The EMBO Journal vol.9 no.3 pp.939-946, 1990

Tandem chromosomal duplications: role of REP sequences in the recombination event at the join-point

Venkatakrishna Shyamala, Erwin Schneider[1] and Giovanna Ferro-Luzzi Ames

Division of Biochemistry and Molecular Biology, University of California, Berkeley, CA 94720, USA

[1]Present address: Universität Osnabrück FB5, Postfach 4469, D-4500, Osnabrück, FRG

Communicated by E.Kellenberger

We show that a family of prokaryotic repetitive sequences, called REP (*r*epetitive *e*xtragenic *p*alindromic), (Stern *et al.*, 1984) is involved in the formation of chromosomal rearrangements such as duplications. The join-points of seven RecA$^+$ tandem duplications previously characterized in *Salmonella typhimurium*, that fuse the *hisD* gene to distant foreign promoters, were cloned and sequenced. In all seven cases they are shown to have originated by recombination between distant REP sequences. Importantly, several join-points had also occurred at REP sequences even in a RecA$^-$ background. Thus, REPs can recombine with each other by a RecA$^-$-independent mechanism involved in the generation of chromosomal rearrangments. While all RecA$^+$ duplications analysed resulted from recombination between REP sequences, some RecA$^-$ duplications did occur also outside of REP sequences, in one case by recombination within a 7 bp homology. Possible roles for the known interaction between DNA gyrase and REP in chromosomal rearrangements are discussed.

Key words: recombination/REP sequences/single specific primer-PCR

Introduction

Chromosomal rearrangements such as deletions, duplications and inversions, are frequent events occurring in the genomes of both prokaryotic and eukaryotic organisms. Some such rearrangements are the result of legitimate recombination, as postulated in the case of some deletion mutations (Albertini *et al.*, 1982). Other events have been ascribed to illegitimate recombination between non-homologous sequences, both in prokaryotes and eukaryotes. Duplication mutations are particularly frequent rearrangements, which are thought to be important early steps in evolutionary processes. Many cases of gene homologies can be ascribed to a sequence of events originating with the duplication of a gene(s) which, by becoming stabilized and therefore redundant, is available for undergoing mutation and evolution. No biochemical mechanism responsible for chromosomal duplications has been uncovered.

We have proposed that prokaryotic REP sequences might be involved in the formation of duplication mutations (Stern *et al.*, 1984; Yang and Ames, 1988). The REP sequences belong to a family of dispersed repetitive DNA which has been found in *Escherichia coli* and *Salmonella typhimurium* (Higgins *et al.*, 1982; Gilson *et al.*, 1984, where they are referred to as palindromic units (P.U.); Gilson *et al.* 1987). A REP sequence is composed of a highly conserved, inverted repeat with the potential to form a stem-loop structure. The entire sequence is about 35 nucleotides long. In most cases REP sequences are clustered in groups of two, three, four or more and the clusters are referred to as REP elements. The REP sequences within an element are always separated by <25 bp and are usually in inverted orientation with respect to each other. We have calculated, by extrapolation, that there are ~ 500 REP sequences per chromosome, organized in 100–200 REP elements and the total amount of DNA occupied by REP sequences is calculated to be $\sim 0.5\%$ of the chromosomal DNA (Stern *et al.*, 1984). Because of the conservation and abundance of these sequences it is reasonable to hypothesize that they perform an important function. Investigations into this aspect have shown that REP sequences do not affect the level of transcription or of transcription termination within an operon (Stern *et al.*, 1984; Gilson *et al.*, 1986b), but can affect translation (i) by stabilizing upstream mRNA against 3′ to 5′ exonuclease attack (Newbury *et al.*, 1987; Stern *et al.*, 1988) and (ii) by interfering with translational efficiency (Stern *et al.*, 1988). However, mRNA stabilization and translational effects do not usually require the maintenance of the strong homology that these sequences display, since any secondary mRNA structure might perform the same function (as, for example, in the case of rho-independent terminators). In contrast, the sequence conservation and the dyad symmetry exhibited by REP sequences suggest that REP might be recognized by a protein, though alternative suggestions have been made (Higgins *et al.*, 1988). That sequence conservation might be the result of recognition by a protein is supported by data from this laboratory showing that the enzyme DNA gyrase specifically binds to REP sequences *in vitro* (Yang and Ames, 1988) and *in vivo* (Yang, Chen and Ames, in preparation). A role for this interaction in chromosomal structure and function, and in the formation of chromosomal rearrangements, has been proposed (Stern *et al.*, 1984; Yang and Ames, 1988, 1990). An unidentified protein was independently shown to interact with REP (Gilson *et al.*, 1986a).

A genetic study of duplications arising within the histidine biosynthetic operon in *S.typhimurium* strains indicated that one end was located very frequently within the intercistronic *his(G-D)* region of this operon, with the other end being located at different distant sites (Anderson and Roth, 1978). Since this intergenic region contains a REP element, we postulated that these particular duplications might have arisen by recombination between this element and distant REP elements within foreign operons (Stern *et al.*, 1984). In this paper we show that this is indeed the case.

Results

Source and genetic nature of duplication mutations

We have utilized duplication-containing strains that had been extensively characterized genetically by Anderson and Roth (1978). In this study, duplication mutations (distinguished by the prefix *pi*) were selected in a histidine auxotrophic strain containing the *Δhis(OG)203*, which deletes the promoter region of the histidine biosynthetic operon, thus inactivating the entire operon. Selection was for growth on the histidine precursor L-histidinol, which can be utilized only if histidinol dehydrogenase, the product of the *hisD* gene, is expressed. Thus, tandem duplication mutations placing the *hisD* gene under the control of a functioning promoter arise as histidinol-utilizing strains. Tandem duplications are characterized by the fusion (the join-point) of the duplicated material to the foreign DNA, which creates a novel DNA sequence; all other DNA sequences are unaltered (Anderson and Roth, 1978). Duplication mutations isolated in a RecA$^+$ background had been grouped into three classes, all of which have one end of the duplicated material in the *his(G-D)* intergenic region. The other end, their join-point, is the distinguishing characteristic: class I has the join-point between 44 and 47 min of the *S.typhimurium* map (Sanderson and Roth, 1988) and includes 27% of the mutants; class II has the join-point between 61.3 and 61.9 min and includes the majority of the mutants (68%); class III contains a single-mutant (5%) with the join-point between 68.2 and 72.8 min. No data was available indicating whether the join-points of individual duplication mutations within each class had occurred by recombination between the same basepairs. Thus, especially within class I, the join-point of which is mapped to within a 3 min interval of the genetic map, the mutations might have entirely different join-points. On the other hand the mutants of class II may have resulted from recombination at a unique, preferred site, since the interval is small and the occurrence of join-points in that area is frequent. Duplications which had originated in a *rec*A strain were also analysed. Figure 1 shows a schematic representation of a duplication mutation with one end in *his(G-D)* and the join-point in a distant region of the chromosome and postulated to contain a REP sequence. Mutations belonging to class I and class II were chosen for analysis of the join-point. TA809 and TA816 were used as control strains not carrying a duplication (Table I).

