Complex Formation between a Putative 66-Residue Thumb Domain of Bacterial Reverse Transcriptase RT-Ec86 and the Primer Recognition RNA*

Masayori Inouye‡§, Haiping Ke†, Arikо Yashio‡, Kunitoshi Yamanaka‡†, Hirofumi Nariya‡‡, Tadashi Shimamoto§, and Sumiko Inouye‡

From the ‡Department of Biochemistry, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, the §Graduate School of Biosphere Science, Hiroshima University, 1-4 Kagamiyama, Higashi-Hiroshima 739-8528, Japan, and the ¶Division of Molecular Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, 4-24-1 Kuhonji, Kumamoto 862-0976, Japan.

Reverse transcriptases (RT) are found in a minor population of Escherichia coli and are responsible for the synthesis of multicopy single-stranded DNA. These RTs specifically recognize 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). Here, we purified the 66-residue, C-terminal fragment of RT-Ec86, RT from E. coli, which is responsible for the synthesis of multicopy single-stranded DNA-Ec86. This fragment, RT-Ec86-(255–320), was found to consist mainly of α-helical structures on the basis of its CD spectrum, which is consistent with the prediction of this region as the thumb domain from the structural alignment of RT-Ec86 with human immunodeficiency virus-1 RT. RT-Ec86-(255–320) was able to bind to a 28-base synthetic RNA consisting of the 5′-end single-stranded RNA containing the branching G residue and the recognition stem-loop structure in the RT-Ec86 primer-template RNA with a Kd value of 5 × 10^-8 M. By stepwise shortening of the 5′-end single-stranded region of the RNA, RT-Ec86-(255–320) was found still to be able to form a stable complex with only the stem-loop structure consisting of an 8-base stem and a 3-base loop. In this stem-loop structure, the UUU loop was essential for the complex formation. RT-Ec73-(251–316) from another E. coli RT could not bind to the 28-base RNA for RT-Ec86 but could bind to its own stem-loop structure having a 3-base AGU loop. These results support the notion that the highly diverse C-terminal regions of bacterial RTs play an important role in recognizing their own specific primer-template RNA structure for the cDNA priming reaction.

Received for publication, July 26, 2004, and in revised form, September 14, 2004
Published, JBC Papers in Press, September 14, 2004, DOI 10.1074/jbc.M408462200

* This work was supported by National Institutes of Health Grant 5 RO1 GM63853. 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.
‡ To whom correspondence should be addressed: Dept. of Biochemistry, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854, Tel.: 732-235-4115; Fax: 732-235-4559; E-mail: inouye@umdnj.edu.
§ The abbreviations used are: RT, reverse transcriptase; msDNA, multicopy single-stranded DNA; SELEX, systematic evolution of ligands by exponential enrichment; HIV, human immunodeficiency virus.

† This paper is available on line at http://www.jbc.org

Printed in U.S.A.
structure having the AGU loop but not the RNA molecules for RT-Ec86-(255–320). The present results support the notion that the highly diversified C-terminal structures of all bacterial RTs were evolved to specifically recognize the highly diverse RNA structures in their cognate primer-template RNA molecules.

EXPERIMENTAL PROCEDURES

Materials—[α-32P]CTP and [α-32P]ATP were purchased from Amersham Biosciences. RNA fragments were synthesized by Integrated DNA Technologies, Inc.

Purification of RT-Ec86-(255–320) and RT-Ec73-(251–316)—For the purification of RT-Ec86-(255–320), the DNA fragment corresponding to the 66-residue, C-terminal region was PCR-amplified and cloned into a T7 expression vector so that an initiation codon was added at the N-terminal end and six His residues were added at the C-terminal end. The resulting clone was termed p86RTΔG. E. coli cells LE392(DE3) (11) transformed with p86RTΔG were grown in LB medium at 37 °C, and 1 mM isopropyl-β-D-thiogalactopyranoside was added at a mid-log phase. After a 3-h induction, cells from 6 liters of culture were harvested by centrifugation and disintegrated by a French press. The cytoplasmic soluble fraction obtained by ultracentrifugation of the cell lysates was fractionated by (NH4)2SO4. The fraction precipitated between 30 and

