MOLECULAR AND CELLULAR BIOLOGY, Feb. 1987, p. 569–577
0270-7306/87/020569-09$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 7, No. 2

Isolation of the Amplified Dihydrofolate Reductase Domain from Methotrexate-Resistant Chinese Hamster Ovary Cells

JAMES E. LOONEY AND JOYCE L. HAMLIN*

Department of Biochemistry and Cell and Molecular Biology Program, University of Virginia School of Medicine, Charlottesville, Virginia 22908

Received 2 September 1986/Accepted 5 November 1986

We isolated overlapping recombinant cosmids that represent the equivalent of two complete dihydrofolate reductase amplicon types from the methotrexate-resistant CHO cell line CHOC400. The type I amplicons are 260 kilobases long, are arranged in head-to-tail fashion, and represent 10 to 15% of the amplicons in the CHOC400 genome. The type II amplicons are 220 kilobases long, are arranged in head-to-head and tail-to-tail configurations, and constituted the majority of the remaining amplicons in CHOC400 cells. The type II amplicon sequences are represented entirely within the type I unit. These are the first complete amplicons to be cloned from a mammalian cell line.

DNA sequence amplification is a phenomenon that has been observed in a variety of transformed mammalian cells, including those of murine, hamster, rat, and human origin (for reviews, see references 15, 26, and 28). Amplification in mammalian systems was first described for drug-resistant Chinese hamster and murine cell lines that had been subjected to incremental increases in the concentration of methotrexate (MTX) over an extended time period (2, 4, 23). Amplification of the gene encoding the target enzyme, dihydrofolate reductase (DHFR), results in increased expression of the enzyme, providing a selective growth advantage in high levels of MTX.

In the last several years, a variety of other inhibitors have been used to select cell lines that have amplified the gene encoding the particular target enzyme (reviewed in references 15, 26, and 28). In addition, it has been shown that several human neoplasms contain multiple copies of known cellular oncogenes, the overexpression of which presumably imparts a selective growth advantage in the context of the organism (1, 6, 8, 27). DNA sequence amplification is therefore a relatively widespread phenomenon that can affect many loci.

In cell lines that are extremely resistant to selective agents, the copy number of the corresponding gene can be as high as 3,000 (31) and the amplified DNA can represent several percent of the total genomic DNA content (4, 5, 21, 23, 30). High levels of amplification are almost invariably accompanied by pronounced chromosomal aberrations in the form of extrachromosomal, acentromeric double minutes or expanded, abnormally banding chromosome arms (reviewed in references 7, 15, and 28). From estimates of the *dhfr* gene copy number and the length of the abnormally banding chromosome regions in MTX-resistant Chinese hamster ovary (CHO) cells, it is clear that the unit of amplification (amplicon) in this system is much larger than the 25-kilobase (kb) *dhfr* gene itself and could be as large as 500 kb in some cases (23). In other drug-resistant cell lines and in some human tumours, estimates for the length of amplicons range as high as 3,000 kb (5).

Very little is known about the mechanisms responsible for DNA sequence amplification in mammalian cells. It has been suggested that amplification could result from unequal sister chromatid exchange, from rolling-circle replication, or from a form of onion skin overreplication in which the extra replicons are eventually arrayed end to end (15, 25, 26, 28). In the unequal sister chromatid exchange and onion skin replication models, recombination events must play a critical role in fixing the extra DNA into tandem arrays, and the joints between amplicons necessarily define the sites at which recombination events occurred.

In an effort to define the structure of amplified DNA sequences, including the joints between amplicons, we began a project several years ago to isolate an amplicon equivalent from an MTX-resistant CHO cell line (CHOC400) that was developed in our laboratory (21). This cell line contains 1,000 to 1,200 *dHFR* genes, which are manifested as stable, abnormally banding chromosomal regions. We first cloned the *dHFR* gene itself, which was isolated intact in a single cosmid by a phenotypic selection method (20). We also isolated clones representing the locus in which DNA replication initiates in the amplified *dHFR* domain (16) and connected this locus to the gene in several overlapping cosmids (22). By chromosome walking, we progressed outward from this core sequence and have now isolated recombinant cosmids representing the equivalent of two complete amplicon types from the CHOC400 cell line that differ in size, end-to-end arrangement, and representation in the genome.

MATERIALS AND METHODS

Preparation and screening of cosmid library and mapping procedures. High-molecular-weight genomic DNA (>100 kb) from cultured CHOC400 cells was prepared by the method of Gross-Bellard et al. (13), partially digested with *Sau*3A1, and fractionated on a 5 to 24% NaCl gradient in 10 mM EDTA, pH 8.0. The gradient was centrifuged in an SW27 rotor at 20,000 rpm for 15 h at 20°C, and the 35- to 45-kb fraction (assessed on an 0.5% agarose gel) was dialyzed and precipitated with ethanol. The digest was suspended and treated with calf intestinal phosphatase (19), followed by heat inactivation, phenol-chloroform extractions, and ethanol precipitation. After resuspension, the genomic digest was ligated to *Bam*HI-cleaved pHC79 at a vector-to-genomic DNA weight ratio of approximately 1:10 (total DNA concentration, 300 μg/ml) for 16 h at 12°C. The ligation mixture was incubated in an in vitro packaging reaction

* Corresponding author.

