Structural insights into the nucleic acid remodeling mechanisms of the yeast THO-Sub2 complex

Sandra K Schuller[†‡], Jan M Schuller[†§], J Rajan Prabu, Marc Baumgärtner, Fabien Bonneau, Jérôme Basquin, Elena Conti*

Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Munich, Germany

Abstract The yeast THO complex is recruited to active genes and interacts with the RNA-dependent ATPase Sub2 to facilitate the formation of mature export-competent messenger ribonucleoprotein particles and to prevent the co-transcriptional formation of RNA:DNA-hybrid-containing structures. How THO-containing complexes function at the mechanistic level is unclear. Here, we elucidated a 3.4 Å resolution structure of *Saccharomyces cerevisiae* THO-Sub2 by cryo-electron microscopy. THO subunits Tho2 and Hpr1 intertwine to form a platform that is bound by Mft1, Thp2, and Tex1. The resulting complex homodimerizes in an asymmetric fashion, with a Sub2 molecule attached to each protomer. The homodimerization interfaces serve as a fulcrum for a seesaw-like movement concomitant with conformational changes of the Sub2 ATPase. The overall structural architecture and topology suggest the molecular mechanisms of nucleic acid remodeling during mRNA biogenesis.

*For correspondence:
conti@biochem.mpg.de

[†]These authors contributed equally to this work

Present address: [‡]Center for Synthetic Microbiology (SYNMIKRO) and Max Planck Institute for Terrestrial Microbiology, Marburg, Germany; [§]Center for Synthetic Microbiology (SYNMIKRO) and Department of Chemistry, Philipps University Marburg, Marburg, Germany

Introduction

The biogenesis of eukaryotic mRNAs in the cell nucleus is an elaborate, multi-step process. As the nascent transcript is synthesized by RNA polymerase II from the corresponding DNA template, it undergoes several chemical modifications, including 5' capping, splicing and 3'-end polyadenylation. Simultaneously, the nascent transcript associates with a cohort of complementary proteins to form a mature messenger ribonucleoprotein particle (mRNP) (*Wende et al., 2019*; *Xie and Ren, 2019*; *Müller-McNicoll and Neugebauer, 2013*; *Meinel and Sträßer, 2015*; *Singh et al., 2015*; *Wegener and Müller-McNicoll, 2018*). The mature mRNP is then shuttled through the nuclear pore to the cytoplasm via specific export factors (*Moore and Proudfoot, 2009*; *Müller-McNicoll and Neugebauer, 2013*; *Rondón et al., 2010*; *Wende et al., 2019*; *Xie and Ren, 2019*). The individual steps in this process are orchestrated by multi-protein complexes. Among them is the transcription-export (TREX) complex, an evolutionary conserved complex containing a multi-protein assembly (transcription-dependent hyperrecombination complex THO), an RNA helicase (Sub2 in yeast/UAP56 in human), and an hnRNP-like protein (Yra1 in yeast/Ref/Aly in human) (*Heath et al., 2016*; *Meinel and Sträßer, 2015*; *Strässer and Hurt, 2001*; *Strässer et al., 2002*; *Xie and Ren, 2019*). Sub2 and Yra1 interplay with the mRNA-export factor Mex67-Mtr2 (TAP-p15 in human) (*Heath et al., 2016*; *Strässer and Hurt, 2001*; *Xie and Ren, 2019*). As the name implies, the TREX complex is believed to play pivotal roles in linking the transcription and nuclear export steps in the mRNP biogenesis pathway, but the molecular mechanisms remain unclear.

The THO complex was first identified in *Saccharomyces cerevisiae* and named after the pheno-types observed upon mutations of its constituent subunits, most notably an increase in recombination events connected to the accumulation of R loops during transcription (*Aguilera and Klein, 1988*; *Chávez et al., 2000*; *Chávez et al., 2001*; *Heath et al., 2016*; *Huertas and Aguilera, 2003*;

Piruat and Aguilera, 1998). R-loops can form behind the elongating RNA polymerase II when the nascent transcript erroneously reanneals with the DNA template to create a three-stranded nucleic acid structure consisting of an RNA:DNA hybrid and an unpaired single DNA strand. R loops are particularly prominent in discrete areas of the genome and can lead to recombination events and, eventually, genomic instability (*Santos-Pereira and Aguilera, 2015*; *Skourti-Stathaki et al., 2014*). Among the subunits of the yeast THO complex, the evolutionary conserved Tho2 and Hpr1 have the most drastic impact on mutation-induced phenotypes epitomized by a DNA recombination rate of about 3000 times that found in wild-type cells (*García-Rubio et al., 2008*; *Chávez et al., 2000*). The mechanisms with which Tho2, Hpr1, and the other yeast THO complex core subunits (Mft1, Thp1, and Tex1) protect against R loop formation are thought to involve the unwinding activity of the Sub2 helicase (*Luna et al., 2019*). Indeed, THO enhances the ATPase properties of Sub2 (*Ren et al., 2017*), but there is currently no high-resolution structure of this complex to allow a mechanistic understanding (*Ren et al., 2017*).

Yeast Sub2 is a conserved RNA-dependent ATPase of the DEAD-box family, a group of enzymes known to destabilize and unwind RNA duplexes (*Linder and Jankowsky, 2011*; *Xie and Ren, 2019*). Biochemical studies have shown that Sub2 interacts tightly with THO both in vivo and in vitro (*Jimeno et al., 2002*; *Luna et al., 2019*; *Peña et al., 2012*; *Ren et al., 2017*; *Strässer et al., 2002*). Furthermore, THO mutant alleles are synthetic lethal with a temperature-sensitive mutant of *sub2* causing a fast onset of nuclear poly(A) RNA accumulation, a hallmark of defective mRNA export (*Jimeno et al., 2002*; *Strässer et al., 2002*; *Wegener and Müller-McNicoll, 2018*). The mRNA export functions of Sub2 are mediated by its interacting partner Yra1, a non-shuttling hnRNP-like protein (*Heath et al., 2016*; *Strässer and Hurt, 2001*; *Stutz et al., 2000*; *Xie and Ren, 2019*). The propensity of Yra1 to bind genomic areas that are prone to R-loop formation (*García-Rubio et al., 2008*) is underscored by its ability to bind both RNA and DNA (*Abruzzi et al., 2004*). Additionally, Yra1 harbors potent RNA:RNA annealing activity (*Portman et al., 1997*). Data from both yeast and human studies have converged on the notion that THO binds Sub2, which in turn recruits Yra1 to form the TREX complex (*Heath et al., 2016*; *Strässer et al., 2002*; *Wende et al., 2019*; *Zenklusen et al., 2002*). This series of events is remarkably conserved in the human orthologues, the DNA–RNA helicase UAP56 (*DDX39B*) and the RNA-binding factor Aly/Ref (*ALYREF*) (*Heath et al., 2016*; *Luo et al., 2001*; *Strässer et al., 2002*; *Wende et al., 2019*). In this study, we identify the molecular mechanisms utilized by yeast THO to activate the unwinding properties of the Sub2 ATPase, thereby intimating how THO-containing complexes can fulfil their functions in R loop prevention and mRNP formation.

Results and discussion

The *S. cerevisiae* THO complex is a dimer upon in vitro reconstitution and upon endogenous purification

S. cerevisiae Tho2 (1597 aa, 184 kDa), Hpr1 (752 aa, 88 kDa), Mft1 (392 aa, 45 kDa), and Thp2 (261 aa, 30 kDa) are all expected to contain folded N-terminal regions that encompass more than two-thirds of their polypeptide chain and to be primarily α-helical in secondary structure as predicted by the program Phyre2 (*Kelley et al., 2015*; *Figure 1A*). Tex1 (422 aa, 47 kDa) is the only non-helical subunit of the THO complex and contains a β-propeller domain (*Kelley et al., 2015*; *Figure 1A*). Since, with the exception of Tex1, it was not possible to accurately predict the domain boundaries between the structured and unstructured regions of THO complex subunits, we co-expressed them as full-length proteins in insect cells. The recombinant THO complex was purified to homogeneity using affinity and size-exclusion chromatography (*Figure 1B*). Full-length Sub2 (446 aa, 50 kDa) was individually expressed, purified, and added to the THO complex. The THO-Sub2 assembly eluted as a single peak by size-exclusion chromatography, albeit at an earlier elution volume than expected for a globular mass of ~450 kDa and rather consistent with the presence of a dimer (*Figure 1B*).

