Structural analysis of a new GC-specific insertion element IS186

Pele Chong*, Ivy Hui, Tip Loo and Shirley Gillam+

Department of Pathology, Faculty of Medicine, 2211 Wesbrook Mall, University of British Columbia, Vancouver, BC V6T 1W5, Canada

Received 23 July 1985

A new insertion sequence, IS186, has been identified from *Escherichia coli* and sequenced. It contains 1336 base pairs with a terminal inverted repeat of 22 nucleotides. A long open reading frame, from an ATG codon at position 55 to a TAG termination codon at 1327, could code for a polypeptide of 424 amino acids. This element recognizes GC-rich regions as target sites for insertion.

Insertion element *Illegitimate recombination* *Inverted repeat sequence* *Recombinational hot spot* *Base-pair duplication* Escherichia coli *plasmid*

1. INTRODUCTION

Bacterial insertion sequences (IS) are small transposable elements that contain no phenotypic markers. Presumably, they carry only the genetic information necessary for transposition [1]. Most bacterial IS that have been analyzed at the nucleotide sequence level range in size from 750 to 1600 bp and contain at least one long open reading frame [1]. These elements have certain unusual features in common [2]: (i) they contain a short terminal inverted repeat of 20–40 bp; (ii) they generate short direct repeats of various sizes (2–20 bp) upon insertion at the host target sequence; (iii) they can enhance or repress the expression of a gene in a bacterial operon. This type of repression is known as a polarity effect.

IS elements can mediate gene rearrangement, deletion and inversion. Precise excision of these elements can restore normal function of the gene [3]. Khorana and his co-workers [4] have demonstrated a high frequency of spontaneous mutations in the bacteriorhodopsin gene in *Halobacterium halobium* generated by IS elements (ISH1 and ISH2). Devos et al. [5] cloned a cDNA copy of RNA phage MS2 in pBR322 by the dA-dT linker method and observed a hot spot for the integration of IS1 near one of these linkers.

We have identified a new insertion sequence, IS186, during the course of mapping and sequencing cDNA clones from rubella virus. These were constructed by a conventional dG/dC tailing procedure at the *Pst*I site of plasmid pBR322 [6]. Of the 6 clones sequenced, 3 were found to contain the new insertion sequence (IS186). IS186 consists of 1336 base pairs with a perfectly matched terminal inverted repeat of 22 nucleotides. It has one long open reading frame (ORF) which could code for a polypeptide of 424 amino acid residues. This insertion element recognizes GC-rich regions and inserts into or near the GC-tail of cDNA clones. Here we describe the structural analysis of IS186.

+ To whom correspondence should be addressed
* Present address: Connaught Research Institute, 1755 Streeles Avenue West, Willowdale, Ontario M2R 3T4, Canada

2. MATERIALS AND METHODS

2.1. *Enzymes and chemicals*

Restriction endonucleases, T_4 DNA ligase,

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

5′-polynucleotide kinase, DNase I, RNase H, DNA polymerase I and the large fragment of DNA polymerase I (Klenow) were purchased from BRL. AMV reverse transcriptase was obtained from Life Science and the terminal deoxynucleotidyl transferase was from Boehringer Mannheim.

Chloramphenicol, ampicillin and tetracycline were purchased from Sigma. dNTPs and ddNTPs were obtained from Pharmacia P-L Biochemicals. [γ-^{32}P]ATP and [α-^{32}P]dATP were obtained from Amersham. Oligonucleotide primers were synthesized by solid-phase synthesis using the phosphite triester method [7] with an Applied Biosystems 380A DNA synthesizer.