Strategies for cloning and sequencing the join-points

We set out to prove our hypothesis that the join-point is due to recombination between REP sequences. In order to characterize the join-point it is necessary to isolate a DNA fragment containing the join-point and sequence it. We devised two strategies for this purpose. One strategy identifies the DNA region enclosing the join-point by taking advantage of the fact that this region is a composite of unknown and known (*hisD* gene) DNA, thus resulting in a new chromosomal restriction fragment as compared with the parent strain. Restriction analysis, Southern transfer and hybridization of chromosomal DNA from both parent and duplication strains, using sequences within the *hisD* gene as a probe, should result in a pattern which differs in one fragment only. An even simpler pattern would be obtained if the join-point is introduced into a strain deleted for the entire intergenic region and *hisD* (Anderson and Roth, 1978); in this case the join-point fragment would be the only one

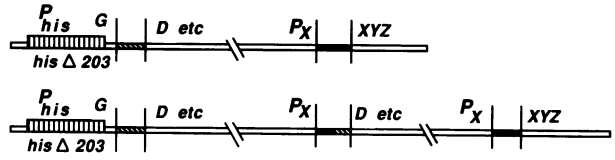

Fig. 1. Schematic representation of a duplication mutation. The top line represents the parental strain, carrying a deletion (*hisΔ203*) of the histidine promoter. The bottom line represents a duplication mutation resulting from recombination between two REP sequences (▥ and ■). The hybrid REP sequence is at the join-point. P*his* and P*x* depict the promoters for the histidine biosynthetic operon and an unknown distant operon respectively. The broken line depicts an unknown length of chromosome.

Table I. Bacterial strains

Strain[a]	Duplication and class	Relevant markers[b]
LT2	wild type	
TT317		*purF1741*::Tn*10*
TA809		*Δhis63*
TA816		*Δhis203*
TR5132	*pi-404*, I	*Δhis63*
TR5142	*pi-414*, I	*Δhis63*
TR5149	*pi-421*, IA[c]	*Δhis63*
TR4879[d]	*pi-401*, II	*Δhis203*
TR5129[d]	*pi-401*, II	*Δhis63*
TR5130	*pi-402*, II	*Δhis63*
TR5131	*pi-403*, II	*Δhis63*
TR5133	*pi-408*, II	*Δhis63*
TR5236	*pi-422*, III[e]	*Δhis63*
TR5237	*pi-423*, IV[e]	*Δhis63*
TR5239	*pi-425*, IV[e]	*Δhis63*
TR5246	*pi-432*, IV[e]	*Δhis63*
TR5242	*pi-428*, V[e]	*Δhis63*
TR5243	*pi-429*, VI[e]	*Δhis63*

[a]TA809 and TA816 are lab stocks. All other strains were obtained from J.Roth's lab and have been characterized genetically (Anderson and Roth, 1978).
[b]*Δhis63* extents through genes *his(OGDC)*; *Δhis203* extends through genes *his(OG)*.
[c]This duplication has been found to extend at least as far as *purF*, contrary to its published classification (see text and Materials and methods).
[d]Duplication *pi-401* was analysed in two different parental backgrounds and by both procedures described in Materials and methods.
[e]These duplications were originally obtained in a parental strain carrying a *recA1* mutation and have been placed in a separate classification (Anderson and Roth, 1978). The duplication mutation was subsequently transferred into a *recA*$^+$ background.

detected by the probe and would be present only in the duplication strain. The nature of the join-point is determined by cloning and sequencing the fragment, starting from the known DNA sequence (the *hisD* gene) and proceeding into the unknown region, thus spanning the join-point. The join-points of *pi-408*, *pi-414* and *pi-421* were identified in this manner. Chromosomal DNA was prepared from duplication strains TR4879, TR5142, and TR5149, and from parental strains TA809 and TA816 (Table I) and the join-point-containing fragment identified as described in Materials and methods. Figure 2 shows the results obtained by digestion with appropriate enzymes and hybridization with a probe containing the *hisD* gene (pAQ1), revealing fragments that are not present in the parent strains. Strain TA809 is deleted for the entire portion of the histidine operon that is carried by pAQ1; thus, its DNA does not hybridize to pAQ1. Only

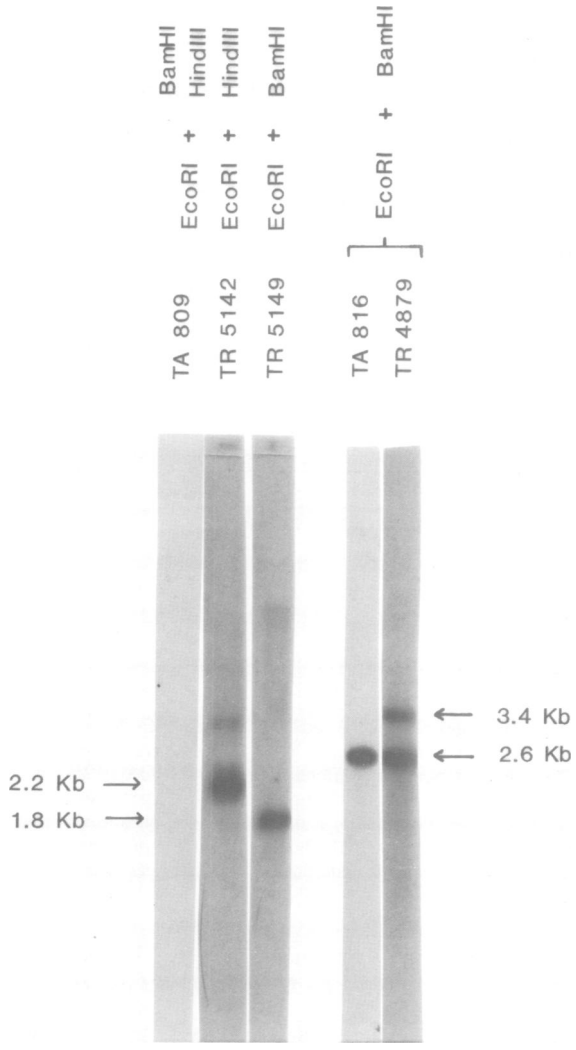

Fig. 2. Identification of join-point fragments. Chromosomal DNA (2 μg) was digested with *Eco*RI (which has a site at the amino terminal end of *hisC* and none within the *hisD* gene), and either *Bam*HI or *Hin*dIII, electrophoresed on 1% agarose gels, transferred to nitrocellulose membranes and hybridized to nick-translated pAQ1 DNA according to standard procedures.