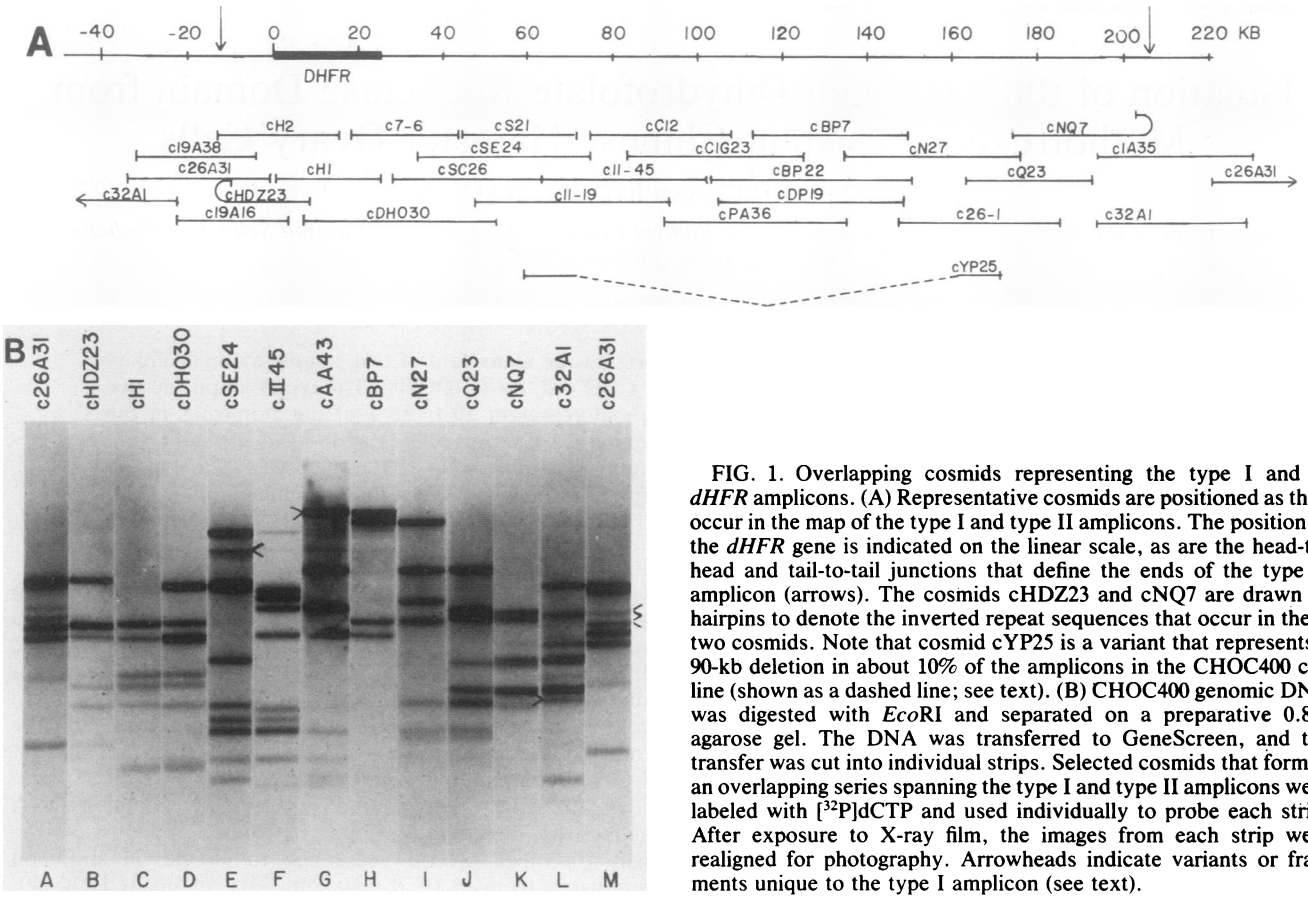

FIG. 1. Overlapping cosmids representing the type I and II *dHFR* amplicons. (A) Representative cosmids are positioned as they occur in the map of the type I and type II amplicons. The position of the *dHFR* gene is indicated on the linear scale, as are the head-to-head and tail-to-tail junctions that define the ends of the type II amplicon (arrows). The cosmids cHDZ23 and cNQ7 are drawn as hairpins to denote the inverted repeat sequences that occur in these two cosmids. Note that cosmid cYP25 is a variant that represents a 90-kb deletion in about 10% of the amplicons in the CHOC400 cell line (shown as a dashed line; see text). (B) CHOC400 genomic DNA was digested with *Eco*RI and separated on a preparative 0.8% agarose gel. The DNA was transferred to GeneScreen, and the transfer was cut into individual strips. Selected cosmids that formed an overlapping series spanning the type I and type II amplicons were labeled with [^{32}P]dCTP and used individually to probe each strip. After exposure to X-ray film, the images from each strip were realigned for photography. Arrowheads indicate variants or fragments unique to the type I amplicon (see text).

mixture (Gigapack; Vector Cloning Systems) according to the recommendations of the supplier, and the packaged cosmids were absorbed to *Escherichia coli* HB101. Colonies were plated, replica filters were prepared on Whatman 541 paper, and the filters were treated, prehybridized, and hybridized with appropriate probes as described by Grunstein and Hogness (14). Colonies giving moderate and strong hybridization signals on X-ray film were streaked onto new plates and rescreened with the same probes. Minilysates were prepared from positive clones by the method of Ish-Horowicz and Burke (17), digested with appropriate enzymes, and electrophoresed on agarose and transferred to GeneScreen (New England Nuclear Corp.). The transfers were hybridized successively with pHC79 vector and with the probe used for screening. Clones with more than one vector were discarded at this stage. A portion of the minilysate from the remaining clones was labeled by nick-translation with [^{32}P]dCTP and used to probe genomic digests of CHO and CHOC400 DNA to determine whether the clones were derived from amplified DNA and to assess the restriction pattern of DNA contained in the cosmid insert. Large-scale cosmid preparations were made by the method of Ish-Horowicz and Burke (17). Cosmids were mapped with *Eco*RI and *Hin*dIII as described previously (20).

Southern transfers and hybridization probes. Genomic DNA, cosmids, or minilysates were digested with appropriate enzymes in the buffers recommended by the supplier (Bethesda Research Laboratories), separated on 0.8% agarose gels, and transferred to GeneScreen as described previously (22). Restriction fragments purified from agarose by electroelution were labeled with [^{32}P]dCTP (3,000 Ci/mmol; ICN Pharmaceuticals) by nick translation (24). Fragments separated on low-melting-point agarose (Bethesda Research Laboratories) were labeled directly in the agarose by the random primer method (10). For mixed-probe experiments, equal amounts of both purified fragments were labeled in the same random primer labeling mixture. Specific radioactivities were in the range of 0.2×10^9 to 2×10^9 dpm/μg of DNA.

RESULTS

Isolation of the type I and type II amplicons. We previously reported the isolation of overlapping cosmid clones representing approximately 150 kb of DNA from the amplified *dHFR* domain in the MTX-resistant cell line CHOC400 (22). We subsequently learned that one of the cosmids used to construct the map in this region contained two different inserts alternating with two pHC79 vector sequences; both inserts derived from amplified genomic DNA, but were shown not to be contiguous in the genome (J. E. Looney, unpublished observations). Thus, with this cosmid, we had made an erroneous connection in the map between cosmids c26-1 and cH2 (Fig. 1A). Elimination of c8-B and c26-1 from the contiguous map reduced its length to approximately 120 kb (Fig. 1A) (22). This error was not detected previously because we did not have other cosmids mapping to the 5′ side of cH1 to confirm the sequence arrangement in this region.

To eliminate such artifacts from subsequent studies, we prepared a new library from a sized, phosphatase-treated, *Sau*3A partial genomic digest of CHOC400 DNA. By this approach, we hoped to avoid the ligation of two genomic fragments and to eliminate the possibility of a cosmid with two vectors and two inserts. Based on the heterogeneity observed in other systems (3, 9, 32), it was also clear that several nonsibling cosmids would have to be isolated in every screening step to obtain a true representation of the average amplicon structure(s) in the CHOC400 cell line.

To clone the entire amplicon, we began by screening this library with small fragments from the 5′ and 3′ ends of the 120-kb consensus region extending from cH2 to cC12 (Fig. 1A). In a typical walking step in which 2,000 to 4,000 colonies were screened with a small end fragment, 1 to 2% of the colonies gave moderate or strong hybridization signals. Minilysates were prepared from positive clones, and restriction digests of these were analyzed in hybridization studies to eliminate cosmids that did not derive from amplified DNA (<5%) or which contained multiple vectors (~5%). The average insert size of the remaining clones was in the range of 35 to 45 kb. The clones isolated in a typical screening step included approximately 40% that extended in the direction of the walk, 40% that extended in the opposite direction, and about 20% that were centered around the hybridization probe. Thus, at each step, five to six clones representing new sequences were isolated and previously isolated sequences were confirmed with five to six additional clones.

