JOURNAL OF BACTERIOLOGY, Aug. 2001, p. 4395–4404 Vol. 183, No. 15
0021-9193/01/$04.00+0 DOI: 10.1128/JB.183.15.4395–4404.2001

Transposition of IS*1397* in the Family *Enterobacteriaceae* and First Characterization of ISKpn*1*, a New Insertion Sequence Associated with *Klebsiella pneumoniae* Palindromic Units

CAROLINE WILDE, SOPHIE BACHELLIER, MAURICE HOFNUNG, AND JEAN-MARIE CLÉMENT*

Unité de Programmation Moléculaire et Toxicologie Génétique, CNRS URA 1444, Institut Pasteur, 75724 Paris Cedex 15, France

Received 23 February 2001/Accepted 15 May 2001

IS*1397* and ISKpn*1* are IS*3* family members which are specifically inserted into the loop of palindromic units (PUs). IS*1397* is shown to transpose into PUs with sequences close or identical to the *Escherichia coli* consensus, even in other enterobacteria (*Salmonella enterica* serovar Typhimurium, *Klebsiella pneumoniae*, and *Klebsiella oxytoca*). Moreover, we show that homologous intergenic regions containing PUs constitute IS*1397* transpositional hot spots, despite bacterial interspersed mosaic element structures that differ among the three species. ISKpn*1*, described here for the first time, is specific for PUs from *K. pneumoniae*, in which we discovered it. A sequence comparison between the two insertion sequences allowed us to define a motif possibly accounting for their specificity.

The chromosome of *Escherichia coli* contains various families of small extragenic sequences repeated from 6 to more than 250 times and representing almost 2% of the bacterial DNA (4). The BIME family (bacterial interspersed mosaic elements) is one of these. The basic motif of BIMEs is the palindromic unit (PU) or repetitive extragenic palindromic sequence (19). PUs are imperfect palindromes of about 40 bp that are transcribed but not translated (4). Five hundred eighty-four PUs are scattered over the chromosome of *E. coli* (6) and have been divided into three classes, Y, Z^1, and Z^2, according to slight variations in sequence and size. BIMEs have been defined as a precise combination of PUs alternating in orientation and type and associated with seven extra-PU motifs (A, B, S, L, s, l, and r), and two major BIME families can be distinguished, called BIME-1 and BIME-2 (3, 18). For a review of PUs and BIMEs, see reference 4 and the *E. coli* short DNA repeats section of the Unit of Molecular Programming and Genetic Toxicologyweb site (http://www.pasteur.fr/recherche/unites/pmtg/repet/index.html). The existence of a general function for BIMEs is still unclear. Some of them can stabilize mRNAs (26, 27) or play a role in transcription termination (14), translational control (38), and genomic rearrangements (36). However, their specific interactions with integration host factor (7, 28), DNA gyrase (12, 42), and DNA polymerase (16) suggest that they may play a role in the functional organization of the bacterial nucleoid.

PUs were detected originally in *E. coli* and *Salmonella enterica* serovar Typhimurium (13, 19) and later in other enterobacteria by Southern blot hybridization and sequence analysis (17). The PU consensus for *S. enterica* serovar Typhimurium differs slightly from the *E. coli* consensus in that there is an additional G between positions 10 and 11 (15). There is another PU motif, D, specific to *Salmonella* and *Klebsiella* (4). *Klebsiella* PUs are closely related to *S. enterica* serovar Typhimurium PUs (2), but they are more GC rich than the rest of the genome. The two species exhibit BIME-like structures, but extra-PU motifs seem to be less conserved than in *E. coli*. There is no L motif in *S. enterica* serovar Typhimurium, and BIMEs containing PUs in direct tandem repeats are present in these enterobacteria.

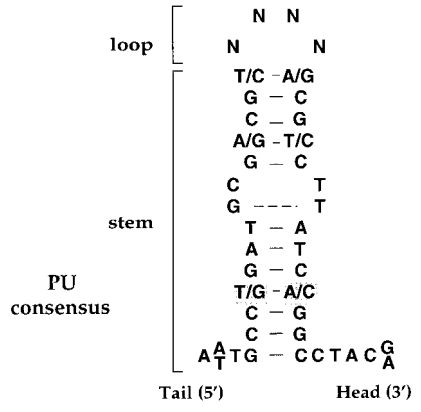
IS*1397* is a 1,432-bp insertion sequence belonging to the IS*3* family (5). It has been found in several natural *E. coli* isolates and is always inserted into the central part of a PU described as the loop (Fig. 1). We have recently shown that IS*1397* is an active insertion sequence that is able to be transposed into *E. coli* from a donor plasmid and that it is inserted specifically into PUs (9). In this study, we analyzed whether IS*1397* has the same target specificity in other *Enterobacteriaceae* species which also contain PUs, i.e., *S. enterica* serovar Typhimurium, *Klebsiella pneumoniae*, and *K. oxytoca*. We also describe a new IS, ISKpn*1*, that was identified during analysis of the PUs of *K. pneumoniae*. This insertion sequence (IS) is closely related to IS*1397* and IS*150* and is also specifically inserted into PUs.

MATERIALS AND METHODS

Plasmids. pNABI, the plasmid used to study the transposition of IS*1397* into the chromosome of *K. pneumoniae*, has already been described (9). Its structure is shown in Fig. 2. This derivative of pACYC184 (8) carries a P15A origin of replication, an *orfAB* artificial gene with a disruption of the palindrome found in the putative frameshift window between *orfA* and *orfB*. However, when we sequenced this region, we found that the expected deletion of an A introduced to create an in-frame fusion between the two genes had not been achieved. As a consequence, only OrfA, and not OrfAB, could be expressed under the control of *Ptac* (isopropyl-β-D-thiogalactopyranoside [IPTG] inducible due to the presence of a functional *lacI*^{q} gene on the plasmid). Overexpression of OrfA is toxic to the cells (J.-M. Clément and C. Wilde, unpublished observations). pNABI also contains IS*1397*, flanked by an interrupted PU sequence with a 4-bp duplication (as naturally found between the *mtlA* and *mtlD* genes in EPEC25 [5]) and with a Km^r-encoding gene inserted downstream of *orfB*.

Another plasmid, pBLOCK, was constructed to study transposition into the chromosomes of *K. oxytoca* and *S. enterica* serovar Typhimurium. It was created to allow selection and quick analysis of transposition in virtually any bacterial species. This derivative of pACYC184 carries the P15A origin of replication and *lacI*^{q}. The Cm^r-encoding gene is placed next to the P15 origin of replication. The

* Corresponding author. Mailing address: Unité de Programmation Moléculaire et Toxicologie Génétique, CNRS URA 1444, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France. Phone: (33) 01 40 61 32 88. Fax: (33) 01 45 68 88 34. E-mail: jclement@pasteur.fr.



BIME structure

FIG. 1. PU structure and BIME organization. (Top) Consensus sequence and hypothetical DNA structure of PUs. Two types of PUs have been defined according to two critical positions of the consensus (7 and 32, boxed in grey). In the Y PUs, these nucleotides are a G and a C, respectively, while in the Z PUs, they are a T and an A. These sequences form an imperfect palindrome, with asymmetry elements which allow orientation of the structure from the tail to the head, and confer a stem-loop structure. (Bottom) BIMEs are composed of successive PUs alternating both in type and in orientation (indicated by black triangles). They are separated by short "extra-PU" motifs located either between two heads (head internal sequences [HIS]) or two tails (tail internal sequences [TIS]). Two tail external sequences (TES), A and B, flanking the tails of the last PUs, are found in some BIMEs (18).

artificial $P_{lac/ara-1}$ promoter from pPROLar.A (Clontech) is found twice, upstream of *orfA* and upstream of *orfAB*, allowing overproduction of the two proteins upon IPTG induction. The last 93 bp of the *lacI*^{q} sequence are found downstream of *orfA*, due to the cloning strategy. OrfA and OrfAB are both toxic to the cell. The transposable module (delineated with arrows in Fig. 2) is not flanked by the interrupted PU and is different from the one found in pNABI. The *orfA* and *orfB* sequences were replaced with the R6K origin of replication (35, 37), which allows stable maintenance of plasmids in strains expressing II protein, such as BW19610. Most of the DNA fragments corresponding to these various parts were generated by PCR. In this case, their sequence was systematically checked after cloning into the recombinant plasmids. As for pNABI, the construction of pBLOCK involved many steps, the details of which will be supplied upon request (jclement@pasteur.fr). We checked that pNABI and pBLOCK enabled the transposition of their respective modules into the *E. coli* chromosome with the same target specificity (J.-M. Clément and F. Le Noanne, unpublished observations).

