The Journal of Microbiology, March 2004, p.56-59 Vol. 42, No. 1

NOTE

Transposition of IntAs into the Conserved Regions of IS*3* Family Elements

Chang-Gyun Han

The Institute of Physical and Chemical Research (Riken), Hirosawa 2-1, Wako-Shi, Saitama, 351-0198, Japan

(Received August 18, 2003 / Accepted December 24, 2003)

Together with the previous reports, my computer survey revealed that several bacteria contain six copies of the type group II intron IntA. The sequence analysis of IntAs showed the high level of homology in the nucleotide sequence (91.9~99.8%). The consensus sequence, 2,270 base pair long, was derived from the nucleotide sequences of all IntA members. The size of the open reading frame *intA* was 502 amino acids long, that is homologous to reverse transcriptase-like proteins encoded within the group II introns. It was reported that EPEC.IntA and Sf.IntA were inserted into IS*911* and IS*629*, respectively. The sequence of the flanking region IntA was analyzed here. The data show the insertion of EC.IntA into IS*629*, the insertion of EHEC.IntA into IS*3*, the insertion of Yp.IntA into IS*904*-like sequence, and the insertion of EK12.IntA into IS*911*. Interestingly, these IS elements nested by IntAs were the members of IS*3* family elements. The sequences of the IS*3* members correspond to the OrfB with the DDE motif conserved in retroviral integrases. Alignment of the flanking sequences of IntAs revealed that the flanking regions -25 to +10 of insertion sites, that are generally believed to be required for the retrohoming, were not strongly conserved. The data presented here suggests that the retrohoming pathway of IntA seems to differ from those of other group II introns.

Key words: group II intron, IntAs, IS*3* family, target-site specificity

Group II introns are mobile genetic elements in fungi, plant, and bacteria (Belfort *et al.*, 1995; Michel and Ferat, 1995). Bacterial group II introns have been found in proteobacteria (*Azotovacter vinelandii*, *Escherichia coli*, *Sinorhizobium meliloti*, *Shigella flexneri*, and *Yersinia pestis*), cyanobacteria (*Calothrix* sp.), and gram-positive bacteria (*Bacillus halodurans* C-125, *Lactococcus lactis*, and *Clostridium difficile*) (Ferat and Michel, 1993; Ferat *et al.*, 1994; Knoop and Brennicke, 1994; Amemura *et al.*, 1995; Mills *et al.*, 1996; Mullany *et al.*, 1996; Shearman *et al.*, 1996; Rajakumar *et al.*, 1997; Hu *et al.*, 1998; Lindler *et al.*, 1998; Martinez-Abarca *et al.*, 1998; Tobe *et al.*, 1999; Takami *et al.*, 2001). Bacterial group II introns are believed to be disseminated by mobile genetic elements (Ferat *et al.*, 1994). IntA and IntB are inserted at two distinct sites in IS*677* (=H-repeat) and IntD is located within IS*629* (=IS*3411*) (Ferat *et al.*, 1994). RmInt1 is inserted within ISRm*2011-2* (Martinez-Abarca *et al.*, 1998). IntA and EPEC.IntA are inserted into IS*629* and IS*911*, respectively (Rajakumar *et al.*, 1997; Tobe *et al.*, 1999). IntC is inserted into IS*679* (Han *et al.*, 2001). The group II intron found in *C. difficile* was found in the conjugative transposon, Tn*5397* (Mullany *et al.*, 1996). Similarly, the lactococcal intron, LI.LtrB, was detected within a transfer gene of conjugative elements (Shearman *et al.*, 1995; Mills *et al.*, 1996). This study focused on a type of group II intron that appears to be present in several bacteria by collecting the members and analyzing their flanking sequences.

The programs, FASTA (Pearson and Lipman, 1988) and BLAST (Altschul *et al.*, 1990), were used for the homology search of the nucleotide sequences in the DDBJ/GenBank/EMBL databases. Multiple sequences were aligned using the program CLUSTAL W version 1.7 (Thompson *et al.*, 1994). The nucleotide sequences were analyzed with the HarrPlot 2.0 and GENETYX-Mac 10.1 system (Software Development Co, Japan).

