# Crystal structure of archaeal RNase HII: a homologue of human major RNase H

Luhua Lai[1,2], Hisao Yokota[3], Li-Wei Hung[3], Rosalind Kim[3] and Sung-Hou Kim[1,3]*

**Background:** RNases H are present in all organisms and cleave RNAs in RNA/DNA hybrids. There are two major types of RNases H that have little similarity in sequence, size and specificity. The structure of RNase HI, the smaller enzyme and most abundant in bacteria, has been extensively studied. However, no structural information is available for the larger RNase H, which is most abundant in eukaryotes and archaea. Mammalian RNase H participates in DNA replication, removal of the Okazaki fragments and possibly DNA repair.

**Results:** The crystal structure of RNase HII from the hypothermophile *Methanococcus jannaschii*, which is homologous to mammalian RNase H, was solved using a multiwavelength anomalous dispersion (MAD) phasing method at 2 Å resolution. The structure contains two compact domains. Despite the absence of sequence similarity, the large N-terminal domain shares a similar fold with the RNase HI of bacteria. The active site of RNase HII contains three aspartates: Asp7, Asp112 and Asp149. The nucleotide-binding site is located in the cleft between the N-terminal and C-terminal domains.

**Conclusions:** Despite a lack of any detectable similarity in primary structure, RNase HII shares a similar structural domain with RNase HI, suggesting that the two classes of RNases H have a common catalytic mechanism and possibly a common evolutionary origin. The involvement of the unique C-terminal domain in substrate recognition explains the different reaction specificity observed between the two classes of RNase H.

Addresses: [1]Department of Chemistry, University of California, Berkeley, CA 94720, USA, [2]College of Chemistry and Institute of Physical Chemistry, Peking University, Beijing 100871, China and [3]Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA.

*Corresponding author.
E-mail: SHKIM@Cchem.Berkeley.edu

**Key words:** archaeal RNase H, DNA replication, removal of Okazaki fragment, RNase H

Received: **22 February 2000**
Revisions requested: **24 March 2000**
Revisions received: **10 April 2000**
Accepted: **16 June 2000**

Published: **28 July 2000**

**Structure** 2000, **8**:897–904

0969-2126/00/$ – see front matter


## Introduction

RNases H are enzymes that specifically degrade the RNA moiety in RNA/DNA hybrids. These enzymes are present in bacteria, archaea, eukarya and retroviruses [1]. Two major types of RNases H of different sizes, levels of activities, and specificities have been found. The genes encoding both types have been cloned from various organisms and there is little sequence similarity between the two types of enzyme [2]. In prokaryotes, RNases H with major activity are termed RNases HI and those with minor activities as RNases HII. Bacteria and retrovirus RNases HI have a molecular weight of about 20 kDa with high sequence homology and their structures and functions have been extensively studied [3–15]. Less is known about RNases HII, which are more abundant in archaea and eukarya and show little sequence similarity to RNases HI. RNase HII in *Escherichia coli* was first discovered in 1990 [16] and those from other organisms were subsequently identified [17,18]. No information on the secondary or tertiary structures and the structure–function relationship of RNase HII is known.

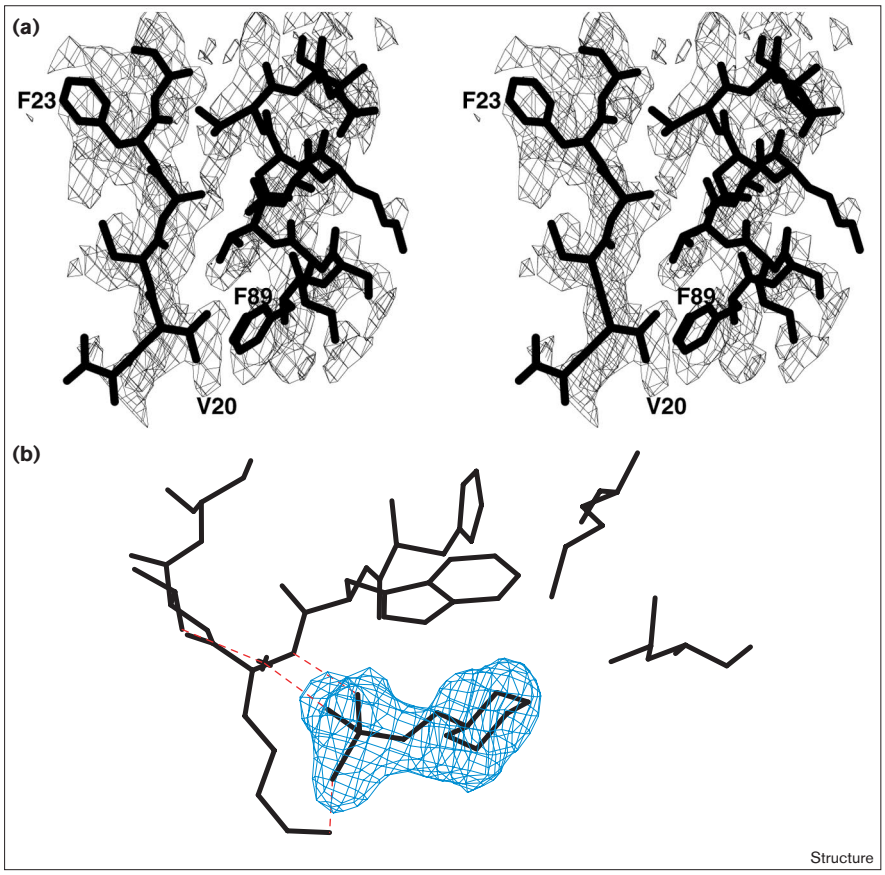
Mammalian RNases H participate in DNA replication, the removal of the Okazaki fragments during the synthesis of the lagging strand at the DNA replication fork (together with flap endonuclease 1), and possibly DNA repair [19]. It was recently found that the large subunit of the major human RNase H and the major RNase H from the yeast *Saccharomyces cerevisiae* are homologous to the prokaryotic RNases HII [1,20], a minor component in prokaryotes. On the basis of sequence homology, here we refer to RNase HI from prokaryotes as the small RNase H, or class I RNase H. The RNase HII from prokaryotes and archaea, and RNase HI from mammalian and yeast (eukaryotic) are referred to as the large RNase H, or class II RNase H. In order to elucidate the structural basis for its biological function, we have cloned, expressed, purified and carried out crystallographic studies on the class II RNase H from *Methanococcus jannaschii* (Mj RNase HII). We have chosen Mj RNase HII because of its similarity to mammalian major RNase H and to bacterial RNase HII.

