

ම

Figures and figure supplements

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

Sandra K Schuller et al

eLife Research article



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.



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.







Figure 2—figure supplement 1 continued

correlation (0.143 – FSC) plot from the final round of refinement for the entire map (THO-Sub2) and the two independently refined maps (rigid protomer and flexible protomer). (**C**) Particle sorting and classification tree used for 3D reconstruction of the THO-Sub2 complex. The major analysis branch (black) outlines the processing of the dimeric THO-Sub2 complex. The rigid and flexible protomers of the dimeric THO-Sub2 complex were refined locally after density subtraction. The minor analysis branch (gray) outlines the processing focused on the rigid protomer of THO-Sub2. The resolution and map quality were improved using local per-particle CTF-correction (see **Figure 2—figure supplement 2**). (**D**) Local resolution analysis of dimeric THO-Sub2, the flexible protomer (**E**), and the rigid protomer (**F**) after density subtraction and local refinement. Maps show variation in local resolution as estimated by cryoSPARC.

eLife Research article



Figure 2—figure supplement 2. High-resolution THO-Sub2 density. (A) Segmented cryo-electron microscopy reconstruction of the rigid THO-Sub2 protomer used for de novo atomic modeling. Front and back views are shown; proteins and domains are colored as in *Figure 2*. (B) Local resolution analysis as estimated by cryoSPARC of the high-resolved rigid THO-Sub2 protomer. (C) 3D FSC and preferred orientation analysis of the rigid THO-Sub2 protomer density calculated using the 'Remote 3DFSC Processing Server' web interface. The red line represents the estimated global FSC of 3.39 Å \pm 1 SD (green dashed lines). The sphericity of 0.983 indicates a uniform isotropic map without preferred particle orientation bias.



Figure 2—figure supplement 3. Model quality. Left: Ramachandran plot of the main-chain φ and ψ torsional angles of the rigid protomer THO-Sub2 atomic model. Areas of favored φ and ψ combinations are defined in dark blue (see also **Supplementary file 1**). Right: Representative regions of the THO-Sub2 (same colors as in **Figure 2**) and surrounding electron density maps are shown. Subunits and residue numbers are specified. Snapshots are shown for the density of both folded regions and extended regions.



Figure 2—figure supplement 4. Comparison of our cryo-electron microscopy structure with previous structural studies. (A) Comparison with a previous negative-stain analyses of the THO complex (*Peña et al., 2012*; *Ren et al., 2017*). Arrows highlight areas with distinctive features in the 2D class-averages and our reconstruction. Overall there is a good agreement between the negative stain electron microscopy analysis and our structure. (B) Comparison with a previously published low-resolution X-ray structure (*Ren et al., 2017*), where the polypeptides for Tho2, Hpr1, Mft1, and Thp2 could not be distinguished, and are therefore all colored the same. The rectangle comprises the portion of the previous model that corresponds to a protomer. The other region of the previous model resulted from misinterpretation of the crystal lattice.



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.



Figure 3—figure supplement 1. Structural features of the THO complex. THO-Sub2 rigid protomer (front view) shown in cartoon backbone representation (colors are as in *Figure 2*). The rectangles highlight the position of the zoom-ins shown in panels C–H (identical to *Figure 3*). Interacting and evolutionary conserved residues (see alignments in *Figure 3—figure supplements 2–5* are shown as sticks and labeled. 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 1–5*), suggesting they may share conserved architectural features.