E. coli strains RR1 and JM101 were used for transformation by pBR322 and M13 vectors, respectively.

2.2. *Construction of rubella virus cDNA library*

Rubella virus (M33 strain, ATCC VR-315) was cultured in Vero cells as reported [8]. Single-stranded cDNA from viral RNA was prepared by using d(T_8CT) as a primer with reverse transcriptase [9], and the second-strand cDNA was synthesized by using RNase H, *E. coli* DNA ligase and DNA polymerase I as described by Okayama and Berg [10]. The dC-tailed double-stranded cDNA was inserted into the *Pst*I site of plasmid pBR322 [6]. Rubella-specific cDNA clones were selected by using the 5′-^{32}P-labelled synthetic oligodeoxyribonucleotide ([^{32}P]dG$_3$A$_2$TCTAGTG) as a probe [9].

2.3. *Mapping and sequencing of IS186*

Fragments from the digests by restriction enzymes (*Bam*HI, *Hae*III, *Pst*I, *Sau*3A1 and *Taq*I) were used for mapping and subcloning into M13 vectors for DNA sequencing. Oligodeoxyribonucleotides (P_3–P_8) were used as primers for DNA sequencing in chain-terminator procedures [11].

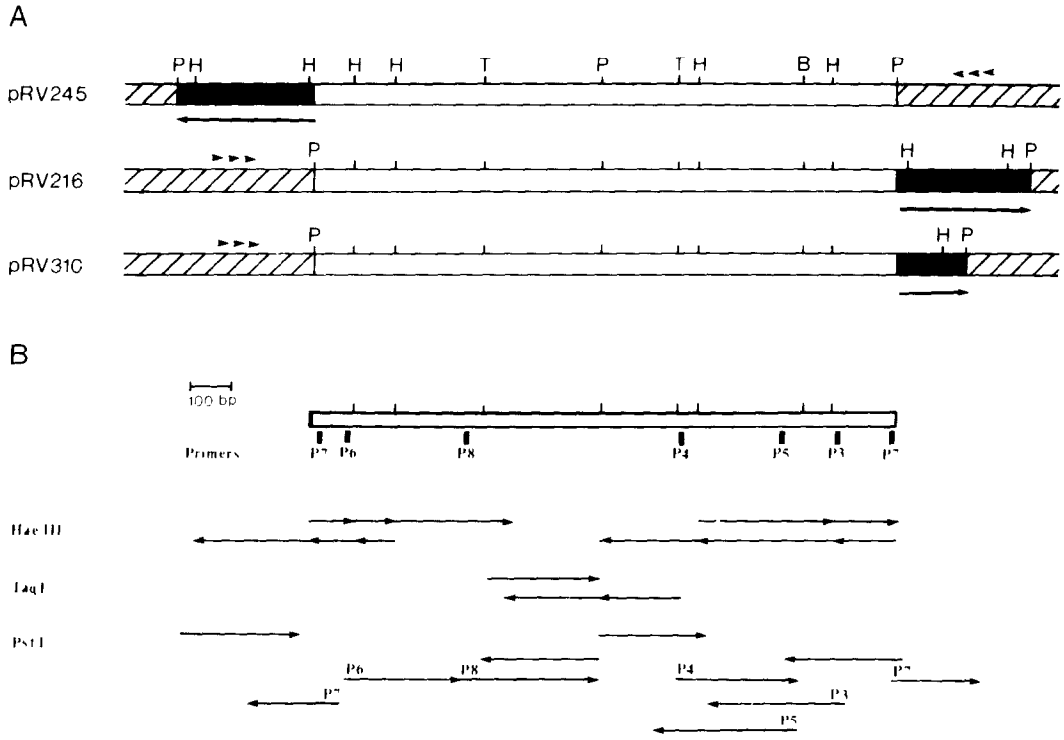

Fig.1. Schematic representation of IS186. (A) Restriction map of IS186. Restriction sites; P (*Pst*I), H (*Hae*III), T (*Taq*I), B (*Bam*HI). The diagonally lined area denotes the ampicillin resistance promoter region and the solid filled area denotes the 3′-end rubella cDNA sequence. The open bar represents the IS186. The arrow indicates the direction of transcription. (B) Strategy used to obtain the DNA sequence of IS186. P3–P8 are the synthetic oligonucleotide primers. (See fig.3 for their sequences.) The horizontal arrows indicate the individual strands of each restriction enzyme fragment that were sequenced.

