Mol Genet Genomics (2001) 266: 216–222
DOI 10.1007/s004380100542

ORIGINAL PAPER

A. Kovařík · M.A. Matzke · A.J.M. Matzke
B. Koukalová

Transposition of IS*10* from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts

Received: 9 January 2001 / Accepted: 5 June 2001 / Published online: 2 August 2001

Abstract During recloning of *Nicotiana tabacum* L. repetitive sequence R8.3 in *Escherichia coli*, a modified clone that differed from the original by the insertion of an IS*10* sequence was unintentionally produced. The insert was flanked by a 9-bp direct repeat derived from the R8.3 sequence, the 9-bp duplication of acceptor DNA in the site of insertion being a characteristic of IS*10* transposition events. A database search using the FASTA program showed IS*10* and other prokaryotic IS elements inserted into numerous eukaryotic clones. Unexpectedly, the IS*10*, which is not a natural component of the *E. coli* genome, appeared to be by far the most frequent contaminant of DNA databases among several IS sequences tested. In the GenEMBL database, the IS*10* query sequence yielded positive scores with more than 500 eukaryotic clones. Insertions of shortened IS*10* sequences having only one intact terminal inverted repeat were commonly found. Most full-length IS*10* insertions (32 out of 40 analyzed) were flanked by 9-bp direct repeats having the consensus 5′-NPuCNNNGPyN-3′ with a strong preference for 5′-TGCTNAGNN-3′. One insertion was flanked by an inverted repeat of more than 400 bp in length. PCR amplification and Southern analysis revealed the presence of IS*10* sequences in *E. coli* strains commonly used for DNA cloning, including some reported to be Tn*10*-free. No IS*10*-specific PCR product was obtained with *N. tabacum* or human DNA. Our data suggest that transposition of IS*10* elements may accompany cloning steps, particularly into large BAC vectors. This might lead to the relatively frequent contamination of DNA databases by this bacterial sequence. It is estimated that one in approximately every thousand eukaryotic clone in the databases is contaminated by IS-derived sequences. We recommend checking submitted sequences for the presence of IS*10* and other IS elements. In addition, DNA databases should be corrected by removing contaminating IS sequences.

Keywords IS*10* transposition · IS elements · DNA cloning · DNA databases · *Escherichia coli*

Introduction

Transposon Tn*10* belongs to the family of composite mobile genetic elements in bacteria (Kleckner 1981). Based on its nucleotide sequence, the total length of Tn*10* is 9147 bp. The flanking inverted repeats, IS*10*-Left and IS*10*-Right (hereafter referred to as IS*10*L and IS*10*R), are each 1329 bp long (Chalmers et al. 2000, GenBank Accession No. AF162223). Tn*10* contains genes for resistance to tetracycline and for transposase, which enables propagation within the genome (Foster et al. 1981). Three additional ORFs have recently been described (Chalmers et al. 2000).

The mobile element of Tn*10* is IS*10*R, which contains the gene for transposase, thus enabling the movement of IS*10*R as an independent entity. IS*10*L is largely defective for this gene (Foster et al. 1981). The distribution of Tn*10*/IS*10* in prokaryotic genomes and plasmids has been studied previously (Matsutani 1991; Chalmers et al. 2000). Solitary IS*10* elements were found to be more frequent than the composite Tn*10*. There are no reports of the presence of Tn*10*/IS*10* sequences in eukaryotic genomes. Tn*10*/IS*10* transposition is accompanied by a 9-bp duplication of acceptor DNA at the site of insertion (Halling and Kleckner 1982; Bender and Kleckner 1992). Another consequence of Tn*10*/IS*10* movement is

---

Communicated by J. Schell

A. Kovařík · B. Koukalová (✉)
Institute of Biophysics,
Academy of Sciences of the Czech Republic,
Královopolská 135, 612 65 Brno, Czech Republic
E-mail: blazena@ibp.cz
Tel.: +420-5-41517199
Fax: +420-5-41211293

M.A. Matzke · A.J.M. Matzke
Institute of Molecular Biology,
Austrian Academy of Sciences,
[2]Billrothstrasse 11, A-5020, Salzburg, Austria

rearrangement of the acceptor genome, including deletions, inversions and other mutations (Raleigh and Kleckner 1984; Bogosian et al. 1993, Chalmers and Kleckner 1996). The frequency of Tn*10*/IS*10* transposition is rather low (10^{-7} per cell per generation according to Chalmers and Blot 1999) and is known to be controlled by several mechanisms (Simons and Kleckner 1983; Roberts et al. 1985). Transposition activity may be increased under certain circumstances, e.g. during further incubation of cells in the stationary phase of growth (Skaliter et al. 1992; Naas et al. 1995) or after UV irradiation (Eichenbaum and Livneh 1998).

