This is an author produced version of a paper published in *Biochemical Society Transactions*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/

**Published paper**

**Rrp47 and the function of the Sas10/C1D domain**

Phil Mitchell

The University of Sheffield, Department of Molecular Biology and Biotechnology, Firth Court, Western Bank, Sheffield S10 2TN, UK

email p.j.mitchell@shef.ac.uk

keywords: Rrp47, C1D, Sas10, Lcp5, neuroguidin, Rrp6
Abstract
The Sas10/C1D domain is found in a small group of eukaryotic proteins that have functions in RNA processing events, translational control and DNA repair mechanisms. The domain is predicted to be helical in nature and comprises approximately 80 amino acid residues. While the Sas10/C1D domain has yet to be functionally characterised, available data suggest that this domain forms a binding surface for specific interactions with other proteins and can concomitantly interact with RNA or DNA. This property of the Sas10/C1D domain may facilitate this family of proteins to dock other proteins onto nucleic acid substrates.

Introduction
The budding yeast (Saccharomyces cerevisiae) protein Rrp47 (also known as Lrp1 or yC1D) and its mammalian homologue C1D are nuclear proteins that bind both DNA and RNA [1-3]. Most work on the yeast protein is based on its physical association with the exosome nuclease complex [4-6] and the role of this complex in RNA processing and degradation pathways (for recent reviews of exosome structure and function, see [7-11]), while C1D was initially characterised as a DNA repair factor. Nevertheless, there is strong evidence of functional conservation of Rrp47/C1D; both proteins interact directly with the catalytic exosome component Rrp6 (known as PM-Sc100 in humans) [2,3] and similar defects in 5.8S rRNA maturation are observed in yeast rrp47-∆ or rrp6-∆ mutants and upon depletion of C1D in mammalian cells [3,5,12], while yeast rrp47-∆ mutants show defects in DNA repair [13,14]. Bioinformatical analyses [15] have identified a domain within Rrp47/C1D, known as the Sas10/C1D domain, that is also found in the U3 small nucleolar RNA (snoRNA)-associated proteins Sas10 (also known as Utp3) and Lcp5, as well as the metazoan protein neuroguidin. Sas10 and Lcp5 are required for the synthesis of 18S rRNA [16,17], while neuroguidin is an eIF4E-binding protein that is required for cytoplasmic polyadenylation-dependent translational control in neuronal cells [18]. Here, I review the data available for the function of members of this protein family and discuss data pertaining to the molecular function of the Sas10/C1D domain.

Structure and expression of Rrp47/C1D
Rrp47 and C1D are small, basic proteins of 21kDa and 16kDa, respectively. There is no detailed structural information currently available but sequence homology and secondary structure prediction
programmes suggest that the N-terminal ~120 residues of Rrp47 comprise a conserved -helical domain, while the C-terminal region is less structured and more variable. The N-terminal domain includes residues 10-89 of Rrp47 and shows homology with other members of the Sas10/C1D family (Fig. 1). The C-termini of Rrp47 homologues are typically very rich in basic residues, particularly lysine. C1D can be phosphorylated in vitro by the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) and is associated with this protein in cell lysates [19]. However, the importance of C1D phosphorylation to its function in vivo remains unclear. Similarly, the yeast protein can also be phosphorylated by the protein kinases Atg1 and Yck1 in vitro [20] but is not observed as a phosphoprotein in yeast cell extracts, either in mass spectrometric analyses of exosome complexes [21] or by in vivo orthophosphate labelling experiments (our unpublished data). Recombinant Rrp47 is expressed in Escherichia coli as a multimeric complex [2]. Consistent with roles in RNA surveillance and DNA repair, both yeast and mammalian proteins are localised to the cell nucleus [1,14,22].

**Protein interactions involving Rrp47/C1D**

Rrp47 is associated with Rrp6-containing exosome complexes [5,6] and interacts directly with Rrp6 in vitro [2]. The interaction occurs between the N-terminal region of Rrp47 that encompasses the Sas10/C1D domain and the N-terminal PMC2NT domain of Rrp6 [2](our unpublished data). Rrp47 levels are severely depleted in the absence of Rrp6 and are normalised upon expression of the PMC2NT domain alone [2], suggesting that Rrp47 stability is dependent upon its interaction with Rrp6. Thus, the depletion of Rrp47 should be taken into account when interpreting data from experiments using *rrp6*-Δ alleles. An interaction is also observed between the mammalian proteins C1D and PMScl-100 [3].

