

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

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>

 Part of the [Medical Biochemistry Commons](#), [Medical Genetics Commons](#), and the [Medical Microbiology Commons](#)

Recommended Citation

Bzik, David J.; Li, Wu-Bo; Horii, Toshihiro; and Inselburg, Joseph, "Molecular Cloning and Sequence Analysis of the Plasmodium Falciparum Dihydrofolate Reductase-thymidylate Synthase gene." (1987). *Open Dartmouth: Faculty Open Access Articles*. 1205.
<https://digitalcommons.dartmouth.edu/facoa/1205>

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) EcoRI fragment on

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.

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 EcoRI 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. EcoRI 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 EcoRI library that was screened with the oligonucleotide probe. These phage each contained a single 3.5-kb EcoRI fragment of *P. falciparum* DNA. One of them, λgt11-TS4a, was sequenced (see below). A sequence beginning at the 5' left end of that EcoRI 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 EcoRI/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

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).

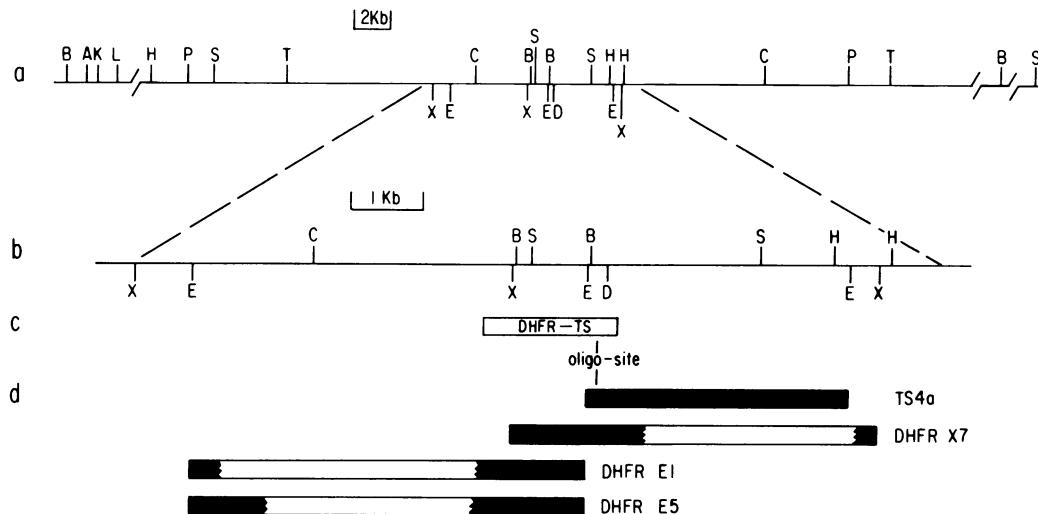


FIG. 1. The restriction map and genomic DNA clones of the DHFR-TS gene. (a) Restriction sites shown are A, *Bam*HI; C, *Bcl*I; B, *Bgl*II; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; L, *Sal*I; S, *Spe*I; T, *Sst*I; X, *Xba*I. The *Dra*I site (D) was determined by DNA sequencing. (b) Enlarged restriction 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. 1a) 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 ATATATATATTTTATTATTTATATTTCTCCTTTATGAGGGAAACAAGTCTGCACGTTTCGATATTATGCCATATGTCATGTTGAAGGTTGAAAGCAAAAT 120
 MetMetGluGlnValCysAspValPheAspIleTyrAlaIleCysAlaCysCysLysValGluSerLysAsn
 30 GAGGGAAAAAAATGAGTTTAATAACTACACATTAGGGCTAGGAAAAGAGTATTACCATGAAATGATAATTCCCTAGATGAAAATTTTGCGATGTTGAAGGTTGAAAGCAAAAT 240
 GluGlyLysAsnGluValPheAsnAsnTyrThrPheArgGlyLeuGlyAsnLysGlyValLeuProTrpLysCysAsnSerLeuAspMetLysTyrPheCysAlaValThrThrTyr
 70 ValAsnGluSerLysTyrGluLysLeuLysTyrLysArgCysLysTyrLeuAsnLysGluThrValAspAsnValAsnAspMetProAsnSerLysLeuGlnAsnValValMet 360
 GTGAATGAATC AAAAATGAGAAAATGAGATGATAATTTAACAAAAGAAACTGTGGAATAGTAAATGCTAATTCTAAAAAATTACAAAAGATTGTTGACTACAT 240
 GlyArgThrAsnTrpGluSerIleProLysLysPheLysProLeuSerAsnArgIleAsnValIleLeuSerArgThrLeuLysGluAspPheAspGluAspValTyrIlelleAsn 480
