10862–10878 *Nucleic Acids Research, 2016, Vol. 44, No. 22*
doi: 10.1093/nar/gkw917

RNA Editing TUTase 1: structural foundation of substrate recognition, complex interactions and drug targeting

Lional Rajappa-Titu[1,†], Takuma Suematsu[2,†], Paola Munoz-Tello[1], Marius Long[1], Özlem Demir[3], Kevin J. Cheng[3], Jason R. Stagno[4], Hartmut Luecke[4], Rommie E. Amaro[3], Inna Aphasizheva[2], Ruslan Aphasizhev[2,5,*] and Stéphane Thore[1,6,7,8,*]

[1]Department of Molecular Biology, University of Geneva, 1211 Geneva, Switzerland, [2]Department of Molecular and Cell Biology, Boston University School of Dental Medicine, Boston, MA 02118, USA, [3]Department of Chemistry & Biochemistry and the National Biomedical Computation Resource, University of California, San Diego, La Jolla, CA 92093, USA, [4]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697, USA, [5]Department of Biochemistry, Boston University School of Medicine, Boston, MA 02118, USA, [6]INSERM, U1212, ARNA Laboratory, Bordeaux 33000, France, [7]CNRS UMR5320, ARNA Laboratory, Bordeaux 33000, France and [8]University of Bordeaux, ARNA Laboratory, Bordeaux 33000, France

Received May 10, 2016; Revised September 27, 2016; Accepted October 04, 2016

ABSTRACT

Terminal uridyltransferases (TUTases) execute 3′ RNA uridylation across protists, fungi, metazoan and plant species. Uridylation plays a particularly prominent role in RNA processing pathways of kinetoplastid protists typified by the causative agent of African sleeping sickness, *Trypanosoma brucei*. In mitochondria of this pathogen, most mRNAs are internally modified by U-insertion/deletion editing while guide RNAs and rRNAs are U-tailed. The founding member of TUTase family, RNA editing TUTase 1 (RET1), functions as a subunit of the 3′ processome in uridylation of gRNA precursors and mature guide RNAs. Along with KPAP1 poly(A) polymerase, RET1 also participates in mRNA translational activation. RET1 is divergent from human TUTases and is essential for parasite viability in the mammalian host and the insect vector. Given its robust *in vitro* activity, RET1 represents an attractive target for trypanocide development. Here, we report high-resolution crystal structures of the RET1 catalytic core alone and in complex with UTP analogs. These structures reveal a tight docking of the conserved nucleotidyl transferase bi-domain module with a RET1-specific C2H2 zinc finger and RNA recognition (RRM) domains. Furthermore, we define RET1 region required for incorporation into the 3′ processome, determinants for RNA binding, subunit oligomerization and processive UTP incorporation, and predict druggable pockets.

INTRODUCTION

Parasitic protists from the order of *Kinetoplastidae* cause devastating human and animal infections such as African sleeping sickness, Nagana, Chagas disease, leishmaniasis and others (1). These early branching eukaryotes share many unique mechanisms of gene expression including multicistronic transcription of nuclear protein-encoding genes by RNA polymerase II and intricate mitochondrial RNA processing pathways (2,3). To enable synthesis of 18 predicted mitochondrial proteins, messenger RNA precursors must be transcribed from maxicircle DNA, 3′ adenylated, often subjected to extensive U-insertion/deletion editing, and further modified by 3′ adenylation/uridylation prior to translation (4–6). The principal role of uridylation in trypanosomal mitochondrial RNA metabolism motivated identification of the first terminal uridyltransferase (TUTase) RET1 (7). In the following decade, the perception of RNA uridylation has evolved from an unusual reaction in an enigmatic organelle into a recognition of U-tailing as the major transcriptome-shaping force in eukaryotes (8).

---

[*]To whom correspondence should be addressed. Tel: +33 5 57 57 15 11; Email: stephane.thore@inserm.fr
Correspondence may also be addressed to Ruslan Aphasizhev. Tel: +1 617 414 1055; Email: ruslana@bu.edu
[†]These authors contributed equally to this work as the first authors.
Present addresses:
Paola Munoz-Tello, Department of Molecular Therapeutics, The Scripps Research Institute-Scripps Florida, FL 33458, USA.
Jason R. Stagno, Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA.

Earlier studies implicated RET1 TUTase in adding 10–15 nt-long U-tails to guide RNAs (gRNAs) and ribosomal RNAs (9,10). The 50–60 nt-long gRNAs bind to and direct the editing of 12 among 18 mitochondrial pre-mRNAs (11). In addition, observed UTP-stimulated *in organello* mRNA degradation hinted at possible involvement of RET1-catalyzed uridylation in mRNA decay (12,13). Finally, sequencing of mRNA 3′ termini revealed that most mRNAs in the actively respiring insect (procyclic) developmental stage of the parasite are modified by addition of 200–300 nt-long A/U-heteropolymers (14), which indicated RET1's involvement in mRNA 3′ processing and translational activation (15). Subsequent work demonstrated that RET1 is not only required for 3′ uridylation on mature gRNAs, ribosomal RNAs and some mRNAs, but also participates in nucleolytic processing of precursors transcribed from maxicircle and minicircle genomes (16,17). In-depth investigation of mitochondrial high molecular mass TUTase-containing complexes revealed that RET1 is mostly sequestered into a stable particle along with DSS1 3'-5' exonuclease and three proteins lacking annotated motifs (16). This ~900 kDa complex, termed the mitochondrial 3' processome (MPsome), is responsible for recognition, uridylation and exonucleolytic processing of 800–1200 nt-long gRNA precursors, and U-tailing of mature gRNAs. Thus, the enzymes with seemingly opposing enzymatic activities, 3' addition and 3'-5' degradation, function as subunits of a single protein complex. This is a profound example of coupling between TUTase and RNase II-like exonuclease, an apparently evolutionarily conserved RNA decay pathway (18–20). Apart from binding to the MPsome, RET1 transiently interacts with a complex of kinetoplast polyadenylation factors 1 and 2 (KPAF1/2) and mitochondrial poly(A) polymerase KPAP1. Pentatricopeptide repeat (PPR)-containing RNA binding factors KPAF1/2 coordinate TUTase and poly(A) polymerase activities in adding A/U-heteropolymers to the mature mRNA 3' termini, which is required for translational activation (15,21).

High resolution atomic structures of apo forms and binary complexes with substrate or non-substrate nucleotides have been determined for three TUTases from *T. brucei*. These include mitochondrial RET2 (22) and MEAT1 (23), and cytosolic TUT4 (24,25) proteins. In addition, the atomic mechanism of UTP recognition by Cid1 TUTase from *Schizosaccharomyces pombe* has been reported (26,27). These studies identified a common enzymatic module, which is formed by fusion of Pol-β DNA polymerase-like catalytic metal binding and UTP-binding domains (28,29). These crystallographic studies were centered on the UTP recognition pocket and the catalytic domain. However, more complex TUTases, such as trypanosomal RET1 or human TUT4 and TUT7, contain multiple unexplored domains involved in oligomerization, RNA–protein and protein–protein interactions (30–32). *In vitro* studies of the recombinant RET1 from related parasite *Leishmania tarentolae* suggest that it likely assembles into a tetramer capable of processively adding hundreds of uridines to a single-stranded RNA (ssRNA); limited U-addition to partially double-stranded RNA (dsRNA) has also been detected (7,33). The essential role of the C2H2 zinc finger domain for activity was established, but its precise function remained unclear (33). The critical and well-understood roles in at least two essential RNA processing pathways, and robust *in vitro* UTP polymerization activity render RET1 an attractive target for trypanocide development. To that end, screening of ~3000-compound library led to identification of competitive inhibitors with K*i* constants in a low micromolar range (34).

Here, we report the high-resolution atomic structure of the RET1 core region obtained by deleting N- and C-terminal extensions that show no similarity to proteins outside *Kinetoplastidae*. The acquired structures of binary complexes with UTP and UTP analogs encompass the enzymatic module plus the zinc finger domain and the putative RRM fold inserted within the catalytic domain. Structural comparison of the zinc finger domain localization and orientation suggests its capacity to bind double-stranded RNA. We also solved the structure of a minimal TUT4 TUTase (24,25) in ternary complex with UTP and RNA-imitating UpU dinucleotide. The TUT4-UTP-UpU structure was used in molecular modeling experiments to reveal the RET1's RNA binding determinants in the vicinity of the active site. In addition, we applied molecular dynamics simulations to determine potentially druggable cavities in UTP-bound RET1. The mutational and biochemical analyses demonstrate that deleting the N-terminal extension does not affect RET1 activity *in vitro*, but hinders protein oligomerization and *in vivo* function. We demonstrate that the C-terminal region and the RRM domains contribute to RNA binding and, therefore, processivity of UTP polymerization reaction *in vitro*. We further show that the N-, C- and N-C terminally truncated RET1 variants successfully incorporate into the mitochondrial 3' processome, but cause a dominant negative phenotype. In summary, our findings reveal the high-resolution structure of the key mitochondrial RNA processing enzyme in trypanosomes and deepen the structure-function understanding of the TUTase family. This work also provides a foundation for structure-based design of TUTase inhibitors with potential trypanocidal activities.

MATERIALS AND METHODS

Protein expression and purification

All reagents and materials were purchased from Thermo Fisher Scientific unless otherwise noted. Full-length RET1 was purified from bacteria as described (35). Gene fragments corresponding to residues 189–976 (RET1ΔN), 1–699 (RET1ΔC) and 189–699 (RET1ΔNC) were amplified by PCR and cloned into the pST0 vector (derived from pET42) as fusion proteins with a N-terminal GST tag. The Tobacco Etch Virus (TEV) protease cleavage site and a 6XHis tag were placed between the GST tag and the protein of interest. The RET1ΔNC D473A and RRR mutants (R387A, R422A, R453A) were produced by site-directed mutagenesis using the QuickChange kit (Stratagene). The recombinant proteins were overexpressed in *Escherichia coli* BL21-Star strain, grown in Terrific Broth media at 37°C for 4 h followed by induction at 18°C for 16 h with 0.1 mM isopropyl-ß-D-thiogalactopyranoside. Cells were lysed in 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 20 mM Imidazole, 0.1% X-Triton 100, DNase I (1 μg/ml), lysozyme

1 μg/ml, 5 mM ß-mercaptoethanol with protease inhibitor cocktail (Roche) by using an Emulsiflex (AVESTIN). Extracts were cleared by centrifugation at 20 000 ×*g* for 40 min and was loaded onto a HiPrep IMAC FF 16/10 column (GE Healthcare). The protein was eluted with buffer containing 250 mM imidazole and TEV protease was added at 1/50 (w/w) ratio, and incubated for 16 h at 4°C. The cleaved protein was separated on 5 ml HiTrap Heparin HP column (GE Healthcare) after diluting the sample 5-fold with buffer containing 50 mM Tris-HCl, pH 7.5 and 5 mM ß-mercaptoethanol. The sample was eluted by a linear gradient of NaCl from 0.1 to 1 M in 50 mM Tris-HCl, pH 7.5 and 5 mM ß-mercaptoethanol. Remaining impurities were removed by gel-filtration on a Superdex 75 column (GE Healthcare) in 20 mM MES pH 6.5, 300 mM NaCl and 1 mM TCEP. Samples were concentrated to 4 mg/ml and used for crystallization trials immediately. Apparent molecular masses of the full-length and truncated RET1 variants were determined by size exclusion chromatography in 20 mM Tris-HCl, pH 7.5, 150 mM KCl, 1 mM DTT using a Superose 12 10/300GL column (GE Life Sciences).

