Cell, Vol. 33, 53–63, May 1983, Copyright © 1983 by MIT 0092-8674/83/050053-11 $02.00/0

A Structure for Amplified DNA

James M. Roberts, Linda B. Buck, and Richard Axel
Columbia University
Institute of Cancer Research
701 West 168th Street
New York, New York 10032

Summary

We have employed gene transfer to generate cell lines in which a chromosomal region consisting solely of defined DNA sequences has undergone gene amplification. We have analyzed recombinant clones from the amplified array to determine the physical structure of amplified DNA in the cell lines. The amplified DNA we have analyzed consists of a tandem array of at least 20 individual repeating units. The individual units are contiguous, and are joined to one another by homologous recombination between repeated sequences. At first approximation, all homologous recombinations are permitted such that crossing-over may occur between any two repeated sequences. Since individual units contain multiple repeated elements, the array is not a regularly repeating structure. The individual units within the array are heterogeneous, both in size and in sequence content. These observations suggest models of gene amplification which involve multiple cycles of unscheduled DNA replication at a single locus, followed by multiple recombination events which serve to link individual units to one another and ultimately to the chromosome.

Introduction

Gene amplification is a frequent mechanism by which cells meet the demand for increased quantities of specific gene products (Brown and Dawid, 1968; Gall, 1974; Alt et al., 1978; Spradling and Mahowald, 1981). In somatic cells, gene amplification may permit cells to overcome deficiencies in essential enzymatic activities. These deficiencies may result from exposure to enzyme inhibitors (Alt et al., 1978; Wahl et al., 1979) or from mutations in the genes encoding the enzyme itself (Roberts and Axel, 1982; Brennand et al., 1982). During development, individual genes or sets of genes may be specifically amplified at precise developmental stages (Gall, 1974; Yao et al., 1974; Spradling and Mahowald, 1981; Glover et al., 1982). The changes in gene number observed in somatic cells either in culture or during development have also been exploited in evolutionary time to generate stable multigene families present in the chromosome of the gamete as well as the somatic cell (Tartoff, 1975; Hood et al., 1975).

In most instances, an increase in the number of genes results in a proportionate increase in the level of specific mRNA and protein. Although the function of amplified DNA is frequently apparent, the mechanisms of gene amplification as well as the precise structure of amplified DNA have been difficult to determine. In cultured cells, the amplified DNA may exist stably integrated within the chromosome (Biedler and Spengler, 1976; Nunberg et al., 1978) or unstably as autonomously replicating elements (double minutes) which lack both centromeres and kinetochores (Kaufman et al., 1981; Brown et al., 1981; Haber and Schimke, 1981). In most instances, the amplified unit is quite large (up to 2000 kb) and may proceed into sequences for which cloned probes are not available (Nunberg et al., 1978; Hamlin et al., 1982). As a consequence, the precise structure of the amplified array or of individual units within the array has not been determined.

We have determined the structure of amplified DNA in transformed cell lines that amplify short segments of DNA consisting of defined sequences for which cloned probes are available. Transformed cells that integrate genes coding for selectable markers also integrate other genes at high frequency (Wigler et al., 1979). Transformed cells therefore contain concatenates of donor DNA, composed of a single selectable gene along with 1000–2000 kb of carrier DNA, at a single locus in the chromosome (Perucho et al., 1980; Robins et al., 1981). If selection is imposed for amplification of a single gene within the transforming element, the process of amplification is likely to extend into adjacent carrier sequences. If the carrier DNA consists of a well-defined collection of cloned DNA sequences, the length and order of these sequences in both parental and amplified cell lines can be mapped with precision. This analysis permits us to describe in detail the physical structure of amplified DNA. Furthermore, the elucidation of this structure suggests mechanisms by which the amplified array is generated.

Results

The Generation of Amplified Genes

In a previous study, we transformed aprt$^-$ tk$^-$ cells with a plasmid containing a wild-type *aprt* gene linked to a truncated *tk* gene from which the promoter had been deleted (Figure 1) (Roberts and Axel, 1982). Transformed cells which integrate a single copy of this plasmid exhibit the aprt$^+$ phenotype but remain tk$^-$. Tk$^+$ subclones were isolated from these transformants and were shown to result from 20 to 50 fold amplification of the linked plasmid along with significant lengths of adjacent carrier DNA. These variants produced aberrant tk transcripts such that multiple *tk* genes are required to produce sufficient enzyme to convert the cell to the tk$^+$ phenotype.

These observations suggested an experimental approach to study the physical structure of amplified DNA. Aprt$^-$ tk$^-$ cells were transformed with the aprt-tk plasmid along with defined carrier DNA consisting solely of 20 random plasmid clones containing 3–15 kb inserts of human DNA. These plasmid clones were constructed by digestion of human genomic DNA with Hind III and Bam HI, and subsequent insertion of the resulting fragments

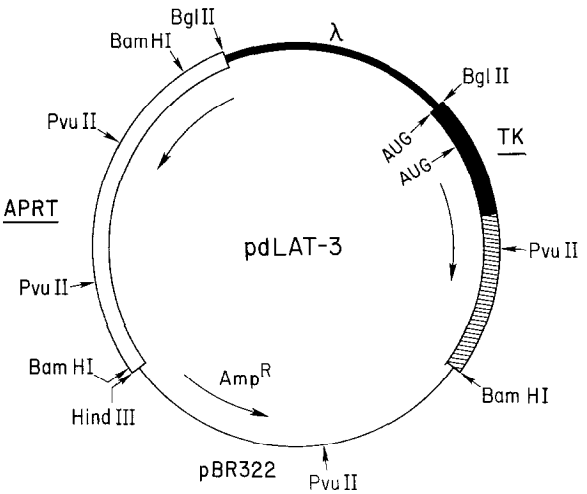

Figure 1. Recombinant Plasmid pdLAT-3 Contains *tk* and *aprt* Gene Sequences

Plasmid pdLAT-3 contains a 2.8 kb Bgl II–Bam HI fragment of the herpes simplex virus genome inserted into the Bam HI site of the vector pBR322. This fragment includes the entire coding sequences of the thymidine kinase gene (▬) as well as 1.6 kb of 3′-flanking DNA (▨), but lacks the promoter and 5′-flanking DNA sequences. The inhibitor AUG is located 50 nucleotides 3′ to the Bgl II site. Additional in-phase AUG codons are noted. This plasmid also contains a 4.3 kb Hind III–Bgl II fragment of hamster DNA, which contains the entire hamster *aprt* gene as well as approximately 700 bp of 5′-flanking DNA (▭). A 2.4 kb Bgl II insert of bacteriophage λ DNA, originating from positions 36489 to 38935 of the phage genome, separates the hamster and herpes virus DNA sequences. Note that the *tk* and *aprt* genes are inserted with opposite polarities.

into the Hind III and Bam HI sites of the pBR322 derivative pAT153 (see Experimental Procedures). These clones therefore possess a single Hind III and Bam HI site. Prior to transformation, the human clones were linearized with Hind III, mixed, and ligated to each other, creating a high molecular weight heteropolymer consisting of a random array of defined DNA sequences. This heteropolymer was combined with the aprt-tk plasmid and introduced into cells as a calcium phosphate precipitate. Cell lines constructed in this way contained a single copy of the aprt-tk plasmid imbedded in defined carrier DNA.

