The EMBO Journal vol.4 no.12 pp.3357–3365, 1985

Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants

Ellen Murphy, Leslie Huwyler and Maria do Carmo de Freire Bastos

Department of Plasmid Biology, The Public Health Research Institute of the City of New York, Inc., 455 First Avenue, New York, NY 10016, USA

Communicated by S.D.Ehrlich

The complete nucleotide sequence of the *Staphylococcus aureus* transposon Tn554, which encodes resistance to erythromycin and spectinomycin, was determined by the dideoxy chain termination method. The transposon was found to be 6691 bp in length and to contain six open reading frames of >125 amino acids. Small insertion and deletion mutations were obtained in each of these by *in vitro* mutagenesis at restriction endonuclease cleavage sites and the mutants characterized with respect to transposition functions and antibiotic resistance markers. Three of the reading frames, designated *tnpA*, *tnpB* and *tnpC*, encode functions that are required for transposition of Tn554; genetic analysis indicated that these three genes define distinct complementation groups of transposition-defective mutants. Two of the open reading frames correspond to the resistance determinants *spc* and *ermA*, the sixth, designated ORF, has no known function. Tn554-specific peptides corresponding to *tnpA*, and *spc* were identified in a coupled transcription-translation system *in vitro*.

Key words: transposon Tn554/transposition/mutagenesis

Introduction

The *Staphylococcus aureus* transposon Tn554, encoding resistances to the macrolide-lincosamide-streptogramin B (MLS) antibiotics and to spectinomycin (Phillips and Novick, 1979), has several unique features that distinguish it from other prokaryotic and eukaryotic transposable elements. Its ends are asymmetric, lacking either inverted or direct terminal repeats, it does not generate a duplication of a target sequence upon transposition, and it is extremely site-specific, always inserting between the same nucleotide pair in the *S. aureus* chromosome (Krolewski *et al.*, 1981; Murphy and Lofdahl, 1984). Transposition to the chromosome occurs in only one orientation; however, when the target site is cloned on a plasmid vector, insertion remains site- but not orientation-specific. The target site contains a 6-bp sequence that matches the terminal 6 bp of the right end of Tn554; transposition into the cloned primary site occurs on one or the other side of this sequence, depending on the orientation (Murphy and Lofdahl, 1984).

These results suggest that Tn554 transposes by a mechanism that is significantly different from those that have been proposed for other transposable elements (Shapiro, 1979). As part of our study of this problem, we have begun to investigate the functions that Tn554 requires for its transposition. Here we report the complete nucleotide sequence of Tn554 and the isolation of mutants affecting transposition functions and antibiotic resistance. The sequence of the *spc* and *ermA* determinants have been reported previously (Murphy, 1985a, 1985b). The results of genetic studies with these mutants (Bastos and Murphy, in preparation) confirm the hypothesis suggested by the nucleotide sequence data, that Tn554 encodes three distinct functions required for its transposition.

Results

Nucleotide sequence of Tn554

The complete nucleotide sequence of Tn554 is shown in Figure 1. 96% of the sequence was verified by determining the sequence of both strands. In two instances (nt 4689–4691 in *ermA* and nt 2695–2698 in *tnpB*) the complementary strands did not agree; both read GGG on one strand and CC on the complement. In both cases the GGG was taken to be the correct sequence and in both cases this allowed an open reading frame to continue. For *ermA* this reading is assumed to be correct because the deduced protein sequence is homologous to that of *ermC* and *ermAM* (Murphy, 1985a); for *tnpB*, the reading is strengthened by the fact that mutants mapping both 5′ and 3′ to the ambiguous region fall within the same complementation group (Bastos and Murphy, in preparation).

Tn554 is 6691 bp in length and contains six major open reading frames, five of them reading from left to right as the sequence is written. Three of these, designated *tnpA*, *tnpB* and *tnpC*, define genes required for transposition (see below); two correspond to the antibiotic resistance markers *spc* (Murphy, 1985b) and *ermA* (Murphy, 1985a). The function (if any) of the sixth reading frame, designated ORF, is unknown. No other potential coding regions of >90 amino acids were found on either strand. These six open reading frames, accounting for 83% of the sequence of Tn554, average 66.5% A+T, compared with 72.2% A+T for the noncoding regions and 67.4% A+T overall (*Staphylococcus aureus* DNA is 68% A+T) (Szybalski, 1968). The main features of the sequences are summarized below and in Table I and Figure 2.

tnpA. The first open reading frame, *tnpA*, located between nt 134 and 1217, potentially specifies a 43-kd protein. The suggested initiation codon, GTG, is preceded by a good Shine-Dalgarno site, AAAGAGGTG, typical of gram-positive organisms (Figure 3) (McLaughlin *et al.*, 1981; Shine and Dalgarno, 1974). The predicted protein sequence is rich in basic amino acids (20.5%; net charge +29), and aromatic amino acids, particularly tyrosine.

tnpB. The second, and largest, open reading frame potentially encodes a 630 amino acid protein of 74 kd. The deduced peptide is, like *tnpA*, highly basic (net charge +19). Its putative ATG start codon overlaps by 2 bp the TGA stop codon of *tnpA* and is preceded by a Shine-Dalgarno site that overlaps the three carboxy-terminal amino acids of *tnpA*.

tnpC. A third, short open reading frame capable of encoding a peptide of 125 amino acids lies immediately 5′ to *tnpB* (nt 3115–3490). Six nucleotides separate the *tnpB* TAA termination codon and a possible ATG initiation codon for *tnpC*.

