

IS421, a new insertion sequence in *Escherichia coli*

Showbu Sato, Yukari Nakada and Akiko Shiratsuchi

Mitsubishi-Kasei Institute of Life Sciences, 11, Minamiooya, Machida-shi, Tokyo 194, Japan

Received 2 March 1989

The nucleotide sequence of a new insertion sequence (IS) in *Escherichia coli*, IS421, was determined. It is 1340 bp long and contains inverted repeats of 22 bp at its termini. It is flanked by 13 bp direct repeats apparently generated upon insertion. There are two ORFs longer than 200 bp in IS421. One can encode a polypeptide of 371 amino acids (aa) and the other, which is on the other strand, can encode a polypeptide of 102 aa. The C-terminal part of the 371 aa polypeptide shows some homology to that of transposases encoded in some other known IS elements. The copy number of IS421 in chromosomal DNA was 4 for *E. coli* K-12 and B, and 5 for *E. coli* C, as determined by the Southern hybridization of restriction fragments.

Insertion sequence; Nucleotide sequence; Transposase gene; (*Escherichia coli*)

1. INTRODUCTION

During the past years, many bacterial ISs have been isolated and characterized. They are small DNA segments that can transpose to other sites in the DNA and are found in multiple copies in the chromosome of prokaryotes and in their plasmids. At low frequency these elements can also mediate inversion, deletion or fusion of DNA segments, as well as change the normal expression of nearby genes by introduction of transcription start or stop signals. In general, they have one or two long ORF frames which are thought to encode a transposase essential for their own transposition. Terminal inverted repeats, and short direct repeats of the host sequence which are generated on insertion, are specific for each IS element [1,2].

We reported previously the cloning of the *trpE*

and *trpG* of *Thermus thermophilus* HB8 in a tryptophan auxotroph of *E. coli* [3]. Aiming at enhancing the expression of these genes, we tried to locate the genes under the *tac* promoter and to remove the attenuator-like structures. During such manipulations, we found a larger plasmid in a tryptophan auxotroph of *E. coli* which grew faster in the absence of tryptophan. This plasmid was about 1400 bp longer than expected and was found to be inserted with an extra DNA segment downstream of the *trpG* gene. The sequence of this DNA segment suggested that it should be a new IS (IS421). We describe here its complete nucleotide sequence and also its copy number in chromosomal DNA of *E. coli* strains.

2. MATERIALS AND METHODS

2.1. Bacterial strains, plasmids and media

E. coli JA221 (*hsdR*, *trpES*, *leuB6*, *lacY*, *recA1*) was kindly provided by Dr K. Nakamura. *E. coli* B wild type and *E. coli* C_{1a} were a gift from Dr A. Kikuchi. pBW11 is a derivative of pBR322 which has the *T. thermophilus trpE* and *trpG* genes between the *Bam*HI and *Pvu*II sites [3]. pKW3316 is a derivative of pKK223-3 containing the thermophile *trpE* and *trpG* and IS421.

2.2. DNA sequencing

pKW3316 was digested with *Sac*I, and the two *Sac*I fragments

Correspondence address: S. Sato, Mitsubishi-Kasei Institute of Life Sciences, 11, Minamiooya, Machida-shi, Tokyo 194, Japan

Abbreviations: aa, amino acid; bp, base pair; IS, insertion sequence; ORF, open reading frame

