Multi-copy single-stranded DNA in Escherichia coli

Xianxing Xie and Ruifu Yang*

Abstract

Multi-copy single-stranded DNA (msDNA) is composed of covalently bound single-stranded DNA and RNA, and synthesized by retron-encoded reverse transcriptase. msDNA-synthesizing systems are thought to be a recent acquisition by Escherichia coli because, to date, only seven types of msDNA, which differ markedly in their primary nucleotide sequences, have been found in a small subset of E. coli strains. The wide use of E. coli in molecular research means that it is important to understand more about these stable, covalently bound, single-stranded DNA or RNA compounds. The present review provides insights into the molecular biosynthesis, distribution and function of E. coli msDNA to raise awareness about these special molecules.

For many years, reverse transcriptase (RT) was thought to exist only in eukaryotic organisms. However, in the late 1980s, several studies reported the existence of multi-copy single-stranded DNA (msDNA) in Myxococcus xanthus, consisting of covalently bound single-stranded DNA and RNA molecules that are synthesized by RT [1–3]. msDNA has been widely identified in myxobacterial strains, suggesting that it may have originated from an ancestral myxobacterium [4, 5]. In contrast, msDNA is rarely found in Escherichia coli, and its function in this species is also not clear. However, recent studies indicate that msDNA could represent a new class of regulatory molecules, possibly involved in directing intestinal colonization of enteropathogenic E. coli by regulating protein abundance [6]. Moreover, a newly developed method allows large-scale annotation of msDNA, by scanning the National Center for Biotechnology Information RefSeq bacterial genome database [7]. Together, these findings will help accelerate research into the physiological role of msDNA.

DISCOVERY OF MSDNA

Myxobacteria are Gram-negative bacteria that have been used as model systems of developmental biology. Total genomic DNA preparations of M. xanthus were occasionally found to contain 500–700 copies per chromosome of a peculiar satellite DNA, identified as an approximately 160 bp single-stranded DNA molecule [8]. Subsequent studies revealed that this molecule, designated msDNA-Mx162, consist of single-stranded RNA and DNA linked by a 2',5'-phosphodiester bond at the 2' position of the 20th rG residue [9]. msDNA-Sa163, found in Stigmatella aurantiaca, showed a high degree of homology to msDNA-Mx162 [4].

Previous studies have ruled out the presence of msDNA in E. coli [8, 9]. However, msDNA-EC67 and msDNA-EC86, which show no primary RNA or DNA sequence homology, were identified in E. coli clinical strains Cl-1 and B, respectively [10, 11]. To date, seven types of msDNA have been reported in E. coli – msDNA-EC48, msDNA-EC67, msDNA-EC73, msDNA-EC78, msDNA-EC83, msDNA-EC86 and msDNA-EC107.

DISTRIBUTION OF MSDNA

Previous screening experiments confirmed the presence of msDNA in 7/113 clinical E. coli isolates (~6 %) and in 9/72 environmental E. coli isolates (~13 %), demonstrating that only a small percentage of E. coli strains carries msDNA [12, 13]. Although the common laboratory strain E. coli K-12 does not harbour msDNA, msDNA was detected in an E. coli K-12 strain transformed with msDNA-synthesizing systems [11]. Furthermore, retronphage ΦR73 integrated its genome into different host strains at the same chromosomal location, thereby enabling the host E. coli to produce msDNA-EC73 [14]. This may answer the question on the origin of retrons, and could be direct proof of horizontal transmission of retrons.

Because of the highly sporadic appearance of msDNA in host genomes, combined with the lack of homology between the different types of molecule, Sun et al. speculated that retrons...
were acquired fairly recently during the evolution of E. coli [13]. Two virtually indistinguishable E. coli strains, ECOR-35 and ECOR-36, which share a very close genetic relationship, did not coordinate with each other in synthesizing msDNA. This indicates that the genes responsible for msDNA synthesis in these two strains were lost in an independent manner, or were only obtained fairly recently [12]. In addition, codon usage within the RT gene mediating msDNA synthesis in a clinical E. coli strain was quite different from that of most other E. coli proteins, supporting the notion that retrons are ‘foreign’ genes within the E. coli genome [11]. The DNA sequences upstream of retron-EC48, which produces the shortest naturally occurring msDNA, showed high identity to bacteriophage P2 and P4 genes, indicating that retron-EC48 may be transmitted by a retropophage [15]. Varmus even proposed that the retron was a part of a prophage and was transmitted between cells by conjugation or transduction [1]. However, retron-EC107 inserts into the E. coli genome by replacing a 34 bp palindromic sequence without introducing any prophage-like sequence [16, 17]. Furthermore, retrons may be clonally inherited because E. coli strains of the same serotype all contained the same msDNA, although not all E. coli strains belonging to a certain serotype harboured msDNA [18]. Comparative study of different msDNA structures and phylogenetic analysis of RT sequences also supported vertical inheritance rather than horizontal transformation [19].

