

# Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor

Álvaro Peña<sup>1,5</sup>, Kamil Gewartowski<sup>2,3,5</sup>,  
Seweryn Mroczek<sup>2,3</sup>, Jorge Cuéllar<sup>1</sup>,  
Aleksandra Szykowska<sup>2,3</sup>, Andrzej  
Prokop<sup>2,3</sup>, Mariusz Czarnocki-Cieciura<sup>2,3</sup>,  
Jan Piwowarski<sup>3</sup>, Cristina Tous<sup>4</sup>,  
Andrés Aguilera<sup>4</sup>, José L Carrascosa<sup>1</sup>,  
José María Valpuesta<sup>1,\*</sup> and  
Andrzej Dziembowski<sup>2,3,\*</sup>

<sup>1</sup>Department of Structure of Macromolecules, Centro Nacional de Biotecnología (CNB-CSIC), Department of Molecular Biology, CSIC, Madrid, Spain, <sup>2</sup>Department of Biophysics, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland, <sup>3</sup>Department of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, Warsaw, Poland and <sup>4</sup>Centro Andaluz de Biología Molecular y Medicina Regenerativa CABIMER, Universidad de Sevilla-CSIC, Sevilla, Spain

**The THO complex is a key factor in co-transcriptional formation of export-competent messenger ribonucleoprotein particles, yet its structure and mechanism of chromatin recruitment remain unknown. In yeast, this complex has been described as a heterotetramer (Tho2, Hpr1, Mft1, and Thp2) that interacts with Tex1 and mRNA export factors Sub2 and Yra1 to form the TRanscription EXport (TRES) complex. In this study, we purified yeast THO and found Tex1 to be part of its core. We determined the three-dimensional structures of five-subunit THO complex by electron microscopy and located the positions of Tex1, Hpr1, and Tho2 C-terminus using various labelling techniques. In the case of Tex1, a  $\beta$ -propeller protein, we have generated an atomic model which docks into the corresponding part of the THO complex envelope. Furthermore, we show that THO directly interacts with nucleic acids through the unfolded C-terminal region of Tho2, whose removal reduces THO recruitment to active chromatin leading to mRNA biogenesis defects. In summary, this study describes the THO architecture, the structural basis for its chromatin targeting, and highlights the importance of unfolded regions of eukaryotic proteins.**

*The EMBO Journal* (2012) 31, 1605–1616. doi:10.1038/emboj.2012.10; Published online 7 February 2012

**Subject Categories:** RNA; structural biology

**Keywords:** electron microscopy; mRNA export; mRNP quality control; THO complex; TRES complex

\*Corresponding author. JM Valpuesta, Department of Structure of Macromolecules, Centro Nacional de Biotecnología (CNB-CSIC), CSIC, Madrid 28049, Spain. Tel: +34 915854690, Fax: +34 915854506; E-mail: jmv@cnb.csic.es or A Dziembowski, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw 02-106, Poland. Tel: +48 225922033; Fax: +48 8237189; E-mail: andrzejd@ibb.waw.pl

<sup>5</sup>These authors contributed equally to this work

Received: 13 September 2011; accepted: 5 January 2012; published online: 7 February 2012

## Introduction

mRNA biogenesis and export is a very complex process involving transient interactions between a large number of proteins and assemblies. During transcription elongation, pre-mRNA molecules are packed into RNA-protein assemblies termed mRNPs (Kohler and Hurt, 2007). All steps leading to the production of translation-competent mRNA in the cytoplasm (transcription, mRNA processing, and export from the nucleus) are tightly coupled, and impairment of any step leads to the activation of the RNA surveillance pathway and the consequent degradation of improper mRNA molecules (Houseley *et al*, 2006).

THO is an evolutionarily conserved macromolecular assembly that functions during transcription facilitating the mRNP packaging and export. The yeast THO complex associates with chromatin in a transcription-dependent manner and is essential for efficient co-transcriptional recruitment of mRNA export factors Yra1 and Sub2 (Strasser *et al*, 2002). Therefore, THO plays an important role in coupling transcription to mRNA export, although its precise function is still elusive. Yeast THO has been described as a four-subunit complex composed of Tho2 (180 kDa), Hpr1 (90 kDa), Mft1 (45 kDa), and Thp2 (30 kDa) (Chavez *et al*, 2000), for which no structural information is available. It has been shown that THO interacts with three other proteins, the RNA helicase Sub2, the RNA-binding protein Yra1, and Tex1 whose function is unknown, forming the TRES complex—from TRanscription and Export (Strasser *et al*, 2002). Moreover, during transcription, THO interacts with mRNA export factor Mex67, and the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 (Strasser *et al*, 2002; Zenklusen *et al*, 2002; Hurt *et al*, 2004). It has also recently been published that TRES interacts with the Prp19 complex, involved in splicing and transcription elongation (Chanarat *et al*, 2011).

Inactivation of the THO subunits results in remarkable molecular phenotypes, which reveals the important role of the complex in mRNA biogenesis and genome stability. Lack of THO causes impairment of mRNP formation leading to defects in transcription elongation and to the formation of RNA/DNA hybrids (R-loops), which in turn causes genomic instability (Huertas and Aguilera, 2003). Another interesting phenotype generated by THO deletion is the formation of large aggregates (called heavy chromatin) composed of transcriptionally active chromatin, nascent transcripts, RNA export machinery, and nuclear pore complexes (NPC) (Rougemaille *et al*, 2008). Finally, the expression level of long GC-rich genes (like *LacZ*) is markedly reduced in *tho* mutants (Chavez *et al*, 2001). All these phenotypes together with the interaction of THO with export factors described above strongly suggest that its activity is directly associated with mRNA transcription, biogenesis, and export.

The interaction between THO and active chromatin, together with biochemical studies of Yra1 and Sub2, have

led to a model in which the THO/TREX complex is recruited to mRNA at the early stages of the export pathway (Strasser *et al*, 2002; Zenklusen *et al*, 2002). Then, mRNA is transmitted to the Mex67/Mtr2 export receptor, which interacts with phenylalanine-glycine (FG) repeat-containing nuclear pore proteins, thus facilitating mRNA translocation through the NPC (Reed and Cheng, 2005; Kohler and Hurt, 2007).

Despite numerous studies, the mechanism of THO function is not well defined. So far, no structural information regarding THO and its subunits has been published and it is unclear whether it interacts directly with nucleic acids and how it is recruited to chromatin. Interestingly, however, a recent report suggests that THO interaction with active chromatin is partially dependent of Syf1—a subunit of the Prp19 complex involved in transcription elongation and splicing (Chanarat *et al*, 2011).

In this work we provide mechanistic insight into the THO function. We demonstrate that Tex1 interacts stably with THO as it co-purifies with the other subunits even at high salt concentrations. We present the three-dimensional reconstruction of the five subunits of the THO complex and the localization, within the structure, of the subunits Tho2, Hpr1, and Tex1. Furthermore, we show that the largest THO subunit Tho2, and in particular its C-terminal domain, is directly responsible for interaction with nucleic acids (ssDNA, dsDNA, and RNA). Deletion of this fragment, while not altering the assembly of the complex, leads to defects in mRNA biogenesis and increases genome instability. Most importantly, the intrinsically unfolded C-terminus of Tho2 is essential for efficient recruitment of the THO complex to chromatin.

## Results

### Purification and structural characterization of the THO complex

To characterize the THO complex, we devised an efficient purification procedure. The native THO complex was purified by IgG affinity chromatography followed by ion exchange chromatography using a *Saccharomyces cerevisiae* strain with a TAP-tagged Tho2 protein (Dziembowski and Seraphin, 2008). The purified complex reproducibly contained not only the four-core THO subunits (Tho2, Hpr1, Mtf2, and Thp2) but also the TREX component Tex1 (Figure 1A; Supplementary Figure S1), while the other TREX components (Yra1 and Sub2) were conspicuously absent. In the present manuscript, we refer to the five-subunit assembly as the THO complex (Jimeno *et al*, 2002; Strasser *et al*, 2002; Rehwinkel *et al*, 2004). The purified THO complex isolated from *S. cerevisiae* was not very soluble (<1 mg/ml), too low to attempt crystallization, but sufficient for structural analysis by electron microscopy. Aliquots of THO were negatively stained and observed by electron microscopy (Figure 1B), which revealed the presence of a homogeneous population of long, thin particles (top gallery in Figure 1B). A total of 14 115 particles were selected, aligned, and classified as described in Materials and methods. A maximum-likelihood classification revealed as the largest population a croissant-like structure ~220 Å long and ~115 Å high, with a flat surface at the base and two large protrusions at the top, one long and thin and the other shorter but wider (Figure 1C). From the tip of the larger protrusion stems a thin and flexible stain-excluding

mass (arrow in Figure 1C), which appears to be sticky as suggested by the presence (~20% of the population) of dimers of THO complexes interacting through this region and forming butterfly like structures (see the bottom gallery in Figure 1B and the average image of 1800 particles in Figure 1D).