Protein crystallization and X-ray data collection

Initial RET1ΔNC crystals appeared in 0.1 M Bicine pH 9.0, 20% polyacrylate 2100, 0.2 M NaCl (MIDAS-Molecular Dimensions) using the sitting-drop vapor diffusion technique with 1:1 ratio at 18°C. These were crushed and used as seeds. After several rounds of optimization, thin plate-like crystals grew in solution containing 0.1 M bicine pH 8.6, 20% polyacrylate 2100, 0.2 M NaI supplemented with non-diluted seeds using the microbatch crystallization technique. Diffraction from these thin plates was observed up to 3.5 Å resolution. In order to obtain better diffracting crystals, thin plates were crushed and used as seeds to perform sitting drop matrix seeding at 18°C with commercial screens. Clusters of small crystals grew in a reservoir containing 0.1 M Tris pH 8.5, 25% PEG 3350, 0.2 M lithium sulfate monohydrate (Index screen – Hampton Research). These clustered crystals were crushed and streak-seeded onto an optimized condition containing 0.1 M Tris pH 8.5, 19% PEG 3350, 0.2 M Lithium sulfate monohydrate that produced crystals of maximum size 0.2 × 0.15 × 0.05 mm in 1 week. Crystals of RET1ΔNC in complex with the UTP analog UMPNPP were obtained by soaking the native crystals grown with lithium chloride instead of lithium sulfate monohydrate for 12 h in crystallization buffer supplemented with 0.5 mM of UMPnPP and 1 mM $MgCl_2$. Crystals of the D473A mutant grew in a condition containing 0.1 M Tris pH 8.5, between 5% and 8% PEG 3350, 0.2 M lithium chloride. Crystals of the complex of D473A with UTP was obtained by soaking crystals of D473A in crystallization buffer supplemented with 1.5 mM UTP and 1 mM $MgCl_2$. Crystals were cryo-protected by passing them into the respective crystallization condition supplemented with 20% sucrose or 20% glycerol before flash freezing in liquid nitrogen. X-ray diffraction data for the native and UMPnPP soaked crystals were collected up to 1.6 Å and 1.8 Å, respectively, at the X06DA beamline of the Swiss Light Source (SLS, Villigen–Paul Scherrer Institute). Diffraction data for D473A in complex with UTP were collected up to 2.7 Å at X06SA of the SLS. The data collection statistics are reported in Table 1. Recombinant TUT4 from *T. brucei* was purified from *E. coli* and crystalized in the presence of UTP followed by soaking in 1 mM UpU solution as described (25).

Structure determination and refinement

The data sets were indexed using the XDS package (36). The structure of RET1ΔNC was determined by molecular replacement using the TUT4 atomic structure (33% sequence identity; PDB code 2IKF) as a search model using the program PHASER from the CCP4 package (37). Additional density corresponding to the RRM domain and the zinc-binding region appeared after the first cycle of refinement. Several cycles of model building in COOT (38) and refinement in PHENIX (39) were necessary to obtain a complete model of RET1ΔNC. Residues 369–380, 624–643 and 689–699 were not visible in the $2F_o$-F_c electron density map, most likely due to flexibility. Once the model for the native data set was complete, it was used to phase the UMPnPP-soaked and D473A-UTP-soaked data sets. Large unidentified peaks of electron density appeared in the active sites of the UMPnPP-soaked and D473A-UTP-soaked structures. The coordinates for the UMPnPP and UTP were obtained from the monomer library of the program COOT and modeled in the $2F_o$-F_c electron density map. The final RET1ΔNC, RET1ΔNC – UMPnPP and RET1ΔNC D473A – UTP models were refined to 1.6 Å, 1.8 Å and 2.2 Å resolution, respectively. The refinement statistics are reported in Table 1. The TUT4:UTP: UpU structure was solved by refining the original TUT4:UTP structure (PDB ID code 2IKF) against the structure factors of the TUT4:UTP:UpU while preserving the test set.

Molecular dynamics simulations

We used the apo and UMPnPP-bound crystal structures to generate starting structures for molecular dynamics (MD) simulations. UMPnPP in the crystal structure was modified computationally to obtain UTP, and the missing residues/loops in the crystal structures were completed with the Prime module of the Schrodinger suite. The Cationic Dummy Atom Model was used for the zinc ion and its coordinating residues (40). Crystallographic water molecules were kept except for those that sterically clash with the built-in residues for the missing loops. The protonation states of residues were determined with the program PropKA (41). The system was solvated in a TIP3P water box forming a 10 Å buffer between the protein and periodic boundary, and Na and Cl ions were added to provide a salt concentration of 0.2 M (42). The Amber FF14SB force field in the Amber 14 suite was used to parameterize the protein (http://ambermd.org/). For UTP, the parameters generated in (43) were used. Three independent copies of 250-ns MD simulations were performed for apo RET1 and UTP-bound RET1 separately (a total of 750 ns for each system). The final system consisted of 103 944 atoms. The final conformation of the UTP-bound RET1 simulation with the most open active site among the three MD copies was used to model 10-mer oligo(U) in the UTP-bound RET1. The

Table 1. Crystallographic Data and Refinement Statistics

Crystal	RET1ΔNC (189–699)	RET1ΔNC - UMPnPP	RET1(D473A)ΔNC - UTP	TUT4 - UTP - UpU
Data collection				
Wavelength (Å)	1.000	1.000	1.000	0.979
Resolution range (Å)	46.49–1.50 (1.54–1.50)	46.4–1.8 (1.85–1.80)	19.7–2.20 (2.26–2.20)	47.13–2.75 (2.85–2.75)
Space group	$P\ 2_1\ 2_1\ 2$	$P\ 2_1\ 2_1\ 2$	$P\ 2_1\ 2\ 2_1$	$P\ 2_1$
Unit cell dimensions	$a = 129.8$	$a = 129.3$	$a = 59.3$	$a = 82.69$
	$b = 58.1$	$b = 58.4$	$b = 66.3$	$b = 42.88$
	$c = 66.6$	$c = 66.6$	$c = 128.9$	$c = 109.62$
	$\alpha = \beta = \gamma = 90^\circ$	$\alpha = \beta = \gamma = 90^\circ$	$\alpha = \beta = \gamma = 90^\circ$	$\beta = 94.05^\circ$
Total reflections	1 062 285	626 853	431 657	71 616
Unique reflections	81 270	47 571	35 049	20 369
Redundancy	13.1 (11.7)	13.2 (12.5)	12.1 (12.1)	3.52 (3.48)
Completeness (%)	99.7 (96.7)	99.9 (99.6)	99.4 (99.8)	99.5 (99.9)
Mean I/ σ	18.7 (1.6)	31.7 (5.76)	22.9 (3.9)	7.1 (2.2)
R_{meas}	0.083 (1.71)	0.056 (0.472)	0.075 (0.683)	0.104 (0.374)
Refinement				
Resolution range (Å)	24.6–1.6	46.4–1.8	19.7–2.2	47.13–2.75
No. of atoms	4510	4353	3975	5470
Protein	3906	3812	3773	5144
Ligands (UMPnPP, UTP, UpU)	-	29	29	132
Magnesium		1	1	5
Zinc	1	1	1	-
Sulfate	10	-	-	-
Chlorine	1	-	-	-
Water	592	510	171	189
Protein residues	476	471	469	641
R_{work}	0.16 (0.24)	0.17 (0.21)	0.24 (0.27)	0.20 (0.31)
R_{free}	0.20 (0.28)	0.21 (0.27)	0.28 (0.34)	0.26 (0.43)
R.m.s. deviations				
Bond lengths (Å)	0.01	0.007	0.003	0.02
Bond angles (°)	1.31	1.08	0.78	0.56
Average B factors (Å^2)				
Protein	22.4	23.1	42.0	48.5
Ligands	-	21.3	49.3	46.6
Ions (Mg^{+2}, Zn^{+2})	23.4	23.8	49.0	34.7
Water	34.8	34.3	41.6	33.6
Ramachandran				
favored (%)	99	98.7	98.0	98
allowed (%)	0.6	1.0	2.1	2
outliers (%)	0.2	0.2	0	0

10U ssRNA was generated from the Nucleic Acid Builder web server (http://structure.usc.edu/make-na). The terminal and penultimate U nucleotides were modeled to mimic the binding pose of a UpU- and UTP-bound TUT4 crystal structure (PDB ID code 5KAL). After standard minimization and equilibration procedures (below), three independent 250 ns MD simulations were performed for each of apo RET1ΔNC and UTP-bound RET1ΔNC systems (a total of 750 ns for each system) while five independent 50-ns MD simulations were performed for RNA/UTP-bound RET1ΔNC system (a total of 250 ns, Supplementary Figure S1). A total of 13 000 steps of energy minimization were carried out to remove artificial contacts and relax the systems. Only the hydrogen atoms were relaxed in the first 2000 steps of minimization, having all other atoms fixed. In the second 2000 steps, all water atoms, ions and hydrogen atoms were relaxed while the protein and ligand were fixed. In the third 2000 steps, all atoms were relaxed while the protein was held fixed. In the following 2000 steps, all atoms were relaxed except for protein backbone atoms. Each of these minimization procedures included a positional restraint weight of 10.0 kcal/ mol Å^2. In the last 5000 steps, the entire system was relaxed with no positional restraints. Following the minimization, 50 ps of heating at constant volume from 0.0 to 100.0 K was performed, followed by 200 ps of heating at constant pressure from 100.0 to 300.0 K with protein backbone atoms constrained with a weight of 5.0 kcal/mol Å^2. Thereafter, an additional 250 ps of heating was performed to prevent structural artifacts from introducing kinetic energy into the system. Subsequently, unconstrained molecular dynamics was performed at 1 bar and at 300 K using a time step of 2 fs. Temperature was maintained constant by Langevin dynamics with a collision frequency of 5 ps^{-1}, and pressure was maintained by a Monte Carlo barostat. Long-range electrostatics were treated by the Particle Mesh Ewald method and a non-bonded cutoff of 10 Å was used (44). The bond stretching freedom of hydrogens was fixed using the SHAKE algorithm to allow for a 2 fs timestep (45). The interatomic distances within the water molecules were fixed using the SHAKE algorithm. The RMSD and *b*-factor values for each residue were computed using the cpptraj module in the Amber suite (46).

Identification of druggable sites in RET1ΔNC

An early version of the upcoming POVME3.0 program was used to cluster the 2500 snapshots extracted with regular 100 ps intervals from RNA/UTP-bound RET1ΔNC simulations based on the active site shape (47). All 19 cluster representative conformations were then evaluated for druggability using the FTMap web server (48). FTMap floods the surface of a protein with 16 organic solvent probe molecules, and locates and ranks the most energetically favorable binding sites. Using a Python script, we then scored each residue based on the number of times it was in the 3.5 Å vicinity of docked organic probes for all 19 conformations.

Ensemble-averaged electrostatics of RET1ΔNC

The ensemble-averaged electrostatic potential of RET1ΔNC was calculated using 250 frames extracted at regular 1 ns intervals of the 10U/UTP-bound RET1ΔNC MD trajectories. For this purpose, a VMD plugin DelEnsembleElec interfacing with DelPhi, a numerical Poisson–Boltzmann-equation-solving suite, was utilized (49). The electrostatic potential was calculated on a 127 × 127 × 127 static grid with an interior dielectric constant of 2, an exterior dielectric constant of 80 and a salt concentration of 0.2 M. The grid scale was set to 1.0 Å and the probe radius defining the dielectric boundary was 1.4 Å. The convergence was achieved once the change of potential decreased below 0.0001 kT/e.

Enzymatic assays

To determine kinetic parameters for UTP and RNA substrates, the assays were carried out in 10 μl of 50 mM HEPES, pH 8.0, 15 mM KCl, 1 mM DTT, 10 mM $Mg(OAc)_2$, 0.01–25 μM of RNA and (α-^{32}P)UTP (0.5–50 μM, ∼4000–10 000 cmp/pmol). The 30-mer RNA substrate terminating with 12Us was chemically synthesized and purified on a 15% acrylamide/8 M urea gel (5′GCUAUGUCUGUCAACUUGUUUUUU UUUUUU3′). Reactions were started by addition of enzyme to 25 nM (or 50 nM for RET1ΔNC-RRR) and incubated at 27°C for 1 to 10 min. For RNA substrate parameters, the UTP concentration was kept constant at 25 μM (∼10 000 cpm/pmol) and protein was added to 25 nM. Reactions were stopped with an equal volume of 10% trichloroacetic acid (TCA) and 1% sodium pyrophoshate solution and the product were transferred onto Whatman 3MM paper (1.5 × 1.5 cm), washed three times per 15 min in 100 ml of ice-cold 10% TCA and 1% sodium pyrophosphate solution, washed once with 95% ice-clod ethanol, dried and counted in a liquid scintillation cocktail (EcoLite (+), MP Biomedicals). The apparent K_m and catalytic rate k_{cat} were obtained by fitting the initial velocities as a function of substrate concentration from three replicate experiments into a standard Michaelis–Menten kinetic model (Supplementary Figure S2). The SigmaPlot Enzyme Kinetics software package was used for calculations. For gel-based assays, reactions were performed in the same buffer with 0.1 or 1 μM of 5′-labeled 12U RNA, 25 μM of UTP and 25 nM of the enzyme. Reactions were stopped by adding 2 volumes of 5 mM EDTA, 95% formamide. The products were separated on a 15% acrylamide/ 8 M urea gel and exposed to a phosphor storage screen.