Here we describe an IS*10* transposition event that occurred incidentally during subcloning of a recombinant plasmid (pUC19::R8.3) carrying a repetitive sequence that originated from *Nicotiana tabacum* L. (Kuhrová et al. 1991). DNA database searches were carried out to investigate whether similar cases of undesired transposition events have occurred during cloning procedures in other laboratories. We found that many published eukaryotic sequences indeed contain IS*10* and also other prokaryotic IS elements.

Materials and methods

Escherichia coli strains

The following *E. coli* K-12 strains were used: JM109, DH1, DH5α, DH10B and XL2-Blue. Cells were cultivated in L-medium.

Isolation of DNA and restriction enzyme digestions

Bacterial DNA was isolated from overnight cultures according to Schleif and Wensink (1981). *Nicotiana tabacum* L. cv. Vielblättriger was used for isolation of tobacco DNA. Total DNA was isolated from leaves using a modified CTAB (cetyltrimethyl ammoniumbromide) method (Kovařík et al. 1997). Human DNA was isolated from blood according to the protocol of Sambrook et al. (1989).

Purified DNA was digested according to the recommendation of the enzyme suppliers. In the case of tobacco DNA an excess of enzyme (5 U/µg DNA) was used.

Southern hybridization

Digested DNA was subjected to electrophoresis on agarose gels. Following electrophoresis, the ethidium bromide-stained gel was photographed and blotted onto a membrane (Hybond N+, Amersham Pharmacia Biotech). To prepare probe about 50 ng of DNA (a 1285-bp PCR product; see below) was labeled with ^{32}P ($\geq 10^8$ dpm/µg DNA, Fermentas, Dekaprime kit). Southern hybridization was carried out in 0.25 M sodium phosphate buffer (pH 7.0), supplemented with 7% SDS at 65°C for 16 h followed by washing at high stringency (twice with 2×SSC, 1% SDS, for 5 min each at room temperature, and twice with 0.2×SSC, 0.1% SDS for 30 min at 65°C) according to Sambrook et al. (1989). Hybridization bands were visualized using a PhosphorImager (Storm, Molecular Dynamics) and by exposure to X-ray film.

PCR

PCR amplification was performed with 50–100 ng of genomic DNA as template, in a reaction volume of 50 µl containing Taq buffer, $MgCl_2$ to a final concentration of 2.5 mM, each nucleotide at 0.4 mM, each primer at 0.5 µM, and 1.0 U of thermostable Taq DNA polymerase (Dynazyme). Reaction mixtures were overlaid with mineral oil in 200-µl tubes. The PCR was run on a MJ Research PTC200 under the following conditions: 3 min initial denaturation at 94°C, 30 cycles of 15 s at 93°C, 30 s at 57°C, 30 s at 72°C, followed by 10 min at 72°C. The IS*10*for (forward) and IS*10*rev (reverse) primers were derived from the published sequence of the IS*10*R element of the Tn*10* transposon (Chalmers et al. 2000). The first nucleotide of IS*10*for (5′-TAATTTCCCCAAAGCGTAAC-3′) corresponds to position 7835; the first nucleotide of IS*10*rev (5′-AAAATCATTAAGTTAAGGTGG-3′) corresponds to position 9119 (reading bottom strand). A map of the primer sites is shown in Fig. 1.

The 1285-bp PCR product was purified by gel electrophoresis and isolated using a gel extraction kit (Qiagen) and used as a probe.

Database searches and data processing

The GenEMBL database was searched for sequences homologous to the Tn*10* transposon and several IS elements (see below) using the FASTA program implemented in the Wisconsin GCG software package. The program allows analysis of sequences flanking regions of homology. Both strands of the query sequence were searched. The search parameters were as follows: gap creation penalty: 16; gap extension penalty: 4; scoring matrix: fastadna.cmp. The GenEMBL databank available at the local terminal contained about 1.2 million sequences (up to the 3 January 2001). The regions flanking the IS*10* insertions were analyzed manually for duplications and inversions. The sequence consensus was calculated from aligned sequences using the COMPARE function.

DNA sequence analysis

The nucleotide sequences of R8.3 (Accession No. AJ292266) and R8.3M (Accession No. AJ293237) were determined as described by Mette and coworkers (1999).

