Cloning and molecular analysis of the bifunctional dihydrofolate reductase-thymidylate synthase gene in the ciliated protozoan Paramecium tetraurelia

Received: 29 June 1995/ Accepted: 4 October 1995

Abstract We have cloned the first bifunctional gene dihydrofolate reductase-thymidylate synthase (DHFR-TS) from a free-living, ciliated protozoan, Paramecium tetraurelia, and determined its macronuclear sequence using a modified ligation-mediated polymerase chain reaction (PCR) that can be of general use in cloning strategies, especially where cDNA libraries are limiting. While bifunctional enzyme sequences are known from parasitic protozoa, none had previously been found in free-living protozoa. The AT-rich (68%) coding region spanning 1386 bp appears to lack introns. DHFR-TS localizes to a ≈ 500 kb macronuclear chromosome and is transcribed as an mRNA of ≈ 1.66 kb, predicted to encode a 53 kDa protein of 462 residues. The N-terminal one-third of the protein is encoded by DHFR, which is joined by a short junctional peptide of ≈ 12 amino acids to the highly conserved C-terminal TS domain. Among known DHFR-TS sequences, the P. tetraurelia gene is most similar to that from Toxoplasma gondii, based on primary sequence and parsimony analyses. The predicted secondary protein structure is similar to those of previously crystallized monofunctional sequences.

Key words Paramecium · Dihydrofolate reductase · Thymidylate synthase · Ciliate · Bifunctional enzyme

Introduction

Dihydrofolate reductase (DHFR; EC 1.5.1.3; Santi and Danenberg 1984; 6,7,8-tetrahydrofolate: NADP+) and thymidylate synthase (TS; EC 2.1.1.45; 5,10-methylene-tetrahydrofolate: dUMP C-methyltransferase) are essential enzymes in folate metabolism, with reduced folate serving as a methyl donor in the de novo production of dTMP. Interest in mono- and bifunctional DHFR-TS enzymes stems from the central role of these reactions in nucleotide metabolism. The folate pathway has been selectively targeted for drug action in microbial (Hitchings and Baccarani, 1984) and parasitic infections (Hughes 1984; Brooks et al. 1987) and in neoplastic disease (Blakley 1984; Santi and Danenberg 1984).

Unlike animals, fungi, and most bacteria, in which DHFR and TS are encoded by separate genes, in protozoa these activities reside on a bifunctional protein, which in its native form is a homodimer of 110-140 kDa (Ferone and Roland 1980; Garrett et al. 1984; Ivanetich and Santi 1990). Bifunctional DHFR-TS genes have been cloned from the parasitic protozoa Leishmania spp. (Beverly et al. 1986; Grumont et al. 1986; Nelson et al. 1990), Plasmodium spp. (Bzik et al. 1987; Cowman and Lew 1989), Crithidia fasciculata (Hughes et al. 1989), and Toxoplasma gondii (Roos, 1993). More recently, fused DHFR-TS genes have been cloned from two plants, Arabidopsis thaliana (Lazar et al. 1993) and Daucus carota (Luo et al. 1993). Fused DHFR-TS enzymes raise the possibility of substrate channeling during dTMP synthesis, providing a novel target for chemotherapeutic intervention. Unfortunately, it is difficult to model the nature of the possible interactions between DHFR and TS, especially given the long junctional region found in all DHFR-TS genes cloned to date. The evolution of fused DHFR-TS enzymes likewise is obscure. These problems may be addressed via an examination of the structure of DHFR-TS in previously unstudied protozoan groups.
Among protozoa, the ciliates are thought to be most closely related to apicomplexans and dinoflagellates (Gagnon et al. 1993). In this study we describe the cloning and sequencing of the DHFR-TS gene from *Paramecium tetraurelia*, which represents the first DHFR-TS sequence from a free-living protozan. We further explore its mRNA, location by macronuclear chromosome size class, and compare the deduced protein structure to other known DHFR-TS enzymes by protein homology and parsimony analyses. For this study, we modified a ligation-mediated polymerase chain reaction (PCR) technique, which has some advantages over inverse PCR and should be of general use, especially where cDNA libraries are limiting.

