Doctoral Dissertation

Studies on the N-terminal substrate binding domains of bacterial RNases H1 and H2

細菌由来 RNase H1 と RNase H2 の N 末端基質結合ドメインに関する研究

Etin Diah Permanasari

June 2015

Graduate School of Engineering

Department Material and Life Science

Osaka University

Contents

Abbreviations

RNase H	: Ribonuclease H
dsDNA	: double-stranded DNA
$dsDNA^{R1}$	: double-stranded DNA containing single ribonucleotide(s)
dsRNA	: double-stranded RNA
JRNase	: junction ribonuclease
$tRNA^{Lys}$	: single-stranded RNA molecule used as a primer for synthesis minus-strand DNA
RT	: reverse transcriptase
HIV-1 RT	: reverse transcriptase from HIV-1
RNH^{HIV}	: RNase H domain of HIV-1 RT isolated from polymerase domain
EcoRNH1	: RNase H1 from *Escherichia coli*
EcoRNH2	: RNase H2 from *E. coli*
BsuRNH2	: RNase H2 from *Bacillus subtilis*
BsuRNH3	: RNase H3 from *B. subtilis*
AaeRNH2	: RNase H2 from *Aquifex aeolicus*
AaeRNH3	: RNase H3 from *A. aeolicus*
CpeRNH3	: RNase H3 from *Chlamydophila pneumonia*
SceRNH1	: RNase H1 from *Saccharomyces cerevisiae*
SceRNH2	: RNase H2 from *S. cerevisiae*
BhaRNH1	: RNase H1 from *Bacillus halodurans*
BhaRNH1C	: C-terminal domain of BhaRNH1
HsaRNH1	: RNase H1 from *Homo sapiens*
HsaRNH1C	: C-terminal domain of HsaRNH1
HsaRNH2	: RNase H2 from *H. sapiens*
HsaRNH2A	: subunit A (catalytic subunit) of HsaRNH2
TthRNH1	: RNase H1 from *Thermus thermophiles*
SIB1 RNase H1	: RNase H1 from *Shewanella* sp. SIB1
StoRNH1	: RNase H1 from *Sulfolobus tokodaii*
TkoRNH2	: RNase H2 from *Thermococcus kodakaraensis*

AfuRNH2	: RNase H2 from *Archaeoglobus fulgidus*
MjaRNH2	: RNase H2 from *Methanococcus jannaschii*
TmaRNH1	: RNase H1 from *Thermotoga maritima*
TmaRNH2	: RNase H2 from *T. maritima*
BstRNH2	: RNase H2 from *Bacillus stearothermophilus*
BstRNH3	: RNase H3 from *B. stearothermophilus*
TamRNH3	: RNase H3 from *Thermovibiro ammonificans*
HBD	: hybrid binding domain
TmaHBD	: HBD of TmaRNH1
RBD	: dsRNA binding domain
SIB1 RBD-RNase H1	: RBD of SIB1 RNase H1
TBP	: TATA-box binding protein
BstTBP	: TBP-like domain of BstRNH3
N-extension	: N-terminal extension
BstNTD	: N-extension of BstRNH2
BstRNH2ΔN	: BstRNH2 lacking N-extension (residues 60-259)
C-extension	: C-terminal extension
TmaRNH2ΔC	: TmaRNH2 lacking C-extension (residues 1-192)
TmaHH	: a hairpin helix structure that corresponds to C-extension of TmaRNH2
BstNTD-TmaRNH2ΔC	: BstNTD is attached to the N-terminus of TmaRNH2ΔC
GRG	: a conserved motif in RNases H2 consists of Gly, Arg and Gly.
pET800TM	: pET25b containting the gene of TmaRNH2
pET600AA	: pET25b containing the gene of AaeRNH2
pET800BS	: pET25b containing the gene of BstRNH2
pET670TMΔC	: pET25b containing the gene of TmaRNH2ΔC
pET620BSΔN	: pET25b containing the gene of BstRNH2ΔN
pET850BSTMΔC	: pET25b containing the gene of BstNTD-TmaRNH2ΔC
CD	: Circular dichroism
GdnHCl	: Guadinide hydrochloride
IPTG	: Isopropyl thio-β-D-galactoside

DTT	: Dithiothreitol
2-ME	: 2-mercapthoethanol
BSA	: Bovine serume albumin
EDTA	: Ethylenediaminetetraacetic acid
PDB	: Protein data bank
D_5-R_1-D_6/D_{12}	: DNA_5-RNA_1-DNA_6/DNA_{12}
D_{15}-R_1-D_{13}/D_{29}	: DNA_{15}-RNA_1-DNA_{13}/DNA_{29}
R_{12}/D_{12}	: RNA_{12}/DNA_{12}
R_{29}/D_{29}	: RNA_{29}/DNA_{29}
K_A	: the association constant
T_m	: the melting temperature
ΔH_m	: the enthalpy change
ΔS_m	: the entropy change
ΔG	: the free energy change
P66 subunit	: a subunit of HIV-1 RT in size of 66 kDa that consists of N terminal DNA polymerase domain and C-terminal RNase H domain
P51 subunit	: a subunit of HIV-1 RT in size of 51 kDa that consists of only N-terminal DNA polymerase domain
TmaHBD-RNH^{HIV}	: TmaHBD is attached at the N-terminus of RNH^{HIV}
BstNTD-RNH^{HIV}	: BstNTD is attached at the N-terminus of RNH^{HIV}
pET-HIVMp66	: pET 25b containing the gene of P66 subunit of HIV-1 RT
pET-RNH^{HIV}	: pET25b containing the gene of RNH^{HIV}
pET-TmaHBD-RNH^{HIV}	: pET25b containing the gene of TmaHBD-RNH^{HIV}
pET-BstNTD-RNH^{HIV}	: pET25b containing the gene of BstNTD-RNH^{HIV}
GFP	: Green fluorescence protein

CHAPTER 1

General Introduction

1.1. Ribonuclease H (RNase H)

RNase H (EC 3.1.26.4) is a non-sequence-specific enzyme that endonucleolytically cleaves the RNA strand of RNA/DNA hybrids [1, 2]. The reaction occurs at the phosphodiester bond of the RNA moiety and requires divalent metal ions, such as Mg^{2+} and Mn^{2+}, to produce the 5'-phosphate and 3'-hydroxyl groups, as shown in Figure 1.1.

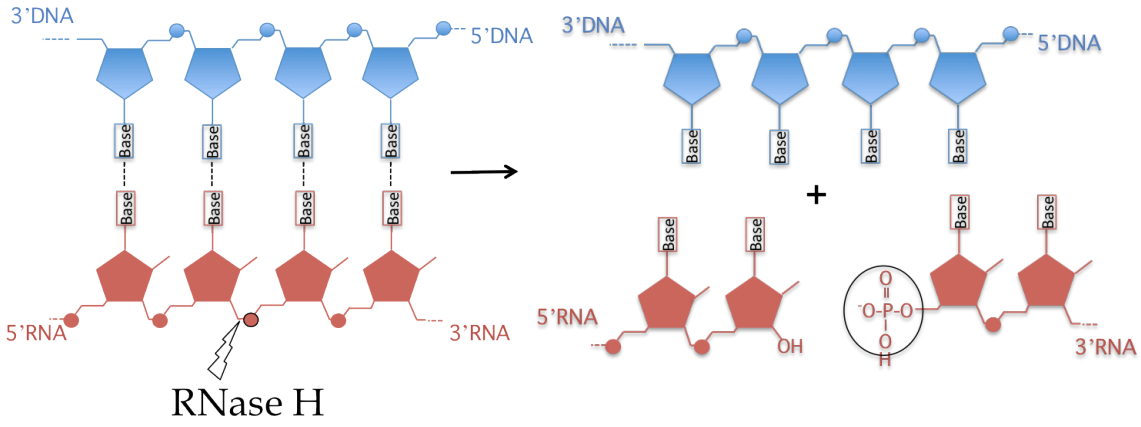


Fig. 1.1. Cleavage of RNA/DNA hybrid with RNase H

RNase H is widely present in all three kingdoms of living organisms, bacteria, archaea and eukaryotes [3, 4]. It is also present in retrovirus as a C-terminal domain of reverse transcriptase (RT) [3, 4]. The first gene encoding RNase H was cloned from *Escherichia coli* in 1983 [5]. Another gene encoding RNase H was cloned from *E. coli* in 1990, by its ability to complement the temperature-sensitive growth phenotype of *E. coli* strain lacking the first gene encoding RNase H [6]. RNases H encoded by the first and second genes cloned from *E. coli* are termed *E. coli* RNase H1 (EcoRNH1) and *E. coli*

RNase H2 (EcoRNH2) respectively. EcoRNH1 and EcoRNH2 show little amino acid sequence similarity with each other, suggesting that they are evolutionarily unrelated. Later, the gene encoding RNase H3 was cloned from *Bacillus subtilis* in 1999 [7]. The *B. subtilis* genome also contains the gene encoding RNase H2 but does not contain the gene encoding RNase H1, although it contains pseudo gene encoding inactive RNase H1 homolog. *B. subtilis* RNase H3 (BsuRNH3) show a weak amino acid sequence similarity to *B. subtilis* RNase H2 (BsuRNH2), suggesting that BsuRNH2 and BsuRNH3 are diverged from common ancestor. Database searches indicate that all RNase H sequences show a similarity to that of EcoRNH1, EcoRNH2 or BsuRNH3. Thus, RNases H are classified into two major families, type 1 and type 2 RNases H, which are evolutionarily unrelated, based on the difference in their amino acid sequences [3, 4]. Type 1 RNases H include bacterial RNase H1, archaeal RNase H1, eukaryotic RNase H1, and retroviral RNase H. Type 2 RNases H include bacterial RNases H2 and H3, archaeal RNase H2, and eukaryotic RNase H2. They are all functional in a monomeric form, except for eukaryotic RNase H2 that is functional in a heterotrimeric form [3, 8].

A single cell of organisms often contains multiple RNases H [4]. For example, the *E. coli* genome contains two genes encoding RNases H1 and H2, the *Aquifex aeolicus* genome contains two genes encoding RNases H2 and H3, and the *Saccharomyces cerevisiae* genome contains two genes encoding RNases H1 and H2, as shown in Figure 1.2. The number and types of the RNase H genes contained in the single genomes are not correlated with the evolutionary relationships of their source organisms determined based on the 16*S* rRNA sequences [4]. Likewise, single eukaryotic genomes often contain two different types of the RNase H genes, which are the genes encoding RNases H1 and H2.

The significance of the multiplicity of the RNases H genes in the single cells is not fully understood. However, they may play different roles in cells as described in section 1.3.

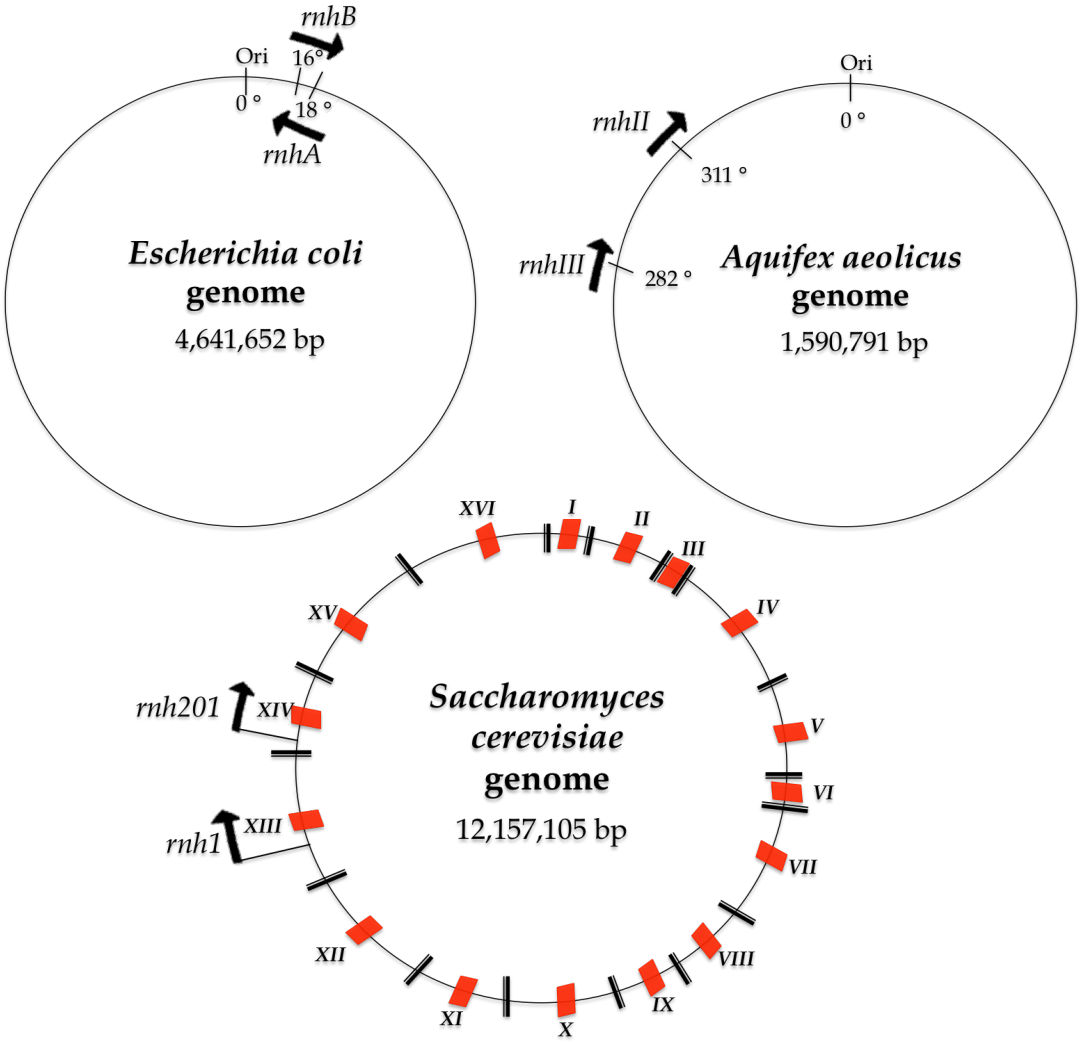


Fig. 1.2. Multiplicity of the RNase H genes in the single genome.

Locations of the genes encoding RNases H on each genome are shown. The *rnhA* and *rnhB* genes of *E. coli* encode RNases H1 and H2 respectively. The *rnhII* and *rnhIII* genes of *A. aeolicus* encode RNases H2 and H3 respectively. The *rnh1* and *rnh201* genes of *S. cerevisiae* encode RNases H1 and subunit A of RNase H2 respectively. The locations of the genes on the *E. coli* genome are 235,535 to 236,002 for *rnhA* and 204,493 to 205,089 for *rnhB*. The locations of the genes on the *A. aeolicus* genome are 1,374,521 to 1,375,111 for *rnhII* and 1,245,528 to 1,246,301 for *rnhIII*. The locations of the genes on the *S. cerevisiae* genome are 740,226 to 741,312 (chromosome XIII) for *rnh1* and 490,317 to 491,240 (chromosome XIV) for *rnh201*. The database search and gene locations are from Kyoto Encyclopedia of Genes and Genome (KEGG) site (http://www.genome.jp/kegg/). An arrow shows the direction of the transcription. Ori indicates the origin of replication.

1.2. Substrate specificity of RNase H

Three different substrates are used to determine the enzymatic activity of RNases H. They are schematically shown in Figure 1.3. RNA-DNA/DNA duplex represents an Okazaki fragment like substrate. dsDNA^{R1} represents double-stranded DNA (dsDNA) containing a single ribonucleotide. RNases H1 and H2 differ in their ability to cleave dsDNA^{R1}. While RNase H2 cleaves dsDNA^{R1} at the (5')DNA-RNA(3') junction, RNase H1 does not cleave it [8-12]. The activity that catalyzes the cleavage of dsDNA^{R1} at the (5')DNA-RNA(3') junction is termed junction ribonuclease (JRNase) activity. This activity is originally defined as the activity that catalyzes the cleavage of an Okazaki fragment-like RNA-DNA/DNA substrate at the 5'-side of the ribonucleotide attached to DNA [10]. RNase H1 does not cleave dsDNA^{R1}, because at least four consecutive ribonucleotides are necessary for effective cleavage of substrate with RNase H1 [12, 13]. Conflicting results have been reported for the ability of RNase H3 to cleave dsDNA^{R1}. BsuRNH3 [12] and *Aquifex aeolicus* RNase H3 (AaeRNH3) [14] do not cleave dsDNA^{R1}, whereas *Chlamydophila pneumonia* RNase H3 (CpeRNH3) cleaves dsDNA^{R1} at the (5')DNA-RNA(3') junction [15, 16].

Fig. 1.3. Substrate of RNase H

In this thesis, the activity determined using RNA/DNA hybrid and dsDNA^{R1} as substrates are defined as RNase H and JRNase activities respectively. It is difficult to

define these activities using RNA-DNA/DNA substrate, because not only JRNase activity but also RNase H activity catalyze the cleavage of this substrate at the 5'-side of the ribonucleotide attached to DNA. Thus, all RNases H exhibit RNase H activity, while only RNase H2 and some RNases H3 exhibit JRNase activity.

1.3. Physiological role of RNase H

Because of the unique RNase H and JRNase activities, RNases H are involved in DNA replication, repair, and transcription [17-34]. They protect cells from undesired

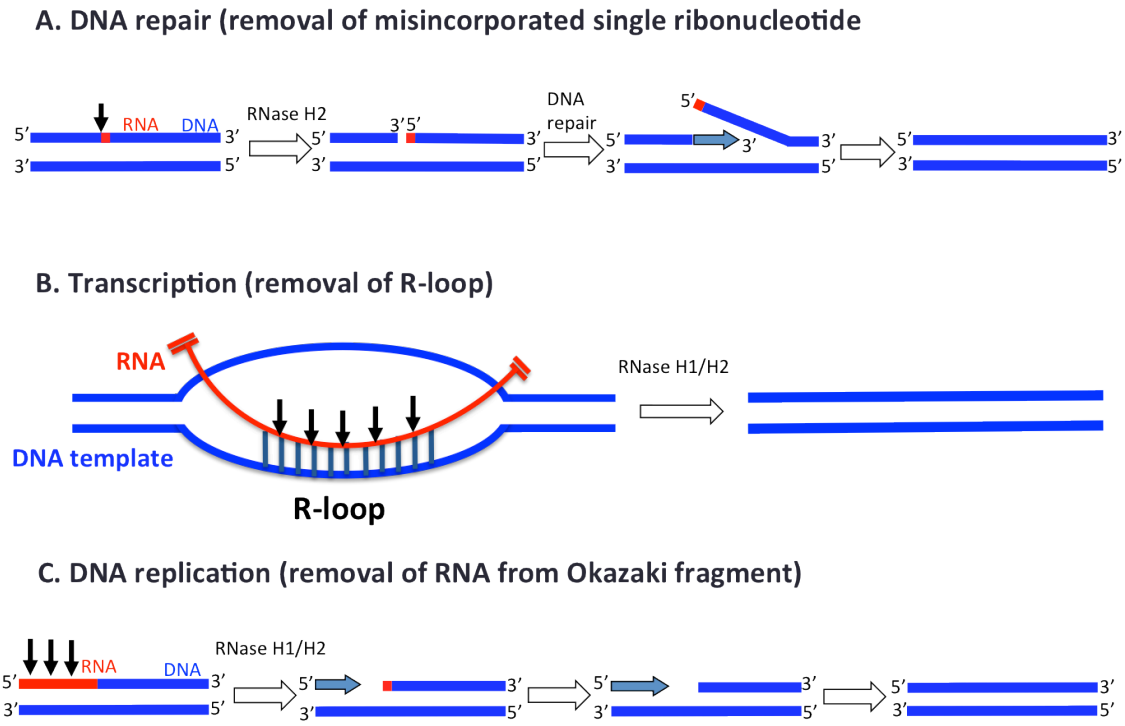


Fig. 1.4. Physiological role of RNases H.

dsDNA^{R1} (A), R-loop (B) and RNA-DNA/DNA duplex (C) are schematically shown. RNA and DNA are colored red and blue respectively. dsDNA^{R1} represents genomic DNA containing single ribonucleotides, which are misincorporated during DNA replication. R-loop is formed during transcription when the newly synthesized RNA strand fails to dissociate from the DNA strand. RNA-DNA/DNA duplex represents an Okazaki fragment produced during DNA replication.

accumulation of nucleic acids. Possible physiological roles of RNases H are shown in Figure 1.4.

dsDNA^{R1} represents genomic DNA containing single ribonucleotides. It is generated by misincorporation of single ribonucleotides into genomic DNA during DNA replication [17-19]. This misincorporation causes genome instability. JRNase activity of RNase H2 is necessary to maintain the genome integrity by removing ribonucleotides misincorporated into DNA through ribonucleotide excision repair (RER) pathway [18-23] (Fig. 1.4-A). The RER pathway is shown in Figure 1.5 in more detail. R-loop is formed when the newly synthesized RNA strand remains associated with the DNA strand [24]. This loop also causes genome instability. RNase H activity of RNases H1 and/or RNases H2 is necessary to maintain the genome integrity by removing an RNA strand of R-loops [24, 25] (Fig. 1.4-B). Okazaki fragment is produced during DNA replication (lagging DNA synthesis). RNase H activity of RNases H1 and/or H2 is thought to be involved in removal of ribonucleotides from Okazaki fragment, although other enzymes, such as DNA polymerase I and FEN1 nuclease, are required to completely remove ribonucleotides from Okazaki fragment [24, 26, 27] (Fig. 1.4-C).

Role of RNases H1-H3 for cell growth has been analyzed by gene disruption. Disruption of single genes encoding RNases H (RNase H1 or H2 for *E. coli* and *S. cerevisiae*, and RNase H2 or H3 for *B. subtilis*) does not affect growth phenotypes of these organisms. However, disruption of both genes encoding RNases H1 and H2 causes temperature-sensitive growth defect in *E. coli* [28] and increased sensitivity to alkylating agents in *S. cerevisiae* [29]. Disruption of both genes encoding RNases H2 and H3 causes a growth defect in *B. subtilis* [28]. In contrast, disruption of the gene encoding RNase H1 [26] or RNase H2 [22, 23] causes embryonic lethality in mouse. Mutation in any subunit of

RNase H2 causes neurological disorder Aicardi-Goutieres syndrome in human [30, 31]. It is probably due to the accumulation of unprocessed RNA/DNA hybrids in the human cells. Moreover, retroviral RNase H is required for proliferation of virus [32, 33], and therefore it is regarded as one of the targets for AIDS therapy [35, 36]. Retroviral RNase H is required to convert single-stranded (ss) genomic RNA to dsDNA by removing genomic RNA after synthesis of minus-strand DNA, $tRNA^{Lys}$, which is used as a primer for synthesis of minus-strand DNA, and polypurine tract, which is used as a primer for synthesis of plus-strand DNA.

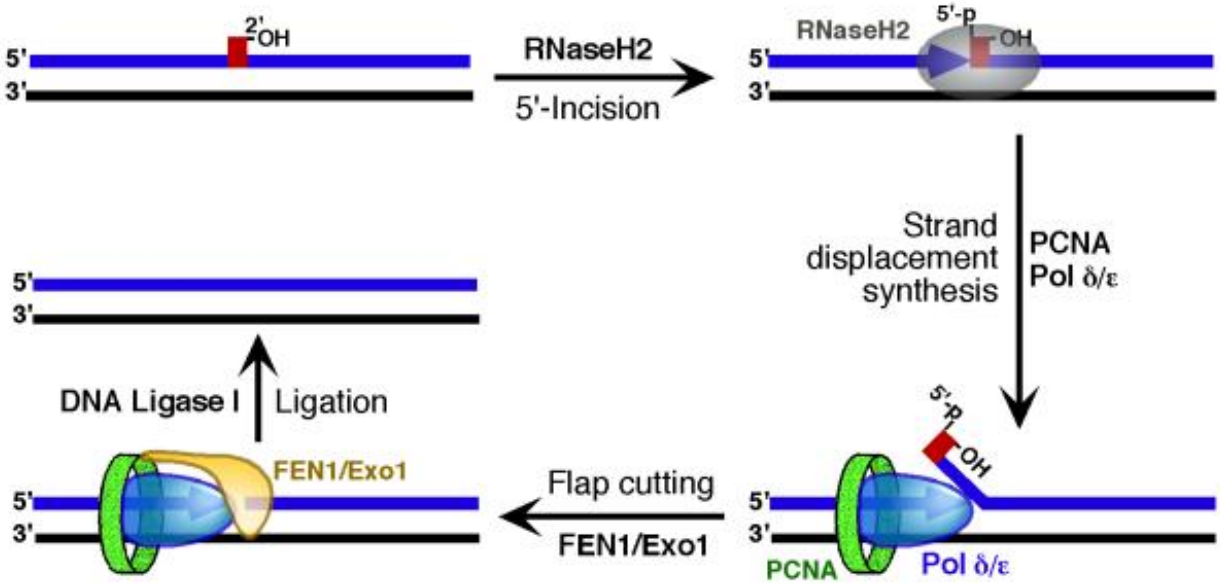


Fig. 1.5. Ribonucleotide Excision Repair (RER) pathway initiated by RNase H2 [20].
RNase H2 incises dsDNA containing a single ribonucleotide at the 5'-side of the ribonucleotide. The mechanism involves other enzymes, such as PCNA, FEN1, Exo1, and DNA ligase, to complete the DNA repair process.

1.4. Structure of RNase H

Type 1 and type 2 RNases H share a similar main-chain fold, termed RNase H-fold, despite their poor amino acid sequence similarities and different substrate specificities. It consists of a five-stranded β-sheet and two α-helices, and contains four acidic active-site residues in the catalytic center [4]. The crystal structure of EcoRNH1 is the first crystal structure determined for RNases H [37, 38]. Since then, many crystal structures of type 1 and type 2 RNases H have been determined. They are the structures of the RNase H domains of HIV-1 RT [39] and moloney murine leukemia viruses RT [40], the catalytic domains of *B. halodurans* RNase H1 (BhaRNH1) [41] and *Homo sapiens* RNase H1 (HsaRNH1) [42], *Thermus thermophiles* RNase H1 (TthRNH1) [43], *Schewanella* sp. MR-1 RNase H1 [44], *Sulfolobus tokodaii* RNase H1 (StoRNH1) [45], metagenome-derived LC9-RNase H1 [46] and LC11-RNase H1 [47, 48], *Halobacterium* sp. NRC-1 RNase H1 [49], *Thermococcus kodakaraensis* RNase H2 (TkoRNH2) [50], *Archaeoglobus fulgidus* RNase H2 (AfuRNH2) [51], *Methanococcus jannaschii* RNase H2 (MjaRNH2) [52], and *Thermotoga maritima* RNase H2 (TmaRNH2) [9], *A. aeolicus* RNase H3 (AaeRNH3) [14], *B. stearothermophilus* RNase H3 (BstRNH3) [53], and *Thermovibiro ammonificans* RNase H3 (TamRNH3) [54]. The catalytic domains of BhaRNH1 and HsaRNH1 are termed BhaRNH1C and HsaRNH1C respectively. The structures of BhaRNH1C, HsaRNH1C, TmaRNH2, and TamRNH3 have been determined as a complex with the substrate. The structures of the representative members of type 1 and type 2 RNases H are shown in Figure 1.6.

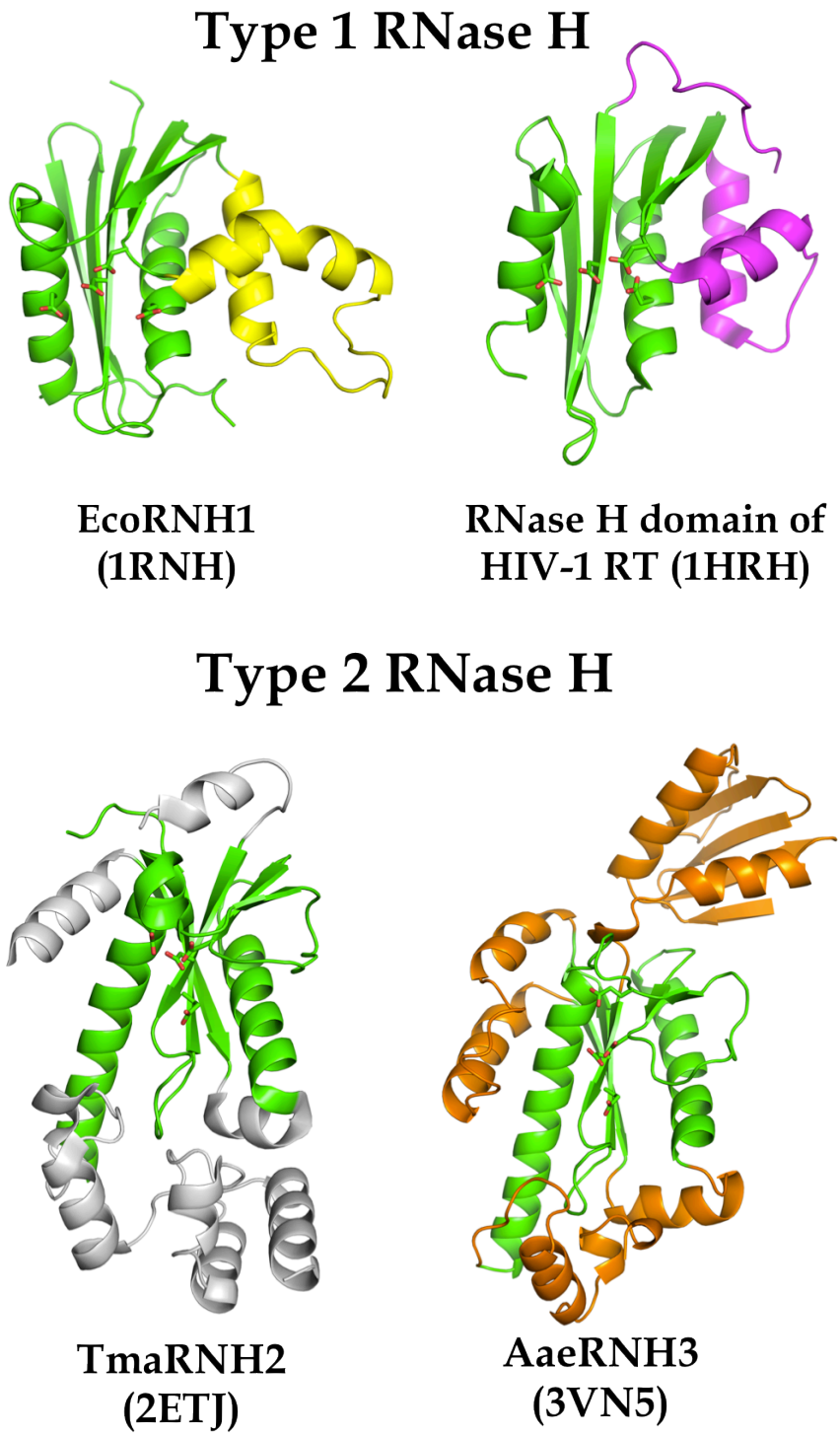


Fig. 1.6. Crystal structures of type 1 and type 2 RNases H.

