Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6025–6029, August 1986
Genetics

Structure and expression of a tandem duplication of the *Drosophila* metallothionein gene

(gene amplification/cadmium resistance/*Mtn* gene)

EDWARD OTTO, JULIE E. YOUNG, AND GUSTAVO MARONI

Department of Biology and Curriculum in Genetics, University of North Carolina, Chapel Hill, NC 27514

Communicated by James F. Crow, April 21, 1986

ABSTRACT **A strain of cadmium-resistant *Drosophila* containing a chromosomal duplication of the metallothionein gene was isolated. This duplication is stably inherited in the absence of selective pressure, and larvae homozygous for it can produce approximately twice as much metallothionein RNA as wild-type larvae. The entire duplication was cloned within a 5.7-kilobase fragment; this fragment contained a direct, tandem repeat of 2.2 kilobases of DNA: 228 bases of 5′ flanking DNA, the entire transcription unit, and 1.4 kilobases of 3′ flanking sequences. The 3′ region of the first repeated unit is joined to the 5′ region of the second unit by a 6-base-pair segment we define as the novel joint. This joint forms part of a 10-base-pair inverted repeat of a segment within the 3′ region of the first unit. Comparison of the sequences of the 5′ and 3′ boundaries revealed no extensive regions of similarity at a position corresponding to the novel joint, thus suggesting that a mechanism other than homologous recombination was involved in the origin of this duplication.**

In multicellular organisms, gene amplification produces multiple gene copies, the fate of which is determined by the setting in which this process occurred. Amplification in terminally differentiated somatic cells is part of a developmentally regulated process and is restricted to a single cell generation. In cells in culture, amplified genes are associated with resistance to antimetabolites or heavy metals and can be stably transmitted or not (1–3). Of more immediate relevance to the evolutionary process, amplification that occurs in the germ line may lead to development of gene families.

The function of amplified DNA is often apparent. However, little information is available about the detailed structure and mechanism of origin of amplified or duplicated sequences. For mammalian cells, structural analyses have been hindered because the amplified units are large, highly repeated, and constantly being rearranged (2, 3). In bacteria and yeast, amplified genes are often contained within smaller, defined segments of DNA and are organized as a tandem array of identical units (4–9). However, details of the DNA sequence at the boundaries of amplified DNA, where repeated and flanking single-copy sequences join, remain unexamined. In only one case, the amplified *ampC* locus of *Escherichia coli*, has the nucleotide sequence of these segments been determined (5).

Unlike mammals, in which multiple genes code for metallothionein (MT), *Drosophila* contains a single MT gene (10). In this paper, we describe the structure of a duplication of this gene in a cadmium-resistant *Drosophila* strain. DNA sequence analyses of the boundaries of this duplication and the junction of the repeated units seem to exclude homologous recombination *per se* as the mechanism of origin. Studies of MT RNA levels and tolerance to cadmium indicate that this duplication is functionally significant.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

***Drosophila* Cultures.** Wild-type strains used were Samarkand, Oregon R, Urbana, Hikone, and Canton S. Forty pairs of flies from each of the wild-type strains were mixed and subdivided among four cultures. The progeny of this first generation were again mixed and subdivided into four cultures, and this process was repeated after each generation. For the first generation, the medium used in each culture was standard yeast/cornmeal/molasses supplemented with 0.1 mM $CdCl_2$. The cadmium concentration was raised to 0.5 mM in the next generation and then raised again, in 0.5 mM increments, when the population appeared adapted to the new level. The highest level we could achieve, after approximately 10 generations, was 2.0 mM, and the population was maintained at this concentration for approximately 15 more generations.

From the resulting cadmium-resistant population, the duplication $Dp(3;3)Mtn^{H22}$ was isolated. $Dp(3;3)Mtn^{H46}$ was isolated from an unselected laboratory strain carrying the mutations *cn* and *bw*. All mutations and balancer chromosomes are described by Lindsley and Grell (11). Except for viability tests, flies were reared in standard yeast/cornmeal/molasses medium in 250-ml bottles.

Viability Tests. Larvae were reared on medium (IDM-YE) composed of 0.2 g of Instant Drosophila Medium (Carolina Biological, Burlington, NC) per ml of a solution of 4% yeast extract. One hundred first-instar larvae from each strain were grown on this medium with or without a supplement of $CdCl_2$. For a given cadmium concentration, viability was calculated as the proportion of individuals that reached pupariation at that concentration as compared to individuals reared on unsupplemented medium (12). At least three replicate vials at each concentration were prepared for each strain.

