MOLECULAR AND CELLULAR BIOLOGY, Mar. 1989, p. 1137–1147
0270-7306/89/031137-11$02.00/0

Vol. 9, No. 3

Identification and Characterization of a Gene That Is Coamplified with Dihydrofolate Reductase in a Methotrexate-Resistant CHO Cell Line

PAMELA K. FOREMAN AND JOYCE L. HAMLIN*

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

Received 19 September 1988/Accepted 16 December 1988

As part of an effort to characterize the spatial and functional relationships among genetic elements within the amplified dihydrofolate reductase (DHFR) domain in Chinese hamster cells, we have used a variation of the differential hybridization approach to identify cDNA clones whose genes are coamplified with DHFR in the methotrexate-resistant cell line, CHOC 400. Our initial screen was successful in isolating both DHFR and non-DHFR cDNAs. One of the non-DHFR cDNA clones, 2BE2121, hybridizes on Northern (RNA) blots to abundant 1,200- and 1,500-nucleotide (nt) transcripts which differ in the lengths of their 3′ untranslated regions. The clone 2BE2121 contains a 789-nt open reading frame but does not appear to be related to any members of the protein or nucleic acid sequence databases. A second larger non-DHFR cDNA, II-19-211, was isolated that is transcribed from the same gene as 2BE2121 but contains only a small carboxyl-terminal portion of the open reading frame. II-19-211 may, therefore, represent either a splicing intermediate or an mRNA transcribed from a cryptic intragenic promoter. Hybridization to cosmids from the DHFR domain shows that 2BE2121 is encoded by a gene ~34 kilobases (kb) long. The 5′-most genomic fragment is less than 4 kb from an interamplicon junction. The 3′ end of the 2BE2121 gene lies ~75 kb downstream from the DHFR gene and ~25 kb downstream from the proximal replication initiation site, and the transcriptional polarity is opposite to that of the leading strand of replication. Thus, both the DHFR and 2BE2121 genes are exceptions to the theory that transcription proceeds in the same direction as the leading strand of the replication fork.

DNA sequence amplification in mammalian cells is a phenomenon frequently associated with resistance to selective agents. The actual mechanism(s) by which amplification occurs is not known, but unequal sister chromatid exchange and disproportionate replication are the two models most frequently invoked (for reviews, see references 19, 37, and 42). The units of amplification (amplicons) are invariably much larger than the gene whose product is selected, with the result that additional unselected genes are likely to occur within an amplified domain and consequently be overexpressed. The fortuitous amplification of linked genes has been reported in the amplicons of *N*-(phosphoracetyl)-L-aspartate-resistant Syrian hamster cell lines (47), in hydroxyurea-resistant Chinese hamster cells (41), and in multidrug-resistant (13, 45) and coformycin-resistant Chinese hamster cells (12). It is not clear what the consequences of fortuitous overexpression of linked genes might be to the cell or what role neighboring genes might play in determining the structural features of the amplicon itself. Presumably, the endpoints of the amplified sequence cannot occur in the body of a colinear, nonselected gene that is essential in two copies for cell viability, since the cell would not survive even the first selection step. However, the overproduction of certain protein products from colinear, nonselected genes could also be detrimental to cells, and amplicons large enough to fortuitously contain such a gene might be selected against. In addition, the relatively open chromatin configuration that characterizes active genes (16) could facilitate the recombination events that integrate amplified DNA into the chromosomes or double minutes, resulting in preferred endpoints near or in the bodies of genes. Thus, a definition of the relationships among functional elements in an amplified domain could shed light on the mechanisms responsible for amplification.

Because the amplicons in mammalian cells are so large, it is also likely that each unit contains at least one origin of replication. The spatial relationships among genes and replication origins may reveal important features of genome arrangement in higher eucaryotic cells. For example, it has been suggested that the direction in which a replication fork passes through a given gene may determine the transcriptional activity of that gene (21, 39, 43). In addition, a link between transcriptional control elements and sequences required for DNA replication has been observed in several viral systems that have served as models for mammalian chromosomal DNA replication (for a review, see reference 15).

In order to study DNA sequence amplification and to facilitate the study of spatial and functional relationships among genetic elements in a large chromosomal domain, we have developed a methotrexate (MTX)-resistant Chinese hamster ovary (CHO) cell line (CHOC 400) that has amplified the dihydrofolate reductase (DHFR) gene and flanking sequences approximately 1,000 times (29). We have recently cloned the equivalent of two complete DHFR amplicon types from this cell line that together account for >80% of the amplicons in the CHOC 400 genome (24). We have previously shown that replication initiates somewhere within a 28-kilobase (kb) locus whose 5′ boundary maps about 15 kb downstream from the 3′ end of the DHFR gene (20), and we have recently obtained evidence that there may, in fact, be two distinct initiation sites within this locus (23).

This system provides a unique opportunity to analyze the spatial relationships among genes, putative origins of repli-

* Corresponding author.

cation, and interamplicon junctions. For this report, we have initiated studies to identify genes that are coamplified with DHFR, and we have discovered at least one other transcription unit whose mRNA is overexpressed in the CHOC 400 cell line. We show that the 5′ end of this gene is located near an interamplicon junction and that transcriptional polarity is opposite to the direction of the leading strand of replication emanating from the proximal initiation site.

(This work constitutes part of the requirements for the Doctor of Philosophy degree in Biochemistry at the University of Virginia for P.F.).

MATERIALS AND METHODS

Cell lines and cell culture. The CHOC 400 cell line was developed from CHO cells in our laboratory (29) and is resistant to 800 μM MTX. The MTX-resistant MK42/400 cell line was also derived from CHO cells by Chasin and colleagues. The original cell line (MK42) contained ~150 copies of the DHFR gene (33). After further increases of MTX in our laboratory (to 800 μM), MK42/400 now contains ~300 copies of the gene. Both cell lines were maintained as previously described (25).

Enzymes and reagents. Unless otherwise indicated, enzymes were purchased from Bethesda Research Laboratories, Inc., Gaithersburg, Md., or Boehringer Mannheim Biochemicals, Indianapolis, Ind. The Ultra Pure CsCl and guanidinium thiocyanate used for RNA isolation, as well as agarose and low-melting-point agarose, were from Bethesda Research Laboratories. All other reagents were from Sigma Chemical Company, St. Louis, Mo.

Preparation of RNA and construction and screening of cDNA library. Total cellular RNA was prepared from CHOC 400 cells by the guanidinium thiocyanate-CsCl sedimentation method of Chirgwin et al. (8). Polyadenylated RNA was isolated by chromatography on oligo(dT)-cellulose (Collaborative Research; 1). The cDNA library was constructed in lambda gt10 by William Vorachek and William Pearson, University of Virginia. First and second strand cDNAs were prepared with an Amersham cDNA synthesis kit. cDNAs were blunt ended with S1 nuclease, were ligated via synthetic *Sma*I-*Eco*RI linker adapters (New England BioLabs, Inc., Beverly, Mass.) to lambda gt10, and were packaged with a Gigapack Gold kit obtained from Stratagene Corp., La Jolla, Calif.) The library was screened by the procedure of Benton and Davis (3) using selected fragments from individual cosmids as radioactive probes (see Results).

Subcloning and sequencing. Lambda phage plate lysate preparation, subcloning, plasmid preparation, and restriction mapping were carried out according to standard techniques (27, 11). cDNAs from recombinant phage were subcloned into the *Sma*I site of pGem-7Zf(+) (Promega Biotech, Madison, Wis.) or into the *Eco*RI site of pEMBL18 (14). Sequencing reactions were performed with a Sequenase kit (United States Biochemical Corp., Cleveland, Ohio) according to the instructions for double-stranded DNA, except that the labeling reaction was allowed to proceed for only 4.5 min at room temperature. Nested deletions for sequencing one strand of cDNA inserts cloned into pGem-7Zf(+) were prepared with an Erase-a-base kit (Promega). Selected synthetic 20-mers complementary to the opposite strands were obtained from Operon Technologies, Inc., San Pablo, Calif., and were used as primers for sequencing the opposite strand. Each of the sequencing reactions was performed at least twice (see Fig. 3). The sequence of genomic fragment k from cosmid II-19 was determined in the neighborhood of the putative splice site utilizing two oligonucleotide primers complementary to the II-19-211 cDNA in a region upstream from the sequence shared by cDNAs II-19-211 and 2BE2121 (see Results).

