

RESEARCH PAPER

 OPEN ACCESS 

Structural basis for Dicer-like function of an engineered RNase III variant and insights into the reaction trajectory of two-Mg²⁺-ion catalysis

Sudhaker Dharavath , Gary X. Shaw , and Xinhua Ji 

Center for Structural Biology, National Cancer Institute, Frederick, MD, USA

ABSTRACT

The RNase III family of dsRNA-specific endonucleases is exemplified by prokaryotic RNase III and eukaryotic Rnt1p, Drosha, and Dicer. Structures of *Aquifex aeolicus* RNase III (AaRNase III) and *Saccharomyces cerevisiae* Rnt1p (ScRnt1p) show that both enzymes recognize substrates in a sequence-specific manner and propel RNA hydrolysis by two-Mg²⁺-ion catalysis. Previously, we created an *Escherichia coli* RNase III variant (EcEEQ) by eliminating the sequence specificity via protein engineering and called it bacterial Dicer for the fact that it produces heterogeneous small interfering RNA cocktails. Here, we present a 1.8-Å crystal structure of a postcleavage complex of EcEEQ, representing a reaction state immediately after the cleavage of scissile bond. The structure not only establishes the structure-and-function relationship of EcEEQ, but also reveals the functional role of a third Mg²⁺ ion that is involved in RNA hydrolysis by bacterial RNase III. In contrast, the cleavage site assembly of ScRnt1p does not contain a third Mg²⁺ ion. Instead, it involves two more amino acid side chains conserved among eukaryotic RNase IIIs. We conclude that the EcEEQ structure (this work) represents the cleavage assembly of prokaryotic RNase IIIs and the ScRnt1p structure (PDB: 4OOG), also determined at the postcleavage state, represents the cleavage assembly of eukaryotic RNase IIIs. Together, these two structures provide insights into the reaction trajectory of two-Mg²⁺-ion catalysis by prokaryotic and eukaryotic RNase III enzymes.

ARTICLE HISTORY

Received 9 May 2022

Revised 1 July 2022

Accepted 5 July 2022

KEYWORDS

RNase III; Rnt1p; Drosha; Dicer; dsRNA; RNA hydrolysis; two-Mg²⁺-ion catalysis; postcleavage complex

Introduction

Discovered in 1968, *Escherichia coli* RNase III (EcRNase III) is the founding member of the RNase III family of dsRNA-specific endonucleases found in all kingdoms of life [1,2]. Representative members of the family include bacterial RNase III, yeast Rnt1p, human Drosha, and human Dicer. A common type of substrate for all RNase III enzymes is stem-loop RNA. Also known as hairpin RNA, stem-loop RNA is an essential secondary structure of primary importance [3]. A less common type of substrate for some RNase IIIs is long duplex RNA. Among the four representative family members, Rnt1p and Drosha process stem-loop RNAs only, whereas RNase III and Dicer also process long dsRNAs. In a successive manner, the processing starts from one end of a long dsRNA and produces small duplex RNAs [4–6]. Dicer typically measures 22 nucleotides for cleavage, producing small interfering RNAs (siRNAs) (Figure 1(a)). In contrast, RNase III typically measures 11 nucleotides for cleavage, resulting in small duplex RNAs of half the length of siRNAs (Figure 1(b)) [7]. In addition to this end-in manner of processing, long dsRNA can also be processed by RNase III in an inside-out scheme, where multiple RNase III molecules bind consecutively to a long dsRNA and cleave the substrate simultaneously [8]. Since the distance between consecutive active centres of adjacent RNase III molecules is 22 nucleotides, the inside-out processing produces siRNA-like small duplex RNAs, indistinguishable from the siRNAs produced

by Dicer (Figure 1(b)). This mechanism has been observed for EcRNase III under special conditions [9].

Since the discovery of EcRNase III, tremendous progress has been made in studies of bacterial enzymes [10,11]. Genetic and functional studies have been performed mainly with EcRNase III, whereas structural and mechanistic studies have been carried out mainly with *Aquifex aeolicus* RNase III (AaRNase III). Crystal structures of AaRNase III have been determined at several distinct catalytic states, providing insights into the mechanisms of substrate recognition, scissile bond selection, two-Mg²⁺-ion catalysis, phosphoryl transfer, and product release [6,11–13]. In addition, a hypothesis that the E38A mutant of EcRNase III (EcE38A) could promote the inside-out processing of long dsRNA led to the discovery of a reagent for the preparation of siRNA cocktails to be used in gene silencing studies (Figure 1(c)) [14]. It offers an economic way for siRNA cocktail preparation since making EcE38A costs much less than the production of human Dicer. Residue E38 is conserved among bacterial RNase IIIs and is located on the RNA-contacting surface distant from the cleavage site. Bacterial RNase III functions as a dimer, and thus removing two negatively charged side chains between the protein and RNA promotes consecutive binding and cleavage of long dsRNA by multiple RNase III dimers (Figure 1(c)).

By next generation sequencing, we previously demonstrated that both EcRNase III and AaRNase III cleave long dsRNAs at preferred sites [7]. The AaRNase III structures

CONTACT Xinhua Ji  jix@mail.nih.gov  Center for Structural Biology, National Cancer Institute, 1050 Boyles Street, Frederick, MD 21702, USA

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15476286.2022.2099650>.

This work was authored as part of the Contributor's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

This is an Open Access article that has been identified as being free of known restrictions under copyright law, including all related and neighbouring rights (<https://creativecommons.org/publicdomain/mark/1.0/>). You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