---

**Fig. 1.** msDNA synthesis and *E. coli* msDNAs. A, biosynthesis of msDNA is shown. Arrangement of msr, msd, and the RT gene are shown by arrows. Inverted repeats, a1/a2 and b1/b2, are labeled by arrows. The G residue, which is used for priming cDNA synthesis, is circled. The msr-msd region (branching G residue) and the RT gene are transcribed under the same promoter. The transcript from the msr-msd region is folded by forming a stem structure between the a1 and a2 inverted repeats as shown. The dotted line indicates the direction of cDNA synthesis from the branching G residue. Thick lines in the mRNA transcript correspond to the RNA molecule in the msDNA, which is essential for the recognition of the primer-template RNA by RT and is termed the recognition RNA. Dashed lines in the msDNA correspond to the DNA molecule. B, structures of msDNAs from *E. coli* retron. msDNA-Ec48 from Mao et al. (2), msDNA-Ec67 from Lampson et al. (3), msDNA-Ec73 from Sun et al. (4), msDNA-Ec78 from Lima and Lim (5), msDNA-Ec83 from Lim (6), msDNA-Ec86 from Lim and Maas (7), and msDNA-Ec107 from Herzer et al. are shown (8).
Bacterial RTs and Recognition of RNA Structures

FIG. 1 — continued
RESULTS

Purification of RT-Ec86-(255–320)—In a previous paper (10), we demonstrated that the C-terminal 91-residue sequence from RT-Ec86 is able to bind the primer-template RNA. In the present study, we further attempted to shorten the C-terminal fragment on the basis of the structural alignment of all E. coli RTs so far identified with HIV-1 RT, of which the x-ray structure has been determined (13, 14). In Fig. 2, the C-terminal halves of all E. coli RTs were aligned, including the highly conserved YXDD box, which is known to be located in the loop region between β9 and β10 strands in HIV-1 RT (13, 14). In this alignment, we assumed that all bacterial RTs have three-dimensional structures similar to HIV-1 RT, which has an RT domain consisting of 14 β strands and 10 α helices (13, 14). With the help of this secondary structure prediction, the C-terminal halves of all E. coli RTs were thus aligned with HIV-1 RT from β9 strand to α helix, the last helix in the RT domain (Fig. 2). Although the thumb domain regions were poorly aligned, we hypothesized that these regions of bacterial RTs also consist of four helices, αG, αH, αI, and αJ, as in the case of HIV-1 RT.

We have previously shown that the C-terminal region of RT plays a crucial role in recognition of its cognate primer-template RNA molecule (10). Therefore, for the primer-template RNA molecule to form the initial complex with its cognate RT for cDNA priming, we hypothesized that the recognition RNA region (shown in bold in Fig. 1A) may be able to interact only with the thumb domain consisting of four helices, αG, αH, αI, and αJ (see Fig. 2). If so, the region encompassing β11, β12, β13, and β14 strands in our previous C-terminal construct, RT-Ec86-(230–320), may be removed without losing the RNA binding activity. On the basis of this assumption, RT-Ec86-(255–320) used in the previous study (10) was further shortened by 25 residues to construct the 66-residue fragment from residue 255 (Ile-255 shown by an arrow in Fig. 2) to residue 320, called RT-Ec86-(255–320).

Purification of RT-Ec86-(255–320)—In a previous paper (10), we demonstrated that the C-terminal 91-residue sequence from RT-Ec86 is able to bind the primer-template RNA. In the present study, we further attempted to shorten the C-terminal fragment on the basis of the structural alignment of all E. coli RTs so far identified with HIV-1 RT, of which the x-ray structure has been determined (13, 14). In Fig. 2, the C-terminal halves of all E. coli RTs were aligned, including the highly conserved YXDD box, which is known to be located in the loop region between β9 and β10 strands in HIV-1 RT (13, 14). In this alignment, we assumed that all bacterial RTs have three-dimensional structures similar to HIV-1 RT, which has an RT domain consisting of 14 β strands and 10 α helices (13, 14). With the help of the second structure prediction, the C-terminal halves of all E. coli RTs were thus aligned with HIV-1 RT from β9 strand to α helix, the last helix in the RT domain (Fig. 2). Although the thumb domain regions were poorly aligned, we hypothesized that these regions of bacterial RTs also consist of four helices, αG, αH, αI, and αJ, as in the case of HIV-1 RT.