 110 150 190 230 270 310 350 390 430 470 510 550 590 630 670 710 750 790 830 870 910 950 990 1030 1070 1110 1150 1190 1230 1270 1310 1350 1390 1430 1470 1510 1550 1590 1630 1670 1710 1750 1790 1830 1870 1910 1950 1990 2030 2070 2110 2150 2190 2230 2270 2310 2350 2390 2430 2470 2510 2550 2590 2630 2670 2710 2750 2790 2830 2870 2910 2950 2990 3030 3070 3110 3150 3190 3230 3270 3310 3350 3390 3430 3470 3510 3550 3590 3630 3670 3710 3750 3790 3830 3870 3910 3950 3990 4030 4070 4110 4150 4190 4230 4270 4310 4350 4390 4430 4470 4510 4550 4590 4630 4670 4710 4750 4790 4830 4870 4910 4950 4990 5030 5070 5110 5150 5190 5230 5270 5310 5350 5390 5430 5470 5510 5550 5590 5630 5670 5710 5750 5790 5830 5870 5910 5950 5990 6030 6070 6110 6150 6190 6230 6270 6310 6350 6390 6430 6470 6510 6550 6590 6630 6670 6710 6750 6790 6830 6870 6910 6950 6990 7030 7070 7110 7150 7190 7230 7270 7310 7350 7390 7430 7470 7510 7550 7590 7630 7670 7710 7750 7790 7830 7870 7910 7950 7990 8030 8070 8110 8150 8190 8230 8270 8310 8350 8390 8430 8470 8510 8550 8590 8630 8670 8710 8750 8790 8830 8870 8910 8950 8990 9030 9070 9110 9150 9190 9230 9270 9310 9350 9390 9430 9470 9510 9550 9590 9630 9670 9710 9750 9790 9830 9870 9910 9950 9990 10030 10070 10110 10150 10190 10230 10270 10310 10350 10390 10430 10470 10510 10550 10590 10630 10670 10710 10750 10790 10830 10870 10910 10950 10990 11030 11070 11110 11150 11190 11230 11270 11310 11350 11390 11430 11470 11510 11550 11590 11630 11670 11710 11750 11790 11830 11870 11910 11950 11990 12030 12070 12110 12150 12190 12230 12270 12310 12350 12390 12430 12470 12510 12550 12590 12630 12670 12710 12750 12790 12830 12870 12910 12950 12990 13030 13070 13110 13150 13190 13230 13270 13310 13350 13390 13430 13470 13510 13550 13590 13630 13670 13710 13750 13790 13830 13870 13910 13950 13990 14030 14070 14110 14150 14190 14230 14270 14310 14350 14390 14430 14470 14510 14550 14590 14630 14670 14710 14750 14790 14830 14870 14910 14950 14990 15030 15070 15110 15150 15190 15230 15270 15310 15350 15390 15430 15470 15510 15550 15590 15630 15670 15710 15750 15790 15830 15870 15910 15950 15990 16030 16070 16110 16150 16190 16230 16270 16310 16350 16390 16430 16470 16510 16550 16590 16630 16670 16710 16750 16790 16830 16870 16910 16950 16990 17030 17070 17110 17150 17190 17230 17270 17310 17350 17390 17430 17470 17510 17550 17590 17630 17670 17710 17750 17790 17830 17870 17910 17950 17990 18030 18070 18110 18150 18190 18230 18270 18310 18350 18390 18430 18470 18510 18550 18590 18630 18670 18710 18750 18790 18830 18870 18910 18950 18990 19030 19070 19110 19150 19190 19230 19270 19310 19350 19390 19430 19470 19510 19550 19590 19630 19670 19710 19750 19790 19830 19870 19910 19950 19990 20030 20070 20110 20150 20190 20230 20270 20310 20350 20390 20430 20470 20510 20550 20590 20630 20670 20710 20750 20790 20830 20870 20910 20950 20990 21030 21070 21110 21150 21190 21230 21270 21310 21350 21390 21430 21470 21510 21550 21590 21630 21670 21710 21750 21790 21830 21870 21910 21950 21990 22030 22070 22110 22150 22190 22230 22270 22310 22350 22390 22430 22470 22510 22550 22590 22630 22670 22710 22750 22790 22830 22870 22910 22950 22990 23030 23070 23110 23150 23190 23230 23270 23310 23350 23390 23430 23470 