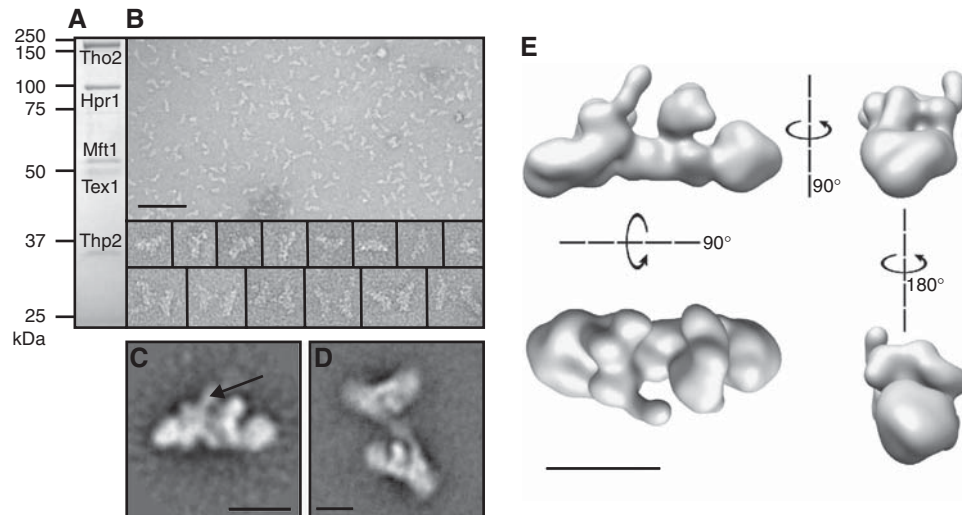
The three-dimensional reconstruction generated using these particles (~17 Å resolution) revealed in full detail the features described above (Figure 1E; Supplementary Figures S2–S5). It is important to point out that the flexible mass stemming from the long protrusion was not observed in the reconstruction, probably due to an averaging out caused by the presence of different conformations of this domain.

### Mapping of Tex1 and Hpr1 into the THO complex

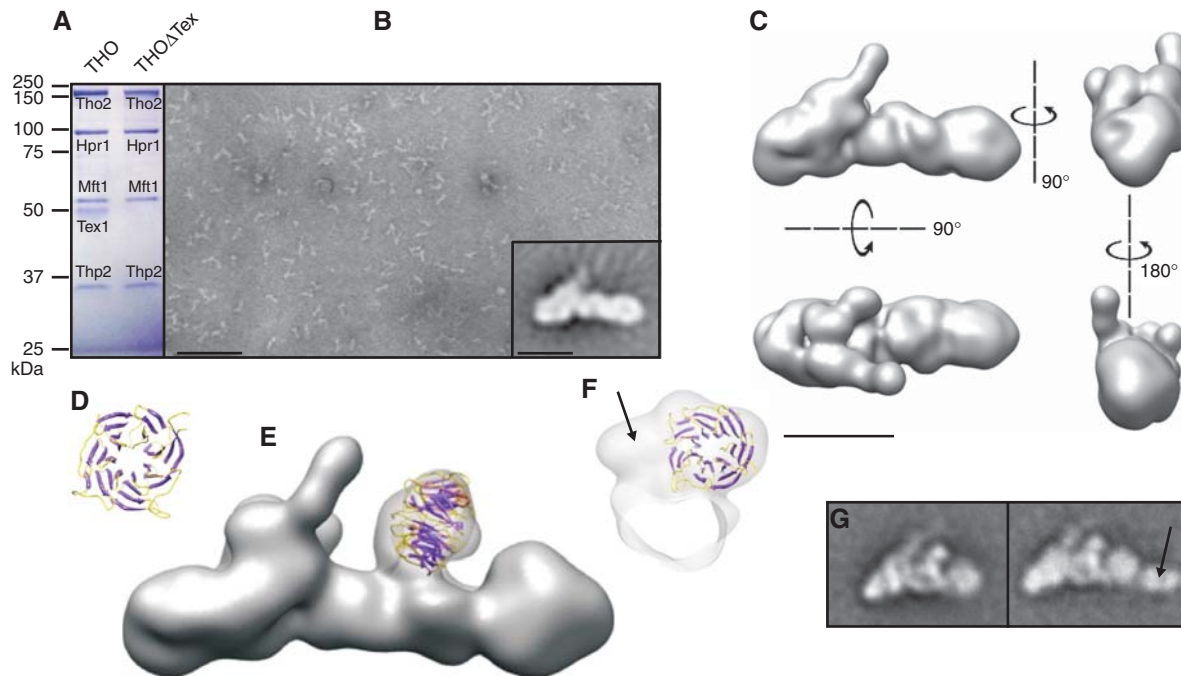
To further characterize the structure of the THO complex, we sought to locate some of the subunits within the complex. We first assessed the position of Tex1. For this purpose, a new strain lacking Tex1 (Tho2TAPΔTex1) was constructed with a TAP-tagged Tho2 subunit. A homogeneous, stoichiometric complex composed of Tho2, Hpr1, Mtf2, and Thp2 proteins was purified using the protocol described above (Figure 2A). Aliquots of the complex were subsequently negatively stained and observed by electron microscopy (Figure 2B), which again revealed the presence of a homogeneous population of long, thin particles. The two-dimensional average image obtained after two-dimensional maximum-likelihood classification and averaging of the largest population selected from 13 273 particles (inset in Figure 2B) revealed a similar overall structure to that obtained for the THO complex but lacking one of the protruding masses. This was confirmed by the three-dimensional reconstruction (~20 Å resolution), which showed the same long, thin, and asymmetric structure, albeit lacking the wider protruding mass in the centre of the THO structure (Figure 2C). Therefore, we assign this protrusion to Tex1.

Tex1 has been described to contain several WD40 domains, which could give rise to a β-propeller structure (Rehwinkel *et al*, 2004). We generated an atomic model of residues 47–371 of Tex1 (residues 1–46 and 372–422 could not be modelled into any known structure), which revealed a seven-blade β-propeller structure (Figure 2D). The atomic model was subsequently docked into the three-dimensional reconstruction of the THO complex. Docking, either manual or automatic, suggested that the atomic model of Tex1 fitted well into the part of the THO volume in which Tex1 was mapped to (Figure 2E and F), leaving only a small region that probably contains the non-modelled C-terminal region of the protein (arrow in Figure 2F).

To locate other subunits of the THO complex we performed immunomicroscopy, using either specific antibodies against the different subunits or epitope-tagged subunits. These approaches produced inconclusive results, with the exception of one polyclonal antibody against the C-terminal region of Hpr1, which stably associated with the THO complex. Aliquots of the immunocomplex were negatively stained and a total of 8250 particles were selected and processed. The two-dimensional average obtained revealed an extra mass (arrow in Figure 2G) that corresponds to the F<sub>ab</sub> domain of the antibody bound to the wider end of the THO structure (Figure 2G), indicating the position of the Hpr1 subunit.



**Figure 1** Three-dimensional reconstruction of the THO complex. (A) SDS-PAGE of the THO complex purified by affinity chromatography followed by ion exchange chromatography, which reproducibly showed the presence in stoichiometric amounts of the four canonical THO subunits (Tho2, Hpr1, Mft1, Thp2) and Tex1. (B) An electron microscopy negatively stained field of THO particles. Bar indicates 1000 Å. The top gallery shows a selection of THO particles and the bottom gallery, a selection of double, butterfly like THO particles. (C) Two-dimensional average image of the THO complex. Arrow points to the flexible region described in the text. (D) Two-dimensional average image of the double, THO particle. Bar indicates 100 Å in (C, D). (E) Four orthogonal views of the three-dimensional reconstruction of the THO complex. Bar indicates 100 Å.

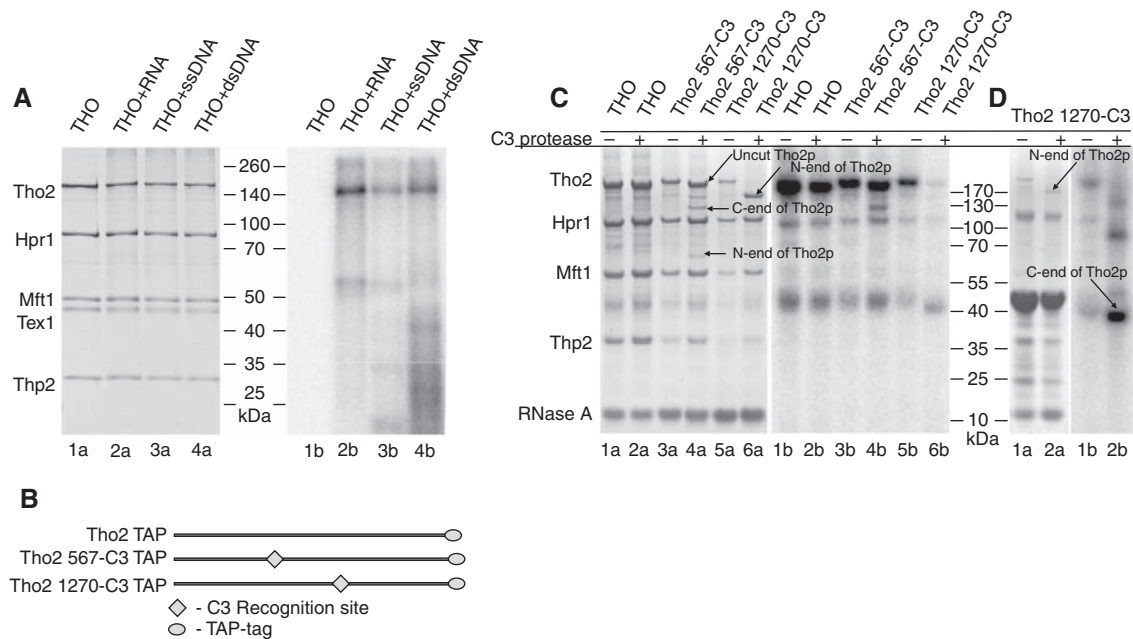


**Figure 2** Three-dimensional reconstruction of the THO $\Delta$ Tex1 complex and mapping of Tex1 and Hpr1 into the THO complex. (A) SDS-PAGE of THO complexes purified by affinity chromatography from a Tho2TAP and Tho2TAP $\Delta$ Tex1 *S. cerevisiae* strains. The right lane shows that THO complex can assemble without Tex1 protein. (B) An electron microscopy field of negatively stained THO $\Delta$ Tex1 particles. The inset shows the two-dimensional average image of the most common view of the complex. Bar indicates 1000 Å in the micrograph and 100 Å in the inset. (C) Four orthogonal views of the three-dimensional reconstruction of THO $\Delta$ Tex1. Bar indicates 100 Å. (D) Atomic model of the N-terminal,  $\beta$ -propeller domain of Tex1. (E) Docking of the atomic model of Tex1 into the corresponding mass of the THO complex. (F) The same docking in an orthogonal, cut section of the complex. The arrow points to the putative region filled by the non-reconstructed, C-terminal region of Tex1. (G) Two-dimensional averages of the THO complex (left) and the immunocomplex formed between THO and the anti-Hpr1 polyclonal antibody (right).

