Highly Specific Recognition of Primer RNA Structures for 2'-OH Priming Reaction by Bacterial Reverse Transcriptases*

(Received for publication, June 15, 1999, and in revised form, August 4, 1999)

Sumiko Inouye, Mei-Yin Hsu, Aiguo Xu, and Masayori Inouye‡

From the Department of Biochemistry, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

A minor population of Escherichia coli contains retro-elements called retrons, which encode reverse transcriptases (RT) to synthesize peculiar satellite DNAs called multicopy single-stranded DNA (msDNA). These RTs recognize specific RNA structures in their individual primer-template RNAs to initiate cDNA synthesis from the 2'-OH group of a specific internal G residue (branching G residue). The resulting products (msDNA) consist of RNA and single-stranded DNA, sharing hardly any sequence homology. Here, we investigated how RT-Ec86 recognizes the specific RNA structure in its primer-template RNA. On the basis of structural comparison with HIV-1 RT, domain exchanges were carried out between two E. coli RTs, RT-Ec86 and RT-Ec73. RT-Ec86 (320 residues) and RT-Ec73 (316 residues) share only 71 identical residues (22%). From the analysis of 10 such constructs, the C-terminal 91-residue sequence of RT-Ec86 was found to be essential for the recognition of the unique stem-loop structure and the branching G residue in the primer-template RNA for retron-Ec86. Using the SELEX (systematic evolution of ligands by exponential enrichment) method with RT-Ec86 and primer RNAs containing random sequences, the identical stem-loop structure (including the 3'-U loop) to that found in the retron-Ec86 primer-template RNA was enriched. In addition, the highly conserved 4-base sequence (UAGC), including the branching G residue, was also enriched. These results indicate that the highly diverse C-terminal region recognizes specific stem-loop structures and the branching G residue located upstream of the stem-loop structure. The present results with seemingly primitive RNA-dependent DNA polymerases provide insight into the mechanisms for specific protein RNA recognition.

Retrons are bacterial retro-elements found in a number of bacteria such as Escherichia coli (1–7), Myxococcus xanthus (8, 9), Stigmatella aurantiaca (10), Rhizobium, Salmonella, Klebsiella, Proteus (11), and Melittangium lichenicola (12). The most unusual feature of the retrons is their product, a small satellite DNA, which links to an internal G residue of a small RNA molecule by a 2',5'-phosphodiester linkage (for reviews, see Refs. 13 and 14). This small DNA-RNA complex is called multicopy single-stranded DNA (msDNA)1; reverse transcriptase (RT) encoded by a retron is essential for the msDNA synthesis. A single RNA transcript (primer-template RNA) from a retron is used as primer as well as template for msDNA synthesis; the 5'-end and the 3'-end regions of the transcript form a stable duplex placing a specific G residue at the end of the duplex. The 2'-OH group of the G residue is used by the retron RT to prime cDNA synthesis using the same RNA strand as template. The template RNA used for cDNA synthesis is removed by ribonuclease H, and cDNA synthesis stops at a specific site on the template for individual msDNAs. As a result, the msDNAs identified so far have the following characteristics: the 5'-end of a single-stranded DNA from 48 to 163 bases in length links to the 2'-OH group of an internal G residue (termed the branching G residue) of a single-stranded RNA of 50 to 120 bases in length (8); both DNA and RNA molecules contain stable secondary structures (10); and the 3'-ends of both DNA and RNA molecules are complementary to each other, forming a DNA-RNA heteroduplex (1) (see Refs. 13 and 14 for reviews).

Notably, individual msDNA synthesis is highly specific to individual retrons. RTs are not exchangeable between two retrons, unless the msr region in a retron is replaced from the same retron of RT (15). The msr region encodes a short single-stranded RNA part of msDNA. This RNA forms one or two stable stem-loop structures; the RNA structures in msDNA-Ec73 and msDNA-ms86 are shown in Fig. 1, A and B, respectively. As can be seen, there are virtually no primary sequence homologies between the two RNA molecules. This unique structural feature is considered responsible for the priming reaction for individual msDNA synthesis.

In the present paper, we have attempted to determine how RT could specifically recognize a primer-template RNA to prime cDNA synthesis from the 2'-OH group of the branching G residue. In particular, using two RTs from E. coli retrons Ec73 (17) and Ec86 (4), we identified the regions that are essential for the recognition of their cognate primer-template structures. These two RTs, RT-Ec73 and RT-Ec86, consist of 316 and 320 amino acid residues, respectively, sharing only 71 identical residues (22% of RT-Ec73). Results from the domain analysis for RNA recognition indicate that presumed primitive DNA polymerases from bacterial retrons have a highly diversified C-terminal region consisting of approximately 90 residues, which plays an essential role in the recognition of a unique stem-loop structure and its upstream sequence containing the branching G residue. It appears that the C-terminal domains of individual bacterial RTs have evolved to recognize highly specific RNA structures for the 2'-OH primer reaction, and they share a structural homology for DNA polymerase common to all RTs from the prokaryotes to the eukaryotes. The
present results also raise interesting questions as to the molecular mechanisms of the priming reaction of bacterial RTs and the mechanisms of RNA-protein interaction.

