Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2

Yihu Xie[a,*], Bradley P. Clarke[a], Yong Joon Kim[b,c], Austin L. Ivey[a], Pate S. Hill[a], Yi Shi[b,c], and Yi Ren[a,*]

[a] Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN, USA;
[b] Department of Cell Biology, University of Pittsburgh, Pittsburgh, PA, USA.
[c] Medical Scientist Training Program, University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA

*Correspondence to: yihu.xie@vanderbilt.edu or yi.ren@vanderbilt.edu

Abstract

The evolutionarily conserved TREX complex plays central roles during mRNP (messenger ribonucleoprotein) maturation and export from the nucleus to the cytoplasm. In yeast, TREX is composed of the THO sub-complex (Tho2, Hpr1, Tex1, Mft1, and Thp2), the DEAD box ATPase Sub2, and Yra1. Here we present a 3.7 Å cryo-EM structure of the yeast THO•Sub2 complex. The structure reveals the intimate assembly of THO revolving around its largest subunit Tho2. THO stabilizes a semi-open conformation of the Sub2 ATPase via interactions with Tho2. We show that THO interacts with the SR-like protein Gbp2 through both the N-terminal domain and RRM domains of Gbp2. Crosslinking mass spectrometry analysis supports the extensive interactions between THO and Gbp2, further revealing that RRM domains of Gbp2 are in close proximity to the C-terminal domain of Tho2. We propose that THO serves as a landing pad to configure Gbp2 to facilitate its loading onto mRNP.

Introduction:

Eukaryotic RNA transcription is carried out in the nucleus by the RNA polymerases. During an early stage of mRNA transcription, a 5' cap is added to the newly synthesized mRNA, which is followed by splicing, 3'-end processing and polyadenylation. Nuclear mRNA biogenesis culminates in their export through the nuclear pore complex to the cytoplasm. Many protein factors including serine-arginine (SR) proteins associate with mRNAs to form mature mRNPs for export (Metkar et al., 2018; Singh et al., 2012). The evolutionarily conserved TRanscript-EXport (TREX) complex plays key roles in the highly coordinated mRNP assembly and export (Carmody & Wente, 2009; Chavez et al., 2000; Luo et al., 2001; Strasser & Hurt, 2001; Strasser et al., 2002; Viphakone et al., 2019; Xie & Ren, 2019; Zhou et al., 2000). TREX is recruited to actively transcribed genes (Cheng et al., 2006; Masuda et al., 2005; Strasser et al., 2002) and impacts transcription especially during elongation (Dominguez-Sanchez et al., 2011; Y. Zhang et al., 2016).

The C-terminal domain of the largest subunit of RNA Pol II is highly phosphorylated on the heptapeptide repeats (YSPTSPS) at the Serine 2 position during the elongation phase of the transcription cycle (Hsin & Manley, 2012). Serine 2 phosphorylation coordinates loading of co-transcriptional 3' end processing factors to the transcription machinery (Ahn et al., 2004). In yeast, the primary RNA Pol II CTD Ser2 kinase is the CTDK-1 complex (Cho et al., 2001; Sterner et al., 1995; Wood & Shilatifard, 2006). Growing evidence links the function of TREX and transcriptional CDKs. The yeast TREX component Mft1 interacts genetically with CTDK-1 (Hurt et al., 2004). In addition to their roles during transcription elongation, TREX and CTDK-1 both influence mRNA 3'-end processing and

56   polyadenylation (Ahn et al., 2004; Rougemaille et al., 2008; Saguez et al., 2008). In

57   humans, the transcriptional kinases are more divergent, at least CDK11, CDK12, and

58   CDK13 are shown to phosphorylate Ser2 on Pol II CTD, all of which have been

59   recognized as tumor suppressors (Cao et al., 2014; Parua & Fisher, 2020). TREX and

60   CDK11 have been shown to interact in human cells and play roles in regulating HIV

61   mRNA 3' end Processing (Pak et al., 2015).

62

63   The coordination of TREX and CTDK-1 is largely unknown. Several lines of evidence

64   suggest that a group of shuttling SR proteins could serve as the link for THO and CTDK-

65   1. SR proteins are well recognized as splicing factors, but they also play important roles

66   in coordinating transcription and mRNA export (Reed & Cheng, 2005). In yeast, there are

67   three shuttling SR proteins, Gbp2, Hrb1, and Npl3, which play roles in mRNA export by

68   interacting with the mRNA export receptor Mex67•Mtr2 (Hackmann et al., 2014). In

69   humans, three SR proteins, SRSF1, SRSF3, SRSF7 also shuttle between the nucleus

70   and the cytoplasm to facilitate mRNA export by serving as adaptors for the human

71   ortholog of Mex67•Mtr2, the NXF1•NXT1 complex (Huang et al., 2003; Huang & Steitz,

72   2005; Muller-McNicoll et al., 2016).

73

74   In yeast cells, TAP-tagged Gbp2 and Hrb1 were shown to associate with the CTDK-1

75   complex (Hurt et al., 2004). Consistent with this observation, using purified recombinant

76   proteins, we recently showed that Gbp2 RRM domains are sufficient to interact with

77   CTDK-1, involving the N-terminal RS domain in its Ctk1 kinase subunit (Xie et al., 2020).

78   We also found that there is a synthetic growth defect when both CTK1 and GBP2 are

4

79   knocked out in yeast. The physical and functional interactions between Gbp2 and CTDK-

80   1 provide a link between Gbp2 function and the transcription machinery. Interestingly, in

81   humans, CDK11 directly interacts with SRSF7 (Hu et al., 2003), and together with TREX,

82   all are implicated in HIV-1 mRNA 3' end processing (Pak et al., 2015; Valente et al., 2009).

83   Among the three yeast shuttling SR-like proteins, Gbp2 and Hrb1, but not Npl3 have been

84   shown to rely on the THO components Hpr1 and Mft1 to load onto mRNPs (Hacker &

85   Krebber, 2004). The different requirements could stem from an interaction between THO

86   and Gbp2 and Hrb1, but not Npl3 (Hurt et al., 2004; Martinez-Lumbreras et al., 2016).

87

88   Despite extensive studies, how TREX, SR proteins, and CTDK-1 coordinately function

89   during mRNA biogenesis is still not clear. To elucidate the molecular mechanisms, we

90   conducted biochemical and structural studies on the yeast TREX complex and Gbp2.

91   Yeast TREX is a ~470 kDa protein complex comprised of the pentameric THO sub-

92   complex (Tho2, Hpr1, Tex1, Mft1, and Thp2), the DEAD box ATPase Sub2, and Yra1.

93   Thus far structural understanding of the TREX complex has been limited to low resolution

94   structures (Pena et al., 2012; Ren et al., 2017). Here we present a 3.7 Å cryo-EM structure

95   of the yeast THO•Sub2 complex to reveal the molecular details of the THO complex

96   assembly and the THO-Sub2 interactions. We demonstrate direct binding between THO

97   and Gbp2 using recombinant proteins and dissect their mode of interaction using *in vitro*

98   binding studies and crosslinking mass spectrometry (XL-MS) analysis of the THO-Gbp2

99   complex. Together, we propose that TREX serves as a landing pad to configure the multi-

100   domain Gbp2 and facilitate its loading onto the mRNP.

101

102 **Results and Discussion**

103 **THO directly interacts with the SR-like protein Gbp2**

104 We began by testing the interaction between the THO complex and Gbp2 using purified

105 recombinant proteins. The ~400 kDa THO complex consisting of full-length Tho2, Hpr1,

106 Tex1, Mft1, and Thp2 subunits (denoted by THO-FL, Figure 1A) was expressed in insect

107 cells. Full-length Gbp2 was expressed in insect cells with an N-terminal GST-tag. Using

108 GST pull down assays, we show that Gbp2 directly interacts with THO-FL (Figure 1B).

109 We next tested the binding of Gbp2 to a THO core complex (denoted by THO*, Figure

110 1A) that contains the ordered regions of all THO's five subunits. We found that THO* is

111 capable of binding to Gbp2, but with reduced interaction compared to THO-FL (Figure

112 1B). These results suggest that multiple regions in THO are involved in Gbp2 recognition,

113 including both the THO core and the potentially flexible regions that are truncated in THO*.

114

115 We next attempted to dissect the domains in Gbp2 that are involved in THO interaction.

116 Gbp2 contains an N-terminal RS domain (NTE) followed by three tandem RRM domains,

117 RRM1, RRM2, and RRM3 (Figure 1A). RRM1 and RRM2 domains are capable of binding

118 to RNA. RRM3 was shown to recognize THO (Martinez-Lumbreras et al., 2016).

119 Interestingly, we found that Gbp2 without RRM3 still binds to THO (Figure 1C). On the

120 other hand, deletion of the N-terminal RS domain of Gbp2 substantially reduced THO

121 interaction, suggesting that the Gbp2 RS domain is required for stable binding to THO.

122

123 Together, our binding studies indicate that THO-Gbp2 interaction involves multiple

124 domains from both THO and Gbp2. To provide insights into the underlying molecular

6

125 mechanisms of the THO-Gbp2 recognition, we take an integrative approach combining

126 cryo-EM structure determination of the THO* core complex and XL-MS analysis of the

127 THO-FL interaction with Gbp2.

128

129 **Cryo-EM structure of the THO*•Sub2 complex at 3.7 Å resolution**

130 The THO complex is an integrated structural and functional unit that regulates the activity

131 of Sub2. We previously determined a THO•Sub2 crystal structure at 6.0 Å resolution (Ren

132 et al., 2017). Here, we carried out single particle cryo-EM studies on THO*•Sub2. For

133 cryo-EM sample preparation, the THO*•Sub2 complex was subjected to crosslinking with

134 glutaraldehyde to obtain a more homogeneous sample. We found that THO*•Sub2 forms

135 a higher ordered assembly composed of four THO*•Sub2 protomers (Figure 2-figure

136 supplement 1). This tetrameric assembly has a two-fold symmetry, and it correlates with

137 the asymmetric unit content in our previously determined THO•Sub2 crystal structure.

138 Two of the four THO*•Sub2 protomers are well ordered, whereas the peripheral two

139 protomers are significantly more flexible. We observed an "arch" and a "bridge" that

140 connect a "rigid" THO*•Sub2 protomer and a "mobile" protomer (Figure 2-figure

141 supplement 1D). Our observations suggest that these two protomers are likely

142 components of the functional assembly observed in THO complex isolated from yeast

143 cells (Pena et al., 2012), which is also consistent with the recently published THO•Sub2

144 structure (Schuller et al., 2020). Details of these features at the THO-THO dimer interface

145 will be discussed in the later section of our XL-MS studies. For obtaining the best quality

146 map for model building, the THO*•Sub2 protomer was extracted from the most ordered

147 two copies and refined to an overall resolution at 3.7 Å (Figure 2-figure supplement 1).

148 The electron density map allows us to build an atomic model of the THO complex de novo

149 (Figure 2A and Figure 2-figure supplement 2). The THO model contains two thousand

150 residues with 90% assigned residue register. Sub2 was modeled using our previously

151 determined crystal structure (Ren et al., 2017). By having the resolution to build an atomic

152 model, we now reveal the molecular details of the structural core of THO and its

153 interaction with Sub2.

154

155 The structure of the THO complex reveals intimate folding of the five subunits (Figure 2B

156 and Figure 2-figure supplement 3). Tho2, the largest subunit spanning the entire length

157 of the elongated THO, plays a critical role in THO assembly. Tho2 can be dissected as

158 "head", "neck", and "trunk" sections. The Tho2 "head" contains an N-terminal helical

159 bundle that clusters with the N-termini of Hpr1, Mft1, and Thp2. Tho2 "neck" is comprised

160 of a helix followed by a loop. The "neck" is embraced by a bi-lobed Hpr1 (lobe A and lobe

161 B). Tho2 "trunk" folds into an alpha solenoid structure, which binds the Tex1 β-propeller

162 at its center and stabilizes a semi-open Sub2 ATPase at its C-terminal end. An extended

163 region at Hpr1 C-terminal region forms a "belt" lining the Tho2 "trunk".

