The structural basis of mRNA recognition and binding by eukaryotic pseudouridine synthase PUS1

Sebastian Grünberg[1,$], Lindsey A. Doyle[2,$], Nan Dai[1], Ivan R. Corrêa Jr.[1], Erbay Yigit[1,*] and Barry L. Stoddard[2,*]

[1] New England Biolabs

240 County Road

Ipswich, MA 01938 USA

[2] Division of Basic Sciences

Fred Hutchinson Cancer Research Center

1100 Fairview Ave. N.

Seattle WA 98109 USA

[$] Co-first authors

[*] Co-corresponding authors (yigit@neb.com; bstoddar@fredhutch.org)

Keywords: RNA modification, mRNA, pseudouridine synthase, x-ray crystallography

ABSTRACT

The chemical modification of RNA bases represents a ubiquitous activity that spans all domains of life. The formation of pseudouridine is the most common RNA modification and is observed within tRNA, rRNA, ncRNA and mRNAs. The catalysts of pseudouridylation, termed pseudouridine synthase or 'PUS' enzymes, include those that rely on guide RNA molecules and others that function as 'stand-alone' enzymes. Among the latter, up to ten are encoded in eukaryotic genomes, including several that modify uracil within mRNA transcripts. Neither the biological purpose of mRNA pseudouridylation, nor the mechanism by which individual mRNA bases are targeted, are well understood. In this study, we describe the high-resolution crystal structure of yeast PUS1 bound to an RNA target that we identified as being a hot spot for recognition, binding, and activity within a model mRNA. The enzyme recognizes RNA structural features corresponding to a base-paired duplex, which appears to act as a docking site leading to subsequent modification of the transcript. The study also allows us to visualize the divergence of related PUS-1 enzymes and their corresponding RNA target specificities, and to speculate on the basis by which this single PUS enzyme can bind and modify mRNA or tRNA substrates.

The cellular transcriptome throughout all domains of life displays a highly complex regulatory network of more than 150 known posttranscriptional RNA modifications that modulate RNA biogenesis, function, specificity, and stability (1-3). Pseudouridine (Ψ), a C5-glycosidic isomer of uridine, was discovered in 1951 and soon after termed the fifth nucleoside (4-6). Almost 20 years later, the first pseudouridine synthase gene, *TruA*, was identified, and found to modify tRNA in bacteria (7,8). More recent advances in transcriptome-wide mapping of Ψ revealed widespread pseudouridylation of mRNAs in eukaryotes at levels comparable to m^6A modifications(9-14). Despite the abundance of Ψ in mRNA, little is known about the purpose of that modification, or if it is installed in a site-specific manner.

Two classes of pseudouridine synthase enzymes, that are jointly responsible for pseudouridylation of RNA, differ in the way that they target their RNA substrates. Guide RNA-dependent PUS enzymes in eukaryotes and archaea (such as Cbf5 in yeast and dyskerin in humans) are part of a box H/ACA ribonucleoprotein complex and utilize small nucleolar RNAs (snoRNAs) as guides that recognize and base-pair with its substrates, which are mostly non-coding RNAs (ncRNA) (15). Alternatively, an RNA recognition mechanism that does not rely on guide RNA factors is employed by stand-alone PUS enzymes that independently recognize and modify their targets (reviewed in (16)). In contrast to the H/ACA snoRNP PUS enzymes, most stand-alone PUS enzymes are conserved throughout eukaryotes and bacteria. These PUS enzymes are classified into six families, which differ in N- or C-terminal extensions flanking a conserved catalytic core domain. That domain is present in all H/ACA snoRNP and stand-alone PUS enzymes and contains a central catalytic motif corresponding to an antiparallel β-sheet that is flanked by α-helices (17,18). A strictly conserved catalytic aspartate residue in that motif is required for the disruption of the glycosidic bond, rotation the base by 180°, and bond reformation (19).

Eukaryotes generally contain approximately ten unique stand-alone PUS enzymes (PUS1-PUS10), which differ in their substrate preference and localization in the cell.

Each of them modifies specific sites in tRNAs, snRNAs, and rRNAs by targeting a sequence and/or structural element in their respective substrates (16,19). The crystal structures of many PUS enzymes have been solved, including structures of eukaryotic/archaeal box H/ACA snoRNPs with guide and substrate RNAs (20-27) and bacterial stand-alone PUS enzymes bound to ncRNAs (28-31). Those studies have revealed mostly highly specific interactions of PUS enzymes with their respective non-coding targets.

It remains somewhat unclear which PUS enzymes are responsible for the pseudouridylation of messenger RNA (mRNA). Some studies have indicated that more than half of the Ψ modifications in mRNA are catalyzed by tRNA-specific PUS enzymes PUS4 and PUS7 through recognition of tRNA-like structures (10,12-14). In contrast, PUS1 (a member of the divergent TruA family of PUS enzymes) is the most promiscuous stand-alone PUS. While PUS1 was previously reported to modify multiple structurally diverse positions in tRNA, U2 and U6 snRNAs, recent data has suggested that PUS1 is the predominant PUS to modify uridines in mRNA (32-35). Although PUS1 sites in mRNA show little sequence similarity, a high-throughput pseudouridylation assay implicated a structure-dependent mRNA target recognition mechanism and suggested that modulation of the RNA structure may play a role in the regulation of mRNA pseudouridylation (14,36). However, it is still unclear how PUS1 selectively binds and modifies mRNA, mostly due to the lack of available structural information of PUS1 enzymes bound to mRNA targets.

Currently two structures of eukaryotic stand-alone PUS enzymes have been solved: the catalytic domains of human PUS1 and human PUS10 (37-39). In addition, a structure of *E. coli* TruA bound to tRNA revealed a mode of substrate binding by the closest bacterial homologue of PUS1 (40). Despite the structural similarities between both proteins, modeling and docking studies of tRNA and the core domain of human PUS1 suggests a significantly different orientation of the tRNA than in the tRNA-complex of *E. coli* TruA (39).

Visualization of how PUS1 binds to mRNA might provide considerable new insight into the basis of its action on such substrates. Therefore, we generated a pair of crystal structures of wild-type *S. cerevisiae* PUS1 and a catalytically inactive PUS1 mutant in complex with short mRNA fragments that were derived from regions of a representative mRNA substrate that were identified as hotspots for PUS1 binding and activity. The structures, from two unrelated crystal forms, both indicate that PUS1 recognizes and binds an RNA duplex, with PUS1 making extensive contacts with each RNA strand and positioning bases that immediately flank the bound duplex located near the enzyme active site. Additional examination and comparison with previously described structural and biochemical studies indicate that (1) while PUS1 and TruA use the same protein surface to interact with their respective target-RNAs, the position and identity of the corresponding contact residues and the position of the bound RNAs differ significantly from one another, and (2) PUS1 likely binds and acts on at least one tRNA target in a manner that is closely related to how it engages with its docking site in mRNA.

MATERIALS AND METHODS

Cloning, Protein Expression and Purification. N-terminally HIS-tagged *S. cerevisae* PUS1 (PUS1; NEB # M0526S) was used for most biochemical assays. A catalytically inactive PUS1 was generated by single point mutation of active site residue aspartate 134 into an alanine ($PUS1_{D134A}$). For biochemical characterization, the mutant PUS1 was expressed for two hours at 37°C in T7 Express lysY competent *E. coli* (NEB # C3010). Cells were harvested, washed with cold 1x PBS, resuspended in HisTrap binding buffer (HTBB; 20 mM Na_2HPO_4, 0.5 M NaCl, 20 mM Imidazole, 1 mM DTT, 20% glycerol) + 1x protease inhibitors (1 mM PMSF, 0.5 nM Leupeptin, 2.75 mM Benzamidine, 2 nM Pepstatin) and stored at -20°C. Cells were thawed, sonicated, and after centrifugation, the supernatant was saved and applied to a 5 ml HisTrap HP column (Cytiva) calibrated with HTBB. $PUS1_{D134A}$ was eluted with HisTrap elution buffer (20 mM Na_2HPO_4, 0.5 M NaCl, 0.5 M Imidazole, 1 mM DTT, 20% glycerol). Fractions containing PUS1 were pooled and dialyzed into HiTrap Heparin binding buffer (10 mM Na_2HPO_4, 100 mM NaCl, 20% glycerol, 1 mM DTT).

The dialyzed protein was applied to a 5 ml HiTrap Heparin column (Cytiva) and eluted with HiTrap Heparin elution buffer (10 mM Na_2HPO_4, 2 M NaCl, 20% glycerol, 1 mM DTT). The fractions containing PUS1 were pooled and dialyzed into HiTrap SP binding buffer (50 mM HEPES, pH 7.0 at 4°C, 50 mM NaCl, 1 mM DTT, 0.1 mM EDTA). The dialyzed PUS1 was applied to a 5 ml HiTrap SP HP cation exchange column and eluted with HiTrap SP elution buffer (50 mM HEPES, pH 7.0 at 4°C, 1 M NaCl, 1 mM DTT, 0.1 mM EDTA). Finally, fractions containing PUS1 were pooled and dialyzed into storage buffer (10 mM Tris/Cl pH 7.4 at 4°C, 200 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 50% Glycerol, 200 µg/ml BSA) and stored at - 80°C.

To generate an untagged version of the PUS1 enzyme for crystallography trials, $PUS1_{D134A}$ lacking the N-terminal HIS-tag was subcloned into expression vector pET21d (EMD Biosciences) using a Gibson Assembly cloning kit and protocol (NEB # E5510S) and sequence verified. A clone encoding tagless wild-type PUS1 was then generated from $PUS1_{D134A}$ by converting alanine 134 back to wild-type aspartate 134 using a QuickChange II XL kit and protocol (Agilent).

Sequence-verified constructs were transformed into BL21-CodonPlus(DE3)-RIL competent cells (Agilent) and expressed using a previously described autoinduction protocol (41). Bacterial pellets from a liter of culture were resuspended in 30 mL of Buffer A (50 mM NaCl, 10 mM Tris pH 7.5, 0.1 mM EDTA, 5% glycerol, 1 mM DTT), and 0.2 mM PMSF and 900 U of Benzonase added. Cells were lysed on ice with a Misonix S-4000 sonicator operating at 70% power; the cell suspension was subjected to 80 seconds of total sonication time over the course of four cycles (each applying 20 second sonication bursts followed by 60 second cooling periods). The cell lysate was clarified by centrifugation in an SS-34 rotor (Sorvall) at 30597 x g for 20 minutes at 4°C, followed by hand filtration through a 5 µm filter. The lysate was loaded onto a 5 mL HiTrap Heparin HP column (Cytiva) and the column was washed with 25 mL of Buffer A. A gradient from 100% Buffer A to 100% Buffer B (Buffer A augmented with 1.0 M NaCl) was run with a total elution volume of 100 mL and 5 mL fractions were collected. Peak fractions were combined, concentrated, filtered, and loaded onto a HiLoad 16/60 Superdex 200 gel filtration column (Cytiva) equilibrated in Buffer SEC (Buffer A

augmented with 200 mM NaCl). Peak fractions were combined and concentrated. The final purification of the protein, including the size exclusion chromatography elution profile of the purified protein, are shown in **Figure 1a.**

PUS1 activity assays. 210 pmol PUS1 was incubated with 84 pmol RNA substrate (synthetic RNA oligos: R164 – R170 (Integrated DNA Technologies), *in vitro* transcript: (IVT) (**Supplementary Table S1**) in 1x NEB buffer 1.1 for 120 min at 30°C. Reactions were stopped by adding 0.8 units of Proteinase K (NEB # P8107S) followed by incubation for 10 min at 37°C. The modified RNA was subsequently column-purified (NEB # T2030), 2 pmol were digested to single nucleosides using the Nucleoside Digestion Mix (NEB # M0649), and the ratio of Ψs versus uridines was determined via tandem quadrupole mass spectrometry (**Supplementary figure S1a**). LC-MS/MS analysis was performed in duplicate by injecting digested DNA on an Agilent 1290 UHPLC equipped with a G4212A diode array detector and a 6490A triple quadrupole mass detector operating in the positive electrospray ionization mode (+ESI). UHPLC was carried out on a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 µm) with the gradient mobile phase consisting of methanol and 10 mM aqueous ammonium formate (pH 4.4). MS data acquisition was performed in the dynamic multiple reaction monitoring (DMRM) mode. Each nucleoside was identified in the extracted chromatogram associated with its specific MS/MS transition: U [M+H]+ at m/z 245→113, Ψ [M+H]+ at m/z 245→209.1. External calibration curves with known amounts of the nucleosides were used to calculate their ratios within the samples analyzed.

