Reverse Transcriptases from Bacterial Retrons Require Specific Secondary Structures at the 5'-End of the Template for the cDNA Priming Reaction*

(Received for publication, August 19, 1992)

Tadashi Shimamoto, Mei-Yin Hsu, Sumiko Inouye, and Masayori Inouye

From the Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey at Rutgers, Piscataway, New Jersey 08854

Multicopy single-stranded DNA (msDNA) is a peculiar molecule consisting of a single-stranded DNA that is branched out from an internal G residue of an RNA molecule (msdRNA) via a 2',5'-phosphodiester linkage. The genetic unit required for msDNA synthesis is designated "retron" and consists of msr (a gene for msdRNA), msd (a gene for msDNA), and a gene for reverse transcriptase (RT) in a single operon. To date, four different msDNAs have been isolated from Escherichia coli. They do not share any primary sequences in either RNA or DNA. To elucidate the specificity of bacterial RT for msDNA synthesis, the msr-msd region from retron-Ec67 was introduced into E. coli cells producing RT-Ec67, or the msr-msd region from retron-Ec73 into E. coli cells producing RT-Ec73. In both cases, msDNA was not synthesized. However, when the msdRNA coding regions (msr) for retron-Ec67 and -Ec73 were mutually exchanged and the chimeric genes were introduced into E. coli cells producing either RT-Ec67 or RT-Ec73, it was thus found that msDNA was produced only when msr and RT were from the same retron. Requirement of the msr region for msDNA synthesis by RT was further investigated by mutations in the msr region for retron-Ec67. These analyses revealed that there is a strict requirement for specific primary sequences as well as the secondary structure in msdRNA. This finding is discussed in relation to the mechanism of the priming reaction of cDNA synthesis by eukaryotic retroviral RTs using tRNAs.

Recent discoveries of bacterial reverse transcriptases (RTs)\(^1\) related to eukaryotic RTs raise interesting questions as to the evolutionary origin of RT and the specificity of the priming reaction of cDNA synthesis (see Inouye and Inouye (1991, 1992a) for reviews). In particular, cDNA synthesized by bacterial RTs has been shown to be linked to the 2'-OH group of an A residue located in the 5'-end region of a template RNA. Thus, the same RNA transcript serves not only as a template, but also as a primer for cDNA synthesis. The product from this reaction is called multicopy single-stranded DNA (msDNA) and consists of a single-stranded DNA that branches out from a single-stranded RNA molecule (msdRNA) via a 2',5'-phosphodiester linkage. msDNA was originally found in myxobacteria (Yee et al., 1984; Dhundale et al., 1985, 1988; Furuichi et al., 1987a, 1987b) and later in a minor population of natural isolates of Escherichia coli. The branched G residue is circled.

*This work was supported in part by National Institutes of Health Grant GM44012. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

\(^{1}\)Supported by Human Frontier Science Program Organization Fellowship LT-225/92.

\(^{1}\)The abbreviations used are: RTs, reverse transcriptases; PCR, polymerase chain reaction; ddATP, dideoxyadenosine triphosphate; msDNA, multicopy single-stranded DNA; msdRNA, RNA molecule attached to msDNA; bp, base pair(s); HIV, human immunodeficiency virus.
reviews). The operon is composed of the coding region for msdrRNA (msr) and the coding region for the single-stranded DNA of msDNA (msd), followed by an open reading frame for RT. In the primary transcript from the operon, there are inverted repeats, one in the msr region near the 5'-end of the primary transcript (a2 in Fig. 1) and the other downstream of the msd region (a1 in Fig. 1). As a result, these inverted repeats allow the transcript to form a stable stem structure placing the branched G residue within the msr region at an end of the stem. The priming reaction of msDNA is considered to be initiated from the 2'-OH group of this G residue, followed by cDNA synthesis using the same RNA transcript as a template. The absolute requirement of the inverted repeats, a1 and a2, has been demonstrated by mutational analysis of these regions, indicating that the formation of the stem structure immediately upstream of the branched G residue is essential for msDNA synthesis (Hsu et al., 1989). In addition, the first base used for the priming reaction has been shown to be variable and to be determined by the complementarity of the base on the template strand corresponding to the 2685

**EXPERIMENTAL PROCEDURES**

**Bacterial Strains and Plasmids—**E. coli CL38 (Lerner and Inouye, 1980), which is an recA derivative of JM63, was used for all experiments in this study. Since this strain is a K12 derivative, it does not have a retron for msDNA production (Lampson et al., 1989). CL38 cells harboring plasmids were grown in L-broth (Miller, 1972) supplemented with appropriate antibiotics. pSP65 (Boehringer Mannheim) was used to close the 0.6-kilobase BamHI-HindIII fragment containing the msr-msd region of retron-Ec67, and the 0.7-kilobase HindII fragment for the msr-msd region of retron-Ec73 was cloned into pUC9 (Vieira and Messing, 1982). pUCK19 (Ohshima et al., 1992) and pBR322 were used for cloning polymerase chain reaction (PCR) fragments of the msr-msd region. For complementation analysis, plasmids carrying only RT genes were used. The 1059-bp XhoI fragment, which contains the RT-Ec73 gene, was cloned from retron-Ec73 (Sun et al., 1991) into the unique XhoI site of pGBl2pp3 (Hsu et al., 1992), a pSC101 derivative harboring the spectinomycin-resistant gene. One of the resulting plasmids carrying the RT gene in the same orientation as the lpp-lac promoter was designated pRT-73. pRT-73 (Hsu et al., 1992) contains the RT-Ec73 gene in pGBl2pp3. pG1-IEPS contains retron-Ec67 (Lampson et al., 1989) and pRT-70 containing the 0.8-kilobase HindII fragment containing the msr-msd region of retron-Ec73 (Sun et al., 1991) were used for PCR to construct chimeric msr-msd genes.

