12-1987

Molecular Cloning and Sequence Analysis of the Plasmodium Falciparum Dihydrofolate Reductase-thymidylate Synthase gene.

David J. Bzik
Dartmouth College

Wu-Bo Li
Dartmouth College

Toshihiro Horii
Osaka University

Joseph Inselburg
Dartmouth College

Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 8360–8364, December 1987
Biochemistry

Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase–thymidylate synthase gene

(genomic DNA cloning/gene structure/homology/evolution/malaria parasite)

DAVID J. BZIK, WU-BO LI, TOSHIHIRO HORII*, AND JOSEPH INSELBURG†

Department of Microbiology, Dartmouth Medical School, Hanover, NH 03756

Communicated by Walter H. Stockmayer, August 17, 1987 (received for review July 17, 1987)

ABSTRACT **Genomic DNA clones that coded for the bifunctional dihydrofolate reductase (DHFR) and thymidylate synthase (TS) (DHFR-TS) activities from a pyrimethamine-sensitive strain of *Plasmodium falciparum* were isolated and sequenced. The deduced DHFR-TS protein contained 608 amino acids (71,682 Da). The coding region for DHFR-TS contained no intervening sequences and had a high A+T content (75%). The DHFR domain, in the amino-terminal portion of the protein, was joined by a 94-amino acid junction sequence to the TS domain in the carboxyl-terminal portion of the protein. The TS domain was more conserved than the DHFR domain and both *P. falciparum* domains were more homologous to eukaryotic than to prokaryotic forms of the enzymes. Predicted secondary structures of the DHFR and TS domains were nearly identical to the structures identified in other DHFR and TS enzymes.**

The effective treatment of *Plasmodium falciparum* malaria with the antifolate pyrimethamine, a specific inhibitor of dihydrofolate reductase (DHFR; 5,6,7,8-tetrahydrofolate: $NADP^+$ oxidoreductase, EC 1.5.1.3) (1), is being compromised by pyrimethamine-resistant (Pyr^r) parasites. Pyr^r strains usually contain normal levels of a mutant DHFR with a reduced affinity for the drug (2–7). The DHFR and thymidylate synthase (TS; 5,10-methylenetetrahydrofolate: dUMP *C*-methyltransferase, EC 2.1.1.45) activities in protozoa often reside on a bifunctional homodimeric polypeptide (DHFR-TS) (8, 9). DHFR-TS catalyzes the sole biosynthetic pathway of dTMP and is essential both for DNA synthesis and for maintaining levels of tetrahydrofolate in *P. falciparum* (10).

We describe here the cloning and sequencing of genomic DNA that encodes the DHFR-TS enzyme of the pyrimethamine-sensitive *P. falciparum* strain FCR3. We compare the primary and secondary structures of the DHFR-TS with known DHFR and TS enzymes. We also consider the opportunities the cloned gene may provide for further study.‡

MATERIALS AND METHODS

Parasite DNA. *P. falciparum* strain FCR3 (7) was the source of parasite DNA. Genomic DNA was purified from *in vitro* cultured trophozoite and schizont stage parasites (7).

Oligonucleotides. A 29-base oligonucleotide mixture 5′ d[GG(A or T)(G or C)T(A or T)CC(A or T)TTTAATAT(A or T)GC(A or T)(A or T)(G or C)(A or T)TATGC] 3′, was synthesized. The oligonucleotides corresponded to a TS consensus sequence, Gly-(Val or Leu)-Pro-Phe-Asn-Ile-Ala-Ser-Tyr-Ala (11). Radioactively end-labeled oligonucleotides (12) hybridized with a 3.5-kilobase (kb) *Eco*RI fragment on Southern blots (13) of restriction enzyme-digested *P. falciparum* DNA (data not shown).

Construction and Screening of Genomic DNA Libraries. Genomic DNA libraries were constructed in λgt11 (14) as an *Eco*RI library and in λZAP (Stratagene Cloning Systems, San Diego, CA) as an *Xba* I library. Bacteriophage were grown in *Escherichia coli* strains Y1088 (15) or XL-1 blue (Stratagene). Recombinant phage DNA was transferred to nitrocellulose as described (12). Genomic DNA fragments were labeled by the oligo-labeling method (16). Hybridization conditions were as described (7). Southern blotting was performed using Zeta probe membranes (Bio-Rad) (17).