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no.Y07501

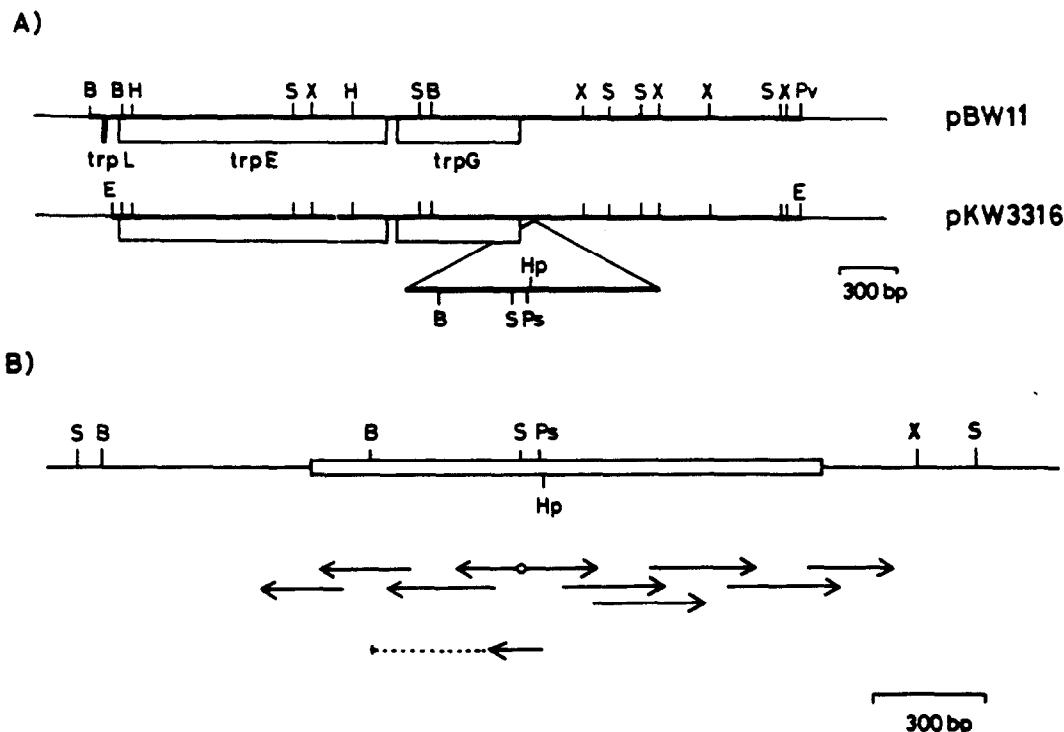


Fig.1. (A) Restriction maps of pBW11 and pKW3316. pKW3316 has an extra DNA segment (IS421), which is shown by the protrudent bold line. In pKW3316, the thermophile DNA starting at 21 bp upstream of the initiation codon of the *trpE* is located downstream of the *tac* promoter of pKK223-3 through the *EcoRI* site. (B) The strategy used for sequencing IS421. Restriction sites: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; Hp, *Hpa*I; Ps, *Pst*I; Pv, *Pvu*II; S, *Sac*I; X, *Xho*I.

```

.....
GCTCTCCCTCAGGCTCAAGGCCACCAAAGCCCCGCCCGCCCGCACCGGGGAGACCTCCCC
-35          -10          ΔΔΔ
ATAAGCGCTAACTTAAGGGTTGTGGTATTACGCCTGATATGATTTAACGTGCCGATGAAT 60
TATTCGCGATTGAATTCCTCAACACCATAATGCGGACTATACTAAATTGCACGGCTACTTA

TACTCTCACGATAACTGGTCAGCAATTCTGGCCCATATTGGTAAGCCCCGAAAGAACTGGAT 120
ATGAGAGTGCTATTGACCAAGTCGTTAAGACCGGGGTATAACCATTTCGGGCTTCTTGACCTA

ACTTCGGCACGTAATGCCGGGGCTCTAACCCGCCCGCCGCGAAATTCGTGATGCTGCAACT 180
TGAAGCCGTGCATTACGGCCCCGAGATTGGGCGGCGCGCTTTAAGCACTACGACGTTGA

CTGCTACGCTCTGGGGCTGGCTTACGGCCCCGGGGGGGATGTTCGTTCATTACGTGAAGTCACT 240
GACGATGCAGACCCCGACCGAATGCCGGGGCCCCCTACAGCAGTAATGCACCTTCAGTGA

GCATGGGCTCAGCTCCATGACGTTGCAACATTATCTGACGTGGCTCTCCTGAAGCGGCTG 300
CGTACCCGAGTCGAGGTACTGCAACGTTGTAATAGACTGCACCGAGAGGACTTCGCCGAC

CGGAATGCCGCGGACTGGTTTGGCATACTTGCCGCACAAACACTTGCTGTACGCGCCGCA 360
GCCTTACGGCGGCTGACCAAACCGTATGAACGGCGTGTGTTGTGAACGACATGCGCGGCGT