Results

Insertion of IS*10* into the cloned tobacco R8.3 sequence in the course of recloning

The R8.3 sequence isolated from the nuclear genome of *N. tabacum* L. was initially described as a 5.5-kb middle repetitive DNA sequence (Kuhrová et al. 1991). Originally, it was cloned into a pUC19 plasmid in the *E. coli* host strain JM109. After recloning of the recombinant plasmid pUC19::R8.3 into the *E. coli* strain DH5α, a modified plasmid designated as pUC19::R8.3M was

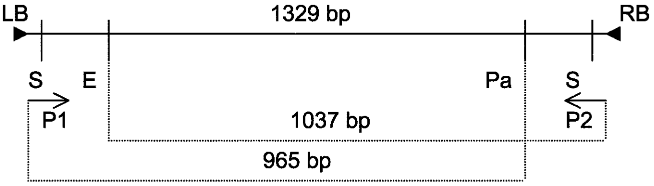

Fig. 1 Schematic representation of the IS*10*R element. The left and right borders with short inverted repeats (*filled arrowheads*) are depicted with respect to the position of IS*10*R in the Tn*10* transposon. The positions of PCR primers P1 and P2 are indicated by the *arrows*. The length of the amplicon is 1285 bp. Restriction enzyme sites: S, *Sau*3AI; E, *Eco*RV; Pa, *Pae*I

obtained. The modification involved an approximately 1.3-kb insertion of non-R8.3 sequence close to the 5′ end (Fig. 2). Sequencing of R8.3M (Accession No. AJ293237) showed that R8.3M contained the IS*10* sequence. The degree of homology between the IS*10* insertion and the published IS*10*R (Accession No. J01829) and IS*10*L (nucleotides 1–1328 in AF162223) sequences was 99.3% and 98.2%, respectively. Ten mismatches were found within the region of homology with IS*10*R. The transposase reading frame was interrupted at several positions, suggesting that the IS*10* inserted into the R8.3M clone cannot code for a functional protein. A 9-bp duplication (5′-GACTAGGCG-3′) of the R8.3 acceptor sequence was generated at the insertion site, which is a characteristic feature of IS*10*/Tn*10* transposition. With the exception of the 9-bp duplication, no major rearrangements of acceptor sequences had occurred.

IS*10* is a rather frequent component of the cloned sequences registered in DNA databases

On searching the GenEMBL database for homologies to the Tn*10*/IS*10* sequence, numerous entries were identified. Surprisingly, a considerable number of hits were in eukaryotic sequences (Table 1). The homology was almost exclusively found within the 1329-bp IS*10* element, except in the case of the clone AC006650, which also included a sequence from the central part of Tn*10*. The level of homology between IS*10* and sequences in the database was high. More than 100 clones yielded 100% homology with IS*10*R (see below).

The 40 randomly selected full-length clones were further analyzed in more detail with respect to homology and integration site (Table 1). Although most clones contained a single insertion, one clone (Accession No. AC009017) contained two IS*10* insertions separated by 1793 bp of a non-Tn*10* sequence. One of these insertions was defective in the CTGA inverted repeat at the end of IS*10*. A 9-bp duplication was found in the flanking sequences of the majority of insertions. Only seven out of 40 insertions did not contain a perfect direct repeat. More than 60% of the 9-bp duplications matched the consensus 5′-TGCTNAGNN-3′. The frequency of occurrence of different bases within the 9-bp duplications is shown in Table 2. In the clone CTC-295J22 (AC011347) a relatively long (~ 400 bp) inverted repeat was found at the site of IS*10* insertion.

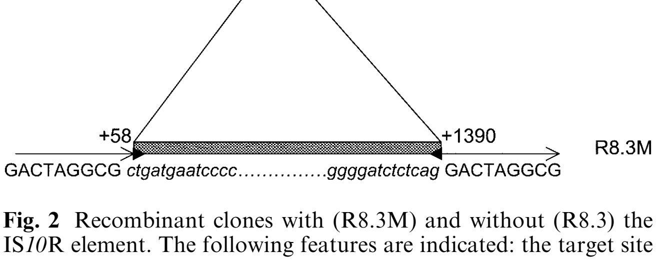

Fig. 2 Recombinant clones with (R8.3M) and without (R8.3) the IS*10*R element. The following features are indicated: the target site (GACTAGGCG), and its duplication (*open arrows*) after the insertion of IS*10*R (*filled box*), the IS*10*R sequence is shown in *italics*, the ends of the inverted terminal repeats are marked by *filled arrows*. The coordinates of the IS*10*R insertion site are in accordance with the R8.3M sequence deposited in the database (AJ293237)