164

165 **Assembly of the THO•Sub2 complex**

166 Tho2 is the main scaffold upon which other THO constituents assemble. Tho2 features a

167 total contact area of ~9000 Å^2 with the other four THO subunits. Tho2 "head" domain

168 binds to a four helix bundle, formed by two pairs of anti-parallel helices contributed by

169 Mft1 and Thp2, respectively (Figure 3A). Tho2 "head" and the helix in its "neck" sandwich

170 the very N-terminal helix of Mft1 (residues 6-17). The opposite side of the Mft1/Thp2 four

8

171   helix bundle runs in parallel with Hpr1 lobe A (residues 1-230). The Tho2 "neck",

172   particularly the loop (residues 167-179), is embraced by the Hpr1 lobe A and lobe B

173   (residues 250-490) (Figure 3B). Although the "neck" is largely buried, it contains multiple

174   hydrophilic residues including K171, N173 and E177. Tho2 and Hpr1 residues at this

175   interface are highly conserved from yeast to human (Figure 2-figure supplement 3).

176

177   The "trunk" of Tho2 (residues 180 to 1200) forms an alpha-solenoid. Hpr1 "belt" contains

178   residue assignment from residues 491 to 535 (Figure 3C). It starts from the beginning of

179   the Tho2 "trunk", featuring aromatic residues at the interface including F511, F515, F518,

180   and W532, and likely extends further to the C-terminus of Tho2 "trunk" as evidenced by

181   our XL-MS studies discussed later. The seven-bladed Tex1 β-propeller sits at the center

182   of the Tho2 "trunk" via blade 4 and 5 (Figure 3D). The loops connecting blade 4/5 (4D5A)

183   and 5/6 (5D6A), as well as the 5BC loop within blade 5 contact a pair of Tho2 helices

184   (residues 626-666), whose opposite side binds to the C-terminal RecA domain of Sub2

185   (Sub2-C). This Tho2-Tex1 interaction is conserved from yeast to human based on the

186   sequence homology (Figure 2-figure supplement 3). In addition, a prominent extension

187   from Tho2 is projected outward perpendicular to the Tho2 "trunk". The C-terminal part of

188   this extension (residues 464-485) forms a hairpin that winds through the bottom face of

189   the Tex1 β-propeller. This additional Tho2-Tex1 binding mode is likely a yeast specific

190   mechanism as human and other metazoan THOs lack this extension (Figure 2-figure

191   supplement 3).

192

193   Regulation of the enzymatic activity of the DEAD-box ATPase is vital to the stepwise

194   remodeling reactions mediated by the TREX complex (Xie & Ren, 2019). We previously

195   showed that THO stimulates the ATPase activity of Sub2 (Ren et al., 2017). The cryo-EM

196   structure provides new insights into the molecular details of their interaction. Overall, THO

197   stabilizes a semi-open conformation of Sub2 by interacting with both RecA domains

198   (Sub2-N and Sub2-C). Comparison of the cryo-EM structure and our previous THO•Sub2

199   crystal structure shows that these two structures are in excellent agreement (Figure 3-

200   figure supplement 1). The cryo-EM structure reveals the atomic details of the THO-Sub2

201   interactions at the Sub2-C interface (Figure 3E). Sub2-C makes contacts with two pairs

202   of Tho2 helices (residues 625-695). The Sub2 loop consisting of residues 304-308 is

203   situated at the center of the interface featuring electrostatic interactions via E305 and

204   N307. In addition, another Sub2 loop consisting of residues 355-358 makes critical

205   contacts via F355 and R358. The importance of this loop is evidenced by our previous

206   mutagenesis studies that show the ATPase activity of Sub2 mutant E356A/K357A/R358A

207   cannot be activated by THO (Ren et al., 2017). This Sub2 activation mechanism is a

208   conserved mechanism shared by several other DEAD-box proteins including Dbp5 which

209   functions at the terminal step of nuclear mRNA export at the cytoplasmic side of the

210   nuclear pore complex (Folkmann et al., 2011; Mathys et al., 2014; Montpetit et al., 2011;

211   Schutz et al., 2008).

212

213   **XL-MS analysis of the THO•Gbp2 complex**

214   The THO complex contains a significant amount of potentially flexible regions including

215   ~400 residues at the Tho2 C-terminal end and ~150 residues at the Hpr1 C-terminal end.

216 These flexible regions are presumably not suitable for structural studies. Our binding

217 studies show that these flexible regions are involved in Gbp2 recognition (Figure 1B). To

218 gain further insights into the THO complex arrangement and the THO-Gbp2 interaction,

219 we took a XL-MS approach (Chait et al., 2016; Leitner et al., 2016; Yu & Huang, 2018) to

220 analyze the complex between THO-FL and Gbp2. We used both EDC and DSS, a

221 carboxyl and amine-reactive crosslinker and an amine specific crosslinker that crosslink

222 residues with $C\alpha$-$C\alpha$ distance less than 17 Å and 30 Å, respectively (Kim et al., 2018; Shi

223 et al., 2014). We obtained a total of 200 unique EDC crosslinks, of which 69 were

224 interprotein crosslinks including 9 crosslinks between Tho2 and Gbp2. We also obtained

225 a total of 133 unique DSS crosslinks to complement the EDC crosslink data, with 53 of

226 these crosslinks being interprotein crosslinks. (Figure 4A, Figure 4A-figure supplement 1,

227 and Table 2). The crosslinking data is highly consistent with the THO structure (Figure 4-

228 figure supplement 1B and 1C). 91% and 100% of the EDC and DSS crosslinks that can

229 be mapped to the structure fall within the expected distance restraint.

230

231 Crosslinks between THO subunits provide insights into the C-terminal domain of THO

232 (Tho2-CTD, residues 1200-1597) downstream of the "trunk" domain and the role it plays

233 on the arrangement of the THO-THO dimer. The Tho2-CTD contains a "bridge" that

234 connects THO to the neighboring THO molecule as indicated by our cryo-EM density map

235 (Figure 4B, Figure 2-figure supplement 1). Comparison with the recently published THO-

236 Sub2 structure reveals that the bridge starts at Tho2 residue 1200 (Schuller et al., 2020).

237 The "bridge" is followed by a structured segment, as suggested by the clustered

238 crosslinking between Tho2 (residues 1260 to 1369) and the Hpr1 lobe B (E297, D434,

11

239   K462, and K467) as well as Mft1 D129 (Figure 4B). In line with our observation, Tho2

240   (residues 1279-1405) was shown to form a rigid core through proteolysis and it folds into

241   a helical structure as indicated by CD spectra (Pena et al., 2012). Importantly, crosslinking

242   involving the structured segment indicates that the Tho2-CTD crosses over to the

243   neighboring THO near its Hpr1 lobe B. The structured segment is followed by a highly

244   flexible tail (residues ~1400-1597), as this region crosslinks to spatially separated

245   residues. For example, Tho2 K1576 crosslinks to both Hpr1 lobe B (E297 and D434) and

246   Tex1 (D341). In support of the flexibility of the Tho2 tail, a previous study showed that

247   Tho2 (1411-1530) was highly sensitive to trypsin digestion (Pena et al., 2012). Our XL-

248   MS data further supports the observed THO-THO dimer interface at the "arch", which is

249   composed of Mft1 and Thp2 from two THO molecules (Schuller et al., 2020). We identified

250   multiple crosslinks involving the "bulge" (Mft1, residues 142-196) including Mft1-

251   K182/Tho2-K1103, Mft1-K165/Tho2-K967, Mft1-K170/Tho2-K967, and Mft1-K174/Tho2-

252   K967 (Figure 2-figure supplement 1D and Table 2). Structural comparison of our cryo-EM

253   structure and the recently published structure reveals significant flexibility in the relative

254   orientation between two THO molecules (Figure 4-figure supplement 2) (Schuller et al.,

255   2020). Of note, as the dimerization of THO is asymmetric, the "bridge" is only observed

256   at the proximal side of the THO dimer. It is conceivable that the Tho2-CTD will exhibit

257   more significant flexibility at the distal side of the THO dimer. Our data also provide

258   insights into the arrangement of the Tex1 C-terminal tail (residues 367-422) and Hpr1-

259   CTD (residues 600-752) (Figure 4-figure supplement 2). The extensive crosslinks

260   observed between Tho2-CTD and Hpr1-CTD suggests that they are spatially close to

261   each other and are likely localized in between two THO molecules. Together, XL-MS

262   results provide critical insights into the regions in THO that are not visible in the cryo-EM

263   structures.

264

265   Crosslinking between Tho2 and Gbp2 indicates that Tho2-CTD is in close proximity to all

266   three Gbp2 RRM domains (Figure 4A and 4C). Each of the three RRM domains crosslinks

267   to the structured segment in Tho2-CTD: RRM1-K190 to Tho2-K1349, RRM2-E241 to

268   Tho2-K1250, and RRM3-D367 to Tho2-K1335. These results suggest that Gbp2 is

269   localized in between two THO molecules near Hpr1 lobe B, as these involved Tho2

270   residues (K1250 and K1335) are crosslinked to Hpr1 lobe B (Figure 4B). Our data also

271   show that each RRM domain crosslinks to the highly flexible tail in Tho2-CTD. It is

272   possible that, in the presence of Gbp2, the Tho2 tail may assume a more specific

273   conformation.

274

275   Our XL-MS results (Figure 4C), together with the *in vitro* binding studies (Figure 1B and

276   1C), demonstrate that Tho2-CTD contributes to Gbp2 interaction. The C-terminal domain

277   of Tho2 also binds to RNA/DNA (Pena, 2012). The function of Tho2-CTD *in vivo* was

278   supported by the growth defect of *tho2-ΔCTD* yeast strains (Pena et al., 2012).

279   Importantly, the synthetic growth defect of *tho2-ΔCTD* and *Δgbp2* strains highlights their

280   functional links (Martinez-Lumbreras et al., 2016).

281

282   As both Gbp2 and Sub2 bind to the C-terminal region of Tho2, we next asked whether

283   Gbp2 and Sub2 can associate with the THO complex together. GST-Gbp2 was used to

284   pull down THO in the presence of Sub2. We found that GST-Gbp2 is able to pull down

13

285   both THO and Sub2 (Figure 4D). In addition, THO and Sub2 appear to be in a

286   stoichiometric amount relative to each other. Our results suggest that THO, Sub2, and

287   Gbp2 can form a THO•Sub2•Gbp2 complex, and therefore Gbp2 could function together

288   with the TREX complex during nuclear mRNP maturation.

289

290   **Working model for coordinated function of TREX and Gbp2**

291   Together with our recent characterization of Gbp2 interaction with the RNA Pol II Ser2

292   kinase CTDK-1 complex, we propose a working model for the coordinated function

293   between TREX, Gbp2, and CTDK-1 (Figure 5). Gbp2 interaction with CTDK-1 provides

294   a means to associate with the transcription machinery (Hurt et al., 2004; Xie et al., 2020).

295   We envision that TREX and Gbp2 function coordinately during nuclear mRNP maturation

296   and surveillance. During transcription, faulty assembly of mRNPs is a threat to genomic

297   stability. If the defective mRNPs persist, they need to be sensed by a surveillance system

298   and degraded. In yeast, Gbp2 and Hrb1 were shown to play key roles in mRNP

299   surveillance (Hackmann et al., 2014). Interactions between Gbp2 and Mex67 for export

300   and between Gbp2 and Mtr4 for degradation through the RNA exosome complex are

301   mutually exclusive. TREX travels with the transcription machinery (Meinel et al., 2013)

302   and its function in mRNP assembly is well documented. In THO/Sub2 mutant yeast cells,

303   mRNP assembly is defective and faulty mRNPs cannot be degraded efficiently, which

304   leads to the formation of heavy chromatin (Rougemaille et al., 2008; Saguez et al., 2008).

305   In humans, depletion of TREX complex components leads to R-loop accumulation,

306   transcriptional elongation defects, and trapped mRNP in nuclear speckles. (Dias et al.,

307   2010; Dominguez-Sanchez et al., 2011; Perez-Calero et al., 2020; Wang et al., 2018).