Alignmer				
S.cerevisiae S.pombe H.sapiens D.rerio	1 MS T I GA V D I L N QK T I T S E - 1 MC A S	VAASV TSK YLQS V TE SL PK PWC SV S SG P S R YV LGMQE YQEMQE	TF SK GN T SH I EDK RI F L F	IHVSSRSHSRFTSTPIT 62 PELKTRDLQGQ 23 RGHSKTREFLAH 54 RNNKSREFPAH 25 PELSTUPEOPEN 27
D.meianogaster		QILEDLKS	YL	KSHSKIREQRSH 27
S.cerevisiae S.pombe H.sapiens D.rerio	63 PN E I L S L K FHV S G S SMA Y S RMD 24 QG P I R S L GWN L S G S R L A S S S S S 55 SAK V H S V A W S C D G R R L A S G S F D 26 SAK V H S V A W S C D G K R L A S G S F D	G S L TVWF I K DA S F – – DK S G S V LVWN S D R L D FK F T T E K TA S V F L L EK D R L – – VK E K TA S V F V L EK D R L – – VK E	V EV Y I PDCC G SDK LA LGN RG YG - L NN Y RGHGD S NN Y RGHGD S	TDLSWNPTSLNQIAVVS 132 EQLVWDPTHSDRLMAVY 89 DQLCWHPSNPDLFVTAS 119 DQLCWHPTNPDLFVTAS 90
D.melanogaster	28 MSKVHSVCWNADGRYLASGSFD	K TVVVYSLERDRF LKG	NTYRGHTAS	DQLCWHRTNPDQFATAS 92
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	133 N S S E I S L L L I N EK S L TA SK L RT 90 A G K M I R F WD F R S A K P I A 120 G D K T I R I WD V R T T K C I A 91 G D K T I R I WD V R T T K C MA 93 G D K T V R I WD I R V G K C V S	L S L G S K – TK V N T C L Y D P L – E I E S N Y E N I Y A – TW S P S – TV N T K G E N I N I – CW S P D – T V S T K G E N I N I – CW S P D – V T N T K G E N I N I – A W S P D	GNWL LAA TK S EK I YI GN YCCA S S RDDML S I GQ T I AV GNK DD V V TI GQ T I AV GNK DD V V TI GK T I AV GNK ED L I TI	F DVKKDHSSVCSLNISD 203 I DA RE R RIMET 148 I DAK TH RSKAE 178 I DAK TH RP RAE 149 I D T R TNK I RVE 151
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	204 I SQEDNDVV YSLAWSNGGSH I F 149 FQQPC E TN ECCWSF SEDLF F 179 EQFK FEVNE I SWNNDNMF F 150 EQFK FEVNE I SWNNDNDF F 152 EPFS FEVNE I SWNNTNDI F	IGFKSGYLAILKAKHGIL MTTGLGTVQIMEWPSLI LTNGNGCINILSYPELI LTNGNGCINILSYPELI LTNGLGCMHISYPSL	EVCTK IKAHTGPITI KRVYDIKAHNSNCFC KPVQSINAHPSNCIC KLIQSINAHPSNCIC EHOMTIKAHPANCIC	IK MD PWG RN F I TG S I DG 275 I E F S PDN RH LA I GGADA 216 I K FD PMGK YFA TG SADA 246 I K FD P TGK YFA TG SADA 217 I F FG P TGK YFA TG SADA 219
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	276 NC YVWNMK SLCC ELI INDLN SA 217 ITSLWD PQELICERSITRMDYP 247 LV SLWDVDELVCV RCF SRLDWP 218 LV SLWNV EELVCV RCF SRLDWP 220 QV SLWDANELACLRMISRLEWP	V TTLDVCHLGK I LG IC TEI I RTLSF S YDS R Y LA SG S EI V RTLSF SHDGKMLA SA S EI V RTLSF SHDGKMLA SA S EI V RTLSF SHDERMIA LA S EI	D EMV YF YD LN SGN LI D RYVD I AD TK TGDQ DHF I DI A EV ETGDK I DHF I DI A EV ETG EK I DLI I DI A FT ETG ER \	HSKSLANYKT 340 WK I PTNG PLNK VAWH PT 288 WEVQC ES PTF TVAWH PK 318 WE I QC ES PTF TVAWH PK 289 TD I HVDA STF TVAWH PK 291
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	341 DPVLKFYPDKSWYIMSGKNDTL 289 319 290 290 292	SNH FVKN EKNLITYWKDM KHILAYAVSE RPLLAFACDDI RPLLAYACDDI QYLLAYACDE	FDN TM I EK RRKNNGC PN	GNNHNK RT SKN TD R I GK 412 SSGLK I FGL 309 SS REAG TVK L FGL 347 SN REAG TVK L FGL 318 RRRDAGNVK I YGF 318
S.cerevisiae	413 D R P S R F N S K K			422
S.pombe				
H.sapiens	348 PNDS			351
D.rerio	319 PNDS			322
D.melanogaster	319 SE			320

Alianment Tev1

Figure 3—figure supplement 2. Tho2 structure-based sequence alignment. The sequence alignment includes Tho2 orthologues from Saccharomyces cerevisiae (P53552), Schizosaccharomyces pombe (Q09779), Homo sapiens (Q8NI27), Danio rerio (F1R5B5), and Drosophila melanogaster (E2QCS8). The secondary structure elements are shown above the sequences, labeled, and colored according to the modules (module-1 to -5) as in Figure 3B Conserved residues are highlighted with color, with darker shades of slate-blue indicating high conservation. This and all other sequence alignments were generated using T-coffee (Notredame et al., 2000).