A family of the group II intron seems to be present in several bacteria (Table 1). Sf.IntA is 2,272 bp in length identified from *S. flexneri*. The *sfiA* gene encodes a putative protein exhibiting the high level of similarity to reverse transcriptase-like proteins encoded within the introns of fungi (GenBank accession no. X55026, U41288, and X57546), plants (GenBank accession no.

* To whom correspondence should be addressed.
(Tel) 81-90-6195-6000; (Fax) 81-48-480-4757
(E-mail) karasu9@brain.riken.go.jp

Table 1. Target IS elements of IntAs

IntAs	Co-ordinates to referred accession no.	Target IS (IS*3* family)	Organism	Referred accession no.
EC.IntA	1~1590	IS*629*	*E. coli* C	D37918
EHEC.IntA	65789~65864	IS*3*	EHEC plasmid pO157	AB011549
EK12.IntA	7985~9748	IS*911*	*E. coli* K-12	AE000133
EPEC.IntA	48555~50824	IS*911*	EPEC EAF plasmid	AB024946
Sf.IntA	C 519~2787	IS*629*	*S. flexneri* SBA1336	U97489
Yp.IntA	54417~55559	IS*904*-like sequece	*Y. pestis* plasmid pMT-1	AF074611

Position begin with C, represents Sf.IntA in minus orientation.
EPEC, Enteropathogenic *Escherichia coli*; EAF, adherent factor; EHEC, Enterohemorrhagic *Escherichia coli*; Y., *Yersinia*; S., *Shigella*.

M68929), and bacteria (GenBank accession no. U7795, U50902, X98606, and X71404) (Rajakumar *et al.*, 1997). EPEC.IntA, 2,270 bp in length, is identified from enteropathogenic *E. coli* (EPEC) (Tobe *et al.*, 1999). *E. coli* K-12 chromosome contains a group II intron truncated at the right end region (Rajakumar *et al.*, 1997; unpublished data). This was termed as group II intron EK12.IntA. EK12.IntA is truncated by IS*5*-mediated deletion (unpublished data). Two group II introns were identified: *S. flexneri* reverse transcriptase (RT) and a RT-like protein. The former was identified from *E. coli* C. The latter was identified from *Y. pestis* (Amemura *et al.*, 1995; Hu *et al.*, 1998; Lindler *et al.*, 1998). These elements were termed as EC.IntA and Yp.IntA, respectively. Using these group II intron-sequences as queries, a new copy was identified from enterohemorrhagic *E. coli* (EHEC) O157:H7 by the computer-aided homology search. Thus, this element was termed EHEC.IntA which was identified from EHEC O157:H7 (accession no. AB011549 and AF 074613).

Although these group II introns were present at several bacteria, the sequences of group II introns had the high level of nucleotide identity (91.9~99.8%) over their entire length to one another. I described these group II introns as IntAs. With the exception of EPEC.IntA and Sf.IntA, other IntAs were truncated at the either side of the end region.

Sf.IntA is 2,272 bp in length and has 431 amino acids (Rajakumar *et al.*, 1997). However, the aligned sequences generated the 2,270-bp consensus sequence from IntAs (data not shown). This consensus sequence contains the open reading frame (ORF) homologous to reverse transcriptase. The size of ORF named *intA* was 502 amino acids. Compared with the 2,270-bp consensus sequence, I found two base insertions in the entire Sf.IntA. One in the ORF of Sf.IntA resulted in the frameshifting that introduces a stop codon. Thus, the *sfiA* (431 amino acids) of Sf.IntA was shorter than *intA* (502 amino acids) of the consensus sequences of IntAs (data not shown). Yp.IntA showed the 92.0% homology to the consensus sequence of IntAs. Comparing with the consensus sequence of IntAs, Yp.IntA contained the deletion of a 16-bp sequence in the ORF and showed 8% nt substitution (data not shown).