Subcloning and DNA Sequencing. *Eco*RI fragments from λgt11 recombinants were subcloned into plasmid pUC18 (18). λZAP recombinants were made into plasmid recombinants by automatic excision (Stratagene). DNA fragments were sequenced using dideoxynucleotide technology (19). Briefly, DNA fragments were purified (20), self-ligated, and sonicated (21). The ends of 0.3- to 0.7-kb DNA fragments were enzymatically repaired (blunted) and cloned into *Sma* I-digested, alkaline phosphatase-treated M13mp8 (22). The DHFR-TS sequence reported here was determined from over 40 kb of randomly derived sequence data. Every base pair in the DHFR-TS coding region was sequenced ≈10 times (average), and both DNA strands were completely sequenced. DNA sequences were reconstructed using the DNA Inspector II programs (23) (a gift of Textco, West Lebanon, NH). The BIONET computer resource for molecular biology (IntelliGenetics, Palo Alto, CA) was used to manipulate DNA sequences and to predict protein secondary structure (24).

RESULTS

Genomic DNA Cloning. Four λgt11 recombinant phage were isolated from the *P. falciparum* genomic *Eco*RI library that was screened with the oligonucleotide probe. These phage each contained a single 3.5-kb *Eco*RI fragment of *P. falciparum* DNA. One of them, λgt11-TS4a, was sequenced (see below). A sequence beginning at the 5′ left end of that *Eco*RI fragment encoded 138 amino acids that corresponded to sequences in other TS genes (Fig. 1 *c* and *d*) (11). The TS4a clone, a 267-base-pair *Eco*RI/*Dra* I fragment contained within the TS4a clone, the clones DHFR X7, DHFR E1, and DHFR E5 (see below) were used to establish the restriction map of genomic DNA (Fig. 1 *a* and *b*) in Southern blotting experiments (data not shown). We were unable to isolate the entire DHFR-TS gene on a single 16-kb *Bcl* I DNA fragment

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: TS, thymidylate synthase; DHFR, dihydrofolate reductase; Pyr^r, pyrimethamine resistant.

*Present address: Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 564, Japan.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03028).

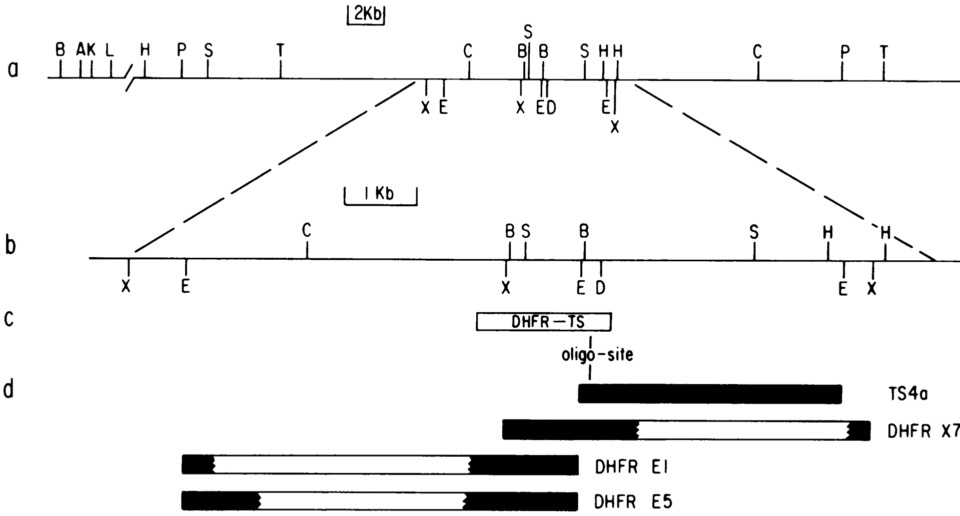

FIG. 1. The restriction map and genomic DNA clones of the DHFR-TS gene. (*a*) Restriction sites shown are A, *Bam*HI; C, *Bcl* I; B, *Bgl* II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn* I; P, *Pst* I; L, *Sal* I; S, *Spe* I; T, *Sst* I; X, *Xba* I. The *Dra* I site (D) was determined by DNA sequencing. (*b*) Enlarged restriction map encompassing DHFR-TS. (*c*) The large open reading frame encoding the DHFR-TS is indicated. The site that corresponds to the binding site for the oligonucleotides is indicated. (*d*) Genomic DNA clones used in the determination of the DHFR-TS nucleotide sequence. Solid portions indicate *P. falciparum* DNA, and open portions indicate deletions of *P. falciparum* DNA (see text).