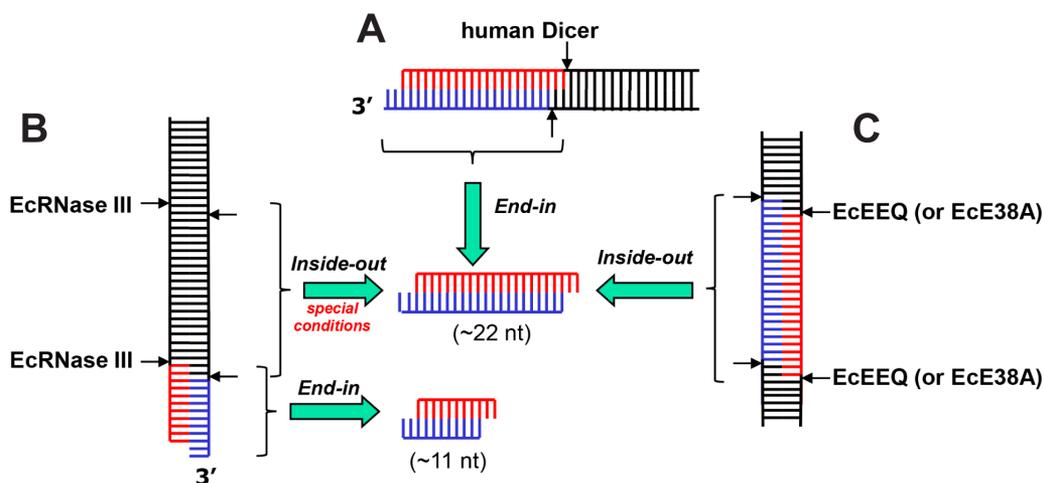


Figure 1. Mechanisms of long dsRNA processing by RNase III enzymes. (a) Human Dicer recognizes the dsRNA termini with a 2-nucleotide 3' overhang, cleaves both strands, and produces a duplex RNA of 22 nucleotides in each strand (The end-in mechanism). (b) An EcrNase III dimer recognizes the dsRNA termini, especially those featuring a 2-nucleotide 3' overhang, cleaves both strands, and produces a short duplex RNA of 11 nucleotides in each strand (The end-in mechanism). Under special conditions, however, two EcrNase III dimers bind to and cleave dsRNA in a cooperative manner, which produces a duplex RNA of 22 nucleotides in each strand. (The inside-out mechanism). (c) Two EcEEQ (the E38A/E65A/Q165A triple mutant of EcrNase III) or EcE38A (the EcE38A single mutant of EcrNase III) dimers bind to and cleave dsRNA in a cooperative manner, which produces a duplex RNA of 22 nucleotides in each strand (The inside-out mechanism). A similar version of this figure was previously published in [7].

suggest that the specificity for a guanine nucleotide at the +3 position near the scissile bond could be eliminated by an alanine substitution at Q165 of EcrNase III. Like E38, residue E65 of EcrNase III is also located on the RNA-contacting surface. Therefore, the E65A mutation in EcrNase III could also assist in promoting the inside-out mechanism. These two predictions prompted us to create a triple mutant (E38A/E65A/Q165A) of EcrNase III (EcEEQ, Figure 1(c)), which indeed produces heterogeneous siRNA cocktails [7]. Therefore, we call this EcrNase III variant bacterial Dicer. It is remarkable that following distinct mechanisms, human Dicer (end-in) and bacterial Dicer (inside-out) produce indistinguishable products (Figure 1). Here, we report the crystal structure of EcEEQ in complex with a stem-loop RNA (RNA6) at 1.80-Å resolution (EcEEQ:RNA6), showing how the E38A and E65A mutations stabilize the protein:RNA complex and how the Q165A mutation eliminates the sequence specificity of EcrNase III.

To enhance substrate specificity and catalytic efficiency, two-Mg²⁺-ion catalysis is used by all RNA and DNA polymerases, and most nucleases and recombinases [15]. The two-Mg²⁺-ion catalysis by RNase III was established by structures of AaRNase III and ScRnt1p, in the form of a postcleavage complex determined at a stage immediately after the cleavage of the scissile bond [12,16]. Intriguingly, the EcEEQ:RNA6 structure reveals the functional role of a third Mg²⁺ ion that is involved in the mechanism of two-Mg²⁺-ion catalysis by bacterial RNase III.

Results and discussion

The EcEEQ:RNA6 is highly homologous to the postcleavage complex of AaRNase III

We started working on EcrNase III in 1996, but we were not able to crystallize the protein, either full-length or truncated.

In parallel with EcEEQ, we performed crystallization trials again with wild-type EcrNase III in this study, but only the EcEEQ:RNA6 complex crystallized, underscoring the significant impact of removing four negatively charged amino acid side chains from RNA-contacting surface of the protein on the stability of the protein:RNA complex. Solving the phase problem by molecular replacement (MR), we first utilized the AaRNase III:RNA6 structure, a protein-product complex containing one Mg²⁺ ion in each cleavage site [6], as the search model, which resulted in multiple solutions of low log-likelihood gain (LLG) and final translation function Z (TFZ) scores (top solution: LLG = 71; TFZ = 6.0). Then, we tested the postcleavage complex of AaRNase III (AaRNase III:RNA9), containing two Mg²⁺ ions in each cleavage site [12], which resulted in a unique solution of much higher LLG and TFZ scores (LLG = 445; TFZ = 15.9). Hence, the EcEEQ:RNA6 structure resembles the postcleavage complex of AaRNase III, indicating that the EcEEQ:RNA6 complex is also a postcleavage complex. The structure is indeed a postcleavage complex with an additional feature. It reveals that a third Mg²⁺ ion is integrated in the cleavage site assembly of two-Mg²⁺-ion catalysis by bacterial RNase IIIs.

The EcEEQ:RNA6 structure is schematically illustrated in Figure 2(a). The complex is composed of two EcEEQ subunits (each containing 226 amino acid residues), two RNA6 molecules (each containing 28 nucleotide residues in the form of a hairpin with a 4-nucleotide capping loop and a 2-nucleotide 3' overhang), six Mg²⁺ ions, 418 water oxygen atoms, and several other ions and molecules from the solvent and cryoprotectant. All three Mg²⁺ ions in the cleavage site of each subunit are well defined with full occupancy (Figure 2(b)). Each EcEEQ molecule contains a specialized endonuclease domain (RIIID, residues 1–147) and a dsRNA-binding domain (dsRBD, residues 156–226, Figure 2(c)). The root-mean-square deviation (RMSD) between the two EcEEQ subunits is 0.17 Å for 196 out of 226 pairs of Ca atoms, and the

