8144–8154 *Nucleic Acids Research, 2012, Vol. 40, No. 16* *Published online 19 June 2012*
doi:10.1093/nar/gks548

How an exonuclease decides where to stop in trimming of nucleic acids: crystal structures of RNase T–product complexes

Yu-Yuan Hsiao[1], Yulander Duh[1,2], Yi-Ping Chen[1], Yi-Ting Wang[1] and Hanna S. Yuan[1,*]

[1]Institute of Molecular Biology, Academia Sinica, Taipei, 11529 and [2]Department of Biotechnology and Laboratory Science in Medicine, National Yang Ming University, Taipei, 11221, Taiwan, ROC

Received April 11, 2012; Revised May 14, 2012; Accepted May 15, 2012

ABSTRACT

Exonucleases are key enzymes in the maintenance of genome stability, processing of immature RNA precursors and degradation of unnecessary nucleic acids. However, it remains unclear how exonucleases digest nucleic acids to generate correct end products for next-step processing. Here we show how the exonuclease RNase T stops its trimming precisely. The crystal structures of RNase T in complex with a stem-loop DNA, a GG dinucleotide and single-stranded DNA with different 3′-end sequences demonstrate why a duplex with a short 3′-overhang, a dinucleotide and a ssDNA with a 3′-end C cannot be further digested by RNase T. Several hydrophobic residues in RNase T change their conformation upon substrate binding and induce an active or inactive conformation in the active site that construct a precise machine to determine which substrate should be digested based on its sequence, length and structure. These studies thus provide mechanistic insights into how RNase T prevents over digestion of its various substrates, and the results can be extrapolated to the thousands of members of the DEDDh family of exonucleases.

INTRODUCTION

Exonucleases are essential enzymes in almost every cellular event associated with nucleic acids, such as DNA replication, DNA repair, RNA maturation and RNA turnover. However, it remains largely unclear how exonucleases select their substrates, bind at the 5′- or 3′-end of nucleic acid chains or cleave 1 nt at a time to the correct site to generate the end product for next-step processes. Exonucleases may have exo-deoxyribonuclease and/or exo-ribonuclease activity targeting DNA and/or RNA substrates (1). Exo-ribonucleases are generally classified into six families, RNR, DEDDh, RBN, PDX, RRP4 and 5PX, within which the DEDDh family is one of the biggest groups, containing more than seven thousand members involved in various aspects of RNA and DNA processing in all kingdoms of life (2). The dysfunction or mutation of several human DEDDh exonucleases is linked to various diseases, such as WRN that is associated with Werner syndrome (3) and TREX1 that is associated with systemic lupus erythematosus, Aicardi-Goutières syndrome, retinal vasculopathy, cerebral leukodystrophy and familial chilblain lupus (4–6). Recently, DEDDh exonucleases have also been found to link viral infection with antiviral mechanism (7–9). It is therefore crucial to understand the molecular mechanism of how DEDDh family exonucleases select and digest their substrates.

Interestingly, most members of the DEDDh exonucleases have distinctive substrate preferences for specific structures or sequences of nucleic acids. Some members of the DEDDh exonuclease family are more specific for double-stranded RNA and/or DNA, such as NP exclusively degrading double-stranded RNA in Lassa fever virus (9) and CRN-4 degrading chromosomal DNA in *Caenorhabditis elegans* (10). Some other members prefer to bind and degrade double-stranded nucleic acids with a 3′-overhang of 2–5 nt, including Snp from *Drosophila* (11), and 3′hExo (12,13) and ERI-1 (14) from *C. elegans*. On the other hand, some members have preferences for single-stranded RNA and/or DNA, including *Escherichia coli* ExoI (15,16) and human ISG20 (17,18), TREX1 and TREX2 (1). Some DEDDh nucleases are restricted by the length of the nucleic acids, such as oligoribonuclease, which only degrades small oligoribonucleotides of 2–5 nt into mononucleotides (19). Besides structural specificity, DEDDh exonucleases may have sequence preferences, such as human poly(A)-specific ribonuclease (PARN) (20) and yeast Pop2 (21), which specifically digest the poly-A tails of mRNA during RNA turnover. Even though numerous DEDDh exonucleases function in diverse cellular events, the working mechanism of these

*To whom correspondence should be addressed. Tel: +886 2 27884151; Fax: +886 2 27826085; Email: hanna@sinica.edu.tw

exonucleases in selection and digestion of their target substrates is mostly unknown.

Among the members of the DEDDh family of exonucleases, RNase T has perhaps the most unique substrate specificity, limited by not only the structure but also by the sequence of a nucleic acid (22). RNase T digests both DNA and RNA, and its activity is inhibited by duplex structure, referred to as the double-strand effect, as well as a 3′-terminal cytosine, referred to as the C effect. A single 3′-terminal C at the end of a single-stranded nucleic acid reduces its activity by 100-fold, and a double 3′-terminal CC almost abolishes its enzymatic activity (22,23). In *E. coli*, RNase T performs the final trimming for a number of small stable RNAs during 3′-end maturation, including tRNA, 5S rRNA and 23S rRNA (24–26). The mature RNAs trimmed by RNase T all have their 5′-end annealed with their 3′-end forming a duplex with a short 3′-overhang in a size ranging from 1 to 4 nt. In the digestion of single-stranded DNA and RNA, RNase T produces an end product of a dinucleotide (22,27).

Our previous crystal structures of RNase T complexed with ssDNA (3′-terminal G versus C) revealed the working mechanism for its C effect (28). The binding of a single-stranded DNA with a 3′-terminal C to RNase T induces an inactive conformational change at the active site and thus inactivates the exonuclease activity. The crystal structure of RNase T in complex with a duplex DNA further showed that the dimeric RNase T has an ideal architecture for binding and degrading a duplex substrate with a short 3′-overhang (28). Because a duplex with a short 3′-overhang of <2 nt binds to RNase T weakly, and also a short 3′-overhang cannot reach the active site for digestion that results in the double strand effect. Moreover, the last AT base pair in the duplex region of the complex crystal structure was melted, suggesting that the duplex with an AT base pair at the end of the duplex can be cleaved more closely to the duplex region, generating a 1-nt 3′-overhang, contrasting with a 2-nt 3′-overhang if the last base pair is a GC. Combining the C effect and double-strand effect, RNase T digests various stable RNAs to produce mature RNA with a precise 3′-overhang: 1 nt for 5S rRNA, 2 nt for 23S rRNA, 4 nt for 4.5S RNA and 4 nt for tRNA.

However, it is still not clear how RNase T decides where to stop trimming its substrates, and in particular why RNase T generates a dinucleotide end product in the digestion of a single-stranded nucleic acid, but conversely generates a 1- to 4-nt 3′-overhang in the digestion of duplex substrates with different sequences. Here we report seven new RNase T–DNA complex crystal structures to provide a complete list of RNase T in complex with not only its single-stranded substrates/products with different 3′-terminal sequences (AA, TA, AT versus AC, CC), but also with the dinucleotide end product (GG), and a stem-loop DNA more closely mimicking natural duplex products. These studies thus present a full spectrum of RNase T in complex with different single- and double-stranded substrates and products, and provide mechanistic insights into how RNase T binds and prevents over digestion of its various substrates. Additionally we derive a general principle for how RNase T trims various sequences of single- and double-stranded nucleic acids. These results can be extrapolated to the molecular mechanisms of thousands of DEDDh family exonucleases.

