcccgATGGCGGCGTCCGCAGCT
hGPN1_pOPINB_re	atggtctagaaagctCTATTTATTGTTTCTCTTCCAG-TATTGTGCCATCGATTC
hGPN3_pOPINB_fw	ctgtttcagggcccgATGCCTCGGTATGCGCAG
hGPN3_pOPINB_re	atggtctagaaagctTCATTCATCCTGGCATTCTTGAAAATATTCG
GFP_N-His_3C_fw	aagttctgtttcagggcccgATGCAGGTTCAACTGGTGGA
GFP_N-His_3C_re	atggtctagaaagctTTATTTAGAGCTCACCGTCACC
438_A_fw	TAACattggaagtggataaccggtcc
438_A_re	aacttccagaccgctgctgtgat
GPN_fw	agcggtctggaagttctgtt
GPN1_438_re	ccacttccaatGTTACTATTTATTGTTTCTCTTCCAGTATTGT
GPN2_438_re	ccacttccaatGTTACTACAGCTGCATGGCTTCCTGC
GPN3_438_re	ccacttccaatGTTATTATTCATCCTGGCATTCTTGAA
KK_addition_SDM_fw	GACGGTGAGCTCTAAAaagaagCATCACCATCACCATCAC
KK_addition_SDM_rev	GTGATGGTGATGGTGATGcttcttTTTAGAGCTCACCGTC
438A_GPN1_His_del_fw	ATGGCGGCGTCCGCAGCTGCCGC
438A_GPN3_His_del_fw	ATGCCTCGGTATGCGCAGCTGG
438A_His_del_rev	CGattggattggaagtacttaataaaggatccgcgcccg
438_C_His_MBP_3C_fw	taacggatccgaattcgagcg
438_C_His_MBP_3C_rev	cgggccctgaaacagaactt
438_C_RPAP2_fw	ctgtttcagggcccgATGGCGGACTTCGCTGG
438_C_RPAP2_rev	aattcggatccgttaTCACTCTGGTAAACAGCTGGTTCTAAATA
438_C_RPAP2_trun(1-215)_rev	aattcggatccgttaTCAACTATCACTGTGAGTGCTAGAAGAAC
438_A_RPAP2mut_fw	TGGTGTTGCTGTCATTACTG
438_A_RPAP2mut_rev	TCTGCACTTTTCTTACTTATGCCTA
438_A_GPN1_mut_GPN3fw	GTCAGTAACCTGACTCGTTC
438_A_GPN1_mut_GPN3_re	ATGTCAGTTTTATTCATGACCACAATG
GDOWN1_S270A_fw	CAGAAAAGCGGTGCCCCGATTAGCAGTGAAGAAC
GDOWN1_S270A_rev	GTTCTTCACTGCTAATCGGGGCACCGCTTTTCTG
GDOWN1_S270E_fw	CAGAAAAGCGGTGAACCGATTAGCAGTGAAGAAC
GDOWN1_S270E_rev	GTTCTTCACTGCTAATCGGTTCACCGCTTTTCTG

List of plasmids

plasmid	source
pOPINE_GFP_nanobody	Addgene (#49172)
pOPINE_GFP_nanobody_with_GGC_for_coupling	Available in lab
pOPINE_GFP_nanobody_2_Lys	In this project
pOPINB_plus_Gdown1_FL	Available in lab
pOPINB_Plus_hGPN1	In this project
pOPINB_Plus_hGPN2	In this project
pOPINB_Plus_hGPN3	In this project
pOPINB_Plus_hIWR1	In this project
pOPINB_T7cons_TIR2_RPAP2trunc(1-215)codonopt	In this project
438-A_Plus_GPN1	In this project
438-A_Plus_GPN2	In this project
438-A_Plus_GPN3	In this project
438_C_His_MBP_3C_RPAP2	In this project
438A_His_GPN1_GPN3	In this project
MBP	Available in lab
hCTD	Kind gift from P. Cramer

Cas9 plus U6 promotor POLR2M guide cassette empty FILLER GFP	Plaschka lab
pLPG-GFP-AID (5′ BlastR-P2A-eGFP AID-3CPOLRM2	Plaschka lab

List of antibodies

antibodies	**vendor**
anti-POLR2M (GDOWN1)	HPA068141, Sigma Aldrich
anti-GFP	A11122, Invitrogen
anti-GFP	sc-9996, Santa Cruz Biotechnology
anti-CTD RPB1	8WG16, produced in lab
anti-GAPDH1	G8795, Sigma Aldrich
anti-U1 snRNP 70	sc-9571, Santa Cruz Biotechnology
anti-dsDNA	sc-545997, Santa Cruz Biotechnology
anti-CPD	CAC-NM-DND-001, Hoelzel Biotech
anti-His-HRP	130-092-785, Miltenyi Biotec

7. References

Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**:493–500 (2024).