Molecular Analyses. Unless otherwise specified, methods reported by Lastowski-Perry *et al.* (13) or Maroni *et al.* (10) were used. Genomic DNA for Southern analyses was isolated by the procedure of Lis *et al.* (14). The DNA sequence of subclones of pDmH22 in M13 vectors was determined by the dideoxy chain-termination method (15). All radioactive probes were prepared by nick-translation of the *Drosophila* MT cDNA cDm51 (13) or the genomic clone λDm13 (10).

RESULTS AND DISCUSSION

Characterization of the Cadmium-Resistant Population. A population of *Drosophila* resistant to 2.0 mM $CdCl_2$ was selected from a mixture of five wild-type strains. Genomic DNA from each of the unselected wild-type strains carried the MT gene (*Mtn*) within a single 3.5-kilobase (kb) *Eco*RI

Abbreviations: MT, metallothionein; bp, base pair(s); kb, kilobase(s).

restriction fragment, as shown by its hybridization (Fig. 1*A*) to a *Drosophila* MT cDNA probe, cDm51 (10). Similar examination of DNA from flies of the cadmium-resistant population revealed that *Mtn* sequences were present in two fragments: 3.5 kb and 5.8 kb (Fig. 1*A*). *In situ* hybridization of a labeled *Drosophila* MT genomic clone, λDm13, to polytene chromosomes from larvae of the resistant population localized all related sequences to the previously identified *Mtn* locus at 85E10-15 in chromosomal arm 3R (ref. 10 and data not shown).

Homozygotes for individual third chromosomes were obtained by appropriate crosses between flies of the cadmium-resistant population and flies with the balancer chromosome *TM3*. These crosses were made in the absence of selective pressure. Southern analysis revealed that DNA from all the resultant homozygous lines contained *Mtn* sequences within either the 3.5-kb or the 5.8-kb fragment but not both (Fig. 1*B*). Flies that were homozygous for the 5.8-kb *Eco*RI fragment were more resistant to cadmium toxicity (LC_{50} = 0.08 mM) than the original wild-type strains (average LC_{50} = 0.03 mM) (Fig. 2).

MT Gene Organization in Cadmium-Resistant Flies. The restriction map in Fig. 3 shows that these cadmium-resistant flies, homozygous for the 5.8-kb *Eco*RI fragment, carry a direct, tandem duplication of a 2.2-kb segment of DNA that includes the entire MT transcription unit. This duplication was designated $Dp(3;3)Mtn^{H22}$. Cytological examination of polytene chromosomes of heterozygotes between $Dp(3;3)Mtn^{H22}$ and a standard chromosome showed no visible abnormality in the 85E region.

The presence of two copies of *Mtn* in $Dp(3;3)Mtn^{H22}$ was verified in a genomic reconstruction experiment. A known amount of total *Drosophila* DNA from homozygous cadmium-resistant or wild-type flies was digested with *Hpa* II and fractionated by gel electrophoresis. Mixtures of digested *E. coli* DNA plus sufficient λDm13 to simulate genomes containing one, two, or four copies of the MT gene were identically digested and fractionated on the same gel. Southern analysis showed that with DNA from $Dp(3;3)Mtn^{H22}$, the intensity of hybridization to cDNA was approximately equivalent to that of the reconstructed genome with two MT genes (Fig. 4).

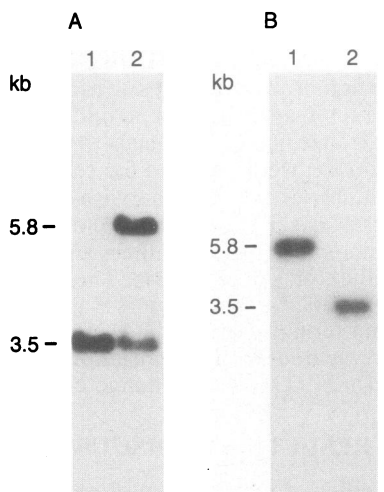


FIG. 1. Blot of total *Drosophila* DNA digested with *Eco*RI and hybridized to a cDm51 probe. DNA was isolated from Samarkand flies (*A*, lane 1), cadmium-resistant flies from the selected population (*A*, lane 2), or flies isogenic for different third chromosomes of that population (*B*, lanes 1 and 2). Results identical to that shown in lane 1 of *A* were obtained with DNA from Oregon R, Urbana, Hikone, or Canton S (data not shown).