**PCR for Construction of Chimeric msr-msd Genes—**As shown in Figs. 2 (A and B), six primers (67a, 67b, 67c, 67d, 67e, and 67f) were individually synthesized to amplify the msr-msd region of retron-Ec67 into three fragments (67AB, 67CD, and 67EF) by PCR using the pG1-IEPS DNA (Lampson et al., 1989) as a template. Primer 67c consists of a 14-base a2 sequence of retron-Ec73 instead of the retron-Ec67 a2 sequence upstream of the G residue (circled) corresponding to the retron-Ec67 branched G residue (see Fig. 2B). Primer 67d consists of the 15-base 3'-end msr sequence from retron-Ec67 plus the 15-base sequence immediately downstream of the msr sequence from retron-Ec73 (see Fig. 2B). Similar sets of primers (73a and 73b, 73c and 73d, and 73e and 73f) were also synthesized as shown in Fig. 2 (A and B). These sets were used to synthesize three PCR fragments (73AB, 73CD, and 73EF) using p73-Hc0.7 as a template.

Fragments 67AB and 73AB contain one of the inverted repeat sequences (a2) from retron-Ec67 and -Ec73, respectively. The a2 sequence corresponds to the region immediately upstream of the branched G residue. Fragments 67AB and 73AB also contain the promoters for retron-Ec67 and -Ec73, respectively. Fragments 67EF and 73EF include the msDNA coding region (msd) and the other inverted repeat sequence (a1) from retron-Ec67 and -Ec73, respectively. Fragments 67CD and 73CD contain part of the msr regions (the sequences downstream of the branched G residue) from retron-Ec67 and -Ec73, respectively, that are flanked with the 15-bp retron-Ec67 and -Ec73 sequences, respectively. These 15-bp sequences are derived from immediately upstream and downstream of the part of the msr regions used (see Fig. 2A). From both retron-Ec67 and -Ec73 DNAs, three PCR fragments (see Fig. 2A) are individually synthesized by PCR and purified by polyacrylamide gel electrophoresis. In the second PCR, 67AB and 67EF were mixed with 73CD, and the amplification was carried out with the 67a and 67f primers (see Fig. 2, A and B) to construct the chimeric genes (msr73-msd67). During the second PCR, fragment 73CD annealed at both its ends to 67AB and 67EF with the 15-bp sequences. In the case of the chimeric msr67-msd73 gene, the three fragments (73AB, 67CD, and 73EF) were used as templates, and PCR was carried out with the 73a and 73f

---

**Fig. 2. Schematic diagram for synthesis of chimeric genes with msr67-msd73 and msr73-msd67 by PCR.** A, msr and msd and their orientations are shown by wide open and filled arrows on thin lines, which represent retron-Ec67, and/or -Ec73, respectively. The branched G residue is circled. Small arrows with numbers and letters are primers used for PCR and are positioned at the locations having DNA sequences to complement primer sequences. Wavy and thick lines attached at the ends of 67c, 67d, 73c, and 73d indicate base sequences derived from other retrons (see text for details). B, nucleotide sequences of primers used for PCR. Boldface and lightface letters represent nucleotides of retron-Ec67 and -Ec73, respectively. Undercase letters represent nucleotides added to create restriction sites, which are underlined.

---

**PCR for Construction of Chimeric msr-msd Genes—**As shown in Figs. 2 (A and B), six primers (67a, 67b, 67c, 67d, 67e, and 67f) were individually synthesized to amplify the msr-msd region of retron-Ec67 into three fragments (67AB, 67CD, and 67EF) by PCR using the pG1-IEPS DNA (Lampson et al., 1989) as a template. Primer 67c consists of a 14-base a2 sequence of retron-Ec73 instead of the retron-Ec67 a2 sequence upstream of the G residue (circled) corresponding to the retron-Ec67 branched G residue (see Fig. 2B). Primer 67d consists of the 15-base 3'-end msr sequence from retron-Ec67 plus the 15-base sequence immediately downstream of the msr sequence from retron-Ec73 (see Fig. 2B). Similar sets of primers (73a and 73b, 73c and 73d, and 73e and 73f) were also synthesized as shown in Fig. 2 (A and B). These sets were used to synthesize three PCR fragments (73AB, 73CD, and 73EF) using p73-Hc0.7 as a template.