Several tk^+ amplified subclones were derived from one $aprt^+$ tk^- cell line. We have previously described a preliminary analysis of the organization of amplified DNA in these cell lines (Roberts and Axel, 1982). In this study, genomic DNA was isolated from a tk^- parent and eight independently derived tk^+ mutant subclones, cleaved, blotted, and probed for pAT153 plasmid sequences (Figure 2). The DNA was cut with Bgl II and Xba I, two enzymes that do not cut within pAT153, such that each integrated plasmid sequence would be displayed as a unique band. As anticipated, the unamplified tk^- parent shows numerous plasmid bands of single copy intensity. This multiplicity of bands in the tk^- parent reflects the composition of the transforming element. All eight tk^+ subclones have amplified a different subset of the pAT bands. We define an amplification unit as those sequences that are reiterated by virtue of the amplification process. Thus despite a

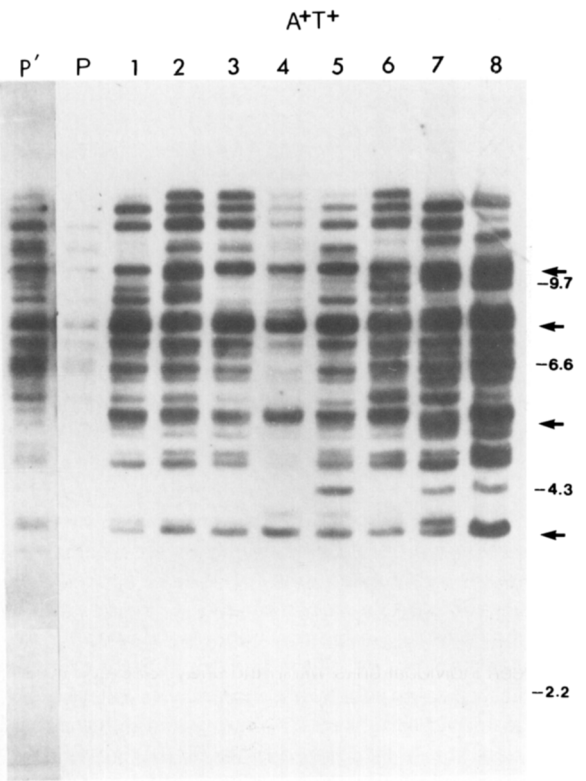

Figure 2. Arrangement of Amplified DNA

Tk^- $aprt^-$ L cells were transformed with pdLAT-3 and cotransformed with 1 μg of carrier DNA consisting entirely of 20 pAT153 clones containing 5–15 kb inserts of human spleen DNA. DNA was isolated from one $aprt^+$ tk^- clone and eight $aprt^+$ tk^+ subclones. Ten micrograms of each DNA were sequentially digested with Bgl II and Xba I, subjected to electrophoresis through a 0.7% agarose gel, blotted to nitrocellulose, and probed with ^{32}P-labeled pBR322 DNA. Lane P is the parental tk^- clone. Lanes 1–8 are the eight tk^+ subclones. Lanes P and 1–8 were exposed for the same length of time; lane P′ is a significantly longer exposure of lane P to display the single copy bands. Arrows indicate the four bands common to all eight amplified subclones.

common genetic background, the amplification unit in each independent tk^+ subclone is unique. The size of the amplification unit varies in individual clones from 40 to more than 200 kb. By counting the amplified pAT bands, we obtain an estimate of the size of the amplification unit. For example, since each pAT sequence is flanked by approximately 10 kb of unique human DNA, mutant 4 has amplified about 40 kb of DNA flanking the $aprt^+$ tk^+ plasmid (Figure 2, lane 4). Mutant 8 has amplified over 20 bands, suggesting that the unit in this cell may be larger than 200 kb (Figure 2, lane 8).

Mapping the Amplified Array

These studies provided a series of cell lines that have amplified the aprt-tk plasmid along with flanking sequences for which cloned probes were available. It was therefore possible to construct genomic libraries both from the parental tk^- line and from one of the amplified tk^+ subclones and to isolate the individual units within the amplified

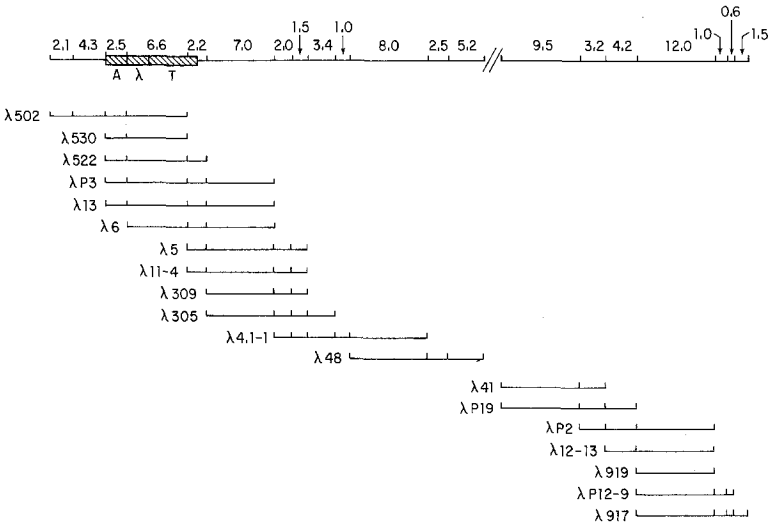

Figure 3. Physical Map of the Amplified Unit

Nineteen overlapping clones together define an 80 kb segment of DNA which becomes amplified in JR4. Eco RI sites are indicated in both the individual phage clones and the composite map. Charon 4 λ clones were isolated from partial Eco RI libraries of either an unamplified parental cell line or an amplified subclone as described in the text. Clones λP3, λ11-4, λ4.1-1, λP19, λP2, λP12-9 all originate from the library of the unamplified parental tk^- cell line. The linear map deduced from these clones was confirmed and supplemented with λ clones isolated from a library of JR4.

array. Comparison of phage clones from parental and amplified cell lines that are homologous to amplified DNA should allow us to determine the physical structure of the individual amplified unit as well as the physical relationship between individual units within the array. One tk^+ mutant, JR4, was selected for analysis, since this cell line possesses the smallest amplified unit among all the tk^+ variants examined (Figure 2). Two partial Eco RI libraries of approximately 1.5×10^6 clones were prepared by inserting DNA from the tk^- parent and the amplified subclone, JR4, into the λ phage vector, charon 4.