The flanking sequences of three members of IntAs have previously been reported. Sf.IntA has been reported to be inserted into the intact IS*629* element (Rajakumar *et al.*, 1997). Either EK12.IntA or EPEC.IntA has been reported to be inserted into IS*911* (Tobe *et al.*, 1999; unpublished data). In addition, I identified and characterized both flanking sequence of other IntAs. Computer survey revealed that EHEC.IntA and EC.IntA were inserted into IS*3* and IS*629*, respectively. In the case of Yp.IntA, the flanking sequence of Yp.IntA was found to be the IS-like sequence. IS*904* (Rauch *et al.*, 1990) was most homologous to the IS-like sequence nested by Yp.IntA. The end region of both sides, however, can not be characterized because the IS-like sequence was severely truncated. Thus, I called this IS-like sequence IS*904*-like sequence.

Unlike other types of group II introns, the result of the analysis of the flanking sequences of IntAs revealed that IntAs were present in several bacterial and inserted into various IS elements such as IS*3*, IS*629*, and IS*911* (Timmerman and Tu, 1985; Matsutani and Ohtsubo, 1990; Prère *et al.*, 1990). Interestingly, these IS elements nested by IntAs belong to the IS*3* family. The IS*3* family is the largest group characterized so far, which includes 28 IS elements such as IS*3*, IS*629*, IS*904*, and IS*911* (Ohtsubo and Sekine, 1996). They are present not only in gram-negative bacteria but also in gram-positive bacteria (Ohtsubo and Sekine, 1996). They greatly differ one another in the nucleotide sequence, but almost all members of this group have two open reading frames corresponding to those in IS*3* (Ohtsubo and Sekine, 1996). Although the predicted amino acid sequences of the ORFAs show little homology, those of ORFBs are significantly related (Schwartz *et al.*, 1988; Fayet *et al.*, 1990; Prére *et al.*, 1990). Comparison of the nucleotide sequences of the IS elements nested by IntAs revealed that IntAs was inserted into the ORFB of IS*3* family. Moreover, the regions of IS elements flanking IntAs were found to correspond to the conserved region of the ORFB of IS*3* family (Fig. 1A), which correspond to the integrase domain shared by retroelements

A

IS*3*	ACCAGAAGTGGGCGGGA**GAC**ATCACGTATT	**EHEC.IntA**	////////////////////
IS*629*	////////////////////////////	**EC.IntA**	TCAGCACATGGCGGGGCTTC
IS*629*	ACCAGCTGTGGGTGGCT**GAT**TTTACTTACG	**Sf.IntA**	TCAGCACATGGCAGGGCTTC
IS*904*-like seq	ACCAGGTGTGGGTGAGC**GAT**ATCACCTATC	**Yp.IntA**	////////////////////
IS*911*	ATCAGGTGTGGTGCGGT**GAT**GTGACCTATA	**EK12.IntA**	////////////////////
IS*911*	ATCAGGTGTGGTGCGGT**GAT**GTGACCTATA	**EPEC.IntA**	TCTGGACGGGTAAGCGCTGG

B

	-30 -20 -10 -1		+1 +10 +20
IS3	ACCAGAAGTGGGCAGGA**GAC**ATCACGTACT	**EHEC.IntA**	TACGTACAGATGAAGGCTGG
IS*629*	ACCAGCTGTGGGTGGCT**GAT**TTTACTTACG	**EC. & Sf.IntA**	TCAGCACATGGCAGGGCTTC
IS*904*-like seq	ACCAGGTGTGGGTGAGC**GAT**ATCACCTATC	**Yp.IntA**	////////////////////
IS*911*	ATCAGGTGTGGTGCGGT**GAT**GTGACCTATA	**EK12. & EPEC.IntA**	TCTGGACGGGTAAGCGCTGG
Consensus seq	A CAG GTGG GA T AC TA		T G AC A GCT

Fig. 1. DNA sequences of IS elements corresponding to the 5' and 3' junctions of IntAs. (A) Target regions of IS elements flanking IntAs, which were identified from databases. (B) Original regions of IS elements flanking IntAs. Both IS elements and IntAs were in plus orientation. The truncated regions were indicated by slashes. The sequences of IS*3* family members flanking IntAs correspond to *OrfB* that has a DDE (Aspartic acid, Aspartic acid, Glutamic acid) motif conserved in retroviral integrases. An acidic amino acid triad is involved in catalysis, and its role is presumably in coordinating divalent metal cations. Nucleotides shown in bold type correspond to the first D of the DDE motif. The first nucleotides from 5' and 3' junctions indicate −1 and +1 positions, respectively. The positions of flanking sequences of IntAs are shown at the top in Arabic numerals. The underlined nucleotide corresponds to the position at the top.

