Dihydrofolate Reductase of Drosophila
CLONING AND EXPRESSION OF A GENE WITH A RARE TRANSCRIPT*

Hong Hao, Michael G. Tyshenko, and Virginia K. Walker‡
From the Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

Traditionally, dihydrofolate reductase (DHFR) has been isolated and the corresponding gene cloned from drug-resistant cell lines which have amplified DHFR genes after selection. A Dhfr sequence has now been obtained by nested polymerase chain reaction (PCR) from Drosophila bearing a single gene copy. Using the PCR-amplified partial cDNA as a probe, Dhfr was cloned by screening a Drosophila genomic library. It consists of regulatory regions as well as a 599-nucleotide coding region with a single 50-base pair (bp) intron and encodes a protein of 182 amino acids. Previously we have shown that the enzyme has kinetic properties characteristic of both "prokaryotic" and "eukaryotic" DHFRs. Here we show that the organization of Drosophila Dhfr is strikingly different from its mammalian counterparts and most similar to that of mosquito. A 790-bp transcript was detected by Northern blot analysis, with a single transcription start site located 27 bp upstream of ATG codon. The Drosophila genome contains a single Dhfr copy at 89F and a selected cell line has not amplified the gene. Confirmation of the identity of this gene has been obtained by kinetic studies of recombinant DHFR over-expressed in Escherichia coli cells.

Dihydrofolate reductase (DHFR, EC 1.5.1.3) is an important enzyme involved in the de novo synthesis of purines, pyrimidines, and glycine. Its critical role in intermediary metabolism has made it a target for anti-folate drugs and a focus for the study of resistance mechanisms and gene amplification. Since the first DHFR gene (Dhfr) was reported from an anti-folate-resistant murine cell line (1), other Dhfr loci have been cloned from a variety of species, including human (2), hamster (3), yeast (4, 5), Escherichia coli (6), Leishmania (7), Lactobacillus casei (8), and mosquito (9). The initial cloning of Dhfr from many organisms has been facilitated by the generation of cell lines with amplified copies of this gene. For example, mouse Dhfr was isolated from a methotrexate (MTX)-resistant S-180 cell line with 200 copies of Dhfr (10); similarly the mosquito gene was isolated from a MTX-resistant cell line with 300 copies. The overall structure and organization of cloned mammalian Dhfr loci shows extensive similarity (11). The mammalian genes (from human, mouse, and hamster) span approximately 30 kb and have five introns. Many transcripts of these genes have been detected and result from multiple initiation sites and/or multiple 3'-polyadenylation sites. The genes have similar intron positions and conserved exon/intron boundaries but show great divergence in intron size, with the exception of the first, and sequence (12). Interestingly, the human, mouse, and Chinese hamster Dhfr loci have bidirectional promoters. One such mouse gene with a 4.0-kb transcript shows homology to the bacterial DNA mismatch repair genes Hex A and Mut S (13).

The purification and characterization of an insect DHFR was reported only recently (14, 15) due to the low level of this protein in these MTX-sensitive organisms. The isolation of the corresponding gene from Drosophila is important not only for the wealth of genetic, cytological, and molecular information available for this insect but because of the opportunity to develop insect-specific antimetabolite agents. We anticipate that such agents, or indeed mutations in this vital gene, will lead to embryonic lethality. In order to study Drosophila Dhfr, MTX-resistant cells were selected over a 4-year period from a S3 cell line. It was expected that selection would result in an overamplification of Dhfr as in other species. Surprisingly, our highly resistant cell line (200 μM MTX) did not show overproduction of DHFR or its mRNA. Therefore, we had to clone Dhfr from a DHFR limited source (representing about 0.001% of the soluble protein; Ref. 14).

MATERIALS AND METHODS

Standard Techniques—Standard procedures were used to carry out phage and plasmid DNA isolation, ligation, and transformation, DNA blotting, and RNA blotting (16). DNA probes were radiolabeled by nick translation (Life Technologies, Inc. labeling kit) and purified as suggested by the manufacturer. DNA bands from low melt agarose gel electrophoresis used for subcloning were purified by Sephasag Band-Prep kit (Pharmacia LKB Biotechnology Inc.).