## Results and discussion

### Overall structure

The structure of Mj RNase HII was determined at 2 Å resolution using multiwavelength anomalous diffraction (MAD) with selenomethionine (SeMet)-substituted protein. The MAD-phased electron density was easily interpretable in the central part of the molecule (Figure 1a).

**Figure 1**




**(a)** A sample of the MAD-phased electron-density map of Mj RNase HII at 2.0 Å resolution. The map is contoured at 1.6σ, with the final model displayed for comparison. **(b)** The simulated annealing omit map ($F_o$–$F_c$) of MES [38]. The electron-density map is contoured at 3.5σ. The MES sulfonate group is hydrogen bonded to the sidechain amine of Lys213, the mainchain amide of Lys213, and the sidechain to Thr214 mediated by water 403.

A crystallographic dimer is present in one asymmetrical unit with the molecules related by a pseudo-twofold symmetry (Figure 2a). Each Mj RNase HII monomer forms a compact structure consisting of two domains (N-terminal domain, residues 1–177; C-terminal domain, residues 178–225) with overall dimensions of 66 Å × 35 Å × 35 Å (Figure 3a). The N-terminal domain contains a central five-stranded mixed β sheet (three antiparallel and two parallel strands) and seven surrounding helices (Figure 3a), with dimensions 40 Å × 36 Å × 18 Å. The long helix 7 (153–177) extends from the N-terminal domain into the C-terminal domain. The C-terminal domain (178–230) has dimensions 22 Å × 16 × 9 Å and consists mainly of two parallel intersecting helices (with an interhelical angle of ~30°) and a connecting loop (Figure 3a). If the direction of the first strand is taken as the main axis of the N-terminal domain and the average of the two helices is taken as the axis of the C-terminal domain, the two domains intersect with an angle of 135° forming an open cleft. Sequence alignments with other large RNases H and the secondary structure assignments are shown in Figure 4.

The two domains are closely packed with a hydrophobic cluster between them. The hydrophobic cluster is formed between the two domains by residues Val14, Leu15, Tyr177, Phe193, Leu194 and Ile207. In addition, hydrogen bonds are formed between the sidechain of Arg209 and the carbonyl oxygen of Gly10, and between the sidechains of Thr210 and Asn73.

The two monomer molecules in one asymmetric unit are related by a twofold pseudo-symmetry axis. The dimeric interface includes hydrophobic interactions (between Leu55 and Tyr51 from either of the two monomers) and several hydrogen bonds. The buried surface is around 1900 Å$^2$. The carbonyl oxygen of Asn47 in molecule A hydrogen bonds to the amide group of Asn58 in molecule B. The sidechain amide of Asn47 in molecule A hydrogen bonds to the carbonyl oxygen of Lys54 in molecule B. The phenol oxygen of Tyr51 in molecule A hydrogen bonds to the carboxyl oxygen of Glu32 in molecule B.

One molecule of 2-(N-morpholino)ethanesulfonic acid (MES) is bound to each protein molecule. The density of the MES molecule is clear in a simulated annealing OMIT map (Figure 1b). The MES molecule lies in the cleft between the two domains and is close to helices 3 and 4, the loop between the first two β strands, helix 9,

and the loop between helices 8 and 9 (Figure 2a). The morpholino ring is surrounded by hydrophobic residues, Ile72, Met76, Leu81 and Trp212. The sulfonate group is hydrogenbonded to the sidechain of Lys213, the main-chain N of Lys213, and the sidechain of Thr214 mediated by a water molecule (Figure 1b).

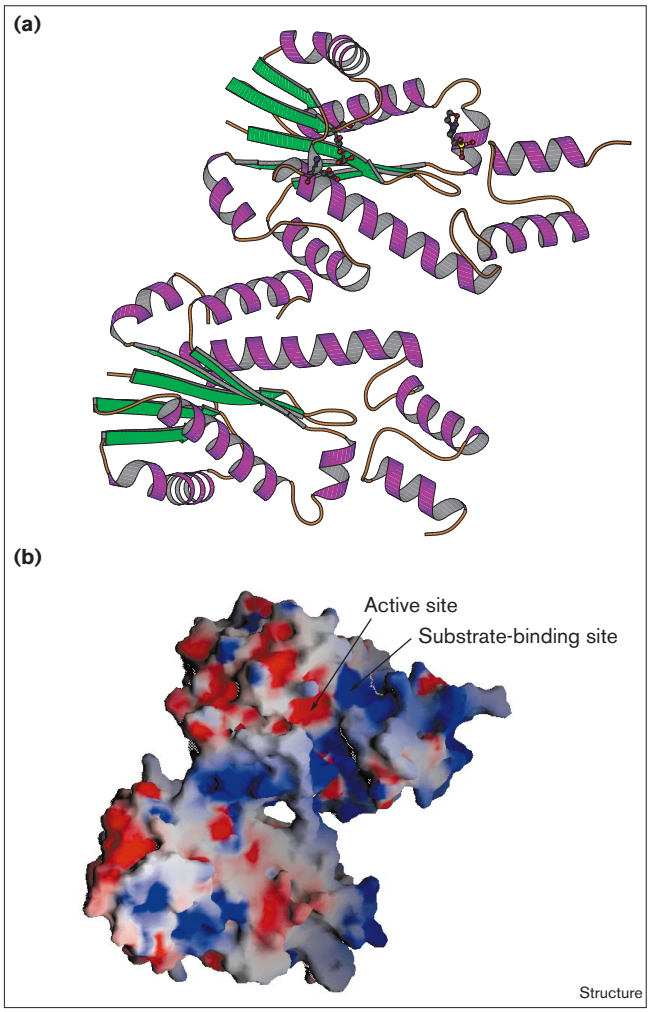
### Comparison of the N-terminal domain of Mj RNase HII with the small *E. coli* RNase H and other protein structures