(Fig. 1*a*) because of the instability of those clones in several genetic backgrounds; we were also unable to isolate DHFR-TS clones from cDNA libraries prepared from trophozoite/schizont stage mRNA.

```
                                                1                   10                  20
ATATATATATTTATTTATTTATTTATATATTTATATTTCTCCTTTTTATGATGGAACAAGTCTGCGACGTTTTCGATATTTATGCCATATGTGCATGTTGTAAGGTTGAAAGCAAAAAT 120
                                                 MetMetGluGlnValCysAspValPheAspIleTyrAlaIleCysAlaCysCysLysValGluSerLysAsn
          30                  40                  50                  60
GAGGGGAAAAAAAAATGAGGTTTTTAATAACTACACATTTAGAGGTCTAGGAAATAAAGGAGTATTACCATGGAAATGTAATTCCCTAGATATGAAATATTTTTGTGCAGTTACAACATAT 240
GluGlyLysLysAsnGluValPheAsnAsnTyrThrPheArgGlyLeuGlyAsnLysGlyValLeuProTrpLysCysAsnSerLeuAspMetLysTyrPheCysAlaValThrThrTyr
          70                  80                  90                  100
GTGAATGAATCAAAATATGAAAAATTGAAATATAAGAGATGTAAATATTTAAACAAAGAAACTGTGGATAATGTAAATGATATGCCTAATTCTAAAAAATTACAAAATGTTGTAGTTATG 360
ValAsnGluSerLysTyrGluLysLeuLysTyrLysArgCysLysTyrLeuAsnLysGluThrValAspAsnValAsnAspMetProAsnSerLysLysLeuGlnAsnValValValMet
          110                 120                 130                 140
GGAAGAACAAACTGGGAAAGCATTCCAAAAAAATTTAAACCTTTAAGCAATAGGATAAATGTTATATTGTCTAGAACCTTAAAAAAAGAAGATTTTGATGAAGATGTTTATATCATTAAC 480
GlyArgThrAsnTrpGluSerIleProLysLysPheLysProLeuSerAsnArgIleAsnValIleLeuSerArgThrLeuLysLysGluAspPheAspGluAspValTyrIleIleAsn
          150                 160                 170                 180
AAAGTTGAAGATCTAATAGTTTTACTTGGGAAATTAAATTACTATAAATGTTTTATTATAGGAGGTTCCGTTGTTTATCAAGAATTTTTAGAAAAGAAATTAATAAAAAAAATATATTTT 600
LysValGluAspLeuIleValLeuLeuGlyLysLeuAsnTyrTyrLysCysPheIleIleGlyGlySerValValTyrGlnGluPheLeuGluLysLysLeuIleLysLysIleTyrPhe
          190                 200                 210                 220
ACTAGAATAAATAGTACATATGAATGTGATGTATTTTTTCCAGAAATAAATGAAAATGAGTATCAAATTATTTCTGTTAGCGATGTATATACTAGTAACAATACAACATTGGATTTTATC 720
ThrArgIleAsnSerThrTyrGluCysAspValPhePheProGluIleAsnGluAsnGluTyrGlnIleIleSerValSerAspValTyrThrSerAsnAsnThrThrLeuAspPheIle
          230                 240                 250                 260
ATTTATAAGAAAACGAATAATAAAATGTTAAATGAACAAAATTGTATAAAAGGAGAAGAAAAAAATAATGATATGCCTTTAAAGAATGATGACAAAGATACATGTCATATGAAAAAATTA 840
IleTyrLysLysThrAsnAsnLysMetLeuAsnGluGlnAsnCysIleLysGlyGluGluLysAsnAsnAspMetProLeuLysAsnAspAspLysAspThrCysHisMetLysLysLeu
          270                 280                 290                 300
ACAGAATTTTACAAAAATGTAGACAAATATAAAATTAATTATGAAAATGATGATGATGATGAAGAAGAAGATGATTTTGTTTATTTTAATTTTAATAAAGAAAAAGAAGAGAAAAATAAA 960
ThrGluPheTyrLysAsnValAspLysTyrLysIleAsnTyrGluAsnAspAspAspAspGluGluGluAspAspPheValTyrPheAsnPheAsnLysGluLysGluGluLysAsnLys
