

# TANDEM GENETIC DUPLICATIONS IN PHAGE AND BACTERIA

◆1714

*R. Philip Anderson and John R. Roth*

Department of Biology, University of Utah, Salt Lake City, Utah 84112

---

## CONTENTS

INTRODUCTION .....	473
<i>The Genetic Consequences of Tandem Duplications</i> .....	474
<i>General Methods for Detecting Tandem Duplications</i> .....	477
SELECTION OF TANDEM DUPLICATIONS .....	478
<i>Gene Dosage</i> .....	478
<i>Coinheritance of Allelic Markers</i> .....	485
<i>Operon Fusion</i> .....	496
GENERAL CONCLUSIONS .....	501
<i>Size and Frequency of Tandem Duplications</i> .....	501
<i>Mechanisms of Tandem Duplication</i> .....	501
<i>Uses of Duplications in Genetic Analysis</i> .....	502
<i>Gene Duplication as a Regulatory Mechanism</i> .....	502

## INTRODUCTION

Within the last several years, methods have been developed for detection and analysis of tandem duplications of genetic material in bacteria and phage. Although the mechanisms for formation of these duplications are not yet completely clear, it seems reasonable to review the accumulating literature on this phenomenon. The nature and frequency of duplications raises the possibility that this mode of gene amplification could play a significant role in bacterial adaptation. It has also been suggested that gene duplication plays an important role in evolution (47, 93, 137). The high frequency of duplications helps to explain a variety of past genetic observations involving unstable mutants and recombinants that revert to their original state

with high frequency. Duplications provide a potential method for genetic analysis, including chromosome mapping and complementation testing in bacterial species that lack a conjugation system. This review considers only tandem duplications, since virtually all duplications analyzed in detail in bacteria appear to be of this type.

### *The Genetic Consequences of Tandem Duplications*

Before discussing experimental data, it is important to consider the behavior and formal genetic aspects of tandem duplications. These considerations lead to predictions that are borne out by the experimental data.

**TANDEM DUPLICATIONS GENERALLY CAUSE NO LOSS OF FUNCTION** The reasons tandem duplications generally cause no loss of function can be seen in the diagram of a tandem duplication presented in Figure 1. In the chromosome carrying the duplication, the only impropriety in sequence is the marked point between the two tandemly repeated copies (*i-d*). This impropriety does not lead to a loss of function because proper versions of these sequences are present elsewhere (*c-d* and *i-j*). For this reason, one would expect no loss of function to result from the duplication mutation. Exceptions to this rule are duplications of material within a single gene, which could inactivate that gene, and duplications of several genes within a single operon, which could generate a polar effect at the join point between duplicated segments upon expression of that operon. Thus, the rule should hold as long as relatively large segments of chromosome are duplicated; this is usually found to be the case.

**TANDEM DUPLICATIONS SHARE MANY PROPERTIES WITH DELETIONS** Formally, both duplications and deletions can be generated by illegitimate, unequal recombination (39, 41) between nonhomologous sequences on homologous chromosomes (114). They could also be generated by legitimate, unequal recombination between identical or similar sequences located at distinct points on these chromosomes. This is illustrated in Figure 2 with recombination between daughter chromo-

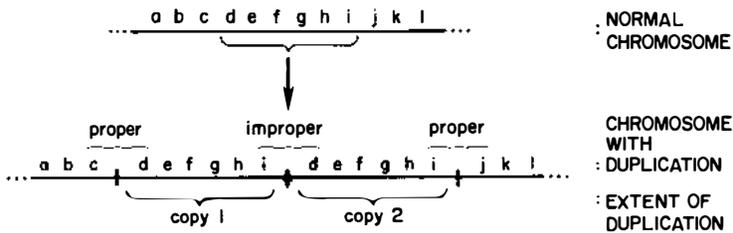


Figure 1 Tandem duplication of the chromosomal segment *defghi*. Lower case letters are nongenetic indications of hypothetical base sequences.

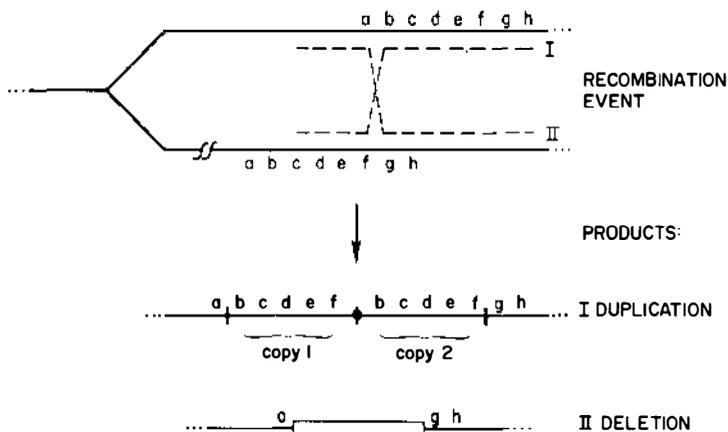


Figure 2 Illegitimate recombination resulting in duplication (I) or deletion (II) of the chromosomal segment *bcdef*. Solid lines indicate double-stranded DNA. Dashed lines represent a reciprocal recombination event.

somes. Reciprocal recombination between the sites *a-b* and *f-g* (which might not involve sequence homology) results in recombinant chromosomes (I) carrying a duplication of the sequence *bcdef* and (II) carrying a deletion of the same sequence. It would not be surprising if at least some duplications and deletions were generated by the same mechanism.

Both duplications and deletions have a single base sequence not found in the chromosome from which they were derived (the "novel joint") (49). In deletions, extensive material has been removed at that point to make the new sequence; in duplications, material has been added at that site to generate the impropriety. These similarities should be kept in mind when considering genetic crosses involving these two mutational types.

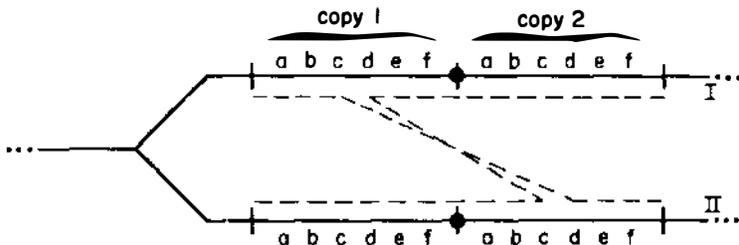
**DUPLICATIONS MAY BE OF UNLIMITED SIZE** Since duplications lead to no loss of function, it seems possible that illegitimate recombination between any two points in the chromosome could give rise to a viable duplication-carrying recombinant. The corresponding deletion mutants would be unlikely to survive if the exchange points were far apart in the chromosome, since essential functions would be lost. The only restriction on the size of a duplication might be the ability of the cell to replicate and segregate this larger chromosome faithfully.

**DUPLICATIONS MAY BE FREQUENT OCCURRENCES** The large number of potential exchanges that yield duplication mutants and the high probability that these mutants will be viable suggest that duplications might be encountered frequently if appropriate conditions were provided for their detection. If duplications and deletions were generated by the same mechanism, one would expect that detectable

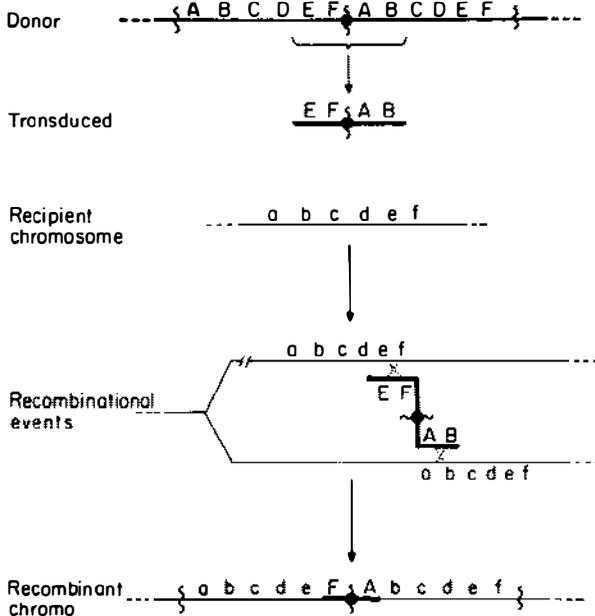
duplications would be much more common than deletions, since only very small deletions would be viable.

**DUPLICATIONS ARE SUBJECT TO FREQUENT LOSS AND FURTHER AMPLIFICATION** The existence of a duplication should be an unstable situation. Normal recombination between the two copies should lead either to loss of the duplication or to further amplification (i.e. triplication) of the chromosomal segment. The events causing this instability are depicted in Figure 3 as recombinational events between nascent daughter chromosomes. Alternatively, loss of the duplication could also be imagined to occur by a reciprocal exchange between the tandemly located copies in a single chromosome. In Figure 3, homologous exchange is seen to give rise to either loss of the duplication (II) or to triplication of the segment (I). Since both of these events involve legitimate recombination between homologous sequences, they might be expected to occur frequently.

**EXTREMELY LARGE DUPLICATIONS CAN BE TRANSDUCED** As mentioned earlier, the only novel base sequence in chromosomes carrying a duplication is the join point between copies of the duplicated segment. At this point, sequences that would be widely separated in a normal chromosome are made contiguous. Transduction of this join point from a donor harboring a duplication into a normal recipient may have surprising results. It is possible to reestablish the donor's duplication state in the recipient cell. This is even possible when the region included in the duplication is much too large to be carried by a transducing phage. Recombination events that account for this behavior are depicted in Figure 4. Once a transduced fragment including the join point enters a recipient cell, recombination events between that fragment and two copies of the homologous region can serve to regenerate the duplication state. In the final recombinant, most of the duplicated material is derived from the recipient chromosome; only material adjacent to the join point is derived from the donor parent. Thus, transduction of large tandem duplications may be detected, provided the selected donor marker and the join point between duplicated material are cotransducible. Transductional events such as those described in Figure 4 were first suggested by Campbell (20). Further evidence in support of these events has been presented (4, 55).



*Figure 3* Legitimate recombination resulting in triplication (I) or haploidization (II) of a tandem duplication.



*Figure 4* A mechanism for transduction of large tandem duplications. Recipient and donor DNAs are light and bold lined, respectively. Dotted lines represent reciprocal recombination events.

### *General Methods for Detecting Tandem Duplications*

Selection schemes designed to detect tandem duplication mutants in a haploid population must capitalize upon some special property conferred by the diploid state. The merodiploid nature of tandem duplications and the sequence impropriety found at the join point between duplicated regions are often traits that lend themselves to selection. All selections described below have utilized these traits in one of three different ways. These represent three basic approaches to the selection and maintenance of tandem duplications.

**GENE DOSAGE** Strains merodiploid for a particular gene should produce approximately twice the amount of that gene product. If such an increase confers a selectable phenotype, strains harboring duplications may be selected from the haploid population. If appropriate selections exist, strains harboring multiple tandem copies may be selected from this merodiploid population. Selections of this general type probably make very few constraints upon the size of duplications generated. Duplication of material adjacent to the gene of interest will likely be of little consequence, provided the extra DNA may be replicated in an orderly fashion. Thus, this selec-

tion method should yield duplications that are the best indicators of the sizes and frequencies of spontaneous duplications.

**COINHERITANCE OF ALLELIC MARKERS** A variety of detection schemes involve genetic crosses that select for simultaneous inheritance of two alleles of a single locus. These alleles may be mutually exclusive (such as two different forms of a single gene product), such that only merodiploid strains may inherit both, or they may be complementing mutations that cannot recombine to produce true wild-type recombinants (such as overlapping complementing deletions). Coinheritance of such allele pairs is possible in heterogenotes merodiploid for the region of interest.

**OPERON FUSION** Analogous to fusion of genes by the deletion of intervening material (85), genes may also be fused by tandem duplication. Figure 5 shows how this may be accomplished. Tandem duplications whose end points lie within different units of transcription lead to fusion of one operon's structural genes to another operon's regulatory elements. An important feature of this selection is that neither operon is destroyed by the tandem duplication, yet a new hybrid one is created. The only requirement is that both operons have the same direction of transcription. Detection methods of this type most frequently select for turn-on of genes whose expression has been shut off, either because of polarity effects or because of inactivation of controlling elements. Unlike methods previously described, selections of this type severely restrict the permissible end points of the duplications obtained. To obtain hybrid gene expression, end points must be near the gene to be expressed and near a functional promoter oriented in the appropriate direction. Thus, sizes and frequencies observed by this method represent only a small subset of the total duplication events that occur.