Overall structures of EcoRNH1 (PDB code 1RNH), RNase H domain of HIV-1 RT (1HRH), TmaRNH2 (2ETJ) and AaeRNH3 (3VN5) are shown. EcoRNH1 and RNase H domain of HIV-1 RT represent type 1 RNases H. TmaRNH2 and AaeRNH3 represent type 2 RNases H. An RNase H-fold is colored green. The four acidic active-site residues are shown in stick models, in which the oxygen atom is colored red.

Four acidic active-site residues are fully conserved in type 1 and type 2 RNases H [4], with a few exceptions [55]. Because they are arranged in the order of Asp, Glu, Asp, and Asp or Glu from the N-terminus, the active-site motif of RNase H is termed DEDD or DEDE [4]. While RNases H1 and H2 have a DEDD active-site motif, RNase H3 has a DEDE active-site motif [4]. Exceptions are the active sites of metagenome-derived LC9-RNase H1 and its homolog [55]. In these RNases H1, the fourth residue of a DEDD active-site motif is replaced by Asn, indicating that these RNases H1 have a DEDN active-site motif. In addition, the first and the second residues of a DEDN active-site motif of these

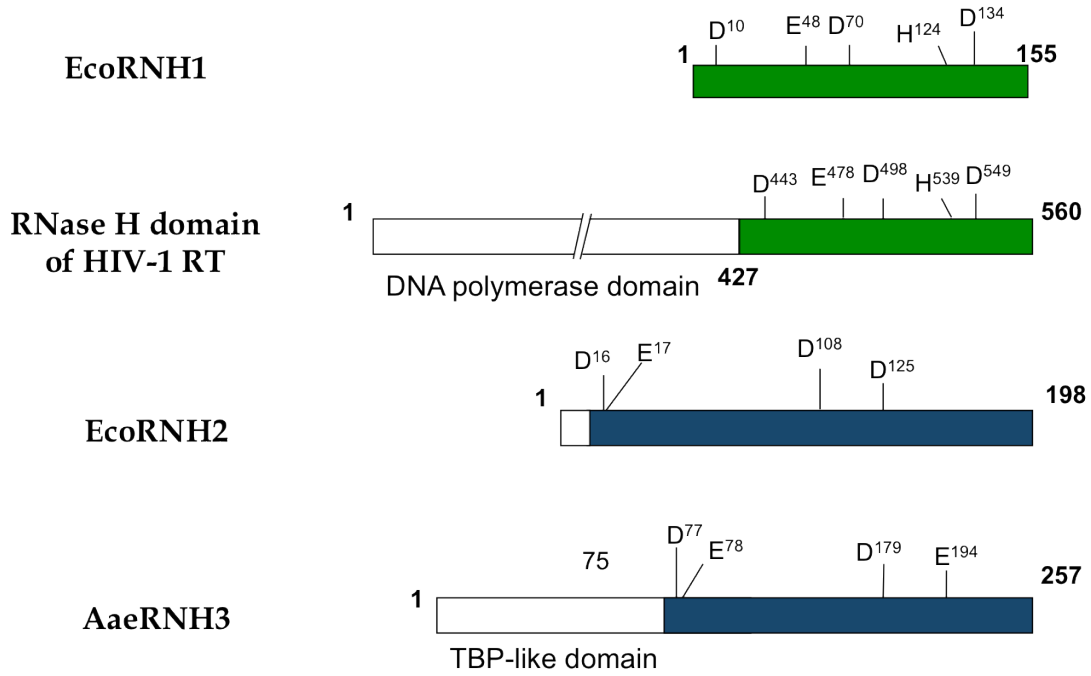


Fig. 1.7. Schematic representation of the primary structures of RNases H.
The primary structures of the representative members of type 1 and type 2 RNases H, which are enzymatically active, are shown. EcoRNH1 and RNase H domain of HIV-1 RT represent type 1 RNases H, while EcoRNH2 and AaeRNH3 represent type 2 RNases H. The catalytic domain of these RNases H are colored green for type 1 RNases H and blue for type 2 RNases H. The positions of the four acidic active-site residues and a histidine residue, which is well conserved in type 1 RNases H, are indicated. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

RNases H1 are adjacent to each other like those of type 2 RNases H. The primary structures of EcoRNH1, RNase H domain of HIV-1 RT, EcoRNH2 and AaeRNH3 are schematically shown in Figure 1.7, together with the positions of the four acidic active-site residues and a histidine residue, which is well conserved in type 1 RNases H. EcoRNH1 represents bacterial, archaeal and eukaryotic RNases H1. The RNase H domain of HIV-1 RT represents retroviral RNases H. EcoRNH2 represents bacterial and archaeal RNases H2, and the catalytic subunits of eukaryotic RNases H2. AaeRNH3 represents bacterial RNases H3. All four residues of a DEDD active-site motif are distantly located with one another in the sequences of type 1 RNases H, whereas the first and second residues of a DEDD or DEDE active-site motif are adjacent to each other in the sequences of type 2 RNases H. The histidine residue is located in a flexible loop near the active site. It has been proposed that this residue is directly involved in the catalytic function as a proton pump [56].

1.5. Catalytic mechanism of RNase H

RNases H1 and H2 share similar steric configurations of the active-site residues, although the positions of these residues in the sequences of type 1 RNases H are different from those in the sequences of type 2 RNases H, suggesting that their catalytic mechanism are similar to each other. A two-metal-ion catalysis mechanism has been proposed for RNase H activity, because the structures of the active-site mutants of BhaRNH1C [41] and HsaRNH1C [42] in complex with the substrate and metal ions contain two divalent metal ions (A and B) at the active site. This mechanism is schematically shown in Figure 1.8. Metal ion A is required for substrate-assisted nucleophile formation. Both metal ions A and B are required to destabilize the enzyme-substrate complex and to promote the

phosphoryl transfer reaction. The distance between metal ions A and B is 4.1 Å when the

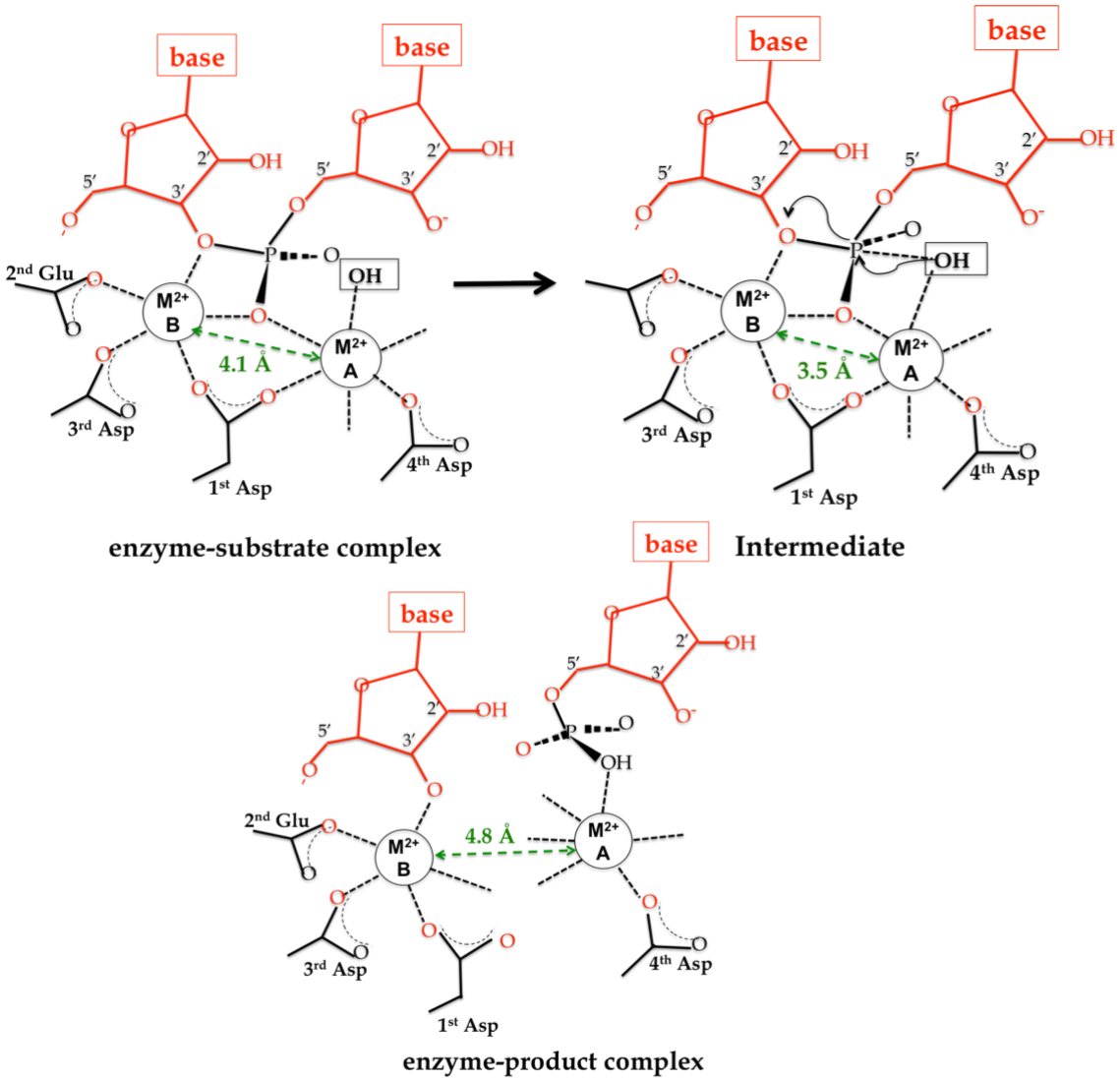


Fig. 1.8. Schematic representation of the two-metal-ion catalysis mechanism of RNase H. The enzyme-substrate complex, an intermediate form, and the enzyme-product complex are schematically shown based on the crystal structures of BhaRNH1C in complex with substrate [41], intermediate-like RNA/DNA duplex [57], and product-like RNA/DNA duplex [57]. The side chains of the four acidic residues that form a DEDD active-site motif, two metal ions (M^{2+} A, M^{2+} B), and a scissile phosphodiester bond of the substrate are shown. The fourth aspartate of a DEDD active-site motif is replaced by Glu and Asn in DEDE and DEDN active-site motifs respectively. Coordination of the metal ions is indicated by black dashed lines. The distances between two metal ions are indicated by green dashed lines. The attacking hydroxyl ion that is coordinated with metal ion A is indicated in the box.

enzyme-substrate complex is formed. However, this distance reduces to 3.5 Å in an intermediate state. Then, this distance increases to 4.8 Å when the P-O3' bond of the substrate is cleaved, probably due to cancellation of a repulsive force between two metal ions.

The number of the metal ions required for activity, however, still remains controversial [58]. The co-crystal structures of BhaRNH1C [41] and HsaRNH1C [42] with the substrate contain two metal ions at the active site, whereas that of TmaRNH2 contains three metal ions at the active site [9]. The number of metal ions bound to the enzyme in a substrate-free form also varies for different metal ions and enzymes. Only single Mg^{2+} ions bind to EcoRNH1 [59] and BstRNH3 [5**3**], whereas two Mn^{2+} ions bind to EcoRNH1 [60] and the RNase H domain of HIV-1 RT [39]. Only single Mn^{2+} ions bind to BstRNH3 [53] and the active site mutant of EcoRNH1 [61].

1.6. Metal preferences of RNase H

RNases H requires divalent metal ions, such as Mg^{2+} and Mn^{2+}, for activity. It exhibits activity in the presence of other metal ions, such as Co^{2+} and Ni^{2+} ions, although these activities are much lower than those determined in the presence of Mg^{2+} or Mn^{2+} ions. RNase H does not exhibit activity in the presence of Ca^{2+} ions, because Ca^{2+} ions usually inhibit activity. However, the most preferable metal ion for activity varies for different enzymes. For example, EcoRNH1 prefers Mg^{2+} to Mn^{2+} ions for RNase H activity [7, 62], whereas EcoRNH2 prefers Mn^{2+} to Mg^{2+} ions for this activity [63]. BsuRNH2 prefers Mn^{2+} to Mg^{2+} ions for RNase H activity, whereas BsuRNH3 prefers Mg^{2+} to Mn^{2+} ions for this activity [7]. TmaRNH2 prefers Mn^{2+} to Mg^{2+} ions for RNase H activity, but does not show this preference for JRNase activity [9]. It exhibits maximal JRNase and

RNase H activities either in the presence of Mg^{2+} or Mn^{2+} ions. BstRNH3 prefers Mg^{2+} to Mn^{2+} ions for RNase H activity [53]. AaeRNH3 does not show this preference [14]. It exhibits maximal RNase H activity either in the presence of Mg^{2+} or Mn^{2+} ions. SceRNH2 prefers Mn^{2+} to Mg^{2+} ions for JRNase activity [64]. In this thesis, the RNase H/JRNase activities determined in the presence of Mg^{2+} or Mn^{2+} ion are designated as Mg^{2+}-dependent RNase H/JRNase activities or Mn^{2+}-dependent RNase H/JRNase activities. The reason why the most preferable metal ion for activity varies for different RNases H remains to be fully understood. Because Mg^{2+} and Mn^{2+} ions are alkaline-earth and transition metal ions respectively, differ in radii, and their coordination geometries are different with each other, slight difference in the configurations of four acidic active-site residues, that coordinates with these metal ions, may be responsible for the difference in metal preference of RNases H.

In the bacterial cells, the concentration of Mg^{2+} ions is 100-fold higher than that of Mn^{2+} ions [65]. However, this difference in cellular concentrations of Mg^{2+} and Mn^{2+} ions does not necessarily indicate that Mg^{2+}-dependent activity is physiologically more important than Mn^{2+}-dependent activity, because the binding of these metal ions to the enzyme are not always similar to each other. For example, Mn^{2+} ions bind to EcoRNH1 much more strongly than Mg^{2+} ions [66], and as a result, EcoRNH1 exhibits maximal Mn^{2+}- and Mg^{2+}-dependent activities in the presence of 0.1 μM $MnCl_2$ and 5 mM $MgCl_2$ [67].

1.7. Substrate binding domains of RNase H

Several RNases H contain a substrate binding domain at the N-terminus, C-terminus or middle of the RNase H domain. The amino acid sequences of representative members of type 1 and type 2 RNases H are shown in Figure 1.9. EcoRNH1 and

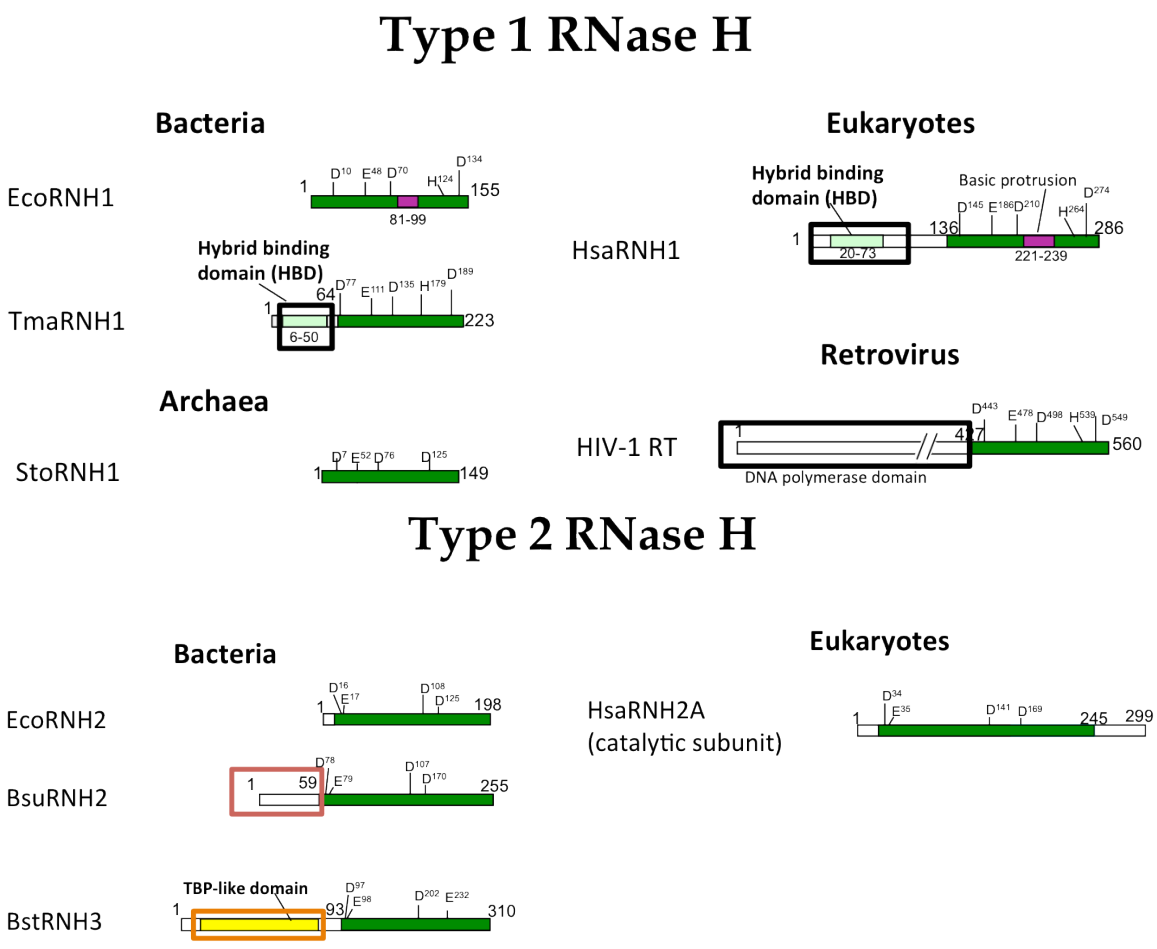


Fig. 1.9. Schematic representation of the primary structures of RNases H.

The primary structures of representative members of RNases H containing a substrate binding domain are shown together with those of RNases H without substrate binding domain. The RNase H domain is colored green. A basic protrusion of type 1 RNases H is colored magenta. HBD is colored cyan. The TBP-like domain of RNase H3 is colored yellow. The active site residues are indicated. The number represents the position of each amino acid residue relative to the N-terminal methionine residue.

HsaRNH1 have a substrate binding domain, termed basic protrusion, at the middle of the RNase H domain. TmaRNH1 and HsaRNH1 have a substrate binding domain, termed hybrid binding domain (HBD), at the N-terminus. AaeRNH3 has a substrate binding domain, termed TBP-like domain, at the N-terminus. HIV-1 RT has an N-terminal DNA polymerase domain. The RNase H domain of HIV-1 RT is correctly folded [39], but loses activity [39, 68-71] or exhibits very weak Mn^{2+}-dependent activity [72], when it is separated from the polymerase domain, indicating that the polymerase domain is required for binding of the RNase H domain to substrate. Because RT always consists of N-terminal polymerase domain and C-terminal RNase H domain, the RNase H domain of RT probably always requires the polymerase domain for substrate binding. However, RNases H do not always contain these substrate binding domains. For example, StoRNH1 and EcoRNH2 contain none of these substrate binding domains (Fig. 1.9). These RNases H may have a unique mechanism by which they efficiently binds to the substrate without the assistance of a substrate binding domain. Importance of a basic protrusion for substrate binding has been shown by the mutational studies of EcoRNH1 [73-75]. Later, the co-crystal structure of HsaRNH1C with the substrate has revealed that a basic protrusion provides DNA binding channel, to which the DNA strand of RNA/DNA hybrid binds [42].

HBD is present in almost all eukaryotic RNases H1 [2], including HsaRNH1, and several bacteria RNases H1, such as TmaRNH1 [76], BhaRNH1 [41] and RBD-RNase H1 from *Shewanella* sp. SIB1 [77]. HBD has been originally termed as dsRNA binding domain (RBD). However, this domain has been shown to bind more strongly to RNA/DNA hybrid than to dsRNA [78]. The amino acid sequences of HBDs of several bacterial and eukaryotic RNases H1 are compared with one another in Figure 1.10. These HBDs show relatively high amino acid sequence similarities with one another. For

example, HBD of TmaRNH1 (TmaHBD) shows the amino acid sequence identities of 43, 42, 33 and 32% to those of SceRNH1, BhaRNH1, SIB1 RBD-RNase H1 and HsaRNH1

Fig. 1.10. Alignment of the amino acid sequences [76].

The amino acid sequences of N-terminal HBDs of several RNases H1 are shown. The HBD sequence of TmaRNH1 (Tma) is compared with those of BhaRNH1 (Bha), SIB1 RBD-RNase H1 (SIB1RBD), SceRNH1 (Sce) and HsaRNH1 (Hsa). The secondary structures of HsaRNH1 (PDB code 3BSU) are shown above the sequences. The amino acid residues, which are highly conserved, are highlighted in black. The amino acid residues that contact the substrate according to the co-crystal structure of HBD of HsaRNH1 with substrate [78] are indicated by open circles. The numbers represent the positions of the amino acid residues to the initiator methionine for each protein.

respectively. The removal of HBD decreases the substrate binding affinity of TmaRNH1 by 50-fold and eliminates its Mg^{2+}-dependent activity and in vivo function without significantly affecting its Mn^{2+}-dependent activity and stability [76]. The removal of HBD also reduces Mg^{2+}-dependent activity of HsaRNH1 without significantly affecting its Mn^{2+}-dependent activity [42] and eliminates its processivity [79]. The substrate binding mechanism of HBD will be described in the next section (section 1.8).

TBP-like domain is always present at the N-terminus of RNase H3. The amino acid sequences of TBP-like domains of several RNases H3 are compared with one another in Figure 1.11. These TBP-like domains do not always show a relatively high amino acid sequence similarities with one another. For example, TBP-like domain of BstRNH3 (BstTBP) shows the amino acid sequence identities of 47, 16 and 19% to those of

BsuRNH3, AaeRNH3 and TamRNH3 respectively. The removal of TBP-like domain greatly reduces substrate binding affinities and Mg^{2+}-dependent activities of BstRNH3 [53] and AaeRNH3 [14] without significantly affecting their Mn^{2+}-dependent activities. Similar

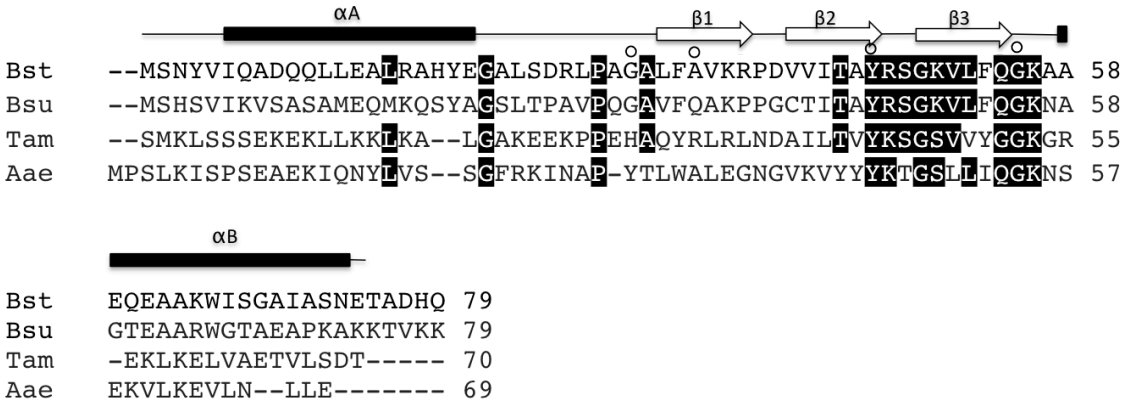


Fig. 1.11. Alignment of the amino acid sequences

The amino acid sequences of TBP-like domains of several RNases H3 are shown. The sequence of TBP-like domain of BstRNH3 (Bst) is compared with those of BsuRNH3 (Bsu), TamRNH3 (Tam) and AaeRNH3 (Aae). The secondary structures of TamRNH3 (PDB code 4PY5) are shown above the sequences. The amino acid residues, which are highly conserved, are highlighted in black. The amino acid residues that contact the substrate according to the co-crystal structure of TamRNH3 with substrate [54] are indicated by open circles. The numbers represent the positions of the amino acid residues to the initiator methionine for each protein.

results have been reported for EcoRNH1 [66, 67, 80] and HIV-1 RNase H domain [81], in which mutation or deletion reduces Mg^{2+}-dependent activity without significantly affecting Mn^{2+}-dependent activity. These results suggest that slight conformational change of the active site does not significantly affect the binding of Mn^{2+} ions but affects binding of Mg^{2+} ions. The substrate binding mechanism of TBP-like domain will be described in the next section (section 1.8).

It remains to be determined whether RNase H2 has a substrate binding domain. The result that HBD and TBP-like domain are only present in RNases H1 and H3 respectively, indicates that RNase H2 does not contain these domains. However, BsuRNH2 has an N-terminal extension (N-extension) as compared to EcoRNH2 (Fig. 1.9). The amino acid sequence of this extension does not show a similarity to any sequence available from database, suggesting that this extension is folded into a unique structure. However, it remains to be determined whether this extension is involved in substrate binding and is conserved in other RNases H2. It also remain to be determined whether RNase H2 has another N- or C-terminal extension that functions as a substrate binding domain.

1.8. Substrate recognition mechanism of RNase H

To reveal mechanism by which RNase H1 recognizes RNA/DNA hybrid as a substrate, the co-crystal structures of BhaRNH1C [41] and HsaRNH1C [42] with RNA/DNA hybrid have been determined. The co-crystal structure of HsaRNH1C with 18bp RNA/DNA hybrid is shown in Figure 1.12. According to this structure, RNA/DNA hybrid binds to the two grooves on the protein surface, which is separated by a ridge, in such a way that the RNA strand fits one groove containing the active site and the DNA strand fits the other groove containing the DNA phosphate binding pocket. The RNA strand contacts the protein in such a way that the 2'-OH groups of two ribonucleotides on each side of the scissile phosphate group form hydrogen bonds with the protein. This is the major reason why RNases H1 does not recognize dsDNA as a substrate. The DNA strand contacts not only the groove containing the phosphate binding pocket but also the DNA binding channel formed in the basic protrusion. The DNA binding channel is formed by Trp221, Trp225 and Ser233. The RNA strand cannot fit this channel because the 2'-OH

group of the RNA strand causes a steric crash with Trp221. This is one of the reasons why RNase H1 does not recognize dsRNA as a substrate. The minor groove width of RNA/DNA hybrid varies from 8 to 10 Å at the region where it contacts the protein. This width is larger than that of dsDNA with a typical B-form conformation (7 Å) and smaller than dsRNA with a typical A-form conformation (11 Å). This is another reason why RNase H does not recognize dsRNA as a substrate.

The crystal structure of eukaryotic RNase H1 or bacterial RNase H1 containing HBD in the intact form remains to be determined. Therefore, it remains to be understood

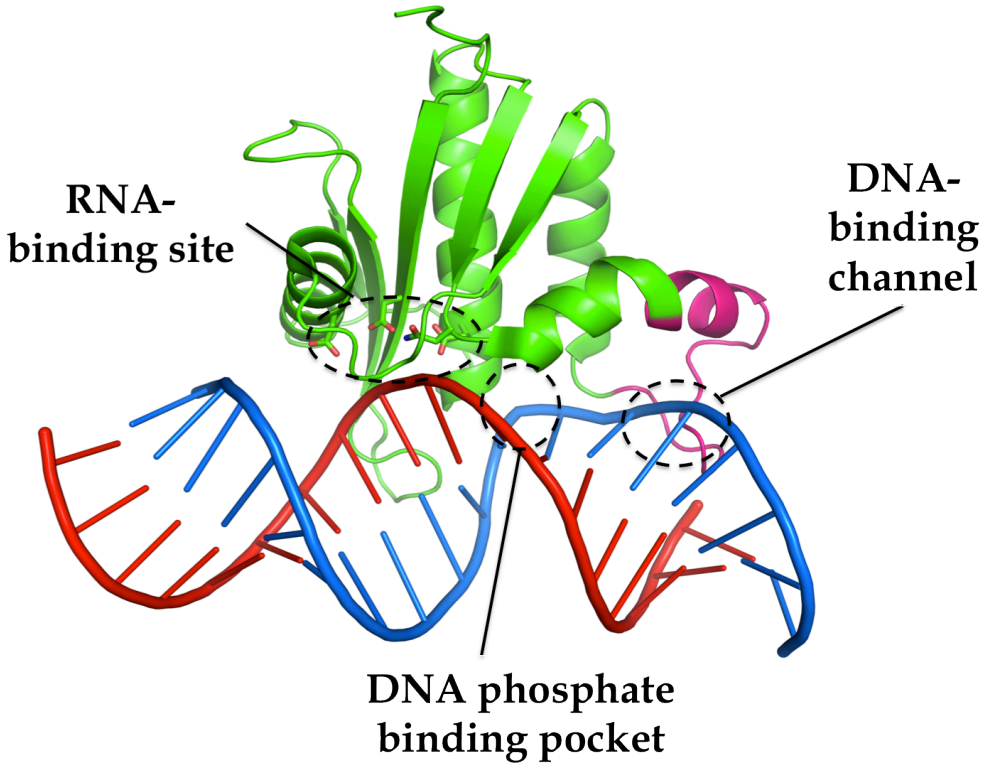


Fig. 1.12. Co-crystal structure of HsaRNH1C with substrate

The crystal structure of HsaRNH1C in complex with 18bp RNA/DNA hybrid [PDB code 2QK9] is shown. DNA is colored blue and RNA is colored red. The RNA binding site, DNA phosphate-binding pocket and DNA-binding channel are indicated. The basic protrusion is indicated magenta. The DNA-binding channel is absent in BhaRNH1C because BhaRNH1C lacks basic protrusion. The active site residues are shown in stick models, in which the oxygen atoms are colored red.

how HBD and the catalytic domain bind to the substrate at the same time. However, the crystal structure of the complex between HBD of HsaRNH1 in an isolated form and RNA/DNA hybrid has been determined [78]. This structure is shown in Figure 1.13-A.

