MOLECULAR AND CELLULAR BIOLOGY, July 1983, p. 1266–1273
0270-7306/83/071266-08$02.00/0

Vol. 3, No. 7

# Organization of a Chinese Hamster Ovary Dihydrofolate Reductase Gene Identified by Phenotypic Rescue

J. D. MILBRANDT, J. C. AZIZKHAN, K. S. GREISEN, AND J. L. HAMLIN*
*Department of Biochemistry, University of Virginia, Charlottesville, Virginia 22908*

Received 10 January 1983/Accepted 21 April 1983

We have constructed a genomic DNA library from a methotrexate-resistant Chinese hamster ovary cell line (CHOC 400) in the cosmid vector pHC79. By utilizing a murine dihydrofolate reductase (DHFR) cDNA clone, we have identified 66 DHFR$^+$ clones among the 11,000 colonies screened by colony hybridization. To isolate a recombinant cosmid containing the entire DHFR gene, we have tested these colonies for their ability to rescue a DHFR$^-$ Chinese hamster ovary cell line, using the spheroplast fusion method of gene transfer developed by W. Schaffner (Proc. Natl. Acad. Sci. U.S.A. 77:2163–2167, 1980). One clone (cH1) was able to transform DHFR$^-$ cells to the DHFR$^+$ phenotype and was shown in hybridization studies to contain all of the gene except a small portion of the 3′ untranslated region. We have mapped cosmid cH1 and several overlapping cosmids with a variety of restriction enzymes and have determined the approximate positions of the five (and possibly six) exons within the DHFR gene. Differences between the sizes of homologous genes in hamster cells (24.5 kilobases [kb]) and in mouse cells (31.5 kb) are shown to reside primarily in the length of the 3′ intron, which is 8 kb in the hamster gene and 16 kb in length in the mouse gene. Our studies confirm the utility of cosmid libraries for the isolation of large genes, as previously shown by R. de Saint Vincent et al. (Cell 27:267–277, 1981). In addition, a cosmid that contains a functional DHFR gene will be a useful vector for the co-amplification and subsequent overexpression of other cloned genes.

When cultured mammalian cells are exposed to stepwise increases in methotrexate (MTX) over the period of many months, drug-resistant cell lines can eventually be isolated that markedly overproduce dihydrofolate reductase (DHFR), the target enzyme for this folate antagonist (1–4, 23, 25). In all cases studied thus far, enzyme overproduction is correlated with amplification of the DHFR gene, along with considerable amounts of flanking DNA. Cells can contain more than 1,000 copies of the amplified sequence (amplicon) per cell (4, 23), and these copies are arranged in tandem, either in homogeneously staining chromosomal regions (3) or in extrachromosomal, acentromeric double minutes (17).

The tremendous enrichment for the DHFR gene and its mRNA in MTX-resistant cells has facilitated the isolation of both the cDNA and the gene itself. Schimke and colleagues have cloned and characterized several full-length DHFR cDNAs (5) and the DHFR gene (7) from an MTX-resistant mouse line (S180). The cDNA clones vary greatly in length, with the differences occurring primarily in the 3′ untranslated region. The mouse DHFR gene is 31.5 kilobases (kb) in length and contains five introns (7).

In our laboratory, we are studying the replication of DHFR amplicons in MTX-resistant Chinese hamster ovary (CHO) cells. One of our cell lines (CHOC 400) contains approximately 1,000 copies of a 135-kb sequence that includes the DHFR gene (23). We have prepared a genomic DNA library from CHOC 400 cells in the cosmid vector pHC79 (15). As part of a project to isolate and characterize the amplicon, we utilized a murine DHFR cDNA to identify cosmids containing DHFR genomic sequences. To facilitate mapping studies, and to generate a selectable, amplifiable marker, we attempted to isolate a clone containing the entire gene. We therefore tested the clones that hybridized to murine DHFR cDNA for their ability to transform a DHFR$^-$ CHO cell line (31) to the DHFR$^+$ phenotype when introduced into these cells by spheroplast fusion (28).

In this paper we describe the preparation and screening of this CHOC 400 genomic library and the identification by phenotypic rescue of a recombinant cosmid containing a functional

DHFR gene. We have used this cosmid and several overlapping clones to construct a restriction map of the Chinese hamster DHFR gene.

## MATERIALS AND METHODS

**Tissue culture.** Monolayer cultures of CHOC 400 were maintained in minimal essential medium (GIBCO Laboratories, Grand Island, N.Y.) supplemented with 10% donor calf serum (GIBCO) and 800 μM MTX (National Cancer Institute). The DHFR$^-$ CHO cell line DXB11 (31) requires glycine, thymidine, and hypoxanthine for growth and was maintained in F12 medium (GIBCO) containing 10% donor calf serum. After spheroplast fusion, DHFR$^+$ transformants were selected by growth in medium lacking these nutrients (minimal essential medium).