MATERIALS AND METHODS

Protein expression and purification

The constructs of the full-length *rnt* gene and E92G mutated *rnt* gene were generated as described (28). The pET28-RNase T plasmid was transformed into *E. coli* BL21-CodonPlus(DE3)-RIPL strain (Stratagene, USA) cultured in LB medium supplemented with 35 μg/ml kanamycin. Cells were grown to an OD_{600} of 0.5 and induced by 0.8 mM IPTG at 18°C for 20 h. The harvested cells were disrupted by a microfluidizer in 50 mM Tris–HCl (pH 7.5) containing 300 mM NaCl. RNase T and mutant proteins were purified by chromatographic methods using the standard Ni^{2+}-chelating protocol (Qiagen Inc., USA), followed by a HiTrap Heparin column (GE Healthcare, USA) and a gel-filtration column (Superdex 75, GE Healthcare, USA). Purified RNase T samples were concentrated to 25–30 mg/ml in 300 mM NaCl and 50 mM Tris–HCl (pH 7.0).

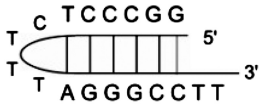
DNase activity assays

Three stem-loop DNA oligonucleotides were used for nuclease activity assays, labeled at the 5′-end with [γ-^{32}P]ATP by T4 polynucleotide kinase. The sequence of stem loop DNA was 5′-GGCCCTCTTTAGGGCCT XX-3′ where XX was AA, AC or CC. The isotope-labeled substrates were purified on a Microspin G-25 column (GE Healthcare, USA) to remove the non-incorporated nucleotides. Purified substrates (20 nM) were incubated with RNase T at various concentrations in a buffer solution of 120 mM NaCl, 2 mM $MgCl_2$ and 50 mM Tris–HCl (pH 7.0) at room temperature for 30 min. The reaction was quenched by the addition of the stop solution (2x Tris–Borate–EDTA buffer) and heating at 95°C for 5 min. Reaction samples were then resolved on 20% denaturing polyacrylamide gels and visualized by autoradiography (Fujifilm, FLA-5000).

Crystallization and crystal structural determination

Wild-type RNase T (25–30 mg/ml) or E92G mutant (25–30 mg/ml) in 300 mM NaCl and 50 mM Tris–HCl (pH 7.0) was mixed with different single-stranded DNA and stem-loop DNA substrates in the molar ratio of 1: 1.2. Detailed crystallization conditions and DNA sequences are listed in Supplementary Table S1. All crystals were cryo-protected by Paraton-N (Hampton Research, USA) for the data collection at 100 K. X-ray diffraction data were collected at BL13B1 and BL13C1 beamlines at NSRRC, Taiwan, or at the BL44XU beamline at SPring-8, Japan. All diffraction data were processed by HKL2000 (29) and the diffraction statistics are listed in Table 1. Structures were solved by the molecular replacement method using the crystal structure of *E. coli* apo RNase T (E92G mutant, PDB accession code: 3NGY)

Table 1. Data collection and refinement statistics for the crystal structures of RNase T–DNA complexes

	Product complex				Substrate complex		
PDB code	3VA3	3VA0	3V9Z	3V9S	3V9X	3V9W	3V9U
	Stem-loop 2nt 3′-overhang	Dinucleotide GG	3′-end AAC	3′-end ACC	3′-end AAA	3′-end TTA	3′-end AAT
RNase T used in crystallization	E92G mutant	Wild-type	Wild-type	Wild-type	Wild-type	E92G mutant	E92G mutant
DNA in crystals (visible in **bold**)		5′-**GG**-3′	5′-TTAC**AAC**-3′	5′-TTAC**ACC**-3′	5′-TTAT**AAA**-3′	5′-GC**TTA**-3′	5′-TTAC**AAT**-3′
Metal ions in active site (per molecular)	none	1 Mg^{2+}	1 Mg^{2+}	1 Mg^{2+}	2 Mg^{2+}	2 Mg^{2+}	2 Mg^{2+}
Data collection							
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P3_2$	$P3_2$	$P3_2$
Cell dimensions							
a, *b*, *c* (Å)	100.27, 112.68, 57.79	95.69, 105.81, 47.07	96.06, 105.84, 46.92	96.42, 106.09, 46.95	46.28, 46.28, 314.38	46.34, 46.34, 315.35	46.25, 46.25, 313.65
Resolution (Å)	30.0–2.7 (2.8–2.7)	30.0–2.2 (2.27–2.20)	50.0–2.1 (2.18–2.10)	30.0–1.8 (1.83–1.80)	30.0–1.9 (1.97–1.90)	30.0–1.7 (1.76–1.70)	30.0–2.3 (2.38–2.30)
R_{sym}	0.102 (0.412)	0.074 (0.232)	0.11 (0.487)	0.091 (0.362)	0.060 (0.353)	0.054 (0.352)	0.071 (0.320)
Average $I/\sigma I$	19.4 (2.24)	10.0 (2.5)	15.9 (3.5)	31.9 (8.6)	25.3 (4.0)	23.9 (3.0)	21.1 (3.9)
Completeness (%)	94.3 (93.5)	97.0 (95.2)	99.4 (99.9)	96.4 (100.0)	98.9 (99.9)	97.3 (99.7)	99.7 (99.8)
Redundancy	3.2 (2.6)	3.4 (3.0)	4.8 (4.7)	5.4 (5.8)	3.0 (3.1)	2.6 (2.7)	3.2 (3.2)
Refinement							
Resolution (Å)	28.9-2.7	28.4-2.2	25.7-2.1	28.5-1.8	23.3-1.9	26.4-1.7	23.2-2.3
No. reflections	17 239/1358	24 224/1879	28 443/2206	43 648/2201	58 768/4567	80 817/6243	33 262/2571
R_{work}/R_{free}	26.3/29.0	20.6/24.3	20.7/24.6	21.1/24.8	18.5/22.3	19.2/22.6	18.5/24.5
R.m.s. deviations							
Bond lengths (Å)	0.003	0.002	0.003	0.006	0.006	0.004	0.004
Bond angles (°)	0.685	0.588	0.712	1.000	0.930	0.872	0.825