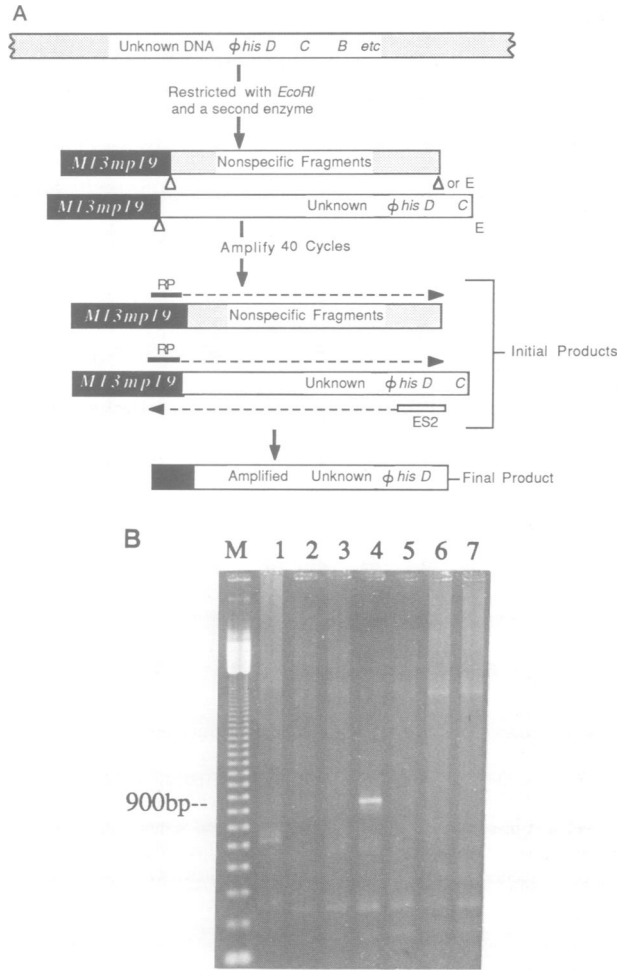

Fig. 3. Amplification of join-point fragment by SSP-PCR. **(A)** Schematic representation of SSP-PCR as applied to duplication mutants. The shaded boxes represent genomic DNA which is mostly nonspecific DNA. The open box represents the specific fragment with the join-point and neighboring DNA. The solid box represents M13mp19 vector DNA. Symbols: E, **Eco**RI site: △, any second enzyme site; ϕ, join-point; ■, M13 reverse sequencing primer (RP); □, *hisD*-specific primer (ES2). **(B)** Amplification of DNA from duplication mutant TR5243 (*pi-429*) by SSP-PCR. TR5243 chromosomal DNA was double-digested with (**lanes**): **1**, *Eco*RI + *Ava*I; **2**, *Eco*RI + *Bam*HI; **3**, *Eco*RI + *Bgl*II; **4**, *Eco*RI + *Hin*dIII; **5**, *Eco*RI + *Pst*I; **6**, *Eco*RI + *Sal*I; **7**, *Eco*RI + *Xho*I, **Lane M**, 123 bp ladder marker. Ligation was into suitably digested M13mp19. Each ligation mix was amplified, the products electrophoresed, and stained with ethidium bromide.

the fragments carrying the join-points in TR5142 (2.2 kb) and TR5149 (1.8 kb) can hybridize to pAQ1. A fragment originating from the portion of the histidine biosynthetic operon carried by pAQ1 lights up in both the parental (TR816) and the duplication (TR4879) strains (2.6 kb), while the duplication strain displays one additional fragment that contains the join-point (3.4 kb). In each case, the join-point-containing fragment was cloned into M13mp19 as a vector as described in Materials and methods. Sequencing of the foreign DNA is achieved by dideoxy sequencing initiated with a primer hybridizing within the amino terminal end of the *hisD* gene. The sequence of these three duplication join-points clearly indicated that they were all the result of recombination between REP sequences (see Figure 4, below).

A second strategy was developed which allows the rapid handling of numerous mutants and which utilizes a modification of the newly-developed method of gene amplification, the polymerase chain reaction (PCR). The modification, single specific primer-PCR (SSP-PCR), was developed in this laboratory to permit sequencing from a known region of DNA into unknown neighboring regions and is thus ideal for characterizing duplication join-points (Shyamala and Ames, 1989b). Figure 3A is a schematic representation of the amplification of a join-point by this procedure; details appear in Materials and methods. In brief, chromosomal DNA is digested with appropriate restriction enzymes, the fragments generated are ligated into a known vector and then amplified by PCR using a primer (ES2) that hybridizes specifically to the known DNA (*hisD*, in this case), and a primer that hybridizes to the vector, M13mp19 (RP). Amplification can occur only in those ligated fragments that span the join-point. The amplified fragment will consist of a small portion of *hisD* DNA and of all of the unknown DNA up to the ligated site, plus a small portion of the vector. The join-point in the amplified fragment is then sequenced