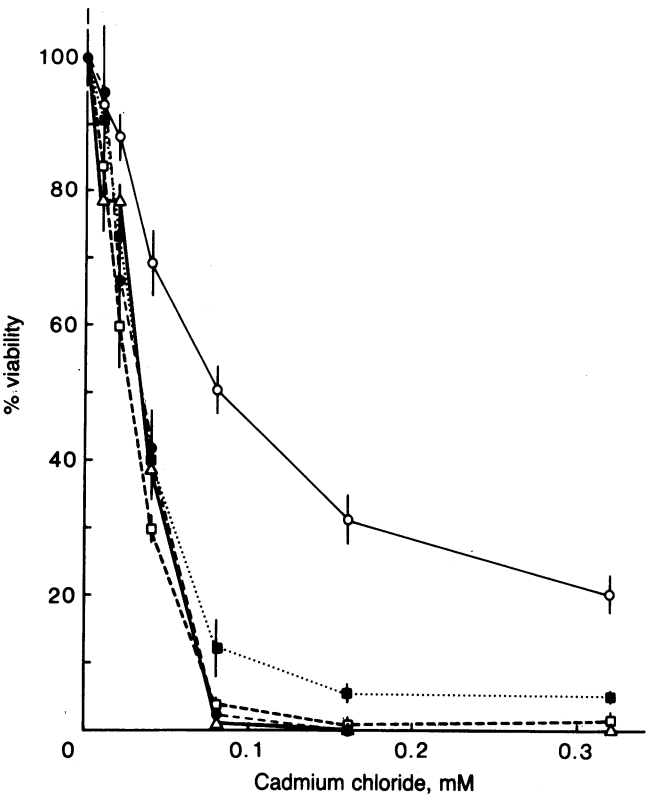


FIG. 2. Viability of wild-type [Samarkand (■), Oregon R (□), Hikone (△), Urbana (●)] and cadmium-resistant [$Dp(3;3)Mtn^{H22}$ (○)] larvae in medium containing $CdCl_2$. Each data point represents the ratio between the average number of larvae that reached pupariation at a given cadmium concentration (0.01, 0.02, 0.04, 0.08, 0.16, or 0.32 mM) and the average number of larvae that reach pupariation in unsupplemented medium. For wild-type strains, n = 3; for cadmium-resistant, n = 4. Vertical bars, ± SEM. The fifth strain, Canton S, is approximately as sensitive to cadmium as the other wild-type strains (data not shown). All strains are more sensitive to cadmium in IDM-YE than in standard medium.

A chromosomal DNA fragment that contained the entire duplication was cloned from a library of fragments produced by a complete digestion of genomic DNA with *Pst* I and *Eco*RI and enriched for fragments of the expected size (Fig. 3*A*) by gel electrophoresis. One clone, pDmH22, contained a 5.7-kb insert that hybridized to the cDNA probe, and its restriction map agreed with the one obtained from digestions of genomic DNA.

DNA Sequence. Subclones of pDmH22 were used to obtain the nucleotide sequence of five regions of the duplication: the junction of repeated units, the 5′ and 3′ boundaries, and the intron within each MT gene (Fig. 3*A*). A portion of these sequences is shown in Fig. 5*A*. The junction of the repeated units was composed of segments from both boundaries joined by a 6-base sequence not present in either boundary. These 6 bases, which defined the novel joint, were present within a perfect 10-bp inverted repeat; the potential stem–loop structure is shown in Fig. 5*B*. Based on this definition of the joint, each repeated unit contained 228 bases of 5′ flanking sequence, the entire *Mtn* transcription unit, and approximately 1.4 kb of 3′ flanking sequences.

In order to test whether this duplication could have arisen by unequal crossing-over between DNA segments of similar sequence, the 5′ and 3′ boundaries were compared. The only significant similarity found starts 28 bases upstream of the 6-base joint and comprises 15 bases, 10 of which match (compare top and bottom lines in Fig. 5*A*). In the more immediate vicinity of the joint, however, there is no appreciable similarity between the two boundaries.

