Doctor Thesis

Structural and biochemical studies of two novel metagenome-derived RNases H1

(メタゲノム法により単離した２種類の新規リボヌクレアーゼ H1 の構造学的、生化学的研究)

Nguyen, Tri-Nhan

April 2013

International Program of Frontier Biotechnology
Division of Advanced Science and Biotechnology
Graduate School of Engineering
Osaka University, Japan

TABLE OF CONTENTS

Abbreviations

Bh-RNase H1:	RNase H1 from *Bacillus halodurans*
Bh-RNases HC:	Bh-RNase H1 derivative without the hybrid binding domain
Bh-RNases HC*:	inactive active-site mutants of Bh-RNases HC
Bst-RNase H3:	RNase H3 from *Bacillus stearothermophilus*
CBB:	Coomassie brilliant blue
CD:	circular dichroism
D13-R4-D12/D29:	29 bp DNA_{13}-RNA_4-DNA_{12}/DNA duplex
D13-R5-D11/D29:	29 bp DNA_{13}-RNA_5-DNA_{11}/DNA duplex
D13-R6-D12/D29:	29 bp DNA_{13}-RNA_6-DNA_{10}/DNA duplex
D15-R1-D13/D29:	29 bp DNA_{15}-RNA_1-DNA_{13}/DNA duplex
DSC:	differential scanning calorimetry
dsRNA:	double-stranded RNA
dsRNase:	dsRNA-dependent RNase
Ec-RNase H1:	RNase H1 from *Escherichia coli*
EDTA:	ethylenediaminetetraacetic acid
Gau-RNase H1:	RNase H1 from *Gemmatimonas aurantiaca*
HBD:	hybrid binding domain
HIV-RNase H:	RNase H domain of HIV-1 reverse transcriptase
Hs-RNase H1:	human (*Homo sapiens*) RNase H1
Hs-RNases HC:	Hs-RNase H1 derivative without the hybrid binding domain
Hs-RNases HC*:	inactive active-site mutants of Hs-RNases HC
IPTG:	isopropyl thio-β-D-galactoside
Mj-RNase H2:	RNase H2 from *Methanococcus jannaschii*
PDB:	Protein data bank

PEG:	polyethyleneglycol
R12/D12:	12 bp RNA/DNA hybrid
R12/R12:	12 bp RNA/RNA duplex
R9-D9/D18:	18 bp RNA_9-DNA_9/DNA duplex
RNase H:	ribonuclease H
SD:	Shine-Dalgarno
Sto-RNase H1:	RNase H1 from *Sulfolobus tokodaii*

CHAPTER 1
General introduction

1.1. Ribonuclease H (RNase H)

RNase H (EC 3.1.26.4) is a sequence-nonspecific endonuclease that cleaves an RNA strand of RNA/DNA hybrids [1]. It hydrolyzes phosphodiester bonds of RNA to yield 3'-hydroxyl and 5'-phosphate ends at cleavage sites (Fig. 1-1).

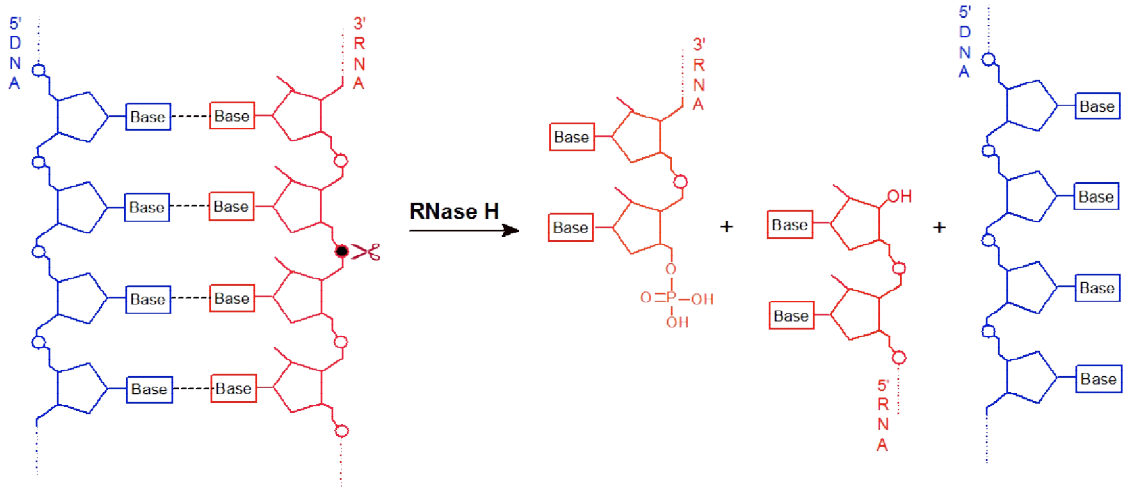

Fig. 1-1. Enzymatic reaction of RNase H.

RNase H is one of the most widely conserved enzymes; it is present in prokaryotes, eukaryotes, and retroviruses. Phylogenetic analyses using RNase H sequences have proposed the classification of two types: type 1 (prokaryotic RNase H1, eukaryotic RNase H1 and retroviral RNase H) and type 2 (prokaryotic RNases H2 and H3, and eukaryotic RNase H2) [2-4]. Two types of RNases H not only show poor amino acid sequence similarities to each other, but also differ in substrate specificity. Type 1 RNases H does not hydrolyze double-stranded DNA (dsDNA) containing a single ribonucleotide [5-7]. They require at least four consecutive ribonucleotides embedded in a dsDNA sequence to cleave [5,6]. In contrast, type 2 RNases H, except for bacterial RNases H3, hydrolyze dsDNA containing a single ribonucleotide at the 5' side of the ribonucleotide [8-12]. Conflicting results indicating the

incapability [6,7] and capability [13,14] of bacterial RNase H3 to cleave this substrate have been reported. All RNases H, except for eukaryotic RNase H2 which is functional in a hetero trimeric form [11-16], are functional in a monomeric form [3].

1.2. Physiological roles of RNase H

Although the physiological functions of RNase H are still not fully understood, this enzyme is thought to play several roles in DNA replication, DNA repair and RNA transcription [17-29]. RNases H may be involved in the removal of the initiator RNA primers from Okazaki fragments during the DNA replication process of lagging strands [17-22]. Recently, replicative DNA polymerases have been found to incorporate significant numbers of ribonucleotides during DNA synthesis [32], and so a role for RNase H, especially RNase H2, in their removal may be crucial to maintain genomic stability [22-28]. Furthermore, removal of R-loops, which accumulate when a transcription bubble collides with a replication fork, can be achieved by overexpressing RNase H [29-31].

Bacteria and eukaryotes usually contain multiple RNases H in a single cell [3]. For example, *Escherichia coli* and *Saccharomyces cerevisiae* contain RNases H1 and H2, and *Bacillus subtilis* contains RNases H2 and H3. While loss of all functional RNases H renders *B. subtilis* unable to grow [33]; it only leads to sensitivity to temperature in *E. coli* [33], and greater sensitivity to alkylating agent in *S. cerevisiae* [34]. In mammals, both type 1 and type 2 RNases H have essential roles. Deletion of RNase H1 in mouse impairs mitochondrial DNA replication causing embryonic lethality [21]. Mutations in human RNase H2 subunits cause Aicardi-Goutières syndrome, a genetic disorder which severely affects the nervous system by activating the innate immune system [35]. In retroviruses, RNase H activity is necessary for their replication [36], and the enzyme has thus been regarded as the target of anti-HIV drug development [37]. Recently, vinylogous ureas, which bind at the subunit interface and inhibit RNase-H activity, have been identified by high-throughput screening [38-40]. RNase H also

plays a crucial role in antisense therapy [41].

1.3. Overall structure and active site of RNase H

The crystal structure of *E. coli* RNase H1 (Ec-RNase H1) determined in 1990 is the first RNase H structure. This structure reveals a novel α/β fold containing four acidic residues (Asp or Glu) in the catalytic center (Fig. 1-2) [42,43]. A similar fold and related active site are found later in the RNase H domain of reverse transcriptase (RT) [44], RNase H2 [45] and RNase H3 [46] (Fig. 1-2A). However, slight differences between two types of RNases H are detected in the positions of the four acidic active-site residues in the primary structures and their steric configurations. In terms of the positions of the active-site residues in the primary structure, the first aspartate and second glutamate residues of type 1 RNases H are located distantly, while those of type 2 RNases H are located adjacently to each other (Fig. 1-3). In terms of the steric configurations of these residues, the spatial positions of the second glutamate and fourth aspartate/glutamate residues are not well conserved between two types of RNases H (Fig. 1-2B). Because the four acidic active-site residues are arranged in the order of Asp, Glu, Asp, and Asp/Glu (Asp or Glu) in almost every RNase H sequence, they are termed DEDD or DEDE active-site motif [3].

1.4. Substrate recognition mechanism of RNase H

To my knowledge, beside the crystal structure of *Thermotoga maritima* RNase H2 (Tm-RNase H2) with a C-terminal truncation in complex with a single ribonucleotide-embedded dsDNA substrate and metal ions [47], the crystal structure of the RNase H enzyme in complex with an RNA/DNA hybrid and metal ions is only available for Bh-RNase HC* and Hs-RNases HC* (Fig. 1-4) [48,49]. Bh- and Hs-RNases HC* represent the inactive active-site mutants of Bh- and Hs-RNases HC, which represent *B. halodurans* RNase H1 and human (*Homo sapiens)* RNase H1 without the hybrid binding domain (HBD),

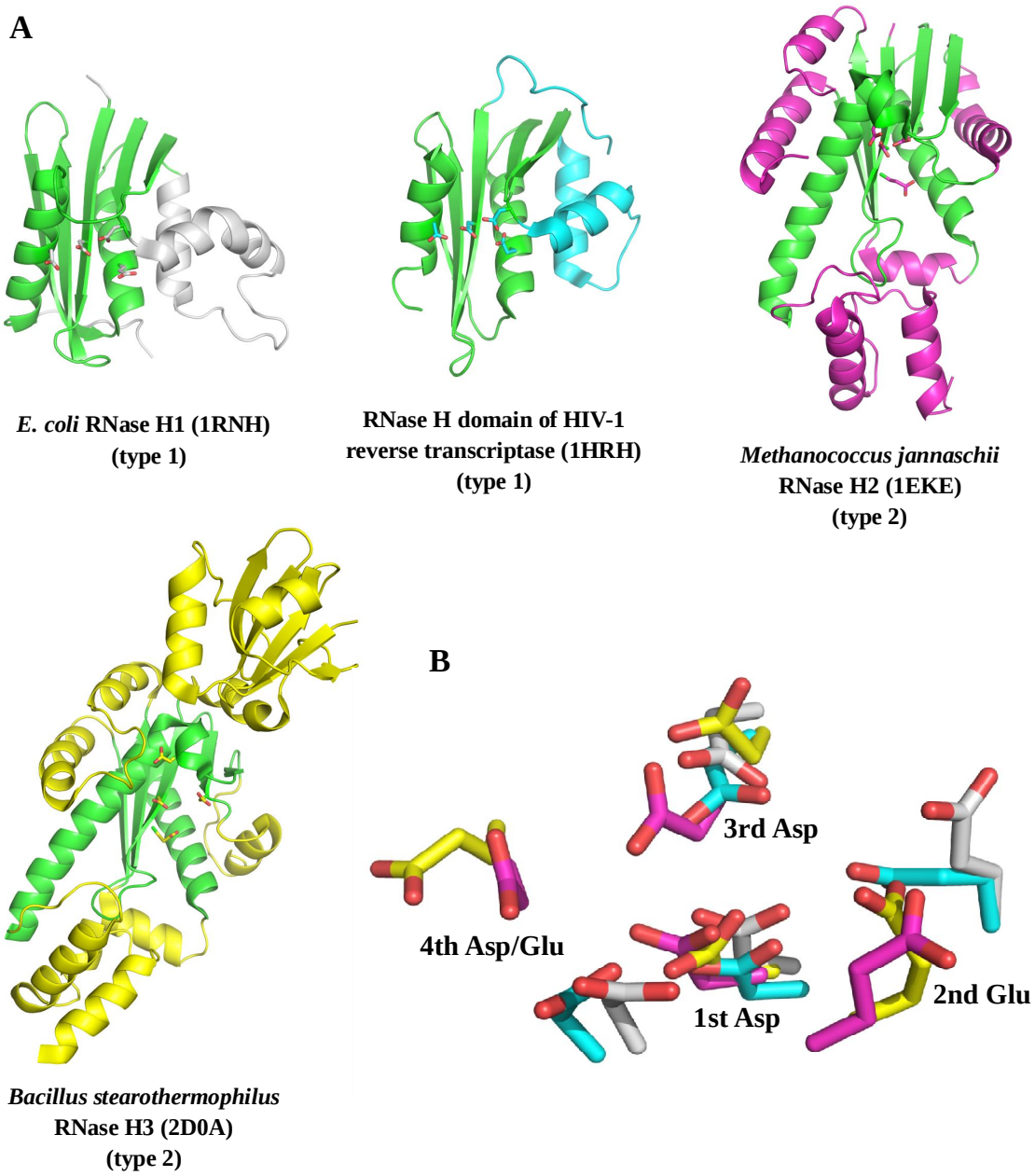

Fig. 1-2. Crystal structures of type 1 and type 2 RNases H. **(A)** Overall structures of *E. coli* RNase H1 (Ec-RNase H1), RNase H domain of HIV-1 reverse transcriptase (HIV-RNase H), *Methanococcus jannaschii* RNase H2 (Mj-RNase H2) and *B. stearothermophilus* RNase H3 (Bst-RNase H3). Ec-RNase H1 and HIV-RNase H represent type 1 RNases H, while Mj-RNase H2 and Bst-RNase H3 represent type 2 RNases H. The common α/β fold shared by these structures (RNase H-fold) is colored green and the four acidic active-site residues are shown in stick models. The oxygen atom is colored red. **(B)** Superimposition of the side chains of the four acidic active-site residues of HIV-RNase H (cyan), Mj-RNase H2 (magenta) and Bst-RNase H3 (yellow) onto those of Ec-RNase H1 (grey). The oxygen atom is colored red. The relative positions of the four acidic active-site residues in the primary structure are shown (1st-4th).

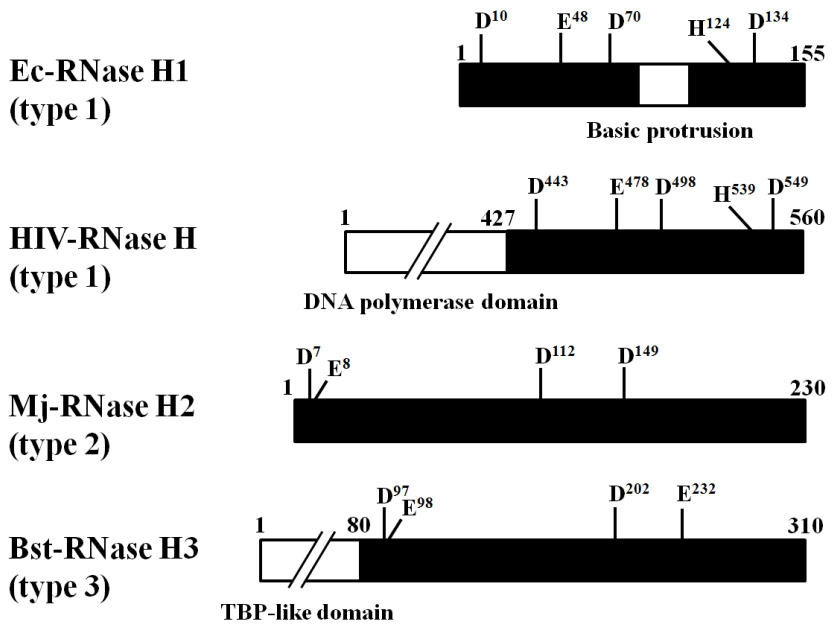

Fig. 1-3. Schematic representation of the primary structures of Ec-RNase H1, HIV-RNase H, Mj-RNase H2 and Bst-RNase H3. Black box represents the RNase H domain. The basic protrusion of Ec-RNase H1, DNA polymerase domain of HIV-1 reverse transcriptase and TATA-box binding protein-like domain (TBP-like domain) of Bst-RNase H3 are shown. The four conserved acidic active-site residues are shown. Another nonconserved active-site residue of type 1 RNases H (His for Ec-RNase H1 and HIV-RNase H) is also shown. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

respectively. HBD is a small (approximately 40 residues) N-terminal domain important for substrate binding [50-52]. In this thesis, the protein-substrate complex is simply designated as the protein complex. According to these protein complex structures, RNase H proteins bind to the minor groove of the substrate, such that one groove containing the active site on the protein surface accommodates the RNA strand and the other containing the phosphate-binding pocket accommodates the DNA strand. Two grooves are separated by a ridge composed of highly conserved residues (Fig. 1-4). The RNA strand of the hybrid is recognized by Bh- or Hs-RNase HC* through direct contacts with five or four consecutive 2'-OH groups, respectively. Whereas, the DNA strand is recognized by its ability to adopt a B

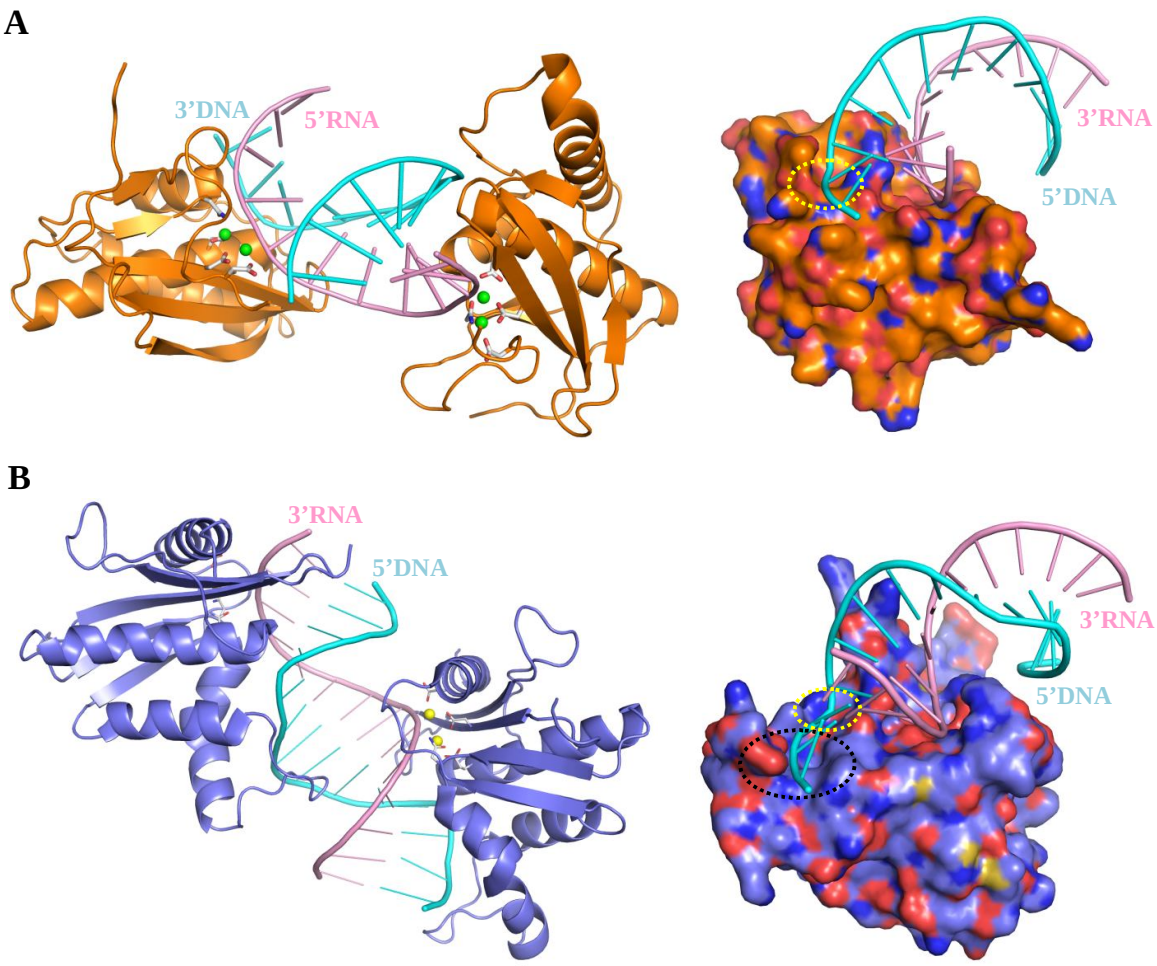

Fig. 1-4. The structures of Bh- and Hs-RNase HC* in complex with the substrate and metal ions. **(A)** The overall structure and protein surface of Bh-RNase HC* (orange) in complex with 12 bp RNA/DNA hybrid and two magnesium ions (green spheres) (PDB code 1ZBI). **(B)** The overall structure and protein surface of Hs-RNase HC* (slate) in complex with 14 bp RNA/DNA hybrid and two calcium ions (yellow spheres) (PDB code 2QKK). RNA and DNA are colored pink and cyan, respectively. The active-site residues are shown by stick models, in which the oxygen atom is colored red. The conserved phosphate-binding pocket is indicated by a yellow dotted-line circle. The DNA-binding channel formed at the basic-protrusion of Hs-RNase HC* is indicated by a black dotted-line circle.

form conformation [48,49]. A conserved binding site of the phosphate two base pairs from the scissile bond, termed phosphate-binding pocket, is observed in both Bh- and Hs-RNase HC* complexes (Fig. 1-4). This pocket is suggested to play a role in anchoring the B form DNA and contributes to the specificity for an RNA/DNA hybrid. In addition to the

phosphate-binding pocket, Hs-RNase HC contains a second DNA-binding site, which is a channel formed at the basic-protrusion (Fig. 1-4B).

1.5. Catalytic mechanism of RNase H

According to the structures of Bh- and Hs-RNases HC* complexes, two magnesium ions and two calcium ions, respectively, bind to the active site. These metal ions are termed metal ions A and B. Based on this observation, a two-metal ion catalysis mechanism has been proposed for RNases H [53]. According to this mechanism, metal ion A is required for substrate-assisted nucleophile formation, metal ion B destabilizes the enzyme-substrate complex, and metal ions A and B together stabilize the transition state and facilitate product release (Fig. 1-5). The crystal structures of Bh-RNase HC* complexed with the reaction

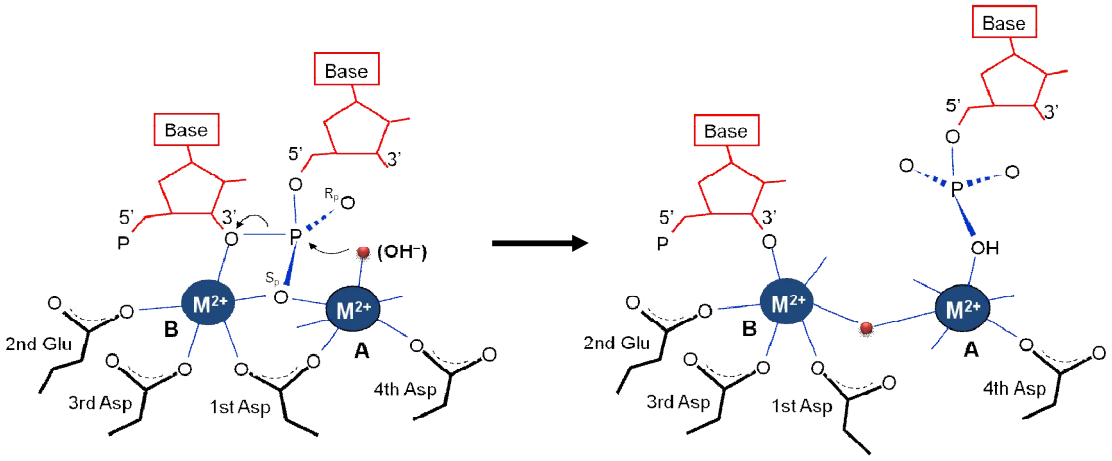

Fig. 1-5. Schematic representation of the two-metal ion catalysis mechanism proposed for RNase H. The side chains of the first aspartate (1st Asp), second glutamate (2nd Glu), third aspartate (3rd Asp) and fourth aspartate (4th Asp) residues, which form the active site of RNase H, are shown. The fourth aspartate residue is replaced by Glu in RNase H3. The attacking hydroxyl ion that is coordinated by metal ion A is indicated as (OH^-). Water molecules are shown as red spheres. In the substrate complex, metal ion A is coordinated by six ligands in an nearly perfect octahedral geometry, while metal ion B is coordinated by five ligands in an irregular geometry. In the product complex, both metal ions A and B are coordinated by six ligands in an octahedral geometry.

product and intermediate mimic revealed that the metal-metal distance and coordination of the metal ions in these complexes are different from those in the substrate complex [54] (Fig. 1-5). The metal-metal distance in the product complex increases as compared to that in the substrate complex, suggesting that reduction of this distance is necessary for the hydrolysis reaction to occur. The transformation from the irregular coordination by five ligands in the substrate complex to the octahedral coordination in the product complex of metal ion B suggests that metal ion B destabilizes the substrate-enzyme complexes and thus reduces the energy barrier between the substrate and product states [54].

In addition to the four conserved acidic residues in the active site, a poorly conserved fifth residue has been reported to play an important but nonessential role in catalysis of various type 1 RNases H [48,49,55,56]. In Bh-RNase HC* complexes, Glu188 has great conformational flexibility and can either participate in metal ion A coordination (Fig. 1-6) or swing out from the active site [48]. Its equivalent in Ec-RNase H1 (His124) also has high mobility [57]. Such mobility is considered to perturb the coordination of metal ion A and

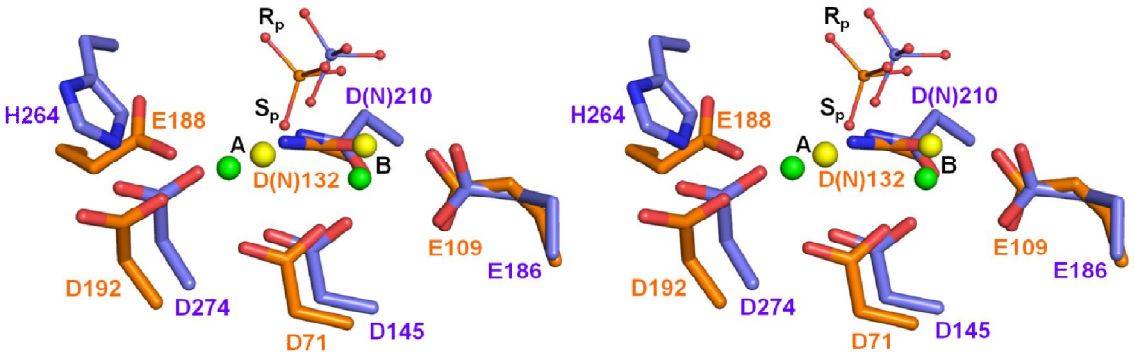

Fig. 1-6. Stereoview of the active-site structures of Hs-RNase HC* (slate) superimposed onto that of Bh-RNase HC* (orange). Five active-site residues are shown by stick models, in which the oxygen and nitrogen atoms are colored red and blue, respectively. The scissile phosphate groups are shown in ball-and-line models, in which the oxygen atom is colored red and the phosphorus atoms of the Hs-RNase HC* and Bh-RNase HC* complexes are colored slate and orange, respectively. Two Mg^{2+} ions A and B in the Bh-RNase HC* complex and two Ca^{2+} ions A and B in the Hs-RNase HC* complex are also shown as green and yellow spheres, respectively. Pro-R_p and pro-S_p oxygens are labeled as R_p and S_p, respectively.

promote product release. The equivalent residue in Hs-RNase HC (His264, Fig. 1-6), which is located on a mobile loop, has been suggested to facilitate product release by dislodging the 5'-phosphate of the reaction product [49].

1.6. Molecular diversities of type 1 RNases H

Phylogenetic analyses of the type 1 RNase H sequences indicate that type 1 RNases H are further divided into four groups; bacterial and eukaryotic RNases H1, retroviral RNase H (RNase H domain of RT) and *Sulfolobus tokodaii* RNase H1 (Sto-RNase H1) homolog [58] (Fig. 1-7). Sto-RNase H1 [58] and *Halobacterium* sp. NRC-1 RNase H1 (Halo-RNase H1) [59], which represent archaeal RNases H1, have been shown to be active both *in vivo* and *in vitro*. However, database searches using the Sto-RNase H1 sequence indicate that not only the archaeal genomes, but also the bacterial and eukaryotic genomes contain the genes encoding a Sto-RNase H1 homolog.

Characteristics common to the amino acid sequences of the Sto-RNase H1 homologs are the lack of a basic protrusion, which has been shown to be important for substrate binding in Ec- and Hs-RNases H1 [49,60,61], and the presence of an arginine residue (Arg118 in Sto-RNase H1 and Arg182 in Halo-RNase H1) equivalent to the fifth active-site residue (e.g. His124 in Ec-RNase H1, Glu188 in Bh-RNase H1 and His264 in Hs-RNase H1), which has been shown to be important for catalytic function of various RNases H1 as discussed above (Fig. 1-8).

Sto- and Halo-RNases H1 exhibit unique enzymatic properties. They cleave the (5')RNA-DNA(3') junction of the Okazaki fragment-like substrate [58,59]. Interestingly, Sto-RNase H1 exhibit dsRNase activity, which degrades the RNA strand of a double-stranded RNA (dsRNA) [58]. It has been reported that retroviral RNases H cleave the (5')RNA-DNA(3') junction [62] and exhibit dsRNase activity [63]. Thus, Sto-RNase H1 homologs are more closely related to retroviral RNases H than bacterial and eukaryotic type 1

RNases H in enzymatic properties.

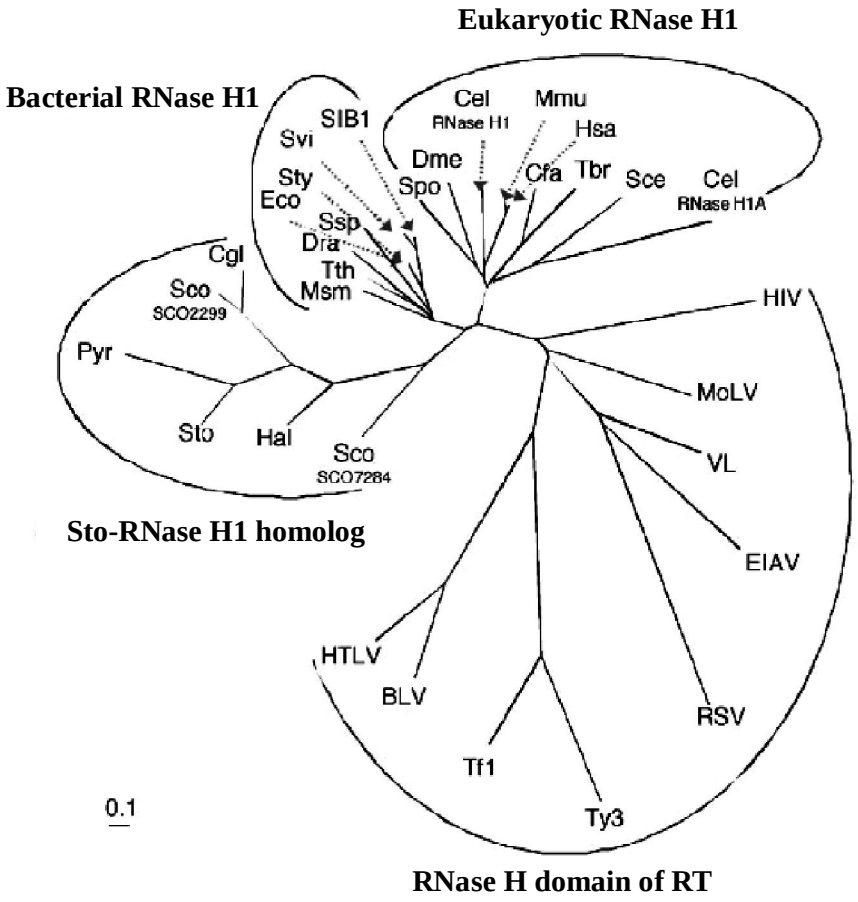

Figure 1-7. Phylogenetic tree derived from type 1 RNase H amino acid sequences [58]. The distance of the branch from a diverging point is proportional to the frequency on average of the amino acid substitution at each position. The scale bar corresponds to this frequency of 0.1. The abbreviations for the organisms are as follows: Eco, *Escherichia coli*; Sty, *Salomonella typhimurium*; SIB1, *Shewanella* sp. SIB1; Svi, *Shewanella violacea*; Syn, *Synechocystis* sp. PCC6803; Tth, *Thermus thermophilus*; Dra, *Deinococcus radiodurans*; Msn, *Mycobacterium smegmatis*; Sce, *Saccharomyces cerevisiae*; Spo, *Schizosaccharomyces pombe*; Cfa, *Crithidia fasciculate*; Tbr, *Trypanosoma brucei*; Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Mmu, *Mus musculus*; Hsa, *Homo sapiens*; Hal, *Halobacterium* sp. NRC-1; Sto, *Sulfolobus tokodaii*; Pae, *Pyrobaculum aerophilum*; Sco, *Streptomyces coelicolor*; Cgl, *Corynebacterium glutamicum*; BLV, bovine leukemia virus; EIAV, equine infectious anemia virus; HIV, human immunodeficiency virus 1; HTLV, human T-cell leukemia virus; MLV, moloney leukemia virus; RSV, rous sarcoma virus; VL, visna lentivirus; Tf1, Tf1 retrotransposon; and Ty3, Ty3 retrotransposon.

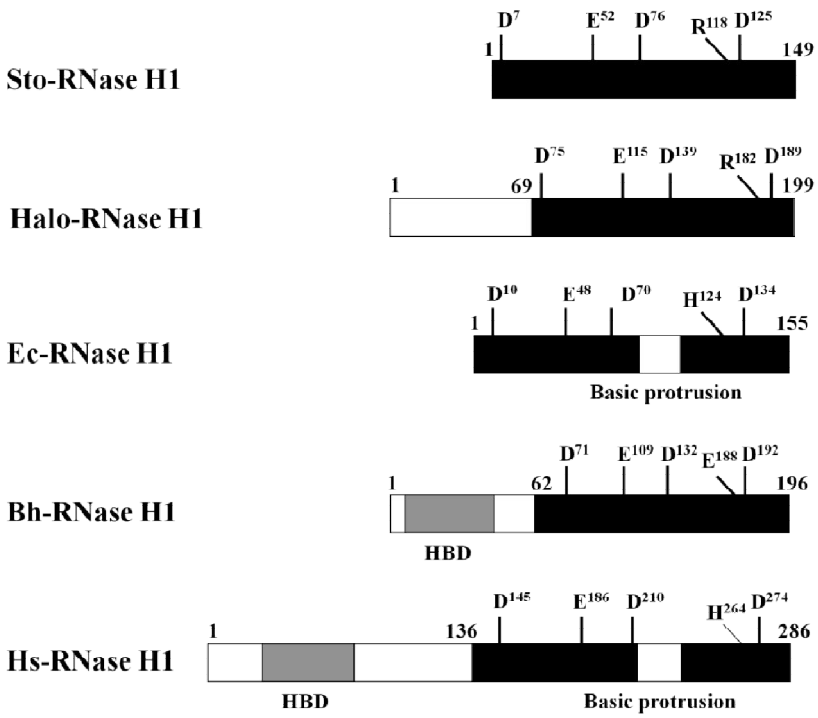

Fig. 1-8. Schematic representation of the primary structures of Sto-, Halo-, Ec-, Bh- and Hs-RNases H1. Solid box represents the RNase H domain and gray box represents the hybrid binding domain (HBD). The basic protrusion, which is located in the middle of the sequences of Ec- and Hs-RNases H1, four conserved acidic active-site residues, and one nonconserved active-site residue, are shown. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

The crystal structure of Sto-RNase H1 is the first structure of Sto-RNase H1 homologs [56] (Fig. 1-9A). It shows high similarity to those of Ec-RNase H1, HIV-RNase H and Bh-RNase HC. The steric configurations of the four acidic active-site residues and the RNA-binding groove are well conserved in the Sto-RNase H1 structure. However, the amino acid residues forming the phosphate-binding pocket are not well conserved in its structure. The weak conservation of the DNA-binding groove has been suggested to account for the ability of Sto-RNase H1 to cleave dsRNA [56]. It is noted that a significant difference is observed at the C-terminal region of Sto-RNase H1 as compared to those of Ec-RNase H1, HIV-RNase H and Bh-RNase HC. While the C-terminal tails of Ec-RNase H1, HIV-RNase H and Bh-RNase HC are almost fully exposed to the solvent, that of Sto-RNase H1 is anchored

to the core region through a disulfide bond, hydrogen bonds and hydrophobic interactions (Fig. 1-9B). Differential scanning calorimetry (DSC) measurements indicate that Sto-RNase H1 is a hyperstable protein with the melting temperature (T_m) of 102°C. The mutations to eliminate the disulfide bond or C-terminal tail greatly destabilize the protein, indicating that anchoring of the C-terminal tail is responsible for hyperstabilization of Sto-RNase H1 [56].