**Preparation of high-molecular-weight DNA.** High-molecular-weight DNA was prepared by a modification of the procedure of Gross-Bellard et al. (13). Cells were grown in 15-cm tissue culture dishes and were washed with phosphate-buffered saline before lysis by the addition of 10 ml of DNA lysis buffer (100 mM Tris, pH 8, 100 mM NaCl; 100 mM EDTA; 0.5% sodium dodecyl sulfate; 150 μg of Proteinase K [Scientific Products] per ml). The lysate was incubated for 12 h at 37°C in this buffer and was extracted directly on the plate with an equal volume of water-saturated phenol for 12 h. The viscous lysate was removed to a screw-cap tube and was phenol extracted twice more, followed by exhaustive dialysis against a buffer containing 10 mM Tris (pH 7.5), 1 mM EDTA, and 10 mM NaCl. The modal size of this DNA was greater than 100 kb. This genomic DNA was digested to a modal size of 30 kb with 0.01 U of *Sau*3A (New England Biolabs) per μg of DNA for 1 h at 37°C. The resulting fragments were treated with 1 U of bacterial alkaline phosphatase (Bethesda Research Laboratories) per μg of DNA for 1 h at 65°C, followed by phenol extraction and ethanol precipitation. The ligatability of these fragments before and after phosphatase treatment was checked on small samples of each preparation.

**Construction and screening of the genomic library.** The cosmid vector pHC79 was generously provided by J. Burke (Imperial Cancer Research Fund, London), and T4 ligase was a gift from R. Kolodner (Sidney Farber Cancer Institute, Boston, Mass.). The cosmid was digested to completion with *Bam*HI (Bethesda Research Laboratories) and was ligated to the phosphatased genomic DNA at a weight ratio of 3:1 for 16 h at 14°C. The ligation mixture contained 7.5 μg of DNA, 40 mM Tris (pH 7.8), 5 mM 2-mercaptoethanol, 8 mM $MgCl_2$, 0.67 mM ATP, and 0.1 U of ligase in a final volume of 15 μl. The ligation mixture was packaged in vitro by the method of Enquist and Sternberg (11). The extract yielded $4 \times 10^7$ plaques per μg of WI30 lambda DNA on *Escherichia coli* HB101. The packaged recombinant DNA was adsorbed to HB101 cells, and the cells were spread onto L plates containing ampicillin (40 μg/ml). Colonies were grown to less than 1 mm in diameter, replica plated, and screened by the method of Grunstein and Hogness (14), except that Whatman no. 541 disks were substituted for nitrocellulose.

Cosmid DNA from positive colonies was prepared according to the method of Ish-Horowicz and Burke (16), digested with *Bam*HI, and examined on agarose gels to determine the size of the inserted DNA.

**Restriction digests, gel electrophoresis, and Southern blotting.** Enzyme digests were performed under the conditions recommended by the supplier (Bethesda Research Laboratories). Double digests were performed by adding both enzymes simultaneously if the salt optima were the same. If the optima were different, the enzyme with the lower salt requirement was added first, and after an adequate incubation the salt concentration was raised and the second enzyme was added. These digests were then electrophoresed on 0.7 and 2% agarose gels to obtain accurate molecular weights of the large and small fragments, respectively. Gels were transferred to Gene Screen (New England Nuclear Corp.) or nitrocellulose (Schleicher and Schuell) by the method of Southern (30), with minor modifications. DNA probes were labeled with [$^{32}$P]-dCTP to a specific activity of $10^8$ dpm/μg by nick translation (27). The dried transfers were hybridized with the indicated DNA probes in the buffer described by Wahl et al. (32); the hybridization mixture contained 1 μg of unlabeled, sonicated CHO DNA per ml to compete with highly repeated sequence elements in the probe. The filters were washed, and autoradiography was performed as previously described (23).

**Spheroplast fusion.** A modification of the method of Schaffner (28) was used to introduce recombinant cosmids into animal cells. CHO DHFR$^-$ cells were plated 16 to 20 h before fusion at a density of $5 \times 10^5$ cells per 60-mm dish. On the same day, 40-ml cultures of strain HB101 harboring recombinant cosmid were grown to an absorbance of 0.6 at 600 nm in 2XYT broth (24) containing 10 μg of ampicillin per ml. Chloramphenicol was added to a final concentration of 10 μg/ml, and the culture was incubated overnight.

Bacterial spheroplasts were prepared by the method of Osborn et al. (26), with minor modifications. Bacteria were harvested by centrifugation at 3,000 × *g* for 10 min, and the pellet was suspended in 20 ml of 30 mM glycyl-glycine (pH 8.0) containing 30% sucrose. The bacterial suspension was transferred to a 125-ml flask, 100 μl of lysozyme (10 mg/ml of water) and 32 μl of 0.25 M EDTA were added, and the flasks were swirled on a rotating platform at 60 rpm for 25 min at 25°c. The spheroplasts were diluted with an equal volume of distilled water and centrifuged as described above, and the pellet was suspended in 6 ml of medium F12 containing serum. Spheroplast formation was monitored by phase-contrast microscopy.

