Structure of the dihydrofolate reductase gene in Chinese hamster ovary cells

Adelaide M. Carothers, Gail Urlaub, Nathan Ellis and Lawrence A. Chasin*

902 Fairchild Center, Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Received 20 January 1983; Revised and Accepted 3 March 1983

ABSTRACT

Overlapping recombinant λ 1059 phages carrying regions of the *dhfr* locus from the amplified Chinese hamster ovary (CHO) cell clone MK42 have been isolated. In addition, *dhfr* cDNAs from this cell line have been cloned into plasmid pBR322. Restriction analysis of these recombinant molecules has led to a map of the Chinese hamster *dhfr* gene. This gene has a minimum size of 26 kb and contains six exons as defined by hybridization to a combination of mouse and CHO cDNA probes. The latter probes reveal 3' exonic sequences that are not present in mouse cDNA. The CHO *dhfr* gene thus extends about 700 bp further 3' than in the mouse, consistent with the larger size of the hamster mRNA. At least five intervening sequences are present, of approximate sizes: 0.3, 2.5, 8.6, 2.6 and 9.4 kb. Four sequences from highly repeated families are situated in introns within the *dhfr* gene. The overall structure of this gene is strikingly similar to that of the mouse. Evolutionary conservation of interrupted gene structure among mammals thus extends to genes that code for household enzymes as well as specialized or structural proteins.

INTRODUCTION

The characterization of mutational lesions which occur in cultured mammalian cells requires both the ability to select specific-locus mutants, and the ability to locate the lesions by fine structure genetic mapping. A selection procedure based on resistance to tritiated deoxyuridine suicide has been devised in order to isolate dihydrofolate reductase (DHFR) negative mutants of Chinese hamster ovary (CHO) cells (1). Analysis of mutations at this locus should aid in determining what aspects of gene structure are necessary for gene expression, and in defining the biological consequences of spontaneous and induced mutation at the DNA level. An examination of the DNA of several independently isolated DHFR-deficient mutant cell lines was carried out by Southern blot analysis (2), using complementary DNA (cDNA) derived from the mouse as a hybridization probe (3). In that preliminary study, alterations were detected in genomic CHO restriction banding patterns at the *dhfr* locus. The most likely origin of those gross changes was considered to be deletions

induced by the ionizing radiation used as a mutagen. In order to define the location and extent of these mutations and others to be analyzed in the future, a complete restriction map of the wild-type dhfr locus in CHO cells is required.

In this paper, we have used the same mouse probe and others to clone and map the CHO dihydrofolate reductase gene. The gene with a small extent of flanking sequences was derived from a CHO cell line, clone MK42, in which the dhfr gene copy number is amplified 150-fold relative to wild-type cells (4). The amplified gene appears similar to the wild-type by blot hybridizations (3). The paper will discuss: 1) cloning of MK42 DNA into bacteriophage and plasmid vectors, 2) cloning of MK42 complementary DNA into a plasmid vector, 3) mapping of the genomic dhfr DNA fragments, 4) locating DHFR exons, 5) locating repeated sequences in the locus, and 6) genomic confirmation of the composite gene map.

MATERIALS AND METHODS

1. Materials

The enzymes used in this work were: E. coli DNA polymerase Klenow fragment, calf thymus terminal deoxynucleotidyl transferase, T4 DNA ligase, and various restriction endonucleases, Bethesda Research Laboratory; various other restriction enzymes, New England BioLabs; DNA polymerase I, Boehringer Mannheim; reverse transcriptase, Dr. J.W. Beard, Life Sciences, Inc.; S1 nuclease, Sigma; polynucleotide kinase from both P.L. Biochemical, Inc. and New England Nuclear; bacterial alkaline phosphatase, Worthington. $\alpha^{32}\text{P}$ -labeled nucleotides were obtained from Amersham and New England Nuclear. $\gamma^{32}\text{P}$ -labeled ATP was purchased from New England Nuclear. Nitrocellulose filter paper (BA85) was purchased from Schleicher and Schuell. Medium components for growth of bacteria and phage were obtained from Difco Laboratories. All other basic reagents were purchased from Sigma.