Trypanosome cell culture, RNAi and RET1 overexpression

RET1 RNAi expression plasmid was generated by cloning a gene fragment (444–1025 bp) into a p2T7-177 vector that allows for tetracycline-inducible knockdown (17). The construct was transfected into a procyclic 29-13 *T. brucei* strain (50), followed by clonal selection of phleomycin-resistant cells. The N-terminal mitochondrial importation signal of MRP1 (gBP21, positions 1–20, (51)) was added to the truncated versions of the RET1 gene, which were then cloned as C-terminally TAP-tagged fusions into a pLew79-based vector (52). Expression of RET1ΔC, RET1ΔN and RET1ΔNC was verified by Western blotting with peroxidase-anti-peroxidase (PAP) reagent (Sigma) (53).

RESULTS

Definition and crystallization of RET1 catalytically active core

Repeated attempts to crystalize the full-length polypeptide did not yield diffracting crystals, which led us to explore the crystallization and enzymatic properties of shortened RET1 variants. We reasoned that the N-terminal 188 amino acids are divergent even among RET1 orthologs in kinetoplastid protozoans while deletion of the zinc finger domain (positions 193–217, *T. brucei* numbering) or zinc chelation inactivated RET1 from *L. tarentolae* (33). Likewise, the highly conserved PAP-associated domain (Pfam03828) shared between TUTases and non-canonical poly(A) polymerases extends to approximately position 664. Analysis of the remaining C-terminal region (664–976) showed its high similarity to RET1 sequences from other trypanosomatids with the most-conserved part extending to position 699 (Figure 1A). The disorder predictions from the web portal PHYRE2 also indicated that fragments 1–188 and 699–976 are likely to be flexible (54). Therefore, we expressed residues 189–699 (RET1ΔNC) as a N-terminally GST-tagged protein in bacteria and purified the target polypeptide to apparent homogeneity while removing the GST tag in the process. RET1ΔNC was crystallized by the sitting-drop vapor diffusion technique with multiple rounds of re-seeding. The structure of the ligand-free protein was solved to 1.6 Å resolution using the model for TUT4 TUTase from *T. brucei* (24). To eliminate potential UTP polymerization activity (33) while preserving nucleotide triphosphate binding in the presence of Mg^{2+} ions, we next determined structures of the enzyme bound to a non-hydrolysable nucleotide UMPnPP, and the catalytically inactive mutant D473A bound to UTP to 1.8 Å and 2.2Å resolution, respectively. Three regions (residues 369–380, residues 624–643 and the C-terminal residues 689–699) missing in the high-resolution electron density maps are likely to be highly flexible. These atomic models allow the definition of three domains: the zinc finger, the enzymatic bi-domain module and the auxiliary fold, previously termed middle domain (22), which resembles the RNA recognition motif (RRM, Figure 1B–E). Interestingly, the zinc finger together with N-terminal residues (190–260) wraps around the C-terminus of the truncated protein, suggesting that the deleted extension encloses the minimal region required for folding and stability of the RET1 core region (Figure 1B).

Domain organization and the mechanism of UTP recognition

The Pol-β DNA polymerase-like catalytic metal binding and UTP-binding domains are separated by a deep groove, at the bottom of which the uridine triphosphate binds via water-mediated hydrogen bonds and direct interactions with conserved residues (Figures 1C and 2A). The nucleotide base is stabilized within the UTP-binding domain by a π-stacking interaction with tyrosine at position 524

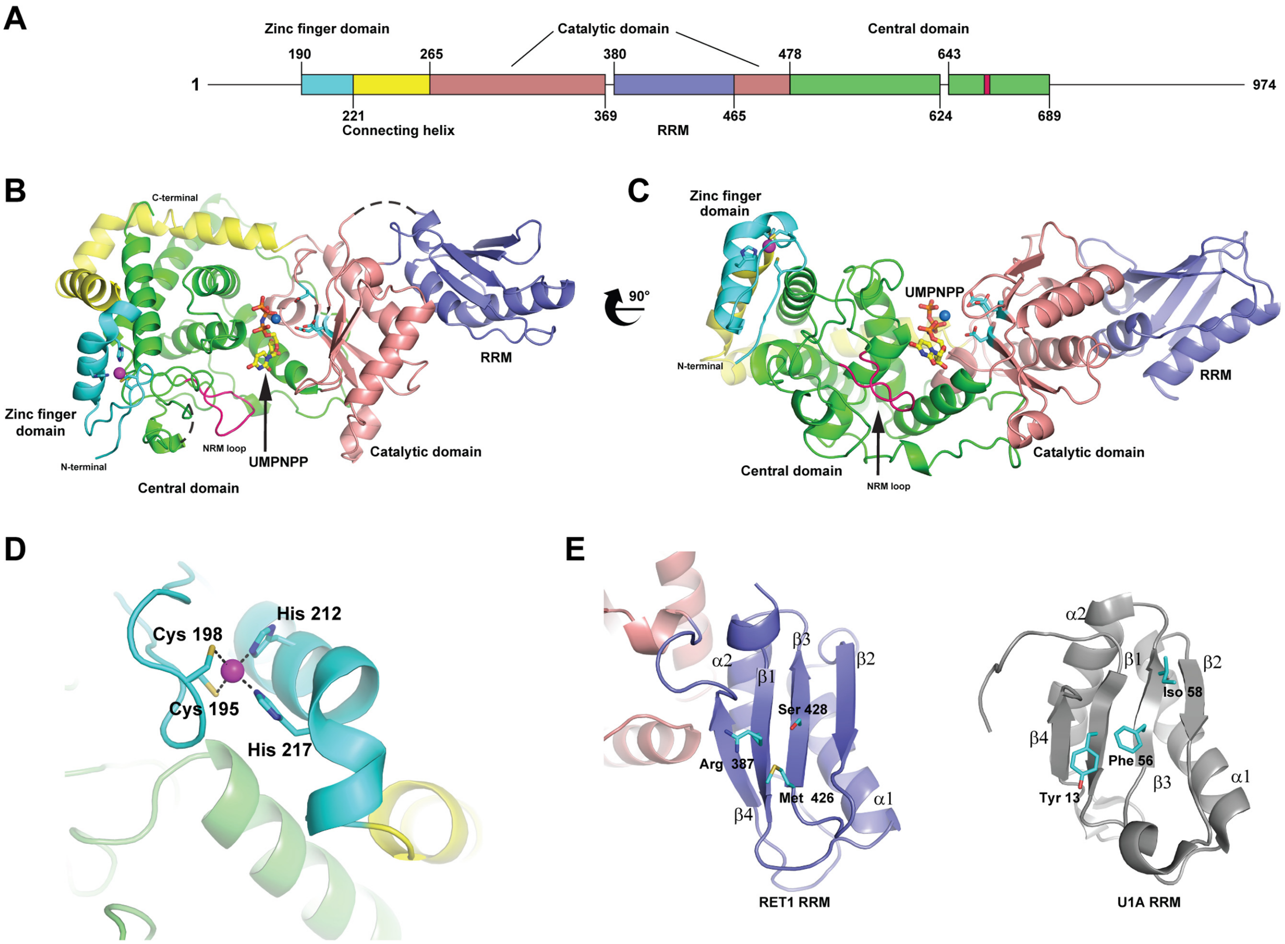


Figure 1. Overall view of the RET1ΔNC atomic structure. (**A**) Linear representation of the RET1 protein with the domains labeled and colored: cyan, zinc finger; yellow, connecting helix; salmon, catalytic domain; purple, middle domain (RRM); green, central domain. (**B**) The atomic structure of RET1ΔNC is represented in cartoon-style with domains colored as in panel A. (**C**) Same as A rotated by 90°. Two missing loops are shown as dashed lines. The non-hydrolyzable UMPnPP compound is shown in stick-style colored according to atom type (yellow, carbon; red, oxygen; blue, nitrogen; gold, phosphate). (**D**) Close view of the zinc finger domain present in the RET1 protein. The zinc binding residues are shown as stick representations and colored according to atom types (cyan, carbon). (**E**) Close view of the middle domain found in RET1. It adopts a fold that resembles the U1A RNA recognition motif. Secondary structure elements are labeled and key residues known to interact with nucleic acid are shown as stick representations and colored according to atom types. Magenta and dark blue spheres represent zinc and magnesium ions, respectively. The molecular graphic figures were generated using PyMOL software (https://www.pymol.org).

and hydrogen bonding with Asn 483 (Figure 2A and B). In addition, glutamic acid 657 in the nucleotide recognition motif (NRM, residues 652–661) and threonine 522 coordinate water molecules to engage positions 3 and 4 of the base, respectively. UTP-bound RET1 models show that the metal-binding triad, aspartic acid residues 310, 312 and 473, is confined to the catalytic domain (Figure 2A). Globally, the bi-domain organization of the active center in which the base recognition and catalysis are spatially separated, is similar to previously studied TUTases (Figure 2C). Several uridylyl transferases and related non-canonical poly(A) polymerases have additional domains either as insertion within the catalytic domain (22), or outside the bi-domain module (55) (Figure 2C). RET1 contains an additional domain which folds like an RRM of typical βαββαβ topology, but lacks the sequence signature of RNA-binding RRM (Figure 1E, (56)). The RRM domain adheres to the enzymatic module through a large hydrophobic interface of 637 Å^2. Although a similar organization is found in the RET2 TUTase, the specific orientation of its middle domain relative to its catalytic and UTP binding domains is different, possibly reflecting distinct RNA binding properties of the two proteins (Figure 2D). Indeed, the preferred substrate of RET1 is a ssRNA, while RET2 recognizes dsRNA formed by gRNA–mRNA hybridization (57,58).

Potential contributions of zinc finger domain to RET1 folding and activity

The presence of a C2H2 zinc finger adjacent to the catalytic domain (Figure 1A and D) is specific to trypanosomal RET1 proteins (59). However, multiple zinc fingers (C2H2 and CCHC types) are found in human ZCCHC6 (TUT7) or ZCCHC11 (TUT4) TUTases. Moreover, it has been shown that one of the C2H2 zinc finger motifs contributes to the Lin28-enhanced uridylation activity of human ZCCHC11 TUTase (32). The zinc finger deletion or chelating tightly-