          310                 320                 330                 340
AATTCTATACATCCAAATGATTTTCAAATATATAATAGCTTGAAATATAAATATCATCCTGAATACCAATATTTAAATATTATTTATGATATTATGATGAATGGAAATAAACAAAGTGAT 1080
AsnSerIleHisProAsnAspPheGlnIleTyrAsnSerLeuLysTyrLysTyrHisProGluTyrGlnTyrLeuAsnIleIleTyrAspIleMetMetAsnGlyAsnLysGlnSerAsp
          350                 360                 370                 380
CGAACGGGAGTAGGTGTTTTAAGTAAATTCGGATATATTATGAAATTTGATTTAAGTCAATATTTCCCATTATTAACTACGAAGAAATTATTTTTAAGAGGAATTATTGAAGAATTGCTT 1200
ArgThrGlyValGlyValLeuSerLysPheGlyTyrIleMetLysPheAspLeuSerGlnTyrPheProLeu[LeuThrThrLysLysLeuPheLeuArgGlyIleIleGluGluLeu]Leu
          390                 400                 410                 420
TGGTTTATTAGAGGAGAAACAAATGGTAATACGTTGTTAAATAAGAATGTAAGGATATGGGAAGCTAATGGTACTAGGGAATTTTTAGATAATAGAAAATTATTTCATAGAGAAGTTAAC 1320
TrpPheIleArgGlyGluThrAsnGlyAsnThrLeuLeuAsnLysAsnValArgIleTrpGluAlaAsnGlyThrArgGluPheLeuAspAsnArgLysLeuPheHisArgGluValAsn
          430                 440                 450                 460
GATTTAGGACCTATTTATGGTTTTCAATGGAGACATTTCGGTGCTGAATATACAAATATGTATGATAATTATGAAAATAAAGGAGTGGATCAATTAAAAAATATAATAAATTTAATTAAA 1440
AspLeuGlyProIleTyrGlyPheGlnTrpArgHisPheGlyAlaGluTyrThrAsnMetTyrAspAsnTyrGluAsnLysGlyValAspGlnLeuLysAsnIleIleAsnLeuIleLys
          470                 480                 490                 500
AATGATCCTACAAGTAGAAGAATTCTTTTGTGTGCATGGAATGTAAAAGATCTTGACCAAATGGCATTACCTCCTTGTCATATTTTATGTCAGTTTTATGTTTTCGATGGAAATTATCA 1560
AsnAspProThrSerArgArgIleLeuLeuCysAlaTrpAsnValLysAspLeuAspGlnMetAlaLeuProPro[Cys]HisIleLeuCysGlnPheTyrValPheAspGlyLysLeuSer
          510                 520                 530                 540
TGTATTATGTATCAAAGATCATGTGATTTAGGGCTAGGAGTACCTTTTAATATTGCTTCTTATTCTATTTTTACTCATATGATTGCACAAGTCTGTAATTTGCAACCTGCGCAGTTCATA 1680
CysIleMetTyrGlnArgSerCysAspLeuGlyLeuGlyValProPheAsnIleAlaSerTyrSerIlePheThrHisMetIleAlaGlnValCysAsnLeuGlnProAlaGlnPheIle
          550                 560                 570                 580
CACGTTTTAGGAAATGCACATGTTTATAATAATCACATTGATAGTTTAAAAATTCAACTTAACAGAATACCCTATCCATTCCCAACACTTAAATTAAATCCAGATATTAAAAATATTGAA 1800
HisValLeuGlyAsnAlaHisValTyrAsnAsnHisIleAspSerLeuLysIleGlnLeuAsnArgIleProTyrProPheProThrLeuLysLeuAsnProAspIleLysAsnIleGlu
          590                 600
GATTTACAATTTCGGATTTTACAATACAAAATTATGTTCATCATGAAAAAATTTCAATGGATATGGCTGCTTAATATTGAAATTAAAAAAATATATGAACAAATGATGACAAAATTACC 1920
AspPheThrIleSerAspPheThrIleGlnAsnTyrValHisHisGluLysIleSerMetAspMetAlaAla***

TGAACGGTTAAGAATTTTTTTTTTTCTTTTTTGTATACACATAAATATATATAATGTAAAATATATATATATATATATTAAGACTATATGGCTATTTTTTTTTTTTTTTTTTTTTTTTTAC 2040

CTGAACATGAAATAATAAATATTTAATATTTACAAAATGTACAGATTATTTATTGTTGTGTATATATAATAGTAAGATATTCTAATTTTTGAAATAATAATTTTTTATTATATTAATTAT 2160
```

