Structure of the archaebacterial transposable element ISH50

Wen-Lian Xu* and W.Ford Doolittle+

Department of Biochemistry, Dalhousie University, Halifax, NS, B3H 4H7, Canada

Received 5 April 1983; Revised and Accepted 24 May 1983

ABSTRACT

We have sequenced in its entirety a new transposable element from the archaebacterium Halobacterium halobium. This 996 bp element shows some features not unlike those of insertion sequences from eubacteria, although it clearly differs from them in sequences which may be involved in transcription and or translation initiation.

INTRODUCTION

Arguments for an early divergence of the archaebacteria from the rest of the prokaryotes (the eubacteria) are now many and compelling. Evidence is accumulating that archaebacterial genome and gene structures may be, in many fundamental ways, unique (1,2). At least one archaebacterial genus, Halobacterium, shows an astonishing degree of genetic instability (3,4). In H. halobium, instability is associated with genetic rearrangements of the type promoted by transposable elements in eubacteria and eukaryotes (5). We have shown that the genome of H. halobium harbors a very large number (≥50) of families of repetitive sequences which effect or are affected by genetic rearrangements occurring at very high frequency (6,7). In order to determine whether such elements exhibit typical "transposon-like" structure, we have sequenced two fragments of the plasmid of H. halobium strain R1 which differ only by a single 996 base pair (bp) element. This element (ISH50) bears 29 bp imperfect inverted terminal repeats and an 819 bp open reading frame on one strand, and a 366 bp potential open reading frame on the other. It is flanked by 8 bp direct repeats of a (target) sequence found only once in the element-free plasmid fragment.

RESULTS AND DISCUSSION

HindIII digestion of the single detectable (50 kbp) plasmid of H. halobium strain R1 yields three fragments of approximately 39, 8 and 7 kbp, res-

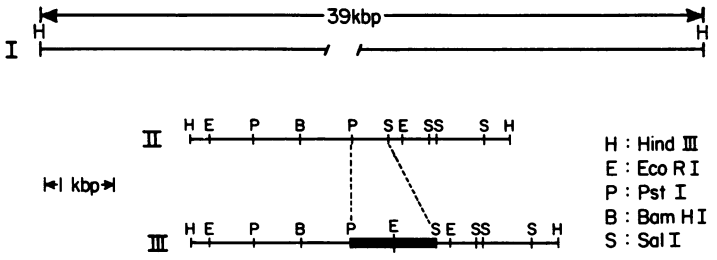

Fig. 1. Hind III fragments of the plasmid of H. halobium strain R1. Covalently closed circular DNA was isolated by ethidium bromide-cesium chloride centrifugation from late logarithmic phase cultures, as described previously (6). The 7 and 8 kbp fragments were purified by agarose gel electrophoresis, and the indicated Pst I - Sal I regions, as well as the Pst I - Eco RI and Eco RI - Sal I regions unique to the 8 kbp fragment, were independently subcloned into pBR322 and physically mapped with the indicated restriction endonucleases. Partially-overlapping Sau 3a, Taq I and Hpa II subclones of these regions in M13 phage derivatives mp8 and mp9 were used for most of the sequencing (8,9).

pectively, and the latter two were cloned into pBR322 (C. Sapienza, unpublished). The cloned fragments showed identical restriction endonuclease digestion maps, except for an approximately 1 kbp insertion (in the 8 kbp fragment) which lies between shared Pst I and Sal I sites (Fig. 1). Southern hybridization experiments revealed that sequences within the insert are present in at least one other position in the genome of strain R1, and in several positions in the genome of strain NRC1, from which R1 was derived (C. Sapienza, unpublished). The Pst I-Sal I region containing the insert in the 8

