Dissertation zur Erlangung des Doktorgrades der Fakultät für Chemie und Pharmazie der Ludwig-Maximilians-Universität München

Structure and function of the GPN-loop GTPase Npa3 and implications for RNA polymerase II biogenesis

Jürgen Gerd Niesser

aus

Laupheim, Deutschland

2015

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28.11.2011 von Herrn Prof. Dr. Patrick Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 10.09.2015

Jürgen Niesser

Dissertation eingereicht am 10.09.2015

1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: PD Dr. Dietmar Martin

Mündliche Prüfung am 30.09.2015

Acknowledgements

Success is not the key to happiness. Happiness is the key to success.
If you love what you are doing, you will be successful.
- Albert Schweitzer (1875-1965) -

People who truly live this spirit are quite rare to find but if they get together a great team is born, both professionally and socially. Undoubtedly, the Gene Center and especially the laboratory of Prof. Dr. Patrick Cramer is such a place and I am very happy that I was part of this family during my PhD thesis.

But it's not only the great people you have selected, *Patrick*, it's the way how you have formed and how you lead the team that makes your lab such a special environment. Based on trust and not control, on motivation and not pressure – you gave me the freedom to develop my own ideas, follow them up and finally solve all these exciting scientific questions. You have provided everything a young scientist can dream of, is it state-of-the-art equipment, a huge scientific network, the chance to attend international conferences in USA and China and the possibility to enable expensive experiments. Your passionate way of tackling scientific projects, your tremendous experience in science and far beyond and the great balance between freedom and support of your people will certainly influence my future carrier. For all those things I want to thank you, *Patrick*!

Further I want to thank the whole group – you guys are one of the main reasons why I loved so much to come to work every day during all those exciting years! Finding such a cheerful and warm-hearted atmosphere in the lab every morning is invaluable and, as Albert Schweitzer already said, the driving force to the success of the whole lab.

But there are some very special scientists, colleagues and friends I especially want to thank. *Wolfgang*, for being an amazing, unique buddy, thanks for our daily discussions about the world including everything on it, for sharing silver medals, your flat, parties, ideas, authorships and lots of fun. *Merle*, thanks for being an unforgettable bench neighbor and synchrotron buddy for sharing ideas, thoughts, sarcasm, your bench and your cheek. *Margaux*, thanks for all those beer garden sessions, coffee breaks, your cute french accent, and for having the most infectious laughter I know. Thanks for countless legendary moments in the lab and far beyond and to keep it short, you're all awesome!

Additionally I want to thank other great scientists, colleagues and friends: *Sarah*, thanks for all your advice and discussions and carefully proofreading this thesis, for sharing a Nature paper and all your experience in crystallography, you have great part in the success of this thesis! Thanks, *Björn* for daily procedures, *Sofia* for all your craziness, *Youwei* for being a brave table tennis opponent, *Michi* for introducing me in the world of bosseln, *Claudia*, for being an amazing lab backbone and always being positive and *Stefan* for Pol II fermentations and soccer discussions. Further I thank *Christoph*, *Gregor,*

Rike, *Andi*, *Thomas*, *Alan*, *Fuensanta*, *Kerstin (K+M)*, *Dirk*, *Simon*, *Hauke*, *Clemens*, *Carlo*, *Schulz,* and *Tobi* for many scientific discussions, advice and help and for the extraordinary atmosphere in the lab.

In particular I also thank *David* and *Kristin* for greatly holding the fort in the Munich satellite lab and also many people of the Hopfner lab.

The 'rooftop girls', *Romy*, *Susi* and *Anja*, I want to thank for many amazing incubation times.

Thanks to all members of my thesis committee: *Dr. Dietmar Martin, Prof. Dr. Mario Halic, Prof. Dr. Klaus Förstemann, Prof. Dr. Roland Beckmann* and *Prof. Dr. Karl-Peter Hopfner* for your support and time. *Dietmar*, I additionally want to thank you for the many times where you showed me how it feels to sit on the other side of the table at exams and your constant interest and support in my ongoing projects.

Thanks to the members of my thesis advisory committee: *Prof. Dr. Aymelt Itzen* and *Prof. Dr. Mario Halic* for very helpful scientific advice, discussions and support.

Further I want to thank many great people of the outstanding Grk1721 PhD program especially *André*, an amazing collaborator, retreat roomie and all kinds of social events friend, *Anindya* our Munich Comedy Club Star, the always laughing *Carina* and *Julia*, *Florian*, *Gabi* and *Matthias* and many others for profound discussions both science-related and not so much, after seminar drinks, beer garden sessions and awesome retreats and also *Petra* for keeping the program running and giving me the chance to be part of the retreat organization team.

I'm also very thankful to *Karina*, *Sabine* and *Laura* of the MPI-bc crystallization facility, *Dr. Andreas Bracher* (MPI-bc) and *Dr. Alexander Bepperling* (Sandoz) for advice concerning chaperone issues and the beamline staff of the synchrotrons DESY (EMBL Hamburg), ESRF (EMBL, France) and SLS (PSI, Switzerland), especially *Dr. Tobias Weinert* for advice on S-SAD data collection strategies.

I also especially want to thank my research students, *Felix*, *Natalie* and *Bronislava* for their enthusiasm, motivation and fun at work. You all did a great job and I'm looking forward to follow your scientific carriers!

Moreover, I owe special thanks to my brother *Maso* and all my friends especially *Daniel* (for making me godfather Jürgen), *Florian* and *David*. We have certainly shaped our personalities throughout our whole lives and shared all the most valuable moments!

Finally, very special thanks to you *Juli*. Thanks for being a great, authentic, loyal personality, woman and girlfriend, for your 'you gonna make it package', all your support and love and simply for being as you are. Thanks for sharing countless unforgettable moments, future plans and life. Your smile always reminds me what's truly important in life!

Ganz besonderer Dank gilt *meinen Eltern*. Ihr habt mich zu dem gemacht was ich heute bin und mich immer in allen Lebenslagen unterstützt. Ohne euch würde ich diese Zeilen heute nicht schreiben und dafür danke ich euch von ganzem Herzen!

Summary

Most eukaryotic proteins form complexes, but how these complexes are assembled in the cell often remains unknown. A prominent eukaryotic protein complex is RNA polymerase II (Pol II), a 12-subunit, 520-kDa enzyme that carries out transcription of protein-coding genes. The structure and function of Pol II has been studied extensively, but little is known about its biogenesis. Assembly of Pol II apparently occurs in the cytoplasm prior to its nuclear import. Pol II biogenesis requires all three members of the recently discovered GPN-loop GTPases, but the function of these enzymes is unknown. The family is characterized by two protein insertions and a highly conserved motif consisting of the amino acids Gly-Pro-Asn (single letter code: GPN) that is suggested to function in GTP hydrolysis. Depletion or mutation in the nucleotide binding site or GPN motif of human GPN1 (also called RPAP4, XAB1, MBD*in*) or its yeast homolog Npa3 leads to cytoplasmic accumulation of Pol II, but it is unknown whether these enzymes are involved in nuclear import and/or assembly. Lack of structural data for any eukaryotic GPN-loop GTPase so far prevented detailed molecular understanding of these essential enzymes.

In this thesis we report crystal structures of the GPN-loop GTPase Npa3 from the yeast *S. cerevisiae*. The enzyme was trapped in a GDP-bound, closed conformation, that shows eukaryote-specific features in both insertion regions at 2.3 Å resolution. Further we show a GTP analog-bound structure at 2.2 Å resolution that reveals a novel, open conformation displaying a conserved hydrophobic pocket distant from the active site. We show that both insertion regions rearrange upon transition from the closed to the open state and provide atomic details of how Npa3 binds the nucleotides. Using site-directed mutagenesis, enzymatic activity assays and molecular modelling we elucidate the molecular mechanism of Npa3 hydrolysis from GTP to GDP that involves the GPN motif.

We further show that Npa3 has chaperone activity and interacts with hydrophobic regions of Pol II subunits that form interfaces in the assembled Pol II complex. Consistent with a function as Pol II assembly chaperone, we show that Npa3 does not interact with mature, assembled Pol II. Biochemical results are in agreement with a model that the hydrophobic pocket binds peptides, and that this can allosterically stimulate GTPase activity and subsequent peptide release.

Thus, our results indicate that GPN-loop GTPases form a new family of assembly chaperones for Pol II and maybe other protein complexes.

Publications

Part of this work is in the process of publication:
Jürgen Niesser, Felix R. Wagner, Dirk Kostrewa, Wolfgang Mühlbacher and Patrick Cramer. “Structure of a GPN-loop GTPase chaperone and RNA polymerase II assembly factor.”
In preparation

Author contributions: J.N. carried out experiments, structure determination and modelling. F.R.W. supported Pol II-Npa3 interaction studies. D.K. and W.M. advised on X-ray data processing and analysis. P.C. initiated and supervised research. J.N. and P.C. prepared the manuscript.

Additional publications:
Wolfgang Mühlbacher, Andreas Mayer, Mai Sun, Michael Remmert, Alan Cheung, **Jürgen Niesser**, Johannes Soeding and Patrick Cramer. “Structure of Ctk3, a subunit of the RNA polymerase II CTD kinase complex, reveals a noncanonical CTD-interacting domain fold.”
Proteins, 2015, 83(10):1849-58.

Sarah Sainsbury, **Jürgen Niesser** and Patrick Cramer. “Structure and function of the initially transcribing RNA polymerase II-TFIIB complex.”
Nature 2013, 493, 437-440.

Contents

1. Introduction

1.1. Transcription

1.1.1. DNA-dependent RNA-Polymerases

Gene transcription by DNA-dependent RNA polymerases is one of the most fundamental processes in all living organisms. Whereas bacteria and archaea rely on a single RNA polymerase, eukaryotes have evolved at least three specialized RNA polymerases (Pol I, II and III) that synthesize distinct subsets of RNA molecules. (Cramer *et al.*, 2008; Werner & Grohmann, 2011). Pol I is located in the nucleoli and produces most ribosomal RNAs (rRNA). In the nucleoplasm, Pol II transcribes messenger RNAs (mRNA) and other small RNAs and Pol III synthesizes transfer RNAs (tRNA), 5S rRNAs and other small RNAs. In plants, two additional polymerases called Pol IV and Pol V transcribe non-coding RNAs required for gene silencing (Matzke *et al.*, 2009). Further, chloroplasts and mitochondria contain their own, phage-related, single-subunit RNA polymerases that specifically transcribe the DNA of these organelles (Cheetham & Steitz, 1999).

Eukaryotic Pol I, II and III consist of 14, 12 and 17 subunits, respectively (Table 1). Five core subunits are shared among all eukaryotic enzymes and two between Pol I and Pol III (Vannini & Cramer, 2012). The crystal structure of the 10 subunit Pol II core (Cramer *et al.*, 2000) provided first structural insights and was later extended to a complete 12 subunit model that also shows the so-called stalk (Armache*, et al.*, 2005). Structural data derived from homology modeling is also available for Pol III (Jasiak *et al.*, 2006) and the crystal structure of Pol I was solved recently (Engel *et al.*, 2013). Even though the multi-subunit RNA polymerases synthesize distinct RNAs, they show great structural conservation (Vannini & Cramer, 2012). The two largest subunits that form the active center cleft are related to those of the bacterial polymerase and two subunits of the eukaryotic enzymes contain the bacterial α motif, respectively (Werner & Grohmann, 2011). Consistent with the structural conservation, eukaryotic Pol I, II and III share common features including a conserved mechanism of transcription initiation that requires interaction with transcription factors.

Although the structure and function of eukaryotic RNA polymerases has been studied extensively during the last decades (Cramer*, et al.*, 2008), only little is known about biogenesis of these essential molecular machines (Wild & Cramer, 2012).

Table 1| Subunit composition of RNA polymerases.
Adapted from (Werner & Grohmann, 2011; Vannini & Cramer, 2012)

	Pol I	Pol II	Pol III	Bacteria
Polymerase core				
- large subunits	A190	Rpb1	C160	β'
	A135	Rpb2	C128	β
- partially shared/α-motif containing	AC40	Rpb3	AC40	α
	AC19	Rpb11	AC19	α
- specific	A12.2	Rpb9	C11	-
- shared	Rpb5	Rpb5	Rpb5	-
	Rpb6	Rpb6	Rpb6	ω
	Rpb8	Rpb8	Rpb8	-
	Rpb10	Rpb10	Rpb10	-
	Rpb12	Rpb12	Rpb12	-
Polymerase stalk	A14	Rpb4	C17	-
	A43	Rpb7	C25	-
TFIIF-like	A49	Tfg1	C37	-
	A34.5	Tfg2	C53	-
Pol III specific	-	-	C82	-
	-	-	C34	-
	-	-	C31	-
Number of subunits	14	12	17	5
M [kDa]	589	514	693	~400

1.1.2. Transcription cycle and regulation of RNA polymerase II

The transcription cycle of Pol II comprises five steps: pre-initiation, initiation, elongation, termination and recycling (Svejstrup, 2004). During pre-initiation Pol II and general transcription factors (GTFs; Table 2) are recruited to the promoter to form the pre-initiation complex (PIC), DNA is melted and the template strand is inserted into the polymerase active site (Sainsbury *et al.*, 2015). Initiation starts with the incorporation of the first RNA nucleotides till the nascent chain reaches a length of 12-13 nucleotides leading to an exchange of initiation to elongation factors and transition to the productive elongation phase. When Pol II reaches the 3' end of the gene termination occurs with the help of termination factors, the newly synthesized RNA and Pol II are released and the enzyme can be recycled for another round of transcription (Svejstrup, 2004).

Transcription regulation is coordinated by the Carboxy-terminal domain (CTD) of Rpb1 which is composed of heptapeptide repeats comprising the amino acids $Y_1S_2P_3T_4S_5P_6S_7$. Hereby, specific and reversible posttranslational modifications of the hydroxy-groups and proline isomerizations provide a unique 'CTD-code' that forms a binding platform for transcription factors only at specific stages of the cycle to modulate transcription (Kim *et al.*, 2010; Mayer *et al.*, 2010; Kubicek *et al.*, 2012; D. W. Zhang *et al.*, 2012; M. Zhang *et al.*, 2012; Eick & Geyer, 2013; Jasnovidova & Stefl, 2013). Additional regulation is obtained by particular secondary structures of nucleic acids (Allen *et al.*, 2004; Lehmann *et al.*, 2007; Aguilera & Garcia-Muse, 2012) and binding of accessory proteins to the DNA, nascent RNA and the Pol II core (Svejstrup, 2004).

1.1.3. General transcription factors in yeast

Pol II transcription initiation requires the general transcription factors TFIIA, -B, -D, -E, -F and -H (Sainsbury*, et al.*, 2015) (Table 2). In the canonical model, initiation starts with binding of the TFIID subunit TATA-binding protein (TBP) to promoter DNA were it induces an up to 90° kink of the DNA (J. L. Kim *et al.*, 1993; Y. Kim *et al.*, 1993). Then TFIIA and TFIIB bind the binary complex and TFIIB bridges between the TBP-DNA complex and Pol II (Bushnell *et al.*, 2004; Kostrewa *et al.*, 2009; X. Liu *et al.*, 2010; Sainsbury *et al.*, 2013), which is recruited to the promoter in complex with TFIIF to form the core PIC. The PIC is completed upon binding of TFIIE and TFIIH leading to an ATP-dependent DNA opening, the formation of a transcription bubble and initial RNA synthesis (Sainsbury*, et al.*, 2015). Once RNA reaches 12-13 nucleotides it clashes with TFIIB, triggering TFIIB displacement and formation of the elongation complex (Sainsbury*, et al.*, 2013).

Table 2| General transcription factors in yeast.
Adapted from (Sainsbury*, et al.*, 2015)

Factor	Subunits	Functions
TFIIA	2	Stabilizes TFIID-DNA complex, counteracts repressive effects of negative co-factors; stimulates constitutive and activated transcription.
TFIIB	1	Required for Pol II recruitment to the promoter, stabilizes TFIID-DNA complex, assists in transcription start site (TSS) selection and stimulates initial RNA synthesis; may aids in DNA-RNA strand separation and is critical for initiation-to-elongation transition.
TFIID	14-15	Nucleates Pol II recruitment and PIC assembly via binding of TBP to TATA-box promoters or TBP associated factors (TAFs) to TATA-less promoters. Further involved in chromatin remodeling and activator binding.
TFIIE	2	Facilitates recruitment of TFIIH to promoters and stimulates ATPase and kinase activities of TFIIH. Functions in promoter opening and stabilization of open DNA by binding to ssDNA.
TFIIF	2-3	Tightly associates with Pol II and stabilizes the PIC. Stimulates early RNA synthesis and aids in TSS selection.
TFIIH	10	Consists of a core that functions in promoter opening as an ATP-dependent translocase and DNA repair, and a kinase module that phosphorylates the CTD to facilitate initiation-to-elongation transition.

1.2. Life cycle of RNA polymerase II

Although the structure and function of Pol II has been studied extensively over the last decades (Cramer*, et al.*, 2008), only little is known about the life cycle of Pol II that comprises assembly, nuclear import and degradation or recycling of the subunits. During the last years a rising number of studies investigated these processes, leading to the identification of required proteins and novel insights into these essential cellular mechanisms. Here, a detailed summary about the current knowledge of the RNA polymerase II life cycle is provided.

1.2.1. Assembly

Biogenesis of Pol II apparently starts in the cytoplasm with the synthesis of its 12 subunits Rpb1-Rpb12, but how they assemble remains poorly understood.

Initial insights into subunit assembly derived from the bacterial RNA polymerase (Ishihama & Ito, 1972) because the eukaryotic Pol II core subunits Rpb1, Rpb2, Rpb3/11 and Rpb6 are homologous to the bacterial subunits β', β, α and ω, respectively (Zhang *et al.*, 1999; Cramer *et al.*, 2001; Vassylyev *et al.*, 2002). Urea-induced complete dissociation of bacterial polymerase subunits followed by stepwise *in vitro* reconstitution of the enzyme and analysis of the assembly intermediates led to the first model of bacterial RNA polymerase assembly (Ishihama & Ito, 1972). The postulated pathway starts with the association of two α-subunits to a αα-dimer which is then bound by the β-subunit to form a $\alpha_2\beta$ assembly intermediate (Figure 1C). The functional polymerase arises from binding of subunit β'. The remaining ω subunit is not essential but is thought to be involved in folding and stabilization of β', possibly joins the $\alpha_2\beta$ intermediate in complex with subunit β' and promotes assembly of both subcomplexes. (Ghosh *et al.*, 2001; Minakhin *et al.*, 2001). Recent NMR data indicates, that β'ω association is restricted to an early stage where the subunits are not completely folded yet (Drogemuller *et al.*, 2015). The postulated *in vitro* assembly pathway was confirmed *in vivo* (Ishihama, 1981).

Assuming a similar pathway for eukaryotic Pol II, the two subunits Rpb3 and Rpb11, that both contain the bacterial α motif (Werner & Grohmann, 2011), would form a complex that subsequently binds to Rpb2. Indeed, dissociation experiments with yeast Pol II revealed a Rpb2/3/11 subcomplex

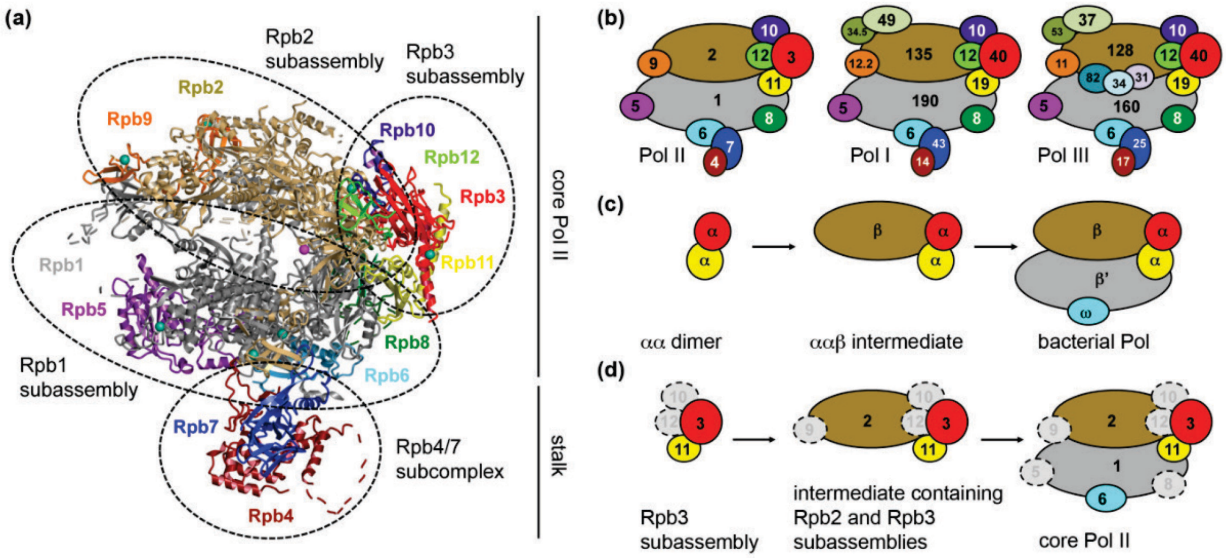


Figure 1| RNA polymerase subunit composition and assemblies.
(A) Crystal structure of Pol II (pdb-code: 1WCM)(Armache*, et al.*, 2005). Potential subassemblies are indicated. Eight zinc ions and the magnesium ion are shown as cyan and pink spheres, respectively. **(B)** Scheme of eukaryotic RNA polymerase subunit composition. Shared or homologous subunits are depicted in the same color. **(C)** Assembly of bacterial RNA polymerase. Subunit are colored according to their eukaryotic homologs as in (B). **(D)** Putative pathway for assembly of the Pol II core. Subunits shown in light grey (with broken outlines) have no homologs in bacterial RNA polymerases. Adapted from (Wild & Cramer, 2012).

that was stable at 6M urea (Kimura *et al.*, 1997). Additionally, Rpb10 was more weakly bound to this complex to form a Rpb2/3/11/10 subassembly at 4M urea. This is consistent with genetic studies, suggesting that Rpb10 association with the Rpb3 and Rpb11 homologs of Pol I and III (AC40 and AC19) has an important role in assembly and also indicates a conserved assembly pathway of the three polymerases (Lalo *et al.*, 1993). For *in vivo* assembly, the abundance of the individual subunits may also play an important role. Quantitative western blot analysis of all Pol II subunits in *S. pombe* revealed that Rpb3 is least abundant, thus limiting for complex formation, followed by Rpb1, Rpb2 and Rpb7 (2 fold more abundant) whereas the other subunits were 4-15 fold more abundant (Kimura *et al.*, 2001). Further the smaller subunits Rpb4-Rpb12 also existed, at least temporarily, in unassembled form. A critical function has been proposed for the common subunit Rpb12 (Rubbi *et al.*, 1999). The structure of the assembled Pol II core shows that Rpb12 interacts with both, Rpb3 and Rpb2, thus supporting a fundamental role of Rpb12 in biogenesis of all three polymerases as it bridges between the conserved second largest and α motif containing subunits (Rubbi*, et al.*, 1999; Cramer*, et al.*, 2000). It is thought that Rpb12 first binds to the Rpb3/11/10 subassembly to form a Rpb3/11/10/12 subcomplex that is then bound by a Rpb2/9 subcomplex (Wild & Cramer, 2012). In the last step of bacterial polymerase assembly the Rpb1 homolog β' binds the corresponding assembly intermediate and the Rpb6 homolog ω was implicated in assembly and stability of the largest subunit (Minakhin*, et al.*, 2001; Drogemuller*, et al.*, 2015). In yeast, temperature-sensitive Rpb1 mutant cells, characterized by transcriptional shutdown, can be rescued by overexpression of Rpb6 (Nonet *et al.*, 1987) underpinning the important role of this subunit in Rpb1 stability. In line with this, mutation of the Rpb6-interacting foot domain of Rpb1 leads to Pol II instability and assembly defects which can also be rescued by overexpression of Rpb6 (Garrido-Godino *et al.*, 2013). Additionally, substoichometric levels of Rpb8 were bound to Rpb1 in dissociation experiments using 4 M urea (Kimura*, et al.*, 1997). Thus, Rpb1 likely binds the other two subassemblies as Rpb1/5/6/8 complex to build the 10 subunit Pol II core (Wild & Cramer, 2012). The whole assembly pathway is further supported by pulse labeling experiments with temperature sensitive subunits Rpb1, Rpb2 and Rpb3 that showed an Rpb2/3 subcomplex that forms early after subunit synthesis and subsequently interacts with Rpb1 (Kolodziej & Young, 1991). The additional, peripheral subunits, such as the stalk, are likely added to the core afterwards as preassembled complexes (Wild & Cramer, 2012). Consistent with this, dissociation from the core has been reported for Rpb4/7 (Pol II) in solution (Edwards *et al.*, 1991), and C25/C17 (Pol III) during native mass spectrometry (Lorenzen *et al.*, 2007). Further the TFIIF-like subcomplex of Pol I (A49/A34.5) can be dissociated with urea (Huet *et al.*, 1975) and native mass spectrometry also revealed dissociation of its Pol III counterpart (C37/C53) as well as the Pol III-specific subunits C82/C34 (Lorenzen*, et al.*, 2007).

Taken together, the current model suggests the formation of an initial subcomplex, comprising subunits Rpb3/11/10/12 that subsequently binds to an Rpb2/9 subassembly (Wild & Cramer, 2012). This

large assembly intermediate is then bound by a third, Rpb1/5/6/8 subcomplex. Finally, the stalk comprising subunits Rpb4 and Rpb7 joins the Pol II core to form the complete functional 12 subunit Pol II complex (Armache*, et al.*, 2005) (Figure 1).

RNA polymerase II assembly factors

In contrast to bacterial RNA polymerase, the eukaryotic enzymes could so far never be reconstituted from individual subunits *in vitro*, indicating that their assembly *in vivo* depends on the help of several factors. Systemic affinity purification screens identified a number of factors that interact with eukaryotic RNA polymerases, both in yeast (Gavin *et al.*, 2002; Krogan *et al.*, 2006) and in human (Jeronimo *et al.*, 2007). However, their cellular function often remains enigmatic and involvement in a wide range of processes like polymerase assembly, cellular transport, transcription, disassembly or degradation is possible. To more specifically screen for putative Pol II assembly factors (Table 3), two recent studies purified human Pol II from conditions that enrich for partially assembled enzymes (Boulon *et al.*, 2010; Forget *et al.*, 2010). Mass spectrometry analysis identified proteins bound to Pol II complexes, including components of the R2TP-prefoldin-like chaperone complex, the large CCT chaperonin complex and all three members of a recently discovered subfamily of GPN-loop GTPases, called GPN1, GPN2 and GPN3 in human (Forget*, et al.*, 2010). In the second study cells were treated with α-amanitin, a transcription inhibitor that leads to degradation of Rpb1 (Nguyen *et al.*, 1996) and Rpb3 accumulation in the cytoplasm, thus allowing affinity purification of Rpb3-bound Pol II subcomplexes (Boulon*, et al.*, 2010). Here, a Pol II subcomplex comprising subunits Rpb2/3/11/10/12 was identified, consistent with previous data (Kimura*, et al.*, 1997; Cramer*, et al.*, 2000). This subcomplex was bound by the GPN-loop GTPases GPN1, GPN2 and GPN3 as well as RPAP1, RPAP2 and Grinl1a (Boulon*, et al.*, 2010) (Figure 2) that were known to bind Pol II (Jeronimo *et al.*, 2004; Jeronimo*, et al.*, 2007). Double treatment of cells with both, α-amanitin and leptomycin B, a specific nuclear export inhibitor of Xpo1 (also Crm1) (Fornerod *et al.*, 1997), led to cytoplasmic accumulation of Rpb1 (Boulon*, et al.*, 2010). This allowed purification of an Rpb1/8 subcomplex and identification of associated components of the R2TP/Prefoldin-like complex. The yeast R2TP/Prefoldin-like complex functions as Hsp90 co-chaperone (Zhao *et al.*, 2005) and its human homologous complex (Te *et al.*, 2007; Boulon *et al.*, 2008) contains the shared polymerase subunit Rpb5 (Sardiu *et al.*, 2008). The R2TP component hSpagh (also RPAP3) binds Hsp90 and likely recruits it to unassembled Rpb1 to maintain free Rpb1 stability and association with other Pol II subunits (Boulon*, et al.*, 2010). Hereby, the CTD heptapeptide repeat of cytoplasmic Rpb1 is mainly unphosphorylated. Further another component of the R2TP/Prefoldin-like complex, the yeast prefoldin Bud27, was implicated in biogenesis of all three RNA polymerases (Miron-Garcia *et al.*, 2013). Bud27 shows genetic interaction with Rpb5 and Rpb6, affects correct assembly of both subunits to Pol I, II and III and leads to growth defects and cytoplasmic accumulation of all three RNA polymerases upon deletion. Additionally, the karyopherin-like protein Rtp1 has been shown to interact with the R2TP

complex, Pol II subunits and nucleoporins but not with the yeast GPN1 homolog Npa3, and a function in assembly by binding to the N-terminal region of Rpb2 has been suggested (Gomez-Navarro *et al.*, 2013). Further, deletion of Rtp1 leads to cytoplasmic accumulation of Pol II subunits but not Pol I and Pol III subunits.

Taken together, *in vivo* assembly of Pol II requires the help of several factors but their molecular function is not very well understood. Lack of structural data for many biogenesis factors, especially the essential GPN-loop GTPases, so far prevented detailed understanding of their function in Pol II biogenesis.

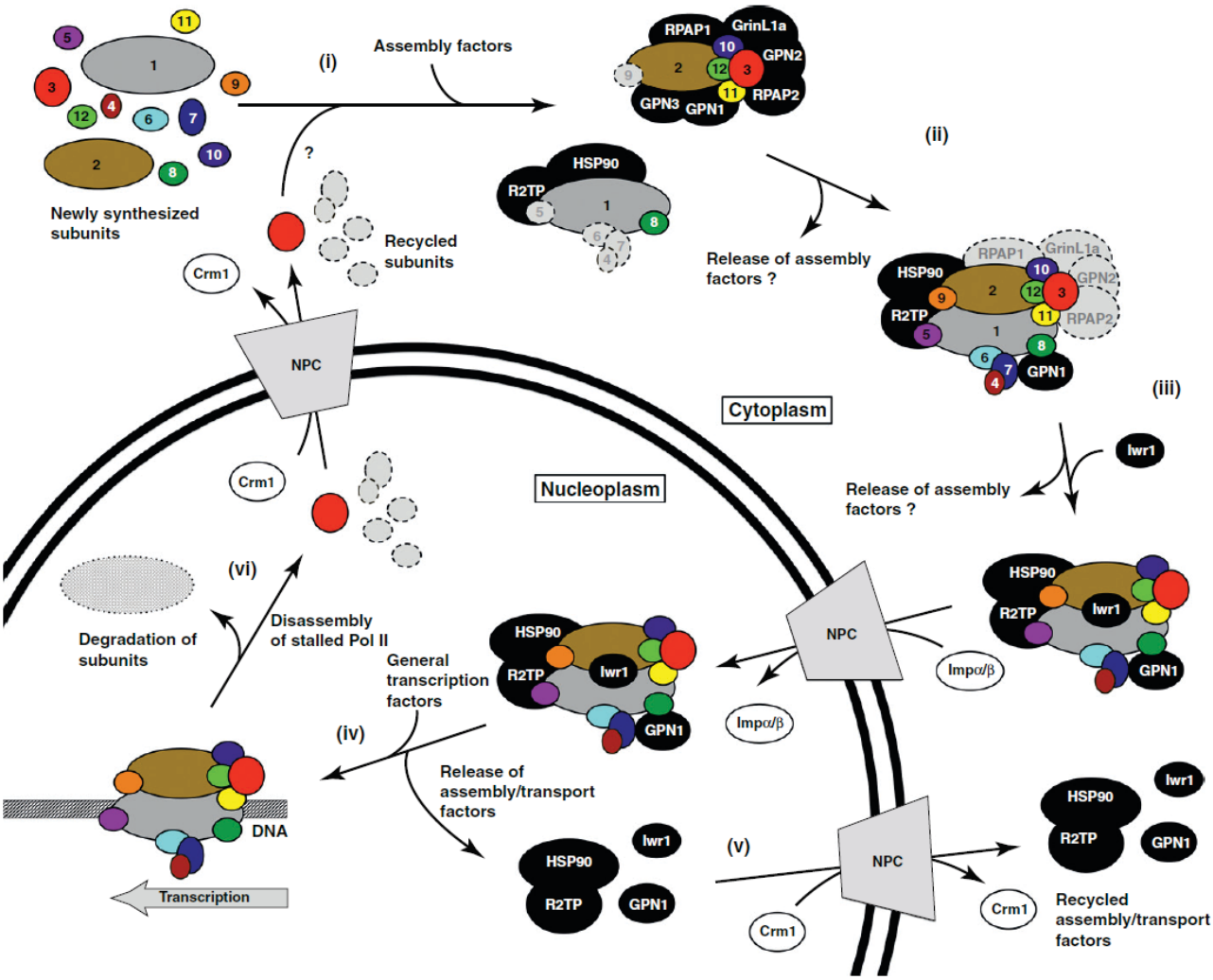


Figure 2| Model of Pol II biogenesis.
The model combines results from yeast and human. **(i)** Subunits derived from protein synthesis and nucleocytoplasmic recycling assemble with the help of assembly factors (black) that may stabilize assembly intermediates. **(ii)** Formation of two major subassemblies comprising subunit Rpb1 and Rpb2, respectively plus additional subunits (in color if determined experimentally or in grey if presumed to be present) and assembly factors (black, human names). Assembly factors shown in grey may leave assembling Pol II. **(iii)** Fully assembled Pol II is bound by the nuclear import adaptor Iwr1 which recruits importin α/β via its NLS for nuclear import. Some assembly factors may stay bound to Pol II. **(iv)** Biogenesis factors are released and general transcription factors bind Pol II for transcription initiation. **(v)** Nucleocytoplasmic recycling of assembly factors in a Crm1/Xpo1-dependant manner. **(vi)** Degradation or nucleocytoplasmic recycling of subunits derived from disassembly of stalled Pol II. Adapted from (Wild & Cramer, 2012).

Table 3| Putative assembly and import factors of RNA polymerase II.
Adapted from (Wild & Cramer, 2012; Forget *et al.*, 2013; Gomez-Navarro*, et al.*, 2013; Minaker *et al.*, 2013; Miron-Garcia*, et al.*, 2013; Guglielmi *et al.*, 2015)

Factor (human/yeast)	Functional data
GPN1/Npa3	Interaction with assembling Pol II and subunits Rpb1, Rpb4 and Rpb7. Depletion or mutation leads to cytoplasmic accumulation of Pol II subunits. Accumulates in the cytoplasm with Rpb1 in disease characterized by protein aggregates where it shows increased expression levels.
GPN2/YOR262W	Interaction with assembling Pol II. Mutation leads to cytoplasmic accumulation of Pol II subunits
GPN3/YLR243W	Depletion or mutation leads to cytoplasmic accumulation of Pol II subunits
GrinL1a/-	Interaction with assembling Pol II
RPAP1/RBA50	Interaction with assembling Pol II
RPAP2/RTR1	Interaction with assembling Pol II and GPN1
Hsp90/Hsp82	Interaction with assembling Pol II and Pol I. May stabilizes Rpb1 and assists in Pol II assembly.
SLC7A6OS/Iwr1	Iwr1 binds Pol II between Rpb1 and Rpb2 and provides NLS for Pol II import in a Kap60/95-dependant manner. Depletion leads to Pol II import defect.
R2TP-Prefoldin complex	Interaction with assembling Pol II. Depletion of the R2TP component hSpagh/RPAP3 leads to decreased levels of cytoplasmic Rpb1 in Pol II assembly defect conditions. Deletion of yeast Bud27 affects Pol I, II and III assembly, leads to growth defect and cytoplasmic accumulation of all polymerases
Rtp1	Interaction with the R2TP complex and Pol II subunits. Affects Pol II nuclear import and may assist assembly of Rpb2/Rpb3 and its binding to Rpb1
CCT complex	Large chaperonin complex that interacts with GPN1 and Pol II subunits
Microtubules	Interference with microtubule integrity leads to cytoplasmic accumulation of Rpb1. Yeast Npa3 mutants are hypersensitive to microtubule depolymerization drug

1.2.2. Nuclear Import

Nuclear import of RNA polymerase II is most likely restricted to fully assembled enzymes, because depletion of any Pol II subunit leads to cytoplasmic accumulation of Pol II (Boulon*, et al.*, 2010). Since none of the 12 Pol II subunits contain a nuclear localization sequence (NLS), additional proteins are required to mediate its nuclear translocation. Here, a critical role is attributed to Iwr1 which contains a N-terminal bipartite NLS (Czeko *et al.*, 2011). Deletion of yeast Iwr1 or mutation of its NLS leads to cytoplasmic accumulation of Pol II subunits (Czeko*, et al.*, 2011). This effect is specific for Pol II because Pol I and Pol III subunits are unaffected. Iwr1 functions as adaptor protein that binds Pol II and recruits importin-α/importin-β (Kap60/95) via its NLS. Electron microscopy revealed that Iwr1 binds Pol II between its two largest subunits, thus restricting Pol II import to fully assembled enzymes. Further nucleocytoplasmic recycling of Iwr1 is required because deletion of its nuclear export sequence (NES) (Peiro-Chova & Estruch, 2009) leads to Pol II import defects in yeast (Czeko*, et al.*, 2011). This is in agreement with observations that treatment with leptomycin B, a specific nuclear export inhibitor of Xpo1 (also Crm1) (Fornerod*, et al.*, 1997), leads to cytoplasmic accumulation of Rpb1, indicating nuclear trapping of the Pol II import factor (Boulon*, et al.*, 2010). The import mechanism of Iwr1 is conserved between yeast and human because human Iwr1 partially rescues Pol II import defects caused by deletion of its yeast counterpart (Czeko*, et al.*, 2011). However, an alternative import pathway may exist since Iwr1 is not essential.

1.2.3. Recycling and Degradation

Disassembly and recycling or degradation of defective nuclear RNA polymerase is important to avoid nuclear aggregation and transcription defects. However, the underlying mechanisms are not very well characterized. Xpo1-dependent nucleocytoplasmic recycling of Rpb3 apparently takes place after disassembly of nuclear Pol II (Boulon*, et al.*, 2010). Indications for this came from experiments were α-amanitin induced specific degradation of nuclear Rpb1 (Nguyen*, et al.*, 1996) leads to cytoplasmic accumulation of Rpb3 (Boulon*, et al.*, 2010). This effect is rather caused by nucleocytoplasmic recycling than import defects of Rpb3 because additional treatment with the Xpo1 exportin inhibitor leptomycin B leads to nuclear accumulation of Rpb3 (Boulon*, et al.*, 2010). The recycled subunit might then be incorporated into a new Pol II complex in the cytoplasm.

Degradation of Pol II likely occurs individually for each subunit rather than *en bloc* because different half-lives were reported for each subunit, ranging from 22 min for Rpb1 to 87 min for Rpb9 whereas the shared subunits Rpb6, Rpb8 and Rpb12 were described as stable in the given time course (Belle *et al.*, 2006). Indeed, ubiquitination and proteasome-mediated degradation of Rpb1 from stalled Pol II has been reported both, in yeast and in human (Somesh *et al.*, 2005; Daulny & Tansey, 2009). The responsible factors including ubiquitin-ligases were identified and their mechanisms were described in detail (Huibregtse *et al.*, 1997; Beaudenon *et al.*, 1999; Somesh*, et al.*, 2005; Ribar *et al.*, 2007; Somesh *et al.*, 2007; Daulny *et al.*, 2008; Harreman *et al.*, 2009; Garrido-Godino*, et al.*, 2013). Further ubiquitination and degradation of Rpb8 upon DNA damage has been reported (Wu *et al.*, 2007). However, the fate of the other subunits is not known and requires further investigations.

1.3. GPN-loop GTPases

1.3.1. General principles and classification of GTPases

GTPases (also called G proteins) are a large and functionally diverse family of guanosine triphosphate (GTP) hydrolyzing enzymes that carry out a wide variety of biological functions such as signal transduction, cellular transport, and macromolecular complex assembly (Wittinghofer & Vetter, 2011). All GTPases share a common 160-180 residue G-domain with an α,β topology that harbors five so called G-motifs (G1-G5) required for nucleotide binding and hydrolysis (Bourne *et al.*, 1991; Wittinghofer & Vetter, 2011). Whereas G4 and G5 bind the guanine base to mediate nucleotide specificity, G1 (also called P-loop or Walker A motif) stabilizes the negative charge of the phosphate ions. The charge is further partially neutralized by an essential, octahedral coordinated Mg^{2+}-ion. Motifs G2 and G3 (also called Walker B motif) are involved in nucleotide sensing, GTP hydrolysis and conformational switching and belong to the most flexible regions in the rather rigid G-domain, termed ‘switch 1’ and ‘switch 2’,

respectively. The G-domain is often extended by protein insertions and additions of sequence elements or whole domains to carry out specific biological functions (Wittinghofer & Vetter, 2011).

The GTPase cycle is often highly regulated by additional factors such as guanine nucleotide exchange factors (GEFs), GTPase-activating proteins (GAPs) and guanine nucleotide dissociation inhibitors (GDIs) (Wittinghofer & Vetter, 2011). High nucleotide affinities (10 pM for Ras) and slow intrinsic dissociation rates (~10^{-4} s^{-1} for Ras) of most G proteins can be overcome by regulation through GEFs. GEFs reduce the affinity and increase dissociation of bound GDP from the cognate GTPase by orders of magnitude, thus facilitating GTP binding and effector protein interaction. The intrinsic GTP hydrolysis rate is usually very low (10^{-4}-10^{-5} s^{-1}) due to the thermodynamically unfavorable nucleophilic attack of the negatively charged γ-phosphate. Hereby GAPs stimulate the intrinsic hydrolysis of their cognate GTPase by orders of magnitude. Many GAPs protrude into the active site of the G protein and provide a catalytic residue and/or stabilize the intrinsic machinery. The most prominent example is the so called 'arginine-finger' which stabilizes the pentavalent transition state of the S_N2 reaction by providing a positively charged guanidinium group to the negatively charged β- and γ-phosphates of GTP. For translational GTPases this function is apparently substituted by binding of a monovalent cation at an equivalent position (Kuhle & Ficner, 2014). Further, some GAPs use a 'Gln/Asn thumb' that resembles the catalytic Gln residue of many G proteins that is thought to stabilize and polarize the nucleophilic water relative to the γ-phosphate (Scheffzek *et al.*, 1997; Vetter & Wittinghofer, 2001; Daumke *et al.*, 2004; Pan *et al.*, 2006; Schuette *et al.*, 2009). In contrast, prenylated G proteins can further be negatively regulated by GDIs that inhibit the cycle by binding to the prenyl groups of the cognate G protein to shuttle it between membrane compartments and inhibit its nucleotide dissociation (Wittinghofer & Vetter, 2011). GTPases activated by dimerization, so-called GADs do not rely on GEFs and GAPs and stimulate GTP hydrolysis by dimerization, most likely in concert with effector protein interaction, which in turn acts as GTPase coregulator (GCR).

Based on a unique set of sequence and structural signatures the superclass of GTPases can be divided into two classes, designated as TRAFAC and SIMIBI (Leipe *et al.*, 2002). The TRAFAC class (after tra nslation fac tors) is characterized by a highly conserved threonine preceding strand β3 (G2 motif), has adjacent anti-parallel strands β3 and β4 and comprises the most prominent examples including translation factors, heterotrimeric G proteins and the Ras superfamily. The SIMIBI class (after si gnal recognition particle, Mi nD and Bi oD) is characterized by a highly conserved aspartate residue after strand β2 (G2 motif), which is adjacent and parallel to the G3 flanking β-strand and involves the signal recognition particle (SRP) and its receptor SR as well as the GPN-loop GTPases (also called XAB1 family) and many others.

1.3.2. Structural insights into an archaeal GPN-loop GTPase

GPN-loop GTPases are characterized by a highly conserved motif consisting of the amino acids Gly-Pro-Asn (single letter code: GPN) and two protein insertions ('insertion 1' and 'insertion 2') of approximately 40 amino acids, respectively (Wittinghofer & Vetter, 2011). Insertion 1 harbors the GPN motif and is reminiscent to the insertion box of the SRP G domain (Freymann *et al.*, 1997). Structural insights came from the archaeal GPN1 homolog Pab0955 from *Pyrococcus abyssi* that shares 27% sequence identity with its human counterpart (Gras*, et al.*, 2007) (Figure 3). A set of crystal structures, both free and in complex with different nucleotides, provide mechanistic snapshots along the hydrolysis pathway. Pab0955 is homodimeric, independent of the bound nucleotide, both in the crystals and in solution

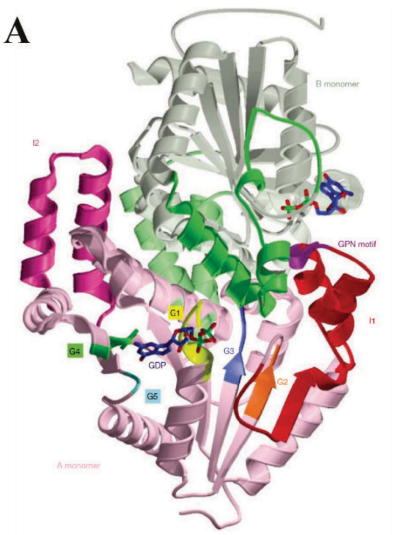


Figure 3| Structure of the archaeal GPN-loop GTPase Pab0955.
(A) Overall structure of the Pab0955·GDP dimer (pdb-code: 1YRA). Monomer's A and B are shown in pink and light green, respectively. GDP is shown as sticks and G1, G2, G3, G4 and G5 are shown in yellow, orange, blue, green and cyan, respectively. The two insertions I1 and I2 are depicted in fully and partially saturated colors, respectively. **(B)** Active site of the Pab0955·P_iGDP structure (pdb-code: 1YR9). Hydrogen bonds are shown as dashed lines and G motifs colored as in (A). The GPN motif is shown in purple and GDP and P_i as sticks. Adapted from (Gras*, et al.*, 2007)

(Gras *et al.*, 2005; Gras*, et al.*, 2007). The structures reveal that the GPN-loop of one monomer protrudes into the active site of the other monomer (Gras*, et al.*, 2007). Hereby, the 'trans'-GPN-loop binds the hydrolyzed orthophosphate in an anion whole formed by backbone amines and the Asn side chain. This neutralizes the negative charge of the phosphate ion and likely plays an essential role in catalysis. Further the Asn side chain may resemble the role of the catalytic Gln (G3 motif) residue of Ras, which is thought to stabilize the phosphate intermediate (Prive *et al.*, 1992). Biochemical data show weak intrinsic GTPase activity (0.012 µM hydrolyzed GTP per min and mg protein), no ATPase activity and autophosphorylation of itself in the presence of GTP and Mg^{2+} (Gras*, et al.*, 2007). Pull-down assays with cellular extracts and screening by surface plasmonic resonance identified the DNA-binding proteincomplexes Topoisomerase VI (subunit B) and the Replication Factor Complex RF-C (small subunit) as potential interaction partners. However, no large conformational changes of the different nucleotide states were observed and the cellular function remained largely unclear.