0		
Scerevisiae	1 MSHECKEDI LEVSDNEGE LOLDASKAA FACETCAATSATECDNNNN TAACDKKCSVVCIHSTCEKDELLKBE	72
Snomhe	I MSHEGEEDLLETSDNEQE IQIDA SKAAEAGE IGAA ISA HEGDNINNI IAAGDKKGSTVGTHSTGFKDFLLKFE	72
U canions	IMA SAGEDLIDTEEELVQUPAGETIPAADTAEUGEKSDKKGSTVGTHSTGFKDFLLKPE	01
n.sapiens Diraria	IMAENDV DNELLDYEDDEVE IAAGGDGAEAPAKKDVKGSYVSTHSSGFRDFLLKPE	55
D.Ierio	1MTENDVDNELLDYEEDEVDAGGAGDAGLGHSDGIISIRKEGVKGSYVSIHSSGFRDFLLKPE	62
D.meianogaster	1	51
)
S.cerevisiae	73 L SRA L DCGEEHPSEVOOHT L POSTHGTDVLCOAK SGLGK TAVEVLSTLOOLDPVPGEVA-VVVLCNARELA	143
S.pombe	62 LI RA I TOSCE EN PSEVOOVCI POSILICITOVI COAK SCMCK TAVEVI STLOOLE PVDCEVS-VI VI CHITRELA	132
H.sapiens		126
D.rerio		120
D melanogaster	05 ELKA IVDCGFEHTSEVQHECTPQA ILQMDVLCQAK SGMGKTAVTVLA ILQQLETVIGQVS-VLVMCHTRELA	122
Differanogaster	32 TERATVDCGFENTSEVQNECTFQAVL0MDTLCQARSGMGRTAVTVLATLQQLEFSDNNTCHVLVMCNTRELA	125
		-
S.cerevisiae	144 YQ I RN E YL R F SK YMPDVK TÂV F YGG T P I SK DA E L LKNKD TA PH I VVA T PG R LKA LV R EK Y I D L SHVKN FV I C	215
S.pombe	133 FQ I KNEYA RF SKYL PDV RTAV FYGG I NI KQDMEA FKDK SK S PH I VVA TPG RLNA LV REK I LKVN SVKH FV LC	204
H.sapiens	127 FQISKEYERFSKYMPNVKVAVFFGGLSIKKDEEVLKKNCPHIVVGTPGRILALARNKSLNLKHIKHFILC	196
D.rerio	134 FQ I SK EYERF SK YMP SVK VAV F FGG L S I KK D E EV LKK E – – S PHV V VG T PG R I LA L S RNK S LN L RH I KH F I L	203
D.melanogaster	124 FQ I SK EYERFSK YMPTVK VAV FFGGMA I QK DEETLK SG TPH I VVG TPG R I LA L I RNKK LN LK L LK H FV L	193
S cerevisiae		-
S nombo	216 ECDK V LEELDMR RDVQETF KATPRDKQVMMFSATLSQETRPTCKKFLQNPLETFVDDEAKLTHGLQQYYK	287
J.poinibe	205 ECDKLLESVDMRRDTQEVFRATPPQRQVMMFSATLSNETRPTCRKFMQNPLETVDDETRLTLHGLQQHVVR	276
n.sapiens D raria	197 ECDKMLEQLDMRRDVQETFRMTPHEKQVMMFSATLSKETRPVCRKFMQDPMETFVDDETKLTL	- 259
D.IEIIU D.malanagastar	204 ECDRMLEQEDMRRDVQETFRMTPHERQVMMFSATLSKETRPVCRKFMQDPMETFVDDETKLTT	266
D.meianogaster	194 ECDKMLEQLDMRRDVQETFRSTPHGKQVMMFSATLSKDTRPVCKKFMQDPMEVYVDDEAKLTT	256
		-
S.cerevisiae	288 LEEREKNRK LAQLLDDLEFNQV Í Í FVK STTRANELTK LLNA SNFPÁ Í TVHGHMK QEER I A RYKA FKDFEK RI	I 359
S.pombe	277 LEEKAKNRK INDLLDSLEFNQVV I FVK SV SRANELDRLLRECNFPSIC I HGGLPQEER I KRYKAFKDFDK RI	I 348
H.sapiens	260 HG LOO Y YVK LKDN EKN RK LFDLLDV LE FNOVV I FVK SVORCIALAOLLV EON FPA IA I H RGMPO E E R LS RYC	331
D.rerio	267 HG LOOYYVK LKDNEKNRK LFDLLDVLEFNOVV LFVK SVORCIALAOLLVEONEPA JA JH RAMPODERLARYC	338
D.melanogaster	257 HG LOOH YVN LKEN EKNKKLFELLDVLEFNOVV I FVK SVORCVALSOLLTEON FPA IG I HRGMTO FERLN RYC	328
-		
C corovicioo		-
S.Cereviside	360 CV S TDV FG RG I D I E R I N LA I N YD L TN EADQYLH RVG RAG RFG TK G LA I S FV S SK ED E EV LAK I Q E RFDVK I A	A 431
S.pombe	349 CVA TDV FG RG I D I E RVN I V I N YDMPD S PD S Y LH RVG RAG R FG TK G LA I T F S S S E E D SQ I LDK I QE R F E V N I T	Г 420
H.sapiens	332 Q F K D F Q R R I L V A T N L F G R G M D I E R V N I A F N Y D M P E D S D T Y L H R V A R A G R F G T K G L A I T F V S D E N D A K I L N D V	/ 403
D.rerio	339 Q F K D F Q R R I L V A T N L F G R G M D I E R V N I A F N Y D M P E D S D T Y L H R V A R A G R F G T K G L A I T F V S D E N D A R T L N D V	/ 410
D.melanogaster	329 Q F K D F Q K R I L V A T N L F G R G M D I E R V N I V F N Y D M P E D S D T Y L H R V A R A G R F G T K G L A I T F V S D E N D A K I L N E V	/ 400
S.cerevisiae		446
S.pombe		434
H.sapiens		434
Drerio		420
D melanonaster		433
Differenciariogaster	401 QUKFUVNISELFEETULSIYIEGK-	424