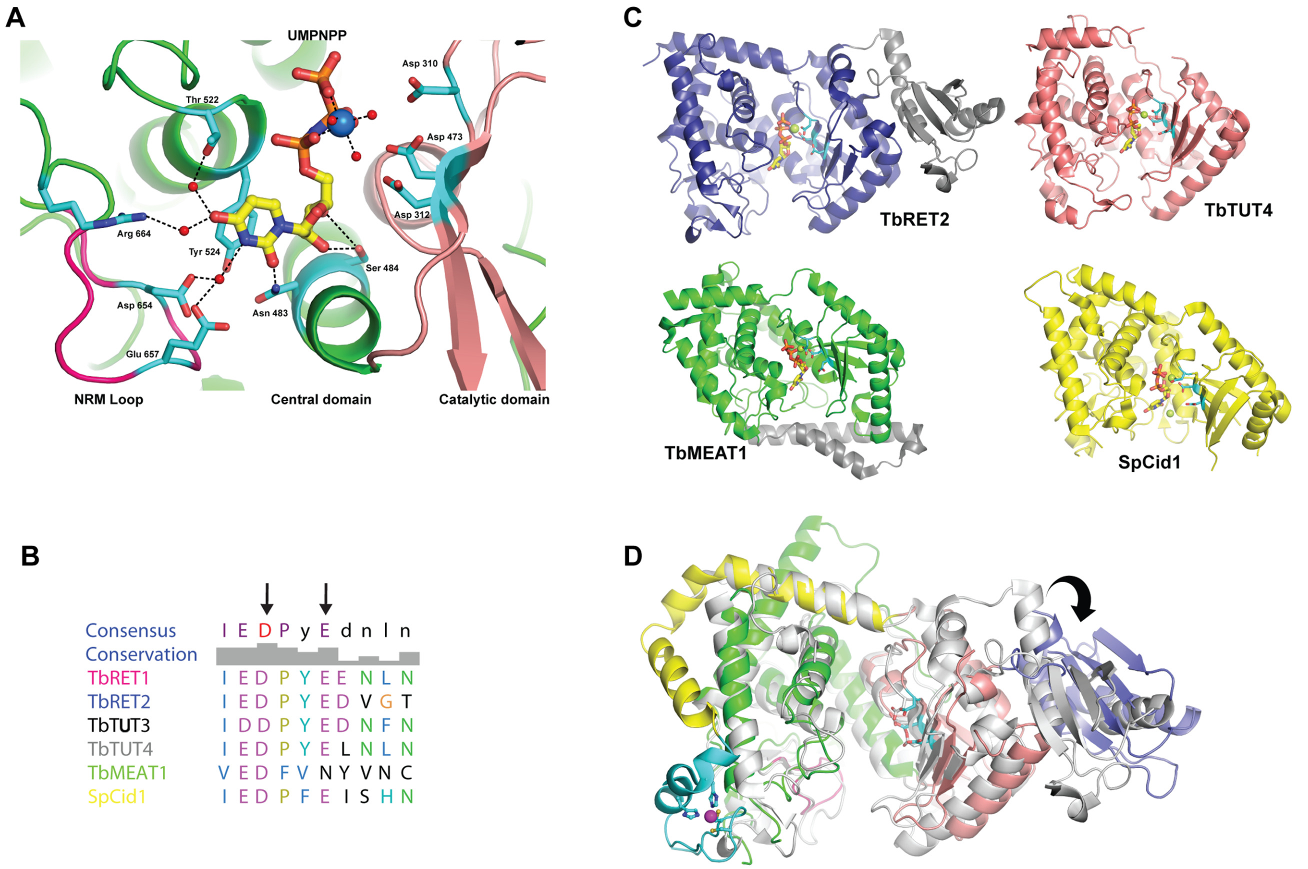


Figure 2. UTP binding and structural similarity. (**A**) Close view of the UMPnPP association in the active site of RET1ΔNC. Key residues are shown as stick representations and colored as in Figure 1. Hydrogen bonds are represented by black dashed lines. Domains are colored as in Figure 1. (**B**) Sequence alignment of the nucleotide recognition motif (NRM) found in trypanosomal and fungal TUTases. The two critical residues Asp 654 and Glu 657 are indicated by arrows. (**C**) Atomic models of other poly(U) polymerases are shown in the same orientation as RET1 in Figure 1. The nucleotidyl transferase centers are shown in color while additional domains are shown in gray. (**D**) The structure of RET2 is closely related to that of RET1. The catalytic domain of RET2 (PDB code 2BV4, shown as gray cartoon) is superimposed onto the RET1ΔNC structure, which is colored as in Figure 1. The relative movement of the RRM is indicated by an arrow.

bound Zn^{2+} ions inactivated RET1 from *Leishmania tarentolae*, but the exact function of this domain remained unclear (33). In an attempt to reveal the zinc finger's role, we superimposed the RET1 structure onto a typical C2H2 zinc finger domain bound to dsDNA using the zinc finger protein ZIF268 complex structure as the template (60). The dsDNA tightly fits the RET1 protein surface with almost no clashes (Figure 3A, RMSD of 0.93 Å between the RET1 zinc finger and the zinc finger 2 of ZIF268). Moreover, several arginine and histidine residues (Arg 228, Arg 615, His 601) are found in the vicinity of the zinc finger at hydrogen-bond distances between their side chain atoms and the phosphodiester backbone of the docked dsDNA (Figure 3B). Such amino acid residues often mediate nucleic acid contacts suggesting that the zinc finger may be involved in RNA binding (Figure 3B). Single R228A and H601D or, double R228A-H601D mutations were introduced into truncated RET1ΔNC background along with C195A–C198A substitutions designed to disrupt zinc coordination. We have been able to purify all, but the zinc binding-impaired proteins to apparent homogeneity (Figure 3C) and compare their processivity profiles to that of the full-length RET1 (Figure 3D). The C195A–C198A mutant was mostly insoluble and refractory to purification, which indicates inability to reach a stable folded conformation (not shown). Although the terminal deletions clearly reduced RET1's processivity, the mutations in the putative dsRNA binding groove exerted no effect on UMP additions to a single-stranded synthetic 30-mer. Indeed, current model of gRNA processing predicts that double-stranded RNA recognition is responsible for mitochondrial processome pausing at a fixed distance from the duplex formed by overlapping transcripts (16). This positions zinc finger as plausible sensor of impending dsRNA region rather than a significant ssRNA binding determinant. To test RET1 activity on a dsRNA that resembles a putative gRNA processing intermediate (double-stranded region with 10–13 nucleotide 3′ overhang (16)), we assembled respective substrates and analyzed reaction products in a high-resolution assay. RET1 enzymatic activity was virtually indistinguishable on ss- or dsRNA with an overhang, while fully dsRNA was unable to support UMP incorporation (Figure 3E). In sum, the zinc finger appears to be indispensable for folding of the catalytic core, but makes little, if any, contribution to uridylation of ssRNA substrate. Further experiments would

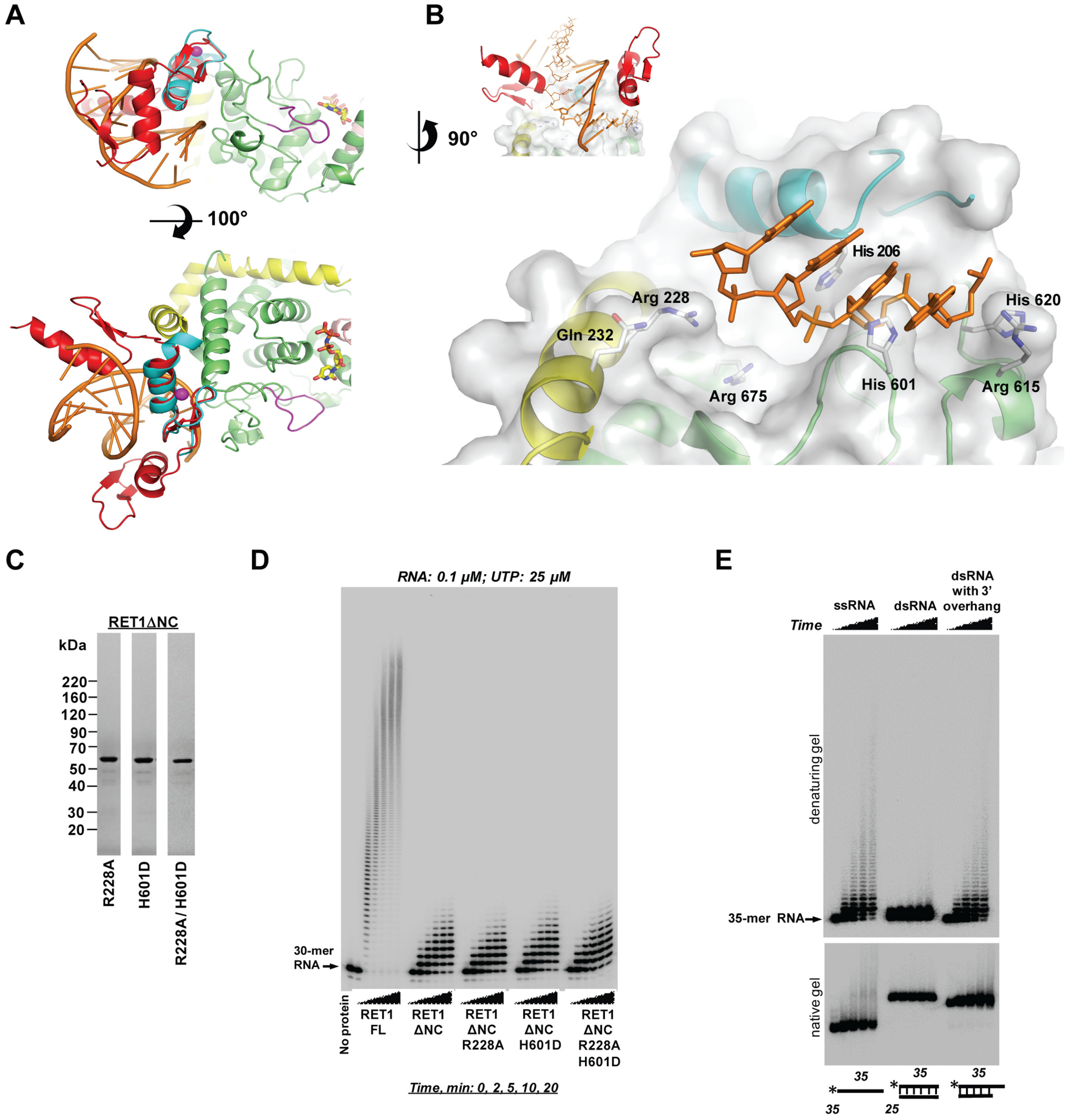


Figure 3. Superposition of the ZIF268/dsDNA complex onto the RET1ΔNC zinc finger domain. (**A**) The second zinc finger of the protein ZIF268 is superimposed onto the zinc finger domain of RET1ΔNC and shown as a red-colored cartoon. The lower panel is a 100° rotation of the superimposition. The ZIF268-bound dsDNA is shown as gold cartoon. (**B**) Four nucleotides lining the RET1 surface are shown as gold-colored stick representations to highlight the absence of clashes and the presence of nearby side chains potentially defining a nucleic acid binding surface of RET1. RET1ΔNC is shown as a cartoon with a transparent surface. Key residues are labeled and shown as stick representations (atoms are colored as in Figure 1; white, carbon). Inset: Overall view of the ZIF268/dsDNA superimposed onto RET1ΔNC, rotated by 90° relative to panel A. Proteins are shown as cartoon and colored as in panel A. DNA is shown as a cartoon and stick representations and colored in gold. (**C**) Purification of truncated RET1 variants bearing mutations in zinc finger domain. Final fractions were separated on 8–16% SDS gel and stained with Sypro Ruby. (**D**) Effects of terminal truncations and mutations in the zinc finger domain on RET1 activity with ssRNA substrate. (**E**) RET1 activity on dsRNA substrates. Single-stranded 5′ radiolabeled 35-mer RNA was folded in reaction buffer, or annealed to a fully-complimentary molecule, or to a 25-mer as diagramed. The resultant ssRNA, dsRNA and dsRNA with 10 nt-long 3′ overhang substrates were incubated with full-length RET1 for 0, 2, 4, 10 or 20 min under the same conditions as in panel A. Reaction products were separated on a denaturing 15% acrylamide/8 M urea gel (upper panel) or native 7% acrylamide gel (lower panel).

be required to fully elucidate its function during gRNA processing.

Oligomerization and processivity determinants

Size-fractionation of the recombinant 120-kDa RET1 from *L. tarentolae* established a molecular mass in the ∼400 kDa range (7,33), while in mitochondria of *T. brucei* this TUTase constitutes an integral subunit of the ∼900 kDa MPsome complex responsible for gRNA processing (16). To assess the oligomeric state of RET1 from *T. brucei*, the full-length protein, the N-, the C- and the NC-terminally truncated variants were expressed and purified (Figure 4A and B). To determine an apparent molecular mass for each protein, final fractions were subjected to size exclusion chromatography along with a set of molecular mass standards (Figure 4C). The resultant values indicated that full-length and C-terminally truncated *T. brucei* proteins ran with a higher apparent molecular weight (between a dimer and trimer), which indicates an assembled multimeric enzyme (Figure 4D). However, the N-terminal deletion, whether combined with the C-terminal deletion or not, produced a near-perfect match between the determined apparent molecular mass and the predicted mass for a monomeric protein (Figure 4C and D). These data implicate the poorly conserved N-terminal extension in RET1 oligomerization.

