

Structure of the Sac3 RNA-binding M-region in the *Saccharomyces cerevisiae* TREX-2 complex

James M. B. Gordon, Shintaro Aibara and Murray Stewart*

MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK

Received December 30, 2016; Revised February 23, 2017; Editorial Decision February 24, 2017; Accepted March 01, 2017

ABSTRACT

Transcription-export complex 2 (TREX-2, or THSC) facilitates localization of actively transcribing genes such as *GAL1* to the nuclear periphery, contributes to the generation of export-competent mRNPs and influences gene expression through interactions with Mediator. TREX-2 is based on a Sac3 scaffold to which Thp1, Sem1, Cdc31 and Sus1 bind and consists of three modules: the N-region (Sac3^{~1-100}), which binds mRNA export factor Mex67:Mtr2; the M-region, in which Thp1 and Sem1 bind to Sac3^{~100-550}; and the CID region in which Cdc31 and two Sus1 chains bind to Sac3^{~720-805}. Although the M-region of Sac3 was originally thought to encompass residues ~250-550, we report here the 2.3Å resolution crystal structure of a complex containing Sac3 residues 60–550 that indicates that the TPR-like repeats of the M-region extend to residue 137 and that residues 90–125 form a novel loop that links Sac3 to Thp1. These new structural elements are important for growth and mRNA export *in vivo*. Although deleting Sac3 residues 1–90 produced a wild-type phenotype, deletion of the loop as well generated growth defects at 37°C, whereas the deletion of residues 1–250 impaired mRNA export and also generated longer lag times when glucose or raffinose was replaced by galactose as the carbon source.

INTRODUCTION

The nuclear export of mature mRNAs through nuclear pore complexes (NPCs) is the culmination of the nuclear phase of the gene expression pathway. However, although the structure of NPCs has been established (1) and the Ran-based nuclear export and import of proteins by karyopherins is understood in considerable detail (2–4), mRNA export is more complex because of the necessity to first complete nuclear processing steps (such as splicing, polyadenylation,

capping, etc.) and the Mex67:Mtr2 export factor is employed rather than a karyopherin (5,6). The generation of export-competent mRNPs is therefore a critical step in the gene expression pathway. Moreover, many of the nuclear steps of the gene expression pathway are coordinated, often near nuclear pores using the TREX-2 complex (6,7).

The TREX-2 complex is conserved across eukaryotes and integrates mRNA export into the gene expression pathway and, in yeast, also mediates the location of actively-expressed genes such as *GAL1* to the nuclear envelope (8–14). In *Saccharomyces cerevisiae*, TREX-2 is based on a Sac3 scaffold to which Thp1, Sem1, Cdc31 and two copies of Sus1 bind (Figure 1) and broadly speaking, the TREX-2 complex can be subdivided into three regions (7): the N-region (Sac3 residues ~1–100), which harbours degenerate FG-like repeats similar to those seen in many nuclear pore proteins (FG nucleoporins) and which bind to the principal mRNA nuclear export factor, Mex67:Mtr2; the M- or proteasome/CSN/eIF3 (PCI)-region, consisting of Sac3 residues ~100–551 bound to Thp1 and Sem1, which forms a nucleic acid binding module as well as docking site for components of the Mediator complex (7,15–16); and the Cdc31-interacting domain (CID)-region, in which Sac3 residues ~720–805 bind to Cdc31 and two Sus1 chains and which, in *S. cerevisiae*, binds to the NPC to tether the complex close to the nuclear basket to facilitate localization of genes such as *GAL1* (17,18). Both the portion of Sac3 in the M-region and Thp1 have similar structures based on a PCI fold (19) in which a series of α -helical tetratricopeptide (TPR)-like repeats are followed by a winged helix motif. The juxtaposition of the winged helix motifs of Sac3 and Thp1, together with a surface patch of positive charge, contributes to the formation of a RNA-binding region in the TREX-2 complex (15,16).

The TREX-2 complex has been implicated in a spectrum of biological roles including preventing genome instability, transcription initiation, gene-gating and mRNA export (15,16,20–22). Previous work using homologues from the thermophile *Chaetomium thermophilum* indicated that the multiple FG-like repeats present on the Sac3 N-terminal region may serve as a docking site for the principal mRNA ex-

*To whom correspondence should be addressed. Tel: +44 1223 267074; Fax: +44 1223 8300; Email: ms@mrc-lmb.cam.ac.uk
Present address: Shintaro Aibara, Science for Life Laboratory, Tomtebodavägen 23A, SE-171 65 Solna, Sweden.

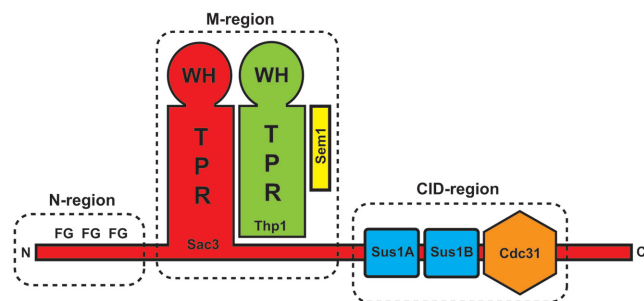


Figure 1. Schematic representation of the TREX-2 complex that is formed around a Sac3 scaffold (red). The TREX-2 complex can be broadly subdivided into three regions. The most N-terminal region of Sac3 (N-region) and its homologues (such as GANP from *Homo sapiens*) contain FG-nucleoporin-like motifs (7). The middle region (M-region, also called PCI-region), in which Sac3 (red) forms a complex with Thp1 (green) and Sem1 (yellow). In this region both Sac3 and Thp1 have PCI folds based on a winged helix domain (WH) and a stack of helical TPR-like motifs (15,16). The juxtaposition of the winged helix domains together with an extended patch of positive charge forms a nucleic acid binding platform (16) and this region also interacts with the Mediator complex (15). In the CID-region, residues ~720–805 of Sac3 interact with one Cdc31 chain (orange) and two Sus1 chains (blue) in *Saccharomyces cerevisiae* and mediates binding to NPCs (17,18).

port factor Mex67:Mtr2 as well as making intra-molecular interactions with the CID-region to form an annular structure (23). However, a recent electron microscopy study of the intact *S. cerevisiae* TREX-2 complex expressed in insect cells indicated that here the CID region was orientated randomly relative to the M-region and showed no evidence to support an interaction between them mediated by the N-region (24). However, although the FG-repeats in the *S. cerevisiae* Sac3 N-terminal region are fewer and less similar to nucleoporin FG repeats, deletion of residues 1–140 generates growth and mRNA export defects (23).