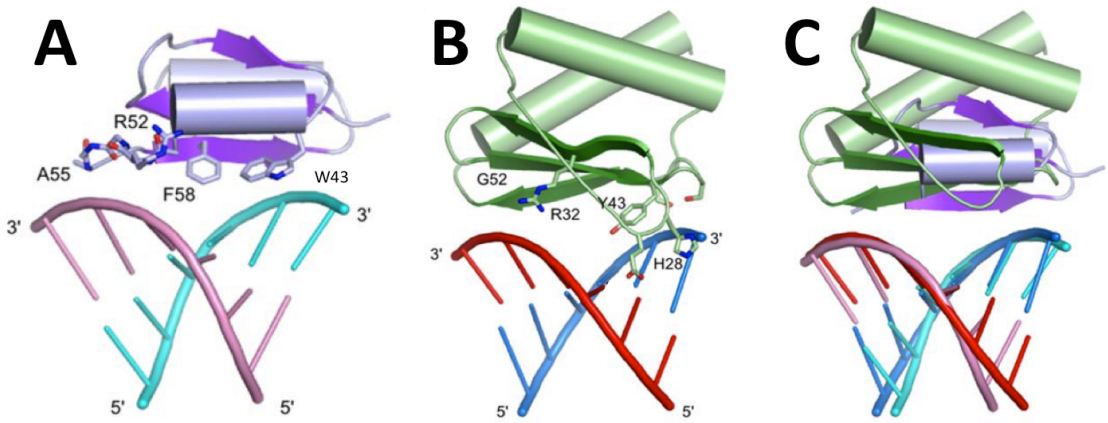


Fig. 1.13. Co-crystal structures of HBD and TBP-like domain with RNA/DNA hybrid. The crystal structures of HBD of HsaRNH1 in an isolated form (PDB code: 3BSU) (A) and TBP-like domain of TamRNH3 (PDB code: 4PY5) (B) in complex with RNA/DNA hybrid are shown. The structure of the complex between TBP-like domain and RNA/DNA hybrid represents the structure of the complex between TamRNH3 and RNA/DNA hybrid around TBP-like domain. The DNA and RNA strands are colored cyan and pink for HsaRNH1, and blue and red for TamRNH3 respectively. The amino acid residues that contact the substrate are shown by stick models, in which the oxygen and nitrogen atom are colored red and blue respectively in panels A and B. The structures shown in panel A is superimposed on that shown in panel B in panel C.

According to this structure, HBD consists of a three-stranded anti-parallel β-sheet (β1- β3) and two α-helices (αA and αB). HBD contacts the minor groove of the substrate mainly at αA helix and β3 strand. RNA-specific interaction is formed between 2'-OH groups of two adjacent RNA nucleotides and two backbone nitrogen atoms of Arg52 and Ala55. The DNA-specific stacking interaction is formed between the deoxyribose rings of two DNA

nucleotides and aromatic rings of Trp43 and Phe58. Trp43 of mouse RNase H1 [79] and Trp22 of TmaRNH1 [76], which correspond to Trp43 of HsaRNH1, have been shown to be important for substrate binding.

To reveal mechanism by which TBP-like domain binds to RNA/DNA hybrid, the co-crystal structure of TamRNH3 with RNA/DNA hybrid has been determined [54]. This structure around TBP-like domain is shown in Figure 1.13-B. According to this structure, TBP-like domain consists of a three-stranded anti-parallel β-sheet (β1- β3) and two α-helices (αA and αB), like HBD. However, the arrangement of two α-helices in the structure of TBP-like domain is different from that in the HBD structure. Two α-helices are located in one side of a β-sheet in TBP-like domain, whereas they are located in both sides of a β-sheet in HBD. TBP-like domain contacts the minor groove of the substrate using a flat surface of the β-sheet. Nevertheless, TBP-like domain contacts the substrate in a similar manner to that of HBD (Fig. 1.13-C). RNA-specific interaction is formed between 2'-OH groups of two adjacent RNA nucleotides and two nitrogen atoms of the protein (side chain nitrogen atom of Arg32 and backbone nitrogen atom of Gly52). The DNA-specific stacking interaction is formed between the deoxyribose rings of two DNA nucleotides and aromatic rings of His28 and Tyr43. Tyr46 of BstRNH3 [82] and Tyr45 of AaeRNH3 [14], which correspond to Tyr43 of TamRNH3, have been shown to be important for substrate binding. This may be the reason why RNase H3 functions as a substitute of RNase H1, as proposed previously [7]. The structures of BstRNH3 [53] and AaeRNH3 [14] in a substrate-free form have also been determined. When these structures are superimposed on that of TamRNH3 in complex with the substrate, the structures of TBP-like domain of BstRNH3 and AaeRNH3 are nearly identical to that of TamRNH3, but their spatial localizations are quite different from that of TamRNH3. This result suggests that TBP-like

domain of BstRNH3 and AaeRNH3 do not contact the substrate unless they greatly change the conformation, as proposed previously [82].

To reveal mechanism by which RNase H2 recognizes dsDNA^{R1} as a substrate, the co-crystal structure of TmaRNH2 with DNA_5-RNA_1-DNA_6/DNA_{12} (D_5-R_1-D_6/D_{12}) duplex has been determined [9]. This structure is shown in Figure 1.14. According to this structure, TmaRNH2 consists of the N-terminal catalytic domain and C-terminal domain. The cleaved strand of the D5-R1-D6/D12 substrate containing single ribonucleotide mostly contacts the catalytic domain, whereas the noncleaved strand mostly contacts the C-

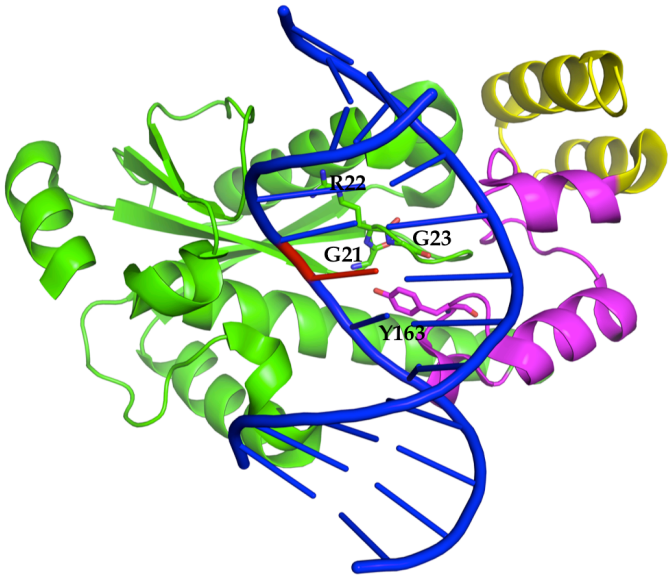


Fig. 1.14. Crystal structure of TmaRNH2 in complex with dsDNA^{R1}

The crystal structures of TmaRNH2 in complex with dsDNA^{R1} (PDB code: 3O3F) is shown. The DNA and RNA strands are colored blue and red respectively. The catalytic domain, C-terminal domain and C-terminal hairpin helix structure are colored green, magenta and yellow respectively. The amino acid residues that interact with the 2'-OH groups of ribonucleotide are shown by stick models, in which the oxygen and nitrogen atom are colored red and blue, respectively.

terminal domain. The 2'-OH group of the single ribonucleotide interacts with the backbone nitrogen atoms of Gly21, Arg22 and Gly23, and side chain OH group of Tyr163. The GRG residues are conserved in RNases H2 and are termed GRG motif. The tyrosine residue is also conserved in RNases H2. The aromatic ring of Tyr163 stacks with the deoxyribose ring of the deoxyribonucleotide at position +2, leading to the deformation of the nucleic acid backbone at the (5')DNA-RNA(3') junction located between positions +1 and +2. This deformation is necessary for the cleavage of dsDNA^{R1} at the (5')DNA-RNA(3') junction. The stacking interaction between the deoxyribonucleotide at position +2 and the conserved tyrosine residue is not formed when deoxyribonucleotide at position +2 is replaced by ribonuleotide, due to the presence of the 2'-OH group. This is the reason why RNase H2 cleaves not only dsDNA^{R1} but also RNA-DNA/DNA substrate at the 5'-side of the ribonucleotide attached to DNA.

1.9. Objective of the study

As described in section 1.7, it remains to be determined whether RNase H2 has a substrate binding domain, although BsuRNH2 has N-extension, which may function as a substrate binding domain. Therefore, the purpose of this study is to identify a novel substrate binding domain of RNase H. To examine whether bacterial RNases H2 have N-extension or C-extension that may function as a substrate binding domain, I compared the amino acid sequences of various bacterial RNases H2 with one another at first. I did not compare the amino acid sequences of the catalytic subunits of eukaryotic RNases H2, because these subunits alone are inactive and require subunits B and C for activity. Comparison of the amino acid sequences of bacterial RNases H2 indicates that some of them have N-extension or C-extension. Some of them have neither extension. None of

them has both N- and C-extensions. For example, BstRNH2 and TmaRNH2 have N-extension and C-extension respectively. AaeRNH2 has neither extension. Because the functions of these extensions still remain unknown, I constructed the BstRNH2 and TmaRNH2 derivatives lacking N-extension and C-extension respectively, and characterized them. I also characterized AaeRNH2 and compared its biochemical properties with those of BstRNH2 and TmaRNH2 to examine whether AaeRNH2 without N- and C-extensions is functional like BstRNH2 and TmaRNH2. Finally, I examined whether substrate binding domains of bacterial RNases H function as a tag for binding of heterologous proteins to RNA/DNA hybrid by constructing the fusion proteins, in which HBD of TmaRNH1 (TmaHBD) or N-extension of BstRNH2 (BstNTD) is attached to the N-terminus of RNH^{HIV} (RNase H domain of HIV-1 RT).

The thesis consists of four chapters. In Chapter 1, I summarize the background of this work.

In Chapter 2, I searched database for bacterial RNases H2 with N- or C-extension. I found that bacterial RNases H2 are classified into three groups, that contain N-extension, C-extension, or neither extension. To examine whether bacterial RNases H2 without these extensions are functional like those containing either N- or C-extension, I chose AaeRNH2, BstRNH2 and TmaRNH2 as representative members of bacterial RNases H2 that contain neither extension, N-extension and C-extension respectively and characterized them. I showed AaeRNH2 exhibited comparable JRNase activity as those of BstRNH2 and TmaRNH2, and was more stable than BstRNH2 and TmaRNH2. Then, to analyze the role of N- and C-extensions of BstRNH2 and TmaRNH2, I constructed the derivatives of BstRNH2 and TmaRNH2 lacking N- and C-extensions respectively, and characterized them. I showed that JRNase activities and RNase H activities of these derivatives as well

as their substrate binding affinities, were decreased as compared to those of their intact partners. These results suggest that N-extension of BstRNH2 and C-extension of TmaRNH2 are important for substrate binding and activity. Moreover, I constructed a fusion protein, BstNTD-TmaRNH2ΔC, in which N-extension of BstRNH2 (BstNTD) is attached to the N-terminus of TmaRNH2 lacking C-extension (TmaRNH2ΔC), and characterized it. I showed that BstNTD partly restored the stability and activity of TmaRNH2ΔC. Thus, BstNTD functions as a potent substrate binding domain.

In Chapter 3, I constructed two fusion proteins, BstNTD-RNH^{HIV} and TmaHBD-RNH^{HIV}, in which BstNTD and TmaHBD are attached to the N-terminus of RNH^{HIV}, and characterized them, to examine whether substrate binding domains of bacterial RNases H function as a tag for binding of heterologous proteins to RNA/DNA hybrid. As described in section 1.7, RNH^{HIV} is inactive or exhibits very weak Mn^{2+}-dependent RNase H activity. I showed that BstNTD and TmaHBD increased substrate binding affinity of RNH^{HIV} and the Mn^{2+}-dependent RNase H activity of RNH^{HIV} without significantly affecting its Mg^{2+}-dependent RNase H activity. This result indicates that BstNTD and TmaHBD functions as a substrate binding domain in an isolated form and can be used as a tag for binding of heterologous proteins to RNA/DNA hybrid.

In Chapter 4, I summarize this work and describe the future remarks.

CHAPTER 2

Role of N-terminal extension of *Bacillus stearothermophilus* RNase H2 and C-terminal extension of *Thermotoga maritima* RNase H2

2.1. Introduction

As described in Chapter 1, RNases H often have a substrate binding domain, although they do not always have it. As shown in Figure 1.9, EcoRNH1 and HsaRNH1 have substrate binding domains, termed basic protrusions, at the middle of the catalytic domains. TmaRNH1 and HsaRNH1 have HBDs at the N-terminus. BstRNH3 has a TBP-like domain at the N-terminus. StoRNH1 and EcoRNH2 do not have a substrate binding domain. However, it remains to be determined whether RNase H2 has a substrate binding domain. It has been reported that BsuRNH2 has a relatively long N-terminal extension (N-extension) as compared to EcoRNH2 [7]. Because the amino acid sequence of this extension does not show a significant similarity to that of HBD or TBP-like domain, this extension is probably folded into a unique structure. However, the role of this extension remains to be analyzed. In addition, it remains to be determined whether this extension is conserved in other RNases H2 and whether RNases H2 with another N- or C-terminal extension exist in nature. These extensions may be folded into a relatively independent structure and may function as a substrate binding domain. Therefore, I decided to compare the amino acid sequences of various bacterial RNases H2 to identify these extensions and analyze their roles. I did not compare the amino acid sequences of eukaryotic RNases H2, because eukaryotic RNase H2 is heterotrimeric and a catalytic subunit requires other two subunits for activity [2].

In this Chapter, I showed that bacterial RNases H2 are classified into three groups that contain N-extension, C-extension and neither extension. I chose BstRNH2, TmaRNH2 and AaeRNH2 as representative members of RNases H2 that contain N-extension, C-extension and neither extension respectively, and characterized them. I showed that AaeRNH2 requires neither extension for activity and stability. I also showed that N-extension of BstRNH2 and C-extension of TmaRNH2 are important for activity, substrate binding and stability. Based on these results, I discuss the roles of N- and C-extensions of bacterial RNases H2.

2.2 Materials and methods

2.2.1. Plasmid construction

Plasmid pET800TM and pET600AA for overproduction of TmaRNH2 and AaeRNH2, respectively, were constructed by polymerase chain reaction (PCR) using the genomic DNA of *T. maritima* and *A. aeolicus* as a template. Plasmid pET800BS for overproduction of BstRNH2 was constructed by ligating the *Nde*I-*Sal*I fragment of pJAL800ST containing the BstRNH2 gene, which was firstly constructed from the genomic DNA of *B. stearothermophilus* CU21, into the *Nde*I-*Sal*I sites of pET25b (Novagen, Madison, WI, USA).

Plasmids pET670TMΔC and pET620BSΔN for overproduction of TmaRNH2ΔC and BstRNH2ΔN, respectively, were constructed by PCR using the KOD-Plus mutagenesis kit (Toyobo, Kyoto, Japan) according to the manufacturer's instructions. Plasmids pET800TM and pET800BS were used as a template. The mutagenic primers were designed such that the C-terminal 46 residues of TmaRNH2 and N-terminal 59 residues of BstRNH2 are removed. Plasmid pET850BSTMΔC for overproduction of BstNTD-TmaRNH2ΔC was constructed by the overlap PCR extension method as described previously for the construction of the plasmid for complementation assay [7]. The DNA oligomers for PCR were synthesized by Hokkaido System Science (Sapporo, Japan). PCR was performed using a GeneAmp PCR system 2400 (Applied Biosystems, Tokyo, Japan). All DNA sequences were confirmed by ABI Prism 310 DNA sequencer (Applied Biosystems).

2.2.2. Overproduction and purification

I used *E. coli* MIC2067(DE3), which lacks the *rnh*A and *rnh*B genes, as a host strain to overproduce the protein and to avoid contamination of RNases H1 and H2 from host-cells [14]. *E. coli* MIC2067(DE3) transformants were grown at 30°C in Luria-Bertani (LB) medium containing 30 mg·L^{-1} chloramphenicol and 50 mg·L^{-1} ampicillin.

For purification of AaeRNH2, cells were harvested by centrifugation at 3,000 *g* for 30 min, suspend in 20 mM sodium acetate (pH 5.5) containing 1 mM EDTA (buffer A), disrupted by French press lysis, and centrifuged at 30,000 *g* for 30 min. The supernatant was dialyzed against buffer A and applied into a Hitrap Q HP column (GE Healthcare, Tokyo, Japan) equilibrated with buffer A. The flow-through containing the protein was pooled and applied to a Hitrap SP HP column equilibrated with the same buffer. The protein was eluted from the column with a linear gradient of NaCl from 0 to 1 M in the same buffer. The fractions containing the protein were collected, dialyzed against buffer A, and loaded into a Hitrap Heparin HP 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 in the same buffer. The final fractions containing the protein were collected and dialyzed against 20 mM sodium acetate (pH 5.5).

TmaRNH2 was overproduced and purified, in which cells were harvested by centrifugation at 3,000 *g* for 30 min, suspend in 20 mM Tris-HCl (pH 8.0) containing 1 mM EDTA (buffer A), disrupted by sonication lysis, and centrifuged at 30,000 *g* for 30 min. The supernatant was subjected to a heat treatment at 75°C for 10 min and centrifuged at 30,000 *g* for 30 min. Then, it was purified as described for AaeRNH2, with slight modifications. It was purified by two, instead of three, chromatographic procedures with a HiTrap SP column and a Hitrap Heparin column. These columns were equilibrated with 20

mM Tris-HCl (pH 8.0) containing 1 mM EDTA, instead of 20 mM sodium acetate (pH 5.5) containing 1 mM EDTA.

TmaRNH2ΔC was overproduced and purified as described for TmaRNH2, except that the heat treatment and chromatography using a HiTrap SP column were not performed, and the gel filtration chromatography using a HiLoad 16/60 Superdex 200pg column (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 8.0) containing 150 mM NaCl was newly performed at the end of the purification procedures. BstNTD-TmaRNH2ΔC was purified as described for TmaRNH2ΔC, except that the chromatography using a HiTrap Q HP column was newly performed at the beginning of the chromatographic procedures and the concentration and pH of the Tris-HCl buffer were changed to 10 mM and pH 7.0, respectively.

BstRNH2 and BstRNH2ΔN were overproduced as described for AaeRNH2, except that the columns of HiTrap Q HP, Hitrap SP, and HiTrap Heparin were equilibrated with 10 mM Tris-HCl (pH 7.0) containing 1 mM EDTA, and a HiLoad 16/60 Superdex 200pg column equilibrated with 10 mM Tris-HCl (pH 7.0) containing 150 mM NaCl were newly performed at the end of the purification procedures.

The purity of the protein was analyzed by SDS/PAGE using a 15% polyacrylamide gel [83], followed by staining with Coomassie Brilliant Blue (CBB). The protein concentration was determined from UV absorption using an A_{280} value of 0.1% solution (1.0 mg·mL^{-1}) of 0.88 cm^{-1} for AaeRNH2, 0.24 cm^{-1} for TmaRNH2, 0.30 cm^{-1} for TmaRNH2ΔC, 0.65 cm^{-1} for BstNTD-TmaRNH2ΔC, 0.91 cm^{-1} for BstRNH2, 0.66 cm^{-1} for BstRNH2ΔN. These values were calculated by using absorption coefficients of 1576 M^{-1}·cm^{-1} for Tyr and 5225 M^{-1}·cm^{-1} for Trp at 280 nm [84].

2.2.3. Enzymatic activity

The JRNase and RNase H activities were determined by using 29-bp DNA_{15}-RNA_1-DNA_{13}/DNA_{29} (D15-R1-D13/D29) and 12-bp RNA_{12}/DNA_{12} (R12/D12) as a substrate respectively. These substrates were prepared by hybridizing 1 μM of 5'-fluorescence-labeled DNA_{15}-RNA_1-DNA_{13} (5'-AATAGAGAAAAAGAAaAAAGATGGCAAAG-3') and RNA_{12} (5'-cggagaugacgg-3') with a 1.5 molar equivalent of the complementary DNA, as described previously [7]. In these sequences, DNA and RNA are represented by uppercase and lowercase letters, respectively. All oligonucleotides were synthesized by Hokkaido System Science. The cleavage of the substrate at 30°C for 15 min and separation of the products on a 20% polyacrylamide gel containing 7 M urea were carried out as described previously [7]. The products were detected by Typhoon 9210 Imager (GE Healthcare) and quantified by ImageQuant 5.2. One unit was defined as the amount of enzyme hydrolyzing 1 nmol of the substrate per min at 30°C. The specific activity was defined as the enzymatic activity per milligram of protein.

The reaction buffers were 10 mM Tris-HCl (pH 8.0) containing 10 mM $MnCl_2$ or $MgCl_2$, 10 mM (for BstRNH2ΔN) or 50 mM (for BstRNH2) NaCl, 1 mM 2-mercaptoethanol (2-ME), and 10 $\mu g \cdot mL^{-1}$ bovine serume albumin (BSA) for BstRNH2 and BstRNH2ΔN; 15 mM Tris-HCl (pH 8.0) containing 1 mM $MnCl_2$ or $MgCl_2$, 50 mM NaCl, 1 mM DTT, and 100 $\mu g \cdot mL^{-1}$ BSA for TmaRNH2, Tma-RNH2ΔC, and BstNTD-Tma-RNH2ΔC; 20 mM Tris-HCl (pH 8.5) containing 10 mM $MnCl_2$ or $MgCl_2$, 100 mM KCl, 1 mM 2-ME, and 100 $\mu g \cdot mL^{-1}$ BSA for AaeRNH2.

2.2.4. CD Spectra

The far-UV (200-260 nm) CD Spectra were measured on a J-725 spectropolarimeter (Japan Spectroscopic, Tokyo, Japan) at 25°C. The protein was dissolved in 10 mM Tris-HCl (pH 8.0) at the concentration of approximately 0.1 $mg{\cdot}mL^{-1}$. A cell with an optical path length of 2 mm was used. The mean residue ellipticity, θ, which has the units of $deg{\cdot}cm^{-2}{\cdot}dmol^{-1}$, was calculated by using a mean amino acid molecular mass of 110 Da.

2.2.5. Analysis of binding to substrate

Binding of the protein to the substrate was analyzed in the absence of the metal cofactor using a Biacore X instrument (Biacore, Uppsala, Sweden) as described previously [76], except that the substrate used to determine the RNase H^{R1} activity (D15-R1-D13/D29) was immobilized on the sensor chip. The protein was dissolved in 10 mM Tris-HCl (pH 8.0) containing 50 mM NaCl, 1 mM 2-ME, 1mM EDTA, 0.005% Tween P20 at various concentrations, and injected onto the sensor chip at 25°C at a flow rate of 20 $\mu L{\cdot}min^{-1}$. The binding surface was regenerated by washing with 2M NaCl. The association constant K_A was estimated from the equilibrium levels of the protein binding to the surface as previously described [14].

2.2.6. Thermal denaturation

Thermal denaturation curves of the proteins were obtained by monitoring the change in CD values at 222 nm as the temperature was increased. The protein was dissolved in 10 mM Tris-HCl (pH 8.0) containing 1.5 M GdnHCl. The protein concentration and optical path length were 0.1 $mg{\cdot}mL^{-1}$ and 2 mm, respectively. The

temperature of the protein solution was linearly increased by approximately 1.0°C·min^{-1}. Thermal denaturation of these proteins was reversible in this condition. The temperature of the mid-point of the transition, T_m, was calculated by curve fitting of the resultant CD values versus temperature data on the basis of a least-squares analysis.

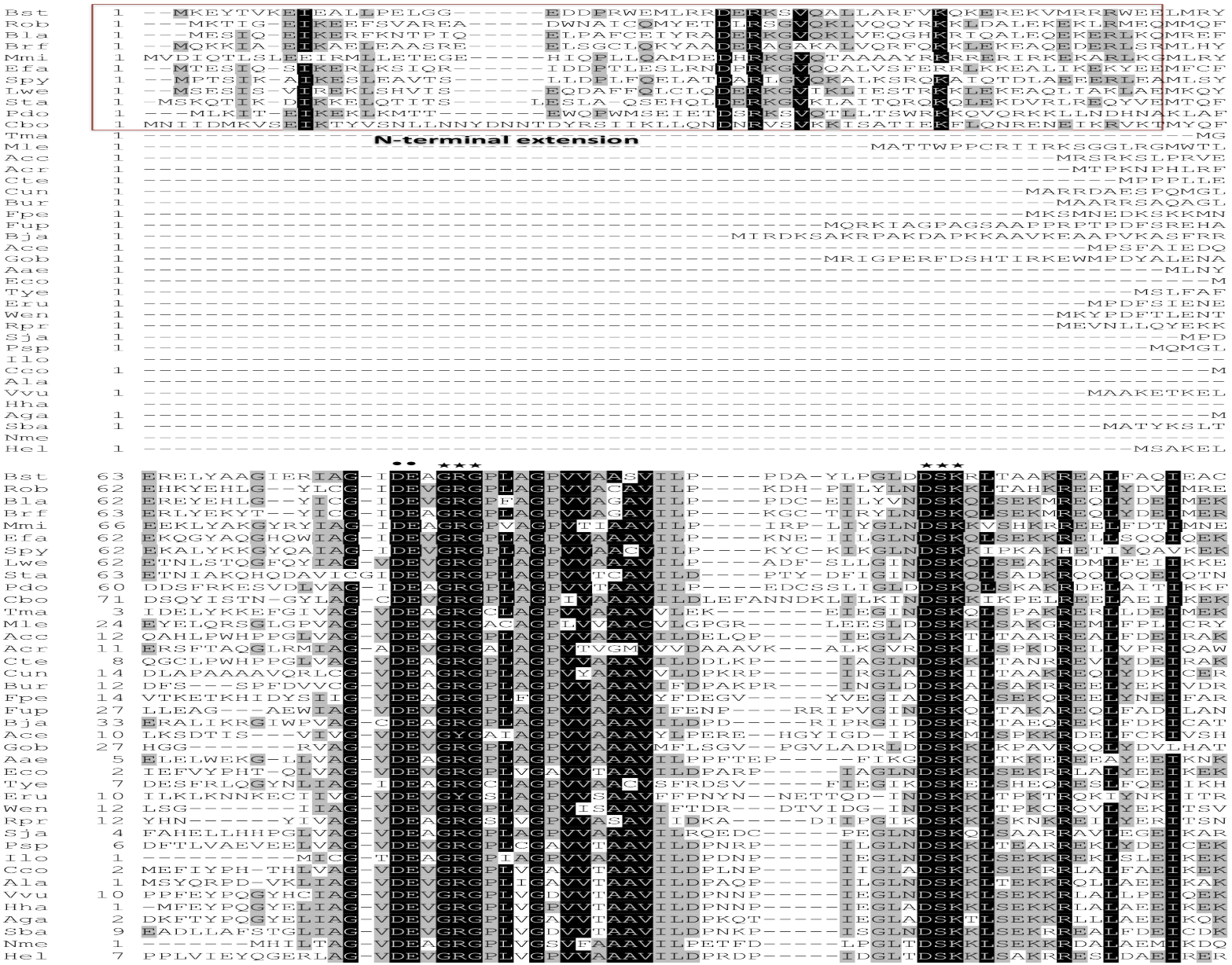
2.2.7. 3D modeling

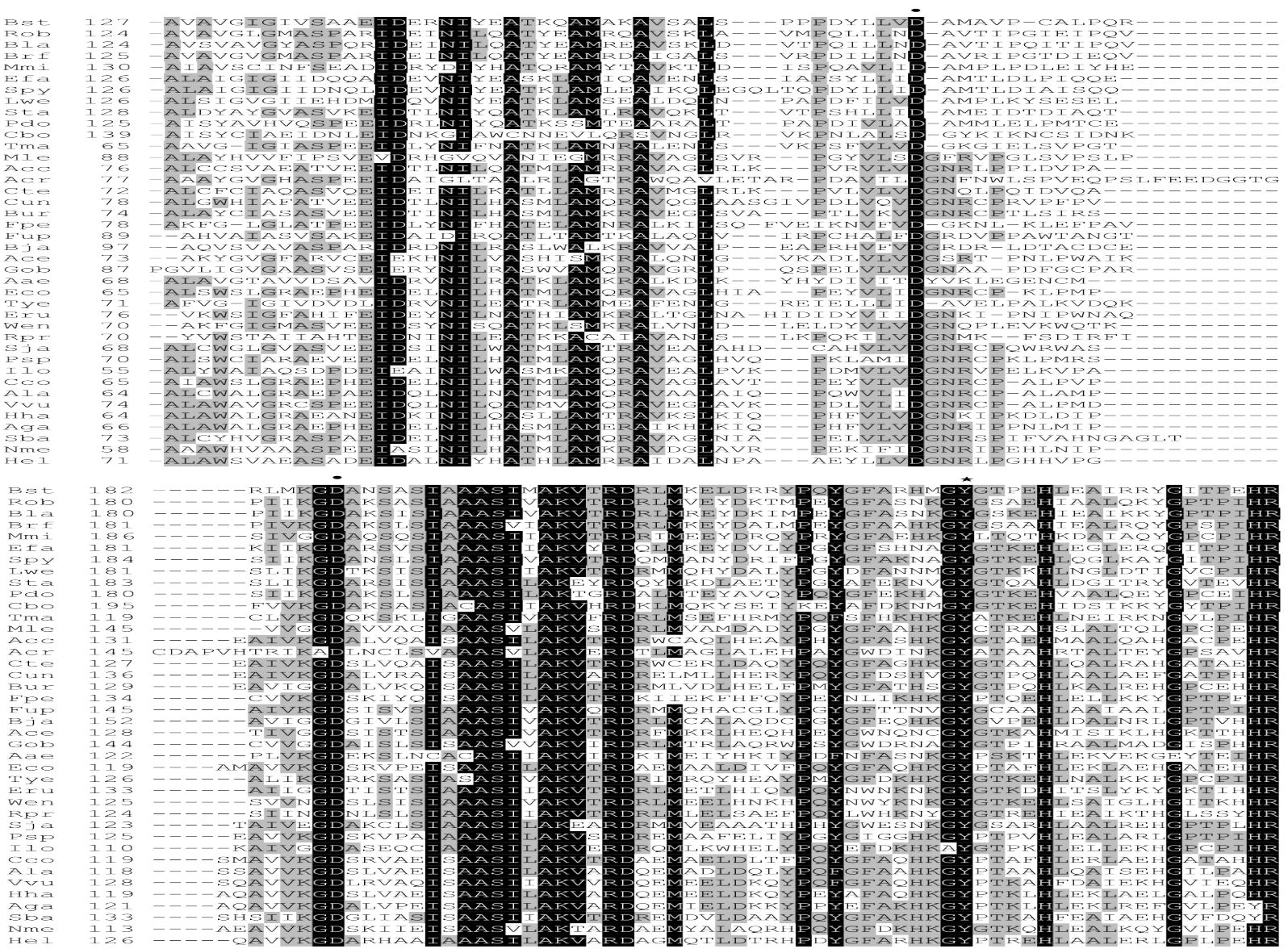
I used online software of the SWISS-MODEL program to build a 3D model for the structures of AaeRNH2, BstRNH2ΔN, and TmaRNH2ΔC [85], based on the crystal structure of TmaRNH2 (PDB code 3O3F) as a template structure. Of various RNases H for which the crystal structures are available, TmaRNH2 shows the highest amino acid sequence identities of 46.8% to AaeRNH2 and 51.6% to BstRNH2.