Search for IS*10* elements in *E. coli* strains and eukaryotic genomes

Data presented in a previous section demonstrated the relatively frequent occurrence of the IS*10* element in cloned DNA sequences. We were therefore interested in the distribution of IS*10* sequence in *E. coli* strains, human and plant DNA. Prior to these experiments, the genotypes were tested for tetracycline resistance/sensitivity. As expected, only XL-2 Blue cells grew on the L-broth supplemented with 10 μg/ml tetracycline. To detect IS*10* elements in genomic DNAs, we employed PCR using specific primers (see Materials and methods). The primers were designed to amplify a 1285-bp region spanning nearly the whole IS*10*R sequence (Fig. 1). A PCR product of the expected length was obtained with all the bacterial DNAs tested (XL2-Blue, DH5α, JM109) but not with human or tobacco DNAs (Fig. 3). The specificity of PCR amplification was further confirmed by digestion of the ~1.3-kb PCR product with *Eco*RV and *Pae*I. These enzymes cut IS*10*R at positions –1064 and –347 bp from the right border of IS*10*R, respectively (Fig. 1). In agreement with the sequence map, two bands of about 1.0 kb and 0.3 kb were obtained with each enzyme (Fig. 3).

To study the genomic organization of IS*10* elements in several *E. coli* strains, we performed Southern hybridizations using IS*10*R as a probe (Fig. 4). DNAs from JM109, DH1, DH5α, DH10B and XL-2 Blue, were digested with *Eco*RV and *Sau*3AI. *Sau*3AI cuts out a large (1253 bp) internal fragment from IS*10*R and IS*10*L. This fragment was visible in lanes loaded with *Sau*3AI-restricted DNA from JM109, DH5α, DH10B and XL-2 Blue but not from DH1. The 0.7-kb *Eco*RV band appeared to be common to all IS*10*-positive DNAs and is probably derived from the IS*10*L internal fragment – there are two *Eco*RV cleavage sites within the IS*10*L and one within the IS*10*R (Matsutani 1991). Therefore the remaining *Eco*RV fragments probably correspond to IS*10*R. It is evident that the number and sizes of the IS*10*R-specific *Eco*RV fragments varied significantly among the strains, indicating that the strains differ with respect to the genomic organization of IS*10*R elements. The *Eco*RV polymorphism seen among the different *E. coli* strains could reflect some rearrangement of IS*10* loci caused, for example, by cultivation conditions. Thus, it is known that maintenance of cells in stab cultures can influence the mobility of IS elements (Naas et al. 1995). No signals were obtained

Table 1 List of selected clones bearing IS*10* insertions

Clone	Source organism	Accession No.[a]	Position	9 bp duplication[b]
R8.3M	Tobacco	AJ293237	58–1390	GACTAGGCG
hRPK.203_M_10	Human	AC006473	77393–78721	TACCCTGCT
CTB-10G5	Human	AC007566	31597–32925	TGCTAAGTC
CTG-2193G5	Human	AC010423	71217–72545	TACAGTGTT
RP11-208N20	Human	AC069122	109403–110731	TACTTCATA
RP23-2M16	Mouse	AC016814	176945–178273	GGCAAAACA
RP23-101G12	Mouse	AC074328	102440–103768	GGCAAAGTT
BcDNA.LD23587	Drosophila	AF181652	2754–4082	CCCTCTAGG
OSJNBs0045I23	Rice	AC025825	43389–44717	GGCGCGGCG
IGF-F14C21	Arabidopsis	AC069144	774651–775982	TGCTGAGTC
FLJ10765	Human	AK001627	422–1750	TGCAGTGCT
COL004457	Human	AK025098	438–1766	TGCGAAGCC
RP11-2N10	Human	AK015666	82786–84144	TGCTTGGTC
RP11-774F2	Human	AP001451	6368–7696	CACAAGACC
CTC-295J22	Human	AC011347	143624–144952	Inverted repeat
R-97N10	Human	AL356800	146096–147431	TGCACAGCC
RP11-681O17	Human	AP001316	90383–91711	TGCTCTGTC
RP110G10	Human	AC053465	116026–117355	TGCTTTGTA
RP11-661O13	Human	AP001793	170236–171564	TGCAGAGCT
BSTsomatotrop.	Bovine	S67119	107–1435	TGCTCAGCA
RP11-770P7	Human	AP001106	147227–148556	TGCTAAGTA
RP11-796c24	Human	AP001846	92294–93622	TGCTGAGCC
DS01630	Drosophila	AC006938	4108–5436	–
359N5	Human	AL158800	149853–151181	TGCAGAGCT
CTC-431G16	Human	AC008496	7485–8813	TGCTAAGTG
CTB-43E15	Human	AC008674	129367–130696	TGCTCAGTA
Xxp1-929G6	Human	AC009017	115339–116665, 112218–113546[c]	–
RP11-354N11	Human	AC044801	86382–87710	TGCTCAGCA
RP11-526K21	Human	AC037457	23142–24470	–
RP11-770P7	Human	AP001896	119786–121114	TGCTCTGTG
XXPAC913I9	Human	AP000452	75086–76412	nd
f43D11-119B8	Human	AP000128	93958–95286	TGTTAAGTG
CTB-47E15	Human	AC008678	138407–139739	TGCATAGTA
RP11-851B10	Human	AP001404	170282–171610	–
AC027380	Mouse	AC027380	146156–147484	TACTGTCAT
RP11-64A1	Human	AC007525	110738–112066	GGCTTTGTG
H_NH017B06	Human	AC079826	148408–149736	GGCTTAGCC
RP23-276B17	Mouse	AC074158	43326–44655	TGCTAAGCA
H505L03s	*C. elegans*	AC006650	9567–12838[d]	–