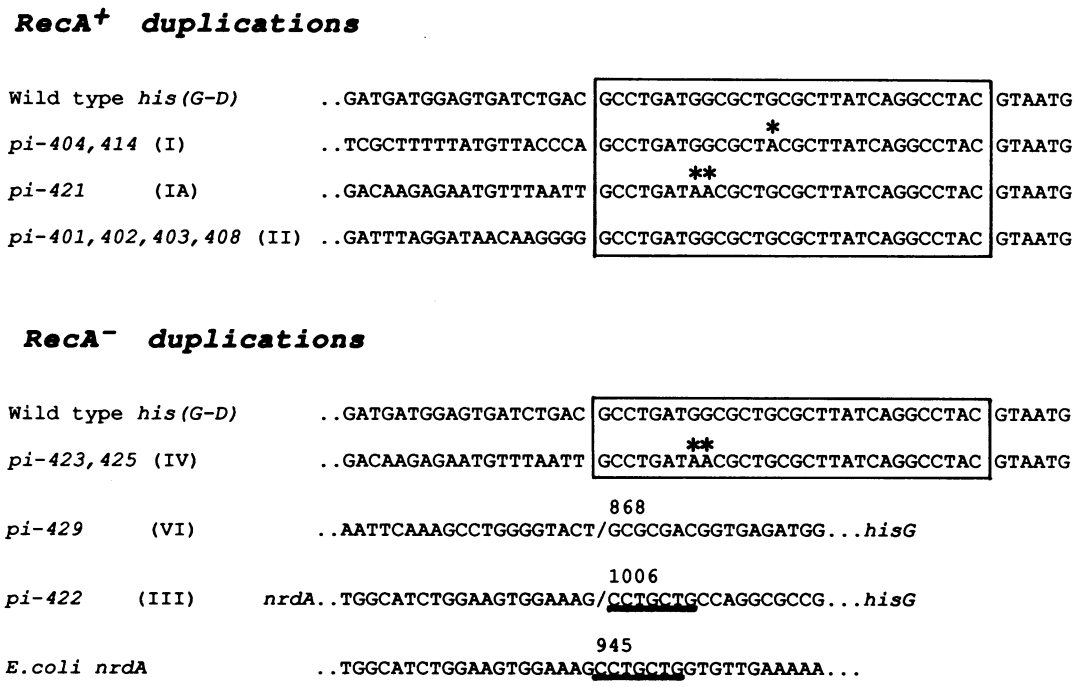

Fig. 4. Sequences of join-point of duplication mutations. Join-point sequences are aligned against part of the sequence of the *his(G-D)* intergenic region of the wild type. The 3′ end of the wild type and of the duplication sequences leads into the *hisD* gene. The roman numerals in parentheses refer to the classification of the duplications (Anderson and Roth, 1978). The REP sequence is boxed in; differences with the wild type REP sequence are marked with asterisks (*). The wild type sequence was obtained from an M13mp19-clone harboring the *hisOGDC'* genes from plasmid pAQ1 and also by asymmetric amplification and sequencing of LT2 DNA with *hisG–hisD* specific primers; it matches the sequence as published by Carlomagno *et al.* (1988). In *pi-422* the 7 bp underlined in *hisG* have a homologous counterpart in *nrdA*. The bottom line is the published *nrdA E.coli* sequence in the region corresponding to the join-point of *pi-422*. The *nrdA* bp to the left of the 7 underlined bp are identical to those in *pi-422*.

using a *hisD*-specific primer close to the join-point. Figure 3B shows the results of the SSP-PCR amplification experiment performed on DNA from strain TR5243. Genomic DNA was digested with *Eco*RI (which has a site in *hisC*) and aliquots of this reaction mix were further digested with each of seven additional restriction enzymes. Visualization by ethidium bromide revealed the presence of a 900 bp fragment only in the *Eco*RI–*Hin*dIII digestion reaction. Presumably, the other reactions produce fragments that are too large to be amplified efficiently. The 900 bp fragment was cloned into M13mp19 and sequenced using a *hisD*-specific primer. The sequence obtained confirmed that amplification had indeed produced a fragment that included *hisD* and proximal foreign DNA and therefore it also included the join-point. The sequence data are all discussed together later (Figure 4). Similar SSP-PCR amplification experiments were performed on all the other mutants listed in Table I (data not shown). In some cases, when the nature of the amplified fragment was in question, the SSP-PCR reaction products were analysed by Southern hybridization with a *hisD*-specific internal oligonucleotide as a probe. In several cases, when the amplified fragments were <400 bp, they were not subcloned in M13 vectors, but rather they were sequenced directly following asymmetric amplification, which produces excess single stranded DNA (Gyllensten and Erlich, 1988; Shyamala and Ames, 1989a).

Duplications generated in recA+ background

Class I duplications. The sequences of all join-points are shown in Figure 4, together with the sequence of the *his(G-D)* intergenic region from the wild-type. The sequences of the join-points of *pi-404, pi-414*, and *pi-421* clearly indicate that a recombination has occurred at the REP element in *his(G-D)*, since the sequence obtained at the 5′ end of the *hisD* gene is identical up to the REP sequence and then continues into a foreign, entirely new sequence. The foreign sequence is the same for *pi-404* and *pi-414*, indicating that for these two strains the recombination has occurred at the same site in these two independent events. In the case of *pi-421*, which has a different foreign sequence, we have ascertained that its join-point does not map to the region around 45 min, as reported, since it has at least two copies of *purF* (located at 47 min; see Materials and methods); thus, it does not belong to class I. We included it in this section because of its initial classification (see later for possible reasons for this discrepancy). We changed its classification to class IA. All three events resulted from a recombination between the *his(G-D)* REP and a foreign REP, as shown by the fact that both REP sequences present at the duplication join-point differ from the present in wild type *his(G-D)*, by one and two basepairs, respectively. This finding excludes the possibility that the event involved a recombination between the extreme 5′ end of the *his(G-D)* REP and any other non-REP sequence elsewhere.

Class II duplications. Four independent duplications were analysed from class II: *pi-401, pi-402, pi-403*, and *pi-408*. Again it is clear that recombination occurred either within REP or at its extreme 5′ end, since a foreign sequence starts immediately 5′ to a REP that is identical to that of the wild type *his(G-D)*. The foreign sequence is the same for all four strains. Thus, many, if not all class II duplications, are the

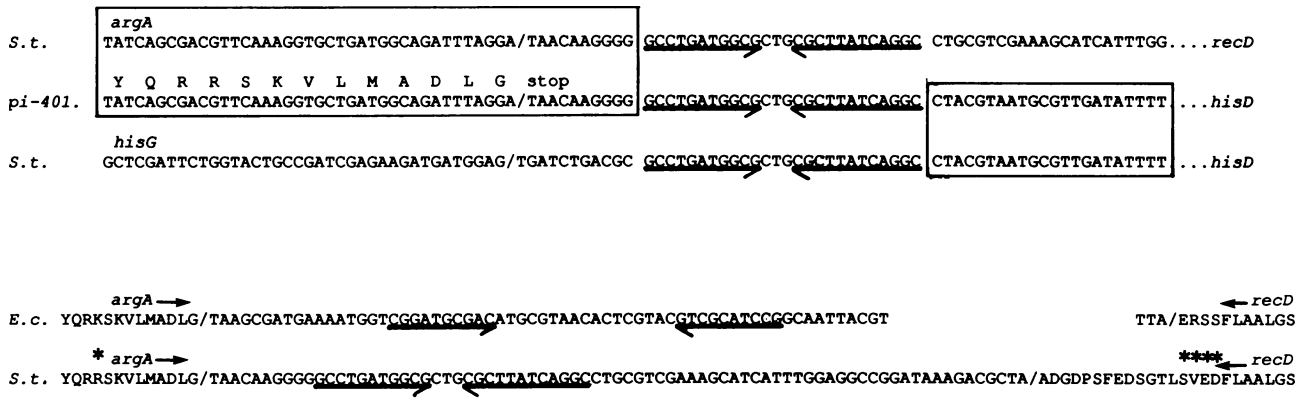

Fig. 5. The join-point of *pi-401* results from recombination between *argA* and *his(G-D)* REPs. In *pi-401* the 5′ and 3′ regions are derived from *argA* and *his(G-D)*, respectively (boxed in). REP elements are marked by arrows. The slashes indicate translation termination. Amino acids are indicated in the second, fourth, and fifth lines, by the single letter code. The *E.coli argA−recD* sequence can be aligned with the *S.typhimurium* sequence with a 49 bp gap (left blank). Asterisks indicate mismatches between *E.coli* and *S.typhimurium*.

result of events at the same site. However, since the REP sequence in these duplication strains is identical to the *his(G-D)* REP sequence, we cannot distinguish whether the recombination has occurred within a particular REP that has sufficient similarity with the *his(G-D)* REP to result in an identical recombined REP sequence, or whether the recombination occurred between the immediate 5′-end of the *his(G-D)* REP and a particular DNA region that carries no REP. These possibilities could be distinguished by sequencing the wild type chromosomal region into which the recombination has occurred.