14

308   The extensive interactions between THO and Gbp2 suggest that THO could serve as a

309   landing pad for Gbp2 loading onto mRNPs to function as a key surveillance factor during

310   mRNP maturation. Interestingly, in human cells, TREX was shown to interact with multiple

311   domains of the mRNP export receptor NXF1•NXT1 to facilitate its loading onto mRNPs

312   (Viphakone et al., 2019). Given that Gbp2 interacts with the yeast export receptor

313   Mex67•Mtr2 (Hackmann, 2014), the interplay between TREX, Gbp2, and Mex67•Mtr2

314   during mRNP biogenesis warrants further study.

315

316 **Methods**

317 **Plasmids and proteins**

318 Both THO-FL and the THO* core complex were expressed in High-Five insect cells by

319 coinfection of recombinant baculoviruses. THO-FL contains full length *S. cerevisiae* Tho2

320 (residues 1-1597), Hpr1 (residues 1-752), Tex1 (residues 1-422), Mft1 (residues 1-392),

321 and Thp2 (residues 1-261) subunits and the former four subunits each contains a TEV

322 cleavable N-terminal His tag. The THO* complex contains *S. cerevisiae* Tho2 (residues

323 1-1257), Hpr1 (residues 1-603), Mft1 (residues 1-256), full length Thp2, and *S. bayanas*

324 Tex1 (residues 1-380) with Tho2 and Hpr1 each containing a TEV cleavable N-terminal

325 His tag. High-Five cells were harvested 48 hr after infection. The cells were sonicated in

326 a lysis buffer containing 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM PMSF,

327 5 mg/L aprotinin, 1 mg/L pepstatin, 1 mg/L leupeptin, and 0.5 mM TCEP. THO complexes

328 were purified by Ni affinity chromatography, followed by TEV digestion to remove His tags.

329 The proteins were then purified on a mono Q column (GE Healthcare) and subjected to

330 further size exclusion purification with a Superose 6 column (GE Healthcare) in 10 mM

331 Tris pH 8.0, 150 mM NaCl, and 0.5 mM TCEP.

332

333 GST tagged Gbp2 (residues 1-427) and Gbp2ΔRRM3 (residues 1-316) were expressed

334 in High-Five cells. Cells were lysed in the same condition as the THO complexes. The

335 GST tagged Gbp2 proteins were purified using glutathione sepharose 4B resin (GE

336 Healthcare) followed by size exclusion chromatography using a Superdex 200 column

337 (GE Healthcare) in 10 mM Tris pH 8.0, 300 mM NaCl, and 0.5 mM TCEP.

338

16

339   Sub2 and Gbp2ΔN (residues 107-427) were expressed in Rosetta *E. coli* cells

340   (Stratagene) with an N-terminal TEV cleavable GST tag. Protein expression was induced

341   at an OD_{600} of 1.0 with 0.5 mM IPTG at 20 °C for 16 hrs. Cells were lysed in the same

342   lysis buffer as mentioned above. Proteins were first purified using glutathione sepharose

343   4B resin. For Sub2, the GST tag was removed by TEV, and the protein was purified on a

344   mono Q column. Untagged Sub2 and GST-Gbp2ΔN were further purified on a Superdex

345   200 column in 10 mM Tris pH 8.0, 150 mM NaCl, and 0.5 mM TCEP.

346

347   All purified proteins were concentrated, flash frozen in liquid nitrogen, and stored at -80°C.

348

349   **Cryo-EM Sample preparation and data collection**

350   Purified THO* and Sub2 were first buffer exchanged to 10 mM HEPES pH 7.0, 100 mM

351   potassium acetate, and 0.5 mM TCEP. THO* was incubated with 3-fold molar excess of

352   Sub2 in the presence of 0.05% glutaraldehyde for 30 min at RT. Crosslinking was

353   quenched with 0.1 M Tris pH 8.0 and the sample was concentrated to 0.5 mg/mL. 1.5 μl

354   of THO*•Sub2 was applied to a glow-discharged UltrAuFoil R 1.2/1.3 grids (Quantifoil).

355   Grids were blotted for 3 s with a blotting force of 3 and 100% humidity at 22 °C and

356   plunged into liquid ethane using an FEI Vitrobot Mark IV (Thermo Fisher).

357

358   Electron micrographs were acquired with a Titan Krios electron microscope (Thermo

359   Fisher) equipped with a Falcon 3 detector (Thermo Fisher). Movies were collected with

360   EPU with a calibrated pixel size of 0.681 Å/pixel. A total of 4907 movies were collected

361   with a defocus range from 0.8 µm to 2.0 µm. Description of the cryo-EM data collection

362   parameters can be found in Table 1.

363

364   **Cryo-EM data processing**

365   Motion correction was performed using MotionCor2 (Zheng et al., 2017). The parameters

366   of the contrast transfer function (CTF) were estimated using Gctf (K. Zhang, 2016). We

367   initially selected 396 K particles from 4907 micrographs with automatic particle picking in

368   RELION-3 (Zivanov et al., 2018). The picked particles were binned by 2 and subjected to

369   reference-free 2D classification. 205 K particles were selected for 3D classification with

370   C2 symmetry using an initial model generated by EMAN2 (Tang et al., 2007). Each

371   particle contains four copies of the THO•Sub2 complex with two copies significantly more

372   flexible than the others. 15 K particles were selected for 3D refinement using a mask

373   covering the two ordered THO•Sub2 molecules with C2 symmetry. The particles were

374   then re-extracted at the original pixel size of 0.681 Å/pixel and subjected to Bayesian

375   polishing, CTF refinement, and 3D refinement. Refinement of the entire four copies of

376   THO•Sub2 molecules generated a map at 4.80 Å resolution. We extracted 30 K

377   THO•Sub2 protomers from the ordered two copies and refinement using a mask covering

378   one THO•Sub2 molecule yielded a map of THO•Sub2 at 3.70 Å resolution with a

379   sharpening B factor of 86 Å^2 as assessed by an FSC threshold of 0.143.

380

381   **Model building**

382   The 3.70 Å THO•Sub2 map was used for model building in COOT (Emsley et al., 2010).

383   The five subunit THO complex was built de novo. Individual RecA domains of Sub2 were

18

384 placed using our previously determined atomic resolution structure (PDB ID 5SUP). The

385 THO•Sub2 model was subjected to real-space refinement in Phenix (Adams et al., 2010).

386 The final THO•Sub2 model contains Tho2 (residues 37 to 913, followed by 10 poly-Ala

387 helices at the C-terminus), Hpr1 (residues 4 to 535), Tex1 (residues 68 to 366), Mft1

388 (residues 5 to 227), Thp2 (residues 8 to 227), and Sub2. Figures were prepared using

389 Chimera (Pettersen et al., 2004) or PyMOL (Molecular Graphics System, Schrodinger,

390 LLC).

391

392 **GST pull-down assays**

393 1 μM of GST or GST-tagged Gbp2 variants was incubated with 1 μM of THO variants or

394 with 1 μM of THO and 2 μM of Sub2 as indicated in the binding buffer (20 mM HEPES

395 pH 7.0, 80 mM NaCl, and 0.5 mM TCEP) at room temperature for 10 min. The reaction

396 mixtures were then added to ~15 μL glutathione resin in an Eppendorf tube and binding

397 was allowed to proceed at room temperature for 30 min with gentle tapping to mix every

398 3-5 min. Beads were washed twice with 500 μL washing buffer containing 20 mM HEPES

399 pH 7.0, 80 mM NaCl (for Figure 1B) or 50 mM NaCl (for Figure 1C and Figure 4D), and

400 0.5 mM TCEP. Bound proteins were eluted with washing buffer supplemented with 20

401 mM glutathione and analyzed using Coomassie-stained SDS-PAGE gels. The

402 experiments were repeated three times independently.

403

404 **Crosslinking mass spectrometry analysis**

405 For EDC crosslinking, 1 μM of THO-FL and 1 μM of GST-Gbp2 were incubated at 10 mM

406 HEPES pH 7.0, 105 mM NaCl, 0.5 mM TCEP in the presence of 20 mM EDC and 0.5 mM

407 sulfo-DHS at room temperature for 40 min. The reaction was quenched at room

408 temperature for 20 min by adding Tris pH 8.0 and β-mercaptoethanol to a final

409 concentration of 50 mM and 20 mM, respectively. DSS crosslinking was performed in the

410 same conditions except that 0.5 mM DSS was used and only Tris pH 8.0 was used for

411 quenching the reaction.

412

413 The DSS and EDC cross-linked samples were directly processed for in-solution Trypsin

414 and Lys-C digestion. The samples were reduced with 5 mM DTT and 5 mM TCEP in 8M

415 urea buffer (50 mM Ammonium bicarbonate), and were then incubated with 30mM

416 iodoacetamide at room temperature for 30 minutes in the dark. 30-45 μg of the purified

417 complex was digested with Trypsin and Lys-C using a 1:100 ratio for each protease upon

418 diluting the sample to 1 M urea. The proteolysis reaction occurred overnight (12-16 hours)

419 at 37°C. After overnight digestion with trypsin, the complex was digested with an

420 additional 1:100 ratio of trypsin at 37°C for 2 hours. The resulting mixture was acidified,

421 desalted by using a C18 cartridge (Sep-Pak, Waters).

422

423 1-2 μg of the trypsin digested crosslinked complex was analyzed with a nano-LC 1200

424 that is coupled online with a Q Exactive™ HF-X Hybrid Quadrupole Orbitrap™ mass

425 spectrometer (Thermo Fisher) (Xiang, Nambulli, et al., 2020; Xiang, Shen, et al., 2020).

426 Briefly, desalted peptides were loaded onto a Picochip column (C18, 1.9 μm particle size,

427 200 Å pore size, 50 μm × 25 cm; New Objective) and eluted using a 60-min liquid

428 chromatography gradient (5% B–8% B, 0–2 min; 8% B–40% B, 2-50 min; 40%B–100%

429 B, 50-60 min; mobile phase A consisted of 0.1% formic acid (FA), and mobile phase B

20

430   consisted of 0.1% FA in 80% acetonitrile). The flow rate was ~350 nl/min. The QE HF-X

431   instrument was operated in the data-dependent mode, where the top 6 most abundant

432   ions (mass range 350 – 2,000, charge state 4 - 8) were fragmented by high-energy

433   collisional dissociation (HCD). The target resolution was 120,000 for MS and 15,000 for

434   tandem MS (MS/MS) analyses. The quadrupole isolation window was 1.6 Th, and the

435   maximum injection time for MS/MS was set at 300 ms.

436

437   After the MS analysis, the data was searched by pLink2 for the identification of cross-

438   linked peptides (Chen et al., 2019). The mass accuracy was specified as 10 and 20 p.p.m.

439   for MS and MS/MS, respectively. Other search parameters included cysteine

440   carbamidomethylation as a fixed modification and methionine oxidation as a variable

441   modification. A maximum of three trypsin missed-cleavage sites were allowed. The

442   crosslink spectra were then manually checked to remove potential false-positive

443   identifications as previously described (Xiang, Shen, et al., 2020). The crosslinking data

444   was analyzed by CX-Circos (http://cx-circos.net). The distance distribution of the

445   crosslinks onto the THO structure was performed with Xlink Analyzer (Kosinski et al.,

446   2015).

447

455

Data availability

457 The cryo-EM density maps will be deposited in the Electron Microscopy Data Bank. The

458 coordinates of the THO•Sub2 complex will be deposited in the Protein Data Bank.