Before fusion, the medium was removed from the cell culture dishes and 2 ml of spheroplast suspension was layered onto each plate. The dishes were centrifuged in a Beckman J6 equipped with a JS4.2 rotor at 2,000 rpm for 20 min. The medium was carefully removed, and 0.75 ml of 40% (vol/vol) polyethylene glycol (Koch-Light; 1,000 molecular weight) was gently pipetted onto the dish. Polyethylene glycol was prepared in calcium- and serum-free minimal essential medium. After 1 min, the polyethylene glycol was carefully diluted with 5 ml of F12 medium containing 10% serum, and the plates were allowed to stand undisturbed for an additional 15 min at 25°C. The plates were washed three times with the same medium and were incubated overnight in nonselective medium (F12). The selective medium (minimal essential medium) was applied 16 h later. Colonies were detected microscopically after 2 weeks.

## RESULTS

**Construction of the CHOC 400 genomic library.** To isolate overlapping recombinant clones spanning the entire 135-kb DHFR amplicon, we elected to clone CHOC 400 genomic DNA in the cosmid vector pHC79. After packaging in vitro, the inserts in cosmids are usually large (35 to 50 kb). Therefore, the number of clones required to map this domain would be small, and the entire DHFR gene might be isolated in a single recombinant cosmid. We treated the *Sau*3A-cleaved DNA with bacterial alkaline phosphatase to prevent the joining of genomic fragments to one another, and then ligated the genomic DNA to an excess of pHC79 digested with *Bam*HI. From 7.5 μg of genomic DNA we obtained a total of 100,000 clones, 70% of which were recombinant. Thus, the resulting library is more than adequate to isolate several equivalents of the DHFR amplicon, which is present approximately 1,000 times per cell, but is not sufficient to contain all single copy sequences (6). The average genomic insert was found to be 25 kb. Some cosmids had inserts as large as 45 kb, and some inserts were much smaller due to the presence of multiple copies of the cosmid.

**Identification of a cosmid containing a functional DHFR gene.** To detect clones containing DHFR sequences, we screened 11,000 colonies with a murine DHFR cDNA (pDHFR21; reference 5). Sixty-six colonies were detected by this procedure, and these were further tested in pools for their ability to transform DHFR$^-$ CHO cells to the DHFR$^+$ phenotype by the spheroplast delivery system. Cultures of each clone were pooled into 11 groups, each containing six clones. Spheroplasts were prepared from cultures of each of the 11 groups and were fused to DHFR$^-$ CHO cells (see above). After 2 weeks, colonies were observed only on the dishes that had been fused to pool 9. Several colonies were observed on each of the dishes fused with this pool; however, only one colony survived in long-term culture. This colony was designated JSH-0 and was used for subsequent characterization of the transformation event. The transient appearance of multiple colonies during the first few weeks of selection may represent the retention of the cosmid in an extrachromosomal state, or the eventual excision of an integrated copy of the cosmid from the host genome. No colonies were observed at any time in plates fused with any of the other 10 pools.

Each of the six individual bacterial clones present in pool 9 was then fused separately to CHO DHFR$^-$ cells. When used individually, cell colonies were present only in the dishes that were fused to clone cH1, thus identifying a cosmid that contains a functional DHFR gene. In this experiment, one colony arose per $2 \times 10^5$ cells plated (12 colonies total), although only 25% of these colonies survived in long-term culture.

We also analyzed each clone in pool 9 for the presence of DHFR sequences. Recombinant cosmid DNA was prepared from each of the six clones in this pool, and the DNAs were digested with *Eco*RI and electrophoresed on agarose gels. The digests were transferred to nitrocellulose and were hybridized with the radioactive murine DHFR cDNA (Fig. 1A). Of the six clones, only two (cH1 and cH2; Fig.1A, lanes e and f) contained DNA sequences that hybridized to the DHFR cDNA probe, and only cH1 contained all of the exons detected in digests of CHOC 400 genomic DNA (lane a). However, one exon did not appear to be present in its entirety in cH1. The 4.4-kb *Eco*RI fragment detected in the genomic CHOC 400 digest