## SELECTION OF TANDEM DUPLICATIONS

### *Gene Dosage*

**LACTOSE OPERON** Experiments by Novick et al (61, 62, 90, 91) have provided the first well-documented case of a gene dosage selection for tandem duplications. *Escherichia coli* strains capable of high rates of  $\beta$ -galactosidase synthesis (up to 25% of total cellular protein) were isolated following long-term growth selection in a chemostat containing limiting lactose as the sole carbon source. Such "hypersynthesizing" strains synthesize  $\beta$ -galactosidase constitutively at a level fourfold higher than normal strains. Apparently, initial selection in the chemostat yielded *lacI*<sup>-</sup> mutants. These were later replaced by strains harboring multiple tandem copies (up to four) of the constitutive *lac* operon. A number of observations support this conclusion. (a) The hypersynthesizing character is unstable. Segregation of stable haploid strains through loss of the extra gene sets is relatively frequent. Moreover, segregants of intermediate expression (indicating an intermediate number of gene copies) may be obtained. (b) The biochemical properties of  $\beta$ -galactosidase produced by hypersynthesizing strains is qualitatively identical in all respects to normal enzyme. (More, not better, enzyme is being made.) (c) During conjugational crosses,

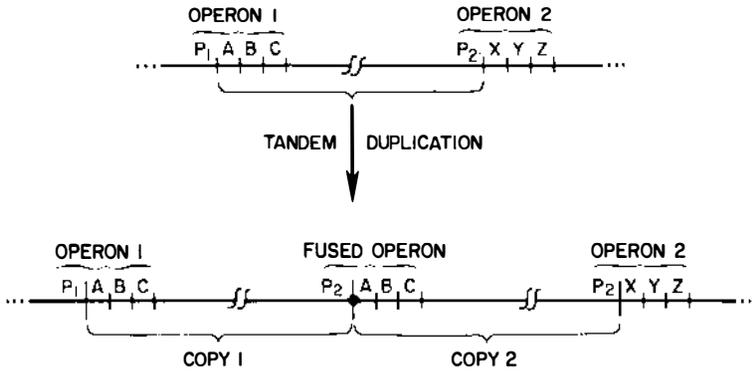


Figure 5 A mechanism for operon fusion based upon tandem duplication of intervening material.

the hypersynthesizing character is transferred at a time expected for *lac* chromosomal genes (i.e. the extra copies are closely linked to *lac*). Also, the "hyper" character is transducible by phage P1. (d) Diploidy for *lac* genes may be demonstrated genetically.

The duplications harbored by these strains are relatively small. Genetic tests indicate that the nearby markers *tsx* and *pro* are not duplicated. Thus, the duplicated material in these strains is less than 3 chromosomal min in length.

The nature of the selection used for isolating these duplications (optimal growth in a chemostat) makes estimation of the duplication frequency difficult. However, it is possible roughly to estimate this frequency in two ways. Data obtained by Horiuchi et al (61) indicate a spontaneous duplication rate of approximately  $10^{-3}$ . While analyzing the recombinants obtained from conjugational crosses between haploid *lac*<sup>+</sup> donors (Hfr *lacI*<sup>-</sup>*Z*<sup>+</sup>*Y*<sup>+</sup>*A*<sup>+</sup>) and haploid *lac*<sup>-</sup> recipients (F<sup>-</sup> *lacI*<sup>+</sup>*Z*<sup>+</sup>*Y*<sup>-</sup>*A*<sup>+</sup>), one unexpected heterogenote was identified among 928 *lac*<sup>+</sup> exconjugants. This strain was concluded to contain a tandem duplication of the *lac* region, each copy of which harbored one parental allele. Since both parents in this cross were haploid, the recovery of 1 duplication per (roughly) 1000 normal haploid recombinants indicates a spontaneous duplication rate (in either the donor or the recipient) of approximately  $10^{-3}$  (see below). The duplicated material in this strain includes both the *tsx* and *pro* loci and, thus, is considerably larger than the duplications studied in the *lac* hypersynthesizing strains.

A second estimate of the *lac* duplication frequency has been obtained more recently by Langridge (73). Sodium lactobionate is not utilized as a carbon source by *E. coli* because it is only poorly bound and hydrolyzed by  $\beta$ -galactosidase. Mutations conferring quantitative increases in  $\beta$ -galactosidase synthesis (probably analogous to hypersynthesizing strains) arise at a spontaneous frequency of approximately  $10^{-4}$ . Most strains isolated in this manner are genetically unstable and hence are likely to be tandem duplications of *lac* genes.

**RIBITOL DEHYDROGENASE** Rigby, Burleigh & Hartley (106) have used a similar selection to isolate *Klebsiella aerogenes* strains, which hypersynthesize ribitol dehydrogenase. Wild-type strains will grow on ribitol as a sole carbon source because of their inducible ribitol permease and ribitol dehydrogenase (RDH) enzymes. Xylitol, which is not normally found in nature, will not support growth of wild-type strains. However, mutants constitutive for RDH will grow slowly on this unnatural pentitol. Metabolism of xylitol by these strains is due to the slight ability of RDH to catalyze dehydrogenation of the unnatural substrate (134). Continuous culture of a RDH constitutive strain, using xylitol as the sole carbon source, has yielded evolved strains that utilize xylitol much faster. The increased fitness exhibited by these strains appears to be due to quantitative increases in RDH activity. In all cases, the increased activity was due to duplication or amplification of the RDH gene, because ribitol dehydrogenase produced by these strains is qualitatively identical to wild-type strains, and because the ability to efficiently utilize xylitol is unstable. These authors suggest that under conditions in which growth is limited by an inefficient enzyme, the most frequent response by populations is amplification of the gene involved. Such amplifications occur at frequencies much higher than mutational alterations of substrate specificity. By this amplification, populations may adapt to short-term environmental stress without stably altering their genetic constitution.

**GLYCYL-tRNA SYNTHETASE** Folk & Berg (37) have described a class of glycine auxotrophs of *E. coli* with an altered glycyl-tRNA synthetase (*glyS*<sup>-</sup> mutants). In such strains, the  $K_m$  of the enzyme for glycine is 20-fold higher than that of wild-type enzyme. Supplying these mutants with a high concentration of exogenous glycine is thought to drive the charging reaction. Most Gly<sup>+</sup> revertants of these mutants harbor duplications (and possibly amplifications) of the *glyS*<sup>-</sup> gene (38). A compelling amount of data leads to this conclusion. (a) Although present in increased amounts, the glycyl-tRNA synthetase produced by these strains is qualitatively identical to parental mutant strains. (b) Such revertants still harbor the parental *glyS*<sup>-</sup> mutation. (c) Their Gly<sup>+</sup> character is unstable; segregants identical to the parental *glyS*<sup>-</sup> mutation are formed at high frequency. However, such instability is dependent upon recombination (*recA*) function. (d) These revertants are often diploid for the nearby loci *xyl*, *malA*, and *glpD*, as determined by their ability to be made heterozygous for these markers. Duplications of the *glyS* region account for most Gly<sup>+</sup> reversion events and arise at a frequency between  $10^{-4}$  and  $10^{-5}$ /cell.

Straus (122) has made careful measurement of the frequency of this duplication event; he estimates the spontaneous frequency to be  $7 \times 10^{-5}$ . The mutagens nitrous acid, ICR-372, niridazole, and ultraviolet light were seen to increase this frequency up to 25-fold.

Genetic tests for inclusion of nearby markers indicate that independent duplication isolates often include different amounts of adjacent material in their duplications. Individual Gly<sup>+</sup> revertants were seen to include various combinations of nearby markers. In two strains, both end points of the duplicated material were

localized; one duplication includes less than 4.9 chromosomal min, and another includes between 4.8 and 5.6 min using the recalibrated linkage map of *E. coli* (9).

Straus & Straus (124) have described the isolation of *glyS*<sup>-</sup> mutants in *Salmonella typhimurium*. As in *E. coli*, all Gly<sup>+</sup> revertants harbor chromosomal duplications, which include the *glyS*<sup>-</sup> region. Such revertants occur at a spontaneous frequency of  $8 \times 10^{-5}$ , comparable to that observed in *E. coli*. However, tests for inclusion of nearby markers indicate that the duplications obtained by this selection are profoundly different than their *E. coli* counterparts. The majority of *Salmonella* Gly<sup>+</sup> duplications appear to include a region comprising 16% of the chromosome. A minority revertant type appears to include 22% of the genome. Thus, these duplications are considerably larger than those isolated in *E. coli*. Most duplications isolated by this technique are formed by a recombination-dependent mechanism. The Gly<sup>+</sup> reversion frequency was found to be 20-fold reduced in a recombination-deficient (*recA*<sup>-</sup>) background.

**HISTIDINE OPERON** *Salmonella* strains harboring tandem duplications of the histidine operon may be selected as those exhibiting increased enzyme expression under conditions that prevent full derepression of the operon (R. P. Anderson and J. R. Roth, manuscript in preparation). Derepression is prevented by the use of strains harboring "down-promoter" (*hisO*<sup>P</sup>) mutations. Because of a promoter defect, such strains (although His<sup>+</sup>) are unable to maximally express *his* enzymes (33). From these strains, increased gene expression is selected as resistance to the histidine analog 3-amino-1,2,4-triazole (AT). This drug is a specific inhibitor of the IGP-dehydratase portion of the *hisB* enzyme (57). Wild-type strains are resistant to the growth-inhibitory properties of this compound because of their ability to derepress *hisB* expression and overcome inhibition. Down-promoter mutants are unable to derepress fully and, therefore, are sensitive to AT. Under appropriate conditions, *hisO*<sup>P</sup> strains give rise to AT-resistant clones whose properties suggest that the mutational event conferring resistance is a tandem duplication of *his* genes. The appropriate conditions consist of selecting for resistance to the lowest concentration of AT necessary for growth inhibition of haploid *hisO*<sup>P</sup> strains; a doubling of *hisB* activity (resulting from a duplication) is sufficient to confer AT resistance. A number of tests indicate that strains isolated by this procedure contain tandem duplications of *his* material: the AT-resistance character of such isolates is genetically unstable (such instability is completely dependent upon a functional recombination system); approximately twofold increases in *his* enzyme levels are observed in these strains; and genetic tests indicate that these strains contain at least two copies of *his* material, as demonstrated through the use of the transposable drug resistance element Tn10 (71). The translocatable element Tn10 is a 11,000-base-pair sequence, which confers tetracycline resistance. Normal haploid strains become His<sup>-</sup> when they inherit *his*<sup>-</sup>::Tn10 insertion mutations. However, AT-resistant strains remain phenotypically His<sup>+</sup> when they inherit such insertions. Thus, AT-resistant strains contain two (or more) copies of the histidine operon; one copy provides the His<sup>+</sup> phenotype and a second copy inherits *his*<sup>-</sup>::Tn10. Such strains are unstable, segregating both *his*<sup>-</sup>::Tn10 and *hisO*<sup>P</sup> haploid progeny.

Although selected as being resistant to an inhibitory concentration of AT, these duplication harboring strains are sensitive to higher AT levels. AT-resistant clones obtained at these higher concentrations appear to contain multiple tandem copies of the *his* operon.

The spontaneous frequency of tandem duplications isolated by this procedure is remarkably high ( $5 \times 10^{-4}$ /cell). Moreover, this frequency is approximately 1000-fold reduced in *recA<sup>-</sup>hisO<sup>P</sup>* strains. Thus, legitimate recombination appears to be involved in formation of these duplications (see below).

**R-FACTOR TRANSITIONING** An excellent example of the adaptive significance tandem duplications may have for bacterial populations has been provided by Rownd and co-workers (46, 100, 101, 108, 109). Drug-resistance factors (R-factors) of the Enterobacteriaceae are circular DNAs composed of two distinguishable units: the resistance transfer factor (RTF), which mediates infectious transfer of the R-factor, and the r-determinant, which harbors a variety of genes conferring drug resistance (26, 48). When harbored by *Proteus mirabilis*, one such R-factor (NR1) has the remarkable property of being able to regulate the number of identical copies of r-determinants associated with each RTF. This regulation occurs in response to drug challenge by antibiotics whose resistance genes reside on the r-determinant. When grown in the absence of antibiotics, these R-factors are composed of a single r-determinant copy attached to a single RTF (100). Following growth in the presence of the appropriate drug, the number of copies of the r-determinant attached to an RTF is seen to increase. The increased r-determinant copies are located in tandem upon the circular R-factor plasmid molecule. This combined plasmid most likely contains one copy of the RTF plus multiple tandem copies (10–15) of the r-determinant. Such selective amplification of the r-determinant portion of the R-factor leads to changes in plasmid density and size. This phenomenon is termed R-factor transitioning.