459

Competing interests

461 The authors declare no competing interests.

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221. doi:10.1107/S0907444909052925

Ahn, S. H., Kim, M., & Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell, 13*(1), 67-76. doi:10.1016/s1097-2765(03)00492-1

Cao, L., Chen, F., Yang, X., Xu, W., Xie, J., & Yu, L. (2014). Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. *BMC Evol Biol, 14*, 10. doi:10.1186/1471-2148-14-10

Carmody, S. R., & Wente, S. R. (2009). mRNA nuclear export at a glance. *J Cell Sci, 122*(Pt 12), 1933-1937. doi:10.1242/jcs.041236

Chait, B. T., Cadene, M., Olinares, P. D., Rout, M. P., & Shi, Y. (2016). Revealing Higher Order Protein Structure Using Mass Spectrometry. *Journal of the American Society for Mass Spectrometry, 27*(6), 952-965. doi:10.1007/s13361-016-1385-1

Chavez, S., Beilharz, T., Rondon, A. G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J. Q., 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. *EMBO J, 19*(21), 5824-5834. doi:10.1093/emboj/19.21.5824

Chen, Z. L., Meng, J. M., Cao, Y., Yin, J. L., Fang, R. Q., Fan, S. B., Liu, C., Zeng, W. F., Ding, Y. H., Tan, D., Wu, L., Zhou, W. J., Chi, H., Sun, R. X., Dong, M. Q., & He, S. M. (2019). A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun, 10*(1), 3404. doi:10.1038/s41467-019-11337-z

Cheng, H., Dufu, K., Lee, C. S., Hsu, J. L., Dias, A., & Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell, 127*(7), 1389-1400. doi:10.1016/j.cell.2006.10.044

Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., & Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev, 15*(24), 3319-3329. doi:10.1101/gad.935901

Dias, A. P., Dufu, K., Lei, H., & Reed, R. (2010). A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat Commun, 1*, 97. doi:10.1038/ncomms1103

Dominguez-Sanchez, M. S., Barroso, S., Gomez-Gonzalez, B., Luna, R., & Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet, 7*(12), e1002386. doi:10.1371/journal.pgen.1002386

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501. doi:10.1107/S0907444910007493

Folkmann, A. W., Noble, K. N., Cole, C. N., & Wente, S. R. (2011). Dbp5, Gle1-IP6 and Nup159: a working model for mRNP export. *Nucleus, 2*(6), 540-548. doi:10.4161/nucl.2.6.17881

Hacker, S., & Krebber, H. (2004). Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *J Biol Chem, 279*(7), 5049-5052. doi:10.1074/jbc.C300522200

508 Hackmann, A., Wu, H., Schneider, U. M., Meyer, K., Jung, K., & Krebber, H. (2014). Quality
509   control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nat Commun,*
510   *5*, 3123. doi:10.1038/ncomms4123
511 Hsin, J. P., & Manley, J. L. (2012). The RNA polymerase II CTD coordinates transcription and
512   RNA processing. *Genes Dev, 26*(19), 2119-2137. doi:10.1101/gad.200303.112
513 Hu, D., Mayeda, A., Trembley, J. H., Lahti, J. M., & Kidd, V. J. (2003). CDK11 complexes
514   promote pre-mRNA splicing. *J Biol Chem, 278*(10), 8623-8629.
515   doi:10.1074/jbc.M210057200
516 Huang, Y., Gattoni, R., Stevenin, J., & Steitz, J. A. (2003). SR splicing factors serve as adapter
517   proteins for TAP-dependent mRNA export. *Mol Cell, 11*(3), 837-843. doi:10.1016/s1097-
518   2765(03)00089-3
519 Huang, Y., & Steitz, J. A. (2005). SRprises along a messenger's journey. *Mol Cell, 17*(5), 613-615.
520   doi:10.1016/j.molcel.2005.02.020
521 Hurt, E., Luo, M. J., Rother, S., Reed, R., & Strasser, K. (2004). Cotranscriptional recruitment of
522   the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX
523   complex. *Proc Natl Acad Sci U S A, 101*(7), 1858-1862. doi:10.1073/pnas.0308663100
524 Kim, S. J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W. Z., Raveh, B., Herricks, T.,
525   Slaughter, B. D., Hogan, J. A., Upla, P., Chemmama, I. E., Pellarin, R., Echeverria, I.,
526   Shivaraju, M., Chaudhury, A. S., Wang, J. J., Williams, R., Unruh, J. R., Greenberg, C. H.,
527   Jacobs, E. Y., Yu, Z. H., de la Cruz, M. J., Mironska, R., Stokes, D. L., Aitchison, J. D.,
528   Jarrold, M. F., Gerton, J. L., Ludtke, S. J., Akey, C. W., Chait, B. T., Sali, A., & Rout, M.
529   P. (2018). Integrative structure and functional anatomy of a nuclear pore complex. *Nature,*
530   *555*(7697), 475-+. doi:10.1038/nature26003
531 Kosinski, J., von Appen, A., Ori, A., Karius, K., Muller, C. W., & Beck, M. (2015). Xlink Analyzer:
532   software for analysis and visualization of cross-linking data in the context of three-
533   dimensional structures. *J Struct Biol, 189*(3), 177-183. doi:10.1016/j.jsb.2015.01.014
534 Leitner, A., Faini, M., Stengel, F., & Aebersold, R. (2016). Crosslinking and Mass Spectrometry:
535   An Integrated Technology to Understand the Structure and Function of Molecular
536   Machines. *Trends in Biochemical Sciences, 41*(1), 20-32. doi:10.1016/j.tibs.2015.10.008
537 Luo, M. L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., & Reed, R. (2001).
538   Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and
539   Aly. *Nature, 413*(6856), 644-647. doi:10.1038/35098106
540 Martinez-Lumbreras, S., Taverniti, V., Zorrilla, S., Seraphin, B., & Perez-Canadillas, J. M. (2016).
541   Gbp2 interacts with THO/TREX through a novel type of RRM domain. *Nucleic Acids Res,*
542   *44*(1), 437-448. doi:10.1093/nar/gkv1303
543 Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., & Reed, R. (2005). Recruitment of the
544   human TREX complex to mRNA during splicing. *Genes Dev, 19*(13), 1512-1517.
545   doi:10.1101/gad.1302205
546 Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski,
547   A., Nowotny, M., Conti, E., & Filipowicz, W. (2014). Structural and biochemical insights
548   to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol*
549   *Cell, 54*(5), 751-765. doi:10.1016/j.molcel.2014.03.036
550 Meinel, D. M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., Cramer,
551   P., Soding, J., Holstege, F. C., & Strasser, K. (2013). Recruitment of TREX to the
552   transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II.
553   *PLoS Genet, 9*(11), e1003914. doi:10.1371/journal.pgen.1003914

554 Metkar, M., Ozadam, H., Lajoie, B. R., Imakaev, M., Mirny, L. A., Dekker, J., & Moore, M. J.
555   (2018). Higher-Order Organization Principles of Pre-translational mRNPs. *Mol Cell, 72*(4),
556   715-726 e713. doi:10.1016/j.molcel.2018.09.012
557 Montpetit, B., Thomsen, N. D., Helmke, K. J., Seeliger, M. A., Berger, J. M., & Weis, K. (2011).
558   A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in
559   mRNA export. *Nature, 472*(7342), 238-242. doi:10.1038/nature09862
560 Muller-McNicoll, M., Botti, V., de Jesus Domingues, A. M., Brandl, H., Schwich, O. D., Steiner,
561   M. C., Curk, T., Poser, I., Zarnack, K., & Neugebauer, K. M. (2016). SR proteins are NXF1
562   adaptors that link alternative RNA processing to mRNA export. *Genes Dev, 30*(5), 553-
563   566. doi:10.1101/gad.276477.115
564 Pak, V., Eifler, T. T., Jager, S., Krogan, N. J., Fujinaga, K., & Peterlin, B. M. (2015). CDK11 in
565   TREX/THOC Regulates HIV mRNA 3' End Processing. *Cell Host Microbe, 18*(5), 560-
566   570. doi:10.1016/j.chom.2015.10.012
567 Parua, P. K., & Fisher, R. P. (2020). Dissecting the Pol II transcription cycle and derailing cancer
568   with CDK inhibitors. *Nat Chem Biol, 16*(7), 716-724. doi:10.1038/s41589-020-0563-4
569 Pena, A., Gewartowski, K., Mroczek, S., Cuellar, J., Szykowska, A., Prokop, A., Czarnocki-
570   Cieciura, M., Piwowarski, J., Tous, C., Aguilera, A., Carrascosa, J. L., Valpuesta, J. M., &
571   Dziembowski, A. (2012). Architecture and nucleic acids recognition mechanism of the
572   THO complex, an mRNP assembly factor. *EMBO J, 31*(6), 1605-1616.
573   doi:10.1038/emboj.2012.10
574 Perez-Calero, C., Bayona-Feliu, A., Xue, X., Barroso, S. I., Munoz, S., Gonzalez-Basallote, V. M.,
575   Sung, P., & Aguilera, A. (2020). UAP56/DDX39B is a major cotranscriptional RNA-DNA
576   helicase that unwinds harmful R loops genome-wide. *Genes Dev, 34*(13-14), 898-912.
577   doi:10.1101/gad.336024.119
578 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., &
579   Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and
580   analysis. *J Comput Chem, 25*(13), 1605-1612. doi:10.1002/jcc.20084
581 Reed, R., & Cheng, H. (2005). TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol,
582   17*(3), 269-273. doi:10.1016/j.ceb.2005.04.011
583 Ren, Y., Schmiege, P., & Blobel, G. (2017). Structural and biochemical analyses of the DEAD-
584   box ATPase Sub2 in association with THO or Yra1. *Elife, 6*. doi:10.7554/eLife.20070
585 Rougemaille, M., Dieppois, G., Kisseleva-Romanova, E., Gudipati, R. K., Lemoine, S., Blugeon,
586   C., Boulay, J., Jensen, T. H., Stutz, F., Devaux, F., & Libri, D. (2008). THO/Sub2p
587   functions to coordinate 3'-end processing with gene-nuclear pore association. *Cell, 135*(2),
588   308-321. doi:10.1016/j.cell.2008.08.005
589 Saguez, C., Schmid, M., Olesen, J. R., Ghazy, M. A., Qu, X., Poulsen, M. B., Nasser, T., Moore,
590   C., & Jensen, T. H. (2008). Nuclear mRNA surveillance in THO/sub2 mutants is triggered
591   by inefficient polyadenylation. *Mol Cell, 31*(1), 91-103. doi:10.1016/j.molcel.2008.04.030
592 Schuller, S. K., Schuller, J. M., Prabu, J. R., Baumgartner, M., Bonneau, F., Basquin, J., & Conti,
593   E. (2020). Structural insights into the nucleic acid remodeling mechanisms of the yeast
594   THO-Sub2 complex. *Elife, 9*. doi:10.7554/eLife.61467
595 Schutz, P., Bumann, M., Oberholzer, A. E., Bieniossek, C., Trachsel, H., Altmann, M., & Baumann,
596   U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase
597   controlled by protein-protein interactions. *Proc Natl Acad Sci U S A, 105*(28), 9564-9569.
598   doi:10.1073/pnas.0800418105