1.3.3. Eukaryotic Npa3/GPN1 and its paralogs GPN2 and GPN3

Eukaryotic cells contain three paralogs of GPN-loop GTPases. In human, these are called GPN1 (also RPAP4, XAB1, or MBD*in*), GPN2, and GPN3 (also Parcs). Archaea contain a single GPN-loop GTPase, and prokaryotes lack homologs. Homo- and heterodimerization of GPN1 and its paralogs were reported (Carre & Shiekhattar, 2011; Alonso *et al.*, 2013; Minaker*, et al.*, 2013; Mendez-Hernandez *et al.*, 2014). Additionally to the characteristic two insertions in the G domain and the GPN motif, GPN1 and its yeast homolog Npa3 contain a C-terminal tail with unknown function and low conservation, which is truncated in its paralogs and lacking in archaea. GPN1/Npa3 contains a nuclear export sequence (NES) (Reyes-Pardo *et al.*, 2012), consistent with the predominant cytoplasmic localization of Npa3 in yeast (Huh *et al.*, 2003; Dez *et al.*, 2004) and GPN1 in human cells (Nitta *et al.*, 2000; Lembo *et al.*, 2003; Forget*, et al.*, 2010; Reyes-Pardo*, et al.*, 2012). However, it may also play nuclear roles because Xpo1-dependent nucleocytoplasmic shuttling of the enzyme was reported (Forget*, et al.*, 2010; Staresincic *et al.*, 2011; Reyes-Pardo*, et al.*, 2012; Forget*, et al.*, 2013). In the yeast *Saccharomyces cerevisiae,* deletion of Npa3 or its paralogs GPN2 (YOR262W) and GPN3 (YLR243W) is lethal (Giaever *et al.*, 2002), indicating essential, non-redundant functions of these enzymes.

GPN1 was initially identified in human cells to bind the DNA repair protein XPA (Nitta*, et al.*, 2000), which functions as assembly platform for the DNA repair machinery (de Laat *et al.*, 1999) and was thus named 'XPA binding protein 1' (XAB1) (Nitta*, et al.*, 2000). However, the suggested role of GPN1 in nuclear import of XPA could not be confirmed (Li *et al.*, 2013). Further, interaction of human GPN1 with MBD2, a component of the large protein complex 'Methyl-CpG-binding protein 1' (MeCP1) was reported, which represses transcription of densely methylated genes and led to the name 'MBD2-interacting protein' (MBD*in*) (Lembo*, et al.*, 2003). Here, the interaction required both, a functional G-domain and the C-terminal acidic tail. A GPN1 homolog was identified in *S. cerevisiae* in a screen for ribosomal biogenesis factors and called 'Nuclear preribosomal-associated protein 3' (Npa3) (Dez*, et al.*, 2004). The corresponding homolog was also found in the yeast *Schizosaccharomyces pombe* (Aves *et al.*, 2002). GPN3 was shown to interact with the oligomerization domain of Apaf-1, which was defective to mediate apoptosome formation and apoptosis induction upon oncogenic stimuli when GPN3 was silenced (Sanchez-Olea *et al.*, 2008). Some cancer cell lines have developed mechanisms to overcome GPN3-depletion defects such as cytoplasmic accumulation of Rpb1, downregulation of transcription, cell cycle arrest and impaired proliferation (Sanchez-Olea*, et al.*, 2008; Calera *et al.*, 2011). GPN-loop GTPases were implicated in Pol II biogenesis because interaction with Pol II subunits and assembly intermediates as well as Pol II biogenesis factors, such as chaperones were shown (Jeronimo*, et al.*, 2007; Boulon*, et al.*, 2010; Forget*, et al.*, 2010). Additionally global downregulation of transcription was observed upon silencing of yeast Npa3 and GPN3 (Jeronimo*, et al.*, 2007). Indeed, mutation in the nucleotide-binding site or GPN motif in human GPN1 leads to cytoplasmic accumulation or decreased

nuclear levels of Rpb1 (Forget, *et al.*, 2010; Carre & Shiekhattar, 2011). This is consistent with Pol II assembly and/or import defects because Pol II is assembled from its subunits in the cytoplasm (Boulon, *et al.*, 2010) prior to its nuclear import (Czeko, *et al.*, 2011). Depletion of human GPN3 (Calera, *et al.*, 2011) or mutation of yeast GPN2 or GPN3 (Minaker, *et al.*, 2013) also leads to cytoplasmic accumulation of Rpb1, indicating a general role of all three GPN-loop GTPases in Pol II biogenesis. Depletion of the GPN1 homolog Npa3 from *S. cerevisiae* leads to cytoplasmic accumulation of Rpb1 and Rpb3 (Staresincic, *et al.*, 2011). Rpb1 accumulation is also observed when Npa3 is mutated in its nucleotide-binding site or GPN motif (Forget, *et al.*, 2010; Staresincic, *et al.*, 2011). Association of yeast Npa3 with Rpb1 is regulated by GTP binding in whole-cell extracts (Staresincic, *et al.*, 2011) and a direct interaction of human GPN1 and GPN3 with recombinant Pol II subunits Rpb4 and Rpb7 and the C-terminal repeat domain (CTD) of Rpb1 has been reported (Carre & Shiekhattar, 2011). Recent studies suggest a role in Pol II assembly rather than import because GPN-loop GTPases lack a nuclear localization signal (NLS), and mutations of GPN2 or GPN3 cannot be rescued by fusion of a NLS to Rpb3, whereas deletion of the import factor Iwr1 is partially rescued (Minaker, *et al.*, 2013). Further GPN1 interacts with the CCT complex (Forget, *et al.*, 2010), a chaperone complex with various functions (Leroux & Hartl, 2000) that interacts with Pol II subunits (Dekker *et al.*, 2008), consistent with a role in cytoplasmic assembly. In patients with myofibrillar myopathies, a neuromuscular disorder characterized by protein aggregates, human GPN1 shows increased expression and accumulates with Rpb1 in the cytoplasm of muscle cells (Guglielmi, *et al.*, 2015).

Whether GPN-loop GTPases are involved in cytoplasmic assembly and/or nuclear import of Pol II is still discussed controversially. The lack of structural data for any eukaryotic GPN-loop GTPase so far prevented detailed molecular understanding of these essential proteins.

1.4. Molecular chaperones

The pioneering experiments of Anfinsen (Anfinsen, 1973) and Caspar and Klug (Caspar & Klug, 1962) state that all information for protein folding and correct assembly with other proteins or nucleic acids are stored within the primary structure of those chains. This was in contrast to findings that the mixture of *Xenopus* histones and DNA in physiological salt concentrations lead to insoluble aggregates rather than spontaneously assembled nucleosomes (Laskey *et al.*, 1978). Experiments showed, that the addition of *Xenopus* egg homogenate prevents aggregation and allows nucleosome assembly. The responsible negatively charged factor, nucleoplasmin, was identified to bind folded histones, thus reducing their basic charge, preventing non-specific aggregation with negatively charged DNA and permitting their inherent self-assembly properties (Laskey, *et al.*, 1978). This principle was then extended 1987 (Ellis, 1987) and lead to the current definition of molecular chaperones which describes any protein that interacts with and aids in the folding and assembly of another macromolecular

structure without being part of its final structure (Kim *et al.*, 2013). The large and highly diverse group of molecular chaperones can be subdivided in various classes dependent on their time of action, mode of action or location of action.

1.4.1. Ribosome-binding chaperones

Emerging polypeptide chains often need to be protected from engaging unfavorable intra- and intermolecular interactions that may cause misfolding and aggregation (Kim*, et al.*, 2013). This is carried out by ribosome-associated chaperones that usually shield exposed hydrophobic regions during translation. In bacteria, the trigger factor (TF) binds at the ribosomal exit tunnel (Kramer *et al.*, 2002; Ferbitz *et al.*, 2004; Merz *et al.*, 2008), interacts with most newly synthesized cytosolic and some secretory proteins (Bukau *et al.*, 2000; Oh *et al.*, 2011; Calloni *et al.*, 2012; Preissler & Deuerling, 2012), binds nascent chains after 60-100 residues (Bukau*, et al.*, 2000; Kaiser *et al.*, 2006; Oh*, et al.*, 2011) and thereby permits prior nascent chain interaction of ribosome-binding targeting factors like the signal recognition particle (Eisner *et al.*, 2003; Ullers *et al.*, 2003) and modifying enzymes like the peptide deformylase (Bingel-Erlenmeyer *et al.*, 2008). ATP-independent TF-release from the nascent chain allows folding or downstream transfer of the polypeptide chain to the Hsp70 chaperone DnaK that likely exhibits functional redundancy with TF (Bukau*, et al.*, 2000; Genevaux *et al.*, 2004; Calloni*, et al.*, 2012). In eukaryotes, a similar role is suggested for the ribosome-associated complex (RAC) and the nascent-chain-associated complex (NAC). RAC comprises the specialized Hsp70-like protein Ssz1 and the co-chaperone zuotin (Hsp40) and cooperates with the ribosome-binding Hsp70 isoform Ssb in fungi to assist nascent chain folding (Bukau*, et al.*, 2000; Gautschi *et al.*, 2002; Raue *et al.*, 2007; Peisker *et al.*, 2008; Koplin *et al.*, 2010; Preissler & Deuerling, 2012). NAC interacts with the ribosome and short nascent chains and exhibits partially redundant function with Ssb in yeast, but its precise role in protein folding and quality control remains elusive (Wegrzyn *et al.*, 2006; Koplin*, et al.*, 2010; Pech *et al.*, 2010; del Alamo *et al.*, 2011; Preissler & Deuerling, 2012).

1.4.2. Chaperones acting downstream of the ribosome

A large number of molecular chaperones function downstream of the ribosome to maintain protein folding, proteome maintenance, macromolecular complex assembly, protein transport, degradation, aggregate dissociation and refolding of stress-denatured proteins (Kim*, et al.*, 2013). In the following section molecular mechanisms of chaperones are explained, including GPN1- and Pol II-interacting members.

Hsp70 chaperones are ATPases that interact with a myriad of nascent and newly synthesized polypeptides without having direct affinity for the ribosome (Calloni*, et al.*, 2012; Niwa *et al.*, 2012). Hsp70 consists of a N-terminal nucleotide binding domain (NBD) that is connected to the C-terminal substrate binding domain (SBD) via a conserved, flexible linker (Bukau & Horwich, 1998; Bertelsen *et al.*,

2009; Mapa *et al.*, 2010; Zuiderweg *et al.*, 2013). ATP binding to the NBD allosterically alters the SBD that opens the peptide binding pocket (Zhuravleva & Gierasch, 2011; Kityk *et al.*, 2012). In the open, ATP-bound state, the SBD binds extended, hydrophobic 5-7-residue long peptides, typically flanked by positively charged amino acids (Bukau & Horwich, 1998). The open, ATP-bound state has high on and off rates for the substrate peptides whereas the rates are low in the closed, ADP-bound state where the peptide is trapped in the pocket (Bukau & Horwich, 1998; Mayer, 2010). In turn, substrate binding allosterically stimulates ATP hydrolysis (Swain *et al.*, 2007; Smock *et al.*, 2010; Zhuravleva & Gierasch, 2011) and stable peptide binding (Bertelsen*, et al.*, 2009; Mapa*, et al.*, 2010). The ATPase cycle of Hsp70 is highly regulated by Hsp40 (J proteins) co-chaperones and nucleotide exchange factors (NEFs) (Hartl & Hayer-Hartl, 2009; Mayer, 2010). Besides its role in protecting nascent chains against aberrant interactions, the Hsp70-Hsp40 system also functions in ATP-dependent co- and posttranslational folding and downstream transfer of polypeptides to other chaperones like the Hsp90 and the chaperonin system (Kim*, et al.*, 2013). Archaea, lacking the Hsp70 system may substitute its function with the ATP-independent prefoldin (PFD, also known as Gim complex), a hexameric α/β complex of 14-23 kDa subunits that binds certain nascent chains, mediates transfer to chaperonins and assists in chaperonin mediated folding (Frydman, 2001; Hartl & Hayer-Hartl, 2002).

Chaperonins (also called Hsp60s) are large double ring complexes of 800-1000 kDa comprising 7-9 60 kDa subunits per ring and form a central cavity where a single client protein is encapsulated for ATP-dependent folding (Hartl, 1996; Bukau & Horwich, 1998; Hartl *et al.*, 2011). They are structurally subdivided into two groups. Group I requires lid-shaped co-chaperones for client protein encapsulation. Group II, including the eukaryotic CCT (also TRiC) contain built-in lids (Horwich *et al.*, 2007; Tang *et al.*, 2007). CCT interacts with nascent chains, assists in posttranslational folding in cooperation with Hsp70 (Etchells *et al.*, 2005; Cuellar *et al.*, 2008), interacts with 5-10% of eukaryotic proteins (Yam *et al.*, 2008), including GPN1 (Forget*, et al.*, 2010) and Pol II subunits (Dekker*, et al.*, 2008) and uses an iris-like closing mechanism enabling encapsulation of large multidomain proteins that don't fit into the cavity entirely (Russmann *et al.*, 2012).

Many eukaryotic proteins are delivered to the Hsp90 system by Hsp70 for ATP-dependent completion of folding and conformational regulation (McClellan *et al.*, 2007; Zhao & Houry, 2007; Taipale *et al.*, 2010). The transfer is mediated by the Hsp90 organizing protein (HOP) that bridges between both key chaperones (Young *et al.*, 2001). Hsp90 is involved in many cellular process, including cell cycle progression, steroid and calcium signaling, protein complex assembly, immune and heat shock response (Young *et al.*, 2003; McClellan*, et al.*, 2007; Taipale*, et al.*, 2010; Makhnevych & Houry, 2012). The chaperone consists of three domains, an N-terminal ATPase domain, essential for function (Young*, et al.*, 2001; Pearl & Prodromou, 2006) with regulatory properties (Hainzl *et al.*, 2009; Tsutsumi *et al.*, 2009), a middle domain required for substrate protein interaction and ATP hydrolysis regulation (Meyer

et al., 2003; Koulov *et al.*, 2010; Retzlaff *et al.*, 2010), and a C-terminal domain for dimerization and co-chaperone interaction (Young*, et al.*, 2001; Harris *et al.*, 2004). The structures of homodimeric Hsp90 from different organisms has been studied extensively (Ali *et al.*, 2006; Shiau *et al.*, 2006; Dollins *et al.*, 2007) and many complex structures with various co-chaperones and inhibitors are available (Young*, et al.*, 2001; Harris*, et al.*, 2004; Roe *et al.*, 2004; Ali*, et al.*, 2006; Pearl & Prodromou, 2006; Vaughan *et al.*, 2006; Southworth & Agard, 2011). The structures together with biochemical data reveal extensive rearrangements (Mayer, 2010; Li *et al.*, 2012) and high flexibility (Krukenberg *et al.*, 2008) consistent with the diversity of its client proteins (Taipale*, et al.*, 2010). The apo state adopts an open V-shaped conformation that closes upon binding of ATP and inactive client proteins. In the next step the N-terminal domains dimerize and the two subunits twist around each other to form the so-called molecular clamp state (Ali*, et al.*, 2006; Southworth & Agard, 2011; Li*, et al.*, 2012). Hydrolysis of ATP to ADP drives substrate activation to completion and leads to N-termini separation, release of the folded substrate, ADP and inorganic phosphate and the closing of the cycle by converting into the open apo state (Kim*, et al.*, 2013). The reaction cycle is regulated by posttranslational modifications and various co-chaperones that accelerate (Aha1) or slow down (HOP, Cdc37 and p23) the cycle at different steps (Kim*, et al.*, 2013). Further, stimulation of ATPase activity by client protein binding was reported (McLaughlin *et al.*, 2002). Additionally to protein folding Hsp90 plays an important role in assembly of various macromolecular complexes, including RNA polymerase II (Makhnevych & Houry, 2012).

1.4.3. Assembly chaperones

Many proteins assemble to macromolecular complexes to carry out their biological function. However, multiprotein complex assembly not only requires universal folding chaperones and rather unspecific 'holdases' like small heat shock proteins (sHsps) that shield hydrophobic regions of a wide range of client proteins to avoid aggregation. Specific assembly chaperones are required to orchestrate the assembly of individual components in a highly ordered manner (Ellis, 2013). The distinction between folding and assembly chaperones is not always absolute because oligomerization often involves conformational rearrangements. A rising number of assembly chaperones has been identified for large molecular machines including the nucleosome (nucleoplasmin, Asf1, CAF-1 etc.) (De Koning *et al.*, 2007; Avvakumov *et al.*, 2011), Rubisco (RbcX, Raf1) (Saschenbrecker *et al.*, 2007; C. Liu *et al.*, 2010; Hauser *et al.*, 2015), the proteasome (PAC1/2/3, hUmp1) (Murata *et al.*, 2009), spliceosomal snRNPs (plCln) (Chari *et al.*, 2008), ATP synthase (PAB) (Mao *et al.*, 2015) and the ribosome (RAC, NAC, Jjj1) (Karbstein, 2010). Assembly chaperones hereby transiently mask surfaces that form interfaces in the

assembled complex thus preventing aggregation and misassembly and opening a time window for association with the native interaction partner (Ellis, 2013). These subunit interfaces often consist of hydrophobic residues and thus are especially prone for aggregation during complex assembly.

1.5. Aims and scope of this study

Biogenesis of RNA polymerase II is not very well understood and requires the help of several factors. Here, the essential, recently discovered GPN-loop GTPases were shown to play a critical role. However, no structural information for any eukaryotic GPN-loop GTPase was available and the cellular function of these enzymes remained largely unclear. In particular, it was not known whether GPN-loop GTPases are involved in assembly or nuclear import of Pol II or both and what precise function they carry out. Further, it was unclear whether these proteins interact with complete assembled Pol II or only with assembly intermediates or single subunits. Additionally, the molecular function of the GPN-loop was only suggested from structural studies of an archaeal GPN-loop GTPase but not confirmed biochemically and remained speculative in the eukaryotic system.

To gain insights into the structure and molecular mechanisms of GPN-loop GTPases we aimed to solve high resolution crystal structures of the yeast GPN1 enzyme Npa3 in complex with various nucleotides to gain snapshots along the hydrolysis pathway. With this we intended to understand how Npa3 binds and hydrolyzes GTP and elucidate the molecular function of its GPN-loop to propose a model for enzymatic Npa3 activity. Further we aimed to confirm this model biochemically by using a combination of structure-guided site-directed mutagenesis and enzymatic activity assays. To unravel the function of Npa3 we aimed to apply biochemical approaches and characterize Pol II interaction sites which might hint to its role in Pol II biogenesis.

2. Materials and Methods

2.1. Materials

2.1.1. Bacterial strains

Table 4| *Escherichia Coli* strains

Strain	Base strain	Genotype	Source
XL1 blue	K12	endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K^- m_K^+)	Stratagene
BL21 Gold RIL	DE3	B; F-; ompT; hsdS(rB- rB-); dcm+; Tetr; gal_(DE3); endA; Hte [argU, ileY, leuW, Camr]	Stratagene
Rosetta	DE3	F$^-$ ompT hsdS$_B$(R_B^- m_B^-) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cam^R)	Novagen

2.1.2. Yeast strains

Table 5| *Saccharomyces cerevisiae* strains

Strain	Genotype	Source
Npa3-C-TAP	S288C; ATCC 201388: MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Thermo Scientific Open Biosystems
Pol II purification strain	MATa or α; ura3-52 trp1Δ leu2Δ1 his3Δ200 pep4::HIS3 prb1Δ1.6R can1Δ GAL rpb3::URA3-N-6xHis-RPB3	Kashlev Lab

2.1.3. Plasmids

Table 6| Plasmids used in this study

Vector ID	Insert/Description	Type	Tag	Primer
B1I	Sc Npa3 1-385 (fl)	pOPINI	N-His	Npa3_Sc_pI_M1_F/Npa3_Sc_pI_N385_R
B2I	Sc Npa3 1-298	pOPINI	N-His	Npa3_Sc_pI_M1_F/Npa3_Sc_pI_K298_R
B3I	Sc Npa3 1-264	pOPINI	N-His	Npa3_Sc_pI_M1_F/Npa3_Sc_pI_K264_R
B4I	Sc Npa3 1-385Δ203-211	pOPINI	N-His	Npa3_Sc_pI_M1_F/Npa3_Sc_Δ203-211_R Npa3_Sc_Δ203-211_F/Npa3_Sc_pI_N385_R
B5I	Sc Npa3 1-298Δ203-211	pOPINI	N-His	Npa3_Sc_pI_M1_F/Npa3_Sc_Δ203-211_R Npa3_Sc_Δ203-211_F/Npa3_Sc_pI_K298_R
B1E	Sc Npa3 1-385 (fl)	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_pE_N385_R
B2E	Sc Npa3 1-298	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_pE_K298_R

B3E	Sc Npa3 1-264	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_pE_K264_R
B4E	Sc Npa3 1-385Δ203-211	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_Δ203-211_R Npa3_Sc_Δ203-211_F/Npa3_Sc_pE_N385_R
B6E	Sc Npa3 1-264Δ203-211	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_Δ203-211_R Npa3_Sc_Δ203-211_F/Npa3_Sc_pE_K264_R
B1E_GPN-AAA	Sc Npa3 1-385_GPN-AAA	pOPINE	C-His	Npa3_Sc_pE_M1_F/ Npa3_Sc_GPN-AAA_R Npa3_Sc_GPN-AAA_F /Npa3_Sc_pE_N385_R
B1E_D40A	Sc Npa3 1-385_D40A	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_D40A_R Npa3_Sc_D40A_F/Npa3_Sc_pE_N385_R
B1E_D106A	Sc Npa3 1-385_D106A	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_D106A_R Npa3_Sc_D106A_F/Npa3_Sc_pE_N385_R
B1E_Q110L	Sc Npa3 1-385_Q110L	pOPINE	C-His	Npa3_Sc_pE_M1_F/Npa3_Sc_Q110L_R Npa3_Sc_Q110L_F/Npa3_Sc_pE_N385_R
B6E_D106A	Sc Npa3 1-264Δ203-211 D106A	pOPINE	C-His	B6E w. Npa3_Sc_pE_M1_F/Npa3_Sc_D106A_R Npa3_Sc_D106A_F/ Npa3_Sc_pE_K264_R
Npa3_265-385	Sc Npa3 265-385	pOPINE	C-His	Npa3_Sc_Q265_pE_F /Npa3_Sc_pE_N385_R
B1Enotag	Sc Npa3 1-385 (fl)	pOPINE	-	Npa3_Sc_pE_M1_F/Npa3_Sc_pE_N385no-tag_R
C1I	Sp Npa3 1-367 (fl)	pOPINI	N-His	Npa3_POM_pI_M1_F/Npa3_POM_pI_Q367_R
C3I	Sp Npa3 1-273	pOPINI	N-His	Npa3_POM_pI_M1_F/Npa3_POM_pI_R273_R
C5I	Sp Npa3 1-301Δ208-215	pOPINI	N-His	Npa3_POM_pI_M1_F/Npa3_POM_Δ208-215_R Npa3_POM_Δ208-215_R/Npa3_POM_pI_K301_R
C1E	Sp Npa3 1-367	pOPINE	C-His	Npa3_POM_pE_M1_F/Npa3_POM_pE_Q367_R
C2E	Sp Npa3 1-301	pOPINE	C-His	Npa3_POM_pE_M1_F/Npa3_POM_pE_K301_R
C3E	Sp Npa3 1-273	pOPINE	C-His	Npa3_POM_pE_M1_F/Npa3_POM_pE_R273_R
A1I	Hs GPN1 1-374 (fl)	pOPINI	N-His	GPN1_pI_M1_F/GPN1_pI_K374_R
A2I	Hs GPN1 1-304	pOPINI	N-His	GPN1_pI_M1_F/GPN1_pI_L304_R
A3I	Hs GPN1 1-275	pOPINI	N-His	GPN1_pI_M1_F/GPN1_pI_R275_R
A5I	Hs GPN1 18-304	pOPINI	N-His	GPN1_pI_H18_F/GPN1_pI_L304_R
A6I	Hs GPN1 18-275	pOPINI	N-His	GPN1_pI_H18_F/GPN1_pI_R275_R
A1E	Hs GPN1 1-374 (fl)	pOPINE	C-His	GPN1_pE_M1_F/GPN1_pE_K374_R
A2E	Hs GPN1 1-304	pOPINE	C-His	GPN1_pE_M1_F/GPN1_pE_L304_R
A3E	Hs GPN1 1-275	pOPINE	C-His	GPN1_pE_M1_F/GPN1_pE_R275_R
A4E	Hs GPN1 18-374	pOPINE	C-His	GPN1_pE_H18_F/GPN1_pE_R275_R
A5E	Hs GPN1 18-304	pOPINE	C-His	GPN1_pE_H18_F/GPN1_pE_L304_R
A6E	Hs GPN1 18-275	pOPINE	C-His	GPN1_pE_H18_F/GPN1_pE_R275_R
D1I	Sc GPN3 1-272 (fl)	pOPINI	N-His	YLR243W_pI_M1_F/ YLR243W_pI_E272_R
D1E	Sc GPN3 1-272 (fl)	pOPINE	C-His	YLR243W_pE_M1_F/ YLR243W_pE_E272_R

E1I	Sp GPN3 1-276 (fl)	pOPINI	N-His	Fet5_pI_F/Fet5_pI_R
E1E	Sp GPN3 1-276 (fl)	pOPINE	C-His	Fet5_pE_F/Fet5_pE_R
E1A_no-tag	Sp GPN3 1-276 (fl)	pOPINA	-	Fet5_pA_noHis_F/ Fet5_pA_noHis_R
F1A	Sc GPN2 1-347 (fl)	pOPINA	C-His	GPN2_Sc_M1_pA_F/GPN2_Sc_R347_pA_R
F2A	Sc GPN2 1-264	pOPINA	C-His	GPN2_Sc_M1_pA_F/GPN2Sc_G264_pA_R
F3A	Sc GPN2_1-243	pOPINA	C-His	GPN2_Sc_M1_pA_F/GPN2Sc_D243_pA_F
F4A	Sc GPN2 1-347Δ206-211	pOPINA	C-His	GPN2_Sc_M1_pA_F/GPN2ScΔI206-S211_pA_R GPN2ScΔI206-S211_pA_F/GPN2_Sc_R347_pA_R
F5A	Sc GPN2 1-264Δ206-211	pOPINA	C-His	GPN2_Sc_M1_pA_F/GPN2ScΔI206-S211_pA_R GPN2ScΔI206-S211_pA_F/GPN2Sc_G264_pA_R
F6A	Sc GPN2_1-243Δ206-211	pOPINA	C-His	GPN2_Sc_M1_pA_F/GPN2ScΔI206-S211_pA_R GPN2ScΔI206-S211_pA_F/GPN2Sc_D243_pA_F
His-N-Iwr1	Sc Iwr1 1-353 (fl)	pET21b	N-His	Cloned by E. Czeko
Iwr1-C-His	Sc Iwr1 1-353 (fl)	pET21b	C-His	Cloned by E. Czeko
N-Cys-Iwr1_9-223	Sc Cys-Iwr1 9-223	pOPINB	N-His-	Iwr1_Sc_pB_Cys-A9_F/Iwr1_Sc_pB_D223_R (N-term. Cys for site-directed fluorescence-labeling)

2.1.4. Oligonucleotides

Table 7| Oligonucleotides used for molecular cloning

Primer name	Sequence (5' to 3')	Target vector
Npa3_Sc_pI_M1_F	ACCATCACAGCAGCGGCAGTCTCAGCACAATCATATG	pOPINI
Npa3_Sc_pI_N385_R	ATGGTCTAGAAAGCTTTAGTTTCTAATATACTTCGCGATATTTTC	pOPINI
Npa3_Sc_pI_K298_R	ATGGTCTAGAAAGCTTTACTTCTCGTTTAACCCTAGATCC	pOPINI
Npa3_Sc_pI_K264_R	ATGGTCTAGAAAGCTTTACTTATAGTATTGGTCGTATTCATCAAC	pOPINI
Npa3_Sc_pE_M1_F	AGGAGATATACCATGAGTCTCAGCACAATCATATG	pOPINE
Npa3_Sc_pE_N385_R	GTGATGGTGATGTTTGTTTCTAATATACTTCGCGATATTTTC	pOPINE
Npa3_Sc_pE_K298_R	GTGATGGTGATGTTTCTTCTCGTTTAACCCTAGATCC	pOPINE
Npa3_Sc_pE_K264_R	GTGATGGTGATGTTTCTTATAGTATTGGTCGTATTCATCAAC	pOPINE
Npa3_Sc_pE_N385no-tag_R	GTGATGGTGATGTTTTTAGTTTCTAATATACTTCGCGATATTTTC	pOPINE
Npa3_Sc_Q265_pE_F	AGGAGATATACCATGCAAGAACGTGAAAAAGCATTG	pOPINE
Npa3_POM_pI_M1_F	ACCATCACAGCAGCGGCATGACAGATAAAGAGAAGAAGCC	pOPINI

Npa3_POM_pI_Q367_R	ATGGTCTAGAAAGCTTTATTGCTTCATACGCTCTGTAAG	pOPINI
Npa3_POM_pI_K301_R	ATGGTCTAGAAAGCTTTATTTAGAAACATGCATGTCTTTC	
Npa3_POM_pI_R273_R	ATGGTCTAGAAAGCTTTATCGTTCCATTTCTGGAAC	pOPINI
Npa3_POM_pE_M1_F	AGGAGATATACCATGACAGATAAAGAGAAGAAGCC	pOPINE
Npa3_POM_pE_Q367_R	GTGATGGTGATGTTTTTGCTTCATACGCTCTGTAAG	pOPINE
Npa3_POM_pE_K301_R	GTGATGGTGATGTTTTTTAGAAACATGCATGTCTTTC	pOPINE
Npa3_POM_pE_R273_R	GTGATGGTGATGTTTTCGTTCCATTTCTGGAAC	pOPINE
GPN1_pI_M1_F	ACCATCACAGCAGCGGCATGGCGGCGTCCG	pOPINI
GPN1_pI_H18_F	ACCATCACAGCAGCGGCCACCCAGTGTGTCTGTTGG	pOPINI
GPN1_pI_K374_R	ATGGTCTAGAAAGCTTTACTATTTATTGTTTCTCTTCCAGTATTGTG	pOPINI
GPN1_pI_L304_R	ATGGTCTAGAAAGCTTTACAAGGCTACAGAACCCATATCTTTTC	pOPINI
GPN1_pI_R275_R	ATGGTCTAGAAAGCTTTAACGTTCATATTCAGGACGATACTC	pOPINI
GPN1_pE_M1_F	AGGAGATATACCATGGCGGCGTCCG	pOPINE
GPN1_pE_H18_F	AGGAGATATACCATGCACCCAGTGTGTCTGTTGG	pOPINE
GPN1_pE_K374_R	GTGATGGTGATGTTTTTTATTGTTTCTCTTCCAGTATTGTG	pOPINE
GPN1_pE_L304_R	GTGATGGTGATGTTTCAAGGCTACAGAACCCATATCTTTTC	pOPINE
GPN1_pE_R275_R	GTGATGGTGATGTTTACGTTCATATTCAGGACGATACTC	pOPINE
GPN2_Sc_M1_pA_F	AGGAGATATACCATGCCCTTCGCTCAGAT	pOPINA
GPN2_Sc_R347_pA_R	GTGGTGGTGGTGTTTCCTAACAAAATCCATTCCTTG	pOPINA
GPN2_Sc_G264_pA_R	GTGGTGGTGGTGTTTGCCGAATATGTAGCCATTTG	pOPINA
GPN2Sc_D243_pA_R	GTGGTGGTGGTGTTTATCCACGGACAAAACCTC	pOPINA
YLR243W_pI_F	ACCATCACAGCAGCGGCTCTCGCGTTGGTGTC	pOPINI
YLR243W_pI_R	ATGGTCTAGAAAGCTTTATTCTTCGACATCTATTTGGTCG	pOPINI
YLR243W_pE_F	AGGAGATATACCATGTCTCGCGTTGGTGTC	pOPINE
YLR243W_pE_R	GTGATGGTGATGTTTTTCTTCGACATCTATTTGGTCG	pOPINE
Fet5_pI_F	ACCATCACAGCAGCGGGTTAAGGTGGCAGCTTTTGTTTG	pOPINI
Fet5_pI_R	ATGGTCTAGAAAGCTTTATTCATCGTCTTCTAAATCATCTGC	pOPINI
Fet5_pE_F	AGGAGATATACCATGGTTAAGGTGGCAGCTTT	pOPINE
Fet5_pE_R	GTGATGGTGATGTTTTTCATCGTCTTCTAAATCATCTGC	pOPINE
Fet5_pA_noHis_F	AGGAGATATACCATGGTTAAGGTGGCAGCTTTTGTTTG	pOPINA
Fet5_pA_noHis_R	GTGGTGGTGGTGTTTCTATTCATCGTCTTCTAAATCATCTGC	pOPINA

Mutagenic Primer

Npa3_Sc_Δ203-211_F	CAAGGAAGATCAAGACGGGTACATGAGCTCATTG	independent
Npa3_Sc_Δ203-211_R	CAATGAGCTCATGTACCCGTCTTGATCTTCCTTG	independent
Npa3_Sc_GPN-AAA_F	GAGAATTACCAGCTAGCTGCGGCCGGTGCCATTGTCACC	independent

Npa3_Sc_GPN-AAA_R	GGTGACAATGGCACCGGCCGCAGCTAGCTGGTAATTCTC	independent
Npa3_Sc_D40A_F	CCATACGTAATCAATCTTGCTCCTGCAGTATTGAGAGTCC	independent
Npa3_Sc_D40A_R	GGACTCTCAATACTGCAGGAGCAAGATTGATTACGTATGG	independent
Npa3_Sc_D106A_F	CAAAACTGCATCATCGCCACTCCAGGCC	independent
Npa3_Sc_D106A_R	GGCCTGGAGTGGCGATGATGCAGTTTTG	independent
Npa3_Sc_Q110L_F	CATCGACACTCCAGGCTTAATCGAATGTTTTGTGTG	independent
Npa3_Sc_Q110L_R	CACACAAAACATTCGATTAAGCCTGGAGTGTCGAT	independent
Npa3_POM_Δ208-215_F	GACTAAAGACGAAGGTGGATATATGGGTTCG	independent
Npa3_POM_Δ208-215_R	CGAACCCATATATCCACCTTCGTCTTTAGTC	independent
GPN2Sc_d_I206-S211_pA_F	CAGGATCTGGATTATTTGGAGCCATATAGTGTACTGGGAAAGAAATATAGCAAG	independent
GPN2Sc_d_I206-S211_pA_R	CTTGCTATATTTCTTTCCCAGTACACTATATGGCTCCAAATAATCCAGATCCTG	independent
Sequencing Primer		
T7f	TAATACGACTCACTATAGGG	All pOPIN
pET-RP	CTAGTTATTGCTCAGCGG	pET28-based
pTriExDown	TCGATCTCAGTGGTATTTGTG	pTriEx-based

2.1.5. Reagents and consumables

Chemicals were obtained from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany) or Sigma-Aldrich (Seelze, Germany) unless stated otherwise. Enzymes and reagents for cloning were obtained from Fermentas (St. Leonrot, Germany), New England Biolabs (Frankfurt am Main, Germany) and Clontech (St. Germain-en-Leye, France). For DNA preparation commercial kits from Qiagen (Hilden, Germany) were used. DNA and RNA oligonucleotides were ordered at ThermoScientific (Ulm, Germany) and Metabion (Planegg, Germany), respectively. Crystallization reagents and tools were ordered at Hampton Research (Aliso Viejo, CA, USA). GTP-derivatives were obtained from Sigma-Aldrich (Seelze, Germany) or Jena Bioscience (Jena, Germany).

2.1.6. Media and additives

Media were usually taken from lab stocks (Table 8). Media additives (Table 9) were sterile filtered.

Table 8| Media for *Escherichia coli* and *Saccharomyces cerevisiae*

Media	Application	Description
LB	*E. coli* culture	1% (w/v) Bacto tryptone, 0.5% (w/v) yeast extract, 8.6 mM NaCl, 2.6 mM NaOH, plates contained 1.5% (w/v) agar
LB X-Gal plates	*E. coli* transformation	1 % (w/v) Bacto tryptone, 0.5% (w/v) yeast extract, 8.6 mM NaCl, 2.6 mM NaOH, 1.5 % (w/v) agar, 0.02% (w/v) 5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid (X-Gal)[1] [2], 1 mM IPTG[2], 0.1% (w/v) ampicillin/kanamycin[2]
YPD	Yeast culture	2% (w/v) peptone, 2% (w/) glucose, 1% (w/v) yeast extract
YPD plates	Yeast plate	YPD, 2% (w/v) agar

[1] stock solution dissolved in DMF; [2]added after autoclave

Table 9| Media additives for *Escherichia coli* and *Saccharomyces cerevisiae*

Additive	Description	Stock solution	Applied concentration
IPTG	*E. coli* induction	1 M in H_2O	0.5 mM
Ampicillin	Antibiotic	100 mg/ml in H_2O	100 µg/ml
Kanamycin	Antibiotic	30 mg/ml in H_2O	30 µg/ml
Chloramphenicol	Antibiotic	50 mg/ml in EtOH	50 µg/ml
Tetracycline	Antibiotic	12.5 mg/ml in EtOH	12.5 µg/ml
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal)	*E. coli* blue/white screening	20% (w/v) in DMF	0.02 (w/v)

2.1.7. Buffers and solutions

Table 10| General buffers, dyes and solutions

Name	Description	Application
10X TAE	50 mM EDTA, pH 8.0, 2.5M Tris-acetate	Agarose gels
6X DNA loading dye	Fermentas	Agarose gels
5X Phusion HF buffer	Finnzymes	PCR
iQ™ SYBR® Green Supermix	BioRad	DNA staining
10X buffer 0	Fermentas	Restriction Digest
10X FastDigest buffer	Fermentas	Restriction Digest
20X MES buffer	NuPAGE® MES SDS Running buffer, Life technologies	SDS-PAGE
20X MOPS buffer	NuPAGE® MOPS SDS Running buffer, Life technologies	SDS-PAGE
5X SDS loading buffer	250 mM Tris-HCl pH 8.0, 50% (v/v) glycerol, 0.5% (w/v) bromophenol blue, 7.5% (w/v) SDS, 500 mM DTT	SDS-PAGE
Protein Ladder	PageRuler™ Prestained Protein Ladder (Fermentas)	SDS-PAGE
Gel staining solution	50% (v/v) ethanol, 7% (v/v) acetic acid, 0.125% (w/v) Coomassie brilliant blue R-250	Coomassie staining
Gel staining solution	Instant blue, Expedeon	Coomassie staining

100X PI	1.42 mg leupeptin, 6.85 mg pepstatine A, 850 mg PMSF, 1650 mg benzamidine dissolved in 50 ml ethanol	Protease inhibitor mix
TFB-1	30 mM KOAc, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$[1], 15% [v/v] glycerol, pH 5.8 (with 0.2 M acetic acid)	Chemically competent *E. coli*
TFB-2	10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl, 15% [v/v] Glycerol, pH 6.5 (with KOH)	Chemically competent *E. coli*
1x Bradford dye	1:5 dilution of Bradford concentrate (BioRad)	Determination of protein concentration

[1] added after pH adjustment

Table 11| Npa3 and GPN2 purification buffers

Name	Description
Lysis buffer	50 mM Tris pH 7.5 (6°C), 5 mM Imidazole, 300 mM NaCl, 5 mM $MgCl_2$, 0.2% (v/v) Tween-20, 2 mM DTT[1], 1X PI[1]
Nickel wash buffer	50 mM Tris pH 7.5 (6°C), 10 mM imidazole, 300 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT[1]
Nickel elution buffer	50 mM Tris pH 7.5 (6°C), 200 mM imidazole, 300 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT[1]
Anion exchange buffer A	50 mM Tris pH 7.5 (6°C), 100 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT[1]
Anion exchange buffer B	50 mM Tris pH 7.5 (6°C), 1 M NaCl, 5 mM $MgCl_2$, 2 mM DTT[1]
SEC buffer 100	10 mM HEPES pH 7.5 (4°C), 100 mM NaCl, 5 mM $MgCl_2$, 10 mM DTT[1]
SEC buffer 200	10 mM HEPES pH 7.5 (4°C), 200 mM NaCl, 5 mM $MgCl_2$, 10 mM DTT[1]

[1] added directly before usage

Table 12| Iwr1 purification buffers

Name	Description
Iwr1 Lysis buffer	20 mM Tris pH 8.0 (4°C), 150 mM NaCl, 5 mM DTT[1], 1X PI[1]
Iwr1 Nickel wash buffer	20 mM Tris pH 8.0 (4°C), 150 mM NaCl, 10/20/30/40 mM imidazole, 5 mM DTT[1]
Iwr1 Nickel elution buffer	20 mM Tris pH 8.0 (4°C), 150 mM NaCl, 250 mM imidazole, 5 mM DTT[1]
Iwr1 Anion exchange buffer A	20 mM Tris pH 8.0 (4°C), 100 mM NaCl, 5 mM DTT[1]
Iwr1 Anion exchange buffer B	20 mM Tris pH 8.0 (4°C), 1 M NaCl, 5 mM DTT[1]
Pol II buffer	5 mM HEPES pH 7.25 (20°C), 40 mM $(NH_4)_2SO_4$, 10 µM $ZnCl_2$, 10 mM DTT[1]

[1] added directly before usage

Table 13| Pol II purification buffers

Name	Description
3X freezing buffer	150 mM Tris pH 7.9 (4°C), 3 mM EDTA, 30 µM $ZnCl_2$, 30 % (v/v) glycerol, 3 % (v/v) DMSO, 30 mM DTT[1], 3X PI[1]
HSB150	50 mM Tris pH 7.9 (4°C), 150 mM KCl, 1 mM EDTA, 10 µM $ZnCl_2$, 10 % (v/v) glycerol, 10 mM DTT[1], 1X PI[1]
HSB1000/7	50 mM Tris pH 7.9 (4°C), 1000 mM KCl, 7 mM imidazole, 1 mM EDTA, 10 µM $ZnCl_2$, 10 % (v/v) glycerol, 2.5 mM DTT[1], 1X PI[1]
Ni buffer 7/50/100	20 mM Tris pH 7.9 (4°C), 150 mM KCl, 7/50/100 mM imidazole
MonoQ 150	20 mM Tris-acetate pH 7.9 (4°C), 150 mM KOAc, 10 % (v/v) glycerol, 0.5 mM EDTA pH 7.9, 10 µM $ZnCl_2$, 10 mM DTT[1]
MonoQ 2000	20 mM Tris-acetate pH 7.9 (4°C), 2 M KOAc, 10 % (v/v) glycerol, 0.5 mM EDTA pH 7.9, 10 µM $ZnCl_2$, 10 mM DTT[1]
Pol II buffer	5 mM HEPES pH 7.25 (20°C), 40 mM $(NH_4)_2SO_4$, 10 µM $ZnCl_2$, 10 mM DTT[1]

[1] added directly before usage

Table 14| Rpb4/7 purification buffers

Name	Description
Ni buffer 0/10/20/50/200	50 mM Tris pH 7.5 (4°C), 150 mM NaCl, 0/10/20/50/200 mM imidazole, 10 mM β-mercaptoethanol[1]
Ni salt buffer	50 mM Tris pH 7.5 (4°C), 2 M NaCl, 10 mM β-mercaptoethanol[1]
SourceQ 100	50 mM Tris pH 7.5 (4°C), 100 mM NaCl, 10 mM β-mercaptoethanol[1]
SourceQ 1000	50 mM Tris pH 7.5 (4°C), 1 M NaCl, 10 mM β-mercaptoethanol[1]
Pol II buffer	5 mM HEPES pH 7.25 (20°C), 40 mM $(NH_4)_2SO_4$, 10 µM $ZnCl_2$, 10 mM DTT[1]

[1] added directly before usage

Table 15| HPLC buffers

Name	Description
Buffer A	50 mM triethylammonium acetate
Buffer B	90 % (v/v) acetonitrile

Table 16| Chaperone assay buffers

Name	Description
TE	50 mM Tris pH8.0, 2 mM EDTA
Incubation buffer	40 mM HEPES-KOH pH7.5, 5 mM $MgCl_2$

Table 17| Biotinylation buffer

Name	Description
BirA dilution buffer	5 mM HEPES-NaOH pH 7.25, 16 mM $MgCl_2$, 10 µM $ZnCl_2$, 10 mM DTT[1]

[1] added directly before usage

2.2. General Methods

2.2.1. Molecular cloning and site-directed mutagenesis

Polymerase chain reaction (PCR) Primers for InFusion cloning (Clontech) (Berrow *et al.*, 2007) were designed using a target vector specific overhang of 15-20 nucleotides, followed by around 20 nucleotides complementary to the gene of interest. Annealing temperatures were calculated using the T_m calculator online tool (NEB). PCR reactions were carried out using the Phusion High Fidelity PCR Master Mix (NEB), 0.5 µM of each primer and variable amounts of either genomic DNA, cDNA or plasmid DNA. Thermocycling programs comprised 30 cycles and were adjusted to the specific needs of the individual reactions in terms of annealing temperature and elongation times (Biometra T3000 Thermocycler). PCR products were separated by agarose gel electrophoresis and purified using the QIAquick gel extraction kit (Qiagen).

