The University Of Sheffield.

Structural and Functional Analysis of TREX

PhD thesis submission

by

Arthur Ekasatrya Holland

Department of Molecular Biology and Biotechnology

University of Sheffield

Sheffield, United Kingdom

September 2014

Acknowledgements

Firstly, I would like to thank my supervisor Professor Stuart Wilson for giving me the opportunity to undertake this project, and for providing support and advice throughout.

I would also like to thank my lab colleagues Guillaume Hautbergue, Nicolas Viphakone, Victoria Porteous, Chung-Te Chang, Matthew Walsh, Marcus Cumberbatch, Simon Lesbirel and Agata Makar. I would particularly like to express my gratitude to Guillaume and Nicolas for their invaluable help and advices with my experiments. Thank you also to Dr Patrick Baker and Fiona Rodgers for their assistance during my X-ray crystallisation experiments. Big thanks also to Kirsty Liversidge for her support during my write-up.

Finally I would like to thank my parents for their support throughout my education/life.

Cheers, guys

Table of Contents

Abstract

The TREX (Transcription-Export) complex is a well-conserved assembly of proteins, which is essential in eukaryotic gene expression. The expression of mRNA involves a number of processing events of primary transcripts to form a mature mRNA. TREX functions at the interface between transcription and export, associating with transcripts at various stages during its processing and packages it into an export-competent mRNP. Export of most mRNAs is dependent on the export receptor, Nxf1. This protein is recruited to the mRNP by directly binding to TREX components. Because the association of TREX with the mRNA is dependent on accurate processing events, it serves as a quality control signal for the recruitment of Nxf1 to promote export. Furthermore, newly identified TREX-associated factors suggest that TREX could integrate into a wide range of cellular functions and mounting evidence indicates that TREX is essential for the maintenance of genome stability. Aberrations in TREX expression cause harmful RNA:DNA hybrid structures, which are susceptible to DNA damage. TREX therefore, represents an important target for studying the regulation of gene expression.

Chapter 1

Introduction

Eukaryotic gene expression requires that chromosomal DNA is transcribed into RNA in the nucleus, which is translated into protein in the cytoplasm. The export of mature, translationally competent mRNA is dependent on a sequence of intimately coupled upstream processing events: 5' end capping, splicing and 3' end processing. The TREX (Transcription/Export) complex is recruited to the mRNA during transcription and functions to link transcriptional and processing events to export.

1.1 The Nuclear Pore

In eukaryotes, the nuclear membrane functions as a physical barrier between the cytoplasm and nucleoplasm, thus spatially separating the processes that occur in each compartment. This separation therefore necessitates the transport of macromolecules across the membrane, and this function is carried out by the nuclear pore complexes (NPCs). NPCs are made up of multiple copies of approximately 30 different proteins called nucleoporins (Cronshaw et al. 2002; Rout et al. 2000), which form an aqueous channel within the nuclear envelope to allow nucleo-cytoplasmic traffic. This translocation of molecules is highly efficient, in some cases allowing at least 100 translocations per second (Ribbeck & Görlich 2001).

1.1.1 NPC Structure

The overall structure of the NPC (Fig. 1.1) is well conserved from yeasts to vertebrates, consisting of a scaffold embedded in the nuclear envelope surrounding a central channel, and two rings (cytoplasmic and nuclear) with protruding filaments (Q. Yang et al. 1998). Based on their function within the

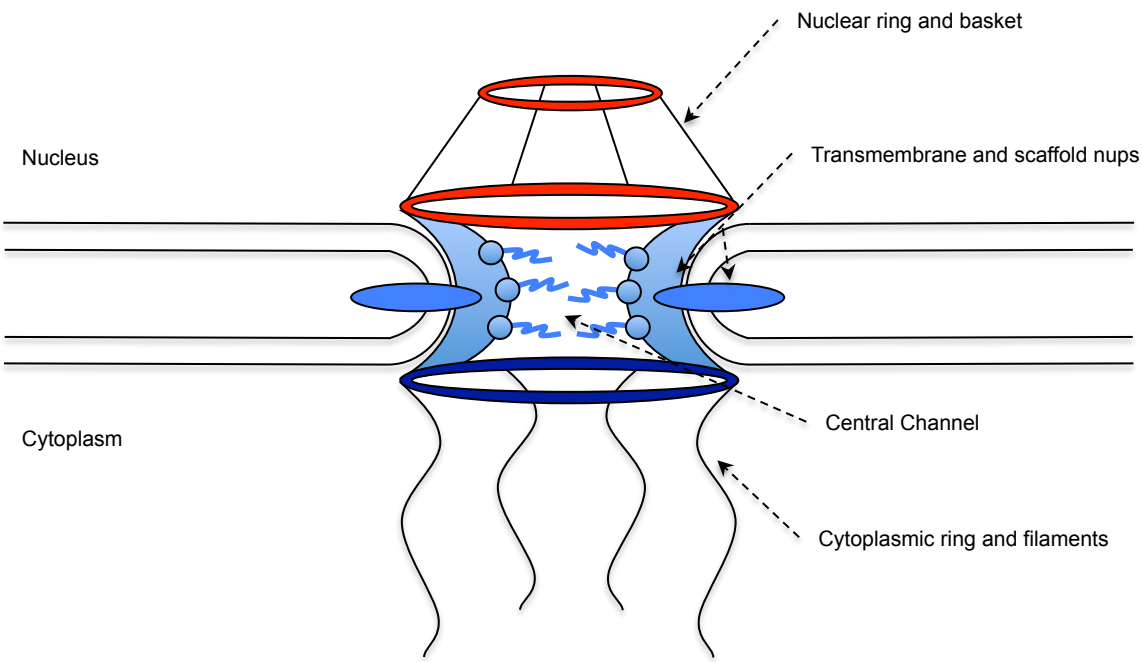


Figure 1.1. Architecture of the Nuclear pore complex. Approximately 30 different types of nucleoporin (nups) are found in the NPC. Scaffold nups are embedded in the nuclear envelope, surrounding a central aqueous channel. The scaffold also serves as an anchor point for peripheral nups, which form the nucleoplasmic and cytoplasmic filaments, and those which extend into the central channel.

nuclear pore complex, nucleoporins (nups) can be broadly divided into three different classes (Rout & Aitchison 2001). The pore membrane proteins (poms), are transmembrane proteins which co-fractionate with the nuclear membrane (Rout et al. 2000). These proteins are thought to stabilise the initial formation of the nuclear pore by fusion of the outer and inner nuclear membranes, and serving as anchor points for other nups to assemble at the NPC. Other nups act as a scaffold to support the central channel structure of the pore and allow dilation of the pore for the transport of molecules (Kiseleva et al. 1998). Finally, FG (phenylalanine-glycine) nucleoporins contain repeats of the FG motif, taking the form of FG, GLFG or FXFG. These FG nups are anchored to the scaffold proteins, extending into the central pore channel and projecting out to form the nucleoplasmic and cytoplasmic filaments. FG nups are involved in interacting with transport receptors and crystal structures have been solved of FG peptides in complex with various export receptors (Bayliss, Littlewood, et al. 2002b; Bayliss, Leung, et al. 2002a). The different types of FG nups exhibit different localisations in the NPC, with FG-nucleoporins showing a bias for the cytoplasmic side, whereas FXFG-nucleoporins display a bias for the nucleoplasmic side of the NPC (Rout et al. 2000), thus they may play distinct roles in mediating translocation of specific transport receptors. Furthermore, the dynamics of the different nup classes vary considerably, with residence times at the NPC ranging from a few seconds to several hours (Rabut et al. 2004). Significantly, nups that compose the scaffold of the complex are extremely stable with residence times of up to 70 hours, reflecting their role as essential components of the NPC structure. By contrast, most FG nups are more dynamic, residing at the NPC for seconds or minutes. This short residence would allow for rapid protein turnover, thus allowing the NPC to adapt in response to cellular changes (D'Angelo & Hetzer 2008).

1.1.2 Nuclear Transport

The transport of most protein and RNA cargos between the nucleus and cytoplasm is mediated by the small GTPase Ran. Ran shuttles between the

nucleus and cytoplasm, altering between GTP and GDP bound states respectively. A gradient of the two Ran states is maintained by regulatory proteins, which are localised to either the cytoplasm or nucleus (Izaurralde et al. 1997). The cytoplasmic protein RanGAP (Ran GTPase activating protein) induces the GTPase activity of Ran thus converting it to the GDP bound state. In the nucleus, RanGEF (Ran guanosine exchange factor) catalyses exchange of the bound GDP for GTP. Cargos are transported by binding to transport receptors, known as exportins and importins. These proteins shuttle between the nucleus and cytoplasm, but their affinity for their cargo is mediated by binding to Ran. In the nucleus, exportins only bind their cargo efficiently in the presence of Ran-GTP. This trimeric complex then traverses the nuclear pore into the cytoplasm, where Ran's GTPase activity is induced, by the cytosolic protein RanGAP. The conversion of Ran to its GDP bound state causes the disassembly of the Ran-exportin-cargo complex and thus the cargo is released into the cytoplasm. Transport of cargo into the nucleus occurs via the importins, which compared to exportins, bind their cargo in the absence of Ran. On entering the nucleus, the importin becomes bound by RanGTP which reduces its affinity for the cargo and induces its release into the nucleus (Mattaj & Englmeier 1998).

In comparison to the described system, export of spliced mRNA appears to occur independently of Ran. This can be demonstrated by introducing Ran mutants which prevent guanosine exchange, or microinjecting RanGTP (normally cytoplasmic) into *Xenopus* nuclei. Both these experiments perturb the normal Ran cycle by allowing RanGDP to accumulate in the nucleus, resulting in a block of export. Significantly, tRNAs and snRNAs are blocked whereas export of spliced mRNA was unaffected (Clouse et al. 2001). Instead, spliced mRNA export is mediated by the nuclear export factor Nxf1, and regulated by a host of proteins associated with mRNA processing.

1.2 Ran-dependent RNA export

Export of several RNA species utilise the RanGTPase pathway. Export of these RNAs (tRNA, snRNA, rRNA and miRNA), are mediated by members of the

karyopherin superfamily of export receptors (exportins). These proteins all share an N-terminal Ran-GTP binding motif (Görlich & Kutay 1999; Köhler & E. Hurt 2007). As described above, association with Ran-GTP allows binding of cargo and subsequent export via association with nucleoporins.

1.2.1 Crm-1 pathways

Human Crm-1 was identified in human cells by its interaction with nucleoporins (Fornerod, van Deursen, et al. 1997b). It was later further characterised as an export receptor for leucine-rich nuclear export signals (NES) and involved in export of snRNAs (Fornerod, Ohno, et al. 1997a) and rRNAs. Crm-1 however, does not bind these RNAs directly, instead it relies on adaptor proteins, which recruit Crm-1 to specific cargoes.

U snRNAs, essential cofactors in the spliceosomal snRNPs, are transcribed by RNA Polymerase II (RNAPII) before being exported to the cytoplasm for assembly into mature, functional snRNPs. This export is dependent on Crm-1, which does not directly interact with the snRNA but requires an additional protein called PHAX (phosphorylated adaptor for RNA export), which contains leucine-rich NES signals important for Crm-1 association (Ohno et al. 2000).

Crm-1 is also implicated in the export of rRNA of both 40S and 60S ribosomal subunits. Inactivation of Crm-1 leads to accumulation of rRNA in the nucleus, with the protein NMD3 being identified as an adaptor for the 60S subunits (Thomas & Kutay 2003; Ho et al. 2000; Gadal et al. 2001). Consistent with its role as an adaptor for Crm-1, NMD3 is a shuttling protein, which contains a leucine-rich NES and binds Crm-1 in a Ran-GTP dependent manner.

Additionally, Crm-1 is involved in the export of viral RNAs and tissue-specific endogenous mRNAs. Firstly, the HIV virus encodes a protein called Rev, which binds to HIV mRNA containing a structural RNA element, the Rev response element (RRE) (Fischer et al. 1995). The Rev protein contains an NES, which allows it to interact with Crm-1 to promote export of the viral mRNA. Secondly, Crm-1 is implicated in export of specific subsets of endogenous mRNAs (Natalizio & Wente 2013). For example the protein Nxf3, a tissue specific factor which is

closely related to the major mRNA export receptor Nxf1, contains an NES that allows it to recruit Crm-1 to mediate export of mRNAs (J. Yang et al. 2001). Crm-1 therefore functions to direct export of a wide variety of RNA species, via its interaction with specific adaptor proteins.

1.2.2 Exportin-t

Exportin-t was identified as a potential tRNA exporter via sequence homology with the yeast protein Los1p (Arts et al. 1998), which had been previously been implicated in tRNA biogenesis (Hopper et al. 1980; D. J. Hurt et al. 1987) and later identified as a Ran-GTP binding protein (Görlich et al. 1997). Unlike Crm-1, Exportin-t is able to directly bind the tRNA substrate and export it without the need for adaptor proteins. (Kutay et al. 1998).

1.2.3 Exportin-5

Exportin-5 was initially identified as an export receptor for the eukaryotic elongation factor 1A (eEF1A) and tRNAs (Calado et al. 2002; Bohnsack et al. 2002). Further studies demonstrated that Exportin-5 mediated export of the adenoviral minihelix-containing RNA, VA1, and that the interaction with this RNA was much higher than that with tRNA, which contain a degenerate minihelix structure. Thus Exportin-5 was proposed to function in an alternative tRNA export pathway (Gwizdek et al. 2003). Interestingly, the structure of this minihelix RNA is also shared with pre-microRNAs (pre-miRNAs). In the presence of Ran-GTP, Exportin-5 specifically binds pre-miRNAs and mediates their export, thus identifying it as an important intermediate in miRNA biosynthesis (Yi et al. 2003) .

Here we have briefly discussed how the majority of RNA species are exported from the nucleus (Fig. 1.2). By comparison, the export of mRNA proceeds through a distinct pathway, independent of Ran-GTP, via the export receptor Nxf1.

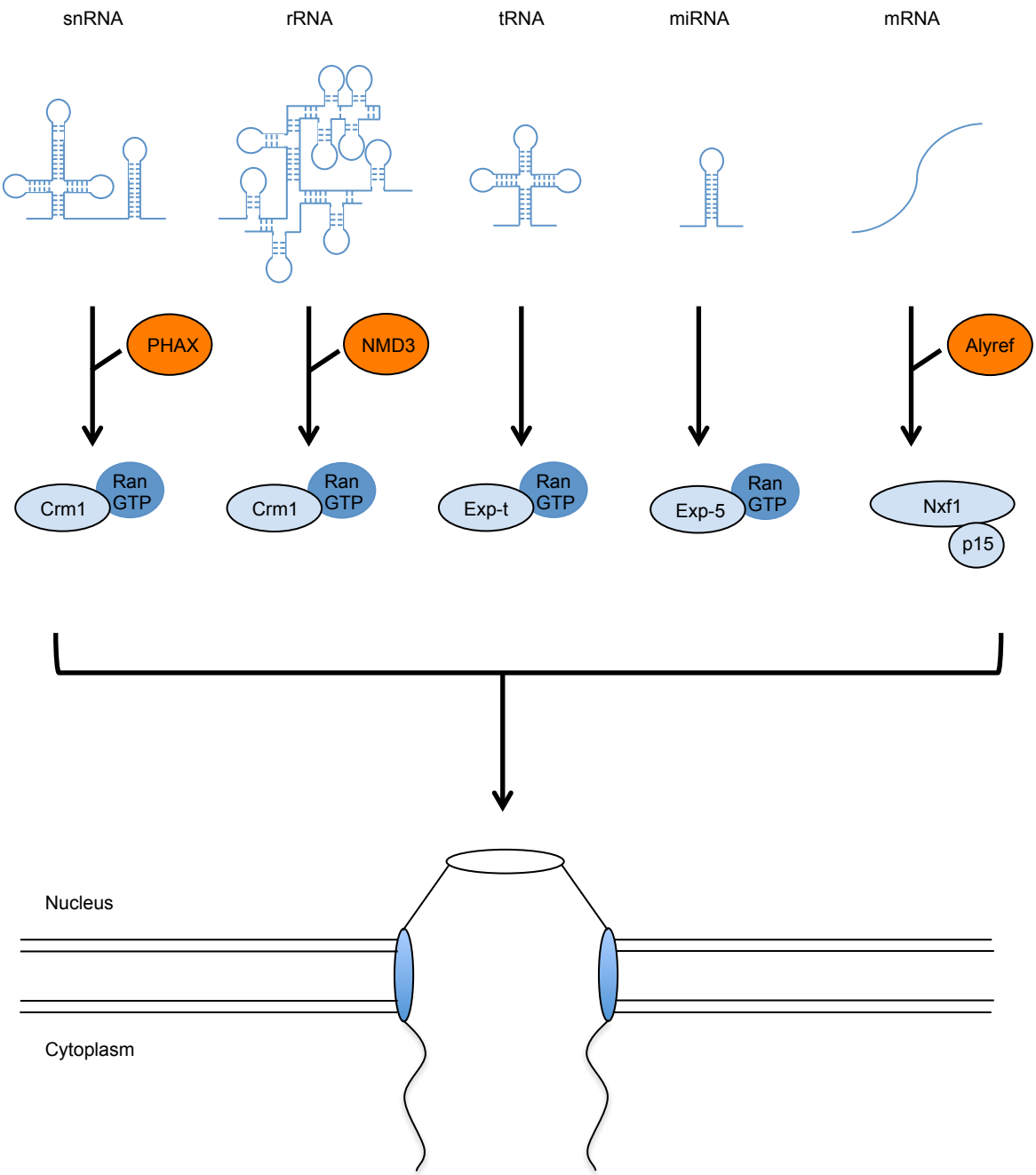


Figure 1.2. RNA export pathways. The major RNA export pathways for snRNA, rRNA, tRNA, miRNA and mRNA. Processed RNAs associate with export adaptors (orange) and export receptors (light blue), then directed to the nuclear pore for export (adapted from: Köhler, A. & Hurt, E., 2007. Exporting RNA from the nucleus to the cytoplasm. Nature reviews. Molecular cell biology, 8(10), pp.761–773.)

1.3 mRNA Export

mRNA processing involves three major events: 5'-end capping, splicing and 3'-end cleavage and polyadenylation. These events are ultimately linked to mRNA export as they serve as signals to recruit export factors. Therefore, improperly processed transcripts do not become associated with these factors and so are not exported from the nucleus but instead targeted for degradation. The first process of 5'-end capping involves the addition of a 7-methylguanosine (m^7G) to the 5' terminus. 5'-end capping occurs co-transcriptionally and has two important implications for the transcript. Firstly, the cap confers stability onto the mRNA by inhibiting 5'-3' exonuclease activity (Shatkin & Manley 2000). Secondly, the m^7G cap is bound by the cap binding complex (CBC). The CBC has been found to be important in recruiting the TREX complex (Cheng et al. 2006), and therefore plays a key role in allowing the mRNA to be efficiently exported. The nascent pre-mRNA is then subject to splicing whereby the non-coding introns are removed and the exons are joined together to produce the coding RNA sequence. During this process, specific proteins, known as the exon junction complex (EJC), are deposited onto the sites where two exons have been joined. Like 5'-end capping, the process of splicing and the proteins associated with it provides a specific signal for the recruitment of TREX complex (Zhou et al. 2000). Finally, the mRNA transcript undergoes polyadenylation at the 3' terminus. The polyadenylation machinery works by first cleaving off the most terminal 3' region of the nascent transcript then catalyses the addition of polyadenine (Proudfoot 2004). This sequence is bound by the polyadenine binding protein (PABP), which protects the mRNA from degredation (Libri et al. 2002). These processing events result in an mRNA molecule, which is properly packaged into a messenger ribonucleoparticle (mRNP), and is ready for export. Unlike proteins and other several other RNA species, which require karyopherins, export of mRNA is mediated by the heterodimeric receptor Nxf1-p15 (Mex67p-Mtr2p in yeast).

1.3.1 Nxf1-p15 mediates bulk mRNA export

Nxf1, also known as TAP (Tip-associated protein) was first identified as a protein, which interacted with the *Herpesvirus saimiri* Tip protein (Yoon et al. 1997). Subsequent studies showed that Nxf1 was an important protein involved in cellular mRNA export. This evidence primarily came from studies on the constitutive transport element (CTE) of Type D retroviruses. CTE is a cis-acting RNA element, which promotes the export of unspliced viral mRNAs by directly binding to Nxf1 (Grüter et al. 1998). Furthermore, microinjection of saturating amounts of CTE resulted in a block of cellular mRNA export, but not of other classes of RNA. This export block could subsequently be overcome by the addition of recombinant Nxf1 (Grüter et al. 1998).

Nxf1 is made up of five distinct domains: an RNA-binding domain (RBD); a pseudo-RNA recognition motif (ΨRRM); a leucine rich region (LRR); an NTF2-like domain; and a UBA (ubiquitin associated) domain (Bachi et al. 2000; Suyama et al. 2000; Liker et al. 2000), each separated by flexible linker regions. The N-terminal region, consisting of the RNP- binding domain and LRR, has been shown to be essential for cargo binding. The CTE RNA binding region has been shown to span both domains, from residue 102-372 (Braun et al. 1999). More detailed study has found that the minimal RNA binding region lies between residues 61-118 (Hautbergue et al. 2008; Zolotukhin et al. 2001). Furthermore, mutation of ten arginine residues within this region to alanine abolished Nxf1 UV-crosslinking activity to mRNA *in vivo* indicating that arginines are essential for mRNA binding (Hautbergue et al. 2008).

The C-terminal half of Nxf1, consisting of the NTF2-like and UBA domain, is required for association with the nuclear pore complex (NPC). Nxf1 forms a stable heterodimer with a small protein, known as p15, via its NTF2-like domain. Formation of this dimer is required for normal Nxf1 function, as depletion of p15 by RNAi results in a block of polyadenylated RNA export (Herold et al. 2001). Structural studies of this region have concluded that p15 contributes indirectly to NPC binding by stabilising the NTF2 fold, which directly associates with the NPC (Fribourg et al. 2001), thus allowing Nxf1 to direct mRNA to the nuclear pore for

export.

Despite being the export receptor for mRNA, Nxf1 exhibits a weak RNA binding activity on its own. Efficient association of Nxf1 with mRNA requires export adaptors, which recruit Nxf1 to the mRNP and enhance its mRNA binding activity.

1.3.2 Export Adaptors

Due to the intrinsically weak RNA binding activity of Nxf1, export adaptors, which bind RNA with high affinity, were originally thought to bridge the interaction between mRNA and Nxf1 (Liker et al. 2000). Instead of bridging the interaction however, adaptors were found to handover mRNA to Nxf1 by mutually exclusive interactions (Hautbergue et al. 2008).

The best studied export adaptor is Alyref (Yra1p in yeast). During splicing, the RNA helicase Uap56 recruits Alyref to the mRNA. Alyref can then in turn recruit Nxf1-p15 by directly binding to the N-terminal region of Nxf1. The association of Nxf1 with the Alyref-mRNA complex, triggers the transfer of mRNA directly to Nxf1. In the presence of Alyref, Nxf1 binds mRNA with an approximately 4-fold higher affinity (Hautbergue et al. 2008).

Yra1p is essential for mRNA export in *Saccharomyces cerevisiae,* with knock down of the protein resulting in an export block (Zenklusen et al. 2001). However, in humans, knock down of Alyref only results in a partial export block (Katahira et al. 2009), and likewise it is not essential for bulk mRNA export in *Drosophila melanogaster* or *Caenorhabditis elegans* (Longman et al. 2003; Gatfield & Izaurralde 2002). This indicates a functional redundancy with other export adaptors present in higher eukaryotes. Indeed, several proteins have been identified that directly bind Nxf1 and act as export adaptors. These include members of the SR (serine/arginine rich) family of splicing factors 9G8, SRp20 and ASF/SF2 (Huang et al. 2003), and the protein UIF (Uap56 interacting factor). UIF was first identified through homology with the Uap56-binding motif of Alyref (Hautbergue et al. 2009). Depletion of Alyref resulted in a significant upregulation of UIF. Consistent with it functioning in a redundant role with

Alyref, simultaneous knock down of Alyref and UIF results in an export block more severe than observed with knock down of the individual proteins.

Additionally, the TREX component Thoc5 was identified as a co-adaptor, which simultaneously binds Nxf1 with Alyref to stimulate export of specific mRNAs (Katahira et al. 2009). More recently, a newly identified TREX component called Chtop was also found to bind to Nxf1 on the NTF2L domain in a mutually exclusive manner with Thoc5 (Chang et al. 2013). Thus, the TREX complex plays an important role in mRNA export, with a number of components making contacts with the export receptor Nxf1.

1.3.3 TREX Complex

The TREX complex is well conserved from yeast to humans and is closely associated with mRNA processing events. It therefore has a key role in coupling these events to mRNA export, helping to ensure that only correctly processed transcripts are efficiently exported. Each component of the complex has a distinct function but they all must act together in order to deliver the mRNA to the export factor Nxf1 (in humans) or Mex67 (in yeast).

The conserved TREX is composed of a core THO complex, which in humans consists of Hpr1, Thoc2, Tex1, Thoc5, Thoc6 and Thoc7, and assembles with Uap56, Alyref and Cip29 in an ATP dependent manner (Masuda et al. 2005; Dufu et al. 2010). In yeast, the Tho complex contains Tho2p, Hpr1p (homologous to Thoc2 and Hpr1), Mft1p and Thp2p (Jimeno et al. 2002) and forms the TREX complex with Sub2, Yra1 and Tho1 (Uap56, Alyref and Cip29). Although well conserved, TREX appears to differ in the way it is recruited to transcripts in these two species. In yeast, TREX is directly associated with genes being actively transcribed, with the THO complex functioning in transcriptional elongation (Chávez et al. 2000; Chávez & Aguilera 1997; Abruzzi et al. 2004). In comparison, the recruitment of metazoan TREX is not directly coupled to transcription, but rather via splicing (Zhou et al. 2000). This difference is likely to reflect the relative abundance of intron-containing genes in humans compared to yeast, which largely lack introns.

Evidence has shown that TREX is recruited to the 5'-end of mRNA. This recruitment is mediated by the interaction between Alyref and the Cap binding complex (Cheng et al. 2006). However, the CBC is also present on unspliced mRNAs as 5'-end capping occurs before splicing, and mRNAs generated from cDNAs are inefficiently exported compared to the same mRNAs generated by splicing *in vivo* (Luo & Reed 1999; Cheng et al. 2006), therefore recruitment of TREX to the 5'-end alone is insufficient to target mRNA for export. Instead, additional coupling to spliced transcripts via Uap56 is necessary. Uap56 has been shown to be an important factor involved in spliceosome assembly (Fleckner et al. 1997), and interacts directly with Alyref (Masuda et al. 2005). Thus, while TREX could be recruited to mRNA via the interaction of Alyref with the 5'-cap, it is essential that further coupling with splicing occurs in order to stabilise TREX on the mRNA (Cheng et al. 2006). Indeed, mutations in Alyref, which prevent it binding to Uap56 result in a block of export, showing that this interaction is essential in coupling TREX recruitment to mRNA export (Masuda et al. 2005)

Recently, a number of putative new TREX components were also identified by mass-spectrometry. These included Zc3h11a, Skar, Elg, and Erh (Dufu et al. 2010). Although the functions of these components are relatively poorly understood, Zc3h11a and Skar were shown to associate with TREX in an ATP dependent manner and Zc3h11a depletion resulted in accumulation of poly(A) RNAs (Dufu et al. 2010). Skar (S6K1 Alyref-like target), an EJC associated protein that has homology to Alyref, has been shown to promote translation of spliced mRNAs and function in regulating cell growth (Ma et al. 2008; Richardson et al. 2004), and interacts with Erh (Smyk et al. 2006). TREX therefore has the potential to play a wider role in regulating cell processes in addition to its core function at the transcription and export interface.

1.3.4 Nuclear envelope budding-mediated mRNA export

Assembled mRNPs vary in size but can exceed the diameter of the nuclear pore, thus may require extensive remodelling to be accommodated by the NPC (Grünwald et al. 2011). Alternatively, a recent study has described an export

mechanism that bypasses the NPC altogether (Speese et al. 2012). During neuromuscular junction (NMJ) synapse development in *Drosophila*, the receptor protein DFrizzled2 is processed to release a C-terminal deletion product (DFz2C), which is transported into the nucleus where it associates into RNP granules containing mRNAs that encode synaptic proteins. These granules are surrounded by LaminC, which is a component of the nuclear lamina associated with the inner nuclear membrane (INM). Export of the granules occurs via envelopment by the INM, passage through the perinuclear space, and then fusion with the outer nuclear membrane to be released into the cytoplasm. This mechanism appears to be similar to that employed during herpes virus capsid egress reviewed in (D. C. Johnson & Baines 2011). Both herpes virus capsids and DFz2C granules are very large RNP complexes that cannot be accommodated by the nuclear pore. The export of viral RNPs in this manner may therefore co-opt an innate host mechanism for the export of mega-RNPs. The extent and functional consequences of which this mechanism is utilised by eukaryotic cells remain to be fully examined, however the authors suggest that the export of DFz2C granules via nuclear membrane budding could serve to target synaptic protein transcripts to the cellular periphery in order to restrict their translation to pre-synaptic sites (Speese et al. 2012).

1.4 Post-Transcriptional Regulation of Gene Expression

1.4.1 Degradation

As described above, expression of mRNA transcripts are subject to various levels of control that are intimately coupled. The post-transcriptional regulation of mRNA stability provides an additional layer of control that can allow rapid changes in the protein content of the cell in response to stimuli. This regulation is mRNA specific, with components of a particular functional group displaying comparable decay rates. For example, the mRNAs of transcriptional factors tend to have very fast decay rates compared to the mRNA of biosynthetic factors (E. Yang et al. 2003). Short mRNA half-lives enable the transcript levels to be adjusted more rapidly, thus allowing for more dramatic changes in the expression levels of the encoded product (E. Yang et al. 2003; Wang et al. 2002) and would be an advantage for regulatory proteins that need to respond to environmental or developmental cues. In comparison, genes involved in core metabolic functions tend to have relatively long-lived and stable mRNAs.

Decay rates are regulated by a host of RNA binding proteins, which associate with specific elements within the transcript sequence (Wu & Brewer 2012). In eukaryotes the major pathway for mRNA degradation is deadenylation-dependent. After deadenylating proteins shorten the poly(A)-tail of the mRNA, it can be degraded by the exosome complex in a 3'-5' direction. Alternatively, decapping enzymes may subsequently remove the 5' cap and allow 5'-3' decay by Xrn1 (Garneau et al. 2007).

The exosome is also involved in the decay of mRNA via surveillance mechanisms of nonsense-mediated decay (NMD) and non-stop decay (NSD). In NMD, the NMD machinery recognises a premature stop codon (PTC), which can arise from errors in splicing. During splicing, the EJC is deposited on the mRNA and several of these components remain bound after export, providing a "memory" of the splicing event. These EJC components provide a binding site for the NMD machinery and mRNAs containing a PTC are directed for degradation (Mitchell & Tollervey 2003; Le Hir et al. 2001). In cases where a transcript doesn't

contain a stop codon, the NSD machinery will target the mRNA for decay. When the ribosome reaches the 3'-end of the transcript, NSD components recognise the lack of stop codon and recruit the exosome to degrade the mRNA (Wu & Brewer 2012)

1.4.2 Microprocessor and miRNA Processing

micro-RNAs (miRNAs) are short ssRNA molecules (approximately 22 nucleotides), which are involved in regulation of gene expression at the post-transcriptional level. miRNA biogenesis involves multiple processing steps in the nucleus and cytoplasm. Most miRNAs are derived from RNAPII transcripts, several kilobases in length called primary-miRNAs (pri-miRNAs), which contain hairpin structures (Lee et al. 2004). These hairpin structures are recognised and processed by the Microprocessor complex, containing the proteins Drosha and DGCR8 (DiGeorge syndrom critical region 8). Drosha, an RNase III-type protein, is the catalytical subunit of the complex, which cleaves the pri-miRNA to release the small hairpin structures of approximately 33bp, known as pre-miRNA (Kim 2005). Drosha activity is dependent on its cofactor DGCR8, which interacts with pri-miRNAs at the junction between the single-stranded and double-stranded regions at the base of the stem-loop structure. It then directs cleavage by Drosha at approximately 11bp away from this junction(Han et al. 2004; Yeom et al. 2006; Han et al. 2006). Pre-miRNAs are then exported to the cytoplasm by Exportin-5. In the cytoplasm, pre-miRNAs are further processed by Dicer to produce mature miRNA molecules. Dicer functions by cleaving off the hairpin loop to release the ~22nt RNA duplex region of the mature miRNA. The interaction of Dicer with two other proteins (TRBP and PACT) promotes the formation of the RNA induced silencing complex (RISC). Formation of the RISC complex involves the loading of the miRNA onto an Argonaute (Ago) protein. During this process, one strand of the miRNA duplex remains bound to Ago to form RISC, whereas the other strand is degraded (Kim 2005; Kim et al. 2009).

The major process of regulation by RISC involves the association of the complex with target mRNAs via the binding of miRNA with complementary

regions within the transcript, resulting in an inhibition of translation (Zeng et al. 2002; Chekulaeva & Filipowicz 2009). Additional evidence also suggests that association of miRNA and mRNA can result in direct degradation of the mRNA (Chekulaeva & Filipowicz 2009).

While the activity of RISC appears to be the major pathway for regulation of gene expression by miRNA, there is emerging evidence that Microprocessor can directly affect gene expression upstream of the formation of mature miRNA. For instance, the DGCR8 mRNA contains two hairpins in the 5' UTR and in the coding sequence and these hairpins are targeted by Microprocessor (Triboulet et al. 2009). Cleavage of these hairpins results in a destabilisation of DGCR8 mRNA and subsequent reduction in protein, which appears to be conserved in both vertebrates and invertebrates (Triboulet et al. 2009; Han et al. 2009; Kadener et al. 2009). DGCR8 was also found to stabilise the Drosha protein, thus Microprocessor is able regulate its own activity via cleavage of the DGCR8 mRNA (Han et al. 2009). Interestingly, many of the known miRNAs are derived from exonic regions of protein-coding transcripts, and depletion of Drosha and DGCR8 was shown to upregulate several mRNAs that were not affected by depletion Dicer and RISC components (Han et al. 2009). Thus, Microprocessor may contribute to regulation of gene expression by directly affecting the stability of mRNAs.

Additionally, a recent study concluded that Microprocessor may regulate gene expression at the transcriptional level, independently of its function in miRNA processing. Drosha and DGCR8 were found to associate with the 5'-end of human genes and depletion of Drosha resulted in downregulation of transcription (Gromak et al. 2013). Drosha was able to co-precipitate with RNAPII, the cap binding complex component Cbp80, and Ars2. Ars2 is implicated in a number of RNA processing pathways and is a component of the TREX complex (Dufu et al. 2010). Microprocessor therefore, appears to regulate gene expression at multiple points in the pathway through stimulation of transcription or cleavage of mRNA, and associates with multiple RNA processing components, which may further broaden its influence.

1.5 Gene Expression and Genome Instability

The process of gene expression requires the unwinding of the DNA double helix to provide access to the template strand for the transcription machinery. This destabilisation of the DNA structure can result in genomic instability, with highly transcribed genes being particularly susceptible to mutagenesis (Nickoloff 1992).

1.5.1 TREX

As well as playing a key role in coupling mRNA processing events to export, TREX also has a significant impact on maintaining genome stability, and aberrations in human TREX expression are associated with several cancers (Domínguez-Sánchez, Sáez, et al. 2011b; Culjkovic-Kraljacic & Borden 2013; Guo et al. 2012). Hpr1 (Hyper-recombination protein 1) was first identified in yeast by a mutation that resulted in intrachromasomal recombination between DNA repeats (Aguilera & Klein 1988), as well as causing transcriptional elongation defects (Chávez & Aguilera 1997). Similarly, mutants of the other components of the yeast THO complex (Tho2, Mft1, Thp2) also stimulate hyper-recombination (Piruat & Aguilera 1998; Chávez et al. 2000). This stimulation of recombination is dependent on transcriptional elongation driven by RNAPII, indicating a strong involvement of the process of transcription in genome instability (Chávez & Aguilera 1997; Prado et al. 1997). The cause of this apparently was due to hybridisation of the nascent mRNA with the template strand of DNA, resulting in a structure known as an R-loop (Huertas & Aguilera 2003). Furthermore, viral factors have been described that recruit TREX components, hijacking this mRNA export pathway to promote expression of viral RNAs (Tian et al. 2013; Schumann et al. 2013). Sequestration of the TREX complex by the Karposi's sarcoma-associated herpesvirus (KSHV) protein, ORF57, induces DNA damage and genome instability as a result of R-loop accumulation (Jackson et al. 2014). Improper assembly of TREX through mutations or sequestration away from host transcription sites by viral factors, therefore has implications not only for mRNP

maturation and export, but also for genome instability via R-loop accumulation (Gómez-González et al. 2011; Domínguez-Sánchez, Barroso, et al. 2011a).

1.5.2 R-loops and Senataxin

RNA:DNA hybrids form naturally during a number of cellular processes involving replication and transcription. For example, during DNA synthesis, short RNA primers (~11 nucleotides) are required to initiate synthesis, as DNA polymerase cannot synthesise DNA strands *de novo*. Additionally, these hybrids necessarily form within the active site of RNAPII at the transcriptional bubble. (Aguilera & García-Muse 2012). R-loops are specific structures, which can form during transcription. They occur when the nascent RNA molecule anneals to the complementary DNA strand exposed during transcription. This process prevents reannealing of the DNA duplex, thus leaving a single stranded region of DNA exposed. R-loops are especially prevalent at G-rich regions downstream of polyadenylation sites and may serve an important function in facilitating transcription termination. In eukaryotes, transcription termination is promoted by the exonuclease Xrn2 (Rat1 in yeast), which degrades the RNA downstream of poly(A) sites. According to the "torpedo" model, Xrn2 (Rat1) rapidly degrades the RNA strand allowing it to catch up with the elongating RNAPII and trigger dissociation of the transcription machinery from the DNA template (S. West et al. 2004; Proudfoot 2011).

Senataxin is a putative RNA:DNA helicase, which functions in transcription termination. Depletion of Senataxin results in increased transcriptional read through and accumulation of R-loops downstream of poly(A) sites (Skourti-Stathaki et al. 2011). Similarly in yeast, the Senataxin homologue, Sen1 functions to promote transcriptional termination and prevent R-loop formation (Mischo et al. 2011). Thus, R-loop formation is an important intermediate in transcription termination. Senataxin (Sen1) functions to resolve R-loops and in doing so, allows the exonuclease Xrn2 (Rat1) to degrade the RNA strand and facilitate dissociation of RNAPII (Skourti-Stathaki et al. 2011).

The formation of a stable RNA:DNA hybrid in an R-loop results in a

displaced region of single stranded DNA, which is exposed and more susceptible to DNA damage (Aguilera 2002). Therefore, while R-loop formation may play a role in transcription termination, prevention of their accumulation by Senataxin is also important in order to maintain genome stability. Depletion of Senataxin in human cells results in various perturbations in RNA metabolism including transcription efficiency, splicing and transcription termination, (Suraweera et al. 2009) and mutations in the gene are associated with the neurodegenerative diseases ALS4 and AOA2 (Chen et al. 2004; Moreira et al. 2004). These mutations are clustered in the N-terminal, protein-protein interaction domain, and C-terminal helicase domain (Chen et al. 2006; Yüce & S. C. West 2013), identifying the importance of these domains for Senataxin function. Additionally, Senataxin was found to associate with RNAPII and transcription factors, and co-localise with components of the DNA damage response (Yüce & S. C. West 2013). Together, the evidence points to Senataxin functioning as a key link between gene expression and genome stability. The precise mechanisms involved, in recruiting Senataxin to sites of transcription and R-loop resolution, remain to be identified.

1.6 Perspectives and Project Aims

The evidence discussed above describes some of the important processes and factors involved during mRNA export. For efficient export to occur, each of these processes must be intimately coupled. This coupling serves as a quality control mechanism by ensuring that each event involved in the export pathway occurs efficiently only if prior upstream events have been successful. This has led to a model where mRNA is successively processed and passed on via a sequence of factors to the export receptor, where it is directed to the nuclear pore and triggering transport into the cytoplasm.