Fragments 67AB and 73AB contain one of the inverted repeat sequences (a2) from retron-Ec67 and -Ec73, respectively. The a2 sequence corresponds to the region immediately upstream of the branched G residue. Fragments 67AB and 73AB also contain the promoters for retron-Ec67 and -Ec73, respectively. Fragments 67EF and 73EF include the msDNA coding region (msd) and the other inverted repeat sequence (a1) from retron-Ec67 and -Ec73, respectively. Fragments 67CD and 73CD contain part of the msr regions (the sequences downstream of the branched G residue) from retron-Ec67 and -Ec73, respectively, that are flanked with the 15-bp retron-Ec67 and -Ec73 sequences, respectively. These 15-bp sequences are derived from immediately upstream and downstream of the part of the msr regions used (see Fig. 2A). From both retron-Ec67 and -Ec73 DNAs, three PCR fragments (see Fig. 2A) are individually synthesized by PCR and purified by polyacrylamide gel electrophoresis. In the second PCR, 67AB and 67EF were mixed with 73CD, and the amplification was carried out with the 67a and 67f primers (see Fig. 2, A and B) to construct the chimeric gene (msr73-msd67). During the second PCR, fragment 73CD annealed at both its ends to 67AB and 67EF with the 15-bp sequences. In the case of the chimeric msr67-msd73 gene, the three fragments (73AB, 67CD, and 73EF) were used as templates, and PCR was carried out with the 73a and 73f

---

**Fig. 2. Schematic diagram for synthesis of chimeric genes with msr67-msd73 and msr73-msd67 by PCR.** A, msr and msd and their orientations are shown by wide open and filled arrows on thin lines, which represent retron-Ec67, and/or -Ec73, respectively. The branched G residue is circled. Small arrows with numbers and letters are primers used for PCR and are positioned at the locations having DNA sequences to complement primer sequences. Wavy and thick lines attached at the ends of 67c, 67d, 73c, and 73d indicate base sequences derived from other retrons (see text for details). B, nucleotide sequences of primers used for PCR. Boldface and lightface letters represent nucleotides of retron-Ec67 and -Ec73, respectively. Undercase letters represent nucleotides added to create restriction sites, which are underlined.
primers (see Fig. 2, A and B). The wild-type gene (msr67-msd67) was produced by one-step PCR using 67a and 67f as primers with retron-Ec67 as a template. Similarly, the msr73-msd73 gene was produced with 73a and 73f as primers using retron-Ec73 as a template. The resulting PCR fragments (67AB-73CD-67EF (msr67-msd67)) and (73AB-67CD-73EF (msr73-msd73)) and the wild-type PCR fragments (67AB-73CD-67EF (msr67-msd67)) and 73AB-67CD-73EF (msr73-msd73) were cloned into pUCK19 (Ohshima et al., 1992). They were designated pUCK(73-67), pUCK(67-73), and pUCK(73-73), respectively.

These msr-msd fragments in pUCK19 were under the control of the lac promoter from pUCK19 and also their own promoters. To avoid the transcription from the lac promoter, the chimeric msr-msd regions were cleaved from the clones by EcoRI and BamHI digestion, and the resulting fragments were recloned into pBR322. In the resultant plasmids, the msr-msd regions were able to be transcribed only from their own promoters. These plasmids were designated pBR(67-67) for msr67-msd67, pBR(73-73) for msr73-msd73, pBR(67-73) for msr67-msd73, and pBR(73-67) for msr73-msd67.

For PCR, 0.1 μg of template DNA (pCl-1EP6 or p73-Hc0.7) was added to 100 μl of PCR mixture containing 50 mM Tris-HCl (pH 9.0), 1.5 mM MgCl2, 20 mM ammonium sulfate, 0.2 mM each dNTP, and 50 pmol of each oligonucleotide primer (Fig. 2B). After heat denaturation of the reaction mixture prepared as described above (100 °C, 3 min) followed by quick chilling on ice, 3 units of Hot Tub™ DNA polymerase (Amersham Corp.) were added to the reaction mixture. PCR was performed for 25 cycles in a programmable thermal controller (MJ Research, Inc.) under the following conditions: 94 °C, 1 min; 55 °C, 2 min; and 72 °C, 2 min. The resulting PCR fragments were purified by polyacrylamide gel electrophoresis and cloned into the Smal site of pUCK19. The cloned PCR fragments were sequenced by the dye-oxide chain termination method (Sanger et al., 1977) with the modified T7 DNA polymerase (Sequenase, United States Biochemical Corp.).

Preparation of RNA and Detection of msDNA by RT Extension—Total RNA fraction was prepared from exponentially growing cells by the method of Chomzynski and Sacchi (1987). RT extension reaction was carried out using avian myeloblastosis virus RT (Molecular Genetic Resources) and [α-32P]dCTP as described previously (Lampson et al., 1989). The labeled samples were electrophoresed on an 8 M urea, 4% polyacrylamide gel.

