Structure of GPN-loop GTPase Npa3 and implications for RNA polymerase II assembly

Jürgen Niesser\textsuperscript{a,b}, Felix R. Wagner\textsuperscript{b}, Dirk Kostrewa\textsuperscript{b}, Wolfgang Mühlbacher\textsuperscript{a} and Patrick Cramer\textsuperscript{a}

\textsuperscript{a}Max-Planck-Institute for Biophysical Chemistry, Department of Molecular Biology, Göttingen, Germany.
\textsuperscript{b}Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany.

Running Head: Structure of GPN-loop GTPase and Pol II assembly

Address correspondence to Patrick Cramer, pcramer@mpibpc.mpg.de
ABSTRACT

Biogenesis of the 12-subunit RNA polymerase II (Pol II) transcription complex requires so-called GPN-loop GTPases, but the function of these enzymes is unknown. Here we report the first crystal structure of a eukaryotic GPN-loop GTPase, the \textit{S. cerevisiae} enzyme Npa3 (a homolog of human GPN1, also called RPAP4, XAB1, MBDin), and analyze its catalytic mechanism. The enzyme was trapped in a GDP-bound, closed conformation, and in a novel GTP analogue-bound, open conformation displaying a conserved hydrophobic pocket distant from the active site. We show that Npa3 has chaperone activity and interacts with hydrophobic peptide regions of Pol II subunits that form interfaces in the assembled Pol II complex. Biochemical results are consistent with a model that the hydrophobic pocket binds peptides, and that this can allosterically stimulate GTPase activity and subsequent peptide release. These results suggest that GPN-loop GTPases are assembly chaperones for Pol II and other protein complexes.

INTRODUCTION

Protein folding was studied mainly for individual proteins (1), but most eukaryotic proteins form complexes, and their assembly is often poorly understood (2). 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 (3), but little is known about its biogenesis. Assembly of Pol II from its 12 subunits apparently occurs in the cytoplasm because depletion of any Pol II subunit leads to cytoplasmic accumulation of the largest
subunit Rpb1 (4). Assembly of the 10-subunit Pol II core likely involves three
subcomplexes, one containing Rpb1 and its associated subunits Rpb5, Rpb6, and
Rpb8, one containing the second largest subunit, Rpb2, and its associated subunit
Rpb9, and one containing Rpb3 and its associated subunits Rpb10, Rpb11, and Rpb12
(5). Assembled Pol II binds the import factor Iwr1 between Rpb1 and Rpb2, and is
imported to the nucleus (6).

Thus far, Pol II could not be reconstituted in vitro, indicating that its assembly in
vivo depends on the help of several factors. Indeed, recent studies identified putative
Pol II assembly factors, including the chaperones Hsp90, the RPAP3/R2TP/Prefoldin-
like complex, and a recently discovered subfamily of GTPases, the so-called GPN-loop
GTPases (4, 7, 8). Eukaryotic cells contain three paralogs of these essential GTPases.
In human, these are called GPN1 (also RPAP4, XAB1, or MBDin), GPN2, and GPN3
(also Parcs). Archaea contain a single GPN-loop GTPase, and prokaryotes lack
homologs. In the yeast S. cerevisiae, deletion of the GPN1 homolog Npa3 or its
paralogs GPN2 (YOR262W) and GPN3 (YLR243W) is lethal (9), indicating essential,
non-redundant functions of these enzymes. Homo- and heterodimerization of GPN1 and
its paralogs were reported (10-12). Npa3 contains a nuclear export sequence (NES,
residues 286-295)(13), consistent with the predominant cytoplasmic localization of Npa3
in yeast (14, 15) and GPN1 in human cells (7, 13, 16, 17).