Class II duplications result from recombination with an identical REP sequence in the argA operon

Such sequence could be obtained by either of the procedures described here for identifying the join-points of duplications, using the cloned class II join-point DNA as a probe. However, since the class II join-point is located near *argA*, around 60 min on the genetic map, where a large segment of *E.coli* DNA has been sequenced (Finch *et al.*, 1986a,b,c; Brown *et al.*, 1987), we compared both the DNA and translated protein sequences of the foreign *S.typhimurium* DNA in class II duplications with those of these *E.coli* sequences. We observed extensive similarity between the foreign DNA in *pi-401* and the 3′ region of *E.coli argA*. The translated sequences are almost identical (Figure 5, second and fourth lines). Thus, it is clear that this is the *E.coli* equivalent of the location of the class II duplication join-point. If the duplications had arisen by recombination at a REP sequence, inspection of this *E.coli* region should contain a REP sequence downstream from *argA*. This was indeed the case (Finch *et al.*, 1986a; Figure 5, fourth line). Since *E.coli* and *S.typhimurium* sequences differ considerably in non-essential regions, particularly in the location and nature of REP elements (Gilson *et al.*, 1984; Stern *et al.*, 1984), the finding of a REP sequence in this *E.coli* region is not definitive evidence that it participated in the duplication-generating event, in particular because the *E.coli argA* REP bears poor similarity to the *S.typhimurium* REP in *his(G-D)* (Figure 5, fourth and third lines). Therefore, it was necessary to sequence the equivalent *S.typhimurium* region downstream from *argA* to determine whether it contains a REP sequence. It was expected that such REP would bear significant homology to the REP in *his(G-D)*, thus explaining the high frequency of incidence of join-points at that site and the resulting unmodified recombinant REP.

The *S.typhimurium* wild type 3′ region of *argA* was amplified by SSP-PCR using an *argA*-specific primer (as obtained from the sequence of the duplication join-point) and the M13 reverse sequencing primer. The product was sequenced following asymmetric amplification (Figure 5, first line). As expected, a portion of it (*argA*) is identical to the foreign DNA in class II duplications; downstream from the structural gene is an extragenic region containing a REP sequence identical to that located in *his(G-D)* (compare the first, second and third lines), thus demonstrating that the frequent recombination at this site is indeed due to the high level of similarity at the recombining sites.

The translated *S.typhimurium argA* sequence is very similar to the sequence of the *E.coli argA* product (fourth and fifth lines). *S.typhimurium* also has a gene downstream from *argA*, which is the equivalent of the *E.coli recD* (transcribed in the opposite direction from *argA* in both species). Thus, these are analogous regions in the two species. Significantly, the intergenic region between *argA* and *recD*, including the REP sequence, displays a lower level of similarity between the two species.

Physiological evidence that the join-point of class II duplications is downstream of *argA* was obtained by showing that in these strains the *hisD* gene has been placed under arginine control, and, therefore, has become part of the *argA* transcription unit. Arginine, which represses expression of *argA*, inhibited growth of TR5129 on histidinol as a histidine source: 10 μmol of arginine gave a zone of inhibition (turbid) of 4 cm diameter on a Petri plate spread with TR5129 and 0.03 mmol of histidinol (data not shown). This inhibition can be ascribed to a lowering of the level of histidinol dehydrogenase by coregulation with *argA*.

Duplications generated in recA background

Duplication mutations had also been generated in a parental strain defective in recombination (*recA*; Anderson and Roth, 1978). These mutations had been characterized genetically to have one end either in the *his(G-D)* region or immediately outside, in the *hisG* gene. It was of interest to determine whether REP sequences were involved also in those *recA* duplications with one end in *his(G-D)*. We analysed TR5236, TR5237, TR5239, TR5242, and TR5246 (carrying *pi-422*,

pi-423, *pi-425*, *pi-428* and *pi-432*, respectively), which have one end in *his(G-D)*, and TR5243 (*pi-429*), which has one end in *hisG*, as a control. Following SSP-PCR and sequencing, the join-point of *pi-429* was shown indeed to be within *hisG*, at 868 bp (Figure 4). Two mutants, *pi-423* and *pi-425*, yielded a SSP-PCR product of 500 bp with *Bam*HI digested DNA. The sequence was identical for the two strains, and it showed that the join-point had been generated by recombination between the *his(G-D)* REP and a foreign REP. Interestingly the foreign sequence was the same as that of *pi-421* (generated in a *recA*$^+$ background), the join-point of which had been placed around 45 min, i.e. in a different location than that of *pi-423* and *pi-425*, around 48 min. We showed that *pi-421* has a join-point at >47 min (see Materials and methods). Given the known extreme instability of duplication mutations, a possible explanation for this conflicting genetic result is that *pi-421* has undergone further rearrangements after mapping and upon storage. Since *pi-421* has all the properties of a duplication, its analysis in this work is justified despite the uncertainty as to the status of the original isolate.

Conflict between genetic and sequence data

The *his(G-D)* region of TR5236 (*pi-422*), TR5242 (*pi-428*), and TR5246 (*pi-432*) was shown to be unaltered, since these strains yielded an amplification product identical to that from a wild type, with primers ES1 and ES2. The join-point of *pi-422* was analysed further and shown to have occurred by recombination within *hisG*, i.e. at 1006 bp, proving that it had an unaltered *his(G-D)* intergenic region. The foreign sequence at the join-point is very similar (90% homologous out of 118 bp compared) to the *E.coli* gene for ribonucleotide reductase 1 (*nrdA*) (Nilsson *et al.*, 1988). A comparison of the *nrdA* and *hisG* sequences indicated a 7 bp identity, at which site the recombination had occurred. This region is underlined in Figure 4. Duplications *pi-428* and *pi-432* were not analysed further. These discrepancies between map location and sequence are possibly due to later rearrangements, as mentioned above.

Discussion

Chromosomal rearrangements giving rise to gene amplification are important evolutionary events (Ohno, 1970) that can increase gene expression by increasing gene dosage and/or placing genes under control of new promoters and that create a source of redundant DNA potentially utilizable for generating genetic diversity. Methods for the selection and detection of bacterial genetic duplications have been developed (last reviewed by Anderson and Roth, 1977). The versatility of bacteria, which have been shown to accommodate tandem duplications that can involve a region as large as 30% of the chromosome, has allowed the genetic analysis of several classes of duplications (Hill *et al.*, 1977; Anderson and Roth, 1978, 1981; Lehner and Hill, 1980; Hoffman *et al.*, 1983; Lin *et al.*, 1984; Tlsty *et al.*, 1984; Sadosky *et al.*, 1989). Gene amplification has been shown to be an important and general phenomenon also in eukaryotes (Schimke, 1982).