To assess the presence of THO dimerization in a cellular environment, we integrated a C-terminal tandem-affinity purification tag (Twin-strep-3C-Protein-A) at the endogenous HPR1 locus in BY4741 (haploid) and BY4743 (diploid) *S. cerevisiae* strains. The haploid strain produces only a tagged version of Hpr1, whereas the diploid strain retains both tagged and untagged versions. The native complex was isolated from both yeast strain extracts by tandem affinity purification (*Figure 1C*). An
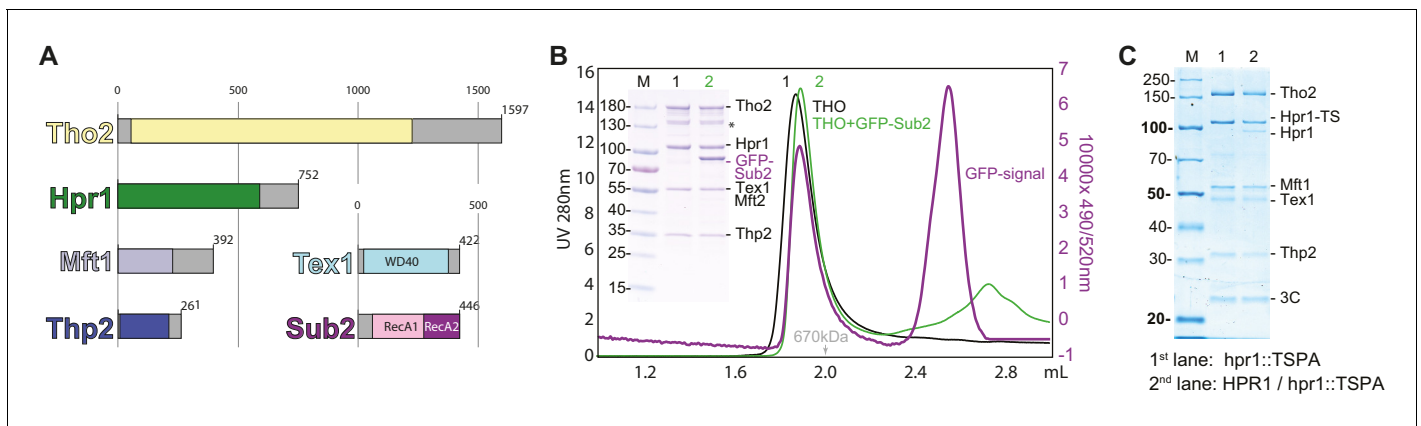
Figure 1. Biochemical reconstitution and native isolation of *Saccharomyces cerevisiae* THO-Sub2. (A) Domain organization of Tho2 (yellow), Hpr1 (green), Mft1 (light blue), Thp2 (dark blue), Tex1 (cyan), and Sub2 (with the RecA1 domain in pink and the RecA2 domain in purple). Gray parts are not resolved in the structural analysis described in the paper and correspond to regions predicted to be mainly unstructured. (B) Analytical size-exclusion chromatography of THO with/without GFP-Sub2. The co-elution of GFP-Sub2 was monitored by fluorescence at excitation 490 and emission 520 nm. The green line is the 280 nm absorbance signal of THO with GFP-Sub2; the purple line is the fluorescence signal of THO with GFP-Sub2. The asterisk indicates a Tho2 degradation product. (C) Tandem affinity isolation of native THO-containing complexes from haploid and diploid yeast. A single allele of HPR1 was tagged C-terminally with a Twin-Strep-3C-Protein-A tag (TSPA) in BY4741 (haploid) or BY4743 (diploid). Eluates resulting from IgG-affinity followed by Strep-tactin (IBA) affinity purification were analyzed on 12% SDS-PAGE stained with Instant Blue (Expedeon). M, molecular weight marker; Hpr1-TS, Hpr1 twin strep.

additional band corresponding to untagged Hpr1 was also observed in the complex originating from the diploid strain, consistent with the presence of a dimer in a physiological context.

Cryo-electron microscopy (cryo-EM) analysis of a yeast THO complex bound to Sub2

For cryo-EM single-particle analysis, we stabilized the THO-Sub2 assembly with the mild crosslinker BS3 to obtain a homogeneous particle distribution suitable for high-resolution reconstructions. After iterative-rounds of particle sorting and refinements, the reconstruction revealed the presence of an asymmetric dimer, with a well-ordered protomer resolved to a resolution of 3.7 Å and a more flexible protomer with a resolution spread of 3.8–7 Å, referred to as rigid and flexible protomers, respectively (*Figure 2A* and *Figure 2—figure supplement 1*; *Supplementary file 1*). Focusing on the classification of the entire dataset on the rigid protomer, we improved the density map to 3.4 Å resolution (*Figure 2—figure supplement 2*), allowing us to build the atomic model of the pentameric THO protomer de novo (*Figure 2B* and *Figure 2—figure supplement 3*; *Supplementary file 1*). The final model includes the large N-terminal structured regions of the five THO-complex subunits as well as a prominent low-complexity sequence of Hpr1 (residue 501–675) (*Figure 1A*). The densities for Sub2 were modeled using the RecA domains from the reported high-resolution crystal structure (*Ren et al., 2017*). The density of the flexible protomer was interpreted using the structure of the rigid protomer (*Figure 2B*). The THO-Sub2 homodimer has a complex intertwined architecture (*Figure 2B and C*). Maximum-likelihood variance analysis of the cryo-EM data revealed that the significant nonuniform distribution of resolution resulted from the remarkable dynamic character of this assembly: the two protomers swivel with respect to each other, alternating the opening and closing of each side of the homodimer (referred to as proximal and distal sides; *Figure 2D*; *Video 1*). Since the sample we characterized by cryo-EM did not contain an energy source, these fluctuations likely reflect thermal motions. In support of our cryo-EM reconstruction (*Figure 2*), our findings are consistent with a previous negative-stain analysis of a native THO complex purified from yeast (*Peña et al., 2012*; *Figure 2—figure supplement 4*). Although a previous THO model based on 6 Å resolution crystallography data (*Ren et al., 2017*) appears different from our structure at first glance, the differences can be reconciled by reinterpreting the crystal lattice (*Figure 2—figure supplement 4*).
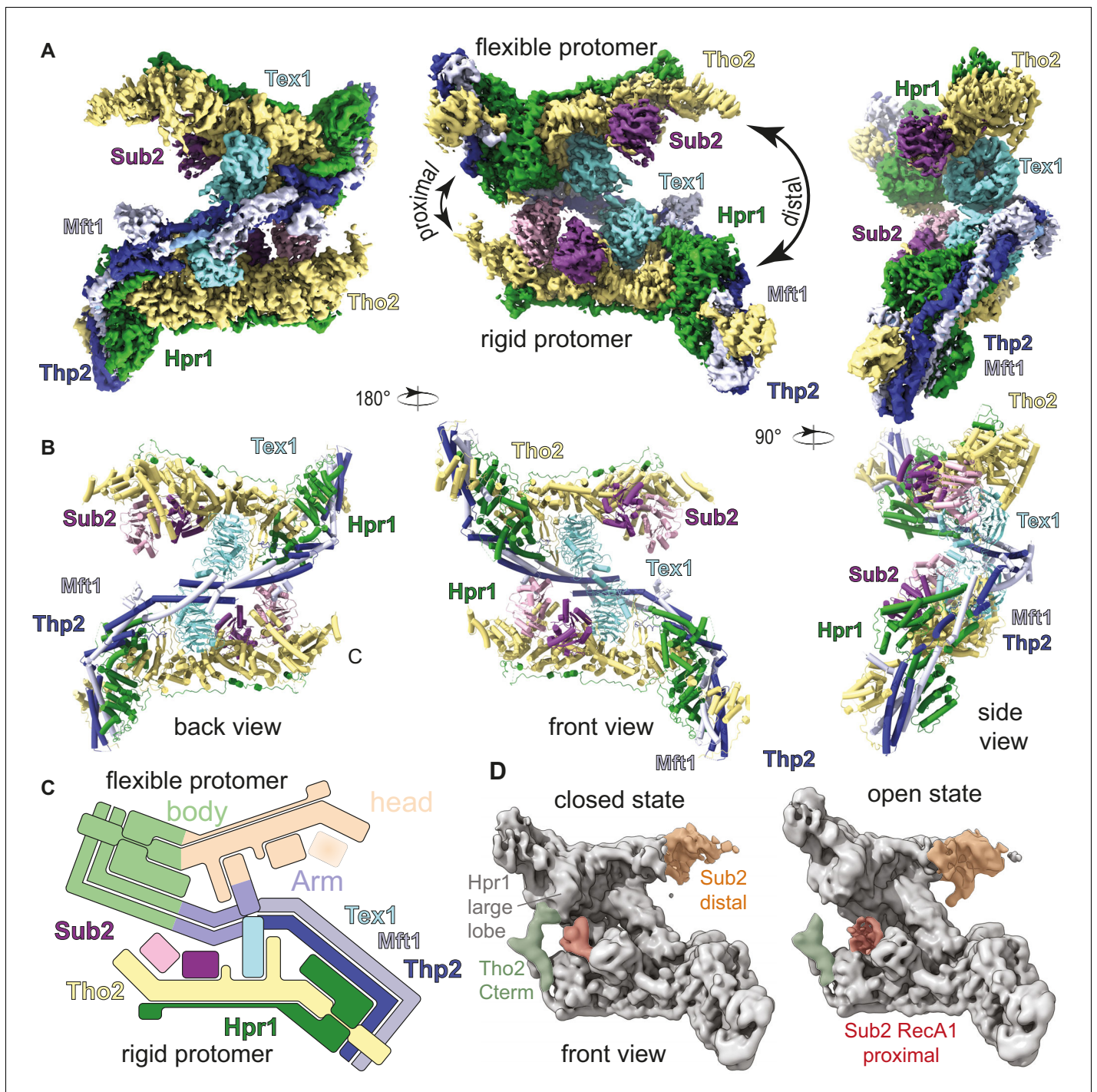
Figure 2. Cryo-electron microscopy (cryo-EM) reconstruction of *Saccharomyces cerevisiae* THO-Sub2 homodimer. (**A**) Segmented cryo-EM reconstruction of the THO-Sub2 dimer. Three different views are shown; proteins and domains are colored as in *Figure 1A*. Features discussed in the text are indicated, including the proximal and distal sides of the asymmetric homodimer, the rigid and flexible protomers, as well as the 'head and 'body' of each protomer. (**B**) Cartoon representation of the structure, shown in the same orientations and colors. Helices are rendered as solid cylinders. (**C**) Schematic representation of the THO-Sub2 complex architecture based on the cryo-EM structure. (**D**) Two frames of the raw cryo-EM data outputs from the variance analysis shown in *Video 1*. The Tho2 C-termini of the two protomers are shown in orange and green. Different conformations are adopted as the dimer switches the proximal and distal sides.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Cryo-electron microscopy analysis.