Aibara S., Dienemann C., Cramer P. Structure of an inactive RNA polymerase II dimer. *Nucleic Acids Research*, 49(18):10747–10755 (2021)

Ali, M. M. U. *et al*. Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).

Aygun O., Svejstrup J. & Liu Y. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *PNAS*. **105**(25): 8580-8584 (2008)

Banerjee S. *et al*. 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science*. **351**:871-87. (2016)

Bernecky, C., Herzog F., Baumeister W. *et al*. Structure of transcribing mammalian RNA polymerase II. *Nature*. **529**: 551-554 (2016)

Boehning M., Duagst-Darzacq C., Rankovic M., *et al*. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Strict Mol Bio.* **25**(9), 833-840. (2018)

Boulon S., Pradet-Balade B., Verheggen C. *et al.* HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol Cell*. **39**(6) (2010)

Bradner J. E., Hnizs D. & Young R. A. Transcriptional addiction in cancer. *Cell*. **168**(4): 629-643 (2017)
Bradner J.E., Hnisz D. and Young R.A. Transcriptional Addiction in Cancer. *Cell.***168**(4): 629-643. (2018)

Bruckner F. & Cramer P. Structural basis of transcription inhibition by α-amanitin and implication for RNA polymerase II translocation. *Nat Struct Mol Biol*. **15**: 811-818 (2008)

Burley S.K., Roeder R.G. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem*. 65:769-799 (1996)

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

Carey M. F., Peterson C. L., Smale S. T. Dignam and Roeder nuclear extract preparation. *Cold Spring Harb Protoc*. **12**, pdb.prot5330 (2009)

Carre C. & Shiekhattar R. Human GTPases Associate with RNA Polymerase II To Mediate Its Nuclear Import. *Mol Cell Biol*. **31**(19), 3953-3962 (2011)

Chen X. *et al.* Structural insights into preinitiation complex assembly on core promoters. *Science.* **372**. (2021)

Cheng B., Li T., Rahl P. B. *et al*. Functional association of Gdown1 with RNA polymerase II poised on human genes. *Mol Cell*. **45**, 38-50 (2012)

Combe C.W., Graham M., Kolbowski L., Fischer L., Rappsilber J. xiVIEW: Visualisation of Crosslinking Mass Spectrometry Data. *J Mol Bio*. **436**(17): 168656. (2024)

Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: Cross-link network maps with residue resolution. *Mol Cell Proteomics.* **14**:1137-1147 (2015)

Conrad T, Ørom UA. Cellular Fractionation and Isolation of Chromatin-Associated RNA. Methods Mol Biol. 1468:1-9. (2017)

Cox J., Hein M., Luber C.A., Paron I., Nagaraj N., Mann M. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ *Mol Cell Proteomics*. 13(9):2513-26 (2014)

Cramer P., Armache K.-J., Baumli S. *et al*. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*. **37**, 337-352 (2008)

Cramer P., Bushnell D.A., Kornberg R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*. 292(5523):1863-76 (2001)

Czeko E., Seizl M., Augsberger C. *et al*. Iwr1 directs RNA polymerase II nuclear import. *Mol Cell* **42**, 261-266 (2011)

Dekker C., Stirling P. C., McCormack E. A. *et al*. The interaction network of the chaperonin CCT. *EMBO J*. **27**, 1827-1839 (2008)

Doblmann J., Dusberger F., Imre R., Hudecz O., Stanek F., Mechtler K., Dürnberger G. apQuant: Accurate Label-Free Quantification by Quality Filtering *J Proteome Res*. 18(1):535-541. (2019)

Domen P.L., Nevens J.R., Mallie A.K., Hermanson G.T, Klenk D.C. Site-directed immobilization of proteins. *J Chromtogr A*. **510**:293-302, 1990.