Although there is no detectable sequence similarity between Mj RNase HII and the small RNase H of *E. coli*, the major domain (N-terminal) of Mj RNase HII shares a similar fold with *E. coli* RNase H, known as the RNase H fold (Figure 5a). The RNase H fold is shared by a group of nucleic-acid-interacting proteins [21,22] such as *E. coli* and HIV RNase H [3–5], retroviral (HIV-1 and ASV) integrase [23,24], Mu transposase [25] and *E. coli* RuvC resolvase [26]. The previously mentioned RNase H fold consists of a central mixed β sheet with the first three β strands connected directly and antiparallel to each other [21] (Figure 5a). However, in Mj RNase HII two helices containing 33 residues are inserted between the second and the third β strands. This is the first example of an RNase H fold that has two extra helices inserted between the second and the third β strands. When the central β sheets were superimposed upon each other, the root mean square deviation (rmsd) of 52 α carbon atoms in Mj RNase HII and *E. coli* RNase H (Protein Data Bank [PDB] code 2RN2 [27]) was 2.21 Å. Only two helices could be superimposed: helix 4 and helix 7 in Mj RNAse HII with helix I and helix V in *E. coli* RNAse H (Figure 5b). When Mj RNase HII was compared to the ASV-1 integrase (PDB code 1VSD [23,24]), the central β sheet aligned well, with an rmsd of 1.86 Å (44 α carbon atoms). Helices 4 and 5 in Mj RNase HII match well with two of the helices in ASV-1 integrase. Helix 7 in Mj RNAse HII is slightly displaced (about 10°) with the corresponding helix in ASV-1 integrase (Figure 5c).

### Active site of Mj RNase HII

Like the small RNase H, Mj RNase HII was also found to be active towards DNA/RNA hybrids (S. Kanaya, personal communication). The specific activity of Mj RNase HII was 0.005 units $mg^{-1}$ (30°C, pH 8.0, 10 mM $MgCl_2$), which is 50 times more active than *E. coli* RNase HII under the same conditions. These observations suggest that *E. coli* RNase HI and RNase HII are functionally and structurally related despite the absence of any significant sequence similarities. However, these two types of enzymes have different substrate specificity; for example, RNase HII from *Pyrococcus kadakaraensis* and *E. coli* can cleave DNA-RNA-DNA/DNA substrate (a duplex formed by a single-strand DNA and a single-strand chimeric DNA-RNA-DNA) at the 5′ end of the last ribonucleotide at the RNA–DNA junction, thereby facilitating the elimination of RNA from an Okazaki substrate [18]. All the large RNases H (*E. coli* RNase HII, *P. kadakaraensis* RNase HII and human RNase HI fragment) were shown to have RNA–DNA junction activity when mixed with chimeric DNA-RNA-DNA/DNA duplex, whereas the small *E. coli* RNase H does not [18].

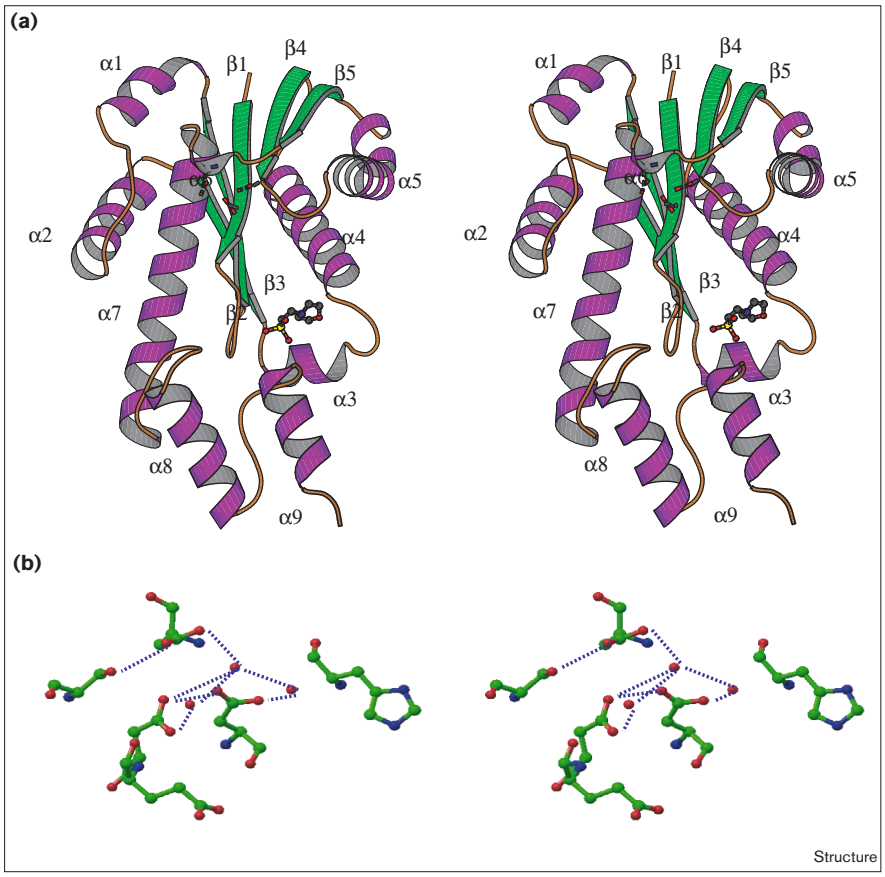
**Figure 2**




**(a)** The orientation of the Mj RNase HII dimer is shown by a secondary structure diagram with the putative catalytic residues and MES shown for molecule B. The putative catalytic residues and MES in molecule A are on the opposite side of the molecule and omitted for clarity. **(b)** The electrostatic molecular surface of the Mj RNase HII dimer without MES. The surface was calculated using GRASP [28] and colored from red to blue in the range −10.9 to 10.0 $K_BT$, where $K_B$ is the Boltzmann constant and $T$ is absolute temperature. For molecule B, the active-site residues form a red negative patch located on one edge of the positive substrate-binding site. The active site and substrate-binding site of molecule A cannot be seen in this orientation.