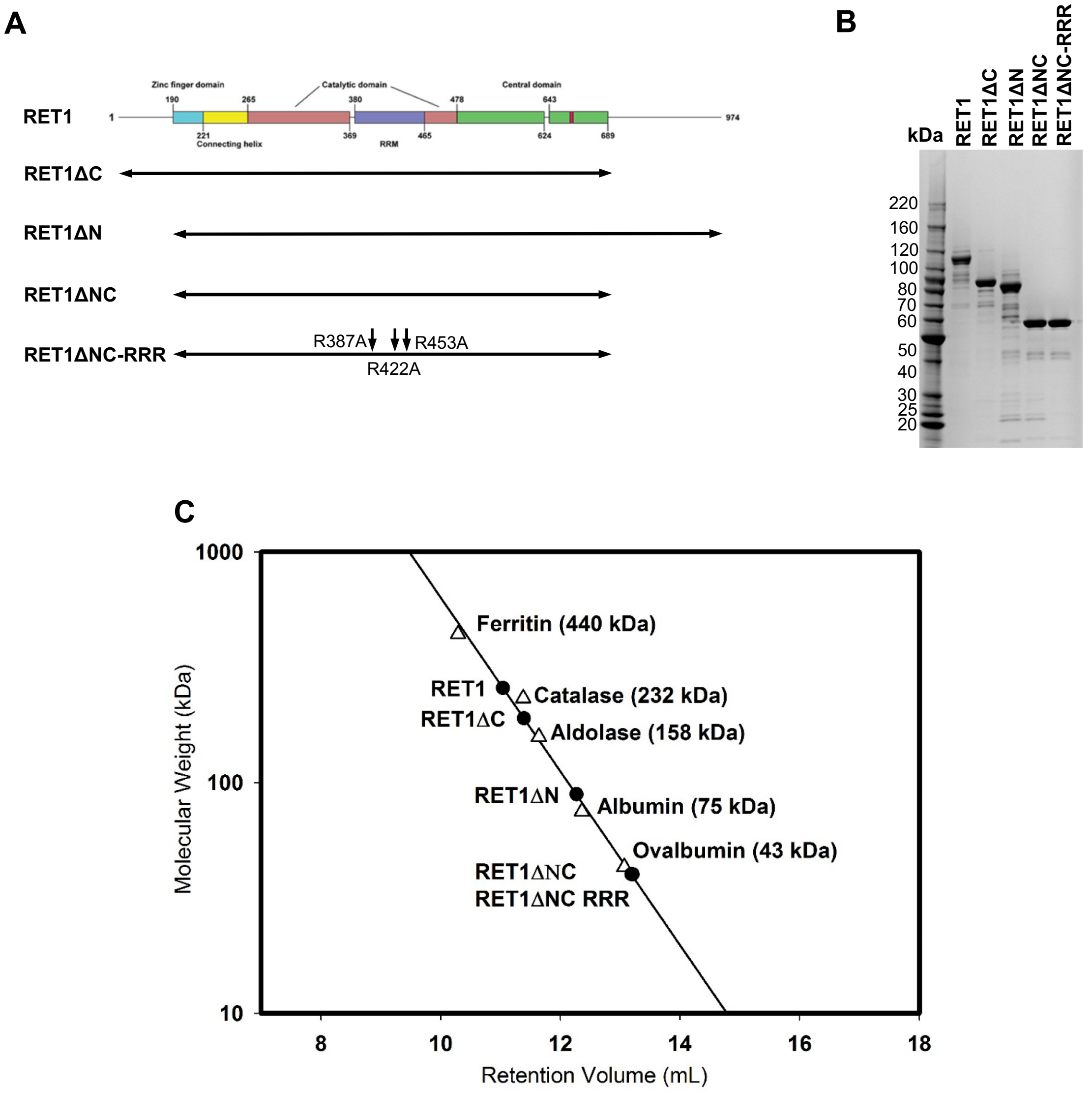
The key amino-acid residues responsible for metal binding, UTP recognition and positioning of the RNA's 3′-end are structurally conserved between RET1 and TUT4, the smallest naturally occurring TUTase composed of only catalytic and UTP-binding domains (24,25). Remarkably, these features are also mostly conserved between trypanosomal TUTases and Cid1 from *S. pombe* (Figure 2), which displays dual specificity toward UTP and ATP (61). TUT4 and Cid1 were extensively investigated by structure-guided mutagenesis and there is a high likelihood that these positions are functionally conserved in RET1 as well (24,62). Conversely, the crystalized catalytic core displays much lower processivity than the full-length RET1 (Figure 3D). Therefore, we next assessed specific contribution of terminal extensions to RET1 enzymatic activity. Because the middle domain adopts an RRM-like fold, we also explored the possibility of its involvement in RNA binding and generated a triple mutant by replacing the arginines in positions 387, 422 and 453 with alanines in the RET1ΔNC (RET1ΔNC-RRR). These mutations exerted no appreciable impact on its solubility or size fractionation properties (Figure 4B and C). A high-resolution uridylyl transferase activity assay in the presence of 5′ labeled synthetic RNA and UTP revealed a significantly reduced processivity of RET1 variants missing the C-terminal extension, while the N-terminally truncated protein was marginally more active than the full-length RET1. The triple mutation further reduced the activity of the RET1ΔNC variant (Figure 5A). Although the C-terminal truncation did not affect the putative oligomerization of RET1, it likely decreased the RET1's capacity to bind RNA substrate. In a primer challenge experiment, we next asked if an increase in RNA substrate concentration could proportionally affects processive and distributive patterns (Figure 5B). Clearly, an excess of RNA primer inhibited processive UMP incorporation by full-length and N-terminally truncated variants, while the activity of C-terminally shortened proteins remained unaffected. Finally, we tested whether trypanosomal RET1 possesses the type of RNA-independent UTP polymerization activity, prominently displayed by its *L. tarentolae* ortholog (33). Surprisingly, no UTP polymerization was detected by incubating the full-length trypanosomal enzyme with increasing concentrations of (α-^{32}P)UTP, whereas a synthetic RNA substrate was efficiently utilized (Figure 5C).

To quantitatively examine the correlation between the N- and C-terminal deletions and changes in RET1's catalytic activity, we determined the K_m, k_{cat} and catalytic efficiency of all RET1 variants for both UTP and RNA substrates (Table 2). As could be expected from the unperturbed folding of the catalytic and central domains, the UTP binding reflected by apparent K_m values and the catalytic efficiency (calculated as turnover number divided by K_m), were affected by less than 2-fold in all mutants. In agreement with the gel-based assay, the N-terminally-truncated RET1 showed increased efficiency of RNA substrate utilization due to its lower K_m (Figure 5A and Table 2). Accordingly, deletion of the C-terminal extension, which is conserved among RET1 orthologs, but is missing from non-processive TUTases such as RET2 (9), TUT4 (24) and MEAT1 (23,63), significantly increased the K_m for the RNA substrate and reduced overall catalytic efficiency by a factor of 20 (Figure 5A and Table 2). Alanine substitutions of arginine residues lining the protein surface toward the putative RRM fold resulted in an approximately 5-fold K_m increase for the RNA substrate as compared to RET1ΔNC. In agreement, the gel-based assay also demonstrated that arginine mutations reduce RET1 catalytic efficiency (Figure 5A).

Overall, biochemical fractionation experiments and enzymatic activity assays established that the N-terminal extension enables RET1 oligomerization, but exerts only a marginal influence on its enzymatic properties. On the contrary, evidence points toward a major contribution of the C-terminal RET1-specific extension to RNA binding and, therefore, efficiency of U-additions.

In vivo functionality of RET1 domains

Considerable evidence suggests that, in mitochondria, the bulk of RET1 protein is sequestered into an ∼900 kDa MPsome complex along with DSS1 3′-5′ exonuclease, and three large proteins lacking annotated motifs (16). Furthermore, it appears that RET1-catalyzed uridine addition does not exceed 15–20 nt *in vivo*, while hundreds of UMP residues can be incorporated *in vitro* (Figure 3). Finally, overexpression of catalytically inactive protein bearing a point mutation in the metal binding site produced a strong dominant negative effect similar to protein downregulation by RNAi (17). Collectively, these findings indicate that incorporation into the MPsome is required for RET1 functions in gRNA processing. To test whether terminal extensions support RET1 incorporation into the MPsome, the RET1ΔC, RET1ΔN and RET1ΔNC gene variants were placed under PARP promoter as C-terminally TAP-tagged proteins in pLew79-MHTAP vector (52). To ensure equal mitochondrial import efficiency, all proteins



D

	Theoretical MW, monomer (kDa)	Retention volume (mL)	Apparent MW (kDa)
RET1	107.6	11.04	257
RET1ΔC	77.4	11.39	190
RET1ΔN	87.6	12.27	89
RE1ΔNC	57.4	13.19	40
RE1ΔNC-RRR	57.4	13.21	40

Figure 4. Size fractionation of RET1 deletion mutants. (**A**) Schematic representation of deletion and site-directed mutagenesis. (**B**) Purified recombinant proteins were separated on a 8–16% gradient SDS gel, stained with Sypro Ruby florescent dye and visualized using an ImageQuant Imager (GE Life Science). (**C**) Retention volumes of RET1 mutants obtained by gel filtration chromatography on a Superose 6 column are plotted against molecular mass standards. (**D**) Theoretical molecular masses of monomeric RET1 derivatives and their experimentally determined apparent molecular masses.

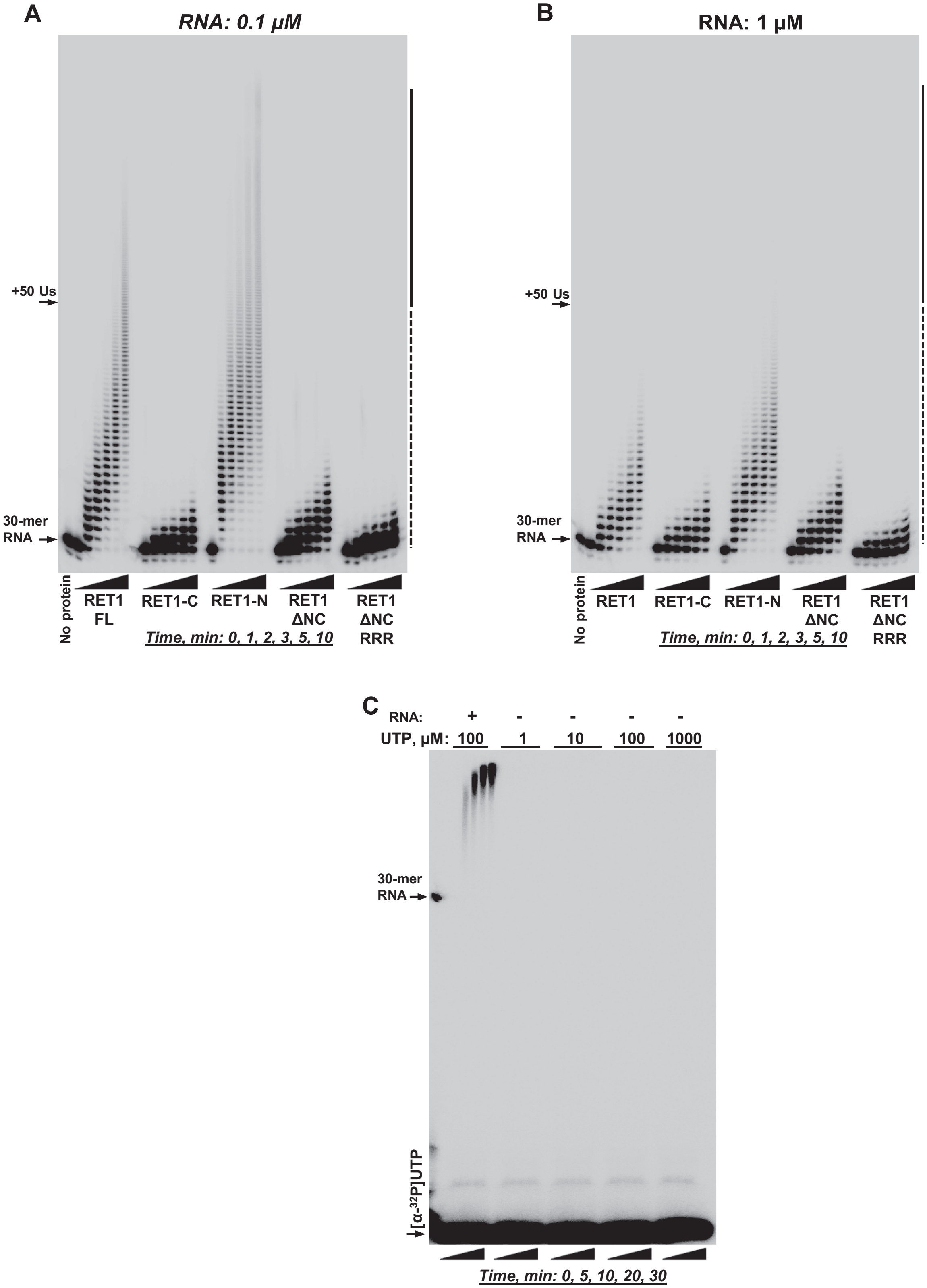


Figure 5. Processivity of UTP polymerization by RET1 mutants. (**A**) RNA uridylyl transferase assays were set up with 0.1 μM of 5′ radiolabeled synthetic RNA, 50 μM of UTP and 5 nM of indicated purified recombinant protein. Reaction products were separated on 12% acrylamide/8 M urea gel and visualized by exposure to a phosphor storage screen. (**B**) The same reactions as in panel A with 1 μM of RNA substrate. (**C**) UTP polymerization reaction in the absence of RNA substrate. Full-length RET1 protein was incubated with increasing concentrations of UTP and 10 nCi/μl of (α-^{32}P) UTP to assess its capacity to polymerize UTP without an RNA primer. A synthetic RNA 30-mer was used as a positive control to demonstrate that the enzyme was active.