23510 23550 23590 23630 23670 23710 23750 23790 23830 23870 23910 23950 23990 24030 24070 24110 24150 24190 24230 24270 24310 24350 24390 24430 24470 24510 24550 24590 24630 24670 24710 24750 24790 24830 24870 24910 24950 24990 25030 25070 25110 25150 25190 25230 25270 25310 25350 25390 25430 25470 25510 25550 25590 25630 25670 25710 25750 25790 25830 25870 25910 25950 25990 26030 26070 26110 26150 26190 26230 26270 26310 26350 26390 26430 26470 26510 26550 26590 26630 26670 26710 26750 26790 26830 26870 26910 26950 26990 27030 27070 27110 27150 27190 27230 27270 27310 27350 27390 27430 27470 27510 27550 27590 27630 27670 27710 27750 27790 27830 27870 27910 27950 27990 28030 28070 28110 28150 28190 28230 28270 28310 28350 28390 28430 28470 28510 28550 28590 28630 28670 28710 28750 28790 28830 28870 28910 28950 28990 29030 29070 29110 29150 29190 29230 29270 29310 29350 29390 29430 29470 29510 29550 29590 29630 29670 29710 29750 29790 29830 29870 29910 29950 29990 30030 30070 30110 30150 30190 30230 30270 30310 30350 30390 30430 30470 30510 30550 30590 30630 30670 30710 30750 30790 30830 30870 30910 30950 30990 31030 31070 31110 31150 31190 31230 31270 31310 31350 31390 31430 31470 31510 31550 31590 31630 31670 31710 31750 31790 31830 31870 31910 31950 31990 32030 32070 32110 32150 32190 32230 32270 32310 32350 32390 32430 32470 32510 32550 32590 32630 32670 32710 32750 32790 32830 32870 32910 32950 32990 33030 33070 33110 33150 33190 33230 33270 33310 33350 33390 33430 33470 33510 33550 33590 33630 33670 33710 33750 33790 33830 33870 33910 33950 33990 34030 34070 34110 34150 34190 34230 34270 34310 34350 34390 34430 34470 34510 34550 34590 34630 34670 34710 34750 34790 34830 34870 34910 34950 34990 35030 35070 35110 35150 35190 35230 35270 35310 35350 35390 35430 35470 35510 35550 35590 35630 35670 35710 35750 35790 35830 35870 35910 35950 35990 36030 36070 36110 36150 36190 36230 36270 36310 36350 36390 36430 36470 36510 36550 36590 36630 36670 36710 36750 36790 36830 36870 36910 36950 36990 37030 37070 37110 37150 37190 37230 37270 37310 37350 37390 37430 37470 37510 37550 37590 37630 37670 37710 37750 37790 37830 37870 37910 37950 37990 38030 38070 38110 38150 38190 38230 38270 38310 38350 38390 38430 38470 38510 38550 38590 38630 38670 38710 38750 38790 38830 38870 38910 38950 38990 39030 39070 39110 39150 39190 39230 39270 39310 39350 39390 39430 39470 39510 39550 39590 39630 39670 39710 39750 39790 39830 39870 39910 39950 39990 40030 40070 40110 40150 40190 40230 40270 40310 40350 40390 40430 40470 40510 40550 40590 40630 40670 40710 40750 40790 40830 40870 40910 40950 40990 41030 41070 41110 41150 41190 41230 41270 41310 41350 41390 41430 41470 41510 41550 41590 41630 41670 41710 41750 41790 41830 41870 41910 41950 41990 42030 42070 42110 42150 42190 42230 42270 42310 42350 42390 42430 42470 42510 42550 42590 42630 42670 42710 42750 