Media and bacterial strains. Luria-Bertani (LB) medium was used to grow all of the strains. Kanamycin (KM) and chloramphenicol (CM) were used at concentrations of 25 and 50 μg/ml, respectively. IPTG was used at a final concentration of 10^{-3} M, and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) was used at 40 μg/ml.

The strains used for plasmid rescue were JM109 [*recA1 endA1 gyrA96 thiA hsdR17*(r_K^- m_K^+) *relA1 supE44* Δ(*lac-proAB*) F′ *traD proAB lacI*^{q}*Z*ΔM15] (43), BW19610 [DE3(*lac*)*X74* Δ*uidA*::*pir-116 recA1* Δ*phoA532* Δ(*phnC?D-P*)33–30] (25), and TOP10F′ {F′ [*lacI*^{q} Tn*10*(Tet^r)] *mrcA* Δ(*mrr-hsdRMS-mrcBC*) ϕ80*lacZ*ΔM15 Δ*lacX74 recA1 deoR araD139* Δ(*ara-leu*)*7697 galU galK rpsL* (Str^r) *endA1 nupG*} (Invitrogen). The strains in which transposition was studied were *K. pneumoniae* subsp. *pneumoniae* ATCC 13883, *K. oxytoca* ATCC 13182, and *S. enterica* serovar Typhimurium 3261. The strains used to study ISKpn*1* distribution were *K. pneumoniae* MGH78578 (a gift from M. McClelland, Genome Sequencing Center [GSC], Washington, University, St. Louis, Mo.), *K. pneumoniae* subsp. *pneumoniae* ATCC 13883, *K. pneumoniae* subsp. *ozonae* ATCC 11296, *K. pneumoniae* subsp. *rhinoscleromatis* ATCC 13884, *K. oxytoca* ATCC 13182, *K. planticola* ATCC 33531, *K. terrigena* ATCC 33257, *Enterobacter aerogenes* ATCC 13048, *K. aerogenes* W70 (20), *E. coli* EPEC 25 (22), *E. coli* C600 (1), *Yersinia pestis* 6.69, *Y. enterocolitica* 8081, and *Y. pseudotuberculosis* 32953. The *Yersinia* strains belong to the Pasteur Institute collection and were a gift from Elisabeth Carniel.

DNA techniques. Restriction enzymes and DNA-modifying enzymes were purchased from New England Biolabs or Boehringer Mannheim and used as recommended. Plasmid DNA manipulations were carried out by using standard procedures (30). Chromosomal DNA extractions were performed with the DNA Easy Tissue kit (Qiagen). PCRs were performed by using the Amersham PCR kit as recommended with a Mastercycler gradient apparatus (Eppendorf).

Oligonucleotides. Oligonucleotides were purchased from Genset. Two oligonucleotides were used to sequence the junctions of IS*1397* chromosomal insertions. Their sequences are as follows: seqIRL, 5′CGGTTGTGGACAACAAG CCAGGG3′ (complementary to a region of the R6K origin of replication); Kmseqout, 5′CACGAGGCAGACCTCAGCGC3′ (corresponding to a region located between the end of the Km^r-encoding gene and the right inverted repeat (IRR) of the module, as found on the plasmid pBLOCK).

Two other oligonucleotides were used for ISKpn*1* cloning, i.e., upKp (5′GC GAATAGCCGGCTGAAAACGTGAG3′) and downKp (5′GGTGGTCATTT CTCAAGGCGAGG3′), flanking the IS of *K. pneumoniae* on contig 840 (May 2000), according to the GSC web site (http://genome.wustl.edu/gsc/index.shtml).

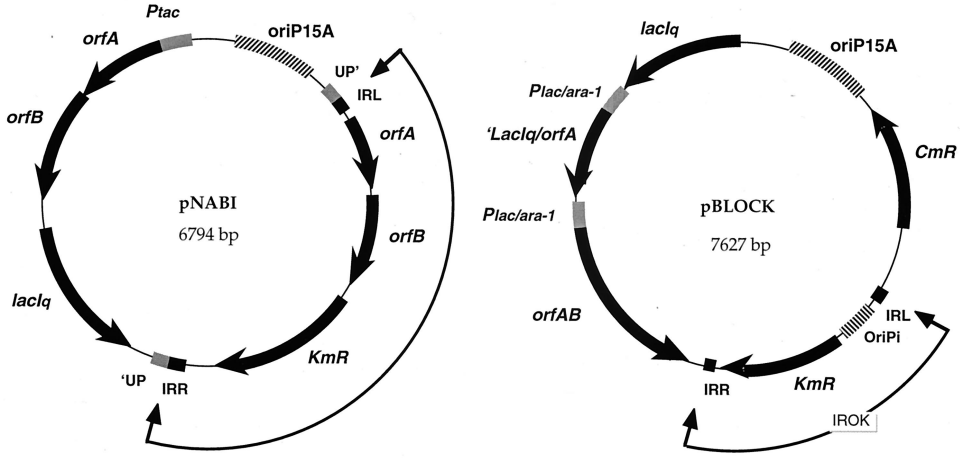


FIG. 2. Plasmids pNABI and pBLOCK. Both plasmids are derived from pACYC184 (p15A origin of replication, hatched box). The orientations of the different genes (thick lines) are indicated by arrows. The transposable module is located in each case, between two arrows (IROK in the case of pBLOCK). IRL and IRR are the left and right inverted repeats from IS*1397*, respectively.

K. pneumoniae PU consensus [a]	TTT SCCGGGTGGCGG CTG ACGCYTTACCSGGC CTACR		
Clone	**insertion site[b]**	**Region structure[c]**	
Within PUs			
Kp 6.1	TAT GCCGGATGGCGG CTA ACGCCTTATCCGGTC CTACC	*yjgA*→	PU→ h ←PU t PU→ h ←PU t PU'(⇐)'PU→ ***
Kp 8.2	AAT GCCGGATGGCGG CTA ACGCCTTATCCGGC CTACC	*yjgA*→	PU'(⇒)'PU → h ←PU t PU→ h ←PU t PU→ ***
Kp 12.1	TCT GCCGGATAGCGG CTT ACGCCTTATCCGGCC TTACG	*narU*→	PU'(⇒)'PU→ *narZ*→
Kp D	TCT GCCGGATAGCGG CTT ACGCCTTATCCGGCC TTACG	*narU*→	PU'(⇐)'PU→ *narZ*→
Kp 7.1	TTC TCCGGATTGGCGGCATAA ATGCCTTATCCGGC CTACG	***	PU→ h ←PU t PU'(⇒)'PU→ h ←PU ***
Kp 9.1	TGT GCCGGATGGCGG CACAA ATGCCTTATCCGGC CTACG	***	←PU t PU→ h ←PU'(⇒)'PU t PU→ h ←PU t PU→ ***
Kp C	TGT GCCGGATGGCGG CATAA ATGCCTTATCCGGC CTACG	***	← PU'(⇒)'PU t PU→ h ←PU t PU'(ISKpn*1*)'PU→ ***
Kp E	AAA GCCGGATGGCGG CTA GCGCCATAT CGGC CTACG	←*yjiX*	PU'(⇒)'PU→ h ←PU ←*yjiY*
Outside PUs			
Kp 10.1	TGG GCCGGATGGGCCGA CAG CAGCTGCAGCTCGCCCGC CTCTG	*bisC*→	(⇒) →*bisC*
Kp G	CCT GCCGGATGACACGT TTACCCGCAGCAGGTCAGGTGC CGCAG	***	(⇒) ***
Kp 11.1	CAG GCCGGATGGAACGG GGT ATGCGGATGCCAGGCCCGCGACG	***	(⇐) ***
Kp P	ACG GCCTTCGGGGGCGA AGTGGCGGTTACCCCATTCCGC CAGCG	***	(⇒) ***
Kp 3.1	ACT TCCACCAGCAACGG AACCGCCAGCGTCGTACTGCTTTCCATC	→*polA*	(⇒) →*polA*
Insertion consensus	WNY GCCGGATGGCGG CWW AYGCCTTATCCGGC CTACS		

FIG. 3. Sites of insertion of the transposable module into the chromosome of *K. pneumoniae*. *a*, The PU consensus is derived from this study. D is A, G, or T; S is G or C; W is A or T; and Y is C or T. *b*, Nucleotides which are duplicated after transposition are underlined. *c*, Single arrows indicate PU and gene orientations. The double arrow represents the oriented transposable module. The letters h and t represent head internal sequences (HIS) and tail internal sequences (TIS), respectively (Fig. 1). Gene names have been assessed by sequence homology to *E. coli* genes (http://genolist.pasteur.fr/Colibri/). Triple asterisks indicate a region for which no homolog has been identified.