Both cryo-EM and crystals obtained from the complete *S. cerevisiae* TREX-2 complex also indicated that the TPR-like repeats of the Sac3 M-region could extend past residue 250 towards the N-terminus, but, although three tubes of electron density consistent with helices could be observed, the limited resolution of these studies frustrated identification of the precise residues involved. Here we describe the 2.3 Å resolution crystal structure of a complex formed by Sac3 residues 60–550 complexed with Thp1 and Sem1. The structural model obtained indicated that the Sac3 TPR-like repeats extend to residue 137 and that residues 90–125 form a novel loop that links the distal portion of the Sac3 M-region to the TPR-like repeats in the N-terminal region of Thp1. Both the additional Sac3 TPR-like region and the novel loop linking Sac3 to Thp1 are important functionally. Deletion of the loop impairs growth at 37°C and also shows slight impairment of mRNA nuclear export, whereas deletion of the additional TPR-like region (residues 125–250) shows a pronounced mRNA export defect together with an increased lag phase when galactose is exchanged for glucose or raffinose as the carbon source. Overall this work has identified two new structural features of the Sac3 M-region that contribute to the function of the TREX-2 complex in yeast.

MATERIALS AND METHODS

Protein expression and purification

Sac3 (residues 60–550) cDNA was cloned into a modified TEV protease-cleavable version of pGEX-4T-1 (GE Healthcare) and co-expressed together with Thp1 and Sem1 cloned into the first and second MCS (multiple cloning site) of RSFDuet-1 (Novagen) in BL21-CodonPlus (DE3)-RIL cells (Stratagene) by isopropyl-β-D-thiogalactoside (IPTG) induction at 20°C as described (16). The Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex was purified from the soluble fraction by sequential chromatography on Glutathione Sepharose 4B and Superdex 200 columns (GE Healthcare).

X-ray crystallography

Crystals of the Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex were obtained by sitting drop vapour diffusion using 200 nl of 20% PEG 3350, 0.2 M Mg formate, 100 mM MES 6.5, mixed with 200 nl of the *S. cerevisiae* Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex at 16.5 mg/ml. Crystals grew over a period of 2 weeks and were harvested and then vitrified by plunging into liquid nitrogen. X-ray diffraction data were collected on beamline I03 at the Diamond Light Source (Didcot, UK). Reflections were indexed and integrated using XDS (25) and then scaled and merged in AIMLESS (26,27). An initial structural model was obtained by molecular replacement using Phaser in PHENIX (28) and the structure of the *S. cerevisiae* Sac3²⁵⁵⁻⁵⁵⁶:Thp1:Sem1 complex (ref. 16; PDB accession code: 3T5V) as a model. Iterative cycles of rebuilding using COOT (29) and refinement using PHENIX (28) were used to generate the final model that contained 8011 non-hydrogen protein atoms and 546 waters and which had excellent geometry (Table 1).

In vivo growth assays

Cells were freshly plated overnight on synthetic complete medium lacking tryptophan (SDC-TRP) plates. On the following day, they were diluted to OD₆₀₀ = 1 after which they were spotted in 10-fold dilutions on SDC-TRP and grown for 3 days at 37°C. Growth kinetics using galactose as a carbon source for cells initially grown in glucose or raffinose were analysed by measuring OD₆₀₀ at 6 min intervals in 96-well microplates at 30°C using a TECAN Spark 10M microplate reader and are reported as means ± SD for six determinations. Statistical significance was assessed using Student's *t*-test. Growth rates and lag times were determined as described (30). Expression levels of the mutants relative to an *ACT1* control were indistinguishable from those of *SAC3* when assessed by RT-qPCR using the SCRIPT one-step RT-qPCR greenmaster system (Jena Biosciences) following the manufacturer's protocol.

Fluorescence *in situ* hybridization

For *in situ* hybridization, cells were grown at 30°C in selective media before being fixed. Analysis of polyA⁺ RNA export by *in situ* hybridization was carried out using Cy3-labelled 30nt oligo(dT) probes essentially as described in

Table 1. Crystallography statistics

Data collection statistics	
Wavelength (Å)	0.9686
Space group	$P2_12_12_1$
Unit cell parameters: a, b, c (Å); α , β , γ (°)	78.37, 86.63, 168.34; 90, 90, 90
Resolution range (outer shell in brackets; Å)	19.95–2.30 (2.37–2.30)
Unique reflections	51 644 (4396)
Total observations	319 802 (27 953)
$\langle I / \sigma(I) \rangle$: all (outer shell)	7.5 (1.9)
$R_{p.i.m.}$: all (outer shell)	0.106 (0.979)
Completeness: all (outer shell) (%)	99.9 (100)
Multiplicity	6.1 (5.4)
Wilson B-factor	33.9
Refinement statistics	
Protein non-hydrogen atoms	8011
Waters	546
Bond length deviation from ideal values (Å)	0.0033
Bond angle deviation from ideal values (°)	0.53
Ramachandran favoured/outliers (%)	98.24/0
All-atom clashscore	1.37
Rwork/Rfree (%)	17.8/21.7
MolProbity (32) score (percentile)	0.87 (100)

Amberg *et al.* (31). Cells were examined at 100X using a Nikon TE2000 microscope. Fractions of cells showing nuclear accumulation of polyA⁺ RNA were calculated by examining at least 300 cells from at least three images and are presented as mean \pm SD.

RESULTS

Structure of the *S. cerevisiae* TREX complex M-domain

Crystals were obtained of the Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex that had $P2_12_12_1$ symmetry with $a = 78.2$ Å, $b = 84.1$ Å, $c = 165.5$ Å and which diffracted to 2.3 Å resolution using synchrotron radiation (Table 1). Phases were obtained using molecular replacement with the crystal structure of the Sac3²⁵⁵⁻⁵⁵⁶:Thp1:Sem1 complex (PDB accession code: 3T5V) and the electron density maps obtained after refinement and rebuilding showed residues 82–547 of Sac3, together with the Thp1 and Sem1 chains observed previously (Figure 2 and Supplementary Movie 1). The final model had an R-factor of 17.8% (Rfree = 21.7%) and a MolProbity (32) score of 0.87 (100th percentile) indicative of excellent geometry (Table 1). Except for some less-well defined regions also present in previous work, such as residues 1–26 and 42–56 of Sem1 and several loops within Sac3 and Thp1 (15,16), the chains of all three proteins could be traced reliably. In the regions that they shared, there were only minor structural differences between the present structure and the structures obtained previously for Sac3²⁵⁵⁻⁵⁵⁶:Thp1:Sem1 (PDB accession code: 3T5V) and Sac3²²²⁻⁵⁷²:Thp1¹⁷⁰⁻⁴⁵⁵:Sem1 (PDB accession code: 4TRQ).