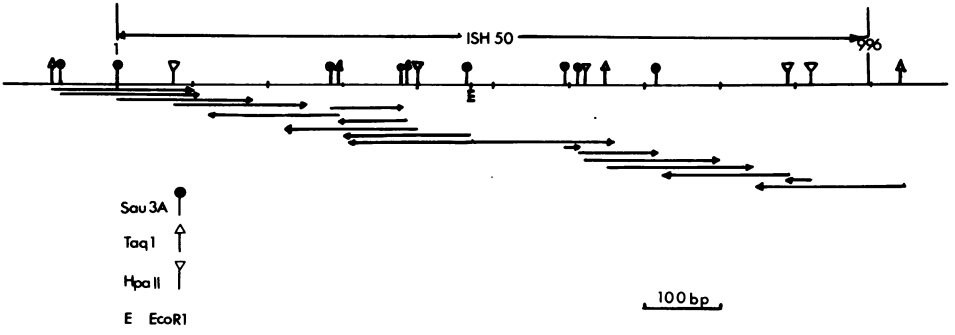

Fig. 2. A schematic representation of subclones used for sequencing. All clones were sequenced at least three times independently.

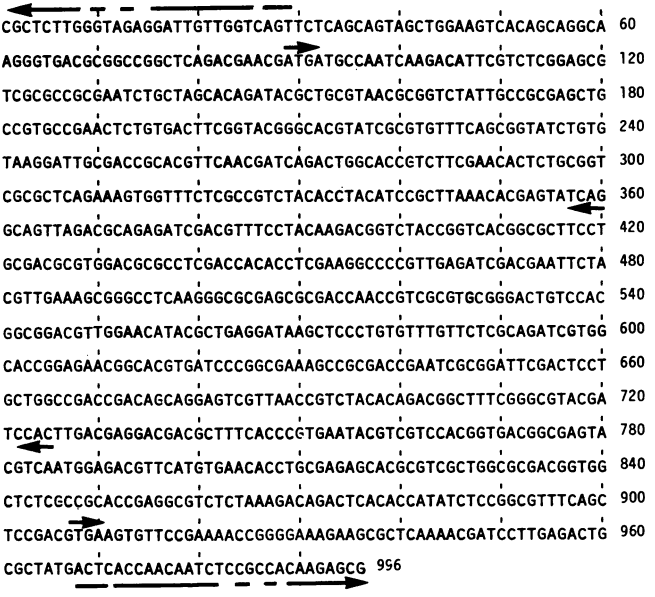

Fig. 3. Complete nucleotide sequence of ISH50. Imperfect terminal repeats are indicated by broken arrows. Initiation/termination signals of the 810 bp open reading frame on the strand whose sequence is presented are indicated by rightward pointing arrows. Initiation/termination signals of the potential 366 bp open reading frame on the complementary strand are indicated by leftward pointing arrows.

kbp fragment and the comparable Pst I-Sal I region of the 7 kbp fragment were sequenced, in part by the method of Maxam and Gilbert (8), and in part by the Sanger dideoxy method, using a total of 18 Sau 3A, Taq I and Hpa II subclones into phages M13 mp8 and mp9[9]. For much of the insert, cloned sequences of both strands were obtained, and were completely complementary. All regions have been sequenced independently at least three times, with no ambiguities in reading and no uncertainty due to "band compression" (Fig. 2).

The 996 bp sequence of ISH50 is shown in Fig. 3. The 29 bp imperfect inverted repeats defining the ends of the element are indicated by arrows. Immediately 5' and immediately 3' to these repeats are direct repeats of the octamer TTGTGGAT, which is present only once in the corresponding region of the 7 kbp fragment lacking ISH50 (Fig. 4). Sequences of the 7 kbp and 8 kbp fragments are otherwise identical for at least 300 bp 5' and 250 bp 3' to the site of ISH50 (Fig. 4 and data not shown). The presence from position 13-24 of ISH50 of a ten of fifteen sequence match with target site nucleotides 5' to

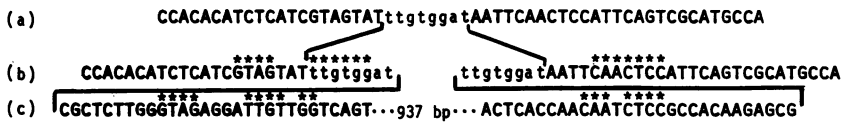

Fig. 4. Structure of the target sequence with and without ISH50. (a) A portion of the sequence of the Pst I - Sal I fragment of the 7 kbp Hind III plasmid fragment (see Fig. 1). (b) The comparable region of the 8kbp fragment. (c) The terminal sequences of ISH50, arrows indicating its position within the linear sequence of the 8 kbp fragment. Lower case letter are used for the target sequence which is duplicated at either end of ISH50. Asterisks indicate regions of potentially significant sequence homology between element termini and adjoining target sequences.