(Fayet *et al.*, 1990; Khan *et al.*, 1991; Kulkosky *et al.*, 1992; Rezsöhazy *et al.*, 1993).

Although the regions of IS*3* family where six IntAs were inserted, were conserved, the sequences of the target regions differed. Alignment of flanking regions -13 to +1 of the target sites of IntAs showed eight positions (-13, -12, -9, -7, -6, -4, -3, and +1) with the common nucleotide sequence (Fig. 1B). This region was considered intron binding site 1 (IBS1) and intron binding site 2 (IBS2) that are complementary to the intron sequence exon binding site 1 (EBS1) and exon binding site 2 (EBS2) located in the region of domain I of the intron RNA (Michel *et al.*, 1989; Ferat and Michel, 1993; Michel and Ferat, 1995). Recently, three group II introns have been investigated extensively to characterize the mechanism of a group II intron retrohoming. In the case of a bacterial group II intron which is identified from a conjugative relaxase gene, *ltrB* (Mills *et al.*, 1996), Ll.ltrB, intron RNA base pairs with positions -13 to +1 of the DNA homing site (Mills *et al.*, 1996; Matsuura *et al.*, 1997). The flanking region -25 to -13 of the insertion site and +2 to +10 of the insertion site were not strongly conserved, indicating seven positions (-23, -22, -21, -20, +4, +6, and +7) with common nucleotide sequences (Fig. 1B). In the case of Ll.ltrB, the flanking regions, -25 to -13 of the insertion site and +2 to +10 of the insertion site, are recognized by the LtrA protein (Cousineau *et al.*, 1998). A single nucleotide change in IBS1 of the target DNA was sufficient to almost completely block the DNA endonuclease and reverse splicing activity (Matsuura *et al.*, 1997). The RNP recognition span from -25 to +10 corresponds to that need for retrohoming (Cousineau *et al.*, 1998). My work suggests that the retrohoming pathway of IntAs differs from those of other group II introns.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.

Belfort, M., M.E. Reaban, T. Coetzee, and J.Z. Dalgaard. 1995. Prokaryotic introns and inteins, a panoply of form and function. *J. Bacteriol.* 177, 3897-3903

Cousineau, B., D. Smith, S. Lawrence-Cavanagh, J.E. Mueller, J. Yang, D. Mills, D. Manias, G. Dunny, A.M. Lambowitz, and M. Belfort. 1998. Retrohoming of a bacterial group II intron, mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* 94, 451-462.

Fayet, O., P. Ramond, P. Polard, M.F. Prere, and M. Chandler. 1990. Functional similarities between retroviruses and the IS*3* family of bacterial insertion sequences? *Mol. Microbiol.* 4, 1771-1777.

Ferat, J.L., M. Le Gouar, and F. Michel. 1994. Multiple group II self-splicing introns in mobile DNA from *Escherichia coli*. C R Acad. Sci. III 317, 141-148.

Ferat, J.L. and F. Michel. 1993. Group II self-splicing introns in bacteria. *Nature* 364, 358-361.

Han, C.-G., Y. Shiga, T. Tobe, C. Sasakawa, and E. Ohtsubo. 2001. Structural and functional characterization of IS*679* and IS*66*-family elements. *J. Bacteriol.* 183, 4296-42304.

Hu. P., J. Elliott, P. McCready, E. Skowronski, J. Garnes, A. Kobayashi, R.R. Brubaker, and E. Garcia. 1998. Structural organization of virulence-associated plasmids of *Yersinia pestis*. *J. Bacteriol.* 180, 5192-5202.

Khan, E., J.P. G. Mack, R.A. Katz, J. Kulkosky, and A.M. Skalka. 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* 19, 851-860.