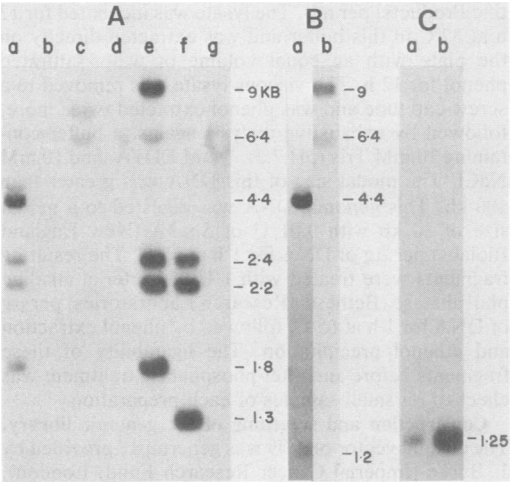


FIG. 1. Hybridization analysis of cosmid clones comprising pool 9. (A) CHOC 400 genomic DNA (lane a) and each cosmid in pool 9 (lanes b through g) were digested with *Eco*RI, electrophoresed on agarose gels, and transferred to nitrocellulose (30). The digests were then hybridized with $^{32}$P-labeled murine DHFR cDNA (pDHFR21; Carothers et al., in press). Cosmids cH1 and cH2 are in lanes e and f, respectively. (B, C) The cosmid cH1 (lanes b) and CHOC 400 DNA (lanes a) were further analyzed by hybridization (B) with a mixture of the two Chinese hamster 3′ DHFR cDNAs (pDCH14 and pDCH18; Carothers et al., in press) or (C) with the murine 5′ genomic clone pDSa7 (7). The 6.4-kb fragments in (A) and (B) are pHC79 monomer cross-reacting with contaminating pBR322 present in the $^{32}$P-labeled DHFR cDNA inserts. Note that cosmid cH1 contains three vector molecules arranged in tandem. Note also that a minor band at approximately 4.8 kb is detected in CHOC 400 genomic digests and may represent a variant generated during the amplification process.

(known to contain the 3′ exon; Milbrandt, unpublished data) is shifted to a new position in cH1 (9 kb). This new fragment also hybridized to pHC79 (not shown), indicating that it contains a vector-genomic junction. To further investigate the 3′ exon in this cosmid, we probed an identical blot with a mixture of two Chinese hamster DHFR cDNA clones, pDCH18 and pDCH14 (A. M. Carothers, G. Urlaub, N. Ellis, and L. A. Chasin, Nucleic Acids Res., in press). These cDNAs overlap one another and together extend 1,900 nucleotides upstream from the polyadenylation site at the 3′ terminus of pDC18. These cDNAs contain an internal *Eco*RI site in the overlapping regions (Carothers et al., in press). In Fig. 1B, it can be seen that this mixed probe hybridizes to two *Eco*RI fragments at 4.4 and 1.2 kb in the CHOC 400 genomic digest, but detects only the 9-kb fragment in cH1. Therefore, the DHFR 3′ exon in this cosmid terminates on the 5′ side of the *Eco*RI site that is located 600 nucleotides from the polyadenylated terminus of pDC18. To ensure that the 5′ end of the gene was present in cosmid H1, we also probed these digests with a murine genomic clone (pDSa7) known to contain the 5′ regulatory and initiation signals of the DHFR gene (7). A 1.25-kb *Eco*RI fragment was detected in both the CHOC 400 and cH1 digests, indicating that this portion of the gene is intact in the cosmid (Fig. 1C).

It is important to note that DHFR mRNAs in both hamster and mouse cells contain a large 3′ untranslated sequence and multiple polyadenylation sites (5; Carothers et al., in press). This could explain why cH1 codes for a functional DHFR enzyme while lacking a portion of the 3′ exon. Alternatively, cH1 may have integrated into either of the partially deleted DHFR genes in the DHFR$^-$ CHO host genome (31), thus restoring a functional gene by site-specific recombination. To test this latter possibility, cH1 was used to probe *Eco*RI genomic digests of CHOC 400, JSH-0 (the rescued cell line), DXB11 (the DHFR$^-$ CHO line), and the cosmid cH1 itself (Fig. 2). When the hybridization patterns of cH1 and CHOC 400 DNA were compared, three additional bands were detected in cH1 that were absent in CHOC 400. All three fragments hybridized to $^{32}$P-labeled pHC79 (not shown) and represent cosmid monomer (6.4 kb) and the two vector-insert junction fragments (1.4 and 9 kb). When JSH-0 DNA was compared with cH1, it was evident that all the bands present in cH1 are conserved in JSH-0 except the 1.4-kb junction fragment. This fragment must therefore represent the site of insertion of cH1 into the genome of the CHO DHFR$^-$ host. Two new fragments were detected in JSH-0 at 1.9 and 4 kb that represent new cosmid-host junctions, and both hybridized to $^{32}$P-labeled vector (not shown). The comparable hybridization signals arising from DHFR sequences in CHO (with two genes) and JSH-0 suggest that between three and four copies of the recombinant cosmid cH1 have been integrated into the genome of DXB11, giving rise to the DHFR$^+$ derivative JSH-0 (Fig. 2). From this analysis, it is apparent that the cosmid has integrated into the host genome by interrupting a sequence in the vector portion of the molecule, thus allowing the genomic insert containing the DHFR gene to be integrated intact. This result argues against a site-specific recombination mechanism for restoration of DHFR activity. Furthermore, upon amplification with MTX, this pattern was maintained, suggesting that this configuration is responsible for producing a functional gene product (22).

