Cell. Mol. Life Sci. (2011) 68:3607–3617
DOI 10.1007/s00018-011-0671-0

Cellular and Molecular Life Sciences

RESEARCH ARTICLE

Retron Se72 utilizes a unique strategy of the self-priming initiation of reverse transcription

Lenka Pilousova · Ivan Rychlik

Received: 23 November 2010 / Revised: 11 January 2011 / Accepted: 24 February 2011 / Published online: 31 March 2011

Abstract Unlike all of the other retrons, the bacterial retron reverse transcriptase RrtE is capable of synthesizing small double-stranded DNA (sdsDNA) from template RNA. In this study, we analyzed the biosynthesis of the sdsDNA by RrtE in detail. We found out that the initiation of reverse transcription was dependent on a novel self-priming mechanism utilizing a free 3′OH of RNA that is reverse-transcribed into sdsDNA. The priming of the sdsDNA synthesis was not dependent on any particular nucleotide being used as a donor of 3′OH (unlike all of the other retrons, which prime from 2′OH of a particular guanosine) or any particular nucleotide being introduced into the sdsDNA first. Due to the relaxed demands for the initiation of reverse transcription, RrtE has the potential to generate dsDNA from different RNA transcripts in vivo.

Keywords Retron · Self-priming mechanism · Reverse transcriptase · *Salmonella*

Introduction

Reverse transcriptases are suggested to be one of the oldest groups of enzymes as they are thought to be involved in the transition of the RNA-based world to the DNA-based world [1]. These enzymes are found in all eukaryotic cells and also in some eukaryotic viruses. When compared to eukaryotes, the prokaryotes only rarely code for these enzymes; however, with the increasing numbers of completely sequenced bacterial genomes, new bacterial reverse transcriptases are being described. In prokaryotes, reverse transcriptases can be found as part of the group II introns, diversity generating elements, and elements known as retrons [2–5].

The retrons are bacterial retroelements that encode retron reverse transcriptases responsible for the production of multicopy single-stranded cDNA (msDNA). Despite remarkable diversity at the sequence level, there are several structural features shared by all retrons. Retrons consist of an *msr* region, *msd* region, and an *rrt* gene coding for the retron reverse transcriptase. After transcription of the whole element, a 5′ nontranslated region containing the *msr* and *msd* regions self-anneals due to the presence of two pairs of inverted repeats: a1/a2 and b1/b2. The a1/a2 repeat forms a stable stem structure that is essential for the retron reverse transcriptase to initiate msDNA synthesis using the 2′OH group of a particular G residue within the *msr* region as a priming site. cDNA (i.e., msDNA) synthesis continues using the *msd* region as a template until the termination point is reached. Except for the 5–8 bp of the RNA/msDNA hybrid protected by the reverse transcriptase, the unprotected reverse transcribed RNA is degraded by host-derived RNase H. As a consequence, single-stranded msDNA complementary to the *msd* region with a 5′ end bound by a 2′–5′ phosphodiester bond to the template RNA and a 3′ end forming a short RNA/msDNA hybrid accumulates in the cytoplasm of bacterial cells. Although the synthesis of msDNA is relatively well described, the biological meaning of this activity remains completely unknown [6].

In our previous study, we identified a novel retron element in *Salmonella enterica* serovar Enteritidis

L. Pilousova · I. Rychlik (✉)
Veterinary Research Institute, Hudcova 70,
621 00 Brno, Czech Republic
e-mail: rychlik@vri.cz

L. Pilousova
e-mail: pilousova@vri.cz

(*S*. Enteritidis) designated as Se72, which, unlike all the other retrons, produced small double-stranded DNA (sdsDNA) [7]. This element consists of the 72 bp long *msd* region followed by three open reading frames (ORFs) essential for sdsDNA synthesis, ORF3, which is surrounded by a 13 bp inverted repeat (IR1), a retron reverse transcriptase (*rrtE*), and a cold shock protein homologue (*csp*) (Fig. 1a). Although the Se72 reverse transcriptase showed sequence similarities to other retron reverse transcriptases (30% identity and 52% similarity at the amino acid level), this retron element also exhibits several specific features. First, unlike all so far described bacterial retrons located chromosomally, this is the only one encoded by a small ColE1 plasmid. Second, the product of its enzymatic activity is not a single-stranded DNA but a double-stranded molecule with single-stranded overhangs (sdsDNA) (see Fig. 1a and also [7]). And third, we were unable to identify the general structures conserved among all previously described retrons, indicating that Se72 may use a different strategy for the initiation and biosynthesis of sdsDNA. In this study, we were therefore interested in the biosynthesis of sdsDNA in detail. We initially determined the 5′ end of the Se72 retron transcript, and after computer modelling of the secondary structure of the transcript, we proposed a model of the synthesis of sdsDNA, which we confirmed experimentally. We found out that unlike all of the other bacterial retrons described so far, the Se72 retron does not branch the template RNA by a 2′–5′ phosphodiester bond and, instead, it uses a free 3′OH end of RNA for initiating the synthesis of sdsDNA. This makes this retron a unique genetic element different from any other prokaryotic element with reverse transcriptase activity described so far.

Materials and methods

Bacterial strains

S. Enteritidis strain 2159 with the wild-type plasmid pI encoding the Se72 retron [8, 9] was used in all of the experiments focused on the identification of native sdsDNA/RNA molecules. For site-directed mutagenesis, a plasmid pI-Amp with an inserted ampicillin resistance gene outside the retron element was used. The *S*. Enteritidis strain 2160 was used as a recipient for the recombinant plasmids [7]. *Salmonella* Enteritidis with inactivated *rrtE* and *rnhA* genes and *Escherichia coli* with retron Ec107 were used as the controls in some of the experiments (see Table 1 for a list of strains). The *rnhA* mutant was constructed by λRed recombination [10], replacing the whole *rnhA* sequence with the *cat* gene coding for chloramphenicol resistance. After the generation of the mutant and its confirmation by PCR using a C1 primer and a primer flanking the *rnhA*, the mutation was transduced into a fresh *S*. Enteritidis 2160 genetic background. All of the strains were routinely grown in LB broth (Difco) at 37°C in a shaking incubator to the mid-exponential phase of growth.