3358

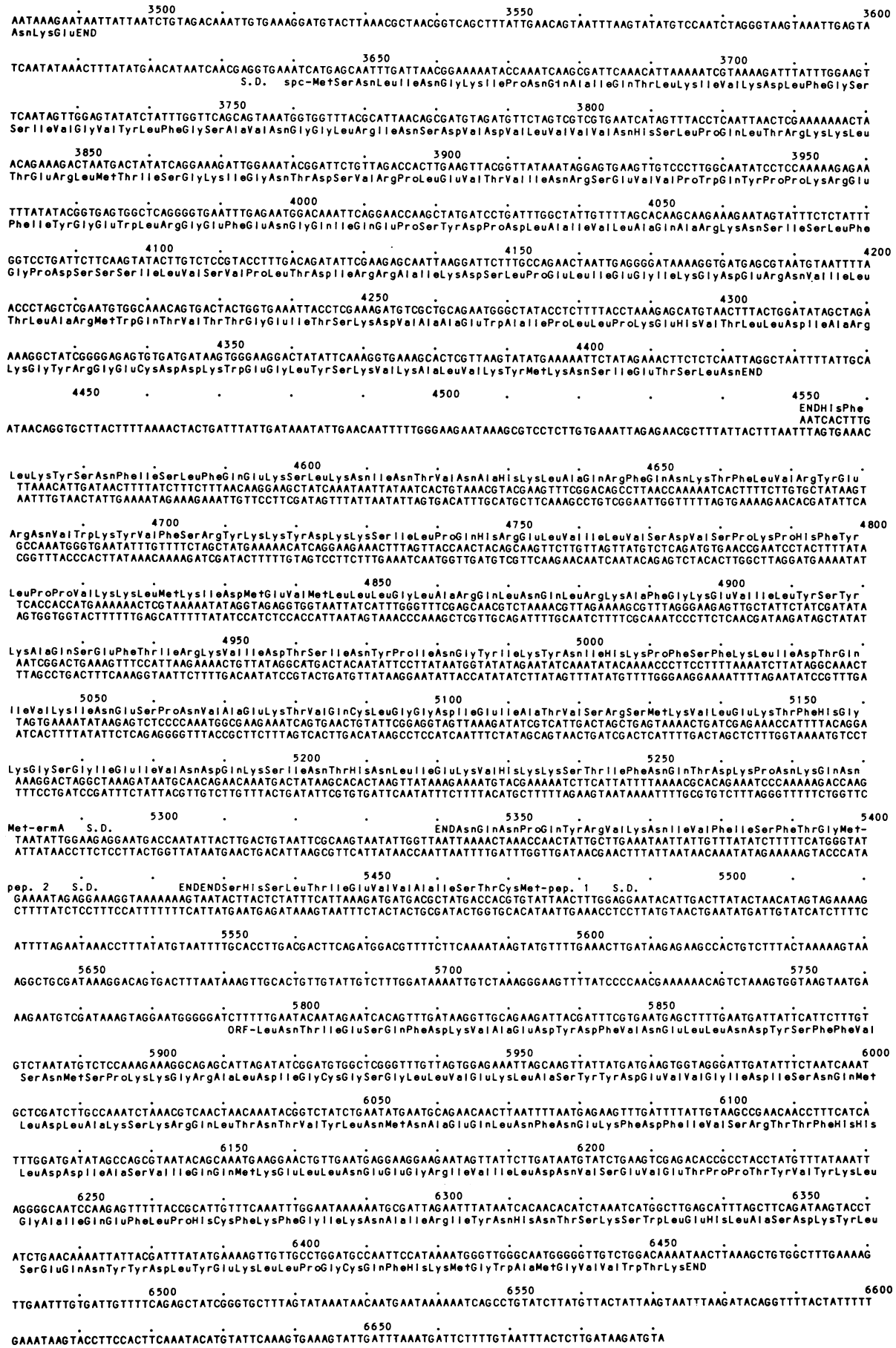

Fig. 1. Nucleotide sequence of Tn554. The sequence is given for one strand, from the left to the right end of the transposon. For *ermA*, which is transcribed off the complementary strand, both strands are shown.

Table I. Proteins deduced from the nucleotide sequence of Tn554

Gene	Size		Start		Stop		%A+T
	Amino acids	M_r, daltons	nt	Codon	nt	Codon	
tnpA	361	42 880	134	GUG	1217	UGA	67.8
tnpB	630	74 373	1216	AUG	3106	UAA	66.1
tnpC	125	14 791	3115	AUG	3490	UAA	70.1
spc	260	28 957	3643	AUG	4423	UAG	63.6
ermA	243	28 380	5283	AUG	4553	UAA	67.5
ORF	220	25 396	5795[a]	UUG	6471	UAA	66.1

[a]Suggested start codon for ORF based on proximity to the best possible Shine-Dalgarno sequence; most probably ORF is not translated.

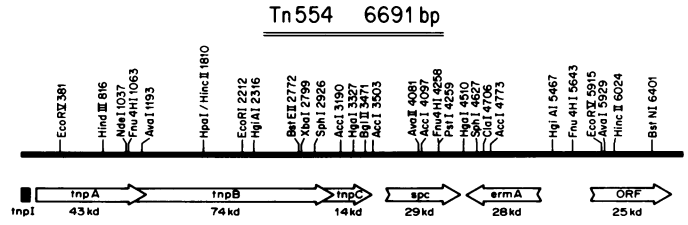

Fig. 2. Functional and restriction map of Tn554. *tnpI* is a locus responsible for trans-inhibition of transposition (Murphy, 1983).

spc. The antibiotic resistance genes have been discussed in detail elsewhere (Murphy, 1985a, 1985b). The *spc* determinant encodes a unique adenyltransferase, AAD(9), that modifies spectinomycin but not streptomycin. The primary amino acid sequence of the deduced 29-kd protein shows significant (35%) homology with a streptomycin-spectinomycin adenyltransferase, AAD(3″) (9), that is widespread among gram-negative bacteria (Hollingshead and Vapnek, 1985; Murphy, 1985b; M. Fling, personal communication).

ermA. *ermA* specifies a 28-kd S-adenosylmethionine-dependent methylase that specifically dimethylates 23S rRNA (Lai and Weisblum, 1971). The amino acid sequence of the *ermA* methylase and the nucleic acid sequence of its regulatory region are homologous to that of the *ermC* rRNA methylase (Gryczan *et al.*, 1980; Murphy, 1985a). Of all the open reading frames, only *ermA* is transcribed from right to left as the map is drawn (Figure 2); thus the *spc* and *ermA* transcripts converge. The 3′ intergenic region contains a potential stem and loop structure that probably functions as a transcription terminator for both the *ermA* and *spc* mRNAs (Murphy, 1985b).

ORF. The right end of Tn554 contains an open reading frame (ORF) sufficient to encode a 225 amino acid protein. Although a very abundant transcript hybridizes to this region (see below), there is no evidence that it is translated: no band corresponding to ORF is evident in the *in vitro* transcription-translation system and the best Shine-Dalgarno site, preceding a TTG start codon, is a relatively poor one. Mutants at three positions in ORF (see below) have no effect upon either the frequency or specificity of transposition, nor does Tn554 confer resistance to any of ~30 antibiotics and heavy metal ions tested (data not shown).

Isolation of mutations at restriction site

Using pEM9631 DNA, small insertion or deletion mutants were created at 19 restriction sites within Tn554 (plasmids used in this study are shown in Table II). Following transformation with the treated and re-ligated DNAs, plasmid DNA was isolated and assayed for the loss of the restriction site in question, and colonies were scored for resistance to spectinomycin and erythromycin and for transposition proficiency (Table III).