C1D has been identified through two hybrid screens as an interactor of the nuclear hormone receptor RevErb [23], DNA-dependent protein kinase [19] and the translin-associated factor X [24], while the homologous protein from *Schizosaccharomyces pombe*, Cti1, was shown to interact with condensin [25]. Analyses of proteins copurifying with TAP-tagged Rrp47 from yeast cell lysates solely identified known exosome components [6], suggesting that the exosome is the only protein complex with which yeast Rrp47 is stably associated. However, this does not eliminate the possibility of additional, transient interactions or complexes involving lower abundance species. Indeed, protein capture experiments using recombinant Rrp47 and yeast cell lysates, in association with analysis of RNA processing phenotypes of *rrp47* mutants, demonstrate functionally relevant interactions between Rrp47 and the snoRNP proteins Nop56 and Nop58 (our unpublished data).
**Rrp47/C1D function as an RNA-binding protein**

Rrp47 binds structured RNA such as poly(A)/(U) or tRNA in vitro but does not bind poly(A) [2]. Similarly, C1D interacts with tRNA and poly(G), but does not bind poly(A), poly(U) or poly(C) [3]. Rrp47 and C1D lack functionally characterised domains or motifs that are known to have RNA binding activity [26,27] and so the molecular basis of this interaction is of particular interest. The highly conserved basic nature of the C-terminus probably makes key interactions in RNA binding [2,28] but would be unable in itself to mediate specific binding to structured RNA. This suggests that RNA binding involves contacts with both the Sas10/C1D domain and the basic C-terminus. It has been shown that Rrp47 can bind both RNA and the PMC2NT domain of Rrp6 concomitantly, demonstrating that the binding surfaces for these ligands do not physically overlap [2].

What is the molecular function of Rrp47 in RNA processing and RNA surveillance? Yeast strains lacking Rrp47 or Rrp6 accumulate cellular RNAs that, in wild-type cells, are normally processed efficiently to mature RNAs or rapidly degraded. These include 3’ extended, incompletely processed precursors to 5.8S rRNA, snoRNAs and small nuclear RNAs (snRNAs), excised RNA fragments generated during RNA processing reactions such as the pre-rRNA 5’ external transcribed spacer (5’ ETS) region, and cryptic unstable transcripts (CUTs) [5,6,29-31]. Rrp47 is not required for the expression of Rrp6 or the binding of Rrp6 with exosome complexes [5]. The patterns of stable RNA precursors observed in *rrp47-Δ* and *rrp6-Δ* strains by northern blot analyses are largely indistinguishable [5] while distinct effects are seen in these mutants upon analysis of global mRNA profiles [14,32], suggesting that Rrp47 functions together with Rrp6 on some, but not all, of its substrates.

Purified recombinant Rrp6 degrades RNA in a distributive manner and degrades stem loop structures slowly [33,34], whilst Rrp47 binds specifically to structured RNA [2]. A plausible mechanism is that Rrp47 promotes the progression of Rrp6 through structured RNA elements (Fig. 2A). By allowing Rrp6 to interact indirectly with structured RNAs, Rrp47 might increase the retention time of substrates on Rrp6. Alternatively, through interacting with the adjacent PMC2NT domain, Rrp47 might position 3’ single-stranded extensions of structured RNA elements close to the catalytic centre of Rrp6. Such a function has been proposed for the C'-terminal HRDC domain of RNase D [35]. Alternatively, but not mutually exclusive with the possible mechanisms noted above, interactions between Rrp47 and other
RNA processing factors may contribute to substrate recognition or processing (Fig. 2B).

**Rrp47/C1D as a DNA-binding protein**

C1D was isolated as a protein with “exceptional DNA affinity”, since it remained bound to DNA under denaturing conditions (alkaline extraction and phenol extraction) [1], and Rrp47 and C1D both bind to circular plasmid or linear double-stranded DNA in vitro [1,2]. The dissociation constant of yeast Rrp47 for DNA is comparable to that of RNA, both being approximately 1µM (calculating the protein concentration based on the predicted molecular weight of the monomeric protein) [2]. RNA is an effective competitor in a DNA binding assay [2], strongly suggesting that the ligands occupy the same binding site on Rrp47. The expression level of Rrp47 is comparable to that of Rrp6 [2], which is reported to be approximately 2,000 molecules per yeast cell [36]. Given the volume of a yeast cell nucleus is approximately $3\mu m^3$ [37], the intracellular concentration of Rrp47 is close to its dissociation constant. Changes in the effective concentration of Rrp47 would therefore have a large impact on its ability to bind substrates.

Rrp47/C1D is well annotated as a DNA repair factor. A key observation is that C1D physically interacts with the catalytic subunit of the major sensor of DNA double strand breaks, DNA-PK, and stimulates its protein kinase activity [19]. DNA-PK binds to the ends of double-stranded DNA and promotes their joining through the nonhomologous end-joining pathway [38]. DNA-PK is thought to trigger a kinase signalling pathway that allows the recruitment of DNA repair factors to the double-strand break (DSB), and C1D is very efficiently phosphorylated by DNA-PK. Yeast rrp47-Δ mutants were subsequently shown to be defective in plasmid-based assays for DNA repair [13]. However, the details of the proposed DNA-PK-dependent signalling pathway are yet to be determined and the role of C1D in DSB repair remains unclear. Moreover, recombinant Rrp47 and C1D bind relaxed and supercoiled, circularised plasmid DNA, just as well as linearised DNA [1,2], and C1D does not require DNA termini to bind to DNA-PKcs or to stimulate its kinase activity [19]. Yeast strains lacking Rrp47 or Rrp6 were reported to be defective in the degradation of specific mRNA transcripts upon exposure to UV irradiation [14]. The rrp6-Δ and rrp47-Δ mutants are defective in the repair of UV irradiation-induced cyclobutane dimers, and rrp6-Δ mutants are synergistically sick in combination with rad26-Δ mutants [14]. One possibility is that the sensitivity of rrp6-Δ and rrp47-Δ mutants to UV irradiation may reflect a quality control function of the exosome to eliminate aberrant transcripts that arise as a result of DNA
damage in transcription-coupled DNA repair [39].