Fully transitioned cultures are more resistant to the antibiotics for which resistance determinants have been amplified (46). Thus, transitioned cells have a growth advantage over nontransitioned cells when cultured in drug-containing media. This selective advantage appears sufficient to explain how entire cultures of R-factor-containing strains become fully transitioned. By a variety of mixing and reconstruction experiments, Hashimoto & Rownd (46) were able to demonstrate selective outgrowth in drug-containing medium by the partially or fully transitioned cells in a mixed population. Furthermore, partially transitioned clones (those exhibiting increased, but not maximal, drug resistance) may be recovered spontaneously in the absence of selection at a frequency of approximately  $10^{-3}$ . The kinetics of transitioning are consistent with selective outgrowth by this spontaneously arising subpopulation; increased amplification can occur following continued selection for increased resistance.

The response such strains make to drug challenge (transitioning) is transient. If at any time during or after the transition process cells are grown in drug-free media, the plasmid DNAs are seen to reduce the level of r-determinant amplification. After

prolonged cultivation of a fully transitioned culture in drug-free media, the population will return to the nontransitioned state in which a single *r*-determinant is attached to a single RTF. Thus, amplification of resistance determinants is a response to environmental challenge (selection) by antibiotics. In the absence of selection, cultures return to the nontransitioned state.

The ability of R-factor populations to respond by transitioning clearly demonstrates the adaptive role gene duplication (and amplification) may play. If, under any circumstances, an increased gene dosage confers an increased fitness, then a subset of the population (the fraction that harbors a spontaneous duplication) will be able to outgrow haploid cells. Thus, they may increase their fraction of the total population. In the absence of selection, it is unnecessary to maintain the extra gene copies, and segregation events can regenerate a predominantly haploid population.

Electron microscopy of R-factor heteroduplex molecules (63, 102) reveals the presence of IS1 sequences (117, 118) at both junctions between the *r*-determinant and the RTF regions of the plasmid. These sequences are present as a direct (nontandem) repeat. The role these sequences play during the amplification process is currently unknown. It is attractive to speculate that such sequences are involved in recombinational exchanges and that these exchanges lead to duplication (or deletion) of intervening material (102).

**MALATE UTILIZATION** Most *Salmonella* strains grow extremely slowly when L-malate is provided as the sole carbon source. Large, fast-growing colonies (*Mte*<sup>+</sup>) may be selected, and most of the isolates appear to contain large tandem chromosomal duplications (123). A gene dosage effect upon expression of some gene (or genes) included within the duplication is probably the basis for this selection; however, the loci responsible are obscure. Since utilization of malate appears to be limited by transport, duplication of the dicarboxylic acid permease gene (*dct*) may be the basis for this selection.

Straus & Hoffmann (123) have shown that *Mte*<sup>+</sup> mutants have a number of properties that suggest their fast growth on malate is due to a tandem duplication: the *Mte*<sup>+</sup> character is unstable; the *Mte*<sup>+</sup> character is transmissible during Hfr conjugation; and *Mte*<sup>+</sup> strains may be made heterozygous for a large number of contiguous chromosomal loci. End-point mapping of this type has shown that all duplications cover the same chromosomal region from *glpT* to *pyrE* (approximately one third of the *Salmonella* genome). Such large duplications arise spontaneously at a selected frequency of  $4 \times 10^{-5}$ – $4 \times 10^{-4}$  per plated cell. (This frequency is likely an overestimate since *Mte*<sup>+</sup> fast growers appear above a background of leaky *Mte*<sup>-</sup> growth.) A second estimate of the frequency of this event has been provided nonselectively. Approximately 0.5% of *His*<sup>+</sup> recombinants arising from a standard conjugational cross (HfrB2 *metA*<sup>-</sup> × *hisG46*) have the *Mte*<sup>+</sup> phenotype. Such *Mte*<sup>+</sup> clones harbor chromosomal duplications identical to those previously described and appear nonselectively following conjugation (both parental strains are *Mte*<sup>-</sup>). The detection of duplications following genetic crosses is most easily explained by the presence of preexisting duplications in a fraction of the recipient

population (see below). Consistent with this explanation, it was observed that most of the duplicated material in *Mte*<sup>+</sup> strains isolated as above was of recipient origin. Thus, a duplication frequency of  $5 \times 10^{-3}$  is indicated.

**BACTERIOPHAGE LAMBDA** Lambda phages harboring tandem duplications of portions of their genome may be selected both genetically and physically from parental phages carrying large deletions of nonessential material (13, 19, 34, 35, 40). Genetic selections are based upon the finding that there is a minimal genome length (about 75–77% of normal lambda) required for phage viability. Deletion mutants of lambda whose genome is very near this critical length have extremely low burst sizes. Large plaque-forming revertants of such deletions usually harbor tandem duplications of an amount of DNA required to confer viability. More recent genetic selections have capitalized upon an *E. coli* host (GL1) that restricts the growth of deletion phages. This strain strongly selects against the growth of such phages and can be used to select addition derivatives arising from them. Physical selections are based upon measurements that relate lambda DNA length to phage buoyant density in CsCl (130). Since lambda phage heads are of approximately fixed volume, the buoyant density of DNA-filled heads depends upon the length of DNA encapsulated. Deletion mutants of phage lambda band at a density characteristically lighter than normal. DNA additions to these deletion phages may be detected on the phage density profile as a shoulder of increased density. Subsequent genetic and physical characterization of the phages found in this shoulder confirm that the majority contain tandem duplications.

One advantage to studying duplications arising on phage genomes is that identification of tandemly duplicated material may be done by both genetic and physical means. Tandem duplications detected by the selections described above have been identified genetically by their instability. Propagation of such strains (in *recA*<sup>+</sup> hosts) yields many progeny phages identical to the parental deletion plus many phages harboring triplication of the region. However, this genetic instability is absent when duplication phages (*int*<sup>-</sup>*red*<sup>-</sup>) are grown on *recA*<sup>-</sup> hosts. Physical identification of duplicated regions has been provided by DNA-DNA heteroduplex analysis. Heteroduplexes involving duplicated and nonduplicated DNAs exhibit a characteristic single-stranded loop (equal to the length of duplicated material), which may be located anywhere along the region that has been repeated. Based upon similar heteroduplex analysis, Emmons & Thomas (36) have mapped the end points of 24 spontaneously occurring duplications. The distribution of such end points was found to be random.

Duplications of the type described above arise at a spontaneous frequency of  $10^{-4}$ – $10^{-5}$  during a single cycle of infection. Furthermore, this frequency is independent of known recombination functions, because identical frequencies were measured under conditions that eliminate all known homologous recombination systems (*int*<sup>-</sup>*red*<sup>-</sup>*recA*<sup>-</sup>). The measured frequencies of duplication in the above two experiments are likely to be underestimates of total duplication events, since each isolated duplication must be large enough to confer viability, yet not large enough to prevent packaging of the entire genome.

### *Coinheritance of Allelic Markers*

It is standard practice in bacterial genetics to set up conditions that select for cells carrying a particular allele of a given gene. In selecting for a mutant allele the conditions permit loss of the wild-type allele. It is expected that this allele will be converted by mutation to the new selected form. Similarly, in genetic crosses, one selects for inheritance of a particular donor allele under conditions that permit the recipient's copy to be lost and replaced by donor material. Contrary to this standard practice, it is sometimes possible to establish conditions that require both acquisition of the new allele and maintenance of the original allele. One can then select cells having two copies of the gene in question. Only recipient cells carrying a duplication of that gene can acquire the new allele (in one copy) and still maintain the other gene copy in its original form. This general method has been used in a variety of guises to detect gene duplications in bacteria. Several examples are presented below.

**glyT SUPPRESSORS** The *glyT* gene of *E. coli* codes for a species of glycine tRNA reading the codons GGA and GGG. Missense suppressor mutations can arise in the *glyT* gene, which cause the mutant tRNA to read the arginine code words AGA and AGG. Work on these suppressor mutations and the tRNA produced has been reviewed by Hill (50).

Strains carrying only the suppressor allele grow poorly on minimal medium, presumably because they lack sufficient tRNA to read the normal glycine codons GGA and GGG. Strains carrying both a *glyT* suppressor mutation and a second, wild-type copy of the *glyT* gene grow well on minimal medium. When selection is made for new *glyT* suppressor mutants (on minimal medium), most arise as slow-growing haploids carrying one (mutant) copy of the *glyT* gene. Approximately 0.1%, however, arise as fast-growing merodiploids carrying a duplication including the *glyT* gene; one copy is wild type, the other is the suppressor allele (53). Following ultraviolet irradiation, as many as 10% of the *glyT* suppressor mutations arise in strains harboring a duplication (51). It appears that strains carrying duplications of the *glyT* gene are frequent (0.1%) in a normal cell population and that their frequency is greatly stimulated by ultraviolet irradiation. Several lines of evidence indicate that the merodiploid strains carry tandem duplications of a chromosomal region including the *glyT* locus: the merodiploid strains are unstable (clones lacking the suppressor arise frequently); this instability is dependent upon recombination ability (introduction of a *recA*<sup>-</sup> mutation allows the stable maintenance of both wild-type and suppressor alleles of *glyT*, which suggests that loss of the suppressor occurs by recombination between duplicate copies; see Figure 3); by construction of appropriate heterozygotes, it can be shown that the unstable suppressor carrying strains are also diploid for one or more chromosomal genes near the *glyT* locus (the extent of several duplications is presented in Figure 6); and the duplication state is cotransducible with chromosomal markers (55). As outlined above (see Figure 4), if one transduces the join point between two copies of a duplicated region into a haploid recipient, the donor's duplication state can be reestablished. When such crosses are done by using donor strains harboring a *glyT* duplication, it can be

shown that the duplication join point is genetically linked to chromosomal markers and to the *glyT* gene. This is very strong evidence that the merodiploid region is a tandem duplication that is part of the chromosome and not plasmid borne.

As can be seen from the *glyT* case, suppressors represent an ideal situation for detection of duplications. The wild-type allele can be selected if it is vital or required for optimal growth. The suppressor allele can be selected by requiring suppression of an appropriate auxotrophic mutation.

Recently, Hill, Grafstrom & Hillman (54) have reported convincing evidence concerning the mechanism of formation of the *glyT* duplications. They propose that these duplications arise by recombination between ribosomal RNA genes (*rrn*) located at distinct points in the chromosome. Their proposal is strongly supported by the isolation and characterization of closed circular DNA formed in duplication strains following recombination between the duplicate copies. These circles are of the size expected if they arose by recombination between *rrn* genes. Furthermore, each class of circles possesses the expected number of *rrn* sequences (54; see Figure 6; C. W. Hill, personal communication). The existence of the *rrnD* gene between *purD* and *metA* was inferred from this work and has been independently demonstrated by Nomura and co-workers (81, 89, 135). If the *glyT* duplications of class IV and V also involve *rrn* sequences, their structure predicts that another *rrn* gene will be found located clockwise from the *metA* gene.

It is striking to note that four or possibly five *rrn* gene copies are located in this rather small region of the genetic map and that all are placed in the same orientation. Thus, they can provide the nontandem, direct sequence repeats needed for *rec*-dependent duplication formation. This arrangement may be of adaptive significance to *E. coli*, since it permits amplification of the number of *rrn* gene copies. Cells carrying such *rrn* amplifications might have a selective advantage under conditions of extremely rapid growth. Duplications of the *glyT* locus are possibly by-products of this amplification process.