Electrophoretic mobility shift assays. 100 nM to 1 µM of the PUS1 wildtype and the D134A mutant enzyme were incubated with 10 pmol of synthetic RNA oligos R169, R194, and R195 as described above. After incubating for 120 min at 30°C and immediately before loading onto the gel, 18 µl of each reaction were transferred to a fresh tube containing 2 µl native PAGE loading dye (0.05% xylene cyanol, 50% glycerol). 18 µl of each sample were then loaded on a 6% Novex™ TBE gel that was running at 150 V at 4°C. After 30 min, the gel was stopped and visualized using the Cy2

channel of an Amersham Typhoon Laser Scanner after staining with SYBR Gold (1/20,000 dilution, Thermo Fisher Scientific # S11494) for 10 min.

Crystallization and Data Collection. RNA targets were complexed with purified protein, with the RNA present in a 1.4-molar excess over the protein. Complexes were screened for crystal grown in 96 well plate formats, with 200 nanoliter drop volumes equilibrated against 100 microliter reservoirs, against multiple commercial crystallization screens, while using a mosquito robot (TTP LabTech). Drops that generated visible crystals were then used to set up subsequent screening and expansion trays by hand, with 2 microliter drops equilibrating against 1000 microliter reservoir volumes.

A mixture of catalytically inactive PUS1 ($PUS1_{D134A}$) at 11.6 mg/mL and a slight molar excess of RNA (‘R263’; 5’-AAA UCG GGA UUC CGG AUA-3’) crystallized at 4° C after equilibration against a reservoir containing 0.05 M Ammonium sulfate, 0.05 M Bis-Tris pH 6.0, and 26% Pentaerythritol ethoxylate. In contrast, a mixture of wild-type PUS1 (scPUS1) at 11 mg/mL in the presence of a similar RNA construct harboring a 5-fluorouracil moiety at a position corresponding to potential enzymatic modification (‘R340’; 5’-[5FU]AA UCG GGA UUC CGG AUA-3’) crystallized at 25°C after equilibration against a reservoir containing 0.2 M Potassium sodium tartrate and 18% PEG 3350 at 25°C.

Crystals were transferred to a cryocooling solution containing either 0.04 M Ammonium sulfate, 0.04 M Bis-Tris pH 6.1, 23% Pentaerythritol ethoxylate, and 22% ethylene glycol or 0.2 M Potassium sodium tartrate, 20% PEG 3350, and 20% ethylene glycol for $PUS1_{D134A}$ or scPUS1, respectively. Crystal were then flash frozen in liquid nitrogen and data was collected on ALS BCSB beam line 5.0.1. Data for $PUS1_{D134A}$ was processed using program HKL2000(42). Data for scPUS1 was automatically processed using program XDS(43).

Phasing and Refinement. Structures were phased via the molecular replacement method, using the catalytic domain of human PUS1 (PDB ID 4J37) (39) as a search model. Molecular replacement searches were conducted using program PHASER (44) within the Phenix crystallographic computational suite (45). The structure was rebuilt and refined using programs COOT (46) and PHENIX.REFINE (47).

RESULTS

Recombinant PUS1 is active in vitro. When incubated with PUS1 as described in Materials and Methods, approximately 24% of uridines of a 1.8 kb mRNA *in vitro* transcribed from the Firefly luciferase (Fluc) gene were converted to Ψs after 90 min of incubation (**Supplemental Figure 1b**). To identify a suitable substrate for PUS1 structural studies, used a sliding window approach, dividing the Fluc mRNA first into 300 nt followed by 60 nt fragments, and identified a region spanning positions 739 to 845 relative to the beginning of the Fluc transcript to be most highly modified.

To further narrow down the RNA substrate sequence, we generated various short synthetic RNA oligos ranging from 29 to 40 nucleotides in length, as well as a short *in vitro* transcript (IVT; 59 nt) (**Figure 1b**). RNA fragments containing Fluc positions 787 to 796 showed high modification rates when incubated with PUS1, while RNA oligos R164 and R165, that did not contain this region, were not significantly modified (**Figure 1c**). RNA oligo R169 and the IVT substrate showed the highest amount of pseudouridylation at approximately 18% pseudoU. Due to the shorter length of the R169 as compared to the IVT RNA, we selected R169 as PUS1 substrate for the following crystallographic studies.

***Active site mutant of PUS1 is catalytically inactive*.** To avoid any enzymatic reaction that could potentially lead to structures of reaction intermediates or substrate release during the crystallization trials, we generated a catalytically inactive PUS1 variant. Yeast PUS1 aspartate residue 134 and the catalytically essential aspartate 146 in human PUS1 (39,48) are conserved. Substituting the aspartate residue 134 with alanine (D134A) generated a catalytically inactive PUS1 variant ($PUS1_{D134A}$) that lacked the ability to pseudouridylate an RNA substrate (**Figure 1d**).

***PUS1 binds to an RNA duplex*.** Despite forming stable complexes in EMSA assays (data not shown), we were unable to grow crystals of $PUS1_{D134A}$ with the R169 substrate, possibly due to flexible portions of that long RNA molecule interfering with

lattice formation. We therefore decided to attempt crystallization in the presence of a shorter single-stranded 18 base RNA oligonucleotide (R263; see **Figure 1b**; sequence provided in methods) that corresponded to the 3' half of R169. As expected, we saw comparable binding of both the wild-type and the D134A mutant to that shorter RNA substrate (**Supplementary Figures S1c and S1d**). We also observed an additional band running slightly slower than the single-stranded R263 oligo that was too small to be caused by an RNA:PUS1 complex. In fact, when using a FAM-labeled version of R263, the higher MW band was the dominant band in the no enzyme control reaction (**Supplementary Figure S1d**, lane 1). We concluded that this particular RNA oligo forms a duplex in solution that is efficiently bound by $PUS1_{wt}$ and by $PUS1_{D134A}$.

Crystals of the R263 RNA in complex with $PUS1_{D134A}$ belonged to space group C2 and diffracted to 2.4 Å resolution (**Table 1**). Modeling and refinement of the enzyme-RNA complex yielded values for the crystallographic R-factors (R_{work} and R_{free}) of 0.222 and 0.265 respectively, with tight protein geometry (rmsd bonds and angles 0.003 Å and 0.58°; 95.73% of residues in favored Ramachandran regions). The average B-values for the protein and the bound RNA were comparable (65.48 and 62.74 $Å^2$, respectively). The electron density maps from molecular replacement and subsequent rounds of refinement displayed well-ordered density for most of the protein chain (the first 70 residues, last 49 residues, and four subsequent surface loops ranging in length from 2 to 19 residues were disordered) and the first 12 bases of the RNA substrate (the final 6 base pairs were also disordered).

The contents of the crystallographic asymmetric unit correspond to a single protein subunit and a single strand of the bound RNA ligand (**Figure 2a**). That complex was observed to be part of a higher-order assembly, comprised of an RNA duplex and two bound copies of PUS-1, that becomes obvious via application of a crystallographic dyad symmetry axis (**Figure 2b**). The base pairs within the RNA duplex (involving positions 2 through 10 in each strand) display both Watson-Crick and non-Watson-Crick interactions with their counterparts, including two central G:G reverse Hoogstein basepairs. The 5'- and 3'-most modeled base on each strand are not engaged in base-paired interactions but are also not flipped out of the strand's duplex conformation

(**Figure 2c**). Immediately proximal to the 5' end of each RNA strand, a well-occupied and tightly coordinated sulfate ion is observed, at a location and distance appropriate to represent an additional backbone phosphate group if the RNA were extended by an additional base at its 5' end (**Figure 2d**). A sulfate ion was also observed and modeled at the same position in the previously described structure of the human PUS-1 apo enzyme (PDB 4J37) (39).

In contrast to the extensive base-paired contacts between the two crystallographically related RNA strands, the corresponding pair of bound protein molecules do not display significant contacts with one another; the few contacts between the two bound protein subunits are limited to two surface-exposed loops (residues 86 to 89 and 95 to 98) in the N-terminal region of the enzyme. Therefore, we believe that the two copies of PUS1 that are associated with opposite sides and ends of the symmetric RNA duplex are bound as individual monomers, independently of one another. That conclusion agrees with the solution behavior of the purified enzyme, which eluted from a size exclusion column as a monomeric species at high micromolar concentrations of protein (**Figure 1a**).

Within the complex between a single bound protein subunit and the RNA duplex (**Figure 3**), at least thirteen amino acid side chains contact numerous atoms on each of the two RNA strands. The residues involved in substrate recognition and binding are largely comprised of two broadly separated clusters of residues within the enzyme's sequence and structure. The first group of seven RNA-contacting amino acids is distributed between residues 92 and 188; they collectively contact one end of the RNA duplex and the adjacent sulfate ion, positioning the first base at the 5'-end of an RNA strand near the entrance to the active site. All but one of those residues are conserved across eukaryotic PUS1 homologues from yeast to humans (**Supplementary Figure S2**). The second group of four additional RNA-contacting amino acids is distributed between residues 362 and 394; they contact a series of bases and backbone atoms further downstream on one of the two RNA strands and are also conserved across PUS1 enzymes.

Finally, two additional residues (K277 and Y459) contact the RNA backbone at two bases near the opposite end of the bound RNA duplex. K277 is located in an insert unique to the yeast enzyme (spanning residues 206 to approximately L279) which is not conserved across eukaryotic PUS1 enzymes. The tyrosine in position 459 is yeast specific. However, in other eukaryotic PUS1 variants the respective position shows a conserved threonine. All the protein-RNA contacts described above are duplicated via symmetry by the second RNA-bound protein subunit (**Figure 3c**); each independently bound protein monomer makes multiple contact to nucleotide bases and to backbone phosphate groups on both RNA strands.

In the structure of catalytically inactive $PUS1_{D134A}$ bound to the RNA construct described above, the unpaired 5' base (A_1) is positioned within several angstroms of residue 134 and appears to be within potential distance to rotate into the enzyme active site. To further examine the interactions of bound RNA with a catalytically competent version of the enzyme (in which the catalytic aspartate at position 134 was restored) we altered the RNA ligand by substituting a 5-fluoro-uracil (5-FU) base at position number 1, reasoning that it might be captured in a suicide complex by the enzyme after being flipped into the active site and subsequent nucleophilic attack by the carboxylate. RNAs harboring such substitutions have previously been demonstrated to act as mechanism-based inhibitors of pseudouridine synthase enzymes and have been used to demonstrate the structural mechanism by which the target base gains access to the enzyme's active site (49).