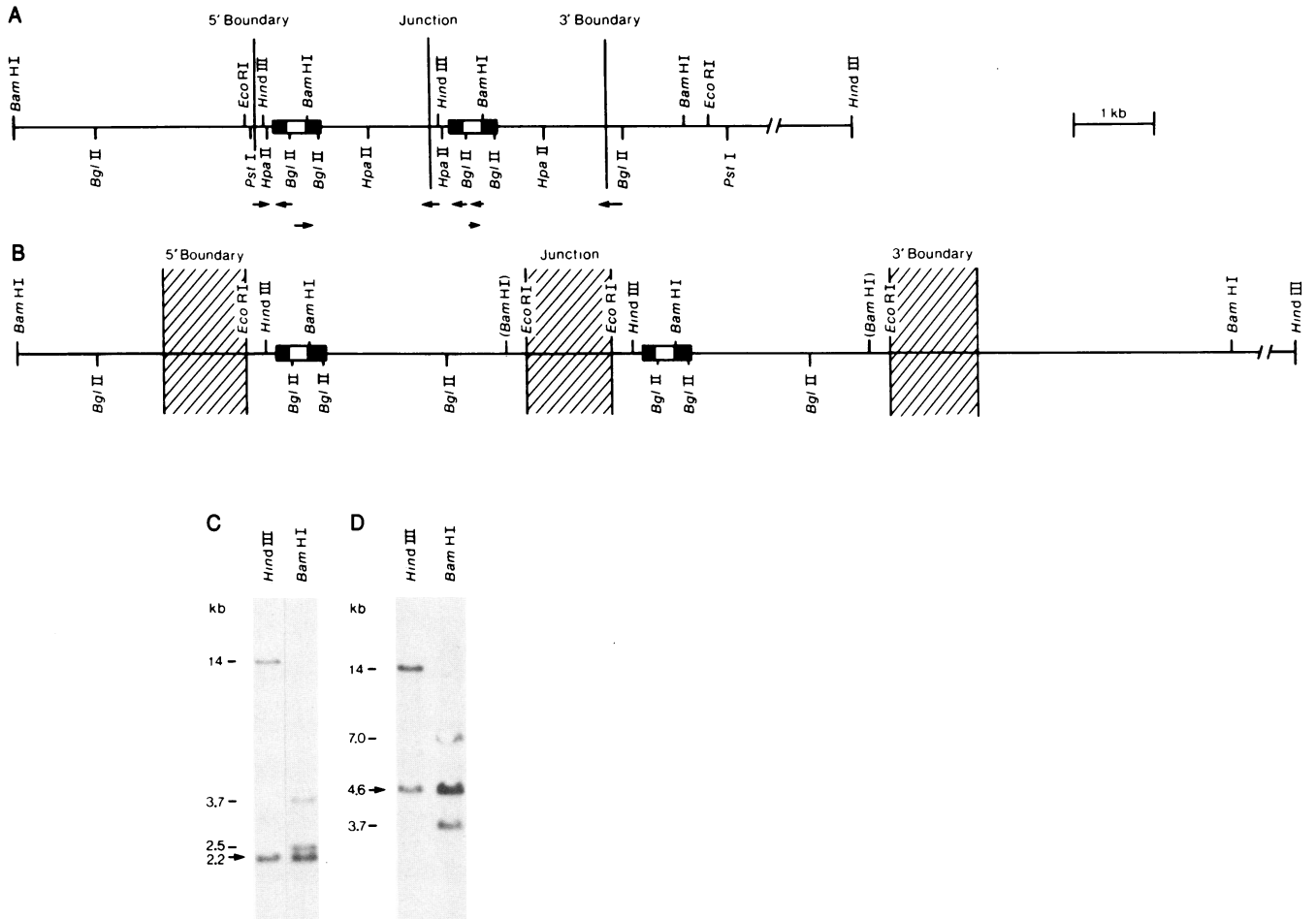


FIG. 3. (*A*) Genomic restriction map of the *Dp(3;3)Mtn*^{H22} region. The map was obtained by hybridization of digested genomic DNA to probes that were derived from cDm51 or from subclones containing only 5′ or 3′ portions of the gene. Each repeated unit is delimited by vertical bars. Boxes represent the transcribed region; exons and introns are indicated in black and white, respectively. The 5.7-kb *Pst* I–*Eco*RI fragment was cloned in pUC9 (pDmH22). Arrows below the map show the direction and extent of sequencing of different subclones of pDmH22. Each subclone was subjected to at least two independent sequence determinations. (*B*) Genomic restriction map of the *Dp(3;3)Mtn*^{H46} region. Crosshatched areas represent the range of uncertainty for the boundaries of the repeated unit. The position of a *Bam*HI site that is present in *A* but is absent in this strain is indicated by parentheses. (*C* and *D*) Southern blots of DNA from flies containing *Dp(3;3)Mtn*^{H22} (*C*) or *Dp(3;3)Mtn*^{H46} (*D*), which was digested with *Hin*dIII or *Bam*HI and hybridized to a cDm51 probe. Arrows indicate the position of the fragment in each lane that contains the novel joint.

To search for sequence divergence between the two units of the duplication, their respective introns were also sequenced. The two introns were identical to each other and both differed from the one in the originally described *Mtn* sequence by a single-base deletion at position 256 (10). The two repeated units are indistinguishable in the 530 bases (including the 264 bases composing the intron) sequenced in both. This suggests that the duplication is of recent origin and could have occurred during the selection process. However, because the flies used for cadmium selection were not genetically marked, contamination by a fly carrying a preexisting duplication cannot be ruled out.

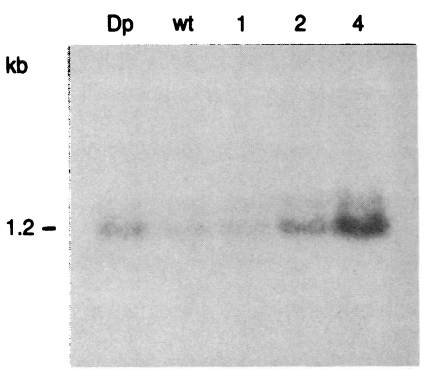


FIG. 4. Genomic reconstruction experiment. Each lane contained either 3 μg of *Drosophila* DNA, from cadmium-resistant *Dp(3;3)Mtn*^{H22} (Dp) or wild-type (wt) larvae, or 3 μg of a mixture of *E. coli* and λDm13 DNA (in proportions of 1, 2, or 4 copies per haploid genome). The filter was hybridized to a cDm51 probe. A *Drosophila* genome size of 1.6×10^8 base pairs (bp) was assumed. The upper band in each lane is the result of partial digestion by *Hpa* II.