DHFR Assay, Purification, and Internal Amino Acid Sequencing—DHFR enzyme activity in cell extracts and purified protein preparations was determined as described previously (14). The procedures for purification of DHFR from Drosophila have been described previously (14). 5 μg of partially purified protein from 18 g of adult flies were separated on a 12% polyacrylamide gel and electroblotted overnight onto a nitrocellulose membrane at 35 V, 4 °C. The protein on the nitrocellulose membrane was subjected to internal amino acid sequencing according to the procedures described by Aebersold et al. (17) at the Harvard Microchemistry Facility.

To purify recombinant DHFR, 100 ml of transformed E. coli were harvested by centrifugation and resuspended in 100 ml potassium phosphate buffer, pH 7.9, 1 mM dithiothreitol, and 1 mM EDTA. Cells were lysed by egg white lysozyme (100 μg/ml) final concentration) on ice for 15 min, followed by three successive freeze-thaw cycles. The enzyme was precipitated by (NH4)2SO4 fractionation (50–70% saturation) and dialyzed against 10 mM KCl, 1 mM EDTA, 2 mM dithiothreitol, and 100 mM Tris-HCl, pH 7.5, overnight. The dialysate was bound to Aff-Gel blue column (Bio-Rad) and eluted with 1 mM KC. The DHFR-containing fractions were dialyzed and 1 mM dithiothreitol and 0.1 mg/ml bovine serum albumin was added for kinetic studies. The K_v values for the substrate and cofactor of recombinant Drosophila DHFR as well as the inhibition and dissociation constant for trimethoprim and [3H]MTX (Amerham Corp.) were determined as described previously (15). These values were determined for three times on at least two preparations of recombinant DHFR.

* This work was supported in part by a Queen's Graduate Scholarship (to H. H.) and by an Operating Grant (to V. K. W.) from the Natural Sciences and Engineering Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequences(s) reported in this paper has been submitted to the GenBankEMBL Data Bank with accession number(s) U06861.

‡ To whom correspondence and reprint requests should be addressed.

Tel: 613-545-6123; Fax: 613-545-6166

1 The abbreviations used are: DHFR, dihydrofolate reductase; MTX, methotrexate; bp, base pair(s); kb, kilobase pair(s).

2 These values were determined for three times on at least two preparations of recombinant DHFR.
Drosophila DHFR Gene

Nested Polymerase Chain Reaction Amplification—Oligonucleotide primers were synthesized on an Applied Biosystems 380 A DNA Synthesizer (Applied Biotics, Inc.). First strand cDNA was produced by RNase H- reverse transcriptase (Moloney murine leukemia virus reverse transcriptase, Life Technologies, Inc.) following the manufacturer's protocol. After completion the volume was made up to 250 μl. 1 μl of first strand cDNA solution or 1 μg of Drosophila genomic DNA was used as a template for the first round of amplification in the presence of 10 mM Tris, pH 9.0, 1.5 mM MgCl₂, 500 mM KCl, 1% Triton X-100, 0.2 mM each deoxyribonucleotide triphosphates, and 50 pmol of each primer. A 1-pl aliquot of the first amplification reaction was added to the nested PCR reaction, which was performed with the same components as the first PCR reaction except using the two inner primers.

Southern Blot Analysis—DNA Sequencing—cDNA and genomic DNA fragments to be sequenced were subcloned into BlueScript SK+ vector. Automated DNA sequencing was carried out by Applied Biosystems 373A fluorescent sequencing apparatus using Sequenase terminator cycle sequencing ready reaction kit (United States Biochemical Corp.) and dideoxynucleotide triphosphates. The sequencing was carried out by Applied Biosystems 373A fluorescent sequencer using Applied Biosystems 373A fluorescent sequencing apparatus using Sequenase terminator cycle sequencing ready reaction kit (United States Biochemical Corp.) and dideoxynucleotide triphosphates. The sequencing was carried out by Applied Biosystems 373A fluorescent sequencer using Applied Biosystems 373A fluorescent sequencing apparatus using Sequenase terminator cycle sequencing ready reaction kit (United States Biochemical Corp.) and dideoxynucleotide triphosphates.