2.3. Results

2.3.1. Identification of N- and C-terminal extensions of RNases H2

The amino acid sequences of various bacterial RNases H2 are shown in Figure 2.1. Comparison of these sequences indicates that bacterial RNases H2 are classified into three groups that contain a relatively long N-terminal extension (N-extension), a relatively long C-terminal extension (C-extension) and neither extension. None of bacterial RNases H2 has both extensions. To examine whether these extensions are structurally and/or functionally important, I chose BstRNH2 and TmaRNH2 as representative members of RNases H2 with N- and C-extensions respectively, constructed the BstRNH2 derivative without N-extension (BstRNH2ΔN) and TmaRNH2 derivative without C-extension (TmaRNH2ΔC) and compared their biochemical properties with those of their intact partners, as described below. I also chose AaeRNH2 as a representative member of bacterial RNases H2 without both extensions and characterize it to examine whether bacterial RNases H2 without both extensions are functional, as described below. I chose BstRNH2, TmaRNH2 and AaeRNH2 as representative members in this study, because their source organisms are thermophiles. Proteins from thermophiles are usually thermostable, and thermostable proteins are more tolerant to deletion or mutation that destabilizes the protein than thermolabile proteins.


N-terminal extension



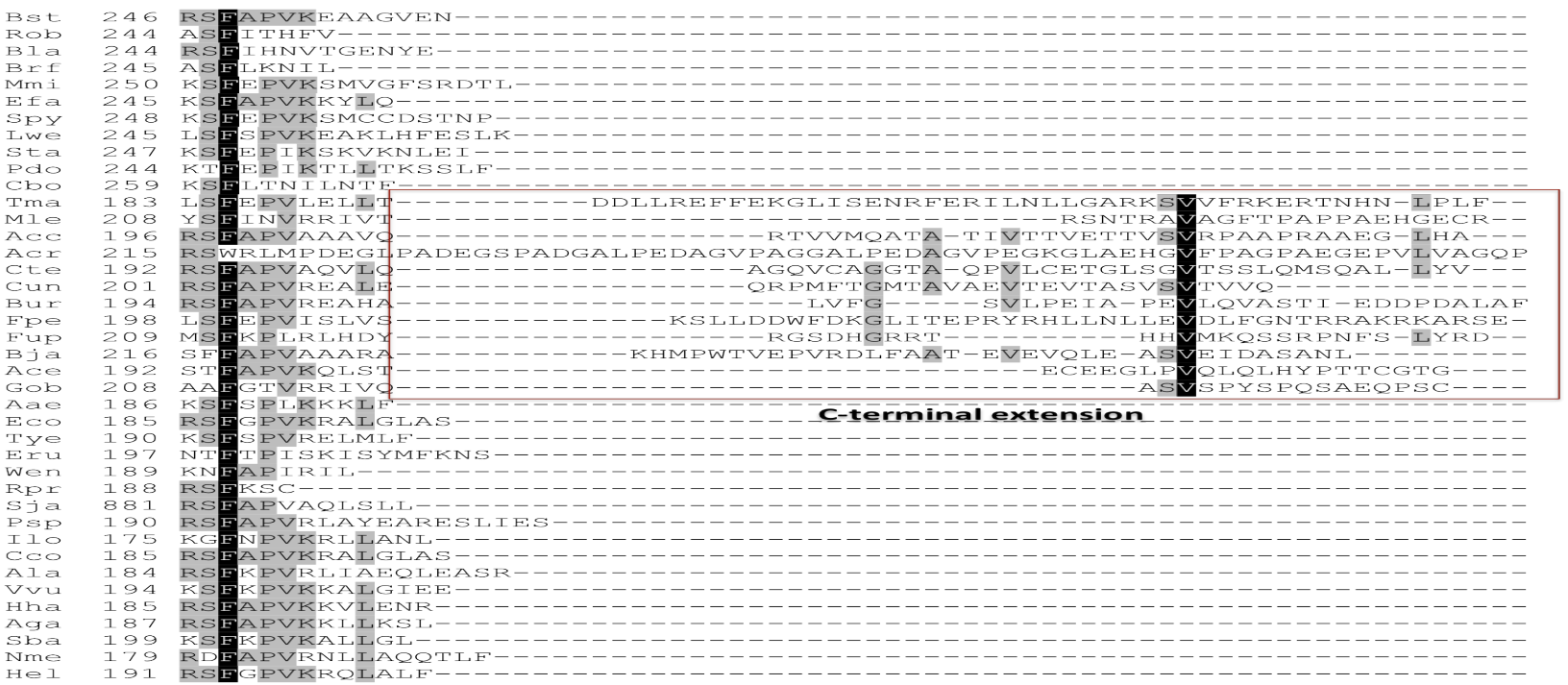


Fig. 2.1. Alignment of the amino acid sequences of bacterial RNases H2.

The amino acid sequences of bacterial RNases H2, which show the amino acid sequence identities of 40-50% to *Thermotoga maritima* RNase H2 (Tma) and are not selected from the same genus redundantly, are compared with one another. They are RNases H2 from *Bacillus stearothermophilus* RNase H2 (Bst), *Ruminococcus obeum* ATCC 29174 (Rob), *Blautia hansenii* DSM 20583 (Bla), *Bryantella formatexigens* DSM 14469 (Brf), *Megasphaera micronuciformis* F0359 (Mmi), *Enterococcus faecium* E980 (Efa), *Streptococcus pyogenes* MGAS8232 (Spy), *Listeria welshimeri serovar* 6b str. SLCC5334 (Lwe), *Staphylococcus arlettae* CVD059 (Sta), *Planococcus donghaensis* MPA1U2 (Pdo), *Clostridium botulinum* D str. 1873 (Cbo), *Mycobacterium leprae* (Mle), *Acidovorax citrulli* AAC00-1 (Acc), *Arthrobacter crystallopoietes* BAB-32 (Acr), *Comamonas testosteroni* ATCC 11996 (Cte), *Cupriavidus necator* N-1 (Cun), *Burkholderia* sp. SJ98 (Bur), *Fervidobacterium pennivorans* (Fpe), *Fulvimarina pelagi* HTCC2506 (Fup),

Bradyrhizobium japonicum USDA 110 (Bja), *Anaplasma centrale* str. Israel (Ace), *Gluconacetobacter oboediens* 174BP2 (Gob), *Aquifex aeolicus* RNase H2 (Aae), *Eschericia coli* (Eco), *Thermodesulfovibrio yellowstonii* (Tye), *Ehrlichia ruminantium* str. Welgevonden (Eru), *Wolbachia endosymbiont* WVITB of *Nasonia vitripennis* (Wen), *Rickettsia prowazekii* (Rpr), *Sphingobium japonicum* UT26S (Sja), *Pseudomonas* sp. GM78 (Psp), *Idiomarina loihiensis* L2TR (Ilo), *Cronobacter condimenti* 1330 (Cco), *Alishewanella aestuarii* B11 (Ala), *Vibrio vulnificus* CMCP6 (Vvu), *Haemophilus haemolyticus* M21639 (Hha), *Aggregatibacter actinomycetemcomitans* D7S-1 (Aga), *Shewanella baltica* OS183 (Sba), *Neisseria meningitidis* CU385 (Nme), and *Halomonas elongata* DSM 2581 (Hel). The accession numbers are BAB91155 for Bst, ZP_04863469 for Rob, ZP_05854594 for Bla, ZP_05345594 for Brf, ZP_07757987 for Mmi, ZP_06681343 for Efa, NP_607244 for Spy, YP_849487 for Lwe, ZP_11000244 for Sta, ZP_08094793 for Pdo, ZP_04863469 for Cbo, AAD35996 for Tma, B8ZRW3 for Mle, YP_970195 for Acc, ZP_24037048 for Acr, ZP_13098281 for Cte, YP_004685800 for Cun, ZP_11402099 for Bur, AFG35810 for Fpe, ZP_01438193 for Fup, NP_769161 for Bja, YP_003328826 for Ace, ZP_08899936 for Gob, AAC07736 for Aae, ACI71167 for Eco, ACI21777 for Tye, CAH57892 for Eru, ZP_08970534 for Wen, ADE29713 for Rpr, NC_014006 for Sja, NZ_AKJF01000065 for Psp, YP_156052 for Ilo, NZ_CAKW01000035 for Cco, ZP_10493190 for Ala, NP_760749 for Vvu, ZP_12492015 for Hha, YP_006286548 for Aga, NZ_CM001435 for Sba, NZ_AEQJ01000005 for Nme, and NC_014532 for Hel. For the RNase H domain, the amino acid residues that are conserved in at least 34 and 15 different proteins are highlighted in black and grey, respectively. For the N- and C-terminal extensions, the amino acid residues that are conserved in at least 9 and 5 different proteins are highlighted in black and grey, respectively. The active site residues are indicated by the filled circles above the sequences. The residues of TmaRNH2 that contact the 2'-OH group of the single ribonucleotide of the substrate (conserved tyrosine residue and the residues forming a GRG motif) [9] are indicated by the stars. The N-terminal and C-terminal extensions are indicated. Gaps are denoted by dashes. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein.

2.3.2. Protein preparation

BstRNH2 and TmaRNH2 have the N-terminal extension with 59 residues (Met1-Leu59) and C-terminal extension with 45 residues (Asp194-Phe238), respectively, as compared to AaeRNH2 (Fig. 2.2). According to the crystal structure of TmaRNH2 in complex with the substrate (duplex DNA with a single ribonucleotide) [9], which is the only structure available for bacterial RNases H2, the C-terminal extension of TmaRNH2

Fig. 2.2. Alignment of the amino acid sequences.

The amino acid sequences of BstRNH2 (Bst), TmaRNH2 (Tma), and AaeRNH2 (Aae) are compared with one another. The accession numbers are BAB91155 for BstRNH2, AAD35996 for TmaRNH2, and AAC07736 for AaeRNH2. The ranges of the secondary structures of TmaRNH2 are shown above the sequences. The amino acid residues that are conserved in at least 2 different proteins are highlighted in black. The active site residues are indicated by the filled circles above the sequences. The residues of TmaRNH2 that contact the 2'-OH group of the single ribonucleotide of the substrate (conserved tyrosine residue and the residues forming a GRG motif) are indicated by the stars. Gaps are denoted by dashes. The numbers represent the positions of the amino acid residues relative to the initiator methionine for each protein. The N-terminal 59 residues of BstRNH2 and C-terminal 46 residues of TmaRNH2, which are truncated to construct BstRNH2ΔN and TmaRNH2ΔC, respectively, are boxed.

assumes a relatively independent helix hairpin structure (helices I and J) (Fig. 2.3). To analyze the role of these extensions, BstRNH2ΔN (Met60-Asn259) and TmaRNH2ΔC (Met1-Leu192) without these extensions were constructed. TmaRNH2ΔC lacks the C-terminal 46, instead of 45, residues. The enzymatic properties of TmaRNH2ΔC may not be seriously changed regardless of whether the C-terminal 45 or 46 residues are truncated.

I used *E. coli* MIC2067(DE3) lacking the *rnhA* and *rnhB* genes to overproduce BstRNH2ΔN and TmaRNH2ΔC, as well as their intact partners. AaeRNH2, which consists

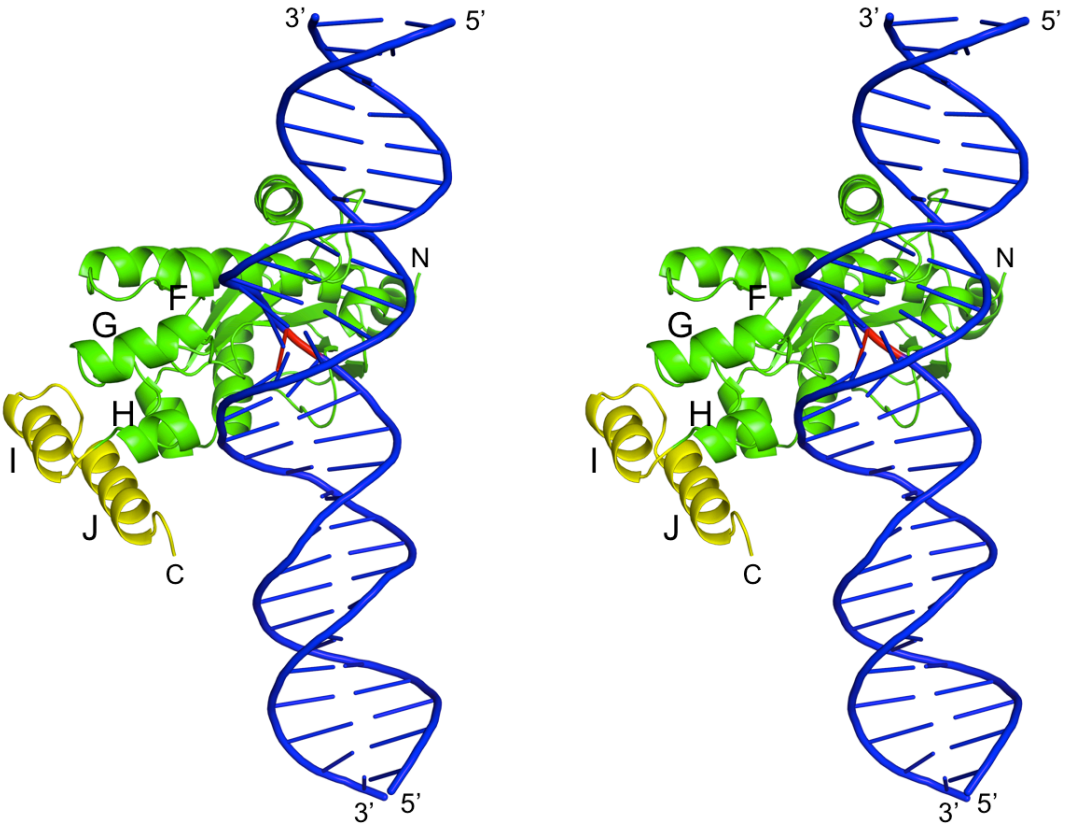


Fig. 2.3. C-terminal helix hairpin structure of TmaRNH2.

A stereoview of the 3D model of TmaRNH2 in complex with 29 bp dsDNA containing single ribonucleotide (D15-R1-D13/D29) is shown in blue and red, respectively. The C-terminal helix hairpin structure truncated to construct TmaRNH2ΔC (Thr193-Phe238) is shown in yellow. Helices F-J, which constitute the C-terminal domain of TmaRNH2, are indicated. 'N' and 'C' represent the N- and C-termini, respectively.

of 196 amino acid residues, was also overproduced in this strain to examine whether it exhibits similar enzymatic activities to those of BstRNH2ΔN and TmaRNH2ΔC. AaeRNH2 shows the amino acid sequence identities of 41.0% to BstRNH2ΔN and 46.8% to TmaRNH2ΔC. Upon overproduction, all proteins accumulated in *E. coli* cells in a soluble form and were purified to give a single band on SDS-PAGE (Fig. 2.4). The amount of the protein purified from 1 L culture was approximately 1 mg for BstRNH2 and BstRNH2ΔN, 10 mg for TmaRNH2, 7 mg for TmaRNH2ΔC, and 2 mg for AaeRNH2. The molecular masses of these proteins were estimated to be 31 kDa for BstRNH2, 24 kDa for BstRNH2ΔN, 27 kDa for TmaRNH2, 22 kDa for TmaRNH2ΔC, and 22 kDa for AaeRNH2 by the gel filtration chromatography. These values are comparable to those calculated from the amino acid sequences (28.8 kDa for BstRNH2, 21.6 kDa for BstRNH2ΔN, 26.6 kDa for TmaRNH2, 21.1 kDa for TmaRNH2ΔC, and 22 kDa for AaeRNH2), suggesting that these proteins exist as a monomer in solution.

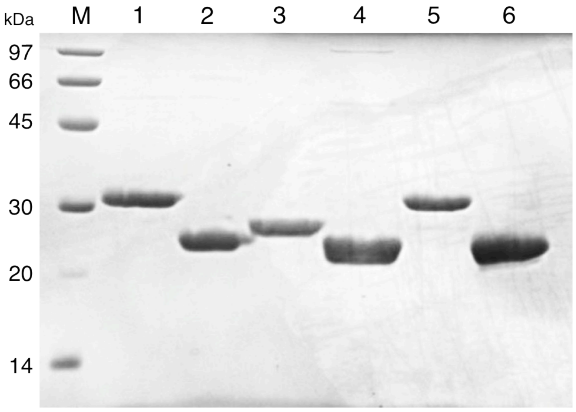


Fig. 2.4. SDS-PAGE of the purified proteins.

BstRNH2 (lane 1), BstRNH2ΔN (lane 2), TmaRNH2 (lane 3), TmaRNH2ΔC (lane 4), BstNTD-TmaRNH2ΔC (lane 5), and AaeRNH2 (lane 6) were subjected to electrophoresis on a 15% polyacrylamide gel in the presence of SDS. After electrophoresis, the gel was stained with Coomassie Brilliant Blue (CBB). Lane M, a low-molecular-weight marker kit (GE Healthcare).

2.3.3. CD spectra

I measured the far- and near-UV CD spectra of the proteins at 25°C and pH 8.0 (Fig. 2.5-A,B). The far-UV CD spectra reveal the secondary structures of the proteins,

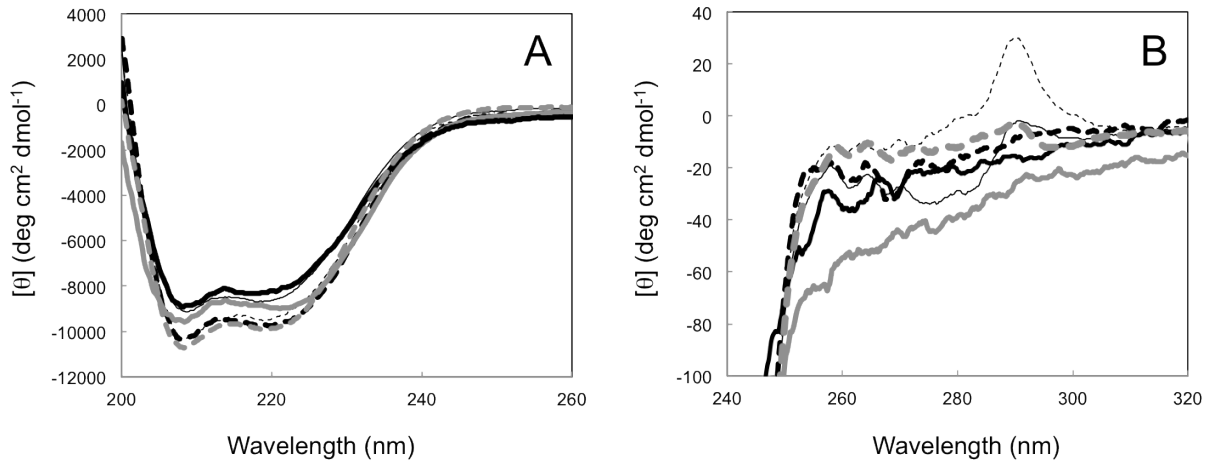


Fig. 2.5. CD spectra. The far-UV (A) and near-UV (B) CD spectra of BstRNH2 (thin dashed dark line), BstRNH2ΔN (thin solid dark line), TmaRNH2 (thick dashed dark line), TmaRNH2ΔC (thick solid dark line), AaeRNH2 (solid gray line), and BstNTD-TmaRNH2ΔC (dashed grey line) are shown. These spectra were measured at 25°C as described in the Section 2.2.4.

whereas the near-UV CD spectra reveal the local conformation around the aromatic residues, such as Trp and Tyr. The far-UV CD spectra of BstRNH2ΔN and TmaRNH2ΔC are similar to those of their intact partners, although these spectra have a shallower trough at 210-220 nm than the spectra of their intact partners. The near-UV CD spectrum of TmaRNH2ΔC is also similar to that of TmaRNH2. In contrast, the near-UV CD spectrum of BstRNH2ΔN is different from that of BstRNH2. These results suggest that the removal of the N- and C-terminal extensions does not seriously affect the overall structures of BstRNH2 and TmaRNH2 respectively. The near-UV CD spectrum of BstRNH2 is changed by the removal of the N-terminal extension, probably because the environment of the

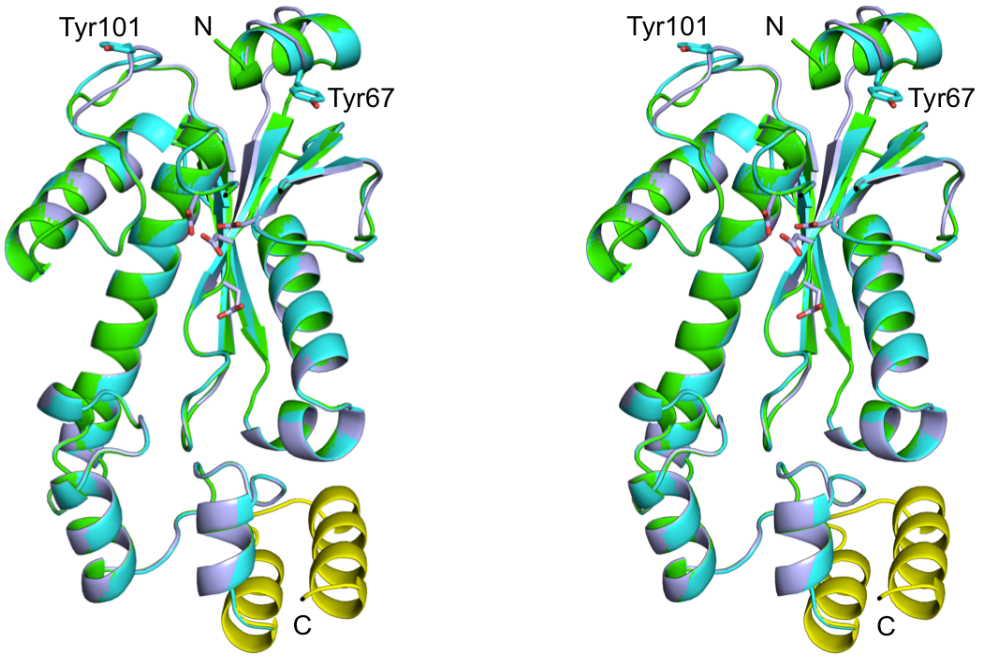


Fig. 2.6. Stereoviews of the 3D models of AaeRNH2 and BstRNH2ΔN. The 3D models of AaeRNH2 (purple) and BstRNH2ΔN (cyan) are superimposed on the crystal structure of TmaRNH2 (green) (PDB code 3O3F). The four acidic active site residues of AaeRNH2 (Asp19, Glu20, Asp11, and Asp127) and two tyrosine residues (Tyr67 and Tyr101) of BstRNH2ΔN, which are presumably located near the interface between the deleted region and RNase H domains of BstRNH2, are shown by stick models. The four acidic active site residues of other proteins are not shown, because their positions are nearly identical with those of AaeRNH2. In the structure of TmaRNH2, the region truncated to construct TmaRNH2ΔC is shown in yellow. "N" and "C" represent the N- and C-termini of TmaRNH2, respectively.

aromatic residues located near the interface between the deleted region and RNase H domain is altered. According to a 3D model of BstRNH2ΔN, two tyrosine residues (Tyr67 and Tyr101) are located near the position where the N-terminal residue is located (Fig. 2.6). These residues, as well as two tryptophan residues (Trp24 and Trp56) in the deleted region, are the candidates of the aromatic residues located near the interface. According to the crystal structure of TmaRNH2, none of the tryptophan and tyrosine residues is located at and around the deleted region.

The far-UV CD spectra of BstRNH2ΔN, TmaRNH2ΔC, and AaeRNH2 are similar

with one another, suggesting that they share a main-chain fold. In contrast, the near-UV CD spectra of these proteins are different from one another, suggesting that the local conformations around the aromatic residues of these proteins are different from one another. BstRNH2ΔN and TmaRNH2ΔC do not contain a tryptophan residue but contain nine and four tyrosine residues, respectively. AaeRNH2 contains one tryptophan and nine tyrosine residues. Only two tyrosine residues are conserved in these proteins. These differences may account for the difference in their near-UV CD spectra.

2.3.4. Enzymatic activities

The JRNase and RNase H activities of all proteins were determined at 30°C using 29-bp DNA_{15}-RNA_1-DNA_{13}/DNA duplex (D15-R1-D13/D29) and 12-bp RNA/DNA hybrid (R12/D12) as a substrate, respectively. Mn^{2+} and Mg^{2+} ions were used as a metal cofactor. All proteins examined cleaved the D15-R1-D13/D29 substrate almost exclusively at the (5')DNA-RNA(3') junction regardless of the metal cofactor (Fig. 2.7-A). In contrast, BstRNH2, TmaRNH2, and AaeRNH2 cleaved the R12/D12 substrate at multiple sites with different sequence preferences (Fig. 2.7-B). The preferable cleavage sites of this substrate with TmaRNH2 were not significantly changed when the metal cofactor was changed and were comparable to those with TmaRNH2ΔC. In contrast, the preferable cleavage sites of this substrate with BstRNH2 were slightly changed when the metal cofactor was changed and were different from those with BstRNH2ΔN. The preferable cleavage sites of this substrate with AaeRNH2 were also slightly changed when the metal cofactor was changed. These results indicate that the sequence preference of TmaRNH2 is not significantly changed by the removal of the C-terminal extension, whereas that of BstRNH2 is changed by the removal of the N-terminal extension.

The specific JRNase and RNase H activities of the proteins were determined by quantifying the amount of the substrate and products separated on urea gel. For simplicity,

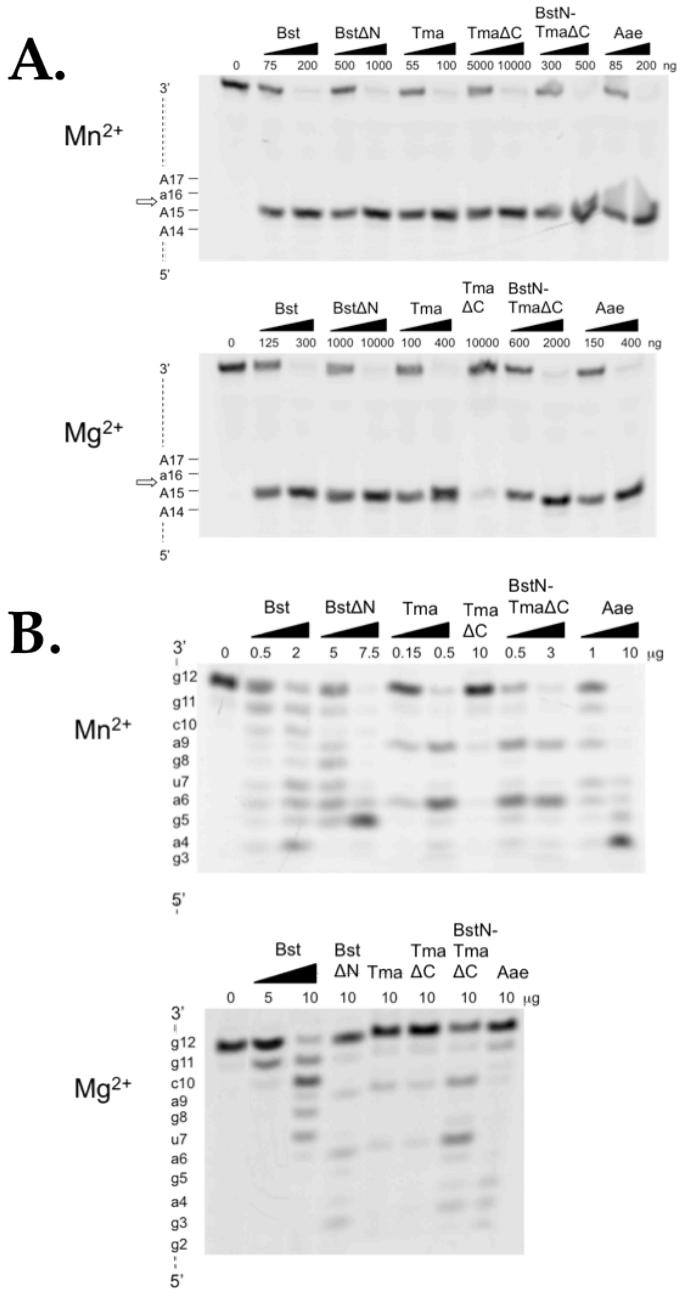


Fig. 2.7. Cleavage of oligomeric substrates by various RNases H. The 5′-end-labeled DNA_{15}-RNA_1-DNA_{13}/DNA_{29} (A) and RNA_{12}/DNA_{12} (B) were hydrolyzed by the enzyme in the presence of either Mn^{2+} or Mg^{2+} at 30 °C for 15 min, and the hydrolysates were separated on a 20% polyacrylamide gel containing 7 M urea, as described in Section 2.2.3. The reaction volume was 10 μL, and concentration of the substrate was 1.0 μm. The sequences of DNA_{15}-RNA_1-DNA_{13} and RNA_{12} around the cleavage sites are indicated along the gel. Deoxyribonucleotides are indicated by uppercase letters, and ribonucleotides are indicated by lowercase letters. Enzymes: Bst, BstRNH2; BstΔN, BstRNH2ΔN; Tma, TmaRNH2; TmaΔC, TmaRNH2ΔC; BstN–TmaΔC, BstNTD–TmaRNH2ΔC; Aae, AaeRNH2. The major site of cleavage of DNA_{15}-RNA_1-DNA_{13} by these enzymes is shown by an arrow. The cleavage sites of RNA_{12} are not shown, because this oligonucleotide is cleaved by these enzymes at all possible sites between g2 and g12.

the JRNase and RNase H activities determined in the presence of Mn^{2+} and Mg^{2+} ions are termed JRNase^{Mn}, JRNase^{Mg}, RNase H^{Mn}, and RNase H^{Mg} activities. The optimum concentrations of the metal ions for these activities were 10 mM for BstRNH2ΔN, BstRNH2, and AaeRNH2, and 1 mM for TmaRNH2ΔC and TmaRNH2, regardless of the metal cofactor (data not shown). The maximal specific JRNase^{Mn}, JRNase^{Mg}, RNase H^{Mn}, and RNase H^{Mg} activities of these proteins are summarized in Table 1.