By this chromosome-walking procedure, we were able to extend the map in the 5′ direction from cH2 for another 5 kb and in the 3′ direction from cosmid cC12 for an additional 95 kb (see map, Fig. 1A). We also rescreened the new library with probes from the previously isolated 120-kb consensus sequence to confirm the map in this region. Thus, the 220-kb sequence from −15 to +205 kb (Fig. 1A) has been overlapped with at least 6 (and usually 10 to 12) nonsibling cosmids at every locus. Except for a few minor variant sequence arrangements (see discussion of variants below), all cosmids isolated in a given screening conformed to the unique map shown in Fig. 1A, indicating that the majority of the amplicons in CHOC400 cells have this sequence arrangement. In addition, all of the sequences from −15 to +205 appeared to be amplified to a similar degree as the gene itself, based on the relative hybridization signals observed when each clone was used to probe digests of CHOC400 genomic DNA (Fig. 1B, lanes B to K, and see below).

When end fragments from cHDZ23 and cNQ7 were used as probes on the cosmid library to walk further in the 5′ and 3′ directions, respectively, we obtained an unexpected result. Although the number of positive clones isolated was still in the range of 1 to 2%, the majority of these clones did not contain new sequences that extended in the direction of the walk. Rather, most of the inserts extended in the opposite direction and were similar to sequences already cloned. Several different screenings gave the same small number of new clones, indicating a sudden drop in the copy number of sequences mapping to the 5′ and 3′ sides of cHDZ23 and cNQ7, respectively. *Eco*RI digests were made of several cosmids that were isolated in walks from cHDZ23 and cNQ7 (Fig. 2). Using an end fragment from cNQ7 (lane G), we obtained 20 cosmids that resembled cNQ7 itself (5 are shown in Fig. 2, lanes B to F) and only 3 cosmids that appeared to contain additional sequences not present in cNQ7 (e.g., c32A1 and c1A40, lanes H and I). Similarly, in walking from the 5′ end of cHDZ23 (lane M), 18 cosmids were obtained that were very similar to cHDZ23 (4 are shown in lanes N to Q) and only 3 that appeared to have significant amounts of new sequence mapping beyond the 5′ end of cHDZ23 (e.g., c26A31 and c19A16, lanes K and L).

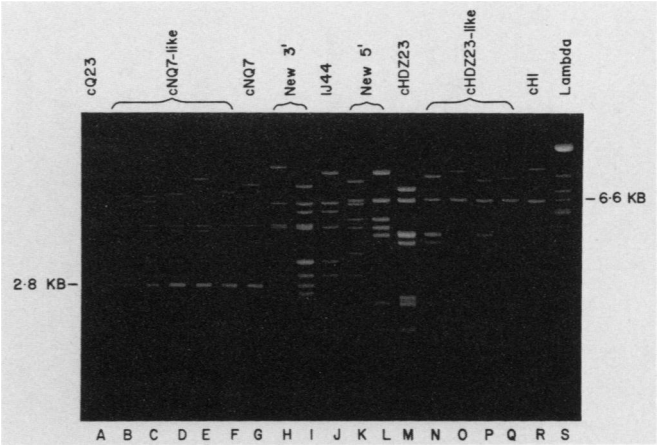

FIG. 2. Cosmids from the junction region in the type I and II amplicons. Representative cosmids recovered in screening to the 3′ side of cNQ7 and the 5′ side of cHDZ23 were digested with *Eco*RI and separated on a 0.8% agarose gel. Lanes: A, cQ23; B, c23A40; C, c21A23; D, c15A16; E, c10A39; F, c10A19; G, cNQ7; H, c32A1; I, cIA40; J, cIJ44; K, c26A31; L, c19A16; M, cHDZ23; N, c27A37; O, c11A37; P, c18A14; Q, c10A42; R, cH1; S, *Eco*RI-*Hin*dIII digest of lambda DNA. The 2.8-kb and 6.6-kb *Eco*RI fragments indicated are duplicated fragments detected in mapping and hybridization studies on these cosmids (see text).

However, the three new cosmids that were isolated from the 5′ end of the amplicon formed an internally consistent map with respect to each other, as did the three new clones that were isolated from the 3′ end of the amplicon (mapping data not show). Thus, there did not appear to be heterogeneity in sequence arrangement at the two ends of the amplicon.

Structure of the type I amplicon. The cosmids c26A31 and c32A1 were among the few new clones obtained initially by walking outward in the 5′ and 3′ directions from cHDZ23 and cNQ7, respectively. The cosmids c26A31 and c32A1 appeared to have two or three similar-sized restriction fragments in common, suggesting that they might be connected in the genome (compare lanes H and K, Fig. 2). This suggestion was confirmed in two additional ways: (i) a 6.5-kb fragment (fragment R; see Fig. 5A) from the 5′ cosmid c26A31 hybridized with a fragment attached to the vector in the 3′ cosmid c32A1 (hybridization data not shown), and (ii) when fragment R from c26A31 and fragment O from c32A1 (see Fig. 5A) were used separately to screen duplicate filters containing clones from the CHOC400 library, about 30% of the clones that hybridized to one probe hybridized to both probes; these cosmids (e.g., cIJ44, lane J, Fig. 2) contained several fragments common to both c26A31 and c32A1 that straddled the region between these two clones (see Fig. 5A).

Representative overlapping cosmids from the amplicon were hybridized individually to *Eco*RI digests of CHOC400 genomic DNA (Fig. 1B). The digests probed with each cosmid appear in the same order in the figure as the cosmids occurred in the map. Each successive cosmid hybridized to fragments that were shared with both the preceding and following clones, and overlaps between cosmids were continuous beginning and ending with c26A31. The map of the type I amplicon has therefore been closed and, by definition, is circularly permuted. It follows that the type I amplicons

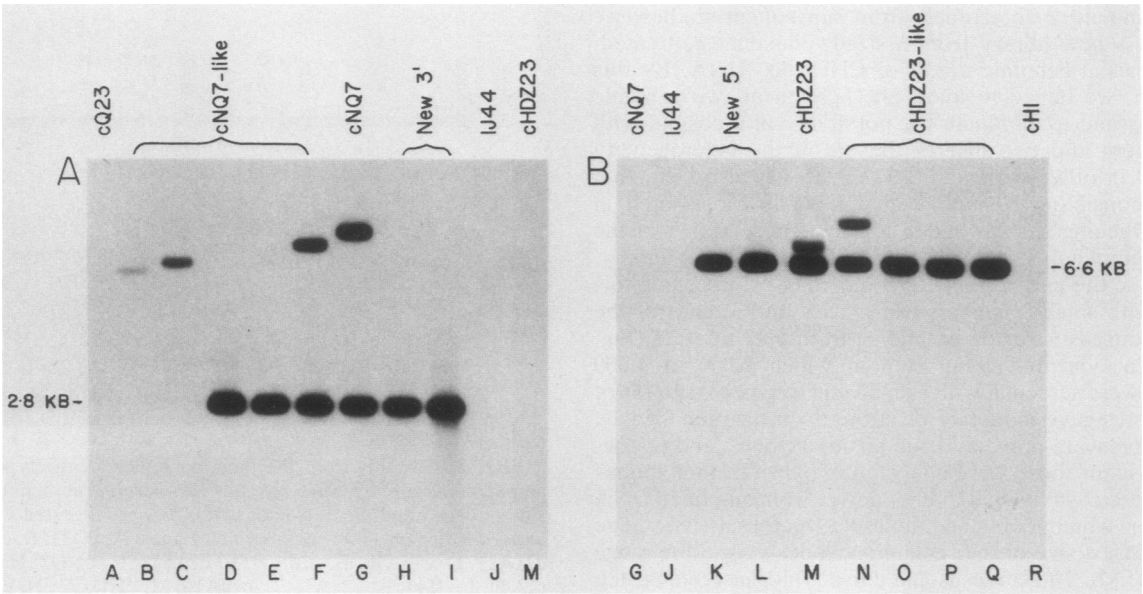

FIG. 3. Cosmids cNQ7 and cHDZ23 contain internal duplications. *Eco*RI digests of the cosmids shown in Fig. 2 were separated on duplicate 0.8% agarose gels and transferred to GeneScreen. A 2.8-kb *Eco*RI fragment from cNQ7 (A) and a 6.6-kb *Eco*RI fragment from cHDZ23 (B) were labeled with [^{32}P]dCTP and hybridized to each transfer. The letters below each lane signify the same cosmids as in Fig. 2.

are arranged in a head-to-tail configuration. By summing the lengths of nonoverlapping fragments in the individual cosmids used to construct the map shown in Fig. 1A, we estimate that the type I amplicon is approximately 260 kb long.