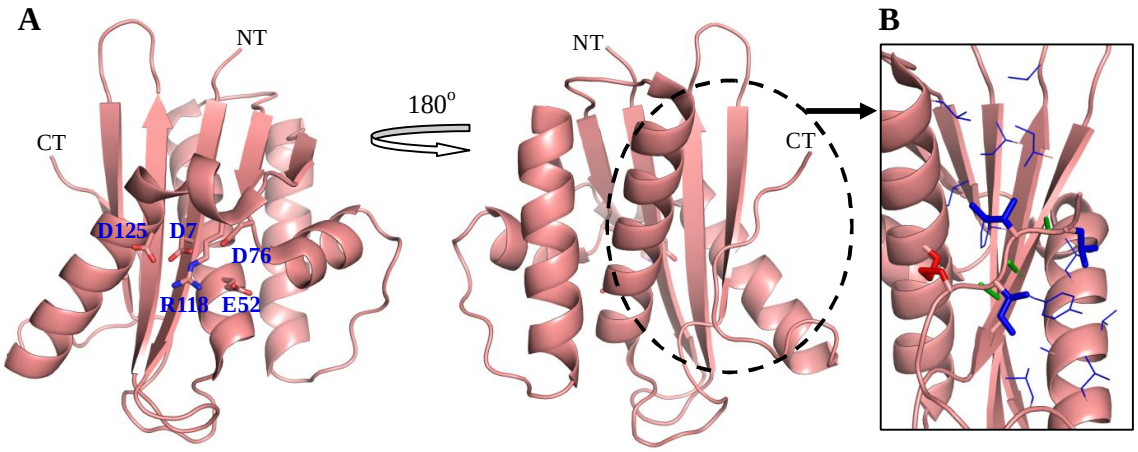

Fig. 1-9. Crystal structure of Sto-RNases H1. (**A)** Overall structures in different view direction (PDB code 2EHG). Five active-site residues of Sto-RNase H1 (Asp7, Glu52, Asp76, Arg118, and Asp125) are shown by stick models, in which the oxygen and nitrogen atoms are colored red and blue, respectively. NT and CT represent N- and C-termini. **(B)** The structure of the C-terminal tail. The side chains of two cystein residues, which form the disulfide bond, are shown as red stick models. The side chains of the hydrophobic residues in the C-terminal tail are shown as blue stick models; and the side chains of the hydrophobic residues located close to these residues are shown as thin blue stick model. The hydrogen bonds formed between C-terminal polypeptide chain and a β strand in the core region are colored green.

1.7. The objective of the study

In this study, I search for novel RNases H with unique structure and/or function, followed by studies on their structural and biochemical properties in order to deepen our understanding on the substrate recognition mechanism, catalytic mechanism and molecular diversities of RNases H.

Over the past decade, our knowledge on the molecular diversities of RNases H has gradually increased, mainly those on prokaryotic RNases H. In the late 1990s, it was reported that the genes encoding type 2 RNases H are present in the genomes from all three kingdoms of living organisms, bacteria, archaea and eukaryotes; whereas those encoding type 1 RNases H are present only in the genomes from bacteria and eukaryotes [2]. It was also reported that the genes encoding RNase H3 are found only in the genomes from bacteria [2]. However, later on, the genes encoding archaeal RNases H1, which are a Sto-RNase H1 homolog, are identified; and these proteins are shown to exhibit RNase H activity both *in vivo* and *in vitro* [58,59]. The gene encoding RNase H3 is also found from the archaeal genome (the euryarchaeota *Methanosphaera stadtmanae*), although its enzymatic activity has not been studied yet [64]. These results encourage us to discover the genes encoding novel RNase H.

The two-metal ion catalysis mechanism has been proposed not only for RNase H [53] but also for many other enzymes catalyzing phosphoryl transfer reactions [65]. However, it is still controversial whether RNase H always uses two metal ions for catalysis, because the amino acid side chains that can serve as general acids and bases are available at the active site. It is noted that nucleotidyl-transfer enzymes, which synthesize, degrade and rearrange DNA and RNA, do not always use two metal ions for catalysis. The ββα-Me and HUH nucleases depend on a single metal ion [58]. Both RNases H and these one-metal ion-dependent enzymes generate 5’-phosphate and 3’-OH products, and share a functionally equivalent metal ion (metal ion B in the two-metal ion catalysis) [66]. Therefore, these results also encourage us to isolate the genes encoding novel RNases H that utilize a single metal ion for

catalysis.

The crystal structures of Bh- and Hs-RNases HC* in complex with the RNA/DNA substrate has provided insight into the substrate recognition mechanism of RNase H. However, both Bh- and Hs-RNases HC* lack HBD, which is important for substrate binding. The structures of these proteins in complex with the substrate may not represent that of the protein in complex with the substrate in a fully functional form. Therefore, it is necessary to determine the structure of the protein without truncation in complex with the RNA/DNA substrate for clarifying and generalizing the substrate recognition mechanism of RNase H. Moreover, the mechanism by which Sto-RNase H1 and retroviral RNases H recognize and cleave dsRNA remains unclear. The isolation of a novel Sto-RNase H1 homolog, followed by the structural and biochemical characterization of this protein may provide insight into the mechanism of dsRNA hydrolysis.

This thesis consists of six Chapters. In Chapter 1, the background of this work is summarized.

In Chapter 2, the metagenomic library was prepared from leaf-branch compost and used to screen for the RNase H genes. Twelve different genes encoding type 1 RNases H (LC1~LC12 RNases H1) were identified. Eleven of them encode novel RNases H, which show the 40-72% amino acid sequence identities to those available from database. Phylogenetic analysis indicates that LC1~LC9-RNases H1 are classified as bacterial RNases H1 and LC10~LC12-RNases H1 are classified as Sto-RNase H1 homologs. LC9-RNase H1 is the first natural enzyme, which lacks a typical DEDD/E active-site motif, but exhibits enzymatic activity.

In Chapter 3, the crystal structure of LC11-RNase H1 was determined. It highly resembles the Sto-RNase H1 structure, except that it lacks the C-terminal tail, which is anchored to the core region of Sto-RNase H1. LC11-RNase H1 did not show a preference to Mg^{2+} or Mn^{2+} ion for activity, unlike Sto-RNase H1, which shows a preference to Mg^{2+} ion.

Unlike Sto-RNase H1, LC11-RNase H1 could not cleave dsRNA. I propose that RNases H or RNase H-like proteins, which do not have a clear DNA binding groove, may cleave dsRNA. LC11-RNase H1 was less stable than Sto-RNase H1 by 37°C in T_m. It is destabilized not only by the removal of the C-terminal tail, but also by the increase in the cavity volume and the number of buried charged residues. However, destabilization by the removal of the C-terminal tail accounts for two thirds of the difference in stability between LC11-RNase H1 and Sto-RNase H1.

In Chapter 4, the crystal structure of LC11-RNase H1 in complex with an RNA/DNA substrate was determined. Beside the common phosphate-binding pocket, LC11-RNase H1 has a unique DNA-binding channel. The active-site residues of LC11-RNase H1 are located farther away from the scissile phosphate group than those of Bh-RNase HC and Hs-RNase HC. LC11-RNase H1 interacts with four non-consecutive 2'-OH groups of the RNA strand, which leads to a dramatic decrease in the ability to cleave a DNA-RNA-DNA/DNA substrate containing four ribonucleotides as compared to those containing five and six ribonucleotides.

In Chapter 5, the crystal structure of metagenome-derived LC9-RNase H1 was determined. The LC9-RNase H1 structure highly resembles the Ec-RNase H1 structure. Non-conserved glutamate and asparagine residues, and two conserved aspartate residues form the active site of LC9-RNase H1. However, the enzymatic properties of LC9-RNase H1 were similar to those of Ec-RNase H1. LC9-RNase H1 may represent bacterial RNases H1 with an atypical DEDN active-site motif, which are evolutionarily distinct from those with a typical DEDD active-site motif.

In Chapter 6, the summary of the novel findings on the study and future remarks are described.

CHAPTER 2

Identification of the genes encoding novel RNases H from compost metagenome

2.1. Introduction

As mentioned in Chapter 1, RNases H share a common DEDD/E (DEDD or DEDE) active-site motif and steric configurations of these four acidic active-site residues (Fig. 1-3, Fig. 1-2B) [3]. The fourth glutamate residue is conserved only in the bacterial RNase H3 sequences. It has been reported for various RNases H that the fourth acidic residue of a DEDD/E active-site motif is not required for activity [11,46,56,67-69]. The mutation of this residue to Ala greatly reduces the enzymatic activity but does not fully abolish it. The mutation of this residue to Asn does not seriously affect the enzymatic activity. It has also been reported for Ec-RNase H1 that the second glutamate residue of this motif is dispensable for Mn^{2+}-dependent activity [69]. Nonetheless, the DEDD/E active-site motif is almost fully conserved in all RNase H sequences available from databases, suggesting that this active-site motif is required for other functions than enzymatic activity. Alternatively, the sequence information is still insufficient to identify novel RNases H, in which all four acidic active-site residues are not necessarily conserved.

Metagenomics is the study of genetic material recovered directly from environmental sources [70]. Because more than 99% of microorganisms in nature cannot be cultivated by the conventional method [71], metagenomics is useful not only to better understand the microbial ecology but also to isolate novel genes [72-79]. Compost is one of the promising genetic resources, because a large variety of micro-organisms inhabit the compost [80-82]. Therefore, it would be informative to examine whether novel RNase H genes can be cloned from metagenomic DNA extracted from the compost.

In this chapter, twelve different RNase H genes were cloned from a metagenome library of leaf-branch compost via their ability to complement the temperature-sensitive (ts) growth

phenotype of *E. coli* MIC3001. Determination of the nucleotide sequences of these RNase H genes, followed by the phylogenetic analyses indicate that one of these RNases H lacks the second and fourth acidic residues of a DEDD/E active-site motif. These results suggest that a great number of RNases H with unique sequences still remain in nature.

2.2. Materials and methods

2.2.1. Cells and plasmids

E. coli MIC3001 [F^-, *supE44, supF58, lacY1 or Δ(lacIZY)6, trpR55, galK2, galT22, metB1, hsdR14*(r_K^- m_K^+), *rnhA399::cat, rnhB270*(Ts)] [83] and *E. coli* MIC2067 [F^-, λ^-, IN(*rnh*D-*rnh*E)1, *rnhA399::cat, rnhB716::kam*] [33] were kindly donated by M. Itaya. Plasmid pBR322 was from Takara Bio Inc. (Shiga, Japan).

2.2.2. Construction of metagenomic library and screening for RNase H gene

The sample was taken from the core (1 m below surface) of the one-year old compost made from leaves and branches (leaf-branch compost) in EXPO Park, Japan. The temperarure and pH of this compost are around 50°C and pH 7. Metagenomic DNA was extracted from 5 g (wet weight) of the compost sample using ISOIL Large Bead Kit (Nippon Gene). The total yield of DNA was 10-15 μg per gram sample. Approximately 0.5 μg of this metagenomic DNA was partially digested with *Sau*3AI, so that the DNA fragments with the size of 0.3-3 kb were obtained, and ligated into the *Bam*HI site of pBR322 (0.13 μg) to construct a metagenome library. This library was used to transform *E. coli* MIC3001 by electroporation with Gene Pulser II Electroporation System (Bio-Rad). By spreading an aliquot of these transformants on Luria-Bertani (LB)-agar plates with 50 $mg{\cdot}L^{-1}$ ampicillin and counting the number of the colonies grown at 30°C, the total number of the transformants was determined to be 3.6×10^5. The insertion check of the 20-30 plasmids isolated from these colonies randomly selected indicated that approximately 80% of these colonies contained the metagenomic DNA fragments, which varied from 0.3 to 3 kb in size. This means that the metagenome library constructed in this study contains approximately 3×10^5 different metagenomic DNA fragments, which vary from 0.3 to 3 kb in size. The rest of the transformants were then spread on LB-agar plates with 50 $mg{\cdot}L^{-1}$ ampicillin, and these plates

were incubated at 42°C. Plasmids were extracted from the colonies grown at 42°C and were used to transform *E. coli* MIC2067 to confirm that these plasmids complement the temperature-sensitive (ts) growth phenotype of this strain. Nucleotide sequences of the cloned DNA fragments were determined by a Prism 310 DNA sequencer (Applied Biosystems). Oligonucleotides for sequencing were synthesized by Hokkaido System Science.

2.2.3. Phylogenetic analysis

The amino acid sequences of 29 RNases H were chosen from those registered in the NCBI database, because they were previously used for phylogenetic analysis of type 1 RNases H [58]. A multiple sequence alignment was created by ClustalW program (ver. 1.83) [84]. The phylogenetic tree was constructed by using the TreeView program [85], based on a distant matrix calculated from the alignment by the phylip program (http://clustalw.ddbj.nig.ac.jp/top-j.html).

2.3. Results

2.3.1. Cloning of DNA fragments containing RNase H genes from metagenome library

The *E. coli* genome contains the *rnhA* and *rnhB* genes encoding RNases H1 and H2, respectively [86]. The *rnhA* mutant strain *E. coli* MIC3001 shows an RNase H-depenednt temperature-sensitive (ts) growth phenotype, and can form colonies on LB-agar plate at 30℃ but cannot form them at 42℃ [83]. Because various RNase H genes have been cloned by their abilities to complement the ts phenotype of *E. coli* MIC3001 [87-92], this strain is suitable to screen a metagenome library for the RNase H genes. The *rnhA*/*rnhB* double mutant strain *E. coli* MIC2067 also shows a ts growth phenotype, which is complemented by functional RNase H genes [33].

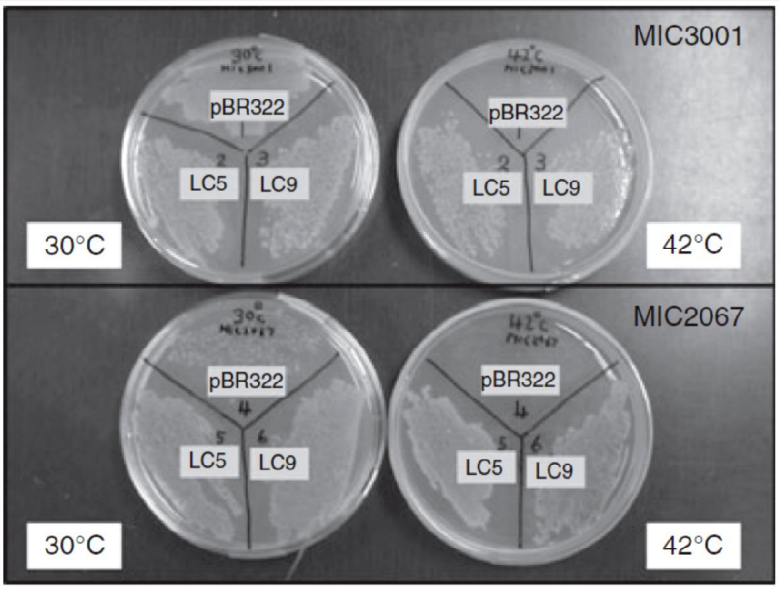

Fig. 2-1. Growth of the *E. coli* MIC3001 and MIC2067 transformants at 30 and 42°C. The *E. coli* MIC3001 and MIC2067 transformants with the pBR322 derivatives harboring the LC5-RNase HI (LC5) and LC9-RNase HI (LC9) genes were grown on a Luria-Bertani plate containing 50 μg/ml ampicillin at 30 and 42°C. The transformants with pBR322 were also grown on this plate as a negative control.

The metagenome library was constructed by ligating the partial *Sau*3AI digests of metagenome extracted from leaf-branch compost into the *Bam*HI site of pBR322 and used to

transform *E. coli* MIC3001. Screening of this metagenome library for the transformants that form colonies on LB-agar plates at 42°C showed that seventeen of the $3x10^5$ transformants could form colonies at this temperature. The plasmids were extracted from these transformants and analyzed for their ability to complement the ts phenotype of *E. coli* MIC2067. As a result, sixteen plasmids were shown to have this ability, indicating that they harbor the functional RNase H genes derived from from metagenomic DNA. The growths of the *E. coli* MIC3001 and MIC2067 transformants with the pBR322 derivatives harbouring two of these functional RNase H genes, termed LC5 and LC9, are shown in Fig. 2-1 as representatives.

2.3.2. Nucleotide sequences of DNA fragments cloned from metagenome library

Determination of the nucleotide sequences of the DNA fragments cloned into sixteen plasmids indicated that eight of them are unique while four of them are duplicated. Accordingly, twelve different DNA fragments, which vary from 619 to 1259 bp in size, were cloned from the metagenome library. All of them contain the genes encoding type 1 RNases H, which vary from 420 to 615 bp in size. These RNases H are termed LC1~LC12-RNases H1 and the plasmids harboring the genes encoding these RNases H1 are termed LC1-LC12. Blast searches of the amino acid sequences of these RNases H1 deduced from their nucleotide sequences indicate that all RNases H1, except for LC6-RNase H1, are novel ones, as summarized in Table 2-1. The amino acid sequence identity between any one of these RNases H1 and that which shows the highest amino acid sequence identity to this RNase H1 varies from 40 to 72%. The amino acid sequence of LC6-RNase H1 is identical with that of *Pseudomonas fluorescens* RNase H1 (accession no. Q3KE7). The nucleotide sequences of LC6-RNase H1 gene and the DNA fragment cloned into plasmid LC6 show 98 and 99% identities to those of the corresponding regions of the *P. fluorescens* genome (accession no.

CP000094), respectively. The nucleotide sequences of the DNA fragments cloned in the plasmids LC1-LC12 are deposited in DDBJ under accession numbers GQ855301-GQ855312, respectively.

Table 2-1. List of RNases H from leaf-branch compost and those with the highest RNase H amino acid sequence identities

RNase H	No. of residues	RNase H with the highest sequence identity		
		Source organism	Accession No.	Identity (%)
LC1-RNase H1	152	*Cytophaga hutchinsonii*	3314281ELC	62
LC2-RNase H1	169	*Syntrophus aciditrophicus*	3310375ENM	72
LC3-RNase H1	149	*Klebsiella pneumoniae*	C4X416	70
LC4-RNase H1	148	*Halorhodospira halophila*	A1WXD7	62
LC5-RNase H1	145	*H. halophila*	A1WXD7	69
LC6-RNase H1	153	*Pseudomonas fluorescens*	Q3KE77	100
LC7-RNase H1	139	*Bordetella pertussis*	Q7VRX8	55
LC8-RNase H1	161	*Bradyrhizobium* sp.	A5EAL2	62
LC9-RNase H1	204	*Gemmatimonas aurantiaca*	C1A5K0	55
LC10-RNase H1	174	*Elusimicrobium minutum*	B2KDX8	40
LC11-RNase H1	140	Uncultured archaeon	Q64AA9	47
LC12-RNase H1	161	*Stigmatella aurantiaca*	Q098P5	44

Localizations of the RNase H1 genes in the DNA fragments cloned in plasmids LC1-LC12 are shown in Fig. 2-2. Potential initiation codons and termination codons of these RNase H1 genes, as well as the nucleotide sequences of their upstream regions (up to 50 bases) are shown in Fig. 2-3. The genes encoding LC3-, LC6-, LC10-, LC11-, and LC12-RNases H1 use the common start codon ATG, while those encoding other seven

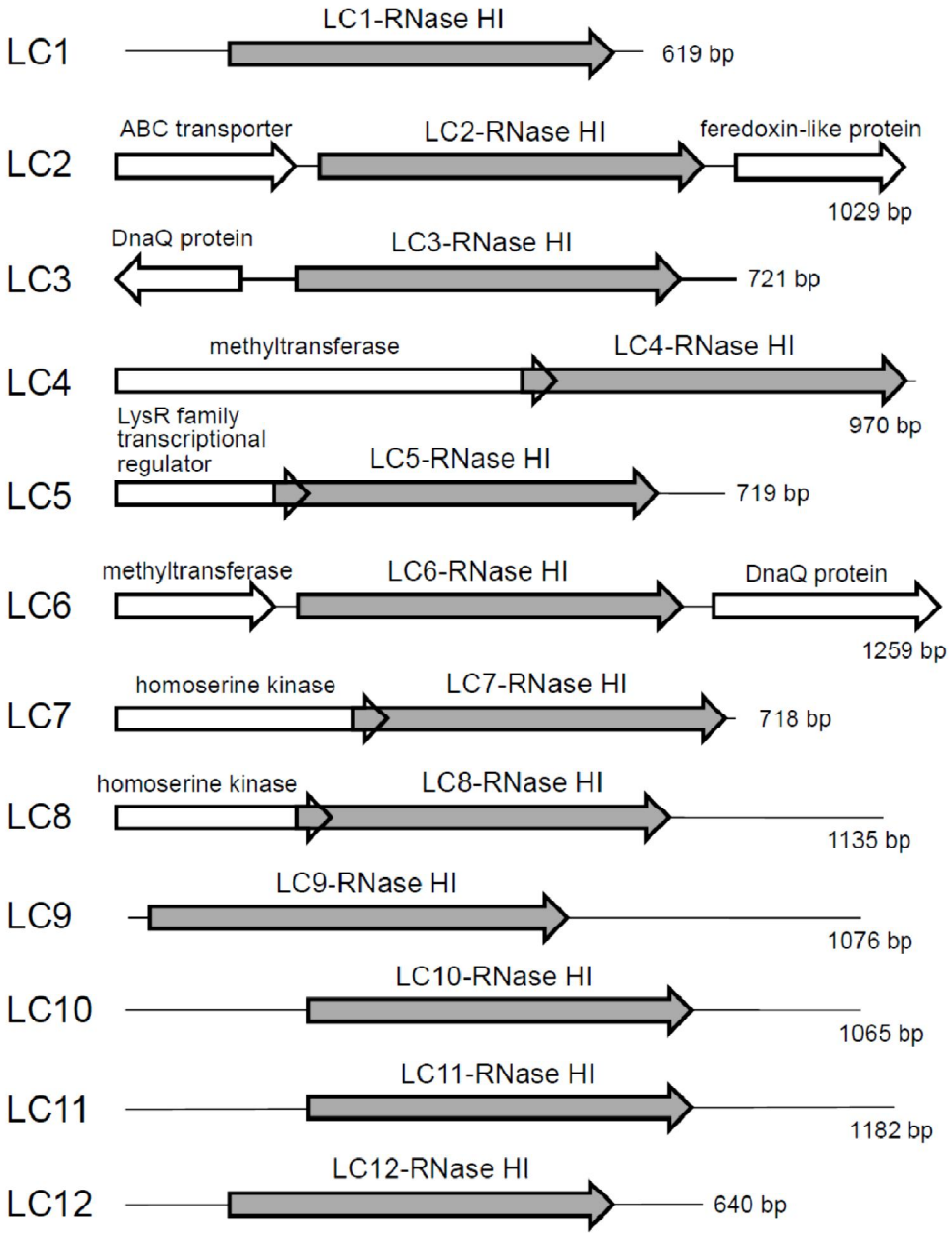

Fig. 2-2. Localization of the RNase H gene. Localizations of the RNase H genes (dark arrows) and other genes (open arrows) in the DNA fragments cloned into plasmids LC1-LC12 are shown. Direction of each arrow represents that of the transcription for each gene. The size of each DNA fragment cloned into the plasmid is shown by base pairs. The protein, which shows the highest amino acid sequence identity to that encoded by each gene, is indicated above the gene.

RNases H1 use the rare start codon GTG or TTG. These codons are selected as potential initiation codons, because none of the DNA fragments cloned in plasmids LC1-LC12 contain

additional ATG codons in the upstream regions of the RNase H1 genes. Potential Shine-Dalgarno (SD) sequence motifs, which are complementary to the 3'-terminal region of 16S rRNA of *E. coli*, are also shown in Fig. 2-3. They are located in the -13 to -3 regions of all RNase H1 genes, except for the LC1-RNase H1, LC6-RNase H1, and LC11-RNase H1 genes. Of various SD motifs, a GGUG motif is more preferred by archaea [93]. For the LC1-RNase H1, LC6-RNase H1, and LC11-RNase H1 genes, leaderless mRNA, which has no SD sequence motif and bind to the ribosome mainly via the codon-anticodon interaction [94], may function in the transcription system.

In addition to the RNase H1 genes, several partial open reading frames were found in the cloned DNA fragments (Fig. 2-2). Homology searches using blast program indicate that the genes located upstream of the RNase H1 genes encode ABC transporter protein in plasmid LC2, DnaQ protein (ε subunit of DNA polymerase III) in plasmid LC3, LysR family transcriptional regulator in plasmid LC5, methyltransferase in plasmids LC4 and LC6, and homoserine kinase in plasmids LC7 and LC8, and those located downstream of the RNase H1 genes encode ferredoxin-like protein in plasmid LC2 and DnaQ protein in plasmid LC6. The directions of the transcriptions of these genes, except for that of the DnaQ gene located upstream of the LC3-RNase H1 gene, are the same as those of the RNase H1 genes.

It is noted that the initiation codons of the genes encoding LC4-, LC5-, LC7-, and LC8-RNases H1 are overlapped with the termination codons of the preceding genes, such that the second and third bases of the initiation codon are shared with the first and second bases of the termination codon (Fig. 2-3). In the resultant tetranucleotide sequence GTGA or TTGA, which is termed stop-start codon, translations of these RNase H1 genes may be directly linked to those of their preceding genes through -1 frame shifting.

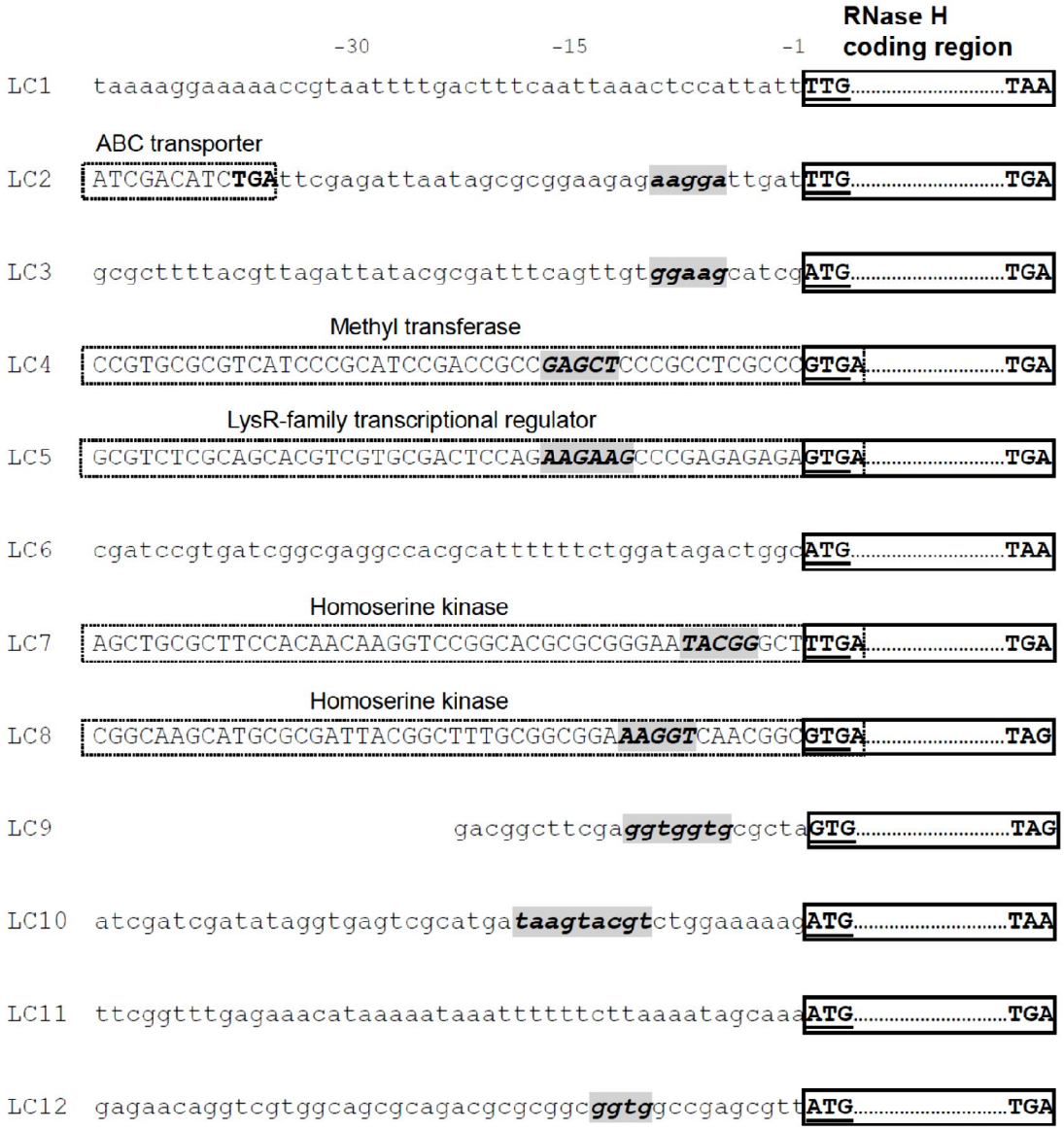

Fig. 2-3. DNA sequences of the 5' upstream regions of the RNase H genes. The RNase H genes and other genes are indicated by open and broken boxes, respectively. The nucleotide sequences for coding and non-coding regions are shown in uppercase and lowercase letters, respectively. The potential start and stop codons are indicated in boldface. The putative SD sequence motifs are indicated by italic and boldface letters.

2.3.3. Amino acid sequences of LC1~LC12-RNases H1

The amino acid sequences of LC1~LC12-RNases H1 are compared with those of

Ec-RNase H1, Sto-RNase H1, Hs-RNase HC, and HIV-RNase H in Fig. 2-4. Ec-RNase H1 represents bacterial type 1 RNases H, Sto-RNase H1 represents Sto-RNase H1 homologs, Hs-RNase H1 represents eukaryotic type 1 RNases H, and HIV-RNase H represents retroviral RNases H [58]. These type 1 RNases H are chosen as representatives, because their structures and functions have been extensively studied [3,36,95]. Their sequences are compared with one another in Fig. 2-4 on the basis of their crystal structures.

LC9-RNase H1 is characterized by the lack of a typical DEDD/E active-site motif. In the LC9-RNase H1 sequence, the second and fourth acidic residues of this motif are replaced by Ala and Asn, respectively. Database searches indicate that the same substitutions are found in the amino acid sequence of RNase H1 from *Gemmatimonas aurantiaca* (Gau-RNase H1) (accession no. C1A5K0), which shows the 55% identity to that of LC9-RNase H1. However, it remains to be determined whether Gau-RNase H1 is enzymatically active or not. The observation that LC9-RNase H1 is functional *in vivo* suggests that Gau-RNase H1 is also functional *in vivo*.

LC10~LC12-RNases H1 are characterized by the lack of a basic protrusion and conserved histidine residue as is Sto-RNase H1. The basic protrusion, which is well conserved in bacterial and eukaryotic type 1 RNases H, provides a DNA binding channel [49] and therefore is important for substrate binding [60]. The histidine residue, which is conserved as His124 for Ec-RNase H1, His264 for Hs-RNase H1, and His539 for HIV-RNase H1, is located in the flexible loop near the active site and is involved in the catalytic function [55]. In the amino acid sequences of LC10~LC12-RNases H1 and Sto-RNase H1, the arginine residue is conserved in the position, where the histidine residue is conserved, suggesting that this arginine residue functions as a substitute of the conserved histidine residue in these RNases H1.

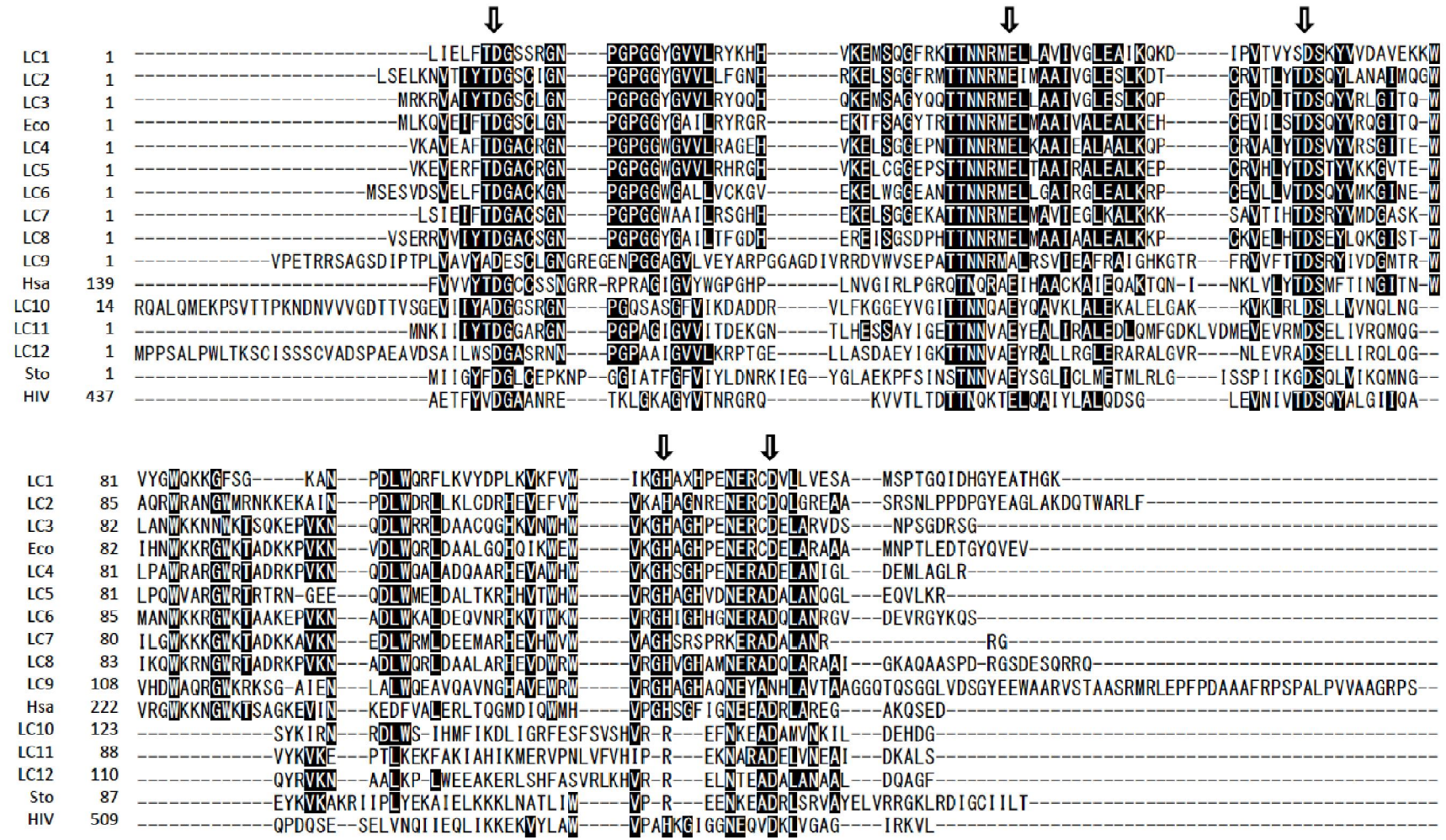

Fig. 2-4. Alignment of the amino acid sequences of type 1 RNases H. The amino acid sequences of LC1~LC12-RNases H1 are compared with those of Ec-RNase H1 (Eco), Hs-RNase H1 (Hsa), Sto-RNase H1 (Sto), and HIV-RNase H. Gaps are denoted by dashes. The amino acid residues, which are conserved in at least seven different proteins, are highlighted in black. The four acidic active-site residues and a conserved histidine or arginine residue involved in catalytic function are indicated by arrows. The numbers represent the positions of the amino acid residues starting from the N terminus of the protein. The accession numbers are P0A7Y4 for Ec-RNase H1, O60930 for Hs-RNase H1, Q973Z8 for Sto-RNase H1, and U53871 for HIV-RNase H.