Identification of the 5′ end of Se72 mRNA and prediction of the secondary structure of the transcript

The 5′ end of Se72 mRNA was determined by primer extension as follows. Total RNA was isolated using an RNeasy Mini kit (Qiagen) and reverse transcribed using the 6-FAM labelled primer P1 5′ GGAAAGCAAGGAAAAA CTC 3′ and M-MLV reverse transcriptase (Invitrogen). The

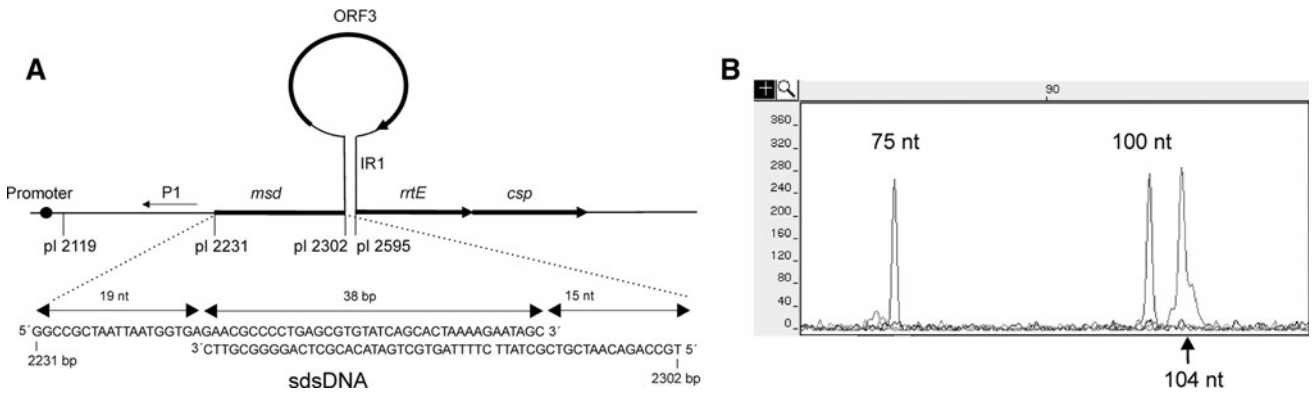

Fig. 1a, b Retron Se72 and identification of the 5′ end of the RNA transcript using primer extension. **a** A schematic drawing of retron Se72 and the sdsDNA locus. **b** Identification of the 5′ end of *rrtE* transcript. Fluorescently labelled primer P1 was extended by M-MLV reverse transcriptase along the *rrtE* transcript. The length of the extension product was determined using capillary electrophoresis as 104 nt (the peak marked with the *arrow*). TAMRA™ Size Standard peaks (75 and 100 nt in size) are also shown

Table 1 List of strains

Strain	Plasmid	Genotype	Reference
S. Enteritidis 2159	pJ and pI	WT	[7]
S. Enteritidis 2160	pJ	WT	[7]
S. Enteritidis 2160	pJ, pI-Amprrt	*rrtE*::Amp	[7]
S. Enteritidis 2160	pJ, pI-Amp	Amp insertion in non-coding region of pI	[7]
E. coli		Ec107	This study
S. Enteritidis 2160 *rnhA*-	pJ and pI-Amp	*rnhA*::Cm	This study
S. Enteritidis 2160 M1	pJ, pI-AmpM1	ATCATA insert between A_{2302} and A_{2303}	This study
S. Enteritidis 2160 M2	pJ, pI-AmpM2	T_{2588}CA replaced by GTG	This study
S. Enteritidis 2160 M3	pJ, pI-AmpM3	A_{2303}TG replaced by CCC	This study
S. Enteritidis 2160 M4	pJ, pI-AmpM4	A_{2591}ACAT replaced by CCACA	This study

resulting cDNA was separated by capillary electrophoresis using an ABI Prism 310 Genetic Analyzer, and the size of the fragment was determined using a TAMRA™ Size Standard (Applied Biosystems). The mFold with default settings (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) was used to predict the secondary structure of the 5′ untranslated region of the Se72 RNA transcript, from the identified transcription start up to the initiation codon of reverse transcriptase.

Northern blot analysis

For Northern blot analysis, RNA was isolated using TRI Reagent extraction according to the manufacturer's instructions (Sigma-Aldrich). Treatment with DNase I and RNase A was performed as described previously [7], and the DbrI 2′–5′ debranching enzyme was used according to Nam et al. [11]. Prior to loading, the samples were mixed with an equal volume of RNA loading buffer (95% formamide, 18 mM EDTA, and SDS, xylene cyanol, and bromphenol blue, each of 0.025% concentration) and incubated for 5 min at 95°C. The samples were then separated in 10% polyacrylamide gel containing 6M urea at 55°C, stained with ethidium bromide, visualized under UV light, and electroblotted onto a nylon membrane (Hybond N, Amersham) using transcript RNA Marker 0.1–1 kb (Sigma-Aldrich) as a molecular weight standard. Synthetic oligonucleotide probes complementary to the RNA sequence of the ORF3 loop [ORF3 probe: 5′ GCTCTTGGTTA CGTTAATGTTGATGTTGATAATAAACATGATGCAACTCCTATTGGTTATGA AATTTTGCCAGACCAGTAGGAAAGCAAGGAAAAACTC 3′], to the (−)sdsDNA strand [(−)sds probe: 5′ TGGTGAGAACGCCCCT GAGCGTGTATCAGCACTAAAAGAATAGCGACGATTGGTGAGAACGCCCC TGAGCGTGTATCAGCACTAAAAGAATAGCGACGA 3′], and to the (+)sdsDNA strand [(+)sds probe: 5′ TCGTCGCTATTCTTTTAGT GCTGATACACGCTCAGGGGCGTTCTCACCAATCGTCGCTATTCTTTTAGT GCTGATACACGCTCAGGGGCGTTCTCACCA 3′] were labelled, hybridized, and detected using Gene Images AlkPhos Direct Labelling and Detection System (Amersham Bioscience). Sequences of both sds probes were duplicated to increase the size of the probe and the efficiency of labelling with alkaline phosphatase.

Extension of sdsDNA strands by reverse transcriptase and PCR

RNeasy Mini kit–purified RNA (~3.5 μg), which had not been treated with DNase I, was subjected to reverse transcription with M-MLV reverse transcriptase (Invitrogen) or AMV reverse transcriptase (AbGene), according to the manufacturer's instructions, except for the fact that no external primer was added to the reaction. This led to an extension from the sdsDNA strands in the native sdsDNA/RNA complex. The resulting cDNA was used as a template for PCR, which was performed in 20 μl volumes using the PCR Master Mix kit (Qiagen) according to the manufacturer's instructions, with primers P2 (5′ GTCCCAAAGAG CCAATAGAG 3′), P3 (5′ AAAGAGAGGTTGGCTCTC TG 3′), P4 (5′ AAGTAATAGATTACGTATCG 3′), and P5 (5′ CAAATTTTTGAACCCAAGAG 3′) combined either with primer P6 (5′ CCAGACAATCGTCGCTATTC 3′) or P7 (5′ TGGCCGCTAATTAATGGTGA 3′). Finally, the PCR products were purified using the QIAquick PCR purification Kit (Qiagen) and sequenced using the Big Dye Terminator v. 3.1 Sequencing kit and the ABI Prism 310 Genetic Analyzer (Applied Biosystems). Sequences of P2/P6 and P2/P7 PCR products were determined with exactly the same results in eight independent experiments with sdsDNA extensions amplified by PCR. If necessary, DNase I treatment was performed with RNase free DNase I (Promega). To control for the presence of intact RNA in the samples, reverse transcription with primer P8 (5′ AAC CAACCAACCAACCAACCCGATTAGTTAAAAAGTT TCG 3′) complementary to the ORF3 region (underlined nucleotides) containing a 20 nt long extension at its 5′ end was performed followed by PCR with primer

P9 (5′ AACCAACCAACCAACCAA 3′), identical to the extension of the P8 primer and the P2 primer.