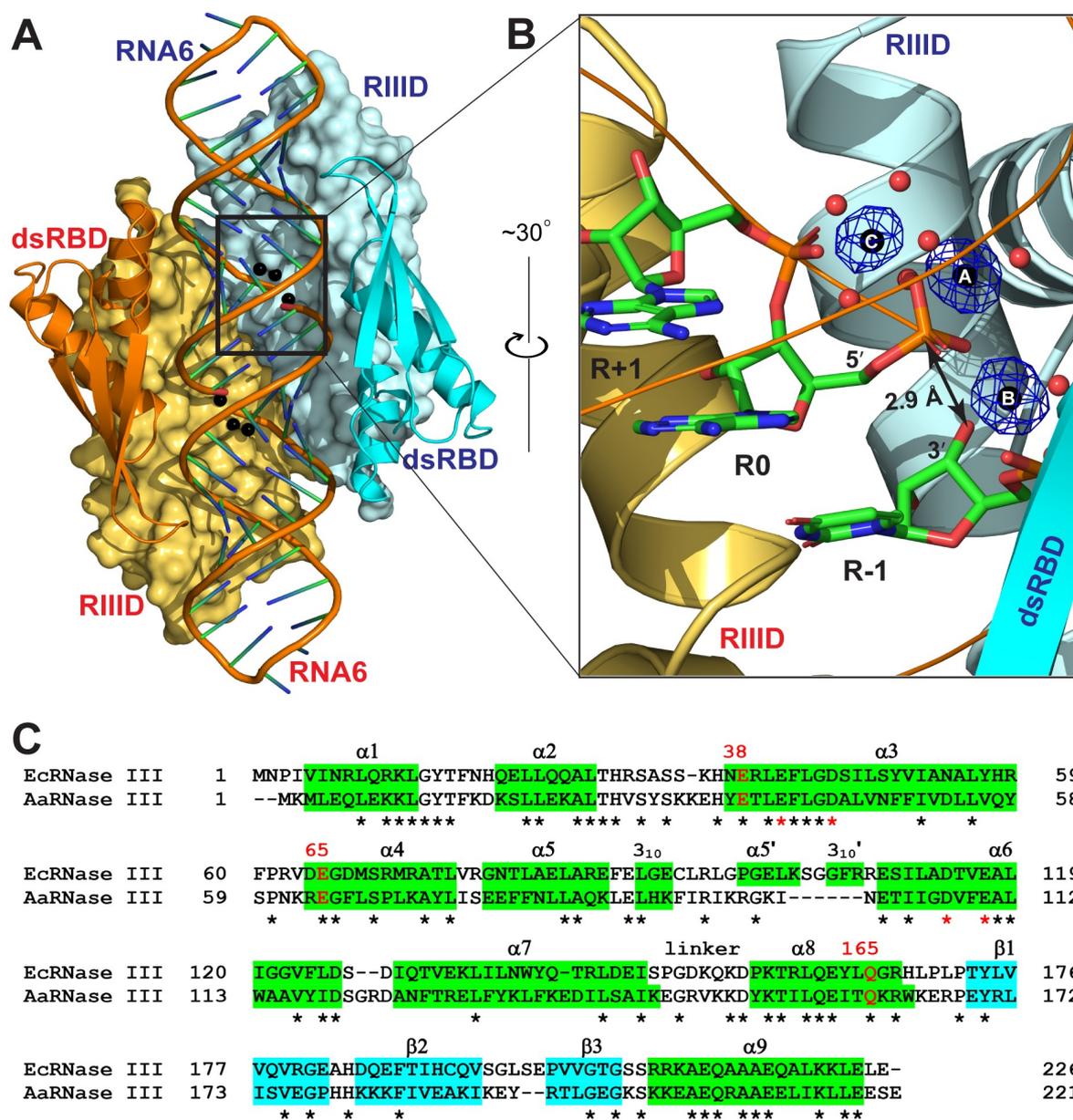


Figure 2. Crystal Structure of the EcEEQ:RNA6 Complex. (a) The two subunits of dimeric EcEEQ are shown in cyan and Orange, respectively. RIIIDs are illustrated as molecular surfaces and dsRBDs as ribbon diagrams (helices as spirals, strands as arrows, and loops as tubes). RNA6 molecules are shown as cartoon models and Mg^{2+} ions as spheres. (b) The architecture of cleavage site is illustrated with nucleotide residues R-1, R0 and R+1 (stick models, N in blue, C in green, O in red, and P in Orange), 3 Mg^{2+} ions (spheres in black), and Mg^{2+} -coordinating water molecules (spheres in red) in one cleavage site assembly. The $F_o - F_c$ omit map (contoured at 6.0 σ , in blue) is shown for the three Mg^{2+} ions. (c) Sequence alignment of EcRNase III (UniProtKB: P0A7Y0) and AaRNase III (UniProtKB: O67082) based on the crystal structures of EcEEQ:RNA6 (this work) and AaRNase III:RNA9 (PDB: 2NUG). Identical residues are indicated with stars under the sequences. Helices are highlighted in green and strands in cyan. The three mutation sites are highlighted in red with residue numbers of EcRNase III shown above the sequences. Conserved residues are indicated with stars under the sequences and the four catalytic residues are highlighted in red.

RMSD between the two RNA6 molecules is 0.14 Å for 587 out of 598 pairs of atoms, underscoring the highly symmetric nature of the EcEEQ:RNA6 complex (Figure 2(a)).

The EcEEQ:RNA6 and AaRNase III:RNA9 structures are highly homologous. The EcEEQ structure exhibits 15 secondary structure elements, of which 13 are shared with AaRNase III (Figure 2(c)). Eleven out of the 13 shared secondary structure elements are of the same length in the two structures, whereas two ($\alpha 7$ and $\alpha 8$) are one residue longer in AaRNase III. Superposition of the EcEEQ:RNA6 and AaRNase III:RNA9 complexes shows that the two structures align very well (Figure 3(a)). The RMSD for 404 out of 417

pairs of Ca atoms between the two RNase III dimers is 1.7 Å. Including RNA, the RMSD for 382 out of 458 pairs of Ca and P atoms is 1.6 Å. The high homology of the two structures offers the convenience of using just the numbering scheme of the focus, i.e. the residue numbers in EcRNase III.

