Nucleic Acids Research, 2022 **1**
https://doi.org/10.1093/nar/gkac1199

Trypanosoma brucei RRP44: a versatile enzyme for processing structured and non-structured RNA substrates

Giovanna Cesaro[1,2], Heloisa Tramontin da Soler[1], Eloise Pavão Guerra-Slompo[1], Ahmed Haouz[3], Pierre Legrand[4], Nilson Ivo Tonin Zanchin[1] and Beatriz Gomes Guimaraes[1,2,*]

[1]Carlos Chagas Institute, Oswaldo Cruz Foundation, FIOCRUZ, Curitiba-PR, Brazil, [2]Biochemistry Postgraduate Program, Federal University of Paraná, Curitiba-PR, Brazil, [3]Institut Pasteur, Plate-forme de cristallographie-C2RT, UMR-3528 CNRS, Paris, France and [4]Synchrotron SOLEIL, Saint-Aubin, France

Received February 14, 2022; Revised November 25, 2022; Editorial Decision November 29, 2022; Accepted December 21, 2022

ABSTRACT

Rrp44/Dis3 is a conserved eukaryotic ribonuclease that acts on processing and degradation of nearly all types of RNA. It contains an endo- (PIN) and an exonucleolytic (RNB) domain and, its depletion in model organisms supports its essential function for cell viability. In *Trypanosoma brucei*, depletion of Rrp44 (TbRRP44) blocks maturation of ribosomal RNA, leading to disruption of ribosome synthesis and inhibition of cell proliferation. We have determined the crystal structure of the exoribonucleolytic module of TbRRP44 in an active conformation, revealing novel details of the catalytic mechanism of the RNB domain. For the first time, the position of the second magnesium involved in the two-metal-ion mechanism was determined for a member of the RNase II family. *In vitro,* TbRRP44 acts preferentially on non-structured uridine-rich RNA substrates. However, we demonstrated for the first time that both TbRRP44 and its homologue from *Saccharomyces cerevisiae* can also degrade structured substrates without 3'-end overhang, suggesting that Rrp44/Dis3 ribonucleases may be involved in degradation of a wider panel of RNA than has been assumed. Interestingly, deletion of TbRRP44 PIN domain impairs RNA binding to different extents, depending on the type of substrate.

INTRODUCTION

The protozoan parasites of the trypanosomatid group include *Trypanosoma brucei*, *Trypanosoma cruzi* and several species of the genus *Leishmania* that cause severe diseases in humans and animals. They belong to the order Kinetoplastida, which exhibits a number of unique molecular features, especially at the genomic organization and RNA processing levels. For instance, the genome is characterized by the absence of RNA polymerase II transcription factors and the protein-coding genes are organized in polycistronic units, which are constitutively transcribed. The polycistronic pre-mRNAs are co-transcriptionally processed by *trans*-splicing and polyadenylation to generate the mature mRNAs, and control of gene expression is predominantly post-transcriptional, through regulation of mRNA processing, translation and decay (reviewed in (1)). In addition, the ribosomal RNA (rRNA) structure and maturation also presents unique and remarkable aspects in trypanosomatids: the rRNA is very rich in methylation and pseudouridylation sites, the large-subunit rRNA of cytoplasmic ribosomes equivalent to the 25S/28S rRNA of other eukaryotes is divided into six fragments and several small nucleolar RNAs mediate cleavages to remove the transcribed spacers during processing of the rRNA precursor (reviewed in (2)). Along with these unique features, one can anticipate distinctive functions for ribonucleases involved in RNA processing and degradation in trypanosomatids.

The exosome is the major eukaryotic 3′-5′ exoribonuclease. It is an essential multi-protein complex acting on maturation, degradation, and quality control of nearly all types of RNA. The catalytic activity of the exosome relies on the distributive exoribonuclease Rrp6/PM-Scl100 and the processive exo- and endoribonuclease Rrp44/Dis3 (reviewed in (3–5)). Rrp44/Dis3 is a conserved multi-domain protein composed by a N-terminal CR3 motif and endoribonuclease domain (PIN domain), two cold shock domains (CSD), an exoribonuclease RNB domain and a C-terminal RNA binding domain S1 (reviewed in (6,7)). The exoribonuclease module, containing the two CSDs, the RNB and the S1 domains, is related to the bacterial RNaseII/R 3′-5′

*To whom correspondence should be addressed. Tel: +55 41 33163225; Email: beatriz.guimaraes@fiocruz.br

exoribonucleases. Two additional homologues are found in humans: DIS3L (*Dis3-Like*) which has a catalytically inactive PIN domain and is associated to the cytoplasmic exosome (8,9) and DIS3L2, which lacks the PIN domain and is a exosome-independent cytoplasmic 3′-5′ exonuclease (10,11). Dis3L2 was also identified in *Schizosaccharomyces pombe* as an essential protein involved in mRNA degradation (12). The function of Rrp44/Dis3 has been extensively studied in *Saccharomyces cerevisiae*. Inactivation of its exoribonuclease catalytic site by a single point mutation (D551N) blocks the yeast exosome activity *in vitro* and *in vivo,* evidenced by the accumulation of partially processed 5.8S rRNA and partially degraded 5′ ETS rRNA (13). The role of the ScRrp44 endoribonuclease activity was also demonstrated, indicating cooperation between endo- and exoribonucleolytic exosome activities (14–16). In addition, the N-terminal CR3 motif and PIN domain were shown to be necessary for ScRrp44 association with the exosome core and complex activity (16,17). The human Dis3 (hDIS3) was also shown to have both endo- and exoribonuclease activities, associated with the nuclear exosome (8).

Studies on the *T. brucei* exosome complex showed the conservation of the subunits, which compose the so-called Exo-9 core: the six RNase PH-related proteins forming a heterohexameric ring and the three S1 domain proteins. The 3′-5′ exoribonuclease RRP6 was identified in association with *T. brucei* exosome. In contrast, the RRP44 subunit did not copurify with the complex (18,19). Similarly, RRP44 was not found in native exosomes purified from *Leishmania tarentolae* (20). Although it is not present in the exosome complex, *T. brucei* RRP44 (TbRRP44) was shown to be essential for cell survival (18,19). RNAi-mediated depletion of TbRRP44 impairs maturation of the 5.8S rRNA (18,21) and causes accumulation of the complete LSU rRNA precursor (21).

Structural and biochemical characterization of Rrp44/Dis3 and of yeast and human exosome complexes have provided extensive information about the association of Rrp44/Dis3 with the exosome core, the RNA interaction mechanisms, substrate selectivity and details of Rrp44/Dis3 activity (22–27). However, the lack of RRP44 interaction with the exosome in trypanosomatids may have implications on RRP44 activity in these organisms, regarding substrate selection and mechanism of catalysis. In a previous work, we have determined the crystal structure of TbRRP44 PIN domain and suggested a molecular basis for absence of TbRRP44 interaction with the exosome. We observed that structural differences in TbRRP44 N-terminal and PIN domain and differences in the *T. brucei* homologues of *S. cerevisiae* Rrp41 and Rrp42, which form the intermolecular contacts between the core exosome and Rrp44, may explain the lack of this association in trypanosomatids (21). In order to obtain further information on TbRPP44 structure and function, we have determined the crystal structure of its exoribonucleolytic module including the two CSDs, the RNB and the S1 domains and performed *in vitro* degradation and interaction assays with several different RNA substrates. The results obtained provide novel details on the RNB exoribonuclease catalytic site and on TbRRP44 endo- and exoribonucleolytic activities on structured and non-structured substrates.

MATERIALS AND METHODS

Cloning

The nucleotide sequences encoding the full-length TbRRP44 (GenBank: AJ308998.1) and a truncated construct comprising residues 232–972 (TbRRP44-ΔNPIN) were synthesized by GeneCust (Boynes, France) with codon optimization for *Escherichia coli* expression and cloned into restriction sites NcoI and XhoI of vector pET28a (Novagen). A TbRRP44 synthetic gene containing G-A mutations at nucleotide positions 417 and 1531 was acquired from Biomatik (Ontario, Canada) cloned into the expression vector pET28a. These mutations replace aspartic acid codons by asparagine codons, creating an endo- and exonuclease catalytic site double mutant (TbRRP44$^{D140N/D511N}$). Otherwise, the nucleotide sequence of this double mutant is identical to the synthetic gene encoding the wild-type TbRRP44. An endonuclease catalytic site single mutant (TbRRP44^{D140N}) was constructed by fusing the NcoI-PstI DNA fragment of the double mutant synthetic gene with the PstI-XhoI fragment of the synthetic gene encoding the wild-type TbRRP44. Conversely, a TbRRP44 exonuclease catalytic site single mutant (TbRRP44^{D511N}) was constructed by fusing the NcoI-PstI DNA fragment of the of the synthetic gene encoding the wild-type TbRRP44 with the PstI-XhoI fragment of the synthetic gene encoding the double mutant. Both single mutants for the endonucleolytic and exonucleolytic sites were constructed in the plasmid pET28a.

The TbRRP44-ΔNPIN^{D511N} variant was constructed by using a PCR fusion strategy using the TbRRP44-ΔNPIN synthetic gene cloned in pET28a as template. For this purpose, we obtained initially a pair of complementary primers bearing a single point mutation replacing aspartic acid 511 by an asparagine (primer F: 5′ GGGCTGTC GTGATATTAATGACGCTCTGCATTGC3′, primer R: 5′ GCAATGCAGAGCGTCATTAATATCACGACAGC CC3′). Subsequently, two PCR products were generated, one using primer R and a primer that anneals at the T7 promoter of plasmid pET28a and, the other using primer F and a primer that anneals at the T7 terminator pET28a. The PCR reactions were performed with the Platinum SuperFi DNA Polymerase (Invitrogen, cat. 12351–010), according to manufacturer's instructions. Each individual PCR product was purified from agarose gels using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany). To reconstitute ΔNPIN/D511N coding sequence, 150 ng of each PCR product were used in a two-step fusion PCR reaction. The first step consisted of 9 cycles of denaturing, annealing and extension by the Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific, cat. F5305) and respective PCR reagents without the amplification primers, followed by addition of the external primers (T7 promoter and T7 terminator) and 30 additional cycles of amplification. The TbRRP44-ΔNPIN^{D511N} full amplicon was purified from agarose gel using the NucleoSpin Gel and PCR Clean-up kit cited above and digested with the

restriction enzymes NcoI and XhoI (Anza system, ThermoFisher Scientific). The pET28a vector was digested in parallel with the same enzymes. Both plasmid and insert DNAs were purified and enzymatically ligated by the T4 DNA Ligase (Invitrogen, cat. 15224–041), according to the suppliers' instructions, and transformed in *E. coli* DH5α for clone selection. The presence of the gene of interest was confirmed in the selected clone by PCR and enzyme digestion, and the single point mutation was confirmed by DNA sequencing with the primer R44seq-R (5′GAGTCGAAGG TCTGGTACGC3′) and the T7 promoter primer.

Vector pGEX-4T1 (former GE Healthcare Life Sciences, current Cytiva Life Sciences, Marlborough, USA) containing the nucleotide sequence encoding *Saccharomyces cerevisiae* Rrp44 (ScRrp44) was kindly provided by Carla C. Oliveira (University of São Paulo). The ScRrp44 gene was transferred to vector pET28a (Novagen) by direct subcloning using the EcoRI and XhoI restriction sites.