To date, we have not been able to determine the position(s) of the junction(s) between the type I amplicons. A junction fragment in the CHOC400 genome would join two parental restriction fragments. Such a junction fragment would therefore normally hybridize to two fragments of different length in CHO DNA. After hybridization of cosmids to parental CHO and CHOC400 genomic DNA digests, we did not detect any fragments with this property in the 195-kb region extending from cH2 to c26-1 (Fig. 1A) (22; J. L. Looney, B. Troutman, and J. L. Hamlin, unpublished data). This analysis is incomplete, however, since small restriction fragments give very low (and sometimes undetectable) hybridization signals on parental DNA when cosmids are used as probes. We also tested all of the individual *Eco*RI fragments >1 kb in length in the remainder of the amplicon (represented by cNQ7, c32A1, and c26A31; see Fig. 5A), and have not yet found a fragment in this region that could be interpreted as a junction fragment. The junction(s) may therefore occur in a smaller fragment(s) that we have not yet tested. Alternatively, it is possible that junctions occur at many different positions in the larger type I amplicons and that we have not been able to isolate cosmids that contain any of these heterogeneous junction fragments.

Structure of the type II amplicon. It is apparent that the hybridization signals from some of the genomic fragments recognized by individual cosmids were much weaker than the signals from other fragments detected by the same cosmid (Fig. 1B). We know from mapping studies that most of the fragments that give lower hybridization signals than expected are truncated in the cosmid by attachment to vector (e.g., the third band in cSE24, the fourth band in cQ23, and the fifth band in c32A1, Fig. 1B). However, the fragments indicated by arrowheads in the hybridization patterns of c32A1 and c26A31 were not attached to the vector. Furthermore, these fragments were unique to c26A31 and c32A1 and mapped to the 5′ and 3′ side of cHDZ23 and cNQ7, respectively (mapping data not shown). They are therefore not as prevalent in the CHOC400 genome as the 220-kb consensus sequence mapping between −15 and +205 kb. This result suggested two possibilities: (i) either the remaining 90% of the amplicons in CHOC400 are heterogeneous in length, so that there are a large number of variably arranged cosmids beyond the consensus region from −15 to +205 kb; or (ii) the predominant amplicon type in CHOC400 cells might be smaller than the type I amplicon, ending near the 5′ and 3′ boundary of cHDZ23 and cNQ7, respectively. We considered the first of these possibilities unlikely because the probes from cHDZ23 and cNQ7 that were used for screening did not detect a heterogeneous family of cosmids extending in the 5′ and 3′ directions. The only cosmids we recovered with these probes either confirmed the previously cloned sequences represented by cHDZ23 and cNQ7 or were derived from the larger type I amplicon.

To test whether cHDZ23 and cNQ7 were connected, forming a smaller amplicon, we asked whether fragments from one cosmid hybridized to digests of the other cosmid (Fig. 3). While none of the fragments from cNQ7 hybridized to any sequences in cHDZ23, a 2.8-kb *Eco*RI fragment from cNQ7 (fragment I, see Fig. 5A) hybridized to a second fragment in an *Eco*RI digest of cNQ7 itself (lane G) and to related cosmids (e.g., c10A19, lane F, Fig. 3A; others not shown). This second fragment was attached to the vector in both cNQ7 and c10A19, accounting for its altered mobility (data not shown). In several other cosmids related to cNQ7 (e.g., c15A16 and c10A39), lanes D and E, Fig. 3A), both copies of fragment I were completely contained within the insert (mapping data not shown), resulting in a 2.8-kb doublet. In these two cosmids, this doublet occurred as part of a triplet, one of which had a different sequence (Fig. 2). However, by comparing the intensity of ethidium bromide staining of the 2.8-kb band in cosmids cNQ7 and c10A19 (lanes F and G) to the triplet in cosmids c15A16 and c10A39 (lanes D and E), an increase in fluorescence of this band was

observed. In all of these clones, one representative of the duplicated sequence either was connected to the vector at the 3′ end or lay very near the 3′ end of the insert (mapping data not shown).

A similar result was obtained when a 6.6-kb fragment from cHDZ23 (fragment V, see Fig. 5A) was hybridized to digests of cNQ7 and to cHDZ23 and related cosmids (Fig. 3B). No sequences in cNQ7 (lane G) were recognized by the cHDZ23 probe. Instead, the probe detected a second fragment in addition to itself in cHDZ23 (lane M) and related cosmids (e.g., c27A37, lane N, Fig. 3B). Again, the fragment was connected to the vector in cHDZ23 and c27A37 and was detected as a doublet in others (e.g., c11A37, c18A14, and c20A42, lanes O to Q, Fig. 3B). This 6.6-kb doublet was also detected in ethidium bromide-stained digests of these cosmids by comparing its intensity with that of larger bands in the same well (lanes O to Q, Fig. 2).

These data showed that the cHDZ23 and cNQ7 families were not connected to each other and further indicated that each cosmid contained either repeated sequence elements or internal duplications. However, both fragment V from cHDZ23 and fragment I from cNQ7 appeared to be single- or low-copy-number sequences when hybridized to parental CHO genomic digests (data not shown). Furthermore, when these probes were used to screen the library in the walking steps outward from cHDZ23 and cNQ7, we did not recover cosmids unrelated to the map in these regions. The probes therefore do not contain either moderately or highly repeated sequence elements that are dispersed throughout the genome. The fact that several other nonsibling cosmids related to cHDZ23 and cNQ7 contained the same duplications (Fig. 2 and 3) also indicates that these sequence arrangements were highly amplified in the CHOC400 genome and were not rearrangements of the cosmids per se. Our results therefore suggest that duplications of the sequences represented by fragments V and I are confined to the amplified DNA represented by the cHDZ23 and cNQ7 families, respectively.

To investigate further the sequence arrangements in cNQ7 and cHDZ23, additional fragments were hybridized to *Eco*RI digests of CHO and CHOC400 genomic DNA and to digests of the cosmids themselves. We found that a 1-kb *Eco*RI fragment from cNQ7 (fragment K/J, see Fig. 5A) hybridized to a highly amplified 1-kb fragment in CHOC400 DNA that was absent from the parental CHO digest, indicating that the fragment was rearranged (Fig. 4A). In addition, fragment K/J recognized a highly amplified 2.5-kb fragment and a less prevalent 0.5-kb fragment in CHOC400 DNA, both of which were present in CHO DNA. The rearranged 1-kb fragment K/J from cNQ7 also hybridized to 2.5-kb and 0.5-kb fragments in cosmid c32A1, but detected only itself and the 2.5-kb fragment in cNQ7 (Fig. 4B). These data indicate that fragment K/J from cNQ7 contains a junction that was formed from fragments K and J, both of which are contained in c32A1 as parental-sized fragments (see map, Fig. 5A).