**STRUCTURE OF MSDNA MOLECULES**

The 2',5'-phosphodiester bond, located between the 2' position of the rG of the single-stranded RNA and the 5' end of the single-stranded DNA, is a hallmark structure in msDNA molecules (Fig. 1). The rG position can differ between various msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded DNA, is a hallmark structure in msDNA molecules (Fig. 1). The rG position can differ between various msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded DNA, is a hallmark structure in msDNA molecules (Fig. 1). The rG position can differ between various msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded DNA, is a hallmark structure in msDNA molecules (Fig. 1). The rG position can differ between various msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded DNA, is a hallmark structure in msDNA molecules (Fig. 1). The rG position can differ between various msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20]. The single-stranded DNA, is a hallmark structure in msDNA molecules (Fig. 1). The rG position can differ between various msDNA molecules, but the replacement of rG by either rC or rA blocked the production of msDNA [20].

Although the 2',5'-phosphodiester bond is the characteristic feature of msDNA molecules, retrons EC78 and EC83 can both produce RNA-less msDNA. In both in vivo and in vitro experiments, two forms of msDNA-EC83 were detected: a 79 bp single-stranded DNA molecule and an 83 bp DNA molecule covalently linked to the RNA molecule as shown in Fig. 2(a, b), respectively [22]. Similar structures were observed for msDNA-EC78, with both a 78 bp RNA-DNA compound and a 74 bp single-stranded DNA molecule being produced [23]. Comparison of the sequences of msDNA-EC78 and msDNA-EC83 showed that they share the same endonucleolytic cleavage site between the fourth and fifth nucleotides of the DNA component of the RNA-DNA compound [22, 23]. Although the mechanism of cleavage is unclear, no proteins other than the RT are involved. Both the RT and MutS binded to the single-stranded DNA, forming a stable complex, rather than the RNA of msDNA [24, 25].

**SYNTHESIS OF MSDNA**

Retrons consist of a single ~2000 bp operon containing three genes: msr, msd and RT (msr-msd-RT) (Fig. 3). The primary transcript folds into a stable stem structure at inverted repeat sequences a1 and a2, placing a G residue at the end of the a1–a2 stem (Fig. 3). The priming reaction for the msDNA is thought to be initiated from the 2'-OH group of this G residue located after the RT protein, followed by cDNA synthesis using the same RNA transcript as a template (Fig. 3) [26, 27]. The RNA template is finally cleaved by RNase H, leaving a heteroduplex structure at the 3' end of the DNA. Other than the a1–a2 stem and the G residue, the only strict requirement is that the msr region and RT are derived from the same retron, as the msd region is exchangeable [26, 28]. This specificity between the msr region and RT may be attributed to the
C-terminal region of the RT, as this plays a major role in the specific recognition of the primer-template RNA [29]. When msDNA is used as a template, this limitation may be ignored, resulting in two major products: a single-stranded DNA extended by the 3′ end of the msDNA using msdRNA as a template, and a double-stranded DNA in which the 3′ end of the msdRNA is extended using msDNA as a template [30].

**FUNCTION OF MSDNA**

Despite the considerable number of studies that have been conducted since the initial identification of msDNA in *E. coli*, its biological function remains unclear. Proteomic analysis revealed that proteins related to the dissimilation of various carbon sources were repressed in msDNA-producing cells with slow growth [31]. Because no obvious disease symptoms were observed in individuals whose stool samples contained msDNA-producing *E. coli*, Sun *et al.* suggested that msDNA may not be associated with any disease phenotype [13]. Based on recent results showing that both msDNA-Ec83 and msDNA-Ec78 are in vivo substrates of exonuclease VII (ExoVII), we propose that high copy numbers of msDNA may block the apoptotic-like cell death caused by ExoVII subunits [32]. In addition, msDNA may have Dam methylase-like activity based on the high degree of sequence identity between Dam methylase and retron-EC67 [33].

The high mutation rates in msDNA-producing cells have attracted considerable attention; however, the molecular mechanism remains elusive. All *E. coli* msDNA molecules, apart from msDNA-EC78, with mismatched base pairs in their stem-loop structures are mutagenic when present in high copy numbers as they titrate out MutS [24, 34]. Studies have shown that starvation may enhance mutation rates by elevating the abundance of msDNA, as RT alone is not mutagenic [35–38]. Furthermore, RT may compete with MutS in binding to msDNA, thereby controlling the mutation rate.

**PERSPECTIVE**

msDNA is widely distributed in myxobacteria isolated from soil lacking in a variety of nutrients. msDNA is thought to be produced by myxobacteria under starvation pressure to elevate mutation rates, with the aim of generating beneficial mutations. If this is the case, are there any molecules equivalent to msDNA that are induced by other stress conditions?

In general, DNA and RNA remain independent from one another within cells because they are synthesized by different enzymes and have their own efficient repair systems. However, in the special case of msDNA, the stable, covalently bound, single-stranded DNA and RNA compounds can prime each other, with RT completing DNA replication and RNA reverse transcription simultaneously. Further
research on msDNA could help to develop a new genetic message delivery system.

**Funding information**
This work was supported by National Special Project on Research and Development of Key Biosafety Technologies Grant 2016YFC1200100.

**Acknowledgements**
We thank Tamsin Sheen, PhD, from Liwen Bianji, Edanz Group China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

**Conflicts of interest**
The authors declare that there are no conflicts of interest.

**References**