**EXPERIMENTAL PROCEDURES**

**Materials**—\([\alpha-\text{\textsuperscript{32}}\text{P}]\text{dCTP}, \text{\textsuperscript{35}}\text{S}\)dATP, and \([\alpha-\text{\textsuperscript{32}}\text{P}])\text{dATP}\) were purchased from Amersham Pharmacia Biotech. Taq polymerase and T4 ligase were purchased from Roche Molecular Biochemicals. Restriction enzymes were from Roche Molecular Biochemicals and New England Biolabs. DNA sequencing was performed using a Sequenase kit from U. S. Biochemical Corp. The TA vector kit was purchased from Invitrogen.

**Bacterial Strains**—E. coli JM83 (19) was used for propagation of plasmids, and E. coli LE392 (DE3), a K-12 strain carrying the gene for T7 polymerase (16), was used for purification of various constructs of RTs.

**Construction of Various Hybrid RTs**—To exchange C-terminal regions between RT-Ec73 and RT-Ec86, an Mspl site was created in their VTG1 coding region by changing the ACA codon for Thr-244 to ACC by site-directed mutagenesis (17). After the DNA sequence was confirmed, the 240-base pair Mspl-BamHI fragments were isolated from pET11RT73 and pET11RT86 and exchanged between them. Other hybrid RTs were constructed using two-step PCR with proper primers by the method described previously (16). All RT genes were cloned in pET11(km) under the control of a T7 promoter and His\(_6\)-tagged at their C-terminal ends except for RT-Ec73. pET73RT(His), in which His\(_6\) was added at the N-terminal end, was constructed previously (16).

**Purification of Various Hybrid RTs**—Purification of RTs was performed using the method described previously (16) with some modifications. Cells were grown in M9 medium supplemented with 0.2% Casamino acids (Difco), 0.4% glucose, 20 μg/ml tryptophan, 2 μg/ml thiamine, 0.8 mM MgSO\(_4\), and 50 μg/ml kanamycin at 37 °C up to 90 klett units. Then, cells were harvested by centrifugation. The mixture was incubated at 95 °C for 15 min, and then 0.5 mM CTP was added. A 1-h incubation, treated with 3 units of RNase-free DNase for 15 min at 37 °C, and then extracted with phenol and chloroform. A 1/10 volume of 3 M sodium acetate (pH 5.1) and 3 volumes of ethanol were added, and the mixture was stored at −70 °C separately until needed. Immediately before use for binding assay, the RNA was precipitated by centrifugation.

**Binding Assay**—RNA prepared was solubilized in an annealing buffer (50 mM Tris-HCl, pH 8.0, 10 mM MgCl\(_2\)) in 1 pmol/μl. The mixture was incubated at 95 °C for 2 min, 37 °C for 30 min, and 4 °C for 30 min for the formation of secondary structures. RT was preincubated in a binding buffer (the final concentration of 10% glycerol, 10 mM Tris-HCl, pH 7.8, 2 mM dithiothreitol, 35 mM MgCl\(_2\), 1 mg/ml RNA) for 10 min at room temperature. After 10 min, 1 μl of an annealed RNA and 1 μl of 10 mM ATP, and H\(_2\)O to make 20 μl was added. The reaction mixture was incubated for 30 min at room temperature. A typical reaction mixture contained 1 pmol of RNA and 4 pmol of RT in 20 μl. The reaction was stopped by adding 5 μl of 5% dye mixture (25% glycerol, 0.1 mM EDTA, 0.025% bromphenol blue, and 0.025% xylene cyanol), and the sample was analyzed by 8% polyacrylamide gel electrophoresis at 4 °C.

**SELEX Method**—Two randomized oligomers, oligo 8081 for RNA I and oligo 7498 for RNA II, 5'-end oligo 8082, and 3'-end oligo 6817, all used for the SELEX method, are listed in Table II. First, a double-stranded DNA was synthesized by the Klenow enzyme using oligo 7498 and oligo 6817 as template and primer, respectively. Two hundred pmol of oligo 7498 and 600 pmol of oligo 6817 were annealed in 50 μl of a buffer containing 10 mM Tris-HCl (pH 7.5), 10 mM MgCl\(_2\), 6 mM NaCl, and 6 mM β-mercaptoethanol, and then 1 μl each of a 2 mM dNTP mixture and the Klenow enzyme were added. After incubation at room temperature for 30 min, the mixture was mixed with 50 μl of a DNA fragmentation buffer (7 mM RT polymerase using ATP, UTP, CTP, and GTP each at 2.5 mM).

The purified RT (100 pmol) was preincubated in 400 μl of a binding buffer, and then 100 μl of an annealed RNA (200 pmol) were added to the RT solution. After incubation for 30 min at room temperature, the reaction mixture was added to an Ni-NTA column, and RNA-RT complex was eluted with 0.2 M imidazole in a buffer (50 mM Tris-HCl, pH 8.0, 1 mM dithiothreitol, and 1 mM EDTA). After treating with phenol
and chlorofom, RNAs were precipitated with ethanol and used for cDNA synthesis by avian myeloblastosis virus RT with oligo 6817. The synthesized cDNAs were amplified by PCR with oligo 8082 and oligo 6817 as primers. The amplified DNA was purified by gel electrophoresis and used for the first cycle of SELEX (27). After six cycles, DNA fragments were cloned into pCRII vector (Invitrogen) and the DNA sequence was determined by the dideoxynucleotide chain termination method (20).