**Materials and methods**

Cells and genomic DNA

*P. tetraurelia* stock 51-S (wild type) was cultured as previously described (Sasner and Van Houten 1989). The ciliated protozoan *P. tetraurelia* possesses two kinds of nuclei, a highly polymorphic vegetative macronucleus and two germinal diploid micronuclei (Freiburg 1988). The macronucleus arises in ciliates during conjugation from a fertilized micronucleus through a process of DNA elimination, rearrangement, and amplification (Karrer 1986). Macronuclear genomic DNA was extracted at 65°C in lysing solution (1% SDS, 0.5 M EDTA, 0.1% TRIS pH 9.5; 100 µg/ml proteinase K) according to Forney et al. (1983).

Oligonucleotides

Degenerate primers TS-430 [GATTTGGATCC(A/T)/GT(A/T)/C T(A/T)/CGGA(A/T/T1/T1)T/AATGGGACCA] and TS-515 [GT(A/T)/C1AAT/G1TGCATA(A/T)/GA1GC(A/G/T) AT(A/G)/TTG1/AAAT(T)/AAGG] corresponding to residues 425–436 and 529–519, respectively, of the *P. falciparum* TS region (Bzik et al. 1987) were used for initial PCR. The following primers were used in ligation-mediated PCR: linker primer VT-1 (GGCGGTGACCTTTAGATCTGGATCC); linker adaptor VT-2 (GGATCCAGAT; note complementarity to VT-1); VT-3 (ATGGAGCTCAATTTGCCATTGGGACA); TS-336 (GTCAGGCTTGGAGACTCCTCTCC); and TS-420 (ATTAGATAGGACATGACTAGGACAG). DHFR-TS sequencing templates and probes were produced using primers DT-463 (GAATCTGATAATATTAATTTGAATAAC) and DT-2007 (AATGAGCTCATTTACATTATAAGGGATT). (See Fig. 1 for primer positions.)

Polymerase chain reaction

PCR was carried out using the conditions of the GeneAmp DNA Amplification Reagent Kit (Perkin-Elmer Cetus, Norwalk, Conn., USA) with some modifications. Ligation-mediated PCR was performed according to Mueller and Wold (1991) with the following changes: genomic DNA was treated with specific restriction enzymes for template preparation based on Southern data, and the EcoRI site of the linker primer and adaptor was changed to BamHI.

Cloning and DNA sequencing

PCR products and cloning vector pBluescript II KS+ (Stratagene Cloning Systems, La Jolla, Calif., USA) were cleaved with the appropriate restriction enzymes, gel purified (Heery et al. 1990), and ligated for transformation of competent Escherichia coli strain XL1-Blue (Stratagene) according to standard protocols (Ausubel et al. 1989). Plasmid DNA was purified on push-columns (Stratagene), alkali denatured, and sequenced with Sequenase version 2.0 modified T7 DNA polymerase (US Biochemicals, Cleveland, Ohio, USA) using commercial sequencing primers (Stratagene). The Genetics Computer Group (GCC; Madison, Wis., USA) programs and databases together with manual alignments were used for DNA and protein sequence analyses. Sequences reported in this paper have been deposited in the EMBL/Genbank database under accession number U03885.

Pulsed-field gel electrophoresis and Southern analysis

Standard protocols were used (Sambrook et al. 1989) for agarose block preparation containing 10000 to 40000 cells previously harvested at 7000 rpm in a microfuge. Cells in agarose blocks were deproteinized twice at 50°C for 24 h each in ESP solution (0.5 M EDTA pH 9.0, 1% lauryl sarcosine, 2 mg/ml proteinase K) and stored as described. Clamped homogeneous electric fields (CHEF) gel electrophoresis was performed according to the manufacturer's recommendations using a CHEF-DRII system (Bio-Rad Laboratories, Hercules, Calif., USA) for electrophoresis at 175 V for 16.5 h at a pulse time of 60 s then for 8.4 h at 90 s. Standard procedures for gel to blot transfer, probe labeling, and blot hybridizations were done as previously described (Nicklas et al. 1991).

RNA extraction and Northern analysis

RNA was isolated from 0.3 ml packed cells in 5 ml lysis solution using the guanidinium thiocyanate/acid phenol procedure (Chomczynski and Sacchi 1987). Poly(A)+ RNA was selected using an mRNA spin column kit (5 Prime R⃝–3 Prime, Boulder, Colo., USA). Denaturing gel electrophoresis, transfer of RNA to nitrocellulose (Schleicher and Schuell), probe labeling, blot hybridizations, post-hybridization washes, and autoradiography were as previously described (Shull et al. 1991).