Labeling and Sequencing of msDNA—msDNA was prepared by the alkaline SDS lysis method used for plasmid DNA (Birnboim and Doly, 1979). For labeling and sequencing of msDNA, msDNA was prepared from a 300-ml overnight culture of CL83 cells harboring a RT gene and a fused msr-msd region. The msDNA fractions were treated with 80 μg/ml RNase A for 15 min at 37 °C and subjected to 10% polyacrylamide gel electrophoresis. After staining the gel with ethidium bromide, the bands corresponding to msDNA were cut out, and the msDNA was electroeluted. About 0.3 μg of msDNA was labeled with [α-32P]dATP and terminal deoxynucleotidyltransferase (International Biotechnologies, Inc.) as follows. The labeled msDNA was added to the labeling mixture (100 μl) containing 0.14 mM potassium cacodylate (pH 7.2), 30 mM Tris, 1 mM CoCl2, and 0.2 mM β-mercaptoethanol. After heat denaturation (100 °C, 5 min) followed by quick chilling on ice, [α-32P]dATP and 30 units of terminal deoxynucleotidyltransferase were added, and the final reaction mixture was incubated at 37 °C for 60 min. The reaction was stopped by adding 25 mM EDTA and the same volume of isopropyl alcohol, and then the msDNA was precipitated by centrifugation. The concentrated samples were then applied to a preparative sequencing gel (6% polyacrylamide gel in 8 M urea). After electrophoresis, each band was cut out, and the labeled msDNAs were eluted from these bands with elution buffer containing 0.3 M sodium acetate (pH 7.0), 0.1% SDS, and 1 mM EDTA. After incubation at 37 °C overnight, the eluted msDNAs were precipitated with ethanol. The DNA sequences of the msDNAs were determined by the method of Maxam and Gilbert (1980).

Results

RESULTS

RTs Are Not Exchangeable between Retron-Ec67 and -Ec73—In a retron, the msr-msd region and the downstream RT gene are under the control of a single promoter located upstream of the msr-msd region and produce msDNA (Fig. 1) (Lampson et al., 1989). However, the RT gene from retron-Ec67 can be expressed under a separate promoter to complement the function of the msr-msd region of retron-Ec67 to synthesize msDNA-Ec67 (Hsu et al., 1992). This was achieved by double transformation of cells with p67-BH0.6 for the msr-msd region of retron-Ec67 in a high-copy number plasmid (pSP65) and with pRT-67 for the RT gene from retron-Ec67 in a low-copy number plasmid (pGB2ppp+) under the lpp-lac promoter (Hsu et al., 1992). The production of msDNA by the cells harboring both p67-BH0.6 and pRT-67 was clearly demonstrated by the RT extension method as shown in Fig. 3 (lane 2). The pattern of the RT extension products (lane 2)
was almost identical to that of the products from the cells harboring pCl-1EP5 that contained the entire retron-Ec67 (lane 1). The size of the major product was ~107 bases in length (including an extra 4-base RNA attached at the 5′-end of DNA) (as shown by dot in both lanes 1 and 2).

A similar experiment was carried out for retron-Ec73 (Sun et al., 1991) with p73-Hc0.7 containing the msr-msd region of retron-Ec73 in pUC9 and with pRT-73 for the RT gene from retron-Ec73 under the lpp-lac promoter in pGB2lpp−5. p73-Hc0.7 and pRT-73 were able to complement each other to produce msDNA-Ec73 (Fig. 3, lane 5) as in the case of the entire retron-Ec73 in p23S3.5 (Sun et al., 1991) (also see Fig. 3, lane 4). The size of the major product (shown by a dot) was 132 bases in length (including an extra 3-base RNA attached at the 5′-end of DNA) for both lanes 4 and 5. However, neither the p67-BH0.6 and pRT-73 complementation (lane 3) nor p73-Hc0.7 and pRT-67 (lane 6) was able to produce msDNA.

**Requirement of RT and msr Region from Same Retron—**
The results described above suggest that there is a specific requirement(s) within the structure of the msr-msd region for an individual RT to synthesize msDNA. RT may require specific primary and/or secondary structures within either msr or msd or both for msDNA synthesis. To examine this possibility, we constructed chimeric msr-msd genes between plasmids used for the experiments were pBR(67-73) for msr67-msd67 and pBR(73-67) for msr73-msd67 (see Fig. 4A). pBR(67-67) for msr67-msd67 and pBR(73-73) for msr73-msd73 were also used as controls.

Fig. 4 (C and D) shows the putative structures of the chimeric msDNAs expected from the structures of the chimeric msr-msd regions as well as the structures of msDNA-Ec67 (Fig. 4A) (Lampson et al., 1989) and msDNA-Ec73 (Fig. 4B) (Sun et al., 1991). The putative secondary structures of the primary transcripts from the chimeric msr-msd regions are also shown in Fig. 5A for msr67-msd73 and in Fig. 5B for msr73-msd67 on the basis of the biosynthetic mechanism of msDNA (Fig. 1) (see Dhundale et al. (1987); also see reviews by Inouye and Inouye (1991, 1992a, 1992b)). The synthesis of msDNA is initiated from the 2′-OH group of the G residue at the end of the a1-a2 stem structure (circled in Fig. 5). For the first base (T residue) linked to the G residue by a 2′,5′-phosphodiester linkage, the A residue at position 126 for msr67-msd73 and the A residue at position 136 for msr73-msd67 are used as templates. msDNA is then elongated from left to right by RT using the RNA strand as templates.