The family of GPN-loop GTPases is characterized by a highly conserved motif
consisting of the amino acids Gly-Pro-Asn (single letter code: GPN). Structural studies
of the archaeal GPN-loop GTPase Pab0955 suggested that the GPN motif functions in
GTP hydrolysis (18). Nucleotide binding and hydrolysis of all GTPases involves five so-
called G motifs (19) (G1-G5). Mutation in the nucleotide-binding site or GPN motif in human GPN1 leads to cytoplasmic accumulation or decreased nuclear levels of Rpb1 (7, 11). Depletion of human GPN3 (20) or mutation of yeast GPN2 or GPN3 (12) 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 the yeast Saccharomyces cerevisiae leads to cytoplasmic accumulation of Rpb1 and Rpb3 (21). Rpb1 accumulation is also observed when Npa3 is mutated in its nucleotide-binding site or GPN motif (7, 21). Association of yeast Npa3 with Rpb1 is regulated by GTP binding in whole-cell extracts (21) 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 (11).

To gain insights into the structure and function of GPN-loop GTPases and their role in Pol II biogenesis we studied the yeast GPN1 enzyme Npa3 with a combination of X-ray crystallography, site-directed mutagenesis, enzymatic activity assays, chaperone assays, and a systemic peptide interaction screen. Our results indicate that Npa3 functions as an assembly chaperone during Pol II biogenesis and binds hydrophobic regions in Pol II subunits that are released upon GTP hydrolysis to form interfaces in the mature polymerase complex.

RESULTS

Domain organization and structure determination
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) (Fig 1A). The GTPase core harbors the motifs G1-G5 that are required for GTP binding and hydrolysis (19). The C-terminal tail is poorly conserved among eukaryotes and is absent in archaea. Because efforts to crystallize full-length Npa3 failed, we removed a non-conserved 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 Npa3∆C∆Loop comprising residues 1-202 and 212-264). We co-crystallized Npa3∆C∆Loop (Figure 1A) with the non-hydrolyzable GTP analogue GMPPCP, but also with GDP, and with GDP and AlF_3 (‘GDP·AlF_3’).

The structure with GDP·AlF_3 was solved using single anomalous diffraction from intrinsic sulphur atoms and a protein model was built and refined to a free R-factor of 23.9% at 1.85 Å resolution (Table 1). 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 1). The electron density of the GDP·AlF_3 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 also showed great stereochemistry, and were used for further analysis, because they showed very well-defined densities for bound nucleotides, indicative of defined enzyme states. 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 1B and Supplemental Figure S1).

Npa3·GDP structure shows eukaryote-specific features

The Npa3 core consists of a central, six-stranded parallel β-sheet surrounded by helices (Figure 1C). 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 (18) (PDB-code 1YRB), but also reveals major differences in insertions 1 and 2 (Supplemental Figure S2) 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°. 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. Helix α3 in insertion 1 of Npa3·GDP is apparently rotated by ~40° towards the core, and helix α4 and the GPN-loop are shifted by ~3 Å towards the G3 motif. Taken together, the core of the eukaryotic enzyme generally resembles its archaeal counterpart, whereas the two insertions differ significantly.

Npa3·GMPPCP structure reveals a novel open conformation
Our Npa3·GMPPCP structure reveals a novel open conformation that differs significantly from the Npa3·GDP structure (Figures 1C and 2). 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; (22)).

In the GMPPCP-bound state, residues D189 of insertion 2 and C179 show peptide bond flips that enable a straight conformation of helices α8 and α9 and the formation of a single helix (Figures 2A and D). Helix α10, which is bent in the Npa3·GDP structure, adopts a straight conformation, and helix α7 is rotated by ~15°. This disrupts a methionine-rich hydrophobic core formed between helices α7, α8, and α9 that kept the pocket closed in the Npa3·GDP structure (Figures 2B and E). 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 2E). Mutation of any one of these residues (E112, Q110, N72) is lethal in yeast (7, 10, 21).

In addition, insertion 1 is changed in the Npa3·GMPPCP structure, including motif G3, which binds the GPN-loop (Figure 2C). 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 2C). 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 (Supplemental Figure S3). 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. 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 analogon GMPPCP is bound.