RESULTS

Bacterial RTs Recognize Only Their Cognate Primer-Template RNAs—A minor population of wild E. coli strains contains highly diverse retrons producing their own specific msDNAs with little homology among them. To date, seven E. coli retrons have been identified: retron-Ec67 (2) retron-Ec86 (4), retron-Ec73 (7), retron-Ec107 (1), retron-Ec83 (3), retron-Ec78 (5), and retron-Ec48 (6). RTs encoded by these retrons are highly diverse; the identities between any two E. coli RTs are less than 30%, and there are virtually no homologies in primary sequences between any two primer-template RNA molecules. The only invariable features among the primer-template RNAs are the branching G residues, located at the end of a stem formed in the primer-template RNA (Fig. 1), and a highly conserved 4-base sequence, UAGC, in which the G residue corresponds to the branching G residue. Downstream of the G residue, there are always stable secondary structures, which have been shown to be essential for RT recognition of its cognate primer-template RNA (15).

To investigate the mechanisms for the highly selective recognition of specific RNA structures by individual bacterial RTs, we first constructed T7 systems to produce two RNA transcripts, Ec73msr-msd30 (Fig. 1A) and Ec86msr-msd48 (Fig. 1B). These RNAs are truncated versions of the primer-template RNAs for RT-Ec73 and RT-Ec86, respectively; the former RNA was produced from a mutated retron-Ec73 containing a 43-base pair deletion in the msd region (the DNA template region of msDNA-Ec73). The mutated retron-Ec73 was still capable of efficiently producing an msDNA, called msDNA-miniEc73, consisting of only 30-base single-stranded DNA in vivo (21). Similarly, the latter RNA was produced from a mutated retron-Ec86 containing a 38-base pair deletion in the msd region. This mutated retron produced msDNA-miniEc86, consisting of a 48-base single-stranded DNA (not shown). These miniprimer-template RNAs were designated Ec73msr-msd30 (114 bases) and Ec86msr-msd48 (133 bases), respectively, and used for the present experiment, because these RNAs were more suitable for cell-free msDNA synthesis and gel mobility shift experiments than the full-length primer-template molecules.

First, purified RTs RT-Ec73 and RT-Ec86 were tested for their specific binding to the miniprimer-template RNAs by gel retardation analysis. As shown in Fig. 1C, when mini-Ec73 RNA was used, only RT-Ec73 bound to the RNA (lane 2), but RT-Ec86 did not (lane 5). Similarly, when mini-Ec86 RNA was used for gel retardation assay, RT-Ec86 bound to the RNA (Fig. 1D, lane 5), but RT-Ec73 did not (lane 2).

Identification of Domains for Specific RNA Recognition—The results presented above demonstrate that bacterial RTs can recognize only their cognate primer-template RNAs. To identify the unique domain structures in the bacterial RTs, which are responsible for the specific RNA recognition, we next attempted to align RT-Ec73 and RT-Ec86 with HIV-1 RT.

Bacterial RTs and eukaryotic RTs have been shown to be evolutionarily related to each other, and phylogenetic interrelationships have also been proposed (22, 23). On the basis of the three-dimensional structure of HIV-1 RT and sequence similarities between bacterial RTs and HIV-1 RT (24, 25), domain assignments were proposed for RT-Ec73 and RT-Ec86 (26). Fig. 2A shows a modified version of the previous alignment, where 15 β-strand structures, from β6 to β14, and 10 α-helical structures, from αA to αD, of HIV-1 RT are aligned to bacterial RTs. In these alignments, there are only one invariant residues among all known RTs, form the catalytic triad essential for DNA polymerase activity (25). Between RT-Ec73 and RT-Ec86, there are 71 identical residues (22% identity) and 46 similar residues. In the structural assignments shown in Fig. 2A, the C-terminal regions, consisting of αG, αH, αI, and αJ, correspond to the thumb domain of HIV-1 RT (25). These regions are particularly highly diversified among bacterial RTs, suggesting that these regions may be involved in recognition of unique primer-template RNA structures for individual RTs. In these alignments, there are two regions, X and Y (boxed in Fig. 2A), that are unique in bacterial RTs. Note that in the Y sequence there is the highly conserved sequence, VTGL (marked by open circles in Fig. 2A), which is found in all bacterial RTs known today (26).