In addition to the previously determined structures of Thp1, Sem1 and Sac3 residues 255–551, Sac3 residues 82–255 were clear and easily traced, but reliable density corresponding to residues 60–81 was not observed, possibly indicating that this region was disordered. Although residues 191, 192 and 244 were in poorly-defined loops, the connectivity and path of the main chain was clear and showed three roughly parallel α -helices together with a novel elongated loop that linked this region of Sac3 to Thp1. Electron microscopy (24) had shown three clear cylinders of additional density, consistent with α -helices, past residue 253 at which the original structural models (PDB accession

codes: 3T5V and 4TRQ) were truncated and had suggested that the Sac3 TPR-like repeat region extended further towards the N-terminus. The Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 crystal structure indicated that the TPR-like repeats of Sac3 indeed extend to residue 137 and identified unequivocally the residues in the three major helices. Sequence analysis using the PSIPRED package (<http://bioinf.cs.ucl.ac.uk/psipred>—ref. 33) had indicated that residues 143–152, 154–177, 206–215 and 233–240 had a high potential for forming α -helices, whereas residues 1–142 showed little potential for forming secondary structure. However, the crystal structure showed that the residues that contributed to the three additional helices were 137–153, 157–177 and 205–213. These three additional helices both extend the TPR-like solenoid of the Sac3 M-region and also cap it. Sac3 residues 126–137 were arranged along one side of the three-helix stack and possibly contributed to its stability. Remarkably, in the Sac3⁶⁰⁻⁵⁵⁶:Thp1:Sem1 complex crystals, Sac3 Arg256 was completely buried between the PCI region determined previously (residues 250–556) and the additional three helices (residues 137–250), where it formed a complex network of H-bonds to Asp293, Ser289 and Met203, and to Arg286, Ser289 and Lys238 through water S213 (Figure 3 and Supplementary Movies 2 and 3) that was augmented with hydrophobic interactions with Ile256, Leu261 and Cys200. Arg256 and residues interacting with it in this region were conserved between different species (Supplementary Figures S1 and 2) as would be anticipated for buried core residues.

In addition to the three additional helices of the TPR-like repeat region, the crystal structure of the Sac3⁶⁰⁻⁵⁵⁶:Thp1:Sem1 complex showed that Sac3 residues 90–125 formed a remarkable extended loop that linked the distal region of the Sac3 TPR-like region to the N-terminal region of Thp1 (Figures 2A, B, 4 and Supplementary Movie 1). With the exception of two short helices (residues 90–96 and 103–106), this loop did not contain extensive secondary structure and was generally in an extended conformation. The loop made extensive contacts with Thp1 (Figure 4) and buried 1528 Å² of surface area. Although the interface was dominated by hydrophobic side chains, notably Thp1 Pro77, Trp78 and Trp81 on Thp1 and Ile87, Ile93, Pro98,

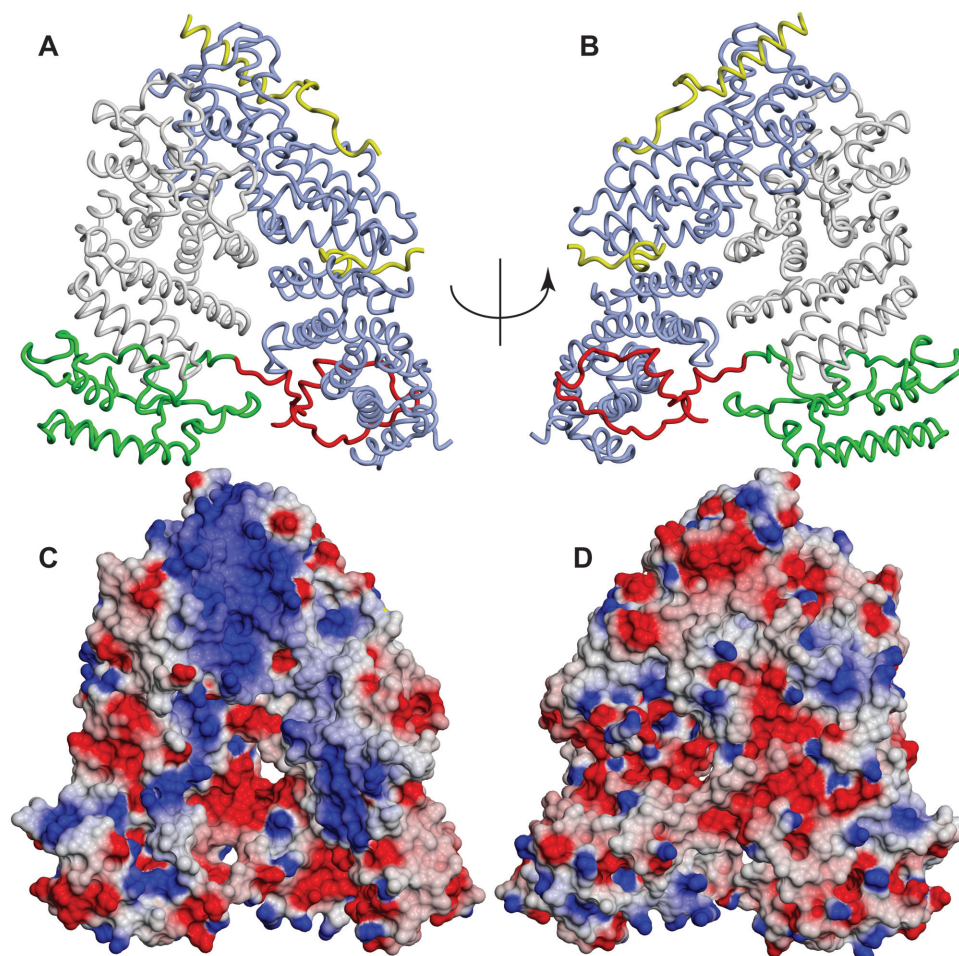


Figure 2. Crystal structure of the Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex. (A and B) In addition to the structures of Sac3 residues 255–551 (grey), Thp1 (blue) and Sem1 (yellow) determined previously (15,16), the crystal structure of the Sac3⁶⁰⁻⁵⁵⁶:Thp1:Sem1 complex showed an extension of the Sac3 TPR repeat region (green) containing three additional helices, together with a novel extended loop (red) that linked the base of the Sac3 TPR region (green) to the N-terminal region of Thp1. (C and D) Charge distribution on the surface of the Sac3⁶⁰⁻⁵⁵⁶:Thp1:Sem1 complex (blue is positive, red is negative charge). There is an extensive region of positive charge stretching from the winged-helix region of both Sac3 and Thp1 (towards the top of the structure) down the distal region of Sac3 together with a distinct positively-charged patch towards the base of Thp1. These positively-charged areas are on the opposite side of the complex to the linker containing Sac3 residues 90–125.