Alignment Sub2

Figure 3—figure supplement 3. Hpr1 structure-based sequence alignment. The sequence alignment includes the structured elements of Hpr1 orthologues from *Saccharomyces cerevisiae* (P17629), *Schizosaccharomyces pombe* (Q9URT2), *Homo sapiens* (Q96FV9), *Danio rerio* (Q7SYB2), and *Drosophila melanogaster* (Q9VNI8). The secondary structure elements are shown above the sequences and colored in green. The HEAT repeats of the small lobe are labeled. Conserved residues are highlighted with color, with darker shades of slate-blue indicating high conservation.



Figure 3—figure supplement 4. Tex1 structure-based sequence alignment. The sequence alignment includes Tex1 orthologues from *Saccharomyces* cerevisiae (P53851), *Schizosaccharomyces pombe* (Q9USL1), *Homo sapiens* (Q96J01), *Danio rerio* (Q6AXK9), and *Drosophila melanogaster* (Q9VHT2). The secondary structure elements are shown above the sequences, labeled and colored in cyan. Conserved residues are highlighted with color, with darker shades of slate-blue indicating high conservation.



Figure 3—figure supplement 5. Sub2 structure-based sequence alignment. The sequence alignment includes Sub2 orthologues from *Saccharomyces cerevisiae* (Q07478), *Schizosaccharomyces pombe* (O13792), *Homo sapiens* (Q13838), *Danio rerio* (Q7SXU7), and *Drosophila melanogaster* (Q27268). The secondary structure elements are shown above the sequences and colored in light pink for RecA1 and dark pink for RecA2. Conserved residues are highlighted with color, with darker shades of slate-blue indicating high conservation.



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).

eLife Research article



Figure 4—figure supplement 1. Superposition of the two THO protomers. The flexible protomer (colored in gray) is aligned on the rigid protomer (colored as in *Figure 2*). The coiled-coils of Mft1-Thp2 start to bend at the point where the Tho2 β-hairpin is close to Mft1 and then cross over to form the dimerization interface.



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.