Table 2. Catalytic parameters of RET1 deletion mutants

	UTP				RNA			
Mutants	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}\mu M^{-1}$)	Relative k_{cat}/K_m	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}\mu M^{-1}$)	Relative k_{cat}/K_m
RET1	8.5 ± 1.3	15.5 ± 0.8	1.8	1	0.53 ± 0.1	23.5 ± 1.7	44.3	1
RET1ΔC	9.8 ± 2.5	22.7 ± 2.1	2.3	1.3	12.4 ± 1.7	29.7 ± 2.1	2.3	0.05
RET1ΔN	11.1 ± 0.8	16.7 ± 0.5	1.5	0.83	0.024 ± 0.006	18.3 ± 1.1	763	17.2
RET1ΔNC	6.6 ± 0.5	23.5 ± 0.6	3.6	2.0	4.5 ± 0.8	32.5 ± 2.0	7.2	0.16
RET1ΔNC RRR	6.1 ± 1.5	9.5 ± 0.7	1.6	0.89	22.6 ± 13.7	19.2 ± 5.3	0.85	0.019

were fused with heterologous mitochondrial importation peptide from a known mitochondrial protein MRB2 (64). Proteins were expressed under *tet*-repressor control in the actively respiring procyclic (insect) developmental form of *T. brucei* (Figure 6A). The mitochondrial fraction was isolated from parental (Figure 6B and C) and respective transgenic (Figure 6D–F) cell lines, extracted with detergent and separated on glycerol gradients. Each fraction was further analyzed by native and denaturing PAGE. The endogenous RET1 was detected by immunoblotting with monoclonal antibody (Figure 6B) while truncated proteins were visualized with antibody against calmodulin binding peptide in the TAP tag (Figure 6D–F). The well-characterized RNA editing core complex was also detected in the parental cell line as an internal separation efficiency control (Figure 6C). In agreement with a previous study (16), most of the endogenous RET1 sedimented in gradient fractions 3–5 that correspond to the ~19S (900 kDa) MPsome complex while a minor fraction extended into larger high molecular mass complexes (fractions 6–12; Figure 6B). The RET1ΔC and RET1ΔN variants were found exclusively in MPsome-like high molecular mass complexes while the polypeptide missing both N- and C-terminal extensions (RET1ΔNC) was evenly distributed between unincorporated (fraction 2) and MPsome-bound (fractions 4–6) forms. These results suggest that the catalytic core lacking terminal extensions may still bind to the MPsome. However, it appears that the N-terminal extension, which is dispensable for activity, but required for oligomerization (Figures 4 and 5, Table 2), may also control the stoichiometry of RET1 subunits within the MPsome.

Because of diametric effects of N- and C-terminal deletions on RET1 oligomerization state (Figure 4) and enzymatic activity (Figure 5, Table 2), and apparent capacity of respective deletion mutants to enter the MPsome complex (Figure 6D–F), we next investigated whether overexpression of truncated TUTase variants would cause a dominant negative effect. Surprisingly, the division rates declined not only in cell lines conditionally expressing mutants with compromised enzymatic activity (RET1ΔC and RET1ΔNC), but also in cells expressing the fully-active RET1ΔN mutation which converts recombinant RET1 into a monomer. Thus, it seems plausible that RET1 stoichiometry in the MPsome is functionally important albeit further work is required to determine the structural basis of the observed phenotypes.

Molecular dynamics simulations of RNA binding by RET1

To gain further insight into RNA binding determinants adjacent to the catalytic center and develop a foundation for rational inhibitor design, we simulated RNA primer bounding to the UTP-bound RET1 using explicitly solvated all-atom molecular dynamics simulations. To this end, we first solved the structure of TUT4 from *T. brucei* with bound UTP and RNA-mimicking dinucleotide UpU (PDB ID 5KAL). Then, we reintroduced the amino acids 370–379, 625–642 and 689–693 that were disordered in the RET1ΔNC crystallographic structure. Next, we simulated the complete RET1ΔNC structure (residue 189–698) in apo and UTP-bound forms, exploring the dynamics of the systems. The RET1ΔNC crystal structure has an active site conformation that does not appear capable of binding larger RNA segments, thus in the simulations, we monitored the dynamics of the active site area and selected the RET1ΔNC conformation with the most open active site. This 'open, relaxed structure' was used to develop a model that included a 10 uracil-long RNA molecule; this 10-U strand was positioned onto the UTP-bound RET1 such that it mimicked the terminal and penultimate uridine nucleosides from the UTP/UpU-bound TUT4 structure (Figure 7A). This ternary configuration was further supported by the excellent correlation observed between the model and crystal structure of Cid1 protein from *S. pombe* bound to ApU (26). The 10-mer RNA proved to be very dynamic during MD simulations, except for the two terminal U nucleotides while the protein turned out to be highly stable as shown by the B-factors values (Figure 7B and C). The 3′ hydroxyl of the terminal U and the first phosphate O atom interact tightly with the Mg^{2+} ion. The terminal RNA U nucleotide base is stabilized by π-stacking interaction with the UTP nucleotide base from one side, and by van der Waals interactions with catalytic domain residues Arg 477 and Val 359 on the other side. Also, Arg 358 from the catalytic domain consistently forms a salt bridge with the second phosphate of the RNA primer in all simulations, in agreement with previously observed interactions between catalytic domain residues and RNA primers (25,26). The penultimate U nucleotide base forms van der Waals interactions with catalytic domain residues Ile 354 and Ala 357, and a dynamic hydrogen bond interaction with Lys 363. The ensemble-averaged electrostatic potential field of the protein (Figure 7D) indicates that the active site is the largest positively charged region, and our 10-U strand binds stably to this region (compare 7C and 7D). Moreover, further away from the active site, the RNA primer appears to line up on the RET1 surface toward the zinc finger domain rather

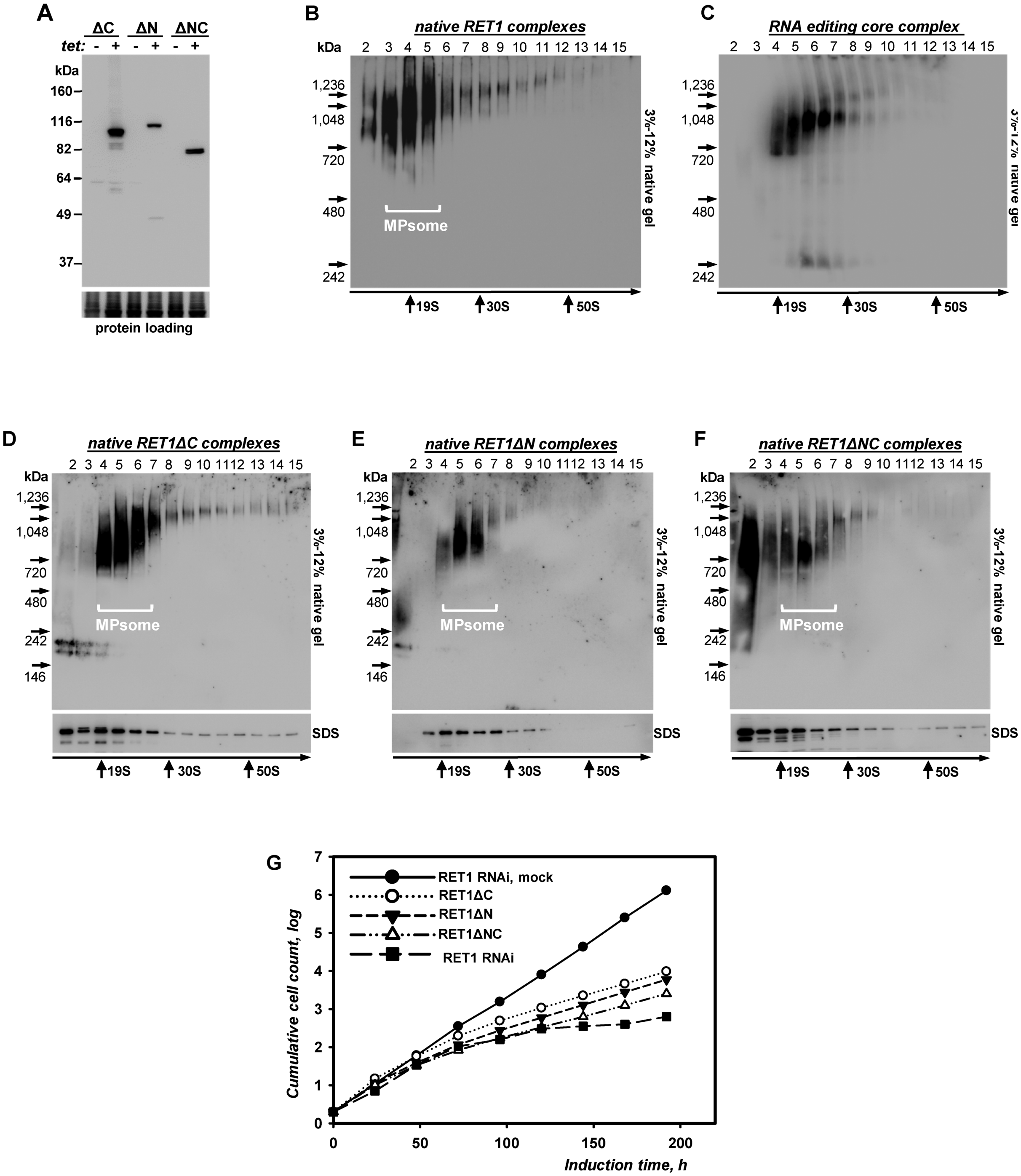


Figure 6. Functional definition of terminal extensions. (**A**) Stable cell lines carrying expression constructs were either mock-induced or grown in the presence of 1 mg/l of tetracycline for three days. Cell lysates were separated on an SDS gel and subjected to immunoblotting with peroxidase-anti-peroxidase (PAP) reagent to detect the C-terminal TAP-tag. (**B**) Mitochondrial fraction was extracted with detergent and soluble contents were separated for 5 h at 178 000 × *g* in a 10–30% glycerol gradient. Each fraction was further resolved on a 3–12% Bis-Tris native gel. Positions of native protein standards are indicated by arrows. Mitochondrial 3′ processome (MPsome) was visualized by immunoblotting with monoclonal antibody against RET1. Thyroglobulin (19S) and bacterial ribosomal subunits were used as S-value standards. (**C**) The well-characterized RNA editing core complex (RECC, ~1.2 MDa) served as an internal separation efficiency marker. RECC was exposed by self-adenylation of RNA ligases REL1 and REL2 in the presence of (α-^{32}P)ATP. (**D–F**) Mitochondrial fractions were isolated from cell lines expressing RET1ΔC, RET1ΔN and RET1ΔNC TAP-tagged fusion proteins for 72 and analyzed as in panel B. Tagged proteins were detected by immunoblotting with antibodies against calmodulin binding peptide. (**G**) Growth kinetics of parasite suspension cultures after mock induction, RET1 deletion mutant proteins induction and RET1 RNAi expression.

