Structure of two retrons of *Escherichia coli* and their common chromosomal insertion site

D. Lim

Department of Microbiology, New York University Medical Center, 550 First Avenue, New York, New York 10016, USA.

Summary

It has been shown that certain strains of myxobacteria and of *Escherichia coli* have a genetic element encoding a reverse transcriptase (RT). This element, called a 'retron', produces a covalently linked RNA–DNA compound (msDNA–RNA). Here, I report the complete nucleotide sequence of retron EC-86, the retron in *E. coli* B, together with its flanking regions. Retron EC-86 contains genes for msDNA–RNA (*msd*, and *msr*), a gene for RT (*ret*) and a gene for an open reading frame whose function is unknown. The upstream junction is composed of the sequence GCGGCCGC, but there are no direct or inverted repeats at the retron–host junctions. It is also shown that another retron of *E. coli*, EC-67, which was isolated originally from the clinical strain CL1 and was later found to be present also in a clinical *E. coli* isolate from Brazil, is inserted at the same chromosomal site as retron EC-86. Retron EC-67 contains only *msd*, *msr*, and *ret*. I suggest that these two retrons were independently inserted into the same site of their host strains via a novel mechanism of integration.

Introduction

Over the last 10 years, it has been shown that many different types of genetic elements utilize reverse transcription in their life cycles. These elements were found in virtually all eukaryotic organisms and they range from plasmids to viruses. Recently, it has been shown that some bacteria, such as *Escherichia coli* B or *Myxococcus xanthus*, have an element which encodes RT, as reviewed by Varmus (1989), Temin (1989), Lim and Maas (1989a; Lampson et al., 1989), were characterized in detail. Although their nucleotide sequences are not related to each other, except for the msDNA–RNA compounds from *Stigmatella aurantiaca* and *M. xanthus* (Dhundale et al., 1987), they have common structural features: msDNA and msdRNA can form stable secondary structures. msDNA is covalently linked to an internal guanine residue of msdRNA by a 2'-5' phosphodiester bond. The 3' ends of msDNA and msdRNA form an RNA–DNA hybrid. Because of this RNA–DNA hybrid region at the 3' ends, msDNA–RNA compounds serve as excellent template-primers for retroviral RTs or for the RTs encoded by *E. coli* (Lim and Maas, 1989a; Lampson et al., 1990).

It has been shown that a 3.5 kb *PstI* fragment of the *E. coli* B chromosome is sufficient for the production of an msDNA–RNA compound in *E. coli* K12 (Lim and Maas, 1989a). The minimum region within this fragment required for the synthesis of this compound consists of genes for msDNA (*msd*), for msdRNA (*msr*) and for an ORF which has sequence homology with retroviral RTs (Lim and Maas, 1989a). We showed that the protein encoded by this ORF has RT activity *in vitro*, and that the msDNA in *E. coli* B is synthesized via a novel mechanism of reverse transcription (Lim and Maas, 1989a). In *E. coli* B, *msr*, *msd*, and the gene for RT, which we shall refer to as *ret*, are transcribed from a single promoter located upstream of *msr*. The primary transcript covering *msr*, *msd*, and *ret* serves both as a mRNA for RT and a template for reverse transcription (Lim and Maas, 1989a). This transcription unit has been called the msDNA operon (Lim and Maas, 1990). In most msDNA synthesis systems, the organization of the genes involved in msDNA synthesis and the mechanism of msDNA synthesis by reverse transcription seem to be similar to the *E. coli* B system summarized above (Lampson et al. 1989; Inouye received 10 January, 1991; revised 27 March, 1991. Tel. (212) 263 1113; Fax (212) 263 8276.)
et al., 1989). Based on these similarities, the term 'retron' was introduced to describe the bacterial reverse transcription system (Temin, 1989). It has been proposed that the ret gene in myxobacteria is as old as the other myxobacterial genes and that retrons may therefore be primitive forms of retroelements (Temin, 1989; Inouye and Inouye, 1991). In contrast to myxobacteria, in E. coli retrons seem to be recent arrivals and can be considered as foreign elements (Lim and Inouye, 1991). To gain a better understanding of the nature of retrons in E. coli, I have analysed the host–retron junctions and functional components of two retrons.