**Restriction map of the DHFR gene.** The restriction sites in cH1 were mapped by subjecting the cosmid DNA to a combination of single- and double-enzyme digests, followed by fragment size determination on agarose gels. Restriction

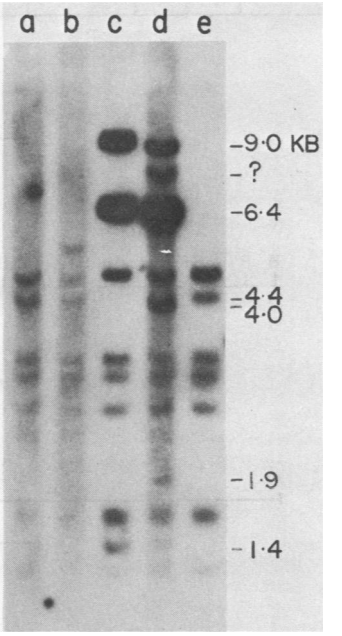


FIG. 2. Integration pattern of cH1 into the JSH-0 genome. Genomic DNA preparations from CHO (lane a), DXB11 (b), JSH-0 (d), and CHOC 400 (e) cell lines, as well as DNA from cH1 (lane c), were cleaved with *Eco*RI. Digests were electrophoresed on agarose gels, transferred to nitrocellulose, and probed with $^{32}$P-labeled cH1 DNA. The lengths of the fragments discussed in the text are indicated. Note that cH1 contains at least three vector molecules. The band in lane d indicated with a question mark is a variant fragment that we have not as yet characterized and is more prominent in this experiment than usually observed (e.g., see Fig. 1A, lane B of reference 22).

enzymes that have GC-rich recognition sequences (*Xho*I, *Sma*I, *Bst*EII, *Kpn*I, and *Bam*HI), and consequently have few recognition sites in the animal cell genome, were used for initial mapping studies. A 5′ mouse genomic clone (pDSa7; reference 7) initially enabled us to orient the map of this gene. Overlapping cosmids containing either the 5′ or 3′ regions of the gene were subsequently identified and mapped to produce the extended map of the DHFR genomic locus shown in Fig. 3.

To localize the coding regions of this large gene, we subcloned four *Hin*dIII fragments that hybridize to the murine cDNA probe (Fig. 3). *Hin*dIII fragments E, L, and U (7.3 , 3.2, and 1.1 kb) and a *Hin*dIII/*Eco*RI fragment (2.0 kb) derived from *Hin*dIII fragment A were further mapped with additional restriction enzymes, and digests were probed with the mouse cDNA. Preliminary placement of the coding regions is shown in Fig. 3. Note that the black boxes within the gene denote the smallest DHFR cDNA binding fragments analyzed thus far, and are larger than the actual exons of the gene. The exact locations and sizes of the exons will require S1 mapping and sequence analysis. From the positions within the map of the 5′ and 3′ cDNA binding fragments, we have estimated that the size of the DHFR gene is 24.5 kb in Chinese hamster cells. We define the 5′ end of the gene as the *Sma*I site located in the first exon-containing fragment (Fig. 3), since this enzyme dissociates nontranscribed sequences from those that are transcribed into mRNA (J. C. Azizkhan, unpublished data). The 3′ end is tentatively positioned approximately 1.5 kb to the right of the last *Bam*HI site in the gene (Fig. 3) and represents the right end of the 3′-most CHO DHFR cDNA used in our studies (pDCH18; Carothers et al., in press).

