# The Organization and Expression of the Human Dihydrofolate Reductase Gene in Methotrexate-Resistant and Methotrexate-Sensitive Cell Lines

Thesis by

Jeffrey Nelson Masters

In Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

California Institute of Technology
Pasadena, California

1985
(submitted July 25, 1984)

This thesis is dedicated to my wife Amy,

whose confidence and love have enabled me to

to get through the hard times

## Acknowledgements

I have made many friends at Caltech and express my gratitude to all of them for helping to round out my graduate career. In particular the members of the Attardi lab have been very supportive and served as great sounding boards for bouncing off new ideas. The following people deserve a special acknowledgement.

Giuseppe, a friend whose commitment to scientific excellence is unsurpassed, taught me how to approach scientific problems in a meaningful manner.

George was a particular friend who enjoyed talking about far - out ideas and problems of monumental proportions, especially while at the B.C.

Jane, after adjusting to a change in gravity, was instrumental in getting my introduction to this thesis coherent.

Carlo helped me to gain confidence in the lab when I first arrived at Caltech and our collaboration was very enjoyable.

Doris was fantastic in making sure that the lab ran smoothly and she magically managed to get those last minute orders delivered very quickly.

The rest of the support staff, Arger, Benetta, Maria and Rosie deserve a lot of credit for what they do. Benetta deserves a special thank you because of her superb handling of the many different cell lines we used.

I also want to thank all of the office people who managed to make sense out of my chicken scratches and make sure that the money flowed smoothly.

And finally, I would like to thank the creative folks of Venice Beach for providing me with a source of provocative tee-shirts.

As a graduate student, I received support from the Earle C. Anthony Graduate Fellowship and a National Institutes of Health Predoctoral Fellowship. Support from the the Jean Wiegle Memorial Fund was used in the preparation of this thesis.

## Abstract

The analysis of several derivatives of the human cell lines HeLa and $VA_2$-B, selected for resistance to methotrexate (MTX), shows a striking variability in the dihydrofolate reductase (DHFR) enzyme levels, chromosome constitution and growth characteristics in the absence of MTX. In contrast to the mouse system, the number of double minutes of the human cells does not correlate with either the increased DHFR levels or the instability of the amplified phenotype.

The isolation of human DHFR cDNA by differential hybridization or by phenotypic expression in *E. coli*, facilitates the characterization of the human DHFR coding region and its multiple mRNAs. Comparison of the DNA sequences of several DHFR cDNAs shows a high degree of homology between the coding regions of the human and mouse genes (89%) with no obvious identity in the 3'-untranslated region. The analysis also demonstrates that cDNAs of the 3 identified mRNAs are colinear in the 3'-end sequence, and that polyadenylation occurs at different sites.

Hybridization of *Eco*RI digested nuclear DNA from the MTX-resistant 6A3 cells with DHFR cDNAs shows *Eco*RI fragments that are either unamplified or amplified relative to the same fragments of human sperm, HeLa cell, and $VA_2$-B cell DNA. Two of the unamplified fragments were isolated from a cosmid library made from human sperm DNA. The DNA sequence analysis shows that these two fragments contain a DHFR intronless pseudogene including the *Eco*RI site found in the DHFR cDNAs. If an RNA intermediate directs the formation of this pseudogene, an RNA larger than the 1.0 kb DHFR mRNA must be involved. In contrast to the unamplified DNA fragments, the amplified fragments contain the exons of the human DHFR gene. The gene is about 29 kb in length, with five introns interrupting the DHFR coding region in the same positions found in the mouse gene. The DHFR mRNAs were mapped as a major 5'-end at position -71 of the human DHFR gene. In addition, six minor 5'-ends of DHFR-specific polysomal RNA were mapped from positions -449 to -480 and represent about 1% of the major transcripts. The upstream transcripts are relatively enriched in the nuclear RNA fraction, indicating a different regulation of expression for these minor transcripts.

## Table of Contents

# CHAPTER 1

## Introduction to Drug Resistance, Gene Amplification and Dihydrofolate Reductase Gene Expression

Dihydrofolic acid reductase (DHFR, EC 1.5.1.3) is the enzyme responsible for reducing folic acid to tetrahydrofolic acid in an NADPH dependent reaction. Tetrahydrofolic acid then undergoes various modifications to act as a single-carbon donor in the cell and is particularly important in the *de novo* synthesis of purines, thymidylate and some amino acids. The enzyme can be inhibited by 4-amino analogs of folic acid, such as amethopterine (methotrexate, MTX) or aminopterine, due to their high affinity binding to the enzyme. This property makes the analogs suitable as chemotherapeutic drugs in the treatment of malignant tumors (Jolivet *et al.*, 1983). However, continued treatment with MTX can result in the tumors becoming resistant to the drug (Bertino *et al.*, 1963). These observations of DHFR led to many studies of drug resistance in mammalian cells. Below is a brief summary of the available data about drug resistance, gene amplification and DHFR gene expression, followed by a summary of the results of this thesis for the human DHFR gene.

## Drug Resistance

Cells cultured *in vitro* can become resistant to folate analogs. In particular, several mammalian cell lines have been selected for resistance to MTX (see Schimke, 1982). Three mechanisms have been identified for this resistance: (i) reduced permeability of the drug (Fisher, 1962; Sirotnak *et al.*, 1968; Flintoff *et al.*, 1976); (ii) an alteration in the DHFR enzyme resulting in a decreased affinity for MTX (Abrect *et al.*, 1972; Flintoff *et al.*, 1976; Flintoff and Essani, 1980; Haber *et al.*, 1981); and (iii) increased DHFR enzyme levels (Fisher, 1961; Hakala *et al.*, 1961; Biedler *et al.*, 1972; Alt *et al.*, 1976; Flintoff *et al.*, 1976; Melera *et al.*, 1980). Little is known about the first mechanism for MTX resistance, other than that it may represent a locus concerned with the active transport of MTX and that it is inducible by mutagens (Flintoff *et al.*, 1976). In regard to the second mechanism, a recent report characterizes an altered DHFR enzyme with a MTX-binding affinity of 1/270th of the wild type enzyme (Haber *et al.*,

1981). A DNA sequence analysis of a DHFR cDNA for the altered enzyme shows that it differs from the wild type enzyme at a single nucleotide of the coding region; a Leu codon (amino acid number 22) in the wild type enzyme is changed to an Arg codon in the mutant enzyme (Simonsen and Levinson, 1983). In this case, a hydrophobic residue is replaced by a basic residue in a position implicated as part of the MTX binding site (Matthews *et al.*, 1978; Volz *et al.*, 1982). The third mechanism, increased DHFR enzyme levels, has been shown to result from an amplification of the DHFR gene (Alt *et al.*, 1978; Nunberg *et al.*, 1978; Dolnick *et al.*, 1979; Melera *et al.*, 1980).

## Gene Amplification

Gene amplification *in vivo* has been shown to be developmentally regulated for *Xenopus* rDNA (Brown and Dawid, 1968) and *Drosophila* chorion genes (Spradling, 1981). A recent report conclusively demonstrates that DHFR gene amplification is responsible for a MTX-resistant tumor in a patient undergoing chemotherapy (Curt *et al.*, 1983). For mammalian cells grown *in vitro*, gene amplification appears to be a general mechanism of resistance to toxic substances. It has been documented for various genes including metallothionein (Beach and Palmiter, 1981), the multifunctional CAD protein (an abbreviation for an enzyme containing the enzymatic activities of carbamylphosphate synthetase, aspartate transcarbamylase and dihydrooratase; Wahl *et al.*, 1979), asparagine synthetase (Andrulis and Siminovitch, 1982), thymidylate synthetase (Rossana *et al.*, 1982), 3-hydroxy-3-methylglutaryl coenzyme A reductase (Luskey *et al.*, 1983) and UMP synthase (Kanalas and Suttle, 1984).

Mutations resulting in lower amounts of gene products or in products of reduced specific activities can be corrected by amplifying the mutant genes; this is a form of phenotypic reversion. Phenotypic reversion of the *Drosophila* "bobbed" mutant occurs by gene amplification of rRNA genes (see Tartof, 1975). Gene amplification-directed phenotypic reversion of altered gene products in mammalian cell cultures has been shown

for various genes. These genes include DHFR (Haber *et al.*, 1981), HGPRT (hypoxanthine-guanine phosphoribosyltransferase; Brennand *et al.*, 1982; Fuscoe *et al.*, 1983) and a defective herpes simplex virus thymidine kinase (TK) gene transfected into a TK-deficient mouse cell line (Roberts and Axel, 1982).

A particularly interesting example of gene amplification comes from analysis of the phenomenon of multiple drug resistance in mammalian cells (Ling, 1982), where resistance to a particular drug confers resistance to other drugs of different chemical composition. This resistance is due to a change in the cell membrane which decreases the permeability to the drug. This membrane change is associated with an increase in the amount of a plasma membrane glycoprotein, which has recently been shown to be associated with a gene amplification event (Roninson *et al.*, 1984). Therefore, drug resistance due to membrane permeability changes (the first mechanism discussed above), could involve gene amplification.

A variety of methods have been used to measure the rates of mutation generating drug-resistant phenotypes. Luria-Delbrück fluctuation analysis (Luria and Delbrück, 1943) of resistance to MTX in mouse cell cultures (Terzi and Hawkins, 1974) gave a mutation rate of $4 \times 10^{-5}$ mutation events/cell division. This rate is similar to those estimated for N-(phosphonoacetyl)-L-aspartate resistance (PALA resistance) in CAD gene amplification in hamster cultured cells ($2\text{-}5 \times 10^{-5}$; Kempe *et al.*, 1976) and phenotypic reversion of a TK-deficient, transformed mouse cell line ($2 \times 10^{-5}$; Roberts and Axel, 1982). In the case of the latter two rates, the majority of events were gene amplification, indicating that these estimates approximate the rates of gene amplification events/cell division. However, a determination of the DHFR gene amplification rate utilizing a fluorescence-activated cell sorter for selection (Johnston *et al.*, 1983), gave a rate of $1 \times 10^{-3}$ amplification events/cell division. This higher rate can be explained by the cell sorter selection technique used, which presumably places less hardship on the cells than selection with MTX. If gene amplification is a random

event such that any gene would have a similar amplification rate, probably more than one gene could be amplified in a given event, depending on the size of the amplified unit. The amplification unit of PALA-resistant hamster cells is estimated to be 500-1000 kb and other genes besides the CAD gene have also been localized to this amplification unit (Brinson *et al.*, 1982; Padgett *et al.*, 1982; Wahl *et al.*, 1984). Coamplification of other genes has not been described in the DHFR system. The DHFR amplification unit has been reported to be very long in mouse cells, in the order of 500-1000 kb (Nunberg *et al.*, 1978; Kaufman *et al.*, 1979) or even as high as 3000 kb (Bostock and Clark, 1980). The size of this unit is much greater than the 30 kb DHFR gene (Crouse *et al.*, 1982), leaving open the question of whether other genes could be coamplified with DHFR.

Amplified genes have either been found to be organized as extrachromosomal elements termed "double minutes" or the genes have been chromosomally localized as "homogeneously staining regions" (HSRs) in mammalian cells (reviewed in Cowell, 1982). Double minutes are small, acentromeric elements rarely found in parental cell lines. On the other hand, HSRs represent expanded regions of chromosomes which were found to stain homogeneously in G-banding preparations of chromosomes from amplified cells. The term HSR is also applied for expanded nonstaining regions or expanded banding regions of chromosomes.

Amplified DHFR genes were first localized to double minutes (Kaufman *et al.*, 1979) and HSRs (Nunberg *et al.*, 1978; Dolnick *et al.*, 1979) by *in situ* hybridization with a mouse DHFR cDNA. A general model concerning the stability of the MTX-resistant phenotype in the absence of selection is that the resistance is stable if the amplified genes are localized to HSRs and unstable if the genes are localized on double minutes (Schimke *et al.*, 1980). Since double minutes lack centromeres, they segregate randomly in cell division. This random segregation, combined with the selective advantage of cells with lower numbers of double minutes having shorter cell-cycle times, results in the loss of double minutes from the cell population. Instability of the resistant phenotype has

been shown to correlate with a decrease in size of an HSR (Biedler *et al.*, 1980), possibly the result of unequal crossing over events between sister chromatids. These results, together with the results presented in Chapter 2, indicate that stable and unstable are relative terms rather than absolute terms.

Chinese hamster ovary cells resistant to MTX tend to develop HSRs relatively quickly when grown in the presence of the drug (50-100 generations; Kaufman and Schimke, 1981). This is in contrast to MTX-resistant cells of mouse origin which rarely develop HSRs, even when grown for 2-3 years in MTX (Kaufman *et al.*, 1981). Double minutes on the other hand, are very prevalent in most of the individual cells from an MTX-resistant mouse cell population but found in less than 5% of the cells from initially unstable MTX-resistant hamster cells (those cells grown for less than 50 generations in MTX). The explanation for this phenomenon is not clear, but it may lie in the fact that mouse cells are typically polyploid in their chromosome constitution, whereas Chinese hamster cells are generally pseudodiploid.

There are presently two models for the amplification process involved in gene amplification (discussed in Cowell, 1982, and Schimke, 1982). These models are unequal crossing over between sister chromatids, as suggested in the gene amplification model for carcinogenesis (Pall, 1981) and "replicon misfiring" resulting in disproportionate DNA replication of a subset of DNA sequences (Varshavsky, 1981a). The latter model is similar to the models proposed for *Drosophila* chorion gene amplification (Spradling, 1981) and for the production of free circular simian virus 40 (SV40) DNA when nonpermissive, SV40-transformed mouse cells are fused with a permissive monkey cell line (Botchan *et al.*, 1978). Recent evidence in support of the replication model comes from studies using DNA synthesis inhibitors or tumor promoters, in combination with MTX selection to increase the frequency of amplification (Varshavsky, 1981b; Tlsty *et al.*, 1982; Brown *et al.*, 1983). Furthermore, Mariani and Schimke (1984) used hydroxyurea, which indirectly inhibits DNA replication by inhibiting the enzyme

ribonucleotide reductase, to amplify the DHFR gene in a single cell cycle. They showed that this amplification event takes place in early S phase of the cell cycle and not at the other phases of the cell cycle. Hence, DNA synthesis inhibitors appear to cause the replication machinery to stall on a given replicon, allowing reinitiation of replication once the inhibition is removed. The cell cycle experiments further suggest that the replication machinery can reinitiate on templates that have been stalled in the act of replication but not on templates that have completed replication.

**DHFR Gene Expression**

In addition to using MTX-resistant cells for studying the mechanisms of gene amplification, MTX-resistant cells have been useful for studying the regulation of expression of the DHFR gene, normally expressed at low levels in wild type cells. Of particular interest is the production of multiple DHFR mRNAs (Setzer *et al.*, 1980), the growth dependent expression of DHFR (Alt *et al.*, 1976; Chang and Littlefield, 1976; Johnson *et al.*, 1978;) and the cell cycle regulation of DHFR (Mariani *et al.*, 1981).

Cytoplasmic poly(A)-containing RNA from both mouse and hamster cells were shown to express multiple DHFR mRNAs by using RNA transfer hybridization experiments with $^{32}$P-labeled DHFR-specific cDNA probes (Setzer *et al.*, 1980). Analysis of the DNA sequences of mouse cDNAs and of the 3'-untranslated sequence of the mouse DHFR gene demonstrated that the four different mouse DHFR-specific mRNAs (750, 1000, 1200 and 1600 nt) are colinear with the gene in the 3'-untranslated region and that the four transcripts differ at the site used for polyadenylation (Setzer *et al.*, 1982). These results are consistent with a model which hypothesizes that the multiple RNAs also derived from the same precursor transcript by differential polyadenylated in the 3'-untranslated region. Only the 1600 nt DHFR mRNA has a consensus polyadenylation sequence (AATAAA; Proudfoot and Brownlee, 1976) 10-15 nt upstream of the polyadenylation site, while the others have a shortened version (ATAA)

15-50 nucleotides upstream of their respective polyadenylation sites. This shortened sequence is found ten times in the 3'-untranslated region, and not all of the sites were associated with polyadenylated transcripts. The hypothesis that polyadenylation sites are signaled by these shortened sequences is, therefore, somewhat dubious.

Multiple RNAs differing in their 3'-untranslated regions are found in other protein-coding genes of higher eukaryotes, including the mouse α-amylase gene (Tosi *et al.*, 1981); the mouse $\beta_2$-microglobulin gene (Parnes and Robinson, 1983); the chicken pro α2(I) collagen gene (Aho *et al.*, 1983); the chicken ovomucoid gene (Gerlinger *et al.*, 1982); and the chicken vimentin gene (Zehner and Paterson, 1983). The vimentin transcripts have two size classes of mRNA, approximately 2.0 and 2.3 kb, that are present in muscle cells, fibroblasts, spinal cord and lens. However, the 2.0 kb species is predominantly expressed in erythroid cells taken from 10- to 15-day-old chicken embryos, in contrast to erythroid cells taken from 4-day-old embryos where both RNAs are expressed in equal but lower amounts (Capetanaki *et al.*, 1983). This is the first example of developmental regulation of 3'-untranslated sequences and raises intriguing questions about the regulation of the other genes expressing RNAs with multiple 3'-ends, and of the possible functional roles that these sequences might play. The functions of these 3'-ends could include directing mRNA to various subcellular compartments or could function in regulating mRNA stability. In the case of DHFR, the 3'-untranslated sequences appear to be important in the regulation of the growth dependent expression of DHFR in mammalian cells at the level of nuclear DHFR RNA turnover (discussed below).

Exponentially growing mouse cells contain 10-20 fold more steady state levels of cytoplasmic DHFR mRNA than stationary cells, a situation which corresponds to an increased abundance of the DHFR enzyme (Kellems *et al.*, 1976). Analysis of cytoplasmic RNA accumulation rates by using $^3$H-uridine labeling of stationary or exponentially growing mouse cells indicates that the exponentially growing cells accumulate DHFR RNA 3-4 times more rapidly, with no apparent change in the

cytoplasmic mRNA stability (Kellems *et al.*, 1979; Hendrickson *et al.*, 1980). This difference in DHFR RNA accumulation is reflected in the nuclear RNA accumulation rate.

An attempt to measure nuclear DHFR RNA synthesis rates in exponentially growing mouse cells and stationary cells using 20 min. pulses of $^{3}$H-uridine showed higher rates of accumulation of nuclear DHFR RNA in exponentially growing cells (Wu and Johnson, 1982). Therefore, transcriptional controls were postulated to be responsible for the observed growth dependent expression of DHFR gene. In conflict with these results, using experiments with shorter $^{3}$H-uridine pulses (1-5 min), Leys and Kellems (1981) found no differences in the nuclear DHFR RNA accumulation rates from samples of the two cell populations. The latter result suggests that the growth dependent regulation of DHFR is a post-transcriptional event rather than transcriptional. The differences in these results could be explained by nuclear RNA turnover becoming an added consideration in the transcription rate experiments using the longer pulses. Unfortunately, neither investigation measured the UTP pool sizes in the two cell populations, which could vary tremendously in exponentially growing cells and stationary cells.

Support for the post-transcriptional basis of the control of growth dependent expression of DHFR comes from gene transfection experiments using a modular DHFR gene and a Chinese hamster DHFR-deficient cell line (Kaufman and Sharp, 1983). The modular gene consisted of the major late adenovirus 2 promoter containing its 5'-untranslated region and a 5' splice site, an immunoglobulin 3' splice site, a mouse DHFR cDNA containing 3 of the 4 polyadenylation sites in its 3'-untranslated sequence and an SV40 polyadenylation site. Analysis of the poly(A)-containing RNA from the Chinese hamster cells tranformed to a DHFR$^{+}$ phenotype showed that the transcripts using polyadenylation sites in the 3'-untranslated region of the DHFR cDNA were growth regulated in their expression. In other words, the steady-state levels of these RNA

species were lower in stationary phase cells compared with exponentially growing cells. In contrast, the steady state levels of the most abundant transcript, which uses the more efficient SV40 polyadenylation site, did not differ between stationary and exponentially growing cells. Transcription experiments using isolated nuclei from stationary cells and exponentially growing cells showed no differences in the amount of DHFR-specific RNA, indicating that the control of the growth dependent expression of the transfected DHFR gene is not at the transcriptional level. These results also indicated that the 3'-untranslated region plays a role in the regulation of the growth dependent expression of DHFR, most likely in post-transcriptional events involving nuclear RNA turnover.

Cell cycle experiments, using a physical procedure to remove mitotic cells from culture dishes for synchronization, demonstrated that the amount of the DHFR enzyme peaked in the early S phase of mouse cells (2-3 fold; Mariani *et al.*, 1981). When stationary cells were stimulated to enter the cell cycle, an increase in the accumulation of DHFR RNA occurred (Johnson *et al.*, 1978) at the same time as DNA synthesis increased. This result demonstrates that growth-dependent expression of DHFR is related to cell cycle regulation. This cell cycle regulation of DHFR gene expression correlates with the need of deoxyribonucleotides for DNA replication and is a general feature of other enzymes involved in DNA synthesis (reviewed in Klevecz and Gerald, 1977).

Since drug resistance is a general feature of mammalian cells, and the gene amplification mechanism of resistance indicates a great plasticity of the genome in these cells, MTX-resistant human cell lines were developed. It seemed reasonable to assume that knowledge of the organization and expression of the human DHFR gene would be required in understanding the mechanism of gene amplification. Therefore, the main focus of this thesis is on determining the DHFR gene structure and analyzing its expression in MTX-resistant and MTX-sensitive cells. A brief summary of the results from the analysis of the human DHFR gene system is described below.

**Chapter 2: Characterization of the MTX-Resistant Cell Lines**

Analysis of various human MTX-resistant cells derived from HeLa or $VA_2$-B cell lines showed that all of the cell lines became resistant to increasing amounts of MTX (final concentration of $1.8x10^{-4}M$) as a result of an increase in the amount of the DHFR enzyme. This increase was most likely due to a DHFR gene amplification event since all of the cells exhibited the type of chromosomal abnormalities associated with gene amplification, HSRs (Barry Maurer, unpublished results) and double minutes.

An interesting feature observed in these human MTX-resistant cell lines was the variability in DHFR enzyme content, chromosome constitution and growth characteristics, despite their resistance to the same level of MTX. This could be attributable to a population phenomenon whereby the DHFR enzyme content of individual cells in a given population varies widely, similar to that found in populations of MTX-resistant rodent cell lines (Brown *et al.*, 1981; Kaufman *et al.*, 1981; Kaufman and Schimke, 1981). The enzyme levels for the human MTX-resistant cell lines varied from 35-275 times the level found in the parental cells. Analysis of the chromosome constitution of these cell types showed that every MTX-resistant cell line contained double minutes. When grown in the absence of MTX, all of the cell lines lost the amplified DHFR enzyme activity. However, the number of double minutes did not correlate with either the amount of DHFR enzyme or the instability of the resistant phenotype in the absence of selection. These results are in contrast to the situation in mouse cell lines, where the number of double minutes correlate with both the enzyme content and the degree of instability of the MTX-resistant phenotype (Kaufman *et al.*, 1979)

**Chapter 3: Cloning of Human DHFR-Specific cDNAs**

When polysomal poly(A)-containing RNA was electrophoresed on denaturing agarose gels and stained with ethidium bromide, a predominant RNA species (3.8 kb) was

observed in the RNA samples from MTX-resistant cells (6A3 or 10B3) but not from the RNA samples from the parental cell lines ($VA_2$-B and HeLa, respectively). This RNA was shown to code for the DHFR enzyme by *in vitro* translation experiments. These results demonstrated that the increased DHFR enzyme content in MTX-resistant cell lines is due to an increase in the amount of DHFR-specific mRNA.

Polysomal poly(A)-containing RNA from 6A3 cells was used to make a cDNA library, and two methods were used to select for DHFR-specific recombinant clones. The first method was differential hybridization (St. John and Davis, 1979), using $^{32}P$-labeled cDNA from 6A3 cells and $VA_2$-B cells to screen replicate filters of recombinant clones. Positive clones containg DHFR inserts were identified by a darker hybridization signal using the 6A3 cDNA probe compared to the $VA_2$-B probe. The second screening procedure involved phenotypic expression of the human DHFR cDNA in order to protect *E. coli* recombinant cells from growth inhibition when exposed to the bacterial DHFR-specific inhibitor, trimethoprim (Chang *et al.*, 1978). Comparison of the DNA sequence of the 5'-end of a phenotypically-expressed cDNA clone (pHD84), with the sequence of a mouse DHFR cDNA clone, confirmed the insert coded for DHFR.

Various DHFR cDNAs were used to probe RNA transfer hybridization experiments, using 6A3 polysomal poly(A)-containing RNA. These results showed that at least three RNAs (3.8, 1.0 and 0.8 kb) exist for human DHFR, and that these RNAs differed in the length of their 3'-untranslated regions.

**Chapter 4: DNA Sequence of the DHFR Coding Region**

The DNA sequence of the complete DHFR coding region was determined from the DHFR cDNA clone, pHD84, a cDNA found to be phenotypically expressed in *E. coli* (Chapter 3). The human DHFR coding sequence (564 nt) showed homology to the mouse DHFR coding sequence with an 89% sequence identity; of the 64 nucleotide differences, 45 represented silent base pair changes. When the derived amino acid sequence of the

human DHFR protein was compared to the homologous proteins from chicken, bovine and mouse, sequence identities of 74%, 81% and 89% were found.

**Chapter 5: Identification of a Human DHFR Pseudogene**

The DNA sequence of four human DHFR cDNAs supports the hypothesis that the multiple DHFR mRNAs are generated from the same transcription unit by differential polyadenylation of the precursor RNA (Setzer *et al.*, 1982). The data indicate that pHD84 is a cDNA of the 0.8 kb mRNA; pHD80 and pHD83 are both cDNAs of the 1.0 kb mRNA and pHD41 is a cDNA derived from a longer transcript, possibly the 3.8 kb mRNA. The sequences of the 3'-untranslated regions are colinear in each of the cDNA clones and differ in the sites of polyadenylation.

When *Eco*RI-digested nuclear DNA from human sperm, HeLa cells, $VA_2$-B cells and MTX-resistant 6A3 cells were used for DNA transfer hybridization experiments with the $^{32}$P-labeled pHD84 cDNA probe, five DHFR-specific fragments were observed to be amplified in the 6A3 DNA samples relative to the other samples (sizes of 13, 6.0, 4.0, 1.8 and 1.5 kb). This result, combined with data using different DHFR cDNA fragments, indicated that these five fragments are part of the amplified DHFR gene (see Chapter 6). At least five additional fragments hybridized to the pHD84 probe in all of the DNA samples, with sizes of 18, 17, 3.8, 3.5 and 2.3 kb. These fragments, in contrast to the DHFR gene segments, were not amplified in 6A3 cell DNA relative to HeLa cell and $VA_2$-B cell DNA.

In order to isolate a complete DHFR gene, a human cosmid library was screened with the $^{32}$P-labeled pHD84 insert. Ten DHFR-positive clones were identified, but these did not contain any of the *Eco*RI fragments found to be amplified in 6A3 cells (described above). However, one of the clones, CosHD37, contained two of the unamplified *Eco*RI fragments (3.5 and 2.3 kb). Restriction enzyme mapping and DNA sequence analysis of this clone demonstrated that these fragments contained a cDNA-like insert of the DHFR

gene, including the *Eco*RI site found in the 5'-end of pHD84. A nucleotide sequence identity of 92% was observed between the cDNA clones and the homologous region of the cosmid clone, with frameshift mutations interrupting the coding sequence. Since this DHFR homologous region cannot code for the DHFR enzyme, the cosmid clone contains a DHFR intronless pseudogene (ψHD1). Chen *et al.* (1982) have identified two additional DHFR pseudogenes that were distinct from ψHD1. Analysis of other gene systems has led to the hypothesis that an RNA intermediate is involved in the formation of intronless pseudogenes (Nishioka *et al.*, 1980; Vanin *et al.*, 1980; Karin and Richards, 1982). The DNA sequence homology between the DHFR cDNAs and ψHD1 extends beyond the polyadenylation site of the 1.0 kb DHFR mRNA. Therefore, if an RNA intermediate is involved in generating ψHD1, then an RNA larger than the 1.0 kb DHFR mRNA must be involved.

Subsequent analysis has localized the human DHFR gene to human chromosome 5, utilizing DNA transfer hybridization experiments with DNA from mouse-human and Chinese hamster-human cell hybrids and the $^{32}$P-labeled pHD84 probe (Maurer *et al.*, 1984). This analysis also showed that the unamplified DHFR-specific *Eco*RI fragments described above, did not segregate with the functional DHFR gene. Thus, all of these unamplified fragments probably represent DHFR pseudogenes or incomplete DHFR genes.

**Chapter 6: Isolation of the Human DHFR Gene**

The complete DHFR gene from 6A3 cells was isolated as overlapping inserts from four lambda genomic libraries made with two different vectors. Analysis by DNA transfer hybridization experiments of nuclear DNAs from 6A3 and HeLa cells showed no differences in the structure of the gene in DHFR amplified and unamplified cells. The gene is approximately 29 kb long, with the 564 bp coding region separated by five introns. The positions of the splice junctions in the DHFR coding sequence are identical

to the positions in the homologous mouse gene (Crouse *et al.*, 1982). However, the corresponding introns of the human, mouse and Chinese hamster DHFR genes show a large variation in their size, except for the first intron. This variation in intron size is greatest for intron five, which varies in length of up to fourfold. In contrast to the intron size variation, the overall lengths of the DHFR genes are equivalent among the three mammalian species.

A sequence comparison between human and mouse DHFR genes reveals some interesting features. There is a relatively high G+C content in the 5'-ends of the human gene (an average of 70%) while there is a relatively low G+C content in the remainder of the gene (an average of 40%). There is no detectable sequence homology in the intron sequences of the human and mouse DHFR genes other than the 15-35 nucleotides directly adjacent to the intron/exon splice junctions. However, the overall G+C composition is conserved. The 5'-noncoding regions of the mouse and human genes also share a significant nucleotide sequence identity (65%) to at least 327 bp upstream of the translation initiation codon. This sequence homology suggests that this region plays an important regulatory role in the expression of DHFR. In contrast to the 5'-end of the DHFR genes from mouse and human cells, the 3'-untranslated regions show no obvious sequence homology. Since the 3'-untranslated sequence of the mouse gene has been implicated in the control of the growth dependent expression of DHFR (discussed above), it would be of interest to determine if a similar mechanism exists for the control of the expression of the corresponding human gene.

**Chapter 7: Mapping of the Major and Minor 5'-ends of DHFR-Specific RNA**

Primer extension and S1 nuclease protection experiments have mapped multiple 5'-ends to DHFR-specific transcripts in the human MTX-resistant cell line, 6A3. One of these ends, position -71 nucleotides from the DHFR translation initiation codon, corresponds to a major DHFR-specific transcript which totals approximately 99% of the

DHFR polysomal poly(A)-containing RNA. This end is most likely associated with the previously identified DHFR mRNA of sizes 3.8, 1.0 and 0.8 kb. The -71 position has also been mapped by Chen *et al.* (1984) and was shown to be a functional transcription initiation site *in vitro*. Six other minor 5'-ends have been mapped to positions -449 to -480 of the DHFR gene and total approximately 1% of the DHFR-related polysomal poly(A)-containing RNA.

Analysis of polysomal poly(A)-containing RNA from HeLa cells and $VA_2$-B cells shows that these upstream initiating transcripts are present in MTX-sensitive cells and are not specific for DHFR amplified cells. Also, the relative levels of the DHFR-specific transcripts varied in relation to the extent of DHFR gene amplification. In contrast to the DHFR polysomal RNA, the upstream initiating transcripts in the nuclear RNA fraction represent approximately 10% of the major transcripts initiating at position -71. The differences in the relative levels of DHFR nuclear RNA and polysomal RNA indicate the existence of a distinct pattern of regulation of expression of the upstream initiating transcripts.

## References

Abrect, A., Biedler, J.L., and Hutchinson, D.J. (1972). Cancer Res. **32**, 1539-1546.

Aho, S., Tate, V. and Boedtker, H. (1983). Nucl. Acids Res. **11**, 5443-5450.

Alt, A.W., Kellems, R.E. and Schimke, R.T. (1976). J. Biol. Chem. **251**, 3063-3074.

Alt, A.W., Kellems, R.E., Bertino, J.R. and Schimke, R.T. (1978).
J. Biol. Chem. **235**, 1357-1370.

Andrulis, I.L., and Siminovitch, L. (1982). In R.T. Schimke (ed.) *Gene Amplification*, Cold Spring Harbor Laboratory. Cold Spring Harbor, New York. Pp. 75-80.

Beach, L.R. and Palmiter, R.D. (1981). Proc. Natl. Acad. Sci., U.S.A. **78**, 2110-2114.

Bertino, J.R., Donohue, D.M., Simmons, B., Gabrio, B.W., Silber, R. and Huennekens, F.M. (1963). J. Clin. Invest. **42**, 446-475.

Biedler, J.L., Albrecht, A.M., Hutchison, D.J., and Spengler, B.A. (1972). Cancer Res. **32**, 153-161.

Biedler, J.L., Melera, P.W. and Spengler, B.A. (1980). Cancer Genet. Cytogenet. **2**, 47-60.

Bostock, C.J. and Clark, E.M. (1980). Cell **19**, 709-715.

Botchan, M., Topp, W. and Sambrook, J. (1978). Cold Spring Harbor Symp. Quant. Biol. **43**, 709-719.

Brennand, J., Chinault, A.C., Konecki, D.S., Melton, D.W. and Caskey, C.T. (1982). Proc. Natl. Acad. Sci., U.S.A. **79**, 1950-1954.

Brinson, O., Adeshir, F. and Stark, G.R. (1982). Mol. Cell. Biol. **2**, 578-587.

Brown, D.D. and Dawid, I.B. (1968). Science, **160**, 272-280.

Brown, P.C., Beverley, S.M. and Schimke, R.T. (1981). Mol. Cell. Biol. **1**, 1077-1083.

Brown, P.C., Tlsty, T.D. and Schimke, R.T. (1983). Mol. Cell. Biol. **3**, 1097-1107.

Capetanaki, Y.G., Ngai, J., Flytzanis, C.N. and Lazarides, E. (1983). Cell, **35**, 411-420.

Chang, S.E. and Littlefield, J.W. (1976) Cell, **7**, 391-396.

Chang, A.C.Y., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schimke, R.T. and Cohen, S.N. (1978). Nature, **275** 617-624.

Chen, M.-J., Shimida, T., Moulton, A.D., Harrison, M. and Nienhuis, A.W. (1982). Proc. Natl. Acad. Sci., U.S.A. **79**, 7435-7439.

Chen, M.-J., Shamida, T., Moulton, A.D., Cline, A., Humphries, R.K., Mazel, J. and Nienhuis, A.W. (1984). J. Biol. Chem. **259**, 3933-3943.

Cowell, J.K. (1982). Ann. Rev. Genet. **16**, 21-39.

Crouse, G.F., Simonsen, C.C., McEwan, R.N. and Schimke, R.T. (1982). J. Biol. Chem. **257**, 7887-7897.

Curt, G.A., Carney, D.N., Cowan, K.H., Jolivet, J., Bailey, B.D., Drake, J.C., Kao-Shan, C.S., Minna, J.D. and Chamber, B.A. (1983). New Engl. J. Med. **308**, 199-202.

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.

Fisher, G.A. (1961). Biochem. Pharmacol. **7**, 75-80.

Fisher, G.A., (1962) Biochem. Pharmocol. **11**, 1233-1237.

Flintoff, W.F. and Essani, K. (1980). Biochemistry. **19**, 4321-4327.

Flintoff, W.F., Davidson, S.V. and Siminovitch, L. (1976). Somatic Cell Genet. **2**, 245-261.

Fuscoe, J.C., Fenwick, R.G. Jr., Ledbetter, D.H. and Caskey, C.T. (1983). Mol. Cell. Biol. **3**, 1086-1096.

Gerlinger, P., Krust, A., LeMeur, M., Perrin, F., Cochet, M., Gannon, F., Dupret, D. and Chambon, P. (1982). J. Mol. Biol. **162**, 345-364.

Haber, D.A., Beverley, S.M., Kiely, M.L., and Schimke, R.T. (1981). J. Biol. Chem. **256**, 9501-9510.

Hakala, M.T., Zakrewski, S.F. and Nichol, C.A. (1961). J. Biol. Chem. **236**, 952-958.

Hendrickson, S.L., Wu, J.-S.R. and Johnson, L.F. (1980). Proc. Natl. Acad. Sci., U.S.A. **77**, 5140-5144.

Johnson, L.F., Fuhrman, C.L. and Wiedemann, L.M. (1978). J. Cell. Physiol. **97**, 397-406.

Johnston, R.N., Beverley, S.M. and Schimke, R.T. (1983). Proc. Natl. Acad. Sci., U.S.A. **80**, 3711-3715.

Jolivet, J., Cowan, K.H., Curt, G.A., Clendeninn, N.J. and Chabner, B.A. (1983). New Engl. J. Med., **309**, 1094-1104.

Kanalas, J.J. and Suttle, D.P. (1984). J. Biol. Chem. **259**, 1848-1853.

Karin, M. and Richards, R.I. (1982). Nature, **299**, 797-802.

Kaufman, R.J., Brown, P.C. and Schimke, R.T. (1979). Proc. Natl. Acad. Sci., U.S.A. **76**, 5669-5673.

Kaufman, R.J., Brown, P.C. and Schimke, R.T. (1981). Mol. Cell. Biol. **1**, 1084-1093.

Kaufman, R.J. and Schimke, R.T. (1981). Mol. Cell. Biol. **1**, 1069-1076.

Kaufman, R.J. and Sharp, P.A. (1983). Mol. Cell. Biol. **3**, 1598-1608.

Kellems, R.E., Alt. F.W. and Schimke, R.T. (1976). J. Biol. Chem. **251**, 6987-6983.

Kellems, R.E., Morhenn, V.B., Pfendt, E.A., Alt, F.W. and Schimke, R.T. (1979). J. Biol. Chem. **254**, 309-318.

Kempe, T.D., Swyryd, E.A., Bruist, M. and Stark, G.R. (1976). Cell, **9**, 541-550.

Klevecz, R.R. and Gerald, L.R. (1977). In G.H. Rothblat and V.J. Cristofalo (eds.) *Growth, Nutrition and Metabolism of Cells in Culture.* Academic, New York. Pp. 149-196.

Leys, E.J. and Kellems, R.E. (1981). Mol. Cell. Biol. **1**, 961-971.

Ling, V. (1982). In N. Bruchovsky and J. H. Goldie (eds.) *Drug and Hormone Resistance in Neoplasia*, **1**, CRC Press, Boca Raton, Fla. Pp. 1-19.

Luria, S.E. and Delbrück, M. (1943). Genetics, **28**, 491-511.

Luskey, K.L., Faust, J.R., Chin, D.J., Brown, M.S. and Goldstein, J.L. (1983). J. Biol. Chem. **258**, 8462-8469.

Mariani B.D., Slate, D.L. and Schimke, R.T. (1981). Proc. Natl. Acad. Sci., U.S.A. **79**, 4985-4989.

Mariani, B.D., and Schimke, R.T. (1984). J. Biol. Chem. **259**, 1901-1910.

Matthews, D.A., Alden, R.A., Bolin, J.T., Filman, D.J., Freer, S.T., Hamlin, R., Hol, W.G.J., Kisliuk, R.L., Pastore, E.J., Plante, L.T., Xuong, N. and Kraut, J. (1978). J. Biol. Chem. **253**, 6946-6954.

Maurer, B.J., Barker, P.E., Masters, J.N., Ruddle, F.H. and Attardi, G. (1984). Proc. Natl. Acad. Sci., U.S.A. **81**, 1484-1488.

Melera, P.W., Lewis, J.A., Biedler, J.L. and Hession, C. (1980). J. Biol. Chem. **255**, 7024-4028.

Nishioka, Y., Leder, A. and Leder, P. (1980). Proc. Natl. Acad. Sci., U.S.A. **77**, 2806-2809.

Nunberg, J.H., Kaufman, R.J., Schimke, R.T., Urlaub, G. and Chasin, L.A. (1978). Proc. Natl. Acad. Sci., U.S.A. **75**, 5553-5556.

Padgett, R.A., Wahl, G.M. and Stark, G.R. (1982). Mol. Cell. Biol. **2**, 293-301.

Pall, M.L. (1981). Proc. Natl. Acad. Sci., U.S.A. **78**, 2465-2468.

Parnes, J.R. and Robinson, R.R. (1983). Nature, **302**, 449-452.

Proudfoot, N.J. and Brownlee, G.G. (1976). Nature, **263**, 211-214.

Roberts, J.M. and Axel, R. (1982). Cell, **29**, 109-119.

Roninson, I.B., Abelson, H.T., Housman, D.E., Howell, N. and Varshavsky, A. (1984). Nature, **309**, 626-628.

Rossana, C., Rao, L.G. and Johnson, L.F. (1982). Mol. Cell. Biol. **2**, 1118-1125.

Schimke, R.T., Brown, P.C., Kaufman, R.J., McGrogan. M. and Slate, D.L. (1980). Cold Spring Harbor Symp. Quant. Biol. **45**, 785-797.

Schimke, R.T. (1982). (ed.) ***Gene Amplification***, Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T. (1980). Cell, **22**, 361-370.

Setzer, D.R., McGrogan, M. and Schimke, R.T. (1982). J. Biol. Chem. **257**, 5143-5147.

Simonsen, C.C. and Levinson, A.D. (1983). Proc. Natl. Acad. Sci., U.S.A. **80**, 2495-2499.

Sirotnak, F.M., Kurita, S. and Hutchinson, D.J. (1968). Cancer Res. **28**, 75-80.

Spradling, A.C. (1981). Cell, **26**, 193-201.

St. John, T. and Davis, R.W. (1979). Cell, **16**, 443-452.

Tartof, K.D. (1975). Ann. Rev. Genet. **9**, 355-385.

Terzi, M. and Hawkins, T.S.C. (1974). Biochem. Exp. Biol. **11**, 245-254.

Tlsty, T., Brown, P.C., Johnston, R. and Schimke, R.T. (1982). In R.T. Schimke, (ed.) *Gene Amplification*, Cold Spring Harbor Laboratory. Cold Spring Harbor, New York. Pp. 231-238.

Tosi, M., Young, R.A., Hagenbüchle, O. and Schibler, U. (1981). Nucl. Acids Res. **9**, 2313-2323.

Vanin, E.F., Goldberg, G.I., Tucker, P.W. and Smithies, O. (1980). Nature, **286**, 222-226.

Varshavsky, A. (1981a). Proc. Natl. Acad. Sci., U.S.A. **78**, 3673-3677.

Varshavsky, A. (1981b). Cell, **25**, 561-572.

Volz, K.W., Matthews, D.A., Alden, R.A. Freer, S.T., Hansch, C., Kaufman, B.T. and Kraut, J. (1982). J. Biol. Chem. **257**, 2528-2536.

Wahl, G.M., Padgett, R.A. and Stark, B.R. (1979). J. Biol. Chem. **254**, 8679-8689.

Wahl, G.M., Vitto, L. and Rubnitz, J. (1984). Mol. Cell. Biol. **3**, 2066-2075.

Wu, J.-S.R. and Johnson, L.F. (1982). J. Cell. Physiol. **110**, 183-189.

Zehner, Z.E. and Paterson, B.M. (1983). Proc. Natl. Acad. Sci., U.S.A. **80**, 911-915.

# CHAPTER 2

## Variable Content of Double Minute Chromosomes Is Not Correlated with Degree of Phenotype Instability in Methotrexate-Resistant Human Cell Lines

MOLECULAR AND CELLULAR BIOLOGY, May 1982, p. 498–507
0270-7306/82/050498-10$02.00/0

Vol. 2, No. 5

# Variable Content of Double Minute Chromosomes Is Not Correlated with Degree of Phenotype Instability in Methotrexate-Resistant Human Cell Lines

JEFFREY MASTERS, BENNETA KEELEY, HELEN GAY, AND GIUSEPPE ATTARDI*

*Division of Biology, California Institute of Technology, Pasadena, California 91125*

Received 3 September 1981/Accepted 3 December 1981

Several variants resistant to $1.8 \times 10^{-4}$ M DL-methotrexate (MTX) have been isolated from the human cell lines HeLa BU25 and $VA_2$-B by exposing them to progressively increasing concentrations of the drug. A striking variability of phenotype and chromosome constitution was observed among the different variants. All resistant cell lines exhibited a greatly increased dihydrofolic acid reductase (DHFR) activity and DHFR content; however, the DHFR activity levels varied considerably among the variants, ranging between about 35 and 275 times the parental level. In the absence of selective pressure, the increased DHFR activity was unstable, and in all cell lines but one was completely lost over a period ranging in different variants between 25 and 200 days. The MTX-resistant cell lines showed anomalies in their chromosome constitution, which involved the occurrence of a duplicated set of chromosomes in most cells of some of the variants and the presence of double minute chromosomes in all cell lines. An analysis of the correlation of loss of double minute chromosomes and loss of DHFR activity in the absence of MTX has given results consistent with the idea that the double-minute chromosomes contain amplified DHFR genes. However, the most significant finding is that, in contrast to what has been reported in the mouse system, the recognizable double-minute chromosomes varied greatly in number in different variants without any relationship to either the level of DHFR activity or the degree of instability of MTX resistance in the absence of selective pressure. These and other observations point to the occurrence in the human MTX-resistant variants of another set of DHFR genes, representing a varied proportion of the total, which is associated with the regular chromosomes, and which may be unstable in the absence of selective pressure.

One of the most striking examples of the plasticity of the genome in animal cells is provided by the selective amplification of the genes coding for proteins with binding specificity for certain metabolic inhibitors or other harmful chemicals, as demonstrated in cultured mammalian cells resistant to such agents (1, 5, 29). The best-studied case of this nature is the dihydrofolic acid reductase (DHFR) gene amplification in mammalian cells resistant to amethopterin (methotrexate, MTX) (1). In several human and experimental tumors and in cultured mouse and hamster cell lines, the development of the resistance to MTX has been found to be due to an increased cell content, up to several hundredfold, of the enzyme (6, 8, 10–12, 15, 22, 26, 27, 30). In cultured mammalian cell lines, the increased enzyme level has been shown to result from an increased rate of synthesis (2, 16), due to a selective amplification of the DHFR genes (1, 10, 23, 26) and to a parallel increase in the amount of DHFR mRNA (1, 9, 18, 23).