Results and discussion

Ligation-mediated PCR and genomic cloning

Protozoan DHFR-TS enzymes, including that of the ciliate *Tetrahymena*, have long been known to be bifunctional polypeptides (Garrett et al. 1984). The corresponding genes have been cloned from several parasitic protists (Ivanetich and Santi 1990; Beverly et al. 1986; Grumont et al. 1986; Nelson et al. 1990; Bzik et al. 1987; Cowman and Lew 1989; Hughes et al. 1989; Roos 1993) and compared to monofunctional sequences (Blakley 1984; Perry et al. 1990). Although there is very little primary sequence conservation of DHFR between species, TS is well conserved. Therefore we used conserved regions in the TS domain of the *P. falciparum* homology alignment (Bzik et al. 1987) to design degenerate primers TS-430 and TS-515 (see
the Materials and methods for primer sequences) assuming *P. tetraurelia* codon bias for initial PCR (PCR 1; Fig. 1). A specific 0.31 kb amplification product was cloned and sequenced; proper alignment with all conserved amino acids found in other species (Bzik et al. 1987) confirmed this PCR product to be part of the thymidylate synthase gene region.

Due to difficulties in generating and maintaining a *P. tetraurelia* cDNA library, perhaps due to the high AT content, we turned to a modified single-sided ligation-mediated PCR method originally devised for footprinting analysis (see the Materials and methods; Mueller and Wold 1991) to directionally clone sequences upstream of PCR 1. This method allows directional amplification, using even large restriction fragments as template, and thus has some advantages over inverse PCR. A fortuitous *EcoRI* site was found near the 3' end of PCR 1 (Fig. 1). A Southern hybridization gel with *EcoRI*-digested DNA was therefore probed with labeled PCR 1 product to localize the next upstream *EcoRI* site, which was 0.61 kb away (Fig. 1). Genomic DNA was cleaved with *EcoRI* and denatured. First-strand synthesis was carried out by extending primer VT-3, and the product was ligated at its blunt ends to an artificial linker (annealed linker primer VT-1 and adaptor VT-2) to add an artificial primer site. Gene-specific primer VT-3 and linker primer VT-1 were then used to amplify the 0.61 kb DNA fragment (PCR 2 in Fig. 1). Similar experiments employing flanking *NsiI* and *BglII* sites were used to clone additional sequences both upstream and downstream in the gene (PCR 3 and PCR 4 in Fig. 1). PCR amplification products were verified by Southern analysis before subcloning and sequencing (not shown). This combination of diagnostic Southern hybridizations and ligation-mediated PCR provides a general method for direct PCR cloning of DNA flanking a known sequence.

Assuming the absence of introns (at the time this work was carried out, no introns had been described in *P. tetraurelia*, or in any bifunctional DHFR-TS), a 2060 nucleotide (nt) region was expected to include the entire DHFR-TS sequence. Preliminary sequence analysis (Fig. 1C) revealed no characteristic promoter sequences within the first 430 nt of this domain, hence only sequences beyond this point were chosen for more careful examination. Several independent full-length PCR products were cloned from reactions using primers DT-463 and DT-2007, and the complete sequence was determined according to the sequencing strategy shown in Fig. 1D. To guard against PCR artifacts, at least three independent clones were sequenced for each nucleotide.

### Southern and Northern analyses

The bifunctional DHFR-TS gene organization was further analyzed by Southern and Northern blotting. Virtually all *P. tetraurelia* chromosomes separated on CHEF gels (> 99%) are of macronuclear origin, and range in size from 50 to 1050 kb (Fig. 2A). The voltage and pulse settings employed here appear to give better resolution of high molecular weight macronuclear chromosomes than the 800 kb upper limit reported previously (Phan et al. 1989; Caron 1992). In contrast to the G surface antigen which is present on two differently sized macronuclear chromosomes due to genomic rearrangements during macronuclear differentiation (Caron 1992), DHFR-TS appears to be localized to a single size-class of macronuclear chromosome(s) of ≥ 500 kb. The bifunctional gene arrangement is supported by the hybridization of DHFR- and TS-specific probes to the same size chromosome(s) (Fig. 2A), and by the expected hybridization pattern of the probes on Southern blots of DNA cleaved by restriction enzymes (not shown).