We have previously shown that the C-terminal region of RT plays a crucial role in recognition of its cognate primer-template RNA molecule (10). Therefore, for the primer-template RNA molecule to form the initial complex with its cognate RT for cDNA priming, we hypothesized that the recognition RNA region (shown in bold in Fig. 1A) may be able to interact only with the thumb domain consisting of four helices, αG, αH, αI, and αJ (see Fig. 2). If so, the region encompassing β11, β12, β13, and β14 strands in our previous C-terminal construct, RT-Ec86-(230–320), may be removed without losing the RNA binding activity. On the basis of this assumption, RT-Ec86-(255–320) used in the previous study (10) was further shortened by 25 residues to construct the 66-residue fragment from residue 255 (Ile-255 shown by an arrow in Fig. 2) to residue 320, called RT-Ec86-(255–320).

This fragment was tagged with six His residues at the C-terminal end and purified as described under “Experimental Procedures.” The purity of the final product (final yield, 2 mg/liter) is shown in Fig. 3A, lane 4. However, the purified RT-Ec86-(255–320) migrated somewhat faster than the one present in the crude extract likely
because of the lipopolysaccharide in the total cell lysate, which was absent in the purified sample. When analyzed by Tricine gel electrophoresis, they migrated at the same position. Its CD spectrum shown in Fig. 3A indicates that the C-terminal fragment has a very high α-helical content, which is consistent with the predicted structure in Fig. 2.

RNA Binding to RT-Ec86-(255–320)—Next we examined the ability of the purified RT-Ec86-(255–320) to bind RNA. On the basis of the previous SELEX experiments (10), a 28-base RNA (RNA 86msr-28; Fig. 4A, a) was synthesized, which consisted of a stem-loop structure identical to that of the recognition RNA for RT-Ec86 with the identical 9-base extension at the 5′-end including the branching G residue. Note that of the two stem-loop structures in the RNA molecule of msDNA-Ec86, the second stem-loop structure downstream of the branching G residue plays the major role for the specific interaction with RT-Ec86 (10). In the gel retardation assay shown in Fig. 4B, lane 1, RNA 86msr-28 was retarded in gel electrophoresis in the presence of RT-Ec86-(255–320). When intact RT-Ec86 was added, the RNA was also retarded with a much slower mobility (data not shown). This slower mobility of the RNA is due to the much higher molecular weight of the intact RT-Ec86 compared with RT-Ec86-(255–320).

To quantitatively analyze the interaction between RT-Ec86-(255–320) and RNA 86msr-28, a filter binding assay was carried out (Fig. 5). Variable amounts of the protein were incubated with a constant amount of the 32P-labeled RNA probe (50 fmol). The apparent Kᵰ value of the binding reaction was defined as the concentration of protein at which half of maximum binding was observed as described by Kajita et al. (15). Fig. 5 shows that RT-Ec86-(255–320) binds to the RNA molecule with a Kᵰ of 5 × 10⁻⁸ M.

Structural Requirement of RNA for Binding to RT-Ec86-(255–320)—Next, we further examined the length requirement of the 5′-end extension by stemwise shortening of RNA 86msr-28 as shown in Fig. 4A. To our surprise, all of the RNA molecules, 86msr-26(b), 86msr-25(c), 86msr-24(d), and 86msr-19(e), were able to bind to RT-Ec86-(255–320) as shown in Fig. 4B, lanes 2-5. These results indicate that the 5′-end single-stranded extension, including the branching G residue, is not required for RT-Ec86-(255–320) binding. It seems that the presumed four-helix bundle structure no longer recognizes the branching G residue, as also evident from the fact that the G to C substitution mutation in RNA 86msr-28 (RNA 86msr-28 m1; Fig. 4A, b) did not affect the gel retardation pattern (Fig. 4B, lane 6).