Structural basis for Dicer-like function of EcEEQ

The first glimpse at the RNase III active site was provided by our structure of *A. aeolicus* RIIID dimer 33 years after the discovery of the enzyme [17,18]. The structure showed that the dimerization of RIIID creates a catalytic valley about 20 Å

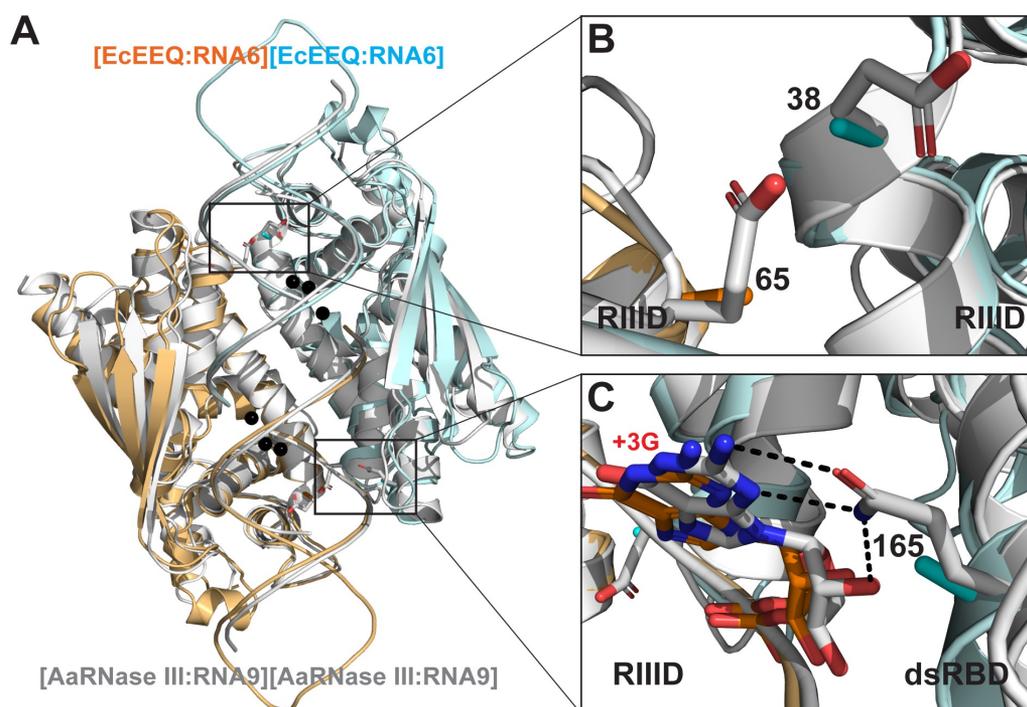


Figure 3. Structural Basis for the Function of EcEEQ, the Bacterial Dicer. (a) Superposition of the EcEEQ:RNA6 (ribbon diagram in cyan and Orange, this work) and AaRNase III:RNA9 (in grey, PDB: 2NUG) structures. Mg²⁺ ions in the EcEEQ:RNA6 structure are shown as black spheres to highlight the two RNA cleavage sites. Selected side chains are shown as stick models in atomic colour scheme (N in blue, O in red, and C in cyan, Orange, or grey). (b) Zoom-in view shows that the E38A and E65A mutations remove negative charges from the catalytic valley of EcRNase III, promoting the inside-out processing of long dsRNA. (c) Zoom-in view shows that the Q165 side chain forms three hydrogen bonds with the +3 G nucleotide, two with the base and one with the 2'-OH group, and that the Q165A mutation abolishes the recognition of the +3 G nucleotide by the enzyme. The length of the three hydrogen bonds ranges between 2.9 and 3.1 Å.

wide and 50 Å long, which can accommodate a dsRNA substrate. Five years later, our AaRNase III:RNA6 structure showed that the proposed catalytic valley is indeed the binding site of a dsRNA substrate [6]. In the middle of the valley are located two cleavage sites, each featuring a cluster of four negatively charged side chains (E41, D45, D114, and E117) [12]. At each end of the valley, two more negatively charged side chains (E38 and E65) are located. These negatively charged side chains are conserved among bacterial RNase III enzymes.

As depicted in Figure 3, panels A and B, residues E38 and E65 are located on the RNA-contacting surface. Therefore, the alanine mutation of both E38 and E65 stabilizes the resulting protein:RNA complex and thereby promotes the inside-out processing of long dsRNAs. This structural implication was validated by *in vitro* cleavage products analysis [7]. The third mutation in EcEEQ is Q165A that is critical for the production of heterogeneous siRNA cocktails. As shown in Figure 3 (c), the Q165 side chain forms three hydrogen bonds with the +3 G nucleotide, two with the base and one with the O2'-hydroxyl group [12]. The length of these hydrogen bonds ranges between 2.9 and 3.1 Å. Hence, the Q165A mutation eliminates these strong hydrogen bonds and thereby abolishes its specificity for the +3 G nt. This structural implication was also validated by *in vitro* cleavage products analysis [7]. The loss of three hydrogen bonds between the protein and RNA results in a less stable protein:RNA complex. This negative impact on the stability is offset by the overwhelming stabilization effect of the E38A and E65A mutations as demonstrated previously by product analysis of *in vitro* cleavage [7].

As mentioned above, the RNA-contacting surface of the catalytic valley is coated by a total of 12 negatively charged side chains. The E38A and E65A mutations neutralize four of them. The remaining eight are catalytic side chains, forming two cleavage sites in the catalytic valley. Interactions between these catalytic side chains and dsRNA are bridged by catalytic Mg²⁺ ions that mitigate electrostatic repulsion between the enzyme and the RNA. No other negatively charged side chains exist between the catalytic valley and the bond dsRNA. Therefore, the impact of E38A and E65A mutations on the stability of the protein:RNA complex is profound. Taken together, these features revealed by the EcEEQ:RNA6 and AaRNase III:RNA9 complexes are structural basis for the function of EcEEQ, the bacterial Dicer that is most suitable for producing heterogeneous siRNA cocktails.

The EcEEQ:RNA6 represents the postcleavage state of RNA hydrolysis by bacterial RNase III

The RIIID dimer hydrolyzes both strands of dsRNA simultaneously with two identical cleavage sites. In each cleavage site of the EcEEQ:RNA6 structure, three Mg²⁺ ions are well defined and fully occupied. As depicted in Figure 4(a), these Mg²⁺ ions (MgA, MgB and MgC) are 'organizers' of the cleavage site assembly by coordinating with four catalytic side chains (E41, D45, D114, and E117), three nucleotide residues (R-1, R 0, and R+1), and eight water molecules around the scissile bond. MgA coordinates with two oxygen atoms of the scissile phosphate group (R 0), two oxygen atoms