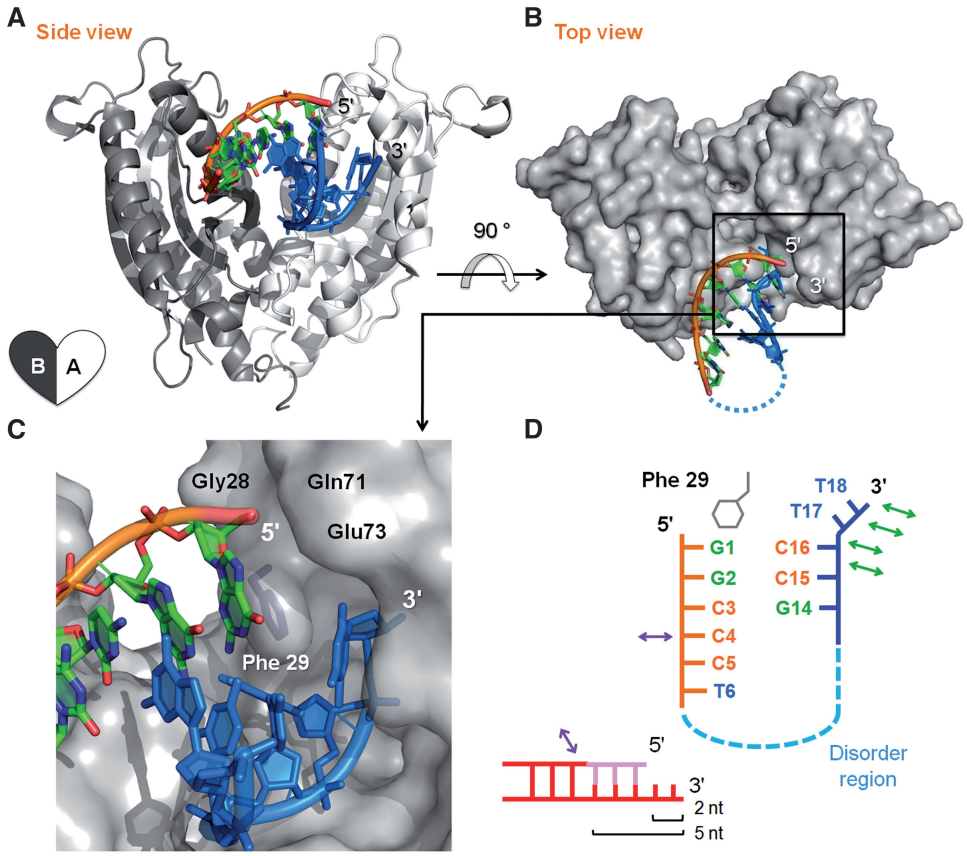


Figure 1. The crystal structure of RNase T bound to a stem-loop DNA product with a 2-nt 3′-overhang. (**A**) The ribbon model of RNase T bound with a stem-loop DNA displayed in a side view. The schematic heart-shaped diagram at the bottom left represents that RNase T is a dimeric protein with two protomers A and B. Nucleotides 1–6 near the 5′-end are shown in orange and nucleotides 14–18 near the 3′-end are shown in blue. (**B**) Molecular surface of RNase T bound to the stem-loop DNA displayed in a top view. Disordered region of the stem-loop DNA is shown by a blue dotted line. (**C**) Phe29 stacking with the 5′-end nucleotide G1 and the 'wall' made of Gly28, Gln71 and Glu73 stop further digestion at the 3′-end of the duplex by RNase T. (**D**) Schematic diagram of the interactions between the stem-loop DNA and RNase T. The interactions between RNase T and DNA are displayed by green arrows (protomer A) and purple arrows (protomer B). The schematic diagram in the bottom shows the rationale for the preference of RNase T in digesting a duplex substrate with a 3′-overhang of 2 to 5 nt.

as the search model by program MOLREP of CCP4 (30). The models were built by Coot (31) and refined by Phenix (32). Structural coordinates and diffraction structure factors have been deposited in the RCSB Protein Data Bank with the PDB ID codes of 3VA3, 3VA0, 3V9Z, 3V9S, 3V9X, 3V9W and 3V9U (Table 1).

RESULTS

Crystal structure of RNase T bound to a stem-loop DNA

Our previous crystal structure of RNase T in complex with duplex DNA (PDB entry: 3NH2) showed that each protomer of the dimeric RNase T binds one strand of DNA (28). However, the duplex DNA with an AT base pair at the end of the duplex in the previous structure was not in a classical Watson–Crick base-paired conformation and therefore cannot fully mimic a duplex substrate. Furthermore, it would be intriguing to confirm that RNase T is unable to melt the GC base pair at the end of the duplex as we suggested previously. We thus further cocrystallized RNase T with an 18-nt stem-loop DNA with a 2-nt 3′-overhang (5′-GGCCCTCTTTAGGGCCTT-3′, paired bases are underlined) in which the last base pair in the duplex region was a GC rather than AT. The RNase T mutant E92G was used here for cocrystallization with the stem-loop DNA because this mutant could be expressed in a large quantity and shared a similar enzyme activity to that of wild-type RNase T. The complex crystal diffracted X-rays to a resolution of 2.7 Å, and the structure was solved by molecular replacement using the structure of apo-RNase T as the searching model (PDB: 3NGY). There were two dimeric RNase T molecules per asymmetric unit, each bound to two stem-loop DNA (nucleotides 1–6 and 14–18 visible in one of the DNAs, and nucleotides 15–18 visible in the other DNA). For clarity, only one stem-loop DNA bound at the one of the active sites of the RNase T dimer is shown in Figure 1. All the diffraction and refinement statistics are listed in Table 1 and the crystallization conditions are listed in Supplementary Table S1.

The overall crystal structure of RNase T bound with the stem-loop DNA showed that the 'non-scissile' strand (orange, nucleotide 1–6) of DNA was bound to

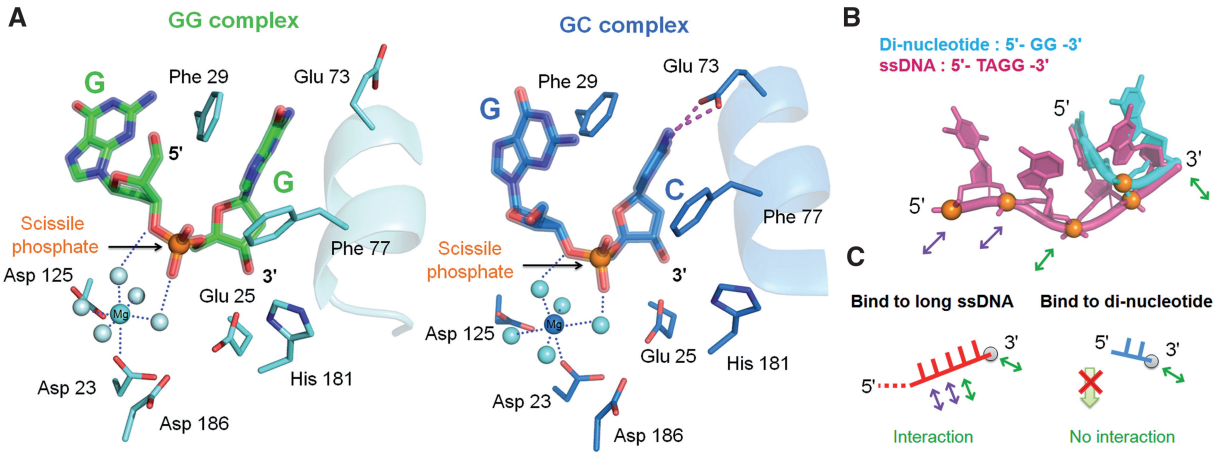


Figure 2. The active site of RNase T-dinucleotide complexes in an inactive conformation. (**A**) Inactive conformation of the active site of RNase T bound to the dinucleotide GG and GC. The two active sites share a similar conformation with the only difference in Glu73, which rotates its side chain to hydrogen bonds (red dotted lines) with the 3′-C in the GC complex. Water molecules are displayed as light blue balls. (**B**) The superposition of the dinucleotide GG onto a longer ssDNA substrate (7 nt) bound at RNase T shows that the dinucleotide shifts its position. The GG dinucleotide (PDB entry: 3VA0) is colored in blue, and the long single-stranded DNA (PDB entry: 3NH1) is colored in red with phosphates colored in orange. The interactions between DNA and RNase T are marked by green (protomer A) and purple (protomer B) arrows. (**C**) The schematic diagram of the interactions between RNase T and a single-stranded DNA, 7 nt versus 2 nt. The dinucleotide interacts with RNase T less extensively, and therefore it shifts its position and cannot be further cleaved into mononucleotides. See Supplementary Figure S2 for a detailed comparison and interactions.

protomer B and the 'scissile' strand (blue, nucleotide 14–18) was bound to protomer A (Figure 1). The 2-nt 3′-overhang was inserted deeply into the narrow active site of protomer A. The major interactions between RNase T and the duplex substrates were located at the 3′-terminal end of the scissile strand in which all the four 3′-terminal residues (C15, C16, T17, T18) interacted with protomer A extensively. The overall binding mode and the DNA conformation were different from those in the previously reported RNase T–DNA duplex structure (PDB entry: 3NH2) (see Supplementary Figure S1 for comparison). The DNA in the current structure had all the visible base pairs of a classical Watson–Crick conformation mimicking more closely a natural duplex substrate.