Protein expression and purification

E. coli BL21 Star (DE3) cells were transformed with the expression vectors and incubated at 37°C in LB medium containing the selection antibiotic. When the cultures reached $OD_{600\,nm}$ ∼0.6, the temperature was reduced to 18°C and protein expression was induced with 0.2 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) for approximately 16 hours. Cells from 1 L culture were harvested by centrifugation at 6000 × g for 10 min, suspended and lysed in 30 ml of buffer A (50 mM Tris–HCl pH 8, 300 mM NaCl, 20 mM imidazole) using a microfluidizer (Microfluidics™ 110L series 0300, Newton, MA, USA). The soluble fraction was separated by centrifugation at 40 000 × g for 25 min at 4°C.

Protein purifications were performed using ÄKTA Pure M25 chromatography systems (former GE Healthcare Life Sciences, current Cytiva Life Sciences, Marlborough, USA). The full-length wild-type TbRRP44, double-mutant and single-point mutants, as well as ScRrp44 were purified following the same protocol. The extracts were loaded onto a 5 ml His-trap HP column (Cytiva, Marlborough, USA) equilibrated with buffer A. The proteins were eluted with 10 column volumes (CV) of a linear gradient from 0 to 100% of buffer B (50 mM Tris–HCl pH 8.0, 300 mM NaCl, 800 mM imidazole). TbRRP44-ΔNPIN and single-point mutant TbRRP44-ΔNPIN^{D511N} were purified using a two-step gradient, a first linear gradient from 0 to 5% of buffer B in 4 CV, followed by linear gradients from 5 to 15% of buffer B in 3 CV. The fractions containing the target protein were pooled and diluted 3 times in 50 mM Tris–HCl pH 8.0 before loading onto a 5 ml Hi-Trap HP Q (Cytiva Life Sciences, Marlborough, USA) equilibrated with 50 mM Tris–HCl pH 8.0. The full-length TbRRP44, double-mutant and single-point mutants and ScRrp44 were eluted with a 20 CV linear gradient from 0 to 100% of buffer B (50 mM Tris–HCl pH 8, 1 M NaCl). TbRRP44-ΔNPIN and TbRRP44-ΔNPIN^{D511N} were eluted in a 10 CV linear gradient from 0 to 45% of buffer B. Fractions containing the target proteins were pooled, concentrated to a final volume of 1 ml, and loaded onto a Superdex 200 16/600 (Cytiva Life Sciences, Marlborough, USA) equilibrated with 50 mM Tris–HCl pH 8.0, 150 mM NaCl). The purified proteins were concentrated to ∼6 mg/ml for crystallization assays or to 0.5 mg/ml and stored at –80°C for the RNA degradation assays.

Crystallization, data collection and processing

Full-length TbRRP44 and TbRRP44-ΔNPIN were submitted to crystallization trials by sitting drop vapor-diffusion method using the commercial screens Crystal Screen 1 and 2, Salt RX and PEG/Ion 1 and 2 (Hampton Research, Aliso Viejo, USA), Wizard 1 and 2 (Emerald Biosystems, Bainbridge Island, USA), Structure screen 1 e 2 (Molecular Dimensions, Sheffield, United Kingdom) and JBScreen Classic 1–8 (Jena Bioscience, Jena, Germany). The crystallization plates were prepared using a Mosquito system (STP Labtech, Melbourn, United Kingdom).

Promising conditions were obtained for TbRRP44-ΔNPIN and optimization was performed manually by varying precipitant, buffer, and protein concentrations. Best crystals were obtained by hanging drop vapor diffusion (total drop volume 2 to 3 μl) in two different conditions, 15% (w/v) PEG 20000, 0.1 M MES pH 6.0 and 12% (w/v) PEG 8000, 0.1 M NaK phosphate pH 5.5, 0.2 M NaCl. These conditions were used for co-crystallization experiments with two oligoribonucleotides (5′ CCAAAAAAAACC 3′ and 5′ GGUUUUUUUUUGG 3′, purchased from Eurogentec (Seraing, Belgium)). Before incubation with the substrates, the protein was incubated with 2 mM EDTA for 30 min. The double mutant TbRRP44$^{D140N/D511N}$ was submitted to crystallization trials after incubation with the oligoribonucleotides described above, by sitting drop vapor-diffusion method using the commercial screens Index HT and PEG/Ion (Hampton Research, Aliso Viejo, USA) and JCSG plus (Molecular Dimensions, Sheffield, United Kingdom). However, no promising hits were obtained from these trials.

TbRRP44-ΔNPIN crystals grown in 15% (w/v) PEG 20000, 0.1 M MES pH 6.0 were cryo-protected with 20% (v/v) glycerol added to the mother liquor prior to flash cooling in liquid nitrogen. X-ray diffraction data were collected at the PROXIMA-1 beam line at Synchrotron SOLEIL using an EIGER 16M detector (DECTRIS, Baden-Daettwil, Switzerland). Data sets collected from two regions of the same crystal were merged during data processing. Diffraction data were processed with XDS (28) using the autoPROC package (29) and anisotropic correction was performed using STARANISO (30).

Structure determination and refinement

The TbRRP44-ΔNPIN structure was determined by molecular replacement with MOLREP (31) using a model generated by the Phyre2 server (32) as the search model. Structure refinement was performed alternating cycles of BUSTER (33) with visual inspection and interactive rebuilding using COOT (34). Model validation was performed using MolProbity (35). Protein–RNA interactions were analysed using the Protein–Ligand Interaction Profiler (PLIP) server (36). The molecular images were generated using Pymol (The PyMOL Molecular Graphics System, Schrödinger, LCC).

RNA degradation assays

5′-End fluorescein-labeled (FAM) RNA oligonucleotides were purchased from Eurogentec (Seraing, Belgium). The polyA, polyU, polyG and polyC substrates are homo 15-mer oligoribonucleotides. In addition, we used a single-stranded AU-rich substrate (polyAU), a substrate designed to form a GNRA stem–loop with an 18-mer AU-rich 3′-end overhang (GNRA18) and a substrate containing a GNRA stem–loop without 3′-end overhang (GNRA0). GNRA tetraloops (N = any nucleotide, R = purine) were included in the structured substrates in order to increase the hairpin stability (37). Both single-stranded polyAU and stem–loop containing substrates are 49-nt long (sequences below, GNRA tetraloops are underlined).

polyAU: 5′FAM-AAUUAUUUAUUAUUUAUUUAUUAUUUAUUUAUUAUUUAUUAUUUAUUUAUUA 3′.

GNRA0: 5′FAM-AUUUAUUAUUAUUUAUUUAUUUAUUAGGGCGGGCGCAAGCCCGCCC 3′.

GNRA18: 5′FAM-AUUUAUUUAGGGCGGGCGCAAGCCCGCCCUAUUUAUUAUUUAUUAUUUA 3′.

All RNA substrates were incubated at 85°C for 4 min and cooled down to room temperature before each assay. The electrophoretic mobility of the RNA substrates was analysed under non denaturing conditions (Supplementary Figure S3), to evaluate the substrates' structure. For single time-point analysis, degradation assays were performed in 15 μl reaction mixtures containing 10 mM Tris–HCl pH 8.0, 75 mM NaCl, 1 mM β-mercaptoethanol, with $MgCl_2$ or $MnCl_2$ ranging from 0 to 4 mM, 0.1 μM RNA substrate and 0.2 μM of protein. Reactions were incubated at 37°C for 30 or 60 min and stopped by addition of the same volume of formamide loading buffer (95% formamide, 5 mM EDTA) and heating at 95°C for 10 min. Time course assays were performed at the same conditions described above, except that 0.1 mM of both $MgCl_2$ and $MnCl_2$ were added in the reaction mixtures. Reaction times are indicated in the figures. Samples were loaded onto a 20% polyacrylamide–8 M urea gel in TBE running buffer (130 mM Tris base pH 8.0, 45 mM boric acid, 2.5 mM EDTA). Electrophoresis was performed in a PROTEAN II xi 20 cm Cell (Bio-Rad Laboratories Inc., Hercules, California, USA). A mixture of xyleno cyanol and bromophenol blue was loaded in a separate lane as a gel migration marker. The gels were visualized on an iBright FL1500 Imaging System (Thermo Fisher Scientific, Waltham, MA, USA).

Quantification of the substrate disappearance was performed using the Gel Analyzer function of the ImageJ software (38). The integrated intensity of individual bands from each reaction time were normalized against the negative control corresponding to the reaction mixture in the absence of protein. In the case of the quantification of product formation (Figure 6C), the same integration area used to estimate the intensity of the product at each reaction time was used to determine the background value for the negative control. This value was subtracted from all individual estimates; therefore, product formation is represented in arbitrary units, considering the negative control as zero. At least three replicas of the degradation assays were used to calculate the average intensity of the bands and standard deviation values. Statistical significance was determined using one-way ANOVA followed by Tukey's multiple comparisons test. Graphs were generated using the GraphPad Prism version 9.0.0 software (GraphPad Software, San Diego, California, USA, www.graphpad.com).

Electrophoretic mobility shift assay

The 5′-end fluorescein-labeled (FAM) substrates polyAU, GNRA0 and GNRA18 (described above) were used in electrophoretic mobility shift assays together with the catalytically inactive variants TbRRP44$^{D140N/D511N}$ and TbRRP44-ΔNPIN^{D511N}. The RNA substrates were incubated at 85°C for 4 min and cooled down to room temperature before the assays. The RNA was diluted in the EMSA buffer (30 mM Tris–HCl pH 8.0, 75 mM NaCl, 1 mM β-mercaptoethanol, 100 μM $MgCl_2$ and 6% (v/v) glycerol) to a final concentration of 0.1 μM. The protein was added to the final concentrations indicated in the figures, the mixture was incubated for 30 min at room temperature and loaded onto a 6% polyacrylamide gel in TBE running buffer. A bromophenol blue solution (2 mg/ml, 20% (v/v) glycerol) was loaded in a separate lane to serve as a migration marker. Electrophoresis was performed in a PROTEAN II xi 20 cm Cell. The gels were visualized on an iBright FL1500 Imaging System (Thermo Fisher Scientific, Waltham, MA, USA).

Analysis of the percentage of binding as a function of protein concentration was based on the intensities of the bands corresponding to the free substrate. The intensity of individual bands were estimated using the ImageJ software (38) and converted to percentage of binding considering the intensity of the band in the absence of protein as zero percent. At least three replicas were used to calculate the standard deviations. Data fitting were performed using the GraphPad Prism version 9.0.0 software (GraphPad Software, San Diego, California USA, www.graphpad.com), applying the Hill model equation.

RESULTS

Overall structure of TbRRP44-ΔNPIN

Our attempts to crystallize the full-length *T. brucei* RRP44 were unsuccessful. However, crystallization assays using a truncated construct (TbRRP44-ΔNPIN) comprising residues 232–972, devoid of the N-terminal and PIN domain, produced crystals suitable for X-ray diffraction.