When the four activities mentioned above are compared with one another for BstRNH2, TmaRNH2, and AaeRNH2, the JRNase^{Mn} activity is always the highest and other activities decrease in the order JRNase^{Mn} > JRNase^{Mg} > RNase H^{Mn} > RNase H^{Mg}. However, the JRNase^{Mg} activity is lower than the JRNase^{Mn} activity by only 40-50% for these proteins. In contrast, the RNase H^{Mg} activity is lower than the RNase H^{Mn} activity by 10 fold for BstRNH2, 100 fold for TmaRNH2, and 50 fold for AaeRNH2. These results indicate that BstRNH2, TmaRNH2, and AaeRNH2 show a weak preference to Mn^{2+} ion for JRNase activity and a strong preference to Mn^{2+} ion for RNase H activity. In addition, The RNase H^{Mn} activity is lower than the JRNase^{Mg} activity by 4 fold for BstRNH2, 1.5 fold for TmaRNH2, and 7 fold for AaeRNH2, indicating that these proteins prefer the D15-R1-D13/D29 substrate to the R12/D12 substrate. Similar results have been reported for TmaRNH2, except that it shows a weak preference to Mg^{2+}, instead of Mn^{2+}, ion for JRNase activity [9]. This discrepancy is probably caused by the difference in the substrate used for assay.

When the four activities of BstRNH2ΔN are compared with those of BstRNH2, they are reduced by 7 fold for JRNase^{Mn} activity, 40 fold for JRNase^{Mg} activity, 10 fold for RNase H^{Mn} activity, and 9 fold for RNase H^{Mg} activity. These results indicate that the N-terminal extension of BstRNH2 is important for activity, especially for JRNase^{Mg} activity.

When the four activities of TmaRNH2ΔC are compared with those of TmaRNH2, they are reduced by 90 fold for JRNase^{Mn} activity, 440 fold for JRNase^{Mg} activity, 500 fold for RNase H^{Mn} activity, and 7 fold for RNase H^{Mg} activity. These results indicate that the C-terminal extension of TmaRNH2 is important for activity, especially for JRNase^{Mn}, JRNase^{Mg}, and RNase H^{Mn} activities.

Table 1. Specific activities of the proteins.

Protein	JRNase^{Mn} activity (unit/mg)	Relative activity (%)	JRNase^{Mg} activity (unit/mg)	Relative activity (%)	RNase H^{Mn} activity (unit/mg)	Relative activity (%)	RNase H^{Mg} activity (unit/mg)	Relative activity (%)
TmaRNH2	6.2 ± 0.44	100	3.4 ± 0.047	55	2.2 ± 0.094	35	0.022 ± 0.0002	0.35
TmaRNH2ΔC	0.070 ± 0.0030	1.1	0.0077 ± 0.0013	0.12 (0.23)	0.0044 ± 0.0012	0.070 (0.20)	0.0033 ± 0.0009	0.053 (15)
BstNTD-TmaRNH2ΔC	1.2 ± 0.047	19	0.56 ± 0.008	9.0 (16)	0.65 ± 0.028	10 (29)	0.025 ± 0.0028	0.40 (110)
BstRNH2	4.7 ± 0.15	76	2.7 ± 0.037	44	0.67 ± 0.019	11	0.067 ± 0.003	1.1
BstRNH2ΔN	0.69 ± 0.019	11 (15)	0.068 ± 0.001	1.1 (2.5)	0.070 ± 0.004	1.1 (10)	0.0071 ± 0.0012	0.11 (11)
AaeRNH2	4.3 ± 0.29	69	2.4 ± 0.16	39	0.33 ± 0.005	5.3	0.0092 ± 0.0011	0.15

Hydrolyses of the D15-R1-D13/D29 (for JRNase activity) and R12/D12 (for RNase H activity) substrates by the protein were performed at 30°C under the conditions described in the Section 2.2.3. The $MnCl_2$ and $MgCl_2$ concentrations were 1 mM for TmaRNH2 and TmaRNH2ΔC, and 10 mM for other proteins, regardless of the metal ions. Experiments were carried out at least twice and the average values are shown together with the errors. The specific activities of the proteins relative to the JRNase^{Mn} activity of TmaRNH2 are shown. The specific JRNase^{Mn}, JRNase^{Mg}, RNase H^{Mn}, and RNase H^{Mg} activities of the mutant proteins (TmaRNH2ΔC, BstNTD-TmaRNH2ΔC, and BstRNH2ΔN) relative to those of the parent proteins are shown in the parentheses.

2.3.5. Complementation of the temperature-sensitive (ts) growth phenotype of an RNase H-deficient *E. coli* strain

E. coli MIC2067(DE3) shows an RNase H-dependent temperature-sensitive (ts) growth phenotype [53]. To examine whether BstRNH2, BstRNH2ΔN, TmaRNH2, TmaRNH2ΔC, and AaeRNH2 complement this phenotypic defect, *E. coli* MIC2067(DE3) transformants for overproduction of these proteins were grown on the LB-agar plates containing 30 $mg \cdot L^{-1}$ chloramphenicol and 50 $mg \cdot L^{-1}$ ampicillin in the absence of IPTG at permissive (30°C) and non-permissive (42°C) temperatures. *E. coli* transformants for overproduction of BstRNH2 and TmaRNH2 formed colonies on these plates both at 30 and 42°C, whereas those for overproduction of other proteins only formed colonies at 30°C (data not shown). These results indicate that BstRNH2 and TmaRNH2 complement the ts phenotype of MIC2067(DE3), whereas other three proteins do not. Because the JRNase activity of AaeRNH2 is comparable to that of BstRNH2 regardless of the metal cofactor (Table 1), the RNase H activities of BstRNH2ΔN, TmaRNH2ΔC, and AaeRNH2 are probably too low to complement the ts phenotype of MIC2067(DE3). It seems unlikely that the production levels of these proteins are too low to complement this phenotype, because the production levels of these proteins upon induction for overproduction are comparable to or rather higher than that of BstRNH2 (data not shown).

2.3.6. Binding to the substrate

To examine whether the removal of the N-terminal extension of BstRNH2 and C-terminal extension of TmaRNH2 affects the substrate binding, I analyzed the binding affinities of Bst-RNH2, Bst-RNH2ΔN, Tma-RNH2, and Tma-RNH2ΔC to the D15-R1-

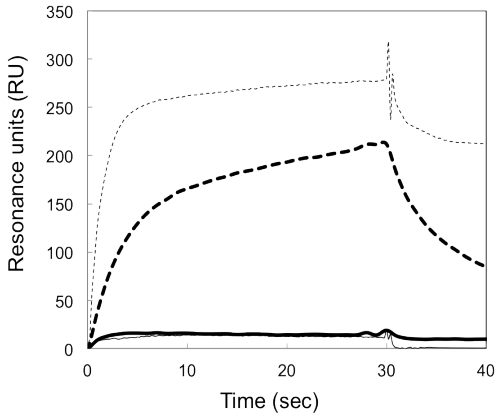


Fig. 2.8. Binding of the proteins to immobilized dsDNA with a single ribonucleotide. The sensorgrams showing the binding of 1 μM of BstRNH2 (thin dashed line), BstRNH2ΔN (thin solid line), TmaRNH2 (thick dashed line), and TmaRNH2ΔC (thick solid line) to the immobilized D15-R1-D13/D29 substrate are shown. Injections were done at time zero for 30 s.

R13/D29 substrate in the absence of the metal cofactor using surface plasmon resonance (Biacore system). These proteins were injected onto the sensor chip, on which the D15-R1-D13/D29 substrate was immobilized. When 1 μM of Bst-RNH2 or Tma-RNH2 was injected onto the sensor chip, a significant increase of RU was detected (Fig. 2.8). In contrast, only a slight increase of RU was detected, when 1 μM of Bst-RNH2ΔN or Tma-RNH2ΔC was injected. The association constants, K_A, of these proteins for binding to the substrate, which were determined by measuring the equilibrium-binding responses at various concentrations of the proteins, are summarized in Table 2. The K_A values of BstRNH2ΔN and TmaRNH2ΔC were lower than those of BstRNH2 and TmaRNH2 by approximately 30 and >100 fold, respectively. These results indicate that the removal of the N-terminal extension of BstRNH2 and C-terminal extension of TmaRNH2 greatly reduces the binding affinities of these proteins to the substrate.

Table 2. Association constants of the proteins for substrate binding

Protein	K_A (M^{-1}) x 10^{-5}	Relative K_A (%)
BstRNH2	190	100
BstRNH2ΔN	6.1	3.2
TmaRNH2	3.4	1.8
TmaRNH2ΔC	< 0.03	< 0.02 (< 0.9)
BstNTD-TmaRNH2ΔC	10	5.3 (290)
AaeRNH2	3.9	2.1

Binding of the proteins to the D15-R1-D13/D29 substrate immobilized onto the sensor chip was analyzed in the absence of the metal cofactor by surface plasmon resonance as described in the Section 2.2.5. The K_A values of the proteins relative to that of BstRNH2 are shown. The K_A values of TmaRNH2ΔC and BstNTD-TmaRNH2ΔC relative to that of TmaRNH2 are shown in the parenthesis.

2.3.7. Thermal stability

To examine whether the removal of the N-terminal extension of BstRNH2 and C-terminal extension of TmaRNH2 affects the protein stability, thermal stability of BstRNH2, BstRNH2ΔN, TmaRNH2, and TmaRNH2ΔC was analyzed by monitoring changes of the CD values at 222 nm. Thermal stability of AaeRNH2 was also analyzed for comparative purpose. In the presence of 1.5 M guanidine hydrochloride (GdnHCl) at pH 8.0, all proteins unfolded in a single cooperative fashion in a reversible manner. The thermal denaturation curves of these proteins are compared with one another in Figure 2.9. The parameters characterizing the thermal denaturation of these proteins are summarized in Table 3. Comparison of these parameters indicates that BstRNH2ΔN and TmaRNH2ΔC are less stable than their intact partners by 11.2 and 19.5°C in T_m, respectively. The ΔH_m and ΔS_m values of these proteins are also decreased as compared to those of their intact partners. These results indicate that the N-terminal extension of BstRNH2 and C-terminal extension of TmaRNH2 greatly contribute to the protein stabilization. Comparison of the

T_m values of AaeRNH2, TmaRNH2, and BstRNH2 indicates that the stability of these proteins decreases in this order.

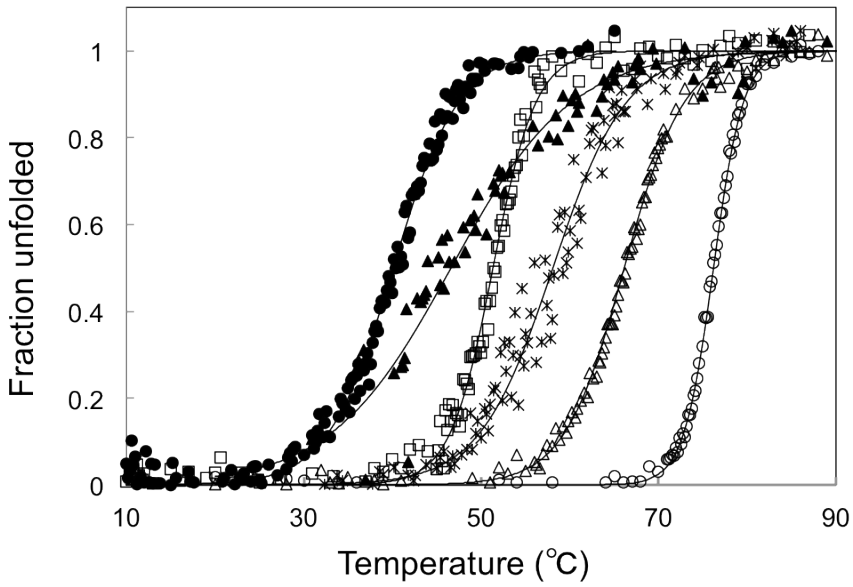


Fig. 2.9. Thermal denaturation curves. Thermal denaturation curves of AaeRNH2 (open circle), TmaRNH2 (open triangle), TmaRNH2ΔC (closed triangle), BstRNH2 (open square), BstRNH2ΔN (closed circle), and BstNTD-TmaRNH2ΔC (star) are shown. These curves were obtained at pH 8.0 in the presence of 1.5 M GdnHCl, as described in the Section 2.2.6. The theoretical curves are drawn on the assumption that the proteins are denatured via a two-state mechanism

AaeRNH2 is more stable than TmaRNH2 and BstRNH2 by 10.1 and 24.9°C in T_m, respectively. The upper limits of the growth temperatures of *A. aeolicus*, *T. maritima*, and *B. stearothermophilus* are 95°C [86], 90°C [87], and 70°C [88], respectively. Thus, the stability of these proteins decreases in proportion to the growth temperatures of their source organisms. It is noted that the ΔH_m values of AaeRNH2, BstRNH2, and TmaRNH2, as well as their ΔS_m values, decrease in same order. The reason why these values do not decrease in proportion to their T_m values remains to be clarified.

Table 3. Parameters characterizing thermal denaturation of the proteins

Protein	T_m (°C)	ΔT_m1 (°C)	ΔT_m2 (°C)	ΔH_m (kJ·mol^{-1})	ΔS_m (kJ·mol^{-1}·K^{-1})
AaeRNH2	76.2 ± 0.03	-	-	601.9 ± 8.6	1.7 ± 0.02
TmaRNH2	66.1 ± 0.05	-10.1	-	282.9 ± 5.2	0.8 ± 0.02
TmaRNH2ΔC	46.6 ± 0.81	-	-19.5	125.6 ± 23.9	0.4 ± 0.07
BstNTD-TmaRNH2ΔC	58.1 ± 0.66	-	-8.0	220.1 ± 41.2	0.7 ± 0.13
BstRNH2	51.3 ± 0.06	-24.9	-	361.1 ± 11.3	1.1 ± 0.03
BstRNH2ΔN	40.1 ± 0.1	-	-11.2	221.9 ± 8.7	0.7 ± 0.03

The parameters were determined from the thermal denaturation curves shown in Figure 2.9. The melting temperature (T_m) is temperature of the midpoint of the thermal denaturation transition. ΔT_m1 is the difference in T_m between AaeRNH2 and TmaRNH2 or BstRNH2 and is calculated as T_m (TmaRNH2 or BstRNH2) – T_m (AaeRNH2). ΔT_m2 is the difference in T_m between TmaRNH2ΔC or BstNTD-TmaRNH2ΔC and TmaRNH2 and between BstRNH2ΔN and BstRNH2, and is calculated as T_m (TmaRNH2ΔC or BstNTD-TmaRNH2ΔC) – T_m (TmaRNH2) or T_m (BstRNH2ΔN) – T_m (BstRNH2). ΔH_m and ΔS_m are the enthalpy and entropy changes of unfolding at T_m calculated by van't Hoff analysis.

2.3.8. Biochemical properties of TmaRNH2ΔC with BstNTD attached at the N-terminus (BstNTD-TmaRNH2ΔC)

To examine whether the N-terminal extension of BstRNH2 (BstNTD) restores the activity and stability of TmaRNH2ΔC, BstNTD-TmaRNH2ΔC, in which BstNTD is attached to the N-terminus of TmaRNH2ΔC, was constructed. BstNTD-TmaRNH2ΔC was overproduced in *E. coli* and purified to give a single band on SDS-PAGE (Fig. 2.4). The production level and purification yield of BstNTD-TmaRNH2ΔC were comparable to those of TmaRNH2ΔC. The molecular mass of this protein estimated by the gel filtration chromatography (31 kDa) was comparable to that calculated from the amino acid sequence (28,3 kDa), suggesting that it exists as a monomer in solution. The far- and near-UV CD

spectra of BstNTD-TmaRNH2ΔC are similar to those of TmaRNH2ΔC (Fig. 2.5), suggesting that the attachment of BstNTD does not seriously affect the structure of TmaRNH2ΔC.

When *E. coli* MIC2067(DE3) transformants for overproduction of BstNTD-TmaRNH2ΔC were grown on the LB-agar plates in the absence of IPTG, they form colonies both at 30 and 42°C (data not shown). This result indicates that BstNTD-TmaRNH2ΔC has an ability to complement the ts phenotype of *E. coli* MIC2067(DE3). Thus, BstNTD restores the RNase H activity of TmaRNH2ΔC. When the enzymatic activity of BstNTD-TmaRNH2ΔC was analyzed using the oligomeric substrates, BstNTD-TmaRNH2ΔC cleaved the D15-R1-D13/D29 and R12/D12 substrates with the same sequence preferences as those of TmaRNH2, regardless of the metal cofactor (Fig. 2.7). The JRNase^{Mn}, JRNase^{Mg}, and RNase H^{Mn} activities of BstNTD-TmaRNH2ΔC were higher than those of TmaRNH2ΔC by 17, 70, and 150 fold, respectively, and lower than those of TmaRNH2 by 5, 6, and 3 fold, respectively (Table 1). The RNase H^{Mg} activity of BstNTD-TmaRNH2ΔC was higher than that of TmaRNH2ΔC by 8 fold and comparable to that of TmaRNH2 (Table 1). These results indicate that BstNTD at least partly restores the enzymatic activity of TmaRNH2ΔC. When the binding affinity of BstNTD-TmaRNH2ΔC to the D15-R1-D13/D29 substrate was analyzed in the absence of the metal cofactor using surface plasmon resonance, it was higher than those of TmaRNH2ΔC and TmaRNH2 by >300 and 3 fold in K_A, respectively (Table 2). This result indicates that BstNTD functions as a potent substrate binding domain and restores the substrate binding affinity of TmaRNH2ΔC beyond the level of TmaRNH2. When thermal stability of BstNTD-TmaRNH2ΔC was analyzed in the presence of 1.5 M GdnHCl by monitoring changes of the CD values at 222 nm, it reversibly unfolded in a single cooperative fashion. The

thermal denaturation curve of this protein is shown in Figure 2.9. BstNTD-TmaRNH2ΔC was more stable than TmaRNH2ΔC by 11.5°C and less stable than TmaRNH2 by 8°C in T_m (Table 3). This result indicates that BstNTD partly restores the stability of TmaRNH2ΔC.

2.4. Discussion

2.4.1. Role of BstNTD

The finding that the removal of the N-terminal extension of BstRNH2 (BstNTD) does not significantly affect the structure, but reduces the activity, substrate binding affinity, and stability of the protein indicates that BstNTD is important for substrate binding and protein stability. BstNTD is probably folded into a relatively independent structure and functions as a substrate binding domain. BstNTD is probably kept folded and functions in an isolated form as well, because the attachment of BstNTD to the N-terminus of TmaRNH2ΔC greatly increases the activity and substrate binding affinity of TmaRNH2ΔC. However, BstNTD contributes to the stabilization of BstRNH2, suggesting that BstNTD interacts with the RNase H domain. BstNTD also contributes to the stabilization of BstNTD-TmaRNH2ΔC, but less significantly than to the stabilization of BstRNH2, probably because the interactions at the interface between BstNTD and RNase H domain that contribute to the stabilization of BstRNH2 are only partially conserved in BstNTD-TmaRNH2ΔC.

It has been reported that BstRNH2ΔN does not bind to RNA/DNA hybrid and therefore does not exhibit RNase H activity in the presence of Mn^{2+} ion [89]. However, in this study, we showed that BstRNH2ΔN exhibits RNase H activity either in the presence of Mn^{2+} or Mg^{2+} ion, although it is lower than that of BstRNH2 by 9-10 fold (Table 1). The reason for this discrepancy is not understood. However, BstRNH2ΔN probably retains weak substrate binding affinity for RNA/DNA hybrid as well, because BstRNH2ΔN retains both weak JRNase activity (Table 1) and weak binding affinity for the D15-R1-D13/D29 substrate (Table 2).

A 3D model of BstRNH2ΔN is highly similar to the crystal structure of TmaRNH2 (Fig. 2.6), suggesting that BstRNH2 contacts the substrate as does TmaRNH2 (Fig. 2.3). In Figure 2.3, a model of the complex between TmaRNH2 and the 29-bp D15-R1-D13/D29 substrate is shown. This model is constructed based on the crystal structure of TmaRNH2 in complex with the short substrate (12-bp D5-R1-D6/D12) [9]. In this model, the substrate is slightly bended at the position where it contacts the enzyme, due to the deformation of the nucleic acid backbone at the RNA-DNA junction [9]. BstNTD probably contacts the noncleaved DNA strand 4-6 bases away (downstream) from the scissile phosphate group. Further structural studies will be required to understand the mechanism by which BstNTD binds to the substrate.

Database searches indicate that not only RNases H2 from the genus *Bacillus*, but also those from other genera, such as *Enterococcus*, *Streptococcus*, *Staphylococcus*, *Ruminococcus*, *Blautia*, *Megasphaera*, *Listeria*, and *Clostridium*, have an N-terminal extension (Fig. 2.1). These extensions show a similarity to BstNTD both in size (50-60 residues) and amino acid sequence (at least 20%), suggesting that they assume a similar structure as that of BstNTD and function as a substrate binding domain. It has been reported that almost every eukaryotic RNases H1 and several bacterial RNases H1 have a hybrid binding domain (HBD) at their N-termini, which is important for substrate binding [2, 76]. Likewise, bacterial RNases H3 have a TBP-like domain at their N-termini, which is important for substrate binding [14, 53]. However, none of the N-terminal extensions of bacterial RNases H2 show a significant amino acid sequence similarity to HBD or TBP-like domain, suggesting that their structures are different from those of HBD and TBP-like domain.

2.4.2. Role of the C-terminal extension of TmaRNH2

According to the crystal structure of the TmaRNH2-substrate complex, TmaRNH2 consists of the catalytic and C-terminal domains, and the substrate binds to a groove on the protein surface between these domains [9]. This interaction is preserved in the model shown in Figure 2.3. The C-terminal domain of TmaRNH2 consists of five helices (helices F-J) and mainly contacts the noncleaved strand through van der Waals interactions or interactions mediated by the backbone of the protein. Helices G and H provide a surface responsible for binding of the noncleaved strand. In contrast, a helix hairpin structure containing helices I and J, which corresponds to the C-terminal extension, does not contact the substrate. This result suggests that the C-terminal helix hairpin structure is not involved in substrate binding. The previous result that TmaRNH2 efficiently hydrolyzes the 12-bp D5-R1-D6/D12 substrate [9], which is too short to contact the helix hairpin structure, supports this hypothesis. Nevertheless, the removal of this helix hairpin structure greatly reduces the binding affinity to the D15-R1-D13/D29 substrate (Table 2) and both the JRNase and RNase H activities (Table 1). This result suggests that the C-terminal helix hairpin structure is required to make the conformation of the C-terminal domain functional. TmaRNH2ΔC may assume a non-native structure, in which the shape of the surface responsible for binding of the noncleaved strand is slightly changed.

The substrate binding affinity and activity of TmaRNH2ΔC are partly restored by the attachment of BstNTD to the N-terminus of TmaRNH2ΔC. The sequence preference of the resultant fusion protein is nearly identical to that of TmaRNH2, indicating that the catalytic domain, instead of BstNTD, determines the sequence preference. These results suggest that the removal of the C-terminal helix hairpin structure causes only a subtle change in the conformation of the C-terminal domain and this conformational change

greatly decreases the interactions between the protein and substrate without significantly altering them. BstNTD probably stabilizes the protein-substrate complex, which is greatly destabilized by the removal of the C-terminal helix hairpin structure. The removal of the C-terminal helix hairpin structure decreases the conformational stability of the protein as well, suggesting that this structure contributes to the stabilization of the protein as well.

Database searches indicate that not only RNases H2 from the genus *Thermotoga*, but also those from other genera, such as *Arthrobacter*, *Comamonas*, *Burkholderia*, *Acidovorax*, *Fervidobacterium*, and *Bradyrhizobium* have a C-terminal extension (Fig. 2.1). However, unlike the N-terminal extensions of bacterial RNases H2, these extensions vary both in size and amino acid sequence, suggesting that they assume variable structures. Because the catalytic domains of these RNases H2 share a structure with that of TmaRNH2, the C-terminal extensions of these RNases H2 are probably required to make the conformation of the C-terminal domain functional.

2.4.3. Activity and stability of AaeRNH2

AaeRNH2 exhibits comparable JRNase activity to those of TmaRNH2 and BstRNH2, despite the lack of an N- or C-terminal extension (Table 1). In addition, AaeRNH2 is more stable than TmaRNH2 and BstRNH2. These results indicate that bacterial RNase H2 does not always require an N- or C-terminal extension to increase activity, substrate binding affinity, and/or stability. Various RNases H2 from mesophilic and psychrophilic bacteria, such as *E. coli*, *Haemophilus haemolyticus*, and *Shewanella baltica*, contain neither an N-terminal nor a C-terminal extension as well (Fig. 2.1), suggesting that efficient functioning without these extensions is not unique to highly thermostable RNases H2.

AaeRNH2 is more stable than TmaRNH2 and BstRNH2 by 10.1 and 24.9°C, respectively in T_m. It has been reported that hyperthermophilic proteins are stabilized by the combination of various factors, such as increased number of ion pairs (ion pair networks) [90, 91], increased number of metal binding sites [92], reduced cavity volume [47], anchoring of C-terminal tail [45], increased interior hydrophobicity [93], increased packing density including aromatic interactions [90], increased molecular compactness [94], oligomerization [95], increased number of proline residues at loop regions [96], and increased number of disulfide bonds [97]. To understand the stabilization mechanism of AaeRNH2, a 3D model of this protein was constructed based on the crystal structure of TmaRNH2. The backbone structure of this model is nearly identical to those of a 3D model of BstRNH2ΔN and crystal structure of TmaRNH2 (Fig. 2.6), suggesting that there is little difference in molecular compactness among them. Comparison of these structures indicates that the cavity volume, interior hydrophobicity, and the number of proline residues at the loop regions do not correlate with the protein stability. For example, the cavity volume, which is calculated using the program CASTp [98], is 343 $Å^3$ for AaeRNH2, 80 $Å^3$ for TmaRNH2, and 289 $Å^3$ for BstRNH2ΔN. The contents of the buried nonpolar and polar residues are 35.7 and 12.2% for AaeRNH2, 38.7 and 10.5% for TmaRNH2, and 38.0 and 11.0% for BstRNH2ΔN, respectively. The number of proline residues at loop regions is five for AaeRNH2, two for TmaRNH2, and ten for BstRNH2ΔN. None of these proteins contains any disulfide bond and metal ion other than those required for activity. However, the number of ion pairs increases as the protein stability increases. The numbers of ion pairs and ion pair networks, which are calculated on the basis of the assumption that ion pairs are formed between two oppositely charged groups located within a distance of 5 Å and ion pair networks are formed by at least four

charged residues, are 11 and 1 for AaeRNH2, 8 and 1 for TmaRNH2 without the C-terminal extension region, and 6 and 0 for BstRNH2ΔN, respectively. In addition, AaeRNH2 contain four aromatic interactions, whereas TmaRNH2 and BstRNH2ΔN do not contain it. These results suggest that stabilization by ion pairs and aromatic interactions at least partly account for the difference in stability between AaeRNH2 and TmaRNH2 or BstRNH2.

2.5. Summary

RNases H greatly differ in domain structures. RNases H1 and H3 contain HBD- and TBP-like-domain, respectively, which are important for activity and substrate binding. Comparison of the amino acid sequences of bacteria RNases H2 indicates that some RNases H2 contain an N- or C-terminal extension, whereas some RNases H2 do not contain these extensions. BstRNH2 represents RNases H2 with N-terminal extension. TmaRNH2 represents RNases H2 with C-terminal extension. AaeRNH2 represents RNases H2 without N- and C-terminal extensions. To analyze the role of these extensions, I overproduced, purified, and characterized these proteins and their derivatives without N- or C-terminal extension, and compared their biochemical properties with those of AaeRNH2.

The removal of a 59-residue N-extension of BstRNH2 (BstNTD) decreased its activity and substrate binding affinity, indicating that BstNTD is required for maximal activity and substrate binding affinity of BstRNH2. This role is similar to those of HBD and TBP-like domain. However, BstNTD does not show an amino acid sequence similarity to HBD or TBP-like domain, suggesting that it assumes a unique structure. The removal of a 46-residue C-extension of TmaRNH2 also decreased the activity, substrate binding affinity, and stability of TmaRNH2. According to the crystal structure of TmaRNH2, however, C-extension of TmaRNH2 assumes a hairpin helix structure, which is not directly involved in substrate binding. Nevertheless, this domain is important for substrate binding and activity. Thus, C-terminal hairpin helix structure of TmaRNH2 is probably required to make the conformation of the C-terminal domain of TmaRNH2 functional. The C-terminal domain of TmaRNH2 is directly involved in substrate binding.

The attachment of BstNTD to the N-terminus of the TmaRNH2 derivative lacking the C-terminal extension (TmaRNH2ΔC), which lost the activity and substrate binding

affinity as compared to TmaRNH2, increased its activity and substrate binding affinity. It suggests that BstNTD can function as a substrate-binding domain. BstNTD directs the enzyme such that the RNase H domain can simultaneously contact the substrate.

The finding that AaeRNH2 is as active as and is more stable than BstRNH2 and TmaRNH2 suggests that bacterial RNases H2 do not always require N- or C-terminal extension to increase the activity, substrate-binding affinity, and/or stability.

CHAPTER 3

Enzymatic activities of RNase H domains of HIV-1 reverse transcriptase with substrate binding domains of bacterial RNase H1 and H2

3.1. Introduction

In Chapter 2, I showed that bacteria RNases H2 have a substrate-binding domain, termed as NTD (N-terminal domain). The NTD of BstRNH2 (BstNTD) was shown to be important for substrate binding, activity, and stability, and its role is similar to those of HBD of RNase H1 and TBP-like domain of RNase H3. The finding that BstNTD could restore the activity and stability of TmaRNH2ΔC suggests that BstNTD acts as a potent substrate-binding domain in an isolated form. However, it remains to be determined whether NTD, HBD or TBP-like domain can restore the activity of another enzyme lacking the RNase H activity.

As described in Chapter 1, RNase H is present in retroviruses, including human immunodeficiency virus type 1 (HIV-1), as a C-terminal domain of reverse transcriptase (RT) (Fig. 1.7 and Fig. 1.9) [33]. Because RNase H activity of HIV-1 RT is required for viral proliferation [32, 33], this activity is regarded as a target for AIDS therapy [35, 36]. HIV-1 RT is a heterodimer consisting of P66 subunit and a P51 subunit [36], as shown in Figure 3.1. It prefers Mn^{2+} to Mg^{2+} ions for RNase H activity [68]. Only P66 subunit retains the RNase H domain at the C-terminus (Fig. 3.1).