The transforming element in the parental cell line consists of perhaps 30–40 human plasmids. However, only a small subset of these sequences is amplified. We wished to isolate only those phage clones which contain sequences from within the amplified array. To identify these phage clones, human DNA inserts were isolated from each of the 20 plasmid clones used as carrier in the construction of the initial cell line. Individual inserts were used as probes in genomic blot hybridizations with DNA from the amplified mutant, JR4. In this manner, we identified those human DNA sequences which had been coamplified with the aprt-tk plasmid. These human DNA sequences were used as probe to screen the libraries of both the unamplified tk^- cell line and the amplified subclone, JR4. Nineteen distinct but overlapping clones were obtained, which permits the construction of a linkage map of the transforming DNA surrounding the *aprt-tk* genes (Figure 3). One set of 12 overlapping clones defines a contiguous 50 kb segment, consisting solely of donor DNA sequences. A second 30 kb block of donor DNA is defined by seven overlapping clones. Indirect mapping data suggest that these two blocks are linked as shown in Figure 2. Unfortunately, we have been unable to obtain clones which physically span the two blocks. The ends of these two blocks contain identical pAT sequences in opposite orientations, creating an inverted repeat, which may be unstable upon cloning in phage vectors.

The use of defined carrier sequences has permitted us to map a large region of the chromosome capable of undergoing amplification without resorting to chromosome walking. The following conclusions may be drawn from the analysis of the physical structure of the transforming DNA in the parental cell line. First, the element we have cloned consists of an array of eight human plasmids into which a single aprt-tk plasmid has integrated. As noted above, the donor human plasmids were cleaved with Hind III and ligated prior to transformation. Each of the eight human plasmid clones remains joined to its neighbors by Hind III linkages. The linearized aprt-tk plasmid has integrated into this array of donor plasmids by two recombination events. One homologous recombination event has occurred between the pAT153 sequences adjacent to human insert 1, and the pBR322 sequences linked to the *aprt-tk* gene (see Figure 3). A second, apparently nonhomologous event has joined the 3′ end of the truncated *tk* gene to human sequence 3 with a concomitant loss of DNA from both recombining sequences.

Clones Derived from the Amplified Array

A comparison of clones derived from the parental and amplified cell lines should permit us to define the size of the amplified unit and the manner in which individual amplified units are linked to one another. From these data, we may deduce the physical structure of the amplified array.

We have analyzed over 50 clones derived from DNA from the amplified cell line. As expected, the majority of the phage clones reveal a sequence arrangement identical with that of clones obtained from the unamplified parental cell line. The amplified unit in JR4 extends for at least 75 kb. Lambda clones from this region with 15–20 kb inserts are likely to fall entirely within the individual amplified units, and will therefore be identical to parental clones. However, a number of clones are unique to amplified DNA and reveal sequence rearrangements which presumably result from the replication or recombination processes required to generate the amplified structure (Figure 4). Thus clones

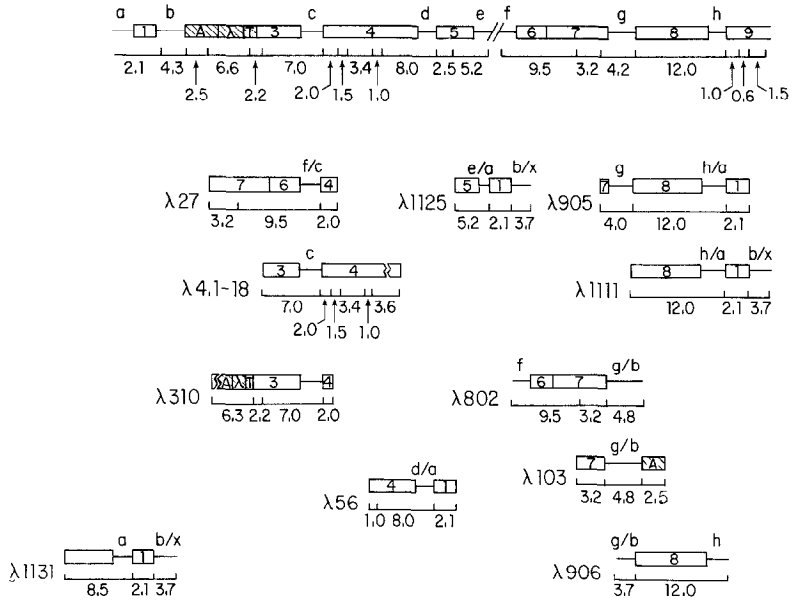

Figure 4. Lambda Clones Reveal DNA Rearrangements Unique to Amplified DNA

The linear map of the region of DNA which undergoes amplification is diagrammed at the top of the figure. Eco RI sites and fragment sizes are indicated. The map displays the linkage arrangement of the transforming sequences which undergo amplification. Plasmid pdLAT-3, which contains the *aprt* (A) and *tk* (T) genes is indicated by the hatched box (▨). The transformation to generate the parental cell was performed with plasmids containing human DNA inserts as the sole carrier. The linear map depicts the linkage of individual human plasmids with one another and with the *aprt-tk* genes. The pAT153 segment of the human plasmids is indicated by letters a–h. The human DNA inserts are indicated by boxes (□) numbered 1–9. Sequence b is actually the pBR322 vector component of pdLAT3.

Below the map are 11 λ clones unique to amplified DNA. Each clone reveals at least one recombination event which links two sequences many kilobases apart in unamplified DNA. Recombination events occur most frequently between two pAT sequences as indicated by a slash. For example, in λ905, recombination between pAT-h and pAT-a is depicted by h/a. pAT sesquences whose origins cannot be identified are indicated by an ×. Recombination within human sequences is indicated with a jagged line (⊏⊐). The human DNA sequence adjacent to insert 1 in λ1131 has not been identified.

are obtained which reveal the linkage of human DNA sequences not adjacent to one another in the parental cell line. These clones contain the junction between two amplified units and presumably reflect the recombination events which join individual units into an amplified array.

Three cogent features of the structure of the amplified array emerge from the analysis of 12 junction clones. First, individual units in the amplified DNA are tandemly joined to one another. No extraneous carrier or recipient cell DNA separates the individual units. Second, units within the amplified array are joined primarily by homologous recombination between repeated DNA sequences. Third, the amplified array is not a regularly repeating structure but consists of a tandem array of units of differing size. These conclusions derive from the analysis of junction clones described in detail below.

Simple Recombinant Clones

Clones from the amplified cell line which differ from those obtained in the parental cell presumably reflect the necessary replication or recombination events required to generate an amplified array. We have analyzed several clones from the amplified cell line which contain closely linked human sequences that are many kilobases apart in the unamplified parental cell line. These clones reflect the junction between one amplified unit and another. Most of these clones appear to result from a single homologous recombination event.

One example is provided by the organization of clone λ103 depicted in Figure 5. In the unamplified parent, human insert 7 is separated from human insert 8 by the plasmid sequences pAT-G. Similarly, the *aprt-tk* genes are separated from human insert 1 by pBR322 sequences. Clone λ103 from the amplified cell line contains human insert 7 linked to the *aprt-tk* genes via a homologous recombination event between pAT-G and pBR322. Restriction analysis of recombinants indicates that the recombination involves a precisely homologous exchange (data not shown). Clone λ103 reflects one of several simple homologous recombinations which serve to link multiple units containing the *aprt-tk* genes into an amplified array.