```

Fig.2. Nucleotide sequence of IS421. The sequence of IS421 is drawn in both strands and that of the flanking regions in only one strand. Terminal inverted repeats of IS421 and the target duplication on pKW3316 are indicated by the arrows and the dotted lines, respectively. The possible promoter sequences, two on the forward and one on the reverse strand, are shown by -35 and -10 on thin lines. The large ORFI on the forward strand is from the open triangles to the filled ones, and the short ORFII on the reverse strand from open to filled squares.

GTTACGGGTTGTACAAGCGGAAAGAGATTGCGTCTTGTGCGATGGAACAGCAATCAGTGCG 420
 CAATGCCCAACATGTTTCGCTTCTCTAACGCAGAACAGCTACCTTGTCGTTAGTCACGC
 CCCGGGGGCGGCAGCGCTGAATGGCGACTACATATGGGATATGATCCTCATACCTGTCAG 480
 GGGCCCCCGCGTCGCGACTTACCGCTGATGTATACCCTATACTAGGAGTATGGACAGTC
 TTCACTGATTTTGAGCTAACCGACAGCAGAGAC TGAACGGCTGGACCGATTGCGCAA 540
 AAGTGACTAAAACTCGATTGGCTGTCTCTCTGACTTGCCGACCTGGCTAAACGCGTT
 ACGGCAGACGAGATACGCATTGCTGACCGGGGATTTCGGTTCGCGTCCCGAATGTATCCGC 600
 TGCCGTCTGCTCTATGCGTAACGACTGGCCCCCTAAGCCAAGCGCAGGGCTTACATAGGCG
 TCACTTGCTTTTGAGAGAAGCTGATTATATCGTCCGGGTTCACTGGCGAGGATTGCGCTGG 660
 AGTGAACGAAAACCTCTTCGACTAATATAGCAGGCCCAAGTGACCGCTCCTAACGCGACC
 TTAACGCGAGAAGGAATGCGCTTTGACATGATGGGTTTTCTGCGCGGGCTGGATTGCGGT 720
 AATTGACGTCTTCTTACGCGAACTGTACTACCCAAAAGACGCGCCCGACCTAACGCCA
 AAGAACGGTGAAACCACTGTAATGATAGGCAATTCAGGTAATAAAAAAGCCGGAGCTCCC 780
 TTCTTGCCACTTTGGTGACATTACTATCCGTTAAGTCCATTATTTTTTCGGCCTCGAGGG
 TTTCCGGCACGTCTCATTTGCCGTATCACTTCCTCCCGAAAAAGCATTAATCAGTAAACC 840
 AAAGGCCGTGCAGAGTAACGGCATAAGTGAAGGAGGGCTTTTCGTAATTAGTCATTTTGG
 □□□
 CGACTGCTCAGCGAGAATCGTTCGAAAAGGACGAGTAGTTTCAGGCGGAAACGCTGGAAGCA 900
 GCTGACGAGTCGCTCTTAGCAGCTTTTCTGCTCATCAAGTCCGCCTTTGCGACCTTCGT
 GCGGGCCATGTGCTATTGCTAACATCATTACCGGAAGATGAATATTAGCAGAGCAAGTG 960
 CGCCCGGTACACGATAACGATTGTAGTAATGGCCTTCTACTTATAAGTCGTCTCGTTAC
 GCTGATTGTTACCGTCTGCGATGGCAAATTGAACTGGCTTTTAAGCGGCTCAAAAGTTTG 1020
 CGACTAACAATGGCAGACGCTACCGTTTAACTTGACCGAAAATTGCGCGAGTTTCAAAC
 -10 -35
 CTGCACCTGGATGCTTTGCGTGCAAAGGAACCTGAACTCGCGAAAGCGTGGATATTTGCT 1080
 GACGTGGACCTACGAAACGCACGTTTCTTTGGACTTGAGCGCTTTTCGCACCTATAAACGA
 -35 -10
 AATCTACTCGCCGATTTTTTAATTGACGACATAATCCAGCCATCGCTGGATTTCCCCCCC 1140
 TTAGATGAGCGGCGTAAAAATTAACGCTGTATTAGGTCGGTAGCGACCTAAAGGGGGG
 AGAAGTGCCGGATCCGAAAAGAAGAACTAAGTTCGTTGTGGAGAATAACAAAAATGGTCAT 1200
 TCTTCACGGCCTAGGCTTTTCTTCTTGATTGAGCAACACCTCTTATTGTTTTTACCAGTA
 CTGGAGCTTACAGGTGGCCATTTCGTGGGACAGTATCCCTGACAGCCTACAAAACGCAATT 1260
 GACCTCGAATGTCCACCGGTAAGCACCCTGTATAGGGACTGTCGGATGTTTTGCGTTAA
 GAAGAACGCGAGGCATCGTCTTAACGAGGCACCGAGGCGTCGCATTCTTCAGATGGTTCA 1320
 CTTCTTGCGCTCCGTAGCAGAATTGCTCCGTGGCTCCGCAGCGTAAGAAGTCTACCAAGT