The msDNA production by the plasmids constructed as described above were examined in E. coli cells producing either RT from retron-Ec67 (RT-Ec67) or from retron-Ec73 (RT-Ec73). For this purpose, E. coli CL83 (Lerner and Inouye, 1990) was first transformed with pBR(67-73), pBR(73-73), pBR(67-67) or pBR(73-73). The transformed cells were then retransformed with the second plasmids, pRT-67 for RT-Ec67 or pRT-73 for RT-Ec73. msDNAs were isolated from the doubly transformed cells by the alkaline SDS method, treated with RNase A, and separated on a 10% polyacrylamide gel. As shown in Fig. 6, msDNA production was observed only in the following combinations: msr67-msd67 and RT-Ec67 (lane 2), msr73-msd73 and RT-Ec73 (lane 8), msr67-msd73 and RT-Ec67 (lane 4), and msr73-msd67 and RT-Ec73 (lanes 10 and 12). The results are summarized in Table 1, and it is evident that msDNAs were synthesized only when msr and RT were from the same retron.

It should be noted that msDNAs without msdRNA attached migrate abnormally on nondenaturing gels because of their extensive secondary structures as can be seen in msDNA-Ec67 (Fig. 6, lane 6) and msDNA-Ec73 (lane 8). In particular, msDNA-Ec73 has an extensive secondary structure (Fig. 4B) (Sun et al., 1991), which is considered to cause faster migration on the gel. A minor band that migrated faster than the major band for msDNA-Ec73 (lane 8) was found to be processed between residues 6 and 7 of msDNA from its DNA sequence (data not shown).

The production of msDNA-(73-67) (Fig. 6, lane 10) was further confirmed using a pUCK19 vector containing the
Fig. 5. Proposed secondary structures of primary transcripts of msr-msd regions from pBR(67-73) and pBR(73-67). A, structure of the msr67-msd73 transcript from pBR(67-73); B, structure of the msr73-msd67 transcript from pBR(73-67). The branched G residue is circled. The inverted repeat sequences (a1 and a2) that form a stem structure in the primary RNA transcript (Hsu et al., 1992) are indicated by the arrows. Arrowheads indicate positions where DNA synthesis terminates.

Fig. 6. Production of msDNA by cells harboring various combinations of msr and msd regions from retron-Ec67 and -Ec73 with RT-Ec67 or RT-Ec73. msDNAs were isolated by the alkaline SDS lysis method (Birnboim and Doly, 1979) from 5-ml late-log cultures. Samples were treated with RNase A (80 μg/ml) and analyzed by 10% polyacrylamide gel electrophoresis followed by staining with ethidium bromide. The origin of the RT, msr, and msd genes used for each lane is shown on the top of each lane; 67 and 73 indicate the origins from retron-Ec67 and -Ec73, respectively. Lane M, pRR322 digested with HaeII as molecular weight marker; numbers to the left represent the molecular sizes in bases. Arrowheads in lane 4 indicate bands H (upper) and L (lower) of msDNA-(67-73) (see text), and the bracket in lane 12 indicates the region extracted for determination of the DNA sequence of msDNA-(73-67).

Fig. 7. Analysis of labeled msDNA-(67-73) and msDNA-(73-67). msDNAs from pBR(67-73) and pRT-67 (Fig. 6, lane 4) and from pUCK(73-67) and pRT-73 (Fig. 6, lane 12) were separated on a preparative gel, extracted, and labeled at the 3'-end with [α-32P]ddATP and terminal deoxynucleotidyltransferase. Lanes 1 and 2, bands H and L of msDNA-(67-73); lane 3, msDNA-(73-67). Bands in each lane are numbered starting with the longest fragment first 1. Lane M, the same molecular weight markers as those described for Fig. 3.

TABLE I
Summary of msDNA production from wild-type and chimeric retrons

|        | RT-Ec67 | RT-Ec73 |
|--------|---------|---------|
| msr    | 67 73   | 67 73   |
| msd    | 67 73   | 67 73   |
| msDNA production* | + - + - | + - + - |

* From Fig. 6.

msr73-msd67 region (pUCK(73-67)). A major band together with a few minor bands were clearly produced when the vector was cotransformed with pRT-73 (lane 12), but not with pRT-67 (lane 11). These bands in lane 12 were found to be a result of different msDNA terminating sites as determined from their DNA sequences (see the below).