2. Construction of a CHO Recombinant Phage Library

Whole genomic Chinese hamster DNA purified (5) from the dhfr⁺ amplified mutant cell line, MK42 (4), was used for the cloning procedure. The phage vector was λ 1059 (6). Digestion of the phage DNA by BamHI, and partial digestion of the MK42 DNA by MboI was performed as described by Karn et al. (6). To ensure a random population of genomic fragments, the amount of MboI was varied in separate 90 min. digestions from 0.5-0.1 units/ μg DNA. The partially digested chromosomal DNA was size fractionated on a 5-20% sodium chloride gradient (7). Aliquots from the gradient were examined by electrophoresis on a 0.5% agarose (SeaKem) gel. Those fractions containing DNA 10-20 kilo-

bases (kb) in size were pooled. The concentration of the digested λ 1059 and genomic DNAs was determined fluorimetrically (8). Next, 1 μ g of appropriately sized MK42 DNA fragments and 2 μ g of BamHI-cleaved λ 1059 DNA were combined with 1.5 Weiss units of T4 DNA ligase and set at 4°C for 18 hr. The ligated mixture was packaged as described by Enquist and Sternberg (9). The packaging system was pretested and yielded λ 1059 plaques on the permissive bacterial strain at an efficiency of 3.8×10^6 PFU/ μ g DNA. In the nonpermissive bacterial strain which selects for recombinant phage, we obtained about 4×10^4 plaques. Since the dhfr gene is amplified roughly 150-fold in the DNA of MK42, we anticipated that this number of plaques would be adequate to find positive clones.

3. Screening for dhfr⁺ Recombinant Phages

Recombinant plaques were screened by the Benton-Davis method (10). Filter hybridizations were performed according to the procedure of Maniatis *et al.* (11). The mouse cDNA probe used for the hybridizations was nick-translated (12) pDHFR21 (13) plasmid DNA. Phage from positive plaques were purified and rescreened twice. Phage growth and DNA preparations were according to standard methods (9). In this way, recombinant phages λ 1059·dhfr2, 3, 6, 11 and 13 were isolated. They will henceforth be referred to as λ D2, etc.

4. Restriction Digestion, Gel Electrophoresis, and DNA Fragment Purification

The conditions for restriction digestion of DNA samples were all as specified by the product manufacturer, either New England BioLabs or Bethesda Research Laboratory. Double digests with enzymes requiring different buffers were done sequentially, first digesting with the enzyme requiring the lower salt, and then adding components necessary to bring the salt to the required concentration plus the second enzyme and continuing the incubation. Gel electrophoresis was performed in a horizontal gel apparatus in tris (hydroxymethyl) aminomethane-acetate-ethylenediamine tetraacetate-sodium chloride buffer (14). Molecular weight standards for the gels were either ³²P-labeled or unlabeled DNA fragments of HindIII-digested wild-type λ or HaeIII-digested ϕ X174 phage (15). Labeled restriction fragments were used when the gel was to be blotted. Labeling of the marker DNA fragments with polynucleotide kinase and γ -³²P-ATP was performed according to Maxam and Gilbert (16). Purified DNA restriction fragments were prepared by electroelution out of agarose gels attached to GelBond NF film (SeaKem). Vertical 4% polyacrylamide gel electrophoresis of restriction fragments under 1 kb was done by standard techniques (16).

5. Subcloning of Restriction Fragments from Recombinant Phages

BamHI or EcoR1 fragments were subcloned from DHFR recombinant λ 1059 phages

(see Results) into the plasmid vectors pBR322 (17) or pBR325 (18), respectively. Selection of recombinants was by drug resistance phenotype followed by agarose gel electrophoresis to screen lysed bacterial colonies for plasmid size (19).

6. Synthesis and Screening of DHFR cDNA Plasmids

Complementary DNA was prepared from total MK42 poly(A⁺) RNA (20) according to published methods (21 and 22). The dC-tailed cDNA was annealed to PstI-cleaved, dG-tailed pBR322 plasmid DNA. After transformation of a bacterial host, colonies displaying the correct phenotype of resistance to tetracycline were treated with chloramphenicol (23). Screening of these plasmid amplified colonies was performed by the Grunstein-Hogness method (24). The insert of pDHFR21 plasmid, purified from a PstI digest, was used as the probe. Northern blot analysis was performed according to Alwine et al. (25) after glyoxal (26) treatment and electrophoresis in 1.25% agarose-Tris-acetate gels; or by the method of Thomas (27) after electrophoresis in 1% agarose-formaldehyde gels (28).

7. Cloning the 8 kb HindIII Fragment Comprising the dhfr 5' End into pBR322

MK42 DNA (200 µg) was digested to completion by HindIII and centrifuged in a 5-25% sucrose gradient (20) to separate the fragments by size. Fractions of 0.3 ml were collected and 10 µl aliquots of each were examined by 0.5% agarose gel electrophoresis. A Southern blot (2) combining the modifications of Wahl et al. (30) of this gel was performed. Hybridization of the filter-bound DNA was to the pDHFR21 plasmid probe. Those fractions displaying exclusively the 8 kb HindIII fragment were pooled. DNA from pBR322 was digested with HindIII, and treated with 0.1 unit/µg of bacterial alkaline phosphatase. After fluorimetric determination of chromosomal and plasmid DNA concentrations, 0.5 µg of the MK42 DNA was ligated to 0.4 µg of plasmid DNA. Screening of ampicillin-resistant transformants by the Grunstein-Hogness method was done as described above.