On the other hand, substitution mutations in the loop region of RNA 86msr-28, either UUU to UGC (RNA 86msr-28 m2; Fig. 4A, g) or UUU to UCU (RNA 86msr-28 m3; Fig. 4A, h), prevented the formation of retarded bands as shown in Fig. 4B, lanes 7 and 8, respectively. We also synthesized two other RNA molecules, one corresponding to the first stem-loop structure of RNA from msDNA-Ec73 (Fig. 4A, i; also see Fig. 1B) and the other corresponding to the second stem-loop structure (j in Fig. 4A; also see Fig. 1B). RT-Ec86-(255–320) bound to neither of them as shown in Fig. 4B, lanes 9 and 10, respectively.

Conversely, when RT-Ec73-(251–316) was used, it did not bind to RNAmsr-28 (Fig. 4A, a), RNAmsr-24 (d), and RNAmsr-28m3 (h) as shown in Fig. 4C, lanes 1-3, respectively. However, RT-Ec73-(251–316) did bind to RNA-73msr-29 (Fig. 4C, lane 5). Note that a faint band observed in the (−) lane is because of a double-stranded RNA dimer formed by two molecules of RNA-73msr-29, which could not be completely removed by heat treatment followed by rapid cooling. It also should be noted that RT-Ec73-(251–316) could not form a complex with RNA-73msr-26 (Fig. 4C, lane 4), indicating that as in the case of RT-Ec73, only the second but not the first stem-loop RNA structure in msDNA-Ec73 serves as the recognition RNA.

DISCUSSION

In the present paper, we identified a minimum domain in RT-Ec86 that is capable of specifically recognizing the secondary structure in the primer-template RNA. The RT-RNA interactions in the same retron are known to be highly specific for the cDNA priming reaction for msDNA synthesis. We found that the C-terminal 66-residue fragment of RT-Ec86 (RT-Ec86-(255–320)) is able to specifically bind to the recognition RNA (RNA-86msr-28), with a Kᵰ value of 5 × 10⁻⁸ M. This value agrees with a preliminary result previously obtained with RT-Ec86 (10). We also determined a minimum RNA structure of 19
bases consisting of only a stem-loop structure (RNA-86msr-19), which was able to bind to RT-Ec86-(255–320). A minimum structure of RT-Ec73 consisting of 66 residues was also determined, which was able to specially recognize the recognition RNA for RT-Ec73.

The C-terminal fragment was predicted to form a four-helix bundle or the thumb domain of RT on the basis of alignment of RT-Ec86 with HIV-1 RT (see Fig. 2). Through the alignments of bacterial RTs, this C-terminal region is highly diverse with few homologous sequences between any two bacterial RTs. Therefore, one may speculate that these diverse regions in bacterial RTs may be responsible for the recognition of highly diverse primer-template RNA structures or the recognition RNA structures except for the branching G residue (see Fig. 1B).

Nevertheless, as shown in Fig. 2, the C-terminal regions of bacterial RTs can be aligned with HIV-1 RT at the YXDD boxes of bacterial RTs and HIV-1 RT, which are located between the β9 and β10 strands. In addition, the G residue between the β12 and β13 strands in HIV-1 RT can be aligned with the G residue in the VTGL(V/I) sequence highly conserved in all bacterial RTs. The β12 and β13 strands are designated as the “primer grip” because of their proximity to phosphate joining the nucleotides at the primer end (14). In bacterial RTs, the VTGL sequence may play an essential role in orienting the 2'-OH group of the branching G residue at the priming site. The thumb domain of bacterial RTs probably participates not only in translocation of the template-primer following nucleotide incorporation as proposed by Jacobo-Molina et al. (14) but also in the specific recognition of the stem-loop structure of the primer-template RNA.