FIG. 2. The DNA sequence encoding the *P. falciparum* DHFR-TS. The predicted amino acid sequence of DHFR-TS is shown below the codons. The nucleotide sequence shown includes 48 bases of 5′ noncoding DNA and 285 bases of 3′ noncoding DNA. The first amino acid residue, Met, is specified by nucleotide numbers 50–52. Nucleotides are numbered in the right column and amino acids are numbered above the codon and amino acid residue. Boxed amino acid residues 369–383 correspond to the proposed folypolyglutamate binding site (see text) and the boxed Cys amino acid residue 490 corresponds to the probable nucleophilic catalyst in the conversion of dUMP to dTMP (see text). The underlined nucleotides correspond to the hybridization site of the oligonucleotides.

We therefore screened a *P. falciparum Xba* I library to isolate the genomic 4.9-kb *Xba* I fragment (Fig. 1 *a* and *b*). A clone (DHFR X7; Fig. 1*d*) that contained a 2.1-kb *Xba* I fragment was isolated, and the DHFR X7 fragment was sequenced (see below). It contained an additional DHFR-TS coding sequence and a 2.8-kb deletion (Fig. 1*d*). To clone the 5′ portion of the coding sequence of the gene, we screened the λgt11 *Eco*RI library with the 1.0-kb *Bgl* II fragment from the DHFR X7 clone so as to isolate a recombinant that contained a 5.4-kb genomic *Eco*RI fragment (Fig. 1 *a* and *b*). Two phage that were isolated and purified, designated DHFR E1 and DHFR E5 (Fig. 1*d*), respectively, had 1.8-kb and 2.4-kb insertions. They had deletions of 3.6 kb and 3.0 kb, which were, respectively, between $(TTTA)_n$ repeat sequences and $(AT)_n$ repeat sequences. Difficulties in cloning other *P. falciparum* genes may be related to the presence of similar repeat sequences. All the cloned fragments containing DHFR-TS coding sequences are shown (Fig. 1*d*).

DNA Sequence Analysis and Sequence Properties. We sequenced the following DNA fragments (Fig. 1): the 3.5-kb *Eco*RI fragment of λgt11 TS4a; the 5′ 1.0-kb *Xba* I/*Bgl* II fragment of DHFR X7; the 5′ 0.8-kb *Eco*RI/*Bgl* II fragment of DHFR E1; and the 5′ 1.4-kb *Eco*RI/*Bgl* II fragment of DHFR E5. The DNA sequence of the DHFR-TS gene is