The biochemical investigations mentioned above have all been carried out in MTX-resistant cell lines of mouse or hamster origin. Recently, several MTX-resistant cell variants have been isolated in this laboratory from two different human cell lines, HeLa BU25 and $VA_2$-B; furthermore, the DHFRs have been isolated from HeLa BU25 cells and from two of the MTX-resistant derivatives and found to be indistinguishable in their physical and enzymatic properties (25). In this paper, we report that the human MTX-resistant cell lines isolated here exhibit a striking variability in phenotypic properties and chromosome constitution, even among sublines. Most significantly, they contain a highly variable number of double minute chromosomes which is not correlated with either the level or the stability of their resistance to MTX.

## MATERIALS AND METHODS

**Materials.** Dihydrofolic acid and NADPH were obtained from Sigma Chemical Co. The MTX used as a

selective agent for the drug-resistant variants was the DL-form from Sigma (consisting of ~50% L-form).

**Cell lines and methods of growth.** The human cell line HeLa BU25, a mutant subline of HeLa S3 deficient in thymidine kinase activity (19), kindly provided by S. Kit, was grown in petri dishes in Eagle medium supplemented with 10% calf serum in the absence or presence of 25 μg of bromodeoxyuridine (BUdR) per ml. The human cell line $VA_2$-B, an azaguanine-resistant subclone of the simian virus 40-transformed line WI-18-$VA_2$ (28), was grown in petri dishes in Eagle medium supplemented with 10% calf serum and 3 μg of 8-azaguanine per ml. The MTX-resistant variants of these two cell lines were grown in the media described above supplemented with 10% extensively dialyzed calf serum and various concentrations of DL-MTX as specified below.

**Chromosome analysis.** Cells were arrested in metaphase by treatment with 0.05 μg of colchicine per ml for 3 to 5 h. Karyotype analysis was performed as described previously (24).

**Cell extraction preparation and enzyme assay.** When a cell extract had to be prepared for enzyme assay, the cells were grown in the absence of MTX for 7 days before harvesting. The preparation of cell extracts and the enzyme assays were carried out essentially as described by Frearson et al. (13), except that in the enzyme assay the concentration of dihydrofolic acid was doubled. One unit of enzyme activity was defined as the amount of enzyme required to reduce 1 nmol of dihydrofolic acid in 15 min at 21°C.

**Electrophoretic analysis.** Electrophoresis of the cell extracts was carried out by using the Laemmli sodium dodecyl sulfate-polyacrylamide gel system (20) as previously described (25).

## RESULTS

**Isolation of MTX-resistant variants.** HeLa BU25 cells previously grown in the absence or presence of BUdR (25 μg/ml) were exposed to $2 \times 10^{-8}$ or $5 \times 10^{-8}$ M DL-MTX in BUdR-free (−BUdR) or BUdR-containing (+BUdR) medium, respectively; similarly, $VA_2$-B cells were treated with either $5 \times 10^{-8}$ or $1 \times 10^{-7}$ M DL-MTX. Almost all of the cells treated with the above concentrations of MTX died in the first 2 weeks. Four colonies of BU25 (−BUdR) cells and four colonies of BU25 (+BUdR) cells resistant to $2 \times 10^{-8}$ M DL-MTX, four colonies of BU25 (+BUdR) cells resistant to $5 \times 10^{-8}$ M DL-MTX, four colonies of $VA_2$-B cells resistant to $5 \times 10^{-8}$ M DL-MTX, and four colonies of the same cell line resistant to $1 \times 10^{-7}$ M DL-MTX were picked up by trypsinization 3 weeks after plating, and each was grown to a large culture. These clones, designated with a number indicating the plate and a letter indicating the clone picked up in each plate (for example, 10A, 10B, 10C, and 10D), were then exposed to increasing DL-MTX concentrations. Steps corresponding to two- to fourfold increases in MTX concentration were used, the final concentration being $1.8 \times 10^{-4}$ M DL-MTX. Considerable cell death occurred after each increase; the surviving cells were grown for several weeks until no obvious cell death occurred before the next increase in MTX concentration in the medium was made. In the course of this process, each variant population was often subdivided into subpopulations, which were exposed to different schedules of increasing MTX concentrations, being thereafter maintained separate (designated by a progressive number following the clone letter designation, for example, 10B1, 10B2, 10B3). The whole process took between 7 and 11 months for different variants. Eleven variant cell lines, six derived from BU25 (−BUdR) cells, one derived from BU25 (+BUdR) cells, and four derived from $VA_2$-B cells, which had acquired resistance to $1.8 \times 10^{-4}$ M DL-MTX, were selected for further investigation (Table 1). Because of the continuous changes occurring in the MTX-resistant variants during growth in the presence of the drug, each variant is designated by its symbol and a subscript indicating the number of months of exposure to $1.8 \times 10^{-4}$ M DL-MTX.

**Growth characteristics of the MTX-resistant variants.** All variants were found to grow more slowly in the presence of the drug than the parental cell lines grew in the absence of the drug. In particular, in the first few months after their adaptation to $1.8 \times 10^{-4}$ M DL-MTX, their population doubling times varied between 35 h and more than 70 h. A marked acceleration in growth rate occurred in the variants after prolonged exposure to the drug, their population doubling time approaching that of the parental cell lines. The slower growth rate of the MTX-resistant variants was not due exclusively to the presence of the drug, as indicated by the observation that in several MTX-resistant variants, in the first 3 weeks after removal of the drug, the rate of cumulative growth was appreciably slower than that of the $VA_2$-B parental cell line, gradually increasing thereafter.

**Levels of DHFR in MTX-resistant and MTX-sensitive cell lines.** All of the MTX-resistant cell lines analyzed here exhibited an increased DHFR activity relative to the parental MTX-sensitive cell lines (Table 1). The increase in DHFR levels in MTX-resistant variants, as measured within the first 10 months after their adaptation to $1.8 \times 10^{-4}$ M DL-MTX, varied between about 35 and 275 times the levels found in the corresponding parental cell lines. It is interesting that a great variation in DHFR activity was observed also among sublines derived from the same original MTX-resistant clone (compare, for example, 7D2 and 7D3).

That the increase in DHFR activity in the MTX-resistant variants analyzed here reflected an increase in enzyme protein was shown by an electrophoretic analysis of the cell extracts.

TABLE 1. DHFR activity and chromosome constitution of MTX-sensitive and MTX-resistant cell lines

| Cell line[a] | DHFR activity (U/mg of protein) | Relative DHFR activity | No. of days for decay of DHFR activity in MTX-free medium | | Chromosome constitution[b] | | |
|---|---|---|---|---|---|---|---|
| | | | To 50% | To 20% | Complement[c] | No. of chromosomes/cell[d] | No. of double minutes per cell[e] |
| Parental | | | | | | | |
| HeLa BU25 (−BUdR) | 50.5 ± 14.7 | 1 | | | 1S (100) | 59 ± 1 | 0.5 (0–6) |
| Variants | | | | | | | |
| $10B1_1$ | 11,500 | 228 | 13 | 33 | 1S (100) | 57 ± 2 | 19 (0–132) |
| $10B2_1$ | 9,100 | 180 | 21 | 38 | 1S (95)<br>2S (5) | 58 ± 3<br>93, 111 | 28 (0–130) |
| $10B3_3$ | 13,800 | 273 | 16 | 31 | 1S (97)<br>2S (3) | 57 ± 4<br>113 | 7 (0–42) |
| $10C1_1$ | 5,400 | 107 | 14 | 21 | 1S (73)<br>2S (24)<br>3S (3) | 55 ± 1<br>106 ± 5<br>166 | 75 (4–257) |
| $10C2_1$ | 5,200 | 103 | 47 | 86 | 1S (19)<br>2S (81) | 55 ± 5<br>112 ± 3 | 73 (5–265) |
| $10C3_4$ | 5,400 | 107 | 11 | 27 | 1S (8)<br>2S (89)<br>4S (3) | 56 ± 2<br>110 ± 5<br>210 | 219 (46–475) |
| Parental | | | | | | | |
| HeLA BU25 (+ BUdR) | 45.8 ± 7.2 | 1 | | | 1S (88)<br>2S (12) | 58 ± 1<br>109 ± 6 | <0.3 |
| Variant | | | | | | | |
| $14D_1$ | 9,600 | 210 | 12 | 49 | 1S (100) | 56 ± 2 | 26 (0–100) |
| Parental | | | | | | | |
| WI-18 $VA_2$-B | 79.3 ± 5.9 | 1 | | | 1S (97)<br>2S (3) | 68 ± 2<br>136 | <0.3 |
| Variants | | | | | | | |
| $6A2_8$ | 11,300 | 142 | 8 | 20 | 1S (97)<br>2S (3) | 58 ± 2<br>115 | 182 (34–841) |
| $6A3_2$ | 7,300 | 92 | 34 (18)[f] | 240 (39)[f] | 1S (94)<br>2S (6) | 56 ± 1<br>106, 112 | 154 (42–391) |
| $7D2_{10}$ | 11,400 | 144 | 18 | 90 | 2S (100) | 118 ± 5 | 98 (2–261) |
| $7D3_2$ | 2,600 | 33 | 29 | 49 | 2S (100) | 122 ± 5 | 120 (9–348) |

[a] The subscripts in the symbols of the variants refer to the time when the analysis of DHFR activity was carried out.

[b] A total of 32 to 40 metaphase spreads were analyzed from each cell line within the first 6 months of maintenance in the presence of $1.8 \times 10^{-4}$ M DL-MTX.

[c] Numbers within parentheses indicate the percentage of cells in each category.

[d] Average ± standard deviation. Whenever only one or two metaphases were found in a given group, the individual values are reported.

[e] Since it was often difficult to identify the pairs of double minutes, especially when present in clusters, and since many of these elements appeared to be unpaired, the values reported here represent the mean and range of numbers of individual minutes per cell. From 45 to 74% of these minutes in different cell lines were in clearly recognizable pairs.

[f] The numbers within parentheses represent the number of days calculated after subtracting the contribution of the relatively stable DHFR fraction to the individual values of overall DHFR activity (see Fig. 6B).

When equal amounts of protein were run through sodium dodecyl sulfate-polacrylamide gels, a band corresponding to a polypeptide with an $M_r$ of ~22,000 comigrating with the purified DHFR (25) was observed in the extracts from the MTX-resistant variants, but not in those from the MTX-sensitive parental cell lines (Fig. 1). The extracts from 10C2 and 10C3 cells exhibited a band corresponding to DHFR which was clearly less pronounced than expected from

their degree of resistance to MTX and from their DHFR activity (Table 1).

**Behavior of increased DHFR activity in the MTX-resistant variants in the absence of MTX.** All of the MTX-resistant cell lines, at the time when they were analyzed (within the first 10 months after their adaptation to $1.8 \times 10^{-4}$ M DL-MTX), were unstable in the absence of MTX (Fig. 2 and Table 1); in fact, they gradually lost all or, in one case, most of their excess DHFR activity over periods ranging in different variants between 25 and 200 days. The cell line 6A3 was unique in that it showed a biphasic loss of the excess DHFR activity: this declined to about 30% in 50 days, and thereafter decreased only very slowly to reach ~10% in 400 days.

The loss of DHFR activity in the MTX-resistant variants in the absence of selection was accompanied by the loss of drug resistance. Thus, 10B3 cells which had been adapted to grow in the presence of $1.2 \times 10^{-4}$ M DL-MTX (DHFR specific activity of $10.4 \times 10^3$ U/mg of protein) and were then maintained in the absence of the drug for 7 weeks (during which time the DHFR activity had dropped to ~17% of the original value) gave only 0.3% survivors when challenged again with $1.2 \times 10^{-4}$ M DL-MTX. These survivors showed the same level of enzyme activity and, upon removal of the drug, the same degree of instability of MTX resistance as the original variants (Fig. 2).

**Chromosome constitution.** An analysis of the chromosome constitution of the MTX-resistant variants again revealed a marked variability. As illustrated in Table 1 and Fig. 3 and 4, which present the results of an analysis carried out within the first 6 months of maintenance of the variants in the presence of $1.8 \times 10^{-4}$ M DL-MTX, some cell lines exhibited predominantly a number of chromosomes close to that of their parental line, and others exhibited predominantly a duplicated number of chromosomes. This variability in chromosome number was also observed among sublines derived from the same original clone, as can be seen for the 10C sublines. One interesting observation is that, in all cases where no duplication of the chromosome complement occurred, the modal number of chromosomes in the MTX-resistant variants was lower than in the corresponding parental line (Table 1 and Fig. 3 and 4).

The most striking feature of the chromosome constitution of the MTX-resistant variants was the presence of double minute chromosomes, small, acentromeric chromosomal elements, which were absent or very rare in the parental

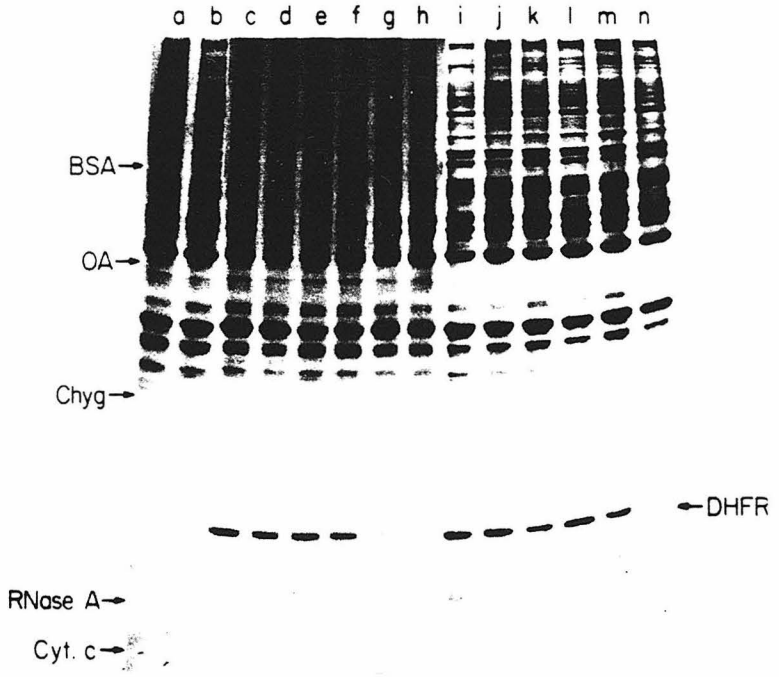



FIG. 1. Electrophoretic patterns in sodium dodecyl sulfate-polyacrylamide gel of the cell extracts from the parental lines and from the eleven MTX-resistant variants described in this work. Lanes: a, BU25 (−BUdR); b, BU25 (+BUdR); c, 10B1; d, 10B2; e, 10B3; f, 10C1; g, 10C2; h, 10C3; i, 14D; j, 6A2; k, 6A3; l, 7D2; m, 7D3; n, $VA_2$-B.

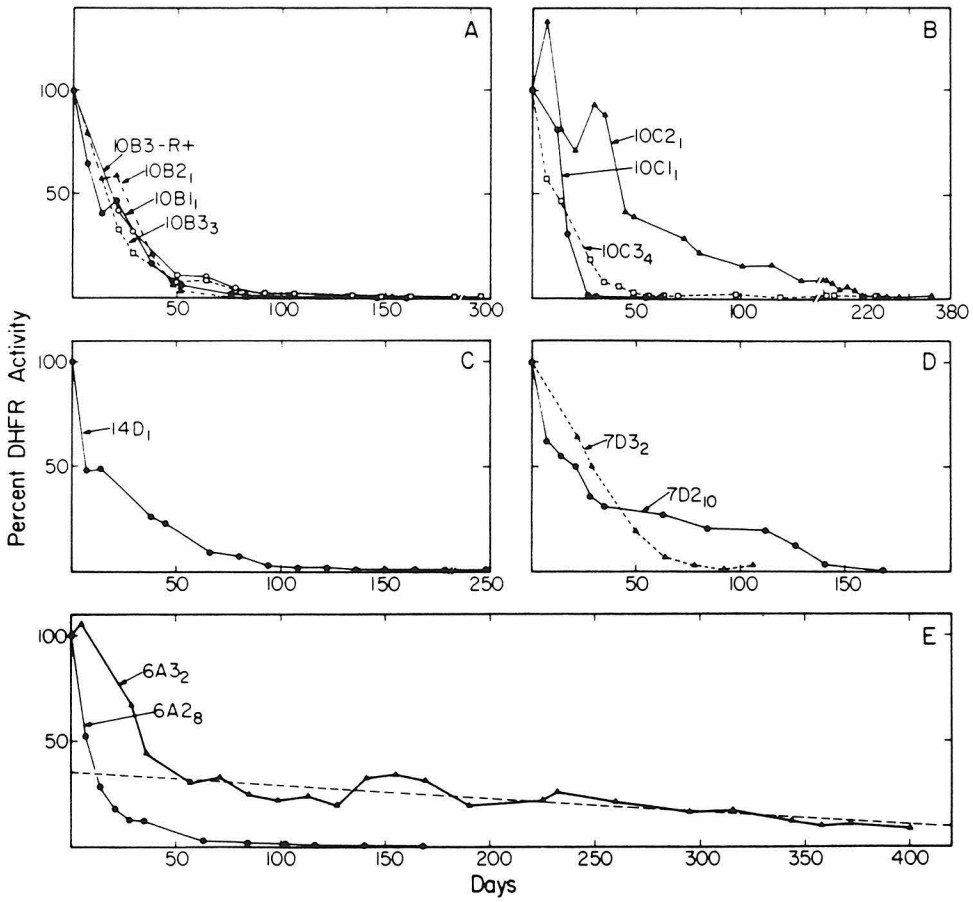



FIG. 2. Quantitative behavior of the DHFR activity of the MTX-resistant variants after removal of the selective pressure. The curve designated 10B3-R+ refers to partial revertants of 10B3 cells resistant to $1.2 \times 10^{-4}$ M DL-MTX which had been maintained for 7 weeks in drug-free medium and which had then survived reexposure to the same concentration of the drug. See text for details. The curve designated $10B3_3$ refers to cells resistant to $1.8 \times 10^{-4}$ M DL-MTX; an identical curve was obtained for cells resistant to $1.2 \times 10^{-4}$ M DL-MTX.

cell lines (Table 1 and Fig. 3, 4, and 5). As indicated by their name, the minute chromosomes appeared typically in pairs, though the distance between the members of each pair was highly variable; furthermore, the minute chromosomes frequently appeared as unpaired elements. In this paper, we refer to the minute chromosomes, whether single or in pairs, as double minute chromosomes or double minutes, although for quantitative purposes the minute chromosomes have been counted individually. They varied considerably in size and stainability with orcein. The predominant size and stainability of the double minutes of the individual cell lines tended to be reproducibly different (compare, for example, the double minutes of the 6A3 [Fig. 5D], 10C3 [Fig. 5E], and 7D3 [Fig. 5F] cells).

The number of double minute chromosomes varied greatly among different MTX-resistant variants, even among sublines. Some variants (10B1, 10B2, 10B3, and 14D) exhibited a relatively small average number per cell of recognizable double minutes (8 to 26) (Fig. 3 and 5C); in these variants, an analysis of the distribution of the number of these elements per cell in each cell population revealed very wide fluctuations, from 0 to more than 100 per cell (Table 1). Other variants (10C3, 6A2, 6A3, and 7D3) had a large average number of double minutes per cell (120 to 220) (Fig. 5D, 5E, and 5F); others (10C1, 10C2 and 7D2) had an intermediate number. In both cases there was again a wide range of values in each cell population (Table 1 and Fig. 3 and 4). An examination of Table 1 shows the absence of any correlation in the various MTX-resistant variants between the average number of double minutes per cell and either the level of DHFR

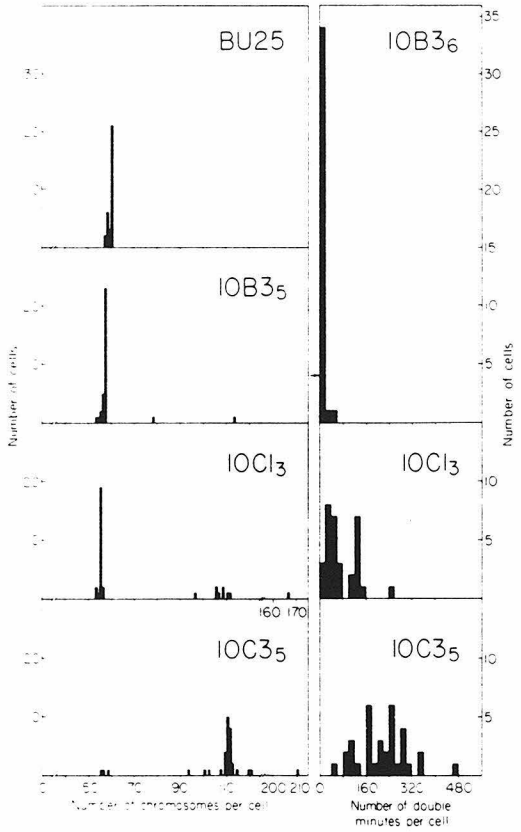



FIG. 3. Frequency distribution of number of chromosomes and number of double minutes per cell in the parental line BU25 (−BUdR) and its MTX-resistant derivatives 10B3, 10C1, and 10C3. The arrow pointing to the ordinate axis in the uppermost right panel indicates the number of cells with no double minutes.

activity or the degree of instability of the increased DHFR level upon removal of MTX. Thus, the variants 10B1, 10B2, and 10B3 had a much smaller number of identifiable double minute chromosomes than did the 10C2 and 7D2 variants, yet they lost their excess DHFR activity much faster. The line 10C3 had three times as many double minutes as the line 10C1, yet they lost their excess DHFR activity at the same rate.

**Correlation of loss of double minute chromosomes and loss of DHFR activity upon removal of MTX.** In mouse cell lines with unstable MTX resistance, a phenotype which in these cell lines appears to be correlated with the presence of double minute chromosomes, removal of the selective pressure leads to the loss of the double minutes (17). The same observation has been made here in the human MTX-resistant variant 6A3. When a suspension culture of these cells was grown for 6 days in the absence of MTX, the average number of double minute chromosomes dropped from 150 to 59. These cells were then tested for the behavior of the DHFR activity and the number of double minute chromosomes upon further growth in the absence of MTX. Both DHFR activity and the number of double minutes declined progressively over the next 100 days (Fig. 6A). However, the kinetics of loss of enzyme activity, which paralleled that observed in cells grown on plates (Fig. 6A), differed significantly from the kinetics of loss of double minute chromosomes, which was likewise similar to that observed in cells grown on plates (data not shown). As mentioned above, the observed loss of DHFR activity from 6A3 cells was biphasic, with a phase of rapid loss superimposed upon a slow loss (Fig. 2). When the contribution of the slow decrease was subtracted from the individual values of DHFR activity, the resulting curve, which represented only the rapid loss of DHFR activity, showed an excellent agreement with the kinetics of loss of double minute chromosomes (Fig. 6B). One could calculate in this experiment that for every minute chromosome lost there was a decrease of about 80 U of DHFR activity.

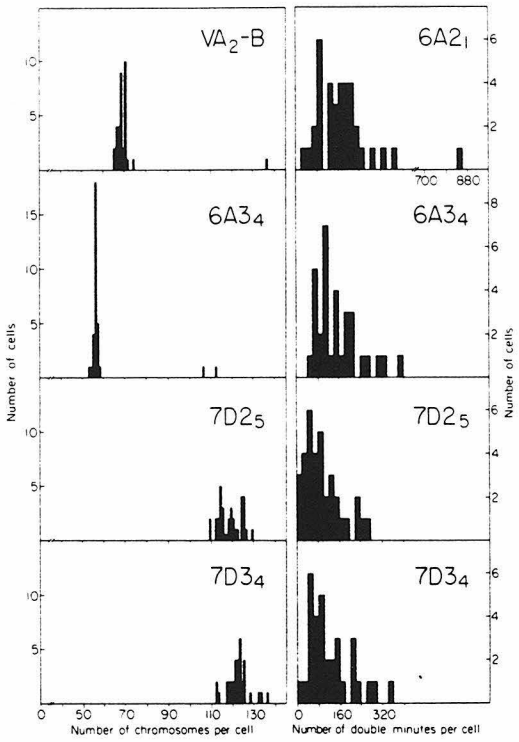



FIG. 4. Frequency distribution of number of chromosomes and number of double minutes per cell in the parental line $VA_2$-B and its MTX-resistant derivatives 6A2, 6A3, 7D2, and 7D3.

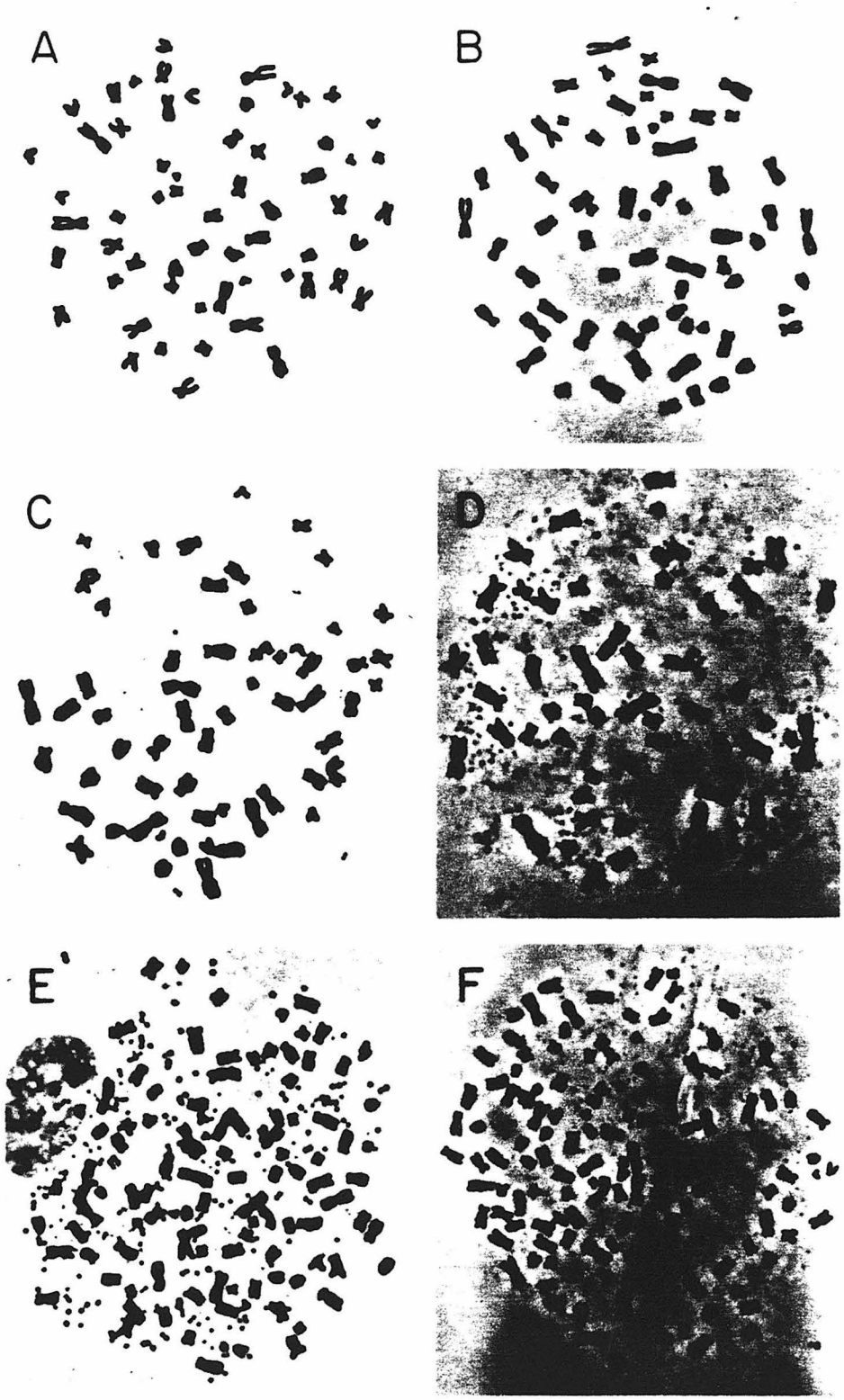



FIG. 5. Representative orcein-stained metaphase spreads from the BU25 (−BUdR) (A) and $VA_2$-B (B) parental cell lines and from the MTX-resistant variants $10B1_{11}$ (C), $6A3_4$ (D), $10C3_4$ (E), and $7D3_4$ (F).

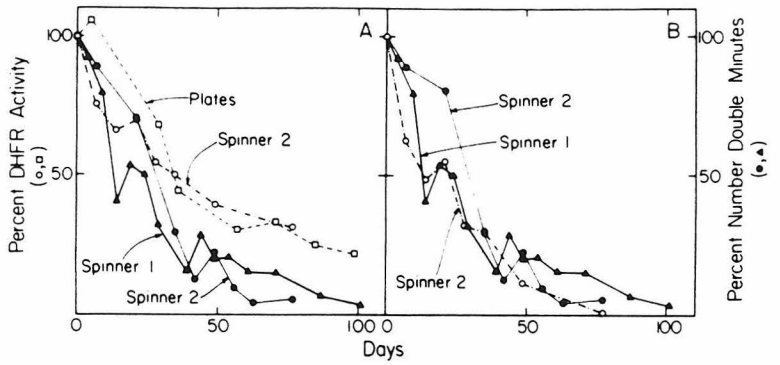



FIG. 6. Quantitative behavior of the DHFR activity and number of double minute chromosomes per cell in the variant 6A3 after removal of the selective pressure. (A) Original data pertaining to cells grown in two spinner cultures and on plates. (B) DHFR data for one spinner culture, corrected by subtracting the contribution of the relatively stable DHFR fraction. See text for details.

## DISCUSSION

In the present work, all the human cell variants isolated for their resistance to $1.8 \times 10^{-4}$ M DL-MTX exhibited a greatly increased DHFR activity. Work reported elsewhere (C. Morandi, J. Masters, M. Mottes, and G. Attardi, J. Mol. Biol., in press) has revealed a large increase in the amount of DHFR-specific mRNA in the several variants which have been analyzed, derived from both the $VA_2$-B and the HeLa BU25 parental lines. Furthermore, a genomic analysis of the 6A3, 7D3, and 10B3 cell lines, using probes derived from cloned DHFR cDNA, has shown a large increase in DHFR-specific DNA sequences as compared to the parental lines (unpublished observations). Therefore, it seems reasonable to conclude that a DHFR gene amplification has occurred in the human variants investigated here.

Although the resistance to high levels of MTX and the underlying basic mechanism appeared to be a common denominator of all of the variants investigated here, a striking diversity in growth behavior, relative level of DHFR activity or DHFR protein, chromosomal constitution, and degree of instability of the phenotype in the absence of selective pressure was observed among the various resistant cell lines analyzed in the first 10 months after their isolation and even among the sublines deriving from the same original MTX-resistant clones. Another aspect of the variability observed in the properties of the human MTX-resistant cell lines was the dynamic character of the changes detected, which was most obvious in their growth behavior and in their chromosome constitution.

In the cell lines 10C2 and 10C3, the detectable amount of DHFR protein was reproducibly much lower than that found in other variants exhibiting a similar increase in DHFR activity; this observation may indicate the occurrence in these variants of a structural gene mutation resulting in a protein product with a reduced affinity for MTX and an increased specific activity.

No completely stable MTX-resistant human variant has been found so far, in contrast to what has been reported for mouse and hamster cell lines. Although we have only limited data on this point, it does not seem that the above-mentioned difference merely reflects the fact the variants analyzed here represent relatively recent isolates. In fact, two of the variants (6A2 and 7D2), which were examined after 8 to 10 months of growth in the presence of $1.8 \times 10^{-4}$ M DL-MTX, exhibited an instability of phenotype comparable to that of the other variants. In some of the mouse MTX-resistant cell lines previously investigated by others, stability of the phenotype has been shown indeed to arise during prolonged growth in selective medium (17); however, other mouse and Chinese hamster cell lines have exhibited a stable resistance to the drug very early after their isolation (10, 26).

The most striking alterations observed in the variant cell lines analyzed here are the anomalies found in their chromosome constitution. These anomalies involved a change in chromosome number and the presence of double minute chromosomes. A third type of alteration, i.e., the appearance in one or more identifiable chromosomes of some variants of a homogeneously staining region (HSR) containing DHFR genes, analogous to that described in Chinese hamster and mouse MTX-resistant cell lines with stably amplified DHFR genes (7, 10, 26), will be described elsewhere (H. Gay, B. Maurer, and G. Attardi, manuscript in preparation).

The occurrence of a duplicated set of chromosomes in most cells of the population, as has been observed here in some of the MTX-resis-

tant variants, has not been reported previously in MTX-resistant cell lines. It seems likely that this chromosome duplication event, which occurred during the development of the resistance to MTX, conferred an advantage upon the cells undergoing it by rapidly doubling the DHFR gene complement and therefore, presumably, the cell content in DHFR. Consistent with this interpretation is the observation that, in variants with a duplicated set of chromosomes, the chromosome(s) containing an HSR is present in duplicate (Gay et al., in preparation).

All MTX-resistant variants investigated here exhibited the presence of double minute chromosomes. Such small, acentromeric elements have been described previously in many experimental and human tumors (for references, see references 3 and 21), where they are relatively stable in the absence of any known selective pressure. Recently, it has been reported that the double minutes from a mouse adrenocortical tumor contain amplified DNA sequences (14). Furthermore, the presence of double minutes has been described in several MTX-resistant mouse cell lines, and evidence has been presented strongly suggesting that these elements contain amplified DHFR genes (26). In the present work, the observation of a parallel loss of double minute chromosomes and DHFR activity in the variant 6A3 is likewise consistent with the idea that, in this variant, such elements contain DHFR genes. Preliminary in situ hybridization experiments with a high-specific-activity probe derived from cloned DHFR cDNA have provided direct evidence of the presence of DHFR-coding sequences in double minute chromosomes from several MTX-resistant human variants (B. Maurer, unpublished observations). We do not know whether and to what extent the differences in size and stainability with orcein of the double minute chromosomes observed in cells of different human variants and among cells of the same variant and within the same metaphase can be accounted for by a variability in chromosome condensation. However, the reproducibility of the differences detected in the appearance of double minutes in the various resistant cell lines suggests the possibility that the amount of DNA can vary among different minutes. Such differences in DNA content may conceivably reflect a variation in the size or number (or both) of the amplified units contained in each minute. On the other hand, the observation that the double minute chromosomes vary greatly in number in different variants having a comparable degree of DHFR gene amplification (as estimated from the increase in DHFR activity [present work] and the increase in DHFR mRNA and DHFR specific sequences in the genome [unpublished observations]) with no obvious inverse relationship between their number and their size or stainability suggests that there must be another set of DHFR genes, representing a varied portion of the total, which is associated with the regular chromosomes. Recent observations indicate that, in some, but not all, variants, these chromosomal genes are located in HSRs associated with identifiable chromosomes (Gay et al., in preparation).

In previous studies on mouse MTX-resistant cell lines, the instability of the resistance in the absence of selective pressure has been found to be associated with the presence of double minute chromosomes; furthermore, in subpopulations of these unstable cell lines, a rough proportionality has been observed between the number of double minutes and the number of amplified DHFR genes (26). More recently, in uptake mutants of neuroblastoma cells, unstable drug resistance has also been correlated with the presence of double minute chromosomes (4). It has been hypothesized that, in the above-mentioned cell lines, unequal distribution of the double minutes at cell division and selective growth advantage for cells with lower numbers of these elements would account for the instability of the phenotype (17). The observations made in the present work may point to a more complicated situation in the human MTX-resistant variants. In fact, the average number of identifiable double minutes per cell in these variants has been found to be totally unrelated to the rate of loss of the excess DHFR activity in the absence of MTX. Our data suggest that instability of MTX resistance may also be associated with a chromosomal localization of the amplified units, and depend on a particular susceptibility of these chromosomal genes to be lost. Preliminary observations indicate that the presence of these unstable chromosomal genes in the human MTX-resistant variants is not obligatorily correlated with the occurrence of morphologically recognizable HSRs. An instability of MTX resistance associated with a reduction in size of the HSR present in an identifiable chromosome has been reported in a hamster cell line (J. L. Biedler and B. A. Spengler, J. Cell Biol. **70**:117a, 1976).

ACKNOWLEDGMENTS

These investigations were supported by Public Health Service grants GM-11726 and T32 GM-07616 from the National Institutes of Health.

The HeLa BU25 cell line was generously provided by Saul Kit (Baylor University, Houston, Tex.). We are very grateful to Anne Chomyn, Carlo Morandi, Claus Doersen, and Barry Maurer for valuable discussions. The technical assistance of Arger Drew is gratefully acknowledged.

LITERATURE CITED

1. **Alt, F. W., R. E. Kellems, J. R. Bertino, and R. T. Schimke.** 1978. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cul-

tured murine cells. J. Biol. Chem. **253**:1357–1370.

2. **Alt, F. W., R. E. Kellems, and R. T. Schimke.** 1976. Synthesis and degradation of folate reductase in sensitive and methotrexate-resistant lines of S-180 cells. J. Biol. Chem. **251**:3063–3074.
3. **Barker, P. E., and T. C. Hsu.** 1979. Double minutes in human carcinoma cell lines, with special reference to breast tumors. J. Natl. Cancer Inst. **62**:257–261.
4. **Baskin, R., R. N. Rosenberg, and V. Dev.** 1981. Correlation of double-minute chromosomes with unstable multidrug cross-resistance in uptake mutants of neuroblastoma cells. Proc. Natl. Acad. Sci. U.S.A. **78**:3654–3658.
5. **Beach, L. R., and R. D. Palmiter.** 1981. Amplification of the metallothioein-I gene in cadmium-resistant mouse cells. Proc. Natl. Acad. Sci. U.S.A. **78**:2110–2114.
6. **Biedler, J. L., A. M. Albrecht, D. J. Hutchison, and B. A. Spengler.** 1972. Drug response, dihydrofolate reductase, and cytogenetics of amethopterin-resistant Chinese hamster cells *in vitro*. Cancer Res. **32**:153–161.
7. **Biedler, J. L., and B. A. Spengler.** 1976. Metaphase chromosome anomaly: association with drug resistance and cell specific products. Science **91**:185–187.
8. **Bostock, C. J., E. M. Clark, N. G. L. Harding, P. M. Mounts, C. Tyler-Smith, V. Van Heyningen, and P. M. P. Walker.** 1979. The development of resistance to methotrexate in a mouse melanoma cell line. I. Characterization of the dihydrofolate reductases and chromosomes in sensitive and resistant cells. Chromosoma **74**:153–177.
9. **Chang, S. E., and J. W. Littlefield.** 1976. Elevated dihydrofolate reductase messenger RNA levels in methotrexate-resistant BHK cells. Cell **7**:391–396.
10. **Dolnick, B. J., R. J. Berenson, J. R. Bertino, R. J. Kaufman, J. H. Nunberg, and R. T. Schimke.** 1978. Correlation of dihydrofolate reductase elevation with gene amplification in a homogeneously staining chromosomal region of L5178Y cells. J. Cell Biol. **83**:394–402.
11. **Fischer, G. A.** 1961. Increased levels of folic acid reductase as a mechanism of resistance to amethopterin in leukemia cells. Biochem. Pharmacol. **7**:75–77.
12. **Flintoff, W. F., S. V. Davidson, and L. Siminovitch.** 1976. Isolation and partial characterization of three methotrexate-resistant phenotypes from Chinese hamster ovary cells. Somatic Cell Genet. **2**:245–261.
13. **Frearson, P. M., S. Kit, and D. R. Dubbs.** 1966. Induction of dihydrofolate reductase activity by SV40 and polyoma virus. Cancer Res. **26**:1653–1660.
14. **George, D. L., and V. E. Powers.** 1981. Cloning of DNA of double minutes of Y1 mouse adrenocortical tumor cells: evidence for gene amplification. Cell **24**:117–123.
15. **Hakala, M. T., S. F. Zakrewski, and C. A. Nichol.** 1961. Relation of folic acid reductase to amethopterin resistance in cultured mammalian cells. J. Biol. Chem. **236**:952–958.
16. **Hängi, U. J., and J. W. Littlefield.** 1976. Altered regulation of the rate of synthesis of dihydrofolate reductase in methotrexate-resistant hamster cells. J. Biol. Chem. **251**:3075–3080.
17. **Kaufman, R. J., P. C. Brown, and R. T. Schimke.** 1979. Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes. Proc. Natl. Acad. Sci. U.S.A. **76**:5669–5673.
18. **Kellems, R. E., F. W. Alt, and R. T. Schimke.** 1976. Regulation of folate reductase synthesis in sensitive and methotrexate-resistant sarcoma S-180 cells. In vitro translation and characterization of folate reductase mRNA. J. Biol. Chem. **251**:6987–6993.
19. **Kit, S., D. R. Dubbs, and P. M. Frearson.** 1966. HeLa cells resistant to bromodeoxyuridine and deficient in thymidine kinase activity. Int. J. Cancer **1**:19–30.
20. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
21. **Levan, A., G. Levan, and F. Milleman.** 1977. Chromosomes and cancer. Hereditas **86**:15–30.
22. **Littlefield, J. W.** 1969. Hybridization of hamster cells with high and low folate reductase activity. Proc. Natl. Acad. Sci. U.S.A. **62**:88–95.
23. **Melera, P. W., J. A. Lewis, J. L. Biedler, and C. Hession.** 1980. Antifolate-resistant Chinese hamster cells. Evidence for dihydrofolate reductase gene amplification among independently derived sublines overproducing different dihydrofolate reductases. J. Biol. Chem. **255**:7024–7028.
24. **Mitchell, C. H., and G. Attardi.** 1978. Cytoplasmic transfer of chloramphenicol resistance in a human cell line. Somatic Cell Genet. **4**:737–744.
25. **Morandi, C., and G. Attardi.** 1981. Isolation and characterization of dihydrofolic acid reductase from methotrexate sensitive and resistant human cell lines. J. Biol. Chem. **256**:10169–10175.
26. **Nunberg, I. H., R. J. Kaufman, R. T. Schimke, G. Urlaub, and L. A. Chasin.** 1978. Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. Proc. Natl. Acad. Sci. U.S.A. **75**:5553–5556.
27. **Perkins, J. P., B. L. Hillcoat, and J. R. Bertino.** 1967. Dihydrofolate reductase from a resistant subline of the L12110 lymphoma. J. Biol. Chem. **242**:4771–4776.
28. **Pontén, J., F. Jensen, and H. Koprowski.** 1963. Morphological and virological investigation of human tissue cultures transformed with SV40. J. Cell. Comp. Physiol. **61**:145–154.
29. **Wahl, G. M., R. A. Padgett, and G. R. Stark.** 1979. Gene amplification causes overproduction of the first three enzymes of UMP synthesis in N-(phosphoacetyl)-L-aspartate-resistant hamster cells. J. Biol. Chem. **254**:8679–8689.
30. **Wiedemann, L. M., and L. F. Johnson.** 1979. Regulation of dihydrofolate reductase synthesis in an overproducing 3T6 cell line during transition from resting to growing state. Proc. Natl. Acad. Sci. U.S.A. **76**:2818–2822.

# CHAPTER 3

## Multiple Forms of Human Dihydrofolate Reductase Messenger RNA: Cloning and Expression in *Escherichia coli* of their DNA Coding Sequence

*J. Mol. Biol.* (1982) **156**, 583–607

# Multiple Forms of Human Dihydrofolate Reductase Messenger RNA

## Cloning and Expression in *Escherichia coli* of their DNA Coding Sequence

CARLO MORANDI†, JEFFREY N. MASTERS, MONICA MOTTES† AND GIUSEPPE ATTARDI

*Division of Biology*
*California Institute of Technology*
*Pasadena, CA 91125, U.S.A.*

*(Received 20 October 1981)*

The programming capacity for the synthesis of human dihydrofolic acid reductase in a rabbit reticulocyte lysate has been found to be greatly enhanced in the polysomal poly(A)-containing RNA from a methotrexate-resistant human cell variant (6A3), as compared to the RNA from its parental line ($VA_2$-B). A major fraction of this promoting activity is associated with a $3{\cdot}8 \times 10^3$ base RNA species detectable as a band in the ethidium bromide-stained electrophoretic pattern of the RNA from 6A3 cells, but not in the RNA from $VA_2$-B cells. Furthermore, sucrose gradient fractionation experiments have indicated that another substantial portion of the messenger activity is associated with RNA components around $10^3$ bases in size. Double-stranded complementary DNA synthesized from total poly(A)-containing RNA of 6A3 cells has been size fractionated, and both large (1400 to 3800 base-pairs) and small size complementary DNA (600 to 1400 base-pairs) species have been used separately to transform *Escherichia coli* $\chi$2282 with pBR322 as a vector. Of 76 transformants obtained with the large size complementary DNA, identified by a differential colony hybridization assay, none has expressed the dihydrofolic acid reductase coding sequence in *E. coli*, as judged by resistance to trimethoprim. By contrast, eight trimethoprim-resistant transformants have been obtained using the small size complementary DNA, and their plasmids have been shown to contain the dihydrofolic acid reductase coding sequence by restriction mapping and DNA sequencing; moreover, immunoautoradiographic experiments have revealed the presence in the extracts of two of these transformants of a protein with the electrophoretic mobility and immunoreactivity of human dihydrofolic acid reductase. Restriction mapping and DNA transfer hybridization experiments have further indicated that the inserts of the chimaeric plasmids conferring trimethoprim resistance upon the host and of those lacking this capacity cover together a complementary DNA region of about $3{\cdot}35 \times 10^3$ base-pairs, in which the 564 base-pair dihydrofolic acid reductase coding stretch is located near the 5′ end of the sense strand. RNA transfer hybridization experiments using different cloned complementary DNA fragments as probes have shown the presence of three species of dihydrofolic acid reductase-specific messenger RNAs, with sizes of $3{\cdot}8 \times 10^3$, $1{\cdot}0 \times 10^3$ and $0{\cdot}8 \times 10^3$ bases, differing in the length of the 3′

† Present address: Istituto di Genetica, Università di Pavia, 27100 Pavia, Italy.

0022–2836/82/110583–25 $03.00/0

untranslated region, in the poly(A)-containing RNA from two methotrexate-resistant variants, 6A3 and 10B3, and, in greatly reduced amounts, in the RNA from their respective parents, $VA_2B$ and HeLa BU25.