It is apparent that some recombination events will link two units, each containing the *aprt* and *tk* genes. Other events may exclude these genes from one or both of the recombined units. An example of one such recombination event is provided by clone λ27 shown in Figure 4. In this clone, a homologous recombination event has occurred between pAT-F and pAT-C, linking human inserts 4 and 6. The linkage generated by such a recombination event excludes the *aprt-tk* genes from both units created by this exchange. This clone is also of interest since the recombining pAT sequences are in opposite orientations in the

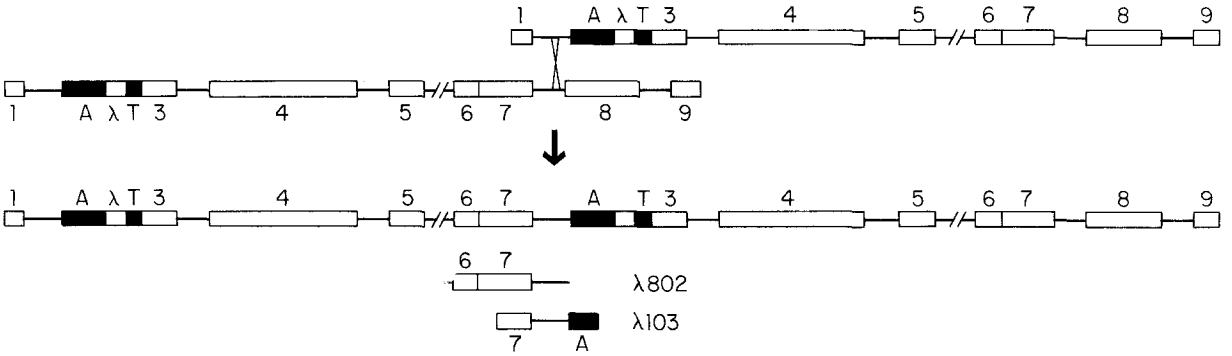

Figure 5. Recombination Joins Two Units into a Linear Array

Model depicting the origin of the sequence arrangement found in clones λ802 and λ103. Two individual amplified units pair so that recombination occurs between pBR322 sequences adjacent to the *aprt* gene and pAT153 sequences adjacent to human insert 7. One of the two recombinant products is a tandem array of two individual units in which human insert 7 becomes closely linked to the *aprt-tk* gene. The unique junction created by this recombination event is contained within λ103 and λ802.

unamplified parent. The recombination event therefore not only links the two units into a tandem array, but inverts the two units with respect to one another.

Analysis of eight additional clones which result from single recombination events indicated that recombination occurs by precise exchanges between the highly repeated pAT and pBR sequences. Two clones, 1111 and 1125, have arisen by two recombination events, each involving homologous exchange. Since recombination may occur at any site of homology through the unit, the ultimate array will consist of units of varying size. Genomic blotting detailed below confirms that the organization of sequences within these λ clones reflects the organization of amplified units in genomic DNA of the amplified cell line.

Recombination within Human DNA

Plasmid sequences are highly repeated within the amplified unit, and therefore are the predominant site of recombination. However, analysis of two clones demonstrates that not all recombinations occur between the large stretches of homology afforded by the repeated plasmid sequences (Figure 4). In λ4.1-18, recombination has occurred within human insert 4. In λ310, a recombination event has occurred at the 3′ end of the *aprt* gene. Since we have not characterized in detail the sequences involved in this recombination event, we do not known whether the recombination within the human insert involves short stretches of homology or nonhomologous exchange. From the collection of clones we have analyzed, it is apparent that recombination within human inserts is a far less frequent event than the precise homologous exchanges observed between plasmid sequences.

Amplification Is a Rapid Event and Occurs within a Single Generation

The amplified array may result from aberrant replication and multiple crossing-over events during a single cell cycle, or from the gradual expansion over a large number of cell generations. The analysis of clones λ103 and λ906 suggests that amplification has occurred in one generation

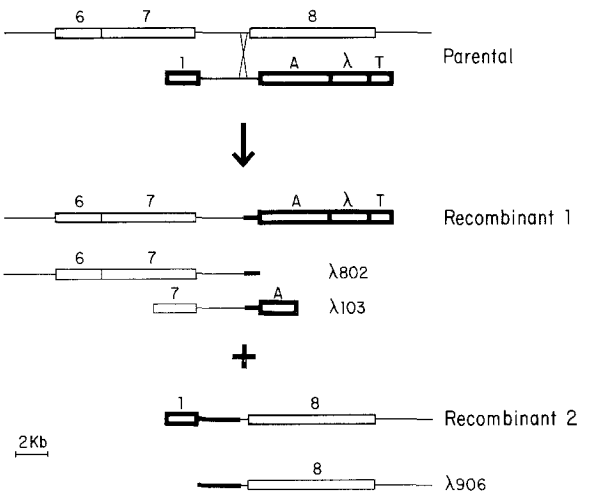

Figure 6. A Single Homologous Recombination Event and the Two Recombinant Products

A single homologous exchange, as depicted, generates two recombinant products. Both products are retained in the DNA of JR4 and are presented in λ clones 802, 103, and 906. For further explanation, see text.

(Figure 6). A single crossing-over event generates two recombinant DNA molecules. If the amplified array is generated by the gradual accumulation of multiple gene copies through unequal exchange, then at each generation the two products of an exchange should segregate at mitosis. In such a scheme, both products cannot be included in the same amplified array. However, clones λ103 and λ906 reflect the two products of a single recombination event. The observation that both products are present in a single array within a single cell suggests that aberrant replication and multiple recombinations are occurring within a single generation and the construction of an array is not a gradual process.

This argument must be tempered by the possibility that the two clones result from two independent crossover events at the same locus, simulating the products of a single recombination event. However, there are over 50 products of pAT–pAT pairings possible from the region we