We found a similar rearrangement in cosmid cHDZ23 when fragment T/U (Fig. 5A) was used to probe genomic digests of CHO and CHOC400 DNA; fragment T/U itself appeared to contain a junction between fragments T and U, both of which are contained as parental-sized fragments in c26A31 (Fig. 5A; hybridization data not shown).

From the *Eco*RI maps of cNQ7 and cHDZ23 (Fig. 5A), it can be seen that the rearranged fragments K/J and T/U are flanked by the duplicated fragments I and V, respectively. We interpret these results to mean that these two cosmids contain head-to-head and tail-to-tail junctions, respectively,

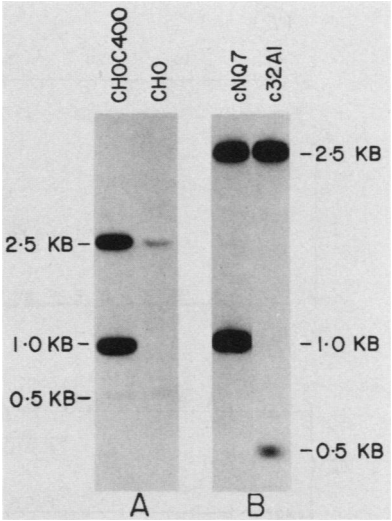

FIG. 4. Cosmid cNQ7 contains a rearranged junction fragment that is highly amplified in the CHOC400 gene. (A) *Eco*RI digests of CHO and CHOC400 genomic DNA were separated on agarose and transferred to GeneScreen. The transfer was probed with a 1.0-kb *Eco*RI fragment (K/J) from cosmid cNQ7 (see map, Fig. 5A). (B) Transfers containing *Eco*RI digests of cosmids cNQ7 and c32A1 were probed with the 1.0-kb *Eco*RI fragment (K/J) from cNQ7. The gel in this panel was run considerably longer than the gel shown in panel A.

between the prevalent type II amplicons. The type II amplicon therefore extends from fragment T/U in the 5′ cosmid cHDZ23 to fragment K/J in the 3′ cosmid cNQ7 (see maps, Fig. 1A and 5A). The type II amplicon is approximately 220 kb long, is represented in its entirety within the type I amplicon, and is missing sequences represented by cIJ44 and part of c26A31.

The proposed sequence arrangements of both the type I and type II amplicons are diagrammed in Fig. 5A, in which a map of the junction region extending from cNQ7 to cHDZ23 is presented (this map is an extension of the right end of the map in Fig. 1A into the next amplicon). It can be seen from the diagram of the type II amplicon that if fragment K/J, which represents the tail-to-tail junction in cNQ7, were hybridized to an *Eco*RI digest of CHOC400 genomic DNA, it would recognize a highly amplified fragment of the same size from the type II amplicon (fragment K/J) in addition to fragment J (also highly amplified) and fragment K (unique to the type I and therefore present less frequently in the genome). A similar result would be obtained with fragment T/U, which represents the proposed head-to-head junction in cHDZ23. In both cases, neither junction fragment would find a counterpart in parental CHO DNA.

In Fig. 5B, we have pictured two of the possible linear arrangements of the two different amplicon types in the CHOC400 genome. In one version, the type I amplicons are grouped together at one chromosomal locus and the type II amplicons are grouped together at another. In the second version, the two types are interspersed.

Representation of the two amplicon types in the CHOC400 genome. We have previously shown that all of the sequences from −10 to +110 kb (Fig. 1A) appear to be amplified to a similar extent in the CHOC400 genome (22). This conclusion derived from experiments in which a fragment from the *dHFR* gene was mixed with a fragment of similar but not

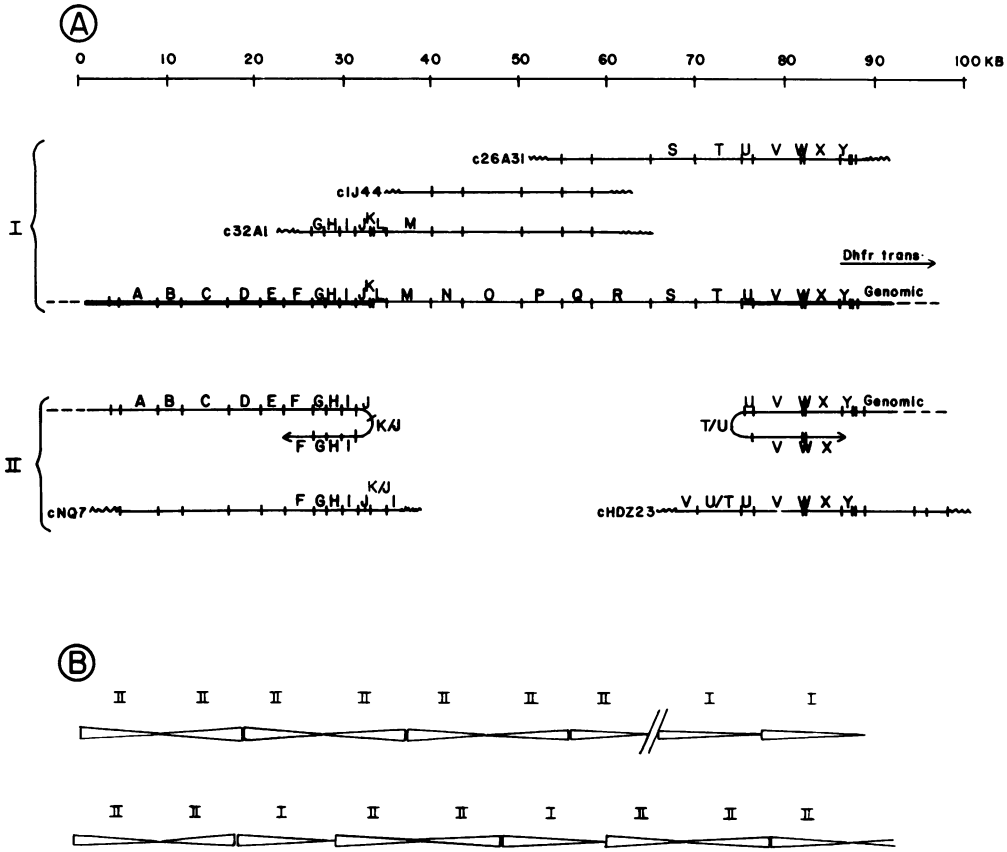

FIG. 5. End-to-end arrangements of the type I and type II amplicons in CHOC400. (A) *Eco*RI map of the region between the 3′ cosmid cNQ7 and the 5′ cosmid cHDZ23 (an extension of the right end of the map in Fig. 1A). The type I map is continuous in this region and was deduced from the sequence arrangement of cosmids c32A1, cIJ44, c26A31, and overlapping clones. The position of transcription initiation from the *dHFR* gene is shown by the arrow above the type I map, and the heavy line in this map indicates sequences shared with the type II amplicon. The type II map shows the inverted sequences that occur at the tail-to-tail and head-to-head junctions between the type II amplicons. This structure was deduced from mapping studies of cosmids cNQ7 and cHDZ23, respectively. (B) Two postulated arrangements of the two amplicon types in the CHOC400 genome are shown. Each arrow symbolizes one amplicon unit, with the point representing the 5′ end or head of the amplicon. In the upper diagram, the type II amplicons are grouped together in a continuous array at one chromosomal location, and the type I amplicons are at another location. In the lower diagram, the two amplicon types are interspersed.

identical size from a distant location; this mixed probe was then hybridized to digests of CHO and CHOC400 genomic DNA. The ratio of the hybridization signals from the two fragments in parental CHO DNA was arbitrarily set to 1, and this ratio was compared with that observed in digests of CHOC400 DNA. With all of the probes extending from −10 to +110 kb, the ratio of signals relative to the gene fragment was similar in parental and in CHOC400 genomic DNA digests (22).