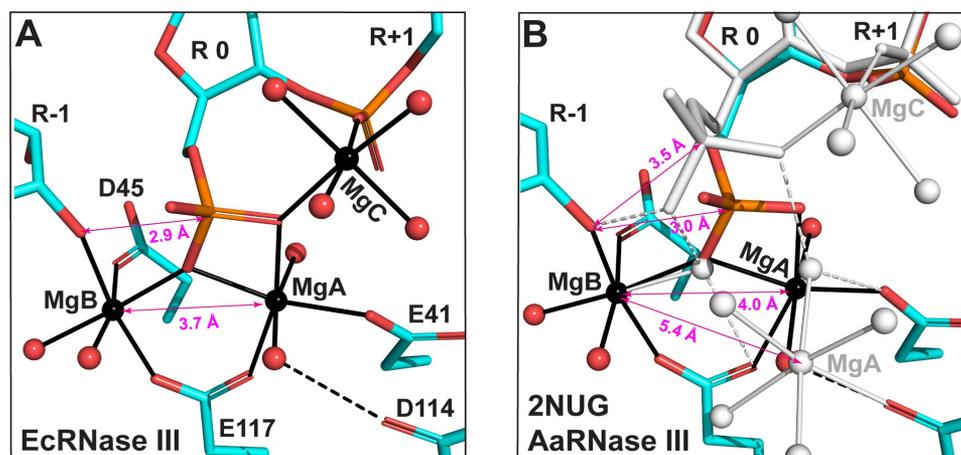


Figure 4. Three Mg^{2+} Ions in the Postcleavage Complex of Bacterial RNase III. (a) The catalytic site assembly of EcRNase III as observed in the EcEEQ:RNA6 structure (this work). Amino acid and nucleotide residues are illustrated as stick models, and Mg^{2+} ions and water oxygen atoms as spheres in atomic colour scheme (C in cyan, N in blue, O in red, P in Orange, and Mg in black). Solid lines indicate coordinate bonds. Dashed lines indicate hydrogen bonds. Double-headed lines indicate contact distances (CDs) between the 3'-oxygen of nucleotide R-1 and the phosphorus of nucleotide R 0 ($CD^{O^{3'},P}$) and between MgA and MgB ($CD^{MgA,MgB}$). (b) The catalytic site assembly of AaRNase III as observed in the AaRNase III:RNA9 structure (PDB: 2NUG), exhibiting a minor conformation (25%, in atomic colour scheme) and a major conformation (75%, in white). The MgC was not observed in the minor conformation most likely due to its low occupancy. For clarity, the four catalytic residues are not labelled in panel b.

of the E41 and E117 side chains, and two oxygen atoms of water molecules. MgB coordinates with one oxygen atom of the scissile phosphate group (R 0), the oxygen atom of 3'-OH group (R-1), two oxygen atoms of the D45 and E117 side chains, and two oxygens of water molecules. MgC coordinates with one oxygen of the scissile phosphate group (R 0), one phosphate oxygen of the R+1 nucleotide, and four oxygen atoms of water molecules. Among the four catalytic side chains, only D114 does not coordinate with Mg^{2+} directly. Bridging negatively charged components, the three Mg^{2+} ions make the cleavage site assembly as compact as needed for catalysis.

The compactness of the assembly can be measured with two contact distances (CDs), one between MgA and MgB ($CD^{MgA,MgB}$) and the other between 3'-oxygen of nucleotide R-1 and phosphorus of nucleotide R 0 ($CD^{O^{3'},P}$). The phosphorus of nucleotide R 0 represents the scissile phosphate group after RNA hydrolysis. Before the reaction, it is covalently bonded to the 3'-oxygen of nucleotide R-1. In the EcEEQ:RNA6 structure, the $CD^{O^{3'},P}$ measures 2.9 Å (Figure 4(a)), which is smaller than the sum of van der Waals (vdW) radii of O (1.52 Å) and P (1.80 Å) [19]. As shown in Figure 4(b), the cleavage site assembly in the AaRNase III:RNA9 structure exhibits two distinct conformations, a conformation of 25% occupancy (minor-AaRNase III:RNA9) and a conformation of 75% occupancy (major-AaRNase III:RNA9) [12]. The $CD^{O^{3'},P}$ in minor-AaRNase III:RNA9 measures 3.0 Å, indicating that 25% of the population represents the postcleavage state, whereas it measures 3.5 Å in major-AaRNase III:RNA9, indicating that 75% of the population represents a state of product release. Whereas two Mg^{2+} ions were observed in minor-AaRNase III:RNA9, three Mg^{2+} ions were observed in major-AaRNase III:RNA9. Neither MgC nor its four coordination water molecules were observed in minor-AaRNase III:RNA9 (Figure 4(b)). It has been previously shown that $CD^{MgA,MgB}$ is about 3.5 Å at the

intermediate state for the two- Mg^{2+} -ion catalysis by either RNase H1 or DNAP η [20,21]. The $CD^{MgA,MgB}$ in the EcEEQ:RNA6 structure is 3.7 Å (Figure 4(a)), suggesting that the EcEEQ:RNA6 structure represents the postcleavage state immediately after the intermediate state of RNA hydrolysis.

Structural insights into two- Mg^{2+} -ion catalysis by bacterial and yeast RNase III enzymes

Nucleophilic attack on phosphorus could produce a relatively long-lived pentacovalent intermediate. And all phosphoryl transfer reactions in DNA and RNA involve such an intermediate and inversion of the stereo configuration at the phosphorus [15]. Based on the EcEEQ:RNA6 structure, models of reaction intermediate and precleavage complex could be readily derived by adjusting the torsion angles along the C4'-C5'-O5'-P-O1P chain and breaking or making the P-O bonds. At the postcleavage state, a scissile-phosphate oxygen is coordinated with both MgA and MgC (Figure 5(a)). This special oxygen atom is in fact the nucleophilic water oxygen at the intermediate (Figure 5(b)) and precleavage states (Figure 5(c)). It appears that MgC plays three important roles in catalysis. First, MgC teams with MgA and MgB to optimize the cleavage site geometry for the formation of the pentacovalent intermediate (Figure 5(b)). Second, MgC synergizes with MgA to activate the nucleophilic water molecule (Figure 5(c)). Third, MgC facilitates the nucleophilic attack and subsequent electron transfer (Figure 5(b)). Coordinated with the scissile phosphate, MgC has been observed not only at the postcleavage state (EcEEQ:RNA6, Figure 4(a)), but also at two distinct snapshots during product release, one after the scissile phosphate moves away from the cleavage site (major-AaRNase III:RNA9, Figure 4(b)) and the other after the scissile hydroxyl also moves away from the cleavage site [12]. We predict that all prokaryotic RNase IIIs employ the third Mg^{2+} ion to assist two- Mg^{2+} -ion catalysis of RNA hydrolysis.