MT RNA Levels. Relative MT RNA levels in wild type (Samarkand) and *Dp(3;3)Mtn*^{H22} were compared by hybridization of the cDNA probe to total nucleic acid from larvae that were grown in various concentrations of $CdCl_2$. The amount of hybridization showed that cadmium-resistant larvae grown at high concentrations of $CdCl_2$, for a time sufficient to allow maximal accumulation of *Mtn* transcripts (13), contained approximately twice as much MT RNA as wild-type larvae reared under identical conditions (Fig. 6). This suggests that both genes are being transcribed, a possibility consistent with heterologous experiments showing that only 130-bp of 5′ flanking sequences of the *Drosophila* MT gene are sufficient for metal-regulated transcription of this gene in hamster cells (J. Allen and R. Palmiter, personal communication). The alternative—en-

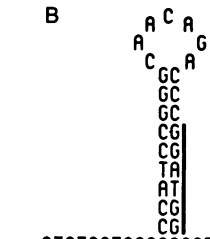
A

5' BOUNDARY caccgccacccccatacatatgtggtacgcaagtaaGAGTGCCTGCGCATGCCCCATGTG

JUNCTION ATCTCCTACGCCCATCCGGGCAACAGACCCGGATGGGAGTGCCTGCGCATGCCCCATGTG

3' BOUNDARY ATCTCCTACGCCCATCCGGGCAACAGACCCttggccagcagtccatgtctgtggagcact

B



ATCTCCTACGCGAGTGCCTGCGCATGCCCCATGTG

FIG. 5. (*A*) A portion of the nucleotide sequence of the 5′ and 3′ boundaries of *Dp(3;3)Mtn*^{H22} and the junction of the repeated units. Sequences are aligned based on maximal similarity. Bases within the junction that are absent from both boundaries are underlined; inverted repeats are overlined. Sections of the boundaries that appear in the junction are indicated by uppercase letters. (*B*) Potential stem–loop structure, which could form by base-pairing of the inverted repeats within the junction sequence. All sequences represent the antisense strand and are displayed 5′ to 3′.

hanced expression of only one gene—cannot be ruled out, however. In any case, it is reasonable to conclude that increased tolerance to cadmium in these larvae is a direct consequence of increased expression of *Mtn* sequences and that the latter is a consequence of the duplication. The correlation between MT mRNA levels and increased resistance to cadmium toxicity has been observed in mammalian cells in culture (16, 17).