Preparation of hybridization probes. The radiolabeled cDNAs that were used in initial studies to screen digests of cosmids from the amplicon were made from polyadenylated RNA essentially as described by Schwarzbauer et al. (38), except that random hexanucleotides (P-L Biochemicals, Inc., Milwaukee, Wis.) were used to prime the reverse transcriptase reaction. DNA probes (either phage DNA, linearized plasmids, or restriction fragments excised from low-melting-point agarose gels) were labeled with ^{32}P-dCTP by the random hexanucleotide primer method (17).

DNA and RNA blotting. Restriction digests of cosmid or plasmid clones were separated on 0.7 to 1.0% agarose gels and, after the gel was soaked in 0.25 M HCl for 5 min, the digests were transferred overnight in 0.4 M NaOH–1.5 M NaCl (35) to GeneScreen Plus (Du Pont Research Products, Boston, Mass.). Polyadenylated RNA was separated on 1.2% agarose gels containing 5 mM methylmercury hydroxide (Organometallics, Inc., E. Hampstead, N.H.) (2) and was transferred to GeneScreen Plus overnight in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Hybridization to both Southern and Northern (RNA) transfers was carried out by the method of Church and Gilbert at 65°C (9). Blots were washed in 2× SSC–1% sodium dodecyl sulfate for 60 min at 65°C and 0.1× SSC–1% sodium dodecyl sulfate for 45 min at room temperature.

RESULTS

Configuration of amplified sequences in CHOC 400 cells. Figure 1 shows schematically the organization and genesis of the two amplicon types that we have cloned from the CHOC 400 genome (Fig. 1, panel A), as well as a representative collection of overlapping cosmids that were used to construct the maps of these amplicons (Fig. 1, panel B [24]). The 273-kb type I amplicon represents only ~5% of all of the amplicons in the highly resistant CHOC 400 cell line (24). The multiple copies of the type I sequence are arranged head to tail in the genome, resulting in the formation of a circularly permuted restriction map that begins and ends at the type I junction indicated by the breaks at the ends of the map in the overlapping cosmids, PA36 and BP7 (Fig. 1, A and B). Except for the junction itself, the map of the type I amplicon is identical to that of the parental CHO DHFR locus (24, 26). The 240-kb type II amplicon arose by a deletion and rearrangement of the type I sequence early in the MTX selection process and represents ~75% of all of the amplicons in the CHOC 400 genome (Fig. 1A [26]). The remaining 20% of the amplicons in this cell line have not been characterized in detail. The replication initiation sites that we have previously defined (20, 23) are present in both the type I and II amplicons, and are indicated with an I on the linear scale in Fig. 1B (note that the positions at which replication forks meet are shown as lying midway between adjacent initiation sites, but this has not been determined experimentally). The cosmids KM96 and KY143 in Fig. 1B were used in studies described below and were isolated from the MTX-resistant cell line A3 (4; C. Ma, unpublished data).

Selection of probes and isolation of cDNAs. In order to determine whether any genes in addition to DHFR are amplified and overexpressed in CHOC 400 cells, radiolabeled cDNAs were prepared from CHOC 400 and CHO polyadenylated RNA and were used individually to probe

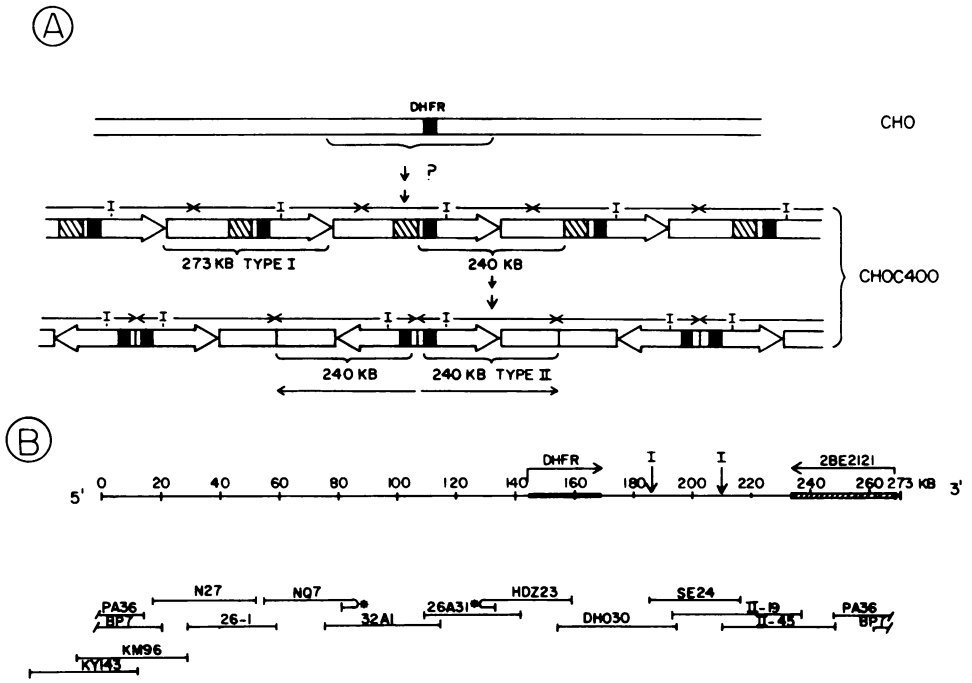

FIG. 1. Organization of the amplified DHFR locus in CHOC 400 cells. (A) The 273-kb type I amplicon in CHOC 400 cells (indicated by the large arrows in the middle panel) arose early during the MTX selection process by an unknown number of steps (indicated by a question mark). In a subsequent amplification step, the type II sequence arose by a truncation and rearrangement of the type I amplicon. Symbols: ■, DHFR gene; ▨, deleted region. The approximate location of the two closely spaced replication initiation loci in the domains is indicated with a single I in each amplicon, and bidirectional replication forks (arrows) are shown emanating from this locus. Note that we arbitrarily show the two forks meeting midway between each pair of initiation sites, but this has not been established experimentally. (B) A linear map of the type I amplicon. Symbols: ■, DHFR genes; ▨, 2BE2121 gene. The two replication sites are indicated with vertical arrows. A representative series of cosmids that spans the type I and type II amplicons is shown below the map. The asterisks at the hairpins in the cosmids HDZ23 and NQ7 indicate the head-to-head and tail-to-tail type II junctions (see text). Note that the type II amplicons extend from the head in HDZ23 through the type I junction to the tail in NQ7. The cosmids 32A1 and 26A31 contain sequences that were deleted from the type I amplicon during formation of the type II amplicon. KM96 and KY143 were cloned from the A3 cell line and represent the parental sequence arrangement in the region of the type I junction represented by cosmid PA36 (see text).

*Eco*RI digests of the overlapping series of cosmids that spans the CHOC 400 type I and II amplicons (Fig. 1B and 2). Overexpression of genes in the CHOC 400 cell line should result in the preferential illumination of genomic restriction fragments by the CHOC 400 cDNA probe if the steady-state levels of the corresponding mRNAs in the cell are high enough. The same *Eco*RI digests were also probed with total CHO genomic DNA in order to identify restriction fragments that contain highly repeated elements.

Several fragments were observed to hybridize to both CHO cDNA and total CHO genomic DNA (Fig. 2B, lanes B and C), suggesting that significant amounts of repetitive sequence elements are contained in CHO transcripts. In addition, however, several other fragments hybridized specifically to labeled cDNA from CHOC 400 cells (Fig. 2B, lanes A). Among these, fragments f, g, and j in cosmid HDZ23 and fragments f and i in cosmid DHO30 correspond to fragments contained in the DHFR gene (7, 28). Thus, an abundant message such as DHFR can be detected by this differential screening method. In addition, fragment c in NQ7, fragments f, i, and k in II-19, and fragments f, i, and j in II-45 hybridized strongly with the CHOC 400 probe. These regions of the genome therefore appear to contain genes in addition to DHFR that are overexpressed in the CHOC 400 cell line. Note that II-19 fragments f, i, and k and II-45 fragments f, i, and j are identical because of significant overlap between these two cosmids (Fig. 1B).