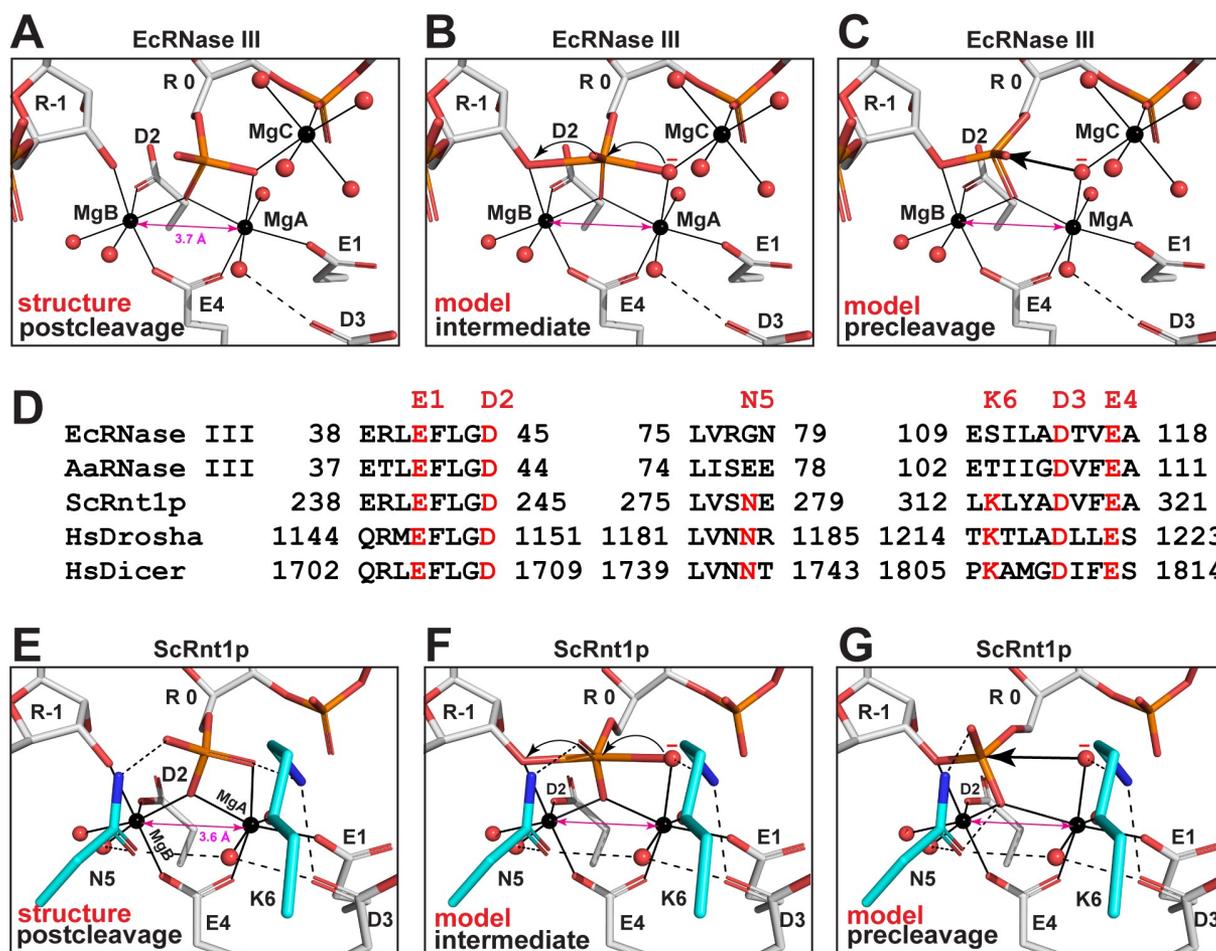


Figure 5. Stepwise Model for the Reaction Trajectory of Two-Mg²⁺-ion Catalysis by RNase III: Distinct Features between Prokaryotic and Eukaryotic Enzymes. (a-c) Cleavage site architecture of bacterial RNase III at the postcleavage (structure), intermediate (model), and precleavage (model) states are represented by the EcEEQ:RNA6 structure (this work) and models derived from the structure. The amino acid side chains and nucleotide residues are shown as stick models and Mg²⁺ ions and water oxygens as spheres (C in grey, N in blue, O in red, P in Orange, and Mg in black). Metal co-ordination bonds are illustrated as solid lines and hydrogen bonds as dashed lines. The nucleotide residue in the middle is numbered 'R 0' and the rest are numbered according to the polarity of the RNA strand. (d) Structure-based sequence alignment of EcRNase III (this work), AaRNase III, (PDB: 2NUG), ScRnt1p (PDB: 4OOG), HsDrosha (PDB: 5B16), and HsDicer (PDB: 5ZAL). Conserved amino acid residues are highlighted in red. Residues E1, D2, D3, and E4 are conserved in all RNase IIIs, whereas N5 and K6 are unique for eukaryotic RNase IIIs. (e-g) Cleavage site architecture of yeast Rnt1p at the postcleavage (structure), intermediate (model), and precleavage (model) states are represented by the structure of ScRnt1p postcleavage complex (PDB: 4OOG) and models derived from the structure, illustrated in the same manner as panels a-c except that the carbon atoms in residues N5 and K6 are highlighted in cyan.

Unlike bacterial RNase IIIs that use four catalytic side chains (E1, D2, D3, and E4) in each event of two-Mg²⁺-ion catalysis, yeast Rnt1p uses two more (N5 and K6) that are conserved among eukaryotic RNase IIIs (Figure 5(d)). As revealed by the Rnt1p postcleavage structure, the side chains of E1, D2, D3, and E4 in Rnt1p assume identical positioning as in bacterial RNase IIIs, whereas those of N5 and K6 are unique for eukaryotes [16]. The side chain of N5 interacts with two water molecules and one oxygen of the 5' phosphate while the ε-amino group of K6 side chain interacts with one oxygen of the 5' phosphate and another from the carboxylic group of D3 (Figure 5(e)). As shown, in the presence of N5 and K6, the CD^{MgA,MgB} is 3.6 Å at the postcleavage state of Rnt1p. Like the MgC in the two-Mg²⁺-ion catalysis of prokaryotic RNase IIIs, the N5 and K6 side chains also play three important roles. First, the K6 side chain recognizes the scissile phosphate group and teams with MgA and MgB to optimize the cleavage site geometry for the formation of the pentacovalent intermediate (Figure 5(f)). Second, K6 side chain synergizes with MgA to activate the nucleophilic

water molecule (Figure 5(g)). Third, K6 and N5 side chains facilitate the nucleophilic attack and subsequent electron transfer (Figure 5(f)). Since K6 is not conserved in one RIIID among eukaryotic RNase III enzymes [8], we predict that all eukaryotic RNase IIIs employ the N5 and most also employ K6 side chains to assist two-Mg²⁺-ion catalysis of RNA hydrolysis. In the presence of K6, the positively charged ε-amino group prevents a positively charged metal ion from binding at the MgC site. In the absence of K6, however, whether a third metal ion would bind at the MgC site remains to be elucidated.