Figure supplement 2. High-resolution THO-Sub2 density.

Figure 2 continued on next page

Overall structure of yeast THO-Sub2 protomer

In each THO protomer, the proteins Tho2 and Hpr1 together create the platform of the complex (*Figure 3A*). The platform is about 200 Å in length, with an oval 'head' and with an extended 'body' formed almost exclusively by alpha-helices. The 'head' is formed by helical repeats of Hpr1 and a loosely structured region of Tho2. Conversely, the 'body' is formed by an extended array of helical repeats of Tho2 (mostly bi-helical HEAT repeats [*Andrade et al., 2001*]) and a largely unstructured segment of Hpr1. The two smallest THO proteins, Mft1 and Thp2, interact to form the 'arm' of the complex, a striking 200 Å long coiled-coil structure that is oriented diagonally with respect to the Tho2-Hpr1 platform. The Tho2-Hpr1 platform can arbitrarily be dissected into five contiguous modules (module-1 to -5) (*Figure 3B*).

The 'head' of the Tho2-Hpr1 platform binds Mft1-Thp2

The 'head' of each protomer comprises the first two Tho2-Hpr1 modules and forms the binding platform for Mft1-Thp2 (*Figure 3A*). Module-1 consists of the N-terminal portions of Tho2-Hpr1 and recognizes the N-terminal portion of the Mft1-Thp2 coiled coil (*Figure 3C* and *Figure 3—figure supplement 1*). Here, a four-helix bundle, consisting of a pair of helices from each Mft1 and Thp2, is sandwiched between a small HEAT-repeat lobe of Hpr1 and a composite helical fold formed by intercalating helices of Tho2 and Mft1 (*Figure 3A*).

Module-1 connects to module-2 via an extended segment of Tho2 and a large network of conserved interactions (*Figure 3—figure supplements 1–3*). Module-2 contains a large helical lobe of Hpr1 packed against the beginning of the helical array of the 'body' (the Tho2 HEAT1-2 bihelical repeats) (*Figure 3A*). Module-2 recognizes the middle portion of Mft1-Thp2, a two-helix coiled coil that is clamped within the two halves of the Hpr1 large helical lobe (*Figure 3D* and *Figure 3—figure supplement 3*). Although Mft1 and Thp2 share little conservation with their metazoan orthologues, the surfaces to which they bind on Tho2-Hpr1 contain evolutionary conserved residues (*Figure 3—figure supplements 1–3*), suggesting that the two predicted orthologues in the metazoan complex (THOC5 and THOC7) may share similar architectural features. In module-2, the Hpr1 large lobe and Mft1 also feature well-structured loops that latch on to module-3 (*Figure 3B and D*).

The 'body' of the Tho2-Hpr1 platform binds Tex1 and Sub2

The 'body' of the complex contains the other three modules organized around the helical repeat array of Tho2 and the extended region of Hpr1 encompassing residues 501–605 (*Figure 3A*). Module-3 is a MIF4G-like fold with five Tho2 helical repeats (HEAT 3–7) lined underneath by Hpr1 residues 501–522 (*Figure 3E* and *Figure 3—figure supplement 2*). This module is characterized by two prominent insertions. A Tho2 β-hairpin insertion (between HEAT 5 and 6) forms a 50 Å protrusion that extends longitudinally on the concave side of the domain (*Figure 3E*). This insertion is buttressed along its length by Mft1 and Hpr1 loops from module-2 and reaches with its tip the C-terminal coiled-coil portion of Mft1-Thp2 (*Figure 3B*). Module-3 provides the Tex1-binding platform: the curved surface of the Tex1 β-propeller (blades 4 and 5) binds the concave side of the MIF4G-like fold (at Tho2 HEAT 6 and 7) with conserved interactions (*Figure 3—figure supplement 1F*). In addition,
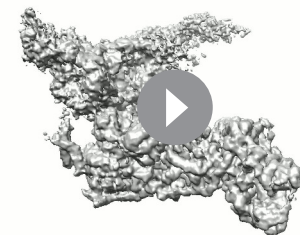
Video 1. Dynamic character of the complex extracted from cryo-electron microscopy data. Variance analysis of the THO-Sub2 complex structure, showing the swiveling motion of the two protomers with respect to each other.
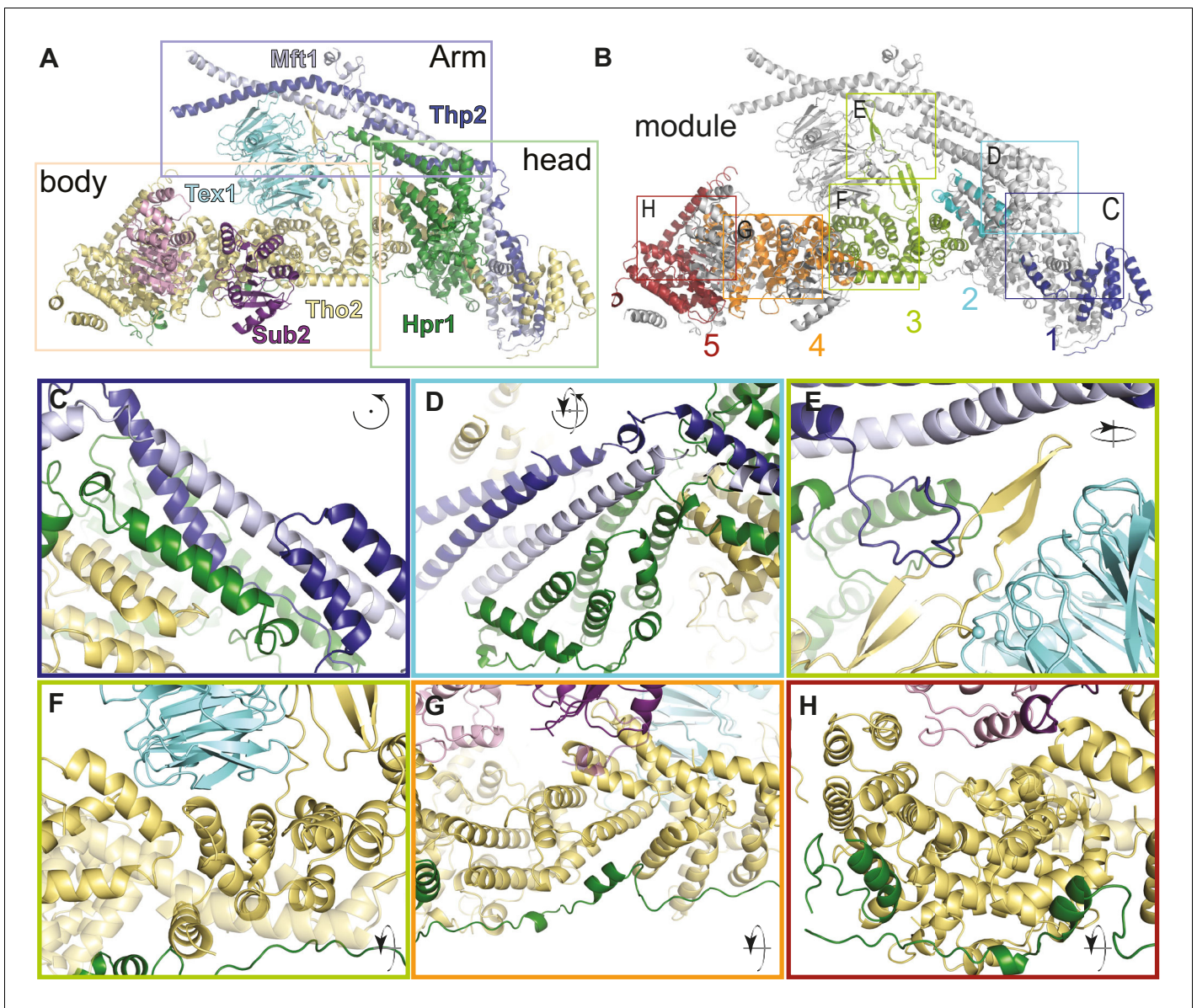https://elifesciences.org/articles/61467#video1