Several of the fragments that specifically hybridized to CHOC 400 cDNA were therefore used as probes, either in groups or singly, to screen a CHOC 400 cDNA library prepared in lambda gt10. (The more faintly hybridizing bands, i.e., h from 26A31 and a and h from HDZ23, were not used in this initial screen.) The probes were as follows: (i) a mixture of the DHFR-specific fragments f from DHO30 and j from HDZ23 to serve as positive controls on the screening protocol; (ii) a mixture of fragments f and i from II-19 and fragment j from cosmid II-45; and (iii) fragment c from cosmid NQ7. In addition, since labeled RNA prepared by in vitro transcription in CHOC 400 nuclei hybridized strongly to fragments a, b, and d in PA36 (data not shown), we selected PA36 fragment a as an additional probe (probe 4). The preferential hybridization of CHOC 400 cDNA to these PA36 fragments was apparently not detectable because of the significant amounts of repetitive DNA that each contains (Fig. 2B, panel PA36, lane C).

The DHFR-specific probe 1 detected two clones (1C111 and 1CD1211) which, in subsequent mapping studies, proved to have inserts that are ~800 and 1,450 base pairs (bp) in length and which correspond to DHFR cDNA. Probe 2 from cosmids II-19 and II-45 detected the cDNA II-19-211, which contains an insert of 1,437 bp. Probe 4 from cosmid PA36 detected the cDNA 2BE2121, which contains a 1,226-bp insert. The restriction maps of 2BE2121 and II-19-211 are shown in Fig. 3, A and B. Two additional cDNAs (3M121 and 5M121) were subsequently identified with probes 2 and 4 whose restriction maps are identical to those of II-19-211

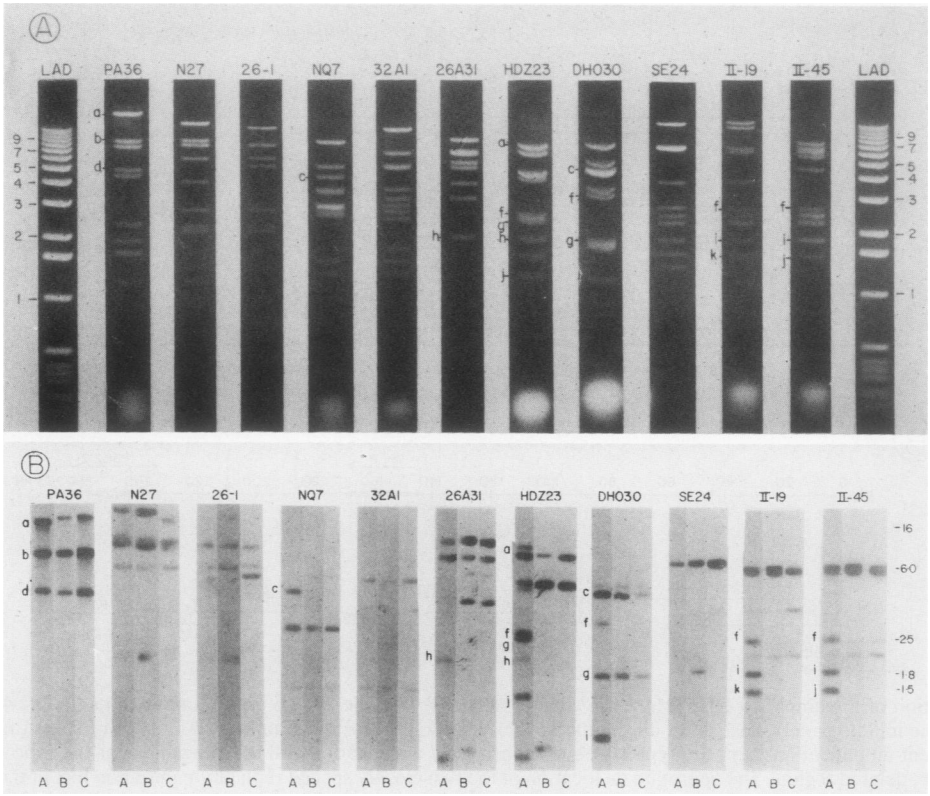

FIG. 2. Differential hybridization of total cellular cDNAs to cosmid fragments. (A) Samples (1.5 μg) of each of the cosmids shown in Fig. 1B (except KY143 and KM96) were digested with *Eco*RI and separated on a 0.8% agarose gel. Fragments in each lane were labeled alphabetically from highest to lowest in molecular weight, and bands discussed in the text are indicated. The size markers on the left and right are from a 1-kb ladder (Bethesda Research Laboratories). (B) Triplicate blots containing 500 ng of each of the *Eco*RI-digested cosmids were probed with one of the following in a total of 10 ml of hybridization buffer: 300 ng of [^{32}P]dCTP-labeled CHOC 400 cDNA (lanes A); 300 ng of labeled CHO cDNA (lanes B); 30 ng of labeled total CHO genomic DNA (lanes C). Blots were hybridized for 48 h at 65°C. Films were exposed for 16 h at −80°C with an intensifying screen.

and 2BE2121 except for slightly different junctions with the vector. Since these two clones were detected in separate screenings, we assume that they represent independent isolates and that the 2BE2121 and II-19-211 mRNA species are therefore relatively abundant in CHOC 400 cells. We have not yet been able to isolate a cDNA using probe 3 (NQ7 fragment c) to screen the library, for reasons that are not clear.

II-19-211 and 2BE2121 represent transcripts from the same genomic locus. In order to ascertain that the cDNA clones 1C111, 2BE2121, and II-19-211 arose from the genomic regions represented by the original probes used to screen the library, each cDNA was hybridized to *Eco*RI digests of the series of cosmids shown in Fig. 2A. Each clone hybridized to a 7- to 12-kb band in each cosmid digest because of the presence of a *cos* site in both the lambda and cosmid vectors (Fig. 4). In addition, 1C111 hybridized to fragments f, g, and j in cosmid HDZ23 and to fragments c, f, and g in cosmid DH030 (Fig. 4, panel A). These fragments were previously shown to contain parts of the DHFR gene (7, 28). Together with restriction mapping data (not shown), this result indicates that 1C111 corresponds to a DHFR cDNA. Essentially the same results were obtained with the independently isolated DHFR-specific clone 1CD1211 (data not shown).

The II-19-211 clone hybridized to fragments f and k in cosmid II-19 and to the corresponding fragments f and j in the cosmid II-45 (i.e., the restriction fragments in this region of the genome that were used to isolate this cDNA [Fig. 4C]). The cDNA 2BE2121 hybridized to fragments a, b, and d from cosmid PA36 but, surprisingly, also hybridized to the same set of fragments illuminated by II-19-211 (Fig. 4B). 2BE2121 and II-19-211 were also found to hybridize to one another (data not shown).

Since 2BE2121 hybridized to the 16-kb *Eco*RI fragment a (*Eco*RI-a) from cosmid PA36, the interesting possibility arose that 2BE2121 might cross the type I interamplicon junction that is contained in this fragment. 2BE2121 would therefore represent an aberrant mRNA that arose by the union of two different genes during the formation of the type I junction. It was also surprising that 2BE2121 hybridized to a larger region of the genome than did II-19-211 (Fig. 4, B and C), even though 2BE2121 has a smaller cDNA insert (see below; Fig. 3C). This result suggested that 2BE2121 and II-19-211 might represent alternatively spliced mRNAs arising from transcription of the same gene. Both of these possibilities were examined in the experiments outlined below.