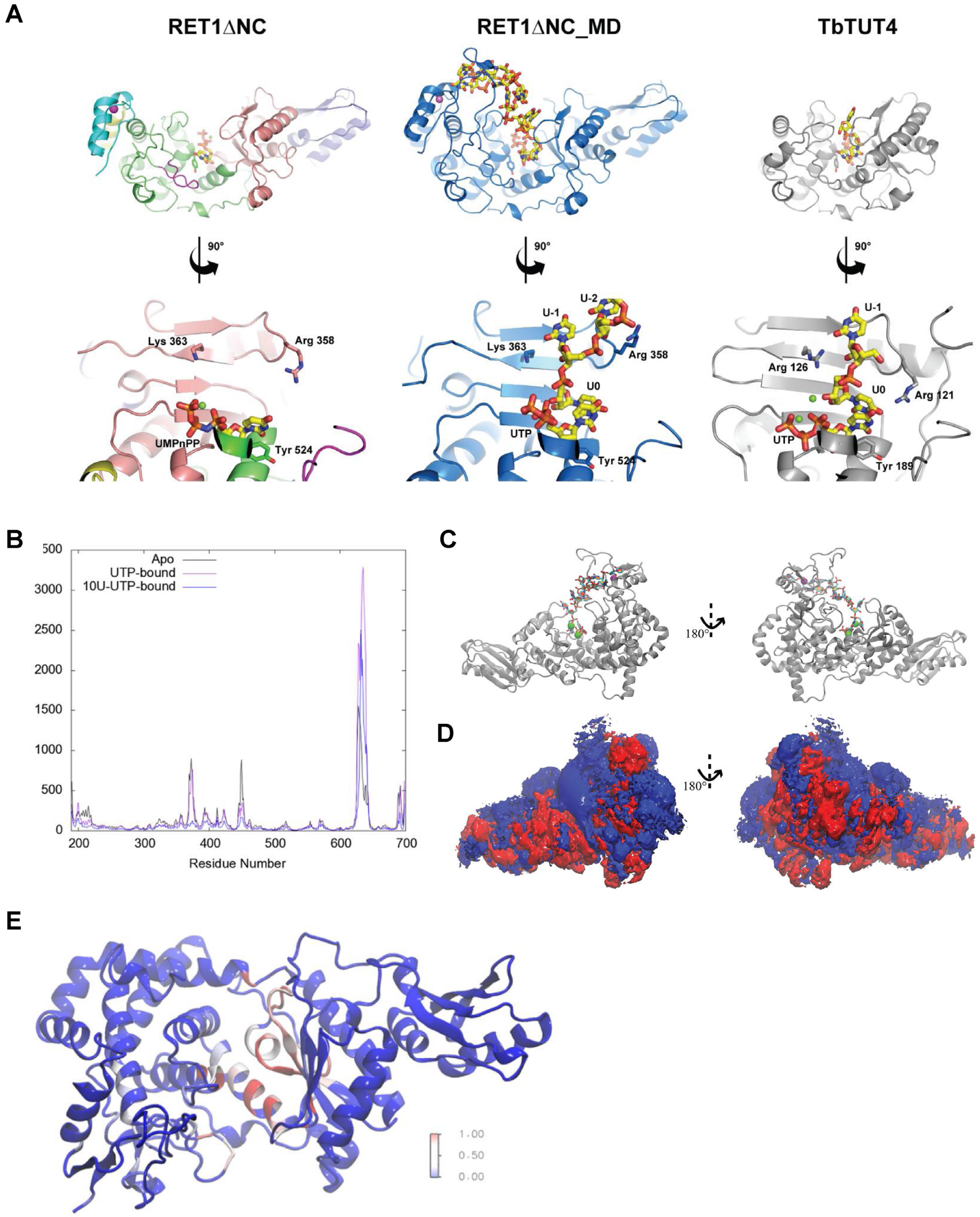


Figure 7. Molecular dynamics (MD) simulations of RNA in the active site and identification of potential druggable regions. (**A**) Overall and close-up view of the RNA binding site illustrating positions of terminal RNA residues in the RET1ΔNC bound to UTP, RET1ΔNC with modeled RNA primer (RET1ΔNC-MD) and TUT4-UTP-UpU. Proteins are depicted as colored ribbons. The RNA primer, UpU and UTP molecules are shown as colored stick representations, and the Mg^{2+} and Zn^{2+} ions are shown as green and orange spheres, respectively. (**B**) B-factors computed for the apo, UTP-bound and RNA primer-UTP- RET1ΔNC simulations. (**C** and **D**) Front and back views of the ensemble-averaged RNA primer –UTP-bound RET1ΔNC shown as stick and ribbon or as electrostatic potential field at positive (blue) and negative (red) contoured at 3 kT/e. (**E**) RET1 druggable sites identified by FTMap evaluation of 19 distinct MD-generated protein conformations. In the ribbon diagram, the color scale extends from blue (least druggable sites) to white to red (most druggable sites).

than toward the RRM (Figure 7C). Overall, however, the large degree of flexibility and persistent motion of the 10-U primer during the MD simulations, especially in the regions that are farther from the active site, indicate that some elements of binding are likely incomplete, suggesting either a non-optimal initial binding mode for the 10-U segment or the possibility that other structural elements not included here (e.g. the C-terminal extension) are required for stable RNA binding.

Designing trypanocides using the RET1 atomic structure

While targeting RET1 for a potential trypanocide, it is helpful to identify and focus on the most druggable areas at the active site. For inhibitor design, the information provided by the static crystal structure can be enriched by MD simulations. Monitoring the active site dynamics through MD, novel transient pockets and alternative conformations of the active site can be identified and integrated into the inhibitor design process. To capture a set of unique active site conformations of RET1ΔNC among the 25 000 variants generated by MD simulations, the POVME program was used to cluster the 10U/UTP-bound RET1 MD trajectories based on the active site shape (see Materials and Methods for details). The RET1 structure colored with respect to these scores illustrates that the most druggable hot spots are at the bottom of the active site cleft, and mostly lie on the RET1 catalytic domain alpha helices (Figure 7E).

DISCUSSION

Since key elements of the enigmatic mitochondrial RNA processing pathway in trypanosomes have begun to emerge, extensive investigation has focused on the adaptation of highly conserved catalytic folds to perform unconventional reactions. The RNA editing core complex provides examples for sequential endonucleolytic gRNA-directed mRNA cleavage, U-addition or removal from the exposed cleavage fragment and gRNA-dependent mRNA re-ligation. These mRNA editing cascade reactions are performed by RNase III-, DNA pol β-, EEP- and Type II RNA ligase-like domains, respectively. These enzymatic domains are embedded into a stable protein complex via trypanosome-specific modules. It appears that complex association may lead to altered enzymatic properties and offer the kinetic advantages of channeling the intermediates between active centers. Likewise, the recent discovery of the mitochondrial processome (MPsome) conceptualized earlier findings of pervasive RET1-catalyzed 3′ uridylation and its profound impact on the mitochondrial transcriptome. The MPsome is composed of RET1 TUTase, DSS1 3′-5′ exonuclease and three large proteins lacking annotated motifs. Remarkably, complex association limits intrinsically high RET1 processivity while enabling efficient RNA degradation by DSS1, an enzyme which is catalytically inactive as an autonomous protein (65). In the current model, uridylation of the initial gRNA precursor by the MPsome-embedded RET1 induces processive DSS1-catalyzed 3’-5’ degradation. Rapid RNA hydrolysis apparently empowers the melting of secondary structure elements along the precursor until the MPsome encounters a relatively stable duplex region formed by the gRNA precursor and the corresponding anti-sense transcript. The ensuing stochastic pausing is thought to enable the secondary uridylation event that disengages the MPsome from the mature gRNA 3′ end. In the latter reaction, RET1 is expected to recognize a 10–13 nt long 3′ overhang extending from the ∼30 nt double-stranded region (16). Therefore, it seems reasonable that *in vivo* the enzyme encounters at least two substrate types: extended single stranded RNA (gRNA precursors) and double-stranded regions with 3′ overhang (gRNA processing intermediates).

In this work, we combined X-ray crystallography, MD simulations and biochemical analysis to determine high-resolution structures of the RET1 catalytic core region, to investigate determinants of UTP and RNA substrate recognition and to study RET1 incorporation into the MPsome. Our data demonstrate a remarkable structural conservation of bi-domain enzymatic module formed by Pol-β DNA polymerase-like catalytic and UTP-binding domains among trypanosomal TUTases. These similarities extend into direct and water-mediated interaction networks that coordinate UTP recognition and RNA positioning in the active site. In contrast to currently available TUTase structures, the RET1 core architecture reveals a previously unseen combination of its enzymatic module with a zinc finger and an RRM domain inserted within the catalytic fold. Both of these modules are shown by point mutations to impact the enzymatic properties of the TUTase although their precise mode of interaction with nucleic acid will required additional studies.

Despite being predicted as disordered regions, RET1-specific terminal extensions apparently represent functional domains that participate in specific biological functions. We show that deleting the N-terminal extension prevents oligomerization of recombinant RET1 in solution without impacting enzymatic activity *in vitro*. However, incorporation of the N-terminally truncated RET1 into the MPsome triggers dominant negative phenotype suggesting a functional role of the N-terminal extension in the context of RET1’s native complex. Within the MPsome, RET1 directly interacts with DSS1 exonuclease and MPSS3 structural protein (16); it remains to be established whether putative RET1 oligomerization region participates in the MPsome complex assembly or determines the RET1 stoichiometry in the complex. On the contrary, the C-terminal deletion has minimal effect on MPsome formation *in vivo*, but greatly reduces RET1 processivity and results in a ∼20-fold lower catalytic efficiency due to an increased apparent K_m for RNA binding. This region’s contribution to RET1 function *in vivo* is, however, significant as indicated by a strong dominant negative phenotype upon RET1ΔC-inducible overexpression. We hypothesize that this phenotype is likely due to the RET1ΔC-impaired RNA binding capacity. Additional studies will be required to determine the structural basis of these phenotypic observations.

Homology modeling based on the TbTUT4-UTP-UpU structure solved in this work and molecular dynamics simulations allowed us to define the RNA binding determinants in the vicinity of the active site and propose potential druggable sites. Considering RET1’s participation in multiple essential RNA processing pathways in a parasite of significant medicinal and economic consequences, these results may lead to future design of RET1 inhibitors with trypanocidal activity.

Collectively, our study reveals a high-resolution structure of the catalytic core part of a critical TUTase from *Trypanosoma brucei*, illuminates some of its function-specific modules *in vitro* and *in vivo* and provides a sound foundation for rational inhibitor design.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the members of our laboratories for helpful discussions. We acknowledge the Paul Scherrer Institut, Villigen, Switzerland for provision of synchrotron radiation beamtime at beamline PXI and PXIII of the SLS and would like to thank V. Olieric for assistance. The atomic coordinates have been deposited in the Protein Data Bank under 5HZD (apo RET1), 5IDO (UTP-bound D473A), 5I49 (UMPnPP-bound) and 5KAL (UpU-UTP-bound TUT4). This crystallographic study was possible thanks to the biostructural platform which was financed via the generous support of the Boninchi foundation, the Schmidheiny foundation, the Swiss National Science Foundation R'equip grant (N°316030-128787) and the University of Geneva. R.E.A. is a co-founder of Actavalon, Inc.

FUNDING

Swiss National Science Foundation [31003A-140924, 31003A-124909 to S.T.]; National Institutes of Health [AI091914, AI101057 to R.A.]; Director's New Innovator Award Program NIH DP2 [OD007237 to REA]; National Biomedical Computation Resource through NIH [P41 GM103426 to R.E.A.]. Funding for open access charge: University of Geneva.

Conflict of interest statement. None declared.