3. RESULTS

Insertion element IS186 was first detected from the plasmid pRV245. The restriction map of IS186 is shown in fig.1A. Plasmid pRV245 has IS186 and 380 bp of the rubella sequence (flanked by GC-tails) inserted into the *Pst*I site of the pBR322 ampicillin resistance gene. The gene order is ampicillin resistance promoter–IS186–rubella sequences. Plasmid pRV216 contains the same insert as pRV245 with the exception that IS186 is inserted in the opposite orientation. Plasmid pRV310 has IS186 inserted in the same orientation as that of pRV216 but contains only 160 bp of the rubella sequence (not shown). The sequencing strategy for IS186 is shown in fig.1B. Most regions of IS186 have been sequenced on both strands using the dideoxynucleotide chain-termination method [11].

The cDNA library was screened with IS186-specific oligonucleotide primers (P_3–P_5 and P7). More than 100 positive cDNA clones were observed, accounting for 7% of the cDNA library. Only 3 clones (pRV216, pRV245 and pRV310) that contain IS186 were sequenced. The integration sites of IS186 in these 3 clones are shown in fig.2. Insertions were found in the GC-tails next to the *Pst*I cloning site, generating a direct duplication of target sequences of 13, 12 and 10 bp in pRV216, pRV245 and pRV310, respectively.

IS186 is 1336 nucleotides long with a perfectly matched terminal inverted repeat of 22 nucleotides (see fig.2). Upon examining the DNA sequence of IS186, 2 significant ORFs were found. The long ORF-1 (fig.3) extends from an ATG start codon at position 55 to a TAG termination codon at position 1327. ORF-1 could encode a polypeptide of 424 amino acids (M_r 48000). The short ORF-2 is found in the opposite orientation and could encode 102 amino acids with GTG as a starting codon at position 787. Several stop codons are encountered in all 3 reading frames within 100 bp of either end of IS186; these may be part of the structure responsible for the polar effect of IS186. Examination of the sequence of IS186 did not yield any good potential Shine-Dalgarno sequence [12] nor ϱ-independent or ϱ-dependent transcription termination signal [13]. However, a possible promoter sequence can be found with the −35 region located at nucleotide 19 and the Pribnow box at nucleotide 43 [13] (fig.4).

To determine the origin of IS186, several genomic Southern blot hybridizations were carried out using radioactively labelled IS186-specific probes on different *E. coli* strains and on tissue