Mapping of the ends of 2BE2121 with respect to the type I interamplicon junction. Further hybridization analyses (data not shown) and sequencing studies (see below) indicated that 2BE2121 and II-19-211 share sequences at their 3′ ends but diverge at their 5′ ends (Fig. 3C). We therefore reasoned that the unique 5′ end of 2BE2121 must contain the region that hybridizes to the additional fragments in cosmid PA36. In order to determine whether 2BE2121 sequences are located on both sides of the type I junction, a 344-bp *Eco*RI-*Pst*I

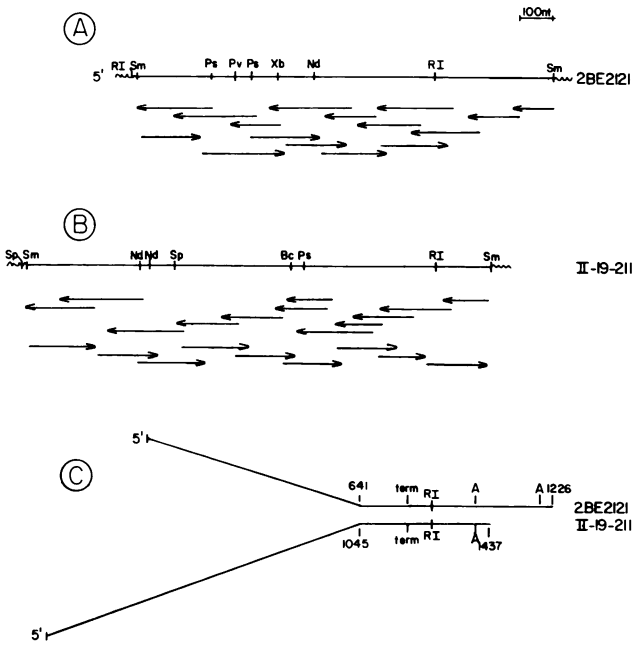

FIG. 3. Maps and sequencing strategies for 2BE2121 and II-19-211. (A and B) The cDNAs 2BE2121 and II-19-211 were each subcloned into the *Sma*I site of pGEM-7Zf(+) and were sequenced according to the schemes shown. Each arrow represents data obtained from a given set of sequencing reactions. The wavy lines at the ends of each insert indicate vector sequences, and the *Sma*I cloning sites define the ends of inserts. Restriction sites: Ps, *Pst*I; Pv, *Pvu*II; Xb, *Xba*I; Nd, *Nde*I; RI, *Eco*RI; Bc, *Bcl*I; Sm, *Sma*I; Sp, *Sph*I. Both the 5'-most *Eco*RI site in 2BE2121 and the 5'-most *Sph*I site in II-19-211 are in vector sequences. (C) Schematic diagram of the sequence relationships between the 2BE2121 and II-19-211 cDNAs. The sequence of II-19-211 is identical to that of 2BE2121 from nucleotide 1045 to its terminus at 1437, corresponding to nucleotides 641 to 1033 in 2BE2121. Abbreviations: term, termination codon in the open reading frames of 2BE2121 and II-19-211; A, consensus polyadenylation signals. Note that these cDNA maps are reversed relative to the genomic map in Fig. 1B.

fragment from the 5' end of 2BE2121 (Fig. 3A) was used to probe *Eco*RI digests of cosmids representing the type I junction region, as well as those of cosmids KM96 and KY143 that represent parental, nonrearranged sequences in this region of the DHFR locus. These cosmids contain all of the sequences in PA36 and BP7 that lie at the 5' end of the type I amplicon (i.e., sequences to the right of the break in cosmid PA36 [Fig. 1B]) but contain none of the sequences from the 3' end of the amplicon (i.e., sequences to the left of the break in PA36). The data in Fig. 5B show that this unique 5' probe hybridized to two fragments each in PA36 (a and d) and BP7 (a and c) but did not hybridize to any fragments in cosmids KM96 and KY143. We therefore conclude that the genomic sequences encoding 2BE2121 must lie solely at the 3' end of the type I amplicon and do not cross the junction. (Note also that a genomic *Eco*RI site must have been spliced out of the message, since the *Eco*RI-*Pst*I cDNA fragment hybridizes to two genomic *Eco*RI fragments in PA36 and BP7.)

The proximity of the end of 2BE2121 to the type I junction was determined more accurately in EcoRI-*Hin*dIII double digests of cosmid PA36. The left-hand lane of Fig. 5C shows the hybridization pattern obtained when an *Eco*RI-*Hin*dIII digest of PA36 was probed with the 16-kb junction-containing *Eco*RI fragment a, which illuminates its cognate fragments in PA36 (see map, Fig. 5A; asterisks in Fig. 5C denote faint cross-reaction with contaminating fragments in the probe).

When the *Eco*RI-*Hin*dIII digest of PA36 was probed with the 5' *Eco*RI fragment from 2BE2121 (Fig. 3A), only the 2.5-kb *Eco*RI-*Hin*dIII fragment and, to a lesser extent, the 3.6-kb *Hin*dIII fragment from *Eco*RI-a were illuminated (Fig. 5C). In addition, 2.7- and 0.7-kb *Hin*dIII-*Eco*RI fragments were detected that arose from *Eco*RI fragment d; this fragment maps on the 5' side of *Eco*RI-a in the genome and therefore also encodes part of the 2BE2121 cDNA that is contained in its 5' *Eco*RI fragment (see maps, Fig. 3 and 4D). Thus, the sequence that encodes the 5' end of the 2BE2121 insert lies somewhere within ~4 kb of the type I junction, which maps in the 0.6-kb *Hin*dIII fragment.

Northern (RNA) analysis of cellular RNA species. The cDNAs 1C111, 2BE2121, and II-19-211 were used to probe Northern RNA blots containing RNA from CHO, CHOC 400, and MK42/400 cells in order to characterize the corresponding cellular mRNA species. MK42/400 is an MTX-resistant CHO cell line that contains ~300 copies of the DHFR locus and is a more resistant derivative of the MK42 cell line originally developed by Chasin and colleagues (33).

When polyadenylated RNA from each of these cell lines was probed with the putative DHFR-specific cDNA, 1C111, three bands ~1,100, 1,800, and 2,400 nucleotides (nt) in length were detected (Fig. 6A). The sizes of these mRNAs correspond to those determined previously for CHO DHFR mRNAs by Carothers et al. (7) and Milbrandt et al. (28) and confirm the fact that 1C111 represents a DHFR mRNA. (Note that with all of the probes used in the experiment in Fig. 6, the signals from parental RNA samples were obtained after much longer film exposures than those from the drug-resistant cell lines [Fig. 6 legend].)

The 2BE2121 cDNA hybridized to 1,500- and 1,200-nt transcripts in all three cell lines (Fig. 6B). The fact that the transcripts are the same size in the CHO, MK42, and CHOC 400 cell lines provides further evidence that the gene encoding 2BE2121 does not cross the type I interamplicon junction, since the much larger amplicons in MK42 do not use this junction and, like CHO cells, contain parental-size fragments in this region of the DHFR locus (25).

In an effort to understand the origins of the two different 2BE2121 transcripts, we also hybridized an *Eco*RI-*Sma*I fragment from the 3' end of 2BE2121 (Fig. 3A) to RNA isolated from CHOC 400 cells (Fig. 6C). Whereas the full-length cDNA hybridized to both 1,500- and 1,200-nt mRNA species (Fig. 6C, left lane), the 3' fragment recognized only the larger transcript (Fig. 6C, right lane). This result showed that the smaller transcript is missing sequence elements present in the 3'-most part of the 2BE2121 insert and suggests that the two different-sized 2BE2121 transcripts arise from the use of alternative polyadenylation sites.

Not surprisingly, the complete II-19-211 cDNA clone also hybridized to 1,500- and 1,200-nt transcripts in all three cell lines (data not shown), since II-19-211 and 2BE2121 hybridize to each other. However, when the 460-bp *Sph*I fragment from the unique 5' end of II-19-211 (Fig. 3B) was used as a probe, transcripts of ~1,500 and ~1,200 nt were again detected in both resistant cell lines (Fig. 6D). Since sequences in this 5' probe are not contained in 2BE2121 (see Fig. 3C), the 1,500- and 1,200-nt bands that it detects must therefore represent RNA species different from the 1,500- and 1,200-nt transcripts detected with 2BE2121. In addition, since the 1,500-nt transcript appears to be more abundant