The crystal structure of TbRRP44-ΔNPIN in complex with an RNA substrate was refined at 2.23 Å resolution to final R_{work}/R_{free} of 0.194/0.265 (Table 1). As expected, the overall structure of the exonucleolytic module of TbRRP44 matches the RNase II fold (39,40), organized into two cold-shock domains (CSD1 and CSD2, residues 232–365 and 367–447), the RNB catalytic domain (residues 448–869) and a C-terminal domain with a S1 RNA-binding fold (residues 871–945) (Figure 1A). The final model lacks residues 315–341 (in the connection between CSD1 β4 and β5 strands), 384–388, 926–932, and the C-terminal residues 946–972, after the S1 domain, which could not be modelled due to the absence of electron density, likely due to the flexibility of these regions.

Table 1. Diffraction data and structure refinement statistics

Data statistics	
Source	SOLEIL-PROXIMA1
Wavelength (Å)	0.9786
Space group	$P2_12_12_1$
Unit cell (Å)	$a = 66.30$, $b = 91.11$, $c = 148.30$
Resolution (Å)	77.63–2.23 (2.55–2.23)
Number of observations	593 016 (22 350)
Number of unique reflections	24 678 (1235)
Completeness	
spherical (%) *	54.9 (8.2)
ellipsoidal (%) *	93.8 (70.6)
Multiplicity	24.0 (18.1)
Mean $(I)/\sigma(I)$	12.1 (1.8)
$CC\ \frac{1}{2}$ (%)	99.8 (63.2)
Refinement statistics	
R_{work}	0.1941
R_{free}	0.2647
RMSD bond lengths (Å)/angles (°)	0.009/0.96
B values	
From Wilson plot (Å^2)	43.8
Mean *B* value (overall, Å^2)	49.6
Protein atoms	5331
Nucleic acid atoms	87
Heterogen atoms	23
Number of water molecules	253
Ramachandran plot	
Favored (%)	96.5
Outliers (%)	0.0

Values in parentheses are for the highest resolution shell. *Diffraction limits and principal axes of ellipsoid fitted to diffraction cut-off surface: $a^* = 2.91$; $b^* = 2.21$; $c^* = 2.96$. Criteria used in determination of diffraction limits: Local $I/\sigma(I) \geq 1.20$

Structural comparison with *S. cerevisiae* Rrp44-ΔNPIN (crystallized without the NPIN domain, PDB code 2vnu, (22)) and human Dis3 (extracted from the exosome complex structure, PDB code 6d6r, (41)) shows the general conservation of the domains (Figure 1B). Superposition of the RNB domains resulted in RMS deviations of 0.555 Å from the structure of ScRrp44 (323 carbon-alpha atoms aligned) and 0.905 Å compared with human Dis3 RNB (337 carbon-alpha atoms aligned). The interdomain orientation shows some variability, as already observed among Rrp44/Dis3 homologues, however, the structure of the individual domains is globally conserved. The exception in TbRRP44 structure is the S1 domain, which presents a different topology, with a shorter connection between strands β3 and β4 when compared with Rrp44/Dis3 homologues (Figure 1C and Supplementary Figure S1). Rrp44/Dis3 homologues usually possess a 3-strand β-sheet forming an 'extra' domain connecting β3 and β4 strands from the typical five-stranded antiparallel beta barrel. The crystal structures of Rrp44/Dis3 homologues in complex with RNA (PDB codes: 6d6r, 4pmw, 4ifd) show that this 'extra' β-sheet does not participate in RNA binding, thus we hypothesize that the shorter β3–β4 connection observed in TbRRP44 has no major impact on substrate interaction.

The TbRRP44-ΔNPIN structure displays the enzyme in action

The TbRRP44-ΔNPIN construct, the only amenable to crystallization, has an intact exonucleolytic active site, devoid of mutations. Despite treatment with EDTA to remove divalent ions essential for catalysis prior to incubation with the substrate for crystallization, the structure surprisingly revealed the protein in an active state, where five out of the originally fourteen-nucleotide long substrate (5′ GGUUUUUUUUUUGG 3′) were observed bound to the active site. In addition, during structure refinement, the electron density maps clearly showed that the 3′-end nucleotide was no longer linked to the RNA substrate, revealing an intermediate state of RNA processing, after the stage of 3′-end nucleotide cleavage but before product release (Figure 2). The substrate was initially modelled assuming an intact phosphodiester bond between the first and second 3′-end nucleotides, as found in homologue structures. However, the strong negative electron density at the position of the phosphate group prompted us to test different refinement models to better describe the reaction state captured in the crystal. In addition to the negative electron density at the position of the phosphate group, a strong positive density observed right beside indicated a structural rearrangement of the phosphate (Figure 2A). We then investigated whether there was a mixture of pre-and post-cleavage states by refining the occupancy of the 3′-end nucleotide in both conformations. The best refinement results (judged by R_{free} values and electron density maps) corresponded to the product (UMP) in a post-cleavage conformation, with the occupancy refined to 0.84. During the refinement, we also observed a positive electron density close to the catalytic D511 side chain. Attempts to fit a water molecule resulted in higher R_{free} and poor electron density. A second magnesium ion was successfully refined at this position (Figure 2B and Supplementary Figure S2). To our knowledge, this is the first structure of a RNase II family member showing two magnesium ions bound to the active site.

The mechanisms of RNA interaction and catalysis are conserved among the RNase II family members (22,39) (reviewed in 39). The bases of the five nucleotides, which occupy the RNA cleft pack together in a parallel stacking conformation and the phosphate groups are involved in hydrogen bonds and salt bridges with neighboring residues. The catalytic site is formed by four acidic residues that, together with two magnesium ions, perform the cleavage of the phosphodiester bond by the two-metal-ion mechanism. A multiple sequence alignment of Rrp44/Dis3 homologues and *E. coli* RNaseII and RNaseR is shown in Supplementary Figure S4 to facilitate comparison of the conserved residues. TbRRP44-ΔNPIN RNA binding and catalytic sites are shown in Figure 3. The RNA bases configuration is stabilized by hydrophobic interactions at the two extremities, with L656 (at the 5′-end) and the conserved Y555 (at the 3′-end). These residues are responsible for 'clamping' the substrate in the catalytic cavity and correspond to *E. coli* RNase II F358 and Y253, respectively. In the post-cleavage state revealed by the TbRRP44-ΔNPIN structure, the base moiety of the product keeps the stacking interaction with tyrosine Y555, while the phosphate group undergoes a structural rearrangement (Figures 2 and 3).

In addition to the four catalytic aspartic acids (D503, D509, D511, D512 in TbRRP44), a conserved arginine residue was shown to be essential for RNase II activity (R500), having a role in stabilizing the phosphodiester bond to be cleaved (39,42,43). The corresponding arginine

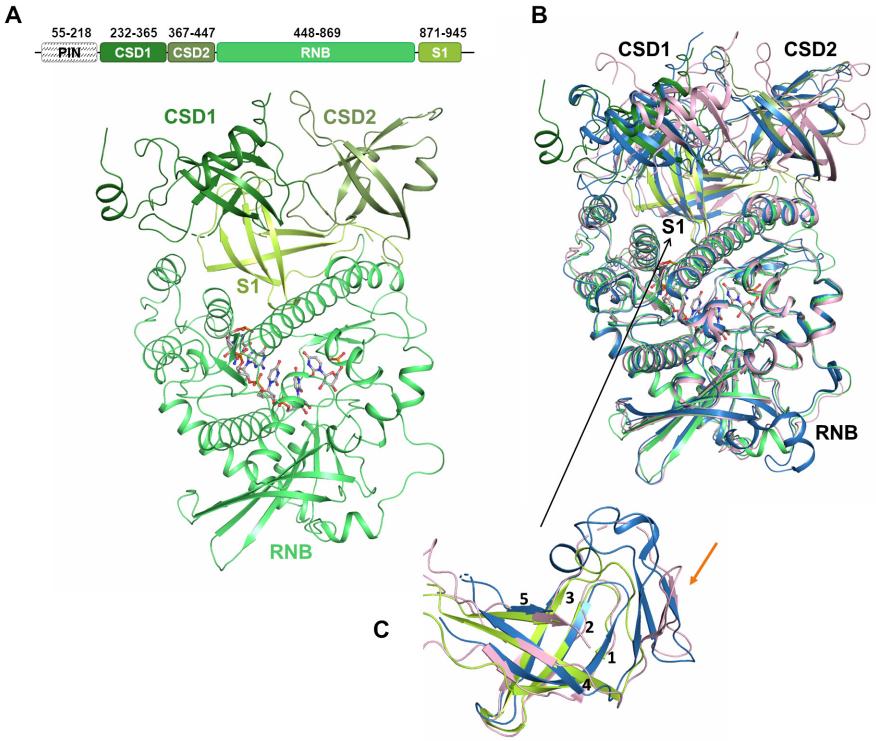

Figure 1. Overall structure of *T. brucei* RRP44-ΔNPIN. (**A**) Schematic representation of TbRRP44 structural domains (top), with the N-terminal domain, absent in the TbRRP44-ΔNPIN deletion mutant, labelled in black. The numbers on the top indicate the domain boundaries. The overall structure (bottom) is represented in cartoon with the RNA bound to the RNB domain shown in sticks. (**B**) Structural superposition based on the RNB domain of TbRRP44-ΔNPIN (green), *S. cerevisiae* Rrp44 (blue, PDB code 2vnu) and *H. sapiens* Dis3 (pink, extracted from the exosome complex structure, PDB code: 6d6r). (**C**) Detail of the S1 domain. A 3-strand β-sheet (indicated by an orange arrow), observed in ScRrp44 (blue) and HsDis3 (pink), is absent in TbRRP44. TbRRP44 has a shorter connection between strands β3 and β4 of the five-stranded antiparallel S1 beta barrel.

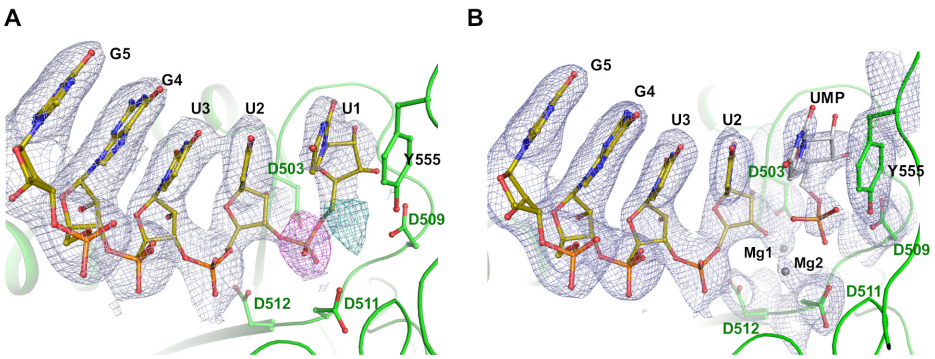

Figure 2. Structure refinement reveals the RNA cleavage by TbRRP44. (**A**) TbRRP44-ΔNPIN structure refined assuming the intact phosphodiester bond between the first and second 3′-end nucleotides (U1 and U2). For clarity, only the electron density maps corresponding to the RNA molecule is shown. The *2mFo-DFc* electron density map is colored blue and contoured at 1.0 sigma whereas the *mFo-DFc* map is contoured at 3.5 sigma and colored magenta (negative) and green (positive). The strong negative peak in the *mFo-DFc* map superimposed to the U1 phosphate group evidences the substrate cleavage. (**B**) Final model with the product (UMP) in a post-cleavage conformation. The *2mFo-DFc* electron density map is colored blue and contoured at 1.0 sigma. The catalytic aspartic acids are identified with green labels. The magnesium ions are represented by grey spheres. The labelled tyrosine residue (Y555) stabilizes the 3′-end nucleotide conformation through a stacking with the base moiety.

in TbRRP44 (R807) is likely to be flexible in TbRRP44-ΔNPIN structure since its sidechain presented weaker electron density than surrounding residues. Given the post-cleavage state captured by the TbRRP44-ΔNPIN structure, we speculate that R807 side chain undergoes a conformational change after the cleavage, which would be linked to the structural rearrangement of the product before its release. R807 would then move back to the previous conformation for another round of catalysis.