DNA sequencing. DNA sequencing was performed either as previously described (9) or by MWG-Biotech AG and ESGS-Cybergene.

Southern blot hybridization. Samples (1 to 5) μg of chromosomal DNAs were loaded on a 1% agarose Tris-acetate gel. DNA transfer onto a Hybond N^+ membrane (Amersham) was performed as previously described (9). Hybridization was carried out by using the DIG Nucleic Acid Labeling and Detection system (Boehringer Mannheim) at 56°C. The probes used were digoxigenin-dUTP labeled with a Promega nick translation kit.

Selection of transposition events. Transposition events in *K. pneumoniae* were selected and studied as previously described for *E. coli* (9). We used a slightly different strategy to study transposition events in *K. oxytoca* and *S. enterica* serovar Typhimurium. Independent clones of *K. oxytoca* and *S. enterica* serovar Typhimurium transformed with pBLOCK were grown overnight at 37°C in liquid LB medium containing KM and CM. A 5-μl volume of each culture was plated on LB medium containing KM and IPTG, incubated overnight at 37°C, and replica plated onto LB medium containing IPTG and either CM or KM. After a 24-h incubation at 37°C, two colonies per plate that grew on KM but not on CM were streaked on LB medium containing KM and IPTG and grown overnight at 37°C in the same liquid medium for chromosomal DNA extraction. Chromosomal DNAs were digested with *Mlu*I (which has a single site in pBLOCK, in the *lacI* gene), and 1 to 5 μg was electrophoresed on a 1% agarose gel and submitted to Southern blot hybridization. The probe used was IROK (Fig. 2), which corresponds to the R6K origin-Km^r module from pBLOCK. For the cloning of chromosomal fragments containing IROK, *Mlu*I-digested chromosomal DNAs were circularized and *E. coli* BW19610 cells were transformed by electroporation. For long fragments, chromosomal DNAs were digested with *Mlu*I and *Bst*EII (none of which have restriction sites in IROK) and ligated before transformation of BW19610 cells. In both cases, recombinant clones were selected on LB medium plates with KM. After DNA sequencing, chromosomal regions flanking the module were identified by using the FASTA software of Infobiogen (http://www.infobiogen.fr/) and the BLAST program of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).

***K. pneumoniae* MGH78578 genomic library.** Fragments of *Pst*I-digested *K. pneumoniae* MGH78578 genomic DNA ranging from 2.3 to 6 kb were purified from an agarose gel (Qiaquick gel extraction kit; Qiagen) and ligated to pUC18 treated with *Pst*I and dephosphorylated with shrimp alkaline phosphatase. TOP10F′ cells were transformed (One Shot Transformation Reaction; Invitrogen) in accordance with the manufacturer's indications with the ligation mixture. We obtained 9,000 white colonies on LB medium plates containing ampicillin, X-Gal, and IPTG. A colony blot hybridization was done with a probe obtained as follows. The IS was amplified by PCR on *K. pneumoniae* chromosomal DNA with the primers downKp and upKp, which are located on each side of the IS on contig 840. The PCR product was cloned into the PCRII-TOPO vector (TOPO TA Cloning; Invitrogen). The probe was the 1,067-bp *Acc*I-*Hin*dIII fragment from ISKpn*1* cloned into PCRII-TOPO. It contained the first 976 bp of the IS and 91 bp from the vector. We obtained one positive clone, which was streaked on LB medium plates with ampicillin and checked for positive hybridization with the same probe. The initial sequencing steps were done with the universal primers forward-40 and reverse, and the sequencing was completed with specific primers.

Distribution of ISKpn*1*. The genomic DNAs of the strains studied were digested with *Hin*dIII and *Mlu*I, and 1 μg of each was loaded onto a 1% agarose gel for Southern blot hybridization with the probe used to screen the *K. pneumoniae* MGH78578 genomic library.

Determination of a global PU consensus for *K. pneumoniae*. We investigated *K. pneumoniae* MGH78578 sequences present in the contigs released by the GSC (http://genome.wustl.edu/gsc/index.shtml) by using the *E. coli* PU consensus and the consensus derived from the transposition sites of IS*1397* in *K. pneumoniae* as query sequences for the BLAST (http://www.ncbi.nlm.nih.gov/) and FASTA (http://www.infobiogen.fr/) programs, and we checked all matches by eye to see whether they were PUs. We aligned the 242 PUs detected with the clustalw software (parameters: gap opening penalty, 10; gap extension penalty, 0.20; gap separation penalty range, 8) and reformatted the result to an msf format with ftmseq in order to use Pretty (with plurality = 3), a Genetics Computer Group, Inc., program. Pretty displayed multiple-sequence alignments and determined the consensus sequence shown in Fig. 3.

Nucleotide sequence accession number. The nucleotide sequence of the IS described here (see Fig. 6) was deposited in the GenBank database under accession no. AF345899.

RESULTS

Transposition events. We studied the transposition of IS*1397* in several bacterial species by using either pNABI (9), which was previously used in *E. coli*, or a derivative, pBLOCK.

Selection of transposition events with pNABI relied on the Km^r $IPTG^r$ $LacI^-$ phenotype, which was the consequence of both transposition of the module into the chromosome and loss of pNABI. However, $IPTG^r$ could result from mutations in *orfAB* and $LacI^-$ phenotype screening of LacZ or $LacI^+$ bacterial strains was not possible. This is why we used pBLOCK, in which the presence of both *orfA* and *orfAB* minimizes the selection of $IPTG^r$ clones due to mutations in these genes. Moreover, loss of the donor plasmid results in loss of Cm^r. It should be noted that the p15A origin of replication is located between *cat* and *orfAB* (Fig. 2), so that a deletion encompassing these genes would not allow the maintenance of an autonomously replicating plasmid. Another advantage of pBLOCK is that the chromosomal insertion sites can propagate after circularization in a Pir^+, strain since they are associated with the R6K origin of replication.

We grew 16 separate liquid cultures of 5×10^8 CFU of independent pNABI-containing *K. pneumoniae* clones, which were plated on LB medium plates containing KM and IPTG. For each culture, we obtained resistant colonies, which were tested for loss of the plasmid; they were streaked on LB medium plates with X-Gal (no IPTG added). In all but one case, we observed blue colonies. In order to check the homogeneity of the transposition events, two blue colonies originating from the same liquid culture were then examined for the presence of an IS*1397*-Km^r insertion in the chromosome. *Bgl*II digests of chromosomal DNA were analyzed by Southern blot hybridization using an internal IS*1397* DNA fragment as a probe (data not shown). We observed a single faint band after transposition and/or several heavily labeled bands when the bacteria had kept the plasmid. In seven cases, the two clones tested displayed different profiles; in four out of these seven cases, one of the clones still had the plasmid and was discarded. In eight cases, the two clones tested exhibited the same profile; in three out of these eight cases, they had kept the plasmid and were discarded. Hence, we retained 15 clones resulting from independent transposition events for further analysis. We successfully cloned 13 different *Bgl*II chromosomal DNA fragments containing the Km^r-encoding gene in pUC18 and sequenced the junctions between the transposition module and the chromosome (Fig. 3).

Transposition events in *K. oxytoca* (*lacZ*$^+$ *lacI*$^+$) and *S. enterica* serovar Typhimurium (Lac^-) were investigated by using pBLOCK. In the presence of IPTG, both OrfA and OrfAB are expressed, which are each lethal to the bacteria. $IPTG^r$ Km^r colonies could be isolated after overnight culture at 37°C (see Materials and Methods); such a phenotype could be due to (i) a deletion or mutation within *orfA* and *orfAB* resulting in nontoxic proteins (and the bacteria would keep the plasmid) or (ii) transposition of IROK into the chromosome with loss of pBLOCK. These two events could be discriminated by checking the sensitivity of the strains to CM, indicating loss of the plasmid.

Approximately 5×10^6 CFU from 12 liquid cultures of independent *K. oxytoca* and *S. enterica* serovar Typhimurium clones containing pBLOCK were plated on LB medium containing IPTG and KM. We obtained about 1,000 resistant colonies per plate that we replica plated on CM-containing medium. Ninety-nine percent of the *S. enterica* serovar Typhimurium colonies and 1% of the *K. oxytoca* colonies were sensitive to CM. Two Km^r $IPTG^r$ Cm^s clones from each plate were examined for the presence of IROK in the chromosome. *Mlu*I digests of chromosomal DNA were analyzed by Southern blot hybridization using IROK as a probe (data not shown). All *S. enterica* serovar Typhimurium candidates but one showed different profiles between the two clones, and one clone still contained the plasmid. No *K. oxytoca* candidates had kept the plasmid, and in three cases, the profiles of the two clones tested were different. We circularized the fragments and transformed *E. coli* BW19610, a strain allowing the replication of the circular fragments containing the R6K origin. We sequenced the junctions between the transposition module and the chromosome for 11 *S. enterica* serovar Typhimurium clones (Fig. 4) and 15 *K. oxytoca* clones (Fig. 5).