On the basis of the above assumption, the β14 strand, the last β strand in the RT domain, can be assigned as shown in Fig. 2. In RT-Ec86, the assumed β14 strand is followed by the GIGRE (residues 254–258) sequence, which is then assumed to be connected to the thumb domain consisting of four helices: G, H, I, and J. Although the assignment of these four helices was carried out on the basis of the secondary structure prediction from the primary structure, the exact locations of these helices are not certain at present. Nevertheless, the purified RT86-(255–320) was found to consist mostly of α-helical structures from its CD analysis as predicted.

It is important to note that there are critical differences in the way the primer-template RNA binds to RT between bacterial RTs and HIV-1 RT. In HIV-1 RT, the 3'-end CCA sequence of tRNA\(^{3\prime-5\prime}\) binds to the primer-binding site of the template, and this RNA-RNA duplex is placed on the palm domain held by the finger and thumb domains (in the right-hand configuration) (16). Therefore, the remaining tRNA structure is positioned at the lower (or downstream) part of the RNA-RNA duplex, which is then recognized by the lower part of the RT
Bacterial RTs and Recognition of RNA Structures

molecule consisting of a part of P51 of the P66/P51 heterodimer of HIV-1 RT. This region is thought to be responsible for the binding of the structure of tRNA\textsubscript{3\textprime} as. In contrast, most bacterial RTs do not have the RNase H domain and thus are likely to exist as a monomer. Furthermore, in the bacterial RT reaction, the primer-template RNA complex is performed before RT interaction, as shown in Fig. 1A, which is then placed on the RT molecule in a similar manner as with the RNA-RNA duplex in HIV-1 RT. Most significantly, the tRNA structure to be recognized by HIV-1 RT locates downstream of the 3' end of the primer site at the lower part of the palm region, whereas in bacterial RTs, the recognition RNA locates upstream of the 3' end of the branching G residue, in which the 2'-OH group is exposed to the active center of the RTs. Therefore, in contrast to HIV-1 RT, the secondary structure to be specifically recognized by bacterial RTs locates at the opposite side of the palm domain in HIV-1 RTs so that the region with which the recognition RNA is able to interact is likely to be only the thumb domain and not the palm domain. It should be noted that for msDNA synthesis the RNA duplex formation immediately upstream of the branching G residue is essential (17). This stem region is assumed to be positioned on the palm domain held by both the finger domain and the thumb domain.

There are a few key features for the RNA-RT-Ec86-(255–320) interaction. First, the 3U loop is essential because the replacement of the UUU loop with UGC or UCU resulted in complete loss of the RNA interaction with the protein. In the SELEX selection for stem-loop structures with intact RT-Ec86 (10), the central U residue was absolutely conserved, although the first and the third U residues were substitutable, albeit with a very low frequency, indicating that the central U residue is essential for the RNA-protein interaction. Secondly, the structure of the upper part of the stem (4 base pairs) is likely to play an important role in the interaction, as 10 of 14 sequences selected from the SELEX experiment had the identical sequence to Ec86 RNA (10). Interestingly, this stem region contains two U-G pairs flanked by a G-C pair, suggesting that structural instability residing in the upper part of the stem due to the U-G pairs may also be important for the RNA-protein interaction. Third, the 5' end single-stranded extension is no longer essential for the RNA-primer interaction with the presumed four-helix bundle thumb domain of RT-Ec86. The previous SELEX screening identified a 5' extension having the highly enriched AGC sequence (G corresponds to the branching G residue; see Ref. 10). The present results suggest that the 5' extension containing the branching G residue may interact with other parts of the RT molecule rather than with the four-helix bundle thumb domain at the C-terminal end. Fourth, the 66-residue, C-terminal domain of RT-Ec86 is highly unique in its amino acid composition. The fragment is very basic, having 16 basic residues (41%), whereas it contains only 6 acidic residues. It is also interesting to note that the fragment contains six aromatic residues. These unique features may be important for the specific interaction of the fragment with Ec86 RNA. Similarly, the 66 residue, C-terminal domain of RT-Ec73 is unique, having 13 basic, 8 acidic, and 7 aromatic residues.