Nucleotide binding and conformational states

The two structures reveal atomic details of how Npa3 binds GMPPCP or GDP using its motifs G1-G5 (Figure 3). In the Npa3·GMPPCP structure, strong electron density corresponding to the γ-phosphate of the nucleotide is observed (Figure 3A). 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) (Fig 3B). After GTP hydrolysis an additional water molecule apparently replaces the coordination sphere previously occupied by the γ-phosphate oxygen of GMPPCP (Figure 3C). 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 analogue.
Catalytic mechanism

The catalytic mechanism for hydrolysis of GTP to GDP and inorganic phosphate by GPN-loop GTPases was suggested for the archaeal enzyme (18). To probe the catalytic mechanism in the eukaryotic system we prepared Npa3 variants with point mutations and tested their ability to hydrolyze GTP (Figure 4). Whereas residue D106 in motif G3 stabilizes the Mg$^{2+}$ ion, residue D40 in motif G2 positions the nucleophilic water molecule (N), as observed for the bacterial GTPase FtsY (23). Consistent with this, the D106A and D40A variants of Npa3 were both inactive (Figure 4A).

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 (18). 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 (Figures 4A and B). 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 (7, 21). 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 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.

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 5A). 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 (Figure 5A). 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 5B).

GPN1 is known to interact with assembling Pol II subunits (4), which expose hydrophobic peptide regions buried in subunit interfaces of the Pol II assembly (24). Therefore, it is likely that the newly observed hydrophobic pocket of Npa3 binds hydrophobic peptides in the cellular context. 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 (1, 2). Indeed, modeling showed that hydrophobic peptides may be accommodated in the pocket in an extended conformation.

**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. In this assay, the temperature-induced aggregation of CS is suppressed when a chaperone is added (25). Indeed, Npa3 was able to suppress temperature-induced aggregation of CS, revealing that this GPN-loop GTPase has chaperone activity (Figure 5C). We next investigated whether the chaperone activity was nucleotide-dependent. Because no GTPase activity was observed in the standard chaperone assay buffer, we changed to buffer C containing 100 mM NaCl (Methods) 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 (Methods). 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 5D). These results indicate 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 (Figures 2C and E) and G4 (Figure 2D) that widens the nucleotide-binding pocket. Indeed, a similar conformational change in Ffh from *T. aquaticus* was proposed to allosterically...
alter the active site via the G4 motif (26). Thus we speculated that peptide binding would allosterically widen the active center, 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 (Figure 5E), providing a link between chaperone and GTPase functions and indicating that the structural changes inferred by crystallography occur in solution. GTP hydrolysis-driven release of CS and subsequent rebinding and pocket opening could thus explain the similar chaperone activity observed in our assay upon GTP addition. 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.

Npa3 binds peptides derived from Pol II subunit interfaces

Because Npa3 was implicated in Pol II biogenesis, and exhibits general chaperone activity, as observed for other assembly chaperones (27, 28), we investigated whether Npa3 may function in the assembly of Pol II. Unfortunately, this could not be tested directly, because Pol II is not available in recombinant form, has never been assembled from its subunits in vitro, 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) covering all regions of all 12 Pol II subunits, for binding Npa3 in the presence of GMPPCP, GTP, or GDP (Figure 6A, Supplemental Table S1, Supplemental Figure S4). In this assay 55 peptides bound Npa3 significantly (signal intensity >3.5) (Supplemental Table S2). The binding efficiency generally did not depend on nucleotides, consistent with the model that opening of the hydrophobic pocket is triggered by peptide binding. Npa3 did not interact with assembled, mature Pol II (not shown), consistent with the model that Npa3 binds Pol II assembly intermediates.