The isolated RNase H domain of HIV-1 RT (P15, termed as RNH^{HIV} in this thesis) is inactive [39, 68-71] or exhibits very weak activity only in the presence of Mn^{2+} ions [72]. Nevertheless, the structure of RNH^{HIV} [39] highly resembles that of the RNase H domain of HIV-1 RT [99], except that a loop containing the conserved histidine residue

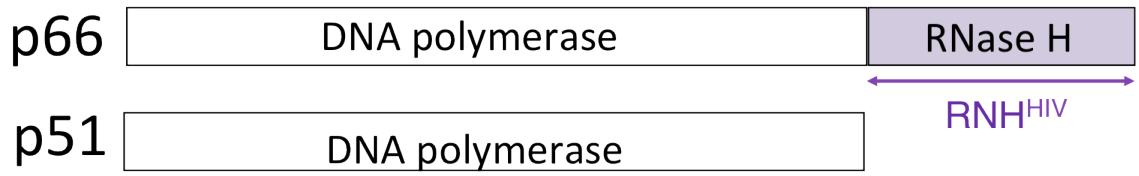


Fig. 3.1. Schematic representation of the subunits in heterodimeric form of HIV-1 RT. Relative positions of the DNA polymerase domain and RNase H domain are shown. The RNase H domain at the C-terminal portion of P66 is shown in purple. This region, which starts from Tyr427, forms a stable domain. The crystal structure of this region is available (PDB code 1HRH) [39].

(His539) is disordered in the RNH^{HIV} structure. It has been suggested that RNH^{HIV} is inactive due to the lack of a substrate binding ability, disorder of a loop containing His539, and increased flexibility [39, 68-71, 100-102].

Insertion of a basic protrusion of EcoRNH1, which is important for substrate binding [73, 74], into the equivalent position of RNH^{HIV} only restores its Mn^{2+}-dependent RNase H activity [68, 71]. Restoration of the Mg^{2+}-dependent activity of RNH^{HIV} requires the presence of the P51 subunit and the thumb and connection domains of the P66 subunit [70, 102]. These results suggest that enhancement of the substrate binding affinity of RNH^{HIV} is not sufficient to restore its Mg^{2+}-dependent RNase H activity. However, a possibility that insertion of a basic protrusion of EcoRNH1 alters the conformation of the active site of RNH^{HIV}, in such a way that it is not suitable for Mg^{2+}-dependent activity, cannot be ruled out. Therefore, it would be informative to examine whether the attachment of NTD, HBD or TBP-like domain to the N-terminus of RNH^{HIV} restores its Mg^{2+}-dependent activity.

In this Chapter, I constructed two chimeric proteins, TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV}, in which HBD of TmaRNH1 (TmaHBD) and NTD of BstRNH2

(BstNTD) are attached to the N-terminus of RNH^{HIV} respectively, and characterized their biochemical properties. I did not construct a chimeric protein, in which TBP-like domain of RNase H3 is attached to the N-terminus of RNH^{HIV}, because TBP-like domain binds to RNA/DNA hybrid in a similar manner to that of HBD, as described in Chapter 1 (Fig. 1.13-C). I showed that the attachment of TmaHBD or BstNTD increases the substrate binding affinity, Mn^{2+}-dependent activity, and stability of RNH^{HIV}, but does not significantly restores its Mg^{2+}-dependent activity.

3.2. Materials and methods

3.2.1. Construction of plasmids

All primers used in this study for PCR are listed in Table 4. For construction of plasmid pET-RNH^{HIV} to overproduce RNH^{HIV}, the gene encoding RNH^{HIV} was amplified by PCR using 5'-primer 1 and 3'-primer 1. Plasmid pET-HIVMp66 to overproduce the P66 subunit of HIV-1 RT [103] was used as a template. The amplified DNA fragment was digested with *Nde*I and *Eco*RI and ligated into the *Nde*I and *Eco*RI sites of pET25b (Novagen, Madison, WI, USA).

Table 4. PCR primers. For 5'- and 3'-primers, underlined bases show the positions of the restriction sites. For 5'- and 3'-fusion primers, the residues shown in *italic* represent those of the RNH^{HIV}.

Primer	Sequence	Restriction site
5'-primer 1	5'-GGAATTCCATATGTATCAACTCGAAAAAGAACCG-3'	*Nde*I
5'-primer 2	5'-GGAATTCCATATGAAGTTGGCAAAAAAATACTACGCTG-3'	*Nde*I
5'-primer 3	5'-GGAATTCCATATGAAGGAGTACACGGTAAAAGAAATC-3'	*Nde*I
3'-primer 1	5'-GGCCGAATTCTTATAAAACTTTACGAATACCTGC-3'	*Eco*RI
5'-fusion primer 1	5'-CTGTCCAGAACTGGACACA*TATCAACTCGAAAAAGAACCGATTGTAGGGGC*-3'	
5'-fusion primer 2	5'-CGGCGTTGGGAGGAGTTG*TATCAACTCGAAAAAGAACCGATTGTAGGGGC*-3'	
3'-fusion primer 1	5'-*CGGTTCTTTTTCGAGTTGATA*TGTGTCCAGTTCTGGACAGATGCACTC-3'	
3'-fusion primer 2	5'-*CGGTTCTTTTTCGAGTTGATA*CAACTCCTCCCAACGCCGCCG-3'	

Plasmid pET-TmaHBD-RNH^{HIV} to overproduce TmaHBD-RNH^{HIV} was constructed by performing PCR twice. In the first PCR, the 210 base pair (bp) DNA fragment containing the gene encoding TmaHBD was amplified with 5'-primer 2 and 3'-fusion primer 1 by using the pET25b derivative to overproduce TmaRNH1 [76] as a template. Likewise, the 420 bp DNA fragment containing the gene encoding RNH^{HIV} was amplified with 5'-fusion primer 1 and 3'-primer 1 by using pET-HIVMp66 as a template.

These two 210 bp and 420 bp DNA fragments were mixed and amplified with 5'-primer 2 and 3'-primer 1. The amplified DNA fragment was digested with *Nde*I and *Eco*RI and ligated into the *Nde*I and *Eco*RI sites of pET25b.

Plasmid pET-BstNTD-RNH^{HIV} to overproduce BstNTD-RNH^{HIV} was also constructed by performing PCR twice. In the first PCR, the 200 bp DNA fragment containing the gene encoding BstNTD was amplified with 5'-primer 3 and 3'-fusion primer 2 by using the pET800BS to overproduce BstRNH2 (Chapter 2) as a template. Likewise, the 420 bp DNA fragment containing the gene encoding RNH^{HIV} was amplified with 5'-fusion primer 2 and 3'-primer 1 by using pET-HIVMp66 as a template. These two 200 bp and 420 bp DNA fragments were mixed and amplified with 5'-primer 3 and 3'-primer 1. The amplified DNA fragment was digested with *Nde*I and *Eco*RI and ligated into the *Nde*I and *Eco*RI sites of pET25b.

All PCR primers were synthesized by Hokkaido System Science (Sapporo, Japan). PCR was performed by a GeneAmp PCR system 2400 (Applied Biosystems) using KOD polymerase (Toyobo, Japan). All DNA sequences were verified with the ABI Prism 310 DNA sequencer (Applied Biosystems).

3.2.2. Preparation of proteins

For overproduction of the proteins, *E. coli* MIC2067(DE3) transformants with the pET25 derivatives were grown and cultivated. When the A_{600} value of the culture reached approximately 0.5, 1 mM IPTG was added to the culture medium and cultivation was continued for an additional 4 h. The subsequent purification procedures were carried out at 4°C.

Purification of RNH^{HIV} was carried out as described for AaeRNH2 in Chapter 2, except that sonication was used for lysis, instead of French press, and the gel filtration chromatography using a HiLoad 16/60 Superdex 200pg column (GE Healthcare) equilibrated with 20 mM sodium acetate (pH 5.5) containing 150 mM NaCl was newly performed at the end of the purification procedures. The fractions containing the protein were collected and used for biochemical characterization.

TmaHBD-RNH^{HIV} was purified as mentioned above for RNH^{HIV}, except that the first HiTrap Q HP column chromatography procedure was eliminated. BstNTD-RNH^{HIV} was also purified as mentioned above for RNH^{HIV}, except that 20 mM Tris-HCl (pH 7.0), instead of 20 mM sodium acetate (pH 5.5), was used for the last gel filtration column chromatography and the same buffer, instead of buffer A, was used for other three column chromatography procedures.

The purity of the protein was analyzed by SDS/PAGE with a 15% polyacrylamide gel [83]. The protein concentration was determined from UV absorption using a cell with an optical path length of 1.0 cm and an A_{280} value of 0.1% solution (1.0 mg·mL^{-1}) of 0.99 for RNH^{HIV}, 1.25 for TmaHBD-RNH^{HIV}, and 1.21 for BstNTD-RNH^{HIV}. These values were calculated by using absorption coefficients of 1576 M^{-1}·cm^{-1} for Tyr and 5225 M^{-1}·cm^{-1} for Trp at 280 nm [84]. HIV-1 RT (0.5 mg·mL^{-1}) was obtained from BioAcademia Inc. (Osaka, Japan).

3.2.3. Determination of enzymatic activity

The RNase H activity was determined by using oligomeric substrates. 5'-Fluorescein-labeled 29 bp RNA/DNA hybrid (R29/D29), 12 bp RNA/RNA duplex (R12/R12), and 12 bp DNA/DNA duplex (D12/D12) were prepared by hybridizing 5'-

fluorescein-labeled 29 b RNA (5'-aauagagaaaaagaaaaaagauggcaaag-3'), 12 b RNA (5'-cggagaugacgg-3'), and 12 b DNA (5'-CGGAGAUGACGG-3') with a 1.5 molar equivalent of the complementary DNA, as described previously [7]. In these sequences, DNA and RNA are represented by capital and lowercase letters respectively. All oligonucleotides were synthesized by Hokkaido System Science. The concentration of the substrate in the reaction mixture (10 μL) was 1 μM. Hydrolysis of the substrate at 30°C for 15 min and separation of the products on a 20% polyacrylamide gel containing 7 M urea were carried out as described previously [7]. The reaction buffers were 20 mM Tris-HCl (pH 8.0) containing 10 mM $MnCl_2$ or 50 mM $MgCl_2$, 25 mM KCl, 1 mM 2-mercaptoethanol (2-ME), and 10 $\mu g \cdot mL^{-1}$ bovine serume albumin (BSA) for RNH^{HIV}, 50 mM Tris-HCl (pH 8.3) containing 10 mM $MnCl_2$ or $MgCl_2$, 50 mM KCl, 3 mM 2-ME, and 10 $\mu g \cdot mL^{-1}$ BSA for HIV-1 RT, 20 mM Tris-HCl (pH 8.5) containing 10 mM $MnCl_2$ or 50 mM $MgCl_2$, 50 mM KCl, 1 mM 2-ME, and 10 $\mu g \cdot mL^{-1}$ BSA for TmaHBD-RNH^{HIV}, and the same buffer for TmaHBD-RNH^{HIV} without 50 mM KCl for BstNTD-RNH^{HIV}. The products were detected with Typhoon 9210 Imager (GE Healthcare), and quantified with IMAGEQUANT 5.2. One unit was defined as the amount of enzyme hydrolyzing 1 nmol of the substrate per min at 30°C**.** The specific activity was defined as the enzymatic activity per μmol of protein.

3.2.4. Measurement of CD spectra

The CD spectra was measured at 25°C as described in Chapter 2. The proteins were dissolved in 20 mM Tris-HCl (pH 8.0) at a concentration of ~0.1 $mg \cdot mL^{-1}$ for far-UV CD spectra and ~1.0 $mg \cdot mL^{-1}$ for near-UV CD spectra.

3.2.5. Binding analysis to substrate

The binding analysis was carried out as described in Chapter 2, except that the R29/D29 was used as a substrate. The proteins were dissolved in 20 mM Tris-HCl (pH 8.0) containing 50 mM NaCl, 1 mM 2-ME, 1 mM EDTA and 0.005% Tween 20 at various concentrations, and injected onto the sensor chip at 25°C with a flow rate of 20 μL·min^{-1}. The binding surface was regenerated by washing with 2 M NaCl. The association constant K_A was calculated from the equilibrium levels of the protein binding to the surface, as previously described [74].

3.2.6 GdnHCl-induced unfolding

The Guanidine hydrochloride (GdnHCl) denaturation curves of the proteins were obtained at 25°C by monitoring the change in CD values at 222 nm as the GdnHCl concentration was increased. The proteins were dissolved in 20 mM Tris-HCl (pH 8.0) containing various concentrations of GdnHCl and incubated for 1 h at 25°C prior to the measurement. The protein concentration was ~0.1 mg·mL^{-1} and the optical path length was 0.2 cm. GdnHCl-induced folding of all proteins examined was fully reversible. On the assumption that the unfolding equilibria of RNH^{HIV} and BstNTD-RNH^{HIV} follow a two-state mechanism, the pre- and post-transition base lines were extrapolated linearly, and the difference in free energy between the folded and unfolded states, ΔG, and the free energy change of unfolding in H_2O, $\Delta G(H_2O)$, were calculated by the equations given by Pace [104].

3.3. Results & Discussion

3.3.1. Preparation of RNH^{HIV} derivatives

T. maritima RNase H1 (TmaRNH1, residues 1-223) consists of a hybrid binding domain (TmaHBD, residues 1-63) and an RNase H domain (residues 64-223). *B. stearothermophilus* RNase H2 (BstRNH2, residues 1-259) consists of an N-terminal substrate binding domain (BstNTD, residues 1-59) and an RNase H domain (residues 60-259). The isolated RNase H domain of HIV-1 RT (RNH^{HIV}, residues 427-560) is correctly folded but is inactive [39]. To examine whether the RNase H activity of RNH^{HIV} is restored by the attachment of TmaHBD or BstNTD, two chimeric proteins, TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV}, in which TmaHBD and BstNTD are attached to the N-terminus of RNH^{HIV} respectively, were constructed. The primary structures of these chimeric proteins are shown in Figure 3.2 in comparison with those of TmaRNH1 and BstRNH2. These chimeric proteins, as well as RNH^{HIV}, were overproduced in *E. coli*

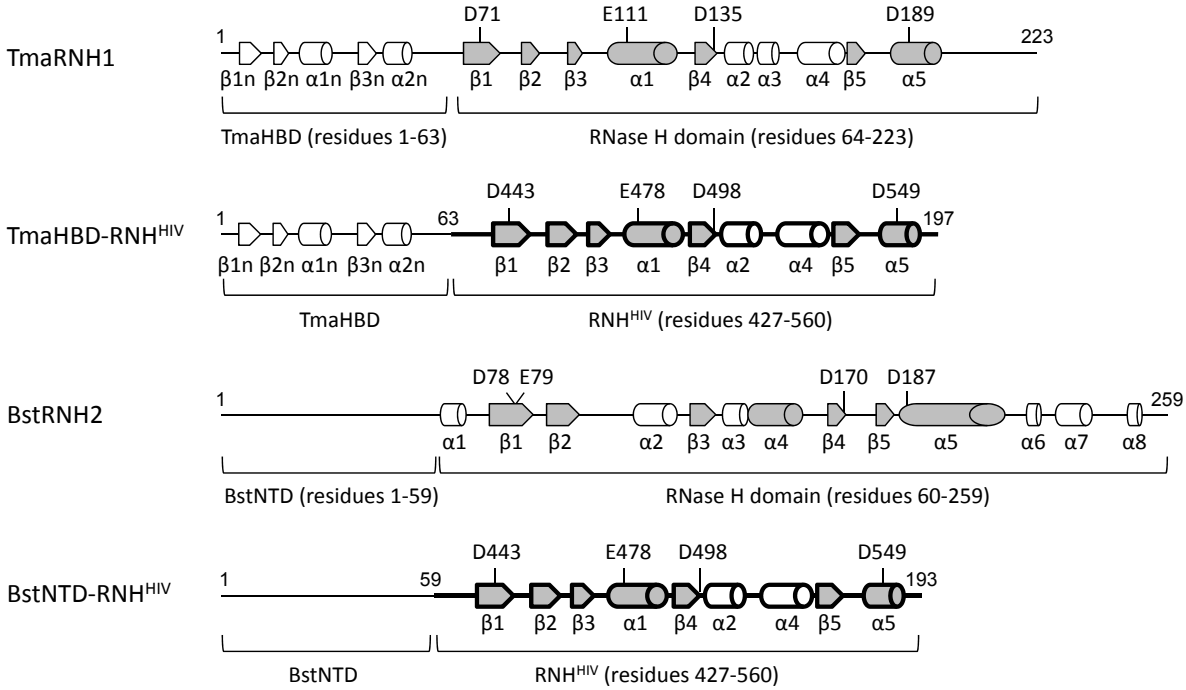


Fig. 3.2. Schematic representation of the primary structures of TmaRNH1, BstRNH2 and two chimeric proteins. The TmaRNH1 and BstRNH2 sequences are indicated by thin line, and the RNH^{HIV} sequence is indicated by thick line. The α-helices and β-strands are represented by cylinders and arrows respectively. These secondary structures of TmaHBD, RNase H domain of TmaRNH1, and RNase H domain of BstRNH2 are arranged based on their tertiary models, which are built with the SWISS-MODEL program [85], with the crystal structures of HBD of human RNase H1 (PDB code 3BSU) [78], RNase H domain of human RNase H1 (PDB code 2QK9) [42], and TmaRNH2 (PDB code 3O3F) [9] respectively, as template structures. The secondary structures of RNH^{HIV} are arranged based on the crystal structure of the isolated RNase H domain of HIV-1 RT (PDB code 1HRH) [39]. Five β-strands and two α-helices that form an "RNase H-fold" and are shared by RNases H1 and H2 are shaded. The positions of the four acidic active site residues are also shown. For the positions of these residues in the TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} sequences, numbering in HIV-1 RT are used. The numbers indicate the positions of the residues relative to the initiator methionine residue for each protein. The ranges of the substrate binding domain and RNase H domain are also shown for each protein.

MIC2067(DE3) in a soluble form and purified to give a single band on SDS-PAGE (Fig. 3.3). I used *E. coli* MIC2067(DE3) as a host strain for overproduction of RNH^{HIV} and its derivatives to avoid contamination by host-derived RNases H, because this strain lacks all functional genes encoding RNases H1 and H2 [63]. The amount of the protein purified from 1 L culture was typically 6 mg for TmaHBD-RNH^{HIV} and RNH^{HIV}, and 4 mg for BstNTD-RNH^{HIV}. All of them were eluted from the column as single peaks in gel filtration column chromatography. The molecular masses of these proteins were estimated to be 23 kDa for TmaHBD-RNH^{HIV}, 25 kDa for BstNTD-RNH^{HIV}, and 12 kDa for RNH^{HIV} from this column chromatography. These values are comparable to those calculated from the amino acid sequences (22,078 Da for TmaHBD-RNH^{HIV}, 22,057 Da for BstNTD-RNH^{HIV}, and 14,773 Da for RNH^{HIV}), suggesting that these proteins exist as a monomer in solution. It has been reported that His-tagged RNH^{HIV} exists as a monomer [107], whereas RNH^{HIV}

with a basic protrusion of EcoRNH1 exists partly as a monomer and partly as a dimer [109] in solution. These results suggest that RNH^{HIV} acquires an ability to form a dimer by the insertion of a basic protrusion of EcoRNH1.

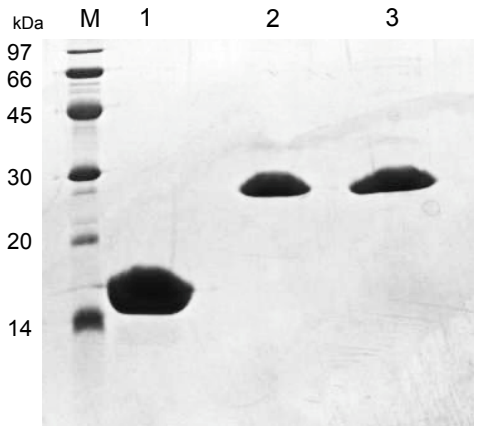


Fig. 3.3. SDS-PAGE of the purified proteins. RNH^{HIV} (lane 1), BstNTD-RNH^{HIV} (lane 2), and TmaHBD-RNH^{HIV} (lane 3) were subjected to electrophoresis on a 15% polyacrylamide gel in the presence of SDS. After electrophoresis, the gel was stained with Coomassie Brilliant Blue. Lane M, a low-molecular-weight marker kit (GE Healthcare).

3.3.2. Complementation of the temperature-sensitive growth phenotype of *E. coli* MIC2067(DE3)

E. coli MIC2067(DE3) shows an RNase H-dependent temperature-sensitive (ts) growth phenotype [53]. A relatively low level of Mg^{2+}-dependent RNase H activity is sufficient to complement the ts growth phenotype of this strain [14, 53, 76]. However, the genes encoding TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} did not complement the ts growth phenotype of this strain (data not shown), suggesting that these proteins do not exhibit Mg^{2+}-dependent RNase H activity or exhibit a very low level of this activity.

3.3.3. CD Spectra

The far-UV CD spectra of proteins mainly reflect their helical contents. These spectra usually give a broad trough with a minimum [θ] value at around 220 nm. As the depth of this trough increases, the helical content of protein increases. The far-UV CD spectrum of TmaHBD-RNH^{HIV} was similar to that of RNH^{HIV}, which was similar to that

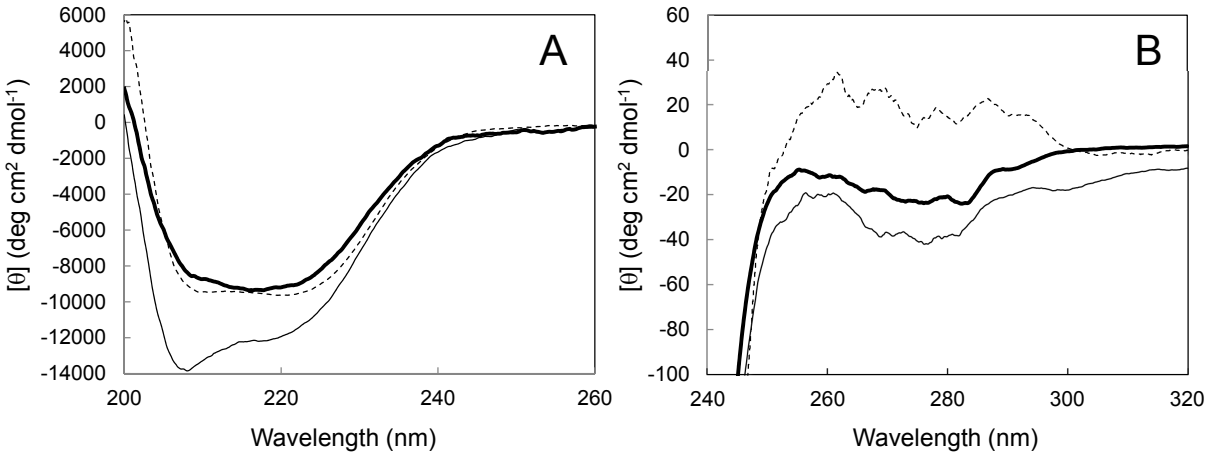


Fig. 3.4. CD spectra. The far-UV (A) and near-UV (B) CD spectra of RNH^{HIV} (thin dashed line), BstNTD-RNH^{HIV} (thin solid line), and TmaHBD-RNH^{HIV} (thick solid line) are shown. These spectra were measured at 25°C as described in Section 3.2.4.

previously reported [68] both in shape and depth of a trough, at 25°C (Fig. 3.4-A). In contrast, the spectrum of BstNTD-RNH^{HIV} was different from that of RNH^{HIV} (Fig. 3.4-A). The spectra of TmaHBD-RNH^{HIV} and RNH^{HIV} give a broad trough with a minimum [θ] value of -9,000 at 215-220 nm, whereas the spectrum of BstNTD-RNH^{HIV} gives a minimum [θ] value of -14,000 at 208 nm, which is accompanied by a shoulder with a [θ] value of -12,000 at 220 nm. These results suggest that the helical contents of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} are similar to and higher than that of RNH^{HIV} respectively. In fact, the helical content of TmaHBD-RNH^{HIV} (35%) is similar to that of RNH^{HIV} (38%) calculated from the crystal structure of RNH^{HIV} [39]. The helical content of TmaHBD-

RNH^{HIV} is calculated on the assumption that the TmaHBD and RNH^{HIV} regions of this protein are independently folded and the structure of TmaHBD is similar to that of HBD of human RNase H1 determined by X-ray crystallography [78]. The helical content of BstNTD-RNH^{HIV} is higher than that of RNH^{HIV}, probably because the helical content of BstNTD is higher than that of RNH^{HIV}. Thus, the structure of RNH^{HIV} may not be significantly changed by the attachment of TmaHBD or BstNTD.

The near-UV CD spectra of proteins reflect their local conformation around the aromatic residues, such as Trp and Tyr. The spectra of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} were different from that of RNH^{HIV} (Fig. 3.4-B). However, this result does not necessarily indicate that the local conformation of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} around the aromatic residues is different from that of RNH^{HIV}, because TmaHBD and BstNTD also contain the aromatic residues. TmaHBD contains one tryptophan and five tyrosine residues. BstNTD contains two tryptophan and one tyrosine residues. Increase in the number of these aromatic residues may be responsible for the change in the near-UV CD spectra of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} as compared to that of RNH^{HIV}.

3.3.4. Binding affinities for substrate

To examine whether the attachment of TmaHBD or BstNTD increases the substrate binding affinity of RNH^{HIV}, binding of HIV-1 RT, RNH^{HIV}, TmaHBD-RNH^{HIV}, and BstNTD-RNH^{HIV} to the R29/D29 substrate was analyzed in the absence of the metal cofactor, with the surface plasmon resonance (Biacore system). These proteins were injected onto the sensor chip, on which the R29/D29 substrate was immobilized. When 1 μM of BSA was injected onto the sensor chip, no increase of RU was detected, whereas when 1 μM of RNH^{HIV}, TmaHBD-RNH^{HIV}, or BstNTD-RNH^{HIV} was injected onto the

sensor chip, an increase of RU was detected (Fig. 3.5). However, the RU value of TmaHBD-RNH^{HIV} or BstNTD-RNH^{HIV} in binding equilibrium was several-fold higher than that of RNH^{HIV}. When 1 μM of HIV-1 RT was injected onto the sensor chip, a great increase of RU was detected (data not shown). This result is not shown in Figure 3.5, because the RU value of HIV-1 RT in equilibrium binding was higher than that of

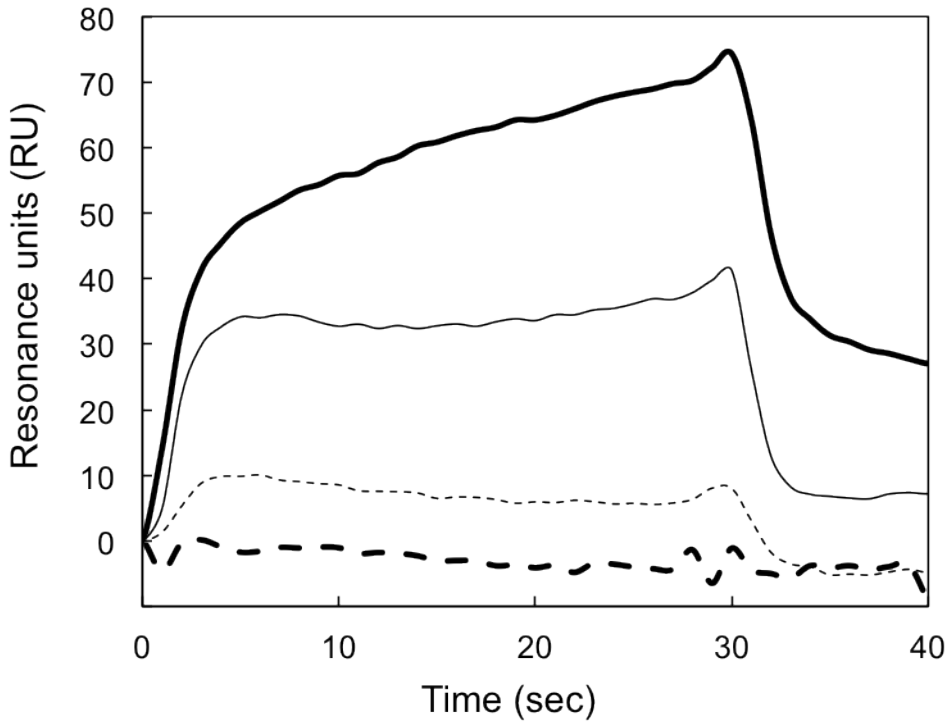


Fig. 3.5. Binding of the proteins to R29/D29 substrate. The sensorgrams showing binding of 1μM of RNH^{HIV} (thin dashed line), BstNTD-RNH^{HIV} (thin solid line), TmaHBD-RNH^{HIV} (thick solid line), and BSA (thick dashed line) to the immobilized R29/D29 substrate are shown. Injections were performed at time zero for 30 s.

TmaHBD-RNH^{HIV} by approximately 200 fold. The association constants, K_A, of these proteins for binding to the substrate, which were determined by measuring the equilibrium-binding responses at various concentrations of the proteins, are summarized in Table 5. The K_A values of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} are lower than that of HIV-1

RT by approximately 4 and 50 fold respectively, but higher than that of RNH^{HIV} by 260 and 22 fold respectively. These results indicate that the attachment of TmaHBD or BstNTD increases the substrate binding affinity of RNH^{HIV}.

The dissociation constants of TmaRNH1 and TmaHBD have been determined to be 0.16 and 0.40 μM respectively, using the same R29/D29 substrate [76]. This means that the K_A values of TmaRNH1 and TmaHBD for this substrate are 6.3 x 10^6 and 2.5 x 10^6 respectively. The K_A value of TmaHBD-RNH^{HIV} is lower than that of TmaHBD by 23 fold, indicating that the binding affinity of TmaHBD for the R29/D29 substrate is greatly reduced when it is attached to the N-terminus of RNH^{HIV}. Because the TmaHBD domain of TmaHBD-RNH^{HIV} governs binding of this protein to the substrate, the RNH^{HIV} domain of this protein may be located in a position, where it suppresses effective binding of the TmaHBD domain to the substrate. In contrast, the K_A value of TmaRNH1 is higher than that of TmaHBD by 2.5 fold. It is also higher than that of the isolated RNase H domain of

Table 5. Association constants of the proteins for substrate binding. [a]Binding of the proteins to R29/D29 substrate immobilized onto the sensor chip was analyzed in the absence of the metal cofactor by surface plasmon resonance as described in Section 3.2.5. [b]The K_A values of the proteins relative to that of HIV-1 RT are shown.