Identification of 5′ ends of both sdsDNA strands present in the RNA/sdsDNA complex

The cDNA extended from the sdsDNA, as described above, was mixed with primer P10 (5′ AGCGACGATTGTATGTGAGTTAAAAGTCCCAA 3′), which hybridized to the 3′ ends of both sdsDNA extension products (nucleotides complementary to 2,303–2,322 bp of plasmid pI are underlined) with a 12 nt long extension on its 5′ end. Next, Taq PCR Master Mix (Qiagen) was added, and the mixture was allowed to extend the 3′ ends for 10 min at 72°C. Extended dsDNA products containing A overhangs were subsequently cloned into the pCR2.1 plasmid using the TOPO TA Cloning kit (Invitrogen) and sequenced. Identical sequences were obtained from ten clones from two independent experiments.

Identification of the 3′ ends of sdsDNA in the RNA/sdsDNA intermediate

To determine the 3′ ends of native sdsDNA, terminal deoxynucleotidyl transferase (TdT) was used to add oligo-dA tails to the 3′ ends of (+) and (−)sdsDNA strands. Purified RNA (~3.5 µg), which had not been treated with DNase I, was mixed with 10 U of the TdT enzyme (New England Biolabs), 5 µl of 2.5 mM $CoCl_2$, 5 µl of 10× reaction buffer (20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, and 1 mM DTT), and 5 µl of 10 mM dATP in a volume of 50 µl. The mixture was incubated for 30 min at 37°C, and the reaction was stopped by heating to 70°C for 10 min. In the next step, oligo-dT primers were used to prime the reverse transcription from the oligo-dA tails. The resulting cDNAs were then used as templates in PCR with the oligo-dT and P2 primers; the PCR products were cloned into pCR2.1 and sequenced as described above.

Ligation-mediated PCR

The RNA from the wild-type strain and the *rrtE* mutant was isolated using an RNeasy Mini kit (Qiagen), mixed with 20 U of T4 RNA Ligase 1 (New England Biolabs), 1 µl of 10× reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT), and 50 pmol of 5′ phosphorylated oligonucleotide O1 (5′ CTCAGTATGGCATCCGTACG 3′), and incubated for 2 h at 37°C. Six microliters of the reaction mixture was then used as a template in the reverse transcription with primer P11 (5′ CGTACGGATGCCATACTGAG 3′), and semi-nested PCR was used to amplify the ligation products prior to cloning. In the first PCR with primers P2 and P11, 1 µl of cDNA was used. Of the PCR product from the first round of PCR, 1 µl was then used as a template in the second round of PCR with primers P3 and P11, and the resulting PCR product was cloned into pCR2.1 and sequenced. The PCR with primers P3 in combination with P12 (5′ GATGCCATACTGAGATG 3′), complementary to the O1 but containing an additional 3 nt at its 3′ end, complementary to $C_{2593}A_{2594}U_{2595}$, was used for the positive selection of clones containing oligonucleotide O1 ligated to U_{2595}.

Site-directed mutagenesis

Plasmid pI-Amp was modified exactly as described previously for the removal of the plasmid pI inverted repeat [7]. The modified target sequences are listed in Table 1.

Results

Identification of the 5′ end of Se72 mRNA

Primer extension of the fluorescently labelled primer P1 with M-MLV reverse transcriptase led to the 104 nt long extension product (Fig. 1). This indicated that the transcriptional start point was at position 2,119 bp of the plasmid pI sequence and, consistent with this finding, a consensus sequence of the σ70 promoter GTGACAG $(N)_{15}$TATAAT was identified 11 bp upstream from the determined transcriptional start point.

Prediction of the secondary structure of the 5′ untranslated region of the Se72 RNA transcript

Identification of the transcriptional start point enabled us to model the secondary structure of the 5′ untranslated region (5′UTR) of retron Se72 mRNA from its transcriptional start point up to the start codon of *rrtE*. The most energetically favorable model predicted sdsDNA and ORF3 loops separated by a region of relaxed conformation in which sequences in positions 2,301–2,306, 2,223–2,227, and 2,590–2,595 bp appeared close to each other (Fig. 2). Among the next 16 predicted structures, an additional 3 met the requirements of the *msd* region, forming an independent loop, and these differed from the most favorable model only in the more relaxed structure of the ORF3 loop. However, neither the structure shown in Fig. 2, nor the remaining 16 predicted ones, suggest a base pairing between the nucleotides marked by asterisks in Fig. 2. Based on our subsequent experiments, the additional base pairing was found to be crucial for sdsDNA synthesis. This discrepancy could be caused by RNA-protein interactions

that may change the RNA structure in the cells, but we did not investigate this in greater detail in this study.

Northern blot hybridization

In order to detect the molecules produced during sdsDNA synthesis, Northern blot hybridization was performed in which four different molecules were identified: the primary RNA transcript, a (+)sdsDNA strand, and ORF3 RNA, either free or with a bound (−)sdsDNA strand. The primary transcript of 5′ UTR was identified as a molecule of approximately 500 nt in size that hybridized with the ORF3 and (+)sds probes and that was sensitive to RNase A and resistant to DNase I treatment (Fig. 3), all consistent with the RNA molecule from the transcription start point up to the initiation codon of *rrtE*. The second molecule only hybridized with the (+)sds probe, and it was resistant to RNase A but sensitive to DNase I treatment (Fig. 3). This corresponds to the (+)sdsDNA strand released from the sdsDNA/RNA complex during electrophoresis under denaturing conditions, indicating that it was noncovalently bound to the complex. Finally, the RNA/(−)sdsDNA molecule was identified as a molecule that hybridized with ORF3 and (−)sds probes. Treatment with DNase I and RNase A released RNA or DNA molecules of expected sizes. Detection of this molecule after electrophoresis under denaturing conditions indicated the presence of a covalent bond between the RNA and the (−)sdsDNA strand, which must have been different from the 2′–5′ bond since it was resistant to the activity of the 2′–5′ debranching enzyme DbrI, whereas the msDNA of retron Ec107, containing the 2′–5′ phosphodiester bond [12], was easily released from the complex with RNA (Fig. 3b). Since DbrI has a weak activity for the debranching of pyrimidine nucleotides [11], mutants with G and A at the 5′ end of the (−)sdsDNA strand were constructed, and RNA isolated from these mutants was processed with DbrI, but even in this case DbrI was unable to release (−)sdsDNA from the complex with RNA (data not shown).

The role of ribonuclease HI in sdsDNA synthesis

Since there was no 2′–5′ linkage between RNA and DNA, it was apparent that the synthesis of sdsDNA differed from msDNA synthesis. In the next step, we therefore tested whether or not the cellular RNase HI involved in msDNA synthesis also participates in the synthesis of sdsDNA.

A mutant lacking the *rnhA* gene encoding for RNase HI did not produce a sufficient enough amount of sdsDNA to be detectable by PAGE (Fig. 4a). However, when RNA from the *rnhA* mutant was subjected to Northern blot hybridization using (+)sdsDNA and (−)sdsDNA probes, two molecules were detected, respectively. The first one corresponded with the primary transcript from the transcriptional start point up to U_{2595}. The second molecule, of a higher molecular weight than the primary transcript, hybridized with both (+)sdsDNA and (−)sdsDNA probes and was present only in the *rnhA* mutant. This molecule corresponded to the RNA transcript with the (−)sdsDNA strand in which the *msd* RNA that had served as the template for reverse transcription was not degraded by the RNase HI. At the same time, a molecule consisting of an ORF3 RNA loop with an (−)sdsDNA strand was absent in the *rnhA* mutant (Fig. 4).