Figure 3. THO complex is built from intertwined conserved interactions. (A) Front view of the THO-Sub2 rigid protomer shown as a cartoon backbone representation. (B) Same view as A with the five modules of the Tho2-Hpr1 platform in different colors. The rectangles highlight the position of the zoom-ins shown in panels C–H. (C–H) Zoom-in views showing the intermolecular interactions between different subunits of a THO-Sub2 protomer as discussed in the text. The cartoon representations show the molecule either in the same view as panel A or after the indicated rotation. Interactions are shown between: (C) Tho2-Hpr1 module-1 and the N-terminal portion of the Mft1-Thp2 coiled-coil; (D) module-2 and central portion of the Mft1-Thp2 coiled-coil; (E) module-3 interactions: Tho2 β-hairpin, bottom surface of Tex1, and loop from the C-terminal Mft1-Thp2 coiled-coil region; (F) Tho2-Hpr1 module-3 and curved surface of Tex1 β-propeller; (G) Tho2-Hpr1 module-4 and Sub2 RecA2 domain; (H) Tho2-Hpr1 module-5 and Sub2 RecA1 domain. The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Structural features of the THO complex.
Figure supplement 2. Tho2 structure-based sequence alignment.
Figure supplement 3. Hpr1 structure-based sequence alignment.
Figure supplement 4. Tex1 structure-based sequence alignment.
Figure supplement 5. Sub2 structure-based sequence alignment.

the bottom surface of the Tex1 β-propeller interacts with the Tho2 β-hairpin (*Figure 3E and F* and *Figure 3—figure supplements 1*, *2,* and *4*). The other prominent insertion in module-3 is a long intra-repeat segment (at Tho2 HEAT 3) that packs along the edge of the MIF4G-like repeats and continues with a loop (referred to as the 'handle') latching on to module-4 (*Figure 3B, F, and G*).

The Sub2-binding platform in the 'body' of the complex is formed by module-4 and module-5 (*Figure 3A*). Module-4 comprises a MIF4G-like fold of Tho2 flanked by tri-helical motifs at both ends (referred to as ARM 8 and ARM 12 and 13) and lined on the convex side by conserved interactions with the Hpr1 extended region spanning residues 523–545 (*Figure 3G* and *Figure 3—figure supplements 1–3*). The RecA2 domain of Sub2 interacts with the concave side of this MIF4G-like fold (at Tho2 ARM 8 and HEAT 9) and contacts the 'handle' from module-3 (*Figure 3G* and *Figure 3—figure supplements 1* and *5*). Module-4 also contains a Tho2 helical insertion (between HEAT 11 and ARM 12) that latches on to module-5. In module-5, Tho2 has a V-shaped fold characterized by a larger side buttressed underneath by Hpr1 residues 546–605 and a smaller side also containing the helical insertion from module-4 (*Figure 3B and H*). In the rigid protomer, the RecA1 domain of Sub2 binds the concave surface of module-5 (Tho2 HEAT 16) and the helical insertion from module-4 (*Figure 3H*). In the flexible protomer, the RecA1 domain does not have ordered density. Another difference between the two protomers is at the ends of the Mft1-Thp2 pairs, which change course roughly at the point where the Tho2 β-hairpin contacts the coiled coil (*Figure 4—figure supplement 1*). Both differences relate to the homodimerization features of the complex, described below.

THO homodimerization is mediated by Thp2-Mft1 and Tex1

In the THO homodimer, the two Tho2-Hpr1 platforms are arranged in an antiparallel fashion, i.e. the 'head' of one protomer faces the 'body' of the other (*Figure 2*). The coiled-coil structures of the two Mft1-Thp2 'arms' extend from opposite ends toward each other forming a chevron-like structure, with the two Tex1 protomers positioned below. The major homodimerization interface in the THO complex is mediated by Mft1-Thp2. The tips of the C-terminal coiled-coil portions of the Mft1-Thp2 pairs intersect at the vertex of the chevron forming a hydrophobic dimerization core (*Figure 4A*). An additional smaller homodimerization interface is mediated by the N-terminal helices of the two Tex1 molecules (*Figure 4B*). These two homodimerization interfaces appear to be the pivot point around which the rest of the Mft1-Thp2 coiled coils and the attached Tho2-Hpr1 platforms swing with a
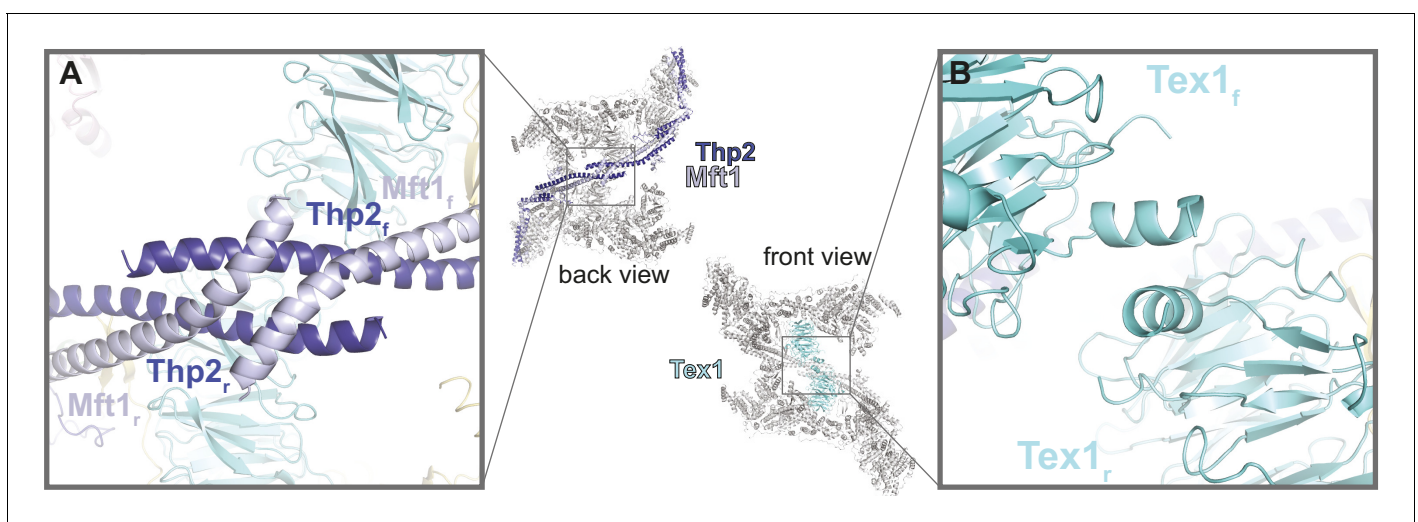

Figure 4. THO homodimerization properties. The central panel shows the back and front views of the THO-Sub2 homodimer, with the whole complex in gray except the dimerization elements highlighted in color: the two Mft1-Thp2 protomers (back view) and the two Tex1 protomers (front view). (**A**) Zoom-in view of the dimerization interface between the C-terminal coiled-coil portions of Mft1 and Thp2 rigid (**r**) and flexible (**f**) protomers (back view of the complex). (**B**) Zoom-in view of the dimerization interface between the N-terminal helices of the two Tex1 protomers (front view of the complex). The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Superposition of the two THO protomers.

seesaw-like movement (*Video 1*; *Figure 2D*). Variance analysis of the cryo-EM data shows that the movement is connected to changes in the two Sub2 proteins.

THO dimer asymmetry is connected to different conformations of two Sub2 ATPases

Like other members of the DEAD-box family of ATPases, Sub2 is expected to undergo conformational changes connected to the nucleic-acid unwinding cycle (*Ozgur et al., 2015b*). In the ATP-RNA-bound active state, the two RecA domains adopt a well-defined closed conformation with a deep ATP-binding crevice and a shallow single-stranded RNA-binding cleft containing a sharp bend that is thought to be important for local RNA unwinding (*Ozgur et al., 2015a*; *Ren et al., 2017*). In the inactive state, DEAD-box proteins generally adopt more variable conformations in which the two RecA domains are spatially restrained by the flexible linker connecting them (*Ozgur et al., 2015b*). Members of the DEAD-box family (eIF4A, Dbp5, and DDX6) have also been shown to adopt an intermediate state upon binding MIF4G-like regulators, which maintain the RecA domains in a semi-closed (activated) conformation that resembles the structure adopted in the active state (*Mathys et al., 2014*; *Montpetit et al., 2011*; *Schütz et al., 2010*).