The last base pair in the duplex region, G1–C16 base pair, was indeed not melted by RNase T. The 5′-end of the 'non-scissile' strand was blocked by a wall made of Gly28, Gln71 and Glu73 (Figure 1C). Moreover, the guanine (G1) at the 5′-end of the duplex was stacked with the aromatic side chain of Phe29 (Figure 1C and D). The stacking caused Phe29 to shift its position, and the movement induced a disrupted conformational change at the active site. As a result, the stem-loop DNA with a 2-nt 3′-overhang was an end product that cannot be further cleaved by RNase T. In summary, the barriers made by the 'wall' and Phe29 block the 5′-end of the 'non-scissile' strand and prevent excessive cleavage of a duplex substrate by RNase T.

Besides the steric hindrance at the 5′-end of the 'non-scissile' strand, only one interaction was observed between the enzyme (Asn141 of protomer B) and the non-scissile strand (the phosphate of the fourth nucleotide C4), suggesting that RNase T recognizes a duplex substrate with a short 3′-overhang of 2–5 nt (Figure 1D). A duplex with a long 3′-overhang of >5 nt should be recognized as a single-stranded substrate, and a duplex with a short 3′-overhang cannot bind well to RNase T due to the loss of the interactions at the 3′-end. This result fully echoes the fact that the natural substrates of RNase T are duplex RNA with a 3′-overhang of 2–5 nt.

Crystal structure of RNase T bound to a dinucleotide end product

Previous activity assays showed that RNase T has length specificity, but does not show any preference for digestion of dinucleotide substrates (22). Single-stranded RNA and DNA were digested by RNase T to generate end products of dinucleotides of any sequences, such as CC, UU or AA (22,23). To understand how a dinucleotide stops the exonuclease activity of RNase T, we further cocrystallized RNase T with a GG dinucleotide and determined the crystal structure at a resolution of 2.2 Å (Table 1).

In the GG dinucleotide complex, one Mg^{2+} ion was bound in the active site, and it was coordinated to five water molecules and Asp23 in an octahedral geometry (Figure 2A). Comparison with the previously solved 7-nt substrate complex with two Mg^{2+} ions bound in the active site (PDB entry: 3NH1) revealed that the 2-nt product complex had an active site in a disrupted conformation. The scissile phosphate was moved up and the general base residue, His181, was shifted away from the active site (Supplementary Figure S2A). The structure of the active site of the GG complex closely resembled the previously solved structure of RNase T in complex with a GC dinucleotide (PDB entry: 3NGZ) (28). The only difference between the two structures was the rotation of the side chain of Glu73, which formed hydrogen bonds with the 3′-terminal C in the GC complex, but not in the GG

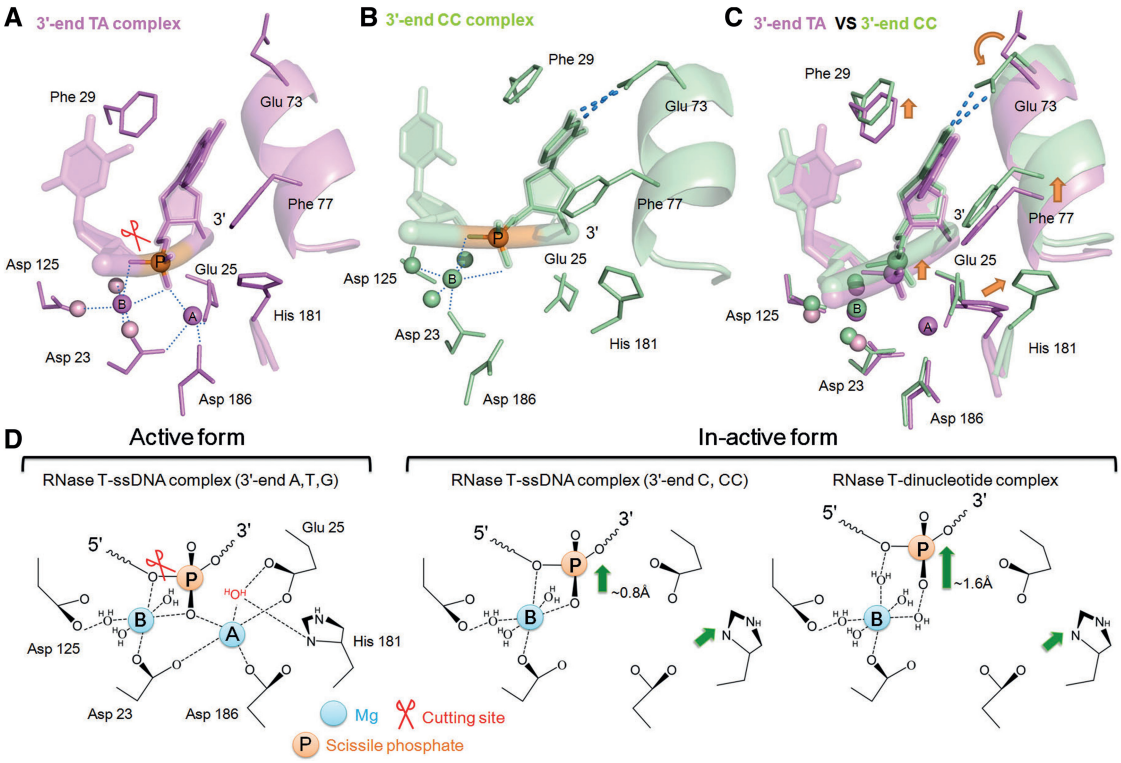


Figure 3. Crystal structures of RNase T bound to single-stranded substrates (3′-end AA, TA or AT) and products (3′-end AC and CC). (**A**) Active site of RNase T bound with a single-stranded DNA with a 3′-end TA reveals an active conformation with two Mg^{2+} ions and a nearby general base His181. The ssDNA with a 3′-end AT and GG bind to RNase T in a similar way as shown in Supplementary Figure S3. The red scissor marks the cutting site. (**B**) Active site of RNase T bound with a single-stranded DNA product with a 3′-end CC reveals an inactive conformation with only one Mg^{2+} ion and a shift-away His181. Hydrogen bonds between Gln73 and the 3′-end C are indicated by blue dotted lines. (**C**) Superposition of the active site of RNase T bound with a single-stranded substrate (3′-end TA) and product (3′-end CC) shows the shifts in the product complex as marked by the orange arrows. (**D**) Schematic diagrams of the three types of active sites in RNase T: substrate-bound active form and product-bound inactive forms (3′-end C and dinucleotide). The nucleophilic water molecule (colored in red) between metal B and the general base His181 was not present in the structures. Green arrows show the shifts induced upon product binding.

complex (Figure 2A). Previously we suggested that the inactive conformation of the active site in the GC complex resulted from the presence of a 3′-terminal C. Here the structure of the GG complex also revealed a similarly inactive conformation, suggesting that the conformational change resulted not from the C effect, but from the binding of a dinucleotide end product. We therefore conclude that dinucleotides with any sequence can induce an inactive conformational change at the active site upon binding to RNase T and thereby block the exonuclease activity of RNase T.