Northern analysis revealed a single 1.66 kb transcript including both DHFR and TS sequences (Fig. 2B). This is a significantly smaller DHFR-TS mRNA than the ≈ 3 kb transcripts reported in parasitic protozoa (Hughes et al. 1989; Kapler and Beverly 1989; Washtien et al. 1985; Sartorius and Franklin 1991). *Paramaecium* DHFR-TS is likely to be a rare transcript, since purified poly(A)⁺ RNA rather than total RNA was required for signal detection (Fig. 2B); no signal was detected in 30 µg total RNA on a Northern filter (not shown).
Deduced DHFR-TS protein analysis

The DHFR-TS protein sequence deduced from sequencing of genomic DNA is shown in Fig. 3. The
preceeding ATG (at position — 80) is out of frame with DHFR and TS homology, while downstream ATG (Met) codons lie beyond conserved amino acid residues (Fig. 4). The 53 kDa predicted protein is 462 amino acids long and as such is the shortest bifunctional DHFR-TS identified to date.

From deduced amino acid sequence homologies, the DHFR domain appears to comprise amino acids 1 to \(\approx 165\). The carboxy-terminal region of DHFR displays very low homology with the sequences in other species (Fig. 4) and its precise extent is therefore a matter of conjecture. DHFR is connected to TS by a short \(\approx 12\) residue junctional peptide from positions \(\approx 166\) to 177 with half of the residues being charged. A charged junctional region is also found in the longer apicomplexan DHFR-TS (Bzik et al. 1987; Roos 1993), which is thought to play a role in secondary structure flexibility between the two domains. By homology, TS is thought to be encoded by residues 178–462.

**DHFR-TS homology**

Despite low primary DHFR sequence conservation across phyla, the PROSITE (Bairoch 1992) signature motif is present (IGyqrnLPW; residues 19–27), as is the ubiquitous Gly-(CR)Gly at positions 102–3. Predicted higher-order structure is consistent with known crystal structures of other DHFR enzymes, and is probably capable of forming all known eukaryotic \(z\)-helical, \(\beta\)-strand, and tight turn elements (Fig. 4; Volz et al. 1982; Oefner et al. 1988; Davies et al. 1990). The ciliate sequence differs from some protozoan DHFR domains in lacking the long insertion found after residue 43 in apicomplexan sequences (Fig. 4).

The predicted TS sequence is highly conserved, as expected from other known TS genes and proteins. The PROSITE signature motif is located at positions 342–353 (LpPChvmsQIV). The predicted secondary protein structure of TS is likely to contain all of the \(\beta\)-strands and at least six of the eight \(z\)-helices found in the *E. coli* enzyme (Perry et al. 1996); helices \(z_3\) and \(z_4\) could not be predicted, as is common in other TS sequences (Roos 1993). Ciliate DHFR-TS is expected to be encoded by amino acids 1 to \(\approx 165\) and TS by residues 178–462.

![Fig. 3 Nucleotide and deduced protein sequence of the *P. tetraurelia* DHFR-TS gene. Amino acids are numbered on the left, nucleotides on the right. Ciliate codon usage is denoted by *underlined* glutamine residues (Q) encoded by UAA and UAG. Underlined positions 16–25 of the nucleotide sequence are found in 5° regions of some other *Paramecium* genes (see the text for details). From comparison to other sequences, DHFR appears to be encoded by amino acids 1 to \(\approx 165\) and TS by residues 178–462.](image-url)
to form a dimer in vivo, as four of the five $\alpha$-helices involved in dimerization of TS (Montfort et al. 1990) are clearly predicted in *P. tetraurelia*.

In comparing primary protein sequences of the ubiquitous enzymes DHFR and TS across a wide spectrum of phyla including eukaryotes, prokaryotes, and viruses, the striking observation is the low conservation of DHFR while TS is very highly conserved (Fig. 4). A bifunctional arrangement of DHFR-TS is unique to protozoa and plants. Figure 4 shows the only cloned non-parasitic protozoan and two plant sequences aligned with a variety of bi- and mono-functional DHFR and TS. The ciliate DHFR is most closely related to that of *T. gondii*, with which it shares 58 amino acids (33%) and least similar to phage T4 (36 identical residues; 20%). Across the ten DHFR sequences shown, only 12 residues (7%) are identical while 33 amino acids (19%) are similar (maximum observed variation of 3 residues). However most residues in the TS domain are conserved; 68 residues (24%) are identical and 108 (38%) are similar across ten phyla (Fig. 4). Identical positions range from 172 (60%) in *T. gondii* to 127 (45%) in phage T4.