Target specificity in enterobacteria. Our previous results obtained with *E. coli* showed that IS*1397* has a tight specificity of insertion into PUs (5, 9). All integrations of the transposable module into the *S. enterica* serovar Typhimurium, *K. pneumoniae*, and *K. oxytoca* chromosomes were transposition events with a 3- or 4-bp duplication. All of the insertions obtained in *S. enterica* serovar Typhimurium occurred within the loop of a PU (Fig. 4), whereas we obtained PU and non-PU integrations for the two other enterobacteria. In the case of *K. pneumoniae* (Fig. 3), we observed eight insertions into PUs; two insertions (Kp 10.1 and Kp G) in which the transposition module was inserted 6 nucleotides after the sequence 5′GCC GGATG3′, which is found in a PU stem; two insertions into extragenic sequences which are not PUs (Kp 11.1 and Kp P); and one intragenic transposition at the end of *polA* (Kp 3.1). In *K. oxytoca* (Fig. 5); there were 12 insertions into PUs and three into genes (Ko 1.2, Ko 2.2, and Ko 5.1), 5 to 7 nucleotides after the 5′GCC(G/T)GAT3′ sequence found in the PU stem. Interestingly, we observed hot spots of insertions in either orientation within and among the three species. There were six independent transposition events (St 2.2, St 6.1, Ko 4.1, Ko 5.2, Ko 11.1, and Ko 12.2) into a PU from the *yjiX-yjiY* intergenic region, four into the *fucR-ygdE* region (Ko 6.1, Ko 7.1, Ko 7.2, and Ko 10.2), three into a PU next to the same open reading frame (Ko 2.1, Ko 3.1, and Ko 8.1), and three (Ko 12.1, Kp 12.1, and Kp D) into a PU from the *narU-narZ* region.

Description of a new IS associated with *K. pneumoniae* PUs. In the course of a computer analysis of the different PU types in *K. pneumoniae* MGH78578, we discovered a new IS that is present in five contigs (at the GSC), always inserted into the loop of a PU with a 3-bp duplication, like IS*1397*. We cloned this IS from a genomic library by using a PCR fragment containing the IS as a probe. The PCR was performed on genomic DNA by using unique sequences flanking the IS on contig 840 as primers and yielded a very small amount of the fragment. We established the definitive sequence of the IS from a library clone that displayed flanking regions different from what was expected, suggesting misassembly of the contigs and explaining the poor yield of the PCR (see Discussion). The IS was named ISKpn*1* (Fig. 6). ISKpn*1* is 1,445 bp long and belongs to the IS*3* family, according to sequence homology and structural fea-

Clone	PU consensus[a]/ insertion site[b]	Region structure[c]		
Y consensus	WTT GCCGGATGGCGGCG TAA ACGCCTTATCCGGC CTACR			
St 3.2	ATT GCCGGATGGCGGC TAA CGCCTAATCCGGC CTACA	*sbp*←	Y'(⇐)'Y→ h ←Z^1	←*pfkA*
St 1.1	TAT GCCGGATAGCGGCG TAA ACGCCTTATCCGGC CTACC	*orf*→	Y'(⇐)'Y→ h ←Z^1	***
Z^1 consensus	ATT GCCTGATGGC GC T RCG CTTATCAGGC CTACR			
St 6.2	TAT GCCGGATGGC GC T TCG CTTATCCGGC CTACA	*caiD*→	Z^1'(⇒)'Z^1→ h ←Y	→*caiE*
St 7.1	ATT GCCGG TGGC GC T TTG CTTATCCGGC CTACA	*gltP*→	Z^1'(⇐)'Z^1→ h ←Y 9 Z^1→	←*yjcO*
St 11.2	ATT GCCGGATGAC GC T TCG CTTATCCGGC CTACG	*gpp*→	Z^1'(⇒)'Z^1→	←*rep*
Z^2 consensus	WTY GCCKGATGGCGACGCTRWMGCGTCTTATCMGGC CTACA			
St 5.2	CAT GCCGGATGGCGACGCTG CGCGTCTTATCCGGC CTACC	*eutN*←	←Z^1 t Z^2'(⇒)'Z^2→	←*eutM*
St 12.2	CGT GCCGGATGGCGACGC GACGCGTCTTATCCGGC CTACA	*panB*→	Z^2'(⇐)'Z^2→ h ←Z^1	→*panC*
D consensus	WWY GCCGGATGGCGGC TT CGCCTTATCCGGC CTACA			
St 2.1	AAT GCCGGATGGCGGC TTTA CCTTATCAGGC CTACA	*orf*→	D'(⇒)'D→ h ←Z^1	←*tpiA*
St 2.2	AAA GCCGGATGGCGGC TT GCGCCTTATCCGGC CTACA	*yjiX*←	D'(⇐)'D→	←*yjiY*
St 6.1	AAA GCCGGATGGCGGC TT GCGCCTTATCCGGC CTACA	*yjiX*←	D'(⇐)'D→	←*yjiY*
St 5.1	ATT GCCGGATAGCGGC TC CGCCTTATCTGGC CTGCG	***	D'(⇐)'D→	→*fdhF*
Insertion consensus	HWW GCCGGATRGCGRC TWA CGCCTTATCMGGC CTACV			

FIG. 4. Sites of insertion of the transposable module into the chromosome of *S. enterica* serovar Typhimurium. *a*, The PU consensuses are taken from reference 4 and the Unit of Molecular Programming and Genetic Toxicology web site (http://www.pasteur.fr/recherche/unites/pmtg/repet/index.html). *b*, Nucleotides which are duplicated after transposition are underlined. H is A, C, or T; M is A or C; R is A or G; V is A, C, or G; and W is A or T. *c*, Single arrows indicate PU and gene orientations. The double arrow represents the oriented transposable module; h and t represent head internal sequences (HIS) and tail internal sequences (TIS), respectively (Fig. 2). Gene names have been assessed by sequence homology to *E. coli* genes (http://genolist.pasteur.fr/Colibri/). Triple asterisks indicate a region for which no homolog has been identified. *orf* indicates a potential coding sequence with no homolog in the databases.

tures. It is flanked with 25-bp imperfect terminal inverted repeats containing six mismatches, a left inverted repeat (IRL) and an IRR, ending with the dinucleotide 5′-CA-3′ (29). ISKpn*1* contains two open reading flames on the same strand, *orfA* and *orfB*, which is in −1 frame with respect to *orfA*. *orfA* extends from ATG (position 50) to TGA (position 571) and is preceded by a ribosome-binding site located 8 bp upstream of the start codon. *orfA* could encode a 173-amino-acid (aa) protein, OrfA, containing a putative α-helix–turn–α-helix (or HTH) motif (aa 23 to aa 40; boxed in Fig. 6) with a probability of 71% (the standard deviation score obtained by comparison with a weight matrix was 3.98 [11]). *orfB* (nucleotides 568 to 1413) could begin with the rarely used CTG start codon located between the frameshift window and the HTH motif (see below) and encode a putative protein of 281 aa, OrfB, which also contains a putative HTH motif (aa 10 to 31, boxed in Fig. 6) with a probability of 71% (standard deviation score, 3.78). ISKpn*1*, like all members of the IS*3* family, is characterized by the presence of a conserved region in OrfB called the D,D(35)E motif that is associated with several additional residues (21, 24, 29). This triad is involved in catalysis and is also present in retroviral integrases. Another characteristic of IS*3* family members is the presence of an A_6G sequence followed by a dyad symmetry, which enables the formation of a fusion protein containing OrfA and OrfB, called OrfAB, the putative transposase, resulting from a −1 translational frameshift. This kind of structure, called a frameshift window, has already been described for IS*3* (33), IS*150* (41), IS*1* (31, 32), and retroviral integrases (40), for example. Screening of the SP-GLOBAL database with the putative OrfA protein sequence gave the best scores with IS*3* family members IS*150*, IS*1397*, and IS*1223*, with 38.5, 37.6, and 35.6% identity, respectively. The same screen performed with the putative OrfB protein sequence gave 63.3 and 56.6% identity with the IS*150* and IS*1397* OrfB proteins, respectively. Screening of the bacterial section of the GenBank database with the nucleotide sequence of ISKpn*1* gave the best score with IS*150* (62.3% identity). Thus, ISKpn*1* is, interestingly, closer in both nucleotide and protein sequences to IS*150* than to IS*1397*, although ISKpn*1* and IS*1397* share the property of insertion into PUs (see Discussion).