Crystals of wild-type PUS1 in complex with the new R340 RNA construct (5'-[*5FU*]AA UCG GGA UUC CGG AUA-3') belonged to a different space group ($P6_122$) and diffracted to 2.9 Å resolution. Although the space group and corresponding lattice packing arrangement of the protein-RNA enzyme complex was unrelated to the packing of the C2 space group of the $PUS1_{D134A}$:RNA complex, the same RNA duplex and the same position of two independently bound protein subunits was observed (**Supplementary Figure S3a**). The RNA duplex and second bound protein molecule is again generated by the application of a crystallographic dyad axis on the contents of the asymmetric unit (which again corresponds to a single protein subunit and a single RNA chain). Structural superposition of an RNA-bound monomer or dimer from the two

crystal structures produced rmsd values of 0.63 Å and 1.47 Å, respectively. The reproducibility of all structural observations described above, in two different crystallization conditions and crystal forms, reinforces the conclusion that the formation and presence of an RNA duplex near the target base, and recognition of elements within each RNA strand by residues from each protein subunit, is a reproducible and mechanistically relevant feature of substrate recognition and activity.

Beyond providing an independent confirmation of the binding of two copies of the enzyme to an RNA duplex in a symmetric arrangement, the resulting electron density map displayed a significant feature of positive difference density in the active site that might be indicative of low occupancy (estimated at less than 10% in refinement) by the 5' 5FU nucleobase of the RNA (**Supplementary Figure S3b**). However, the combination of lower resolution and mixture of density features surrounding that base prevented us from unambiguously modeling the bound RNA in a state corresponding to a trapped catalytic complex.

The structure of the RNA-bound PUS1 enzyme was found to be similar to both the previously solved structure of human PUS1 in the absence of bound RNA (PDB 4J37) and the structure of bacterial tRNA pseudouridine synthase TruA (PDB 2NR0), with a rmsd across all comparable alpha carbons of approximately 2 Å in both pairwise superpositions (**Figure 4**). PUS1 and TruA employ the same protein surface, spanning the majority of their primary sequences, to contact their respective RNA targets; however, the position and identity of the corresponding contact residues and the conformation and orientation of their bound RNAs differ significantly from one another (**Supplementary Figure S4)**.

DISCUSSION

Previous biochemical studies have provided solid evidence indicating how PUS enzymes might target and interact with their ncRNA substrates. However, it is still unknown how stand-alone PUS enzymes interact with mRNA, or if PUS-mediated mRNA pseudouridylation is site specific or random. Our data present crystal structures of a guide RNA-independent eukaryotic PUS enzyme bound to a mRNA fragment and

shed light on the molecular mechanism underlying mRNA recognition and binding by such enzymes.

While bacterial TruA has been reported to crystalize as a homodimer (with each subunit binding an individual tRNA), other PUS enzymes – including human PUS1 (37,39) – function as monomers. Therefore, the recruitment of two PUS1 to the RNA-duplex was initially surprising. However, our data and the crystal structures (solved in two different space groups) collectively indicate that yeast PUS1 does in fact interact with RNA substrates as a monomer, and that the observation of two bound enzyme subunits to the RNA duplex in a symmetric manner is a (fortuitous) biochemical and crystallographic artifact. First, the PUS1 enzyme is a monomer in solution as confirmed by size-exclusion chromatography (SEC) at high protein concentrations (**Figure S1a**). Second, despite binding to opposing ends and sides of the highly symmetric RNA-duplex used for crystallization, the PUS1 subunits do not display an extensive interface that would be expected for a functional protein dimer. The intermediate complexes observed in the EMSA experiments are likely formed by different combinations of single-stranded or duplexed RNA binding to either a single PUS1 subunit or to two independently bound PUS1 subunits. In this scenario, the binding preference of PUS1 to the respective RNA substrate would be determined by the secondary structure of the latter. However, we cannot exclude the possibility that binding of PUS1 to a single stranded RNA is required for or supports the formation of a stable RNA-duplex for some RNA substrates *in vitro*.

A question that immediately arises from our analyses is if the base-paired RNA duplex observed in the crystal structures represents a physiological and/or mechanistical state that influences PUS1 specificity and activity on mRNA. Our data show that each individual PUS1 subunit makes extensive contacts to both strands in the RNA duplex. Those RNA-contacting residues, which are well-conserved conserved across PUS1 homologues, appear to enable binding of an RNA double helix and position the 5' flanking region of the bound RNA strand close to the active site of a PUS1 subunit. The importance of double-stranded stem RNA-structures for mRNA pseudouridylation by TruB1 and PUS7 has recently been reported by the Schwartz laboratory through high-throughput Ψ mapping (10). Both enzymes modify structure motifs in mRNA that match

important motifs in their tRNA-substrates, suggesting a similar mechanism of site-recognition.

In contrast, PUS1 and its bacterial counterpart TruA are fairly promiscuous in their target selection, modifying multiple RNAs with divergent sequences in the region of modification (50). Accordingly, Illumina-based mapping of Ψ installed by PUS1 in yeast and human cells revealed a weak HRU sequence motif, but the authors concluded that this motif alone was not enough to explain PUS1 specificity (14). Using an approach that combined high-throughput kinetic analysis with computational prediction and mutational analysis, the same authors recently reported that most mRNA targets of PUS1 share a similar structural motif that regulates pseudouridylation by PUS1 (36). They showed that an RNA-duplex in form of a stem, connected by a short loop is essential for PUS1 binding, and that the length and stability of the stem directly impacts the pseudouridylation rate. Their findings agree with previous data showing that a loop flanked by stem structures is preferably targeted by human PUS1 (37).

The observation that a single copy of PUS1 interacts with both RNA strands in a bound RNA duplex reinforces the concept that a base paired structure within a potential mRNA substrate is important for enzyme binding and eventual positioning of a pseudouridylation site near the active site of the enzyme. Our biochemical data support a hypothesis for structure-driven pseudouridylation by PUS1.

One explanation for yeast PUS1's less restricted mode of structure recognition could be the location of its thumb loop (Tyr 83 – Thr 99). In the PUS1:mRNA structure, the loop is ordered and pointing away from the active site. This orientation results in a significantly wider opening of the active site cleft compared to TruA or human PUS1(39), which may allow for the binding of more diverse RNA substrates.

It is noteworthy that none of the uridines within the crystallization oligonucleotide are located near the enzyme's active site in the crystal structures but are instead involved in duplex formation via base pairing interactions between RNA strands. The structure suggests that an RNA-duplex with a minimum length of 10 to 12 nt is required for all RNA:PUS1 interactions to stably form and to position the RNA correctly in the active site. This agrees with a 11 bp median length of double stranded stem structures found

in PUS1 mRNA targets (36). Our PUS1-dsRNA structure suggests that a minimum RNA-length is required for the formation of a sufficient stem structure for PUS1-binding near single stranded RNA regions in which the target uridines are accessible.

Comparison of the structures solved in this study with that of bacterial tRNA pseudouridine synthase TruA (29) (a representative of the structural family from which PUS1 is derived) bound to its tRNA substrate indicates how RNA recognition mechanisms can and have diverged dramatically. Both enzymes display contacts to their respective RNA targets involving approximately 15 residues distributed across their N- and C-terminal domains, and in so doing position their active sites near bases that are potential sites of modification (individual bases located 5' of the bound RNA duplex for PUS1; bases 38, 39 and 40 of the tRNA anticodon stem loop for TruA). Of the residues involved in direct RNA contacts from each enzyme, several equivalent positions in each enzyme's N-terminal region contact RNA atoms but involve quite different protein residues ($\underline{H}_{89}$GM$\underline{Q}$Y$\underline{N}$PP$\underline{N}_{97}$ in PUS1; $\underline{Y}_{24}$GW$\underline{Q}$R$\underline{QNEV}_{32}$in TruA) that contact unique nucleotide identities and conformations in their corresponding RNA targets.

Conversely, a separate examination and comparison of the structures solved in this study against a tRNA known to be modified within its anticodon loop by PUS1 (35,39) indicates that the enzyme might be able to do so through a binding mode and interactions with that substrate's anticodon stem loop that are similar to the binding and interactions observed in our crystal structure. Superposition of the tRNA substrate, via its anticodon stem loop, onto the RNA duplex from the crystal structure (**Figure 5)** positions a base in the tRNA anticodon loop that is known to be modified by the enzyme immediately adjacent to the enzyme's active site. The corresponding docked position of the tRNA further positions the previously identified RNA-contacting residues at similar distances and potential interactions with atoms along the tRNA backbone. As well, the superposition avoids significant clash between the C-terminal helical region of the PUS1 enzyme and the remainder of the tRNA molecule.

While the structures presented here provide the first comprehensive mechanistic insight into the interaction between a eukaryotic stand-alone PUS1 and its mRNA substrate,

some questions remain unanswered. First, is the underlying RNA sequence within a given RNA stem loop in an mRNA important for pseudouridylation activity, or is any such structure, regardless of base pair identity, sufficient for enzyme activity? Second, if guide RNA-independent PUS enzymes recognize and bind to defined RNA structural motifs, and if such motifs are indeed responsible for the majority of mRNA pseudouridylation, are they specifically placed in mRNA to have regulatory functions? Ψs are not randomly distributed within mRNAs, and have been shown to be enriched in the 3'-UTRs and coding region(14), suggesting a functional role. The structural basis for target recognition by PUS1 presented here is a first step towards understanding how PUS enzymes select their targets in mRNA, and if and how this process may be regulated.

ACKNOWLEDGEMENTS

The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The Pilatus detector was funded under NIH grant S10OD021832. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169

DATA AVAILABILITY

The structures described in this manuscript have been deposited in the RCSB protein database and are available for public download and examination (PDB ID codes 7R9F and 7R9G). The original source data and raw images corresponding to the biochemical analyses of PUS1 behavior and activity have been uploaded to a public repository

(at https://dataverse.harvard.edu/dataverse/scPUS1/) and are also available upon request from the authors.

CONFLICT OF INTEREST STATEMENT

SG, ND, IC and EY are employees of New England Biolabs, a for-profit biotech company that manufactures enzymes and reagents such as PUS1 for commercial sale. BLS is a paid consultant for New England Biolabs, which also funded this work in his laboratory.

REFERENCES

1. Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crecy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A. *et al.* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res*, **46**, D303-D307.
2. Jones, J.D., Monroe, J. and Koutmou, K.S. (2020) A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. *Wiley Interdiscip Rev RNA*, **11**, e1586.
3. McCown, P.J., Ruszkowska, A., Kunkler, C.N., Breger, K., Hulewicz, J.P., Wang, M.C., Springer, N.A. and Brown, J.A. (2020) Naturally occurring modified ribonucleosides. *Wiley Interdiscip Rev RNA*, **11**, e1595.
4. Cohn, W.E. (1959) 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochim Biophys Acta*, **32**, 569-571.
5. Davis, F.F. and Allen, F.W. (1957) Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem*, **227**, 907-915.
6. Cohn, W.E. and Volkin, E. (1951) Nucleoside-5'-Phosphates from Ribonucleic Acid. *Nature*, 483–484.
7. Kammen, H.O., Marvel, C.C., Hardy, L. and Penhoet, E.E. (1988) Purification, structure, and properties of Escherichia coli tRNA pseudouridine synthase I. *J Biol Chem*, **263**, 2255-2263.
8. Ciampi, M.S., Arena, F. and Cortese, R. (1977) Biosynthesis of pseudouridine in the in vitro transcribed tRNATyr precursor. *FEBS Lett*, **77**, 75-82.
9. Tardu, M., Jones, J.D., Kennedy, R.T., Lin, Q. and Koutmou, K.S. (2019) Identification and Quantification of Modified Nucleosides in Saccharomyces cerevisiae mRNAs. *ACS Chem Biol*, **14**, 1403-1409.
10. Safra, M., Nir, R., Farouq, D., Vainberg Slutskin, I. and Schwartz, S. (2017) TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res*, **27**, 393-406.
11. Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F. and Yi, C. (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol*, **11**, 592-597.
12. Lovejoy, A.F., Riordan, D.P. and Brown, P.O. (2014) Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One*, **9**, e110799.
13. Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S. *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, **159**, 148-162.
14. Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M. and Gilbert, W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143-146.
15. Kiss, T., Fayet-Lebaron, E. and Jady, B.E. (2010) Box H/ACA small ribonucleoproteins. *Mol Cell*, **37**, 597-606.