RESULTS

1. Construction of a dhfr Restriction Map

A relatively small recombinant DNA library was constructed starting with DNA from clone MK42, a highly methotrexate-resistant CHO derivative in which the dhfr locus has been amplified approximately 150-fold (4). Randomly cleaved genomic DNA fragments were generated by partial MboI digestion and inserted into the bacteriophage lambda vector, λ1059 (6). Viral plaques were screened by hybridization to the mouse DHFR cDNA plasmid pDHFR21 (13). Twenty-four

Figure 1. The *dhfr* locus of MK42 CHO cells.

A restriction map of the gene is diagrammed showing the location of *EcoRI* (E), *BamHI* (B), and *HindIII* (H) cleavage sites. The exons are represented as filled squares. The 3'-most exon has been drawn larger than the others based on analysis of the homologous cDNAs. Repeated sequences are represented as open circles. The locations of all symbols have been drawn to scale. Below the map, the approximate intron sizes are indicated. The map was generated by restriction analysis of several *dhfr* recombinant phages. They are designated λD6, 13, 2, 3, and 11. The size and location of their inserts are drawn below the corresponding genomic regions. The diagonal bars on λD13 indicate that further 3' of this position, the DNA insert is not colinear with the the genome, presumably because of a recombination that took place during cloning. The dotted line extending 5' on λD11 indicates that we have not performed genomic confirmations for the region beyond the terminal *HindIII* site present on the map. Below the phages on the right side of the illustration, are shown two cloned CHO cDNAs, pDCH14 and pDCH18. At the bottom, are presented the genomic inserts of recombinant plasmids. All were subcloned from λD6 or λD13, except for pMH8, whose derivation is described in the text. The size of each insert and its location with respect to the map are indicated. Cloned genomic segments deleted for repeated sequences were used for genomic confirmations (pMB5, pB6-14, pB61H3, and pB67HI). Exon-containing cloned inserts were used for fine mapping studies (pMB5, pE11-15, pB6-14, and pB6-1).

independent recombinant DNA phages were isolated in this manner, and eight of these have been examined extensively. Many of those selected for study spanned overlapping regions of the gene. A restriction map was constructed for each of the recombinant phage by comparing single and double enzyme digestions. The results for five phages are shown in Figure 1, in which *HindIII*, *BamHI*, and *EcoRI* cutting sites are indicated. The approximate positions of exons within the emerging map were determined by Southern blot hybridizations of the restricted phage DNA using the mouse cDNA probe, pDHFR21. A representa-

Figure 2. BamHI digested λD6 DNA.

Lane A shows the results of agarose gel electrophoresis of BamHI digested λD6 DNA. The sizes of the ethidium bromide stained fragments are: >20 kb (the left phage arm), 10.6 kb (the right phage arm), 6.8 kb, 5.3 kb, 4.6 kb and 0.98 kb. The autoradiograph of the Southern blot appears in lane B. The fragments that hybridize with the mouse cDNA probe, pDHFR21 are: >20 kb, 6.8 kb, 5.3 kb and 4.6 kb. The numbers on the right refer to the sizes in kb of the molecular weight markers used: HindIII digested wild-type λ DNA.

tive analysis for recombinant phage λD6 cut with BamHI is shown in Figure 2.

One phage, λD11, did not overlap with any of the other phages defining the composite phage map. An exon-containing 2.7 kb EcoRI fragment was subcloned from this phage into pBR325. This plasmid (pE11-15), when used as a probe for a Southern blot of MK42 genomic DNA, hybridized to an 8 kb HindIII fragment. With the idea that this HindIII fragment would establish continuity in the map, we proceeded to clone it directly from a HindIII digest of MK42 DNA. In this case, size-fractionated DNA was cloned into pBR322 rather than a bacteriophage, and colonies were screened using pDHFR21 insert as a probe. A positive clone, designated pMH8, which carries an 8.1 kb insert contains

Figure 3. Overlap of pMH8 with the composite phage map.

DNA from pMH8 (lane A) and pB6-14 (lane B) was digested with BamHI and HindIII and electrophoresed in a 0.8% agarose gel. A Southern blot of this gel was then hybridized to a fragment representing the 5' end of the pB6-14 insert (2.7 kb EcoRI-BamHI fragment). The 0.45 kb BamHI-HindIII band indicated is common to both recombinant plasmids.

