
Complete sequence of an IS element present in pSC101

Alberto Bernardi and Françoise Bernardi

Laboratoire d'Enzymologie, C.N.R.S., 91190 Gif-sur-Yvette, France

Received 15 April 1981

ABSTRACT

Recently a new insertion element (IS102) has been described in plasmid pSC101. We have determined its complete sequence : it consists of 1057 bp ; 338 bp at one end are identical to those already determined for the kanamycin resistance transposon Tn903. It is not flanked by any direct repeat. Its coding capabilities are discussed, and compared to those of IS903.

INTRODUCTION

The absence of inverted repeats as detected by electron microscopy analysis had suggested that plasmid pSC101 (1, 2) was devoid of IS elements or transposons. However, recently two IS elements were revealed by the biological activities normally associated with IS elements : replicon fusion (3, 4) and deletion formation (5, 6). We have determined the complete sequence of one of these elements (IS102).

Replicon fusion between pSC101 and the phage f1 has allowed the group of Zinder to demonstrate the presence of an IS element of 209 bp, named IS101 (7). The instability of this bireplicon leads to a full transposition yielding a phage f1 with a complete IS101 element flanked by a 5 bp direct repeat and pSC101 still containing the element. This transposition has been shown to be under the control of the $\gamma\delta$ element of the sex factor F (8).

More recently a second IS element was discovered in pSC101, independently by Ohtsubo's group and ourselves. We have shown that a hybrid plasmid formed by pSC101 and a fragment of phage λ carrying the immunity region (pSC101::f2), undergoes large deletions starting at the same nucleotide in pSC101 (6), reminiscent of the now well established features of the IS mediated deletions (9). In order to establish the presence of this element we determined the nucleotide sequence of pSC101 upstream of the common endpoint of the deletions : preliminary results on this pSC101 sequence (6) showed

identity over 250 bp with the ends of the kanamycin resistant transposon Tn903 (10, 11) already sequenced.

Ohtsubo *et al.* reached the same conclusion on the presence of an IS element, closely related to Tn903, from their results on replicon fusion between a thermosensitive replication mutant of pSC101 and ColEI (4) ; this process involved an element of approximately 1000 bp present in pSC101. Their sequence analysis of 100 bp at each end of the element agrees with the data on Tn903 and shows a 9 bp repeat of the target site in ColEI on both sides of the element. This element is identical to the one we have described (6) and we shall call it IS102 according to Ohtsubo's proposal (4).

Tn903 is a constituent of plasmid R6-5 and has been shown to be composed of two inverted repeats of a 1000 bp sequence which flanks a kanamycin resistance gene (12). Recently Grindley and Joyce (13) have determined the sequence of one of the inverted repeats and have shown that these inverted repeats are IS elements which were named IS903. It was therefore interesting to know if there was a complete identity between IS102 and IS903.

MATERIALS AND METHODS

They were already described in detail in Ref. 5. The Maxam and Gilbert (14) technique was used for sequence work.

DNA ligase and polynucleotide kinase from T₄ infected cells were a gift of Dr. F. Rougeon (Institut Pasteur, Paris).

RESULTS AND DISCUSSION

The ends of IS102 are defined on pSC101 in the following way : one end is clearly identified as the common endpoint of all IS102-mediated deletions (6) and the other end is its perfect, 18 bp long inverted repeat (13).

The complete sequence of IS102 was determined partly on pSC101 directly and partly on the largest HindII fragment of pSC101 subcloned in an amplifiable plasmid ; this fragment carries also the replication origin of pSC101 (Fig. 1). For this purpose we have ligated pSC101 degraded to completion with HindII to the vector pBR322 linearized by partial HindII degradation : the ligated mixture was used to transfect the strain poIA ts214 (kindly provided by Dr. A. Goze, IRBM, Paris) and the recombinants were examined on ampicillin (Amp) or tetracycline (Tet) containing plates at 42°C : at this temperature the composite plasmid replicates under pSC101 control in this strain (15) and therefore only recombinants containing the largest pSC101 fragment are viable. The plasmid prepared from one of these Amp^R recombinants was used to trans-