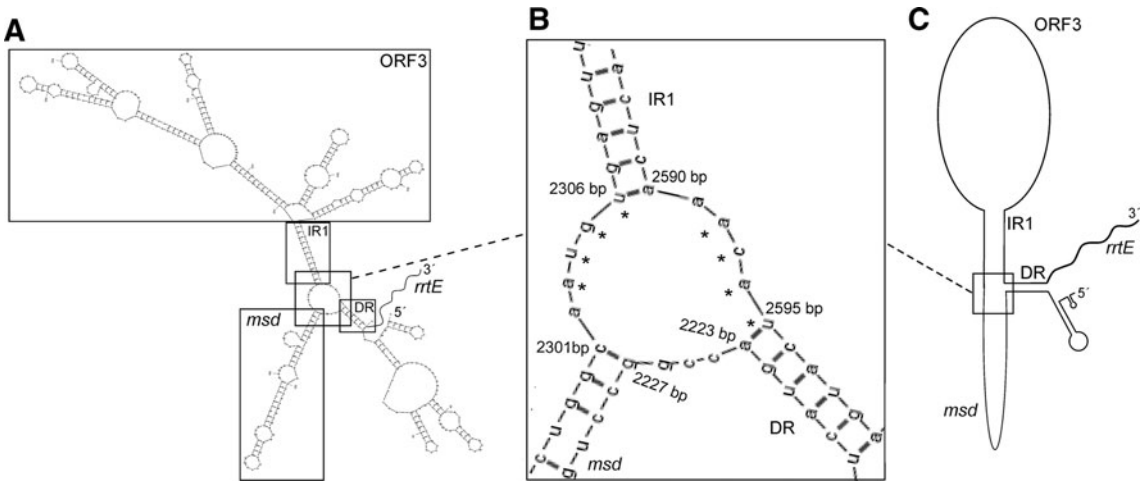

Fig. 2a–c Prediction of the secondary structure of 5′ UTR of the Se72 transcript. **a** Predicted secondary structure of 5′ UTR of the Se72 transcript. **b** In between the inverted repeats (IR1) and the *msd* loop (template for sdsDNA synthesis), a relaxed structure with possible alternative base pairing, indicated by *asterisks*, was identified. **c** A schematic drawing of the secondary structure based on the mFold predicted model. *DR* Direct repeat, *rrtE* sequence translated to RrtE

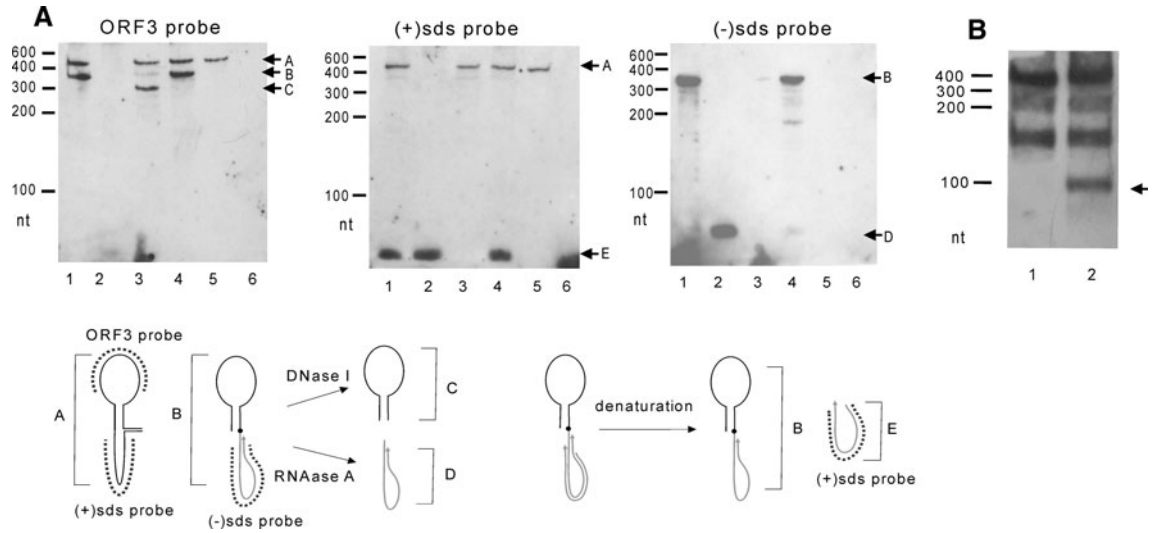

Fig. 3a, b Denaturing Northern blot analysis of the intermediate molecules in sdsDNA synthesis. **a** Detection of the intermediate molecules in sdsDNA synthesis. RNA was purified from the wild-type *S*. Enteritidis 2159, separated by denaturing PAGE, and electroblotted and hybridized with synthetic oligonucleotide probes. The molecular sizes determined using the ssRNA standard are shown on the *left*. *Lane 1* Untreated RNA, *lane 2* RNA treated with RNase A, *lane 3* RNA treated with DNase I, *lane 4* RNA treated with DbrI, *lane 5* untreated RNA from the *rrtE* mutant, *lane 6* RNA from *S*. Enteritidis 2160 lacking pI, used as a negative control. *Below* there is a scheme characterizing the identified molecules. A continuous *black line* represents the RNA transcript, *gray lines* represent sdsDNA strands with the 3′ ends pointed out by *arrows*, a *black circle* represents the 3′–5′ linkage, and *dotted lines* represent the oligonucleotide probes used for the detection of individual molecules. *Molecule A* represents the primary transcript from the transcription start point to the start codon of *rrtE*. *Molecule B* corresponds to (−)sdsDNA/RNA, *molecule C* corresponds to the RNA part of (−)sdsDNA/RNA, *molecule D* corresponds to the DNA part of (−)sdsDNA/RNA, and *molecule E* corresponds to the (+)sdsDNA strand. **b** RNA from *E. coli* harboring the retron Ec107. *Lane 1* Untreated RNA, *lane 2* RNA treated with DbrI; msDNA released from the msDNA/RNA complex, released by DbrI, is indicated by the *arrow*