Site-directed mutagenesis Point mutations and deletions were introduced by overlap-extension-PCR (Higuchi *et al.*, 1988). Mutagenic primers were designed with a melting temperature of approximately 78°C, containing the desired mismatches in the middle of the primer. To increase the annealing stability nucleotides G or C were used at the 5' and 3' end of the primer if possible. Two overlapping PCR products were synthesized in separate reactions using one primer containing the respective overhang for InFusion cloning (Clontech) (Berrow*, et al.*, 2007) and one primer containing the desired mutation, respectively. In a second PCR reaction 40 ng of each purified PCR product were used as template to amplify the gene of interest with the desired mutation. PCR reactions were carried out as descried above. PCR products were separated by agarose gel electrophoresis and purified using the QIAquick gel extraction kit (Qiagen).

Electrophoretic separation of DNA Electrophoretic separation of DNA was carried out in horizontal 1X TAE agarose gels containing 0.7 µg/ml Sybr Safe (Invitrogen) and 0.5-2% agarose, depending on the DNA length to be separated. Electrophoresis was carried out in PerfectBlue Gelsystem electrophoresis chambers (Peqlab). Samples were mixed with 6X loading dye (Fermentas) prior to loading and DNA was visualized and documented using a Safe Imager blue light transilluminator (Invitrogen, λ=470 nm).

Enzymatic restriction cleavage 2 µg of the desired vector were digested per 100 µl reaction prior to InFusion cloning using restriction endonucleases (NEB or Fermentas) as recommended by the respective manufacturer. Cleaved vectors were purified using the QIAquick PCR purification kit (Qiagen).

Restriction- and ligation-free InFusion cloning InFusion cloning allows restriction- and ligation-free cloning (Berrow*, et al.*, 2007). For this 100 ng of the linearized vector were incubated with 3X molar excess of the gene of interest (amplified with the respective InFusion overhangs) and the InFusion Dry-Down PCR cloning mixture (Clontech) as recommended by the manufacturer. Then 2.5 µl of the InFusion product were used for transformation of the plasmid into chemically competent *E. coli* XL1 blue cells as described in 2.2.2.

Isolation and verification of plasmid DNA After transformation of plasmids into *E. coli* cells (2.2.2.), single colonies were picked from selective plates and used for colony PCR. For this, colonies were inoculated in 30 µl H_2O and mixed at RT for 10 min in a thermomixer. Then 6 µl of the mixture were used as template for a PCR reaction carried out as described above. Colonies resulting in positive PCR reactions were inoculated in 5 mL LB medium and incubated at 37°C shaking overnight. Plasmids were isolated using the QIAprep Spin Miniprep kit (Qiagen) as recommended by the manufacturer. DNA sequences were verified by sequencing (GATC).

2.2.2. Preparation and transformation of competent *E. coli* cells

Preparation of chemically competent *E. coli* cells *E. coli* cells were grown in 5 ml LB medium overnight at 37°C and 140 rpm. The pre-culture was diluted 1:100 with LB medium containing appropriate antibiotics and cells were grown for 2-3 hours at 37°C and 140 rpm until OD_{600}=0.25 - 0.3. Then, cultures were chilled on ice for 5 min and centrifuged at 4,000 x *g* for 5 min at 4°C. After resuspending in 50 ml Transformation buffer 1 (TFB-1) (Table 10) per 250 ml culture, the cells were again incubated on ice for 5 min followed by centrifugation at 4,000 x *g* and 4°C for 5 min. The supernatant was discarded and the pellet was gently resuspended in 5 ml Transformation buffer 2 (TFB-2) (Table 10) per 250 ml culture. After another incubation on ice for 15 min the competent cells were aliquoted (50 µl), flash frozen in liquid nitrogen and stored at -80°C.

Transformation of chemically competent *E. coli* cells For transformation cells were thawed on ice and 100 ng vector or 2.5 µl InFusion product were added and incubated on ice for 5 min. The competent cells were heat shocked at 42°C for 45 s and cooled on ice for 5 min. Then 450 µl LB medium was added and the cells were grown in a thermomixer (Qiagen) at 37°C for 1 hour at 600 rpm. Finally, cells were plated on LB plates supplemented with appropriate antibiotics and incubated overnight at 37°C.

Cotransformation of electro-competent *E. coli* cells For cotransformation, electro-competent *E. coli* BL21 Gold RIL cells were thawed on ice and 100 ng of the respective vectors were added and incubated in a cuvette on ice for 5 min. Electroporation was carried out at 2.5 kV in a MicroPulser Electroporater (Biorad). Cells were then cooled on ice for 5 min and grown in 300 µl LB medium at 37°C for 1 hour at 600 rpm in a thermomixer (Qiagen). Finally, cells were plated on selective plates and incubated overnight at 37°C.

2.2.3. Recombinant protein expression in *E. coli*

Expression and purification screening New constructs of the proteins of interest were screened on ideal expression temperature, cell density and IPTG concentration for induction as well as protein solubility and affinity-tag accessibility prior to large-scale purification. For this, *E. coli* precultures were grown in 1 ml LB medium supplemented with appropriate antibiotics at 37°C shaking overnight. Main cultures were inoculated 1:100 in 5 ml LB medium supplemented with appropriate antibiotics and grown to varying OD_{600} values at 37°C shaking. Induction was induced with varying amounts of IPTG and protein expression was carried out at different temperatures at 160 rpm overnight. Cells were harvested by centrifugation at 4,000 x *g* for 10 min and the pellets were frozen at -20°C. Then, cells were resuspended in 1 ml of the respective lysis buffer and lysed by sonication. Cell debris and insoluble proteins were removed by centrifugation at 15,000 rpm and 4°C for 15 min in a microcentrifuge. The supernatant was incubated for 25 min at 8°C with 10 µl Magnetic Ni-NTA agarose beads (Life Technologies) pre-equilibrated with the respective wash buffer. The sample was washed four times with 500 µl of the respective wash buffer before protein elution was carried out with 50 µl of the respective elution buffer. Samples were taken from every step for SDS-PAGE analysis as described in 2.2.4.

Protein expression in *E. coli* For recombinant protein expression from a single vector chemically competent *E. coli* Rosetta (DE3) cells were used whereas coexpression from two different plasmids was carried out in electro-competent *E. coli* Codon-Plus (DE3) RIL cells. For this, the plasmids encoding the proteins of interest were transformed as described in 2.2.2. Precultures were grown in 20 ml LB medium supplemented with appropriate antibiotics at 37°C shaking overnight. Main cultures were inoculated 1:100 in LB medium supplemented with appropriate antibiotics and grown at 37°C and 160 rpm till OD_{600}=0.6-0.8. Expression was induced by addition of varying amounts of IPTG and carried out at 18-20°C and 160 rpm overnight. Cells were harvested by centrifugation at 5,000 rpm for 10 min and pellets were stored at -20°C.

2.2.4. Protein analysis

Determination of protein concentration Total protein concentrations were usually determined by measuring the absorption at 280 nm using a ND-1000 (NanoDrop) spectrometer. Molar absorption coefficients of individual proteins were calculated using the ProtParam software (Gasteiger E, 2005). If the protein didn't contain UV absorbing residues or the sample was contaminated with large amounts of non-protein molecules absorbing at 280 nm, such as nucleic acids and detergents, protein concentrations were determined by Bradford assay (Bradford, 1976). For this, the dye reagent (BioRad) was used as recommended by the manufacturer and samples were measured at 595 nm. Reference curves were generated for every new batch of the dye reagent using bovine serum albumin (Fraction V, Roth).

SDS-Polyacrylamide gel electrophoresis Electrophoretic separation of proteins was carried out using NuPAGE Novex 4-12% Bis-Tris Mini gels (Invitrogen) as recommended by the manufacturer. Vertical electrophoresis was carried out in X-Cell Sure Lock tanks (Invitrogen) using either NuPAGE MES or NuPAGE MOPS SDS running buffer (Table 10), dependent on the size of proteins to be separated. Samples were mixed with 5X SDS loading buffer (Table 10) prior to loading. Gels were stained with either Coomassie staining solution or InstantBlue (Expedeon) followed by destaining with H_2O.

TCA precipitation If protein concentrations were too low to be visible on an SDS gel, TCA precipitation was carried out before. For this, the sample was mixed with an equal volume of 20% (v/v) trichloroacetic acid and incubated on ice for 10 min. Then the sample was centrifuged at 15,000 rpm and 4°C for 10 min and the pellet was washed with 1 ml ice-cold acetone. The sample was centrifuged again at 15,000 rpm and 4°C for 10 min, the supernatant was discarded and the pellet was boiled at 95°C until completely dry. Then, the sample was resuspended in 1X SDS loading buffer (Table 10) for SDS-PAGE analysis. If the sample turned yellow, 5 µl Tris pH 8 were added to neutralize the acid.

Limited proteolysis Limited proteolysis time courses were performed to identify stable and compactly folded protein regions. 50 µg/ml of the protein of interest were digested with either 1 µg/ml trypsin, 1 µg/ml chymotrypsin, 0.2 µg/ml subtilisin A or 0.1 µg/ml proteinase K at 37°C. Samples were taken at 0, 1, 2, 4, 8, 15 and 60 min, respectively and the reaction was immediately stopped by the addition of 2X SDS loading dye (Table 10) followed by subsequent boiling at 95°C for 10 min. The degraded protein bands were separated by SDS-PAGE, excised and analyzed by mass spectrometry.

Mass spectrometry Unknown protein bands were identified by peptide mass fingerprinting by the Zentrallabor für Proteinanalytik (ZfP), Adolf-Butenandt-Institut, Munich using LC-MS/MS followed by a database search to identify the corresponding proteins.

2.2.5. Crystallization screening

Pre-Crystallization Test The Pre-Crystallization Test (Hampton Research) was used to determine appropriate protein concentrations prior to crystallization screening. The assay was performed as recommended by the manufacturer.

Initial crystallization screening Initial screening was performed at the Crystallization Facility of the Max-Planck-Institute of Biochemistry, Martinsried. Using various commercially available as well as in-house produced screens, sitting drop 96 well plates were set up at either 20°C or 4°C with a drop size ranging from 100 nl to 500 nl. Protein to reservoir ratios were adjusted to the specific needs and ranged between 1:1 and 3:1. Initial crystals were refined by varying pH values, concentrations of components and precipitants, temperature and drop-size either by self-designed user sitting-drop screens at the MPI or in-house using 24-well sitting drop plates or 15 well hanging-drop plates (Qiagen).

2.2.6. Bioinformatic tools

Protein and gene sequences were retrieved from NCBI or *Saccharomyces cerevisiae* genome databases (SGD). Sequence data was visualized and processed using ApE (Davis & Hammarlund, 2006) and Chromas (Technelysium). Multiple sequence alignments were generated using ClustalW2 or ClustalOmega (Larkin *et al.*, 2007; Goujon *et al.*, 2010; Sievers *et al.*, 2011; McWilliam *et al.*, 2013) and displayed using ESPript (Gouet *et al.*, 1999) or Jalview (Waterhouse *et al.*, 2009). Protein secondary structures were predicted by HHpred (Soding *et al.*, 2005) and psipred (Jones, 1999).

2.3. Expression and purification of specific proteins and protein complexes

2.3.1. Npa3 and variants

Full-length Npa3 from *Saccharomyces cerevisiae* was amplified from genomic DNA and subcloned into the pOPINI vector (provided by OPPF-UK) containing an N-terminal hexahistidine tag as described in 2.2.1. Truncated variants of Npa3 were amplified from full-length Npa3 plasmid DNA and either cloned into pOPINI vector containing an N-terminal hexahistidine tag or pOPINE vector (provided by OPPF-UK) containing a C-terminal hexahistidine tag. Deletions or/and mutations were introduced by overlap extension PCR as described in 2.2.1. All variants of Npa3 were transformed and expressed in *E. coli* Rosetta (DE3) (Novagen) as described in 2.2.2. and 2.2.3., respectively. Expression was induced with 0.5 mM IPTG for 20 h at 20°C. Cells were lysed by sonication in lysis buffer (Table 11). After centrifugation at 24,000 x *g* for 30 min, the cleared lysate was loaded onto a 2 ml Ni-NTA column

(Qiagen), pre-equilibrated with nickel wash buffer (Table 11). The column was washed with 10 CV nickel wash buffer before elution of the bound protein with 3 CV nickel elution buffer (Table 11). The conductivity of the eluate was adjusted to match that of the anion exchange buffer A (Table 11) and applied to a MonoQ 10/100 GL column (Amersham) equilibrated with anion exchange buffer A. The protein was eluted with a linear gradient over 15 CV from 100 mM to 1 M NaCl in anion exchange buffer B (Table 11). Fractions containing Npa3 were pooled and the sample was concentrated for a final size exclusion step. Size exclusion chromatography was carried out using either a HiPrep Sephacryl S-300 HR column (GE Healthcare) or a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) pre-equilibrated in SEC buffer 100 (Table 11) for all variants except Npa3(1-264Δ203-211) were this step was performed in SEC buffer 200 (Table 11). Fractions containing Npa3 were pooled and full-length Npa3 was concentrated to 2-100 mg/ml, flash frozen in liquid nitrogen and stored as 5-20 µl aliquots at -80°C. Npa3(1-264Δ203-211) was concentrated to 3.7 mg/ml and used for crystallization without freezing.

2.3.2. Npa3-GPN2 complexes

Npa3 variants were cloned into pOPINE vector (provided by OPPF-UK) and a stop codon was introduced at the 3' end of the gene to avoid expression of the plasmid encoded C-terminal hexahistidine tag. GPN2 variants were cloned into pOPINB vector (provided by OPPF-UK) with an N-terminal hexahistidine tag and both vectors were cotransformed into *E. coli* BL21 (DE3) RIL (Stratagene) cells by electroporation as described in 2.2.2. Protein expression was carried out as described (2.2.3) using 0.5 mM IPTG for induction. Purification was performed as described for full-length Npa3 (2.3.1).

2.3.3. Iwr1

Full-length Iwr1, cloned into a pET21b vector with an N-terminal hexahistidine tag was transformed into *E. coli* Rosetta (DE3) (Novagen) cells as described (2.2.2). Protein expression was induced with 0.5 mM IPTG for 20 h at 18°C as described (2.2.3). Cells were lysed by sonication in Iwr1 Nickel Lysis buffer (Table 12). After centrifugation at 24,000 x *g* for 30 min, the cleared lysate was loaded onto a 2 ml Ni-NTA column (Qiagen) equilibrated with Iwr1 Nickel Lysis buffer. The column was washed four times with 10 CV Iwr1 Nickel wash buffer sequentially containing 10, 20, 30 and 40 mM imidazole (Table 12), respectively and eluted with 3 CV Iwr1 Nickel elution buffer (Table 12). The eluate was applied to a MonoQ 10/100 GL column (Amersham) equilibrated with Iwr1 anion exchange buffer A (Table 12) and eluted over a linear gradient of 15 CV from 100 mM NaCl to 1 M NaCl in Iwr1 anion exchange buffer B (Table 12). After concentration with a partial buffer exchange using three volumes Pol II buffer (Table 12), the sample was applied to a Superdex 75 10/300 or Superose 12 10/300 size exclusion column (GE Healthcare) equilibrated with Pol II buffer. Fractions containing Iwr1 were pooled and concentrated to 2-4 mg/ml, flash frozen in liquid nitrogen and stored as 20 µl aliquots at -80°C.

2.3.4. Endogenous RNA polymerase II from *S. cerevisiae*

Yeast cell lysis Fermented yeast cell pellets of the Pol II purification strain (Table 5) were provided by Stefan Benkert and pellets were thawed in a water bath at a maximum of 30°C. 1 ml 100X PI (Table 10) was added to the cell suspension and cells lysis was carried out using beat beaters (Hamilton Beach). For this, 200 ml of the cell mixture were added to a metal chamber together with 200 ml of glass beads (0.5 mm diameter, BioSpec Products). Air bubbles were removed by stirring gently with a glass rod. The chamber was completely filled up with HSB150 (Table 13) before assembly with the impeller. Lysis was carried out for 80 min at 4°C with cycles of 30 s on and 90 s off. To avoid warming of the cell mixture the lysis chamber was covered with a salt-ice mixture. The lysate was then separated from the beads by filtering through a mesh funnel and washed with HSB150. Finally, the lysate was centrifuged twice for 30 min at 13,690 x *g* at 4°C.

Protein purification The cleared lysate was subjected to ultracentrifugation for 90 min at 76,220 x *g* and 4°C. The aqueous phase was collected and protein precipitation was carried out with 50% (w/v) $(NH_4)_2SO_4$ stirring overnight at 4°C. The solution was then centrifuged twice for 45 min at 34,200 x *g* and 4°C and the precipitate was dissolved in 140 ml HSB150 per 100 g pellet by stirring for 1 to 2 h at 4°C. The conductivity was adjusted to that of HSB1000/7 (Table 13). The sample was then applied to pre-equilibrated Ni-NTA resin (Qiagen) and washed with 5 CV HSB1000/7 and 3 CV Ni buffer 7 (Table 13) before elution of bound protein with 3 CV Ni buffer 50 and 3 CV Ni buffer 100 (Table 13). Elution fractions were pooled and conductivity was adjusted to that of MonoQ150 buffer (Table 13). The sample was loaded on a MonoQ 10/100 GL column (GE Healthcare) pre-equilibrated with MonoQ150 and eluted with a linear gradient over 12 CV to 75% buffer MonoQ2000 (Table 13). Fractions containing Pol II were pooled and diluted with 3 volumes Pol II buffer (Table 13) before concentration to around 1 ml. Then 4-fold molar excess of recombinant Rpb4/7 (for purification protocol see 2.3.5.) were added and incubated on ice for 45 min. Finally, the complete assembled Pol II complex was applied to a Superose 6 10/300 column (GE Healthcare) equilibrated with Pol II buffer and fractions containing Pol II were concentrated to 3-3.8 mg/ml, flash frozen in liquid nitrogen and stored in 20 µl aliquots at -80°C.

2.3.5. Rpb4/7

Protein expression was performed as described in 2.2.3 and expression was induced with 0.5 mM IPTG for 16 hours at 20°C. Cells were lysed by sonication in Ni buffer 0 (Table 14) containing 1X PI (Table 10). After centrifugation at 16,000 x *g* for 20 min, the cleared lysate was applied to a pre-equilibrated (Ni buffer 0) Ni-NTA agarose column (Qiagen) (Table 14). The column was washed with 5 CV Ni buffer 0, 3 CV Ni salt buffer, 3 CV Ni buffer 10 and 3 CV Ni buffer 20 (Table 14). Protein was eluted with 3 CV Ni buffer 50 followed by 6 CV Ni buffer 200. Fractions containing Rpb4/7 were pooled and applied to a Source15Q 16/10 column (GE Healthcare) pre-equilibrated with SourceQ 100 buffer (Table 14). The

column was washed with 10 CV SourceQ 100 buffer and the protein was eluted with a linear gradient over 10 CV to 1 M NaCl in SourceQ 1000 buffer (Table 14). Finally, fractions containing Rpb4/7 were pooled and a size exclusion step was performed using a HiLoad 26/60 Superdex75 pg column (GE Healthcare), pre-equilibrated in Pol II buffer (Table 14). Protein was concentrated to 6 mg/ml and 40 µl aliquots were flash frozen in liquid nitrogen and stored at -80°C.

2.4. Crystallization

2.4.1. Crystallization of Npa3ΔCΔLoop·GMPPCP

Npa3ΔCΔLoop (comprising residues 1-264Δ203-211) was freshly purified as described in 2.3.1 and nucleotide exchange was carried by incubation with 10 mM GMPPCP at 8°C overnight. Crystals were grown in a 1.5 ml Eppendorf tube in SEC buffer 200 (Table 11) for 1-3 days at 8°C. Crystals were fished directly from the Eppendorf tube and transferred to a spot plate containing 100 µl mother solution. Cryo protection was carried out at 8°C as described in 2.4.4.

2.4.2. Crystallization of Npa3ΔCΔLoop·GDP·AlF_x

A fresh 50 mM AlF_x stock solution was prepared by mixing equal volumes of 1 M NaF and 100 mM $AlCl_3$ (both dissolved in SEC buffer 200 (Table 11)) followed by an incubation for 30 min at RT. Then, 5 mM GDP and 10 mM AlF_x were incubated with freshly purified Npa3ΔCΔLoop (comprising residues 1-264Δ203-211) overnight at 8°C. Crystals were grown for 1-3 days by hanging drop vapor diffusion at 8°C with SEC buffer 200 supplemented with 10 mM AlF_x as reservoir solution. Cryo protection procedure was carried out at 8°C as described in 2.4.4.

2.4.3. Crystallization of Npa3ΔCΔLoop·GDP

Npa3ΔCΔLoop (comprising residues 1-264Δ203-211) was freshly purified as described in 2.3.1 and the sample was incubated with 10 mM GDP at 8°C overnight. Crystals were grown for 6-15 days by sitting drop vapor diffusion at 20°C with 9 mM HEPES pH 7.0, 45 mM NaCl, 5 mM $MgCl_2$ and 5% (v/v) Jeffamine M-600 as reservoir solution. Cryo protection was carried out at 20°C as described in 2.4.4.

2.4.4. Cryo-protection and freezing

Crystals were transferred to a spot plate containing 100 µl mother solution. Cryo-protection was carried out by step-wise transferring to mother solution containing 35% (v/v) glycerol (steps comprised incubation with 20, 40, 60, 80 and 100% of the cryo solution, respectively). Crystals were incubated for 30 min between every cryo-step before harvesting and flash-cooling in liquid nitrogen.

2.5. Data collection and X-ray structure determination

2.5.1. Data collection

Diffraction data of crystals Npa3ΔCΔLoop·GDP·AlF_x were collected at beamline X06DA of the Swiss Light Source (Villigen, Switzerland) at 100 K and 1.0001 Å, respectively (Table 18). Diffraction data of crystals Npa3ΔCΔLoop·GMPPCP and Npa3ΔCΔLoop·GDP were collected at beamline MX1 at EMBL/DESY (Hamburg, Germany) at 100 K and 0.99988 Å, respectively (Table 18).

2.5.2. Experimental phasing

A single anomalous diffraction experiment from intrinsic sulphur atoms (S-SAD) was performed on the same Npa3ΔCΔLoop·GDP·AlF_x crystal as in 2.5.1 at beamline X06DA of the Swiss Light Source in Villigen, Switzerland. Diffraction data were collected at 100 K and a wavelength of 2.066 Å at three different χ angles (0°, 10° and 20°) as described recently (Weinert *et al.*, 2015) (Table 18). For each dataset 7,200 images were collected with an increment of 0.1°. Raw data were processed with XDS and the three datasets were merged using XSCALE (Kabsch, 2010b; 2010a). The programs SHELXC/D/E of the SHELX suite (Sheldrick, 2008) were used for detection of the sulphur atoms and for SAD phasing.

2.5.3. Molecular Replacement

The Npa3ΔCΔLoop·GDP structure was solved by molecular replacement using PHASER (McCoy *et al.*, 2005) with the Npa3ΔCΔLoop·GDP·AlF_x core structure as search model. The number of molecules per asymmetric unit were estimated using the program matthews from the CCP4 suite (Collaborative Computational Project, 1994).

2.5.4. Model building and refinement

The Npa3ΔCΔLoop·GMPPCP structure could be refined using the Npa3ΔCΔLoop·GDP·AlF_x structure with phenix.refine (Afonine *et al.*, 2005), because the crystals had the same space group and almost identical unit cell parameters.

Atomic models of all structures were iteratively built with COOT (Emsley & Cowtan, 2004) and refined with phenix.refine. Figures were prepared with PyMOL (deLano Scientific).

2.6. Functional characterization of Npa3

2.6.1. Analysis of GTPase activity

GTPase activity was measured in 96-well plates using the Malachite Green Phosphate Assay Kit (BioAssay Systems) as recommended by the manufacturer. 0.15 µM of Npa3 variants were incubated with 100 µM GTP (Fermentas) at 37°C and 600 rpm in SEC buffer 200 (Table 11) in a thermomixer. Orthophosphate concentrations from at least three independent experiments were determined at various time points by measuring the absorption at 620 nm at four different positions within each well using an Infinite M1000 plate reader (Tecan) and the values were averaged. Control experiments were performed to determine orthophosphate contaminations in the individual solutions and data was corrected for this value.

GTPase activity in the presence of partially unfolded citrate synthase (CS) was measured at 43°C in SEC buffer 100 (Table 11) with the pH value adjusted to 7.5 at 43°C.

2.6.2. Isolation of bound nucleotides

GTPase-bound nucleotides were isolated either to check nucleotide loading after purification or to monitor nucleotide exchange over time after *in vitro* incubation with varying amounts of GTP-derivatives.

For this 125 µM of the purified GTPase were used and, for the latter one, incubated with excess of the desired nucleotide *in vitro* for different time periods at 4°C. Here, unbound nucleotides were removed by loading the sample onto a Mini Bio-Spin Chromatography column (BioRad), pre-equilibrated with the respective protein buffer, and elution was carried out by centrifugation for 4 min at 1,000 x *g* in a microcentrifuge. Afterwards, the bound nucleotide was released from the protein by denaturation at 95°C for 5 min. Denatured protein was removed by centrifugation at 15,000 rpm for 10 min and the supernatant containing the released nucleotide was used for HPLC analysis (2.6.3).

2.6.3. High performance liquid chromatography (HPLC)

Identification and quantification of nucleotides was performed by HPLC. Here, a 50 µl sample containing 125 µM nucleotide was loaded on a C18 reversed phase column (25 cm x 4.6 mm, 5 µm particle size, Sigma-Aldrich) pre-equilibrated in buffer A (Table 15). Elution was carried out at 1 ml/min for 55 min before very strongly bound components were eluted with buffer B (Table 15). Retention times for nucleotide standards were determined experimentally using 250 µM GTP, GMPPCP, GMPPNP, GDP or GMP, respectively. Nucleotides were detected by measuring the absorption at 260 nm. Comparative quantification of eluted samples was carried out by peak integral calculation and comparison with known standards.

2.6.4. Chaperone assay

Thermal aggregation of citrate synthase (CS), a general chaperone substrate protein, was essentially carried out as described (Buchner *et al.*, 1998) with modifications. An ammonium sulfate solution of CS from porcine heart (Sigma) was dialyzed against TE buffer (Table 16) overnight at 4°C. CS was concentrated to 17 mg/ml and aliquots were flash frozen in liquid nitrogen. For each experiment an aliquot of CS was supplemented with TE buffer to yield a concentration of 30 µM. Possible precipitates were removed by centrifugation at 14,000 rpm for 30 min at 4°C and protein concentrations were measured again prior to experimental procedure.

Varying amounts of Npa3 were added to a quartz cuvette containing either 800 µl preheated (43°) incubation buffer (Table 16) in the absence of nucleotides or 800 µl pre-heated (43°C) SEC buffer 100 (Table 11) (pH 7.5 at 43°C) and 1 mM GMPPCP, GTP or GDP, respectively. The cuvette was transferred to a thermostated Fluoromax 3 fluorometer (HORIBA Scientific) and light scattering was measured at 360 nm and 43°C to check whether Npa3 aggregates under the assay conditions. If a stable baseline was observed, CS (monomer) was added to reach a final concentration of 0.15 µM and CS aggregation was measured every 0.2 s for 50 min by light scattering at 360 nm. Data was normalized on the baseline and measurement of at least three independent experiments were averaged.

2.7. Interaction studies

2.7.1. *In vitro* Biotin-Pulldown

Biotinylation of RNA polymerase II RNA polymerase II containing a biotin acceptor peptide at the N-terminus of Rpb3 was purified as described (2.3.4). Enzymatic biotinylation was carried out *in vitro* in 200 µl reactions containing 270 µg of Pol II, 15 µg BirA (thankfully provided by Dr. Laurent Larivière), 2 mM ATP pH7 and 100 µM biotin supplemented with BirA dilution buffer (Table 17). The mixture was incubated for 2.5 h in a thermomixer at 20°C and 600 rpm before buffer exchange to SEC buffer 100 (Table 11) was carried out using Micro Bio-Spin™ Chromatography columns (BioRad) as recommended by the manufacturer. Aliquots of biotinylated Pol II were prepared, flash frozen in liquid nitrogen and stored at -80°C.

Pull-Down of RNA polymerase II associated proteins 3-5 µg of biotinylated Pol II were incubated with 5-15 times molar excess of the purified putative binding protein at 8°C and 600 rpm overnight (and 1 mM of the respective nucleotide if required). The mixture was then incubated with 20 µl of pre-equilibrated magnetic streptavidin beads (Invitrogen) in a thermomixer at 22°C and 750 rpm for 30 min and unbound proteins were removed by washing three times with SEC buffer 100 (Table 11) containing 0.05 % (v/v) NP-40. The supernatant was removed and elution was carried out by adding 15 µl of 1x SDS buffer (Table 10) followed by incubation at 95°C for 2 min. The elution fraction was analyzed by SDS-PAGE as described (2.2.4).

2.7.2. Native gel electrophoresis

3 µg of RNA polymerase II were incubated with 3-10 times molar excess of the purified putative binding proteins at 8°C overnight. Native gel electrophoresis was carried out at 100 V using the NativePAGE™ Novex® Bis-Tris Gel System (Life Technologies) as recommended by the manufacturer. Gels were stained using Instant blue (Expedeon).

2.7.3. Analytical size exclusion chromatography

Purified, putatively interacting proteins were incubated in 200 µl reactions in the respective protein buffers at 20°C for 30 min or at 8°C overnight. The sample was then loaded on a Superose 6 or Superose 12 column (GE Healthcare) depending on the size of the proteins. Protein fractions were TCA precipitated (2.2.4) if necessary and analyzed by SDS-PAGE as described (2.2.4).

2.7.4. Coexpression and His-Affinity purification

Interactions between various constructs of the GPN-loop GTPases as well as the Pol II nuclear import factor Iwr1 were analyzed by coexpression followed by His-tag affinity purification. Therefore electro-competent BL21 Gold (RIL) cells were cotransformed (2.2.2) with only one of the proteins containing a hexahistidine-tag. Protein expression was performed as described (2.2.3) and the cleared cell lysate was incubated at 8°C for 25 min with 10 µl magnetic Ni-NTA beads (Life Technologies) equilibrated with Lysis buffer (Table 11). The beads were then washed four times with 500 µl Nickel wash buffer (Table 11) before elution was performed by incubation with 50 µl Nickel elution buffer (Table 11) for 3 min on ice. Samples were then analyzed by SDS-PAGE (2.2.4).

2.7.5. Tandem affinity purification

TAP-tagged yeast strains were obtained from Thermo Scientific Open Biosystems (Table 5). Tandem affinity purifications (TAP) from 2l yeast cultures were done as described (Puig *et al.*, 2001) by Anja Kieser from the Sträßer Lab.

2.7.6. Immobilized peptide microarrays

An array of 1,139 15meric peptides covering the complete sequence of RNA polymerase II with an overlap of 11 amino acids were synthesized and triplicates were N-terminally immobilized on a glass surface via a Ttds-linker by JPT Peptide Technologies, Berlin, Germany. Purified N-terminal hexahistidine tagged wild-type Npa3 was pre-incubated with either 10 mM GMPPCP or 10 mM GDP overnight at 4°C in SEC buffer 100 (Table 11). Incubation with 10 mM GTP was started 5 min prior to experimental procedure. Microarray incubations and data analysis was carried out by JPT Peptide Technologies. Microarrays were blocked with blocking buffer (Pierce International, Superblock TBST20) for 60 min, washed with TBS containing 0.1% Tween20 (TBS-T) and incubated with 1 mg/ml of the respective Npa3 sample for 60 min at 4°C. After an additional TBS-T washing step 0.4 µg/ml Penta His antibody Alexa 647 (Qiagen) were incubated for 45 min followed by washing with 3 mM SSC buffer (JPT Peptide Technologies, Berlin, Germany). The microarrays were dried and fluorescence signal was analyzed using a Genepix Scanner 4200AL (Molecular Devices) and GenePix spot-recognition software by JPT Peptide Technologies. In case of false positive binding, neighboring overlapping peptides containing partially the same sequence were also not taken into consideration.

3. Results and Discussion

3.1. Structure and function of Npa3-nucleotide complexes

3.1.1. Npa3 domain organization and crystallization

Npa3 contains 385 amino acid residues and consists of a central GTPase core (residues 1-41, 83-184, and 225-263), two protein insertions ('insertion 1' and 'insertion 2' containing residues 42-82 and 185-224, respectively), and a C-terminal tail (residues 264-385) (Figure 4A). The GTPase core harbors the motifs G1-G5 that are required for GTP binding and hydrolysis (Bourne*, et al.*, 1991). The C-terminal tail is poorly conserved among eukaryotes and is absent in archaea. Because efforts to crystallize full-length Npa3 failed, we removed the C-terminal tail and part of a loop (residues 203-211 in insertion 2) that is predicted to be disordered and absent in most eukaryotic homologs (variant Npa3ΔCΔLoop, comprising residues 1-202 and 212-264). We co-crystallized Npa3ΔCΔLoop (Figure 4) with the non-hydrolyzable GTP analog GMPPCP, but also with GDP, and with GDP and AlF_x ('GDP·AlF_x'), which mimics the pentavalent transition state (Wittinghofer, 1997).

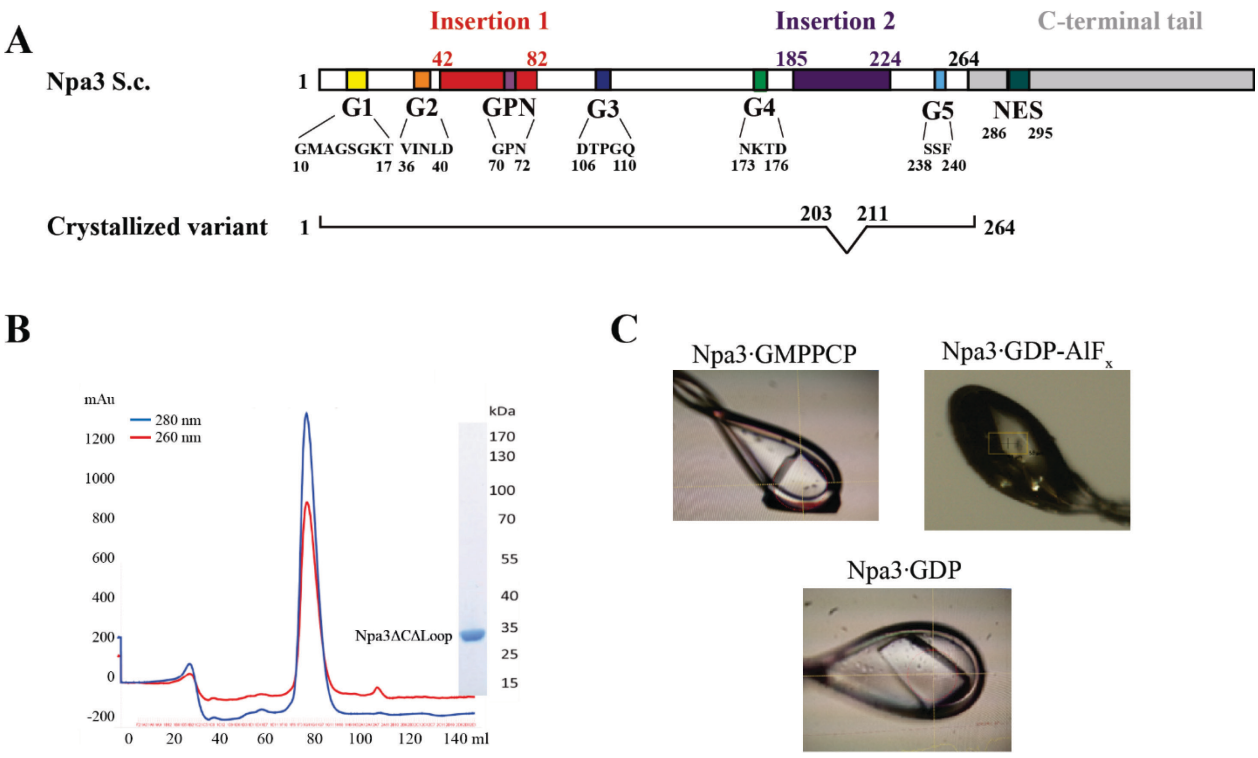


Figure 4| Npa3 domain organization, purification and crystallization.
(A) Schematic representation of Npa3 from the yeast *Saccharomyces cerevisiae*. NES: Nuclear export sequence. The color-code is used throughout the figures. **(B)** Purification of Npa3ΔCΔLoop. The chromatogram of the final gel filtration step (HiPrep Sephacryl S300) and the corresponding SDS gel of the peak fraction is shown. **(C)** Crystals of Npa3 in complex with various nucleotides.

3.1.2. Structure determination of Npa3 complexes

The structure with GDP·AlF_x was solved using single anomalous diffraction from intrinsic sulfur atoms (Figure 5) and a protein model was built and refined to a free R-factor of 23.9% at 1.85 Å resolution (Table 18). The model was used to solve the structures containing GMPPCP, or GDP, which were refined to free R-factors of 22.7% and 27.6% at 2.2 Å and 2.3 Å resolution, respectively (Table 18). The electron density of the GDP·AlF_x complex was not clearly interpretable in the active site region, where it showed a mixture of different nucleotide configurations. Thus this structure was discarded, although the protein model was excellent. The structures containing GMPPCP or GDP showed very well defined densities for bound nucleotides, represented defined enzyme states, showed great stereochemistry, and were used for further analysis. The structures are relevant for all eukaryotic orthologs because Npa3 is highly conserved, with 50% of the residues in the crystallized variant being identical between *S. cerevisiae* and human enzymes (Figure 6A and C). The structures revealed a closed GDP-bound and open GMPPCP-bound state (Figure 6B) that are described in the following sections (3.1.3 and 3.1.4).

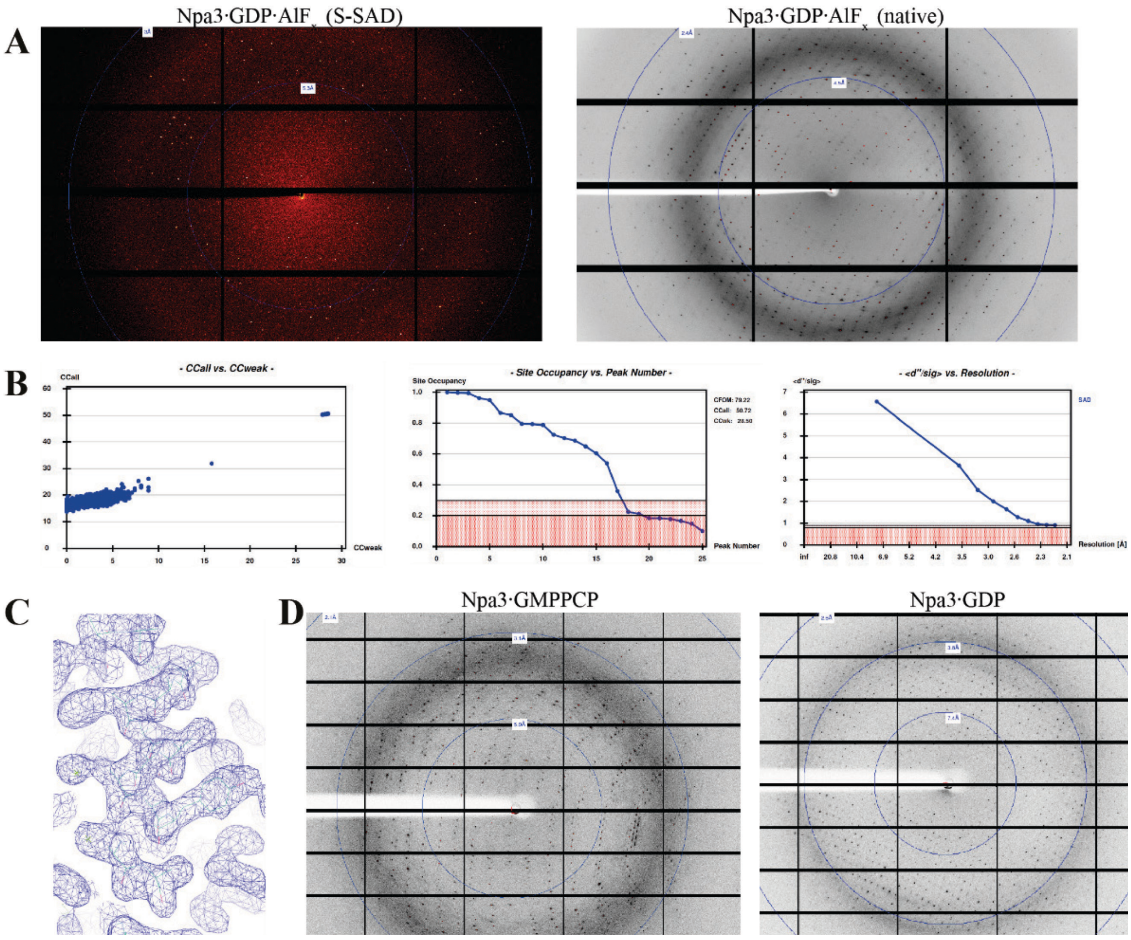


Figure 5| X-ray diffraction and structure solution.
(A) Diffraction image of Npa3ΔCΔLoop·GDP·AlF_x in the S-SAD (left panel) and native diffraction experiment (right panel). **(B)** Structure solution of Npa3ΔCΔLoop·GDP·AlF_x using SHELX revealed a clear solution (CC_{all}=50.72, CC_{weak}=28.5, left panel). The occupancy of 17 sulfur sites is shown in the middle panel (red areas, with occupancies below 30% were discarded as noise or bound solvent molecules). The anomalous signal vs resolution is shown in the right panel and data >2.7 Å (d''/sig > 1.2) were used for further processing. **(C)** Initial density of the S-SAD experiments shows clearly defined densities for side-chains. Shown is a representative portion of the initial density before refinement with the high-resolution native-data set. **(D)** Diffraction images of GMPPCP- (left panel) and GDP- bound (right panel) crystals.

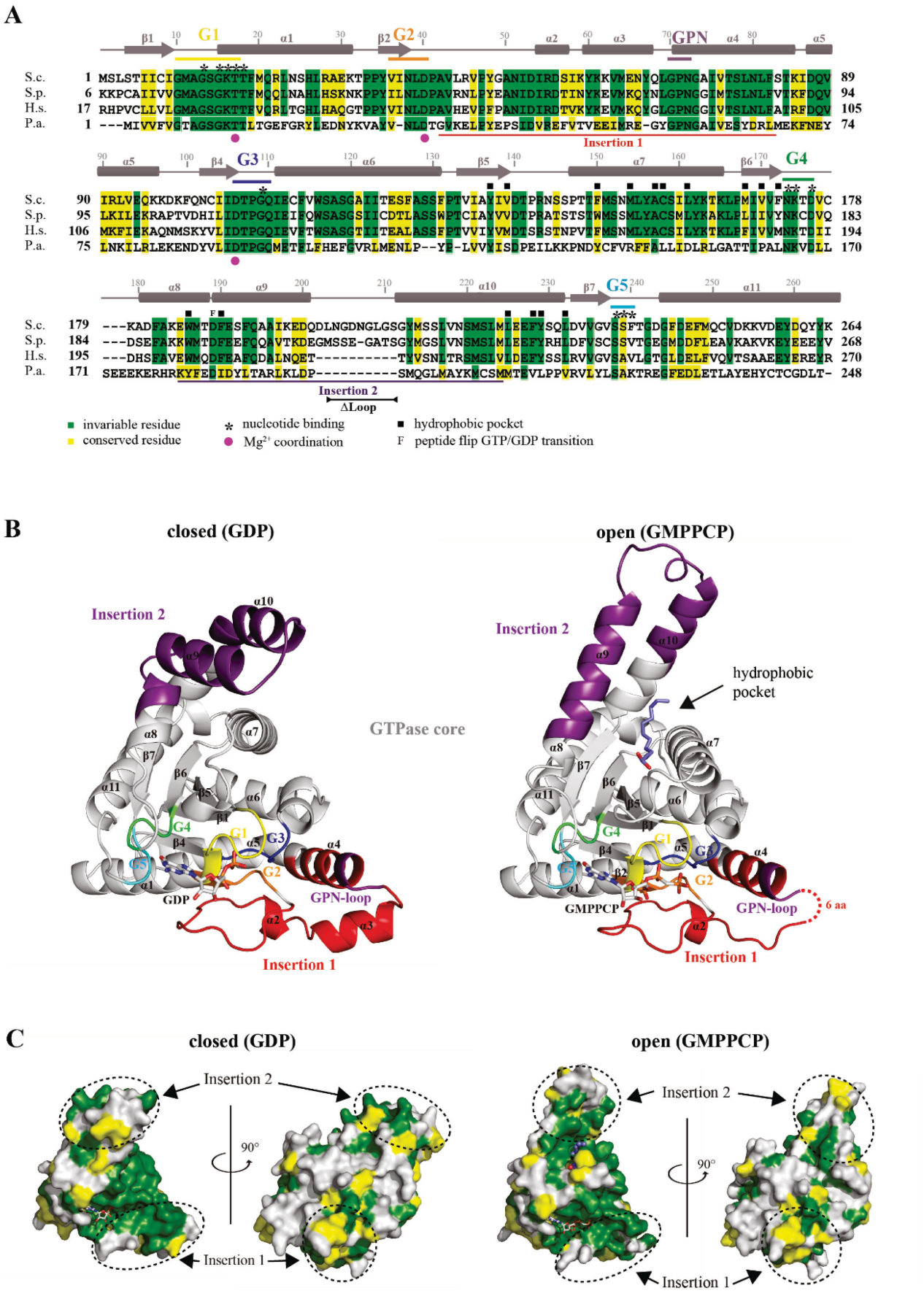


Figure 6| Crystal structures of Npa3 in GDP- (closed) and GMPPCP-bound (open) forms.
(A) Amino acid sequence alignment of Npa3 from *S. cerevisiae* (*S.c.*), with eukaryotic homologs from *S. pombe* (*S.p.*),

GPN1 from *H. sapiens* (*H.s.*) and the archaeal homolog Pab0955 from *P. abyssi* (*P.a.*). Secondary structure elements are indicated above the sequence (cylinders: α-helices; arrows: β-strands). Amino acid numbering above the sequence corresponds to Npa3 from *S.c.* Invariant residues are in green and conserved residues in yellow. Motifs G1-G5 and insertions 1 and 2 are marked with bars, residues of the hydrophobic pocket with black squares, residues involved in nucleotide-binding with asterisks, and residues involved in magnesium binding with pink spheres. **(B)** Ribbon representation of the closed (GDP-bound, left) and open (GMPPCP-bound, right) conformation. The G motifs and insertion regions are colored according to (A). A fatty acid bound to the hydrophobic pocket that is opened in the Npa3·GMPPCP structure is shown as slate blue sticks. Missing residues are indicated with dashed lines. **(C)** Surface conservation of Npa3 in the closed, GDP-bound (left panels) and open, GMPPCP-bound (right panels) state. Invariant residues are shown in green, conserved residues in yellow, variable residues in grey.