**The C-terminal region of the Tho2 protein interacts with nucleic acids**

THO has been shown to associate with RNA and DNA *in vitro* (Jimeno *et al*, 2002), but its function and the subunit(s)

involved in this interaction are not known. To characterize these interactions, we performed UV cross-linking experiments using the highly purified complex and  $^{32}$ P-labelled oligonucleotides (either RNA, ssDNA, or dsDNA)



**Figure 3** The C-terminal region of Tho2 interacts with nucleic acids. (A) Tho2 cross-links to RNA, ssDNA, or dsDNA. Purified THO was incubated with radiolabelled oligonucleotides (either RNA, ssDNA, or dsDNA) and UV cross-linked separated by SDS-PAGE. Proteins were visualized by Coomassie blue staining (left) while cross-linked radiolabelled nucleic acid was visualized using autoradiography (right). (B) Diagram of a modified Tho2 protein showing the position of the C3 protease site and the TAP-tag. (C) The C-terminal region of Tho2 is essential for Tho2-RNA interaction. A C3 cleavage site was introduced in Tho2 after residues 567aa (Tho2 567-C3 TAP) and 1270 (Tho2 1270-C3 TAP), and the complex subsequently treated with C3 protease and purified by ion exchange chromatography. Afterwards, the complex was UV cross-linked to  $^{32}$ P-labelled *in vitro* transcribed RNA. Products of cross-linking reactions were treated with RNase A and separated by SDS-PAGE. Proteins were visualized by Coomassie blue staining (lanes a) while cross-linked radiolabelled nucleic acid was visualized using autoradiography (lanes b). The Tho2 proteolysis products and visible Tho2 cross-links are indicated. (D) The C-terminal region of Tho2 cross-links to RNA. The THO complex with the Tho2 1270-C3 TAP mutant was C3 protease digested and processed as described above. Novel cross-linking product with molecular weight corresponding to the small C-terminal fragment of the Tho2 protein is indicated.

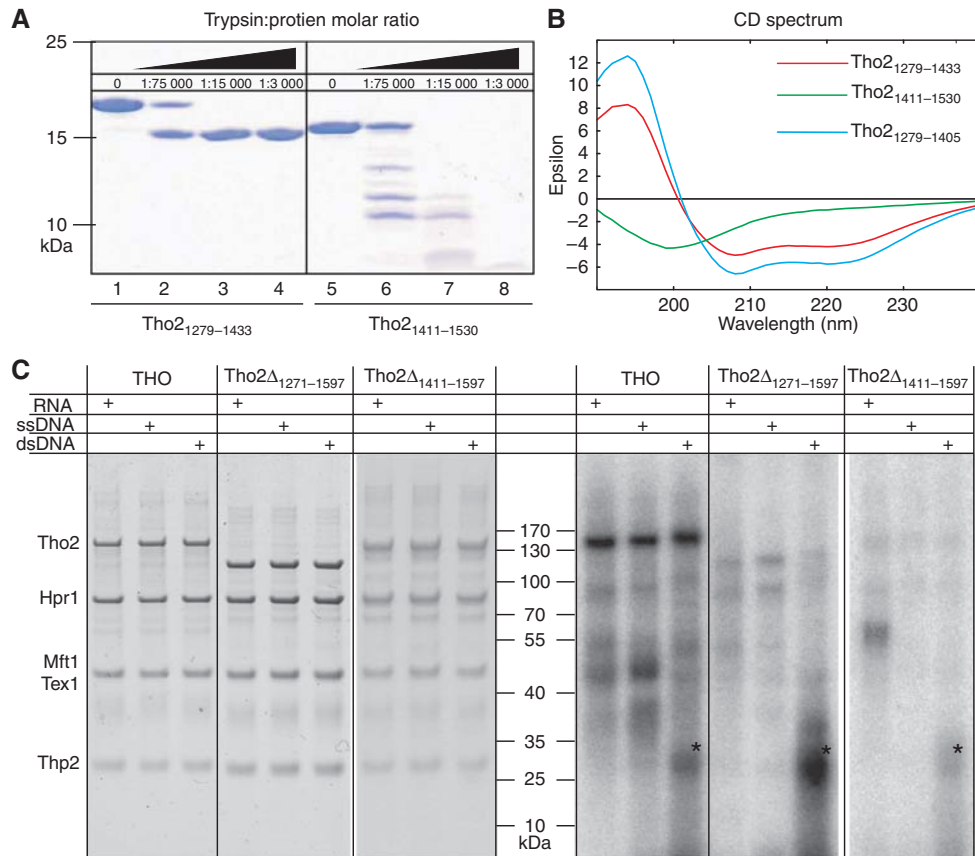
(Figure 3A). The three types of oligonucleotides were cross-linked to the largest subunit of the THO complex, Tho2. These interactions were also confirmed by an electrophoretic mobility shift assay (data not shown).

The analysis of the Tho2 sequence did not reveal any canonical RNA-binding domains and secondary structure predictions analyses suggested that it is composed mostly of  $\alpha$ -helices. Therefore, to identify which part of the very large Tho2 polypeptide interacts with nucleic acids, we performed cross-linking experiments combined with site-specific protease digestion of this protein. For this purpose, we introduced C3 protease cleavage sites at positions 567 (Tho2 567-C3 TAP) or 1270 (Tho2 1270-C3 TAP) of the Tho2 polypeptide in the Tho2TAP strain (see diagram in Figure 3B). The modified THO complexes were purified by IgG affinity chromatography followed by C3 protease cleavage and ion exchange chromatography and then subjected to UV cross-linking (Figure 3C). In the case of Tho2 567-C3 TAP, only a small percentage of Tho2 was proteolyzed, and the RNA remained associated with the intact Tho2 and to the C-terminal fragment excised (Figure 3C, lanes 4a and 4b), which points to this region as involved in RNA binding. In contrast, the digestion of Tho2 1270-C3 TAP was complete, but only the large N-terminal fragment remained associated with the THO complex while the C-terminal 327 aa fragment was lost (Figure 3C, lane 6). Interestingly, RNA-Tho2 cross-link disappeared when the C-terminal fragment was removed, which reinforces the notion of this part of the Tho2 protein being involved in the interaction with RNA.

To confirm that the C-terminal region of the Tho2 indeed interacts with RNA, we altered the order of the procedure and performed protease digestion after all the chromatography steps (Figure 3D). This ensures that both fragments appearing after proteolysis are present in the cross-linking solution. After UV exposure in case of Tho2 1270-C3 TAP, a new cross-linking species was visible at a molecular weight corresponding to the small C-terminal fragment of the Tho2 protein. We conclude that the C-terminal fragment of Tho2 interacts with nucleic acids.

#### **The C-terminal region of Tho2 constitutes a basic unfolded tail not essential for complex integrity**

We set out to characterize the C-terminal, nucleic acid binding region (residues 1279–1597) of Tho2. A bioinformatics analysis of this region using several Tho2 sequences revealed a poorly conserved, highly positively charged and partly disordered region (see the multiple alignment of Tho2 sequences; Supplementary Figure S6A). This region was insoluble when expressed in *Escherichia coli*, so in order to locate the secondary structure elements and the unstructured regions, we combined trypsin digestion and CD analysis of the fragments. We generated 44 constructs encompassing different fragments of the nucleic acid binding region, which exhibited different degrees of solubility, but most of which were highly sensitive to trypsin digestion. The 1411–1530 region was exceptionally sensitive, containing no stable fragments at all (Figure 4A) and its CD spectrum showed minimal values for wavelengths below 200 nm (Figure 4B),



**Figure 4** The C-terminal region of Tho2 forms a basic unfolded tail essential for THO complex nucleic acids binding. **(A)** Limited proteolysis experiments of the recombinant Tho2 protein fragments. **(B)** Circular dichroism spectra of the recombinant Tho2 protein fragments. Notice that Tho2<sub>1411-1530</sub> fragment generates a minimum below 200 nm characteristic of disordered proteins, while Tho2<sub>1279-1404</sub> and Tho2<sub>1279-1433</sub> have minima around 210 and 230, suggestive of a high  $\alpha$ -helix content. **(C)** UV cross-linking between nucleic acids and THO complexes isolated from Tho2TAP; Tho2Δ<sub>1271-1597</sub> and Tho2Δ<sub>1408-1597</sub> strains. Purified THO complexes were incubated with radiolabelled RNA, ssDNA, or dsDNA. After cross-linking, the proteins were separated by SDS-PAGE. Proteins were visualized by Coomassie blue staining (left) while cross-linked radiolabelled nucleic acid was visualized using autoradiography (right). Fast migrating radioactive species (marked with asterisks) represent unbound dsDNA particles.

suggesting a high level of disorder. In addition, we acquired CD spectra at higher temperatures to see if there are any secondary structures to be destabilized by heat, but did not detect any significant differences (Supplementary Figure S6B). In contrast, the CD spectrum of the 1279-1433 fragment showed a high content of  $\alpha$ -helices (Figure 4B) and limited proteolysis (Figure 4A) combined with mass spectrometry indicated that it forms a stable fragment between residues 1279-1405. This region, when expressed in *E. coli*, was highly soluble and folded correctly, as expected (Figure 4B). According to the CD spectra of both the 1279-1433 and 1279-1405 fragments, their secondary structures were stable up to 45°C and completely melted at 60°C (Supplementary Figure S6B). Therefore, we suggest that the disordered domain in the C-terminal region of Tho2 is located at the very end of the polypeptide chain (residues 1405-1597).

