

I. Martha Schlichtherle · David S. Roos
Judith L. Van Houten

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-methylenetetrahydrofolate: 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 Bacanari, 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.

Communicated by B. J. Kilbey

I. M. Schlichtherle¹ · D. S. Roos² · J. L. Van Houten (✉)
Department of Biology, University of Vermont,
Burlington, VT 05405-0086, USA

Present addresses:

¹ Microbiology and Tumor Biology Center,
Karolinska Institutet and Smittskyddsinstitutet,
Box 280, 17177 Stockholm, Sweden

² Department of Biology, University of Pennsylvania,
Philadelphia, PA 19104-6018, USA

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 protozoan. 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 polyploid 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)TA(T/C)GG(A/T)TT(C/T)TAATGGAGACA] and TS-515 [GT(T/C)AA(T/C)AATGCATA(A/T)GAIGC(A/G/T)AT(A/G)TT(G/A)AA(T/A)GG] 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(GCGGTGACCTCTTAGATCTGGATCC); linker adaptor VT-2 (GGATCCAGATC; note complementarity to VT-1); VT-3 (ATGGAGCTCAATTTGCCATTTGCCACA); TS-336 (GTCAAGCTTTGAGACTCCCTCCTTC); and TS-420 (ATAGATAGGAACATGACTTAGGACCAG). DHFR-TS sequencing templates and probes were produced using primers DT-463 (GAATCTAGATAATATTAATTGAATAAC) and DT-2007 (AATGAGCTCATTACATTATAAGGGATTA). (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 *Bam*HI.

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 (GCG; 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 10 000 to 40 000 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-DR II 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 → 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. We therefore 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 *BglIII* 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). *Paramecium* 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 μ g total RNA on a Northern filter (not shown).

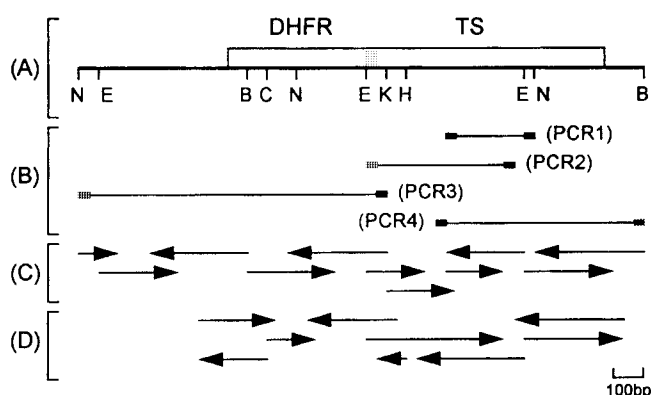


Fig. 1A-D Cloning, sequencing strategy, and restriction map of *Paramecium tetraurelia* dihydrofolate reductase-thymidylate synthase (DHFR-TS). **A** Partial genomic restriction map of the DHFR-TS gene. The coding region is boxed, with the DHFR/TS junctional region shaded. Restriction sites: B, *BglIII*; C, *ClaI*; E, *EcoRI*; H, *HindIII*; K, *KpnI*; N, *NsiI*. **B** Polymerase chain reaction (PCR) strategy. Black boxes symbolize gene-specific primers; grey boxes represent linker primers. PCR 1 (primers TS-430 and TS-515) was based on conserved TS domains and predicted *P. tetraurelia* sequence. PCR 2: ligation-mediated PCR from synthetic VT1/2 primers ligated to the *EcoRI* site to VT3 (based on PCR 1 sequence). PCR 3: ligation-mediated PCR from VT1/2 (ligated to *NsiI* site) to TS-336 (based on PCR 2 sequence). PCR 4: ligation-mediated PCR from synthetic VT1/2 (ligated to *BglIII* site) to TS-420 (based on PCR 2 sequence). **C** Sequencing strategy of ligation-mediated PCR templates. **D** Final sequencing strategy of region amplified using primers DT-463 and DT-2007 (see the Materials and methods for primer sequences)