**Comparison of the parental CHO gene with the CHOC 400 amplified gene.** Since the restriction

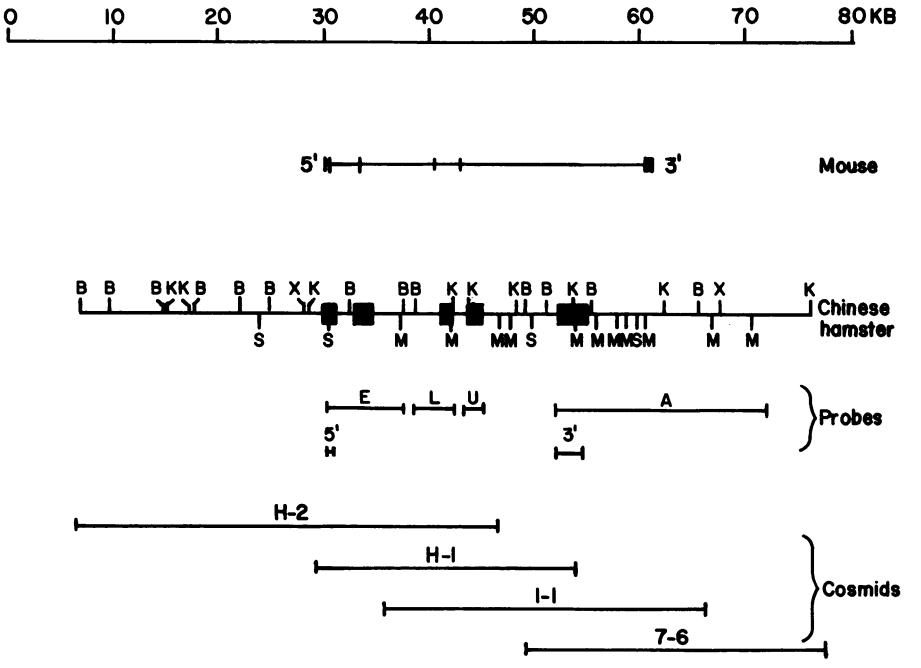


FIG. 3. Restriction map of the Chinese hamster DHFR gene locus. The four cosmids indicated were analyzed by single- and double-enzyme digests, as described in the text. Fragments less than 0.5 kb may not have been detected in this analysis. To determine the location of the exons in the DHFR gene, digests were probed with the following DNA sequences: (i) murine DHFR cDNA (pDHFR21); (ii) the murine 5′ genomic clone (pDSa7); (iii) three CHO *Hin*dIII genomic clones that cross-react with murine cDNA (E, L, and U in figure); and (iv) a 2-kb CHO *Hin*dIII/*Eco*RI genomic subclone of *Hin*dIII fragment A that hybridizes to murine DHFR cDNA and to the 3′ CHO cDNAs (labeled 3′ in figure). To locate exons more precisely, the *Hin*dIII fragments E and L were mapped with additional restriction enzymes, and digests were probed with murine DHFR cDNA. The black boxes shown within the gene denote the smallest DHFR cDNA-hybridizing fragments so far analyzed, and necessarily are larger than the actual exons. The locations of the exons in the mouse DHFR gene are shown for comparison. In this case, the size and location of the exons have been determined from sequencing data (7). Symbols used: B, *Bst*EII; K, *Kpn*I; M, *Bam*HI; S, *Sma*I; X, *Xho*I.

map of the DHFR gene was constructed from clones derived from the amplified gene in CHOC 400, we wanted to confirm that the gene had not been altered during the amplification process. We therefore compared restriction digests of genomic DNA from the parental CHO and the MTX-resistant CHOC 400 cell lines by hybridization to $^{32}$P-labeled cH1 DNA. Since CHOC 400 has at least 500-fold more DHFR gene copies than does the parental CHO cell line, a 50-fold excess (5 μg) of CHO DNA was loaded on the gel so that the hybridization signals from both cell lines could be detected in the same autoradiographic exposure. To eliminate slight differences in mobility between the two tracks due to DNA concentration differences, an equal amount of the CHO digest (5 μg) was added to 100 ng of the CHOC 400 digest. Note that under the conditions employed here, single copy sequences cannot be detected in 100 ng of CHOC 400 DNA.

When $^{32}$P-labeled cH1 DNA was hybridized to a transfer of this gel, all the bands in CHOC 400 that are recognized by this probe were present in the parental CHO genome, with a few exceptions (Fig. 4). In the *Kpn*I and *Bcl*I digests (Fig. 4, lanes e through h), faint bands (indicated by arrows) were present in CHOC 400 that were apparently absent in CHO. These bands could represent variants that have appeared during the amplification process. We estimate the copy number of these variant fragments to be less than 5% of the predominant amplified bands. In total, these hybridization data suggest that the amplified DHFR gene in CHOC 400 is a faithful representation of one of the DHFR alleles in CHO cells, and that the cloned gene represented by cH1 and overlapping cosmids represents a parental DHFR gene.

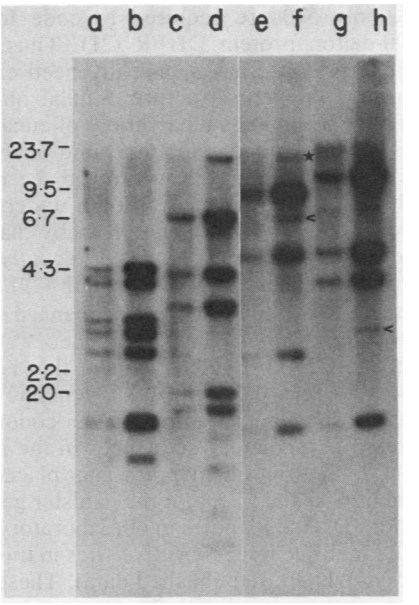


FIG. 4. Comparison of the DHFR genes in the parental CHO and MTX-resistant CHOC 400 cell lines. Genomic DNAs from CHO and CHOC 400 were digested with *Eco*RI (lanes a, b), *Hin*dIII (lanes c, d), *Bcl*I (lanes e, f), and *Kpn*I (lanes g, h); 5 μg of a CHO digest (left member of each pair) was compared with 100 ng of a CHOC 400 digest mixed with 5 μg of CHO digest. After separation on agarose gels and transfer to nitrocellulose, the digests were hybridized to $^{32}$P-labeled cH1 DNA. Positions of a *Hin*dIII digest of lambda DNA are indicated. Arrowheads indicate positions of bands discussed in text. The starred fragment in lane f is an artifact of this particular digest.