599   Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S. J., Williams, R., Schneidman-
600   Duhovny, D., Sali, A., Rout, M. P., & Chait, B. T. (2014). Structural characterization by
601   cross-linking reveals the detailed architecture of a coatomer-related heptameric module
602   from the nuclear pore complex. *Mol Cell Proteomics, 13*(11), 2927-2943.
603   doi:10.1074/mcp.M114.041673
604   Singh, G., Kucukural, A., Cenik, C., Leszyk, J. D., Shaffer, S. A., Weng, Z., & Moore, M. J. (2012).
605   The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR
606   Protein Nexus. *Cell, 151*(4), 915-916. doi:10.1016/j.cell.2012.10.032
607   Sterner, D. E., Lee, J. M., Hardin, S. E., & Greenleaf, A. L. (1995). The yeast carboxyl-terminal
608   repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. *Mol
609   Cell Biol, 15*(10), 5716-5724. doi:10.1128/mcb.15.10.5716
610   Strasser, K., & Hurt, E. (2001). Splicing factor Sub2p is required for nuclear mRNA export through
611   its interaction with Yra1p. *Nature, 413*(6856), 648-652. doi:10.1038/35098113
612   Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon,
613   A. G., Aguilera, A., Struhl, K., Reed, R., & Hurt, E. (2002). TREX is a conserved complex
614   coupling transcription with messenger RNA export. *Nature, 417*(6886), 304-308.
615   doi:10.1038/nature746
616   Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007).
617   EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol,
618   157*(1), 38-46. doi:10.1016/j.jsb.2006.05.009
619   Valente, S. T., Gilmartin, G. M., Venkatarama, K., Arriagada, G., & Goff, S. P. (2009). HIV-1
620   mRNA 3' end processing is distinctively regulated by eIF3f, CDK11, and splice factor 9G8.
621   *Mol Cell, 36*(2), 279-289. doi:10.1016/j.molcel.2009.10.004
622   Viphakone, N., Sudbery, I., Griffith, L., Heath, C. G., Sims, D., & Wilson, S. A. (2019). Co-
623   transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. *Mol
624   Cell, 75*(2), 310-323 e318. doi:10.1016/j.molcel.2019.04.034
625   Wang, K., Wang, L., Wang, J., Chen, S., Shi, M., & Cheng, H. (2018). Intronless mRNAs transit
626   through nuclear speckles to gain export competence. *J Cell Biol, 217*(11), 3912-3929.
627   doi:10.1083/jcb.201801184
628   Wood, A., & Shilatifard, A. (2006). Bur1/Bur2 and the Ctk complex in yeast: the split personality
629   of mammalian P-TEFb. *Cell Cycle, 5*(10), 1066-1068. doi:10.4161/cc.5.10.2769
630   Xiang, Y., Nambulli, S., Xiao, Z., Liu, H., Sang, Z., Duprex, W. P., Schneidman-Duhovny, D.,
631   Zhang, C., & Shi, Y. (2020). Versatile and multivalent nanobodies efficiently neutralize
632   SARS-CoV-2. *Science*. doi:10.1126/science.abe4747
633   Xiang, Y., Shen, Z., & Shi, Y. (2020). Chemical Cross-Linking and Mass Spectrometric Analysis
634   of the Endogenous Yeast Exosome Complexes. *Methods Mol Biol, 2062*, 383-400.
635   doi:10.1007/978-1-4939-9822-7_18
636   Xie, Y., Lord, C. L., Clarke, B. P., Ivey, A. L., S., Hill P., W., McDonald H., Wente, S. R., & Ren,
637   Y. (2020). Structure and activation mechanism of the yeast RNA Pol II CTD kinase CTDK-
638   1 complex. *Proc Natl Acad Sci U S A (manuscript in press)*.
639   Xie, Y., & Ren, Y. (2019). Mechanisms of nuclear mRNA export: A structural perspective. *Traffic,
640   20*(11), 829-840. doi:10.1111/tra.12691
641   Yu, C., & Huang, L. (2018). Cross-Linking Mass Spectrometry: An Emerging Technology for
642   Interactomics and Structural Biology. *Analytical Chemistry, 90*(1), 144-165.
643   doi:10.1021/acs.analchem.7b04431

644   Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol, 193*(1), 1-12.
645   doi:10.1016/j.jsb.2015.11.003
646   Zhang, Y., French, S. L., Beyer, A. L., & Schneider, D. A. (2016). The Transcription Factor THO
647   Promotes Transcription Initiation and Elongation by RNA Polymerase I. *J Biol Chem,*
648   *291*(6), 3010-3018. doi:10.1074/jbc.M115.673442
649   Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017).
650   MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron
651   microscopy. *Nat Methods, 14*(4), 331-332. doi:10.1038/nmeth.4193
652   Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E., & Reed, R. (2000). The protein Aly links
653   pre-messenger-RNA splicing to nuclear export in metazoans. *Nature, 407*(6802), 401-405.
654   doi:10.1038/35030160
655   Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., & Scheres, S.
656   H. (2018). New tools for automated high-resolution cryo-EM structure determination in
657   RELION-3. *Elife, 7*. doi:10.7554/eLife.42166
658
659
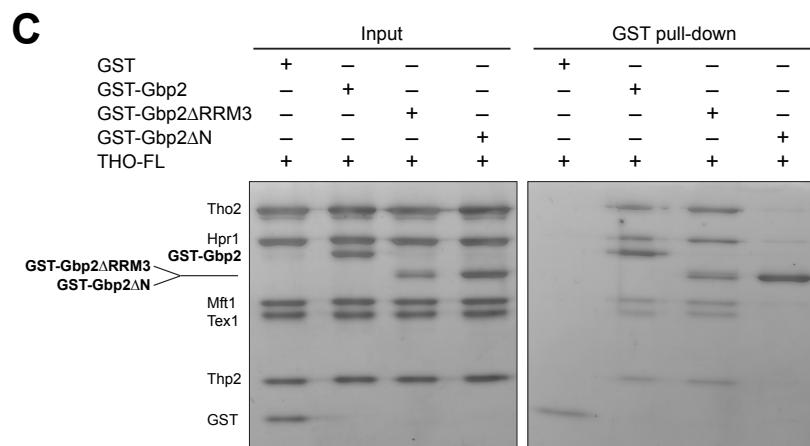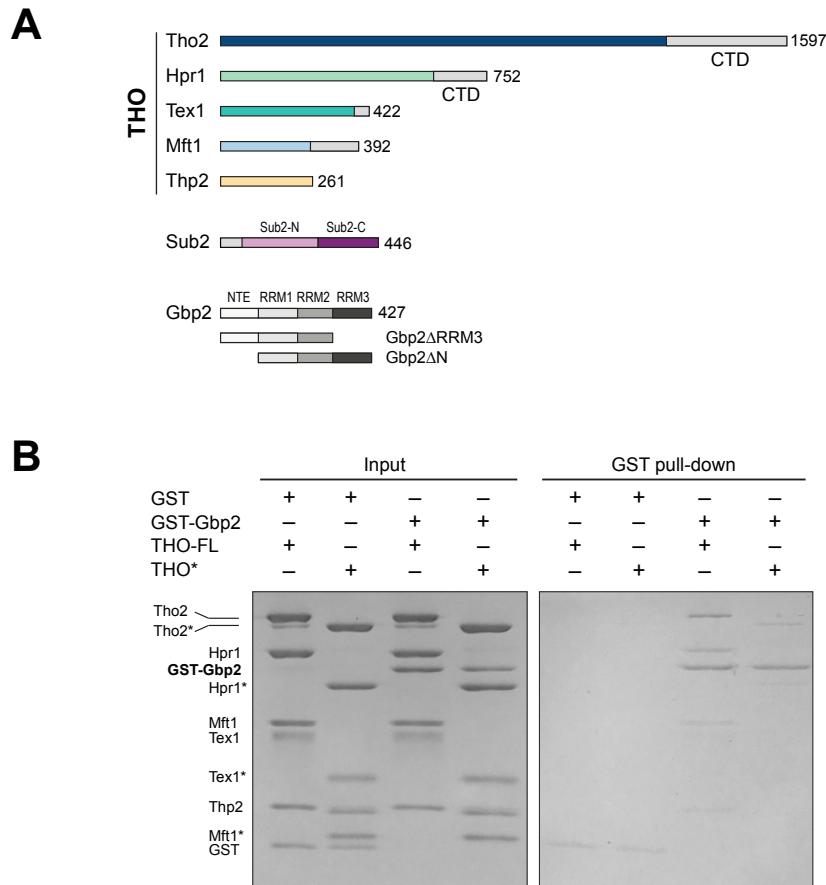
660 **Figure 1. The THO complex directly interacts with the SR-like protein Gbp2. A)**

661 Domain organization of the THO complex, Sub2, and Gbp2. Within THO, the protein

662 regions that are included in the core THO* complex are colored (Tho2 in blue, Hpr1 in

663 green, Tex1 in cyan, Mft1 in light blue, and Thp2 in yellow). Sub2 is colored in pink (Sub2-

664 N) and purple (Sub2-C). Gbp2 contains an N-terminal extension (NTE) followed by three

665 RRM domains. **B)** THO directly interacts with Gbp2. *In vitro* GST-pull down assays with

666 purified recombinant proteins show that both THO-FL and THO* bind to Gbp2 with the

667 former exhibiting stronger interaction. **C)** THO binding to Gpb2 requires the N-terminal

668 extension of Gbp2.
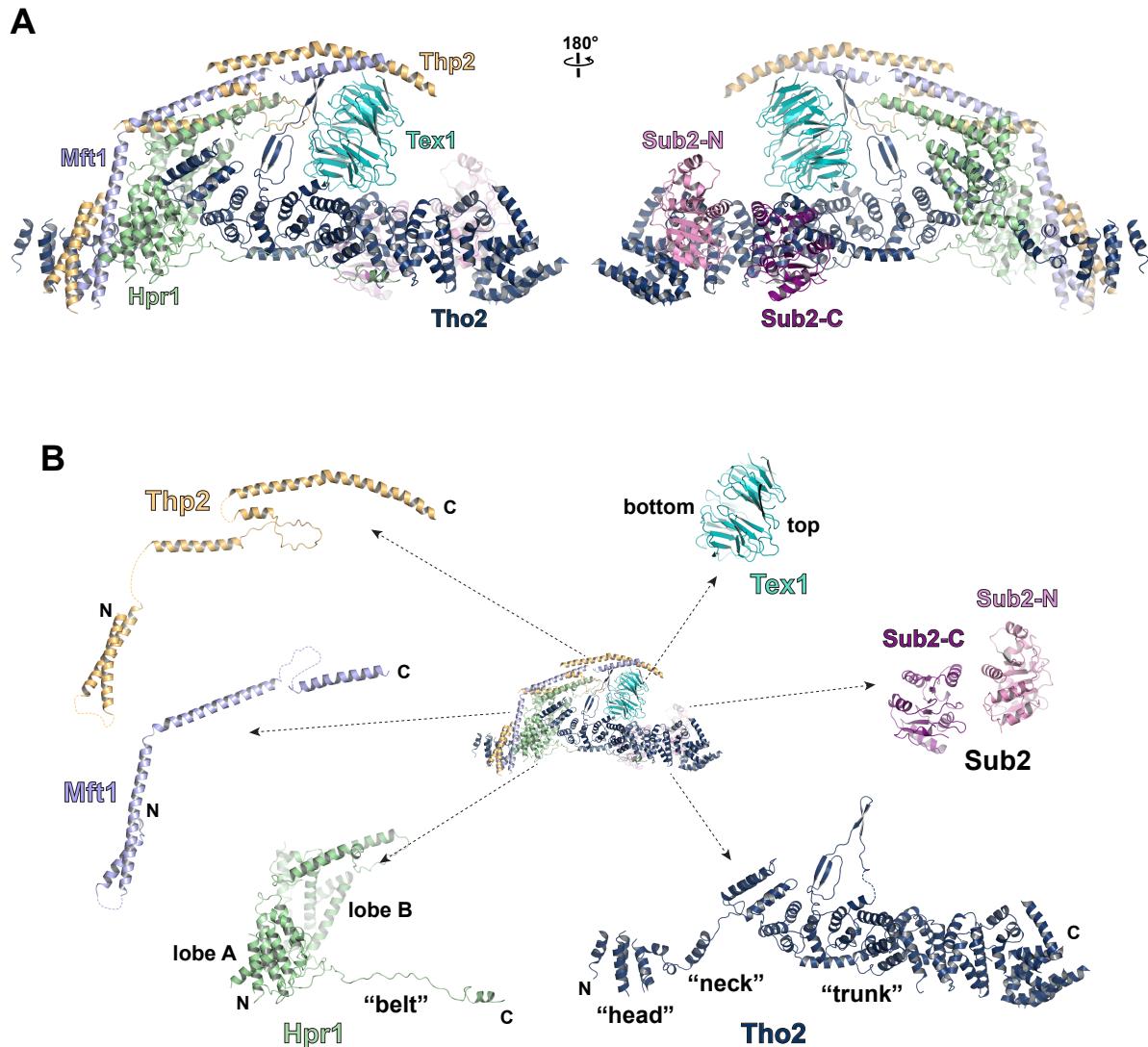
Figure 2. Cryo-EM structure of the THO*•Sub2 complex at 3.70 Å resolution. A) Overall architecture of the THO*•Sub2 complex in front and back views. **B)** Dissected view of the THO*•Sub2 complex subunits. The largest THO subunit, Tho2, contains a "head", a "neck", and an α-solenoid "trunk". Hpr1 contains lobe A, lobe B, followed by an extended "belt".
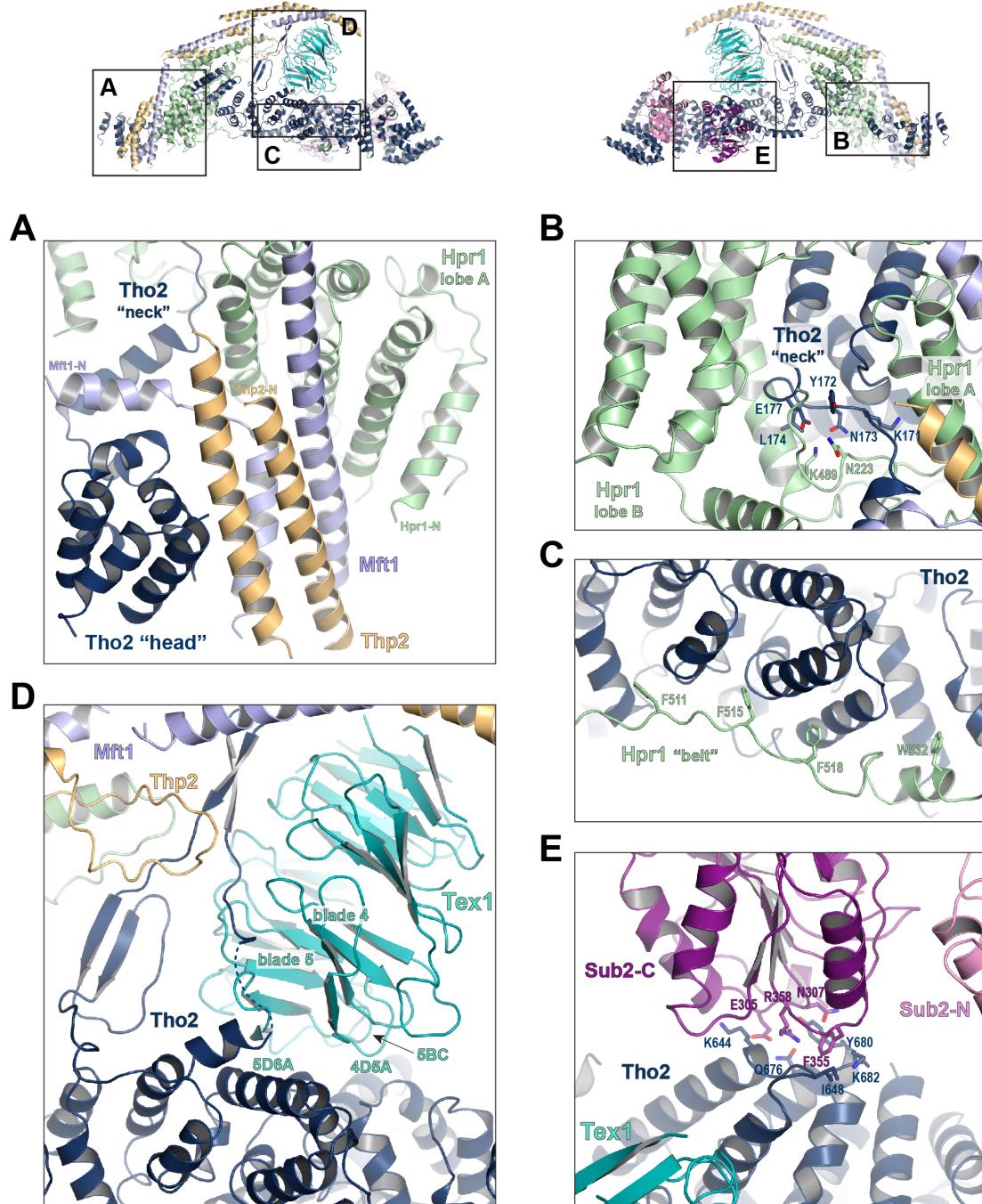
Figure 3. Key interactions in the THO*•Sub2 complex. **A)** A highly intimate interface involving the Tho2 "head". **B)** The Tho2 "neck" is embraced by the two lobes of Hpr1. **C)** The Hpr1 exhibits an extended "belt" lining the Tho2 "trunk". **D)** The Tex1 beta propeller sits at the center of the Tho2 "trunk". **E)** The interface between Tho2 and Sub2-C.