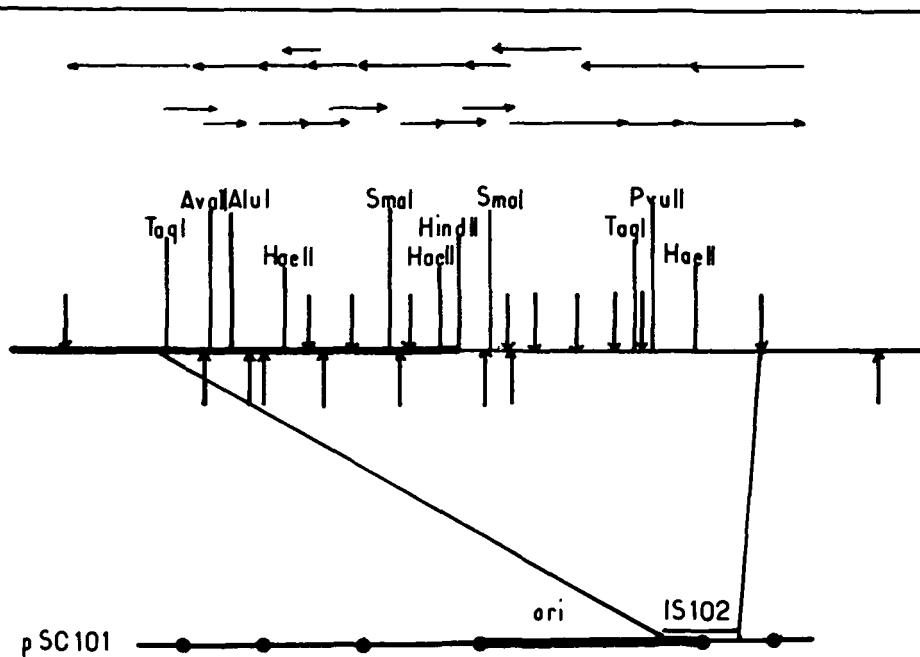
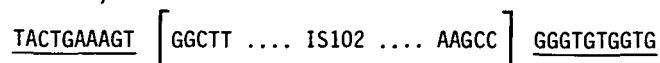


Figure 1. On the top line arrows indicate the extent and direction of the sequence determination. The middle line is an enlarged restriction map of IS102. Downward arrows indicate HpaII sites, upward arrows indicate HinfI sites. On the lowest line the position of IS102 on pSC101 linearized at its EcoRI site is shown. The heavy line indicates the largest HindII fragment from pSC101 which has been subcloned in pBR322. Dots represent HindII sites.

fect the recipient strain HB101. One clone so obtained (AB159) was used for sequencing purposes.

The restriction map of IS102 and its position on pSC101 are shown in Fig. 1. The complete sequence of the element is given in Fig. 2 : its total length is 1057 bp.

We have also sequenced the junctions of the IS102 element in pSC101 and found that the direct repeats normally present on each side of an element after transposition (9) are absent in this case. During replicon fusion IS102 has been shown to generate a 9 bp repeat (4), whereas the flanking sequences in pSC101 (underlined) are :



It appears therefore that for both elements present on pSC101, IS101 (7) and IS102, the direct repeats are missing. The most likely explanation