[a] All sequences were retrieved from GenEMBL database

[b] -, the insertion is not flanked by a 9-bp repeat; nd, not determined

[c] 2-bp deletion of IS*10* terminal inverted repeat

[d] The insertion included part of the Tet resistance genes of Tn*10*

Table 2 Consensus sequences of IS*10* integration sites for clones listed in Table 1

Percentage of sequences that conform to the consensus	Consensus (9 bp)[a]
60	TGCTNAGNN
65	TGCNNNGNN
75	NNCNNNGNN
100	NNCNNNNNN

[a] The program CONSENSUS, implemented in the GCG software package was used for alignment

from tobacco and human DNAs, even after long exposure times (not shown).

Discussion

In this paper, we describe an example of IS*10*R transposition from the bacterial chromosome into a recombinant plasmid during cloning of the R8.3 sequence from tobacco in *E. coli*. The sequence analysis revealed a single integration of IS*10*R with a perfect 9-bp duplication of R8.3 sequence at the insertion site. The cloning steps were carried out in Tn*10*-free *E. coli* strains, indicating that this insertion could not be attributed to the composite Tn*10* transposon. There are only a few reports on the distribution of IS*10* elements in prokaryotic genomes (Matsutani 1991, Chalmers et al. 2000). Our PCR and Southern hybridization data demonstrated the presence of IS*10* sequences in *E. coli* strains that are generally considered to be Tn*10*-free (Figs. 3 and 4). The RFLP analysis (Fig. 4) suggested that the genomic orzganization of IS*10* elements varied in individual *E. coli* strains, possibly as a consequence of their transposition activity.

An interesting question concerns the origin of IS*10* elements in some *E. coli* strains, as it is known that Tn*10* is not a common component of *E. coli* genomes (Deonier

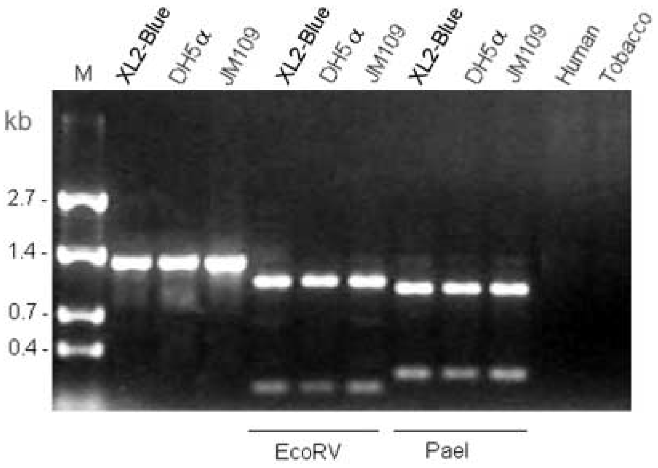

Fig. 3 PCR amplification of IS*10*R sequences from genomic DNAs. The 1.3-kb PCR products were obtained by amplification of DNA isolated from the XL-2 Blue, DH5α and JM109 strains of *E. coli* K-12. Digestion of the products with restriction enzymes yielded fragments of the expected lengths (see Fig. 1). M, size markers (pUC19 digested with *Taq*I)

1996). Bogosian et al. (1993) reported residual copies of IS*10* in the chromosome of *E. coli* LBB84 after its precursor strain had been cured of Tn*10* with fusaric acid. Subsequently, IS*10* transposed in *trans* from the chromosome to some recombinant plasmids grown in this host strain. JM109, DH5α and DH10B could have acquired IS*10* elements in a similar manner. Indeed, Tn*10* transposition occurred twice in the course of construction of the JM109 strain (Yanisch-Perron et al. 1985). DH1 was the only strain (out of five tested) that, according to our Southern hybridization data, did not contain IS*10*. Perhaps this strain has never been manipulated in the presence of Tn*10*.