The obtained 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 6B), 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 (Figures 6C and D; Supplemental Table S2, Supplemental Figure S4). 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 were bound by Npa3 (Figures 6C and D). We identified numerous Npa3-binding peptides within the interface between the two largest Pol II subunits, Rpb1 and Rpb2. 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) (Figure 6C). We also detected three Npa3-binding peptides (193, 194, 201) in the funnel domain of Rpb1.
Npa3 also interacted with peptides from subunits Rpb8, Rpb9 and Rpb11 that are located in critical interfaces with Rpb1 (Figure 6D). Three overlapping Npa3-binding peptides (peptides 1,037-1,039) contained a hydrophobic region in Rpb8 (117-SFGGLLMR-124) that forms an interface with 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 and 1,108-1,110). Peptide 1,109 covers most interface residues and shows the strongest Npa3 binding among Rpb11-derived peptides. 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 6D), consistent with a reported interaction of human GPN1 with Rpb4 and Rpb7 (11). 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. Unfortunately, attempts to use identified peptides for interaction analysis in solution and for co-crystallization trials with Npa3 failed due to insolubility of such synthesized peptides. Nevertheless, our data show that Npa3 binds numerous Pol II-derived hydrophobic peptides that are located at subunit interfaces, consistent with a function of Npa3 in Pol II assembly.

DISCUSSION

Many macromolecular complexes were shown to require assembly chaperones for their biogenesis (2), including the nucleosome (29, 30), Rubisco (31), the proteasome (32), spliceosomal snRNPs (33), and the ribosome (34). 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 (4), a chaperone that is involved in the assembly of several protein complexes (35), but additional factors are likely involved in Pol II assembly.

Here we study the structure and function of the new family of GPN-loop GTPases and provide evidence that these enzymes assist in Pol II assembly. 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 general chaperone activity, and that a chaperone substrate protein can stimulate its GTPase activity. The latter observation is reminiscent to the reported stimulation of ATPase activity of chaperones Hsp70 (36-38) and Hsp90 (39) by substrate binding. GTPases were also shown to play a key role during ribosome assembly in bacteria (40) and in eukaryotes (41).

Together with published data, our results suggest a model for Pol II biogenesis (Figure 7). In this model, GPN-loop GTPases are involved in the correct assembly of Pol II in the cytoplasm. Assembled Pol II would then be recognized by Iwr1 and imported into the nucleus (6). Several lines of evidence argue for a role of GPN-loop GTPases in Pol II assembly. First, GPN1, GPN2 and GPN3 interact with Pol II assembly intermediates (4). Second, GPN1 interacts with the CCT complex (7), a chaperone complex with various functions (42) that interacts with Pol II subunits (43). Third, 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 (44), consistent with a chaperone function of GPN1.
In our model, an exposed hydrophobic peptide region in a newly synthesized Pol II subunit would trigger opening of the Npa3 pocket in its GDP-bound state. Npa3 would then trap exposed hydrophobic regions of Pol II subunits, preventing their misassembly, and opening a time window for association with the cognate Pol II subunit (Figure 7). Peptide binding would then alter the active site allosterically, decreasing its affinity for GDP, provoking GDP displacement, and increasing its affinity for GTP, to facilitate GTP rebinding. Subsequent GTP hydrolysis would then lead to release of the bound Pol II subunit, to enable association with cognate subunits and Pol II assembly.

Assembled Pol II would then be recognized by Iwr1, which binds between the two largest polymerase subunits and serves as an adaptor for import of Pol II into the nucleus (6).

Our model does not include any nuclear roles of Npa3, although they may exist, because nucleocytoplasmic shuttling of GPN1/Npa3 has been reported (7, 13, 21, 45). However, 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 (12), consistent with the main function of these enzymes being cytosolic. In the future, a system to co-express all 12 Pol II subunits may be derived that would allow one to test whether in the presence of GPN-loop GTPases Pol II assembly is facilitated. Future research on the GPN-loop family of essential GTPases is guided by our structure-function analysis of Npa3 and its apparent function as a chaperone in assembly of protein complexes such as Pol II.