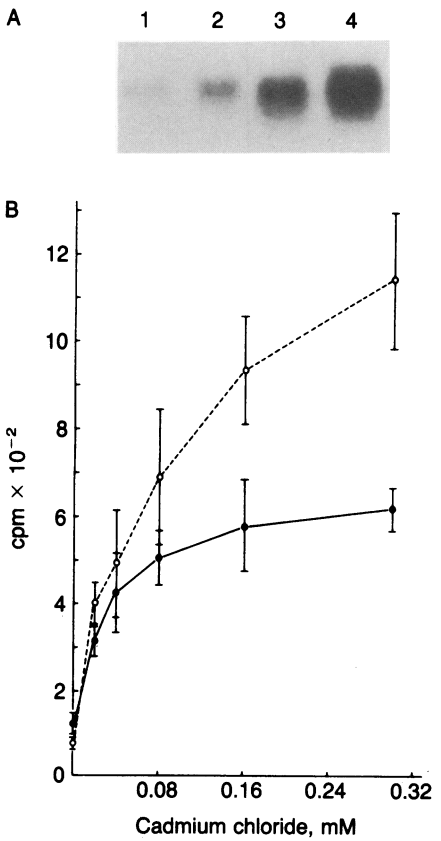


FIG. 6. MT RNA levels. (*A*) Autoradiograph of blot of electrophoretically fractionated RNA hybridized to a cDm51 probe. Each lane contained 6 μg of total nucleic acid from wild-type (lanes 1 and 3) or *Dp(3;3)Mtn*^{H22} (lanes 2 and 4) larvae which were uninduced (lanes 1 and 2) or treated with 0.16 mM $CdCl_2$ for 36 hr (lanes 3 and 4). (*B*) MT RNA levels of wild-type (●) or cadmium-resistant (○) larvae that were treated with various concentrations of $CdCl_2$. The level of hybridization was measured by scintillation counting of filters from RNA blots as in *A*. Data points for each concentration are means of the amount of hybridization (cpm) to nucleic acid from uninduced larvae (n = 6) or from larvae that were treated for 24 or 36 hr (n = 4). Vertical bars indicate ± SEM.

Another Duplication of the MT Gene. An unselected laboratory strain containing another duplication of the MT gene, *Dp(3;3)Mtn*^{H46}, was also identified. Restriction enzyme analysis of DNA from homozygous flies showed that they contained a 4.6-kb repeated DNA segment (Fig. 3*D*) and that both boundaries of this unit were different from those in *Dp(3;3)Mtn*^{H22} (Fig. 3*B*).

The Origin of Duplicated and Amplified Genes. Two mechanisms have been proposed most often for gene amplification: saltatory replication and unequal crossing-over (1–3). The best example of the former is the development-specific replication of DNA containing the chorion genes of *Drosophila*. Replication of these sequences is disproportionate and results in an onion skin structure in which the degree of amplification of a given sequence is directly related to its proximity to the origin of replication (18–20). In contrast, unequal crossing-over is proposed for cases where amplified DNA is organized as a linear, tandem array of identical units in which all repeated sequences seem to be amplified to the same degree (4–9). According to this model, the generation of a duplication, which perhaps represents the primary event in gene amplification, would require that regions of similar sequence flank both sides of the unamplified unit. Evidence from studies in bacteria showed that large regions of similarity, such as rRNA genes, are sites for recombination and that this process, which is RecA-dependent, results in large chromosomal duplications (21, 22). Edlund and Normark (5) have determined the nucleotide sequences for the 5′ and 3′ flanking regions of the chromosomal *ampC* locus as well as for the joint of repeated units in an amplified locus from an ampicillin-resistant mutant. They observed that these sequences share a 12-bp segment of perfect homology and they suggest that the original duplication was generated by recombination within these short segments. Other instances of gene duplications have been described in which the mechanism does not seem to involve legitimate recombination (21), although the arguments are less direct because nucleotide sequences are not available.