Edelheit, O., Hanukoglu, A. & Hanukoglu, I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol* **9**, 61 (2009)

Edwards A. M., Kane C. M., Young R. A. *et al*. Two dissociable subunits of yeast RNA polymerase II stimulate initiation of transcription at a promoter in vitro. *J Biol Chem*. **266**, 71-75 (1991)

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell* **45**, 111–122. (2012)

Fagerberg L., Hallström B. M., Oksvold, P. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. **13**, 397-406 (2014)

Farnung L., Ochmann M., Garg G., Vos S.M, Cramer P. Structure of a backtracked hexasomal intermediate of nucleosome transcription. *Mol Cell*. **82**(17): 3126-3134 (2022)

Fianu I., Chen Y., Denemann C., Dybkov O., Linde A., Urlaub H. and Cramer P. Structural basis of Integrator-mediated transcription regulation. Science **374**, 883-887. (2021)

Fianu I., Dienemann C., Aibara S. *et al*. Cryo-EM structure of mammalian RNa polymerase II in complex with human RPAP2. *Commun Biol*. **4**, 606 (2021)

Fianu, I., Ochmann, M., Walshe, J.L. *et al.* Structural basis of Integrator-dependent RNA polymerase II termination. *Nature* **629**, 219-227 (2024)

Finka, A., Mattoo, R. U. H. & Goloubinoff, P. Experimental milestones in the discovery of molecular chaperones as polypeptide unfolding enzymes. *Annu. Rev. Biochem.* **85**, 715–742 (2016).

Forget D., Lacombe A. A. & Cloutier P. The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. *Mol Cell Proteomics*, **9**, 2827-2839 (2010)

Forget D., Lacombe A.-A., Cloutier P. *et al*. Nuclear import of RNA polymerase II is coupled with nucleocytoplasmic shuttling of the RNA polymerase II-associated protein 2. *Nucleic Acid Res*. **41**(14): 6881-6891 (2013)

Forsberg, B.O., Shah, P.N.M. & Burt, A. A robust normalized local filter to estimate compositional heterogeneity directly from cryo-EM maps. Nat Commun 14, 5802 (2023).

Gao M., Guo G., Huang J. *et al*. Lou Z. DOCK7 protects against replication stress by promoting RPA stability on chromatin. *Nucleic Acids Res.* **49**(6): 3322-3337. (2021)

Gaspari M., Larsson NG., Gustafsson CM. The transcription machinery in mammalian mitochondria. *BBA*. **1659**: 2–3 (2004).

Girbig M., Misiaszek A. D., Vorländer M. K. *et al.* Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states. *Nat Struct Mol Biol* **28,** 210-219 (2021)

Goloubinoff, P., Sassi, A.S., Fauvet, B. *et al.* Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins. *Nat Chem Biol* **14**, 388–395 (2018).

Gong Z., Ye S.X, Nie Z.F., Tang C. The conformational preference of chemical cross-linkers determines the cross-linking probability of reactive protein residues. J Phys Chem B. 124:4446-4453. (2020)

Goodrich J.A., Tjian R. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **77**: 145–156. (1994)

Grantham J., Brackley K.I., Willison K.R. Substantial cct activity is required for cell cycle progression and cytoskeletal organization in mammalian cells. Exp. Cell Res. 2006;312:2309–2324.

Gras S., Chaumont V, Fernandez B. *et al.* Structural insights into a new homodimeric self-activated GTPase family. *EMBO Rep*. **8**, 569-575 (2007)

Guan H., Wang Y., Yu T., Huang Y., Li M., Saeed A.F.U.H., *et al*. Cryo-EM structures of the human PA200 and PA200-20S complex reveal regulation of proteasome gate opening and two PA200 apertures. *PLoS Biol* **18**(3): e3000654. (2020)

Guo J., Turek M. E., Price D. H. Regulation of RNA polymerase II termination by phosphorylation of Gdown1. *J Biol Chem*. **289**(18), 12657-12665 (2014)

Harlen K. M. & Churchman L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol*. **18**: 263-273 (2017)

He Y., Yan C., Fang J. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533,** 359-365 (2016)

Heidemann M., Hintermair S., Voβ K. *et al*. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochem Biophys Acta*. **1829**: 55-62 (2013).