In the THO-Sub2 cryo-EM reconstruction, the Sub2 protomer at the proximal side of the homodimer adopts an intermediate semi-closed conformation, similar to that observed in other DEAD-box proteins in the activated state (*Mathys et al., 2014*; *Montpetit et al., 2011*; *Schütz et al., 2010*; *Figure 5* and *Figure 5—figure supplement 1*), rationalizing the Sub2 ATPase-activating properties of THO (*Ren et al., 2017*). The interactions that keep Sub2 in the activated conformation appear to be more extensive as compared to those in other DEAD-box proteins, as Sub2 binds at three sites in the 'body' of the rigid protomer (module-4, -5, and the latch; *Figure 5*) and appears to be additionally stabilized by contacts with the 'head' of the flexible protomer. These contacts are not present at the distal side of the homodimer, where the distance between the two protomers is larger and where indeed there is a higher degree of flexibility in the density map (*Figure 2D*). Thus, in our cryo-EM reconstruction of a THO complex in a resting state (e.g. without substrate or energy source), one Sub2 is kept in an activated conformation at the proximal side and the other Sub2 is in a flexible conformation at the distal site. Conformational changes of the Sub2 RecA domains upon substrate binding are expected to trigger changes in the connections with which the modules are
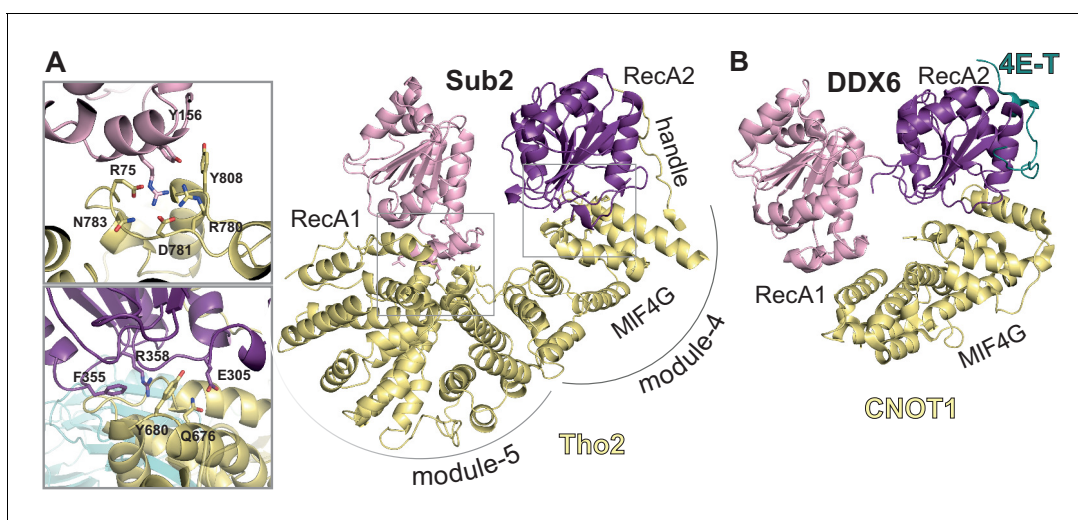

Figure 5. Sub2-activated conformation at the proximal side of the THO homodimer. (**A**) Sub2-Tho2 interaction at the proximal side. The zoom views show a subset of conserved interacting residues. See also *Figure 3—figure supplements 2* and *5*. (**B**) Structure of DDX6-CNOT1-4ET (*Ozgur et al., 2015a*) shown in the same orientation as Sub2-Tho2 in panel A after superposition of their RecA2 domains. Note that the Tho2 'handle' binds RecA2 at the equivalent position as protein 4E-T.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Conformational states of Sub2.

sequentially latched on to each other, thus imparting directionality to the intrinsic seesaw-like movements that we observed in the resting state of the complex.

Mechanistic model for THO/TREX function

In all DEAD-box proteins, the active ATP consumption step depends on the concomitant presence of RNA (*Linder and Jankowsky, 2011*; *Zingler et al., 2008*). In addition, Sub2 requires the last subunit of the TREX complex, Yra1, to efficiently bind RNA (*Ren et al., 2017*). Yra1 has an unusual domain architecture with a central RRM (RNA recognition motif) domain connected by low-complexity sequences to similar and conserved motifs at the N- and C-termini (N-box and C-box motifs): the C-box specifically binds the Sub2 RecA1 domain while the N-box can also bind Sub2 with a similar affinity (*Figure 6—figure supplement 1A*; *Ren et al., 2017*; *Strässer and Hurt, 2001*; *Stutz et al., 2000*). Given the dimeric nature of THO-Sub2, a single Yra1 molecule could thus, in principle, bridge the Sub2 molecules at the proximal and distal sides of the complex (*Figure 6* and *Figure 6—figure supplement 1*). We envision that the RNA-dependent ATP binding and hydrolysis steps of Sub2 allow the complex to travel along the nascent mRNA and prevent the formation of aberrant secondary structures. It is also possible that the Sub2 ATPase cycle may impact the loading of Yra1 onto the mRNA.

The dimeric nature of THO-Sub2 also has important implications for how the complex may function in preventing R loop formation. Like all other helicases, Sub2 binds an RNA strand with fixed polarity, namely with the 3′ end at RecA1 and the 5′ end at RecA2 (*Ren et al., 2017*). In addition, the human Sub2 orthologue UAP56 has been shown to bind and unwind RNA:DNA duplexes (*Pérez-Calero et al., 2020*). Since the two Sub2 proteins are kept in opposite orientation by the two antiparallel Tho2-Hpr1 platforms in the THO dimer, the interacting nucleic acid strands would in turn need to have an opposite orientation. The topology of Sub2 within the complex is therefore compatible with the topology of R-loops, where the two strands of the RNA:DNA hybrid have opposite
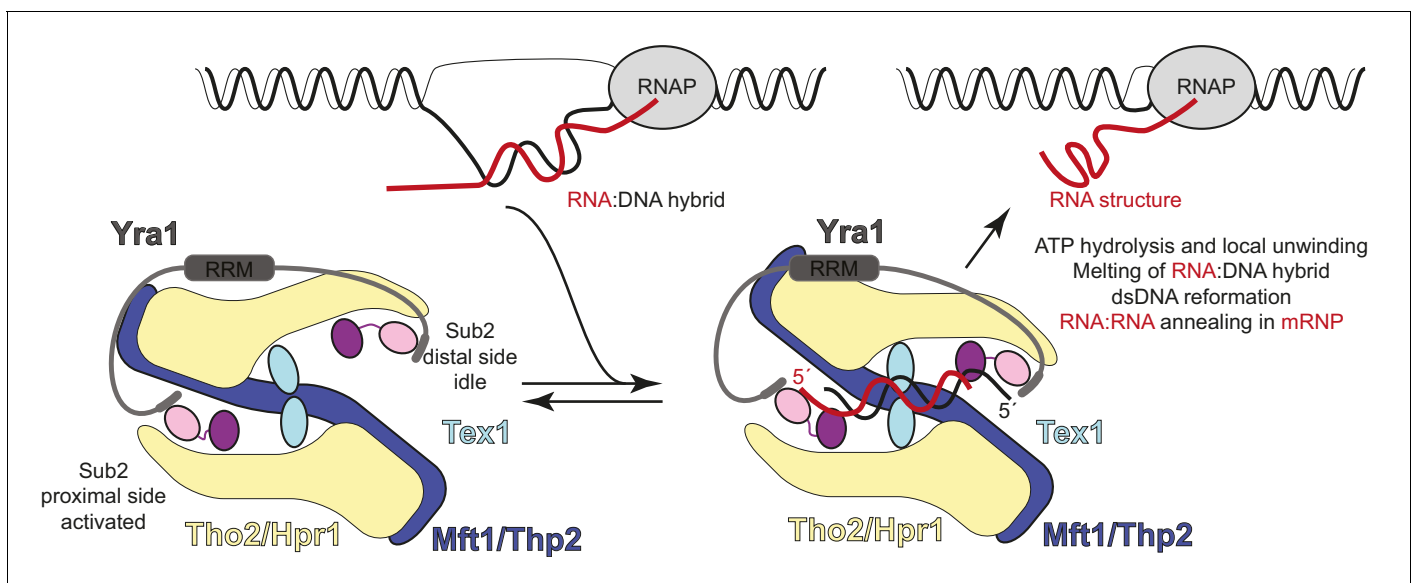

Figure 6. Hypothetical model of transcription-export (TREX) molecular mechanisms. Schematic depicts the TREX complex: on the left at resting state, with Yra1 bound to the Sub2 RecA1 domains via its N-box and C-box motifs (*Ren et al., 2017*); on the right in a substrate-binding state, with an RNA:DNA hybrid positioning the 5′ ends of the RNA (red) and DNA (black) strands at the two opposite RecA1 domains (see also *Figure 6—figure supplement 1*). Binding of the RNA strand to the activated Sub2 would require changes in their relative orientation. In this hypothetical model, the energy released at the RNA-dependent ATP hydrolysis step is harnessed in a mechanical movement as the separated strands are dissociated and the complex returns to the resting state. The involvement of such mechanical force explains how incorporation in the complex may allow Sub2 to resolve RNA:DNA hybrids that would otherwise be too long to be melted by a DEAD-box protein in isolation (*García-Pichardo et al., 2017*; *Linder and Jankowsky, 2011*).