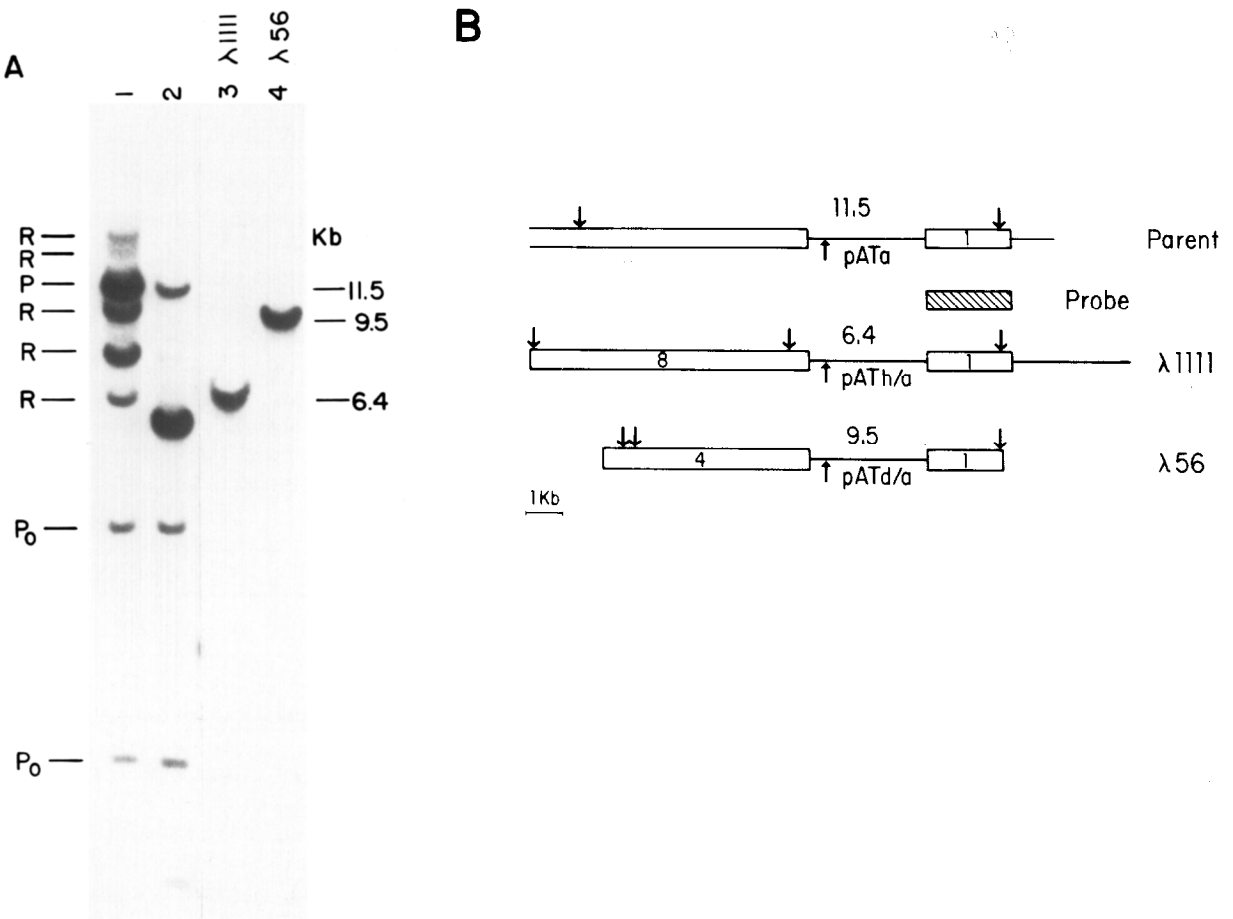

Figure 7. Analysis of the Amplification Unit in Genomic DNA

(A) DNA was isolated from the amplified cell line JR4 or two λ clones from a library of JR4, cleaved, electrophoresed through 0.8% agarose, blotted, and probed with ^{32}P-labeled human insert 1 DNA.

(Lane 1) Sac I digest of 10 μg of JR4 DNA. (Lane 2) Sac I/Sac I digest of 10 μg of JR4 DNA. (Lane 3) Sac I digest of 500 pg of λ1111. (Lane 4) Sac I digest of 500 pg of λ56.

(B) Map of unamplified parental DNA in the region of human insert 1 and two recombinant phage clones λ111 and λ56 derived from JR4. These λ clones correspond to those analyzed in (A). Arrows indicate Sac I sites (↓) and Sal I sites (↑).

have mapped, suggesting that repeated exchange at identical loci is unlikely.

Analysis of the Amplification Unit in Genomic DNA

The physical structure of the amplified array described thus far derives largely from a comparison of recombinant clones obtained from the libraries of parental and amplified cell DNA. We have used cloned sequences from the amplified array as probes to examine further the organization of the amplified DNA by genomic blot hybridization. Our analysis of phage clones is restricted to only 10 out of 20 or 30 possible junctions, which comprise the entire amplified array. Genomic blotting permits us to identify additional recombination events not represented in the λ clones we have analyzed. Furthermore, these experiments allow us to demonstrate that the organization of sequences in the λ clones accurately reflects the organization of the amplified DNA in the cellular genome and does not result from artifacts in the cloning process.

We have therefore used blot hybridization to map the individual amplified units in genomic DNA. Figure 7 depicts the organization of sequences surrounding human insert 1 in parental DNA. An asymmetrically placed Sac I site is present within human insert 1. If genomic DNA is cleaved with Sac I and probed with this insert, only fragments to the left of this site retain sufficient homology with the probe to be detected in blot hybridization. Cleavage of amplified DNA with Sac I will therefore generate three classes of fragments: a fragment (labeled P in Figure 7) from within the amplified array that is identical with the 11.5 kb parental fragment, reflecting individual units which have not recombined at this locus; fragments (labeled R in Figure 7) from within the amplified array that differ in size from the parental fragment, reflecting recombinations at this locus; and frag-

ments (labeled P_0 in Figure 7) homologous to human insert 1 that are present in unamplified parental DNA, but reside beyond the limits of amplification. Cleavage of genomic DNA from the amplified cell line with Sac I generates nine discrete fragments of different intensities. The two low molecular weight fragments (P_0 in Figure 7) derive from outside the region of amplification. The most intense fragment, 11.5 kb (P in Figure 7), reflects the parental DNA arrangement in this region. We expect this band to be most intense, since individual units which do not recombine within this region will preserve this Sac I fragment. The remaining six bands (R in Figure 7) result from recombination within the Sac I fragment to generate the junction between tandem units.

We have isolated two λ clones which reveal recombination within the plasmid sequences immediately adjacent to human insert 1. Each clone generates a different sized Sac I fragment (Figure 7), and each of these fragments is detected within genomic DNA. These data indicate that the recombinations revealed by analysis of λ clones indeed reflect rearrangements present within the genome; the recombinant phages do not result from cloning artifacts.

Four additional bands are observed which presumably reflect recombination events not identified in the clones we have isolated. Thus genomic blotting experiments identify four additional joints in this region.

Analysis of recombinant clones indicates that most joints linking individual units result from homologous recombination between plasmid sequences. If this is the case, the heterogeneity of Sac fragments observed in genomic blotting experiments should result from recombination in pAT-A sequences, creating novel linkages between human insert 1 and other human inserts. These exchanges should therefore preserve the linkage between human insert 1 and plasmid sequence.

A single Sac I site is present in human insert 1 and a single Sal I is present in the adjacent pAT sequence. Although cleavage of amplified DNA with Sac I generates a heterogeneous set of bands, cleavage with the enzymes Sac I + Sal I should generate a single fragment. Such a result would indicate that all recombinations within this region result from homologous exchange within pAT sequences. This prediction is confirmed in the blotting experiments shown in Figure 7. An intense 5 kb Sac I–Sal I band is observed which reflects the prevalence of homologous exchange. The additional band at 11.5 kb may result from a nonhomologous exchange in this region. The two low molecular bands, as noted above, derive from outside the region of amplification.