Herz, H.-M., Hu D. & Shilatifard A. Enhancer malfunction in cancer. *Mol Cell*. **53**, 859-866 (2014)

Holstege F.C.P., van der Vliet P.C., Timmers H.T.M. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J* **15**:1666–1677. (1996)

Horwich, A. L. & Fenton, W. A. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q. Rev. Biophys.* **42**, 83–116 (2009).

Hou T. Y., Kraus W.L. Come one, come all? Re-evaluating RNA polymerase II pre-initiation complex assembly using single-molecule microscopy. Mol Cell. **81**(17):3443-3445. (2021)

Hsu, P.L., Yang, F., Smith-Kinnaman, W., Yang, W., Song, J.E., Mosley, A.L., and Varani, G. Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD. *J Mol Biol*. **426**, 2970-2981. (2014)

Hu X., Malik S., Negroiu C. C. *et al*. A Mediator-responsive form of metazoan RNA polymerase II. *PNAS* **103**(25), 9506-9511 (2006)

Imbalzano A.N., Zaret K.S., Kingston R.E.. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J Biol Chem* **269**: 8280–8286. (1994)

Ishida R, Okamoto T, Motojima F, Kubota H, Takahashi H, Tanabe M, Oka T, Kitamura A, Kinjo M, Yoshida M, Otaka M, Grave E, Itoh H. Physicochemical Properties of the Mammalian Molecular Chaperone HSP60. *Int J Mol Sc*i.19(2):489. (2018)

Jeronimo C., Forget D., Bouchard A. *et al*. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol Cell*. **27**(2): 262-274 (2007)

Jiang Y., Liu M., Spencer C.A., Price D.H. Involvement of Transcription Termination Factor 2 in Mitotic Repression of Transcription Elongation. 14(3):375-386. (2004)

Jishage M., Malik S., Wagner U. *et al*. Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. *Mol Cell*. **45**, 51-63 (2012)

Jishage M., Yu X., Shi Y. *et al.* Architecture of Pol II(G) and molecular mechanism of transcription regulation by Gdown1. *Nat Struct Mol Biol* **25,** 859–867 (2018)

Kao, A. Chiu C., Vellucci D., Yang Y., Patel V.R., Guan S., Randall A., Baldi P., Rychnovsky S.D, Huang L. Development of a Novel Cross-linking Strategy for Fast and Accurate Identification of Cross-linked Peptides of Protein Complexes. *Mol Cell Proteom*. **10**(1) (2011)

Kassube S. A., Jinek M., Fang J. *et al.* Structural mimicry in transcription regulation of human RNA polymerase II by the DNA helicase RECQL5. *Nat Struct Mol Bio*. **20**(7): 892-899 (2013)

Kim Y, Geiger JH, Hahn S, Sigler PB. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**:512–520 (1993)

Kimura M., Ishiguro A. & Ishihama A. RNA polymerase II subunits 2, 3 and 11 form a core subassembly with DNA binding activity. *J Biol Chem*. **272**, 25851-25855 (1997)

Klinge, S. & Woolford, J.L. Ribosome assembly coming into focus. *Nat Rev Mol Cell Biol* **20,** 116–131 (2019)

Koch, B. *et al*. Generation and validation of homozygous fluorescent knock-in cells using CRISPR–Cas9 genome editing. *Nat. Protoc.* 13:1465–1487 (2018).

Kosugi S., Hasebe M., Tomita M. *et al*. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins prediction of composite motifs. *Proc Natl Acad Sci USA*. **106**, 10171-10176 (2009)

Kubala M. H., Kovtun O., Alexandrov K. *et al*. Structural and thermodynamic analysis of the GFP:GFP-nanobody complex. *Protein Sci*. **19**(12), 2389-2401 (2010)

Lagerwaard I.M., Albanese P., Jankevics A., Scheltema R.A. Xlink Mapping and AnalySis (XMAS)—Smooth Integrative Modeling in ChimeraX. bioRxiv [Internet]: 2022.04.21.489026. (2022)

Langer, T., Lu, C., Echols, H. *et al.* Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356**, 683–689 (1992).