 ACCCTTAAGTTAGCGCTTATGGGAGACCTCCCCCGCCATCAGGGCCCGCATGACCTCGTA
 TGGGAATTCAATCGCGAATA
 GGCCTCCTCTTCCTCCAAAACCTCGCCCAGAATGGCCTTCTTCA

of about 1150 bp were subcloned onto pUC12 at the *SacI* site. Deletion plasmids were prepared from the two kinds of pUC12-derivatives by the method of Frischaut et al. [4] using

the *XbaI* site. All the deletion plasmids were digested with *XbaI*, labeled at the 5'- or the 3'-end, and subjected to the method of Maxam and Gilbert [5]. To overlap the two *SacI*

fragments, the *Bam*HI-*Hpa*I fragment of about 500 bp was subjected to the chemical degradation method [5]. The whole sequence was analyzed for both strands.

2.3. Southern hybridization

The *Bam*HI-*Hpa*I fragment within IS421 was labeled with 32 P by nick-translation and used as a probe. Chromosomal DNA was exhaustively digested with a single or two restriction enzymes, and the fragments were electrophoretically resolved on vertical 0.9% agarose slab gels and blotted to nylon membrane.

3. RESULTS AND DISCUSSION

3.1. Nucleotide sequence of IS421

A plasmid, pKW3316, was unexpectedly obtained during gene manipulation of pBW11 for improving expression of the thermophile *trp* genes. It contained an extra DNA segment of about 1400 bp downstream of the *trpG* (fig.1A). Sequencing of this segment was carried out as shown in fig.1B and the result was shown in fig.2. The orientation of the sequence is opposite to that of the physical map in fig.1A to read the large ORF from left to right, and the right flanking sequence ends at the complementary and inverted termination codon (TCA) of the *trpG*. The exact length of the sequence is 1340 bp and it has the 22 bp inverted repeats defining the end of the element. At 5' as well as 3' immediately to the inverted repeats are direct repeats of a 13-mer, GGGAGGTCTCCCC, which is present only once in the corresponding region of pBW11. These features suggest that this DNA segment is an IS, and it was named IS421.