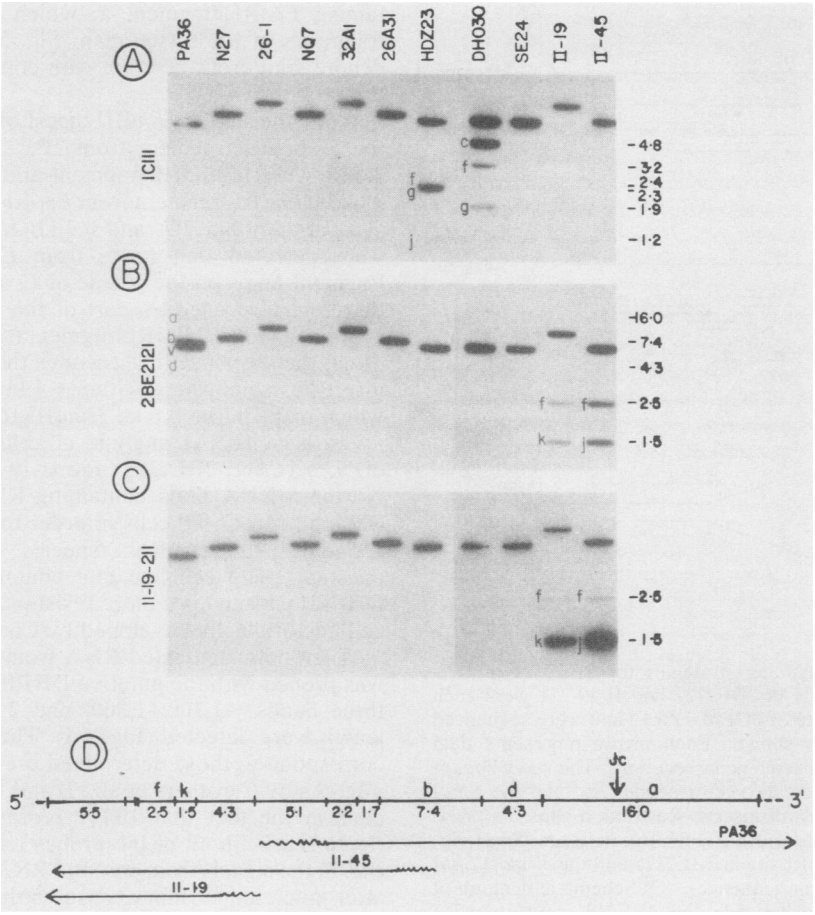

FIG. 4. Hybridization of isolated cDNAs to cosmid digests. (A through C) The lambda cDNA clones indicated to the left of each panel were radiolabeled and used to probe transfers of *Eco*RI digests of each of the cosmids spanning the amplified region. Fragments correspond to those identified in Fig. 2A. Each digest contains one pHC79 cosmid vector fragment that hybridizes strongly to the *cos* site in the lambda gt10 vector in each probe. Fragment sizes were determined from the 1-kb ladder in the ethidium bromide-stained gel before transfer. (D) An *Eco*RI map of the genomic region contained in the overlapping cosmids II-19, II-45, and PA36. *Eco*RI fragments are lettered as in Fig. 2 and 4. Note that this map has the same sense as the genomic map in Fig. 1B but a reading sense opposite to that of the cDNA maps in Fig. 3A and B.

than the 1,200-nt species, a more 3′ polyadenylation site may be used preferentially in the formation of these two transcripts. An extremely faint signal at the 1,500-nt position was also detected in CHO RNA with the II-19-211 5′ probe but cannot be visualized in the reproduction in Fig. 6D. Because of the faint signals in parental CHO RNA, we are therefore uncertain as to whether the 1,200-nt transcript is also present. Note that the transfers probed with the II-19-211 5′ probe in Fig. 6D were exposed to X-ray film for a considerably longer time than any of the others in this experiment, even though the probes all had roughly equal specific radioactivities. The II-19-211 transcripts must therefore be much less prevalent than either DHFR or 2BE2121 mRNAs in all three cell lines.

Sequence analysis of 2BE2121 and II-19-211. II-19-11 and 2BE2121 were sequenced according to the strategies outlined in Fig. 3, A and B. A terminal tract of A residues in 2BE2121 identifies the 3′ end of the clone (Fig. 7). There are two copies of the consensus polyadenylation signal, AATAAA (5, 10), in 2BE2121 that begin at positions 998 and 1194 in the 2BE2121 sequence (boxed in Fig. 7). However, since the 3′ *Eco*RI-*Sma*I fragment from 2BE2121 did not illuminate the 1,200-nt transcript in hybridization studies (Fig. 6C), the termination site for this species must lie very close to the *Eco*RI site at position 868, upstream from both of these consensus sites (Fig. 3A and 7).

An open reading frame begins with a serine residue at the 5′ end of 2BE2121 and terminates with a TGA at position 790 in the 2BE2121 sequence, upstream from the single *Eco*RI site at 868 (Fig. 7 and 3A). Several ATG codons that could represent the amino terminus of the predicted protein lie in the same reading frame in 2BE2121, including two at positions 106 and 136. However, the sequences surrounding these methionine codons do not resemble the optimal ACCATGG sequence for initiation of eucaryotic ribosomes (22). It therefore seems likely that 2BE2121 is missing the 5′ end of its cognate mRNA and that the actual translation initiation site lies further upstream. However, based on the estimated sizes of the corresponding mRNAs on Northern (RNA) blots, 2BE2121 is fewer than ~300 nt shorter than its mRNA, which distance could be accounted for in large part by the presence of a long poly(A) tract at the 3′ end of the transcript. We therefore propose that 2BE2121 encodes a polypeptide of slightly greater than 263 amino acids with a molecular mass of ~31,000 daltons. A search of the protein and nucleic acid databases found no significant similarities to any polypeptides described previously.

The sequence of the II-19-211 cDNA is compared with

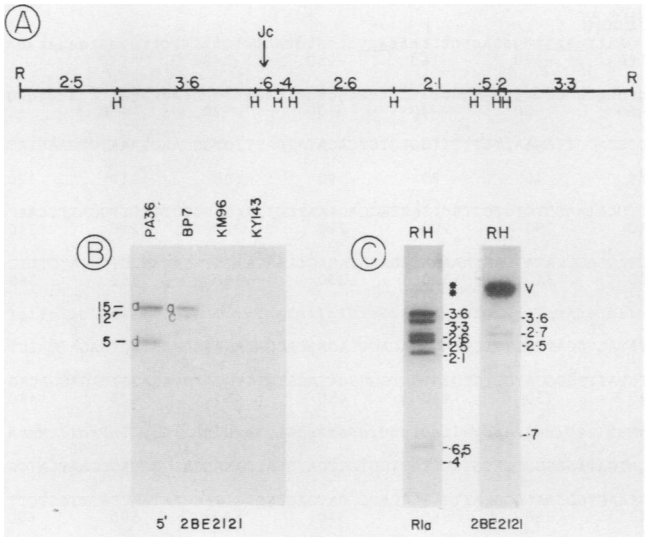

FIG. 5. Hybridization of 2BE2121 to sequences in the region of the type I junction. (A) The *Hin*dIII cleavage sites within the 16-kb *Eco*RI fragment a contained in cosmids PA36 and BP7 were previously mapped (26). Sizes of the fragments in kilobases are shown. The position of the type I junction in this fragment is indicated (Jc). (B). A 344-bp *Eco*RI-*Pst*I fragment from the 5′ end of the 2BE2121 insert (see Fig. 3A) was used to probe *Eco*RI digests of the cosmids PA36, BP7, KM96, and KY143 (see map, Fig. 1B). The probe reacts with *Eco*RI-a in both PA36 and BP7. PA36 fragment d and BP7 fragment c are also illuminated and represent the same genomic fragment (attached to vector in BP7). This 4.3-kb fragment maps to the DHFR-proximal side of the junction (i.e., between coordinates 245 and 273 in Fig. 1B). (C) *Eco*RI-*Hin*dIII digests of the cosmid PA36 were transferred to GeneScreen Plus and were hybridized with the 15-kb *Eco*RI a fragment isolated from PA36 (left-hand lane) or a subclone containing an 870-bp *Eco*RI fragment from the 5′ end of 2BE2121 (right-hand lane; see Fig. 3A). The sizes of bands in lanes 1 and 2 are indicated and correspond to fragments in the map in panel A. Asterisks in lane 1 indicate faint bands resulting from slight contamination of the probe with other restriction fragments in the PA36 *Eco*RI digest from which the probe was isolated. V, Vector sequences in the cosmid PA36 that hybridize with the vector in the cloned fragment.