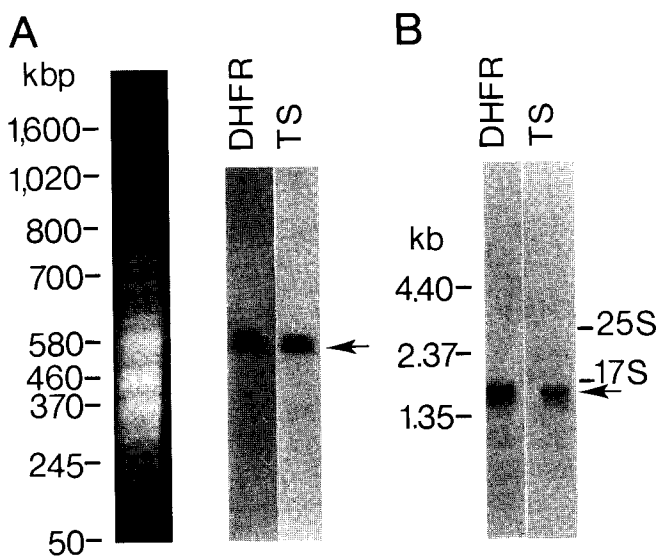


Fig. 2A, B Molecular analyses of the *P. tetraurelia* DHFR-TS gene. **A** Pulsed-field gel electrophoresis and Southern analysis of DHFR-TS. A 1% agarose clamped homogeneous electric fields (CHEF) gel of fractionated DNA from 10 000 deproteinized cells (see the Materials and methods) was stained with ethidium bromide (left), and blots were probed with DNA regions corresponding to DHFR and TS, respectively. The arrow indicates the \approx 500 kb macronuclear chromosome harboring DHFR-TS as compared to yeast chromosomal markers. **B** Northern analysis of DHFR-TS transcript. Poly(A)⁺ RNA (10 μ g) from late-log phase cells was separated on a formaldehyde/agarose gel, immobilized on nitrocellulose, and probed with DHFR- and TS-specific probes. The arrow denotes the 1.66 kb mRNA

DHFR-TS genomic DNA sequence analysis

The sequence encompassing the DHFR-TS gene is shown in Fig. 3. Based on amino acid homologies (see below), a coding region of 1386 nt is thought to start with AUG at position 121 and continue through base 1506 followed by a UGA stop codon. Upstream flanking sequences from ligation-mediated PCR and final sequencing contain no eukaryotic consensus promoter sequences (TATAA-, CAAT-, and GC-box motifs). Likewise, the GC-boxes found in mammalian (Dynam et al. 1986), *Crithidia* (Hughes et al. 1989), and *Toxoplasma* (Roos 1993) DHFR promoter regions are not found in the *Paramecium* sequence. However, the 71% A + T-rich 5' region does contain a GC-rich heptanucleotide sequence of dyad symmetry (GGGAGGG) at positions 102–108 located near the presumed initiator AUG. Although this sequence is not found in other *Paramecium* genes, it is appropriately placed to play a role in transcriptional initiation, as untranslated transcribed regions are typically very short in this organism (Preer et al. 1987; Scott et al. 1993; Yamauchi et al. 1992). The 10 mer sequence TAATTATATT, found in most cloned *Paramecium* surface antigens and proposed to be involved in transcriptional regulation (Godiska 1987), also appears in the DHFR-TS 5' region

at positions 16–25 (underlined in Fig. 3). In the DHFR-TS gene this 10 mer sequence is followed by C rather than the consensus A. No consensus translational initiation sequence was found around either AUG near the beginning of the coding region.