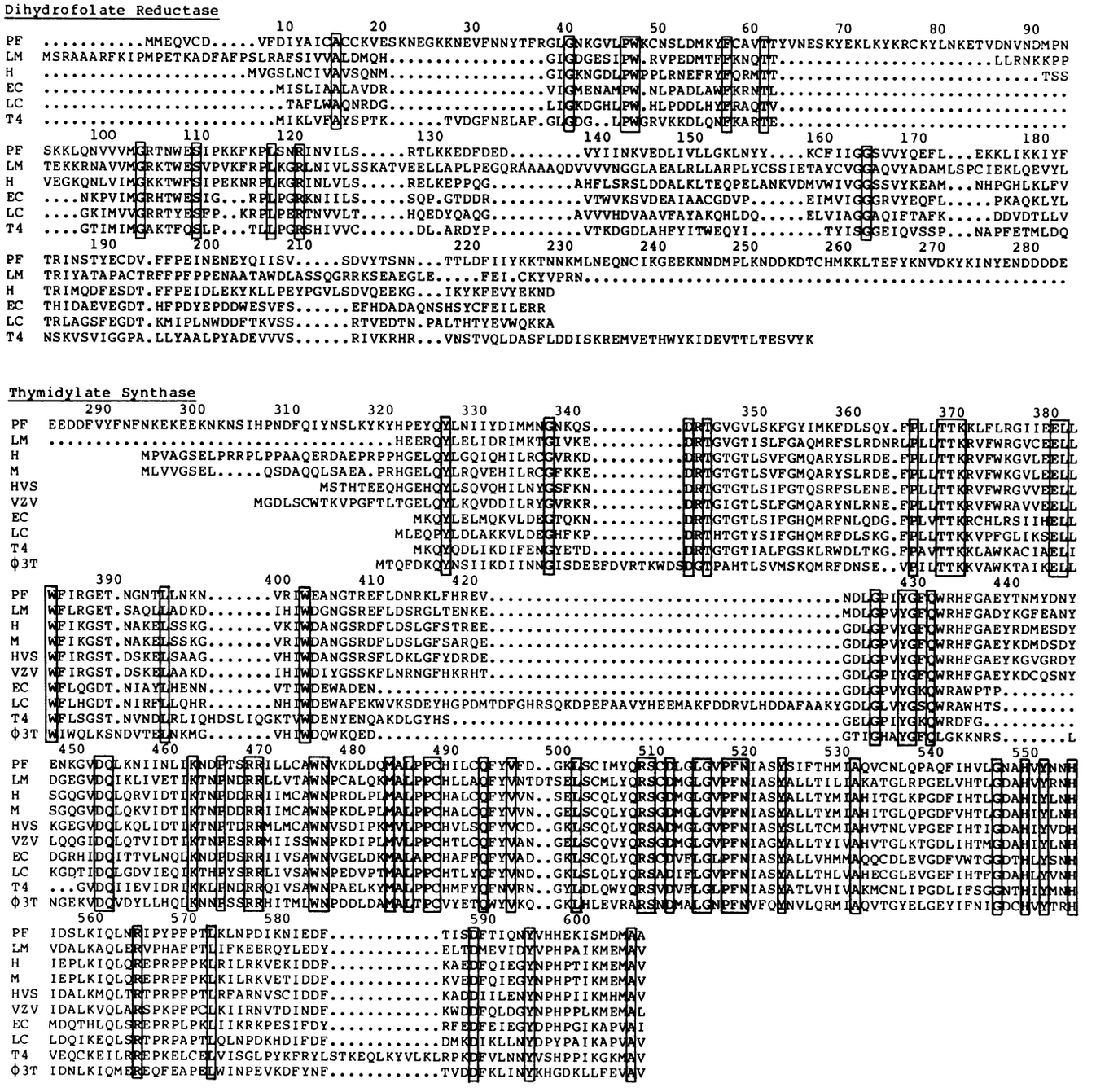

FIG. 3. Comparison of the primary structure of *P. falciparum* DHFR-TS with enzymes from other sources. The enzymes were DHFR (amino acids 1–228), junctional sequence (amino acids 229–322), and TS (amino acids 323–608). EC, *Escherichia coli* (27, 28); HVS, *Herpesvirus saimiri* (29); H, *Homo sapiens* (human) (30–32); LC, *Lactobacillus casei* (33, 34); LM, *Leishmania major* (35); M, mouse (36); T4, phage T4 (37, 38); *ϕ*3T, phage *ϕ*3T (39); PF, *P. falciparum*; VZV, varicella zoster virus (40). The sequence of the *Leishmania tropica* DHFR-TS (11) is nearly identical to the *L. major* sequence and was not used in the alignment. Boxes show locations of invariant amino acids. Amino acids are identified by the single-letter code.

shown (Fig. 2). The sequence begins in the $(AT)_n$ ladder in the 5′ noncoding region of the gene. The location to which the oligonucleotides hybridized is underlined. A single long open reading frame began at the AUG codon at nucleotide 49 and ended at the UAA stop codon at nucleotides 1873–1875. No intervening sequences were identified. The assigned reading frame was based on homology of the predicted amino acid sequence to other known DHFR-TS gene sequences (see below). The primary DHFR-TS protein contained 608 amino acids and had a calculated molecular mass of 71,682 Da, which corresponded with the measured mass of 70 kDa (7) and was higher than the 63 kDa previously reported (9).

The coding sequence for the DHFR-TS gene resembled all known *P. falciparum* coding regions (25) in that (*i*) it had a lower A+T content (75%) than the flanking sequences [94% (5′) and 87% (3′)]; (*ii*) the mRNA sense strand had a higher A (41%) than T (34%) content; (*iii*) the A+T content of the first three coding positions increased in that order (65%, 76%, and 84%); (*iv*) the *S* value (26), 16.3, a measure of codon preference, was nearly identical to the average value of 16.2 for *P. falciparum* coding regions (25).

Protein Homology. The alignment of the protein sequence of *P. falciparum* DHFR-TS was based on the locations of invariant amino acids and the structural equivalence of the proteins in those enzymes from other organisms (Fig. 3). While all invariant amino acids of DHFR and TS are indicated (Fig.3), the many conservative amino acid changes were not indicated. The *P. falciparum* protein contained three structural domains: an amino-terminal DHFR domain (≈27 kDa), a junctional domain (≈11 kDa), and a carboxyl-terminal TS domain (≈34 kDa).

DHFR Homology. We defined the DHFR domain to include amino acids 1–228 based on both amino acid homology and predictions of secondary structure of the DHFR enzymes (see below). The *P. falciparum* DHFR domain had little amino acid homology to other forms of DHFR in amino acids 1–10 and 201–228, while amino acids 11–200 had significant homology except for several amino acid insertions (Fig. 3).

Eleven amino acids were invariant between all forms and 35 other amino acids were conserved between four or more of the different forms of DHFR (Fig. 3). Many of these conserved amino acids are thought to participate in substrate binding (41). The *P. falciparum* DHFR domain was most homologous to the two eukaryotic ["vertebrate" (11, 41)] forms of DHFR, sharing 61 amino acids with the human DHFR and 56 amino acids with *L. major* (Table 1). Eukaryotic DHFR could be distinguished from prokaryotic (and prokaryotic viral) DHFR by amino acid insertions that form loops connecting elements of secondary structure (11, 41). On the basis of insertions and amino acid homology, the *P. falciparum* enzyme resembled the eukaryotic form of DHFR (Fig. 3; Table 1).

Secondary structures in other DHFRs were also conserved in the *P. falciparum* enzyme (41). The conserved eight prominent β-strands (designated βA–βH) found in all forms of DHFR except *L. tropica* (11) were predictable in the *P. falciparum* DHFR domain (amino acids 1–228). The β-strands βA–βF, involved in substrate binding, were identified in the first 188 amino acids. The four α-helices found in other forms of DHFR were predictable in *P. falciparum*. The eukaryotic-specific αE′ helix, however, was not predictable in the *P. falciparum* DHFR.