RESULTS

Cloning of Drosophila Dhfr and Northern Analysis—A number of strategies were employed in the attempt to clone Dhfr from wild type Drosophila. Degenerate oligonucleotide probes based on the amino-terminal sequence of Drosophila DHFR (14) were hybridized to both Drosophila genomic and cDNA libraries and used to prime first strand cDNA for the rapid amplification of cDNA ends (RACE-PCR). The Drosophila libraries were also hybridized with heterologous Dhfr cDNAs from mouse and mosquito under low stringency conditions. Additionally, a Drosophila cDNA expression library was screened with two Drosophila DHFR-specific antibodies (from purified native DHFR and denatured DHFR). The analysis of these results are presented under "Discussion."

The failure of these standard methods led us to employ an alternative way to clone Dhfr. DHFR preparations which were eluted from an Affi-Gel blue column (14) were electroblotted onto a nitrocellulose membrane followed by in situ protease digestion. The resulting peptide fragments were separated by reverse-phase high performance liquid chromatography and subjected to sequence analysis. Four degenerate oligonucleotide primers were synthesized according to the internal and confirmed amino terminal sequence of DHFR (Fig. 1A). PCR was performed using first strand cDNA as template for amplification by primers 1 and 4 (Fig. 1A, lane 1); an aliquot of this reaction was then reamplified with one of two combinations involving nested primers (primers 1 and 3, lane 2; primers 2 and 3, lane 3). In both cases, a unique amplified band with the expected molecular weight of about 500 bp was obtained (Fig. 1A, lane 2 and 3). Analysis of the sequenced amplification product revealed homology with other Dhfrs.

Cloning of a full-length Dhfr, the PCR amplified Dhfr cDNA was used as a probe to screen Drosophila genomic and cDNA libraries. No positive cDNA recombinant phage were observed; however, genomic library hybridization yielded two phage (D2, D3). A 3.5-kb SalI fragment from D2 (Fig. 2A) hybridized to the Dhfr cDNA probe and was subcloned into pBluescript SK+ vector and sequenced using the strategy shown in Fig. 2B.

To determine the copy number of Drosophila Dhfr, DNA from adult flies was digested to completion with a variety of restriction enzymes (Fig. 3) and hybridized with PCR-amplified Dhfr. A unique labeled band was observed in each lane of the Southern blot. A single hybridization band was localized to the right arm of the third chromosome at position 89E in polytene chromosomes after hybridizing with biotin-labeled Dhfr genomic DNA from the phage D2 insert (Fig. 4).

Northern analysis of Drosophila Dhfr was performed using 20 pg of poly(A)+ RNA from adult flies and hybridized with PCR-amplified Dhfr cDNA. A 790-bp transcript was detected (Fig. 5), sufficient to encode DHFR. Primer extension analysis using 10 μg of poly(A)+ RNA resulted in a 100-bp extension product (Fig. 6), whereas 50 μg of total RNA was not sufficient to yield a detectable extension product signal. No minor bands were observed even after longer exposure, indicating the absence of minor transcription start sites.

Organization of Drosophila Dhfr—Both strands of the 3.5-kb genomic SalI fragment were sequenced (Fig. 2B), as well as the cDNA amplified by nested PCR. The coding region of Drosophila Dhfr spans 599 nucleotides and encodes a 182-amino acid protein (Fig. 7). Comparison of the genomic Dhfr sequence

2 M. Wolfner, personal communication.
Drosophila DHFR Gene

Fig. 1. A, nested PCR amplification of a Dhfr cDNA fragment. First strand cDNA from adults was amplified using the two outer primers (primers 1 and 4) at 40 cycles of 94 °C, 1 min; 46 °C, 1 min; and 72 °C, 1 min (lane 1). A 1-μl aliquot of the PCR reaction was reamplified with primers 1 and 3 (lane 2), primers 2 and 3 (lane 3), primer 1 alone (lane 4), primer 2 alone (lane 5), and primer 3 alone (lane 6), in cycles of 94 °C, 1 min; 56 °C, 1 min; and 72 °C, 1 min, repeated 30 times. HindIII-digested λ DNA marker is shown on the right.