31
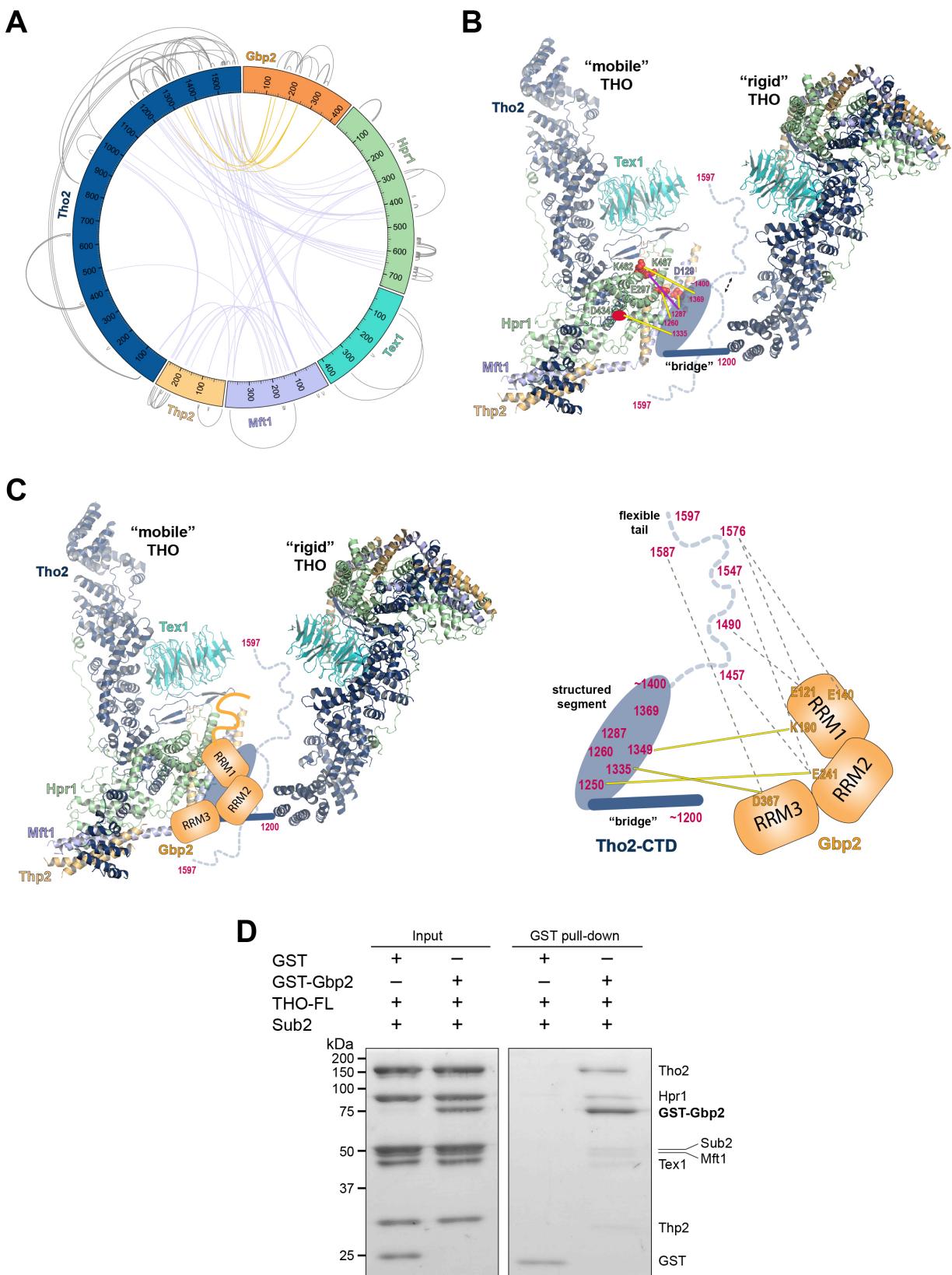
678 **Figure 4. Chemical cross-linking and mass spectrometry reveals THO-Gbp2**

679 **interactions. A)** Circular plot showing the intermolecular crosslinking sites with EDC

680 cross-linker. Each THO•Gbp2 complex subunit is represented as a colored segment with

681 the amino acid residues indicated. Inter-molecular crosslinks are mapped inside the circle

682 and the intra-molecular crosslinks are mapped outside the circle. The crosslinks between

683 Tho2 and Gbp2 are colored in orange. **B)** Schematics of the arrangement of the Tho2-

684 CTD, which contains a "bridge" connecting two THO molecules, followed by a structured

685 segment and a flexible tail (residues ~1400-1597). The EDC crosslinks between the

686 structured Tho2-CTD fragment and Hpr1 (E297, D434, and K462) as well as Mft1 (D129)

687 are indicated by yellow lines. The DSS crosslink between Tho2-CTD and Hpr1-K467 is

688 indicated by a purple line. **C)** Schematics of the THO-Gbp2 interactions (left) and the

689 identified crosslinking sites between Tho2-CTD and Gbp2 RRM domains. **D)** *In vitro* GST-

690 pull downs show that Gbp2 binds to the THO•Sub2 complex.

691
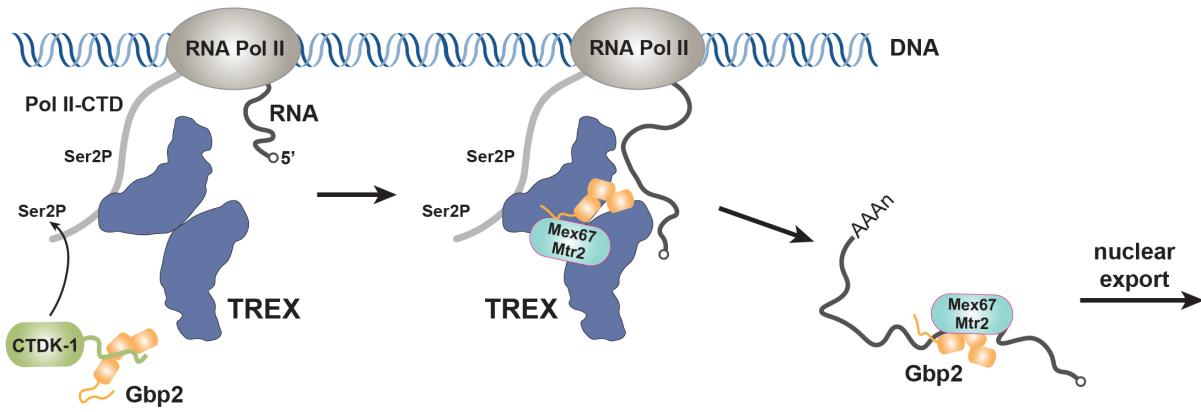
Figure 5. Working model of coordinated function of TREX and Gbp2.

During transcription, the yeast CTDK-1 complex phosphorylates Ser2 of the RNA Pol II CTD. The N-terminal extension in CTDK-1's kinase subunit Ctk1 recognizes the RRM domains of Gbp2, connecting Gbp2 to the transcription machinery. TREX travels along with the transcription machinery and recognizes multiple domains of Gbp2, possibly facilitating its loading onto the maturing mRNP. Both TREX and Gbp2 are involved in subsequent loading of the export receptor Mex67•Mtr2 to generate export competent mRNPs.