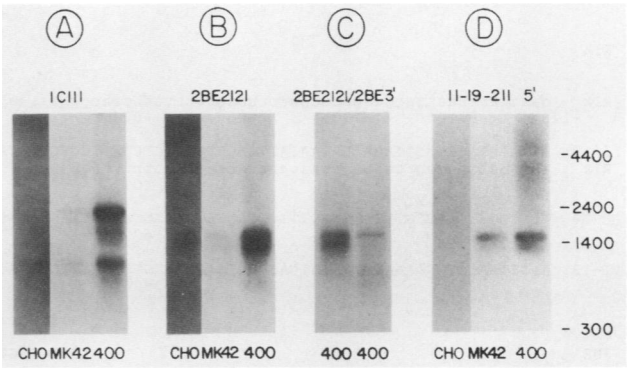

FIG. 6. Northern analysis of cDNAs. Polyadenylated RNA was prepared from the cell lines indicated under each panel, and 4 μg was separated in each of several wells of a 1.2% methylmercury hydroxide-containing agarose gel. After transfer to GeneScreen Plus, the samples were hybridized with one of the following probes: 1C111 cDNA (panel A); 2BE2121 cDNA (panel B); 2BE2121 cDNA (panel C, lane 1); the 350-bp *Eco*RI-*Sma*I fragment from the 3′ end of 2BE2121 (panel C, lane 2 [see Fig. 3A]); the 460-bp *Sph*I fragment from the 5′ end of the cDNA II-19-211 (panel D). The CHO samples in lane 1, panels A and B, were each exposed for 48 h at −80°C with an intensifying screen. Lanes 2 and 3 in panels A and B and both lanes in panel C were exposed for 18 h at 25°C without a screen. The CHO sample and the MK42 and CHOC 400 samples in panel D were exposed for 3 weeks and 3 days, respectively, at −80°C with an intensifying screen. Size markers to the right of panel D were determined from an RNA ladder obtained from Bethesda Research Laboratories.

that of 2BE2121 in Fig. 7. No poly(A) tract was observed at either end of II-19-211. The 5′ and 3′ ends of the insert were thus tentatively assigned so that the sequence shared with 2BE2121 had the same transcriptional polarity. II-19-211 is identical to 2BE2121 from nucleotide 1045 to its 3′ end but lacks the 193 nt at the terminus of 2BE2121 (shown schematically in Fig. 3C). The two clones, therefore, share the last 146 nucleotides of the 2BE2121 open reading frame. The II-19-211 sequence appears not to contain any open reading frames longer than 146 nt in any other reading frame when assessed on either DNA strand. There is also a stop codon (TGA) in the 2BE2121 reading frame that lies 12 nt upstream from the shared region. Thus, II-19-211 could not encode a protein product longer than 48 residues.

It seemed likely that the point at which the sequences of 2BE2121 and II-19-211 diverge represents, or is close to, a splice junction and that the two different mRNAs might result from differential splicing. In order to investigate this possibility, we sequenced the genomic fragment that must contain the presumptive splice acceptor site (*Eco*RI fragment k from cosmid II-19 [Fig. 4D]). The order of the relevant cDNA-hybridizing *Eco*RI fragments in this region of the genome reading 5′ to 3′ in Fig. 4D is f, k, b, d, and a (note that this map has the same sense as the genomic map in Fig. 1B but is reversed relative to the cDNA maps in Fig. 3C). Fig. 3C shows that an *Eco*RI site lies within the sequence shared by 2BE2121 and II-19-211. Since both cDNAs illuminated fragments f and k (Fig. 4, B and C), we assumed that this *Eco*RI site lies between these two fragments and that the sequences diverge in the more 5′ fragment k.

An oligonucleotide primer was therefore prepared for a region contained in II-19-211 cDNA (positions 887 to 906) that lies just upstream from the point of divergence from 2BE2121. We found that the sequence of fragment k in this region of the genome is contiguous with that of II-19-211 throughout the length of the cDNA (Fig. 7). II-19-211 is therefore not spliced in this region. II-19-211 cDNA sequences 5′ to the shared region therefore may lie within an intron in the gene that encodes 2BE2121. The sequence of the 2BE2121 cDNA diverges from the genomic sequence at approximately the same point that it diverges from the sequence of II-19-211 and therefore must be spliced at this point (see arrow, Fig. 7). The sequence of the genomic DNA close to the point of divergence of the two cDNAs contains a 3′ splice junction consensus beginning at position 1046 in the II-19-211 cDNA sequence (CAG ↓ G preceded by a pyrimidine-rich region [31]), and the sequence of the 2BE2121 cDNA at this point (position 644-645) is consistent with the joining of a splice donor to a splice acceptor sequence (CAGG [Fig. 7]).

The genomic sequence upstream from the 5′ end of II-19-211 in cosmid II-19 fragment k was also determined (beginning 182 nt upstream from the 5′ end of II-19-211 [Fig. 7, first three lines]). Interestingly, several sequence motifs resembling those of promoter elements were observed in this region. The sequence TATAA is located 51 nt upstream from the beginning of the II-19-211 cDNA at position −51 (under-

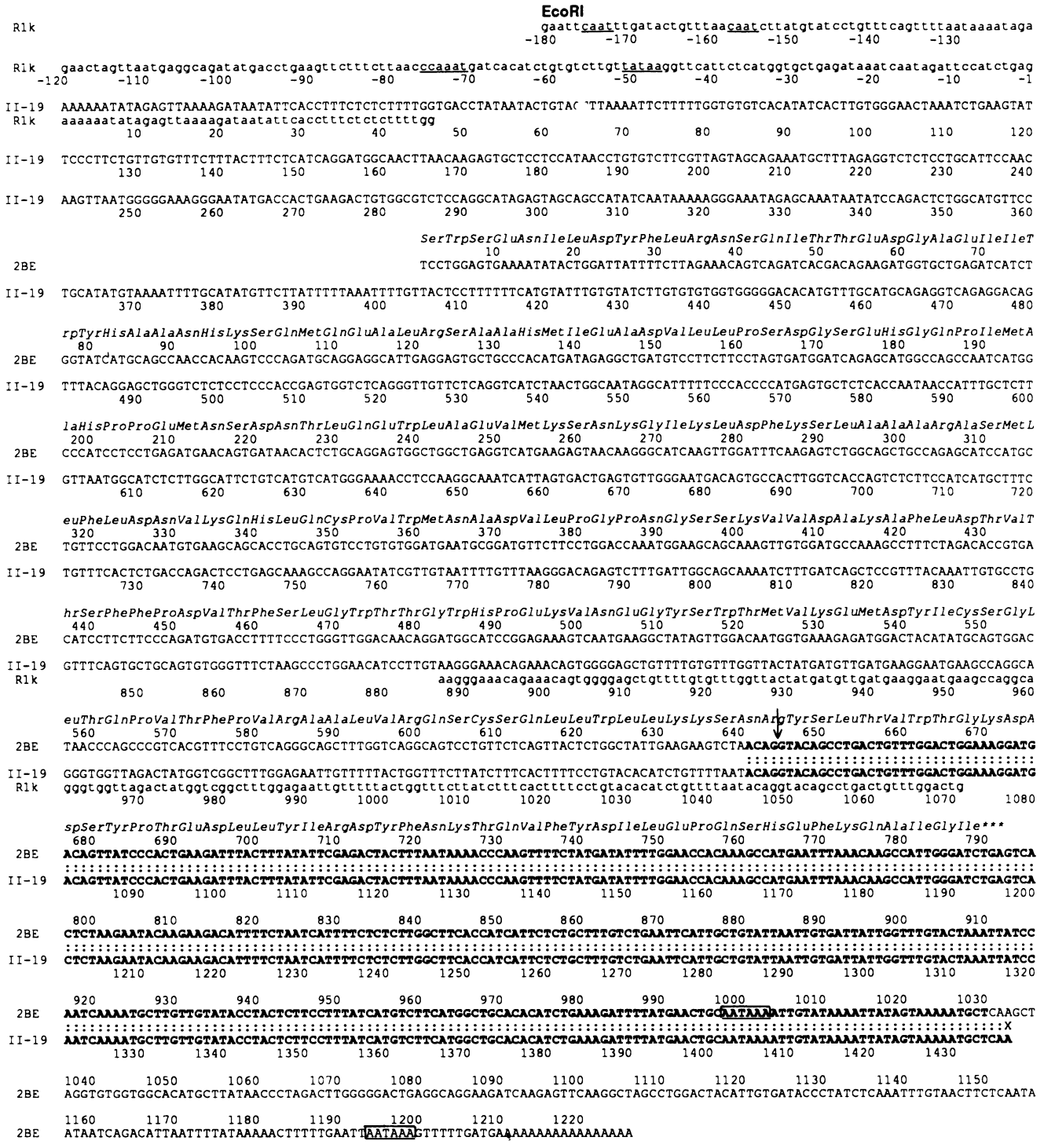

FIG. 7. Nucleotide sequences of the cDNAs 2BE2121 and II-19-211 and relevant genomic regions. The nucleotide sequences of the subcloned 2BE2121 and II-19-211 cDNAs were determined according to the strategies described in the legend to Fig. 3 A and B and are aligned to show the region of identity between the two clones (bold letters and double dots). The deduced amino acid sequence of the long open reading frame beginning at the 5′ end of 2BE2121 is shown above in italics. The consensus polyadenylation signals at positions 998 and 1194 in the 2BE2121 sequence (AATAAA) are boxed. The arrow at 2BE2121 position 644 to 645 indicates the proposed site of splicing in this cDNA. Part of genomic fragment k was also sequenced near the probable splice junction (see text), and the sequence is aligned under residues 887 to 1071 of II-19-211. Note the consensus splice site beginning at position 1046 in the II-19-211 sequence. In addition, the part of the upstream sequence from genomic fragment k that contains the presumed 5′ end of the II-19-211 cDNA was subcloned into pGEM-7Zf(+) and partially sequenced. The first 220 nucleotides are shown (first three lines), and the sequence begins 182 nucleotides from the 5′ end of fragment k. Sequences resembling consensus promoter elements (TATAAT and CCAAT) are underlined.