Figure 4. Hybridization of a 5'-specific mouse cDNA probe to CHO dhfr genomic sequences.

The probe for the Southern blot hybridization was the nick-translated 0.46 kb BglII-SacI fragment of mouse pDHFR11 (28). Lane A shows BglII digested total genomic DNA from the CHO clone, MK42. Lanes B, C, and D show λ D11 DNA digested with BglII, EcoRI and HincII, respectively.

the same EcoRI fragments that are present on λ D11 (see Fig. 1). Southern blot analysis showed that pMH8 overlaps with pB6-14 (see Fig. 1), a plasmid subclone of λ D6. This overlap was localized to a 0.45 kb HindIII-BamHI fragment common to both plasmids, as shown in Figure 3.

The final composite map is derived from the restriction maps of pMH8 and phages λ D6 and λ D13. It contains all the exons defined by mouse and Chinese hamster cloned cDNAs (see below) and extends approximately 26 kb.

2. Orientation of the CHO dhfr Map

Mouse cDNA probes specific for the 5' and for the 3' ends of the mRNA were used to determine the 5' to 3' polarity of the restriction map. The mouse cDNA clone pDHFR11 has previously been described (31) and represents a complete copy of the predominant large mouse DHFR messenger RNA except for less than 50 nucleotides at the 5' end. A 0.46 kb BglII-SacI fragment of pDHFR11 was electrophoretically purified and nick-translated. This probe is specific for those sequences that cross-hybridize with the 5' mouse exons I, II and III (32). Figure 4 shows the results of hybridization of this pDHFR11 5' probe to BglII-cleaved MK42 genomic DNA and λ D11 phage DNA. All exons detected by this probe are found on a 3.4 kb BglII fragment. This fragment is located near the left end of the map shown in Fig. 1; its position is indicated by the subclone pMB5 in that figure. In contrast to these results with the 5' mouse probe, the pDHFR21 cDNA probe, which lacks the 5' mouse mRNA sequences (13), hybridizes to 3' exons (Fig. 2) but not to the 5'-most 1.3 kb EcoRI fragment (data not shown). Thus, the map shown in Fig. 1 is oriented with the 5' end of the transcribed region at the left.

3. Identification of the 3'-most Exon using Chinese Hamster cDNA Probes

In order to accurately define exon positions in the Chinese hamster *dhfr* gene, homologous Chinese hamster cDNA probes are required. Therefore, simultaneous with our efforts to isolate recombinant genomic sequences, we also cloned double-stranded cDNA using poly(A⁺) RNA from the amplified clone MK42 as a template. In the MK42 cell line, three over-produced molecular weight species of DHFR mRNA can be detected. The two predominant species are approximately 2400 and 1100 base pairs in length, as indicated in Figure 5. Multiple species of DHFR mRNAs have previously been demonstrated in amplified cell lines derived from mouse (33, 34 and 35) and Chinese hamster lung fibroblasts (36). We can also detect the same two predominant species in unamplified CHO cell lines (P. Mitchell, unpublished results).

We first isolated cDNA clones by standard procedures (21, 22) using oligo-dT as a primer and dG-dC tailing for insertion into pBR322. The two longest clones from this experiment, pDCH18 (1200 bp) and pDCH14 (1400 bp), were found to carry overlapping sequences that hybridize exclusively to the 3'-most exon shown in Fig. 1. The combined insert length of these two cDNA clones totals 1900 bp. Allowing as much as 100 bp for G-C tails and 100 bp for poly(A) sequences that may have been copied, this leaves 1700 bp as the approximate minimum size of the 3'-most exon. Thus the 3'-most exon is by far the largest exon, and most of it could only be present in the largest mRNA species. As can be seen in Fig. 5, whereas pDCH14 hybridizes to both species of mRNA, pDCH18 hybridizes only to the large mRNA. We conclude from these data that the smaller DHFR mRNA species lacks 3' sequences present in the larger mRNA. The size of the smaller species could be accounted for by a polyadenylation site located near the BamHI site in the 3'-most exon (Fig. 1). Multiple colinear DHFR mRNAs in the mouse have been shown to involve alternative polyadenylation sites (33).