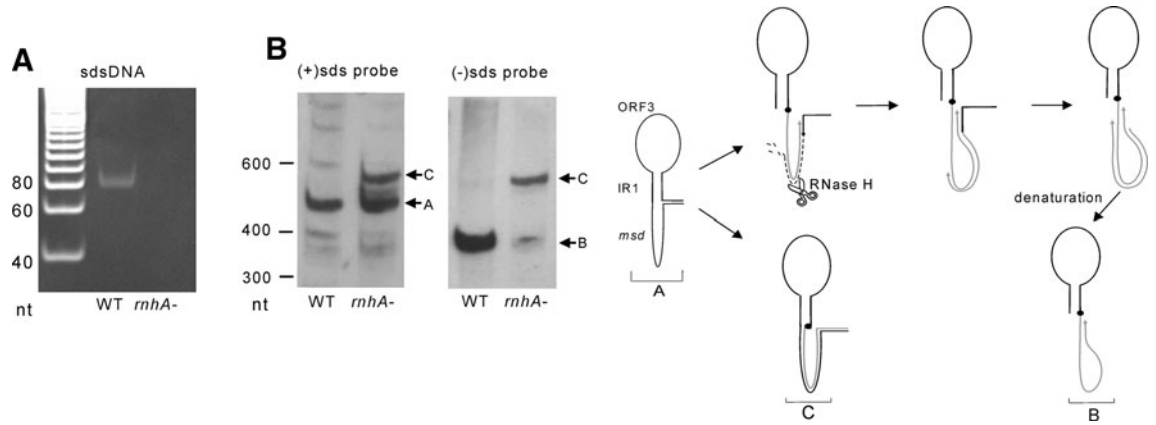

Fig. 4a, b Northern blot analysis of the *rnhA* mutant. **a** Detection of sdsDNA extensively treated with RNase A and separated by nondenaturing PAGE. **b** Northern blot analysis of RNA purified from the wild-type *S*. Enteritidis 2159 and the *rnhA* mutant and separated by denaturing PAGE and hybridized with oligonucleotide probes. The molecular sizes determined using the ssRNA standard are shown on the *left*. *Molecule A* was present in both RNA samples and represents the primary transcript from the transcription start point to the start codon of *rrtE*. *Molecule B* was only detected in RNA isolated from the WT strain, whereas *molecule C* was present only in the *rnhA* mutant

Extension of sdsDNA by reverse transcriptase

Using the predicted sdsDNA/RNA structure shown in Figs. 3 and 4, we speculated that an extension of both sdsDNA strands with commercial reverse transcriptases should result in cDNA molecules that could easily be detected by PCR with primers identical to either (−)sdsDNA or (+)sdsDNA strands in a combination with primers designed along the ORF3 loop. Furthermore, the relative positions and orientations of such primer pairs should exclude amplification from plasmid pI, eventually contaminating RNA samples. To test this hypothesis, RNA

from *S.* Enteritidis strain 2159 was isolated, subjected to reverse transcription without the addition of any external primer, and the obtained cDNA was used as a template in PCR. As expected, PCR with either primer P6 identical to the (−)sdsDNA strand or primer P7 identical to the (+)sdsDNA strand resulted in positive amplifications when combined with primers P2, P3, and P4, identical to the ORF3 loop sequence, whereas no amplification was observed when primer P5, located downstream from the ORF3 loop, was used in combinations with P6 or P7 (Fig. 5). When the resulting PCR products were sequenced, the sequence of the P2/P6 PCR product consisted of a sequence complementary to the *msd* region from primer P6 up to C_{2224} [i.e., it was identical to the (−)sdsDNA strand], a dinucleotide CA, and after this dinucleotide the sequence continued with a sequence complementary to the ORF3 loop, from U_{2595} up to the primer P2 (Fig. 5). The sequence of the P2/P7 PCR product was identical with the *msd* region up to A_{2302} [i.e., the sequence was identical to the (+)sdsDNA strand], and then it continued with a sequence complementary to the ORF3 loop from U_{2595} up to the primer P2. Sequences of P2/P6 and P2/P7 PCR products were determined with the exactly same results in eight independent experiments. No amplification occurred when DNase I treatment preceded the reverse transcription of RNA isolated from the wild-type strain or when the cDNA from the *rrtE* mutant was used as a template (Fig. 5). The last control experiment showed that the same results were obtained when either M-MLV or AMV reverse transcriptase was used for sdsDNA extension (data not shown).

Identification of 3′ and 5′ ends of sdsDNA molecules in the RNA/DNA complex

The sequence of sdsDNA after RNase A treatment has already been identified (see [7] or Fig. 1a); however, the ability of M-MLV reverse transcriptase to extend both sdsDNA strands along the ORF3 region allowed us to determine the sequence of native sdsDNA in a complex with RNA. The 3′ ends of the (−) and (+)sdsDNA strands of the native sdsDNA bound to RNA were identified as G complementary to C_{2224} of the plasmid pI and T identical to T_{2293} of the plasmid pI, respectively. This means that the native 3′ ends of the sdsDNA strands shown in Fig. 6a differed from those determined previously [7] and shown in Fig. 1a. The 5′ ends of the (−) and (+)sdsDNA strands were identified as T complementary to A_{2302} and G identical to G_{2231}, respectively, exactly as determined earlier for the purified sdsDNA ([7], see also Fig. 6a). Subsequent analysis of these results showed that the sequence of the native 5′ single-stranded overhang of (−)sdsDNA strand allowed an imperfect base pairing with its own 3′ end, which explained the introduction of the dinucleotide CA into the extension product of the (−)sdsDNA strand (Fig. 5). However it is possible that base pairing, due to the presence of mismatched nucleotides, represents only one factor enabling the extension by M-MLV reverse transcriptase, and other factors facilitating the extension by commercial reverse transcriptase may include a specific three-dimensional structure of the whole complex.

Detection of the free 3′ end of the RNA transcript serving as the primer

Comparing the sequence of P2/P6 and P2/P7 PCR products together with the 5′ end of (−)sdsDNA indicated that U_{2595} is the most probable nucleotide used for the priming of (−)sdsDNA synthesis. Since in the Northern blot experiments we excluded a 2′–5′ linkage between the RNA and (−)sdsDNA, we expected that the synthesis of the (−)sdsDNA strand was primed from the 3′OH of U_{2595}. If this assumption was correct, we would be able to detect free 3′OH at U_{2595} by ligation-mediated PCR. To prove this, RNA isolated from the strain *S.* Enteritidis 2159 was ligated with the oligonucleotide O1, and after ligation, PCR with primer P9 complementary to O1 and primer P2 was performed followed by cloning the PCR products. When eight clones were randomly picked and sequenced, the oligonucleotide O1 was found to be ligated to positions U_{2567}, U_{2584}, A_{2586}, C_{2589}, $A_{2590}(2\times)$, A_{2591}, and A_{2594}. Therefore, we tested the insert sizes by PCR in 100 colonies and found that these contained inserts of an expected size or shorter. Based on this, we thought that a nonspecific degradation of the 3′ end of the RNA transcript must have occurred during RNA isolation or the ligation reaction. Therefore, in the repeated experiment, we used primers designed to enable positive amplification only in the case of the ligation of O1 to U_{2595}. An additional set of 16 clones was tested, and among these, we found one clone in which O1 was ligated to U_{2595}, confirming the presence of an RNA transcript ending with U_{2595} and with free 3′OH available for the initiation of sdsDNA synthesis by RrtE.