Leu99, Ile100, Phe108 and Phe123 on Sac3, 12 putative H-bonds also contributed to the interaction.

Previous studies identified a large surface patch of positive charge located over the winged helix regions of both Thp1 and Sac3 that extended somewhat towards the distal region of the Sac3 PCI fold and which has been implicated in interactions with nucleic acids (16) and the Mediator complex (15). This patch continued over the additional helices (Figure 2C and D) and so is more extensive than thought. However, the second positively charged cluster centered on Sac3 Arg256 that has been proposed to be involved in the interaction with Mediator (15) was buried by the residues in the additional TPR region below residue 250 and so was inaccessible (Figure 3) in the complete M-region structure. Generally the regions of positive surface charge were located on the opposite side of the complex to the loop formed by Sac3 residues 90–125 (Figure 2A and B) indicating that this loop may not interact directly with nucleic acids.

In summary, the crystal structure of the Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex identified two new structural features in the Sac3 M-region. In one, residues 137–250 fold into three additional helices that continue the stack of TPR-like helices previously identified in residues 255–550 and which appeared to be stabilized by residues 126–136 that lie along one side of these helices. In the second, residues 90–125 formed a remarkable loop that reaches out from Sac3 to connect it to the N-terminal region of Thp1.

Functional roles of the new Sac3 elements

Sac3 deletion mutants were used to explore the functional role of the extended loop (residues 90–125) and the continuation of the TPR region (residues 125–250) to function in terms of growth rate, *GAL* induction, and mRNA export.

Deletion of the Sac3 extended loop (*sac3Δ125*) impaired cell growth at 37°C to levels comparable to deletion of *SAC3* (Figure 5A). Although deletion of *SAC3* generates a growth

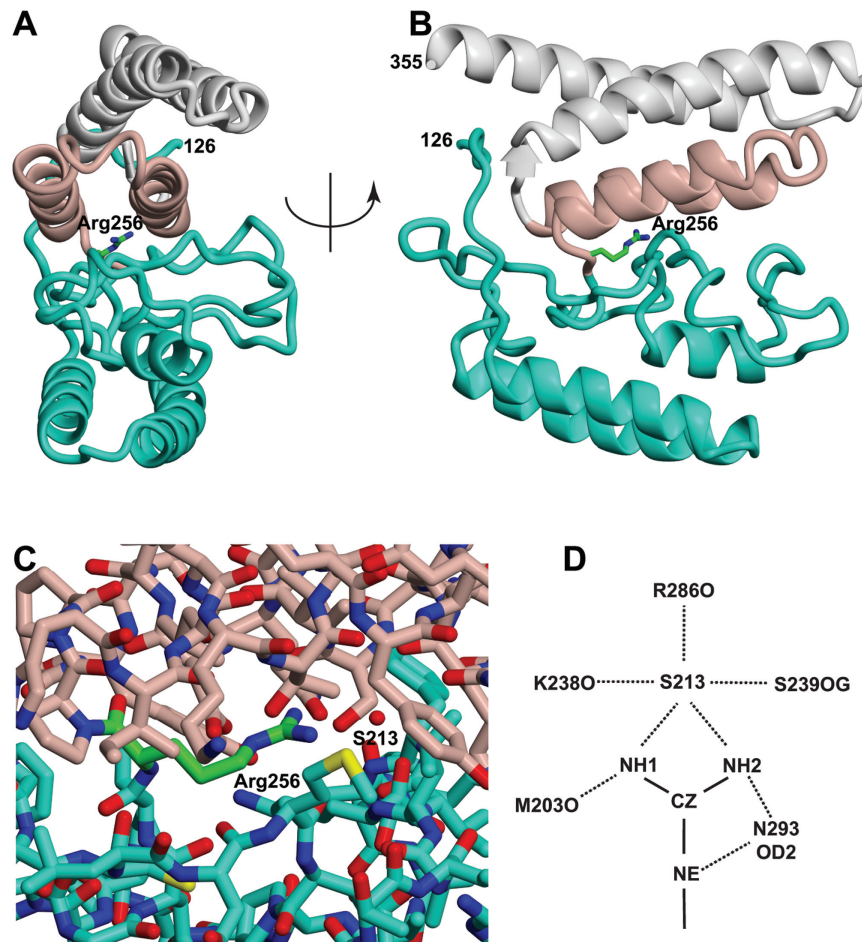


Figure 3. Arg256 is buried in the interface between the Sac3 TPR helices and the three additional helices. (A and B) Illustration of how Sac3 Arg256 is sandwiched between the two helices of the TPR repeat that contains residues 255–269 and 280–296 (pink) and the additional module (residues 126–250, green). Arg256 and the Sac3 255–269 and 280–296 helices formed lattice contacts in previous crystal structures and were proposed to form an interaction interface (15), but in the more complete structure of the Sac3^{60–556}:Thp1:Sem1 complex they are buried and inaccessible. (C) Detail of the residues surrounding Arg256 and water S213 that are buried in the interface between the 255–296 TPR-like repeat and residues 125–250. (D) Hydrogen bond network linking Arg256 to Met203, Lys238, Arg286, Ser289, Asn293 and water S213. See also Supplementary Movies 2 and 3.

defect at 37°C (10,13,34), previous work had shown that deletion of residues 1–60 did not impair growth at 37°C, whereas deletion of residues 1–140 (23), the R256D mutation (15) or mutation of positively-charged residues in the winged-helix domains of either Sac3 or Thp1 (16), generated severe growth defects at 37°C. We found that, whereas deletion of residues 1–90 (*sac3*Δ90) had a negligible impact on growth at 37°C, deletion of these residues together with the extended loop (*sac3*Δ125) showed a considerable growth defect that was comparable to that seen in *sac3* NULL cells (Figure 5A).