Recently, Hsu et al. (1990) localized the retron of the clinical E. coli strain, CL1 (retron EC-67, for nomenclature, see Lampson et al., 1989), at a chromosomal site that corresponds to minute 19 on the chromosome of E. coli K12. The element inserted at that site is 34 kb long and is flanked by 26-bp direct repeats. These authors suggested that the 34 kb segment constitutes the retron and that it was inserted by a mechanism similar to transposon insertion. We have shown that the retron of E. coli B (retron EC-86) is also inserted at minute 19 of the E. coli chromosome (Lim and Maas, 1990). In this paper I show that retrons in EC-67 and EC-86 are only a small part of the DNA segment inserted at minute 19. My results suggest that the retrons were inserted independently into the same site of the 34 kb element by an unknown mechanism.

Results

Some clinical E. coli strains, but not E. coli K-12, have an insertion site for the E. coli B retron

As a first step towards understanding the nature of the element encoding RT in E. coli I decided to define the precise length and the functional components of the element encoding RT in E. coli B. The ideal approach would be to compare the DNA sequences of two strains which are isogenic except for the presence of the element in its chromosome. Initially the chromosomal region adjacent to RT was compared with the corresponding region of E. coli K12, assuming that E. coli K12 is a RT-negative reference strain. During this study, it was found that the msDNA operon of E. coli B is present in a DNA segment which is inserted into minute 19 of the E. coli K12 genetic map (Lim and Maas, 1990). It was shown that this inserted segment is longer than 15 kb, which is much longer than the msDNA operon itself (about 1.3 kb, see Lim and Maas, 1989a). Therefore, at this point it was not clear what portion of this inserted segment is functionally and evolutionarily related to the retroelement itself. Subsequently it was found that the DNA fragment flanking the E. coli B ret gene hybridized to DNA isolated from several clinical E. coli strains, although the same probe did not hybridize to DNA from the K12 strain (see below). In view of these observations, I decided to compare the ret locus of E. coli B with the corresponding region of clinical strains, rather than with that of the K12 strain.

Since the 3.5 kb PstI fragment in the original clone pDB808 (Lim and Maas, 1989a) was not big enough for this purpose (see below), a bigger fragment surrounding the ret gene was cloned into a lambda vector (Lim and Maas, 1990). The restriction map of the E. coli B chromosome bordering the ret gene and the location of the probes used for the Southern hybridizations are shown in Fig. 1A. Chromosomal DNA from 21 different E. coli strains (representative strains of our collection, see Table 1) were isolated, digested with EcoRI and transferred to a membrane after agarose gel electrophoresis. Three different probes were used for hybridizations (Fig. 1A): probe I (1.2 kb fragment from position 2110 to 3353 of Fig. 2D) contains the minimal DNA fragment required for msDNA synthesis in E. coli K12 (Lim and Maas, 1989a). Probe II is the 3.5 kb PstI fragment, which covers 2.5 kb upstream of the ret gene. Probe III is a 5.8 kb XhoI–EcoRI fragment, which covers the downstream region of the ret gene. Probe I hybridized only with the DNA isolated from E. coli B (see Fig. 1B), showing that the nucleotide sequence of msd, msr, and ret of E. coli B is unique among the strains tested. Interestingly, although neither probe hybridized with DNA from the K12 strain, probes II and III hybridized with the DNA isolated from five E. coli strains (Fig. 1C). Since none of these strains hybridized with probe I, as shown in Fig. 1B, positive hybridization with these two probes is due to the presence of DNA fragments homologous to the upstream (in the case of