The first important event is the recruitment of the TREX complex to the nascent mRNA transcript. Proper recruitment of this complex is dependent on a number of mRNA processing events such as 5'-end capping and importantly, in metazoans, splicing. The spliceosome deposits a several proteins onto the mRNA including the export factor Alyref (Le Hir et al. 2001), which is recruited by the spliceosome-associated Uap56. Uap56 is an ATP-dependent DEAD-box helicase and recruitment of Alyref has been shown to be dependent on the ATP-bound state of Uap56 (Taniguchi & Ohno 2008). The binding of Uap56 to both Alyref and mRNA cooperatively promotes its ATPase activity, as does the TREX component Cip29. This stimulation of Uap56 activity is thought to drive the remodelling of TREX into a mature complex allowing it to stably associate on the mRNP (Chang et al. 2013). Subsequently, the ADP-bound state of Uap56 has a lower affinity for mRNA (Taniguchi & Ohno 2008), thus triggering its release. Alyref is left on the mRNA and acts to recruit Nxf1 to the mRNP.

Nxf1 is the general export receptor for most mRNAs. However, the inherent affinity of Nxf1 for RNA is weak, indicating the need for adaptor proteins to facilitate the interaction. Instead of bridging the interaction between Nxf1 and RNA, adaptor proteins directly hand over RNA to Nxf1 and enhance its binding affinty (Hautbergue et al. 2008). A recent study from the lab determined that Nxf1 exhibits an intramolecular interaction between its RNA binding domain (RBD) and NTF2-like domains (NTF2L). The TREX components Alyref and Thoc5 bind to these domains respectively and act to perturb the intramolecular

interaction, switching Nxf1 into an open conformation that binds RNA efficiently (Viphakone et al. 2012). This leads to a model where TREX is recruited to the mRNP via mRNA processing events and serves to integrate these processing events to export by switching Nxf1 to a conformation capable of binding mRNA. Further investigation would be of significant interest in understanding the structural and mechanistic details of this intramolecular interaction.

Apart from the conserved TREX complex, a number of putative components have recently been identified (Dufu et al. 2010). The precise roles of these proteins are still poorly understood but two of these, Zc3h11a and Skar, are known to affect mRNA export and processing. Skar itself is an EJC-associated protein that functions in splicing and cell growth. How these new proteins associate with TREX, and the functional consequences of this association remain to be fully realised.

TREX also has an important role in the maintenance of the genome. Disregulation of a number of TREX proteins has been associated with various human cancers (Guo et al. 2012; Domínguez-Sánchez, Sáez, et al. 2011b; Culjkovic-Kraljacic & Borden 2013). The cause of this appears to be dependent on the formation of R-loops, a specific RNA:DNA structure, which leaves a region of single stranded DNA susceptible to damage. Senataxin is an RNA:DNA helicase, which associates with several mRNA processing factors (Yüce & S. C. West 2013) and functions to prevent R-loop accumulation (Skourti-Stathaki et al. 2011). Whether or not there is a functional link between TREX and Senataxin in the maintenance of genome instability however, has not been explored.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial Strains

The following bacterial strains were used during this project:

Strain	Use
DH5α	Molecular cloning
BL21(RP)	Recombinant protein production

Growth Media

All media was sterilised by autoclaving and stored at room temperature before use. All recipes taken from Sambrooks's Molecular Cloning: A Laboratory Guide (Sambrook, 1989).

- **Luria Bertani (LB):** 10% Tryptone, 10% NaCL, 5% Yeast Extract
- **LB Agar:** As above, with the addition of 2% Agar (w/v)
- **Terrific Broth (TB):** 12g/L Trypton, 24g/L Yeast Extract, 4mL/L Glycerol, 2.31g/L KH_2PO_4, 12.54g/L K_2HPO_4

Antibiotic Selection in E. coli

To select for *E. coli* cells containing a particular plasmid, antibiotics were used as selective agents. The final concentrations of antibiotics in the media are listed:

Antibiotic	Final Concentration
Ampicillin	100 µg/ml
Kanamycin	50 µg/ml (30 µg/ml in LB Agar)
Chloramphenicol	34 µg/ml

2.1.2 Tissue Culture

Cell Lines

- **HEK-293T -** Human embryonic kidney transformed with SV40, which expresses the large T antigen
- **Flp-In 293 -** Contains a single integrated FRT site at a transcriptionally active genomic locus

Growth Media

HEK-293T cells were maintained in Dulbeccos Modified Eagle Media (DMEM), supplemented with 10% fetal calf serum (FCS), and 1% penecillin-streptomycin (all from Life Technologies).

Flp-In 293 stable cell lines were maintained as above, but with tetracyclin-free FCS (Life Technologies) and with the addition of hygromycin and blasticydin (Life Technologies).

2.1.3 Vectors

Plasmid	Features
pET-24b	Prokaryotic expression vector, T7 promoter, kanamycin resistance, 3' 6x His-tag. Used for expression of His-tagged recombinant proteins in *E. coli*, induced by IPTG
pGEX-6P1	Prokaryotic expression vector, ampicillin resistance, 5' GST-tag. Used for expression of GST-tagged recombinant proteins in *E.coli*, induced by IPTG
pCOLADuet-1	Bacterial expression vector containing two multiple cloning sites (MCS). Used for expression of two target genes. MCS1 is flanked by an N-terminal His-Tag. Used to co-express His-tagged Nxf1(371-619) and untagged p15
pcDNA5-FRT-TO	Mammalian expression vector containing an FRT site for Flp recombinase-dependent stable integration into the Flp-In 293 genome. CMV promoter, 2 tetracycline operator sequences, ampicillin resistance for plasmid selection, hygromycin resistance for stable integration selection

2.1.4 Buffers

DNA/RNA Buffers

- **TE (Tris-EDTA):** 10 mM Tris-HCl (pH 8), 1 mM EDTA (pH 8)
- **6x DNA Loading Buffer:** 0.25% (w/v) Xylene Cyanol, 0.25% (w/v) Bromophenol Blue, 30% (v/v) Glycerol
- **1x TBE (Tris-Borate-EDTA):** 90 mM Tris-HCl, 90 mM Boric acid, 2.5 mM EDTA (pH 8)

Protein Purification

- **His-tagged Protein Purification Buffers**

  Cobalt Loading Buffer: 50 mM Tris-HCl (pH 8), 1 M NaCl, 5 mM Imidazole, 0.5% Triton X-100, 10% Glycerol

  Cobalt Wash Buffer: 50 mM Tris-HCl (pH 8), 1 M NaCl, 10 mM Imidazole, 0.5% Triton X-100, 10% Glycerol

  Cobalt Elution Buffer: 50 mM Tris-HCl (pH 8), 0.5 M NaCl, 200 mM Imidazole, 10% Glycerol

- **RB100 Buffer:** 25 mM HEPES-KOH (pH 7.5), 100 mM KOAc, 10 mM $MgCl_2$, 1 mM DTT, 0.05% Triton X-100, 10% Glycerol
- **GSH Elution Buffer:** 50 mM Tris, 40 mM reduced glutathione, 100 mM NaCl
- **IP (Immunoprecipitation) Lysis Buffer:** 50 mM HEPES-NaOH (pH 7.5), 100 mM NaCl, 1 mM EDTA (pH 8), 1 mM DTT, 0.5% Triton X-100, 10% (v/v) Glycerol
- **PreScission Protease (PSP) Cleavage Buffer:** 20 mM Tris-HCl (pH 7), 150 mM NaCl, 1 mM DTT, 0.5 mM EDTA
- **TEV Protease Cleavage Buffer:** 50 mM Tris-HCl (pH 8), 150 mM NaCl, 1 mM DTT, 0.5 mM EDTA

SDS-PAGE and Western Buffers

- **4x SDS-PAGE Loading Buffer:** 200 mM Tris-HCl (pH 6.8), 1% (w/v) Bromophenol Blue, 50% (v/v) Glycerol, 10% (w/v) SDS, 10% (v/v) β-mercaptoethanol
- **4x SDS-PAGE Resolving Gel Buffer:** 1.5 M Tris-HCl (pH 8.8), 0.15% (w/v) SDS
- **4x SDS-PAGE Stacking Gel Buffer:** 0.5 M Tris-HCl (pH 6.8), 0.15% (w/v) SDS
- **SDS-PAGE Running Buffer:** 25 mM Tris, 250 mM Glycine, 0.1% (w/v) SDS

- **Coomassie Brilliant Blue Stain:** 0.1% (w/v) Coomassie Brilliant Blue R-250, 45% (v/v) Methanol, 10% (v/v) Acetic Acid
- **Destain Solution:** 45% (v/v) Methanol, 10% (v/v) Acetic Acid
- **Transfer Buffer:** 39 nM Glycine, 48 nM Tris, 0.037% (w/v) SDS, 20% (v/v) Methanol
- **ECL (Enhanced Chemi-Luminescence) Solution 1:** 100 mM Tris-HCl (pH 8.5), 2.5 mM Luminol, 400 µM p-Coumaric Acid
- **ECL Solution 2:** 100 mM Tris-HCl (pH 8.5), 5.3 mM Hydrogen Peroxide

Miscellaneous Buffers

- **1x PBS (Phosphate Buffered Saline):** 137 mM NaCl, 2.7 mM KCl, 4.3 mM NaH_2PO_4, 1.47 mM KH_2PO_4 (pH to 7.4 with HCl). Add 0.1% (v/v) Tween-20 to make **PBST**
- **1x TBS (Tris Buffered Saline):** 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. Add 2% (v/v) Tween-20 to make **TBST**
- **Transformation Buffer 1 (TFB1):** 30 mM K-acetate, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% glycerol (v/v)
- **Transformation Buffer 2 (TFB2):** 10 mM MOPS, 75 mM $CaCl_2$, 100 mM $RbCl_2$, 15% glycerol (v/v)

2.1.5 Molecular Biology Kits

- **DNA extraction from agarose gel:** QIAGEN gel extraction kit
- **Small scale plasmid purification:** QIAGEN mini-prep kit
- **Mid scale plasmid purification:** QIAGEN midi-prep kit
- ***In vitro transcription/translation*:** Promega TNT quick coupled transcription/translation system

2.2 Methods

2.2.1 Molecular Cloning

Molecular cloning is the technique of inserting a DNA sequence of interest into a plasmid vector so that it can be replicated by the host cell (*E. coli*). The sequence of interest (insert) is in most cases generated by *Polymerase Chain Reaction (PCR)*. The primers used were designed for each specific reaction, incorporating the required restriction sites and adding an additional 6 bases to the 5'-ends to provide a binding platform for the restriction endonucleases. After amplification, *Agarose Gel Electrophoresis* is carried out on the PCR reaction before visualising the DNA by UV light. The relevant DNA band is then cut out of the gel and purified by a gel extraction kit. *DNA Restriction Digestion* is then performed on the purified DNA to expose the compatible overhangs. The digested DNA is then purified from the reaction by *Phenol-Chloroform Iasoamyl-Alcohol DNA Purification* to be used in a *Ligation* reaction with a vector, which has been digested with complementary restriction enzymes. The cut vector has been treated with calf intestinal alkaline phosphatase (CIAP), which prevents the vector self-ligating by removing the 5'-phosphate group. After CIAP treatment, the reaction is stopped and the vector purified by *Phenol-Chloroform Purification.* The ligation reaction is then used to *Transform* competent *E. coli* cells which are plated on LB-agar supplemented with selective antibiotics. After incubation, the colonies that have grown are picked and used to inoculate liquid LB and the resulting culture used for *Plasmid Purification*.

Polymerase Chain Reaction (PCR)

Typically, a 50 µl PCR reaction contained 10-100 ng DNA template, 1X reaction buffer, 200 µM each dNTP, 0.5 µM forward primer, 0.5 µM reverse primer, 1 unit DNA polymerase. Reactions were then cycled according to the table below:

Step	**Cycles**	**Temperature (^{o}C)**	**Time (min)**
1	1	96-98 (Initial Denaturation)	0.5-1
2	25-30	96-98 (Denaturation)	0.5
		56-60 (Annealing)	0.5
		72 (Extension)	0.5/kb product length
3	1	72 (Final Extension)	10

Agarose Gel Electrophoresis

Agorose gels were prepared by adding 0.5-2% (w/v) agarose to 1x TBE buffer and then heated to dissolve the powder. Ethidium bromide was then mixed into the molten agarose to a final concentration of 20 µg/ml and then left to cool and set in a gel tray. Once set, the gel was placed into an electrophoresis tank which was then filled with 0.5x TBE buffer. 6x DNA loading buffer was then added to DNA samples to a 1x final concentration before loading into the wells. DNA bands were separated by running a current of 100 Volts across the gel and visualised by UV light. If the DNA is to be used for downstream experiments such as cloning, the bands can be excised from the gel and purified using the QIAGEN gel extraction kit.

DNA Restriction Digestion

Digests were carried out using the optimal buffer as recommended by the manufacturer and using a volume of enzyme not exceeding 10% of the total reaction volume to avoid issues caused by excess glycerol. For plasmids, digests were performed at 37^{o}C for 2-4 hours. For PCR products, digests were incubated overnight (approximately 16 hours) at 37^{o}C.

Phenol-Chloroform Iasoamyl-Alcohol DNA Purification

One volume of Phenol-chloroform isoamyl-alcohol (25:24:1, pH 6.8) was added to one volume of DNA solution then vortexed for 1 minute before centrifuging at 16000 rcf for 3 minutes. The upper, aqueous layer was then transferred into a fresh eppendorf tube and mixed with 1 M sodium acetate (pH 5.3) to a final concentration of 10% (v/v). 3 volumes of ethanol were added and mixed by inversion before incubating for 20 minutes at -20^{o}C to precipitate the DNA. The solution was then centrifuged at 16000 rcf for 20 minutes and the supernatant discarded. The resulting DNA pellet was washed by addition of 1 ml 70% ice-cold ethanol to remove excess salt, then centrifuged again for 10 minutes at 16000 rcf. The supernatant was discarded completely and the DNA pellet air dried for 10-15 minutes before resuspension in the required volume of dH_2O. Generally, 30 µl of dH_2O was used to resuspend 5 µg of digested vector, and 10 µl used to resuspend a digested PCR reactions.

DNA Ligation

Cut vectors were treated for 15 minutes at 37^{o}C with 1U calf alkaline phosphatase to prevent re-circularisation. 100 ng of cut vector was then incubated with purified DNA insert (at a molar ratio of approximately 3:1 insert:vector) in the presence of 1U T4 DNA ligase and 1X ligation buffer (Roche). Incubations were performed at room temperature for 2 hours, or at 16^{o}C overnight.

Making Chemically Competent E. coli

Overnight cultures were prepared by inoculating the appropriate strain in 10 ml LB then incubated at 37^{o}C. This culture was used to inoculate 500 ml LB to a final cell density of OD_{600} = 0.05 and incubated at 37^{o}C until the OD_{600} = 0.48. The cells were placed on ice for 10 minutes before being centrifuged at 6000 rcf at 4^{o}C for 10 minutes in pre-sterilised centrifuge pots. The resulting pellets were resuspended in 250 ml, ice cold, filter-sterilised TFB1 and then left on ice for 10 minutes before being centrifuged again. The resulting pellet was then gently

resuspended in 25 ml ice cold, filter sterilised TFB2 and left on ice for 10 minutes. The suspension was then aliquoted into sterile, pre-cooled eppendorf tubes before being flash frozen in liquid nitrogen, and stored at -80^{o}C.

Transforming Chemically Competent E. coli

Cells were defrosted on ice and then 50-100 µl of cells were combined with half a ligation reaction, or 10-100 ng of purified plasmid DNA. This mix of cells and DNA was incubated on ice for 20 minutes before heat-shocking at 37^{o}C for 5 minutes. The mix was then placed back on ice for 2 minutes before the addition of 1 ml LB, and then incubated at 37^{o}C for 1 hour. After incubation, the cells were centrifuged at 4000 rcf for 1 minute. 900 µl of the supernatent was discarded, with the remaining media used to gently resuspend the bacterial pellet. The suspension was then spread onto pre-warmed LB-agar containing the relevant antibiotic and incubated at 37^{o}C for 16 hours to allow colonies to grow.

Plasmid Purification from E. coli

Colonies of DH5α cells were picked from the plate and used to inoculate 1.5 ml or 50 ml (mini-prep or midi-prep respectively) LB containing the relevant antibiotic, then incubated at 37^{o}C for 16 hours in a shaking incubator (200 rpm). The following day, the inoculation was centrifuged at 4000 rcf and the supernatant discarded completely. The resulting pellet was used to purify the plasmid DNA using the appropriate QIAGEN kit according to the manufacturer's protocol.

Plasmid Construction

pGEX-6P1 (GST) Vectors

The pGEX-6P1 vector was used to produce GST-tagged, PreScission Protease (PSP) cleavable proteins. pGEX-6P1 vectors containing full length Nxf1, Nxf1(NTF2L), Hpr1, Tex1, Alyref, ThoC5, ThoC6, ThoC7, Uap56, Chtop were previously created in the laboratory. Truncations of Nxf1(RBD) were amplified

from full length Nxf1. Full length Ars2, Zc3h11a, Skar and Erh were amplified from I.M.A.G.E. cDNA clones purchased from BioSource. pGEX-6P1 was cut using *BamHI and XhoI* sites.

pET24b and pET24b-GB1 Vectors

The pET24b vector was used to produce C-terminal 6x histidine-tagged proteins. pET24b-GB1 was a vector created in the laboratory to add a N-terminal GB1 solubility tag followed by a 6x histidine tag. Full length Ars2, Zc3h11a, Skar and Erh were amplified from the I.M.A.G.E. cDNA clones listed above and ligated into a pET24b cut with *NdeI* and *XhoI* sites.

pcDNA5/FRT/TO-3xFLAG Vectors

The pcDNA5/FRT/TO-3xFLAG vector was used for transient or stable expression of N-terminally FLAG-tagged proteins in mammalian cells. Full length Senataxin was amplified from cDNA provided by M. Lavin. Full length Alyref was amplified from I.M.A.G.E. clone cDNA (BC052302) Inserts were cloned into a *NotI* and *XhoI* cut vector.

2.2.2 Molecular Biology

Expression of Recombinant Proteins in E. coli

Expression vectors (pGEX-6P1, pET24b) were transformed into competent *E. coli* strain BL21(RP) and plated on selective LB-agar. Single colonies were picked the next day and used to inoculate TB supplemented with selective antibiotics and incubated overnight at 37^{o}C whilst shaking at 200 rpm. The following day, overnight cultures were used to inoculate 2 L flasks containing 750 ml TB containing the selective antibiotics (chloramphenicol was not included as it inhibits protein synthesis). The medium was inoculated to a final OD_{600} = 0.05 and then left in a shaking incubator at 37^{o}C until OD_{600} = 0.6 - 1.0. Once this OD was reached, expression was induced by the addition of IPTG to a final concentration of 300 mM. Induction times were determined empirically for different proteins, but typically for 4-6 hours at 37^{o}C or 16 hours at 20^{o}C. After

induction, cultures were centrifuged at 6000 rcf for 15 minutes. The resulting pellet was stored at -20^{o}C until required.

Recombinant Protein Purification from E. coli

GST-Tagged Proteins

Cell pellets were resuspended in 1x PBS + 0.1% Tween (PBST) using at least 5 times the pellet volume. The suspension was supplemented with protease inhibitors before sonicating on ice at 15 microns to lyse the cells. This crude lysate was then centrifuged at 16000 rcf at 4^{o}C to pellet the cell debris and the supernatant was applied to glutathione sepharose beads. The supernatant and beads were incubated at 4^{o}C on a rotating wheel for 1 hour. The beads were then pelleted by centrifugation at 400 rcf and the supernatant discarded. The beads were then washed 3 times with ice cold PBST. After the final wash was discarded, bound proteins were eluted by the addition of GSH elution buffer at 2.5 times the bed volume of the beads and then incubated on ice for 15 minutes with regular agitation. After centrifugation at 400 rcf, the supernatant containing the purified protein was transferred to a fresh tube, being careful not to transfer any beads, and then stored at 4^{o}C or -80^{o}C depending on the downstream application.

6xHistidine-Tagged Proteins

Cell pellets were resuspended in cobalt loading buffer and then supplemented with protease inhibitors before sonication and centrifugation as described above. The supernatant was applied to Cobalt sepharose beads and incubated for 1 hour at 4^{o}C on a rotating wheel. The beads were then washed as described above with cobalt wash buffer. Bound protein was eluted by the addition of cobalt elution buffer and incubated with agitation for 30 minutes on ice. The supernatant containing the purified protein was transferred to a fresh tube and stored at 4^{o}C or -80^{o}C.

Removal of Tags by Protease Cleavage

Proteins were purified as described above by on GSH or Co^{2+} resins. After washing, the beads are equilibrated in cleavage buffer by washing an additional five times in PSP cleavage buffer (for GST-tagged proteins) or TEV cleavage buffer (6xHis-tagged proteins). Cleavage reactions are performed in 1.5 x volume of resin used with the addition of 1 µg protease per 10 µg of purified protein. Reactions were typically carried out overnight on a rotating wheel. After incubation, the samples were centrifuged at 400 rcf to pellet the beads. The resulting supernatant, containing the eluted protein was transferred to a fresh tube. The remaining beads, which contain the bound tags and protease were discarded.

SDS-PAGE Electrophoresis

Protein samples were separated according to their sizes by SDS-PAGE (Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis). Gels consisted of a lower, resolving gel and an upper, stacking gel. The percentage of acrylamide used in the resolving gel was typically 12%. For resolution of large proteins (more than 200 kDa), an 8% gel was used. Resolution of smaller proteins (15 kDa or less) a 15% gel was used. A 5% stacking gel was used for all applications. Gels were prepared by diluting the relevant 4x gel buffer in water and mixing in the appropriate volume of 30% acrylamide:bisacrylamide. The polymerisation reaction started by the addition of 0.2% (v/v) TEMED and 0.1% (w/v) ammonium persulphate before pouring into the assembled gel casts (BIORAD Mini-Protean II system). The gel was left to set for 15 minutes before pouring the 5% stacking gel into which a comb was inserted. After the stacking gel had set, the comb was removed and the wells gently rinsed with deionised water. The gel was fitted into an electrophoresis tank and the tank was filled with SDS-PAGE running buffer. Samples were prepared by boiling in 1x SDS-PAGE loading buffer for 3 minutes to denature the proteins. Samples were then loaded into the wells alongside a protein marker (PageRuler Plus Prestained Protein Ladder, Fermentas). Gels were run at 35 mA until the dye front reached the end. The gels

were then removed from their casts and the protein bands visualised by incubating them in coomassie brilliant blue stain solution whilst shaking for up to 1 hour. The stain was then removed and incubated in destain solution with shaking. The destain solution was replaced several times until the gel was clear and the protein bands visible. If the proteins are to be visualised by Western blotting, coomassie staining was not performed.

Western Blotting & Enhanced Chemi-Luminescence Detection

After proteins were separated by SDS-PAGE electrophoresis, they could be transferred to a nitrocellulose membrane, where they were detected by the use of specific antibodies. To set up the transfer, 3 pieces of Whatman paper were soaked in transfer buffer and layered on the positive electrode of a Biometra semi-dry western blot apparatus. The membrane was then soaked in transfer buffer and added on top of the paper stack. The gel cast was then disassembled, the gel placed onto the nitrocellulose membrane before adding another stack of 3 soaked papers. Using a 10 ml pipette, air bubbles were removed from the assembled stack by gently rolling across it. The negative electrode was then pressed down on top of the stack and a constant current of $2mA/cm^2$ applied to start the transfer. Transfers were carried out for 1 hour in general, but this could be increased to up to 2 hours for large proteins (larger than 130 kDa). Once the transfer was complete, the stack was disassembled and the gel discarded. The membrane was then blocked by adding a solution of 5% milk (w/v) dissolved in 1xTBST. The membrane was then incubated in the blocking solution for 1 hour at room temperature whilst shaking. After blocking, the membrane was incubated in the blocking solution, containing the appropriate dilution of primary antibody, for 1 hour with shaking. Subsequently, the primary antibody solution was discarded and the membrane was washed 3 times in TBST, for 5 minutes each wash. The membrane was then incubated for a further 30 minutes in blocking solution containing the appropriate dilution of secondary antibody. The membrane was then again washed extensively in TBST for 30 minutes with several changes of buffer.

After washing, the proteins could be detected by enhanced chemi-luminescence. This method works because the secondary antibody is conjugated to a horseradish peroxidase (HRP) molecule which catalyses the breakdown of hydrogen peroxidase (H_2O_2). The ECL solutions contain H_2O_2 and luminol, and during the peroxidase reaction the luminol will emit light at this location. To detect the light emitted, the two ECL solutions were mixed in a 1:1 ration then added onto the membrane and incubated for 30 seconds. The membrane was then exposed using the BioRad Gel-doc system. The exposure image could then be aligned against an image of the membrane showing the prestained markers to determine protein sizes.

GST Pull-down Assay

GST-tagged proteins were purified on glutathione sepharose beads as described above and used as bait for radiolabelled proteins produced by the T7 Quick Coupled Transcription/Translation system (Promega).

The transcription/translation reactions were set up on ice in RNase-free eppendorf tubes as follows: 1.5 µl plasmid DNA (200 ng), 0.5 µl ^{35}S-methionine, 8 µl rabbit reticulocyte lysate. The reaction was mixed by pipetting then incubated at 30^{o}C for 1.5 hours. During the incubation, the GST proteins were purified as previously described. The cell lysates were added to 30 µl of glutathione-sepharose beads per reaction. After washing, the beads were dried using a gel-loading tip before adding 400 µl of buffer. 8 µl of the completed transcription/translation reactions were then added into this 400 µl of buffer on the beads and incubated for another hour at 4^{o}C with rotation. All pull-downs were performed in the presence of 10 µg/ml RNaseA to ensure interactions were direct and not bridged via RNA molecules. The remaining 2 µl of the reaction was saved for the input and kept on ice for the duration of the experiment. The beads were then pelleted by centrifugation at 400 rcf, the supernatant was discarded and the beads washed 3 times as described previously. After the final wash, the beads were dried completely using a gel-loading tip before the addition of 60 µl GSH elution buffer. The beads were incubated on ice for 15

minutes with agitation to elute the bound protein. Once the incubation was complete, the beads were pelleted and the supernatant, containing the purified GST protein and binding partners, was transferred to a fresh eppendorf tube being careful not to carry over any beads.

The elutions were run alongside 0.5 µl of the input reactions, on an SDS-PAGE gel and stained with coomassie as previously described. After de-staining the gel sufficiently, the gel was dried. To detect radiolabelled binding partners the dried gel was used to expose a phosphor screen, typically for at least 16 hours. The screen was then scanned in a phosphorimager (BioRad) to produce a phosphor image.

Immunoprecipitation of FLAG-Tagged Proteins

FLAG-tagged proteins were purified from human cells by immunoprecipitation using FLAG-M2 agarose beads (Sigma). The beads were prepared by blocking with 1% BSA in IP lysis buffer and incubated for at least 2 hours prior to loading of the cell extracts, at 4^{o}C with rotation. After blocking, the beads were washed 3 times with 1 ml IP lysis buffer. Protein expression was induced by the addition of tetracycline for stable cell lines, or the appropriate plasmid was transfected into 293T cells. After 42-72 hours the growth media was removed and the cells gently washed with PBS. Cells were then lysed by the addition of ice cold IP lysis byffer, supplemented with protease inhibitors. The lysed cells and crude lysate were collected with a cell scraper and transferred into pre-cooled eppendorf tubes before centrifuging at 16000 rcf for 5 minutes at 4^{o}C. The supernatants were then transferred onto 30 µl of FLAG-M2 agarose slurry, which had previously been blocked. Beads and cell extract were incubated for 1 hour at 4^{o}C on a rotating wheel. After incubation, tubes were centrifuged at 400 rcf for 1 minute to pellet the beads and the supernatant discarded. The beads were then washed 3 times with 1 ml IP lysis buffer then dried with a gel loading tip. To elute the purified FLAG-tagged protein, 60 µl of IP lysis buffer containing 100 µg/ml 3X FLAG peptide (Sigma) was added to the dried beads and incubated with constant agitation for 30 minutes at 4^{o}C. After incubation the

beads were pelleted by centrifugation at 400 rcf and the supernatant, containing the eluted proteins transferred to a fresh eppendorf, being very careful not to transfer any beads. The eluates were then run on SDS-PAGE gels and the purified protein visualized by Western blot.

In Vivo Co-Immunoprecipitation

To find proteins that are stably associated with the target FLAG-tagged protein *in vivo,* an immunoprecipitation experiment was performed as described above. Flp-In T-REX 293 cells were grown as a negative control alongside the cell lines expressing the relevant FLAG-tagged protein, which were induced for 72 hours with 10 µg/ml tetracycline. Cells were lysed with IP lysis buffer supplemented with protease inhibitors and 10 µg/ml RNaseA to ensure that protein complexes were not bridged via RNA molecules. The protein concentrations of the cleared lysates were measured before loading equal amounts of protein from the different conditions onto the beads. After incubating and washing the beads, the eluted proteins were separated by SDS-gel electrophoresis and then transferred to nitrocellulose membrane. The membrane was then probed with antibodies to various proteins of interest.

Total RNA Extraction

To extract samples of total RNA from human cells, 293T or Flp-In stable cell lines were grown in 6 cm plates. The media was discarded and the cells washed with 5 ml pre-cooled 1X PBS per plate. The PBS was then removed and the cells lysed by addition of 200 µl Reporter Gene Lysis buffer to each dish (supplemented with protease and RNase inhibitors and pre-cooled on ice). The cells were then scraped and homogenised by gently pipetting up and down, before transferring the lysate into eppendorf tubes pre-prepared with 750 µl Trizol LS. This solution was homogenised by gently pipetting up and down and then incubated for 5 mins at room temperature. After incubation, 200 µl chloroform was added to each tube and the solution shaken vigorously for 30 secs, then centrifuged for 5 mins at 16 000 rcf at 4^{o}C. The upper aqueous phase

(~500-600 µl) was then transferred to fresh eppendorf tubes, being careful not to take the interphase or lower organic phase. Isopropanol was then added at equal volume and homogenised using the pipette, before addition of 1 µl glycogen and stored at -20^{o}C for at least 30 mins. The solution was then centrifuged again as described above for 20 mins, and the supernatant discarded. The pellet was washed with ice-cold 70% ethanol and then centrifuged again for 10 mins. The supernatant was discarded, being careful not to suck up the pellet, and the pellet left to air dry for 5-10 mins. Pellets were then left to soak in RNase-free DEPC water for 30-60 mins on ice before resuspension by pipette. RNA samples were then DNase treated to remove any contaminating DNA. DNaseI (Roche) was added to the RNA samples, as per the manufacturer's manual, then incubated for 1 hour at 37^{o}C. DNaseI was then inactivated and removed by phenol-chloroform isoamyl-alcohol purification as described previously.

Reverse Transcription-PCR (RT-PCR)

RT-PCR was carried out to produce cDNA copies of the extracted RNA. 0.4-2 µg of RNA (maximum 10 µl) was combined with 5 µM oligo(dN)$_6$ and topped up to 14 µl with RNase-free H_2O. This mixture was denatured for 5 minutes at 70^{o}C in a PCR machine, then immediately placed back on ice for at least 1 minute. Meanwhile, a master mix for the reverse transcription reaction was made. This contained, per tube, 4 µl 5x RTase buffer (Bioline), 1 µl dNTPs (10 mM stock), 0.5 µl RNase inhibitors, 0.5 µl Bioscript Reverse Transcriptase (Bioline). The mixture was homogenised then 6 µl distributed into each reaction. For the cDNA synthesis, the reactions were incubated in a PCR machine for 10 minutes at 25^{o}C, 1 hour at 42^{o}C, then 5 minutes at 85^{o}C to inactivate the transcriptase. The resulting cDNA samples were then stored at -20^{o}C.

Quantitative RT-PCR (qRT-PCR)

qRT-PCR was used to detect the relative levels of expression of specific transcripts. cDNA was produced as described above, then diluted 5x by the

addition of 80 µl H_2O. Primer stocks were prepared at 1 µg/µl. The forward and reverse primers for each specific transcript were combined to form a primer pair mixture, each primer diluted by 20x in the final mix to a concentration of 50 ng/µl. Each qPCR reaction contained 1 µl cDNA, 5 µl 2x Sensimix with SYBR Green, 1 µl primer mix and 3 µl H_2O, and were assembled on ice. The reactions were then incubated as follows:

Step	**Cycles**	**Temperature (^oC)**	**Time (min:sec)**
1	1	95 (Initial Denaturation)	10:00
2	45	95 (Denaturation)	00:10
		59 (Annealing)	00:15
		72 (Extension)	00:25

2.2.3 Cell Biology

Mammalian Tissue Culture

All cell lines were cultured in DMEM supplemented with 10% FCS, 1% (v/v) penicillin/streptomycin and grown incubated at 37^{o}C with 5% CO_2. Stable cell lines were cultured using tetracycline-free FCS, with blasticydin and hygromycin in addition to the penicillin/streptomycin selection.

Cells were maintained in T75 flasks and passaged every 3-4 days (depending on the cell line). For each passage, the media was removed and the cells washed gently with 10 ml 1x PBS. The PBS was then removed and the cells detached by the addition of 1 ml 0.25% Trypsin-EDTA solution (pre-warmed to 37^{o}C). Cells were incubated with Trypsin-EDTA for 5 mins, periodically tapping the flask to encourage the cells to detach. The trypsin was then neutralised by the addition of fresh, pre-warmed DMEM (typically 2-4 ml depending on the cell line) and cell clumps broken up by gently pipetting up and down. 1ml of this cell suspension was used to seed a new flask, which was then topped up with 13 ml fresh, pre-warmed DMEM containing the appropriate supplements for the cell line.

Transient Transfections

One day before transfection, cells were split into the required plates so that they would reach ~80% confluency at the time of transfection (for 293T cells, 200-300 µl of the cell suspension was used per 6 cm plate). Plates were then topped up with the appropriate volume of medium. To set up the transfection mixture, the appropriate amounts of pre-warmed media and DNA (see table below) were combined and homogenised by pipetting up and down. The turbofect (pre-warmed to room temperature) was then added and mixed by gently flicking the tube before incubating at room temperature for 20 mins. After incubation, the mixture was distributed drop-wise and evenly around the plate. The plate was then gently moved back and forth to evenly distribute the transfection mixture within the media, and then the plates incubated for 48 hours.

Plate Size	**Final Volume**	**DNA Plasmid**	**Medium**	**Turbofect**
6 cm	5 ml	6 µg	600 µl	12 µl

Generation of Stable Cell Lines: Flp-In T-REX 293 system

pcDNA5-FRT-TO-3xFLAG plasmid was digested with NotI/XhoI then purified by phenol-chloroform isoamyl-alcohol as described above. Alyref and Senataxin coding regions were amplified by PCR, digested and purified as described above before ligating into the cut vector. Flp-In T-REX 293 cells in 6 cm plates were then co-transfected with each plasmid construct and a plasmid encoding the FLP recombinase, at a ratio of 2.4 µg : 3.6 µg (FRT vector : FLP recombinase) as described above. The cells were incubated for 24 hours in DMEM containing tetracycline-free FCS. After 24 hours this media was removed and refreshed. After another 24 hours, this media was removed and the cells washed with 1 ml PBS before detaching the cells by addition of 500 µl pre-warmed trypsin-EDTA per plate. The trypsin was neutralised with 1 ml media and the cell suspension split between two 10 cm dishes, each topped up with 9 ml Tet-free media. The cells were allowed to adhere to the plate by incubating for at

least 4 hours before replacing the media with Tet-free media containing the selective antibiotics blasticydin and hygromycin. After two days, dead cells are removed by replacing the media and left for another five days before replacing the media again to refresh the selective antibiotics. After 1 week, most of the cells had died leaving cells that contained the stably integrated DNA. Colony formation was then monitored over the next 1-2 weeks until they were large enough to fill the objective lens under 10X magnification. At this point, the cells were trypsinised and transferred to a 10 cm plate, allowed to grow until they reached confluency before trypsinising again and transferring into a T75 flask for passaging.

Chapter 3

Characterisation of the Nxf1 intramolecular interaction.

In eukaryotic cells, the Nxf1/p15 heterodimer serves as the mRNA export receptor for fully processed mRNA. The Nxf1 protein consists of five domains: an N-terminal RNA binding domain (RBD) followed by a pseudo-RNA recognition motif (ΨRRM); a leucine-rich region (LRR); an NTF2-like domain (NTF2L); and a C-terminal Ubiquitin associated (UBA) domain (Fig. 3). Nxf1 binds mRNA through its N-terminal RBD (RNA Binding Domain) via 10 essential arginines, and then transports it through the nuclear pore. Nxf1 has an inherently weak RNA binding activity, however in the presence of TREX components Alyref and Thoc5, it's affinity for RNA is enhanced. TREX is recruited to the mRNP during transcription and mRNA processing events, thus the requirement for Alyref and Thoc5 for efficient mRNA binding to Nxf1 serves as a quality control ensuring that only processed mRNA is directed for export. During this study, we identified that the RBD was able to interact with the NTF2L domain of Nxf1 (Viphakone et al. 2012). Gel filtration analysis of Nxf1/p15 indicated that it existed as a heterodimer, thus indicating that the RBD:NTF2L interaction is intramolecular. Additionally, Alyref and Thoc5 simultaneously bind directly to Nxf1 on the RBD and NTF2L domains respectively. This investigation aimed to further characterise the RBD:NTF2L interaction and provide structural insight into the molecular switch undergone by Nxf1, between low and high mRNA binding modes.

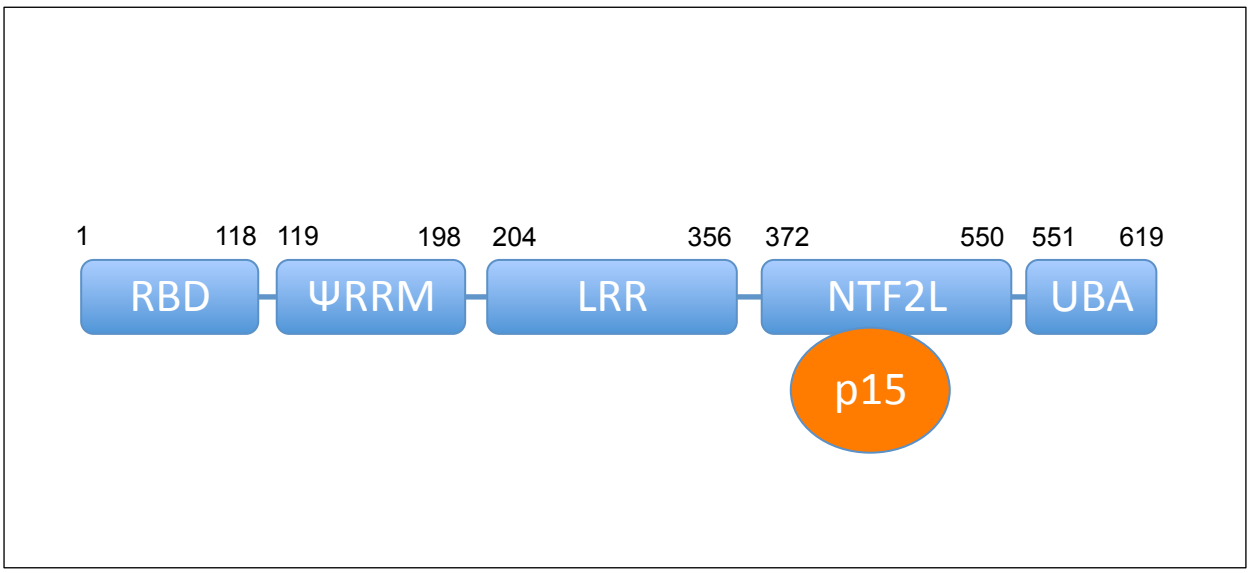


Figure 3. Domain organisation of Nxf1. Schematic of the Nxf1 domains. RBD (RNA binding domain), ΨRRM (pseudo-RNA recognition motif), LRR (Leucine-rich region), NTF2L (NTF2-like), UBA (Ubiquitin associated). Nxf1 functions as a heterodimer with p15, which binds on the NTF2L domain.