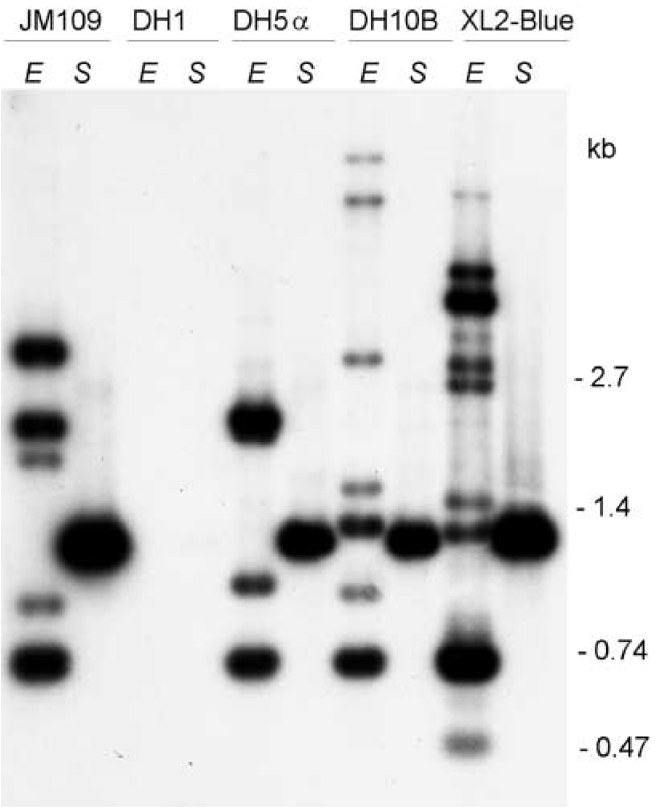

Fig. 4 Southern analysis of total DNAs with the IS*10*R probe. The DNAs purified from the *E .coli* K-12 strains JM109, DH1, DH5α, DH10 and XL-2 Blue were extensively digested with *Sau*3AI (S) and *Eco*RV (E). The digested DNAs (about 1 μg per lane) were fractionated by electrophoresis on an agarose gel and subjected to hybridization using the ^{32}P-labeled probe. The blot was then exposed to X-ray film (Medix). With the exception of DH1 all strains tested gave hybridization signals with the probe. Only the XL-2 Blue is described as a Tn*10*-containing strain and proved to be tetracycline-resistant

The database search also showed contamination with other simple prokaryotic IS elements, listed in Table 3. Strikingly, the highest scores were obtained for IS*10*R; the other IS elements were much less abundant. This result is quite unexpected, since IS*5* has been reported to show relatively high transposition activity (10^{-5}/generation) in *E. coli*, while IS*10*R was much less active (transposing at rates of about 10^{-7}/generation). Also the frequency of insertion hot spots would favor transposition of IS*5* (which has a 5-bp recognition motif) and IS*1* (which inserts at random sites) rather than of IS*10*R (9 bp motif).

Table 3 Result of database search for IS elements in eukaryotic clones

IS element^{a}	Number of homologous clones at E values ≤ 10^{-4}		
	Accession number	≥95% sequence overlap	≤ 95% sequence overlap^{b}
IS*1*	V00609	29	72
IS*2*	Z23101	0	74
IS*3*	X02311	3	28
IS*5*	J01734	18	63
IS*10R*	J01829	203^{c}	356
IS*30*	X00792	2^{d}	12
IS*150*	X07037	9	43
IS*200*	L25845	0	0

[a] The IS sequences were aligned with more than 1.2 million sequences in GenEMBL database using the FASTA program

[b] The majority of IS insertions contained one functional inverted repeat at the end of IS

[c] 101 clones displayed 100% homology to IS*10R*

[d] Both IS*30* insertions are flanked by the same sequence (~800 bp) of *E. coli* origin and a short inverted repeat

So we are left with the question of why IS*10* elements preferentially contaminate plasmids during cloning procedures. Several explanations may be proposed.