Table 18| X-ray diffraction data collection and refinement statistics

		Native		S-SAD
	GDP[a]	GMPPCP[a]	GDP-AlF_x[b]	GDP-AlF_x[b, c]
Data collection				
Space group	$C222_1$	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions				
a, *b*, *c* (Å)	108.0, 119.2, 347.5	116.2, 116.2, 56.8	116.2, 116.2, 55.9	116.2, 116.2, 55.9
Wavelength (Å)	0.99988	0.99888	1.0001	2.066
Resolution (Å)	50-2.3 (2.36-2.3)[d]	50-2.2 (2.26-2.20)[d]	80-1.85 (1.90-1.85)[d]	50-2.15 (2.21-2.15)[c]
R_{sym} (%)	5.7 (126.9)	4.5 (155)	3.8 (145.6)	5.2 (89.8)
$I/\sigma I$	20.16 (1.88)	28.50 (1.94)	38.32 (2.16)	66.37 (2.00)
Completeness (%)	99.7 (99.8)	99.5 (98.0)	100 (100)	98.2 (79.9)
Redundancy	7.15 (6.91)	14.11 (13.12)	14.39 (14.42)	63.07 (7.31)
$CC_{(1/2)}$[e] (%)	100 (69.0)	100 (83.2)	100 (72.0)	100 (62.3)
Refinement				
Resolution (Å)	45.3-2.3	36.8-2.2	46.2-1.85	
No. reflections	99,343	20,172	33,217	
R_{work}/ R_{free} (%)	23.8/27.6	21.4/22.7	20.1/23.9	
No. atoms				
Protein	12065	1998	2036	
Nucleotide	168	32	28	
AlF_x	-	-	4	
Mg^{2+}	6	1	1	
Lauric acid	-	14	14	
Glycerol	24	12	12	
Water	121	5	108	
B-factors[f] (Å^2)				
Protein	83.5	121.2	59.4	
Nucleotide	69.6	123.6	57.6	
AlF_x	-	-	107.1	
Mg^{2+}	64.2	166.6	104.7	
Lauric acid	-	108.0	49.7	
Glycerol	107.7	145.2	63.8	
Water	60.1	110.7	53.5	
R.m.s deviations				
Bond lengths (Å)	0.011	0.007	0.008	
Bond angles (°)	1.29	1.15	1.18	

[a]Diffraction data were collected at beamline MX1 at EMBL/DESY, Hamburg, Germany and processed with XDS (Kabsch, 2010b).

[b]Diffraction data were collected at beamline X06DA, Swiss Light Source, Switzerland and processed with XDS (Kabsch, 2010b).

[c]Data are merged form three data sets, measured at different χ angles (0°, 10°, 20°) from a single crystal.

[d]Numbers in parenthesis refer to the highest resolution shell.

[e]$CC_{1/2}$ = percentage of correlation between intensities from random half-datasets (Karplus & Diederichs, 2012).

[f]The average over all copies in the asymmetric unit is given.

3.1.3. Npa3·GDP structure shows eukaryote-specific features

The Npa3 core consists of a central, six-stranded parallel β-sheet surrounded by helices (Figure 6B and Figure 7). The two insertion regions extend the Npa3 core, resulting in an overall L-like shape of the enzyme. The asymmetric unit of the Npa3·GDP crystals contains six enzymes that differ only slightly in regions forming crystal contacts. The Npa3·GDP structure resembles the homologous archaeal structure Pab0955·MgGDP (Gras*, et al.*, 2007) (PDB-code 1YRB), but also reveals major differences in insertions 1 and 2 (Figure 7) and in regions connecting the core to insertion 2. A single residue links strand β7 in the core to the C-terminal end of helix α10, which is bent by ~50° (Figure 7A and B). In the archaeal structure, strand β7 is shorter and the linker region to α10 comprises five residues, allowing for a straight conformation of the helix. Whereas helices α9 and α10 in Npa3·GDP lie side by side, they are tilted in the archaeal structure (Figure 7A and C). Helix α3 in insertion 1 of Npa3·GDP is apparently

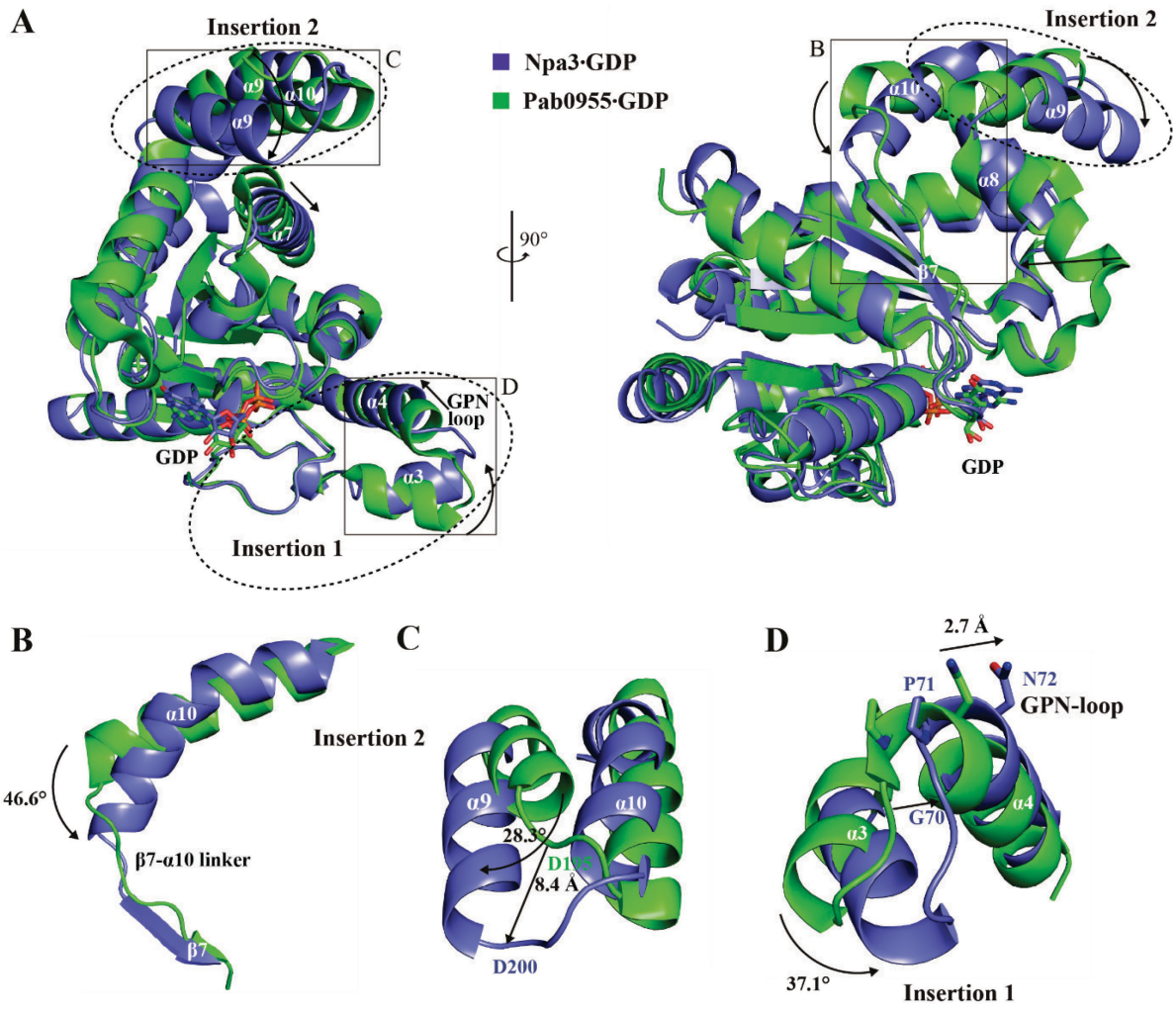


Figure 7| Superposition of Npa3·GDP (blue) and the archaeal GPN-loop GTPase Pab0955·GDP (green) reveals eukaryote-specific structural features.
(A) Superposition of overall structures from Npa3·GDP and Pab0955·GDP (pdb code: 1YRB,(Gras *et al.*, 2007)) shown as front view (left) and side view (right). Significant differences between the two structures are indicated with black arrows and the insertion regions are shown with dashed lines. Boxes highlight areas shown in B-D. **(B)** The linker between strand β7 and helix α10 is shorter in Npa3 leading to a strong bending of helix α10. **(C)** Helix α9 of Npa3 is rotated, leading to significant differences of insertion 2 compared to the archaeal GPN-loop GTPase. **(D)** Helix α3 of insertion 1 is rotated and the GPN-loop is shifted.

rotated by 37° towards the core, and helix α4 and the GPN-loop are shifted by ~3 Å towards the G3 motif (Figure 7A and D).

Taken together, the core of the eukaryotic enzyme generally resembles its archaeal counterpart, whereas the two insertions differ significantly.

3.1.4. Npa3·GMPPCP structure reveals novel open conformation

Our Npa3·GMPPCP structure reveals a novel open conformation that differs significantly from the Npa3·GDP structure (Figure 6B and Figure 8). Both insertions are rearranged and the position of helix α7 is changed, opening an extended hydrophobic pocket. Insertion 2 adopts a different fold, and remotely resembles the helical bundle of the N domain of the prokaryotic GTPase Ffh from *T. aquaticus* (pdb-code 1LS1) (Ramirez *et al.*, 2002). In the GMPPCP-bound state, residues D189 of insertion 2 and C179 show peptide flips (Figure 8B and E) that enable a straight conformation of helices α8 and α9 and the formation of a single helix. Helix α10, which is bent in the Npa3·GDP structure, adopts a straight conformation, and helix α7 is rotated by ~15° (Figure 8A-C and F). This disrupts a methionine-rich hydrophobic core formed between helices α7, α8, and α9 that kept the pocket closed in the Npa3·GDP structure (Figure 8C). Rotation of helix α7 enables it to hydrogen bond to residue E112, which no longer interacts with Q110 in motif G3 and N72 in the GPN-loop, as observed in the closed Npa3·GDP structure (Figure 8F). Mutation of any one of these residues (E112, Q110, N72) is lethal in yeast (Forget*, et al.*, 2010; Staresincic*, et al.*, 2011; Alonso*, et al.*, 2013). In addition, insertion 1 is changed in the Npa3·GMPPCP structure, including motif G3, which binds the GPN-loop (Figure 8D). Insertion 1 is more flexible than in the GDP-bound structure. Its hydrogen bonds with Y48 to D106 (G3) and with Y60 to the carbonyl of P71 (GPN-loop) are lost, helix α3 is unfolded, and residues 64-69 are mobile. Positioning of insertion 1 apparently involves a novel “DIRD” motif that comprises the invariant residues D53, I54, R55, and D56 (Figure 8D). The DIRD motif is, along with the GPN motif, the most highly conserved region within the sequence of the yeast Npa3 paralogs GPN2 and GPN3 (Figure 9). Residue N38 in motif G2 bridges between D106 in motif G3 and D53 in the DIRD motif to keep helix α2 and thus insertion 1 in close proximity to the core (Figure 8D). The side chain of D56 in the DIRD motif is flipped in the Npa3·GMPPCP structure and hydrogen bonds to D53, possibly stabilizing the changed position of insertion 1. Taken together, Npa3 can adopt two very different conformations, a closed conformation observed with bound GDP, and a novel open conformation observed when the GTP analog GMPPCP is bound.

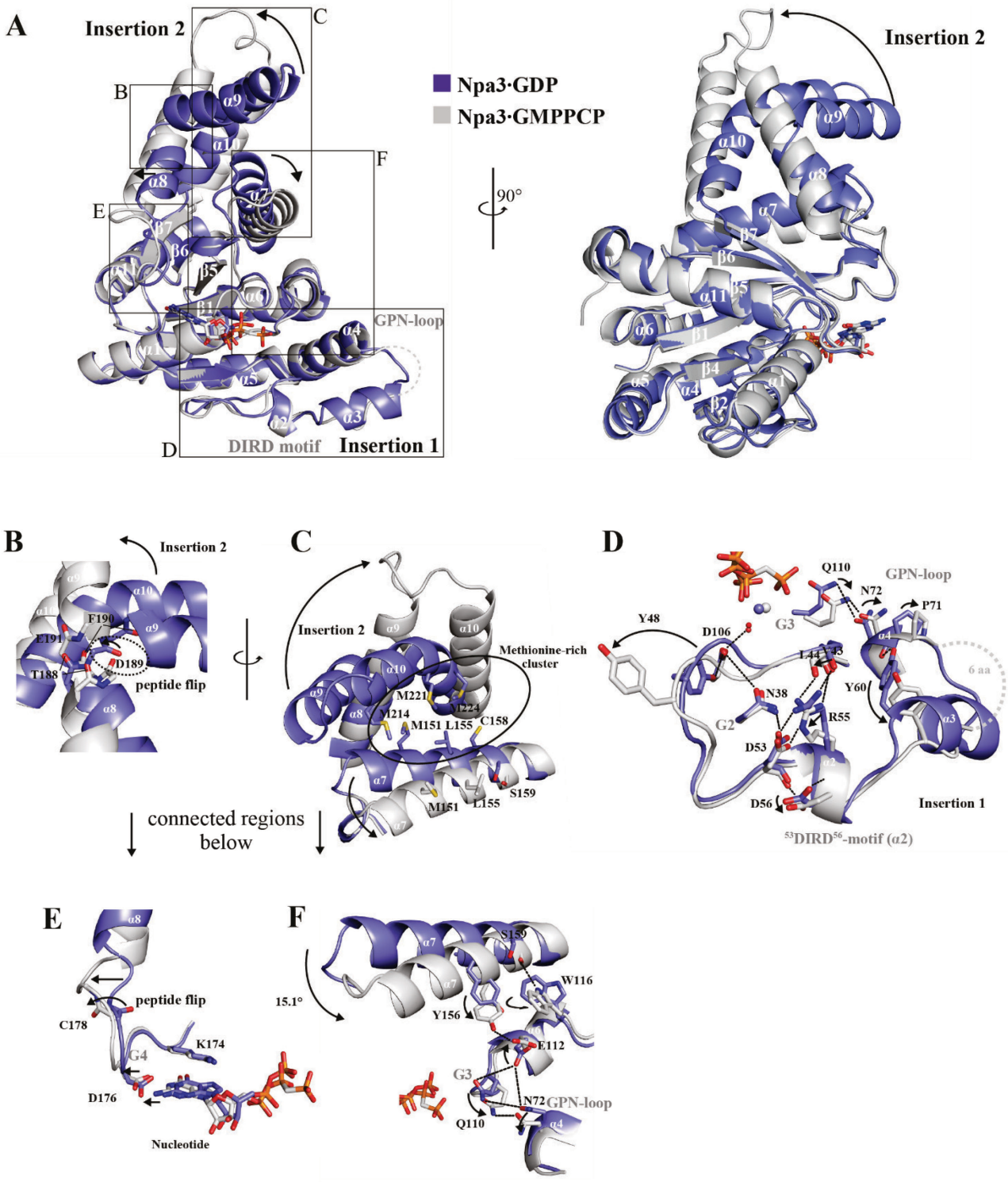


Figure 8| Superposition of closed Npa3·GDP (blue) and open Npa3·GMPPCP (grey) structures.
Magnesium ions are shown as blue (Npa3·GDP) or grey (Npa3·GMPPCP) spheres, water molecules as small red spheres and hydrogen bonds as dashed lines. **(A)** Superposition of overall structures from Npa3·GDP and Npa3·GMPPCP shown as front view (left) and side view (right). Significant differences between the two structures are indicated with black arrows and boxes highlight areas shown in B-F **(B)** Peptide flip of D189 enables pocket opening and the formation of a single helix from helices α8 and α9. **(C)** Conformational changes in insertion 2 and helix α7 facilitate the opening of an extended, hydrophobic pocket. **(D)** A set of residues in insertion 1 rearranges, including the GPN-loop and a DIRD-motif leading to increased flexibility of this region in the GMPPCP-bound state. **(E)** Pocket opening allosterically alters the active site via the G4 motif. **(F)** Conformational changes in helix α7 are linked to the G3 motif and the GPN-loop.

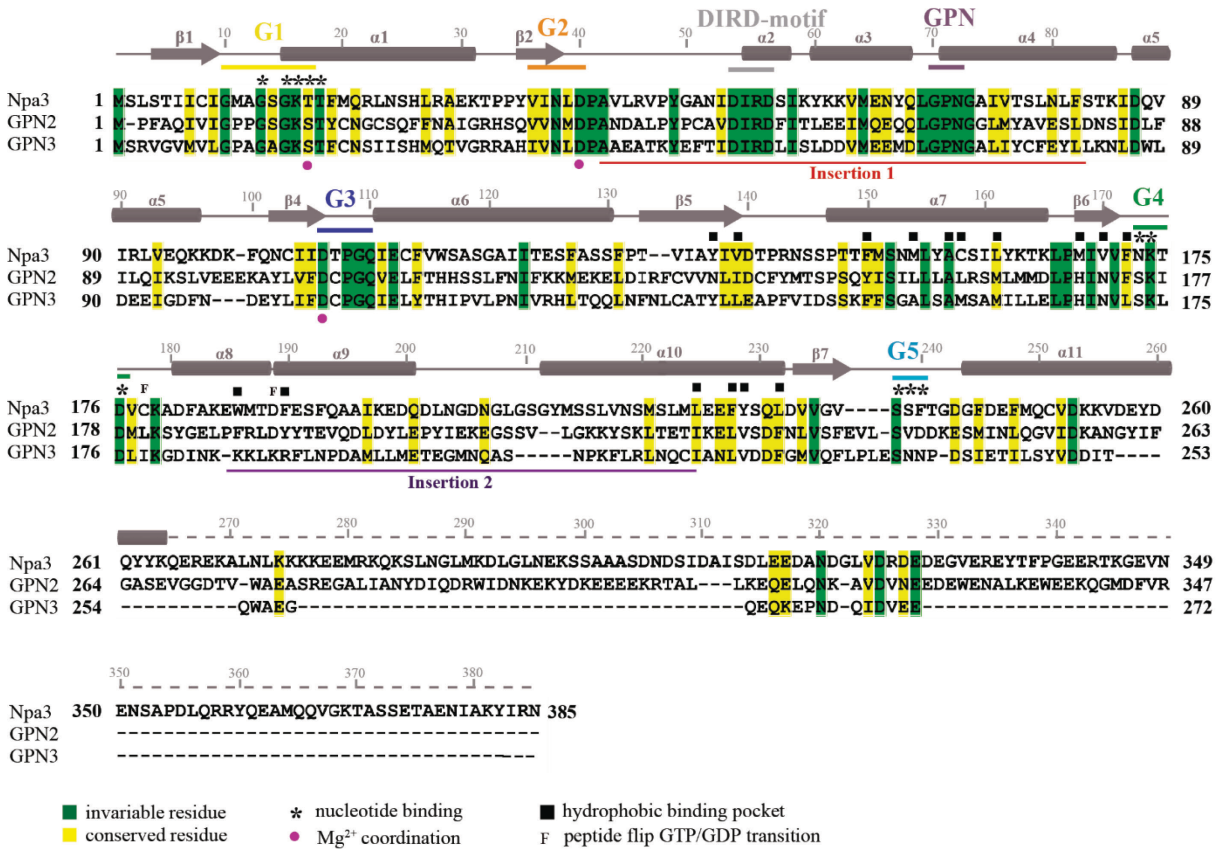


Figure 9| Amino acid sequence alignment of GPN-loop GTPases Npa3 from *S. cerevisiae*, and its paralogs GPN2 and GPN3. Secondary structure elements are indicated above the sequence (cylinders: α-helices; arrows: β-strands; dashed line: no structural data available). Numbering above the sequence corresponds to Npa3. Invariant residues are in green whereas conserved residues are in yellow. The G-domains G1-G5 and insertion regions are marked with bars, residues of the hydrophobic pocket with black squares, residues involved in nucleotide binding with asterisks and residues involved in magnesium binding are marked with pink spheres.

3.1.5. Nucleotide binding and conformational states

The two structures reveal atomic details of how Npa3 binds GMPPCP or GDP using its motifs G1-G5 (Figure 10). In the Npa3·GMPPCP structure, strong electron density corresponding to the γ-phosphate of the nucleotide is observed (Figure 10A). The octahedral coordination of the magnesium ion by oxygen atoms of the β- and γ-phosphate of GMPPCP and Thr17 of the G1 motif (P-loop, 10-GMAGSGKT-17) is completed by three water molecules, hydrogen bonded by D40 in the G2 motif (36-VINLD-40) and D106 and T107 in the G3 motif (106-DTPGQ-110) (Figure 10B). After GTP hydrolysis an additional water molecule apparently replaces the coordination sphere previously occupied by the γ-phosphate oxygen of GMPPCP (Figure 10C). Residues in the G1 motif stabilize the negative charge of the phosphate ions. Guanine specificity is conferred by hydrogen bonds of the Watson-Crick edge of GMPPCP to motifs G4 (173-NKTD-176) and G5 (238-SSF-240). F240 in motif G5 stacks against the guanine base. Thus our structures explain how Npa3 specifically binds GDP and the GTP analog.

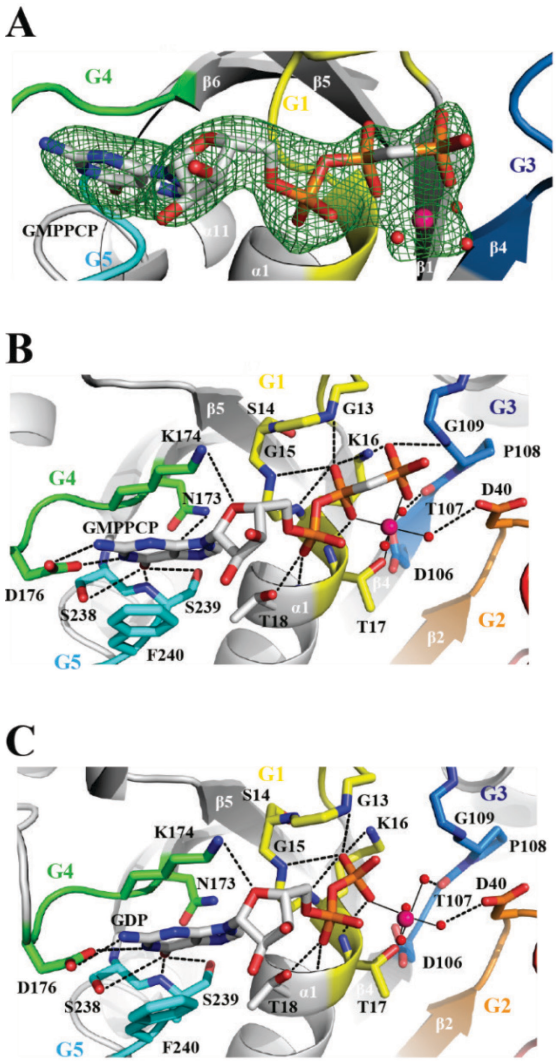


Figure 10| Nucleotide-binding pocket and active site.
(A) Initial unbiased F_o-F_c difference electron density of GMPPCP and the magnesium ion, including coordinated water molecules contoured at 3σ (green mesh). The final Npa3·GMPPCP model is superimposed. Motifs G1-G5 are color-coded as in Fig. 4A. Water molecules are shown as small red spheres, magnesium ions as pink spheres and hydrogen bonds as dashed lines. **(B-C)** Nucleotide interaction network of Npa3 with GMPPCP **(B)** and GDP **(C)**. Metal-ion-ligand interactions are shown as solid black lines.

3.1.6. Catalytic mechanism of Npa3

The catalytic mechanism for hydrolysis of GTP to GDP and inorganic phosphate by GPN-loop GTPases was suggested for the archaeal enzyme (Gras*, et al.*, 2007). To probe the catalytic mechanism in the eukaryotic system we prepared Npa3 variants with point mutations and tested their ability to hydrolyze GTP (Figure 11A). Whereas residue D106 in motif G3 stabilizes the Mg2+ ion, residue D40 in motif G2 positions the nucleophilic water molecule (N) (Figure 11B), as observed for the bacterial GTPase FtsY (Voigts-Hoffmann *et al.*, 2013). Consistent with this, the D106A and D40A variants of Npa3 were both inactive (Figure 11A).

The catalytic mechanism was suggested to also involve the GPN-loop. In the archaeal enzyme dimer, the GPN-loop of one monomer protrudes into the active site of the other monomer, where it binds the hydrolyzed GTP γ-phosphate (Gras*, et al.*, 2007) (Figure 3). Indeed, Npa3 variants with a mutation of the GPN-loop (^{70}GPN^{72} changed to ^{70}AAA^{72}) lacked GTPase activity, supporting the suggested mechanism (Figure 11). In addition, a single Q110L mutation, predicted to disrupt buttressing of the GPN-loop by residue Q110 in motif G3, was also inactive. All mutated residues shown here to be involved in catalysis *in vitro* are essential *in vivo* (Forget, et al., 2010; Staresincic, et al., 2011). Further, purified Npa3 formed a dimer, and in the Npa3·GDP crystals, two symmetry-related complexes formed a dimer that contained the GPN-loop of one monomer in the active site of the other monomer, as observed in the archaeal structure.

From these results emerges the catalytic mechanism of GPN-loop GTPases. The nucleophilic water molecule (N) is positioned in-line of the scissile phosphodiester bond at the γ-phosphate of GTP by residue D40 and a buttressing water molecule bound to the backbone of residue G109 of motif G3, which stabilizes the GPN-loop via residue Q110 and positions it in the active site of monomer B. The negative charge in the transition state of the S_N2 reaction is partially neutralized by a magnesium ion that is positioned also by D106 in motif G3. Modeling of the Npa3 dimer, based on the archaeal dimer structure, shows that the GPN-loop of one monomer protrudes into the active site of the other monomer even in the open enzyme conformation, and could bind the hydrolyzed orthophosphate.

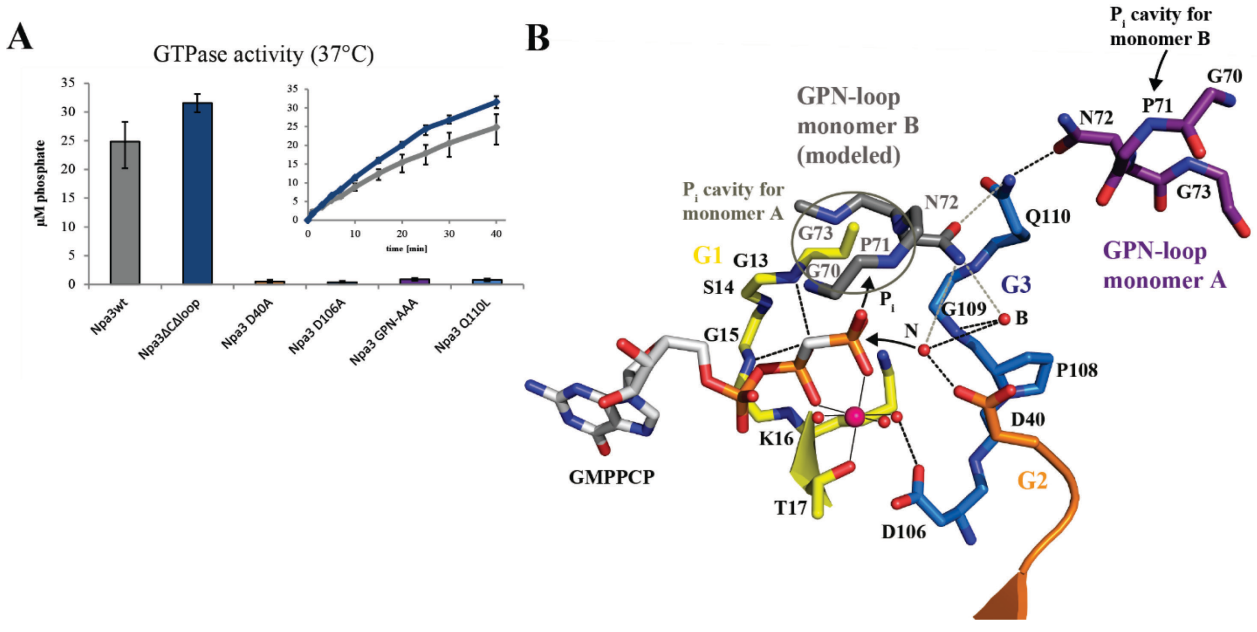


Figure 11| Catalytic mechanism of Npa3.
(**A**) GTPase activity of Npa3 mutants. Bars represent free orthophosphate concentrations after 40 min at 37°C. Kinetics are shown for wild-type Npa3 (grey) and the crystallized variant Npa3ΔCΔLoop (blue). (**B**) Schematic mechanism of GTP hydrolysis. The active site of Npa3·GMPPCP is shown. The GPN-loop of monomer B is modeled by superpositioning of two Npa3·GMPPCP enzymes on the archaeal Pab0955 dimer (PDB-code: 1YRB) (Gras*, et al.*, 2007). The nucleophilic water (N) that attacks the γ-phosphate and the buttressing water (B) are superimposed from the Npa3·GDP structure. Hydrogen bonds are shown as dashed black lines, potential hydrogen bonds derived from dimer modeling as dashed grey lines, water molecules as small red spheres, the magnesium ion as pink sphere and metal ion-ligand interactions as solid black lines.

3.1.7. A putative peptide binding pocket

In the Npa3·GMPPCP structure we observed a strong, extended electron density in the hydrophobic pocket that is created upon the transition from the closed to the open state of the enzyme (Figure 12A). This density could be explained by a molecule of lauric acid, a C12 fatty acid that may have been derived from hydrolysis of Tween20, a lauric acid ester that was present in the lysis buffer. The fatty acid carboxyl group binds residues N144 and W186, whereas the acid tail is close (within 5 Å) to the hydrophobic residues Y137, V139, F150, M154, A157, C158, L161, M168, V170, F172, W186, F190, L225, F228, Y229 and L232. The surface of the pocket is highly conserved between species (Figure 12B).

In the cellular context, it is likely that the newly observed hydrophobic pocket of Npa3 binds hydrophobic peptides. GPN1 is known to interact with assembling Pol II subunits (Boulon*, et al.*, 2010; Forget*, et al.*, 2010), which expose hydrophobic peptide regions, because large hydrophobic subunit interfaces are present in the functional Pol II assembly (Armache*, et al.*, 2005). We therefore considered that the hydrophobic pocket naturally binds to hydrophobic protein regions as described for molecular chaperones that prevent misassembly and aggregation of multisubunit complexes (Ellis, 2013; Kim*, et al.*, 2013). Indeed, modeling showed that hydrophobic peptides may be accommodated in the pocket in an extended conformation.

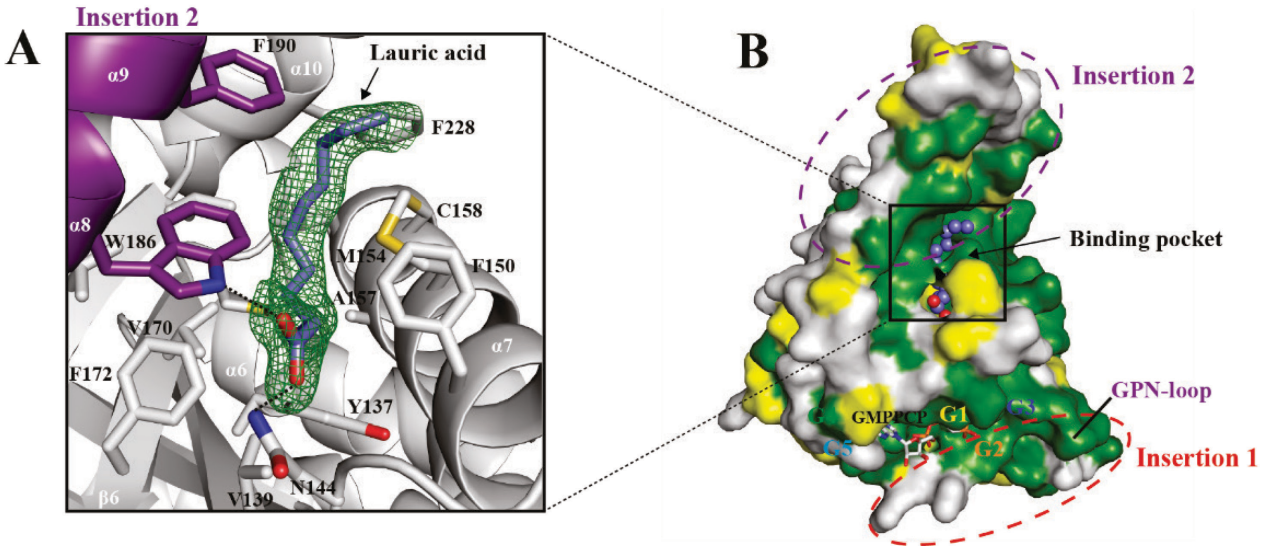


Figure 12| A highly conserved, putative peptide-binding pocket is exposed upon transition from the closed to the open state.
(A) Initial unbiased Fo-Fc difference electron density for lauric acid (slate blue sticks) bound to the putative peptide-binding pocket of Npa3, contoured at 2σ (green mesh). **(B)** Highly conserved surface of the putative peptide-binding pocket. Invariant residues are in green, conserved residues in yellow, and variable residues in grey. Insertion regions and G motifs are depicted. Lauric acid is shown as slate blue spheres.

3.1.8. Npa3 has GTPase stimulating chaperone activity

The existence of a putative hydrophobic peptide-binding pocket prompted us to ask whether Npa3 has chaperone activity. We tested for chaperone-like activity of Npa3 *in vitro* with a standard assay that uses citrate synthase (CS) as a general chaperone substrate protein, whose temperature-induced aggregation is suppressed when a chaperone is added (Buchner*, et al.*, 1998). Indeed, Npa3 was able to suppress temperature-induced aggregation of CS (Figure 13A). We next investigated whether this effect was nucleotide-dependent. Because no GTPase activity was observed in the standard chaperone assay buffer (not shown), we changed to SEC buffer 100 (Table 11) and adjusted the pH value to 7.5 at 43°C. This buffer supported GTP hydrolysis, albeit at about 5-fold reduced levels, apparently due to the increased temperature of 43°C (compare wild-type Npa3 in Figure 11A and Figure 13C). Chaperone-like activity was readily observed in the presence of nucleotides, but no significant differences occurred after the addition of GMPPCP, GTP, or GDP (Figure 13B).

These data indicated that the open Npa3 conformation is not induced by GTP binding but rather by binding of hydrophobic unfolded protein peptide regions. Pocket opening in Npa3 goes along with a conformational change in motifs G3 (Figure 8C and F) and G4 (Figure 8B and E) that widens the nucleotide-binding pocket. Indeed, a similar conformational change in Ffh from *T. aquaticus*, one of the closest folding homologs of Npa3, was proposed to allosterically alter the active site via the G4 motif (Freymann*, et al.*, 1997). Thus we speculated that peptide binding would allosterically widen the active center, may promote GDP displacement, and facilitate GTP binding and thus may lead to an increased GTP turnover rate.

To investigate this, we analyzed the GTPase activity of Npa3 both in the presence and in the absence of partially unfolded CS at 43°C. Indeed, the presence of CS stimulated GTPase activity more than four-fold, providing a link between chaperone and GTPase functions (Figure 13C).

Taken together, our results show that Npa3 exhibits chaperone-like activity and indicate that it can bind unfolded protein peptide regions without GTP addition. They also suggested that peptide binding triggers opening of the pocket, promotes GDP displacement and thus facilitates GTP rebinding, whereas peptide release is regulated by GTP hydrolysis.

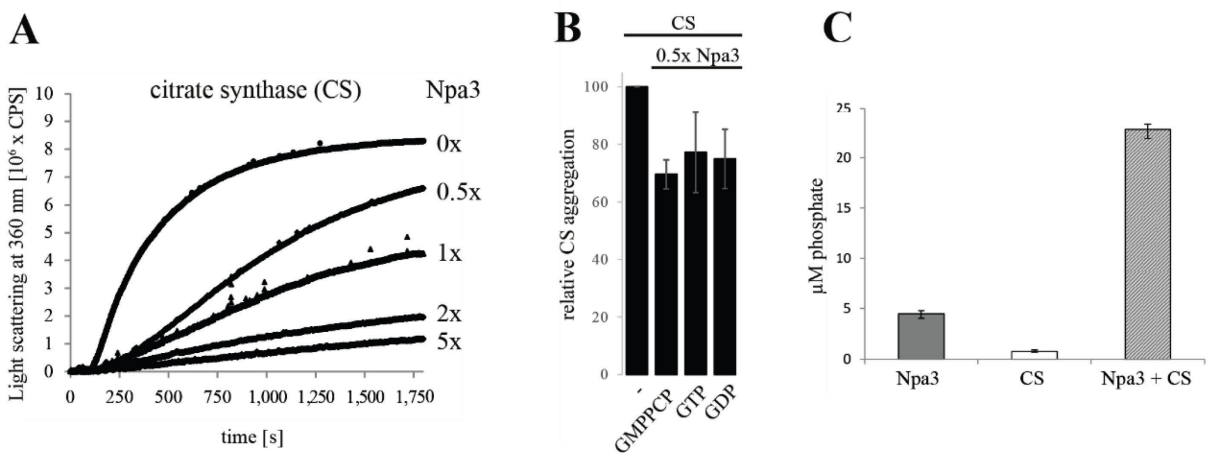


Figure 13| Npa3 has chaperone-like activity that stimulates GTPase activity.
(A) Npa3 has chaperone-like activity *in vitro*. Npa3 suppresses thermally induced (43°C) aggregation of the general non-native chaperone-substrate protein citrate synthase (CS). Different amounts of wild-type Npa3 were added as indicated on the right. **(B)** Chaperone-like activity of Npa3 is independent of added nucleotides under limiting conditions. 1 mM of GMPPCP, GTP, or GDP, were added and relative CS aggregation in the presence of 0.5 x molar amounts of Npa3 was determined after 30 min. **(C)** GTPase activity of Npa3 is stimulated > 4-fold in the presence of the non-native chaperone-substrate protein citrate synthase. Free orthophosphate concentrations were determined after 40 min at 43°C (compare Methods).

3.2. Analysis of Npa3 interactions with RNA polymerase II

3.2.1. Npa3 does not interact with complete assembled Pol II complexes

It is not clear whether Npa3 functions in assembly or nuclear import of Pol II (Wild & Cramer, 2012). Binding of Npa3/GPN1 to Pol II subunits and subassemblies has been shown (Jeronimo*, et al.*, 2007; Boulon*, et al.*, 2010; Forget*, et al.*, 2010; Carre & Shiekhattar, 2011; Staresincic*, et al.*, 2011) but it's unknown whether it also binds the mature, complete assembled polymerase. Because only complete assembled Pol II is imported into the nucleus (Boulon*, et al.*, 2010; Czeko*, et al.*, 2011), a nuclear import factor likely interacts with the assembled Pol II complex. To investigate this, we tested for Npa3 binding to complete assembled, purified, endogenous Pol II from *S. cerevisiae* with both analytical gel filtration and pull-down experiments (Figure 14). Npa3 did neither interact with assembled, mature Pol II (Figure 14A and B), nor with a Pol II complex bound to the nuclear import factor Iwr1 (Figure 14B).

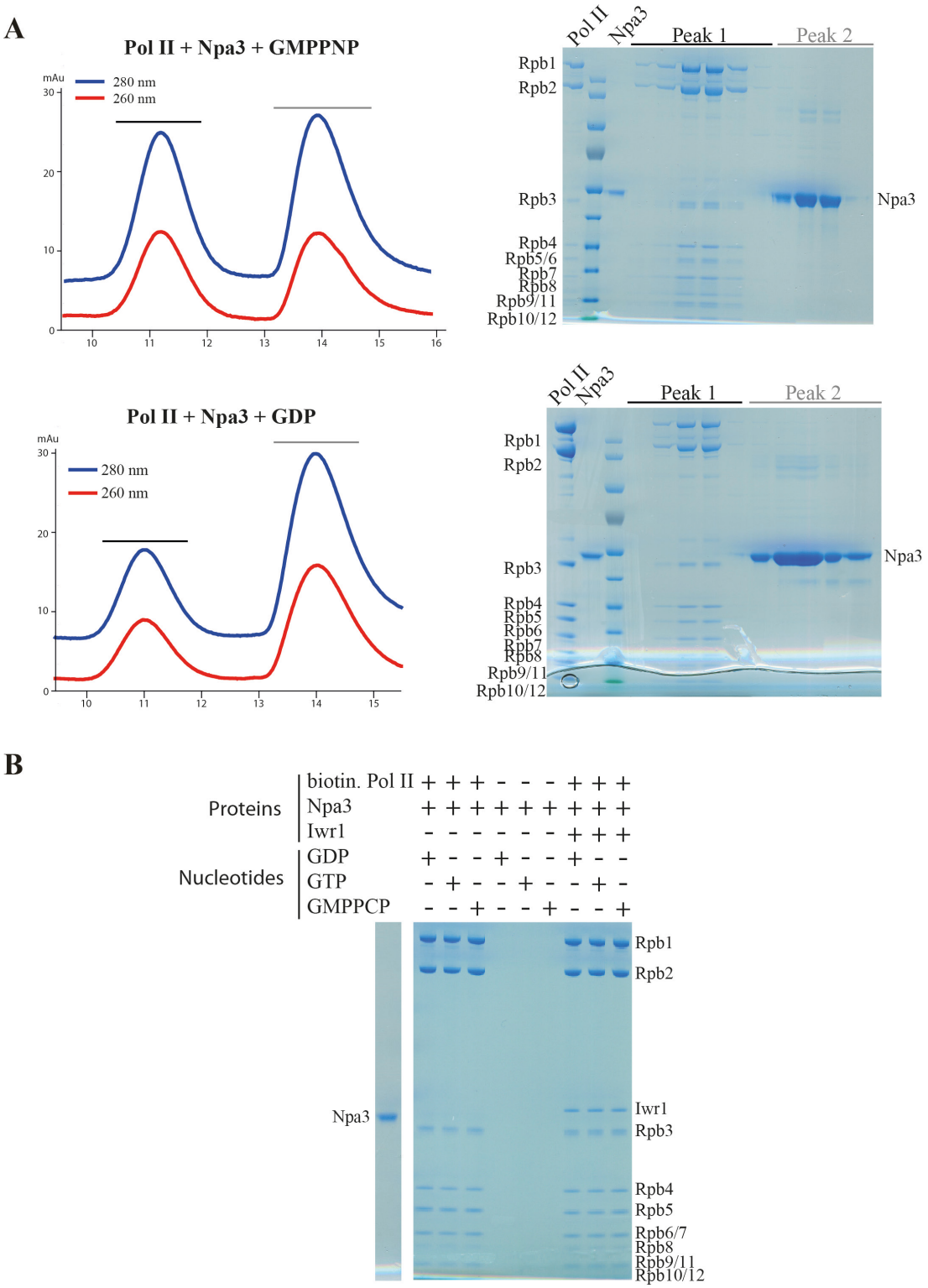


Figure 14| Npa3 does neither interact with assembled Pol II nor with a Pol II-Iwr1 complex.
(A) Analytical gel filtration (Superose 6 10/300, GE healthcare) of Pol II and Npa3-nucleotide complexes. Purified Pol II (12 µg), 5x molar excess of purified Npa3 and 1 mM GMPPNP (upper panel) or 1 mM GDP (lower panel) were incubated overnight at 4°C for nucleotide exchange and potential complex assembly. After gel filtration the fractions were TCA precipitated and analyzed by SDS-PAGE. Marker: (top to bottom: 170, 130, 100, 70, 55, 40, 35, 25, 15, 10 kDa) **(B)** Streptavidin pull-down of biotinylated Pol II and associated factors. Purified Pol II was biotinylated at Rpb3 and 4.2 µg Pol II were incubated with 15x molar excess of Npa3 and 1 mM GMPPCP or GDP overnight at 8°C or with 1 mM GTP for 30 min at RT and 600 rpm. Pol II-Iwr1 complexes were preformed by incubation with 5x molar excess of Iwr1 over Pol II at RT for 45 min before Npa3 was added.