3.1 The interaction between the RBD and NTF2L domains of Nxf1 requires amino acids 50-98

The first stage of this investigation was to define the minimal region of the Nxf1-RBD, which is required to bind to the NTF2L domain. GST tagged Nxf1-NTF2L/p15 was used as bait to pull down various truncations of the Nxf1-RBD domain as well as a mutant RBD (1-198RA), which doesn't bind RNA (Hautbergue et al. 2008). The truncations and mutant proteins were expressed in reticulocyte lysate as ^{35}S-methionine labeled proteins as described in the materials and methods. The results show that the mutant RBD has a weak interaction with the NTF2L domain, indicating that the ten arginines, which are essential for RNA binding, are also required for the interaction of the RBD with the NTF2L domain (Figure 3.1.1 A, lane 8). Significantly, these arginines are all located within the region 1-118, which also binds to NTF2L (Figure 3.1.1 A, lane 3). However, Nxf1 (61-118), which contains these arginines and was shown to be sufficient to bind RNA (Hautbergue et al. 2008), does not interact with the NTF2L domain indicating that this feature of the Nxf1 RBD alone is not sufficient for the interaction with NTF2L. Further truncations of amino acids 1-118 were produced in order to better define the residues required for the interaction with NTF2L. These truncations were produced as GST-tagged fusions and used to pull down ^{35}S-methionine labeled Nxf1-NTF2L (372-550). The results appear to confirm the important role of the arginines in the interaction with the NTF2L domain as truncations that lack them exhibit weaker binding (Figure 3.1.1 B, lanes 2-4). However, as illustrated in Figure 3.1.1 A, the arginines are required but not sufficient for the interaction with Nxf1-NTF2L as deleting the first 60 residues of the RBD abolishes binding. Deletion of up to the first 50 residues however, still maintains the interaction with Nxf1-NTF2L (Figure 3.1.1 B, lanes 8 & 9). Strikingly, this region of the RBD (amino acids 50-60) contains a cluster of acidic residues (Figure 3.1.2 highlighted in blue; 53E, 54E, 55D 56D, 58D), which may play an important role in the interaction, thus we identified amino acids 50-98 as the minimal region required for the Nxf1-RBD to bind Nxf1-NTF2L domain.

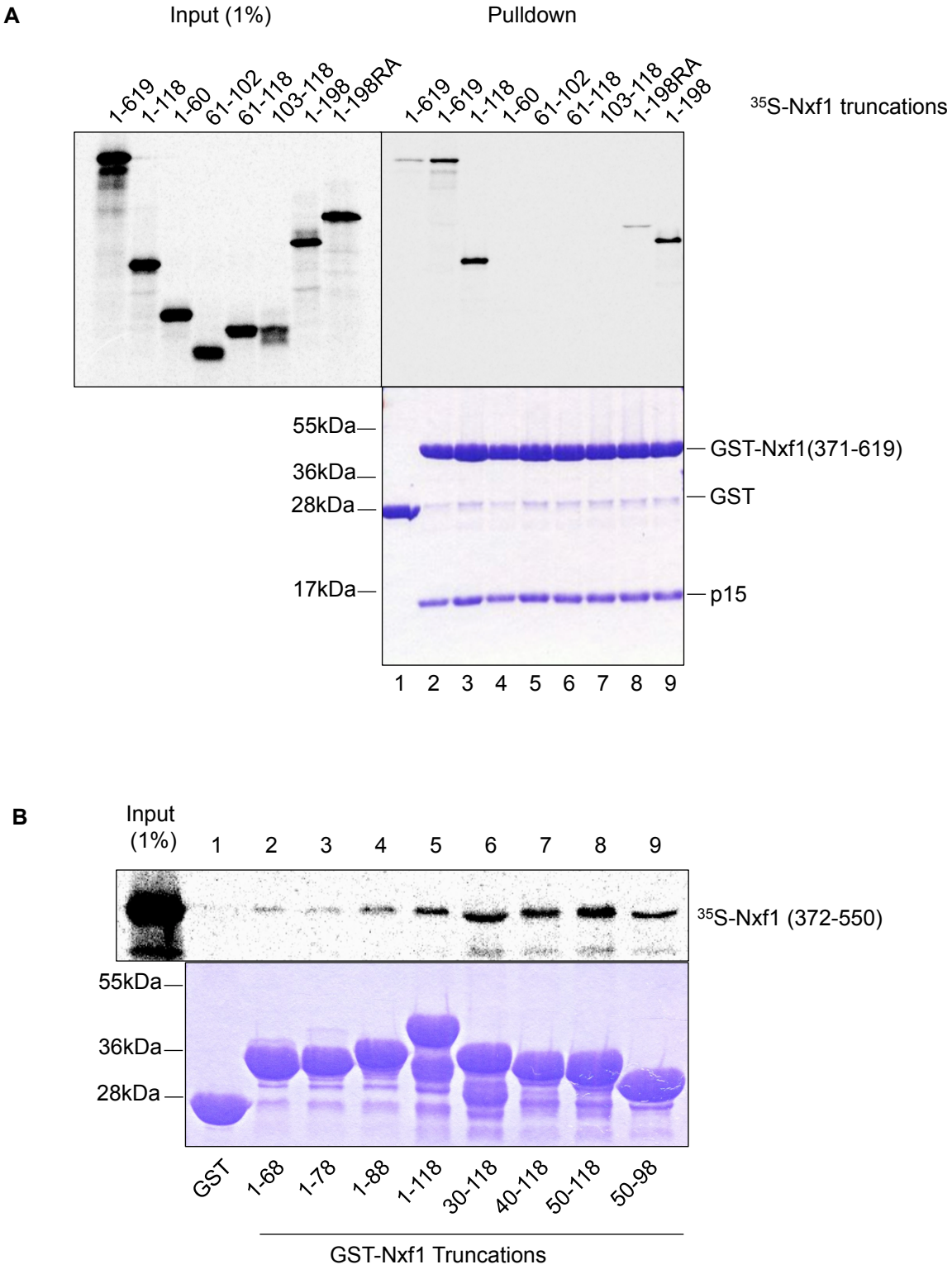


Figure 3.1.1. Identification of the minimal NTF2L-binding region of Nxf1. (A) The minimal binding region lies within aa1-118 and requires the 10 arginines necessary for RNA association. **(B)** Further truncations of Nxf1(1-118) were used to identify aa50-98 as the minimal region required to interact with Nxf1-NTF2L.

C

```
1                                                  50
MADEGKSYSE HDDERVNFPQ RKKKGRGPFR WKYGEGNRRS RGGGSGIRSS RLEEDDGDVA

61                                          98                  118
MSDAQDGPRV RYNPYTTRPN RRGDTWHDRD RIHVTVRRDR APPERGGAGT SQDGTSKN
```

Figure 3.1.2 Identification of the minimal NTF2L-binding region of Nxf1. Sequence of Nxf1(1-118). Binding of this region to the NTF2L domain requires aa50-98. Nxf1(50-98) contains 8 of the essential arginines required for both RNA and NTF2L interaction (highlighted in blue). aa51-60 contain an acidic patch of residues that also appear to be necessary for the interaction with NTF2L (highlighted in red)

Together with the observation that the RBD arginines are required for interaction with Nxf1-NTF2L, the acidic patch of residues suggests that there is an ionic component involved in this interaction.

3.2 The NTF2L domain uses a patch of basic residues to bind RBD

Given the apparent requirement for charged residues on the RBD domain for the interaction with NTF2L domain, we next looked at defining the NTF2L residues required for this interaction. By studying the 3D crystal structure of the NTF2L domain (Fribourg et al. 2001), we were able to identify a number of residues whose charged side chains were exposed and not involved in interactions within the core structure of the domain (Fig. 3.2.1, highlighted). Significantly, these residues were also previously identified, via alanine substitutions, as important for the interaction between Nxf1 and Thoc5 (Katahira et al. 2009). Additionally, work in in the laboratory has shown that one of these alanine mutants (NTF2L 456-459A) enhances the interaction with Nxf1-RBD (Viphakone et al. 2012). To assess the impact that these charged residues had on the RBD:NTF2L interaction, mutations were generated in the NTF2L domain to reverse their charges. The R453D mutant and the K456D/R459D double mutant both showed significantly reduced binding to the GST-RBD (1-118) (Figure 3.2.2 B, lanes 5 and 6). Together with the evidence that the mutant NTF2L(456-459A) displays an enhanced interaction with RBD (Viphakone et al. 2012), we have identified this surface of the NTF2L domain as the specific platform for binding RBD, thus key for the regulation of Nxf1 intramolecular interaction.

3.3 Crystallisation trials with Nxf1-NTF2L/p15 and RBD

With the aim of producing a structural model of the Nxf1 intramolecular interaction, we employed a number of strategies to purify Nxf1 for crystallisation screens. Firstly, 6xHis-Nxf1(371-619)/p15 was expressed and purified on cobalt beads before cleaving off the histidine tag using TEV protease as described in the

A

```
      408  412                              434  438                  453 456   459      468
      |    |                                |    |                    |   |     |        |
GDRQGLLDAYHDGACCSLSIPFIPQNPARSSLAEYFKDSRNVKKLKDPTLRFRLLKHTRLNVVAFLNELP
                                                                  AAAA  AAAA
```

B

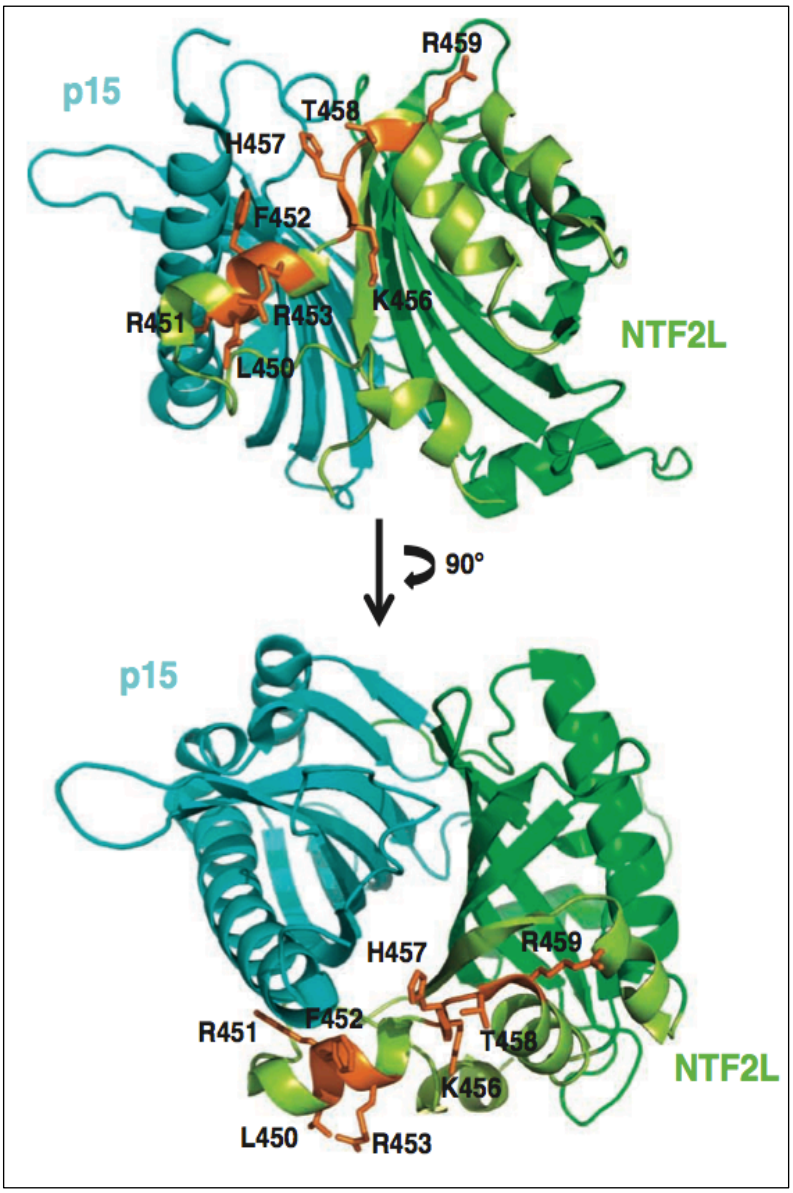


Figure 3.2.1. The Thoc5 binding region of Nxf1-NTF2L. (A) Sequence of the NTF2L region associated with Thoc5 binding. Highlighted residues were targeted for mutagenesis. aa450-453 and 456-459 (underlined) were previously identified as being crucial for Thoc5 interaction via Alanine mutations (Katahira, J. et al., 2009) and 456-459A enhances the interaction with RBD (Viphakone, N. et al., 2012). **(B)** Positions of the highlighted residues on the NTF2L:p15 structure (Fribourg, S. et al., 2001). (Figure B from, Viphakone, N. et al., 2012)

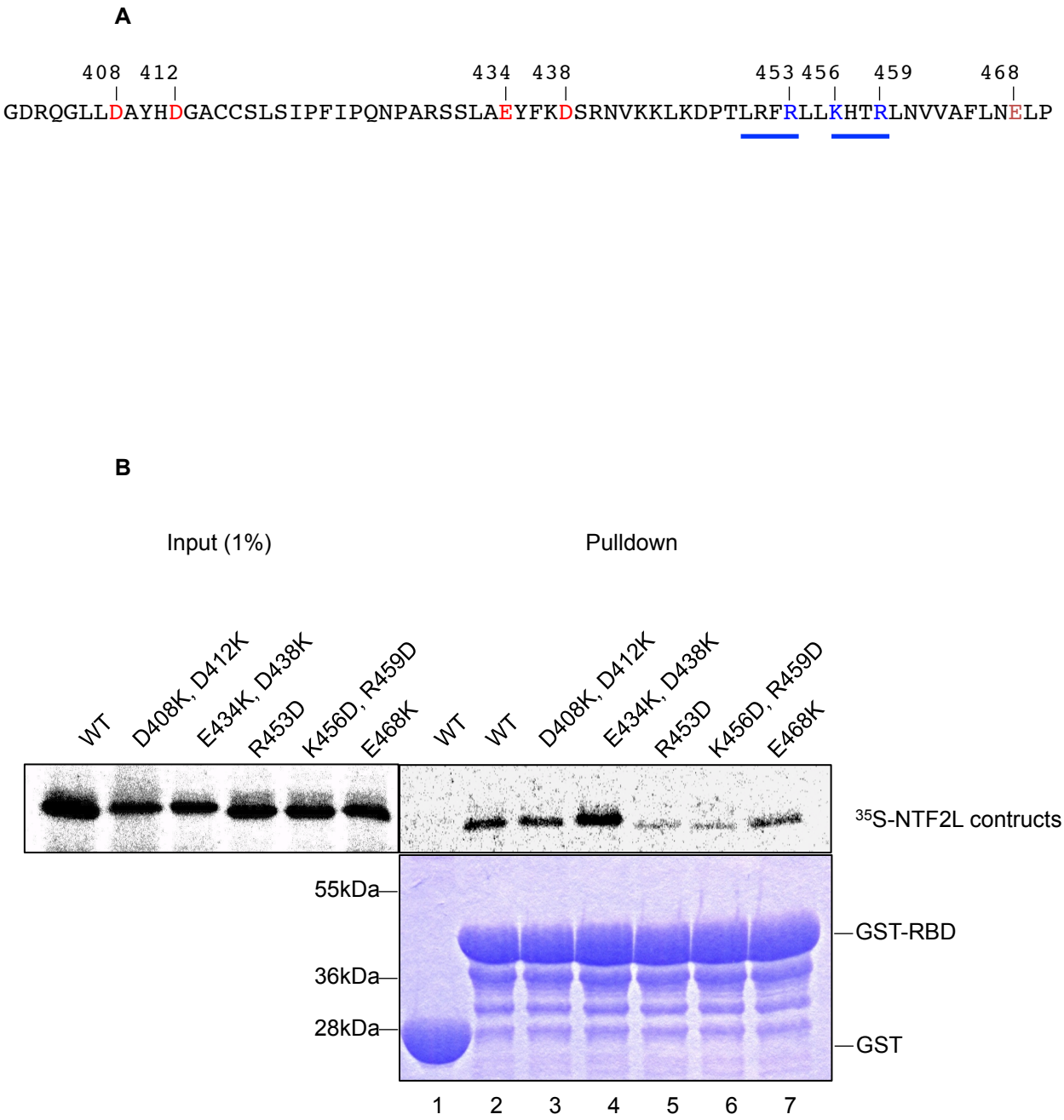


Figure 3.2.2 Identification of the NTF2L residues involved in binding RBD. (A) Sequence of NTF2L region, with the residues targeted for mutagenesis highlighted. Underlined are the residues essential for Thoc5 binding. **(B)** Pull down of NTF2L mutants with GST-RBD. Mutation of R453 and R456 + R459 perturb the interaction with Nxf1-RBD, indicating that the Thoc5 and RBD binding region overlap.

materials and methods. The resulting eluate containing Nxf1(371-619) and co-purified p15 was further purified on a MonoQ anion exchange column (Fig. 3.3.1 A). The fractions containing the Nxf1/p15 complex were concentrated in a Vivaspin centricon (30 kDa cut off) to a final concentration of 12 mg/ml. This was then used in hanging drop crystallisation screens, using a range of conditions around those previously published (Fribourg et al. 2001). The minimal binding region Nxf1(50-98) was also purified with the aim of using in a soaking procedure with any crystals that developed to form the complex. The Nxf1(50-98) protein was however highly insoluble when cleaved from the GST-tag and precipitated on a MonoS cation exchange column (GE Healthcare). Nxf1(1-118) proved to be more soluble and was successfully purified on the MonoS column (Fig. 3.3.1 B) but precipitated and was lost upon concentration.

We attempted to overcome this issue of precipitation by having two short RBD domain peptides synthesized, which proved sufficiently soluble for crystallisation screen. One peptide was the isolated acidic patch (EEDDGD) the other spanning the region containing the essential arginine residues (81-100), described above. Nxf1(371-619)/p15 was purified as before, then used in robot crystallisation screens with or without each peptide at a 40x molar excess, using the PACT and JCSG suites (QIAGEN). The screens were incubated for two weeks before checking for crystal formation, however all conditions described were unsuccessful.

In addition to the above strategies, the plasmid containing the Nxf1(371-619)/p15 used above was mutated to insert a restriction site into which the Nxf1(1-118) sequence plus a C-terminal linker of nine residues was cloned. The aim was to produce a fusion protein of Nxf1(1-118) and Nxf1(371-619) separated by a short linker sequence, GGGGSGGGG (Assenberg et al. 2009) thus eliminating the need for soaking or co-crystallisation experiments. To test whether the RBD and NTF2L components of the fusion protein behaved as expected in this context, it was expressed in reticulocyte lysate as a radiolabelled protein and used in a pull down against either GST-Nxf1(50-118) or GST-Nxf1(371-619)/p15. Binding was observed in both cases confirming that the two domains were

functioning correctly (Fig. 3.3.2). The expression of this protein from *E. coli* was, however, very low and contained many contaminants after purification thus was not suitable for crystallisation screens.

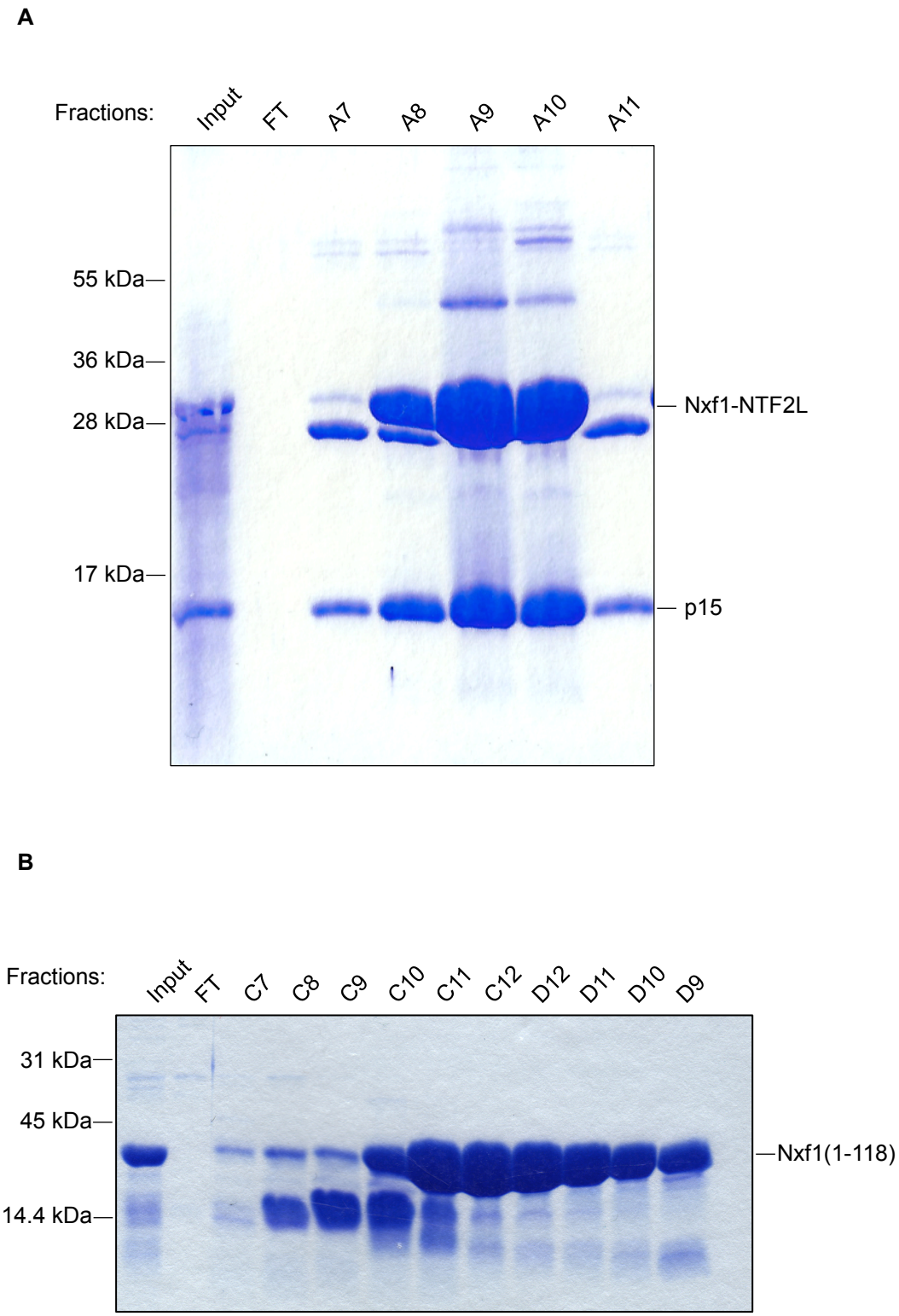


Figure 3.3.1 Purification of Nxf1 domains for crystallographic study. (A) Fractions obtained from anion exchange of Nxf1-NTF2L. Protein was purified under the conditions described in {Fribourg:2001vp}. **(B)** Purification of Nxf1(1-118) by cation exhange in 20 mM MES (pH6), 100 mM NaCl.

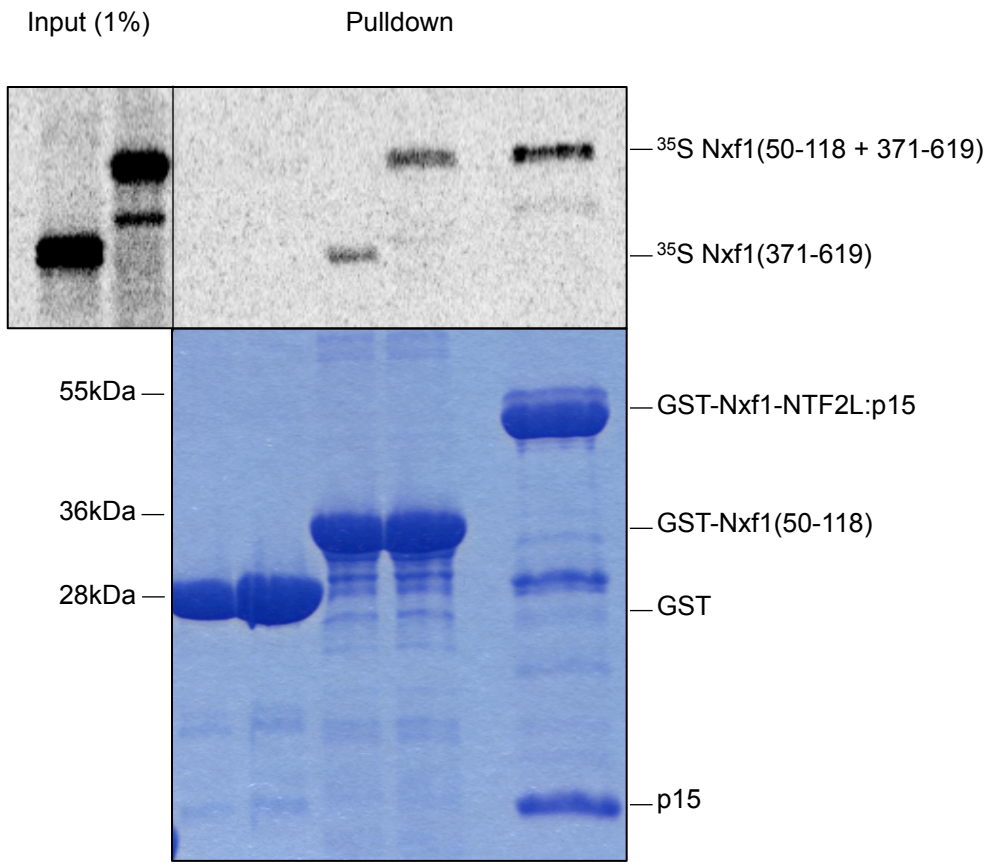


Figure 3.3.2. Construction of an RBD-NTF2L fusion protein for crystallographic study. A fusion protein consisting of the minimal NTF2L binding domain (aa50-118) and NTF2L domain. The fusion protein was used in a pull down against GST-Nxf1-NTF2L and GST-Nxf1(50-118) to confirm that the individual domains were functional

3.4 Summary

In this chapter, we have identified some key features involved in the Nxf1 intramolecular interaction. The pull down experiments, indicate that the minimal region of the RBD required to interact with the NTF2L domain lies within 50-98. Within this region lie two key features which appear to play an important role in the interaction; an acidic patch of residues (EEDDGD) and the arginines essential for the RNA binding activity of Nxf1. The fact that these arginines are involved in the interaction underlies the inherently weak association between Nxf1 and RNA, as the binding of RBD onto NTF2L may occlude these binding sites for the RNA. Additionally, the mutational analysis of the NTF2L domain identifies that the sites for RBD and Thoc5 binding overlap. Consequently, Thoc5 can disrupt the Nxf1 intramolecular interaction and in conjunction with Alyref, will fully unlock Nxf1 to allow efficient mRNA binding. The apparent ionic nature of the intramolecular interaction is striking, however the observation that the NTF2L domain requires a patch of basic residues is counterintuitive to the fact that the minimal RBD region contains multiple basic arginines. The RBD region of Nxf1 is absent in any published structures of the protein, indicating a highly unstructured or flexible region. It is, therefore, conceivable that the intramolecular interaction of Nxf1 is more complex than RBD binding to NTF2L directly via the charged residues and may require a structural rearrangement of the RBD itself, which is favorable to bind NTF2L.

Structural data of Nxf1 would elucidate the exact nature of the intramolecular interaction, however, published structures and the work from this investigation highlights the difficulty of this task. The flexible nature of Nxf1 may play a role in the unsuccessful crystal screens, as dynamic molecules will hinder the crystallisation process due to an entropic barrier. The inability to replicate crystals of Nxf1-NTF2L from published conditions though, may suggest that optimisation of the purification and crystallisation is required, but due to time constraints this could not be explored further.

Chapter 4

The Human TREX Complex

The TREX (Transcription-Export) complex is comprised of the stable core Tho complex and associated proteins. Many of the proteins are well conserved from yeast to humans, underlining its functional importance. In humans, the conserved TREX complex comprises of Alyref, Uap56, Cip29 and the core Tho complex (Hpr1, Thoc2, Tex1, Thoc5, Thoc6 and Thoc7). Recently, a host of other proteins have been identified as putative TREX components in humans (Dufu et al. 2010). Furthermore, the EM structure of the yeast Tho was solved providing the architecture and localisation of the various components within the complex (Peña et al. 2012). Despite evidence from IPs showing the stable association of human Tho, the architecture of this complex remains unsolved.

4.1 Assembly of the human Tho complex

With the aim of purifying a recombinant human Tho complex for EM structural analysis, two protein purification strategies were used. Firstly, we used a baculovirus expression system, called MultiBac, which had been previously set up in the lab. This system allows for the insertion of multigene cassettes into a bacterial artificial chromosome (BAC), which is then integrated into a baculovirus genome for expression of multiprotein complexes by infection of insect cells (Trowitzsch et al. 2010). For our experiments, a MultiBac baculovirus was used to express Hpr1, Thoc5, Thoc6, Thoc7 and Tex1 (ΔThoc2). Thoc2 was expressed as a 6xHis-tagged protein in a separate baculovirus. ΔThoc2 and 6xHis-Thoc2 were expressed by infecting separate cultures of SF9 cells. Cell lysates were then mixed and incubated to allow complex formation, before incubating on cobalt beads to pull down the 6xHis tagged Thoc2. After elution, complex formation was assessed by coomassie stain and Western blot. The results show that co-

purification of any subunits with ThoC2 could not be observed by a coomassie stained gel (Fig. 4.1.1 A). The Western blots, however, do show that Hpr1, Thoc5 and ThoC7 were pulled down with ThoC2 (Fig. 4.1.1 B). Additionally, no difference was observed in the ability of these subunits to bind full length ThoC2 or ThoC2 (1-1186). ThoC6 and Tex1 were not detectable by Western blot.

The second strategy was to express the Tho components (apart from ThoC2, which was expressed from SF9 cells as described above) as GST-fusions. We noted a significant contaminating band of approximately 57 kDa co-purified with GST-Thoc6 and Tex1 (Fig. 4.1.2). This was likely to be an *E. coli* chaperone protein, which may interfere with protein assembly downstream. We therefore aimed to remove this contaminating protein by an additional wash step using ATP (Thain et al. 1996). This additional wash step was able to reduce the amount of the co-purifying protein (Fig. 4.1.2, lanes 3 & 4), although it was not completely removed and when protein expression was up-scaled, a significant amount of it remained. GST fusions of Thoc5, Thoc6, Thoc7 and Tex1 were purified and cleaved from their tags before mixing and combined together with Thoc2 expressed in SF9 cells onto GSH beads, which contained bound, uncleaved Hpr1. After incubation and washing, GST-Hpr1 was cleaved using PreScission protease to release Hpr1 and any bound proteins, which were visualised by coomassie stain and Western blot. The coomassie gel indicated that most of the proteins were not pulled down with Hpr1, being detectable in the flow through fraction as well as subsequent washes (Fig. 4.1.3 A). This was confirmed by the Western blot which showed strong signals in the flow through. Some Thoc7 was detected in the cleaved fraction but this was not associated with Thoc5 (Fig. 4.1.3 B). The inability for us to assemble a complete, stoichiometric Tho complex could be explained by the low protein expression, as seen in the baculovirus expression system, or contaminating proteins such as those that co-purified with GST-Thoc6 and GST-Tex1, which may interfere with protein-protein interactions. Additionally, other TREX components may be involved in Tho assembly, which we did not include in our experiment.

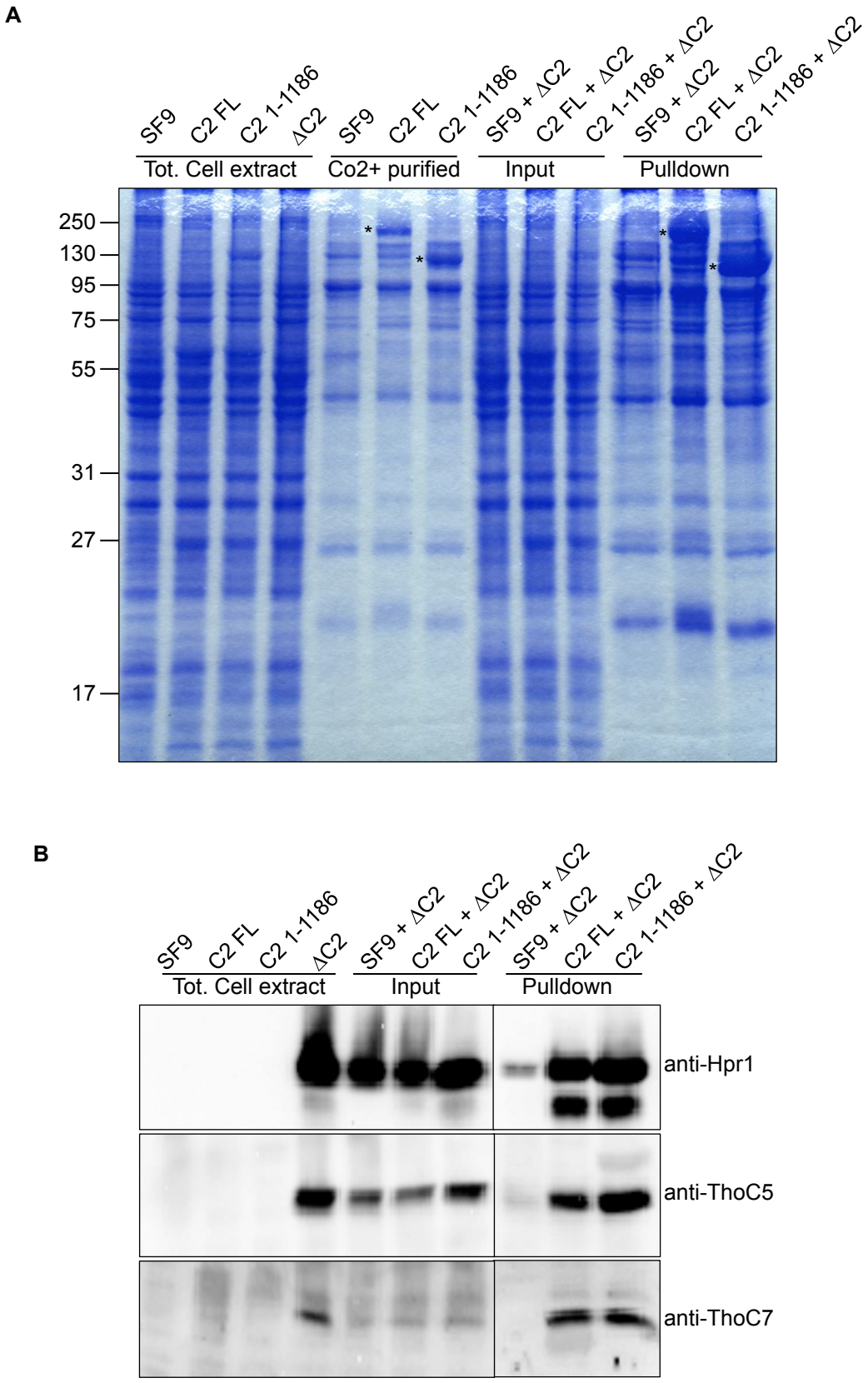
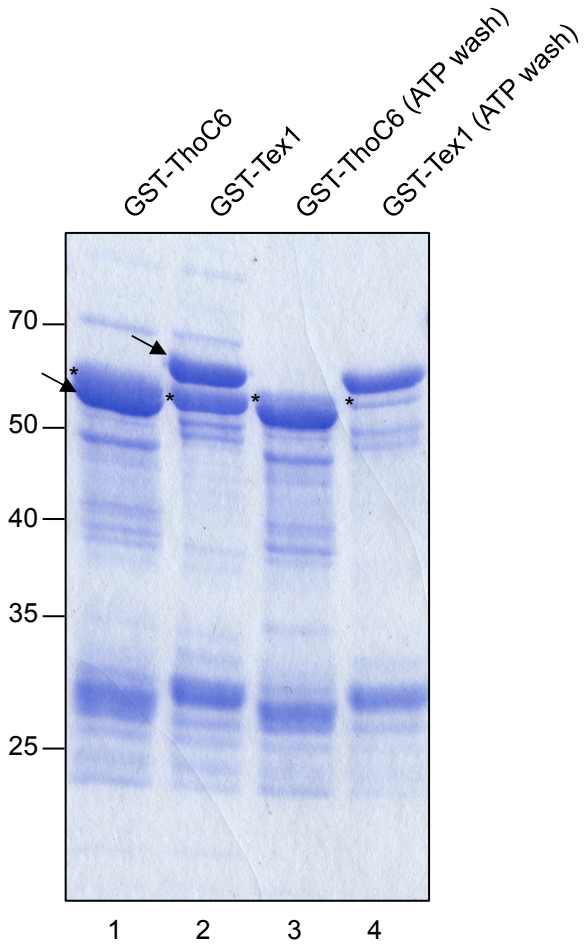


Fig. 4.1.1 Assembly of the human THO complex expressed from the Baculovirus system. **(A)** ΔC2 (Hpr1, Tex1, Thoc5, Thoc6, Thoc7) lysate was combined with lysates from SF9 cells, or SF9 cells expressing 6xHis-tagged Thoc2 or Thoc2(1-1186). Thoc2 proteins (asterisked) were pulled down on Co^{2+} resin. Co-purifying ΔTho components were undectable by Coomassie stain. **(B)** Western blot identified presence of co-purifying Hpr1, Thoc5, Thoc7.



<u>Fig. 4.1.2</u> Removal of GST-Thoc6 and GST-Tex1 contaminating proteins. A protein of approximately 57 kDa (asterisked) was observed to co-purify with GST-Thoc6 and Tex1 (respective positions indicated by arrows). Inclusion of an additional wash step with an ATP wash buffer (50mM Tris 7.5, 150mM NaCl, 20mM MgCl, 5mM ATP, 1% TritonX) reduced the amount of contaminant that co-purified

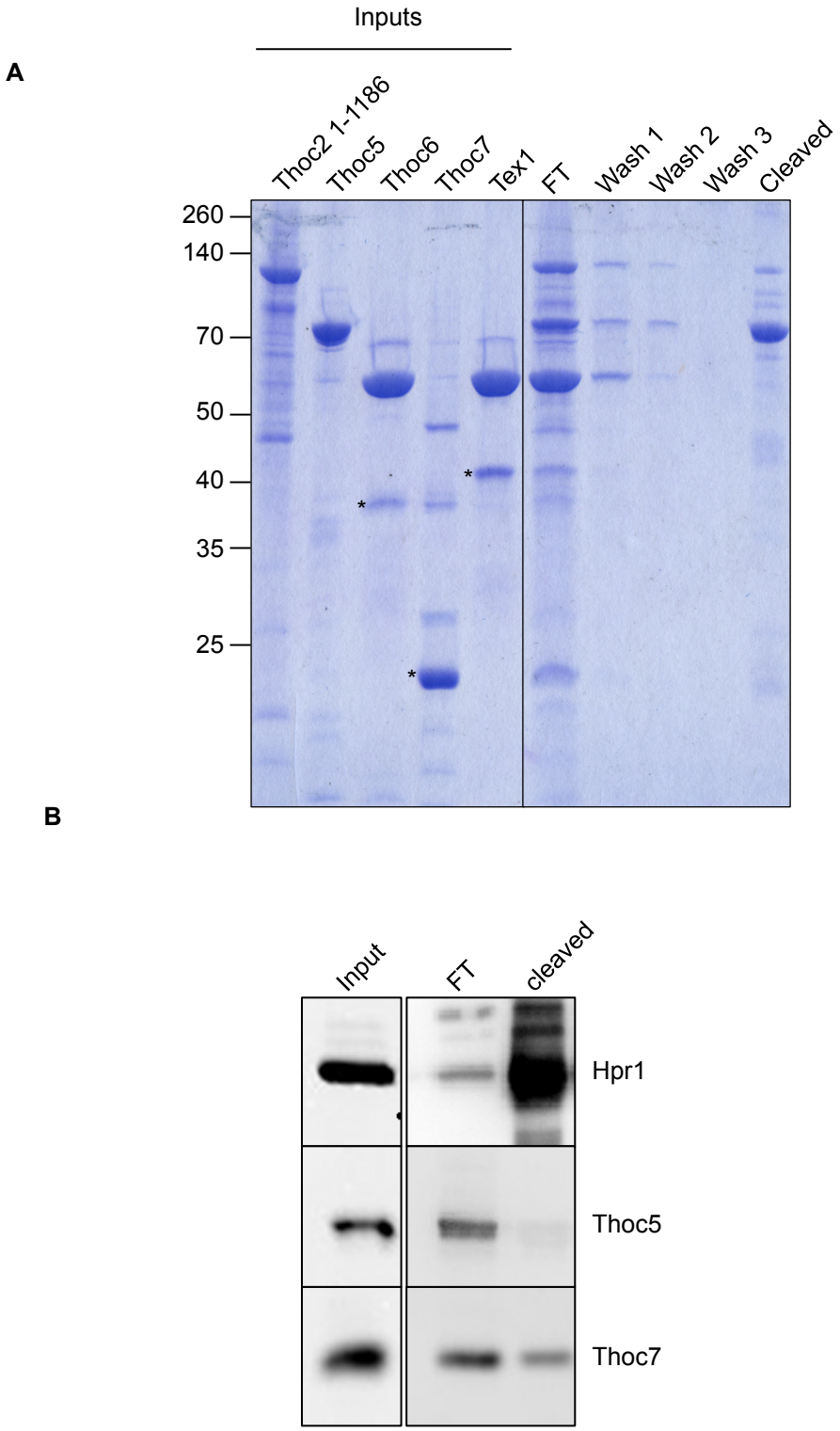


Fig. 4.1.3 Assembly of human THO using purified recombinant proteins. **(A)** *E.coli*-expressed, GST-tagged Thoc5, Thoc6,Thoc7 and Tex1 were purified and cleaved from GST tags, combined with Thoc2(1-1186) purified from baculovirus before incubating with GST-Hpr1 (bands corresponding to Thoc6, Thoc7 and Tex1 are asterisked). Hpr1 was cleaved from GST tag to elute complex. Coomassie stained gel indicates majority of proteins lost in flow-through. **(B)** Western blot on Hpr1, Thoc5 and Thoc7.