From comparisons of deduced protein homology, *Paramecium* genomic DHFR-TS appears to lack introns, as is common in this organism. A recent analysis of 50 introns in *P. tetraurelia* shows an extremely short intron length, when introns are present, and also a strict adherence to the use of 5' and 3' GT and AG splice sites (Russell et al. 1994). Examination of the DHFR-TS sequence revealed no GT sites located 20–30 bases upstream from an AT site, nor any loss of reading frame after a GT sequence, which argues against the possibility that introns occur in the DHFR-TS gene. Furthermore, all conserved amino acids align correctly with those of other species (Fig. 4). The only DHFR-TS sequences known to contain introns are those of *Toxoplasma gondii* (Roos 1993) and *Arabidopsis thaliana* (Lazar et al. 1993). The coding region is A + T-rich (68%), a feature common to cloned *Paramecium* genes (Preer et al. 1987; Godiska 1987).

In common with other *Paramecium* genes but unique to cloned DHFR-TS sequences are the 11 UAA and 4 UAG codons which appear as Q in Fig. 3. In the vast majority of organisms these are stop codons, but in ciliates they encode glutamine (Preer et al. 1985; Caron and Meyer 1985; Helftenbein 1985; Horowitz and Gorovsky 1985). Three such codons are found in the DHFR domain, the remainder are within TS, of which five (positions 211, 310, 350, 364, 418) are found at conserved positions in various species (Fig. 4). Other codon usage is also typical of known *Paramecium* genes (Martindale 1989) except for the prominence of GUG (Val) relative to alternative codons, and the lack of AGC (Ser) and ACG (Thr) codons, commonly found in surface antigen genes (Martindale 1989). The A + T third-base codon bias is somewhat lower in DHFR-TS (72%) than is typically found in *Paramecium* (80%; Martindale 1989; Kink et al. 1990).

Downstream A + T-rich (80%) flanking regions do not reveal a consensus eukaryotic polyadenylation signal. The transcript is thought to be polyadenylated, however, since poly(A)⁺ RNA selection was employed to detect the mRNA on a Northern blot (Fig. 2). Untranslated 3' regions of transcripts appear to extend 60–100 nt beyond the translational stop in *Paramecium* (Yamauchi et al. 1992; Godiska 1987; Kink et al. 1990). The run of T residues \approx 60 nt beyond the translational stop (positions 1564–1569) may serve as the transcriptional termination signal.

Deduced DHFR-TS protein analysis

The DHFR-TS protein sequence deduced from sequencing of genomic DNA is shown in Fig. 3. The

preceding 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 ≈ 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 ≈ 12 residue junctional peptide from positions ≈ 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 (IGyqnrLPW; 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 α -helical, β -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 (LpPCHvmsQFfV). The predicted secondary protein structure of TS is likely to contain all of the β -strands and at least six of the eight α -helices found in the *E. coli* enzyme (Perry et al. 1990); helices α_G and α_H could not be predicted, as is common in other TS sequences (Roos 1993). Ciliate DHFR-TS is expected

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 ≈ 165 and TS by residues 178–462