3.2.2. Npa3 binds peptides derived from hydrophobic Pol II subunit interfaces

Because Npa3 was implicated in biogenesis of Pol II, we investigated whether Npa3 may act as an assembly chaperone for Pol II. Unfortunately, this could not be tested directly, because Pol II is not available in recombinant form and endogenous Pol II is not amenable to chaperone assays. We could however ask whether Npa3 would be able to bind peptides derived from Pol II subunits that may be exposed during Pol II assembly. To address this in an unbiased fashion, we used peptide arrays to screen a total of 1,139 Pol II-derived 15 residue long peptides (overlap of 11 residues, respectively; Table 20) covering all regions of all 12 Pol II subunits, for binding Npa3 in the presence of GMPPCP, GTP, or GDP. In this assay 55 peptides bound Npa3 significantly (signal intensity >3.5) (Table 19). The binding efficiency generally did not depend on nucleotides (Figure 16-Figure 20 and Figure 25), consistent with the model that opening of the hydrophobic pocket is triggered by peptide binding. The 55 Npa3-binding peptides stemmed from all Pol II subunits except Rpb3, Rpb6, and Rpb12. When we mapped the Npa3-binding peptides onto the Pol II structure (Figure 15), we found that 42 of the 55 peptides were at least partially located in interfaces between Pol II subunits and were enriched in hydrophobic residues (Table 19, Figure 16-Figure 20).

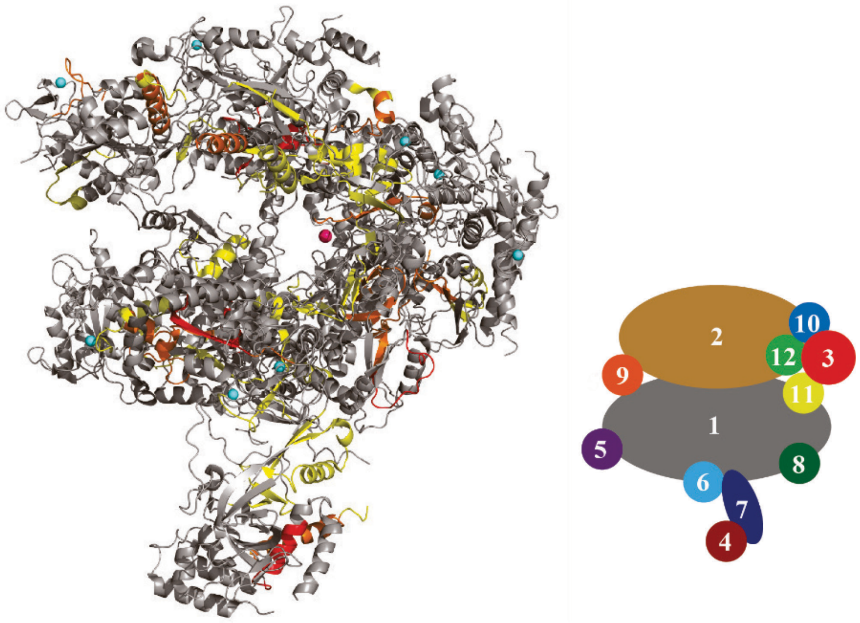


Figure 15| Location of Npa3-binding peptides in the assembled Pol II complex.
Npa3 binding to Pol II peptides is depicted in yellow (signal intensity 3.5-3.75), orange (3.75-4) and red (>4) whereas unbound regions are in grey (<3.5). Zinc atoms are shown as cyan spheres and the magnesium ion as pink sphere. A schematic representation of the 12 Pol II subunits Rpb1-Rpb12 in the folded Pol II complex is shown on the right. (pdb-code 1WCM) (Armache *et al.*, 2005).

Table 19| Npa3-binding Pol II-derived peptides

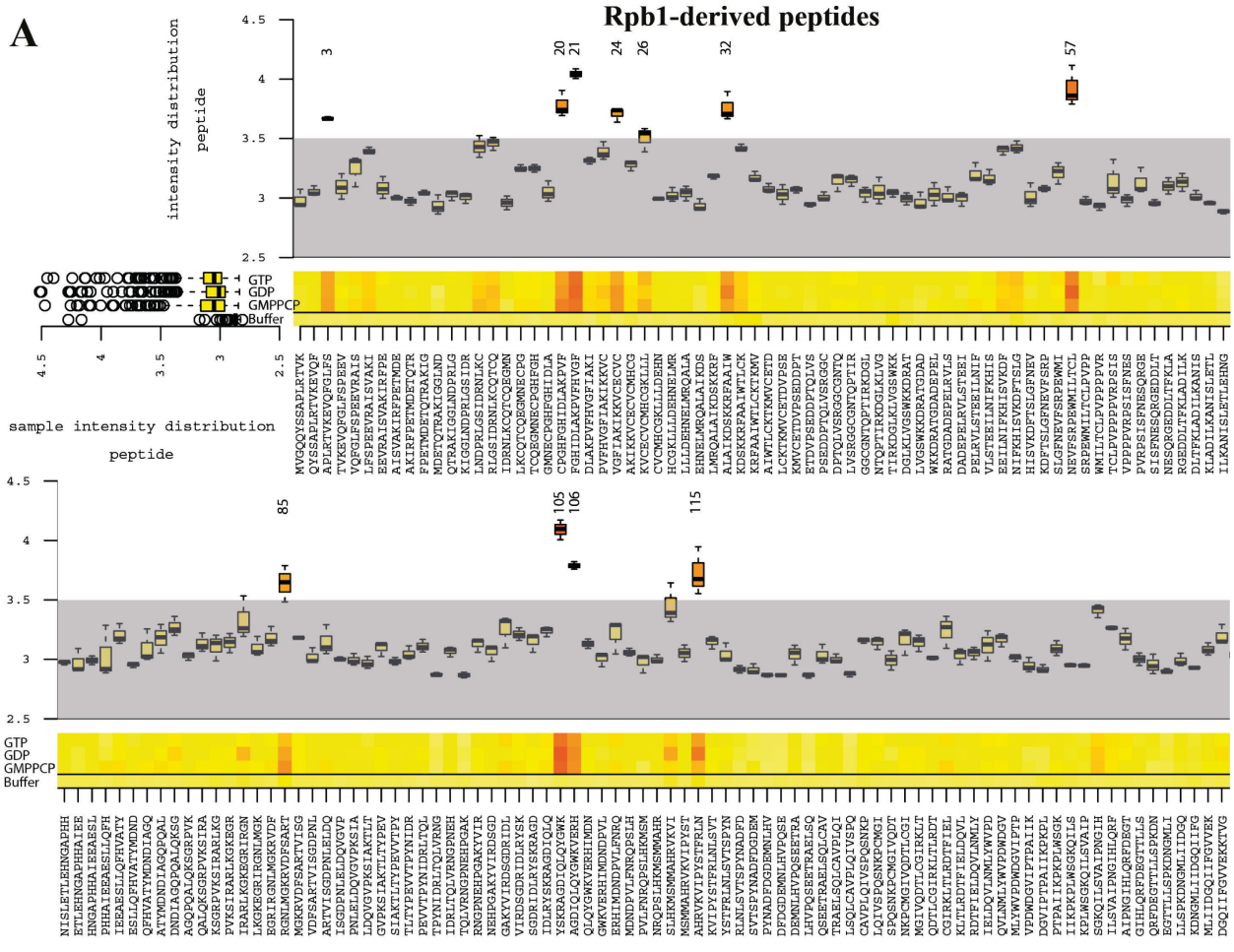
Pol II subunit	Peptide No	Start a.a.	End a.a.	Sequence	Interface
Rpb1	3	9	23	APLRTVKEVQFGLFS	Yes
	20	77	91	CPGHFGHIDLAKPVF	Yes
	21	81	95	FGHIDLAKPVFHVGF	Yes
	24	93	107	VGFIAKIKKVCECVC	Yes
	26	101	115	KVCECVCMHCGKLLL	No
	32	125	139	ALAIKDSKKRFAAIW	No
	57	225	239	NEVFSRPEWMILTCL	Yes
	85	337	351	RGNLMGKRVDFSART	Yes
	105	417	431	YSKRAGDIQLQYGWK	Yes
	106	421	435	AGDIQLQYGWKVERH	Yes
	115	457	471	AHRVKVIPYSTFRLN	Yes
	193	769	783	SVEGKRIAFGFVDRT	Yes
	194	773	787	KRIAFGFVDRTLPHF	Yes
	201	801	815	ENSYLRGLTPQEFFF	Yes
	234	933	947	YKQLVKDRKFLREVF	Yes
	261	1041	1055	AFDWVLSNIEAQFLR	Yes
	275	1097	1111	GVPRLKEILNVAKNM	No
	305	1217	1231	KQTFKNDLFVIWSED	No
	375	1497	1511	DSGSNDAMAGGFTAY	Lacking in 1WCM
Rpb2	463	116	130	EARLRNLTYSSGLFV	Yes
	464	120	134	RNLTYSSGLFVDVKK	No
	465	124	138	YSSGLFVDVKKRTYE	No
	481	188	202	DLYKLKECPFDMGGY	Yes
	482	192	206	LKECPFDMGGYFIIN	Yes
	525	364	378	ITQLEGFESRKAFFL	No
	527	372	386	SRKAFFLGYMINRLL	No
	539	420	434	LFKKLTKDIFRYMQR	No
	547	452	466	TITSGLKYALATGNW	No
	630	784	798	NYNVRMDTMANILYY	Yes
	645	844	858	SSIDRGLFRSLFFRS	Yes
	646	848	862	RGLFRSLFFRSYMDQ	Yes
	647	852	866	RSLFFRSYMDQEKKY	Yes
	730	1184	1198	GCDNKIDIYQIHIPY	Yes
Rpb4	833	54	68	EARLVIKEALVERRR	Yes
	834	58	72	VIKEALVERRRAFKR	Yes
	837	70	84	FKRSQKKHKKKHLKH	Yes
	855	142	156	KNTMQYLTNFSRFRD	Yes
Rpb5	892	69	83	ISKFPDMGSLWVEFC	No
	921	185	199	ALYLGLKRGEVVKII	No
Rpb7	970	11	25	ITLHPSFFGPRMKQY	Yes
	971	15	29	PSFFGPRMKQYLKTK	Yes
	975	31	45	LEEVEGSCTGKFGYI	Yes
	976	35	49	EGSCTGKFGYILCVL	Yes
	977	39	53	TGKFGYILCVLDYDN	Yes
	984	67	81	SAEFNVKYRAVVFKP	Yes
Rpb8	1036	104	118	FEEVSKDLIAVYYSF	Yes
	1037	108	122	SKDLIAVYYSFGGLL	Yes
	1038	112	126	IAVYYSFGGLLMRLE	Yes
	1039	116	130	YSFGGLLMRLEGNYR	Yes
	1043	132	146	LNNLKQENAYLLIRR	Yes
Rpb9	1047	2	16	TTFRFCRDCNNMLYP	Yes
Rpb11	1106	46	60	IRAELLNDRKVLFAA	Yes
	1108	54	68	RKVLFAAYKVEHPFF	Yes
	1109	58	72	FAAYKVEHPFFARFK	Yes
	1110	62	76	KVEHPFFARFKLRIQ	Yes

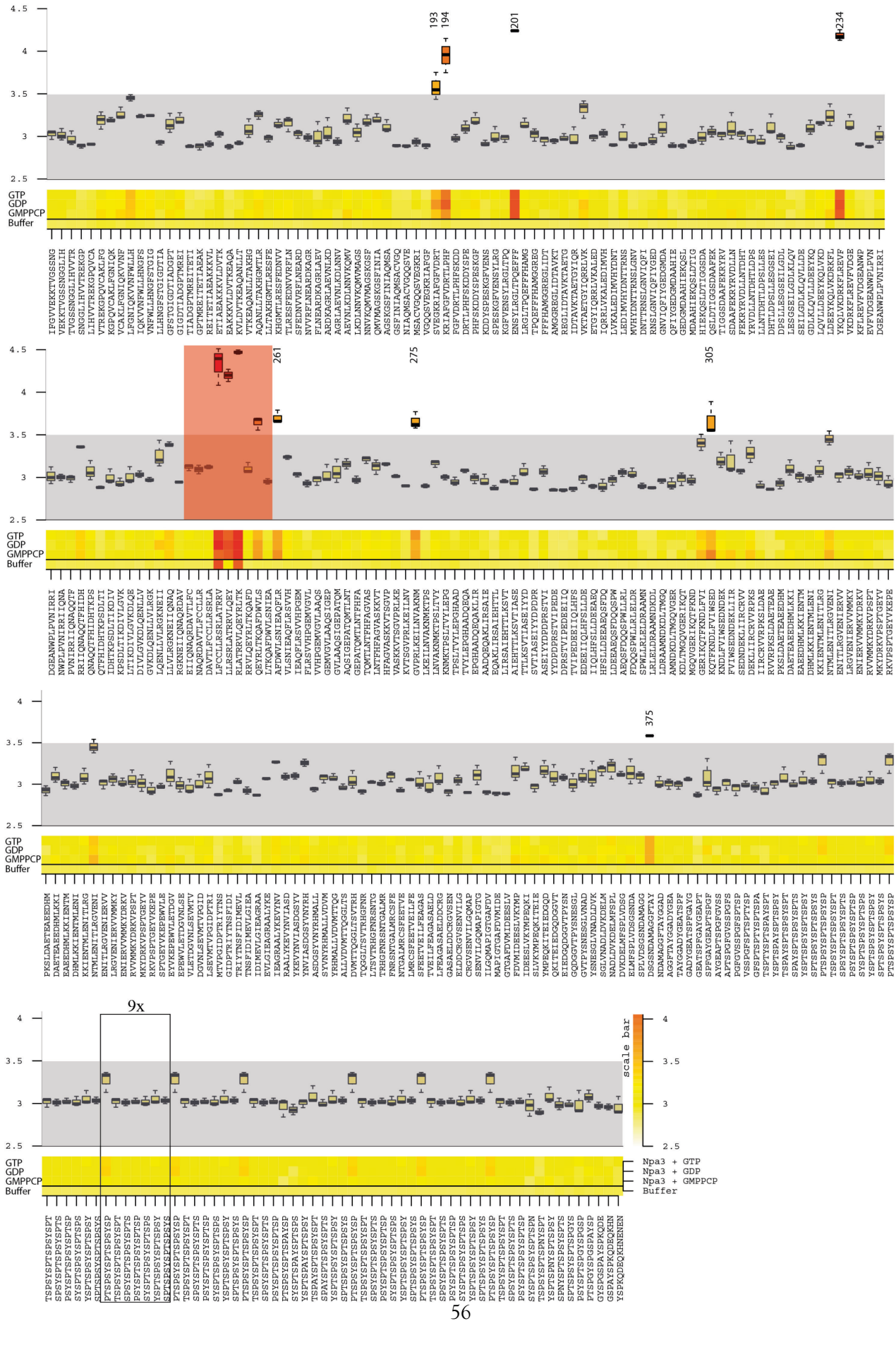
We therefore mapped residues in subunit interfaces in the mature Pol II complex using CoCoMaps (https://www.molnac.unisa.it/BioTools/cocomaps), and compared them with peptides that are bound by Npa3. In the following sections a detailed analysis of Npa3 binding peptides derived from Rpb1, Rpb8, Rpb11, Rpb4 and Rpb7 is provided.

3.2.3. Npa3 binds peptides derived from Rpb1 interfaces

We identified numerous Npa3-binding peptides within the interface between the two largest Pol II subunits, Rpb1 and Rpb2 (Figure 16). In particular, many peptides mapped to the extended Rpb1-Rpb2 interface within the clamp domain of Pol II (Rpb1 peptides 3, 20, 21, 24, 57, 85, and Rpb2 peptide 730). We also detected three Npa3-binding peptides (193, 194, 201) in the funnel domain of Rpb1.

Npa3 interacted also with peptides from subunits Rpb8, Rpb9 and Rpb11 that are located in critical interfaces with Rpb1.




4.5
4
3.5
3
2.5
193
194
201
234
GTP
GDP
GMPPCP
Buffer
261
275
305
375
9x
scale bar
Npa3 + GTP
Npa3 + GDP
Npa3 + GMPPCP
Buffer

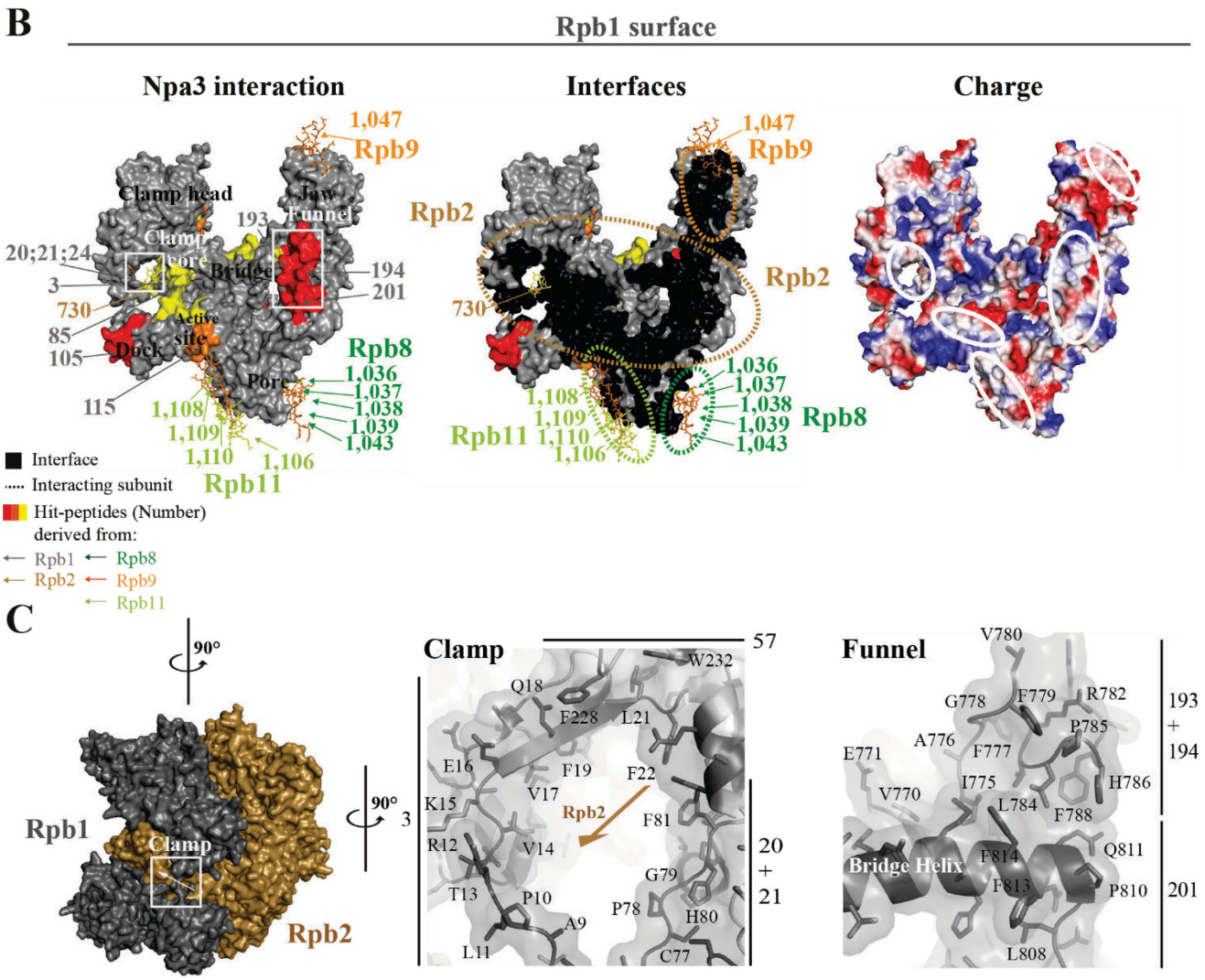


Figure 16| Npa3 binds peptides derived from Rpb1 interfaces.
(**A**) Boxplot representation of the heatmap describing the Npa3 peptide-binding landscape of Rpb1-derived peptides in the presence of GTP, GDP and GMPPCP, respectively. Control experiments were performed without Npa3 and nucleotides to test cross-reactivity of the anti-His antibody. Intensity distribution is shown in logarithmic scale. Peptides with signal intensity <3.5 were defined as unbound (grey area). Red boxes indicate false-positive binding of the antibody to the peptides, either directly or in overlapping regions. These peptides were not used for further analysis. The sample intensity distribution corresponds to all Rpb1-derived peptides. Peptide numbers are shown for Npa3-binding peptides above the boxplots. Box labeled with 9x shows a representative portion of nine identical regions of CTD peptides **(B)** Surface representations of Rpb1 show that Npa3 interacts with hydrophobic peptides located at Rpb1 interfaces. Numbers correspond to the peptide numbers from the array (see Table 19 for Npa3-binding peptides and Table 20 for all tested peptides). Left panel: Npa3 binding to Rpb1-peptides is depicted in yellow (signal intensity 3.5-3.75), orange (3.75-4) and red (>4) whereas unbound regions are in grey (<3.5). Npa3 binding peptides derived from other Pol II subunits that form interfaces with Rpb1 are shown as sticks and colored according to their signal intensity as described (compare Figure 17, Figure 18 and Figure 25 for analysis of these subunits). Middle panel: Rpb1 interfaces are shown in black. Dashed lines indicate positions of interacting subunits in the assembled Pol II complex. Right panel: Surface charge of Rpb1. White solid lines highlight hydrophobic interface regions. **(C)** Rpb1-Rpb2 interfaces bound by Npa3 in the peptide array. Left panel: Interface between Rpb1 (grey) and Rpb2 (brown). The extended clamp interface is highlighted. The clamp and funnel domain is shown in the middle and right panel, respectively. Residues involved in the interface are shown as sticks and Npa3-binding interface residues are labeled. The corresponding peptide numbers are shown at the sides. All structural figures were made using the Pol II structure pdb-code:1WCM (Armache, *et al.*, 2005).

3.2.4. Npa3 binds Rpb8-derived peptides at the interface to Rpb1

Three overlapping Npa3-binding peptides (peptides 1,037; 1,038; 1,039) contained a hydrophobic region in Rpb8 (117-SFGGLLMR-124) that forms an interface with Rpb1.

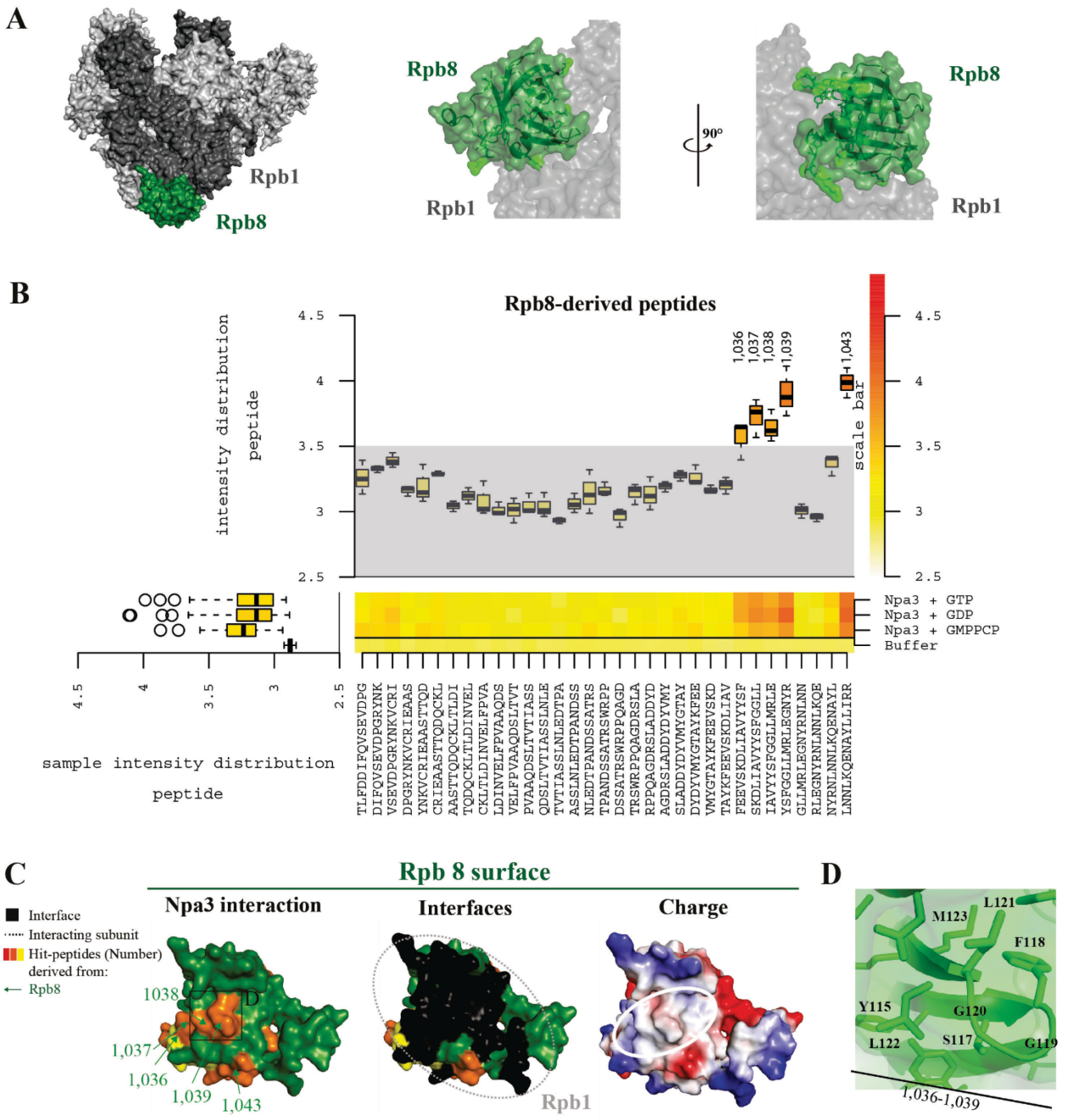


Figure 17| Npa3 binds Rpb8-derived peptides located at the subunit interface to Rpb1.
(A) Position of Rpb8 (green) and its interacting subunit Rpb1 (dark grey) in the assembled Pol II complex. (**B**) Boxplot representation of the heatmap describing Npa3 peptide-binding landscape of Rpb8-derived peptides in the presence of GTP, GDP and GMPPCP, respectively. Control experiments were performed without Npa3 and nucleotides to test cross-reactivity of the anti-His antibody. Intensity distribution is shown in logarithmic scale. Peptides with signal intensity <3.5 were defined as unbound (grey area). (**C**) Surface representations of Rpb8 show that Npa3 interacts with hydrophobic Rpb8-derived peptides located at the subunit interface to Rpb1. Left panel: Npa3 binding to Rpb8 peptides is depicted in yellow (signal intensity 3.5-3.75), orange (3.75-4) and red (>4) whereas unbound regions are in green (<3.5). Middle panel: Subunit interfaces are shown in black. Dashed lines indicate Rpb1 position in the assembled Pol II complex. Right panel: Surface charge of Rpb8. White solid lines highlight hydrophobic interface regions. Numbers correspond to the peptide numbers from the array (see Table 19 for Npa3-binding peptides and Table 20 for all tested peptides) **(D)** Putative Npa3 binding region of Rpb8. Residues involved in the interface are shown as sticks and Npa3 binding interface residues are labeled. All structural figures were made using the Pol II structure pdb-code:1WCM (Armache, *et al.*, 2005)

3.2.5. Npa3 binds Rpb11-derived peptides at the interface to Rpb1

In Rpb11, 18 of a total of 25 residues that form the interface with Rpb1 were found in overlapping Npa3-binding peptides (peptides 1,106;1,108-1,110). Peptide 1,109 covers most interface residues and shows the strongest Npa3 binding among Rpb11-derived peptides.

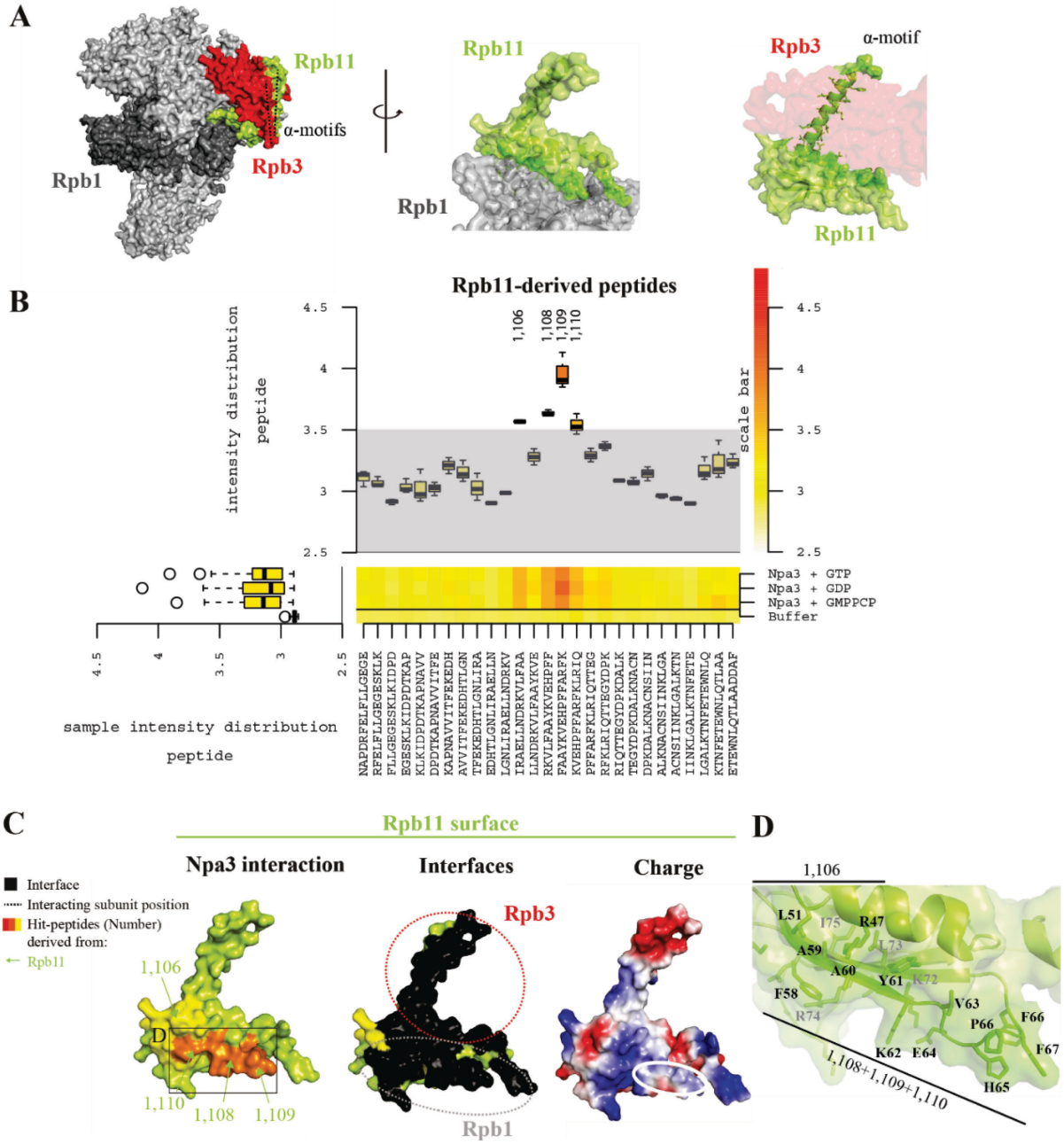


Figure 18| Npa3 binds Rpb11-derived peptides located at the subunit interface to Rpb1 but not to Rpb3.
(A) Position of Rpb11 (green) and its interacting subunits Rpb1 (dark grey) and Rpb3 (red) in the assembled Pol II. The Rpb1 and Rpb3 interface of Rpb11 is shown in the middle and right panel, respectively and Rpb11 interface residues are highlighted as dark green sticks. (**B**) Boxplot representation of the heatmap describing Npa3 peptide-binding landscape of Rpb11-derived peptides in the presence of GTP, GDP and GMPPCP, respectively. Control experiments were performed without Npa3 and nucleotides to test cross-reactivity of the anti-His antibody. Intensity distribution is shown in logarithmic scale. Peptides with signal intensity <3.5 were defined as unbound (grey area). (**C**) Surface representations of Rpb11 show that Npa3 interacts with hydrophobic Rpb11-derived peptides located at the subunit interface to Rpb1. Left panel: Npa3 binding to Rpb11 peptides is depicted in yellow (signal intensity 3.5-3.75), orange (3.75-4) and red (>4) whereas unbound regions are in green (<3.5). Middle panel: Subunit interfaces are shown in black. Dashed lines indicate subunit positions in the assembled Pol II complex. Right panel: Surface charge of Rpb11. White solid lines highlight hydrophobic interface regions. Numbers correspond to the peptide numbers from the array (Table 19 and Table 20) **(D)** Putative Npa3 binding region of Rpb11. Residues involved in the subunit interface are shown as sticks and Npa3 binding interface residues are labeled. All structural figures were made using the Pol II structure pdb-code:1WCM (Armache, *et al.*, 2005).

3.2.6. Npa3 binds Rpb4- and Rpb7-derived peptides at their subunit interface

We also identified interactions of Npa3 with ten peptides in the Rpb4-Rpb7 interface (Rpb4 peptides 833, 834, 837, 855; Rpb7 peptides 970, 971, 975, 976, 977, 984) (Figure 19 and Figure 20), consistent with a reported interaction of human GPN1 with Rpb4 and Rpb7 (Carre & Shiekhattar, 2011). Npa3 did not interact with a purified yeast Rpb4-Rpb7 complex (not shown), supporting the model that Npa3 binds peptide regions located at the subunit interface that are only accessible during subcomplex assembly.

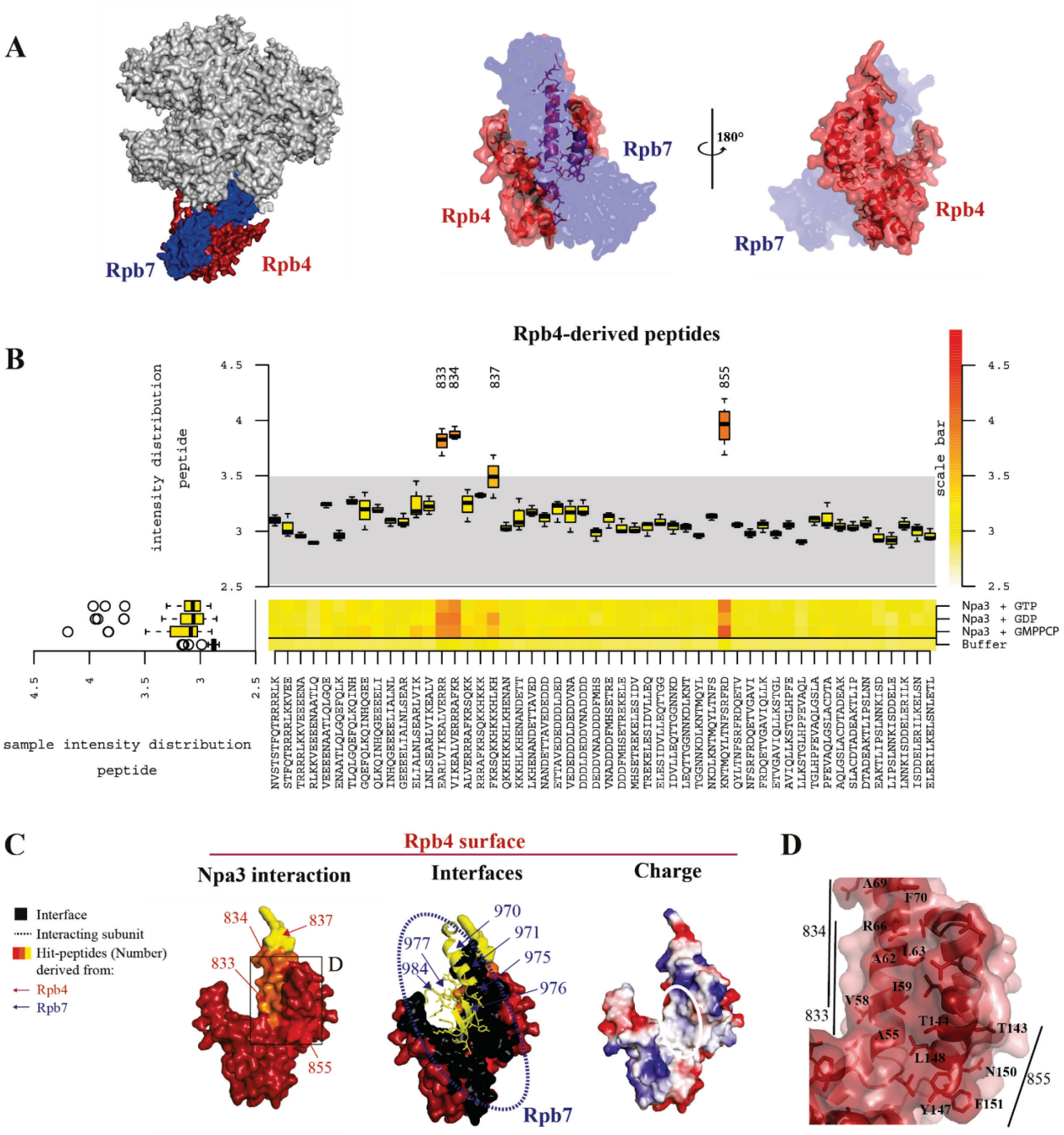


Figure 19| Npa3 binds Rpb4-derived peptides located at the subunit interface to Rpb7.
(A) Position of Rpb4 (dark red) and its interacting subunit Rpb7 (blue) in the assembled Pol II complex. The Rpb4/7 subcomplex is shown in the middle and right panel and Rpb4 interface residues are highlighted as dark red sticks. **(B)** Boxplot representation of the heatmap describing Npa3 peptide-binding landscape of Rpb4-derived peptides in the presence of GTP, GDP and GMPPCP, respectively. Control experiments were performed without Npa3 and nucleotides to test cross-reactivity

of the anti-His antibody. Intensity distribution is shown in logarithmic scale. Peptides with signal intensity <3.5 were defined as unbound (grey area). **(C)** Surface representations of Rpb4 show that Npa3 interacts with hydrophobic Rpb4 peptides located at the subunit interface to Rpb7. Left panel: Npa3 binding to Rpb4 peptides is depicted in yellow (signal intensity 3.5-3.75), orange (3.75-4) and red (>4) whereas unbound regions are in dark red (<3.5). Middle panel: Subunit interfaces of Rpb4 are shown in black. Npa3-binding Rpb7 peptides at the same interface are shown. Dashed lines indicate subunit positions in the assembled Pol II complex. Right panel: Surface charge of Rpb4. White solid lines highlight hydrophobic interface regions. Numbers correspond to the peptide numbers from the array (Table 19, Table 20). **(D)** Putative Npa3 binding region of Rpb4. Residues involved in the interface are shown as sticks and Npa3 binding interface residues are labeled. All structural figures were made using the Pol II structure pdb-code:1WCM (Armache*, et al.*, 2005).

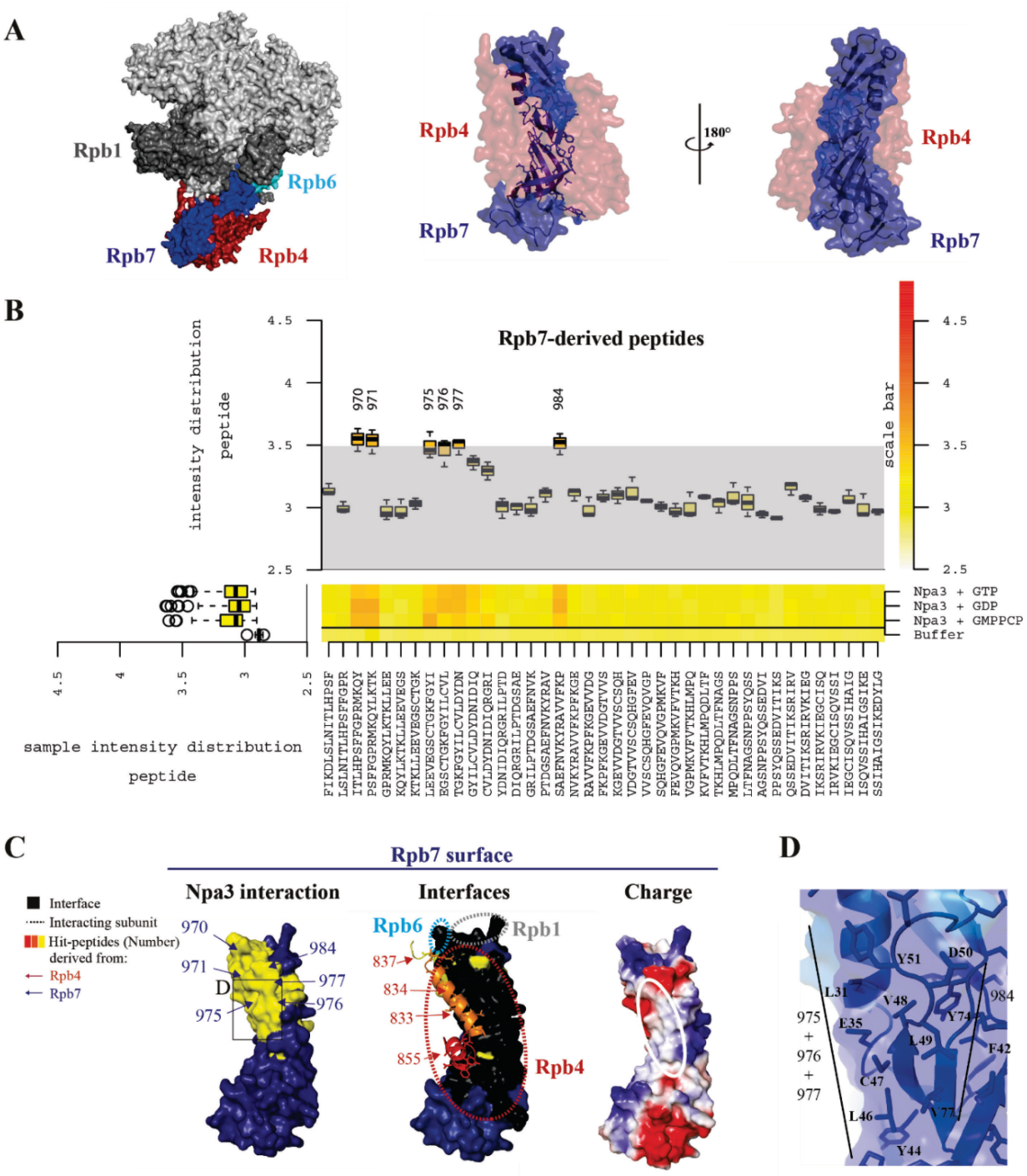


Figure 20| Npa3 binds Rpb7-derived peptides located at the subunit interface to Rpb4.
(A) Position of Rpb7 (blue) and its interacting subunits in the assembled Pol II complex. The Rpb4/7 subcomplex is shown in the middle and right panel and Rpb7 interface residues are highlighted as sticks. **(B)** Boxplot representation of the heatmap describing Npa3 peptide-binding landscape of Rpb7-derived peptides in the presence of GTP, GDP and GMPPCP, respectively. Control experiments were performed without Npa3 and nucleotides to test cross-reactivity of the anti-His antibody. Intensity

distribution is shown in logarithmic scale. Peptides with signal intensity <3.5 were defined as unbound (grey area). **(C)** Surface representations of Rpb7 show that Npa3 interacts with hydrophobic Rpb7 peptides located at the subunit interface to Rpb4. Left panel: Npa3 binding to Rpb4 peptides is depicted in yellow (signal intensity 3.5-3.75), orange (3.75-4) and red (>4) whereas unbound regions are in blue (<3.5). Middle panel: Subunit interfaces of Rpb7 are shown in black. Npa3 binding Rpb4 peptides at the same interface are shown. Dashed lines indicate subunit positions in the assembled Pol II complex. Right panel: Surface charge of Rpb7. White solid lines highlight hydrophobic interface regions. Numbers correspond to the peptide numbers from the array (Table 19, Table 20). **(D)** Putative Npa3 binding region of Rpb7. Residues involved in the interface are shown as sticks and Npa3 binding interface residues are labeled. All structural figures were made using the Pol II structure pdb-code: 1WCM (Armache*, et al.*, 2005).

Taken together, these data show that Npa3 binds numerous Pol II-derived hydrophobic peptides that are located at subunit interfaces, consistent with a function of Npa3 as an assembly chaperone for Pol II complex formation.

3.2.7. Discussion and model for RNA polymerase II biogenesis

Many macromolecular complexes were shown to require assembly chaperones for their biogenesis (Ellis, 2013), including the nucleosome (De Koning*, et al.*, 2007; Avvakumov*, et al.*, 2011), Rubisco (Liu*, et al.*, 2010), the proteasome (Murata*, et al.*, 2009), spliceosomal snRNPs (Chari*, et al.*, 2008), and the ribosome (Karbstein, 2010). In contrast, biogenesis of the 12-subunit Pol II complex remains poorly understood. It was shown that Pol II biogenesis requires the R2TP/Prefoldin-like complex and the ATPase Hsp90 (Boulon*, et al.*, 2010), a chaperone that is involved in the assembly of several protein complexes (Makhnevych & Houry, 2012), but additional assembly chaperones are likely required to enable correct Pol II assembly.

Here we provide evidence that the new family of GPN-loop GTPases are Pol II assembly chaperones. We present the first structure of a eukaryotic GPN-loop GTPase, the GPN1 homolog Npa3 from yeast. We show that Npa3 can adopt an open state with a hydrophobic pocket, that it can bind peptides derived from Pol II subunit interfaces, that it has chaperone activity, and that a chaperone substrate protein can stimulate its GTPase activity, which apparently triggers closing of the pocket. The latter observation is reminiscent to the reported stimulation of ATPase activity of chaperones Hsp70 (Swain*, et al.*, 2007; Smock*, et al.*, 2010; Zhuravleva & Gierasch, 2011) and Hsp90 (McLaughlin*, et al.*, 2002) by substrate binding. GTPases were also shown to play a key role during ribosome assembly in bacteria (Britton, 2009) and in eukaryotes (Kressler *et al.*, 2010).