		Spacing	ΔG
tnpA	uauuuAAAGAGGUGggaacauGUGgag	12	-15.4
tnpB	tcgagAGAAAGGAGcauaagaaAUGaau	10	-18.4
tnpC	aguuuGAcGGGGUAAUuaucaAUGgau	11	-14.8
spc	uaaucAAcGAGGUGAaaucAUGagc	10	-15.2
peptide 1	pppacauAAGGAGGUuucaauuAUGugc	11	-19.0
peptide 2	aaaugGAAAGGAGauaaaaguAUGggu	10	-16.2
ermA	uaaccAGuAAGGAGaagguuauaAUGuuc	11	-14.2
ORF	aguagGAAuGGgGGAUCuuuUUGaau	9	-15.4

Fig. 3. Postulated Shine-Dalgarno sites of Tn554. Nucleotides in upper case are those that match the 3′ end of *Bacillus subtilis* 16S rRNA. 'Spacing' is from the base suggested to pair with the U of the CCUCC of the rRNA to the initiation codon. Free energies (Kcal) were calculated according to Tinoco *et al.* (1973).

Table II. Strains and plasmids

Strain[a]	Plasmid	Genotype	Reference
RN4470	–	1.4-kb Cm^r fragment replacing 900 bp of target site in chromosome	Murphy and Löfdahl, 1984
RN4565	pEM9634	2.8-kb chromosomal target insert in pT181*repC*-ts (pRN8057)	Murphy and Löfdahl, 1984
RN4652	pEM9700	171-bp target insert in pT181*cop623*	
RN4491	pEM9631	Tn554*ermI1*→pEM9634	
RN4689	pEM9698	Tn554 (wild-type)→pEM9634	
RN4934	pEM9717	Tn554*ermI1*→pEM9700	
BD404	–	*B. subtilis* deficient in ATP-dependent nuclease	Doley *et al.*, 1974

[a]All *Staphylococcus aureus* strains are derivatives of NCTC 8325-4.

Transposition-defective mutants. With a single exception, all mutations created at any of 10 restriction sites located between nt 381 (*Eco*RV) to 3471 (*Bgl*II) abolished the ability of Tn554 to transpose, while mutations mapping to the right of the *Bgl*II site did not affect transposition. These mutations include those at four restriction sites within each of the *tnpA* and *tnpB* coding regions and two in *tnpC*. Typical transduction frequencies for the mutants were 10^4- to 10^5-fold lower than wild-type (Table III), which corresponds to fewer than 10 transductant colonies per plate. Southern blot hybridization of whole-cell DNA from some of these colonies indicated that Tn554 DNA was present at a secondary chromosomal location rather than in the cloned plasmid target (data not shown), suggesting that they did not arise by a normal transposition event.

The exceptional mutant was *tnpA27*, obtained after limited *Bal*31 digestion at the *Eco*RV site at nt 381, which contains a 15 base, in-frame deletion of nucleotides 386–400. *TnpA27* transposed at 10% of the wild-type frequency and retained wild-type site and orientation specificity (Figure 4). Frameshift mutations at this site, such as *tnpA44* and *tnpA26*, were completely transposition-defective.

In four instances (*tnpA60, tnpB31, tnpB18* and *tnpB29*) transposition-defective mutants which were identified on the basis of phenotype still contained the restriction site in question. Sequence analysis verified that these contained wild-type sequences in the region of the restriction site. These mutants were mapped

Table III. Characteristics of restriction-site mutants of Tn554

Mutant	Location	Sequence change[a]	Phenotype	Transposition frequency[b]	Reversion frequency
tnpA26	*Eco*RV 381	Δ123bp	tnp^-	N.T.	
tnpA27[c]	*Eco*RV 381	Δ15bp	tnp^+	5.1×10^{-5}	
tnpA44	*Eco*RV 381	Δ64bp	tnp^-	2.8×10^{-8}	
tnpA60	645	Δ1bp	tnp^-	5.3×10^{-8}	
tnpA41	*Hind*III 816	(Ω4bp)	tnp^-	1.2×10^{-8}	
tnpA7	*Nde*I 1037	(Ω2bp)	tnp^-	4.0×10^{-8}	
tnpA4	*Ava*I 1194	Ω4bp	tnp^-	1.5×10^{-8}	
tnpA85	*Ava*I 1194	(Ω4bp)	tnp^-	N.T.	
tnpB31	1385	G→A	tnp^-	1.4×10^{-7}	
tnpB3	*Hpa*I 1810	Δ1bp	tnp^-	4.4×10^{-9}	
tnpB12	*Eco*RI 2212	(Ω4bp)	tnp^-	1.4×10^{-8}	
tnpB15	*Bst*EII 2772	Ω5bp	tnp^-	1.4×10^{-8}	
tnpB164	*Sph*I 2925	(Δ4bp)	tnp^-	N.T.	
tnpB177	*Sph*I 2925	(Δ4bp)	tnp^-	N.T.	
tnpB18	?		tnp^-	3.3×10^{-8}	
tnpB29	?		tnp^-	1.5×10^{-8}	
tnpC80	*Acc*I 3190	Ω2bp	tnp^-	6.2×10^{-8}	
tnpC45	*Acc*I 3190-3503	Δ312bp	tnp^-	1.8×10^{-8}	
tnpC12	*Bgl*II 3471	Ω4bp	tnp^-	7.8×10^{-8}	
Acc128	*Acc*I 3503	(Ω2bp)	tnp^+	N.T.	
Acc136	*Acc*I 3503	(Ω2bp)	tnp^+	N.T.	
spc23	*Acc*I 4097	(Ω2bp)	Sp^s		2.2×10^{-7}
spc59	*Acc*I 4097	(Ω2bp)	Sp^s		1.1×10^{-7}
spc37	*Pst*I 4259	(Δ4bp)	Sp^s		1.9×10^{-7}
spc59	*Pst*I 4259	(Δ4bp)	Sp^s		3.1×10^{-7}
ermA33	*Sph*I 4627	(Δ4bp)	Em^s		$<7.5 \times 10^{-9}$
ermA56	*Sph*I 4627	(Δ4bp)	Em^s		$<3.6 \times 10^{-9}$
ermA18	*Cla*I 4706	(Ω2bp)	Em^s		$<7.7 \times 10^{-9}$
ermA23	*Cla*I 4706	(Ω2bp)	Em^s		$<5.8 \times 10^{-9}$
ermA450	*Acc*I 4773	(Ω2bp)	Em^s		$<3.4 \times 10^{-9}$
EcoRV81	*Eco*RV 5915	(Δ)	tnp^+	N.T.	
Ava34	*Ava*I 5921	(Ω4bp)	tnp^+	N.T.	
Ava57	*Ava*I 5921	Ω4bp	tnp^+	N.T.	
BstN4	*Bst*NI 6401	(Δ1bp)	tnp^+	2.5×10^{-4}	
RN2864	(Wild-type Tn554)		tnp^+	6.2×10^{-4}	
RN451	(No Tn554)				1.2×10^{-7} (Sp) $<7.9 \times 10^{-9}$ (Em)