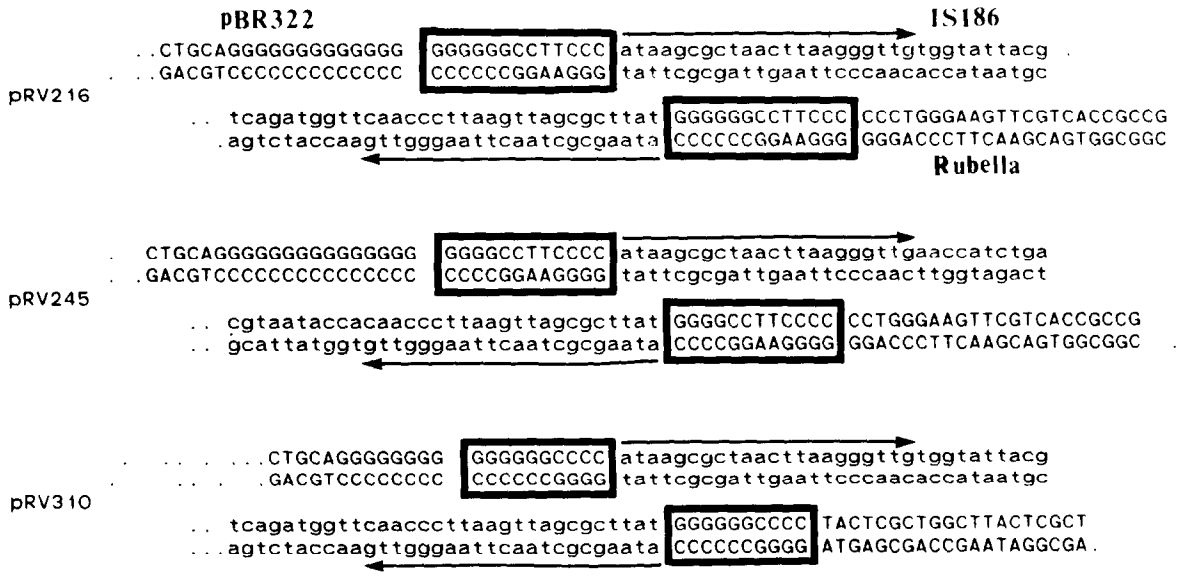

Fig.2. Comparison of the integration sites of IS186. Boxed area indicates the duplicated target site and the arrow indicates the inverted repeat sequence of IS186.