The large abundant DHFR mRNA species in the mouse is only 1600 bp (33 and 34); hence, it must lack some sequences present in Chinese hamster DHFR mRNA. In fact, the mouse cDNA probe pDHFR21, which extends to the 3' end of the large mouse DHFR mRNA (13), fails to hybridize to the 0.7 kb EcoRI fragment of λ D13. In contrast, the Chinese hamster cDNA probes do hybridize to this fragment (data not shown) which includes the 3' end of the 3'-most exon (see Fig. 1). Thus, most if not all of the sequences that are specific to the Chinese hamster are located at the extreme 3' end of this exon. This specificity could be due to a divergence in sequence homology in this 3' untranslated region and/or to a relative lack of exonic sequence in the mouse.

Figure 5. Northern blot analysis of MK42 mRNA.

Each channel of the gel contained 1 μ g of MK42 poly(A⁺) RNA. The probes were nick-translated mouse pDHFR21 (lane A), CHO pDCH14 (lane B), and CHO pDCH18 (lane C) DNAs. The numbers on the left refer to size estimates of the major DHFR mRNA species in kilobases. Size standard were HindIII fragments of λ DNA plus CHO ribosomal RNA.

4. Location of Exons Within Subcloned Fragments

To map more accurately the location of the coding sequences, we subcloned several appropriate exon-containing fragments into pBR322 and pBR325. Restriction mapping was done by single and double digestion of purified inserts or whole plasmid DNAs using a variety of endonucleases. The products of these digestions were subjected to electrophoresis on polyacrylamide gels in order to better resolve their sizes. Southern blots of selected digests were probed for exon-containing fragments using either the 1.5 kb PstI insert of pDHFR21 or the 0.46 kb BglI-SacI fragment of pDHFR11. The results of this work identified the location of exons I, III, IV, V and VI as shown in Figure 6. The location of exon II was identified by the ability of the 0.26 kb HincII fragment of pMB5 DNA to hybridize to DHFR mRNA in Northern blots (data not shown). This location for exon II predicts that a 0.25 kb EcoR1 and a 0.26 kb HincII fragment should hybridize to a mouse 5' cDNA probe. Our failure to detect these fragments in Southern blots (Fig. 4, lanes C and D) was probably due to the very small size of this exon (50 bp in the mouse, ref. 32) and the stringency of the hybridization conditions.)

5. Location of Repeated Sequences Within the dhfr Gene

We expected that a gene of this size would probably contain a number of dispersed highly repeated sequences. Such sequences are prevalent in eukaryotic genomes (37). It was necessary to determine the location of these sequences in order to construct genomic probes that would be specific for the dhfr locus. We mapped the repeated sequences by hybridizing nick-translated total genomic wild-type CHO (K1) DNA to Southern blots of restricted recombinant phage

Figure 6. Fine structure map of dhfr exon-containing fragments.

The map shows relative locations and sizes of exons I-VI. The exons have been established only by Southern blot analysis. Thus, the ends are jagged denoting that the precise limits are not yet known. Names of the restriction enzymes have been abbreviated. They are: E, EcoRI; Bg, BglII; R, RsaI; Sm, SmaI; Ss, SstII; EII, EcoRII; HII, HincII; Sp, SphI; Bs, BstEII; A, AvaI; Hp, HphI; Sa, SacI; Pv, PvuII; Hf, HinfI; HIII, HindIII; M, MspI; C, CfoI; Ps, PstI; and K, KpnI.

Figure 7. Representative Southern blot identifying the locations of repeated sequences.

On the left is the ethidium bromide stained agarose gel showing pMH8 DNA digested with EcoRI (A), EcoRI-BglII (B), BglII (C), BamHI-BglII (D), and BamHI-EcoRI (E). In lane M was run the molecular weight standards (λ HindIII and ϕ X174 HaeIII). On the right is the autoradiograph of the blotted gel after hybridization to nick-translated total genomic K1 DNA. The specific activity of the probe was 4×10^8 cpm/ μ g. For the hybridization, 10^7 cpm/ml was used.

Figure 8. Genomic confirmation of the map derived from recombinant DNA.

Total DNA from the amplified line MK42 was digested with various restriction enzymes and subjected to 0.8% agarose gel electrophoresis. Blot hybridization was then carried out using as probes various plasmids containing subcloned fragments from different regions of the cloned *dhfr* gene. The numbers in the map at the bottom indicate the location of each probe's insert in the *dhfr* gene. The lower case letters indicate the location of restriction fragments that hybridize to a given probe. Lane A: *Hind*III, probe 1; B: *Bam*HI, probe 2; C: *Eco*RI, probe 2; D: *Bam*HI, probe 3; E: *Hind*III, probe 4; F: *Eco*RI, probe 5; G: *Bam*HI, probe 5; H: *Eco*RI, probe 6; M: size standards, 3' end-labeled (16) *Hind*III digested λ DNA plus *Hae*III digested ϕ X174 RF DNA. Probes (see Fig. 1): 1, pMB5; 2, pB6-14; 3, pB61H3; 4, pB67H1; 5, pB13-7; 6, pB13-6.