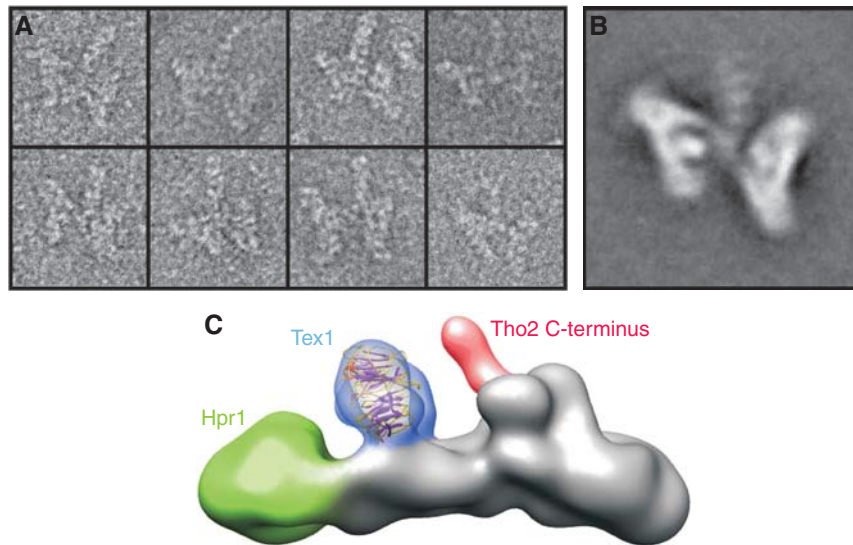
Taken together, the results presented above strongly suggest that the region of Tho2 responsible for the interaction with nucleic acids is partially disordered. Although the amino-acid sequence of this fragment is not evolutionary conserved, all the Tho2 sequences analysed contain a large number of basic residues, suggesting that its function may be preserved in other eukaryotes.

To further analyse the role of the Tho2 C-terminal region, we constructed yeast strains expressing two shortened,

TAP-tagged versions of Tho2: Tho2Δ<sub>1271-1597</sub> TAP, where the entire region was removed, and Tho2Δ<sub>1408-1597</sub> TAP, where only the unstructured part was deleted. When purified by ion exchange chromatography, the THO complexes with deletion mutants of Tho2 eluted from the ion exchange column at the same salt concentration as the full-length version (Supplementary Figure S1). Also, the composition of the complexes was unaltered as revealed by SDS-PAGE. These results strongly indicated that the C-terminal region of Tho2 is dispensable for complex assembly. However, the shortened versions of Tho2 were virtually unable to bind any types of nucleic acids (Figure 4C). In contrast, Tex1 appeared not to be involved in THO interaction with nucleic acids, as there was no difference between cross-linking of nucleic acids using the standard THO and the THOΔTex1 complex (Supplementary Figure S6C).

**The C-terminal region of the Tho2 protein is located at the tip of the narrow protrusion within the THO complex structure**

We then set out to locate the C-terminal region of Tho2 in the THO complex by performing three-dimensional reconstructions of the two complexes containing Tho2Δ<sub>1271-1597</sub> and Tho2Δ<sub>1408-1597</sub>. We expected to see a volume missing from the original structure of the complex; however, we observed



**Figure 5** Localization of the C-terminal region of Tho2. (A) A gallery of double DID-THO particles interacting through the C-terminal region of Tho2, as pointed by the DID–Dyn2 label bound tagged to this region. (B) A two-dimensional average image of these complexes. (C) Localization of Tex1, Hpr1, and the C-terminal domain of Tho2 within the THO complex.

no significant differences between the three-dimensional reconstructions of these two complexes compared with that of wild-type THO. This result further reinforces the notion that the C-terminal region of Tho2 is unstructured.

We thus adopted a different tagging strategy to locate Tho2: we replaced the unstructured C-terminal region (aa 1408–1597) with the tag developed by Flemming *et al* (2010). This tag consists of a dynein light chain-interacting domain (DID) composed of six dynein light chain (Dyn2) binding domains that can bind six Dyn2 homodimers, which must be supplied to the solution. The complex was reinforced by another DID protein that was added to the solution. The whole label, once formed, has a molecular mass of 130 kDa and a rod-like shape of  $\sim 250$  Å length. The THO was purified as described above with DID-tagged Tho2. The DID and Dyn2 proteins were then added to the THO solution and aliquots of this preparation were negatively stained and subjected to electron microscopy. To our surprise, whereas no DID–Dyn2 label was observed in the DID-THO particles, most of the butterfly like, double DID-THO particles present in the solution revealed the rod-like structure described for this label (Figure 5A). The DID-tags of the two particles seem to contribute to the formation of the DID–(Dyn2)<sub>6</sub>–DID heterodimer, and this is supported by the fact that the label always protrudes from the region connecting the two THO particles, clearly visible in the average image generated with 1715 particles (Figure 5B). This clearly points to the previously mentioned sticky, non-structured region as the C-terminal domain of Tho2 (Figure 5C).

#### **Truncation of Tho2 impairs gene expression**

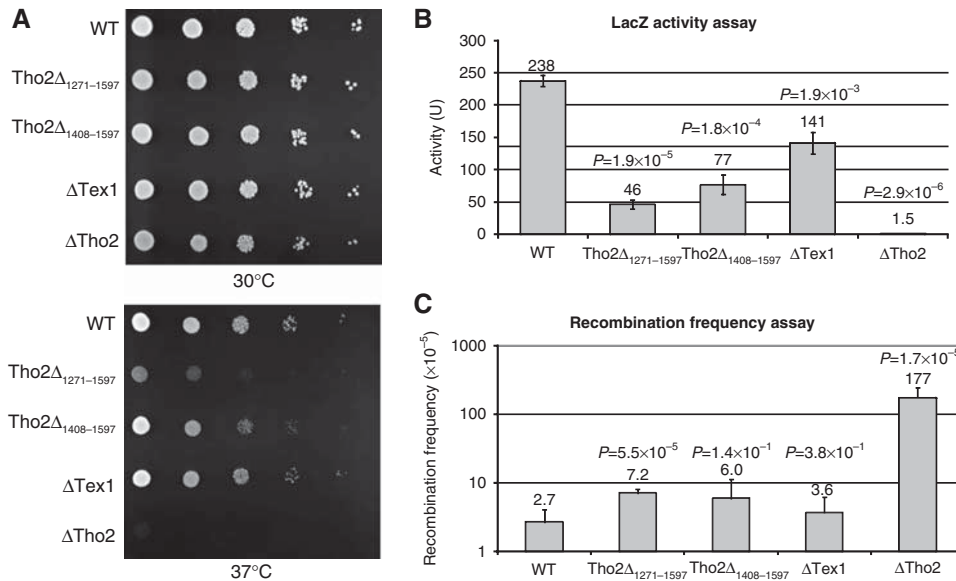
It has been previously described that depletion/knockout of the THO complex leads to inhibition of transcription elongation (Chavez and Aguilera, 1997) but also to an increase in recombination due to the formation of DNA–RNA hybrids (Garcia-Rubio *et al*, 2008). We tested *in vivo* whether THO lacking the C-terminal, nucleic acid binding domain of Tho2 leads to the phenotypes characteristic for *tho2Δ* strains.

All phenotype tests were performed for five isogenic yeast strains: WT, Tho2 $\Delta_{1408-1597}$ , Tho2 $\Delta_{1271-1597}$ ,  $\Delta$ Tex1, and  $\Delta$ Tho2.

Since strains lacking Tho2 or Hpr1 have been reported not to grow at 37°C (Piruat and Aguilera, 1998), we tested all strains for growth at both 30 and 37°C and observed that—while the growth of the  $\Delta$ Tex1 strain is not affected—the Tho2 $\Delta_{1408-1597}$  and Tho2 $\Delta_{1271-1597}$  strains are temperature sensitive (Figure 6A). At 37°C, the Tho2 $\Delta_{1408-1597}$  strain grows slightly slower than the wild-type strain but the Tho2 $\Delta_{1271-1597}$  strain barely grows at all. This suggests that the C-terminal region of the Tho2 protein encompassed between the amino acids 1271 and 1597 is important for cell survival at restrictive temperatures; Tex1 does not seem to be determinant for cell survival at 37°C.