Because the structure of Mj RNase HII contains an RNase H fold, its active-site geometry may be similar to other small RNases H. When comparing the structures of HIV integrase, RuvC and the small *E. coli* RNase H, Yang and Steitz [21] pointed out that these three enzymes have remarkably similar active sites that consist of a set of three

**Figure 3**



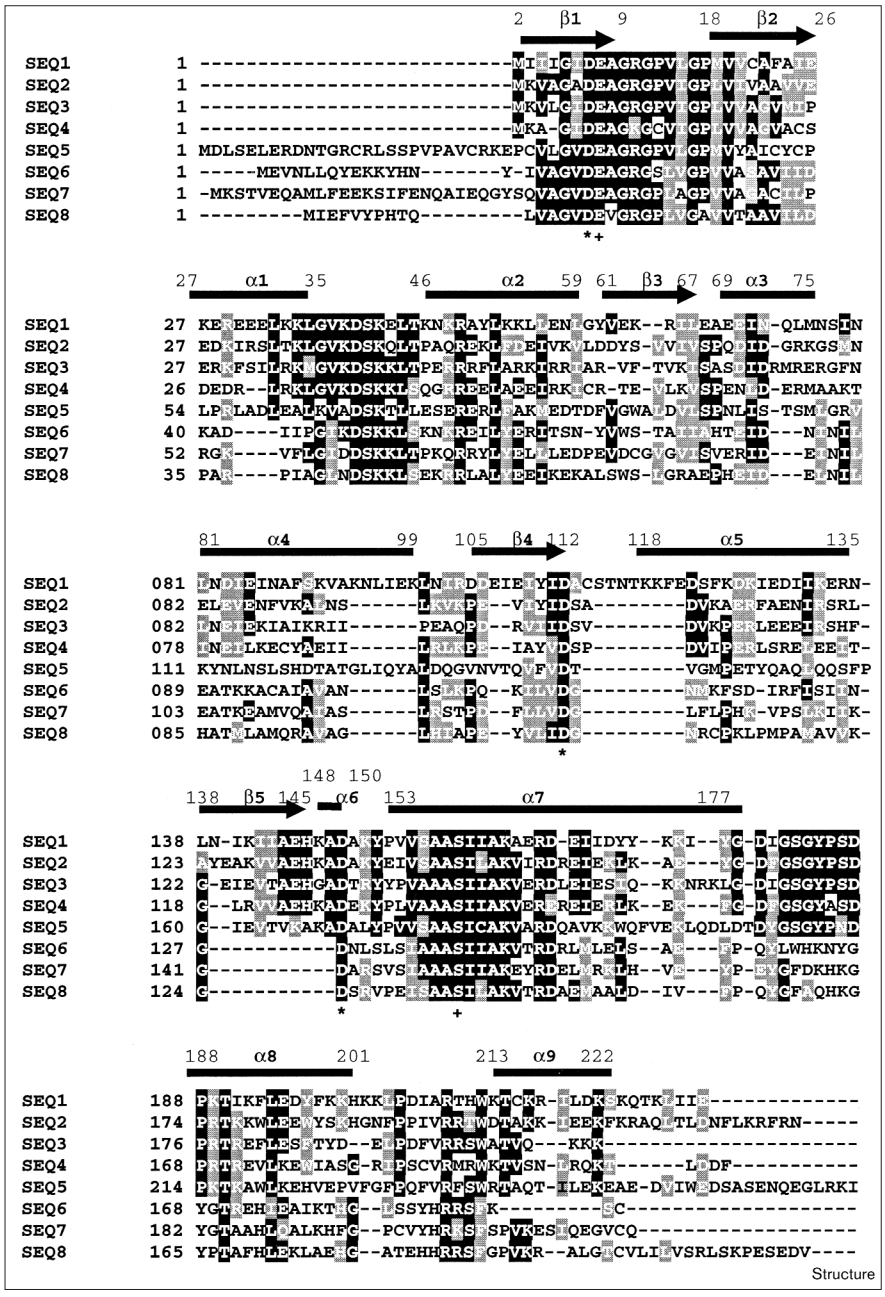

**(a)** Stereoview of the overall structure and active site of the Mj RNase HII monomer. The helices are colored purple, β strands green and loops orange. The N and C termini are labeled. The secondary structure assignment was based on PROCHECK [39]. The three active-site residues (Asp7, Asp112 and Asp149) and the bound MES are shown in ball-and-stick form. The figure was drawn using MOLSCRIPT [40]. **(b)** Stereoview of the active site in RNase HII, showing residues Asp7, Asp112, Asp149, three water molecules and nearby residues Ser159, Glu8 and His146. Hydrogen bonds are shown as dashed lines.

highly conserved carboxylates that are absolutely required for catalytic activity. Two of the three carboxylates are located at the same positions in space in all three proteins: one in the first β strand and the other at the C terminus of the fourth β strand. The location of the third carboxylate varies; when the three structures are aligned by superimposing their β sheets and the first two carboxylates, the third carboxylate in *E. coli* RNase H is on the opposite side of the β sheet to those in HIV integrase and RuvC.