or plasmid DNAs. Only highly repeated sequences are in abundance sufficient to produce a signal under the hybridization conditions (38). The locations of these sequences are indicated by open circles on the map of Figure 1. Figure 7 shows the results of a representative blot in which the K1 probe was hybridized to several different digests of pMH8 DNA. *Eco*RI cleavage of the plasmid DNA separated two repeated sequences (Fig. 7, lane A). Double digestion by *Eco*RI and other enzymes (*Bam*HI and *Bgl*II are shown) further defined the location of the repeats. Similar analyses led us to position additional repeated sequences as shown in Figure 1. One probe containing repeated sequence(s), the 1.2 kb *Hind*III

fragment of pB6-1, hybridizes to the other three fragments that contain repeated sequences, indicating that at least two of these sequences are related.

6. Genomic Confirmation of the dhfr Map

Southern blotting analysis of Chinese hamster genomic DNA was performed in order to confirm the validity and linearity of the map presented in Figure 1. The probes used for this work are diagrammed at the bottom of this figure and are: pMB5 (3.4 kb BglII), pB6-14 (5.3 kb BamHI), pB61H3 (0.9 kb HindIII), pB67HI (3.2 kb HindIII-BamHI), pB13-7 (2.1 kb BamHI), and pB13-6 (0.7 BamHI). On each gel, EcoR1, BamHI, or HindIII digested samples of MK42 genomic DNA were electrophoresed and blotted. Figure 8 shows the resulting autoradiographs. In every case tested, a given probe revealed a fragment of the size predicted by the map in Fig. 1. The fragments displayed in Figure 7 represent sequences that together include the entire 28 kb of DNA that has been mapped. Since each sequence tested overlaps with the next, the possibility of a deletion in the recombinant phages is eliminated.

DISCUSSION

The 28 kb of cloned and mapped DNA shown in Fig. 1 contains most, if not all, of the information specifying DHFR mRNA in CHO cells. The six independently isolated recombinant DNA molecules (five phage and one plasmid) represent overlapping sequences for 27 out of the 28 kb mapped. Since the restriction maps of these overlapping sequences are the same, it is unlikely that any part of the map in these regions could have arisen by an artifactual recombination event during cloning. The 1 kb region that is present uniquely in phage λ D6 (in the 9.4 kb intron) is flanked on both sides by exons within the inserted fragment. This finding argues against artifactual recombination here, as well. The most definitive experiment identifying this map as that of the dhfr locus is provided by the genomic confirmation results shown in Fig. 8. In this experiment, a series of overlapping fragments spanning the entire 28 kb region was shown to be present in the total DNA of clone MK42, the amplified mutant which served as the source of the recombinant DNA library. The analysis of overlapping fragments in this experiment precludes the presence of any additional sequences in this gene at any joint between tested fragments. It is probable that all exons are present within the 28 kb of mapped DNA, since preliminary DNA sequencing data places the AUG translation start codon within exon 1. Experiments aimed at identifying the transcriptional start site(s) are in progress.

The dhfr gene in CHO cells is relatively large, having at least five in-

tervening sequences separating six exons that code for the 2400 base pairs of the largest DHFR mRNA species. The length of the dhfr gene from the beginning of exon I to the end of exon IV is 26 kb, a length similar to that recently reported for the mouse dhfr gene (31 kb, ref. 32). In contrast, the dhfr locus in the Syrian hamster (BHK cells) has been reported to span only 10.7 kb (30). This last figure was derived from Southern blot analysis of total genomic DNA using a mouse cDNA probe, and it is possible that some exons were missed due to lesser sequence homology, particularly in untranslated regions.

Mapping of genomic sequences in wild-type CHO cells indicates the same overall gene structure (3 and Urlaub *et al.*, in preparation). Thus, clone MK42 apparently contains amplified copies of the wild-type dhfr gene. This conclusion is not surprising since the size of the repeated unit containing the dhfr gene has been estimated to be several hundred kb (4). This size is much larger than the dhfr structural gene itself. The map of the CHO gene we have cloned is also consistent with restriction fragments derived from one of two polymorphic dhfr sequences present in Chinese hamster lung cell DNA (40).

Four regions homologous to highly repeated CHO DNA sequences were found within the dhfr gene. While two of these are located just 5' of exons V and VI, it is not yet clear whether this placement is of any significance.