the element, and from positions 977 to 984 of a seven of eight sequence match with nucleotides 3' to the element appears more than coincidental, and may indicate some site specificity for insertion (see below). The ATG beginning at position 89 and the TGA beginning at position 907 define an 819 bp open reading frame sufficient to code for a polypeptide of 35,900 daltons: (273 amino acids). In the complementary strand (not shown) there is an open reading frame sufficient to code for a protein of 122 amino acids, defined by a potential GTG initiation codon beginning at position 725 and a TGA termination codon beginning at position 359. There is no obvious "Pribnow" or "TATA-box"-like sequence 5' to the ATG in the large open reading frame, although there is, surprisingly, a sequence GGTCAGTTCT bearing reasonable homology to the consensus eukaryotic "CAT-box" (GGPyCAATCT) some 66 bp 3' to it. However, no such sequence appears in the comparable region 5' to the smaller potential open reading frame in the complementary strand, nor are there any other regions of striking homology (other than the terminal inverted repeats) in regions 5' to these open reading frames.

ISH50 clearly bears overall structural similarity to many eubacterial insertion sequences. It also bears structural similarity to the only other sequenced archaebacterial transposable element, ISH1 (10). This 1118 bp element also generates 8 bp flanking duplications of target site sequences, is terminated by imperfect inverted repeats (although of only 8 bp), exhibits at least one open reading frame of 810 bp on one strand, a potential open reading frame of 402 bp on the complementary strand, and shows, at its termini, some homology to sequences near the site of insertion, suggestive of site-specific transposition. There is, however, no extensive primary sequence homology between these two elements, even in regions 5' to their open reading frames,

where signals for initiation of transcription and translation might be expected. There are only scattered regions of significant amino acid sequence homology between the proteins which these open reading frames could encode. A third element, ISV1, present in both H. halobium and H. volcanii also shows little obvious sequence homology to either ISH1 or ISH50 (J. Hofman, unpublished). It seems likely that among the 50 or more transposable elements borne by H. halobium, there may be many which bear only remote evolutionary resemblance to each other.

It is of some interest that archaebacteria, whose divergence from eubacteria is very ancient indeed (1,2), harbor tranposable elements whose overall structural features are so similar to those of most eubacterial and some eukaryotic transposable elements. It is tempting to suggest that these structural similarities reflect constraints on the ways in which transposable elements can be constructed and maintained, rather than common descent.

This work was supported by a grant from the Medical Research Council of Canada. We are grateful to Carmen Sapienza for performing some of the experiments described here. W.-L. Xu was supported by a graduate student fellowship from the Dalhousie Faculty of Graduate Studies.

*Permanent address: The Basic Medical Institute, Beijing, China

+To whom correspondence should be addressed

REFERENCES

1. Woese, C.R. (1981) Scient. Am. 244 (6), 94-106.
2. Kandler, O. (1982) (ed.) Archaebacteria, Gustav Fischer Verlag, Stuttgart.
3. Weidinger, G., Klotz, G. and Goebel, W. (1979) Plasmid 2, 377-386.
4. Pfeifer, F., Weidinger, G. and Goebel, W. (1981) J. Bact. 145, 375-381.
5. Movable Genetic Elements, (1981) Cold Spring Harbor Symp. Quant. Biol. 45.
6. Sapienza, C. and Doolittle, W.F. (1982) Nature 295, 384-389.
7. Sapienza, C. and Doolittle, W.F. (1982) Nature 299, 182-185.
8. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
9. Messing, J., Crea, R. and Seeburg, P.H. (1981)Nucleic Acids Res. 9, 309-321.
10. Simsek, M., DasSarma, S., RajBhandary, U.L. and Khorana, H.G. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 7268-7272.