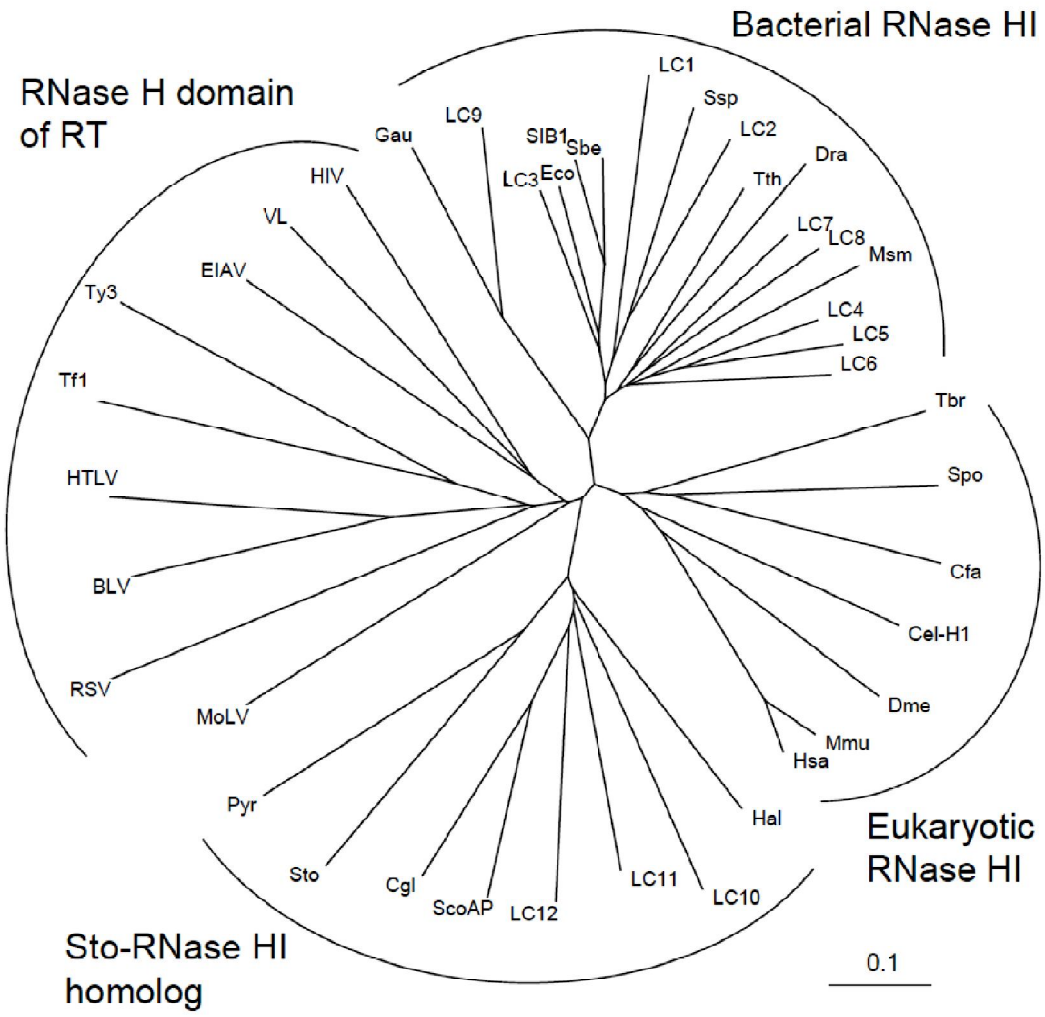

Fig. 2-5. Phylogenetic tree derived from RNase H sequences. Phylogenetic tree derived from the amino acid sequences of 32 RNases H registered in the database and LC1~LC12-RNases H1 were constructed as described under Materials and methods. The distance of the branch from a diverging point is proportional to the frequency of the amino acid substitution at each position in average. The scale bar corresponds to this frequency of 0.1. The abbreviations for the sources of various RNases H1 are as follows (accession numbers in parentheses). Msm, *Mycobacterium smegmatis* (Q07705); Tth, *Thermus thermophilus* (P29253); Dra, *Deinococcus radiodurans* (Q9RVX2); SIB1, *Shewanella sp.* (2805307A); Sbe, *Shewanella benthica* (A9DIN3); Eco, *Escherichia coli* (3505351JM); Ssp, *Synechocystis sp.* (Q55801); Gau, *Gemmatimonas aurantiaca* (C1A5K0); Spo, *Schizosaccharomyces pombe* (Q9UST8); Cfa, *Crithidia fasciculate* (2706382A); Tbr, *Trypanosoma brucei* (O00870) ; Mmu, *Mus musculus* (2502296A); Hsa, *Homo sapiens* (2505262A) ; Dme, *Drosophila melanogaster*, (2404291A); Cel-H1, *Caenorhabditis elegans* (Q86MG7) ; Hal, *Halobacterium salinarium* (Q9HSF6) ; ScoAP, *Streptomyces coelicolor* (Q9L014); Cgl, *Corynebacterium glutamicum* (Q79VE1); BLV, *bovine leukemia virus* (A7KWZ1) ; HTLV,

adult T-cell leukemia virus (0907253C); Sto, *Sulfolobus tokodaii* (3103422A); Pyr, *Pyrobaculum aerophilum* (Q8ZWG9); EIAV, *equine infectious anemia virus* (1202369B); VL, *Visna Lentivirus* (1110153C) ; HIV, *Human immunodeficiency virus* 1 (Q75395); MoLV, *Moloney murine leukemia virus* (0711245A); Tf1, *Tf1 retrotransposon* (Q05654); Ty3, *Ty3 retrotransposon* (1613350B); RSV, *Rous sarcoma virus* (O92956).

2.3.4. Phylogenetic analysis of LC1~LC12-RNase H1 sequences

A phylogenetic tree was constructed based on the amino acid sequences of various type 1 RNases H including LC1~LC12-RNases H1 (Fig. 2-5). LC1~LC9-RNases H1 are classified as bacterial RNases H1 and LC10~LC12-RNases H1 are classified as Sto-RNase H1 homologs. When the amino acid sequences of LC1~LC12-RNases H are individually compared with those of Ec-RNase H1, Sto-RNase H1, Hs-RNase H1, and HIV-RNase H1, LC1-, LC2-, LC3-, LC4-, LC5-, LC6-, LC7-, LC8-, and LC9-RNases HI show the highest amino acid sequence identities of 44, 49, 67, 57, 53, 53, 55, 56, and 37% to Ec-RNase H1, respectively. LC10-, LC11-, and LC12-RNases H1 show the highest amino acid sequence identities of 40, 43, and 41% to Sto-RNase H1, respectively (Fig. 2-5).

2.4. Discussion

2.4.1. Strategy to screen for target genes from metagenome

Two strategies are used to screen for the target genes from metagenome. One is the sequence-based method, in which the target genes are amplified by PCR using primers designed based on the conserved sequences [75,76]. The other is the function-based method. This method needs metagenome library, which is constructed by ligating DNA fragments of metagenome into appropriate vectors. In this method, the production of enzymes is detected either by the change in color, fluorescence, or turbidity of the indicator plates due to enzymatic reaction (*in vitro* method) [73,96-98] or by their ability to complement the unusual growth phenotype of host strain (*in vivo* method) [99-102]. The *in vivo* (genetic) method is superior to the *in vitro* method, because there is no limitation in the library size for screening. In this study, I showed that novel RNase H genes are effectively cloned from metagenome extracted from leaf-branch compost by the genetic method. Universal distribution of RNase H in various organisms and needlessness of chaperone for their folding may make the cloning of these RNase H genes relatively easy. The observation that eight of the twelve RNase H1 genes encode novel bacterial RNases H1 and three of them encode novel Sto-RNase H1 homologs suggests that leaf-branch compost contain a variety of bacteria and archaea.

2.4.2. LC9-RNase H1 lacking a typical active-site motif

LC9-RNase H1 is the first natural enzyme, which lacks a typical active-site motif, but exhibits enzymatic activity. In this enzyme, the second and fourth residues of the DEDD/E active-site motif are replaced by Ala and Asn, respectively. This result indicates that a DEDD/E active-site motif is not fully conserved in RNases H. The second glutamate residue coordinates with metal ion B, which is required for promotion of the catalytic reaction by reducing the energy barrier between the substrate and product, while the fourth acidic residue

coordinates with metal ion A, which is required for substrate-assisted nucleophile formation and product release [48,54]. Replacement of the fourth aspartate residue with Asn may not seriously affect the coordination of metal ion A and therefore may not seriously affect the enzymatic activity, because the corresponding residue of Bh-RNase H1 (Asp192) coordinates with metal ion A regardless of whether it is mutated to Asn [48]. However, it is unlikely that replacement of the second glutamate residue with Ala does not seriously affect the coordination of metal ion B. LC9-RNase H1 may therefore adopt alternative catalytic mechanism, in which only single metal ion is required for activity. Further biochemical and structural studies will be required to understand the catalytic mechanism of LC9-RNase H1.

2.4.3. LC10~LC12-RNases H1

LC10~LC12-RNases H1 are members of the Sto-RNase H homolog family. Archaeal RNases H1, Halo-RNase H1 [59] and Sto-RNase H1 [58], which represent this family, cleave RNA-DNA/DNA substrates not only at the phosphodiester bonds between ribonucleotides, but also at an (5')RNA-DNA(3') junction, which is not cleaved by the members of other RNase H families. Therefore, it would be informative to examine whether LC10~LC12-RNases H1 cleave an RNA-DNA junction.

Comparison of the amino acid sequences of LC10~LC12-RNases H1 with that of Sto-RNase H1 indicates that LC10- and LC12-RNases H1 have an N-terminal extension of around 30 residues, while LC11-RNase H1 does not have it (Fig. 2-4). Halo-RNase H1 also has a 70-residue extension at the N-terminus with unknown function, as compared to Sto-RNase H1. In addition, LC10~LC12-RNases H1 lack a short C-terminal extension. It has been reported that Sto-RNase H1 is a highly thermostable enzyme with the T_m value of 102°C and is stabilized by anchoring its C-terminal tail (residues 143-149) to main body through hydrophobic interactions, a disulfide bond, and hydrogen bonds [56]. Deletion of this tail greatly destabilizes the protein by 24°C in T_m. These results suggest that

LC10~LC12-RNases H1 are much less stable than Sto-RNase H1, because they do not have this tail. The temperature of the compost, from which these genes are cloned, is around 50°C. Therefore, it is likely that the source organisms of these genes are moderate thermophiles, while that of Sto-RNase H1 is a hyperthermophile. Identification of Sto-RNase H1 homologs lacking the C-terminal tail in nature may suggest that Sto-RNase H1 adapts to high temperatures by attaching the C-terminal tail. It has also been reported that the mutation of the arginine residue conserved in Sto-RNase H1 homologs at the position, in which the histidine residue is conserved in bacterial RNases H1, eukaryotic RNases H1, and retrovial RNases H, considerably decreases the enzymatic activity of Sto-RNase H1 [56]. This arginine residue may be involved in catalytic function by facilitating the product release by dislodging the cleaved 5'-phosphate, as proposed for the conserved histidine residue [49].

2.4.4. Rare start codons UUG and GUG

AUG and alternative codons GUG, UUG AUU and AUA are known to be selected as the start codon by 30S ribosomal subunit [103]. Based on the complete genome sequence of *Escherichia coli* K-12, it is predicted that 14% of the *E. coli* genes use GUG as the start codon, and another 3% of the genes use UUG [104]. Laursen *et al.* (2005) reported that the initiation codons in *E. coli* occur at a frequency of 90, 8, and 1% for AUG, GUG, and UUG, respectively [94]. The common start codon AUG constitutes 70-90% of all start codons predicted for most archaeal genes as well. However, three start codons AUG, GUG, and UUG are exceptionally used with equal frequencies in some hyperthermophiles [93]. It has been reported that the start codon GUG is used for the genes encoding type 1 RNases H from *Thermus thermophilus* HB8 [88] and *Mycobacterium smegmatis* [105]. The observation that seven of the twelve RNase H genes cloned in this study use a start codon GUG or UUG might suggest rather high utilization of rare start codons in the uncultured microbes from the leaf-branch compost.

2.4.5. Stop-start codons UUGA and GUGA

Overlapping stop-start codons are widely distributed. The first type of the stop-start codon is UAAUG (UAA + AUG) [106]. The second type is AUGA (AUG + UGA) and GUGA (GUG + UGA) [107,108]. These overlapping stop-start codons are often observed in the bacterial polycistronic genes [103]. It has been reported that AUGA is involved in the effective translational coupling, in which the second open reading frame is translated by the same ribosome worked for the first open reading frame [107]. Two potential stop-start codons UUGA and GUGA found in the RNase H genes from the leaf-branch compost may also be involved in the translational coupling.

2.5. Summary

Metagenome library was prepared from leaf-branch compost and used to screen for the RNase H genes by their abilities to complement the temperature-sensitive growth phenotype of the *rnhA* mutant *E. coli* strain MIC3001. Of the 3 x 10^5 clones, seventeen clones could grow at non-permissive temperature 42°C. Isolation of the plasmids from these clones and determination of the nucleotide sequences of the cloned DNA fragments allowed us to identify twelve different genes encoding type 1 RNases H (LC1~LC12 RNases H1). Eleven of them encode novel RNases H, which show the 40-72% amino acid sequence identities to those available from database. Seven of them apparently utilize rare start codons GTG and TTG. Of them, three GTG and one TTG codons overlap with the termination codon TGA of a preceding open reading frame in the tetranucleotide sequences GTGA and TTGA, respectively. Phylogenetic analysis indicates that LC1~LC9-RNases H1 are classified as bacterial RNases H1 and LC10~LC12-RNases H1 are classified as Sto-RNase H1 homologs. LC9-RNase H1 lacks a typical DEDD/E active-site motif, which is well conserved in various RNases H. These results indicate that direct screening of environmental DNA without cultivation of microbes is a useful procedure to isolate novel RNase H genes.

CHAPTER 3

Activity, stability, and structure of metagenome-derived LC11-RNase H1, a homolog of *Sulfolobus tokodaii* RNase H1

3.1. Introduction

LC11-RNase H1 is one of the twelve RNases H1 enzymes (LC1~LC12-RNases H1) isolated from leaf-branch compost by a metagenomic approach. It is a Sto-RNase H1 homolog, which is phylogenetically different from bacterial RNases H1, eukaryotic RNases H1, and retroviral RNases H [58]. LC11-RNase H1 is the smallest RNase H enzyme with 140 amino acid residues (molecular mass of 15.6 kDa) so far examined and shows the amino acid sequence identity of 34% to Sto-RNase H1. It lacks a C-terminal tail, which is not important for enzymatic function but is responsible for hyperstabilization of Sto-RNase H1 [56]. Sto-RNase H1 is unique in that it can cleave not only the RNA strand of an RNA/DNA hybrid but also the RNA strand of a dsRNA [51]. None of other RNases H, except for some retroviral RNases H, cleave dsRNA. Therefore, it would be informative to examine whether LC11-RNase H1 exhibits dsRNA-dependent RNase (dsRNase) activity.

In this chapter, I overproduced LC11-RNase H1 in *E. coli*, purified, and compared its activity, stability, and structure with those of Sto-RNase H1. Unlike Sto-RNase H1, LC11-RNase H1 did not exhibit dsRNase activity. LC11-RNase H1 was less stable than Sto-RNase H1 by 37°C in T_m. I discuss the structural features for these differences.

3.2. Materials and methods

3.2.1. Cells and plasmids

The *E. coli* strain MIC2067(DE3) [56], plasmid pBR322-LC11 containing the gene encoding LC11-RNase H1, and plasmids pET-Sto [109] and pET-StoΔC6 [56] for overproduction of Sto-RNase H1 and Sto-RNase H1ΔC6, respectively, were previously constructed in our laboratory. Plasmid pET25b was obtained from Novagen.

3.2.2. Plasmid construction, overproduction and purification

Plasmid pET-LC11 for overproduction of LC11-RNase H1 was constructed by amplifying the gene encoding LC11-RNase H1 by PCR and ligating it into the *Nde*I-*Bam*HI sites of pET25b. Plasmid pBR322-LC11 was used as a template. All DNA oligomers for PCR were synthesized by Hokkaido System Science. PCR was performed with a GeneAmp PCR system 2400 (Applied Biosystems). The DNA sequences were confirmed by a Prism 310 DNA sequencer (Applied Biosystems).

For overproduction of LC11-RNase H1, the *E. coli* MIC3001(DE3) transformants with the pET25b derivatives were grown at 30°C. When the absorbance at 600 nm reached approximately 0.5, 1 mM isopropyl thio-β-D-galactoside (IPTG) was added to the culture medium and cultivation was continued at 30°C for an additional 30 min. Then, the temperature of the growth medium was shifted to 25°C and cultivation was continued at 25°C for an additional 16 h. The subsequent purification procedures were carried out at 4°C. Cells were harvested by centrifugation at 8000 *g* for 10 min, suspended in 10 mM Tris/HCl (pH 8.0) containing 1 mM EDTA and 1 mM dithiothreitol (DTT) (buffer A), disrupted by sonication lysis, and centrifuged at 30,000 *g* for 30 min. The supernatant was collected, dialyzed against 10 mM sodium phosphate (pH 7.0), and applied to a column (5 mL) of DE52 (GE Healthcare) equilibrated with the same buffer. The protein was eluted from the column with a linear gradient of NaCl from 0 to 1 M. The fractions containing the protein were

collected, dialyzed against 20 mM Tris-HCl (pH 9.0), and loaded onto a HiTrap Q column (GE Healthcare) equilibrated with the same buffer. The protein was eluted from the column with a linear gradient of NaCl from 0 to 1 M. The fractions containing the protein were collected, dialyzed against 20 mM Tris-HCl (pH 7.5) containing 50 mM NaCl, and loaded onto the Hiload 16/600 Superdex 200pg column (GE Healthcare) equilibrated with the same buffer. The fractions containing the protein were collected and dialyzed against 10 mM Tris-HCl (pH 7.5).

Overproduction of Sto-RNase H1 and Sto-RNase H1ΔC6 using *E. coli* MIC2067(DE3) transformants with pET-Sto and its derivatives and purification of these proteins were performed as described previously [109].

The purity of the protein was analyzed by SDS/PAGE on a 15% polyacrylamide gel [110], followed by staining with Coomassie Brilliant Blue. The protein concentration was determined from UV absorption using an A_{280} value of 0.30 for LC11-RNase H1, 0.97 for Sto-RNase H1, and 1.0 for Sto-RNase H1ΔC6 for a 0.1% (1.0 $mg \cdot ml^{-1}$) solution. These values were calculated by using ε of 1576 $M^{-1} \cdot cm^{-1}$ for Tyr and 5225 $M^{-1} \cdot cm^{-1}$ for Trp at 280 nm [111].

3.2.3. Enzymatic activity

The RNase H activity was determined by measuring the amount of the acid-soluble digestion product from the substrate, ^{3}H-labeled M13 DNA/RNA hybrid, accumulated upon incubation at 37°C for 15 min, as described previously [50]. The buffer used for assay was 10 mM Tris-HCl (pH 8.5) containing 10 mM $MgCl_2$, 10 mM NaCl, 1 mM 2-mercaptoethanol, and 50 $\mu g \cdot ml^{-1}$ bovine serum albumin. One unit was defined as the amount of enzyme producing 1 μmol of acid-soluble material per min at 37°C. The specific activity was defined as the enzymatic activity per milligram of protein.

For cleavage of the oligomeric substrates, 5'-fluorescein-labeled 12 bp RNA/DNA

hybrid (R12/D12), 12 bp RNA/RNA duplex (R12/R12), 29 bp DNA_{15}-RNA_1-DNA_{13}/DNA duplex (D15-R1-D13/D29), 29 b RNA and 29 b DNA, and 3'-fluorescein-labeled 18 bp RNA_9-DNA_9/DNA duplex (R9-D9/D18) were prepared as described previously [50]. Hydrolysis of the substrate at 37°C for 15 min and separation of the products on a 20% polyacrylamide gel containing 7 M urea were carried out as described previously [50]. The products were detected by Typhoon 9240 Imager (GE Healthcare). The reaction buffers were the same as those for the hydrolysis of the M13 DNA/RNA hybrid. The products were identified by comparing their migration on the gel.

3.2.4. DSC

The DSC measurement was performed on a high sensitivity VP-DSC (Microcal Inc.) and controlled by the VPVIEWERTM software package (Microcal Inc.) at scan rate of 1°C·min^{-1}. The proteins were dissolved in 20 mM Gly-HCl (pH 3.0) at a concentration of approximately 0.5 mg·ml^{-1}. Before performing the measurement, the proteins were degassed in a vacuum for 5 min.

3.2.5. Crystallization

The crystallization conditions were initially screened using crystallization kits from Hampton Research (Crystal Screens I and II) and Emerald Biostructures (Wizard I and II). The conditions were surveyed using sitting-drop vapour-diffusion method at 4°C. Drops were prepared by mixing 1 µl of protein solution at the concentration of 7 mg·ml^{-1} and 1 µl of reservoir solution and were vapour-equilibrated against 100 µl reservoir solution. Native LC11-RNase H1 crystals appeared after three weeks using Wizard I solution No. 20 (0.1 M imidazole pH 8, 0.2 M NaCl, 0.4 M NaH_2PO_4, 1.6 M K_2HPO_4). The crystallization conditions were further optimized by varying the volume and concentration of the protein; and a few crystals appeared when the drop was prepared by mixing 1.5 µl protein solution at

the concentration of 10 mg·ml^{-1} and 1.5 µl of the reservoir solution, after 2-3 weeks. Despite the same crystallization conditions, only three crystals were obtained from 48 crystallization wells.

3.2.6. X-ray diffraction data collection and structure determination

Table 3-1. Data collection and refinement statistics

Data collection	
Wavelength (Å)	0.9
Space group	$P2_1$
Unit cell	
a, b, c (Å)	54.3, 104.9, 57.2
α, β, γ (°)	90.0, 110.3, 90.0
Resolution (Å)	50.00 – 1.40 (1.42 – 1.40)
Observations	1394496
Unique reflections	115103
Completeness (%)	97.4 (85.8)[a]
R_{merge} (%)[b]	6.6 (45.1)[a]
Average I/s (I)	31.7 (2.8)[a]
Refinement statistics	
Resolution limits (Å)	47.76 – 1.40
No. of molecules	4
No. of protein atoms	4434
No. of water atoms	904
R_{work} (%)	17.2
R_{free} (%)[c]	19.2
Deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.061
Mean B factors (Å^2)	
Protein atoms	13.2
Water atoms	24.6

[a] Values in parentheses are for the highest resolution shell. [b] $R_{merge} = \sum | I_{hkl} - \langle I_{hkl} \rangle | / \sum I_{hkl}$, where I_{hkl} is the intensity measurement for reflection with indices *hkl* and $\langle I_{hkl} \rangle$ is the mean intensity for multiply recorded reflections. [c] R_{free} was calculated using 5 % of the total reflections chosen randomly and omitted from refinement.

X-ray diffraction data set of LC11-RNase H1 was collected at -173°C using synchrotron radiation on the BL44XU station at SPring-8 (Hyogo, Japan). All data sets were indexed, integrated, and scaled using the HKL2000 program suite [112]. The structure was solved by the molecular replacement method using MOLREP [113] in the CCP4 program suite. The crystal structure of RNase H1 from *Chlorobium tepidum* (Protein Data Bank entry 3H08) was used as a starting model. The statistics for data collection and refinement are summarized in Table 3-1. The figures were prepared by PyMol (http://pymol.sourceforge.net/).

3.2.7. Accession number

The coordinates and structure factors for LC11-RNase H1 have been deposited in the Protein Data Bank under accession codes 3U3G.

3.3. Results and discussion

3.3.1. Overproduction and purification

I used the *rnhA*/*rnhB* double mutant strain *E. coli* MIC2067(DE3) as a host strain for overproduction of LC11-RNase H1 to avoid contamination of host-derived RNases H (RNases H1 and H2). Upon induction for overproduction, the recombinant protein accumulated in the cells in a soluble form. It was purified to give a single band on SDS-PAGE (Fig. 3-1). The amount of the protein purified from 1 L of culture was approximately 10 mg. The molecular mass of LC11-RNase H1 was estimated to be 16 kDa by SDS-PAGE and 18 kDa by gel filtration column chromatoghraphy. Both values are comparable to that calculated from the amino acid sequence (15.6 kDa), indicating that LC11-RNase H1 exists in a monomeric form, like other RNases H1.

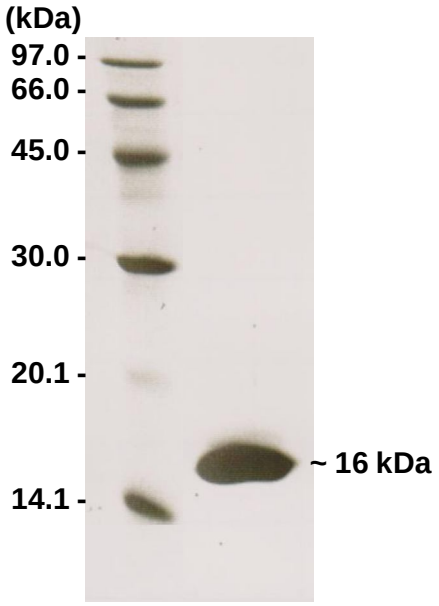

Figure. 3-1. SDS/PAGE analyses for the purified protein of LC11-RNase H1. The lane on the left is low-molecular-weight marker kit (Pharmacia Biotech); the other lane is LC11-RNase H1 protein.

3.3.2. Enzymatic activity

The dependencies of the LC11-RNase H1 activity on pH, salt, and metal ion were analyzed at 37°C by changing one of the conditions used for assay [10 mM Tris-HCl (pH 8.5), 10 mM NaCl, and 10 mM $MgCl_2$]. The M13 DNA/RNA was used as a substrate. The enzymatic activity was determined at this temperature (37°C), which would be lower than the optimum one, because all the substrates used in this study are not fully stable at ≥50°C.

When the enzymatic activity was determined over the pH range of 4-11, LC11-RNase H1 exhibited the highest activity at pH 10 (Fig. 3-2). However, the RNase H activity was determined at pH 8.5, at which LC11-RNase H1 exhibits approximately 80% of the maximal enzymatic activity, because the stability of the substrate and the solubility of the metal cofactor decrease as the pH increases beyond 9.0.

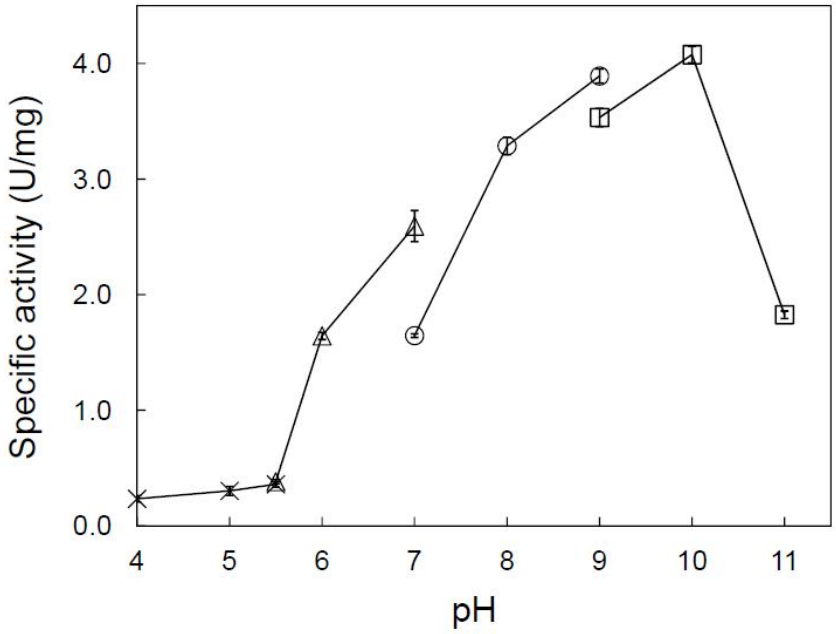

Figure 3-2. pH dependency of LC11-RNase H1. The enzymatic activity of LC11-RNase H1 was determined in the presence of 10 mM $MgCl_2$ at 37°C over the pH range of 4-11 using M13 DNA/RNA hybrid as a substrate. The buffers used for assay were 10 mM sodium acetate (H 4.0-5.5) (cross), 10 mM 2-(*N*-morpholino)ethanesulfonic acid (MES)-NaOH (pH 5.5-7.0) (open triangle), 10 mM Tris-HCl (pH 7.0-9.0) (open circle), and *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS)-NaOH (pH 9.0-11.0) (open square). Experiments were carried out at least twice and the average values are shown together with the errors.

When the enzymatic activity was determined at various concentrations of NaCl, LC11-RNase H1 exhibited the highest activity in the presence of 10 mM NaCl (Fig. 3-3). Its activity gradually decreased as the salt concentration increased. As a result, LC11-RNase H1 exhibited 50 and 10% of the maximal activity in the presence of 0.4 and 1 M NaCl,

respectively. A similar result was obtained when KCl was used as a salt (Fig. 3-3).

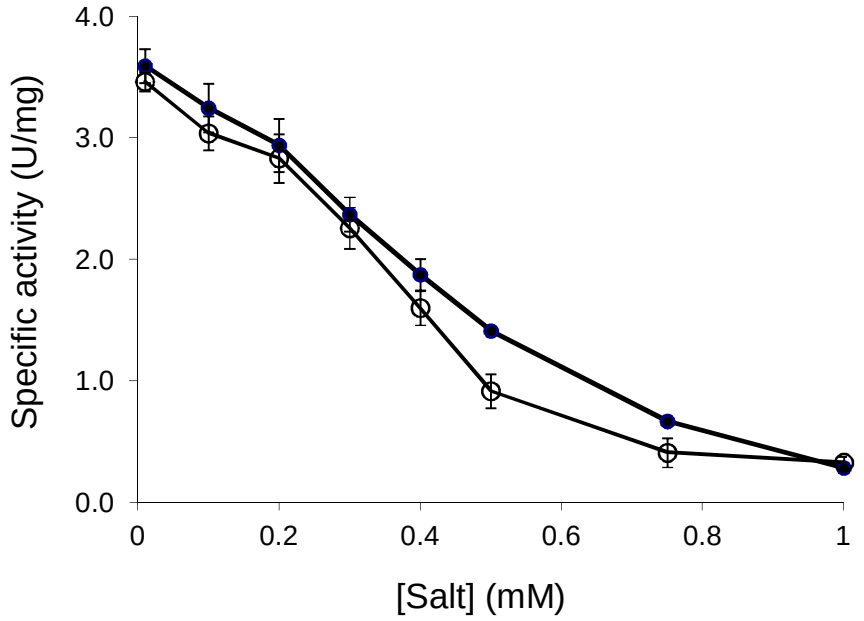

Figure 3-3. Salt dependency of LC11-RNase H1. The enzymatic activity of LC11-RNase H1 were determined at 37°C in 10 mM Tris-HCl (pH 8.5) containing NaCl (open circle) and KCl (closed circle) using M13 DNA/RNA hybrid as a substrate. Experiments were carried out at least twice and the average values are shown together with the errors.

When the enzymatic activity was determined in the presence of various concentrations of $MgCl_2$, LC11-RNase H1 exhibited the highest activity at 10 mM $MgCl_2$ (Fig. 3-4). It also exhibited activity in the presence of $MnCl_2$, $CoCl_2$, $NiCl_2$, $ZnCl_2$, and $CuCl_2$ with the maximum at 10, 0.1, 10, 1, and 1 mM, respectively, but did not exhibit activity in the presence of $CaCl_2$ (Fig. 3-4). The maximal Mn^{2+}-dependent activity was slightly lower than but comparable to the maximal Mg^{2+}-dependent activity, while the maximal Co^{2+}-, Ni^{2+}-, Zn^{2+}-, and Cu^{2+}-dependent activities were lower than the maximal Mg^{2+}-dependent activity by 8-15 fold. The specific activities of LC11-RNase H1 and Sto-RNase H1 were determined to be 3.5 and 2.2 units/mg, respectively, in the presence of 10 mM $MgCl_2$, and 3.4 and 0.8 units/mg, respectively, in the presence of 10 mM $MnCl_2$. It has been reported that Sto-RNase H1 exhibits activity in the presence of Co^{2+} and Ni^{2+} as well, but does not exhibit activity in

the presence of Cu^{2+}, Zn^{2+}, and Ca^{2+} [58]. The maximal Co^{2+}-and Ni^{2+}-dependent activities of Sto-RNase H1 are lower than its maximal Mg^{2+}-dependent activity by 10-20 fold. Thus, LC11-RNase H1 shows a similar metal ion preference for activity to that of Sto-RNase H1, although its preference to Mg^{2+} ion is weaker than that of Sto-RNase H1.

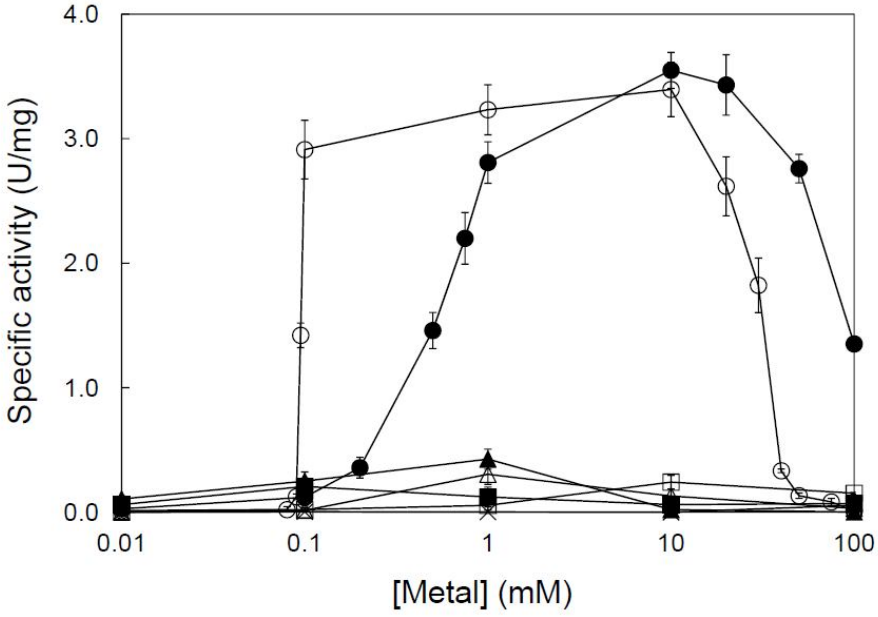

Figure 3-4. Metal ion dependency of LC11-RNase H1. The enzymatic activity of LC11-RNase H1 were determined at 37°C in 10 mM Tris-HCl (pH 8.5) containing various concentrations of $MgCl_2$ (closed circle), $MnCl_2$ (open circle), $CuCl_2$ (closed triangle), $ZnCl_2$ (open triangle), $CoCl_2$ (closed square), $NiCl_2$ (open square), and $CaCl_2$ (cross) using M13 DNA/RNA hybrid as a substrate. Experiments were carried out at least twice and the average values are shown together with the errors.

3.3.3. Substrate and cleavage-site specificities

The substrate and cleavage-site specificities of LC11-RNase H1 were analyzed by using 12 bp RNA/DNA hybrid (R12/D12), 12 bp RNA/RNA duplex (R12/R12), 29 b RNA, 29 b DNA, 29 bp DNA_{15}-RNA_1-DNA_{13}/DNA duplex (D15-R1-D13/D29), and 18 bp RNA_9-DNA_9/DNA duplex (R9-D9/D18) as a substrate. For comparative purposes, these substrates were cleaved by Sto-RNase H1 as well. D15-R1-D13 is the chimeric oligonucleotide, in which single ribonucleotide is flanked by 15 and 13 b DNAs at the 5' and

3’ sides, respectively. R9-D9/D18 is an Okazaki fragment-like substrate, in which the 18 b RNA_9-DNA_9 chimeric oligonucleotide is hybridized to the 18 b complementary DNA.