[a]Changes in parentheses are predicted from the type of restriction site ends; others were sequenced. Δ, deletion; Ω, insertion.
[b]Transposition frequency expressed as Em^rCm^r transductants per plaque-forming unit. N.T. not tested.
[c]*tnpA27* is indistinguishable from wild-type in the screening assay.

by complementation analysis (Bastos and Murphy, in preparation) and in two cases the genes involved were sequenced in their entirety and found to have single base changes unrelated to the targeted restriction site (Table III). One of these, *tnpB31*, is a mis-sense mutation with a transduction frequency ~10-fold higher than the frameshift mutants; in this case transposition to the correct location was observed (not shown).

Antibiotic-sensitive mutants. Mutants at the *Acc*I site (bp 4097) and *Pst*I sites (4259) abolished resistance to spectinomycin; those located at the *Sph*I (4627), *Cla*I (4706) or *Acc*I (4773) were Em^s. The frequencies of reversion of Sp^r or Em^r, respectively, were not significantly different from those of RN451, a control strain lacking any Tn554 or plasmid sequences (Table III). These results are fully consistent with the placement of the *spc* and *ermA* genes based on the sequences and on previous genetic considerations (Murphy, 1983), and additionally, the inability of these mutants to revert to antibiotic resistance bears out the expection that the method used here should generate frameshift mutations.

Silent mutations. Mutations at four sites [*Acc*I (3503), *Eco*RV (5915), *Ava*I (5929) and *Bst*NI (6401)] displayed no detectable phenotype. The first of these lies in the intergenic region between *tnpC* and *spc*, but within the 5′ untranslated region of a *spc* transcript (Murphy, 1985b). The remainder are located within ORF. All were scored as $Em^rSp^rtnp^+$; representative mutants were analyzed by blot hybridization and were found to have inserted at the unique chromosomal insertion site in the proper orientation (Figure 4). One of the *Ava*I mutants, Ava57, was sequenced to verify that the predicted frameshift mutation was actually present.

Analysis of transcripts

RNA prepared from cells with (RN4689) or without (RN4565) Tn554 was subjected to Northern analysis using as probes isolated fragments of DNA prepared from different regions of the transposon. The most abundant transcript specified by Tn554 was a 900 base mRNA that hybridized to a 486 bp probe that lies completely within the predicted ORF reading frame (Figure 5).

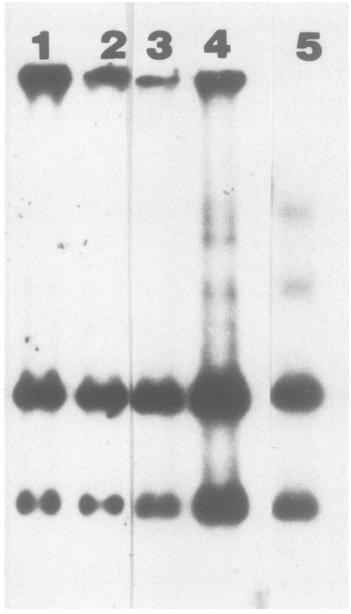

Fig. 4. Southern blot hybridization following transposition of wild-type and mutant Tn554 elements to the chromosome. Chromosomal DNA (1 – 2 μg) was digested with *Hind*III and applied to an 0.7% agarose gel, transferred to nitrocellulose (Southern, 1975) and hybridized to ^{32}P-labeled pRN4175 (pII147::Tn554) DNA at 57°C overnight. Filter washes were twice in 2 × SSC, 0.1% SDS at 37°C and twice in 0.1 × SSC at 57°C. **Lane 1**, RN2864 (wild-type); **lanes 2 – 4**, ORF mutants EcoRV81, Ava57 and BstN4, respectively; **lane 5**, *tnpA27*.

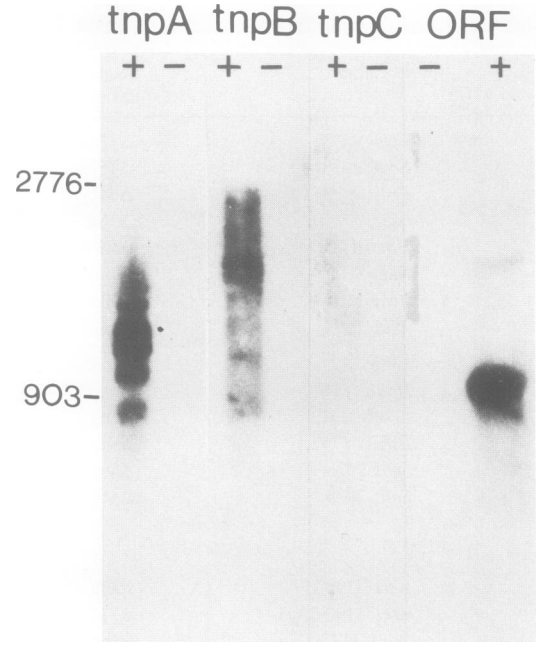

Fig. 5. Analysis of Tn554-specific transcripts. RNA (10 μg) from RN4689 (containing Tn554, lanes labeled '+') or RN4565 (no Tn554, '−') was displayed on a 1.5% agarose-formaldehyde gel (Lehrach *et al.*, 1977), transferred to nitrocellulose strips and hybridized with 4 – 8 × 10^6 c.p.m. of nick-translated probes. Probes were isolated fragments prepared from Tn554-containing plasmids: *tnpA*, nt 381 – 1063 (*Eco*RV-*Fnu*4HI); *tnpB*, nt 1385 – 2597 (*Hinf*I-*Hinf*I); *tnpC*, nt 3190 – 3471 (*Acc*I-*Bgl*II); ORF, nt 5915 – 6401 (*Eco*RV-*Bst*NI). Hybridization was in 40% formamide, 5 × SSC, 50 mM $NaPO_4$, 0.1% SDS, 5 × Denhardt's solution, 1 mM EDTA at 51°C (T_m – 15°) for 18 h; the filters were washed twice in 2 × SSC, 0.1% SDS at room temperature and twice in 0.1 × SSC at 55°C (T_m – 12°C). Standards indicated on the left are restriction fragments of pT181 DNA.