4.2 Developing a binary interaction network within TREX

To create a map of the direct binary interactions within TREX, the recently identified putative TREX components were cloned into pGEX6P1 and pET24b vectors. pGEX6P1 plasmids were used to express GST fusions which were used as the bait in pull downs against reticulocyte expressed pET24b plasmids. Interactions were detected by exposing a phosphor screen as detailed in the materials and methods. To detect ThoC2 interacting partners, the GST fusions were used to pull down ThoC2 purified from SF9 insect cells and the presence of ThoC2 detected by Western blot.

As a negative control, GST was used to pull down radiolabelled TREX components (Fig. 4.2.1 A). We noted that non-specific interactions were detected with Alyref and Skar, and was taken into account for the analysis of the following pulldowns (Fig. 4.2.1 B-Q).

Using the results from the pull downs, we generated a protein-protein interaction network in the cytoscape platform. We sorted the interactions as high or medium confidence. High confidence interactions were defined by reciprocal binding between two components, for example GST-Alyref pulls down Uap56 and GST-Uap56 pulls down Alyref. A medium confidence network was generated for interactions that occurred in at least one condition.

Consistent with the current literature, the results indicate a strong link between the subunits of the core TREX complex (Fig. 4.2.2). The Tho components largely exhibit strong interactions primarily with the Tho subunits or the conserved TREX components Uap56 and Alyref, and fewer strong interactions with the peripheral components. For this analysis, Hpr1 was included as a high confidence interaction as it is known to associate with TREX, despite a lack of reciprocal binding. Likewise, Thoc2 is included, as reciprocal binding cannot be examined due to the nature of the experiment. Both Thoc2 and Hpr1 are known integral components of Tho, and interestingly, Thoc2 and Hpr1 interact extensively with much of the TREX components. This may be particularly relevant given that Tho2 and Hpr1 depletions cause the most severe defects in yeast, thus

may be essential for proper THO/TREX assembly.

Conversely, the high confidence network of the wider TREX complex (Fig. 4.2.3) appears to indicate that peripheral, putative TREX components display extensive interactions between themselves, Uap56 and Alyref, and less with the core Tho subunits. Indeed, the interaction network of the wider TREX complex appears to highlight Uap56 and Alyref in a central role within TREX, as a focal point for many interactions with the core Tho and putative TREX components. We also generated a medium confidence interaction, which included the non-reciprocal binding events (Fig. 4.2.4). This network expands the interactions between peripheral TREX components and the core Tho subunits.

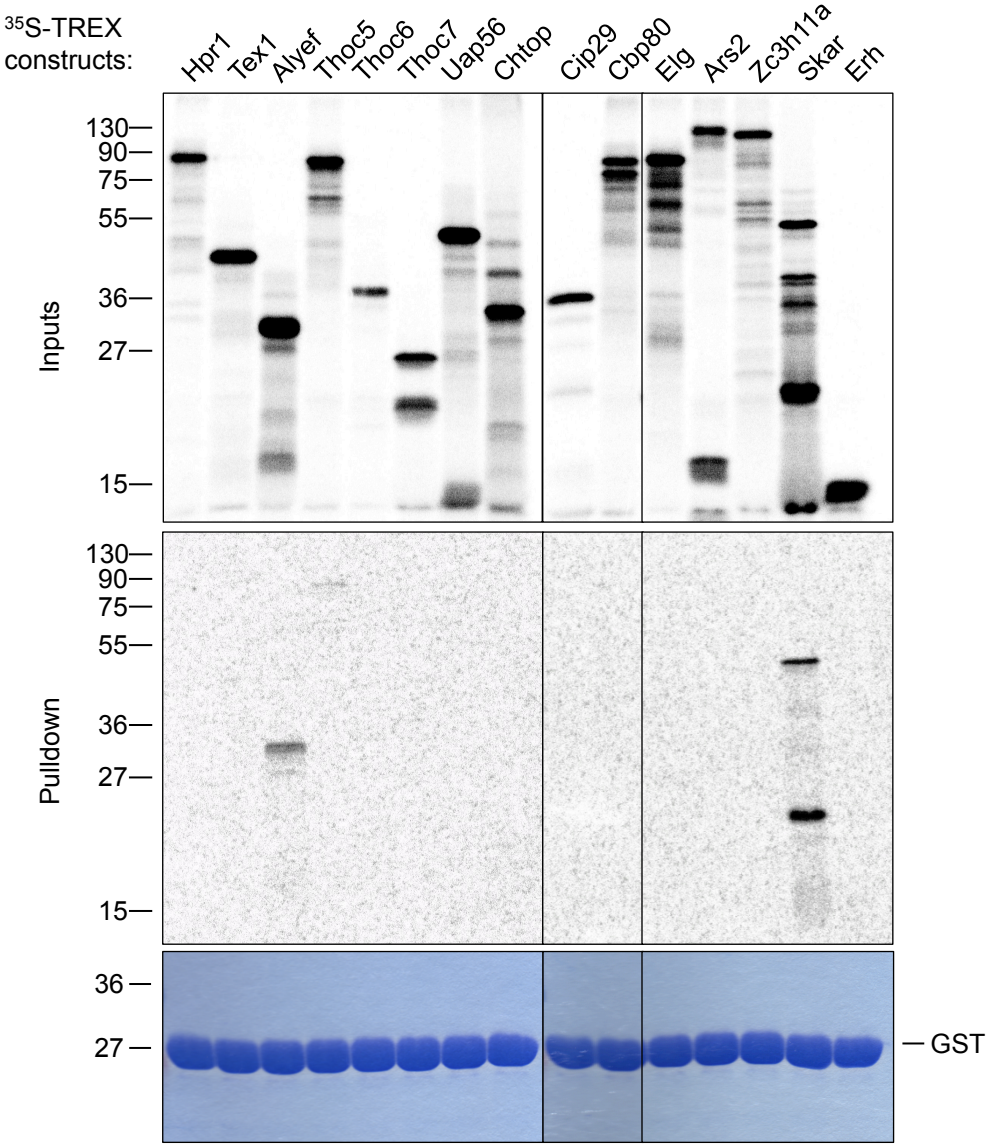
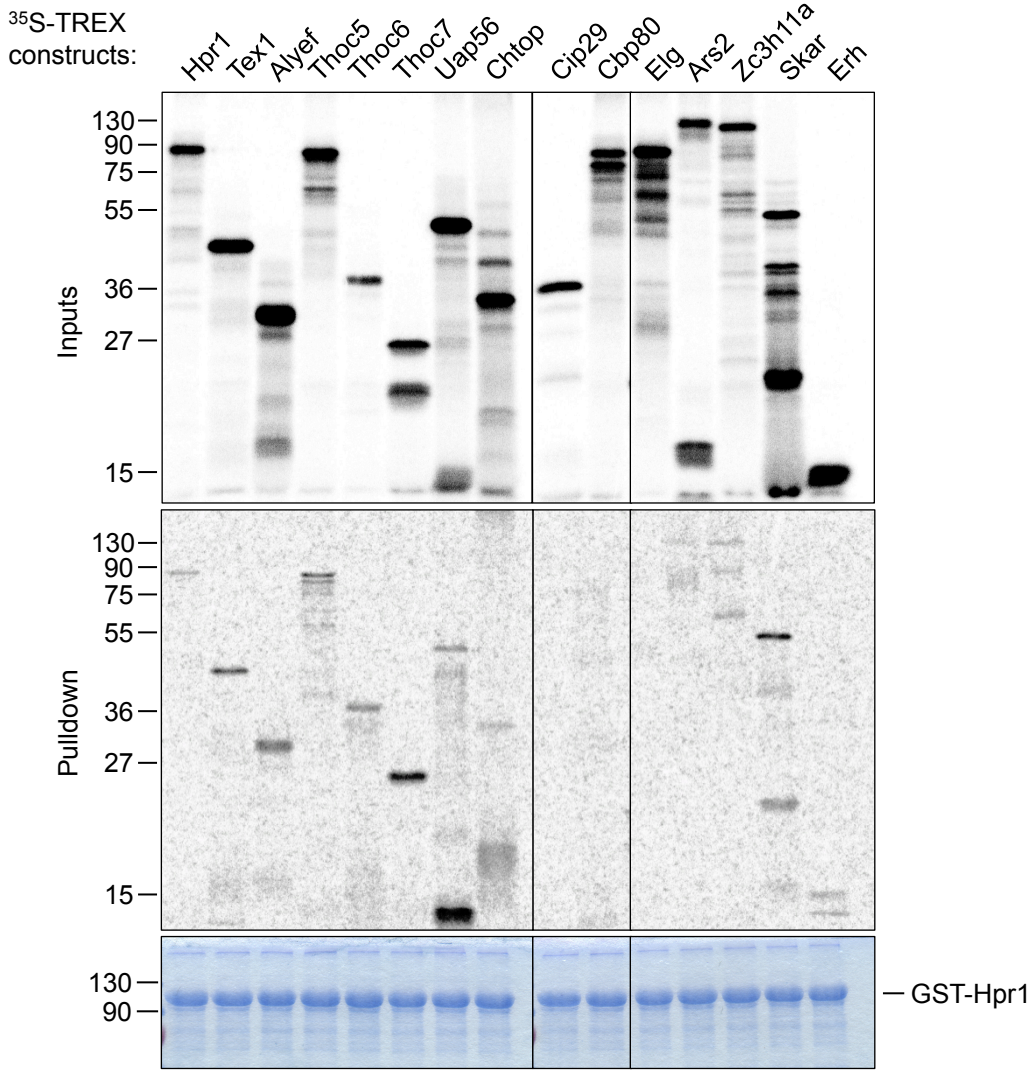


Fig. 4.2.1 A TREX binary interactions *in vitro*. (A) GST-control. Alyref and Skar display some non-specific interaction with GST



<u>Fig. 4.2.1 B</u> TREX binary interactions *in vitro*. (B) GST-Hpr1. GST-Hpr1 displays weak but specific interactions with Tho components Tex1, Thoc5, Thoc6 and a stronger interaction with Thoc7.

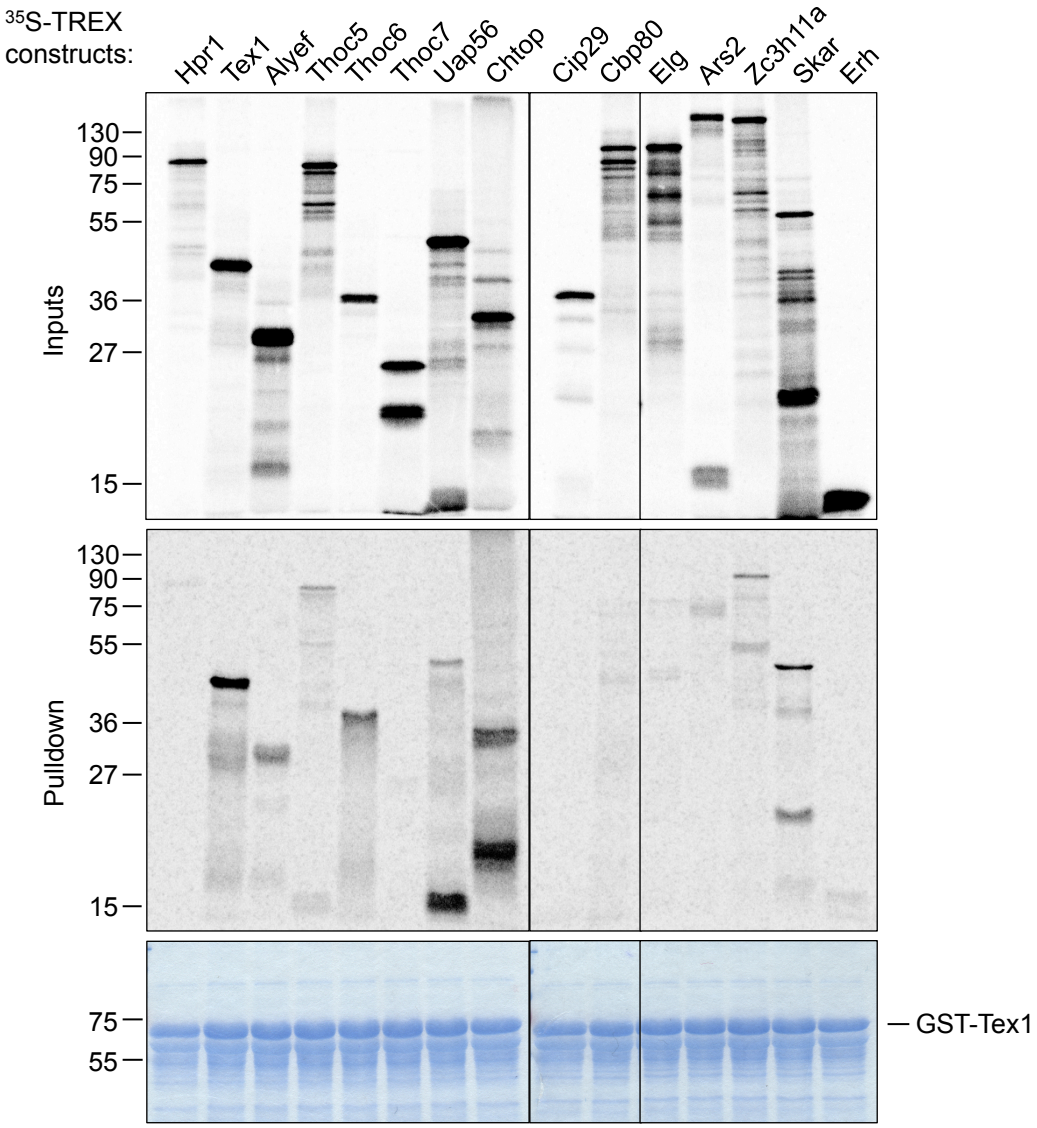


Fig. 4.2.1 C TREX binary interactions *in vitro*. (C) GST-Tex1. GST-Tex1 displays a significant self interaction. Binding is observed with Thoc5, Thoc6, Uap56, Chtop and Zc3h11a. The binding of Alyref and Skar are likely a non-specific interaction with GST

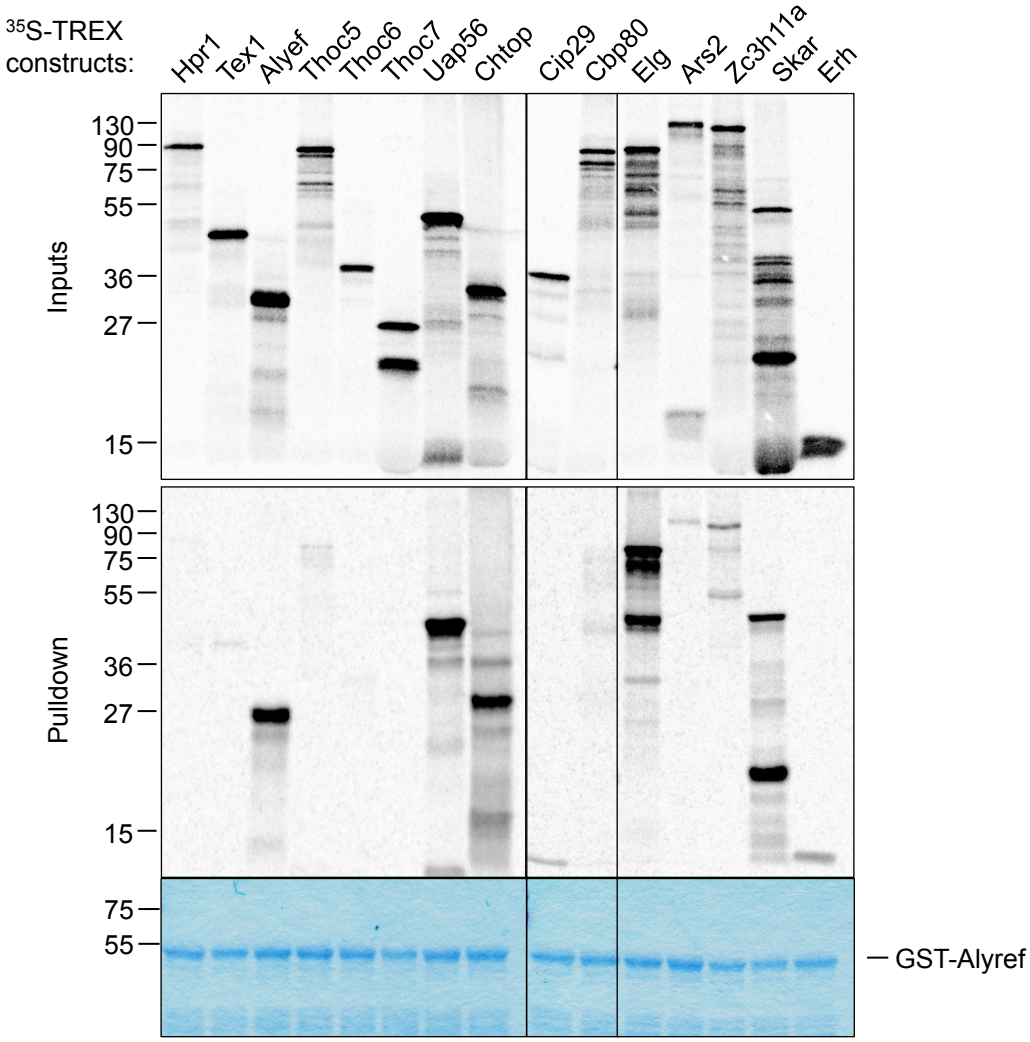


Fig. 4.2.1 D TREX binary interactions *in vitro*. (D) GST-Alyref. GST-Alyref displays strong interactions with Uap56, Chtop, and Elg. Weaker interactions are also observed with Ars2 and Zc3h11a.

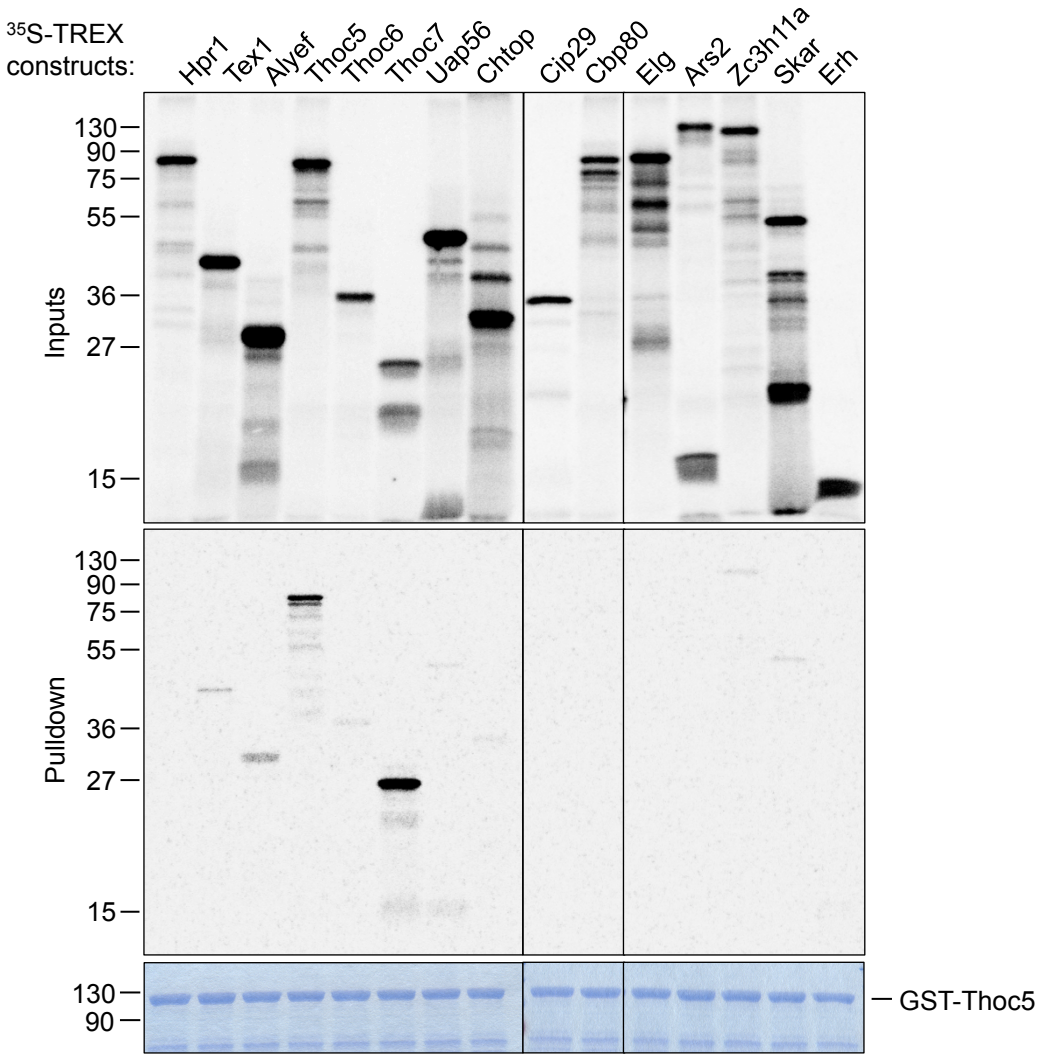


Fig. 4.2.1 E TREX binary interactions *in vitro*. (E) GST-Thoc5. GST-Thoc5 displays a strong interaction with Thoc7 as well as itself. There is a weak interaction visible with Tex1.

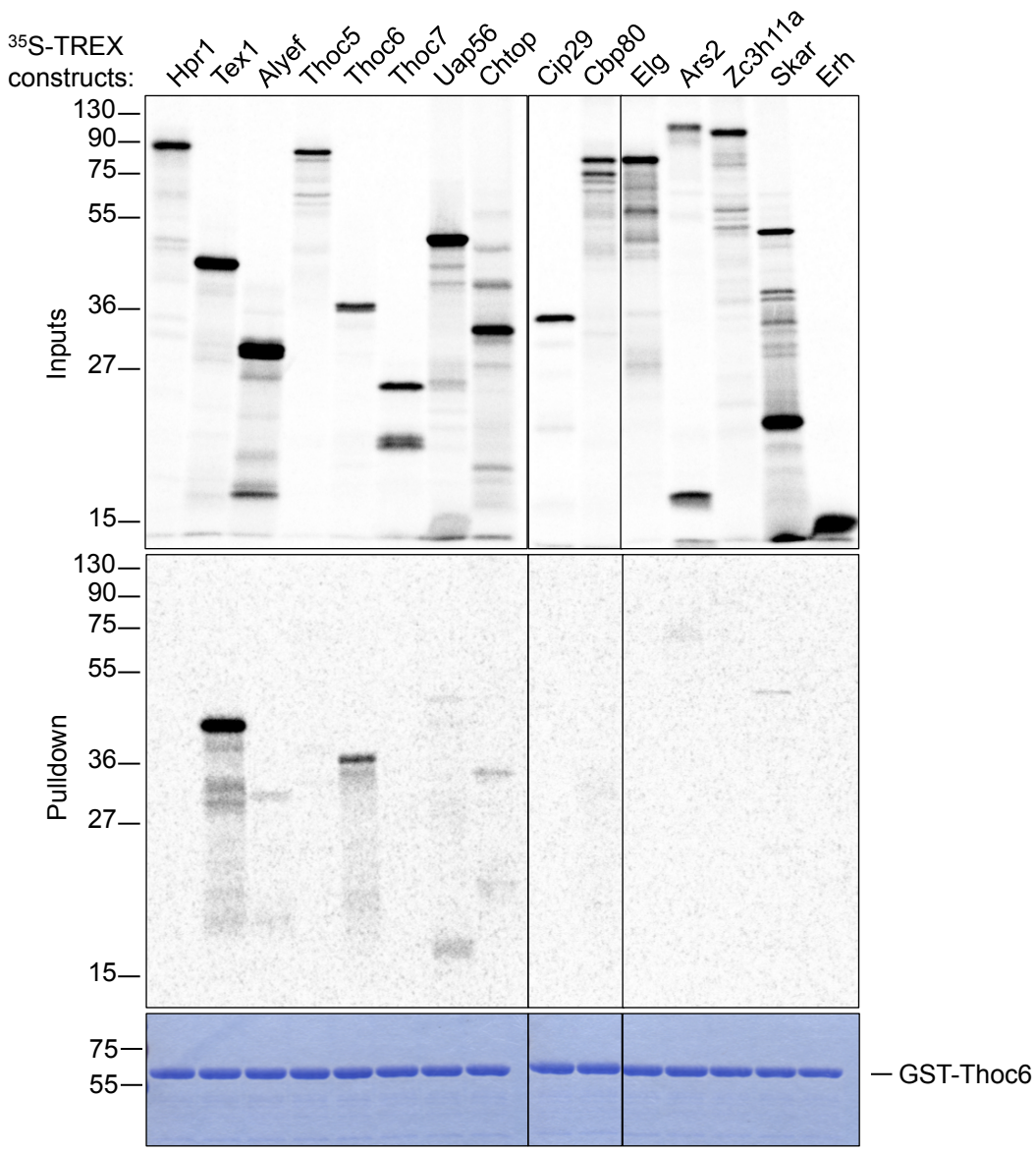


Fig. 4.2.1 F TREX binary interactions *in vitro*. (F) GST-Thoc6. GST-Thoc6 interacts strongly with Tex1 and itself. Weak interactions with Uap56 and Chtop are also observed.

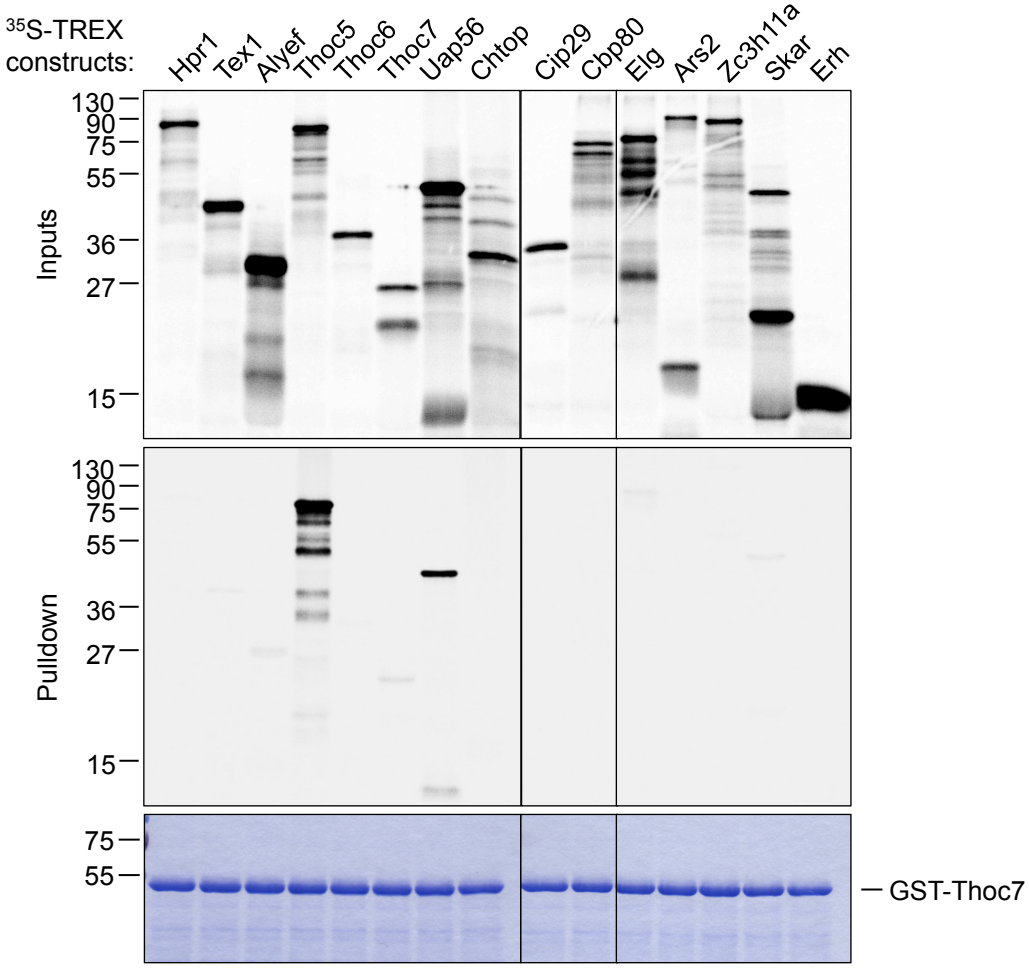


Fig. 4.2.1 G TREX binary interactions *in vitro*. (G) GST-Thoc7. GST-Thoc7 interacts specifically with Thoc5 and Uap56.

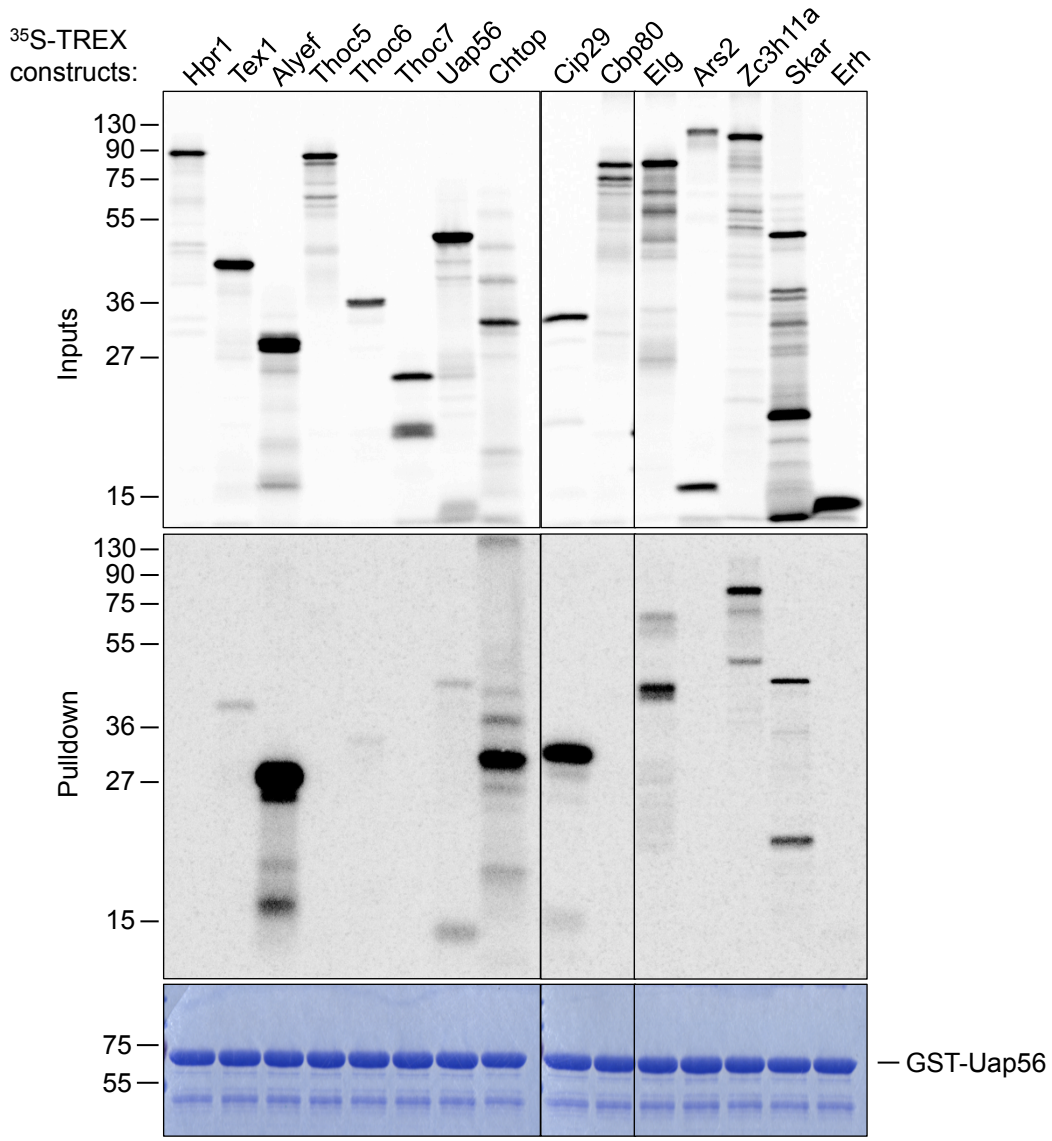


Fig. 4.2.1 H TREX binary interactions *in vitro*. (H) GST-Uap56. GST-Uap56 strongly interacts with Alyref, Chtop, Cip29 and Zc3h11a. Weaker interactions with Tex1 and Elg are also observed.

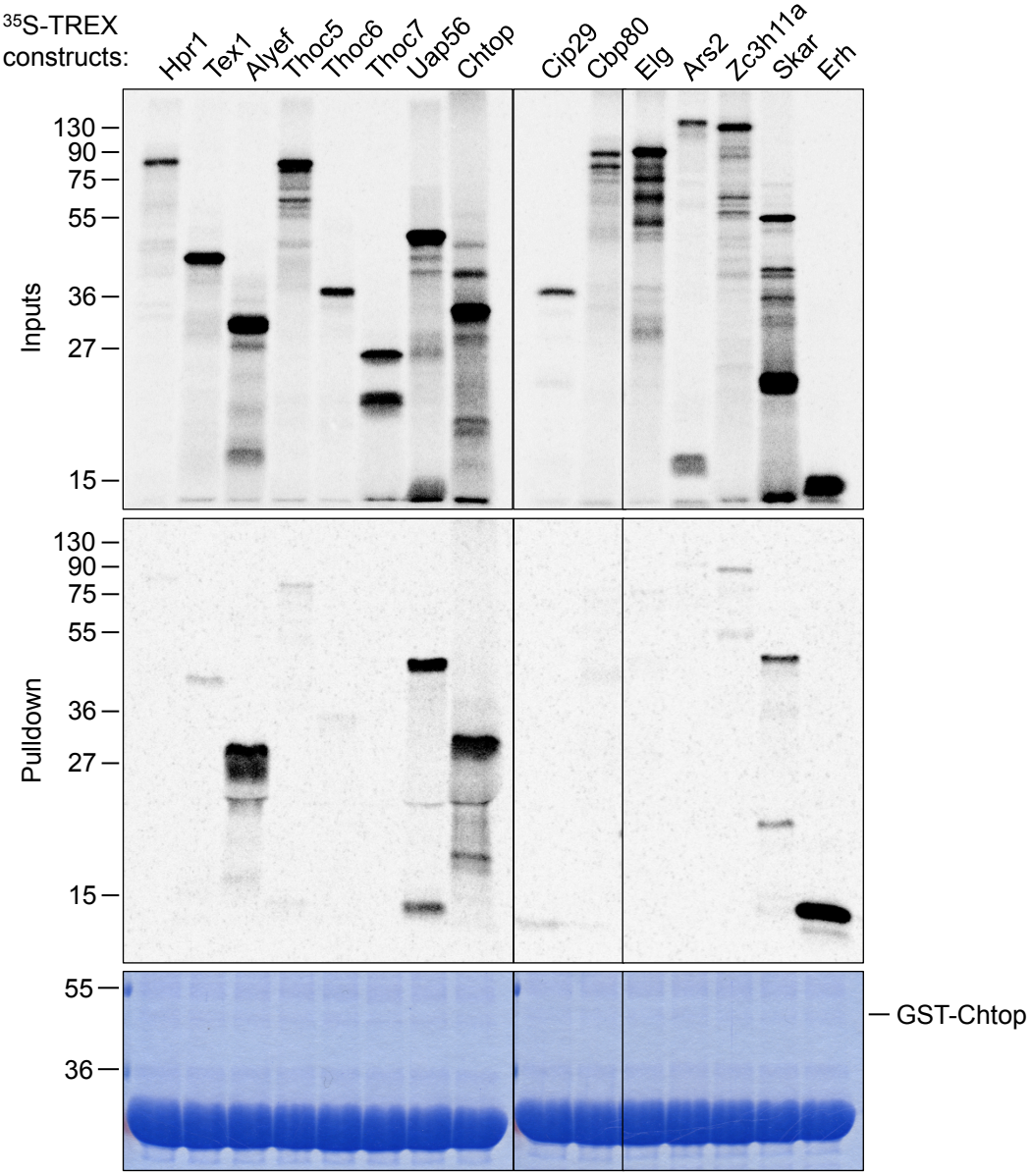


Fig. 4.2.1 I TREX binary interactions *in vitro*. (I) GST-Chtop. GST-Chtop interacts strongly with Alyref, Uap56 and Erh. Weaker but specific interactions are also observed with Tex1, Thoc5, and Zc3h11a.

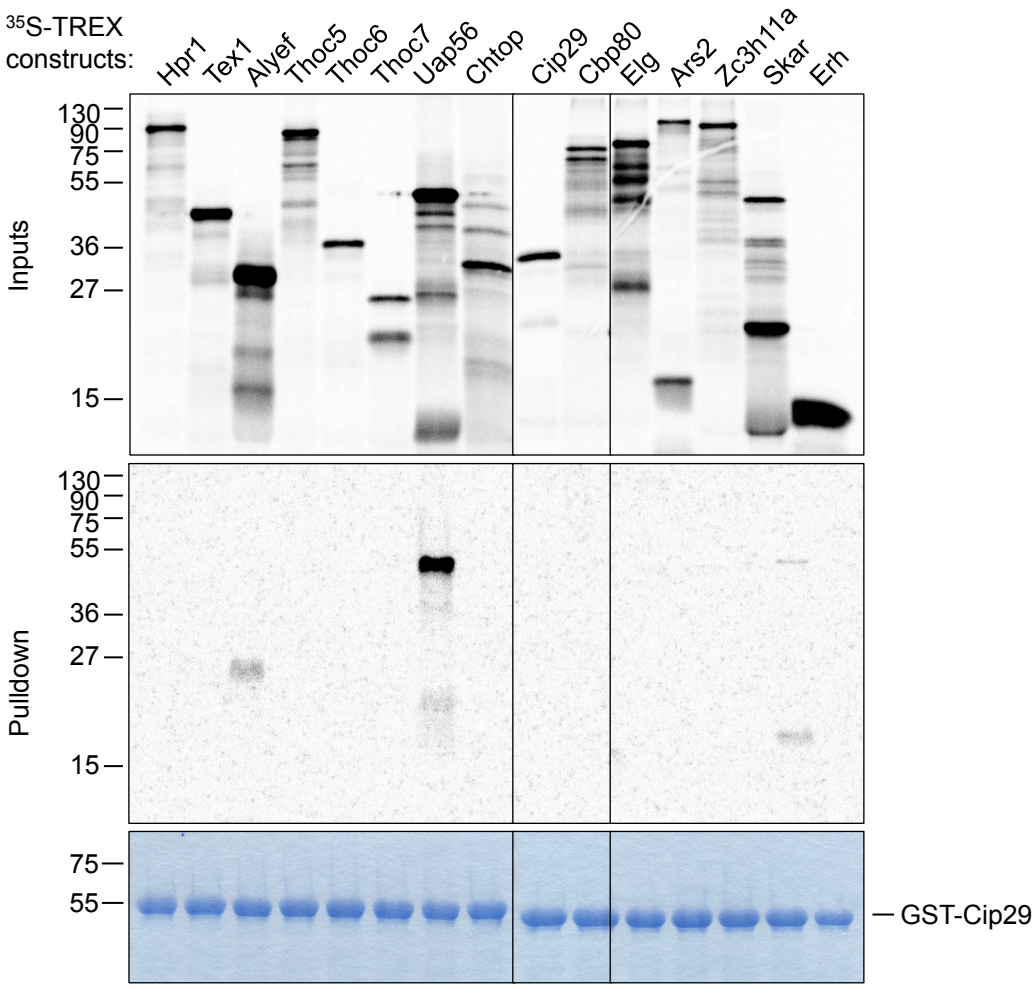


Fig. 4.2.1 J TREX binary interactions *in vitro*. (J) GST-Cip29. GST-Cip29 interacts specifically Uap56.