Although the extended loop (residues 90–125) did not appear to influence the lag time when galactose was substituted for glucose or raffinose as the carbon source, deletion of both the loop and the TPR extension (*sac3*Δ250) resulted in a substantial increase in the lag time (Figure 5B). The lag time for cells initially grown in glucose was longer than for cells grown in raffinose because, although in both cases it is necessary to induce the *GAL* genes needed for galactose metabolism, it is also necessary to remove the repression of the *GAL* genes present when glucose is the car-

bon source. Budding yeast grow best using glucose as a carbon source, but can use galactose if glucose is absent. In the presence of glucose, the genes of the galactose (*GAL*) network are repressed by several different mechanisms (including recruitment of the Mig1 DNA-binding repressor and the Cyc8:Tup1 co-repressor complex) acting in concert (35,36). However, this repression is removed when an alternative carbon source, such as galactose, raffinose or glycerol, is substituted for glucose and is facilitated by desumoylation of Cyc8 and Tup1 by Ulp1 that is located at NPCs (37). Although in glucose medium the *GAL7-10-1* gene cluster is distributed throughout the nucleus, this cluster becomes localized to the nuclear envelope in either raffinose or galactose medium (13), bringing it into close proximity to Ulp1 facilitating removal of the Mig1-based repression (37). However, in the presence of raffinose or glycerol, Gal80 still remains bound to the Gal4 activator and represses its activity. In the presence of galactose, Gal80 is released from Gal4, activating the *GAL* genes. Consequently, the lag phase when cells are transferred directly from glucose to galactose medium is reduced when cells are initially grown in

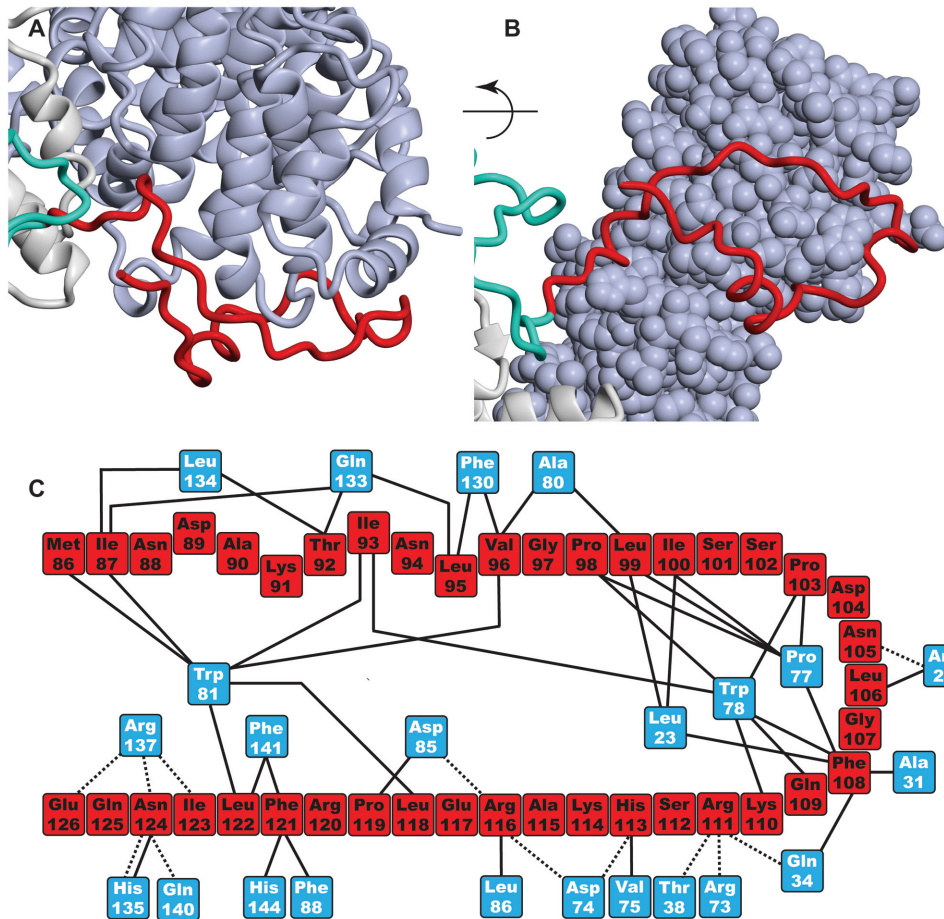


Figure 4. Sac3 residues 90–125 form an extended loop linking the distal end of the TPR region to Thp1. (A and B) Illustration of how the loop (red) stretches out from Sac3 (grey) to encircle the N-terminal region of the Thp1 PCI region (blue) with which it makes an extensive number of contacts and buries 1528 Å² of surface area. Except for a small helix (residues 89–95), the residues in the loop lack regular secondary structure. (C) Schematic illustration of the interactions between residues in the Sac3 extended loop (red) and Thp1 (blue).

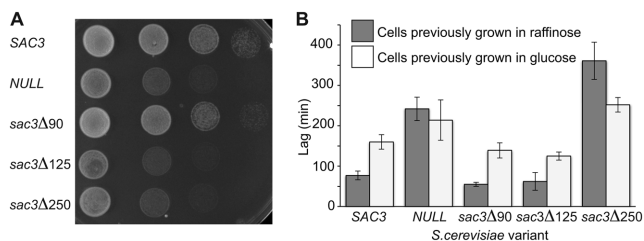


Figure 5. Influence of different Sac3 regions on growth. (A) Serial dilutions of *sac3* NULL cells carrying plasmid-borne *SAC3* or different *SAC3* truncations. Cells were grown at 37°C for 3 days. Although deletion of residues 1–90 (*sac3*Δ90) grew similarly to *SAC3* cells, deletion of the extended loop (*sac3*Δ125) produced a growth defect similar to *sac3* NULL cells. (B) Lag times for *Saccharomyces cerevisiae* cells transferred from either glucose or raffinose to galactose medium. Cells previously grown in glucose showed longer lag times than cells previously grown in raffinose. *SAC3*, *sac3*Δ90 and *sac3*Δ125 showed similar lag phases, whereas *sac3* NULL cells and *sac3*Δ250 cells showed longer lag times ($P < 0.05$). Bars represent SDs from six determinations.

raffinose before being transferred to galactose. We found that both *sac3* NULL and *sac3*Δ250 cells showed significantly ($P < 0.05$) longer lag times when raffinose was re-

placed by galactose as the carbon source, whereas *sac3*Δ90 and *sac3*Δ125 cells showed lag times of the order of 50 min, similar to that observed for *SAC3* cells. The lag phase with *sac3*Δ250 cells was considerably longer ($P < 0.05$) even than that seen with *sac3* NULL cells (Figure 5B) perhaps as a result of its sequestering other cellular components as seen with other Sac3 mutants (34). A similar pattern was seen in cells transferred directly from glucose to galactose medium, although in this case the lag times were generally longer (Figure 5B). Again only *sac3* NULL and *sac3*Δ250 cells showed lag times significantly ($P < 0.05$) longer than wild-type cells. Taken together, these results indicate a role for Sac3 residues 126–250 (that correspond to the extension of the Sac3 TPR domain) in the expression of the *GAL* genes.