The requirements for a specific sequence for initiating sdsDNA synthesis

In order to get a deeper insight into the role of the sequences involved in initiating reverse transcription, we constructed four different mutants in the sequences expected to be required for priming sdsDNA synthesis. In the M1 mutant, we inserted an ATCATA sequence in between A_{2302} [the first nucleotide transcribed into the native (−)sdsDNA strand] and A_{2303}. The sdsDNA from this mutant was visualized in PAGE as a molecule of slightly higher molecular weight (Fig. 6b). When the sdsDNA isolated from the M1 mutant was directly cloned

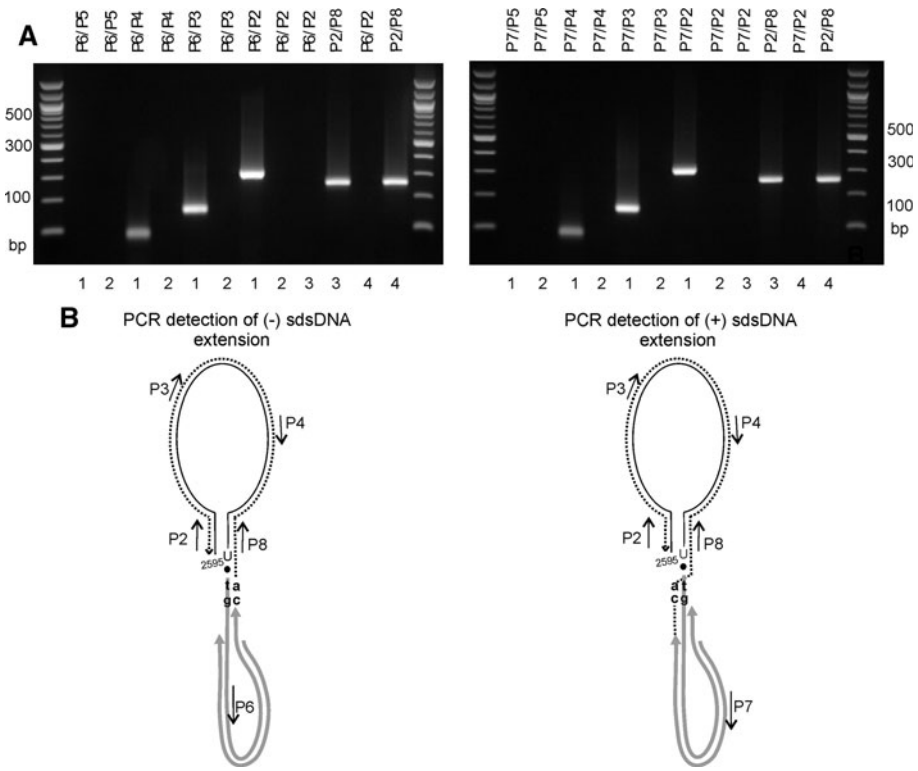

Fig. 5a, b Extension of sdsDNA strands by reverse transcriptase followed by PCR. **a** PCR amplification of the cDNA obtained by extension from the 3′ ends of both sdsDNA strands. In the lanes marked as *1*, RNA originating from the wild-type *S*. Enteritidis strain 2159 was reverse transcribed into cDNA and used as a template in PCR. In the lanes marked as *2*, plasmid DNA from the wild-type *S*. Enteritidis strain 2159 was used as a template in PCR. In lanes marked as *3*, RNA originating from the wild-type *S*. Enteritidis strain 2159 was extensively treated with DNase I prior to reverse transcription and PCR. In the lanes marked as *4*, RNA originating from the *S*. Enteritidis *rrtE* mutant was reverse transcribed into cDNA and used as a template in PCR. **b** Schematic drawings of sdsDNA M-MLV RT extension products detected by PCR. A continuous *black line* represents the RNA transcript, *gray lines* represent sdsDNA strands with the 3′ ends pointed out by *arrows*, the *dotted line* represents the extension product, and a *black circle* represents 3′–5′ linkage

after Taq polymerase extension, as described previously [7], the inserted nucleotides were found at the 5′ end of the (−)sdsDNA strand. Parallel sequencing of the P2/P6 PCR product after M-MLV extension of sdsDNA produced by the M1 mutant showed that the inserted nucleotides appeared immediately downstream of the CA dinucleotide, further confirming the hypothesis on the origin of the CA nucleotide (Fig. 6). Replacement of the TCA sequence at position 2,588–2,590 bp in the M2 mutant with GTG sequences did not affect the preferred base pairing in the priming center, and the sequence of sdsDNA was the same as in the wild-type strain (Fig. 6). In the M3 mutant, we replaced the ATG sequence at position 2,303–2,305 bp with a CCC. This modification decreased sdsDNA synthesis below the detection limit of PAGE; however, sequencing of the P2/P6 PCR product indicated that the nucleotides CCC had become part of the sdsDNA (Fig. 6c). Similarly, when nucleotides AACAT at position 2,591–2,595 bp were replaced by CCACA in mutant M4, sdsDNA could not be detected by PAGE, although sequencing of the P2/P6 PCR product showed that the sdsDNA was one nucleotide longer than that found in the wild-type strain, corresponding to the initiation of reverse transcription from an alternative hairpin structure that had originated due to a slightly different base pairing in the priming center (Fig. 6). These experiments showed that although the synthesis of sdsDNA is most efficient in the wild-type strain, there was no strict requirement for a specific priming nucleotide or for a particular nucleotide to be inserted first into the (−)sdsDNA strand.

Discussion

In this study, we characterized the biosynthesis of sdsDNA, produced by the Se72 retron, and found out that the Se72 retron reverse transcriptase utilizes a completely different strategy for priming sdsDNA synthesis than any other retron reverse transcriptase. In the first step of sdsDNA biosynthesis, the 5′ UTR of the Se72 transcript folds into