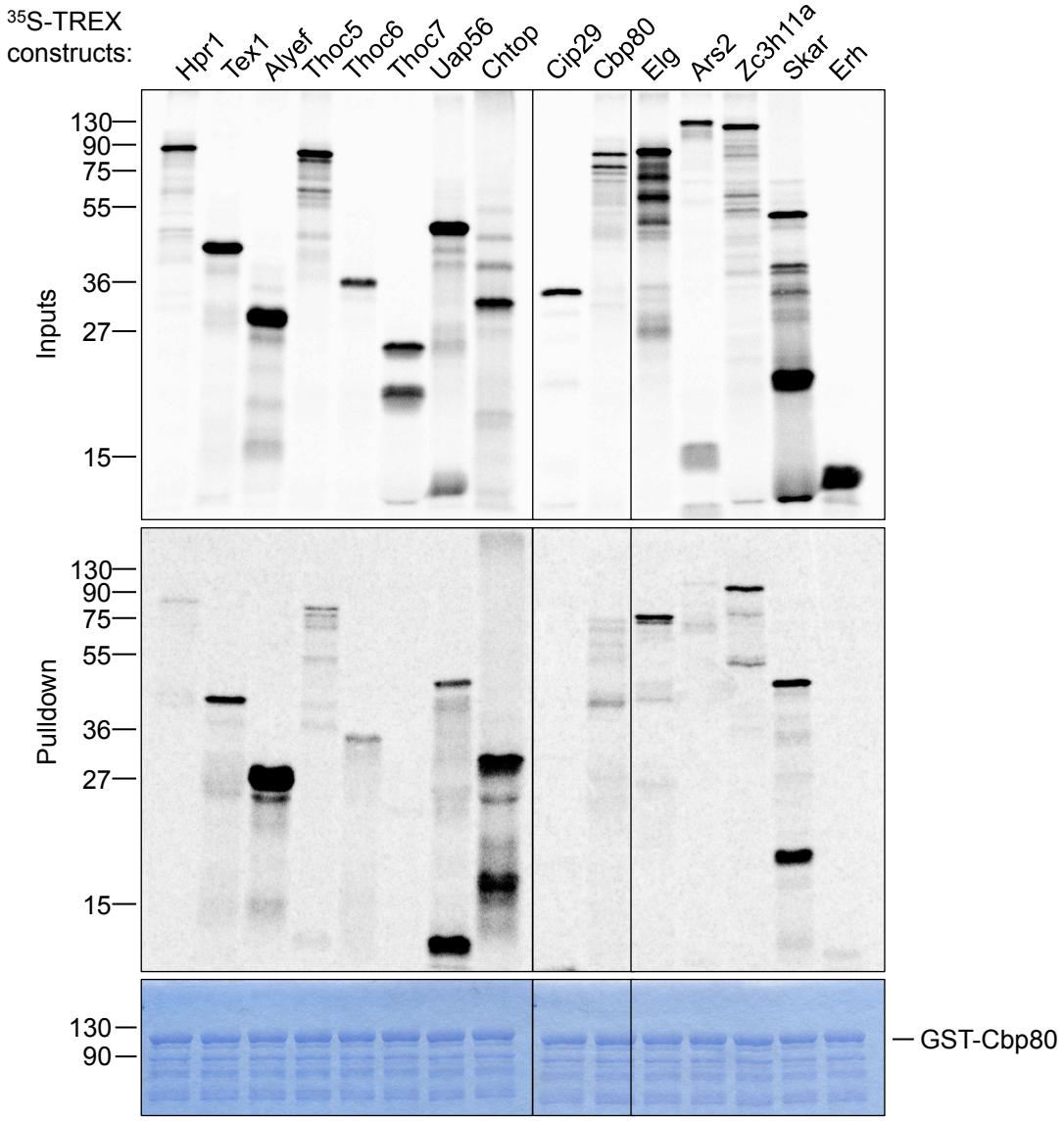
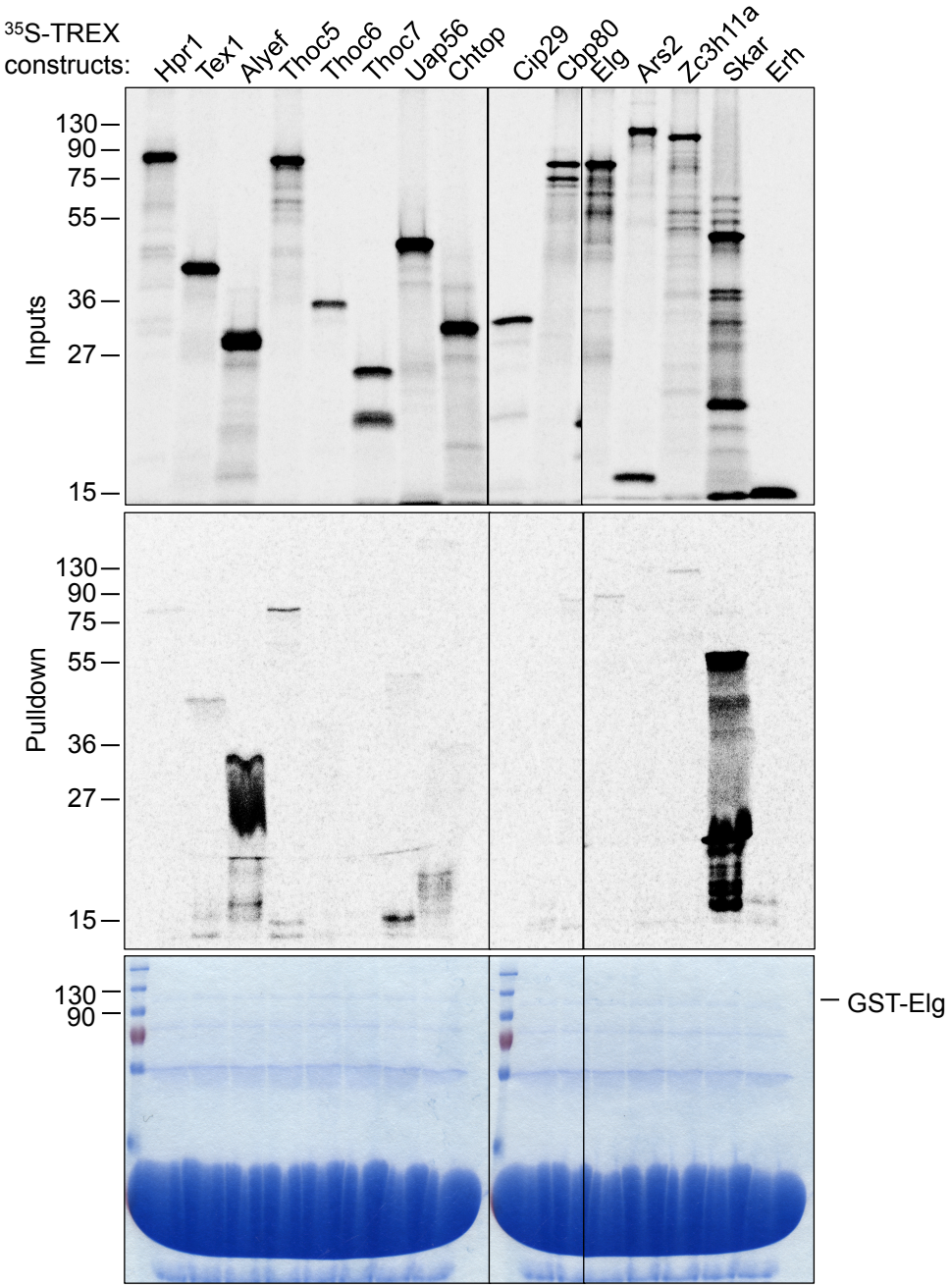


Fig. 4.2.1 K TREX binary interactions *in vitro*. (K) GST-Cbp80. GST-Cbp80 significantly binds Tex1, Alyref, Chtop, Elg and Zc3h11a. Weaker but specific interactions are also observed with Thoc5, Thoc6 and Ars2.



<u>Fig. 4.2.1 M</u> TREX binary interactions *in vitro*. (L) GST-Elg. GST-Elg was very poorly expressed and a large amount of GST was present in the samples. The strongest interactions observed were with Alyref and Skar, but due to their non-specific interaction in the negative control this is likely a result of the large amount of GST. Potential specific interactions are observed with Thoc5, Tex1, Zc3h11a and Erh.

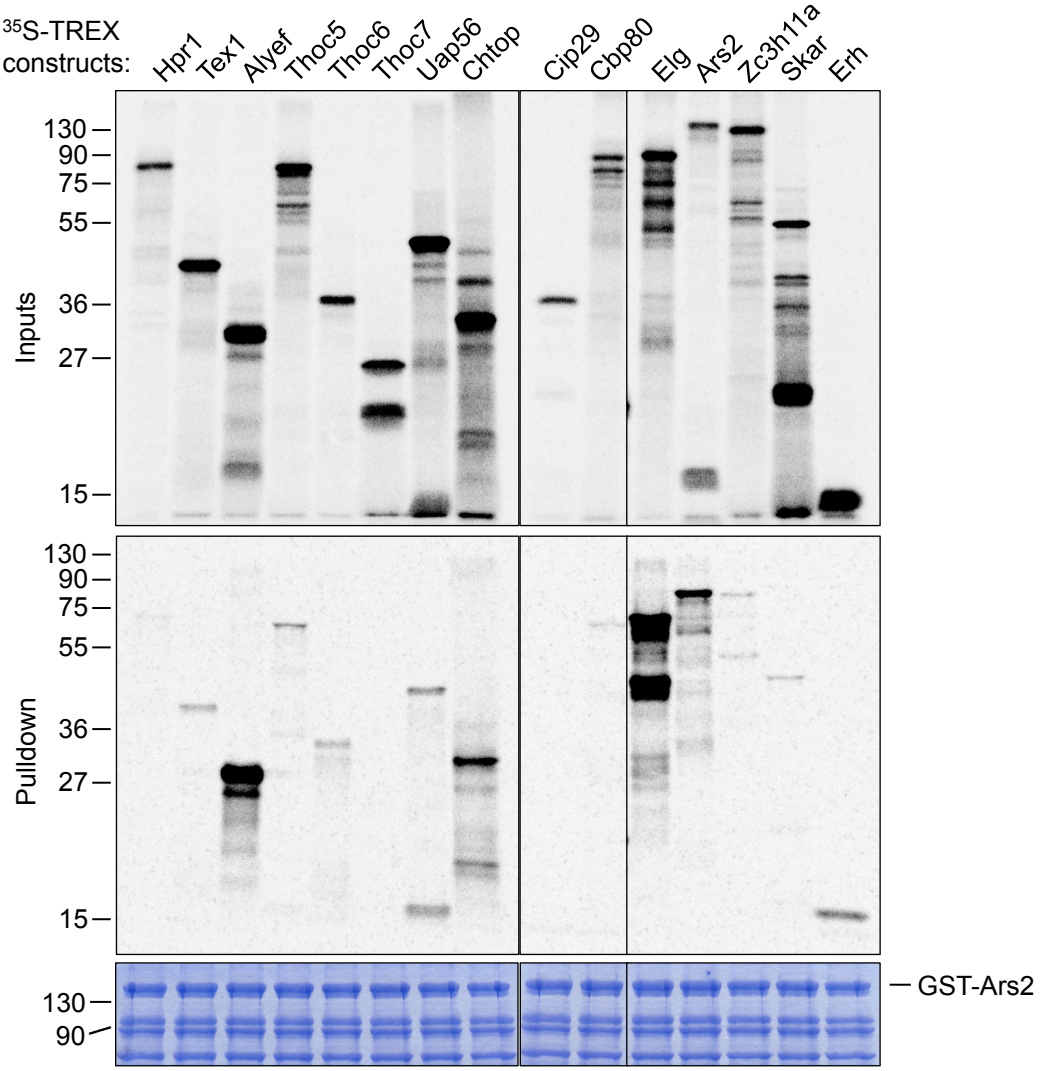


Fig. 4.2.1 N TREX binary interactions *in vitro*. (M) GST-Ars2. GST-Ars2 interacts strongly with Alyref, Chtop, Elg, and itself. Weaker, specific interactions are observed for Tex1, Thoc5, Thoc6, Uap56, Cbp80, Zc3h11a and Erh.

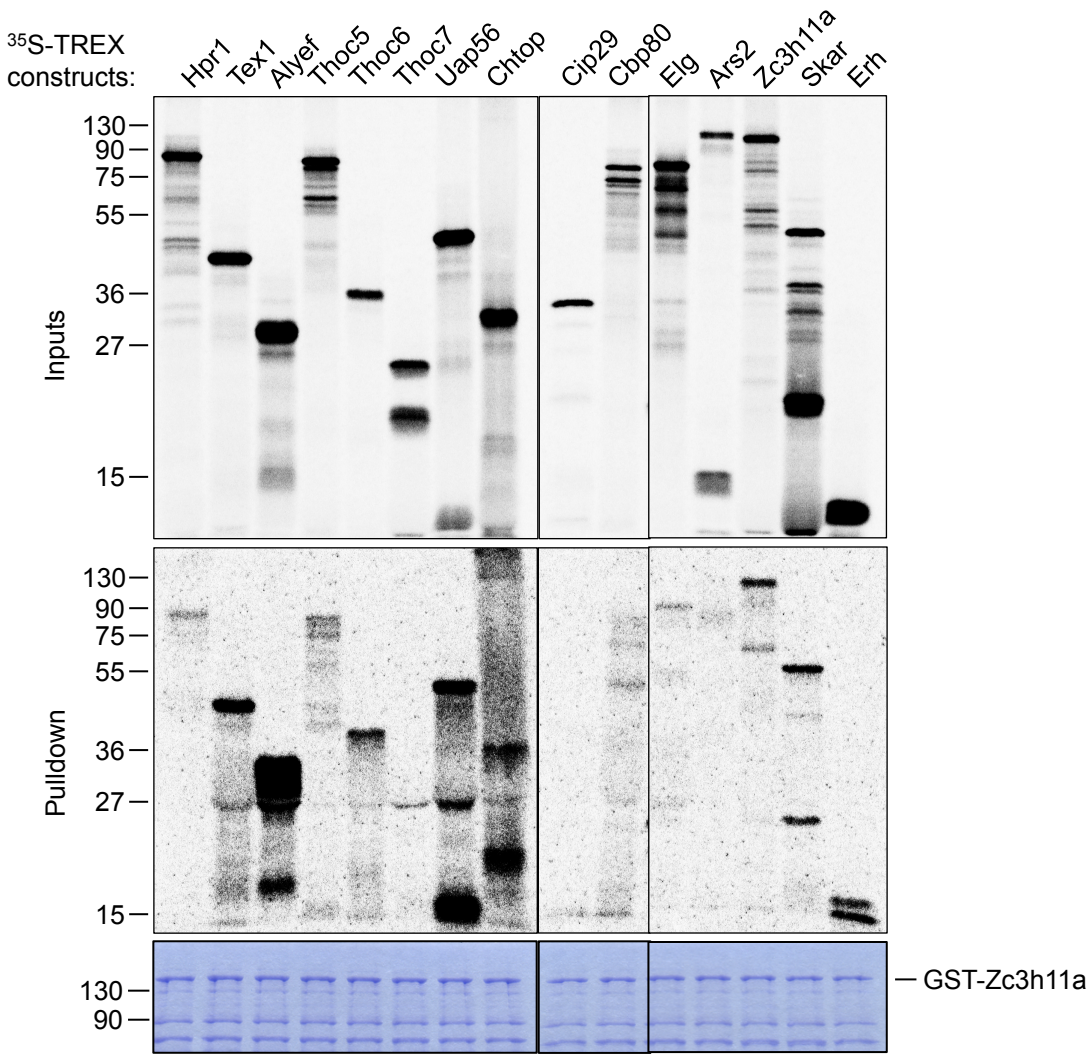
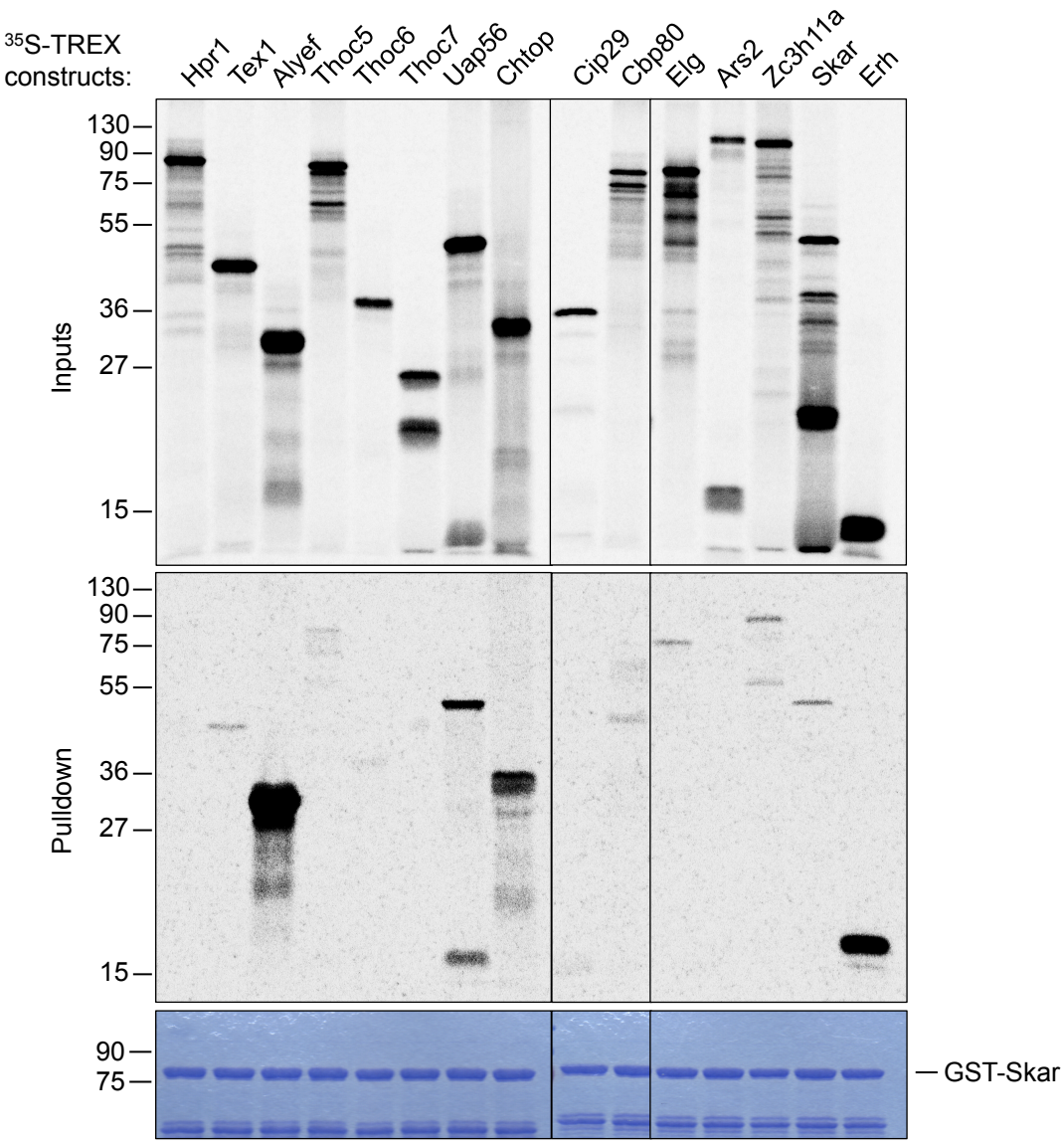


Fig. 4.2.1 O TREX binary interactions *in vitro*. (N) GST-Zc3h11a. GST-Zc3h11a binds strongly to Tex1, Alyref, Thoc6, Uap56, Chtop, Erh and itself. Weaker specific interactions are observed with Hpr1, Thoc5, and Elg.



<u>Fig. 4.2.1 P</u> TREX binary interactions *in vitro*. (O) GST-Skar. GST-Skar binds strongly to Alyref, Uap56, Chtop and Erh. Weaker but specific interactions are observed for Tex1, Thoc5, Elg and Zc3h11a.

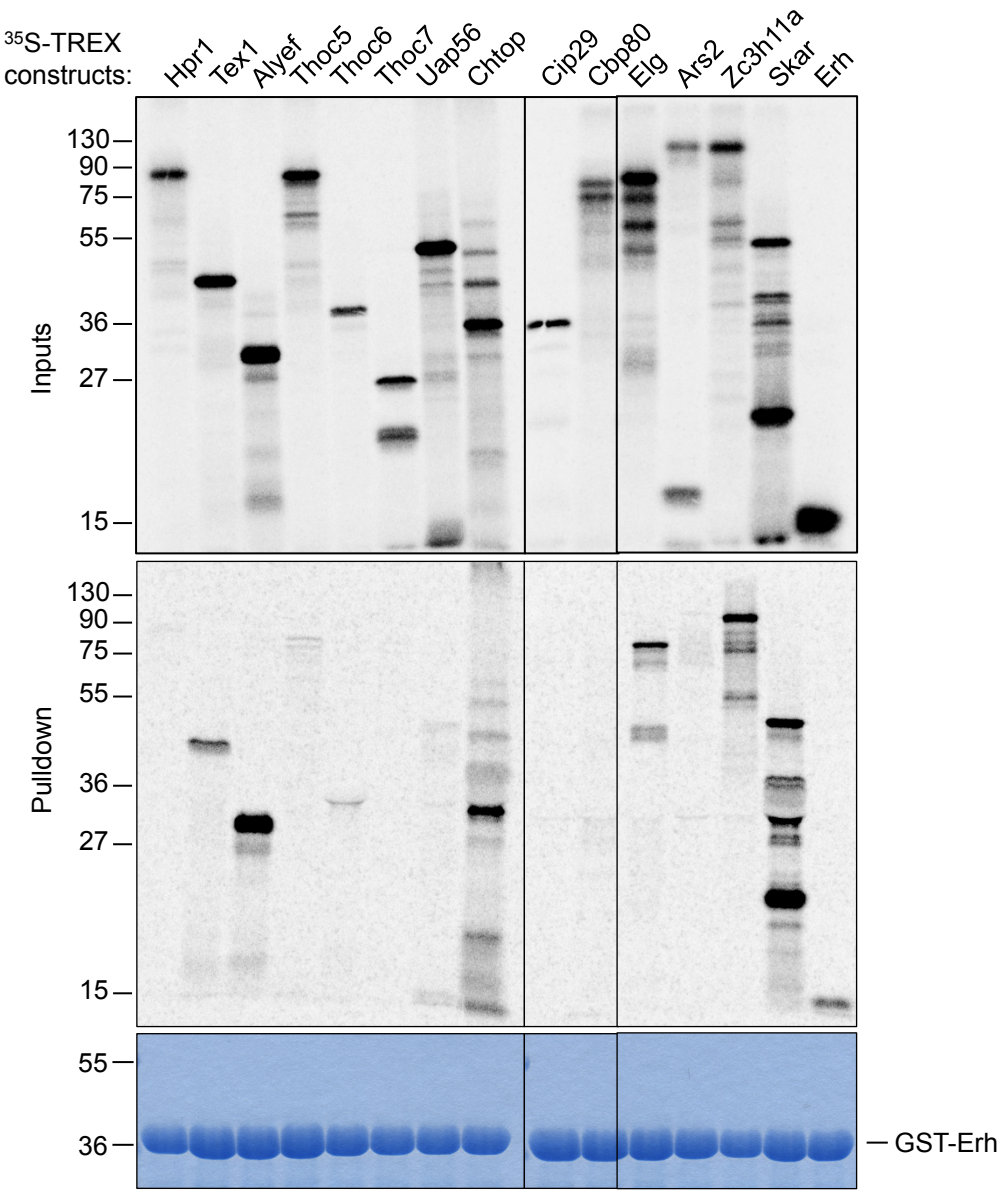


Fig. 4.2.1 Q TREX binary interactions *in vitro*. (P) GST-Erh. GST-Erh binds strongly to Tex1, Alyref, Chtop, Elg, Zc3h11a and Skar. Weak interactions are observed with Thoc5, Thoc6 and Uap56.

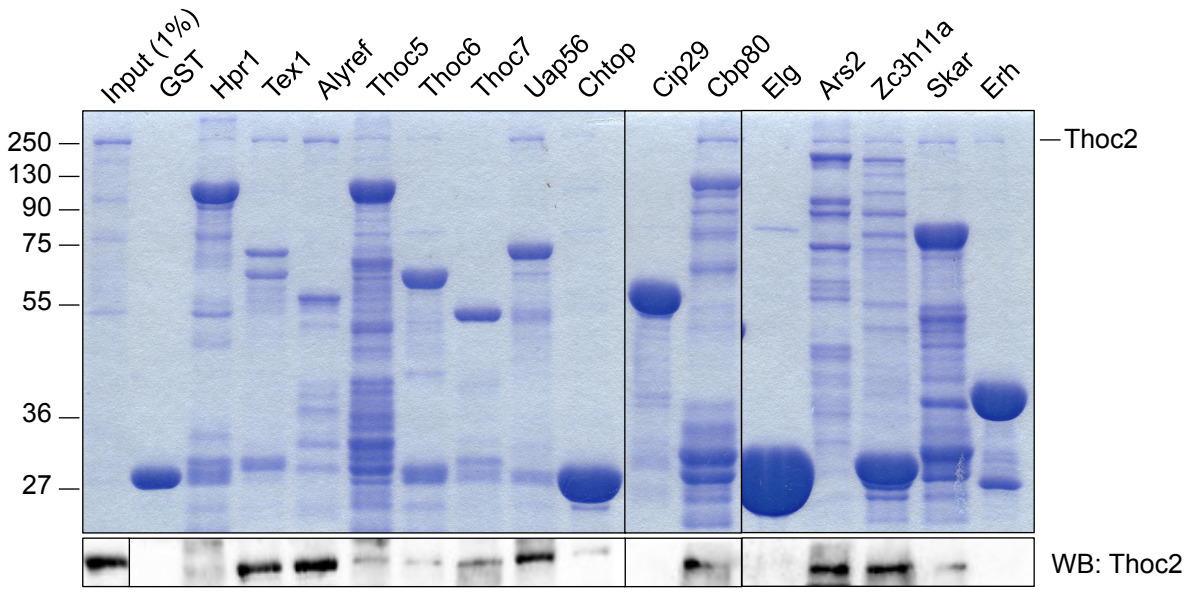


Fig. 4.2.1 R TREX binary interactions. (Q) GST-TREX proteins used to pull down Thoc2 expressed from baculovirus. Interactions were detected by coomassie stain and Western blot. Specific interactions are observed with Tex1, Alyref, Thoc5, Thoc6, Thoc7, Uap56, Chtop, Cbp80, Ars2, Zc3h11a and Skar. Coomassie staining indicates interactions with Hpr1 and Erh, which were not confirmed by Western blot.

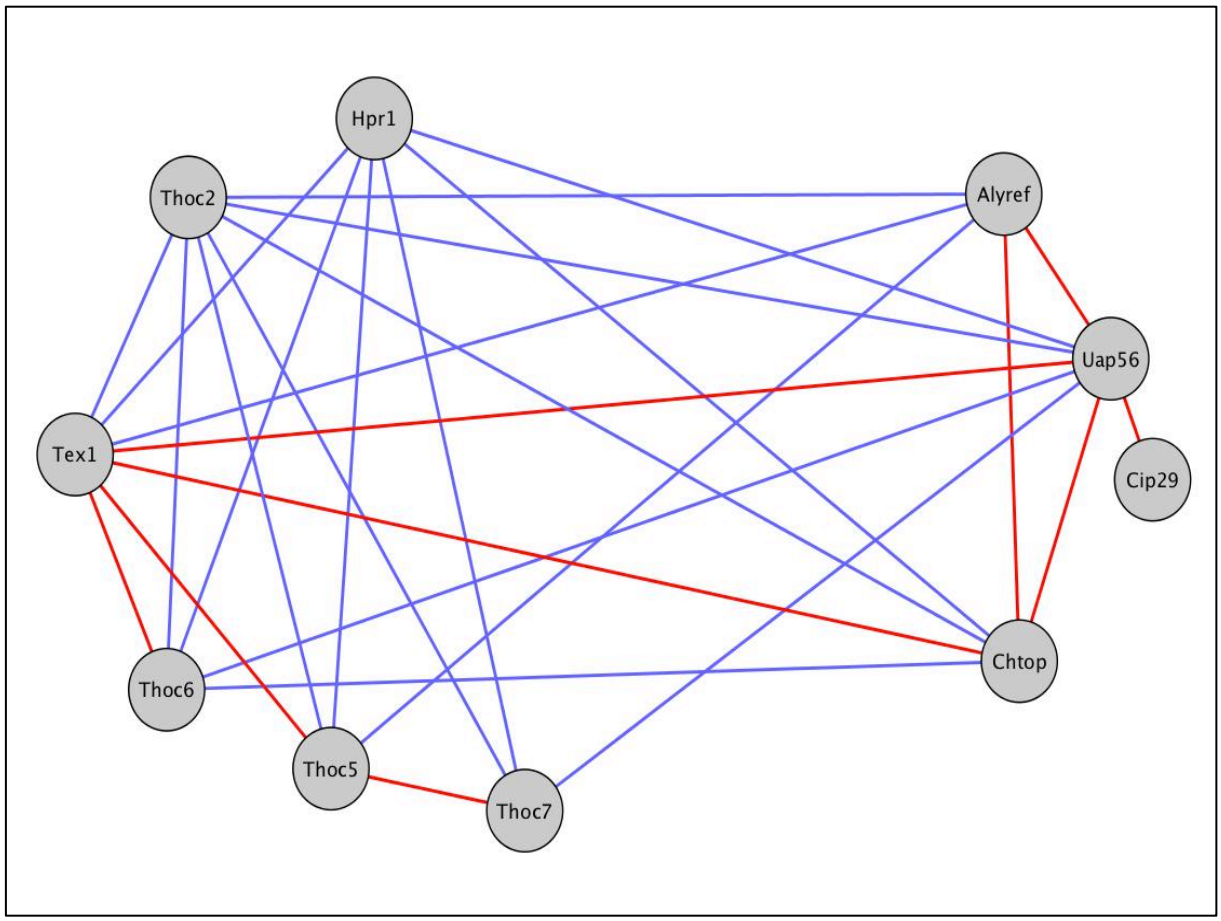


Fig. 4.2.2 Binary interaction network of the core TREX complex. Red lines indicate high confidence interactions from reciprocal binding of components, blue lines indicates non-reciprocal interactions.

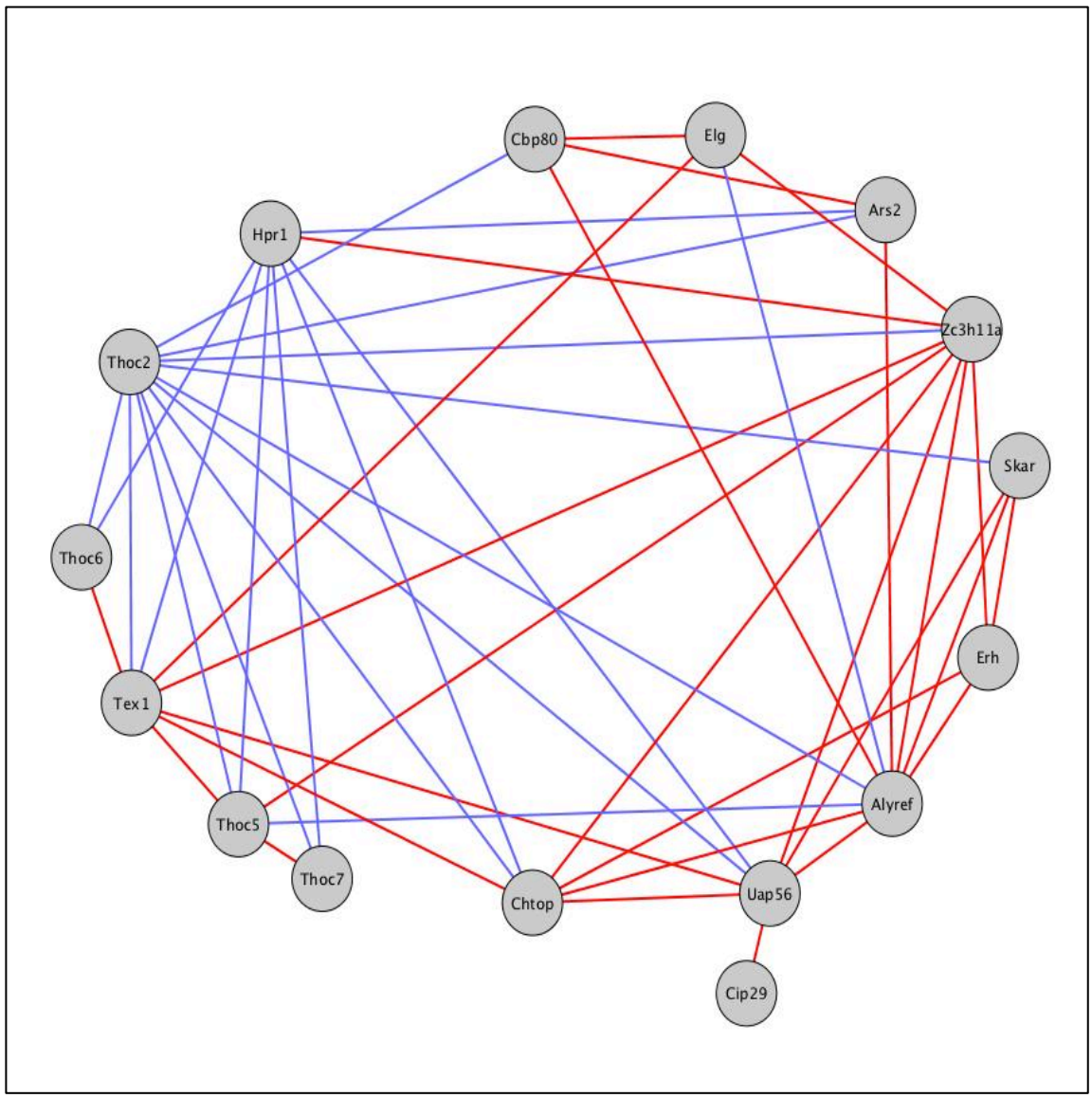


Fig. 4.2.3 High confidence interaction map of TREX. High confidence reciprocal interactions are largely observed between the putative TREX proteins with Chtop, Uap56 and Alyref and fewer with the core Tho proteins.

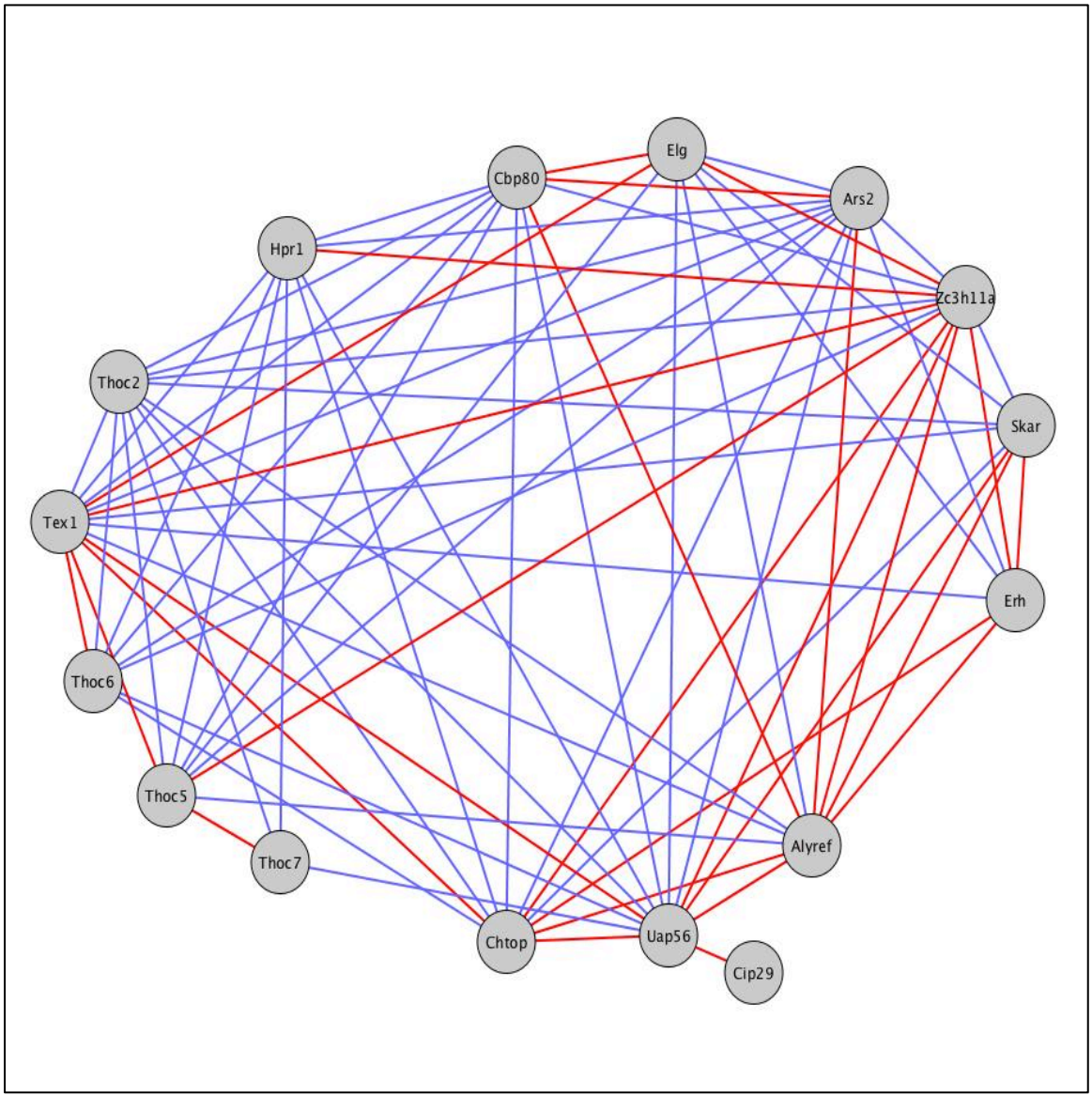


Fig. 4.2.4 Medium confidence interaction map of TREX. Interaction network within TREX including non-reciprocal binding events.

4.3 Identification of minimal binding regions between Cip29 and Uap56

Cip29 and Alyref bind to Uap56 to form a trimeric complex, and stimulate the ATPase and helicase activities of Uap56 to facilitate remodelling of TREX (Dufu et al. 2010; Chang et al. 2013). The Cip29 structure contains an N-terminal SAP domain, a putative DNA binding domain found in various nuclear proteins. Whilst the Uap56 and Alyref interaction has been well characterised (Golovanov et al. 2006), the interaction between Cip29 and Uap56 remains relatively unresolved.