To identify the RT regions required to specifically recognize the individual primer-template RNA, a number of hybrid proteins between RT-Ec73 and RT-Ec86 were constructed as shown in Fig. 2B. First, we exchanged the C-terminal region at the highly conserved VTGL sequence between RT-Ec73 and RT-Ec86 by using a MspI site created within the VTGL-corresponding region. In RT-Ec73/86, the C-terminal fragment of RT-Ec73 downstream of the VTGL residue (238–241) was exchanged with the corresponding fragment from RT-Ec86 (see Fig. 2B, construct a). Similarly, in RT-Ec86/73 the C-terminal fragment downstream of the VTGL sequence (residue 243–246) was exchanged with the corresponding fragment from RT-Ec73 (see Fig. 2B, construct b). The resulting hybrid RTs, RT-Ec73/86 and RT-Ec86/73 (see Fig. 2B, constructs a and b, respectively), were tested for binding to miniprimer-template RNAs (Fig. 1A and B). Gel retardation assays are shown in Fig. 1, C and D, when the mini-Ec73 RNA was used, RT-Ec86/73 but not RT-Ec73/86 bound to the RNA (Fig. 1C, lanes 4 and 3, respectively). Conversely, when the mini-Ec86 RNA was used, RT-Ec73/86 but not RT-Ec86/73 bound to RNA (Fig. 1D, lanes 3 and 4, respectively). These results indicate that the Y sequence and/or the C-terminal fragment following the Y sequence play a crucial role in the specific RNA recognition by RTs. In addition to these sequences, however, the N-terminal fragment upstream of the Y sequence is required for msDNA synthesis in vivo and also seems to participate in the specific recognition of the RNA structure as discussed below.

Specificity of Recognition of Stem-Loop Structures—We next attempted to dissect its primer-template RNA using RT-Ec86 and its primer-template RNA (Ec86msr-msd48 RNA; see Fig. 1B) to identify the region(s) required for the binding to RT-Ec86. For this purpose, we used only the primer RNA portion of Ec86msr-msd48 from base 1 to 82. Thus, the resulting RNA, called Ec86msr RNA, consists of 82 bases with two stem-loop structures, termed a and b, as shown in Fig. 3A. This RNA was retarded in gel electrophoresis in the presence of RT-Ec86 (Fig. 3B, lane 2) but not in the presence of RT-Ec73 (Fig. 3B, lane 3).
3D, lane 2) as the original primer-template RNA (Ec86msr-msd48 RNA; see Fig. 1D, lane 5). Next, stem-loop structures a and b were separately eliminated from Ec86msr RNA to form either Ec86msrD a RNA (deletion of the entire 25-base stem-loop structure a) or Ec86msrD b RNA (deletion of the entire 17-base stem-loop structure b) (see Fig. 3, B and C, respectively). These RNAs were used for gel retardation assay with RT-Ec86. As shown in Fig. 3E, Ec86msrΔa RNA (1 pmol) was completely retarded by RT-Ec86 (4 pmol) (lane 3) under the identical condition used for Ec86mrs RNA, shown in Fig. 3D. In contrast, when Ec86msrΔb RNA was used, RT-Ec86 hardly bound to the RNA (Fig. 3F). Even if the RNA amount was increased to 8 pmol (lane 6), only a faint retarded band was observed. Note that the mobility of Ec86msrΔa RNA was not retarded at all by bovine serum albumin (BSA; Fig. 3E, lane 2). The retardation of Ec86msrΔa RNA by RT-86 was completely inhibited in the presence of a 10- or 30-fold excess of nonradioactive Ec86msr-msd48 RNA (Fig. 3E, lanes 4 and 5, respectively) but not in the presence of Ec73msr-msd30 RNA (lanes 6 and 7). These results demonstrate that stem-loop structure b is specifically recognized by RT-Ec86 and that stem structure a has a weak requirement for recognition. Preliminary experiments revealed that the $K_d$ values of RT-Ec86 and RT-Ec73/86 for Ec86msr RNA were at the level of $10^{-8}$ M as measured by a gel retardation assay (not shown).

Roles of X and Y Sequences in RNA Recognition—Earlier, we showed that the Y sequence and/or the C-terminal fragment downstream of the Y sequence are responsible for specific RNA recognition (Fig. 1). To further characterize the role of the Y and X sequences, these sequences or the segments containing these sequences were exchanged between RT-Ec73 and RT-Ec86, resulting in eight new constructs, from c to j, as shown in Fig. 2B. These hybrid RTs were purified and then tested for their abilities to bind to Ec73msr-msd30 RNA (Fig. 1A) and Ec86msr RNA (Fig. 3A). The results are shown in Fig. 5, A and B.