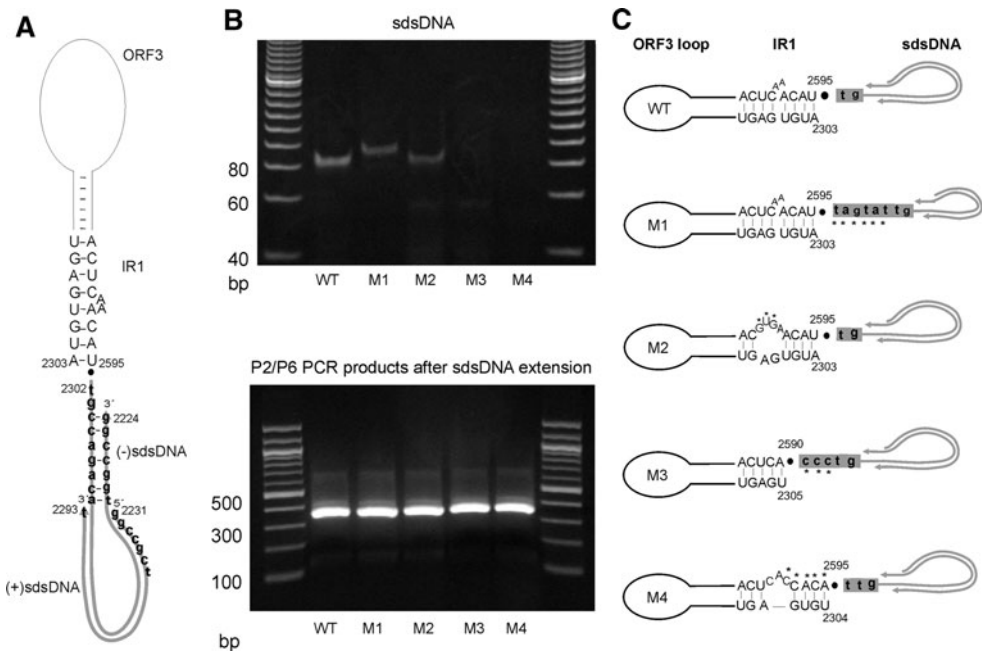

Fig. 6a–c A detailed model of the RNA/sdsDNA complex and analysis of the initiation of (−)sdsDNA synthesis in mutant strains. **a** sdsDNA in the native RNA/sdsDNA complex. A continuous *black line* represents the RNA transcript, *gray lines* represent sdsDNA strands with the 3′ ends pointed out by *arrows*, and a *black circle* represents 3′–5′ linkage. The *capital letters* indicate ribonucleotides forming IR1 and *lowercase letters* indicate deoxyribonucleotides forming sdsDNA. **b** Detection of sdsDNA extensively treated with RNase A and separated by nondenaturing PAGE and the PCR products obtained from sdsDNA extension products by agarose gel electrophoresis in mutants with a modified primary center. **c** An sdsDNA molecule in the native RNA/sdsDNA complex of wild type and mutant strains. *Gray boxes* highlight the nucleotides at the 5′ ends of the (−)sdsDNA strands that serve as a template for the extension of the 3′ ends of the same molecules by M-MLV reverse transcriptase. Nucleotides changed by the mutation are marked with *asterisks*

the secondary structure, leaving a free U_{2595} at the 3′ end of IR1 available for priming. Next, RrtE initiates the synthesis of a (−)sdsDNA strand from the 3′OH group of U_{2595} and proceeds along to the *msd* region. Template RNA from the RNA/(−)sdsDNA hybrid is degraded by cellular RNase HI, leaving only a short RNA/DNA hybrid at the 3′ end of (−)sdsDNA. Finally, the 3′ end of the remnant RNA serves as a primer for (+)sdsDNA synthesis using the (−)sdsDNA strand as a template (Fig. 4). The formation of double-stranded DNA using an RNA primer resistant to RNase H degradation is similar to the genome replication of retroviruses and long terminal repeats (LTR) containing retrotransposons. However, these elements utilize a polypurine tract RNA primer, 20–25 bp long, which is resistant to RNase H due to its specific sequence [13]. On the other hand, the RNA/DNA hybrid in the msRNA-msDNA complex is 5–8 bp long with no preference for purine ribonucleotides, and it is protected against RNase H degradation by retron reverse transcriptase [14]. The same mechanism we propose also for the (−)sdsDNA/RNA hybrid. Our results also showed that the sdsDNA never appears free of the RNA part of the complex. The structure shown in Fig. 6a therefore represents the biologically relevant molecule that accumulates in the cytoplasm of the Se72 retron host cell. On the other hand, the sdsDNA molecule shown in Fig. 1a represents an artificial molecule obtained after extensive in vitro treatment with RNase A.

The priming strategy of (−)sdsDNA synthesis is different from any other strategy described so far for bacterial retrons because it does not branch RNA by 2′–5′ phosphodiester bonds [6], and it uses a 3′OH end of its own RNA for priming instead. The priming nucleotide was confirmed by two independent protocols—initially by sequencing the extension products of sdsDNA in which U_{2595} was always found next to T_{2302}, the first nucleotide of the (−)sdsDNA strand, and then by the ligation-mediated PCR in which the synthetic oligonucleotide O1 was ligated to U_{2595}. Although in the second experiment we also observed ligation of the oligonucleotide O1 to RNA molecules shorter than those ending with U_{2595}, we never observed ligation to any position downstream of U_{2595}, and therefore we assumed that the ligation to shorter molecules was a consequence of nonspecific exonuclease degradation of the RNA molecule during its purification or the ligation. Furthermore, in the RNA isolated from the strain producing sdsDNA, the majority of RNA transcripts ending with 3′OH in U_{2595} should have been linked to the (−)sdsDNA strand, which also explains the rather rare ligation of oligonucleotide O1 to U_{2595}, especially when compared with the degradation products. How the free 3′ end suitable for

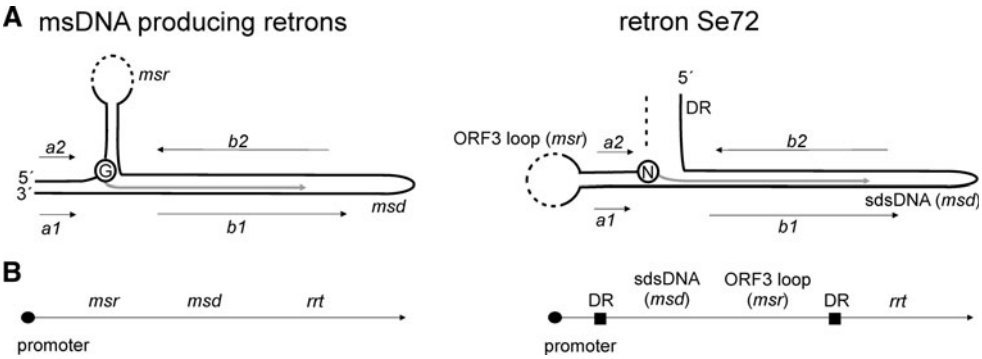

Fig. 7a, b A comparison of the conserved structure of msDNA-producing retrons and retron Se72. **a** Secondary structures of msDNA/RNA and sdsDNA/RNA molecules. A continuous *black line* represents the RNA transcript, and the *gray line* represents cDNA. **b** Genomic structures of msDNA-producing retrons and retron Se72

priming is generated remains unclear. Its generation is not dependent on RrtE or RNase HI, because Northern blot analysis showed that an RNA molecule of an appropriate size was also present in the *rrtE* and *rnhA* mutants. It is therefore likely that the 3′OH end of the RNA molecule is generated during transcription using the inverted repeats as a signal for transcription termination, although in some cases the transcription must extend into the *rrtE* gene sequence to allow for subsequent RrtE translation.

The priming strategy of the Se72 retron is similar to that of the LTR retrotransposon Tf1, which, unlike most other LTR retroelements, does not prime (−)cDNA synthesis by using tRNA. Instead, intramolecular self-complementarity near the 5′ end of its RNA transcript allows the formation of an 11-bp RNA duplex that is cleaved at the end of the duplex to produce a 3′ end suitable for priming (−)cDNA synthesis. However, additional characteristics of reverse transcription in the Tf1 retrotransposon are different from those of retron Se72, including the generation of the priming 3′OH, which in the case of the Tf1 retrotransposon is a function of the RNase H domain of its reverse transcriptase [15]. Since Tf1 and related LTR retrotransposons [16, 17] diverged early in the evolution of LTR retrotransposons and long before retroviruses, this mechanism was speculated to represent a tRNA-independent early form of reverse transcription initiation [18]. Although the Tf1 retrotransposon remains as the only retroelement that was confirmed to self-prime the cDNA strand synthesis from an intramolecular RNA duplex in vivo, the ability of reverse transcriptases to prime reverse transcription from the intramolecular RNA duplexes in vitro is quite well known and it is used, for example, for detecting the Dengue virus [19]. The Se72 retron is therefore the second genetic element to be found using the 3′ end self-priming strategy in vivo, and the first one of this type found in prokaryotes.