Knoop, V. and A. Brennicke. 1994. Evidence for a group II intron in *Escherichia coli* inserted into a highly conserved reading

frame associated with mobile DNA sequences. *Nucleic Acids Res.* 22, 1167-1171.

Kulkosky, J., K.S. Jones, R.A. Katz, J.P. G. Mack, and A.M. Skalka. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* 12, 2331-2338.

Lindler, L.E., G.V. Plano, V. Burland, G.F. Mayhew, and F.R. Blattner. 1998. Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. *Infect. Immun.* 66, 5731-5742.

Martinez-Abarca, F., S. Zekri, and N. Toro. 1998. Characterization and splicing in vivo of a *Sinorhizobium meliloti* group II intron associated with particular insertion sequences of the IS*630*-Tc*1*/IS*3* retroposon superfamily. *Mol. Microbiol.* 28, 1295-306.

Matsutani, S. and E. Ohtsubo. 1990. Nucleotide sequence of insertion element IS*629*. *Nucleic Acids Res.* 18, 1899.

Matsuura, M., R. Saldanha, H. Ma, H. Wank, J. Yang, G. Mohr, S. Cavanagh, G.M. Dunny, M. Belfort, and A.M. Lambowitz. 1997. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: Biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* 11, 2910-2924.

Michel, F. and J.L. Ferat. 1995. Structure and activities of group II introns. *Annu. Rev. Biochem.* 64, 435-461.

Michel, F., K. Umesono, and H. Ozeki. 1989. Comparative and functional anatomy of group II catalytic introns-a review. *Gene* 82, 5-30.

Mills, D.A., D.A. Manias, L.L. McKay, and G.M. Dunny. 1997. Homing of a group II intron from *Lactococcus lactis* subsp. lactis ML3. *J. Bacteriol.* 179, 6107-6111.

Mills, D.A., L.L. McKay, and G.M. Dunn. 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J. Bacteriol.* 178, 3531-3538.

Mullany, P., M. Pallen, M. Wilks, J.R. Stephen, and S. Tabaqchali. 1996. A group II intron in a conjugative transposon from the gram-positive bacterium, *Clostridium difficile*. *Gene* 174, 145-150.

Ohtsubo, E. and Y. Sekine. 1996. Bacterial insertion sequence. *Curr. Top Microbiol. Immunol.* 204, 1-26.

Pearson, W.R. and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85, 2444-2448.

Prère, M.F., M. Chandler, and O. Fayet,. 1990. Isolation and analysis of IS*911*, a new member of the IS*3* group of insertion sequences. *J. Bacteriol.* 172, 4090-4099.

Rajakumar, K., C. Sasakawa, and B. Adler. 1997. Use of a novel approach, termed island probing, identifies the *Shigella flexneri she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect. Immun.* 65, 4606-4614.

Rauch, P.J., M.M. Beerthuyzen, and W.M. de Vos. 1990. Nucleotide sequence of IS*904* from Lactococcus lactis subsp. lactis strain NIZO R5 *Nucleic Acids Res.* 18, 4253-4254

Rezsöhazy, R., B. Hallet, J. Delcour, and J. Mahillon. 1993. The IS*4* family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* 9, 1283-1295.

Schwartz, E., M. Kroeger, and B. Rak. 1988. IS150: distribution, nucleotide sequence and phylogenetic relationships of a new *E. coli* insertion element. *Nucleic Acids Res.* 16, 6789-6802.

Shearman, C., J.J. Godon, and M. Gasson. 1996. Splicing of a group II intron in a functional transfer gene of *Lactococcus lactis*. *Mol. Microbiol.* 21, 45-53.

Takami, H., C.-G. Han, Y. Takaki, and E. Ohtsubo. 2001. Identification and distribution of new insertion sequences in the genome of alkaliphilic *Bacillus halodurans* C-125. *J. Bacteriol.* 183, 4345-4356.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680

Timmerman, K.P. and C.-P. Tu. 1985. Complete sequence of IS*3*. *Nucleic Acids Res.* 13, 2127-2139.

Tobe, T., T Hayashi, C.-G. Han, G.K. Schoolnik, E. Ohtsubo, and C. Sasakawa. 1999. Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. *Infect. Immun.* 67, 5455-5462.