We expanded this analysis to include sequences from the remainder of the amplicon, and representative autoradiograms are presented in Fig. 6. Fragments A to J and U to Y, which are contained in cosmids cNQ7, c32A1, cHDZ23, and c26A31 (Fig. 5A, type I map), were each individually mixed with a second reference fragment of similar but not identical size isolated from cC12, which is in the center of the type II amplicon (Fig. 1A). Each of these mixed probes hybridized to the two fragments in each case with similar relative intensities in parental and in CHOC400 genomic digests (e.g., fragments I and V, Fig. 6A and D). However, fragments L to S, which are unique to the type I amplicon, were greatly underrepresented in the CHOC400 genome relative to the reference fragment from c12 (results for fragments L and S are shown in Fig. 6B and C). The copy number therefore drops off sharply between the two fragments that straddle the junctions at each end of the type II amplicon (from approximately 100% relative to the cC12 probe to approximately 10%). The sequences that are unique to the type I amplicon (represented by cosmid cIJ44 and part of c26A31, Fig. 5A) were therefore present at 10 to 15% of the copy number of the type II amplicons that extend from cHDZ23 to cNQ7 (see Fig. 1A and 5A). This conclusion is also compatible with results obtained in the walking steps outward from cHDZ23 and cNQ7, in which only 10 to 15% of the cosmids recovered extended beyond these clones into the sequences unique to the type I amplicon.

Minor variant sequence arrangements. In each of the chromosome-walking steps involved in the isolation of the 220-kb consensus region shared by the type I and type II amplicons (Fig. 1A), almost every cosmid recovered at each step conformed to the consensus sequence. One exception was a small family of cosmids that were recovered as part of a larger group in screening to the 5′ side of c26-1. A probe from the 5′ end of this cosmid detected about 25 clones of the cN27 and cBP22 types (Fig. 1A), but also hybridized with four clones representing a variant structure (the cYP25 family). One additional clone of the cYP25 type was also isolated in a screening to the 3′ side of cS21. cYP25 and

related cosmids contained a variant 8-kb *Eco*RI fragment that hybridized with an 11-kb fragment in cS21 and a 4.8-kb fragment in c26-1 (hybridization data not shown). This variant was also detected when cSE24 (which overlaps cS21) was hybridized to CHOC400 DNA (Fig. 1B, lane E, arrowhead). The union of sequences in cS21 and c26-1 represents a deletion of approximately 90 kb in about 10% of the amplicons in CHOC400, based on relative hybridization signals on genomic blots. Since this 8-kb variant fragment occurs within the region common to both the type I and type II amplicons, it is not possible at present to determine from which amplicon type it derived.

It could be argued that this variant fragment represents the legitimate junction between the type I amplicons. However, we have shown that it did not appear during the development of the CHOC400 cell line until an MTX resistance level of 50 μg/ml was attained (M. Montoya-Zavala and J. L. Hamlin, unpublished data). Since the larger type I amplicon can be detected at much lower levels of drug resistance (unpublished data), the real junction fragment is probably located elsewhere in the type I map. This 8-kb fragment is one of the few variants in the genome that we have detected in hybridization studies with cosmids, which suggests that our cloning strategy would have recovered variants that occur in >5 to 10% of the amplicons.

Another 20-kb *Eco*RI variant fragment was detected in cCIG23 and related cosmids (e.g., cAA43, arrowhead, lane G, Fig. 1B). This fragment was highly amplified in the CHOC400 genome, but absent as such from parental CHO DNA (data not shown). Instead, the 20-kb fragment hybridized to a 9-kb and an 11-kb fragment in CHO digests, and fragments of this size were also detected as faint bands in *Eco*RI digests of CHOC400 DNA probed with cCIG23 or cAA43 (e.g., lane G, Fig. 1B, bands just below the 2-kb fragment). When the cosmids cCIG23 and cAA43 were used to probe other restriction enzyme digests of CHOC400 and CHO DNA, however, the restriction patterns appeared to be identical (data not shown). We therefore assume that the 20-kb *Eco*RI fragment resulted from the loss of an *Eco*RI site sometime early in the amplification process and is not related to rearrangements involved in amplification per se.

We also occasionally detected minor variant fragments when other cosmids were used to probe CHOC400 DNA (see faint bands in the *Eco*RI digests of CHOC400 DNA probed with cBP7, lane H, Fig. 1B). We have not yet determined whether any of these fragments are also present in CHO digests, which would indicate that they represent moderately repeated, dispersed sequence elements. However, it is likely that some minor bands represent legitimate alterations in sequence arrangement in some of the amplicons in the CHOC400 genome.

DISCUSSION

Several groups have isolated recombinant clones representing amplified DNA from drug-resistant cell lines and human neoplasms (3, 9, 12, 18, 32). To date, however, no amplicon equivalent has been cloned in its entirety from any mammalian cell line. The problem has been made more difficult by the apparent large size of the units of amplification in some cell lines. Furthermore, in two notable cases [the *dHFR* amplicons in an MTX-resistant murine cell line and the CAD amplicons in *N*-(phosphonacetyl)-L-aspartate (PALA)-resistant BHK cells], the amplicons in a single cell line appeared to be quite variable. Not surprisingly, the selectable gene in each case (*dHFR* and CAD) appeared to be identical in structure in the majority of amplicons, but sequences flanking the gene were shown to be highly rearranged (3, 9). This finding indicates that in these cell lines the amplicons are not uniform in length and may even contain sequences that originally derived from other unrelated loci but became joined to the amplicon by complex recombination events (9). Since attempts were made to construct single amplicon maps from clones that derived from different, nonidentical amplicons, the resulting maps exhibited multiple branch points, and cloning an entire amplicon equivalent would have been extremely difficult.

The lack of homogeneity among the amplicons in the MTX-resistant murine cell lines studied by Federspiel et al. (9) could have been because the amplicons in most of these cell lines are located in unstable, extrachromosomal double minutes, which may have undergone frequent illegitimate recombination events that created heterogeneity in amplicon structure. However, the highly rearranged CAD amplicons observed in PALA-resistant Syrian hamster cells are located in stable, abnormally banding chromosomal regions (3), as are the *dHFR* amplicons in CHOC400 cells (21). This incongruity suggests that other more subtle factors, such as cell type or drug treatment regimen, may be contributing to the heterogeneity observed in these systems.