Cleavage of the R12/D12 substrate with LC11-RNase H1 and Sto-RNase H1 is summarized in Figure 3-5. Both proteins cleaved this substrate at multiple sites, but with different cleavage-site specificities. LC11-RNase H1 cleaved this substrate most preferably at a6-u7 and a9-c10 in the presence of 10 mM $MgCl_2$ and nearly equally at all sites between g5 and g12 in the presence of 10 mM $MnCl_2$, whereas Sto-RNase H1 cleaved this substrate at all sites between a4 and g12, most preferably at g5-a6, in the presence of 10 mM $MgCl_2$ and most preferably at a4-g5 and less preferably at g3-a4, u7-g8, c10-g11, and g11-g12 in the presence of 10 mM $MnCl_2$.

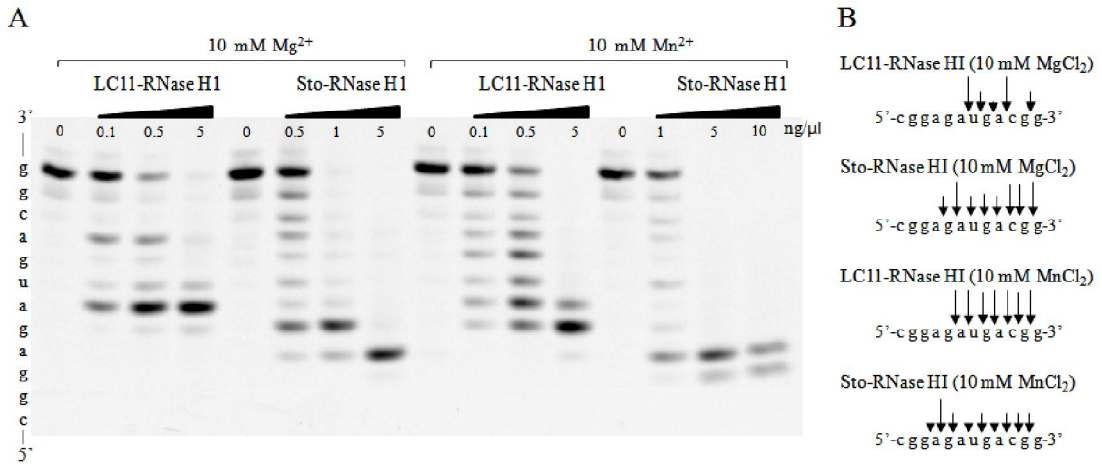

Figure 3-5. Cleavage of the 12 bp RNA/DNA substrate with LC11-RNase H1 and Sto-RNase H1. **(A)** Separation of the hydrolysates by urea gel. The 5’-end labeled 12 bp RNA/DNA substrate was hydrolyzed by the enzyme at 37°C for 15 min and the hydrolysates were separated on a 20% polyacrylamide gel containing 7 M urea. The concentration of the substrate was 1.0 μM. The concentration of the enzyme in the reaction mixture is indicated above each lane. The sequence of the RNA strand of the substrate is indicated along the gel. **(B)** Schematic representation of the sites and extents of cleavage by LC11-RNase H1 and Sto-RNase H1. Cleavage sites are shown by arrows. The differences in the lengths of the arrows reflect relative cleavage intensities at the position indicated. These lengths do not necessarily reflect the amount of the products accumulated upon complete hydrolysis of the substrate.

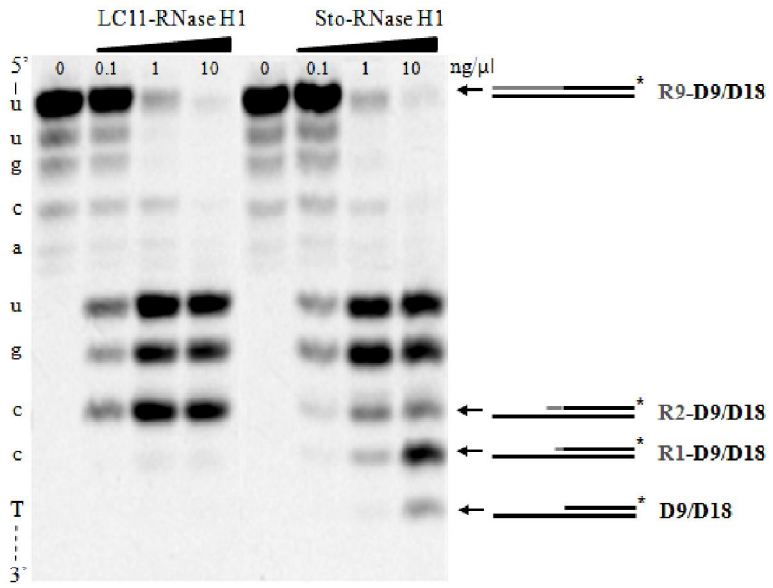

Figure 3-6. Cleavage of the 18 bp RNA-DNA/DNA substrate with LC11-RNase H1 and Sto-RNase H1. The 3'-end labeled 18 bp RNA_9-DNA_9/DNA substrate was hydrolyzed by the enzyme in the presence of 10 mM $MgCl_2$ and the hydrolysates were separated by urea gel as in Fig. 3-5A. The partial sequence of the RNA_9-DNA_9 strand of the substrate is indicated along the gel, in which deoxyribonucleotides are indicated by capital letters and ribonucleotides are indicated by lowercase letters. The substrate and the products containing no, one, and two ribonucleotides at the 5'-termini are schematically shown on the right side of the gel. The asterisk indicates the fluorescence-labeled site.

Cleavage of the R9-D9/D18 substrate with LC11-RNase H1 and Sto-RNase H1 in the presence of 10 mM $MgCl_2$ is summarized in Figure 3-6. Both proteins cleaved this substrate at multiple sites, but with different cleavage-site specificities. LC11-RNase H1 preferably cleaved this substrate at three sites (a5-u6, u6-g7, and g7-c8) with similar efficiencies, whereas Sto-RNase H1 preferably cleaved this substrate at five sites with different efficiencies (most preferably at a5-u6 and u6-g7 and less preferably at g7-c8, c8-c9, and c9-T10). This result is consistent with the previous one that Sto-RNase H1 can cleave the (5')RNA-DNA(3') junction of the Okazaki fragment-like substrate [58]. The cleavage-site specificities of both proteins for this substrate were not seriously changed in the presence of 10 mM $MnCl_2$ (data not shown). These results indicate that LC11-RNase H1 cannot cleave

the (5')RNA-DNA(3') junction. It is noted that Sto-RNase H1 can cleave the (5')RNA-DNA(3') junction, but very poorly even after extensive hydrolysis, suggesting that its RNase H activity is not sufficient for complete removal of RNA from Okazaki fragment *in vivo*.

Cleavage of the R12/R12 substrate with LC11-RNase H1 and Sto-RNase H1 in the presence of 10 mM $MnCl_2$ is summarized in Figure 3-7. Sto-RNase H1 cleaved this substrate at all sites between g5 and g12, most preferably at a9-c10, as reported previously [58], whereas LC11-RNase H1 did not cleave this substrate. Neither LC11-RNase H1 nor Sto-RNase H1 cleaved this substrate in the presence of 10 mM $MgCl_2$ (data not shown). These results indicate that, unlike Sto-RNase H1, LC11-RNase H1 does not have an ability to cleave double-stranded (ds) RNA.

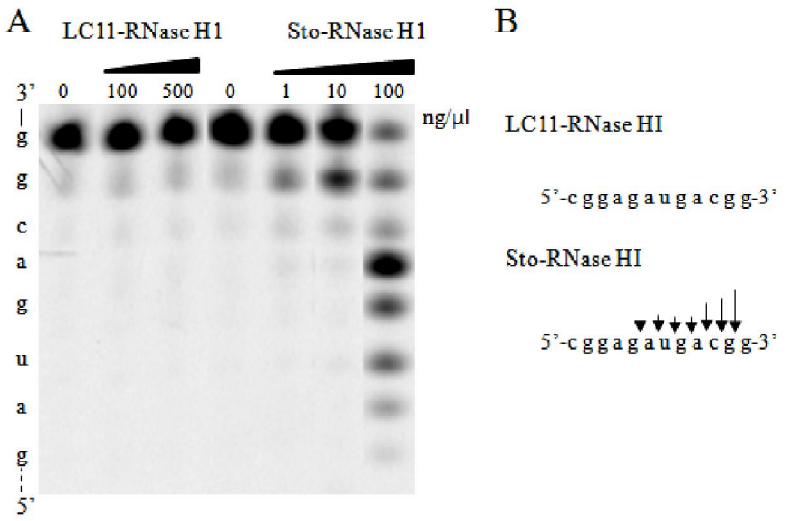

Figure 3-7. Cleavage of the 12 bp RNA/RNA substrate with LC11-RNase H1 and Sto-RNase H1. **(A)** The 5'-end labeled 12 bp RNA/RNA substrate was hydrolyzed by the enzyme in the presence of 10 mM $MnCl_2$ and the hydrolysates were separated by urea gel as in Fig. 3-5A. **(B)** The sites and extents of cleavage by LC11-RNase H1 and Sto-RNase H1 are schematically shown as in Fig. 3-4B.

Both LC11-RNase H1 and Sto-RNase H1 did not cleave D15-R1-D13/D29, 29 b RNA, and 29 b DNA either in the presence of 10 mM $MgCl_2$ or $MnCl_2$, indicating that both proteins

do not have an ability to cleave dsDNA, single ribonucleotide-embedded dsDNA, single-stranded (SS) RNA, and ssDNA.

3.3.4. Thermal stability

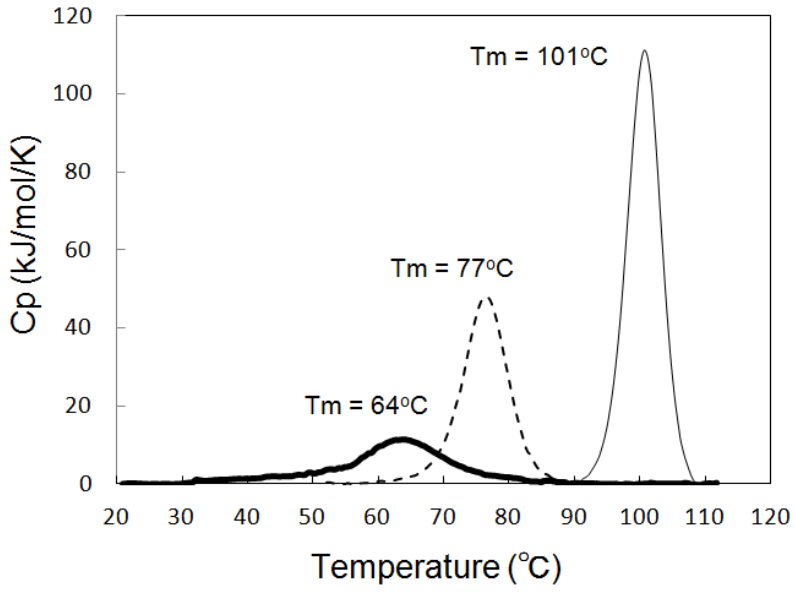

Figure 3-8. DSC curves. The DSC curves of LC11-RNase H1 (thick solid line), Sto-RNase H1□C6 (dashed line), and Sto-RNase H1 (thin solid line) were measured at a scan rate of 1°C /min. These proteins were dissolved in 20 mM Gly-HCl (pH 3.0) at approximately 0.5 mg/ml.

Heat-induced unfolding of LC11-RNase H1, Sto-RNase H1, and the Sto-RNase H1 derivative lacking the C-terminal six-residue tail (Sto-RNase H1ΔC6) was analyzed by differential scanning calorimetry (DSC) at pH 3.0. All DSC curves were reproduced by repeating thermal scans, indicating that thermal unfolding of these proteins is reversible. The denaturation curves of all proteins showed single transitions (Fig. 3-8). The peak of this curve represents the T_m of the protein, which was 64°C for LC11 RNase H1, 101°C for Sto-RNase H1, and 77°C for Sto-RNase H1ΔC6. The T_m values of Sto-RNase H1 and Sto-RNase H1ΔC6 are identical with those previously reported [56]. Thus, LC11-RNase H1 is less stable than Sto-RNase H1 and Sto-RNase H1ΔC6 by 37°C and 13°C in T_m, respectively. These

results indicate that destabilization by the removal of the C-terminal tail accounts for two thirds of the difference in T_m between LC11-RNase H1 and Sto-RNase H1.

3.3.5. Crystal structure

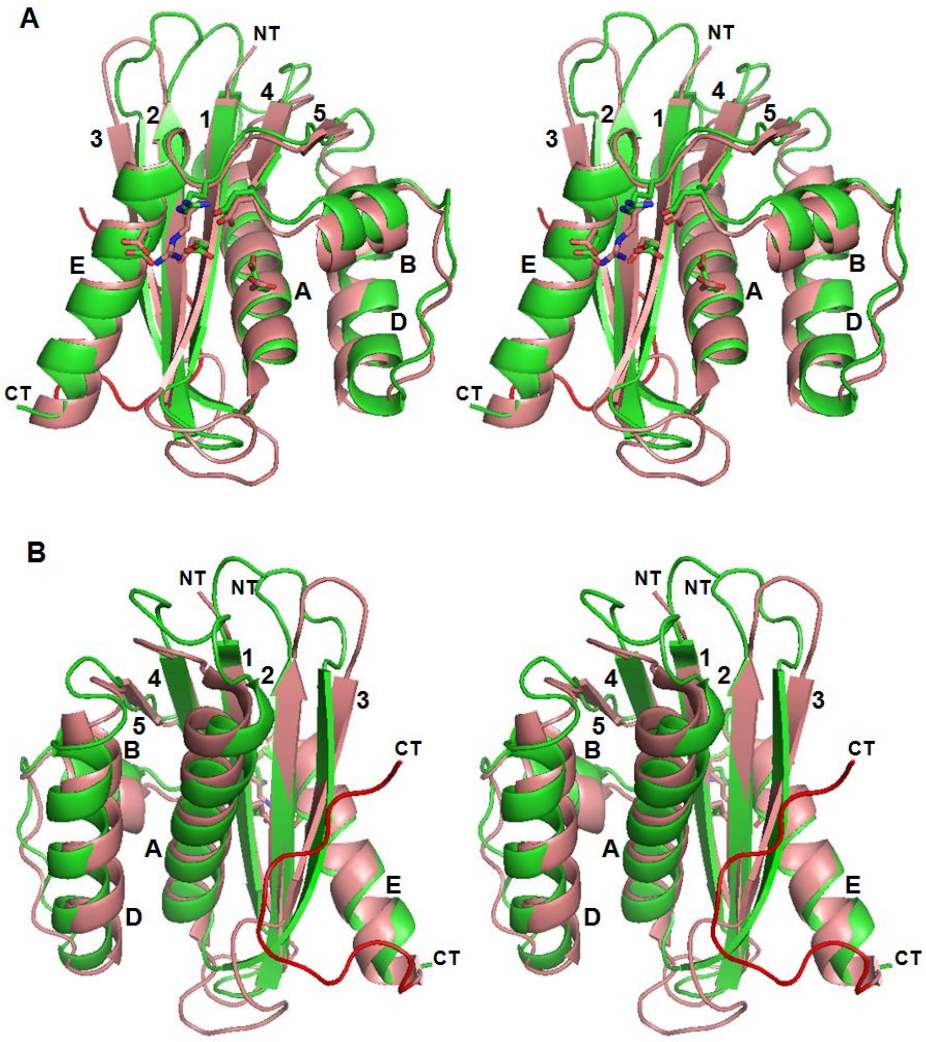

Figure 3-9. Stereoview of the structure of LC11-RNase H1 superimposed on that of Sto-RNase H1 (PDB code 2EHG). The LC11-RNase H1 and Sto-RNase H1 structures are colored green and salmon, respectively. Five active-site residues of LC11-RNase H1 (Asp9, Glu49, Asp77, Arg121, and Asp128) and the corresponding residues of Sto-RNase H1 (Asp7, Glu52, Asp76, Arg118, and Asp125) are shown by stick models. The C-terminal tail of Sto-RNase H1 is colored red. NT and CT represent N- and C-termini. The structures shown in (A) and (B) differ in the view direction.

The crystal structure of LC11-RNase H1 was determined at 1.4 Å resolution. The asymmetric unit of the crystal structure consists of four protein molecule (A-D). The structures of these four protein molecules are virtually identical with one another with root-mean-square deviation (RMSD) values of 1.00 Å between molecules A and D, 1.04 Å

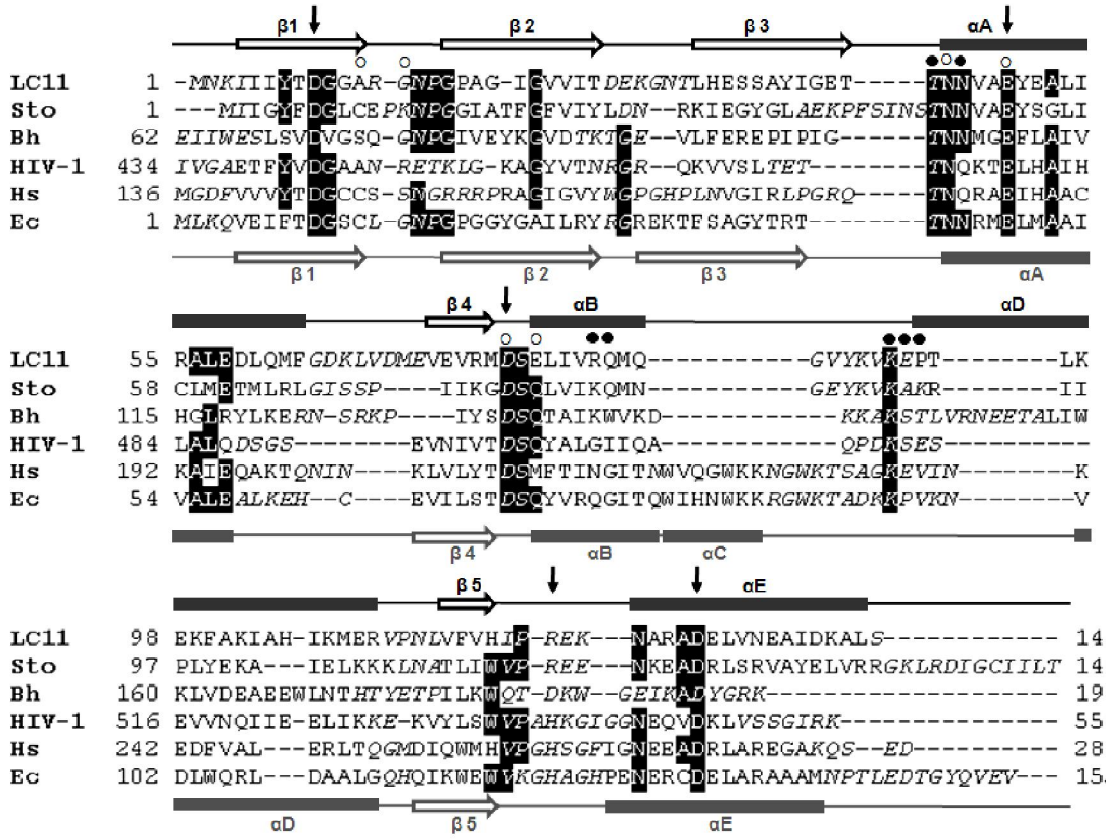

Figure 3-10. Alignment of the amino acid sequences. The amino acid sequences of LC11-RNase H1 (LC11), Sto-RNase H1 (Sto), Bh-RNase HC (Bh), HIV-RNase H (HIV-1), Hs-RNase HC (Hs) and Ec-RNase H1 (Ec) are compared with one another on the basis of their crystal structures. The ranges of the secondary structures of LC11-RNase H1 and Ec-RNase H1 are shown above and below the sequences, respectively. The amino acid residues that are not located within the α-helices or β-strands are shown in italic. The amino acid residues that are conserved in at least four different proteins are highlighted in black. The positions of the five active-site residues are indicated by arrows. The amino acid residues that contact the RNA and DNA strands of the substrate in the cocrystal structure of Bh-RNase HC* with the substrate are indicated by the open and closed circles above the sequences, respectively. Gaps are denoted by dashes. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

between molecules B and D, and 0.50 Å between molecules C and D for 137 Cα atoms. I used the structure of molecule D in this study.

The overall structure of LC11-RNase H1 consists of a five-stranded mixed β-sheet with three antiparallel (β1-β3) and two parallel (β4 and β5) strands and four α-helices (αA, αB, αD and αE) (Fig. 3-9A, B). This structure is highly similar to that of Sto-RNase H1 [56] with the RMSD value of 1.49 Å for 126 Cα atoms (Figs. 3-9A, B). However, this structure lacks a C-terminal tail, which is responsible for hyperstabilization of Sto-RNase H1 [56]. In addition, several loops, especially that between β3-strand and αA-helix vary in size for these proteins.

The amino acid sequence of LC11-RNase H1 is compared with those of Sto-RNase H1, Bh-RNase H1, HIV-RNase H, Hs-RNase H1 and Ec-RNase H1 on the bases of their crystal structures in Figure 3-10. The sequence of LC11-RNase H1 is 34% identical to Sto-RNase H1, 16% identical to Bh-RNase HC, 22% identical to HIV-RNase H, 24% identical to Hs-RNase HC, and 26% identical to Ec-RNase H1.

3.3.6. Catalytic center

The four acidic residues, Asp9, Glu49, Asp77 and Asp128, form the active site of LC11-RNase H1. The corresponding residues are Asp7, Glu52, Asp76 and Asp125 for Sto-RNase H1, and Asp71, Glu109, Asp132 and Asp192 for Bh-RNase H1. Superimposition of the active-site structures of LC11-RNase H1 and Sto-RNase H1 on the cocrystal structure of Bh-RNase HC* (inactive active-site mutant of Bh-RNase H1 without the HBD) with the substrate and metal cofactor [48] indicates that the steric configuration of the active-site residues of LC11-RNase HI is very similar to those of Sto-RNase H1 and Bh-RNase H1 (Fig. 3-11). Therefore, these four acidic residues of LC11-RNase H1 probably form the metal binding sites A and B, as seen in the cocrystal structure of Bh-RNase HC*. In addition, Arg121 of LC11-RNase H1 is located at the similar position where Arg118 of Sto-RNase H1 and Glu188 of Bh-RNase HC* are located (Fig. 3-11). This residue robably promotes a

product release by perturbing the coordination of the metal ion A, as proposed for Glu188 of Bh-RNase H1 [48].

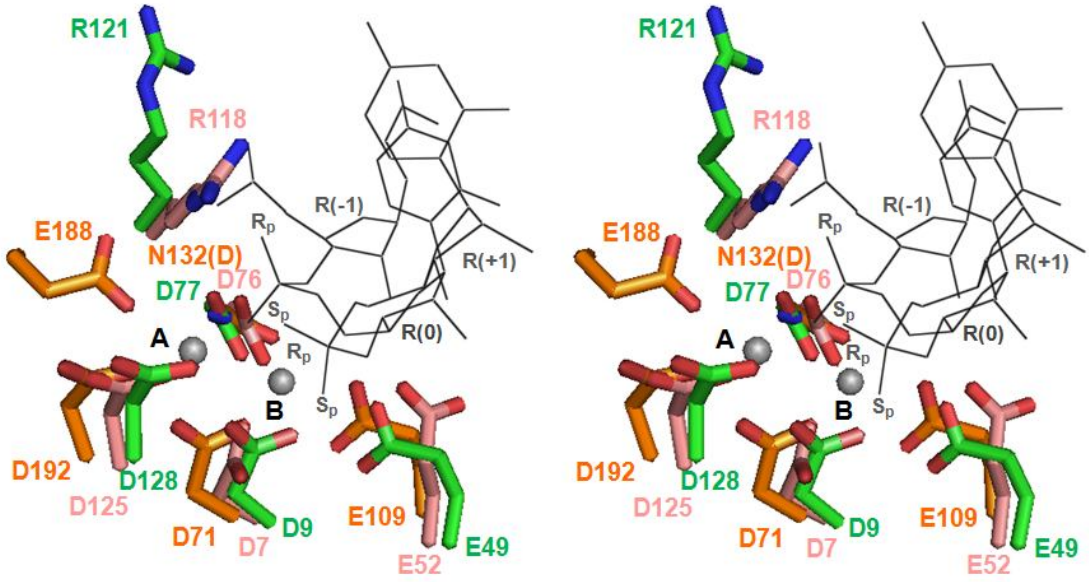

Fig. 3-11. Stereoview of the active-site structures of RNases H. The side chains of the active-site residues of LC11-RNase H1 (green) and Sto-RNase H1 (salmon) are superimposed onto those in the cocrystal structure of Bh-RNase HC* with the substrate and metal ions (orange, PDB code 1ZBI). The positions of the RNA strand of the substrate with the scissile phosphate group between R(-1) and R(0) and two metal ions A and B are also shown.

3.3.7. Substrate binding grooves on protein surface

According to the cocrystal structures of Bh-RNase HC* [48] and Hs-RNase HC* (inactive active-site mutant of Hs-RNase H1 without the HBD) [49] with the RNA/DNA substrate, these enzymes bind to the minor groove of the substrate by two grooves on protein surface, such that one groove containing the active site interacts with the RNA backbone and the other groove containing the phosphate-binding pocket interacts with the DNA backbone. Hs-RNase HC* has a DNA binding channel in the basic protrusion in addition to these grooves. dsRNA may not fit to these grooves and channel on protein surface because of its wide-groove A-form conformation, and therefore may not be recognized as a substrate by

Bh-RNase H1 and Hs-RNase H1. This may be the reason why these enzymes cannot cleave dsRNA.

To examine whether LC11-RNase H1 also has these grooves on protein surface, a model for the complex between LC11-RNase H1 and the substrate (12 bp RNA/DNA hybrid) was constructed based on the cocrystal structure of Bh-RNase HC* with the substrate (Fig. 3-12A, B). According to this model, RNA and DNA binding grooves are present on the surface of LC11-RNase H1. Lys92 and Glu93 apparently form humps, which separate inside and outside of the DNA-binding groove. These humps may interfere with the binding of LC11-RNase H1 to dsRNA. A similar DNA binding groove is present on the surface of Bh-RNase HC* (Fig. 3-12B). Lys146 and Thr148 form humps as Lys92 and Glu93 of LC11-RNase H1. However, according to a model for the Sto-RNase H1-substrate complex, a DNA-binding groove is not clearly detected on protein surface (Fig. 3-12B). Lys92 of LC11-RNase H1 and Lys146 of Bh-RNase H1 are conserved as Lys91 in Sto-RNase H1. However, this residue does not form a clear hump to separate inside and outside of the DNA-binding groove. HIV-RNase H also does not have this DNA-binding groove (data not shown). These results suggest that Sto-RNase H1 and HIV-RNase H1 can cleave dsRNA because they do not have a sharply-defined DNA-binding groove and therefore can accommodate dsRNA. However, it has been reported that the mutation of Lys91 of Sto-RNase H1 to Ala significantly reduces the dsRNase activity without seriously affecting the RNase H activity [56]. This result suggests that electrostatic interactions are responsible for efficient binding of Sto-RNase H1 to dsRNA. I found that the PIWI-containing module of the argonaute protein, a key catalytic component of the RNA-induced silencing complex, also does not have a clear binding groove for the guide strand and has positively charged residues at the guide strand binding site. According to the complexed structure of *Thermus thermophilus* argonaute protein (Tt-Ago) with 5’-phosphorylated 21-base guide DNA and a 12-nucleotide target RNA [114], and the complexed model of *Neurospora crassa* (Nc)

QDE-2 MID-PIWI lobe [115] with the substrate from the Tt-Ago complexed structure (data not shown), no binding groove is present on the surface of Tt-Ago MID-PIWI module or Nc QDE-2 MID-PIWI lobe at the DNA guide strand binding site. However they have the binding pocket for the 5' end of the DNA guide strand, which is located in the MID domain and contains several positively charged residues (Arg446 and Lys457 in Tt-Ago; Lys590, Lys617

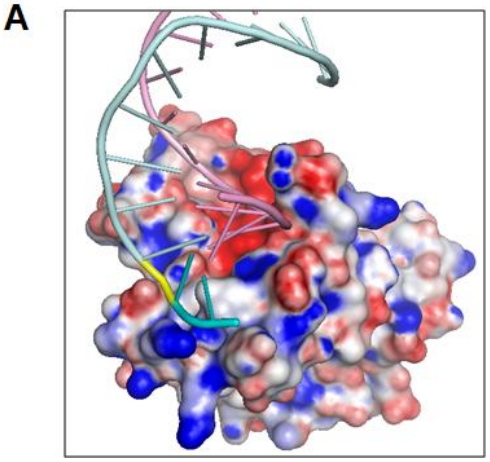

Fig 3-12A. A model for the complex between LC11-RNase H1 and RNA/DNA hybrid. Negative and positive surface potentials of LC11-RNase H1 are shown in red and blue, respectively. The electrostatic potential value ranges from –15 to +15 kT/e. The RNA and DNA strands of RNA/DNA hybrid are shown in magenta and cyan, respectively. The DNA phosphate group located two-base pairs away from the scissile bond is shown in yellow. The highly negatively charged region represents the active site.

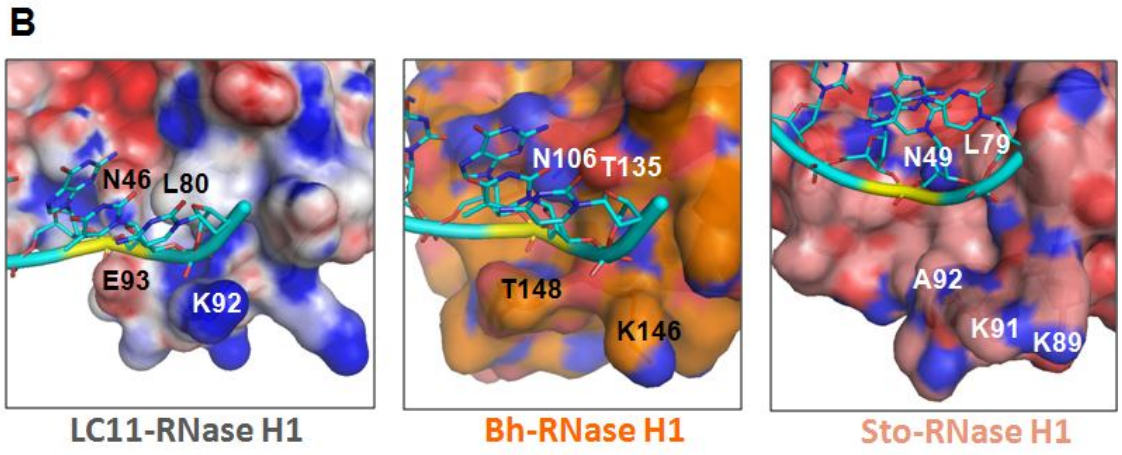

Fig. 3-12B. The structures around the DNA binding grooves containing the phosphate binding pockets. A model of the LC11-RNase H1-substrate complex, the structure of the Bh-RNase HC*-substrate complex, and a model of the Sto-RNase H1-substrate complex are shown, where LC11-RNase H1, Bh-RNase HC*, and Sto-RNase H1 are colored gray, orange, and salmon, respectively. The DNA strands of RNA/DNA hybrid is shown in cyan. The DNA phosphate group located two-base pairs away from the scissile bond is shown in yellow.

and Lys634 in Nc QDE-2). I propose that only the RNase H or RNase H-fold containing proteins, which do not have a clear DNA binding groove but have positively charged residues near the DNA binding site, can cleave dsRNA.

3.3.8. Structural features for the difference in stability between LC11-RNase H1 and Sto-RNase H1

Comparison of the crystal structure of LC11-RNase H1 with that of Sto-RNase H1ΔC6 [116] indicates that the number of buried non-polar residues and the number of ion pairs of LC11-RNase H1 are comparable to those of Sto-RNase H1ΔC6, while the number of buried charged residues and the total volume of cavities of LC11-RNase H1 are higher than those of Sto-RNase H1ΔC6 (Table 3-1). These results suggest that LC11-RNase H1 is destabilized as compared to Sto-RNase H1 not only by the removal of the C-terminal anchoring tail but also by the increase in the cavity volume and the number of buried charged residues. It has been reported that increased cavity volume [117,118] and burial of non-hydrogen-bonded polar or charged groups [119] contribute to protein instability.

Table 3-2. Structural features of LC11-RNase H1 and Sto-RNase H1ΔC6

	LC11-RNase H1	Sto-RNase H1ΔC6
No. of amino acids	140	141
T_m (^oC)	64	77
Content of residues (%)		
Buried charged	18	10
Buried non-polar	48	49
No. of ion pairs (≤5.0 Å)	15	12
Total volume of cavities (Å^3)	223	20

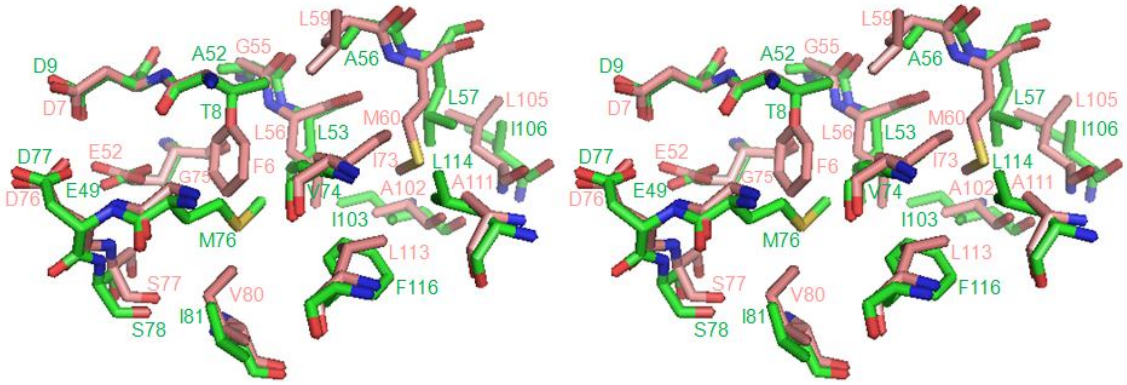

Fig. 3-13. Stereoview of the core structure of LC11-RNase H1 superimposed on that of Sto-RNase H1ΔC6 (PDB code 3ALY). The amino acid residues located at the hydrophobic cores of LC11-RNase H1 (green) and Sto-RNase H1ΔC6 (salmon) are shown by stick models.

The difference in cavity volume at the hydrophobic core between LC11-RNase H1 and Sto-RNase H1ΔC6 is shown in Fig. 3-13 as a typical example. Calculation of the cavity volume using the program CASTp [120] indicates that two cavities exist around Thr8 and Val74 of LC11-RNase H1.The total volume of these cavities is 158 $Å^3$. These cavities do not exist in the hydrophobic core of Sto-RNase H1ΔC6, because these cavities are filled by the amino acid substitutions from Thr8 to Phe6, Ile81 to Val80, Leu57 to Met60, and Val74 to Ile73.

3.4. Summary

LC11-RNase H1 is a homolog of Sto-RNase H1. It lacks a C-terminal tail, which is responsible for hyperstabilization of Sto-RNase H1. Sto-RNase H1 is characterized by its ability to cleave not only an RNA/DNA hybrid but also a double-stranded RNA (dsRNA). To examine whether LC11-RNase H1 also exhibits both RNase H and dsRNase activities, LC11-RNase H1 was overproduced in *E. coli*, purified, and characterized. LC11-RNase H1 exhibited RNase H activity, but did not exhibit dsRNase activity. LC11-RNase H1 was less stable than Sto-RNases H1 and its derivative lacking the C-terminal tail (Sto-RNase H1ΔC6) by 37 and 13°C in T_m, respectively. To understand the structural bases for these differences, the crystal structure of LC11-RNase H1 was determined at 1.4 Å resolution. The LC11-RNase H1 structure is highly similar to the Sto-RNase H1 structure. However, LC11-RNase H1 has two grooves on protein surface, one containing the active site and the other containing the DNA-phosphate binding pocket, while Sto-RNase H1 has one groove containing the active site. In addition, LC11-RNase H1 contains more cavities and buried charged residues than Sto-RNase H1. I propose that LC11-RNase H1 does not exhibit dsRNase activity because dsRNA cannot fit to the two grooves on protein surface and that LC11-RNase H1 is less stable than Sto-RNase H1ΔC6 because of the increase in cavity volume and number of buried charged residues.