The opportunity to form stable base pairing between homologous regions is thought to be an important feature of the mechanism that gives rise to tandem duplications. Short, randomly occurring homologous DNA sequences have been shown to mediate recombination (Edlund and Normark, 1981; Albertini *et al.*, 1982), and therefore might play an important role in the generation of tandem duplications. Indeed, in one study, the join-point of one duplication mutation has been shown to have arisen by recombination between two identical 12 bp sequences (Edlund and Normark, 1981). A duplication mutation within the histidine transport operon has been shown to have arisen by recombination between two identical 17 bp sequences (Speiser and Ames, in preparation). Our own findings in this paper show that a 7 bp homology (*pi-422*) is sufficient to result in recombination between distant and otherwise unrelated regions of the chromosome. Unequal recombination between multiple copies of a gene can also lead to chromosomal rearrangements, as has been shown in the case of the ribosomal RNA (*rrn*) and the rearrangement hot spots (*rhs*) operons (Lehner and Hill, 1980; Anderson and Roth, 1981; Sadosky *et al.*, 1989). It has been postulated that the RecA system is responsible for the generation of these rearrangements since the duplication mutations arose at lower frequency in *recA*$^-$ strains. Similarly, the resolution of duplication mutations to yield wild type was postulated to be RecA$^-$ dependent. While RecA involvement would not be surprising in cases concerning gene families (*rrn* and *rhs*) which share large segments of fairly well conserved or complete homology, the case of short homologies is puzzling, since available data suggest that the RecA system requires at least 40 bp of homology for function (Smith, 1988).

The finding by Anderson and Roth that the majority of duplications in *Salmonella* strains that fuse the *hisD* gene to a foreign promoter have one end in the intergenic region between the *hisG* and *hisD* genes (Anderson and Roth, 1978), caused the authors to speculate that a sequence in *his(G-D)* might exist that is at least in part homologous to a sequence at the other endpoint of duplications, thereby allowing recombination to occur. The later discovery that a REP element is located in the *his(G-D)* intergenic region (Higgins *et al.*, 1982), gave support for this hypothesis. The results presented here confirm that recombination between REP sequences can indeed result in duplication formation. We have sequenced seven duplication join-point (from class I, IA, and II from *recA*$^+$ strains) and we have shown that all have arisen by recombination between REP sequences. In addition, all four members analysed from class II are identical. Thus, class II may be derived from a unique and frequent event of recombination occurring at a REP sequence immediately downstream of the *argA* gene. The *argA* REP is identical with that of *his(G-D)*, including the usually variable nucleotides in the loop of the REP palindrome, thus resulting in a stretch of 28 bp identical in the two operons. Since this class of isolates is conspicuously absent in a *recA* strain, it has been hypothesized that it was generated by legitimate recombination (Anderson and Roth, 1978). Our results support this hypothesis, since (possibly) all members of class II are identical, having arisen by recombination within a segment of identity sufficiently long to be compatible with present knowledge of RecA requirements.

The two representatives of class I indicate that class I may also be homogeneous. Since class I spans 3 min of the genetic map, it is somewhat surprising that it would include a single join-point. Since several duplications originating in a *recA* background have join-points in the same genetic map region as that of class IA *recA*$^+$ duplications, it is possible that they might all arise by *recA*-independent processes. Indeed,

duplications arose in general in a *recA* strain at 1/6 the frequency of a *recA*$^+$ strain (Anderson and Roth, 1978), indicating that recombination-independent events may be responsible for an important fraction of all duplication events (RecA-dependent events are usually decreased several orders of magnitude in *recA* strains; Smith, 1988). Because of this reasoning and because recombination clearly occurred between non-identical REP sequences in the RecA$^+$ class IA duplication mutation, plus the fact the REP sequences may be too short and not sufficiently homologous to participate in RecA-mediated recombination (Smith, 1988), the possibility should be considered that, even in a RecA$^+$ background, a protein(s), other than the RecA system, that specifically recognizes REP sequences is involved. The fact that duplications *pi-423* and *pi-425* resulting from recombination between non-identical REP sequences were generated in a RecA$^-$ background (two out of five examined) could be taken as supportive evidence. If a system other than RecA were responsible for these events, DNA gyrase should be considered as a possible candidate, since it can bind to REP sequences (Yang and Ames, 1988) and it can cause illegitimate recombination (Naito *et al.*, 1984). It is tempting to speculate that DNA gyrase could bring together distantly-located REP sequences by binding to them individually, followed by protein−protein interaction, thus allowing their recombination. On the other hand, it is also possible that the RecA system can be involved in a site-specific reaction which is dependent on DNA gyrase participation, since the interaction of DNA gyrase with REP is obviously a complex one, involving other proteins as well, such as the histone-like protein, HU (Yang and Ames, 1990). Recombination events with the properties of site-specific recombination have been hypothesized to occur between very short homologies in a *recA* strain (Galas, 1978).

If REP sequences are frequent sites of recombination, we would expect a much larger variety of duplication join-points. The limited number of classes obtained in the Anderson and Roth 1978 study might reflect (i) the need for the proper orientation of the recombining REPs, (ii) the appropriate level of expression of the resulting recombinant operon and (iii) an incomplete spectrum of possible duplication events. It may also indicate that additional factors, such as neighboring sequences, may be important. However, we did not recognize any obvious common sequence pattern in the neighborhood of the join-points.

Clearly, recombination between REP sequences is not the only possible mechanism generating duplications as selected in this system, since end-points at sites within the histidine operon other than the REP sequence do indeed arise in a *recA* background, as shown by genetic mapping (Anderson and Roth, 1978) and by our sequence data. In addition, it should be kept in mind that the duplications chosen for this study belong to a somewhat restricted group since one endpoint must fall within ~1000 bp of the beginning of the *hisD* gene (Anderson and Roth, 1978). Thus, the selection procedure may have resulted in an enrichment of duplication mutations specifically derived by REP−REP recombination. Duplications selected in other systems have been shown to occur by recombination between DNA other than REP sequences.

It should also be noted that while the initial recombination event might occur by one mechanism (e.g. by DNA gyrase action), subsequent amplification events under continued selective pressure, alternating with subsequent reduction in copy number, may be due to a different mechanism(s). Thus, it is not easy to evaluate the separate involvement of potential recombination systems in the various stages of creation and maintenance of a duplication mutation.