## 1. Introduction

In recent years, several mammalian cultured cell systems have been described that have developed resistance to high concentrations of the folic acid analog methotrexate as a result of exposure to gradually increasing concentrations of the drug (Hakala *et al.*, 1961; Littlefield, 1969; Biedler *et al.*, 1972; Flintoff *et al.*, 1976; Alt *et al.*, 1976; Nunberg *et al.*, 1978; Bostock *et al.*, 1979). The most common mechanism of this resistance has been shown to be a selective amplification of the structural gene for dihydrofolic acid reductase, the target enzyme of the drug (Alt *et al.*, 1978; Nunberg *et al.*, 1978; Dolnick *et al.*, 1979; Melera *et al.*, 1980), with resulting overproduction of the DHFRase† messenger RNA (Chang & Littlefield, 1976; Alt *et al.*, 1978; Melera *et al.*, 1980) and of the enzyme (Hakala *et al.*, 1961; Littlefield, 1969; Biedler *et al.*, 1972; Flintoff *et al.*, 1976; Alt *et al.*, 1976; Nunberg *et al.*, 1978). The content of the latter in the above-mentioned methotrexate-resistant cell lines is increased up to several hundred-fold, reaching levels exceeding the amount inactivated by the drug.

The recent availability of cloned DHFRase specific cDNA from methotrexate-resistant mouse cells (Chang *et al.*, 1978) has allowed a detailed analysis of the mouse DHFRase mRNA (Setzer *et al.*, 1980). One intriguing result of this analysis has been the demonstration of the occurrence of four functional DHFRase mRNAs, ranging in size from 750 to 1600 nt, in cultured mouse cells with or without amplified DHFRase genes and in mouse liver. These mRNAs appear to differ primarily in the size of the 3′ untranslated region. Multiple DHFRase mRNAs have also been detected in Chinese hamster ovary cells resistant to methotrexate (Setzer *et al.*, 1980). The origin and significance of this multiplicity of forms of DHFRase mRNA are unknown.

We have isolated and partially characterized several methotrexate-resistant cell variants derived from two different human lines, HeLa BU25 and $VA_2$-B (Masters *et al.*, 1982). In these methotrexate-resistant variants, the DHFRase has been shown to be increased to levels comparable to those observed in methotrexate-resistant mouse and Chinese hamster cell lines. In the work reported here, a 3·8 kb RNA component, i.e. a species about seven times as long as the DHFRase coding stretch, has been detected in the ethidium bromide-stained electrophoretic pattern of the polysomal poly(A)-containing RNA from these cells, but not in the pattern of RNA from the parental cells. This RNA has been functionally characterized as DHFRase mRNA; furthermore, the occurrence of smaller size DHFRase-specific mRNAs, besides the 3·8 kb component, has been demonstrated in a methotrexate-resistant variant by translation assays *in vitro*. cDNA copies of these mRNAs have been cloned in *Escherichia coli*, and the sequence coding for the human DHFRase

† Abbreviations used: DHFRase, dihydrofolic acid reductase; cDNA, complementary DNA; kb, $10^3$ bases or base-pairs; bp, base-pairs; nt, nucleotides; IgG, immunoglobulin G; ds-, double-stranded; ss-, single-stranded.

shown to be expressed in the bacterial host. By using these cloned cDNA probes, it has been found that there are at least three different forms of human DHFRase-specific mRNAs, which have in common the polypeptide coding sequence near their 5′ end, and differ in the length of the 3′ untranslated region.

## 2. Materials and Methods

### (a) *Materials*

DL-Methotrexate (consisting of ~50% L-form) and trimethoprim were from Sigma, oligo(dT)$_{12-18}$ from Miles Lab. Inc., oligo(dT)-cellulose (T3) from Collaborative Research, methylmercuric hydroxide from Alpha products, restriction enzymes from New England Biolabs or from Bethesda Research Laboratories, *Aspergillus orizae* $S_1$ nuclease from Sigma, *E. coli* DNA polymerase I from Boehringer-Mannheim, calf thymus terminal transferase from PL Biochemicals, reticulocyte lysate from Bethesda Research Laboratories, Pansorbin from Calbiochem., and Seaplaque Agarose from Marine Colloids.

### (b) *Cell lines and methods of growth*

The human cell line $VA_2$-B, an azaguanine-resistant subclone of the simian virus 40 (SV40) transformed line WI-18-$VA_2$ (Pontén *et al.*, 1963) was grown in suspension in Eagle's phosphate medium supplemented with 3 μg 8-azaguanine/ml and 5% (v/v) calf serum. The human cell line HeLa BU25, a mutant subline of HeLa S3 deficient in thymidine kinase activity (Kit *et al.*, 1966), kindly provided by Dr S. Kit, was grown in suspension in Eagle's phosphate medium supplemented with 10% (v/v) calf serum. The methotrexate-resistant variants of the $VA_2$-B cell line (6A3) and of the BU25 cell line (10B3, 10C1, 10C2 and 10C3) were grown in suspension in the media described above, supplemented with 10% (v/v) extensively dialyzed calf serum and $1{\cdot}8 \times 10^{-4}$ M-DL-methotrexate.

### (c) *Extraction and analysis of cytoplasmic RNA*

Cytoplasmic polysomal RNA was prepared essentially as described by Padgett *et al.* (1979). The polysome pellet, separated from the 12,000 ***g*** supernatant of the cell lysate by centrifugation through a 1 M-sucrose cushion, was resuspended in 0·01 M-Tris (pH 7·0), 0·1 M-NaCl, 0·001 M-EDTA, 0·5% (w/v) sodium dodecyl sulfate (0·6 to 1 ml per gram of cells). After incubation for 1 h at room temperature with 100 μg of self-digested Pronase/ml, the RNA was extracted 3 times with phenol/chloroform/isoamyl alcohol (25 : 24 : 1, by vol.). Following precipitation with ethanol, the Pronase/phenol extraction procedure was repeated. Poly(A)-containing RNA was isolated by passage through an oligo(dT)-cellulose column according to Amalric *et al.* (1978).

Fractionation of the poly(A)-containing RNA by electrophoresis through a 1·2% (w/v) agarose/$CH_3HgOH$ slab gel was carried out as described by Bailey & Davidson (1976), except that the concentration of $CH_3HgOH$ in the gel was 20 mM and the electrophoresis buffer was 50 mM-sodium borate (pH 8·2), 0·1 mM-EDTA. The gels were stained with ethidium bromide (1 μg/ml in 0·5 M-ammonium acetate). For the isolation from the gel of an RNA cut containing the 3·8 kb DHFRase mRNA, the method of Wieslander (1979), utilizing low melting temperature Seaplaque agarose for the preparation of the gel and extraction of the RNA with phenol from the melted agarose, was followed.

For preparative fractionation in a sucrose gradient under denaturing conditions, 200 μg of poly(A)-containing RNA were denatured with 20 mM-$CH_3HgOH$ in a 500-μl volume and run through a 5% to 20% (w/v) sucrose gradient in 1 mM-Tris (pH 6·7), 1 mM-NaCl, 1 mM-EDTA (Cantatore & Attardi, 1980) in a Beckman SW41 rotor for 13 h at 35,000 revs/min at 2°C.

(d) *Preparation of antibodies against human DHFRase*

The human DHFRase was purified from methotrexate-resistant $VA_2$-B 6A3 cells as described by Morandi & Attardi (1981). Two rabbits were injected subcutaneously in the back 4 times, at 1 month intervals, with 1 mg of pure enzyme in 50 mM-potassium phosphate (pH 7·0), emulsified with an equal volume of Freund's complete adjuvant. Blood (40 ml) was collected from each rabbit every 2 weeks for 4 months, starting 4 weeks after the 4th injection. The IgG fraction was purified as described by Palmiter *et al.* (1971); the DHFRase-specific IgG content was estimated to be about 3% of the total IgG from the precipitin curve (Kabat & Mayer, 1948).

(e) *Translation* in vitro

The rabbit reticulocyte *in vitro* translation system was used under the conditions suggested by the vendor, using 10-$\mu$l reaction mixtures and [$^{35}$S]methionine. For the analysis of the sucrose gradient fractions, an equal volume of each fraction was used. For indirect immunoprecipitation (Kessler, 1975), 5 $\mu$l of each reaction mixture were diluted with 95 $\mu$l of NET buffer (50 mM-Tris (pH 7·4), 150 mM-NaCl, 5 mM-EDTA, 0·25% (w/v) sodium dodecyl sulfate, 0·75% (v/v) Triton X-100) containing 2 mM-methionine and 100 $\mu$g ovalbumin/ml. Then 5 $\mu$l of pre-immune serum were added and the samples left overnight at 4°C. After addition of 30 $\mu$l of 10% (v/v) Pansorbin (fixed *Staphylococcus aureus* cells), the mixtures were left on ice for 20 min and centrifuged for 30 s in an Eppendorf microfuge. The supernatants were transferred to new tubes, each mixed with 5 $\mu$l of anti-DHFRase antiserum (see below), incubated at 4°C for 3 h and then treated with Pansorbin as before. The Pansorbin pellets were washed 3 times with NET buffer, and the proteins eluted with 0·063 M-Tris (pH 6·8), 0·002 M-methionine, 2% (w/v) sodium dodecyl sulfate at room temperature for 1 h. After centrifuging the insoluble residues, the supernatants were removed and electrophoresed using the Laemmli sodium dodecyl sulfate/polyacrylamide gel system (Laemmli, 1970), with 15% (w/v) polyacrylamide in the separating gel and 5% (w/v) polyacrylamide in the stacking gel (acrylamide to bisacrylamide, 30 : 0·8 (w/w) in both gels). The gels were treated for fluorography according to Laskey (1980).

(f) *Cloning of double-stranded cDNA*

Double-stranded cDNA was synthesized following published procedures (Efstratiadis & Villa-Komaroff, 1979; Friedman & Rosbach, 1977; Buell *et al.*, 1978). For single-stranded cDNA synthesis, vanadyl ribonucleoside complex (BRL) was included in the reaction mixture at a concentration of 4 mM as an RNAase inhibitor (Berger & Birkenmeier, 1979; Rice & Strauss, 1981). The concentration of poly(A)-containing RNA was 500 $\mu$g/ml, and the ratio of 5 units of avian myeloblastosis virus reverse transcriptase per $\mu$g of RNA was found to be optimal for the synthesis of large cDNA.

The double-stranded cDNA was fractionated on a 5% to 20% linear sucrose gradient in 10 mM-Tris·HCl (pH 8·0), 300 mM-sodium acetate, 1 mM-EDTA, by centrifugation in a Spinco SW41 rotor at 37,000 revs/min for 14 h at 19°C.

The pBR322 DNA was linearized by *Pst*I, electrophoresed through an agarose gel in 40 mM-Tris·HCl (pH 7·8), 5 mM-sodium acetate, 1 mM-EDTA, eluted according to Yang *et al.* (1979), and then dG-tailed according to Roychoudhury & Wu (1980); an average of about 13 dGMP molecules was added to each 3′ end. The double-stranded cDNA was dC-tailed in the same way, with 9 to 16 dCMP molecules being added to each 3′ end. Equimolar amounts of tailed cDNA and tailed pBR322 were then annealed according to Steinmetz *et al.* (1981).

(g) *Transformation of* E. coli

*E. coli* strains $\chi$2282, a thy$^+$ variant of $\chi$1776 (Chang *et al.*, 1978), and HB101 (Boyer & Roulland-Dussoix, 1969) were used with the indicated plasmids. All work was done in

accordance with the National Institutes of Health Guidelines (1980) for recombinant DNA research. $\chi$2282 *E. coli* cells were transformed with pBR322 containing human DHFRase cDNA inserts according to a procedure described by D. Hanahan (personal communication). The recombinant colonies, which were obtained at a frequency of approximately 10 per ng of DNA, were replica plated in duplicate onto nitrocellulose filters, and those containing DHFRase sequences were identified by differential hybridization screening (St. John & Davis, 1979; Wahl *et al.*, 1979). For this purpose, twin replica filters were incubated with $5 \times 10^5$ cts/min per filter of $^{32}$P-labeled cDNA prepared from total poly(A)-containing RNA of either $VA_2$-B or 6A3 cells, for 20 h at 42°C in 50% (v/v) formamide, 50 mM-sodium phosphate (pH 6·8), 5 × SSC (SSC is 0·15 M-NaCl, 0·015 M-sodium citrate, pH 7), 10% (w/v) dextran sulfate, 1 × Denhardt's solution (0·02% (w/v) Ficoll, 0·02% (w/v) polyvinylpyrrolidone, 0·02% (w/v) bovine serum albumin (Denhardt, 1966)), 0·1% (w/v) sodium dodecyl sulfate and 100 $\mu$g of heat-denatured, sonicated salmon sperm DNA/ml. The filters were rinsed once for 20 min at 42°C in hybridization buffer, then 3 times for 30 min at 68°C in 2 × SSC, 0·1% (w/v) sodium dodecyl sulfate, and once for 30 min at 68°C in 0·5 × SSC. For direct phenotypic selection of clones that synthesize human DHFRase, the tetracycline-resistant transformants were replica plated on agar plates in M9 medium supplemented with diaminopimelic acid (100 $\mu$g/ml), tetracycline (10 $\mu$g/ml) and trimethoprim (4 $\mu$g/ml). HB101 cells were transformed according to the procedure described by Cohen *et al.* (1972).

### (h) *Plasmid DNA preparation*

Plasmid-containing *E. coli* $\chi$2282 were grown to saturation in LB broth, supplemented with 100 $\mu$g diaminopimelic acid/ml and 15 $\mu$g tetracycline/ml at 37°C, under vigorous shaking. Cells were harvested and processed as described by Haseltine *et al.* (1980), with the modifications that the cleared Triton X-100 bacterial lysate was incubated at 65°C for 15 min and the denatured proteins pelleted, and subsequently the macromolecular components of the supernatant were concentrated by precipitation with 10% (w/v) polyethylene glycol 6000 for at least 2 h in ice before the CsCl/EthBr density-gradient centrifugation.

Plasmids harbored by *E. coli* HB101 were prepared in the same way, except that plasmid duplication during growth was induced by chloramphenicol (Clewell, 1972). For rapid plasmid DNA isolation, the tetracycline-resistant plasmids were grown in 10 ml of LB broth culture, as described by Davis *et al.* (1980).

### (i) *Restriction enzyme analysis*

Plasmid DNA was treated with restriction enzymes under conditions recommended by the vendors and the products were electrophoresed on either agarose gels using Tris-acetate/EDTA buffer (McDonnell *et al.*, 1977) or polyacrylamide gels using Tris-borate/EDTA buffer (Maniatis *et al.*, 1975*b*), and stained with ethidium bromide. DNA restriction fragments were isolated from slab gels by the electroelution method of Yang *et al.* (1979).

### (j) *Preparation of radioactive probes*

$^{32}$P-labeled, random primed cDNA was prepared from total poly(A)-containing RNA of $VA_2$-B or 6A3 cells according to Mullins *et al.* (1980), using 100 ng of RNA, 50 $\mu$Ci of [$\alpha$-$^{32}$P]dCTP (800 Ci/mmol) and the other unlabeled dNTPs at 100 $\mu$M, 15 units of reverse transcriptase and 4 $\mu$g of calf thymus primer DNA, kindly provided by Dr J. Casey (Dept. of Biochemistry, Louisiana State University Medical Center), in a reaction volume of 25 $\mu$l. Total plasmid DNA and isolated restriction fragments were nick-translated according to Maniatis *et al.* (1975*a*) in 25-$\mu$l reaction volumes containing 100 ng of DNA, 1 $\mu$g DNAase I/ml, 2 units of *E. coli* DNA polymerase I, 25 $\mu$Ci of [$\alpha$-$^{32}$P]dCTP (2000 to 3000 Ci/mmol) and the other unlabeled dNTPs at 20 $\mu$M.

(k) *RNA and DNA transfer techniques*

RNA transfer to nitrocellulose was done according to Thomas (1980), using the formaldehyde gel system of Rave *et al.* (1979): 5 $\mu$g of RNA per sample were used. Quantitative transfer of RNAs as large as 28 S ribosomal RNA was accomplished without treating the gel before transfer. Hybridization with probes $^{32}$P-labeled by nick translation was done in 50% (v/v) formamide, 0·02 M-sodium phosphate (pH 6·8), 5 × SSC, 0·1% (w/v) sodium dodecyl sulfate, 10% (w/v) dextran sulfate, 10 × Denhardt's solution at 42°C for 10 to 15 h. The filters were then washed for 1 h at 42°C with 5 × SSC, 0·1% sodium dodecyl sulfate, 10 × Denhardt's solution, and then at 68°C for 2 h with 2 × SSC, 0·1% sodium dodecyl sulfate, and for 2 h with 0·5 × SSC, with 2 buffer changes in each case.

(l) *DNA sequencing*

Sequencing reactions were carried out as described by Maxam & Gilbert (1980). The pHD84 plasmid was cut with *Eco*RI at the single pBR322 site and at the single site in the insert located at 137 nt from the closer *Pst*I site. After fill-in labeling the 3′ ends with [$\alpha$-$^{32}$P]dNTPs and the Klenow fragment of *E. coli* DNA polymerase I or labeling the 5′ ends with [$\gamma$-$^{32}$P]ATP and phage T4 polynucleotide kinase subsequently to treatment with bacterial alkaline phosphatase, the labeled fragments were digested with *Pst*I. The 137 nt fragment of the insert, $^{32}$P-labeled at the *Eco*RI-produced end in one or the other strand, was then isolated on a polyacrylamide gel, eluted and subjected to the sequencing reactions.

(m) *Immunoautoradiography of* E. coli *extracts*

*E. coli* $\chi$2282 carrying tetracycline-resistant and trimethoprim-resistant plasmids were grown overnight in 50 ml of M9 medium supplemented with 0·5% (w/v) Casamino acids (DIFCO), 0·2% (w/v) glucose, 20 $\mu$g thiamine/ml, 10 $\mu$g biotin/ml; tetracycline was used at 15 $\mu$g/ml, trimethoprim, when appropriate, at 0·1 mg/ml or 1 mg/ml. The cells were harvested, washed with saline, resuspended in 1·5 ml of 50 mM-potassium phosphate (pH 7·0) and sonicated with a Branson sonifier equipped with a microtip at maximum power setting 3 times for 5 s. After spinning down the cell debris, the supernatant was run through a sodium dodecyl sulfate/12·5% (w/v) polyacrylamide gel (acrylamide to bisacrylamide, 37·5 : 1). Incubation of the slab gel with anti-DHFRase IgG and $^{125}$I-labeled protein A was carried out as described by Granger & Lazarides (1979), using an IgG solution at 175 $\mu$g/ml and ~10 $\mu$Ci of $^{125}$I-labeled protein A. The dried gel was exposed for 1 month at −60°C using a preflashed film and a DuPont Cronex intensifying screen.

## 3. Results

(a) *Detection and assay of DHFRase mRNA* in vitro

Preliminary to the cloning of the DHFRase mRNA, an investigation of the DHFRase mRNA in methotrexate-sensitive and resistant cell lines was carried out. In previous experiments involving sodium dodecyl sulfate/polyacrylamide gel electrophoresis of cell extracts of methotrexate-sensitive and resistant human cell lines, a pronounced band corresponding to DHFRase was observed in the Coomassie blue-stained patterns of methotrexate-resistant, but not in those of the methotrexate-sensitive cell extracts (Morandi & Attardi, 1981; Masters *et al.*, 1982). In order to investigate whether a corresponding accumulation of DHFRase-specific mRNA in the form of discrete band(s) could be detected in the RNA from the resistant variants, cytoplasmic polysomal RNA was isolated from the parental cell lines $VA_2$-B and HeLa Bu25 and from their respective methotrexate-resistant variants 6A3 and 10B3. When the poly(A)-containing RNA

fractions of these preparations were electrophoresed through agarose/$CH_3HgOH$ slab gels and stained with ethidium bromide, a band corresponding to a molecular size of 3·8 kb was seen in the patterns from the two methotrexate-resistant cell lines, but not in those from the parental lines (Fig. 1). This 3·8 kb RNA was also observed in the other variant cell lines (Masters *et al.*, 1982) analyzed so far (10C1, 10C2, 10C3, data not shown).

The total poly(A)-containing RNA from $VA_2$-B and 6A3 cells was used to program a rabbit reticulocyte lysate translation system *in vitro*, and the products were electrophoresed through a sodium dodecyl sulfate/polyacrylamide gel. A band with the electrophoretic mobility of human DHFRase was observed among the

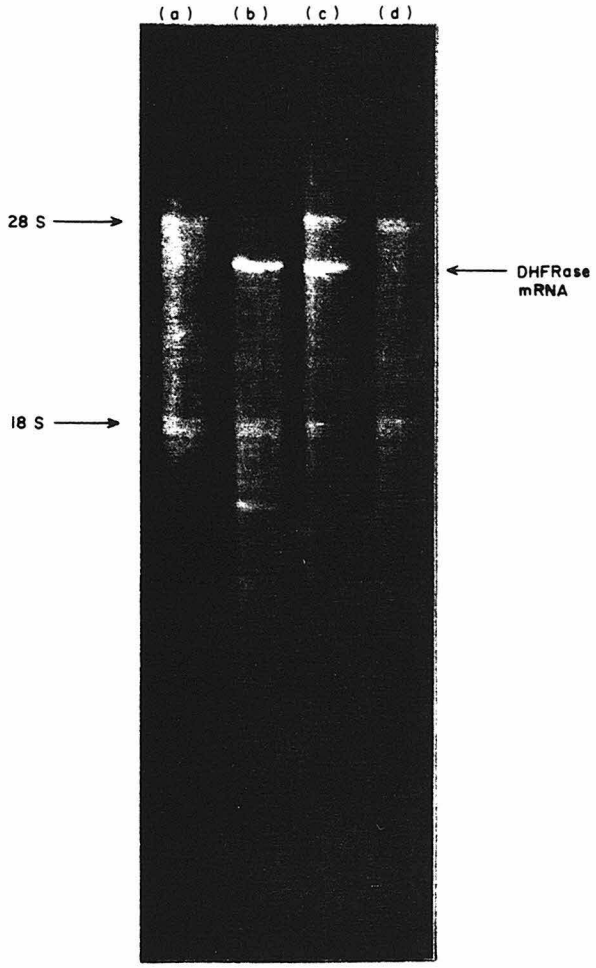



FIG. 1. Detection of the 3·8 kb poly(A)-containing RNA in methotrexate-resistant human cell lines. Samples (5 $\mu$g) of poly(A)-containing RNA from methotrexate-sensitive $VA_2$-B(a) and HeLa Bu25(d) cells and from their respective methotrexate-resistant variants 6A3(b) and 10B3(c) were electrophoresed through an agarose/$CH_3HgOH$ slab gel, and stained with ethidium bromide.

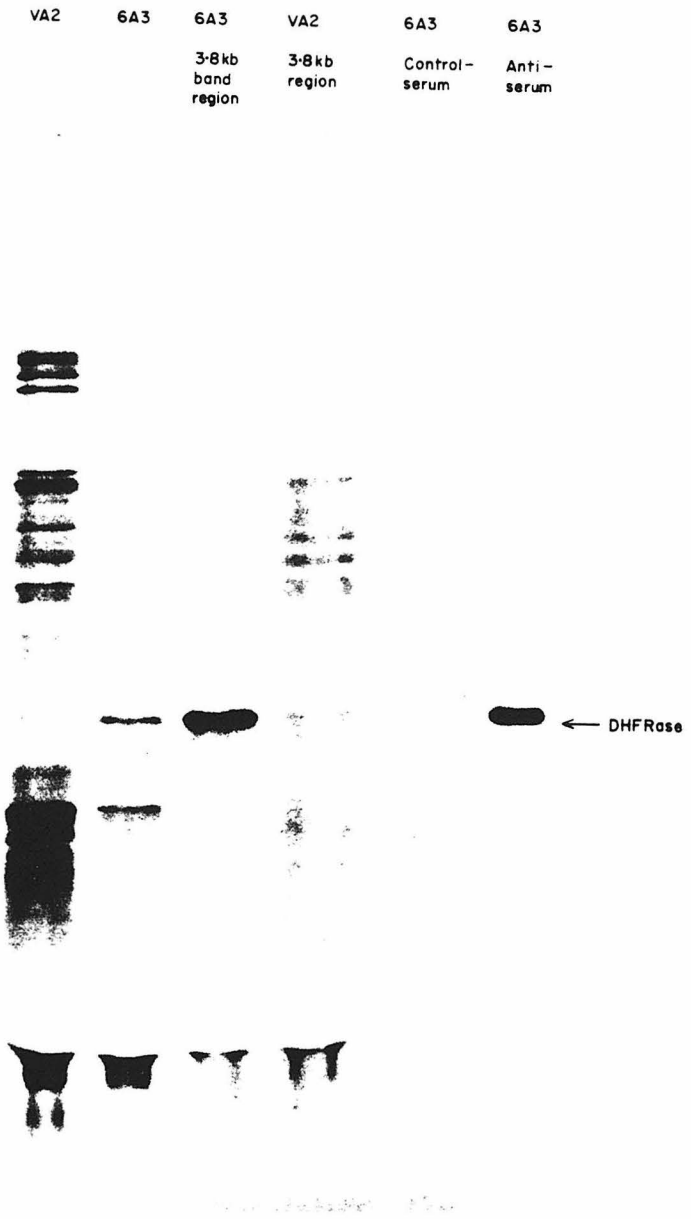



FIG. 2. Synthesis of human DHFRase *in vitro*. Samples of poly(A)-containing RNA were used to program translation reactions *in vitro* in a reticulocyte lysate. the products were electrophoresed through a 15% sodium dodecyl sulfate/polyacrylamide gel and the gel was subjected to fluorography. From left to right. the first 2 lanes show the translation products synthesized under the direction of equal amounts of RNA from $VA_2$-B and 6A3 cells; the next 2 lanes contain the translation products synthesized using equal amounts of RNA recovered from the 3·8 kb region of the 2 lanes of a gel containing the poly(A)-containing RNA of 6A3 and $VA_2$-B cells; the last 2 lanes show the products formed *in vitro* immunoprecipitated indirectly. using pre-immune or anti-human DHFRase serum. from a translation reaction *in vitro* programmed with RNA from 6A3 cells. The translation products obtained *in vitro* without added RNA are similar to those shown in the first lane of Fig. 3 (data not shown).

products synthesized under the direction of the 6A3 RNA, but was not recognizable among the $VA_2$-B RNA-programmed products (Fig. 2). There is a somewhat more slowly migrating band in the latter pattern that probably does not represent DHFRase, and which is, in any event, of a much lower intensity. Indirect immunoprecipitation of the products *in vitro* with antiserum against human DHFRase showed that the specific band in the pattern of 6A3 RNA-programmed products does correspond to human DHFRase (Fig. 2). To test whether the 3·8 kb component in the 6A3 RNA pattern codes for DHFRase, a segment of the gel containing it and a comparable segment of the gel in which the $VA_2$-B RNA had been fractionated were cut out and the RNA was extracted from them according to Wieslander (1979). When the two RNA samples were tested in the translation system *in vitro*, a component with the mobility of DHFRase was found to be synthesized in the 6A3 RNA-directed reaction: a considerably smaller amount of an identically migrating component was detected among the products of the $VA_2$-B RNA directed reaction (Fig. 2).

In order to investigate whether other forms of DHFRase mRNA, besides the 3·8 kb form, occur in the methotrexate-resistant cell lines, the poly(A)-containing polysomal RNA from 6A3 cells was fractionated on a 5% to 20% sucrose gradient in low ionic strength buffer, and alternate or successive fractions, as shown in Figure 3, were tested by the translation assay *in vitro*. Two broad peaks of DHFRase synthesis were observed, one centered around a position corresponding to the 3·8 kb DHFRase mRNA (fractions 9 to 18), and the other, less pronounced, centered around a position corresponding to about 1 kb (fractions 20 to 24). It is clear that the translation products *in vitro* in fractions 13 to 16 are enriched in DHFRase as compared to fractions 22 to 23; from the analysis of less exposed fluorograms, it could be estimated that this enrichment is three- to sixfold. The nature of the RNA with which the latter peak of DHFRase synthesis promoting activity is associated is examined further below.

### (b) *Cloning of DHFRase cDNA*

In previous work by others (Chang *et al.*, 1978) on the cloning in *E. coli* of the DNA sequence coding for the mouse DHFRase, the much lower sensitivity of the mammalian enzyme to the antimetabolite trimethoprim as compared to the bacterial enzyme (Burchall & Hitching, 1965) was exploited to select for transformants that phenotypically expressed the DNA sequence. This approach, besides facilitating the screening of the transformants carrying plasmids with DHFRase cDNA inserts, automatically identified the resistant clones as harboring the DHFRase coding sequence. In the present work, the 3·8 kb form of DHFRase mRNA described above, for its abundance, appeared to be a suitable template for preparing cDNA to be cloned. On the other hand, because of its unusually large size and of the probability that the DHFRase coding sequence would be located near its 5′ end, as in the mouse DHFRase mRNA (Chang *et al.*, 1978), it was anticipated that there would be difficulties in obtaining a complete copy of the DHFRase coding stretch with reverse transcriptase. The smaller size RNA with DHFRase synthesis promoting activity, which had been detected in the poly(A)-containing

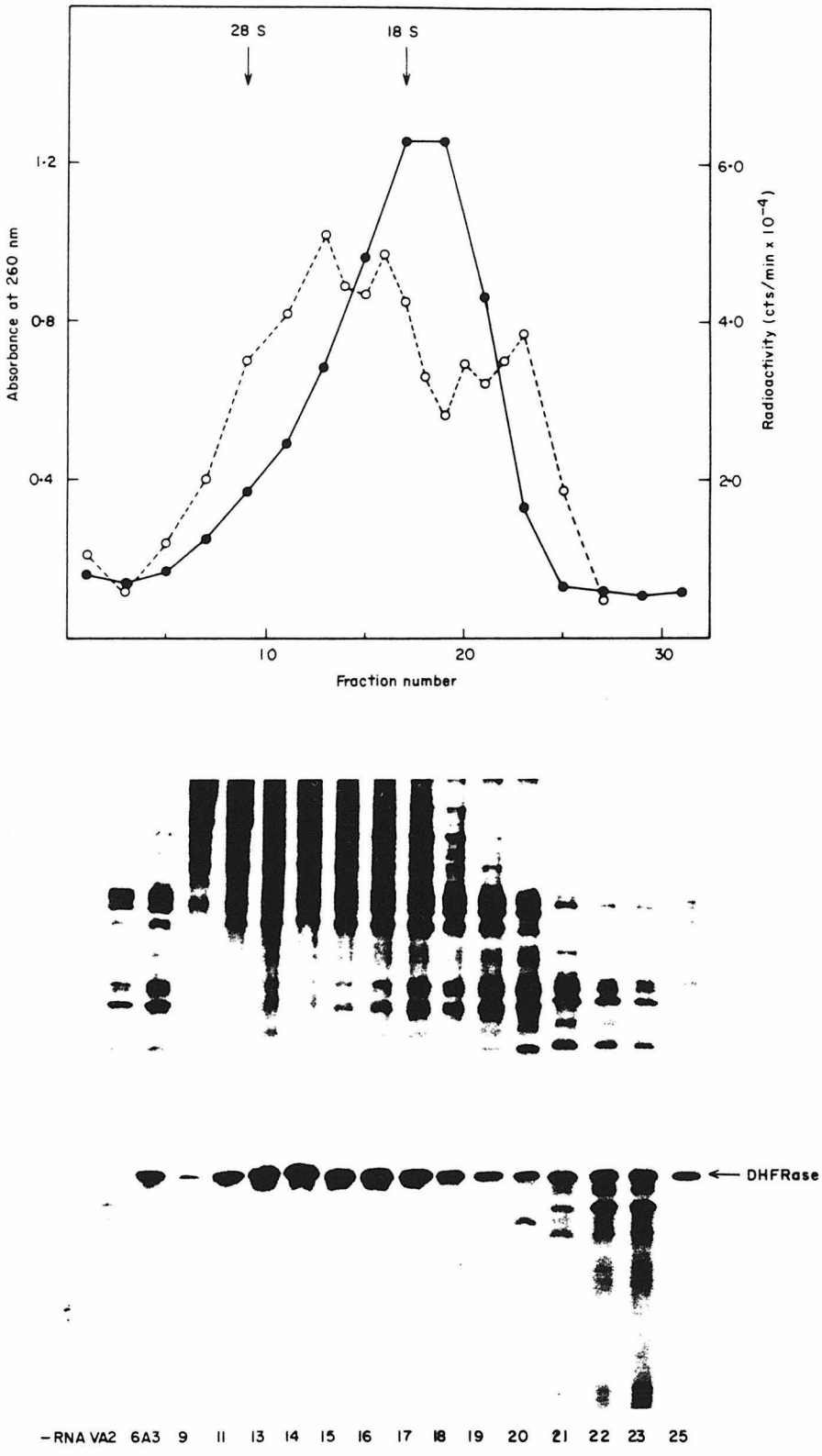

28 S
18 S
1·2
0·8
0·4
Absorbance at 260 nm
6·0
4·0
2·0
Radioactivity (cts/min x 10⁻⁴)
10
20
30
Fraction number
← DHFRase
–RNA VA2 6A3 9 11 13 14 15 16 17 18 19 20 21 22 23 25

FIG. 3.

RNA from 6A3 cells fractionated on sucrose gradients, was expected to be a better, though less abundant, source of cDNA for cloning the DHFRase coding sequence and for obtaining expression of this sequence in *E. coli*. Furthermore, since the human DHFRase coding sequence was expected to have a size similar to that of the mouse enzyme coding sequence (Chang *et al.*, 1978), it seemed reasonable to expect that even relatively small double-stranded cDNA molecules (0·6 to 1·0 kb) synthesized from any complete single-stranded cDNA copies of the 3·8 kb RNA might contain the DHFRase coding sequence. On the basis of the arguments presented above, both the larger size (>1·4 kb) and the smaller size cDNA (<1·4 kb) transcribed from the total poly(A)-containing RNA were used independently in the cloning experiments described below. The approaches followed are illustrated in schematic form in Figure 4.

Double-stranded cDNA was prepared, as described in Materials and Methods, from total poly(A)-containing RNA of the methotrexate-resistant cell line 6A3. Lane (b) in Figure 5 shows the electrophoretic pattern of the ss-cDNA in an alkaline agarose gel, while lanes (c) and (d) show the patterns in the same gel of the ds-cDNA before and after digestion with $S_1$. It is clear that the bulk of the ss-cDNA has sizes ranging between 400 and 3000 nt, with evidence of the presence of larger molecules up to at least 4000 nt: the non-$S_1$-treated, denatured ds-cDNA reveals an approximate doubling of its molecular length. The $S_1$-treated ds-cDNA, covering in the alkaline agarose gel a size range between 300 and about 4000 nt (lane (d)), was fractionated on a 5% to 20% neutral sucrose gradient to separate fragments smaller than 600 bp, which would compete during bacterial transformation, from the rest (Fig. 5). The ds-cDNA larger than 600 bp was separated into three cuts, with sizes ranging from 600 to 1400 bp, 1400 to 3000 bp and 3000 to 3800 bp, as estimated from their migration in the sucrose gradient relative to that of known restriction fragments, and as verified by electrophoretic analysis in an alkaline agarose gel (Fig. 5, lanes (e), (f) and (g)).

The two fractions of ds-cDNA containing molecules larger than 1400 bp were pooled to yield ~0·25 μg of ds-cDNA, which was then dC-tailed, annealed with *Pst*I digested and dG-tailed pBR322, and utilized to transform *E. coli* χ2282 (Chang *et al.*, 1978). About 4000 tetracycline-resistant transformants were obtained that were also ampicillin sensitive, due to the insertional inactivation of the β-lactamase gene. These transformants were replica plated on agar plates in minimal medium supplemented with 4 μg of trimethoprim/ml, in order to test for the possible

---

FIG. 3. Fractionation of poly(A)-containing RNA from 6A3 cells by sedimentation through a sucrose gradient and analysis of the programming capacity for protein synthesis *in vitro* of the individual fractions. The upper panel shows the $A_{260}$ profile in a sucrose gradient of the poly(A)-containing RNA of 6A3 cells (●) and the profile of the acid-precipitable radioactivity incorporated into protein in translation reactions *in vitro* (○) programmed with equal volumes of the given fractions.

The lower panel shows the electrophoretic patterns of the products of the translation reactions *in vitro* carried out using portions of the various fractions. The lanes designated $VA_2$ and 6A3 contain the products of reactions programmed with the total poly(A)-containing RNA from $VA_2$-B and 6A3 cells; the lane designated −RNA shows the products obtained with no added RNA in the translation mixture *in vitro*. The overexposure of the fluorogram has artificially accentuated the difference in amount of radioactivity between gel tracks 11 to 17 and gel tracks 21 to 23.

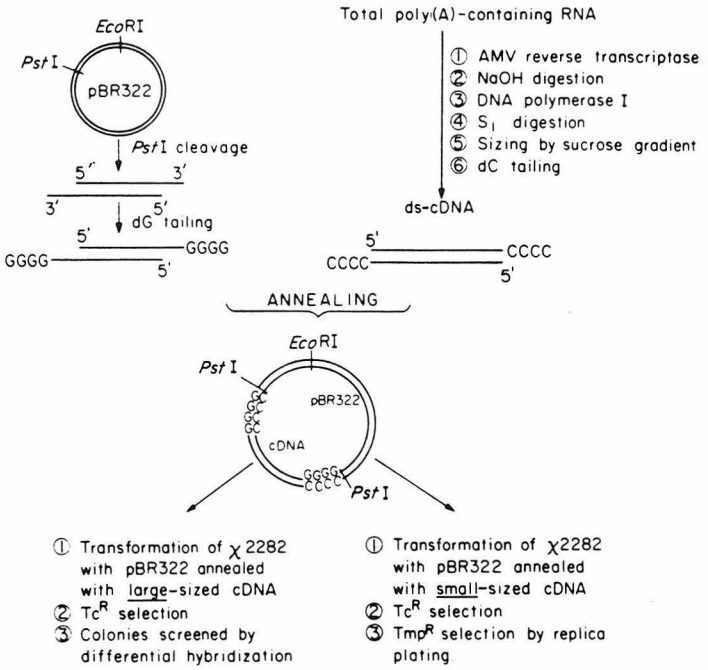



FIG. 4. Scheme illustrating the experimental approaches followed for cloning the human DHFRase cDNA and detecting its expression in *E. coli*. AMV, avian myeloblastosis virus. Sequence hyphens have been omitted for clarity.

expression in these bacteria of the human DHFRase sequence; however, none of the colonies exhibited trimethoprim resistance.

In order to detect the presence in the transformants of any recombinant plasmid molecules containing a DNA insert complementary to the 3·8 kb DHFRase mRNA, 2000 transformants were screened using the differential colony hybridization technique of St. John & Davis (1979), as modified by Wahl *et al.* (1979). For this approach, advantage was taken of the fact that the putative mRNAs for DHFRase are present in the total population of poly(A)-containing RNA from methotrexate-resistant 6A3 cells at a much higher level than in the parental cell line $VA_2$-B (Fig. 1). The colonies were replica plated in duplicate onto nitrocellulose filters, and one filter of each pair was incubated with $^{32}$P-labeled cDNA synthesized from the total poly(A)-containing RNA of 6A3 cells, and the other filter with an equal amount of radioactivity of cDNA from the parental $VA_2$-B cells: 76 colonies exhibited a much stronger hybridization with the 6A3 cDNA probe. Figure 6 shows a typical pair of nitrocellulose filters subjected to the differential colony hybridization test: one can recognize several clones showing a clearly different hybridization with the two probes.

The smaller cDNA fragments, with sizes ranging between 600 and 1400 bp, separated in the sucrose gradient fractionation of Figure 5, were also dC-tailed,

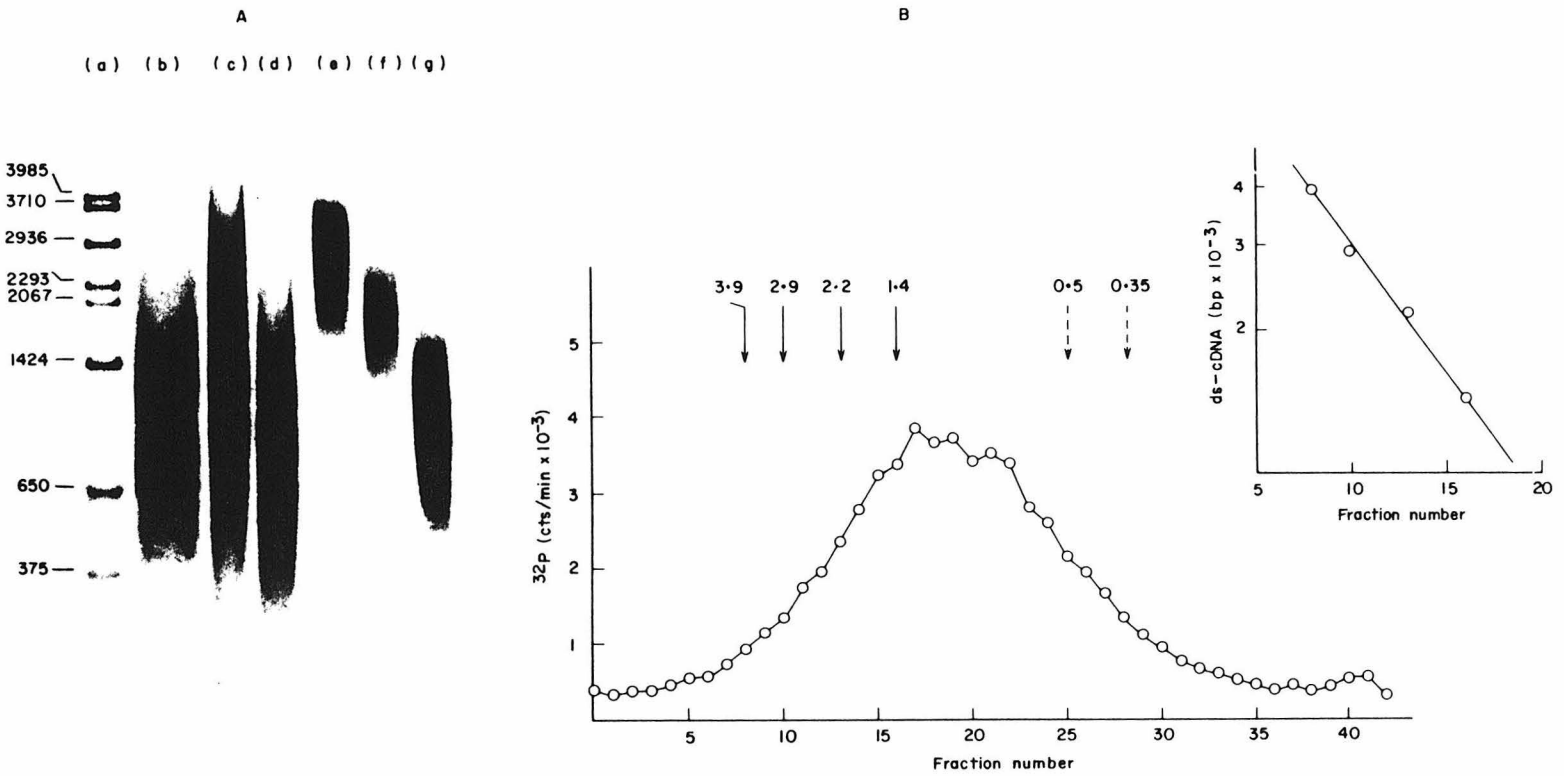



FIG. 5. Fractionation of human DHFRase cDNA.

A. Autoradiogram, after electrophoresis of DHFRase cDNA through an alkaline 1% agarose slab gel. Lane (a), size standards (in bp) represented by $^{32}$P-end labeled restriction fragments of pBR322 digested with *Bam*HI-*Eco*RI, or *Scl*I-*Eco*RI, or *Ava*I-*Eco*RI, or *Pvu*II-*Eco*RI. Lane (b), ss-cDNA synthesized using the total poly(A)-containing RNA as a template. Lanes (c) and (d), ds-cDNA before (c) and after (d) treatment with nuclease $S_1$. Lanes (e) to (g), cDNA in pooled fractions 7 to 10 (e), 11 to 15 (f) and 16 to 25 (g) from the sucrose gradient shown in B.

B. Fractionation through a sucrose gradient of DHFRase ds-cDNA. The unbroken arrows indicate fractions corresponding to the size standards run in a parallel gradient; the broken arrows indicate fractions corresponding to sizes extrapolated from the linear plot of the marker size *versus* migration (inset).