16. Rintala-Dempsey, A.C. and Kothe, U. (2017) Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol*, **14**, 1185-1196.
17. Koonin, E.V. (1996) Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res*, **24**, 2411-2415.
18. Mueller, E.G. and Ferre-D'Amare, A.R. (2009) In Grosjean, H. (ed.), *In DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin, TX, pp. 363-376.
19. Hamma, T. and Ferre-D'Amare, A.R. (2006) Pseudouridine synthases. *Chem Biol*, **13**, 1125-1135.
20. Czudnochowski, N., Ashley, G.W., Santi, D.V., Alian, A., Finer-Moore, J. and Stroud, R.M. (2014) The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs. *Nucleic Acids Res*, **42**, 2037-2048.
21. Liang, B., Xue, S., Terns, R.M., Terns, M.P. and Li, H. (2007) Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat Struct Mol Biol*, **14**, 1189-1195.
22. Duan, J., Li, L., Lu, J., Wang, W. and Ye, K. (2009) Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell*, **34**, 427-439.
23. Liang, B., Zhou, J., Kahen, E., Terns, R.M., Terns, M.P. and Li, H. (2009) Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat Struct Mol Biol*, **16**, 740-746.
24. Li, S., Duan, J., Li, D., Yang, B., Dong, M. and Ye, K. (2011) Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev*, **25**, 2409-2421.
25. Manival, X., Charron, C., Fourmann, J.B., Godard, F., Charpentier, B. and Branlant, C. (2006) Crystal structure determination and site-directed mutagenesis of the Pyrococcus abyssi aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res*, **34**, 826-839.
26. Hamma, T., Reichow, S.L., Varani, G. and Ferre-D'Amare, A.R. (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol*, **12**, 1101-1107.
27. Rashid, R., Liang, B., Baker, D.L., Youssef, O.A., He, Y., Phipps, K., Terns, R.M., Terns, M.P. and Li, H. (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell*, **21**, 249-260.
28. Hoang, C., Chen, J., Vizthum, C.A., Kandel, J.M., Hamilton, C.S., Mueller, E.G. and Ferre-D'Amare, A.R. (2006) Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Mol Cell*, **24**, 535-545.
29. Hoang, C. and Ferre-D'Amare, A.R. (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929-939.
30. Phannachet, K., Elias, Y. and Huang, R.H. (2005) Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry*, **44**, 15488-15494.

31. Pan, H., Agarwalla, S., Moustakas, D.T., Finer-Moore, J. and Stroud, R.M. (2003) Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proc Natl Acad Sci U S A*, **100**, 12648-12653.
32. Basak, A. and Query, C.C. (2014) A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast. *Cell Rep*, **8**, 966-973.
33. Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H. and Branlant, C. (1999) Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol*, **19**, 2142-2154.
34. Behm-Ansmant, I., Massenet, S., Immel, F., Patton, J.R., Motorin, Y. and Branlant, C. (2006) A previously unidentified activity of yeast and mouse RNA:pseudouridine synthases 1 (Pus1p) on tRNAs. *RNA*, **12**, 1583-1593.
35. Motorin, Y., Keith, G., Simon, C., Foiret, D., Simos, G., Hurt, E. and Grosjean, H. (1998) The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA*, **4**, 856-869.
36. Carlile, T.M., Martinez, N.M., Schaening, C., Su, A., Bell, T.A., Zinshteyn, B. and Gilbert, W.V. (2019) mRNA structure determines modification by pseudouridine synthase 1. *Nat Chem Biol*, **15**, 966-974.
37. Huet, T., Miannay, F.A., Patton, J.R. and Thore, S. (2014) Steroid receptor RNA activator (SRA) modification by the human pseudouridine synthase 1 (hPus1p): RNA binding, activity, and atomic model. *PLoS One*, **9**, e94610.
38. McCleverty, C.J., Hornsby, M., Spraggon, G. and Kreusch, A. (2007) Crystal structure of human Pus10, a novel pseudouridine synthase. *J Mol Biol*, **373**, 1243-1254.
39. Czudnochowski, N., Wang, A.L., Finer-Moore, J. and Stroud, R.M. (2013) In human pseudouridine synthase 1 (hPus1), a C-terminal helical insert blocks tRNA from binding in the same orientation as in the Pus1 bacterial homologue TruA, consistent with their different target selectivities. *J Mol Biol*, **425**, 3875-3887.
40. Hur, S. and Stroud, R.M. (2007) How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. *Mol Cell*, **26**, 189-203.
41. Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, **41**, 207-234.
42. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*, **276**, 307-326.
43. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr*, **66**, 125-132.
44. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
45. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, **75**, 861-877.
46. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.

47. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*, **68**, 352-367.
48. Sibert, B.S., Fischel-Ghodsian, N. and Patton, J.R. (2008) Partial activity is seen with many substitutions of highly conserved active site residues in human Pseudouridine synthase 1. *RNA*, **14**, 1895-1906.
49. Spenkuch, F., Motorin, Y. and Helm, M. (2014) Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol*, **11**, 1540-1554.
50. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res*, **26**, 148-153.

FIGURE CAPTIONS

Figure 1. Recombinant PUS1 modifies mRNA. ***Panel a:*** Wildtype PUS1 ($PUS1_{wt}$, orange) and catalytically inactive PUS1 ($PUS1_{D134A}$, blue) run as monomers over HiLoad 16/60 Superdex 200. Standards (Bio-Rad Gel Filtration Standards) are overlaid in dotted gray and appropriate peaks labeled. **Inset** SDS-PAGE of the final purification product of $PUS1_{wt}$ without (-) or with (+) dithiothreitol. ***Panel b:*** Schematic view of the RNA substrates used (sequences in **Supplementary Table S1)**. Firefly Luciferase (Fluc) mRNA positions 739, 787, 796, and 845 are indicated. RNA oligos used in biochemical and structural analyses are highlighted by color (R263 in purple, R169 in red). ***Panel c:*** Level of pseudouridylation as determined by LC-MS/MS (% pseudoU) of RNA substrates by PUS1 after incubation for two hours. Each data set contains at least two replicas. The length of the RNA substrates is indicated above/inside of each column. ***Panel d:*** A $PUS1_{D134A}$ enzyme is catalytically inactive. The rate of pseudouridylation (% pseudoU) by increasing concentrations of PUS1 wildtype (left) and D134A mutant (right) with R168 substrate is shown. Data of two replicas is shown.

Figure 2. Model and representative electron-density for RNA-bound PUS1. ***Panel a:*** The contents of the asymmetric unit, for both structures that were solved, corresponds to a single protein subunit (colored as a spectrum, from the blue N-terminal end of the refined model to the red C-terminal end) bound to a single RNA oligonucleotide (grey bases). The model of the catalytically inactive D134A enzyme is shown in two orientations related by a 90° rotation around the x-axis. ***Panel b:*** In both structures, the application of a crystallographic 2-fold dyad rotation axis generates a dimeric complex in which two subunits are independently bound to an RNA duplex. The second protein subunit and second RNA strand are colored in dark teal and lighter grey, respectively. ***Panel c:*** Representative 2Fo-Fc electron density contoured across the RNA duplex and ***Panel d:*** at the region of protein-RNA contacts observed at the 5' end of one RNA strand. The structural features illustrated in this figure are replicated for the wild-type enzyme bound to a closely related RNA complex, which was solved in an

unrelated crystallographic space group and lattice (**Supplementary Figure S3**). The position of the bound sulfate ion mirrors a similarly placed sulfate in the previously described structure of unbound human PUS1.

Figure 3. Protein-RNA contacts. ***Panel a:*** The first twelve nucleotides of each RNA strand are visible in the crystal structure, while the last six nucleotides are disordered and unobservable. Each protein subunit displays identical contacts, related by two-fold symmetry, to RNA bases and backbone atoms and to tightly coordinated sulfate ion immediately upstream of the upstream base in each bound RNA (contacts made by only one protein subunit are displayed for clarity). Each protein subunit contacts atoms from each RNA strand, implying that the formation and recognition of an RNA duplex near the active site is a mechanistically relevant feature of base selection. ***Panel b:*** Distribution of contacts in the protein-subunit interface for one enzyme subunit. ***Panel c:*** Distribution of contacts around the RNA duplex for both enzyme subunits in the dimeric assemblage.

Figure 4. Comparison of yeast PUS1 to *E. coli* TruA and human PUS1. ***Panel a:*** Human PUS1 apo-enzyme (4J37). The structure includes two bound sulfate ions, one of which aligns with a single sulfate ion near the enzyme active site that was also observed in RNA-bound *S. cerevisiae* PUS1. ***Panel b*****:** *S. cerevisiae* $PUS1_{D134A}$ bound to duplex RNA. A large insertion in the yeast enzyme, spanning residues S206 to approximately L279 (indicated by the oval), is unique as compared to its homologues in other eukaryotes (**Supplementary Figure S2**). It contains two RNA-contacting residues that are unique to the yeast enzyme. ***Panel c:*** *E. coli* TruA (2NR0) bound to tRNA. TruA utilizes an equivalent surface to bind its respective target but in a considerably different manner from PUS1.

Figure 5. Superposition of the crystal structure of PUS1 bound to the RNA duplex described in this study with a human tRNA known to be modified by the same enzyme within its anticodon loop. The protein is colored; the RNA duplex from the crystal structure is dark grey; the tRNA substrate that is docked onto the crystal structure is

light grey. The position of the active site D134 residue is indicated with light blue spheres; the position of the site of uracil modification is indicated with red spheres.