S1 nuclease digestion according to the Berk and Sharp (1977) procedure using as probe a 273-base 5′ end-labeled, strand-separated fragment (*Fnu*4HI, 5643 to *Eco*RV, 5915) yielded a protected fragment of 207 – 209 bp, placing the start of the transcript at 5709 – 5711 (Figure 6). Two other abundant transcripts corresponding to *spc* and *ermA* have been previously defined (Murphy, 1985a, 1985b).

Two major transcripts of ~1150 and 1250 bases were detected using a probe specific for *tnpA*. In addition, minor transcripts of 800, 1000 and 1425 bases were observed in variable amounts (Figure 5). Specific bands corresponding to *tnpB* and *tnpC* were not detected, but a diffuse signal was always observed for both genes. This signal was specific for RNA prepared from Tn554-containing strains, however, and was never detected in RNA from RN4565. No transcript large enough to encode all three *tnp* genes was observed.

In vitro protein synthesis

The proteins specified by Tn554 in an *in vitro* transcription-translation system from *Bacillus subtilis* are shown in Figure 7. Bands of ~43, 30 and 29 kd, corresponding to the products of *tnpA*, *ermA* and *spc*, respectively, were identified by their absence when a corresponding frameshift mutant was used as template. No other Tn554-specified peptides could be identified. Truncated peptides of the expected sizes were observed for the mutants *ermA450* (19.8 kd, lanes 3 and 5) and *spc23* (17.6-kd, lane 4) but not for *tnpA44* (lane 11). The wild-type *ermA* product co-migrates with ErmC, and slightly more slowly than the *spc* product although the latter is predicted to be slightly larger. Both the wild-type and the truncated *ermA* products synthesized *in vitro* cross-react with antibody directed against ErmC (C.Denoya and D.Dubnau, personal communication). The *spc* product, AAD(9), was most clearly visible in those lanes containing little or no wild-type ErmA.

pEM9631 and its mutant derivatives (Tn554 inserted in the 2.8-kb target fragment) directed the synthesis of large amounts of a 22-kd peptide, which was not present in Tn554-containing derivatives of pEM9700 (171-bp target insert) such as RN4934 (pEM9717, lane 1), indicating that the peptide is not a Tn554 product. Plasmid pEM9634, containing the uninterrupted 2.8-kb insert (lane 6), but neither pEM9700 nor pT181 (not shown), synthesized a 25-kd peptide. It is likely that the 25-kd peptide is encoded by the 2.8-kb fragment cloned on pEM9634, and that transposition of Tn554 into this coding region produces a truncated, 22-kd peptide. The successful construction of a host with a 900 bp, chromosomal deletion encompassing this region, and the ability of Tn554 to transpose in this background (into the 171-bp cloned target site) indicates that the intact 25-kd peptide is required neither for cell viability nor for transposition.

Discussion

The nucleotide sequence data and mutational analysis presented here imply that Tn554 specifies three functions, designated *tnpA*, *tnpB* and *tnpC*, that are required for its transposition, in addition to the two antibiotic resistance determinants *ermA* and *spc*. A detailed genetic analysis of the former will be published elsewhere. Proteins corresponding to *tnpA*, *ermA* and *spc* were identified in a *Bacillus subtilis* S-30 system; a 25 kd peptide, specified by the chromosomal target DNA and insertionally inactivated upon Tn554 transposition, was also observed. Our

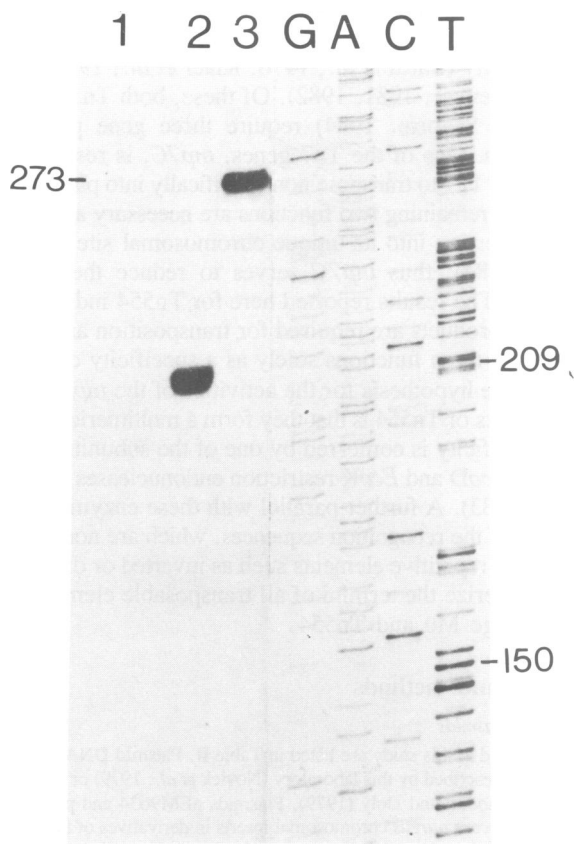

Fig. 6. Determination of the 5′ end of ORF-mRNA by S1 nuclease digestion (Berk and Sharp, 1977). RNA was hybridized in 80% formamide at 42°C with 5′ end-labeled, single-stranded probe (nt 5643, *Fnu*4HI to 5915, *Eco*RV), and diluted into S1 buffer containing 500 u/ml endonuclease (Berk and Sharp, 1977; Murphy, 1985a). **Lane 1**, tRNA; **lane 2**, 50 μg RN4689 RNA; **lane 3**, RN4689 RNA, no S1; GACT, M13 sequencing ladder of an unrelated DNA.