MATERIALS AND METHODS
**Npa3 expression and purification.** Wild-type Npa3 from *Saccharomyces cerevisiae* was amplified from genomic DNA and subcloned into pOPINI vector (46) (provided by OPPF-UK) containing an N-terminal hexahistidine tag. Mutations were introduced by overlap extension PCR from wild-type Npa3 plasmid DNA and mutants were subcloned into pOPINE vector (46) (provided by OPPF-UK) containing a C-terminal hexahistidine tag. Variants of Npa3 were expressed in *E. coli* Rosetta (DE3) (Novagen). The culture was grown in LB medium at 37°C until absorbance at 600 nm of 0.6, 0.5 mM IPTG were added and the culture was grown for a further 20 h at 20°C. Cells were harvested by centrifugation and frozen at -20°C. Cells were lysed by sonication in buffer A (50 mM Tris pH7.5, 300 mM NaCl, 5 mM MgCl₂, 2 mM DTT supplemented with 5 mM imidazole, 0.2% (v/v) Tween20 and 1x protease inhibitors (100x stock: 1.42 mg leupeptin, 6.85 mg pepstatin A, 850 mg PMSF, 1.685 mg benzamidine in 50 ml ethanol). After centrifugation at 24,000g for 30 min, the cleared lysate was loaded onto a 2 ml Ni-NTA column (Qiagen), pre-equilibrated with buffer A containing 10 mM imidazole. The column was washed with 10 column volumes buffer A containing 10 mM imidazole before elution of the bound protein with buffer A containing 200 mM imidazole. The conductivity of the eluate was adjusted to match that of buffer B (50 mM Tris pH7.5, 100 mM NaCl, 5 mM MgCl₂, 2 mM DTT) and applied to a MonoQ 10/100 GL column (Amersham) equilibrated in buffer B. The protein was eluted with a linear gradient from 100 mM to 1 M NaCl. After concentration, the sample was applied to a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with buffer C (10 mM Hepes pH7.5, 5 mM MgCl₂, 10 mM DTT) either containing 100 mM NaCl for wild-type and full-
length Npa3 mutants or 200 mM NaCl for Npa3ΔCΔLoop. Peak fractions were pooled and concentrated as desired.

**Crystallization and X-ray structure analysis of Npa3ΔCΔLoop.** The protein was concentrated to approximately 3.7 mg/ml and incubated at 8°C overnight with either 10 mM GDP (Sigma-Aldrich), 5 mM GDP + 100 mM NaF/10 mM AlCl$_3$ or 10 mM GMPPCP (Jena Bioscience). Npa3ΔCΔLoop·GDP crystals were grown at 20°C by sitting drop vapour diffusion after 6-15 days using 9 mM HEPES pH7, 45 mM NaCl, 4.5 mM MgCl$_2$ and 5% (v/v) Jeffamine M-600 as reservoir solution. Npa3ΔCΔLoop·GDP·AlFx was crystallized at 8°C by hanging-drop vapour diffusion with buffer C containing 200 mM NaCl, 100 mM NaF and 10 mM AlCl$_3$ as reservoir solution. Npa3ΔCΔLoop·GMPPCP crystals were grown over night at 8°C in a 1.5 ml Eppendorf tube in buffer C containing 200 mM NaCl. Crystals grew to a maximum size of approximately 0.25x0.25x0.1 mm under all conditions. Cryo-protection was carried out by step-wise transferring to mother solution containing 35% (v/v) glycerol before flash-cooling in liquid nitrogen. A single anomalous diffraction experiment from intrinsic sulphur atoms (S-SAD) was performed on crystals Npa3ΔCΔLoop·GDP·AlFx at beamline X06DA of the Swiss Light Source in Villigen, Switzerland (Table 1). 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 (47). A high resolution native data set of Npa3ΔCΔLoop·GDP·AlFx was collected from the same crystal at the same beamline. Diffraction data for GDP- and GMPPCP-bound Npa3ΔCΔLoop were collected at 100 K at beamline MX1 at EMBL/DESY in Hamburg, Germany (Table 1). Raw data were
processed with XDS (48). The programs SHELXC/D/E from the SHELX suite (49) were used for detection of 17 sulphur atoms and for SAD phasing. The GDP-bound structure was solved by molecular replacement using PHASER (50) with the Npa3ΔCΔLoop·GDP·AlF₆ core structure as search model. The GMPPCP-bound structure was refined using the Npa3ΔCΔLoop·GDP·AlF₆ structure with phenix.refine (51). Models were iteratively built with COOT (52) and refined with phenix.refine. Figures were prepared with PyMOL (deLano Scientific).