1. The first possibility is based on the assumption that different IS elements differ in their propensity to transpose in *trans* (i.e., from chromosome to plasmid). Though the frequency of canonical transposition of IS*10* from chromosome to plasmid is comparable to that of intrachromosomal transposition (Eichenbaum and Livneh 1998), the rate of non-canonical, inside-out, transposition has been reported to be of several orders greater for Tn*10* than for Tn*5* (Chalmers and Blot 1999). Non-canonical transposition events in the plasmid clones are documented by the absence of terminal duplication (Table 1), non-symmetrical direct repeats and the integration of truncated IS elements (Tables 1 and 3).
2. The second possibility is that some situations that arise during cloning into *E. coli* may stress host cells, thus promoting a specific IS*10* movement. Such stresses could include the treatment of cells with transformation buffers that often contain high concentrations of divalent cations (Hanahan 1983). In this context, it is interesting to note that some Tn*10* transposition steps are strongly dependent on magnesium ions and that $Mn2^+$ and possibly other metals could relax target specificity (Junop and Haniford 1997). Generally, stress conditions imposed upon the cells can be expected to induce transposition of genetic mobile elements and DNA rearrangement (McClintock 1984).
3. As a third possibility one might consider the role of relatively recent integration of IS*10* into the chromosome of some host *E. coli* strains. Genetic and epigenetic controls of the mobility of such an artificially delivered transposon might not yet be fully established, as is the case for other IS elements that have been present in *E. coli* for a very long time. Consequently IS*10* may be more active under certain conditions (e.g. stress) than other "ancient" IS elements. In this context, it would be interesting to know how transposon activity changes over time.

According to Bender and Kleckner (1986), IS*10* transposition is non-replicative and its frequency is rather low. The integration sites of Tn*10* have been extensively studied and found to depend on the symmetrical GCTNAGC sequence, as well as on the 6–9 bp flanking this heptanucleotide motif (Bender and Kleckner 1992). We have mapped IS*10* integration sites in a number of randomly selected clones present in the database (Table 2). Most, but not all, sequences matched the published integration hot spot NGCTNAGCN (Halling and Kleckner 1982). More than 60% of integrations occurred into the 9-bp acceptor sequence TGCTNAGNN. Our consensus sequences differed from the published hot-spot sequence in several features. In our alignment there was a strong preference for T in the position of the first nucleotide. Interestingly, T was also present at this position in the integration site reported by Bogosian et al. (1993). Furthermore, the eighth nucleotide was always cytosine or thymine. The first half of the 9-bp repeat seemed to be more conserved than the second. This is in contrast to the purely symmetrical consensus published by Halling and Kleckner (1982). In seven cases, no 9-bp duplication was found (Table 1). The loss of direct repetitions might be indicative of non-canonical transposition (Chalmers and Kleckner 1996). Besides full-length insertions many clones in the database (Table 3) contained homologies with only a part of IS sequences. Inspection of several clones of this type revealed that the homology was limited to the ends of IS and was high (over 95%). It is known that the ends of IS elements contain inverted repeats that serve as recognition sequences for transposase. According to in vitro studies, the IS*10* transposase shows a degree of flexibility with respect to the utilization of two participating transposon ends: they can occur in inverted or direct orientation on the same and also on two different molecules. Moreover, pseudo-transposon ends may be utilized as well (Chalmers and Kleckner 1996). We speculate that non-canonical transpositions, e.g. incomplete IS*10* insertions, might have occurred by intersynaptic rearrangements promoted by intersynapsis of IS*10* and a pseudo-end present in a cloned sequence. If this is the case, larger rearrangements of both plasmid and genomic DNA would be expected.

In conclusion, we have found that IS*10* can transpose relatively frequently during cloning procedures, thus accounting for the appearance of this prokaryotic sequence in eukaryotic clones in databases. With the exception of R8.3M and several other clones, most IS*10* sequences in the database are integrated into recombinant BACs. These vectors are known to carry DNA inserts up to 1 Mb in length and perhaps the occurrence of IS*10* elements in these clones may be related to the frequency of insertion hotspots and cryptic transposon ends within any given sequence. The frequency of IS insertions in nucleotide databases can be roughly estimated from the known number of sequences in a given database and number of positive scores (Table 3). In our analysis we obtained almost 1000 IS entries among eukaryotic sequences in the GenEMBL database. At the time of analysis the database contained 1,201,370 sequences of both eukaryotic and prokaryotic origin. It follows that approximately one clone in every thousand in the database is contaminated by an IS. We therefore recommend that cloned DNA sequences should be checked for the presence of IS*10* and other mobile IS elements. The presence of characteristic terminal duplications (4–9 bp) would be an advantageous marker of a perfect transposition event that is not accompanied by larger DNA rearrangements. These sequence artefacts could be relatively easily corrected by elimination of IS and duplication of acceptor sequence. On the other hand, non-canonical transpositions (marked by e.g. a lack of terminal duplication, long inverted repeats of flanking DNA, truncated IS) would require more thorough analysis to decipher the correct sequence order. The compilation of insertion sites can provide further information about the mechanism of IS element transposition.