5'	-CTCTATTTATACATATAATATATCTTAAATTTATATGAATGCAGATAATATTAATGAATAACAAAAGTTTATAAATTTGGATATTTAAATAATAACAAGGGAGGGAATTA AAAAAGA	120
1	ATGACTACTAATAAAACATTTAGTAGTGTATTAGCAATGACTCTCAATGGAGGCATAGGGTATTAGAATAGATTGCCCTTGAAAATTAAGGAAGATCTTCAGAGATTCAAGAAAATCACA	240
	M T T N K T F S M V L A M T L N G G I G Y Q N R L P W K L K E D L Q R F K K I T	
41	ACAGGCGGGATTTGTGATAATGGGAAGAAAGACATTCGAATCGATGAACCTCTAACCAACAGAGTGAATGTAGTGATTTCAAAAATATGAAATCATCTAATGAAGTTCAAGTG	360
	T G G I V I M G R K T F E S M N S K P L P N R V N V V I S K N M K S S N E V Q V	
81	TTTCTAGAAATAGAAGATGCATTAACAATTTTACAATACATCTCAGCAAAATATATTTAATAGGAGAAAGAAATCTTTGAAGAAGGTTTGAACAATGTTAGATGTTTCAT	480
	F P R I E D A L Q F Y N T S H Q K L Y L I G G K R I F E E G L A T D K C S D V H	
121	TTGACTAGAAATAGGAGTAGAACTAAGTCCGATGTGTATTGAAACAAAATATTTTCTCTACATTTTAAAGTGAATAAGACAAGTTAAACTAAAAGTGAAGTGGATCAATATGATTAT	600
	L T R I G V E T K C D V Y L N K N I F S T F Q V N K T S Q T K S E N G I N Y D Y	
161	CAACATTTAATTAATAAGAATTTCTCAGCAACAATCATATATAGATGAAGAACATTAAGAATAAATCAATTTAGATATGATAACAAAATAATGAAGAGGGAGTCTCAAAGGATGACAGA	720
	Q H L I N K N S H E Q S Y I D E E H Q E N Q Y L D M I T K I M K E G V S K D D R	
201	ACTGGAGTAGGTACCATGTCTATTTTGGGTAAACTATGAGATTTAATTTGGCCATTCCTTCTCTTATTAAACAACAAGAAAGTATCTTTCTGGAGTTGTTGAGGAATTCCTATGG	840
	T G V G T M S I F G Q T M R F N L A Q S F P L L T T T K K V F F R G V V E E L L W	
241	TTTTAAGAGGGAATACAAATGGTAAATTTCTTTGGATAAGGGAGTTAAGATATGGGAAGGAAATGGCACCCGAGAATATTTGGATACCATTTGGTTTATAACATAGATAGGAACATGAC	960
	F L R G N T N G K L L L D K G V K I W E G N G T R E Y L D T I G L Q H R Q E H D	
281	TTAGGACCACTATGGATTCAATGGAGACATTTGGAGCTAAATATAAGGATGTTAGACTGATTATTCCAATTAAGGAGTGGATTAAAGTGAAGGAGATTATACAATTTACTGAAAAAC	1080
	L G P V Y G F Q W R H F G A K Y K D C Q T D Y S N Q G V D Q V K E I I Q L L K N	
321	AATCCAGATAGTAGAAGAATAATATATCAGCTTGGAAATCCAATGATTAAAATAAATGGCCATGGCCACCATGTCTATGATGTCATAATCTTTGGTGGCAATGGCAAAATGAGTTGC	1200
	N P D S R R I I L S A W N P I D L K Q M A L P P C H V M S Q F F V A N G K L S C	
361	ATGATGATTTAGAGATCATGTGACTTTGGTTTGGGAATTCCTTCAATATTTGCTAGTTATGATGCTCACTTACATGTTGGCAAGGAATGCAATTTGAATTTGGGAGAAATTTGTTTCAT	1320
	M M Y Q R S C D F G L G I P F N I A S Y A L L T Y M L A K E C N L N L G E F V H	
401	GTTTTAGGAGATACTCATATTTATAGTAATCATGTGGAGGCTTTAAGAAATAAATTTAGAGGGTCTCTATCTTCCCACTTCTGAAAATCAAGGGTAATAAGAGTTTGTTTGATTAC	1440
	V L G D T H I Y S N H V E A L K K Q I E R V P Y P F P L L K I K G N K S L F D Y	
441	ACTTATGAGATTTTGAATTTGGTTGGATATAATGCTCATGATAAAAATAGAAATGAAAATGGCTGTTTGAATTTTACAATTTATCTTAAAATTTTATTACAATTCAAATTTAAATTAAC	1560
	T Y E D F E L V G Y N A H D K I E M K M A V End	
	TAATTTTTTAATTTCTACTTAATCCCTTATAATGTAAAT-3' 1600	

to form a dimer in vivo, as four of the five α -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 CH₂-H₄-folate or an analogue thereof (Montfort et al. 1990); 5 of these positions are shared by the ciliate sequence.