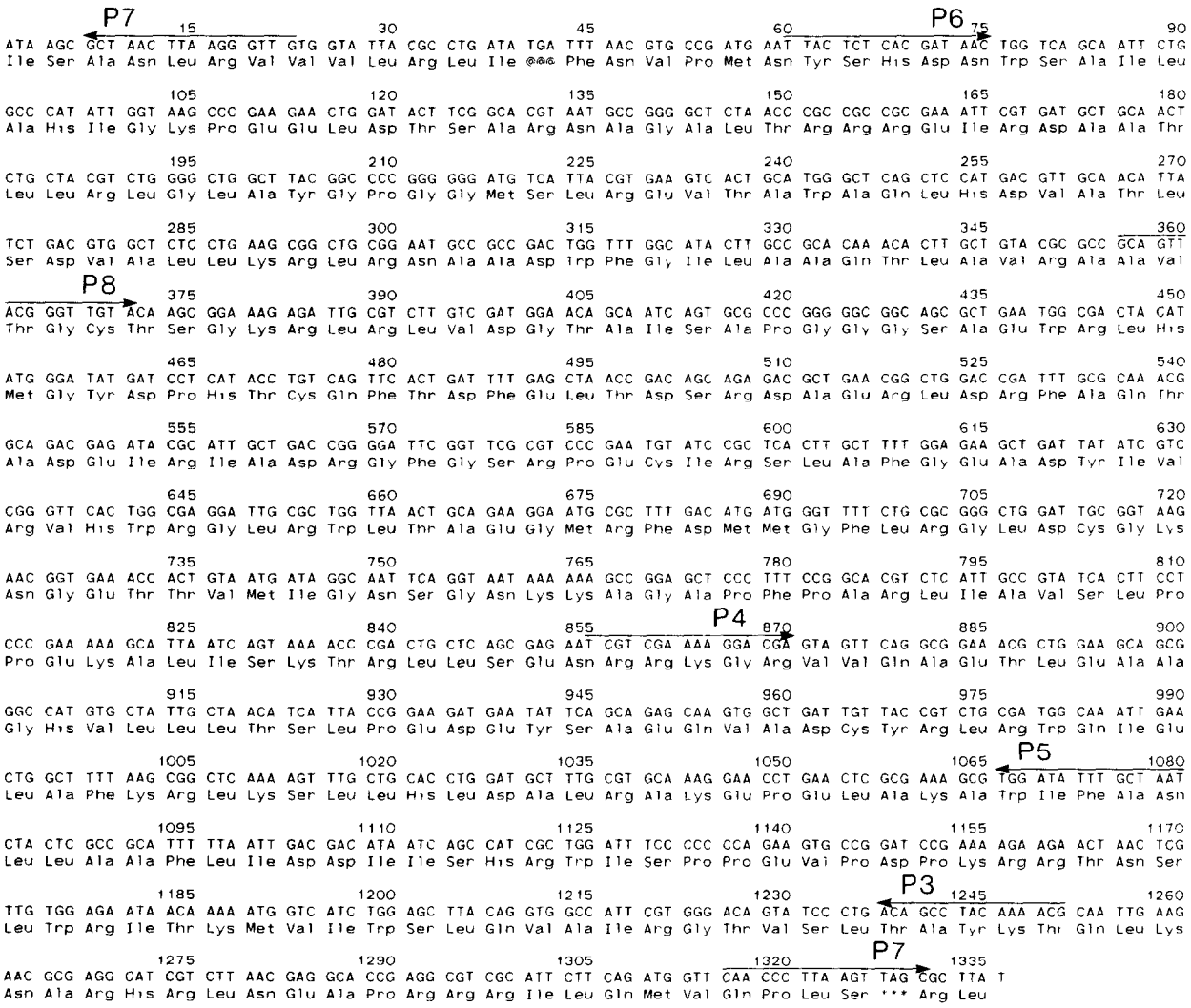

Fig.3. Nucleotide sequence of IS186. ORF 1, position 55–1327; ORF 2, position 787–481 in opposite orientation.

culture cell lines (RK and Vero cells) used for the propagation of rubella virus. IS186 sequences were found in all *E. coli* strains tested but not in the tissue culture cell lines (fig.5). Besides IS186, we also found inserts of IS1 [14] and IS4 [15] in our rubella cDNA clones (not shown).

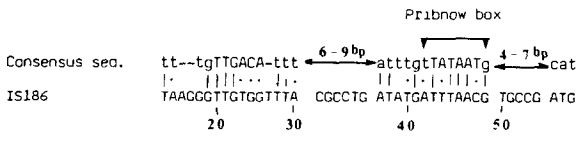

Fig.4. Sequence preceding the large open reading frame of IS186.

4. DISCUSSION

The evidence presented here indicates that the DNA segment inserted in rubella molecular chimeras is a new translocatable element. Candido and his co-workers (Dept of Biochemistry, University of British Columbia, Vancouver BC, Canada) also observed IS186 inserted into their rainbow trout heat shock gene cDNA library, which was also constructed by dG-dC-tailing at the *Pst*I site of pBR322 (personal communications).

A number of special features were observed for IS186:

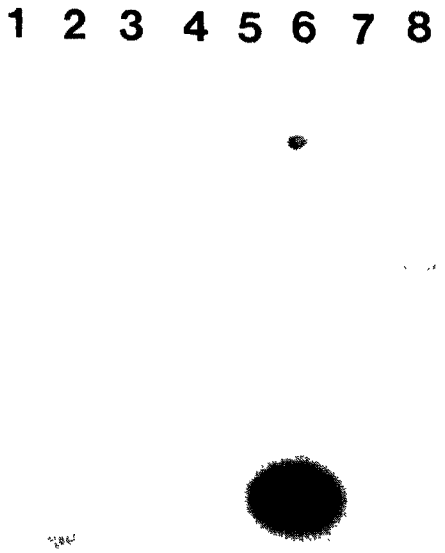

Fig.5. Southern blot hybridization of IS186 sequence to *Pst*I and *Bam*HI digests of DNA from different strains of *E. coli* and animal cells. A nick-translated 0.68 kbp fragment from *Pst*I digest of IS186 was used as probe. Total DNA from RRI, JM83, JM101, Vero, RK, IS186 fragment, JC8111 and RY1088 (lanes 1–8, respectively) was used for digestion, and fractionated by 1% agarose gel electrophoresis.

(i) it shows a high frequency of insertion into the cDNA library of rubella virus (7%);
(ii) it recognizes GC-rich regions as target sites for insertion;
(iii) it can duplicate 10, 12 or 13 nucleotides at the site of insertion.

The high frequency of insertion of IS186 into the rubella cDNA library suggests that there may be hot spots in the rubella cDNA sequence for recombination. It is reasonable to suggest that the 3′-end sequence of the rubella viral genome contains hot spots for recombination because 3′-end sequences of most viruses known to date usually contain signals to regulate the replication of the virus. It is possible that cloning of rubella cDNA into the *Pst*I site of pBR322 may lead to the expression of β-lactamase-rubella fusion proteins which are detrimental to the bacterial host. Thus the presence of IS186 next to the cloned rubella cDNA may constitute a selective advantage.