CHAPTER 4

Crystal structure of LC11-RNase H1 in complex with RNA/DNA hybrid

4.1. Introduction

As mentioned in Chapter 1, up to now the crystal structure of the RNase H1 enzyme in complex with RNA/DNA hybrid is only available for Bh-RNase HC* [48] and Hs-RNases HC* [49]. Bh- and Hs-RNases HC* represent the inactive active-site mutants of Bh- and Hs-RNases HC, which represent Bh-RNase H1 and Hs-RNase H1 without the hybrid binding domain (HBD), respectively. HBD, which consists of approximately 40 amino acid residues, is commonly present at the N-terminal regions of eukaryotic RNases H1 [95]. It is also present in several bacterial RNases H1 including Bh-RNase H1, but is not present in LC11-, Sto- and Ec-RNases H1 (Fig. 4-1). HBD is important not only for substrate binding [50-52,95] but also for Mg^{2+}-dependent activity [48-50]. According to the structures of Bh- and Hs-RNases HC* in complex with the substrate, these proteins bind to the minor groove of the substrate, such that one groove containing the active site on the protein surface accommodates the RNA strand and the other containing the phosphate-binding pocket accommodates the DNA strand.

Bh- and Hs-RNases HC are regarded as representatives of bacterial and eukaryotic RNases H1. However, it remains to be determined whether other bacterial RNases H1, which do not contain an HBD and show relatively low amino acid sequence similarities to Bh-RNase HC and Hs-RNase HC, recognize the substrate by the similar mechanism. In addition, it remains to be determined whether these RNases H1 change their conformation upon substrate binding. LC11-RNase H1 does not contain an HBD and shows the amino acid sequence identities of only 16% to Bh-RNase HC. The amino acid residues of Bh-RNase HC that interact with the substrate are not well conserved in the LC11-RNase H1 structure as mentioned in Chapter 3. In addition, LC11-RNase H1 does not show a preference to Mg^{2+} or

Mn^{2+} ions for activity, unlike Bh- and Hs-RNases HC. Bh- and Hs-RNases HC show a strong preference to Mn^{2+} ions for activity, because the removal of the HBD greatly reduces the Mg^{2+}-dependent activities of Bh- and Hs-RNases H1 without seriously affecting their Mn^{2+}-dependent activities [48,49]. It has been suggested that the removal of the HBD affects the interaction between the enzyme and substrate at the active site [50]. Therefore, it would be informative to determine the crystal structure of LC11-RNase H1 in complex with the substrate.

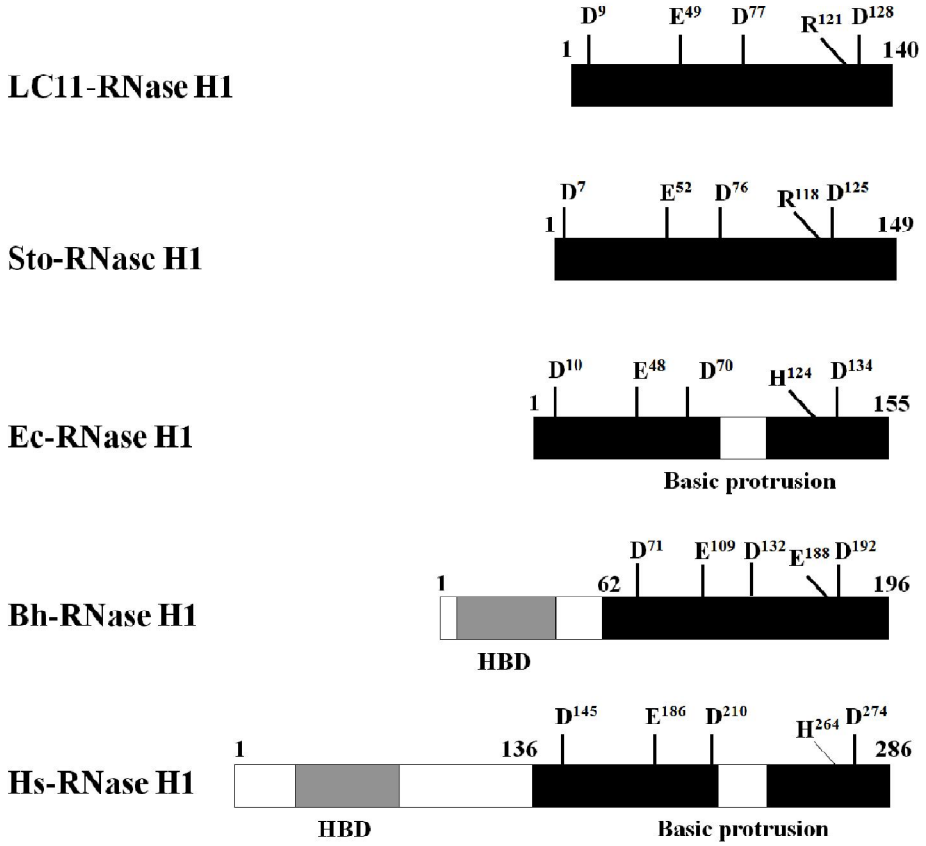

Figure 4-1. Schematic representation of the primary structures of LC11-, Sto-, Ec-, Bh- and Hs-RNases H1. Solid box represents the RNase H domain and grey box represents the hybrid binding domain (HBD). The basic protrusion, which is located in the middle of the sequences of Ec- and Hs-RNases H1, four conserved acidic active-site residues, and one non-acidic active-site residue, which is well conserved as His, are shown. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

In this chapter, I determined the crystal structure of LC11-RNase H1 in complex with the RNA/DNA hybrid using the active-site mutant, LC11-D77N. The structure reveals the difference in substrate conformation and protein-substrate interactions between the LC11-RNase H1 complex and Bh- or Hs-RNase HC* complex. Based on the structure of the LC11-RNase H1 complex and the structure-based mutational studies, I discuss the structural features for the difference in the ability to cleave dsRNA between LC11-RNase H1 and Sto-RNase H1.

4.2. Materials and methods

4.2.1. Plasmid construction, overproduction, and purification

Plasmids pET-D77N, pET-K92A and pET-E93A for overproduction of LC11-D77N, LC11-K92A and LC11-E93A, respectively, and plasmids pET-A92E/K93P, pET-K89A, pET-K91A and pET-K89A/K91A for overproduction of Sto-A92E/K93P, Sto-K89A, Sto-K91A and Sto-K89A/K91A, respectively, were constructed by PCR using the KOD-Plus Mutagenesis kit (Toyobo, Japan) according to the manufacturer's instructions. Plasmid pET-LC11 containing the gene encoding LC11-RNase H1 (Chapter 3) and plasmid pET-Sto containing the gene encoding Sto-RNase H1 [109] were used as a template. For the construction of LC11-RNase H1 mutants, the mutagenic primers were designed such that the GAC codon for Asp77 is changed to AAC for Asn, the AAG codon for Lys92 is changed to GCG for Ala, and the GAA codon for Glu93 is changed to GCA for Ala. For the construction of Sto-RNase H1 mutants, the mutagenic primers were designed such that the AAA codon for Lys89 and Lys91 is changed to GCA for Ala, the GCT codon for Ala92 is changed to GAG for Glu, and the AAA codon for Lys93 is changed to CCA for Pro. All DNA oligomers for PCR were synthesized by Hokkaido System Science (Japan). PCR was performed with a GeneAmp PCR system 2400 (Applied Biosystems, Japan). The DNA sequences were confirmed by a Prism 310 DNA sequencer (Applied Biosystems).

To avoid contamination of host-derived RNases H1 and H2, overproduction of LC11-RNase H1 mutants and Sto-RNase H1 mutants were also performed in *E. coli* MIC2067(DE3). LC11-RNase H1 mutants were purified by anion exchange column chromatography (DE52 and HiTrap Q, GE Healthcare, Japan) and gel filtration (Superdex 200 column, GE Healthcare), as described previously in Chapter 3. Sto-RNase H1 mutants were purified by cation exchange cloumn chromatography (SP HP column, GE Healthcare) and affinity chromatography (Heparin HP column, GE Healthcare), as described previously

[101]. The buffers for these chromatography were 10 mM sodium phosphate (pH 7.0) for DE52, 20 mM Tris-HCl (pH 9.0) for HiTrap Q, 20 mM Tris-HCl (pH 7.5) containing 50 mM NaCl for Superdex 200, 10 mM Tris-HCl (pH 8.0) containing 1 mM EDTA for HiTrap SP HP, and 10 mM Tris-HCl (pH 7.5) containing 1 mM EDTA for HiTrap Heparin HP. LC11-RNase H1 mutants were stored in 10 mM Tris-HCl (pH 7.5), while Sto-RNase H1 mutants were stored in 10 mM sodium acetate (pH 4.5). The purity of the protein was analyzed by SDS-PAGE using a 15% polyacrylamide gel [103], followed by staining with Coomassie Brilliant Blue (CBB). The protein concentration was determined from UV absorption using a cell with an optical path length of 1 cm, on the assumption that the A_{280} values for 0.1% (1.0 mg·ml^{-1}) solution of LC11-RNase H1 (0.30) and Sto-RNase H1 (0.97) [56] are not changed by the mutations.

4.2.2. Crystallization

Prior to crystallization, LC11-D77N was mixed with 6, 12, or 14 bp RNA/DNA hybrid at a 1:1.2 protein:substrate molar ratio. The final protein concentration was 5 mg·ml^{-1}. The crystallization conditions were initially screened using crystallization kits from Hampton Research (USA) (Crystal Screens and Index). The conditions were surveyed using sitting-drop vapour-diffusion method at 18°C. Drops were prepared by mixing the complexes with the reservoir solution at equal volume (0.2 µl) using a phoenix crystallization robot (Art Robbins Instruments, USA). Native crystals only appeared in the mixture of LC11-D77N with 14 bp RNA/DNA hybrid after one day at various conditions which contain PEG 3350. The crystallization conditions were further optimized and the crystals were obtained when the drop was prepared by mixing 1.2 µl complex solution at the final protein concentration of 7 mg·ml^{-1} with 1.2 µl reservoir solution [0.2 M ammonium acetate, 0.1 M Tris-HCl (pH 8.5), 25% PEG 3350] against 100 µl reservoir solution at 20°C. Crystals grew to final size (0.4 - 0.5mm) in 10 - 14 days. Since the crystal was obtained in the reservoir solution containing

25% PEG 3350, it was used directly for diffraction data collection without further cryoprotection.

4.2.3. X-ray diffraction data collection and structure determination

X-ray diffraction data set of LC11-RNase H1 in complex with 14 bp RNA/DNA hybrid was collected at a wavelength of 0.9 Å at -173°C using synchrotron radiation on the BL44XU station at SPring-8 (Hyogo, Japan). The data set was indexed, integrated and scaled using the program HKL2000 [112]. Although I got the reliable results for I/sigma and R_{merge} at 2.1 Å resolution, the completeness was relatively low. Therefore, the structure was determined at 2.3 Å resolution. The crystal belonged to the space group $P2_1$, with unit cell parameters a = 42.613, b = 90.758 and c = 55.605 Å. The complex structure of LC11-RNase H1 with the substrate was determined by molecular replacement using MOLREP [113]. The crystal structure of LC11-RNase H1 (PDB code 3U3G) was used as a search model for the protein moiety. The crystal structure of a 12 bp RNA/DNA substrate in complex with Bh-RNase HC* (PDB code 1ZBI) was used as a search model for the substrate moiety. The structure solution of a 12 bp RNA/DNA substrate was further built into a 14 bp RNA/DNA substrate by adding two more base pairs using Coot [121]. Automatic model building and also automatic placement of water molecules were then done using Arp/wArp [122], followed by manually checking on correct assignment. Iteration between model refinement using REFMAC5 [123] and manual adjustments using Coot was done until no further decrease of the R_{free} value was observed. The statistics for data collection and refinement are summarized in Table 4-1. The figures were prepared using PyMol (http://www.pymol.org).

4.2.4. Protein data bank accession number

The coordinates and structure factors for the complex between LC11-D77N and 14 bp RNA/DNA hybrid have been deposited in the Protein Data Bank under accession code 4H8K.

Table 4-1. Data collection and refinement statistics

Data collection	
Wavelength (Å)	0.9
Space group	$P2_1$
Unit cell	
a, b, c (Å)	42.6, 90.8, 55.6
α, β, γ (°)	90.0, 97.8, 90.0
Resolution (Å)	55.09 – 2.30 (2.34 – 2.30)[a]
Observations	132683
Unique reflections	17868 (782)[a]
Completeness (%)	95.4 (82.1)[a]
R_{merge} (%)[b]	4.5 (19.1)[a]
Average I/s (I)	43.6 (5.7)[a]
Refinement statistics	
Resolution limits (Å)	38.28 – 2.30
Number of atoms	
Protein	2156
RNA/DNA	579
Water	94
R_{work} (%)	20.7
R_{free} (%)[c]	27.8
Deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.053
Overall mean B factors (Å^2)	52.05
Ramachandran analysis (%)	
Favored regions	95.2
Allowed regions	4.8

[a] Values in parentheses are for the highest resolution shell.

[b] $R_{merge} = \sum | I_{hkl} - \langle I_{hkl} \rangle | / \sum I_{hkl}$, where I_{hkl} is the intensity measurement for reflection with indices hkl and $\langle I_{hkl} \rangle$ is the mean intensity for multiply recorded reflections.

[c] R_{free} was calculated using 5 % of the total reflections chosen randomly and omitted from refinement.

4.2.5. Enzymatic activity

The enzymatic activity was determined at 37°C using 5'-fluorescein-labeled 12 bp RNA/DNA hybrid (R12/D12), 12 bp RNA/RNA duplex (R12/R12), 29 bp DNA_{13}-RNA_4-DNA_{12}/DNA_{29} duplex (D13-R4-D12/D29), 29 bp DNA_{13}-RNA_5-DNA_{11}/DNA_{29} duplex (D13-R5-D11/D29), and 29 bp DNA_{13}-RNA_6-DNA_{10}/DNA_{29} duplex (D13-R6-D10/D29) as a substrate, as described in Chapter 3. The sequence of D13-R4-D12/D29 is 5'-AATAGAGAAAAAGaaaaAAGATGGCAAAG-3', in which DNA and RNA are represented by uppercase and lowercase letters, respectively. The sequences of D13-R5-D11 and D13-R6-D10 are identical to that of D13-R4-D12, except that one and two "A"s located downstream of the tetraribonucleotides are changed to "a", respectively. These oligomers were synthesized by Hokkaido System Science. The reaction buffer used for assay was 10 mM Tris-HCl (pH 8.5) containing 10 mM $MgCl_2$ or $MnCl_2$, 10 mM NaCl, 1 mM 2-mercaptoethanol, and 50 $\mu g \cdot ml^{-1}$ bovine serum albumin. The products were detected by Typhoon 9240 Imager (GE Healthcare).

4.3. Results and discussion

4.3.1. Construction of active-site mutant

As mentioned in the previous chapter, four acidic residues, Asp9, Glu49, Asp77, and Asp128, form the active site of LC11-RNase H1. These residues are conserved as Asp71, Glu109, Asp132, and Asp192 for Bh-RNase H1, Asp145, Glu186, Asp210, and Asp274 for Hs-RNase H1, Asp7, Glu52, Asp76, and Asp125 for Sto-RNase H1, and Asp10, Glu48, Asp70, and Asp134 for Ec-RNase H1, respectively (Fig. 4-1). It has been reported for Ec-RNase H1 that Asp10 and Asp70 are indispensable for activity, whereas Glu48 and Asp134 are dispensable for it [124]. Indispensability of the first and/or third residue and dispensability of the fourth residue of a DEDD active-site motif for activity have also been reported for Bh-RNase H1 [48], Hs-RNase H1 [49], and Sto-RNase H1 [58]. Therefore, I constructed the active-site mutant of LC11-RNase H1, LC11-D77N, in which the third residue of a DEDD active-site motif, Asp77, is replaced by Asn, to prevent degradation of the RNA/DNA substrate during crystallization. However, when the RNase H activity of LC11-D77N was determined using 12 bp RNA/DNA hybrid (R12/D12) as a substrate, LC11-D77N exhibited 20% of the activity of LC11-RNase H1 either in the presence of Mg^{2+} or Mn^{2+} ions (data not shown). Asp77 of LC11-RNase H1 may coordinate with metal ion B, regardless of whether it is mutated to Asn, as reported for the third residue of a DEDD active-site motif of Bh-RNase HC [48].

4.3.2. Structure determination

Crystallization of the LC11-D77N-substrate complex was attempted by using various substrates with different sizes. Crystals were obtained only from the protein complexed with the 14 bp RNA/DNA substrate in the absence of metal ions. Crystals of the protein-substrate complex were not obtained in the presence of metal ions due to the degradation of the

substrate during crystallization, in agreement with the biochemical results which showed residual activity of LC11-D77N. The crystal structure of LC11-D77N complexed with the 14 bp RNA/DNA substrate was determined at 2.3 Å resolution. Data collection and refinement statistics are shown in Table 4-1. Two LC11-D77N molecules bind to the minor groove of the substrate, such that each molecule binds to one half of the substrate (Fig. 4-2A). Their structures are virtually identical with root-mean-square deviation (RMSD) value of 0.40 Å for 138 Cα atoms. The overall structure of the LC11-D77N-substrate complex is similar to that of Bh-RNase HC* (inactive Bh-RNase H1 derivative without the HBD and with the mutation of Asp132 to Asn) complexed with the 12 bp RNA/DNA substrate and two Mg^{2+} ions [48] (Fig. 4-2A). It is also similar to that of Hs-RNase HC* (inactive Hs-RNase H1 derivative without the HBD and with the mutation of Asp210 to Asn) complexed with the 14 bp RNA/DNA substrate and two Ca^{2+} ions [49].

4.3.3. Active-site structure

I designate Asp9, Glu49, Asn77, and Asp128 of LC11-D77N and the corresponding residues of Bh-RNase HC* and Hs-RNase HC* as four conserved active-site residues, or simply active-site residues, in this study, although Bh-RNase HC* and Hs-RNase HC* are nearly inactive. I also designate the protein-substrate complex simply as the protein complex. The steric configurations of the four conserved active-site residues of LC11-D77N are similar to those of Bh-RNase HC* and Hs-RNase HC* (Fig. 4-2B). However, the position of the scissile phosphate group in the LC11-D77N complex is shifted away from the active site by 1.9 Å and 1.1 Å as compared to those in the Bh-RNase HC* and Hs-RNase HC* complexes, respectively. As a result, the distances of the oxygen atom of the scissile phosphate group from the side chains of the active-site residues in the LC11-D77N complex (4.5 Å for Asp9, 4.3 Å for Glu49, 3.9 Å for Asn77, and 4.7 Å for Asp128) increase as compared to the corresponding ones in the Bh-RNase HC* complex (3.2 Å for Asp71, 3.6 Å for Glu109,

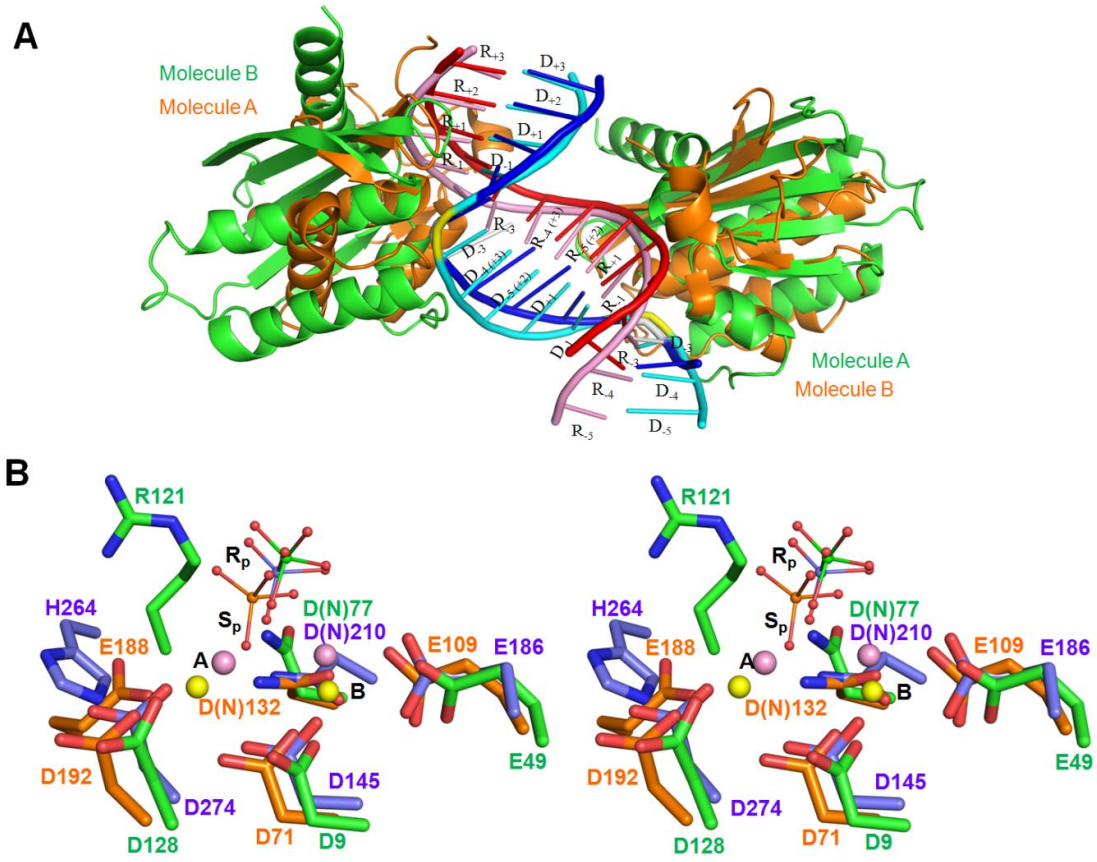

Figure 4-2. Comparision of the structures of LC11-D77N, Hs-RNase HC* and Bh-RNase HC* in complex with the substrate. **(A)** The overall structure of LC11-D77N (green) in complex with 14 bp RNA/DNA hybrid superimposed onto that of Bh-RNase HC* (orange) in complex with 12 bp RNA/DNA hybrid (PDB code 1ZBI). RNA and DNA in the LC11-D77N complex are colored pink and cyan, respectively, while those in the Bh-RNase HC* complex are colored red and blue, respectively. The positions of deoxyribonucleotides and ribonucleotides are indicated by D_n and R_n, respectively, where "n" represents the nth residue from the scissile phosphate group. The phosphate group between R_{-1} and R_{+1} contacts the active site. The D_{-3} phosphates in the LC11-D77N and Bh-RNase HC* complexes are highlighted in yellow and grey, respectively. **(B)** Stereoview of the active-site structures of LC11-D77N (molecule A, green) and Hs-RNase HC* (molecule A, slate) superimposed onto that of Bh-RNase HC* (molecule B, orange). Five active-site residues are shown by stick models, in which the oxygen and nitrogen atoms are colored red and blue, respectively. The scissile phosphate groups are shown in ball-and-line models, in which the oxygen atom is colored red and the phosphorus atoms of the LC11-D77N, Hs-RNase HC* and Bh-RNase HC* complexes are colored green, slate and orange, respectively. Two Mg^{2+} ions A and B in the Bh-RNase HC* complex are also shown as yellow spheres. Pro-R_p and pro-S_p oxygens are labeled as R_p and S_p, respectively.

3.5 Å for Asn132, and 4.5 Å for Asp192) and Hs-RNase HC* complex (3.5 Å for Asp145, 3.9 Å for Glu186, 4.0 Å for Asn210, and 4.5 Å for Asp274). The side chain of the fifth active-site residue of Bh-RNase HC*, Glu188, is located 3.6 Å away from the scissile phosphate group (Fig. 4-2B). The corresponding residue, Glu122, of LC11-D77N is located far away from this phosphate group (9.1 Å), whereas Arg121 of LC11-D77N is located 3.5 Å away from it (Fig. 4-2B), suggesting that Arg121 functions as a substitute of Glu188 of Bh-RNase HC. In fact, the corresponding residue of Sto-RNase H1, Arg118, has been shown to be important for catalytic function [56].

A shift of the scissile phosphate group has also been observed in the structure of the Hs-RNase HC* complex with a single Ca^{2+} ion, which is considered as a "nonproductive complex" [49]. However, this shift is large (up to 3.4 Å) and the RNA strand is also greatly displaced from the active site, while the RNA strand of the LC11-D77N complex is superimposed well on that of Hs-RNase HC* complex with two Ca^{2+} ions (Fig. 4-3A), suggesting that the LC11-D77N complex is possibly considered as a "productive complex".

It remains to be determined whether a similar shift of the scissile phosphate group is observed in the structure of the LC11-D77N complex in a metal-bound form. However, binding of the metal ions to the active site may not seriously affect the distance between the scissile phosphate group and the active-site residues, because comparison of the structures of the Hs-RNase HC* complex with two Ca^{2+} ions and that without Ca^{2+} ions shows that the distances of the scissile phosphate group from the active-site residues are not seriously changed regardless of the presence or absence of the Ca^{2+} ions (Fig. 4-3B). LC11-RNase H1 (as mentioned in Chapter 3), Bh-RNase HC [48], and Hs-RNase HC [49] exhibit activity either in the presence of Mg^{2+} or Mn^{2+} ions. However, LC11-RNase H1 does not show a preference to Mg^{2+} or Mn^{2+} ions, whereas Bh-RNase HC and Hs-RNase HC show a strong preference to Mn^{2+} ions. Full-length Bh-RNase H1 and Hs-RNase H1 prefer Mg^{2+} to Mn^{2+} ions for activity, but only their Mg^{2+}-dependent activities are greatly reduced by the removal

of the HBD. Therefore, the relative distances of the active-site residues to the scissile phosphate group may account for the difference in metal ion preference between LC11-RNase H1 and Bh- or Hs-RNase HC. Alteration in these distances may not seriously affect the binding of the Mn^{2+} ions to the active site, but may significantly affect the binding of the Mg^{2+} ions, because Mn^{2+} is a transition metal having coordinates with a different geometry than Mg^{2+}. However, further structural studies of LC11-RNase H1 and full-length Bh- or Hs-RNase H1 will be necessary to understand the structural basis for the difference in metal ion preference between LC11-RNase H1 and Bh- or Hs-RNase HC.

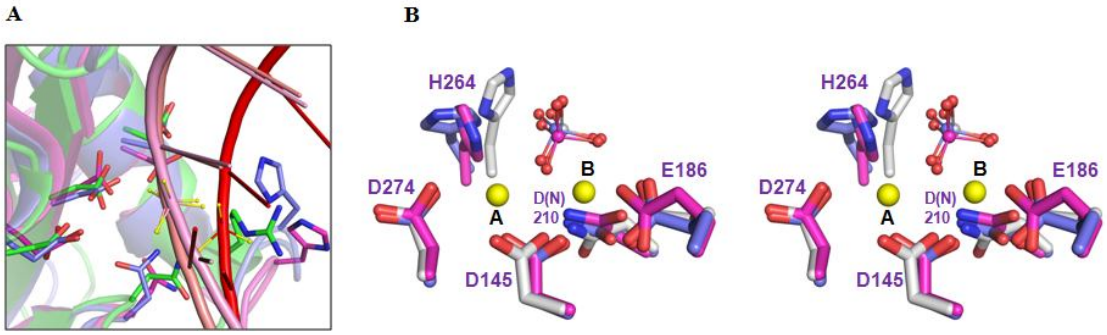

Fig. 4-3. (A) Comparision of the structures around the scissile phosphate group of LC11-D77N (green), Hs-RNase HC* with two Ca^{2+} ions (slate) and Hs-RNase HC* with one Ca^{2+} ion (light magenta) in complex with the substrates. Five active-site residues are shown by stick models, in which the oxygen and nitrogen atoms are colored red and blue, respectively. The RNA strand in the LC11-D77N complex are colored pink, while those in the Hs-RNase HC* with two Ca^{2+} ions and Hs-RNase HC* with one Ca^{2+} ion complexes are colored salmon and red, respectively. The DNA strands were not displayed. The scissile phosphates are indicated by yellow ball-and-line models. **(B)** Stereoview of the active site of Hs-RNase HC*. The side chains of the active-site residues in the Hs-RNase HC* complexes with 18 and 20 bp RNA/DNA and without Ca^{2+} ions (grey, PDB code 2QK9, and light magenta, PDB code 2QKB, respectively) are superimposed onto those in the Hs-RNase HC* complex with 14 bp RNA/DNA hybrid and two Ca^{2+} ions (slate, PDB code 2QKK). Five active-site residues are shown by stick models, in which the oxygen and nitrogen atoms are colored red and blue, respectively. The scissile phosphate groups are shown in ball-and-line models, in which the oxygen is colored red and the phosphorus atoms of the 14-mer, 18-mer and 20-mer complexes are colored slate, light magenta and grey, respectively. Two Ca^{2+} ions A and B in the 14-mer complex are also shown as yellow spheres.

4.3.4. Substrate conformation and binding

At the protein-substrate interface, the backbones of the RNA/DNA hybrid fit into two grooves on the protein surface, which is separated by a ridge (Fig. 4-4A). The RNA strand binds to the groove containing the active site, whereas the DNA strand binds to the groove containing the phosphate binding pockets. In the structure of the protein-substarte complex, the position of the ribonucleotide is indicated by R_n, where the suffix "n" represents the nth residue from the scissile phosphate group (negative for the upstream region and positive for the downsteam reagion). For example, R_{-1} and R_{+1} represent the ribonucleotides immediately 5' and 3' to the scissile phosphate group, respectively. The deoxyribonucleotide that forms a base pair with R_n is indicated by D_n. The R_n or D_n phosphate represents the 5' phosphate group of R_n or D_n. For example, the DNA phosphate two base pairs away from the scissile phosphate group is D_{-3} phosphate.

The RNA strand of the RNA/DNA hybrid is in the A form with C3'-endo sugar puckers, as is those in the Bh-RNase HC* [43] and Hs-RNase HC* [49] complexes. However, unlike Bh-RNase HC* or Hs-RNase HC*, which makes direct contacts with five or four consecutive 2'-OH groups, respectively, LC11-RNase H1 interacts with four non-consecutive 2'-OH groups, two on each side of the scissile phosphate group (Fig. 4-5). The backbone atoms of Ala12 and Gly14, which are adjacent to the active-site residue Asp9, contact the 2'-OH groups of R_{+1} and R_{+2}, respectively. These contacts are well compared with those made by Ser74 and Gly76 adjacent to the active-site residue Asp71 of Bh-RNase HC* or Cys148 and Ser150 adjacent to the active-site residue Asp145 of Hs-RNase HC*. Likewise, Glu79 adjacent to the active-site residue Asn77 contacts the 2'-OH groups of R_{-2} and R_{-3}. These contacts are well compared with those made by Gln134 adjacent to the active-site residue Asn132 of Bh-RNase HC* or Met212 adjacent to the active-site residue Asn210 of Hs-RNase HC*, although the contact between the protein and 2'-OH group of R_{-3} is missing in the Hs-RNase HC* complex. However, unlike the active-site residues Glu109 of Bh-RNase HC*

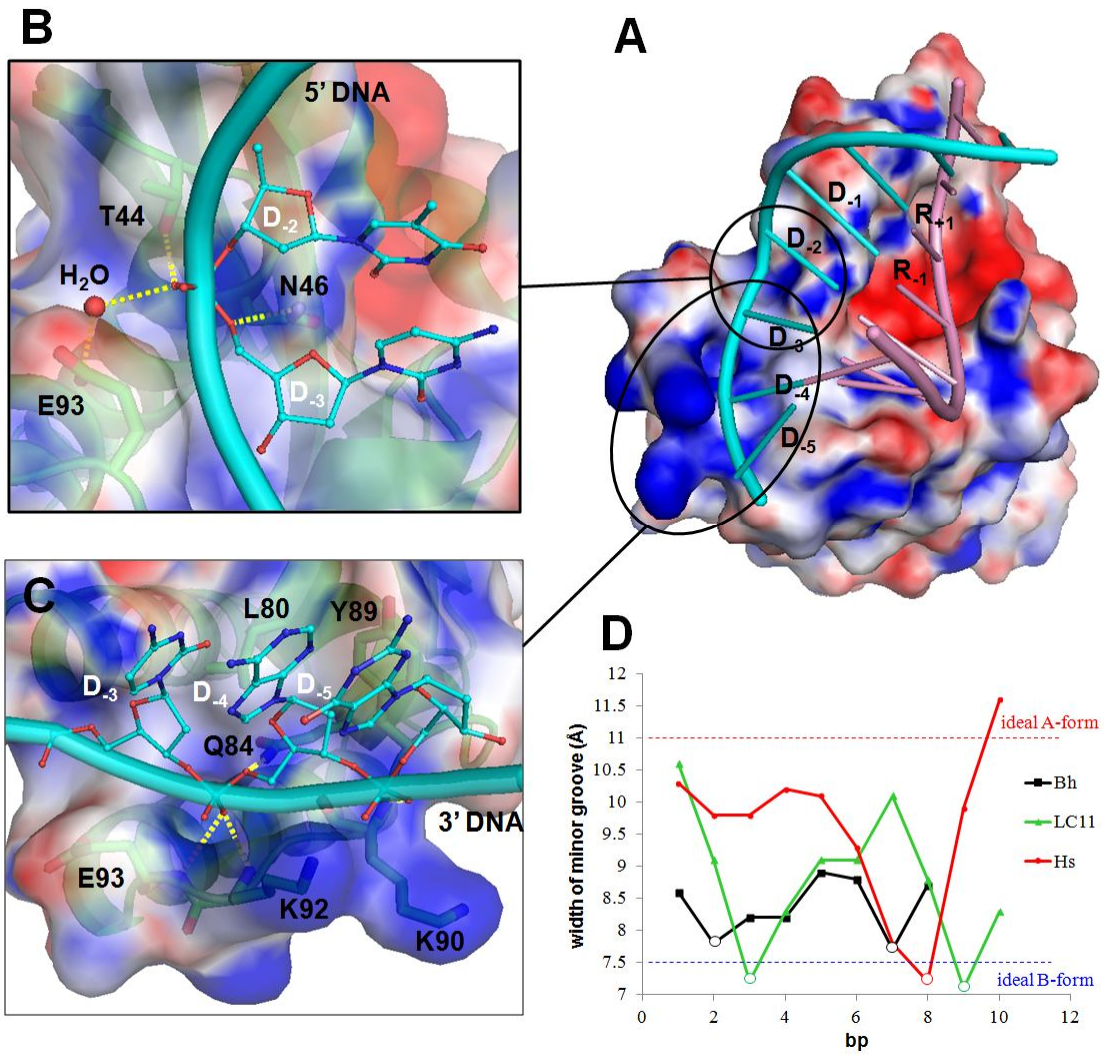

Figure 4-4. The conformation of RNA/DNA hybrid and its interaction with LC11-D77N. **(A)** The surface potential representation of LC11-D77N (molecule A) in complex with the 14-bp RNA/DNA hybrid in the tube-and-stick model. Negative and positive surface potentials of LC11-D77N are shown in red and blue, respectively. The electrostatic potential value ranges from –10 to +10 kT/e. The RNA and DNA strands are colored pink and cyan, respectively. **(B, C)** The structures around the first (B) and second (C) phosphate-binding sites. The interactions between the D_{-3} phosphate and protein (Thr44, Asn46, and Glu93), between the D_{-4} phosphate and protein (Gln84, Lys92, and Glu93), and between the D_{-5} phosphate and Lys90 are indicated by yellow dotted lines. The oxygen atoms are colored red. The water molecule is shown as a red sphere. **(D)** Plot of the minor groove width of the 12-, 14- and 18-bp RNA/DNA hybrids in complex with Bh-RNase HC*, LC11-D77N, and Hs-RNase HC*, respectively. The minor groove widths of ideal A and B form helices are indicated. D_{-3} phosphates are indicated by open circles.

and Glu186 of Hs-RNase HC*, the corresponding residue Glu49 of LC11-D77N cannot contact the 2'-OH group of R_{-1} due to a shift of its position. Glu49 is located 4.0 Å away from this 2'-OH group.