**trpT SUPPRESSORS** Another example of the use of suppressors to detect duplications is the case of lethal amber and UGA suppressor mutations that arise by mutations in the *trpT* gene of both *E. coli* and *Salmonella*. The *trpT* gene codes

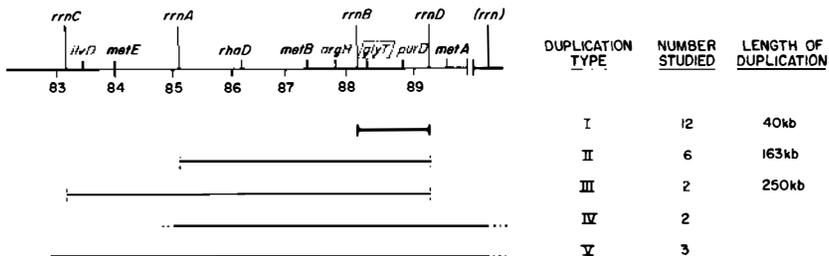


Figure 6 End points of tandem duplications that include that *glyT* region of *Escherichia coli* (54).

for the single tryptophan tRNA and, thus, is essential (58, 115, 136). The first description of mutations at this locus is probably that of Schwartz, who studied unstable suppressors of certain *lac* mutations (111, 112). The *lac* mutations used by Schwartz proved to be UGA mutations (107). Later *trpT* mutations suppressing either amber or UGA mutations were studied in both *Salmonella* and *E. coli*. In these later studies, suppressor mutations were selected in F'-merodiploid strains possessing two copies of the wild-type *trpT* gene. Suppressor mutations could convert one of the two *trpT* gene copies to an amber or UGA suppressor form and still leave the cell with one wild-type copy of the *trpT* gene. Since *trpT* codes for an essential tRNA, it is vital that the cells maintain this wild-type allele. When attempts were made to transduce these recessive lethal suppressor mutations into a haploid recipient strain, it was expected that no transductants would be recovered, since no extra copy of the *trpT* gene was provided. In practice, however, transductants were obtained with surprisingly high frequency. In both *E. coli* (116) and *Salmonella* (84), haploids show less than a fivefold reduction in the frequency of transductants carrying the lethal suppressor (compared to F'-merodiploid recipients with two copies of the relevant region). In both species, all such suppressor-carrying transductants obtained with haploid recipients harbor duplications including both the *ilv* and the *trpT* genes. Since neither the donor nor the recipient in these crosses was known to carry a duplication, it is unclear how these transductants were formed. It seems possible that 20% of the recipient cell population carried a pre-existing duplication, and that transductants arose in this subpopulation by recombination into one of the two copies of the duplicated region. Such a high frequency of duplications seems surprising. However, the *trpT* gene, like the *glyT* gene discussed above, is located between copies of the rRNA cistrons (9). These neighboring genes might provide homologous sequences that could recombine to generate the duplications detected. A high level of duplications for this region might account for the unstable UGA suppressors isolated by Schwartz (111, 112). It seems likely that these are lethal suppressors at the *trpT* locus, which arose in strains carrying a pre-existing duplication of that gene.

Many other examples of unstable suppressor mutations have been reported (28, 30, 43, 56). Like the *glyT* and *trpT* cases described above, these unstable suppressors are likely to be deleterious or lethal when haploid, but they can arise in strains that carry pre-existing duplications. These duplications permit the cells to maintain both the suppressor mutation and its wild-type allele.

T4rII Crosses between complementary mutations with limited opportunity for recombination provide another way of detecting duplications. This method has been exploited extensively in the rII region of phage T4. Here, because of a peculiarity of the rII region, overlapping deletion mutations exist that are unable to recombine but that can complement to provide an rII<sup>+</sup> phenotype. Crosses between such mutants cannot yield wild-type progeny but, at low frequency, can yield phenotypically rII<sup>+</sup> progeny that carry a duplication of the rII region; one of the complementing rII mutations is carried by each copy of the duplicated region. Such partially diploid heterozygotes are unstable and generate segregants of the two

parental  $rII^-$  genotypes. This selection method is made possible by the fact that a portion of the  $rIIB$  cistron is not essential for  $rIIB$  function (21, 29). Deletions affecting  $rIIA$  and the nonessential region of  $rIIB$  still provide  $rIIB$  function. Deletions that affect both essential and nonessential portions of  $rIIB$  provide  $rIIA$  function. Thus,  $rIIA$  and  $rIIB$  deletions that overlap in the nonessential portion of the  $rIIB$  cistron complement, yet are unable to recombine. Crosses between such deletions (selecting  $rII^+$ ) yield tandem duplications, each copy of which harbors one of the parental mutations. It is inferred that these duplications are formed by illegitimate recombination between the two mutant T4 genomes.

The increased genome size of such duplications may have serious consequences for T4 viability, if the "headful" packaging mechanism can no longer generate a terminal redundancy. For this reason, most duplications in T4 must be accompanied by a deletion mutation that removes nonessential material elsewhere in the genome (97, 131). This permits formation of the terminal redundancy. The necessity for this second event reduces the apparent frequency of duplication formation.

This selection method for T4 duplications was first described by Weil, Terzaghi & Crasemann (132). Since then, these duplications have been studied by Symonds and co-workers (27, 125, 128, 129), with special emphasis on analysis of T4 recombination pathways, and by Van de Vate & Symonds (127) and Hornyk & Weil (59), with emphasis on the non-essential regions affected by the associated deletion mutations. The nature of the duplications themselves has been investigated in most detail by Parma and co-workers (97-99), who have made the following observations. (a) Duplication heterozygotes segregate triplication mutants at a frequency approximating the frequency of  $rII^-$  segregants. This supports the tandem nature of the duplications and suggests that duplications segregate by recombination between different genomes rather than between copies of the duplicated region in a single genome (see Figure 3). (b) There is a direct relationship between size of the duplicated region and the frequency of  $rII^-$  segregant formation. (c) The end points of over 50 duplications are located at a wide variety of sites, suggesting that many different illegitimate exchanges are capable of generating duplications. However, the distribution of end points is not random; given the map region of one end of a duplication, it is possible to predict the region of the map in which the other end is most likely to fall. There seems to be a tendency of duplicated regions to terminate within structural genes. No explanation for either of these observations is currently available. (d) The size of the duplications appear to be on the order of 1000-5000 base pairs. The spectrum of duplication types is limited by the requirement that both copies of the  $rII$  region must function and that the duplications cannot exceed a size that can be compensated by a deletion of nonessential material. (e) Constraints upon the nature of detectable duplications make it difficult to estimate their frequency, as discussed by Parma et al (97). However, it has been estimated that the minimum frequency of duplications per genome in T4 is between  $2.5 \times 10^{-6}$  and  $5 \times 10^{-4}$ .

**HISTIDINE OPERON.** Crosses between closely linked, complementary mutations have also provided a way of identifying chromosomal duplications in *Salmonella typhimurium* (4). For example, when two closely linked but complementing histi-

dine-requiring mutations are crossed, two distinct types of His<sup>+</sup> recombinants result. The first type is the expected wild-type recombinant, which arises following a genetic exchange in the small region between the two *his* mutations. The second type of recombinant arises when the recipient cell carries a duplication of the *his* region. In such a cell, the donor's mutant *his* operon can be inherited in one copy of the duplicated region whereas the recipient *his* region is retained in the other copy. Since the two mutations involved complement, the resulting recombinant is a phenotypically His<sup>+</sup> merodiploid, harboring a different *his*<sup>-</sup> mutation in each of its two *his* regions.

The relative frequency of His<sup>+</sup> merodiploid and true *his*<sup>+</sup> recombinants depends upon the frequency of recombination between the two mutations involved and the frequency of duplication-carrying cells in the recipient population. By choosing extremely closely linked mutations, the frequency of true recombinants can be made quite low. In such cases, almost all His<sup>+</sup> recombinants recovered are the complementing merodiploids. The frequency of this recombinant type can be used to estimate the frequency of duplications in the recipient culture. By this means, the frequency of cells carrying duplications of the *Salmonella his* region has been estimated to be approximately 10<sup>-3</sup>–10<sup>-4</sup>. The segments of chromosome included in such duplications can be extremely large; several have been estimated to include as much as 25% of the chromosome (4).

**UNSTABLE RECOMBINANT CLONES** If one assumes that duplications are frequent events and that they can include large segments of the bacterial chromosome, it is clear that whenever complementary auxotrophic mutations are crossed and selection is made for phototrophic recombinants, unstable merodiploids can be expected.

There are many reports of unstable and mixed recombinant clones arising following conjugational crosses (5, 17, 24, 76, 77, 80, 96, 133). Frequently these findings have been discussed as evidence for several rounds of recombination between the recipient chromosome and the transferred chromosomal segment. It seems likely that many reports of unstable recombinant clones can be accounted for in terms of the phenomenon described above. Such cases could be due to inheritance of the donated segment in one of two copies of a duplicated region carried by the recipient cell. The resulting merodiploid cell could then divide repeatedly before undergoing recombination between copies. Eventually, recombination would lead to a mixed clone composed of several haploid segregant types.

**GALACTOSE OPERON** One example of an unstable conjugational recombinant has been analyzed by Campbell (20). He found recombinant clones that were heterozygous for the *gal* region. Subsequent analysis suggested that the merodiploidy was not associated with an F'-factor but was likely to be maintained as a tandem duplication. The two *gal* regions, present in the partially diploid cell, were not cotransduced by P1 phage. Thus, if tandem duplication is involved, the duplicated segment is larger than a P1 genome. However, the unstable characteristic of the duplication was found to be cotransducible with the *gal* region. Campbell suggested

that transduction of the join point between tandem copies of the duplicated region would be sufficient to re-establish the duplication state of the donor (20). Such recombination events must occur between the transduced fragment and each of two copies of the recipient chromosome. His model (outlined in Figure 4) has been used by Hill, Schiffer & Berg (55) to explain their transduction results with the *glyT* duplications. Recombination events such as those depicted in Figure 4 account for many instances in which the merodiploid state of unstable strains has been found to be transducible.

**HETEROCLONES IN STREPTOMYCES** Partially diploid conjugational progeny are frequently encountered in *Streptomyces coelicolor*. Extensive use has been made of these progeny (termed *heteroclones*) in mapping and complementation testing of this organism (see 60). The means of obtaining heteroclones is exactly parallel to one of the methods described above for obtaining duplications of the histidine region of *Salmonella*. Strains carrying closely linked, complementing auxotrophic mutations are crossed, and selection is made for prototrophy. Such crosses yield two recombinant types. Wild-type recombinants result from a genetic exchange between the selected markers; such recombinants are recovered in low yield because of the small region in which the exchange must occur. In addition to these standard recombinants, clones are frequently encountered that have inherited two copies of the entire region. These heteroclones maintain both donor and recipient alleles of the two genes under selection and are often diploid for large segments of chromosome extending well beyond the immediate neighborhood of the selected markers. The structure and segregation behavior of these merodiploids has been used to infer the map order and intervening map distances of widely separated genes in the *Streptomyces* chromosome. The structure and mode of inheritance of these added chromosomal segments has never been fully explained. It seems possible, in view of the high frequency of duplications in other bacteria, that heteroclones result from recombination events occurring in recipient cells that carry pre-existing tandem duplications. The size of the merodiploid region would be determined by the pre-existing duplication, and the heterozygosity would result from replacement of material in one of the two copies by material inherited from the donor parent.

**TRANSFORMATION STUDIES** Transforming systems have also provided examples of unstable recombinants that may represent tandem duplications. These have been observed in *Haemophilus* (78) and *Streptococcus* (8, 15) in situations involving unstable but simultaneous inheritance of two capsular types. Hotchkiss and co-workers (69, 70, 74, 75) have analyzed an unstable mutation in *Streptococcus* conferring sulfonamide resistance (*sul<sup>r</sup>-c*). The resistance property is lost with rather high frequency. During transformation, this unstable resistance characteristic can be transferred by a single DNA fragment, suggesting that the duplicated region may be small or that the duplication state can be re-established by a mechanism such as that outlined in Figure 4. The reasons this drug resistance mutation must be carried as a duplication are not clear. It seems possible that the mutation has a recessive lethal effect and that selection is being made for simultaneous maintenance

of both a wild-type and a mutant copy of the gene involved. Another possibility is that drug resistance is dependent upon the simultaneous presence of two mutant forms of the locus. This seems less likely but could account for the behavior of these strains.

Ravin et al (103, 104) have described erythromycin and streptomycin mutations of *Streptococcus* that behave similarly to the *sul<sup>r</sup>-c* mutations. The erythromycin resistance mutations (*ery<sup>R</sup>*) have been most extensively analyzed. When particular *ery<sup>R</sup>* mutations are transferred into recipient strains by transformation, the resulting recombinants are unstable and segregate both stably resistant clones and fully sensitive clones. Similar to *sul<sup>r</sup>-c*, it is not clear why these particular mutations are inherited frequently as merodiploids. The high efficiency with which merodiploid transformants arise suggests that something in the structure of the donor chromosome promotes duplication formation. Existence of a duplication in the donor could explain this efficiency (see Figure 4). Stable *ery<sup>r</sup>* segregants also give rise to merodiploid transformants when used as donor strains. It is possible that these stable strains still carry a duplication, but its instability is not apparent because of the acquisition of a recessive lethal mutation within one copy of the duplicated region. Most haploid segregants would be nonviable and, thus, not detected.