lined in Fig. 7). The location of this sequence would be consistent with a TATA box function if the cDNA were missing ~20 nucleotides that are present at the 5′ end of its cognate mRNA. In addition, near matches to the CCAAT box promoter element (CAAT and CCAAAT) are found at −176, −158, and −76 nt from the 5′ end of II-19-211 (underlined in Fig. 7). It is also therefore possible that II-19-211 is transcribed from a cryptic promoter lying within a 2BE2121 intron.

DISCUSSION

We have used a variation of the differential hybridization approach to isolate two cDNA clones (2BE2121 and II-19-211) corresponding to mRNAs that are coamplified with DHFR in CHOC 400 and in MK42/400 cells. Several other *Eco*RI fragments from the amplicon, in addition to those that hybridized to 2BE2121, II-19-211, and DHFR cDNAs, were also detected by this method (Fig. 2B). We have recently isolated an additional small cDNA that hybridizes to fragment i in cosmids II-19 and II-45, but this clone has not yet been characterized. Whether the remaining CHOC 400 cDNA-specific amplicon fragments also represent amplified genes is yet to be determined; these include fragment c in cosmid NQ7, for which we were unable to isolate a cDNA, fragments a and h in cosmid HDZ23, and fragment h in 26A31, which is identical to h in HDZ23. Except for fragment c, these fragments have not yet been used to screen the CHOC 400 cDNA library.

Mitchell et al. have previously identified an additional gene that is co-amplified with DHFR (30). This gene is thought to initiate transcription in the same promoter region as the DHFR gene but is transcribed in the opposite direction and encodes a 4-kb mRNA. It is possible that fragments a and h in HDZ23, which hybridized specifically to CHOC 400 cDNA but which were not used as probes to screen the library, correspond to this transcript.

Many of the fragments that hybridized to CHOC 400 cDNA and virtually all of the fragments that hybridized to CHO cDNA contain repetitive sequences (Fig. 2B). These fragments were not selected as probes for screening the cDNA library even though they may also represent parts of additional transcription units. The approach we used may therefore have failed to identify all of the genes contained in the amplified DHFR domain in CHOC 400 cells. It is possible that the use of unlabeled CHO genomic DNA to block repetitive elements in the radioactive cDNA probe could allow the identification of additional transcripts. The advantage of this differential hybridization approach is that it allows us to focus on cDNAs derived from genes that are de facto linked to the DHFR gene.

Two possible models may explain the relationship between the two cDNAs that we have described in this report. First, II-19-211 could represent a splicing intermediate in the formation of 2BE2121, since 1,048 nt at the presumptive 5′ end of II-19-211 appear to lie within an intron of the 2BE2121 gene. The 5′ sequences unique to 2BE2121 probably represent one or more additional exons that lie upstream from this region in the gene. Since we have shown previously that DHFR transcripts represent 25 to 30% of the total cellular RNA in CHOC 400 cells (28, 29) and since the 2BE2121 genes appear to transcribe similarly large amounts of mRNA (Fig. 6), the splicing apparatus may become saturated with these extra transcripts, resulting in the accumulation of intermediates. Intermediates of the proposed II-19-211 type (i.e., cleaved at a 5′ splice site but unligated) have been observed in mutated beta-globin RNAs in vitro (18). In addition, polyadenylated splicing intermediates of ovalbumin and ovomucoid RNAs have been observed in vivo (43). Interestingly, we have occasionally observed DHFR transcripts of higher molecular weight than that of the mature DHFR mRNAs in Northern (RNA) blots of CHOC 400 RNA (28), suggesting that DHFR precursor RNAs may also accumulate in these cells. II-19-211 type intermediates would be expected to occur in the form of a lariat, however, based on the studies of Padgett et al. (34) and Ruskin et al. (36), and it is unclear how a cDNA could be generated from a molecule that would contain a 2′ to 5′ phosphodiester-linked branch.

Alternatively, II-19-211 could be the result of transcription from a cryptic promoter lying within an intron that is spliced out to form 2BE2121 cDNA. Sequences within this region of the genome that correspond to the 5′ end of II-19-211 are consistent with this possibility (Fig. 7). If this is the case, however, it seems likely that II-19-211 results from aberrant transcription initiation, since the II-19-211 transcript does not encode a long open reading frame on either strand. However, we cannot completely rule out the possibility that this cDNA encodes a short polypeptide. It should also be noted that we have not unambiguously demonstrated the transcriptional polarity of II-19-211, and it is therefore possible that it could represent an antisense RNA. Antisense RNAs have recently been shown to be transcribed from several mammalian genes and could be involved in their regulation (32, 40, 48).

Hybridization with probes from the 3′ end of 2BE2121 showed that the 1,500- and 1,200-nt species on Northern (RNA) blots probably arise from the use of different 3′ polyadenylation signals, one of which correlates with the presence of the consensus hexanucoleotide polyadenylation signal, AATAAA, at position 1194 in the 2BE2121 sequence (Fig. 7). The DHFR gene also produces multiple-size mRNA molecules differing in the positions of their 3′ termini (46). II-19-211 also appears to utilize two termination sites, since a probe unique to this cDNA hybridized to both 1,500- and 1,200-nt bands on Northern (RNA) blots. Because the II-19-211 cDNA lacks most of its 3′ terminus, we were unable to correlate these termination sites to those used in 2BE2121.

Hybridization to cosmids spanning the predominant DHFR amplicons in CHOC 400 cells shows that the gene that encodes 2BE2121 is ~34 kb in length, and the sequences corresponding to the 5′ end of 2BE2121 are less than 4 kb from the type I interamplicon junction (Fig. 5). However, since the 2BE2121 cDNA is probably 50 to 100 bp short of being full length, the real 5′ end of this mRNA could, in principle, have been transcribed from the 5′ end of a second gene lying on the other side of the type I junction. It is also possible that altered chromatin structures associated with the 5′ end of the gene played a role in the formation of the junction. The site of the type I junction does not represent a recombinational hotspot, however, since none of the several other MTX-resistant Chinese hamster cell lines that we have examined contain interamplicon junctions at this site (25).

Surprisingly, we have not detected any polypeptides in the predicted molecular mass range whose steady-state levels are as high as might be expected based on the prevalence of the 2BE2121 mRNA in CHOC 400 cells (29), although we have not yet ruled out the presence of small amounts of a protein with a very short half-life. CHOC 400 cells may therefore utilize posttranscriptional regulation to control the level of the 2BE2121 product. Additional studies may reveal important features of the regulation of 2BE2121 gene expres-

sion that place limitations on its overexpression. An understanding of such limitations could be important in studies designed to overexpress unselected genes that have been linked to a selectable marker and artificially amplified for the purpose of generating large amounts of a protein product.

Sequencing analysis and orientation of the polyadenylated terminus of 2BE2121 with respect to the genome show that 2BE2121 and DHFR are convergently transcribed, with the replication initiation sites in this domain situated approximately midway between the two genes (Fig. 1B). Both genes thus appear to be transcribed in a direction opposite to that in which they are replicated. A similar opposing replication-transcription pattern was found for the murine heavy-chain constant-region locus in a study by Braunstein et al. (6). Our results are therefore exceptions to the proposal that DNA replication in the transcriptional direction may be a necessary condition for gene expression (39, 43).