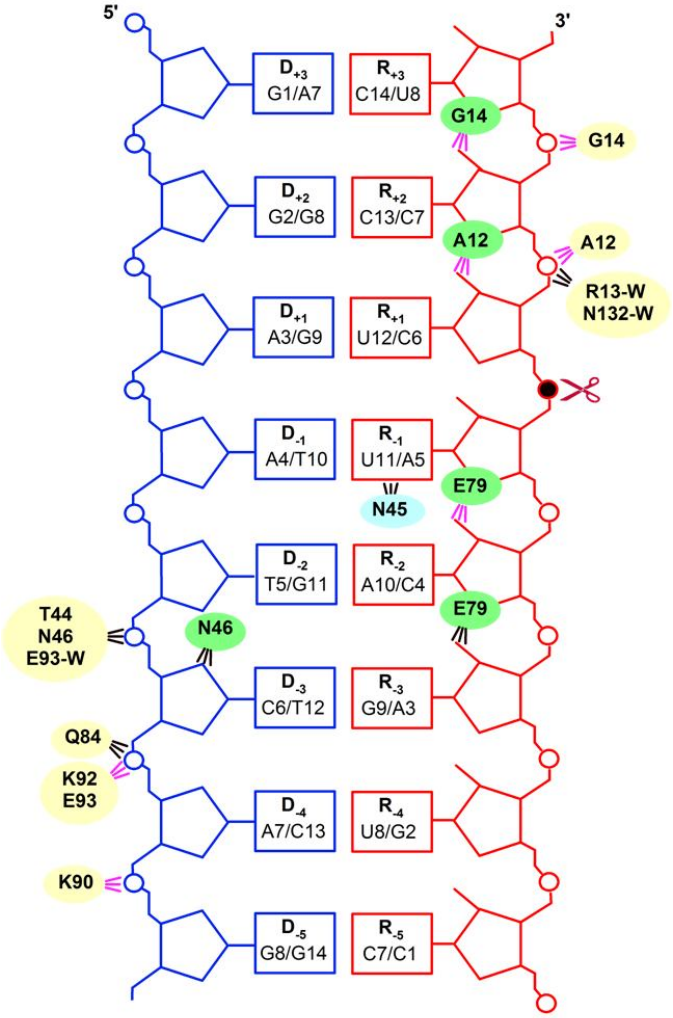

Fig. 4-5. The interactions between the RNA/DNA hybrid and LC11-D77N. The residues interacting with phosphate, sugars, and bases are shown in yellow, green and cyan highlight, respectively. Pink lines indicate interactions made by the protein backbone, and black lines indicate interactions made by side chains. Water-mediated interactions are indicated by -W.

LC11-D77N makes direct contact with three consecutive phosphate groups of the DNA strand (D_{-3}-D_{-5}) in the DNA-binding groove (Fig. 4-5). The first phosphate-binding site, to which the D_{-3} phosphate binds, is formed by Thr44/Asn46/Glu93 (Fig. 4-4B). This site is

similar to those of Bh-RNase HC* [43] and Hs-RNase HC* [44], which are formed by Thr104/Asn106/Ser147/Thr148 and Arg179/Thr181/Asn240, respectively. In contrast, the second phosphate-binding site, to which the D_{-4} phosphate binds, is unique to LC11-D77N. This site is formed by Gln84/Lys92/Glu93 (Fig. 4-4C). These residues, as well as Leu80 and Tyr89, form a channel, to which the phosphate group and ribose of D_{-4} bind. HS-RNase HC* does not have this channel, but has a DNA-binding channel, to which the D_{-5} phosphate binds, in the basic protrusion [44]. This channel is formed by Trp221/Trp225/Ser233. LC11-D77N does not have a basic protrusion, but has a site, to which the D_{-5} phosphate binds. The backbone of Lys90 directly contacts with this phosphate group.

In order to place the D_{-3} and D_{-4} phosphates in the first and second phosphate-binding sites of LC11-D77N, the backbone torsion angles (gamma and alpha) of DNA are dramatically distorted by ~150^{o} from the ideal values. Only DNA can accecpt such a large distorsion of the backbone torsion angles. Noticiably, while sugar puckers of most of the deoxynucleotides in the DNA strand are B-like (mainly C2'-endo or O1'-endo), those of D_{-4} and D_{-5} are in A-form or between A and B forms (C3'-endo or C4'-exo). It has been reported that the deoxynucleotide in the DNA-binding channel of Hs-RNase H1 is also between A and B forms (C4'-exo) [44]. As do the minor groove widths of the RNA/DNA hybrid in the Hs-RNase HC* complex, those in the LC11-D77N complex vary from ~7 Å at the D_{-3} phosphate to ~10 Å at the D_{-5} phosphate (Fig. 4-4D). Neither this large variation of sugar pucker conformations nor the minor groove width is not observed in the Bh-RNase HC* complex, because Bh-RNase HC* does not have a DNA-binding channel, to which the D_{-4} or D_{-5} phosphate binds. As a result, two Bh-RNase HC* molecules bind to the two D_{-3} phosphates with four phosphates in between, whereas two LC11-D77N or Hs-RNase HC* molecules bind to the two D_{-3} phosphates with five phosphates in between (Fig. 4-2A). When an RNA strand is modeled into the DNA-binding groove of LC11-D77N based on the DNA backbone positions, two 2'-OH groups clash with the side chain of Leu80 and the benzene ring of Tyr89. In the corresponding model of Hs-RNase HC*, one 2'-OH group clashes with

the indole ring of Trp221 [44]. Thus, as proposed for Hs-RNase H1, LC11-RNase H1 probably recognizes the DNA strand by the absence of 2'-OH groups, which cause steric clashes at the DNA-binding groove, and by the flexible B form conformation.

4.3.5. Cleavage of DNA-RNA-DNA/DNA substrates containing short ribonucleotides

To examine whether LC11-RNase H1 interacts with four nonconsecutive 2'-OH groups of the RNA strand as indicated by the crystal structure of the LC11-D77N complex, the activity of LC11-RNase H1 was analyzed using the DNA_{13}-RNA_4-DNA_{12}/DNA_{29} (D13-R4-D12/D29), DNA_{13}-RNA_5-DNA_{11}/DNA_{29} (D13-R5-D11/D29) and DNA_{13}-RNA_6-DNA_{10}/DNA_{29} (D13-R6-D10/D29) duplexes as a substrate, and compared with that of Ec-RNase H1. In these substrates, four (a14-a17), five (a14-a18), and six (a14-a19) ribonucleotides (adenosines) are flanked by 13 and 10-12 b DNA at the 5' and 3' sides, respectively. According to a model of the Ec-RNase H1 complex constructed based on the structure of the Hs-RNase HC* complex, Ec-RNase H1 forms hydrogen bonds with four consecutive 2'-OH groups of the RNA strand, two on each side of the scissile phosphate group. It has been reported that the ability of Ec-RNase H1 to cleave the DNA-RNA-DNA/DNA substrate containing three ribonucleotides is greatly reduced as compared to that to cleave the substrate containing four ribonucleotides [6]. Ec-RNase H1 cleaves the substrate containing four ribonucleotides preferentially at the middle of the ribonucleotides. These results indicate that complete formation of the hydrogen bonds between the protein and the substrate is required for efficient cleavage of the substrate by the enzyme. Therefore, if LC11-RNase H1 forms hydrogen bonds with four nonconsecutive 2'-OH groups of the substrate, it would cleave the D13-R4-D12/D29 substrate much less efficiently than the D13-R5-D11/D29 and D13-R6-D10/D29 substrates, because only the D13-R5-D11/D29 and D13-R6-D10/D29 substrates permit full contacts with the protein.

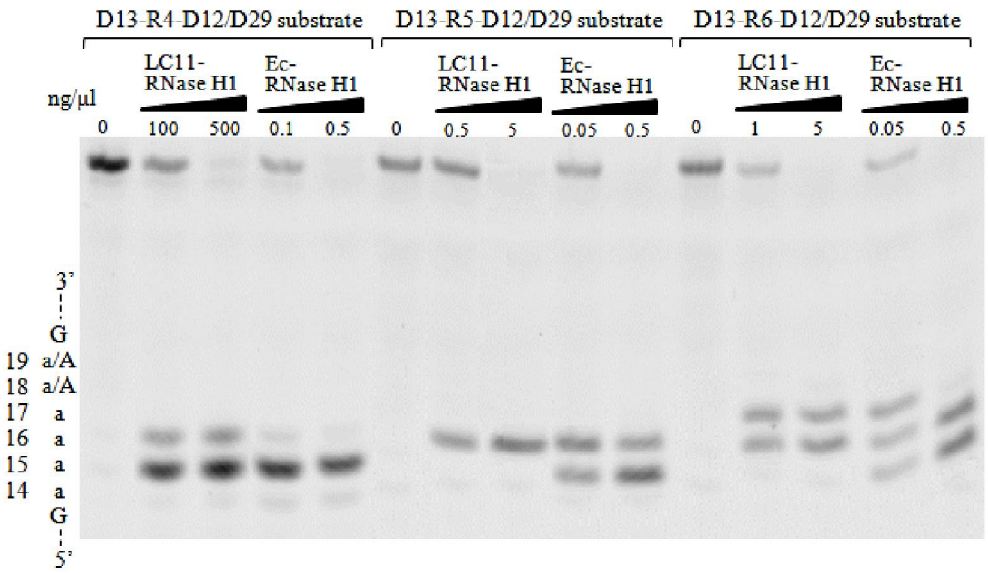

Figure 4-6. Cleavage of the D13-R4-D12/D29, D13-R5-D11/D29 and D13-R6-D10/D29 substrates with LC11-RNase H1 and Ec-RNase H1. The 5'-end labeled D13-R4-D12/D29, D13-R5-D11/D29 and D13-R6-D10/D29 substrates were hydrolyzed by the enzymes in the presence of 10 mM $MgCl_2$ at 37°C for 15 min and the hydrolysates were separated on a 20% polyacrylamide gel containing 7 M urea. The concentration of the substrate was 1.0 μM. The concentration of the enzyme in the reaction mixture is indicated above each lane. The enzymes and substrates are also shown above the gel. The sequences of the D13-R4-D12, D13-R5-D11 and D13-R6-D10 strands around the cleavage sites are indicated at the left side of the gel. The 18th nucleotide is "A" (DNA) for D13-R4-D12 and "a" (RNA) for D13-R5-D11 and D13-R6-D10, and the 19th one is "A" for D13-R4-D12 and D13-R5-D11 and "a" for D13-R6-D10.

LC11-RNase H1 cleaved the D13-R4-D12/D29, D13-R5-D11/D29 and D13-R6-D10/D29 substrates at two sites (a15-a16 and a16-a17), single site (a16-a17), and two sites (a16-a17 and a17-a18), respectively (Fig. 4-6). However, the amount of the enzyme required to cleave 70% of the substrate containing four ribonucleotides was higher than that required to cleave other two substrates by approximately 100 fold. This result indicates that the ability of LC11-RNase H1 to cleave the substrate containing four ribonucleotides is reduced by 100 fold as compared to those to cleave the substrates containing five and six ribonucleotides. LC11-RNase H1 cleaved the RNA_{12}/DNA_{12} (R12/D12) substrate as efficiently as the D13-R5-D11/D29 and D13-R6-D10/D29 substrates (data not shown).

Hydrogen bonds between the protein and 2'-OH groups of the substrate are fully formed only when LC11-RNase H1 cleaves the D13-R5-D11/D29 and D13-R6-D10/D29 substrate at the sites mentioned above. Therefore, LC11-RNase H1 poorly cleaved the substrate containing four ribonucleotides, probably because only three nonconsecutive hydrogen bonds are formed between the protein and 2'-OH groups of the substrate. It efficiently cleaved the substrates containing five and six ribonucleotides at the sites mentioned above, probably because all four nonconsecutive hydrogen bonds are formed between the protein and 2'-OH groups of the substrate.

In contrast, Ec-RNase H1 cleaved all three substrates with comparable efficiencies (Fig. 4-6). It also cleaved the R12/D12 substrate as efficiently as these substrates (data not shown). Ec-RNase H1 cleaved the DNA-RNA-DNA/DNA substrates containing four, five, and six ribonucleotides at single (a15-a16), two (a15-a16 and a16-a17), and three (a15-a16, a16-a17, and a17-a18) sites, respectively. Ec-RNase H1 efficiently cleaved these substrates at these sites, probably because all four consecutive hydrogen bonds are formed between the protein and 2'-OH groups of the substrate. These results strongly suggest that LC11-RNase H1 forms hydrogen bonds with four nonconsecutive 2'-OH groups of the RNA strand and therefore the structure of the LC11-D77N complex determined in a metal-free form may not be seriously different from the "productive complex".

4.3.6. Comparison with structure in substrate-free form

Comparison of the structure of LC11-D77N in a substrate-bound form with that in a substrate-free form shows that these structures are virtually identical with each other with the RMSD value of 1.1 Å for 136 Cα atoms (Fig. 4-7A). No significant difference is observed in the steric configurations of the active-site residues as well. The major difference is observed in the loop between αA and β4 (6.7 Å shift at the Cα atom of Lys66), which is the most disordered part in the electron density reflecting its intrinsic flexibility. Other differences are

observed in the loop between β1 and β2 (2.7 Å shift at the Cα atom of Asn15), and the loop between αB and αD (2.8 Å shift at the Cα atom of Lys92). As a result of these shifts, Ala12, Gly14, Asn45, and Glu79 move closer to and make contacts with the RNA strand, whereas Asn15, Gln84, Lys92, and Glu93, which make contacts with the DNA strand, move away from the substrate to avoid of the steric clash (Fig. 4-7B).

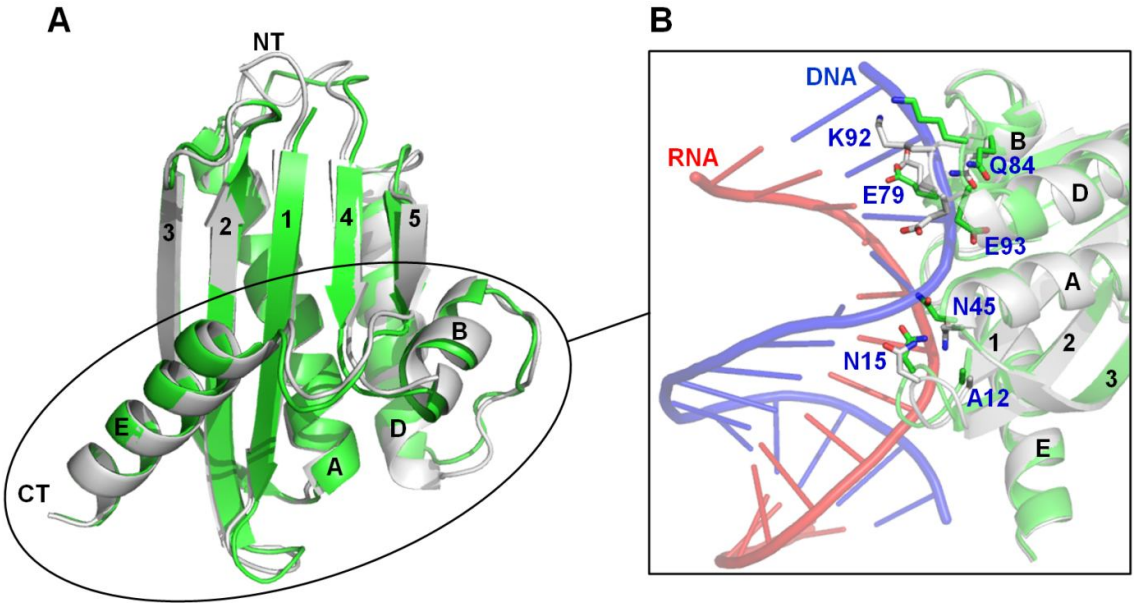

Figure 4-7. Comparison of the structures of LC11-RNase H1 in substrate-bound and substrate-free forms. The structure of the substrate-bound form of LC11-D77N is superimposed onto that of the substrate-free form of LC11-RNase H1 (PDB code 3U3G). The substrate-bound and substrate-free forms are colored green and grey, respectively. **(A)** Superimposition of the overall structures of both forms. **(B)** The displacement of the amino acid residues interacting with the RNA/DNA hybrid. The side chains of the amino acid residues interacting with the RNA/DNA hybrid in the substrate-bound form (green) and those of the corresponding residues in the substrate-free form (grey) are shown by stick models, in which the oxygen and nitrogen atoms are colored red and blue, respectively.

3.7. Modeling of Sto-RNase H1 complex

Despite a high structural similarity, LC11-RNase H1 and Sto-RNase H1 differ in their abilities to cleave dsRNA. LC11-RNase H1 does not cleave dsRNA, whereas Sto-RNase H1 cleaves it. On the assumption that the DNA-binding groove of Sto-RNase H1 is altered as

compared to that of LC11-RNase H1, such that it can accommodate not only the DNA strand of RNA/DNA hybrid but also the RNA strand of dsRNA, a model for the complex between Sto-RNase H1 and RNA/DNA hybrid was constructed based on the structure of the LC11-D77N complex (Fig. 4-8A). This model is more valid than that previously constructed based on the structure of the Bh-RNase HC* complex, because the conformation of the substrate and the interface between the protein and substrate in the LC11-D77N complex are different from those in the Bh-RNase HC* complex as mentioned above.

According to this model, Sto-RNase H1 also interacts with three consecutive phosphate groups of the DNA strand (D_{-3}-D_{-5}), as does LC11-D77N (Fig. 4-8B). However, the first phosphate-binding site is formed only by Thr47 and Asn49, which are conserved as Thr44 and Asn46 in LC11-RNase H1. The extra phosphate ligand (Glu93 via a water molecule) in LC11-RNase H1 is absent in Sto-RNase H1, because Glu93 in LC11-RNase H1 is replaced by Ala (Ala92) in Sto-RNase H1. It has been suggested that the absence of this extra phosphate ligand accounts for the capability of HIV-1 RNase H to cleave dsRNA [48]. To examine whether Sto-RNase H1 loses an ability to cleave dsRNA by mutating Ala92 and Lys93 to the corresponding residues of LC11-RNase H1 (Glu93 and Pro94, respectively) and LC11-RNase H1 acquires this ability by mutating Glu93 to the corresponding residue of Sto-RNase H1 (Ala92), the double mutant protein of Sto-RNase H1, Sto-A92E/K93P, and single mutant protein of LC11-RNase H1, LC11-E93A, were constructed. The dsRNase and RNase H activities of these mutant proteins were determined using the R12/R12 and R12/D12 substrates, respectively. However, both activities of Sto-A92E/K93P were rather slightly higher than those of Sto-RNase H1 (Fig. 4-9). Likewise, LC11-E93A did not exhibit dsRNase activity, but exhibits comparable RNase H activity to that of LC11-RNase H1 (data not shown). These results suggest that the extra phosphate ligand plays no role in the specific cleavage of RNA/DNA hybrids.

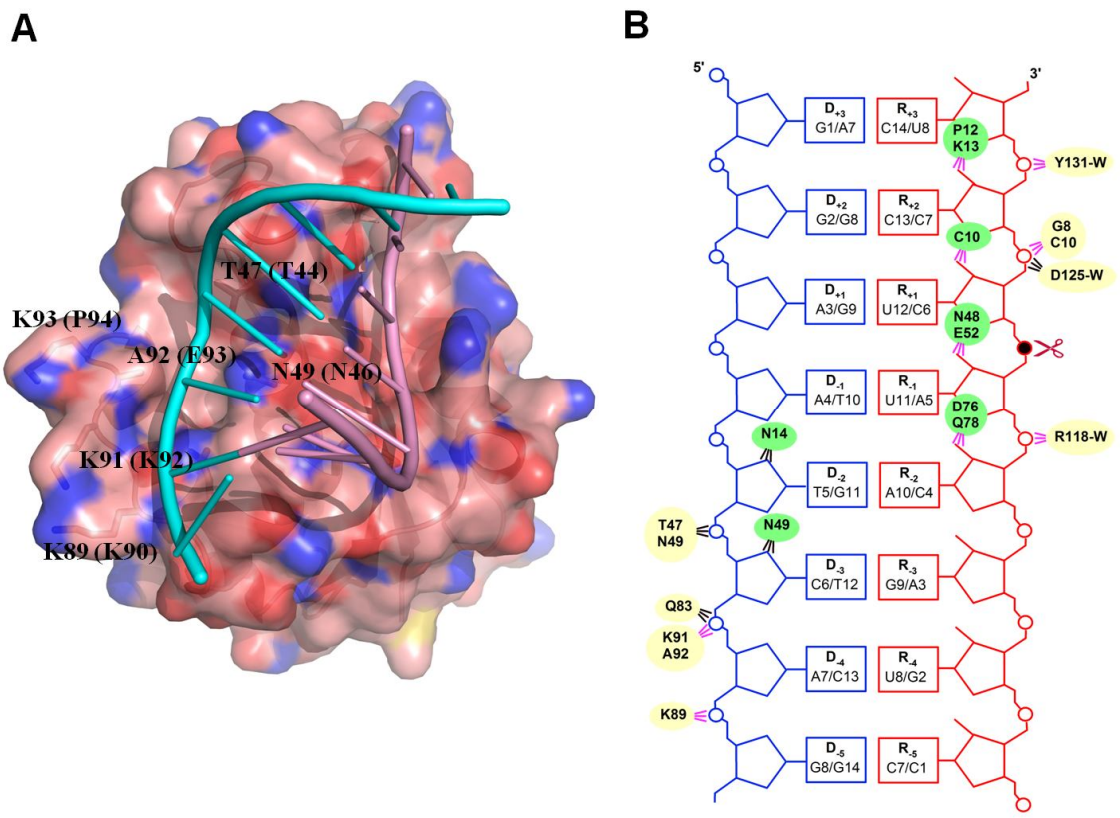

Figure 4-8. Model of Sto-RNase H1-substrate complex. This model was constructed by superimposing the Sto-RNase H1 structure onto the cocrystal structure of LC11-D77N with the substrate. **(A)** The molecular surface of Sto-RNase H1 in complex with the RNA/DNA hybrid. The oxygen, nitrogen, and sulfur atoms on the protein surface are colored red, blue, and yellow, respectively. The positions of Thr47, Asn49, Gln83, Lys89, Lys91, Ala92, and Lys93 are indicated. The corresponding residues in LC11-RNase H1 are also indicated in parentheses. **(B)** Possible interactions between the RNA/DNA hybrid and Sto-RNase H1. The residues interacting with phosphate and sugars are shown in yellow and green highlight, respectively. Pink lines indicate interactions made by the protein backbone, and black lines indicate interactions made by side chains. Water-mediated interactions are indicated by -W.

The second phosphate-binding site is formed by Gln83/Lys91/Ala92. Of these residues, Gln83 and Lys91 are conserved as Gln84 and Lys92 in LC11-RNase H1, whereas Ala92 is replaced by Glu (Glu93) in LC11-RNase H1. However, this replacement may not seriously affect the interaction with the phosphate group, because the main chain atoms contact the phosphate group in this site. At the third phosphate-binding site, Lys89 contacts the D_{-5}

phosphate, as Lys90 of LC11-RNase H1 does. Therefore, the interaction between Sto-RNase H1 and the DNA strand may not be significantly different from that between LC11-RNase H1 and the DNA strand. One possible structural basis for the difference in substrate specificity between Sto-RNase H1 and LC11-RNase is the absence of a clear DNA-binding groove on the surface of Sto-RNase H1. Lys92 of LC11-RNase H1 forms a clear hump to sepatate inside and outside of the DNA-binding channel (Fig. 4-4A), whereas the corresponding residue of Sto-RNase H1, Lys91, does not form it (Fig. 4-8A).

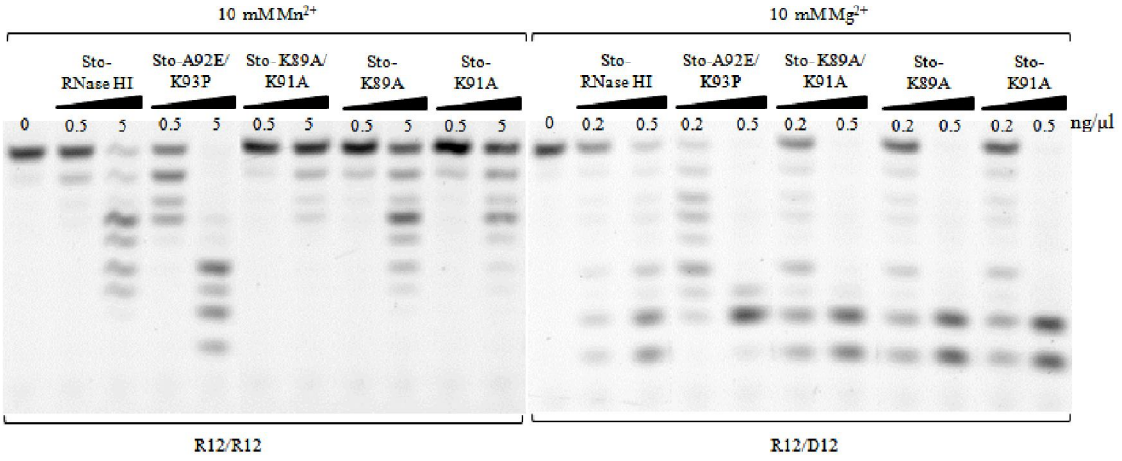

Figure 4-9. Cleavage of the R12/R12 and R12/D12 substrates with Sto-RNase H1 mutants. The 5'-end labeled R12/R12 and R12/D12 substrates were hydrolyzed by the enzyme in the presence of 10 mM $MnCl_2$ and $MgCl_2$, respectively, at 37°C for 15 min, and the hydrolysates were separated on a 20% polyacrylamide gel containing 7 M urea. The concentration of the substrate was 1.0 □M. The concentration of the enzyme in the reaction mixture is indicated above each lane. The sequence of the R12 strand is shown along the gel.

4.3.8. Importance of Lys89 and Lys91 for dsRNase activity of Sto-RNase H1

To examine whether LC11-RNase H1 acquires an ability to cleave dsRNA by removing the hump formed by Lys92, the mutant protein LC11-K92A, in which Lys92 of LC11-RNase H1 is replaced by Ala, was constructed. However, LC11-K92A did not exhibit dsRNase activity, but exhibited comparable RNase H activity to that of LC11-RNase H1 (Fig. 4-10).

The hump formed by Lys92 may be removed by the mutation of Lys92 to Ala, but LC11-RNase H1 does not acquire dsRNase activity by this mutation, probably because the absence of a clear DNA-binding groove is not sufficient to alter the substrate specificity of LC11-RNase H1, so that LC11-RNase H1 accommodates dsRNA. The presence of positively charged residues may also be important for dsRNase activity of Sto-RNase H1.

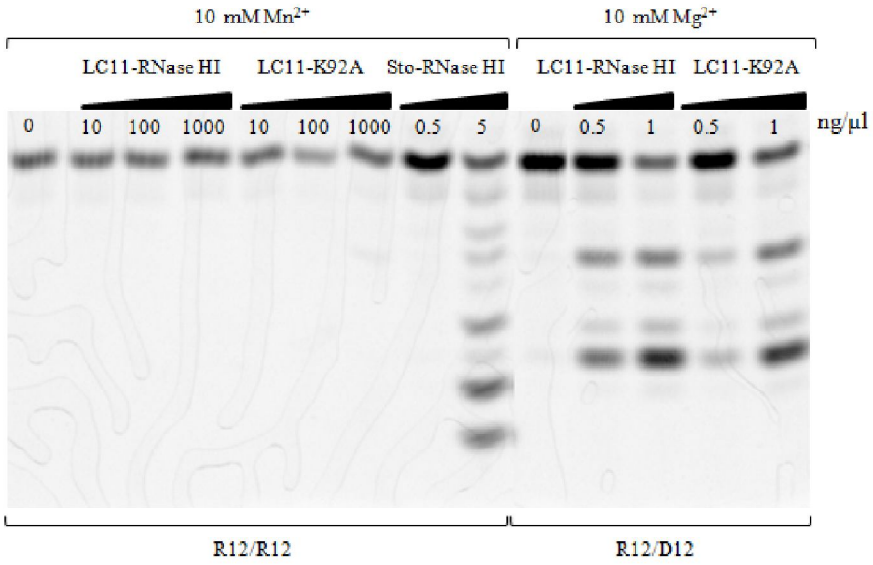

Figure 4-10. Cleavage of the R12/R12 and R12/D12 substrates with LC11-RNase H1 mutants. The 5'-end labeled R12/R12 and R12/D12 substrates were hydrolyzed by the enzyme in the presence of 10 mM $MnCl_2$ and $MgCl_2$, respectively, at 37°C for 15 min, and the hydrolysates were separated on a 20% polyacrylamide gel containing 7 M urea. The concentration of the substrate was 1.0 □M. The concentration of the enzyme in the reaction mixture is indicated above each lane. The sequence of the R12 strand is shown along the gel.

To confirm the importance of positively charged residues near the non-cleavable RNA-binding site for dsRNase activity, the mutant proteins of Sto-RNase H1, Sto-K89A, Sto-K91A, and Sto-K89A/K91A, in which Lys89 and/or Lys91 of Sto-RNase H1 are replaced by Ala, were constructed. The dsRNase activities of Sto-K89A, Sto-K91A, and Sto-K89A/K91A were reduced by approximately 30%, 34%, and 68%, respectively as compared to that of Sto-RNase H1, whereas their RNase H activites were nearly identical

with that of Sto-RNase H1 (Fig. 4-9). These results indicate that Lys89 and Lys91 are important for dsRNase activity of Sto-RNase H1. Both Lys89 and Lys91 of Sto-RNase H1 are conserved as Lys90 and Lys92 in LC11-RNase H1, respectively. Nevertheless, LC11-RNase H1 cannot cleave dsRNA, indicating that local conformation of these residues are also important for dsRNase activity of Sto-RNase H1.

For better understanding of the dsRNA recognition mechanism of Sto-RNase H1, a model for the complex with dsRNA is more useful than that with RNA/DNA hybrid. However, when one strand of dsRNAs, for which the three-dimensional structure is available, is placed in the RNA binding groove of Sto-RNase H1, the other strand is always displaced from the protein. Therefore, it is likely that dsRNA is also distorted when it binds to Sto-RNase H1. Crystallization of Sto-RNase H1 in complex with dsRNA is now under study.

4.4. Summary

In this chapter, the crystal structure of LC11-RNase H1 in complex with an RNA/DNA substrate was determined. Unlike Bh-RNase HC and Hs-RNase HC, LC11-RNase H1 interacts with four non-consecutive 2'-OH groups of the RNA strand. The lack of interactions with four consecutive 2'OH groups leads to a dramatic decrease in the ability of LC11-RNase H1 to cleave the DNA-RNA-DNA/DNA substrate containing four ribonucleotides as compared to those to cleave the substrates containing five and six ribonucleotides. The interaction of LC11-RNase H1 with the DNA strand is also different from those of Bh-RNase HC and Hs-RNase HC. Beside the common phosphate-binding pocket, LC11-RNase H1 has a unique DNA-binding channel. Furthermore, the active-site residues of LC11-RNase H1 are located farther away from the scissile phosphate group than those of Bh-RNase HC and Hs-RNase HC. Modeling of Sto-RNase H1 in complex with the 14 bp RNA/DNA substrate, together with the structure-based mutational analyses, suggest that the ability of Sto-RNase H1 to cleave double-stranded RNA is dependent on the local conformation of the basic residues located at the DNA binding site.

CHAPTER 5

Crystal structure of metagenome-derived LC9-RNase H1 with atypical DEDN active-site motif

5.1. Introduction

LC9-RNase H1 is the most unique protein among twelve RNases H1 isolated from leaf-branch compost by the metagenomic approach (Chapter 2). LC9-RNase H1 lacks two of the four acidic residues that are fully conserved in almost every RNase H1 sequence and form the active sites of these RNases H1. In the amino acid sequence of LC9-RNase H1, the second and fourth residues of the DEDD active-site motif are replaced by Ala and Asn, respectively as mentioned in Chapter 2. Then, the question arises whether the steric configurations of the active-site residues of LC9-RNase are different from those of other RNases H1 with a typical DEDD active-site motif.

In this chapter, I determined the crystal structure of LC9-RNase H1. Based on the structure-based mutational analyses, I identified four acidic active-site residues, including non-conserved glutamate and asparagine residues. I propose that LC9-RNase H1 represents bacterial RNases H1 with an atypical DEDN active-site motif, which are evolutionarily distinct from those with a typical DEDD active-site motif.

5.2. Materials and methods

5.2.1. Plasmid construction, overproduction, and purification

Plasmid pET-LC9 for overproduction of LC9-RNase H1 was constructed by amplifying the gene encoding LC9-RNase H1 by PCR using pBR322-LC9 (Chapter 2) as a template and ligating it into the *Nde*I-*Bam*HI sites of pET25b (Novagen). Plasmids pET-D22A, -E23A, -D96A and -N159A for overproduction of LC9-D22A, -E23A, -D96A and -N159, respectively, were constructed by PCR using the KOD-Plus Mutagenesis kit (Toyobo) according to the manufacturer's instructions. Plasmid pET-LC9 was used as a template. All DNA oligomers for PCR were synthesized by Hokkaido System Science. PCR was performed with a GeneAmp PCR system 2400 (Applied Biosystems). The DNA sequences were confirmed by a Prism 310 DNA sequencer (Applied Biosystems).

Overproduction of LC9-RNase H1 and its derivatives using *E. coli* BL21(DE3) transformants, disruption of the cells by sonication lysis, and centrifugation were performed as described previously [50]. The proteins were purified by cation exchange chromatography using a HiTrap SP column (GE Healthcare) equilibrated with 20 mM sodium phosphate (pH 6.8), followed by hydrophobic interaction chromatography using a Hitrap Phenyl HP column (GE Healthcare) equilibrated with 50 mM sodium phosphate (pH 7.0) containing 1 M $(NH_4)_2SO_4$, and stored in 10 mM sodium acetate (pH 4.5).

The purity of the protein was analyzed by SDS/PAGE as mentioned in previous chapters. Gel filtration chromatography was performed using a TSK-GEL G3000SW$_{XL}$ column (Sigma-Aldrich) equilibrated with 50 mM Tris-HCl (pH 7.5) containing 0.5 M NaCl. The protein concentration was determined from UV absorption using an A_{280} value of 2.2 for LC9-RNase H1 and its derivatives for a 0.1% (0.1 mg·ml^{-1}) solution. These values were calculated by using ε of 1576 $M^{-1}\cdot cm^{-1}$ for Tyr and 5225 $M^{-1}\cdot cm^{-1}$ for Trp at 280 nm [111].

5.2.2. Circular dichroism (CD) spectra

The far-UV CD spectra were measured on a J-725 spectropolarimeter (Japan Spectroscopic) at 25°C. The proteins were dissolved in 10 mM sodium acetate (pH 4.5). The protein concentration and optical path length were 0.1 $mg \cdot ml^{-1}$ and 2 mm, respectively. The mean residue ellipticity, θ, which has the units of $deg \cdot cm^{-2} \cdot dmol^{-1}$, was calculated by using an average amino acid molecular weight of 110.

5.2.3. Enzymatic activity

The RNase H activity was determined by measuring the amount of the acid-soluble digestion product from the substrate, ^{3}H-labeled M13 DNA/RNA hybrid, accumulated upon incubation at 37°C for 15 min, as described in Chapter 3. The reaction mixture contained 10 mM Tris-HCl (pH 8.5), 10 mM $MgCl_2$, 10 mM NaCl, 1 mM 2-mercaptoethanol, 50 $\mu g \cdot ml^{-1}$ bovine serum albumin, 1.0 μM M13 DNA/RNA hybrid (RNA nucleotide phosphate concentration), and appropriate amount of enzyme. One unit was defined as the amount of enzyme producing 1 μmol of acid-soluble material per min. The specific activity was defined as the enzymatic activity per milligram of protein.