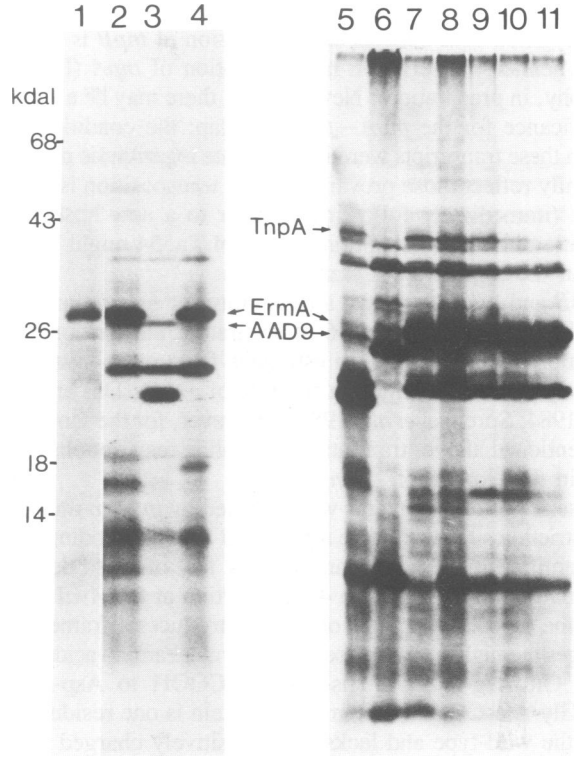

Fig. 7. SDS-polyacrylamide gel electrophoresis of Tn554 peptides synthesized *in vitro*. **Lane 1**, RN4934 (pEM9717); **lane 2**, Ava57; **lane 3**, *ermA450*; **lane 4**, *spc23;* **lane 5**; *ermA450;* **lane 6**, pEM9634; **lane 7**, Ava 57; **lane 8**, *tnpC45;* **lane 9**, *spc23;* **lane 10**, *tnpB3;* **lane 11**, *tnpA44.* Locations of ^{14}C-labeled protein standards (BRL) are indicated on the left. All the Tn554 elements used in these experiments carry the *ermI1* mutation and express *ermA* constitutively.

failure to detect specific transcripts or peptides corresponding to *tnpB* and *tnpC* is most likely a reflection of a low level of expression of transposases in general rather than an indication that these reading frames do not encode products, since mutations in *tnpB* and *tnpC*, as well as in *tnpA*, abolish Tn554 transposition. Application of the Fickett (1982) test for coding *versus* noncoding regions gave a 'noncoding' verdict for ORF, 'coding' for *spc*, and 'no opinion' for *ermA*, *tnpA*, *tnpB* and *tnpC*. The 'RNY' test (Shepherd, 1983) predicts the proposed reading frames in all six cases, but with relatively weak periodicity; stronger periodicity is detected in some non-coding regions. These statistical tests may be less useful for DNA that is very A+T rich, where other constraints are likely to affect base positioning and codon usage.

Comparison of the Tn554 nucleic acid and deduced amino acid sequences to the Genbank and Dayhoff data bases, respectively, using the fastn and fastp homology programs (Wilbur and Lipman, 1983), revealed no significant homologies other than those previously identified for the *ermA* and *spc* determinants (Murphy, 1985a, 1985b). A computer search also failed to locate any region with homology to the helix-turn-helix domain postulated to be common to a variety of DNA binding proteins (Sauer *et al.*, 1982).

Probable translational initiation sites for Tn554 are compared in Figure 3. Gram-positive Shine-Dalgarno sites are characterized by a greater complementarity to the 3′ end of 16S rRNA than those of gram-negative organisms, with an average free energy of −16.7 Kcal/mol, compared with −11.7 Kcal/mol for *E. coli* SD sites (Hager and Rabinowitz, 1985). All of the six major reading frames, and the two regulatory peptides encoded by the 5′ end of the *ermA* mRNA (Murphy, 1985a) have energies of −14.0 Kcal/mol or less, with spacing to the initiation codon of 9−12 bases. Two of these are postulated to initiate with a UUG or GUG codon, which represent about one-third of initiation codons in gram-positive organisms and which are apparently no less efficient than AUG codons (Hager and Rabinowitz, 1985). Although ORF possesses a good SD site by these criteria, the calculation is critically dependent upon the energy assigned to the G-U pair [−1.3 Kcal, according to Salser (1977)]. As noted above, ORF is not translated *in vitro* by *B. subtilis* S-30 extracts although it is abundantly transcribed *in vivo*.

TnpA and *tnpB* exhibit the potential for translational coupling *via* an ATGA overlap of the *tnpA* termination codon and the *tnpB* initiation codon. Such overlaps have been found in a number of operons in which polar effects are observed, and can lead to a re-initiation rate of ~10% in the absence of a functional SD site for the distal gene (Das and Yanofsky, 1984; Sprengel *et al.*, 1985). Even if a SD sequence is present, as it is between *tnpA* and *tnpB*, translational coupling may account for up to 70% of the expression of the distal gene (Das and Yanofsky, 1984). However, the available evidence suggests that *tnpA* and *tnpB* are not translationally coupled: analysis of the mRNA species hybridizing to *tnpA* suggests that the two major transcripts are just long enough to encode *tnpA*; there is no indication of a transcript long enough to include both *tnpA* and *tnpB*. Specific

tnpB and *tnpC* transcripts were not detected. Furthermore, complementation tests suggest that expression of *tnpB* is dependent upon neither transcription nor translation of *tnpA* (Bastos and Murphy, in preparation). Nevertheless, there may be a functional significance for the *tnpA*−*tnpB* overlap; the conditions under which these transcripts were isolated (late logarithmic phase) may not truly reflect those prevailing when transposition is most frequent (immediately following transfer to a new host), and an alteration in the transcription pattern of Tn554 might be important in the regulation of transposition.

tnpB and *tnpC* also have a possibility for translational coupling; in this case there are 6 nt separating the termination and initiation codons. Based on estimates from other systems, this is likely to allow reinitiation at a rate of ~1% (Das and Yanofsky, 1984; Sprengel *et al.*, 1985). However, for the same reasons as mentioned above, translational coupling may not play a major role in the expression of *tnpC*.

A series of *tnp*$^-$ mutants were isolated by *in vitro* mutagenesis at restriction sites within the *tnpA*, *tnpB* and *tnpC* coding regions, as identified by sequence analysis. A few of these deserve individual mention. *tnpA4*, a 4-bp insertion at the *Ava*I site very near the carboxy-terminus of *tnpA*, introduces a frameshift that changes the highly charged seven C-terminal amino acids residues from Glu-Arg-Lys-Glu-His-Lys-Lys-COOH to Asp-Arg-Glu-Lys-Gly-Ala-COOH. The mutant protein is one residue shorter than the wild-type and lacks three positively charged residues; this mutant is totally defective for transposition. Similarly, mutations at the carboxy-terminus of *tnpC* in which the C-terminus is changed from Leu-Tyr-Asn-Lys-Glu-COOH to Arg-Ser-Val-COOH are also *tnp*$^-$. Thus for both proteins, the C-termini appear to be critical for proper functioning. A spontaneous, conservative substitution in *tnpB31* (Arg52→His) results in transposition activity that is 5- to 10-fold greater than the frameshift mutants but still 1000-fold lower than the wild type. *tnpA27* is only partially defective (10%) although it has suffered a deletion of amino acid residues 85−89.