Acknowledgements We thank Prof. Dr. M. Bezdék for helpful discussions and valuable comments and H. van der Winden for assistance with DNA sequencing. We are also grateful to one of the anonymous referees for comments which help to improve the manuscript. The work was supported by Grant Agency of the Czech Republic (Grants GAČR 521/01/0037 and IGA AVCR S 5004010) (to A.K. and B.K.) and the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (Grant No. Z21-MED) (A.M. and M.M.).

References

Bender J, Kleckner N (1986) Genetic evidence that Tn*10* transposes by a nonreplicative mechanism. Cell 45:801–815

Bender J, Kleckner N (1992) Tn*10* insertion specificity is strongly dependent upon sequences immediately adjacent to the target-site consensus sequence. Proc Natl Acad Sci USA 89:7996–8000

Bogosian G, Bilyeu K, O'Neil JP (1993) Genome rearrangement by residual IS*10* elements in strains of *Escherichia coli* K-12 which have undergone Tn*10* mutagenesis and fusaric acid selection. Mol Gen Genet 233:17–22

Chalmers R, Blot M (1999) Insertion sequences and transposons. In: Charlesbois RL (ed) Organization of the prokaryotic genome. ASM Press, Washington, D.C., pp 151–169

Chalmers R, Kleckner N (1996) IS*10*/Tn*10* transposition efficiently accommodates diverse transposon end configurations. EMBO J:5112–5122

Chalmers R, Sewitz S, Lipkow K, Crellin P (2000) Complete nucleotide sequence of Tn*10*. J Bacteriol 182:2970–2972

Deonier RC (1996) Native insertion sequence elements: locations, distributions and sequence relationships. In: Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (eds) *Escherichia coli* and *Salmonella*: cellular and molecular biology (2nd edn)., vol. 2. ASM Press, Washington, D.C., pp 2000–2012

Eichenbaum Z, Livneh Z (1998) UV light induces IS*10* tranposition in *Escherichia coli*. Genetics 149:1173–1181

Foster TJ, Davis MA, Roberts DE, Takeshita K, Kleckner N (1981) Genetic organization of transposon Tn*10*. Cell 23:201–213

Halling SM, Kleckner N (1982) A symmetrical six-base-pair target site sequence determines Tn*10* insertion specificity. Cell 28:155–163

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580

Junop MS, Haniford DB (1997) Factors responsible for target site selection in Tn*10* transposition: a role for the DDE motif in target DNA capture. EMBO J 15:2646–2655

Kleckner N (1981) Transposable elements in prokaryotes. Annu Rev Genet 15:341–404

Kovařík A, Koukalová B, Bezděk M, Opatrný Z (1997) Hypermethylation of tobacco heterochromatic loci in response to osmotic stress. Theor Appl Genet 95:301–306

Kuhrová V, Bezděk M, Vyskot B, Koukalová B, Fajkus J (1991) Isolation and characterization of two middle repetitive DNA sequence of nuclear tobacco genome. Theor Appl Genet 81:740–744

Matsutani S (1991) Multiple copies of IS*10* in the *Enterobacter cloacae* MD36 chromosome. J Bacteriol 173:7802–7809

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Mette MF, van der Winden J, Matzke MA, Matzke AJM (1999) Production of aberrant promoter transcripts contribute to methylation and silencing of unlinked homologous promoters in *trans*. EMBO J 18:241–248

Naas T, Blot M, Fitch WM, Arber W (1995) Dynamics of IS-related genetic rearrangement in resting *Escherichia coli* K-12. Mol Biol Evol 12:198–207

Raleigh EA, Kleckner N (1984) Multiple IS*10* rearrangements in *Escherichia coli*. J Mol Biol 173:437–461

Roberts D, Hoopes BC, McClure WR, Kleckner N (1985) IS*10* transposition is regulated by DNA adenine methylation. Cell 43:117–130

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Schleif RF, Wensink PC (1981) Practical methods in molecular biology. Springer-Verlag, New York-Heidelberg-Berlin

Simons RW, Kleckner N (1983) Translation control of IS*10* transposition. Cell 34:683–691

Skaliter R, Eichenbaum Z, Shwartz H, Ascarelli-Goell R, Livneh Z (1992) Spontaneous transposition in the bacteriophage λ *cro* gene residing on a plasmid. Mutat Res 267:139–151

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119