There are two ORFs longer than 200 nucleotides in IS421. One, ORFI, begins with GTG at position 49–51 or ATG at 55–57, and ends with TAA at position 1168–1170. The other, ORFII, is found on the complementary strand, beginning with GTG at position 790–788 and ending with TGA at 485–483. According to the Fickett's equation [6], ORFI may possibly be a protein-coding sequence while ORFII is not. The presence of one long ORF on one strand with one short ORF on the other strand is a common feature of some other IS elements.

Three possible promoter regions were found in IS421 with the aid of a computer. One of them precedes ORFI, and its –35 sequence, TTGTGG, is located at position 20–25 and the corresponding –10 sequence is TTTAAC at position 43–48 or TAACGT at position 45–50 with spacing of 17 or

19 bp, respectively. Since the first start signal (GTG) of ORFI at position 49–51 is located too close to or involved in the –10 sequence, we tentatively take the first ATG at position 55–57 for the initiation signal of ORFI although no Shine-Dalgarno sequence precedes the ATG. It is known that translation occurs without the Shine-Dalgarno sequence albeit with low efficiency [7]. Thus, ORFI is 1113 bp long, encoding a 371 aa polypeptide. *E. coli* JA221 harboring pKW3316 produced more anthranilate synthase I and II, gene products of *trpE* and *trpG*, than that harboring a pKW3316-like plasmid devoid of IS421 (unpublished). How IS421 promotes the expression of its upstream genes is unclear at present. However, there is a sequence in the reverse strand of position 988–960 (fig.2) which resembles the ρ -dependent termination sequence of λ phage t_R [8], a sequence able to form a stem and loop followed by CAAUCA.

3.2. Comparison of IS421 and other known IS elements

The nucleotide sequence of IS421 was compared with that of other known IS elements, IS1 [9], IS2 [10], IS3 [11], IS4 [12], IS5 [13], IS26 [14], IS30 [15], IS102 [16], IS10R [17], IS50R [18] in *E. coli* IS231 [19] in *Bacillus thuringiensis* and ISH1 [20],

	10	20	30	40
IS421	TLEAAGHVLLTSLPDEYSAEQVADCYRLRWQIELAFKR			
IS4	ERKGVCHTSMTDAMRFPGGENGCLSHHETCYRE			
IS10R	SASKEPSVAAHVEIRTPKQLVNIYSKMMETTERD			
IS50R	GETPLKWLGTGEPVNSLAQQLAIDITHRRSEFHA			
IS231	SKRLTINIVYSNTEGIVPMIHPFSSSSIIET			
ISH1	SESTRPLIXHREQTQLQAHARMNEDNQMSRTGSQ			
			*	*
	50	60	70	
LKSL---LNLDALRAKEPELAKAWIFANLLAAFL	(279-349/371)			
IQTMQRSRE-TSEKVEQLWGVVYNE	(280-352/442)			
PAYGGERNSSTSSSRFDIMLLIAHMLQLT	(258-331/402)			
WETGAGAEQRNEEPDNLRMVSIISFVAVRLQ	(277-350/461)			
WFIHVV---QNIQRLRLECHVYGRHIF	(315-385/478)			
EDDGEKRSRGPPV-TRKCIHTQAS	(197-270/270)			
			*	

Fig.3. Comparison of a protein encoded by the ORFI of IS421 with transposases of other IS elements, IS4 [12], IS10R [17], IS50R [18], IS231 [19] and ISH1 [20] in their C-terminal amino acid sequences. The position of the amino acid sequence shown in the figure and the total number of amino acid residues of each protein are indicated in the parentheses. Amino acid residues identical with the IS421 protein are shaded, and the residues common throughout all the transposases are indicated by asterisks.