Alignment Tho2

S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	1 MA EQTLLSK - LNALSQKV I P	75 57 68 65 26
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	76DENSPLKLSDVASFTNELVNHER-QV SQASIVGK-MFIAVSSTV-PNINDLTTISLCKLIPSLHEELFKFSWISSKLLNKEQTTLLTHLLKKSKYELKKYNLLVENSVGY 1 68 IFSLAFERLEEKNTLASYVIDTLWLFDTEWINNFHEGS-HDKAEKRL-VSIGKG-KEFLEBEWLLSRLDCKFENINVVPNGD	182 178 172 169 231
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	183 GOVALLILA YVDP-ONFSKVSÄYLKE I YHIMGK YS DS I RTLOVILOVSS OF I TECYK FFIAL I RKSDSWPSSHVANNSNSS-NEGOMIAAN I SENL 2 183 GOVALLILA YVDPONFSKVSÄYLKE I YHIMGK YS DS I RTLOVILOVSS OF I TECYK FFIAL I RKSDSWPSSHVA	282 304 252 247 309
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	283 SQYNE EV DK EN VERYMOMCC I LLK NG FVN FYST WONVK PEMEFLQE Y I QN LET ELE EESTK GVENPLAMAAA - LSTENE TD EDNA LV	394 396 336 331 392
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	395 TQQ-DILLFGKIKLERELIHGCVIPVIHVIKQYPKVLYVSESLSRYLGRVFEYLLNPLYTSMTSSGESK-DMATAL-MITRIDNGILÄHK PRLIHK VK THEPFE-SLELNSSYVFYYSE 5 397 - EIKKAK PSQKLGLLKSLLSIGDLSSELILGRYPFLRAYPELSNLYHKLHHISISSIYANYSPLKLLPNDVRERLKQPKFIPEDSRLREITLRPPKENLVFSLDPFADRFNKTESEVFYYFE 5 317 - VE-KPPDNQKLGLLEALLKIGDWQHAQNIMDQMPPYYAA-SHKLIALAICKLIHITIEPLYRRVJVPKGAKG-SPVNALQNK	510 520 415 410 475
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	12 IV - NSNLTPFASVNDLFENSH I VLS I I GPVELG RI PTELSK I ŠKI GVADI OKNHOSE - SL PVTI DKWI DVVKK I I FPATSLONN PI AT SEVYELMK FPFEK RYFI I VLS I I GPVEK KTOR PTELSK I ŠKI GVADI OKNHOSE - SL PVTI DKWI DVVKK I I FPATSLONN PI AT SEVYELMK FPFEK RYFI I VLS I I GPVEK KTOR PTELSK I SKI GVADI OKNHOSE - SL	529 534 539 533 600
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	THEAT LONG TO THE AT LONG TO THE AT LONG TO THE AT LONG TO THEAT LONG TO THE AT LONG TO TH	758 764 668 662 729
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	ANNIERI 759 HNGN I IAVSILKELI I I TVGG I RDLNEVNIK QLLMLN SG SPLKQYARHLI YD F RDDNSVI SG KITS FF TDQSA I SE I I LLYLY-LNLKAN TQNSH YK I LSTRO E DANN TLLWSFI ELI KHC KKGKRFEE 755 KVNQMFDLV I LK ELI SOMTGLO PWTNLSONQI QGAAGG PVLRQLSI SLI YENPDVX KS SMRLFNT I QKNGLATQLLVLLSGKYSTC I YDV TDENSH LKLI SSLQDEGSDVL YLLMEFLMHVC SFK SYYR 669 KAGK SFDLLI LK EVVQK MAG I EI TDEMTY EQLEAMTGGEQLKA E GGYFGQ I RNTKK SS QRLKDA LLDHDLALPLCLLMAQQRNGVI F SEGG-EKHLKLVGKLYDQCHDTLVQFGGFLASNLSTEDI YK 763 KAGK SFDLLI LK EVVQK MAG I EI TDEMTY EQLEAMTGGEQLKA E GGYFGQ I RNTKK SSNRLK EA LANNDLAVA I CLLMAQQRNGVI F SEGG-EKHLKLVGKLYDQCHDTLVQFGGFLASNLSTEDI YK 773 OK SGK SLDLLI LKE VVQK MAG I EI TDEMTY EQLEAMTGGEQLKA E GGYFGQ I RNTKK SSNRLK EA LANNDLAVA I CLLMAQQRNGVI F SEGG-EKHLKLVGKLYDQCHDTLVQFGGFLASNLSTEDI YK 774 KSGK SLDLLI LKE VVQK MAG VESCE EMTHDDLQAMCGGEQLKGEGGYFGQ I RNTKK SSNRLK EA LANNDLAVA I CLLMAQQRNGVI F SEGG-EKHLKLVGKLYDQCDTLVQFGFLASNLSTEDI YK 775 KSGK SLDLI LKE VVQK MAG VESCE EMTHDDLQAMCGGEQLKGEGGYFGQ V RNTKK SSNRLK EA LANNDLAVA I CLLMAQQKHCVI Y RETAAHSHLKLVGNLYDQCQDTLVQFGFLGSTY SVDE YVE 775 KSGK SLDLI LKE VVQK MAG VESCE EMTHDDLQAMCGGEQLK F SUGY RYK SSNRLK EA LANNDLAVA I CLLMAQQKHCVI Y RETAAHSHLKLVGNLYDQCQDTLVQFGFLGSTY SVDE YVE 776 KSGK SLDLI LKE VVGK MAG VESCE EMTHDDLQAMCGGEQLK SDV RYK SSNRLK EA LANNDLAVA I CLLMAQQKHCVI Y RETAAHSHLKLVGNLYDQCQDTLVQFGFLGSTY SVDE YVE 776 KSGK SLDLI LKE VYGK MAG VESCE EMTHDDLQAMCGGEQL RGEAGYFGQ V RYK SSNRLK F F ANNDLAVA I CLLMAQKHCVI Y RETAAHSHLKLVGNLYDQCQDTLVQFGFLGSTY SVDE YVE F 777 KSGK STUDI LKE VYGK MAG V SSCE F F F F F F F F F F F F F F F F F F F	384 394 795 789 857