When Ec73msr-msd30 RNA was used, the following four constructs caused mobility shift as judged from the formation of new upper bands accompanied with density reduction at the position of free RNA: RT-Ec73 (Fig. 4A, lane 1); RT-Ec73X86 (Fig. 2B, construct f; only the X region of RT-Ec73 was exchanged with that of RT-Ec86; lane 3); RT-Ec86/73 (construct b; lane 6); RT-Ec73/86/73 (construct c; the sequence of RT-Ec73 between the X and Y regions was exchanged with that from RT-Ec86; lane 7); and RT-Ec86(230–320) (construct j; the C-terminal 91 residue fragment of RT-Ec86; lane 11). Except for the RT-Ec86(230–320) fragment, all hybrid proteins that caused gel mobility shift contained the Y sequence and the....
The following 52-residue C-terminal fragment, both of which are derived from RT-Ec73. These results indicate that the entire C-terminal region of RT-Ec73, including its Y sequence, is required for the recognition of the template-primer RNA. Interestingly, the C-terminal 91-residue fragment from RT-Ec73 and RT-Ec86, respectively, and the C-terminal region downstream of the Y sequence (RT-Ec73 and RT-Ec86) were carried out within the region of the highly conserved Asp residue immediately downstream of the X sequence (see legend for A and “Experimental Procedures”), and thus RT-Ec73/86 and RT-Ec86/73 were constructed. Additional exchanges of the N-terminal fragments, including the X sequence (RT-Ec73/86/73 and RT-Ec86/73/86), were carried out within the region of the X sequence from Ile-228 to Ser-264 of RT-Ec73 was deleted. The X sequence exchanges were carried out between the X sequence from Leu-55 to Lys-83 of RT-Ec73 and the Y sequence from Ile-228 to Ser-264 of RT-Ec73 was deleted. The X sequence exchanges were carried out between the X sequence from Leu-55 to Lys-83 of RT-Ec73 and the Y sequence from Leu-55 to Lys-83 of RT-Ec73 and the Y sequence from Ile-228 to Ser-264 of RT-Ec73 was deleted. The C-terminal 9-residue fragment of RT-Ec86 (RT-Ec86(230–320)) contained the fragment downstream of Lys-230. All of the constructs were His-tagged and expressed in a T7 vector as described under “Experimental Procedures.”
recognition, and that both sequences have to be derived from the same RT. Interestingly, however, the N-terminal domain upstream of the Y sequence either from the same RT protein or a different RT protein is also involved in RNA recognition specifically.

**Recognition of Unique RNA Structures by RT-Ec86**—To de-
**RNA Recognition by Bacterial RT for 2'-OH Priming Reaction**

**Fig. 5. Determination of RNA binding specificity of RT-Ec86 and RT-Ec73/86 by the SELEX method.** The structure of Ec86msr RNA used for the SELEX method is shown at the top. Two different RNA molecules were used for the SELEX method; in RNA I the 10-base sequence from base 9 to 18 was randomized, and in RNA II the 11-base sequence from base 23 to 33 was randomized. These sequences are boxed and are marked I and II, respectively. The branching G residue is circled. A, sequence enrichment in RNA I with RT-Ec86 by the SELEX method. After six cycles, 9 clones were randomly picked up and their DNA sequences were subsequently determined. Highly conserved AGC or AG sequences and the highly conserved G residue at the second position are shown in bold. The sequence of Ec86msr RNA, corresponding to the randomized region, is shown in **bold** and boxed at the bottom. B, sequence enrichment in RNA II with RT-Ec86 after six cycles of the SELEX method. Fourteen random clones were sequenced, and their DNA sequences were determined as shown. Those bases, which are identical to the wild-type sequence shown at the bottom, are shown in **bold**. C, sequence enrichment in RNA II with RT-Ec73/86 after 6 cycle of the SELEX method. Fourteen random clones were sequenced as described for B. Note that sequences 7 and 14 have a 1-base insertion and a 1-base deletion, respectively. Those sequences for Ec86 RNA (bold) and Ec73 RNA are boxed at the bottom.

- **RNA I**
  - 1: UUGGAGAAGGC
  - 2: UUGGAGAAGGC
  - 3: UUGGAGAAGGC
  - 4: UUGGAGAAGGC
  - 5: UUGGAGAAGGC
  - 6: UUGGAGAAGGC
  - 7: UUGGAGAAGGC
  - 8: UUGGAGAAGGC
  - 9: UUGGAGAAGGC
  - 10: UUGGAGAAGGC
  - 11: UUGGAGAAGGC
  - 12: UUGGAGAAGGC
  - 13: UUGGAGAAGGC
  - 14: UUGGAGAAGGC

- **RNA II**
  - 1: GUUGUGUGCCGC
  - 2: GUUGUGUGCCGC
  - 3: GUUGUGUGCCGC
  - 4: GUUGUGUGCCGC
  - 5: GUUGUGUGCCGC
  - 6: GUUGUGUGCCGC
  - 7: GUUGUGUGCCGC
  - 8: GUUGUGUGCCGC
  - 9: GUUGUGUGCCGC
  - 10: GUUGUGUGCCGC
  - 11: GUUGUGUGCCGC
  - 12: GUUGUGUGCCGC
  - 13: GUUGUGUGCCGC
  - 14: GUUGUGUGCCGC

---

determine how RT-Ec86 recognizes specific structures in the Ec86msr RNA, the SELEX method (27) was applied using two differently randomized RNA molecules derived from Ec86msr RNA. In the first RNA (RNA I), region I, encompassing a 10-base sequence from the C residue 5 bases upstream of the branching G residue to the U residue 4 bases downstream of the branching G residue, was randomized (boxed in Fig. 5; Table II); and in the second RNA (RNA II), the 11-base stem-loop structure of stem-loop b was randomized (boxed in Fig. 5; Table II). After six cycles of SELEX using RNA I with RT-Ec86, 9 sequences were determined as shown in Fig. 5A. Of 9 sequences, 6 (A-1–A-6) contained the highly conserved 3-base sequence AGC (shown in **bold**), which includes the branching G residue. The AGC sequence at the branching G is found in 10 of 11 msDNAs so far identified: Mx162 (8), Ec86 (4), Ec87 (2), Mx85 (9), Ec73 (7), Sa163 (10), Ec107 (1), Ec83 (3), MI162 (12), Ec78 (5), and Ec48 (21). In msDNA-Ec67, this sequence is replaced with AGA (2). The position of the branching G residue was quite accurately determined at the 5th base upstream of stem b together with the C residue at the 9th base, A at the 6th and C at the 4th residue. In two cases, the AGC sequence (A-7) or the AGA sequence (A-8) were closer to stem b by one base.