The position of one of the magnesium ions (labelled Mg1 in Figure 3) is conserved when compared with RNaseII/Dis3/Rrp44 structures, despite the absence of the coordinating waters described in previously reported structures. The second magnesium (labelled Mg2 in Figure 3),

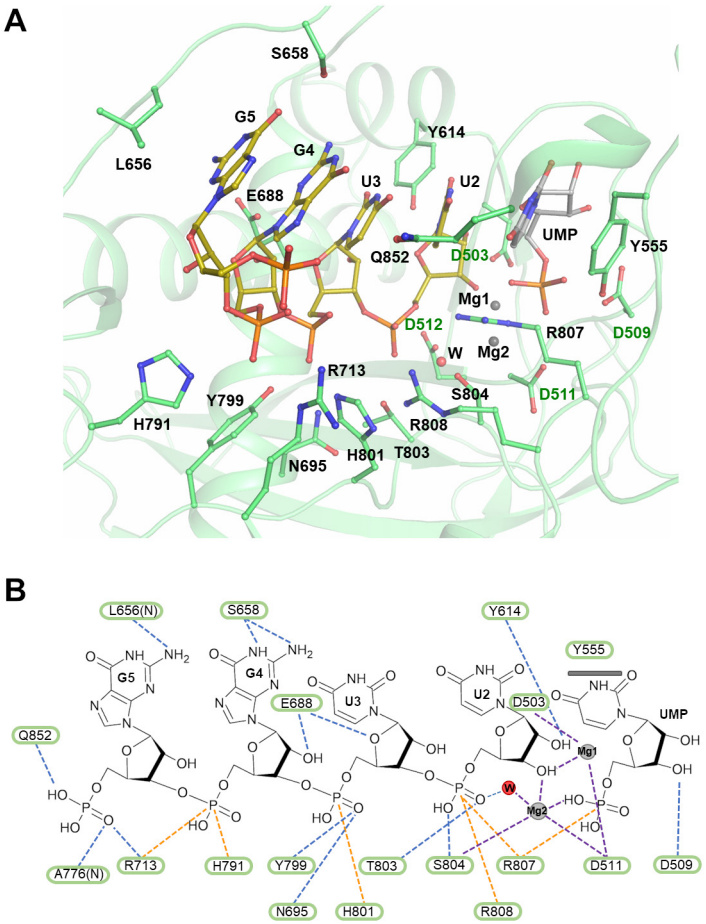

Figure 3. Detail of TbRRP44 exoribonucleolytic active site. (**A**) The ribonucleotides at positions 2 to 5 are shown in yellow and the bases are numbered. The product (UMP) in a post-cleavage conformation is shown in grey. Residues involved in RNA interaction are shown in sticks and labelled. The catalytic aspartic acids are identified with green labels. The magnesium ions are represented with grey spheres and labelled. A water molecule (W) involved in interactions within the catalytic site is shown as a red sphere. (**B**) Schematic representation of the RNA interactions. Hydrogen bonds and salt bridges are shown in blue and orange dashed lines, respectively. Magnesium ion coordination are shown in purple. The stacking between Y555 and the UMP base moiety is represented with a grey bar.

which is not observed in the structures obtained after mutation of the catalytic aspartic acid, is coordinated by the phosphate group of the UMP product, the 3′-OH of the second nucleotide ribose moiety, a water molecule, the catalytic D511 and the neighbour S804 (Figure 3). The two magnesium ions are 3.8 Å apart. We hypothesize that the position of this second magnesium may change during catalysis and the TbRRP44-ΔNPIN structure revealed the position corresponding to the state immediately after substrate cleavage.

Substrate entry route

The structures of *S. cerevisiae* and human exosome complexes have revealed details of the RNA interaction and distinct paths employed by the enzymes to drive the substrate to the Rrp44/Dis3 catalytic site (24–26,41,44). Considering that interaction of trypanosomatid RRP44 with the exosome was not detected yet (18–20), the role of the exosome core in channelling the RNA substrate towards TbRRP44 can be disregarded. We investigated whether the TbRRP44-ΔNPIN structure could provide further information about the RNA interaction and substrate entry. Structural comparison with human Dis3, *S. cerevisiae* Rrp44 and *Mus musculus* Dis3L2, the exosome-independent Dis3 homologue devoid of PIN domain, suggests that the substrate path in TbRRP44 might be similar to the one described for ScRrp44 (Figure 4). In the human Dis3 structure, determined in complex with the exosome (41), and MmDis3L2 (45), the RNA enters by a channel on the face of the RNB, formed by CSD1, CSD2 and S1 domains, which is 'closed' in TbRRP44-ΔNPIN structure due to the conformation of CSD2 and S1 domains (Figure 4A,B). Instead, in ScRrp44, the substrate enters by a 'side' path open between CSD1 and the top of the RNB domain (Figure 4C). Two arginine residues at the top of the RNB domain (R811/R896) stabilize the conformation of the sixth nucleotide (from the 3′-end), which is adjacent to the bend of the RNA strand. Although in the TbRRP44-ΔNPIN structure only 5 nucleotides could be modelled in the active site, the structure is very similar to the *S. cerevisiae* homologue, including the side chain conformation of these argi-

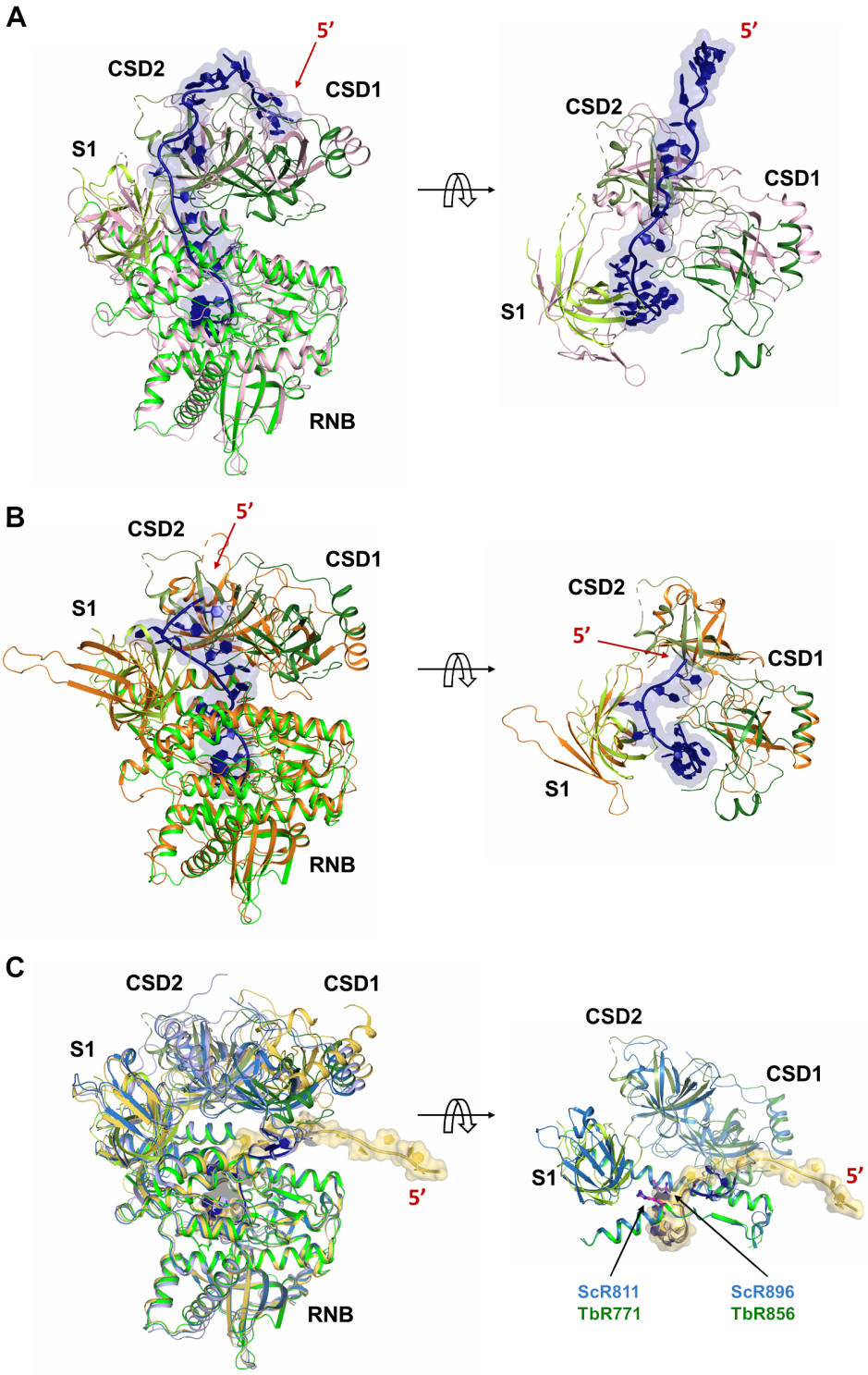

Figure 4. Analysis of the substrate entry path. Structural superposition of *T. brucei* RRP44-ΔNPIN (green) with (**A**) *H. sapiens* Dis3 (pink, extracted from the exosome complex structure, PDB code 6d6r); (**B**) *Mus musculus* DisL2 (orange, PDB code: 4pmw). RNA molecules bound to HsDis3, and MmDisL2 are represented in dark blue. In (**C**) TbRRP44-ΔNPIN (green) is superposed to different S. *cerevisiae* Rrp44 structures. ScRrp44-ΔNPIN (blue, PDB code: 2vnu) with bound RNA in dark blue, ScRrp44 in the open conformation (protein and RNA in light blue, extracted from the exosome complex structure, PDB code: 5c0w) and ScRrp44 in the close conformation (protein and RNA in yellow, extracted from the exosome structure, PDB code: 4ifd). The 5′-end of the bound RNAs is indicated in red and the catalytic site is located at the opposite end, buried in the RNB domain. The right panels show the detail of the substrate entry, by a 'top channel' in HsDis3 and MmDisL2 (A, B) and a 'side entry' in ScRrp44 (**C**). In (C), the ScRrp44 arginine residues which stabilize the conformation of the sixth nucleotide, adjacent to the bend of the RNA strand are highlighted in pink and indicated by arrows (corresponding TbRRP44 residues are also indicated). The right panel in (C) only shows TbRRP44 (green) and ScRrp44-ΔNPIN (blue) structures, for reasons of clarity, and the bound RNA molecules are represented in the same colors as in the left panel.

nine residues (R771/R856 in TbRRP44, Figure 4C). Interestingly, ScRrp44 structures either in a free state (22) or associated with the exosome in an open or close conformation (24,25) show the same 'side' entry path for the RNA (Figure 4C). Our structural analysis suggests that TbRRP44 might adopt the same mechanism for RNA entry, through the 'side' path between the CSD1 and the top of the RNB domain.