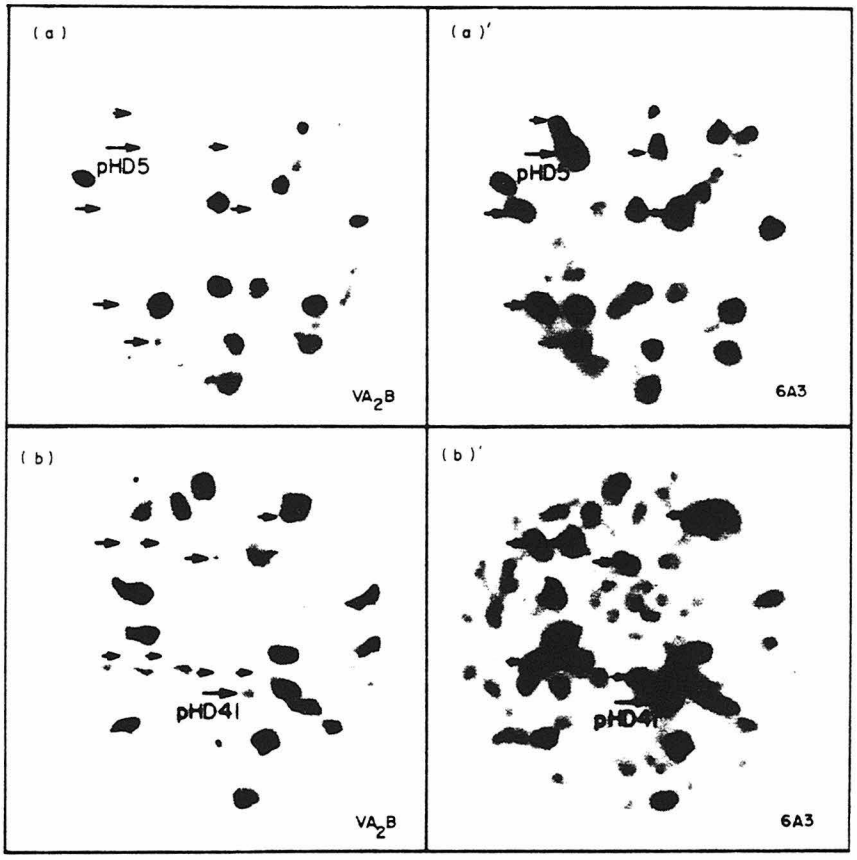



FIG. 6. Differential colony hybridization. A total of 112 tetracycline-resistant and ampicillin-sensitive transformants of *E. coli* χ2282 were replica plated on each pair of nitrocellulose filters ((a), (a)' and (b), (b)') and hybridized with $^{32}$P-labeled cDNA from total poly(A)-containing RNA of the parental $VA_2$-B cells ((a), (b)) or of the methotrexate-resistant 6A3 cells ((a)', (b)'). The arrows point to colonies giving an enhanced signal with the 6A3 probe.

annealed with *Pst*I-cut and dG-tailed pBR322, and utilized to transform *E. coli* χ2282 cells. From 0·1 μg of cDNA, 2600 tetracycline-resistant and ampicillin-sensitive colonies were obtained. Among them, eight were able to grow, although with different efficiencies, in the presence of 4 μg of trimethoprim/ml. These clones were then tested for resistance to higher doses of trimethoprim (10, 25, 50, 75, 100, 500, 1000 μg/ml) in order to determine the maximum concentration that allowed growth. As shown in Table 1, this varied between about 4 μg/ml and more than 1 mg/ml, with only one clone (subsequently shown to harbor the plasmid pHD84) exhibiting resistance to the latter concentration. In order to check whether trimethoprim resistance in the eight clones did indeed result from the

TABLE 1

*Properties of recombinant plasmids containing DHFRase cDNA inserts*

| Recombinant plasmid number | Size of insert† (no. bp) | Colony hybridization test‡ | Trimethoprim resistance test | Maximum trimethoprim concentration allowing growth of host bacteria (μg/ml) |
|---|---|---|---|---|
| A. *Clones selected by differential hybridization (pHD1–76)* | | | | |
| pHD1 | 1600 | + | − | <4 |
| pHD5 | 2250 | + | − | <4 |
| pHD23 | 940 | + | − | <4 |
| pHD25 | 2030 | + | − | <4 |
| pHD35 | 860 | + | − | <4 |
| pHD41 | 2300 | + | − | <4 |
| pHD43 | 960 | + | − | <4 |
| pHD56 | 2080 | + | − | <4 |
| pHD58 | 1000 | + | − | <4 |
| Remaining plasmids | 750–1500 | + | − | <4 |
| B. *Clones selected by trimethoprim resistance* | | | | |
| pHD77 | ~1000 | + | + | 4 |
| pHD78 | ~1000 | + | + | 4 |
| pHD79 | 1030 | + | + | 4 |
| pHD80 | 980 | + | + | 25 |
| pHD81 | ~1000 | + | + | 25 |
| pHD82 | 980 | + | + | 100 |
| pHD83 | 1010 | + | + | 4 |
| pHD84 | 720 | + | + | >1000 |

† Estimated from electrophoretic mobility in agarose gels.

‡ Refers to the differential hybridization assay for the clones of group A. and to the direct hybridization assay for the clones of group B.

transformation of *E. coli* χ2282 with recombinant plasmids containing DHFRase cDNA sequences, a colony hybridization test was performed using, as a probe, the nick-translated 2300 bp insert of one of the plasmids (pHD41) that had previously given a positive result in the differential hybridization experiment, and thus presumably contained at least a portion of the 3·8 kb mRNA sequence. All of the eight clones gave positive signals in this test.

Finally, in order to prove that the trimethoprim-resistance phenotype was plasmid-borne in the transformants, and at the same time to introduce the plasmids into a host easier to grow, the two plasmids that had conferred upon *E. coli* χ2282 the highest trimethoprim resistance (pHD82 and pHD84) were purified and utilized to transform the trimethoprim-sensitive *E. coli* HB101 strain. With both plasmids, 100% of the tetracycline-resistant transformants were found also to be trimethoprim-resistant.

(c) *Physical characterization of the cloned cDNA segments*

Table 1 summarizes the data concerning the size of the inserts in the plasmids carrying DHFRase cDNA sequences that have been described in the previous section, and the assays utilized for the identification of the clones transformed with these plasmids. The plasmids harbored in bacteria identified by differential hybridization have inserts varying in size between 750 and 2300 bp; those whose bacterial hosts were identified on the basis of the phenotypic expression of DHFRase cDNA sequences have inserts with sizes ranging from 750 to 1000 bp. Five of the latter plasmids (pHD79, pHD80, pHD82, pHD83 and pHD84), five of the plasmids originally identified by differential hybridization (pHD1, pHD5, pHD25, pHD41 and pHD56), and another group of plasmids from the same set (pHD23, pHD35, pHD43 and pHD58), which were subsequently shown to give a positive signal in a colony hybridization test using the nick-translated insert of pHD84, were used to construct a physical map of the cloned DHFRase cDNA sequences by using a variety of restriction enzymes (Fig. 7). On the basis of the distribution of the restriction sites, the inserts of the plasmids mentioned above were found to form a set of overlapping sequences covering a region that extends over approximately 3350 bp. This size is fairly close to that of the 3·8 kb DHFRase mRNA, suggesting that most of the sequences of this mRNA are represented in the cloned segments. The insert of plasmid pHD5 was found to contain a *Pvu*II site that is absent from the overlapping segment of the insert of pHD41. This could be due to heterogeneity in the RNA or to a copying error by the reverse transcriptase or to some event that has occurred in the bacterial host. From the alignment of the inserts of the plasmids analyzed here and from their orientation relative to neighboring restriction sites in the vector, it could be established that, in the plasmids of the clones that expressed trimethoprim resistance (pHD79, pHD80,

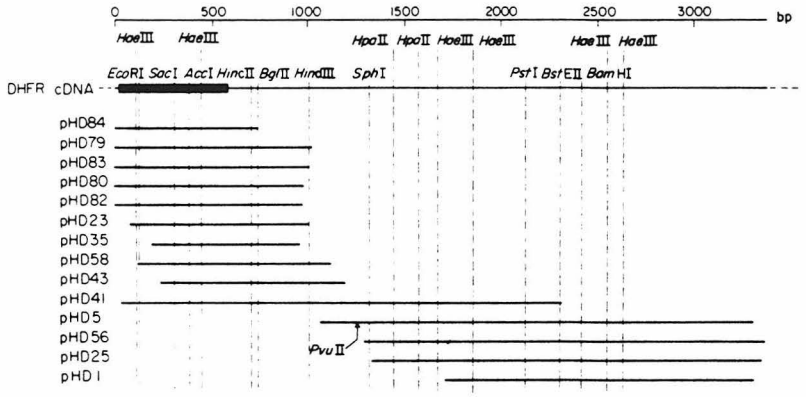



FIG. 7. Restriction map of the cDNA synthesized from the human DHFRase mRNA, as constructed by analysis of the inserts of the recombinant plasmids cloned in *E. coli*. In the map, the filled bar indicates the limits of the DHFRase coding sequence, as determined from the DNA sequence (unpublished results). The pHD5 insert exhibits a *Pvu*II restriction site absent from pHD41 (see the text). See the text for the orientation of the inserts in the recombinant plasmids.

pHD82, pHD83 and pHD84) and in plasmids pHD5, pHD43 and pHD58, the cDNA is oriented in the opposite direction as in plasmids pHD1, pHD23, pHD35 and pHD41 (data not shown).

The sequences of the inserts of the plasmids phenotypically expressed in *E. coli* and therefore presumed to contain the complete DHFRase coding region, are located near one extremity of the restriction map (Fig. 7). This extremity was expected to correspond to the 5′ end of the strand with the same polarity as the 3·8 kb mRNA, by analogy with the situation described in the mouse system, where the main DHFRase mRNA, ~1600 nt long, contains the 564 nt long coding sequence near its 5′ end (Nunberg *et al.*, 1980). This expectation has been fully confirmed by DNA sequence analysis (see below). The distribution of restriction sites pointed to the occurrence of only one DHFRase coding stretch in the 3350 bp long span of cDNA sequences. A DNA transfer hybridization experiment in which the plasmids pHD5, 41 and 56, digested with *Pst*I, and the plasmid pHD41, doubly digested with *Pst*I and *Bgl*II, were hybridized with the nick-translated *Pst*I insert of pHD84, the plasmid with the strongest phenotypic expression, gave results that clearly confirmed the conclusion of the restriction enzyme analysis (data not shown).

### (d) *DNA sequence analysis*

Both strands of the small 137 bp *Pst*I-*Eco*RI fragment of pHD84 were sequenced in order to determine the orientation of the coding sequence. The sequence of the sense strand is presented in Figure 8. It shows a reading frame, starting at 23 nt downstream of the poly(dG) tail, that can be recognized as DHFRase coding sequence by its homology to the mouse DNA sequence (Chang *et al.*, 1978; Nunberg *et al.*, 1980), which is also shown in Figure 8. These results confirm the previously stated assumption that the coding region is at the 5′ end of the 3·8 kb mRNA. The homology between the mouse and human DHFRase coding sequences in the first 75 nucleotides (i.e. those known so far for this portion of the mouse reading frame (Chang *et al.*, 1978)) is 82%, while the amino acid sequence homology between the mouse (Stone *et al.*, 1979) and human enzyme in the first 32 amino acid residues is 94%. A nucleotide sequence homology of about 80% is found between the mouse and human DNA sequences also in the 12 nucleotides preceding the coding stretch while, further upstream, the sequence homology decreases. There is, at 16 nt from the initiator AUG, a 6 nt sequence, 5′ . . . G-G-A-G-G-T . . . 3′, which shows a perfect base complementarity to the Shine–Dalgarno sequence, 3′ . . . C-C-U-C-C-T . . . 5′ (Shine & Dalgarno, 1974).

From the sequence data of Figure 8 and the above-mentioned orientation of the cDNAs in the vector, it can be inferred that, in the plasmids of the trimethoprim-resistant clones analyzed here (pHD79, pHD80, pHD82, pHD83 and pHD84), the DHFRase coding sequence is oriented in the same direction as the coding sequence of the β-lactamase gene; the homologous strand in plasmids pHD5, pHD43 and pHD58 is also oriented in the direction of reading of the β-lactamase gene, while it is oriented in the opposite direction in plasmids pHD1, pHD23, pHD35 and pHD41.

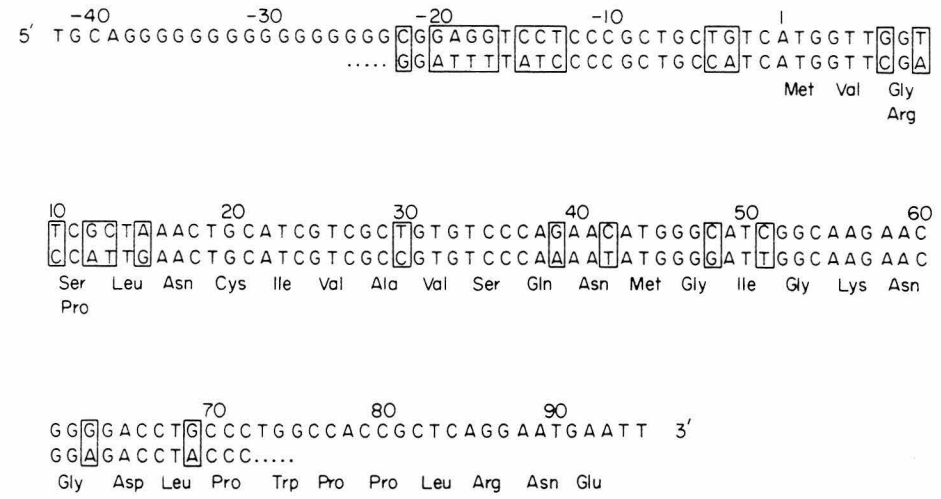



FIG. 8. DNA sequence of the 137 nt *Pst*I-*Eco*RI fragment of pHD84. The upper DNA sequence is that of a segment of the sense strand of pHD84, the lower sequence is that available for the mouse DHFRase cDNA (Chang *et al.*, 1978; Nunberg *et al.*, 1980). The numbering starts from the ATG initiator codon. The identity of the first 4 nucleotides was deduced from the sequence of the opposite strand. Boxed regions represent the non-homologous nucleotides. Below the DNA sequences is the amino acid sequence deduced from them; where 2 amino acids are indicated at the same position, the upper refers to the human DHFRase and the lower to the mouse DHFRase (Chang *et al.*, 1978; Stone *et al.*, 1979). The sequence hyphens have been omitted for clarity.

### (e) *Detection of human DHFRase in* E. coli *χ2282 transformants*

The previously discussed evidence, concerning the expression of trimethoprim resistance in *E. coli* χ2282 transformants carrying plasmids whose inserts contain the DHFRase coding sequence, strongly suggested that the basis for the drug resistance was the synthesis of the human enzyme in the bacterial host. It was in fact expected that the human DHFRase, as other mammalian DHFRases, is less sensitive to trimethoprim than the bacterial enzyme (Burchall & Hitching, 1965). In order to verify the presence of the human enzyme in *E. coli* χ2282, an immunoautoradiographic assay was carried out. For this purpose, *E. coli* χ2282 cell extracts harboring different plasmids were fractionated on a 12·5% polyacrylamide slab gel; the latter was then incubated with IgG prepared from rabbit anti-human DHFRase antiserum and then with $^{125}$I-labeled protein A, following a published procedure (Granger & Lazarides, 1979). Figure 9 shows the autoradiographic patterns of the cell extracts from *E. coli* χ2282 strains harboring pHD84 or pHD82 plasmids, grown with and without trimethoprim. In the four samples, a band migrating with the same mobility as the purified human DHFRase can be observed. This band is, on the contrary, absent from a cell extract from *E. coli* χ2282 harboring pHD5, a plasmid whose insert lacks the DHFRase coding sequence (Fig. 7). In the autoradiography in Figure 9, one can see in the extracts from the trimethoprim-resistant strains a specific band migrating slightly faster than the human DHFRase. The possible significance of this polypeptide is discussed below.

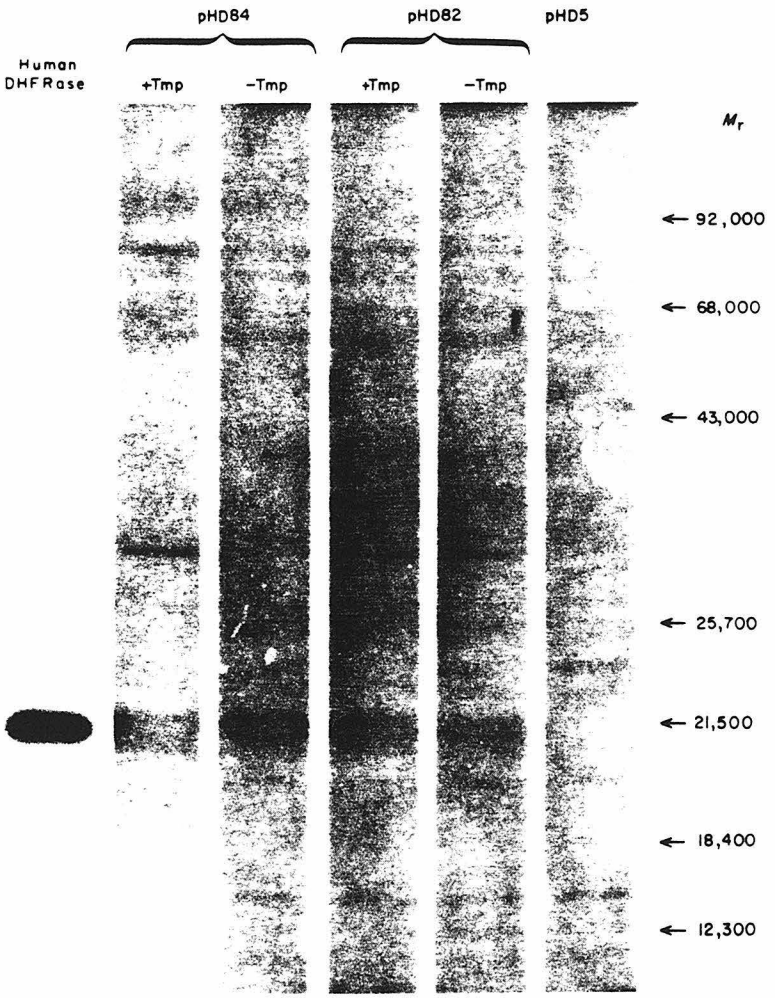



FIG. 9. Immunoautoradiography after sodium dodecyl sulfate/polyacrylamide slab gel electrophoresis of extracts of bacteria harboring plasmids with DHFRase cDNA inserts. *E. coli* χ2282 harboring pHD84 or pHD82 plasmids were grown in the presence (+Tmp) or absence (−Tmp) of trimethoprim (1 or 0·1 mg/ml). The extract of *E. coli* χ2282 harboring pHD5 and not expressing trimethoprim resistance was run as a control. On each slot, approximately 85 μg of *E. coli* extract were loaded. In the left-hand lane, 0·25 μg of pure human DHFRase (Morandi & Attardi, 1981) was run for comparison of its migration with that of the protein synthesized in bacteria. The arrows and $M_r$ values on the right represent the positions of migration and molecular weights of the protein standards run in a separate lane.

### (f) *Multiple forms of RNA homologous to DHFRase cDNA*

The results shown in Figure 3 had suggested the possibility that the 6A3 cells may contain more than one RNA species coding for DHFRase. In order to investigate this possibility further, equal amounts of poly(A)-containing RNA from the methotrexate-sensitive cell lines $VA_2$-B and HeLa Bu25 and from their

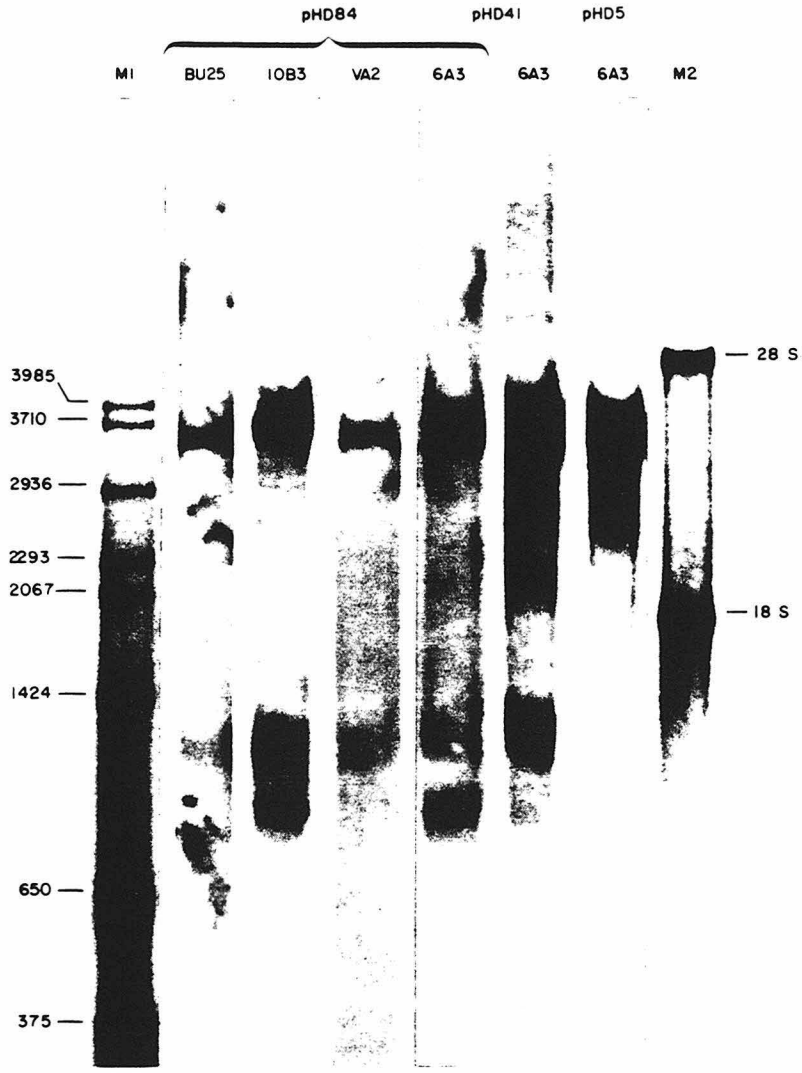



FIG. 10. Poly(A)-containing RNA from methotrexate-sensitive and resistant cells transferred to nitrocellulose paper and hybridized with different nick-translated DHFRase cDNA probes. Samples (5 $\mu$g) of poly(A)-containing RNA from each of the methotrexate-sensitive $VA_2$-B and HeLa Bu25 cell lines and of their methotrexate-resistant derivatives 6A3 and 10B3 were electrophoresed through a formaldehyde/agarose gel, transferred to nitrocellulose paper and hybridized with nick-translated plasmid DNA, as indicated. Lane M1 contains 3′-end labeled, digested pBR322 DNA size standards, and lane M2 contains total HeLa cell cytoplasmic RNA $^{32}$P-labeled *in vivo*. The autoradiograms were exposed at −70°C with intensifier screens for 11 h for the HeLa Bu25 lane, and for 2 h for the other lanes.

respective methotrexate-resistant variants 6A3 and 10B3 were electrophoresed on formaldehyde/agarose gels, transferred to nitrocellulose paper and hybridized with the nick-translated plasmid DNAs pHD5, pHD41 and pHD84 (Fig. 10). The results obtained with the pHD84 and pHD41 probes show that at least three RNA species

hybridizing with DHFRase-specific cDNA, with molecular sizes of about 3·8, 1·0 and 0·8 kb, are present in the methotrexate-resistant variants and also, in much reduced amounts, in their parental cell lines. In contrast to pHD84 and pHD41, the plasmid pHD5 hybridized only with the large DHFRase mRNA (Fig. 10). The same result as with pHD5 was obtained when nick-translated pHD1, pHD25 and pHD56 DNAs were used to probe the 6A3 RNA (data not shown). These observations indicate that the 1·0 and 0·8 kb RNAs must contain sequences corresponding to the 5′ end proximal half of the 3·8 kb mRNA and, in particular, must overlap the DHFRase coding sequence at least in part.

## 4. Discussion

Among the transformants obtained in the present work using plasmids with inserts deriving from cDNA molecules larger than 1400 bp, 76 out of 2000 analyzed revealed the presence of DHFRase cDNA sequences by differential colony hybridization, but none of them ($<5\times10^{-4}$) proved to be trimethoprim resistant. In agreement with this result, restriction enzyme mapping and colony hybridization data indicated that 71 out of 76 chimaeric plasmids did not contain in their insert any portion of the DHFRase coding sequence (among them, pHD1, pHD5, pHD25 and pHD56), while four contained an incomplete coding sequence (pHD23, pHD35, pHD43 and pHD58). The seventy-sixth plasmid, pHD41, was found to contain an insert possibly including a complete DHFRase coding sequence; however, the orientation of this coding sequence in the opposite direction to that of the $\beta$-lactamase gene would probably have allowed it to be expressed only weakly, even if it were complete (Chang *et al.*, 1980). The mapping positions of the inserts of the nine above-mentioned recombinant plasmids were consistent with the view that they may have resulted from incomplete extension by the reverse transcriptase of the single-stranded cDNA to the 5′ end of the 3·8 kb RNA and/or from incomplete synthesis of the second strand by *E. coli* DNA polymerase I.

In contrast to the negative results, in terms of phenotypic expression, obtained with the plasmids containing large size cDNA-derived inserts, eight clones out of 2600 transformants ($\sim3\times10^{-3}$) harboring plasmids with inserts derived from cDNA molecules in the range 600 to 1400 bp were found to be trimethoprim resistant. A restriction enzyme mapping of the inserts of these plasmids revealed that they all contained the DHFRase coding stretch, as identified in one of the plasmids (pHD84) by DNA sequencing. It is not possible to say whether this DHFRase coding sequence derived from cDNA synthesized from the 5′-end portion of the 3·8 kb mRNA or from the 5′-end portion of smaller DHFRase mRNAs. The available evidence indicates that the trimethoprim resistance in these clones was due to the expression of the exogenous DHFRase coding sequence, as previously shown for *E. coli* cells transformed with plasmids carrying the mouse DHFRase coding sequence (Chang *et al.*, 1978). Both the above-mentioned restriction mapping and DNA sequencing data, and the observation of the presence of a protein immunoreactive with anti-human DHFRase in extracts of these transformants, but not in extracts of clones harboring plasmids that lacked the

DHFRase coding stretch, containing instead other segments of DHFRase cDNA, strongly support the above conclusion. According to this interpretation, the great variability in degree of trimethoprim resistance observed in these clones presumably reflects differences in the rate of expression of the DHFRase coding sequence. As previously suggested for the mouse DHFRase coding sequence cloned in *E. coli* (Chang *et al.*, 1980), it seems very likely that the above differences result from variations in the degree of homology between the S-D sequence at the 3′ end of the 16 S rRNA and the ribosome attachment site in the mRNA on the 5′ side of the coding stretch, and/or variations in the distance between this ribosome binding site and the initiator codon. However, this point has not been investigated. In the chimaeric plasmid pHD84, which conferred upon the host the highest resistance to trimethoprim (>1 mg/ml), comparable to that of the "strong expressors" described among the bacteria transformed with mouse DHFRase cDNA (Chang *et al.*, 1978), the DNA sequence analysis has revealed, at a distance of 16 nucleotides from the start codon AUG, the presence of the sequence 5′ . . . G-G-A-G-G-T . . . 3′, showing a perfect homology to the S-D sequence 3′ . . . C-C-U-C-C-A . . . 5′ (Fig. 8).

The significance of the polypeptide present in the extracts from trimethoprim-resistant transformants, which migrated on polyacrylamide gel electrophoresis slightly faster than the human DHFRase, and which showed immunoreactivity to anti-human DHFRase antibodies, is not clear. It is conceivable that it represents a slightly shorter form of the enzyme, synthesized starting at the GUG at positions 31 to 33 in the sequence shown in Figure 8. There is indeed, at a distance of 18 nucleotides upstream of this GUG, a sequence (5′ . . . A-T-G-G-T-T-G-G-T-T-C-G-C . . . 3′) showing homology of nine nucleotides (underlined), within a 13-nucleotide segment, with the sequence surrounding and including the S-D sequence: a similar degree of homology has previously been interpreted to be adequate for expression of the mouse DHFRase coding sequence in transformed bacteria (Chang *et al.*, 1980). Further work is needed, however, to verify the possibility mentioned above concerning the nature of the faster migrating polypeptide.

The sequence reported here of the first 32 amino acid residues of the human DHFRase shows a 94% homology with the sequence of the mouse enzyme (Stone *et al.*, 1979). This finding is fully consistent with the available evidence, indicating the high degree of sequence conservation among DHFRases of vertebrate origin (Kumar *et al.*, 1980). The complete sequence of the human DHFRase will be reported elsewhere.

The most significant observations made in the present work concern the presence, in human methotrexate-resistant and sensitive cells, of multiple forms of DHFRase mRNA differing in the length of the 3′ untranslated region. These observations confirm and extend those made by Setzer *et al.* (1980) in mouse cells. The most abundant DHFRase mRNA species in the human cells analyzed here is a 3·8 kb poly(A)-containing RNA, which has a coding stretch of 564 nucleotides (unpublished observations) near its 5′-end and an apparently untranslated 3′ tail of ~3·2 kb. Never before has an mRNA with such a long 3′ untranslated segment relative to the coding stretch been observed. This mRNA appears to be functional, because it is found in polysomes and because it is capable of programming a reticulocyte lysate for the synthesis of DHFRase.

Two other forms of poly(A)-containing RNA having sequences of the DHFRase coding stretch, with sizes of 1·0 and 0·8 kb, have been detected in the RNA transfer hybridization experiments. There is strong, although indirect, evidence that these two poly(A)-containing RNAs are also functional forms of DHFRase mRNA. Their size corresponds well to the slower sedimenting peak of DHFRase synthesis *in vitro*: promoting activity in the sucrose gradient fractionation of the poly(A)-containing RNA from the methotrexate-resistant 6A3 cells; furthermore, the reproducibility of their size and relative amounts in both methotrexate-sensitive cell lines and methotrexate-resistant variants with amplified DHFRase genes would speak against their arising from degradation of the 3·8 kb mRNA; also, the observation that these RNA species contain both a poly(A) tail and at least a portion of the DHFRase coding stretch argues against degradative phenomena being responsible for their presence. The most plausible interpretation of the present observations is that these species represent forms of functional DHFRase mRNA containing the coding stretch near their 5′ end as the 3·8 kb mRNA, but differing from the latter in the length of the 3′ untranslated segment.

The significance of the existence of multiple forms of DHFRase mRNA observed here in human cells, and previously reported in mouse and hamster cells (Setzer *et al.*, 1980), is obscure. No essential function for protein synthesis of the 3′ non-coding stretch of mRNA has been identified. The experiments by Kronenberg *et al.* (1979) tend to exclude an absolute requirement of this region for translation *in vitro*. Furthermore, in mammalian mitochondria, the mRNAs utilized for mitochondrial translation lack completely a 3′ non-coding stretch (Ojala *et al.*, 1981). Therefore, it is not surprising that multiple forms of DHFRase mRNA having a different length of the 3′ non-coding stretch may be utilized *in vivo* for DHFRase synthesis in mammalian cells. On the other hand, the persistence of these multiple forms in distinct cell lines maintained *in vitro* for a long time and having a different origin and history (like the HeLa BU25 and $VA_2$-B cell lines) and, in the case of the mouse DHFRase mRNA, the observation of the same multiplicity of species in cultured cells and in liver strongly suggest that this plurality of RNA species has a functional significance. What this possible significance is, and whether the different species of DHFRase mRNA are transcribed from different genes or, on the contrary, derive from alternative pathways of transcription or processing of RNA transcripts of the same gene, remains to be established.

The multiple DHFRase mRNA species were found to be greatly increased in amount in the poly(A)-containing RNA fraction from the two methotrexate-resistant variants 10B3 and 6A3, as compared to their respective parental lines, BU25 and $VA_2$-B. This observation is consistent with the previous evidence indicating a greatly increased DHFRase activity and content in the above-mentioned methotrexate-resistant cell lines, as compared to the parental lines (Masters *et al.*, 1982), and with more recent findings (unpublished) concerning a large increase in DHFRase-specific DNA sequences in their genome: taken together, the available evidence points to a DHFRase gene amplification as the basis for methotrexate-resistance in these cell lines.

These investigations were supported by National Institutes of Health grants GM-11726 and T32 GM-07616. The *E. coli* strain $\chi$2282 was generously provided by Stanley N. Cohen. We are very grateful to M. Steinmetz and C. M. Rice for valuable discussions, to I. Sandoval and E. Lazarides for a gift of [$^{125}$I]protein A, to J. Casey for a gift of calf thymus DNA primer, and to Keith Hughes for help in the preparation and analysis of plasmids. The technical assistance of Arger Drew is gratefully acknowledged.

## REFERENCES

Alt, F. W., Kellems, R. E. & Schimke, R. T. (1976). *J. Biol. Chem.* **251**, 3063–3074.

Alt, F. W., Kellems, R. E., Bertino, J. R. & Schimke, R. T. (1978). *J. Biol. Chem.* **253**, 1357–1370.

Amalric, F., Merkel, C., Gelfand, R. & Attardi, G. (1978). *J. Mol. Biol.* **118**, 1–25.

Bailey, J. M. & Davidson, N. (1976). *Anal. Biochem.* **70**, 75–85.

Berger, S. L. & Birkenmeier, C. S. (1979). *Biochemistry*, **18**, 5143–5149.

Biedler, J. L., Albrecht, A. M., Hutchison, D. J. & Spengler, B. A. (1972). *Cancer Res.* **32**, 153–161.

Bostock, C. J., Clark, E. M., Harding, N. G. L., Mounts, P. M., Tyler-Smith, C., Van Heyningen, V. & Walker, P. M. P. (1979). *Chromosoma*, **74**, 153–177.

Boyer, H. W. & Roulland-Dussoix, D. (1969). *J. Mol. Biol.* **41**, 459–472.

Buell, G. N., Wickens, M. P., Payvar, F. & Schimke, R. T. (1978). *J. Biol. Chem.* **253**, 2471–2482.

Burchall, J. J. & Hitching, G. H. (1965). *Mol. Pharmacol.* **1**, 126–136.

Cantatore, P. & Attardi, G. (1980). *Nucl. Acids Res.* **8**, 2605–2625.

Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. & Cohen, S. N. (1978). *Nature (London)*, **275**, 617–624.

Chang, A. C. Y., Erlich, H. A., Gunsalus, R. P., Nunberg, J. H., Kaufman, R. J., Schimke, R. T. & Cohen, S. N. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 1442–1446.

Chang, S. E. & Littlefield, J. W. (1976). *Cell*, **7**, 391–396.

Clewell, D. B. M. (1972). *J. Bacteriol.* **110**, 667–676.

Cohen, S. N., Chang, A. C. Y. & Hsu, L. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 2110–2114.

Davis, R. W., Thomas, M., Cameron, J., St. John, T. P., Scherer, S. & Padgett, R. A. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 404–411, *Academic Press, New York.*

Denhardt, D. (1966). *Biochem. Biophys. Res. Commun.* **23**, 641–646.

Dolnick, B. J., Berenson, R. J., Bertino, J. R., Kaufman, R. J., Nunberg, J. H. & Schimke, R. T. (1979). *J. Cell Biol.* **83**, 394–402.

Efstratiadis, A. & Villa-Komaroff, L. (1979). In *Genetic Engineering: Principles and Methods* (Setlow, J. K. & Hollander, A., eds), vol. 1, pp. 15–36, Plenum Press, New York.

Flintoff, W. F., Davidson, S. V. & Siminovitch, L. (1976). *Som. Cell Genet.* **2**, 245–261.

Friedman, E. Y. & Rosbash, M. (1977). *Nucl. Acids Res.* **4**, 3455–3471.

Granger, B. L. & Lazarides, E. (1979). *Cell*, **18**, 1053–1063.

Hakala, M. T., Zakrewski, S. F. & Nichol, C. A. (1961). *J. Biol. Chem.* **236**, 952–958.

Haseltine, W. A. M., Lindan, C. P., D'Andrea, A. D. & Johnsrud, L. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 235–248, Academic Press, New York.

Kabat, E. A. & Mayer, M. M. (1948). *Experimental Immunochemistry*, C. C. Thomas, Springfield.

Kessler, S. (1975). *J. Immunol.* **115**, 1617–1624.

Kit, S., Dubbs, D. R. & Frearson, P. M. (1966). *Int. J. Cancer*, **1**, 19–30.

Kronenberg, H. M., Roberts, B. E. & Efstradiatis, A. (1979). *Nucl. Acids Res.* **6**, 153–166.

Kumar, A. A., Blankenship, D. T., Kaufman, B. T. & Freisheim, J. H. (1980). *Biochemistry*, **19**, 667–678.

Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Laskey, R. A. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 363–371. Academic Press, New York.
Littlefield, J. W. (1969). *Proc. Nat. Acad. Sci., U.S.A.* **62**, 88–95.
Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975*a*). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 1184–1188.
Maniatis, T., Jeffrey, A. & Van de Sande, H. (1975*b*). *Biochemistry*, **14**, 3787–3794.
Masters, J., Keeley, B., Gay, H. & Attardi, G. (1982). *Mol. Cell. Biol.* in the press.
Maxam, A. M. & Gilbert, G. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 499–560. Academic Press, New York.
McDonell, M. W., Simon, M. N. & Studier, F. W. (1977). *J. Mol. Biol.* **110**, 119–146.
Melera, P. W., Lewis, J. A., Biedler, J. L. & Hession, C. (1980). *J. Biol. Chem.* **255**, 7024–7028.
Morandi, C. & Attardi, A. (1981). *J. Biol. Chem.* **256**, 10169–10175.
Mullins, J. I., Casey, J. W., Nicolson, M. O. & Davidson, N. (1980). *Nucl. Acids Res.* **8**, 3287–3305.
Nunberg, J. H., Kaufman, R. J., Schimke, R. T., Urlaub, G. & Chasin, L. A. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 5553–5556.
Nunberg, J. H., Kaufman, R. J., Chang, A. C. Y., Cohen, S. N. & Schimke, R. T. (1980). *Cell*, **19**, 355–364.
Ojala, D., Montoya, J. & Attardi, G. (1981). *Nature (London)*, **290**, 470–474.
Padgett, R. A., Wahl, G. M., Coleman, P. F. & Stark, G. R. (1979). *J. Biol. Chem.* **254**, 974–980.
Palmiter, R. D., Oka, T. & Schimke, R. T. (1971). *J. Biol. Chem.* **246**, 724–737.
Pontén, J., Jensen, F. & Koprowski, H. (1963). *J. Cell. Comp. Physiol.* **61**, 145–154.
Rave, N., Crkvenjakov, R. & Boedtker, H. (1979). *Nucl. Acids Res.* **6**, 3359–3567.
Rice, C. H. & Strauss, J. H. (1981). *J. Mol. Biol.* **154**, 325–348.
Roychoudhury, R. & Wu, R. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 43–62. Academic Press, New York.
Setzer, D. R., McGrogan, M., Nunberg, J. H. & Schimke, R. T. (1980). *Cell*, **22**, 361–370.
Shine, J. & Dalgarno, L. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 1342–1346.
Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Wiessman, S. M., Vehara, H., Nathenson, S. & Hood, L. (1981). *Cell*, **24**, 125–134.
St. John, T. & Davis, R. W. (1979). *Cell*, **16**, 443–452.
Stone, D., Paterson, S. J., Raper, J. H. & Phillips, A. W. (1979). *J. Biol. Chem.* **254**, 480–488.
Thomas, P. S. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 5201–5205.
Wahl, G. M., Padgett, R. A. & Stark, G. R. (1979). *J. Biol. Chem.* **254**, 8679–8689.
Wieslander, L. (1979). *Anal. Biochem.* **98**, 305–309.
Yang, R. C.-A., Lis, J. & Wu, R. (1979). In *Methods in Enzymology* (Wu, R., ed.), vol. 68, pp. 176–182. Academic Press, New York.

*Edited by S. Brenner*

# CHAPTER 4

## The Nucleotide Sequence of the cDNA Coding for the Human Dihydrofolic Acid Reductase

*Gene*, 21 (1983) 59–63
Elsevier Biomedical Press



# The nucleotide sequence of the cDNA coding for the human dihydrofolic acid reductase

(Recombinant DNA; Maxam-Gilbert sequencing; comparison with mouse, bovine and chicken DHFR; sequence conservation)

**Jeffrey N. Masters and Giuseppe Attardi**

*Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125 (U.S.A.)*

(Received October 13th, 1982)
(Accepted October 15th, 1982)

SUMMARY

The nucleotide sequence of the human dihydrofolic acid reductase (DHFR) reading frame has been derived from the analysis of human DHFR cDNA. This sequence and the corresponding amino acid sequence have been compared with those available for the enzyme and its coding segment from other organisms. There is an 89% nucleotide sequence homology between the human DHFR reading frame and the mouse coding sequence. Furthermore, amino acid-sequence homologies of 74%, 81% and 89% has been found between human DHFR and chicken, bovine and mouse DHFR, respectively.

## INTRODUCTION

DHFR is the enzyme responsible for reducing dihydrofolic acid to tetrahydrofolic acid in an NADPH-dependent reaction. It is the target enzyme of 4-amino analogs of folic acid, like amethopterin (Mtx), which have been used extensively in cancer chemotherapy (Bertino and Johns, 1972) and in the selection of cell lines with DHFR gene amplification (Schimke, 1981). We have previously described the cloning of the human DHFR cDNA from Mtx-resistant cells (Morandi et al., 1981).

In this communication we report the nucleotide sequence of the human DHFR reading frame derived from cDNA analysis and the corresponding amino acid sequence, and we compare these sequences with those available for the enzyme and its coding segment from other organisms.

## MATERIALS AND METHODS

The isolation of pHD80, pHD83 and pHD84 plasmids containing as inserts segments of the human DHFR cDNA was carried out as described previously (Morandi et al., 1981). The method of Maxam and Gilbert (1980) was used for sequencing. The strategy followed is described in the legend of Fig. 1.

Abbreviations: bp, base pairs; cDNA, DNA complementary to mRNA; DHFR, dihydrofolic acid reductase; Mtx, methotrexate; SDS, sodium dodecyl sulfate.

60

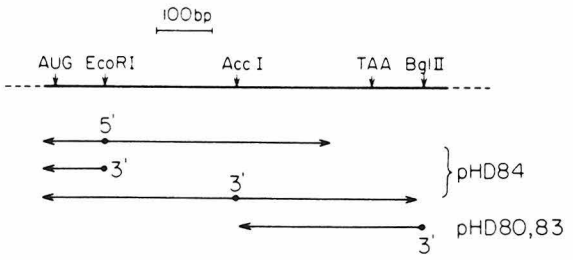



Fig. 1. Strategy used for sequencing the coding segment of human DHFR cDNAs. The single restriction sites of the inserts of pHD84, pHD80 and pHD83 which have been used are indicated. The plasmid DNAs were cut at one of the sites indicated above, and either 5'-end labeled with [$\gamma$-$^{32}$P]ATP and polynucleotide kinase after alkaline phosphatase treatment, and then digested with *Pst*I (there is one such site at each end of the insert in the three plasmids; Morandi et al., 1982), or digested with *Pst*I and then fill-in labeled at the 3'-ends with the Klenow fragment of *E. coli* DNA polymerase I, and [$\alpha$-$^{32}$P]dNTPs, as previously described (Morandi et al., 1982): in both cases, the fragments of interest, labeled in only one strand, were isolated by electrophoresis on a polyacrylamide gel. The arrows indicate the direction and extent of sequence determined.

## RESULTS AND DISCUSSION

Fig. 1 shows a portion of the physical map of the DHFR cDNA illustrating the positions of the restriction sites used for DNA sequencing relative to the translation initiation and termination codons, and the sequencing strategy employed here. Either 5'- or 3'-end labeling, or both, were performed at the sites indicated. We used three plasmids expressing the DHFR coding sequence in *E. coli*: pHD84, which confers resistance to > 1 mg/ml trimethoprim upon the host, and whose insert ends just before the *Bgl*II site in Fig. 1; pHD80 and pHD83, which confer upon *E. coli* cells resistance to 25 $\mu$g/ml and 4 $\mu$g/ml trimethoprim, respectively, and which contain inserts extending past the *Bgl*II site. No sequence differences were found in the overlapping coding regions of these plasmids.

The nucleotide sequence of the human DHFR cDNA coding region is shown in Fig. 2, in parallel

```
                10                  20                  30                  40                  50                  60
ATGGTTGGTTCGCTAAACTGCATCGTCGCTGTGTCCCAGAACATGGGCATCGGCAAGAAC
ATGGTTCGACCATTGAACTGCATCGTCGCCGTGTCCCAAAATATGGGGATTGGCAAGAAC

            70                  80                  90                  100                 110                 120
GGGGACCTGCCCTGGCCACCGCTCAGGAATGAATTCAGATATTTCCAGAGAATGACCACA
GGAGACCTACCCTGGCCTCCGCTCAGGAACGAGTTCAAGTACTTCCAAAGAATGACCACA

        130                 140                 150                 160                 170                 180
ACCTCTTCAGTAGAAGGTAAACAGAATCTGGTGATTATGGGTAAGAAGACCTGGTTCTCC
ACCTCTTCAGTGGAAGGTAAACAGAATCTGGTGATTATGGGTAGGAAAACCTGGTTCTCC

    190                 200                 210                 220                 230                 240
ATTCCTGAGAAGAATCGACCTTTAAAGGGTAGAATTAATTTAGTTCTCAGCAGAGAACTC
ATTCCTGAGAAGAATCGACCTTTAAAGGACAGAATTAATATAGTTCTCAGTAGAGAACTC

250                 260                 270                 280                 290                 300
AAGGAACCTCCACAAGGAGCTCATTTTCTTCCAGAAGTCTAGATGATGCCTTAAAACTT
AAAGAACCACCACGAGGAGCTCATTTTCTTGCCAAAAGTTTGGATGATGCCTTAAGACTT

            310                 320                 330                 340                 350                 360
ACTGAACAACCAGAATTAGCAAATAAAGTAGACATGGTCTGGATAGTTGGTGGCAGTTCT
ATTGAACAACCGGAATTGGCAAGTAAAGTAGACATGGTTTGGATAGTCGGAGGCAGTTCT

370                 380                 390                 400                 410                 420
GTTTATAAGGAAGCCATGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATG
GTTTACCAGGAAGCCATGAATCAACCAGGCCACCTTAGACTCTTTGTGACAAGGATCATG

        430                 440                 450                 460                 470                 480
CAAGACTTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGAGAAATATAAACTTCTG
CAGGAATTTGAAAGTGACACGTTTTTCCCAGAAATTGATTTGGGGAAATATAAACTTCTC

    490                 500                 510                 520                 530                 540
CCAGAATACCCAGGTGTTCTCTCTGATGTCCAGGAGGAGAAAGGCATTAAGTACAAATTT
CCAGAATACCCAGGCGTCCTCTCTGAGGTCCAGGAGGAAAAAGGCATCAAGTATAAGTTT

550                 560
GAAGTATATGAGAAGAATGATTAA
GAAGTCTACGAGAAGAAAGACTAA
```

Fig. 2. Nucleotide sequence of the human DHFR cDNA coding region (top lines). The sequence of the coding segment of DHFR derived from S180 M50 mouse cells (Chang et al., 1978; Nunberg et al., 1980; Crouse et al., 1981) is shown for comparison (bottom lines). The non-homologous regions are boxed.

with the nucleotide sequence of the DHFR coding region of a cDNA clone and genomic clones derived from Mtx-resistant S180 M50 mouse cells (Chang et al., 1978; Nunberg et al., 1980; Crouse et al., 1982).

The human DHFR reading frame is 564 nucleotides long and exhibits an 89% homology with the mouse DHFR coding stretch. Of the 64 nucleotide differences, 45 represent silent changes and 7 produce codons specific for analogous or isosteric amino acids. Therefore, there is a functional homology of 98% between the human and mouse coding sequences.