Considering that expression of  $\beta$ -galactosidase is strongly inhibited in THO deletion strains (Chavez *et al*, 2001), we also tested the ability of the analysed strains to express exogenous  $\beta$ -galactosidase (Figure 6B). Remarkably,  $\beta$ -galactosidase activity was reduced five-fold in yeast with THO complex lacking the RNA/DNA-binding domain compared with the wild-type, whereas it was 500-fold less in  $\Delta$ Tho2 and two-fold in  $\Delta$ Tex1 cells. Interestingly, in contrast to the  $\Delta$ Tho2 strain for which inhibition of  $\beta$ -galactosidase expression correlated with strongly reduced mRNA steady-state levels, shortening Tho2 as well as Tex1 deletion did not decrease the mRNA abundance (data not shown).

The third phenotype we tested was transcription-associated recombination, which is highly increased in yeast with disrupted THO (Garcia-Rubio *et al*, 2008). Deletion of Tho2 resulted in a strong hyper-recombination phenotype (a 65-fold increase in recombination compared with wild-type), while only a modest phenotype ( $\sim 2$ -fold increase) was observed in Tho2 $\Delta_{1408-1597}$  and Tho2 $\Delta_{1271-1597}$ . Recombination measurements in the Tho2 $\Delta_{1408-1597}$  strain were too inconsistent to observe a statistically significant difference ( $P < 0.05$ ) between this strain and the wild-type one. No hyper-recombination was observed in the Tex1 $\Delta$



**Figure 6** Tho2 shortening increases recombination and decrease  $\beta$ -galactosidase expression. Phenotypes of five different *S. cerevisiae* strains were analysed: wild-type (1), strains with shortened Tho2 protein: Tho2 $\Delta_{1271-1597}$  (2) and Tho2 $\Delta_{1408-1597}$  (3),  $\Delta$ Tex1 (4), and  $\Delta$ Tho2 (5) strains. (A) Growing test on YPD plates on 30 and 37°C. (B)  $\beta$ -Galactosidase activity test. Yeast strains with plasmid located lacZ gene under galactose-induced promoter were cultured for 2 h at 37°C on YP medium with 2% galactose.  $\beta$ -Galactosidase activity was measured with the use of ONPG. (C) Recombination analysis. Recombination frequencies were calculated as described in Garcia-Rubio *et al* (2008) using the L-PHO5 plasmid. Experiments were repeated three times and respective *P*-values are presented.

strain (Figure 6C), consistent with previous data (Luna *et al*, 2005). These observations suggest that the C-terminal region of Tho2 is not essential in the maintenance of genetic stability.

In all the tests performed, the difference between wild-type and *tho2*Δ strains was much higher than between wild-type and Tho2-truncated strains. This is explained by the fact that shortening of the Tho2 protein does not affect THO complex formation, whereas deletion of any of the THO subunits completely prevents it: purification of the complex from strains missing any of its subunits (data not shown) proved impossible, consistent with the observation that removal of any THO subunit caused a destabilization of the other components, whereas this was not the case for the Sub2 component of TREX whose deletion did not affect the stability of the four THO subunits (Huertas *et al*, 2006).

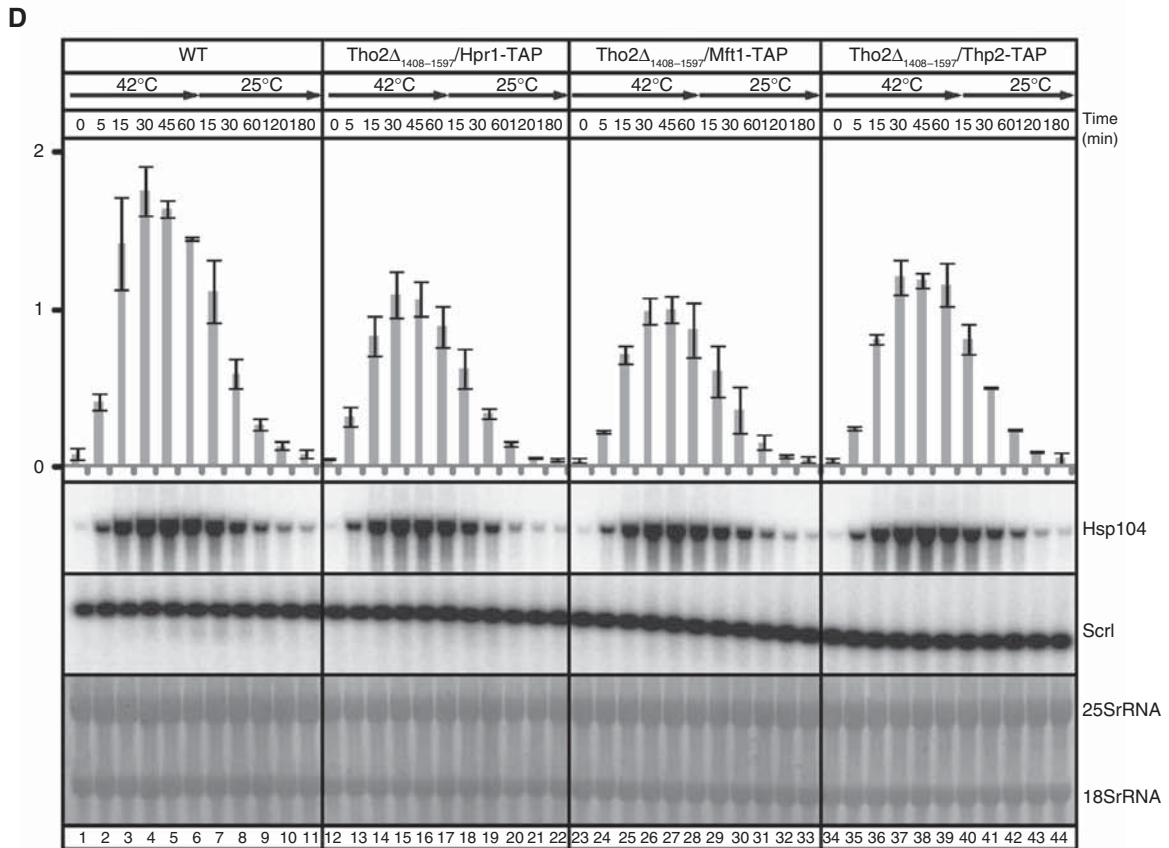
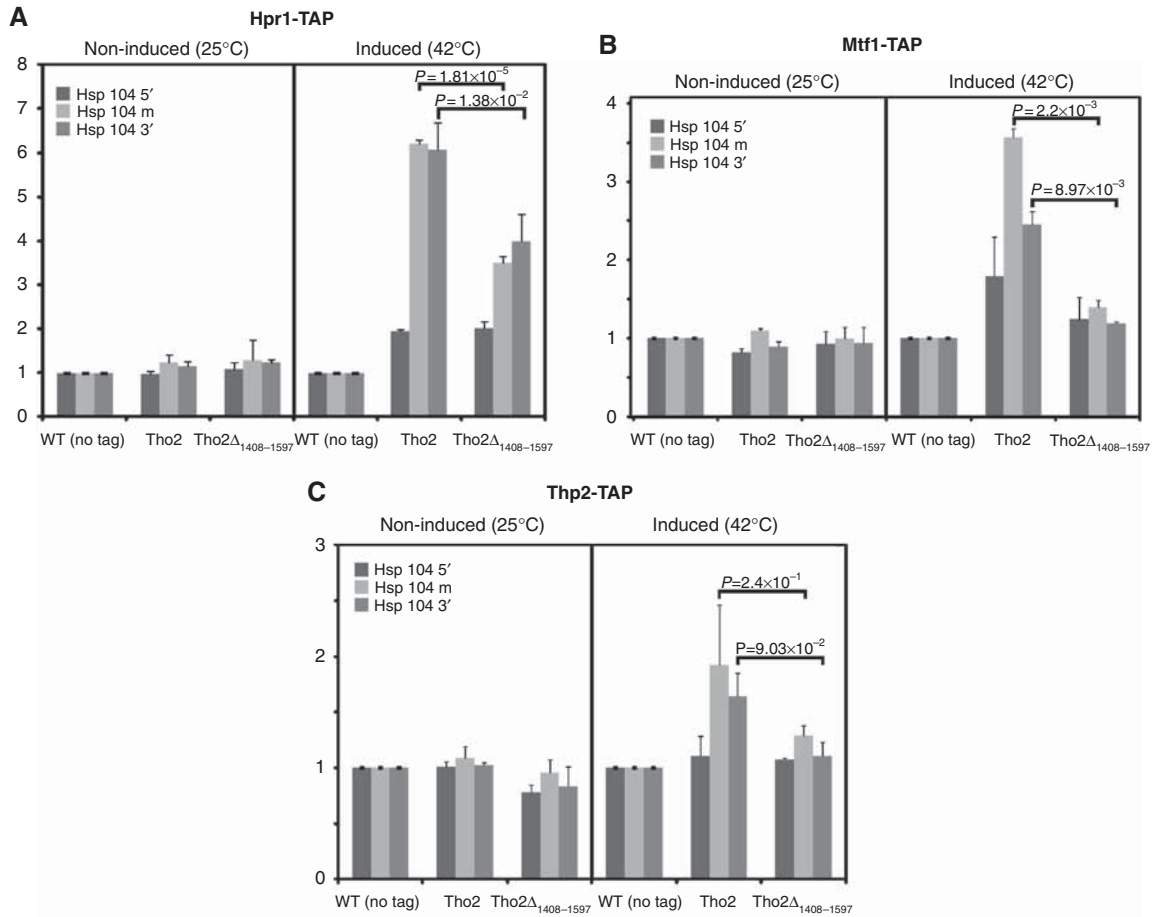
**The nucleic acid interacting domain of Tho2 is involved in the association of THO with active chromatin**