42790 42830 42870 42910 42950 42990 43030 43070 43110 43150 43190 43230 43270 43310 43350 43390 43430 43470 43510 43550 43590 43630 43670 43710 43750 43790 43830 43870 43910 43950 43990 44030 44070 44110 44150 44190 44230 44270 44310 44350 44390 44430 44470 44510 44550 44590 44630 44670 44710 44750 44790 44830 44870 44910 44950 44990 45030 45070 45110 45150 45190 45230 45270 45310 45350 45390 45430 45470 45510 45550 45590 45630 45670 45710 45750 45790 45830 45870 45910 45950 45990 46030 46070 46110 46150 46190 46230 46270 46310 46350 46390 46430 46470 46510 46550 46590 46630 46670 46710 46750 46790 46830 46870 46910 46950 46990 47030 47070 47110 47150 47190 47230 47270 47310 47350 47390 47430 47470 47510 47550 47590 47630 47670 47710 47750 47790 47830 47870 47910 47950 47990 48030 48070 48110 48150 48190 48230 48270 48310 48350 48390 48430 48470 48510 48550 48590 48630 48670 48710 48750 48790 48830 48870 48910 48950 48990 49030 49070 49110 49150 49190 49230 49270 49310 49350 49390 49430 49470 49510 49550 49590 49630 49670 49710 49750 49790 49830 49870 49910 49950 49990 50030 50070 50110 50150 50190 50230 50270 50310 50350 50390 50430 50470 50510 50550 50590 50630 50670 50710 50750 50790 50830 50870 50910 50950 50990 51030 51070 51110 51150 51190 51230 51270 51310 51350 51390 51430 51470 51510 51550 51590 51630 51670 51710 51750 51790 51830 51870 51910 51950 51990 52030 52070 52110 52150 52190 52230 52270 52310 52350 52390 52430 52470 52510 52550 52590 52630 52670 52710 52750 52790 52830 52870 52910 52950 52990 53030 53070 53110 53150 53190 53230 53270 53310 53350 53390 53430 53470 53510 53550 53590 53630 53670 53710 53750 53790 53830 53870 53910 53950 53990 54030 54070 54110 54150 54190 54230 54270 54310 54350 54390 54430 54470 54510 54550 54590 54630 54670 54710 54750 54790 54830 54870 54910 54950 54990 55030 55070 55110 55150 55190 55230 55270 55310 55350 55390 55430 55470 55510 55550 55590 55630 55670 55710 55750 55790 55830 55870 55910 55950 55990 56030 56070 56110 56150 56190 56230 56270 56310 56350 56390 56430 56470 56510 56550 56590 56630 56670 56710 56750 56790 56830 56870 56910 56950 56990 57030 57070 57110 57150 57190 57230 57270 57310 57350 57390 57430 57470 57510 57550 57590 57630 57670 57710 57750 57790 57830 57870 57910 57950 57990 58030 58070 58110 58150 58190 58230 58270 58310 58350 58390 58430 58470 58510 58550 58590 58630 58670 58710 58750 58790 58830 58870 58910 58950 58990 59030 59070 59110 59150 59190 59230 59270 59310 59350 59390 59430 59470 59510 59550 59590 59630 59670 59710 59750 59790 59830 59870 59910 59950 59990 60030 60070 60110 60150 60190 60230 60270 60310 60350 60390 60430 60470 60510 60550 60590 60630 60670 60710 60750 60790 60830 60870 60910 60950 60990 61030 61070 61110 61150 61190 61230 61270 61310 61350 61390 61430 61470 61510 61550 61590 61630 61670 61710 61750 61790 61830 61870 61910 61950 61990 