The production of the chimeric msDNA molecules was further confirmed for each msDNA preparation used in Fig. 6 by the RT extension method. Because of the RNA-DNA hybrid structures at their 3'-end (see Fig. 4), msDNAs serve as a substrate for retroviral RTs (Lampson et al., 1989; Lim and Maas, 1989). [α-32P]dCTP and three nonradioactive dNTPs (dGTP, dATP, and dTTP) were incorporated by RT from avian myeloblastosis virus for msDNA-Ec67 (Fig. 3, lanes 1 and 2) and for msDNA-Ec73 (lanes 4 and 5). Similarly, msDNAs were detected for msr67-msd73 with RT-Ec67 and for msr73-msd67 with RT-Ec73 (data not shown). However,
msDNAs were not detected for msr67-msd73 with RT-Ec73 and for msr73-msd67 with RT-Ec67 (data not shown). These results are consistent with the results in Fig. 6, demonstrating that msDNA is produced only when RT and the msr region are from the same retron.

**DNA Sequences of Chimeric msDNA Molecules**—To determine the DNA sequences of the chimeric msDNAs, the msDNAs from cells harboring pBR(67-73) and pRT-67 and pUCK(73-67) and pRT-73 were isolated by preparative polyacrylamide gel electrophoresis. For msDNA-(67-73), the bands corresponding to the upper and lower bands indicated by the arrowheads in Fig. 6 (lane 4) were individually isolated and designated bands H and L, respectively. For msDNA-(73-67), the bands indicated by the bracket in Fig. 6 (lane 12) were eluted from the gel. These msDNAs were labeled at the 3'-end with [\(\alpha\]P]dATP using terminal deoxynucleotidyltransferase as described under “Experimental Procedures.” As shown in Fig. 7, the labeled products were separated into a few bands by gel electrophoresis on 8 M urea-polyacrylamide gel. The DNA fraction from msDNA-(67-73)H resulted in two bands, H1 and H2 (lane 1); msDNA-(67-73)L in two bands, L1 and L2 (lane 2); and msDNA-(73-67) in three bands (lane 3). All the bands were again individually extracted from the gel, and their DNA sequences were determined by the method of Maxam and Gilbert (1980).

As shown in Fig. 8A, band H1 (msDNA-(67-73)H1) consists of 111 bases with the 3'-end DNA sequence (-CCAACAAAT-3') completely complementary to the RNA sequence immediately downstream of the branched G residue as predicted for the structure of msDNA-(67-73) (Fig. 4C). The entire 111-base DNA sequence can be read in Fig. 8A. This reveals that the DNA structure consists of the 5'-end 68-base sequence from msDNA-Ec73 (bases 1–68 of msDNA-Ec73) (see Fig. 4B), the 7-base sequence corresponding to the RNA-DNA hybrid region of msDNA-Ec67 (bases 61–67 of msDNA-Ec67) (see Fig. 4A and box in Fig. 8A), and the 36-base sequence complementary to msdRNA from msDNA-Ec67 (bases 16–51 of msRNA-Ec67) (see Fig. 4A). This result completely agrees with the predicted structure of msDNA-(67-73) shown in Fig. 4C. The DNA sequence of msDNA-(67-73)H2 (Fig. 7, lane 1, band 2) was found to be shorter by 1 base at the 3'-end than that of msDNA-(67-73)H1. Otherwise, the DNA sequence of msDNA-(67-73)H2 was identical to that of msDNA-(67-73)H1 (data not shown).

msDNA-(67-73)L1 and -L2 were found to consist of 75 and 74 bases, respectively. The entire DNA sequence of msDNA-
(67-73)L2 is shown in Fig. 8B and is composed of the 68-base sequence identical to that of residues 1-68 of msDNA-Ec73 (Fig. 4B) plus the 6-base sequence at the 3'-end identical to the 3'-end sequence of msDNA-Ec67 (Fig. 4A). The DNA sequence of msDNA-(67-73)L1 was longer by a C residue at the 3'-end than that of msDNA-(67-73)L2; otherwise, the DNA sequences of msDNA-(67-73)L1 and -L2 were identical (data not shown). Thus, the structure of msDNA-(67-73)L1 is identical to that predicted for msDNA-(67-73) (Fig. 4C). These results indicate that termination of msDNA for msDNA-(67-73) was less stringent than that for msDNA-Ec73: sometimes termination occurs earlier by 1 base or cannot occur at the regular site to extend the DNA strand almost all the way to the branched G residue. These termination sites are indicated by the arrowheads on the primary transcript for msr67-msd73 in Fig. 5A.

The DNA sequence of msDNA-(73-67)-band 3, the major band of the three bands in Fig. 7 (lane 3), is shown in Fig. 8C. It consists of the 60-base sequence identical to the sequence from residues 1 to 60 of msDNA-Ec67 (Fig. 4A) plus the 6-base sequence at the 3'-end (Fig. 8C). This 6-base sequence is identical to the 5-base sequence of the 3'-end of msDNA-Ec73 (Fig. 4B) plus an additional A residue at the 3'-end. This 3'-end A residue most likely resulted from the extension of msDNA synthesis by 1 base at the 3'-end, which adds an A residue (see Fig. 4, A and D). Thus, the structures of msDNA-(73-67)-band 2 and-band 1 are likely to have further extensions at the 3'-end (2 A residues and 3 A residues for band 2 and 3 DNAs, respectively). These termination sites are indicated by the arrowheads on the primary transcript for msr73-msd67 in Fig. 5B. Otherwise, the structure of msDNA-(73-67) is identical to that predicted in Fig. 4D.