TbRRP44 activity on structured and non-structured substrates

In order to investigate the activity of both TbRRP44 ribonucleolytic centers on different RNA substrates, we have produced four TbRRP44 protein variants (Supplementary Figure S5). The full-length TbRRP44 (from now on named TbRRP44^{WT}) includes the complete polypeptide chain without any modification. TbRRP44^{D140N} has a single-point mutation (D140N) designed to inactivate the endonuclease site, whereas TbRRP44^{D511N} was designed to be devoid of exonuclease activity. In addition, the TbRRP44-ΔNPIN construct allowed us to investigate the structural role of the PIN domain on TbRRP44 activity.

TbRRP44^{WT} efficiently degrades single-stranded polyU and polyAU substrates in the presence of different concentrations of Mg^{2+} or Mn^{2+} ions (Figure 5). In contrast, TbRRP44^{WT} degrades polyA (Figure 5), polyC and polyG substrates (data not shown) at much lower rates. Moreover, TbRRP44^{WT} did not show a clear preference for Mg^{2+} or Mn^{2+} when the cofactor was added to the reaction. The two-metal-ion mechanism for nucleic acid phosphodiester cleavage is well known (46,47). In the case of Rrp44/Dis3 homologues, it has been shown that the PIN domain (endonucleolytic domain) would preferentially bind manganese (14,16,21) whereas magnesium ions bind to the RNB active site (22). We observe that TbRRP44^{WT} efficiently degrades polyU and polyAU substrates even without addition of Mg^{2+} or Mn^{2+} ions to the reaction, likely due to the presence of these cofactors stably bound to the protein and retained from the expression host. In fact, the structure of TbRRP44-ΔNPIN presented in this work show magnesium ions in the RNB active site and we have previously observed manganese bound to the active site of TbRRP44 PIN domain (21) without addition of these compounds in any step of protein preparation. However, when added, the cofactors stimulated TbRRP44^{WT} activity, especially towards the polyU substrate, with the effect being noticed at 2 μM or higher concentrations. Moreover, high concentrations of $MgCl_2$ (but not $MnCl_2$) inhibit TbRRP44^{WT} activity (Figure 5). This effect is similar to the inhibition of ScRrp44 activity by high concentrations of $MgCl_2$ reported in the context of the exosome complex (13).

We then investigated the activity of TbRRP44 variants on non-structured and structured substrates. The non-structured substrate consisted of a 49-nucleotide long polyAU. Two different structured RNA substrates containing stem–loops were tested, one forming a GNRA stem–loop without 3′-end overhang (GNRA0), and another forming a GNRA stem–loop with an 18-mer AU-rich 3′-end overhang (GNRA18). As the polyAU, both GNRA0 and GNRA18 are 49-nucleotide long substrates. Recombinant *S. cerevisiae* Rrp44 (ScRrp44) was used in parallel assays for comparison (Figure 6). TbRRP44^{D511N} (inactive RNB) was not able to degrade any of the substrates under the conditions tested (Supplementary Figure S6). TbRRP44^{D140N} (inactive PIN) and the wild-type enzyme degraded the single-stranded polyAU substrate at similar rates (Figure 6A). Interestingly, TbRRP44-ΔNPIN was less active on the polyAU substrate than the TbRRP44^{D140N} variant (Figure 6A), indicating that the PIN domain could have a structural role, contributing to the catalysis carried out by the exoribonucleolytic module.

Compared to the polyAU substrate, the structured RNAs were degraded at lower rates by TbRRP44^{WT}, TbRRP44^{D140N} and TbRRP44-ΔNPIN (Figure 6B and C). Degradation of the GNRA18 substrate results in intermediate bands (Figure 6B) that are not observed with the GNRA0 substrate (Figure 6C). The GNRA18 intermediate products indicate that, in presence of a 3′ overhang, the exonuclease module would promptly engage the substrate, but the internal stem–loop hinders the enzyme activity. Surprisingly, considering the final product formation, our results show that TbRRP44^{WT}, TbRRP44^{D140N} and TbRRP44-ΔNPIN slowly degraded the GNRA0 structured RNA lacking a 3′overhang (Figure 6C). The absence of intermediate products of GNRA0 suggests that, although the degradation is slow, once the 3′-end is released the substrate is continuously degraded.

In contrast to polyAU degradation, analysis of the activity of TbRRP44 on the structured substrates did not show a significant difference between wild-type and TbRRP44^{D140N} or TbRRP44-ΔNPIN variants' efficiencies (Figure 6B and C). In the case of the GNRA18 substrate, TbRRP44^{D140N} appeared to be even slightly more efficient that the wild-type enzyme (Figure 6B), but we believe this result comes from the intrinsic variability of the assays.

S. cerevisiae Rrp44 shows the same degradation pattern observed for TbRRP44 towards the three substrates tested, but at much higher rates. It also degraded the structured RNA in the absence of a 3′overhang and degradation of the GNRA18 substrate also resulted in intermediate bands, although the disappearance of intact substrate was faster than that observed for TbRRP44.

Taken together, our results indicate that TbRRP44 activity is essentially exonucleolytic. However, comparison between TbRRP44^{WT} and TbRRP44-ΔNPIN activities on the polyAU substrate does not allow ruling out the contribution of the PIN domain for substrate degradation, most likely playing a structural role.

The role of TbRRP44 NPIN domain in substrate binding

To investigate the role of TbRRP44 NPIN domain in substrate binding, we have produced two variants devoid of enzymatic activity, the full-length double-mutant TbRRP44$^{D140N/D511N}$ and the truncated single-point mutant TbRRP44-ΔNPIN^{D511N} (Supplementary Figure S5). We performed electrophoretic mobility shift assays (EMSA) to evaluate the interaction of both variants with the 49-nucleotide long non-structured polyAU and stem–loop containing substrates GRNA0 and GNRA18, substrates previously used in degradation assays (Figure 7).

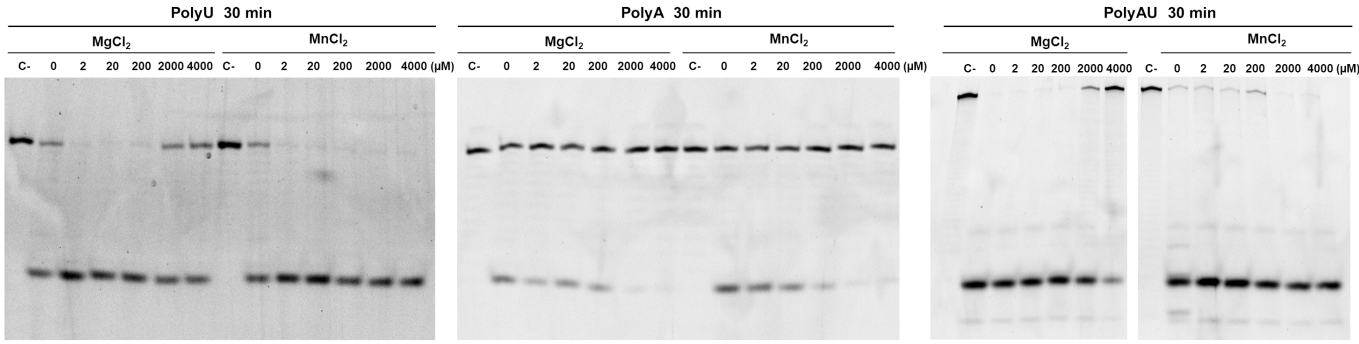

Figure 5. Activity of *T. brucei* RRP44 *in vitro*. Gel-electrophoresis analysis of products generated by the incubation for 30 min of TbRRP44^{WT} with 5′-FAM labelled oligoribonucleotides in the presence of different concentrations of $MgCl_2$ and $MnCl_2$, as indicated. C- stands for negative control (reaction mixture in the absence of protein). TbRRP44 preferentially degrades uracil-containing substrates.

The estimated dissociation constants show that deletion of PIN domain did not significantly affect TbRRP44 interaction with either polyAU or GNRA18 substrates, in contrast to the impact on GRNA0 binding. Regarding the substrates, dissociation constant for TbRRP44–GNRA18 interaction was lower than for polyAU under the conditions tested (Figure 7A and B), however, the most significant difference is observed for the GNRA0 substrate. The absence of single-stranded 3′-end severely affects the substrate interaction with both full-length and truncated variants (Figure 7C), the loss of affinity being more pronounced with the TbRRP44-ΔNPIN^{D511N} construct. At high concentrations of protein, we observe additional slower migrating forms corresponding to protein-RNA complexes, mainly for the complexes involving the full-length construct. We hypothesize that non-specific oligomers may be formed, most likely due to non-functional protein-protein interactions. Interestingly, the data could not be satisfactorily fitted by nonlinear regressions used for quantification of interactions without cooperativity. Instead, data were correctly fitted using the Hill model, as verified by the *R*-squared values. The Hill coefficients indicate an interaction mechanism between TbRRP44 and the RNA substrates involving positive cooperativity ($n > 1$). This result agrees with a model where RNA binding sites located at the PIN domain and at the exonucleolytic module domains play a coordinated role in substrate interaction.

DISCUSSION

Rrp44/Dis3 together with the homologues Dis3L and Dis3L2 have been extensively characterized due to their role in multiple RNA maturation and degradation pathways. The relevance of this group of RNases is also evidenced by the association of mutations in the human homologues to serious diseases (reviewed in (6,48)). Rrp44/Dis3 share the domain organization of bacterial RNase II/RNase R exoribonucleases with the addition of a N-terminal endonuclease PIN domain. The Dis3-like proteins are distinguished by the inactive PIN domain, in the case of Dis3L (8,9), or by its absence in Dis3L2 (10), which have lost the ability to associate with the exosome complex (11). *T. brucei* RRP44 is peculiar in the sense that it conserves the structure and catalytic centers of Rrp44/Dis3 homologues, but it was not found associated with the exosome (18,19).

Despite sharing the general function of degrading RNA in the 3′-5′ sense in a processive manner, members of the RNase II superfamily present specific features regarding their mechanisms and substrate selection. For instance, RNase II is not capable of degrading structured RNA, in contrast to RNase R (49,50). Association of *S. cerevisiae* Rrp44 with the exosome is essential for both endo- and exonucleolytic activities of the complex (51,52). In this work, we sought to investigate the structure and activity of *T. brucei* RRP44, including the roles of the endo- and exonucleolytic domains in the degradation of different RNA substrates.

The three-dimensional structure of the exoribonucleolytic module of TbRRP44, devoid of mutations, revealed unprecedented features of the catalytic site, including the position of the second magnesium ion, involved in the two-metal mechanism for catalysis. We believe that, although our protocol of incubation with EDTA did not inactivate the protein, probably due to its high affinity to the magnesium ions, it severely reduced the reaction rate, and with the time required for crystallization, the protein was fortuitously captured in an intermediate stage of the catalytic cycle, right after the cleavage of the 3′-end nucleotide. It is interesting to notice that the post-cleavage stage observed in the TbRRP44-ΔNPIN structure corresponds, in principle, to the last reaction cycle, since four nucleotides usually remain in the Rrp44/Dis3 catalytic cleft. We hypothesize that the second magnesium ion may have a role in the step of removal of the nucleoside monophosphate and displacement of the RNA substrate for the next round of catalysis. The absence of an excess of magnesium and the binding of the final four-nucleotide RNA chain may have contributed to the slowing of the product release and allowed the observation of its post-cleavage conformation.