To define the binding regions involved in the Uap56:Cip29 interaction, truncations of Cip29 were cloned as GST fusions and used to pull down recombinant full length Uap56, and two truncations comprising the N-terminal (aa1-254) and C-terminal (aa255-427) RecA-like domains. The results show that Cip29 specifically interacts with the C-terminal domain of Uap56 (255-427) and that this interaction does not require amino acids 1-50 of Cip29, which contains the SAP domain (Fig. 4.3). Importantly, the N-terminal domain of Uap56 binds preferentially to Alyref (Chung-Te Chang, unpublished data), thus the observation that it is not involved in the interaction with Cip29 delineates the architecture of the Alyref:Uap56:Cip29 trimeric complex.

4.4 Crystallisation Trials of the Alyref:Uap56:Cip29 Complex

With the aim of crystallising Uap56 in complex with Alyref and Cip29, we purified 6xHis-tagged Uap56 as described in the materials and methods before further purifying on a Suparose12 gel filtration column under the conditions used in (Shi et al. 2004). As shown in Figure 4.4.1 A, the eluted fractions containing Uap56 co-purified with a band of approximately double its size. This band increased in intensity upon concentration, which was not accompanied by an increase in Uap56 concentration (Fig. 4.4.1 B). The protein sample precipitated rapidly upon concentration and would not concentrate beyond 5 mg/ml and was therefore unsuitable for crystallisation screens. The co-purified band was

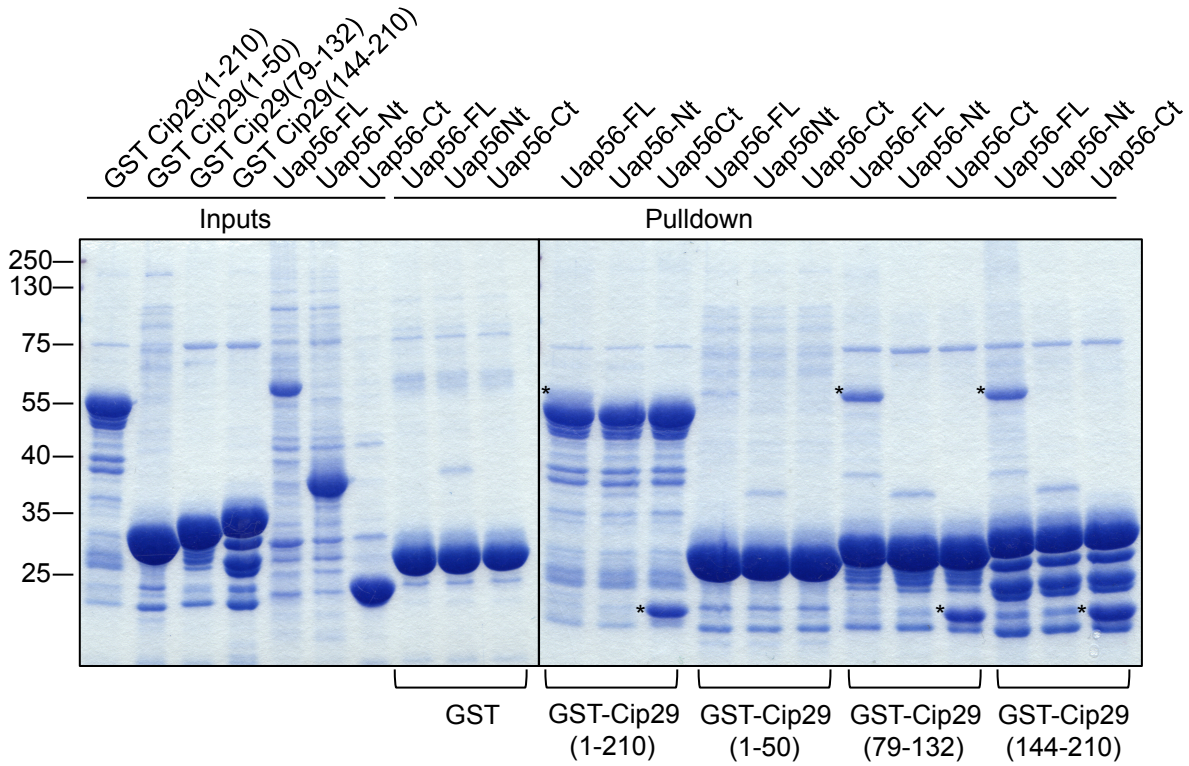
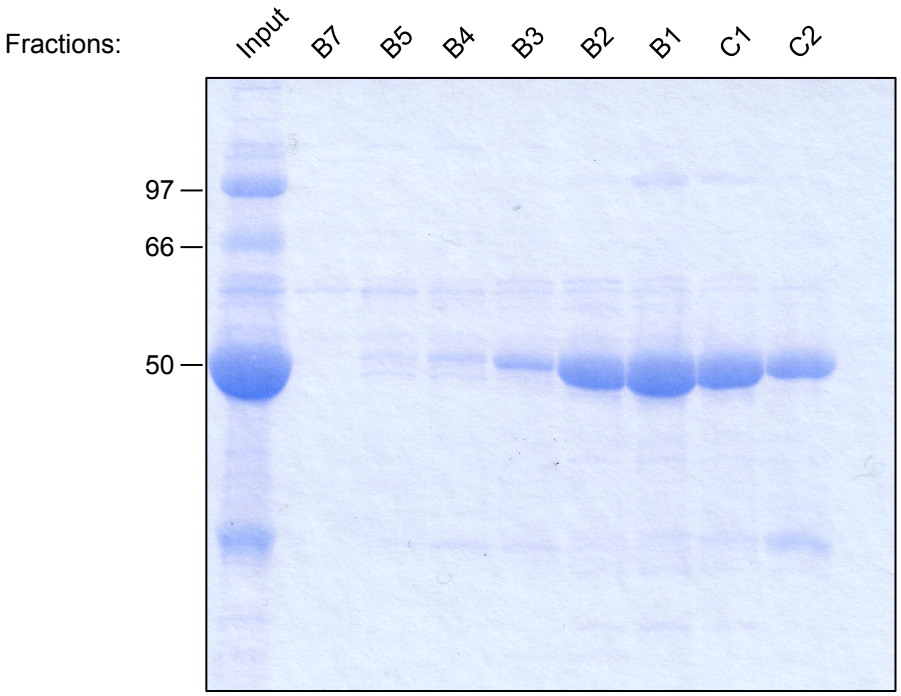
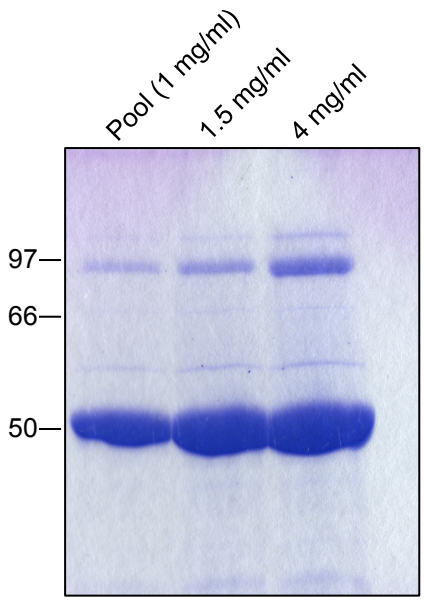


Fig. 4.3 Mapping the Uap56 and Cip29 interacting regions. GST-tagged Cip29 truncations used to pull down Uap56 truncations. Positions Uap56 truncations pulled down by Cip29 are highlighted with asterisks. Cip29 interacts specifically with the C-terminal domain of Uap56. This interaction is dependent on the Cip29 region within amino acids 79-210.

A



B

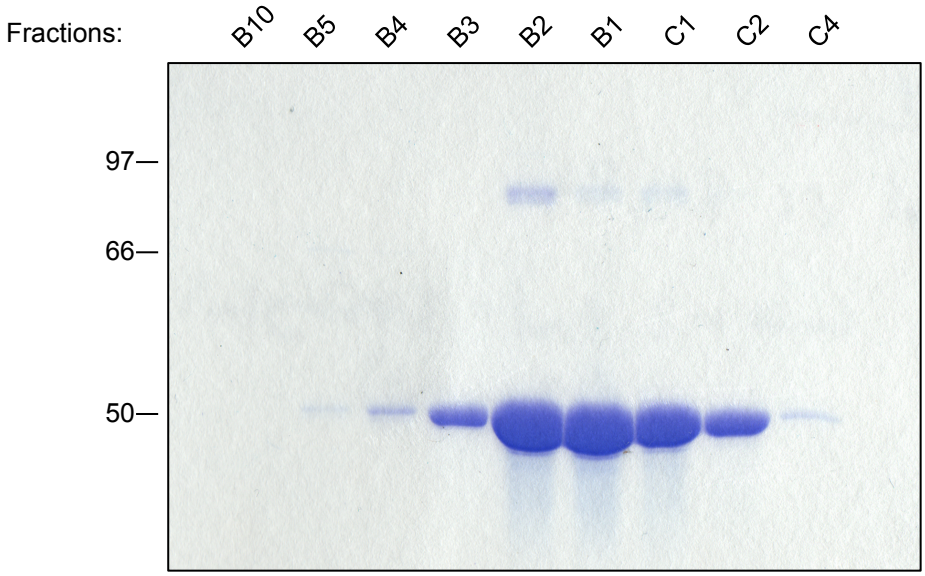


<u>Fig. 4.4.1</u> Gel filtration of Uap56. (A) Purification of Uap56 by gel filtration. Eluted fractions B2, B1 and C1 were pooled and concentrated. **(B)** Concentration of Uap56. Concentration favoured an increase in the high molecular weight band, which was not associated with an increase in Uap56, and the sample precipitated at 4 mg/ml

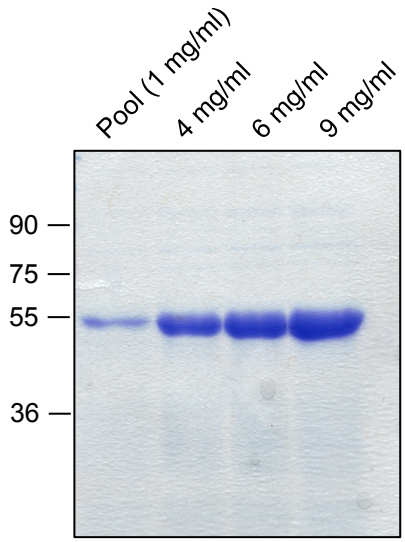
assumed to be a dimer of Uap56, which may have nucleated the precipitation as the sample was concentrated. We therefore attempted to overcome this by purifying Uap56 in the presence of 50 mM Arg/Glu, which had been previously used to help solubilise Alyref for NMR analysis (Golovanov et al. 2006). This appeared to help prevent the formation of the dimer, and the purified sample was concentrated to 9 mg/ml (Fig. 4.4.2). The concentrated sample was then used in crystal screens using the PACT and JCSG suites. No crystal formation was observed after several weeks of incubation. Additionally, there were very few samples containing precipitate suggesting that the Arg/Glu buffer had solubilised the protein to an extent where aggregation of the molecules was unfavorable, thus the use of this method may not be suitable for protein crystallography.

We next concentrated on crystallisation trials using the N-terminal and C-terminal domains of Uap56 in order to study the interaction with Alyref or Cip29 respectively. Alyref is known to be very insoluble, thus we aimed to use the minimal Uap56 binding peptide to soak into crystals of N-terminal Uap56. 6xHis tagged Uap56-N was purified as previously described by cobalt affinity purification and subsequent gel filtration. After gel filtration the fractions containing the protein were pooled and concentrated. As with the full length protein, Uap56-N had similar solubility issues and was unable to be concentrated above 5 mg/ml without precipitating, therefore was not used for crystallisation trials (Fig. 4.4.3). With the aim of producing structural data of the Uap56:Cip29 interaction, full length and C-terminal Uap56 was pulled down by the interacting GST-Cip29 truncations described above. The GST tag was then cleaved off to release Cip29 and the co-purified Uap56. The untagged Cip29 (79-132 and 144-210) were very insoluble however as observed by significant smearing on the coomassie gel and presence of precipitate in the solution (Fig. 4.4.4). To overcome this issue, a longer Cip29 truncation, spanning the two smaller binding regions identified (aa79-210), was cloned as a GST fusion and used to pull down Uap56 (Fig. 4.4.5). As shown in the figure, this truncation behaved as expected, binding specifically to the C-terminal domain of Uap56. Additionally, after cleavage, this truncation was much more stable and the complex formed with C-

A

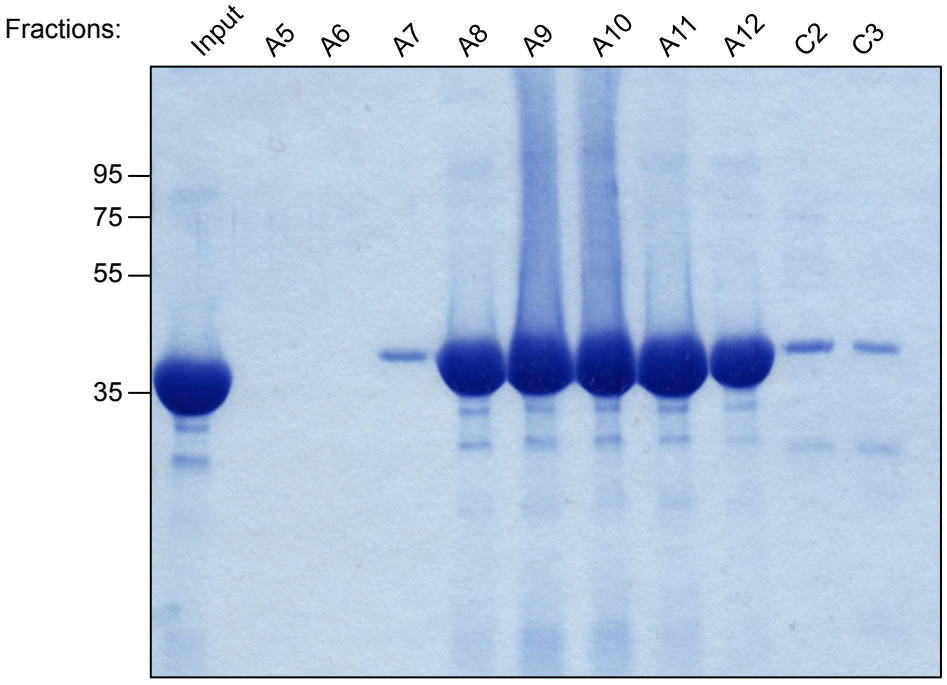


B

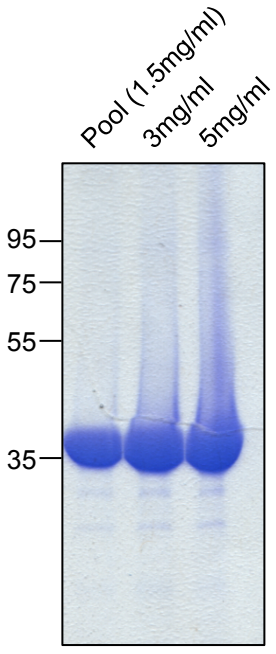


<u>Fig. 4.4.2</u> Gel filtration of Uap56 with Arg/Glu buffer. (A) Frections B2, B1, C1 and C2 of purified Uap56 were pooled and concentrated. **(B)** In the presence of 50 mM Arg/Glu, formation of the higher molecular weight band was not observed during concentration, and concentration of Uap56 reached 9 mg/ml

A



B



<u>Fig. 4.4.3</u> Gel filtration of Uap56-N. (A) Freactions eluted from gel filtration column. Fractions A8, A9, A10, A11 and A12 were pooled for concentration. **(B)** Concentration of Uap56-N.

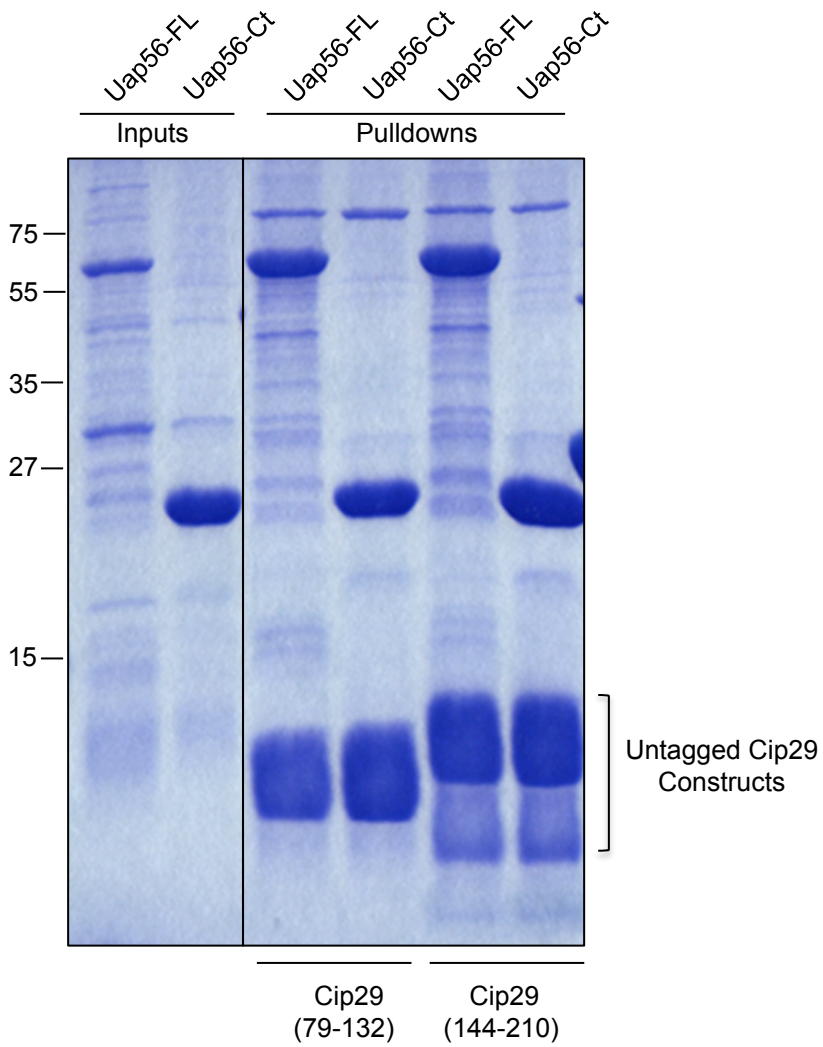


Fig. 4.4.4 Co-purification of Uap56 and Cip29. The Uap56 interacting regions of Cip29 were purified on GSH resin and used to pull down Uap56. The formed complex was eluted by cleavage of the GST tag from Cip29 proteins.

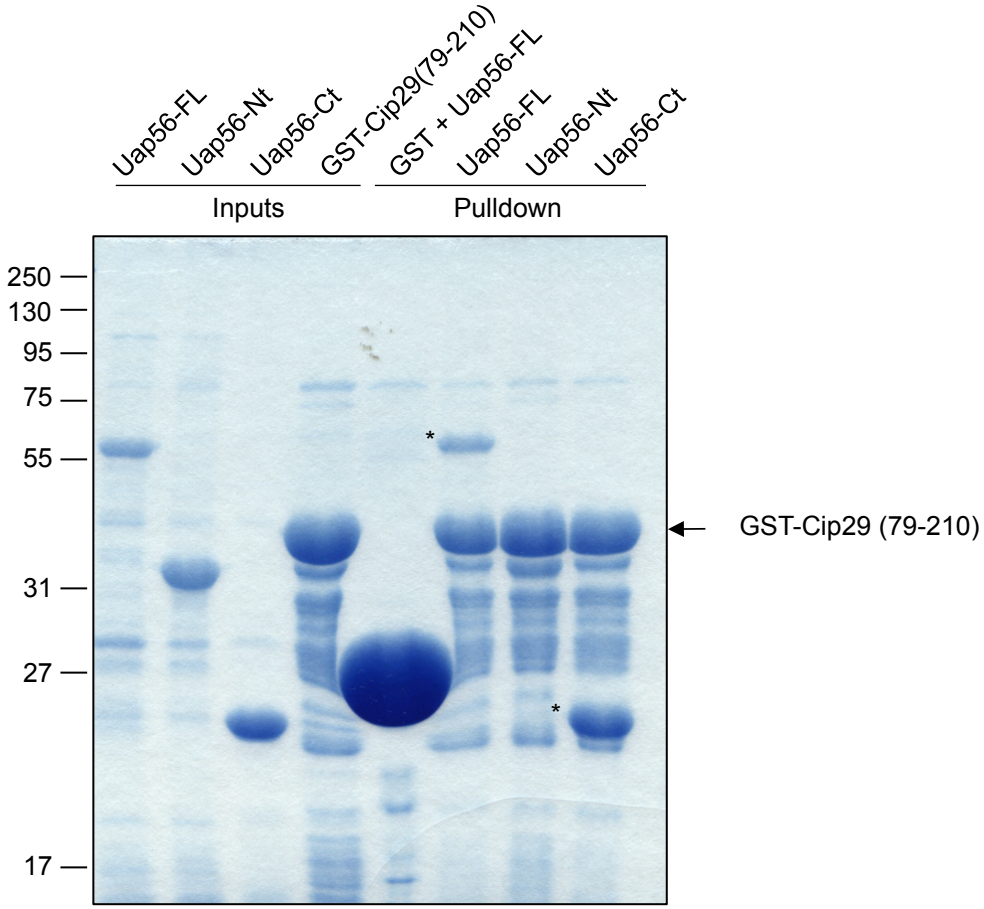


Fig. 4.4.5 Co-purification of Uap56 and Cip29. GST-Cip29(79-210) spanning the two Uap56 binding regions was used to pull down Uap56 truncations. Cip29:Uap56 complex was eluted by cleavage of Cip29 from the GST tag, positions of copurified Uap56 indicated by asterisks.

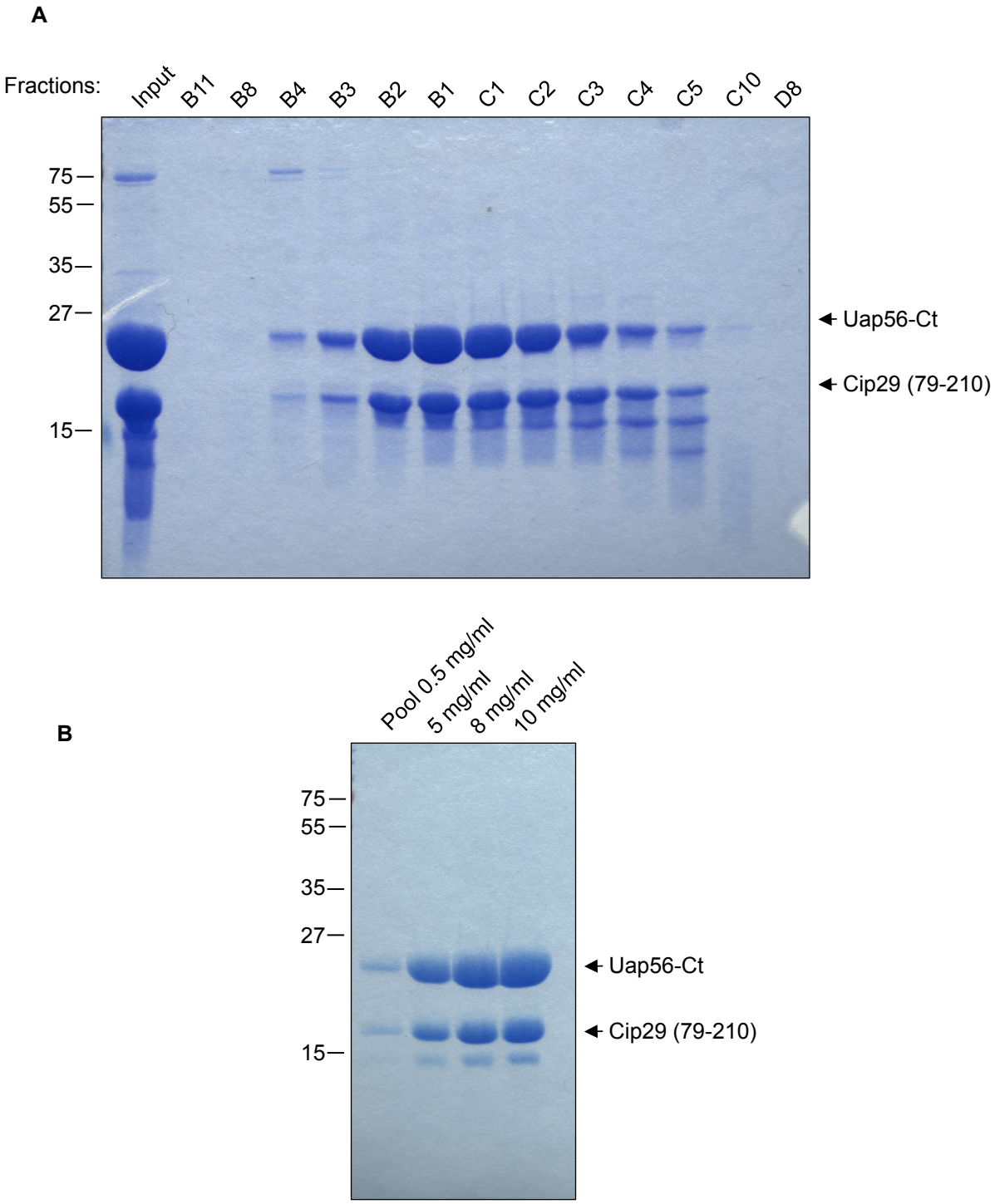


Fig. 4.4.6 Gel filtration of Uap56-C and Cip29. (A) The Cip29(79-210):Uap56-Ct complex was purified by gel filtration. Eluted fractions B2, B1, C1 and C2 were pooled for concentration. **(B)** Concentration of the complex to 10 mg/ml

terminal Uap56 was further purified by gel filtration using the same conditions described above (Fig. 4.4.6). The pooled fractions containing the protein complex were then concentrated to 10 mg/ml and used in crystal screens with the PACT and JCSG suites. After incubation for several weeks, no crystal formation was observed.

4.5 TREX and Microprocessor

Evidence gathered in the lab from microarray data (unpublished data) suggested a link between TREX and microRNA processing. Knock down of TREX components Alyref and Thoc5 resulted in an up-regulation of genes involved in the microRNA processing pathway: DGCR8 and Ars2. A recent study has shown that the knock down of Microprocessor results in an up-regulation of endogenous mRNAs, indicating that Microprocessor may play a direct role in regulating mRNA stability, independently of miRNAs. Additionally, this study also found that the majority of DGCR8 binding sites were protein-coding mRNAs (Macias et al. 2012). Interestingly, the mRNA with DGCR8 binding sites significantly overlapped with the list of mRNAs in the microarray analysis, which were upregulated upon TREX depletion. This provided an intriguing possibility that TREX is involved in recruitment of Microprocessor to mRNA transcripts.

GST-Alyref is able to pull down the Microprocessor component, Drosha, expressed in 293T as a FLAG-tagged protein (Fig. 4.5.1). GST-TREX proteins were purified and incubated with 293T total cell extract expressing Flag-Drosha, therefore an indirect interaction between Alyref and Drosha via other proteins cannot be ruled out. However, from the TREX binary interaction data we gathered above, Ars2 is able to directly interact with the core Tho proteins Thoc2, Tex1, Thoc6 and Thoc5, as well as Alyref, Uap56 and Chtop. Ars2 is a CBC-associated protein involved in miRNA processing (Gruber et al. 2009) indicating an extensive association between TREX and this miRNA processing factor. These results provide evidence of a clear link between TREX and Microprocessor, though the functional consequence of this is unclear. Further experiments were

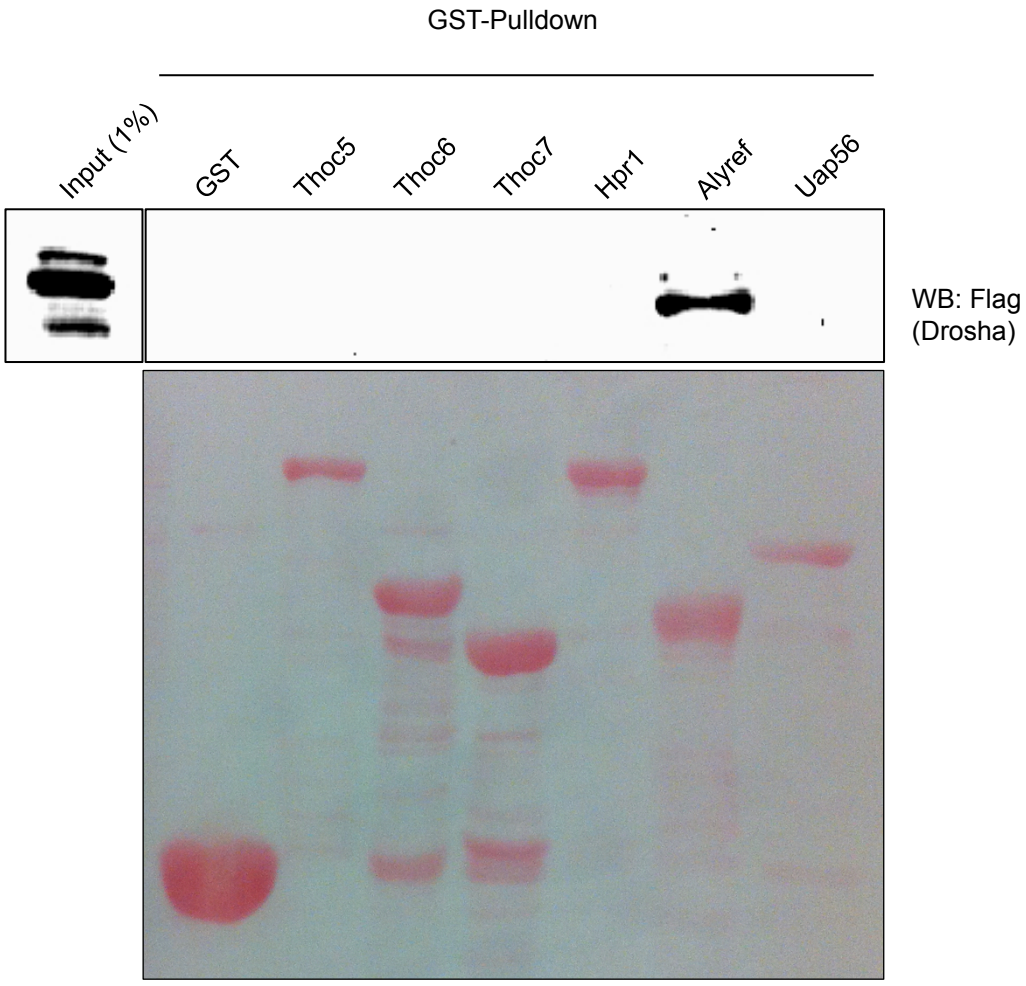
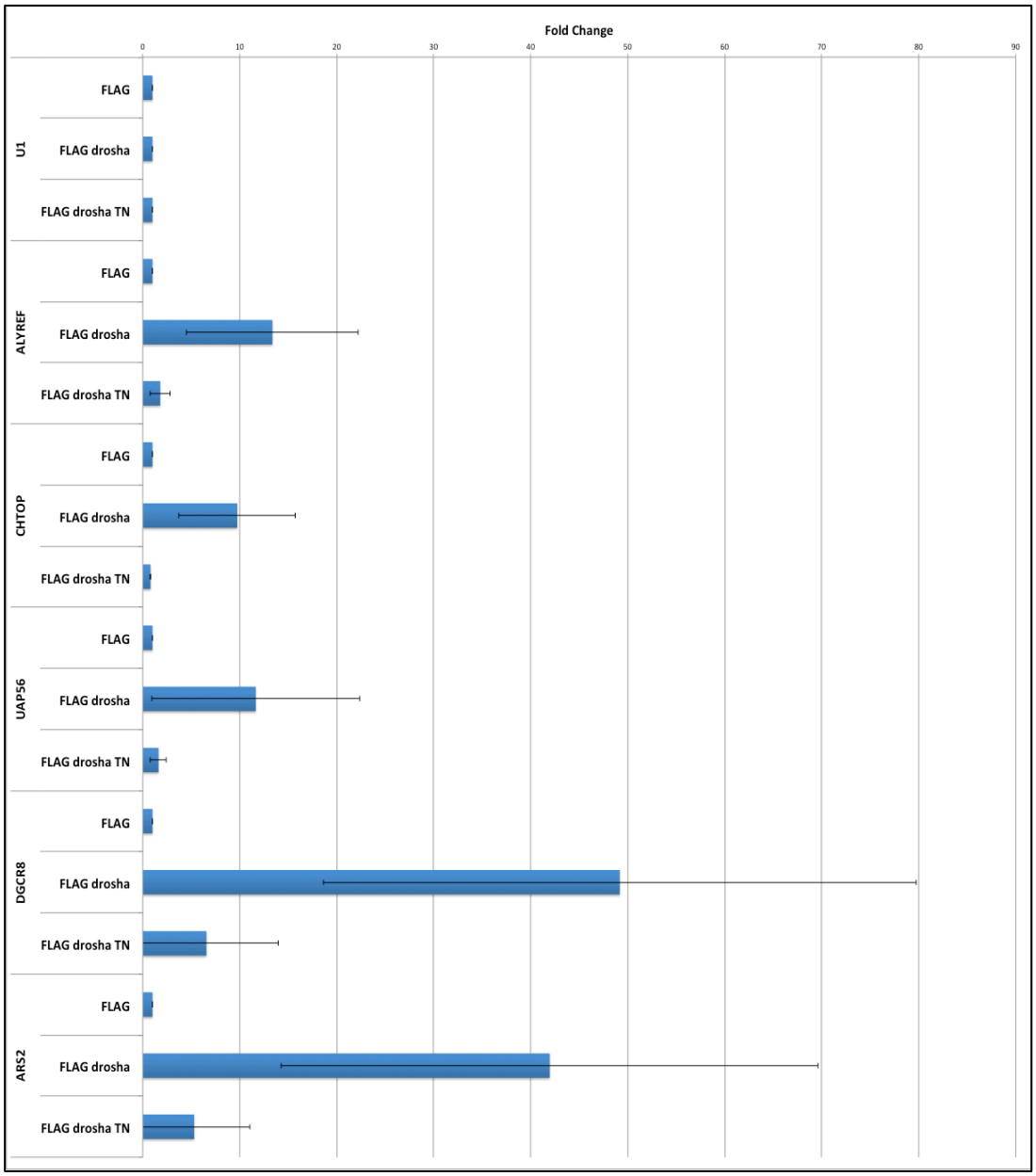


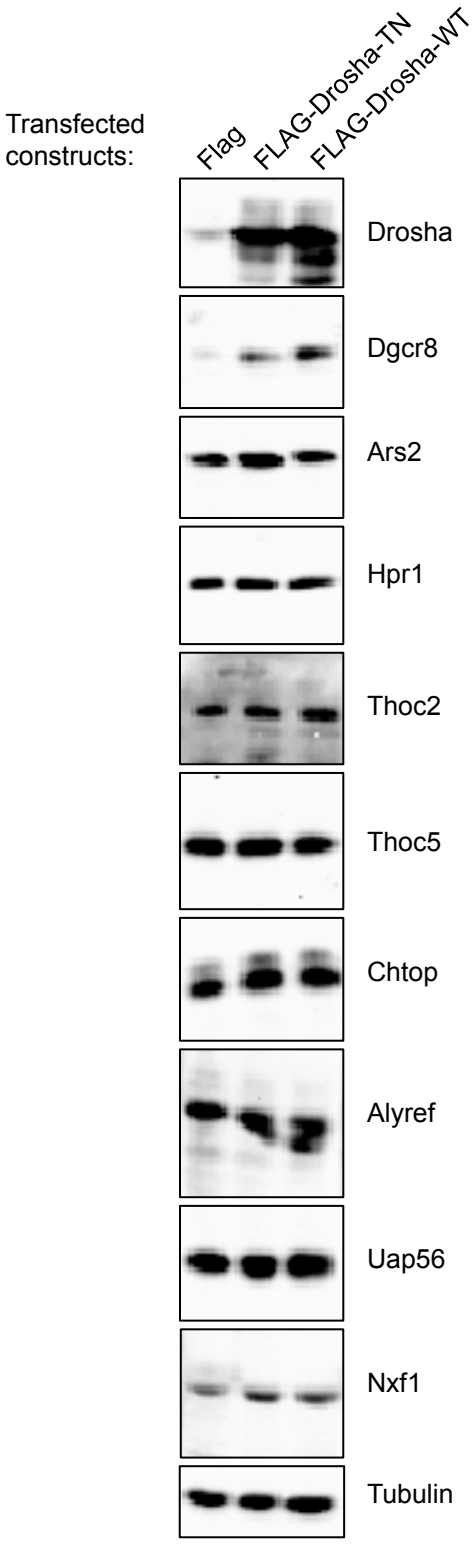
Fig. 4.5.1 GST-Alyref associates with FLAG-Drosha. GST-TREX proteins were incubated with 293T cell extract expressing FLAG-Drosha. GST-Alyref was specifically able to pull down FLAG-Drosha, detected by Western blot with Flag antibody

carried out to validate the array data and define the relationship between TREX and Microprocessor.

Drosha-TN is a catalytic mutant, which is unable to process pri-miRNAs. Overexpression of this mutant has been shown to block miRNA processing and result in an increase of DGCR8 RNA and protein levels (Han et al. 2009). To test whether TREX components were similarly regulated by Drosha, we over-expressed a FLAG- Drosha(TN) or FLAG-Drosha(WT), in 293T cells and compared the expression levels of TREX components at the protein and RNA levels. Contrary to the literature cited above, the RNA levels of DGCR8 and the TREX components tested were increased when FLAG-Drosha(WT) was overexpressed. Alyref, Chtop and Uap56 all showed an approximately 10-fold increase of averaged RNA levels when cells were transfected with FLAG-Drosha(WT) compared to cells transfected with FLAG. By comparison, when cells were transfected with FLAG-Drosha(TN), these components did not show a such a significant change in RNA levels. DGCR8 RNA levels showed an average increase of nearly 50-fold upon overexpression of Drosha-WT compared to an approximately 6-fold increase when Drosha-TN was overexpressed. Similarly, Ars2 RNA levels were increased by approximately 40-fold in the FLAG-Drosha(WT) condition compared to a 5-fold increase in the Drosha(TN) condition (Fig. 4.5.2). At the protein level, a modest increase of DGCR8 was observed in the FLAG-Drosha transfected cells compared to FLAG transfected cells, with a larger increase in the FLAG-Drosha(WT) condition. However, for Alyref, Chtop, Uap56 and Ars2, the changes observed at the RNA level did not translate to changes in the protein levels. Similarly, the protein levels of the TREX components Hpr1, ThoC2, Thoc5 and Nxf1 showed no significant response to FLAG-Drosha(WT) or FLAG-Drosha(TN) (Fig. 4.5.3). Subsequent sequencing of the Drosha-TN construct, provided by a collaborator, showed that it was not the full catalytic mutant, which contained a point mutation in each of the RNaseIII domains (Han et al. 2009). Instead, a point mutation was only present in the first RNaseIII domain. This may be the cause of the discrepancies observed between our results and previous studies, and therefore, we are unable to conclude whether



<u>Fig. 4.5.2</u> Comparison of RNA levels between cells overexpressing Drosha-WT or Drosha-TN by qPCR. Large increase in RNA levels of Alyref, Chtop, Uap56, DGCR8 and Ars2 are observed when Drosha-WT is overexpressed but not Drosha-TN.



<u>Fig. 4.5.3</u> Effect on TREX protein expression by Drosha-TN catalytic mutant. Western blot on human cell extracts overexpressing FLAG, FLAG-Drosha, and FLAG-Drosha-TN. Antibodies used are indicated to the right of the panels. Increase in expression of Dgcr8 is observed upon overexpression of Drosha-TN. No significant change in expression of other components is observed

or not Drosha activity affects the stability of TREX mRNA.