<table>
<thead>
<tr>
<th>Strain no.</th>
<th>Serotype</th>
<th>msDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>K12</td>
<td>?</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>?</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>317</td>
<td>RH2</td>
<td>+</td>
</tr>
<tr>
<td>331</td>
<td>O119:H6</td>
<td>+</td>
</tr>
<tr>
<td>32</td>
<td>O119:H6</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>O119:H6</td>
<td>-</td>
</tr>
<tr>
<td>67</td>
<td>O119:H6</td>
<td>-</td>
</tr>
<tr>
<td>323</td>
<td>O55:H+</td>
<td>+</td>
</tr>
<tr>
<td>40</td>
<td>O55:H6</td>
<td>+</td>
</tr>
<tr>
<td>41</td>
<td>O55:H7</td>
<td>+</td>
</tr>
<tr>
<td>57</td>
<td>O26:H+</td>
<td>+</td>
</tr>
<tr>
<td>64</td>
<td>O126:H9</td>
<td>+</td>
</tr>
<tr>
<td>45</td>
<td>O127:H6</td>
<td>-</td>
</tr>
<tr>
<td>322</td>
<td>O92:H33</td>
<td>+</td>
</tr>
<tr>
<td>332</td>
<td>O142:H6</td>
<td>+</td>
</tr>
<tr>
<td>49</td>
<td>O142:H6</td>
<td>-</td>
</tr>
<tr>
<td>68</td>
<td>O142:H6</td>
<td>-</td>
</tr>
<tr>
<td>70</td>
<td>O126:H27</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>O86:H34</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>O110:H+</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. E. coli strains used in this study.
Fig. 1. A. Restriction map of the msDNA operon of *E. coli* B and its flanking regions. The msDNA operon of *E. coli* B is present in a DNA fragment which is inserted at minute 19 of the *E. coli* K12 chromosome (Lim and Maas, 1990). The solid bar indicates the minimum fragment required for msDNA synthesis. The locations of the three probes used are shown. Restriction enzymes are: H, HindIII; P, PstI; X, XhoI; R, EcoRI.

B. Southern hybridizations with probe I. Southern hybridizations were performed with DNA isolated from the strains which were positive in the hybridizations with probe II and III.

C. Southern hybridizations with probes II and III. Chromosomal DNA was isolated from 21 *E. coli* strains (indicated by numbers or letters above each lane), digested with EcoRI, electrophoresed in a 0.7% agarose gel, and transferred to a membrane. First the membranes were used for the hybridization with probe II (upper panel) and then with probe III (lower panel) after washing off probe II. The plus or minus sign on each lane indicates the presence or absence of msDNA. Approximate sizes of hybridized bands are indicated on the left-hand side of each gel in kb. In strain 49, the sizes of the bands recognized by probe II and probe III are different (7 kb versus 30 kb), showing that an EcoRI recognition site(s) is present between these two bands. Note that the sizes of bands in the lanes of strains 331 and 25 (both strains belong to serotype O119:H6) are identical, but they are different from the sizes of bands in the lanes of strain 317.
hybridization with probe II) and the downstream (in the case of hybridization with probe III) regions of the E. coli B ret gene. Among these hybridization-positive strains, three produce msDNA (strain 317, serotype R:H2, and strains 331 and 25, both serotype O119:H6). We have shown that the msDNAs in these strains are different from that of E. coli B (Lim et al., 1990), which is in agreement with the present negative hybridization results with probe I, which contains only the msDNA operon of E. coli B (Fig. 1B). Six out of eight msDNA-positive clinical strains did not hybridize with either probe, showing that in these strains not only the ret genes but also their flanking sequences are different from those of the E. coli B ret gene. Although none of the probes hybridized with DNA from the K12 strain, two of the msDNA-negative strains (49 of O142:H6 and 70 of O126:H27) clearly hybridized with probes II and III (Fig. 1C). From these results I conclude that the hybridization-positive strains have an insertion site for the E. coli B retron which is not present in E. coli K12. It was expected that a comparison of the DNA sequences flanking the ret gene of E. coli B with the corresponding DNA sequences hybridizing with probes II and III would clarify the insertion site and the organization of the E. coli B retron. Strain 317 was chosen for comparison, since it gave the strongest signal in the hybridizations.

Retrons EC-86 and EC-67 are inserted at the same site

A genomic library of strain 317 was constructed in a λ vector and screened by plaque hybridization using the 3.5 kb PstI fragment of pDB808 as a probe (probe II in Fig. 1A). Of the positive clones, two recombinant lambda phages were analysed in detail. The restriction map of the chromosomal fragment derived from the restriction maps of the two recombinant phages is shown in Fig. 2A. Since probe II and probe III hybridized with a 2 kb HindIII fragment and a 4 kb HindIII–EcoRI fragment, respectively, as shown in Fig. 2A, the nucleotide sequences flanking the HindIII site within the ret gene of 317 were determined. The nucleotide sequences of the upstream and the downstream flanking regions of the E. coli B ret gene were also determined. The nucleotide sequences derived from these two strains were compared, with the results shown in Fig. 2 (B–E).