The location of exons within the CHO dhfr gene is very similar to that in the mouse. Our estimates of the sizes of the five introns in the CHO gene are (from 5' to 3') 0.3, 2.5, 8.6, 2.6, and 9.4 kb; whereas the corresponding lengths in the mouse are 0.3, 2.4, 7.6, 2.3, and 16.5 kb (31). The overall organization of this locus has been conserved evolutionarily in these two rodent species. It is likely that the exact placement of introns within the gene will be conserved, as well, thus extending this generalization from specialized genes to household genes such as dhfr.

The CHO and mouse dhfr genes do differ at their 3' ends. The mouse cDNA probe pDHFR21 extends fully to the 3' end of the large mouse mRNA, yet fails to hybridize with the terminal fragment (EcoR1) of exon VI. The Chinese hamster genome should contain additional exonic sequences relative to the mouse since the largest abundant DHFR mRNA species is longer than that of the mouse. In the mouse, exon VI is 950 bp (33). In CHO, the size of exon VI can be estimated from the lengths of the cDNA clones pDCH14 and pDCH18 (Fig. 1). These overlapping clones comprise 1.9 kb of insert sequence yet hybridize exclusively to genomic sequences in the exon VI region. Subtraction of a generous 200

bp to account for poly(C) tails added during cloning and for a poly(T) tract copied from poly(A) still leaves 1.7 kb as a minimum estimate for exon VI. This additional 750 bp as compared to the mouse is consistent with the figure (800 bp) that would be predicted on the basis of the size difference between the largest abundant mouse DHFR mRNA species (1600 nucleotides, ref. 32) and our measurement of 2400 nucleotides for the largest abundant CHO DHFR mRNA. Notwithstanding its larger size, CHO exon VI is uninterrupted by introns, as is the case for the 3' untranslated regions of most eukaryotic genes analyzed.(41).

Multiple mRNA species coding for the same household enzyme have now been observed not only for mouse (32, 33) and Chinese hamster DHFR (Fig. 5 and ref. 40, 42), but also for hypoxanthine phosphoribosyltransferase in Chinese hamster cells (43) and alcohol dehydrogenase in yeast (44). The physiological significance of these multiple mRNAs is not clear. It has been suggested that multiple polyadenylation sites may provide alternative targets for differential regulation of dhfr gene expression (45). In the case of DHFR another possibility is raised by the recent findings of Reddy and Pardee. These workers have reported that while approximately half of the cellular DHFR activity is apparently free in the cytoplasm, the remainder is associated in a multi-enzyme aggregate (termed "replitase") involved in nucleotide metabolism and DNA synthesis (46, 47). It is tempting to speculate that different DHFR mRNAs may code for enzyme molecules that are destined for different compartments in the cell. Many of these possibilities may be tested by the analysis of dhfr mutations produced in cultured cells (1) as well as in cloned dhfr DNA (48, 49).

ACKNOWLEDGEMENTS

We thank Robert Schimke for providing pDHFR21 and Christian Simonsen for providing pDHFR11 and unpublished mouse DNA sequence data. We are grateful for the valuable experimental assistance of Emmanuel Kas, Ted Widlanski, and Peter Karczmar. We thank Sally Lewing for preparation of this manuscript. This work was supported by NIH Grant No. GM34537 to L.A.C. and by a Damon Runyon - Walter Winchell Fellowship (DRG-446-F) to A.M.C.

* To whom to address correspondence.