Several peptide motifs and individual amino acids have been identified in binding substrates or their analogues of DHFR and TS. A photoaffinity labeled analogue of the folate antagonist methotrexate labels a 15-residue stretch in mouse DHFR (Price et al. 1987) identical to the *Lactobacillus casei* folylpolyglutamate binding site (Maley et al. 1982). In the relevant region of the *P. tetraurelia* sequence (amino acids 48–60), 8 positions are identical to the mouse and *L. casei* sequences, and 2 others are similar. Crystallographic studies have identified 17 residues of human DHFR important in water-bridged or direct ligand binding (Oefner et al. 1988; Davies et al. 1990). Eleven of these positions are either identical or similar in the ciliate sequence, including the acidic residue Asp-32 implicated in enzyme catalysis (reviewed in Brown and Kraut 1992). Similarly, 29 residues are involved in NADP$^+$ binding to *E. coli* DHFR (Bystroff et al. 1990) of which 12 are identical and 5 conserved with *P. tetraurelia*.

The pyrimidine binding motif of *L. casei* (amino acids 193–202) is shared by *P. tetraurelia* TS at 6 positions (residues 341–346), including the catalytic Cys-345 (Santi and Danenberg 1984). The 12 amino acids identified by X-ray crystallography as involved in dUMP binding (Montfort et al. 1990) are all shared with the ciliate TS. Two of the three lysines implicated in direct ligand binding are present in ciliate TS. Six *E. coli* residues have been identified by crystallography in binding of CH$_2$H$_2$-folate or an analogue thereof (Montfort et al. 1990); 5 of these positions are shared by the ciliate sequence.

**Fig. 4** Alignments of mono- and bi-functional DHFR-TS protein sequences using the single-letter amino acid code. Numbering is based on the *P. tetraurelia* sequence. The junctional region (residues $\approx$166–177) has arbitrarily been included with DHFR. Alignments are based on primary sequences and on secondary structures derived from crystallographic studies of human DHFR and *Escherichia coli* TS (see the text for details). Asterisks indicate identical residues and 'plus' symbols denote conserved amino acids (maximum of 3 variations observed). Secondary structures in the human DHFR and *E. coli* TS are indicated below the sequences with an inverted 'v' for $\beta$-strands and dashes for $\alpha$-helices. Bifunctional sequences are represented as: Pte, *P. tetraurelia*; Tgo, *Toxoplasma gondii* (Roos 1993); Pfa, *P. falciparum* (Bzik et al. 1987); Lma, *Leishmania major* (Beverly et al. 1986) Ath, *Arabidopsis thaliana* (Lazar et al. 1993). Dca, *Daucus carota* (Luo et al. 1993). Monofunctional sequences are: Hsa, *Homo sapiens* (Masters and Attardi 1983; Takeishi et al. 1985); Eco, *E. coli* (Belfort et al. 1983; Smith and Calvo 1980); pT4, bacteriophage T4 (Chu et al. 1984; Purohit and Mathews 1984).

**Fig. 5** TS phylogeny based on parsimony analysis of all available TS sequences followed by branch pruning to retain the species shown. Heuristic searching algorithms gave rise to the single most parsimonious tree from multiple runs under various starting conditions using midpoint rooting. Branch lengths are proportional to relative distances between TS sequences. Sequences used are given in Fig. 4.
Leishmania clusters with plants rather than with the other protozoa analyzed. Similar results were obtained by Lazar and co-workers (Lazar et al. 1993) during parsimony analysis of plant and protozoan sequences, although the significance of this observation is unclear.

Acknowledgments. We thank Drs. Yvonne Janssen, Malcolm Lippert, and Brain Foley for their respective assistance in Northern analysis, pulsed field gel electrophoresis and Southern analysis, and computer analysis. This work was supported by NIH and the Vermont Cancer Center to J.V.H., and Mac Arthur Fdn., Burroughs Wellcome, Merck Research Labs., NHL and NSF to D.S.R.

References