As shown schematically in Fig. 2A, the nucleotide sequences derived from these two strains are composed of an upstream homologous region, a unique middle region (indicated by closed bars in Fig. 2A) and a downstream homologous region. The unique region of strain 317 (nucleotide sequence from 1152 to 3226, Fig. 2E) contains a long ORF that codes for a protein of 665 amino acids. A homology search of GenBank showed that this ORF is identical to the previously reported RT ORF of the clinical strain, E. coli CL1 (Lampson et al., 1989). Subsequently, it was found that the entire nucleotide sequence of the unique region is identical to the sequence of strain CL1. However, differences between strain CL1 and strain 317 are found in the homologous flanking regions: an insertion of a guanine at position 1072, and two insertions (adenine and thymine) at positions 3313 and 3319 (Fig. 2B and 2C). It was found that E. coli JM109 harbouring a plasmid containing the 6.5 kb Xhol–EcoRI fragment of strain 317 (Fig. 2A, the Xhol site is in the vector) produces an msDNA (data not shown), showing that, as in strain CL1, this region is involved in msDNA synthesis. According to the studies with strain CL1 (Lampson et al., 1989), the unique region of strain 317 codes for msdRNA (from position 1214 to 1271), msDNA (from position 1265 to 1331), and RT (starting from the ATG at position 1383, which is the third ATG within the ORF).

Within the unique region of E. coli B (nucleotide sequence from 1152 to 3325, see Fig. 2D), besides msd, msr, and ret, one more ORF is found. This ORF codes for a protein composed of 223 amino acids (ORF223). I could not find any homologous proteins in a protein database by means of computer searches with programs FASTP (Lipman and Pearson, 1985) or Wordsearch (Devereux et al., 1984). At present, I do not know the function of the protein encoded by this ORF. It is not directly involved in msDNA synthesis since we have shown that the 1.2 kb fragment (from 2110 to 3353 in Fig. 2D) containing only msd, msr, and ret is sufficient for msDNA synthesis in E. coli K12. It is interesting that, in spite of this extra ORF, the entire length of the unique region of strain B is remarkably similar to that of strain 317 (2174 bp in strain B versus 2075 bp in strain 317); the presence of ORF223 is compensated for by the presence of the short RT ORF in E. coli B (665 amino acids in strain 317 versus 320 amino acids in strain B). It is worth mentioning that, in contrast to the RT of retron EC-67 (see Lampson et al., 1989), the RT of retron EC-86 does not have an RNase H domain. So far it is the smallest RT known among RTs, including other RTs for msDNA synthesis.

The upstream junction between the homologous and the unique sequences is located at position 1152 and contains sequence GGGGGCGC (Fig. 2B). At present, the biological significance of this palindromic junctional sequence is unknown. In the upstream homologous region, an ORF composed of 111 amino acids was found (from position 746 to position 414 in Fig. 2B). This ORF is in opposite orientation from that of the RT ORF. In the two strains, the nucleotide sequences encoding ORF111 are identical, suggesting that the function of this ORF may be important. In fact, the overexpression of ORF111 efficiently eliminates the element producing the msDNA–RNA compound (see below). The nucleotide sequence between ORF111 and the upstream junction (from 746 to
1151), is also well-conserved relative to the downstream homologous region. There are five repeats of a TGCATG sequence in this region (Fig. 2B), whose significance is presently unknown.

The overall homology observed in the downstream flanking region is lower than that observed in the upstream flanking sequence. Downstream, a short, moderately homologous region, 72 nucleotides long (with 73% identity between the two strains, underlined in Fig. 2C), is present between the unique region (no homology at all) and the highly homologous region (92% identity, starting at position 3399 of the E. coli B sequence in Fig. 2C). In contrast to the upstream junction, this gradual increase in homology in the downstream region makes it difficult to locate the homologous-unique junction, but it should be between positions 3325 and 3399 (Fig. 2C). I believe it to be located at position 3325, since strong homology in the amino acid sequence of ORF336 between the two strains starts at this position. ORF336 in strain B and in 317 share 93% identity in terms of their amino acid sequences. We do not know the function of this protein, although its conservation in the two strains suggests that it is functional. A computer search with data base NBRF did not reveal any protein showing significant homology with ORF336.