## DISCUSSION

Our purpose in constructing the CHOC 400 genomic library in a cosmid vector was to obtain relatively long genomic inserts in order to study the organization of the amplified DHFR gene locus and the process of amplification itself. To date, we have been able to isolate and map approximately 130 kb of the DHFR amplicon in seven overlapping cosmids (Milbrandt, unpublished data). This region includes the gene itself and an origin of replication (N. H. Heintz, J. D. Milbrandt, K. S. Greisen, and J. L. Hamlin, Nature [London], in press). Because of the relatively large inserts in our library, we have been able to identify a recombinant cosmid (cH1) that contains a functional DHFR gene by utilizing the spheroplast method of gene transfer (27). In the study reported here we utilized this cosmid and several overlapping clones representing regions flanking the gene to construct a restriction map of 70 kb of DNA included in the DHFR amplicon. We have used a murine DHFR cDNA, a genomic fragment containing the 5′ region of the mouse gene, and two Chinese hamster DHFR cDNA clones to define the boundaries of the Chinese hamster DHFR gene, as well as the approximate positions of the exons and introns. This Chinese hamster gene is estimated to be 24.5 kb in length, compared to a size of 31.5 kb in mouse cells (7). The hamster gene contains at least four introns, and, by analogy to the mouse gene, may contain a fifth small intron interrupting the single 5′ exon depicted in Fig. 3. The large size difference between the two homologous gene lies primarily in the 3′ intron, which is 8 kb in hamster and 16.5 kb in mouse. The distance from the start of each gene (as defined by the position of the 5′ probe utilized in Fig. 1) to the 5′ end of the 3′ intron is approximately 13 kb, and the exons are positioned similarly within this region in both genes. However, the restriction maps are entirely different throughout the length of the gene. This is not surprising, since, of the 24,500 nucleotides in the

gene, only 570 are required to code for the 21,000-dalton protein DHFR (21). Thus, little selective pressure has apparently been exerted on most of the gene structure. Similar observations have been made on a variety of eucaryotic genes, including hemoglobin (9) and ovalbumin (19). The small exon/intron ratio in the mouse and hamster genes is presently not understood. However, the DHFR gene in baby hamster kidney cells has been analyzed recently, and may be as small as 10.5 kb in length (29).

Although we have not yet sequenced the 5′ end of the Chinese hamster gene, we have isolated a 1.25-kb *Eco*RI fragment that is complementary to a 0.56-kb *Sau*3A genomic clone know to contain the initial AUG codon and upstream 5′ regulatory sequences in the mouse gene (7). Restriction mapping has placed this fragment near the 5′ end of the hamster gene. In addition, DNase I studies in our laboratory have located multiple hypersensitive sites in this fragment (Azizkhan, unpublished data). These sites have been correlated with the 5′ start signals of other transcriptionally active genes (10). Preliminary in vitro transcription run-off experiments have identified two alpha-amanitin-sensitive start sites and one *Pol*III start site within the 7.3-kb *Hin*dIII clone that includes this *Eco*RI fragment (N. Heintz, unpublished data).

The cosmid cH1 is particularly interesting because it contains a functional DHFR gene. Thus, we have been able to introduce this cosmid into DHFR$^-$ CHO cells and rescue cell lines that have been transformed to the DHFR$^+$ phenotype. This approach has also been used to identify a cosmid clone that contains the 28-kb Syrian hamster CAD gene by de Saint Vincent et al. (8). In both cases, the genes are very large and could not have been isolated intact in other cloning vectors.

The advantages of a clone containing a functional DHFR gene are numerous. When we began these studies, we had a truncated DHFR cDNA with which to locate the 3′ end of the gene, but we did not have a probe that defined the 5′ end. However, since cH1 was able to rescue DHFR$^-$ CHO cells, we were reasonably certain that it contained all those sequences at the 5′ end required for gene expression. Second, we were able to map essentially the entire gene from this single cosmid. In cases where there are more than one allelic version of a gene, which has been suggested for DHFR in Chinese hamster cells (20), defining the structure of a single allele would be complicated if the gene sequences were distributed among several clones. However, the results of our hybridization data with cH1 suggest that CHO cells contain only one allelic version of the DHFR gene (Fig. 4). Although variants comprising a small percentage of the amplified DHFR genes in CHOC 400 have been detected by this analysis, it is apparent that the "consensus" DHFR gene in the MTX-resistant CHOC 400 cell line is equivalent to a parental CHO gene.

JSH-0, the cell line that has been stably transformed by the cosmid cH1, is also interesting. It has been shown in the mouse system that there are multiple polyadenylation signals in the 3′ untranslated region of the DHFR gene (Carothers et al., in press). It is apparent from hybridization analyses presented here that cells transformed with cH1 produce a functional DHFR enzyme, even though the cosmid cH1 is lacking sequences in the 3′ untranslated portion of this gene (Fig. 1 and 2). This suggests that the CHO DHFR gene may also contain multiple polyadenylation and transcription termination signals, some of which may be included in cH1. The mRNA from this cell line should provide an interesting tool for studying the importance of multiple 3′ termination signals in the regulation of mRNA transcription.

Recently, DHFR mini-genes have been constructed and transferred to CHO DHFR$^-$ cells (12, 18). In these experiments it was possible to amplify these mini-genes by applying MTX, but in several of the cell lines, the amplified genes were unstable (18). Co-amplification and overexpression of simian virus 40 small t antigen has been achieved when this gene was linked to a DHFR mini-gene construct (18). In the accompanying paper (22) we show that cH1 can also be amplified after stable introduction into DHFR$^-$ CHO cells by spheroplast fusion. Thus, cH1 could serve as a selectable, amplifiable marker for the co-amplification and overexpression of a colinear gene.

## ACKNOWLEDGMENTS

We thank Larry Chasin and Robert Schimke for generously providing us with several clones. We also thank Carlton White for expert technical assistance, and Nicholas Heintz for helpful discussions.