The nucleotide sequence codes for a protein of 186 amino acids (excluding the initiator methionine which has not been detected in the DHFRs sequenced so far) (Kumar et al., 1980), with a calculated $M_r$ of 21 297; this value agrees well with the $M_r$ of 21 000 estimated for the enzyme from its electrophoretic mobility on an SDS-polyacrylamide gel (Morandi and Attardi, 1981). Fig. 3 shows the human DHFR amino acid sequence and the positions where the sequences of the enzymes from chicken liver (Kumar et al., 1980), bovine liver (Lai et al., 1979), and mouse lymphoma L1210 cells (Stone et al., 1979), differ. The homologies between the human DHFR and the enzymes from chicken, bovine and mouse are 74, 81, and 89%, respectively. It is interesting to note that the cDNA derived amino acid sequence of the S-180 cells DHFR differs from the directly determined sequence of the lymphoma cell enzyme in four amino acid residues (122-Gln, 123-Glu, 127-Gln and 173-Lys), this representing a homology of 97%.

Matthews et al. (1978, 1979) have recently solved at 2.5 Å resolution the X-ray crystallographic structure of the ternary complex of the *L. casei* DHFR with Mtx and NADPH, identifying the amino acid residues involved in the three-dimensional binding of these molecules to the enzyme. More recently, the X-ray structure at 2.9 Å resolution of the ternary complex of the chicken liver DHFR with NADPH and a phenyltriazine inhibitor has similarly been determined (Voltz et al.,

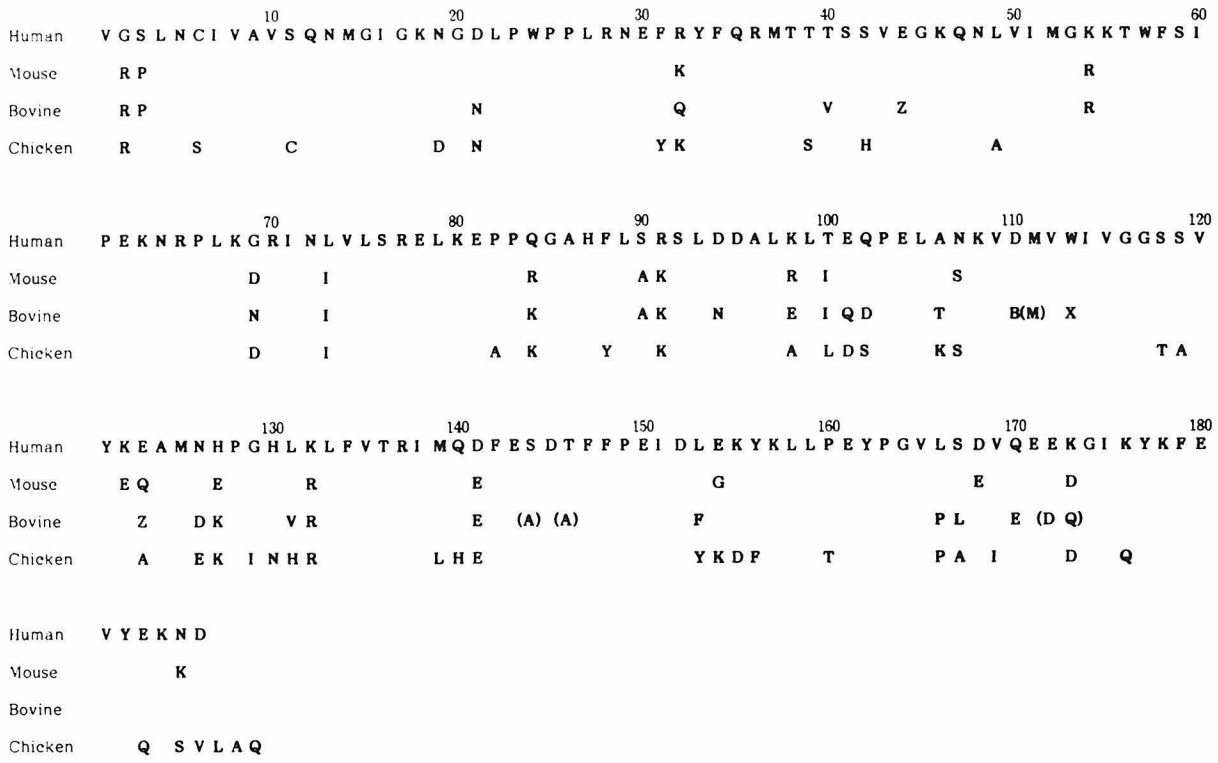



Fig. 3. Amino acid sequence of the human DHFR derived from the nucleotide sequence of Fig. 2. The nonhomologous amino acids in the sequence of the enzymes from mouse L120 cells (Stone et al., 1979), bovine liver (Lai et al., 1979) and chicken liver (Kumar et al., 1980) are indicated.

1982). This investigation has revealed that the overall backbone chain folding of the chicken enzyme is very similar to that of the bacterial enzyme, and that the binding of NADPH to the chicken DHFR involves interactions analogous or identical to those occurring in the *L. casei* DHFR-NADPH-Mtx ternary complex. Interestingly, an alignment of the *L. casei* DHFR amino acid sequence with the sequences of the homologous enzymes from other bacteria and from three mammalian sources (chicken liver, bovine liver and mouse L1210 cells) has revealed a high degree of conservation throughout the evolution of the amino acid residues involved in Mtx and NADPH binding (Kumar et al., 1980). Thus, 5 out of the 13 residues participating in Mtx binding in the *L. casei* enzyme (Ala-9, Phe-34, Leu-67, Arg-70 and Thr-136, using the mammalian numbering system) were found to be invariant, and two others (Leu-22 and Ser-59) almost invariant in all enzymes for which sequence data were available. Asp at position 30, also implicated in Mtx binding, was found to be invariant in all bacterial enzymes and replaced by Glu in the mammalian enzymes. A scanning of Fig. 3 shows that the human DHFR does not provide an exception to any of these rules. In the same study referred to above, of the 24 amino acid residues involved in NADPH binding in the *L. casei* enzyme, eight (Ala-9, Ile-16, Gly-17, Try-24, Gly-53, Thr-56, Leu-75 and Gly-117) were found to be invariant in the enzymes from all species analyzed, three others (Leu-22, Ser-59 and Gly-20) appear to be common to the three mammalian enzymes and many bacterial enzymes, and one (Thr-146) was found in all enzymes except in that from bovine liver, where it was substituted by Ala. The residues 54 and 55, which are also involved in NADPH binding, were found to be represented by either Arg or Lys in different enzymes. Again, it appears from Fig. 3 that the human DHFR sequence exhibits the invariant or semi-invariant features mentioned above, with Lys residues being present at positions 54 and 55 and Thr at position 146. The striking conservation from bacteria to man of the primary sequence of DHFR involved in the formation at the Mtx and NADPH binding sites obviously reflects the strong structural constraints which have accompanied the evolution of the enzyme. The knowledge of the primary sequence of human DHFR will hopefully be useful for the development of improved agents for cancer chemotherapy.

## ACKNOWLEDGEMENTS

This investigation was supported by NIH grants GM-11726 and T32 GM-07616.

## REFERENCES

Bertino, J.R. and Johns, D.: Folate antagonists, in Brodsky, I. (Ed.), Cancer Chemotherapy. Grune and Stratton, New York, 1972. pp. 9–22.

Chang, A.C.Y., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schimke, R.T. and Cohen, S.N.: Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. Nature 275 (1978) 617–624.

Crouse, G.F., Simonsen, C.C., McEwan, R.N. and Schimke, R.T.: Structure of amplified normal and variant dihydrofolate reductase genes in mouse sarcoma S180 cells. J. Biol. Chem. 257 (1982) 7887–7897.

Kumar, A.A., Blankenship, D.T., Kaufman, B.T. and Freisheim, J.H.: Primary structure of chicken liver dihydrofolate reductase. Biochemistry 19 (1980) 667–678.

Lai, P.-H., Pan, H.-C., Gleisner, J.M., Peterson, D.L., and Blakley, R.L.: Primary sequence of bovine liver dihydrofolate reductase, in Kisliuk, R.L. and Brown, G.M. (Eds.), Chemistry and Biology of Pteridines. Elsevier/North Holland Publishing Co., Amsterdam, 1979. pp. 437–440.

Maxam, A.M. and Gilbert, W.: Sequencing end labeled DNA with base specific chemical cleavages in Grossman, L. and Moldave, K. (Eds.), Methods in Enzymology 65. Academic Press, New York, 1980. pp. 499–560.

Matthews, D.A., Alden, R.A., Bolin, J.T., Filman, D.J., Freer, S.T., Hamlin, R., Hol, W.G.J., Kisliuk, R.L., Pastore, E.J., Plante, L.T., Xuong, N. and Kraut, J.: Dihydrofolate reductase from *Lactobacillus casei*: X-ray structure of the enzyme-methotrexate-NADPH complex. J. Biol. Chem. 253 (1978) 6946–6954.

Matthews, D.A., Alden, R.A., Freer, S.T., Xuong, N. and Kraut, J.: Dihydrofolate reductase from *Lactobacillus casei*: stereochemistry of NADPH binding. J. Biol. Chem. 254 (1979) 4144–4151.

Morandi, C. and Attardi, G. Isolation and characterization of dihydrofolic acid reductase from methotrexate-sensitive and -resistant human cell lines. J. Biol. Chem. 256 (1981) 10169–10175.

Morandi, C., Masters, J., Mottes, M. and Attardi, G. Multiple forms of human dihydrofolic acid reductase messenger RNA. J. Mol. Biol. 156 (1982) 583–607.

Nunberg, J.H., Kaufman, R.J., Chang, A.C.Y., Cohen, S.N. and Schimke, R.T.: Structure and genomic organization of the mouse dihydrofolate reductase gene. Cell 19 (1980) 355–364.

Schimke, R.T. (Ed.): Gene Amplification, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Stone, D., Phillips, A.W. and Burchall, J.J.: The amino acid sequence of the dihydrofolate reductase of a trimethoprim-resistant strain of *Escherichia coli*. Eur. J. Biochem. 72 (1979) 613–624.

Volz, K.W., Matthews, D.A., Alden, R.A., Freer, S.T., Hansch, C., Kaufman, B.T. and Kraut, J.: Crystal structure of avian dihydrofolate reductase containing phenyltriazine and NADPH. J. Biol. Chem. 257 (1982) 2528–2536.

Communicated by A.D. Riggs.

# CHAPTER 5

## A Human Dihydrofolate Reductase Pseudogene and its Relationship to the Multiple Forms of Specific Messenger RNA

*J. Mol. Biol.* (1983) **167**, 23–36

# A Human Dihydrofolate Reductase Pseudogene and its Relationship to the Multiple Forms of Specific Messenger RNA

JEFFREY N. MASTERS, JIYOUNG K. YANG, ALESSANDRA CELLINI AND GIUSEPPE ATTARDI

*Division of Biology, California Institute of Technology Pasadena, Calif. 91125, U.S.A.*

*(Received 15 December 1982)*

The presence of dihydrofolate reductase (DHFRase)-specific sequences that, in contrast to the normal DHFRase gene, are not amplified in a methotrexate-resistant cell line, has been detected in the DNA from human sperm and from several human cell lines. DNA fragments containing some of these sequences have been isolated from a cosmid library of human sperm DNA. One of these fragments contains a DHFRase pseudogene ($\psi$HD1) that completely lacks introns, has 92% sequence homology to the corresponding region of normal DHFRase complementary DNA, but exhibits several alterations that make it non-functional.

The sequence analysis of the inserts of four different plasmids containing the reading frame and varying lengths of the 3′ non-coding regions of human DHFRase-specific cDNAs has revealed that the 3′ non-coding segments all are colinear in their corresponding portions. Furthermore, the data indicate that the cDNA of one of the plasmids is probably derived from the smallest of the three main human DHFRase messenger RNAs, the $0{\cdot}8 \times 10^3$ base (0·8 kb) mRNA, the cDNA of two others, from the 1·0 kb mRNA, and the cDNA of the fourth, from a longer mRNA. These results are consistent with the idea that the multiple forms of DHFRase mRNA in human cells derive from the same gene by different transcription or RNA-processing events. Moreover, the sequence comparison between the $\psi$HD1 and the different DHFRase cDNAs clearly indicates that, if an mRNA intermediate has participated in the formation of this pseudogene, a form of mRNA larger than the 1·0 kb mRNA, probably the 3·8 kb mRNA, must have been involved.

## 1. Introduction

Studies on eukaryotic genome organization in several laboratories have recently revealed the occurrence of copies of specific genes that, because of their structural alterations, are probably non-functional; indeed, they have not yet been found to be transcribed. The alterations in question range from gross deletions or insertions to small deletions, insertions or single base changes that result in frame shifts, formation of stop codons or perturbation of RNA processing. These apparently non-functional genes, which are usually referred to as pseudogenes, have been



0022-2836/83/170023-14 $03.00/0 

found either near the loci of the corresponding normal genes (Jacq *et al.*, 1977; Miller *et al.*, 1978; Proudfoot & Maniatis, 1980; Lacy & Maniatis, 1980; Jahn *et al.*, 1980; Cleary *et al.*, 1981) or dispersed to other regions of the genome, often in different chromosomes (Hollis *et al.*, 1982; Leder *et al.*, 1981). First observed in the 5 S RNA gene family of *Xenopus laevis* (Jacq *et al.*, 1977; Miller *et al.*, 1978), structurally altered copies have subsequently been described for a variety of structural genes, like the α-globin (Nishioka *et al.*, 1980; Vanin *et al.*, 1980; Proudfoot & Maniatis, 1980) and β-globin genes (Lacy & Maniatis, 1980; Jahn *et al.*, 1980; Cleary *et al.*, 1981) of several mammalian species, the human β-tubulin genes (Wilde *et al.*, 1982*a*,*b*), the mouse immunoglobulin genes (Hollis *et al.*, 1982), the human metallothionein genes (Karin & Richards, 1982), and the human small nuclear RNA genes (Van Arsdell *et al.*, 1981). The recently reported dispersed copies of sea urchin histone genes (orphons) probably also include pseudogenes (Childs *et al.*, 1981).

One interesting subclass of pseudogenes related to structural gene families is represented by genes from which one or more of the introns are clearly missing, as if they had been excised with the same precision and specificity that occurs in the splicing of the mRNA precursors (Nishioka *et al.*, 1980; Vanin *et al.*, 1980; Wilde *et al.*, 1982*a*,*b*; Hollis *et al.*, 1982; Karin & Richards, 1982). This feature and the observation that in several of these genes the homology with their normal homologues ends precisely at the poly(A) addition site, the sequence of the pseudogene in some cases continuing in a poly(A) tract, have led to the suggestion that an RNA transcript may be an intermediate in their formation. In particular, these genes, which have been designated "processed" genes, may have resulted from reverse transcription of a processed RNA into complementary DNA, followed by reintegration of this into the genome at a different site from that of the original gene. A cDNA† intermediate has also been thought to be involved in the dispersion in the human genome of members of the *Alu* gene family (Van Arsdell *et al.*, 1981; Jagadeeswaran *et al.*, 1981) and of the small nuclear RNA gene family (Van Arsdell *et al.*, 1981). In this paper, we report on the occurrence in the human genome of an intronless copy of the dihydrofolate reductase gene: this pseudogene exhibits multiple alterations that would make it non-functional for DHFRase production.

## 2. Materials and Methods

### (a) *Materials*

Restriction enzymes were from New England Biolabs. *Escherichia coli* DNA polymerase I from Boehringer–Mannheim and New England Nuclear. DNA polymerase I large fragment (Klenow) from New England Biolabs and New England Nuclear. benzoylated-naphthoylated DEAE(BND)-cellulose and DL-methotrexate (MTX, consisting of 50% of the L-form) from Sigma.

### (b) *Cell lines and methods of growth*

Suspension cultures of the S3 clonal strain of HeLa cells were grown as previously described (Amaldi & Attardi, 1968). The human cell line $VA_2$-B, an azaguanine-resistant

† Abbreviations used: cDNA, complementary DNA; DHFRase, dihydrofolate reductase; kb, $10^3$ bases or base-pairs.

subclone (Weiss *et al.*, 1968) of simian virus 40-transformed line WI-18-VA$_2$ (Pontén *et al.*, 1963) and its MTX-resistant derivative 6A3 (Masters *et al.*, 1982) were grown as previously described (Morandi *et al.*, 1982).

(c) *DNA preparation*

Genomic DNA was isolated by the method of Gross-Bellard *et al.* (1973). Cosmid and plasmid DNAs were in general prepared from 1 liter cultures by alkaline lysis (Birnboim & Doly, 1979) followed by CsCl/ethidium bromide centrifugation: the DNA of pUC9 plasmids containing fragments of cosmid DNA (see below) was purified by alkaline lysis of the bacteria and BND-cellulose chromatography of the DNA as described by Kiger & Sinsheimer (1969). Human sperm DNA was a gift from Jay Ellison (California Institute of Technology, Pasadena, California).

(d) *DNA transfer and hybridization*

DNA samples were digested to completion with restriction enzymes (completion of digestion of the genomic DNA was tested by adding lambda DNA to a portion of the reaction mixture and continuing the incubation until the lambda DNA gave a complete digestion pattern upon agarose gel electrophoresis), electrophoresed through an agarose gel, transferred to a nitrocellulose filter and hybridized with a particular probe, $^{32}$P-labeled by nick-translation as described by Morandi *et al.* (1982). The hybridization was carried out in 6 × SSC (SSC is 0·15 M-NaCl, 0·05 M-sodium citrate), 0·1% sodium dodecyl sulfate, 0·1% sodium pyrophosphate, 10 × Denhardt's solution (0·2% Ficoll, 0·2% polyvinylpyrrolidone, 0·2% bovine serum albumin), 10% dextran sulfate, 50 μg polycytidylic acid/ml, and 100 μg heat-denatured salmon sperm DNA/ml, for 15 h at 68°C. The filter was then washed once with 6 × SSC, 10 × Denhardt's solution, 0·1% sodium pyrophosphate, 0·1% sodium dodecyl sulfate, once with 2 × SSC, 0·1% sodium dodecyl sulfate, and 3 times with 0·5 × SSC, for 1 h each at 68 °C. The filter was finally dried and exposed to Kodak XAR-5 film with a Dupont Lightning Plus intensifying screen at −70°C.

(e) *Cosmid library screening*

A human cosmid DNA library (provided by Jay Ellison and Leroy Hood, California Institute of Technology, Pasadena, California) was constructed by preparing a partial *Mbo*I digest of human sperm DNA and ligating it to the *Bgl*II site of pTL5 (Hohn & Collins, 1980), according to previously published procedures (Lund *et al.*, 1982; Steinmetz *et al.*, 1982). This library was screened, by the method of Grosveld *et al.* (1981), with the insert of pHD84, $^{32}$P-labeled by nick-translation. *Eco*RI fragments of cosmid DNA were subcloned using the plasmid pUC9 (Vieira & Messing, 1982) as a vector.

(f) *DNA sequencing*

Individual restriction fragments were labeled at their 5′ or 3′ ends, and cleaved with an appropriate restriction endonuclease to produce fragments labeled at only one end, which were then isolated by polyacrylamide gel electrophoresis, as previously described by Morandi *et al.* (1982). The method of Maxam & Gilbert (1980) was used for sequencing.

## 3. Results

(a) *Distribution of DHFRase-specific sequences in human DNA*

Figure 1 shows the distribution of DHFRase-specific sequences in a sample of human sperm DNA and in the DNA from the human cell lines HeLa, VA2-B (Weiss *et al.*, 1968), and a MTX-resistant derivative of VA$_2$-B, 6A3 (Masters *et al.*, 1982). Similar amounts of the various DNA samples were digested with *Eco*RI, submitted to electrophoresis through an agarose slab gel, transferred to

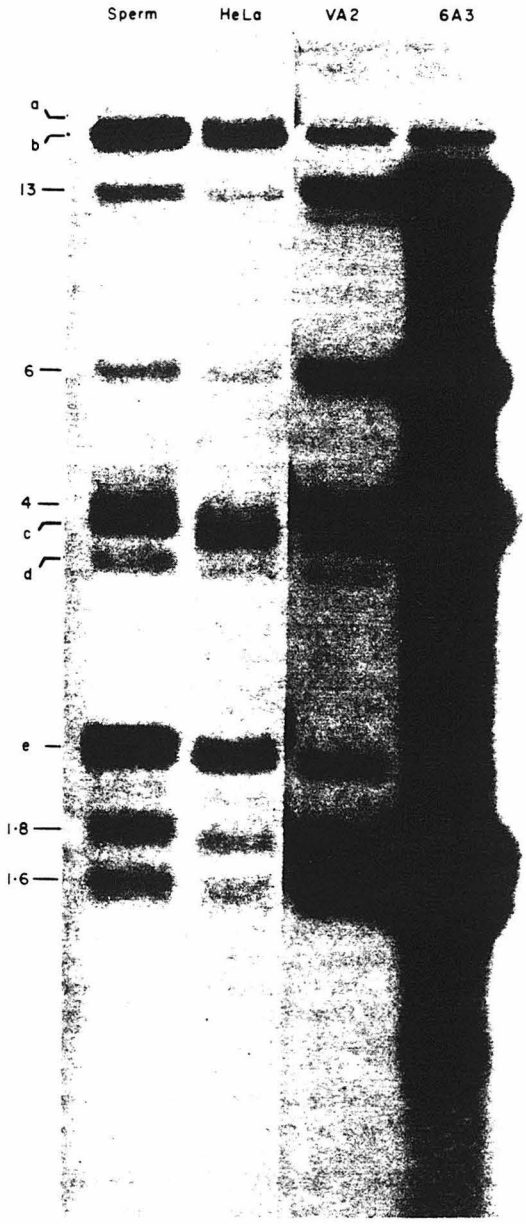



FIG. 1. Distribution of DHFRase-specific sequences in *Eco*RI-digested DNA from human sperm and from the human cell lines HeLa, $VA_2$-B and its MTX-resistant derivative 6A3. Samples (7 $\mu$g) of the sperm and HeLa DNAs and 5 $\mu$g samples of $VA_2$-B and 6A3 DNAs were digested to completion with *Eco*RI, submitted to electrophoresis through a 0·7% agarose gel, transferred to nitrocellulose and hybridized with the insert of pHD84 $^{32}$P-labeled by nick-translation. The numbers represent the sizes (kb) of the normal DHFRase gene fragments, which are amplified in $VA_2$-B and, to a higher degree, in 6A3 cells; the letters indicate fragments of putative variant DHFRase genes or pseudogenes (see the text).

nitrocellulose and hybridized with the insert of pHD84, $^{32}$P-labeled by nick-translation. This plasmid is a human DHFRase cDNA clone, which confers upon *E. coli* resistance to trimethoprim at 1 mg/ml (Morandi *et al.*, 1982). All lanes show two sets of hybridization bands. One set of bands, which correspond to fragments of sizes 13, 6, 4, 1·8 and 1·6 kb, appears to be more intense in the $VA_2$-B DNA sample, and much more so in the 6A3 DNA, than in the sperm and HeLa DNAs. We interpret these fragments as belonging to the normal DHFRase gene, which appears to be moderately amplified in $VA_2$-B cells and greatly amplified in 6A3 cells. The fragments described above have been recently isolated from a Ch4A library of 6A3 DNA and shown to be fragments of the amplified DHFRase gene and to contain the whole or almost the whole coding sequence (J. K. Yang, J. N. Masters & G. Attardi, unpublished data). The apparently lower amount of these fragments in the HeLa DNA as compared to the sperm DNA in the experiment shown in Figure 1 may be a consequence of the heteroploidy of this cell line. The other set of bands in Figure 1, designated a, b, c, d and e, correspond to fragments of sizes of 18, 17, 3·8, 3·5 and 2·3 kb, respectively, and do not show any increase in intensity in the $VA_2$-B and 6A3 DNA samples as compared to the sperm and HeLa DNAs, but rather a decrease. (Band a, which is very faint here in the $VA_2$-B and 6A3 DNA samples, appears clearly in other blots using DNA from HeLa and two human fibroblast lines (not shown).) These bands could represent fragments of allelic copies of the DHFRase gene, other variant DHFRase genes or pseudogenes. The strong intensity of some of these bands, i.e. b, c and e, suggested that they could correspond to intronless "processed" genes. Further analysis, as described below, has shown this interpretation to be correct at least for bands d and e.

(b) *Isolation of cosmid clones containing DHFRase-specific sequences*

During the screening of a cosmid library of sperm DNA (the same as that analyzed in Fig. 1) for the purpose of isolating cosmids containing the whole or most of the DHFRase gene, several cosmids were isolated that contained some of the non-amplified DHFRase-specific sequences discussed above. The library had been prepared with a partial *Mbo*I digest of human sperm DNA using the vector pTL5 (Hohn & Collins, 1980), and screened with the $^{32}$P-labeled insert of pHD84. Although ten positive clones were found, none of the cosmids contained any of the amplifiable DHFRase gene fragments shown in Figure 1: we have no explanation for this observation. Dot hybridization experiments showed that the DNA of nine of these cosmids hybridized with each of three subfragments of the insert of pHD84 corresponding to the 5′-end proximal, middle and 3′-end proximal portions (probes $\alpha$, $\beta$ and $\gamma$, respectively, in Fig. 2) (not shown). This is in contrast to the behavior of the normal DHFRase gene fragments in the genomic blots: all of these, in fact, hybridized with the whole pHD84 probe, while only the 13 kb, 1·6 kb and 4·0 kb fragments hybridized with the pHD43 probe: therefore, they contain sequences from the middle and 3′-end proximal segments of the insert of pHD84 (Fig. 2). These results indicate that these cosmids contained a segment corresponding to the whole or almost the whole DHFRase coding sequence. Upon

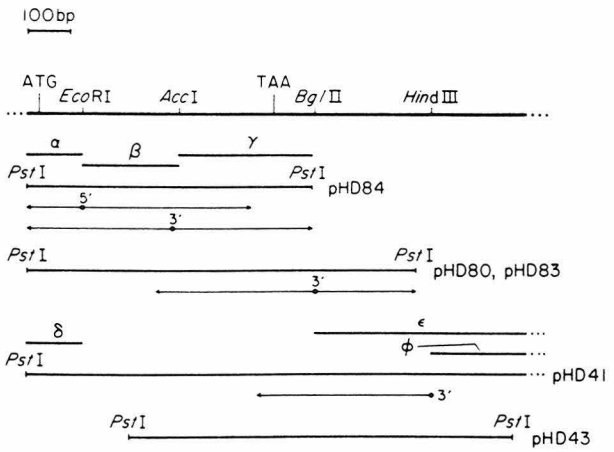



FIG. 2. Representation of the DHFRase cDNA probes used in the DNA and RNA transfer hybridization experiments and the strategy used for sequencing the DHFRase cDNAs. These cDNAs were prepared from poly(A)-containing polysomal RNA from the MTX-resistant cell line 6A3 and were cloned into the single *Pst*I site of pBR322 (Morandi *et al.*, 1982). The top line represents a portion of the cDNA of the 3·8 kb DHFRase mRNA (Morandi *et al.*, 1982): the positions of the initiator and terminator codons of the DHFRase reading frame and of the restriction sites used are indicated. Below this, the individual cDNAs used and their fragments have been aligned, and the sequencing strategy has been shown for each clone, the extent and direction of sequence determined being indicated by arrows. bp, base-pairs.

*Eco*RI digestion of the DNA from these cosmids and DNA transfer hybridization with a pHD84 probe, this segment was found to be located, in every case, in the largest fragment (9 to 28 kb).

The tenth cosmid, cosHD37, has a size of ~41 kb , and its *Eco*RI-digested DNA yielded, among others, a 3·5 kb and a 2·3 kb fragment, both of which hybridized with pHD84 (Fig. 3(a)). It seemed possible that the 3·5 kb and 2·3 kb fragments corresponded to the fragments of similar size designated d and e in the genomic blots of Figure 1. The identity of electrophoretic migration of the 3·5 kb and 2·3 kb fragments of cosHD37 with bands d and e, respectively, was confirmed in an experiment in which cosHD37 DNA and HeLa DNA were subjected to electrophoresis, alone or in a mixture, through an agarose slab gel, transferred to nitrocellulose and subjected to hybridization with a pHD84 probe (Fig. 3(b)).

cosHD37 was chosen for further analysis. In DNA transfer hybridization experiments (not shown), it was observed that the 5′-end proximal fragment of pHD41 (probe δ) hybridized exclusively with the 3·5 kb fragment of cosHD37, while pHD43 hybridized only with the 2·3 kb fragment of cosHD37. These and the previously discussed data were compatible with the idea that the 3·5 kb and 2·3 kb fragments were the two fragments of a variant DHFRase gene or a DHFRase pseudogene separated by the *Eco*RI site existing 89 bp downstream from the ATG in the cDNA coding sequence. This idea received support from the restriction map of cosHD37 DNA (Fig. 4). This showed that the two *Eco*RI fragments are linked, and that the positions of the *Acc*I and *Hinc*II sites relative to the *Eco*RI site, in the portion of the 2·3 kb fragment adjacent to the 3·5kb fragment, are very

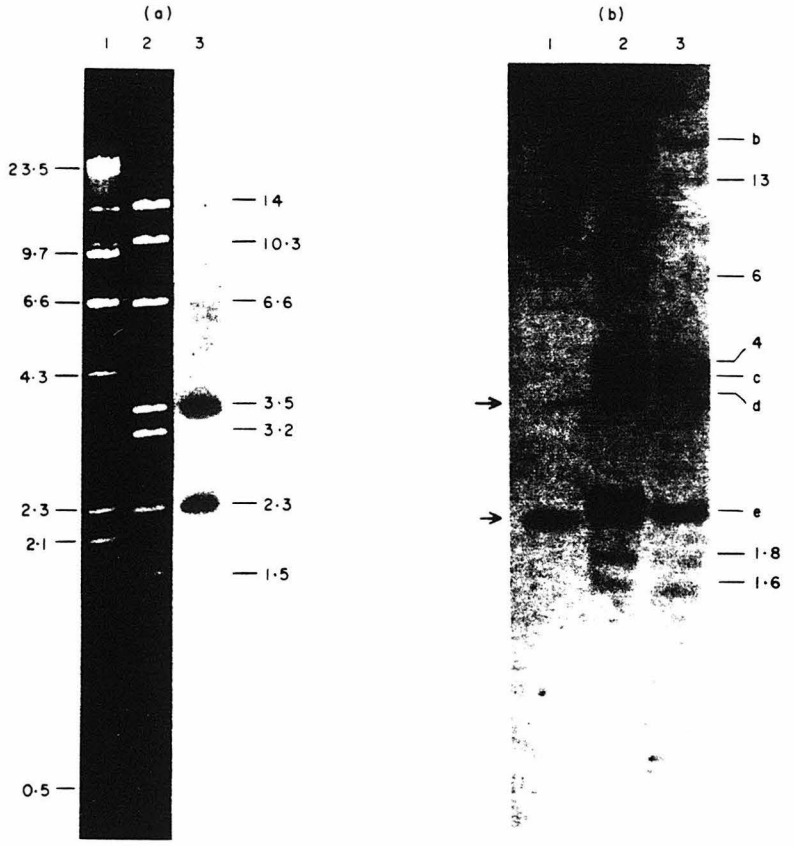



FIG. 3. (a) Detection of DHFRase-specific sequences in cosHD37 DNA. Lane 1. *Hin*dIII-digested λ DNA; lane 2. *Eco*RI-digested cosHD37 DNA; lane 3. autoradiogram of lane 2 after transfer of the DNA to nitrocellulose and hybridization to the nick-translated insert of pHD84. (b) DNA transfer hybridization analysis of *Eco*RI-digested HeLa DNA and cosHD37 DNA using a DHFRase cDNA probe. *Eco*RI-digested DNAs were submitted to electrophoresis through a 0·5% agarose gel, transferred to nitrocellulose, hybridized to $^{32}$P-labeled pHD84. Lane 1. cosHD37 DNA, 0·6 ng; lane 2. cosHD37 DNA, 0·6 ng, plus HeLa DNA, 5 μg; lane 3. HeLa DNA, 5 μg. The faint band at 14 kb in lane 3 of (a), and lanes 1 and 2 of (b), is due to hybridization of contaminating pBR322 sequences in the probe with the pBR322 sequences present in the cosmid DNA (see Fig. 4).

similar to those of the same sites in the insert of pHD84. Conclusive evidence that cosHD37 contains a human DHFRase pseudogene (designated ψHD1) has come from a DNA sequence analysis of portions of the 3·5 kb and 2·3 kb fragments of this cosmid.

### (c) *DNA sequence analysis of 3′ non-coding segments of DHFRase cDNAs*

Figure 5 shows the previously determined sequence (sense strand) of the DHFRase reading frame and the upstream region in the insert of pHD84 (Morandi *et al.*, 1982; Masters & Attardi, 1982), as well as the newly determined

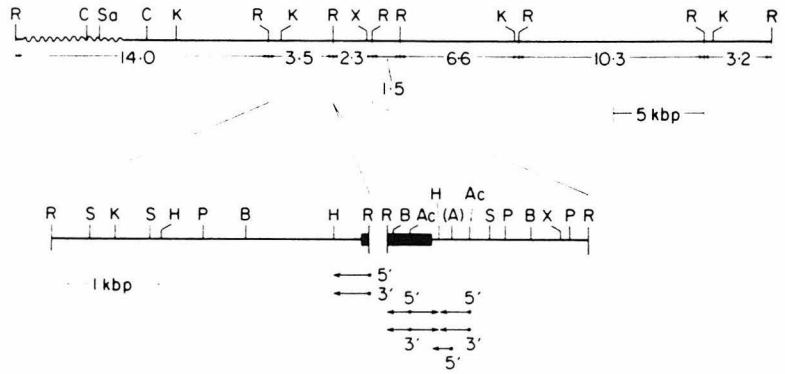



FIG. 4. Restriction map of cosHD37 DNA and strategy used for sequencing the ψHD1 pseudogene contained in cosHD37. The upper line shows the restriction map of cosHD37 DNA. The lengths of the *Eco*RI fragments are shown below the map in kbp ($10^3$ base-pairs). The wavy line indicates pTL5 vector sequences. The lower line shows the detailed restriction map of the two *Eco*RI fragments, with sizes of 3·5 and 2·3 kbp, which were subcloned separately using pUC9 as a vector. The filled bar indicates the limits of the DNA sequence homologous to the normal DHFRase coding sequence, as shown in Fig. 5. The restriction enzymes used are designated as follows: Ac, *Acc*I; B, *Bgl*II; C, *Cla*I; H, *Hin*cII; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; Sa, *Sal*I; S, *Sph*I; X, *Xho*I. In the lower part of the figure, the arrows show the direction and extent of sequence determined. (A) is an *Alu*I site, which was deduced from sequence data and was used to sequence through the *Hin*cII site immediately upstream.

sequence of the DNA segments lying downstream from the reading frame present in pHD84, pHD80, pHD83 and pHD41. The segments sequenced and the sequencing strategy used are shown in Figure 2. As shown in Figure 5, the sequences of the inserts of pHD84, pHD80, pHD83 and pHD41 were found to be colinear in their corresponding segments, as determined relative to the reading frame; this, in all the inserts, lies near the 5′ end (Morandi *et al.*, 1982). Furthermore, the sequence of the insert of pHD84 was found to contain a stretch of 12 A residues, starting at position 648 (just preceding the poly(C) tail used for cloning). Similarly, the inserts of both pHD80 and pHD83 were found to contain a stretch of A residues starting at position 900 (52 residues long in pHD80 and >70 residues long in pHD83, just preceding the poly(C) tail of these plasmids). A reasonable interpretation of these results is that the cDNA of pHD84 is derived from reverse transcription of the 0·8 kb mRNA, while the cDNAs of pHD80 and pHD83 are derived from the 1·0 kb mRNA (Morandi *et al.*, 1982). The cDNA of pHD41 must have been derived from reverse transcription of a longer mRNA, probably the 3·8 kb mRNA (Morandi *et al.*, 1982). The above conclusion and the colinearity of the corresponding segments of the three types of cDNAs support the idea that the three main forms of mRNA previously detected are derived from transcription of the same sequence (although not necessarily of the same template) extending over different lengths. In agreement with this idea are the results of RNA transfer hybridization experiments using the polysomal poly(A)-containing RNA from 6A3 cells. When the whole plasmid pHD41 was used as a probe, all three species of mRNA hybridized. When a 3′-end proximal subfragment of this plasmid was used (probe ϵ in Fig. 2), beginning at the *Bgl*II site that lies 5 base-pairs downstream from the poly(A) addition site identified in


FIG. 5. DNA sequences of human DHFRase cDNA clones and of $\psi$HD1. The upper sequence represents the combined sequences (sense strand) of the segments of pHD84, pHD80, pHD83 and pHD41 indicated in Fig. 2, extending from the 5′ nucleotide adjacent to the poly(C) tail of pHD84 to the *Hind*III site of pHD41, with the numbering starting at the ATG initiator codon. The corresponding portions of the sequences were identical. The lower sequence is that of the segment of cosHD37 that is homologous to the sequenced cDNA. Gaps have been introduced into the cDNA sequence and the cosHD37 sequence to maximize the homology between them. (·) represents homologous nucleotides. The long unbroken arrows indicate the positions of the exon junctions of the human DHFRase gene identified so far, the long broken arrows mark the positions of the other splice points expected from the probable homology with the mouse DHFRase gene, and the short arrows indicate sites where a stretch of A residues occurs in the cDNA indicated, just prior to the poly(C) tail. The translation initiation and termination codons of the cDNA are boxed, and an ochre termination codon brought into frame by a deletion in the cosmid sequence is underlined.

the cDNA of pHD84, and terminating at the pBR322 *Eco*RI restriction site (which lies downstream from the DHFRase coding sequence), the 0·8 kb mRNA did not hybridize, while the 1·0 kb and the 3·8 kb mRNAs did. When a different subfragment was used (probe $\phi$ in Fig. 2), beginning at the *Hin*dIII site of pHD41 and ending at the pBR322 *Eco*RI site, only the 3·8 kb mRNA hybridized (not shown).

Previous work has shown that the human DHFRase coding sequence has an 89% nucleotide sequence homology and an 89% amino acid sequence homology to the mouse DHFRase reading frame (Masters & Attardi, 1982). A homology of 25% exists between the 3′ non-coding region of the mouse 1·6 kb mRNA (Setzer *et al.*, 1982) and the available 3′ non-coding sequence of the human gene.

### (d) *Identification of a human DHFRase pseudogene ($\psi$HD1)*

Figure 5 also shows the comparison of the cDNA sequence described above with the homologous region of $\psi$HD1. Gaps have been introduced into the cDNA sequence and the $\psi$HD1 sequence to maximize the homology between them. The sequence of $\psi$HD1 begins in the 3·5 kb *Eco*RI fragment and continues in the 2·3 kb *Eco*RI fragment, extending beyond the poly(A) addition site identified in the cDNA of pHD80 and pHD83. It exhibits an overall 92% homology with the cDNA sequence, with no significant difference in the degree of divergence between the region corresponding to the DHFRase coding sequence (93% homology) and the 3′ non-coding segment (90% homology). Several small insertions, deletions and single base changes are observed, distributed fairly uniformly throughout the sequence. In particular, there are five insertions, between nucleotides −16 and −15, 364 and 365, 640 and 641, 658 and 659, and 920 and 921, of the cDNA sequence; and two deletions, corresponding to nucleotides 243 and 301 to 304 of the cDNA sequence. The deletion of nucleotide 243 causes a frameshift in the DHFRase coding region, which results in an ochre termination codon beginning at a position corresponding to nucleotide 293 in the cDNA sequence.

The deletion of nucleotide 243 in $\psi$HD1 is particularly interesting since this is the first nucleotide of the fourth exon of the mouse DHFRase gene (Crouse *et al.*, 1982). In order to determine whether the missing nucleotide also corresponds to a splice junction of human DHFRase gene, the exon region of the cloned 1·6 kb fragment of the human gene (Fig. 1), which from restriction mapping data was expected to contain the coding sequence corresponding to the fourth exon in the mouse gene (Yang *et al.*, unpublished data), was sequenced. The sequence data clearly confirmed this expectation, thus indicating that the 3/4 and 4/5 exon junctions of the mouse gene correspond precisely to splice junctions in the human gene (Yang *et al.*, unpublished data). The sequence data of Figure 5 show that, apart from the missing nucleotide 243, the region immediately surrounding the two splice junctions in the human gene (indicated by long, unbroken arrows) shows perfect homology to the corresponding segments of the $\psi$HD1 pseudogene. The normal gene and the $\psi$HD1 sequences also show perfect homology to each other in the sites corresponding to the other exon junctions identified in the

mouse gene (Setzer *et al.*, 1982; Crouse *et al.*, 1982) (indicated by broken arrows in Fig. 5).

## 4. Discussion

The DHFRase pseudogene identified in the present work in the sperm DNA from a human individual and in the genomes from two human established cell lines has revealed the typical structure of an intronless gene. The intervening sequences that interrupt the reading frame of the human DHFRase gene (Yang *et al.*, unpublished data) are missing. The precise positions of two of these intervening sequences have been identified (Yang *et al.*, unpublished data), and they appear to coincide with the sites of two introns in the mouse gene (Crouse *et al.*, 1982). Assuming that the other introns also occur at the same positions in the human DHFRase reading frame as in the mouse DHFRase coding sequence, as seems likely, it appears that the apparent "excision" of introns from ψHD1 has been absolutely precise. The exon junction at nucleotides 242/243 of the pHD84 cDNA insert is a possible exception, since a deletion of the first nucleotide of the downstream exon (nucleotide 243) appears to have occurred in the pseudogene.

The portion of the ψHD1 sequence corresponding to the DHFRase reading frame has a very high degree of homology to it (~93%, excluding the deletions and insertions). However, the presence of small deletions and insertions in the pseudogene has introduced frameshifts, the first one beginning at a position corresponding to nucleotide 222 of the normal reading frame, and stop codons, the first of which is at a position corresponding to nucleotide 271 of the normal reading frame. These alterations would make ψHD1 non-functional, at least as a DHFRase coding gene.

Particularly interesting is the relationship of ψHD1 to the three main human DHFRase mRNAs previously identified, which differ in the length of their 3′ untranslated regions (Morandi *et al.*, 1982). The sequence analysis of the 3′ non-coding segments of the cDNA inserts of plasmids pHD84, pHD80 and pHD83 has provided evidence indicating that the first cDNA is derived from the 0·8 kb mRNA, and the other two from the 1·0 kb mRNA. Furthermore, the sequences of these cDNAs and of the pHD41 cDNA were found to be colinear in their corresponding segments. These observations and the results of RNA transfer hybridization experiments are consistent with the idea that the two forms of mRNA, and probably also the largest (the 3·8 kb mRNA), are derived from the same gene by different transcription or RNA-processing events. A similar conclusion has been reached concerning the origin of the multiple DHFRase mRNAs in the mouse system (Setzer *et al.*, 1980, 1982). The hexanucleotide A-A-U-A-A-A, which is found 15 to 20 nucleotides upstream from the polyadenylation site in most eukaryotic mRNAs (Proudfoot & Brownlee, 1976) and is considered to be a recognition site for polyadenylation (Fitzgerald & Shenk, 1981), is absent in the 0·8 and 1·0 kb mRNAs, although possible variants of this sequence are present at the expected positions. This observation indicates that, as previously suggested for the mouse system (Setzer *et al.*, 1982), the hexanucleotide A-A-U-A-A-A is not required for polyadenylation of at least two of the human DHFRase mRNAs.

The sequence of ψHD1 downstream from the segment corresponding to the DHFRase reading frame also exhibits a high degree of homology (~90%) to the 3′ untranslated region of the cDNAs sequenced here, up to a position past the putative poly(A) addition site of the 1·0 kb mRNA. However, it does not reveal any evidence for the presence of a poly(A) tract, although it cannot be excluded that a poly(A) tract may be present somewhere downstream. In the 5′ flanking region of ψHD1, we have likewise no evidence, as yet, as to the existence of homology to the initiation point(s) of transcription of the DHFRase mRNAs. In spite of this, the structure of ψHD1 strongly suggests that an RNA intermediate has participated in its formation, as suggested previously for other intronless pseudogenes (Nishioka *et al.*, 1980; Vanin *et al.*, 1980; Wilde *et al.*, 1982*a*,*b*; Hollis *et al.*, 1982; Karin & Richards, 1982). If the RNA intermediate postulated here is derived from reverse transcription of a DHFRase mRNA, a form of mRNA larger than the 1·0 kb mRNA, probably the 3·8 kb mRNA, must have been involved. Recent preliminary evidence indicates that ψHD1 is located in a different chromosome from that containing the normal DHFRase gene (B. Maurer, P. Barker, J. Masters, P. d'Eustachio, F. Ruddle & G. Attardi, unpublished observations), a result that is consistent with the idea that an RNA intermediate, whether an mRNA or a retrovirus genome copy, was responsible for its formation and translocation. Whatever event led to the formation of ψHD1, the limited degree of divergence from the DHFRase coding sequence suggests that it must have been a recent one. The great divergence between human and mouse DHFRase in the 3′ non-coding sequence places the origin of this pseudogene in evolutionary time after the separation of the lines leading to primates and rodents.

ψHD1 is probably not the only DHFRase pseudogene present in the human genome. The preliminary evidence obtained on the structure of several cloned fragments isolated from the cosmid library suggests that the other bands giving a hybridization signal with a pHD84 probe in the genomic blots analyzed here (a, b and c in Fig. 1) could represent allelic DHFRase genes, other variant genes or pseudogenes. The existence of allelic DHFRase genes with different restriction patterns and specifying protein products of different electrophoretic mobilities has been demonstrated in Chinese hamster cells (Flintoff *et al.*, 1982; Lewis *et al.*, 1982).

The significance of the moderate DHFRase gene amplification that has occurred in the human cell line $VA_2$-B in the absence of methotrexate selection is unknown. This cell line is a subclone resistant to 8-azaguanine (Weiss *et al.*, 1968) derived from the line WI-18-$VA_2$ transformed by simian virus 40 (Pontén *et al.*, 1963). The DHFRase gene amplification could have occurred at any time during the life of this cell line. Besides the DHFRase gene dosage, the level of the three DHFRase mRNAs is also increased in these cells (Morandi *et al.*, 1982); by contrast, the DHFRase activity appears to be only slightly increased (Masters *et al.*, 1982).

These investigations were supported by National Institutes of Health grants GM11726 and T32 GM07616. We are very grateful to A. Chomyn for helpful discussions and to J.

Ellison for a gift of human sperm DNA and to J. Ellison and L. Hood for providing us with the human DNA cosmid library.