Requirement of Secondary Structure of msdRNA for msDNA Synthesis—The results described above demonstrated that the msr region is essential for msDNA synthesis. msdRNA coded by the msr region contains extensive inverted sequences that are believed to form stable secondary structures as shown in Fig. 4. Thus, we next examined the importance of the putative secondary structures of msdRNA for msDNA synthesis using msDNA-Ec67 (Fig. 4A). Using PCR, various mutations were incorporated in the putative secondary structure of msDNA-Ec67 as depicted in Fig. 9.

The 2-base substitution mutation (67r-1) that changed 2 U residues at positions 18 and 19 to 2 C residues did not alter the amount of msDNA produced (Table II). This mutation is believed to tighten the stem structure by changing two U-G pairs to two C-G pairs (Fig. 9). The G to C substitution mutation at position 24 (67r-2) that destroyed the G-C base pair in the stem (see Fig. 9) resulted in a substantial reduction of msDNA production (Table II). On the other hand, the resumption of the base pair in the same position of the stem by altering the C residue at position 38 to a G residue in the msr region containing mutation 67r-2 (67r-3) (see Fig. 9) restored the normal level of msDNA production (Table II). This result indicates that msdRNA indeed forms the secondary structure proposed in Fig. 9.

We also constructed three deletion mutations: deletion of the U residue at position 25 (67r-4), deletion of 2 U residues at positions 25 and 26 (67r-5), and deletion of 2 U residues of 67r-5 together with 2 A residues at positions 36 and 37 (67r-6). The first two mutations cause 1 and 2 mismatched A residues in the stem for mutations 67r-4 and 67r-5, respectively (see Fig. 9), whereas mutation 67r-6 shortens the stem structure by 2 base pairs. None of these mutations were able to produce msDNA (Table II). Furthermore, the flipping of the entire inverted repeat sequences in the msr region (67r-7) also resulted in no msDNA production (Fig. 9 and Table II). The msdRNA with this mutation still maintains a stem structure of the same length as that of the wild-type msdRNA and of a similar stability as that of the wild-type msdRNA (Fig. 9).

The 67r-7 mutation flips not only the stem structure, but also the sequence of the loop from AAU to UAA. The conversion of this UAA sequence of the 67r-7 mutation back to AAU (67r-8) (see Fig. 9) did not restore msDNA production (Table II). The RNA sequence of the loop was also found to be essential for msDNA synthesis; the AAU sequence of the wild-type loop to AGU (67r-9), UAA (67r-10), and CCA (67r-11) resulted in no msDNA production (Fig. 9 and Table II). The opening of the stem structure on the top of the stem by altering the G-C base pair to the G-G mismatch (67r-12) also resulted in no msDNA production (Table II). This mutation changed the 3-base loop of msdRNA-Ec73 to a 7-base loop by reducing the stem length by 2 base pairs.

Taken together, the mutational analysis of the secondary structure reveals that there is a very stringent requirement of the secondary structure in msdRNA for msDNA synthesis.

DISCUSSION

This study demonstrates that the specificity of msDNA synthesis by bacterial RTs is determined by the structure of the msr region of a retron; between two retroons, the msd regions encoding the single-stranded DNA are exchangeable, whereas the msr regions encoding msdRNAs are not. To initiate msDNA synthesis, the 3'-OH group of an G residue located at the 5'-end region of a template RNA, a stem structure has to be formed between two inverted repeat sequences within a template, one immediately upstream of the cDNA initiation site (the branched G residue) at the 5'-end region of a template (a2 sequence) (see Fig. 5, A and B) and the other at the 3'-end region of the same template (a1 sequence). It has been demonstrated that formation of the stem structure is essential, although no requirement for specific sequences immediately upstream of the branched G residue in the stem structure is needed (Hsu et al., 1989).
is also confirmed by this study since the stem structure from retron-Ec73 can be used for msDNA synthesis by RT-Ec67.

For the initiation of retroviral cDNA synthesis, a specific tRNA is required for the priming reaction. It appears that a retroviral RT generally recognizes a specific tRNA to unfold it, allowing the tRNA to hybridize to the primer-binding site (Varmus and Swanstron, 1985) that is formed in the viral RNA and is complementary to the 3' end of the primer tRNA. It has been suggested that the interaction between template RNA forming secondary structures and primer tRNA is important for the initiation of reverse transcription in all type C and D retroviruses (Aiyar et al., 1992). The cDNA priming reaction is initiated from the A residue of the 3' end CCA sequence of the tRNA. In the msDNA priming reaction, one can consider that the A1 sequence corresponds to the retroviral primer-binding site sequence and the A2 sequence to the 3'-end sequence of the primer tRNA. However, unlike the tRNA primer, the mismatched G residue immediately downstream of the A2 sequence is used for the msDNA priming reaction (see Fig. 5, A and B). Furthermore, in contrast to the retroviral tRNA primer, there is no requirement for a secondary structure upstream of the A2 primer sequence. Instead, a specific RNA sequence (msdRNA) downstream of the A2 sequence was found to be required for msDNA synthesis. The mutational analysis of the msDNA for msDNA-Ec67 revealed that the secondary structure formed by the msDNA is essential for the production of msDNA-Ec67.