To identify the potential active site of Mj RNase HII, we compared its tertiary structures with those of small *E. coli* RNase H and ASV integrase. When the central β sheets are superimposed, Asp7 aligns well with the first aspartate residue in the active sites of *E. coli* RNase H and ASV integrase, whereas Asp112 aligns well with the second aspartate in the active site (Figures 5b,c). Because the first two aspartates are highly conserved, Asp7 and Asp112 were assigned as active site residues. The third residue in the active site is not as clear as the first two. There are two carboxylates, Glu8 and Asp149, that are conserved in sequence (Figure 4) and are located close to Asp7 and Asp112 in space (Figure 3b). The two residues do not superimpose with any of the carboxylates in *E. coli* RNase H. On comparison with ASV integrase, however, Asp149 is close to Glu157, the third residue in the active site of integrase (Figure 5c). As discussed earlier, the overall topology of Mj RNase HII is more like integrase than *E. coli* RNase H. It is therefore reasonable that the third active site residue is in a similar position in the Mj RNase HII and integrase tertiary structures and on the opposite side of the β sheet compared with *E. coli* RNase H (Figures 5b,c).

The three possible active-site residues, Asp7, Asp112 and Asp149, form a tight hydrogen-bond network together with three well-defined water molecules (Figure 3b). As the crystal was grown under low $MgCl_2$ concentration (1 mM), no metal site can be assigned unambiguously in the electron-density map. There is also a conserved serine residue, Ser159 (Figures 3b,4), close to the acidic triad that forms a hydrogen bond with the sidechain carboxylate oxygen of Asp149. Glu8 and His146 are also in close proximity to the proposed active site. We suggest that the active site of Mj RNase HII is composed of an acid triad formed by Asp7, Asp112 and Asp149 (Figure 3b). Compared with the reaction mechanism of *E. coli* RNase H [10], Glu8 may act as the fourth carboxylate. Ser159 and His146 may also be involved in the catalytic reaction.

**Figure 4**



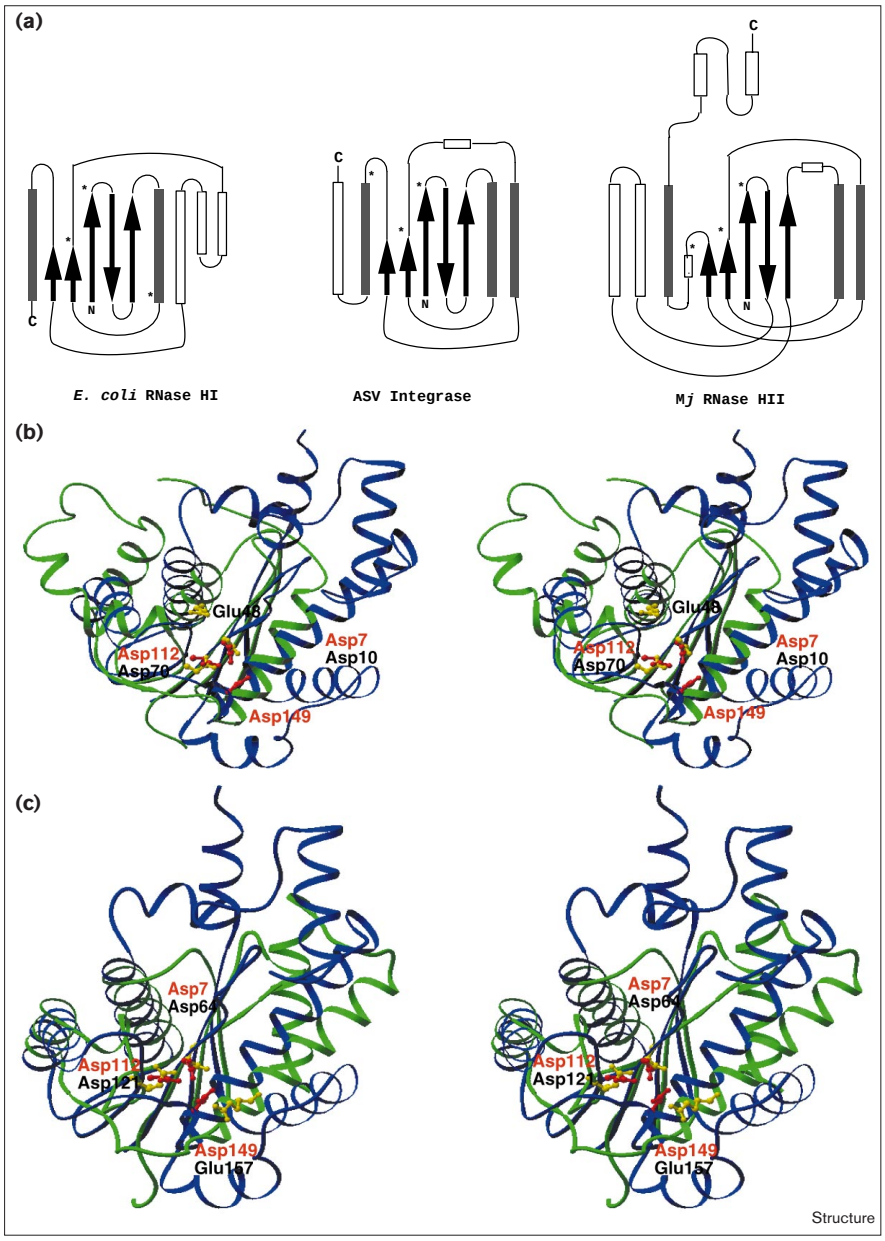

Multiple sequence alignment of various RNases HII. Secondary structure locations were also aligned with the sequences. The active site triad residues are denoted by an asterisk, and the nearby serine and glutamate are denoted by '+'. SEQ1, *M. janaschii* RNase HII; SEQ2, *Pyrococcus abyssi* RNase HII; SEQ3, *M. thermoautotrophicum* RNase HII; SEQ4, *Archaeoglobus fulgidus* RNase HII; SEQ5, the large subunit of human RNase HI; SEQ6, *Ricckettsia prowazekii* RNase HII; SEQ7, *Chlamydia trachomatis* RNase HII; SEQ8, *E. coli* RNase HII. The alignment was carried out using ClustalW [41]. Residues that are identical are shown as white letters on a black background; residues that are conserved are shown as white letters on a gray background.