We next asked how a dinucleotide can induce such an inactive conformational change at the active site of RNase T upon binding. To answer this question, we further superimposed the dinucleotide complex with the substrate complex of RNase T bound with a 7-nt ssDNA (PDB entry: 3NH1). In the substrate complex, RNase T interacted extensively with the long single-stranded DNA, binding not only the 3′-end nucleotide, but also the second, third, fourth and fifth nucleotides from the 3′-end (Figure 2B, C and Supplementary Figure S2B). In contrast, in the dinucleotide product complex, only four Phe residues stacked with the penultimate and the 3′-end nucleotides. Because the interactions at the third, fourth and fifth nucleotides were missing, the dinucleotide was shifted and not bound at a position appropriate for digestion. The binding of a dinucleotide to RNase T thus induces a disrupted conformational change at the active site and inhibits enzyme activity.

Crystal structures of RNase T bound to single-stranded substrates and products

We have reported the crystal structure of RNase T in complex with a single-stranded DNA with a sequence of TTATAGG (PDB entry: 3NH1). In this substrate complex, two Mg^{2+} ions were bound in the DEDDh active site (28). Since RNase T prefers to digest single-stranded nucleic acids with a 3′-end G, A and T, but is inhibited by 3′-terminal C and CC, we further determined five complex crystal structures of RNase T bound with single-stranded DNA (5–7 nt) with a 3′-end AA, TA or AT (substrate complexes), and AC or CC (product complexes). The crystal structures were determined at a high resolution of 1.7–2.2 Å, revealing defined electron density for the bound metal ions and nucleotides (Table 1).

The three substrate complexes with a 3′-end sequence of AA, TA and AT all had structures similar to the

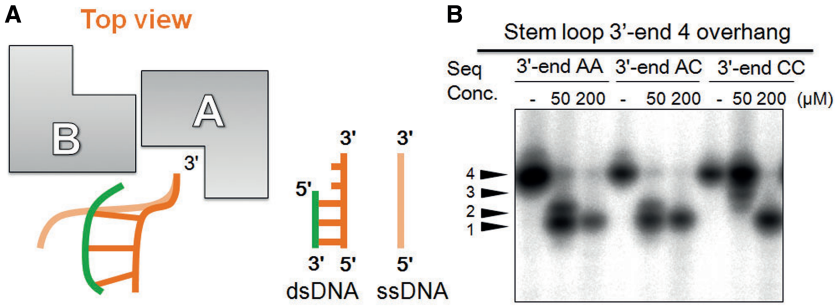


Figure 4. The different binding and cleavage modes of RNase T for single- and double-stranded substrates. (**A**) In binding a single-stranded substrate, both protomers A and B of RNase T interact with the same strand of the substrate, whereas in binding a double-stranded substrate, protomer A and B interact with different strands of the substrate. (**B**) Stem-loop DNA with a 4-nt 3′-overhang of different sequences were digested by RNase T. The stem-loop DNA with a 3′-end CC was most resistant to digestion as compared to the ones with a 3′-end AA or AC. The length of the 3′-overhang of the stem-loop DNA is indicated by arrows to the left of the figure.

previously reported structure of RNase T bound to a single-stranded DNA with a 3′-end GG (PDB entry: 3NH1). The five DEDDh catalytic residues, Asp23, Glu25, Asp125, His181 and Asp186, were in a classical DEDDh exonuclease geometry with two metal ions bound in the active site (Figure 3A and Supplementary Figure S3A). These two metal ions were assigned as Mg^{2+} based on their geometries which were different from those of monovalent ions (33). The nucleophilic water molecule between His181 and metal ion A was missing in these structures, likely due to the protonation of the general base residue His181 by the acidic conditions used for the cocrystallization experiments, and thus the three DNA substrates were not cleaved in the crystals. A similar result was reported for the general base His162, which was protonated (without a bound water) at pH 5.8, but deprotonated (with a bound water) at pH 8.5, in the DEDDh active site of *E. coli* ε subunit of DNA Polymerase III (34).

In contrast, the crystal structures of the two product complexes (with a 3′-end AC and CC) revealed only one Mg^{2+} ion bound at the active site (Figure 3B and Supplementary Figure S3B). Comparison with the substrate complexes revealed that the major difference was that Glu73 rotated to hydrogen bond with the 3′-end C, and that the helix containing the Phe29 and Phe77 residues moved up so that these two residues could stack with the 3′-end C (differences marked by arrows in Figure 3C). The scissile phosphate also moved up accordingly, but the shifted distance of ~0.8 Å was less than that found in the RNase T-dinucleotide product complex of ~1.6 Å (see the schematic diagram for the active site conformation in Figure 3D). Due to the short distance, the Mg^{2+} ion in the complex was coordinated to fewer water molecules, slightly different from the one in the dinucleotide complex (three versus five water molecules). Because of the movement of the scissile phosphate, the second Mg^{2+} was not bound to the active site, and as a result the general base residue His181 was also shifted away from the active site. Superposition of the two AC and CC product complexes showed that the induced conformational change in the active sites was almost identical (Supplementary Figure S3B). This result thus provides a structural basis showing that a single 3′-end C in a single-stranded DNA is sufficient to induce the C effect of RNase T.

Different binding modes of RNase T in digestion of single- and double-stranded substrates

We showed here that a single C in the 3′-end of a single-stranded DNA can fully induce the C effect. However, our previous studies suggested that a 3′-end C in the 3′-overhang of a duplex substrate is not sufficient, but rather a dinucleotide CC is required, to induce the C effect (28). We wondered why single-stranded and double-stranded substrates should exhibit such a difference. Our previous results also showed that a single-stranded DNA with either a 3′-end C or a 3′-end CC bound to RNase T with similar affinity; however, a double-stranded DNA with a 3′-end CC overhang bound to RNase T with significantly reduced affinity than one with a 3′-end C. To further dissect the molecular basis for the differences in substrate binding, we compared the binding modes of RNase T in binding to a single-stranded and a duplex DNA with a 3′-overhang. Interestingly, we found that a single-stranded DNA is bound by both protomers A and B with its 3′-end inserting into the active site of protomer A. On the other hand, a duplex substrate is bound separately by protomers A and B with one protomer interacting with only one strand of the duplex (Figure 4A). The difference in the C effect for single- and double-stranded substrates likely results from the two different binding modes.

We further performed RNase T digestion assays using different duplex substrates with 3′-end AA, AC or CC. We found that indeed the duplex substrate with a 3′-end CC overhang was more resistant to digestion by RNase T than the substrate with a 3′-end AA or AC overhang (Figure 4B). This result confirmed that a dinucleotide CC in the 3′-end overhang region is required to induce the C effect when RNase T digests a duplex substrate with a short 3′-overhang. These findings also explain how RNase T can remove a single C in the overhang region near the duplex structure in 5S and 23S rRNA precursors (24,25,35).