**Efficient curing of the retron by overexpression of ORF111**

Since there was an indication from the sequencing data that ORF111 might encode an integrase, I carried out the following experiments. If the ORF111 encodes a site-specific recombinase (integrase), it is to be expected that the overexpression of this protein may lead to excision of the element. In order to test this hypothesis, two plasmids were constructed: the first plasmid, pDB1110, was created by insertion of a 950 bp \( Hpal-HindIII \) fragment (from the HindIII site in Fig. 2A to the \( Hpal \) site at position 803 in Fig. 2B) into the HindIII–HincII sites of vector pTZ18R. In this insertion, intact ORF111 is expressed under the control of the lac promoter of the vector. The second plasmid, pDB1109, was constructed by insertion of an 870 bp \( SspI-HindIII \) fragment (nucleotide sequence from 737 to HindIII) into the same sites of the vector. In this control plasmid, two \( N \)-terminal amino acids of ORF111 are deleted and the remaining portion is not in frame with lacZ\( ^{\prime} \) of the vector. Therefore, no expression of ORF111 is expected.

Both plasmids were transformed into the *E. coli* B strain, AC2514, and independent transformants were tested for the production of msDNA–RNA after the induction of the lac promoter, as described in the Experimental procedures. The results of one such test are shown in Fig. 3B. In this particular experiment, 6 out of 7 transformants containing pDB1110 no longer produce msDNA–RNA. Overall, among the 18 transformants tested, 14 colonies did not produce the msDNA–RNA compound. In contrast, all 18 colonies transformed by the control plasmid pDB1109 produced the msDNA–RNA compound. In order to verify that the failure of msDNA synthesis in these transformants is due to the absence of genes rather than to changes in the control of msDNA synthesis by the ORF111 product, Southern hybridization was performed. As shown in Fig. 3C, the chromosomal DNA isolated from five 'cured' strains does not hybridize with a DNA fragment containing msd, msr, and ret genes of *E. coli* B (probe I of Fig. 1A). Interestingly, the DNA isolated from cured strains does not hybridize even with probe II or III (data not shown). This shows that the loss of msDNA production in the cured strains is due to a deletion that is much larger than the msDNA operon.
Discussion

Integration of two retrons at the same site

Since the discovery of RT in bacteria, the question of whether a RT in bacteria is a product of a cellular gene, as proposed for *M. xanthus* (Inouye et al., 1989), or part of a separate genetic element, as proposed for *E. coli* (Lim and Maas, 1989a), is an interesting one. In the experiments presented in this paper, it is assumed that the gene for *E. coli* B RT is part of a separate genetic element that is inserted into an *E. coli* genome. With this hypothesis in mind, experiments were performed to determine the host–parasite junctions, the precise length of the element, and the functional components in the element encoding *E. coli* B RT.

Since it was found that *E. coli* K12 was not appropriate as a reference strain for this purpose, Southern hybridizations were performed to search for strains that might fulfill this purpose. From one strain the DNA fragment homologous to both the upstream and the downstream flanking regions of the retron of *E. coli* B was cloned and sequenced. When the nucleotide sequence derived from this strain was compared with that of *E. coli* B, it was found that in both strains a unique sequence of similar size is flanked by homologous sequences. Subsequently, it was shown that in both strains the unique region codes for the genes involved in msDNA synthesis, showing that two different retrons (retron EC-86 in *E. coli* B and retron EC-67 in strain CL-1 or in strain 317) are inserted into the same site of the chromosome. We do not know how these different retrons were integrated at this site.