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Hypothetical model of THO-Sub2-Yra1 (transcription-export) in binding RNA:DNA hybrids.

polarity (*Figure 6* and *Figure 6—figure supplement 1*). We propose that after the RNA:DNA hybrid is recognized and melted by binding to the two Sub2 molecules in the complex, the individual strands are released upon ATP hydrolysis and can engage in appropriate interactions due to the vicinity of the other DNA strand and of the RNA-annealing activity of Yra1 (*Figure 6*). Such a mechanism reconciles the intrinsic duality of the TREX complex in mRNA biogenesis, to dissolve RNA:DNA hybrids that can form upon mRNA synthesis, thus preventing R-loop formation, and to favor the annealing of RNA:RNA structures, thus chaperoning the formation of mature mRNPs.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background (*Escherichia coli*)	BL21 Star (DE3) pRARE	EMBL Heidelberg Core Facility		Electrocompetent cells
Cell line (*Spodoptera frugiperda*)	IPLB-Sf21-AE	Gibco		
Cell line (*Trichoplusia ni*)	BTI-Tn-5B1-4	Gibco		
Strain, strain background (*Saccharomyces cerevisiae*)	BY4741 (MATa) yeast	Euroscarf	Y00000	
Strain, strain background (*Saccharomyces cerevisiae*)	BY4743 (MATa/α) yeast	Euroscarf	Y20000	
Antibody	Anti-protein-A IgG (mouse, monoclonal)	Sigma Aldrich	P2921	(1:333 dilution)
Recombinant DNA reagent	Tho2	This paper (Materials and methods)	Uniprot P53552	pFastBac Hta-Tho2 Conti Lab
Recombinant DNA reagent	Hpr1	This paper (Materials and methods)	Uniprot P17629	pFastBac Hta-Hpr1 Conti Lab
Recombinant DNA reagent	Mft1	This paper (Materials and methods)	Uniprot P33441	pFastBac Hta-Mft1 Conti Lab
Recombinant DNA reagent	Thp2	This paper (Materials and methods)	Uniprot O13539	pFastBac Hta-Thp2 Conti Lab
Recombinant DNA reagent	Tex1	This paper (Materials and methods)	Uniprot P53851	pFastBac Hta-Tex1 Conti Lab
Recombinant DNA reagent	Sub2	This paper (Materials and methods)	Uniprot Q07478	3C-GST-fusion Conti Lab
Recombinant DNA reagent	GFP-Sub2	This paper (Materials and methods)		Conti Lab
Commercial assay or kit	Bac-to-Bac Baculovirus Expression System	ThermoFisher Scientific		
Software, algorithm	SerialEM	https://bio3d.colorado.edu/SerialEM/	SerialEM_3-8-0beta8_64 & SerialEM_3-8-0beta11_64	
Software, algorithm	Focus	https://focus.c-cina.unibas.ch/wiki/doku.php	v 1.1.0	
Software, algorithm	cryosparc	doi: 10.1038/nmeth.4169	Cryosparc2	
Software, algorithm	CTFfind4	doi: 10.1016/j.jsb.2015.08.008		
Software, algorithm	TOPAZ	doi: 10.1038/s41592-019-0575-8		
Software, algorithm	UCSF Chimera	UCSF, https://www.cgl.ucsf.edu/chimera/		

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Software, algorithm	UCSF ChimeraX	UCSF, https://www.rbvi.ucsf.edu/chimerax/		
Software, algorithm	COOT	http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/	0.9	
Software, algorithm	Phenix	https://www.phenix-online.org/	PHENIX 1.18	
Software, algorithm	Molprobity	Duke Biochemistry, http://molprobity.biochem.duke.edu/		
Software, algorithm	PyMol 2	PyMOL Molecular Graphics System, Schrodinger LLC	PyMOL 2.1	

Biochemical reconstitution

S. cerevisiae Tho2 (1597 aa, 184 kDa), Hpr1 (752 aa, 88 kDa), Mft1 (392 aa, 45 kDa), and Thp2 (261 aa, 30 kDa) were co-expressed as full-length proteins in insect cells using the pFastBac system. Tex1 (422 aa, 47 kDa) was C-terminally truncated (residues 1–380) to increase protein stability. The proteins were tagged with an N-terminal TEV-cleavable His tag. Cells were infected with 1% (v/v) virus and harvested 72 hr after infection. The cells were pelleted, resuspended, and lysed with a Dounce homogenizer in lysis buffer containing 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol, and 30 mM imidazole, supplemented with complete protease inhibitor (Roche) and benzonase. THO complex was affinity purified using nickel-based affinity chromatography (IMAC, HIS-Select resin from Sigma-Aldrich). After washing with 20 column volumes (CVs) of lysis buffer, the bound THO complex was eluted by increasing the imidazole concentration to 300 mM. For tag-cleavage TEV protease was added and the complex was dialyzed overnight in 50 mM Tris-HCl (pH 7.5), 250 mM NaCl, 5% glycerol, and 5 mM β-mercaptoethanol. For tag-removal the cleaved complex was once passed over IMAC beads and the beads were washed with an extra CV of dialysis buffer. As a final purification step the complex was concentrated with an Amicon 30 kDa cut-off filter and purified via size-exclusion chromatography (Superose6, equilibrated in 25 mM HEPES [pH 7.5], 150 mM NaCl, 5% glycerol, and 2 mM DTT).

Sub2 was expressed as an N-terminal 3C cleavable GST-fusion protein in *E. coli* STAR pRARE cells. After cell lysis by sonication and pelleting of cell debris the supernatant was incubated with GSH affinity beads in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5% glycerol, and 2 mM DTT, supplemented with benzonase and AEBSF. After removal of the GST tag via 3C protease with on-column cleavage, the proteins were further purified via a heparin column and injected onto a size-exclusion chromatography column (Superdex200 16/60) equilibrated in 20 mM HEPES pH 7.5, 200 mM NaCl, and 2 mM DTT. Purified proteins were concentrated, flash frozen in liquid nitrogen, and stored at −80°C. GFP-Sub2 was purified according to the same protocol.

The analytical size-exclusion chromatography to assess the reconstitution of THO and THO-Sub2 shown in *Figure 1A* was carried out on a Thermo Scientific Vanquish Ultra High-Performance Liquid Chromatography (UHPLC) platform, a system with improved separation and equipped with a fluorescence detector. Using this system, the reconstituted THO complex eluted at a molecular mass expected for a dimer. Since addition of Sub2 resulted only in a minor shift of the elution peak, the analytical experiment to confirm complex formation was carried out with GFP-Sub2 and monitored at the appropriate wavelength showing that Sub2-GFP indeed co-migrated with the THO complex. The same reconstitution conditions with untagged Sub2 were used for the sample subjected to cryo-EM single particle analysis.

Native complex isolation

A sequence coding for a C-terminal Twin-Strep-3C-Protein-A tandem affinity tag was inserted into the endogenous HPR1 locus in *S. cerevisiae* strains BY4741 (MATa) and BY4743 (MATa/α) using standard yeast genetics techniques. Yeast were grown in 500 ml YPD to OD600 = 1, harvested by filtration and frozen in liquid nitrogen until processing. Frozen yeast nuggets were lysed with a cryo-mill (SPEX), resuspended in 1 ml ice-cold purification buffer (50 mM potassium phosphate pH 8 and 0.1% NP40) and incubated for 30 min with protein-G dynabeads (Life Technologies) coated with

anti-protein-A IgG (Sigma-Aldrich). Beads were separated from the lysate with a magnet, washed four times in 1 ml purification buffer, resuspended in 200 µl purification buffer containing 250 ng 3C protease, and rotated for 30 min at 4˚C. 3C-eluates were incubated for 30 min at 4˚C with Mag-StrepXT magnetic beads (IBA) before washing beads three times in 1 ml purification buffer and eluting in 20 µl SDS-loading dye. Whole SDS-eluates were separated on 12% SDS-PAGE and stained with Instant Blue (Expedeon).