Protein	K_A x 10^{-4} (M^{-1})	Relative K_A[b] (%)
HIV-1RT	41.9	100
RNH^{HIV}	0.04	0.1
BstNTD-RNH^{HIV}	0.9	2.2
TmaHBD-RNH^{HIV}	10.9	26

TmaRNH1 by 50 fold [76]. The TmaHBD and RNase H domains of TmaRNH1 cooperatively increase the substrate binding affinity of TmaRNH1, probably because these domains are arranged optimally for effective binding of the substrate.

3.3.5. Stability

To examine whether the attachment of TmaHBD or BstNTD increases the stability of RNH^{HIV}, stability of RNH^{HIV}, TmaHBD-RNH^{HIV}, and BstNTD-RNH^{HIV} against GdnHCl-induced unfolding was analyzed at 25°C by monitoring the change in CD values at 222 nm as the GdnHCl concentration was increased. The GdnHCl-induced unfolding curves of these proteins are shown in Figure 3.6. To obtain these curves, the proteins were

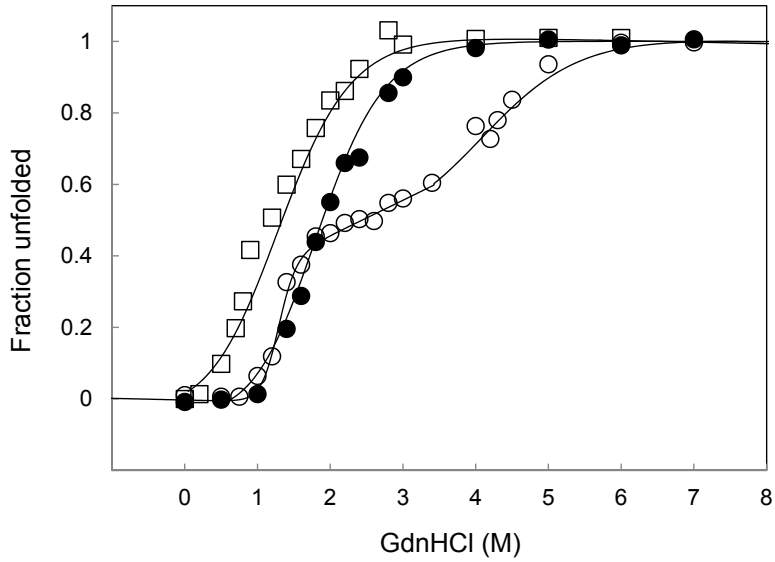


Fig. 3.6. GdnHCl-induced unfolding of the proteins. The GdnHCl-induced unfolding curves of RNH^{HIV} (open squares), BstNTD-RNH^{HIV} (closed circles), and TmaHBD-RNH^{HIV} (open circles) measured at 25 °C are shown. The apparent fraction of unfolded protein is shown as a function of GdnHCl concentration. The theoretical lines of RNH^{HIV} and BstNTD-RNH^{HIV} are drawn on the assumption that these proteins are unfolded *via* a two-state mechanism. The line of TmaHBD-RNH^{HIV} represents the optimal fit to the data.

incubated in the presence of various concentrations of GdnHCl for 1 h prior to the measurement. The curves shown in Figure 3.6 were not significantly changed when this incubation time was increased to 18 h, indicating that the unfolding reactions of these proteins reach equilibrium within 1 h. GdnHCl-induced unfolding of RNH^{HIV} and BstNTD-RNH^{HIV} showed a two-state transition. The parameters characterizing GdnHCl-induced unfolding of these proteins are summarized in Table 6. The apparent free energy change of unfolding in the absence of denaturant, $\Delta G(H_2O)$, and midpoint of the unfolding curve, C_m, of BstNTD-RNH^{HIV} were higher than those of RNH^{HIV} by 2.79 $kJ \cdot mol^{-1}$ and 0.73 M respectively. GdnHCl-induced unfolding of TmaHBD-RNH^{HIV} did not show a two-state transition. Apparently, TmaHBD-RNH^{HIV} unfolds through an intermediate state. In this state, only the RNH^{HIV} region may be fully unfolded, because it has been reported that TmaHBD is fully stable at 25°C in the presence of 3 M GdnHCl [76]. Further studies will be required to understand the unfolding reaction of this protein. However, the result that TmaHBD-RNH^{HIV} is almost fully folded in the presence of 1.0 M GdnHCl, in which RNH^{HIV} is unfolded by 50%, indicates that TmaHBD-RNH^{HIV} is more stable than RNH^{HIV}. Thus, the attachment of TmaHBD or BstNTD increases the stability of RNH^{HIV}.

Table 6. Parameters characterizing GdnHCl-induced unfolding of RNH^{HIV} and BstNTD-RNH^{HIV}

Protein	C_m (M)	m (kJ·mol^{-1}·M^{-1})	$\Delta G(H_2O)$ (kJ·mol^{-1})	$\Delta\Delta G(H_2O)$ (kJ·mol^{-1})
RNH^{HIV}	1.01	6.39	6.45	-
BstNTD-RNH^{HIV}	1.74	5.31	9.24	2.79

[a]GdnHCl-induced unfolding of these proteins was reversible in this condition. The midpoint of the GdnHCl-induced unfolding curve (C_m), the measurement of the dependence of ΔG on the GdnHCl concentration (m), and the free energy change of unfolding in H_2O ($\Delta G(H_2O)$) were calculated from the GdnHCl-induced unfolding curves shown in Figure 3.7. The difference in $\Delta G(H_2O)$ ($\Delta\Delta G(H_2O)$) between RNH^{HIV} and BstNTD-RNH^{HIV} was calculated as $\Delta G(H_2O)$ of BstNTD-RNH^{HIV} - $\Delta G(H_2O)$ of RNH^{HIV}. Errors are within ±0.07 M for C_m, ±0.58 kJ.mol^{-1}.M^{-1} for m, and ±0.8 kJ.mol^{-1} for $\Delta G(H_2O)$.

3.3.6. Enzymatic activities

To examine whether the attachment of TmaHBD or BstNTD increases the enzymatic activity of RNH^{HIV}, the activities of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} were determined at 30°C using the R29/D29 substrate. The activities of RNH^{HIV} and HIV-1 RT were also determined for comparative purpose. The R29/D29 substrate is a substrate for non-specific RNase H assay, instead of that for tRNA^{Lys3} removal specificity assay (tRNA^{Lys3} substrate). This substrate is identical to that used to analyze the substrate-binding affinities of the proteins as mentioned above. This substrate has also been used to analyze the substrate-binding affinity of TmaRNH1 [76] and the activity of *Thermococcus kodakarensis* RNase H2 (TkoRNH2) [64]. The activities of all proteins were determined at their optimum conditions, which were pH 8.5, 10 mM $MnCl_2$, and 50 mM KCl for TmaHBD-RNH^{HIV}, pH 8.5 and 10 mM $MnCl_2$ or 50 mM $MgCl_2$ for BstNTD-RNH^{HIV}, pH

8.0, 10 mM $MnCl_2$, and 25 mM KCl for RNH^{HIV}, and pH 8.3, 10 mM $MgCl_2$ or 10 mM $MnCl_2$, and 50 mM KCl for HIV-1 RT. The optimum pHs, metal ion concentrations, and salt concentrations for the activities of TmaHBD-RNH^{HIV}, BstNTD-RNH^{HIV}, and RNH^{HIV} were determined by changing either one of these conditions. The optimum condition for the activity of HIV-1 RT has been reported previously [72]. The results are shown in Figure 3.7.

HIV-1 RT cleaved this substrate most effectively at a18-a19 and less effectively at a9-a10, a10-a11, a17-a18, and a19-a20 only in the presence of Mg^{2+} ions. In contrast, RNH^{HIV} cleaved this substrate very weakly mainly at g7-a8, c25-a26, and a26-a27 only in the presence of Mn^{2+} ions, indicating that it only exhibits a very weak Mn^{2+}-dependent activity, as reported previously [72]. TmaHBD-RNH^{HIV} cleaved this substrate most effectively at g7-a8, c25-a26, and a26-a27, and less effectively at g5-a6, a6-g7, a8-a9, a10-a11, g13-a14, and a16-a17 only in the presence of Mn^{2+} ions. BstNTD-RNH^{HIV} cleaved this substrate most effectively at g7-a8, c25-a26, and a26-a27, and less effectively at a4-g5, g5-a6, a8-a9, a9-a10, and g13-a14, either in the presence of Mn^{2+} ions or Mg^{2+} ions, but more effectively in the presence of Mn^{2+} ions than in the presence of Mg^{2+} ions.

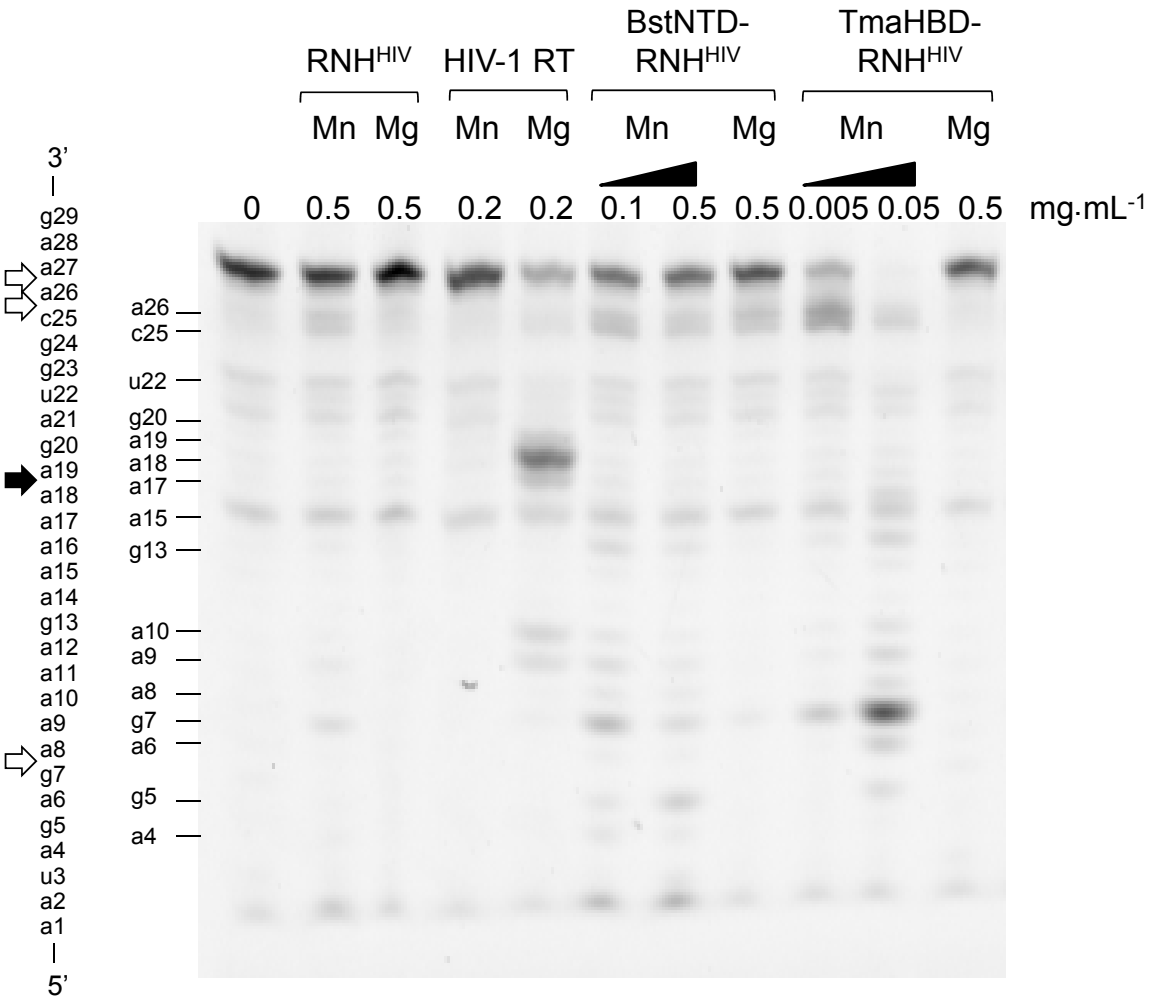


Fig. 3.7. Cleavage of the R29/D29 substrate with RNH^{HIV}, HIV-1RT, BstNTD-RNH^{HIV}, and TmaHBD-RNH^{HIV}. The 5'-end labeled R29/D29 substrate was hydrolyzed by the enzyme in the presence of 10 mM $MnCl_2$ (Mn) or 10 or 50 mM $MgCl_2$ (Mg) at 30°C for 15 min, and the hydrolysates were separated on a 20% polyacrylamide gel containing 7 M urea, as described in 3.2.6. The concentration of the substrate was 1.0 μM. The concentration of the enzyme in the reaction mixture is indicated above each lane. The positions of the products of the enzymatic reaction and their 3'-terminal residues are shown along the gel. These products were identified by comparing their migration in the gel with that of the products of EcoRNH1 in the presence of Mn^{2+} ions. Partial cleavage of the R29/D29 substrate with EcoRNH1 in the presence of Mn^{2+} ions produces all possible oligonucleotides due to cleavage of the substrate at all possible sites between a2 and a28 (data not shown). The entire sequence of the RNA strand of the R29/D29 substrate is shown at the left side of the gel. In this sequence, the most effective cleavage sites with HIV-1 RT (solid arrow) and RNH^{HIV}, HIV-1RT, or BstNTD-RNH^{HIV} (open arrow) are also shown.

The specific activities of HIV-1 RT, RNH^{HIV}, TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} were calculated by estimating the amount of the substrate cleaved by these proteins from the gel shown in Figure 3.7. Because these proteins greatly differ in

Table 7. Specific activities and kinetic parameters of the proteins

Protein	Metal	Specific activity (unit·μmol^{-1})	Relative activity [a] (%)	K_m (μM)	k_{cat} (min^{-1})
HIV-1RT	10 mM $MgCl_2$	460 ± 8	100	< 0.08	0.48 ± 0.03 [c]
	10 mM $MnCl_2$	< 12			
RNH^{HIV}	50 mM $MgCl_2$	< 1.5			
	10 mM $MnCl_2$	3.0 ± 0.8	0.7	> 2.3	>0.006
BstNTD-RNH^{HIV}	50 mM $MgCl_2$	3.5 ± 0.3	0.8		
	10 mM $MnCl_2$	40 ± 1.5	8.7	0.47 ± 0.05	0.059 ± 0.007
TmaHBD-RNH^{HIV}	50 mM $MgCl_2$	< 2.0			
	10 mM $MnCl_2$	1500 ± 11	330	< 0.08	1.58 ± 0.11 [b]

Hydrolyses of the R29/D29 substrate by the proteins were performed at 30°C under the conditions described in 3.2.6. For determination of the specific activities and kinetic parameters, the substrate concentrations were 1.0 μM and varied from 0.08 to 2.3 μM respectively. Experiments were carried out at least twice and the average values are shown together with the errors. [a]The specific activities of the proteins relative to that of HIV-1 RT determined in the presence of 10 mM $MgCl_2$ are shown. [b]The k_{cat} values calculated from the specific activities determined at the substrate concentration of 2.3 μM are shown.

molecular masses (117 kDa for HIV-1 RT, 15 kD for RNH^{HIV}, and 22 kDa for TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV}), but contain only one active site, the specific activities of these proteins were calculated based on their molar numbers, instead of the weights. The results are summarized in Table 7. The Mn^{2+}-dependent activity of RNH^{HIV} was lower than the Mg^{2+}-dependent activity of HIV-1 RT by 150 fold. The Mn^{2+}-dependent activity of TmaHBD-RNH^{HIV} was higher than that of RNH^{HIV} by 500 fold and the Mg^{2+}-dependent

activity of HIV-1 RT by 3.3 fold. The Mn^{2+}-dependent activity of BstNTD-RNH^{HIV} was higher than that of RNH^{HIV} by 13 fold and lower than that of the Mg^{2+}-dependent activity of HIV-1 RT by 12 fold. The Mg^{2+}-dependent activity of BstNTD-RNH^{HIV} was lower than that of HIV-1 RT by 11 fold and the Mg^{2+}-dependent activity of HIV-1 RT by 130 fold. These results indicate that the attachment of TmaHBD or BstNTD greatly increases the Mn^{2+}-dependent activity of RNH^{HIV}, but does not significantly restores its Mg^{2+}-dependent activity. The Mn^{2+}-dependent RNase H activity of HIV-1 RT, and Mg^{2+}-dependent activities of RNH^{HIV} and TmaHBD-RNH^{HIV} were determined to be <12, <1.5, and <2.0 units·μmol^{-1} respectively, because these enzymes exhibited a background level of activity even when their concentrations in the reaction mixtures for assay were increased to 0.5 mg·mL^{-1} for RNH^{HIV} and TmaHBD-RNH^{HIV}, and 0.2 mg·mL^{-1} for HIV-1 RT.

To examine whether the attachment of TmaHBD or BstNTD only affects substrate binding of RNH^{HIV} or affects its catalysis as well, the enzymatic activities of HIV-1 RT, RNH^{HIV}, TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} were determined in the presence of various concentrations (0.08-2.3 μM) of the R29/D29 substrate in their optimum conditions at 30°C. However, only the cleavage of this substrate with BstNTD-RNH^{HIV} followed Michaelis-Menten kinetics, and the K_m and k_{cat} values were determined to be 0.47μM and 0.059 min^{-1} from the Lineweaver-Burk plot (Table 7). The cleavages of this substrate with other three proteins did not follow Michaelis-Menten kinetics, because the activity of RNH^{HIV} increased as the substrate concentration increased up to 2.3 μM, and the activities of HIV-1 RT and TmaHBD-RNH^{HIV} were nearly unchanged regardless of the substrate concentration. The activities of HIV-1 RT and TmaHBD-RNH(HIV) were not determined at lower substrate concentrations than 0.08 μM, because the products were not clearly detected at these substrate concentrations. These results suggest that the K_m values

of these proteins are >2.3 μM for RNH^{HIV} and <0.08 μM for HIV-1 RT and TmaHBD-RNH^{HIV} (Table 7). The k_{cat} values of HIV-1 RT and TmaHBD-RNH^{HIV} were calculated to be 0.48 and 1.58 min^{-1} respectively, on the assumption that the specific activities of these proteins determined at the substrate concentration of 2.3 μM represent their V_{max} values (Table 7). These values are higher than that of BstNTD-RNH^{HIV}, which could be higher than or comparable to that of RNH^{HIV}. Comparison of the K_m values of the four proteins suggests that the binding affinity of BstNTD-RNH^{HIV} is higher than that of RNH^{HIV} and lower than that of HIV-1 RT or TmaHBD-RNH^{HIV}. This result is consistent with that determined by surface plasmon resonance. However, the k_{cat} values also vary for different proteins, suggesting that the removal of the polymerase domain of HIV-1 RT or the attachment of an alternative substrate binding domain affects turnover number (catalysis) of RNH^{HIV} as well. The removal or attachment of an additional domain may alter the conformation of the RNase H active site of HIV-1 RT.

HIV-1 RT has an ability to cleave double-stranded (ds) RNA in the presence of Mn^{2+} ions [105]. To examine whether RNH^{HIV}, TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} recognize only RNA/DNA hybrid as a substrate, their abilities to cleave dsRNA, dsDNA, single-stranded (ss) RNA and ssDNA were analyzed using the R12/R12, D12/D12, R12, and D12 substrates respectively either in the presence of Mg^{2+} ions or Mn^{2+} ions. None of these substrates was cleaved by these proteins at any condition examined, indicating that these proteins recognize only RNA/DNA hybrid as a substrate. This result indicates that the polymerase domain is required for dsRNA-cleaving activity of HIV-1 RT. This domain may be required for binding of dsRNA to HIV-1 RT and/or to make the conformation of the RNase H active site of HIV-1 RT suitable for cleavage of dsRNA.

3.3.7. Cleavage-site specificities

The cleavage-site specificities of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV}, which are similar to each other and to that of RNH^{HIV}, are different from that of HIV-1 RT (Fig. 3.7). This result suggests that the removal of the thumb and connection domains of the P66 subunit and the P51 subunit of HIV-1 RT alters the interaction between the substrate and the RNase H domain, in such a way that HIV-1 RT and RNH^{HIV} cleave the R29/D29 substrate at different sites. However, the RNH^{HIV} derivative with a basic protrusion of EcoRNH1, RNH^{HIV}(HEH), shows the same cleavage-site specificity as that of HIV-1 RT when the $tRNA^{Lys3}$ substrate is used [68]. The interaction between RNH^{HIV} and substrate may be restored by the insertion of a basic protrusion of EcoRNH1. Alternatively, RNH^{HIV} and HIV-1 RT may cleave the $tRNA^{Lys3}$ substrate at the same site, due to limitation of possible cleavage sites in this RNA-DNA/DNA substrate or strong preference of RNH^{HIV} and HIV-1 RT for the 5'-side of the ribonucleotide of the (5')RNA-DNA(3') junction. The observation that the metal ion preference of RNH^{HIV} is different from that of HIV-1 RT [68] supports the hypothesis that interaction between RNH^{HIV}(HEH) and substrate is different from that between the RNase H domain of HIV-1 RT and substrate. The attachment of TmaHBD or BstNTD greatly increases the substrate binding affinity and enzymatic activity of RNH^{HIV}. Nevertheless, it does not significantly alter the cleavage-site specificity of RNH^{HIV}, probably because the RNH^{HIV} domain of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} governs the cleavage-site specificity.

3.3.8. Metal preference

When the RNase H activity was determined using the R29/D29 substrate, RNH^{HIV} only exhibits a very weak Mn^{2+}-dependent activity and only this activity is greatly

increased by the attachment of TmaHBD or BstNTD. BstNTD-RNH^{HIV} exhibits Mg^{2+}-dependent activity, but very weakly. This activity is too low to complement the *ts* growth phenotype of *E. coli* MIC2067(DE3). In contrast, HIV-1 RT only exhibits Mg^{2+}-dependent activity. Only Mn^{2+} ions or Mg^{2+} ions can support the RNase H activities of these proteins, probably because Mn^{2+} is a transition metal ion whereas Mg^{2+} is an alkaline-earth metal ion, and therefore they differ in ionic radius and coordination geometry. Because not only the active site residues but also the substrate provide ligands for coordination of these metal ions [41], RNH^{HIV}, TmaHBD-RNH^{HIV}, and BstNTD-RNH^{HIV} may interact with the substrate, in such a way that only Mn^{2+} ions can coordinate with the substrate and the active site residues. Likewise, HIV-1 RT may interact with the substrate, in such a way that only Mg^{2+} ions can coordinate with the substrate and the active site residues. Our result that the attachment of an N-terminal substrate binding domain of bacterial RNases H1/H2 does not restore the Mg^{2+}-dependent activity of RNH^{HIV} suggest that enhancement of the substrate binding affinity of RNH^{HIV} by the attachment of an alternative substrate binding domain is not sufficient to restore its Mg^{2+}-dependent activity. Similar results have been reported for RNH^{HIV}, in which the attachment of an N-terminal His-tag [107, 108] and the insertion of a basic protrusion of EcoRNH1 [68, 71] do not restore its Mg^{2+}-dependent activity. Because the R29/D29 substrate binds to HIV-1 RT much more strongly than to RNH^{HIV}, TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV}, and RNA/DNA hybrid bound to HIV-1 RT is bended between the polymerase and RNase H domains [42], the polymerase domain is probably required for strong binding and bending of the substrate, which are responsible for Mg^{2+}-dependent RNase H activity of HIV-1 RT. This domain is also probably required to make the conformation of the RNase H active site of HIV-1 RT functional in the presence of Mg^{2+} ions.

It has been reported that HIV-1 RT cleaves a substrate for nonspecific RNase H assay either in the presence of Mg^{2+} ions or Mn^{2+} ions, but slightly more effectively in the presence of Mn^{2+} ions than in the presence of Mg^{2+} ions [68]. HIV-1 RT cleaves the $tRNA^{Lys3}$ substrate either in the presence of Mn^{2+} ions [68, 70] or Mg^{2+} ions [102]. The result that HIV-1 RT does not cleave the R29/D29 substrate in the presence of Mn^{2+} ions suggests that the preferable metal ion for the RNase H activity of HIV-1 RT is changed when the substrate is changed.

3.4. Summary

RNase H is present in the C-terminal portion of P66 subunit of HIV-1 RT. The isolated RNase H domain of HIV-1 RT (RNH^{HIV}) is nearly inactive, possibly due to the lack of substrate-binding affinity, disorder of a loop containing His539, and increased flexibility. To examine whether the activity of RNH^{HIV} is restored by the attachment of TmaHBD or BstNTD, the RNH^{HIV} derivatives with TmaHBD and BstNTD at the N-terminus were constructed and characterized. Both RNH^{HIV} derivatives bound to RNA/DNA hybrid more strongly than RNH^{HIV} and exhibited much higher Mn^{2+}-dependent RNase H activity than it. Both derivatives did not exhibit or exhibited a very weak Mg^{2+}-dependent RNase H activity. These results indicate that TmaHBD and BstNTD function as an RNA/DNA hybrid-binding tag, and greatly increase the substrate-binding affinity and Mn^{2+}-dependent RNase H activity of RNH^{HIV}. However, they do not restore the Mg^{2+}-dependent RNase H activity of RNH^{HIV}, indicating that enhancement of the substrate binding affinity of RNH^{HIV} is not enough to restore its Mg^{2+}-dependent RNase H activity.

CHAPTER 4

General discussion and future remarks

4.1. General discussion

RNase H is an endoribonuclease that cleaves the RNA moiety of RNA/DNA hybrids. Its activity is considered to be essential for maintaining genome integrity in the cell. It is involved in the DNA repair process, removal of R-loop, and removal of RNA from Okazaki fragment during DNA replication. Some RNases H have substrate binding domains. However, these domains are not always present in RNases H and vary in structures. Therefore, the studies on the substrate binding domains of RNases H are necessary to increase our knowledge on structure-function relationships of RNases H.

When I started this work, it was known that some RNases H1 have substrate binding domains, termed basic protrusion and HBD, and all RNases H3 have them, termed TBP-like domain. These domains have been shown to be important for substrate binding, activity, and/or stability. However, little was known on substrate binding domains of RNases H2. I showed for the first time that some bacterial RNases H2 have a substrate binding domain, termed NTD, at the N-terminus. Although, the structure of NTD remains to be determined, poor amino acid sequence similarity between NTD and HBD or TBP-like domain suggests that NTD has a unique structure. Identification of this novel substrate binding domain allows me to propose that RNases H are structurally and functionally diverged by acquiring various substrate binding domains, such as HBD, basic protrusion, TBP-like domain and NTD.

The role of C-extensions of RNases H2 was analyzed using TmaRNH2. According to its crystal structure in complex with dsDNA[R1], TmaRNH2 consists of the catalytic

domain and C-terminal domain, which contact the cleaved and non-cleaved strands of dsDNA^{R1} respectively [9]. The C-terminal domain consists of helices F-J, in which helices I and J correspond to C-extension of TmaRNH2 and are folded into a hairpin helix structure (TmaHH). Helix G and H make a direct contact with the substrate in the upstream region of the scissile phosphate group, whereas TmaHH is not involved in substrate binding (Fig. 2.3). Nevertheless, the removal of TmaHH decreased activity and stability of TmaRNH2. This result suggests that the removal of TmaHH causes a slight conformational change that may greatly decrease the substrate binding affinity of the C-terminal domain. Thus, C-extension of TmaRNH2 is probably necessary to make the conformational of the C-terminal domain functional.

The finding that AaeRNH2 is as active as and is more stable than BstRNH2 and TmaRNH2 suggests that AaeRNH2 does not require NTD or C-terminal hairpin helix structure to increase its activity and stability. However, the activities and stability of BstRNH2 and TmaRNH2 are greatly reduced by the removal of BstNTD and TmaHH respectively, indicating that the catalytic domains of these enzymes require NTD or C-terminal hairpin helix structure for maximal activity and stability. The 3D model of AaeRNH2 highly resembles those of the catalytic domains of BstRNH2 and TmaRNH2. However, AaeRNH2 may have additional interactions that increase its substrate binding affinity and stability, as compared to those of the catalytic domains of BstRNH2 and TmaRNH2. The catalytic domains of BstRNH2 and TmaRNH2 may lose these interactions, because BstNTD or TmaHH may compensate for the loss of these interactions. Comparison of the various stabilizing factors suggests that stabilization by ion pairs and aromatic interactions at least partly account for the difference in stability between AaeRNH2 and TmaRNH2 or BstRNH2.

The fact that the isolated RNase H domain of HIV-1 RT (RNH^{HIV}) is inactive raises a question whether the activity of this domain is restored by the attachment of N-terminal substrate binding domain of bacterial RNase H. I found that the removal of the DNA polymerase domain of P66 subunit and entire P51subunit of HIV-1 RT alters the interaction between its RNase H domain and the substrate. Thus, RNH^{HIV} cleaves RNA/DNA substrate at the different position from that cleaved by HIV-1 RT, but only in the presence of Mn^{2+} ions. Yet, the Mn^{2+}-dependent activity of RNH^{HIV} was very low. The attachment of TmaHBD or BstNTD at the N-terminus of RNH^{HIV} does not alter its interaction with the substrate, but greatly increases the Mn^{2+}-dependent activity of RNH^{HIV} by 500-fold and 13-fold, respectively. This results suggests that TmaHBD and BstNTD greatly increase the binding affinity RNH^{HIV} by stabilizing the interaction between RNH^{HIV} and the substrate. Thus, this is the first study to my knowledge showing that the N-terminal substrate binding domain of RNase H can be used as an RNA/DNA hybrid tag.

RNases H are diverged by acquiring substrate binding domain. We have found that the metal dependencies can be modified by acquiring substrate binding domain. Thus, the mode of Mn^{2+} or Mg^{2+}-dependent activity caused by acquiring substrate binding domain can also be a remarkable example of parallel evolution, in which organisms can differ from their common ancestor by their requirement for a divalent metal ion for enzymatic activity.

4.2. Future remarks

BstRNH2 and AaeRNH2 were biochemically characterized in this study. However, their substrate recognition mechanisms remain to be understood. TmaRNH1 has been biochemically characterized previously [76]. However, its substrate recognition mechanism also remain to be understood. Therefore, it will be necessary to determine the

crystal structures of BstRNH2, AaeRNH2 and TmaRNH1 in complex with the substrate to understand these mechanisms. To understand the mechanism by which the attachment of BstNTD and TmaHBD to the N-terminus of RNH_{HIV} increases the substrate binding affinity and Mn^{2+}-dependent activity of NH^{HIV}, it will also be necessary to determine the crystal structures of TmaHBD-RNH^{HIV} and BstNTD-RNH^{HIV} in complex with the substrate.