In addition to increasing the lag time when galactose was substituted for glucose or raffinose as a carbon source, the *sac3*Δ250 variant showed polyA⁺ RNA export defects, with 88% of cells showing a level of nuclear retention of polyA⁺ RNA when examined by FISH (Figure 6). The *sac3*Δ125 variant showed a slight impairment, with 19% of cells showing some nuclear accumulation in FISH, which, although statistically significant ($P < 0.05$), was marginal when compared to the defect seen with *sac3* NULL or *sac3*Δ250 cells.

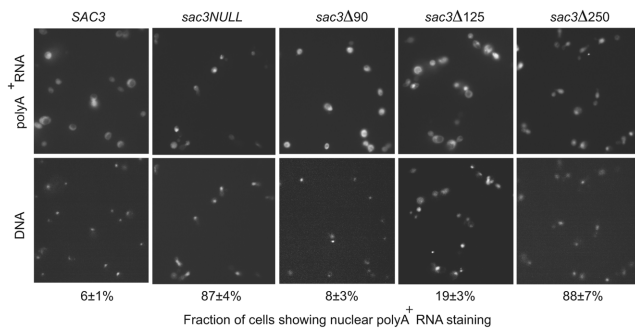


Figure 6. Contributions of different regions of Sac3 to mRNA nuclear export. Levels of nuclear accumulation of polyA⁺ RNA with different Sac3 variants were assessed using FISH with Cy3-labelled 30 nt oligo-(dT) probes. *Sac3*Δ90 showed 8% of cells with increased nuclear staining, which was similar to *SAC3* cells (6%); *sac3*Δ125 cells showed a slight increase in nuclear staining (19%), whereas most *sac3*Δ250 cells (88%) showed nuclear accumulation of polyA⁺ RNA similar to that seen with *sac3* NULL cells (86%), indicating that the nuclear export of mRNA had been impaired by this deletion.

The *sac3*Δ90 variant showed only 8% of cells with nuclear accumulation of polyA⁺ RNA and was similar to wild-type cells that showed 6% (Figure 6).

In summary, the data obtained with the deletion mutants indicate that, although the loop containing residues 90–125 is important for growth, its deletion does not impair mRNA export greatly nor the lag times seen when galactose is substituted as the carbon source, whereas the TPR extension is important for both of these functions.

DISCUSSION

Although previous studies proposed that Sac3 residues 220–250 are inherently disordered (15), results obtained by cryo-EM and crystallography pointed towards there being considerable secondary and tertiary structure in this region and indicated that there are three helices present in the Sac3 chain before residue 255 (24). Moreover, even when they are not in secondary structure elements, the positions of residues 90–250 were well defined and clearly not disordered. The 2.3 Å resolution crystal structure of the Sac3⁶⁰⁻⁵⁵⁶:Thp1:Sem1 complex demonstrated that the TPR-like repeats of the PCI fold of the Sac3 M-region indeed extend past residue 255 to residue 137, and also showed the presence of a novel loop (residues 90–125) that links the Sac3 TPR region to Thp1 (Figures 2 and 7 and Supplementary Movie 1). The sequences of Sac3 homologues from both yeast (Supplementary Figure S1) and vertebrates (Supplementary Figure S2) were conserved between residues 90 and 250, consistent with both the helices and the extended loop being retained.

The Sac3^M:Thp1:Sem1 region forms an interaction platform by which the TREX-2 complex interacts with both mRNAs and the Mediator complex (15,16) and so the demonstration that the TPR-repeat region extends substantially below residue 255 indicates that the interaction interface between TREX-2 and these components may be more extensive than previously thought. Moreover, the new structural information indicated that results obtained previously with some Sac3 truncation constructs should be inter-

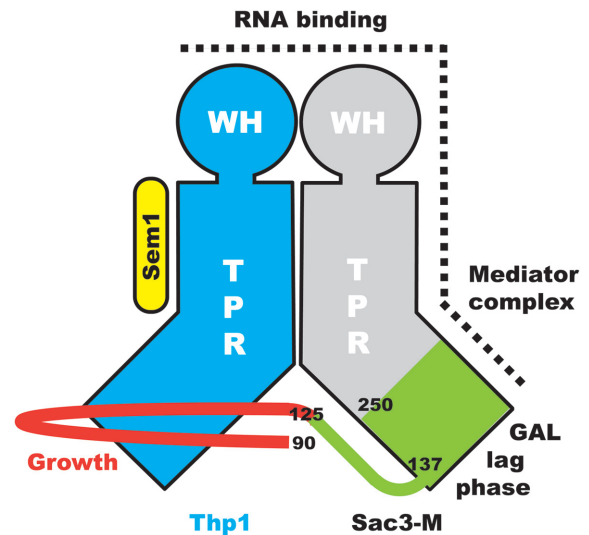


Figure 7. Schematic illustration of the structure of the TREX-2 M-region. Both Sac3 and Thp1 have PCI folds based on a series of α -helical TPR repeats together with a winged helix (WH) domain, whereas Sem1 lies along one side of Thp1. The structure of the Sac3⁶⁰⁻⁵⁵⁶:Thp1:Sem1 complex has identified two new structural features in Sac3: an extension of the TPR region from residue 250 to residue 127 (green) and an extended loop comprising residues 90–125 that links the distal portion of the Sac3 TPR region with Thp1. A cluster of positively-charged residues that extends from the Thp1 winged helix domain and down one side of Sac3 appears to be important for binding RNA. The extended loop (residues 90–125) is important for growth, whereas residues 126–250 are important for the induction of *GAL* genes when cells are moved to galactose as a carbon source. Other studies (15) have indicated that positive residues in the Sac3 TPR region also function in conjunction with the mediator complex.

preted cautiously because the truncations were sometimes located within secondary structural elements and so could have disrupted the conformation locally. Thus, for example, although deletion of Sac3 residues 1–140 generates both growth and mRNA export defects that are similar to those observed in a *SAC3 null* strain (23), residues 137–140 are contained within a helix and residues 126–136 appear to contribute to the stability of the distal TPR repeat. Moreover, this construct lacked the 90–125 loop and so, because the *Sac3*Δ125 mutant showed a similar growth defect (Figure 5B), the growth defect observed could derive from this source. Similarly, although deletion of residues 1–221 generated crystals in which Sac3 residues 222–252 were proposed to be disordered (15), this disorder could also have resulted from deletion of the three additional helices (residues 137–153, 157–177 and 205–213). Residues 222–252 were not disordered when the longer Sac3⁶⁰⁻⁵⁵⁰ construct was employed (Figure 2A, B and Supplementary Movie 1). It has also been proposed (15) that because the Sac3 helices encompassing residues 258–269 and 280–296 contributed to crystal packing contacts in two different crystal forms, they could represent a physiologically relevant interaction site. However, this interface is buried by the extension of the TPR domain in the Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 crystals (Figure 3 and Supplementary Movies 2 and 3) and so is inaccessible in the complete M-domain. It is, however, likely that in constructs in which residues 1–250 were absent, this exposed

interface could have been somewhat sticky and could have contributed to crystal packing contacts as a result.