34

A

B

Final resolution = 3.70 Å

C

D

THO•Sub2 cyro-EM structure
high contour level

THO•Sub2 cyro-EM structure
low contour level

THO•Sub2 crystal structure at 6.0 Å

35
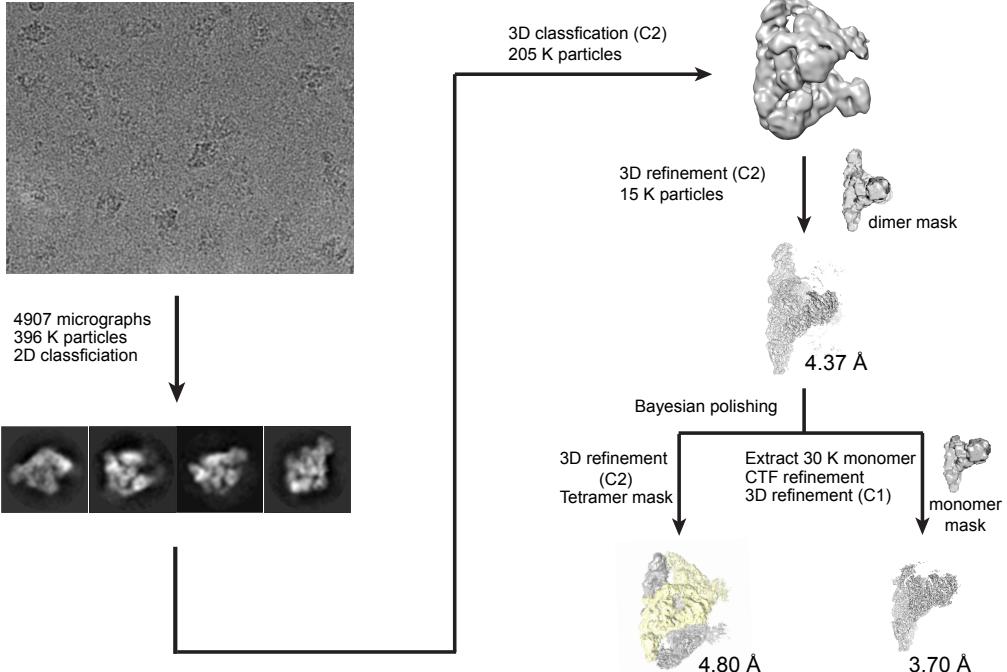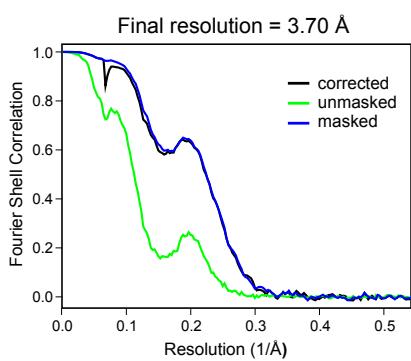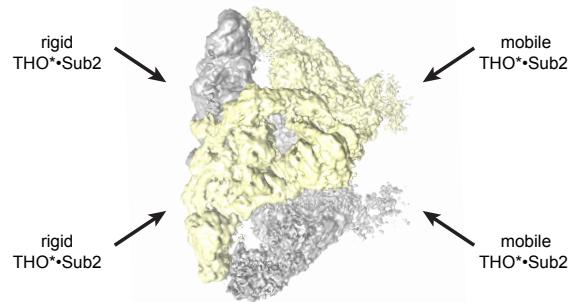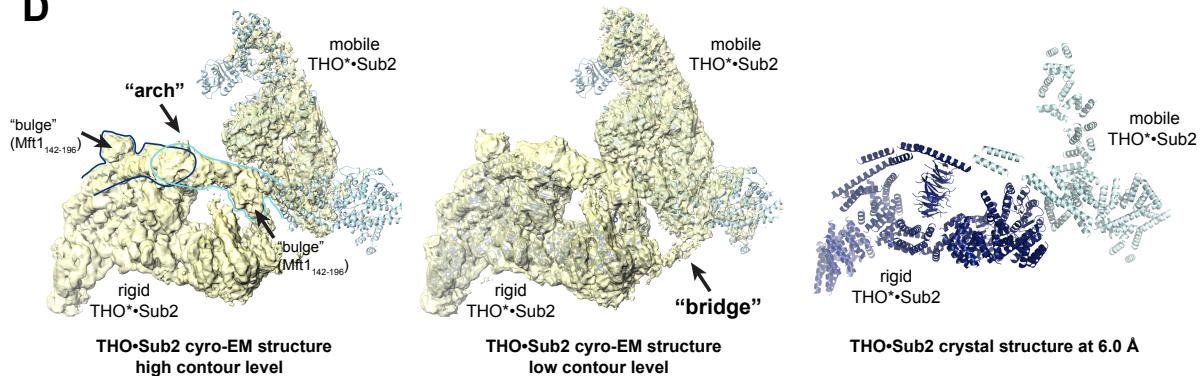
701 **Figure 2-figure supplement 1. Cryo-EM data processing. A)** Cryo-EM data processing

702 workflow. **B)** Fourier shell correlation (FSC) curves between the two half maps of the

703 THO*•Sub2 protomer. **C)** Individual particles of the cryo-EM sample contain four copies

704 of the THO*•Sub2 protomer. **D)** An "arch" and a "bridge" are observed between a rigid

705 THO*•Sub2 protomer and a mobile THO*•Sub2 protomer. Comparison of the cryo-EM

706 map and our previous crystal structure (PDB ID 5SUQ) reveals that the crystal structure

707 corresponds to a rigid THO*•Sub2 protomer and partial structure of a mobile THO*•Sub2

708 protomer.

709
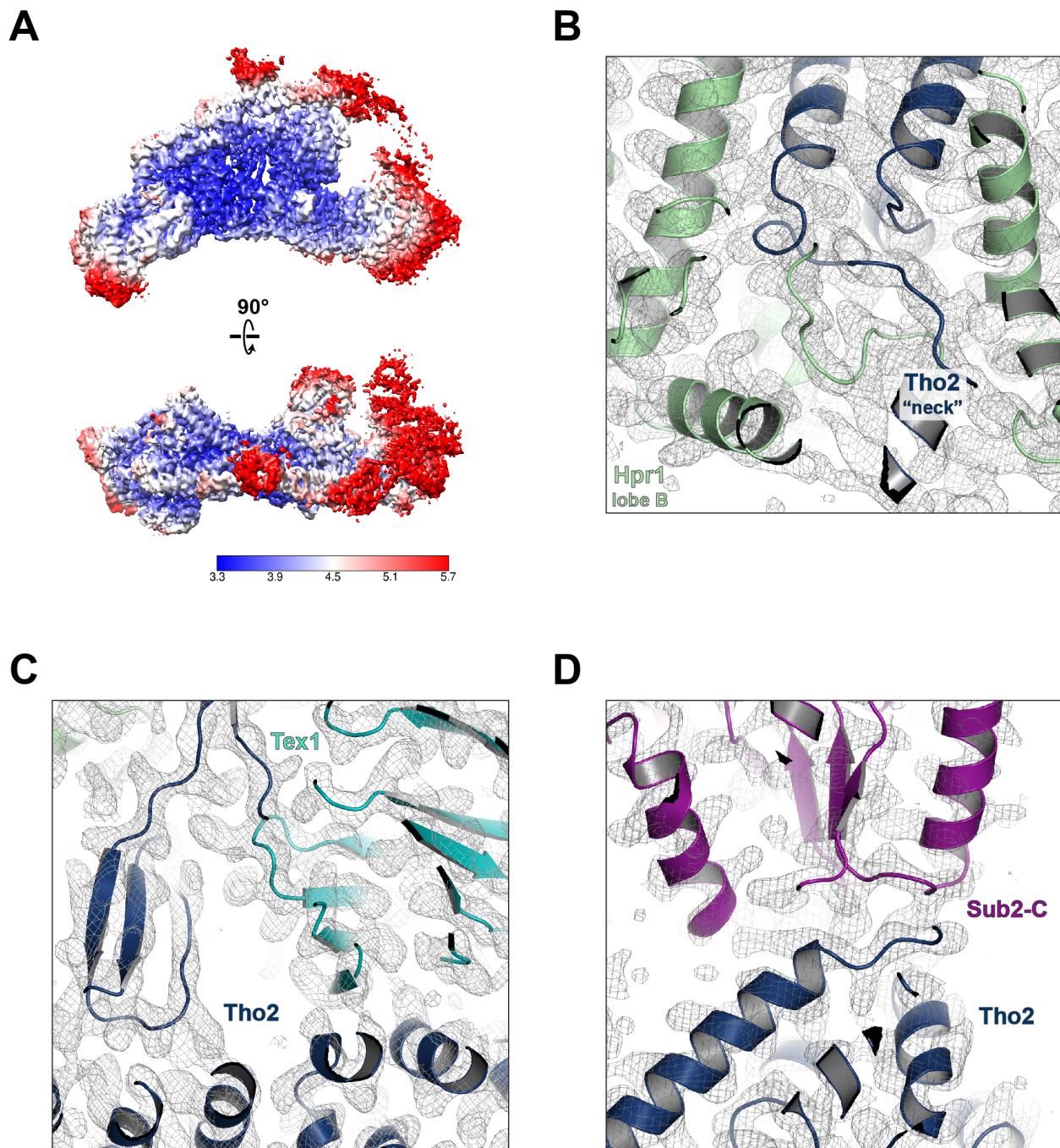
Figure 2-figure supplement 2. Cryo-EM model building. A) Local resolution of the final reconstruction calculated by Relion. Electron density maps at the Tho2 "neck" (**B**), the Tho2-Tex1 interface (**C**), and the Tho2-Sub2 interface (**D**).
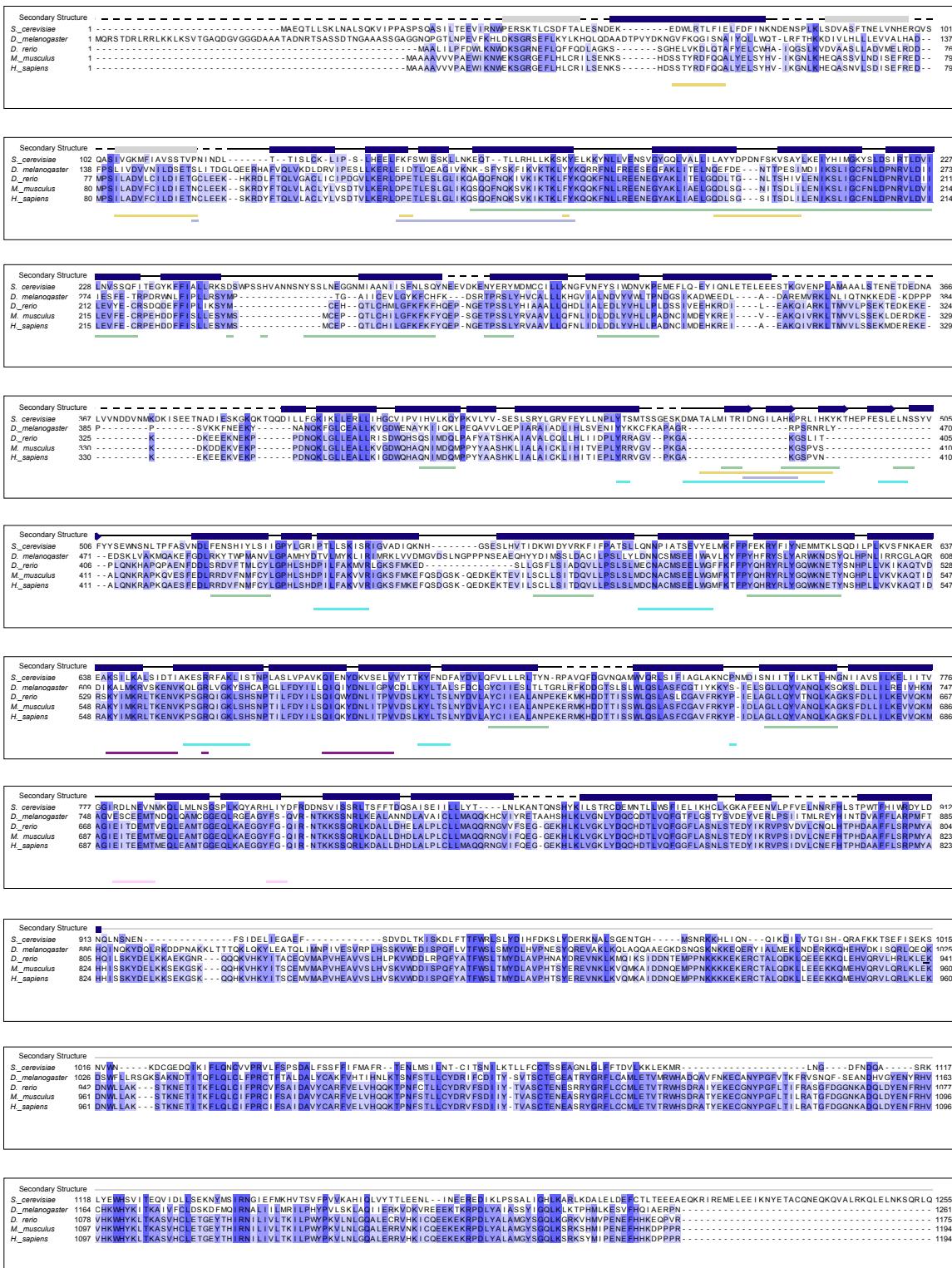
715 **Figure 2-figure supplement 3. Sequence alignment of Tho2 homologues.**