Table 1. Crystallographic Data and Refinement Statistics

	scPUS1_D134A + sgR263	**scPUS1 + sgR340**
PDB ID	***7R9G***	***7R9F***
Data Collection		
Space group	C2	P 6_1 2 2
Unit cell		
a, b, c	104.2, 98.3, 72.2	134.5, 134.5, 158.1
alpha, beta, gamma	90, 102.8, 90	90, 90, 120
Wavelength (Å)	0.9774	0.9774
Resolution range (Å)	50-2.4 (2.44-2.4)	50-2.89 (2.993-2.89)
Unique reflections	27637	19497
R-merge	0.072 (0.632)	0.026 (0.730)
R-meas	0.078 (0.698)	0.036 (1.033)
R-pim		0.026 (0.730)
CC1/2	(0.813)	(0.502)
I/sigma(I)	22.2 (1.6)	15.62 (1.09)
Chi^2	0.74	
Multiplicity	6.5 (5.3)	2.0 (2.0)
Completeness (%)	99.6 (95.1)	99.9 (99.84)
Wilson B-factor	53.3	93.4
Refinement		
R-work	0.2217	0.2537
R-free	0.2647	0.2964
Number of non-hydrogen atoms	3234	2962
macromolecules	3178	2938
ligands	6	24
solvent	50	0
Protein residues	385	379
RMS(bonds)	0.003	0.004
RMS(angles)	0.58	0.75
Ramachandran favored (%)	95.73	89.95
Ramachandran allowed (%)	3.47	7.34
Ramachandran outliers (%)	0.8	2.72
Clashscore	6.51	9.17
Average B-factor	65.36	91.61
macromolecules	65.48	91.63
ligands	62.74	88.32
solvent	58.11	N/A

Fig. 1

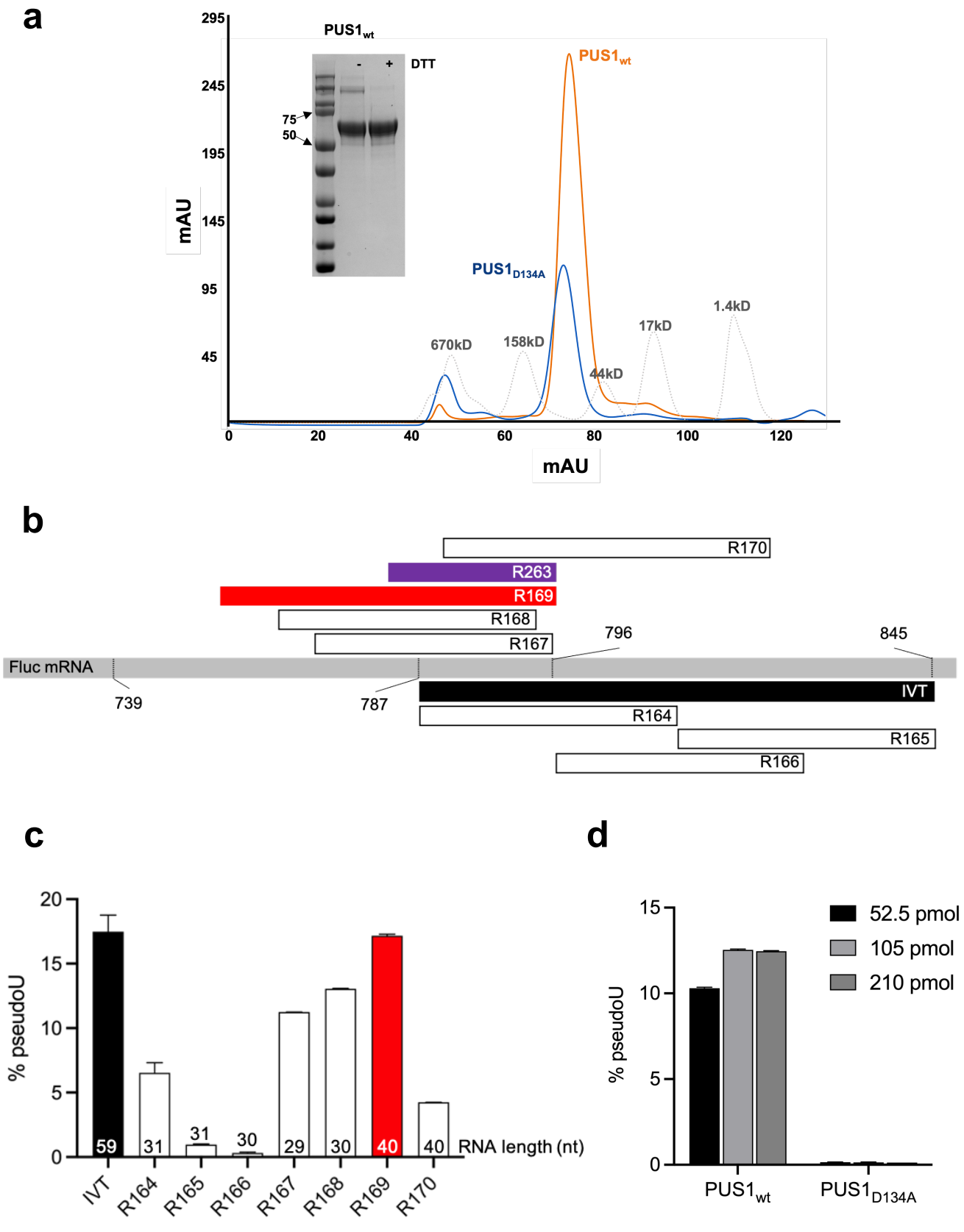

a
PUS1wt
DTT
-
+
75
50
295
245
195
145
95
45
mAU
PUS1wt
PUS1D134A
670kD
158kD
44kD
17kD
1.4kD
0
20
40
60
80
100
120
mAU
b
R170
R263
R169
R168
R167
796
845
Fluc mRNA
739
787
IVT
R164
R165
R166
c
% pseudoU
20
15
10
5
0
59
31
31
30
29
30
40
40
RNA length (nt)
IVT
R164
R165
R166
R167
R168
R169
R170
d
% pseudoU
15
10
5
0
52.5 pmol
105 pmol
210 pmol
PUS1wt
PUS1D134A

Fig. 2

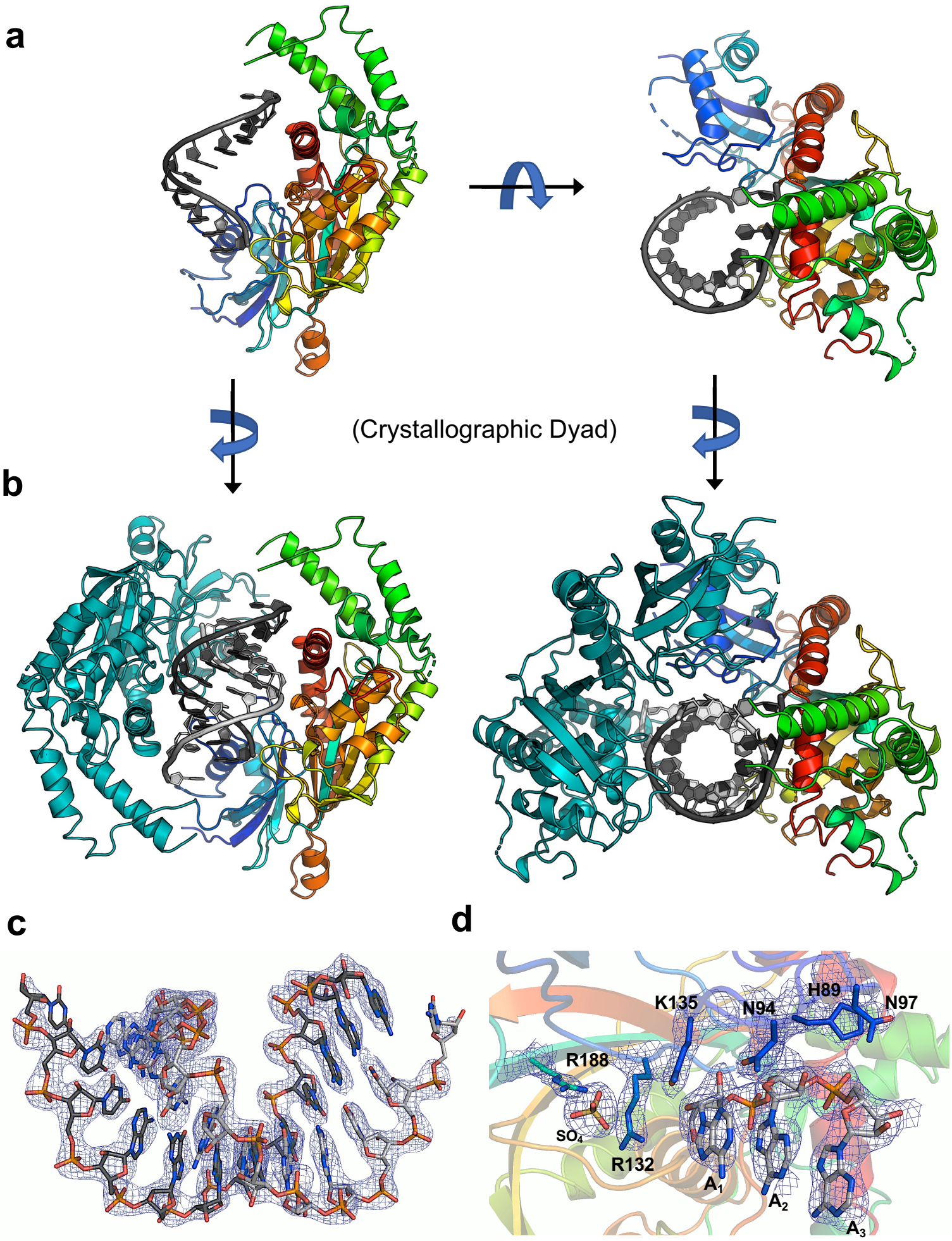

a
(Crystallographic Dyad)
b
c
d
K135
N94
H89
N97
R188
SO4
R132
A1
A2
A3

Fig. 3

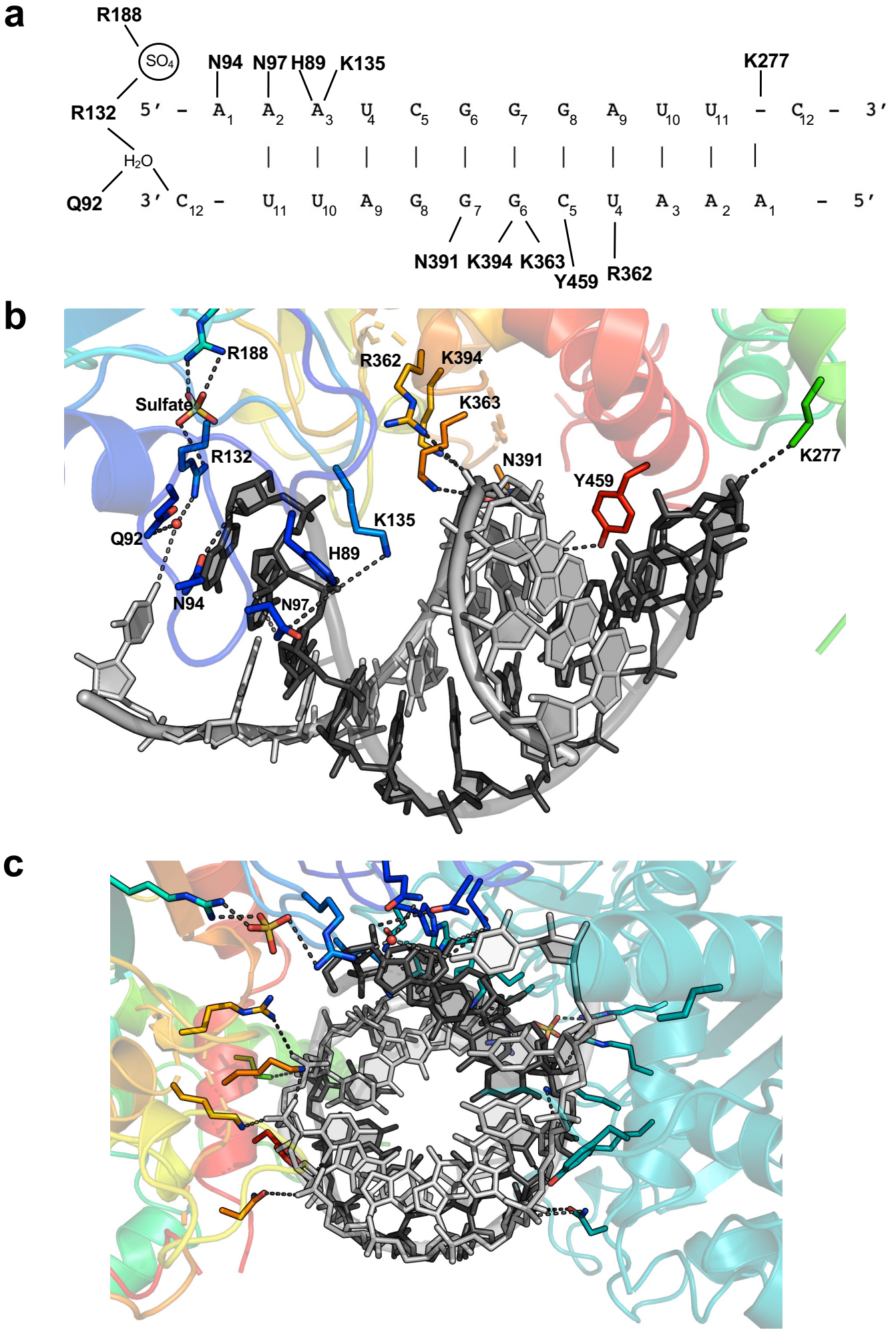

a
R188
SO4
N94
N97
H89
K135
K277
R132
5′ – A1 A2 A3 U4 C5 G6 G7 G8 A9 U10 U11 – C12 – 3′
H2O
Q92
3′ C12 – U11 U10 A9 G8 G7 G6 C5 U4 A3 A2 A1 – 5′
N391
K394
K363
Y459
R362
b
R188
Sulfate
R132
Q92
N94
N97
H89
K135
R362
K394
K363
N391
Y459
K277
c