Although a full explanation of the above phenomena is not yet available, they are likely to involve tandem duplications, since recombination is essential to their segregation and since the unstable state can be cotransferred with chromosomal markers.

An additional case of duplication in a transforming system is the work of Anagnostopoulos and co-workers in *Bacillus subtilis* (6, 7, 126). This is discussed later.

**CROSSES INVOLVING LOW HOMOLOGY** For unknown reasons, unstable recombinants are frequently recovered from crosses in which there is reduced homology for recombination. One example of this situation is interspecies crosses in which the two DNAs involved share little sequence homology. Another example is seen in crosses involving recipient strains carrying large deletions. If selection is made for inheritance of a donor gene by a recipient strain that carries a large deletion removing material homologous to that donor gene, then available sites for recombination between donor and recipient DNA are severely limited. In such situations, unstable recombinant clones are often recovered. These two examples are discussed below, and a model is presented that we feel can explain these observations in terms of tandem duplications present in the donor DNA.

In conjugational crosses between *E. coli* and *Salmonella*, recombinants are recovered with greatly reduced frequency. The genetic maps of these organisms are virtually identical (9, 110), but at the base sequence level most regions of the chromosome show greatly reduced homology. From such crosses, unstable recombinants are frequently recovered. Some of these probably carry F' episomes. Other recombinants, however, seem to have inherited the selected markers by addition at a chromosomal site near the normal map location of the selected marker. The nature of such interspecies crosses has been reviewed by Middleton & Mojica-a (83). Additional reports of similar findings have appeared (67, 68, 86).

One of the best documented cases is the report of Johnson et al (68). Crosses were made between a *Salmonella* Hfr and an *E. coli* F<sup>-</sup> recipient strain. Selection was made for inheritance of the donor's *argH*<sup>+</sup> gene. Many of the Arg<sup>+</sup> recombinants seemed to have inherited the *argH*<sup>+</sup> gene at a chromosomal site independent of the F-factor. Analysis of these recombinants indicated that a block of *Salmonella* material including the *rha*<sup>+</sup>, *metB*<sup>+</sup>, and *argH*<sup>+</sup> genes were inserted near the *E. coli argH* gene, possibly generating a tandem repeat of this chromosomal region. One possible explanation of the occurrence of such insertions is discussed below.

Homology between donor and recipient can also be reduced by use of a deletion mutation. In several cases, unstable recombinants have been observed following transductional crosses in which the recipient strain carries a large deletion of the region whose inheritance was selected. The large size of the deletion reduces the extent of homology shared by the recipient and the transduced fragment. In such cases, transductants arise with greatly reduced frequency and, in some cases, seem to have inherited the selected material by unstable addition to the chromosome at a site near the recipient's deletion. It is likely that these unstable recombinants carry tandem duplications.

Two examples of the above situation have been described in detail. One example, discussed below, involves the *lac* operon of *E. coli* (119–121). Another involves a threonine deaminase mutant of *B. subtilis* studied by Anagnostopoulos and co-workers (65, 66).

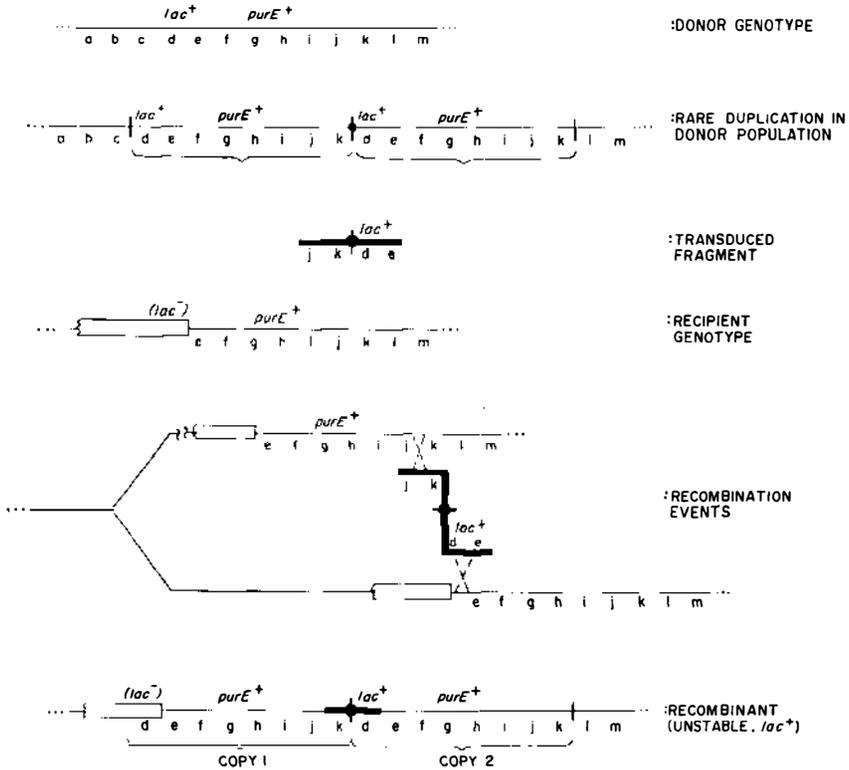
Stodolsky, Rae & Mullenbach (121) used a large *lac* deletion ( $\Delta 111$ ) as recipient in transductional crosses. This deletion removes the *proA* and *proB* genes as well as the entire *lac* operon. Since no stable *lac*<sup>+</sup> recombinants are formed in these crosses, it is thought that the deletion is too large to be spanned by a single transducing fragment. Thus, standard recombination events at the ends of the deletion cannot account for inheritance of the donor *lac* region. The few *lac*<sup>+</sup> recombinants that arise appear to inherit the donor *lac*<sup>+</sup> region by addition to a site near the end of the recipient deletion. The transductants are unstable and segregate *lac*<sup>-</sup> clones that are not distinguishable from the original parent deletion mutant ( $\Delta 111$ ). The transductants become stable after introduction of a *rec*<sup>-</sup> mutation. It seems likely that these strains carry tandem duplications of the *lac* region; one copy carries the recipient's deletion and the other copy carries the functional *lac*<sup>+</sup> operon. The duplicated segment can also include the *proC*, *lon*, and *purE* loci. Since *purE* is not contrasducible with *lac*, much of the duplicated material must be of recipient origin.

Stodolsky proposes a novel model to account for these findings (120). This model involves replication of a partially completed recombinational intermediate followed by illegitimate fusion of two DNA ends. Although the model is plausible and accounts neatly for the observations, we would like to propose a slightly different mechanism, which involves more standard events. We propose that some transductional events depend upon the existence of tandem duplications in the donor population. In Figure 7, the transductional data of Stodolsky is interpreted in this way. This interpretation requires that duplications are present at a low but detectable level in the donor strain. (Evidence that duplications are present at such frequency

has been discussed previously.) When a duplication event generates a linkage between *lac* and a sequence to the right of *purE*, the resulting sequence *jk-de* (see Figure 7) is formed. Transduction of this novel chromosome fragment into a recipient can permit regeneration of the donor duplication and provide a means of inheriting the *lac*<sup>+</sup> region. The resulting Lac<sup>+</sup> transductants carry a duplication of recipient chromosomal material extending from *lac* to a point to the right of *purE*; from the donor, they have inherited only *lac*<sup>+</sup> and material immediately adjacent to the join point of the donor duplication (heavy lines in Figure 7). The only requirement for this sort of transductional event is a duplication in the donor strain, which leaves the *lac*<sup>+</sup> gene cotransducible with the duplication join point. The other end of the duplication (*j-k* in Figure 7) can be anywhere in the chromosome to the right of the *lac* region. This mode of inheritance of *lac*<sup>+</sup> is nearly identical to transductional behavior of duplications described earlier; the only difference is that in this case the duplication would be carried by relatively rare individuals in the donor population.

Events described above may permit inheritance of genes in many cases of restricted homology. If the donor and recipient share few regions of homology, duplication events in the donor can serve to bring these regions close to the selected donor markers. Such rearrangements may be necessary to permit inheritance of that marker. Events of this sort may help to explain the frequent recovery of unstable recombinant clones following crosses between distantly related species.

Transductional shortening of F' episomes can also be accounted for in terms of events similar to those outlined above. The basic observation has been made that F' episomes can be transduced into F<sup>-</sup> recipients by P1 phage (92, 94). Frequently F' transductants so obtained carry episomes that are shorter than the donor F' episome. It appears that only a portion of the donor F' episome is transduced into the recipient. Yet, this portion is able to circularize to generate a new, smaller F' plasmid. The events that lead to circularization are unclear. One possible explanation is that duplications in the donor F' episome provide a terminal redundancy, which allows for circularization of the transduced F' fragment by means of recombination. The only requirement for this would be existence of duplications in the F' episome that are transducible and that include genes essential to plasmid replication. If the transduced fragment included all of one copy of such a duplicated region and a small portion of the second copy, a terminal redundancy would be generated. Recombination between these terminal redundancies would allow circularization of the fragment in the recipient. If this explanation were correct, then the structure of the shortened episome would reflect duplication events that occurred in the donor strain. This mechanism for transductional shortening should require recombination. In the experiments of Ohtsubo (94), shortened F' episomes were formed at 50-fold higher frequency in *recA*<sup>+</sup> than in *recA*<sup>-</sup> recipients. This recombination dependence supports the idea that shortened F' episomes circularize by means of terminal redundancy. The above explanation of F' shortening gains credibility from the observed high frequency of duplication mutations in bacteria. We have recently measured the frequency of tandem duplications of the *his* region on an F'*his* episome (R. P. Anderson and J. R. Roth, manuscript in preparation); 1 in 1000 F'



**Figure 7** Tandem duplications in a donor population allow the formation of *lac*<sup>+</sup> merodiploid recombinants using a nontransducible *lac*<sup>-</sup> recipient deletion (116). Recipient and transduced donor DNAs are light and bold lines, respectively.

episomes carries such a duplication. This high frequency is sufficient to provide an explanation for transductional shortening of F' episomes.

**DUPLICATION FORMATION IN STRAINS WITH TRANSPOSITIONS** Anagnostopoulos et al (6, 7, 126) have described a novel strain of *Bacillus subtilis* in which duplications are formed with high efficiency following transductional or transformational crosses. The unique properties of this strain are associated with a particular mutation (*trpE26*). When strains carrying this mutation are used as recipients in transduction or transformation crosses, the Trp<sup>+</sup> recombinants obtained are unstable and can be shown to carry a large duplication including one quarter to one third of the entire chromosome. The *trpE*<sup>+</sup> gene is included in the duplicated segment.

The unusual behavior of this strain has been shown to be a result of a transposition that generates the *trpE* mutation. The transposition moves a segment of the chromosome (including part of the *trpE* gene) from one point on the chromosome

to another; this rearrangement leaves a haploid cell with a disrupted *trpE* gene (*trpE26*). Any wild-type *trpE*<sup>+</sup> chromosomal segment transduced or transformed into this recipient is homologous to widely separated sequences in the *trpE26* chromosome. The *trpE*<sup>+</sup> fragment must recombine with both of these homologous sequences to permit inheritance of *trpE*<sup>+</sup>. The situation is diagrammed in Figure 8. The events are completely analogous to those described earlier (20, 55); the only difference is that now the illegitimate join point is carried by the recipient.