The heterogeneous Sac I bands are not of uniform intensity. The most intense band reflects the parental arrangement. Three other bands from within the array, at 16, 13, and 6.4 kb, are, as expected, of single copy intensity. However, two bands at 9.5 and 7 kb are present at approximately five copies per genome. These observations suggest that the replicative process is not complete before recombination occurs. Rearrangements that link units early in the amplification process will themselves be amplified by additional rounds of replication and recombination. Conclusions drawn from Southern blotting about the region including human insert 1 have been confirmed by analogous experiments probing the region surrounding human insert 7 (data not shown).

Finally, it is important to demonstrate that the irregular tandemly repetitious structure of amplified DNA is not a consequence of cellular heterogeneity within the cell line JR4, but truly describes the structure of amplified DNA in each cell of the JR4 population. Therefore, four subclones of JR4 were isolated and analyzed by genomic blotting. All four subclones were identical to the JR4 cell line (data not shown). Thus the pattern we observe does not reflect a heterogeneous population of cells, each containing different but regularly repeating units, but rather a homogeneous population of cells containing an irregular repeating array of amplified units.

The Existence of an Amplification Gradient

The observation that individual units within the array differ from one another prompted us to ask whether all sequences within the array are amplified to the same extent. Genomic DNA from the amplified cell line was digested with two different pairs of restriction enzymes that do not cut the plasmid sequences, blotted, and probed with pAT153. This procedure displays each distinct pAT from the array as a unique band. From the intensity of the hybridizing bands, we can obtain an accurate estimate of the extent of amplification of each pAT sequence within the entire array. The blotting data clearly indicate that different pAT sequences hybridize with different intensities. The intensity of hybridization, and therefore the extent of amplification, clearly depends upon the position within the unit. The extent of amplification of individual sequences along the physical map is shown in Figure 8. The pattern of hybridizations suggests a gradient of amplification, with centrally located pAT sequences amplified three or four times more than those sequences located at the periphery of the unit. This gradient of amplification may have its origin either in the replication or in the recombination steps required in the amplification process.

Discussion

The Physical Structure of Amplified DNA

Amplified DNA may exist in two physical states within the nucleus; either integrated into the chromosome or as an autonomously replicating extrachromosomal element. Integration frequently results in the clustering of amplified genes in expanded homogeneously staining regions on one of two homologous chromosomes (Biedler and Spengler, 1976; Nunberg et al., 1978; Dolnick et al., 1979). Amplified DNA can also be found on small paired extrachromosomal elements known as double minutes. Such elements are genetically unstable since they lack both centromeres and kinetochores (Kaufman et al., 1981;

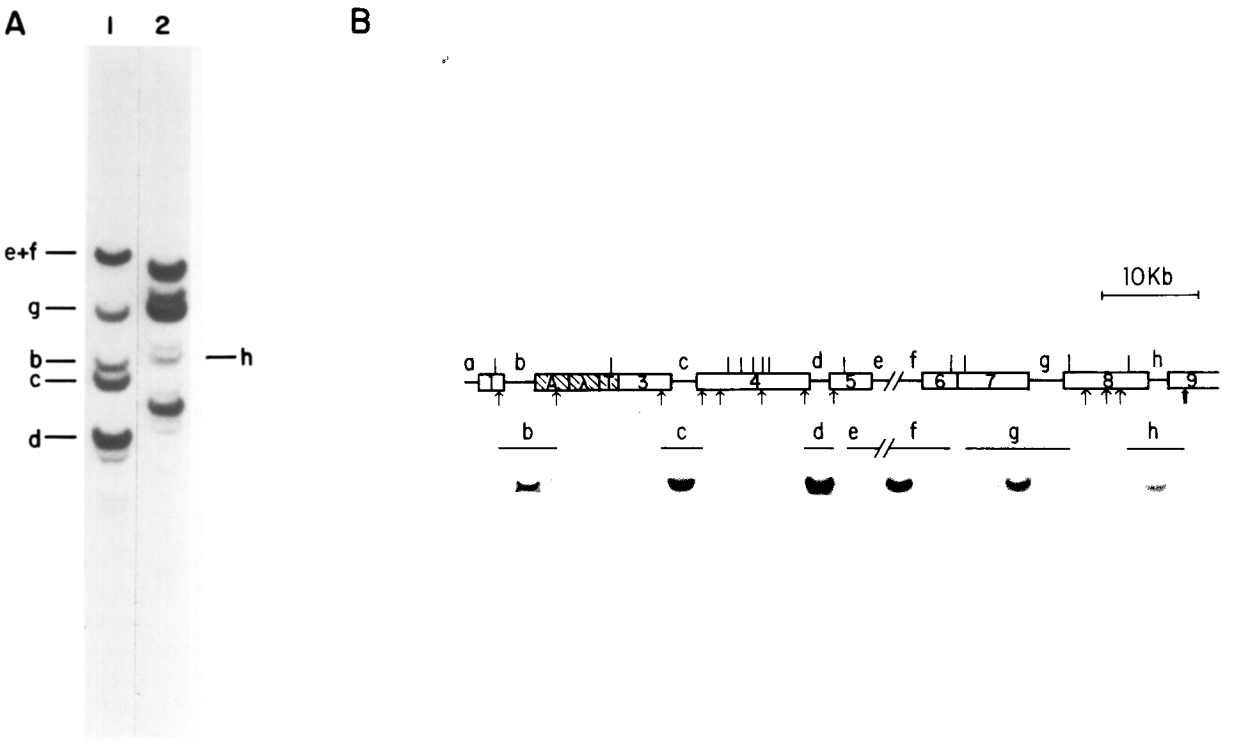

Figure 8. The Amplification Gradient

(A) Ten micrograms of high molecular weight genomic DNA from JR4 was cleaved with Sac I + Xba (lane 1) or Sac I + Kpn I (lane 2) (see 8B), electrophoresed through 0.5% agarose, blotted, and probed with ^{32}P-labeled pAT153 DNA. The bands are labeled b–h, according to their position within the amplified unit (as depicted in B).

(B) Map of the amplified unit. Sac I (↑), Xba (■), and Kpn (↑) sites are indicated. pAT bands from the Southern blot in (A) have been placed below their appropriate location within the amplified unit to illustrate the gradient of amplification.

Brown et al., 1981), and their presence may impose a selective disadvantage on the growth rate of the host cell (Haber and Schimke, 1981). Whether the amplified genes exist within or apart from the chromosome, they appear physically clustered in a contiguous array. A more complete picture of the molecular structure of amplified DNA has been difficult to obtain, since individual amplified units are frequently large, and may extend hundreds to thousands of kilobases beyond the selectable marker into sequences for which probes are unavailable (Schimke, 1982). Junctions which link individual units therefore reside large distances from the selected gene and are inaccessible without resorting to long distance chromosomal walking.

We have employed gene transfer to generate an amplified array consisting solely of a small number of cloned sequences. The small size of the amplified unit, along with the availability of cloned probes for every sequence within the array, permits us to determine the physical organization of amplified DNA in our transformed cell lines.