DHFR-TS evolution

Phylogenetic relatedness of DHFR-TS sequences was deduced by parsimony analysis of TS sequences using PAUP 3.0 (Swofford 1991). DHFR phylogenies are consistent in segments which can be unambiguously aligned, although the small size of these regions precludes detailed analysis. A comparable tree was produced by distance methods (not shown). Figure 5 shows TS of *P. tetraurelia* to be most similar to *T. gondii*, a finding also substantiated by protein alignments (see Fig. 4). A further grouping of the ciliate sequence with *Plasmodium* is not surprising since common ancestry has previously been suggested for these groups (Gajadhar et al. 1991). Although these three protozoa cluster together, their branching shows great depths of divergence as protists, reflecting the current view that protozoa encompass the vast bulk of eukaryotic diversity (Sogin 1989).

Plants also appear to have bifunctional DHFR-TS, and in parsimony analysis, TS of the protozoan

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 β -strands and dashes for α -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).

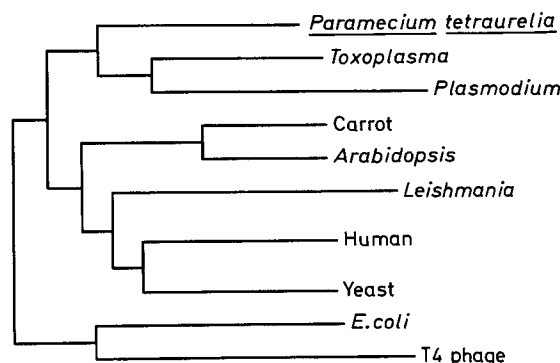


Fig. 5 TS phylogeny based on parsimony analysis of all available TS sequences followed by branch pruning to retain the species shown. Heuristic searching algorithms gave rise to the single most parsimonious tree from multiple runs under various starting conditions using midpoint rooting. Branch lengths 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.

Acknowledgements We thank Drs. Yvonne Janssen, Malcolm Lippert, and Brian 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., NHI and NSF to D.S.R.