Fig. 4

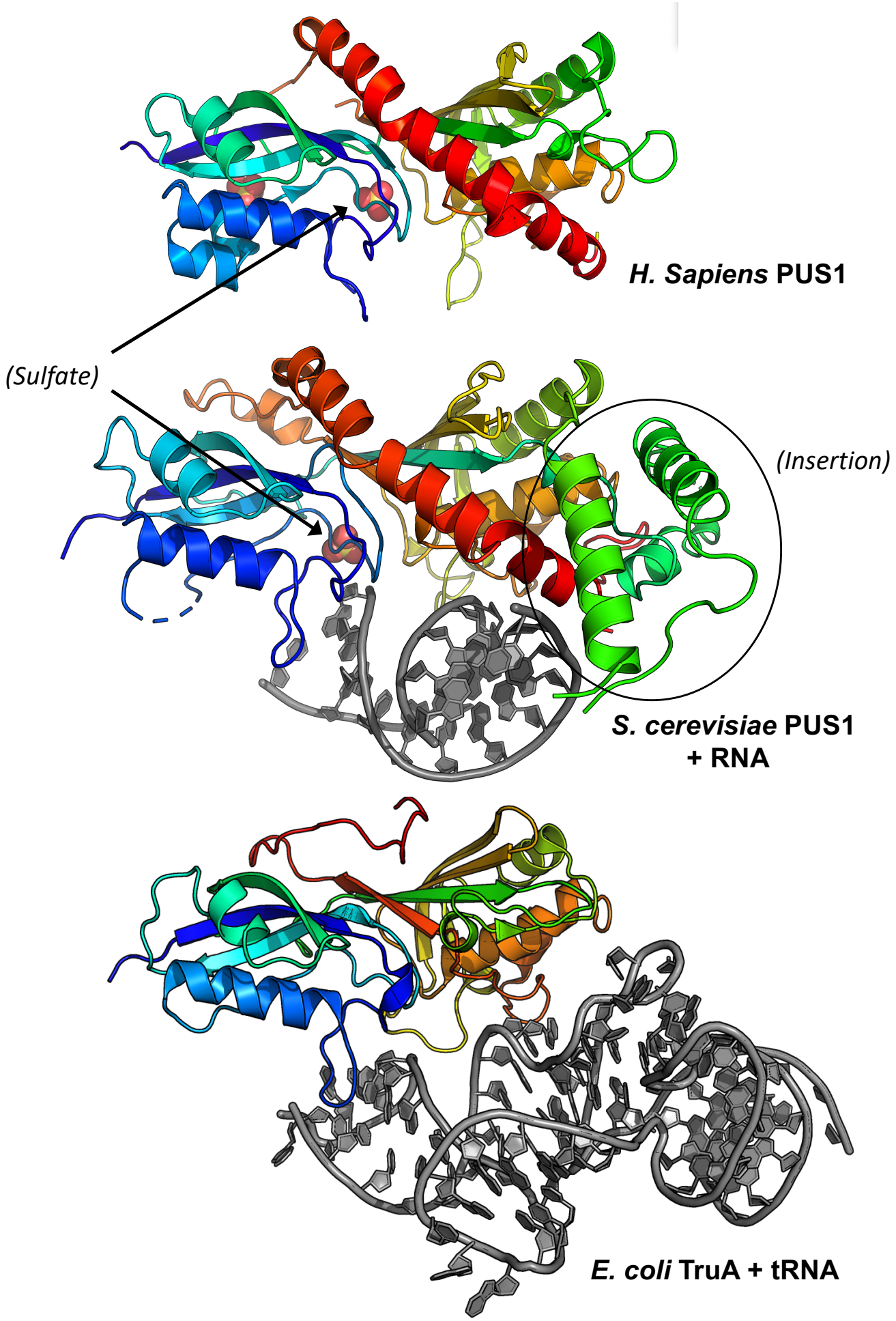
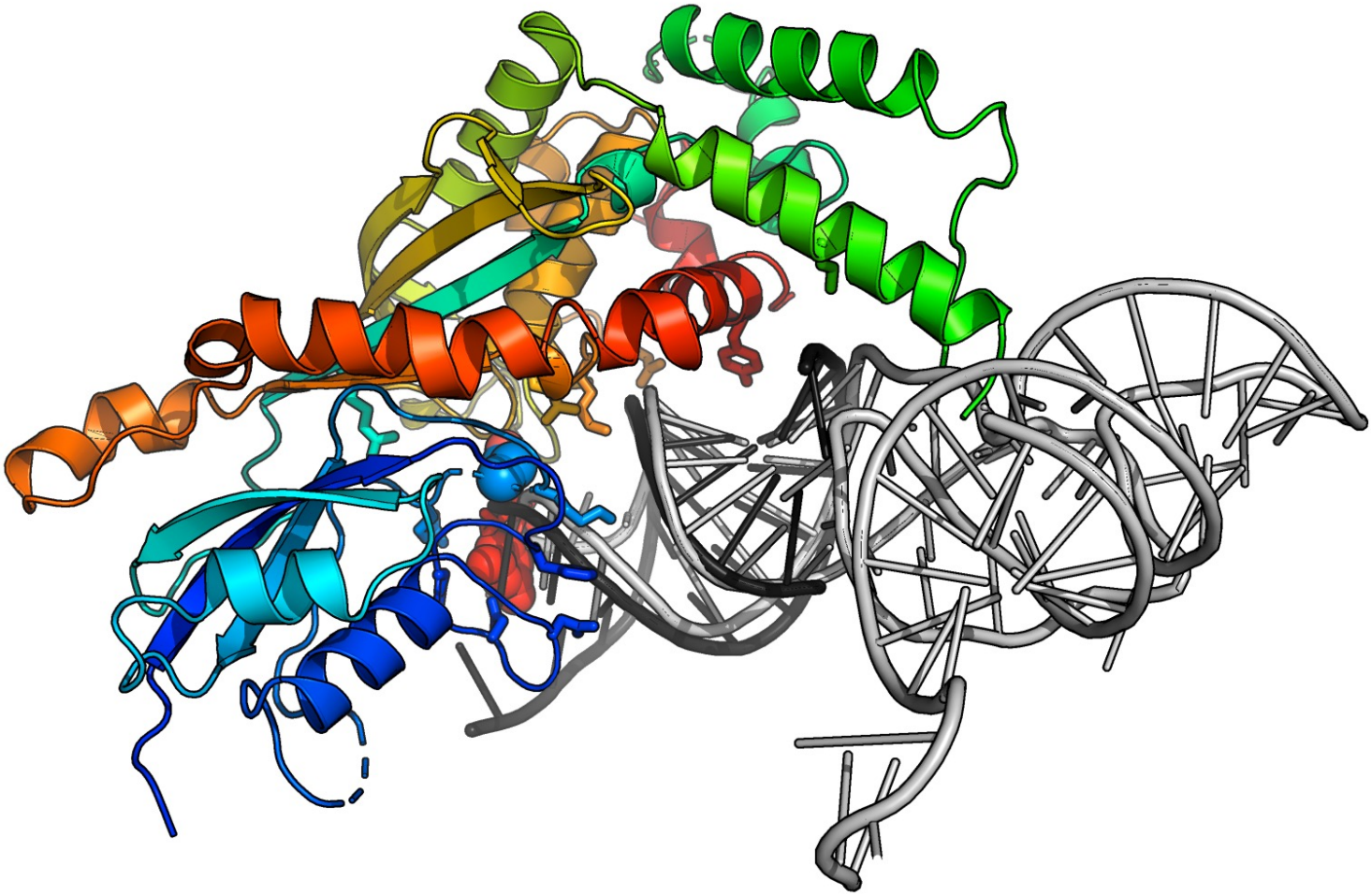

H. Sapiens PUS1
(Sulfate)
(Insertion)
S. cerevisiae PUS1
+ RNA
E. coli TruA + tRNA

Fig. 5



Supplemental Figure Captions

Fig. S1. PUS1 wildtype and D134A mutant bind R263 substrate. ***Panel a:*** Schematic representation of the PUS1 activity assay and analysis. ***Panel b***: EMSAs of 10 pmol 18 nt R263 with either increasing concentrations of PUS1wt (**b**) or $PUS1_{D134A}$ (**c**). The band showing the single-stranded and duplexed form of R263 is indicated. Wildtype PUS1 reactions were run on a native 10% TBE gel, while samples containing the D134A variant were run on a native 6% TBE gel.

Fig. S2. Sequence alignment of eukaryotic PUS1 homologues. Amino acid residues that we found to interact with the substrate RNA are highlighted by red boxes.

Fig. S3. Comparison of PUS1 / RNA complex structures. A Superposition and side-by-side comparison of $PUS1_{D134A}$ / RNA complex (colored protein backbones) solved in crystallographic space group C2 and wild-type PUS1 / RNA complex (grey protein backbones) solved in an unrelated, $P6_122$ space group. In both structures, two monomers of the enzyme are independently bound in a symmetric arrangement to an RNA duplex that is generated via a crystallographic 2-fold dyad symmetry axis. **B** Model and electron density for wild-type PUS1 and the region of the RNA containing the 5' 5-fluorouracil base. A 2Fo-Fc map (gray) is contoured at 1s. In the corresponding refinement, the base is estimated to display greater than 90% occupancy of the unflipped conformation, as modeled and shown.

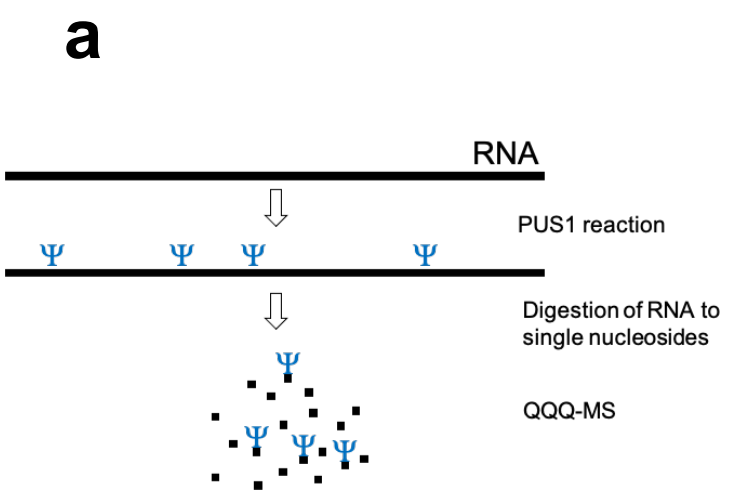
Fig. S4. Sequence alignment and RNA interactions of *S. cerevisiae* PUS1 and *E. coli* TruA. The conserved catalytic aspartate at position 134 in PUS1 and position 60 in TruA is indicated by a yellow star and box. RNA-interacting residues are highlighted in green (PUS1) and magenta (TruA).