Cryo-EM sample preparation and data collection

The THO-Sub2 complex was assembled by incubating a fivefold molar excess of Sub2 together with purified THO complex and separating the excess helicase using size-exclusion chromatography. Fractions containing the full complex were pooled and the complex stabilized by mild cross-linking using 1 mM BS3 (bissulfosuccinimidyl suberate) for 15 min at RT. The sample was quenched with ammonium bicarbonate and concentrated to 1 mg/ml using a 30 kDa cut-off table-top concentrator. For cryo-EM sample preparation, 4.0 µl of the purified complex were applied to glow discharged Quantifoil 2/1 grids, blotted for 3.5 s with force '4' in a Vitrobot Mark IV (Thermo Fisher) at 100% humidity and 4˚C, and plunge frozen in liquid ethane cooled by liquid nitrogen.

Electron micrographs were acquired with a FEI Titan Krios transmission electron microscope (ThermoFisher) using SerialEM software (*Schorb et al., 2019*). Movie frames were recorded at a nominal magnification of 64,000 × (calibrated physical pixel size: 1.38 Å/px) using a K3 direct electron detector (Gatan) and a GIF quantum energy filter (Gatan) at 20 eV slit width. The total electron dose of approximately 55 electrons per $Å^2$ was distributed over 40 frames. Cryo-EM micrographs were processed on-the-fly using the Focus software package (*Biyani et al., 2017*). For Tho-Sub2, 6689 micrographs were collected.

Cryo-EM data processing

The THO-Sub2 dataset was processed entirely in CryoSparc (*Punjani et al., 2017*). Dose-fractionated movies were gain-normalized, aligned, and dose-weighted using Patch Motion correction. The contrast transfer function (CTF) was determined using CTFfind4 (*Rohou and Grigorieff, 2015*). A total of 1000 particles were picked manually and used to train a model that was subsequently used to pick the entire dataset using TOPAZ (*Bepler et al., 2019*). A total of 922,935 candidate particles were extracted and cleaned using iterative-rounds of reference-free 2D classification. The 298,569 particles after 2D classification were used for ab initio model reconstruction with the SGD algorithm to prevent model bias. The particles were further iteratively classified in 3D using heterogenous refinement. The 113,076 particles belonging to the best-aligning particles were subsequently subjected to nonuniform 3D refinement, yielding 3.93 Å global resolution and a B-factor of -81.8 $Å^2$. Next, we carried out focused local refinement after signal subtraction on both protomers of the THO-Sub2 dimer. The reconstructions for the rigid and flexible protomers were both significantly improved, indicating a flexible dimeric interface. The flexible and rigid protomers yielded 4.01 Å with a B-factor of -73.6 $Å^2$ and 3.69 Å with a B-factor of -71.1 $Å^2$, respectively.

To further investigate the dynamic motion of the protomers with respect to each other we performed a variance analysis on three modes. For a cleaner visualization and to limit the influence of high-frequency noise, the resolution was filtered to 5 Å. Using the 'simple' output, a linear movie of 20 volumes along each mode was calculated. The volume series along the second mode is shown to contain the swiveling motion anchored at the Tex1-Tex1 interface of the THO-Sub2 dimer (*Video 1*).

The major heterogeneity within the dimer lies in the flexibility between the two protomers. Thus, we reanalyzed the dataset with a mask around the better-resolved protomer, yielding an improved map of the rigid protomer at 3.5 Å. Per-particle local CTF refinement improved the resolution to 3.4 Å with a temperature factor of -108.6 $Å^2$ after nonuniform refinement. This map was used for de novo model building of the THO complex.

Model building

The reconstructed density was interpreted using COOT (*Emsley et al., 2010*). Model building was iteratively interrupted by real-space refinements using Phenix (*Liebschner et al., 2019*). Statistics assessing the quality of the final model were generated using Molprobity (*Chen et al., 2010*). Images of the calculated density and the built model were prepared using UCSF Chimera

(*Pettersen et al., 2004*), UCSF ChimeraX (*Goddard et al., 2018*), and PyMOL. The model we built de novo for the 5-subunit THO complex has very good stereochemistry (see Ramachandran plot, *Figure 2—figure supplement 4A*).

Acknowledgements

Additional information

Author contributions

Sandra K Schuller, Conceptualization, Investigation, Visualization, Writing - original draft; Jan M Schuller, Conceptualization, Formal analysis, Validation, Visualization, Methodology, Writing - original draft; J Rajan Prabu, Conceptualization, Resources; Marc Baumgärtner, Resources, Investigation; Fabien Bonneau, Conceptualization, Investigation, Methodology; Jérôme Basquin, Conceptualization, Supervision, Validation, Investigation, Methodology, Project administration; Elena Conti, Conceptualization, Supervision, Funding acquisition, Writing - original draft, Project administration

Author ORCIDs

Sandra K Schuller https://orcid.org/0000-0002-1800-8014
Jan M Schuller https://orcid.org/0000-0002-9121-1764
J Rajan Prabu https://orcid.org/0000-0002-7726-9310
Fabien Bonneau http://orcid.org/0000-0001-8787-7662
Elena Conti https://orcid.org/0000-0003-1254-5588

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.61467.sa1
Author response https://doi.org/10.7554/eLife.61467.sa2

Additional files

Supplementary files

• Supplementary file 1. Cryo-electron microscopy data collection, refinement, and validation statistics.

• Transparent reporting form

Data availability

Cryo-EM maps are available in the Electron Microscopy Data Bank (11859 and 11871). Atomic models are available in the Protein Data Bank (7APX and 7AQO).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Schuller SK, Schuller JM, Prabu RJ, Baumgartner M, Bonneau F, Basquin J, Conti E	2020	yeast THO-Sub2 complex dimer	https://www.rcsb.org/structure/7AQO	RCSB Protein Data Bank, 7AQO
Schuller SK, Schuller JM, Prabu RJ, Baumgartner M, Bonneau F, Basquin J, Conti E	2020	yeast THO-Sub2 complex	https://www.rcsb.org/structure/7APX	RCSB Protein Data Bank, 7APX
Schuller SK, Schuller JM, Prabu RJ, Baumgartner M, Bonneau F, Basquin J, Conti E	2020	THO-Sub2 highRes	https://www.ebi.ac.uk/pdbe/entry/emdb/EMD-11859	Electron Microscopy Data Bank, 11859
Schuller SK, Schuller JM, Prabu RJ, Baumgartner M, Bonneau F, Basquin J, Conti E	2020	THO-Sub2 dimer	https://www.ebi.ac.uk/pdbe/entry/emdb/EMD-11871	Electron Microscopy Data Bank, 11871

References

Abruzzi KC, Lacadie S, Rosbash M. 2004. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *The EMBO Journal* **23**:2620–2631. DOI: https://doi.org/10.1038/sj.emboj.7600261, PMID: 15192704

Aguilera A, Klein HL. 1988. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae* I. isolation and genetic characterization of hyper-recombination mutations. *Genetics* **119**:779–790. PMID: 3044923

Andrade MA, Petosa C, O'Donoghue SI, Müller CW, Bork P. 2001. Comparison of ARM and HEAT protein repeats. *Journal of Molecular Biology* **309**:1–18. DOI: https://doi.org/10.1006/jmbi.2001.4624, PMID: 11491282

Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, Berger B. 2019. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods* **16**:1153–1160. DOI: https://doi.org/10.1038/s41592-019-0575-8, PMID: 31591578

Biyani N, Righetto RD, McLeod R, Caujolle-Bert D, Castano-Diez D, Goldie KN, Stahlberg H. 2017. Focus: the interface between data collection and data processing in cryo-EM. *Journal of Structural Biology* **198**:124–133. DOI: https://doi.org/10.1016/j.jsb.2017.03.007, PMID: 28344036

Chávez S, Beilharz T, Rondón AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A. 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *The EMBO Journal* **19**:5824–5834. DOI: https://doi.org/10.1093/emboj/19.21.5824, PMID: 11060033

Chávez S, García-Rubio M, Prado F, Aguilera A. 2001. Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **21**:7054–7064. DOI: https://doi.org/10.1128/MCB.21.20.7054-7064.2001, PMID: 11564888

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **66**:12–21. DOI: https://doi.org/10.1107/S0907444909042073, PMID: 20057044

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. *Acta Crystallographica. Section D, Biological Crystallography* **66**:486–501. DOI: https://doi.org/10.1107/S0907444910007493, PMID: 20383002

García-Pichardo D, Cañas JC, García-Rubio ML, Gómez-González B, Rondón AG, Aguilera A. 2017. Histone mutants separate R loop formation from genome instability induction. *Molecular Cell* **66**:597–609. DOI: https://doi.org/10.1016/j.molcel.2017.05.014, PMID: 28575656

García-Rubio M, Chávez S, Huertas P, Tous C, Jimeno S, Luna R, Aguilera A. 2008. Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Molecular Genetics and Genomics* **279**:123–132. DOI: https://doi.org/10.1007/s00438-007-0301-6, PMID: 17960421