TS Homology. We defined the TS domain to include amino acids 323–608 based on amino acid homology and secondary structure of other TS enzymes (see below). The *P. falciparum* TS domain was very homologous to eukaryotic forms of the TS enzyme based on amino acid homology (Fig. 3; Table 2). The two protozoan enzymes shared fewer amino acids with each other than either enzyme did with the human TS (Table 2). While the two protozoan enzymes shared 153 amino acids, these were encoded by 75% A+T-rich DNA in *P. falciparum* compared with 38% A+T-rich DNA for *L. major* (35).

Fifty-one invariant amino acids occurred between the 10 characterized forms of TS. At 137 locations, seven or more amino acids out of 10 were identical. Between the 10 forms of TS, there was ≈68% amino acid homology if amino acid insertions were disregarded. Many of the amino acid changes were conservative and revealed a remarkable overall chemical homology. Conservation of structural elements throughout the TS domain was therefore implied.

Eight sites of amino acid insertions in all TS occurred outside of predicted structural elements and represented loops in domains that probably do not influence the core structure (42). The amino acid insertions in the *P. falciparum* enzyme were identical in size in all eukaryotic forms except for *L. major*. The *P. falciparum* enzyme was most homologous to the eukaryotic enzymes based on both amino acid homology and insertions (Fig. 3; Table 2).

The *P. falciparum* TS domain contained essentially all of the elements found in the three-dimensional structure of the *L. casei* TS enzyme (42). The five β-strands (βi–βv), which formed the dimer interface in the *L. casei* TS, were present and 10 of the 11 α-helices were predicted. The corresponding α-C helix in *P. falciparum* was not predicted and had a weak α-helical nature due to the presence of several Asn and Gly residues (amino acids 392, 393, 394, 398, and 400).

Amino acids 370–372 were invariant and formed part of the sequence identified as binding folate cofactors (Fig. 3) (43). The Lys (amino acid 372) and Lys or Arg (amino acid 373) are thought to play a role in binding folypolyglutamate (42, 43). The three-dimensional structure of TS from *L. casei* led to the assignment of amino acids that form the active site and the elements of structure thought to be important in the activities of the enzyme (42). The Cys at amino acid 490 (Fig. 3) is the probable binding site for FdUMP and dUMP (27, 37, 42, 44) and has been proposed as the essential nucleophilic catalyst

Table 1. Numbers of amino acids shared by different DHFRs

	PF	LM	H	EC	LC	T4
PF	228	56	61	47	38	24
LM		226*	59	48	40	24
H			186	47	42	29
EC				159	43	34
LC					162	32
T4						193

The alignments used to calculate shared amino acids are shown in Fig. 3. Organisms (and abbreviations) used for comparison are shown in Fig. 3.
*Includes DHFR and junctional amino acids.

Table 2. Numbers of amino acids shared by different TSs

	PF	LM	H	M	HVS	VZV	EC	LC	T4	φ3T
PF	286*	153	161	161	160	142	115	127	114	102
LM		289*	174	174	172	161	124	133	116	97
H			313	279	207	204	141	139	128	104
M				307	205	205	142	139	126	105
HVS					294	192	130	148	113	107
VZV						301	124	138	120	101
EC							264	156	126	95
LC								316	122	104
T4									286	90
φ3T										279

The alignments used to calculate shared amino acids are shown in Fig. 3. Organisms (and abbreviations) used for comparison are shown in Fig. 3.
*Based on the TS domain starting at *P. falciparum* amino acid 323 in Fig. 3.

in conversion of dUMP to dTMP (11, 42). That Cys is embedded within a highly conserved sequence Met-Ala-(Leu or Val)-Leu-(Pro or Ala or Thr)-Pro-Cys-(His or Val).

Junctional Sequence. We defined amino acids 229–322 as a junctional sequence joining the DHFR and TS domains. They had essentially no homology to DHFR or TS enzymes of other organisms (Fig. 3). The A+T content of this sequence (81%) is higher than the DHFR domain (75%) and TS domain (73%). This sequence contained 41 charged amino acids and 33 hydrophilic amino acids. A secondary structure analysis predicted a random coil structure for most of this sequence. The junctional sequence resembled that of *L. major* with respect to having many charged and hydrophilic amino acids and random coil structure, although the *L. major* junction (37 amino acids) was smaller (35).

DISCUSSION

The successful cloning and sequence analysis of the *P. falciparum* DHFR-TS gene now permits a careful examination of pyrimethamine resistance. One Pyr^r mutant (5, 7) had a gene duplication of DHFR-TS sequence and a disproportionate overproduction of the DHFR-TS protein (7). Study of that gene may provide insight into regulation of the enzyme's synthesis. A known mutant DHFR [FCR3-D8 (5, 7)] that confers a high level of parasite drug resistance could be used as a selectable genetic marker for establishing a parasite gene transformation system. Such a system would be invaluable to further analyze other forms of drug resistance and to study parasite antigens that play a role in inducing natural parasite immunity.