Our evidence indicates that homologous recombination, by itself, cannot account for duplication *Dp(3;3)Mtn*^{H22}. This is based on two observations. First, the only region of some sequence similarity ends 14 bases away from the joint. Second, given the high degree of sequence conservation between the two units of the duplication, it is reasonable to assume that the 6-base segment unique to the joint was introduced during the duplication process. The occurrence of an inverted repeat (of which the 6 bases are part) at the point of the novel joint may be significant to the mechanism of duplication, but further speculation should be deferred.

In mammals, MT exists as two major isoforms, which are closely related in amino acid sequence; they are encoded by multigene families whose members contain nucleotide se-

quences that are, likewise, related. In mouse, there are two genes; they are arranged in tandem and transcribed in the same direction (23). In humans, there are several functional genes at least some of which are linked (24, 25). This gene organization may indicate that these families have arisen by tandem duplications of the type described in this paper.

We have previously shown (10) that there is a single MT gene in *Drosophila*. The presence of duplications in at least two laboratory strains (presented here) and in samples from several natural populations (data not shown) indicate that *Drosophila melanogaster* may be at a transitional stage in the evolution of this gene, during which duplications are present but not yet fixed.

We are very grateful to T. D. Tlsty, J. C. Lucchesi, and D. F. Kubai for valuable suggestions and to J. Allen and R. Palmiter for sharing their unpublished data. We thank S. Whitfield for the artwork. This work was supported by U.S. Public Health Service Grant ES02654 to G.M. and National Institutes of Health Training Fellowship T32-GM07092 to E.O.

1. Schimke, R. T. (1982) *Gene Amplification* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
2. Stark, G. R. & Wahl, G. M. (1984) *Annu. Rev. Biochem.* **53,** 447–491.
3. Schimke, R. T. (1984) *Cell* **37,** 705–713.
4. Grundstrom, T., Bjork, G. R. & Normark, S. (1980) *Mol. Gen. Genet.* **180,** 249–257.
5. Edlund, T. & Normark, S. (1981) *Nature (London)* **292,** 269–271.
6. Fogel, S. & Welch, J. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5342–5346.
7. Karin, M., Najarian, R., Haslinger, A., Valenzuela, P., Welch, J. & Fogel, S. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 337–341.
8. Albertini, A. M., Hofer, M., Calos, M., Tlsty, T. D. & Miller, J. H. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 841–850.
9. Tlsty, T. D., Albertini, A. M. & Miller, J. H. (1984) *Cell* **37,** 217–224.
10. Maroni, G., Otto, E. & Lastowski-Perry, D. (1986) *Genetics* **112,** 493–504.
11. Lindsley, D. L. & Grell, E. H. (1968) *Genetic Variations of Drosophila melanogaster* (Oak Ridge National Laboratory, Oak Ridge, TN).
12. Maroni, G. & Watson, D. (1985) *Insect Biochem.* **15,** 55–63.
13. Lastowski-Perry, D., Otto, E. & Maroni, G. (1985) *J. Biol. Chem.* **260,** 1527–1530.
14. Lis, J. T., Simon, J. A. & Sutton, C. A. (1983) *Cell* **35,** 403–410.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Beach, L. R. & Palmiter, R. D. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2110–2114.
17. Gick, G. G. & McCarty, K. S., Sr. (1982) *J. Biol. Chem.* **257,** 9049–9053.
18. Spradling, A. C. & Mahowald, A. P. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1096–1100.
19. Spradling, A. (1981) *Cell* **27,** 193–201.
20. Osheim, Y. N. & Miller, O. L. (1983) *Cell* **33,** 543–553.
21. Anderson, R. P. & Roth, J. R. (1977) *Annu. Rev. Microbiol.* **31,** 473–505.
22. Anderson, R. P. & Roth, J. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3113–3117.
23. Searle, P. F., Davison, B. L., Stuart, G. W., Wilkie, T. M., Norstedt, G. & Palmiter, R. D. (1984) *Mol. Cell. Biol.* **4,** 1221–1230.
24. Karin, M. & Richards, R. I. (1982) *Nature (London)* **299,** 797–802.
25. Schmidt, C. J., Hamer, D. & McBride, O. W. (1984) *Science* **24,** 1104–1106.