References

- Ausubel FM, Brent R, Kingston RE, Moore DD, Seichman JG, Smith JA, Struhl K (eds.) (1989) Current protocols in molecular biology. Wiley, New York, pp 1.8.1 - 1.8.3, 3.14.2
- Bairoch A (1992) PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res* 20: 2013–2018
- Belfort M, Maley G, Pedersen-Lane J, Maley F (1983) Primary structure of the *Escherichia coli* *ThyA* gene of thymidylate synthetase. *Proc Natl Acad Sci USA* 80: 4914–4918
- Beverly SM, Ellenberger TE, Cordingley JS (1986) Primary structure of the gene encoding the bifunctional dihydrofolate reductase-thymidylate synthase of *Leishmania major*. *Proc Natl Acad Sci USA* 83: 2584–2588
- Blakley RL (1984) Dihydrofolate Reductase. In: Blakley RL, Benkovic SJ (eds.) Folates and pterines. Wiley, New York, pp 191–253
- Brooks RG, Remington JS, Luft BJ (1987) Antimicrob Agents Annu 2: 297–306
- Brown KA, Kraut J (1992) Exploring the molecular mechanism of dihydrofolate reductase. *Faraday Disc* 93: 217–224
- Bystroff C, Oatley SJ, Kraut J (1990) Crystal structures of *Escherichia coli* dihydrofolate reductase: the NADP⁺ holoenzyme and the folate-NADP⁺ ternary complex substrate binding and a model for the transition state. *Biochemistry* 29: 3263–3277
- Bzik DJ, Li W, Horii T, Inselburg J (1987) Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene. *Proc Natl Acad Sci USA* 84: 8360–8364
- Caron F (1992) A high degree of macromolecular chromosome polymorphism is generated by variable DNA rearrangements in *Paramecium primaurelia* during macronuclear differentiation. *J Mol Biol* 225: 661–678
- Caron F, Meyer E (1985) Does *Paramecium primaurelia* use a different genetic code in the macronucleus? *Nature* 314: 186–188
- Caron F, Meyer E (1989) Molecular basis of surface antigen variation in paramecia. *Annu Rev Microbiol* 43: 23–42
- Caron F, Ruiz F (1992) A method for the amplification of *Paramecium* micronuclear DNA by PCR and its application to the central repeats of *Paramecium* G surface antigen genes. *J Eukaryotic Microbiology* 39: 312–318
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162: 156–159
- Chu FK, Maley GF, Maley F, Belfort M (1984) The intervening sequence in the thymidylate synthase gene of bacteriophage T4. *Proc Natl Acad Sci USA* 81: 3049–3053
- Cowman AF, Lew AM (1989) Antifolate drug selection results in duplication and rearrangement of chromosome 7 in of *Plasmodium chabaudi* under antifolate selection. *Mol Cell Biol* 9: 5182–5188
- Davies JF, Delcamp TJ, Prendergast NJ, Ashford VA, Freisheim JH, Kraut J (1990) Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. *Biochemistry* 29: 9467–9479
- Dynan WS, Sazer S, Tijan R, Schimke RT (1986) Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. *Nature* 319: 246–248
- Ferone R, Roland S (1980) Dihydrofolate reductase: thymidylate synthase, a bifunctional polypeptide from *Crithidia fasciculata*. *Proc Natl Acad Sci USA* 77: 5802–5806
- Forney J, Epstein LM, Preer LB, Rudman BM, Widmayer DJ, Klein WH, Preer JR (1983) Structure and expression of genes for surface proteins in *Paramecium*. *Mol Cell Biol* 3: 466–474
- Freiburg M (1988) Organization and expression of the nuclear genome. In: Görtz H-D (ed.) *Paramecium*. Springer, Berlin, pp 141–145