**GTPase activity assay.** GTPase activity was measured in 96 well plates using the Malachite Green Phosphate Assay Kit (BioAssay Systems, Hayward, CA) as recommended by the manufacturer. 0.15 µM of Npa3 variants were incubated with 100 µM GTP (Fermentas, St. Leonrot, Germany) at 37°C in buffer C containing 200 mM NaCl. GTPase activity in the presence of partially unfolded citrate synthase (CS) was measured at 43°C in buffer C containing 100 mM NaCl, pH 7.5 (43°C). Orthophosphate concentrations from three independent experiments were determined at the indicated time points by measuring the absorption at 620 nm using an Infinite M1000 plate reader (Tecan, Männedorf, Switzerland). Control experiments were performed to determine orthophosphate contaminations in the individual solutions and data was corrected for this values.

**Analysis of chaperone activity.** Thermal aggregation of citrate synthase (CS) was essentially carried out as described (25) with modifications. 0.15 µM CS (monomer) were either incubated in pre-heated (43°C) incubation buffer IB (40 mM HEPES-KOH
pH 7.5 (43°C), 5 mM MgCl₂) in the absence of nucleotides or in pre-heated (43°C) buffer C containing 100 mM NaCl, pH 7.5 (43°C) in the presence of 1 mM GMPPCP, GTP or GDP, respectively and varying amounts of Npa3 were added. For experiments where GMPPCP or GDP was added, Npa3 samples were incubated with the respective nucleotide overnight at 4°C for nucleotide exchange. CS aggregation was measured by light scattering at 360 nm using a Fluoromax 3 fluorometer (Horiba, Unterhaching, Germany). Data represents the average of three independent experiments. Control experiments were performed that showed that Npa3 does not aggregate under the assay conditions (not shown).

**Pol II peptide-binding assay.** 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 buffer C containing 100 mM NaCl. 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 TBS T20) 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) was 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.

Accession code. Coordinates of the Npa3·GDP and Npa3·GMPPCP complex structure have been deposited with the protein data bank under accession codes XXXX and YYYY, respectively.

ACKNOWLEDGEMENTS

We thank the crystallization facility at the Max Planck Institute of Biochemistry, Martinsried, Germany. Part of this work was performed at the Swiss Light Source at the Paul Scherrer Institut, Villigen, Switzerland and the Deutsches Elektronen Synchrotron (DESY) and the European Molecular Biology Laboratory (EMBL), Hamburg, Germany. We thank Aymelt Itzen, Andreas Bracher, Alexander Bepperling and Andre Heuer for advice and help.

FUNDING INFORMATION

J.N. was supported by the Deutsche Forschungsgemeinschaft (Graduiertenkolleg 1721). P.C. was supported by the Deutsche Forschungsgemeinschaft, (SFB860, GRK1721), the Advanced Grant ‘TRANSIT’ of the European Research Council, and the Volkswagen Foundation. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.
AUTHOR CONTRIBUTIONS

J.N. carried out experiments, structure determination and modelling. F.R.W. supported Pol II-Npa3 interaction studies (data not shown). D.K. and W.M. advised on X-ray data analysis. P.C. initiated and supervised research. J.N. and P.C. prepared the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES


**Figure 1** - Crystal structures of Npa3 in GDP- (closed) and GMPPCP-bound (open) forms.

A  Schematic representation of yeast (S.c.) Npa3 domain organization. The color code is used throughout the figures.

B  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, 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.

C  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.

**Figure 2** - 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.
Peptide flip of D189 enables pocket opening and the formation of a single helix from helices α8 and α9.

Conformational changes in insertion 2 and helix α7 facilitate the opening of an extended, hydrophobic pocket.

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.

Pocket opening allosterically alters the active site via the G4 motif.

Conformational changes in helix α7 are linked to the G3 motif and the GPN-loop.

**Figure 3** - Nucleotide-binding pocket and active site.

Initial unbiased Fo-Fc 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 1a. Water molecules are shown as small red spheres, magnesium ions as pink spheres and hydrogen bonds as dashed lines.