These suggestions provide clues for further mutation and biochemical experiments in order to clearly define the active site and catalytic mechanism for class II RNases H.

### Substrate-binding site

In the present model of Mj RNase HII, a buffer molecule, MES, was found bound to the protein. Because MES contains a six-membered ring and a sulfonate connected by -$(CH2)_2$-, it resembles a nucleic acid monophosphate. MES binds to the cleft between the N- and C-terminal domains and is about 16 Å from the active site. The long cleft connecting the active site and the MES-binding site is mostly composed of conserved residues, including those in the loop between the first and the second strands (residues 10–17), the loop between the first and the second helix (residues 35–45), the N terminus of helix 7, helix 3, and the loop between helix 8 and helix 9. Residues 35–45 form a highly mobile loop but contain many conserved residues with positive charges, the conformation which may become rigid after binding to the substrate. When the electrostatic surface was calculated [28], a hollow positive surface with a negative patch

**Figure 5**




**(a)** Topology of *E. coli* RNase HI, ASV integrase and Mj RNase HII. β Strands are represented by black arrows, helices that are well superimposed are denoted by dark rectangles and helices that cannot be superimposed are represented by white rectangles. **(b)** Stereo diagram of Mj RNase HII superimposed on *E. coli* RNase HI. Mj RNase HII is in blue, and *E. coli* RNase HI in green. Sidechains of the active-site residues are drawn in ball-and-stick form. Active-site residues in Mj RNase HII (Asp7, Asp112 and Asp149) are shown in red. Active-site residues in *E. coli* RNase HI (Asp10, Asp70 and Glu48) are shown in yellow. Although Asp7 and Asp112 of the Mj protein can be superimposed onto Asp10 and Asp70 of the *E. coli* protein, Asp149 of the former and Glu48 of the latter are on the opposite side of the β sheet. **(c)** Stereo diagram of Mj RNase HII superimposed on ASV integrase. Mj RNase HII is shown in blue and ASV integrase in green. Sidechains of the active site residues are drawn in ball-and-stick form. Active-site residues in Mj RNase HII (Asp7, Asp112 and Asp149) are red. Active-site residues in ASV integrase (Asp64, Asp121 and Glu157) are yellow. The three residues at the active site can superimpose well. (b) and (c) were drawn using the program Ribbons [42].

(contributed by the active site) in the middle of one edge was found inside of the cleft (Figure 2b), a feature compatible with being an active site for nuclease activities.

It is interesting to note that there is also a buffer molecule, HEPES in the structure of integrase [23]. HEPES is similar to MES in that both contain a six-membered ring and a sulfonate connected by -$(CH2)_2$-. The location of the HEPES molecule is also similar to MES, as both of them are close to the loop formed by the first and the second strands. However, because RNase HII has an extra C-terminal domain compared with integrase and RNase H, the orientation of MES and HEPES is different. The HEPES molecule in integrase basically lies on the surface of the enzyme, while MES in RNase HII is between the N-terminal (RNase H fold) and the C-terminal domains. This implies that RNase HII may also have unique features for substrate binding and catalysis.

## Biological implications

**Mammalian RNase H (class II RNase H) participates in several important cellular processes such as DNA replication, the removal of the Okazaki fragments, and possibly DNA repair. Inhibiting human RNase H activity is also a challenge to antisense gene therapy. However, little information is available for this class of enzymes.**

**Table 1**

**Statistics for data collection.**

| | Number of reflections | | Completeness (%) | | $R_{sym}$ (%) | |
|---|---|---|---|---|---|---|
| Wavelength (Å) | Measured | Unique | All | Last shell | All | Last shell (2.03–2.00 Å) |
| 1.00 (remote) | 173,694 | 29,727 | 98.6 | 94.6 | 0.059 | 0.225 |
| 0.9801 (inflection) | 177,644 | 29,927 | 98.6 | 95.1 | 0.068 | 0.291 |
| 0.9799 (peak) | 176,247 | 29,930 | 98.9 | 97.6 | 0.071 | 0.258 |
| 0.9686 (remote) | 173,337 | 29,919 | 98.4 | 93.9 | 0.071 | 0.386 |

**The crystal structure of RNase HII from *Methanococcus jannaschii* at 2 Å resolution reveals two compact domains and, despite the absence of sequence similarity, the large N-terminal domain has a fold similar to that of *E. coli* and retrovirus RNase HI. The active site of RNase HII contains primarily three aspartates, Asp7, Asp112 and Asp149, and suggests a common catalytic mechanism between the two classes of enzyme. The nucleotide-binding site is located in the cleft between the N- and C-terminal domains. As the C-terminal domain only exists among the class II RNases H, this unique feature may explain the fact that the class II RNases H can specifically cut the RNA–DNA junctions in the chimerical DNA-RNA-DNA/DNA duplex, wheras the class I cannot.**

**The structure of archaeal RNase HII and the suggested model of the active site and substrate-binding site may help us to understand the biochemical function of the major mammalian RNase H. The structure may serve as a template for building homologous mammalian RNases H structure models for inhibitor design.**

## Materials and methods

*Expression and purification*

The coding region for RNase HII (Mj0135) of *M. jannaschi* [29] was amplified using the polymerase chain reaction (PCR), cloned into the pET-21a vector (Novagen), and transformed into the *E. coli* host cell B834 (DE3)/pSJS1244 [30] for isopropyl-β-D-thiogalactoside-dependent protein expression. For the wild-type protein, the cells were grown at 37°C in Luria–Bertani medium containing 50 μg $ml^{-1}$ ampicillin and 50 μg $ml^{-1}$ spectinomycin. The yield was approximately 8 mg purified Mj RNase HII per liter of culture. For the SeMet-containing protein, the cells were grown in M9 medium supplemented with SeMet. Because the protein is thermally stable, the cell extract was incubated at 80°C for 40 min to denature most of the *E. coli* proteins. The soluble fraction was purified by cation exchange chromatography over a Hi-Trap SP column (Pharmacia). The protein was applied in a buffer containing 50 mM MES (pH 6.2), 2 mM $MgCl_2$, 4 mM dithiothreitol (DTT) (buffer A) and eluted in a linear gradient of 0–0.5 M NaCl in buffer A. The protein eluted at 0.3 M NaCl.