The RNase H activity of HIV-1 RT is regarded as a target for development of antiviral drugs [35, 36]. Initial screenings for the inhibitors have been made by using the intact protein of HIV-1 RT. However, because of the complexity in the structure and function of HIV-1 RT, it is not easy to identify the lead compounds that inhibit the RNase

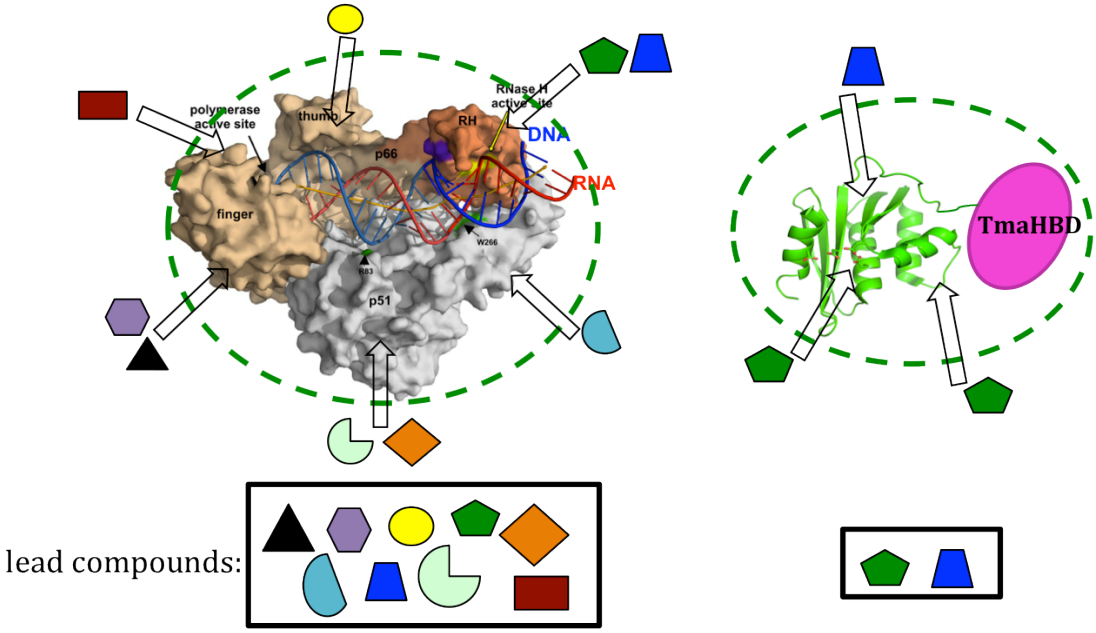


Fig. 4.1. Schematic representation of initial and second screenings for the inhibitors of RNase H activity of HIV-1 RT. Intact HIV-1 RT is usually used for an initial screening of RNase H inhibitors. However, because of the complexity of the structure of HIV-1 RT, these inhibitors may not directly interact with the RNase H domain of HIV-1 RT. By using TmaHBD-RNH^{HIV} with more simple structure for second screening, it may be possible to screen for the lead compounds that directly bind to the RNase H domain and specifically inhibit the RNase H activity.

H activity of HIV-1 RT by directly contacting its RNase H domain. It has been reported that p51-G-TCR, which lacks the polymerase domain of the P66 subunit but retains Mg^{2+}-dependent RNase H activity, is useful to identify these inhibitors in second screenings [102]. However, p51-G-TCR still contains the thumb and connection domains of the P66 subunit and the entire P51 subunit. Because TmaHBD-RNH^{HIV} contains only the RNase H domain of HIV-1 RT and exhibits higher RNase H activity than HIV-1 RT in the presence of Mn^{2+} ions, this protein may be useful to identify the inhibitors specific for the RNase H domain of HIV-1 RT in second screenings, as shown in Figure 4.1.

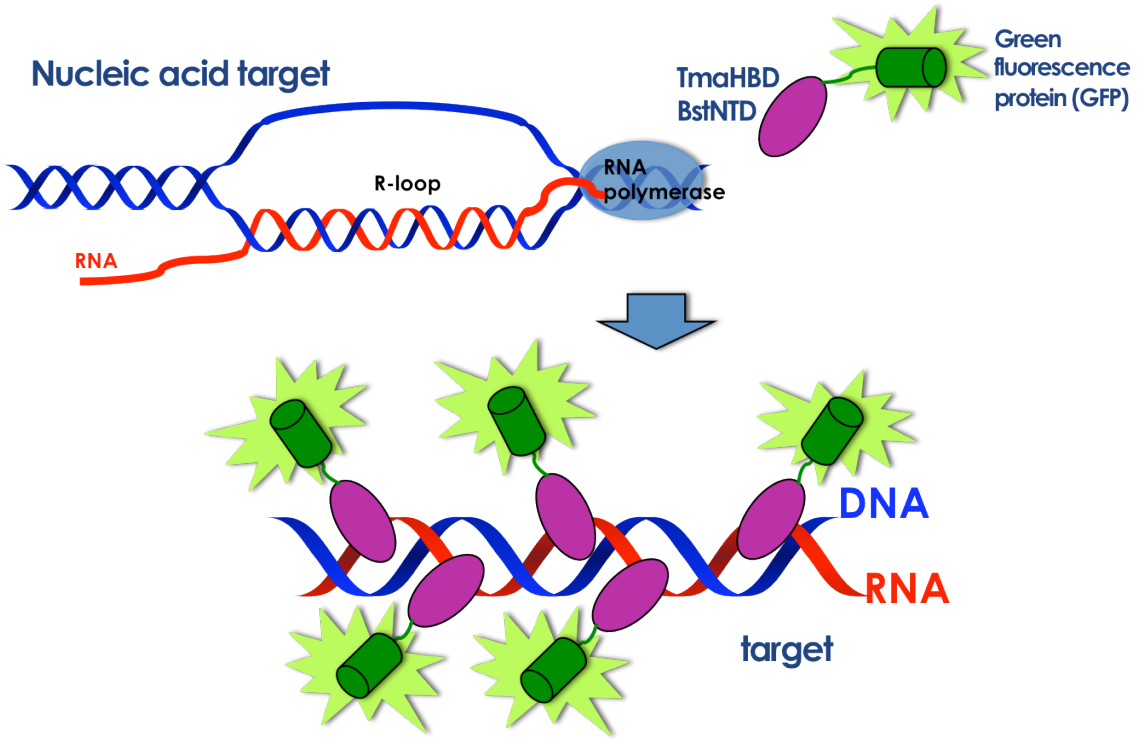


Fig. 4.2. Schematic representation of R-loop mapping in the genome. R-loop is formed during transcription when the newly synthesized RNA strand fails to dissociate from the DNA strand. This R-loop can be detected if the fusion protein, in which TmaHBD is attached to the N-terminus of a reported proteins (GFP as an example), is expressed in the cell, because this fusion protein specifically binds to the RNA/DNA hybrid region of R-loop.

RNA/DNA hybrid structure, termed R-loop, is thought to be a rare byproduct of transcription. The accumulation of R-loops in the genome threat the genome integrity and is a potential cause of disease. Recent evidence suggests that R-loop is also involved in molecular mechanisms of neurological disease and cancer [106]. Then, the detection of R-loops in the genome is probably necessary. Since TmaHBD and BstNTD can be used as an RNA/DNA hybrid tag, they may also be used for R-loop mapping in the genome. If the fusion protein, in which TmaHBD or BstNTD is attached to the N-terminus of a fluorescent reporter protein, is expressed in the cell, it is expected that this fusion protein specifically binds to targeted R-loop. Then, R-loop can be detected, as shown in Figure 4.2.

Furthermore, this study provides valuable information on protein engineering. Protein engineering is a technology to alter protein function. Because the enzymatic properties of BstRNH2 and TmaRNH2 are greatly changed by the removal of NTD and TmaHH, and the enzymatic properties of RNHHIV are greatly changed by the attachment of BstNTD or TmaHBD, domain engineering may be a promising protein engineering technology to generate enzymes with novel functions, which are useful for industrial purpose.

References

1. Crouch, R. J., & Dirksen, M. L. (1982). Ribonuclease H. *In Nuclease* (Linn, S. M., & Roberts, R. J., ed.), pp. 211-241, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. *FEBS Journal*, **276**, 1494-1505.
3. Ohtani, N., Haruki, M., Morikawa, M., & Kanaya, S. (1999). Molecular diversities of RNases H. *Journal of Bioscience and Bioengineering*, **88**, 12-19.
4. Tadokoro, T., & Kanaya, S. (2009). Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS Journal*, **276**, 1482-1493.
5. Kanaya, S., & Crouch, R. J. (1983). DNA sequence of the gene encoding for *Escherichia coli* ribonuclease H. *Journal of Biological Chemistry*, **258**, 1276-1281.
6. Itaya, M. (1990). Isolation and characterization of a second RNase H (RNase H II) of *Escherichia coli* K-12 encoded by the *rnh*B gene. *Proceeding of the National Academy of Science USA*, **87**, 8587-8591.
7. Ohtani, N., Haruki, M., Morikawa, M., Crouch, R. J., Itaya, M., & Kanaya, S. (1999). Identification of the genes encoding Mn^{2+}-dependent RNase HIII from *Bacillus subtilis*: classification of RNases H into three families. *Biochemistry*, **38**, 605-618.
8. 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 Research*, **32**, 407-414.

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

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

11. Rydberg, B., & Game, J. (2002). Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proceeding of the National Academy of Science USA*, **99**, 16654-16659.

12. 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 Letters*, **531**, 204-208.

13. 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. *Journal of Biological Chemistry*, **265**, 5561-5566.

14. Jongruja, N., You, D. J., Angkawidjaja, C., Kanaya, E., Koga, Y., & Kanaya, S. (2012). Structure and characterization of RNase H3 from *Aquifex aeolicus*. *FEBS Journal*, **279**, 2737-2753.

15. 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. *Molecular Microbiology*, **83**, 1080-1093.

16. Lu, Z., Hou, J., Wang, Y., & Liu, J. (2012). Involvement of Ser94 in RNase HIII from *Chlamydophila pneumonia* in the recognition of a single ribonucleotide

misincorporated into double-stranded DNA. *Biochimica et Biophysica Acta*, **1824**, 859-865.

17. Nick McElhinny, S. A., Watts, B. E., Kumar, D., Watt, D. L., Lundström, E. B., Burgers, P. M., Johansson, E., Chabes, A., & Kunkel, T. A. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proceeding of the National Academy of Science USA*, **107**, 4949-4954.

18. Nick McElhinny, S. A., Kumar, D., Clark, A. B., Watt, D. L., Watts, B. E., Lundström, E. B., Johansson, E., Chabes, A., & Kunkel, T. A. (2010). Genome instability due to ribonucleotide incorporation into DNA. *Nature Chemical Biology*, **6**, 774-781.

19. McDonald, J. P., Vaisman, A., Kuban, W., Goodman, M. F., & Woodgate, R. (2012). Mechanisms employed by *Escherichia coli* to prevent ribonucleotide incorporation into genomic DNA by Pol V. *PLoS Genetics*, **8**, e1003030.

20. 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. *Molecular Cell*, **47**, 980-986.

21. 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. *Molecular Cell*, **45**, 99-110.

22. Reijns, M. A., Rabe, B., Rigby, R. E., Mill, P., Astell, K. R., Lettice, L. A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., Devenney, P. S., Sexton, D., Grimes, G., Holt, I. J., Hill, R. E., Taylor, M. S., Lawson, K. A., Dorin, J. R., & Jackson, A.

P. (2012). Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*, **149**, 1008-1022.

23. Hiller, B., Achleitner, M., Glage, S., Naumann, R., Behrendt, R., & Roers, A. (2012). Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *Journal of Experimental Medicine*, **209**, 1419-1426.

24. Kogoma, T., & Foster, P. L. (1998). Physiological functions of *E. coli* RNase HI. In *Ribonucleases H* (Crouch RJ & Toulme JJ eds), pp. 39-66, INSERM, Paris.

25. Drolet, M., Phoenix, P., Menzel, R., Massé, E., Liu, L. F., & Crouch, R. J. (1995). Overexpression of RNase H partially complements the growth defect of an Eschericia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proceeding of the National Academy of Science USA*, **92**, 3526-3530.

26. 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. *Molecular Cell*, **11**, 807-815.

27. 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. *Molecular and Cellular Biology*, **19**, 8361-8371.

28. 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. *Journal of Bacteriology*, **181**, 2118-2123.

29. 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.

30. Crow, Y. J., Leitch, A., Hayward, B. E., Garner, A., Parmar, R., *et. al*., & Jackson, A. P. (2006). Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. *Nature Genetics*, **38**, 910-916.

31. Reijns, M. A., & Jackson, A. P. (2014). Ribonuclease H2 in health and disease. *Biochemical Society Transactions*, **42**, 717-725.

32. Tisdale, M., Schulze, T., Larder, B. A., & Moelling, K. (1991). Mutations within the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase abolish virus infectivity. *Journal of General Virology*, **72**, 59-66.

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

34. Wahba, L., Amon, J. D., Koshland, D., & Vuicca-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Molecular Cell*, **44**, 978-988.

35. Corona, A., Di Leva, F. S., Thierry, S., Pescatori, L., Cuzzucoli Crucitti, G., Subra, F., Delelis, O., Esposito, F., Rigogliuso, G., Costi, R., Cosconati, S., Novellino, E., Di Santo, R., & Tramontano, E. (2014). Identification of Highly Conserved Residues Involved in Inhibition of HIV-1 RNase H Function by Diketo Acid Derivatives. *Antimicrobial Agents Chemotherapy*, **58**, 6101-6110.

36. Le Grice, S. F. (2012). Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. *Journal of Biological Chemistry*, **287**, 40850-40857.

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

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

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

40. Lim, D., Gregorio, G. G., Bingman, C., Martinez-Hackert, E., Hendrickson, W. A., Goff, S. P. (2006). Crystal structure of the moloney murine leukemia virus RNase H domain. *Journal of Virology*, **80**, 8379-8389.

41. 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.

42. 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. *Molecular Cell*, **28**, 264-276.

43. Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H., & Morikawa, K. (1993). Crystal structure of ribonuclease HI from *Thermus*

thermophilus HB8 refined at 2.8 Å resolution. *Journal of Molecular Biology*, **230**, 529-542.

44. Tadokoro, T., You, D-J., Abe, Y., Chon, H., Matsumura, H., Koga, Y., Takano, K., & Kanaya, S. (2007). Structural, thermodynamic, and mutational analyses of a psychrotrophic RNase HI. *Biochemistry*, **46**, 7460-7468.

45. 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 Arginine 118 and C-terminal anchoring. *Biochemistry*, **46**, 11494-11503.

46. 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 Letters*, **587**, 1418-1423.

47. 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 Science*, **21**, 553-561.

48. 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. *Journal of Structural Biology*, **182**, 144-154.

49. You, D.-J., Jongruja, N., Tannous, E., Angkawidjaja, C., Koga, Y., & Kanaya, S. (2014). Structural basis for salt-dependent folding of ribonuclease H1 from halophilic archaeon *Halobacterium* sp. NRC-1. *Journal of Structural Biology*, **187**, 119-128.

50. 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 Science*, **10**, 707-714.

51. Chapados, B. R., Chai, Q., Hosfield, D. J., Shen, B., & Tainer, J. A. (2001). Structural biochemistry of a type 2 RNase H: RNA primer recognition and removal during DNA replication. *Journal of Molecular Biology*, **307**, 541-556.

52. Lai, L., Yokota, H., Hung, L. W., Kim, R., and Kim, S. H. (2000). Crystal structure of archaeal RNase HII: a homologue of human major RNase H. *Structure*, **8**, 897-904.

53. 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. *Journal of Molecular Biology*, **356**, 165-178.

54. Figiel, M., & Nowotny, M. (2014). Crystal structure of RNase H3-substrate complex reveals parallel evolution of RNA/DNA hybrid recognition. *Nucleic Acid Research*, **42**, 9285-9294.

55. 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. *Journal of Applied Microbiology*, **109**, 974-983.

56. Oda, Y., Yoshida, M., & Kanaya, S. (1993). Role of histidine 124 in the catalytic function of ribonuclease HI from *Escherichia coli*. *Journal of Biological Chemistry*, **268**, 88-92.

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

58. Ho, M. H., De Vivo, M., Peraro, M. D., & Klein, M. L. (2010). Understanding the effect of magnesium ion concentration on the catalytic activity of ribonuclease H through computation: does a third metal binding site modulate endonuclease catalysis?. *Journal of the American Chemical Society*, **132**, 13702-13712.

59. Katayanagi, K., Okumura, M., & Morikawa, K. (1993). Crystal structure of *Escherichia coli* RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for a single Mg^{2+}-binding site. *Proteins*, **17**, 337-346.

60. Goedken, E. R., & Marqusee, S. (2001). Co-crystal of *Escherichia coli* RNase HI with Mn^{2+} ions reveals two divalent metals bound in the active site. *Journal of Biological Chemistry*, **276**, 7266-7271.

61. Tsunaka, Y., Takano, K., Matsumura, H., Yamagata, Y., & Kanaya, S. (2005). Identification of single Mn^{2+} binding sites in the mutant proteins of *E. coli* RNase HI at Glu^{48} and/or Asp^{134} by X-ray crystallography. *Journal of Molecular Biology*, **345**, 1171-1183.

62. 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 structure. *Journal of Biological Chemistry*, **248**, 5914-5921.

63. Ohtani, N., Haruki, M., Muroya, A., Morikawa, M., & Kanaya, S. (2000). Characterization of ribonuclease HII from *Escherichia coli* overproduced in a soluble form. *Journal of Biochemistry*, **127**, 895-899.

64. Rohman, M. S., Koga, Y., Takano, K., Chon, H., Crouch, R. J., & Kanaya, S. (2008). Effect of the disease-causing mutations identified in human ribonuclease (RNase) H2 on the activities and stabilities of yeast RNase H2 and archaeal RNase HII. *FEBS Journal*, **275**, 4836-4849.

65. Medicis, E. D., Paquette, J., Gauthier, J. J., & Shapcott, D. (1986). Magnesium and manganese content of halophilic bacteria. *Applied and Environmental Microbiology*, **52**, 567-573.

66. Tsunaka, Y., Haruki, M., Morikawa, M., Oobatake, M., & Kanaya, S. (2003). Dispensability of Glu^{48} and Asp^{134} for Mn^{2+}-dependent activity of *E. coli* ribonuclease HI. *Biochemistry*, **42**, 3366-3374.

67. Keck, J. L., & Marqusee, S. (1996). The putative substrate recognition loop of *Escherichia coli* ribonuclease H is not essential for activity. *Journal of Biological Chemistry*, **271**, 19883-19887.

68. Stahl, S. J., Kaufman, J. D., Vikić-Topić, S., Crouch, R. J., & Wingfield, P. T. (1994). Construction of an enzymatically active ribonuclease H domain of human immunodeficiency virus type 1 reverse transcriptase. *Protein Engineering*, **7**, 1103-1108.

69. Hostomsky, Z., Hostomska, Z., Hudson, G. O., Moomaw, E. W., & Nodes, B. R. (1991). Reconstitution in vitro of RNase H activity by using purified N-terminal and C-terminal domains of human immunodeficiency virus type 1 reverse transcriptase. *Proceeding of the National Academy of Science USA*, **88**, 1148-1152.

70. Smith, J. S., Gritsman, K., & Roth, M. J. (1994). Contribution of DNA polymerase subdomains to the RNase H activity of human immunodeficiency virus type 1 reverse transcriptase. *Journal of Virology*, **68**, 5721-5729.

71. Keck, J. L., & Marqusee, S. (1995). Substitution of a highly basic helix/loop sequence into the RNase H domain of human immunodeficiency virus reverse transcriptase restored its Mn^{2+}-dependent RNase H activity. *Proceeding of the National Academy of Science USA*, **92**, 2740-2744.

72. Hang, J. Q., Rajendran, S., Yang, Y., Li, Y., In, P. W., Overton, H., Parkes, K. E., Cammack, N., Martin, J. A., & Klumpp, K. (2004). Activity of the isolated HIV RNase H domain and specific inhibition by N-hydroxyimides. *Biochemical and Biophysical Research and Communications*, **317**, 321-329.

73. Iwai, S., Wakasa, M., Ohtsuka, E., Kanaya, S., Kidera, A., & Nakamura, H. (1996). Interaction of the basic protrusion of *Escherichia coli* RNase H1 with its substrate. *Journal of Molecular Biology*, **263**, 699-706.

74. 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. *Journal of Biological Chemistry*, **272**, 22015-22022.

75. 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. *Journal of Biological Chemistry*, **266**, 11621-11627.

76. 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 Journal*, **277**, 4474-4489.

77. Tadokoro, T., Chon, H., Koga, Y., Takano, K., & Kanaya, S. (2007). Identification of the gene encoding a type 1 RNase H with an N-terminal double-stranded RNA binding domain from a psychrotropic bacterium. *FEBS Journal*, **274**, 3715-3727.

78. Nowotny, M., Cerritelli, S. M., Ghirlando, R., Gaidamakov, S. A., Crouch, R. J., & Yang, W. (2008). Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. *EMBO Journal*, **27**, 1172-1181.

79. Gaidamakov, S. A., Gorshkova, I. I., Schuck, P., Steinbach, P. J., Yamada, H., Crouch, R. J., & Cerritelli, S. M. (2005). Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. *Nucleic Acids Research*, **33**, 2166-2175.

80. Goedken, E. R., Raschke, T. M., & Marqusee, S. (1997). Importance of the C-terminal helix to the stability and enzymatic activity of *Escherichia coli* ribonuclease H. *Biochemistry*, **36**, 7256-7263.

81. Cirino, N. M., Cameron, C. E., Smith, J. S., Rausch, J. W., Roth, M. J., Benkovic, S. J., & Le Grice, S. F. (1995). Divalent cation modulation of the ribonuclease functions of human immunodeficiency virus reverse transcriptase. *Biochemistry* **34**, 9936-9943.

82. Miyashita, S., Tadokoro, T., Angkawidjaja, C., You, D-J., Koga, Y., Takano, K., & Kanaya, S. (2011). Identification of the substrate binding site in the N-terminal TBP-like domain of RNase H3. *FEBS Letters*, **585**, 2313-2317.

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

84. Goodwin, T. W., & Morton, R. A. (1946). The spectrophotometric determination of tyrosine and tryptophan in proteins. *Biochemical Journal*, **40**, 628-632.

85. Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, **31**, 3381-3385.

86. Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., & Swanson, R. V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*, **392**, 353-358.

87. Huber, R., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B., & Stetter, K. O. (1986). *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Archives of Microbiology*, **144**, 324-333.

88. Imanaka, T., Fujii, M., Aramori, I., & Aiba, S. (1982). Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. *Journal of Bacteriology*, **149**, 824-830.

89. Muroya, A., Nakano, R., Ohtani, N., Haruki, M., Morikawa, M., & Kanaya, S. (2002). Importance of an N-terminal extension in ribonuclease HII from *Bacillus stearothermophilus* for substrate binding. *Journal of Bioscience and Bioengineering*, **93**, 170-175.

90. Hashimoto, H., Inoue, T., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., & Kai, Y. (1999). Hyperthermostable protein structure maintained by intra and inter-helix ion pairs in archaeal O^6-methylguanine-DNA methyltransferase. *Journal of Molecular Biology*, **292**, 707-716.

91. Karshikoff, A., & Ladenstein, R. (2001). Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends in Biochemical Science*, **26**, 550-556.

92. Uehara, R., Takeuchi, Y., Tanaka, S., Takano, K., Koga, Y., & Kanaya, S. (2012). Requirement of Ca^{2+} ions for the hyperthermostability of Tk-subtilisin from *Thermococcus kodakarensis*. *Biochemistry*, **51**, 5369–5378.

93. 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. *Journal of Molecular Biology*, **408**, 514-528.

94. Thompson, M. J., & Eisenberg, D. (1999). Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *Journal of Molecular Biology*, **290**, 595-604.

95. Thoma, R., Hennig, M., Sterner, R., & Kirschner, K. (2000). Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Structure*, **8**, 265-276.

96. Watanabe, K., Chishiro, K., Kitamura, K., & Suzuki, Y. (1991). Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6-glucosidase from *Bacillus thermoglucosidasius* KP1006. *Journal of Biological Chemistry*, **266**, 24287-24294.

97. Boutz, D. R., Cascio, D., Whitelegge, J., Perry, L. J., & Yeates, T. O. (2007). Discovery of a thermophilic protein complex stabilized by topologically interlinked chains. *Journal of Molecular Biology*, **368**, 1332-1344.

98. 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 Research*, **34**, W116-118 (Web Server issue).

99. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., & Steitz, T. A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, **256**, 1783-1790.

100. Bohlayer, W. P., & DeStefano, J. J. (2006). Tighter binding of HIV reverse transcriptase to RNA-DNA versus DNA-DNA results mostly from interactions in the polymerase domain and requires just a small stretch of RNA-DNA. *Biochemistry*, **45**, 7628-7638.

101. Zuniga, R., Sengupta, S., Snyder, C., Leon, O., & Roth, M. J. (2004). Expression of the C-terminus of HIV-1 reverse transcriptase p66 and p51 subunits as a single polypeptide with RNase H activity. *Protein Engineering Design and Selection*, **17**, 581-587.

102. Farias, R. V., Vargas, D. A., Castillo, A. E., Valenzuela, B., Cote, M. L., Roth, M. J., & Leon, O. (2011). Expression of an Mg^{2+}-dependent HIV-1 RNase H construct for drug screening. *Antimicrobial Agents and Chemotherapy*, **55**, 4735-4741.

103. Konishi, A., Shinomura, M., & Yasukawa, K. (2013). Enzymatic characterization of human immunodeficiency virus type 1 reverse transcriptase for use in cDNA synthesis. *Applied Biochemistry and Biotechnology*, **169**, 77-87.

104. Pace, C. N. (1990). Measuring and increasing protein stability. *Trends in Biotechnology*, **8**, 93-98.

105. Ben-Artzi, H., Zeelon, E., Gorecki, M., & Panet, A. (1992). Double-stranded RNA-dependent RNase activity associated with human immunodeficiency virus type 1 reverse transcriptase. *Proceeding of the National Academy of Science USA*, **89**, 927-931.

106. Groh, M., & Gromak, N. (2014). Out of balance: R-loops in human disease. *PLOS Genetics*, **10**, e1004630.

107. Smith, J. S., & Roth, M. J. (1993). Purification and characterization of an active human immunodeficiency virus type 1 RNase H domain. *Journal of Virology*, **67**, 4037-4049.

108. Evans, D. B., Fan, N., Swaney, S. M., Tarpley, W. G., & Sharma, S. K. (1994). An active recombinant p15 RNase H domain is functionally distinct from the RNase H domain associated with human immunodeficiency virus type 1 reverse transcriptase. *Journal of Biological Chemistry*, **269**, 21741-21747.

109. Christen, M. T., Menon, L., Myshakina, N.S., Ahn, J., Parniak, M. A., & Ishima, R. (2012). Structural basis of the allosteric inhibitor interaction on the HIV-1 reverse transcriptase RNase H domain. *Chemical Biology & Drug Design*, **80**, 706-716.

List of publications

1. Permanasari, E. D., Angkawidjaja, C., Koga, Y., & Kanaya, S. (2013). Role of N-terminal extension of *Bacillus stearothermophilus* RNase H2 and C-terminal extension of *Thermotoga maritima* RNase H2. *FEBS Journal*, **280**, 5065-5079.
2. Permanasari, E. D., Yasukawa, Y., & Kanaya, S. (2015). Enzymatic activities of RNase H domains of HIV-1 reverse transcriptase with substrate binding domains of bacterial RNases H1 and H2. *Molecular Biotechnology*, **57**, 526-538.

Acknowledgements

First of all, I would like to express my sincere gratitude and appreciation for all the people who have helped, inspired me, and made this doctor thesis possible.

My most sincere gratitude is extended to my supervisor, Prof. Dr. Shigenori Kanaya, for his kindly consideration in accepting me as a student in the lab five years ago, and the immeasurable amount of support throughout this study. His endless energy and enthusiasm in the research extremely motivated me during my study. His valuable suggestion and constructive critisims become one of challenge during completion of this study. He provided me with the encouragement and improved my ability in research writing. I greatly benefit for his help and guidance during completion of my research topic and manuscripts. Again, I truly convey my heartfelt gratitude to “Kanaya Sensei” of every incredible opportunity during my stay as a graduate student in his lab for these five years. This thesis would not have been possible without his extraordinary support.

I would like to express my special gratitude to Prof. Dr. Kiichi Fukui and Prof. Dr. Kazuhito Fujiyama for their valuable comments, advices, and suggestions.

I am deeply grateful to the Japanese Ministry Education, Culture, Sports, Science and Technology for the financial support in completing the study, and to the International Program of Frontier Biotechnology for providing the opportunity of 5 years Master and Doctoral program in Osaka University.

I express my sincere gratitude to Assoc. Prof. Yuichi Koga, Assist. Prof. Clement Angkawidjaja, Dr. Dong-Ju Yu, and Dr. Eiko Kanaya for their supports, helps, suggestions, discussions, and creative criticisms.

My special gratitude to Ms. Reiko Matsumoto for her warm concern and encouragement during my life in Japan, particularly in the lab. Thank you for our time together out side the lab. She has been like “a mother” for all international students in Kanaya Lab. I am very grateful for her careness and support.

I offer my special thanks to my senior colleagues Dr. Nujarin Jongruja and Dr. Nguyen Tri Nhan for their support, help, discussion, and assistance in the experiment during my first year in the lab. My special thanks to Dr. Saifur Rohman and Dr. Cahyo Budiman for their valuable discussion, share, and wisdom advises during completion of my study. Thanks a lot Pak Saifur and Pak Cahyo. Also my special thank is to Dr. Sintawee Sulaiman for our endless friendship during my stay in Kanaya Lab. Thank you “Aoi”.

To Indah, my best friend in Indonesia, thank you for your careness and time, and cheering me up whenever I need some supports. I wish to thank everybody with whom I have shared experiences in life and all my Indonesian friends who inspired me to complete the study. I am tempted to individually thank all of my friends. I will simply say thank you very much to you all.

Last but not least, my very special thank is to my husband and family. To my husband, Andre Mahindra, thank you for your support, blessing, strength, and everything. Thank you Ayah for accompanying me during my life study in Japan. I cannot wait to start another life chapter of us. I thank to my family especially my parents (especially my Mom), my little brother, and sisters. They have always supported and encouraged me to do my best in all matters of life. To them I dedicate this thesis. Above all I thank God Almighty for the blessings and confidence shower on me to reach the goal.

Etin Diah Permanasari