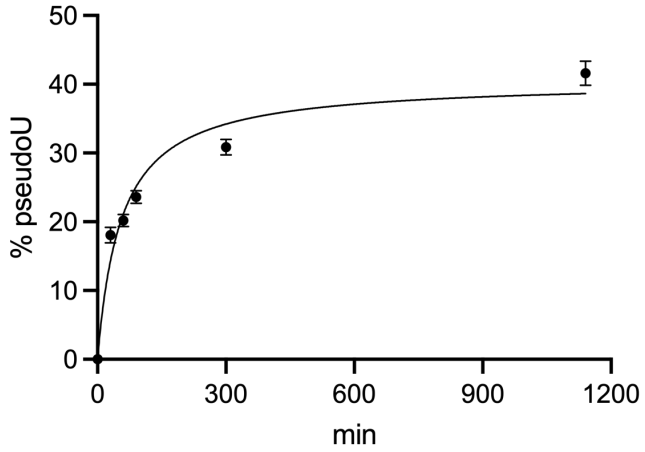
Table S1. Sequences of the RNA oligos used in this study.

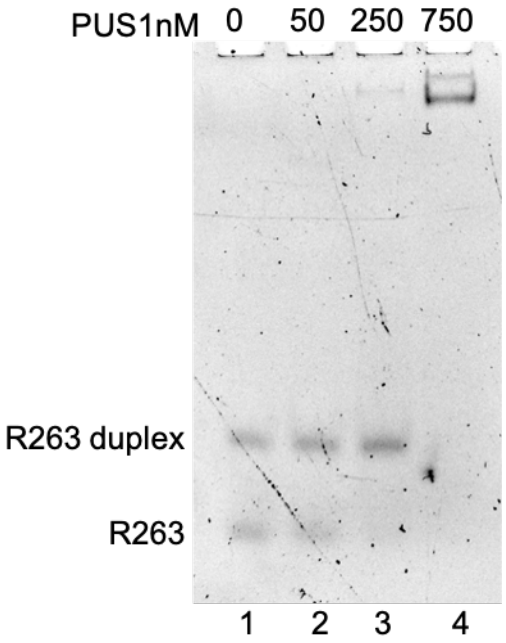
References

1. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008;36(Web Server issue):W70-4.

Fig. S1


a
RNA
PUS1 reaction
Digestion of RNA to single nucleosides
QQQ-MS


b
% pseudoU
min
0
10
20
30
40
50
0
300
600
900
1200


c
PUS1nM 0 50 250 750
R263 duplex
R263
1 2 3 4

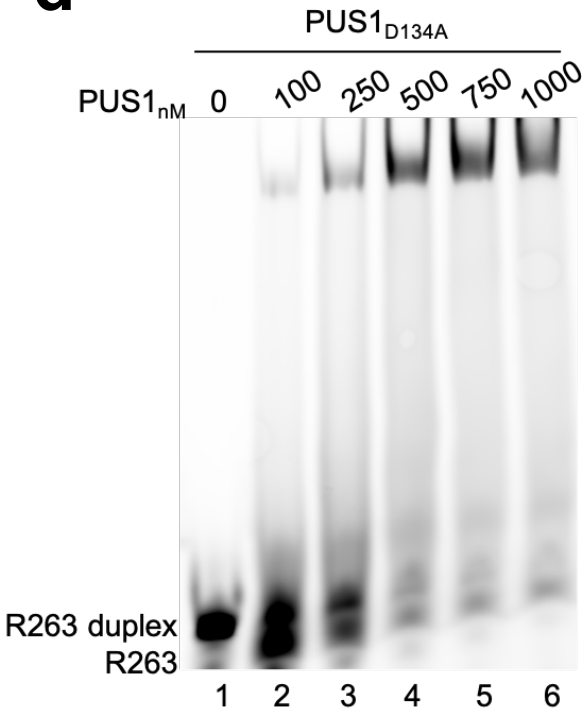
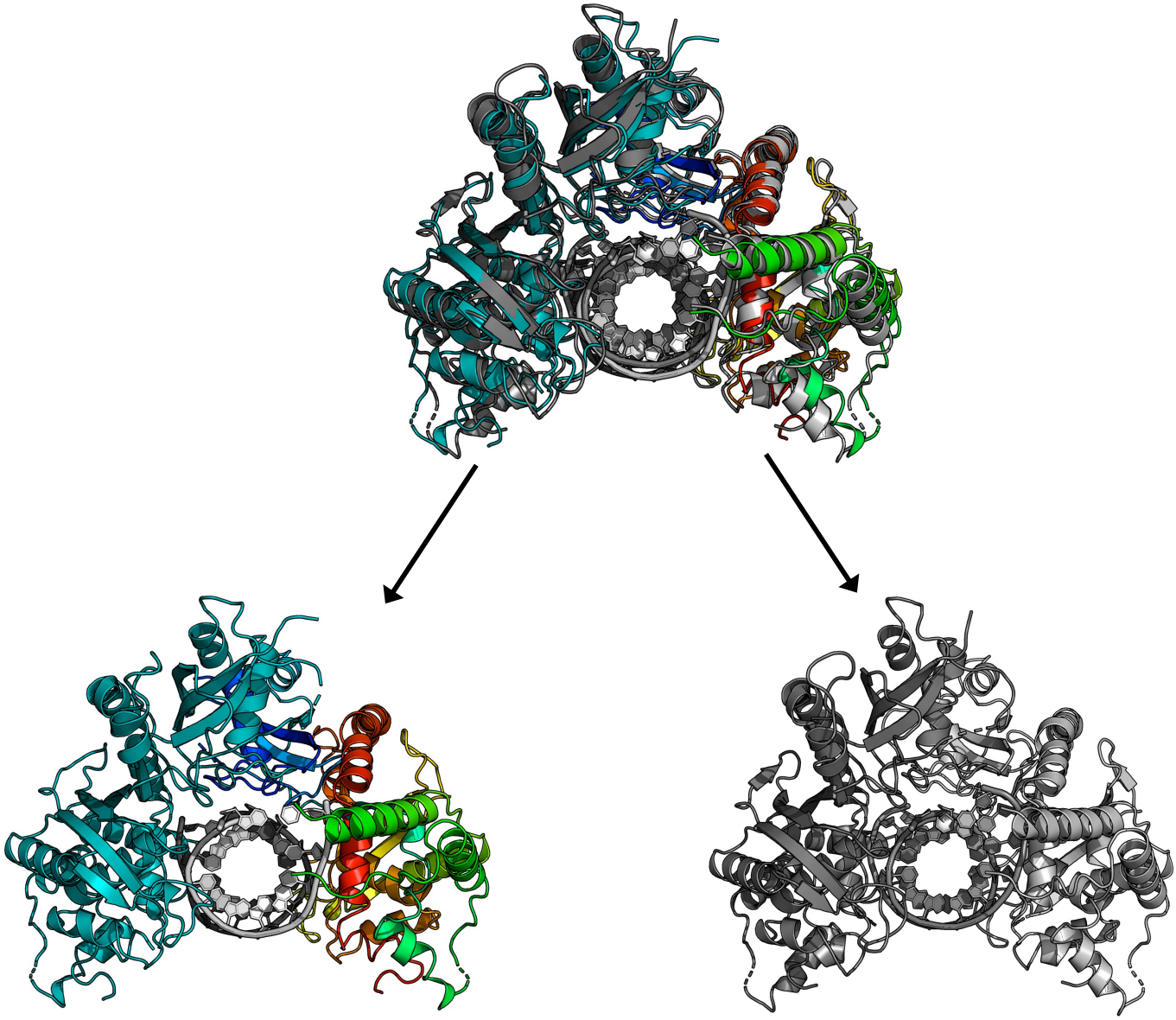

d
PUS1D134A
PUS1nM 0 100 250 500 750 1000
R263 duplex
R263
1 2 3 4 5 6

Fig. S2

```
                  -------XXXXXXXXXXXXXXRXXXXXXXXXXXXXXXXGXXXXXXXXRKXXXXXXXXXXXXXXXHXXKKX
S. cerevisiae     ---------------MSEENLRPAYDDQVNE--DVYKRGAQSKLTKARKADFDDEKDKKKDNDKHIDKRP   53
H. sapiens        MGLQLRALLGAFGRWTLRLGPRPSCSPRMAGNAEPPPAGAAC-PQDRRSCSGRAGGDRVWEDGEHPAKKL   69
X. laevis         ---------------------------MADLEKETLLETVT-PGEVRKASVLQEESGAQEENGHQAKKF   41
D. melanogaster   -------MSEALDKLEVEKEAR---------KEAKELEGEAVKDANRIKRTL---KRKKWVDWKEQDEKD   51

                  XX---XXXXGXXXXXXXXXPKRKXXXLXXYXGXGYHGMQRNXGXXXXXTIEXXLXXAXVXXGXIXXXXXX
S. cerevisiae     KSGPRLDENGNPLPKEPRLPKRKVAVMVGYCGTGYHGMQYNPPN---PTIESALFKAFVEAGAISKDNSN  120
H. sapiens        KS-------GGDEERREKPPKRKIVLLMAYSGKGYHGMQRNVGSSQFKTIEDDLVSALVRSGCIPENHGE  132
X. laevis         -K-------GDQEELCKKHPKRKVVLLMAYSGKGYHGMQRNVGSSQFKTIEDELVQALVSSGCIPDNHAD  103
D. melanogaster   AA---NGVKRAPFDPADRIKRKKSAILLSYCGANYYGMQRNPGM---QTIEEELFKAMLKHKWITEDSFE  115

                  XXXKXXFQRXARTDKGVSAAGQVXSLKXXXXDX-XXEKINXXLPXXIRXXGXXRVTXXFNXKXXCXARTY
S. cerevisiae     DLKKNGFMRAARTDKGVHAGGNLISLKMIIEDPDIKQKINEKLPEGIRVWDIERVNKAFDCRKMCSSRWY  190
H. sapiens        DMRKMSFQRCARTDKGVSAAGQVVSLKVWLIDD-ILEKINSHLPSHIRILGLKRVTGGFNSKNRCDARTY  201
X. laevis         EMKKMSFQRCARTDKGVSAAGQVVSLKIWLIDN-VVEKINENLPSNIRILGLKRVTGKFNSKNTCDARTY  172
D. melanogaster   QIQISCFQRAARTDKGVSAARQVCSVKLPE-EL-DLEAFNADLPQQIRLFGVERVTKGFNAKDQCNARTY  183

                  XYXLPTXAFXXXXXX-------------------------------------------------------
S. cerevisiae     EYLLPTYSLIGPKPGSILYRDIEESKTELPGVLDEDLESKEFWEEFKKDANEKFSTEEIEAILAYVPPAR  260
H. sapiens        CYLLPTFAFAHKDRD-------------------------------------------------------  216
X. laevis         SYTLPTFAFSHKDKE-------------------------------------------------------  187
D. melanogaster   TYTLPTVAFAPFEEK-------------------------------------------------------  198

                                                  ----------------XXXXXXXXRXSXEXLXXVXXXLXXYXGTHNFHNFTSXKXXXDPSAXRXIXX
S. cerevisiae     DEFDINEELYQKVKKYKQLENAHRRRYRISAAKLAKFRASTSQYLGAHNFHNFTLGKDFKEPSAIRFMKD  330
H. sapiens        ------------------V--QDETYRLSAETLQQVNRLLACYKGTHNFHNFTSQKGPQDPSACRYILE  265
X. laevis         ------------------T--QDENFRLSQETLNRVNELLALYKGTHNFHNFTSQKGPKDPSAKRYIME  236
D. melanogaster   ----------------VDDVHDTFRISPELLQKVKETLKLYEGTKNFHNFTSKKSFLDPSSKRFIMS  249

                  XXXXXPFXXX-XXXEXXXIKVKGQSFMXHQIRKMVGLXXAIVXGXXPXXXXERXXGXEKXDXPKAPGLGL
S. cerevisiae     IKVSDPFVIGDAQTEWISIKIHGQSFMLHQIRKMVSMATLITRCGCPVERISQAYGQQKINIPKAPALGL  400
H. sapiens        MYCEEPFVR--EGLEFAVIRVKGQSFMMHQIRKMVGLVVAIVKGYAPESVLERSWGTEKVDVPKAPGLGL  333
X. laevis         MLCEPPFQQ--GGLELAVIKVKGQSFMMHQIRKMIGLVIAVVKGFAPKSIIERSWGEEKVDIPKAPGLGL  304
D. melanogaster   FTSSEPFRSP-QDIEFVTLKVKGQSFMLHQIRKMVGLAIAIVRGNTTAATLERALTEERLDLPMAPGLGL  318

                  VLEXVHFEXYNXRXGXDGXHXXLXWXXXEXXVXXFKXXHIYXXIXXTEXXEXXMXXWXXTLXXHXXXXXX
S. cerevisiae     LLEAPVFEGYNKRLEQFGYK-AIDFSKYQDEVDKFKMKHIYDKIYKEEVDENVFNAFFSYIDSFNKVTG-  468
H. sapiens        VLERVHFEKYNQRFGNDGLHEPLDWAQEEGKVAAFKEEHIYPTIIGTERDERSMAQWLSTLPIHNFSATA  403
X. laevis         VLERVHFEKYNKRFGNDGLHDSLNWVEEEEKIEAFKKEHIYPTIIQTEIEEKSMITWLATLPIHDYDATS  374
D. melanogaster   VLDTVHYERYNDRYGKDGIHNPLTWQAQEAQVQEFIEREIFSQIYKTEAEQRNMLDWIGTLHYHSYDTRT  388

                  XXXXXXXXXXXXX-----------XXGXXXXXDXX-----------------------------------
S. cerevisiae     -AQGEETADKSGPAVQKSIFEFLTAKGIPGLTDAPESNKKIKQRKRMEEEEAASKKAEISSTTQSNEPEV  537
H. sapiens        LTAGGTGAKVPSP-----------LEGSEGDGDTD-----------------------------------  427
X. laevis         LQQNQAEEEEEE-------------------SD-------------------------------------  388
D. melanogaster   EDAPPPSSEDK-K-----------VKGGDDDNDE-----------------------------------  410

                  -------
S. cerevisiae     QPEAAAN  544
H. sapiens        -------  427
X. laevis         -------  388
D. melanogaster   -------  410
```