*Crystallization and data collection*

Crystallization conditions were screened by the sparse matrix method [31] using a PEG/Ion screen (Hampton Research, Laguna Hills, CA). Needle-shaped crystals were obtained by hanging-drop vapor diffusion at 22°C. The drop was made by mixing the protein solution (30 mg $ml^{-1}$ Mj RNase HII, 50 mM MES, 2 mM $MgCl_2$, 4 mM DTT, pH 6.2) with the well solution (0.2 M KSCN, 22% PEG 3350, pH 6.8) at a 1:1 ratio. These were subsequently crushed and streak-seeded into a drop containing 15 mg $ml^{-1}$ protein, 25 mM MES (pH 6.2), 1 mM $MgCl_2$, 2 mM DTT, 0.1 M KSCN, 9% PEG 3350 and equilibrated against 0.2 M KSCN, 18% PEG 3350 at pH 6.8 in the well. Rod-like crystals with dimensions 0.2 mm × 0.3 mm × 0.4 mm were grown within three days. Crystals of the SeMet-containing protein were obtained under similar conditions and seeding procedures, but the crystals were smaller. The crystals belong to space group $P2_1$ with unit-cell parameters of a = 61.451 Å, b = 55.418 Å, c = 72.148 Å, β = 111.83° at 100K and contain two protein molecules per asymmetric unit. The crystals were flash-frozen in solution containing 40% glycerol, 20% PEG 3350, 0.2 M KSCN, pH 6.8, before exposure to X-ray.

X-ray diffraction data sets (Table 1) of the SeMet protein were collected at four wavelengths at the MCF beamline 5.0.2 at the Advanced Light Source of the EO Lawrence Berkeley National Laboratory. The data were processed with DENZO and reduced with SCALEPACK [32]. The SOLVE program package was used to carry out heavy-atom searching and phasing [33,34]. Six selenium sites were found in one asymmetric unit, corresponding to three expected SeMets (including the one at the N termini) per protein monomer.

*Structure determination*

The electron-density map was modified by solvent flattening using DM [35] with a solvent content of 45% and followed by twofold noncrystallographic symmetry (NCS) averaging. At this stage, a β sheet and several helices could be seen clearly in the map, while the connections between the secondary structures were not clear. Initial models were built into the map using O program [36], followed by multiple rounds of solvent envelope calculation, NCS averaging, model extension and rebuilding. Refinement was carried out using CNS program [37,38] against the low-remote (1 Å wavelength) data at 100K. Noncrystallographic symmetry restraints

**Table 2**

**Refinement statistics of Mj RNase HII at 2.0 Å resolution.**

| | |
|---|---|
| Resolution (Å) | 20–2 |
| R factor (%) | 21.0 |
| $R_{free}$ (5% total data) (%) | 26.2 |
| $R_{free}$ last shell (2.00–2.05Å) (%) | 26.9 |
| Number of nonhydrogen atoms | 3,952 |
| Number of 'substrate' atoms | 2 MES |
| Number of water molecules | 341 |
| Bond lengths rmsd (Å) | 0.008 |
| Bond angles rmsd (°) | 1.258 |
| Average B factor ($Å^2$) | |
| Protein | 32.98 |
| MES | 43.98 |
| Water | 38.93 |
| Ramachandran plot | |
| Most favored (%) | 92.7 |
| Additional allowed (%) | 7.3 |
| Generously allowed (%) | 0.0 |
| Disallowed (%) | 0.0 |

were initially applied and gradually released when the $R_{free}$ value dropped below 0.30. In the final refinement, no NCS restraints were used. Five per cent of the total data were used to monitor the $R_{free}$ throughout the refinement. After the $R_{free}$dropped to 0.30, 313 water molecules and two MES molecules were identified at several stages in the electron-density map and added using the WATERPICK module in CNS and visual validation. The final model, refined using 2 Å resolution data, has an R value of 21% and $R_{free}$ of 26%. The model was validated by PROCHECK [39]. The structure has a stereochemistry, with all residues in the allowed regions of the Ramachandran plot (Table 2). The central core of the structure has very clear electron density. The five C-terminal residues (226–230) are disordered in both molecules. Loop residues 40–42 in molecule A have very weak density and are not included in the model.

### *Accession numbers*

The coordinates have been deposited in the PDB with the accession code 1EKE.

## Acknowledgements

We are grateful to S Kanaya of Osaka University, Japan for the activity assay of Mj RNAse HII on an RNA/DNA hybrid. This work was supported by the Director, Office of Science, Office of Biological and Enviromental Reasearch under the US Department of Energy under contract No. DE-AC03-76SF00098.