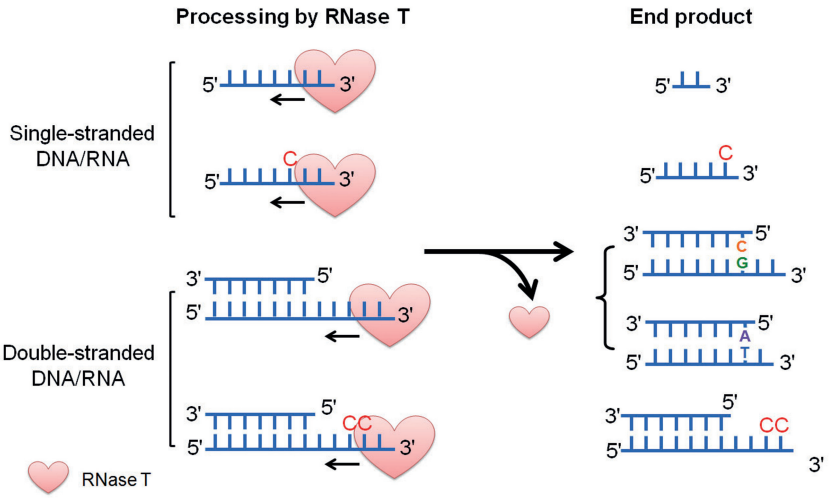


Figure 5. General rules for RNase T in trimming of nucleic acids. RNase T digests single-stranded nucleic acids to generate either a dinucleotide (without C in the middle of the chain) or an end product with a 3′-end C (with a C in the middle of the chain). Double-stranded nucleic acids with a 3′-overhang of 2–5 nt are processed by RNase T to generate an end product with a 3′-overhang of 1–4 nt, depending on the sequence of the substrate in double-stranded and overhang regions. Without a CC sequence in the 3′-overhang region, a 1-nt overhang is generated if the last base pair is AT, or a 2-nt overhang is generated if the last base pair is GC, whereas with a CC sequence in the 3′-overhang region, the trimming stops at CC.

DISCUSSION

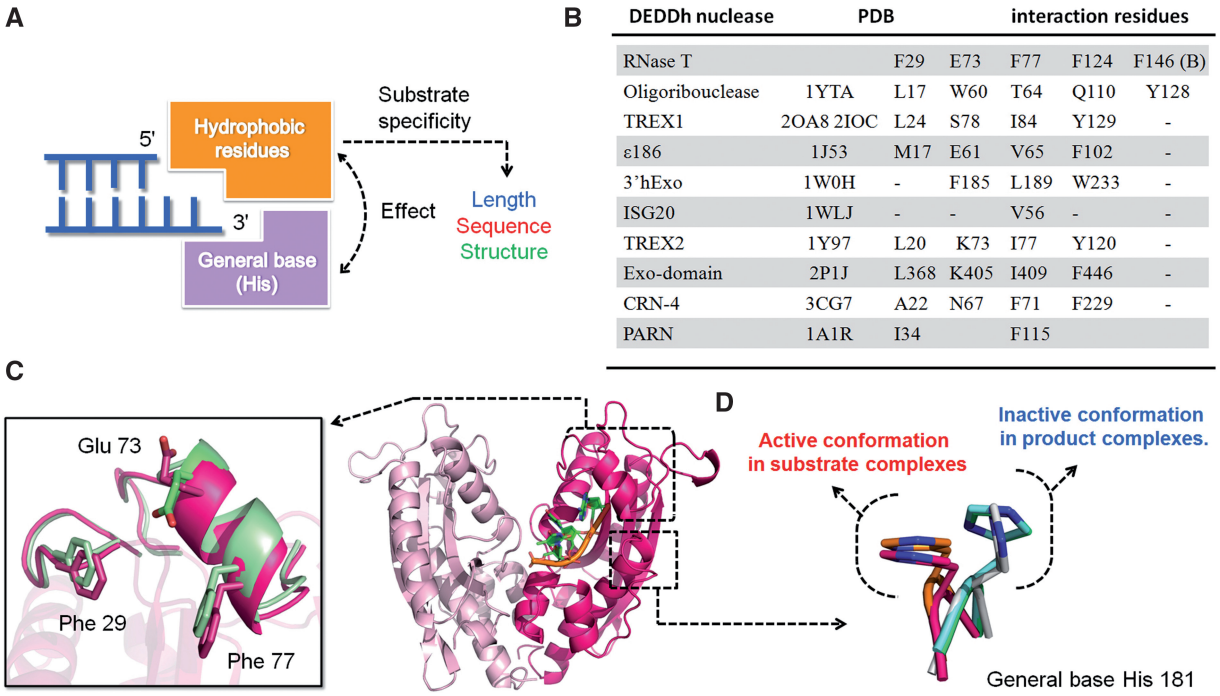
General rules of RNase T in substrate binding and cleavage

The four RNase T–product complex structures and three RNase T-substrate complex structures described here, in combination with the previously reported five complex structures, fully demonstrate the general rules for RNase T in substrate binding and cleavage. For a single-stranded substrate without a C, RNase T can remove 1 nt at a time from the 3′- to 5′-end to produce a dinucleotide end product, whereas for a single-stranded substrate with a C in the chain, RNase T is stopped by C and generates an end product with a 3′-end C (Figure 5). This specific inhibition mechanism may prevent RNase T from randomly digesting single-stranded RNA or DNA, which usually contain C in the middle of the chain. It should be noted here that RNase T digests both DNA and RNA with similar sequence and structure specificity so the rule listed in Figure 5 can be applied to both RNA and DNA substrates. This highly regulated exonuclease activity of RNase T for single-stranded nucleic acids may also be linked to possible roles in DNA repair and replication (36–38).

For a duplex substrate with a 3′-overhang of 2-5 nt without a CC sequence, RNase T can process the substrates and generate a 1-nt or 2-nt 3′-overhang, depending on the last base pair composition in the duplex region, such as 1 nt in 5S rRNA (last base pair: AT) and 2 nt in 23S rRNA (last base pair: GC) (28). However, for a duplex with a 3′-overhang containing a CC sequence, RNase T is stopped by CC and generates an end product with a longer 3′-overhang ending with CC, such as 4.5S RNA with a 4-nt 3′-overhang of ACCC (Figure 5). This unique specificity of RNase T for duplex substrates can be linked to its cellular functions. Previous studies showed that RNase T and RNase PH are both required for tRNA 3′-end maturation in *E. coli* (26,39,40). RNase T trims most of the tRNA precursors up to CC-3′, and the over-trimmed tRNAs are modified by re-addition of a 3′-terminal A by nucleotidyltransferase to generate the 3′-end CCA sequence (41,42). This re-addition step is believed to be a quality control process for tRNA maturation (43,44). However, some 3′-downstream sequences of tRNA precursors contain tandem cytosines, such as $tRNA^{Asp}$ and $tRNA^{Val}$ with sequences of CCA-CCCTA-3′ and CCA-CCGGG-3′, respectively (45). RNase T cannot trim these tRNA precursors, and they should be processed by RNase PH. Northern blot analysis of exonuclease knockout strains showed that RNase PH is indeed a more effective exonuclease than RNase T in $tRNA^{Val}$ maturation (26). A similar phenomenon was observed in the 3′-end mutation process of 23S rRNA in *E. coli* showing that RNase PH initiates the 3′-end mutation, and RNase T performs the final trimming for 23S rRNA (24,46). Why is RNase T only involved in the final trimming of 23S rRNA 3′-end maturation? This is likely because the 23S rRNA precursor containing a CC dinucleotide sequence in the 3′-overhang region is a substrate only for RNase PH, but not for RNase T. Therefore after the initial processing by RNase PH and RNase III, RNase T further removes the extra nucleotides in the 3′-overhang and generate the final mature 23S rRNA (24,46,47).