## REFERENCES

Amaldi, F. & Attardi, G. (1968). *J. Mol. Biol.* **33**, 737–755.
Birnboim, H. C. & Doly, J. (1979). *Nucl. Acids Res.* **7**, 1513–1523.
Childs, G., Maxson, R., Cohn, R. H. & Kedes, L. (1981). *Cell.* **23**, 651–663.
Cleary, M. L., Schon, E. A. & Lingrel, J. B. (1981). *Cell.* **26**, 181–190.
Crouse, G. F., Simonsen, C. C., McEwan, R. N. & Schimke, R. T. (1982). *J. Biol. Chem.* **257**, 7887–7897.
Fitzgerald, M. & Shenk, T. (1981). *Cell.* **24**, 251–260.
Flintoff, W. F., Weber, M. K., Nagainis, C. R., Essani, A. K., Robertson, D. & Salser, W. (1982). *Mol. Cell. Biol.* **2**, 275–285.
Gross-Bellard, M., Oudet, P. & Chambon, P. (1973). *Eur. J. Biochem.* **36**, 32–38.
Grosveld, F. G., Hans-Henrik, M. D., de Boer, E. & Flavell, R. A. (1981). *Gene.* **13**, 227–237.
Hohn, B. & Collins, J. (1980). *Gene.* **11**, 291–298.
Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D. & Leder, P. (1982). *Nature (London).* **296**, 321–325.
Jacq, C., Miller, J. R. & Brownlee, G. G. (1977). *Cell.* **12**, 109–120.
Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. (1981). *Cell.* **26**, 141–142.
Jahn, C. L., Hutchinson, C. A. III, Phillips, S. J., Weaver, S., Haigwood, N. L., Voliva, C. F. & Edgell, M. H. (1980). *Cell.* **21**, 159–168.
Karin, M. & Richards, R. I. (1982). *Nature (London).* **299**, 797–802.
Kiger, J. A. & Sinsheimer, R. L. (1969). *J. Mol. Biol.* **40**, 467–490.
Lacy, E. & Maniatis, T. (1980). *Cell.* **21**, 545–553.
Leder, A., Swan, D., Ruddle, F., D'Eustachio, P. & Leder, P. (1981). *Nature (London).* **293**, 196–200.
Lewis, J. A., Davide, J. P. & Melera, P. W. (1982). *Proc. Nat. Acad. Sci. U.S.A.* **79**, 6961–6965.
Lund, T., Grosveld, F. G. & Flavell, R. A. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 520–524.
Masters, J. N. & Attardi, G. (1983). *Gene.* **21**, 59–63.
Masters, J. N., Keeley, B., Gay, H. & Attardi, G. (1982). *Mol. Cell Biol.* **2**, 498–507.
Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–559.
Miller, J. R., Cartwright, E. M., Brownlee, G. G., Fedoroff, N. V. & Brown, D. D. (1978). *Cell.* **13**, 717–725.
Morandi, C., Masters, J. N., Mottes, M. & Attardi, G. (1982). *J. Mol. Biol.* **156**, 583–607.
Nishioka, Y., Leder, A. & Leder, P. (1980). *Proc. Nat. Acad. Sci. U.S.A.* **77**, 2806–2809.
Pontén, J., Jensen, F. & Koprowski, H. (1963). *J. Cell. Comp. Physiol.* **61**, 145–154.
Proudfoot, N. J. & Brownlee, G. G. (1976). *Nature (London).* **263**, 211–214.
Proudfoot, N. J. & Maniatis, T. (1980). *Cell.* **21**, **537**–544.
Setzer, D. R., McGrogan, M., Nunberg, J. H. & Schimke, R. T. (1980). *Cell.* **22**, 361–370.
Setzer, D. R., McGrogan, M. & Schimke, R. T. (1982). *J. Biol. Chem.* **257**, 5143–5147.
Steinmetz, M., Astar, W., Minard, K. & Hood, L. (1982). *Cell.* **28**, 489–498.
Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T. & Gesteland, R. F. (1981). *Cell.* **26**, 11–17.
Vanin, E. F., Goldberg, G. I., Tucker, P. W. & Smithies, O. (1980). *Nature (London).* **286**, 222–226.
Vieira, J. & Messing, J. (1982). *Gene.* **19**, 259–268.
Weiss, M. C., Ephrussi, B. & Scaletta, L. J. (1968). *Proc. Nat. Acad. Sci., U.S.A.* **59**, 1132–1135.

Wilde, C. D., Crowther, C. E. & Cowan, N. J. (1982*a*). *Science*, **217**, 549–552.
Wilde, C. D., Crowther, C. E., Cripe, T. P., Lee, M. G.-S. & Cowan, N. J. (1982*b*). *Nature (London)* **297**, 83–84.

*Edited by S. Brenner*

*Note added in proof*: Recently, the occurrence in the human genome of two DHFR intronless pseudogenes distinct from that described here has been reported (Chen *et al.* (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 7435–7439).

# CHAPTER 6

## Human Dihydrofolate Reductase Gene Organization: Extensive Conservation of the (GC)-Rich 5' Noncoding Sequence and Strong Intron Size Divergence from Homologous Mammalian Genes

# Human Dihydrofolate Reductase Gene Organization: Extensive Conservation of the (GC)-Rich 5' Noncoding Sequence and Strong Intron Size Divergence from Homologous Mammalian Genes

Jiyoung K. Yang, Jeffrey N. Masters and Giuseppe Attardi

Published in the Journal of Molecular Biology

## Abstract

The complete human dihydrofolate reductase (DHFR) gene has been cloned from four recombinant lambda libraries constructed with the DNA from a methotrexate resistant human cell line with amplified DHFR genes. The detailed organization of the gene has been determined by restriction mapping of the cloned fragments and DNA sequencing of all the protein coding regions and adjacent intron segments, and shown to correspond to that of the native human DHFR gene. The gene spans a length of approximately 29 kb from the ATG initiator codon to the end of the 3' untranslated region, and contains 5 introns which interrupt the protein coding sequence. The number and positions of introns are identical to those found in the mouse gene. By contrast, the size of the homologous introns, with the exception of the first one, varies greatly, up to several fold, in the genes from man, mouse and Chinese hamster; the intron sequences also exhibit a great divergence, except in the junction regions. A striking sequence homology, extending over several hundred nucleotides, exists between the human and mouse gene 5' non-coding regions. These regions are characterized by an unusually high G+C content, 72 and 66% in the human and mouse gene respectively, which is maintained in the first coding segment and first intron, and sharply contrasts with the relatively low G+C content, ~40%, of the remainder of the gene.

## Introduction

Recent work from this laboratory has led to the isolation of several human cell variants, derivatives of the cell lines HeLa BU25 and $VA_2$-B, which are resistant to high concentrations of methotrexate (MTX) as a result of the hyperproduction of dihydrofolate reductase (DHFR), the target enzyme of the drug (Masters *et al.*, 1982). These variants have been characterized as to the stability of their phenotype in the absence of selective pressure and to their chromosomal constitution (Masters *et al.*, 1982), and two of the variants, 10B3 and 6A3, in the physical and enzymatic properties of the hyperproduced DHFR (Morandi and Attardi, 1981) and in the pattern of transcription (Morandi *et al.*, 1982). Furthermore, the cloning and characterization of the DHFR-specific cDNAs derived from the 6A3 cell line (Morandi *et al.*, 1982; Masters and Attardi, 1983; Masters *et al.*, 1983) have permitted an analysis of the distribution of the DHFR-specific sequences in the human genome by DNA transfer-hybridization experiments, the identification of a DHFR pseudogene (Masters *et al.*, 1983), and the chromosomal localization of the human native DHFR gene and three pseudogenes (Maurer *et al.*, 1983, and unpublished data).

In this paper, we describe the organization of the complete human DHFR gene and compare it to that of the homologous genes from mouse (Crouse *et al.*, 1982) and Chinese hamster sources (Carothers *et al.*, 1983; Milbrandt *et al.*, 1983). Although the three mammalian genes have a similar size and a similar overall structure as concerns the number and positions of the introns, there has been a strong divergence in the size and sequence of the introns, especially marked for the fifth intron. By contrast, a striking sequence homology has been detected in the 5' non-coding regions of the human and mouse genes, which have an unusually high G+C content: this homology extends up to at least 327 bp upstream of the initiator codon.

## Materials and Methods

### (a) Materials.

Restriction enzymes and *Bal*31 were from New England Biolabs, Bethesda Research Laboratories, or Boehringer Mannheim; T4 DNA ligase from Bethesda Research Laboratories; *Escherichia coli* DNA polymerase I from Boehringer Mannheim; DNA polymerase I large fragment (Klenow) from New England Biolabs or New England Nuclear; ($\alpha^{32}$P)-dNTPs from Amersham; BND-Cellulose from Sigma.

### (b) Cell Lines and Chromosome Isolation.

The human cell line 6A3 is a MTX-resistant variant derived from $VA_2$-B cells (Masters *et al.*, 1982). Suspension cultures of 6A3 cells were grown as previously described (Morandi *et al.*, 1982), and arrested in metaphase by treatment with vinblastin sulfate (50 ng/ml) for 20 hr. Chromosomes of normal size range (referred to in this paper as "normal" chromosomes) and large and small double minute chromosomes (indicated as large and small double minutes) were separated by differential centrifugation at 4,100 xg for 30 min, at 16,300 g for 30 min and at 70,000 g for 60 min, respectively, in a high pH buffer system (Wray *et al.*, 1972).

### (c) Preparation of $C_0t$ Fractionated DNA.

HeLa cell DNA was isolated by the method of Gross-Bellard *et al.*, (1973), sheared to an average size of ~600 nt using a Virtis homogenizer (Britten *et al.*, 1974), denatured, allowed to reanneal to $C_0t$ 100 to renature the bulk of the repetitive sequences (Schmid and Deininger, 1975), and then subjected to hydroxylapatite chromatography to separate the dsDNA fraction, which represented 44% of the total, as measured by the absorbance at 260 nm. This DNA was then tested to determine the minimum amount needed to quench the signal from $^{32}$P-labeled probes containing repetitive elements. It was found

that addition of 10 μg/ml of $C_0t$ 100 DNA to the hybridization mixture in the plaque screening assays was sufficient to obtain this effect.

**(d) λ Library Construction and Screening.**

Cha4A libraries were constructed as previously described (Maniatis *et al.*, 1978), using sucrose gradient fractionated partial *Eco*RI digests of DNAs from 6A3 chromosomes or from large double minutes or small double minutes. λJ1 (kindly provided by J. Mullins) was derived from λL47.1 (Loenen and Brammar, 1980) by replacement of the 1.37 kb and 2.83 kb *Eco*RI-*Bam*HI fragments with the 100 bp *Eco*RI-*Bam*HI polylinker sequence [derived from πVX; (Maniatis *et al.*, 1982)]. A partial *Mbo*I digest of 6A3 chromosomal DNA was fractionated by sucrose gradient centrifugation and cloned into the λJ1 arms generated by *Bam*HI. The libraries were screened by *in situ* plaque hybridization (Benton and Davis, 1977) with pHD84, a plasmid which contains a human DHFR cDNA segment comprising the whole coding sequence (Morandi *et al.*, 1982), or with pBH31R1.8 or pUH112PR1.3 (two plasmids carrying human DHFR gene fragments), $^{32}$P-labeled by nick-translation. The recombinant phages were grown in DP50supF (Blattner *et al.*, 1977), and DNAs were prepared as previously described (Williams *et al.*, 1977).

**(e) Subcloning into Plasmids.**

DNA fragments were isolated by electrophoresis through a low melting (Seaplaque) agarose gel, and subcloned into pBR325 (Bolivar, 1978) or pUC9 (Vieira and Messing, 1982), using as a host the *E. coli* B derivative HB101 (Boyer and Roulland-Dussoix, 1969) or, respectively, TB1 [a JM83 (Messing, 1979) derivative deficient in the host restriction system and cured of a P1 prophage, T. O. Baldwin, personal communication]. The plasmid pUH31PRO.8 was constructed by ligating a *Pst*I-*Eco*RI digest of pBH31R1.8 to a *Pst*I-*Eco*RI digest of pUC9 [which had been spermine precipitated (Hoopes and McClure, 1981) to remove the small linker fragment]. A *Pst*I digest of pUH31PR0.8 was digested

for various lengths of time with the exonuclease *Bal*31 (Lau and Gray, 1979), treated with the large fragment of DNA polymerase I to make blunt ends (Maniatis *et al.*, 1982), then digested with *Eco*RI and ligated to a *Sma*I-*Eco*RI digest of pUC9 to produce pUH31Δ7 and pUH31Δ16. The recombinant clones were screened either by restriction enzyme analysis of DNA rapidly isolated from randomly picked colonies (Birnboim and Doly, 1979) or by colony hybridization (Grunstein and Hogness, 1975). Large-scale preparations of plasmid DNAs were made by the alkaline lysis method (Birnboim and Doly, 1979), followed by benzoylated-naphtoylated-DEAE (BND)-cellulose chromatography (Kiger and Sinsheimer, 1969) or by CsCl-ethidium bromide gradient centrifugation.

**(f) DNA Transfer and Hybridization.**

DNA samples were digested with restriction enzymes as suggested by the vendors, fractionated by agarose gel electrophoresis, transferred to a nitrocellulose filter and hybridized with an appropriate $^{32}$P-labeled probe as previously described (Masters *et al.*, 1983). $C_0t$ fractionated HeLa cell DNA was used in some experiments as carrier DNA to quench the signal given by repetitive elements in particular probes.

**(g) DNA Sequencing.**

Individual plasmid DNAs were digested with a restriction enzyme cutting within the region of interest, and the ends thus generated were either 5'-labeled, using T4 polynucleotide kinase and ($\gamma$-$^{32}$P) ATP following treatment with calf intestinal alkaline phosphatase, or 3'-labeled, using the large fragment (Klenow) of *E. coli* DNA polymerase I and an appropriate ($\alpha$-$^{32}$P)-dNTP: this approach allowed the sequencing of both strands from the same end of the fragment. The DNA was then digested with a second enzyme to produce the fragment of interest labeled at only one end. The $^{32}$P-labeled fragments were in general purified by polyacrylamide gel electrophoresis (Maxam and Gilbert, 1980); however, when the second digestion resulted in an unwanted $^{32}$P-labeled fragment

less than 50 bp long, the DNA was spermine precipitated (Hoopes and McClure, 1981) to remove the bulk of the small fragment, and then sequenced directly. The sequencing reactions were carried out as described by Maxam and Gilbert (1980).

## Results

### (a) Construction and Screening of DNA Libraries from a Human Cell Line with Amplified DHFR Genes.

6A3, a MTX-resistant human cell line derived from $VA_2$-B cells (Masters *et al.*, 1982), contains amplified DHFR genes which are in part associated with "normal" chromosomes, in part with double minute chromosomes (Maurer and Attardi, unpublished data). Because of the existence of amplified genes in the two types of chromosomal structures, the "normal" chromosomes, the large double minute chromosomes and the small double minute chromosomes were separated by differential centrifugation and utilized independently to construct λCha4A libraries. About $1.0 \times 10^6$ recombinant phages from the library of chromosomal DNA, $1.7 \times 10^5$ phages from the library of large double minute DNA and $7.5 \times 10^4$ phages from that of small double minute DNA were screened with $^{32}$P-labeled pHD84, a plasmid containing a cDNA segment comprising the entire human DHFR coding region (Morandi *et al.*, 1982), by *in situ* plaque hybridization; ~0.2% of the phages from each of the three libraries gave positive signals with the probe. Ninety-eight positive clones (62 from the chromosomal DNA library, 22 from the large double minute DNA library and 14 from the small double minute DNA library) were analyzed by DNA transfer hybridization experiments using the pHD84 probe for the presence of the *Eco*RI fragments containing the DHFR coding sequence, which are found in digests of human DNA and which are amplified in 6A3 cells (Masters *et al.*, 1983): these fragments have sizes of 13, 6, 4, 1.8 and 1.6 kb. Ten groups of phages differing in the inserts were found among the 98 positive clones. Seven of the 10 groups of phages could be shown to

contain in their inserts one or two of the above listed *Eco*RI fragments reactive with pHD84, and could be classified as follows: 13 kb, 13 kb and 4 kb, 4 kb and 1.6 kb, 1.6 kb, 6 kb, 6 kb and 1.8 kb, and 1.8 kb. The remaining three groups of recombinant phages contained rearranged DHFR gene fragments producing abnormal *Eco*RI digestion products, and will be described below. The recombinant phages which contained the 6 kb or the 6 kb and 1.8 kb *Eco*RI fragments turned out to be very unstable and to undergo rearrangement during phage growth, so that they could not be analyzed further.

Figure 1A shows the alignment of the phage DNA inserts containing *Eco*RI fragments of the amplified DHFR gene. One can see that the inserts cover a stretch of >35 kb interrupted by a gap. We were unable to identify any stable recombinant phage whose insert overlapped the inserts of λCha4A-$DM_L$31 and λCha4A-$DM_L$29 among a total of $1.3x10^6$ recombinants analyzed. It was conceivable that by a partial digestion of genomic DNA with a more frequently cutting enzyme, one might split the sequences in a way that would allow their cloning (a situation which has indeed occurred in the case of human mtDNA, unpublished results). Therefore, we partially digested the chromosomal fraction of 6A3 DNA with *Mbo*I, fractionated the DNA on a sucrose gradient and ligated the 15-20 kb fraction to the purified arms of *Bam*HI digested λJ1 DNA. About $5x10^4$ recombinant phages were screened first with $^{32}$P-labeled pBH31R1.8, a plasmid carrying the 1.8 kb *Eco*RI genomic fragment which lies next to the gap between the inserts of λCha4A-$DM_L$31 and λCha4A-$DM_L$29 (Fig. 2A), and then with $^{32}$P-labeled pHD84. The DNA from two phages gave positive signals with both probes [λJ1-CM15, Figure 1B, and λJ1-CM12, a phage containing rearranged DHFR gene sequences (see below)], the DNA from 13 phages hybridized only with pBH31R1.8 (not shown), and the DNA from 12 phages hybridized only with pHD84 [three phages containing normal DHFR gene sequences (λJ1-CM112, λJ1-CM113, λJ1-CM114, Fig. 1B), four phages containing rearranged DHFR gene fragments, to be described below, and four unstable phages (not shown)]. The inserts of λJ1-CM15 and λJ1-CM112 reduced somewhat, at each end, the

gap existing in the restriction map of the DHFR gene. In a final attempt to bridge the remaining gap, ~$1.2\times10^5$ recombinant λJ1 recombinant phages were screened with a probe containing the 1.3 kb 5'-end sequences of λJ1-CM112 subcloned in pUC9 (pUH112PR1.3, Fig. 2A). The 1.3 kb fragment was shown to contain repetitive sequences by its hybridization to nick translated HeLa DNA (data not shown). However, by using $C_0t$ fractionated HeLa DNA in the hybridization mixture, as described in Materials and Methods, we were able to quench the signal from the repetitive DNA. Among the recombinants, 25 phages gave a hybridization signal with the pUH112PR1.3 probe. Analysis of the DNA from these phages enabled us to find one clone (λJ1-CM211) whose insert overlaps that of λJ1-CM15 by 275 bp, as determined by DNA sequencing (see below), thus bridging the gap in the restrictiom map of the DHFR gene (Fig. 1B).

**(b) Organization of the Amplified DHFR Gene.**

The inserts of five recombinant phages from the λCha4A libraries and five phages from the λJ1 library (Fig. 1) were aligned and analyzed in detail by restriction enzyme mapping and DNA transfer hybridization experiments using $^{32}$P-labeled pHD84. Their orientation was determined by dot hybridization experiments with 5'- and 3'-specific probes from cDNA clones (Morandi *et al.*, 1982; Masters *et al.*, 1983).

The precise location of the protein coding regions in the DHFR gene was further investigated by subcloning the various fragments containing these regions (Fig. 2A), by analyzing these fragments by restriction enzyme digestion and DNA transfer hybridization experiments, and finally by DNA sequencing (see below). As observed in the cloning of the 6 kb *Eco*RI fragment of the DHFR gene into λ vectors, difficulties were experienced in the subcloning of this fragment. Thus, the region of the fragment containing the coding region III was particularly resistant to subcloning into pUC9 using JM83 or a restriction deficient and P1 cured derivative of this strain (TB1, kindly provided by T. O. Baldwin) as a host. As another approach to solve this problem, the 6 kb

*Eco*RI fragment from λJ1-CM211 was digested with several blunt-end cutting enzymes (*Sma*I, *Pvu*II, *Hin*cII, or *Hpa*I) and ligated to *Sma*I and *Eco*RI digested pUC9. One of the subclones (pUH211RS1.2), containing an *Eco*RI-*Sma*I fragment, was mapped near the 5'-end of the 6 kb EcoRI fragment and utilized for DNA sequencing (Fig. 2B). Another subclone was found to contain the coding region III and was utilized to localize this region by restriction mapping (pUH211PvR2.1, Fig. 2B). In another experiment, the 6 kb *Eco*RI fragment was completely digested with *Mbo*I, the mixture was ligated into the *Bam*HI site of pUC9, and the recombinants were screened by colony hybridization using the pHD84 probe. One such clone, pUH211M0.6, containing the coding region III (Fig. 2B) was utilized for DNA sequencing.

**(c) DNA Sequencing.**

Figure 2B shows the segments of the DHFR gene and of the 5' non-coding region which were sequenced and the sequencing strategies. The sequences of all the protein coding regions, starting at the ATG, of the adjacent intron regions, including the whole first intron, and of the 3' untranslated region up to the *Bgl*II site are shown in Figure 3, together with the homologous regions of the mouse DHFR sequence (Crouse *et al.*, 1982; Setzer *et al.*, 1982). Sequencing of the whole human DNA insert of pUH15RO.5 (Fig. 2B) yielded the sequence of the major portion of the second coding region and of the adjacent segment of the second intron (position 457 to 881) (Fig. 3). The sequence from position 607 to 881 was also found in the 5'-end proximal segment of the insert of pUH211RS1.2 (Fig. 2B). The overlap by 275 bp of the two sequences thus appeared to bridge the gap in the restriction map of the DHFR gene, strongly suggesting that the whole gene had been cloned. This conclusion was corroborated by the results of genomic blotting experiments (see below).

The protein coding sequence is complete, and is identical to that determined from the human DHFR cDNA (Masters and Attardi, 1983). Also the sequence of the 3'

untranslated region, up to the *Bgl*II site, matches perfectly that of the cDNA (Masters *et al.*, 1983). The perfect alignment of the restriction sites of this region in λJ1-CM113 (Fig. 2B) with the restriction sites in the cDNA (Morandi *et al.*, 1982) indicates that probably there are no introns in this portion of the gene at least up to the *Bam*HI site (pUH113Sm5.6, Fig. 2B). The intron-exon junctions are conserved as compared to the mouse sequence; their positions follow the GT-AG rule (Breathnach *et al.*, 1978), although the only unambiguous junction is between coding regions V and VI.

The available human and mouse gene sequences show, as previously reported (Masters *et al.*, 1983), high homology (89%) in the protein coding regions; the homology drops to less than 30% in the introns at 15 to 35 bp from the splice junctions in all cases where both the human and the mouse sequences are available. A dot-matrix plot of the intron sequences did not show any obvious homology of these regions in the two genes (not shown).

The 5' non-coding sequence of the human DHFR gene, ending with the ATG translation initiation codon, is shown in Figure 4, together with the available 5' non-coding sequence of the mouse gene (Crouse *et al.*, 1982), as well as with the homologous regions of two recently isolated human pseudogenes, ψHD1 ($\psi_1$, Masters *et al.*, 1983) and hDHFR-$\psi_2$ ($\psi_2$, Chen *et al.*, 1982). The homology to the mouse 5' non-coding region is 65% and extends to at least -327 bp from the ATG initiator codon. Both pseudogenes exhibit good homology, upstream of the initiator codon, to the human 5' non-coding region up to position -61 (79% for ψHD1 and 85% for hDHFR-$\psi_2$), and then rapidly diverge. ψHD1 has an additional 176 nt at the 5' end which shows no homology to the available 5' non-coding sequence of the human gene (data not shown)

The alignment of the first coding segment and 5' non-coding region of the human and mouse DHFR gene was determined by the best-fit plotting program of T. Hunkapiller and L. Hood (manuscript in preparation) and is shown in Figure 5. This plot best illustrates the extensive nucleotide homology of the human and mouse 5' non-coding

sequences. In particular, it shows clearly that the human 5' non-coding region contains a 45 bp stretch at position -91 to -135, which is homologous to the fourfold 45-48 bp repeat of the 5' non-coding region of the mouse sequence (also shown in Fig. 4). This 45 bp human sequence is not repeated; however, a portion of it, from position -127 to -110 (GGGGCGGGGGGGCGGGGC), has a 16 out of 18 homology to the sequence from position -93 to -76 (GGGACGAGGGGGCGGGGC).

Figure 4 also shows that the human 5' non-coding sequence is particularly rich in G+C (~72% G+C from position -395 to the ATG codon). In order to compare the distribution of G+C in the available sequences of the human and mouse DHFR genes, the variation in % G+C along the two genes was plotted as shown in Figure 6. The high G+C content found in the 5' non-coding region of the human gene is maintained in the first coding region and almost the entire first intron (67% G+C), whereas the rest of the available human sequence, including the 3' non-translated region, is A-T rich (36% in coding regions II to VI and adjacent intron segments, and 41% in the 3' non-coding region). This unusual pattern of G+C distribution appears to be conserved in the mouse gene (66% in 5' non-coding region, 60% in first coding region and first intron, 37% in coding regions II to VI and adjacent intron segments, and 41% in the 3' non-coding region).

**(d) The Organization of the Amplified DHFR Gene is Identical to That of the Native Gene.**

In order to test whether the complete DHFR gene had been cloned, and to investigate whether the organization of the human DHFR gene isolated from the MTX-resistant 6A3 cells had been altered during amplification, genomic blot hybridization experiments were carried out using *Eco*RI, *Hin*dIII, *Sst*I, and *Kpn*I digests of HeLa cell DNA and 6A3 cell DNA with the pHD84 probe. As shown in Figure 7A, all the restriction fragments which were expected to hybridize to pHD84 on the basis of the restriction map of the cloned

gene (Fig. 7B) were found in both the 6A3 cell DNA and the HeLa cell DNA samples. The 8.8 kb *Sst*I fragment hybridized very weakly to pHD84 because it contains only 19 bp of coding sequence. These results clearly confirmed the conclusion of the restriction mapping studies indicating that the whole amplified DHFR gene had been cloned, and furthermore indicated that the gene organization determined here is identical to that of the native DHFR gene. The extra fragments present in HeLa cell DNA probably represent pseudogenes or other DHFR-related sequences (Masters *et al.*, 1983), which are unlinked to the DHFR gene in chromosome 5 (Maurer *et al.*, 1983).

Blot hybridization of *Eco*RI digested DNAs from HeLa and 6A3 cells with $^{32}$P-labeled λJ1-CM15 (Fig. 1B) showed that the 5' flanking sequences of the amplified gene, over at least 30 kb upstream of the coding sequence, are also present upstream of the native gene in HeLa cells (Figs. 7A and 7B). It appears, therefore, that in 6A3 cells, these upstream sequences had been amplified together with the DHFR gene without any significant rearrangement during selection in the presence of MTX.

**(e) Phage Recombinants Containing Rearranged DHFR Gene Fragments.**

In the course of this work, some λCha4A and λJ1 recombinant phages were isolated, whose DNA hybridized with pHD84, but which contained *Eco*RI fragments not found in the normal DHFR gene. The constitution of the inserts of these clones is shown in Table 1. The 1.8 kb and 6 kb *Eco*RI fragments of λJ1-CM12 hybridized with pHD84, and were thought to be the 1.8 kb and 6 kb *Eco*RI fragments of the native and the amplified gene observed in genomic blots (Fig. 7). However, a detailed analysis of the subcloned 6 kb *Eco*RI fragment of λJ1-CM12 revealed that this fragment differs in restriction pattern from the 6 kb *Eco*RI fragment of the normal DHFR gene (Fig. 7B and unpublished data). Whether the 1.8 kb *Eco*RI fragment of λJ1-CM12 and λCha4A-CM95 represents

the 1.8 kb fragment of the normal gene has not been determined in this study.

λCha4A-DMς6 contains normal DHFR gene sequences except for the insertion of a 3.5 kb *Eco*RI fragment between the 2.5 kb and 1.6 kb *Eco*RI fragments (Fig. 7B). The 3.5 kb fragment is apparently derived from fusion of the adjacent 2.5 kb and 1.6 kb *Eco*RI fragments with a small deletion, as indicated by restriction mapping of the three fragments (unpublished data). The other recombinant phages containing rearranged DHFR gene fragments which are listed in Table 1 have not been analyzed in detail in this study.

## Discussion

The complete human DHFR gene amplified in 6A3 cells has been cloned using four different lambda phage libraries, and its organization determined and shown to correspond to that of the native human DHFR gene. This gene spans approximately 29 kb beginning at the ATG translation initiation codon and extending up to the end of the 3' untranslated region represented in the 3.8 kb mRNA (Morandi *et al.*, 1982). The intron/exon junctions of the gene are at positions identical to those found in the mouse gene, and the corresponding sequences show good homology to the consensus donor and acceptor sequences of the majority of nuclear protein-coding genes of higher eucaryotes (Mount, 1982). The data of Chen *et al.*, (1982) on the structure of the 3' portion of the human DHFR gene agree with our results, as concerns the size of the fourth and fifth introns and the probable absence of introns in the 3' untranslated region.

### (a) Intron Divergence.

In contrast to the conservation of the intron-exon junction positions, a comparison of the human, mouse (Crouse *et al.*, 1982) and Chinese hamster genes (Carothers *et al.*, 1983; Milbrandt *et al.*, 1983) shows a strong divergence of the lengths of the introns in these

species, except for intron I (Fig. 8). Especially striking is the difference in size, which is almost fourfold, between human intron V and the homologous intron of the mouse and Chinese hamster genes. This degree of divergence in intron size is much more marked than previously observed among homologous globin genes in man, mouse and rabbit (Van Ooyen *et al.*, 1979; Nishioka and Leder, 1979; Liebhaber *et al.*, 1980). In spite of the variation in intron size, the overall length of the DHFR gene is very similar in man, mouse and Chinese hamster.

Previous work (Masters and Attardi, 1983) had shown that the human and mouse DHFR reading frames have ~89% nucleotide sequence homology. By contrast, in the present study, the analysis of the available intron sequences of the human and mouse genes have revealed a strong divergence of these sequences, as previously observed for the introns of other split genes (Breathnach and Chambon, 1981). In fact, no significant nucleotide sequence homology has been detected between corresponding DHFR introns in mouse and man, except in short (15 to 35 bp) stretches adjacent to the coding sequences at the 5'-end and 3'-end.

**(b) G+C Distribution Pattern.**

Another feature of considerable interest revealed by the present work is the unusual G+C distribution along the human DHFR gene. According to G+C content, the gene appears to be divided into two portions: a 5'-end portion rich in G+C (~70%), which includes ~400 bp of non-coding sequences, the first coding region and almost the whole first intron, and the main body of the gene, including the 3' untranslated region, with a low G+C content (~40%). This G+C distribution pattern is conserved in the mouse gene, pointing to the existence of different evolutionary constraints in nucleotide substitution rates in different portions of this gene. This marked unevenness of G+C content within the DHFR gene may reflect the subdivision of the genome in blocks of different G+C content, without correspondence to gene organization and without direct functional

significance. A more interesting possibility is that the G+C distribution observed here has its functional correlate in the existence of different chromatin domains within the DHFR gene, possibly important for the regulation of gene expression. In view of the proposed role of methylation in gene regulation, it is interesting that the CpG dinucleotides represent 11.7% of the dinucleotides in the 5' non-coding region of the human gene, a figure which is fairly close to the frequency of 12.7% expected from the base composition. By contrast, in the (A+T)-rich portion of the gene (including the 3' noncoding region), the CpG frequency drops to 0.8%, a value which is significantly lower than the expected one of 3.5%. The latter frequencies of CpG are similar to the averages observed and expected in the mammalian genome (1 and 4%, respectively) (McClelland and Ivarie, 1982). Whether the lack of selection against CpG in the 5' non-coding region of the DHFR gene is significant from the point of view of control of gene expression remains to be determined.

**(c) 5' Noncoding Region.**

Previous work (Morandi *et al.*, 1982) has led to the identification in human cells of three main DHFR mRNAs which differ in the length of their 3' untranslated regions. We do not know yet where the 5' ends of these DHFR mRNAs map. It seems possible that at least one transcription initiation site is located around position -61 of the human sequence, i.e., the position where two human pseudogenes, which are likely to derive from mRNA intermediates (Chen *et al.*, 1982; Masters *et al.*, 1983), start diverging in sequence. This region lies immediately downstream of the mouse fourfold repeat which has been suggested to contain a DHFR mRNA 5' end (Crouse *et al.*, 1982). In the human system, preliminary primer extension experiments using reverse transcriptase and polysomal poly(A)-containing RNA from 6A3 cells have mapped a putative 5' end of DHFR mRNA around position -65 (Masters, unpublished observations). A possible TATA-like sequence, CACAAATA, exists 32 bp 5' to position -61. This sequence lies within and near the 3' end

of the 45 bp stretch which is homologous to the fourfold 45-48 bp repeat present in the mouse 5' non-coding region (Crouse *et al.*, 1982). The 125 bp upstream of this putative TATA box have an extremely high G-C content (84%). The significance of the very biased base composition of this region is not known, but it may play a role in the possible cell cycle dependent (Hendrickson *et al.*, 1980; Mariani *et al.*, 1981) and other types of regulation of this gene. (G+C)-rich sequence motifs, which play an important role in transcription, have been detected in the 5' flanking regions of the herpes simplex virus thymidine kinase gene (McKnight and Kingsbury, 1982) and the SV40 large T antigen gene (Benoist and Chambon, 1981; Fromm and Berg, 1982; Everett *et al.*, 1983); mutation of these regions have been shown to reduce transcription efficiency up to ten- to twentyfold.

The 5' non-coding region shows considerable homology (~65%) in the two human and mouse genes up to at least 327 bp upstream of the initiator codon. However, the 45-48 bp fourfold repeat of the mouse gene (Crouse *et al.*, 1982) is present in only one copy in the human sequence. The significance of this difference is not clear. A sequence conservation of the 5' non-coding region has been previously observed in several eukaryotic genes, although not so extensive as that reported here (Van Ooyen *et al.*, 1979; Breathnach and Chambon, 1981). The only exception is the ~365 bp nucleotide sequence homology of the 5' non-coding region in the human and rat insulin genes (Bell *et al.*, 1982). It is generally believed that the sequence conservation in the 5' non-coding region underlies the importance of this region in the control of gene expression. In several systems, such as the SV40 early region (Benoist and Chambon, 1981), the sea urchin histone H2A gene (Grosschedl and Birnstiel, 1980), the globin genes (Efstratiadis *et al.*, 1980), the ovalbumin gene (Mulvihill *et al.*, 1982; Compton *et al.*, 1983), and the yeast CYC1 (Guarente and Ptashne, 1981), HIS4 (Donahue *et al.*, 1983), and GAL10 (Guarente *et al.*, 1982) genes, besides the signals close to the initiation sites of transcription, regulatory sequences further upstream have been described. In the case of

the CYC1 promoter region, the heme regulatory site is located at about 275 bp upstream of the transcription initiation site (Guarente and Mason, 1983). Specific binding sites for progesterone receptor have been located at 135 to 300 bp upstream of the mRNA stating site for the ovalbumin gene (Mulvihill *et al.*, 1982; Compton *et al.*, 1983). It seems possible that conserved sequences located in the region extending up to several hundred bp upstream of the DHFR reading frame play an important regulatory function in the expression of this gene.

**(d) 3' Untranslated Region.**

In contrast to 5' non-coding region, the 3' untranslated region of the DHFR gene shows no obvious homology in the human and mouse genes over at least 770 bp downstream of the coding sequence (Masters *et al.*, 1983, and unpublished data). It is, therefore, particularly intriguing that there is evidence that the 3' untranslated region of the DHFR gene is at least partly responsible for the regulation of the DHFR mRNA level in growth-stimulated cells vs. stationary cells (Kaufman and Sharp, 1983). Whatsoever the 3'-end signals involved in this growth-dependent regulation may be, it would appear that, in this case, a high rate of nucleotide sequence divergence is compatible with their functional conservation.

## Acknowledgements

These investigations were supported by National Institutes of Health grants GM11726 and T32 GM07616. We are very grateful to Dr. A. Chomyn for carrying out the chromosome preparation and fractionation, B. Maurer for carrying out some of the genomic blot hybridizations, Dr. F. Amalric (Centre de Récherche de Biochimie et de Génetique Cellulaires du C.N.R.S., Toulouse, France) for help in some of the early mapping studies and valuable discussions, Dr. J. Mullins (Dept. of Microbiology, Harvard University) for providing the phage λJ1, Dr. T. O. Baldwin (Texas A & M University, College Station,

Texas) for providing the strain TB1, and T. Hunkapiller and Dr. L. Hood for help in the computer program analysis of the data and for making the best-fit plotting program available to us prior to publication. The excellent assistance of Ms. Doris Finch and Ms. Arger Drew is gratefully acknowledged.

## References

Bell, G. I., Selby, M. J. and Rutter, W. J. (1982). Nature **295**, 31-35.

Benoist, C. and Chambon, P. (1981). Nature **290**, 304-310.

Benton, W. D. and Davis, R. W. (1977). Science **196**, 180-182.

Birnboim, H. C. and Doly, J. (1979). Nucl. Acids Res. **7**, 1513-1523.

Blattner, F. R., Williams, B. G., Blechl, A. E., Thompson, K. D., Faber, H. E., Furlong, L. A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L. and Smithies, O. (1977). Science **196**, 161-169.

Bolivar, F. (1978). Gene **4**, 121-136.

Boyer, H. W. and Roulland-Dussoix, D. (1969). J. Mol. Biol. **41**, 459-472.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978). Proc. Natl. Acad. Sci. USA **75**, 4853-4857.

Breathnach, R. and Chambon, P. (1981). Ann. Rev. Biochem. **50**, 349-383.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Methods Enzymol. **29**, 363-418.

Carothers, A. M., Urlaub, G., Ellis, N. and Chasin, L. A. (1983). Nucl. Acids Res. **11**, 1997-2012.

Chen, M., Shimada, T., Moulton, A. D., Harrison, M. and Nienhuis, A. W. (1982). Proc. Natl. Acad. Sci. USA **79**, 7435-7439.

Compton, J. G., Schrader, W. T. and O'Malley, B. W. (1983). Proc. Natl. Acad. Sci. USA **80**, 16-20.

Crouse, G. F., Simonsen, C. C., McEwan, R. N. and Schimke, R. T. (1982). J. Biol. Chem. **257**, 7887-7897.

Donahue, T. F., Daves, R. S., Lucchini, G. and Fink, G. R. (1983). Cell **32**, 89-98.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980). Cell **21**, 653-668.

Everett, R. D., Baty, D. and Chambon, P. (1983). Nucl. Acids. Res. **11**, 2447-2464.

Fromm, M. and Berg, P. (1982). J. Mol. Appl Genet. **1**, 457-481.

Gross-Bellard, M., Oudet, P. and Chambon, P. (1973). Eur. J. Biochem. **36**, 32-38.

Grosschedl, R. and Birnstiel, M. L. (1980). Proc. Natl. Acad. Sci. USA **77**, 7102-7106.

Grunstein, M. and Hogness, D. (1975). Proc. Natl. Acad. Sci. USA **72**, 3961-3965.

Guarente, L. and Ptashne, M. (1981). Proc. Natl. Acad. Sci. USA **78**, 2199-2203.

Guarente, L., Yocum, R. R. and Gifford, P. (1982). Proc. Natl. Acad. Sci. USA **79**, 7410-7414.

Guarente, L. and Mason, T. (1983). Cell **32**, 1279-1286.

Hendrickson, S. L., Jin-Shybun, R. W. and Johnson, L. F. (1980). Proc. Natl. Acad. Sci. USA **77**, 5140-5144.

Hoopes, B. C. and McClure, W. R. (1981). Nucl. Acids Res. **9**, 5493-5504.

Kaufman, R. J. and Sharp, P. A. (1983). Mol. Cell. Biol. **3**, 1598-1608.

Kiger, J. A. and Sinsheimer, R. L. (1969). J. Mol. Biol. **40**, 467-490.

Lau, P. P. and Gray, H. B. Jr. (1979). Nucl. Acids Res. **6**, 331-357.

Liebhaber, S. A., Goossens, M.J. and Kan, Y. W. (1980). Proc. Natl. Acad. Sci. USA **77**, 7054-7058.

Loenen, W. A. M. and Brammar, W. J. (1980). Gene **10**, 249-259.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C. and Quon, D. (1978). Cell **15**, 687-701.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). in *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory.

Mariani, B. D., Slate, D. L. and Schimke, R. T. (1981). Proc. Natl. Acad. Sci. USA **78**, 4985-4989.

Masters, J., Keeley, B., Gay, H. and Attardi, G. (1982). Mol. Cell. Biol. **2**, 498-507.

Masters, J. N. and Attardi, G. (1983). Gene **21**, 59-63.

Masters, J. N., Yang, J. K., Cellini, A. and Attardi, G. (1983). J. Mol. Biol. **167**, 23-36.

Maurer, B., Barker, P. E., Masters, J. N., Ruddle, F. H. and Attardi, G. (1983). Proc. Natl. Acad. Sci. USA. In press.

Maxam, A. M. and Gilbert, W. (1980). Methods Enzymol. **65**, 499-560.

McClelland, M. and Ivarie, R. (1982). Nucl. Acids Res. **10**, 7865-7877.

McKnight, S. L. and Kingsbury, R. (1982). Science **217**, 316-324.

Messing, J. (1979). Recombinant DNA Technical Bulletin, NIH Publication No. 79-99, 2, No. 2 (1979) pp. 43-48.

Milbrandt, J. D., Azizkhan, J. C., Greisen, K. S. and Hamlin, J. L. (1983). Mol. Cell. Biol. **3**, 1266-1273.

Morandi, C. and Attardi, G. (1981). J. Biol. Chem. **256**, 10169-10175.

Morandi, C., Masters, J. N., Mottes, M. and Attardi, G. (1982). J. Mol. Biol. **156,** 583-607.

Mount, S. M. (1983). Nucl. Acids. Res. **10,** 459-473.

Mulvihill, E. R., LePennec, J.-P. and Chambon, P. (1982). Cell **24,** 621/632.

Nishioka, Y. and Leder, P. (1979). Cell **18,** 875-882.

Schmid, C. W. and Deininger, P. L. (1975). Cell **6,** 345-358.

Setzer, D. R., McGrogan, M. and Schimke, R. T. (1982). J. Biol. Chem. **257,** 5143-5147.

Van Ooyen, A., Van Den Berg, J., Mantel, N. and Weissmann, C. (1979). Science **206,** 337-344.

Vieira, J. amd Messing, J. (1982). Gene **19,** 259-268.

Williams, B. G., Blattner, F. R., Jaskunas, S. R. and Nomura, M. (1977). J. Biol. Chem. **252,** 7344-7354.

Wray, W., Stubblefield, E. and Humphrey, R. (1972). Nature New Biol. **238,** 237-238.

**Table 1**

Recombinant phages containing rearranged DHFR gene sequences.

| Recombinant phages | Sizes (kb) of *Eco*RI fragments in the insert[+] |
|---|---|
| λJ1.CM12 | 4.7, 1.5, 0.8, 1.8*, 6.0* |
| λCha4A.CM95 | 8.0, 4.0, 1.8* |
| λCha4A.DM$_S$6 | ***0.5***, ***2.4***, 3.5*, ***1.6***, ***4.0**** |
| λCha4A.DM$_L$3 | 12* |
| λJ1.CM104 | 10.5, 2.3*, 1.6, 1.3 |
| λJ1.CM107 | 16* |
| λJ1.CM108 | 14.5* |
| λJ1.CM109 | 11.5*, 2.5, 1.6* |
| λJ1.CM224 | 14.5*, 2.2 |
| λJ1.CM233 | 10, 4.2, 2.1*, 0.68 |

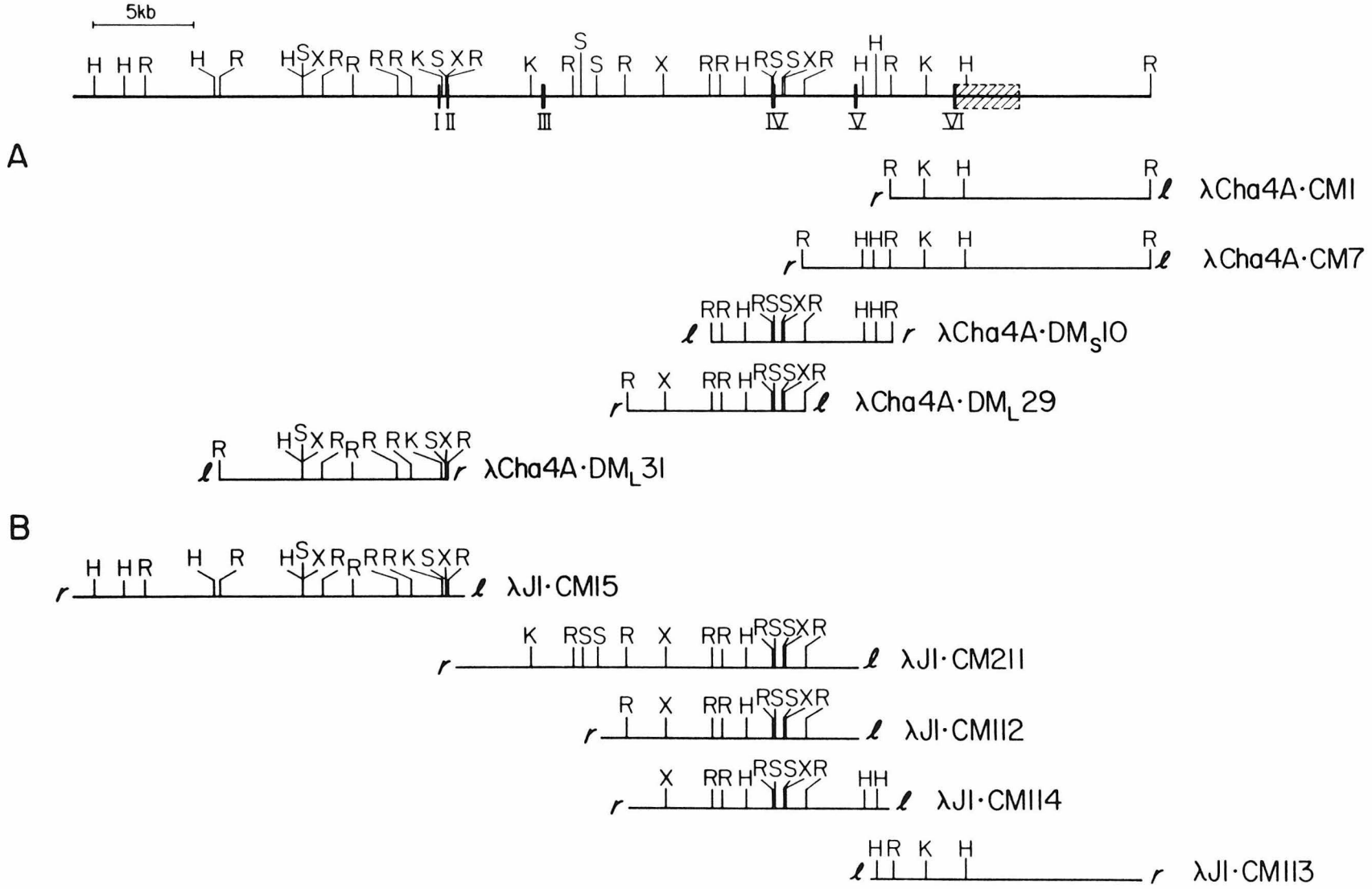
The asterisked numbers represent fragments which hybridized with pHD84.