The amplified DNA we have analyzed consists of a tandem array of at least 20 individual repeating units. The individual units are contiguous, and are joined to one another by homologous recombination between repeated sequences. At first approximation, all homologous recombinations are permitted, such that crossing-over may occur between any two repeated sequences. Since individual units contain multiple repeated elements, the array is not a regularly repeating structure; the individual units within the array are heterogeneous both in size and in sequence content. Finally, the extent of amplification is not equal throughout the unit. A gradient of amplification exists, with sequences at the center of the unit amplified 3–4 times more than those at the periphery. These structural features of amplified DNA are illustrated in Figure 8.

Is Amplification a Rapid or Gradual Process?

What can we deduce about the mechanisms of gene amplification from the structure of amplified DNA? Two general mechanisms have been invoked to explain the events responsible for gene amplification. Gradual processes, such as unequal crossing-over between sister chromatids, have been suggested to explain the origin and evolution of the tandem arrays of repeating units, which, for example, characterize the ribosomal RNA genes of most vertebrate species (Brown and Blackler, 1972; Smith, 1973; Tartoff, 1975). Similar mechanisms have been invoked to explain the slow step-wise selection required to generate amplified cell lines resistant to specific inhibitors of enzyme activity (Alt et al., 1978).

Rapid amplification events may involve localized bursts of replication which escape normal control mechanisms and allow a sequence to be recognized and copied by the replicative machinery more than once per generation (Alt et al., 1978; Wahl et al., 1979; Botchan et al., 1978; Roberts and Axel, 1982). One unique example involves the "rolling circle" replication which amplifies ribosomal DNA during Xenopus oogenesis (Brown and Blackler, 1972; Hourcade et al., 1973). Two observations suggest that the array of amplified *aprt-tk* genes we have generated results from a rapid mechanism of DNA amplification. First, the unamplified $aprt^+$ tk^- parent cell is grown in BUdR to ensure the maintenance of the tk^- phenotype. The cells are removed from BUdR only three generations prior to imposing selection for gene amplification (Roberts and Axel, 1982). Second, if the amplified array is generated by the gradual accumulation of multiple gene copies through unequal crossing-over, we would predict that the two products of an unequal exchange would segregate during mitosis, and therefore both could not be included in a single amplified array. In at least one instance, we observe that the two products of a single exchange are both present in the amplified array. The observation that segregation is not occurring suggests that multiple recombinations have occurred in a single generation.

Mechanisms of Gene Amplification

What molecular mechanisms can generate the amplified structure we have characterized? Models which result in the generation of multiple identical copies through simple replicative processes perhaps analogous to the "rolling circle" are inconsistent with our findings. The array is irregular and its construction is likely to involve two events: multiple cycles of unscheduled DNA replication at a single locus, followed by multiple recombination events which serve to link individual units to one another and, ultimately, to the chromosome.

One particularly attractive model invokes "local polytenization" to generate an intermediate structure resembling an onion skin (Botchan et al., 1978). Thus an origin of replication would escape normal growth control and permit a localized region of DNA to undergo multiple rounds of replication within a single cell cycle. The resultant structure, illustrated in Figure 9, would consist of multiple unintegrated copies of DNA held together only by hydrogen bonding. This structure must ultimately resolve into a tandem array prior to the next mitosis. Resolution therefore requires multiple recombination events between individual duplexes, linking the individual units into a linear structure. This structure can integrate through recombination with homologous sequences in the chromosome. We have demonstrated that exchanges linking the individual units almost always involve homologous sequences. If the amplified DNA consists of multiple regions of homology, then recombination may occur at any one of several repeated sequences within this unit, resulting in an irregularly repeated structure. Alternatively, recombination within a single DNA duplex would generate free circles, which may be the precursor of extrachromosomal elements or double minutes. This may be analogous to the mechanism of amplification and excision of circular viral DNA from an integrated copy of a viral genome (Botchan et al., 1978).

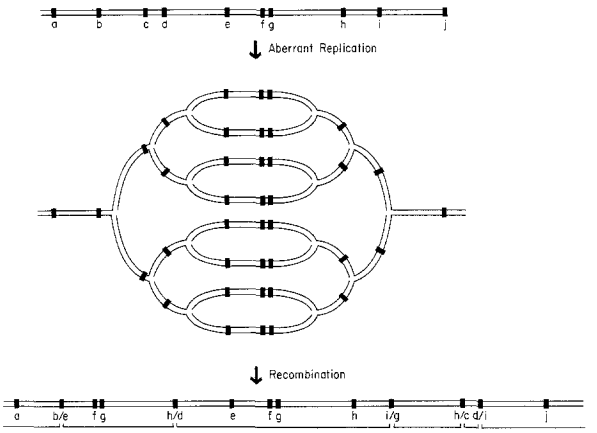

Figure 9. Schematic Model of Gene Amplification

Gene amplification proceeds through two steps. Aberrant replication results in the replication of a restricted chromosomal region many times in one cell cycle. Three rounds of replication are depicted here. The resulting structure resolves into an irregular linear array via homologous recombination between repeated DNA sequences (■) (see text).

The processes of replication and recombination may occur simultaneously. The generation of an amplified array does not appear to proceed by first replicating the full complement of amplified units and subsequently recombining them to one another. Rather, the replication and recombination steps are intermingled. Junction fragments linking individual units are not all amplified to the same extent. These results suggest that recombinations which occurred early in the amplification process will generate novel joints which will be further amplified by additional rounds of replication.

The Amplification Gradient

Not all sequences within the array are equally amplified. Hybridization experiments reveal a gradient of amplification, with centrally located plasmid sequences amplified four times more than those sequences located at the periphery of the unit. What is the origin of this amplification gradient? The gradient may result from repeated replication events initiating from a fixed point or origin but terminating at varying distances from this origin. This will generate a nested set of replication forks in which the individual units of replication vary in length by virtue of terminations at random distances from the origin. This model has been proposed to explain the gradient observed upon chorion gene amplification during Drosophila development. Spradling (1981) has demonstrated that sequences proximal to a putative chromosomal origin of replication are amplified to a greater extent than those more distal to this site.

The gradient of amplification need not reflect the replicative steps, but may result from recombination events as

well. Consider the possibility that replication generates multiple units which initiate and terminate identically. If recombination is permitted with equal probability at any of multiple repeated elements within the unit, modeling predicts a tandem array which will also generate a gradient of amplification proceeding outward from the center of the unit. At present, we cannot determine the potential contributions of the replicative or recombination events in the generation of the gradient.

In conclusion, we have exploited gene transfer to construct a small amplified array consisting of defined DNA sequences. The amplification is rapid, occurring in a single cell division, and requires two interrelated processes, replication and recombination. These events generate a tandem array of irregularly repeated units linked together through homologous, but unequal, crossing-over events. We have determined the physical structure of amplified DNA which has arisen subsequent to gene transfer. It is likely that amplification of endogenous cellular genes may proceed via similar mechanisms, involving repetitive genomic sequences, to generate a structure resembling the array we have described. Finally, we emphasize that the elucidation of the structure of amplified DNA does not permit us to deduce the precise mechanisms through which it is generated. Numerous problems remain. What initiates the aberrant firing of a single origin several times per cell cycle? Is a unique origin employed to generate all DNA copies within the repeating array? Does amplification proceed from a unique point to fixed or variable termini? What is the nature of the recombination events required to resolve the array, and finally, what determines whether amplified sequences will reside extrachromosomally or integrated with the host chromosome?