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	85 NV L PFV ELNNRFHLS TPWT HI WRDY - LDNQ LNSN - EN FS TOE LI - EGA EFSDVD LTK I SKDLFT FWRLSLYD I HFDK SLYDERKNALSGENT 9 85 LI PS FEQLI (QOFH (QOFWA FY LS RYKNLDH SLTGSN TEDAMD I D YEN	375 307 311 905 976
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	976 GHMSN RKK H LIQNQIKDILVTGISHQRAFKKTSEFIS - EKSN-VWN	J78 133 127 121 395
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	1079 LFCCTSSEAGNLGL FFTDVLKKLEKMRLNGDFNYGFLTIERASGF0GGNKADQLDYEN FRUVKEVKWKYY ITGYFESCLLSTEFYMHIYNSV FPVVKAHDILDYYTTLEENL 11 1134 IFSMTQREADNFGRFLYELVDITSWYRDX ILVERECLANGALPGFRLYWSDEQNO-PDLSAVLPYMKFVLLFSKWMKYY ISYFESCLLSTEFYMHIYNSVIILEK LPCFPLIESGSALKRAAERKL 1134 IFSMTQREADNFGRFLYELVDITSWYRDX ILVERECLANGALPGFRLYWSDEQNO-PDLSAVLPYMKFVLLFSKWMKYY ISYFESCLLSTEFYMHIYNSVIILEK LPCFPLIESGSALKRAAERKL 1134 IFSMTQREADNFGRFLCCMLETVTRWHSDRATYEKEGGNYPGFLTIERATGFDGGNKADQLDYENFRHVVHKWHYKLTKASVHCLETGEYTHIRNLIULTKILPWYPKVLNLGQALEERVHKIC 1132 VASCTENESRRYGRFLCCMLETVTRWHSDRATYEKEGGNYPGFLTIERASGFDGGNKADQLDYENFRHVVHKWHYKLTKASVHCLETGEYTHIRNLIULTKILPWYPKVLNLGQALEERVHKIC 1136 VTSCTEGEATRYGRFLCAMLETVTRWHADQAVFNKECANYPGFLTIERASGFDGGNKADQLDYENFRHVVHKWHYKLTKASVHCLETGEYTHIRNLIULTKILPWYPKVLNLGQALEERVHKIC 1136 VTSCTEGEATRYGRFLCAMLETVTRWHADQAVFNKECANYPGFVTKFRVSNQF-SEANDHVGYENYRHVCHKWHYKLTKASVHCLETGEYTHIRNLIULTKILPWYPKVLSKLAQIIERKVDKVRI	173 260 152 146 219
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	1174 IN - EE REDIK VEDSALIGH KARLKDALELDEFCTLTE-EEAEQK R-I-REME-LEEIKNYETACQNEQKQVALRKQLEL-NKSQRLQND-PPKSVASGSAGLNSKDRY 12 126 JDEEKREDIK VEDSALIGH KARLKDALELDEFCTLTE-EEAEQK R-I-QR-PQQLSVAATSAVDSKTASJSEQAKIDKQKVAL-NPSAPEFVPDSTPSDAVASETDNKN 12 126 JDEEKREDIK VEDSALIGH KARLKDALENDFPNSSFSGTVR - PSNSEK -LQR-PQQLSVAATSAVDSKTASJSEQAKIDKQKVAL-NPSAPEFVPDSTPSDAVASETDNKN 12 125 JQEEKEKRPDLYALAGYSQLKSKKSWI PENEFHHKDP - PPRNAV - ASVQNCPGGGPSSSSIGSASKSDESSTEETDKSRERSQC-GVKAVNKASSTTPKCNSSNGNSG 12 127 QEEKEKRPDLYALAMGYSGQLKSKKVHMV PENEFHHKDP - PPRNAV -ASVQNCPGOGPSSSSIGSASKSDESSTEETDKSRERSQC-GVKAVNKASSTTPKCNSSNGNSG 12 127 QEEKEKRPDLYALAMGYSGQLKSKKVHMV PENEFHHKEP-PVRSTTTGTLQNCPGNIGKTSTTTAASAGKTEDGVVEDSEKSKNSQA-AQKTASKNSSTTPKCNSSSNGNSG 12 1220 EEEKKRPDLYALAASSYIGQLKLKTPHMLKESVEHQIAERPNKESP-TS-VGAPAATRSDKLSPTSPSGN	275 364 259 256 303
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster	1276 TYŠ RNE PVIPTK PSS SQWSYSK V TRHVDDINHYLA TNHLQKA I SLVENDDE TRNL KKLŠKOM PIF – DFRN – 19 1365 L – VENKAVE – – – – – – – – – – – – – – – – – – –	34 5 40 2 30 5 30 0 40 0
S.cerevisiae S.pombe H.sapiens D.rerio D.melanogaster S.cerevisiae	1346 STLE IF RRY FRTLION PON PD FAKIDSLKRYIK. N - SREP Y PD TTS SYSEAAA PEYTKRS SRE 1403 RTRT RTNO NO. RTRT RTNO NO. SREP Y PD TTS SYSEAAA PEYTKRS SREP Y PD TTS SYSEAAA PEYTKRS 1403 RTRT RTNO NO. NO. SREP Y PD TTS SYSEAAA PEYTKRS SREP Y PD TTS SYSEAAA PEYTKRS 1304 RTRT RTNO NO. NO. NO. SYSEAAA PEYTKRS SREP Y PD TTS SYSEAAA PEYTKRS 1305 ER SKEK PT ANA KEM KSKEN VK EK F K KK EK F ANK - KK K EK PK - KK NK NK MK KEM - VK CI SKEK PT ANA KEM KSKEN VK CI SKEK PT SKEK F F SR EK PI DE SKEK PT SKEK F F SK EK F SK EK F SK EK F SK EK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PT SKEK F SK EK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK EK PI DE SKEK PT SKEK F F SK	407 457 381 360 491
S.pombe H.sapiens D.rerio D.melanogaster S.cerevisiae	1930 - A NA AR EHE SQK SDRIV OUCHVIN HAR RY SNNINS TNY GRE BS SANN RTSN DNK AD EV TECK DON K ROD	498 498 473 582 592
S.pombe H.sapiens D.rerio D.melanogaster S.cerevisiae S.nombe	1540 RAC R SNGSNRGNDS RDA DG R R 5 TH YA SNK RP RS - SD SQ 5 P - SN L R E DE R E NS R R RA RQD R RD RDS RQ OR D - P P - RD R T	527 588 555 631
H.sapiens D.rerio D.melanogaster	1589 - S DKH R 1556 - S EKH R 1632 K S A R S S RVN Y	