The TREX-2 complex functions in a number of different contexts (including mRNA export, integrating nuclear steps of the gene expression pathway, transcription of genes such as *GAL*) and different regions of the complex appear to contribute to each. The localization of TREX-2 to NPCs, which appears to be important for all of these functions, appears to derive primarily from the interaction of the CID domain (Sac3 residues 720–805, together with Cdc31 and two Sus1 chains) with Nup1, a component of the NPC nuclear basket (10,17,18), whereas the M- and N- regions appear to be important for growth, mRNA export, localization of genes such as *GAL 7-10-1* cluster and the integration of many of the nuclear steps in the gene expression pathway. Previous work has implicated positively charged residues in the Sac3 M-region with the interaction with both RNAs (16,38) and with the Mediator complex (15). In the latter case, it is unlikely that one of the Sac3 residues proposed as interacting with the Mediator complex, Arg256, is involved directly because it is inaccessible (Figure 3 and Supplementary Movies 2 and 3). Mutating Arg256 to Asp would be expected to disrupt the interface between the residues in the distal helical region of the Sac3 PCI fold (residues 125–250) and those in the next TPR repeat and the local loss of structure could instead account for the this variant not interacting with the Mediator complex. Although further work will be required to define the precise role of the R256D mutant, because Sac3 residues 255–557 alone form a well-ordered structure, any disruption due to mutating Arg256 might be restricted to the distal region of the Sac3 TPR region and so account for the ability of this mutant to still bind to Thp1.

Although further work will be necessary to define the precise function of the extended loop formed by Sac3 residues 90–125, this loop is clearly important and deleting residues 1–125 (*sac3 Δ 125*) generated a growth defect at 37°C similar to that observed with the *sac3 NULL* strain (Figure 5A), whereas deletion of residues 1–90 (*sac3 Δ 90*) displayed wild-type growth. The considerable surface area buried by this loop's binding to Thp1 (1528 Å²) would be consistent with this interaction being relatively strong and it could, for example, function to retain the position of the distal region of the Sac3 M-region (residues 125–250) relative to Thp1. The lack of clusters of positive charge in this region (Figure 2C and D) suggests that it is probably not involved directly in interactions with nucleic acids and its small impact on mRNA export was probably a result of its deletion altering the structure of Sac3 residues 125–250, especially since impairment of mRNA export by TREX-2 variants does not necessarily translate into any reduction in growth rate (see, e.g. 20, 24, 35). Thus, the phenotype of the *sac3 Δ CID* strain shows a slightly reduced growth rate and a moderate mRNA export defect when compared with the *sac3* null strain, which is strongly impaired in mRNA export. However, deletion of residues 1–250 showed more pronounced impairment of mRNA export indicating that, in addition to the major positive patch centered on the winged helix regions of Sac3 and Thp1, residues in the distal region of Sac3 past residue 250 that encompassed the three additional helices made an important contribution to this function. Indeed, as seen in Figure 2C, the positively-charged region did

extend somewhat towards the base of the Sac3 TPR domain and would be consistent with an analogous extension of the mRNA binding region.

Although the loop encompassing Sac3 residues 90–125 impaired growth at 37°C (Figure 5A), it did not alter the lag times observed at 30°C when galactose was substituted as a carbon source with cells grown initially in either glucose or raffinose (Figure 5B). *Sac3 NULL* cells showed increased lag times for cells grown initially in either glucose or raffinose and remarkably *sac3 Δ 250* cells showed an even greater increase, especially in cells initially grown in raffinose. Because the lag phase when raffinose is replaced by galactose requires only activation of *GALI*, it is shorter than when glucose is replaced by galactose because then both derepression and activation are necessary. Previous work has shown that TREX-2 facilitates the localization of *GAL* genes to NPCs when either raffinose or galactose is used as a carbon source and, in addition to derepressing/activating the *GAL* genes, this localization also appears to generate a negative feedback loop that is important when galactose is removed (13). Although the localization of TREX-2 itself to NPCs is mediated primarily by an interaction between the CID region and proteins of the nuclear basket such as Nup1 (10,17), Schneider *et al.* (15) have shown that the M- region of TREX-2 is also required for localization of the *GAL7-10-1* cluster to NPCs and the increase in lag times seen with *sac3 Δ 250* cells is consistent with the region containing Sac3 residues 125–250 contributing to this function.

In summary, the 2.3 Å resolution the crystal structure of the Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex has demonstrated that the TPR-like repeat region of the Sac3 chain extends further towards the N-terminus than had been proposed previously and, in addition to three additional helices that extend the TPR-like domain, Sac3 residues 90–125 form an extended loop that links the distal regions of the Sac3 and Thp1 PCI domains (Figure 7). Deletion mutants indicate that the additional TPR-like region of Sac3 is important for both mRNA export and for orchestrating the expression of *GAL* genes when galactose is substituted for either glucose or raffinose as a carbon source, whereas deletion of the Sac3⁹⁰⁻¹²⁵ loop results in a growth defect at 37°C.

ACCESSION CODE

Coordinates and structure factors for the Sac3⁶⁰⁻⁵⁵⁰:Thp1:Sem1 complex have been deposited in the protein data bank with accession code 5UBP.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to our colleagues in Cambridge for their assistance, insightful comments and criticisms, and to John O'Neill, Nick Barry, Jonathan Howe and Mathias Pasche for technical assistance. We also thank the Diamond Light Source and staff of beamline I03 for assistance with collection of the X-ray crystallography data.

FUNDING

Medical Research Council (MRC) [MC_U105178939 to M.S.]. Funding for open access charge: MRC [MC_UP_1201/6].

Conflict of interest statement. None declared.