Distribution of ISKpn*1* among several enterobacterial species. We studied the distribution of ISKpn*1* among several *Klebsiella* and *Yersinia* species and two *E. coli* strains by Southern blot hybridization (Fig. 7). Chromosomal DNAs were digested by *Mlu*I and *Hin*dIII, and the probe (see Materials and Methods) was specific for ISKpn*1* since it did not hybridize with *E. coli* K-12, which contains IS*150*, or cross-hybridize with IS*1397* (the very slight signal is nonspecific, since the molar ratio of the IS*1397* fragment [Fig. 7, lane 16] to the genomic DNAs ([Fig. 7, lanes 2 to 15] is 500). Figure 7 shows that ISKpn*1* is present only in *K. pneumoniae* MGH78578, *K. pneumoniae* subsp. *pneumoniae* (used to study the transposition of IS*1397*), *K. pneumoniae* subsp. *ozonae*, and *K. aerogenes*. ISKpn*1* is present in fewer than 10 copies on the chromosomes of these strains, in contrast to IS*1397*, which was present in high copy numbers in certain strains, such as EPEC25 (at least 25 times) and ECOR49 (about 15 times) (5). ISKpn*1* is absent from *K. pneumoniae* subsp. *rhinoscleromatis* and in *K. oxytoca*. Since *Hin*dIII cuts into the IS (Fig. 6), the smallest fragment

K. oxytoca PU consensus[a]	TC- T TG G AC GA - G C CCCGG GGCGG CTG CGC CT CCGGG CTAC ctc a -- a -- tt g C			
Clones	**insertion site[b]**	**Region structure[c]**		
Within PUs				
Ko 2.1 & 8.1	TCG GCCGGATGGCGG CTGAG GCCTTATCCGGC CGACA	*orf*→	PU'(⇒)'PU→	***
Ko 3.1	TCG GCCGGATGGCGG CTGAG GCCTTATCCGGC CGACA	*orf*→	PU'(⇐)'PU→	***
Ko 4.1 & 11.1	AAA GCCGGATGGCGG CT AGC GCCTCATCCGGC CTACA	*yjiX*←	PU'(⇒)'PU→ h ←PU	←*yjiY*
Ko 5.2 & 12.2	AAA GCCGGATGGCGG CT AGC GCCTCATCCGGC CTACA	*yjiX*←	PU'(⇐)'PU→ h ←PU	←*yjiY*
Ko 6.1, 7.2 & 10.2	AGC GCCGGATAGCGG CTTAAC ACCTTATCCGGC CTACA	*fucR*→	PU'(⇒)'PU→	←*ygdE*
Ko 7.1	AGC GCCGGATAGCGG CTTAAC ACCTTATCCGGC CTACA	*fucR*→	PU'(⇐)'PU→	←*ygdE*
Ko 12.1	TTA GC GGATGGCGG CTGA C GCCTTATCCGGC AAACG	*narU*→	PU'(⇒)'PU→	→*narZ*
Outside PUs				
Ko 1.2	AGT GCCTGATGGGCGGCATGCTGGCCCACGCGCTGGTCTAT	*bglA*→	(⇒)	→*bglA*
Ko 2.2	ACC GCCGGATCGCGG CGTACGGTATCCGGCGTGCCGCT	*livG*←	(⇒)	←*livG*
Ko 5.1	GAC GCCGGATACCTGC GCGCCAATATACGGCCCCAGCGG	*tauD*←	(⇒)	←*tauD*
Insertion consensus	WVV GCCGGATGGCGG CTKARC RCCTYATCCGGC CTACA			

FIG. 5. Sites of insertion of the transposable module into the chromosome of *K. oxytoca*. *a*, The PU consensus is taken from reference 2. Lowercase letters are indicative of less frequent occurrences. *b*, Nucleotides which are duplicated after transposition are underlined. K is G or T; R is A or G; V is A, C, or G; W is A or T; and Y is C or T. *c*, Single arrows indicate PU and gene orientations. The double arrow represents the orientation of the transposable module. The letters h and t represent head internal sequences (HIS) and tail internal sequences (TIS), respectively (Fig. 2). Gene names have been assessed by sequence homology to *E. coli* genes (http://genolist.pasteur.fr/Colibri/). Triple asterisks indicate a region for which no homolog has been identified. *orf* indicates a potential coding sequence without homolog in the databases.

which can hybridize to the probe is 976 bp long and the last 469 bp of the IS do not match the probe. In *K. pneumoniae* subsp. *ozonae* and *K. pneumoniae* subsp. *pneumoniae*, we observed one fragment between 992 and 1,164 bp, indicating the presence of a *Hin*dIII or *Mlu*I restriction site close to the ISKpn*1* IRL in these fragments. These bands are heavily labeled. Such a phenomenon has already been described in the case of IS*1397* in several natural *E. coli* isolates and has been explained by a possible overrepresentation of IS-containing fragments identical or similar in size (5).

DISCUSSION

pNABI and pBLOCK harbor a transposable module containing a selectable marker (a Km^r-encoding gene) flanked by IS*1397* inverted repeats. The two plasmids enabled us to select for transposition of these modules into the chromosomes of four enterobacteria: *E. coli*, *S. enterica* serovar Typhimurium, *K. pneumoniae*, and *K. oxytoca*. In all instances, a 3- or 4-bp duplication at the site of insertion was characteristic of a bona fide transposition event. The difference between pBLOCK and pNABI is that the OrfAB transposase is, in the first case, expressed from an engineered gene which is located outside of the transposable module (Fig. 2), whereas the wild-type IS*1397* *orfA-orfB* genes are located within the module in pNABI. Both plasmids were efficient donor plamids for transposition in *E. coli* (not shown here). This shows that low expression of OrfAB in *cis* (pNABI) or its overexpression in *trans* (pBLOCK) can promote transposition. The same conclusion had been formulated in the case of IS*903* (10). Another interesting feature is that OrfA is expressed from both pBLOCK and pNABI. The expression of OrfA has been shown to prevent transposition in the case of IS*3* (34) or IS*1* (23, 44). On the contrary, coexpression of OrfA and OrfAB increases transposition in the case of IS*911* (39). As already mentioned (9), since we cannot accurately calculate transposition rates from our results, we do not know whether the expression of IS*1397* OrfA actually reduced or increased transposition efficiency. We can only conclude that the phenomenon was not abolished. This point is currently being checked in more detail.

As previously described in the case of *E. coli* (9), we found a strong target specificity for IS*1397* transposition in the three other species examined, since a large majority (80%) of the transposition sites were identified as PUs. PUs have been precisely described and classified in *E. coli* (3). The situation is not that clear and less well documented in the other species. Nonetheless, PU consensuses have been proposed for *S. enterica* serovar Typhimurium, *K. pneumoniae*, and *K. oxytoca* (2, 15). If PU sequences look very much conserved, clear differences between species can be observed and can be considered signatures of these different organisms. For instance, four PU types (Y, Z^1, Z^2, and D) can be distinguished in *S. enterica* serovar Typhimurium, whereas only three (Y, Z^1, and Z^2) can be found in *E. coli*. PUs in *K. pneumoniae* and *K. oxytoca* are more distantly related. We investigated *K. pneumoniae* sequences present in the contigs currently released by the GSC (http://genome.wustl.edu/gsc/index.shtml) by using the *E. coli* PU consensus and the consensus derived from the transposition sites of IS*1397* in *K. pneumoniae* as query sequences for the BLAST (http://www.ncbi.nlm.nih.gov/) and FASTA