B. DNA sequencing strategy. The arrows indicate the direction and extent of each sequencing run.

C. Internal Sequence

VAPDSDMLPLQOEGGKFEYK

with its cDNA revealed a single 50-bp intron which interrupts the Lys codon at amino acid position 27. The location of the intron is conserved compared with the single mosquito intron and with the first intron in mammals. The splice junctions with a GT dinucleotide in the 5' junction sequence and an AG dinucleotide in the 3' junction sequence conform to the "Chamber" rule (23). The putative branch-point sequence CTAAA in the intron is consistent with the consensus branch-point sequence CTAAT in Drosophila genes (24).

Potential promoter elements were identified within the 933 bp upstream of the ATG translation start site. Primer extension revealed that the single transcription start site was located at 27 bp upstream of the ATG start codon, which is 23 bp downstream of a TATA sequence (Fig. 7). Although several putative TATA boxes were detected at further upstream regions: positions -218 to -223, -233 to -237 and -362 to -367, there was no experimental evidence of their function. The 5'-flanking region of Drosophila Dhfr appears to lack the 48-bp GC box which is conserved in the promoter region of mammalian Dhfr loci. A classic AATAAA polyadenylation signal was not found in the 3'-flanking sequence, but like murine Dhfr (10) polyadenylation, may occur at an alternate sequence, TAAAAT, which is
Drosophila DHFR Gene

Fig. 4. Chromosome location of Drosophila Dhfr. Polytene chromosome squashes from third instar larvae were hybridized to biotin-labeled D2 phage insert which contains full-length Dhfr. The arrow points to the hybridization signal at band 89E.

Fig. 5. Northern analysis. A single 790-bp transcript from Dhfr was detected. 20 μg of poly(A)+ RNA isolated from adult Canton S flies was electrophoresed, transferred to a Hybond-N membrane, and hybridized to a 32P-labeled Dhfr cDNA probe. The Northern blot was exposed for 7 days at -70 °C.

adjacent to the translation stop codon (Fig. 7).

Drosophila DHFR Is Homologous to the DHFRs from Other Species—The DHFR sequence from Drosophila was compared with the DHFR from human, mouse, chicken, yeast, mosquito, and E. coli (Fig. 8). Drosophila DHFR has 49% identity with mammalian DHFR, 27% identity with E. coli DHFR, 34% identity with yeast DHFR, and 58% identity with mosquito DHFR. Examining the conserved residues in the alignment reveals a total of 24 residues conserved between Drosophila and other species. Of these, it is thought that 9 residues are involved in NADPH binding, 5 residues are involved in MTX binding, and 3 residues are involved in both NADPH and MTX binding.

Expression of Drosophila DHFR in E. coli—In order to produce large quantities of DHFR for structural studies and kinetic characterization as well as to confirm the identity of the gene, Drosophila Dhfr cDNA was cloned into a pEYcHis expression vector and transformed into TOP10 E. coli cells. After isopropyl-1-thio-β-D-galactopyranoside induction an increasingly intense band with a molecular mass of 27 kDa was observed on Western blots using Drosophila DHFR-specific antisera (Fig. 9A). A few lower molecular weight bands were also detected and were probably due to site-specific protease degradation of DHFR. The higher molecular weight of the expressed recombinant DHFR is due to the polyhistidine leader peptide fused to the amino terminus of Drosophila DHFR. Increased DHFR activity in cell lysates at various times after induction was seen compared with a control (TOP10 cells transformed with expression vector alone) (Fig. 9B), and this increase correlated well with the more intense protein band on the Western blot. DHFR specific activity was 3800 times greater in transformed E. coli cells as compared with crude homogenates of wild type flies.