REFERENCES

- Hoelz,A., Glavy,J.S. and Beck,M. (2016). Toward the atomic structure of the nuclear pore complex: when top down meets bottom up. *Nat. Struct. Mol. Biol.* **23**, 624–630.
- Matsuura,Y. (2016). Mechanistic insights from structural analyses of Ran-GTPase-driven nuclear export of proteins and RNAs. *J. Mol. Biol.* **428**, 2025–2039.
- Christie,M., Chang,C.W., Róna,G., Smith,K.M., Stewart,A.G., Takeda,A.A., Fontes,M.R., Stewart,M., Vértessy,B.G., Forwood,J.K. and Kobe,B. (2016). Structural Biology and Regulation of Protein Import into the Nucleus. *J. Mol. Biol.* **428**, 2060–2090.
- Vetter,I.R. and Wittinghofer,A. (2001). The guanine nucleotide-binding switch in three dimensions. *Science*, **294**, 1299–1304.
- Strässer,K., Bassler,J. and Hurt,E. (2000). Binding of the Mex67p/Mtr2p heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export. *J. Cell Biol.* **150**, 695–706.
- Köhler,A. and Hurt,E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nature Rev. Mol. Cell Biol.* **8**, 761–773.
- Fischer,T., Strässer,K., Rác,A., Rodríguez-Navarro,S., Oppizzi,M., Ihrig,P., Lechner,J. and Hurt,E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *EMBO J.*, **21**, 5843–5852.
- Rodríguez-Navarro,S., Fischer,T., Luo,M.-J., Antúnez,O., Bretschneider,S., Lechner,J., Pérez-Ortín,J.E., Reed,R. and Hurt,E. (2004). Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell*, **116**, 75–86.
- Casolari,J.M., Brown,C.R., Komili,S., West,J., Hieronymus,H. and Silver,P.A. (2004). Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell*, **117**, 427–439.
- Cabal,G.G., Genovesio,A., Rodríguez-Navarro,S., Zimmer,C., Gadál,O., Lesne,A., Buc,H., Feuerbach-Fournier,F., Olivo-Marin,J.C., Hurt,E.C. and Nehrass,U. (2006). SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, **441**, 770–773.
- Dieppl,G., Iglesias,N. and Stutz,F. (2006). Cotranscriptional recruitment to the mRNA export receptor Mex67p contributes to nuclear pore anchoring of activated genes. *Mol. Cell Biol.* **26**, 7858–7870.
- Taddei,A., Van Houwe,G., Hediger,F., Kalck,V., Cubizolles,F., Schober,H. and Gasser,S.M. (2006). Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature*, **441**, 774–778.
- Green,E.M., Jiang,Y., Joyner,R. and Weis,K. (2012). A negative feedback loop at the nuclear periphery regulates GAL gene expression. *Mol. Biol. Cell*, **23**, 1367–1375.
- García-Oliver,E., García-Molinero,V. and Rodríguez-Navarro,S. (2012). mRNA export and gene expression: the SAGA-TREX-2 connection. *Biochim. Biophys. Acta*, **1819**, 555–565.
- Schneider,M., Hellerschmid,D., Schubert,T., Amlacher,S., Vinayachandran,V., Reja,R., Pugh,B.F., Clausen,T. and Köhler,A. (2015). The nuclear pore-associated TREX-2 complex employs mediator to regulate gene expression. *Cell*, **162**, 1016–1028.
- Ellisdon,A.M., Dimitrova,L., Hurt,E. and Stewart,M. (2012). Structural basis for the assembly and nucleic acid binding of the TREX-2 transcription-export complex. *Nat. Struct. Mol. Biol.* **19**, 328–336.
- Jani,D., Lutz,S., Marshall,N.J., Fischer,T., Köhler,A., Ellisdon,A.M., Hurt,E. and Stewart,M. (2009). Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export. *Mol. Cell*, **33**, 727–737.
- Jani,D., Valkov,E. and Stewart,M. (2014). Structural basis for binding the TREX2 complex to nuclear pores, GAL1 localisation and mRNA export. *Nucleic Acids Res.*, **42**, 6686–6697.
- Ellisdon,A.M. and Stewart,M. (2012). Structural biology of the PCI-protein fold. *Bioarchitecture*, **2**, 118–123.
- Gallardo,M., Luna,R., Erdjument-Bromage,H., Tempst,P. and Aguilera,A. (2003). Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J. Biol. Chem.* **278**, 24225–24232.
- Faza,M.B., Kemmler,S., Jimeno,S., González-Aguilera,C., Aguilera,A., Hurt,E. and Panse,V.G. (2009). Sem1 is a functional component of the nuclear pore complex-associated messenger RNA export machinery. *J. Cell Biol.* **184**, 833–846.
- Bhatia,V., Barroso,S.I., García-Rubio,M.L., Tumini,E., Herrera-Moyano,E. and Aguilera,A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, **511**, 362–365.
- Dimitrova,L., Valkov,E., Aibara,S., Flemming,D., McLaughlin,S.H., Hurt,E. and Stewart,M. (2015). Structural characterization of the *Chaetomium thermophilum* TREX-2 complex and its interaction with the mRNA nuclear export factor Mex67:Mtr2. *Structure*, **23**, 1246–1257.
- Aibara,S., Bai,X. and Stewart,M. (2016). The Sac3 TPR-like region in the *Saccharomyces cerevisiae* TREX-2 complex is more extensive but independent of the CID region. *J. Struct. Biol.* **195**, 316–324.
- Kabsch,W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.
- Evans,P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.
- Evans,P.R. and Murshudov,G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214.
- Adams,P.D., Afonine,P.V., Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
- Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
- Hall,B.G., Acar,H., Nandipati,A. and Barlow,M. (2014). Growth rates made easy. *Mol. Biol. Evol.* **31**, 232–238.
- Amberg,D.C., Goldstein,A.L. and Cole,C.N. (1992). Isolation and characterization of RAT1: an essential gene of *Saccharomyces cerevisiae* required for the efficient nucleocytoplasmic trafficking of mRNA. *Genes Dev.* **6**, 1173–1189.
- Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.
- Buchan,D.W., Minneci,F., Nugent,T.C., Bryson,K. and Jones,D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.*, **41**, W349–W357.
- Fischer,T., Rodríguez-Navarro,S., Pereira,G., Rác,A., Schiebel,E. and Hurt,E. (2004). Yeast centrin Cdc31 is linked to the nuclear mRNA export machinery. *Nat. Cell Biol.*, **6**, 840–848.
- Keleher,C.A., Redd,M.J., Schultz,J., Carlson,M. and Johnson,A.D. (1992). Ssn6-Tup1 is a general repressor of transcription in yeast. *Cell*, **68**, 709–719.
- Tzamaras,D. and Struhl,K. (1994). Functional dissection of the yeast Cyc8-Tup1 transcriptional co-repressor complex. *Nature*, **369**, 758–761.
- Texari,L., Dieppl,G., Vinciguerra,P., Contreras,M.P., Groner,A., Letourneau,A. and Stutz,F. (2013). The nuclear pore regulates GAL1 gene transcription by controlling the localization of the SUMO protease Ulp1. *Mol. Cell*, **51**, 807–818.
- Valkov,E. and Stewart,M. (2015). 1.25 Å resolution structure of an RNA 20-mer that binds to the TREX2 complex. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 1318–21.