Mechanistic insights into the substrate specificity of RNase T

It is intriguing to discover how RNase T achieves its unique cleavage specificity for its diverse single- and double-stranded substrates with different sequences. Based on our structural studies of RNase T, we found that one of the most important determinants in substrate selection is unexpectedly centered on the hydrophobic residues near the active site. Previous studies showed that hydrophobic



DEDDh nuclease	PDB	interaction residues				
RNase T		F29	E73	F77	F124	F146 (B)
Oligoribouclease	1YTA	L17	W60	T64	Q110	Y128
TREX1	2OA8 2IOC	L24	S78	I84	Y129	-
ε186	1J53	M17	E61	V65	F102	-
3'hExo	1W0H	-	F185	L189	W233	-
ISG20	1WLJ	-	-	V56	-	-
TREX2	1Y97	L20	K73	I77	Y120	-
Exo-domain	2P1J	L368	K405	I409	F446	-
CRN-4	3CG7	A22	N67	F71	F229	-
PARN	1A1R	I34		F115		

Figure 6. Hydrophobic residues play key roles in regulating the substrate specificity and enzyme activity of the DEDDh family of exonucleases. (**A**) A number of hydrophobic residues of RNase T interact with nucleic acid substrates and regulate the activity of RNase T in digestion of nucleic acids of different lengths, sequences and structures. (**B**) A table lists the putative substrate-binding residues in the DEDDh exonucleases that correlate with Phe29, Glu73, Phe77, Phe124 and Phe146 in RNase T by structural alignment. (**C**) Crystal structures of RNase T–DNA complexes show that the hydrophobic residues, including Phe29 and Phe77, located in a flexible helix, shift their positions upon binding different types of substrates/products. Red and green molecules are the apo- and substrate-bound form of RNase T, respectively (ssDNA with a 3′-end C). (**D**) The general base H181 is located in a flexible loop region with two conformations: active conformation in apo-form (orange) and all the substrate complexes (red: 3′-end TA), and inactive conformation in all the product complexes (light green: 3′-end CC; light blue: dinucleotide; gray: duplex DNA with 2 nt 3′-overhang).

residues are involved in sequence-specific binding in protein–RNA and protein–DNA interactions (48,49). Our studies show that several key aromatic residues in RNase T, including Phe29 and Phe77, play important roles in differentiating a digestible or non-digestible substrate. When a preferred substrate is bound to RNase T, such as a single-stranded DNA with a 3′-end A, G or T, these two phenylalanine residues move to a position to stack with the 3′-end base and the movement is associated with an 'active' conformation at the active site where two metal ions are bound with a nearby general base His181. In contrast, when a product is bound to RNase T, including a dinucleotide, a ssDNA with a 3′-end C or a duplex DNA with a short 3′-overhang, the two phenylalanine residues shifted more to stack either with the 3′-end base or with the 5′-end base (for duplex DNA), and as a result, the movement induces an 'inactive' conformational change at the active site where only one metal ion is bound and the general base His181 is shifted away (Figure 6C and D). These hydrophobic residues are therefore tightly associated with the regulation of the exonuclease activity of RNase T and they construct a precise machine to determine which substrate should be digested based on its sequence, length and structure.

This substrate selection mechanism induced by hydrophobic residues probably is conserved in the DEDDh family of exonucleases. We superimposed the structure of RNase T with a number of DEDDh exonucleases and identified the residues corresponding to the same position as those involved in substrate interactions in RNase T, including Phe29, Glu73, Phe77, Phe124 and Phe146 (Figure 6B). We found that the residues identified are mostly hydrophobic residues, such as leucine, isoleucine, tryptophan and tyrosine, suggesting that they may also involve hydrophobic interactions with nucleic acid substrates and direct substrate specificity. Moreover, the general base histidine residue of these DEDDh exonucleases are all located in a flexible loop region with high temperature factors, which is different from other enzymes that usually have rigid active-site residues with low temperature factors (Supplementary Figure S4). This result indicates that the general base histidine residue in these DEDDh exonucleases may be shifted to an active or an inactive position according to the bound substrate in a manner similar to that found in RNase T. Thus, even though the general base histidine does not interact with the substrate directly, its conformation is indirectly regulated by substrate binding. Our profound structural study of RNase T thus provides a good testing model for understanding how the DEDDh exonucleases select and cleave their substrates in a variety of processes.

ACCESSION NUMBERS

Protein Data Bank with the PDB ID codes of 3VA3, 3VAO, 3V9Z, 3V9S, 3V9X, 3V9W and 3V9U.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–4.

FUNDING

Academia Sinica and the National Science Council, Taiwan, R.O.C. Portions of this research were carried out at the National Synchrotron Radiation Research Center [BL-13B1 and BL-13C1], a national user facility supported by the National Science Council of Taiwan, ROC, and SPring-8 (BL44XU), Japan. Funding for open access charge: Academia Sinica, Taiwan.

Conflict of interest statement. None declared.