Fig. S3

a



b

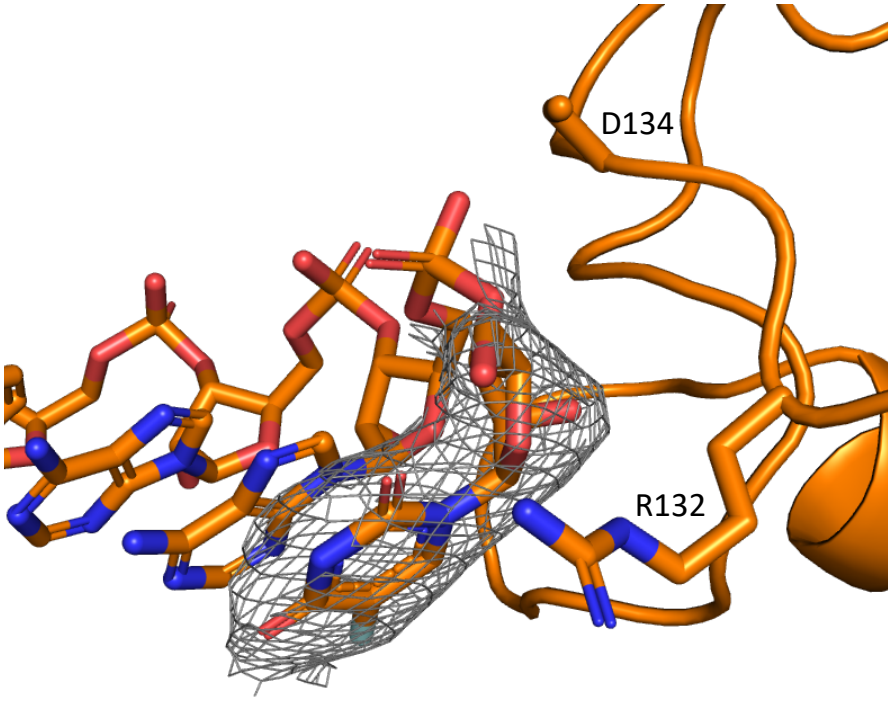


Fig. S4

```
scPUS1      MSEENLRPAYDDQVNEDVYKRGAQSKLTKARKADFDDEKDKKKDNDKHIDKRPKSGPRLD    60
ecTruA      ------------------------------------------------------------    0

scPUS1      ENGNPLPKEPRLPKRKVAVMVGYCGTGYHGMQYNPPNPTIESALFKAFVEAGAISKDNSN    120
ecTruA      -----MSDQQQPPVYKIALGIEYDGSKYYGWQRQNEVRSVQEKLEKALSQVA--------    47
                 : .: : *  *:*: : * *: *:* * :    :::. * **: :..

scPUS1      DLKKNGFMRAARTDKGVHAGGNLISLKMII--EDPDIKQKINEKLPEGIRVWDIERVNKA    178
ecTruA      -NEPITVFCAGRTDAGVHGTGQVVHFETTALRKDAAWTLGVNANLPGDIAVRWVKTVPDD    106
              :   .: *.*** ***. *::: ::     :*   .  :* :** .* *  :: * .

scPUS1      FDCRKMCSSRWYEYLLPTYSLIGPKPGSILYRDIEESKTELPGVLDEDLESKEFWEEFKK    238
ecTruA      FHARFSATARRYRYIIYNHRLRP-------------------AVLSKGV--THFYEPLDA    145
            *..*  .::* *.*:: .: *                     .**.:.:  ..*:* :.

scPUS1      DANEKFSTEEIEAILAYVPPARDEFDINEELYQKVKKYKQLENAHRRRYRISAAKLAKFR    298
ecTruA      ERM-------------------------------------------------------H    149
            :                                                         :

scPUS1      ASTSQYLGAHNFHNFTLGKDF-KEPSAIRFMKDIKVSDPFVIGDAQTEWISIKIHGQSFM    357
ecTruA      RAAQCLLGENDFTSFRAVQCQSRTPWRNVMHINVTRHGPY---------VVVDIKANAFV    200
             ::.  ** ::* .*   :   : *    :  ::.  .*:         : :.*:.::*:

scPUS1      LHQIRKMVSMATLITRCGCPVERISQAY-----------------------GQQKINIPK    394
ecTruA      HHMVRNIVGSLMEVGAHNQPESWIAELLAAKDRTLAAATAKAEGLYLVAVDYPDRYDLPK    260
             * :*::*.    :   . * . *::                            :: ::**

scPUS1      APALGLLLEAPVFEGYNKRLEQFGYKAIDFSKYQDEVDKFKMKHIYDKIYKEEVDENVFN    454
ecTruA      PPMGPLFLAD-------------------------------------------------    270
             *   *:*

scPUS1      AFFSYIDSFNKVTGAQGEETADKSGPAVQKSIFEFLTAKGIPGLTDAPESNKKIKQRKRM    514
ecTruA      ------------------------------------------------------------    270

scPUS1      EEEEAASKKAEISSTTQSNEPEVQPEAAAN 544
ecTruA      ------------------------------ 270
```

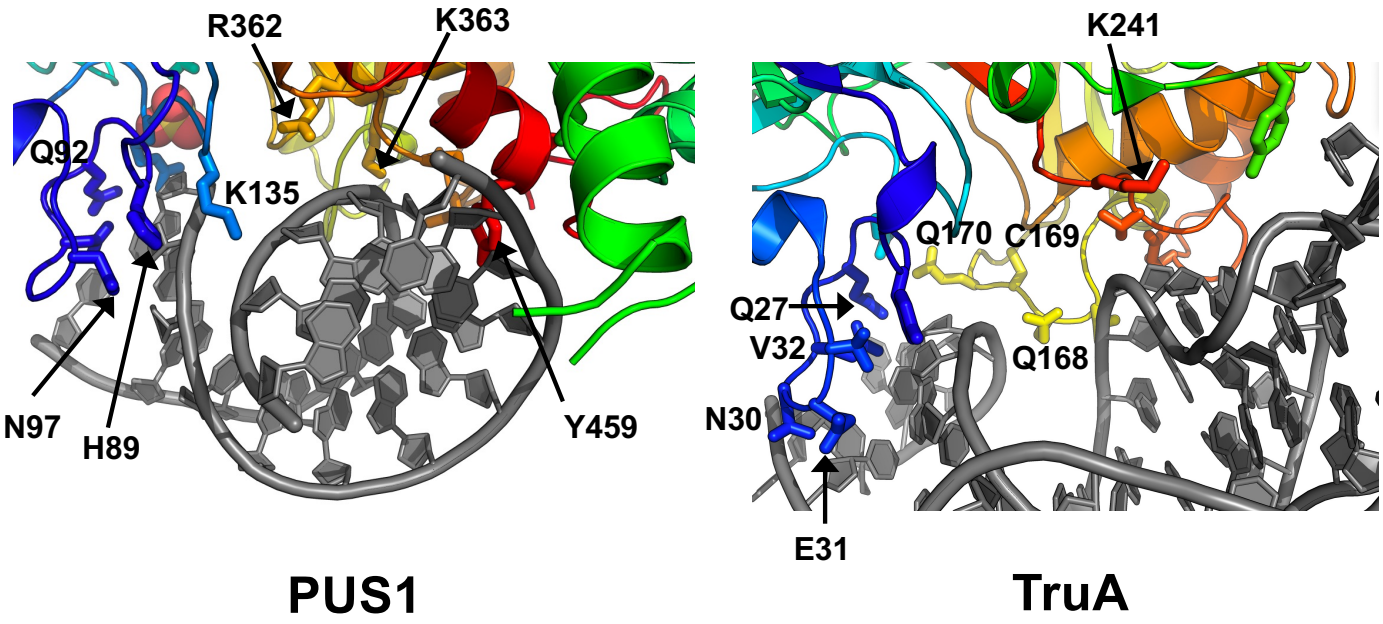


Table S1

name	sequence
IVT	AUUCCGGAUACUGCGAUUUUAAGUGUUGUUCCAUUCCAUCACGGUUUUGGAAUGUUUAC
R164	GGGAUUCCGGAUACUGCGAUUUUAAGUGUUG
R165	UUCCAUUCCAUCACGGUUUUGGAAUGUUUAC
R166	CGAUUUUAAGUGUUGUUCCAUUCCAUCACG
R167	UUUUUGGCAAUCAAAUCGGGAUUCCGGAUA
R168	GGGUUUUUGGCAAUCAAAUCAUUCCGGAUA
R169	AGAGAUCCUAUUUUUGGCAAUCAAAUCGGGAUUCCGGAUA
R194	AGAGAACCGAUUUUUGGCAAACAAAGCGGGAGCCCGGACA
R195	AGAGAACCGAACCGAGGCAAGCAAAUCGGGAUUCCGGAUA
R263	AAAUCGGGAUUCCGGAUA
R340	/i5F-U/AAUCGGGAUUCCGGAUA