- Gagnon S, Levesque RC, Sogin ML, Gajadhar AA (1993) Molecular cloning, complete sequence of the small subunit ribosomal RNA coding region and phylogeny of *Toxoplasma gondii*. *Mol Biochem Parasitol* 60: 145–148
- Gajadhar AA, Marquardt WC, Hall R, Gunderson J, Ariztia-Carmona EV, Sogin ML (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Teileria annulata*, and *Cryptosporidium parvum* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol Biochem Parasitol* 45: 147–154
- Garrett C, Coderre J, Meek T, Garvey E, Claman D, Beverly S, Santi D (1984) A bifunctional thymidylate synthase-dihydrofolate reductase in protozoa. *Mol Biochem Parasitol* 11: 257–265
- Godiska R (1987) Structure and sequence of the H surface protein gene of *Paramecium* and comparison with related genes. *Mol Gen Genet* 208: 529–536
- Grumont R, Washtien WL, Caput D, Santi DV (1986) Bifunctional thymidylate synthase-dihydrofolate reductase from *Leishmania tropica*: sequence homology with the corresponding monofunctional proteins. *Proc Natl Acad Sci USA* 83: 5387–5391
- Heery DM, Gannon F, Powell R (1990) A simple method for subcloning DNA fragments from gel slices. *Trends Genet* 6: 173
- Heltenbein E (1985) Nucleotide sequence of macronuclear DNA molecule coding for α -tubulin from the ciliate *Stylonichia lemnae*. Special codon usage: TAA is not a translation termination codon. *Nucleic Acids Res* 13: 415–433
- Hitchings GH, Baccanari DP (1984) In: Sirotnak FM, Burchall JJ, Ensminger WD, Montgomery JA (eds.) Folate antagonists as therapeutic agents. Academic Press, New York, pp 151–172
- Horowitz S, Gorovsky MA (1985) An unusual genetic code in nuclear genes of *Tetrahymena*. *Proc Natl Acad Sci USA* 82: 2452–2455
- Hughes DE, Shonekan OA, Simpson L (1989) Structure, genomic organization and transcription of the bifunctional dihydrofolate reductase-thymidylate synthase gene from *Crithidia fasciculata*. *Mol Biochem Parasitol* 34: 155–166
- Hughes WT (1984) Folate antagonists as antiprotozoan agents: Clinical studies. In: Sirotnak FM, Burchall JJ, Ensminger WD, Montgomery JA (eds.) Folate antagonists as therapeutic agents. Academic Press, New York, pp 251–269
- Ivanetich K, Santi D (1990) Bifunctional thymidylate synthase-dihydrofolate reductase in protozoa. *FASEB J* 4: 1591–1597
- Kapler GM, Beverly SM (1989) Transcriptional mapping of the amplified region encoding the dihydrofolate reductase-thymidylate synthase of *Leishmania major* reveals a high density of transcripts, including overlapping and antisense RNAs. *Mol Cell Biol* 9: 3959–3972
- Karrer KM (1986) The nuclear DNAs of holotrichous ciliates. In: Gall JG (ed.) The Molecular biology of ciliated protozoa. Academic Press, Orlando, FL, pp 85–110
- Kink JA, Maley ME, Preston RR, Ling K-Y, Wallen-Friedman MA, Saimi Y, Kung C (1990) Mutations in *Paramecium calmodulin* indicate functional differences between the C-terminal and N-terminal lobes *in vivo*. *Cell* 62: 165–174
- Lazar G, Zhang H, Goodman HM (1993) The origin of the bifunctional dihydrofolate reductase-thymidylate synthase isogenes of *Arabidopsis thaliana*. *Plant J* 3: 657–668