REFERENCES

1. Shevelev,I.V. and Hubscher,U. (2002) The 3′-5′ exonucleases. *Nat. Rev. Mol. Cell. Biol.*, **3**, 364–376.
2. Zuo,Y. and Deutscher,M.P. (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.*, **29**, 1017–1026.
3. Uhrhammer,N.A., Lafarge,L., Dos Santos,L., Domaszewska,A., Lange,M., Yang,Y., Aractingi,S., Bessis,D. and Bignon,Y.J. (2006) Werner syndrome and mutations of the WRN and LMNA genes in France. *Hum. Mutat.*, **27**, 718–719.
4. Kavanagh,D., Spitzer,D., Kothari,P.H., Shaikh,A., Liszewski,M.K., Richards,A. and Atkinson,J.P. (2008) New roles for the major human 3′-5′ exonuclease TREX1 in human disease. *Cell Cycle*, **7**, 1718–1725.
5. Crow,Y.J., Hayward,B.E., Parmar,R., Robins,P., Leitch,A., Ali,M., Black,D.N., van Bokhoven,H., Brunner,H.G., Hamel,B.C. *et al.* (2006) Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat. Genet.*, **38**, 917–920.
6. Haaxma,C.A., Crow,Y.J., van Steensel,M.A., Lammens,M.M., Rice,G.I., Verbeek,M.M. and Willemsen,M.A. (2010) A de novo Asp18Asn mutation in TREX1 in a patient with Aicardi-Goutieres syndrome. *Am. J. Med. Genet. A*, **152A**, 2612–2617.
7. Yan,N., Regalado-Magdos,A.D., Stiggelbout,B., Lee-Kirsch,M.A. and Lieberman,J. (2010) The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. *Nat. Immunol.*, **11**, 1005–1013.
8. Geijtenbeek,T.B. (2010) Host DNase TREX1 hides HIV from DNA sensors. *Nat. Immunol.*, **11**, 979–980.
9. Hastie,K.M., Kimberlin,C.R., Zandonatti,M.A., MacRae,I.J. and Saphire,E.O. (2011) Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3′ to 5′ exonuclease activity essential for immune suppression. *Proc. Natl Acad. Sci. USA*, **108**, 2396–2401.
10. Hsiao,Y.Y., Nakagawa,A., Shi,Z., Mitani,S., Xue,D. and Yuan,H.S. (2009) Crystal structure of CRN-4: implications for domain function in apoptotic DNA degradation. *Mol. Cell. Biol.*, **29**, 448–457.
11. Kupsco,J.M., Wu,M.J., Marzluff,W.F., Thapar,R. and Duronio,R.J. (2006) Genetic and biochemical characterization of *Drosophila* Snipper: A promiscuous member of the metazoan 3′hExo/ERI-1 family of 3′ to 5′ exonucleases. *RNA*, **12**, 2103–2117.
12. Yang,X.C., Purdy,M., Marzluff,W.F. and Dominski,Z. (2006) Characterization of 3′hExo, a 3′ exonuclease specifically interacting with the 3′ end of histone mRNA. *J. Biol. Chem.*, **281**, 30447–30454.
13. Dominski,Z., Yang,X.C., Kaygun,H., Dadlez,M. and Marzluff,W.F. (2003) A 3′ exonuclease that specifically interacts with the 3′ end of histone mRNA. *Mol. Cell*, **12**, 295–305.
14. Kennedy,S., Wang,D. and Ruvkun,G. (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature*, **427**, 645–649.
15. Lu,D., Myers,A.R., George,N.P. and Keck,J.L. (2011) Mechanism of Exonuclease I stimulation by the single-stranded DNA-binding protein. *Nucleic Acids Res.*, **39**, 6536–6545.
16. Breyer,W.A. and Matthews,B.W. (2000) Structure of *Escherichia coli* exonuclease I suggests how processivity is achieved. *Nat. Struct. Biol.*, **7**, 1125–1128.
17. Nguyen,L.H., Espert,L., Mechti,N. and Wilson,D.M. 3rd (2001) The human interferon- and estrogen-regulated ISG20/HEM45 gene product degrades single-stranded RNA and DNA in vitro. *Biochemistry*, **40**, 7174–7179.
18. Horio,T., Murai,M., Inoue,T., Hamasaki,T., Tanaka,T. and Ohgi,T. (2004) Crystal structure of human ISG20, an interferon-induced antiviral ribonuclease. *FEBS Lett.*, **577**, 111–116.
19. Ghosh,S. and Deutscher,M.P. (1999) Oligoribonuclease is an essential component of the mRNA decay pathway. *Proc. Natl Acad. Sci. USA*, **96**, 4372–4377.
20. Wu,M., Reuter,M., Lilie,H., Liu,Y., Wahle,E. and Song,H. (2005) Structural insight into poly(A) binding and catalytic mechanism of human PARN. *EMBO J.*, **24**, 4082–4093.
21. Thore,S., Mauxion,F., Seraphin,B. and Suck,D. (2003) X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO Rep.*, **4**, 1150–1155.
22. Zuo,Y. and Deutscher,M.P. (2002) The physiological role of RNase T can be explained by its unusual substrate specificity. *J. Biol. Chem.*, **277**, 29654–29661.
23. Zuo,Y. and Deutscher,M.P. (1999) The DNase activity of RNase T and its application to DNA cloning. *Nucleic Acids Res.*, **27**, 4077–4082.
24. Li,Z., Pandit,S. and Deutscher,M.P. (1999) Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T. *RNA*, **5**, 139–146.
25. Li,Z. and Deutscher,M.P. (1995) The tRNA processing enzyme RNase T is essential for maturation of 5S RNA. *Proc. Natl Acad. Sci. USA*, **92**, 6883–6886.
26. Li,Z. and Deutscher,M.P. (1996) Maturation pathways for *E. coli* tRNA precursors: a random multienzyme process in vivo. *Cell*, **86**, 503–512.
27. Zuo,Y. and Deutscher,M.P. (2002) Mechanism of action of RNase T. II. A structural and functional model of the enzyme. *J. Biol. Chem.*, **277**, 50160–50164.
28. Hsiao,Y.Y., Yang,C.C., Lin,C.L., Lin,J.L., Duh,Y. and Yuan,H.S. (2011) Structural basis for RNA trimming by RNase T in stable RNA 3′-end maturation. *Nat. Chem. Biol.*, **7**, 236–243.
29. Otwinowski,Z. and Minor,W. (1997) ''Processing of X-ray diffraction data collected in oscillation mode'', methods in enzymology. *Macromol. Crystallogr. Part A*, **276**, 307–326.
30. Potterton,E., Briggs,P., Turkenburg,M. and Dodson,E. (2003) A graphical user interface to the CCP4 program suite. *Acta. Cryst. D*, **59**, 1131–1137.
31. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta. Cryst. D*, **60**, 2126–2132.
32. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta. Cryst. D*, **58**, 1948–1954.
33. Brucet,M., Querol-Audi,J., Bertlik,K., Lloberas,J., Fita,I. and Celada,A. (2008) Structural and biochemical studies of TREX1 inhibition by metals. Identification of a new active histidine conserved in DEDDh exonucleases. *Protein Sci.*, **17**, 2059–2069.
34. Hamdan,S., Carr,P.D., Brown,S.E., Ollis,D.L. and Dixon,N.E. (2002) Structural basis for proofreading during replication of the Escherichia coli chromosome. *Structure*, **10**, 535–546.
35. Ibrahim,H., Wilusz,J. and Wilusz,C.J. (2008) RNA recognition by 3′-to-5′ exonucleases: the substrate perspective. *Biochimica et Biophysica Acta.*, **1779**, 256–265.
36. Padmanabha,K.P. and Deutscher,M.P. (1991) RNase T affects *Escherichia coli* growth and recovery from metabolic stress. *J. Bacteriol.*, **173**, 1376–1381.
37. Kelly,K.O. and Deutscher,M.P. (1992) The presence of only one of five exoribonucleases is sufficient to support the growth of *Escherichia coli*. *J. Bacteriol.*, **174**, 6682–6684.

38. Viswanathan,M., Lanjuin,A. and Lovett,S.T. (1999) Identification of RNase T as a high-copy suppressor of the UV sensitivity associated with single-strand DNA exonuclease deficiency in *Escherichia coli*. *Genetics*, **151**, 929–934.
39. Li,Z. and Deutscher,M.P. (1994) The role of individual exoribonucleases in processing at the 3′ end of *Escherichia coli* tRNA precursors. *J. Biol. Chem.*, **269**, 6064–6071.
40. Schurer,H., Schiffer,S., Marchfelder,A. and Morl,M. (2001) This is the end: processing, editing and repair at the tRNA 3′-terminus. *Biol. Chem.*, **382**, 1147–1156.
41. Weiner,A.M. (2004) tRNA maturation: RNA polymerization without a nucleic acid template. *Curr. Biol.*, **14**, R883–R885.
42. Deutscher,M.P. and Ghosh,R.K. (1978) Preparation of synthetic tRNA precursors with tRNA nucleotidyltransferase. *Nucleic Acids Res.*, **5**, 3821–3829.
43. Dupasquier,M., Kim,S., Halkidis,K., Gamper,H. and Hou,Y.M. (2008) tRNA integrity is a prerequisite for rapid CCA addition: implication for quality control. *J. Mol. Biol.*, **379**, 579–588.
44. Hou,Y.M. (2010) CCA addition to tRNA: implications for tRNA quality control. *IUBMB life*, **62**, 251–260.
45. Lowe,T.M. and Eddy,S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
46. Gutgsell,N.S. and Jain,C. (2012) Role of precursor sequences in the ordered maturation of *E. coli* 23S ribosomal RNA. *RNA*, **18**, 345–353.
47. Allas,U., Liiv,A. and Remme,J. (2003) Functional interaction between RNase III and the *Escherichia coli* ribosome. *BMC Mol. Biol.*, **4**, 8.
48. Ellis,J.J., Broom,M. and Jones,S. (2007) Protein-RNA interactions: structural analysis and functional classes. *Proteins*, **66**, 903–911.
49. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.