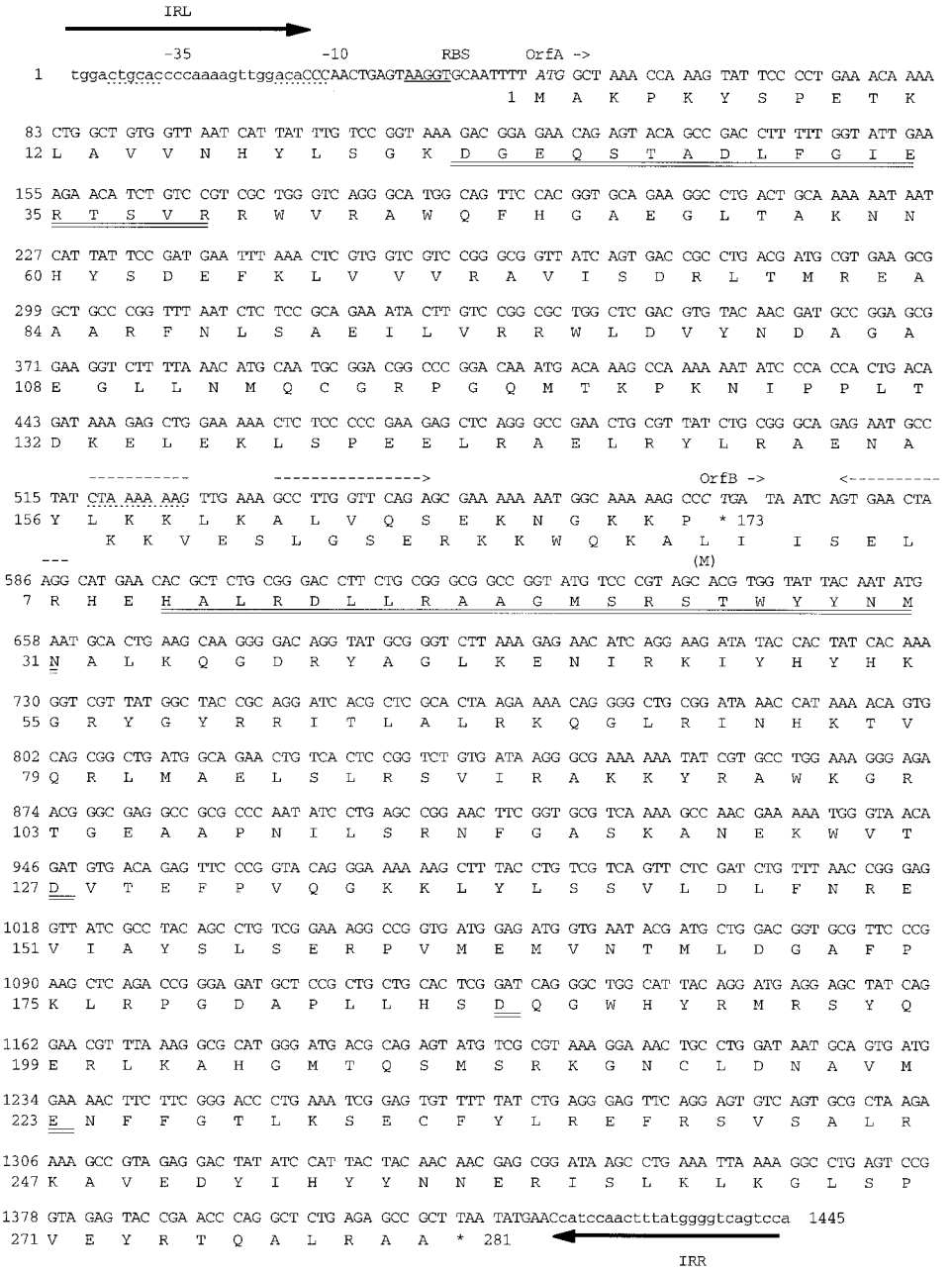


FIG. 6. Nucleotide and amino acid sequences of the two putative ORFs of ISKpn*1*. The IRL and IRR are in lowercase and indicated by black arrows. The ribosome-binding site (RBS) is underlined. The initiation codons of the two ORFs are in italics, and their termination codons are indicated by stars. Nucleotide and amino acid numbers are indicated on the left. The two HTH motifs are doubly underlined. The frameshift window is overlined and followed by a palindrome indicated by dashed arrows. The DDE motif (OrfB residues 127, 187, and 223) is doubly underlined.

(http://www.infobiogen.fr/) programs. The emerging consensus (Fig. 3 and 8) is actually rather different from *E. coli. K. oxytoca* PUs are less well documented due to the poor representation of sequences in currently available data banks. However, we could identify a general PU consensus for this species. This consensus is different from those of *E. coli, S. enterica* serovar Typhimurium, and even *K. pneumoniae*. A striking peculiarity of IS*1397* insertion sites in the *Klebsiella* chromosome is that the target sequences are much closer to the consensus found in *E. coli* than to the general consensus found in these species. In particular, the sequence GCCGGATG is most commonly found upstream of the insertion point in eight cases) and GCCGGATA is found less often in two cases). Both sequences are found in the first part of the stem of the *E. coli* and *Salmonella* Y PU. It should also be mentioned that the equivalent region of the *E. coli* Z PU (GCCTGATG) has been found only once as a transposition target in *K. oxytoca* (Ko 1.2 in a non-PU site; Fig. 5), whereas it was the target in 9 out of 29 cases in *E. coli* (9). The relatively higher GC content of the *Klebsiella* genomes could explain this bias. However, this could not hold for *S. enterica* serovar Typhimurium, where we, interestingly, found five cases of insertions into Z PUs all containing

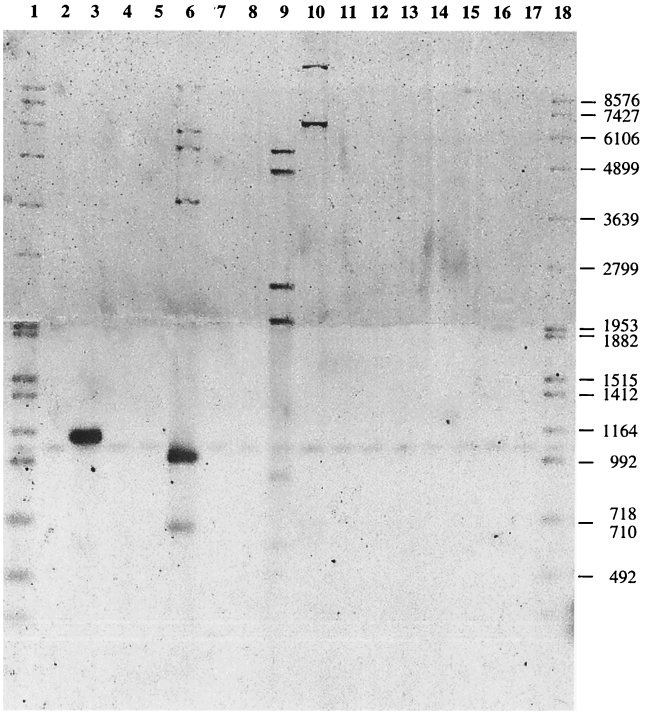


FIG. 7. Distribution of ISKpn*1*. Genomic DNAs were digested by *Hin*dIII and *Mlu*I and loaded on to a 1% agarose gel for Southern blot hybridization using an ISKpn*1* fragment as a probe. Lanes: 2, *K. planticola*; 3, *K. pneumoniae* subsp. *ozonae*; 4, *K. oxytoca*; 5, *K. pneumoniae* subsp. *rhinoscleromatis*; 6, *K. pneumoniae* subsp. *pneumoniae*; 7, *E. aerogenes*; 8, *K. terrigena*; 9, *K. pneumoniae* MGH78578; 10, *K. aerogenes*; 11, *E. coli* EPEC 25; 12, *E. coli* C600; 13, *Y. pestis*; 14, *Y. pseudotuberculosis*; 15, *Y. enterocolitica*; 1 and 18, digoxigenin-labeled DNA molecular weight marker VII (50 ng); 16, pNABI digested by *Eco*RV (100 ng); 17, pNABI digested by *Eco*RV (10 ng). The values on the right are molecular weights.

GCCGGATG and never GCCTGATG, which was expected with equal probability. However, the small number of cases does not allow us to draw firm conclusions on this point. Interestingly, even in the few "aberrant" *Klebsiella* transpositions (i.e., not located in a PU, sometimes within an ORF; Fig. 3 and 5), GCCGGATG (exactly or with slight variations) is found upstream of the insertion point. The exact recognition site of IS*1397* could thus be restricted to GCCGGATR or GCCTG ATR. In *Klebsiella*, these sequences are not really expected to be found more frequently in PUs than in the rest of the chromosome since they are not part of the consensus. However, they are clearly preferred as transposition sites when they belong to a PU (62 to 81% of the cases), almost as frequently as in *E. coli* (9) or *S. enterica* serovar Typhimurium (86 and 100% of the cases, respectively).

From our results, we can conclude that IS*1397* is specific for *E. coli* PUs. These PUs can be found frequently in *S. enterica* serovar Typhimurium and less frequently in *Klebsiella* and can be efficient targets for IS*1397* in these cases. It will be interesting to study transposition in an enterobacterium which does not contain PUs but is phylogenetically close to *E. coli*. This is why we undertook the same type of work with *Yersinia*.