Organization of retrons EC-86 and EC-67: retron as a cassette

The unique region of *E. coli* B contains the genes *msd*, *msr*, and *ret*. In addition to these genes, it has an extra ORF (ORF223). It is believed that the protein encoded by this ORF is not involved directly in msDNA–RNA synthesis, in view of the following facts. Previously, we have shown that *E. coli* K12 harbouring a plasmid containing *msd*, *msr*, and *ret* efficiently produces the msDNA–RNA compound. Since probe II, which covers ORF223, did not hybridize to *E. coli* K12 DNA, the possibility that the host strain complements this protein in msDNA synthesis can be ruled out. However, the following observations strongly suggest that with regard to its origin ORF223 is related to the *ret* gene. First, the GC content of the ORF223 coding region is 39%, which is similar to the GC content of the genes for msDNA, msdRNA, and RT (37%). In contrast to this, the GC content of the flanking homologous regions is about 48%, which is similar to the GC content of the *E. coli* genome (49%). Second, the codon usage of ORF223 is similar to that of the RT ORF, rather than to that of cellular genes or to that of ORFs in the flanking homologous regions (data not shown). It is known that the codon usage of *E. coli* RT is different from that of other chromosomal genes (Lim and Maas, 1989a; Lampson et al., 1989). This suggests that the genes in the unique region of *E. coli* B are evolutionarily related to each other and that the entire unique region originated separately from the homologous flanking regions. On the basis of these facts, it is concluded that the unique region defines retron EC-86.

The GC content of the unique region of strain 317 is also significantly different from that of the flanking homologous region (39% in the unique region versus 49% in the homologous region). Again, the codon usage of RT in the unique region is significantly different from that of ORFs in the flanking region. The unique region in this strain contains only genes encoding msDNA synthesis. These results show that the unique region of strain 317 is evolutionarily and functionally separate from the homologous flanking region. Thus, the element encoding RT in strain...
shown in Fig. 2. Although the retron EC-86 has an extra ORF (ORF223), the entire length of the retrons in the two strains is remarkably similar. The presence of an extra ORF in retron EC-86 is compensated for by the presence of a small RT ORF. The E. coli B RT does not have an RNase H domain and, thus, cellular RNase H is required for msDNA synthesis (D. Lim, in preparation). This similarity in length in the two retrons suggests that there may be a size limitation (packaging?) during the integration of a retron into the E. coli chromosome.

**Different retrons in E. coli have different origins**

There could be two possible explanations for the generation of the homologous-unique-homologous arrangement observed in strains B and 317. First, it could occur by insertion of different retrons independently ('independent insertion model') or, second, it could be generated by 'hyper-mutability' of an ancestral element during the evolution of the two strains. For example, if repeated cycles of reverse transcription and insertion of a reverse transcription product into the chromosome via homologous recombination occurred, the retron could diverge much faster than the flanking regions, since such cycles would occur many times per generation, and both transcription by RNA polymerase and reverse transcription by RT would generate more errors than the replication by DNA polymerase. This model predicts that, first, the middle of the retrons must diverge faster than the distal parts, since in general homologous recombination between the chromosome and the reverse transcription product, the chance of replacement of the middle region is higher than that of the distal part. Second, the mutation rate of the element should be greater than that of the flanking regions. The first prediction is made unlikely by the abrupt change from an almost identical sequence to the unique sequence at the 5' junction. The second prediction is not supported by my data. The nucleotide sequence of the strain 317 retron shown here is identical with the previously reported nucleotide sequence of the strain CL1 msDNA element, whereas there are differences in the flanking regions. From these considerations, it is clear that the retrons in strain 317 and strain B were independently inserted after separation of their host strains. This independent insertion model is further supported by observations in other clinical strains; most msDNA elements in our clinical strains have flanking sequences different from those of strain B or strain 317 (Fig. 1C), showing that these other retrons were inserted independently from retrons in strains B and 317. The present data suggest that retrons utilize a novel mechanism of integration. First, we do not see any direct or inverted repeats at the element-host junctions (see Fig. 2). Therefore there is no target duplication, which is frequently observed in transposon insertions. Second, there seems to be a size limitation for the inserted elements since the sizes of retrons in strain B and strain 317 are similar.

At present, it is unknown whether retrons in E. coli have the capacity to move horizontally between strains. The presence of identical retrons in strains 317 and CL1 suggests that retrons may be mobile, but it has been shown that the two strains actually belong to the same clone (T. Whittam, unpublished data). This strongly suggests that the common ancestor of strains CL1 and 317 carried retron EC-67 and that subsequently it was vertically transmitted. At present it is unknown where EC-67 and other retrons originated and how they were inserted into E. coli.