Similarities in structure and properties were found between IS186 and IS4. Of the IS known, IS186 and IS4 are the only ones that generate a variable length of duplication at the target site. IS4 generates an 11 or 12 bp duplication; whereas IS186 generated a 10, 12 or 13 bp duplication upon insertion. Both IS186 and IS4 contain 2 ORFs (one long and one short). Within the long ORF, there are several basic amino acid residue domains which may play an important role in DNA binding (table 1).

ACKNOWLEDGEMENTS

This work was supported jointly by grants from the Medical Research Council of Canada, British Columbia Health Care Research Foundation, Alberta Heritage Foundation for Medical Research and Health and Welfare Canada. P.C. is an MRC Postdoctoral Fellow.

Table 1

Similarity of the amino acid basic residue regions of IS186 and IS4[a]

IS186	IS4
Arg^{33}-Arg-Arg-Glu-Ile-Arg^{38}	Arg^{45}-Lys-Arg-Arg-Leu-Pro^{50}
Leu^{78}-Lys-Arg-Leu-Arg-Asn^{83}	Arg^{105}-Arg-Val-Phe-Thr-Lys^{110}
Gly^{108}-Lys-Arg-Leu-Arg-Leu^{113}	Arg^{244}-Lys-Leu-Gly-Lys-Gly^{249}
Arg^{268}-Arg-Lys-Gly-Arg-Val^{273}	Arg^{262}-Lys-Lys-Trp-Pro-Gly^{267}
Phe^{315}-Lys-Arg-Leu-Lys-Ser^{320}	Arg^{281}-Lys-Gly-Lys-Val-Lys^{286}
Lys^{367}-Arg-Arg-Thr-Asn-Ser^{392}	Leu^{331}-Arg-Ser-Lys-Lys-Pro^{306}
Lys^{402}-Asn-Ala-Arg-His-Arg^{807}	Arg^{415}-Arg-Glu-Arg-Ala-Phe^{420}
Pro^{412}-Arg-Arg-Arg-Ile-Leu^{417}	Arg^{422}-Val-Val-Lys-Glu-Arg^{427}

[a] Amino acid sequences are predicted from the DNA sequence of IS4 [15]

REFERENCES

[1] Iida, S., Meyer, J. and Arber, W. (1983) in: Mobile Genetic Elements (Shapiro, J.A. ed.) pp.159–221, Academic Press, New York.
[2] Kleckner, N. (1981) Annu. Rev. Genet. 15, 341–404.
[3] Calos, M.P. and Miller, J.H. (1980) Cell 20, 579–595.
[4] Simsek, M., DasSarma, S., RajBhandary, U.L. and Khorana, G.H. (1982) Proc. Natl. Acad. Sci. USA 79, 7268–7272.
[5] Devos, R., Contreras, R., Van Emmelo, J. and Fiers, W. (1979) J. Mol. Biol. 128, 621–632.
[6] Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naker, S.P., Chick, W.L. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 3727–3731.
[7] Atkinson, T. and Smith, M. (1984) in: Oligonucleotide Synthesis: A Practical Approach (Gait, M.J. ed.) pp.83–114, IRL Press.
[8] Chong, P. and Gillam, S. (1985) J. Virol. Methods 10, 261–268.
[9] Gillam, S., Chong, P., Loo, T. and Trudel, M. (1984) Abstract, Canadian Society of Microbiologists, June, Kingston, Ontario, Canada.
[10] Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161–170.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[13] Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319–354.
[14] Ohtsubo, H. and Ohtsubo, E. (1978) Proc. Natl. Acad. Sci. USA 75, 615–619.
[15] Klaer, R., Kuhn, S., Tillmann, E., Frits, H.-J. and Starlinger, P. (1981) Mol. Gen. Genet. 181, 169–175.