Experimental Procedures

Cell Culture

Cell lines were selected, cloned, and maintained as previously described (Roberts and Axel, 1982). Briefly, tk^- $aprt^-$ L cells were transformed with 20 ng of Sal I–linearized pdAT3, and cotransformed with 1 μg of carrier DNA. The carrier DNA was prepared in the following manner: Human spleen DNA was cut to completion with Hind III and Bam HI, and randomly cloned into Hind III–Bam HI–cut pAT153. From the resulting plasmid library, 20 clones containing 3–15 kb inserts were isolated, cut with Hind III, and ligated together, creating a random high molecular weight heteropolymer of plasmid and human DNA sequences. This DNA was then used as carrier in the above described transformation.

Library Construction and Screening

Charon 4 libraries of the parental and amplified cell lines were prepared from high molecular weight cellular DNA as described by Scheller et al. (1982) and Maniatis et al. (1978).

Blot Hybridization

Total cellular DNA was electrophoresed, blotted, and probed as previously described (Roberts and Axel, 1982).

Acknowledgments

We wish to thank Julie DeMartino, Jane Onken, and John Fisher for technical assistance. This work is supported by a fellowship from the Cancer Research Institute, Inc., and the Seath Sprague Educational and Charitable Foundation, New York, New York (L. B. B.), and grants from the National Institutes of Health (R. A.).

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

Received January 11, 1983; revised March 8, 1983

References

Alt, F. W., Kellems, R., Bertino, J., and Schimke, R. T. (1978). Selective multiplication of dihydrofolate reductase genes in methotrexate resistant variants of cultured murine cells. J. Biol. Chem. *253,* 1357–1370.

Biedler, J., and Spengler, B. (1976). Metaphase chromosome anomaly: association with drug resistance and cell-specific products. Science *191,* 185–187.

Botchan, M., Topp, W., and Sambrook, J. (1978). Studies on simian virus 40 excision from cellular chromosomes. Cold Spring Harbor Symp. Quant. Biol. *43,* 708–720.

Brennand, J., Chinault, A. C., Konecki, D. S., Melton, D. W., and Caskey, C. T. (1982). Cloned cDNA sequences of the hypoxanthine/guanine phosphoribosyltransferase gene from a mouse neuroblastoma cell line found to have amplified genomic sequences. Proc. Natl. Acad. Sci. USA *79,* 1950–1954.

Brown, D., and Blackler, A. (1972). Gene amplification proceeds by a chromosome copy mechanism. J. Mol. Biol. *63,* 75–83.

Brown, D., and Dawid, I. (1968). Specific gene amplification in oocytes. Science *160,* 272–280.

Brown, P., Beverley, S., and Schimke, R. (1981). Relationship of amplified dihydrofolate reductase genes to double minute chromosomes in unstably resistant mouse fibroblast cell lines. Mol. Cell. Biol. *1,* 1077–1083.

Dolnick, B., Berenson, R., Bertino, J., Kaufman, R., Nunberg, J., and Schimke, R. (1979). Correlation of dihydrofolate reductase elevation with gene amplification in a homogeneously staining chromosomal region in LS1784 cells. J. Cell. Biol. *83,* 394–402.

Gall, J. (1974). Free ribosomal RNA genes in the macronucleus of *Tetrahymena.* Proc. Nat. Acad. Sci. USA *71,* 3078–3081.

Glover, D., Zaha, A., Stocker, A., Santelli, R., Pueyo, M., Toledo, S., and Lara, F. (1982). Gene amplification in *Rhyorchosciara* salivary gland chromosomes. Proc. Nat. Acad. Sci. USA *79,* 2947–2751.

Haber, D. A., and Schimke, R. T. (1981). Unstable amplification of an altered dihydrofolate reductase gene associated with double-minute chromosomes. Cell *26,* 355–362.

Hamlin, J., Montoya-Zavala, M., Heintz, N., Millbrandt, J., and Azizkahn, J. (1982). Studies on the mechanism of dihydrofolate reductase gene amplification in Chinese hamster ovary cells. In Gene Amplification, R. Schimke, ed. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 155–159.

Hood, L., Campbell, J., and Elgin, S. (1975). Multigene families. Ann. Rev. Genet. *9,* 305–354.

Hourcade, D., Dressler, D., and Wolfson, J. (1973). The nucleolus and the rolling circle. Cold Spring Harbor Symp. Quant. Biol. *38,* 537–550.

Kaufman, R., Brown, P., and Schimke, R. (1981). Loss and stabilization of amplified dihydrofolate reductase genes in mouse sarcoma S-180 cell lines. Mol. Cell. Biol. *1,* 1084–1093.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell *15,* 687–701.

Nunberg, J., Kaufman, R., Schimke, R., Urlaub, G., and Chasin, L. (1978). Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate resistant Chinese hamster ovary cell line. Proc. Nat. Acad. Sci. USA *75,* 5553–5556.

Perucho, M., Hanahan, D., and Wigler, M. (1980). Genetic and physical linkage of exogenous sequences in transformed cells. Cell *22,* 309–317.

Roberts, J. M., and Axel, R. (1982). Gene amplification and gene correction in somatic cells. Cell *29,* 109–119.

Robins, D. M., Ripley, S., Henderson, A. S., and Axel, R. (1981). Transforming DNA integrates into the host chromosome. Cell *23,* 29–39.

Scheller, R. H., Jackson, J. F., McAllister, L. B., Schwartz, J. H., Kandel, E. R., and Axel, R. (1982). A family of genes that codes for ELH, a neuropeptide eliciting a stereotyped pattern of behavior in Aplysia. Cell *28,* 707–719.

Schimke, R. (1982). Gene Amplification, (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Smith, G. (1973). Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. *38,* 507–513.

Spradling, A. C. (1981). The organization and amplification of two chromosomal domains containing Drosophila chorion genes. Cell *27,* 193–201.

Spradling, A. C., and Mahowald, A. P. (1981). A chromosome inversion alters the pattern of specific DNA replication in Drosophila follicle cells. Cell *27,* 203–209.

Tartoff, K. (1975). Redundant genes. Ann. Rev. Genet. *9,* 355–387.

Wahl, G., Padgett, R., and Stark, G. (1979). Gene amplification causes overproduction of the first three enzymes of UMP synthesis in N-(phosphoroacetyl)-L-aspartate resistant hamster cells. J. Biol. Chem. *254,* 8679–8689.

Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S., and Axel, R. (1979). Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell *16,* 777–785.

Yao, M., Kimmel, A., and Gorovsky, M. (1974). Comparison of the sequences of macro and micronuclear DNA of *Tetrahymena pyriformis*. Chromosoma *48,* 1–18.