The secondary structure formed in msDNA-Ec67 consists of a 12-bp stem structure with a 3-base loop. We found that there were hardly any mutations in the secondary structure that did not affect msDNA production, indicating that RT-Ec67 stringently requires not only the 12-bp stem structure of the unique sequence, but also the AUA loop sequence for the synthesis of msDNA-Ec67. This stringent requirement for the secondary structure is most likely to be necessary for the msDNA priming reaction from the 2'-OH of the G residue at the 3'-end of the A2 sequence (see Fig. 5A). Note that like retroviral RTs, RT-Ec67 is able to elongate a DNA chain using a non-specific RNA or DNA template with the use of a non-specific RNA or DNA primer in a cell-free system (Lampson et al., 1990).

Our finding that both bacterial and retroviral RTs strictly require a specific secondary structure (see Fig. 8) for the cDNA priming reaction is intriguing particularly in light of the recent elucidation of the three-dimensional structure of HIV-1 RT (Aurhold et al., 1992; Kohlstaedt et al., 1992). Since both RTs are evolutionarily related (Xiong and Eckbush, 1990), the three-dimensional structural domains of bacterial RTs are considered to be also similar to that of HIV-1 RT, in which the terminus of the carboxyl-terminal RNAase H domain. This sequence, designated the connection domain, forms a subdomain that plays essential roles not only for the p51 conformation, but also for the p51 and p66 subunit interaction. Except for RT-Ec67, all the other known bacterial RTs lack the connection domain (see Inouye and Inouye (1992a, 1992b) for reviews), which raises an interesting question as to whether bacterial RTs form a dimer and, if so, how the structure equivalent to p51 is formed without the connection domain.

Acknowledgments—We thank Drs. Monica Roth, Lynne Vales, and Joseph Dougherty for critical reading of this manuscript.

REFERENCES

Aiyar, A., Cobrinik, D., Ge, Z., Kung, H.-J., and Leis, J. (1992) J. Virol. 66, 2461-2474.
Arnold, E., Jacobo-Molina, A., Nanni, R. G., Williams, R. L., Lu, X., Ding, J., Clark, A. D., Jr., Zhang, A., Ferris, A. L., Clark, P., Hizi, A., and Hughes, S. H. (1992) Nature 357, 85-88.
Birboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156-159.
Dundasle, A. R., Furuchii, T., Inouye, S., and Inouye, M. (1985) J. Bacteriol. 164, 914-917.
Dundasle, A., Lampson, B., Furuchii, T., Inouye, M., and Inouye, S. (1987) Cell 51, 1105-1112.
Dundasle, A., Inouye, M., and Inouye, S. (1988) J. Biol. Chem. 263, 9055-9058.
Furuchii, T., Dundasle, A., Inouye, M., and Inouye, S. (1987a) Cell 48, 47-53.
Herrera, P. J., Inouye, S., and Whittam, T. S. (1990) J. Bacteriol. 172, 6175-6181.
Herrera, P. J., Inouye, S., and Inouye, M. (1992) Mol. Microbiol. 6, 345-354.
Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Gene (Amst.) 77, 11-39.
Inouye, M., and Inouye, S. (1989) J. Biol. Chem. 264, 6218-6219.
Inouye, M., and Inouye, S. (1992a) J. Bacteriol. 174, 2419-2424.
Inouye, M., and Inouye, S. (1992b) in Reverse Transcriptase (Goff, S., and Skolka, A., eds) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, in press.
Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992) Science 256, 1783-1790.
Lampson, B. C., Sun, J., Hsu, M.-Y., Vallejo-Ramirez, J., Inouye, S., and
Y-OH Priming Reaction for msDNA Synthesis

Lampson, B. C., Viswanathan, M., Inouye, M., and Inouye, S. (1990) J. Biol. Chem. 265, 8490-8496
Lerner, C. G., and Inouye, M. (1990) Nucleic Acids Res. 18, 4631
Lim, D., and Maas, W. N. (1989) Cell 54, 991-994
Maxam, A., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Ohshima, A., Inouye, S., and Inouye, M. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 1016-1020
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
Sun, J., Inouye, M., and Inouye, S. (1991) J. Bacteriol. 173, 4171-4181
Varmus, H., and Swanstrom, R. (1985) in RNA Tumor Viruses (Weiss, N., Teich, H., Varmus, H., and Coffin, J., eds) Vol. 2, pp. 369-512, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Vieira, J., and Messing, J. (1982) Gene (Amst.) 19, 259-268
Xiong, Y., and Eickbush, T. H. (1990) EMBO J. 9, 3353-3362
Yee, T., Furuichi, T., Inouye, S., and Inouye, M. (1984) Cell 38, 203-209