ACKNOWLEDGMENTS

We thank our colleagues Bill Pearson, Kevin Lynch, and Tom Parsons for lengthy and very helpful discussions, and particularly Bill Pearson for instructing and aiding us in all aspects of computer analysis of sequence data. Special thanks are extended to Bill Vorachek for preparation of the cDNA library and to Larry Chasin (Columbia University, New York, N.Y.) and June Biedler (Memorial Sloan-Kettering Cancer Center) for the MK42 and A3 cell lines. Our technicians, Carlton White and Kevin Cox, provided expert and dedicated assistance throughout this project.

This work was supported by a Public Health Service grant (GM 26108) to J.L.H. from the National Institutes of Health and by a grant from the American Cancer Society. J.L.H. was the recipient of an American Cancer Society Faculty Research Award.

LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
2. **Bailey, J. M., and N. Davidson.** 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal. Biochem. **70**:75–85.
3. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
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. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**:349–359.
6. **Braunstein, J. D., D. Schulze, T. DelGiudice, A. Furst, and C. L. Schildkraut.** 1982. The temporal order of replication of murine immunoglobulin heavy chain constant region sequences corresponds to their linear order in the genome. Nucleic Acids Res. **10**:6887–6902.
7. **Carothers, A. M., G. Urlaub, N. Ellis, and L. A. Chasin.** 1983. Structure of the dihydrofolate reductase gene in Chinese hamster ovary cells. Nucleic Acids Res. **11**:1997–2012.
8. **Chirgwin, J. M., A. E. McDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
9. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
10. **Cole, C. N., and T. P. Stacy.** 1985. Identification of sequences in the herpes simplex virus thymidine kinase gene required for efficient processing and polyadenylation. Mol. Cell. Biol. **5**: 2104–2113.
11. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics: a manual for genetic engineering. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
12. **Debatisse, M., B. R. de Saint Vincent, and G. Buttin.** 1984. Expression of several amplified genes in an adenylate-deaminase overproducing variant of Chinese hamster fibroblasts. EMBO J. **3**:3123–3127.
13. **deBruijn, M. H. L., A. M. Van der Bliek, J. L. Biedler, and P. Borst.** 1986. Differential amplification and disproportionate expression of five genes in three multidrug-resistant Chinese hamster lung cell lines. Mol. Cell. Biol. **6**:4717–4722.
14. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single-stranded plasmids. Nucleic Acids Res. **11**: 1645–1655.
15. **DePamphilis, M. L.** 1988. Transcriptional elements as components of eukaryotic origins of DNA replication. Cell **52**:635–638.
16. **Elgin, S. C. R.** 1981. DNAseI hypersensitive sites of chromatin. Cell **27**:413–415.
17. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
18. **Furdon, P. J., and R. Kole.** 1986. Inhibition of splicing but not cleavage at the 5′ splice site by truncating human beta-globin pre-mRNA. Proc. Natl. Acad. Sci. USA **83**:927–931.
19. **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.
20. **Heintz, N. H., and J. L. Hamlin.** 1982. An amplified chromosomal sequence that includes the gene for dihydrofolate reductase initiates replication within specific restriction fragments. Proc. Natl. Acad. Sci. USA **79**:4083–4087.
21. **James, C. D., and M. Leffak.** 1986. Polarity of DNA replication through the avian alpha-globin locus. Mol. Cell. Biol. **6**:976–984.
22. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
23. **Leu, T.-H., and J. L. Hamlin.** 1989. High-resolution mapping of replication fork movement through the amplified dihydrofolate reductase domain in CHO cells by in-gel renaturation analysis. Mol. Cell. Biol. **9**:523–531.
24. **Looney, J. E., and J. L. Hamlin.** 1987. Isolation of the amplified dihydrofolate reductase domain from methotrexate-resistant Chinese hamster ovary cells. Mol. Cell. Biol. **7**:569–577.
25. **Looney, J. E., C. Ma, T.-H. Leu, W. B. Flintoff, W. B. Troutman, and J. L. Hamlin.** 1988. The dihydrofolate reductase amplicons in different methotrexate-resistant Chinese hamster cell lines share at least a 273-kb core sequence, but some amplicons are much larger and remarkably uniform in structure. Mol. Cell. Biol. **8**:5268–5279.
26. **Ma, C., J. E. Looney, T.-H. Leu, and J. L. Hamlin.** 1988. Organization and genesis of dihydrofolate reductase amplicons in the genome of a methotrexate-resistant Chinese hamster ovary cell line. Mol. Cell. Biol. **8**:2316–2327.
27. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
28. **Milbrandt, J. D., J. C. Azizkhan, and J. L. Hamlin.** 1983. Amplification of a cloned Chinese hamster ovary dihydrofolate reductase gene after transfection into a dihydrofolate reductase-deficient cell line. Mol. Cell. Biol. **3**:1274–1282.
29. **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 kb region that includes the gene for dihydrofolate reductase. Proc. Natl. Acad. Sci. USA **78**: 6043–6047.
30. **Mitchell, P. J., A. M. Carothers, J. H. Han, J. D. Harding, E. Kas, L. Veniola, and L. A. Chasin.** 1986. Multiple transcription start sites, DNase I-hypersensitive sites, and an opposite-strand exon in the 5′ region of the CHO DHFR gene. Mol. Cell. Biol. **6**:425–440.
31. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
32. **Nepveu, A., and K. B. Marcu.** 1986. Intragenic pausing and antisense transcription within the murine c-myc locus. EMBO J. **5**:2859–2865.
33. **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. USA **75**:5553–5556.

34. **Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp.** 1984. Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. Science **225**:898–903.
35. **Reed, K. C., and D. A. Mann.** 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13**: 7207–7221.
36. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell **38**:317–331.
37. **Schimke, R. T.** 1984. Gene amplification in cultured animal cells. Cell **37**:705–713.
38. **Schwarzbauer, J. E., J. W. Tamkun, I. R. Lemischka, and R. O. Hynes.** 1983. Three different fibronectin mRNAs arise by alternative splicing within the coding region. Cell **35**:421–431.
39. **Smithies, O.** 1982. The control of globin and other eukaryotic genes. J. Cell. Physiol. Suppl. **1**:137–143.
40. **Soma, G.-I., M. Murata, N. Kitahara, T. Gatanaga, H. Shibai, H. Morioka, and T. Andoh.** 1985. Detection of a countertranscript in promyelocytic leukemia HL60 cells during early differentiation by TPA. Biochem. Biophys. Res. Commun. **132**: 100–109.
41. **Srinivasan, P. R., P. N. Tonin, E. J. Wensing, and W. H. Lewis.** 1987. The gene for ornithine decarboxylase is co-amplified in hydroxyurea-resistant hamster cells. J. Biol. Chem. **262**:12871–12878.
42. **Stark, G. R., and G. M. Wahl.** 1984. Gene amplification. Annu. Rev. Biochem. **53**:447–491.
43. **Trempe, J. P., Y. I. Lindstrom, and M. Leffak.** 1988. Opposite replication polarities of transcribed and nontranscribed histone H5 genes. Mol. Cell. Biol. **8**:1657–1663.
44. **Tsai, M.-J., A. C. Ting, J. L. Nordstrom, W. Zimmer, and B. W. O'Malley.** 1980. Processing of high molecular weight ovalbumin and ovomucoid precursor RNAs to messenger RNA. Cell **22**:219–230.
45. **Van der Bliek, A. M., T. Van der Velde-Koerts, V. Ling, and P. Borst.** 1986. Overexpression and amplification of five genes in a multidrug-resistant Chinese hamster ovary cell line. Mol. Cell. Biol. **6**:1671–1678.
46. **Veniola, L., G. Urlaub, and L. A. Chasin.** 1987. Polyadenylation of Chinese hamster dihydrofolate reductase genomic genes and minigenes after gene transfer. Somatic Cell Mol. Genet. **13**: 491–504.
47. **Wahl, G. M., L. Vitto, and J. Rubnitz.** 1983. Co-amplification of rRNA genes with CAD genes in *N*-(phosphonacetyl)-L-aspartate-resistant Syrian hamster cells. Mol. Cell. Biol. **3**:2066–2075.
48. **Williams, T., and M. Fried.** 1986. A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature (London) **322**:275–279.