REFERENCES

1. Bilbe,G. (2015) Infectious diseases. Overcoming neglect of kinetoplastid diseases. *Science*, **348**, 974–976.
2. Kramer,S. (2012) Developmental regulation of gene expression in the absence of transcriptional control: The case of kinetoplastids. *Mol. Biochem. Parasitol.*, **181**, 61–72.
3. Aphasizhev,R. and Aphasizheva,I. (2011) Mitochondrial RNA Processing In Trypanosomes. *Res. Microbiol.*, **162**, 655–663.
4. Aphasizheva,I. and Aphasizhev,R. (2015) U-Insertion/Deletion mRNA-Editing Holoenzyme: Definition in Sight. *Trends Parasitol.*, **32**, 144–156.
5. Aphasizhev,R. and Aphasizheva,I. (2013) Emerging roles of PPR proteins in trypanosomes: Switches, blocks and triggers. *RNA Biol.*, **10**, 1495–1500.
6. Aphasizhev,R. and Aphasizheva,I. (2011) Uridine insertion/deletion mRNA editing in Trypanosomes: A playground for RNA-guided information transfer. *Wiley Interdiscip. Rev.*, **2**, 669–685.
7. Aphasizhev,R., Sbicego,S., Peris,M., Jang,S.H., Aphasizheva,I., Simpson,A.M., Rivlin,A. and Simpson,L. (2002) Trypanosome mitochondrial 3′ terminal uridylyl transferase (TUTase): The key enzyme in U-insertion/deletion RNA Editing. *Cell*, **108**, 637–648.
8. Munoz-Tello,P., Rajappa,L., Coquille,S. and Thore,S. (2015) Polyuridylation in Eukaryotes: A 3′-End modification regulating RNA life. *Biomed. Res. Int.*, **2015**, 968127.
9. Aphasizhev,R., Aphasizheva,I. and Simpson,L. (2003) A tale of two TUTases. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 10617–10622.
10. Ernst,N.L., Panicucci,B., Igo,R.P. Jr, Panigrahi,A.K., Salavati,R. and Stuart,K. (2003) TbMP57 is a 3′ terminal uridylyl transferase (TUTase) of the Trypanosoma brucei editosome. *Mol. Cell*, **11**, 1525–1536.
11. Blum,B., Bakalara,N. and Simpson,L. (1990) A model for RNA editing in kinetoplastid mitochondria: 'Guide' RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell*, **60**, 189–198.
12. Ryan,C.M. and Read,L.K. (2005) UTP-dependent turnover of Trypanosoma brucei mitochondrial mRNA requires UTP polymerization and involves the RET1 TUTase. *RNA*, **11**, 763–773.
13. Kao,C.Y. and Read,L.K. (2005) Opposing effects of polyadenylation on the stability of edited and unedited mitochondrial RNAs in Trypanosoma brucei. *Mol. Cell. Biol.*, **25**, 1634–1644.
14. Etheridge,R.D., Aphasizheva,I., Gershon,P.D. and Aphasizhev,R. (2008) 3′ adenylation determines mRNA abundance and monitors completion of RNA editing in T. brucei mitochondria. *EMBO J.*, **27**, 1596–1608.
15. Aphasizheva,I., Maslov,D., Wang,X., Huang,L. and Aphasizhev,R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. *Mol. Cell*, **42**, 106–117.
16. Suematsu,T., Zhang,L., Aphasizheva,I., Monti,S., Huang,L., Wang,Q., Costello,C.E. and Aphasizhev,R. (2016) Antisense transcripts Delimit Exonucleolytic activity of the mitochondrial 3′ processome to generate guide RNAs. *Mol. Cell*, **61**, 364–378.
17. Aphasizheva,I. and Aphasizhev,R. (2010) RET1-catalyzed Uridylylation Shapes the Mitochondrial Transcriptome in *Trypanosoma brucei*. *Mol. Cell. Biol.*, **30**, 1555–1567.
18. Lim,J., Ha,M., Chang,H., Kwon,S.C., Simanshu,D.K., Patel,D.J. and Kim,V.N. (2014) Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell*, **159**, 1365–1376.
19. 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.
20. Chang,H.M., Triboulet,R., Thornton,J.E. and Gregory,R.I. (2013) A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature*, **497**, 244–248.
21. Aphasizheva,I., Maslov,D.A., Qian,Y., Huang,L., Wang,Q., Costello,C.E. and Aphasizhev,R. (2016) Ribosome-associated pentatricopeptide repeat proteins function as translational activators in mitochondria of trypanosomes. *Mol. Microbiol.*, **99**, 1043–1058.
22. Deng,J., Ernst,N.L., Turley,S., Stuart,K.D. and Hol,W.G. (2005) Structural basis for UTP specificity of RNA editing TUTases from Trypanosoma brucei. *EMBO J.*, **24**, 4007–4017.
23. Stagno,J., Aphasizheva,I., Bruystens,J., Luecke,H. and Aphasizhev,R. (2010) Structure of the mitochondrial editosome-like complex associated TUTase .1 reveals divergent mechanisms of UTP selection and domain organization. *J. Mol. Biol.*, **399**, 464–475.
24. Stagno,J., Aphasizheva,I., Rosengarth,A., Luecke,H. and Aphasizhev,R. (2007) UTP-bound and Apo structures of a minimal RNA uridylyltransferase. *J. Mol. Biol.*, **366**, 882–899.
25. Stagno,J., Aphasizheva,I., Aphasizhev,R. and Luecke,H. (2007) Dual Role of the RNA Substrate in Selectivity and Catalysis by Terminal Uridylyl Transferases. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14634–14639.
26. Munoz-Tello,P., Gabus,C. and Thore,S. (2014) A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure. *Nucleic Acids Res.*, **42**, 3372–3380.
27. Yates,L.A., Fleurdepine,S., Rissland,O.S., De Colibus,L., Harlos,K., Norbury,C.J. and Gilbert,R.J. (2012) Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat. Struct. Mol. Biol.*, **19**, 782–787.
28. Holm,L. and Sander,C. (1995) DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem. Sci.*, **20**, 345–347.
29. Martin,G. and Keller,W. (2007) RNA-specific ribonucleotidyl transferases. *RNA*, **13**, 1–16.
30. Hagan,J.P., Piskounova,E. and Gregory,R.I. (2009) Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.*, **16**, 1021–1025.

31. Heo,I., Joo,C., Cho,J., Ha,M., Han,J. and Kim,V.N. (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell*, **32**, 276–284.
32. Thornton,J.E., Du,P., Jing,L., Sjekloca,L., Lin,S., Grossi,E., Sliz,P., Zon,L.I. and Gregory,R.I. (2014) Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). *Nucleic Acids Res.*, **42**, 11777–11791.
33. Aphasizheva,I., Aphasizhev,R. and Simpson,L. (2004) RNA-editing terminal uridylyl transferase 1: identification of functional domains by mutational analysis. *J. Biol. Chem.*, **279**, 24123–24130.
34. Cording,A., Gormally,M., Bond,P.J., Carrington,M., Balasubramanian,S., Miska,E.A. and Thomas,B. (2016) Selective inhibitors of trypanosomal uridylyl transferase RET1 establish druggability of RNA post-transcriptional modifications. *RNA Biol.*, 1–9.
35. Aphasizhev,R. and Aphasizheva,I. (2007) RNA editing uridylyltransferases of trypanosomatids. *Methods Enzymol.*, **424**, 51–67.
36. Kabsch,W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
37. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
38. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
39. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
40. Pang,Y.P., Xu,K., Yazal,J.E. and Prendergas,F.G. (2000) Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.*, **9**, 1857–1865.
41. Rostkowski,M., Olsson,M.H., Sondergaard,C.R. and Jensen,J.H. (2011) Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.*, **11**, 6.
42. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, **79**, 926–935.
43. Demir,O. and Amaro,R.E. (2012) Elements of nucleotide specificity in the trypanosoma brucei mitochondrial RNA editing enzyme RET2. *J. Chem. Inf. Model*, **52**, 1308–1318.
44. Darden,T., Perera,L., Li,L.P. and Pedersen,L. (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure*, **7**, R55–R60.
45. Miyamoto,S. and Kollman,P.A. (1992) Settle - an analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, **13**, 952–962.
46. Salomon-Ferrer,R., Gotz,A.W., Poole,D., Le Grand,S. and Walker,R.C. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Particle Mesh Ewald. *J. Chem. Theory Comput.*, **9**, 3878–3888.
47. Durrant,J.D., Votapka,L., Sorensen,J. and Amaro,R.E. (2014) POVME 2.0: An enhanced tool for determining pocket shape and volume characteristics. *J. Chem. Theory Comput.*, **10**, 5047–5056.
48. Kozakov,D., Grove,L.E., Hall,D.R., Bohnuud,T., Mottarella,S.E., Luo,L., Xia,B., Beglov,D. and Vajda,S. (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.*, **10**, 733–755.
49. Votapka,L.W., Czapla,L., Zhenirovskyy,M. and Amaro,R.E. (2013) DelEnsembleElec: Computing ensemble-averaged electrostatics using DelPhi. *Commun. Comput. Phys.*, **13**, 256–268.
50. Wirtz,E., Leal,S., Ochatt,C. and Cross,G.A. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Mol. Biochem. Parasitol.*, **99**, 89–101.
51. Koller,J., Muller,U.F., Schmid,B., Missel,A., Kruft,V., Stuart,K. and Goringer,H.U. (1997) Trypanosoma brucei gBP21. An arginine-rich mitochondrial protein that binds to guide RNA with high affinity. *J. Biol. Chem.*, **272**, 3749–3757.
52. Jensen,B.C., Kifer,C.T., Brekken,D.L., Randall,A.C., Wang,Q., Drees,B.L. and Parsons,M. (2007) Characterization of protein kinase CK2 from Trypanosoma brucei. *Mol. Biochem. Parasitol.*, **151**, 28–40.
53. Aphasizheva,I., Zhang,L. and Aphasizhev,R. (2016) Investigating RNA editing factors from trypanosome mitochondria. *Methods*, **107**, 23–33.
54. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
55. Lapkouski,M. and Hallberg,B.M. (2015) Structure of mitochondrial poly(A) RNA polymerase reveals the structural basis for dimerization, ATP selectivity and the SPAX4 disease phenotype. *Nucleic Acids Res.*, **43**, 9065–9075.
56. Maris,C., Dominguez,C. and Allain,F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
57. Ringpis,G.E., Aphasizheva,I., Wang,X., Huang,L., Lathrop,R.H., Hatfield,G.W. and Aphasizhev,R. (2010) Mechanism of U insertion RNA editing in Trypanosome Mitochondria: The Bimodal TUTase activity of the core complex. *J. Mol. Biol.*, **399**, 680–695.
58. Ringpis,G.E., Stagno,J. and Aphasizhev,R. (2010) Mechanism of U-insertion RNA editing in Trypanosome Mitochondria: Characterization of RET2 functional domains by mutational analysis. *J. Mol. Biol.*, **399**, 696–706.
59. Aphasizhev,R. (2005) RNA uridylyltransferases. *Cell Mol. Life Sci.*, **62**, 2194–2203.
60. Elrod-Erickson,M., Rould,M.A., Nekludova,L. and Pabo,C.O. (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171–1180.
61. Rissland,O.S. and Norbury,C.J. (2008) The Cid1 poly(U) polymerase. *Biochim. Biophys. Acta*, **1779**, 286–294.
62. Munoz-Tello,P., Gabus,C. and Thore,S. (2012) Functional implications from the Cid1 poly(U) polymerase crystal structure. *Structure*, **20**, 977–986.
63. Aphasizheva,I., Ringpis,G.E., Weng,J., Gershon,P.D., Lathrop,R.H. and Aphasizhev,R. (2009) Novel TUTase associates with an editosome-like complex in mitochondria of Trypanosoma brucei. *RNA*, **15**, 1322–1337.
64. Aphasizhev,R., Aphasizheva,I., Nelson,R.E. and Simpson,L. (2003) A 100-kD complex of two RNA-binding proteins from mitochondria of Leishmania tarentolae catalyzes RNA annealing and interacts with several RNA editing components. *RNA*, **9**, 62–76.
65. Mattiacio,J.L. and Read,L.K. (2008) Roles for TbDSS-1 in RNA surveillance and decay of maturation by-products from the 12S rRNA locus. *Nucleic Acids Res.*, **36**, 319–329.