In the light of published data, our results close a gap in understanding Pol II biogenesis. In our model, GPN-loop GTPases first enable correct Pol II assembly in the cytoplasm (Figure 21). Assembled Pol II would then be recognized by Iwr1 and imported into the nucleus (Czeko*, et al.*, 2011). Several lines of evidence argue for a role of GPN-loop GTPases in Pol II assembly rather than nuclear import of Pol II. First, GPN1, GPN2 and GPN3 interact with Pol II assembly intermediates (Boulon*, et al.*, 2010). Second, GPN1 interacts with the CCT complex (Forget*, et al.*, 2010), a chaperone complex with various functions

(Leroux & Hartl, 2000) that interacts with Pol II subunits (Dekker*, et al.*, 2008). Third, GPN-loop GTPases lack a nuclear localization signal (NLS), and mutations of GPN2 or GPN3 cannot be rescued by fusion of a NLS to Rpb3, whereas deletion of Iwr1 is partially rescued (Minaker*, et al.*, 2013). Forth, in patients with myofibrillar myopathies, a neuromuscular disorder characterized by protein aggregates, human GPN1 shows increased expression and accumulates with Rpb1 in the cytoplasm of muscle cells (Guglielmi*, et al.*, 2015), consistent with a Pol II-specific chaperone function of GPN1. However, Npa3 may additionally play nuclear roles, because nucleocytoplasmic shuttling of GPN1/Npa3 has been reported (Forget*, et al.*, 2010; Staresincic*, et al.*, 2011; Reyes-Pardo*, et al.*, 2012; Forget*, et al.*, 2013).

From these results emerges the molecular basis of GPN-loop GTPase function. In our model (Figure 21), an exposed hydrophobic peptide region in a newly synthesized Pol II subunit triggers the opening of the Npa3 pocket in its GDP-bound state. Npa3 then traps exposed hydrophobic regions of Pol II subunits, preventing their misassembly and opening a time window for association with the cognate Pol II subunit. Peptide binding allosterically alters the active site, decreasing its affinity for GDP, to provoke GDP displacement, and increasing its affinity for GTP, to facilitate GTP rebinding. Subsequent GTP hydrolysis would lead to release of the bound Pol II subunit, to enable association with cognate subunits and Pol II assembly. Assembled Pol II is then recognized by Iwr1, which binds between the two largest polymerase subunits (Czeko*, et al.*, 2011), and serves as an adaptor for import of Pol II into the nucleus.

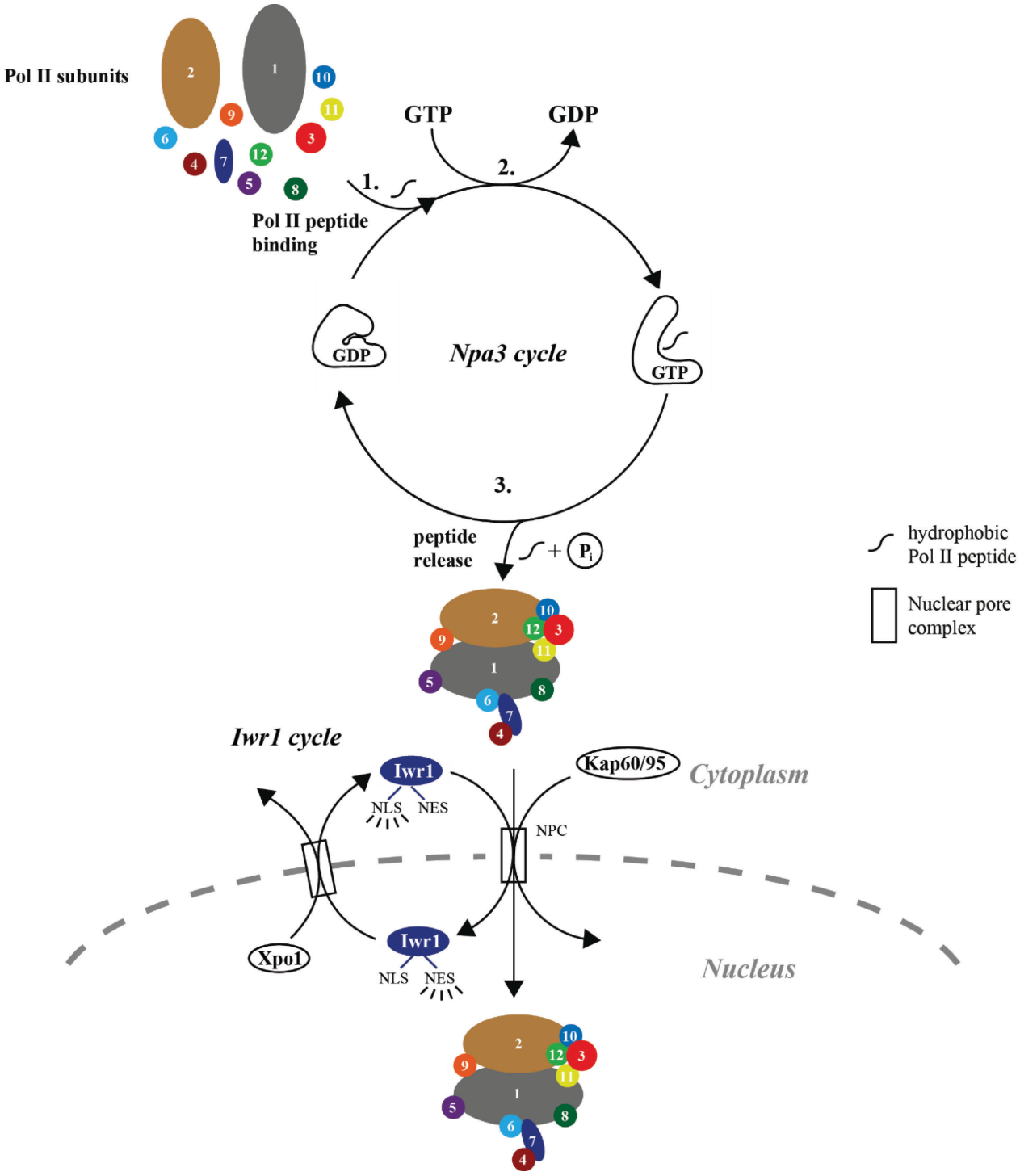


Figure 21| Model for RNA polymerase II biogenesis.
Whereas the 'Npa3 cycle' drives cytoplasmic assembly of Pol II, the 'Iwr1 cycle' drives Pol II nuclear import. In the Npa3 cycle, pocket opening of Npa3·GDP is induced by binding of hydrophobic regions of Pol II subunits that form interfaces in the assembled Pol II complex, thereby preventing misassembly (step 1). Pocket opening allosterically communicates with the active site, stimulates GDP displacement, and thereby facilitates GTP rebinding (step 2). GTP hydrolysis leads to release of Pol II peptides, facilitating formation of Pol II subunit interfaces and assembly of Pol II in the cytoplasm. In the Iwr1 cycle, assembled Pol II is recognized by Iwr1, which provides an import adaptor for nuclear import via its nuclear localization sequence (NLS). Iwr1 is recycled with the use of its nuclear export signal (NES).

3.3. Further analysis of Npa3 and Npa3/GPN2 complexes

Data presented in this chapter have been obtained during this thesis, but are not part of a publication.

3.3.1. Additional Npa3ΔCΔLoop crystallization conditions

In addition to the crystallization conditions that led to the described structures (3.1), other conditions were identified to promote crystal growth of Npa3ΔCΔLoop in complex with GDP or GMPPCP, respectively (Figure 22). Even though the crystals were very small and did not diffract synchrotron radiation, optimization of the conditions could may lead to bigger, diffracting crystals. The crystals showed different macroscopic shapes than the ones described and may serve as a starting point if a different crystal packing is desired. This could for instance be potentially interesting to analyze homodimeric Npa3 assemblies.

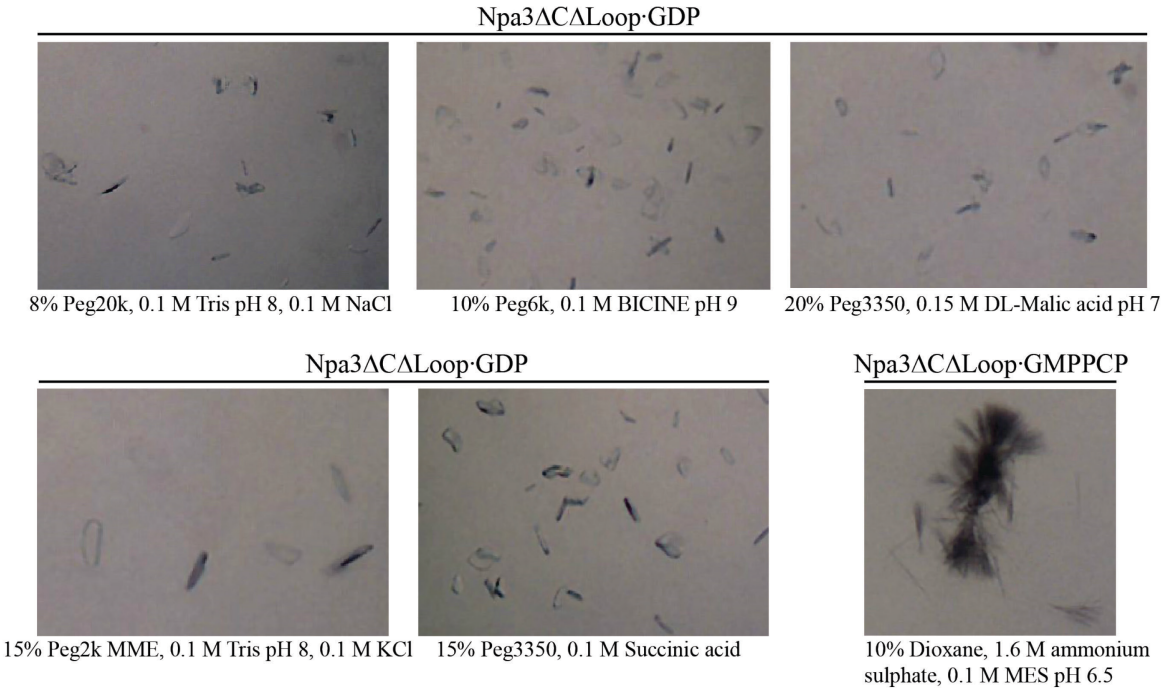


Figure 22| Additional Npa3ΔCΔLoop crystallization conditions.
Npa3ΔCΔLoop was purified as described for wild-type Npa3 (2.3.1) (lower salt concentration than in 'standard' protocol for this construct). This prevents crystallization in the 'standard' condition and crystal shape but allows only low protein concentrations (< 1mg/ml) due to stability problems. Npa3ΔCΔLoop was concentrated to 0.4 mg/ml and crystal were obtained by sitting drop vapor diffusion in 200 nl drops at 20°C in the indicated conditions.

3.3.2. Npa3 preferentially heterodimerizes with GPN2 and is required for its stable expression

Attempts to recombinantly express various GPN2 constructs from *S. cerevisiae* in *E. coli* did hardly yield any protein neither in the soluble nor in the insoluble fraction (not shown) indicating that GPN2 is degraded within the cells. Recent studies reported that GPN2 is not able to homodimerize but heterodimerizes with Npa3 (Minaker*, et al.*, 2013). We therefore co-transformed *E. coli* cells with

plasmids encoding both Npa3 without a tag and GPN2 with an N-terminal hexahistidine tag, respectively. Expression and subsequent Ni^{2+}-affinity purification isolated stoichiometric Npa3/GPN2 complexes that could be purified to homogeneity (Figure 23) with yields, equivalent to those of Npa3 alone (approx. 100 mg per 4 liter *E. coli* culture). Notably, co-expression of N-terminal hexahistidine tagged Npa3 with untagged GPN2 also led to almost stoichiometric Npa3/GPN2 complexes (not shown), indicating that Npa3 has higher affinities for hetero- than for homodimerization. This is consistent with observations for GPN1 and GPN3 in mammalian cells (Mendez-Hernandez*, et al.*, 2014).

3.3.3. The Npa3/GPN2 complex does not interact with assembled Pol II

Interaction of GPN2 with Pol II subunits has been reported (Boulon*, et al.*, 2010; Forget*, et al.*, 2010). We thus investigated whether the Npa3/GPN2 complex interacts with mature, complete assembled, purified Pol II using both analytical gel filtration and pull-down experiments with biotinylated Pol II. As observed for Npa3 alone (Figure 14) the Npa3/GPN2 complex did not interact with assembled Pol II (Figure 23) indicating that GPN2 interaction is restricted to the assembling polymerase.

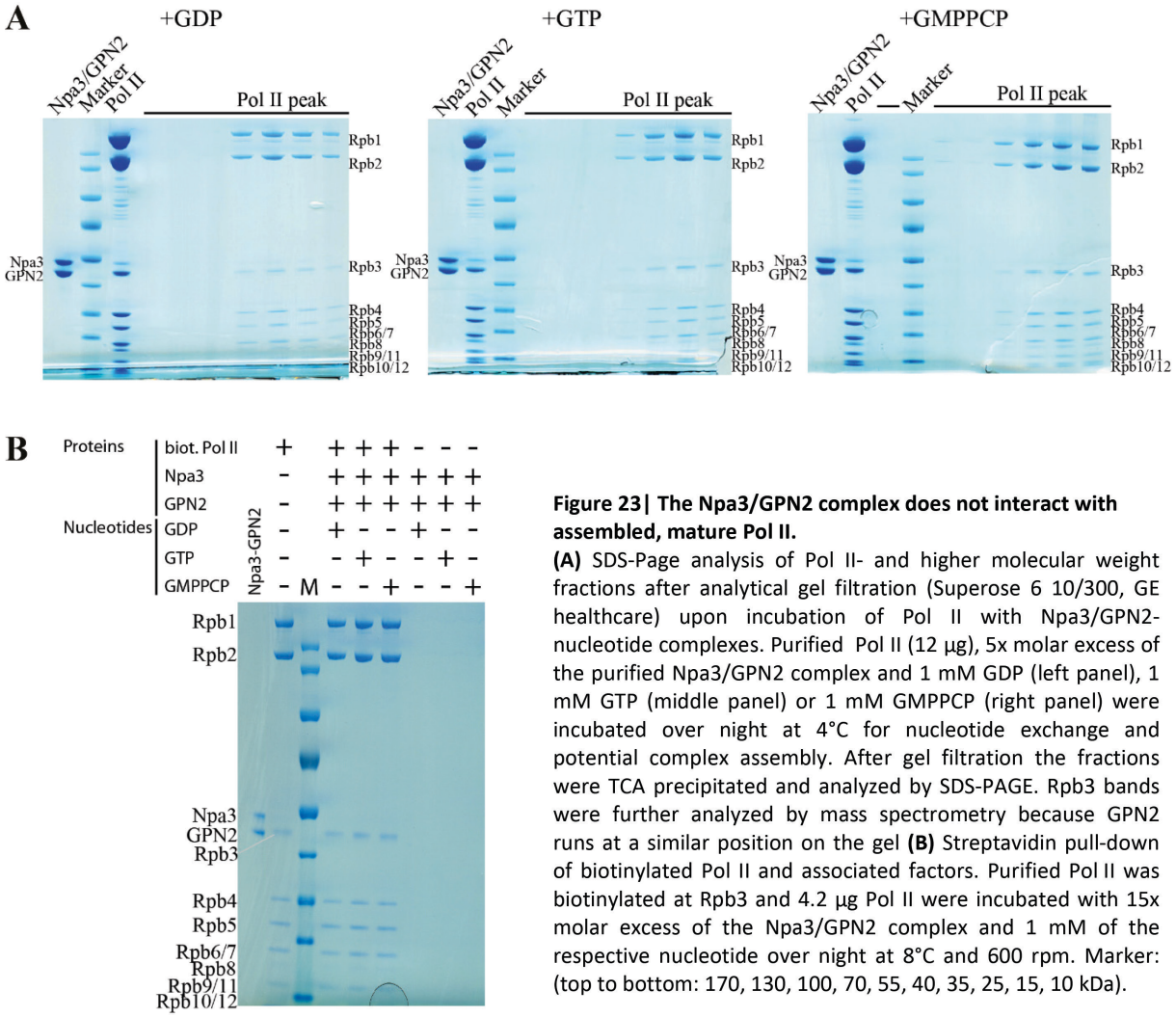


Figure 23| The Npa3/GPN2 complex does not interact with assembled, mature Pol II.
(A) SDS-Page analysis of Pol II- and higher molecular weight fractions after analytical gel filtration (Superose 6 10/300, GE healthcare) upon incubation of Pol II with Npa3/GPN2-nucleotide complexes. Purified Pol II (12 µg), 5x molar excess of the purified Npa3/GPN2 complex and 1 mM GDP (left panel), 1 mM GTP (middle panel) or 1 mM GMPPCP (right panel) were incubated over night at 4°C for nucleotide exchange and potential complex assembly. After gel filtration the fractions were TCA precipitated and analyzed by SDS-PAGE. Rpb3 bands were further analyzed by mass spectrometry because GPN2 runs at a similar position on the gel **(B)** Streptavidin pull-down of biotinylated Pol II and associated factors. Purified Pol II was biotinylated at Rpb3 and 4.2 µg Pol II were incubated with 15x molar excess of the Npa3/GPN2 complex and 1 mM of the respective nucleotide over night at 8°C and 600 rpm. Marker: (top to bottom: 170, 130, 100, 70, 55, 40, 35, 25, 15, 10 kDa).

3.3.4. The C-terminal tail is not required for Npa3/GPN2 complex formation

In order to investigate the function of the C-terminal tail in Npa3/GPN2 complex formation we performed co-expression experiments using both, full-length and C-terminal deletion variants of hexahistidine-tagged GPN2 and untagged Npa3. Therefore we prepared a GPN2 variant lacking both the C-terminal tail (residues 265-347) and part of an internal loop (residues 206-211), (GPN2ΔCΔLoop comprising residues 1-205 and 212-264), equivalent to the crystallized Npa3ΔCΔLoop construct. Here, co-expression of untagged, full-length Npa3 with tagged GPN2ΔCΔLoop and subsequent Ni^{2+}-affinity purification isolated stoichiometric heterodimers, indicating that the C-terminal tail of GPN2 is not required for complex formation (Figure 24A). Vice versa, co-expression of tagged, full-length GPN2 with untagged Npa3ΔCΔLoop also led to heterodimerization (Figure 24B). We then co-expressed both deletion variants, tagged GPN2ΔCΔLoop and untagged Npa3ΔCΔLoop (Figure 24C). Here, a single band was identified on the SDS gel due to the similar molecular weight of both constructs and analysis by mass-spectrometry revealed that both proteins co-eluted from the affinity column, indicating that none of the proteins requires the C-terminal tail for heterodimerization.

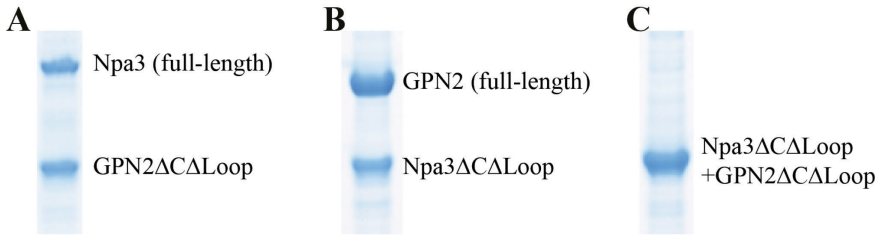


Figure 24| The C-terminal tail of both, Npa3 and GPN2 is not required for heterodimerization of the GPN-loop GTPases. SDS-PAGE analysis of Ni^{2+} affinity pull-downs after co-expression experiments using hexahistidine-tagged GPN2 variants and untagged Npa3 variants as described in the text. (A) Ni^{2+}-affinity pull-down of hexahistidine-tagged GPN2ΔCΔLoop (lacking the C-terminal tail and part of an internal loop) and full-length Npa3. (B) Ni^{2+}-affinity pull-down of hexahistidine-tagged full length GPN2 and Npa3ΔCΔLoop (lacking the C-terminal tail and part of an internal loop). (B) Ni^{2+}-affinity pull-down of hexahistidine-tagged GPN2ΔCΔLoop and untagged Npa3ΔCΔLoop (both lacking the equivalent parts of the C-terminal tail and an internal loop).

3.3.5. Crystallization trials of Npa3/GPN2 complexes

Because intense attempts to crystalize a full-length Npa3/GPN2 complex did not reveal crystals we tested various truncated constructs and combinations, including an Npa3ΔCΔLoop/GPN2ΔCΔLoop heterodimer (3.3.4). This complex represents the equivalent heterodimer to the crystallized Npa3 variant. However, crystallization was hampered also in conditions were Npa3ΔCΔLoop alone crystallized.

4. Conclusion and outlook

In this thesis a combination of X-ray crystallography, site-directed mutagenesis, enzymatic activity assays, chaperone assays, protein-protein interaction techniques and a systemic Pol II peptide interaction screen were used to investigate the structure and function of the GPN-loop GTPase Npa3. In previous studies, functional characterization of Npa3 or its human homolog GPN1 mainly focused on cellular localization of Pol II subunits (Wild & Cramer, 2012), but biochemical and structural data were lacking and the question whether Npa3 functions in assembly or nuclear import of Pol II remained unanswered. In this thesis we report the first structure of a eukaryotic GPN-loop GTPase, show that it can adopt a closed GDP-bound state that reveals eukaryote-specific features, and a novel open GTP-analog-bound state that exposes a conserved hydrophobic pocket. We further show that Npa3 has chaperone activity and that a chaperone substrate protein can stimulate its GTPase activity, which apparently triggers closing of the pocket. Using a systemic Pol II peptide interaction screen covering all regions of all 12 Pol II subunits we describe for the first time putative, particular interaction sites of a GPN-loop GTPase with Pol II subunits. We show that these regions are located at Pol II subunit interfaces and enriched in hydrophobic residues, consistent with a function of Npa3 as assembly chaperone. Such peptide arrays were successfully used in the past to characterize assembly chaperone interactions sites (Saschenbrecker*, et al.*, 2007). The 15-residue long peptides can form small secondary structures, mimic partially folded Pol II subunit regions, and allow accessibility to Pol II subunit interfaces that are hard to grasp with conventional methods and bound by assembly chaperones.

Our studies provide a framework for future validation and characterization of these interactions. Attempts to co-crystallize Npa3 with Rpb1- or Rpb11-derived peptides identified in our screen failed because they are hydrophobic and aggregated in the crystallization condition (not shown). Crystallographic studies of a fusion protein consisting of the crystallized Npa3 variant and a C-terminally fused Pol II peptide, separated by a short, flexible linker could solve the solubility problem and should allow peptide-binding to the nearby hydrophobic pocket. However, screening for new crystallization conditions may be required because crystal contacts separate the pocket from the C-terminus in the current crystal packing. Here, other crystallization conditions identified in this thesis to promote Npa3 crystals growth (3.3.1) could provide a promising starting point for the identification of a different crystal packing. Further, here identified Npa3 interactions with Pol II subunits could for instance be verified with yeast two-hybrid assays as done for other assembly chaperones (Mao*, et al.*, 2015) or recombinant co-expression experiments to additionally enable complex analysis. Structure-guided site-directed mutagenesis of the hydrophobic Npa3 pocket could be used and impaired subunit interactions, Pol II assembly defects and growth defects could be analyzed.

In addition, structural characterization of Npa3 in complex with Pol II subunits or subassemblies should be of major interest for future Pol II biogenesis research. Our model (Figure 21) shows that Pol II subunit release from Npa3 is regulated by GTP hydrolysis and our high resolution crystal structures in combination with site-directed mutagenesis provide atomic details of how to abolish both processes and trap the complexes. For this, yeast cells expressing both, a functional Npa3 variant for cell viability and a tagged GTP hydrolysis-deficient mutant variant should allow purification of Npa3-bound Pol II subcomplexes. Further urea- (Kimura*, et al.*, 1997) or α-amanitin (Boulon*, et al.*, 2010) treatment of Pol II may be considered for subcomplex production. Chemical crosslinking and mass-spectrometry in combination with our Npa3 and the Pol II (Cramer*, et al.*, 2000; Armache*, et al.*, 2005) crystals structures could now be used to determine the complex architecture.

Notably, no structure of any assembly factor in complex with Pol II subassemblies is available so far. Here, cryo-electron microscopy could be increasingly important in the future because recent advances in imaging hardware and processing software (Kuhlbrandt, 2014) may allow high resolution cryo-EM structures of Npa3-bound Pol II subassemblies and may also shed light into the structure and function of the C-terminal tail of Npa3 which is lacking in the crystal structure.

In this study we showed that Npa3 is a Pol II assembly chaperone but the structure and function of its paralogs GPN2 and GPN3 is still unknown. However, they likely have similar, non-redundant functions and thus Npa3 may represents the founding member of a new family of assembly chaperones. Heterodimerization of Npa3/GPN1 with both paralogs was reported (Minaker*, et al.*, 2013; Mendez-Hernandez*, et al.*, 2014) and likely needs to be considered to completely understand the function of Npa3. Interestingly, Npa3/GPN1 seems to preferentially heterodimerize with its paralogs because coexpression of human GPN1 and GPN3 (Mendez-Hernandez*, et al.*, 2014) and recombinant co-expression of yeast Npa3 and GPN2 (3.3.2) mainly leads to heterodimers. This, together with the proposed enzymatic mechanism that likely requires the 'trans'-GPN-loop of a GPN-loop GTPase dimer for GTP hydrolysis (3.1.6; (Gras*, et al.*, 2007)) and Pol II subunit release, may indicate an intriguing mechanism for Pol II assembly: Given that the other GPN-loop GTPases also function as assembly chaperones, they may bind Pol II regions at the opposite site of the interface than Npa3. In this case, GPN-loop GTPases could bring interacting subunits together by heterodimerization, orchestrate their assembly in a highly regulated manner and get released by GTP hydrolysis upon dimer formation to facilitate subunit association. To initially test this hypothesis, the Npa3/GPN2 complex could be used in our Pol II peptide array to analyze whether the presence of GPN2 leads to additional interactions on Pol II subunit regions that form interfaces with the ones identified for Npa3 binding. Further, this Npa3/GPN2 complex should be tested for chaperone activity, GPN2 hydrolysis-deficient mutants could be used for *in vivo* subunit pull-downs and the complex structures of GPN-loop GTPase heterodimers

should be investigated. Here, homology modeling of GPN2 and GPN3 with our Npa3 structure and analysis of their putative substrate binding pocket may give first insights into substrate specificities.

Over the last decades, all efforts to recombinantly express and *in-vitro* assemble Pol II failed making analysis of its assembly steps including associated factors potentially challenging. Here, the addition of GPN-loop GTPases could be a prerequisite for subunit stability and interface formation and a step towards *in vitro* assembly. This would in fact allow many new possibilities to identify and analyze assembly intermediates, characterize assembly factor functions, and determine the chronological order of subunit assembly and chaperone action. For instance, it remains elusive whether Npa3 already interacts with nascent chains or at a later state of Pol II subunit maturation.

Further, regulation of Pol II biogenesis should be investigated in more detail because Npa3 is phosphorylated by Pcl1-PHO85, a cyclin-CDK complex involved in cell cycle progression (Keniry *et al.*, 2004; Albuquerque *et al.*, 2008; Holt *et al.*, 2009), indicating that Pol II biogenesis is linked to the cell cycle. This may regulate transcription by controlling the amount of assembled, functional Pol II, dependent on the specific needs of the cell. Additionally Pol I and Pol III biogenesis may need to be considered since they share subunits with Pol II (Vannini & Cramer, 2012) and biogenesis of all three polymerases is may mutually dependent.

In this thesis we propose the cytoplasmic function of Npa3 but nucleocytoplasmic shuttling was reported (Forget*, et al.*, 2010; Staresincic*, et al.*, 2011; Forget*, et al.*, 2013) indicating additional nuclear functions. One option would be that Npa3 not only assists cytoplasmic Pol II assembly but also nuclear disassembly of defective Pol II. Further, nucleocytoplasmic recycling of Pol II subunits may couple disassembly and assembly as observed for Rpb3 (Boulon*, et al.*, 2010) and may be considered for Npa3 function. To investigate disassembly, degradation and recycling of Pol II subunits, α-amanitin, a transcription inhibitor that specifically leads to degradation of nuclear Rpb1 has been shown to be an effective tool (Nguyen*, et al.*, 1996; Jung & Lippard, 2006; Boulon*, et al.*, 2010) and could be used to analyze Npa3 function in this processes.

Taken together, the structural and functional data presented in this thesis provide novel insights into the role of GPN-loop GTPases in Pol II biogenesis and provide a framework for future analysis towards the understanding of this fundamental process.

5. Supplementary Information

Table 20| Pol II-derived peptides tested for Npa3 binding.

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence	Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	1	1	15	MVGQQYSSAPLRTVK	**Rpb1**	42	165	179	GGCGNTQPTIRKDGL
	2	5	19	QYSSAPLRTVKEVQF		43	169	183	NTQPTIRKDGLKLVG
	3	9	23	APLRTVKEVQFGLFS		44	173	187	TIRKDGLKLVGSWKK
	4	13	27	TVKEVQFGLFSPEEV		45	177	191	DGLKLVGSWKKDRAT
	5	17	31	VQFGLFSPEEVRAIS		46	181	195	LVGSWKKDRATGDAD
	6	21	35	LFSPEEVRAISVAKI		47	185	199	WKKDRATGDADEPEL
	7	25	39	EEVRAISVAKIRFPE		48	189	203	RATGDADEPELRVLS
	8	29	43	AISVAKIRFPETMDE		49	193	207	DADEPELRVLSTEEI
	9	33	47	AKIRFPETMDETQTR		50	197	211	PELRVLSTEEILNIF
	10	37	51	FPETMDETQTRAKIG		51	201	215	VLSTEEILNIFKHIS
	11	41	55	MDETQTRAKIGGLND		52	205	219	EEILNIFKHISVKDF
	12	45	59	QTRAKIGGLNDPRLG		53	209	223	NIFKHISVKDFTSLG
	13	49	63	KIGGLNDPRLGSIDR		54	213	227	HISVKDFTSLGFNEV
	14	53	67	LNDPRLGSIDRNLKC		55	217	231	KDFTSLGFNEVFSRP
	15	57	71	RLGSIDRNLKCQTCQ		56	221	235	SLGFNEVFSRPEWMI
	16	61	75	IDRNLKCQTCQEGMN		57	225	239	NEVFSRPEWMILTCL
	17	65	79	LKCQTCQEGMNECPG		58	229	243	SRPEWMILTCLPVPP
	18	69	83	TCQEGMNECPGHFGH		59	233	247	WMILTCLPVPPPPVR
	19	73	87	GMNECPGHFGHIDLA		60	237	251	TCLPVPPPPVRPSIS
	20	77	91	CPGHFGHIDLAKPVF		61	241	255	VPPPPVRPSISFNES
	21	81	95	FGHIDLAKPVFHVGF		62	245	259	PVRPSISFNESQRGE
	22	85	99	DLAKPVFHVGFIAKI		63	249	263	SISFNESQRGEDDLT
	23	89	103	PVFHVGFIAKIKKVC		64	253	267	NESQRGEDDLTFKLA
	24	93	107	VGFIAKIKKVCECVC		65	257	271	RGEDDLTFKLADILK
	25	97	111	AKIKKVCECVCMHCG		66	261	275	DLTFKLADILKANIS
	26	101	115	KVCECVCMHCGKLLL		67	265	279	KLADILKANISLETL
	27	105	119	CVCMHCGKLLLDEHN		68	269	283	ILKANISLETLEHNG
	28	109	123	HCGKLLLDEHNELMR		69	273	287	NISLETLEHNGAPHH
	29	113	127	LLLDEHNELMRQALA		70	277	291	ETLEHNGAPHHAIEE
	30	117	131	EHNELMRQALAIKDS		71	281	295	HNGAPHHAIEEAESL
	31	121	135	LMRQALAIKDSKKRF		72	285	299	PHHAIEEAESLLQFH
	32	125	139	ALAIKDSKKRFAAIW		73	289	303	IEEAESLLQFHVATY
	33	129	143	KDSKKRFAAIWTLCK		74	293	307	ESLLQFHVATYMDND
	34	133	147	KRFAAIWTLCKTKMV		75	297	311	QFHVATYMDNDIAGQ
	35	137	151	AIWTLCKTKMVCETD		76	301	315	ATYMDNDIAGQPQAL
	36	141	155	LCKTKMVCETDVPSE		77	305	319	DNDIAGQPQALQKSG
	37	145	159	KMVCETDVPSEDDPT		78	309	323	AGQPQALQKSGRPVK
	38	149	163	ETDVPSEDDPTQLVS		79	313	327	QALQKSGRPVKSIRA
	39	153	167	PSEDDPTQLVSRGGC		80	317	331	KSGRPVKSIRARLKG
	40	157	171	DPTQLVSRGGCGNTQ		81	321	335	PVKSIRARLKGKEGR
	41	161	175	LVSRGGCGNTQPTIR		82	325	339	IRARLKGKEGRIRGN

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	83	329	343	LKGKEGRIRGNLMGK
	84	333	347	EGRIRGNLMGKRVDF
	85	337	351	RGNLMGKRVDFSART
	86	341	355	MGKRVDFSARTVISG
	87	345	359	VDFSARTVISGDPNL
	88	349	363	ARTVISGDPNLELDQ
	89	353	367	ISGDPNLELDQVGVP
	90	357	371	PNLELDQVGVPKSIA
	91	361	375	LDQVGVPKSIAKTLT
	92	365	379	GVPKSIAKTLTYPEV
	93	369	383	SIAKTLTYPEVVTPY
	94	373	387	TLTYPEVVTPYNIDR
	95	377	391	PEVVTPYNIDRLTQL
	96	381	395	TPYNIDRLTQLVRNG
	97	385	399	IDRLTQLVRNGPNEH
	98	389	403	TQLVRNGPNEHPGAK
	99	393	407	RNGPNEHPGAKYVIR
	100	397	411	NEHPGAKYVIRDSGD
	101	401	415	GAKYVIRDSGDRIDL
	102	405	419	VIRDSGDRIDLRYSK
	103	409	423	SGDRIDLRYSKRAGD
	104	413	427	IDLRYSKRAGDIQLQ
	105	417	431	YSKRAGDIQLQYGWK
	106	421	435	AGDIQLQYGWKVERH
	107	425	439	QLQYGWKVERHIMDN
	108	429	443	GWKVERHIMDNDPVL
	109	433	447	ERHIMDNDPVLFNRQ
	110	437	451	MDNDPVLFNRQPSLH
	111	441	455	PVLFNRQPSLHKMSM
	112	445	459	NRQPSLHKMSMMAHR
	113	449	463	SLHKMSMMAHRVKVI
	114	453	467	MSMMAHRVKVIPYST
	115	457	471	AHRVKVIPYSTFRLN
	116	461	475	KVIPYSTFRLNLSVT
	117	465	479	YSTFRLNLSVTSPYN
	118	469	483	RLNLSVTSPYNADFD
	119	473	487	SVTSPYNADFDGDEM
	120	477	491	PYNADFDGDEMNLHV
	121	481	495	DFDGDEMNLHVPQSE
	122	485	499	DEMNLHVPQSEETRA
	123	489	503	LHVPQSEETRAELSQ
	124	493	507	QSEETRAELSQLCAV
	125	497	511	TRAELSQLCAVPLQI
	126	501	515	LSQLCAVPLQIVSPQ
	127	505	519	CAVPLQIVSPQSNKP
	128	509	523	LQIVSPQSNKPCMGI

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	129	513	527	SPQSNKPCMGIVQDT
	130	517	531	NKPCMGIVQDTLCGI
	131	521	535	MGIVQDTLCGIRKLT
	132	525	539	QDTLCGIRKLTLRDT
	133	529	543	CGIRKLTLRDTFIEL
	134	533	547	KLTLRDTFIELDQVL
	135	537	551	RDTFIELDQVLNMLY
	136	541	555	IELDQVLNMLYWVPD
	137	545	559	QVLNMLYWVPDWDGV
	138	549	563	MLYWVPDWDGVIPTP
	139	553	567	VPDWDGVIPTPAIIK
	140	557	571	DGVIPTPAIIKPKPL
	141	561	575	PTPAIIKPKPLWSGK
	142	565	579	IIKPKPLWSGKQILS
	143	569	583	KPLWSGKQILSVAIP
	144	573	587	SGKQILSVAIPNGIH
	145	577	591	ILSVAIPNGIHLQRF
	146	581	595	AIPNGIHLQRFDEGT
	147	585	599	GIHLQRFDEGTTLLS
	148	589	603	QRFDEGTTLLSPKDN
	149	593	607	EGTTLLSPKDNGMLI
	150	597	611	LLSPKDNGMLIIDGQ
	151	601	615	KDNGMLIIDGQIIFG
	152	605	619	MLIIDGQIIFGVVEK
	153	609	623	DGQIIFGVVEKKTVG
	154	613	627	IFGVVEKKTVGSSNG
	155	617	631	VEKKTVGSSNGGLIH
	156	621	635	TVGSSNGGLIHVVTR
	157	625	639	SNGGLIHVVTREKGP
	158	629	643	LIHVVTREKGPQVCA
	159	633	647	VTREKGPQVCAKLFG
	160	637	651	KGPQVCAKLFGNIQK
	161	641	655	VCAKLFGNIQKVVNF
	162	645	659	LFGNIQKVVNFWLLH
	163	649	663	IQKVVNFWLLHNGFS
	164	653	667	VNFWLLHNGFSTGIG
	165	657	671	LLHNGFSTGIGDTIA
	166	661	675	GFSTGIGDTIADGPT
	167	665	679	GIGDTIADGPTMREI
	168	669	683	TIADGPTMREITETI
	169	673	687	GPTMREITETIAEAK
	170	677	691	REITETIAEAKKKVL
	171	681	695	ETIAEAKKKVLDVTK
	172	685	699	EAKKKVLDVTKEAQA
	173	689	703	KVLDVTKEAQANLLT
	174	693	707	VTKEAQANLLTAKHG

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	175	697	711	AQANLLTAKHGMTLR
	176	701	715	LLTAKHGMTLRESFE
	177	705	719	KHGMTLRESFEDNVV
	178	709	723	TLRESFEDNVVRFLN
	179	713	727	SFEDNVVRFLNEARD
	180	717	731	NVVRFLNEARDKAGR
	181	721	735	FLNEARDKAGRLAEV
	182	725	739	ARDKAGRLAEVNLKD
	183	729	743	AGRLAEVNLKDLNNV
	184	733	747	AEVNLKDLNNVKQMV
	185	737	751	LKDLNNVKQMVMAGS
	186	741	755	NNVKQMVMAGSKGSF
	187	745	759	QMVMAGSKGSFINIA
	188	749	763	AGSKGSFINIAQMSA
	189	753	767	GSFINIAQMSACVGQ
	190	757	771	NIAQMSACVGQQSVE
	191	761	775	MSACVGQQSVEGKRI
	192	765	779	VGQQSVEGKRIAFGF
	193	769	783	SVEGKRIAFGFVDRT
	194	773	787	KRIAFGFVDRTLPHF
	195	777	791	FGFVDRTLPHFSKDD
	196	781	795	DRTLPHFSKDDYSPE
	197	785	799	PHFSKDDYSPESKGF
	198	789	803	KDDYSPESKGFVENS
	199	793	807	SPESKGFVENSYLRG
	200	797	811	KGFVENSYLRGLTPQ
	201	801	815	ENSYLRGLTPQEFFF
	202	805	819	LRGLTPQEFFFHAMG
	203	809	823	TPQEFFFHAMGGREG
	204	813	827	FFFHAMGGREGLIDT
	205	817	831	AMGGREGLIDTAVKT
	206	821	835	REGLIDTAVKTAETG
	207	825	839	IDTAVKTAETGYIQR
	208	829	843	VKTAETGYIQRRLVK
	209	833	847	ETGYIQRRLVKALED
	210	837	851	IQRRLVKALEDIMVH
	211	841	855	LVKALEDIMVHYDNT
	212	845	859	LEDIMVHYDNTTRNS
	213	849	863	MVHYDNTTRNSLGNV
	214	853	867	DNTTRNSLGNVIQFI
	215	857	871	RNSLGNVIQFIYGED
	216	861	875	GNVIQFIYGEDGMDA
	217	865	879	QFIYGEDGMDAAHIE
	218	869	883	GEDGMDAAHIEKQSL
	219	873	887	MDAAHIEKQSLDTIG
	220	877	891	HIEKQSLDTIGGSDA

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	221	881	895	QSLDTIGGSDAAFEK
	222	885	899	TIGGSDAAFEKRYRV
	223	889	903	SDAAFEKRYRVDLLN
	224	893	907	FEKRYRVDLLNTDHT
	225	897	911	YRVDLLNTDHTLDPS
	226	901	915	LLNTDHTLDPSLLES
	227	905	919	DHTLDPSLLESGSEI
	228	909	923	DPSLLESGSEILGDL
	229	913	927	LESGSEILGDLKLQV
	230	917	931	SEILGDLKLQVLLDE
	231	921	935	GDLKLQVLLDEEYKQ
	232	925	939	LQVLLDEEYKQLVKD
	233	929	943	LDEEYKQLVKDRKFL
	234	933	947	YKQLVKDRKFLREVF
	235	937	951	VKDRKFLREVFVDGE
	236	941	955	KFLREVFVDGEANWP
	237	945	959	EVFVDGEANWPLPVN
	238	949	963	DGEANWPLPVNIRRI
	239	953	967	NWPLPVNIRRIIQNA
	240	957	971	PVNIRRIIQNAQQTF
	241	961	975	RRIIQNAQQTFHIDH
	242	965	979	QNAQQTFHIDHTKPS
	243	969	983	QTFHIDHTKPSDLTI
	244	973	987	IDHTKPSDLTIKDIV
	245	977	991	KPSDLTIKDIVLGVK
	246	981	995	LTIKDIVLGVKDLQE
	247	985	999	DIVLGVKDLQENLLV
	248	989	1003	GVKDLQENLLVLRGK
	249	993	1007	LQENLLVLRGKNEII
	250	997	1011	LLVLRGKNEIIQNAQ
	251	1001	1015	RGKNEIIQNAQRDAV
	252	1005	1019	EIIQNAQRDAVTLFC
	253	1009	1023	NAQRDAVTLFCCLLR
	254	1013	1027	DAVTLFCCLLRSRLA
	255	1017	1031	LFCCLLRSRLATRRV
	256	1021	1035	LLRSRLATRRVLQEY
	257	1025	1039	RLATRRVLQEYRLTK
	258	1029	1043	RRVLQEYRLTKQAFD
	259	1033	1047	QEYRLTKQAFDWVLS
	260	1037	1051	LTKQAFDWVLSNIEA
	261	1041	1055	AFDWVLSNIEAQFLR
	262	1045	1059	VLSNIEAQFLRSVVH
	263	1049	1063	IEAQFLRSVVHPGEM
	264	1053	1067	FLRSVVHPGEMVGVL
	265	1057	1071	VVHPGEMVGVLAAQS
	266	1061	1075	GEMVGVLAAQSIGEP