When RNA II was used with RT-Ec86, highly unique sequences were enriched as shown in Fig. 5B. Of 14 sequences determined, 10 (71%) were identical to the wild-type sequence (B-1–B-10). All of the remaining 4 sequences, despite base substitution to the wild-type sequence, still retain 4-base palindromic structures so that they are able to form a 4-base pair stem structure, except for the B-12 sequence, which has two mismatches in the stem. The wild-type sequence also forms a 4-base pair stem with a 3-base loop. Importantly, all but two (B-13 and B-14) have an identical loop structure consisting of 3 U residues. These results demonstrate that RT-Ec86 highly specifically recognizes the stem-loop structure found in the wild-type primer (structure b in Fig. 3A). It is interesting to note that the preferential sequences include the triple U sequence in the loop as well as two GU pairs rather than GC pairs in the stem.

The SELEX screening on RNA II was also performed with a hybrid RT, RT-Ec73/86 (construct a in Fig. 2B). Of 14 sequences determined, all but one (C-14 in Fig. 5C) again retained the ability to form a 4-base pair stem as in the case with RT-Ec86 (Fig. 5B). However, in most of these sequences, one or two base pairs were replaced with different base pairs. Such replacements occurred most frequently at the second base pair from the bottom; the UG base pair was replaced with either GU (C-1, C-5, C-6, C-7, and C-10), AU (C-8, C-9, and C-13), or CG (C-11 and C-12). It is interesting to note that no matter what kinds of replacements occurred, again there were always two GU pairs maintained in the stem structure except for C-4 and C-9, with one GU plus one AU pair, and C-10, with one GU plus two AU pairs. As for the loop structure, the UUU sequence preferred for RT-Ec86 was found in only 1 sequence (C-1). In 8 of 13 sequences, the first U residue was replaced with A, and only 3 sequences retained U at this position. Interestingly, the primer RNA for RT-Ec73 has an A residue at this position (see Fig. 5C). This A residue may be preferred because of the N-terminal domain of RT-Ec73/86 derived from RT-Ec73. Intriguingly, a U residue, without exception, occupied the second position; U is used for RT-Ec86, whereas G is used for RT-Ec73. The third position was mostly U, which is preferred by both RTs. These results confirmed the notion that the C-terminal domain including the Y region is responsible for the recognition of the specific secondary RNA structure.

**cDNA Synthesis with Ec86msd-msr48 RNA and Its Inhibition by Ec86msrΔα RNA**—In a cell-free system for msDNA synthesis with RT-Ec86, Ec86msdr-msd48 RNA (see Fig. 1B) can serve as a primer and a template (as shown in Fig. 6A). The cDNA synthesis was highly dependent upon the addition of dGTP and TTP when [α-32P]dTTP was used for labeling (lane 3). In the presence of only dATP and dCTP (lane 1) or dATP, TTP, and dCTP (lane 2) no msDNA synthesis was observed. The dependence of dCTP incorporation upon the addition of dGTP and TTP, but not upon the addition of dCTP, is due to the template structure of Ec86msd-msd48 RNA. 3' CAGUCUXXXX (1 base after the a1 sequence in Fig. 1B). Thus, the results obtained in Fig. 6A are consistent with the notion that the cDNA synthesis was initiated from the 2'-OH group of the branching G residue from which the cDNA, GTCAGA3, was elongated.

Next, we examined the inhibitory effect of Ec86msrΔα RNA on the msDNA synthesis with Ec86msd-msd48. As shown in
Fig. 6A, as more Ec86msrRNA RNA was added to the same reaction mixture used in Fig. 6A, mini-msDNA synthesis was inhibited more severely; when 5 pmol (5-fold) of Ec86msrRNA RNA was added to the reaction mixture, the msDNA synthesis was inhibited by more than 90% (lane 4). At 10 pmol of Ec86msrRNA RNA, the msDNA was almost completely blocked, indicating that Ec86msrRNA RNA indeed binds to RT-Ec86, competitively inhibiting the msDNA synthesis from Ec86msrRNA RNA.