One relevant observation is that transposition hot spots were found. In *K. oxytoca*, we found an unidentified region three times (Ko 2.1, Ko 3.1, and Ko 8.1) and *fucR-ygdE* four times (Ko 6.1, Ko 7.1, Ko 7.2, and Ko 10.2). In *K. pneumoniae*, the region located after *yjgA* was found twice (Kp 6.1 and Kp 8.2). Such a phenomenon has already been observed in *E. coli* (9). It was more surprising that hot spots were also found between species, despite the lack of similitude between these intergenic regions. *narU-narZ* was found twice in *K. pneumoniae* (Kp 12.1 and Kp D) and once in *K. oxytoca* (Ko 12.1); *yjiX-yjiY* was found twice in *S. enterica* serovar Typhimurium (St 2.2 and St 6.1), four times in *K. oxytoca* (Ko 4.1, Ko 5.2, Ko 11.1, and Ko 12.2), and once in *K. pneumoniae* (Kp E). This suggests that mere sequence recognition by IS*1397* transposase is not the only factor which determines transposition. Some local chromosomal features must exist, possibly shared by the three species investigated in this study, and these features are probably not related to PUs. If one considers the case of *yjiX-yjiY*, this intergenic region contains a solitary D PU in *S. enterica* serovar Typhimurium and is composed of two convergent PUs in *K. oxytoca* and *K. pneumoniae* that are similarly organized but

K. pneumoniae	**PU consensus**	TTT	SCCGGGTGGCGGC	TGA	CGCYTTACCSGGC	CTACR
	IS*1397* insertion consensus	WNY	GCCGGATGGCGGC	WWA	YGCCTTATCCGGC	CTACS
K. oxytoca	**PU consensus**	CYY	CCCGGW GGCGGT	GCTGRCG	CACCTKW CCGGG	CTACS
	IS*1397* insertion consensus	WVV	GCCGGATGGCGGC	TKAR	CRCCTYATCCGGC	CTACA
S. typhimurium	**PU consensus**	WWY	GCCKGATGGCGRCGC	TDAM	GCGCCTTATCMGGC	CTACR
	IS*1397* insertion consensus	HWW	GCCGGATRGCGRC	TWA	CGCCTTATCMGGC	CTACV
E. coli	**PU consensus**	HNN	GCCGGATRGCGRC	TDAR	YRCCTTATCMGGC	CTACV
	IS*1397* insertion consensus	AWT	GCCKGATGRCGC	TDMR	CGYCTTATCMGGC	CTACR

FIG. 8. Summary of PU consensuses and the transposable module insertion target consensus in *Enterobacteriaceae*. Data are taken from Fig. 3 to 5. The *E. coli* PU consensus is taken from reference 4. D is A, G, or T; H is A, C, or T; K is G or T; M is A or C; N is A, C, G, or T; S is C or G; R is A or G; V is A, C, or G; W is A or T; and Y is C or T.

K. pneumoniae PU consensus[a]	TTT SCCGGGTGGCGG CTGAC GCYTTAC CSGGC CTACR		
K. pneumoniae contig/clone[b]	ISKpn*1* insertion site[c]	Region structure[d]	
654-840	ATT GCCCGGCGGCG CTGC GC TTGC CGGGC CTACG	*ygfF*→	PU'(⇐)'PU→ h ←PU ***
642-654	CTGC GC TTGCACGGGC CTACG	*proK*→	?(⇐)'PU→ h ←PU t PU→ h ←PU t PU→ h ←PU t PU→ ***
230-642	ATT GCCCGGCGGCG CTGC GC TTGC CGGGC CTACG	*dld*→	PU→ h ←PU t PU'(⇐)'PU→ h ←PU t PU→ h ←PU ←*pbpG*
823-230	ATC GCCCGGCGGCG CTGC GC TTGCACGGGC CTACG	*pheV*→	PU→ h ←PU t PU'(⇐)'PU→ h ←PU t PU→ h ←PU t PU→ h ←PU ***
840-823	GAT GCCCGGCGGCG CTGC GC TTGCGCGGGC CTACG	*dppA*←	PU'(⇐)'PU→ ←*proK*
Kp C	ATC GCCCGGCGGCG CTGC	***	←PU t PU→ h ←PU t PU'(⇒)'PU→ ***
Insertion consensus	RWY GCCCGGCGGCG CCTGC GC TTGCRCGGGC CTACG		

FIG. 9. Sites of insertion of ISKpn*1* into the chromosome of *K. pneumoniae. a*, The PU consensus is derived from this study. *b*, The *K. pneumoniae* strain MGH78578 contig numbers were found at the GSC web site in May 2000 (http://genome.wustl.edu/gsc/index.shtml). The contigs have been split and reassembled by PCR analysis and BIME sequence comparison (see Discussion). In the case of 642–654, the beginning of the sequence is not available at the GSC web site. Clone Kp C is another example we sequenced after transposition of IS*1397* in *K. pneumoniae* strain ATCC 13883 (cf. Fig. 3). *c*, Nucleotides which are duplicated after transposition are underlined. R is A or G, W is A or T, and Y is C or T. *d*, Single arrows indicate PU and gene orientations. Double arrows represent the orientation of ISKpn*1*; h and t represent head internal sequences (HIS) and tail internal sequences (TIS), respectively (Fig. 1). Gene names have been assessed by sequence homology to *E. coli* genes (http://genolist.pasteur.fr/Colibri/). Triple asterisks indicate a region for which no homolog has been identified.

differ in sequence. The nature of the potential for "attracting" IS*1397* in these regions remains totally unknown.

In this study, we identified a new insertion sequence in the *K. pneumoniae* MGH78578 genome (sequenced in St. Louis), named ISKpn*1*. This sequence was always found inserted into *K. pneumoniae* PUs (see below). We detected five copies of the IS in the contigs available in the database (May 2000). However, the flanking genes of the copy of ISKpn*1* we cloned did not correspond to any of these contigs, suggesting misassemblies. We therefore checked the structure of IS flanking regions by PCR analysis. We tested each possible combination of primers and obtained fragments for only five combinations (Fig. 9). We also checked the sequences of flanking PUs and extra-PU motifs, which are extremely well conserved in each BIME. We then concluded that there were actually five copies of the IS with the same 3-bp duplication, TGC. Another example (with a partial sequence) was found in *K. pneumoniae* strain ATCC 13883, in which we studied IS*1397* transposition. Indeed, IS*1397* had been transposed into a BIME which already contained ISKpn*1* (clone C; Fig. 3). Southern blot hybridization showed that ISKpn*1* is only present in *K. pneumoniae* subsp. *pneumoniae, K. pneumoniae* subsp. *ozonae*, and *K. aerogenes*. This indicates that it is restricted to a small subset of species closely related to *K. pneumoniae*. The same observations have been made for the distribution of IS*1397* in *E. coli* isolates (5).

Like IS*1397*, ISKpn*1* is inserted into the central part of PUs with a 3-bp duplication. ISKpn*1* is found in PUs which are specific to *K. pneumoniae* (Fig. 9) with a very high GC content and an additional C at position 7. On the contrary, as discussed above, IS*1397* has been systematically transposed into *K. pneumoniae* PUs, which are closer to the *E. coli* consensus. The case of clone C (Fig. 3) is particularly interesting since it deals with a BIME where typical *K. pneumoniae* PUs alternate with PUs which are close to the *E. coli* consensus and which were the targets for ISKpn*1* and IS*1397*, respectively.

ISKpn*1* is a new member of the IS*3* family, as shown by sequence comparisons. OrfB is known to contain the catalytic domain of the transposase, and OrfA has been demonstrated to recognize the inverted repeats of the IS specifically (see reference 24 for a review). We think that OrfA may confer the specificity of insertion. Figure 10 shows an alignment of the

FIG. 10. Alignment of OrfA proteins from IS*150*, IS*1397*, and ISKpn*1*. IS*150*, ISKpn*1*, and IS*1397* were aligned by using the DNA strider alignment program (BLOCKS). Residues found to be common to all three ISs are boxed in black, and residues found to be common to of the two ISs are boxed in grey. Regions of homology among the three proteins are framed with solid lines, and regions of homology between IS*150* and ISKpn*1* or ISKpn*1* and IS*1397* are framed with dotted lines.