## References

1. Frank, P., Braunshofer-Reiter, C., Wintersberger, U., Grimm, R. & Busen, W. (1998). Cloning of the cDNA encoding the large subunit of human RNase HI, a homologue of the prokaryotic RNase HII. *Proc. Natl. Acad. Sci. USA* **95**, 12872-12877.
2. Ohtani, N., *et al.*, & Kanaya, S. (1999). Identification of the genes encoding $Mn^{2+}$-dependent RNase HII and $Mg^{2+}$-dependent RNase HIII from *Bacillus subtilis*: classification of RNases H into three families. *Biochemistry* **38**, 605-618.
3. Davies, J.F.d., 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.
4. 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.
5. Katayanagi, K., *et al.*, & Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* **347**, 306-309.
6. Yamasaki, K., Akasako-Furukawa, A. & Kanaya, S. (1998). Structural stability and internal motions of *Escherichia coli* ribonuclease HI: $^{15}N$ relaxation and hydrogen-deuterium exchange analyses. *J. Mol. Biol.* **277**, 707-722.
7. Haruki, M., Noguchi, E., Kanaya, S. & Crouch, R.J. (1997). Kinetic and stoichiometric analysis for the binding of *Escherichia coli* ribonuclease HI to RNA-DNA hybrids using surface plasmon resonance. *J. Biol. Chem.* **272**, 22015-22022.
8. Kanaya, S., Oobatake, M. & Liu, Y. (1996). Thermal stability of *Escherichia coli* ribonuclease HI and its active site mutants in the presence and absence of the $Mg^{2+}$ ion. Proposal of a novel catalytic role for Glu48. *J. Biol. Chem.* **271**, 32729-32736.
9. Iwai, S., *et al.*, & Nakamura, H. (1996). Interaction of the basic protrusion of *Escherichia coli* ribonuclease HI with its substrate. *J. Mol. Biol.* **263**, 699-706.
10. Kashiwagi, T., *et al.*, & Morikawa, K. (1996). Proposal for new catalytic roles for two invariant residues in *Escherichia coli* ribonuclease HI. *Protein Eng.* **9**, 857-867.
11. Dabora, J.M., Pelton, J.G. & Marqusee, S. (1996). Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry* **35**, 11951-11958.
12. Raschke, T.M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat. Struct. Biol.* **6**, 825-831.
13. Smith, C.M., Smith, J.S. & Roth, M.J. (1999). RNase H requirements for the second strand transfer reaction of human immunodeficiency virus type 1 reverse transcription. *J. Virol.* **73**, 6573-6581.
14. Krynetskaia, N.F., Krynetski, E.Y. & Evans, W.E. (1999). Human RNase H-mediated RNA cleavage from DNA–RNA duplexes is inhibited by 6-deoxythioguanosine incorporation into DNA. *Mol. Pharmacol.* **56**, 841-848.
15. Werner, S. & Wohrl, B.M. (1999). Soluble Rous sarcoma virus reverse transcriptases α, αβ, and β purified from insect cells are processive DNA polymerases that lack an RNase H 3′→5′ directed processing activity. *J. Biol. Chem.* **274**, 26329-26336.
16. Itaya, M. (1990). Isolation and characterization of a second RNase H (RNase HII) of *Escherichia coli* K-12 encoded by the *rnhB* gene. *Proc. Natl. Acad. Sci. USA* **87**, 8587-8591.
17. Zhang, Y.B., Ayalew, S. & Lacks, S.A. (1997). The *rnhB* gene encoding RNase HII of *Streptococcus pneumoniae* and evidence of conserved motifs in eucaryotic genes. *J. Bacteriol.* **179**, 3828-3836.
18. Haruki, M., *et al.*, & Kanaya, S. (1998). Gene cloning and characterization of recombinant RNase HII from a hyperthermophilic archaeon. *J. Bacteriol.* **180**, 6207-6214.
19. Turchi, J.J., Huang, L., Murante, R.S., Kim, Y. & Bambara, R.A. (1994). Enzymatic completion of mammalian lagging-strand DNA replication. *Proc. Natl. Acad. Sci. USA* **91**, 9803-9807.
20. Frank, P., Braunshofer-Reiter, C., Karwan, A., Grimm, R. & Wintersberger, U. (1999). Purification of *Saccharomyces cerevisiae* RNase H(70) and identification of the corresponding gene. *FEBS Lett.* **450**, 251-256.
21. Yang, W. & Steitz, T.A. (1995). Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3**, 131-134.
22. Rice, P., Craigie, R. & Davies, D.R. (1996). Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **6**, 76-83.
23. Bujacz, G., *et al.*, & Skalka, A.M. (1995). High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* **253**, 333-346.
24. Bujacz, G., *et al.*, & Skalka, A.M. (1996). The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* **4**, 89-96.
25. Rice, P. & Mizuuchi, K. (1995). Structure of the bateriophage Mu transposase core: a common structural motif for DNA transposition and retroviral intergration. *Cell* **82**, 209-220.
26. Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H., & Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. *Cell* **78**, 1063-1072.
27. Katayanagi, K., *et al.*, & Morikawa, K. (1992). Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* **223**, 1029-1052.
28. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
29. Bult, C.J., *et al.*, & Venter, J.C. (1996). Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. *Science* **273**, 1058-1073.
30. Kim, R., *et al.*, & Kim, S.H. (1998). Overexpression of archaeal proteins in *Escherichia coli*. *Biotechnol. Lett.* **20**, 207-210.
31. Jancarik, J. & Kim, S.H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **21**, 916-924.
32. Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing.* (Sawyer, L., Isaacs, N. & Bailey, S., eds.) pp. 56-62, SERC Daresbury Laboratory, UK.
33. Terwilliger, T.C. & Berendzen, J. (1997). Bayesian correlated MAD phasing. *Acta Crystallogr. D* **53** , 571-579.
34. Terwilliger, T.C. (1997). Multiwavelength anomalous diffraction phasing of macromolecular structures: analysis of MAD data as single isomorphous replacement with anomalous scattering data using the MADMRG Program. *Methods Enzymol.* **276**, 530-537.
35. Cowtain, K.D. & Main, P. (1993). Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr. D* **49**, 148-157.
36. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaad, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
37. Brünger, A.T., Adams, P.D. & Rice, L.M. (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr. Opin. Struct. Biol.* **8**, 606-611.
38. Brünger, A.T., *et al.*, & Warren, G.L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
39. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). Procheck – a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
40. Kraulis, P.J. (1991). Molscript – a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
41. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). Clustal W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
42. Carson, M. (1991). Ribbons 2 0. *J. Appl. Crystallogr.* **24**, 958-961.