62030 62070 62110 62150 62190 62230 62270 62310 62350 62390 62430 62470 62510 62550 62590 62630 62670 62710 62750 62790 62830 62870 62910 62950 62990 63030 63070 63110 63150 63190 63230 63270 63310 63350 63390 63430 63470 63510 63550 63590 63630 63670 63710 63750 63790 63830 63870 63910 63950 63990 64030 64070 64110 64150 64190 64230 64270 64310 64350 64390 64430 64470 64510 64550 64590 64630 64670 64710 64750 64790 64830 64870 64910 64950 64990 65030 65070 65110 65150 65190 65230 65270 65310 65350 65390 65430 65470 65510 65550 65590 65630 65670 65710 65750 65790 65830 65870 65910 65950 65990 66030 66070 66110 66150 66190 66230 66270 66310 66350 66390 66430 66470 66510 66550 66590 66630 66670 66710 66750 66790 66830 66870 66910 66950 66990 67030 67070 67110 67150 67190 67230 67270 67310 67350 67390 67430 67470 67510 67550 67590 67630 67670 67710 67750 67790 67830 67870 67910 67950 67990 68030 68070 68110 68150 68190 68230 68270 68310 68350 68390 68430 68470 68510 68550 68590 68630 68670 68710 68750 68790 68830 68870 68910 68950 68990 69030 69070 69110 69150 69190 69230 69270 69310 69350 69390 69430 69470 69510 69550 69590 69630 69670 69710 69750 69790 69830 69870 69910 69950 69990 70030 70070 70110 70150 70190 70230 70270 70310 70350 70390 70430 70470 70510 70550 70590 70630 70670 70710 70750 70790 70830 70870 70910 70950 70990 71030 71070 71110 71150 71190 71230 71270 71310 71350 71390 71430 71470 71510 71550 71590 71630 71670 71710 71750 71790 71830 71870 71910 71950 71990 72030 72070 72110 72150 72190 72230 72270 72310 72350 72390 72430 72470 72510 72550 72590 72630 72670 72710 72750 72790 72830 72870 72910 72950 72990 73030 73070 73110 73150 73190 73230 73270 73310 73350 73390 73430 73470 73510 73550 73590 73630 73670 73710 73750 73790 73830 73870 73910 73950 73990 74030 74070 74110 74150 74190 74230 74270 74310 74350 74390 74430 74470 74510 74550 74590 74630 74670 74710 74750 74790 74830 74870 74910 74950 74990 75030 75070 75110 75150 75190 75230 75270 75310 75350 75390 75430 75470 75510 75550 75590 75630 75670 75710 75750 75790 75830 75870 75910 75950 75990 76030 76070 76110 76150 76190 76230 76270 76310 76350 76390 76430 76470 76510 76550 76590 76630 76670 76710 76750 76790 76830 76870 76910 76950 76990 77030 77070 77110 77150 77190 77230 77270 77310 77350 77390 77430 77470 77510 77550 77590 77630 77670 77710 77750 77790 77830 77870 77910 77950 77990 78030 78070 78110 78150 78190 78230 78270 78310 78350 78390 78430 78470 78510 78550 78590 78630 78670 78710 78750 78790 78830 78870 78910 78950 78990 79030 79070 79110 79150 79190 79230 79270 79310 79350 79390 79430 79470 79510 79550 79590 79630 79670 79710 79750 79790 79830 79870 79910 79950 79990 80030 80070 80110 80150 80190 80230 80270 80310 80350 80390 80430 80470 80510 80550 80590 80630 80670 80710 80750 80790 80830 80870 80910 80950 80990 81030 81070 81110 81150 81190 81230 8127