During the course of these experiments, we noticed that there is another band migrating at a position lower than mini-msDNA, as indicated by arrowhead a in Fig. 6B. The synthesis of this band is reciprocal to the synthesis of mini-msDNA-Ec86, as its highest production was observed when 10 pmol of Ec86msrRNA RNA was used. Because this band production depends on the incorporation of [α-32P]dCTP, it was speculated that Ec86msrRNA RNA may serve as primer and template to produce the band a product. Indeed, the sequence 3′-GAGU5′ (from the 6th G residues to the 9th U residue from the 3′-end of the RNA) can form a duplex (see Fig. 6C), which allows the 2′-OH priming reaction at the branching G residue (circled) from the 11th U residue as template. Therefore, we tested the possible cDNA synthesis on the presence of Ec86msrRNA RNA only. Fig. 6D demonstrates that the cDNA synthesis hardly occurs with only dATP and [α-32P]dCTP (lane 1) or with dGTP, TTP, and [α-32P]dCTP (lane 3). However, with dATP, TTP, and [α-32P]dCTP, a reasonable amount of cDNA is produced (lane 2). This cDNA is likely to be a tetrancelotidte consisting of 5′-ATTCC5′ (see Fig. 6C). When all 4 bases are added, [α-32P]dCTP was extremely well incorporated, the oligonucleotide was further extended by several bases (lane 4). Again this is consistent with the secondary structure for E86msrRNA RNA proposed in Fig. 6C. These results provide additional strong support for the specific biochemical function of E86msrRNA RNA as essential for RT-Ec86 recognition.

**DISCUSSION**

Retrons are unevenly distributed in bacterial genomes. Retron-Mx162 from *M. xanthus* is the first retro-element ever found in the prokaryotes (1, 28), and it has been shown that all natural isolates of *M. xanthus* contain retron-Mx162, which produces msDNA-Mx162 (8). In contrast, in *E. coli* only a minor population of natural isolates contains retrons, which are highly diverse. Among seven retrons so far identified in *E. coli* (retron-Ec67 (2), retron-Ec86 (4), retron-Ec73 (7), retron-Ec107 (1), retron-Ec83 (3), retron-Ec78 (5), and retron-Ec48 (21)), there are virtually no sequence homologies in their msDNAs except for a few bases upstream of the branching G residue. RTs encoded by these retrons are also highly diverse, and RT-Ec86 and RT-Ec73, which are studied in the present paper, share only 22% identity. In other enterobacteria, retrons are also found in minor populations (11), indicating that the retro-elements in enterobacteria were only recently integrated in their genomes, after these species had been established, whereas the retro-elements in *M. xanthus* were integrated into its genome before this species was established, during its evolution.

The msr region encoding a stem-loop structure(s) immediately downstream of the branching G residue in msDNA has been shown to be essential for the priming reaction for the RT encoded by the same retron that contains the msr region (15). If the msr region is exchanged with another retron’s msr region, msDNA cannot be synthesized. Consistent with this finding, the present results clearly demonstrate that bacterial RTs bind quite tightly to their cognate msr-encoded RNA but not to RNA from another retron. Interestingly, of two stem-loop structures formed immediately downstream of the branching G residue in Ec86msrRNA RNA (see Fig. 3A), only the second stem-loop structure is recognized by RT-Ec86. At present, it is not clear what the role of the first stem-loop structure is in msDNA synthesis.

In the present paper, we have demonstrated that in the
RNA Recognition by Bacterial RT for 2'-OH Priming Reaction

specific recognition of the second stem-loop structure by RT-Ec86, the C-terminal region of the RT plays the major role. This region is assigned by sequence alignment with HIV-1 RT to encode the thumb domain of the RT molecule, consisting of β12, β13, β14, αG, αH, αI, and αJ (24, 25). Notably, this region is most diverse in bacterial RTs, which is consistent with the notion that this region is responsible for the recognition of the specific primer structure in the individual primer-temperate RNAs.

It is also interesting to note that the Y sequence contains two β-strands, β12 and β13, which form an element designated as the “primer grip” because of its proximity to phosphate joining the nucleotides at the primer end (24). Between these two β-strands is the VTGL sequence, which is highly conserved in bacterial RTs. It is tempting to speculate that the VTGL sequence indeed forms the loop between strands β12 and β13 and plays an essential role in orienting the branching G residue at the priming site. The enriching of the branching G residue by the SELEX method (Fig. 5) also suggests that the structure provides specific recognition at this site for a G residue.

Importantly, however, this G residue is not the 3'-end of the primer RNA for msDNA synthesis. The 3'-OH group of this G residue is already occupied by being connected to a long RNA strand by a 3'-5'-phosphodiester linkage. Therefore, the cDNA synthesis has to be initiated from the 2'-OH group of the G residue. The appropriate positioning of the G residue is essential for the priming reaction; it must be positioned in such a way that its 2'-OH group is close and exposed to the catalytic triad of three invariant Asp residues. For this configuration, the stem-loop structure downstream of the G residue is likely to play the essential role. On the basis of the three-dimensional structure of HIV-1 RT (24, 25), the structure of the priming complex of RT-Ec86 with Ec86msr-msd48 RNA (Fig. 1B) may be formed as follows; the primer-template a2-a1 duplex lays down at the foot of the thumb so that the branching G residue (circled in Fig. 1B) is positioned at the primer grip site. The RNA strand downstream of the G residue forms a stable complex with the thumb so that the thumb is sandwiched by the RNA primer-template duplex from the inside and by the stable stem-loop structure formed in the RNA strand downstream of the branching G residue from the outside. This would bind the primer RNA strand at the branching G residue to make its 2'-OH accessible for the priming reaction. We propose that the interaction between the stem-loop structure and the thumb determines the highly specific mechanisms of msDNA priming reaction. Of the four helical structures in the thumb (αG, αH, αI, and αJ), the αJ helical structure is located at the external part of the thumb and may be primarily responsible for the interaction with the stem-loop structure of the primer RNA. The bacterial RT’s αJ sequences, as aligned in Fig. 2A, are highly diverse and contain many basic residues, which may be important for the specific recognition with a specific RNA secondary structure (for example, as shown for the bacteriophage λ N peptide and box B RNA complex (29)). It remains to be answered why and how such custom-made thumbs have evolved to regulate the cDNA priming reaction in this highly specific manner.