Materials and methods

Protein expression and purification

The expression vector of EcEEQ was constructed and His₆-MBP tagged EcEEQ (His₆-MBP-EcEEQ) was overproduced in *E. coli* BL21(DE3) Codon Plus-RIL cells (ThermoFisher Scientific, Waltham, MA) as described [7] with limited

modifications. Briefly, the cells were cultivated in Luria-Bertani (LB) broth containing 100 $\mu\text{g ml}^{-1}$ ampicillin and 30 $\mu\text{g ml}^{-1}$ chloramphenicol at 37°C, induced by the addition of isopropyl β -d-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and shaken for 4–6 hr at 37°C. The cells were harvested by centrifugation at 4000 g for 10 min at 4°C and lysed in 30 mM Tris (pH 7.4), 1 M NaCl, and 100 $\mu\text{l l}^{-1}$ 2-mercaptoethanol by sonication at 45 kHz. After removal of insoluble cell debris by centrifugation at 12,000 rpm for 30 min, the supernatant was filtered through a 0.45- μm cellulose acetate membrane and applied to a HisTrap FF column (GE Healthcare Life Sciences, Pittsburgh, PA). Equilibration and washing were performed using a buffer containing 30 mM Tris (pH 7.4), 1 M NaCl, 20 mM imidazole, and 100 $\mu\text{l l}^{-1}$ 2-mercaptoethanol, and elution was carried out in 30 mM Tris (pH 7.4), 1 M NaCl, 400 mM Imidazole, and 100 $\mu\text{l l}^{-1}$ 2-mercaptoethanol. Fractions containing His₆-MBP-EcEEQ were pooled. The His₆-MBP tag was removed by cleaving the fused protein with 0.5 mg ml⁻¹ TEV protease [22] at a designed site in the linker and passing the digested protein through a reverse HisTrap FF column to yield the recombinant EcEEQ protein. The EcEEQ protein was further purified with a HiLoad (26/60) Superdex 200 size exclusion column (GE Healthcare Life Sciences). The final product, in 25 mM Tris (pH 7.4), 200 mM NaCl, and 100 $\mu\text{l l}^{-1}$ 2-mercaptoethanol, was concentrated to 12 mg ml⁻¹ (determined spectrometrically using a molar extinction coefficient of 14,440 M⁻¹ cm⁻¹), aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

Crystallization and X-Ray diffraction data collection

Previously, a total of 12 RNA oligos (RNA1 through RNA12) were used and/or observed in the structures of AaRNase III:RNA complexes [11]. RNA6, a 28-nucleotide stem-loop RNA derived from a canonical substrate of EcrNase III [6], was purchased from Dharmacon RNA Technologies (Chicago, IL) for this study. Prior to crystallization, the protein and RNA were incubated at room temperature for 30 min in a solution consisting of 7.6 mg ml⁻¹ EcEEQ, 0.4 mM RNA6, 50 mM MgCl₂, 300 mM NaCl, and 25 mM Tris-HCl (pH 7.4). The crystallization screening was carried out with a Mosquito crystallization robot (SPT Labtech Ltd., Hertfordshire, UK) by sitting drop vapour diffusion and the plates were incubated at 19 ± 1°C. Micro crystals appeared after 3 days in drops containing the protein-RNA solution and an equal volume of well solution (25% PEG 3350 and 0.2 M KBr in 100 mM Hepes buffer, pH 7.5) and reached suitable size for X-ray diffraction after 1–2 weeks. The crystals were flash-frozen after being soaked briefly in a cryoprotection solution containing 75% (v/v) reservoir solution and 25% (v/v) ethylene glycol. X-ray diffraction data were collected at 100 K at the Southeast Regional Collaborative Access Team (SER-CAT) insertion device beamline 22 (22-ID) of the Advanced Photon Source, Argonne National Laboratory. The data was indexed, integrated, and scaled with the HKL3000 suite [23]. Data collection and processing statistics are summarized in Table 1.

Table 1. Data collection and structure refinement statistics.

Data Collection	
Space group	P2 ₁
Cell constants	
a, b, c (Å)	56.93, 65.75, 84.52
α , β , γ (°)	90.0, 102.09, 90.0
Resolution (Å)	30.00–1.80 (1.86–1.80)*
Completeness (%)	99.5 (98.4)
Total/Unique reflections	362,165/55,901
Redundancy	6.5 (6.1)
I/ σ (I)	14.4 (3.3)
R _{merge}	0.121 (0.881)
R _{pim}	0.053 (0.504)
CC _{1/2}	80.6 (60.4)
Refinement	
Resolution (Å)	28.31–1.80 (1.90–1.80)
No. of reflections	55,856 (7,711)
R _{work} /R _{free}	0.171/0.202
No. of atoms/B-factors (Å ²)	
Protein	3,617/26.4
RNA	1,196/27.84
Water	418/32.6
Mg ²⁺ , K ⁺ , Cl ⁻	15/29.69
Ethylene glycol, Tris	72/40.50
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.018
Ramachandran plot (%)	
Favoured	98.88
Allowed	1.12
Outliers	0

*Values in parentheses are for the highest-resolution shell.