1 GGCTTTGTTG AATAAATCGA ACTTTTGCTG AGTTGAAGGA TCAGATCAGC
 51 CATCCTCCCG ACAACACAGA CCATTCCGTG GCAAAGCAAA AGTTCAGAAT
 101 CACCAACTGG TCCACCTACA ACAAAGCTCT CATCAACCGT GGCTCCCTCA
 151 CTTTCTGGCT GGATGATGAG GCGATTCAGG CCTGGTATGA GTCGGCAACA
 201 CCTTCATCAC GAGGAAGGCC CCAGCGCTAT TCTGATCTCG CCATCACCAC
 251 CGTTCTGGTG ATTAACGCG TATTCCGGCT GACCCTGCGG GCTGCGCAGG
 301 GTTTTATTGA TTCCATTTTT GCCCTGATGA ACGTTCCGTT GCGCTGCCCG
 351 GATTACACCA GTGTCAGTAA GCGGGCAAAG TCGGTTAATG TCAGTTTCAA
 401 AACGTCCACC CGGGGTGAAA TCGCACACCT GGTGATTGAT TCCACCGGGC
 451 TGAAGGTCTT TGGTGAAGGC GAATGGAAAG TCAGAAAGCA CGGCAAAGAG
 501 CGCCGTGCGTA TCTGGCGAAA GTTGCATCTT GCTGTTGACA GCAACACACA
 551 TGAAGTTGTC TGTGCAGACC TGTCGCTGAA TAACGTCACG GACTCAGAAG
 601 CCTTCCC GGG CTTATCCGG CAGACTCACA GAAAAATCAG GGCAGCCGCG
 651 GCAGACGGGG CTTACGATAC CCGGCTCTGT CACGATGAAC TGCGCCGCAA
 701 AAAAAATCAGC GCGCTTATTC CTCCCCGAAA AGGTGCGGGT TACTGGCCCCG
 751 GTGAATATGC AGACCGTAAC CGTGCAGTGG CTAATCAGCG AATGACCGGG
 801 AGTAATGCGC GGTGGAAATG GACAACAGAT TACAACCGTC GCTCGATAGC
 851 GGAAACGGCG ATGTACCGGG TAAACAGCT GTTCGGGGGT TCACTGACGC
 901 TCGTGACTA CGATGGTCAG GTTGCAGGAG CTATGGCCCT GGTACGAGCG
 951 CTGAACAAAA TGACGAAAGC AGGTATGCCT GAAAGCGTGC GTATTGCCTG
 1001 AAAACACAAC CCGCTACGGG GGAGACTTAC CCGAAATCTG ATTTATTCAA
 1051 CAAAGCC

(7) is that rearrangements have already occurred which have removed the direct repeats of both pSC101 elements.

We have made a detailed analysis of the sequences of the two elements IS903 (13) and IS102. The two ends of the elements are conserved : there is identity for 54 nucleotides at one end and 338 nucleotides at the other end ; in the internal sequence, 665 nucleotides long, 60 single base changes are found. These nucleotide discrepancies account for the loss or appearance of restriction sites as is the case for *PvuII* which has two sites in IS903 and one in IS102. In other instances (*BstN*, *HaeII*, *HinfI*) the number of sites is the same but the location is different. Grindley and Joyce (13) pointed out for IS903 two short internal sequences identical to the ends of the element ; these sequences are conserved in IS102 (see Fig. 2). However the most interesting features arise from the coding capacities of the two elements. Two open reading frames potentially capable of coding for two polypeptide chains are present in IS903 : the largest one (307 codons) initiates at the position 78 with GTG and ends at position 999, the smallest one (114 codons) starts at position 551 with ATG and ends at position 209.

Using a program constructed by J.P. Dumas (IRBM, Paris) we have analysed the coding possibilities of the IS102 sequence : numerous stop codons are found on both strands and in all reading frames. Information for two polypeptide chains, 224 and 114 codons long respectively, is present (Fig. 2) ; there are however no well defined RNA polymerase binding sites (15) or 16S RNA matching sequences (16).

The smallest polypeptide chain (114 codons) has the same length and location in IS903 and IS102, however 24 codons are different. The largest one (224 codons) initiates in IS102 at position 327 with ATG and ends at position 999 which is the same location as in IS903. However Grindley and Joyce favoured a larger reading frame starting with the less frequently used

Figure 2. Complete IS102 nucleotide sequence. 60 nucleotides indicated below the main line show the differences with the IS903 sequence. 11 nucleotides which provoke a mutation in the largest polypeptide chain (224 codons) are underlined. The start and stop codons for the polypeptide chain are indicated ; the largest reading frame, number I, starts at position 327 and terminates at position 999 ; the smallest reading frame, number II, starts at position 551 and terminates at position 209. The two 18 bp inverted repeats are underlined by arrows ; the arrow at position 119-127 corresponds either as inverse or direct repeat to the sequence of the 18 bp inverted repeat (2-10 and 1047-1056) ; the arrow at position 19-30 indicates a sequence that is inversely repeated at position 85-95 as indicated. These features were already reported by Grindley and Joyce (13). Asterisks indicate methylation in C residue.