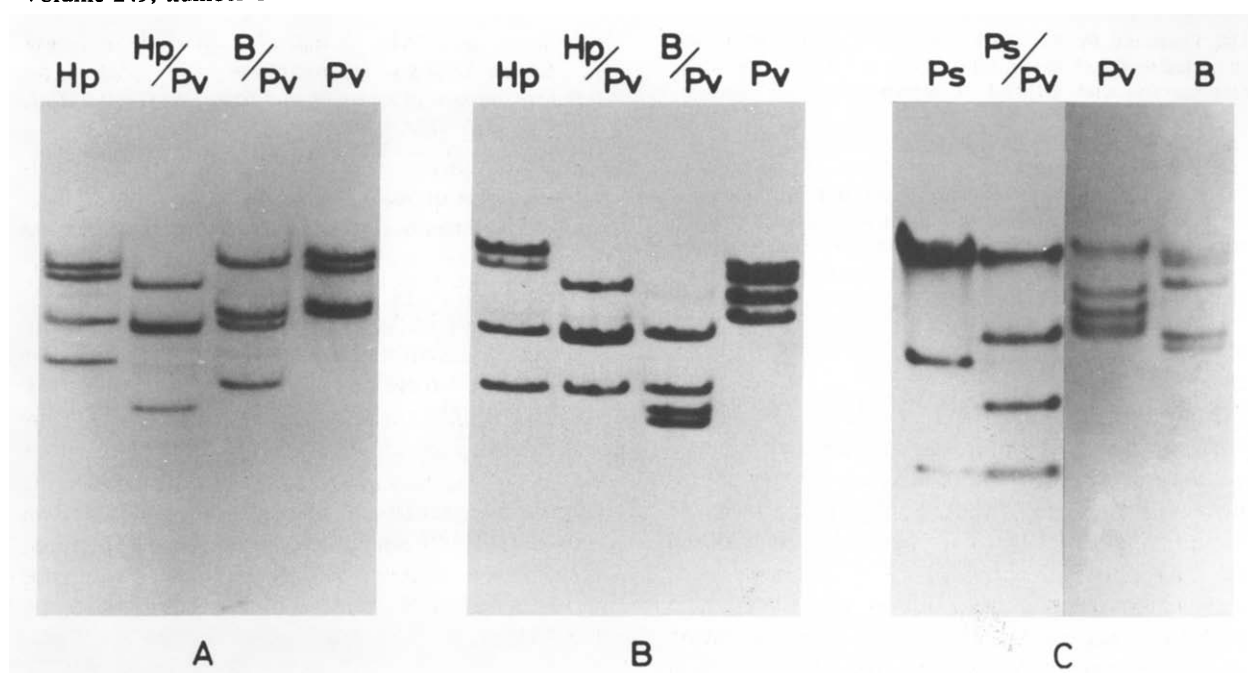


Fig.4. Detection of IS421-homologous sequences in restriction fragments of chromosomal DNA of three *E. coli* strains by the Southern hybridization. (A) Chromosomal DNA of *E. coli* JA221 (a derivative of K-12), (B) *E. coli* B wild type and (C) *E. coli* C_{1a}. Hp, *Hpa*I; B, *Bam*HI; Pv, *Pvu*II; Ps, *Pst*I.

ISH2 [21] and ISH50 [22] in *Halobacterium*. We have found no striking similarity in general, but homologies slightly more frequent than expected for random sequence were found in IS5, IS50R and ISH50. ISs are supposed to encode a transposase, and Mahillon et al. [19] suggested that transposases of some ISs had some homology to one another in the C-terminal region. The protein encoded by ORF1 of IS421 also showed some homology to those transposases in the C-terminal region (fig.3).