We hypothesized that the newly discovered nucleic acid interacting domain of Tho2 may be involved in the recruitment of the THO complex to chromatin during transcription. To assess this, we analysed the efficiency of THO recruitment to transcriptionally active genes by chromatin immunoprecipitation (ChIP) experiments using the Tho2 $\Delta_{1408-1597}$  mutant, in which the unfolded region of the protein responsible for nucleic acid recognition was removed. To exclude the possibility that a tag attached to the RNA-binding region of Tho2 may produce a steric hindrance, a TAP-tag was placed at the C-terminus of other subunits of the THO complex, that is, Hpr1, Thp2, and Mft1, in both the wild-type and Tho2 $\Delta_{1408-1597}$  isogenic strains. In all cases, the tagged subunits were expressed at comparable levels, as revealed by quantitative western analysis (data not shown). This is in

agreement with observations that the assembly of the THO complex is not affected by introducing a tag to any of the subunits (results not shown).

To analyse the transcription-dependent recruitment of THO, we used a stress-induced *HSP104* gene as a model system. We confirmed by northern analysis that *HSP104* transcription was activated upon heat shock and by ChIP that RNAPII associated with DNA along the gene (data not shown). Moreover, steady-state mRNA levels and polymerase occupancy along the *HSP104* gene did not change significantly after Tho2 protein shortening. In agreement with previous studies, induction of *HSP104* expression resulted in association of the THO complex subunits along this gene (Figure 7) (Abruzzi *et al*, 2004). In addition, as already reported, ChIP combined with RNase treatment revealed that association of Hpr1 at *HSP104* is RNA independent (Abruzzi *et al*, 2004) (Supplementary Figure S7). Furthermore, all the THO subunits were recruited mostly to the middle and 3' regions of the *HSP104* gene (Figure 7), consistent with a recent genome-wide analysis (Gomez-Gonzalez *et al*, 2011). The level of DNA enrichment varied depending on the tagged gene and was highest in the Hpr1-TAP strain (Figure 7A), but in all cases THO recruitment was reduced by deletion of the Tho2 C-terminal region (Figure 7). We conclude that the Tho2 C-terminal, nucleic acid binding region facilitates recruitment of the THO complex to chromatin.

To determine the effect of the reduced levels of THO recruitment on *HSP104* expression, we performed northern blot quantifications of the kinetics of *HSP104* induction and its decay after shifting cells to non-restrictive temperature. In agreement with the ChIP experiments, it could clearly be seen that Tho2 shortening lowers the induction level but does not change the kinetic of mRNA appearance and decay (Figure 7D).





## Discussion

It has been established that THO plays a central role in co-transcriptional formation of export-competent mRNP molecules, but the mechanism of its function is still unknown. In the present study, we have performed a structural and biochemical characterization of this macromolecular assembly, which has provided the first three-dimensional structure of the complex and relevant mechanistic insight into its function.

### The THO complex architecture

The *S. cerevisiae* THO complex has been classically described as a four-subunits complex (Tho2, Hpr1, Tho2, and Mft1), capable of interacting with a plethora of proteins, among others Tex1, Yra1, and Sub2, with which it forms the TREX complex ((Strasser *et al*, 2002; Hurt *et al*, 2004). However, during our purification of THO, we reproducibly co-purify Tex1, even under the stringent conditions, in contrast to the other TREX components, Yra1 and Sub2, which dissociate during the purification procedure (Figure 1). We decided to describe THO as the five-subunits complex.

Using electron microscopy and image processing techniques, we have determined the three-dimensional structure of both THO and THO $\Delta$ Tex1 complexes (Figures 1 and 2). Both complexes have a long, croissant-like shape structure with a flat surface on one side and a more corrugated one on the other. On the latter surface, whereas THO $\Delta$ Tex1 has a single, thin protrusion, THO reveals two protrusions, the same thin one observed in the THO $\Delta$ Tex1 complex, and a wider one, which we assign to Tex1 (Figures 2 and 5C). Fold recognition analysis clearly points to Tex1 as a WD40, seven-blade  $\beta$ -propeller; the atomic model we generated assumes such a structure and docks into the volume assigned to Tex1 in the THO structure (Figure 2D–F). Tex1 is one of the most evolutionarily conserved subunits of the THO complex, yet the phenotypes caused by its deletion are non-existing or barely detectable in yeast (Luna *et al*, 2005) (Figure 6). We observed that the lack of Tex1 has no effect on the THO complex assembly and binding to nucleic acids. Furthermore, *tex1* $\Delta$  strains exhibit only a very mild decrease of LacZ expression and no increase in recombination rate, in agreement with previous results (Luna *et al*, 2005). This is consistent with the observation that removal of any subunit of the core THO complex causes strong gene expression and recombination phenotypes that is accompanied by destabilization of the rest of THO components in the cell (Huertas *et al*, 2006) and an impossibility to purify the complex (data not shown). Consistently, Tex1 is irrelevant for THO integrity and structure. This is in contrast to other systems such as plants, in which the Tex1 orthologue is essential for the synthesis of one class of endogenous miRNA/siRNAs, tasiRNA (Jauvion *et al*, 2010; Yelina *et al*, 2010). We speculate that the reason for the small impact of Tex1 on RNA

metabolism in yeast is linked to the lack of the RNAi machinery (Jauvion *et al*, 2010; Yelina *et al*, 2010). WD40 proteins have classically been involved in protein–protein interaction (Smith *et al*, 1999), so it cannot be ruled out that Tex1 plays a role in the interaction of THO with other proteins. This is in agreement with previously data that places THO as a central player in the transcription process that interacts with many factors coupling the mRNA processing, packaging and export (Strasser *et al*, 2002; Luna *et al*, 2005; Masuda *et al*, 2005).

We have mapped the position of the C-terminal fragment of Hpr1 (Figures 2G and 5C), the most evolutionary conserved and best functionally characterized THO subunit (Gwizdek *et al*, 2006; Hobeika *et al*, 2007, 2009; Iglesias *et al*, 2010), into the THO structure. The original hyper-recombinant *hpr1-1* mutation isolated encodes indeed a truncated Hpr1 protein lacking the C-terminal 559–752 fragment, clearly indicating the *in vivo* biological relevance of the Hpr1 C-terminus (Aguilera and Klein, 1990). This region (amino acids 548–752) (Figures 2G and 5C) is ubiquitinated in a transcription-dependent manner and the ubiquitin moiety is recognized by the C-terminal UBA domain of the export receptor Mex67. Structural studies have indicated that ubiquitinated Hpr1 and the NPC subunits, FG nucleoporins (Nups), may bind to the Mex67 UBA in a mutually exclusive manner (Iglesias *et al*, 2010). This suggests that the Mex67 export factor may be recruited to the nascent transcript via ubiquitinated Hpr1 and then transferred to nuclear pore FG Nups. Further structural studies involving some of the components of the export machinery mentioned above would allow a better understanding of the mRNA export pathway.

The C-terminal, nucleotide acid binding region of Tho2 has been localized at the tip of the narrow protrusion of the THO complex; thanks to DID–Dyn2 electron microscopy labelling experiment. This region shares characteristics of intrinsically unstructured proteins (Uversky and Dunker, 2010), consistent with the following: first, the volume of this region is not visible by electron microscopy, as the three-dimensional reconstructions obtained for the THO complex with either full-length Tho2 or several C-terminal deletion variants are essentially identical; second, the recombinant soluble region of this fragment is very sensitive to protease digestion; third, the CD spectra acquired with fragments of this region reveal a large content of disordered polypeptide. All these evidences are in agreement with secondary structure predictions that point to a high level of disorder for this part of the protein (Supplementary Figure S6D).

### Interaction between the THO complex and the nucleic acids: chromatin recruitment

Combination of RNA/DNA–protein cross-linking experiments with site-specific proteolytic digestion reveals that THO complex directly interacts with nucleic acids through the C-terminal, unstructured fragment of Tho2. Interestingly,

**Figure 7** Association of THO to transcriptionally active chromatin depends significantly on the presence of Tho2 C-terminal domain. ChIP analysis of Hpr1-TAP (A), Mft1-TAP (B), and Thp2-TAP (C) protein binding to the heat-shock-induced Hsp104 gene in wt and Tho2 $\Delta_{1408-1597}$  strains. ChIP values were calculated as described in Materials and methods. Graphs show the average value of three independent experiments with standard deviation. Respective *P*-values are presented. (D) Northern analysis of Hsp104 mRNA during heat-shock induction (60 min) and after shifting into non-restrictive temperature (180 min) at the indicated time points. Graph shows arbitrary values of the Hsp104 mRNA levels normalized to Scrl RNA. The average and standard deviation values of two independent experiments are shown.

deletion of the nucleic acid binding domain diminished recruitment of the THO/TREX complex to chromatin, strongly suggesting that this naturally unfolded region plays an important role in the chromatin recruitment process. Since it was demonstrated that THO deletion severely impairs association of the export factors Sub2 and Yra1 mRNA with active chromatin (Zenklusen *et al*, 2002), our results shed light into the mechanism of co-transcriptional assembly of export-competent mRNPs. The nucleic acid binding properties of the Tho2 C-terminus suggests that this region helps recruit chromatin to the THO complex by interaction with DNA or both DNA and the nascent RNA. Interactions with the RNA are however not crucial for THO recruitment, since our data, in agreement with previous studies, show that THO ChIP is not sensitive to RNase treatment (Abruzzi *et al*, 2004).