REFERENCES

1. Urlaub, G., and Chasin, L.A. (1980) Proc. Natl. Acad. Sci. USA 77, 4216-4220.

2. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
3. Graf, L.H. Jr., and Chasin, L.A. (1982) *Mol. Cell Biol.* 2, 93-96.
4. Nunberg, J.H., Kaufman, R.J., Schimke, R.T., Urlaub, G., and Chasin, L.A., (1978) *Proc. Natl. Acad. Sci. USA* 75, 5333-5556.
5. Pellicer, A., Wigler, M., Axel, R., and Silverstein, S. (1978) *Cell* 14, 133-141.
6. Karn, J., Brenner, S., Barnett, L., and Cesareni, G. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5172-5176.
7. Liu, C.-P., Tucker, P.W., Mushinski, J.F., and Blattner, F.R. (1980) *Science* 209, 1348-1353.
8. Kapuscinski, J. and Skoczylao, B. (1977) *Anal. Biochem.* 83, 252-257.
9. Enquist, L., and Sternberg, N. (1979) *Methods in Enzymology* (Academic Press, Inc., New York) 68, 281-298.
10. Benton, W.D., and Davis, R.W. (1977) *Science* 196, 180-182.
11. Maniatis, T., Hardison, R.C., Lacy, E., Laver, J., O'Connell, C., and Quon, D. (1978) *Cell* 15, 687-701.
12. Rigby, P.W.J., Dieckman, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
13. Chang, A.C.Y., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schimke, R.T., and Cohen, S.N. (1978) *Nature* 175, 617-624.
14. Dugaiczky, A., Boyer, H.W., and Goodman, H.M. (1975) *J. Mol. Biol.* 96, 171-184.
15. Sanger, F., Air, G.M., Barnell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977) *Nature* 265, 687-697.
16. Maxam, A.M., and Gilbert, W. (1980) *Methods in Enzymology* (Academic Press, Inc., New York) 65, 499-563.
17. Bolivar, F., Rodriguez, R.L., Greene, P.J., Bettach, M.C., Heyneker, H.L., and Boyer, H.W. (1977) *Gene* 2, 95-113.
18. Bolivar, F. (1978) *Gene* 4, 121-136.
19. Eckhardt, T. (1978) *Plasmid* 1, 584-588.
20. Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408-1412.
21. Buell, G.N., Wickens, M.P., Payvar, F., and Schimke, R.T. (1978) *J. Biol. Chem.* 253, 2471-2482.
22. Wickens, M.P., Buell, G.N., and Schimke, R.T. (1978) *J. Biol. Chem.* 253, 2483-2495.
23. Hanahan, D., and Meselson, M. (1980) *Gene* 10, 63-67.
24. Grunstein, M., and Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961-3965.
25. Alwine, J.C., Kemp, D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R., and Wahl, G.M. (1979) *Methods in Enzymology* (Academic Press, Inc., New York) 68, 220-242.
26. McMaster, G.K., and Carmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4835-4839.
27. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
28. Rave, N., Ckvenjakov, R., and Boedtker (1979) *Nucleic Acids Res.* 6, 3559-5567.
29. Grosveld, F.G., Dahl, H.M., deBoer, E., and Flavell, R.A. (1981) *Gene* 13, 227-237.
30. Wahl, G.M., Stern, M., and Stark, G.R. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3683-3688.
31. Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohen, S.N., and Schimke, R.T. (1980) *Cell* 19, 355-364.
32. Crouse, G.F., Simonsen, C.C., McEwan, R.N., and Schimke, R.T. (1982) *J. Biol. Chem.* 257, 7887-7897.
33. Setzer, D.R., McGrogan, M., and Schimke, R.T. (1982) *J. Biol. Chem.* 257, 5143-5147.

34. Setzer, D.R., McGrogan, M., Nunberg, J.H., and Schimke, R.T. (1982) *Cell* 22, 361-370.
35. Dolnick, B.J., Berenson, R.J., Bertino, J.R., Kaufman, R.J., Nunberg, J.H., and Schimke, R.T. (1979) *J. Cell. Biol.* 83, 394-402.
36. Lewis, J.A., Kurtz, D.T., and Melera, P.W. (1981) *Nucleic Acids Res.* 9, 1311-1322.
37. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L., and Schmid, C.W. (1980) *Proc. Natl. Acad. Sci. USA* 77, 1398-1402.
38. Baralle, F.E., Shoulders, C.C., Goodbourn, S., Jeffreys, A., and Proudfoot, N.J. (1980) *Nucleic Acids Res.* 8, 4393-4404.
39. Schoner, R.G., and Littlefield, J.W. (1981) *Nucleic Acids Res.* 9, 6601-6614.
40. Lewis, J.A., Davide, J.P., and Melera, P.W. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6961-6965.
41. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
42. Melera, P.W., Hession, C.A., Davide, J.P., Scotto, K.W., Biedler, J.L., Meyers, M.B., and Shanske, S. (1982) *J. Biol. Chem.* 257, 12939-12949.
43. Konecki, D.S., Brennaud, J., Fuscoe, J.C., Caskey, C.T., and Chinault, A.C. (1982) *Nucl. Acids Res.* 10, 6763-6775.
44. Bennetzen, J.L., and Hall, B.D. (1981) *J. Biol. Chem.* 257, 3018-3025.
45. Kaufman, R.J., and Sharp, P.A. (1982) *Molec. Cell Bio.* 2, 1304-1319.
46. Reddy, G.P.V., and Pardee, A.B. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3312-3316.
47. Reddy, G.P.V., and Pardee, A.B. (1982) *J. Biol. Chem.* 257, 12436-12531.
48. Gasser, C.S., Simonsen, C.C., Schilling, J.W., and Schimke, R.T. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6522-6526.
49. Crouse, G.F., McEwan, R.N., and Pearson, M.L. (1982) *Mol. Cell Biol.*, in press.