For cleavage of the oligomeric substrates, R12/D12, R12/R12, D15-R1-D13/D29, 29 b RNA and 29 b DNA were prepared as described in Chapter 3. Hydrolysis of the substrate at 37°C for 15 min and separation of the products on a 20% polyacrylamide gel containing 7 M urea were carried out as described previously. The products were detected by Typhoon 9240 Imager (GE Healthcare). The reaction buffers were the same as those for the hydrolysis of the M13 DNA/RNA hybrid. The products were identified by comparing their migration on the gel.

5.2.4. Crystallization

The crystallization conditions were initially screened using crystallization kits from

Hampton Research (Crystal Screens I and II) and Emerald Biostructures (Wizard I and II). The conditions were surveyed using sitting-drop vapour-diffusion method at 4°C. Drops were prepared by mixing the protein solution at the concentration of 7 $mg \cdot ml^{-1}$ with the reservoir solution at equal volume (1 µl), and were vapour-equilibrated against 100 µl reservoir solution. Native LC9-RNase H1 crystals appeared after two weeks in Crystal Screens II No. 6 (100 mM Tris-HCl pH 8.5, 200 mM $MgCl_2.6H_2O$ and 30% PEG 4000). The crystallization conditions were further optimized and crystals were obtained when the drop was prepared by mixing 1.5 µl protein solution at the concentration of 8.5 $mg \cdot ml^{-1}$ with 1 µl reservoir solution (100 mM Tris-HCl pH 8.5, 200 mM $MgCl_2.6H_2O$ and 30% PEG 4000) at 4°C. Crystals grew to final size (0.3 - 0.4 mm) in 6 - 8 weeks. Since the crystal was obtained in the reservoir solution containing 30% PEG 4000, it was used directly for diffraction data collection without further cryoprotection.

5.2.5. X-ray diffraction data collection and structure determination

X-ray diffraction data set of LC9-RNase H1 was collected at a wavelength of 0.9 Å at -173°C using synchrotron radiation on the BL44XU station at SPring-8 (Hyogo, Japan). The data set was indexed, integrated and scaled using the program HKL2000 [112]. The structure was solved by the molecular replacement method using MOLREP [113] in the CCP4 program suite. The crystal structure of Sto-RNase H1 (Protein Data Bank entry 2EHG) was used as a starting model. Automatic model building and also automatic placement of water molecules were then done using Arp/wArp [122], followed by manually checking on correct assignment. Iteration between model refinement using REFMAC5 [123] and manual adjustments using Coot [121] was done until no further decrease of the R_{free} was observed. The statistics for data collection and refinement are summarized in Table 5-1. The figures were prepared using PyMol (http://www.pymol.org).

Table 5-1. Data collection and refinement statistics

Data collection	
Wavelength (Å)	0.9
Space group	$P2_1$
Unit cell	
a, b, c (Å)	41.6, 39.9, 50.3
α, β, γ (°)	90.0, 102.6, 90.0
Resolution (Å)	50.00 – 1.62 (1.65 – 1.62)
Observations	305097
Unique reflections	20537
Completeness (%)	99.2 (96.2)[a]
R_{merge} (%)[b]	7.7 (34.9)[a]
Average I/s (I)	43.6 (4.8)[a]
Refinement statistics	
Resolution limits (Å)	49.04 – 1.62
Number of atoms	
Protein	1613
RNA/DNA	153
R_{work} (%)	17.4
R_{free} (%)[c]	21.6
Deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.027
Overall mean B factors (Å^2)	25.89
Ramachandran analysis (%)	
Favored regions	99.04
Allowed regions	0.96

[a] Values in parentheses are for the highest resolution shell. [b] $R_{merge} = \sum | I_{hkl} - \langle I_{hkl} \rangle | / \sum I_{hkl}$, where I_{hkl} is the intensity measurement for reflection with indices hkl and $\langle I_{hkl} \rangle$ is the mean intensity for multiply recorded reflections. [c] R_{free} was calculated using 5 % of the total reflections chosen randomly and omitted from refinement.

5.2.6. Protein Data Bank accession number

The coordinates and structure factors for LC9-RNase H1 have been deposited in the Protein Data Bank under accession code 4IBN.

5.3. Results and discussion

5.3.1. Protein preparation

Upon overproduction, LC9-RNase H1 accumulated in the cells in a soluble form. It was purified to give a single band on SDS/PAGE. The amount of LC9-RNase H1 purified from 1 L of culture was approximately 3 mg. The molecular mass of LC9-RNase H1 was estimated to be 24 kDa by SDS/PAGE and 21 kDa by gel filtration chromatography. These values are comparable to that calculated from the amino acid sequence (23.8 kDa), indicating that LC9-RNase H1 exists in a monomeric form, like other type 1 RNases H1.

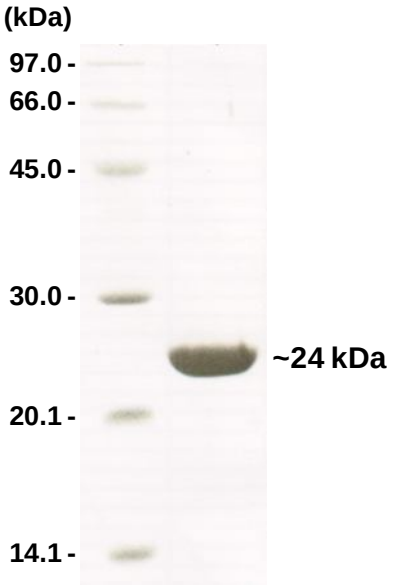

Figure. 5-1. SDS/PAGE analyses for the purified protein of LC9-RNase H1

5.3.2. Crystal structure

The crystal structure of LC9-RNase H1 was determined at 1.6 Å resolution. The N-terminal nine residues are not visible in the electron density map, possibly due to a structural disorder. Details of the data-collection statistics and refinement are summarized in Table 5-1. The overall structure of LC9-RNase H1 is highly similar to that of Ec-RNase H1 (PDB code 1RNH) with the RMSD value of 1.50 Å for 127 Cα atoms (Fig. 5-2). It shares a five-stranded mixed β-sheet with three antiparallel (β1- β3) and two parallel (β4 and β5) strands and five α-helices (αA, αB, αC, αD and αE) with Ec-RNase H1. However, LC9-RNase H1 has a longer C-terminal tail than Ec-RNase H1. This tail (Gln169-Ser220), which is extended from the C-terminus of αE-helix and contains two additional helices (αF and αG), is longer than that of Ec-RNase H1 by 40 residues. This tail mostly assumes an

extended structure and is anchored to the core region through hydrogen bonds and hydrophobic interactions. According to a model of the LC9-RNase H1-substrate complex constructed based on the structure of the catalytic domain of Hs-RNase HC* in complex with the substrate [49] suggests that RNA/DNA hybrid binds to the protein such that the RNA strand contacts the active site and the DNA strand contacts αC-helix and the following loop. Because, the C-terminal tail is apparently located far from the substrate binding site, it may not be important for enzymatic activity. However, further mutational and structural studies will be required to understand the role of this tail.

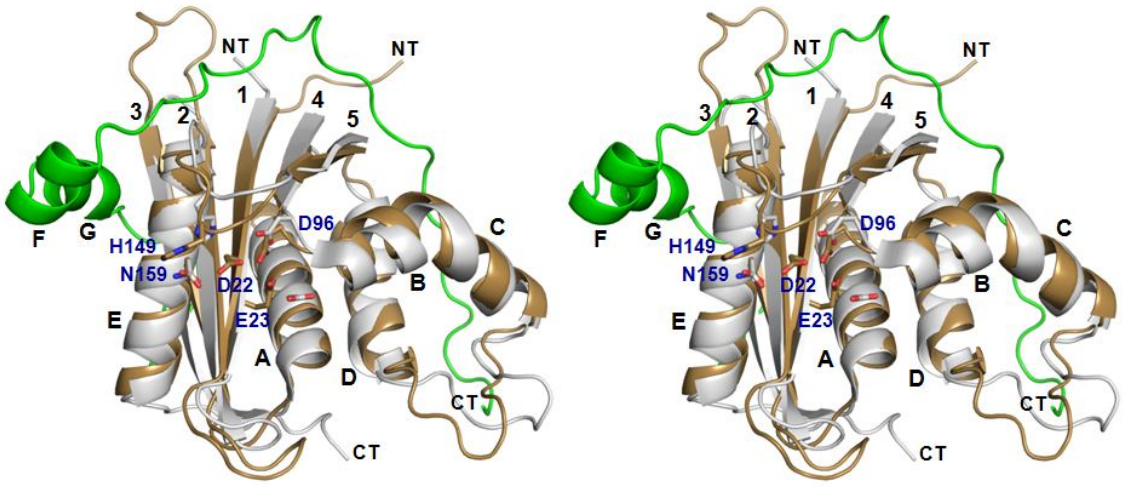

Figure 5-2. Stereoview of the structure of LC9-RNase H1 superimposed on that of Ec-RNase H1 (PDB code 1RNH). The LC9-RNase H1 and Ec-RNase H1 structures are colored brown and grey, respectively. Five active-site residues of LC9-RNase H1 (Asp22, Glu23, Asp96, His149, and Asn159) and Ec-RNase H1 (Asp10, Glu48, Asp70, His124, and Asp134) are shown by stick models, in which the oxygen and nitrogen atoms are colored red and blue, respectively. The C-terminal tail of LC9-RNase H1 (Gln169-Ser220) is colored green. NT and CT represent N- and C-termini.

The amino acid sequence of LC9-RNase H1 is compared with those of Ec-RNase H1 and Hs-RNase HC on the bases of their crystal structures in Figure 5-3. The sequence of LC9-RNase H1 is 37.4 and 31.8% identical to those of Ec-RNase H1 and Hs-RNase HC, respectively.

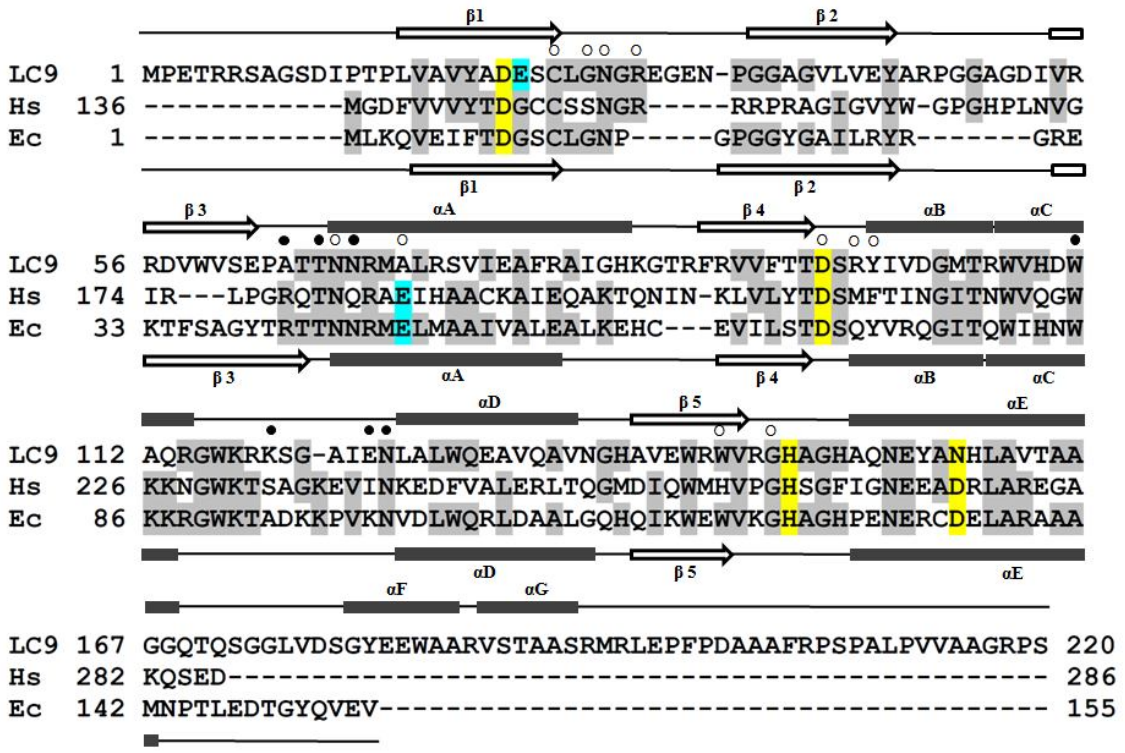

Figure 5-3. Alignment of the amino acid sequences. The amino acid sequences of LC9-RNase H1 (LC9), Ec-RNase H1 (Ec), and Hs-RNase HC (Hs) are compared with one another on the basis of their crystal structures. The ranges of the secondary structures of LC9-RNase H1 and Ec-RNase H1 are shown above and below the sequences, respectively. The amino acid residues that are conserved in at least two different proteins are highlighted in grey. The second glutamate residues of a DEDN or DEDD active-site motif are highlighted in cyan, while the other active-site residues are highlighted in yellow. The amino acid residues that contact the RNA and DNA strands of the substrate in the cocrystal structure of Hs-RNase HC* with the substrate are indicated by the open and closed circles above the sequences, respectively. Gaps are denoted by dashes. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

5.3.3. Active site

The four conserved acidic residues (Asp10, Glu48, Asp70, and Asp134 for Ec-RNase H1, and Asp145, Glu186, Asp210, and Asp274 for Hs-RNase H1) form the active sites of Ec-RNase H1 and Hs-RNase H1 (Fig. 1-8). In contrast, the two conserved acidic residues, Asp22 and Asp96, one non-conserved glutamate residue (Glu23), and one non-conserved

asparagines residue (Asn159) form the active site of LC9-RNase H1 (Fig. 5-2, Fig. 5-3). Superimposition of the active-site structures of LC9-RNase H1 and Ec-RNase H1 on the cocrystal structure of Hs-RNase HC* with the substrate and metal ions [49] indicates that the steric configurations of the four active-site residues of LC9-RNase HI are very similar to those of Ec-RNase H1 and Hs-RNase HC* (Fig. 5-4). Therefore, these four residues of LC9-RNase H1 probably form the metal binding sites A and B, as seen in the cocrystal structure of Hs-RNase HC*. In addition, His149 of LC9-RNase H1 is located at the similar position where His124 of Ec-RNase H1 and His264 of Hs-RNase HC* are located (Fig. 5-4). This residue may facilitate product release by dislodging the 5'-phosphate, as proposed for His264 of Hs-RNase H1 [49]. It is noted that the first and second active-site residues of LC9-RNase H1 (Asp22 and Glu23) are located adjacently in the same β-strand (β1), whereas the corresponding residues of Ec-RNase H1 (Asp10 and Glu48) and Hs-RNase H1 (Asp145 and Glu186) are located distantly in β1-strand and αA-helix, respectively (Fig. 5-3). Similar

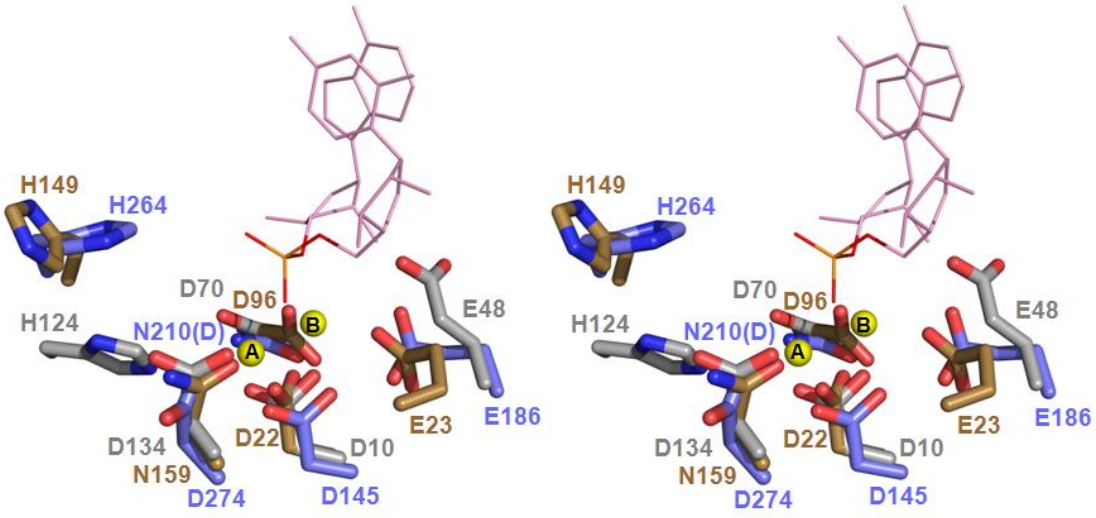

Fig. 5-4. Stereoview of the active-site structures of RNases H. The side chains of the active-site residues of LC9-RNase H1 (brown) and Ec-RNase H1 (grey) are superimposed onto those in the cocrystal structure of Hs-RNase HC* with the 14 bp RNA/DNA substrate and metal ions (slate, PDB code 2QKK). The positions of the RNA strand of the substrate with the scissile phosphate group and two metal ions A and B are also shown.

arrangement of the active-site residues is also observed in the amino acid sequences of RNases H2, in which the first (Asp) and second (Glu) active-site residues are located adjacently in the same β-strand [2,68]. However the steric configurations of the active-site residues of LC9-RNase H1 are more similar to those of RNases H1 than to those of RNases H2.

5.3.4. Enzymatic activity of LC9-RNase H1

The optimum condition for activity of LC9-RNase H1 was analyzed at 37°C by changing one of the conditions used for assay [10 mM Tris-HCl (pH 8.5), 10 mM NaCl, and 10 mM $MgCl_2$]. The M13 DNA/RNA was used as a substrate. LC9-RNase H1 exhibited activity at the wide pH range between 6 and 11 with the maximum at pH 10 (Fig. 5-5), like Ec-RNase H1 [125,126]. Nevertheless, the activity was determined at pH 8.5, at which LC9-RNase H1

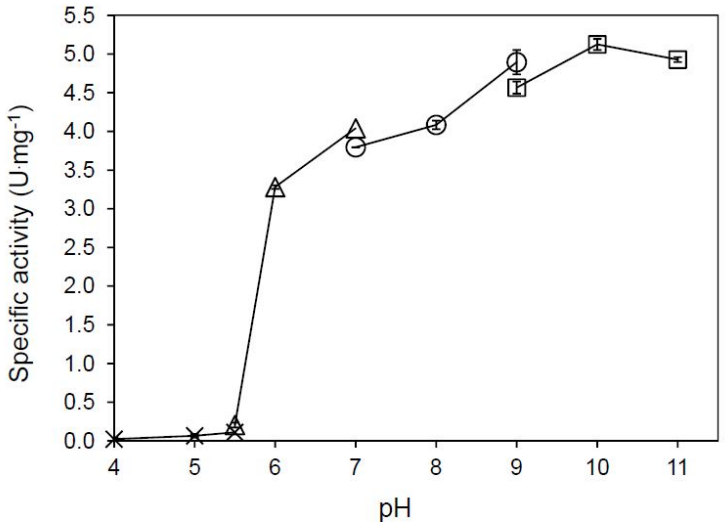

Fig. 5-5. pH dependency of LC9-RNase H1. The enzymatic activity of LC9-RNase H1 was determined in the presence of 10 mM $MgCl_2$ at 37°C using M13 DNA/RNA hybrid as a substrate. The buffers used for assay were 10 mM sodium acetate (pH 4.0-5.5) (cross), 10 mM 2-(*N*-morpholino)ethanesulfonic acid (MES)-NaOH (pH 5.5-7.0) (open triangle), 10 mM Tris-HCl (pH 7.0-9.0) (open circle), and *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS)-NaOH (pH 9.0-11.0) (open square). Experiments were carried out at least twice and the average values are shown together with the errors.

exhibits approximately 80% of the maximal activity, because the stability of the substrate and the solubility of the metal cofactor decrease as the pH increases beyond 9.0. LC9-RNase H1 preferred Mg^{2+} to Mn^{2+} ions for activity (Fig. 5-6), like Ec-RNase H1 [126]. It exhibited maximal activity at 10 mM $MgCl_2$ and 0.1 mM $MnCl_2$. The specific activities of LC9-RNase H1 and Ec-RNase H1 were determined to be 4.3 and 12 units/mg, respectively, in the presence of 10 mM $MgCl_2$, and 0.4 and 1.1 units/mg, respectively, in the presence of 0.1 mM $MnCl_2$.

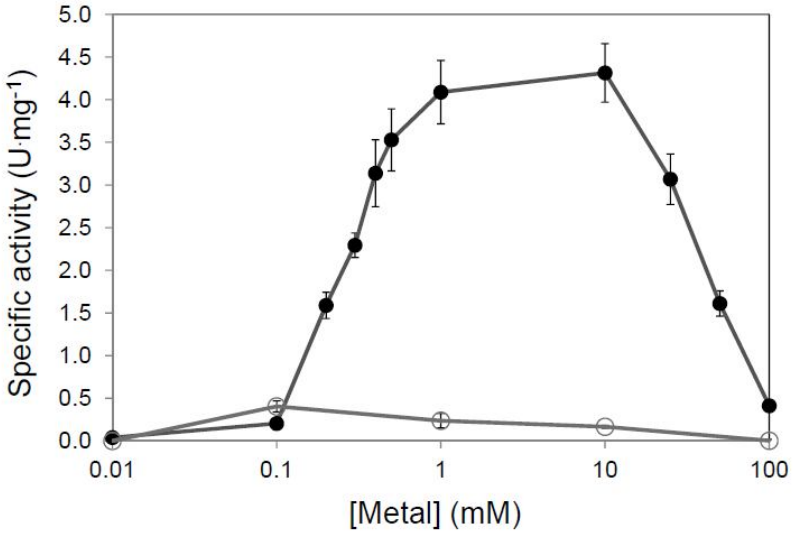

Fig. 5-6. Metal dependency of LC9-RNase H1. The enzymatic activity of LC9-RNase H1 was determined at 37°C in 10 mM Tris-HCl (pH 8.5) containing various concentrations of $MgCl_2$ (closed circle) and $MnCl_2$ (open circle) using M13 DNA/RNA hybrid as a substrate. Experiments were carried out at least twice and the average values are shown together with the errors.

The substrate and cleavage-site specificities of LC9-RNase H1 were analyzed by using 12 bp RNA/DNA hybrid (R12/D12), 12 bp RNA/RNA duplex (R12/R12), 29 b RNA, 29 b DNA, and 29 bp DNA_{15}-RNA_1-DNA_{13}/DNA duplex (D15-R1-D13/D29), and compared with those of Ec-RNase H1. D15-R1-D13 is a 29 b DNA containing a single ribonucleotide. LC9-RNase H1 cleaved the R12/D12 substrate at multiple sites, most preferably at a9-c10, in the presence of 10 mM $MgCl_2$, as did Ec-RNase H1 (Fig. 5-7). Both enzymes did not cleave

other substrates (data not shown), indicating that they do not cleave dsDNA, single ribonucleotide-embedded dsDNA, dsRNA, ssRNA and ssDNA. Thus, the enzymatic properties of LC9-RNase H1 are similar to those of Ec-RNase H1, although its specific activity is lower than that of Ec-RNase H1 by 3 fold.

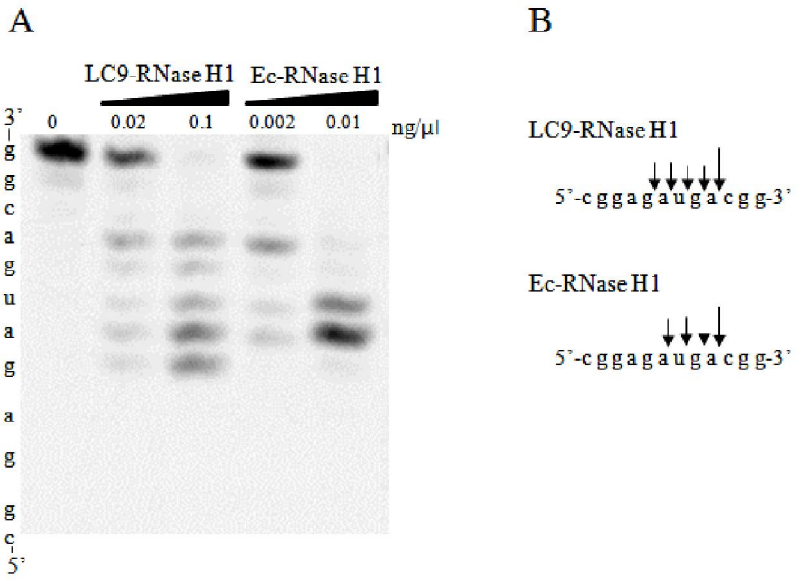

Figure 5-7. Cleavage of the 12 bp RNA/DNA substrate with LC9-RNase H1 and Ec-RNase H1. **(A)** Separation of the hydrolysates by urea gel. The 5'-end labeled 12 bp RNA/DNA substrate was hydrolyzed by the enzyme in the presence of 10 mM $MgCl_2$ at 37°C for 15 min and the hydrolysates were separated on a 20% polyacrylamide gel containing 7 M urea. The concentration of the substrate was 1.0 μM. The enzyme is indicated above each lane together with its concentration in the reaction mixture. The sequence of the RNA strand of the substrate is indicated along the gel. **(B)** Schematic representation of the sites and extents of cleavage by LC9-RNase H1 and Ec-RNase H1. Cleavage sites are shown by arrows. The differences in the lengths of the arrows reflect relative cleavage intensities at the position indicated. These lengths do not necessarily reflect the amount of the products accumulated upon complete hydrolysis of the substrate.

5.3.5. Mutation of active-site residues

To examine whether all four active-site residues are required for activity of LC9-RNase H1, I constructed the four mutant proteins, LC9-D22A, LC9-E23A, LC9-D96A, and LC9-N159A, in which Asp22, Glu23, Asp96, and Asn159 are replaced by Ala, respectively.

The far-UV CD spectra of all mutant proteins were nearly identical with that of LC9-RNase H1, indicating that none of the mutations does not significantly affect the tertiary structure of the protein (Fig. 5-8). LC9-D22A, LC9-E23A, and LC9-D96A exhibited the background level of activity either in the presence of Mg^{2+} or Mn^{2+} ions (Table 5-2), indicating that Asp22, Glu23, and Asp96 are required for activity. These residues are probably required for coordination with metal ion A and/or metal ion B. LC9-N159A also exhibited greatly decreased activity, but retained 2.5 and 20% of the activity of LC9-RNase H1 in the presence of Mg^{2+} and Mn^{2+} ions, respectively (Table 5-2), indicating that Asn159 is important for activity but is not required for activity. Dispensability of the corresponding aspartate residues for activity has been reported for several RNases H1 [46,59,68].

Table 2. Specific activities of LC9-RNase H1 and its variants[a]

Protein	Metal	Specific activity (U/mg)	Relative activity[b] (%)
LC9-RNase H1	$MgCl_2$	4.30	100
	$MnCl_2$	0.40	100
LC9-D22A	$MgCl_2$	$< 4 \times 10^{-4}$	< 0.01%
	$MnCl_2$	$< 4 \times 10^{-4}$	< 0.1%
LC9-E23A	$MgCl_2$	$< 4 \times 10^{-4}$	< 0.01%
	$MnCl_2$	$< 4 \times 10^{-4}$	< 0.1%
LC9-D96A	$MgCl_2$	$< 4 \times 10^{-4}$	< 0.01%
	$MnCl_2$	$< 4 \times 10^{-4}$	< 0.1%
LC9-N159A	$MgCl_2$	0.11	2.6
	$MnCl_2$	0.08	20

[a] The enzymatic activity was determined at 37°C and pH 8.5 in the presence of 10 mM $MgCl_2$ or 0.1 mM $MnCl_2$ using M13 DNA/RNA hybrid as a substrate. The experiment was carried out at least twice. Errors, which represent the 67% confidence limits, are all at or below ±20% of the values reported.

[b] The Mg^{2+}- and Mn^{2+}-dependent specific activities of the LC9-RNase H1 variants relative to those of LC9-RNase H1.

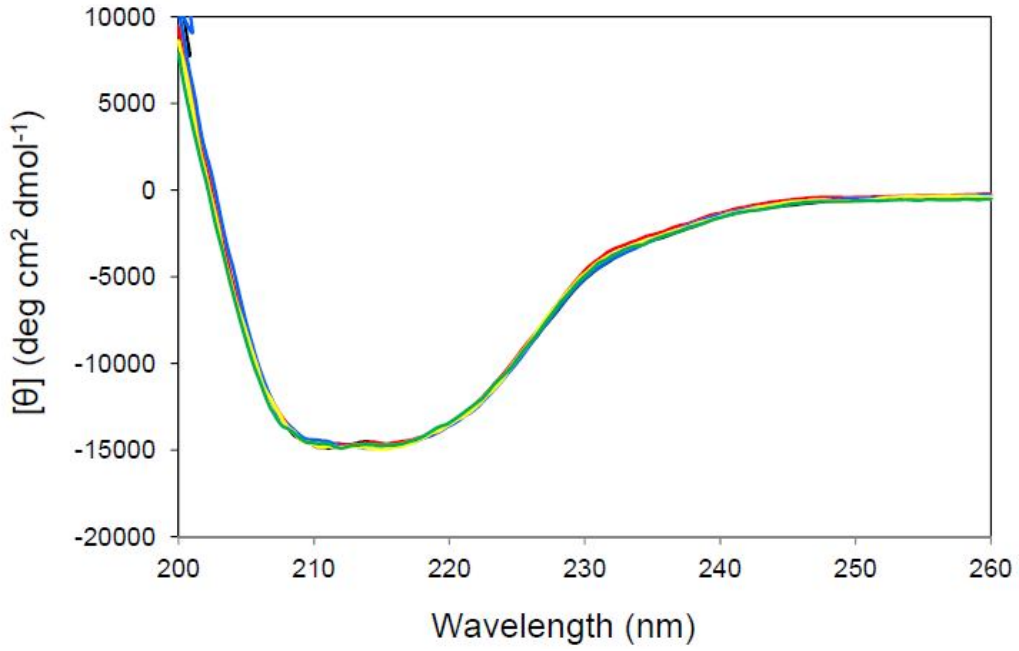

Fig. 5-8. CD spectra of LC9-RNase H1 and its variants. The far-UV CD spectra of LC9-RNase H1 (black), LC9-D22A (red), LC9-E23A (blue), LC9-D96A (yellow), and LC9-N159A (green) are shown. These spectra were measured at 25°C as described in Materials and methods.

5.3.6. LC9-RNase H1 homologs

Identification of Asp22, Glu23, Asp96, and Asn159 as the active-site residues of LC9-RNase H1 indicates that LC9-RNase H1 has an atypical DEDN active-site motif, in which the first and second residues are consecutive and the fourth acidic residue of a typical DEDD active-site motif is replaced by Asn. Database searches indicate that LC9-RNase H1 is not the only enzyme with this motif. RNase H1 from *Gemmatimonas aurantiaca* (Gau-RNase H1) (accession no. C1A5K0), which shows the amino acid sequence identity of 55% to LC9-RNase H1, also has this motif (Fig. 5-9) (Chapter 2). Phylogenetic analysis using representative members of bacterial RNases H1 indicates that a common ancestor of LC9-RNase H1 and Gau-RNase H1 is diverged from that of all bacterial RNases H1 before all other bacterial RNases H1 are diverged from it (Fig. 2-5). Because *G. aurantiaca* is a representative member of the bacterial phylum *Gemmatimonadetes* newly identified recently

[127], LC9-RNase H1 and its homologs with an atypical DEDN active-site motif may be functionally closely related to, but evolutionarily distantly related to other RNases H1 with a typical DEDD active-site motif.

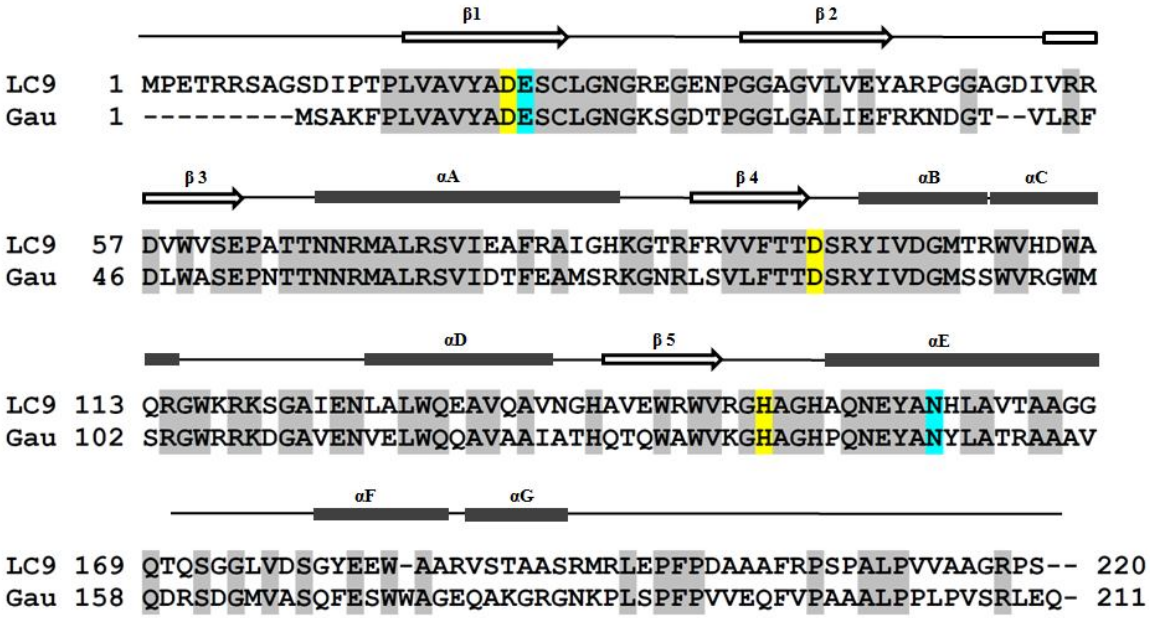

Figure 5-9. Alignment of the amino acid sequences of LC9-RNase H1 (LC9), and Gau-RNase H1 (Gau). The ranges of the secondary structures of LC9-RNase H1 are shown above. Other indications are described previously in Fig. 5-3.

5.4. Summary

The crystal structure of metagenome-derived LC9-RNase H1 was determined. LC9-RNase H1 lacks two of the four conserved acidic residues that form the active site of various RNases H. Nevertheless, the enzymatic properties of LC9-RNase H1 were similar to those of Ec-RNase H1. The structure-based mutational analyses showed that, in addition to the two conserved aspartate residues, non-conserved glutamate and asparagine residues form the active site of LC9-RNase H1. I propose that LC9-RNase H1 represents bacterial RNases H1 with an atypical DEDN active-site motif, which are evolutionarily distinct from those with a typical DEDD active-site motif.

CHAPTER 6

General discussion and future remarks

6.1. General discussion

This study broadens our knowledge on molecular diversities, catalytic mechanism and substrate recognition mechanism of RNase H.