This work was supported by Public Health Service grants GM26108 and AM22125 from the National Institutes of Health and by a March of Dimes Basil O'Connor Starter Grant. J.D.M. was supported by a National Institutes of Health postdoctoral fellowship, J.C.A. was supported by National Institutes of Health grant AM22125, and J.L.H. is the recipient of an American Cancer Society Faculty Research Award.

## LITERATURE CITED

1. **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.
2. **Bertino, J. R., S. Srimatkandada, D. Engel, W. D. Medina, D. I. Scheer, B. A. Moroson, and S. Dube.** 1982. Gene amplification in a methotrexate-resistant human leukemia line, K-562, p. 23–28. *In* R. T. Schimke (ed.), Gene amplification. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
3. **Biedler, J. L., and B. A. Spengler.** 1976. A novel chromosome abnormality in human neuroblastoma and antifolate-

resistant Chinese hamster cell lines in culture. J. Natl. Cancer Inst. 57:683–695.
4. **Bostock, C. J., and E. M. Clark.** 1980. Satellite DNA in large marker chromosomes of methotrexate-resistant mouse cells. Cell 19:709–715.
5. **Chang, A. C. Y., J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, and S. N. Cohen.** 1978. Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. Nature (London) 275:617–624.
6. **Clark, L., and J. Carbon.** 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell 9:91–99.
7. **Crouse, G. F., C. C. Simonsen, R. N. McEwan, and R. T. Schimke.** 1982. Structure of amplified normal and variant dihydrofolate reductase genes in mouse sarcoma S180 cells. J. Biol. Chem. 257:7887–7897.
8. **de Saint Vincent, R., S. Delbruck, W. Eckhart, J. Meinkoth, L. Vitto, and G. M. Wahl.** 1981. Cloning and reintroduction into animal cells of a functional CAD gene, a dominant amplifiable genetic marker. Cell 27:267–277.
9. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human beta-globin gene family. Cell 21:653–668.
10. **Elgin, S. C. M.** 1981. DNAase I-hypersensitive sites of chromatin. Cell 27:413–416.
11. **Enquist, L., and N. Sternberg.** 1979. In vitro packaging of lambda Dam vectors and their use in cloning DNA fragments. Methods Enzymol. 68:281–298.
12. **Gasser, C. S., C. C. Simonsen, J. W. Schilling, and R. T. Schimke.** 1982. Expression of abbreviated mouse dihydrofolate reductase genes in cultured hamster cells. Proc. Natl. Acad. Sci. U.S.A. 79:6522–6526.
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. U.S.A. 72:3961–3965.
15. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene 11:291–298.
16. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. 9:2989–2998.
17. **Kaufman, R. J., P. C. Brown, and R. T. Schimke.** 1979. Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes. Proc. Natl. Acad. Sci. U.S.A. 76:5669–5673.
18. **Kaufman, R. J., and P. A. Sharp.** 1982. Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary DNA gene. J. Mol. Biol. 159:601–621.
19. **Lai, E. C., S. L. C. Woo, A. Dugarczyk, and B. O'Malley.** 1979. The ovalbumin gene: alleles created by mutations in the intervening sequences of the natural gene. Cell 16:201–211.
20. **Lewis, J. A., J. P. Davide, and P. W. Melera.** 1982. Selective amplification of polymorphic dihydrofolate reductase gene loci in Chinese hamster lung cells. Proc. Natl. Acad. Sci. U.S.A. 79:6961–6965.
21. **Melera, P. W., D. Wolgemuth, J. L. Biedler, and C. Hession.** 1980. Antifolate-resistant Chinese hamster cells. J. Biol. Chem. 255:319–322.
22. **Milbrandt, J. D., J. C. Azizkhan, and J. L. Hamlin.** 1983. Amplification of a cloned Chinese hamster dihydrofolate reductase gene after transfer into a dihydrofolate reductase-deficient cell line. Mol. Cell. Biol. 3:1274–1282.
23. **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. U.S.A. 78:6043–6047.
24. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 433. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Nunberg, J. H., R. J. Kaufman, R. T. Schimke, G. Urlaub, and L. A. Chasin.** 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. Natl. Acad. Sci. U.S.A. 75:5553–5556.
26. **Osborn, M. J., J. E. Gander, E. Parisi, and J. Carson.** 1972. Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. J. Biol. Chem. 247:3962–3972.
27. **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.
28. **Schaffner, W.** 1980. Direct transfer of cloned genes from bacteria to mammalian cells. Proc. Natl. Acad. Sci. U.S.A. 77:2163–2167.
29. **Schoner, R. G., and J. W. Littlefield.** 1981. The organization of the dihydrofolate reductase gene of baby hamster kidney fibroblasts. Nucleic Acids Res. 9:6001–6014.
30. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
31. **Urlaub, G., and L. A. Chasin.** 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc. Natl. Acad. Sci. U.S.A. 77:4216–4220.
32. **Wahl, G. M., M. Stern, and G. R. Stark.** 1976. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. 76:3683–3687.