During the course of this study, a number of spontaneous *tnp*$^-$ mutants were isolated. Such mutants were generally found when, in the initial screening, phenotypic analysis preceded restriction analysis. (This was done when the restriction enzyme in question was one that poorly digested DNA prepared by the rapid alkaline procedure.) Thus, *tnpA60, tnpB18* and *tnpB29* were isolated in several initial, unsuccessful attempts to generate a mutation at the *Sph*I site at bp 2925 (although mutations at 4627 in *ermA* were readily obtained in these experiments). We have no explanation for the surprisingly high frequency of spontaneous mutation except to note that it occurred with specific batches of restriction enzymes and may thus have been a result of a contaminating activity.

Most transposable elements encode a single, usually large protein, termed a transposase, that is required for the insertion of that element (Gill *et al.*, 1979; Grindley and Joyce, 1980; Johnson *et al.*, 1982). Bacteriophage Mu replication requires two proteins, gpA and gpB, both of which are now known to be required for the initial conservative integration event following infection (Chaconas *et al.*, 1985) as well as for replicative transposition (Wijffelman *et al.*, 1974; Wijffelman and Lotterman, 1977). The relative amounts of the two proteins, or possibly their availability to interact with host proteins, affects the choice of recombinational pathway (Harshey, 1983); the C-terminal domain of gpB may be a crucial factor in this decision (Chaconas *et al.*, 1985). IS1 transposition also requires two proteins, *insA* and *insB* (Machida *et al.*, 1984), whose individual functions are not known.

Among prokaryotic transposable elements only Tn554, Tn7 and IS4 exhibit a marked site preference as well as an orientation specificity (Barth *et al.*, 1976; Klaer *et al.*, 1981; Lichtenstein and Brenner, 1981, 1982). Of these, both Tn554 and Tn7 (Hauer and Shaprio, 1984) require three gene products for transposition. One of the Tn7 genes, *tnp7C*, is responsible for the ability of Tn7 to transpose non-specifically into plasmid DNAs whereas the remaining two functions are necessary and sufficient for Tn7 insertion into its unique chromosomal site (Hauer and Shapiro, 1984); thus *tnp7C* serves to reduce the insertional specificity. The results reported here for Tn554 indicate that all three gene products are required for transposition and therefore that none of them functions solely as a specificity determinant. An attractive hypothesis for the activities of the *tnpA*, *tnpB* and *tnpC* products of Tn554 is that they form a multimeric transposase whose specificity is conferred by one of the subunits, similar to the *Eco*B, *Eco*D and *Eco*K restriction endonucleases (Gough and Murray, 1983). A further parallel with these enzymes concerns the nature of the recognition sequences, which are non-symmetric and lack the repetitive elements such as inverted or direct repeats that characterize the termini of all transposable elements except bacteriophage Mu and Tn554.

Materials and methods

Strains and plasmids

The strains used in this study are listed in Table II. Plasmid DNA was prepared as previously described by this laboratory (Novick *et al.*, 1979) or by the alkaline method of Birnboim and Doly (1979). Plasmids pEM9634 and pEM9700 contain *Staphylococcus aureus* chromosomal inserts in derivatives of the tetracycline (Tc) resistance plasmid pT181 (Khan and Novick, 1983) of 2.8 kb and 171 bp, respectively; these inserts contain the unique insertion site of Tn554. The host strain RN4470 is a construct lacking 900 bp surrounding the chromosomal insertion site and carrying in its place a non-homologous, 1.4 kb fragment encoding chloramphenicol (Cm) resistance (Murphy and Löfdahl, 1984). Tn554-containing plasmids were derived by transduction of one of these strains with a ϕ11 lysate prepared on a Tn554-containing strain, with selection for erythromycin (Em) or spectinomycin (Sp) resistance, as appropriate. Plasmid pEM9634 and its Tn554-containing derivatives carry a thermosensitive *repC* allele (Iordanescu and Surdeanu, 1980); consequently they do not replicate at 43°C.

Nucleotide sequencing

Sequences were obtained by the dideoxy chain termination method (Sanger *et al.*, 1977) using restriction fragments cloned into the polylinker restriction sites of the M13 derivatives mp10 and mp11, and *E. coli* JM105 (Messing, 1983). Most of the sequence was obtained using *Mbo*I, *Hinf*I and *Taq*I fragments randomly cloned from plasmids containing all or part of Tn554; when necessary, specific clones were identified by plaque hybridization (Messing, 1983). Buffer gradient gels (8% polyacrylamide) and [^{35}S]dATP (New England Nuclear) were used according to Biggin *et al.* (1983).

In vitro mutagenesis

Mutations were isolated by digesting pEM9631 with restriction enzymes for which Tn554 contains one or a few sites, isolating the linear plasmid DNA (in the case where partial digestions were necessary), and filling in or trimming the ends as required prior to re-ligation. Enzyme manipulations were carried out according to Maniatis *et al.* (1982). Thus, most 5′ overhanging ends were filled using Klenow fragment; 3′ overhangs (*Sph*I, *Pst*I) were digested using the 3′→5′ exonuclease activity of T4 DNA polymerase in the absence of deoxynucleotide triphosphates, and flush ends (*Eco*RV, *Hpa*I) were subjected to limited digestion with *Bal*-31 (Legerski *et al.*, 1978) to generate small deletions. The DNA was then ligated and transferred to *S. aureus* by protoplast transformation (Chang and Cohen, 1979; Murphy, 1983) with selection for the vector marker, Tc. Colonies were assayed for transposition (see below) and for Sp and Em resistances, and plasmid DNA was prepared for restriction endonuclease analysis.

Transposition

Transposition proficiency was tested in a qualitative assay by scoring isolated colonies for the retention of Em resistance following three cycles of overnight growth without selection at 43°C, at which temperature the plasmid vector cannot replicate and is lost from the cells. Only those cells containing a *tnp*$^+$ Tn554 element that can transpose to the chromosome will give rise to Em^{r}Tc^s colonies. For each transformant, 40−50 individual colonies were tested; those remaining

Tc^r still contained the vector and were not included in the analysis. Because a single transposition event per colony is sufficient to generate a positive result in this test, Tn554 elements that contain mutations that reduce, but not abolish, transposition will be scored as *tnp*$^+$. For quantitation of transposition frequencies, a chromosomally inserted Tn554 element was used as a donor for a second round of transposition. Chromosomal insertions of *tnp*$^-$ mutants were obtained by complementation. Transducing lysates were prepared and used to infect RN4652; Em^r transductants arising *via* transposition of Tn554 to pEM9700 remain Cm^r and contain an 11.3 kb plasmid, whereas those resulting from chromosomal recombination involving homologous flanking sequences displace the Cm^r marker and contain the original pEM9700 target plasmid (4.6 kb). In some cases the presence and location of Tn554 DNA was assessed by blotting of whole-cell DNA prepared from Em^{r}Cm^r transductant colonies.