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	267	1065	1079	GVLAAQSIGEPATQM
	268	1069	1083	AQSIGEPATQMTLNT
	269	1073	1087	GEPATQMTLNTFHFA
	270	1077	1091	TQMTLNTFHFAGVAS
	271	1081	1095	LNTFHFAGVASKKVT
	272	1085	1099	HFAGVASKKVTSGVP
	273	1089	1103	VASKKVTSGVPRLKE
	274	1093	1107	KVTSGVPRLKEILNV
	275	1097	1111	GVPRLKEILNVAKNM
	276	1101	1115	LKEILNVAKNMKTPS
	277	1105	1119	LNVAKNMKTPSLTVY
	278	1109	1123	KNMKTPSLTVYLEPG
	279	1113	1127	TPSLTVYLEPGHAAD
	280	1117	1131	TVYLEPGHAADQEQA
	281	1121	1135	EPGHAADQEQAKLIR
	282	1125	1139	AADQEQAKLIRSAIE
	283	1129	1143	EQAKLIRSAIEHTTL
	284	1133	1147	LIRSAIEHTTLKSVT
	285	1137	1151	AIEHTTLKSVTIASE
	286	1141	1155	TTLKSVTIASEIYYD
	287	1145	1159	SVTIASEIYYDPDPR
	288	1149	1163	ASEIYYDPDPRSTVI
	289	1153	1167	YYDPDPRSTVIPEDE
	290	1157	1171	DPRSTVIPEDEEIIQ
	291	1161	1175	TVIPEDEEIIQLHFS
	292	1165	1179	EDEEIIQLHFSLLDE
	293	1169	1183	IIQLHFSLLDEEAEQ
	294	1173	1187	HFSLLDEEAEQSFDQ
	295	1177	1191	LDEEAEQSFDQQSPW
	296	1181	1195	AEQSFDQQSPWLLRL
	297	1185	1199	FDQQSPWLLRLELDR
	298	1189	1203	SPWLLRLELDRAAMN
	299	1193	1207	LRLELDRAAMNDKDL
	300	1197	1211	LDRAAMNDKDLTMGQ
	301	1201	1215	AMNDKDLTMGQVGER
	302	1205	1219	KDLTMGQVGERIKQT
	303	1209	1223	MGQVGERIKQTFKND
	304	1213	1227	GERIKQTFKNDLFVI
	305	1217	1231	KQTFKNDLFVIWSED
	306	1221	1235	KNDLFVIWSEDNDEK
	307	1225	1239	FVIWSEDNDEKLIIR
	308	1229	1243	SEDNDEKLIIRCRVV
	309	1233	1247	DEKLIIRCRVVRPKS
	310	1237	1251	IIRCRVVRPKSLDAE
	311	1241	1255	RVVRPKSLDAETEAE
	312	1245	1259	PKSLDAETEAEEDHM
Rpb1	313	1249	1263	DAETEAEEDHMLKKI
	314	1253	1267	EAEEDHMLKKIENTM
	315	1257	1271	DHMLKKIENTMLENI
	316	1261	1275	KKIENTMLENITLRG
	317	1265	1279	NTMLENITLRGVENI
	318	1269	1283	ENITLRGVENIERVV
	319	1273	1287	LRGVENIERVVMMKY
	320	1277	1291	ENIERVVMMKYDRKV
	321	1281	1295	RVVMMKYDRKVPSPT
	322	1285	1299	MKYDRKVPSPTGEYV
	323	1289	1303	RKVPSPTGEYVKEPE
	324	1293	1307	SPTGEYVKEPEWVLE
	325	1297	1311	EYVKEPEWVLETDGV
	326	1301	1315	EPEWVLETDGVNLSE
	327	1305	1319	VLETDGVNLSEVMTV
	328	1309	1323	DGVNLSEVMTVPGID
	329	1313	1327	LSEVMTVPGIDPTRI
	330	1317	1331	MTVPGIDPTRIYTNS
	331	1321	1335	GIDPTRIYTNSFIDI
	332	1325	1339	TRIYTNSFIDIMEVL
	333	1329	1343	TNSFIDIMEVLGIEA
	334	1333	1347	IDIMEVLGIEAGRAA
	335	1337	1351	EVLGIEAGRAALYKE
	336	1341	1355	IEAGRAALYKEVYNV
	337	1345	1359	RAALYKEVYNVIASD
	338	1349	1363	YKEVYNVIASDGSYV
	339	1353	1367	YNVIASDGSYVNYRH
	340	1357	1371	ASDGSYVNYRHMALL
	341	1361	1375	SYVNYRHMALLVDVM
	342	1365	1379	YRHMALLVDVMTTQG
	343	1369	1383	ALLVDVMTTQGGLTS
	344	1373	1387	DVMTTQGGLTSVTRH
	345	1377	1391	TQGGLTSVTRHGFNR
	346	1381	1395	LTSVTRHGFNRSNTG
	347	1385	1399	TRHGFNRSNTGALMR
	348	1389	1403	FNRSNTGALMRCSFE
	349	1393	1407	NTGALMRCSFEETVE
	350	1397	1411	LMRCSFEETVEILFE
	351	1401	1415	SFEETVEILFEAGAS
	352	1405	1419	TVEILFEAGASAELD
	353	1409	1423	LFEAGASAELDDCRG
	354	1413	1427	GASAELDDCRGVSEN
	355	1417	1431	ELDDCRGVSENVILG
	356	1421	1435	CRGVSENVILGQMAP
	357	1425	1439	SENVILGQMAPIGTG
	358	1429	1443	ILGQMAPIGTGAFDV

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	359	1433	1447	MAPIGTGAFDVMIDE
	360	1437	1451	GTGAFDVMIDEESLV
	361	1441	1455	FDVMIDEESLVKYMP
	362	1445	1459	IDEESLVKYMPEQKI
	363	1449	1463	SLVKYMPEQKITEIE
	364	1453	1467	YMPEQKITEIEDGQD
	365	1457	1471	QKITEIEDGQDGGVT
	366	1461	1475	EIEDGQDGGVTPYSN
	367	1465	1479	GQDGGVTPYSNESGL
	368	1469	1483	GVTPYSNESGLVNAD
	369	1473	1487	YSNESGLVNADLDVK
	370	1477	1491	SGLVNADLDVKDELM
	371	1481	1495	NADLDVKDELMFSPL
	372	1485	1499	DVKDELMFSPLVDSG
	373	1489	1503	ELMFSPLVDSGSNDA
	374	1493	1507	SPLVDSGSNDAMAGG
	375	1497	1511	DSGSNDAMAGGFTAY
	376	1501	1515	NDAMAGGFTAYGGAD
	377	1505	1519	AGGFTAYGGADYGEA
	378	1509	1523	TAYGGADYGEATSPF
	379	1513	1527	GADYGEATSPFGAYG
	380	1517	1531	GEATSPFGAYGEAPT
	381	1521	1535	SPFGAYGEAPTSPGF
	382	1525	1539	AYGEAPTSPGFGVSS
	383	1529	1543	APTSPGFGVSSPGFS
	384	1533	1547	PGFGVSSPGFSPTSP
	385	1537	1551	VSSPGFSPTSPTYSP
	386	1541	1555	GFSPTSPTYSPTSPA
	387	1545	1559	TSPTYSPTSPAYSPT
	388	1549	1563	YSPTSPAYSPTSPSY
	389	1553	1567	SPAYSPTSPSYSPTS
	390	1557	1571	SPTSPSYSPTSPSYS
	391	1561	1575	PSYSPTSPSYSPTSP
	392	1565	1579	PTSPSYSPTSPSYSP
	393	1569	1583	SYSPTSPSYSPTSPS
	394	1573	1587	TSPSYSPTSPSYSPT
	395	1577	1591	YSPTSPSYSPTSPSY
	396	1581	1595	SPSYSPTSPSYSPTS
	397	1585	1599	SPTSPSYSPTSPSYS
	398	1589	1603	PSYSPTSPSYSPTSP
	399	1593	1607	PTSPSYSPTSPSYSP
	400	1597	1611	SYSPTSPSYSPTSPS
	401	1601	1615	TSPSYSPTSPSYSPT
	402	1605	1619	YSPTSPSYSPTSPSY
	403	1609	1623	SPSYSPTSPSYSPTS
	404	1613	1627	SPTSPSYSPTSPSYS

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb1	405	1617	1631	PSYSPTSPSYSPTSP
	406	1621	1635	PTSPSYSPTSPSYSP
	407	1625	1639	SYSPTSPSYSPTSPS
	408	1629	1643	TSPSYSPTSPSYSPT
	409	1633	1647	YSPTSPSYSPTSPSY
	410	1637	1651	SPSYSPTSPSYSPTS
	411	1641	1655	SPTSPSYSPTSPSYS
	412	1645	1659	PSYSPTSPSYSPTSP
	413	1649	1663	PTSPSYSPTSPAYSP
	414	1653	1667	SYSPTSPAYSPTSPS
	415	1657	1671	TSPAYSPTSPSYSPT
	416	1661	1675	YSPTSPSYSPTSPSY
	417	1665	1679	SPSYSPTSPSYSPTS
	418	1669	1683	SPTSPSYSPTSPSYS
	419	1673	1687	PSYSPTSPSYSPTSP
	420	1677	1691	PTSPSYSPTSPSYSP
	421	1681	1695	SYSPTSPSYSPTSPN
	422	1685	1699	TSPSYSPTSPNYSPT
	423	1689	1703	YSPTSPNYSPTSPSY
	424	1693	1707	SPNYSPTSPSYSPTS
	425	1697	1711	SPTSPSYSPTSPGYS
	426	1701	1715	PSYSPTSPGYSPGSP
	427	1705	1719	PTSPGYSPGSPAYSP
	428	1709	1723	GYSPGSPAYSPKQDE
	429	1713	1727	GSPAYSPKQDEQKHN
	430	1717	1731	YSPKQDEQKHNENEN
Rpb1/2	431			QDEQKHNENENSRMS
	432			KHNENENSRMSDLAN
	433			NENSRMSDLANSEKY
	434			RMSDLANSEKYYDED
Rpb2	435	4	18	LANSEKYYDEDPYGF
	436	8	22	EKYYDEDPYGFEDES
	437	12	26	DEDPYGFEDESAPIT
	438	16	30	YGFEDESAPITAEDS
	439	20	34	DESAPITAEDSWAVI
	440	24	38	PITAEDSWAVISAFF
	441	28	42	EDSWAVISAFFREKG
	442	32	46	AVISAFFREKGLVSQ
	443	36	50	AFFREKGLVSQQLDS
	444	40	54	EKGLVSQQLDSFNQF
	445	44	58	VSQQLDSFNQFVDYT
	446	48	62	LDSFNQFVDYTLQDI
	447	52	66	NQFVDYTLQDIICED
	448	56	70	DYTLQDIICEDSTLI
	449	60	74	QDIICEDSTLILEQL
	450	64	78	CEDSTLILEQLAQHT

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb2	451	68	82	TLILEQLAQHTTESD
	452	72	86	EQLAQHTTESDNISR
	453	76	90	QHTTESDNISRKYEI
	454	80	94	ESDNISRKYEISFGK
	455	84	98	ISRKYEISFGKIYVT
	456	88	102	YEISFGKIYVTKPMV
	457	92	106	FGKIYVTKPMVNESD
	458	96	110	YVTKPMVNESDGVTH
	459	100	114	PMVNESDGVTHALYP
	460	104	118	ESDGVTHALYPQEAR
	461	108	122	VTHALYPQEARLRNL
	462	112	126	LYPQEARLRNLTYSS
	463	116	130	EARLRNLTYSSGLFV
	464	120	134	RNLTYSSGLFVDVKK
	465	124	138	YSSGLFVDVKKRTYE
	466	128	142	LFVDVKKRTYEAIDV
	467	132	146	VKKRTYEAIDVPGRE
	468	136	150	TYEAIDVPGRELKYE
	469	140	154	IDVPGRELKYELIAE
	470	144	158	GRELKYELIAEESED
	471	148	162	KYELIAEESEDDSES
	472	152	166	IAEESEDDSESGKVF
	473	156	170	SEDDSESGKVFIGRL
	474	160	174	SESGKVFIGRLPIML
	475	164	178	KVFIGRLPIMLRSKN
	476	168	182	GRLPIMLRSKNCYLS
	477	172	186	IMLRSKNCYLSEATE
	478	176	190	SKNCYLSEATESDLY
	479	180	194	YLSEATESDLYKLKE
	480	184	198	ATESDLYKLKECPFD
	481	188	202	DLYKLKECPFDMGGY
	482	192	206	LKECPFDMGGYFIIN
	483	196	210	PFDMGGYFIINGSEK
	484	200	214	GGYFIINGSEKVLIA
	485	204	218	IINGSEKVLIAQERS
	486	208	222	SEKVLIAQERSAGNI
	487	212	226	LIAQERSAGNIVQVF
	488	216	230	ERSAGNIVQVFKKAA
	489	220	234	GNIVQVFKKAAPSPI
	490	224	238	QVFKKAAPSPISHVA
	491	228	242	KAAPSPISHVAEIRS
	492	232	246	SPISHVAEIRSALEK
	493	236	250	HVAEIRSALEKGSRF
	494	240	254	IRSALEKGSRFISTL
	495	244	258	LEKGSRFISTLQVKL
	496	248	262	SRFISTLQVKLYGRE

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb2	497	252	266	STLQVKLYGREGSSA
	498	256	270	VKLYGREGSSARTIK
	499	260	274	GREGSSARTIKATLP
	500	264	278	SSARTIKATLPYIKQ
	501	268	282	TIKATLPYIKQDIPI
	502	272	286	TLPYIKQDIPIVIIF
	503	276	290	IKQDIPIVIIFRALG
	504	280	294	IPIVIIFRALGIIPD
	505	284	298	IIFRALGIIPDGEIL
	506	288	302	ALGIIPDGEILEHIC
	507	292	306	IPDGEILEHICYDVN
	508	296	310	EILEHICYDVNDWQM
	509	300	314	HICYDVNDWQMLEML
	510	304	318	DVNDWQMLEMLKPCV
	511	308	322	WQMLEMLKPCVEDGF
	512	312	326	EMLKPCVEDGFVIQD
	513	316	330	PCVEDGFVIQDRETA
	514	320	334	DGFVIQDRETALDFI
	515	324	338	IQDRETALDFIGRRG
	516	328	342	ETALDFIGRRGTALG
	517	332	346	DFIGRRGTALGIKKE
	518	336	350	RRGTALGIKKEKRIQ
	519	340	354	ALGIKKEKRIQYAKD
	520	344	358	KKEKRIQYAKDILQK
	521	348	362	RIQYAKDILQKEFLP
	522	352	366	AKDILQKEFLPHITQ
	523	356	370	LQKEFLPHITQLEGF
	524	360	374	FLPHITQLEGFESRK
	525	364	378	ITQLEGFESRKAFFL
	526	368	382	EGFESRKAFFLGYMI
	527	372	386	SRKAFFLGYMINRLL
	528	376	390	FFLGYMINRLLLCAL
	529	380	394	YMINRLLLCALDRKD
	530	384	398	RLLLCALDRKDQDDR
	531	388	402	CALDRKDQDDRDHFG
	532	392	406	RKDQDDRDHFGKKRL
	533	396	410	DDRDHFGKKRLDLAG
	534	400	414	HFGKKRLDLAGPLLA
	535	404	418	KRLDLAGPLLAQLFK
	536	408	422	LAGPLLAQLFKTLFK
	537	412	426	LLAQLFKTLFKKLTK
	538	416	430	LFKTLFKKLTKDIFR
	539	420	434	LFKKLTKDIFRYMQR
	540	424	438	LTKDIFRYMQRTVEE
	541	428	442	IFRYMQRTVEEAHDF
	542	432	446	MQRTVEEAHDFNMKL

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb2	543	436	450	VEEAHDFNMKLAINA
	544	440	454	HDFNMKLAINAKTIT
	545	444	458	MKLAINAKTITSGLK
	546	448	462	INAKTITSGLKYALA
	547	452	466	TITSGLKYALATGNW
	548	456	470	GLKYALATGNWGEQK
	549	460	474	ALATGNWGEQKKAMS
	550	464	478	GNWGEQKKAMSSRAG
	551	468	482	EQKKAMSSRAGVSQV
	552	472	486	AMSSRAGVSQVLNRY
	553	476	490	RAGVSQVLNRYTYSS
	554	480	494	SQVLNRYTYSSTLSH
	555	484	498	NRYTYSSTLSHLRRT
	556	488	502	YSSTLSHLRRTNTPI
	557	492	506	LSHLRRTNTPIGRDG
	558	496	510	RRTNTPIGRDGKLAK
	559	500	514	TPIGRDGKLAKPRQL
	560	504	518	RDGKLAKPRQLHNTH
	561	508	522	LAKPRQLHNTHWGLV
	562	512	526	RQLHNTHWGLVCPAE
	563	516	530	NTHWGLVCPAETPEG
	564	520	534	GLVCPAETPEGQACG
	565	524	538	PAETPEGQACGLVKN
	566	528	542	PEGQACGLVKNLSLM
	567	532	546	ACGLVKNLSLMSCIS
	568	536	550	VKNLSLMSCISVGTD
	569	540	554	SLMSCISVGTDPMPI
	570	544	558	CISVGTDPMPIITFL
	571	548	562	GTDPMPIITFLSEWG
	572	552	566	MPIITFLSEWGMEPL
	573	556	570	TFLSEWGMEPLEDYV
	574	560	574	EWGMEPLEDYVPHQS
	575	564	578	EPLEDYVPHQSPDAT
	576	568	582	DYVPHQSPDATRVFV
	577	572	586	HQSPDATRVFVNGVW
	578	576	590	DATRVFVNGVWHGVH
	579	580	594	VFVNGVWHGVHRNPA
	580	584	598	GVWHGVHRNPARLME
	581	588	602	GVHRNPARLMETLRT
	582	592	606	NPARLMETLRTLRRK
	583	596	610	LMETLRTLRRKGDIN
	584	600	614	LRTLRRKGDINPEVS
	585	604	618	RRKGDINPEVSMIRD
	586	608	622	DINPEVSMIRDIREK
	587	612	626	EVSMIRDIREKELKI
	588	616	630	IRDIREKELKIFTDA
Rpb2	589	620	634	REKELKIFTDAGRVY
	590	624	638	LKIFTDAGRVYRPLF
	591	628	642	TDAGRVYRPLFIVED
	592	632	646	RVYRPLFIVEDDESL
	593	636	650	PLFIVEDDESLGHKE
	594	640	654	VEDDESLGHKELKVR
	595	644	658	ESLGHKELKVRKGHI
	596	648	662	HKELKVRKGHIAKLM
	597	652	666	KVRKGHIAKLMATEY
	598	656	670	GHIAKLMATEYQDIE
	599	660	674	KLMATEYQDIEGGFE
	600	664	678	TEYQDIEGGFEDVEE
	601	668	682	DIEGGFEDVEEYTWS
	602	672	686	GFEDVEEYTWSSLLN
	603	676	690	VEEYTWSSLLNEGLV
	604	680	694	TWSSLLNEGLVEYID
	605	684	698	LLNEGLVEYIDAEEE
	606	688	702	GLVEYIDAEEEESIL
	607	692	706	YIDAEEEESILIAMQ
	608	696	710	EEEESILIAMQPEDL
	609	700	714	SILIAMQPEDLEPAE
	610	704	718	AMQPEDLEPAEANEE
	611	708	722	EDLEPAEANEENDLD
	612	712	726	PAEANEENDLDVDPA
	613	716	730	NEENDLDVDPAKRIR
	614	720	734	DLDVDPAKRIRVSHH
	615	724	738	DPAKRIRVSHHATTF
	616	728	742	RIRVSHHATTFTHCE
	617	732	746	SHHATTFTHCEIHPS
	618	736	750	TTFTHCEIHPSMILG
	619	740	754	HCEIHPSMILGVAAS
	620	744	758	HPSMILGVAASIIPF
	621	748	762	ILGVAASIIPFPDHN
	622	752	766	AASIIPFPDHNQSPR
	623	756	770	IPFPDHNQSPRNTYQ
	624	760	774	DHNQSPRNTYQSAMG
	625	764	778	SPRNTYQSAMGKQAM
	626	768	782	TYQSAMGKQAMGVFL
	627	772	786	AMGKQAMGVFLTNYN
	628	776	790	QAMGVFLTNYNVRMD
	629	780	794	VFLTNYNVRMDTMAN
	630	784	798	NYNVRMDTMANILYY
	631	788	802	RMDTMANILYYPQKP
	632	792	806	MANILYYPQKPLGTT
	633	796	810	LYYPQKPLGTTRAME
	634	800	814	QKPLGTTRAMEYLKF

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb2	635	804	818	GTTRAMEYLKFRELP
	636	808	822	AMEYLKFRELPAGQN
	637	812	826	LKFRELPAGQNAIVA
	638	816	830	ELPAGQNAIVAIACY
	639	820	834	GQNAIVAIACYSGYN
	640	824	838	IVAIACYSGYNQEDS
	641	828	842	ACYSGYNQEDSMIMN
	642	832	846	GYNQEDSMIMNQSSI
	643	836	850	EDSMIMNQSSIDRGL
	644	840	854	IMNQSSIDRGLFRSL
	645	844	858	SSIDRGLFRSLFFRS
	646	848	862	RGLFRSLFFRSYMDQ
	647	852	866	RSLFFRSYMDQEKKY
	648	856	870	FRSYMDQEKKYGMSI
	649	860	874	MDQEKKYGMSITETF
	650	864	878	KKYGMSITETFEKPQ
	651	868	882	MSITETFEKPQRTNT
	652	872	886	ETFEKPQRTNTLRMK
	653	876	890	KPQRTNTLRMKHGTY
	654	880	894	TNTLRMKHGTYDKLD
	655	884	898	RMKHGTYDKLDDDGL
	656	888	902	GTYDKLDDDGLIAPG
	657	892	906	KLDDDGLIAPGVRVS
	658	896	910	DGLIAPGVRVSGEDV
	659	900	914	APGVRVSGEDVIIGK
	660	904	918	RVSGEDVIIGKTTPI
	661	908	922	EDVIIGKTTPISPDE
	662	912	926	IGKTTPISPDEEELG
	663	916	930	TPISPDEEELGQRTA
	664	920	934	PDEEELGQRTAYHSK
	665	924	938	ELGQRTAYHSKRDAS
	666	928	942	RTAYHSKRDASTPLR
	667	932	946	HSKRDASTPLRSTEN
	668	936	950	DASTPLRSTENGIVD
	669	940	954	PLRSTENGIVDQVLV
	670	944	958	TENGIVDQVLVTTNQ
	671	948	962	IVDQVLVTTNQDGLK
	672	952	966	VLVTTNQDGLKFVKV
	673	956	970	TNQDGLKFVKVRVRT
	674	960	974	GLKFVKVRVRTTKIP
	675	964	978	VKVRVRTTKIPQIGD
	676	968	982	VRTTKIPQIGDKFAS
	677	972	986	KIPQIGDKFASRHGQ
	678	976	990	IGDKFASRHGQKGTI
	679	980	994	FASRHGQKGTIGITY
	680	984	998	HGQKGTIGITYRRED

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb2	681	988	1002	GTIGITYRREDMPFT
	682	992	1006	ITYRREDMPFTAEGI
	683	996	1010	REDMPFTAEGIVPDL
	684	1000	1014	PFTAEGIVPDLIINP
	685	1004	1018	EGIVPDLIINPHAIP
	686	1008	1022	PDLIINPHAIPSRMT
	687	1012	1026	INPHAIPSRMTVAHL
	688	1016	1030	AIPSRMTVAHLIECL
	689	1020	1034	RMTVAHLIECLLSKV
	690	1024	1038	AHLIECLLSKVAALS
	691	1028	1042	ECLLSKVAALSGNEG
	692	1032	1046	SKVAALSGNEGDASP
	693	1036	1050	ALSGNEGDASPFTDI
	694	1040	1054	NEGDASPFTDITVEG
	695	1044	1058	ASPFTDITVEGISKL
	696	1048	1062	TDITVEGISKLLREH
	697	1052	1066	VEGISKLLREHGYQS
	698	1056	1070	SKLLREHGYQSRGFE
	699	1060	1074	REHGYQSRGFEVMYN
	700	1064	1078	YQSRGFEVMYNGHTG
	701	1068	1082	GFEVMYNGHTGKKLM
	702	1072	1086	MYNGHTGKKLMAQIF
	703	1076	1090	HTGKKLMAQIFFGPT
	704	1080	1094	KLMAQIFFGPTYYQR
	705	1084	1098	QIFFGPTYYQRLRHM
	706	1088	1102	GPTYYQRLRHMVDDK
	707	1092	1106	YQRLRHMVDDKIHAR
	708	1096	1110	RHMVDDKIHARARGP
	709	1100	1114	DDKIHARARGPMQVL
	710	1104	1118	HARARGPMQVLTRQP
	711	1108	1122	RGPMQVLTRQPVEGR
	712	1112	1126	QVLTRQPVEGRSRDG
	713	1116	1130	RQPVEGRSRDGGLRF
	714	1120	1134	EGRSRDGGLRFGEME
	715	1124	1138	RDGGLRFGEMERDCM
	716	1128	1142	LRFGEMERDCMIAHG
	717	1132	1146	EMERDCMIAHGAASF
	718	1136	1150	DCMIAHGAASFLKER
	719	1140	1154	AHGAASFLKERLMEA
	720	1144	1158	ASFLKERLMEASDAF
	721	1148	1162	KERLMEASDAFRVHI
	722	1152	1166	MEASDAFRVHICGIC
	723	1156	1170	DAFRVHICGICGLMT
	724	1160	1174	VHICGICGLMTVIAK
	725	1164	1178	GICGLMTVIAKLNHN
	726	1168	1182	LMTVIAKLNHNQFEC

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb2	727	1172	1186	IAKLNHNQFECKGCD
	728	1176	1190	NHNQFECKGCDNKID
	729	1180	1194	FECKGCDNKIDIYQI
	730	1184	1198	GCDNKIDIYQIHIPY
	731	1188	1202	KIDIYQIHIPYAAKL
	732	1192	1206	YQIHIPYAAKLLFQE
	733	1196	1210	IPYAAKLLFQELMAM
	734	1200	1214	AKLLFQELMAMNITP
	735	1204	1218	FQELMAMNITPRLYT
	736	1208	1222	MAMNITPRLYTDRSR
Rpb2/3	737			ITPRLYTDRSRDFMS
	738			LYTDRSRDFMSEEGP
	739			RSRDFMSEEGPQVKI
	740			FMSEEGPQVKIREAS
Rpb3	741	4	18	EGPQVKIREASKDNV
	742	8	22	VKIREASKDNVDFIL
	743	12	26	EASKDNVDFILSNVD
	744	16	30	DNVDFILSNVDLAMA
	745	20	34	FILSNVDLAMANSLR
	746	24	38	NVDLAMANSLRRVMI
	747	28	42	AMANSLRRVMIAEIP
	748	32	46	SLRRVMIAEIPTLAI
	749	36	50	VMIAEIPTLAIDSVE
	750	40	54	EIPTLAIDSVEVETN
	751	44	58	LAIDSVEVETNTTVL
	752	48	62	SVEVETNTTVLADEF
	753	52	66	ETNTTVLADEFIAHR
	754	56	70	TVLADEFIAHRLGLI
	755	60	74	DEFIAHRLGLIPLQS
	756	64	78	AHRLGLIPLQSMDIE
	757	68	82	GLIPLQSMDIEQLEY
	758	72	86	LQSMDIEQLEYSRDC
	759	76	90	DIEQLEYSRDCFCED
	760	80	94	LEYSRDCFCEDHCDK
	761	84	98	RDCFCEDHCDKCSVV
	762	88	102	CEDHCDKCSVVLTLQ
	763	92	106	CDKCSVVLTLQAFGE
	764	96	110	SVVLTLQAFGESEST
	765	100	114	TLQAFGESESTTNVY
	766	104	118	FGESESTTNVYSKDL
	767	108	122	ESTTNVYSKDLVIVS
	768	112	126	NVYSKDLVIVSNLMG
	769	116	130	KDLVIVSNLMGRNIG
	770	120	134	IVSNLMGRNIGHPII
	771	124	138	LMGRNIGHPIIQDKE
	772	128	142	NIGHPIIQDKEGNGV

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb3	773	132	146	PIIQDKEGNGVLICK
	774	136	150	DKEGNGVLICKLRKG
	775	140	154	NGVLICKLRKGQELK
	776	144	158	ICKLRKGQELKLTCV
	777	148	162	RKGQELKLTCVAKKG
	778	152	166	ELKLTCVAKKGIAKE
	779	156	170	TCVAKKGIAKEHAKW
	780	160	174	KKGIAKEHAKWGPAA
	781	164	178	AKEHAKWGPAAAIEF
	782	168	182	AKWGPAAAIEFEYDP
	783	172	186	PAAAIEFEYDPWNKL
	784	176	190	IEFEYDPWNKLKHTD
	785	180	194	YDPWNKLKHTDYWYE
	786	184	198	NKLKHTDYWYEQDSA
	787	188	202	HTDYWYEQDSAKEWP
	788	192	206	WYEQDSAKEWPQSKN
	789	196	210	DSAKEWPQSKNCEYE
	790	200	214	EWPQSKNCEYEDPPN
	791	204	218	SKNCEYEDPPNEGDP
	792	208	222	EYEDPPNEGDPFDYK
	793	212	226	PPNEGDPFDYKAQAD
	794	216	230	GDPFDYKAQADTFYM
	795	220	234	DYKAQADTFYMNVES
	796	224	238	QADTFYMNVESVGSI
	797	228	242	FYMNVESVGSIPVDQ
	798	232	246	VESVGSIPVDQVVVR
	799	236	250	GSIPVDQVVVRGIDT
	800	240	254	VDQVVVRGIDTLQKK
	801	244	258	VVRGIDTLQKKVASI
	802	248	262	IDTLQKKVASILLAL
	803	252	266	QKKVASILLALTQMD
	804	256	270	ASILLALTQMDQDKV
	805	260	274	LALTQMDQDKVNFAS
	806	264	278	QMDQDKVNFASGDNN
	807	268	282	DKVNFASGDNNTASN
	808	272	286	FASGDNNTASNMLGS
	809	276	290	DNNTASNMLGSNEDV
	810	280	294	ASNMLGSNEDVMMTG
	811	284	298	LGSNEDVMMTGAEQD
	812	288	302	EDVMMTGAEQDPYSN
	813	292	306	MTGAEQDPYSNASQM
	814	296	310	EQDPYSNASQMGNTG
	815	300	314	YSNASQMGNTGSGGY
	816	304	318	SQMGNTGSGGYDNAW
Rpb3/4	817			NTGSGGYDNAWMNVS
	818			GGYDNAWMNVSTSTF

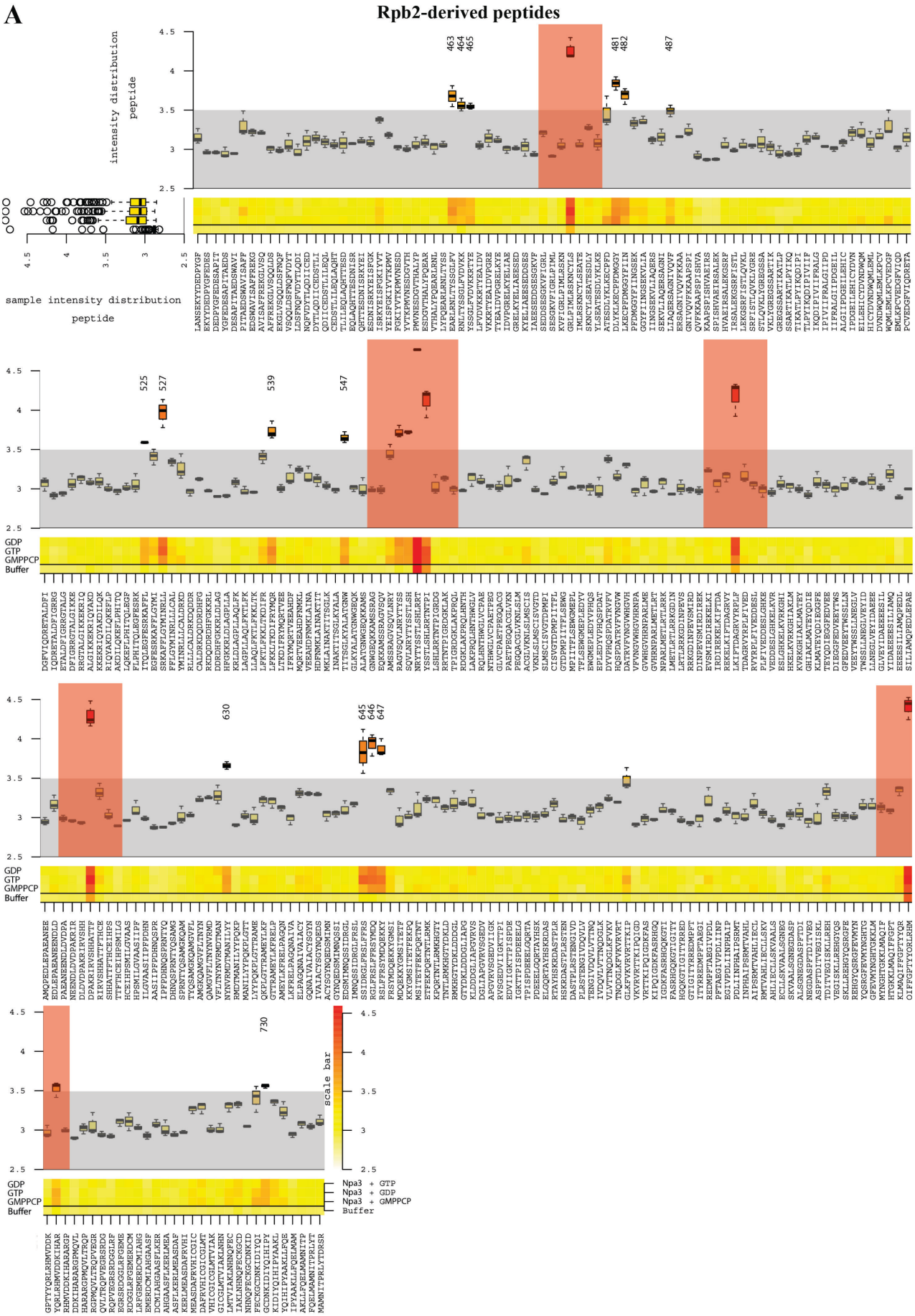
Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
	819			NAWMNVSTSTFQTRR
Rpb4	820	2	16	NVSTSTFQTRRRRLK
	821	6	20	STFQTRRRRLKKVEE
	822	10	24	TRRRRLKKVEEEENA
	823	14	28	RLKKVEEEENAATLQ
	824	18	32	VEEEENAATLQLGQE
	825	22	36	ENAATLQLGQEFQLK
	826	26	40	TLQLGQEFQLKQINH
	827	30	44	GQEFQLKQINHQGEE
	828	34	48	QLKQINHQGEEEELI
	829	38	52	INHQGEEEELIALNL
	830	42	56	GEEEELIALNLSEAR
	831	46	60	ELIALNLSEARLVIK
	832	50	64	LNLSEARLVIKEALV
	833	54	68	EARLVIKEALVERRR
	834	58	72	VIKEALVERRRAFKR
	835	62	76	ALVERRRAFKRSQKK
	836	66	80	RRRAFKRSQKKHKKK
	837	70	84	FKRSQKKHKKKHLKH
	838	74	88	QKKHKKKHLKHENAN
	839	78	92	KKKHLKHENANDETT
	840	82	96	LKHENANDETTAVED
	841	86	100	NANDETTAVEDEDDD
	842	90	104	ETTAVEDEDDDLDED
	843	94	108	VEDEDDDLDEDDVNA
	844	98	112	DDDLDEDDVNADDDD
	845	102	116	DEDDVNADDDDFMHS
	846	106	120	VNADDDDFMHSETRE
	847	110	124	DDDFMHSETREKELE
	848	114	128	MHSETREKELESIDV
	849	118	132	TREKELESIDVLLEQ
	850	122	136	ELESIDVLLEQTTGG
	851	126	140	IDVLLEQTTGGNNKD
	852	130	144	LEQTTGGNNKDLKNT
	853	134	148	TGGNNKDLKNTMQYL
	854	138	152	NKDLKNTMQYLTNFS
	855	142	156	KNTMQYLTNFSRFRD
	856	146	160	QYLTNFSRFRDQETV
	857	150	164	NFSRFRDQETVGAVI
	858	154	168	FRDQETVGAVIQLLK
	859	158	172	ETVGAVIQLLKSTGL
	860	162	176	AVIQLLKSTGLHPFE
	861	166	180	LLKSTGLHPFEVAQL
	862	170	184	TGLHPFEVAQLGSLA
	863	174	188	PFEVAQLGSLACDTA
	864	178	192	AQLGSLACDTADEAK
	865	182	196	SLACDTADEAKTLIP
	866	186	200	DTADEAKTLIPSLNN
	867	190	204	EAKTLIPSLNNKISD
	868	194	208	LIPSLNNKISDDELE
	869	198	212	LNNKISDDELERILK
	870	202	216	ISDDELERILKELSN
	871	206	220	ELERILKELSNLETL
Rpb4/5	872			ILKELSNLETLYMDQ
	873			LSNLETLYMDQENER
	874			ETLYMDQENERNISR
Rpb5	875	1	15	MDQENERNISRLWRA
	876	5	19	NERNISRLWRAFRTV
	877	9	23	ISRLWRAFRTVKEMV
	878	13	27	WRAFRTVKEMVKDRG
	879	17	31	RTVKEMVKDRGYFIT
	880	21	35	EMVKDRGYFITQEEV
	881	25	39	DRGYFITQEEVELPL
	882	29	43	FITQEEVELPLEDFK
	883	33	47	EEVELPLEDFKAKYC
	884	37	51	LPLEDFKAKYCDSMG
	885	41	55	DFKAKYCDSMGRPQR
	886	45	59	KYCDSMGRPQRKMMS
	887	49	63	SMGRPQRKMMSFQAN
	888	53	67	PQRKMMSFQANPTEE
	889	57	71	MMSFQANPTEESISK
	890	61	75	QANPTEESISKFPDM
	891	65	79	TEESISKFPDMGSLW
	892	69	83	ISKFPDMGSLWVEFC
	893	73	87	PDMGSLWVEFCDEPS
	894	77	91	SLWVEFCDEPSVGVK
	895	81	95	EFCDEPSVGVKTMKT
	896	85	99	EPSVGVKTMKTFVIH
	897	89	103	GVKTMKTFVIHIQEK
	898	93	107	MKTFVIHIQEKNFQT
	899	97	111	VIHIQEKNFQTGIFV
	900	101	115	QEKNFQTGIFVYQNN
	901	105	119	FQTGIFVYQNNITPS
	902	109	123	IFVYQNNITPSAMKL
	903	113	127	QNNITPSAMKLVPSI
	904	117	131	TPSAMKLVPSIPPAT
	905	121	135	MKLVPSIPPATIETF
	906	125	139	PSIPPATIETFNEAA
	907	129	143	PATIETFNEAALVVN
	908	133	147	ETFNEAALVVNITHH
	909	137	151	EAALVVNITHHELVP
	910	141	155	VVNITHHELVPKHIR

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
	911	145	159	THHELVPKHIRLSSD
	912	149	163	LVPKHIRLSSDEKRE
	913	153	167	HIRLSSDEKRELLKR
	914	157	171	SSDEKRELLKRYRLK
	915	161	175	KRELLKRYRLKESQL
	916	165	179	LKRYRLKESQLPRIQ
	917	169	183	RLKESQLPRIQRADP
	918	173	187	SQLPRIQRADPVALY
	919	177	191	RIQRADPVALYLGLK
	920	181	195	ADPVALYLGLKRGEV
	921	185	199	ALYLGLKRGEVVKII
	922	189	203	GLKRGEVVKIIRKSE
	923	193	207	GEVVKIIRKSETSGR
	924	197	211	KIIRKSETSGRYASY
	925	201	215	KSETSGRYASYRICM
Rpb5/6	926			SGRYASYRICMMSDY
	927			ASYRICMMSDYEEAF
	928			ICMMSDYEEAFNDGN
Rpb6	929	2	16	SDYEEAFNDGNENFE
	930	6	20	EAFNDGNENFEDFDV
	931	10	24	DGNENFEDFDVEHFS
	932	14	28	NFEDFDVEHFSDEET
	933	18	32	FDVEHFSDEETYEEK
	934	22	36	HFSDEETYEEKPQFK
	935	26	40	EETYEEKPQFKDGET
	936	30	44	EEKPQFKDGETTDAN
	937	34	48	QFKDGETTDANGKTI
	938	38	52	GETTDANGKTIVTGG
	939	42	56	DANGKTIVTGGNGPE
	940	46	60	KTIVTGGNGPEDFQQ
	941	50	64	TGGNGPEDFQQHEQI
	942	54	68	GPEDFQQHEQIRRKT
	943	58	72	FQQHEQIRRKTLKEK
	944	62	76	EQIRRKTLKEKAIPK
	945	66	80	RKTLKEKAIPKDQRA
	946	70	84	KEKAIPKDQRATTPY
	947	74	88	IPKDQRATTPYMTKY
	948	78	92	QRATTPYMTKYERAR
	949	82	96	TPYMTKYERARILGT
	950	86	100	TKYERARILGTRALQ
	951	90	104	RARILGTRALQISMN
	952	94	108	LGTRALQISMNAPVF
	953	98	112	ALQISMNAPVFVDLE
	954	102	116	SMNAPVFVDLEGETD
	955	106	120	PVFVDLEGETDPLRI
	956	110	124	DLEGETDPLRIAMKE

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
	957	114	128	ETDPLRIAMKELAEK
	958	118	132	LRIAMKELAEKKIPL
	959	122	136	MKELAEKKIPLVIRR
	960	126	140	AEKKIPLVIRRYLPD
	961	130	144	IPLVIRRYLPDGSFE
	962	134	148	IRRYLPDGSFEDWSV
	963	138	152	LPDGSFEDWSVEELI
Rpb6/7	964			SFEDWSVEELIVDLM
	965			WSVEELIVDLMFFIK
	966			ELIVDLMFFIKDLSL
	967			DLMFFIKDLSLNITL
Rpb7	968	3	17	FIKDLSLNITLHPSF
	969	7	21	LSLNITLHPSFFGPR
	970	11	25	ITLHPSFFGPRMKQY
	971	15	29	PSFFGPRMKQYLKTK
	972	19	33	GPRMKQYLKTKLLEE
	973	23	37	KQYLKTKLLEEVEGS
	974	27	41	KTKLLEEVEGSCTGK
	975	31	45	LEEVEGSCTGKFGYI
	976	35	49	EGSCTGKFGYILCVL
	977	39	53	TGKFGYILCVLDYDN
	978	43	57	GYILCVLDYDNIDIQ
	979	47	61	CVLDYDNIDIQRGRI
	980	51	65	YDNIDIQRGRILPTD
	981	55	69	DIQRGRILPTDGSAE
	982	59	73	GRILPTDGSAEFNVK
	983	63	77	PTDGSAEFNVKYRAV
	984	67	81	SAEFNVKYRAVVFKP
	985	71	85	NVKYRAVVFKPFKGE
	986	75	89	RAVVFKPFKGEVVDG
	987	79	93	FKPFKGEVVDGTVVS
	988	83	97	KGEVVDGTVVSCSQH
	989	87	101	VDGTVVSCSQHGFEV
	990	91	105	VVSCSQHGFEVQVGP
	991	95	109	SQHGFEVQVGPMKVF
	992	99	113	FEVQVGPMKVFVTKH
	993	103	117	VGPMKVFVTKHLMPQ
	994	107	121	KVFVTKHLMPQDLTF
	995	111	125	TKHLMPQDLTFNAGS
	996	115	129	MPQDLTFNAGSNPPS
	997	119	133	LTFNAGSNPPSYQSS
	998	123	137	AGSNPPSYQSSEDVI
	999	127	141	PPSYQSSEDVITIKS
	1000	131	145	QSSEDVITIKSRIRV
	1001	135	149	DVITIKSRIRVKIEG
	1002	139	153	IKSRIRVKIEGCISQ

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
	1003	143	157	IRVKIEGCISQVSSI
	1004	147	161	IEGCISQVSSIHAIG
	1005	151	165	ISQVSSIHAIGSIKE
	1006	155	169	SSIHAIGSIKEDYLG
Rpb7/8	1007			AIGSIKEDYLGAIMS
	1008			IKEDYLGAIMSNTLF
	1009			YLGAIMSNTLFDDIF
	1010			IMSNTLFDDIFQVSE
Rpb8	1011	4	18	TLFDDIFQVSEVDPG
	1012	8	22	DIFQVSEVDPGRYNK
	1013	12	26	VSEVDPGRYNKVCRI
	1014	16	30	DPGRYNKVCRIEAAS
	1015	20	34	YNKVCRIEAASTTQD
	1016	24	38	CRIEAASTTQDQCKL
	1017	28	42	AASTTQDQCKLTLDI
	1018	32	46	TQDQCKLTLDINVEL
	1019	36	50	CKLTLDINVELFPVA
	1020	40	54	LDINVELFPVAAQDS
	1021	44	58	VELFPVAAQDSLTVT
	1022	48	62	PVAAQDSLTVTIASS
	1023	52	66	QDSLTVTIASSLNLE
	1024	56	70	TVTIASSLNLEDTPA
	1025	60	74	ASSLNLEDTPANDSS
	1026	64	78	NLEDTPANDSSATRS
	1027	68	82	TPANDSSATRSWRPP
	1028	72	86	DSSATRSWRPPQAGD
	1029	76	90	TRSWRPPQAGDRSLA
	1030	80	94	RPPQAGDRSLADDYD
	1031	84	98	AGDRSLADDYDYVMY
	1032	88	102	SLADDYDYVMYGTAY
	1033	92	106	DYDYVMYGTAYKFEE
	1034	96	110	VMYGTAYKFEEVSKD
	1035	100	114	TAYKFEEVSKDLIAV
	1036	104	118	FEEVSKDLIAVYYSF
	1037	108	122	SKDLIAVYYSFGGLL
	1038	112	126	IAVYYSFGGLLMRLE
	1039	116	130	YSFGGLLMRLEGNYR
	1040	120	134	GLLMRLEGNYRNLNN
	1041	124	138	RLEGNYRNLNNLKQE
	1042	128	142	NYRNLNNLKQENAYL
	1043	132	146	LNNLKQENAYLLIRR
Rpb8/9	1044			KQENAYLLIRRMTTF
	1045			AYLLIRRMTTFRFCR
	1046			IRRMTTFRFCRDCNN
Rpb9	1047	2	16	TTFRFCRDCNNMLYP
	1048	6	20	FCRDCNNMLYPREDK
	1049	10	24	CNNMLYPREDKENNR
	1050	14	28	LYPREDKENNRLLFE
	1051	18	32	EDKENNRLLFECRTC
	1052	22	36	NNRLLFECRTCSYVE
	1053	26	40	LFECRTCSYVEEAGS
	1054	30	44	RTCSYVEEAGSPLVY
	1055	34	48	YVEEAGSPLVYRHEL
	1056	38	52	AGSPLVYRHELITNI
	1057	42	56	LVYRHELITNIGETA
	1058	46	60	HELITNIGETAGVVQ
	1059	50	64	TNIGETAGVVQDIGS
	1060	54	68	ETAGVVQDIGSDPTL
	1061	58	72	VVQDIGSDPTLPRSD
	1062	62	76	IGSDPTLPRSDRECP
	1063	66	80	PTLPRSDRECPKCHS
	1064	70	84	RSDRECPKCHSRENV
	1065	74	88	ECPKCHSRENVFFQS
	1066	78	92	CHSRENVFFQSQQRR
	1067	82	96	ENVFFQSQQRRKDTS
	1068	86	100	FQSQQRRKDTSMVLF
	1069	90	104	QRRKDTSMVLFFVCL
	1070	94	108	DTSMVLFFVCLSCSH
	1071	98	112	VLFFVCLSCSHIFTS
	1072	102	116	VCLSCSHIFTSDQKN
	1073	106	120	CSHIFTSDQKNKRTQ
Rpb9/10	1074			FTSDQKNKRTQFSMI
	1075			QKNKRTQFSMIVPVR
	1076			RTQFSMIVPVRCFSC
	1077			SMIVPVRCFSCGKVV
Rpb10	1078	4	18	PVRCFSCGKVVGDKW
	1079	8	22	FSCGKVVGDKWESYL
	1080	12	26	KVVGDKWESYLNLLQ
	1081	16	30	DKWESYLNLLQEDEL
	1082	20	34	SYLNLLQEDELDEGT
	1083	24	38	LLQEDELDEGTALSR
	1084	28	42	DELDEGTALSRLGLK
	1085	32	46	EGTALSRLGLKRYCC
	1086	36	50	LSRLGLKRYCCRRMI
	1087	40	54	GLKRYCCRRMILTHV
	1088	44	58	YCCRRMILTHVDLIE
	1089	48	62	RMILTHVDLIEKFLR
	1090	52	66	THVDLIEKFLRYNPL
	1091	56	70	LIEKFLRYNPLEKRD
Rpb10/11	1092			FLRYNPLEKRDMNAP
	1093			NPLEKRDMNAPDRFE
	1094			KRDMNAPDRFELFLL