We thank the Dartmouth Molecular Genetics Center for synthesizing the oligonucleotide mixture used in this work. This work was supported by National Institutes of Health Grant AI 20437 to J.I.

1. Ferone, R, Burchall, J. J. & Hitchings, G. H. (1969) *Mol. Pharmacol.* **5,** 49–59.
2. Diggers, S. M., Gutteridge, W. E. & Trigg, P. I. (1970) *Nature (London)* **228,** 579–580.
3. Sirawaraporn, W. & Yuthavong, Y. (1984) *Mol. Biochem. Parasitol.* **10,** 355–367.
4. McCutchan, T. R., Welsh, J. A., Dame, J. B., Quakyi, I. A., Graves, P. M., Drake, J. C. & Allegra, C. J. (1984) *Antimicrob. Agents Chemother.* **26,** 656–659.
5. Banyal, H. S. & Inselburg, J. (1986) *Exp. Parasitol.* **62,** 61–70.
6. Walter, R. D. (1986) *Mol. Biochem. Parasitol.* **19,** 61–66.
7. Inselburg, J., Bzik, D. J. & Horii, T. (1987) *Mol. Biochem. Parasitol.*, in press.
8. Ferone, R. & Roland, S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5802–5806.
9. Garrett, C. E., Coderre, J. A., Meek, T. D., Garvey, E. P., Claman, D. M., Beverley, S. M. & Santi, D. V. (1984) *Mol. Biochem. Parasitol.* **11,** 257–265.
10. Sherman, I. (1979) *Microbiol. Rev.* **43,** 453–495.
11. Grumont, R., Washtien, W. L., Caput, D. & Santi, D. V. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5387–5391.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
13. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
14. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1194–1198.
15. Young, R. A. & Davis, R. W. (1983) *Science* **222,** 778–782.
16. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
17. Reed, K. C. & Mann, D. A. (1986) *Nucleic Acids Res.* **13,** 7207–7221.
18. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
20. Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112,** 295–298.
21. Deininger, P. L. (1983) *Biochemistry* **129,** 216–223.
22. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
23. Gross, R. H. (1986) *Nucleic Acids Res.* **14,** 591–597.
24. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276.
25. Weber, J. L. (1987) *Gene* **52,** 103–109.
26. Ikemura, T. (1985) *Mol. Biol. Evol.* **2,** 13–34.
27. Belfort, M., Maley, G., Pederson-Lane, J. & Maley, F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4914–4918.
28. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8,** 2255–2274.
29. Honess, R. W., Bodemer, W., Cameron, K. R., Niller, H.-H., Fleckenstein, B. & Randall, R. E. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3604–3608.
30. Masters, J. N. & Attardi, G. (1983) *Gene* **21,** 59–63.
31. Chen, M.-J., Shimada, T., Moulton, A. D., Cline, A., Humphries, R. K., Maizel, J. & Nienhuis, A. W. (1984) *J. Biol. Chem.* **259,** 3933–3943.
32. Takeishi, K., Kaneda, S., Ayusawa, D., Shimizu, K., Gotoh, O. & Seno, T. (1985) *Nucleic Acids Res.* **13,** 2035–2043.
33. Freisheim, J. H., Bitar, K. G., Reddy, A. V. & Blankenship, D. T. (1978) *J. Biol. Chem.* **253,** 6437–6444.
34. Maley, G. F., Bellisario, R. L., Guarino, D. V. & Maley, F. (1979) *J. Biol. Chem.* **254,** 1301–1304.
35. Beverley, S. M., Ellenberger, T. E. & Cordingley, J. S. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 2584–2588.
36. Perryman, S. M., Rossana, C., Deng, T., Vanin, E. F. & Johnson, L. F. (1986) *Mol. Biol. Evol.* **3,** 313–321.
37. Chu, F. K., Maley, G. F., Maley, F. & Belfort, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3049–3053.
38. Purohit, S. & Mathews, C. K. (1984) *J. Biol. Chem.* **259,** 6261–6266.
39. Kenny, E., Atkinson, T. & Hartley, B. S. (1985) *Gene* **34,** 335–342.
40. Thompson, R., Honess, R. W., Taylor, L., Morran, J. & Davison, A. J. (1987) *J. Gen. Virol.* **68,** 1449–1455.
41. Blakley, R. L. (1984) in *Folates and Pteridines*, eds. Blakley, R. L. & Benkovic, S. J. (Wiley, New York), Vol. 1, pp. 191–253.
42. Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O., Santi, D. V. & Stroud, R. M. (1987) *Science* **228,** 448–455.
43. Maley, G. F., Maley, F. & Baugh, C. M. (1982) *Arch. Biochem. Biophys.* **216,** 551–558.
44. Bellisario, R. L., Maley, G. F., Galivan, J. H. & Maley, F. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 1848–1852.