**Other members of the Sas10/C1D family**
The Sas10/C1D domain is a bioinformatically defined protein domain (Pfam domain PF04000)[15] of approximately 80 amino acid residues that is found in organisms throughout the eukaryotic kingdom (Fig. 1). Almost all of the members of this family contain one or two Sas10/C1D domains and most members lack any other functionally characterised domain. This protein family comprises Sas10, Lcp5 and neuroguidin, in addition to Rrp47/C1D.

Sas10 was initially characterised as a yeast protein that, when overexpressed, partially suppressed silencing of transcriptionally repressed chromatin [40]. Genetic analyses revealed that this suppression functions independently of the Sir protein silencing machinery. Mutants of the exosome complex and its associated cofactor TRAMP also show increased expression of genes encoded by repressed chromatin in a Sir-independent manner [41,42]. One explanation is that overexpression of Sas10 may (partially) titrate out the complex between Rrp47 and Rrp6. Sas10 has been shown to be a component of the U3 snoRNA-containing “small subunit processome” that is required for processing of 18S rRNA from the 35S pre-rRNA [17]. Lcp5 is also required for 18S rRNA processing [16] and is a component of the same 90S preribosomes [43-45].

Neuroguidin, like maskin, binds to both the translation initiation factor eIF4E and the cytoplasmic polyadenylation element (CPE)-binding protein CPEB, and inhibits translation of specific mRNAs in a CPE-dependent manner [18]. Mutational analyses of sequence motifs within neuroguidin that are required for eIF4E binding demonstrate that the Sas10/C1D homology region is not responsible for interaction with eIF4E. Whether neuroguidin binds CPEB via the Sas10/C1D domain is not clear. By analogy with maskin, translation activation is coupled to polyadenylation by Gld2 [46]. Gld2 is a noncanonical poly(A) polymerase related to the TRAMP components Trf4 and Trf5 that promote RNA degradation by the exosome. Thus, both Rrp47 and neuroguidin function in concert with related poly(A) polymerases via bridging proteins (Rrp6 or CPEB). Rrp6, like CPEB, is phosphorylated [47] but the functional relevance of this event is unclear.

Sas10, Lcp5 and neuroguidin all function with RNA but whether these proteins interact directly with RNA requires further direct experimentation. Notably, Sas10, Lcp5 and neuroguidin also have basic
regions at the C-terminus in addition to the Sas10/C1D domain that may contribute to RNA binding.

**Conclusions**

The Sas10/C1D domain is found in eukaryotic proteins that function in aspects of RNA processing, RNA quality control and translational control. At least in the case of Rrp47, this domain provides contacts in both protein-protein interactions and in the binding of nucleic acids that are key to protein function. Future research will address whether the Sas10/C1D domain represents a general dual functional domain for protein- and RNA interactions, and discern the molecular nature of these interactions and their potential regulation through phosphorylation events.

**Funding**

This work was funded by a research grant from the Wellcome Trust [088306/Z/09/Z].
References


and molecular activity. Biochim Biophys Acta 1779, 558-565


12 Briggs, M., Burkard, K. and Butler, J. (1998) Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. J Biol Chem 273, 13255-13263


42 Houseley, J., Kotovic, K., El Hage, A. and Tollervey, D. (2007) Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. EMBO J 26, 4996-5006


Figures

Figure 1. Alignment of the Sas10/C1D domain.
Sequences were taken from the full alignment for Pfam family PF04000 (http://pfam.sanger.ac.uk). Rrp47 is synonymous with Lrp1. Residues that showed conservation within the full alignment were colour-coded by hand, using the following key: glycine, brown; proline, yellow; small or hydrophobic residues (A,V,L,I,M,F,W), turquoise; charged amino acids (D,E,R,K), dark green; histidine or tyrosine, bright green.

Figure 2. Roles of Rrp47 in Rrp6-mediated RNA degradation.
(A) Rrp6 is able to degrade nonstructured RNA efficiently but progresses slowly through structural elements. Rrp47 is able to interact with structured elements within RNA whilst remaining bound to Rrp6. Rrp47 may enable Rrp6 to progress more readily through structural elements by retaining the enzyme on the RNA and/or by directly presenting the free 3’ end of the RNA to the catalytic domain of Rrp6. (B) Protein-protein interactions between Rrp47 and other processing factors (labelled “X”) allow targeting of Rrp6 to the 3’ end of its substrates or to recognise assembled ribonucleoprotein (RNP) complexes.