In our chromosome-walking strategy in the CHOC400 genomic library, we were necessarily also constructing a consensus map from recombinant clones that derived from different amplicons in the same cell. Our success in isolating the equivalent of two complete amplicon types is most likely due to the relatively small size of the amplicons in this cell line (220 to 260 kb) and the relative homogeneity of these structures in the CHOC400 genome. In the course of the walking procedure, we did identify a significant (>5%) variant structure (the cYP25 family), but this exceptional arrangement corresponded to a variant fragment that we had detected previously by hybridizing cosmid cS21 to CHOC400 genomic DNA digests (22). We cannot presently rule out the possibility that there are also a large number of random variations scattered throughout the amplicons, any one of which occurs too infrequently to be cloned or detected by hybridization of cosmids to genomic digests. The type I and type II maps in Fig. 1A and 5A should therefore be considered maps of the average amplicons in the CHOC400 genome.

The map of the type I amplicon is circularly permuted, indicating that these amplicons are arranged in head-to-tail arrays, although we have not as yet identified the junction fragments between the type I amplicons that will allow us to formally demonstrate this proposal. However, we are using a series of CHO mutants that have deleted large segments of the *dHFR* locus (29) in a strategy that may allow us to determine the location of the junctions unambiguously.

Especially surprising is the apparently homogeneous arrangement of the interamplicon junctions between the type II amplicons. We isolated several cosmids representing both the 5′ end (cHDZ23 family) and the 3′ end (cNQ7 family) of this amplicon, and all of the cosmids within each group contained the same rearranged fragments, which undoubtedly represent the joints between the type II amplicons. Thus, once the amplification and recombination events occurred that gave rise to these rearrangements of parental fragments, the rearrangements seem to have been propagated in subsequent amplification steps.

The highly amplified, rearranged fragments detected in cHDZ23 and cNQ7 are flanked on both sides by duplicated fragments. This finding suggests that the cHDZ23 and cNQ7

cosmid families, which define the ends of the type II amplicon, contain the inverted repeat sequences that would be expected of head-to-head and tail-to-tail unions between amplicons. Our results are thus compatible with the observations of Ford and Fried (11), who used a snap-back hybridization procedure to demonstrate that extensive inverted repeat sequences occur in the amplified DNA of a PALA-resistant BHK cell line and in the DNA of four human tumor cell lines containing amplified c-*myc* genes. It is also interesting that in all of the cosmids derived from the suspected joint regions between the type II amplicons in CHOC400, the center of the inverted sequence (joint) is located close to the end of the cosmid insert. None were found in which the rearranged fragment is located in the middle of the insert. This result would be expected if larger inverted repeat sequences were not tolerated by the bacterial host, were recombined out, and were therefore not included in our analysis.

The question arises why there are two different amplicon types in the CHOC400 genome. The *dHFR* amplicons in this cell line are located in abnormally banding chromosome regions on three different chromosomes (one major and two minor sites [21]). One explanation is that amplification of the type I amplicon may have initiated at only one of these sites, but a fragmentation event that gave rise to secondary sites may have involved a truncation of the type I amplicon at the 5′ end, giving rise to the type II amplicon. In situ hybridization studies with probes from the region unique to the type I repeated sequence should help us to test this hypothesis, since these probes should hybridize to only one or two of the abnormally banding sites in the CHOC400 karyotype.

The type I and type II amplicons are considerably longer than our original estimate of 135 kb, which was based on the sum of the lengths of observable restriction fragments separated on ethidium bromide-stained agarose gels (21). This discrepancy is most likely due to the fact that doublets and triplets are often not detected in this method, leading to an underestimate of the total amplicon length. However, the minimum 220-kb and 260-kb lengths that we have calculated from the map of the clones are reasonably compatible with estimates based on the size of the abnormally banding regions on chromosomes and the *dHFR* gene copy number. The abnormally banding regions in CHOC400 represent ~5% of the total condensed mitotic chromosome length (21). With a genome size of 6×10^9 base pairs (bp), this represents ~3×10^8 bp. We estimate that there are 1,000 to 1,200 copies of the *dHFR* amplicon per cell, which would give an average amplicon size of 250 to 270 kb. It should be noted, however, that these size estimates do not take into account the 90-kb deletion that occurred between cS21 and c26-1 in about 10% of the amplicons (represented by cYP25, see Fig. 1A). Since we do not know in which amplicon type this deletion occurred, it is possible that the type I amplicon is only 170 kb long or that 10% of the type II amplicons are only 130 kb long. Alternatively, the 90-kb deletion may be distributed between both amplicon types.

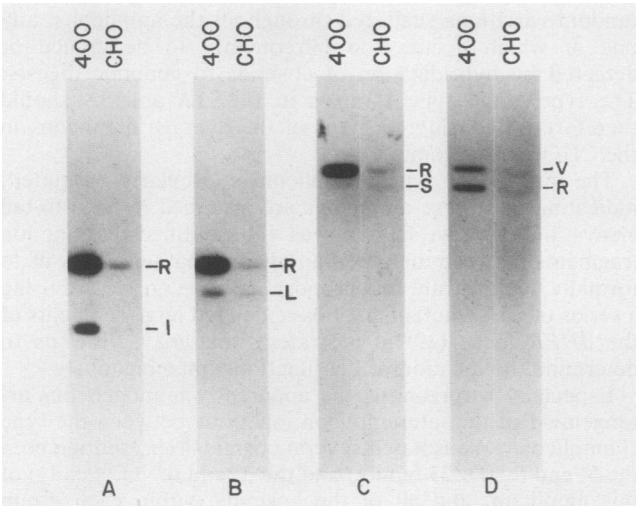

FIG. 6. Representation of the two amplicon types in the CHOC400 genome. *Eco*RI digests of CHO and CHOC400 genomic DNA were separated on agarose gels and transferred to GeneScreen. The digests were hybridized to a mixed probe containing a reference *Eco*RI fragment (R) from the cosmid cC12 (see Fig. 1A) and a second fragment from the junction region shown in Fig. 5A. (A) A 2.6-kb reference fragment (R) and the 1.3-kb fragment I. (B) A 2.6-kb reference fragment (R) and the 1.8-kb fragment L. (C) A 6.0-kb reference fragment (R) and the 5.3-kb fragment S. (D) A 6.0-kb reference fragment (R) and the 6.6-kb fragment V.

Our data do not presently allow us to discriminate between the re-replication and unequal sister chromatid exchange models for DNA sequence amplification in this system. The initial unit of amplification might be equivalent to a parental replicon, as we have suggested previously (22). The cosmid cS21, which contains the initiation locus, would therefore be expected to lie close to the center of the type I amplicon, which is the structure detected at early stages of amplification (unpublished data). This would fix the junction between the type I amplicons in the region of cNQ7, near the tail-to-tail junctions in the type II amplicons (Fig. 5A). We will not be able to test this hypothesis until we have determined the position of the junction(s) between the type I amplicon. However, in a forthcoming report, we will show that three other independently derived methotrexate-resistant Chinese hamster cell lines have amplified the entire 260-kb type I amplicon (J. E. Looney and J. L. Hamlin, manuscript in preparation). This finding is interesting in light of the fact that the *dHFR* gene itself is presumably the only sequence whose amplification is selected by the drug treatment regimen. We suggest that the cell may be mechanistically constrained, at least initially, to amplify an entire replicon, whose average boundaries are fixed in the genome, possibly at preferred topoisomerase II binding sites or other sites that are involved in maintaining chromosome structure or function. However, any proposed mechanism must somehow explain the head-to-tail, head-to-head, and tail-to-tail arrangements that we have observed in this system.