Duplications such as those described above can now be generated for virtually any region of the chromosome of *Salmonella* and *E. coli* (72). This is made possible through the use of transposable drug resistance genes such as the Tn10 element (14, 71). A translocation situation analogous to that described above for *B. subtilis* can be provided by use of strains carrying Tn10 insertions at different points on the chromosome, thus providing homology at these separated sites. Recombination between these separated homologies yields duplications that include all material between the sites of Tn10 insertion. This method has been tested for the *his* region

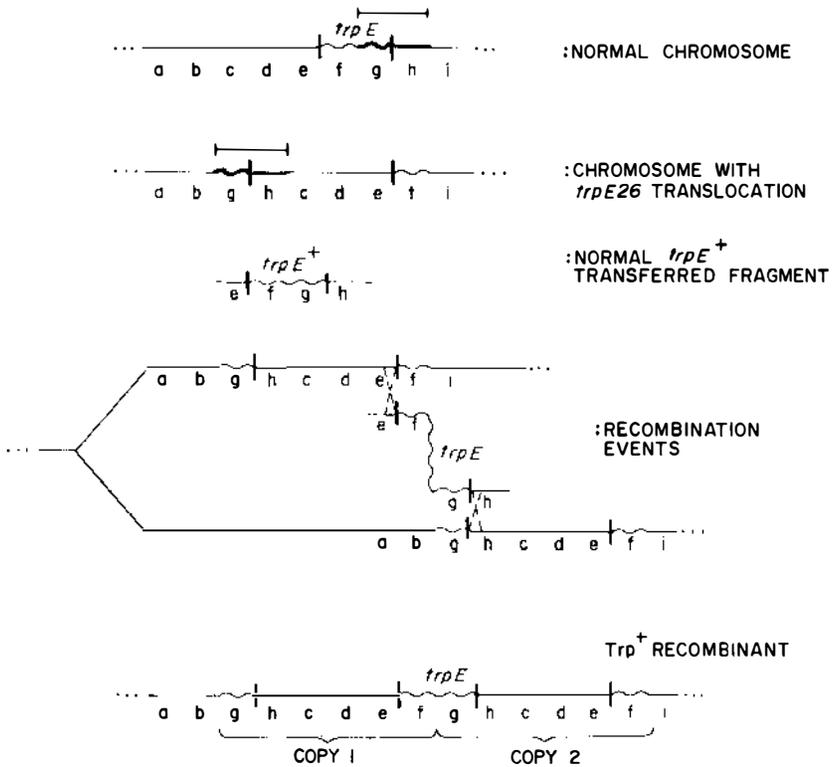


Figure 8 The *trpE26* mutation results from transposition of the chromosomal segment *gh* (bold lined) and promotes tandem duplication when used as a recipient (122).

of *Salmonella* (F. Chumley; unpublished results). Duplications of predetermined extent can be constructed by such transductional crosses. Formation of these duplications is completely analogous to the situation described by Anagnostopolous and coworkers (6, 7, 126).

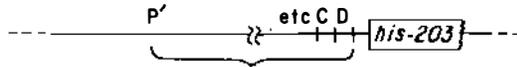
### *Operon Fusion*

A variety of selections involving what we term "operon fusion" have allowed detection of tandem duplications. Figure 5 depicts the basic scheme. Tandem duplications whose end points lie within different units of transcription lead to formation of a hybrid operon in which genes of one transcriptional unit are fused to promoter elements of another. The only restriction on such fusions is that both operons must have the same direction of transcription. Since no material is destroyed by the duplication event, widely separated operons may be subject to such fusion. This is quite unlike deletion fusions (85), in which the preservation of essential genetic material requires that the operons to be fused must be in rather close proximity.

In most cases, selections of this type demand expression of genes that are structurally intact but that have been turned off either by polarity effects or by inactivation of natural promoters. Thus, the selections detect tandem duplication events that fuse unexpressed genes to functional promoters. In the following sections, details of individual selections are considered.

**HISTIDINE OPERON** Operon fusions in the *his* region of *Salmonella* may be selected as follows: mutation *his-203* deletes the entire operator-promoter region and a portion of the first structural gene of the *his* operon. In this strain, the remaining *his* genes are structurally intact, yet unexpressed. Selection is made for turn-on of one of these remaining genes (*hisD*); strains with a functional *hisD*<sup>+</sup> gene are able to utilize the intermediate histidinol as a source of histidine. These *hisD*<sup>+</sup> revertants arise in *his-203* at a spontaneous frequency of approximately 10<sup>-9</sup>. It appears that many of these strains harbor tandem duplications, which fuse the *hisD* gene to foreign regulatory elements. Figure 9 shows how this may be accomplished. At some distance from the histidine operon are promoters (termed P') capable of expressing the *hisD* gene. Tandem duplication of material from a point within the *his* operon to a point within the transcription unit P' leads to expression of the duplicated *hisD* gene by this foreign promoter. Of 21 independent, unstable histidinol-utilizing revertants of *his-203* that we have investigated, all may be accounted for by duplication events such as that shown here (R. P. Anderson and J. R. Roth, manuscript in preparation).

**Transducibility** As discussed above, transduction of the merodiploid state of large tandem duplications may be detected, provided the selected donor marker and the join point between duplicated segments are cotransducible (20, 55). According to Figure 9, the *hisD*<sup>+</sup> gene is necessarily located at the join point between duplicated regions. Thus, the *hisD*<sup>+</sup> character of all such unstable strains is transducible. All transductants obtained in this manner are unstably *hisD*<sup>+</sup> and segregate *hisD*<sup>-</sup> clones of the recipient genotype. Thus, the merodiploid state of these revertants is transducible (see Figure 4).

PHENOTYPEHISTIDINOL<sup>-</sup>

TANDEM DUPLICATION

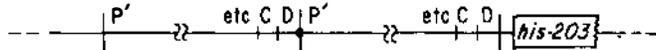
HISTIDINOL<sup>+</sup>

Figure 9 The selection for operon fusions of the *his* region of *Salmonella*.

**Stability** The instability of these strains is dependent upon recombination; *recA*<sup>-</sup> derivatives of the unstable revertants are stably *hisD*<sup>+</sup>. When such stable *recA*<sup>-</sup> strains are used as donors, the *hisD*<sup>+</sup> recombinants obtained with recipient *his-203* are all unstable. Thus, the stable (*recA*<sup>-</sup>) strains still harbor the duplication, and recombination is required for instability of the *hisD*<sup>+</sup> activity.

**Merodiploidy** Unstable strains isolated by this procedure may be shown genetically to be diploid for large chromosomal regions. The extent of duplicated material in 21 independent *hisD*<sup>+</sup> unstable revertants of *his-203* is shown in Figure 10. These different classes of revertants are evident. Class I revertants (5 of 21 isolates) are merodiploid for the nearby loci *metG* and *udk* but not for *cysA*. Class II revertants (15 of 21 isolates) are merodiploid for each of 10 loci tested in the region from *his* through *argB*. Thus, these strains harbor duplications of approximately 16% of the *Salmonella* genome. The join point in class II duplications may be shown to be linked to the *argB* gene by the following experiment: when any class II strain is used as a donor and *argB*<sup>-</sup> *his-203* is used as a recipient, 25% of *Arg*<sup>+</sup> recombinants inherit *hisD*<sup>+</sup> nonselectively. Such recombinants are unstable for both their *hisD*<sup>+</sup> and their *argB*<sup>+</sup> characters. The most likely explanation for these observations involves inheritance of a tandem duplication by events such as those outlined in Figure 4. The single class III isolate is merodiploid for each of 14 markers tested in the region from *his* through *argE*. Thus, this strain harbors a duplication of approximately 26% of the chromosome.

The selection system and unstable *hisD*<sup>+</sup> revertants discussed above were originally described by Ames, Hartman & Jacob (3). The phenomenon was further studied by Levinthal & Yeh (79). Instability of *hisD*<sup>+</sup> revertants was originally interpreted as evidence that the functional *hisD*<sup>+</sup> gene was translocated to an extrachromosomal element (termed the pi-histidine plasmid). The more recent results outlined above suggest strongly that the unstable revertants carry tandem duplications that fuse the *hisD*<sup>+</sup> gene to a functional chromosomal promoter. The most convincing evidence for this conclusion is that recombination is required for both instability and inheritance of the *hisD*<sup>+</sup> character and that the *hisD*<sup>+</sup> gene in class II revertants is cotransducible with a chromosomal *argB*<sup>-</sup> gene. These properties would not be expected of a plasmid-borne *hisD*<sup>+</sup> gene (10, 94).

Class II revertants of *his-203* appear to be generated by a recombination-dependent mechanism, since duplications of this type are not found when the selection is performed in a *recA*<sup>-</sup> background (see below).

**ARGININE OPERON** A number of chromosomal duplications carrying one or more of the four genes clustered in the *argECBH* bipolar operon have been isolated by selecting for reversal of the polar effect caused by certain *arg* mutations (12, 23, 31, 32, 44). Strongly polar *argC*<sup>-</sup> or *argB*<sup>-</sup> mutants are unable to utilize ornithine as an arginine source because of polar effects upon the *argH* (argininosuccinase) enzyme. At a frequency of 10<sup>-7</sup>–10<sup>-8</sup>, strains able to utilize ornithine may be selected. Approximately 10% of these are unstable for their Orn<sup>+</sup> phenotype and are concluded to contain tandem duplications that fuse the *argH* gene to unrelated promoter elements. This conclusion is based upon observed recombination dependence for segregation of the Orn<sup>+</sup> character and upon genetic tests, which demonstrate chromosomal merodiploidy of nearby genes. Cotransduction of Orn<sup>+</sup> with the nearby chromosomal markers *ppc* and *purD* indicate that some *arg* duplications are relatively short (1–2 genes), whereas others are somewhat larger (10–30 genes). The formation of such tandem duplications occurs by a recombination-independent mechanism, since the observed frequency is approximately the same in Rec<sup>+</sup> and Rec<sup>-</sup> genetic backgrounds.

**TRYPTOPHAN OPERON** Jackson & Yanofsky have described the isolation of duplications that fuse the *trp* operon to nearby promoters (64). Such duplications arise from a selection designed to provide relief from the polar effects of *trpE* nonsense mutations. The selection is as follows: the tryptophan synthetase  $\beta_2$  enzyme found in *trpA*<sup>-</sup> strains has but 3% of the activity of the tryptophan synthetase  $\alpha_2\beta_2$  complex in conversion of indole to tryptophan. However, because of the ability of cells to derepress *trpB* ( $\beta_2$ ) activity, such *trpA*<sup>-</sup> strains will grow on indole media. A strain such as *trpE*<sup>-</sup> (amber) *trpE*<sup>-</sup> (ochre) *trpA*<sup>-</sup>, which contains two strong polar mutations in the *trp* operon operator-proximal to *trpB*, is unable to grow on indole media. Selection for indole growers from this strain selects for mutations that relieve the polar effect of the *trpE* nonsense mutations. In addition to indole growth, selection was made for 5-methyltryptophan resistance. Since this analog represses expression of a wild-type *trp* operon, growth in its presence indicates constitutive expression of the *trp* enzymes.

The majority of mutants isolated by this procedure (which arise at a frequency between 10<sup>-7</sup> and 10<sup>-8</sup>) are unstable for their selected phenotype. They are concluded to contain duplications (or translocations) of the *trp* operon, which lead to constitutive expression of *trpB* activity. This conclusion is based upon the instability exhibited by such mutants and a demonstration that these strains are merodiploid for the *trp* operon. In each case, the *trpB*<sup>+</sup> character of these strains does not exhibit the tight linkage to *cysB* found for wild-type strains. Rather, in most cases *trpB*<sup>+</sup> is rather weakly linked to *cysB*. Thus, the *trpB* gene has been moved to a location nearby (but at some distance from) the *trp* locus. Genetic tests indicate that the *trpC*, *trpB*, and *trpA* genes (but not *trpD*) are included in the duplicated region.

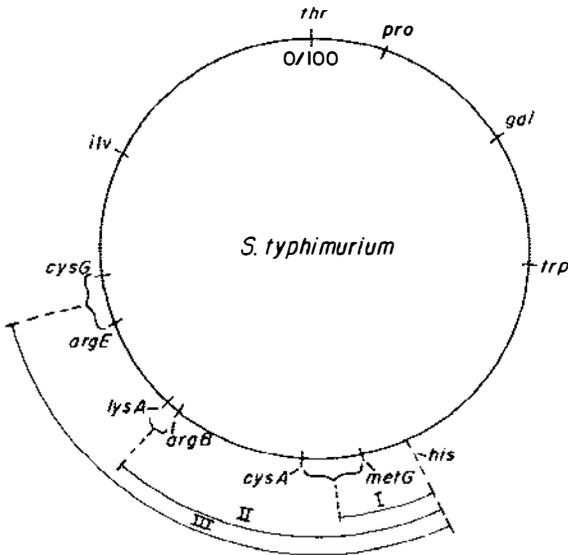


Figure 10 End points of tandem duplications selected by operon fusion in the *his* region of *Salmonella*.

One anomalous behavior of these strains is not easily explained by the simple sort of tandem duplication depicted in Figure 5. In addition to being unstable for their *trpB*<sup>+</sup> character, these strains may also be unstable for other unselected markers; *trpB*<sup>-</sup> segregants are often (but not always) Cys<sup>-</sup>, or they require other unidentified supplements. The simplest duplication and/or translocation models for *trpB*<sup>+</sup> reversion do not account for these results.