codon GTG at position 78 on the basis of the abolished transposition caused by an insertional mutation downstream of the GTG codon. Anyway this larger reading frame exists also in IS102. However the abolition of IS903 transposition by the insertional mutation (13) could be due to a modification of an internal resolution site (IRS) as already described for Tn3 (17) rather than to a change in the reading frame as postulated (13) and therefore the 224 codons long polypeptide could be the functional transposase.

As far as the longest polypeptide chain is concerned it should be noted that discrepancies at the nucleotide level affect only slightly this peptide : there are only 11 changes over 224 codons, all located in the first 108 codons (Fig. 2). The others 28 single base changes are silent. One could speculate that the conserved part of the transposase could be involved in the recognition of the ends of the element.

The biological activities of IS102 so far well documented are deletion formation (6) and replicon fusion (4) while no full transposition has been observed. IS102 mediated deletions have been shown to extend in one direction (6).

Meacock and Cohen (18) have described pSC101 replication deficient mutants arising from large deletions ; seven out of eight mutants so formed have a common endpoint which according to their restriction map could be IS102 mediated deletions but extending in the opposite direction as compared to the ones we have described. The deletion that does not seem to be associated with IS102 could well be provoked by the IS101 (see Introduction) as judged by the close similarity between one endpoint of the deletion and the localization of the IS101 element on pSC101 (7).

More genetical data will be necessary to decide whether IS102 is a transposition deficient derivative of IS903 which requires other specific bacterial functions.

ACKNOWLEDGMENTS

We wish to thank Jeremy Nathans for helpful discussions during this work, and Mrs. Jocelyne Mauger for typing the manuscript.

This research was supported by C.N.R.S. grant ATP 3532 to A.B.

REFERENCES

1. Cohen, S.N. and Chang, A.C.Y. (1973) Proc. Natl. Acad. Sci. USA 70, 1293-1297.
2. Cohen, S.N. and Chang, A.C.Y. (1977) J. Bact. 132, 734-737.

3. Ravetch, V.J., Ohsumi, M., Model, P., Vovis, G.F., Fischhoff, D. and Zinder, N.D. (1979) *Proc. Natl. Acad. Sci. USA* 76, 2195-2198.
4. Ohtsubo, H., Zenilman, M. and Ohtsubo, E. (1980) *J. Bact.* 144, 131-141.
5. Bernardi, F. and Bernardi, A. (1980) *Gene* 9, 13-25.
6. Bernardi, A. and Bernardi, F. (1981) *Gene* 13, 103-109.
7. Fischhoff, D.A., Vovis, G.F. and Zinder, N.D. (1980) *J. Mol. Biol.* 144, 247-265.
8. Miller, A.C. and Cohen, S.N. (1980) *Nature* 285, 577-579.
9. Calos, M.P. and Miller, J.H. (1980) *Cell* 20, 579-595.
10. Nomura, N., Yamagishi, H. and Oka, A. (1978) *Gene* 3, 39-51.
11. Oka, A., Nomura, N., Sugimoto, K., Sugisaki, H. and Takanami, M. (1978) *Nature* 276, 845-847.
12. Timmis, K.N., Cabello, F. and Cohen, S.N. (1978) *Molec.gen. Genet.* 162, 121-137.
13. Grindley, N.D.F. and Joyce, M.C. (1980) *Proc. Natl. Acad. Sci. USA* 77, 7176-7180.
14. Maxam, A.M. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldave, K. Eds., Vol 65 part 1, pp. 499-560.
15. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* 72, 784-788.
16. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342-1346.
17. Heffron, F., McCarthy, J.B., Ohtsubo, H. and Ohtsubo, E. (1979) *Cell* 18, 1153-1163.
18. Meacock, P.A. and Cohen, S.N. (1979) *Molec. gen. Genet.* 174, 135-147.