3.3. Copy numbers of IS421 in *E. coli* chromosomal DNA

IS421 was found during the manipulation of *T. thermophilus* genes using *E. coli* JA221, which suggests that it could originate from the host *E. coli*. We determined copy number of IS421 in *E. coli* JA221 (originated K-12) and other strains, *E. coli* B and *E. coli* C (fig.4). For *E. coli* JA221 (fig.4A), four bands with apparently uniform intensity were detected in the *Hpa*I fragments, *Pvu*II fragments, and *Bam*HI/*Pvu*II fragments. The *Pvu*II/*Hpa*I digestion gave three bands, but the second band looked doubled in intensity. Similar

banding patterns of restriction enzyme fragments hybridized to the IS421 fragment were obtained for *E. coli* B wild type although their mobility on the electrophoresis was different from that with *E. coli* JA221 DNA (fig.4B). From these results, *E. coli* JA221 and *E. coli* B wild type were inferred to have four copies of IS421. As for *E. coli* C_{1a}, the maximum number of fragments that hybridized to the IS421 fragment is five, which are found in the *Pvu*II and the *Bam*HI digests (fig.4C). Since the five bands in each digest seem to have similar strength to one another, we estimated that *E. coli* C_{1a} has five copies of IS421.

Acknowledgements: We thank Dr T. Tanaka for helpful discussion. This work was partly supported by a special fund of the Science and Technology Agency of Japan.

REFERENCES

- [1] Kleckner, N. (1981) *Annu. Rev. Genet.* 15, 341–404.
- [2] Iida, S., Meyer, J. and Arber, W. (1983) in: *Mobile Genetic Elements* (Shapiro, J.A. ed.) pp.159–221, Academic Press, New York.
- [3] Sato, S., Nakada, Y., Kanaya, S. and Tanaka, T. (1988) *Biochim. Biophys. Acta* 950, 303–312.

- [4] Frischauf, A.M., Garoff, H. and Lehrach, H. (1980) *Nucleic Acids Res.* 8, 5541–5549.
- [5] Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499–560.
- [6] Fickett, J.W. (1982) *Nucleic Acids Res.* 10, 5303–5318.
- [7] Ptashne, M., Backman, K., Humayum, M.Z., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, R.T. (1976) *Science* 194, 156–161.
- [8] Rosenberg, M., Court, D., Shimatake, H., Brady, C. and Wulff, D.L. (1978) *Nature* 272, 414–423.
- [9] Ohtsubo, H. and Ohtsubo, E. (1978) *Proc. Natl. Acad. Sci. USA* 75, 615–619.
- [10] Ghosal, D., Sommer, H. and Saedler, H. (1979) *Nucleic Acids Res.* 6, 1111–1122.
- [11] Timmerman, K.P. and Tu, C.-P.D. (1985) *Nucleic Acids Res.* 13, 2127–2139.
- [12] Klaer, R., Kuhn, S., Tillmann, E., Fritz, H.-J. and Starlinger, P. (1981) *Mol. Gen. Genet.* 181, 169–175.
- [13] Engler, J.A. and Van Bree, M.P. (1981) *Gene* 14, 155–163.
- [14] Mollet, B., Iida, S., Shepherd, J. and Arber, W. (1983) *Nucleic Acids Res.* 11, 6319–6330.
- [15] Dalrymple, B., Caspers, P. and Arber, W. (1984) *EMBO J.* 3, 2145–2149.
- [16] Bernardi, A. and Bernardi, F. (1981) *Nucleic Acids Res.* 9, 2905–2911.
- [17] Halling, S.M., Simons, R.W., Way, J.C., Walsh, R.B. and Kleckner, N. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2608–2612.
- [18] Auerswald, E.-A., Ludwig, G. and Schaller, H. (1981) *Cold Spring Harbor Symp. Quant. Biol.* 45, 107–113.
- [19] Mahillon, J., Seurinck, J., Van Rampuy, L., Delcour, J. and Zabeau, M. (1985) *EMBO J.* 4, 3895–3899.
- [20] Simsek, M., DasSarma, S., RajBhandary, U.L. and Khorana, H.G. (1982) *Proc. Natl. Acad. Sci. USA* 79, 7268–7272.
- [21] DasSarma, S., RajBhandary, U.L. and Khorana, H.G. (1983) *Proc. Natl. Acad. Sci. USA* 80, 2201–2205.
- [22] Xu, W.-L. and Doolittle, W.F. (1983) *Nucleic Acids Res.* 11, 4195–4199.