Strains harboring tandem duplications of the *trp* operon may be similarly selected in *Salmonella typhimurium* (10a, 11, 82). Positive selection for *supX* deletions (88) has yielded a large collection of strains in which all or part of the *trp* operon has been deleted. Some of these deletions appear to remove only the operator-promoter region of the operon (leaving the structural genes intact), whereas others extend varying amounts into the structural genes. For many of these deletions, expression of "silent" genes may be selected as Trp<sup>+</sup> revertants (in the former case) or as strains able to utilize anthranilic acid or indole as a tryptophan source (in the latter cases). In each instance, some of the revertants are unstable for their selected phenotype. Fusion of remaining *trp* genes to nearby promoters appears the most likely explanation for such strains. Preliminary evidence has been presented that indicates this event occurs by a recombination (*recA*)-dependent process (11).

**GALACTOSE OPERON** The *E. coli* mutation *gal3* results from the linear insertion of an 1100-base-pair insertion sequence into the operator-promoter region of the *gal* operon (2). As a result, genes *galETK* are unexpressed. Among the Gal<sup>+</sup> revertants of *gal3* is a class that is unstable and that synthesizes constitutive levels

of the enzymes (52, 87). It seems likely that these unstable strains are the result of tandem duplications that fuse the *gal* enzymes to unrelated promoters (1); instability of these strains is dependent upon recombination (*recA*) function; unstable revertants are merodiploid for *gal*; LFT lysates prepared from revertants lysogenic for phage lambda contain few  $\lambda gal^+$  transducing particles. However, P1 generalized transduction indicates linkage between *gal*<sup>+</sup> in these strains and the lambda prophage site. Thus, the *gal*<sup>+</sup> character of these strains is located near the original *gal3* site but not close enough to be packaged by a single  $\lambda dgal$  particle.

A rough estimate of the size of the duplicated region can be obtained from this data. Since  $\lambda dgal^+$  particles cannot be formed, yet phage P1 can cotransduce *gal*<sup>+</sup> with the lambda prophage, it appears the duplications include 1–3 min of chromosomal material.

**T4rIIB** Tandem duplications that fuse the bacteriophage T4rIIB region to secondary promoters may be selected as pseudorevertants of frameshift mutations occurring in the promoter proximal region of the rIIB cistron (42). The amino acid sequence coded by the first part of the rIIB cistron is not critical for gene function (21, 29). Thus, revertants of frameshift mutations need not preserve this sequence. Rather, events that restore the reading frame suffice to confer the rIIB<sup>+</sup> phenotype. Among such revertants are strains whose rIIB<sup>+</sup> phenotype is genetically unstable; rIIB<sup>-</sup> segregants identical to the parental strain accumulate in stocks of these revertants. Genetic tests indicate that these unstable strains are merodiploid for the rIIB cistron. Thus, they are concluded to contain tandem duplications that fuse the essential region of rIIB to more distal promoters. This event must occur such that the correct reading frame for the rIIB cistron is encountered at the join point between duplicated regions. Segregation data indicate that these duplications are relatively small, including very little material beyond the rIIB cistron. Headful packaging of DNA necessitates this (see above).

**BACTERIOPHAGE P2** An unusual clear-plaque mutant of bacteriophage P2 (P2*vir37*) has been shown to result from tandem duplication of a small portion of the phage genome (16, 22). This event leads to expression of the early genes *A* and *B* in a manner insensitive to immunity repressor. Thus, it is similar to the types of operon fusions discussed above in that expression of these early genes is no longer under normal control elements. The nature of this mutational event is well documented. (a) The virulent nature of P2*vir37* is unstable. Turbid (wild type) plaques are present in lysates of this mutant at high frequency, despite repeated plaque purification. These segregants appear to arise by a recombinational mechanism, since ultraviolet irradiation of P2*vir37* leads to an increased proportion of turbid segregants. (b) P2*vir37* contains additional DNA. This is indicated by increased phage buoyant density in CsCl and increased heat lability of phage particles. (c) Merodiploidy for gene *B* can be demonstrated genetically. (d) Electron microscope analysis of heteroduplex molecules reveals a tandem duplication of 2.8% of the genome in strain P2*vir37*. This duplication occurs in the genome at a site very near the early functions *A* and *B*.

It is not possible to distinguish whether the duplication event has attached the early genes to a new repressor-insensitive promoter or whether the event has merely separated the normal promoter from its adjacent operator site. In either case, events analogous to operon fusion are involved.

## GENERAL CONCLUSIONS

### *Size and Frequency of Tandem Duplications*

Duplications have been observed in a wide variety of situations. Common to most of these cases is the finding that duplications occur with an unexpectedly high frequency and can include rather large segments of the bacterial chromosome. In retrospect, these properties may not be so surprising. The fact that large duplications cause no loss of function means that there may be virtually no limit to the extent of the duplicated segment. This, in turn, means that a large number of sites may be available for the illegitimate exchanges that lead to duplication. It is likely that the high frequency and large size of duplications reflect these considerations.

### *Mechanisms of Tandem Duplication*

Based upon the cases described above, it seems likely that duplications can be formed by at least two distinct mechanisms. We have termed these legitimate and illegitimate duplications.

Legitimate duplications are those thought to arise by normal recombination between identical or nearly identical sequences located at distinct points in the chromosome. Events of this sort depend upon a functional recombination system and show strong preference for particular end points. Since homologous recombination occurs at a fairly high frequency, it is not surprising that duplications of this type are quite common. It will be interesting to learn what sorts of sequences provide the homology for these events. In the case of *glyT*, it seems clear that the several *rrn* genes are sites for recombination. In other regions of the chromosome, identical or nearly identical tRNA genes may serve. It is also possible that some fortuitous homologies could be sufficient to permit recombination. The possible role of IS sequences will be discussed below.

Illegitimate duplications are those that arise independent of the *rec* system. These include many of the duplications that fuse genes to new promoters. Since it is unlikely that a sequence within a given gene will be repeated elsewhere in the genome, duplications with end points within genes would most likely require an illegitimate recombination, but such events appear to be rare; therefore, one would expect that illegitimate duplications might be rare. In general, this is confirmed by the cases studied. High frequency illegitimate exchange has been observed, however, in cases of site-specific recombination (45). Illegitimate exchanges catalyzed by IS sequences may be examples of such events. In these situations, *rec*-independent duplications might arise at high frequency.

Over the last several years, data has accumulated on a series of transposable base sequences, the IS sequences (18, 117, 118). These sequences can transpose from place to place in the genome and generate, with high frequency, deletions of material

adjacent to the site of insertion (25, 105). Wild-type bacterial strains are thought to carry multiple copies of these IS sequences in their genomes. The sites of such sequences are largely unknown, but some have been identified as points of F-factor insertion (25, 95, 113). IS sequences have many properties that could allow them to play an important role in the formation of both legitimate and illegitimate duplications. For legitimate duplications, IS elements could provide the homology necessary at separated sites on the chromosome. Normal (*rec* dependent) recombination between IS sequences at distant chromosomal positions could be an important contribution to the high frequency of *rec*-dependent duplication formation.

For illegitimate duplications, IS specific events could provide the source of illegitimate exchange. The deletion-generating property of IS elements can be thought of as resulting from an illegitimate exchange between the IS element and a different point in the same chromosome; such exchanges loop out and remove the segment of chromosome between these sites. If such events occur between the IS element and a point on a sister chromosome (see Figure 2), then IS-generated events could form illegitimate duplications.

#### *Uses of Duplications in Genetic Analysis*

Theoretically, duplications provide a means of performing complementation testing and chromosome mapping in bacterial species for which no conjugation system is available.

Mapping is possible since each duplication includes a contiguous block of genes in the chromosome, and since some of the duplications are quite large. Generation of a set of duplications by any of the methods outlined above and determination of what genes are located within the duplicated segment should allow the construction of chromosomal maps. The resolution of this technique is limited only by the number of duplication end points available. If the heteroclones described in *Streptomyces* (see above) are, in fact, the result of duplications, the efficacy of this mapping method may have already been demonstrated.

The use of duplications in complementation testing requires maintaining the duplication state while moving desired mutations into each of the two copies. This could be a problem in some situations, but the use of transposable drug resistance elements as dominant selective markers (72) should provide a means of circumventing such difficulties. Again, the work in *Streptomyces* (see above) may have already provided a demonstration of the general method.

#### *Gene Duplication as a Regulatory Mechanism*

The high frequency of duplication suggests the possibility that cells might use this process as a form of gene amplification. Cells carrying duplications of various regions of the chromosome are present as a reasonably high fraction of the population. Therefore, any environmental condition that favored cells with two (or more) copies of a particular chromosome segment would permit those cells to outgrow the rest of the population in a short time. If the conditions that favored a particular duplication changed, the cell population would not be committed to its new genotype, since the instability of the duplication would permit return to a predominantly haploid state. The kinetics of these populational responses have been described in

theoretical terms (72a). In a sense, this might be thought of as a rather crude regulatory mechanism, which gains its specificity from natural selection. Although we know of no direct evidence demonstrating that such amplification is important to bacterial chromosomes, its role in R-factor transitioning has been clearly demonstrated (see above). A more widespread use of this process for amplification of chromosomal genes would not seem surprising. The finding of Hill et al (54) that many rRNA cistrons are located close together and in the same orientation is suggestive. Duplications arising between these cistrons would amplify rRNA genes; some such duplications would also amplify the ribosomal protein genes that lie between rRNA genes. Amplification of this region might be of selective advantage under conditions of fast growth rate. The findings of Nomura and co-workers (81, 89, 135) that essential tRNA genes are located within some of the *rrn* transcripts could be interpreted as a mechanism of preventing loss of *rrn* loci following unequal recombination events.

#### Literature Cited

1. Ahmed, A. 1975. *Mol. Gen. Genet.* 136: 243-53
2. Ahmed, A., Scraba, D. 1975. *Mol. Gen. Genet.* 136:233-42
3. Ames, B. N., Hartman, P. E., Jacob, F. 1963. *J. Mol. Biol.* 7:23-42
4. Anderson, R. P., Miller, C. G., Roth, J. R. 1976. *J. Mol. Biol.* 105:201-18
5. Anderson, T. F. 1958. *Cold Spring Harbor Symp. Quant. Biol.* 23:47-58
6. Audit, C., Anagnostopoulos, C. 1973. *J. Bacteriol.* 114:18-27
7. Audit, C., Anagnostopoulos, C. 1975. *Mol. Gen. Genet.* 137:337-51
8. Austrian, R., Bernheimer, H. P. 1959. *J. Exp. Med.* 110:571-84
9. Bachmann, B. J., Low, K. B., Taylor, A. L. 1976. *Bacteriol. Rev.* 40:116-67
10. Bastarrachea, F., Willetts, N. S. 1968. *Genetics* 59:153-66
- 10a. Basu, S. K., Margolin, P. 1972. *Genetics* 71:Suppl., p. 3
11. Basu, S. K., Margolin, P. 1973. *Abstr. Annu. Meet. Am. Soc. Microbiol.*, p. 71
12. Beefinck, F., Cunin, R., Glansdorff, N. 1974. *Mol. Gen. Genet.* 132:241-53
13. Bellett, A. J. D., Busse, H. G., Baldwin, R. L. 1971. *The Bacteriophage Lambda*, pp. 501-13. Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab.
14. Berg, D. E., Davies, J., Allet, B., Rochaix, J. D. 1975. *Proc. Natl. Acad. Sci. USA* 72:3628-32
15. Bernheimer, H. P., Wermundsen, I. E. 1969. *J. Bacteriol.* 98:1073-79
16. Bertani, G., Bertani, L. E. 1974. *Proc. Natl. Acad. Sci. USA* 71:315-19
17. Bresler, S. E., Lanzov, V. A., Blinkova, A. A. 1967. *Genetics* 56:105-16
18. Bukhari, A. I., Shapiro, J. A., Adhya, S., eds. 1977. *DNA Insertion Elements, Plasmids, and Episomes*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab. 450 pp.
19. Busse, H. G., Baldwin, R. L. 1972. *J. Mol. Biol.* 65:401-12
20. Campbell, A. 1965. *Virology* 27:329-39
21. Champe, S. P., Benzer, S. 1967. *J. Mol. Biol.* 4:288-301
22. Chatteraj, D. K., Inman, R. B. 1974. *Proc. Natl. Acad. Sci. USA* 71:311-14
23. Cunin, R., Elseviers, D., Glansdorff, N. 1970. *Mol. Gen. Genet.* 108:154-57
24. Curtiss, R. III. 1964. *Genetics* 50: 679-94
25. Davidson, N., Deonier, R. C., Hu, S., Ohtsubo, E. 1975. *Microbiology* 1: 56-71
26. Davies, J. E., Rownd, R. 1972. *Science* 176:758-68
27. Davis, K. J., Symonds, N. 1974. *Mol. Gen. Genet.* 132:173-80
28. Dawson, G. W. P., Smith-Keary, P. F. 1963. *Heredity* 18:1-20
29. Drake, J. W. 1963. *Genetics* 48:767-76
30. Eggertsson, G., Adelberg, E. A. 1965. *Genetics* 52:319-40
31. Elseviers, D., Cunin, R., Glansdorff, N. 1969. *FEBS Lett.* 3:18-20
32. Elseviers, D., Cunin, R., Glansdorff, N., Baumberg, S., Ashcroft, E. 1972. *Mol. Gen. Genet.* 117:349-66
33. Ely, B. 1974. *Genetics* 78:593-606
34. Emmons, S. W. 1974. *J. Mol. Biol.* 83:511-25
35. Emmons, S. W., MacCosham, V., Baldwin, R. L. 1975. *J. Mol. Biol.* 91:133-46
36. Emmons, S. W., Thomas, J. O. 1975. *J. Mol. Biol.* 91:147-52