Structure solution and refinement

The structure of the EcEEQ:RNA6 complex was solved by molecular replacement (MR) using PHASER [24]. As mentioned above, the search model was the AaRNase III:RNA9 structure [12] after solvent molecules and ions were removed. The sequence identity between AaRNase III and EcEEQ is 33%. Each RNase III:RNA complex contains two RNase III and two RNA molecules. RNA9 contains 22 nucleotide residues. Therefore, 44 nucleotides are present in the AaRNase III:RNA9 search model, whereas 56 nucleotides are present in the EcEEQ:RNA6 structure. Nonetheless, the MR solution was unique with high LLG and TFZ scores (LLG = 445; TFZ = 15.9). Starting with the MR solution, the phases were improved with phenix_mr_rosetta [25] by density- and energy-guided model optimization and iterative model rebuilding. Although the current version of phenix_mr_rosetta did not work on nucleic acids structures, the phasing power of the EcEEQ protein were so strong that the difference Fourier electron density revealed the structure of two RNA6 molecules in their entirety. The initial RNA6 molecules that we built with COOT [26] were remodelled with ERRASER (Enumerative Real-Space Refinement ASsisted by Electron density under Rosetta) maintained at the ROSIE server [27,28]. Further adjustment and refinement of the structure were carried out with COOT [26] and PHENIX [29]. The quality of the final structure was validated on the Worldwide PDB (wwPDB) Validation Server [30]. The structure refinement statistics are summarized in Table 1. Illustrations were prepared using PyMOL (Schrodinger, LLC).

Based on the EcEEQ:RNA6 structure, model complexes for the intermediate and precleavage states of EcrNase III were derived by adjusting the torsion angles along the C4'-C5'-O5'

-P-O1P chain and breaking or making the P-O bonds [12]. Similarly, model complexes for the intermediate and precleavage states of ScRnt1p were derived based on the structure of its postcleavage complex [16]. No adjustments were made to other components of the cleavage assemblies because the $CD^{MgA, MgB}$ values (3.6 or 3.7 Å) mimic that at the intermediate state (3.5 Å) observed for RNase H1 and DNAP η [20,21].

Acknowledgments

We thank Lan Jin and He Song for discussion and Joshua Rose and Alexander Wlodawer for reading the manuscript. X-ray diffraction data were collected at the SER-CAT 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability statement

The atomic coordinates and structure factors for the EcEEQ:RNA6 complex has been deposited with the Protein Data bank (PDB: 7R97). <https://www.rcsb.org/>

Funding

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

ORCID

Sudhaker Dharavath  <http://orcid.org/0000-0002-5973-0214>
 Gary X. Shaw  <http://orcid.org/0000-0002-2510-3538>
 Xinhua Ji  <http://orcid.org/0000-0001-6942-1514>

References

- [1] Robertson HD, Webster RE, Zinder ND. Purification and properties of ribonuclease III from *Escherichia coli*. *J Biol Chem*. 1968;243:82–91.
- [2] Nicholson AW. RNAi: a guide to gene silencing. Hannon GJ, ed. New York: Cold Spring Harbor Laboratory Press, Cold Spring Harbor; 2003. p. 149–174.
- [3] Svoboda P, Di Cara A. Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci*. 2006;63:901–908.
- [4] Court DL, Belasco JG, Brawerman G, editors. Control of messenger RNA stability. RNA processing and degradation by RNase III. New York: Academic Press; 1993. p. 71–116.
- [5] Liu Z, Wang J, Cheng H, et al. Cryo-EM structure of human Dicer and its complexes with a pre-miRNA substrate. *Cell*. 2018;173:1191–1203 e1112.
- [6] Gan J, Tropea JE, Austin BP, et al. Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*. 2006;124:355–366.
- [7] Jin L, Song H, Tropea JE, et al. The molecular mechanism of dsRNA processing by a bacterial Dicer. *Nucleic Acids Res*. 2019;47:4707–4720.
- [8] Weinberg DE, Nakanishi K, Patel DJ, et al. The inside-out mechanism of Dicers from budding yeasts. *Cell*. 2011;146:262–276.
- [9] Yang D, Buchholz F, Huang Z, et al. Short RNA duplexes produced by hydrolysis with *Escherichia coli* RNase III mediate effective RNA interference in mammalian cells. *Proc Natl Acad Sci USA*. 2002;99:9942–9947.
- [10] Nicholson AW. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley interdisciplinary reviews. RNA*. 2014;5:31–48.
- [11] Court DL, Gan J, Liang Y-H, et al. RNase III: genetics and function; structure and mechanism. *Annu Rev Genet*. 2013;47:405–431.
- [12] Gan J, Shaw G, Tropea JE, et al. A stepwise model for double-stranded RNA processing by ribonuclease III. *Mol Microbiol*. 2008;67:143–154.
- [13] Gan J, Tropea JE, Austin BP, et al. Intermediate states of ribonuclease III in complex with double-stranded RNA. *Structure (Camb)*. 2005;13:1435–1442.
- [14] Xiao J, Feehery CE, Tzertzinis G, et al. E. coli RNase III(E38A) generates discrete-sized products from long dsRNA. *RNA*. 2009;15:984–991.
- [15] Yang W, Lee JY, Nowotny M. Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell*. 2006;22:5–13.
- [16] Liang YH, Lavoie M, Comeau MA, et al. Structure of a eukaryotic RNase III postcleavage complex reveals a double-ruler mechanism for substrate selection. *Mol Cell*. 2014;54:431–444.
- [17] Blaszczyk J, Tropea JE, Bubunenkov M, et al. Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure*. 2001;9:1225–1236.
- [18] Zamore PD. Thirty-three years later, a glimpse at the ribonuclease III active site. *Mol Cell*. 2001;8:1158–1160.
- [19] Mantina M, Chamberlin AC, Valero R, et al. Consistent van der Waals radii for the whole main group. *J Phys Chem A*. 2009;113:5806–5812.
- [20] Nowotny M, Yang W. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J*. 2006;25:1924–1933.
- [21] Batra VK, Beard WA, Shock DD, et al. Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure*. 2006;14:757–766.
- [22] Kapust RB, Tözsér J, Fox JD, et al., Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. 2001;Protein Eng. 14:993–1000.
- [23] Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997;276:307–326.
- [24] McCoy AJ, Grosse-Kunstleve RW, Adams PD, et al. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40:658–674.
- [25] Terwilliger TC, Dimaio F, Read RJ, et al. phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J Struct Funct Genomics*. 2012;13:81–90.
- [26] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. 2004;Acta Crystallogr D. 60:2126–2132.
- [27] Chou FC, Sripakdeevong P, Dibrov SM, et al. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods*. 2013;10:74–76.
- [28] Sripakdeevong P, Kladwang W, Das R. An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proc Natl Acad Sci USA*. 2011;108:20573–20578.
- [29] Adams PD, Afonine PV, Bunkoczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*. 2010; 66: 213–221.
- [30] Gore S, Sanz Garcia E, Hendrickx PMS, et al. Validation of structures in the protein data bank. *Structure*. 2017;25:1916–1927.