Figure 5—figure supplement 1. Conformational states of Sub2. Comparison of the activated semi-closed Sub2 structure as found at the proximal side of the THO homodimer cryo-electron microscopy structure with the conformation of active Sub2, bound to RNA, nucleotide, and the C-terminal motif of Yra1 (C-box) (*Ren et al., 2017*). All structures are aligned on the N-terminal RecA1 domain. The RecA1 domain is colored in light-pink, the RecA2 in purple, and Yra1 in gray. RNA is shown in black and nucleotide in yellow.



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*).

Alignment I	Hpr1
S.cerevisiae	1 MSNTEEL I
S.pombe	1 MEVOKGLI EAFYN TYPLE KAKELDK SPLC SEYELF I KELWPS I VE - SFHN STEF 53
H.sapiens	1 MS PT PP LFS LP EA RTRFTK ST REALNNKN IK PLLSTFSQV PGS - EN EKKCTL 51
D.rerio	1 MS PP S - H F D F I EA RD K F T VA TK NAVD TRNCK PL TTA F SH L PGN - E TEKKATL 50
D.melanogaster	1 MS TAV EVAN TA PEK LANY FALQS A F E KALELA I TDGK VELLVKEYN RFPAN TEHDK RLPM 60
Scarovisiaa	
S.nombe	47 EWE - PLEKNISA IH DKDSLIDITLK KFITD SMINATEDEEENNLEKGLINSCIGLDFV 103
H.sapiens	52 DOAEPCHEEFLINKSSCEWIALL_SLAIC_CVTECICTAST
D.rerio	51 DOAL RGV L FEO LYNOKY
D.melanogaster	61 DHA F RV LLMK RLD ED
S.cerevisiae	104 YN S R FN R SN PA SWGN T F F E L F S T I I D L L N S P S T F L K FW P YA E S R I EW FK MN T S V E P V S L G E S N L I S Y 170
S.pombe	102 PFLILEELMDIHT-VNECAKLYEYFETRPSLMKGIVSNRGR 141
n.sapiens Direrio	93
D.nelanogaster	
Dimenanogaster	
S.cerevisiae	171 KQPLYEKLRHWNDILAKLENNDILNTVKHYNMKYKLENFLSELLPINEESNFNRSASISALQESDNE 237
S.pombe	142 GPVLLRISNELLRRLSRQENSS FCGRIDILLSKAFPPEERSGANLRGDYNTVHS 195
H.sapiens	133 KNYLLRMCNDLLRRLSKSQ NTV FCGRIQLFLARLFPLSEKSGLNLQSQFNLENVT - V 188
D.rerio	133 KNYLLRMCNDLLRRLSK SQ NTV FCGRIQLFLARLFPLSEK SGLNLQSQFNLDNIT V 188
D.melanogaster	140KNNILRMCNDLLRRLSRTQNTVFCGRIQLFLSKFFPFSERSGLNIVSEFNLDNFTE 195
S.cerevisiae	
S.pombe	
H.sapiens	189 FN TN EQESTLGQKH TED REEGMDV EEGEMGD EEA P TTC S I P I D YN L Y
D.rerio	189 FNK N EQDS T L G Q H T E V K E E G M D V E E G E M G D E D A P A P S S I P I D Y N L Y 235
D.melanogaster	196 YG LD SK DHD E S DNK E L E D TA E D I P LK I D YD L Y 227
S.cerevisiae	278 LEYKNDVDRSLSPLLDAILEIEENFYSKIKMNNRTRYSLEEALNTEYYANYDVMTPKLPV YMKHSNA 344
S.pombe Hisanians	225TAYWDLQCMC - SNPPKLLASDT
D rerio	236 KK FWSLQDYF - KN PV QC YEK I S WK I FLK YSE - 265
D.melanogaster	200
5	
S.cerevisiae	345 MK MD RN F FWAN LON LK F SDD
S.pombe	255KYI-TLDKGEPSKYIYS RSL 306
H.sapiens	266EV LAV FK S YK LDD TQA S RKK MEEELK T - GG EHV Y FAK FL T S EK L 307
D.rerio	266EALAVFKSFKLDDMQASKKKLEEALAVFKSGDHVYFAKFLTSEKL 308
D.melanogaster	258 N I LQS F S S F K L E D V RQS S N E NA S G T D Q A M D V D E E A V D T T A V S S V - I K A N H F F A K F L T N P K L 317
Scarovisiaa	
S nombe	3/2 MDTSLSNTTCLYKQLTQEDDDYYRKQFTLQLCFTTNLTRNTSSDE-TRNFYKSCYLRENPLSD 434
H.sapiens	307 FEYQL
D.rerio	309 MD L 0 L
D.melanogaster	318 LALQL
S.cerevisiae	435 IDFENLDEVNKKRGLNLCSYICDNRVLKFYKIKDPDFYRVIRKLMSSDEKFTTAKIDGFKEFQNFRI 501
S.pombe	361 EDTSK LNELSK EA YSFLHTARCGSVQRTIK EIIHIEGNWK LWK GLGC PSLEK PLV 415
H.sapiens	348 EQSLWIEDTTKSVYQLLSENPPDGERFSKMVEHILNTEENWNSWKNEGCPSFVKERT 404
D.rerio D.molanogastor	349 DQS LWI ED TTK LV YQL LK E I PP DGDK FG SMV EH I LN TE ENWN SWK NEGC PS FVK E RP 405
D.meianogaster	358 DQADFTKETES RVYKLLEETPPYGK RFS RTVYHMLA REEMWNNWKNEGCK EFKKPEE 414
S.cerevisiae	502 SKEK I P PPA EDETE-KKETE IKMGNKI INNVWKI - PTG-LDKI EQEVKK P 548
S.pombe	416 DKAA I DEAV E - GLKK LTN TPVK - L RFAMGNAALS RLWEOAGEN TLDDLKK FERYRI PSP 472
H.sapiens	405 SDTK PTR I I RK RTA PEDFLGKG - PT - KK I LMGN EELTRLWN LC PDN - MEACK SETREHMPTL 463
D.rerio	406 A E TK P I R P S RK RQA P E D F L G K G - P D - RK I L MGN D E L T R L WN L N P D N - MEA C K S E N R E F M P S L 464
D.melanogaster	415 PTLSEEDSK PTPNK RPRRPLGDALRDASRS-GK FYLGNDNLTRLWNYSPDN-LQACK SEQRNFLPLL 479
Scerevisiae	
S.pombe	343 EGVITEAAT CARWESKISSEISGEARDEIIKQUILKELKSKILDPUKVNE-KIGVDGLEEEPK 0.1
H.sapiens	464 EFFFEA- I FOADPENMY EN EYKAVN NSNYGWRALRIAR SHFFOPTNO-O
D.rerio	455 EDFFEEA - I EQADPANMV EDEYKVV R N SNYGWRALRLLS RRSPHFFOPTNO - 0 515
D.melanogaster	480 E T Y L E T P - H E K V D P