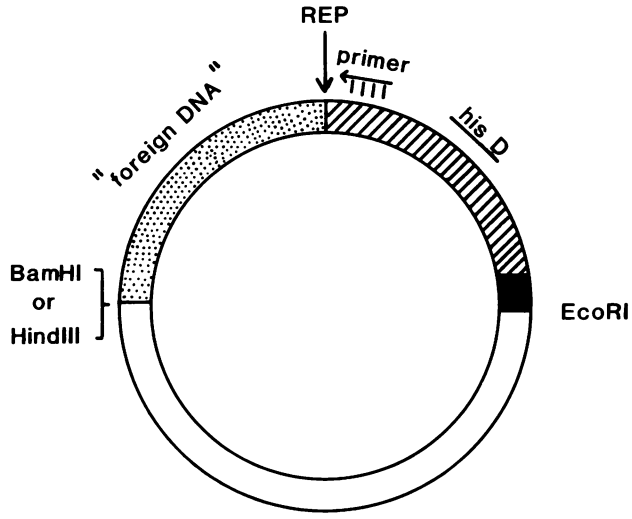

Fig. 6. Schematic representation of clones harboring a duplication join-point. The open box represents M13mp19 RF DNA, which is ligated to the *Eco*RI site in *hisC* (solid box) at one end and to the 'foreign DNA' (dotted box) at the other end. The position and direction of the *hisD*-specific primer used for sequencing is indicated.

Finally, it is particularly intriguing that the eukaryotic enzyme topoisomerase II, which is evolutionary related to prokaryotic DNA gyrase to which it is homologous, also mediates illegitimate recombination in an *in vitro* prokaryotic system (Bae *et al.*, 1988), and it has been implicated in illegitimate recombination *in vivo* at the *MAR* sites ('matrix association regions'; also referred to as *SAR* sites: 'scaffold attached regions'; Gasser and Laemmli, 1986) (Sperry *et al.*, 1989). *MARs* contain topoisomerase II consensus sequences and do indeed interact specifically with this enzyme. In addition, some have been shown to be hot spots of *in vitro* DNA cleavage by topoisomerase II, and *MARs* sites have been localized at sites of chromosomal DNA insertion, deletion, and translocation (Sperry *et al.*, 1989). These data provide support for the hypothesis that topoisomerase II and DNA gyrase are involved in chromosomal rearrangements through their illegitimate recombination activity.

This work is part of our investigation on the function of REP sequences. The finding that they can participate in recombination reactions giving rise at least to duplication mutations, suggests that, whatever their origin and other present day functions, they have also been utilized by the cell to participate in the important mechanism of chromosomal rearrangements. The involvement of DNA gyrase in this phenomenon will be explored with the use of gyrase mutations and inhibitors and with *in vitro* recombination assays.

Materials and methods

DNA preparation, Southern transfer, hybridization and subcloning

Chromosomal DNA was extracted as described (Owen and Borman, 1987), digested with *Eco*RI (which has a site at the amino terminal end of *hisC* and none within the *hisD* gene), and either *Bam*HI or *Hin*dIII, electrophoresed on 1% agarose gels, transferred to nitrocellulose membranes, and hybridized to nick-translated pAQ1 DNA according to standard procedures (Maniatis *et al.*, 1982). Plasmid pAQ1 was used as a probe for detecting join-point fragments; it contains a 3053 bp fragment (*hisOGDC'*) cloned between the

*Pst*I and *Eco*RI sites of pBR322 and it was obtained from B.Ames. In some cases a *hisD*-specific internal oligonucleotide (ES3) was used as the probe.

The portion of the gel containing the join-point fragment was cut out, the DNA isolated with 'Geneclean' (Bio101 Inc., LaJolla, CA), ligated into M13mp19 RF DNA linearized with the same set of restriction enzymes, and transformed into JM103 (Messing, 1983). Positive clones were identified by plaque transfer to nylon membranes (Genescreen, NEN, West Germany) and hybridization using pAQ1 as a probe. Figure 6 shows a schematic representation of the resulting clone.

Polymerase chain reaction

Primers used: ES1=AGCTCGATTCTGGTACTGCCGAT-1136 (T); ES2=TAATACTGTCAGAGGCGGAA-1376 (B); ES3=ATAAGGCG-GAACCCTGTGATG-1247 (T); ES5=TATGAATTC/TACAGAACC-CAAAAT-1241 (B, nucleotides to the left of the slash constitute *Eco*RI site); ES=GTTCAGGGCTACAGCTGTTC-1125 (B); Reverse primer: RP=CAGGAAACAGCTATGACCATGATTA- (T); A1=TATCAGC-GACGTTCAAAGGTGCTGAT (T).

For the conventional polymerase chain reaction, oligomers were designed for the two ends and amplification was carried out as described (Shyamala and Ames, 1989a). To identify the unknown end of a duplication join-point, DNA was amplified by the technique of SSP-PCR (Shyamala and Ames, 1989b). Chromosomal DNA was initially digested with *Eco*RI and then aliquots of this (2.0 μg each) were digested with an array of second enzymes. The doubly-digested DNA was diluted to a concentration of 6 μg/ml for ligation. Vector DNA was suitably digested and used at a concentration of 180 μg/ml in a final volume of 10 μl. Ligation was carried out at 15°C overnight with 5 units of enzyme.

A 1 μl aliquot of the ligation mix was used for amplification in a total volume of 25 μl. The M13 reverse primer (RP) served as the vector-specific generic primer for all amplifications. The second primer specific for *hisD* was ES2 (1376) and for *argA*−*recD* in wild type, *argA* specific primer A1 was used. An aliquot (4 μl) of the amplification mix was electrophoresed on 1% agarose in Tris (0.1 M), borate (0.1 M) EDTA (1 mM), and visualized by staining in ethidium bromide. Any band unique to a digest mixture was cut out and amplified with one primer only to determine specificity. If the amplified DNA was truly the product of both (RP and ES2) primers, depending on the size of the DNA, it was directly sequenced following asymmetric amplification or was sequenced by cloning into M13 vectors (Shyamala and Ames, 1989a).

Sequencing

The sequence of the DNA across the join-point was obtained from the phage clones by the chain termination method using an oligonucleotide primer corresponding to *Salmonella hisD* gene. For dideoxy sequencing according to Sanger *et al.* (1977), a kit from Boehringer, Mannheim (West Germany) was used, replacing dGTP with 7-deaza-dGTP in all reaction mixtures (Barr *et al.*, 1986). The oligonucleotide primer was annealed to the template DNA by incubation at 56°C for five min, and subsequently transferred to 37°C for 15 min. The sequencing reactions were carried out at 37°C and the sequencing gel was run at 60 watts. Alternatively, Sequenase kit was employed (Shyamala and Ames, 1989a).