716 Tho2 sequences from *S. cerevisiae*, *D. melanogaster*, *D. rerio*, *M. musculus*, and *H.*

717 *sapiens* were aligned with ClustalW. Shading indicates the degree of conservation across

718 homologues. Tho2 secondary structural features from our cryo-EM structure are shown

719 above the sequence alignment with α-helices represented as blue bars and β-sheets

720 represented as blue arrows. The model of Tho2 contains ten poly-Ala helices at its C-

721 terminus represented by gray line. Three poly-Ala helices at the N-terminus of Tho2 are

722 represented by gray bars. Sequences that are not present in the model (loops and highly

723 flexible regions) are represented by dotted lines. The regions of Tho2 that form interaction

724 interfaces with each of the other THO Complex subunits were identified using PISA, and

725 these interaction regions are shown below the sequence alignment. The coloration for the

726 interaction regions is consistent with the color scheme throughout the manuscript with

727 Hpr1 in green, Tex1 in cyan, Mft1 in light blue, Thp2 in yellow, Sub2-N in pink and Sub2-

728 C in purple.

729

39

730
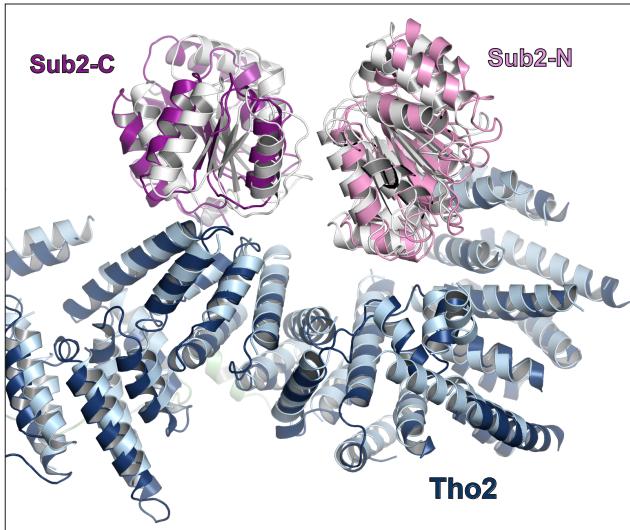
Figure 3-figure supplement 1. Comparison of the cryo-EM structure and our previous crystal structure of THO•Sub2 at the THO-Sub2 interface. Cryo-EM structure is colored in blue (Tho2), pink (Sub2-N), and purple (Sub2-C). Crystal structure is colored in light blue (Tho2) and gray (Sub2-N and Sub-C).

735

41
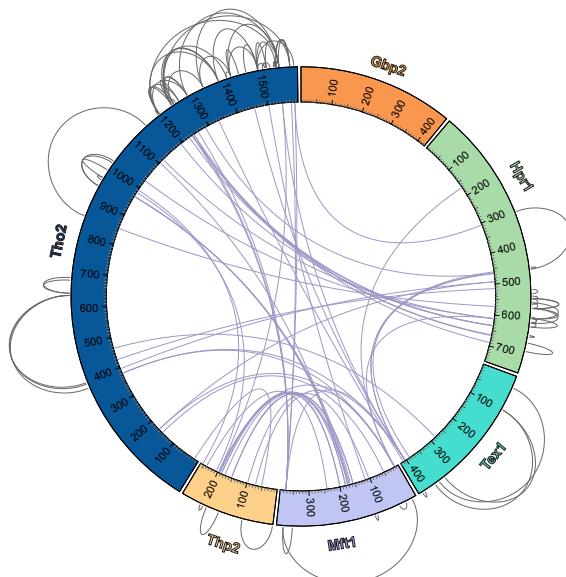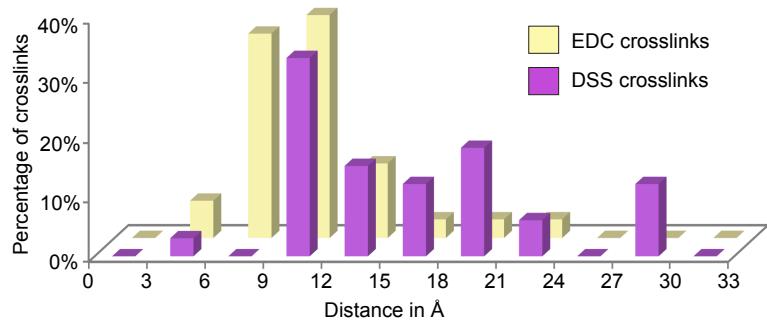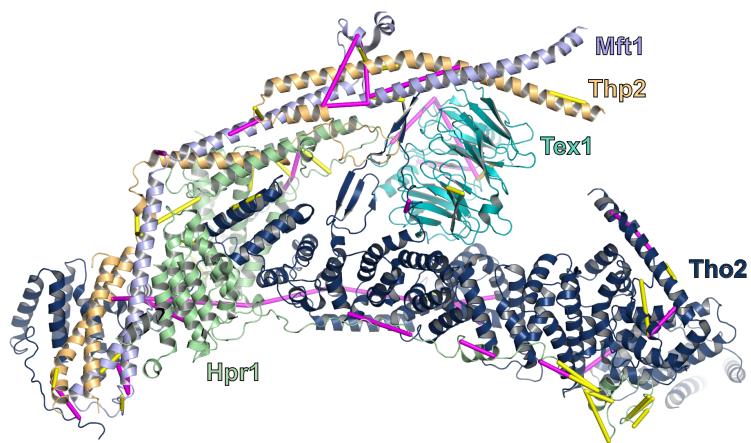
736 **Figure 4-figure supplement 1. Analyses of the XL-MS data. A)** Circular plot showing

737 the intermolecular crosslinking sites with DSS cross-linker. **B)** Distance distribution of the

738 crosslinks. We mapped the $C\alpha$-$C\alpha$ distances between cross-linked residues onto the

739 dimeric THO structure (PDB ID 7AQO). 91% of the EDC crosslinks and 100% of the DSS

740 crosslinks that can be mapped to the structure fall within the expected threshold of 17 Å

741 and 30 Å. **C)** EDC (yellow) and DSS (purple) crosslinks are mapped on the THO structure

742 (PDB ID 7APX).
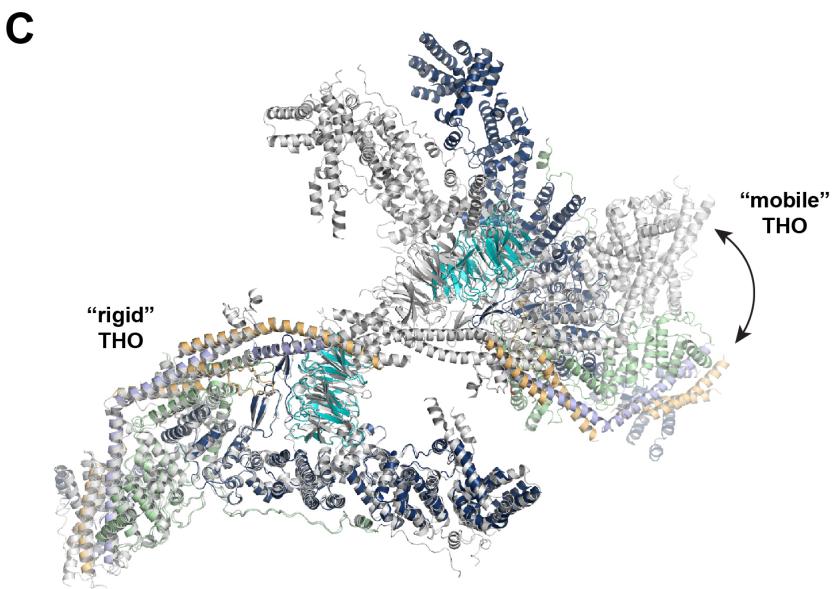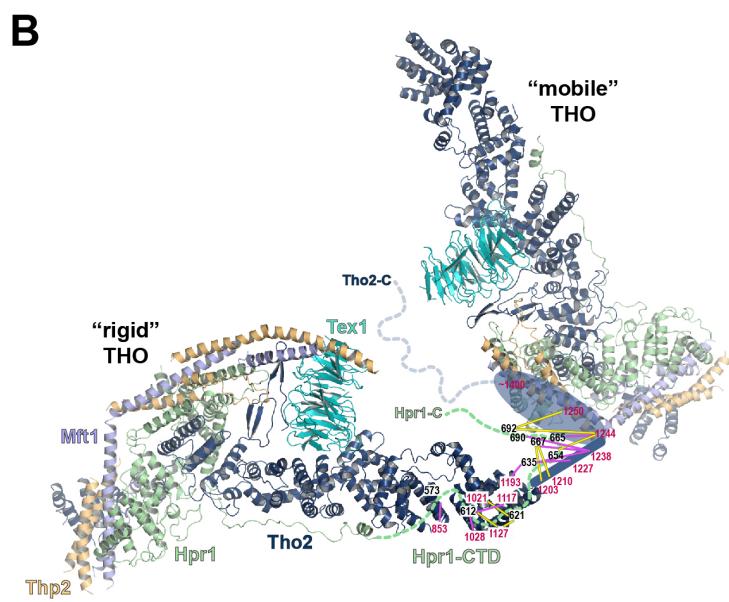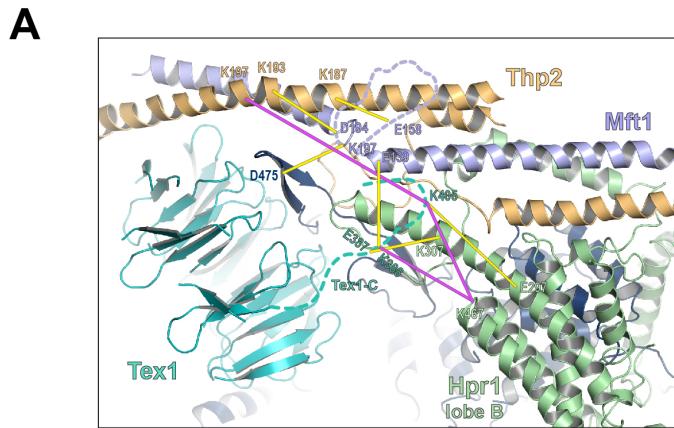
743 **Figure 4-figure supplement 2. XL-MS data indicate the arrangement of the C-termini**

744 **of Tex1 and Hpr1. A)** The C-terminal tail of Tex1 (residues 367-422) is localized near

745 Hpr1 lobe B. The crosslinking sites are indicated by yellow lines. **B)** The Hpr1-CTD binds

746 to the C-terminus of the Tho2 "trunk" and is localized close to the neighboring THO

747 molecule. **C)** Structural alignment of the dimeric THO assembly between our cryo-EM

748 structure and the recently reported THO-Sub2 structure (PDB ID 7AQO). Sub2 is omitted

749 for clarity. The structures are aligned using one THO molecule, revealing significant

750 flexibility in the relative orientation between the two THO molecules.

Table 1. Cryo-EM data collection, refinement and validation statistics

	THO-Sub2 protomer (EMDB-xxxx) (PDB xxxx)	THO-Sub2 tetramer (EMDB-xxxx)
Data collection and processing		
Microscope/Camera	Titan Krios/Falcon 3EC	
Voltage (kV)	300	
Electron exposure (e–/Å^2)	50	
Defocus range (μm)	0.8 to 2.0	
Pixel size (Å)	0.681	
Symmetry imposed	C1	
Initial particle images (no.)	396 K	
Final particle images (no.)	30 K	15 K
Resolution at 0.143 FSC (masked, Å)	3.70	4.80
Map sharpening B factor (Å^2)	86	145
Refinement		
Model resolution at 0.5 FSC (Å)	3.87	
Model composition		
Protein residues	2378	
B factors (Å^2)		
Protein	109.6	
R.m.s. deviations		
Bond lengths (Å)	0.008	
Bond angles (°)	1.11	
Validation		
MolProbity score	2.14	
Clashscore	11.86	
Poor rotamers (%)	0.15	
Ramachandran plot		
Favored (%)	90.1	
Allowed (%)	9.7	
Disallowed (%)	0.2	