37. Folk, W. R., Berg, P. 1970. *J. Bacteriol.* 102:193-203
38. Folk, W. R., Berg, P. 1971. *J. Mol. Biol.* 58:595-610
39. Franklin, N. C. 1967. *Genetics* 55:699-707
40. Franklin, N. C. 1967. *Genetics* 57: 301-18
41. Franklin, N. C. 1971. *The Bacteriophage Lambda*, pp. 175-82. Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab.
42. Freedman, R., Brenner, S. 1972. *J. Mol. Biol.* 69:409-19
43. Gallucci, E., Garen, A. 1966. *J. Mol. Biol.* 15:193-200
44. Glandsdorff, N., Sand, G. 1968. *Genetics* 60:257-68
45. Gottesman, M. F., Weisberg, R. A. 1971. *The Bacteriophage Lambda*, pp. 113-38. Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab.
46. Hashimoto, H., Rownd, R. H. 1975. *J. Bacteriol.* 123:56-58
47. Hegeman, G. D., Rosenberg, S. L. 1970. *Ann. Rev. Microbiol.* 24:429-62
48. Helinski, D. R. 1973. *Ann. Rev. Microbiol.* 27:437-70
49. Hershey, A. D. 1970. *Carnegie Inst Washington Yearb.* 69:717-22
50. Hill, C. W. 1975. *Cell* 6:419-27
51. Hill, C. W., Combriato, G. 1973. *Mol. Gen. Genet.* 127:197-214
52. Hill, C. W., Echols, H. 1966. *J. Mol. Biol.* 19:38-51
53. Hill, C. W., Foulds, J., Soll, L., Berg, P. 1969. *J. Mol. Biol.* 39:563-81
54. Hill, C. W., Grafstrom, R. H., Hillman, B. S. 1977. *DNA Insertion Elements, Plasmids, and Episomes*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab. 450 pp.
55. Hill, C. W., Schiffer, D., Berg, P. 1969. *J. Bacteriol.* 99:274-78
56. Hill, R. F. 1963. *J. Gen. Microbiol.* 30:289-97
57. Hilton, J. L., Kearney, P. C., Ames, B. N. 1965. *Arch. Biochem. Biophys.* 112:544-47
58. Hirsh, D. 1971. *J. Mol. Biol.* 58:439-58
59. Homyk, T., Weil, J. 1974. *Virology* 61:505-23
60. Hopwood, D. A. 1967. *Bacteriol. Rev.* 31:373-403
61. Horiuchi, T., Horiuchi, S., Novick, A. 1963. *Genetics* 48:157-69
62. Horiuchi, T., Tomizawa, J., Novick, A. 1962. *Biochim. Biophys. Acta* 55:152-63
63. Hu, S., Ohtsubo, E., Davidson, N., Saedler, H. 1975. *J. Bacteriol.* 122:764-72
64. Jackson, E. N., Yanofsky, C. 1973. *J. Bacteriol.* 116:33-40
65. Jamet, C., Anagnostopoulos, C. 1969. *Mol. Gen. Genet.* 105:225-42
66. Jamet-Vierney, C., Anagnostopoulos, C. 1975. *Genetics* 81:437-58
67. Johnson, E. M., Craig, W. G. Jr., Wohlhieter, J. A., Lazere, J. R., Synenki, R. M., Baron, L. S. 1973. *J. Bacteriol.* 115:629-34
68. Johnson, E. M., Placek, B. P., Snellings, N. J., Baron, L. S. 1975. *J. Bacteriol.* 123:1-6
69. Kashmiri, S. V. S., Hotchkiss, R. D. 1975. *Genetics* 81:9-19
70. Kashmiri, S. V. S., Hotchkiss, R. D. 1975. *Genetics* 81:21-31
71. Kleckner, N., Chan, R. K., Tye, B.-K., Botstein, D. 1975. *J. Mol. Biol.* 97: 561-75
72. Kleckner, N., Roth, J., Botstein, D. 1977. *J. Mol. Biol.* In press
- 72a. Koch, A. L. 1977. *Genetics*. In press
73. Langridge, J. 1969. *Mol. Gen. Genet.* 105:74-83
74. Ledbetter, M. S., Hotchkiss, R. D. 1975. *Genetics* 80:667-78
75. Ledbetter, M. S., Hotchkiss, R. D. 1975. *Genetics* 80:679-94
76. Lederberg, J. 1949. *Proc. Natl. Acad. Sci. USA* 35:178-84
77. Lederberg, J. 1957. *Proc. Natl. Acad. Sci. USA* 43:1060-65
78. Leidy, G., Hahn, E., Alexander, H. E. 1953. *J. Exp. Med.* 97:467-82
79. Levinthal, M., Yeh, J. 1972. *J. Bacteriol.* 109:993-1000
80. Lotan, D., Yagil, E., Bracha, M. 1972. *Genetics* 72:381-91
81. Lund, E., Dahlberg, J. E., Lindahl, L., Jaskunas, S. R., Dennis, P. P., Nomura, M. 1976. *Cell* 7:165-77
82. Margolin, P., Bauerle, R. H. 1966. *Cold Spring Harbor Symp. Quant. Biol.* 31: 311-20
83. Middleton, R. B., Mojica-a, T. 1971. *Adv. Genet.* 16:53-76
84. Miller, C. G., Roth, J. R. 1971. *J. Mol. Biol.* 59:63-75
85. Miller, J. H., Reznikoff, W. S., Silverstone, A. E., Ippen, K., Signer, E. R., Beckwith, J. R. 1970. *J. Bacteriol.* 104:1273-79
86. Mojica-a, T., Middleton, R. B. 1972. *Genetics* 71:491-505
87. Morse, M. L. 1967. *Genetics* 56:331-40
88. Mukai, F. H., Margolin, P. 1963. *Proc. Natl. Acad. Sci. USA* 50:140-48
89. Nomura, M. 1976. *Cell* 9:633-44
90. Novick, A. 1961. *Purdue Growth Sym-*

- posium. Growth in Living Systems*, New York: Basic Books
91. Novick, A., Horuichi, T. 1961. *Cold Spring Harbor Symp. Quant. Biol.* 26:234-45
  92. Novick, R. P. 1969. *Bacteriol. Rev.* 33:210-63
  93. Ohno, S. 1970. *Evolution by Gene Duplication*, p. 160. New York: Springer-Verlag.
  94. Ohtsubo, E. 1970. *Genetics* 64:189-97
  95. Ohtsubo, E., Deonier, R. C., Lee, H. J., Davidson, N. 1974. *J. Mol. Biol.* 89:565-84
  96. Ou, J. T. 1975. *Genetics* 80:401-19
  97. Parma, D. H., Heath, G. T., Che, C.-C., Annest, J. L. 1977. *Genetics*. In press
  98. Parma, D. H., Ingraham, L. J., Snyder, M. 1972. *Genetics* 71:319-35
  99. Parma, D. H., Snyder, M. 1973. *Genetics* 73:161-83
  100. Perlman, D., Rownd, R. H. 1975. *J. Bacteriol.* 123:1013-34
  101. Perlman, D., Twose, T. M., Holland, M. J., Rownd, R. H. 1975. *J. Bacteriol.* 123:1035-4
  102. Ptashne, K., Cohen, S. N. 1975. *J. Bacteriol.* 122:776-81
  103. Ravin, A. W., Ma, M. 1975. *Genetics* 80:421-43
  104. Ravin, A. W., Takahashi, E. A. 1970. *J. Bacteriol.* 101:38-52
  105. Reif, H. J., Saedler, H. 1975. *Mol. Gen. Genet.* 137:17-27
  106. Rigby, P. W. J., Burleigh, B. D., Hartley, B. S. 1974. *Nature* 251:200-4
  107. Roth, J. R. 1970. *J. Bacteriol.* 102:467-75
  108. Rownd, R., Mickel, S. 1971. *Nature New Biol.* 234:40-43
  109. Rownd, R., Perlman, D., Hashimoto, S., Mickel, S., Appelbaum, E., Taylor, D. 1973. *Cellular Modification and Genetic Transformation*, pp. 115-28. Baltimore: Johns Hopkins Univ. Press
  110. Sanderson, K., Hartman, P. E. 1977. *Bacteriol. Rev.* In press
  111. Schwartz, N. M. 1965. *J. Bacteriol.* 89:712-17
  112. Schwartz, N. M. 1967. *Genetics* 57:505-12
  113. Sharp, P. A., Hsu, M.-T., Ohtsubo, E., Davidson, N. 1972. *J. Mol. Biol.* 71:471-97
  114. Smithies, O., Connell, G. E., Dixon, G. H. 1962. *Nature* 196:232-36
  115. Soll, L., Berg, P. 1969. *Nature* 223:1340-42
  116. Soll, L., Berg, P. 1969. *Proc. Natl. Acad. Sci. USA* 63:392-99
  117. Starlinger, P., Saedler, H. 1972. *Biochemie* 54:177-96
  118. Starlinger, P., Saedler, H. 1976. *Curr. Top. Microbiol. Immunol.* 75:111-23
  119. Stodolsky, M. 1973. *Genetics* 73:Suppl. pp. 65-66
  120. Stodolsky, M. 1974. *Genetics* 78:809-22
  121. Stodolsky, M., Rae, M. E., Mullenbach, E. 1972. *Genetics* 70:495-510
  122. Straus, D. S. 1974. *Genetics* 78:823-30
  123. Straus, D. S., Hoffmann, G. R. 1975. *Genetics* 80:227-37
  124. Straus, D. S., Straus, L. D. 1976. *J. Mol. Biol.* 103:143-54
  125. Symonds, N., Van den Ende, P., Durston, A., White, P. 1972. *Mol. Gen. Genet.* 116:223-38
  126. Trowsdale, J., Anagnostopoulos, C. 1975. *J. Bacteriol.* 122:886-98
  127. Van de Vate, C., Symonds, N. 1974. *Genet. Res. Cambridge* 23:87-105
  128. Van de Vate, C., Van den Ende, P., Symonds, N. 1974. *Genet. Res. Cambridge* 23:107-13
  129. Van den Ende, P., Symonds, N. 1972. *Mol. Gen. Genet.* 116:239-47
  130. Weigle, J., Meselson, M., Paigen, K. 1959. *J. Mol. Biol.* 1:379-86
  131. Weil, J., Terzaghi, B. 1970. *Virology* 42:234-37
  132. Weil, J., Terzaghi, B., Crasemann, J. 1965. *Genetics* 52:683-93
  133. Wood, T. H. 1967. *Science* 157:319-21
  134. Wu, T. T., Lin, E. C. C., Tanaka, S. 1968. *J. Bacteriol.* 96:447-56
  135. Yamamoto, M., Strycharz, W. A., Nomura, M. 1976. *Cell* 7:179-90
  136. Yaniv, M., Folk, W. R., Berg, P., Soll, L. 1974. *J. Mol. Biol.* 86:245-60
  137. Zipkas, D., Riley, M. 1975. *Proc. Natl. Acad. Sci. USA* 72:1354-58