Figure 6—figure supplement 1. Hypothetical model of THO-Sub2-Yra1 (transcription-export) in binding RNA:DNA hybrids. (A) The scheme on the top left shows the domain organization of Yra1. The N-terminal and C-terminal regions of Yra1 bind to Sub2 with similar affinities as measured in the Figure 6—figure supplement 1 continued on next page

Figure 6—figure supplement 1 continued

fluorescence anisotropy experiment on the right panel. In this experiment, we used fluorescent peptides corresponding to the N-terminal and Cterminal regions flanking the RRM, with sequences shown at the bottom right panel, in an alignment that highlights the similarities between the two motifs. Stars indicate residues that are involved in binding the C-box to Sub2 (*Ren et al., 2017*). The peptides were labeled with a 5'-6-carboxyfluorescein (6-FAM), dissolved to a concentration of 10 nM and incubated with Sub2 for 20 min at different concentrations in a buffer containing 20 mM Tris-HCl (pH 7.4), 250 mM NaCl, and 2 mM dithiothreitol, at 20°C in 50 µl reactions on an Infinite (Tecan). The excitation and emission wavelengths were 485 and 535 nm, respectively. Each titration point was measured three times using 10 reads with an integration time of 40 µs. The data were analyzed by nonlinear regression fitting using the BIOEQS software. Both peptides bound Sub2 with a similarity affinity, with a Kd of 2.9 µM. (B) Hypothetical model of RNA:DNA hybrid recognition by THO. The cryo-EM structure of THO and Sub2 RecA2 is presented in surface representation. In the cartoon surface representation, the model is based on the superposition with the Sub2–RNA–ADP BeFx structure with Yra1 C-box (in gray, [*Ren et al., 2017*]) and fitted with an RNA:DNA hybrid (pdb code 1hys) as it positions the 5' ends of the RNA (red) and DNA (black) strands at the two opposite RecA1 domains. The duplex could bind Sub2 similarly to what has been reported for the DEAD-box protein Mss16 (*Mallam et al., 2012*) and consistently with the properties of the human orthologue (*Pérez-Calero et al., 2020*).