Hybridization

Southern and Northern blot hybridizations were carried out as described in Maniatis *et al.* (1982). RNA was prepared by precipitation with guanidine isothiocyanate (Chirgwin *et al.*, 1979) and purified by centrifugation on a 6 M CsCl cushion (Glisin *et al.*, 1974). Nick-translated (Rigby *et al.*, 1977) or 5′-end labeled (Maxam and Gilbert, 1980) probes were prepared from plasmids carrying Tn554 or fragments isolated from acrylamide gels.

In vitro protein synthesis

S-30 extracts were prepared from *Bacillus subtilis* strain BD404, a mutant lacking the ATP-dependent nuclease (Doley *et al.*, 1974), according to the Zubay procedure (Zubay, 1973) as modified for *B. subtilis* (C.Narayanan and D.Dubnau, personal communication). The reaction mixtures (50 μl) contained 8 μg DNA, 1−2 μl S-30 extract, 15.5 units RNasin (Promega), 10 μCi[^{35}S]methionine and other components as described (Zubay, 1973). Samples were precipitated in acetone, resuspended in loading buffer and displayed on a 12.5% SDS-polyacrylamide gel (Laemmli, 1970). Gels were treated with Enlightning (New England Nuclear), dried and exposed at −70°C.

Acknowledgements

We thank C.Narayanan for help with the S30 extracts. This work was supported by grant GM27253 from the National Institutes of Health to E.M. M.B. was supported by the National Council for Scientific and Technological Development (CNPq-Brazil).

References

Barth,P.T., Datta,N., Hedges,R.W. and Grinter,N.J. (1976) *J. Bacteriol.*, **125**, 800-810.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963-3965.
Birnboim,H.C. and Doly,I. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.
Chaconas,G., Giddens,E.B., Miller,J.L. and Gloor,G. (1985) *Cell*, **41**, 857-865.
Chang,S. and Cohen,S.N. (1979) *Mol. Gen. Genet.*, **168**, 111-115.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.
Das,A. and Yanofsky,C. (1984) *Nucleic Acids Res.*, **12**, 4757-4768.
Doley,J., Sasarman,E. and Anagnostopoulos,C. (1974) *Mutat. Res.*, **22**, 15-23.
Fickett,J.W. (1982) *Nucleic Acids Res.*, **10**, 5303-5318.
Gill,R.E., Heffron,F. and Falkow,S. (1979) *Nature*, **282**, 797-801.
Glisin,V., Crkvenjakov,R. and Byus,C. (1974) *Biochemistry (Wash.)*, **13**, 2633-2637.
Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1-19.
Grindley,N.D.F. and Joyce,C.M. (1980) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 125-133.
Gryczan,T.J., Grandi,G., Hahn,J., Grandi,R. and Dubnau,D. (1980) *Nucleic Acids Res.*, **8**, 6081-6097.
Hager,P.W. and Rabinowitz,J.C. (1985) in Dubnau,D. (ed.), *The Molecular Biology of the Bacilli*, Vol. **2**, pp. 1-32.
Harshey,R.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2012-2016.
Hauer,B. and Shapiro,J.A. (1984) *Mol. Gen. Genet.*, **194**, 149-158.
Hollingshead,S. and Vapnek,D. (1985) *Plasmid*, **13**, 17-30.
Iordanescu,S. and Surdeanu,M. (1980) *Plasmid*, **4**, 1-17.
Johnson,R.C., Yin,J.C.P. and Reznikoff,W.S. (1982) *Cell*, **30**, 873-882.
Khan,S.A. and Novick,R.P. (1983) *Plasmid*, **10**, 251-259.
Klaer,R., Kuhn,S., Fritz,H.J., Tillmann,E., Saint-Girons,I., Haberman,P., Pfeifer,D. and Starlinger,P. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 215-224.
Krolewski,J.J., Murphy,E., Novick,R.P. and Rush,M.G. (1981) *J. Mol. Biol.*, **152**, 19-33.
Laemmli,U.K. (1970) *Nature*, **227**, 680-685.
Lai,C. and Weisblum,B. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 856-860.
Legerski,R.J., Hodnett,J.L. and Gray,H.B.,Jr. (1978) *Nucleic Acids Res.*, **5**, 1445-1464.
Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry (Wash.)*, **16**, 4743-4751.
Lichtenstein,C. and Brenner,S. (1981) *Mol. Gen. Genet.*, **183**, 380-387.
Lichtenstein,C. and Brenner,S. (1982) *Nature*, **297**, 601-603.
Machida,Y., Machida,C. and Ohtsubo,E. (1984) *J. Mol. Biol.*, **177**, 229-245.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, NY.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
McLaughlin,J.R., Murray,C.L. and Rabinowitz,J.C. (1981) *J. Biol. Chem.*, **256**, 11283-11291.
Messing,J. (1983) *Methods Enzymol.*, **101**, 10-89.
Murphy,E. (1983) *Plasmid*, **10**, 260-269.
Murphy,E. and Löfdahl,S. (1984) *Nature*, **307**, 292-294.
Murphy,E. (1985a) *J. Bacteriol.*, **162**, 633-640.
Murphy,E. (1985b) *Mol. Gen. Genet.*, **200**, 33-39.
Novick,R.P., Murphy,E., Gryczan,T.J., Baron,E. and Edelman,I. (1979) *Plasmid*, **2**, 109-112.
Philips,S. and Novick,R.P. (1979) *Nature*, **278**, 476-478.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.
Salser,W. (1977) *Cold Spring Harbor Symp. Quant. Biol.*, **42**, 985-1002.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Sauer,R.T., Yocum,R.R., Doolittle,R.F., Lews,M. and Pabo,C.O. (1982) *Nature*, **298**, 447-451.
Shapiro,J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1933-1937.
Shepherd,J.C.W. (1983) *Proc. Natl. Acad. Sci. USA*, **78**, 1596-1600.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342-1346.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Sprengel,R., Reiss,B. and Schaller,H. (1985) *Nucleic Acids Res.*, **13**, 893-909.
Szybalski,W. (1968) *Methods Enzymol.*, **12**, 330-360.
Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature New Biol.*, **246**, 40-41.
Wijffelman,C., Gassler,M., Stevens,W.F. and van de Putte,P. (1974) *Mol. Gen. Genet.*, **131**, 85-96.
Wijffelman,C. and Lotterman,B. (1977) *Mol. Gen. Genet.*, **151**, 169-174.
Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 726-730.
Zubay,G. (1973) *Ann. Rev. Genet.*, **7**, 267-287.

Received on 23 August 1985