Nucleotide interaction network of Npa3 with GMPPCP (B) and GDP (C). Metal-ion-ligand interactions are shown as solid black lines.

**Figure 4** - Catalytic mechanism.

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(18)). 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.

Figure 5 - Putative peptide-binding pocket and GTPase-stimulating chaperone activity.

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.

C 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.

D 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.

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).

Figure 6 - Npa3 binds Pol II-derived peptides located at subunit interfaces.

A Boxplot representation of a representative portion of the heatmap describing Npa3 peptide-binding landscape. Tested were a total of 1,139 15-meric Pol II-derived peptides, covering the complete sequence of the 12-subunit Pol II in the presence of GMPPCP, GTP and GDP, 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).

B Location of Npa3-binding peptides in the assembled Pol II complex (pdb-code 1WCM(24)). 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). Schematic representation of the 12 Pol II subunits Rpb1-Rpb12 in the folded Pol II complex is shown on the right.

C-D Surface representations of Pol II subunits show that Npa3 interacts with hydrophobic Pol II-derived peptides located at subunit interfaces (left panels). Peptides in Pol II subunit surfaces involved in interaction with other subunits are colored as in (b). Numbers correspond to the peptide numbers from the array.
(see Supplementary Table S1 for list of all peptides, Supplementary Table S2 for list of all peptides that interact with Npa3 and Supplementary Fig S 4 for original peptide interaction analysis) and are color-coded according to the subunit they are derived from as shown on the left and the schematic Pol II subunit organization from (B). Npa3-bound peptides from other subunits that interact with Rpb1 are shown as sticks. Residues involved in subunit interfaces are shown in black (middle panels). White solid lines highlight hydrophobic interface regions, bound by Npa3 (right panels). Dashed lines indicate where other Pol II subunits are positioned in the assembled complex.

**Figure 7 - Model for Pol 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 (step 3). 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).
Table 1 X-ray diffraction data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Native</th>
<th>S-SAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C222₁</td>
<td>P4₁2₁₂</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a, b, c (Å)</td>
<td>108.0, 119.2, 347.5</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.99888</td>
<td>0.99888</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-2.3 (2.36-2.3)³</td>
<td>50-2.2 (2.26-2.2)³</td>
</tr>
<tr>
<td>R₁/</td>
<td></td>
<td>of (%)</td>
</tr>
<tr>
<td>Redundancy (%)</td>
<td>97.1 (99.8)</td>
<td>99.5 (98.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7</td>
<td>99.5</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7</td>
<td>99.5</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;</td>
<td>23.9/27.6</td>
<td>21.4/22.7</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>23.9/27.6</td>
<td>21.4/22.7</td>
</tr>
<tr>
<td>No. atoms</td>
<td>Protein</td>
<td>12,065</td>
</tr>
<tr>
<td></td>
<td>Nucleotide</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>AlF₄&lt;sup&gt;⁻&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Mg²⁺</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Lauric acid</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Glycerol</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>121</td>
</tr>
<tr>
<td>B-factors (Å²)</td>
<td>Protein</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td>Nucleotide</td>
<td>69.6</td>
</tr>
<tr>
<td></td>
<td>AlF₄&lt;sup&gt;⁻&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Mg²⁺</td>
<td>64.2</td>
</tr>
<tr>
<td></td>
<td>Lauric acid</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Glycerol</td>
<td>107.7</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>60.1</td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td>Bond lengths (Å)</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>Bond angles (°)</td>
<td>1.29</td>
</tr>
</tbody>
</table>

¹Diffraction data were collected at beamline MX1 at EMBL/DESY, Hamburg, Germany and processed with XDS (48).

²Diffraction data were collected at beamline X06DA, Swiss Light Source, Switzerland and processed with XDS (48).

³Data are merged form three data sets, measured at different χ angles (0°, 10°, 20°) from a single crystal.

⁴Numbers in parenthesis refer to the highest resolution shell.

⁵³CC<sub>1/2</sub> = percentage of correlation between intensities from random half-datasets (53).