OrfA proteins from ISKpn*1*, IS*150*, and IS*1397*. ISKpn*1* is closer in both its nucleotide and protein sequences to IS*150* than to IS*1397*, although ISKpn*1* and IS*1397* are both inserted into PUs but not IS*150*. The C termini of the OrfA proteins from IS*1397* and ISKpn*1* share a conserved 13-aa motif, ELR YLRAENAYLK, which is not found in IS*150*. Thus, we can speculate that these amino acids play a role in the selection of PUs as transposition targets.

ACKNOWLEDGMENTS

We thank A. Cahen and F. Le Noanne for technical assistance. We thank the GSC, Washington University, St. Louis, Mo., for *K. pneumoniae* strain MGH78578 and for communication of DNA sequence data prior to publication. We are also thankful to P. A. D. Grimont for the other *Klebsiella* strains and the *Salmonella* strains and to E. Carniel for the *Yersinia* strains.

This work was supported by a grant from the Ministère Français de l'Education Nationale, de la Recherche et de la Technologie.

REFERENCES

1. **Appleyard, R. K.** 1954. Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *E. coli* K12. Genetics **39:**440–452.
2. **Bachellier, S., D. Perrin, M. Hofnung, and E. Gilson.** 1993. Bacterial interspersed mosaic elements (BIMEs) are present in the genome of *Klebsiella*. Mol. Microbiol. **7:**537–544.
3. **Bachellier, S., W. Saurin, D. Perrin, M. Hofnung, and E. Gilson.** 1994. Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). Mol. Microbiol. **12:**61–70.
4. **Bachellier, S., E. Gilson, M. Hofnung, and C. W. Hill.** 1996. Repeated sequences, p. 2012-2040. *In* F. C. Neidhardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Shaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
5. **Bachellier, S., J.-M. Clément, M. Hofnung, and E. Gilson.** 1997. Bacterial interspersed mosaic elements (BIME) are a major source of sequence polymorphism in *Escherichia coli* intergenic regions including specific associations with a new insertion sequence. Genetics **145:**551–562.
6. **Bachellier, S., J.-M. Clément, and M. Hofnung.** 1999. Short palindromic repetitive DNA elements in enterobacteria: a survey. Res. Microbiol. **150:**627–639.
7. **Boccard, F., and P. Prentki.** 1993. Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3′ end of transcription units. EMBO J. **12:**5019–5027.
8. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:**1141–1156.
9. **Clément, J.-M., C. Wilde, S. Bachellier, P. Lambert, and M. Hofnung.** 1999. IS*1397* is active for transposition into the chromosome of *Escherichia coli* K-12 and inserts specifically into palindromic units of bacterial interspersed mosaic elements. J. Bacteriol. **181:**6929–6936.
10. **Derbyshire, K. M., and N. D. Grindley.** 1996. *cis* preference of the IS*903* transposase is mediated by a combination of tranposase instability and inefficient translation. Mol. Microbiol. **21:**1261–1272.
11. **Dodd, I. B., and J. B. Egan.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:**5019–5026.
12. **Espéli, O., and F. Boccard.** 1997. In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site. Mol. Microbiol. **26:**767–777.
13. **Gilson, E., J.-M. Clément, D. Brutlag, and M. Hofnung.** 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. **3:**1417–1421.
14. **Gilson, E., J. P. Rousset, J.-M. Clément, and M. Hofnung.** 1986. A subfamily of *E. coli* palindromic units implicated in transcription termination? Ann. Inst. Pasteur/Microbiol. (Paris) **137:**259–270.
15. **Gilson, E., J.-M. Clément, D. Perrin, and M. Hofnung.** 1987. Palindromic units: a case of highly repetitive DNA sequences in bacteria. Trends Genet. **3:**226–230.
16. **Gilson, E., D. Perrin, and M. Hofnung.** 1990. DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic units highly repetitive DNA: implications for bacterial chromosome organization. Nucleic Acids Res. **18:**3941–3952.
17. **Gilson, E., S. Bachellier, S. Perrin, D. Perrin, P. A. D. Grimont, F. Grimont, and M. Hofnung.** 1990. Palindromic unit highly repetitive DNA sequences exhibit species specificity within *Enterobacteriaceae*. Res. Microbiol. **141:**1103–1116.
18. **Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung.** 1991. Palindromic units are part of a new bacterial interspersed mosaic element (BIME). Nucleic Acids Res. **19:**1375–1383.
19. **Higgins, C. F., G. Ferro-Luzzi Ames, W. M. Barnes, J.-M. Clément, and M. Hofnung.** 1982. A novel intercistronic regulatory element of prokaryotic operons. Nature **298:**760–762.
20. **Katsuragi, N., N. Takizawa, and Y. Murooka.** 1987. Entire nucleotide sequence of the pullulanase gene of *Klebsiella aerogenes* W70. J. Bacteriol. **169:**2301–2306.
21. **Kulkosky, F., K. S. Jones, R. A. Katz, J. P. 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.
22. **Lim, D., T. A. T. Gomes, and W. K. Maas.** 1990. Distribution of msDNA among serotypes of enteropathogenic *Escherichia coli* strains. Mol. Microbiol. **4:**1711–1714.
23. **Machida, C., and Y. Machida.** 1989. Regulation of IS*1* transposition by the *insA* gene product. J. Mol. Biol. **208:**567–574.
24. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
25. **Metcalf, W. W., W. Jiang, and B. L. Wanner.** 1994. Use of the rep technique for allele replacement to construct new *Escherichia coli* hosts for maintenance of R6K gamma origin plasmids at different copy numbers. Gene **138:**1–7.
26. **Newbury, S. F., N. H. Smith, and C. F. Higgins.** 1987. Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell **51:**1131–1143.
27. **Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgins.** 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48:**297–310.
28. **Oppenheim, A. B., K. E. Rudd, I. Mendelson, and D. Teff.** 1993. Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*. Mol. Microbiol. **10:**113–122.
29. **Polard, P., and M. Chandler.** 1995. Bacterial transposases and retroviral integrases. Mol. Microbiol. **15:**13–23.
30. **Sambrook, J., E. F. Frisch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
31. **Sekine, Y., and E. Ohtsubo.** 1992. DNA sequences required for translational frameshifting in production of the transposase encoded by IS*1*. Mol. Gen. Genet. **235:**325–332.
32. **Sekine, Y., H. Nagasawa, and E. Ohtsubo.** 1992. Identification of the site of translational frameshifting required for production of the transposase encoded by insertion sequence IS*1*. Mol. Gen. Genet. **235:**317–324.
33. **Sekine, Y., N. Eisaki, and E. Ohtsubo.** 1994. Translational control in production of transposase and in transposition of insertion sequence IS*3*. J. Mol. Biol. **235:**1406–1420.
34. **Sekine, Y., K. Izumi, T. Mizuno, and E. Ohtsubo.** 1997. Inhibition of transpositional recombination by OrfA and OrfB proteins encoded by insertion sequence IS*3*. Genes Cells **2:**547–557.
35. **Shafferman, A., R. Kolter, D. Stalker, and D. R. Helinski.** 1982. Plasmid R6K DNA replication. III. Regulatory properties of the pi initiation protein. J. Mol. Biol. **161:**57–76.
36. **Shyamala, V., E. Schneider, and G. Ferro-Luzzi Ames.** 1990. Tandem chromosomal duplications: role of REP sequences in the recombination event at the joinpoint. EMBO J. **9:**939–946.
37. **Stalker, D. M., R. Kolter, and D. R. Helinski.** 1982. Plasmid R6K DNA replication. I. Complete nucleotide sequence of an autonomously replicating segment. J. Mol. Biol. **161:**33–43.
38. **Stern, M. J., E. Prossnitz, and G. Ferro-Luzzi Ames.** 1988. Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *Salmonella typhimurium*: involvement of REP sequences. Mol. Microbiol. **2:**141–152.
39. **Ton-Hoang, B., L. Polard, C. Haren, C. Turlan, and M. Chandler.** 1999. IS*911* transposon circles give rise to linear forms that undergo integration in vitro. Mol. Microbiol. **32:**617–627.
40. **Varmus, H., and P. Brown.** 1989. Retrovirus, p. 55–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
41. **Vögele, K., E. Schwartz, C. Welz, E. Schiltz, and B. Rak.** 1991. High-level ribosomal frameshifting directs the synthesis of IS*150* gene products. Nucleic Acids Res. **19:**4377–4385.
42. **Yang, Y., and G. Ferro-Luzzi Ames.** 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. Proc. Natl. Acad. Sci. USA **85:**8850–8854.
43. **Yanish-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
44. **Zerbid, D., P. Polard, J. M. Escoubas, D. Galas, and M. Chandler.** 1990. The regulatory role of the IS*1*-encoded InsA protein in transposition. Mol. Microbiol. **4:**471–477.