To validate the array data, and test whether TREX may be involved in recruiting Microprocessor to mRNAs, a stable RNAi cell line was used to knock down Alyref and Thoc5 for up to 72 hours before testing the expression levels of various proteins by Western blot compared to a control RNAi cell line (Fig.4.5.4). Consistent with the array data, the results show that knock down of TREX components Alyref and Thoc5 result in a significant increase of Ars2 protein expression at 48 hours after knock down. The increase of DGCR8 expression on the other hand is more modest though appears to be more significant at the 72 hour time point. Additionally, Drosha displays an upregulation after 48 hours of knockdown, although this was not observed by the microarray data.

Next, the expression levels of the transcripts from these cell lines were measured by qRT-PCR. The hypothesis that TREX is involved in recruiting Microprocessor to promote mRNA cleavage would predict that depletion of TREX components would result in an upregulation of mRNAs targeted by Microprocessor. The results show that contrary to the protein expression level, DGCR8 transcript levels are slightly reduced. This also contrasts the microarray data, which indicated an increase in DGCR8 expression by 2.3-fold. Ars2 transcript levels were also increased by 2-fold according to the microarray data, however we only observed a modest average increase of Ars2 expression by qRT-PCR and there was significant variance between the replicate samples as illustrated by the error bars, thus there is unlikely to be an general change in Ars2 expression in response to Alyref and Thoc5 knock down (Fig. 4.5.5).

Microarray analysis was additionally carried out using the RNA extracted from the Control and Alyref + Thoc5 RNAi cell lines, to assess whether there was a change in level of miRNA processing upon knock down of Alyref and Thoc5. Total RNA from Alyref + Thoc5 RNAi cell lines in biological triplicates, and miRNAs were detected by Affymetrix GeneChip miRNA array before analysis using the Partek data analysis suite. The results indicated that there was no difference in the amount of fully processed miRNA and pre-miRNA between the control and knock down samples. If TREX were to be involved in recruiting Microprocessor to

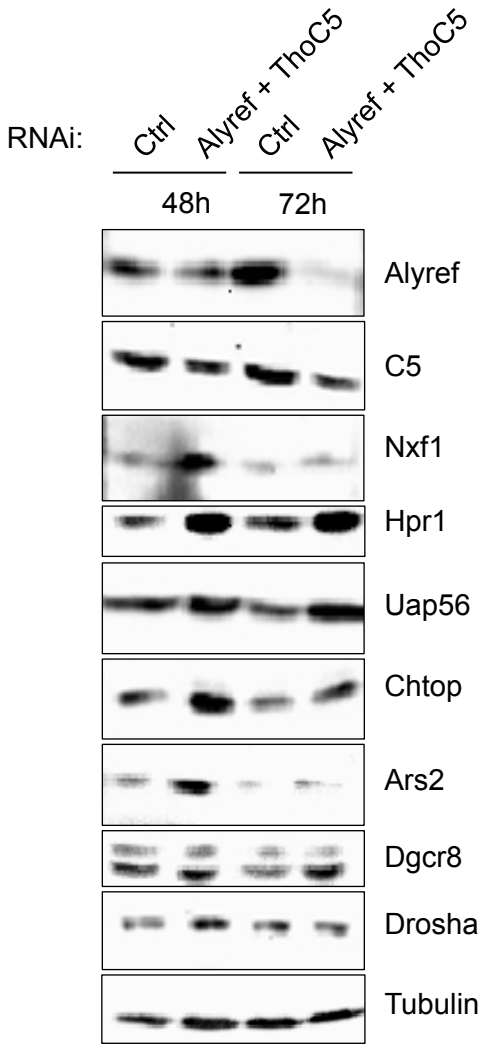


Fig. 4.5.4 Effect of Alyref + Thoc5 RNAi on protein expression of TREX and Microprocessor components. Western blot on human cell extracts depleted of Alyref and Thoc5. Antibodies used are indicated on the right of the panels. Depletion of Alyref and Thoc5 is associated with an upregulation of TREX proteins. Ars2, DGCR8 and Drosha levels are also increased.

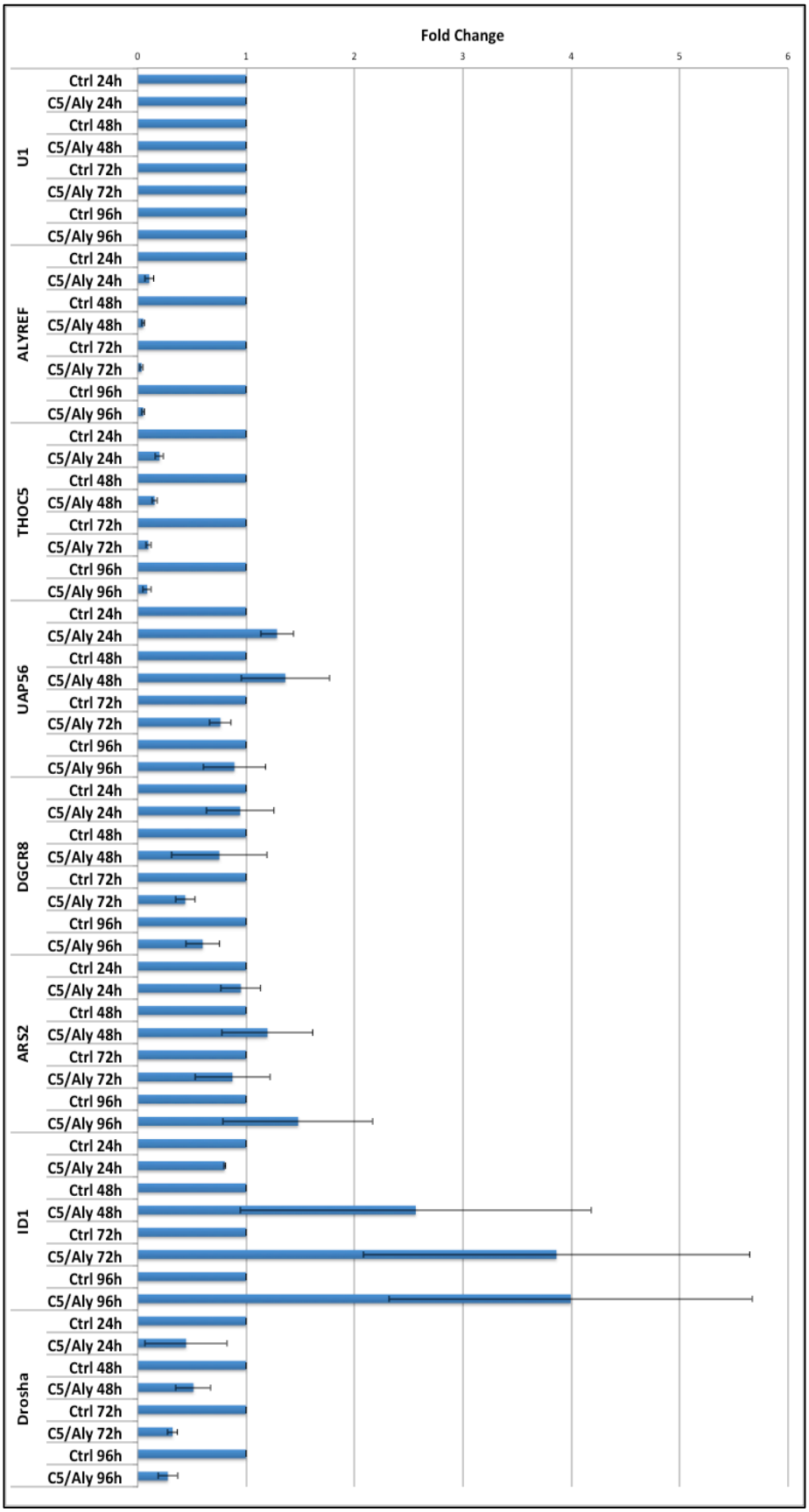


Fig. 4.5.5 Effect on RNA levels by Alyref + Thoc5 RNAi. Increase of Uap56 and Ars2 RNA levels observed on Alyref + Thoc5 depletion. In contrast to protein levels, DGCR8 and Drosha RNA levels are decreased upon depletion of Alyref + Thoc5

transcripts, one would expect a decrease in the amount of fully processed miRNAs and a concomitant increase of unprocessed pre-miRNAs in the knock down condition compared to control.

4.6 Summary

The recently solved EM structure of yeast Tho identified the architecture of the components of this complex, providing some structural insight into the function of Tho. The structure of human Tho, however, remains unsolved and given the disparity of some components between yeast and human Tho, the structure of human Tho is still of significant interest. The stability of the assembled Tho complex isolated from human cells is well documented, we were unable to assemble a complete Tho complex in the systems used. The low expression of ΔC2 components relative to ThoC2 in the baculovirus system may have contributed to the inability to form a stoichiometric Tho complex however, the pull down experiment designed to overcome the issue of stoichiometry also resulted in only a partial reconstitution of the complex. It is plausible therefore, that Tho assembly is dependent on processes *in vivo* that cannot be bypassed in a recombinant system. For instance, Uap56 is suggested to remodel TREX and stabilise it on the mRNP (Chang et al. 2013; Cheng et al. 2006). This process itself may also be favorable for the formation of a stable Tho complex. Additionally, the identification of new putative TREX components (Dufu et al. 2010), provides the possibility that some of these may be involved in stabilisation of the Tho complex.

We then investigated the direct, binary interactions within the wider TREX complex. Several new putative TREX components have recently been reported, but their function, within the context of TREX, remains largely uncharacterised. By performing pulldowns with the repertoire of TREX proteins, we were able to develop an overview of the interaction network within TREX. As expected from the stable nature of the conserved TREX complex, there was a significant network of strong interactions between these components. Significant

note is made of the interactions of Hpr1 and ThoC2. Hpr1 binds to all of the core Tho proteins, whilst ThoC2 binds to all, bar ThoC7, as well as Uap56 and Alyref. Hpr1 was identified in yeast because of the hyper-recombination effects of its mutants. ThoC2 is known to be a multi-copy suppressor of these mutants, thus our observation of extensive contacts between Hpr1 and ThoC2 with the rest of Tho further highlights their importance within the complex. Additionally, there are many binding events between the newly identified TREX components with Alyref and Uap56. The precise functions of many of these proteins are still largely not understood, but Alyref and Uap56 may serve to bridge the association between these proteins and Tho, thus integrating more functional diversity within TREX. A notable discrepancy in our data is the inability for Alyref and Cbp80 to interact with each other with high confidence. This association is well documented, however evidence for it comes from indirect binding assays such as IPs. In this context, Cbp80 would be associated with Cbp20, forming the cap-binding complex. This therefore indicates that Alyref may only interact with a mature cap-binding complex and not with the individual components. On the other hand, the reported association between Alyref and the cap-binding complex, could be bridged by Ars2, as Ars2 is known to form a complex with the CBC (Hallais et al. 2013) and directly interacts with Alyref in our pull down experiments.

The trimeric complex of Uap56, Alyref, and Cip29 has recently been shown to be key in the remodeling of TREX components to form a mature complex. Here we define the binding regions in Uap56:Cip29 interaction, which involve the C-terminal domains of both proteins. Together with unpublished data from the lab, this helps define the architecture involved in the trimer whereby Alyref and Cip29 bind to the N-terminal and C-terminal domains of Uap56 respectively. Crystollographic study of this trimer would be of significant interest. In particular, given the ATP-dependent nature of TREX remodeling by Uap56, and the stimulatory effect of Alyref and Cip29 on Uap56 ATPase activity, it would be relevant to study how the structure of the complex is affected in the presence or absence of ATP.

Experiments previously performed in the lab had demonstrated a potential link between TREX and Microprocessor. This investigation set out to determine whether TREX is involved in recruiting Microprocessor to transcripts to promote miRNA maturation and regulate mRNA stability by cleavage of mRNA transcripts. Despite validating a clear relationship between TREX and the expression of Microprocessor components, the microarray data suggests that this is unlikely to play a role in miRNA processing. In fact, a recent paper found that Drosha regulates gene expression independently of its role in miRNA processing, promoting efficient transcription elongation through its interaction with RNA Polymerase II and Cbp80 (Gromak et al. 2013). TREX has been shown to be recruited to the 5'-end of mRNA via the interaction of Alyref and Cbp80 (Cheng et al. 2006), thus in this context, rather than being recruited by TREX, Microprocessor may recruit TREX through the interaction between Alyref and Cbp80.

Chapter 5

Investigating the association between Alyref and Senataxin

Gene expression involves unwinding of the DNA double helix, exposing the template strand to allow transcription of RNA. During this process, the transcribed RNA can base pair with the template strand, forming an RNA:DNA hybrid called an R-loop. The non-template strand is left exposed as a single stranded region of DNA, which is susceptible to DNA damage, resulting in genome instability. R-loops form naturally at the G-rich regions at the 3' end of transcribed genes and slow down polymerase movement along the gene, allowing the exoribonuclease Xrn2 (yeast Rat1) to degrade the RNA strand and promote transcription arrest. Senataxin is a putative RNA:DNA helicase, which functions to unwind RNA:DNA hybrids, resolving the R-loop and allowing the RNA strand to be degraded. How Senataxin is recruited to these sites is, however, unclear. A recent paper (Nedea et al. 2008) found that the Alyref orthologue, Yra1, was able to co-purify with Sen1 (Senataxin), representing a potential role for TREX in the regulation of Senataxin activity.

5.1 Senataxin directly binds Alyref via a CH-like domain

Setx belongs to the SF1 superfamily of helicases and shares significant homology to Upf1. The domain organisation of Upf1 consists of a CH (cysteine-histidine-rich) domain followed by two RecA domains forming the helicase region. Alignment of Setx and Upf1 confirms that significant homology is concentrated at the C-terminal regions, particularly in the helicase domains (Fig. 5.1.1 highlighted in blue). Additionally, the alignment identified Setx homology with the CH-domain of Upf1 (Fig. 5.1.1 highlighted in red). A recent study solved

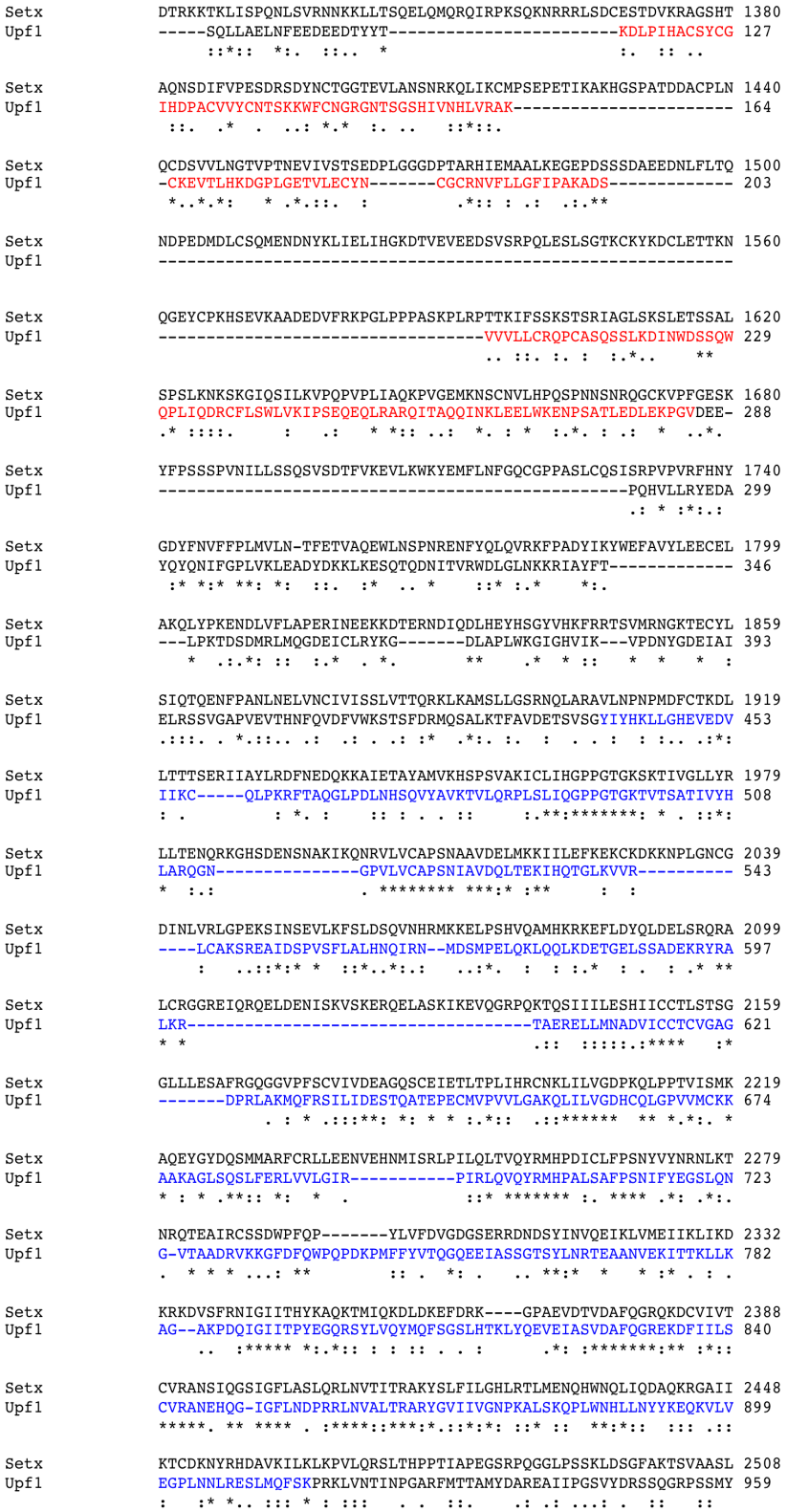


Figure 5.1.1. Alignment of human Senataxin and Upf1 proteins. Senataxin and Upf1 share significant homology with the C-terminal helicase domains (Upf1 helicase region highlighted in blue). Senataxin also aligned to the Upf1 region comprising its CH domain (highlighted in red)

the crystal structure of Upf1 and its enzymatic activator, Upf2. The CH-domain inhibits the ATPase activity of Upf1, and this inhibition is released on Upf2 binding. Analysis of the structures revealed that the CH domain undergoes a significant conformational change upon binding to Upf2. This conformational change involves the displacement of the CH domain from the RecA2 domain, which in turn activates the helicase activity of Upf1 (Fig. 5.1.2, from (Chakrabarti et al. 2011)).

In yeast, Sen1 copurifies with the Alyref homologue Yra1 (Nedea et al. 2008), however whether this interaction is direct or conserved in humans was unclear. We thus investigated whether Senataxin can directly bind Alyref by using GST-Senataxin truncations (Suraweera et al. 2009) to pull down ^{35}S-labelled Alyref. We found that Alyref specifically bound a number of regions within Senataxin, and most intriguingly displayed a strong interaction with truncation S5 (Fig. 5.1.3, lane 6). By alignment with the Upf1 sequence, truncation S5 (aa1341-1673) was identified as spanning the Upf1-CH (aa115-281) region (Fig. 5.1.1 highlighted in red), and is therefore referred to as Setx-CH hereafter.

5.2 The interaction between Senataxin CH domain and Alyref disrupts the binding of CH to the helicase region

Given that Alyref was able to specifically bind to Senataxin-CH, we next wanted to ask whether Alyref could function to displace the CH domain from the bound helicase domain, comparable to the Upf1-Upf2 model. We therefore performed a displacement assay by pulling down radiolabelled helicase domain with GST-CH, then adding a titration of Alyref, or BSA as a control. The amount of the Setx-helicase that remained bound to GST-CH was then assessed by phosphorimage. The results show that the helicase domain can specifically bind to the GST-CH domain (Fig. 5.2.1. lane 2). The addition of increasing amounts of Alyref results in a displacement of the helicase domain as observed from the decrease in signal on the phosphorimage (Fig. 5.2.1, lanes 6-7), which is not observed when BSA is used as a competitor (lanes 8-10). The increasing amounts

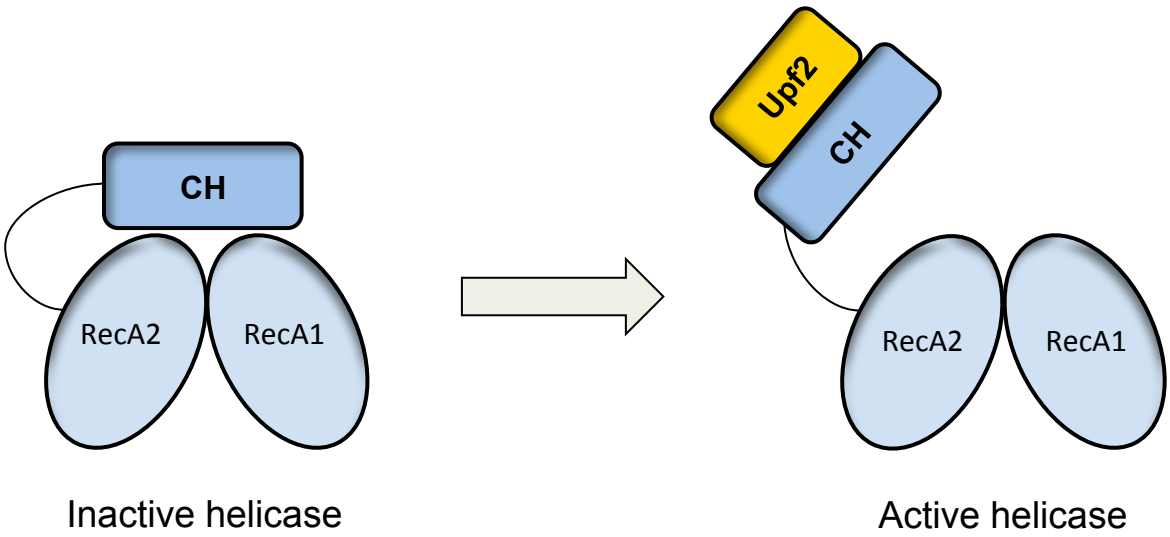


Figure 5.1.2. Model of Upf1 regulation. The association between Upf1 CH and helicase domains inhibits helicase activity. The binding of Upf2 to Upf1-CH triggers a conformational change that displaces the CH domain, and stimulating helicase activity (adapted from: Chakrabarti, S. et al., 2011)

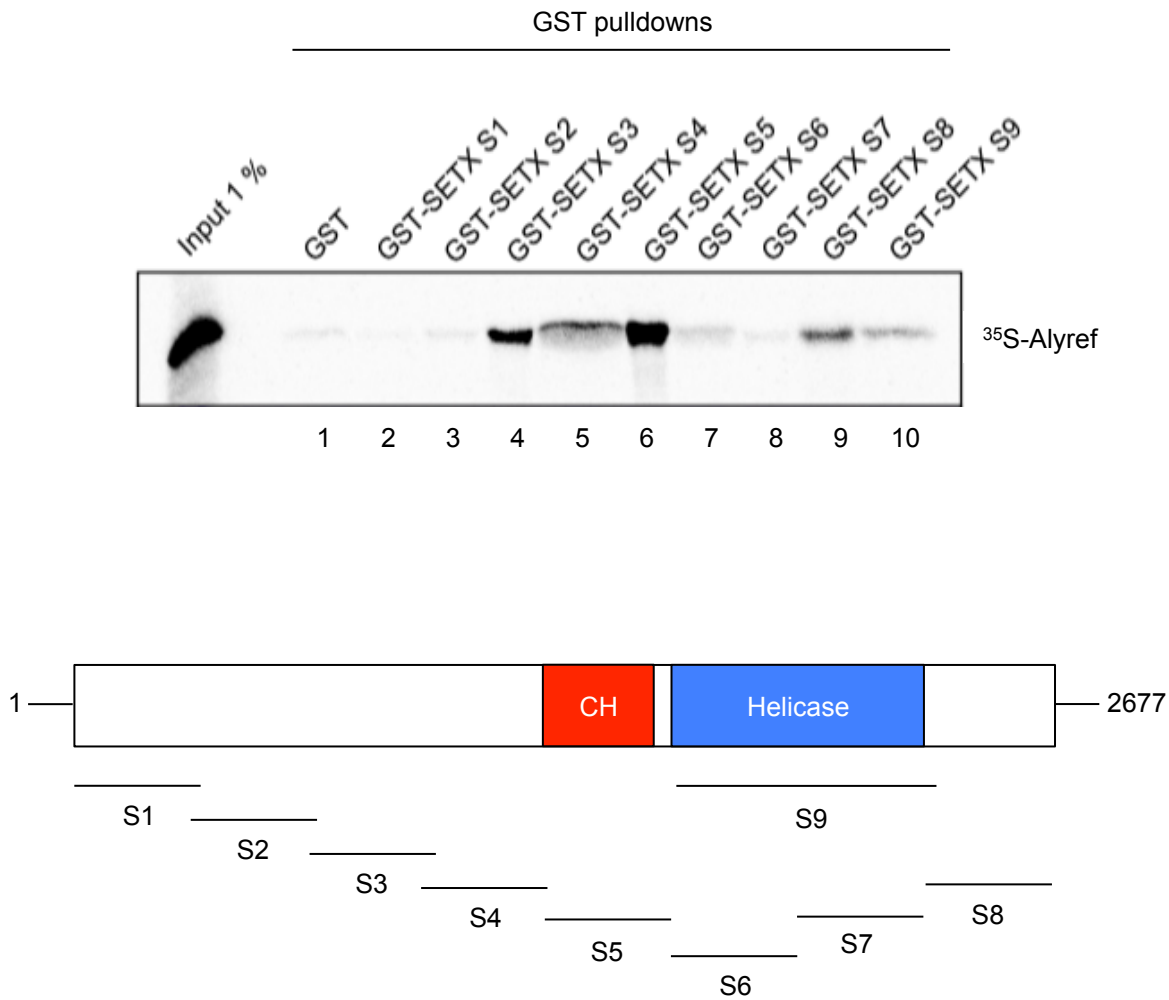


Figure 5.1.3. Alyref directly interacts with Senataxin. GST-Senataxin truncations (Suraweera, A. et al., 2009) were used to pull down radiolabelled Alyref. Alyref interacts strongly with truncation S5 (lane 6), which corresponds to the Upf1-CH domain by sequence alignment.

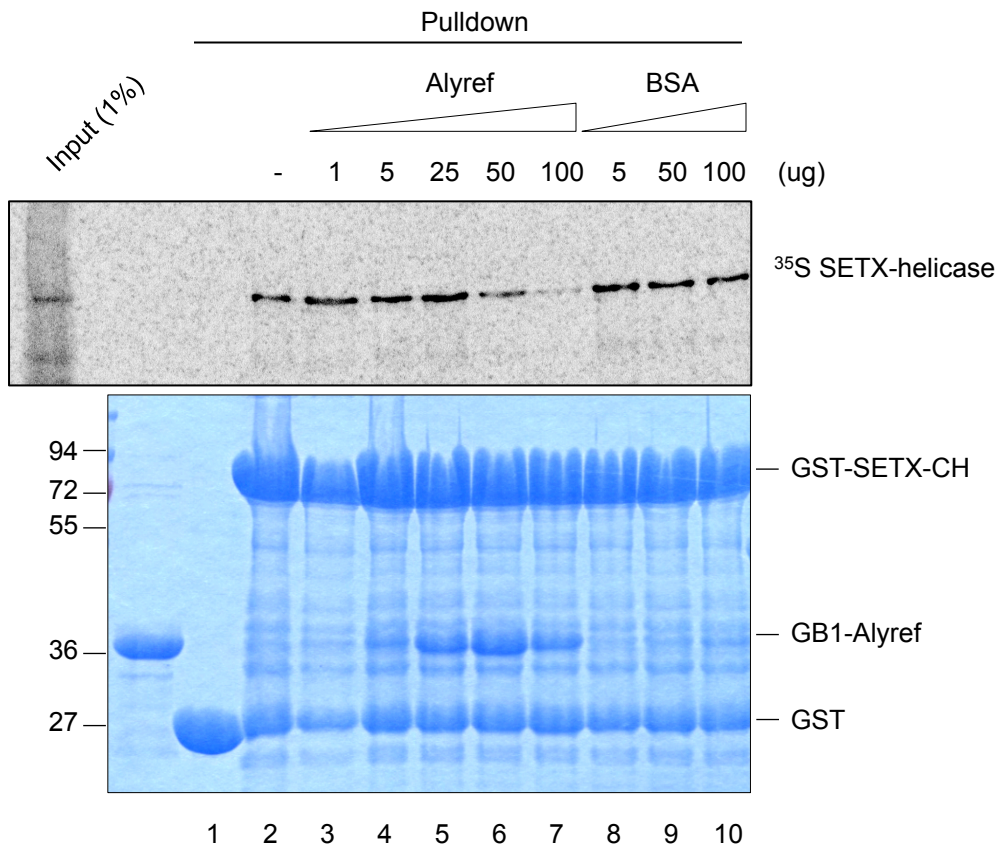


Figure 5.2.1. The interaction between Senataxin CH and helicase domains is disrupted by Alyref. GST-Setx-CH was able to directly interact with the helicase domain (lane 2). After this complex was formed, increasing amounts of Alyref was added resulting in a displacement of the helicase domain due to the interaction between Alyref and Setx-CH (lanes 3-7). BSA was used as a control to show that this effect was specific to Alyref (lanes 8-10).

of Alyref, which accompanies the loss of signal (seen in the coomassie), verifies that it is the interaction between Setx-CH and Alyref that causes the displacement of Setx-helicase from the CH domain.

To further investigate the Alyref interaction with Setx-CH domain, GST Setx-CH was used to pull down various truncations of radiolabelled Alyref. The results indicate that the region of Alyref required for the interaction with Setx-CH lies within amino acids 16-73 (Fig. 5.2.2 A). Of particular note, within this region is the N-terminal arginine-rich region that associates with RNA, and is required for binding to Nxf1. Additionally, the binding regions for Uap56 also overlaps with that of Setx-CH (Fig. 5.2.2 B), thus interaction between Alyref and Setx or Uap56 may be mutually exclusive. Given the recruitment of Alyref and Setx to the 3'-end of transcripts, this may indicate that Setx could serve as an intermediate between Alyref association with Uap56.

5.3 Senataxin does not interact with Alyref *in vivo*

Yeast Sen1 had previously been shown to co-purify with the Alyref homologue, Yra1 (Skourti-Stathaki et al. 2011) however, whether this interaction is conserved in humans had not been investigated. To identify whether Senataxin can interact with Alyref, full length Setx was cloned as a 3xFLAG-tagged construct, which was expressed in 293T cells by transient transfection, as well as used to build a stable tetracycline-inducible stable cell line as described in materials and methods. The expression of FLAG-Setx from total cell extracts was tested by Western blot using FLAG and Setx antibodies (Fig. 5.3.1). FLAG-Setx was immunoprecipitated on FLAG-agarose beads as described in the materials and methods and the eluate probed for co-purified Alyref, Uap56 and Hpr1. As shown in the results, Alyref was able to co-IP specifically with FLAG-Setx, which was over-expressed by transient transfection in 293T cells (Fig. 5.3.2 A). Additionally, a co-IP with Uap56 or Hpr1 was not observed, indicating that the fraction of Alyref that is associated with Senataxin is in a separate complex to that associated with Uap56 and Hpr1 and suggests that Setx is not associated

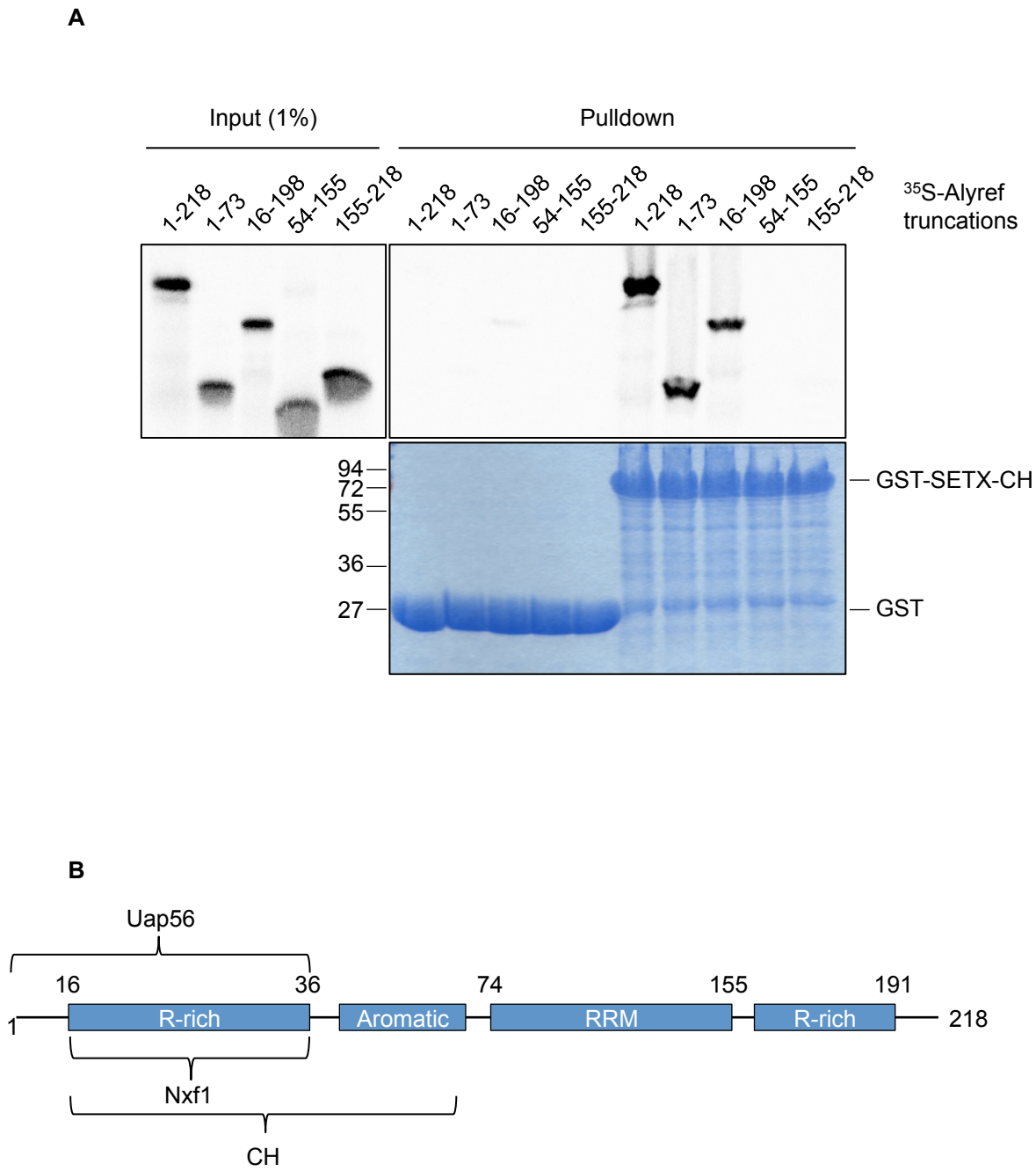


Figure 5.2.2. Indentification of the CH-binding region of Alyref. (A) GST-Setx-CH was used to pull down radiolabelled Alyref truncations. Setx-CH interacts specifically with the amino acids within the region 16-73. **(B)** Schematic of the Alyref domains (R-rich, Arginine-rich; RRM, RNA recognition motif). The Setx-CH interacting region overlaps with that required for Uap56 and Nxf1binding.

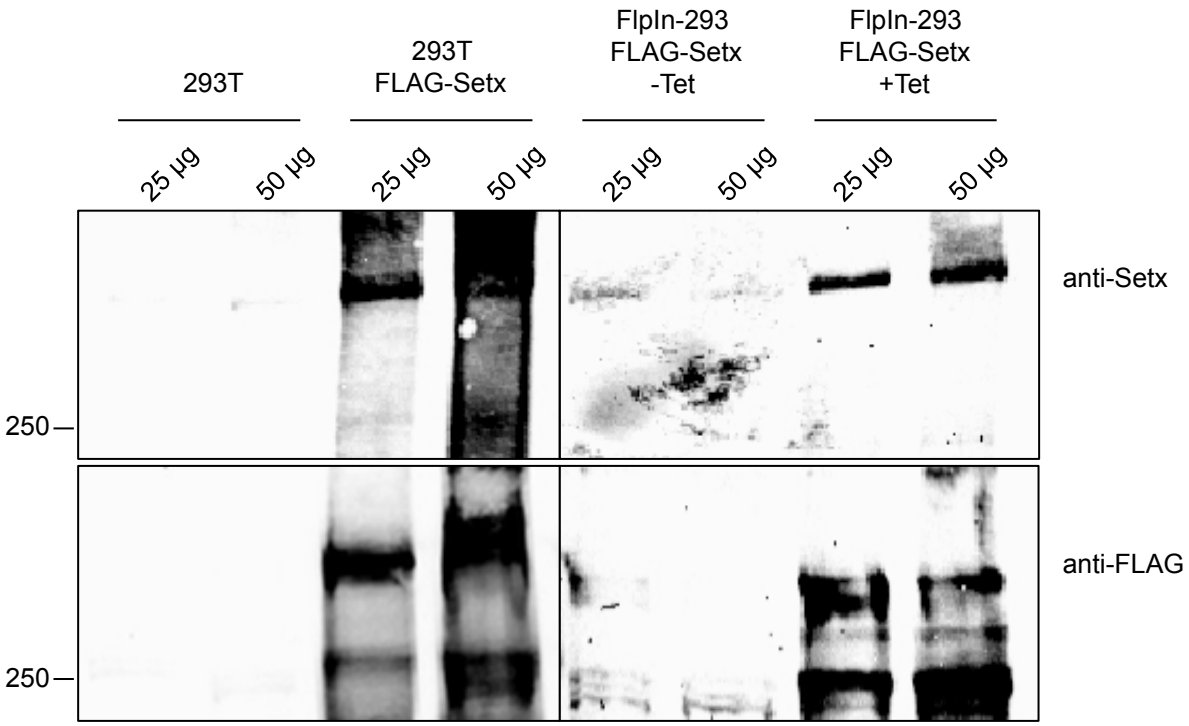


Figure 5.3.1. Construction of a FLAG-Setx expression plasmid and stable cell line. Full length Senataxin was cloned and expressed by transient transfection in 293T cells (panels on the left) or used to generate a tetracycline-inducible stable cell line (panels on the right).

with the TREX complex. We next performed the same experiment using stably expressed FLAG-Setx. In contrast to the IP with transiently expressed Setx, Alyref did not co-IP with stably expressed FLAG-Setx (Fig. 5.3.2 B). This result would suggest that the co-IP of Alyref with transiently expressed FLAG-Setx was biased in some way due to the over-expression of Setx. It is possible that the interaction between Setx and Alyref is a very transient event, thus the overexpression of Setx allowed the interaction to be observed, whereas expression of Setx at more endogenous levels meant that fewer molecules of Setx were available to bind to Alyref. Therefore, although the interaction between Setx and Alyref may be specific, it is unlikely to be a favorable event *in vivo*.

5.4 Investigating functional association between Alyref and Senataxin

We hypothesised that if Alyref was indeed involved in recruiting Senataxin to sites of transcription, knock down of Alyref would inhibit the association of Setx with the mRNP. To test this, we performed an mRNP capture assay with either Control or Alyref RNAi cell lines, then probed for Setx by Western blot (Fig. 5.4 A). As previously shown, knock down of Alyref does not affect the association of Uap56 with the mRNP due to it functioning upstream of Alyref in the mRNA export pathway. The Setx blot shows that the protein is associated with mRNA in the -UV condition where crosslinking between protein and RNA will not have taken place, and Setx is not enriched in the +UV condition. This would indicate that Setx is non-specifically associated with the oligo(dT) beads and therefore we were unable to draw any conclusions regarding its recruitment to the mRNP.

Indicative of TREX components functioning within the same pathway, knockdown of proteins within the complex results in an up-regulation of several other components (Chang et al. 2013). To test whether expression of Setx was regulated in a similar fashion, we used a stable RNAi cell line to knock down Alyref for 24, 48 and 72 hours and monitored the expression of Setx by Western blot. As expected, the depletion of Alyref is accompanied by an increase of

Uap56 expression. However, the expression of Setx remained the same across the 72 hours (Fig. 5.4 B). This result is interesting given the involvement of TREX in maintaining genome stability. The recruitment of TREX to the mRNP via Alyref means that depletion of Alyref should prevent proper assembly of TREX and result in genetic stress from R-loop formation. The expression of Setx therefore, does not appear to be regulated by this process, however, as Alyref depletion results in only a mild export block due to functional redundancy with other export adaptors, it would be interesting to explore whether depletion of other TREX components, which cause more stress on the cell (for example Alyref + UIF, or Alyref + ThoC5) would result in an upregulation of Setx.