Since we have now cloned the equivalent of two entire amplicons from the CHOC400 genome, we hope to be able to determine whether either of these units has the same boundaries as the parental replicon in the CHO genome. The isolation of a complete replicon will allow us to study the structural organization of this functional domain in the nucleus.

ACKNOWLEDGMENTS

We thank the members of our laboratory and Bill Pearson for many helpful discussions. We also extend our appreciation to Valerie Frazier, Kevin Cox, and Carlton White for their invaluable technical assistance.

This work was supported by Public Health Service grants GM-26108 and GM-32481 (to J.L.H.) and by National Research Service Award GM-09742 (to J.E.L.) from the National Institutes of Health and by an American Cancer Society Faculty Research Award to J.L.H.

LITERATURE CITED

1. **Alitalo, K., M. Schwab, C. C. Lin, H. E. Varmus, and J. M. Bishop.** 1983. Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-*myc*) in malignant neuroendocrine cells from a human colon carcinoma. Proc. Natl. Acad. Sci. USA **80**:1707–1711.
2. **Alt, F. W., R. E. Kellems, J. R. Bertino, and R. T. Schimke.** 1978. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. J. Biol. Chem. **253**:1357–1370.
3. **Ardeshir, F., E. Giulotto, J. Zieg, O. Brison, W. S. L. Liav, and G. R. Stark.** 1983. Structure of amplified DNA in different Syrian hamster cell lines resistant to *N*-(phosphonacetyl)-L-aspartate. Mol. Cell. Biol. **3**:2076–2088.
4. **Biedler, J. L., and B. A. Spengler.** 1976. A novel chromosome abnormality in human neuroblastoma and antifolate-resistant Chinese hamster cells in culture. J. Natl. Cancer Inst. **57**:683–695.
5. **Bostock, C. J., and E. M. Clark.** 1980. Satellite DNA in large marker chromosomes of methotrexate-resistant mouse cells. Cell **19**:709–715.
6. **Collins, S., and M. Groudine.** 1982. Amplification of endogenous *myc*-related DNA sequences in a human myeloid leukaemia cell line. Nature (London) **298**:679–681.
7. **Cowell, J. K.** 1982. Double minutes and homogeneously staining regions: gene amplification in mammalian cells. Annu. Rev. Genet. **16**:21–59.
8. **Dalla Favera, R., F. Wong-Staal, and R. C. Gallo.** 1982. *onc* gene amplification in promyelocytic leukaemia cell line HL-60 and primary leukaemic cells of the same patient. Nature (London) **299**:61–63.
9. **Federspiel, N. A., S. M. Beverley, J. W. Schilling, and R. T. Schimke.** 1984. Novel DNA rearrangements are associated with dihydrofolate reductase gene amplification. J. Biol. Chem. **259**:9127–9140.
10. **Feinberg, A. P., and B. Vogelstein.** 1983. High specific activity labeling of DNA restriction endonuclease fragments. Anal. Biochem. **132**:6–13.
11. **Ford, M., and M. Fried.** 1986. Large inverted duplications are associated with gene amplification. Cell **45**:425–430.
12. **Gros, P., J. Croop, I. Roninson, A. Varshavsky, and D. E. Housman.** 1986. Isolation and characterization of DNA sequences amplified in multidrug-resistant hamster cells. Proc. Natl. Acad. Sci. USA **83**:337–341.
13. **Gross-Bellard, M., P. Oudet, and P. Chambon.** 1978. Isolation of high molecular weight DNA from mammalian cells. Eur. J. Biochem. **36**:32–38.
14. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
15. **Hamlin, J. L., J. D. Milbrandt, N. H. Heintz, and J. C. Azizkhan.** 1984. DNA sequence amplification in mammalian cells. Int. Rev. Cytol. **90**:31–82.
16. **Heintz, N. H., J. D. Milbrandt, K. S. Greisen, and J. L. Hamlin.** 1983. Cloning the initiation region of a mammalian chromosomal replicon. Nature (London) **302**:439–441.
17. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9**:2989–2998.
18. **Kanda, N., R. Schreck, F. Alt, G. Bruns, D. Baltimore, and S. Latt.** 1983. Isolation of amplified DNA sequences from IMR-32 human neuroblastoma cells: facilitation by fluorescence-activated flow sorting of metaphase chromosomes. Proc. Natl. Acad. Sci. USA **80**:4069–4073.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 133–134. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Milbrandt, J. D., J. C. Azizkhan, K. S. Greisen, and J. L. Hamlin.** 1983. Organization of a Chinese hamster ovary dihydrofolate reductase gene identified by phenotypic rescue. Mol. Cell. Biol. **3**:1266–1273.
21. **Milbrandt, J. D., N. H. Heintz, W. C. White, S. M. Rothman, and J. L. Hamlin.** 1981. Methotrexate-resistant Chinese hamster ovary cells have amplified a 135 kilobase pair region that includes the dihydrofolate reductase gene. Proc. Natl. Acad. Sci. USA **78**:6043–6047.
22. **Montoya-Zavala, M., and J. L. Hamlin.** 1985. Similar 150-kilobase DNA sequences are amplified in independently derived methotrexate-resistant Chinese hamster cells. Mol. Cell. Biol. **5**:619–627.
23. **Nunberg, J. H., R. J. Kaufman, R. T. Schimke, G. Urlaub, and L. A. Chasin.** 1978. Amplified dihydrofolate reductase genes are localized to a homogenously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. Proc. Natl. Acad. Sci. USA **75**:5553–5556.
24. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
25. **Schimke, R. T. (ed.).** 1982. Gene amplification, p. 317–333. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Schimke, R. T.** 1984. Gene amplification in cultured animal cells. Cell **37**:705–713.
27. **Schwab, M., K. Alitalo, H. E. Varmus, J. M. Bishop, and D. George.** 1983. A cellular oncogene (c-Ki-*ras*) is amplified, overexpressed, and located within karyotypic abnormalities in mouse adrenocortical tumour cells. Nature (London) **303**:497–501.
28. **Stark, G. R., and G. M. Wahl.** 1984. Gene amplification. Annu. Rev. Biochem. **53**:447–491.
29. **Urlaub, G., and L. A. Chasin.** 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc. Natl. Acad. Sci. USA **77**:4216–4220.
30. **Wahl, G. M., L. Vitto, R. A. Padgett, and G. R. Stark.** 1982. Single copy and amplified CAD genes in Syrian hamster chromosomes localized by a highly sensitive method for in situ hybridization. Mol. Cell. Biol. **2**:308–319.
31. **Yeung, C. Y., L. G. Frayne, M. R. Al-Ubaidi, A. G. Hook, D. A. Wright, and R. E. Kellems.** 1983. Amplification and molecular cloning of murine adenosine deaminase gene sequences. J. Biol. Chem. **258**:15179–15185.
32. **Zieg, J., C. E. Clayton, F. Ardeshir, E. Giulotto, E. A. Swyryd, and G. R. Stark.** 1983. Properties of single-step mutants of Syrian hamster cell lines resistant to *N*-(phosphonacetyl)-L-aspartate. Mol. Cell. Biol. **3**:2089–2098.