**Retrons in strains B and 317 may be inserted into a genetic element**

From the comparison of the RT locus of the clinical strain CL1 with the corresponding region of the K12 chromosome, Hsu et al. (1990) showed that the msDNA operon in strain CL1 is present in a 34 kb DNA fragment inserted into a K12-like E. coli chromosome (Hsu et al., 1990). They proposed that this entire fragment is a retroelement, and that the 34 kb fragment present in strain CL1 is integrated into the E. coli chromosome via a reverse transcription process. However, this is unlikely since the 34 kb element is present in two msDNA-negative strains (Fig. 1, strains 49 and 70).

In view of the data presented in this paper and elsewhere (Hsu et al., 1990), it is likely that two genetic elements are involved in the formation of msDNA operon loci of strains B, 317, and CL1: one, a retron, is defined as a unique region in the nucleotide sequence in strains B and 317, and the other, the 34 kb element, is defined as a unit of insertion into a K12-like E. coli chromosome. Although we do not know how much DNA is deleted in the 'cured' strains, preliminary analysis showed that the deletion is longer than the unique region. It will be interesting to see if the 'curing' deletes the entire 34 kb insertion and regenerates a chromosomal structure similar to that of the E. coli K12 chromosome.

**Experimental procedures**

**Strains**

The E. coli strains used in this study are listed in Table 1, together with their serotypes and details of the presence or absence of msDNA; they are representative strains of our collection used to test for the production of msDNA. They were isolated in Brazil from intestinal infections and the presence or absence of msDNA among these strains has been reported (Lim et al., 1990). E. coli B strain AC2514 was described previously (Lim and Maas, 1989a).
Southern hybridizations

About 10 μg of DNA isolated from different strains was electrophoresed on a 0.7% agarose gel after digestion with EcoRI. DNA was transferred onto a GeneScreen plus membrane (New England Nuclear). Hybridization was performed in 6x SSC, 50% formamide at 37°C and the final wash was with 0.1x SSC, 0.5% sodium dodecyl sulphate at 42°C. The probes were made by nick translation of a fragment isolated by agarose gel electrophoresis.

Cloning and sequencing

The genomic library of *E. coli* B has been described by Lim and Maas (Lim and Maas, 1990). The genomic library of strain 317 was constructed in Lambda Gem-11 (Promega) as described by Lim and Maas (1990). Two recombinant lambda clones (Lambda 3171 and Lambda 3172) were selected by plaque hybridization using the 3.5 kb *PstI* fragment of pDB808 (probe II of Fig. 1) and their inserts were analysed by restriction mapping and Southern hybridization. The *XhoI*–*EcoRI* fragment (see Fig. 1A) was subcloned in *SalI* and *EcoRI* sites of plasmid pTZ19R (United States Biochemicals). DNA sequencing was performed by the chain termination method with plasmid DNAs and Sequenase (United States Biochemicals) after generation of nested deletions starting from the *HindIII* site using ExoIII nuclease and ExoVII nuclease (Yanisch-Perron et al., 1985).

'Curing' of the msDNA element

Plasmid pDB110 was constructed by insertion of the 950bp *HindIII*–*HpaI* fragment (Fig. 1) into a *HindIII*–*HincII* sites of pTZ18R (United States Biochemicals). Plasmid pDB1109 was constructed by insertion of the 870bp *HindIII*–*SspI* fragment (see Fig. 1) into the same vector. Both plasmids were transformed into *E. coli* AC2514. Individual transformants were inoculated into 5 ml of LB containing 100 μg ml⁻¹ of ampicillin and 1 mM isopropyl–β-D-thiogalactoside. Incubation was continued for 16 h at 37°C. A single colony from each 5 ml culture was isolated by streaking on LB ampicillin (50 μg ml⁻¹) plates. The isolated colonies were tested for msDNA production by the RT-extension method (Lim and Maas, 1989a).

References


Lampson, B.C., Viswanathan, M., Inouye, M., and 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.