The numbers in bold italic type indicate *Eco*RI fragments also found in the normal gene; those underlined by a single line have a size very similar to that of normal fragments but they were not analyzed further.

## Figure 1

Construction of the Human DHFR Restriction Map.

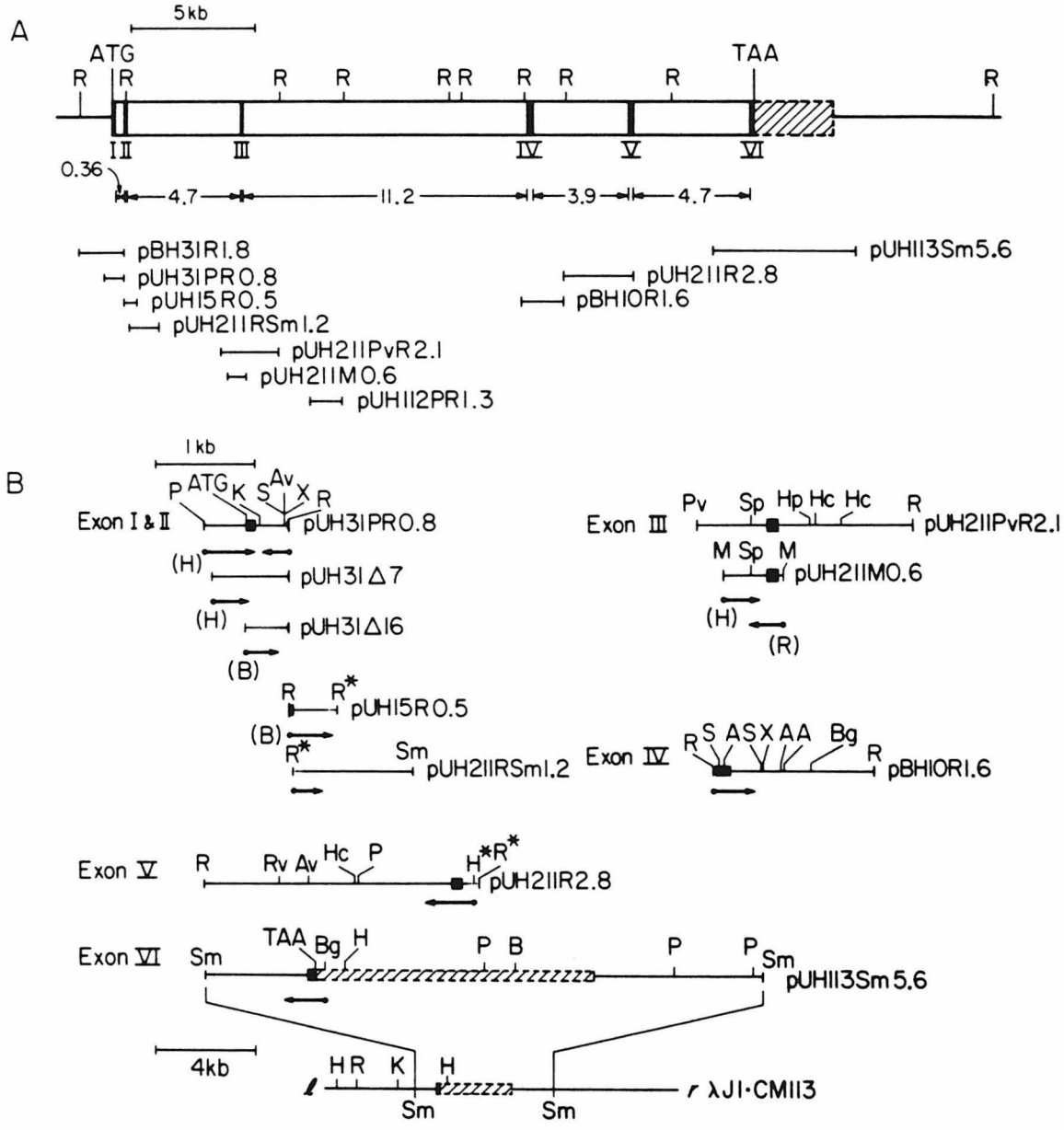
The human DHFR gene and its flanking sequences were isolated from four different λ libraries [three λCh4A (A) and one λJ1 (B)] constructed with the DNA from the "normal" chromosomes (CM) or the large double minutes ($DM_L$) or the small double minutes ($DM_S$) of the MTX-resistant human cell line 6A3, as described in Materials and Methods. The restriction map was constructed by analysis of the DNA from ten overlapping recombinant λ phages by single or double digestion with 5 different restriction enzymes, and by Southern blot hybridization of restriction digests with $^{32}P$-labeled pHD84. The coding sequences are indicated as solid boxes and the 3'-untranslated sequences, as a cross-hatched box. Sequence data show that λJ1-CM15 and λJ1-CM211 contain 275 bp of overlapping sequences at the 3'- and 5'-end of the insert, respectively. The orientation of the inserts with respect to λ vector arms is indicated as r (right arm) and l (left arm). The restriction enzymes used are indicated as follows: R, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sst*I; X, *Xho*I.


5kb
A
λCha4A·CM1
λCha4A·CM7
λCha4A·DMs10
λCha4A·DML29
λCha4A·DML31
B
λJ1·CM15
λJ1·CM211
λJ1·CM112
λJ1·CM114
λJ1·CM113

## Figure 2

Organization of the Human DHFR Gene (A) and Sequencing Strategy (B).

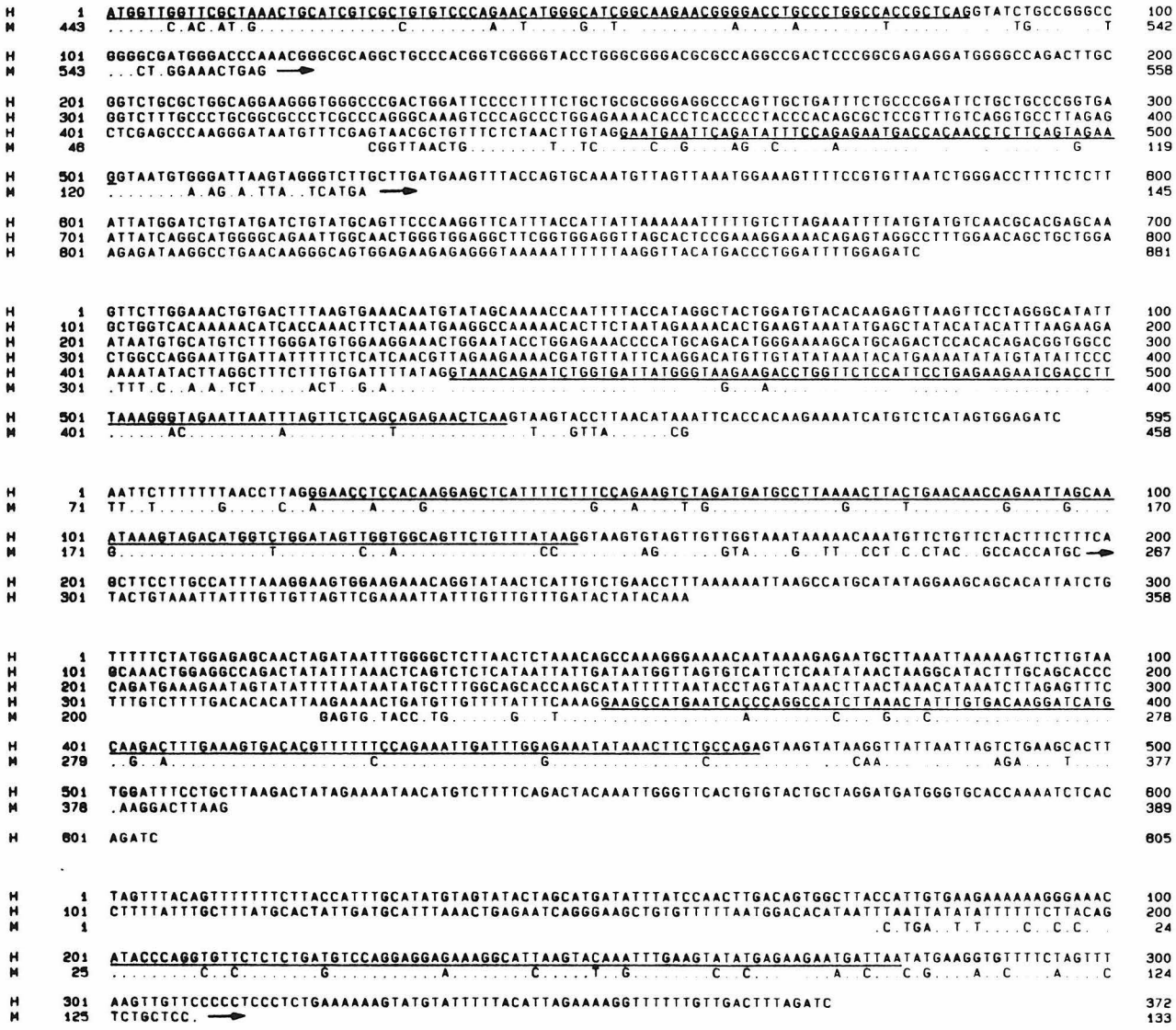
A. The various segments of the human DHFR gene are represented as boxes: solid boxes, DHFR coding sequences; empty boxes, intervening sequences; hatched box, 3'-untranslated sequences. The positions of the initiation codon (ATG), termination codon (TAA), and *Eco*RI sites (R) are shown on the top of sequences. The coding regions are designated by Roman numerals (I to VI), and the sizes of the introns are indicated in kb. Also shown are the subcloned fragments, which were utilized for the screening of λJ1 library, fine restriction mapping, and the sequencing of the protein coding regions. In each designation, the first number represents the number of the recombinant phage from whose DNA the cloned fragment was derived, and the last number, the size of the fragment; the letters between the two numbers represent the restriction enzyme site(s) utilized for cloning (R, *Eco*RI; P, *Pst*I; Pv, *Pvu*II; M, *Mbo*I; Sm, *Sma*I). B. Fine restriction maps of the subcloned gene fragments containing coding regions, and sequencing strategies. The extent and direction of the sequences determined are indicated by arrows. In each case, both strands of a given segment were sequenced. Restriction enzyme designations are as follows: A, *Acc*I; Av, *Ava*I; B, *Bam*HI; Bg, *Bgl*II; H, *Hin*dIII; Hc, *Hin*cII; Hp, *Hpa*I; K, *Kpn*I; M, *Mbo*I; P, *Pst*I; Pv, *Pvu*II; R, *Eco*RI; Rv, *Eco*RV; S, *Sst*I; Sm, *Sma*I; Sp; *Sph*I; X, *Xho*I. The restriction sites in parentheses belong to the plasmid DNA, those with asterisks to phage DNA; the phage DNA segments are indicated by a dotted line.


A
5kb
ATG
TAA
R
I II
III
IV
V
VI
0.36
4.7
11.2
3.9
4.7
pBH31R1.8
pUH31PR0.8
pUH15R0.5
pUH211RSm1.2
pUH211PvR2.1
pUH211M0.6
pUH112PR1.3
pUH113Sm5.6
pUH211R2.8
pBH10R1.6
B
1kb
P ATG K S Av X R
Exon I & II
pUH31PR0.8
(H)
pUH31Δ7
pUH31Δ16
(B)
R R*
pUH15R0.5
R* Sm
pUH211RSm1.2
Exon III
Pv Sp Hp Hc Hc R
pUH211PvR2.1
M Sp M
pUH211M0.6
(R)
Exon IV
R S AS X AA Bg R
pBH10R1.6
Exon V
R Rv Av Hc P H* R*
pUH211R2.8
Exon VI
Sm TAA Bg H P B P P Sm
pUH113Sm5.6
4kb
H R K H
Sm
Sm
λJ1·CM113

## Figure 3

DNA Sequence of the Protein Coding Segments and Adjacent Intron Regions and of the 3'-Untranslated Segment of the Human DHFR Gene and a Comparison with the Homologous Mouse Gene Sequence.

The portions of the human gene (H) sequence shown correspond, in the order, to coding regions I and II with the complete first intron and 5'-end proximal segment of second intron, then to the individual coding regions III to VI with their respective flanking intron regions, and finally to the 3' untranslated region up to the *Bgl*II site. The coding sequences are underlined and delimited according to the GT-AG rule (Breathnach ***et al.***, 1978). The portions of the mouse DHFR gene (M) which are shown (coding regions and adjacent intron segments) are numbered according to Crouse ***et al.***, (1982) and Setzer ***et al.***, (1982). Arrows at the 3'-ends of the mouse gene segments indicate that the corresponding sequences continue but are not shown. Dots (.) indicate nucleotides identical to those of the human sequence.



## Figure 4

Comparison of the 5' Non-coding Sequences of the Human and Mouse DHFR Genes and of the Homologous Regions of Two Human Pseudogenes.

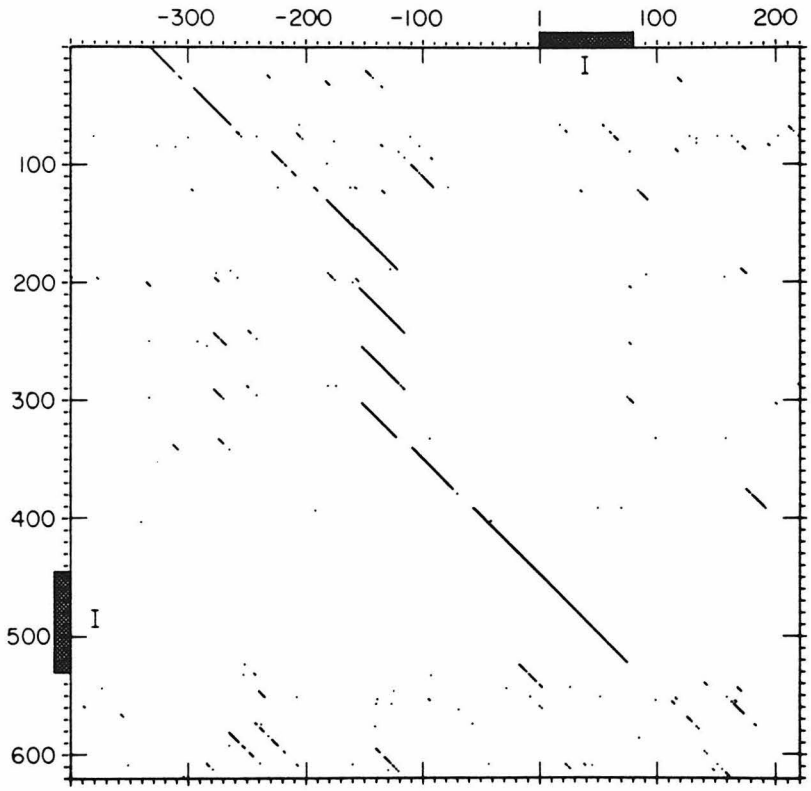
The human DNA sequence (H) is indicated with progressive negative numbers relative to the initiator codon ATG. The available mouse sequence (M) is aligned below and numbered according to Crouse *et al.*, (1982). The first nucleotides of the four 45-48 bp repeats of the mouse 5' non-coding sequence are indicated by circled numbers under the mouse sequence. The homologous segments of two recently isolated pseudogenes ψHD1 (ψ1) (Masters *et al.*, 1983) and hDHFR-$\psi_2$ ($\psi_2$) (Chen *et al.*, 1982) are also shown. Dots represent identical nucleotides; gaps have been introduced to maximize homologies.


H -394 CTGCAGCGCCAGGGTCCACCTGGTCGGCTGCACCTGTGGAGGAGGAGGTGGATTTCAGGCTTCCCGTAGACTGGAAGAATCGGCTCAAAACCGCTTGCCT -295
M 1 ....CT.....C....CT.G.....T..... 33
H -294 CGCAGGGGCTGAGCTGGAGGCAGCGAGGCCGCCCGACGCAGGCTTCCGGCGAGACATGGCAGGGCAAGGATGGCAGCCCGGCGGCAGGGCCCGGCGAGGA -195
M 34 .C.C...CT..G.GGC.GA.T..A ...T.....G.AT....C.GC .....TT.CA.TGT..AGAAGA..C.. ..T..T.... 118
H -194 GCGCGAACCCGCGGCCGCAGTTCCCAGGCGTCTGCGGGCGCGAGCACGCCGCGACCCTGCGTGCGCCGGGGCGGGGGGGCGGGGCCTCGCCTGCACAAAT -95
M 119 ..... GG.....C..T.TC.C ..CT..CG.C.....T.A... ..T.G.G.A.C..C.TT.....A....TAG..A..... 211
① -94
H AGGG -91
M 212 ..AATGCGCGGCGGGCCTTGGTGGGGGCGGGGCCTTAGCTGCACAAATAGGATGCGCGGCGGGCCTTGGTAGGGGCGGAGCCTTAGCTGCACAAATAGGA 311
② ③
H -90 ACGAGGGGGCGGGGCGGCCACAATTTCGCGCCAAACTTGACCGCGCGTTCTG -39
M 312 TGCGCGGCGGGCCTTGGTGGGGGCGGGGCCTAAGCTGCGCAAGTGGTACAC..CTCA G...TG.G.....G..AATCCTA. 406
④
$\Psi_1$ AA.AT.AAATAG.....T....G...T..A.CT..
$\Psi_2$ AA.A.CC...CG.....T....G.....C...
H -38 CTGTAACGAGCGGGCTCGGAGGTCCTC CCGCTGCTGTCATG 3
M 407 .GTG..G.C T.G.A...TT.TA..C.....CA..... 445
$\Psi_1$ G....GT....A.....G...AT.....
$\Psi_2$ G....G.....T.....AG...A.....

## Figure 5

Best-fit Plot of the 5' Non-coding Regions of the Human and Mouse DHFR Genes.

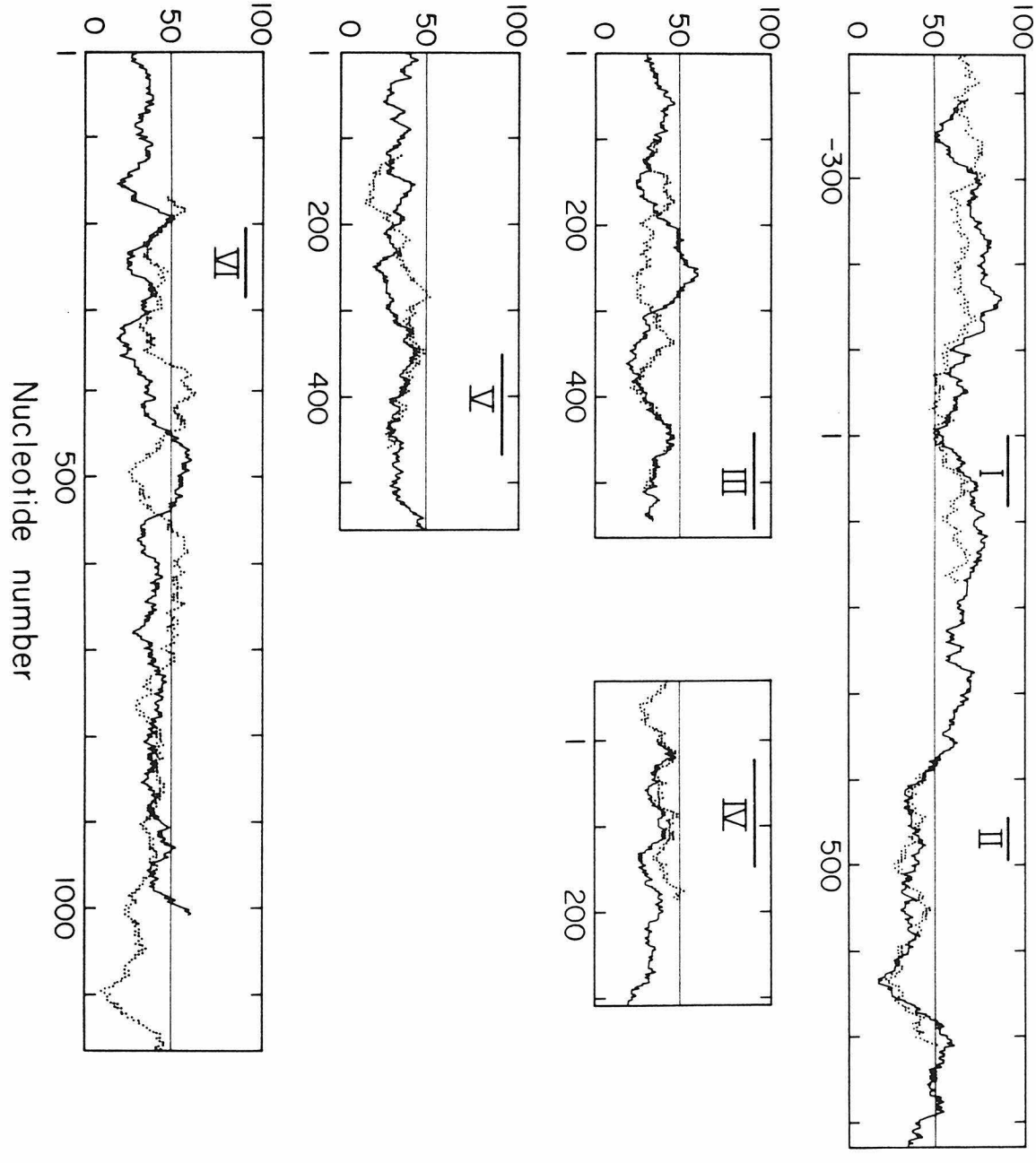
The human sequence is along the x-axis and the mouse sequence, along the y-axis; the diagonal lines represent the best alignment of the two sequences achieved by a computer program developed by T. Hunkapiller and L. Hood (manuscript in preparation). In this plot, the alignment is based on scanning one sequence with successive 50 nt blocks, with 1 nt steps, of the other sequence, a dot being placed where a homology of >50% exists. The positions of the first coding regions of the human and mouse genes are indicated.


-300
-200
-100
1
100
200
I
100
200
300
400
500
600
I

## Figure 6

The Conservation of the of G+C Distribution Pattern Along the Human and Mouse DHFR Genes.

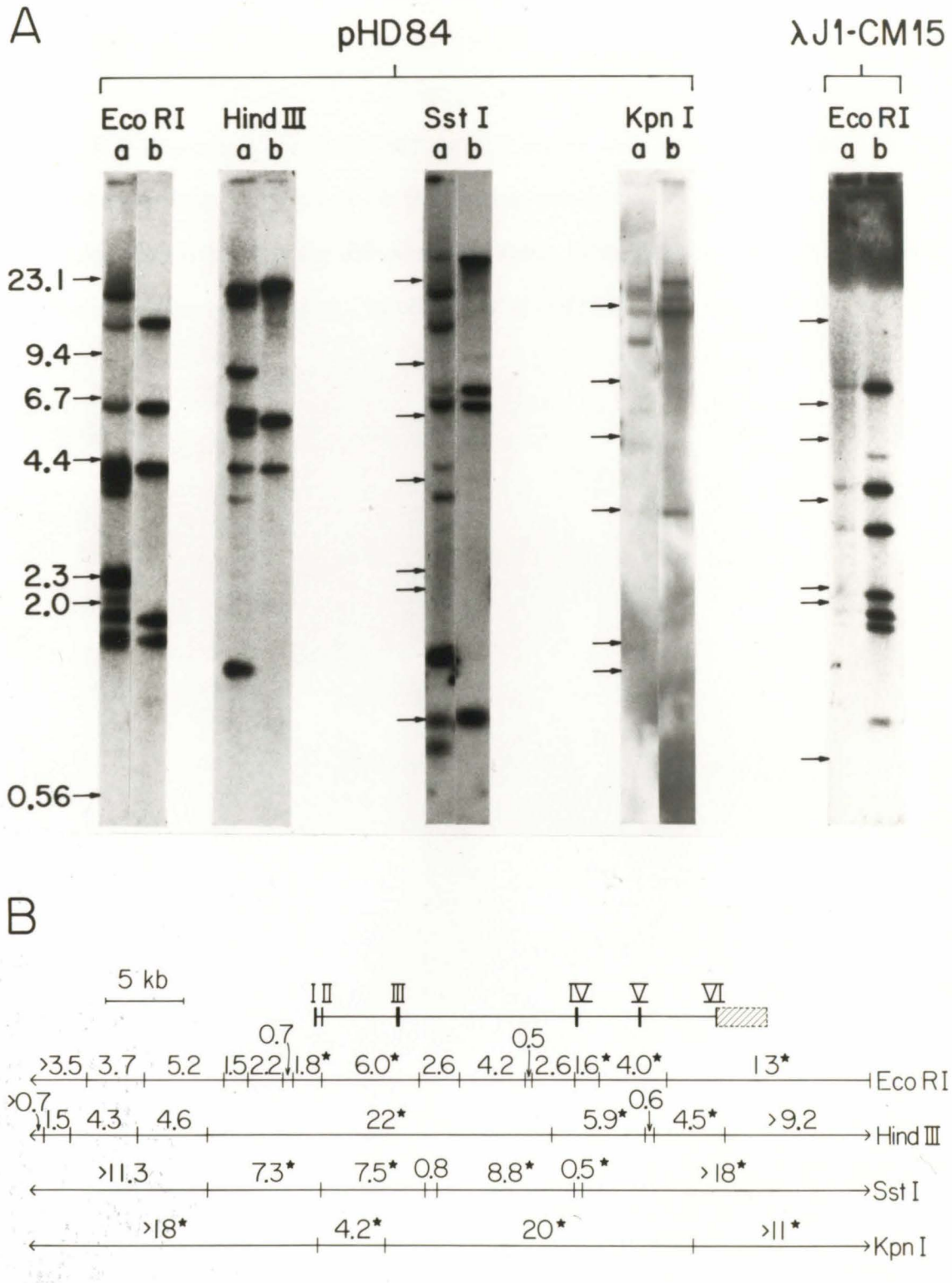
The variation in G+C content along the sense strand of the human and mouse DHFR genes was determined by analyzing the % G+C of the two sequences in successive 50 nt stretches with 1 nt steps. The human sequence is indicated by a solid line and the mouse sequence (Crouse *et al.*, 1982; Setzer *et al.*, 1982), by a dotted line. The positions of the protein coding regions are indicated by horizontal bars and identified by Roman numerals. The numbering on the abscissae is according to the human sequence (Figs. 3 and 4).


Percent GC
100
50
0
-300
1
500
I
II
III
200
400
IV
V
VI
1000
Nucleotide number

## Figure 7

Patterns of Hybridization with Human DHFR Specific Probes of Restriction Enzyme Digests of Genomic DNA from HeLa Cells and 6A3 Cells.

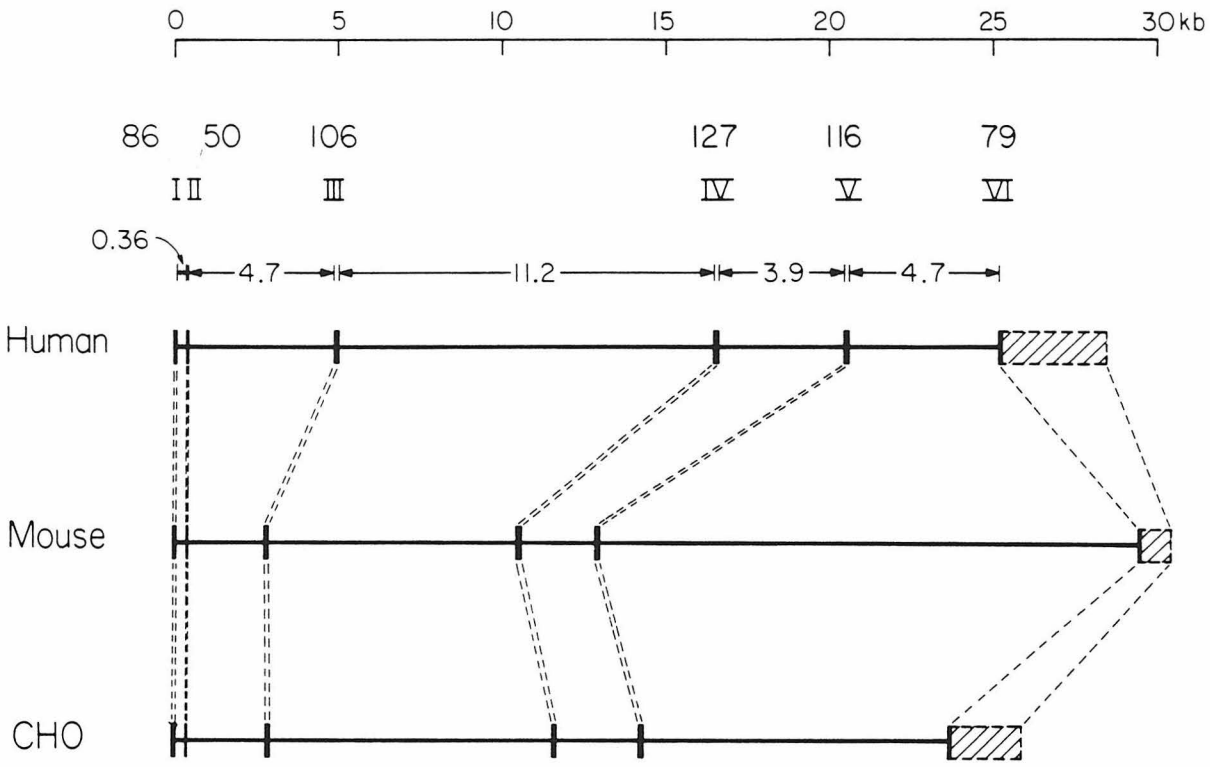
A. Samples of 10 μg of HeLa S3 DNA (except in the experiment with the λJ1-CM15 probe, where ~2 μg were used) (lanes a), and 0.5 μg of 6A3 DNA (lanes b) were digested with the indicated restriction enzyme and fractionated by electrophoresis through an agarose gel. The DNAs were transferred to a nitrocellulose paper and hybridized to $^{32}$P-labeled pHD84 or λJ1.CM15 (using $C_0t$ fractionated DNA as a carrier in the latter case), as previously described (Masters *et al.*, 1983). B. Restriction maps of the human DHFR gene and adjacent genomic regions, with the sizes of the fragments being indicated in kb. An asterisk indicates fragments which contain protein coding sequences.


A
pHD84
λJ1-CM15
Eco RI
Hind III
Sst I
Kpn I
Eco RI
a b
23.1
9.4
6.7
4.4
2.3
2.0
0.56
B
5 kb
I II III IV V VI
0.7
0.5
>3.5 3.7 5.2 1.5 2.2 1.8* 6.0* 2.6 4.2 2.6 1.6* 4.0* 13*
Eco RI
>0.7 1.5 4.3 4.6 22* 5.9* 0.6 4.5* >9.2
Hind III
>11.3 7.3* 7.5* 0.8 8.8* 0.5* >18*
Sst I
>18* 4.2* 20* >11*
Kpn I

## Figure 8

Comparison of the Organization of the DHFR Genes from Man, Mouse, and Chinese Hamster.

The coding sequences are indicated as solid boxes and the 3'-untranslated sequences, as a cross-hatched box. The sizes of the coding sequences in each exon (in bp) and those of the human gene introns (in kb) are indicated. The sizes of the Chinese hamster gene introns are those reported by Carothers *et al.*, (1983).


0
5
10
15
20
25
30kb
86
50
106
127
116
79
I II
III
IV
V
VI
0.36
4.7
11.2
3.9
4.7
Human
Mouse
CHO

# CHAPTER 7

## Dihydrofolate Reductase Gene Related Transcripts Present in both Polysomal and Nuclear RNA of Human Cell Lines Map with their 5'-Ends Upstream of the Main mRNA Cap Site

## Abstract

The 5'-ends of dihydrofolate reductase (DHFR) specific transcripts have been mapped in the 5'-flanking region of the DHFR gene of the human methotrexate-resistant cell line 6A3 by primer extension and S1 protection experiments. The main 5'-end, at position -71 relative to the first nucleotide of the DHFR reading frame, corresponds to a major class of trancripts representing approximately 99% of the DHFR-specific polysomal poly(A)-containing RNA; in particular, an analysis of this RNA fractionated on an agarose-methylmercuric hydroxide gel has revealed that the -71 end is associated with the previously described DHFR mRNA of sizes 3.8, 1.0 and 0.8 kb, and corresponds to the recently identified transcription initiation site of the DHFR gene. Six other minor 5'-ends have been mapped to nucleotide positions -449 to -480 upstream of the DHFR gene, and correspond to approximately 1% of the DHFR-specific polysomal poly(A)-containing RNA. Analysis of polysomal RNA from $VA_2$-B cells, the parental line of 6A3, and from HeLa cells has shown that the upstream initiating transcripts are not specific for cells with amplified genes, although their amount is increased in the cell lines containing amplified genes in about the same proportion as the three identified DHFR RNAs. The upstream initiating DHFR-specific transcripts are also present in nuclear RNA; in this fraction, they are indeed relatively more abundant than in the polysomal RNA, representing approximately 11% of the major transcripts initiating at position -71: this may point to a different regulation of these DHFR-related upstream initiating transcripts.

## Introduction

The analysis of the mode of expression of the dihydrofolate reductase (DHFR) gene in mammalian cells, which has been greatly facilitated by the availability of cell lines with amplified genes, has revealed an unsuspected complexity both in these and the parental cell lines, in the form of the existence of multiple species of DHFR mRNAs. Thus, in mouse cells, four major polyadenylated species of DHFR mRNA (Setzer *et al.*, 1982), in Chinese hamster 3 species (Lewis *et al.*, 1982; Setzer *et al.*, 1982), and in human cells, three main species have been identified (Morandi *et al.*, 1982). These multiple mRNA species differ mainly in the length of the 3'-untranslated tail (Setzer *et al.*, 1982; Carothers *et al.*, 1983; Masters *et al.*, 1983), which can be up to several fold longer than the reading frame, as is the case for the major human DHFR mRNA (3.8 kb mRNA). Furthermore, there is good evidence indicating that the 3'-end tails of the multiple mRNAs are colinear (Setzer *et al.*, 1982; Masters *et al.*, 1983). These observations would be compatible with the idea that the same transcript could produce different forms of mRNA by processing and polyadenylation at alternative sites.

The significance of the existence of multiple forms of DHFR mRNA is unknown. The conservation of this feature in different mammalian species strongly suggests that the multiplicity of mRNAs reflects the existence of a physiologically significant control mechanism. For a full understanding of the possible functional role of the multiplicity of DHFR mRNA, it would be important to know whether there is any regulation of expression of the different DHFR mRNAs at the level of initiation of transcription. In the present work, an extensive mapping study of the 5'-ends of DHFR-specific transcripts in human cell lines with an amplified or normal DHFR gene complement has revealed, unexpectedly, the presence, besides a main transcription initiation point at position -71 relative to the first nucleotide of the DHFR reading frame, of at least six additional minor sites upstream of the DHFR gene, which correspond to 5'-ends of DHFR-related transcripts. These upstream initiating transcripts are found both in polysomal RNA and

in the nuclear RNA; in the latter fraction, they are enriched about 10-fold over the transcripts initiating at the main start point as compared to the polysomal fraction. While this work was in progress, a report has appeared (Chen *et al.*, 1984) describing the identification of a main initiation point at position -71 for human DHFR specific transcripts, on the basis of primer extension and S1 protection experiments utilizing total HeLa cell RNA and of *in vitro* transcription assays.

## Materials and Methods

**Cell lines and methods of growth.** HeLa S3 cells were grown in suspension as described (Amaldi and Attardi, 1968). $VA_2$-B, an 8-azaguanine resistant derivative (Weiss *et al.*, 1968) of the SV40 transformed line WI-18-$VA_2$ (Pontén *et al.*, 1963), and its methotrexate-resistant variant 6A3 (Masters *et al.*, 1982) were grown as described (Morandi *et al.*, 1982).

**RNA preparation.** Polysomal RNA was prepared from HeLa, $VA_2$-B, and 6A3 cells (Morandi *et al.*, 1982), and the poly(A)-containing fraction was separated from the non-poly(A)-containing fraction by oligo(dT)-cellulose chromatography (Amalric *et al.*, 1978).

For the preparation of nuclear RNA, nuclei were isolated from 6A3 cells (~8 ml packed cell volume) by a hypotonic shock procedure (Jeanteur *et al.*, 1968), resuspended by vortexing in 4 ml of 6 M urea, 3M LiCl (Auffray and Rougeon, 1980), 10 mM Na acetate, pH 5, 0.2 mg/ml heparin, 0.1% SDS, and sonicated at 4 mA for 60 sec with a Branson sonifier. The RNA was allowed to precipitate overnight at 4°C, and the precipitate was washed with 8 M urea, 4 M LiCl, dissolved in water, extracted with phenol/chloroform/isoamylalcohol (25:24:1) and ethanol precipitated. The RNA was then digested with DNaseI (50 μg/ml) in the presence of $CaCl_2$ (0.01 M) and proteinase K (10 μg/ml) (Tullis and Rubin, 1980), phenol extracted, ethanol precipitated and fractionated by oligo(dT)-cellulose chromatography. The total nuclear RNA recovered

(1.3 mg) was estimated to be 4% of total RNA (assuming 30 pg RNA/cell (Philipson, 1961) $1.5 \times 10^8$ cells/ml packed cells), and contained an amount of 18S rRNA corresponding to approximately 1% of that present in the cytoplasmic fraction, as estimated by scanning with a Joyce Loebl microdensitometer a negative from an ethidium bromide stained gel (with various amounts of cytoplasmic and nuclear non-poly(A)-containing RNA).

**DNA preparation and labeling.** Plasmid DNAs were prepared by alkaline lysis (Birnboim and Doly, 1979), followed by CsCl-ethidium bromide centrifugation. For production of specific primers or fragments to be used in S1 protection experiments, 10-50 µg plasmid DNA was first digested with a restriction enzyme producing the desired end, treated with calf intestinal alkaline phosphatase (Boeringer-Mannhiem), and labeled at the 5'-ends with T4 polynucleotide kinase (Bethesda Research Laboratories) and ($\gamma$-$^{32}$P)-ATP (ICN); the DNA was then spermine precipitated (Hoopes and McClure, 1981) to remove the bulk of unincorporated nucleotides, digested with a second restriction enzyme, and the fragments of interest, labeled at only one end, were isolated by elution from polyacrylamide gels (Maxam and Gilbert, 1980). When the first restriction digest contained many fragments, the fragment of interest was isolated from the rest by elution from an agarose (Yang *et al.*, 1979) or polyacrylamide gel (Maxam and Gilbert, 1980). DNA was sequenced according to Maxam and Gilbert (1980). $^{32}$P-labeled markers were prepared by 3'-end labeling *Hpa*II digests of pBR322 or human mtDNA in 50 mM Tris, pH 8.0, 5 mM $MgCl_2$, 1 mM ß-mercaptoethanol, 0.1 mg/ml BSA, 20 mM dATP and dTTP at 20°C for 60 min with the large fragment of *E. coli* DNA polymerase I and 1.5-fold molar excess of ($\alpha$-$^{32}$P)-dCTP and -dGTP (2000-3000 Ci/mmole, Amersham) over the number of available sites to be filled in at the ends. Unlabeled dCTP and dGTP were then added to 50 mM, and the reaction was continued for 30 min to ensure complete filling of all of the ends.

**Primer extension and S1 protection experiments.** RNA-DNA hybridizations in both primer extension and S1 protection experiments were carried out in 10 µl of 0.1 M NaCl,

0.02 M Tris, pH 8.0 (25°C), 0.001 M EDTA at 60-75°C for 1 h, using the desired DNA fragment from 0.2 μg of plasmid DNA and polysomal poly(A)-containing RNA from 6A3 cells (1 μg), or HeLa cells (10 μg), or $VA_2$-B cells (10 μg), and stopped by quick cooling in ice-water.

The primer extension reactions were performed in 50 μl of 0.08 M Tris, pH 8.3, 0.05 M KCl, 0.01 M MgCl, 0.01 M dithiothreitol, 4 mM sodium pyrophosphate, 20 units RNasin (Promega Biotec), in the presence of 10 units AMV reverse transcriptase (Life Science), at 42°C for 30 min, and were followed by phenol extraction, ethanol precipitation and electrophoresis through 20 cm, 7 M urea-5% polyacrylamide gels or 80 cm 8.3 M urea-5% polyacrylamide gels; the gels were then dried and exposed to sensitized Kodak XAR-5 film with Dupont intensifying screens (Laskey and Mills, 1977).

The S1 nuclease digestions were carried out in 100 μl of 0.3 M NaCl, 0.03 M Na acetate, 0.001 M $ZnSO_4$, using 80-200 units S1 nuclease (Sigma) at 37°C for 30-60 min, and were followed by phenol extraction, ethanol precipitation and electrophoresis, as described above.

## Results

**Multiple 5'-ends of human DHFR specific RNAs.** Figure 1 shows the DHFR specific cDNA and genomic restriction fragments, and figure 2 the experimental strategies used in the 5'-end mapping experiments. The first primer extension reactions were carried out using 6A3 polysomal poly(A)-containing RNA as a template and the 53 nucleotide *Eco*RI-*Hin*fI fragment (Fig. 1A) derived from pHD84 (Morandi *et al.*, 1982; Masters *et al.*, 1983) as a primer (Fig. 2A). When the products of this reaction were electrophoresed on a denaturing polyacrylamide gel, a major band corresponding to reverse transcripts ~210 nucleotides (nt) long ($\alpha_1$) and a minor band corresponding to reverse transcripts 10 nt shorter ($\alpha_2$) were seen (Fig. 3A, lane 1). The 3'-end of the ~210 nt reverse

transcript corresponds in the DNA sequence to a position ~70 bp upstream of the DHFR gene initiator codon (Fig. 2A), where the major transcription initiation site for the human DHFR gene of HeLa cells had been recently mapped (Chen *et al.*, 1984); this result, therefore, indicated a similarly positioned initiation site for the DHFR mRNAs from 6A3 cells.

A longer exposure of the same autoradiogram (Fig. 3A, lane 2) revealed another minor band (β) corresponding to a size of ~620 nt, as well as other bands migrating below the ~210 nt band. While the latter bands probably represented premature termination products of the reverse transcriptase reaction, the upper band conceivably corresponded to reverse transcripts of either DHFR gene-related RNAs of low abundance or an unrelated RNA which had hybridized with the pHD84 derived primer. In order to distinguish between the latter two alternatives, a large scale primer extension reaction using 10 μg of 6A3 polysomal poly(A)-containing RNA was carried out to isolate component β from an agarose-10 mM $CH_3HgOH$ gel. This component was used in an S1 protection experiment with 6A3 polysomal poly(A)-containing RNA. If component β had been reverse transcribed from a DHFR gene-related RNA (i.e., containing at least part of the DHFR coding sequence), it was expected that hybridization with the highly predominant DHFR mRNAs would protect from S1 digestion a segment of β equivalent to ~210 nt (Fig. 2C) or shorter (depending upon the length of the homologous sequences). [The minor portion of β protected by the fully homologous RNA would probably not be detected.] Under the alternative hypothesis that β derived from an RNA unrelated to DHFR, an S1 protected band corresponding to somewhat less than 620 nt was expected. The results of this analysis showed a major protected segment corresponding to the ~210 nt band ($\alpha_1$) and a minor segment corresponding to the $\alpha_2$ band detected in the primer extension experiments (compare lanes 2 and 3 of Fig. 3B).