It is interesting to note that msDNA synthesis stops at a certain base in a highly specific manner leaving a RNA-DNA duplex at the 3'-ends of both molecules (1) (also see Fig. 1B). This accurate termination of cDNA synthesis in the middle of a template RNA by bacterial RTs appears to occur before RT reaches the recognition RNA structure: 3 bases for msDNA-Ec86 to 13 bases for msDNA-Ec73 (see Fig. 1B). However, when a retroviral RT is added to msDNA isolated from cDNA, synthesis can continue all the way to the branching G residue in all msDNA thus far tested (1). Therefore, it is tempting to speculate that the stable interaction of the recognition RNA stem-loop structures with the thumb domains of individual RTs prevents RTs from proceeding further with the cDNA synthesis at a certain distance from the recognition RNA structure.

The thumb domain of retroviral RTs has been proposed to participate in translocation of the template-primer duplex following the addition of a nucleotide at the β active center (14). It is rather astounding that the thumb domains of bacterial RTs evolved to acquire a seemingly additional function to recognize specific RNA structures. Because it is considered that retroviral RTs evolved from a common ancestral RT with bacterial RTs (1), retroviral RTs might have lost the function of the thumb domain for RNA recognition during evolution. Structural elucidation of the precise interaction between RNA and
the thumb domain will no doubt shed light onto the general molecular mechanism of RNA recognition by a protein consisting of a four-helix bundle and may provide important insights into the versatility of a four-helix bundle protein for the recognition of a large number of unique RNA structures. It will also be of great interest to examine whether the thumb domains of at least some eukaryotic RTs may retain the ability to recognize specific RNA structures.

Acknowledgments—We are grateful to Dr. M. Roth and Dr. Sangita Phadtare for their critical reading of this manuscript.

REFERENCES
1. Yamanaka, K., Shimamoto, T., Inouye, S., and Inouye, M. (2002) in Mobile DNA II (Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A.) pp. 784–795, ASM Press, Washington, D. C.
2. Mao, J. R., Inouye, S., and Inouye, M. (1997) J. Bacteriol. 179, 7865–7868
3. Lampson, B. C., Sun, J., Hsu, M.-Y., Vallejo-Ramirez, J., Inouye, S., and Inouye, M. (1989) Science 243, 1033–1038
4. Sun, J., Inouye, M., and Inouye, S. (1991) J. Bacteriol. 173, 4171–4181
5. Lima, T. M., and Lim, D. (1997) Plasmid 38, 25–33
6. Lim, D. (1992) Mol. Microbiol. 6, 3531–3542
7. Lim, D., and Maas, W. K. (1989) Cell 56, 891–904
8. Hertz, P. J., Inouye, S., and Inouye, M. (1992) Mol. Microbiol. 6, 345–354
9. Shimamoto, T., Hsu, M. Y., Inouye, S., and Inouye, M. (1993) J. Biol. Chem. 268, 2684–2692
10. Inouye, S., Hsu, M. Y., Xu, A., and Inouye, M. (1999) J. Biol. Chem. 274, 31296–31244
11. Shimamoto, T., Inouye, M., and Inouye, S. (1995) J. Biol. Chem. 270, 581–588
12. Phadtare, S., and Inouye, M. (1999) Mol. Microbiol. 33, 1004–1014
13. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992) Science 256, 1783–1790
14. 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., Hizi, A., Hughes, S. H., and Arnold, E. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 6320–6324
15. Rajita, Y., Nakayama, J., Aizawa, M., and Ishikawa, P. (1995) J. Biol. Chem. 270, 22167–22175
16. Steitz, T. A. (1999) J. Biol. Chem. 247, 17395–17398
17. Hsu, M.Y., Inouye, S., and Inouye, M. (1989) J. Biol. Chem. 264, 6214–6219