RNA degradation assays showed a preference of TbRRP44 for uracil-containing substrates and, surprisingly, revealed the capacity of both TbRRP44 and the yeast homologue to degrade structured RNAs lacking a 3′ overhang. ScRrp44 presented the same degradation pattern but higher RNA degradation rates, compared with the *T. brucei* homologue, towards the three substrates tested. Together with the RNA interaction assays, comparison

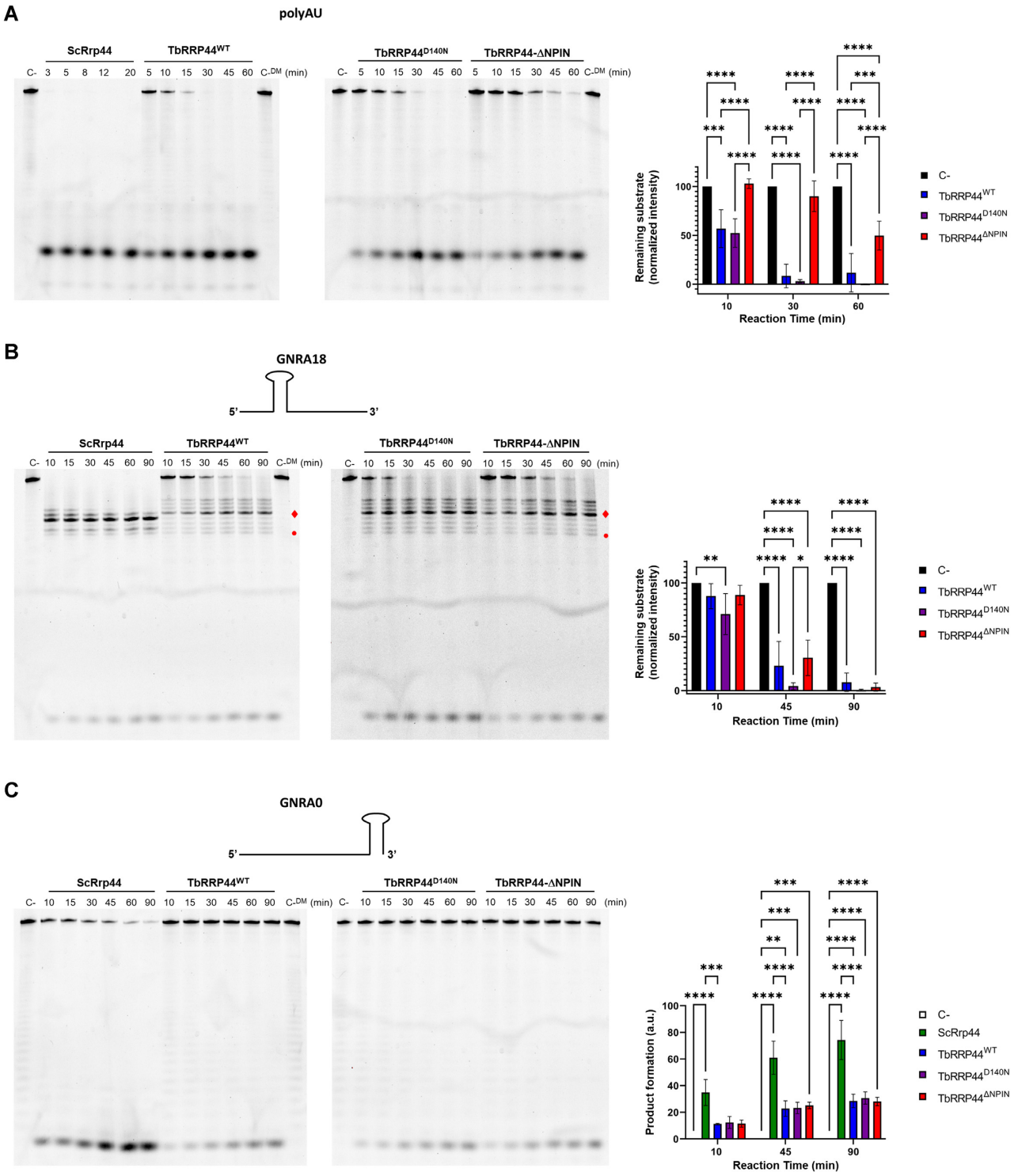

Figure 6. Activity of *T. brucei* RRP44 wild-type and mutants on non-structured and structured RNA. Gel-electrophoresis analysis of products generated by the incubation of TbRRP44 protein variants with 5′-FAM labelled oligoribonucleotides for the durations indicated. Yeast Rrp44 (ScRrp44) activity was analysed in parallel. The gels presented are representative of at least three replicates. Two negative controls were used, C- corresponds to the reaction mixture in the absence of protein and C-^{DM} stands for double mutant, the catalytically inactive variant TbRRP44$^{D140N/D511N}$, which was incubated with the substrate for 90 min. Substrates used are (**A**) single-stranded AU-rich substrate (polyAU), (**B**) GNRA stem–loop with an 18-mer AU-rich 3′-end overhang (GNRA18) and (**C**) substrate containing a GNRA stem–loop without 3′-end overhang (GNRA0). The bar graphs show the quantification of the disappearance of the intact substrate in (A) and (B) and the appearance of the final product in (**C**). Statistical significance is indicated by: **** *P*-value < 0.0001; *** *P*-value 0.0001 to 0.001; ** *P*-value 0.001 to 0.01; * *P*-value 0.01 to 0.05. In the gels of figure (B) the diamond indicates the stronger intermediate product of the GNRA18 degradation, which is ∼7 nucleotides shorter (8-nt for ScRrp44) than the intact substrate and thus corresponds to the stem–loop with ∼11-nt 3′ overhang. The dot indicates the last visible intermediate, corresponding to the substrate containing approximately 7-nt 3′ overhang.

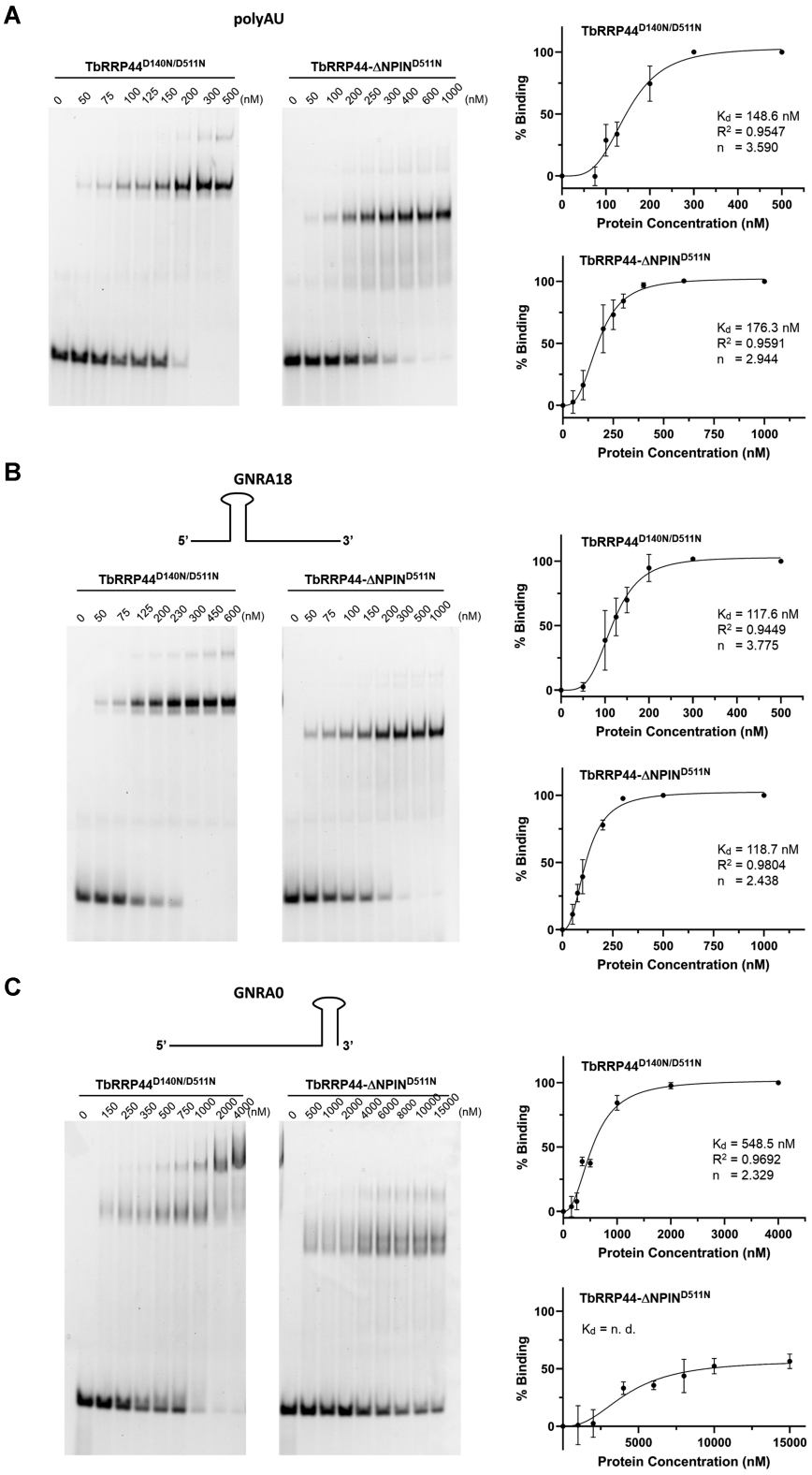

Figure 7. Interaction of full-length and truncated *T. brucei* RRP44 with structured and non-structured RNA. Electrophoretic mobility shift assays were performed using 5′-end fluorescein-labeled (FAM) RNA substrates at 100 nM. The gels presented are representative of at least three replicates. Substrates used are (**A**) single-stranded AU-rich substrate (polyAU), (**B**) GNRA stem–loop with an 18-mer AU-rich 3′-end overhang (GNRA18) and (**C**) substrate containing a GNRA stem–loop without 3′-end overhang (GNRA0). The protein variants and concentrations used in the assays are indicated on the top of the gel. The graphs on the right represent the percentage of binding as a function of protein concentration, with data fitting based on the Hill model equation. Dissociation constants (Kd), R-squared and the Hill coefficients (*n*) are indicated. Few outliers were excluded from some data sets. In (C), quantification of the interaction between TbRRP44-ΔNPIN^{D511N} and the GNRA0 substrate could not be accurately performed due to their very low affinity, as observed in the corresponding gel.

of the activities of the mutants produced in this work on structured and non-structured substrates led us to consider distinct mechanisms that could be employed by the enzyme. To degrade non-structured substrates, TbRRP44 appears to rely exclusively on the exonucleolytic activity, although the PIN domain has a role in the enzyme efficiency, likely participating in the substrate interaction. Wild-type TbRRP44 was capable of degrading stem–loop-containing RNA without a 3′overhang, although with low efficiency. It seems somehow obvious to consider that the 3′-end would be released through endonucleolytic cleavages allowing the exonuclease domain to take over. Nonetheless, our results show that the mutants devoid of endonuclease activity (TbRRP44^{D140N} and TbRRP44-ΔNPIN) still degrade stem–loop RNAs without 3′-end overhang at similar rates to that of the wild-type enzyme. This result suggests that the exonucleolytic module is also able to destabilise RNA structures.