Pol II subunit	Peptide No.	Start a.a.	End a.a.	Sequence
Rpb11	1095	2	16	NAPDRFELFLLGEGE
	1096	6	20	RFELFLLGEGESKLK
	1097	10	24	FLLGEGESKLKIDPD
	1098	14	28	EGESKLKIDPDTKAP
	1099	18	32	KLKIDPDTKAPNAVV
	1100	22	36	DPDTKAPNAVVITFE
	1101	26	40	KAPNAVVITFEKEDH
	1102	30	44	AVVITFEKEDHTLGN
	1103	34	48	TFEKEDHTLGNLIRA
	1104	38	52	EDHTLGNLIRAELLN
	1105	42	56	LGNLIRAELLNDRKV
	1106	46	60	IRAELLNDRKVLFAA
	1107	50	64	LLNDRKVLFAAYKVE
	1108	54	68	RKVLFAAYKVEHPFF
	1109	58	72	FAAYKVEHPFFARFK
	1110	62	76	KVEHPFFARFKLRIQ
	1111	66	80	PFFARFKLRIQTTEG
	1112	70	84	RFKLRIQTTEGYDPK
	1113	74	88	RIQTTEGYDPKDALK
	1114	78	92	TEGYDPKDALKNACN
	1115	82	96	DPKDALKNACNSIIN
	1116	86	100	ALKNACNSIINKLGA
	1117	90	104	ACNSIINKLGALKTN
	1118	94	108	IINKLGALKTNFETE
	1119	98	112	LGALKTNFETEWNLQ
	1120	102	116	KTNFETEWNLQTLAA
	1121	106	120	ETEWNLQTLAADDAF
Rpb11/12	1122			NLQTLAADDAFMSRE
	1123			LAADDAFMSREGFQI
	1124			DAFMSREGFQIPTNL
Rpb12	1125	2	16	SREGFQIPTNLDAAA
	1126	6	20	FQIPTNLDAAAAGTS
	1127	10	24	TNLDAAAAGTSQART
	1128	14	28	AAAAGTSQARTATLK
	1129	18	32	GTSQARTATLKYICA
	1130	22	36	ARTATLKYICAECSS
	1131	26	40	TLKYICAECSSKLSL
	1132	30	44	ICAECSSKLSLSRTD
	1133	34	48	CSSKLSLSRTDAVRC
	1134	38	52	LSLSRTDAVRCKDCG
	1135	42	56	RTDAVRCKDCGHRIL
	1136	46	60	VRCKDCGHRILLKAR
	1137	50	64	DCGHRILLKARTKRL
	1138	54	68	RILLKARTKRLVQFE
	1139	56	70	LLKARTKRLVQFEAR


A
Rpb2-derived peptides
intensity distribution
peptide
sample intensity distribution
peptide
463
464
465
481
482
487
525
527
539
547
630
645
646
647
730
scale bar
GDP
GTP
GMPPCP
Buffer
Npa3 + GTP
Npa3 + GDP
Npa3 + GMPPCP
Buffer

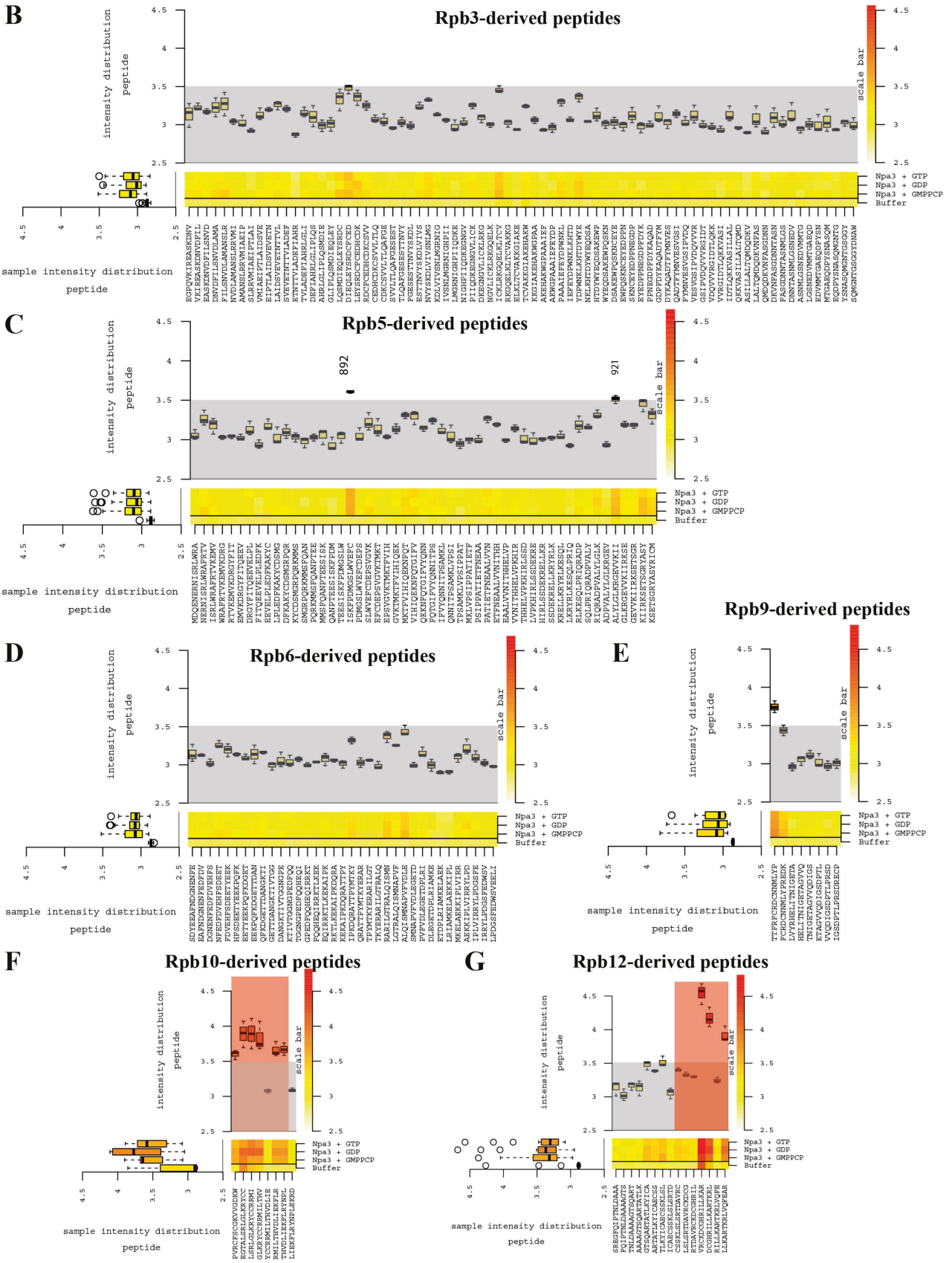


Figure 25| Analysis of Npa3 binding to peptides derived from Rpb2, Rpb3, Rpb5, Rpb6, Rpb9, Rpb11 and Rpb12.
Boxplot representation of the heatmap describing the Npa3 peptide-binding landscape in the presence of GTP, GDP and GMPPCP, respectively. Control experiments were performed without Npa3 and nucleotides to test cross-reactivity of the anti-His antibody. Intensity distribution is shown in logarithmic scale. Peptides with signal intensity <3.5 were defined as unbound (grey area). Red boxes indicate false-positive binding of the antibody to the peptides, either directly or in overlapping regions. These peptides were not used for further analysis. Peptide numbers are shown for Npa3 binding peptides above the boxplots.

6. References

Afonine PV, Grosse-Kunstleve RW and Adams PD (2005) A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr D Biol Crystallogr* 61 (Pt 7):850-5.

Aguilera A & Garcia-Muse T (2012) R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 46 (2):115-24.

Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J and Zhou H (2008) A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol Cell Proteomics* 7 (7):1389-96.

Ali MM, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C and Pearl LH (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* 440 (7087):1013-7.

Allen TA, Von Kaenel S, Goodrich JA and Kugel JF (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 11 (9):816-21.

Alonso B, Beraud C, Meguellati S, Chen SW, Pellequer JL, Armengaud J and Godon C (2013) Eukaryotic GPN-loop GTPases paralogs use a dimeric assembly reminiscent of archeal GPN. *Cell Cycle* 12 (3):463-72.

Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181 (4096):223-30.

Armache KJ, Mitterweger S, Meinhart A and Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* 280 (8):7131-4.

Aves SJ, Hunt C, Xiang Z, Lyne MH, Wood V, Rajandream MA, Skelton J, Churcher CM, Warren T, Harris D, Gwilliam R and Barrell BG (2002) The mei3 region of the Schizosaccharomyces pombe genome. *Yeast* 19 (6):521-7.

Avvakumov N, Nourani A and Cote J (2011) Histone chaperones: modulators of chromatin marks. *Mol Cell* 41 (5):502-14.

Beaudenon SL, Huacani MR, Wang G, McDonnell DP and Huibregtse JM (1999) Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in Saccharomyces cerevisiae. *Mol Cell Biol* 19 (10):6972-9.

Belle A, Tanay A, Bitincka L, Shamir R and O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* 103 (35):13004-9.

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI and Owens RJ (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* 35 (6):e45.

Bertelsen EB, Chang L, Gestwicki JE and Zuiderweg ER (2009) Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc Natl Acad Sci U S A* 106 (21):8471-6.

Bingel-Erlenmeyer R, Kohler R, Kramer G, Sandikci A, Antolic S, Maier T, Schaffitzel C, Wiedmann B, Bukau B and Ban N (2008) A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. *Nature* 452 (7183):108-11.

Boulon S, Marmier-Gourrier N, Pradet-Balade B, Wurth L, Verheggen C, Jady BE, Rothe B, Pescia C, Robert MC, Kiss T, Bardoni B, Krol A, Branlant C, Allmang C, Bertrand E and Charpentier B (2008) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J Cell Biol* 180 (3):579-95.

Boulon S, Pradet-Balade B, Verheggen C, Molle D, Boireau S, Georgieva M, Azzag K, Robert MC, Ahmad Y, Neel H, Lamond AI and Bertrand E (2010) HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol Cell* 39 (6):912-24.

Bourne HR, Sanders DA and McCormick F (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature* 349 (6305):117-27.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72 248-54.

Britton RA (2009) Role of GTPases in bacterial ribosome assembly. *Annu Rev Microbiol* 63 155-76.

Buchner J, Grallert H and Jakob U (1998) Analysis of chaperone function using citrate synthase as nonnative substrate protein. *Methods Enzymol* 290 323-38.

Bukau B, Deuerling E, Pfund C and Craig EA (2000) Getting newly synthesized proteins into shape. *Cell* 101 (2):119-22.

Bukau B & Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* 92 (3):351-66.

Bushnell DA, Westover KD, Davis RE and Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* 303 (5660):983-8.

Calera MR, Zamora-Ramos C, Araiza-Villanueva MG, Moreno-Aguilar CA, Pena-Gomez SG, Castellanos-Teran F, Robledo-Rivera AY and Sanchez-Olea R (2011) Parcs/Gpn3 is required for the nuclear accumulation of RNA polymerase II. *Biochim Biophys Acta* 1813 (10):1708-16.

Calloni G, Chen T, Schermann SM, Chang HC, Genevaux P, Agostini F, Tartaglia GG, Hayer-Hartl M and Hartl FU (2012) DnaK functions as a central hub in the E. coli chaperone network. *Cell Rep* 1 (3):251-64.

Carre C & Shiekhattar R (2011) Human GTPases associate with RNA polymerase II to mediate its nuclear import. *Mol Cell Biol* 31 (19):3953-62.

Caspar DL & Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27 1-24.

Chari A, Golas MM, Klingenhager M, Neuenkirchen N, Sander B, Englbrecht C, Sickmann A, Stark H and Fischer U (2008) An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell* 135 (3):497-509.

Cheetham GM & Steitz TA (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286 (5448):2305-9.

Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50 (Pt 5):760-3.

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, Jawhari A, Jennebach S, Kamenski T, Kettenberger H, Kuhn CD, Lehmann E, Leike K, Sydow JF and Vannini A (2008) Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* 37 337-52.

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR and Kornberg RD (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288 (5466):640-9.

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

Cuellar J, Martin-Benito J, Scheres SH, Sousa R, Moro F, Lopez-Vinas E, Gomez-Puertas P, Muga A, Carrascosa JL and Valpuesta JM (2008) The structure of CCT-Hsc70 NBD suggests a mechanism for Hsp70 delivery of substrates to the chaperonin. *Nat Struct Mol Biol* 15 (8):858-64.

Czeko E, Seizl M, Augsberger C, Mielke T and Cramer P (2011) Iwr1 directs RNA polymerase II nuclear import. *Mol Cell* 42 (2):261-6.

Daulny A, Geng F, Muratani M, Geisinger JM, Salghetti SE and Tansey WP (2008) Modulation of RNA polymerase II subunit composition by ubiquitylation. *Proc Natl Acad Sci U S A* 105 (50):19649-54.

Daulny A & Tansey WP (2009) Damage control: DNA repair, transcription, and the ubiquitin-proteasome system. *DNA Repair (Amst)* 8 (4):444-8.

Daumke O, Weyand M, Chakrabarti PP, Vetter IR and Wittinghofer A (2004) The GTPase-activating protein Rap1GAP uses a catalytic asparagine. *Nature* 429 (6988):197-201.

Davis MW & Hammarlund M (2006) Single-nucleotide polymorphism mapping. *Methods Mol Biol* 351 75-92.

De Koning L, Corpet A, Haber JE and Almouzni G (2007) Histone chaperones: an escort network regulating histone traffic. *Nat Struct Mol Biol* 14 (11):997-1007.

de Laat WL, Jaspers NG and Hoeijmakers JH (1999) Molecular mechanism of nucleotide excision repair. *Genes Dev* 13 (7):768-85.

Dekker C, Stirling PC, McCormack EA, Filmore H, Paul A, Brost RL, Costanzo M, Boone C, Leroux MR and Willison KR (2008) The interaction network of the chaperonin CCT. *EMBO J* 27 (13):1827-39.

del Alamo M, Hogan DJ, Pechmann S, Albanese V, Brown PO and Frydman J (2011) Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biol* 9 (7):e1001100.

Dez C, Froment C, Noaillac-Depeyre J, Monsarrat B, Caizergues-Ferrer M and Henry Y (2004) Npa1p, a component of very early pre-60S ribosomal particles, associates with a subset of small nucleolar RNPs required for peptidyl transferase center modification. *Mol Cell Biol* 24 (14):6324-37.

Dollins DE, Warren JJ, Immormino RM and Gewirth DT (2007) Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. *Mol Cell* 28 (1):41-56.

Drogemuller J, Strauss M, Schweimer K, Wohrl BM, Knauer SH and Rosch P (2015) Exploring RNA polymerase regulation by NMR spectroscopy. *Sci Rep* 5 10825.

Edwards AM, Kane CM, Young RA and Kornberg RD (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem* 266 (1):71-5.

Eick D & Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* 113 (11):8456-90.

Eisner G, Koch HG, Beck K, Brunner J and Muller M (2003) Ligand crowding at a nascent signal sequence. *J Cell Biol* 163 (1):35-44.

Ellis J (1987) Proteins as molecular chaperones. *Nature* 328 (6129):378-9.

Ellis RJ (2013) Assembly chaperones: a perspective. *Philos Trans R Soc Lond B Biol Sci* 368 (1617):20110398.

Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60 (Pt 12 Pt 1):2126-32.

Engel C, Sainsbury S, Cheung AC, Kostrewa D and Cramer P (2013) RNA polymerase I structure and transcription regulation. *Nature* 502 (7473):650-5.

Etchells SA, Meyer AS, Yam AY, Roobol A, Miao Y, Shao Y, Carden MJ, Skach WR, Frydman J and Johnson AE (2005) The cotranslational contacts between ribosome-bound nascent polypeptides and the subunits of the hetero-oligomeric chaperonin TRiC probed by photocross-linking. *J Biol Chem* 280 (30):28118-26.

Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E and Ban N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* 431 (7008):590-6.

Forget D, Lacombe AA, Cloutier P, Al-Khoury R, Bouchard A, Lavallee-Adam M, Faubert D, Jeronimo C, Blanchette M and Coulombe B (2010) The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1 and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. *Mol Cell Proteomics* 9 (12):2827-39.

Forget D, Lacombe AA, Cloutier P, Lavallee-Adam M, Blanchette M and Coulombe B (2013) Nuclear import of RNA polymerase II is coupled with nucleocytoplasmic shuttling of the RNA polymerase II-associated protein 2. *Nucleic Acids Res* 41 (14):6881-91.

Fornerod M, Ohno M, Yoshida M and Mattaj IW (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* 90 (6):1051-60.

Freymann DM, Keenan RJ, Stroud RM and Walter P (1997) Structure of the conserved GTPase domain of the signal recognition particle. *Nature* 385 (6614):361-4.

Frydman J (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* 70 603-47.

Garrido-Godino AI, Garcia-Lopez MC and Navarro F (2013) Correct assembly of RNA polymerase II depends on the foot domain and is required for multiple steps of transcription in Saccharomyces cerevisiae. *Mol Cell Biol* 33 (18):3611-26.

Gasteiger E HC, Gattiker A, Duvaud S, Wilkins M R, Appel R D and Bairoch A (2005) Protein Identification and Analysis Tools on ExPASy Server. *Humana Press*

Gautschi M, Mun A, Ross S and Rospert S (2002) A functional chaperone triad on the yeast ribosome. *Proc Natl Acad Sci U S A* 99 (7):4209-14.

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G and Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415 (6868):141-7.

Genevaux P, Keppel F, Schwager F, Langendijk-Genevaux PS, Hartl FU and Georgopoulos C (2004) In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep* 5 (2):195-200.

Ghosh P, Ishihama A and Chatterji D (2001) Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the alpha2beta subassembly. *Eur J Biochem* 268 (17):4621-7.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW and Johnston M (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature* 418 (6896):387-91.

Gomez-Navarro N, Peiro-Chova L, Rodriguez-Navarro S, Polaina J and Estruch F (2013) Rtp1p is a karyopherin-like protein required for RNA polymerase II biogenesis. *Mol Cell Biol* 33 (9):1756-67.

Gouet P, Courcelle E, Stuart DI and Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15 (4):305-8.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J and Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38 (Web Server issue):W695-9.

Gras S, Chaumont V, Fernandez B, Carpentier P, Charrier-Savournin F, Schmitt S, Pineau C, Flament D, Hecker A, Forterre P, Armengaud J and Housset D (2007) Structural insights into a new homodimeric self-activated GTPase family. *EMBO Rep* 8 (6):569-75.

Gras S, Fernandez B, Chaumont V, Carpentier P, Armengaud J and Housset D (2005) Expression, purification, crystallization and preliminary crystallographic analysis of the PAB0955 gene product. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 61 (Pt 2):208-11.

Guglielmi V, Marini M, Masson EF, Malatesta M, Forget D, Tomelleri G, Coulombe B and Vattemi G (2015) Abnormal expression of RNA polymerase II-associated proteins in muscle of patients with myofibrillar myopathies. *Histopathology*

Hainzl O, Lapina MC, Buchner J and Richter K (2009) The charged linker region is an important regulator of Hsp90 function. *J Biol Chem* 284 (34):22559-67.

Harreman M, Taschner M, Sigurdsson S, Anindya R, Reid J, Somesh B, Kong SE, Banks CA, Conaway RC, Conaway JW and Svejstrup JQ (2009) Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc Natl Acad Sci U S A* 106 (49):20705-10.

Harris SF, Shiau AK and Agard DA (2004) The crystal structure of the carboxy-terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site. *Structure* 12 (6):1087-97.

Hartl FU (1996) Molecular chaperones in cellular protein folding. *Nature* 381 (6583):571-9.

Hartl FU, Bracher A and Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. *Nature* 475 (7356):324-32.

Hartl FU & Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295 (5561):1852-8.

Hartl FU & Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* 16 (6):574-81.

Hauser T, Bhat JY, Milicic G, Wendler P, Hartl FU, Bracher A and Hayer-Hartl M (2015) Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat Struct Mol Biol*

Higuchi R, Krummel B and Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* 16 (15):7351-67.

Holt LJ, Tuch BB, Villen J, Johnson AD, Gygi SP and Morgan DO (2009) Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* 325 (5948):1682-6.

Horwich AL, Fenton WA, Chapman E and Farr GW (2007) Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol* 23 115-45.

Huet J, Buhler JM, Sentenac A and Fromageot P (1975) Dissociation of two polypeptide chains from yeast RNA polymerase A. *Proc Natl Acad Sci U S A* 72 (8):3034-8.

Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS and O'Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* 425 (6959):686-91.

Huibregtse JM, Yang JC and Beaudenon SL (1997) The large subunit of RNA polymerase II is a substrate of the Rsp5 ubiquitin-protein ligase. *Proc Natl Acad Sci U S A* 94 (8):3656-61.

Ishihama A (1981) Subunit of assembly of Escherichia coli RNA polymerase. *Adv Biophys* 14 1-35.

Ishihama A & Ito K (1972) Subunits of RNA polymerase in function and structure. II. Reconstitution of Escherichia coli RNA polymerase from isolated subunits. *J Mol Biol* 72 (1):111-23.

Jasiak AJ, Armache KJ, Martens B, Jansen RP and Cramer P (2006) Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* 23 (1):71-81.

Jasnovidova O & Stefl R (2013) The CTD code of RNA polymerase II: a structural view. *Wiley Interdiscip Rev RNA* 4 (1):1-16.

Jeronimo C, Forget D, Bouchard A, Li Q, Chua G, Poitras C, Therien C, Bergeron D, Bourassa S, Greenblatt J, Chabot B, Poirier GG, Hughes TR, Blanchette M, Price DH and Coulombe B (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol Cell* 27 (2):262-74.

Jeronimo C, Langelier MF, Zeghouf M, Cojocaru M, Bergeron D, Baali D, Forget D, Mnaimneh S, Davierwala AP, Pootoolal J, Chandy M, Canadien V, Beattie BK, Richards DP, Workman JL, Hughes TR, Greenblatt J and Coulombe B (2004) RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits. *Mol Cell Biol* 24 (16):7043-58.

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292 (2):195-202.

Jung Y & Lippard SJ (2006) RNA polymerase II blockage by cisplatin-damaged DNA. Stability and polyubiquitylation of stalled polymerase. *J Biol Chem* 281 (3):1361-70.

Kabsch W (2010a) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66 (Pt 2):133-44.

Kabsch W (2010b) Xds. *Acta Crystallogr D Biol Crystallogr* 66 (Pt 2):125-32.

Kaiser CM, Chang HC, Agashe VR, Lakshmipathy SK, Etchells SA, Hayer-Hartl M, Hartl FU and Barral JM (2006) Real-time observation of trigger factor function on translating ribosomes. *Nature* 444 (7118):455-60.

Karbstein K (2010) Chaperoning ribosome assembly. *J Cell Biol* 189 (1):11-2.

Karplus PA & Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336 (6084):1030-3.

Keniry ME, Kemp HA, Rivers DM and Sprague GF, Jr. (2004) The identification of Pcl1-interacting proteins that genetically interact with Cla4 may indicate a link between G1 progression and mitotic exit. *Genetics* 166 (3):1177-86.

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC and Bentley DL (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 17 (10):1279-86.

Kim JL, Nikolov DB and Burley SK (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365 (6446):520-7.

Kim Y, Geiger JH, Hahn S and Sigler PB (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365 (6446):512-20.

Kim YE, Hipp MS, Bracher A, Hayer-Hartl M and Hartl FU (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* 82 323-55.

Kimura M, Ishiguro A and Ishihama A (1997) RNA polymerase II subunits 2, 3, and 11 form a core subassembly with DNA binding activity. *J Biol Chem* 272 (41):25851-5.

Kimura M, Sakurai H and Ishihama A (2001) Intracellular contents and assembly states of all 12 subunits of the RNA polymerase II in the fission yeast Schizosaccharomyces pombe. *Eur J Biochem* 268 (3):612-9.

Kityk R, Kopp J, Sinning I and Mayer MP (2012) Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Mol Cell* 48 (6):863-74.

Kolodziej PA & Young RA (1991) Mutations in the three largest subunits of yeast RNA polymerase II that affect enzyme assembly. *Mol Cell Biol* 11 (9):4669-78.

Koplin A, Preissler S, Ilina Y, Koch M, Scior A, Erhardt M and Deuerling E (2010) A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. *J Cell Biol* 189 (1):57-68.

Kostrewa D, Zeller ME, Armache KJ, Seizl M, Leike K, Thomm M and Cramer P (2009) RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462 (7271):323-30.

Koulov AV, LaPointe P, Lu B, Razvi A, Coppinger J, Dong MQ, Matteson J, Laister R, Arrowsmith C, Yates JR, 3rd and Balch WE (2010) Biological and structural basis for Aha1 regulation of Hsp90 ATPase activity in maintaining proteostasis in the human disease cystic fibrosis. *Mol Biol Cell* 21 (6):871-84.

Kramer G, Rauch T, Rist W, Vorderwulbecke S, Patzelt H, Schulze-Specking A, Ban N, Deuerling E and Bukau B (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature* 419 (6903):171-4.

Kressler D, Hurt E and Bassler J (2010) Driving ribosome assembly. *Biochim Biophys Acta* 1803 (6):673-83.

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A and Greenblatt JF (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* 440 (7084):637-43.

Krukenberg KA, Forster F, Rice LM, Sali A and Agard DA (2008) Multiple conformations of E. coli Hsp90 in solution: insights into the conformational dynamics of Hsp90. *Structure* 16 (5):755-65.

Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S and Stefl R (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26 (17):1891-6.

Kuhlbrandt W (2014) Biochemistry. The resolution revolution. *Science* 343 (6178):1443-4.

Kuhle B & Ficner R (2014) A monovalent cation acts as structural and catalytic cofactor in translational GTPases. *EMBO J* 33 (21):2547-63.

Lalo D, Carles C, Sentenac A and Thuriaux P (1993) Interactions between three common subunits of yeast RNA polymerases I and III. *Proc Natl Acad Sci U S A* 90 (12):5524-8.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23 (21):2947-8.

Laskey RA, Honda BM, Mills AD and Finch JT (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275 (5679):416-20.

Lehmann E, Brueckner F and Cramer P (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* 450 (7168):445-9.

Leipe DD, Wolf YI, Koonin EV and Aravind L (2002) Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol* 317 (1):41-72.

Lembo F, Pero R, Angrisano T, Vitiello C, Iuliano R, Bruni CB and Chiariotti L (2003) MBDin, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. *Mol Cell Biol* 23 (5):1656-65.

Leroux MR & Hartl FU (2000) Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. *Curr Biol* 10 (7):R260-4.

Li J, Soroka J and Buchner J (2012) The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim Biophys Acta* 1823 (3):624-35.

Li Z, Musich PR, Cartwright BM, Wang H and Zou Y (2013) UV-induced nuclear import of XPA is mediated by importin-alpha4 in an ATR-dependent manner. *PLoS One* 8 (7):e68297.

Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV, Berninghausen O, Mielke T, Hartl FU, Beckmann R and Hayer-Hartl M (2010) Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* 463 (7278):197-202.

Liu X, Bushnell DA, Wang D, Calero G and Kornberg RD (2010) Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* 327 (5962):206-9.

Lorenzen K, Vannini A, Cramer P and Heck AJ (2007) Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* 15 (10):1237-45.

Makhnevych T & Houry WA (2012) The role of Hsp90 in protein complex assembly. *Biochim Biophys Acta* 1823 (3):674-82.

Mao J, Chi W, Ouyang M, He B, Chen F and Zhang L (2015) PAB is an assembly chaperone that functions downstream of chaperonin 60 in the assembly of chloroplast ATP synthase coupling factor 1. *Proc Natl Acad Sci U S A* 112 (13):4152-7.

Mapa K, Sikor M, Kudryavtsev V, Waegemann K, Kalinin S, Seidel CA, Neupert W, Lamb DC and Mokranjac D (2010) The conformational dynamics of the mitochondrial Hsp70 chaperone. *Mol Cell* 38 (1):89-100.

Matzke M, Kanno T, Daxinger L, Huettel B and Matzke AJ (2009) RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* 21 (3):367-76.

Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J and Cramer P (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17 (10):1272-8.

Mayer MP (2010) Gymnastics of molecular chaperones. *Mol Cell* 39 (3):321-31.

McClellan AJ, Xia Y, Deutschbauer AM, Davis RW, Gerstein M and Frydman J (2007) Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* 131 (1):121-35.

McCoy AJ, Grosse-Kunstleve RW, Storoni LC and Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61 (Pt 4):458-64.

McLaughlin SH, Smith HW and Jackson SE (2002) Stimulation of the weak ATPase activity of human hsp90 by a client protein. *J Mol Biol* 315 (4):787-98.

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP and Lopez R (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* 41 (Web Server issue):W597-600.

Mendez-Hernandez LE, Perez-Mejia AE, Lara-Chacon B, Barbosa-Camacho AA, Pena-Gomez SG, Martinez-Sanchez M, Robledo-Rivera AY, Sanchez-Olea R and Calera MR (2014) Gpn1 and Gpn3 associate tightly and their protein levels are mutually dependent in mammalian cells. *FEBS Lett* 588 (21):3823-9.

Merz F, Boehringer D, Schaffitzel C, Preissler S, Hoffmann A, Maier T, Rutkowska A, Lozza J, Ban N, Bukau B and Deuerling E (2008) Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *EMBO J* 27 (11):1622-32.

Meyer P, Prodromou C, Hu B, Vaughan C, Roe SM, Panaretou B, Piper PW and Pearl LH (2003) Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol Cell* 11 (3):647-58.

Minaker SW, Filiatrault MC, Ben-Aroya S, Hieter P and Stirling PC (2013) Biogenesis of RNA polymerases II and III requires the conserved GPN small GTPases in Saccharomyces cerevisiae. *Genetics* 193 (3):853-64.

Minakhin L, Bhagat S, Brunning A, Campbell EA, Darst SA, Ebright RH and Severinov K (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* 98 (3):892-7.

Miron-Garcia MC, Garrido-Godino AI, Garcia-Molinero V, Hernandez-Torres F, Rodriguez-Navarro S and Navarro F (2013) The prefoldin bud27 mediates the assembly of the eukaryotic RNA polymerases in an rpb5-dependent manner. *PLoS Genet* 9 (2):e1003297.

Murata S, Yashiroda H and Tanaka K (2009) Molecular mechanisms of proteasome assembly. *Nat Rev Mol Cell Biol* 10 (2):104-15.

Nguyen VT, Giannoni F, Dubois MF, Seo SJ, Vigneron M, Kedinger C and Bensaude O (1996) In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Res* 24 (15):2924-9.

Nitta M, Saijo M, Kodo N, Matsuda T, Nakatsu Y, Tamai H and Tanaka K (2000) A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. *Nucleic Acids Res* 28 (21):4212-8.

Niwa T, Kanamori T, Ueda T and Taguchi H (2012) Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc Natl Acad Sci U S A* 109 (23):8937-42.

Nonet M, Scafe C, Sexton J and Young R (1987) Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol* 7 (5):1602-11.

Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, Nichols RJ, Typas A, Gross CA, Kramer G, Weissman JS and Bukau B (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* 147 (6):1295-308.

Pan X, Eathiraj S, Munson M and Lambright DG (2006) TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. *Nature* 442 (7100):303-6.

Pearl LH & Prodromou C (2006) Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu Rev Biochem* 75 271-94.

Pech M, Spreter T, Beckmann R and Beatrix B (2010) Dual binding mode of the nascent polypeptide-associated complex reveals a novel universal adapter site on the ribosome. *J Biol Chem* 285 (25):19679-87.

Peiro-Chova L & Estruch F (2009) The yeast RNA polymerase II-associated factor Iwr1p is involved in the basal and regulated transcription of specific genes. *J Biol Chem* 284 (42):28958-67.

Peisker K, Braun D, Wolfle T, Hentschel J, Funfschilling U, Fischer G, Sickmann A and Rospert S (2008) Ribosome-associated complex binds to ribosomes in close proximity of Rpl31 at the exit of the polypeptide tunnel in yeast. *Mol Biol Cell* 19 (12):5279-88.

Preissler S & Deuerling E (2012) Ribosome-associated chaperones as key players in proteostasis. *Trends Biochem Sci* 37 (7):274-83.

Prive GG, Milburn MV, Tong L, de Vos AM, Yamaizumi Z, Nishimura S and Kim SH (1992) X-ray crystal structures of transforming p21 ras mutants suggest a transition-state stabilization mechanism for GTP hydrolysis. *Proc Natl Acad Sci U S A* 89 (8):3649-53.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M and Seraphin B (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24 (3):218-29.

Ramirez UD, Minasov G, Focia PJ, Stroud RM, Walter P, Kuhn P and Freymann DM (2002) Structural basis for mobility in the 1.1 A crystal structure of the NG domain of Thermus aquaticus Ffh. *J Mol Biol* 320 (4):783-99.

Raue U, Oellerer S and Rospert S (2007) Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence. *J Biol Chem* 282 (11):7809-16.

Retzlaff M, Hagn F, Mitschke L, Hessling M, Gugel F, Kessler H, Richter K and Buchner J (2010) Asymmetric activation of the hsp90 dimer by its cochaperone aha1. *Mol Cell* 37 (3):344-54.

Reyes-Pardo H, Barbosa-Camacho AA, Perez-Mejia AE, Lara-Chacon B, Salas-Estrada LA, Robledo-Rivera AY, Montero-Moran GM, Lara-Gonzalez S, Calera MR and Sanchez-Olea R (2012) A nuclear export sequence in GPN-loop GTPase 1, an essential protein for nuclear targeting of RNA polymerase II, is necessary and sufficient for nuclear export. *Biochim Biophys Acta* 1823 (10):1756-66.

Ribar B, Prakash L and Prakash S (2007) ELA1 and CUL3 are required along with ELC1 for RNA polymerase II polyubiquitylation and degradation in DNA-damaged yeast cells. *Mol Cell Biol* 27 (8):3211-6.

Roe SM, Ali MM, Meyer P, Vaughan CK, Panaretou B, Piper PW, Prodromou C and Pearl LH (2004) The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). *Cell* 116 (1):87-98.

Rubbi L, Labarre-Mariotte S, Chedin S and Thuriaux P (1999) Functional characterization of ABC10alpha, an essential polypeptide shared by all three forms of eukaryotic DNA-dependent RNA polymerases. *J Biol Chem* 274 (44):31485-92.

Russmann F, Stemp MJ, Monkemeyer L, Etchells SA, Bracher A and Hartl FU (2012) Folding of large multidomain proteins by partial encapsulation in the chaperonin TRiC/CCT. *Proc Natl Acad Sci U S A* 109 (52):21208-15.

Sainsbury S, Bernecky C and Cramer P (2015) Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* 16 (3):129-43.

Sainsbury S, Niesser J and Cramer P (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 493 (7432):437-40.

Sanchez-Olea R, Ortiz S, Barreto O, Yang Q, Xu CJ, Zhu H and Yuan J (2008) Parcs is a dual regulator of cell proliferation and apaf-1 function. *J Biol Chem* 283 (36):24400-5.

Sardiu ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, Florens L and Washburn MP (2008) Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc Natl Acad Sci U S A* 105 (5):1454-9.

Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU and Hayer-Hartl M (2007) Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* 129 (6):1189-200.

Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F and Wittinghofer A (1997) The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* 277 (5324):333-8.

Schuette JC, Murphy FVt, Kelley AC, Weir JR, Giesebrecht J, Connell SR, Loerke J, Mielke T, Zhang W, Penczek PA, Ramakrishnan V and Spahn CM (2009) GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J* 28 (6):755-65.

Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64 (Pt 1):112-22.

Shiau AK, Harris SF, Southworth DR and Agard DA (2006) Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* 127 (2):329-40.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD and Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7 539.

Smock RG, Rivoire O, Russ WP, Swain JF, Leibler S, Ranganathan R and Gierasch LM (2010) An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Mol Syst Biol* 6 414.

Soding J, Biegert A and Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33 (Web Server issue):W244-8.

Somesh BP, Reid J, Liu WF, Sogaard TM, Erdjument-Bromage H, Tempst P and Svejstrup JQ (2005) Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell* 121 (6):913-23.

Somesh BP, Sigurdsson S, Saeki H, Erdjument-Bromage H, Tempst P and Svejstrup JQ (2007) Communication between distant sites in RNA polymerase II through ubiquitylation factors and the polymerase CTD. *Cell* 129 (1):57-68.

Southworth DR & Agard DA (2011) Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90:Hop complex. *Mol Cell* 42 (6):771-81.

Staresincic L, Walker J, Dirac-Svejstrup AB, Mitter R and Svejstrup JQ (2011) GTP-dependent binding and nuclear transport of RNA polymerase II by Npa3 protein. *J Biol Chem* 286 (41):35553-61.

Svejstrup JQ (2004) The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim Biophys Acta* 1677 (1-3):64-73.

Swain JF, Dinler G, Sivendran R, Montgomery DL, Stotz M and Gierasch LM (2007) Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Mol Cell* 26 (1):27-39.

Taipale M, Jarosz DF and Lindquist S (2010) HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol* 11 (7):515-28.

Tang YC, Chang HC, Hayer-Hartl M and Hartl FU (2007) SnapShot: molecular chaperones, Part II. *Cell* 128 (2):412.

Te J, Jia L, Rogers J, Miller A and Hartson SD (2007) Novel subunits of the mammalian Hsp90 signal transduction chaperone. *J Proteome Res* 6 (5):1963-73.

Tsutsumi S, Mollapour M, Graf C, Lee CT, Scroggins BT, Xu W, Haslerova L, Hessling M, Konstantinova AA, Trepel JB, Panaretou B, Buchner J, Mayer MP, Prodromou C and Neckers L (2009) Hsp90 charged-linker truncation reverses the functional consequences of weakened hydrophobic contacts in the N domain. *Nat Struct Mol Biol* 16 (11):1141-7.

Ullers RS, Houben EN, Raine A, ten Hagen-Jongman CM, Ehrenberg M, Brunner J, Oudega B, Harms N and Luirink J (2003) Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. *J Cell Biol* 161 (4):679-84.

Vannini A & Cramer P (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 45 (4):439-46.

Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S and Yokoyama S (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* 417 (6890):712-9.

Vaughan CK, Gohlke U, Sobott F, Good VM, Ali MM, Prodromou C, Robinson CV, Saibil HR and Pearl LH (2006) Structure of an Hsp90-Cdc37-Cdk4 complex. *Mol Cell* 23 (5):697-707.

Vetter IR & Wittinghofer A (2001) The guanine nucleotide-binding switch in three dimensions. *Science* 294 (5545):1299-304.

Voigts-Hoffmann F, Schmitz N, Shen K, Shan SO, Ataide SF and Ban N (2013) The structural basis of FtsY recruitment and GTPase activation by SRP RNA. *Mol Cell* 52 (5):643-54.

Waterhouse AM, Procter JB, Martin DM, Clamp M and Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25 (9):1189-91.

Wegrzyn RD, Hofmann D, Merz F, Nikolay R, Rauch T, Graf C and Deuerling E (2006) A conserved motif is prerequisite for the interaction of NAC with ribosomal protein L23 and nascent chains. *J Biol Chem* 281 (5):2847-57.

Weinert T, Olieric V, Waltersperger S, Panepucci E, Chen L, Zhang H, Zhou D, Rose J, Ebihara A, Kuramitsu S, Li D, Howe N, Schnapp G, Pautsch A, Bargsten K, Prota AE, Surana P, Kottur J, Nair DT, Basilico F, Cecatiello V, Pasqualato S, Boland A, Weichenrieder O, Wang BC, Steinmetz MO, Caffrey M and Wang M (2015) Fast native-SAD phasing for routine macromolecular structure determination. *Nat Methods* 12 (2):131-3.

Werner F & Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9 (2):85-98.

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

Wittinghofer A (1997) Signaling mechanistics: aluminum fluoride for molecule of the year. *Curr Biol* 7 (11):R682-5.

Wittinghofer A & Vetter IR (2011) Structure-function relationships of the G domain, a canonical switch motif. *Annu Rev Biochem* 80 943-71.

Wu W, Nishikawa H, Hayami R, Sato K, Honda A, Aratani S, Nakajima T, Fukuda M and Ohta T (2007) BRCA1 ubiquitinates RPB8 in response to DNA damage. *Cancer Res* 67 (3):951-8.

Yam AY, Xia Y, Lin HT, Burlingame A, Gerstein M and Frydman J (2008) Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat Struct Mol Biol* 15 (12):1255-62.

Young JC, Hoogenraad NJ and Hartl FU (2003) Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. *Cell* 112 (1):41-50.

Young JC, Moarefi I and Hartl FU (2001) Hsp90: a specialized but essential protein-folding tool. *J Cell Biol* 154 (2):267-73.

Zhang DW, Rodriguez-Molina JB, Tietjen JR, Nemec CM and Ansari AZ (2012) Emerging Views on the CTD Code. *Genet Res Int* 2012 347214.

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K and Darst SA (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* 98 (6):811-24.

Zhang M, Wang XJ, Chen X, Bowman ME, Luo Y, Noel JP, Ellington AD, Etzkorn FA and Zhang Y (2012) Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS Chem Biol* 7 (8):1462-70.

Zhao R, Davey M, Hsu YC, Kaplanek P, Tong A, Parsons AB, Krogan N, Cagney G, Mai D, Greenblatt J, Boone C, Emili A and Houry WA (2005) Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* 120 (5):715-27.

Zhao R & Houry WA (2007) Molecular interaction network of the Hsp90 chaperone system. *Adv Exp Med Biol* 594 27-36.

Zhuravleva A & Gierasch LM (2011) Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones. *Proc Natl Acad Sci U S A* 108 (17):6987-92.

Zuiderweg ER, Bertelsen EB, Rousaki A, Mayer MP, Gestwicki JE and Ahmad A (2013) Allostery in the Hsp70 chaperone proteins. *Top Curr Chem* 328 99-153.

Abbreviations

Δ	deletion
°C	degree celsius
Å	Ångstrom
aa	amino acid(s)
AlF	aluminium fluoride
Amp	ampicillin
ATP	adenosine 5′-triphosphate
BICINE	2-(Bis(2-hydroxyethyl)amino)acetic acid
bp	base pair(s)
BSA	bovine serum albumin
CAM	chloramphenicol
CC	correlation coefficient
cDNA	complementary DNA
CPS	counts per second
CTD	Carboxy-terminal domain
CV	column volume
Da	dalton
DESY	Deutsche Elektronen Synchrotron
DMF	N,N-Dimehtlyformamide
DMSO	dimethylsulfoxide
DNA	deoxribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetraacetate
EMBL	European Molecular Biology Laboratories
et al.	and others [Latin: et alii]
EtOH	ethanol
Fig	figure
fl	full-length
g	gram or gravitational force
GDP	guanosine 5′-diphosphate
GTF	general transcription factor
GTP	guanosine 5′-triphosphate
GTPase	GTP hydrolase
GMPPCP	βγ-Methyleneguanosine 5′-triphosphate
GMPPNP	5'-Guanylyl imidodiphosphate
GPN	glycine-proline-asparagine
h	hour(s)
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
Hs	*Homo sapiens*
Hsp	heat-shock protein
IPTG	isopropyl β-D-1-thiogalactopyranoside
Iwr1	interacting with RNA polymerase II protein 1
k	kilo or thousand
l	liter
LB	lysogeny broth
M	molar (mol/liter)

mAu	milli absorbance unit
Min	minute(s)
MOPS	3-(*N*-morpholino)propanesulfonic acid
mRNA	messanger RNA
MW	molecular weight
NCBI	National Center for Biotechnology Information
Ni-NTA	nickel-nitrilotriacetic acid
Npa3	nucleolar preribosomal associated protein 3
nt	nucleotide(s)
OD_{600}	optical density at a wavelength of 600 nm
Pa	*Pyrococus abyssi*
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein data bank
PEG	poly(ethylene glycol)
PI	protease inhibitor
P_i	inorganic phosphate
PIC	pre-initiation complex
PMSF	phenylmethylsulfonylfluorid
Pol	DNA-dependant RNA polymerase
R-factor	residual factor
RNA	ribonucleic acid
RMSD	Root mean square deviation
RPAP	RNA polymerase II-associated proteins
Rpb	RNA polymerase B subunit
rpm	rounds per minute
rRNA	ribosomal RNA
RT	room temperature
s or sec	second(s)
S-SAD	sulfur-SAD
SAD	single-wavelength anomalous diffraction
Sc, S. cerevisiae	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulfate
SEC	size exclusion chormatography
SLS	Swiss Light Source
TAE	tris-acetate-EDTA
TAP	tandem-affinity purification
TBP	TATA-binding protein
TBS	tris buffered saline
TCA	trichloroacetic acid
TE	tris-EDTA
tet	tetracyclin
TF	transcription factor
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
TSS	transcription start site
U	unit(s)
UV	ultraviolet
V	volt
vs	versus
v/v	volume per volume
wt	wildtype
w/v	weight per volume
YPD	yeast extract peptone dextrose

List of figures

List of tables