Kinetic Characterizations of Recombinant Drosophila DHFR—To further characterize the recombinant Drosophila DHFR, kinetic studies were carried out using purified enzyme preparations. The $K_m$ values for NADPH and dihydrofolate were determined from primary and secondary Hanes plots and shown to be 11.28 and 4.71 μM, respectively. The $K_i$ value for trimethoprim, a competitive inhibitor of DHFR, was determined to be 95.95 μM. The dissociation constant for MTX was measured by equilibrium dialysis of the enzyme preparation against [3H]MTX, the $K_d$ value was 0.30 nm.

DISCUSSION

Our interest in cloning Dhfr from Drosophila was to study this important “housekeeping” gene in a model insect, to understand the molecular genetic basis of MTX resistance in a Drosophila cell line, and further to use this as a model for the
eventual development of insect-specific antimetabolite agents. Cloning of Dhfr from other species has been greatly facilitated by cell lines in which this gene has undergone amplification. Although the Drosophila cell lines were resistant to more MTX than the control cell lines, there was no evidence of increased DHFR activity and no Dhfr gene amplification (not shown). Thus, without amplified cell lines the isolation of such a meagerly expressed gene was extremely difficult. Initial attempts at cloning Dhfr failed. In hindsight, the heterologous Dhfr probes from mouse and mosquito show low overall sequence homology with Drosophila. The eventual development of insect-specific antimetabolite agents.

Third, there were two discrepancies in the amino-terminal amino acids as determined by Edman degradation and nucleotide sequencing, resulting in the synthesis of two oligonucleotide probes with mismatches.

FIG. 7. Nucleotide sequence and deduced amino acid sequence of Drosophila Dhfr. The sequence shown was determined on both strands of the genomic and cDNA clones. The first base of the initiation codon is designated as nucleotide 1. The TATA box is underlined, and the transcription start site is indicated by a bold arrow. The amino acid initiation codon (ATG) and termination codon (TAA) are boxed. The deduced protein sequence is shown in single-letter code and aligned above the nucleotide sequence. The gap between the protein sequence indicates the presence of the 50-bp intron.

FIG. 8. Alignment of Drosophila DHFR with other DHFRs. The sequence of other DHFRs are from human (2), mouse (25), chicken (26), yeast (5), mosquito (9), and E. coli (6). The Drosophila DHFR sequence is shown in bold. Residues that are identical to human DHFR are indicated with an underline.

By determining an internal sequence as well as the amino-terminal sequence of Drosophila DHFR and performing nested PCR, we have cloned the cDNAs from a low abundance DHFR source. We have determined some oligonucleotide probes with mismatches.
mosquito. Both insect Dhfrs are about 1 kb long and contain a single intron, compared with mammalian Dhfrs which are about 30 kb in length and include five introns. Dhfr has an intron of 50 bp, one of the smallest Drosophila introns found to date (24). In contrast to mammalian Dhfr with introns which show virtually no sequence homology except around splice junctions (2, 12), the Drosophila Dhfr intron has 58% sequence homology with the mosquito intron, equal to the identity shared by the exons of these two Dhfrs. Thus, it seems that the exons and intron of Dhfr in these two insects are evolving at the same rate from an ancestral dipteran Dhfr.

A striking difference between the Drosophila and mammalian Dhfr is at the 5' promoter region of the gene. TATA-like sequences have been identified in Drosophila Dhfr, but mammalian Dhfrs lack a TATA box, replacing it with 48 bp of tandem repeated GC boxes (2, 28, 29). These GC boxes have binding sites for transcription factor Sp1 and function as bidirectional promoter elements. Only a short divergent open reading frame located from -356 to -727 (372 bp long) was uncovered by computer analysis of the Drosophila Dhfr, and this putative transcript has no significant homology with other sequences in GenBank (data not shown). Mammalian Dhfrs usually have multiple mRNA species due to heterogeneity at both the 5'- and 3'-nontranslational regions of mRNAs (2, 29-31). Mosquito Dhfr also shows multiple transcription initiation sites at the 5' end, but has a single polyadenylation site (32).