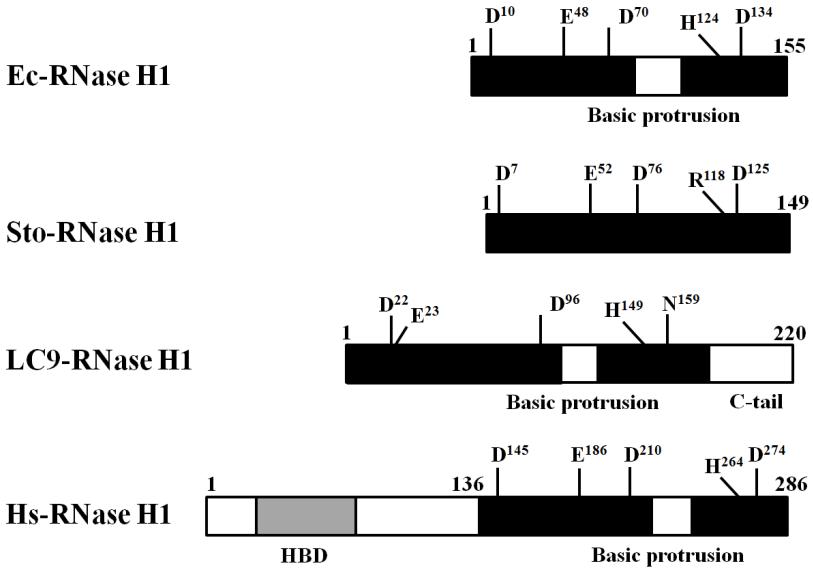

Figure 6-1. Schematic representation of the primary structures of Ec-, Sto-, LC9- and Hs-RNases H1. Solid box represents the RNase H domain and grey box represents the hybrid binding domain (HBD). The basic protrusion, which is located in the middle of the sequences of Ec-, LC9- and Hs-RNases H1, the C-terminal tail in LC9-RNase H1, and five active-site residues are shown. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

This is the first study to my knowledge to discover a novel RNase H with an atypical DEDN active-site motif, which may represent a new subfamily of RNases H1. Fig. 6-1 summarizes the difference in primary structure between LC9-RNase H1 with an atypical DEDN active-site motif and representative members of RNases H1 with a typical DEDD active-site motif. Ec-RNase H1 represents bacterial RNases H1, Sto-RNase H1 represents Sto-RNase H1 homologs and Hs-RNase H1 represents eukaryotic RNases H1. These RNases

H have a non-conserved active-site residue (His for Ec- and Hs-RNases H1 and Arg for Sto-RNase H1) besides the four conserved acidic active-site residues. LC9-RNase H1 differs from other RNases H1 in the position of the second residue and type of the fourth residue of a DEDD/N active-site motif. It also differs from other RNases H1 by the presence of a C-terminal extension. However, it is similar to Ec- and Hs-RNases H1 in that it has a basic protrusion and the histidie residue at the active site. Phylogenetic analyses indicate that LC9-RNase H1 and its homologs are evolutionarily distantly related to other RNases H1, suggesting that LC9-RNase H1 represents a new group of RNases H1.

A two-metal ion catalysis mechanism has been proposed for RNase H. When LC9-RNase H1 without a typical DEDD active-site motif was identified, I thought it may employ an alternative catalytic mechanism, because four conserved acidic active-site residues provide ligands for coordination of two metal ions according to this mechanism. However, determination of the crystal structure of LC9-RNase H1 indicates that a non-conserved glutamate residue acts as a substitute of the second glutamate residue of a DEDD active-site motif. Thus, not only RNases H1 with a typical DEDD active-site motif but also those with an atypical DEDN active-site motif probably employ a two-metal ion catalysis mechanism.

Numerous experiments on substrate specificity of Ec-RNase H1 have established that RNases H1 require at least four ribonucleotides to cleave the RNA/DNA substrate. This study, however, shows that LC11-RNase H1 requires at least five ribonucleotides to cleave the substrate efficiently. A complete formation of four hydrogen bonds between the RNase H protein and the 2'-OH groups of ribonucleotides of the substrate, either consecutively or non-consecutively (Fig. 6-2), are required for efficient cleavage of the substrate with the RNase H enzyme. Therefore, it may be more appropriate to describe that RNases H1 require at least 4 interactions with the 2'-OH groups of ribonucleotides to cleave the RNA/DNA hybrid.

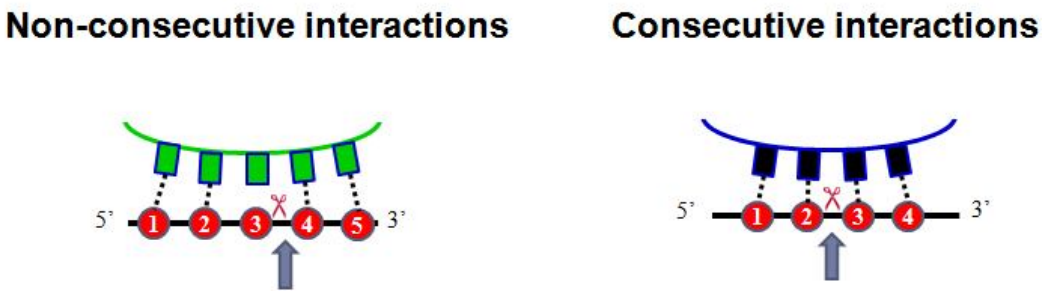

Figure 6-2. Schematic representation of the interactions between RNase H and the 2'-OH groups of ribonucleotides in the substrate. The red circles represent the ribonucleotides in the cleaved strand of the substrate, whereas either the green boxes or the black boxes represent the amino acid residues in the RNase H protein which interact with the 2'-OH groups of ribonucleotides in the substrate. The cleavage site is indicated by the arrow.

As far as I know, the structure of LC11-RNase H1 in complex with the RNA/DNA hybrid is the first structure of a full-length RNase H protein bound to the substrate. Similar to the structures of Bh- and Hs-RNases HC* in complex with the substrate, that of LC11-RNase H1-substrate complex shows that the RNA strand of the hybrid is recognized by the protein through extensive contacts with 2'-OH groups; and the DNA strand is recognized by its ability to adopt a B form conformation. However, beside the common phosphate-binding pocket, LC11-RNase H1 has a unique DNA-binding channel, which differs from the DNA-binding channel formed at the basic-protrusion of Hs-RNase HC in complex with the substrate. Noticeably, the relative distances of the active-site residues to the scissile phosphate group in the LC11-D77N complex increase as compared to those in the Bh- and Hs-RNase HC* complexes. They may account for the difference in metal ion preference between LC11-RNase H1 and Bh- or Hs-RNase HC.

It is still unclear how Sto-RNase H1 or HIV-RNase H regconizes and cleaves dsRNA, although the weak conservation of the residues in the DNA-binding site has been suggested to account for that ability. This study examines the structural features of the DNA-binding site that allow RNase H protein to cleave dsRNA. My findings show that some RNases H and

RNase H-fold containing proteins, which do not have a clear DNA binding groove on the protein surface but have positively charged residues near the DNA binding site, can cleave dsRNA. It is likely that both the presence of positively charged residues around the DNA binding site and their local conformations play important roles for the ability of RNase H to cleave dsRNA.

The discovery of a novel RNase H1, which may represent a new group of RNases H1, and a new Sto-RNase H1 homolog with unique substrate recognition mechanism, demonstrates the potential of metagenomic approach to increase our knowledge on the molecular diversity of a given protein and deepen our understanding on the structure-function relationships of proteins. Accumulation of this information will facilitate understanding of the physiological role of RNase H.

6.2. Future remarks

Although the important role of positively charged residues near the DNA-binding site to the dsRNase activity of RNase H is revealed in this study, the dsRNA recognition mechanism of RNase H still remains to be understood. For better understanding of the mechanism, crystallization of Sto-RNase H1 in complex with dsRNA is necessary. Likewise, the relative distances of the active-site residues to the scissile phosphate group is proposed to account for the difference in metal ion preference between LC11-RNase H1 and Bh- or Hs-RNase HC. However, further structural study of LC11-RNase H1-structure complex in the presence of metal ions, and structural studies of full-length Bh- or Hs-RNase H1 will be necessary to understand the structural basis for their differences in metal ion preference.

The crystal structure of LC9-RNase H1 reveals that LC9-RNase H1 differs from other RNases H1 by the presence of a C-terminal extension, which is anchored to the core region through hydrogen bonds and hydrophobic interactions. However the role of this extension

remains to be understood. Further mutational and structural studies will be required to understand the role of this extension.

Last but not least, the crystal structure of LC9-RNase H1 reveals its atypical DEDN active site, in which the second glutamate residue (Glu23) is shifted from the conserved position of all RNases H1 (the position of Ala71) to the position adjacent to the first aspartate residue, and the fourth conserved aspartate residue is replaced by Asn (Asn159). However, the remaining questions are (1) whether replacement of the fourth catalytic residue Asn by Asp, thus making the conventional DEDD, would render LC9-RNase H1 more active; and (2) how many mutations are necessary to revert LC9-RNase H1 to the conventional RNase H1, and whether double mutations E23G and A71E are sufficient. These mutational studies would be informative to understand whether RNases H1 with an atypical active site motif are diverged from those with a typical active site motif by Darwin's theory of evolution (mutations and selections). Thereby, it potentially improves our understanding of RNase H evolution, as well as general protein evolution.

References

1. Crouch, R.J., Dirksen, M.-L., 1982. Ribonucleases H. *In Nuclease* (Linn, S. M., & Roberts, R. J., ed.), pp. 211-241, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Ohtani, N., Haruki, M., Morikawa, M. & Kanaya, S. (1999) Molecular diversities of RNases H. *J Biosci Bioeng* 88, 12–19.
3. Tadokoro, T. & Kanaya, S. (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J* 276, 1482–1493.
4. Ohtani, N., Haruki, M., Morikawa, M., Crouch, R.J., Itaya, M. & Kanaya, S. (1999) Identification of the genes encoding Mn^{2+}-dependent RNase HII and Mg^{2+}-dependent RNase HIII from *Bacillus subtilis*: classification of RNases H into three families. *Biochemistry* 38, 605–618.
5. Hogrefe, H.H., Hogrefe, R.I., Walder, R.Y. & Walder, J. A. (1990) Kinetic analysis of *Escherichia coli* RNase H using DNA-RNA-DNA/DNA substrates. *J Biol Chem* 265, 5561–5566.
6. Haruki, M., Tsunaka, Y., Morikawa, M. & Kanaya, S. (2002) Cleavage of a DNA-RNA-DNA/DNA chimeric substrate containing a single ribonucleotide at the DNA-RNA junction with prokaryotic RNases HII. *FEBS Lett* 531(2), 204–208.
7. Jongruja, N., You, D-J., Angkawidjaja, C., Kanaya, E., Koga, Y., and Kanaya, S. (2012) Structure and characterization of RNase H3 from *Aquifex aeolicus*. *FEBS J* 279, 2737–2753.
8. Eder, P.S. & Walder, J.A. (1991) Ribonuclease H from K562 human erythroleukemia cells: Purification, characterization, and substrate specificity. *J Biol Chem* 266, 6472–6479.

9. Eder, P.S., Walder, R.Y. & Walder, J.A. (1993) Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie* 75, 123–126.

10. Ohtani, N., Tomita, M. & Itaya, M. (2008) Junction ribonuclease activity specified in RNases HII/2. *FEBS J* 275, 5444–5455.

11. Jeong, H.S., Backlund, P.S., Chen, H.C., Karavanov, A.A. & Crouch, R.J. (2004) RNase H2 of *Saccharomyces cerevisiae* is a complex of three proteins. *Nucleic Acids Res* 32, 407–414.

12. Shaban, N.M., Harvey, S., Perrino, F.W. & Hollis, T. (2010) The structure of the mammalian RNase H2 complex provides insight into RNA-DNA hybrid processing to prevent immune dysfunction. *J Biol Chem* 285, 3617–3624.

13. Lu, Z., Liang, R., Liu, X., Hou, J. & Liu, J. (2012) RNase HIII from *Chlamydophila pneumoniae* can efficiently cleave double-stranded DNA carrying a chimeric ribonucleotide in the presence of manganese. *Mol Microbiol* 83, 1080–1093.

14. Lu, Z., Hou, J., Wang, Y. & Liu, J. (2012) Involvement of Ser94 in RNase HIII from *Chlamydophila pneumoniae* in the recognition of a single ribonucleotide misincorporated into double-stranded DNA. *Biochim Biophys Acta* 1824, 859–865.

15. Figiel, M., Chon, H., Cerritelli, S. M., Cybulska, M., Crouch, R. J. & Nowotny, M. (2011) The structural and biochemical characterization of human RNase H2 complex reveals the molecular basis for substrate recognition and Aicardi-Goutieres syndrome defects. *J Biol Chem* 286, 10540-10550.

16. Reijns, M.A., Bubeck, D., Gibson, L.C., Graham, S.C., Baillie, G.S., Jones, E.Y. & Jackson, A.P. (2011) The structure of the human RNase H2 complex defines key interaction interfaces relevant to enzyme function and human disease. *J Biol Chem* 286, 10530-10539.

17. Ogawa, T. & Okazaki, T. (1984) Function of RNase H in DNA replication revealed by RNase H defective mutants of *Escherichia coli*. *Mol Gen Genet* 193(2), 231-237.

18. Kogoma, T., & Foster, P.L. (1998) Physiological functions of *E. coli* RNase HI. Crouch, R.J. & Toulmé, J.J. (eds.), *In Ribonucleases H*, INSERM, Paris, pp. 39–66.

19. Qiu, J., Qian, Y., Frank, P., Wintersberger, U. & Shen, B. (1999) *Saccharomyces cerevisiae* RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Mol Cell Biol* 19, 8361-71.

20. Sato, A., Kanai, A., Itaya, M. & Tomita, M. (2003) Cooperative regulation for Okazaki fragment processing by RNase HII and FEN-1 purified from a hyperthermophilic archaeon, *Pyrococcus furiosus*. *Biochem Biophys Res Commun* 309(1), 247-252.

21. Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E. & Crouch, R.J. (2003) Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol Cell* 11(3), 807-815.

22. Bubeck, D., Reijns, M.A., Graham, S.C., Astell, K.R., Jones, E.Y. & Jackson, A.P. (2011) PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res* 39, 3652-3666.

23. Rydberg, B. & Game, J. (2002) Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc Natl Acad Sci U S A* 99(26), 16654-16659.

24. Nick McElhinny, S.A., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundstrom, E.B., Johansson, E., Chabes, A. & Kunkel, T.A. (2010) Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol* 6, 774-781.

25. Kim, N., Huang, S. N., Williams, J. S., Li, Y. C., Clark, A. B., Cho, J. E., Kunkel, T. A., Pommier, Y. & Jinks-Robertson, S. (2011) Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science* 332, 1561-1564.

26. Lazzaro, F., Novarina, D., Amara, F., Watt, D.L., Stone, J.E., Costanzo, V., Burgers, P.M., Kunkel, T.A., Plevani, P. & Muzi-Falconi, M. (2012) RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Mol Cell* 45, 99-110.

27. Sparks, J.L., Chon, H., Cerritelli, S.M., Kunkel, T.A., Johansson, E., Crouch, R.J. & Burgers, P.M. (2012) RNase H2-initiated ribonucleotide excision repair. *Mol Cell* 47(6), 980-986.

28. Hiller, B., Achleitner, M., Glage, S., Naumann, R., Behrendt, R. & Roers, A. (2012) Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *J Exp Med* 209(8),1419-1426.

29. Drolet, M., Phoenix, P., Menzel, R., Masse, E., Liu, L.F. & Crouch, R.J. (1995) Overexpression of RNase H partially complements the growth defect of *an Escherichia coli* delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci U S A* 92(8), 3526-3530.

30. Cheng, B., Rui, S., Ji, C., Gong, V.W., Van Dyk, T.K., Drolet, M. & Tse-Dinh, Y.C. (2003) RNase H overproduction allows the expression of stress-induced genes in the absence of topoisomerase I. *FEMS Microbiol Lett* 221(2),237-242.

31. Broccoli, S., Rallu, F., Sanscartier, P., Cerritelli, S.M., Crouch, R.J. & Drolet, M. (2004) Effects of RNA polymerase modifications on transcription-induced negative supercoiling and associated R-loop formation. *Mol Microbiol* 52, 1769–1779.

32. Nick McElhinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundstrom, E.B., Burgers, P.M., Johansson, E., Chabes, A. & Kunkel, T.A. (2010) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Nat Acad Sci U S A* 107, 4949–4954.

33. Itaya, M., Omori, A., Kanaya, S., Crouch, R.J., Tanaka, T., Kondo, K. (1999) Isolation of RNase H genes that are essential for growth of *Bacillus subtilis* 168. *J Bacteriol* 181, 2118–2123.

34. Arudchandran, A., Cerritelli, S., Narimatsu, S., Itaya, M., Shin, D.Y., Shimada, Y. & Crouch, R.J. (2000) The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. *Genes Cells* 5, 789–802.

35. Crow, Y.J., Leitch, A., Hayward, B.E., Garner,A., Parmar,R., et al., (2006) Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nat Genet* 38, 910-916.

36. Champoux, J.J., & Schultz, S.J. (2009) Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription. *FEBS J* 276, 1506–1516.

37. Moelling, K. (2012) Targeting the retroviral ribonuclease H by rational drug design. *AIDS* 26(16), 1983-1993.

38. Wendeler, M., Lee, H.F., Bermingham, A., Miller, J.T., Chertov, O., Bona, M.K., Baichoo, N.S., Ehteshami, M., Beutler, J., O'Keefe, B.R., Götte, M., Kvaratskhelia, M. & Le Grice, S. (2008) Vinylogous ureas as a novel class of inhibitors of reverse transcriptase–associated ribonuclease H activity. *ACS Chem Biol* 3, 635–644.

39. Chung, S., Wendeler, M., Rausch, J.W., Beilhartz, G., Gotte, M., O'Keefe, B.R., Bermingham, A., Beutler, J.A., Liu, S., Zhuang, X. & Le Grice, S.F. (2010) Structure-activity analysis of vinylogous urea inhibitors of human immunodeficiency virus-encoded ribonuclease H. *Antimicrob Agents Chemother* 54, 3913–3921.

40. Chung, S., Miller, J.T., Johnson, B.C., Hughes, S.H. & Le Grice, S.F. (2012) Mutagenesis of human immunodeficiency virus reverse transcriptase p51 subunit defines residues contributing to vinylogous urea inhibition of ribonuclease H activity. *J Biol Chem* 287, 4066–4075.

41. Dean, N.M. & Bennett, C.F. (2003) Antisense oligonucleotide-based therapeutics for cancer. *Oncogene* 22, 9087–9096.

42. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1990) Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* 347, 306–309.

43. Yang, W., Hendrickson, W.A., Crouch, R.J. & Satow, Y. (1990) Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science* 249, 1398–1405.

44. Davies II, J.F., Hostomska, Z., Hostomsky, Z., Jordan, S.R., Matthews, D. (1991) Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, 252, 88–95.

45. Lai, L., Yokota, H., Hung, L.W., Kim, R. & Kim. S.H. (2000) Crystal structure of archaeal RNase HII: a homologue of human major RNase H. *Struc Fold Des* 8, 897–904.

46. Chon, H., Matsumura, H., Koga, Y., Takano, K. & Kanaya, S. (2006) Crystal structure and structure-based mutational analyses of RNase HIII from *Bacillus stearothermophilus*: a new type 2 RNase H with TBP-like substrate-binding domain at the N terminus. *J Mol Biol* 356, 165–178.

47. Rychlik, M.P., Chon, H., Cerritelli, S.M., Klimek, P., Crouch, R.J. & Nowotny, M. (2010) Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage. *Mol Cell* 40(4), 658–670.

48. Nowotny, M., Gaidamakov, S.A., Crouch, R.J. & Yang, W. (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* 121, 1005–1016.

49. Nowotny, M., Gaidamakov, S.A., Ghirlando, R., Cerritelli, S.M., Crouch, R.J. & Yang, W. (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell* 28, 264–276.

50. Jongruja. N., You, D-J., Kanaya, E., Koga, Y., Takano, K. & Kanaya, S. (2010) The N-terminal hybrid binding domain of RNase HI from *Thermotoga maritima* is important for substrate binding and Mg^{2+}-dependent activity. *FEBS J* 277, 4474–4489.

51. Lima, W.F., Wu, H., Nichols, J.G., Prakash, T.P., Ravikumer, V., et al. (2003) Human RNase H1 uses one tryptophan and two lysines to position the enzyme at the 3'-DNA/5'-RNA terminus of the heteroduplex substrate. *J Biol Chem* 278, 49860–49867.

52. Nowotny, M., Cerritelli, S.M., Ghirlando, R., Gaidamakov, S.A., Crouch, R.J., et al., (2008) Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. *EMBO J* 27, 1172–1181.

53. Yang, W., Lee, J.Y. & Nowotny, M. (2006) Making and breaking nucleic acids: two-Mg^{2+}-ion catalysis and substrate specificity. *Mol Cell* 22, 5–13.

54. Nowotny, M. & Yang, W. (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J* 25, 1924–1933.

55. Oda, Y., Yoshida, M. & Kanaya, S. (1993) Role of histidine 124 in the catalytic function of ribonuclease HI from *Escherichia coli*. *J Biol Chem* 268, 88–92.

56. You, D-J., Chon, H., Koga, Y., Takano, K. & Kanaya, S. (2007) Crystal structure of type 1 Ribonuclease H from hyperthermophilic archaeon *Sulfolobus tokodaii*: role of Arg118 and C-terminal anchoring. *Biochemistry* 46, 11494–11503.

57. Mandel, A.M., Akke, M. & Palmer, A.G. III (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* 246, 144–163.

58. Ohtani, N., Yanagawa, H., Tomita, M. & Itaya, M. (2004) Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, *Sulfolobus tokodaii* 7. *Nucleic Acids Res* 32, 5809–5819.

59. Ohtani, N., Yanagawa, H., Tomita, M. & Itaya, M. (2004) Identification of the first archaeal type 1 RNaseH gene from *Halobacterium* sp. NRC-1: archaeal RNase HI can cleave an RNA–DNA junction. *Biochem J* 381,795–802.

60. Kanaya, S., Katsuda-Nakai, C. & Ikehara, M. (1991) Importance of the positive charge cluster in *Escherichia coli* ribonuclease HI for the effective binding of the substrate. *J Biol Chem* 266, 11621–11627.

61. Haruki, M., Noguchi, E., Kanaya, S. & Crouch, R.J. (1997) Kinetic and stoichiometric analysis for the binding of *Escherichia coli* ribonuclease HI to RNA–DNA hybrids using surface plasmon resonance. *J Biol Chem* 272, 22015–22022.

62. Gotte, M., Maier, G., Onori, A.M., Cellai, L., Wainberg, M.A. & Heumann, H. (1999) Temporal coordination between initiation of HIV(+)-strand DNA synthesis and primer removal. *J Biol Chem* 274, 11159–11169.

63. Hughes, S.H., Arnold, E. & Hostomsky, Z. (1998) RNase H of retroviral reverse transcriptases. *In Ribonucleases H* (Crouch RJ & Toulme JJ, eds), pp. 195–224. INSERM, Paris.

64. Kochiwa, H., Tomita, M. & Kanai, A. (2007) Evolution of ribonuclease H genes in prokaryotes to avoid inheritance of redundant genes. *BMC Evol Biol* 7, 128–140.

65. Sträter, N., Lipscomb, W.N., Klabunde, T. & Krebs, B (1996) Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew Chem Int Ed Engl* 35, 2024–2055.

66. Yang, W. (2008) An equivalent metal ion in one- and two-metal-ion catalysis. *Nat Struct Mol Biol,* 15(11), 1228–1231.

67. Chai, Q., Qiu, J., Chapados, B.R. & Shen, B. (2001) Archaeoglobus fulgidus RNase HII in DNA replication: enzymological functions and activity regulation via metal cofactors. *Biochem Biophys Res Commun* 286, 1073–1081.

68. Muroya, A., Tsuchiya, D., Ishikawa, M., Haruki, M., Morikawa, M., Kanaya, S. & Morikawa, K. (2001) Catalytic center of an archaeal Type 2 Ribonuclease H as revealed by X-ray crystallographic and mutational analyses. *Protein Sci* 10, 707–714.

69. Tsunaka, Y., Haruki, M., Morikawa, M., Oobatake, M. & Kanaya, S. (2003) Dispensability of Glu48 and Asp134 for Mn2+-dependent activity of *E. coli* ribonuclease HI. *Biochemistry* 42, 3366–3374.

70. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. & Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5, R245–R249.

71. Amann, R.I., Ludwig, W. & Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59, 143–169.

72. Streit, W.R. & Schmitz, R.A. (2004) Metagenomics - the key to the uncultured microbes. *Curr Opin Microbiol* 7, 492–498.

73. Lorenz, P. & Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev* 3, 510–516.

74. Rajendhran, J. & Gunasekaran, P. (2008) Strategies for accessing soil metagenome for desired applications. *Biotechnol Adv* 26, 576–590.

75. Morimoto, S. & Fujii, T. (2009) A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Appl Microbiol Biotechnol* 83, 389–396.

76. Kotik, M. (2009) Novel genes retrieved from environmental DNA by polymerase chain reaction: current genome-walking techniques for future metagenome applications. *J Biotechnol* 144, 75–82.

77. Simon, C. & Daniel, R. (2009) Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol* 85(2), 265–276.

78. Simon, C. & Daniel, R. (2011) Metagenomic analyses: past and future trends. *Appl Environ Microbiol* 77(4), 1153–1161.

79. Xing, M.N., Zhang, X.Z. & Huang, H. (2012) Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnol Adv* 30(4), 920–929.

80. Lammle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H. & Rupp, S. (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol* 127(4), 575-592

81. Pang, H., Zhang, P., Duan, C.J., Mo, X.C., Tang, J.L. & Feng, J.X. (2009) Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr Microbiol* 58(4), 404-408.

82. Kim, Y.H., Kwon, E.J., Kim, S.K., Jeong, Y.S., Kim, J., Yun, H.D. & Kim, H. (2010) Molecular cloning and characterization of a novel family VIII alkaline esterase from a compost metagenomic library. *Biochem Biophys Res Commun* 393(1), 45-49.

83. Itaya, M. & Crouch, R.J. (1991) A combination of RNase H (*rnh*) and recBCD or sbcB mutations in *Escherichia coli* K12 adversely affects growth. *Mol Gen Genet* 227, 424–432.

84. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673–4680.

85. Page, R.D.M. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12, 357–358.

86. Itaya, M. (1990) Isolation and characterization of a second RNase H (RNase HII) of *Escherichia coli* K-12 encoded by the *rnhB* gene. *Proc Natl Acad Sci U S A* 87, 8587–8591.

87. Itaya, M., McKelvin, D., Chatterjie, S.K. & Crouch, R.J. (1991) Selective cloning of genes encoding RNase H from *Salmonella typhimurium*, *Saccharomyces cerevisiae* and *Escherichia coli rnh* mutant. *Mol Gen Genet* 227, 438–445.

88. Itaya, M. & Kondo, K. (1991) Molecular cloning of a ribonuclease H (RNase HI) gene from an extreme thermophile *Thermus thermophilus* HB8: a thermostable RNase H can functionally replace the *Escherichia coli* enzyme *in vivo*. *Nucleic Acids Res* 19, 4443–4449.

89. Campbell, A.G. & Ray, D.S. (1993) Functional complementation of an *Escherichia coli* ribonuclease H mutation by a cloned genomic fragment from trypanosomatid *Crithidia fasciculata*. *Proc Natl Acad Sci U S A* 90, 9350–9354.

90. Hesslein, D.G.T. & Campbell, A.G. (1997) Molecular cloning and expression of a ribonuclease H from the kinetoplastid, *Trypanosoma brucei*. *Mol Biochem Parasitol* 86, 121–126.

91. Zhang, Y.-B., Ayalew, S. & Lackes, S.A. (1997) The *rnhB* gene encoding RNase HII of *Streptococcus pneumoniae* and evidence of conserved motifs in eucaryotic genes. *J Bacteriol* 179, 3828–3836.

92. Haruki, M., Hayashi, K., Kochi, T., Muroya, A., Koga, Y., Morikawa, M., Imanaka, T. & Kanaya, S. (1998) Gene cloning and characterization of recombinant ribonuclease HII from a hyperthermophilic archaeon. *J Bacteriol* 180, 6207–6214.

93. Torarinsson, E., Klenk, H.P. & Garrett, R.A. (2005) Divergent transcriptional and translational signals in archaea. *Environ Microbiol* 7, 47–54.

94. Laursen, B.S., Sørensen, H.P., Mortensen, K.K. & Sperling-Petersen, H.U. (2005) Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* 69, 101–123.

95. Cerritelli, S.M. & Crouch, R.J. (2009) Ribonuclease H: the enzymes in eukaryotes. *FEBS J* 276, 1494–1505.

96. Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.-E. & Streit, W.R. (2003) Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol* 69, 6235–6242.

97. Pang, H., Zhang, P., Duan, C.J., Mo, X.C., Tang, J.L. & Feng, J.X. (2009) Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr Microbiol* 58, 404–408.

98. Jeon, J.H., Kim, J.T., Kim, Y.J., Kim, H.K., Lee, H.S., Kang, S.G., Kim, S.J. & Lee, J.H. (2009) Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl Microbiol Biotechnol* 81, 865–874.

99. Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W.R. (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* 67, 89–99.

100. Majernik, A., Gottschalk, G. & Daniel, R. (2001) Screening of environmental DNA libraries for the presence of genes conferring Na+(Li+)/H+ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J Bacteriol* 183, 6645–6653.

101. Li, Y., Wexler, M., Richardson, D.J., Bond, P.L. & Johnston, A.W.B. (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. *Environ Microbiol* 7, 1927–1936.

102. Wexler, M., Bond, P.L., Richardson, D.J. & Johnston, A.W.B. (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* 7, 1917–1926.

103. Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361, 13–37.

104. Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D. et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1474.

105. Dawes, S.S., Crouch, R.J., Morris, S.L. & Mizrahi, V. (1995) Cloning, sequence analysis, overproduction in *E. coli* and enzymatic characterization of the RNase HI from *Mycobacterium smegmatis*. *Gene* 165, 71–75.

106. Zheng, Y., Roberts, R.J., Kasif, S. & Guan, C. (2005) Characterization of two new aminopeptidases in *Escherichia coli*. *J Bacteriol* 187, 3671–3677.

107. Sprengel, R., Reiss, B. & Schaller, H. (1985) Translationally coupled initiation of protein synthesis in *Bacillus subtilis*. *Nucleic Acids Res* 13, 893–909.

108. Lee, J.H. & Kim, Y.T. (2006) Cloning and characterization of the astaxanthin biosynthesis gene clusterfrom the marine bacterium *Paracoccus haeundaensis*. *Gene* 370, 86–95.

109. You, D.J., Chon, H., Koga, Y., Takano, K. & Kanaya, S. (2006) Crystallization and preliminary crystallographic analysis of type 1 RNase H from the hyperthermophilic archaeon *Sulfolobus tokodaii* 7. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62, 781–784.

110. Laemmli, U.K. (1970) Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* 227, 680–685.

111. Goodwin, T.W. & Morton, R.A. (1946) The spectrophotometric determination of tyrosine and tryptophan in proteins. *Biochem J* 40, 628–632.

112. Otwinowski, Z. & Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography Pt A* 276, 307–326.

113. Vagin, A. & Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* 30, 1022–1025.

114. Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T. & Patel, D.J. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* 461(7265), 754–761.

115. Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E. & Weichenrieder, O. (2011) Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proc Natl Acad Sci U S A* 108(26), 10466–10471.

116. Takano, K., Okamoto, T., Okada, J., Tanaka, S., Angkawidjaja, C., Koga, Y. & Kanaya, S. (2011) Stabilization by fusion to the C-terminus of hyperthermophile *Sulfolobus tokodaii* RNase HI: a possibility of protein stabilization tag. *PLoS One* 6(1), e16226.

117. Ishikawa, K., Nakamura, H., Morikawa, K. & Kanaya, S. (1993) Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry* 32, 6171–6178.

118. Kadonosono, T., Chatani, E., Hayashi, R., Moriyama, H., Ueki, T. (2003) Minimization of cavity size ensures protein stability and folding: structures of Phe46-replaced bovine pancreatic RNase A. *Biochemistry* 42, 10651–10658.

119. Pace, C.N., Fu, H.L., Fryar, K.L., Landua, J., Trevino, S.R., Shirley, B.A., Hendricks, M.M., Iimura, S., Gajiwala, K., Scholtz, J.M. & Grimsley, G.R. (2011) Contribution of hydrophobic interactions to protein stability. *J Mol Biol* 408, 514–528.

120. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34, W116–W118.

121. Emsley, P. & Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126–2132.

122. Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3, 1171–1179.

123. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240–255.

124. Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., et al. (1990) Identification of the amino acid residues involved in an active site of *Escherichia coli* ribonuclease H by site-directed mutagenesis. *J Biol Chem* 265, 4615–4621.

125. Kashiwagi, T., Jeanteur, D., Haruki, M., Katayanagi, K., Kanaya, S. & Morikawa, K. (1996) Proposal for new catalytic roles for two invariant residues in *Escherichia coli* ribonuclease HI. *Protein Eng* 9, 857–867.

126. Berkower, I., Leis, J. & Hurwitz, J. (1973) Isolation and characterization of an endonuclease from *Escherichia coli* specific for ribonucleic acid in ribonucleic acid-deoxyribonucleic acid hybrid structures. *J Biol Chem* 248, 5914–5921.

127. Zhang, H., Sekiguchi, Y., Hanada, S., Hugenholtz, P., Kim, H., Kamagata, Y. & Nakamura, K. (2003) *Gemmatimonas aurantiaca* gen. nov., sp. nov., a gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes* phyl. nov. *Int J Syst Evol Microbiol* 53, 1155–1163.

List of publications

1. Kanaya, E., Sakabe, T., Nguyen, T.N., Koikeda, S., Koga, Y., Takano, K. & Kanaya, S. (2010) Cloning of the RNase H genes from a metagenomic DNA library: identification of a new type 1 RNase H without a typical active-site motif. *J Appl Microbiol* 109, 974-983.
2. Nguyen, T.N., Angkawidjaja, C., Kanaya, E., Koga, Y., Takano, K. & Kanaya, S. (2012) Activity, stability, and structure of metagenome-derived LC11-RNase H1, a homolog of *Sulfolobus tokodaii* RNase H1. *Protein Sci* 21, 553-561.
3. Nguyen, T.N., You, D.J., Matsumoto, H., Kanaya, E., Koga, Y. & Kanaya, S. (2013) Crystal structure of metagenome-derived LC11-RNase H1 in complex with RNA/DNA hybrid. *J Struct Biol*, http://dx.doi.org/10.1016/j.jsb.2013.02.018.
4. Nguyen, T.N., You, D.J., Kanaya, E., Koga, Y. & Kanaya, S. (2013) Crystal structure of metagenome-derived LC9-RNase H1 with atypical DEDN active site motif. *FEBS Lett*, http://dx.doi.org/10.1016/j.febslet.2013.03.020.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor, Prof. Shigenori Kanaya, for supporting me during these past more or less five years. His guidance helped me in all the time of my research, and I truly appreciate his careful reading and correcting my writing since my first manuscript until the time of this thesis. With his immense knowledge, he offered me very valuable comments and suggestions; thereby I could finish my study earlier than the schedule. He is strict in work, but humorous and very friendly. Honestly, I could not have imagined having a better advisor and mentor for my PhD study. Once again, thank you very much “sensei”!

I also would like to thank several other individuals who in one way or another contributed to the preparation and completion of this study.

The Japanese Ministry of Education, Culture, Sports, Science and Technology, for the financial support throughout the period of my staying in Japan.

Assist. Prof. Clement Angkawidjaja, Dr. Dong-Ju You and Dr. Marcin Nowotny, for introducing, teaching, guiding and helping me for all the protein crystallography works. My very first steps into this field were led by them.

Dr. Eiko Kanaya and Tomoko Sakabe for helping me in metagenome work, and Hiroyuki Matsumoto for Sto-RNase H1 study.

Prof. Hajime Watanabe and Prof. Takeharu Nagai, for valuable comments, hard questions and constructive criticism, that helped me to make this thesis more completed.

Ms. Reiko Matsumoto, former secretary of Kanaya laboratory (Molecular Biotechnology laboratory), who always has kind concern and consideration regarding our academic requirements or even our daily life, not only during the time she was working at our lab, but also until now (long after her job transfer).

Prof. Kazufumi Takano, Assoc. Prof. Yuichi Koga, Tita, Ploy, Elias and many other Kanaya lab members, for the stimulating discussions and helpful ideas.

My deep thanks also go to my friends in Japan for being around and generously lending me a hand whenever I am in need of help. To Ilma, words cannot express my thankfulness to you, for being very helpful, understanding, cheering me up and sometimes being tolerant of my teases. To Maria, thank you for your support whenever I need your help, especially anything related to Japanese. To Rina, Nobby, Hansky, Fitri, Farah, pa Tuệ, Tú Anh, bé Bò, ku Thắng, anh Nam, anh Ninh, ku Vũ and many other Vietnamese friends (too many to list here but you know who you are), thank you guys for all the fun we had and all the sadness we shared, that helped me get through all stressful and depressing moments.

I especially thank my family and my friends for their moral support although they are far away from Japan. To my cousins, my uncles and my aunts, bạn Uyên, bạn Trọng Nam, chị Phương Thảo, chị Ca, bạn Na and more, thank you for showing your consideration for me all the time and bringing me warmth especially when I was feeling down.

And above all, I dedicate this thesis to my Mother who has constantly provided unconditional love and care for me! I would not have made it this far without her.

To the memory of Mợ Sáu.