Leitner A., Joachimiak L.A., Unverdorben P., *et al*. Aebersold R. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. Natl Acad Sci USA. 111:9455-9460. (2014)

Levine M., Cattoglio C. & Tijan R. Looping back to leap froward: transcription enters a new era. *Cell*. **157**, 13-25 (2014)

Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology Vol. 3, Art. 3, (2004)

Liu M., Z. Xie, Price D.H. A human RNA polymerase II transcription termination factor is a SWI2/SNF2 family member. *J. Biol. Chem*., 273.:25541-25544 (1998)

Lu F. H., Weinmann R. & Reinberg D. The nonphosphorylated form of RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci USA*. **88**: 10004-8 (1991)

Luse D. S. The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled? *Transcription*. **5**(1), e27050 (2014)

Makhnevych T. & Houry W. A. The role of Hsp90 in protein complex assembly. *Biochem Biophys Acta*. **1823**(2), 674-682 (2012)

Malik S. & Roeder R. G. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci*. **30**(5), 256-263 (2005)

Mapp A. K., Pricer R. & Sturlis S. Targeting transcription is no longer a quixotic quest. *Nat Chem Biol*. **11**(12): 891-894 (2015)

Martino F., Pal M., Munoz-Hernandez H. *et al*. RPAP3 provides a flexible scaffold for coupling HSP90 to the human R2TP co-chaperone complex. *Nat Commun*. **9**, 1501 (2018)

Max, T., Søgaard, M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* **282**:14113–14120 (2007)

Mayeda, A., Krainer, A.R. Preparation of Hela Cell Nuclear and Cytosolic S100 Extracts for In Vitro Splicing. *Humana Press*. In: Haynes, S.R. (eds) RNA-Protein Interaction Protocols. Methods in Molecular Biology™, **118**. (1999)

Mayer M. Gymnastics of molecular chaperones. Mol Cell. 39:321-331 (2010)

McClellan A. J., *et al.* Diverse cellular functions of the Hsp90 molecular chaperone uncovered using system approaches. *Cell*. **131**, 121-135 (2007)

Méndez-Hernández E., Pérez-Mejía A., Lara-Chacón B. *et al*. and Calera M. Gpn1 and Gpn3 associate tightly and their protein levels are mutually dependent in mammalian cells. FEBS Letters. **588**: 3823-3829. (2014)

Minaker S. W., Filiatrault M. C., Ben-Aroya S. *et al.* Biogenesis of RNA polymerases II and III requires the conserved GPN small GTPases in Saccharomyces cerevisiae. *Genetics*. **193**(3), 853-864 (2013)

Munoz, I. G. *et al*. Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nature Struct. Mol. Biol.* **18**, 14–19 (2011).

Nguyen V. T., Giannoni F., Dubois M. F. *et al.* In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Res*. **24**, 2924–2929 (1996)

Nishimura K., Fukagawa T., Takisawa H. *et al*. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods.* **6**, 917-922 (2009)

Pacheco-Fiallos, B., Vorländer, M.K., Riabov-Bassat, D. *et al*. mRNA recognition and packaging by the human transcription–export complex. *Nature* **616**:828–835 (2023).

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

Plaschka C., Hantsche M., Dienemann C., Burzinski C., Plitzko J., Cramer P. Transcription initiation complex structures elucidate DNA opening. *Nature* **533**:353–358. (2016)

Pühringer T., Hohmann U., Fin L. *et al*. Structure of human core transcription-export complex reveals a hub for multivalent interactions. *eLife*. **9**:e615053 (2020)

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

Ramsey E. P., Abascal-Palacios G., Daiß J. L. *et al.* Structure of human RNA polymerase III. *Nat Comm*. **11**:6409 (2020)

Roginski R. S., Mohan Raj B. K., Birditt B. *et al*. The human GRINL1A gene defines a complex transcription unit, an unusual form of gene organization in eukaryotes. *Genomics* **84**(2), 265-276 (2004)

Salema V., Fernandez L. A. High yield purification of nano bodies from the periplasm of E. coli as fusions with the maltose binding protein. *Protein Expr Purif.* **91**(1), 42-48 (2013)

Sato S., Tomomori-Sato C., Parmely T. J. *et al*. A set of consensus mammalian Mediator subunits identified by multidimensional protein identification technology. *Mol Cell*, **14**, 685-691 (2004)

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

Schier A.C., Taatjes D.J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* 34:465-488. (2022)

Schwanhäusser B., Busse D., Li N. *et al*. Global quantification of mammalian gene expression control. *Nature* 473:337–342 (2011)

Shahbabian K., Chartrand P. Control of cytoplasmic mRNA localization. *Cell Mol Life Sci*. **69**(4):535–552. (2011)

Sharma S.K., De los Rios P., Christen P., Lustig A., Goloubinoff P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nature Chem Biol*. 6:914-920 (2010).