Strains and growth conditions

Table I lists all strains used. Duplication-containing strains were maintained on VBC minimal medium (Roth, 1970), supplemented with glucose (0.4%) and L-histidinol (1 mM), and subcultured to either exponential or stationary phase in either minimal-histidinol or LB medium (Miller, 1972) immediately prior to DNA extraction. As a test for whether a duplication had included *purF*, TR5132, TR5129, TR5149, TR5237, and TR5239 were transduced to tetracycline resistance with P22 phage prepared on TT317 and the transductants were tested for purine auxotrophy (Anderson and Roth, 1978). TR5129, TR5149, TR5237, and TR5239 gave only purine prototrophs, indicating that they carried two copies of *purF*; TR5132 gave only purine auxotrophs, as expected.

Acknowledgements

We are grateful to J.R.Roth and C.Connor for supplying all of the duplication strains used in this study. We also thank D.Irwin and M.Morgan for many helpful discussions. This work was supported by: National Institutes of Health Grant DK12121 to G.F.-L.A.; American Cancer Society Grant MV-397 to G.F.-L.A.; and a postdoctoral fellowship of the Deutsche Forschungsgemeinschaft to E.S.

References

Albertini,A.M., Hofer,M., Calos,M.P. and Miller,J.H. (1982) *Cell*, **29**, 319−328.

Anderson,R.P. and Roth,J.R. (1977) *Annu. Rev. Microbiol.*, **31**, 473−505.

Anderson,R.P. and Roth,J.R. (1978) *J. Mol. Biol.*, **119**, 147−166.

Anderson,R.P. and Roth,J.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3113−3117.

Bae,Y.-S., Kawasaki,I., Ikeda,H. and Liu,L.F. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2076−2080.

Barr,P.J., Thayer,R.M., Laybourn,P., Najarian,R.C., Seela,F. and Tolan,D.R. (1986) *Biotechniques*, **4**, 428−432.

Brown,K., Finch,P.W., Hickson,I.D. and Emmerson,P. (1987) *Nucleic Acids Res.*, **15**, 10586.

Carlomagno,M.S., Chiariotti,L., Alifano,P., Nappo,A.G. and Bruni,C.B. (1988) *J. Mol. Biol.*, **203**, 585−606.

Edlund,T. and Normark,S. (1981) *Nature*, **292**, 269−271.

Finch,P.W., Storey,A., Brown,K., Hickson,I.D. and Emmerson,P. (1986a) *Nucleic Acids Res.*, **14**, 8583−8594.

Finch,P.W., Storey,A., Chapman,K.E., Brown,K., Hickson,I.D. and Emmerson,P.T. (1986b) *Nucleic Acids Res.*, **14**, 8573−8582.

Finch,P.W., Wilson,R.E., Brown,K., Hickson,I.D., Tomkinson,A.E. and Emmerson,P.T. (1986c) *Nucleic Acids Res.*, **14**, 4437−4451.

Galas,D.J. (1978) *J. Mol. Biol.*, **126**, 847−857.

Gasser,S.M. and Laemmli,U.K. (1986) *Cell*, **46**, 521−530.

Gilson,E., Clément,J.-M., Brutlag,D. and Hofnung,M. (1984) *EMBO J.*, **3**, 1417−1421.

Gilson,E., Perrin,D., Clément,J.-M., Szmelcman,S., Dassa,E. and Hofnung,M. (1986a) *FEBS Lett.*, **206**, 323−328.

Gilson,E., Rosset,J.P., Clément,J.M. and Hofnung,M. (1986b) *Ann. Inst. Pasteur Microbiol.*, **137B**, 259−270.

Gilson,E., Perrin,D., Saurin,W. and Hofnung,M. (1987) *J. Mol. Evol.*, **25**, 371−373.

Gyllensten,V.B. and Erlich,H.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7652−7656.

Higgins,C.F., Ames,G.F.-L., Barnes,W.M., Clément,J.M. and Hofnung,M. (1982) *Nature*, **298**, 760−762.

Higgins,C.F., McLaren,R.S. and Newbury,S.F. (1988) *Gene*, **72**, 3−14.

Hill,C.W., Grafstrom,R.H., Harnish,B.W. and Hillman,B.S. (1977) *J. Mol. Biol.*, **116**, 407−428.

Hoffmann,G.R., Walkowicz,M.J., Mason,J.M. and Atkins,J.F. (1983) *Mol. Gen. Genet.*, **190**, 183−188.

Lehner,A.F. and Hill,C.W. (1980) *J. Bacteriol.*, **143**, 492−498.

Lin,R.-J., Capage,M. and Hill,C.W. (1984) *J. Mol. Biol.*, **177**, 1−18.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Messing,J. (1983) *Methods Enzymol.*, **101**, 10−89.

Miller,J.H. (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. pp. 352−355.

Naito,A., Naito,S. and Ikeda,H. (1984) *Mol. Gen. Genet.*, **193**, 238−243.

Newbury,S.F., Smith,N.H., Robinson,E.C., Hiles,I.D. and Higgins,C.F. (1987) *Cell*, **48**, 297−310.

Nilsson,O., Aberg,A., Lundquist,T. and Sjoberg,B.-M. (1988) *Nucleic Acids Res.*, **16**, 4174.

Ohno,S. (1970) *Evolution by Gene Duplication.* Springer Verlag, New York.

Owen,R.J. and Borman,P. (1987) *Nucleic Acids Res.*, **15**, 3631−3633.

Roth,J. (1970) *Methods Enzymol.*, **17**, 3−35.

Sadosky,A.B., Davidson,A., Lin-R.-J. and Hill,C.W. (1989) *J. Bacteriol.*, **171**, 636−642.

Sanderson,K.E. and Roth,J.R. (1988) *Microbiol. Rev.*, **52**, 485−532.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.

Schimke,R.T. (ed.) (1982) *Gene Amplification.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Shyamala,V. and Ames,G.F.-L. (1989a) *J. Bacteriol.*, **171**, 1602−1608.

Shyamala,V. and Ames,G.F.-L. (1989b) *Gene*, **84**, 1−8.

Smith,G.R. (1988) *Microbiol. Rev.*, **52**, 1−28.

Sperry,A.O., Blasquez,V.C. and Garrard,W.T. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5497−5501.

Stern,M., Ames,G.F.-L., Smith,N.H., Robinson,E.C. and Higgins,C.F. (1984) *Cell*, **37**, 1015−1026.

Stern,M., Prossnitz,E. and Ames,G.F.-L. (1988) *Mol. Microbiol.*, **2**, 141−152.

Straus,D.S. and Straus,L.D'A. (1976) *J. Mol. Biol.*, **103**, 143−153.

Tlsty,T.D., Albertini,A.M. and Miller,J.H. (1984) *Cell*, **37**, 217−224.

Yang,Y. and Ames,G.F.-L. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8850−8854.

Yang,Y. and Ames,G.F.-L. (1990) In Riley,M. and Drlica,K. (eds), *The Bacterial Chromosome.* Am. Soc. Microbiol., Washington DC, pp. 211−225.

Received on November 14, 1989; revised on December 28, 1989