The Se72 retron appears to be completely different from all other bacterial retrons known to date. It lacks the generally conserved structure of overlapping *msd–msr* regions surrounded by an a1/a2 inverted repeat, it produces double-stranded DNA instead of single-stranded DNA, and it is plasmid encoded. Furthermore, it does not prime reverse transcription from the 2′OH group of a particular G residue within the RNA strand, but from the 3′ end of the RNA strand. Interestingly, in positions 2,218–2,223 bp and 2,595–2,600 bp, i.e., immediately upstream of the sdsDNA loop and downstream of the ORF3 loop, respectively, we identified 6 bp long sequences which, due to their palindromic structure, can be understood as being both direct and indirect repeats. The presence of this repeat (DR) may either indicate a recent inversion of the ORF3 and sdsDNA loops or it may indicate a recent insertion of the sequence.

If an inversion occurred within the direct repeat, the sdsDNA and ORF3 loops would exchange their orientation. After such a rearrangement, the Se72 retron would become structurally very similar to other bacterial retrons if ORF3 was understood as the *msr* region, the sdsDNA loop as the *msd* region, and IR1 as the a1/a2 repeat (Fig. 7) [20]. And since the reverse transcriptases of msDNA-producing retrons were described as being able to prime reverse transcription from the 3′ ends of either RNA or DNA molecules even if using their own retron-encoded msDNA molecule as a primer in vitro [21], it is possible that the Se72 retron differs from other bacterial retrons only in the relative positions of the *msd* and *msr* regions, leading to its unusual priming strategy in vivo.

Equally interesting is the second possibility, i.e., that the ORF3 and sdsDNA loops were inserted upstream of the Se72 retron reverse transcriptase relatively recently during evolution. This leads to a speculation that there must have been a plasmid that spread easily throughout bacterial populations and that only coded for the retron reverse transcriptase. Since we have shown the relatively weak preference of the Se72 retron reverse transcriptase for a sequence preceding the priming nucleotide, and because many prokaryotic genes use inverted repeats as signals for the termination of RNA transcription, the Se72 retron reverse transcriptase could catalyze the synthesis of different double-stranded DNAs that can be randomly inserted into a chromosome [22] or captured by different integron structures [23].

Acknowledgments This work was supported by project LC06030 and AdmireVet project CZ.1.05/2.1.00/01.0006-ED0006/01/01 from the Czech Ministry of Education and project MZE0002716202 from the Czech Ministry of Agriculture. We thank Scott W. Stevens (University of Texas, Austin) for providing the DbrI enzyme.

References

1. Joyce GF (1989) RNA evolution and the origins of life. Nature 338:217–224
2. Chiang CC, Kennell JC, Wanner LA, Lambowitz AM (1994) A mitochondrial retroplasmid integrates into mitochondrial DNA by a novel mechanism involving the synthesis of a hybrid cDNA and homologous recombination. Mol Cell Biol 14:6419–6432
3. Fedorova O, Zingler N (2007) Group II introns: structure, folding and splicing mechanism. Biol Chem 388:665–678
4. Liu MS, Gingery M, Doulatov SR, Liu YC, Hodes A, Baker S, Davis P, Simmonds M, Churcher C, Mungall K, Quail MA, Preston A, Harvill ET, Maskell DJ, Eiserling FA, Parkhill J, Miller JF (2004) Genomic and genetic analysis of *Bordetella* bacteriophages encoding reverse transcriptase-mediated tropism-switching cassettes. J Bacteriol 186:1503–1517
5. Dhundale A, Lampson B, Furuichi T, Inouye M, Inouye S (1987) Structure of msDNA from *Myxococcus xanthus*—evidence for a long, self-annealing RNA precursor for the covalently linked, branched RNA. Cell 51:1105–1112
6. Lampson BC, Inouye M, Inouye S (2005) Retrons, msDNA, and the bacterial genome. Cytogenet Genome Res 110:491–499
7. Rychlik I, Sebkova A, Gregorova D, Karpiskova R (2001) Low-molecular-weight plasmid of *Salmonella enterica* serovar Enteritidis codes for retron reverse transcriptase and influences phage resistance. J Bacteriol 183:2852–2858
8. Rychlik I, Karpiskova R, Faldynova M, Sisak F (1998) Computer-assisted restriction endonuclease analysis of plasmid DNA in field strains of *Salmonella enteritidis*. Can J Microbiol 44:1183–1185
9. Rychlik I, Svestkova A, Karpiskova R (2000) Subdivision of *Salmonella enterica* serovar Enteritidis phage types PT14b and PT21 by plasmid profiling. Vet Microbiol 74:217–225
10. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA 97:6640–6645
11. Nam K, Hudson RHE, Chapman KB, Ganeshan K, Damha MJ, Boeke JD (1994) Yeast lariat debranching enzyme—substrate and sequence specificity. J Biol Chem 269:20613–20621
12. Herzer PJ, Inouye S, Inouye M (1992) Retron-Ec107 is inserted into the *Escherichia coli* genome by replacing a palindromic 34 bp intergenic sequence. Mol Microbiol 6:345–354
13. Rausch JW, Le Grice SF (2004) 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. Int J Biochem Cell Biol 36:1752–1766
14. Furuichi T, Inouye S, Inouye M (1987) Biosynthesis and structure of stable branched RNA covalently linked to the 5′ end of multicopy single-stranded DNA of *Stigmatella aurantiaca*. Cell 48:55–62
15. Levin HL (1996) An unusual mechanism of self-primed reverse transcription requires the RNase H domain of reverse transcriptase to cleave an RNA duplex. Mol Cell Biol 16:5645–5654
16. Butler M, Goodwin T, Simpson M, Singh M, Poulter R (2001) Vertebrate LTR retrotransposons of the Tf1/sushi group. J Mol Evol 52:260–274
17. Malik HS, Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. J Virol 73:5186–5190
18. Le Grice SF (2003) "In the beginning": initiation of minus strand DNA synthesis in retroviruses and LTR-containing retrotransposons. Biochemistry 42:14349–14355
19. Tuiskunen A, Leparc-Goffart I, Boubis L, Monteil V, Klingström J, Tolou HJ, Lundkvist A, Plumet S (2010) Self-priming of reverse transcriptase impairs strand-specific detection of dengue virus RNA. J Gen Virol 91:1019–1027
20. Hsu MY, Inouye S, Inouye M (1989) Structural requirements of the RNA precursor for the biosynthesis of the branched RNA-linked multicopy single-stranded DNA of *Myxococcus xanthus*. J Biol Chem 264:6214–6219
21. Lampson BC, Viswanathan M, Inouye M, Inouye S (1990) Reverse transcriptase from *Escherichia coli* exists as a complex with msDNA and is able to synthesize double-stranded DNA. J Biol Chem 265:8490–8496
22. Lampson BC, Rice SA (1997) Repetitive sequences found in the chromosome of the myxobacterium *Nannocystis exedens* are similar to msDNA: a possible retrotransposition event in bacteria. Mol Microbiol 23:813–823
23. Recchia GD, Hall RM (1997) Origins of the mobile gene cassettes found in integrons. Trends Microbiol 5:389–394