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 polyglutamate 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 λ gt11 *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 λ gt11 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

<u>Dihydrofolate Reductase</u>												
PF	MMEQVCD	VFDIYAI	CAC	KVES	KNEG	KKNEV	NNY	TER	G	EN	KGVLE
LM	MSRAAARFK	KIP	MPM	PET	KAD	FAP	PSL	RAF	SIV	V	ALD	MQH
H
EC	TSS
LC
T4
	100	110	120	130	140	150	160	170	180			
PF	SKKLQN	VVV	M	CRT	NWES	I	P	K	F	K	S	N
LM	TEKKR	NAV	V	M	R	K	T	W	E	V	P	V
H	VEGK	QLV	M	G	K	T	W	E	S	P	G	V
EC	...NPKV	IMGR	HRT	WES	IG...	RPL	KPR	NIL	VLS	...REL	KPQG...	...
LC	...GKIM	VV	GRRT	Y	ES	FPP	.KR	H	PER	TVN	VLT	...
T4	...GT	IM	MAK	TF	SLP	.T	U	P	SH	IVVC
	190	200	210	220	230	240	250	260	270	280		
PF	TRIN	STYE	CDV	FF	PEI	NE	YQ	I	S
LM	TRYI	AT	AP	AT	REF	P	P	PA	NA	TAW	LASS	QRRK
H	TRIM	QDF	ESDT	..F	PEI	DLK	Y	K	V	PE	PG	VLS
EC	THIDAE	VE	GDT	.H	FPD	YEP	DW	ES	VFS	EFH	DAQNSH
LC	TRLAGS	FEG	DT	K	MI	PLN	WDD	FT	KVSS	RT	VEDTN
T4	TKSV	VIGG	PA	L	LYA	ALP	YADE	VVVS	RIV	KRHR
	290	300	310	320	330	340						
<u>Thymidylate Synthase</u>												
PF	EEDDF	VYF	NFN	KEE	KN	NSI	H	P	NDF	Q	I	YNSL
LM
H	MPV	AGSEL	P	R	P	L	P	P	A	Q	R	AA
M	MLV	V	SEL
HVS	MS	T	H	T	E	Q	H	E	H	N	Y	CFK
VZV	MGDL	S	CWT	K	V	P	G	T	L	Q	V	DK
EC	MKQ	Q	ELM	K	Q	V	L	D	E	G	N	Q
LC	M	LEQ	P	Y	L	D	A	K	V	D	E	Y
T4	M	KQ	Q	D	L	I	K	F	E	G	Y	E
ϕ 3T	MTQFD	K	Q	I	S	I	I	D	I	N	G	S
	390	400	410	420								
PF	WFIRGET	.NGNT	LNKN	VR	I	WE	ANG	TRE	FLD	NR	KLFHREV
LM	WFIRGET	.SA	QI	ADKD	I	H	WDGNS	REF	DLR	SGL	TENKE
H	WF	IKGST	.NA	KE	SSKG	V	K	WD	ANG	R	DFLDSLGFSTREE
M	WF	IKGST	.NA	KE	SSKG	VR	I	WD	ANG	R	DFLDSLGP SARQE
HVS	WF	IRGST	.DS	KE	LAAG	V	H	WD	ANG	R	FLDKLGFYDRDE
VZV	WF	IRGST	.DS	KE	LAAG	I	H	I	YGS	KFLN	RNGFHKRHT
EC	WF	LOGDT	.NIA	Y	LEHNN	V	T	I	DE	WADE	...
LC	WF	LHGDT	.N	IRE	LLHR	NH	I	WD	EFA	KWVKS	DEYHGPDMDTDFGHR
T4	WF	LSG	ST	NVN	DLR	I	QHDSL	I	QGK	T	W	DENYQAKDLYGHS
ϕ 3T	WI	QWLKS	NDV	T	LNKMG	V	H	WD	QKD	E	...
	450	460	470	480	490	500	510	520	530	540	550	
PF	ENKG	V	DQ	LN	I	N	D	S	T	R	I	YMDNY
LM	DGE	CV	DQ	I	K	L	V	E	T	K	L	MDG
H	SGQ	CV	DQ	LR	Q	V	I	D	T	S	Q	VY
M	SGQ	CV	DQ	LR	Q	V	I	D	T	S	Q	VY
HVS	KGE	CV	DQ	LN	Q	I	T	K	N	TC	Y	RDMSDY
VZV	LQQG	I	DQ	LN	Q	I	T	K	N	TC	Y	GDQSY
EC	DGRH	DQ	IT	T	V	N	Q	N	TC	Y	GDQSY	WRFGA
LC	KGDT	I	DQ	DN	Q	I	T	K	N	TC	Y	GDQSY
T4	...G	DQ	I	I	E	V	I	E	I	I	I	WRF
ϕ 3T	NGEKV	DQ	Y	D	V	I	N	Q	N	TC	Y	GDQSY
	560	570	580	590	600							
PF	IDSLK	I	QLM	R	IP	P	T	K	LN	PDI	K	NMD
LM	VDA	KAQ	LE	R	V	H	A	P	T	I	Y	DNH
H	IEPLK	I	QLM	R	E	P	P	K	I	R	K	V
M	IEPLK	I	QLM	R	E	P	P	K	I	R	K	V
HVS	IDALK	M	Q	L	T	P	P	T	R	A	N	Y
VZV	IDALK	V	Q	L	T	P	P	T	R	A	N	Y
EC	MDO	THL	Q	L	S	P	R	P	T	R	A	N
LC	LDO	I	KEQ	L	S	T	P	R	P	T	R	A
T4	VEQ	CKE	I	LR	R	E	P	K	E	R	I	N
ϕ 3T	IDNL	KI	QM	R	E	P	A	E	P	V	T	N

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 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 β A– β H) found in all forms of DHFR except *L. tropica* (11) were predictable in the *P. falciparum* enzyme.

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.

falciparum DHFR domain (amino acids 1–228). The β -strands β A– β F, 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 α E' 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 \approx 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 (β i– β v), 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 2. Numbers of amino acids shared by different TSs

	PF	LM	H	M	HVS	VZV	EC	LC	T4	ϕ 3T
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
ϕ 3T										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. Banyai, 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.