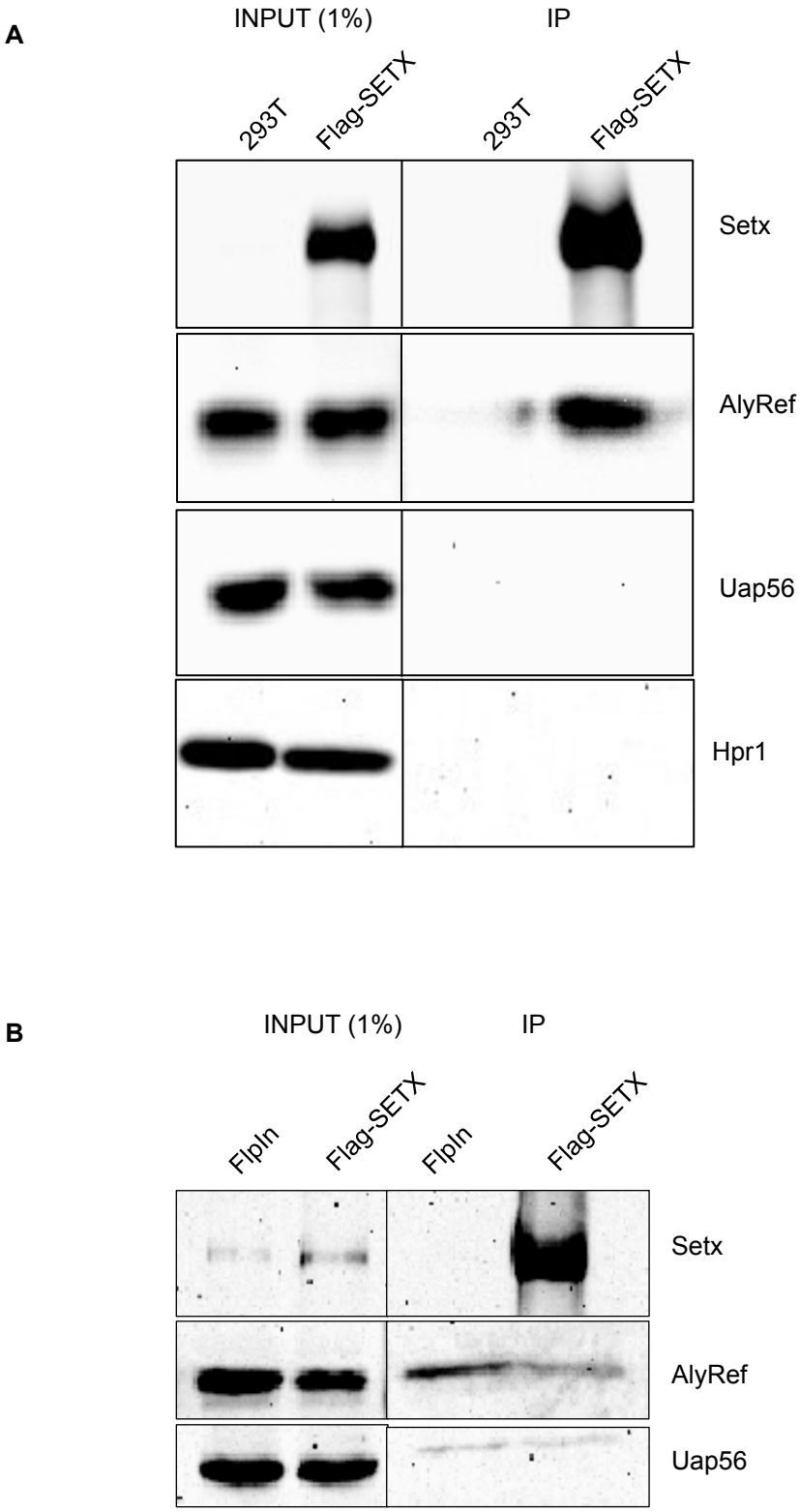


Figure 5.3.2. Interaction studies of Senataxin *in vivo*. (A) FLAG-Setx, overespressed by transient transfection was able to associate with Alyref in a Co-IP experiment. Associations with Uap56 and Hpr1 were not detected. **(B)** Co-IP of Alyref with Senataxin expressed from the stable cell line was not detected

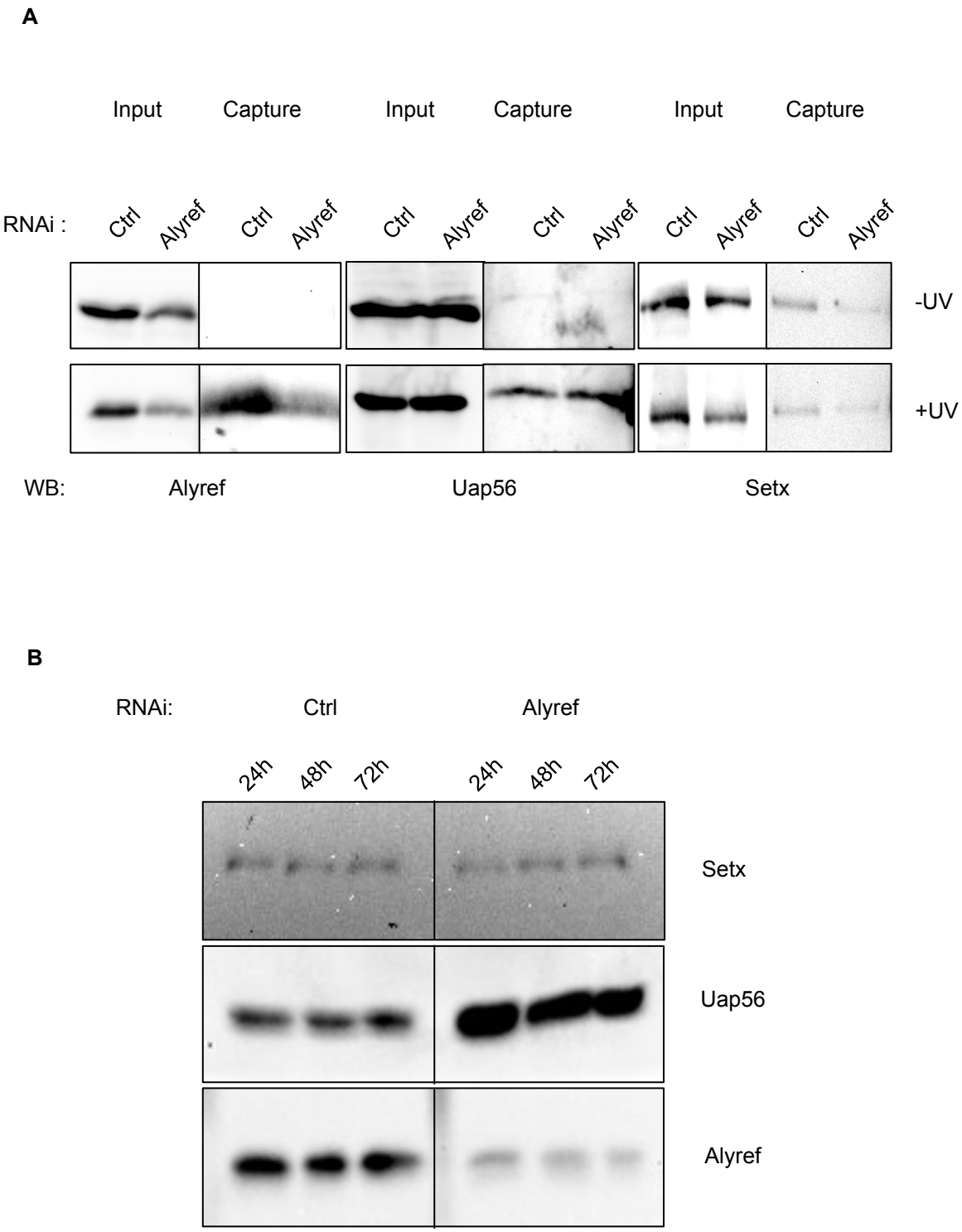


Figure 5.4. **(A)** mRNP capture in control and Alyref RNAi conditions. Senataxin is non-specifically pulled down on oligo(dT) resin independently of crosslinking by UV irradiation. **(B)** Senataxin protein expression is not affected by depletion of Alyref, as would be expected for a component that functions in the same pathway, such as Uap56, which is upregulated on Alref knock down.

5.5 Summary

During this investigation, we have identified that human Senataxin shares substantial sequence homology to Upf1. Given that the regulatory CH-domain of Upf1 is displaced upon binding Upf2, this was particularly intriguing in light of the observation that Alyref could directly interact with Setx-CH. Here, we observe for the first time that Senataxin is able to exhibit an intramolecular interaction between its CH and helicase domains as in Upf1, and further investigation demonstrated that that Alyref can disrupt this interaction by binding to Setx-CH. Alyref therefore, has the potential to regulate Setx activity in a similar fashion to Upf2 regulation of Upf1.

We next aimed to confirm whether the interaction between Setx and Alyref was relevant *in vivo.* Initial experiments indicated that Alyref was able to co-IP with transiently overexpressed Setx. Association between Setx and the TREX components Uap56 and Hpr1 was not observed. This suggested that the interaction of Alyref with Setx occurred upstream of the recruitment of Alyref to Uap56 and association with TREX. The Setx:Alyref interaction however, was not confirmed in IPs where Setx was stably expressed at near endogenous levels. The specific interaction between Setx and Alyref therefore, may be an artefact of using overexpressed proteins. Although this does not necessarily mean Setx and Alyref don't associate *in vivo*, the low expression of Setx may cause difficulty in observing the interaction under conditions where Setx is expressed at endogenous levels. Additionally, the Yra1 ChIP (chromatin IP) profile indicates that it is enriched at the 3' end of open reading frames (S. A. Johnson et al. 2011). By comparison, Senataxin is enriched at the 5' of pause sites, downstream of the poly(A) signal (Skourti-Stathaki et al. 2011). The recruitment of Alyref and Senataxin to genes, therefore do not overlap, further suggesting that Alyref is unlikely to associate with Senataxin *in vivo*. Although Alyref may not function to regulate Senataxin activity as we initially hypothesised, another protein may act in this capacity to disrupt the Senataxin CH:helicase interaction. For example,

Senataxin associates with RNAPII and several RNA processing factors (Suraweera et al. 2009; Yüce & S. C. West 2013), which may provide a regulatory function on Senataxin activity.

Chapter 6

Discussion

Characterisation of the Nxf1 intramolecular interaction

Nxf1 is the major export receptor for RNA, featuring a modular domain organisation with an N-terminal RNA-binding domain (RBD), pseudo RNA-recognition motif (ΨRRM), leucine-rich region (LRR), NTF2-like domain (NTF2L) and ubiquitin-associated domain (UBA). Inherently, it exhibits a weak RNA binding activity, which is enhanced dramatically upon association with export adaptor proteins, which directly handover mRNA to the Nxf1 RBD (Hautbergue et al. 2008). Furthermore, Nxf1 appears to regulate its own RNA binding activity by forming an intramolecular interaction between its RBD and NTF2L domains, which is perturbed by the association of the adaptor Alyref and co-adaptor Thoc5 (Viphakone et al. 2012). In this study, we have defined a number of important features involved in the RBD:NTF2L interaction. The RBD contains a sequence of ten arginines, which are essential for RNA interaction (Hautbergue et al. 2008). Here we show that these arginines appear to also be required for the interaction with NTF2L, thus providing competition between NTF2L and RNA binding. Though these arginines are necessary for the interaction, they were not sufficient to bind in the absence of an N-terminal patch of acidic residues, suggesting an ionic component involved in the RBD:NTF2L interaction. By studying the 3D structure of NTF2L (Fribourg et al. 2001), we were able to identify a number of charged residues on the surface of the domain. Importantly, many of these residues had previously been shown to be essential for Thoc5 binding (Katahira et al. 2009). Through mutagenesis of these residues, we were able to conclude that the RBD and Thoc5 binding sites overlap on the NTF2L domain. The recruitment of Nxf1 by TREX therefore, serves to unlock Nxf1 into an efficient RNA binding conformation through the direct interaction with Alyref and Thoc5.

We had aimed to define the RBD:NTF2L interaction in closer detail through crystallographic studies. The solubility issues encountered with the RBD constructs presented a significant barrier to this investigation, however we were also unable to produce crystals of the NTF2L-UBA region, of which the crystal structure has been solved (Fribourg et al. 2001; Liker et al. 2000). Nxf1 contains flexible linker regions that separate each domain (Fribourg et al. 2001), which are clearly important for the conformational changes involved in Nxf1 function. These same linkers however, may cause the difficulty in Nxf1 crystallisation. Obtaining crystallographic data requires the protein molecules to pack together and align in the same orientation in order to diffract X-rays in a consistent manner. Flexible regions are often missing from crystal structures, as observed with the Nxf1 structures cited above, since they do not diffract the X-rays in a consistent pattern.

Assembly of the human THO complex

Recently, the EM structure for the *S. cerevisiae* THO was solved providing a view of the architecture of components within the complex (Peña et al. 2012). Human THO contains homologues to the yeast Tho2p, Hpr1p and Tex1, but also contains additional components with no known counterparts in yeast. We were therefore interested in obtaining structural information of the human THO complex. To this end, we expressed the THO components in a baculovirus or *E. coli* system and then attempted to assemble a complete complex. The results showed that only a partial reconstitution of the complex was observed. A number of factors may have contributed to this, in particular, the low expression of some components would have been unfavorable to the assembly of a stoichiometric complex, and the presence of contaminating proteins may have interfered with protein interactions. Additionally, the EM structure of the yeast complex was obtained by purification of the native complex (Peña et al. 2012), whereas we were reconstituting the separate components from recombinant systems. This suggests that processes and/or components present *in vivo* may be involved in the assembly and stabilisation of THO, which were unaccounted for in

our recombinant system. For example, the association of THO with TREX, of which several new components have been identified (Dufu et al. 2010), may facilitate THO assembly.

Investigating a functional link between TREX and Microprocessor

Microprocessor consists of the proteins Drosha and DGCR8, which function in nuclear processing of miRNAs. Evidence from a number of recent studies have indicated that this complex is also associated with regulating gene expression, independently of its miRNA processing function (Macias et al. 2012; Gromak et al. 2013). Microprocessor was found to regulate its own expression through a feedback loop involving the stabilisation of Drosha by the DGCR8 protein, and cleavage of DGCR8 mRNA by Drosha (Triboulet et al. 2009; Han et al. 2009). Further studies found that the majority of Microprocessor targets were found in protein-coding mRNAs, and that Microprocessor activity affected the expression of these mRNAs (Macias et al. 2012). Significantly, depletion of Microprocessor resulted in an upregulation of these mRNAs, a large proportion of which overlapped with mRNAs upregulated upon TREX depletion (unpublished data). This led to the intriguing possibility that TREX was involved in recruitment of Microprocessor to transcripts, and regulate gene expression by promoting mRNA cleavage. By knocking down TREX components Alyref and Thoc5, we examined the effect of TREX depletion on mRNA stability. As expected, the results showed an increase in protein expression of several TREX components. Consistent with the effect seen in Drosha depleted cells (Han et al. 2009), DGCR8 also displayed an upregulation upon Alyref and Thoc5 knock down. However, qPCR analysis revealed that DGCR8 RNA was downregulated when Alyref and Thoc5 were knocked down, indicating that the increased protein expression observed was not due to stabilisation of its mRNA. Additionally, miRNA array analysis revealed that Alyref and Thoc5 knock down did not affect the processing of miRNA. Together, this evidence appears to refute our hypothesis that TREX recruits Microprocessor to mRNAs.

Recent work demonstrated that Drosha functions independently of RNA

cleavage by promoting transcription (Gromak et al. 2013). In this work, the Microprocessor components were enriched at 5' promoter regions of human genes, and depletion of Drosha was associated with transcriptional defects. Furthermore, this function was dependent on the interaction between Drosha and the CBC component Cbp80. Similarly, TREX is recruited to the 5' end of genes through the interaction between Alyref an Cbp80 (Cheng et al. 2006) and functions in transcriptional elongation (Domínguez-Sánchez, Barroso, et al. 2011a). The overlap in the association of Microprocessor and TREX to the 5' end of genes may explain the functional link between these complexes. In this context, Microprocessor may be involved in the early stage of transcription, and then hand over to Alyref and TREX upon their recruitment to the CBC.

Genome instability and Senataxin

Depletion of TREX proteins is associated with genome instability in both yeast and humans (García-Rubio et al. 2007; Domínguez-Sánchez, Barroso, et al. 2011a; Chávez & Aguilera 1997). The apparent cause of this was due to the formation of R-loops (Huertas & Aguilera 2003), which are RNA:DNA hybrid structures formed between the nascent mRNA transcript and the complementary DNA strand, and a displaced strand of DNA that is susceptible to damage. Senataxin (Setx) is a putative RNA:DNA helicase that functions to prevent R-loop accumulation and protect the genome (Skourti-Stathaki et al. 2011). In yeast, the Setx homologue was found to co-purify with the TREX component Yra1 (Alyref in humans), indicating a potential role of TREX in regulating Setx activity to maintain genome stability. Alignment of Setx with a related helicase, Upf1, revealed that Setx contained a CH-like domain. In Upf1, this domain binds to its own helicase domain to attenuate its activity. The interaction between Upf1-CH and a co-factor, Upf2, displaces the CH domain to promote helicase activity. Our results found that Setx also displays an intramolecular interaction between its CH and helicase domain, which is perturbed by the interaction of Alyref with Setx-CH. Thus, Alyref could potentially function analogously to Upf2, displacing Setx-CH from the helicase

domain and release the inhibition of helicase activity. Subsequent experiments however, indicated that Setx and Alyref did not associate under *in vivo* conditions. Furthermore, depletion of Alyref did not affect the expression of Setx, indicating the two proteins are functionally separated.

Future Work

One of the key aims of this project was to gather structural data on components within the TREX complex in order to closely define the interactions involved. Despite successfully purifying proteins for crystallisation screens, we were unable to obtain crystals for further structural analysis. A significant obstacle to the crystallisation process is the flexibility of protein molecules, as the movement of the molecules presents an entropic barrier toward the formation of ordered molecules required for crystal formation. The flexible regions between the Nxf1 domains is well documented (Fribourg et al. 2001; Liker et al. 2000) and may therefore contribute to the difficulty of crystallising this protein. Additionally, the NMR structure of Alyref demonstrated that the flexible domains of this protein are essential for its function (Golovanov et al. 2006). Flexible domains may therefore be a general characteristic of RNA binding proteins and thus present significant difficulty for crystallography of TREX components. Further structural analysis using NMR may therefore provide a more suitable medium and provide a more dynamic picture of TREX interactions.

We were also interested in obtaining structural data of the human Tho complex to define the architecture of the components within. A recent study had solved the structure of yeast Tho, by purifying the native complex directly from yeast cells (Peña et al. 2012). We were however, unable to assemble a complete Tho complex using a recombinant protein approach. This difficulty may be due to the low expression of proteins or co-purifying contaminants, but may also indicate that Tho assembly is dependent on biological processes or components that were not included in our recombinant approach. For example, the recently identified TREX associated proteins, some of which directly associate with Tho components, may be involved in the assembly or stabilisation of a mature Tho

complex. Though the precise functions of these putative TREX proteins are not fully characterised, there is evidence that Skar and Zc3h11a are associated with mRNA processing and export (Folco et al. 2012; Ma et al. 2008). Purifying a native TREX complex directly from cultured cells therefore, would ensure these factors are accounted for and may provide a good route for providing structural data. However, more interesting would be to identify which, if any, of these new TREX proteins are essential for the assembly of the complex. This could be achieved for example, by performing knock down screens of the various components then assessing the integrity of the TREX complex by IP. As well as identifying key components within TREX, the results could also inform the strategies to be used in acquiring structural data of the complex. Furthermore, the proteins Skar an Zc3h11a associate with TREX in an ATP dependent manner (Folco et al. 2012) as Alyref and Cip29 do (Dufu et al. 2010). The ATPase cycle of Uap56 has been suggested to play a role in the remodelling and maturation of the TREX complex (Chang et al. 2013). Inclusion of Uap56 and hydrolysable ATP may therefore also be a prerequisite for assembling a TREX complex *in vitro*, and future studies may focus on identifying what effect the putative TREX components have on the ATPase cycle and TREX assembly.

The TREX complex also appears to have a functional association with Microprocessor. Unpublished data from the lab had identified an upregulation of components in the miRNA processing pathway upon TREX depletion. Additionally, there was a significant overlap in the mRNAs upregulated in this data, and the mRNAs upregulated upon Microprocessor depletion (Macias et al. 2012). We found that TREX is able to associate with components involved in miRNA processing (Drosha and Ars2) and examined whether TREX functioned to recruit Microprocessor. The evidence we compiled however, found that though depletion of TREX affects the expression of Microprocessor, it was not in a manner indicative of the recruitment of Microprocessor to mRNA transcripts. A recent study suggested a role for Microprocessor in transcriptional elongation, independent of its RNA cleavage function (Gromak et al. 2013). In this model, Microprocessor associates to the 5' end of genes involving an interaction with

Cbp80 to promote efficient transcription. TREX is also reported to be recruited to the 5' end of genes via the interaction between Alyref and Cbp80, thus it would be interesting to examine whether Microprocessor is involved in the recruitment of TREX, and how mRNA export is affected by this association. For instance, if Microprocessor recruits TREX to the RNA, we would predict a decrease in the amount of RNA bound to TREX upon depletion of Microprocessor. This could be assessed for by RNA immunoprecipitation assays coupled with RT-PCR to quantify the amount of RNA bound to TREX.

Finally, we investigated the potential role of TREX in maintaining genome stability in the context of the RNA:DNA helicase, Senataxin. Setx functions to resolve R-loops that form during transcription and prevent DNA damage. Depletion of Setx results in accumulation of R-loops (Skourti-Stathaki et al. 2011) and this is also observed when TREX subunits are mutated (Gómez-González et al. 2011), however the relationship between Setx and TREX had not been examined. Our observation that Alyref can bind directly to the Setx-CH and disrupt the intramolecular interaction between Setx-CH and helicase domains presented an intriguing possibility that TREX could regulate Setx helicase activity. However, further investigation suggested that the Alyref:Setx interaction was not observable under *in vivo* conditions. The first focus of continued investigation should be to determine that Setx is a true helicase. Although Setx clearly plays a role in preventing R-loop formation, evidence for this is indirectly observed from R-loop accumulation upon Setx depletion. Using *in vitro* helicase assays, for example, would prove directly that Setx is a bona fide helicase. It would then be important to investigate the biological significance of the Setx intramolecular interaction, specifically whether or not the CH domain does in fact inhibit helicase activity. Furthermore, if Setx does regulate its own activity in this manner then it would be interesting to characterise how this inhibition of helicase activity is released. The strong association of Setx with RNAPII and several RNA processing factors (Suraweera et al. 2009; Yüce & S. C. West 2013) provides enticing targets to screen for potential co-factors of Setx that could function in this role. If Setx activity is regulated in a similar fashion to the Upf1

model, its helicase activity would be stimulated upon addition of such a co-factor. Additionally, the *in vivo* relevance of the Setx intramolecular interaction could be investigated by engineering an internal deletion mutant lacking the CH domain. In theory, if the CH domain inhibits helicase activity then this mutant would be constitutively active, which could be assessed for by measuring the impact on R-loop formation.

References

Abruzzi, K.C., Lacadie, S. & Rosbash, M., 2004. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *The EMBO Journal*, 23(13), pp.2620–2631.

Aguilera, A., 2002. The connection between transcription and genomic instability. *The EMBO Journal*, 21(3), pp.195–201.

Aguilera, A. & García-Muse, T., 2012. R loops: from transcription byproducts to threats to genome stability. *Molecular Cell*, 46(2), pp.115–124.

Aguilera, A. & Klein, H.L., 1988. Genetic control of intrachromosomal recombination in Saccharomyces cerevisiae. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics*, 119(4), pp.779–790.

Arts, G.J., Fornerod, M. & Mattaj, I.W., 1998. Identification of a nuclear export receptor for tRNA. *Current biology : CB*, 8(6), pp.305–314.

Assenberg, R. et al., 2009. Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. *Journal of virology*, 83(24), pp.12895–12906.

Bayliss, R., Leung, S.W., et al., 2002a. Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats. *The EMBO Journal*, 21(12), pp.2843–2853.

Bayliss, R., Littlewood, T., et al., 2002b. GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta. *The Journal of biological chemistry*, 277(52), pp.50597–50606.

Bohnsack, M.T. et al., 2002. Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. *The EMBO Journal*, 21(22), pp.6205–6215.

Braun, I.C. et al., 1999. TAP binds to the constitutive transport element (CTE) through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus. *The EMBO Journal*, 18(7), pp.1953–1965.

Calado, A. et al., 2002. Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *The EMBO Journal*, 21(22), pp.6216–6224.

Chakrabarti, S. et al., 2011. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Molecular Cell*, 41(6), pp.693–703.

Chang, C.-T. et al., 2013. Chtop is a component of the dynamic TREX mRNA export complex. *The EMBO Journal*.

Chávez, S. & Aguilera, A., 1997. The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes & Development*, 11(24), pp.3459–3470.

Chávez, S. et al., 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *The EMBO Journal*, 19(21), pp.5824–5834.

Chekulaeva, M. & Filipowicz, W., 2009. Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Current Opinion in Cell Biology*, 21(3), pp.452–460.

Chen, Y.-Z. et al., 2004. DNA/RNA Helicase Gene Mutations in a Form of Juvenile Amyotrophic Lateral Sclerosis (ALS4). *The American Journal of Human Genetics*, 74(6), pp.1128–1135.

Chen, Y.-Z. et al., 2006. Senataxin, the yeast Sen1p orthologue: characterization of a unique protein in which recessive mutations cause ataxia and dominant mutations cause motor neuron disease. *Neurobiology of disease*, 23(1), pp.97–108.

Cheng, H. et al., 2006. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, 127(7), pp.1389–1400.

Clouse, K.N. et al., 2001. A Ran-independent pathway for export of spliced mRNA. *Nature cell biology*, 3(1), pp.97–99.

Cronshaw, J.M. et al., 2002. Proteomic analysis of the mammalian nuclear pore complex. *The Journal of Cell Biology*, 158(5), pp.915–927.

Culjkovic-Kraljacic, B. & Borden, K.L.B., 2013. Aiding and abetting cancer: mRNA export and the nuclear pore. *Trends in cell biology*, 23(7), pp.328–335.

D'Angelo, M.A. & Hetzer, M.W., 2008. Structure, dynamics and function of nuclear pore complexes. *Trends in cell biology*, 18(10), pp.456–466.

Domínguez-Sánchez, M.S., Barroso, S., et al., 2011a. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS genetics*, 7(12), p.e1002386.

Domínguez-Sánchez, M.S., Sáez, C., et al., 2011b. Differential expression of THOC1 and ALY mRNP biogenesis/export factors in human cancers. *BMC cancer*, 11, p.77.

Dufu, K. et al., 2010. ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes & Development*, 24(18), pp.2043–2053.

Fischer, U. et al., 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*, 82(3), pp.475–483.

Fleckner, J. et al., 1997. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes & Development*, 11(14), pp.1864–1872.

Folco, E.G. et al., 2012. The proteins PDIP3 and ZC11A associate with the human TREX complex in an ATP-dependent manner and function in mRNA export. *PLoS ONE*, 7(8), p.e43804.

Fornerod, M., Ohno, M., et al., 1997a. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, 90(6), pp.1051–1060.

Fornerod, M., van Deursen, J., et al., 1997b. The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. *The EMBO Journal*, 16(4), pp.807–816.

Fribourg, S. et al., 2001. Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. *Molecular Cell*, 8(3), pp.645–656.

Gadal, O. et al., 2001. Nuclear export of 60s ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Molecular and Cellular Biology*, 21(10), pp.3405–3415.

García-Rubio, M. et al., 2007. Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Molecular genetics and genomics : MGG*, 279(2), pp.123–132.

Garneau, N.L., Wilusz, J. & Wilusz, C.J., 2007. The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology*, 8(2), pp.113–126.

Gatfield, D. & Izaurralde, E., 2002. REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *The Journal of Cell Biology*, 159(4), pp.579–588.

Golovanov, A.P. et al., 2006. The solution structure of REF2-I reveals interdomain interactions and regions involved in binding mRNA export factors and RNA. *RNA*, 12(11), pp.1933–1948.

Gómez-González, B. et al., 2011. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *The EMBO Journal*, 30(15), pp.3106–3119.

Görlich, D. & Kutay, U., 1999. Transport between the cell nucleus and the cytoplasm. *Annual review of cell and developmental biology*, 15, pp.607–660.

Görlich, D. et al., 1997. A novel class of RanGTP binding proteins. *The Journal of Cell Biology*, 138(1), pp.65–80.

Gromak, N. et al., 2013. Drosha regulates gene expression independently of RNA cleavage function. *Cell reports*, 5(6), pp.1499–1510.

Gruber, J.J. et al., 2009. Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell*, 138(2), pp.328–339.

Grünwald, D., Singer, R.H. & Rout, M., 2011. Nuclear export dynamics of RNA-protein complexes. *Nature*, 475(7356), pp.333–341.

Grüter, P. et al., 1998. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Molecular Cell*, 1(5), pp.1–11.

Guo, S., Liu, M. & Godwin, A.K., 2012. Transcriptional regulation of hTREX84 in human cancer cells. *PLoS ONE*, 7(8), p.e43610.

Gwizdek, C. et al., 2003. Exportin-5 mediates nuclear export of minihelix-containing RNAs. *The Journal of biological chemistry*, 278(8), pp.5505–5508.

Hallais, M. et al., 2013. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nature structural & molecular biology*.

Han, J. et al., 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5), pp.887–901.

Han, J. et al., 2009. Posttranscriptional crossregulation between Drosha and DGCR8. *Cell*, 136(1), pp.75–84.

Han, J. et al., 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development*, 18(24), pp.3016–3027.

Hautbergue, G.M. et al., 2008. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proceedings of the National Academy of Sciences of the United States of America*, 105(13), pp.5154–5159.

Hautbergue, G.M. et al., 2009. UIF, a New mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. *Current biology : CB*, 19(22), pp.1918–1924.

Herold, A., Klymenko, T. & Izaurralde, E., 2001. NXF1/p15 heterodimers are essential for mRNA nuclear export in Drosophila. *RNA*, 7(12), pp.1768–1780.

Ho, J.H., Kallstrom, G. & Johnson, A.W., 2000. Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *The Journal of Cell Biology*, 151(5), pp.1057–1066.

Hopper, A.K., Schultz, L.D. & Shapiro, R.A., 1980. Processing of intervening sequences: a new yeast mutant which fails to excise intervening sequences from precursor tRNAs. *Cell*, 19(3), pp.741–751.

Huang, Y. et al., 2003. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular Cell*, 11(3), pp.837–843.

Huertas, P. & Aguilera, A., 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell*, 12(3), pp.711–721.

Hurt, D.J. et al., 1987. Cloning and characterization of LOS1, a Saccharomyces cerevisiae gene that affects tRNA splicing. *Molecular and Cellular Biology*, 7(3), pp.1208–1216.

Izaurralde, E. et al., 1997. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. *The EMBO Journal*, 16(21), pp.6535–6547.

Jackson, B.R., Noerenberg, M. & Whitehouse, A., 2014. A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. *PLoS pathogens*, 10(5), p.e1004098.

Jimeno, S. et al., 2002. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *The EMBO Journal*, 21(13), pp.3526–3535.

Johnson, D.C. & Baines, J.D., 2011. Herpesviruses remodel host membranes for virus egress. *Nature reviews. Microbiology*, 9(5), pp.382–394.

Johnson, S.A. et al., 2011. The export factor Yra1 modulates mRNA 3' end processing. *Nature structural & molecular biology*, 18(10), pp.1164–1171.

Kadener, S. et al., 2009. Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. *RNA*, 15(4), pp.537–545.

Katahira, J. et al., 2009. Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. *The EMBO Journal*, 28(5), pp.556–567.

Kim, V.N., 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nature reviews. Molecular cell biology*, 6(5), pp.376–385.

Kim, V.N., Han, J. & Siomi, M.C., 2009. Biogenesis of small RNAs in animals. *Nature reviews. Molecular cell biology*, 10(2), pp.126–139.

Kiseleva, E. et al., 1998. Active nuclear pore complexes in Chironomus: visualization of transporter configurations related to mRNP export. *Journal of cell science*, 111 (Pt 2), pp.223–236.

Köhler, A. & Hurt, E., 2007. Exporting RNA from the nucleus to the cytoplasm. *Nature reviews. Molecular cell biology*, 8(10), pp.761–773.

Kutay, U. et al., 1998. Identification of a tRNA-specific nuclear export receptor. *Molecular Cell*, 1(3), pp.359–369.

Le Hir, H. et al., 2001. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *The EMBO Journal*, 20(17), pp.4987–4997.

Lee, Y. et al., 2004. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, 23(20), pp.4051–4060.

Libri, D. et al., 2002. Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Molecular and Cellular Biology*, 22(23), pp.1–13.

Liker, E. et al., 2000. The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. *The EMBO Journal*, 19(21), pp.5587–5598.

Longman, D., Johnstone, I.L. & Cáceres, J.F., 2003. The Ref/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans. *RNA*, 9(7), pp.881–891.

Luo, M.J. & Reed, R., 1999. Splicing is required for rapid and efficient mRNA export in metazoans. *Proceedings of the National Academy of Sciences of the United States of America*, 96(26), pp.1–6.

Ma, X.M. et al., 2008. SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell*, 133(2), pp.303–313.

Macias, S. et al., 2012. DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nature structural & molecular biology*, 19(8), pp.760–766.

Masuda, S. et al., 2005. Recruitment of the human TREX complex to mRNA during splicing. *Genes & Development*, 19(13), pp.1512–1517.

Mattaj, I.W. & Englmeier, L., 1998. Nucleocytoplasmic transport: the soluble phase. *Annual review of biochemistry*, 67, pp.265–306.

Mischo, H.E. et al., 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular Cell*, 41(1), pp.21–32.

Mitchell, P. & Tollervey, D., 2003. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3“-->5” degradation. *Molecular Cell*, 11(5), pp.1405–1413.

Moreira, M.-C. et al., 2004. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature genetics*, 36(3), pp.225–227.

Natalizio, B.J. & Wente, S.R., 2013. Postage for the messenger: designating routes for nuclear mRNA export. *Trends in cell biology*, 23(8), pp.365–373.

Nedea, E. et al., 2008. The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Molecular Cell*, 29(5), pp.577–587.

Nickoloff, J.A., 1992. Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Molecular and Cellular Biology*, 12(12), pp.5311–5318.

Ohno, M. et al., 2000. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell*, 101(2), pp.187–198.

Peña, A. et al., 2012. Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *The EMBO Journal*, 31(6), pp.1605–1616.

Piruat, J.I. & Aguilera, A., 1998. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *The EMBO Journal*, 17(16), pp.4859–4872.

Prado, F., Piruat, J.I. & Aguilera, A., 1997. Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. *The EMBO Journal*, 16(10), pp.2826–2835.

Proudfoot, N., 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Current Opinion in Cell Biology*, 16(3), pp.272–278.

Proudfoot, N.J., 2011. Ending the message: poly(A) signals then and now. *Genes & Development*, 25(17), pp.1770–1782.

Rabut, G., Lénárt, P. & Ellenberg, J., 2004. Dynamics of nuclear pore complex organization through the cell cycle. *Current Opinion in Cell Biology*, 16(3), pp.314–321.

Ribbeck, K. & Görlich, D., 2001. Kinetic analysis of translocation through nuclear pore complexes. *The EMBO Journal*, 20(6), pp.1320–1330.

Richardson, C.J. et al., 2004. SKAR is a specific target of S6 kinase 1 in cell growth control. *Current biology : CB*, 14(17), pp.1540–1549.

Rout, M.P. & Aitchison, J.D., 2001. The nuclear pore complex as a transport machine. *The Journal of biological chemistry*, 276(20), pp.16593–16596.

Rout, M.P. et al., 2000. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *The Journal of Cell Biology*, 148(4), pp.635–651.

Schumann, S. et al., 2013. Kaposi's sarcoma-associated herpesvirus ORF57 protein: exploiting all stages of viral mRNA processing. *Viruses*, 5(8), pp.1901–1923.

Shatkin, A.J. & Manley, J.L., 2000. The ends of the affair: capping and polyadenylation. *Nature structural biology*, 7(10).

Shi, H. et al., 2004. Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), pp.17628–17633.

Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N., 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular Cell*, 42(6), pp.794–805.

Smyk, A. et al., 2006. Human enhancer of rudimentary is a molecular partner of PDIP46/SKAR, a protein interacting with DNA polymerase delta and S6K1 and regulating cell growth. *The FEBS journal*, 273(20), pp.4728–4741.

Speese, S.D. et al., 2012. Nuclear envelope budding enables large ribonucleoprotein particle export during synaptic Wnt signaling. *Cell*, 149(4), pp.832–846.

Suraweera, A. et al., 2009. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Human Molecular Genetics*, 18(18), pp.3384–3396.

Taniguchi, I. & Ohno, M., 2008. ATP-dependent recruitment of export factor Aly/REF onto intronless mRNAs by RNA helicase UAP56. *Molecular and Cellular Biology*, 28(2), pp.601–608.

Thain, A. et al., 1996. A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends in genetics : TIG*, 12(6), pp.209–210.

Thomas, F. & Kutay, U., 2003. Biogenesis and nuclear export of ribosomal subunits in higher eukaryotes depend on the CRM1 export pathway. *Journal of cell science*, 116(Pt 12), pp.2409–2419.

Tian, X. et al., 2013. The interaction of the cellular export adaptor protein Aly/REF with ICP27 contributes to the efficiency of herpes simplex virus 1 mRNA export. *Journal of virology*, 87(13), pp.7210–7217.

Triboulet, R. et al., 2009. Post-transcriptional control of DGCR8 expression by the Microprocessor. *RNA*, 15(6), pp.1005–1011.

Trowitzsch, S. et al., 2010. New baculovirus expression tools for recombinant protein complex production. *Journal of structural biology*, 172(1), pp.45–54.

Viphakone, N. et al., 2012. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nature communications*, 3, p.1006.

Wang, Y. et al., 2002. Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), pp.5860–5865.

West, S., Gromak, N. & Proudfoot, N.J., 2004. Human 5“ --> 3” exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016), pp.522–525.

Wu, X. & Brewer, G., 2012. The regulation of mRNA stability in mammalian cells: 2.0. *Gene*, 500(1), pp.10–21.

Yang, E. et al., 2003. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome research*, 13(8), pp.1863–1872.

Yang, J. et al., 2001. Two closely related human nuclear export factors utilize entirely distinct export pathways. *Molecular Cell*, 8(2), pp.397–406.

Yang, Q., Rout, M.P. & Akey, C.W., 1998. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Molecular Cell*, 1(2), pp.223–234.

Yeom, K.-H. et al., 2006. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic acids research*, 34(16), pp.4622–4629.

Yi, R. et al., 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development*, 17(24), pp.3011–3016.

Yoon, D.W. et al., 1997. Tap: a novel cellular protein that interacts with tip of herpesvirus saimiri and induces lymphocyte aggregation. *Immunity*, 6(5), pp.571–582.

Yüce, O. & West, S.C., 2013. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Molecular and Cellular Biology*, 33(2), pp.406–417.

Zeng, Y., Wagner, E.J. & Cullen, B.R., 2002. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Molecular Cell*, 9(6), pp.1327–1333.

Zenklusen, D. et al., 2001. The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Molecular and Cellular Biology*, 21(13), pp.4219–4232.

Zhou, Z. et al., 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature*, 407(6802), pp.401–405. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=11014198&retmode=ref&cmd=prlinks.

Zolotukhin, A.S. et al., 2001. U2AF participates in the binding of TAP (NXF1) to mRNA. *The Journal of biological chemistry*, 277(6), pp.3935–3942.