The phenotype generated by deletion of the Tho2 C-terminus demonstrates that chromatin recruitment of the THO complex facilitates expression of some of the target genes. However, deletion of whole Tho2 generates a more severe phenotype leading to strong inhibition of transcription elongation for the subset of genes. Our data suggest that Tho2 C-terminus is important for gene expression but the effect on transcription may vary depending on the target gene. In addition, since THO association with chromatin is not completely abolished by the Tho2 C-terminus deletion, alternative pathways of THO recruitment must exist. One possibility is that other THO components and not only Tho2 contribute to chromatin recruitment. Another possibility is that recruitment could be mediated by other proteins. One candidate would be Syf1, a component of the Prp19 splicing complex that has been proposed to be involved in the THO complex recruitment (Chanarat *et al*, 2011). However, when Syf1 is mutated, about 50% of normal level of TREX is bound to chromatin, sufficient for proper mRNA export (Chanarat *et al*, 2011).

Further experimental evidence is needed to fully explain the mechanism of recruitment of the THO complex to chromatin and the way it discriminates between silent and transcriptionally active states. Nevertheless, our data support a model in which THO complex associates with actively transcribed chromatin and thus provides a signal to Sub2 and Yra1 to assemble mRNP and proceed with mRNA export.

## Materials and methods

### Yeast strains

All *S. cerevisiae* strains used in this study (see Supplementary Table S1) were derived from MGD453-13D (*MATa ade2 arg4 leu2-3,112 trp1-289 ura3-52*) with one exception—Tho2 $\Delta_{1408-1597}$ . Did1TAP strain was derived from  $\Delta$ Dyn2 Ds1-2b strain kindly provide from Ed Hurt's laboratory (*MAT $\alpha$  leu2- $\Delta$ 1 trp1- $\Delta$ 63 his3- $\Delta$ 200 ura3-52 dyn2::kanMX4*). All mutant strains were constructed via the standard method—PCR products were introduced into the host strain via lithium acetate transformation, followed by selection and confirmation by PCR and sequencing.

### THO complex purification

All THO purifications were performed through two chromatographic steps, an IgG sepharose beads and ion exchange column with AKTA purifier FPLC (see Supplementary data for details).

### RNA/DNA UV cross-linking assay

Two types of radiolabelled RNA were used: *in vitro* transcribed Lsm1 cDNA fragment from *Arabidopsis thaliana* (about 400 nt long)

or synthetic 44 nt long oligonucleotide. A synthetic 44 nt oligonucleotide was used as DNA template (see Supplementary data for sequences). Purified THO complex (about 2  $\mu$ g) was incubated for 5 min with hot RNA/DNA (5 pmol) at 30°C in buffer GET (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% glycerol, 0.1 mM EDTA). Cross-linking was carried out for 1 min with the UVC cross-linker (Stratagene). When 400 nt long RNA was used, RNase A was added for 1 h to digest excess of RNA. After cross-linking, proteins were precipitated and separated by SDS-PAGE. Gels were stained with Coomassie blue for protein visualization and radiolabelled RNA/DNA were visualized by autoradiography.

### $\beta$ -Galactosidase assay

$\beta$ -Galactosidase activity was analysed and calculated as described (see also Supplementary data and Gietz *et al*, 1997). Experiments were repeated three times and the statistical significance was calculated using *t*-test and ANOVA (analysis of variance) test and calculated *P*-values are presented.

### Recombination analysis

Recombination frequencies were analysed and calculated as described in Garcia-Rubio *et al* (2008) using the L-PHO5 plasmid, carrying a *leu2* direct repeat. For each genotype, the recombination frequencies were calculated based on at least three independent transformations using a minimum of three colonies from each transformation. Statistical significance was measured as described above.

### Chromatin immunoprecipitation

Heat-shock inductions were performed as follows: Cells were grown in a YPD at 25°C to a late exponential phase ( $OD_{600} \sim 0.6-0.7$ ). Half of the cultures were quickly centrifuged and suspended immediately in a medium preheated to 42°C, and incubated in water bath for 30 min. The rest of the batch was immediately cross-linked and then used as 'non-induced' control. Cross-linking, ChIP, and qPCR was performed as described (El Hage *et al*, 2008) with a few modifications (see Supplementary data for details). Experiments were repeated at least three times and statistical significance was measured as described above.

### Northern blots

All details are presented in the Supplementary data.

### Mass spectrometry

Bands of interest were excised from coomassie-stained gels and digested as described in Shevchenko *et al* (1996) with minor variations. MALDI-TOF/TOF (time of flight) analysis, were automatically acquired in an ABI 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems) in positive ion reflector mode. Data were analysed using MASCOT software v.2.2.04 (Matrix Science). See Supplementary data for details.

### Expression and analysis of Tho2 fragments

In all, 44 constructs from C-terminal domain of Tho2 were generated and expressed in *E. coli* BL21. Overexpressed proteins were purified on Ni-NTA column (GE Healthcare) and by gel filtration on Superdex 75 (GE Healthcare) on AKTA purifier FPLC. For cloning and purification details see Supplementary data.

Limited proteolysis was carried out for peptides Tho2 $_{1279-1433}$  and Tho2 $_{1411-1530}$ . A variable factor were the trypsin:peptide molar ratios used (1:75 000, 1:15 000, 1:3000). Reactions were performed in 20 mM phosphate buffer pH 7.4 and 150 mM NaF for 30 min at room temperature and stopped with 1 mM PMSF. After proteolysis, degradation products were visualized by 16% Tricine SDS-PAGE and Coomassie staining.

Excised bands were analysed by electrospray TOF mass spectrometry (TOF MS ESI+) and Edman sequencing. For the MS analysis, trypsin:peptide 11:1 ratio and control without protease were used. To each sample, 1 mM PMSF final concentration was added after half an hour to stop the proteolysis reaction. For Edman sequencing, after 16% Tricine SDS-PAGE, peptides were transferred to membrane and analysed using the standard procedure.

### Circular dichroism

In all, 2–6  $\mu$ M solutions of Tho2 $_{1279-1433}$ , Tho2 $_{1279-1433}$ , and Tho2 $_{1279-1404}$  proteins were prepared in 150 mM NaF, mM phosphate buffer pH 7.4. The CD spectra were collected in 0.2 cm

quartz cuvettes in a JASCO J-815 CD Spectropolarimeter in the range of 270–190 nm with a data pitch of 1.0 nm. The bandwidth was set to 1.0 nm and digital integration time to 2 s. CD data were collected in three accumulations in constant temperature (25, 35, 45, 60, or 90°C) controlled by PTC-423S single position Peltier. Protein concentrations were corrected by extinction coefficient ( $\epsilon_{\lambda 280}$ ) calculated with use of ProtParam. The percentages of the different secondary structures were estimated by CDNN program (Bohm *et al*, 1992).

### Electron microscopy and image processing

Samples (either THO, THO $\Delta$ Tex1, or the labelled complexes) were applied onto carbon-coated copper grids previously glow-discharged and stained with 2% uranyl acetate. Micrographs were taken under minimal dose conditions on Kodak SO-163 film, in a JEOL JEM1200EXII microscope with a tungsten filament operated at 100 kV and  $\times 60\,000$  magnification. Micrographs were digitized in a Zeiss SCAI scanner with a sampling window corresponding to 2.33 Å/pixel for all the specimens. Individual particles were manually selected using XMIPP software package (Marabini *et al*, 1996). Image classification was performed using a free-pattern maximum-likelihood multi-reference refinement (Scheres *et al*, 2005). When appropriate, the particles were subjected to Kohonen's self-organizing feature maps (Marabini and Carazo, 1994). Homogeneous populations were obtained and averaged to final 2D characterization. For 3D reconstruction, reference models and first refinements steps were performed using the EMAN software package (Ludtke *et al*, 1999), until a volume with the general shape of the complex became evident. Three different volumes were used initially, one generated from a common line approach, another from noise, and a third one from a Gaussian blob, and the three rendered similar results (Supplementary Figures S3 and S4). The XMIPP software package (Scheres *et al*, 2008) was used in the subsequent iterative angular refinement procedure. The resolution of the reconstructions was determined by the FSC 0.5 criterion for the Fourier shell correlation coefficient between two independent reconstructions (Supplementary Figure S5) and these values were used to low-pass filter the final volumes. The handedness of the reconstructed volumes was chosen arbitrarily because of the intrinsic ambiguity generated by the electron microscopy reconstruction procedure.

### Generation and docking of the atomic model of Tex1 into the THO complex

The structure prediction was carried out in using a threading procedure. Tex1 from different organisms always mapped to seven-sword  $\beta$ -propeller proteins. Protein alignments to WDR5 protein (pdb2g9a) was prepared and validated using 3D-Jury (Ginalski *et al*, 2003). The dimensional model of Tex1 (residues 47–371) was generated with Modeller (Sali *et al*, 1995) based on manually curated, high confidence sequence-to-structure alignments. Docking of the atomic model of Tex1 into the three-dimensional reconstruction of the THO complex was carried out manually and optimized using COLACOR, an off-lattice correlation maximizer distributed with Situs 2.2, based on the local optimization of COLORES (Chacón and Wrighers, 2002).