Unlike its mammalian and mosquito counterparts, Drosophila Dhfr has only one detectable transcription initiation site located 27 bp upstream of the ATG codon. This transcription start site lies 23 bp downstream of a TATA sequence. Based on Northern and primer extension analysis, the 3' end of the Drosophila Dhfr transcript is about 210 bp long. DHFR assays and Western analysis confirmed the identity of the recombinant DHFR. There were some differences in the kinetic data for the substrate, cofactor, and inhibitors of the recombinant Drosophila enzyme; the $K_m$ for NADPH and DHFR was 2- and 16-fold higher, respectively, than Drosophila DHFR, and the $K_m$ for MTX was 3-fold lower than the fly enzyme. We attribute these differences to the extra 36 amino acid residues on the relatively hydrophobic amino terminus of DHFR. These additional amino acids may affect normal post-translational modifications or folding of the enzyme and result in the altered kinetic properties.

A comparison of the deduced amino acid sequences of DHFRs from several species (Fig. 7) using a CLUSTAL program of PC-Gene (IntelliGenetics) showed, not surprisingly, that Drosophila DHFR was most closely related to the mosquito enzyme. The hypothetical tree (not shown) also indicates that the insect DHFRs are more similar to those of vertebrates than to those from yeast and E. coli. Curiously, however, Drosophila DHFR showed some characteristics which were typical of non-vertebrate enzymes (15). In addition, insect DHFRs appear to be less sensitive to MTX than other DHFRs. The MTX dissociation constant ($K_d$) of Drosophila DHFR is 34-6615 times higher than that determined for mammalian DHFRs (15). All 17 residues involved in MTX binding (4) are conserved between Drosophila and mosquito DHFR, but only 12 residues are conserved between insect and mammalian DHFRs. Therefore, the other 5 residues in mammals (equivalent to Leu-30, Lys-31, Ser-34, Gly-58, and Val-59 of Drosophila DHFR) may interact more strongly with MTX. The identification of these important residues should facilitate future molecular modeling studies and contribute to the understanding of the structure-function relations of this crucial enzyme.

Drosophila DHFR has both typically vertebrate and prokaryotic physical and kinetic characteristics (15). This dual feature of Drosophila DHFR has now also been demonstrated in its gene structure. As in many Drosophila genes, Dhfr is compact like the lower eukaryotic Dhfr but has more homology with the vertebrate DHFR sequence than that from prokaryotes or lower eukaryotes. Despite these differences between DHFRs from different species, most of the residues involved in the interaction of the protein with its substrates and/or inhibitors are conserved, which reflects the strong evolutionary constraints in the enzyme's structure and function.

Acknowledgments—We are grateful to Tina Schaef er (Core Facility, Insect Biotech Canada) for performing automated DNA sequencing. We thank M. Wolfer and A. Hilliker for their advice on in situ hybridization and acknowledge all the help and encouragement given by our laboratory colleagues.

REFERENCES
1. Alt, F. W., Kellems, R. E., Bertino, J. R., and Schimke, R. T. (1978) J. Biol. Chem. 253, 1357-1370
2. Chen, M. J., Shimada, T., Moulton, A. D., Cline, A., Humphries, R. K., Maizel, J., and Niemuth, A. W. (1984) J. Biol. Chem. 259, 3933-3943
3. Carothers, A. M., Urlaub, G., Ellis, N., and Chasin, L. A. (1983) Nucleic Acids Res. 11, 1997-2012
4. Barclay, B. J., Huang, T., Nagel, G. M., Misener, V. L., Game, J. C., and Wahl, G. M. (1988) Gene (Amst.) 63, 175-185
5. Lugasisky, P. A., Taylor, G. R., and Haynes, R. H. (1987) Nucleic Acids Res. 15, 10555-10571
6. Smith, D. R., and Calvo, J. M. (1980) Nucleic Acids Res. 8, 2255-2274
7. Beverley, S. M., Ellenberger, T. E., and Cordingly, J. S. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 2584-2588
8. Andrews, J., Clare, G. M., Davies, R. W., Gronenborn, A. M., Gronenborn, B., Kalderon, D., Papadopoulos, P. C., Schafer, S., Sims, F. F. G., and Stan-