**Detailed mapping of the major 5'-ends of the human DHFR specific RNAs.** A combination of S1 protection and primer extension analysis was carried out to map more

precisely the major 5'-ends of the human DHFR specific RNAs, using the 199 nt *Hpa*II-*Bst*NI fragment and the 99 nt *Ban*II-*Bst*NI fragment of pBH31R1.8 (Fig. 1B), respectively. These probes were expected to produce extended primers and S1 protected fragments ~75 nt shorter than those observed with the pHD84 derived 53R-H primer, as diagramed in Figures 2C and 2D. The actual fragment sizes observed with these probes, as determined by electrophoresis on a 20 cm 7 M urea-5% polyacrylamide gel, corresponded to 130, 140 and 540 nt, i.e., the sizes expected for extended primers and S1 protected fragments having the same 3'-ends as $\alpha_2$, $\alpha_1$ and $\beta$, respectively (data not shown). [For simplicity, these shortened forms of $\alpha_1$, $\alpha_2$ and $\beta$, will also be referred to as $\alpha_1$, $\alpha_2$ and $\beta$.] Figure 4A shows the extended primers and S1 protected fragments from the same experiment, separated on an 80 cm sequencing gel in parallel with the products of the sequencing reactions carried out on the 199 nucleotide fragment. Both the S1 protection products (lanes 2 and 4) and the primer extension products (lanes 1 and 3) exhibit a pronounced band representing a fragment which falls with its 3'-end slightly behind the G at position -72 upstream of the human DHFR reading frame (Fig. 5; Chen *et al.*, 1984; Yang *et al.*, 1984): This fragment, therefore, corresponds to $\alpha_1$ in Figure 3. Since the sequencing reaction products lack the 3' base by which they are defined, and, furthermore, contain both 5'- and 3'-phosphates, in contrast to the extended primer and the S1-resistant fragment which have 3'-hydroxyls, the alignment of the sequence with these fragments is off by $1\frac{1}{2}$ nucleotides (Sollner-Webb and Reeder, 1979). Therefore, the 3'-ends of the extended primer and of the S1-resistant fragment map in correspondence to the T at position -71 upstream of the human DHFR reading frame. The band corresponding to a fragment one nucleotide longer than $\alpha_1$ in Figure 4, lane 3, may derive from reverse transcription of a minor portion of the DHFR mRNAs starting one nucleotide upstream of the major initiation point; however, the lack of a significant S1-resistant fragment of corresponding size would argue against it. For the same reason, the bands corresponding to fragments migrating below $\alpha_1$, of which those 2, 3, and 9 nt

shorter than $\alpha_1$ are particularly strong, probably represent premature termination points of the reverse transcriptase reaction, rather than reverse transcripts of RNAs starting downstream of the main initiation point. One exception may be the band corresponding to $\alpha_2$ in Figure 3, which represents a fragment 9 nt shorter than $\alpha_1$. There is, in fact, an S1-resistant fragment moving one nucleotide behind this extended primer; the production of this S1-resistant fragment in the S1 reactions carried out both at 20°C (lane 1) and at 37°C (lane 3) would argue against its being due to nibbling of the RNA-DNA hybrids in a region which is AT-rich (see below). (Such nibbling, on the contrary, probably accounts for the weaker bands between $\alpha_1$ and $\alpha_2$, which are clearly more pronounced at 37°C than at 20°C). A possible explanation for the 1 nt difference between the $\alpha_2$ extended primer and the putatively corresponding S1-resistant fragment is the presence of cap structure at the 5'-end of the mRNA which may prevent, under the conditions of S1 digestions used here, the complete trimming of the hybridized DNA strand. Further work, however, will be needed to determine whether the band corresponding to $\alpha 2$ corresponds to a true initiation point.

**Detailed mapping of the minor 5'-ends of the human DHFR specific RNAs.** The same analysis as described above was carried out to map the minor, upstream 5'-ends of the human DHFR specific RNAs, using the 1125 nucleotide *Eco*RI-*Bst*NI fragment and the 60 nucleotide *Pvu*II-*Bst*NI fragment of pBH31R1.8 (Fig. 1B) for S1 protection and primer extension experiments, respectively (diagramed in Figs. 2E and 2F). When the products of these reactions were separated on a short denaturing polyacrylamide gel (Fig. 4B), a series of six major bands were observed both in the lane containing the extended primer (lane 1) and in that containing the S1-resistant fragments (lane 2). These bands, which are labeled $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, $\beta_5$ and $\beta_6$, represent fragments with sizes of 54, 59, 63, 68, 72 and 80 nt, and correspond to band $\beta$ in Figure 3: in fact, the range of sizes of the fragments is reasonably close to that expected from the use of restriction fragments which should give extended primers and S1-resistant products ~520 nt shorter

than those obtained using the pHD84 derived 53R-H primer (Fig. 2). Separation of the extended primers and S1-resistant fragments from the same experiment on a thin sequencing gel in parallel with the products of the sequencing reactions carried out on the 1125 nucleotide fragment (Fig. 4c) showed that the bands $\beta_1$ to $\beta_6$ correlate with bands at positions -449, -456, -460, -466, -470, and -480 upstream of the human DHFR reading frame (again correcting for the one and a half nucleotide difference, as described above).

**Further characterization of the major and minor 5'-ends.** In order to determine whether the upstream 5'-ends of DHFR specific RNAs identified above are specific for the cell lines 6A3 with amplified DHFR genes or are general features of all cell lines, poly(A)-containing RNA samples from HeLa cells, $VA_2$-B cells and 6A3 cells were hybridized with the 1351 nucleotide *Eco*RI-*Sfa*NI fragment of pBH31R1.8 (Fig. 1B) and digested with S1 (Fig. 2G). The bands corresponding to $\alpha_1$, $\alpha_2$ and $\beta$ were observed with the RNA from all three cell lines (Fig. 6). Furthermore, it is clear that the RNAs with upstream 5'-ends are 5-10 times more abundant in $VA_2$-B cells than in HeLa cells, and ~20 times more abundant in 6A3 cells than in $VA_2$-B cells, thus behaving like the three DHFR mRNAs (Morandi *et al.*, 1982). S1 protection and primer extension experiments using 6A3 cell poly(A)-containing RNA fractionated on an agarose-$CH_3HgOH$ gel showed that the major 5'-end (corresponding to $\alpha_1$) is associated with the three previously identified DHFR mRNAs (3.8, 1.0 and 0.8 kb; Morandi *et al.*, 1982); by contrast, the minor 5'-ends (corresponding to the $\beta$-bands) appear to be associated with RNAs of different sizes, in the range of 1 to 4.5 kb (data not shown).

A primer extension analysis of nuclear RNA from 6A3 cells using the 53 nucleotide *Eco*RI-*Hin*fI fragment of pHD84 as a primer (Fig. 1b) is shown in Figure 7 (lanes 1 and 3), in parallel with the same type of analysis carried out on the polysomal RNA fraction (lanes 2 and 4). The results show that the upstream 5'-ends are primarily present in the poly(A)-containing fraction of both polysomal and nuclear RNA, that the RNA species

with the upstream 5'-ends are enriched in the nuclear fraction relative to the RNA species exhibiting the major 5'-end, and that the nuclear RNA preparation exhibits other minor bands which may or may not result from the occurrence of DHFR specific RNAs with different 5'-ends. A quantitation of the results made by cutting out and eluting the $^{32}$P-labeled bands corresponding to $\alpha_1$, $\alpha_2$ and $\beta$ and measuring their radioactivity showed that, in the polysomal RNA fraction, the RNA species carrying the upstream 5'-ends represent about 1% of the RNA species exhibiting the major 5'-end, and that this ratio is about 11% in the nuclear RNA fraction.

## Discussion

Polysomal poly(A)-containing RNA from human cell lines has previously been shown to contain at least three distinct species of DHFR mRNA with sizes of 3.8, 1.0 and 0.8 kb, differing in the length of the 3'-untranslated segment (Morandi *et al.*, 1982). These RNAs are assumed to derive from the transcript of a single gene by distinct cleavage and poly(A)-addition events occurring in the 3' untranslated region (Masters *et al.*, 1983; Chen *et al.*, 1984), as has been suggested to be the case for the mouse DHFR mRNAs (Setzer *et al.*, 1982). In the present work, by using S1 protection and primer extension experiments, we have shown that, while the majority of the DHFR gene transcripts map with their 5'-end at the recently identified transcription initiation site (Chen *et al.*, 1984), a minor portion of the DHFR specific RNAs map with their 5'-ends at multiple sites upstream of the main mRNA cap site.

The major 5'-end, which was found in approximately 99% of the DHFR-specific transcripts, was mapped at position -71 in the 5'-flanking region of the human DHFR gene (Yang *et al.*, 1984), in agreement with the data of Chen *et al.* (1984). In the latter mentioned work, evidence was presented indicating that this 5'-end is a major transcription initiation site. Consistent with this conclusion are our observations indicating that the 5'-ends of the three predominant DHFR mRNAs map in

correspondence to the site at position -71. Thirty nucleotides upstream of this site, at positions -101 to -94, there exists a possible "TATA-like" sequence (Goldberg, 1979), CACAAATA. This sequence lies within and near the 3'-end of the 45 bp stretch that is homologous to the fourfold 45-48 bp repeat present in the mouse gene 5'-noncoding region (Crouse *et al.*, 1982). Three of the four mouse repeats contain the same "TATA-like" sequence and it has been suggested that this fourfold repeat contains a DHFR mRNA 5'-end (Crouse *et al.*, 1982). The 5'-flanking region of the human DHFR gene does not have any sequence homologous to the "CAAT" box (Efstratiadis *et al.*, 1980); the latter is a sequence present in the 5'-flanking regions of various eukaryotic genes at positions -70 to -80 nucleotides from the transcription initiation site, and is thought to be part of eucaryotic promoters (Breathnach and Chambon, 1980). Flanking on both sides of the proposed "TATA-like" sequence, there is an almost perfect direct repeat of a purine-rich sequence at nucleotide positions -127 to -111 (GGGGCGGGGGGGCGGGG) and positions -93 to -77 (GGGACGAGGGGGCGGGG). The presence of homopolymer purine or homopolymer pyrimidine tracts has been shown to correlate with S1 hypersensitive sites near the 5'-ends of transcribed genes (Larsen and Weintraub, 1982; Schon *et al.*, 1983) and it has been suggested that they play a role in transcription initiation of eukaryotic genes. It is possible that the purine-rich sequence flanking the "TATA-like" sequence of the human DHFR gene has a similar role.

The most significant observation made in the present work is the identification of 6 minor 5'-ends of human DHFR related transcripts at positions -480, -470, -466, -460, -456 and -449 (Chen *et al.*, 1984; Yang *et al.*, 1984); these minor 5'-ends pertain to approximately 1% of the human DHFR transcripts found in the polysomal poly(A)-containing RNA. The most upstream of these sites (position -480) maps to the 3' side of a sequence (CCCCGCCCCCCCAGCCCC, positions -497 to -480), which is an almost perfect inverted repeat of the purine-rich sequence flanking the major 5'-end at positions -93 to -77. Again, it is interesting to speculate that this pyrimidine-rich sequence may

play a role in the transcription initiation of these upstream sequences. In contrast to the major 5'-end at position -71, no "TATA-like" sequence is found upstream of any of the 6 minor 5'-ends discussed here.

Transcripts of relatively low abundance have previously been found to initiate at various positions upstream (170-4500 bp) of the major transcription initiation site for the mouse β-globin gene (Hofer and Darnell, 1981) and the human β-globin gene (Carlson and Ross, 1983; Grindlay *et al.*, 1984), γ-globin gene (Grindlay *et al.*, 1984) and ε-globin gene (Allan *et al.*, 1983). The upstream sites for the human β-globin gene have been shown to be transcribed by RNA polymerase III (Pol III; Carlson and Ross, 1983); furthermore, at least one ε-globin transcript originates in an Alu repeat (Allan *et al.*, 1983), raising the possibility that it is also transcribed by Pol III, since Pol III can initiate transcription at Alu repeat sequences *in vitro* (Duncan *et al.*, 1979; Pan *et al.*, 1981; Di Segni *et al.*, 1981; Haynes and Jelinek, 1981; Manley and Colozzo, 1982). The 1.8 kb *Eco*RI fragment containing the 5'-end of the DHFR gene does not exhibit any Alu sequences (Chen *et al.*, 1984; Yang *et al.*, 1984). However, the sequence at positions -507 to -497 (TACCCACAAGC) exhibits a reasonable homology (9/11 match) to the upstream control sequence ($5'\ T^{A}_{G}N\,^{C}_{T}NN^{AA}_{GG}NGG\ 3'$) of the Pol III promoter consensus sequence (Sharp *et al.*, 1981), and the sequence at positions -459 to -450 (GACAGAACCT) shows an 8/9 match to the downstream control sequence of the same promoter ($5'\ G^{A}_{T}TC^{G}_{A}ANNCN\ 3'$). These observations suggest the possibility that at least some of the DHFR-related trancripts are transcribed by Pol III.

Two translation initiation codons are present in the region between the upstream and the major 5'-end of the human DHFR specific transcripts, one at positions -(239-237), and the other, at positions -(225-223). The initiating codon at positions -(239-237) (singly underlined in Fig. 5) is not in frame with the DHFR reading frame, and the corresponding reading frame, potentially coding for a polypeptide of 48 amino acids, terminates at the TAG at position -(112-110). The other initiation codon at positions -(225-223) (doubly

underlined in Fig. 5) is in frame with the DHFR coding region, and its utilization could add 75 amino acids to the N-terminus of the DHFR polypeptide. A third translation initiation codon is found at positions -(478-476), 2 nucleotides down from the most upstream of the minor 5'-ends detected in this work. Its significance is doubtful, in view of the lack of an appreciable 5'-noncoding region in the corresponding transcript and of the low abundance of this transcript; however, if this initiating codon could be utilized, a peptide of 50 amino acids would be made before terminating at the TAG at positions -(328-326).

Our data show no evidence for any RNA-splicing events in the region between positions -480 and -71 and suggest that the transcripts starting at the upstream initiation sites are spliced identically to the major DHFR specific RNAs, at least up to the 5'-end proximal side of the third DHFR exon. At this time we do not know where the 3'-ends of the upstream initiated RNAs map; therefore, we cannot say whether these upstream initiated transcripts contain the complete DHFR coding sequence.

The analysis of the upstream initiating transcripts in nuclear RNA has suggested that these transcripts may be regulated differently from the major DHFR specific RNAs; in fact, the steady-state level of the upstream initiating transcripts, as estimated from the relative amounts of extended primers, is approximately 11% of that of the major DHFR specific transcripts in the nuclear fraction, while they represent only about 1% of the major transcripts in the cytoplasmic fraction. At what level this difference arises is not known, but the mechanism involved could conceivably operate at the transcriptional level, at the level of nuclear RNA processing and transport, or at the level of RNA stability. The observation that a substantial portion of these minor transcripts are present in the polyadenylated fraction of polysomal RNA suggests that they are utilized as mRNAs. Such possibility, if confirmed, would raise intriguing questions concerning the nature of the translation products of these mRNAs and their possible regulatory role. A relevant observation in this respect may be that the amount of the upstream

initiating DHFR transcripts is increased in the cell lines with amplified DHFR genes ($VA_2$-B, 6A3) approximately in the same proportion as the main DHFR mRNAs.

## References

Allan, M., Lanyon, W.G. and Paul, J. (1983) Multiple origins of transcription in the 4.5 kb upstream of the ε-globin gene. Cell **35**, 187-197.

Amaldi, F. and Attardi, G. (1968) Partial sequence analysis of ribosomal RNA from HeLa cells. I. Oligonucleotide pattern of 28S and 18S RNA after pancreatic RNase digestion. J. Mol. Biol. **33**, 737-755.

Amalric, F., Merkel, C., Gelfand, R. and Attardi, G. (1978) Fractionation of mitochondrial RNA from HeLa cells by high resolution electrophoresis under denaturing conditions. J. Mol. Biol. **122**, 301-319.

Auffray, C. and Rougeon, F. (1980) Purification of mouse immunoglobin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**, 303-314.

Birnboim, H. C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**, 1513-1523.

Breathnach, R. and Chambon, P. (1980) Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. **50**, 349-383.

Carlson, D. P. and Ross, J. (1983) Human β-globin promotor and coding sequences transcribed by RNA polymerase III. Cell **34**, 857-864.

Carothers, A. M., Urlaub, G, and Chasin, L. A. (1983) Structure of the dihydrofolate reductase gene in Chinese hamster ovary cells. Nucl. Acids. Res. **11**, 1997-2012.

Chen, M.-J., Shimada, T., Moulton, A. D., Cline, A., Humphries, R. K., Maizel, J. and Neinhuis, A. W. (1984) The functional human dihydrofolate reductase gene. J. Biol. Chem. **254**, 3933-3943.

Crouse, G. F., Simonsen, C. C., McEwan, R. N. and Shimke, R. T. (1982) Structure of amplified normal and variant dihydrofolate reductase genes in the mouse sarcoma S180 cells. J. Biol. Chem. **257**, 7887-7897.

Di Segni, G., Cararra, G., Tocchini-Valentini, G. R., Shoulders, C. C. and Baralle, F. E. (1981) Selective *in vitro* transcription of one of the two Alu family repeats in the 5' flanking region of the human ε-globin gene. Nucl. Acids Res. **9**, 6709-6722.

Duncan, C., Biro, P. A., Choudary, P. V., Elder, J. T., Wang, R. R. C., Forget, B. G., DeRiel, J. K. and Weissman, S. M. (1979) RNA polymerase III transcriptional units are interspersed among human non-α-globin genes. Proc. Nat. Acad. Sci. U.S.A. **76**, 5095-5099.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeReil, J. K., Forget, B. G., Weissman, S. M., Slighton, J. L., Blechl, A. C., Smithies, O., Barelle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980) Structure and evolution of the human β-globin gene family. Cell **21**, 653-658.

Goldberg, M. (1979) Ph.D. Thesis, Stanford University, Stanford California

Grindlay, G. J., Lanyon, W. G., Allan, M. and Paul, J. (1984) Alternative sites of transcription initiation upstream of the canonical cap site in human γ-globin and β-globin genes. Nucl. Acids Res. **12**, 1811-1820.

Haynes, S. R. and Jelinek, W. R. (1981) Low molecular weight RNAs transcribed *in vitro* by RNA polymerase III from Alu-like dispersed repeats in Chinese hamster DNA are also found *in vivo*. Proc. Nat. Acad. Sci. U.S.A. **78**, 6130-6134.

Hofer, E. and Darnell, J. E. (1981) The primary transcription unit of the mouse B-major globin gene. Cell **23**, 585-593.

Hoopes, B. C. and McClure, W. R. (1981) Studies on the selectivity of DNA precipitation by spermine. Nucl. Acids Res. **9**, 5493- 5504.

Jeanteur, P., Amaldi, F. and Attardi, G. (1968) Partial sequence analysis of ribosomal RNA from HeLa cells. II. Evidence for sequences of nonribosomal type in 45S and 32S ribosomal RNA precursors. J. Mol. Biol. **33**, 757-775.

Larsen, A. and Weintraub, H. (1982) An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. Cell **29**, 609-622.

Laskey, R. A. and Mills, A. D. (1977) Enhanced autoradiographic detection of $^{32}P$ and $^{125}I$ using intensifying screens and hypersensitized film. Febs. Lett. **82**, 314-316.

Lewis, J. A., Davide, J. P. and Melera, P. W. (1982) Selective amplification of polymorphic dihydrofolate reductase gene loci in Chinese hamster lung cells. Proc. Nat. Acad. Sci., U.S.A. **79**, 6961-6965.

Manley, J. L. and Colozzo, M. T. (1982) Synthesis *in vitro* of an exceptionally long RNA transcript promoted by an AluI sequence. Nature **300**, 376-379.

Masters, J. N., Keeley, B., Gay, H. and Attardi G. (1982) Variable content of double minute chromosomes is not correlated with degree of instability in methotrexate-resistant human cell lines. Mol. Cell. Biol. **2**, 498-507.

Masters, J. N., Yang, J. K., Cellini, A. and Attardi, G. (1983) A human dihydrofolate reductase pseudogene and its relationship to the multiple forms of specific messenger RNA. J. Mol. Biol. **167**, 23-36.

Maxam, A. M. and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. **65**, 499-560.

Morandi, C., Masters, J., Mottes, M. and Attardi, G. (1982) Multiple forms of human

dihydrofolate reductase messenger RNA; cloning and expression in *E. coli* of their DNA coding sequences. J. Mol. Biol. **156,** 583-607.

Pan, J., Elder, J. T., Duncan, C. H. and Weissman, S. M. (1981) Structural analysis of interspersed repetitive polymerase III transcription units in human DNA. Nucl. Acids Res. **9,** 1151-1170.

Philipson, L. (1961) Chromatographic separation and characteristics of nucleic acids in HeLa cells. J. Gen. Phys. **44,** 899-910.

Pontén, J., Jensen, F. and Koprowski, H. (1963) Morphological and virological investigation of human tissue cultures transformed with SV40. J. Cell. Comp. Phys. **61,** 145-154.

Schon, E., Evans, T., Welsh, J. and Efstratiadis, A. (1983) Conformation of promoter DNA: fine structure mapping of S1 hypersensitive sites. Cell **35,** 837-848.

Setzer, D. R., McGrogan, M. and Schimke, R.T. (1982) Nucleotide sequence surrounding multiple polyadenylation sites in the mouse dihydrofolate reductase gene. J. Biol. Chem. **257,** 5143-5147.

Sharp, S., DeFranco, D., Dingerman, T., Farrell, P and Söll, D. (1981) Internal control regions for transcription of eukaryotic tRNA genes. Proc. Nat. Acad. Sci. U.S.A. **78,** 6657-6661.

Sollner-Webb, B. and Reeder, R. H. (1979) The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis.* Cell **18,** 485-499.

Tullis, R. H. and Rubin, H. (1980) Calcium protects DNase I from proteinase K: a new method for removal of contaminating RNase from DNase I. Anal. Biochem. **107,** 260-264.

Weiss, M. C., Ephrussi, B. and Scaletta, L. J. (1968) Loss of T-antigen from somatic

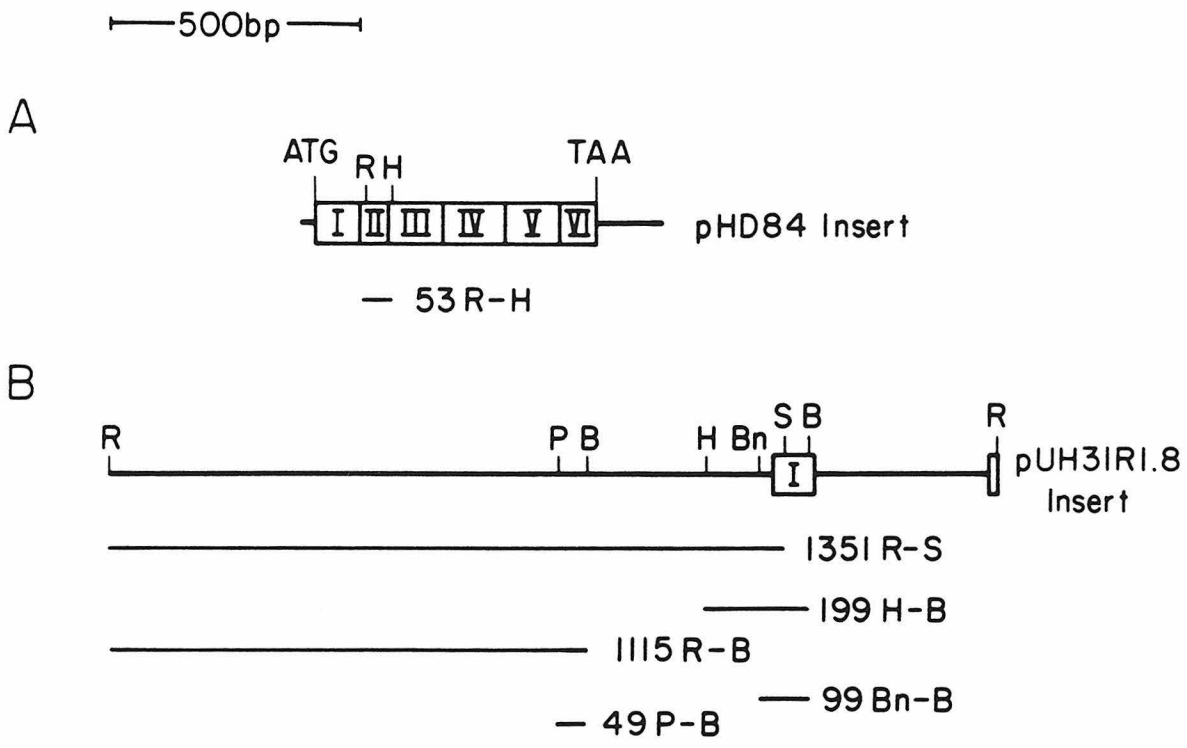
hybrids between mouse cells and SV40-transformed human cells. Proc. Nat. Acad. Sci., U.S.A. **59**, 1132-1135.

Yang, R. C.-A., Lis, J. and Wu, R. (1979) Elution of DNA from agarose gels after electrophoresis. Meth. Enzymol. **68**, 176-182.

Yang, J. K., Masters, J. N. and Attardi, G. (1984) Human dihydrofolate reductase gene organization: extensive conservation of the (GC)-rich 5' noncoding sequence and strong intron size divergence from homologous mammalian cells. J. Mol. Biol., in press.

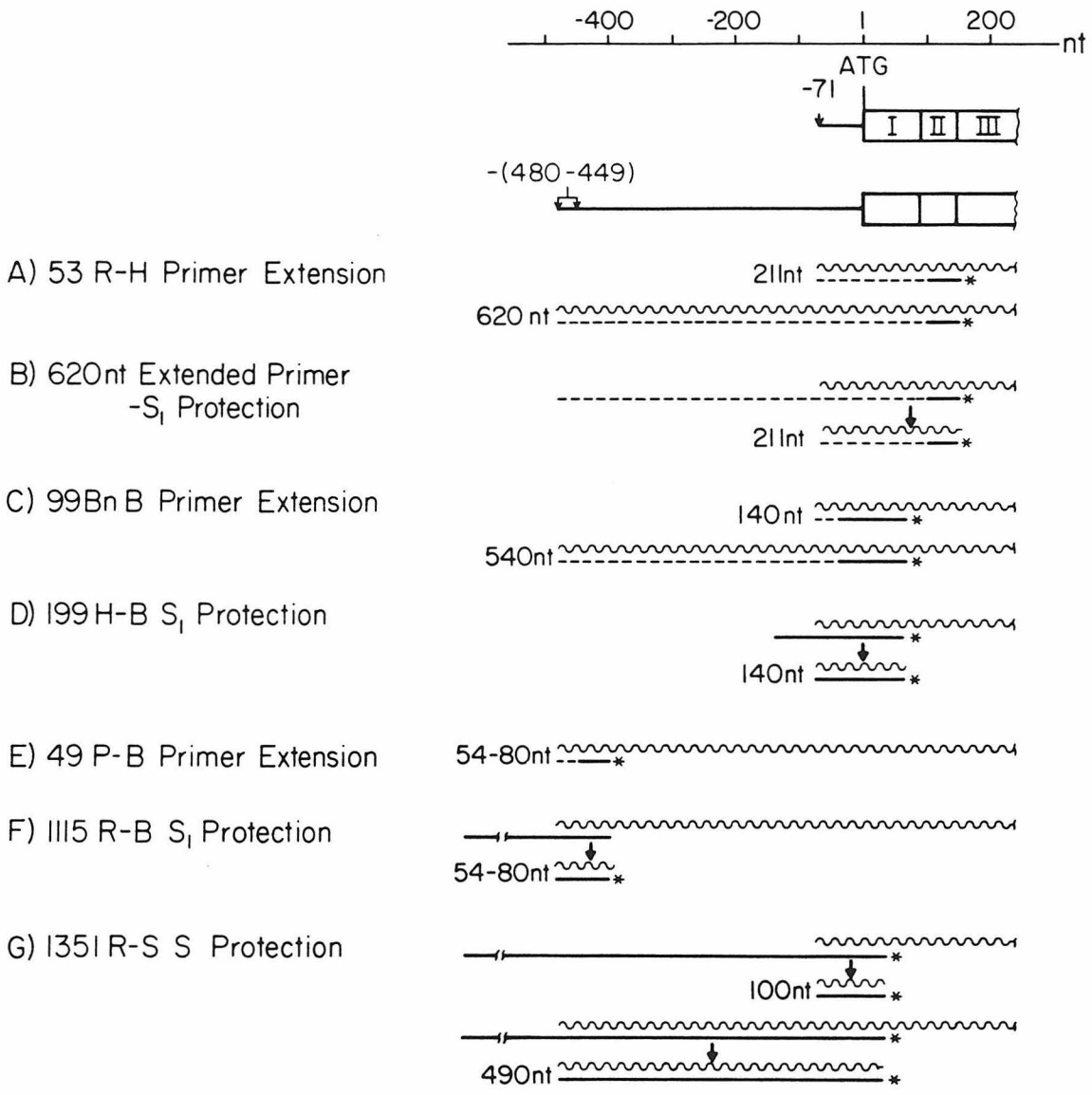
## Figure 1

Restriction fragments used in the primer extension and S1 protection experiments. (a) pHD84 cDNA insert: the DHFR coding regions contained in the six exons are shown. The restriction sites utilized to isolate the 53 bp fragment (53R-H) used for primer extension analysis are indicated, and the fragment is aligned below. (b) The pBH31R1.8 insert is diagramed to show the DHFR coding regions that it contains and the restriction sites used to generate the primers utilized in the extension experiments (99 Bn-B, 49 P-B) and the fragments utilized in the S1 protection experiments (199 H-B, 1115 R-B, 1351 R-S): these fragments are aligned below. For the purpose of the above mentioned experiments, each fragment was labeled with $^{32}P$ at the 5'-end of the non-sense strand. Restriction site designations: B, *Bst*NI; Bn, *Ban*II; H, *Hin*fI; P, *Pvu*II; R, *Eco*RI; S, *Sfa*NI.


500bp
A
ATG
R H
TAA
I II III IV V VI
pHD84 Insert
53 R-H
B
R
P B
H Bn
S B
I
R
pUH3IR1.8
Insert
1351 R-S
199 H-B
1115 R-B
99 Bn-B
49 P-B

## Figure 2

Diagrammatic representation of the primer extension and S1 protection experiments utilizing the fragments shown in Figure 1. The transcripts to which the probes hybridize and the sizes of the extended primers and S1-resistant products obtained are shown. The scale at the top represents the nucleotide numbering of the human DHFR reading frame and upstream region. The DHFR transcripts corresponding to the major and minor 5'-ends are diagramed below. Wavy lines represent RNA molecules , solid lines, the DNA strands (with a star indicating the labeled end), and dashed lines, the extended primers.


-400
-200
1
200
nt
ATG
-71
I
II
III
-(480-449)
A) 53 R-H Primer Extension
211nt
620 nt
B) 620nt Extended Primer
-S1 Protection
211nt
C) 99Bn B Primer Extension
140nt
540nt
D) 199 H-B S1 Protection
140nt
E) 49 P-B Primer Extension
54-80nt
F) 1115 R-B S1 Protection
54-80nt
G) 1351 R-S S Protection
100nt
490nt

## Figure 3

Detection of minor human DHFR specific transcripts mapping with their 5'-ends upstream of the major transcripts. (a) Lane 1: 1 μg of 6A3 cell polysomal poly(A)-containing RNA was reverse transcribed using the pHD84 derived 53R-H primer, and the products were electrophoresed through a denaturing 5% polyacrylamide gel. Lane 2 is a longer exposure of lane 1 to show a slower migrating band designated β. P indicates the unreacted primer and $P^*$ indicates the portion of the primer which reannealed with its complement strand and was extended four nucleotides by reverse transcriptase. M1 shows human mtDNA, and M2, pBR322 DNA, both digested with *Hpa*II and 3'-end labeled with ($\alpha$-$^{32}$P)-dCTP and -dGTP and the large fragment of *E. coli* DNA polymerase I. (b) 10 μg of 6A3 cell polysomal poly(A)-containing RNA were used in a primer extension reaction with the 53R-H primer; after fractionation of the products through a denaturing agarose gel, the material corresponding to the band β was eluted, and one-half was used in an S1 protection experiment with 6A3 RNA. The S1-resistant products were then electrophoresed through a denaturing polyacrylamide gel (lane 2), in parallel with a sample of the products of the primer extension reaction (lane 1) and with the other half of the isolated β fragment (lane 3). M1 and M2 contain the same markers as in (a).

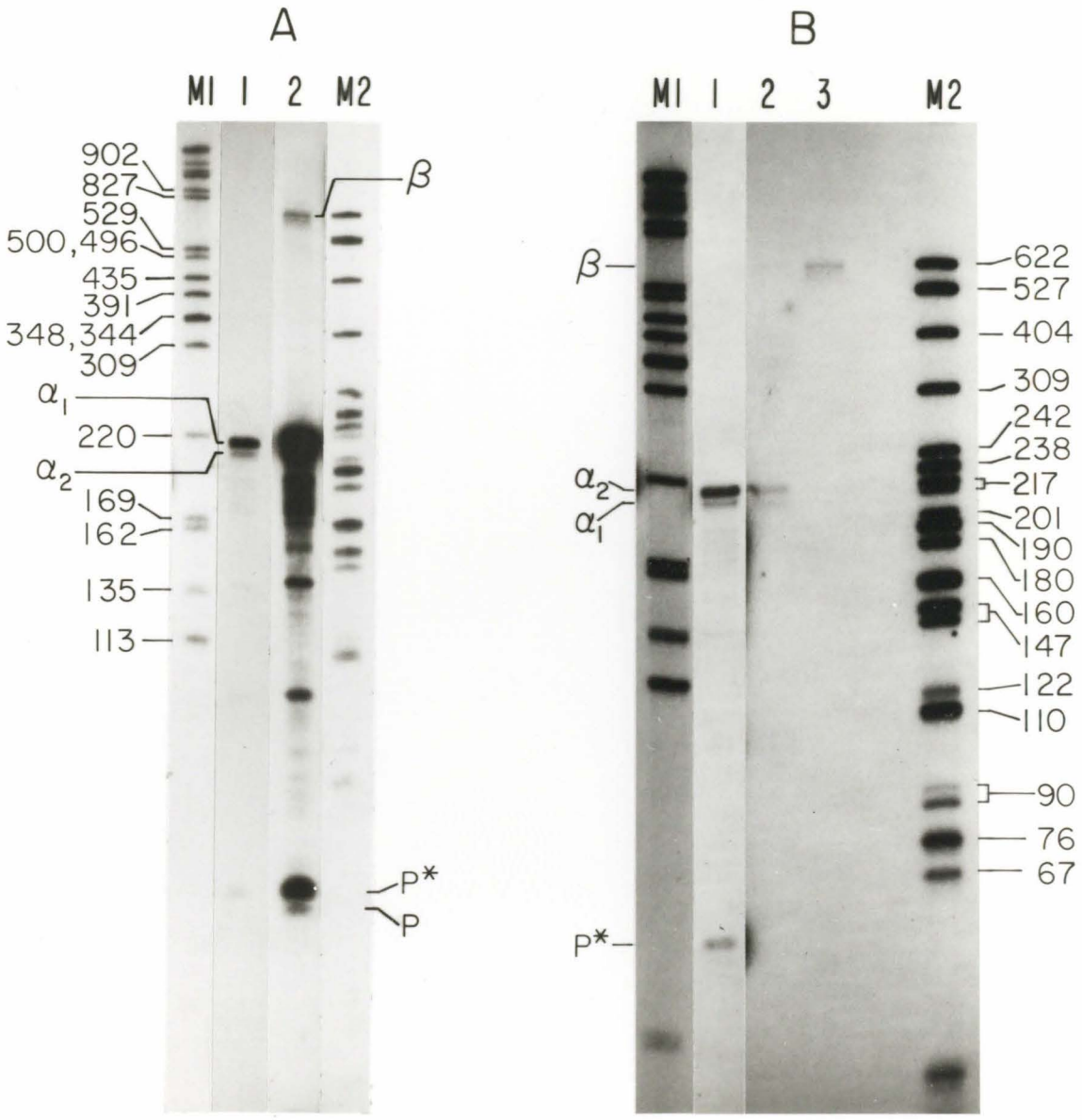
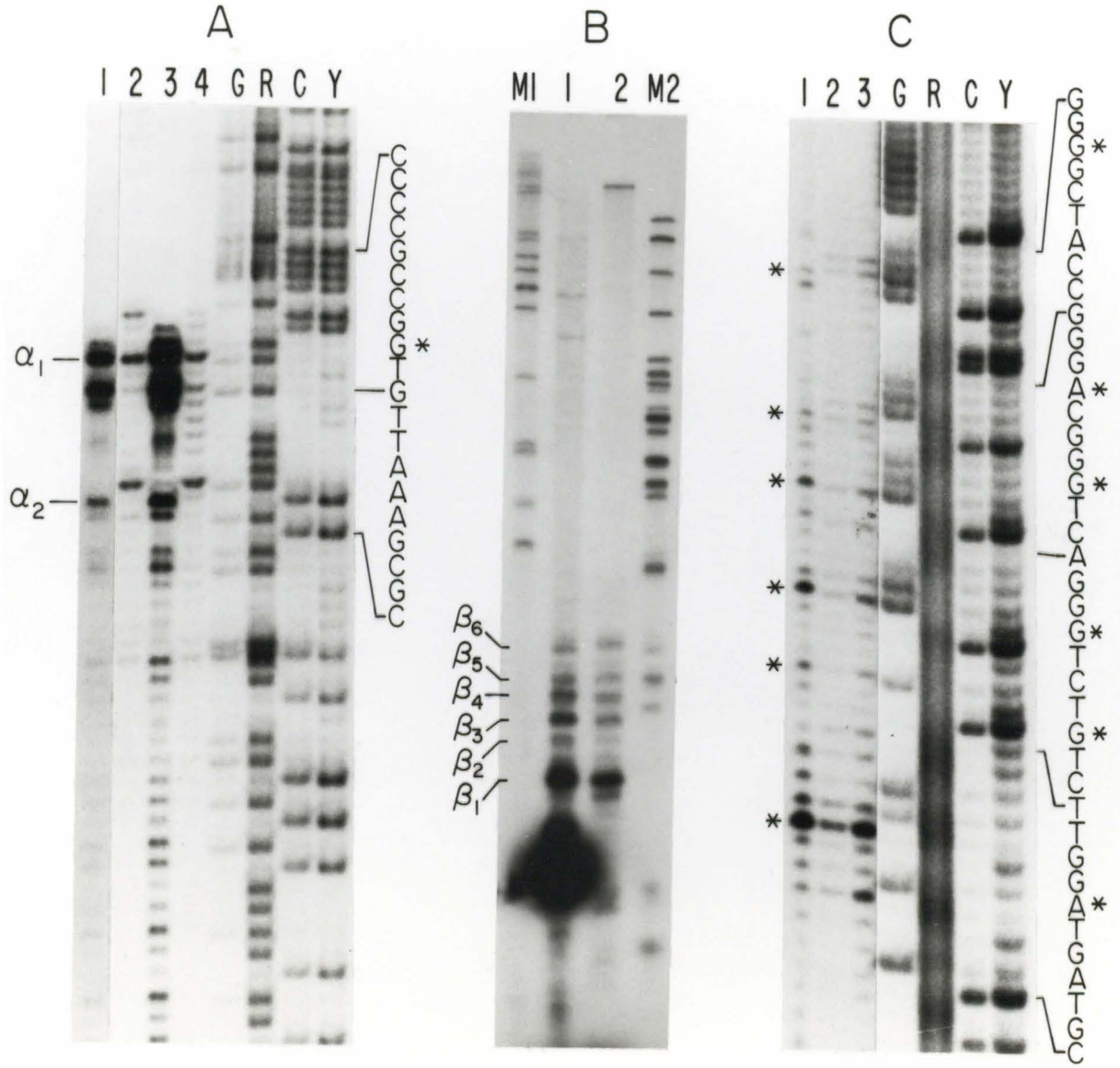

A
M1
1
2
M2
902
827
529
500,496
435
391
348,344
309
α1
220
α2
169
162
135
113
β
P*
P
B
M1
1
2
3
M2
β
α2
α1
P*
622
527
404
309
242
238
217
201
190
180
160
147
122
110
90
76
67

**Figure 4**

Detailed mapping of the 5'-ends of the RNA species corresponding to the bands $\alpha_1$, $\alpha_2$ and $\beta$. (a) Electrophoresis through an 80 cm sequencing gel of the primer extension and S1-resistant products obtained using 6A3 cell polysomal poly(A)-containing RNA and the 199 nt and the 99 nt probes, respectively, (Fig. 1b). Lane 1: extended primers produced using the 99 nt fragment as a primer. Lane 2: S1-resistant products obtained with the 199 nt fragment using 80 units of S1 nuclease at 20°C. Lane 3: long exposure of lane 1. Lane 4: S1-resistant products obtained with the 199 nt fragment using 80 S1 units at 37°C. G, R, C, and Y represent the products of the sequencing reactions (G, G+A, C and C+T) carried out on the 199 nt fragment. The sequence determined is shown, with an asterisk indicating the position of the 5'-end corresponding to $\alpha_1$. (b) The S1 protection and primer extension products obtained with 6A3 cell polysomal poly(A)-containing RNA, using the 1115 nt and the 49 nt fragments of pBH31R1.8, respectively, were electrophoresed through a 20 cm denaturing polyacrylamide gel. Lane 1: products of the primer extension reaction using the 49 nt probe. Lane 2: S1-resistant products obtained with the 1115 nt fragment using 200 units of S1 nuclease at 37°C. M1 and M2 contain the same markers as in Figure 3. (c) Samples of the S1 protection and primer extension products from the experiments shown in panel B were electrophoresed through an 80 cm sequencing gel. Lane 1: A portion of the sample analyzed in panel B, lane 1. Lane 2: S1-resistant products obtained with the 1115 nt fragment using 100 units of S1 at 37°C. Lane 3: A portion of the sample analyzed in panel B, lane 2. G, R, C and Y represent the products of the sequencing reactions (G, G+A, C and C+T) carried out on the 1115 nt fragment. The sequence is shown, with asterisks indicating the positions of the 5'-ends corresponding to $\beta_1$ to $\beta_6$.


A
1 2 3 4 G R C Y
α1
α2
B
MI 1 2 M2
β6
β5
β4
β3
β2
β1
C
1 2 3 G R C Y

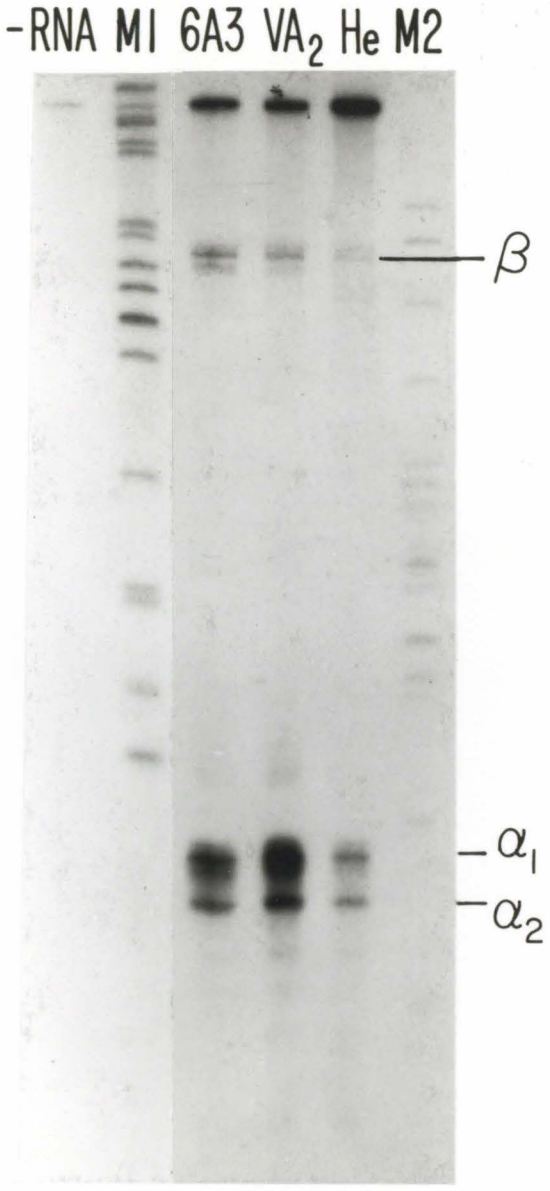
## Figure 5

Alignment of the major and minor 5'-ends of DHFR specific RNAs with the sequence upstream of the human DHFR reading frame. The sequence is numbered according to Yang *et al.* (1984); the sequence preceding the *Pst*I site at position -394 is from Chen *et al.* (1984). The major 5'-end at position -71 is indicated by the arrow, and the minor 5'-ends are indicated by asterisks. The "TATA-like" sequence beginning at position -101 is indicated by a bracket over the sequence. The G-rich direct repeats at positions -127 to -111 and -93 to -77 are underlined with a dashed line, and an inverted repeat of the sequence at positions -93 to -77 is underlined at positions -497 to -480. The DHFR translation initiation codon is boxed; another ATG in the upstream sequence, in frame with the DHFR coding sequence (position -225), is doubly underlined. A third ATG in the upstream sequence (position -239) , not in frame with the DHFR coding sequence, as well a termination codon (TAG, position -95) in frame with this ATG, are singly underlined.


-524 CCCGCCTCCGCCCGCCTTACCCACAAGCCCCGCCCCCCCAGCCCCGATGGCCCTGCCCAG -465
-464 TCCCAGACAGAACCTACTACGTGCGGCGGCAGCTGGGGCGGGAAGGCGGGCGCTGGGGGC -405
-404 GCTGCGGCCGCTGCAGCGCCAGGGTCCACCTGGTCGGCTGCACCTGTGGAGGAGGAGGTG -345
-344 GATTTCAGGCTTCCCGTAGACTGGAAGAATCGGCTCAAAACCGCTTGCCTCGCAGGGGCT -285
-284 GAGCTGGAGGCAGCGAGGCCGCCCGACGCAGGCTTCCGGCGAGACATGGCAGGGCAAGGA -225
-224 TGGCAGCCCGGCGGCAGGGCCCGGCGAGGAGCGCGAACCCGCGGCCGCAGTTCCCAGGCG -165
-164 TCTGCGGGCGCGAGCACGCCGCGACCCTGCGTGCGCCGGGGGCGGGGGGGGCGGGGCCTCGC -105
-104 CTGCACAAATAGGGACGAGGGGGCGGGGCGGCCACAATTTCGCGCCAAACTTGACCGCGC -45
-44 GTTCTGCTGTAACGAGCGGGCTCGGAGGTCCTCCCGCTGCTGTCATG 3

## Figure 6

Occurrence of the major and minor 5'-ends of DHFR specific RNAs in cell lines with amplified DHFR genes ($VA_2$-B cells, 6A3 cells) and with non-amplified genes (HeLa cells). Samples of polysomal poly(A)-containing RNA from HeLa cells (10 µg), $VA_2$-B cells (10 µg) and 6A3 cells (0.5 µg) were hybridized with the 1351 nt fragment from pBH31R1.8 (Fig. 1B), digested with 80 units of S1 at 37°c, and electrophoresed through a denaturing 5% polyacrylamide gel. M1 and M2 contain the same markers as in Figure 3; -RNA: S1-resistant products formed without added RNA.


-RNA
M1
6A3
VA2
He
M2
β
α1
α2

## Figure 7

Primer extension analysis of nuclear and polysomal RNA from 6A3 cells. the RNA samples were hybridized with the 53R-H fragment (Fig. 1B), and the primer extension products were electrophoresed through a denaturing polyacrylamide gel. Lane 1: 1 μg nuclear poly(A)-containing RNA; lane 2: 1 μg polysomal poly(A)-containing RNA; lane 3, 5 μg nuclear non-poly (A)-containing RNA; lane 4: 5 μg polysomal non-poly (A)-containing RNA.


1
2
3
4
β
α1
α2