### Localization of Hpr1 subunit into the THO three-dimensional structure

The localization of Hpr1 into the THO volume was carried out by immunomicroscopy. Antibodies were designed against the

C-terminal region of Hpr1 attending to the fact that it is ubiquitinated during the transcription elongation process and that it also interacts with the export factor Mex67 *in vivo* (Abruzzi *et al*, 2004; Hobeika *et al*, 2007; Gomez-Gonzalez *et al*, 2011), thus suggesting that it is exposed to the surface of the complex. Different epitopes were selected but only the antibody with an epitope composed by the sequence LQDAREYKIGKERRKRA (positions between the residues 636–653 of the Hpr1 sequence) could finally form a stable complex with THO. The antibody was produced by Pacific Immunology Company in rabbit. Specificity of the anti-Hpr1 antibody was tested by western and dot blot analysis. Aliquots of the immunocomplex were negatively stained in carbon-coated grids and a total of 8250 particles were selected and processed.

### Localization of C-terminal domain of Tho2 subunit into the THO three-dimensional structure

To localize the RNA/DNA-binding domain of Tho2, a recently described tagging method using a complex composed of Dyn2 and DID (Flemming *et al*, 2010) was used. We prepared a protein fusion in yeast strain with a Dyn2 deletion: Tho2 $\Delta_{408-1597}$ -DID1-TEV-ProteinA. The THO complex from this strain was purified on IgG sepharose and incubated 2 h at 4°C with heterologously expressed DID2 and Dyn2 proteins. The final purification step was performed on a Resource Q ion exchange column. Yeast strain with Dyn2 deletion and plasmids for DID2, Dyn2 expression were kindly provided by Dr Ed Hurt.

### Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

## Acknowledgements

We thank Joanna Kufel, Aleksander Chlebowski for critical comments on the manuscript and members of the AD and JMV laboratories for stimulating discussions. Krystian Stodue for help with protein purifications and Grażyna Goch for advise on CD analysis. This work was supported by the Foundation for Polish Science Team Programme co-financed by the EU European Regional Development Fund, EMBO installation grant (to AD) and the Spanish Ministry of Science and Innovation Grants BFU2010-15703/BMC (to JMV) and BFU2006-05260 (to AA). This work was also funded by the EU-Grant '3D repertoire'.

*Author contributions:* KG performed all biochemical experiments on the native complex and constructed all the yeast strains. AP carried out most of the electron microscopy and image processing. JC and JLC carried out the labelling experiments with DID–Dyn2. SM performed ChIP experiments and northern blots. AS generated and analysed expression constructs. APr analysed yeast phenotypes. MC performed CD analysis. CT and AA generated the antibody against Hpr1 and provided the recombination assays. JP developed THO complex purification procedure. AD and JMV conceived the project, supervised the experiments and, with contribution from all the authors, wrote this paper.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Abruzzi KC, Lacadie S, Rosbash M (2004) Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J* **23**: 2620–2631
- Aguilera A, Klein HL (1990) HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the *Saccharomyces cerevisiae* TOP1 gene. *Mol Cell Biol* **10**: 1439–1451
- Bohm G, Muhr R, Jaenicke R (1992) Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng* **5**: 191–195
- Chacón P, Wrighers W (2002) Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol* **317**: 375–384
- Chanarat S, Seizl M, Strasser K (2011) The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev* **25**: 1147–1158
- Chavez S, Aguilera A (1997) The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes Dev* **11**: 3459–3470
- Chavez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Sveistrup JQ, Lithgow T, Aguilera A (2000) A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *EMBO J* **19**: 5824–5834

- Chavez S, Garcia-Rubio M, Prado F, Aguilera A (2001) Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol* **21**: 7054–7064
- Dziembowski A, Seraphin B (2008) Sample preparation for protein complex analysis by the TAP method. In: *Sample Preparation for Proteomics*, vonHagen J (ed). Weinheim: VCH Wiley
- El Hage A, Koper M, Kufel J, Tollervey D (2008) Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes Dev* **22**: 1069–1081
- Flemming D, Thierbach K, Stelter P, Bottcher B, Hurt E (2010) Precise mapping of subunits in multiprotein complexes by a versatile electron microscopy label. *Nat Struct Mol Biol* **17**: 775–778
- Garcia-Rubio M, Chavez S, Huertas P, Tous C, Jimeno S, Luna R, Aguilera A (2008) Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Mol Genet Genomics* **279**: 123–132
- Gietz RD, Triggs-Raine B, Robbins A, Graham KC, Woods RA (1997) Identification of proteins that interact with a protein of interest: applications of the yeast two-hybrid system. *Mol Cell Biochem* **172**: 67–79
- Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **19**: 1015–1018
- Gomez-Gonzalez B, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marin A, Foiani M, Aguilera A (2011) Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* **30**: 3106–3119
- Gwizdek C, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G, Stutz F, Dargemont C (2006) Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proc Natl Acad Sci USA* **103**: 16376–16381
- Hobeika M, Brockmann C, Gruessing F, Neuhaus D, Divita G, Stewart M, Dargemont C (2009) Structural requirements for the ubiquitin-associated domain of the mRNA export factor Mex67 to bind its specific targets, the transcription elongation THO complex component Hpr1 and nucleoporin FXFG repeats. *J Biol Chem* **284**: 17575–17583
- Hobeika M, Brockmann C, Iglesias N, Gwizdek C, Neuhaus D, Stutz F, Stewart M, Divita G, Dargemont C (2007) Coordination of Hpr1 and ubiquitin binding by the UBA domain of the mRNA export factor Mex67. *Mol Biol Cell* **18**: 2561–2568
- Houseley J, LaCava J, Tollervey D (2006) RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* **7**: 529–539
- Huertas P, Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**: 711–721
- Huertas P, Garcia-Rubio ML, Wellinger RE, Luna R, Aguilera A (2006) An hpr1 point mutation that impairs transcription and mRNP biogenesis without increasing recombination. *Mol Cell Biol* **26**: 7451–7465
- Hurt E, Luo MJ, Rother S, Reed R, Strasser K (2004) Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. *Proc Natl Acad Sci USA* **101**: 1858–1862
- Iglesias N, Tutucci E, Gwizdek C, Vinciguerra P, Von Dach E, Corbett AH, Dargemont C, Stutz F (2010) Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes Dev* **24**: 1927–1938
- Jauvion V, Elmayan T, Vaucheret H (2010) The conserved RNA trafficking proteins HPR1 and TEX1 are involved in the production of endogenous and exogenous small interfering RNA in *Arabidopsis*. *Plant Cell* **22**: 2697–2709
- Jimeno S, Rondon AG, Luna R, Aguilera A (2002) The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J* **21**: 3526–3535
- Kohler A, Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* **8**: 761–773
- Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**: 82–97
- Luna R, Jimeno S, Marin M, Huertas P, Garcia-Rubio M, Aguilera A (2005) Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol Cell* **18**: 711–722
- Marabini R, Carazo JM (1994) Pattern recognition and classification of images of biological macromolecules using artificial neural networks. *Biophys J* **66**: 1804–1814
- Marabini R, Masegosa IM, San Martin MC, Marco S, Fernandez JJ, de la Fraga LG, Vaquerizo C, Carazo JM (1996) Xmipp: An image processing package for electron microscopy. *J Struct Biol* **116**: 237–240
- Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R (2005) Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* **19**: 1512–1517
- Piruat JI, Aguilera A (1998) A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *EMBO J* **17**: 4859–4872
- Reed R, Cheng H (2005) TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol* **17**: 269–273
- Rehwinkel J, Herold A, Gari K, Kocher T, Rode M, Ciccarelli FL, Wilm M, Izaurralde E (2004) Genome-wide analysis of mRNAs regulated by the THO complex in *Drosophila melanogaster*. *Nat Struct Mol Biol* **11**: 558–566
- Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F, Libri D (2008) THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association. *Cell* **135**: 308–321
- Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**: 318–326
- Scheres SH, Nunez-Ramirez R, Sorzano CO, Carazo JM, Marabini R (2008) Image processing for electron microscopy single-particle analysis using XMIPP. *Nat Protoc* **3**: 977–990
- Scheres SH, Valle M, Nunez R, Sorzano CO, Marabini R, Herman GT, Carazo JM (2005) Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol* **348**: 139–149
- Shevchenko A, Wilm M, Vorm O, Jensen ON, Podtelejnikov AV, Neubauer G, Mortensen P, Mann M (1996) A strategy for identifying gel-separated proteins in sequence databases by MS alone. *Biochem Society Trans* **24**: 893–896
- Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* **24**: 181–185
- Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, Hurt E (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**: 304–308
- Uversky VN, Dunker AK (2010) Understanding protein non-folding. *Biochim Biophys Acta* **1804**: 1231–1264
- Yelina NE, Smith LM, Jones AM, Patel K, Kelly KA, Baulcombe DC (2010) Putative *Arabidopsis* THO/TREX mRNA export complex is involved in transgene and endogenous siRNA biosynthesis. *Proc Natl Acad Sci USA* **107**: 13948–13953
- Zenklusen D, Vinciguerra P, Wyss JC, Stutz F (2002) Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol Cell Biol* **22**: 8241–8253