- Luo M, Piffanelli P, Tastelli L, Cella R (1993) Molecular cloning and analysis of a cDNA coding for the bifunctional dihydrofolate reductase-thymidylate synthase of *Daucus carota*. *Plant Mol Biol* 22: 427–435
- Maley GF, Maley F, Baught CM (1982) Studies on identifying the folylpolyglutamate binding sites of *Lactobacillus casei* thymidylate synthase. *Arch Biochem Biophys* 216: 551–558
- Martindale DW (1989) Codon usage in *Tetrahymena* and other ciliates. *J Protozool* 36: 29–34
- Masters JM, Attardi G (1983) The nucleotide sequence of the cDNA coding for the human dihydrofolate reductase. *Gene* 21: 59–63
- Montfort WR, Perry KM, Fauman EB, Finer-Moore JS, Maley GF, Hardey L, Maley F, Stroud RM (1990) Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding dUMP and an anti-folate. *Biochemistry* 29: 6964–6977
- Mueller PR, Wold B (1991) Ligation-mediated PCR for genomic sequencing and footprinting. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds.) *Current protocols in molecular biology*. Wiley, New York, pp 15.5.1–15.5.16
- Nelson K, Alonso G, Langer PJ, Beverly SM (1990) Sequence of the dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene in *Leishmania amazonensis*. *Nucleic Acids Res* 18: 2819
- Nicklas JA, Hunter TC, O'Neill JP, Albertini RJ (1991) Fine structure mapping of the hypoxanthine-guanine phosphoribosyl (HPRT) gene region of the human X chromosome (Xq26). *Am J Hum Genet* 49: 267–278
- Oefner C, D'Arcy A, Winkler FK (1988) Crystal structure of human dihydrofolate reductase complexed with folate. *Eur J Biochem* 174: 377–385
- Perry KM, Fauman EB, Finer-Moore JS, Montfort WR, Maley GF, Maley F, Stroud RM (1990) Plastic adaptation toward mutations in proteins: structural comparisons of thymidylate synthases. *Proteins* 8: 315–333
- Phan HL, Forney J, Blackburn EH (1989) Analysis of *Paramecium* macronuclear DNA using pulsed field gel electrophoresis. *J Protozool* 36: 402–408
- Preer JR, Preer LB, Rudman BH, Barrett AJ (1985) Deviation from the universal genetic code shown by the gene for surface protein 51A in *Paramecium*. *Nature* 14: 188–190
- Preer JR, Preer LB, Rudman B, Barnett AJ (1987) Molecular biology of the genes for immobilization antigens in *Paramecium*. *J Protozool* 34: 418–423
- Price EL, Smith PL, Klein TE, Freisheim JH (1987) Photoaffinity analogues of methotrexate as folate antagonist binding probes. 1. Photoaffinity labeling of murine L1210 dihydrofolate reductase and amino acid sequencing of the binding region. *Biochemistry* 26: 4751–4756
- Purohit S, Mathews CK (1984) Nucleotide sequence reveals overlap between T4 and phage genes encoding dihydrofolate reductase and thymidylate synthase. *J Biol Chem* 259: 6261–6266
- Roos DS (1993) Primary structure of the dihydrofolate reductase-thymidylate synthase gene from *Toxoplasma gondii*. *J Biol Chem* 268: 6269–6279
- Russell CB, Fraga D, Hinrichsen RD (1994) Extremely short 20–33 nucleotide introns are the standard length in *Paramecium tetraurelia*. *Nucleic Acids Res* 22: 1221–1225
- Sambrook J, Fritsch EF, Maniatis T (1989) In: *Molecular cloning, a laboratory manual* vol 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 1.102, 6.53–6.54, 9.38, A.1, A.3, A.6, B.15, E.2
- Santi D, Danenberg PV (1984) Folates in pyrimidine nucleotide biosynthesis. In: Blakley RL, Benkovic SJ (eds.) *Folates and pterins*, vol 1. Wiley, New York, pp 345–398
- Sartorius C, Franklin RM (1991) Hybridization arrest of cell-free translation of the malarial dihydrofolate reductase/thymidylate synthase mRNA by anti-sense oligodeoxyribonucleotides. *Nucleic Acids Res* 19: 1613–1618
- Sasner JM, Van Houten JL (1989) Studies of the *Paramecium* folate chemoreceptor. *Chem Senses* 14: 587–595
- Scott J, Leec C, Forney J (1993) Molecular and genetic analysis of the B type surface protein gene from *Paramecium tetraurelia*. *Genetics* 133: 189–198
- Shull S, Heintz NH, Periasamy M, Manohar M, Janssen YMW, Marsh JP, Mossman BT (1991) Differential regulation of anti-oxidant enzymes in response to oxidants. *J Biol Chem* 266: 24398–24403
- Smith DR, Calvo JM (1980) Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. *Nucleic Acids Res* 8: 2255–2274
- Sogin ML (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. *Am Zool* 29: 487–499
- Swofford DL (1991) PAUP: phylogenetic analysis using parsimony, version 3.0. Computer program distributed by the Illinois Natural History Survey, Champaign, IL
- Takeishi K, Kaneda S, Ayusawa D, Shimizu K, Gotoh O, Seno T (1985) Nucleotide sequence of a functional cDNA for human thymidylate synthase. *Nucleic Acids Res* 13: 2035–2043
- Volz KW, Matthews DA, Alden RA, Freer ST, Hansch C, Kaufman BT, Kraut J (1982) Crystal structure of avian dihydrofolate reductase containing phenyltriazine and NADPH. *J Biol Chem* 257: 2528–2536
- Washtien WL, Grumont R, Santi DV (1985) DNA amplification in antifolate resistant *Leishmania*. *Proc Natl Acad Sci USA* 60: 7809–7812
- Yamauchi K, Ochiai T, Usuki I (1992) The unique structure of the *Paramecium caudatum* hemoglobin gene: the presence of one intron in the middle of the coding region. *Biochim Biophys Acta* 1171: 81–87