It is also important to note that the X region seems to play an essential role in msDNA synthesis, as the exchanging of this region between RT-Ec73 and RT-Ec86 resulted in no msDNA synthesis (see Fig. 2B). This region is located between β5b and β6 strands, and it is assumed to be close to the joint of the thumb in bacterial RTs. This region seems also to be specific for individual RTs and to be involved in the formation of the priming complex, probably coordinating the initial interaction of primer-template RNA with RT. The determination of the three-dimensional studies of bacterial RTs will provide important insights into these questions.

Acknowledgments—We thank Dr. K. Yamanaka for a critical reading of the manuscript and A. Yashio for preparing Fig. 2A.

REFERENCES

1. Herzer, P. J., Inouye, S., and Inouye, M. (1992) Mol. Microbiol. 6, 345–354
2. Lampson, B. C., Sun, J., Hsu, M.-Y., Vallejo-Ramirez, J., Inouye, S., and Inouye, M. (1989) Science 243, 1033–1038
3. Lim, D. (1992) Mol. Microbiol. 6, 3531–3542
4. Lim, D., and Maas, W. K. (1989) Cell 56, 891–904
5. Lima, T. M., and Lim, D. (1997) Plasmid 38, 25–33
6. Mao, J. R., Inouye, S., and Inouye, M. (1997) J. Bacteriol. 179, 7865–7868
7. Sun, J., Inouye, M., and Inouye, S. (1991) J. Bacteriol. 174, 4171–4181
8. Dhnudale, A., Lampson, B., Furuchi, T., Inouye, M., and Inouye, S. (1987) Cell 51, 1105–1112
9. Inouye, S., Herzer, P. J., and Inouye, M. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 942–945
10. Furuchi, T., Dhnudale, A., Inouye, M., and Inouye, S. (1987) Cell 48, 47–53
11. Rice, S. A., and Lampson, B. C. (1993) J. Bacteriol. 175, 4250–4254
12. Rice, S. A., and Lampson, B. C. (1993) J. Bacteriol. 177, 37–45
13. Inouye, S., and Inouye, M. (1993) Curr. Opin. Genet. & Dev. 3, 713–718
14. Inouye, S., and Inouye, M. (1995) in Reverse Transcriptase, (Skalka, A., and Gaff, S., eds) pp. 391–410, Cold Spring Harbor Press, Cold Spring Harbor, NY
15. Shimamoto, T., Hsu, M.-Y., Inouye, S., and Inouye, M. (1993) J. Biol. Chem. 268, 2684–2692
16. Shimamoto, T., Inouye, M., and Inouye, S. (1995) J. Biol. Chem. 270, 581–588
17. Inouye, S., and Inouye, M. (1991) in Directed Mutagenesis, A Practical Approach (McPherson, M. J., ed) pp. 71–82, Oxford University Press, Oxford
18. Miller, J. H. (1972) A Short Course in Bacterial Genetics Cold Spring Harbor Laboratory, pp. 25.1–25.6, Cold Spring Harbor, NY
19. Vieira, J., and Messing, J. (1982) Gene 19, 259–268
20. Sanger, F., Nicklen, S., and Coulson, A. S. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
21. Mao, J.-R., Shimada, M., Inouye, S., and Inouye, M. (1995) J. Biol. Chem. 270, 19684–19687
22. Inouye, S., Hsu, M.-Y., Eagle, S., and Inouye, M. (1989) Cell 56, 709–717
23. Xiong, Y., and Eckehaus, T. (1990) EMBO J. 9, 3353–3362
24. Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Huiz, A., Hughes, S. H., and Arnold, E. (1995) Proc. Natl. Acad. Sci. U. S. A. 90, 6320–6324
25. Kohlsteadt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Stitz, T. A. (1992) Science 256, 1783–1789
26. Inouye, S., and Inouye, M. (1996) Virus Genes 11, 81–94
27. Tuerck, C., and Gold, L. (1990) Science 249, 505–510
28. Yee, T., Furuchi, T., Inouye, S., and Inouye, M. (1984) Cell 38, 203–209
29. Legault, P., Li, J., Mogridge, J., Ray, L. E., and Greenblatt, J. (1998) Cell 93, 289–299
30. Stitz, T. A., Smerdon, S. J., Jager, J., and Joyce, C. M. (1994) Science 266, 2022–2025