Lorentzen and colleagues showed that yeast Rrp44 was inactive towards RNA duplexes with a small (2 nt) 3′ overhang. Additionally, deletion of PIN domain affected ScRrp44 efficiency in degrading RNA duplexes with 4–5 nt overhang (22). We hypothesize that the differences in ScRrp44 activity observed in our work are associated to the distinct type and composition of the substrate used. We believe that Rrp44/Dis3 from different organisms may be able to degrade a broader spectrum of substrates, including those without a 3′ overhang and their activity depends on parameters such as substrate length, nucleotide composition, as well as type of tertiary structure.

RNA interaction assays help to build a model considering a possible cooperation between the endonuclease domain and the exonucleolytic module, depending on the substrate nature. The absence of a 3′-end overhang affects the protein-substrate interaction, for both full-length and ΔNPIN TbRRP44 variant, however, for substrates lacking a single-stranded 3′-end, the NPIN domain appears to have a critical role in binding (Figure 7). In the case of non-structured substrates or structured RNA with 3′-end overhang, depletion of the NPIN domain did not markedly affect the substrate interaction. However, in all cases, our results indicated a degree of cooperativity involved in the interaction. An aspect to consider is the length of the 5′-end single-strand. In addition to distinct 3′-ends, GNRA0 and GNRA18 substrates have distinct 5′-overhangs (29-nt and 7-nt, respectively). We hypothesize that TbRRP44-RNA interaction involves binding of the single-stranded RNA 3′ segment to the exonuclease module and binding of an upstream RNA segment to the NPIN domain. The 3′-end is directed to exonucleolytic site, whereas the PIN domain would contribute to the processing mechanism by anchoring the region towards the 5′ of the substrate. In the absence of a 3′-overhang, the RNA binding would mainly rely on the PIN domain interaction with the 5′-overhang. Both PIN and the exonucleolytic module may have intrinsic functions in destabilizing structured regions of the substrate and the exonucleolytic module efficiently processes the released single strand. Taken together our results indicate that the exonucleolytic/RNase II module of TbRRP44 alone can degrade structured and non-structured substrate, however, the PIN domain contributes to degradation efficiency of certain types of substrates.

The role of the PIN domain in TbRRP44 activity correlates with *in vivo* data obtained in a genetic complementation study (Guerra-Slompo et al., submitted in parallel to NAR). This parallel study showed that the exonucleolytic activity is essential for TbRRP44 function, while inactivation of the endonucleolytic site by single point mutation has no impact on cell proliferation neither in pre-rRNA maturation. Deletion of the PIN domain, on the other hand, is lethal. This shows that the physical presence of the PIN domain is essential for TbRRP44 function and is consistent with the hypothesis that the PIN domain plays a structural function possibly contributing for RNA interaction. A cooperative mechanism between the endonucleolytic and the exonucleolytic domains seems to be conserved in all eukaryotes. Previous evidence for this mechanism was reported for *S. cerevisiae* and human cells. Deleterious phenotypes are exacerbated when mutation of the endonucleolytic site, which does not cause phenotypes, is combined with non-catalytic site mutants of the exonuclease domain that cause phenotypes (13–16,53,54). In *S. cereviaise*, this synergistic effect can be attributed to the role of the PIN domain in RNA binding (16,23).

Studies on RNase R exoribonucleases have shown that its CSD1, CSD2 and S1 domains have a role on RNA substrate selection, preventing the degradation of blunt-end dsRNA molecules (55). Chu and colleagues showed that a tri-helix region in the RNB domain of RNase R is involved in RNA unwinding which is essential for degradation of structured substrates (56). Our structural analysis and the intermediates observed on the degradation of the substrate GNRA18 lead us to consider possible mechanisms for RNA unwinding and degradation by TbRRP44. Approximately 8 nucleotides is the distance between the catalytic site and the tri-helix region. Approximately 11 nucleotides is the distance to the side entry, located between the RNB and CSD1 domains (Figure 8A), which also corresponds to the 3′ overhang of the GNRA18 intermediate resistant to degradation (Figure 6B). Our results suggest that the stem–loop hinders the enzyme activity until it eventually reaches the tri-helix region where unwinding may occur. Interestingly, the substrate top entry identified in human Dis3 (41) and *M. Musculus* Dis3L2 (45) would also be compatible with the distances to reach the tri-helix and the catalytic site (Figure 8B). Although the conformation of the S1 and CSD2 domains 'closes' this entry in the TbRRP44 structure described in this work, interdomain reorientation is expected upon RNA binding. A model for the full-length TbRRP44 was recently made available on the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/entry/Q95Z12) (57,58). Interestingly, the structure prediction indicates a flexible linker connecting the NPIN domain to the TbRRP44 exoribonucleolytic module (Supplementary Figure S7). Considering the role of the NPIN domain in RNA binding (Figure 7), we speculate that a mechanism by which the protein selects the substrate route depending on the nature of the substrate might be associated to a large reorientation of the PIN domain (Figure 8B). A model where both top and side channels could be employed to direct the substrate for degradation

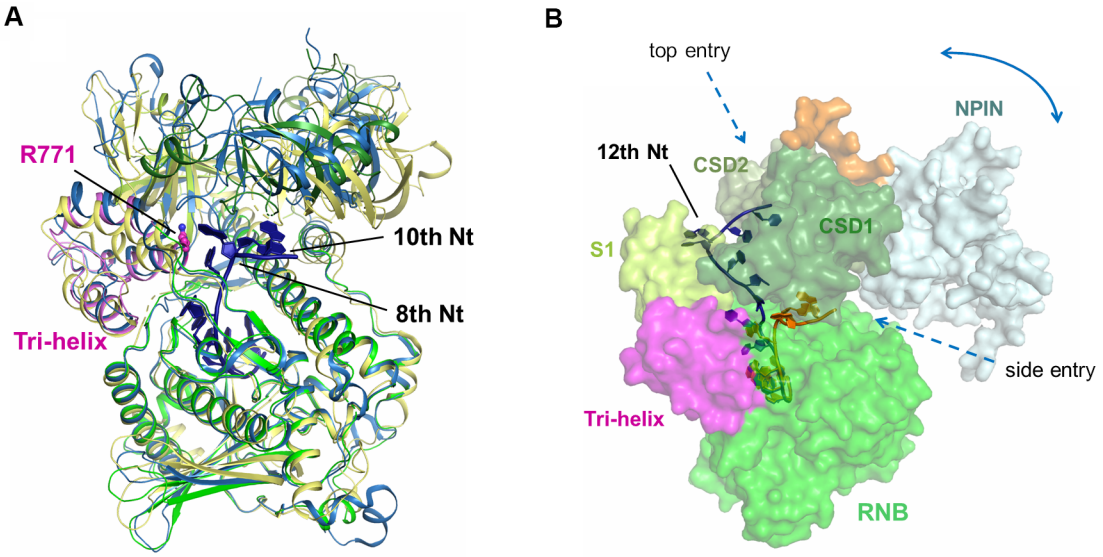

Figure 8. Model for RNA binding. (**A**) Structural superposition, based on the RNB domain, of TbRRP44-ΔNPIN (green), *S. cerevisiae* Rrp44 in complex with RNA (blue, RNA in dark blue, PDB code: 2vnu) and *E. coli* RNase R (yellow, PDB code: 5xgu). The RNB tri-helix region shown to be involved in substrate unwinding in RNase R (56) is shown in pink. The conserved R771 residue (TbRRP44 numbering) located at the C-terminal of the third helix of the tri-helix region is represented in spheres and labelled. The eighth nucleotide (3′–5′ sense) located at the top of the RNB domain in the ScRrp44 structure is indicated. (**B**) Surface representation of full-length TbRRP44 based on the model available at the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/entry/Q95Z12) (57,58). The NPIN domain (absent in the crystal structure) is shown in light blue and the linker between NPIN and CSD1 domains is shown in orange. The CSDs, RNB and S1 domains keep the color code used in previous figures, with the tri-helix region represented in pink. The superposed RNA substrates (cartoon) were extracted from ScRrp44 structure (PDB code 2vnu, side entry path in orange) and from *Mus musculus* DisL2 structure (PDB code 4pmw, top entry path, in dark blue). We propose that a reorientation of the PIN domain may be associated to a mechanism to direct the RNA substrate to either the top or the side channel, depending on the nature of the substrate.

was proposed for RNase R, whose structure showed both channels open in the absence of substrate (56). In the context of the exosome complex, a structural rearrangement of Rrp44 has been associated to distinct mechanisms for substrate recruitment (24,25). The RNA substrate may access the Rrp44 active site by passing through the Exo-9 ring or directly, without going through the exosome central channel. The Rrp44 exonuclease domain undergoes major conformational changes between the direct access conformation and the central channel conformation of the RNA exosome, forming two distinct interfaces with the exosome core subunits. These distinct Rrp44 conformations and substrate recruitment mechanisms have been associated to specific functions of the exosome (59). In a similar manner, in the case of trypanosomatids, where the interaction of RRP44 with the exosome was not demonstrated, conformational changes involving the PIN domain of RRP44 could be associated with the recruitment mechanism of different types of substrates.

In conclusion, the crystal structure of the exonuclease module of TbRRP44 in an active conformation includes the two magnesium ions and provided novel information on the RNA cleavage mechanism. The RNA degradation and interaction assays provided evidence that TbRRP44 presents different catalytic efficiencies according to the RNA primary and tertiary structures. The PIN domain modulates the RNB domain activity most probably by contributing for RNA interaction. Our results showing that both *T. brucei* and yeast Rrp44 are able to degrade structured substrates without 3′-end overhang suggest that the panel of substrates targeted by Rrp44/Dis3 may be larger than what is presently assumed.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structure have been deposited at the Protein Data Bank under accession code 7TUV.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Clemens Vonrhein from Global Phasing Ltd for advice in data processing and model refinement. We are grateful to Carla C. Oliveira from the University of São Paulo for the gift of the plasmid pGEX4T1-ScRrp44. The authors acknowledge the FIOCRUZ program of technical platforms through the Platform for Protein Purification and Characterization (RPT-15A) and the Platform of Crystallography at Pasteur Institute. The authors are grateful to Synchrotron SOLEIL for beam time awarded at PROXIMA1 beam line (Proposal number 20191165) and all the support they benefited as SOLEIL's users.

FUNDING

ICC-CNPq-PROEP [442323/2019-0]; Fiocruz-INOVA [3501948026]; CAPES-COFECUB [CAPES 862/2015 –

COFECUB Me862-15]; Nilson I.T. Zanchin and Beatriz G. Guimarães are CNPq research fellows [304167/2019-3; 304788/2018-0]; studentships for the graduate students Giovanna Cesaro and Heloisa T. da Soler were provided by CNPq and ICC-Fiocruz, respectively. Funding for open access charge: FIOTEC-Fundação de apoio à Fiocruz.
Conflict of interest statement. None declared.