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Science* **27**:14–25. DOI: https://doi.org/10.1002/pro.3235, PMID: 28710774

Heath CG, Viphakone N, Wilson SA. 2016. The role of TREX in gene expression and disease. *Biochemical Journal* **473**:2911–2935. DOI: https://doi.org/10.1042/BCJ20160010, PMID: 27679854

Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell* **12**:711–721. DOI: https://doi.org/10.1016/j.molcel.2003.08.010, PMID: 14527416

Jimeno S, Rondón AG, Luna R, Aguilera A. 2002. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *The EMBO Journal* **21**:3526–3535. DOI: https://doi.org/10.1093/emboj/cdf335, PMID: 12093753

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**:845–858. DOI: https://doi.org/10.1038/nprot.2015.053, PMID: 25950237

Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, et al. 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in phenix. *Acta Crystallographica Section D Structural Biology* **75**:861–877. DOI: https://doi.org/10.1107/S2059798319011471, PMID: 31588918

Linder P, Jankowsky E. 2011. From unwinding to clamping - the DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology* **12**:505–516. DOI: https://doi.org/10.1038/nrm3154, PMID: 21779027

Luna R, Rondón AG, Pérez-Calero C, Salas-Armenteros I, Aguilera A. 2019. The THO complex as a paradigm for the prevention of cotranscriptional R-Loops. Cold Spring Harbor Symposia on Quantitative Biology 105–114. DOI: https://doi.org/10.1101/sqb.2019.84.039594

Luo ML, Zhou Z, Magni K, Christoforides C, Rappsilber J, Mann M, Reed R. 2001. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and aly. *Nature* **413**:644–647. DOI: https://doi.org/10.1038/35098106, PMID: 11675789

Mallam AL, Del Campo M, Gilman B, Sidote DJ, Lambowitz AM. 2012. Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* **490**:121–125. DOI: https://doi.org/10.1038/nature11402, PMID: 22940866

Mathys H, Basquin J, Ozgur S, Czarnocki-Cieciura M, Bonneau F, Aartse A, Dziembowski A, Nowotny M, Conti E, Filipowicz W. 2014. Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Molecular Cell* **54**:751–765. DOI: https://doi.org/10.1016/j.molcel.2014.03.036, PMID: 24768538

Meinel DM, Sträßer K. 2015. Co-transcriptional mRNP formation is coordinated within a molecular mRNP packaging station in *S. cerevisiae*. *BioEssays* **37**:666–677. DOI: https://doi.org/10.1002/bies.201400220, PMID: 25801414

Montpetit B, Thomsen ND, Helmke KJ, Seeliger MA, Berger JM, Weis K. 2011. A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* **472**:238–242. DOI: https://doi.org/10.1038/nature09862, PMID: 21441902

Moore MJ, Proudfoot NJ. 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**:688–700. DOI: https://doi.org/10.1016/j.cell.2009.02.001, PMID: 19239889

Müller-McNicoll M, Neugebauer KM. 2013. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nature Reviews Genetics* **14**:275–287. DOI: https://doi.org/10.1038/nrg3434, PMID: 23478349

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**:205–217. DOI: https://doi.org/10.1006/jmbi.2000.4042, PMID: 10964570

Ozgur S, Basquin J, Kamenska A, Filipowicz W, Standart N, Conti E. 2015a. Structure of a human 4E-T/DDX6/CNOT1 complex reveals the different interplay of DDX6-Binding proteins with the CCR4-NOT complex. *Cell Reports* **13**:703–711. DOI: https://doi.org/10.1016/j.celrep.2015.09.033, PMID: 26489469

Ozgur S, Buchwald G, Falk S, Chakrabarti S, Prabu JR, Conti E. 2015b. The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS Journal* **282**:850–863. DOI: https://doi.org/10.1111/febs.13198, PMID: 25645110

Peña A, Gewartowski K, Mroczek S, Cuéllar J, Szykowska A, Prokop A, Czarnocki-Cieciura M, Piwowarski J, Tous C, Aguilera A, Carrascosa JL, Valpuesta JM, Dziembowski A. 2012. Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *The EMBO Journal* **31**:1605–1616. DOI: https://doi.org/10.1038/emboj.2012.10, PMID: 22314234

Pérez-Calero C, Bayona-Feliu A, Xue X, Barroso SI, Muñoz S, González-Basallote VM, Sung P, Aguilera A. 2020. UAP56/DDX39B is a major cotranscriptional RNA-DNA helicase that unwinds harmful R loops genome-wide. *Genes & Development* **34**:898–912. DOI: https://doi.org/10.1101/gad.336024.119, PMID: 32439635

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera–a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**:1605–1612. DOI: https://doi.org/10.1002/jcc.20084, PMID: 15264254

Piruat JI, Aguilera A. 1998. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *The EMBO Journal* **17**:4859–4872. DOI: https://doi.org/10.1093/emboj/17.16.4859, PMID: 9707445

Portman DS, O'Connor JP, Dreyfuss G. 1997. YRA1, an essential *Saccharomyces cerevisiae* gene, encodes a novel nuclear protein with RNA annealing activity. *RNA* **3**:527–537. PMID: 9149233

Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**:290–296. DOI: https://doi.org/10.1038/nmeth.4169, PMID: 28165473

Ren Y, Schmiege P, Blobel G. 2017. Structural and biochemical analyses of the DEAD-box ATPase Sub2 in association with THO or Yra1. *eLife* **6**:e20070. DOI: https://doi.org/10.7554/eLife.20070, PMID: 28059701

Rohou A, Grigorieff N. 2015. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**:216–221. DOI: https://doi.org/10.1016/j.jsb.2015.08.008, PMID: 26278980

Rondón AG, Jimeno S, Aguilera A. 2010. The interface between transcription and mRNP export: from THO to THSC/TREX-2. *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1799**:533–538. DOI: https://doi.org/10.1016/j.bbagrm.2010.06.002, PMID: 20601280

Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics* **16**:583–597. DOI: https://doi.org/10.1038/nrg3961, PMID: 26370899

Schorb M, Haberbosch I, Hagen WJH, Schwab Y, Mastronarde DN. 2019. Software tools for automated transmission electron microscopy. *Nature Methods* **16**:471–477. DOI: https://doi.org/10.1038/s41592-019-0396-9, PMID: 31086343

Schütz P, Wahlberg E, Karlberg T, Hammarström M, Collins R, Flores A, Schüler H. 2010. Crystal structure of human RNA helicase A (DHX9): structural basis for unselective nucleotide base binding in a DEAD-box variant protein. *Journal of Molecular Biology* **400**:768–782. DOI: https://doi.org/10.1016/j.jmb.2010.05.046, PMID: 20510246

Singh G, Pratt G, Yeo GW, Moore MJ. 2015. The clothes make the mRNA: past and present trends in mRNP FashionGuramrit. *Annual Review of Biochemistry* **84**:325–354. DOI: https://doi.org/10.1146/annurev-biochem-080111-092106, PMID: 25784054

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**:436–439. DOI: https://doi.org/10.1038/nature13787, PMID: 25296254

Strässer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondón AG, Aguilera A, Struhl K, Reed R, Hurt E. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**:304–308. DOI: https://doi.org/10.1038/nature746, PMID: 11979277

Strässer K, Hurt E. 2001. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413**:648–652. DOI: https://doi.org/10.1038/35098113, PMID: 11675790

Stutz F, Bachi A, Doerks T, Braun IC, Séraphin B, Wilm M, Bork P, Izaurralde E. 2000. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**:638–650. DOI: https://doi.org/10.1017/S1355838200000078, PMID: 10786854

Wegener M, Müller-McNicoll M. 2018. Nuclear retention of mRNAs - quality control, gene regulation and human disease. *Seminars in Cell & Developmental Biology* **79**:131–142. DOI: https://doi.org/10.1016/j.semcdb.2017.11.001, PMID: 29102717

Wende W, Friedhoff P, Sträßer K. 2019. Mechanism and regulation of Co-transcriptional mRNP assembly and nuclear mRNA export. *Advances in Experimental Medicine and Biology* **1203**:1–31. DOI: https://doi.org/10.1007/978-3-030-31434-7_1, PMID: 31811629

Xie Y, Ren Y. 2019. Mechanisms of nuclear mRNA export: a structural perspective. *Traffic* **20**:829–840. DOI: https://doi.org/10.1111/tra.12691, PMID: 31513326

Zenklusen D, Vinciguerra P, Wyss JC, Stutz F. 2002. Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Molecular and Cellular Biology* **22**:8241–8253. DOI: https://doi.org/10.1128/MCB.22.23.8241-8253.2002, PMID: 12417727

Zingler N, Solem A, Pyle AM. 2008. Protein-Facilitated ribozyme folding and catalysis. Nucleic Acids Symposium Series 67–68. DOI: https://doi.org/10.1093/nass/nrn034