Shiau, A. K., Harris, S. F., Southworth, D. R. & Agard, D. A. Structural analysis of *E. coli* hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **127**, 329–340 (2006).

Siwaszek A., Ukleja M., Dziembowski A. Proteins involved in the degradation of cytoplasmic mRNA in the major eukaryotic model systems. *RNA Biol*. **11**(9):1122-1136. (2014)

Somesh B. P., Reid J., Liu W.-F., *et al*. Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell*. **121**(6), 913-923 (2005)

Stahl K., Graziadei A., Dau T. *et al.* Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nat Biotechnol* **41**:1810–1819 (2023)

Staresincic L., Walker J., Dirac-Svejstrup A. B. *et al*. GTP-dependent binding and nuclear transport of RNA polymerase II by Npa3 protein. *J Biol Chem*. **286**, 35553-35561 (2011)

Stark H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol*. **481**, 109-126 (2010)

Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Rev. Mol. Cell Biol.* **11**, 515–528 (2010).

Tang G., Peng L., Baldwin P.R. *et al.* EMAN2: an extensible image processing suite for electron microscopy, *J Struct Biol*. **157**(1), 38-46 (2007)

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

Verma R., Oania R., Fang R. *et al.* Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. *Mol Cell.* **41**(1), 82-92 (2011)

Vos S.M., Farnung L., Urlaub H. *et al.* Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* **560**, 601–606 (2018)

Wang R. & Brattain M. G. The maximal size of protein to diffuse through the nuclear pore is larger than 60 kDa. *FEBS Lett*. **581**(17), 3164-3170 (2007)

Wang X., Qi Y., Wang Z. *et al*. Chen X. RPAP2 regulates a transcription initiation checkpoint by inhibiting assembly of pre-initiation complex. *Cell Rep*. 39(4): 110732 (2022)

Watabe-Uchida M., John K.A., Janas J.A., Newey S.E, Van Aelst L. The Rac Activator DOCK7 Regulates Neuronal Polarity through Local Phosphorylation of Stathmin/Op18. *Neuron.* 51(6):727-739 (2006)

Wild T. & Cramer P. Biogenesis of multisubunit RNA polymerases. *Trends Biochem Sci* **37**(3), 99-105 (2012)

Wilson M. D., Harreman M., Svejstrup J.Q. Ubiquitylation and degradation of elongating RNA polymerase II: The last resort. *Biochimica Et Biophysica Acta*. (2012)

Wong K. H, Jin Y. & Struhl K. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter complex. *Mol Cell*. **54**, 601-612 (2014)

Xiang, K., Manley, J.L., and Tong, L. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. *Nat Commun* 3,946. (2012)

Yang L., Mali P., Kim-Kisalek C. *et al*. CRISPR-Cas-mediated targeted genome editing in human cells. *Methods Mol Biol*. **1114**, 245-267 (2014)

Yu M. *et al*., RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386 (2015).

Zecha J., Satpathy S., Kanashova T. *et al*. TMT labeling of masses: A robust and cost-efficient, in-solution labelling approach. *Mol Cell Proteomics*. 18(7): 1468-1478 (2019)

Zeng F., Hua Y., Liu X. *et al.* Gpn2 and Rba50 directly participate in the assembly of the Rpb3 subcomplex in the biogenesis of RNA polymerase II. *Mol Cell Biol*. **38**(13), e00091-00018 (2018)

Zhu Z., Liu J, Feng H., Zhang Y., Huang R., Pan Q., Nan J., Miao J. and Cheng B. Overcoming the cytoplasmic retention of GDOWN1 modulates global transcription and facilitates stress adaptation. *eLife* **11**:e79116 (2022)

Zivanov, J. *et al*. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166. (2018).