REFERENCES

1. Clayton,C. (2019) Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol.*, **9**, 190072.
2. Rajan,K.S., Chikne,V., Decker,K., Waldman Ben-Asher,H. and Michaeli,S. (2019) Unique Aspects of rRNA Biogenesis in Trypanosomatids. *Trends Parasitol.*, **35**, 778–794.
3. Schneider,C. and Tollervey,D. (2013) Threading the barrel of the RNA exosome. *Trends Biochem. Sci.*, **38**, 485–493.
4. Mitchell,P. (2014) Exosome substrate targeting: the long and short of it. *Biochem Soc Trans*, **42**, 1129–1134.
5. Januszyk,K. and Lima,C.D. (2014) The eukaryotic RNA exosome. *Curr. Opin. Struct. Biol.*, **24**, 132–140.
6. Reis,F.P., Pobre,V., Silva,I.J., Malecki,M. and Arraiano,C.M. (2013) The RNase II/RNB family of exoribonucleases: putting the 'Dis' in disease. *Wiley Interdiscip. Rev. RNA*, **4**, 607–615.
7. Robinson,S.R., Oliver,A.W., Chevassut,T.J. and Newbury,S.F. (2015) The 3′ to 5′ exoribonuclease DIS3: from structure and mechanisms to biological functions and role in human disease. *Biomolecules*, **5**, 1515–1539.
8. Tomecki,R., Kristiansen,M.S., Lykke-Andersen,S., Chlebowski,A., Larsen,K.M., Szczesny,R.J., Drazkowska,K., Pastula,A., Andersen,J.S., Stepien,P.P. *et al.* (2010) The human core exosome interacts with differentially localized processive RNases: HDIS3 and hDIS3L. *EMBO J.*, **29**, 2342–2357.
9. Staals,R.H.J., Bronkhorst,A.W., Schilders,G., Slomovic,S., Schuster,G., Heck,A.J.R., Raijmakers,R. and Pruijn,G.J.M. (2010) Dis3-like 1: a novel exoribonuclease associated with the human exosome. *EMBO J.*, **29**, 2358–2367.
10. Astuti,D., Morris,M.R., Cooper,W.N., Staals,R.H.J., Wake,N.C., Fews,G.A., Gill,H., Gentle,D., Shuib,S., Ricketts,C.J. *et al.* (2012) Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nat. Genet.*, **44**, 277–284.
11. Lubas,M., Damgaard,C.K., Tomecki,R., Cysewski,D., Jensen,T.H. and Dziembowski,A. (2013) Exonuclease hDIS3L2 specifies an exosome-independent 3′-5′ degradation pathway of human cytoplasmic mRNA. *EMBO J.*, **32**, 1855–1868.
12. Malecki,M., Viegas,S.C., Carneiro,T., Golik,P., Dressaire,C., Ferreira,M.G. and Arraiano,C.M. (2013) The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. *EMBO J.*, **32**, 1842–1854.
13. Dziembowski,A., Lorentzen,E., Conti,E. and Séraphin,B. (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat. Struct. Mol. Biol.*, **14**, 15–22.
14. Lebreton,A., Tomecki,R., Dziembowski,A. and Séraphin,B. (2008) Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature*, **456**, 993–996.
15. Schaeffer,D., Tsanova,B., Barbas,A., Reis,F.P., Dastidar,E.G., Sanchez-Rotunno,M., Arraiano,C.M. and Van Hoof,A. (2009) The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat. Struct. Mol. Biol.*, **16**, 56–62.
16. Schneider,C., Leung,E., Brown,J. and Tollervey,D. (2009) The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic. Acids. Res.*, **37**, 1127–1140.
17. Schaeffer,D., Reis,F.P., Johnson,S.J., Arraiano,C.M. and Van Hoof,A. (2012) The CR3 motif of Rrp44p is important for interaction with the core exosome and exosome function. *Nucleic Acids Res.*, **40**, 9298–9307.
18. Estévez,A.M., Kempf,T. and Clayton,C. (2001) The exosome of *Trypanosoma brucei*. *EMBO J.*, **20**, 3831–3839.
19. Estévez,A.M., Lehner,B., Sanderson,C.M., Ruppert,T. and Clayton,C. (2003) The roles of intersubunit interactions in exosome stability. *J. Biol. Chem.*, **278**, 34943–34951.
20. Cristodero,M., Böttcher,B., Diepholz,M., Scheffzek,K. and Clayton,C. (2008) The *Leishmania tarentolae* exosome: purification and structural analysis by electron microscopy. *Mol. Biochem. Parasitol.*, **159**, 24–29.
21. Cesaro,G., Carneiro,F.R.G., Ávila,A.R., Zanchin,N.I.T. and Guimarães,B.G. (2019) Trypanosoma brucei RRP44 is involved in an early stage of large ribosomal subunit RNA maturation. *RNA Biol.*, **16**, 133–143.
22. Lorentzen,E., Basquin,J., Tomecki,R., Dziembowski,A. and Conti,E. (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol. Cell*, **29**, 717–728.
23. Bonneau,F., Basquin,J., Ebert,J., Lorentzen,E. and Conti,E. (2009) The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell*, **139**, 547–559.
24. Makino,D.L., Baumgärtner,M. and Conti,E. (2013) Crystal structure of an rna-bound 11-subunit eukaryotic exosome complex. *Nature*, **495**, 70–75.
25. Makino,D.L., Schuch,B., Stegmann,E., Baumgärtner,M., Basquin,C. and Conti,E. (2015) RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature*, **524**, 54–58.
26. Zinder,J.C., Wasmuth,E.V. and Lima,C.D. (2016) Nuclear RNA exosome at 3.1 Å reveals substrate specificities, RNA paths, and allosteric inhibition of Rrp44/Dis3. *Mol. Cell*, **64**, 734–745.
27. Gerlach,P., Schuller,J.M., Bonneau,F., Basquin,J., Reichelt,P., Falk,S. and Conti,E. (2018) Distinct and evolutionary conserved structural features of the human nuclear exosome complex. *Elife*, **7**, e38686.
28. Kabsch,W. (2010) 1 XDS. *Acta Crystallogr. Sect. D*, **66**, 125–132.
29. Vonrhein,C., Flensburg,C., Keller,P., Sharff,A., Smart,O., Paciorek,W., Womack,T. and Bricogne,G. (2011) Data processing and analysis with the autoPROC toolbox, *Acta Cryst.*, **D67**, 293–302.
30. Tickle,I.J., Flensburg,C., Keller,P., Paciorek,W., Sharff,A., Vonrhein,C. and Bricogne,G. (2018) STARANISO.
31. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 22–25.
32. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
33. Bricogne,G., Blanc,E., Brandl,M., Flensburg,C., Keller,P., Paciorek,W., Roversi,P, Sharff,A., Smart,O.S., Vonrhein,C. *et al.* (2019) BUSTER version 2.10.3.
34. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **60**, 2126–2132.
35. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 12–21.
36. Adasme,M.F., Linnemann,K.L., Bolz,S.N., Kaiser,F., Salentin,S., Haupt,V.J. and Schroeder,M. (2021) PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Res.*, **49**, W530–W534.
37. Heus,H.A. and Pardi,A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*, **253**, 191–194.
38. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
39. Frazão,C., McVey,C.E., Amblar,M., Barbas,A., Vonrhein,C., Arraiano,C.M. and Carrondo,M.A. (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature*, **443**, 110–114.
40. Zuo,Y., Vincent,H.A., Zhang,J., Wang,Y., Deutscher,M.P. and Malhotra,A. (2006) Structural basis for processivity and single-strand specificity of RNase II. *Mol. Cell*, **24**, 149–156.
41. Weick,E.M., Puno,M.R., Januszyk,K., Zinder,J.C., DiMattia,M.A. and Lima,C.D. (2018) Helicase-dependent RNA decay illuminated by a Cryo-EM structure of a human nuclear RNA exosome-MTR4 complex. *Cell*, **173**, 1663–1677.

42. Arraiano,C.M., Matos,R.G. and Barbas,A. (2010) RNase II: the finer details of the Modus operandi of a molecular killer. *RNA Biol.*, **7**, 276–281.
43. Barbas,A., Matos,R.G., Amblar,M., López-Viñas,E., Gomez-Puertas,P. and Arraiano,C.M. (2009) Determination of key residues for catalysis and RNA cleavage specificity: ONE MUTATION TURNS RNase II INTO A "SUPER-ENZYME". *J. Biol. Chem.*, **284**, 20486–20498.
44. Liu,J.J., Bratkowski,M.A., Liu,X., Niu,C.Y., Ke,A. and Wang,H.W. (2014) Visualization of distinct substrate-recruitment pathways in the yeast exosome by em. *Nat. Struct. Mol. Biol.*, **21**, 95–102.
45. Faehnle,C.R., Walleshauser,J. and Joshua-Tor,L. (2014) Mechanism of Dis3L2 substrate recognition in the Lin28/let-7 pathway. *Nature*, **514**, 252–256.
46. Steitz,T.A. and Steitz,J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 6498–6502.
47. Palermo,G., Cavalli,A., Klein,M.L., Alfonso-Prieto,M., Dal Peraro,M. and De Vivo,M. (2015) Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.*, **48**, 220–228.
48. M,S., PJ,daC., SC,V. and CM,A. (2019) The implication of mRNA degradation disorders on human disease: focus on DIS3 and DIS3-like enzymes. *Adv. Exp. Med. Biol.*, **1157**, 85–98.
49. Cheng,Z.F. and Deutscher,M.P. (2005) An important role for RNase R in mRNA decay. *Mol. Cell*, **17**, 313–318.
50. Cheng,Z.F. and Deutscher,M.P. (2002) Purification and characterization of the *Escherichia coli* exoribonuclease RNase R. Comparison with RNase II. *J. Biol. Chem.*, **277**, 21624–21629.
51. Wasmuth,E.V. and Lima,C.D. (2012) Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. *Mol. Cell*, **48**, 133–144.
52. Drążkowska,K., Tomecki,R., Stoduś,K., Kowalska,K., Czarnocki-Cieciura,M. and Dziembowski,A. (2013) The RNA exosome complex central channel controls both exonuclease and endonuclease Dis3 activities in vivo and in vitro. *Nucleic. Acids. Res.*, **41**, 3845–3858.
53. Tomecki,R., Drazkowska,K., Kucinski,I., Stodus,K., Szczesny,R.J., Gruchota,J., Owczarek,E.P., Kalisiak,K. and Dziembowski,A. (2014) Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Res.*, **42**, 1270–1290.
54. Kobyłecki,K., Drążkowska,K., Kuliński,T.M., Dziembowski,A. and Tomecki,R. (2018) Elimination of 01/A′-A0 pre-rRNA processing by-product in human cells involves cooperative action of two nuclear exosome-associated nucleases: RRP6 and Dis3. *RNA*, **24**, 1677–1692.
55. Matos,R.G., Barbas,A. and Arraiano,C.M. (2009) RNase R mutants elucidate the catalysis of structured RNA: rNA-binding domains select the RNAs targeted for degradation. *Biochem. J.*, **423**, 291–301.
56. Chu,L.Y., Hsieh,T.J., Golzarroshan,B., Chen,Y.P., Agrawal,S. and Yuan,H.S. (2017) Structural insights into RNA unwinding and degradation by RNase R. *Nucleic Acids Res.*, **45**, 12015–12024.
57. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
58. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A. *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
59. Han,J. and van Hoof,A. (2016) The RNA exosome channeling and direct access conformations have distinct in vivo functions. *Cell Rep.*, **16**, 3348–3358.